



**T.C.
İSTANBUL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**



DOKTORA TEZİ

**DİYABETİK AYAK HASTALARININ AMPÜTASYON RİSKİNİN
YAPAY ZEKÂ TEKNİKLERİ İLE ÖNGÖRÜLMESİ**

Denizhan DEMİRKOL

Enformatik Anabilim Dalı

Enformatik Programı

**DANIŞMAN
Doç. Dr. Çiğdem EROL**

Nisan, 2022

İSTANBUL

Bu çalışma, 25.04.2022 tarihinde ařağıdaki jüri tarafından Enformatik Anabilim Dalı, Enformatik Programında Doktora tezi olarak kabul edilmiştir.

Tez Jürisi

Doç. Dr. Çiğdem EROL(Danışman)
İstanbul Üniversitesi
Fakülte

Prof. Dr. Şamil AKTAŞ
İstanbul ÜNİVERSİTESİ
Tıp Fakültesi

Prof. Dr. Sevinç GÜLSEÇEN
İstanbul Üniversitesi
Enformatik Bölümü

Doç. Dr. Tuncay ÖZCAN
İstanbul Teknik Üniversitesi
İşletme Fakültesi

Doç. Dr. Aşkın DEMİRAG
Yeditepe Üniversitesi
Ticari Bilimler Fakültesi

20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince; Bu Lisansüstü teze, İstanbul Üniversitesi’nin aboneli olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü’nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.

Bu tez, 2214/A (Yurtdışı Doktora Sırası Araştırma Bursu) kapsamında **TÜRKİYE BİLİMSEL VE TEKNOLOJİK ARAŞTIRMA KURUMU (TÜBİTAK)** tarafından **1059B142000477** numaralı proje ile desteklenmiştir.

Bu tez, **FRANSA TÜRKİYE Büyükelçiliği - Institute Français Turquie ve Campus France** tarafından tahsis edilen Doktora Öğrencileri için **2022 Araştırma Bursları** kapsamında desteklenmiştir.

ÖNSÖZ

Bu doktora tezinin önsözü alışılmış doktora tezlerinin önsözlerinden biraz farklı olmasını istiyorum. Elbette her doktora tezi yoğun emek ve çalışma sonuunda hazırlanmıştır fakat bu tez çalışması sıra dışı bir emek, gayret ve fedakârlıkla oluşturulmuş bir çalışmadır. Bu çalışmanın gerçekleştirilmesinde en başta Danışman Hocam Sayın **Doç. Dr. Çiğdem SELÇUKCAN EROL'un** araştırmamıza olan inancı olağanüstü anlayışı, destekleri, adeta bir anne gibi sarıp sarmalaması ve doktora eğitimi kadar zorlu bir süreci kolaylaştırması yatmaktadır. Bu Önsöz vesilesi ile kendisine şükranlarımı sunuyorum. Bununla birlikte yine bölüm başkanımız Sayın **Prof. Dr. Sevinç GÜLSEÇEN** Hocama doktora sürecine başladığım ilk günden sonuna kadarki anlayışı, destekleri ve her zaman yanımda olduğu için sonsuz teşekkürlerimi sunuyorum. Bilindiği üzere bilimsel bir araştırmanın ve doktora tezi gibi bilim insanının kariyerindeki mihenk taşı niteliğindeki bir çalışmanın ve araştırmanın lâıyık veçhile gerçekleştirilmesi, eldeki donelere yani verilere bağlıdır. İstanbul Üniversitesi Tıp Fakültesi Su Altı Hekimliği ve Hiperbarik Tıp Anabilim Dalı'nda öğretim üyesi olan Sayın **Prof. Dr. Şamil AKTAŞ'a** tezimin her sürecindeki katkıları, araştırmanın başarı ile gerçekleştirileceğine dair olan inancı ve de verinin toplanması sürecinden yaptığımız bütün çalışmalarda sağlamış olduğu tüm imkanlar için bütün kalbimle teşekkür ediyorum. Elbette veri toplama sürecinde gerek mesai saatlerinde gerekse de gece nöbetlerinde bile bana hiçbir karşılık beklemeden yardım eden Uzman Dr. Özdiñ ACARLI'ya da bu vesile ile teşekkür etmek isterim. Ayrıca doktora tez izleme komitemde yer alan **Sayın Doç.Dr. Tuncay ÖZCAN'a** her zamanki desteğı için tekrar teşekkür etmek istiyorum. Yeditepe Üniversitesi Yönetim Bilişim Sistemleri Bölüm Başkanı **Doç Dr. Aşkın Demirağ** hocama da 2008 yılından beri akademik kariyerimin her aşamasında yanımda olduğu için en içten şükranlarımı sunarım.

Bu doktora tez çalışması yoğun araştırma faaliyetleri, önüme konan çeşitli engellere rağmen bitmek bilmeyen araştırma hevesi, mükemmellik tutkusu sonucunda ortaya çıkmış bir tez çalışmasıdır. **Fransa-Sorbonne Université- LIMICS laboratuvarında** sekiz ay süre ile araştırma faaliyetlerinde bulunurken kıymetli bilgi birikimini paylaşan **Prof. Xavier TANNIER (Ph.D.) ve Directrice Prof. Marie-Christine JAULENT (Ph.D.)'e** ve bana 8 ay süre ile bu imkânı sağlayan, araştırmamıza fon desteğinde bulunan **TÜRKİYE Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK)'a** minnettarım.

Ve... FRANSA TÜRKİYE Büyükelçiliğı'nin 2022 Araştırma Bursları Kapsamında TÜRKİYE çapında yalnızca "**8**" (**sekiz**) Doktora Öğrencisine layık görülen Araştırma Bursunu tarafıma tahsis eden **FRANSA Hükümetine, Fransa Türkiye Büyükelçiliğine ve Institute Français Turquie** komitesine ve de **Campus France Paris** ekibine en içten teşekkürlerimi ve şükranlarımı sunuyorum. Elbette ki bu olağanüstü prestijli dış kaynaklı hükümet bursuna layık görüldüğüm için onur duyuyorum ve bir ömür bu haklı gururu taşıyacağımı beyan etmek istiyorum.

Özellikle belirtmek isterim ki Fransa'da bulunduğum süre boyunca ilk günden araştırmamın sonuna kadar desteklerini esirgemeyen ve beni kendi evlatları/kardeşlerinden ayırmayan yüce gönüllü **Ayla NALLAR** Teyzeme ve canım **Müge NALLAR BOUILLON** ablacığma şükranlarımı özellikle belirtmek istiyorum...

Bununla birlikte Enformatik Bölümünde tanımaktan mutluluk duyduğum bu süreçte manevi, desteğini gördüğüm arkadaşım Dr. Öğretim Üyesi Fatma Önay KOÇOĞLU'na en özel teşekkürlerimi sunmak istiyorum. Ayrıca yine gerek Enformatik bölümünde gerekse de Fransa'da bulunduğum süre boyunca hep sıcak ve dost canlısı tutumlarını ve desteklerini gördüğüm Dr. Öğretim Üyesi Zeki Ozen 'e de en içten teşekkürlerimi arz ederim. Dahası, Enformatik Bölümünde bulunduğum sürece hep iyiliğini gördüğüm Dr. Fidan BAYRAKTUTAN'a da teşekkür etmek isterim. Enformatik Bölümünde geçirdiğim süre boyunca hep desteğini gördüğüm Dr. Öğretim Üyesi Zerrin AYVAZ REİS'e özel teşekkürlerimi de iletmek istiyorum. Dahası, Dr. Murat GEZER'e ve Dr. Öğretim Üyesi İnci ZAİM GÖKBAY'a da şahsıma her zaman göstermiş oldukları nazik ve sıcak tutumları için teşekkür ediyorum. Bununla birlikte yine tanımaktan mutluluk duyduğum Doç.Dr. Emre Akadal'a ve Dr. Serra ÇELİK'e de bu vesile ile teşekkür etmek istiyorum. Ayrıca, Enformatik Bölümünde tanıdığım tüm akademik ve idari personele ve başta Nesim Şükrü YUMLU olmak üzere Sibel GÜVELİ'ye de tez sürecimdeki göstermiş oldukları iyi niyet ve sabırları için en içten teşekkürlerimi sunuyorum. **İstanbul Üniversitesi Fen Bilimleri Enstitüsünde bulunan tüm akademik ve idari personele** de böylesine zorlu bir süreçteki destekleri için bu vesile ile teşekkür ederim.

Akademik camiayı biraz da kortikosteriod ilaçlar gibi iki ucu keskin kılıca benzetmişimdir. Bıçak sırtında hiç sekmeden ve fire vermeden yürümek araştırmacının göstereceği fedakârlıklara bağlıdır. Benimle bu fedakârlıkları göstermekten imtina etmeyen ve en önemlisi bu doktora tezini adadığım canım annem **Nurten GÜVENÇ DEMİRKOL** ve değerli babam **Mehmet SİRAC DEMİRKOL**'a teşekkür etmek için söyleyeceğim her kelime kifayetsiz kalıyor...

Nisan 2022

Denizhan DEMİRKOL

İÇİNDEKİLER

Sayfa No

ÖNSÖZ	i
İÇİNDEKİLER.....	iii
ŞEKİL LİSTESİ	ix
TABLO LİSTESİ.....	xiii
SİMGE VE KISALTMA LİSTESİ	xxi
ÖZET	xxvii
SUMMARY	xxix
1. GİRİŞ	1
2. GENEL KISIMLAR.....	4
2.1 DIABETES MELLITUS	4
2.2.1 Tanım.....	4
2.2.2 Epidemiyoloji	4
2.1.3 Teşhis.....	5
2.1.4 Sınıflandırma	6
2.1.5 Komplikasyonlar	7
2.1.5.1 Akut Komplikasyonlar	7
2.1.5.2 Kronik Komplikasyonlar	7
2.1.5.3 Diyabetik Nöropati.....	8
2.1.5.4 Diyabetik Ayak Yaraları.....	8
2.2 DİYABETİK AYAK SINIFLANDIRMA SİSTEMLERİ	9
2.2.3 Meggit Wagner Sınıflandırması	9
2.2.2 Teksas San Antonio Üniversitesi Diyabetik Ayak Sınıflandırması	10
2.2.4 S(AD) SAD Sınıflandırması	10
2.2.4 PEDIS Sınıflandırması	11
2.3 DİYABETİK AYAK YARALARININ TEDAVİSİNDE KULLANILAN YÖNTEMLER	13
2.3.1 Cerrahi Tedavi	13
2.3.2 Yükten Kurtarma	13
2.3.3 Antibiyoterapi.....	14
2.3.4 Negatif Basıncılı Yara Örtüsü	14

2.3.5	Hiperbarik Oksijen Tedavisi.....	14
2.3.6	Büyüme Faktörleri.....	15
2.4	YAPAY ZEKÂ.....	16
2.4.1	Tanım.....	16
2.4.2	Veri Tabanlarında Bilgi Keşfi	16
2.4.2.1	Seçim- Verinin Toplanması.....	17
2.4.2.2	Veri Ön İşleme	17
2.4.2.3	Veri Entegrasyonu	18
2.4.2.4	Veri Dönüştürme.....	19
2.4.2.5	Veri İndirgeme	20
2.4.2.6	Veri Madenciliği Sürecinde Yapay Zekâ- Makine Öğrenmesi Algoritmalarının Kullanılması.....	22
2.5	MAKİNE ÖĞRENMESİ ALGORİTMALARI.....	23
2.5.1	Danışmanlı Öğrenme	23
2.5.2	K en Yakın Komşu Algoritması	26
2.5.3	Karar Ağaçları	28
2.5.4	Rastgele Orman Algoritması (“Random Forest Algorithm”).....	32
2.5.5	Basit-Sade Bayes Algoritması.....	33
2.5.6	Lojistik Regresyon.....	35
2.5.7	Destek Vektör Makineleri	39
2.5.8	Yapay Sinir Ağları.....	46
2.5.9	Aşırı Öğrenme Makineleri.....	56
2.5.10	Topluluk Öğrenme Yöntemleri	58
2.5.11	Extreme Gradient Boosting (“Xgboost”) Algoritması	63
2.5.12	Danışmansız Öğrenme.....	65
2.5.12.1	Kümeleme	65
2.5.12.2	Birliktelik Kuralları- “Association Rules”	66
2.5.13	Pekiştirmeli Öğrenme	66
2.6	MAKİNE ÖĞRENMESİNDE HİPERPARAMETRE OPTİMİZASYON TEKNİKLERİ	66
2.6.1	Algoritmaların Hiperparametre Arama Uzayı.....	68
2.6.2	Model Kullanmayan Algoritmalar	69
2.6.2.1	Izgara Araması	69
2.6.3	Meta Sezgisel Optimizasyon Algoritmaları	70
2.6.3.1	Genetik Algoritma.....	71

2.6.3.2 Parçacık Sürü Optimizasyonu Algoritması	73
2.7 YORUMLAMA.....	76
2.7.1 Model performans değerlendirme yöntemleri	76
2.7.2 Model performans değerlendirme ölçütleri	79
2.8 DİYABETİK AYAĞIN AMPÜTASYON RİSKİNİN ÖNGÖRÜLMESİNE YÖNELİK LİTERATÜR TARAMASI	84
3 MALZEME VE YÖNTEM.....	116
3.1 PROBLEM VE TEZ ÇALIŞMASININ AMACI.....	116
3.2 DOKTORA TEZ ÇALIŞMASININ ÖZGÜNLÜĞÜ.....	116
3.3 DİYABETİK AYAK VERİ SETİNİN OLUŞTURULMASI VE VERİNİN TOPLANMASI	117
3.3.1 Orijinal Veri Seti	118
3.4 VERİ ÖN İŞLEME.....	120
3.4.1 Orijinal Veri setinin Özellikleri.....	124
3.4.2 Veri Dönüştürme	128
3.4.2.1 Nümerik Veri Seti.....	128
3.4.2.2 Temel Bileşenler Analizi Sonuçları	130
3.4.2.3 Kategorik Veri Seti.....	134
3.5 VERİ MADENCİLİĞİ- YAPAY ZEKÂ – MAKİNE ÖĞRENMESİ MODELLERİNİN KURULMASI	137
3.5.1 Orijinal Veri Setinin Alt Kırımları ile Birlikte Elde Edilen Alt Kırılım Veri Setleri ve Uygulanan Algoritmalar	137
3.6 MODEL PERFORMANS DEĞERLENDİRME YÖNTEMLERİ & ÖLÇÜTLERİ.....	138
3.7 KULLANILAN ARAÇLAR	140
4 BULGULAR.....	142
4.1 VERİYİ ANLAMA VE GERÇEKLEŞTİRİLEN İSTATİSTİKSEL ANALİZLER.....	142
4.2 BOX-PLOT GRAFİĞİ VE NORMALLİK TESTLERİ.....	147
4.3 SPEARMAN RHO KORELASYON ANALİZİ TESTİ SONUÇLARI	157
4.4 VERİ MADENCİLİĞİ SÜRECİNDE MAKİNE ÖĞRENMESİ SINIFLANDIRMA ALGORİTMALARININ UYGULANMASI SONUCU ELDE EDİLEN BULGULAR – İKİLİ SINIFLANDIRMA	158
4.4.1 K- En Yakın Komşu Algoritmasından Elde Edilen Bulgular (İkili Sınıflandırma).....	159
4.4.1.1 Orijinal Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular	159

4.4.1.2 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular	160
4.4.2 Sade Bayes Sınıflandırıcısından Elde Edilen Bulgular (İkili Sınıflandırma)	161
4.4.2.1 Orijinal Veri Setinden Edilen Bulgular	161
4.4.2.2 Orijinal Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinden Elde Edilen Bulgular	162
4.4.2.3 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılımından Elde Edilen Bulgular	163
4.4.2.4 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinden Elde Edilen Bulgular	164
4.4.3 Sınıflandırma ve Regresyon Ağaçları (“CART”) Algoritmasından Elde Edilen Bulgular (İkili Sınıflandırma)	165
4.4.3.1 Orijinal Veri Setinden Edilen Bulgular	165
4.4.3.2 Orijinal Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinden Elde Edilen Bulgular	167
4.4.3.3 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılımından Elde Edilen Bulgular	168
4.4.3.4 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinden Elde Edilen Bulgular	170
4.4.4 Rastgele Orman Algoritması Elde Edilen Bulgular (İkili Sınıflandırma)	170
4.4.4.1 Orijinal Veri Setinden Elde Edilen Bulgular	170
4.4.4.2 Orijinal Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinden Elde Edilen Bulgular	172
4.4.4.3 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılımından Elde Edilen Bulgular	173
4.4.4.4 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinden Elde Edilen Bulgular	176
4.4.5 XGBOOST Algoritmasından Elde Edilen Bulgular (İkili Sınıflandırma)	177
4.4.5.1 Orijinal Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular	177
4.4.5.2 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular	178
4.4.6 Yapay Sinir Ağları Algoritmasından Elde Edilen Bulgular (İkili Sınıflandırma)	179
4.4.6.1 Orijinal Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular	179
4.4.6.2 Çok Katmanlı Yapay Sinir Ağları Parametrelerinin Genetik Algoritma Kullanılarak Optimize Edilmesi	180

4.4.6.3. Çok Katmanlı Yapay Sinir Ağları Parametrelerinin Parçacık Sürü Optimizasyonu Kullanılarak Optimize Edilmesi.....	181
4.4.6.4 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular	182
4.4.6.4 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılım Veri Setinde Çok Katmanlı Yapay Sinir Ağları Parametrelerinin Genetik Algoritma Kullanılarak Optimize Edilmesi	184
4.4.6.5 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılım Veri Setinde Çok Katmanlı Yapay Sinir Ağları Parametrelerinin Parçacık Sürü Optimizasyonu Kullanılarak Optimize Edilmesi.....	185
4.4.7 Aşırı Öğrenme Makineleri Algoritmasından Elde Edilen Bulgular (İkili Sınıflandırma)	186
4.4.7.1 Orijinal Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular	186
4.4.7.2 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular	187
4.4.8 Destek Vektör Makineleri Algoritmasından Elde Edilen Bulgular (İkili Sınıflandırma)	189
4.4.8.1 Orijinal Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular	189
4.4.8.2 DVM Algoritmasının Hiperparametrelerinin Genetik Algoritma Kullanılarak Optimize Edilmesi.....	190
4.4.8.3 DVM Algoritmasının Hiperparametrelerinin Parçacık Sürü Optimizasyonu Kullanılarak Optimize Edilmesi.....	191
4.4.8.4 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular	191
4.4.8.5 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılımında DVM Algoritmasının Hiperparametrelerinin Genetik Algoritma Kullanılarak Optimize Edilmesi.....	192
4.4.8.6 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılımında DVM Algoritmasının Hiperparametrelerinin Parçacık Sürü Optimizasyonu Kullanılarak Optimize Edilmesi.....	193
4.4.9 Lojistik Regresyon Analizinden Elde Edilen Bulgular (İkili –“Binary” Sınıflandırma)	194
4.4.9.1 Orijinal Veri Setinden Edilen Bulgular.....	194
4.4.9.2 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılımından Elde Edilen Bulgular.....	195
4.4.10 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılım Veri Setinde En iyi Performansı Gösteren İkili Ayırma ve Üçlü Ayırma Performans Değerlendirme Yöntemleri ile Elde Edilen Performans Değerlendirme Ölçütleri (İkili Sınıflandırma)	200

4.5 VERİ MADENCİLİĞİ SÜRECİNDE MAKİNE ÖĞRENMESİ SINIFLANDIRMA ALGORİTMALARININ UYGULANMASI SONUCU ELDE EDİLEN BULGULAR – ÇOK KATEGORİLİ SINIFLANDIRMA.....	202
4.5.1 Orijinal Veri Setinden Elde Edilen Bulgular	202
4.5.1.1 Orijinal Veri Setinin %80'ini Oluşturan “Eğitim” Veri Setinde Veri Seti Dengeleme İşlemi Gerçekleştirilmeden Elde Edilen Bulgular	202
4.5.1.2 Orijinal Veri Setinin %80'ini Oluşturan “Eğitim” Veri Setinde Veri Setinin “Oversampling” Metodu ile Dengeleme İşlemi Gerçekleştirdikten Sonra Elde Edilen Bulgular	205
4.5.2 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Niteliklerden Oluşan Alt Kırılımından Elde Edilen Bulgular.....	208
4.5.2.1 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Niteliklerden Oluşan Alt Kırılımının %80'ini Oluşturan “Eğitim” Veri Setinde, <u>Veri Seti Dengeleme İşlemi Gerçekleştirilmeden Elde Edilen Bulgular</u>	208
4.5.2.2 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Niteliklerden Oluşan Alt Kırılımının %80'ini Oluşturan “Eğitim” Veri Setinde, <u>Veri Seti Dengeleme İşlemi Gerçekleştirildikten Sonra Elde Edilen Bulgular</u>	211
4.5.3 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Ayırıklaştırma Prosedürü Sonucunda Tamamen Kategorik Niteliklerden Oluşan Alt Kırılımından Elde Edilen Bulgular.....	214
4.5.3.1 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Ayırıklaştırma Prosedürü Sonucunda Tamamen Kategorik Niteliklerden Oluşan Alt Kırılımının %80'ini Oluşturan “Eğitim” Veri Setinde, <u>Veri Seti Dengeleme İşlemi Gerçekleştirilmeden Elde Edilen Bulgular</u>	214
4.5.3.2 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Ayırıklaştırma Prosedürü Sonucunda Tamamen Kategorik Niteliklerden Oluşan Alt Kırılımının %80'ini Oluşturan “Eğitim” Veri Setinde, <u>Veri Seti Dengeleme Prosedürü Gerçekleştirildikten Sonra Elde Edilen Bulgular</u>	217
4.6 DİYABETİK AYAK HASTALARININ MAJÖR AMPÜTASYON RİSKİNİN “ONE VERSUS REST” YAKLAŞIMI KULLANILARAK ÖNGÖRÜLMESİ	220
4.7 GELİŞTİRİLEN MODELLERİN R SHINY KULLANICI ARA YÜZÜNE ENTEGRE EDİLMESİ	226
5 TARTIŞMA VE SONUÇ	228
6. KAYNAKLAR	234
7. EKLER	249
EK-1: ETİK KURUL ONAY BELGESİ	249
EK-2: ARAŞTIRMADA KULLANILAN DİYABETİK AYAK İZLEM FORMU	250
8. ÖZGEÇMİŞ	253

ŞEKİL LİSTESİ

Sayfa No

Şekil 2. 1: İstanbul Üniversitesi İstanbul Tıp Fakültesi Sualtı Hekimliği ve Hiperbarik Tıp Anabilim Dalı'nda tedavi için başvuran bir diyabetik ayak hastasına ait ve parmak ucunda sınırlı gangren bulguları olan bir diyabetik ayak fotoğrafı	9
Şekil 2. 2: Meggit- Wagner Diyabetik Ayak Sınıflandırması (Meggit,1976; Wagner, 1981).....	10
Şekil 2. 3: Teksas San Antonio Üniversitesi Diyabetik Ayak Sınıflandırması (Lawrence ve diğ. 1996; Mutluoğlu, 2013).....	10
Şekil 2. 4: S(AD) SAD Sınıflandırması (Macfarlane ve Jeffcoate, 1999; Mutluoğlu, 2013).....	11
Şekil 2. 5: PEDIS Sınıflandırması (Shaper, 2004; Mutluoğlu, 2013).....	12
Şekil 2. 6: Veri tabanlarında Bilgi Keşfi (Fayyad ve diğ. 1996; Akpınar, 2014)	17
Şekil 2. 7 : Makine Öğrenmesi Yöntemleri- Kagiya ve diğ. (2019)	23
Şekil 2. 8: Danışmanlı öğrenmede iş akış süreci Mahesh (2020)	24
Şekil 2. 9: Danışmanlı öğrenmede iş akış süreci (Ayodele, 2010)	25
Şekil 2.10: Rastgele Orman Algoritması'nın şematik gösterimi (Liu ve diğ., 2012)	32
Şekil 2. 11: Birden fazla doğrusal sınıflandırıcı -hiperdüzlem “hyperplanes” verileri ayrılabilir (Olson ve Delen, 2008).	40
Şekil 2. 12: Hiperdüzlemlerin maksimum ayrımı (Olson ve Delen, 2008)	41
Şekil 2. 13: Doğrusal olmayan DVM için haritalama (Lee ve Kim, 2010)	44
Şekil 2. 14: Biyolojik bir sinir hücresinin yapısı Olson ve Delen (2008)	47
Şekil 2. 15: Yapay bir sinir hücresi ve elemanları (Haykin, 2009).....	47
Şekil 2. 16: Yumuşak geçişli Sigmoid aktivasyon fonksiyonu (Harrington, 2012).....	49
Şekil 2. 17: Sert geçişli “Step function” aktivasyon fonksiyonu (Harrington, 2012).....	50
Şekil 2. 18: Hiperbolik tanjant aktivasyon fonksiyonu (Namin ve diğ., 2009)	50
Şekil 2. 19: Tek ve çok katmanlı ileri beslemeli YSA mimarileri (Haykin, 2009).....	52
Şekil 2. 20: Geri Beslemeli YSA Mimarisi (Johnson, 2022).....	53

Şekil 2. 21: Gizli veya çıktı katmanı birimi: Ağırlıklı toplam oluşturmak için girdiler karşılık gelen ağırlıklarıyla çarpılır ve net girdiye de doğrusal olmayan bir aktivasyon fonksiyonu uygulanır (Han ve diğ., 2012)	55
Şekil 2.22: Torbalama- “ <i>Bagging</i> ” Topluluk Öğrenme Metodunun şekilsel olarak gösterimi (Nagauri, 2020)	59
Şekil 2.23: Artırma- “ <i>Boosting</i> ” Topluluk Öğrenme Yöntemi’nin şekilsel olarak gösterimi (Patel, 2020)	60
Şekil 2.24: Artırma- “ <i>Boosting</i> ” Topluluk Öğrenme Yöntemi’nin şekilsel olarak gösterimi (Patel, 2020)	61
Şekil 2.25: İstifleme – “ <i>Stacking</i> ” Topluluk Öğrenme Yönteminin şekilsel gösterimi (Ceballos, 2020)	62
Şekil 2. 26: Çoğunluk Oylaması – “ <i>Majority Voting</i> ” şekilsel gösterimi (Raschka, 2020).....	63
Şekil 2.27: Model parametreleri ve hiperparametreler arasındaki farkın şekilsel gösterimi (Agrawal, 2021).....	67
Şekil 2.28: Izgara Arama yöntemi kullanılarak hiperparametre optimizasyonunun gerçekleştirilmesi (Syarif ve diğ, 2016)	70
Şekil 2. 29: Evrimsel Algoritma kullanılarak hiper-parametre optimizasyonu süreci (Syarif ve diğ., 2016).....	72
Şekil 2. 30: PSO'daki temel hız güncelleme mekanizmasının şekilsel gösterimi Kiranyaz ve diğerleri (2014)	74
Şekil 2.31: PSO algoritması için Bağış (2009) tarafından oluşturulmuş akış diyagramı.....	75
Şekil 2. 32: Üçlü Ayırma – “three way split” yönteminin şekilsel gösterimi	78
Şekil 2. 33: K- kat çapraz geçerleme (Olson ve Delen, 2008).....	78
Şekil 2. 34: Makine Öğrenmesi Sınıflandırma Algoritmalarının Performansını Değerlendirmek Amacıyla Kullanılan Performans Değerlendirme Ölçütlerinin oluşturulmasında Yararlanılan Karışıklık Matrisi	80
Şekil 2. 35: ROC uzayı (Japkowicz ve Shah, 2011)	83
Şekil 3. 1: Orijinal Diyabetik Ayak Veri Setinin R Proglama Dilinde Alınmış Özeti	122
Şekil 3. 2: Orijinal Diyabetik Ayak Veri Setinin R Proglama Dilinde Alınmış Özeti (Devamı).....	123

Şekil 3. 3: Orijinal Diyabetik Ayak Veri Setinin R Proglama Dilinde Alınmış Özeti (Devamı).....	124
Şekil 3. 4: “knnimputation” fonksiyonu kullanılarak eksik verilerin doldurulmasından sonra veri setinin R dilindeki özeti	126
Şekil 3. 5 “knnimputation” fonksiyonu kullanılarak eksik verilerin doldurulmasından sonra veri setinin R dilindeki özeti (devam).....	126
Şekil 3.6: Tüm nümerik niteliklerin min-max normalizasyon yöntemi ile normalizasyon işleminde sonra veri setinin R dilindeki özeti	127
Şekil 3.7: Tüm nümerik niteliklerin min-max normalizasyon yöntemi ile normalizasyon işleminde sonra veri setinin R dilindeki özeti (devam)	128
Şekil 3.8: Tüm nümerik niteliklerin min-max normalizasyon yöntemi ile normalizasyon işleminde sonra veri setinin R dilindeki özeti (devam)	128
Şekil 3. 9: Tüm niteliklerin nümerik olduğu veri setinin R dilindeki özeti	129
Şekil 3.10: Tüm niteliklerin nümerik olduğu veri setinin R dilindeki özeti (devam)	129
Şekil 3. 11: Tüm niteliklerin nümerik olduğu veri setinin R dilindeki özeti (devam)	130
Şekil 3. 12: Tüm niteliklerin nümerik olduğu veri setinin R dilindeki özeti (devam)	130
Şekil 3. 13: Değişkenlerin PCA sonucunda şekilsel gösterimi	131
Şekil 3. 14: Değişkenlerin PCA sonucunda şekilsel gösterimi (renklendirilmiş versiyon)	132
Şekil 3. 15: PCA sonucunda ortaya çıkan “ <i>scree-plot</i> ” grafiği	132
Şekil 3. 16: PCA sonucunda ortaya çıkan değişkenlerin ortalama katkısını gösteren grafik.....	133
Şekil 3. 17: Tüm boyutlarda değişkenler cos2 değerlerinin görselleştirilmesi	134
Şekil 3. 18: Kategorik veri setinin R dilinde alınmış özeti-1	135
Şekil 3. 19: Kategorik veri setinin R dilinde alınmış özeti-2 (devam)	136
Şekil 3. 20: Kategorik veri setinin R dilinde alınmış özeti-3 (Devam).....	136
Şekil 3. 21: Kategorik veri setinin R dilinde alınmış özeti-4 (Devam).....	137
Şekil 3. 22: Hedef niteliğin ikili “ <i>binary</i> ” olduğu durum için oluşturulan veri setleri ve kullanılan algoritmalar.....	137
Şekil 3. 23: Hedef niteliğin hastaların majör ve minör amputasyon statülerini de içeren “ <i>multi-class</i> ” veri seti ve kullanılan algoritmalar	138
Şekil 4. 1: Hastaların Genel Özelliklerine ait Değerlerin Box-Plot Grafiği	147

Şekil 4. 2: Hastaların Genel Özelliklerinin Logaritmik Dönüşümlü Değerlerine ait Box-Plot Grafiği	147
Şekil 4. 3: Nümerik ve Kategorik Değişkenlerden Oluşan Orijinal Veri Setinde Sınıflandırma ve Regresyon Ağaçları Algoritmasından Elde Edilen “Budama Yapılmış” Karar Ağacı Yapısı.....	166
Şekil 4. 4: Rastgele Orman Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılım Veri Setinin %80’i Kullanılarak Oluşturulmuş Modelin Grafıksel Gösterimi Çıktısı	175
Şekil 4. 5: Izgara arama ile gerçekleştirilen parametre optimizasyonunda ROC değerlerine göre değişen “ <i>decay</i> ” ve “ <i>hidden units</i> ” değerleri.....	179
Şekil 4. 6: Izgara arama ile gerçekleştirilen parametre optimizasyonunda ROC değerlerine göre değişen “ <i>decay</i> ” ve “ <i>hidden units</i> ” değerleri.....	183
Şekil 4. 7: Lojistik Regresyon Analizi Modelinin R Programlama Dilinde Alınmış Ekran Görüntüsü.....	197
Şekil 4. 8: Linear Model vs. Lojistik Model	198
Şekil 4. 9: Lojistik Regresyon Modelinden Elde Edilen ROC Eğrisi	199
Şekil 4. 10: Diyabetik Ayak Hastalarının Genel – “ <i>Overall</i> ” Ampütasyon Riskinin Tahmini için Geliştirilen R-Shiny Uygulaması ve Örnek Değerlerin Girilmesi ile Verilen Tahmin Çıktısı	227

TABLO LİSTESİ

	Sayfa No
Tablo 2. 1: Diyabet Tanı Kriterleri (Ulusal Diyabet Konsensus Grubu, 2019).....	6
Tablo 2. 2: Diyabetin Sınıflandırılması	6
Tablo 2.3: Literatürde sık kullanılan ve en bilinen γ, r, τ ve d çekirdek parametrelerini içeren çekirdek fonksiyonları (Scholkopf ve Smola, 2001; Öz, 2019)	45
Tablo 2. 4: Birleştirme (Toplama) Fonksiyonları.....	48
Tablo 2.5: Hiperparametre optimizasyon yapılabilen makine öğrenmesi algoritmaları, hiperparametre tipi ve arama uzayı (Yang ve Shami, 2020)	68
Tablo 3. 1: Orijinal veri setinde bulunan nitelikler.....	118
Tablo 3. 2: Orijinal veri setinde bulunan nitelikler (devam)	119
Tablo 3. 3: Orijinal veri setinde bulunan nitelikler (devam)	120
Tablo 3. 4: Doktora tez çalışmasında R programlama dilinde kullanılan paket ve kütüphaneler	140
Tablo 4. 1: Diyabetik Ayak Hastalarına ait Sosyo-Demografik ve Genel Özelliklere İlişkin Bilgiler.....	143
Tablo 4. 2: Hastalara Ait Sosyo-Demografik ve Genel Özelliklere Ait Dağılım.....	144
Tablo 4. 3: Hastalara Ait Sosyo-Demografik ve Genel Özelliklere Ait Dağılım.....	145
Tablo 4. 4: Hastaların Genel Özelliklerine ait Değerlerin Tanımlayıcı İstatistiksel Analizi	146
Tablo 4. 5: Hastaların Genel Özelliklerine Ait Normallik Testi Analizleri	148
Tablo 4. 6: Hastaların Genel Özelliklerin Logaritmik Dönüşümlü Değerlerine Ait Normallik Dağılım Analizleri.....	149
Tablo 4. 7: Hastaların Genel Özellikleri (Kategorik Değişkenler) ile Karar Aşaması Arasındaki İlişki Analizi.....	150
Tablo 4. 8: Hastaların Genel Özellikleri (Kategorik Değişkenler) ile Ampütasyon Statüsü arasındaki İlişki Analizi	151

Tablo 4. 9: Hastaların Ampütasyon Statülerine göre Genel Özelliklerinin (Nümerik Değişkenler) Karşılaştırılması	152
Tablo 4. 10: Hastaların Minör ve Majör Ampütasyon (Üçlü Kategorik - “ <i>Multi Class</i> ”) Statülerine Göre Genel Özelliklerinin (Nümerik Değişkenler) Karşılaştırılması	154
Tablo 4. 11: Hastaların Genel Özellikleri ile Genel “ <i>Overall</i> ” Ampütasyon Statüsü Arasındaki İlişkinin Değerlendirmesinde Spearman Rho Korelasyon Analizi	157
Tablo 4. 12: Orijinal Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular	159
Tablo 4. 13: K-En Yakın Komşu Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular	160
Tablo 4. 14: Sade Bayes Algoritması için Orijinal Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	161
Tablo 4. 15: Sade Bayes Algoritması için Orijinal Veri Setinde Çapraz Geçerleme Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	162
Tablo 4. 16: Sade Bayes Algoritması için Orijinal Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	162
Tablo 4. 17: Sade Bayes Algoritması için Orijinal Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinde Çapraz Geçerleme Performans Değerlendirme Yöntemi Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	163
Tablo 4. 18: Sade Bayes Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Değerlerden Oluşan Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	163
Tablo 4. 19: Sade Bayes Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Değerlerden Oluşan Alt Kırılım Veri Setinde Çapraz Geçerleme Performans Değerlendirme Yöntemi Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	164

Tablo 4. 20: Sade Bayes Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinden Elde Edilen Bulgular	164
Tablo 4. 21: Sınıflandırma ve Regresyon Ağaçları (“CART”) Algoritması için Orijinal Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	165
Tablo 4. 22: Sınıflandırma ve Regresyon Ağaçları (“CART”) için Orijinal Veri Setinde Çapraz Geçerleme Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	167
Tablo 4. 23: Sınıflandırma ve Regresyon Ağaçları (“CART”) Algoritması için Orijinal Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinin Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	167
Tablo 4. 24: Sınıflandırma ve Regresyon Ağaçları (“CART”) için Orijinal Veri Setinde Çapraz Geçerleme Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	168
Tablo 4. 25: Sınıflandırma ve Regresyon Ağaçları (“CART”) Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılımının Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	168
Tablo 4. 26: Sınıflandırma ve Regresyon Ağaçları (“CART”) Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılımının Çapraz Geçerleme Performans Değerlendirme Yöntemi Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	169
Tablo 4. 27: Sınıflandırma ve Regresyon Ağaçları (“CART”) Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış ve Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinin Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	170
Tablo 4. 28: Rastgele Orman Algoritması için Orijinal Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	170

Tablo 4 29: Rastgele Orman Algoritması için Orijinal Veri Setinde Çapraz Geçerleme Performans Değerlendirme Yöntemi Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	172
Tablo 4. 30: Rastgele Orman Algoritması için Orijinal Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	172
Tablo 4. 31: Rastgele Orman Algoritması için Orijinal Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinde Çapraz Geçerleme Performans Değerlendirme Yöntemi Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	173
Tablo 4. 32: Rastgele Orman Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	173
Tablo 4. 33: Rastgele Orman Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılım Veri Setinde Çapraz Geçerleme Performans Değerlendirme Yöntemi Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	175
Tablo 4. 34: Rastgele Orman Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	176
Tablo 4. 35: XGBOOST Algoritması için Orijinal Veri Setinin Nümerik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	177
Tablo 4. 36: XGBOOST Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	178
Tablo 4. 37: Yapay Sinir Ağları Algoritması için Orijinal Veri Setinin Nümerik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	180

Tablo 4. 38: Çok Katmanlı Yapay Sinir Ağları Parametrelerinin Genetik Algoritma Kullanılarak Optimize Edilmesi Sonucunda Elde Edilen Performans Değerlendirme Ölçütleri	181
Tablo 4. 39: Çok Katmanlı Yapay Sinir Ağları Parametrelerinin Parçacık Sürü Optimizasyonu Kullanılarak Optimize Edilmesi Sonucunda Elde Edilen Performans Değerlendirme Ölçütleri	181
Tablo 4. 40: Yapay Sinir Ağları Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri.....	182
Tablo 4. 41: Çok Katmanlı Yapay Sinir Ağları Parametrelerinin Genetik Algoritma Kullanılarak Optimize Edilmesi Sonucunda Elde Edilen Performans Değerlendirme Ölçütleri.....	184
Tablo 4. 42: Çok Katmanlı Yapay Sinir Ağları Parametrelerinin Parçacık Sürü Optimizasyonu Kullanılarak Optimize Edilmesi Sonucunda Elde Edilen Performans Değerlendirme Ölçütleri	185
Tablo 4. 43: Aşırı Öğrenme Makineleri Algoritması için Orijinal Veri Setinin Nümerik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri.....	187
Tablo 4. 44: Aşırı Öğrenme Makineleri Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri.....	188
Tablo 4. 45: Destek Vektör Makineleri Algoritması için Orijinal Veri Setinin Nümerik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri.....	189
Tablo 4. 46: Destek Vektör Makineleri Algoritmasının Genetik Algoritma Kullanılarak Optimize Edilmesi Sonucunda Elde Edilen Performans Değerlendirme Ölçütleri.....	190
Tablo 4. 47: Destek Vektör Makineleri Algoritmasının Parçacık Sürü Optimizasyonu Kullanılarak Optimize Edilmesi Sonucunda Elde Edilen Performans Değerlendirme Ölçütleri.....	191

Tablo 4. 48: DVM Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	192
Tablo 4. 49: Destek Vektör Makineleri Algoritmasının, Genetik Algoritma Kullanılarak Optimize Edilmesi Sonucunda Elde Edilen Performans Değerlendirme Ölçütleri.....	193
Tablo 4. 50: Destek Vektör Makineleri Algoritmasının, Parçacık Sürü Optimizasyonu Kullanılarak Optimize Edilmesi Sonucunda Elde Edilen Performans Değerlendirme Ölçütleri.....	194
Tablo 4.51: Lojistik Regresyon için Orijinal Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri.....	194
Tablo 4. 52: Lojistik Regresyon Algoritması için Orijinal Veri Setinde Çapraz Geçerleme Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri	195
Tablo 4. 53: Lojistik Regresyon Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri.....	195
Tablo 4. 54: Lojistik Regresyon analizi sonucunda elde edilmiş olan Nagelkerke R Square değeri	199
Tablo 4. 55: Lojistik Regresyon Analizi -Elde Edilmiş olan Hosmer and Lemeshow Test Sonucu	199
Tablo 4. 56: Lojistik Regresyon Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılım Veri Setinde Çapraz Geçerleme Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri.....	200
Tablo 4 57: Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılım Veri Setinde En iyi Performansı Gösteren İkili Ayırma ve Üçlü Ayırma Performans Değerlendirme Yöntemleri ile Elde Edilen Doğruluk ve F-Ölçütü Performans Değerlendirme Ölçütleri	201
Tablo 4. 58: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Minör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen	

Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları	203
Tablo 4. 59: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları	204
Tablo 4. 60: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Minör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları	205
Tablo 4. 61: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları	207
Tablo 4. 62: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Minör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları	209
Tablo 4. 63: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları	210
Tablo 4. 64: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Minör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları	211
Tablo 4. 64: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Minör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları (devam)	212
Tablo 4. 65: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen	

Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları	213
Tablo 4. 66: Hedef Niteliğin 3 Kategoriden Oluştugu (“Multi-Class”) Diyabetik Ayak Hastalarının “Minör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (“Three-Way Split”) Performans Sonuçları	215
Tablo 4. 67: Hedef Niteliğin 3 Kategoriden Oluştugu (“Multi-Class”) Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (“Three-Way Split”) Performans Sonuçları	216
Tablo 4. 68: Hedef Niteliğin 3 Kategoriden Oluştugu (“Multi-Class”) Diyabetik Ayak Hastalarının “Minör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (“Three-Way Split”) Performans Sonuçları	218
Tablo 4. 69: Hedef Niteliğin 3 Kategoriden Oluştugu (“Multi-Class”) Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (“Three-Way Split”) Performans Sonuçları	219
Tablo 4. 70: Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskinin “One-versus-All” metodolojisi, Eğitim Veri Setinde Veri Seti Dengeleme İşlemi Yapılarak ve Üçlü Ayırma (%80 Eğitim, %10 Validasyon,%10 Test) Performans Değerlendirme Yöntemi Kullanılarak Öngörülmesi	221
Tablo 4. 71: Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskinin “One-versus-All” metodolojisi, Eğitim Veri Setinde Veri Seti Dengeleme İşlemi Yapılarak ve Üçlü Ayırma (%70 Eğitim, %20 Validasyon,%10 Test) Performans Değerlendirme Yöntemi Kullanılarak Öngörülmesi	223
Tablo 4. 72: Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskinin “One-versus-All” metodolojisi, Eğitim Veri Setinde Veri Seti Dengeleme İşlemi Yapılarak ve Üçlü Ayırma (%60 Eğitim,%20 Validasyon,%20 Test) Performans Değerlendirme Yöntemi Kullanılarak Öngörülmesi	224

SİMGE VE KISALTMA LİSTESİ

Simgeler	Açıklama
a_i	: Lagrange Çarpanları
b	: Offset Parametresi
B_k	: K Adet Sınıftan Oluşan Sınıflar Kümesi
β	: Beta Katsayısı
C	: Ceza Parametresi
c_i	: Hızlanma (Öğrenme) Sabiti
Err_j	: Çıktı Katmanındaki Bir Birim j için Hata
$f(x)$	: Ateşleme Fonksiyonu
H	: Gizli Katman Matrisi
H^+	: H Matrisinin <i>Moore-Penrose</i> Genelleştirilmiş Ters
I_j	: j Birimindeki Net Girdi
j	: Gizli Katman Birimi
$K()$	: Kernel Fonksiyonu
k	: Komşu Sayısı
$L()$	: Lagrange Fonksiyonu
l	: Öğrenme Oranı
$\log(x)$	: Logit Fonksiyonu
n	: Kişi/Hasta Sayısı
Obj	: Modeli Değerlendirmek için Kullanılan Fonksiyon
O_i	: Önceki Katmandan Gelen i Biriminin Çıktısı
O_j	: j'nin Gerçek Çıktısı
θ_j	: Birimin Etkinliğini Değiştiren Bir Eşik Değer Olarak Görev Yapan Önyargı (" <i>Bias</i> ") Değeri
$\emptyset$	: Haritalama Fonksiyonu
$P(x)$	: Olasılık Fonksiyonu

P	: Olasılık Deęeri
r	: Spearman Rho Korelasyon Katsayısı
sim	: Benzerlik Fonksiyonu
T_j	: Verilen eğitim örneęinin bilinen sınıf etiketine dayalı olan gerçek çıktı
t	: Bağımsız örneklem T Testi
T	: Çıktı Matrisi
tanh(x)	: Hiperbolik Tanjant Tipi Aktivasyon Fonksiyonu
t_i	: Veritabanındaki Bir Kayıt
μ	: Lineer Tahmin Edici
w	: Atalet Ağırlığı
λ	: Düzeltme Parametresi
ξ	: Aylak Deęişken
$\bar{X}$	: Ortalama
$x_{aj}(t)$	: a Parçacığının t Zamanındaki Konumunun j. Boyutlu Bileşeni
χ^2	: Ki-Kare
w	: Ağırlık Vektörü
ω_{ij}	: Önceki Katmandaki i Biriminden j Birimine Olan Bağlantının Ağırlığı
$v_{aj}(t)$	: a Parçacığının t Zamanındaki hızının j. Boyutlu Bileşeni
$y_{aj}(t)$	: a Parçacığının t Anında En İyi (<i>pbest</i>) Konumunun j. Boyutlu Bileşeni
$\hat{y}_j(t)$	: t Zamanında Sürünün “ <i>global</i> ” En İyi Konumunun j. Boyutlu Bileşeni
γ	: Düzeltme Parametresi

Kısaltmalar	Açıklama
-------------	----------

ABI	: “Ankle Brachial Index”
ACC	: Doğruluk
ACE	: Anjiyotensin Dönüştürücü Enzim
APVS	: Ayak Bileęi Tepe Sistolik Hızı

ATA	: “1 atmosfer absolute”
AUC	: ROC Eğrisi Altında Kalan Alan
BMI	: Vücut Kitle İndeksi (VKI)
CART	: Sınıflandırma ve Regresyon Ağaçları
CFI	: Kritik Ayak İskemisi
CI	: Confidence Interval (Güven Aralığı)
CN	: Charcot Nöroartropatisi
CNN	: Convolutional Neural Network
CREDIT	: Türkiye Kronik Böbrek Hastalığı Prevalansı Araştırması (Chronic Renal Disease in Turkey)
CRP	: C-Reaktif Protein
CSI	: Standartlaştırılmış Hibrit Görüntü Tabanlı Puanlama Sistemi
CT	: Bilgisayarlı Tomografi
CWD	: Sürekli Doppler Ölçümleri
D.M	: Diabetes Mellitus
Df	: Serbestlik Derecesi
DFI	: Diyabetik Ayak Enfeksiyonları
dn	: Doğru Negatif
DOR	: Tanısal üstünlük oranı
dp	: Doğru Pozitif
DPP4-İ	: Dipeptidil Peptidaz 4 İnhibitörleri, Gliptinler
DVM	: Destek Vektör Makineleri
EGF	: Epidermal Büyüme Faktörü
EGF-IL	: Rekombinant İnsan Epidermal Büyüme Faktörü
eGFR	: Tahmini Glomerüler Filtrasyon Hızı
ELM	: Aşırı Öğrenme Makineleri
ERR	: Hata Oranı
ESR	: Eritrosit Sedimentasyon Hızı
EVT	: Endovasküler Tedavi
F	: F Ölçütü
FGF	: Fibroblast Kaynaklı Büyüme Faktörü
FNR	: Yanlış Negatif Oranı
FPR	: Yanlış Pozitif Oranı

GA	: Genetik Algoritma
GBYSA	: Geri Beslemeli Yapay Sinir Ağları
GDM	: Gestasyonel Diabetes Mellitus
GLP-1	: Glukagon Benzeri Peptid-1 Reseptör Agonistleri
HBA1C	: Glikozillenmiş Hemoglobin
HBOT	: Hiperbarik Oksijen Tedavisi
HDL	: Yüksek Yoğunluklu Lipoprotein
HT	: Hipertansiyon Varlığı
ID3	: “ <i>Iterative Dichotomiser 3</i> ”
IWGDF	: Uluslararası Diyabetik Ayak Çalışma Grubu
KAH	: Koroner Arter Hastalığı
kNN	: K En Yakın Komşu Algoritması,
LDL	: Düşük Yoğunluklu Lipoprotein
LEA	: Hastada Alt Ekstremitte Ampütasyonu
LP-	: Negatif Olabilirlik Oranı
LP+	: Pozitif Olabilirlik Oranı
MAC	: Medial Arteriyel Kalsinoz
MDRO	: Çoklu İlaça Dirençli Organizmalar
MRI	: Manyetik Rezonans Görüntüleme
MRSA	: Metisiline Dirençli Staphylococcus Aureus
MWT	: Nemli Yara Bakım Tedavisi
NDF	: Diyabetik Olmayan Ayak Ülseri Grubu
NLO	: Nötrofil-Lenfosit Oranı
NPV	: Negatif Öngörü Değeri
NPWT	: Negatif Basıncılı Yara Örtüsü Tedavisi
OAD	: Oral Antidiyabetik İlaçlar
OGTT	: Oral Glukoz Tolerans Testi
OR	: Olasılık Oranı (Odds Ratio)
PAH	: Periferik Arter Hastalığı
PCA	: Temel Bileşenler Analizi
PDG	: Trombosit Kaynaklı Büyüme Faktörü
PEDIS	: Perfüzyon (<i>Perfusion</i>), Boyut (<i>Extent</i>), Derinlik (<i>Depth</i>), İnfeksiyon (<i>Infection</i>) ve Duyu (<i>Sensation</i>)

PLR	: Trombosit Lenfosit Oranı
PLT	: Trombosit
PNL	: Nötrofil
PPV	: Pozitif Öngörü Değeri
PPV	: Pozitif Öngörü Değeri
PSO	: Parçacık Sürü Optimizasyonu
PTA	: Perkütan Translüminal Anjiyoplasti İşlemleri
PWAR	: Yara Alanı Küçülme Yüzdesi
QP	: Karesel Programlama
RF	: Rastgele Orman Algoritması
ROC	: Alıcı İşlem Karakteristik Eğtisi
S(AD) SAD	: (<i>Size</i>): yara boyutu, (<i>Area</i>): alan, (<i>Dept</i>): derinlik, (<i>Sepsis</i>): enfeksiyon şiddeti, (<i>Arteriopathy</i>): arteriyel dolaşım, (<i>Denervation</i>)
S.D.	: Standart Sapma
SCI	: <i>Science Citation Index</i>
SCI-Exp	: <i>Science Citation Index-Expanded</i>
SD	: Standart Sapma
SLFNs	: Tek Gizli Katmanlı Aşırı Öğrenme Makineleri
SPC	: Belirleyicilik
SPCD	: Superpixel Renk Tanımlayıcıları
SPECT	: Tek Foton Emisyonlu Bilgisayarlı Tomografi
SPP	: Deri Perfüzyon Basıncı
SSCI	: “ <i>Social Sciences Citation Index</i> ”
SVM	: “ <i>Support Vector Machines</i> ”
SWME	: Semmes-Weinstein Monofilaman İncelemesi
TBI	: “ <i>Toe Brachial Index</i> ”
TcPO₂	: Transkutanöz Oksijen
TEKHARF	: “ <i>Turkish Adult Risk Factor Survey</i> ”
TEMĐ	: Türkiye Endokrinoloji ve Metabolizma Derneđi
TEMĐ	: Diabetes Mellitus Çalışma ve Eğitim Grubu
TPR	: Duyarlılık
TURDEP	: Türkiye Diyabet, Hipertansiyon, Obezite ve Endokrinolojik Hastalıklar Prevalans Çalışması

TWC	: “ <i>University of Texas</i> ” Sınıflandırma Sistemi
UHMS	: Sualtı - Hiperbarik Tıp Topluluğu
UT	: Teksas Üniversitesi
VPT	: Yapılandırılmış Ayak Muayenesi ve Biyoteziyometre
VTBK	: Veri Tabanlarında Bilgi Keşfi
WBC	: Lökosit – “ <i>White Blood Cell</i> ”
WHI	: Yara İyileşme İndeksi
WHO	: Dünya Sağlık Örgütü
XGBOOST	: “ <i>Extreme Gradient Boosting</i> ” Algoritması
yn	: Yanlış Negatif
yp	: Yanlış Pozitif
YSA	: Yapay Sinir Ağları
3GCR-EC	: Üçüncü Nesil Sefalosporine Dirençli Enterobacteriaceae

ÖZET

DOKTORA TEZİ

Diyabetik Ayak Hastalarının Ampütasyon Riskinin Yapay Zekâ Teknikleri ile Öngörülmesi

Denizhan DEMİRKOL

İstanbul Üniversitesi

Fen Bilimleri Enstitüsü

Enformatik Anabilim Dalı

Danışman : Doç. Dr. Çiğdem EROL

Diyabetik ayak sendromuna bağlı gelişen diyabetik ayak ülserleri, yüksek morbidite ve mortalite ile karakterize olup hastalar için amputasyon gibi hayati sonuçları olabilen diyabetin en önemli komplikasyonları arasında bulunmaktadır. Bu doktora tez çalışmasının temel amacı diyabetik ayak hastalarının amputasyon riskinin yapay zekâ-makine öğrenmesi sınıflandırma algoritmaları kullanılarak öngörülmesidir. Çalışmada kullanılan veri seti, İstanbul Üniversitesi Tıp Fakültesi Sualtı Hekimliği ve Hiperbarik Tıp Anabilimdalı'nda diyabetik ayak infeksiyonu tanısıyla Ocak 2009 – Eylül 2019 tarihleri arasında tedavi görmüş ve takip süreçleri sonlanmış 407 hastayı içermektedir. Bu tez çalışmasında, diyabetik ayak hastaları retrospektif olarak değerlendirilmiş, hastaların başvuru anındaki demografik ve klinik özellikleri ile takip ve tedavi sonucu sağlanan fayda ve amputasyon ilişkisi yapay zekâ teknikleri ile açığa çıkarılmıştır. Araştırmada yöntem olarak Veri Tabanlarında Bilgi Keşfi süreci takip edilmiştir. Veri setindeki diyabetik ayak hastalığına ilişkin en önemli nitelikler Temel Bileşenler Analizi yoluyla belirlenmiştir. Bu tez çalışmasında diyabetik ayak hastalarının hem genel hem de majör ve minör amputasyon riskinin öngörülmesi için çeşitli tahmin/sınıflandırma modelleri oluşturulmuştur. Bu modellerin kurulmasında, k-En Yakın Komşu algoritması, Sade Bayes Sınıflandırıcı, Sınıflandırma ve Regresyon Ağaçları algoritması (CART), Rastgele Orman Algoritması, Yapay Sinir Ağları, Aşırı Öğrenme Makineleri, Destek Vektör Makineleri, Lojistik Regresyon ve XGBoost algoritmalarından yararlanılmıştır. Hiperparametre optimizasyonu için ayrıca Genetik Algoritma ve Parçacık Sürü Optimizasyonu algoritmaları da kullanılmıştır. Model performans değerlendirme yöntemleri olarak; çapraz geçерleme, ikili

ayırma yöntemi, “*bootstrap sampling*” ve üçlü ayırma yöntemleri kullanılmıştır. Veri seti ve alt kırılım veri setlerinde uygun makine öğrenmesi sınıflandırma algoritmalarının uygulanması hem nitelik seçimi işlemi gerçekleştirilmeden hem de gerçekleştirildikten sonra tekrarlanmıştır. Sonuç olarak nitelik seçimi yapılarak oluşturulmuş, kategorik ve nümerik değerlerden oluşan alt kırılım veri setinde daha iyi test performansı gösteren modeller; temelinde topluluk öğrenme metodolojisi bulunan “*ensemble*” bir yöntem olan “Rastgele Orman Algoritması” (doğruluk: 0,78; duyarlılık: 0,67; belirleyicilik: 0,87; kesinlik: 0,80; Negatif Öngörü Değeri: 0,77; F-Ölçütü:0,73) ve “Lojistik Regresyon” (doğruluk: 0,90; duyarlılık: 0,94; belirleyicilik: 0,86; kesinlik: 0,84; Negatif Öngörü Değeri: 0,95; F-Ölçütü: 0,89) algoritmaları olup, seçilen deney koşullarında diğer algoritmalara kıyasla daha iyi performans göstermiştir. R-Shiny kullanıcı ara yüzüne bu nedenle genel amputasyonu tahmin etmek için oluşturulmuş ve ilgili deney koşullarında en yüksek performansı gösteren bu iki model entegre edilerek diyabetik ayak tedavisi ile uğraşan hekim ve sağlık personeline yardımcı bir karar destek sistemi geliştirilmiştir.

Nisan 2022, 284 sayfa.

Anahtar kelimeler: Diyabet, Diyabetik Ayak, Amputasyon, Yapay Zekâ, Makine Öğrenmesi

SUMMARY

Ph.D. THESIS

Prediction of Amputation Risk of Patients with Diabetic Foot by Artificial Intelligence Techniques

Denizhan DEMİRKOL

İstanbul University

Institute of Graduate Studies in Sciences

Department of Informatics

Supervisor : Assoc. Prof. Dr. Çiğdem EROL (Ph.D.)

Diabetic foot ulcers due to diabetic foot syndrome are characterized by high morbidity and mortality and are among the most important complications of diabetes mellitus, which can have vital consequences such as amputation for patients. The main purpose of this doctoral thesis is to predict the amputation risk of diabetic foot patients using artificial intelligence-machine learning classification algorithms. The dataset used in the doctoral thesis includes 407 patients who were treated with the diagnosis of diabetic foot infection in Istanbul University Faculty of Medicine, Department of Underwater and Hyperbaric Medicine between January 2009 and September 2019 and their follow-up processes were terminated. In this research, diabetic foot patients were evaluated retrospectively and it is aimed to determine with artificial intelligence techniques the relationship between the demographic and clinical features of the patients at the time of admission, and the benefits of the treatment and amputation results obtained through follow-up. In the research, the process of Knowledge Discovery in Databases created by Fayyad, Shapiro and Smyth (1996) has been followed. The most important features related to the amputation risk of diabetic foot patients in the data set were determined through Principal Component Analysis. In this doctoral thesis, various prediction/classification models were created to predict the risk of both general – “overall” and major and minor amputation – “multi class” risk of diabetic foot patients. In the creating of these models, k-Nearest Neighbor algorithm, Naive Bayes Classifier, Classification and Regression Trees Algorithm (CART, Random Forest algorithm, Artificial Neural Networks, Extreme Learning Machines, Support

Vector Machines, Logistic Regression and XGBoost Algorithms were used. Genetic Algorithm and Particle Swarm Optimization algorithms were also used for hyperparameter optimization for specific algorithms. As model performance evaluation methods; cross validation, hold-out method, bootstrap sampling and three way split methods were used. The application of appropriate machine learning classification algorithms on the dataset and sub-dimension datasets was repeated both before and after the feature selection process. As a result, the models that show the best performance in the sub-dimension data set consisting of categorical and numerical values, which were created by feature selection; the "Random Forest Algorithm" (accuracy: 0,78; sensitivity: 0,67; specificity: 0,87 ; precision: 0,80 ; negative predictive value: 0,77; F-Measure:0,73) which is an based on ensemble learning methodology and "Logistic Regression" algorithms (accuracy: 0,90; sensitivity: 0,94; specificity: 0,86; precision: 0,84; negative predictive value: 0,95; F-Measure: 0,89), performed better than other algorithms in the defined experimental conditions. Therefore, a decision support system has been developed to help physicians and healthcare professionals dealing with diabetic foot treatment by integrating these two prediction models, which show the highest performance in the relevant experimental conditions to the R-Shiny user interface.

April 2022, 284 pages.

Keywords: Diabetes Mellitus, Diabetic Foot, Amputation, Artificial Intelligence, Machine Learning

1. GİRİŞ

Bilgi, yüzyıllardır karar verme açısından kilit öneme sahip olan bir kavramdır. Steward (1998) içinde yaşadığımız yeni çağda, zenginliğin bilginin ürünü olduğunu ve zenginlik yaratmak için ihtiyaç olan sermaye varlıkları olarak kabul edilen arazi, bedensel emek, imalat aletleri ve fabrikaların yerini bilginin almış olduğunu ifade etmiştir. Günümüzdeki organizasyonların büyük çoğunluğu bilgi yoğun işletmelerdir. İçinde bulunduğumuz devir de bilgi toplumu olarak literatürde yerini almıştır (Tonta ve Küçük, 2005).

Günümüzde değişmekte olan yeni dünya düzeninde, dijitalizasyon ile birlikte çeşitli bilim dallarında çalışan uzmanların bilgi birikimlerinin bilgisayar ortamına aktarılması yapay zekâ metodları ve teknikleri ile gerçekleştirilmektedir (Gülseçen ve Demirkol, 2019). Makine öğrenmesi algoritmaları yapay zekânın bir alt çalışma alanıdır. Veri madenciliğinde de makine öğrenmesi algoritmaları kullanılabilmektedir. Yapay zekânın insanlığa en çok katkısı olacak uygulamaları arasında **sağlıkta yapay zekâ uygulamaları** olduğu söylenilebilir.

Dünyada yaşlanan nüfus ile birlikte diyabet, kanser, demans gibi kronik hastalıklar artış gösterme eğilimine girmiş ve hasta ile birebir ilişkide olan hekimler, hemşireler ve hatta sağlık sektöründe tüm çalışanlar bu tip hastalarla karşı karşıya geldiklerinde hastanın durumuna adapte olmak için ciddi efor harcamaktadırlar. Bu gibi durumlarda yapay zekânın bir karar destek sistemi olarak oldukça faydalı olacağı da ifade edilmektedir (Bulut, 2019). İstatistiki yöntemlerle tespit edilmesi güç olan büyük boyutlu veri kümeleri arasından kıymetli olan bilgiyi ortaya çıkarma işi veri madenciliği süreciyle ve makine öğrenmesi algoritmaları ile gerçekleştirilebilmektedir.

Diyabet (*Diabetes mellitus*), pankreas tarafından insülin üretiminde veya kalıtsal ve/veya edinilmiş eksiklikten veya üretilen insülinin etkisizliğinden kaynaklanan kronik bir hastalıktır. Bu tür bir eksiklik kandaki glikoz konsantrasyonlarının artmasına neden olur ve bu da vücudun birçok sistemine, özellikle de kan damarlarına ve sinirlere zarar verir. Son zamanlarda derlenen veriler, yaklaşık 150 milyon insanın dünya çapında diyabete sahip olduğunu ve bu sayının 2025 yılına kadar iki katına çıkabileceğini göstermektedir. Bu artışın özellikle gelişmekte olan ülkelerde gerçekleşeceği öngörülmekte ve bunun nedenleri arasında ise nüfus artışı, yaşlanma, sağlıksız diyetler, obezite ve sedanter yaşam tarzları olarak gösterilmektedir. Diyabetin,

diyabetik retinopati, diyabetik nefropati- böbrek yetmezliği, kalp damar hastalıkları, diyabetik ayak gibi birçok komplikasyonu bulunmaktadır.

Diyabetik ayak, kan damarlarındaki ve sinirlerdeki değişikliklere bağlı olarak, sıklıkla ülserasyona ve sonradan uzuv amputasyonuna yol açmaktadır. Bu komplikasyon, diyabetin en ciddi ve maliyetli komplikasyonlarından biridir. Hem vasküler hem de nörolojik hastalık süreçlerinden kaynaklanır. Diyabet, düzenli muayene ve ayağın iyi bakımı ile önlenilecek olan alt uzuvdaki travmatik olmayan amputasyonun en yaygın nedenidir (WHO, 2010).

Bu tez çalışmasının amacı, Türkiye’de ve Dünyada prevalansı yüksek olan diyabet hastalığının en önemli komplikasyonlarından biri olan amputasyonun ve ilişkili faktörlerin yapay zekâ – makine öğrenmesi sınıflandırma algoritmaları kullanılarak önceden tahmin edilmesine dayanmaktadır. Bu amaçla yapılacak çalışmada diyabetik ayak hastalarının amputasyon risklerinin makine öğrenmesi sınıflandırma algoritmaları kullanılarak tahmin modeli oluşturmak ve bu modeli de bir kullanıcı arayüzüne entegre ederek daha geniş kitlelerce kullanma imkanının sağlanması amaçlanmıştır.

Bu tez çalışmasının GENEL KISIMLAR bölümünde öncelikle, çalışmanın iki ana başlığı olan diyabet ve ilişkili diyabetik ayak hastalığı, yapay zekâ ve ilişkili makine öğrenmesi, veri madenciliği kavramları açıklanmıştır. İlk ana başlık olarak diyabet hastalığı ve ilgili olarak tanım, tanı koymada kullanılan kriterler, bulgular, tedavi ile ilgili genel bilgiler verilmiştir. Sonrasında ise diyabetin komplikasyonları, nöropati ve diyabetik ayak infeksiyonu ile ilgili bilgi ve uygulanan tedaviler anlatılmıştır. Aynı zamanda diyabetik ayağın öngörüsü ile ilgili araştırmalara yer verilen literatür çalışması bölümün sonunda yer almaktadır. Aynı bölümde ikinci ana başlık olarak yapay zekâ ve veri madenciliği kavramı ele alınmıştır. Makine öğrenmesi algoritmaları ve veri madenciliğinin tanımı, Fayyad, Shapiro ve Smyth, (1996) tarafından oluşturulan bilgi keşfi süreç tanımları, analizlerde büyük önem taşıyan ön işleme süreci detaylı olarak ifade edilmiştir. Ardından makine öğrenmesi algoritmaları teorik olarak açıklanmış; bu çalışma kapsamında kullanılan sınıflandırma algoritmaları, model performans değerlendirme yöntemleri, model performans değerlendirme ölçütleri ile ilgili bilgiler yine bu bölümde yer almıştır.

MALZEME VE YÖNTEM bölümü, bir önceki bölümde bahsedilmiş olan veri madenciliği sürecine paralel olarak oluşturulmuş, bu süreçteki aşamalar takip edilerek analizler gerçekleştirilmiştir. Seçilen sürece paralel olarak beş başlık altında yapılan işlemlerden

bahsedilmiştir. Verinin toplanması ve veri seçimi, veri ön işleme, veri dönüştürme, veri madenciliği ve yorumlama başlıklarından oluşan bu bölümde veri seti ile ilgili bilgiler, veri setinin orijinal hali, ön işleme aşamasında yapılan değişiklikler ve oluşturulan alt kırılım veri setleri ile ilgili detaylı bilgiler ve makine öğrenmesi analizlerinde kullanılan algoritmalar, performans değerlendirme yöntemleri ve ölçütleri ile ilgili açıklamalara yer verilmiştir.

BULGULAR bölümünde, R programlama dili RStudio ve Python programlama dili ve Anaconda – Jupiter Lab, Scikit-Learn kütüphanesi kullanılmak sureti ile makine öğrenmesi sınıflandırma algoritmaları ve ilgili algoritmaların parametre/hiperparametre ayarlaması (“tuning”) için çeşitli optimizasyon teknikleri (ızgara arama ve meta-sezgisel optimizasyon teknikleri) de kullanılarak yapılan analizlerin performans değerlendirmeleri karşılaştırmalı olarak verilmiştir.

TARTIŞMA VE SONUÇ bölümünde ise bulgular tartışılmış ve gelecek araştırmalara yönelik öneriler geçmiş çalışmalarla kıyaslanarak ortaya konulmuştur.

2. GENEL KISIMLAR

2.1 DIABETES MELLITUS

2.2.1 Tanım

Diyabet (Diyabetes Mellitus: şeker hastalığı) insülin sekresyonu, insülinin etkisindeki eksiklik veya her ikisindeki kusurlardan kaynaklanan hiperglisemi ile karakterize bir grup metabolik hastalık olarak tanımlanmaktadır (American Diabetes Association, 2014).

2.2.2 Epidemiyoloji

Diyabet insidansı obezite ve hareketsiz yaşamın yaygınlaşması ile birlikte günden güne artmakta olup, diyabet günümüzde küresel bir sağlık sorunu hâline gelmiştir. Hatta diyabet hastalığı için yüzyılın sağlık krizi de denilebilir (Coşansu, 2009).

Uluslararası Diyabet Federasyonu'nun 2019 yılındaki 9. Diyabet Atlası verisine göre, 2000'deki ilk raporundan bu yana, tahmini diyabet yaygınlığı (tip 1 ve tip 2 kombine hem teşhis edilmiş hem de teşhis edilmemiş) 20-79 yaşlarındaki insanlarda 151 milyondan (o sırada küresel nüfusun %4,6'sı) 463 milyona (%9,3) yükselmiştir. Bu raporda, 2019 yılında 20 ile 79 yaş aralığındaki 463 milyon kişinin (her 11 kişiden 1'nin) diyabet hastası olduğu tahmin edilmekte ve bu sayının 2030'da 578 milyona ve bu pandemi için yeterli eylem planı uygulanmaz ise, bu sayının 2045'te 700 milyona (%10,9) ulaşacağı öngörülmüştür. Bununla birlikte diyabetli her 2 kişiden 1'inin tanı almamış diyabet hastası (232 milyon insan) olduğu bulunmuş, diyabetli her 5 kişiden 1'inin ise 65 yaş ve üzeri olduğu tespit edilmiştir. Ayrıca, 2019 yılında diyabet ve komplikasyonlarından kaynaklanan ölüm sayısının da 4,2 milyon olduğu tahmin edilmektedir. Aynı raporda global sağlık giderlerinin yüzde 10'unun diyabet için harcanmakta olduğu belirtilirken, diyabetli her 4 kişiden 3'ünün (79%) düşük-orta gelirli ülkelerde yaşamakta olduğunun altı çizilmiştir (International Diabetes Federation, 2019). 2021 yılında aynı kurum tarafından yayınlanan rapora göre 2021'de 20-79 yaşları arasında diyabet hastalığına sahip insan sayısının dünya genelinde 537 milyona ulaştığı ve 6,7 milyondan fazla insanın diyabet ve diyabete bağlı komplikasyonlardan ölmüş olduğu tahmin edilmekte olup, gerekli önleyici tedbirlerin alınmaması durumunda ise 2045 yılında bu sayının %46'lık bir artışla 783 milyon kişiye ulaşacağı öngörülmektedir. Bu raporun bulguları 21.yüzyılda küresel çaptaki en önemli sağlık sorunlarından birinin diyabet hastalığı olduğunu doğrulamaktadır. Sonuç olarak bu

hastalığın endişe verici seviyelere ulaştığının altı çizilmiştir (International Diabetes Federation, 2021).

Dünya genelinde diyabet sıklığının artmasıyla birlikte, Türkiye’de de diyabet hızla artmaktadır. Türkiye’de diyabet prevalansını belirlemek amacıyla yapılan çalışmalardan ilki Satman ve diğ. (2002) tarafından yapılan TURDEP I (Türkiye Diyabet, Hipertansiyon, Obezite ve Endokrinolojik Hastalıklar Prevalans Çalışması – I) çalışmasıdır. Bu araştırmaya 20 yaş ve üzerinde toplamda 24.778 kişi katılmış ve tümüne 75 gram glukozlu OGTT (Oral Glukoz Tolerans Testi) uygulanmıştır. Çalışma sonuçlarına göre diyabet prevalansı % 7,2 olarak bulunmuştur (Satman ve diğ. 2002).

Ülke çapındaki diyabet prevalansının son 12 yıldaki değişiminin açıklanması amacıyla yapılan TEKHARF çalışmasının sonuçlarına göre, diyabet genel prevalansının %80 oranında yükseldiği tespit edilmiştir ki bu, yıllık %5 artışa denk düşmektedir (Onat ve diğ. 2014).

Türkiye Kronik Böbrek Hastalığı Prevalansı Araştırması (Chronic Renal Disease in Turkey – CREDIT) sonuçlarına göre, diyabet sıklığı kadınlarda %14,3, erkeklerde %10,9 ve genelde ise %12,7 olarak belirtilmiştir (Süleymanlar ve diğ. 2011).

TURDEP- II çalışması ise açlık kapiller glukoz ile birlikte OGTT yöntemi kullanılarak TURDEP-I çalışmasının devamı olarak planlanmış ve aynı merkezlerde yine Satman ve diğ. (2013) tarafından gerçekleştirilmiştir. Çalışma sonuçlarına göre, 20 yaş ve üzeri yetişkinlerde diyabet prevalansı %16,5 olarak tespit edilmiş ve 2010 yılı verilerine göre Türkiye’de 6,5 milyon civarında diyabet hastası olduğu belirtilmiştir. Araştırmanın en çarpıcı bulgulardan birisi de TURDEP-I popülasyonuna göre standardize edildiğinde, diyabet prevalansının son 12 yılda %90 artış ile %13,7’ye ulaştığı hesaplanmıştır (Satman ve diğ. 2013).

Bu araştırma sonuçları göstermektedir ki, diyabet; Dünyadaki artışla birlikte Türkiye’de de hızlı bir artış eğilimindedir. Sonuç olarak hem diyabetin hem de diyabetin komplikasyonlarının önceden tespit edilmesi ve önlenmesi kritik önem arz etmektedir.

2.1.3 Teşhis

Türkiye Diyabet Vakfı ve Ulusal Diyabet Konsensus Grubu’nun 2019 yılında yayınladığı Diyabet Tanı ve Tedavi Rehberi’ne göre diyabetin tanı kriterleri Tablo 2.1’deki gibidir. Bu tabloda yer alan kriterlerden sadece birinin bulunması tanı için yeterli görülmektedir.

Tablo 2. 1: Diyabet Tanı Kriterleri (Ulusal Diyabet Konsensus Grubu, 2019)

Açlık Plazma Glukozu	≥ 126 mg/dl
Rastlantısal Plazma Glukozu + diyabet semptomları	≥ 200 mg/dl
Oral Glukoz Tolerans Testi (OGTT)'nde 2st plazma glukozu	≥ 200 mg/dl
HbA1c	$\geq 6,5$

2.1.4 Sınıflandırma

Türkiye Endokrinoloji ve Metabolizma Derneği (TEMĐ) tarafından 2018 yılında yayınlanan güncellenmiş Diabetes Mellitus ve Komplikasyonlarının Tanı, Tedavi ve İzlem Kılavuzu'na göre diyabet Tablo 2.2'deki gibi sınıflandırılabilir (TEMĐ Diabetes Mellitus Çalışma ve Eğitim Grubu, 2018).

Tablo 2. 2: Diyabetin Sınıflandırılması

Tip 1 Diyabet	Genellikle mutlak insülin eksikliğine sebep olan β hücre hasarı bulunmaktadır.
Tip 2 Diyabet	Temelinde insülin direnci bulunan ve ilerleyici insülin sekresyon defekti ile karakterizedir.
Gestasyonel diabetes mellitus (GDM)	Gebelik sırasında ortaya çıkar ve genellikle doğumla birlikte düzelir.
Diğer spesifik diyabet tipleri	Genetik defektler, pankreasın ekzokrin doku hastalıkları, endokrinopatiler, ilaç veya kimyasal ajanlar, immun aracılıklı nadir diyabet formları ve infeksiyonlar bu grup arasında sınıflandırılabilir.

Diyabetin ilk alt sınıfı, tip I veya insüline bağımlı diabetes mellitus, genellikle klinik olarak ani semptomların başlangıcı, insülinopeni ve yaşamı sürdürmek için enjekte edilen insüline bağımlılık ve ketozise eğilim ile karakterizedir. Klasik olarak, bu hastalık türü çocuklarda görülmektedir bu nedenle daha önceleri çocuk diyabeti olarak adlandırılmıştır. Bununla birlikte, herhangi bir yaşta ilk kez tanı alabilir ve semptomatik olabilir; bu nedenle, başlangıç yaşına göre tanının uygun olmadığı tespit edilmiştir.

İkinci tip diyabet, tip II veya insüline bağımlı olmayan diabetes mellitus, sıklıkla diyabetin metabolik komplikasyonlarına karşı kabul edilebilir minimal semptomlar gösterir veya hiç göstermez. Bununla birlikte, açlık hiperglisemisi eğer diyet veya oral ajanların kullanımı ile düzeltilemez ise, bu durumun tedavisi için insülin kullanımı da gerekebilir. Bu tür hastalarda, Gestasyonel diabetes mellitus infeksiyonlar veya travma ile şiddetli stres gibi özel koşullar altında ketozis gelişebilir. Tip II diyabetli hastalar yıllar veya on yıllar boyunca asemptomatik olabilir ve hastalığın sadece yavaş ilerlemesini gösterir. Bununla birlikte, diyabetle ilişkili kronik komplikasyonları, yani makroanjyopati, mikroanjyopati, nöropati ve katarakt bu tipteki hastalarda görülebilmektedir (National Diabetes Data Group, 1979).

2.1.5 Komplikasyonlar

Diyabetin hem akut hem de kronik komplikasyonları bulunmaktadır. Bunlar detaylı olarak alt başlıklarda açıklanmıştır.

2.1.5.1 Akut Komplikasyonlar

Diyabetin dolayısıyla hipergliseminin belirtileri arasında poliüri, polidipsi, kilo kaybı ve bulanık görme bulunur. Kronik hiperglisemiye belirli infeksiyonlara yatkınlık da eşlik edebilir. Kontrolsüz diyabetin akut, hayatı tehdit eden sonuçları bulunmaktadır (American Diabetes Association, 2014). Hipoglisemi (düşük kan şekeri), ketoasidoz (diyabetik koma), laktik asidoz, bakteri/ mantar infeksiyonları, hiperglisemik nonketotik koma diyabetin akut komplikasyonları arasında bulunmaktadır (Uludağ, 2010).

2.1.5.2 Kronik Komplikasyonlar

Kronik hiperglisemi ile karakteristik olan diyabet uzun vadede, göz, böbrek, kalp, kan damarları ve sinirler başta olmak üzere çeşitli organların hasarına ve disfonksiyonuna yol açabilir ve sonuç olarak da hasar gören organların yetmezliği ile ilişkilidir. Diyabetin uzun vadeli komplikasyonları arasında potansiyel görme kaybı olan retinopati; böbrek yetmezliğine yol açan nefropati, ayak ülseri, ampütasyon ve Charcot eklemleri riski olan periferik nöropati ve gastrointestinal, genitoüriner ve kardiyovasküler semptomlara ve cinsel işlev bozukluğuna neden olan otonom nöropati sayılabilir. Diyabetli hastalarda artmış aterosklerotik kardiyovasküler, periferik arter hastalığı ve serebrovasküler hastalık insidansı vardır. Ayrıca, hipertansiyon ve lipoprotein metabolizmasının anormallikleri genellikle diyabetli kişilerde bulunur (American Diabetes Association, 2014). Bununla birlikte, uygun diyabet yönetimi

sağlanırsa, bu ciddi komplikasyonlar geciktirilebilir veya tamamen önlenebilir (International Diabetes Federation, 2019).

2.1.5.3 Diyabetik Nöropati

Diyabetin en sık görülen uzun dönemli komplikasyonlarından biri olan nöropati, önemli bir morbidite ve mortalite nedenidir (Türkiye Diyabet Programı, 2015). Periferik nöropati, diyabetle ilişkili nöropatinin en yaygın şeklidir (International Diabetes Federation, 2019). Diyabeti olan hastaların %60'ında periferik nöropati olduğu belirtilmiştir (Saltoglu ve diğerleri, 2016a). Diyabetik nöropati ekstremitelerin distal sinirlerini, özellikle ayakların sinirlerini etkiler. (T.C. Sağlık Bakanlığı- Türkiye Halk Sağlığı Kurumu, 2015). Diyabetle ilişkili nöropati, ilerleyici hissizliğe neden olan simetrik duyu işlevini değiştirmektedir (International Diabetes Federation, 2019).

Hiperglisemi, glikosile edilmiş nihai ürünlerin birikmesine ve sonuçta oksidatif stres, aksonal kayıp ve demiyelinizasyona neden olarak sinir fonksiyon bozukluğuna yol açmaktadır. Fazla glikoz aldoz redüktaz enzimi tarafından sorbitole dönüştürülür. Sorbitol de myoinositol sinir büyüme faktörü düzeylerini azaltarak diyabetik nöropatiye yol açmaktadır. Bu nedenle optimal glikoz kontrolü birincil önleyici tedbir olarak görülmektedir. (Brownlee, 2005).

2.1.5.4 Diyabetik Ayak Yaraları

Diyabetik ayak yaraları, dünyada her 30 saniyede bir ayak kaybına neden olan tedavisi güç bir hastalık nedeni olmakla birlikte diyabetin en çok hastanede yatmaya neden olan komplikasyonudur. Bu komplikasyon iş gücü kaybına, sakatlıklara, psikososyal travmaya ve sağlık bakım maliyetlerinin artmasına neden olmaktadır. Sağlık Bakanlığı verilerine göre ülkemizde yapılan amputasyon sayısı yılda 12 bin civarında olup bunun önemli bir kısmını diyabetin komplikasyonları nedeniyle yapılan amputasyonlar oluşturmaktadır (Saltoglu ve diğ. 2016).



Şekil 2. 1: İstanbul Üniversitesi İstanbul Tıp Fakültesi Sualtı Hekimliği ve Hiperbarik Tıp Anabilim Dalı'nda tedavi için başvuran bir diyabetik ayak hastasına ait ve parmak ucunda sınırlı gangren bulguları olan bir diyabetik ayak fotoğrafı

Diyabetlilerde periferik nöropati, periferik arter hastalığı ve enfeksiyona yatkınlık nedeniyle ayak ülserleri ve nihayetinde amputasyonların sık görülmekte olup bunun en önemli nedeni ise diyabetik nöropatinin yarattığı hissizlik sonucu tekrarlayan travmaların fark edilmemesi olarak belirtilmiştir (Türkiye Diyabet Programı, 2015).

2.2 DİYABETİK AYAK SINIFLANDIRMA SİSTEMLERİ

Diyabetik ayak sınıflandırmaları özellikle tedavinin standartlaştırılması ve hastalığın prognozu hakkında bilgi verebilmesi nedeniyle önem taşımaktadır. Günümüzde klinikte en çok kullanılan sınıflandırma sistemi Meggit Wagner olup akademik çalışmalarda PEDİS ve Texas San Antonio Üniversitesi Diyabetik Ayak Sınıflandırması öne çıkmaktadır (Memiş, 2015).

2.2.3 Meggit Wagner Sınıflandırması

Diyabetik ayak ile ilgili bilirliliği en yüksek sınıflandırmadır. İlk olarak 1976 yılında Meggit tarafından ileri sürülmüş, daha sonra 1981 yılında Wagner tarafından geliştirilmiştir.(Meggit,

1976; Wagner, 1981). Bu sınıflandırma temel olarak yara derinliğini esas alır. (Mutluoğlu, 2013). Şekil 2.2’de Meggit-Wagner diyabetik ayak sınıflandırma sistemi görülmektedir.

Grade 0	Sağlam cilt
Grade I	Yüzeyel ülser
Grade II	Genellikle infeksiyon ile seyreden ve tendon gibi cilt altı dokulara uzanım gösteren ülser
Grade III	Abse ya da osteomyelit ile seyreden derin ülser
Grade IV	Sınırlı gangren
Grade V	Yaygın gangren

Şekil 2. 2: Meggit- Wagner Diyabetik Ayak Sınıflandırması (Meggit,1976; Wagner, 1981)

2.2.2 Teksas San Antonio Üniversitesi Diyabetik Ayak Sınıflandırması

Teksas Üniversitesi diyabetik ayak sınıflandırma sistemi, diyabetik ayakta hastalığın gidişatının temel olarak infeksiyon, iskemi ve yara derinliği ile ilişkili olduğunu dikkate alarak geliştirilmiştir (Lawrence ve diğ. 1996). Şekil 2.3’te ilgili sınıflandırma sistemi görülmektedir.

	Grade 0	Grade I	Grade II	Grade III
A	Yara yok	Yüzeysel yara	Tendon ya da kapsül tutulumu	Eklem ya da kemik tutulumu
B	Yara yok, infeksiyon mevcut	Yüzeysel infekte yara	İnfeksiyon ile seyreden tendon ya da kapsül tutulumu	İnfeksiyon ile seyreden eklem ya da kemik tutulumu
C	Yara yok, iskemi mevcut	İskemi ile seyreden yüzeysel yara	İskemi ile seyreden tendon ya da kapsül tutulumu	İskemi ile seyreden eklem ya da kemik tutulumu
D	Yara yok, iskemi ve infeksiyon mevcut	İskemi ve infeksiyon ile seyreden yüzeysel yara	İskemi ve infeksiyon ile seyreden tendon ya da kapsül tutulumu	İskemi ve infeksiyon ile seyreden eklem ya da kemik tutulumu

Şekil 2. 3: Teksas San Antonio Üniversitesi Diyabetik Ayak Sınıflandırması (Lawrence ve diğ. 1996; Mutluoğlu, 2013).

2.2.4 S(AD) SAD Sınıflandırması

S(AD) SAD sınıflandırması 1999 yılında İngiliz araştırmacılar tarafından mevcut sınıflandırma sistemlerinin ya yeterince spesifik olamayacak kadar basit ya da hatırd tutulamayacak kadar

komplike olması ileri sürülerek geliştirilmiştir (Macfarlane ve Jeffcoate, 1999). Şekil 2.4’te ilgili sınıflandırma sistemi görülmektedir.

Grade	Alan	Derinlik	Sepsis	Arteriel yetmezlik	Denervasyon
0	Sağlam cilt	Sağlam cilt	İnfeksiyon yok	Pedal nabızlar normal	Normal VAE
I	< 10 mm ²	Cilt&Cilt altı	Yüzeysel	Zayıf pedal nabız, ya da tek taraflı yok	VAE artmış
II	10-30 mm ²	Tendon, Eklem kapsülü	Sellülit	Her iki pedal nabız yok	Belirgin nöropati
III	> 30 mm ²	Eklem aralığı, Kemik	Osteomiyelit	Gangren	Charcot ayağı

Şekil 2. 4: S(AD) SAD Sınıflandırması (Macfarlane ve Jeffcoate, 1999; Mutluoğlu, 2013).

S(AD) SAD sınıflandırması her biri bir alt başlık olan faktörlerin baş harflerinin bir araya getirilmesinden oluşan bir sınıflandırmadır. Buna göre; (“Size”): yara boyutu, (“Area”): alan, (“Dept”): derinlik, (“Sepsis”): infeksiyon şiddeti, (“Arteriopathy”): arteriyel dolaşım, (“Denervation”): duyu muayenesini tarif etmektedir (Mutluoğlu, 2013).

2.2.4 PEDIS Sınıflandırması

PEDIS sınıflandırması Uluslararası Diyabetik Ayak Çalışma Grubu (IWGDF) tarafından geliştirilmiştir (Schaper, 2004). Bu sınıflandırmada, yaraların perfüzyon (“Perfusion”), boyut (“Extent”), derinlik (“Depth”), infeksiyon (“Infection”) ve duyu (“Sensation”) durumları dikkate alınarak oluşturulmuştur. Her kategori ayrı bir başlık altında, söz konusu risk faktörünün şiddetine göre derecelendirilmektedir (Mutluoğlu, 2013). Şekil 2.5’te ilgili sınıflandırma sistemi görülmektedir.

Perfusion (perfüzyon)	
Grade I	Etkilenen ayakta PAH ile ilişkili belirti ve bulgu yoktur. Ayrıca dorsalis pedis ve tibialis posterior arterleri palpabldır ya da Ayak bileği brakial indeks 0,9 ile 1,10 arasındadır ya da Ayak başparmak/brakial indeks 0,6 üzerindedir ya da Transkütan oksijen kısmi basıncı (TcPO ₂) 60 mmHg üzerindedir.
Grade II	PAH belirti ve bulguları mevcuttur ancak kritik bacak iskemisi belirti ve bulguları yoktur. Aralıklı topallama varlığı Ayak bileği-brakial indeksi < 0,9; fakat ayak bileği basıncı >50 mmHg ya da Ayak başparmak-brachial indeksi <0,6; fakat sistolik ayak başparmak kan basıncı >30 mmHg ya da 30 ile 60 mmHg arasında TcPO ₂ ya da İnvazif olmayan girişimlerde PAH ile uyumlu ancak kritik bacak iskemisi düzeyinde izlenmeyen bozukluklar
Grade III	Kritik bacak iskemisi Sistolik ayak bileği kan basıncı <50 mmHg ya da Sistolik ayak başparmak kan basıncı <30 mmHg ya da TcPO ₂ <30 mmHg.
Yaranın boyutu (Extend):	
Yaranın boyutu, debritleme sonrası tercihen santimetrekare cinsinden kaydedilmeli, bu yapılmıyorsa, yaranın en uzun çapı ile buna dik geçen diğer çapın çarpılması ile ölçülmelidir.	
Derinlik (Depth)	
Grade I	Tam kat yüzeysel ülser. Dermis altında tutulum yoktur.
Grade II	Dermis altında, fasya, kas ya da tendon tutulumu vardır.
Grade III	Kemik ve/veya eklem aralığının tutulumu vardır.
İnfeksiyon (Infection)	
Grade I	İnfeksiyon belirti ve bulgusu yok.
Grade II	Subkutan dokuların tutulumu olmadan cilt seviyesinde sınırlanan infeksiyon vardır. Sistemik bulgu yoktur ve aşağıdakilerden en az iki tanesi mevcuttur:
	Lokal ödem veya endüasyon Ülser çevresinde >0,5 ile 2 cm çap arasında eritem Lokal hassasiyet ya da ağrı Lokal sıcaklık artışı Pürülan akıntı
Grade III	İki santimetre çapın üzerinde eritem varlığı ile yukarıdaki bulgulardan en azından birinin bulunması ya da sistemik inflamatuvar yanıt olmaksızın apse, osteomyelit, septik artrit, fasiit gibi derin doku infeksiyonu varlığı.
Grade IV	Sistemik enflamatuvar yanıt ile seyreden infeksiyon tablosu. Aşağıdakilerden 2 veya daha fazlası mevcuttur: >38 ya da <36 °C'nin altında ateş varlığı. Kalp atımı >90 atım/dk Solunum sayısı >20 /dk PaCO ₂ <32 mmHg Beyaz küre sayısı >12.000 ya da <4.000/cu mm 10% immatür band formasyonunun bulunması
Duyu (Sensation)	
Grade I	Etkilenen ayakta duyu kaybı yoktur.
Grade II	Etkilenen ayakta aşağıda tanımlanan yöntemler ile tespit edilen duyu kaybı mevcuttur: 10-g'lık Semmes-Weinstein monofilaman muayenesinde, ayağın 3 farklı bölgesine uygulanan basıncın en az 2 noktada hasta tarafından algılanmaması. Ayak başparmağı üzerinde 128-Hz'lik diapazonla oluşturulan vibrasyonun hasta tarafından algılanmaması

Şekil 2. 5: PEDIS Sınıflandırması (Shaper, 2004; Mutluoğlu, 2013)

2.3 DİYABETİK AYAK YARALARININ TEDAVİSİNDE KULLANILAN YÖNTEMLER

Diyabetik ayak, diğer kronik yara çeşitleri arasında özel bir yere sahiptir ve gerek akut gerek kronik yaraların tedavisi, **temel olarak cerrahi yöntemlerle** gerçekleştirilir. Fakat diyabetik ayak gibi kronik yaralarda yalnızca cerrahi yöntemler yeterli olmamaktadır. Bu gibi durumlarda, tedavinin başarılı olması için çeşitli yardımcı tedavi yöntemleri de gerekmekte olduğu bu alanda çalışan çeşitli otoritelerce belirtilmiştir.

Kronik yara iyileşmesinde, iyileşmeyi sağlamanın yolu, antibiyotikler gibi kimyasal, kök hücre tedavisi gibi hücresel, kurtçuk tedavisi gibi biyolojik, büyüme faktörleri uygulamak gibi sitokin çerçevesinde olabilmektedir (Aktaş, 2015).

Diyabetik ayak infeksiyonunda genel tedavi hedeflerinin, yara iyileşmesini sağlayabilmek ve bacağı kurtarmak için genel tedavi yaklaşımı, acil ve agresif debridmanlarla ölü ve enfekte dokuların uzaklaştırılması, uygun antibiyotik tedavisi, metabolik kontrol, yükten ve basıdan kurtarma ("*off-loading*"), Periferik Arter Hastalığı (PAH)'nın tanısı ve uygun şekilde tedavisi ve bununla birlikte ayağın işlevinin kazandırılması (rekonstrüksiyon) olarak belirtilmiştir (Saltoğlu ve diğ. 2016).

Aşağıdaki alt başlıklarda diyabetik ayak infeksiyonlarında kullanılan en önemli tedaviler genel hatlarıyla anlatılmıştır.

2.3.1 Cerrahi Tedavi

Diyabetik Ayak Yarası ve İnfeksiyonunun Tanısı, Tedavisi ve Önlenmesi-Ulusal Uzlaş Raporu'na göre, diyabetik ayak ülserlerinde temel cerrahi hedefleri ve uygulanabilecek işlemler dört grupta toplanmıştır. Bunlar, tedavinin öncelikli hedefi olan infeksiyonlu bir diyabetik ayak yarasını infeksiyonsuz ayak yarasına çevirmek üzere basit debridmanlardan amputasyon operasyonlarına uzanan yelpazede yer alan cerrahi seçeneklerden oluşmaktadır (Saltoglu ve diğ. 2016).

2.3.2 Yükten Kurtarma

Diyabetik ayak ülserlerinde "*off – loading*" yani yükten kurtarma, yara iyileşmesini destekleyerek ülserlerin tekrarlamasını böylece amputasyonu önlemektedir (Kılıçoğlu ve diğ. 2018).

2.3.3 Antibiyoterapi

Diyabetik ayak infeksiyonunda uygun doku örnekleme ile klinik infeksiyonun hemen tanımlanması geniş spektrumlu ampirik antibiyotik kullanımını sınırlandıracak ve bilinçli antibiyotik kullanımını sağlayacak belirtilmektedir. Dar spektrumlu, yüksek dozlu, kısa süreli antimikrobiyal tedavinin ideal olduğu belirtilmiştir. Gram pozitif organizmaların diyabetik ayak infeksiyonunda en yaygın patojenleri temsil etmekle birlikte, antimikrobiyal direnç sonucunda antibiyotik kullanılabilirliğinde zorluklar ortaya çıkmaktadır. Antibiyotik tedavisi seçiminde etkinlik, organizma duyarlılığı ve yan etki profili gibi çeşitli etmenler göz önünde bulundurulur. Bununla birlikte infeksiyonun kemiğe kadar ilerlemesi (osteomiyelit) amputasyon riskini ve buna bağlı morbidite ve mortaliteyi de arttıracak belirtilmiştir diyabetik ayak infeksiyonlarında osteomiyelitin tanınmasının ve tedavisinin önemi vurgulanmaktadır (Ertuğrul, 2020).

2.3.4 Negatif Basıncılı Yara Örtüsü

Yaraların steril biçimde kapatıldıktan sonra üzerlerine sürekli veya aralıklı olarak negatif basınç uygulanarak tedavi edilmesine negatif basınçlı yara örtüsü tedavisi (NPWT) denilmektedir. NPWT tedavisinin ana kullanım alanlarından biri de diyabetik ayakdır. Bu cihazın kullanımının bildirildiği birçok klinik çalışma bulunmaktadır. Bu tedavi ile yara iyileşmesinin değişik basamaklarında yararlı etkiler lokal kan akımı artışı, ödemde azalma ve eksudanın uzaklaştırılması, bakteri yükünde azalma, anjiogeneze ve granülasyon dokusu oluşumunda artış, yara kontraksiyonu ve epitelizeasyon artışı gibi yararlı etkiler sağlanmaktadır (Aktaş, 2010).

2.3.5 Hiperbarik Oksijen Tedavisi

Sualtı - Hiperbarik Tıp Topluluğu (UHMS) tanımına göre Hiperbarik Oksijen Tedavisi (HBOT) hastaya kapalı bir basınç odasında, deniz seviyesindeki atmosferik basınçtan (1 ATA = 1 atmosfer absolute = 1 Bar = 760 mmHg) daha yüksek basınç ortamında, aralıklı olarak %100 oksijen solutulması şeklinde uygulanan sistematik bir tedavidir. Tedavi süresi ve basıncı klinik koşullara göre değişkenlik göstermekle birlikte mevcut bilgiler uygulanan tedavinin HBOT olarak nitelendirilebilmesi için, 1,4 ATA veya daha yüksek basınçlarda uygulanması gerektiği yönündedir (Undersea & Hyperbaric Medical Society, 2014).

Hiperbarik oksijen tedavisi sistemik bir tedavidir ve topikal olarak uygulanan oksijen tedavileri HBOT tedavisi olarak kabul edilmemektedir. HBOT tedavisi hastaların basınçlı ortamda oksijen solumalarına olanak sağlayan tek kişilik basınç odalarında veya çok kişilik basınç odalarında uygulanabilmektedir.

Türü ne olursa olsun, kronik yaraların ortak özellikleri doku hipoksisidir ve - yara iyileşmesi sürecinin tamamı enerjiye, başka bir ifade ile oksijene bağımlı olduğu ifade edilmiştir. Diyabetik ayak lezyonlarında, makro ve mikrovasküler sorunlara ek olarak infeksiyon ve ödeme bağlı olarak da doku hipoksisi bulunmaktadır. Bu da yara iyileşmesini durdurmakta, infeksiyon tablosunu ağırlaştırmakta ve dokuların gangrene gitmesine sebep olmaktadır. HBOT'un diyabetik ayak infeksiyonlarında yardımcı bir tedavi olarak kullanılmasının sağladığı yararları vurgulayan randomize çalışmalar bulunmaktadır. HBOT, diyabetik ayak infeksiyonlarında; iyileşme oranlarını arttırmakta, iyileşme süresini kısaltmakta, amputasyon oranlarını düşürmektedir (Aktaş, 2015).

2.3.6 Büyüme Faktörleri

Hem in vivo hem de in vitro veriler, büyüme faktörlerinin yara iyileşmesini hızlandırmadaki etkinliğini göstermiştir. En çok incelenen büyüme faktörleri trombosit kaynaklı büyüme faktörü ("*platelet derived growth factor-PDGF*"), fibroblast kaynaklı büyüme faktörü ("*fibroblast growth factor- FGF*"), dönüştürücü büyüme faktörü β -1 ("*transforming growth factor- β 1*") ve epidermal büyüme faktörü ("*epidermal growth factor -EGF*") olarak belirtilmiştir (Tsang ve diğ. 2003). Deneysel çalışmalar EGF'nin yara iyileşmesini destekleme potansiyelini göstermiştir. EGF, hayvan eksizyonel ve termal yaralanma modellerinde epidermal ve dermal onarımı açıkça uyarmaktadır (Mustoe ve diğ. 1991).

Rekombinant insan epidermal büyüme faktörü (EGF-IL) intra lezyonel enjeksiyonunun diyabetik ayak sendromunun tedavisinde kullanılması önemli terapötik yaklaşımlardan biridir. EGF, 53 amino asidin bir polipeptididir ve diyabetik ayak yaralarının tedavisinde yardımcı bir araç olarak sunulmuştur. Aktaş ve diğ. (2016) EGF-IL'nin diyabetik ayak ülserlerindeki etkinliğini belirlemek amacıyla yaptıkları araştırmada, İstanbul Tıp Fakültesi, Kronik Yara Konseyi'nde değerlendirilen, 11 diyabetik ayak sendromlu 12 diyabetik ülsere uygulanmış 75 μ g interlezyonel EGF-IL enjeksiyonunun tedavi sonuçlarını retrospektif olarak incelemişlerdir. 2 hastaya amputasyon yapılırken, geri kalan 9 hastanın 10 lezyonu tamamen iyileştiği tespit edilmiştir. Çalışma sonuçları, EGF'nin intralezyonel uygulanmasının ileri

düzeydeki iskemik özellikteki diyabetik ayak ülseri olgularında amputasyonu önleyebileceğini kanıtlamaktadır (Aktaş ve diğ. 2016).

2.4 YAPAY ZEKÂ

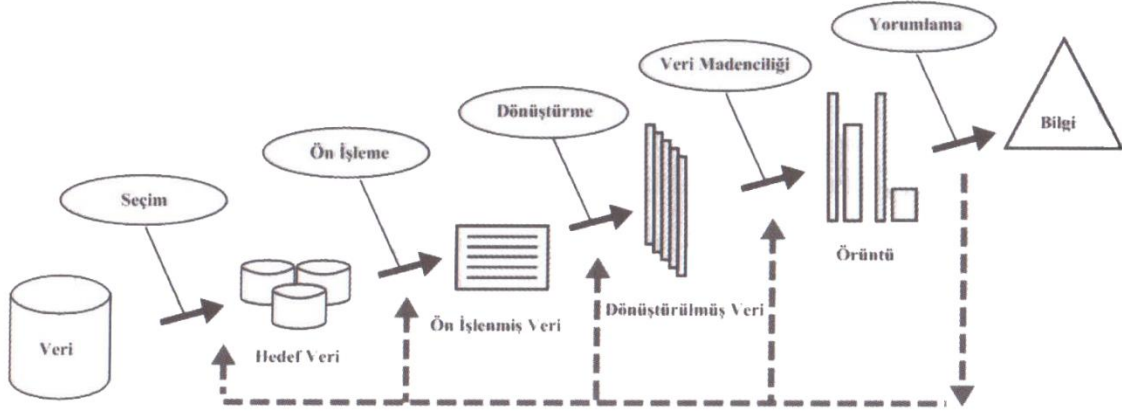
2.4.1 Tanım

Yapay zekâ, bir bilgisayarın ya da bilgisayar denetimli bir makinenin genellikle insana özgü nitelikler olduğu varsayılan akıl yürütme, anlam çıkartma, genelleme ve geçmiş deneyimlerden öğrenme gibi yüksek zihinsel süreçlere ilişkin görevleri yerine getirme yeteneği olarak tanımlanmaktadır (Nabiyev, 2012).

2.4.2 Veri Tabanlarında Bilgi Keşfi

Veri madenciliği, büyük miktardaki verilerden ilginç örüntüler ve bilgi keşfetme süreci olarak tanımlanmaktadır (Han ve diğ. 2012). Yapay zekâ, makine öğrenmesi algoritmaları ile birlikte veri madenciliği yaklaşımının tıpta uygulamasının günümüzün en önemli araştırma alanları arasında bulunmaktadır. Veri madenciliği yaklaşımı tıpta; diyabet, kanser, crohn hastalığı, böbrek yetmezliği, epilepsi rahatsızlığı gibi çok çeşitli hastalıklarda başarıyla uygulanmıştır (Delen, 2009; Kartal, 2015; Pekel ve Özcan, 2018; Davazdahemami ve Delen, 2019; Delen ve diğ., 2019; Emre ve diğ., 2019; Lin ve diğ., 2019; Reddy ve diğ., 2019; Akgül ve diğ., 2020; Özmen ve Özcan, 2020; Özmen ve Özcan, 2022).

Veri madenciliğinde yaygın olarak kullanılan süreçlerden birisi de Fayyad, Shapiro ve Smyth (1996) tarafından oluşturulan Veri Tabanlarında Bilgi Keşfi (VTBK) sürecidir. Veri setlerinden anlamlı ve faydalı olabilecek bilgi elde etme süreci veri tabanlarında bilgi keşfi olarak adlandırılmaktadır. Makine öğrenmesi algoritmalarının uygulanması da veri madenciliği/bilgi keşfi sürecinin aşamalarından biridir. Bilgi keşfi süreci kapsamında her bir aşamaya ayrıca bakıldığında, her bir aşamanın birbirini tamamlayıcı şekilde bir bütün oluşturmakta olduğu saptanmıştır. Bilgi keşfinin bu adımlara göre açıklamaları aşağıda detaylı olarak ele alınmıştır (Fayyad ve diğ. 1996b; Han ve diğ. 2012)



Şekil 2. 6: Veri tabanlarında Bilgi Keşfi (Fayyad ve diğ. 1996; Akpınar, 2014)

2.4.2.1 Seçim- Verinin Toplanması

Verinin toplanması, yapay zekâ- makine öğrenmesi algoritmalarının kullanılması ile oluşturulacak olan tahmin modelinin eğitilmesi ve test edilmesi için gerekli olan verinin elde edilmesini ifade etmektedir. Başka bir deyişle, veri seti ile yapılması planlanan analizler için hedef olarak kullanılacak ve analiz edilecek verilerin temin edilmesi olarak da tanımlanabilir.

2.4.2.2 Veri Ön İşleme

Veri ön işleme aşaması ise eldeki verilerin analiz için hazır hale getirilmesini ifade eder. Veri setlerindeki değerler her zaman analizlere doğrudan dahil etmeye elverişli olmamaktadır. Bu sebeple veri setinin elden geçirilerek analizlere hazır hâle getirilmesi gerekmektedir. Bu amaçla gerçekleştirilen çok çeşitli veri ön işleme teknikleri bulunmaktadır. Veri ön işleme, veri analizi madenciliği süratle tamamlanabilmesi ve sonuçlarının kalitesinin artırılması amacı ile veri üzerinde gerçekleştirilen veri temizleme ("*data cleaning*"), veri entegrasyonu ("*data integration*"), veri dönüştürme ("*data transformation*") ve veri indirgeme ("*data reduction*") işlemleri olarak tanımlanmaktadır (Akpınar, 2017; García ve diğ. 2015).

- **Veri Temizleme**

Kirli veri yanıltıcı, doğru olmayan veya genel formata uymayan, imla veya noktalama hataları içeren, yanlış bir veri alanına girilmiş olan veya tekrarlayan ("*duplicate*") veridir. Veri temizleme süreci genel olarak; eksik değerlerin ("*missing value*") tamamlanması, aykırı veya

başka bir deyişle sıra dışı değerlerin belirlenerek gürültünün (*noise*) azaltılması ve verilerdeki tutarsızlıkların giderilmesi adımlarından oluşmaktadır (Han ve diğ. 2012).

Veri temizlemede kullanılan teknikler aşağıdaki gibi özetlenebilir:

- **Eksik Verilerin Tamamlanması**

Eksik verilerle çok çeşitli yöntemlerle başa çıkılabilmektedir. Bu yöntemler:

- 1- Eksik değer içeren özniteliklerin silinmesi
- 2- Eksik değerlere maksimum olabilirlik prosedürü kullanılarak atama yapılması
- 3- Öznitelikler arasındaki ilişkilerin tespit edilerek eksik değerlerin tahmin edilmesi yoluyla tamamlanması

olarak özetlenmiştir (Farhangfar ve diğ. 2007; García ve diğ. 2015).

- **Veride Parazit- Gürültü Azaltımı**

Anlam içermeyen, veri girişi sırasında veya donanım ve yazılım hataları sonucunda ortaya çıkan sorunlu veri, sinyal işlemede ve enformatikte parazit (*noise*) olarak isimlendirilir.

Verideki gürültü ile başa çıkabilmek için kutulama yöntemleri, kümeleme yaklaşımları, insan ve bilgisayarın kombinasyonu ile ortak denetimi ve regresyon analizi gerçekleştirilerek veri setindeki gürültünün saptanması olarak özetlenebilir (Han ve diğ. 2012).

- **Veri Tutarsızlıklarının Saptanması**

Belirli bir özneliliğin farklı veritabanlarında farklı adlara sahip olabilmesinden mütevellit veri tutarsızlıkları olabilmektedir. Veri tutarsızlıklarının saptanması ve giderilmesi, manuel olarak veya bilgi mühendisliği araçları kullanılarak gerçekleştirilebilmektedir.

2.4.2.3 Veri Entegrasyonu

Birden çok veri kaynağından alınan verilerin birleştirilmesini içerir. Bu işlem sonucunda elde edilen veri setindeki fazlalıkları ve tutarsızlıkları önlemek için bu işlem sırasında dikkat edilmesi gereken bazı hususlar vardır.

- Bunlardan ilki **şema birleştirme hatasıdır**. Sadece bir veritabanında, farklı veri tabanlarında tutulan verilerin birleştirilmesi nedeniyle şema birleştirme hataları meydana gelebilmektedir (Han ve diğ. 2012; Oğuzlar, 2003).
- İkinci önemli husus ise “*redundancy*” olarak literatürde tanımlanan fazlalıktır. Bir öznelik başka bir tablodan türetilabiliyorsa fazlalık problemi ortaya çıkabilmektedir. Korelasyon analizi gerçekleştirilerek veri setindeki fazla özneliklerin ortaya çıkarılabilir (Han ve diğ. 2012).

2.4.2.4 Veri Dönüştürme

Veri dönüştürme aşaması veri setlerinin ideal bir şekilde analizlere hazırlanması açısından önem arz etmektedir. Veri dönüştürme; normalizasyon, yumuşatma, birleştirme ve genelleştirme adımlarından oluşmaktadır (Han ve diğ. 2012).

- **Normalizasyon**

Veri setinde bulunan niteliklerin belli bir teknik kullanılarak standartlaştırılması yani belli bir ölçü birim aralığına indirgenmesi, yapılacak analizler sonucunda elde edilecek performans sonuçlarını etkileme potansiyeli bulunmaktadır. Normalizasyon için kullanılan teknikler minimum-maksimum normalizasyon, z-score normalizasyon ve ondalık ölçekleme sıralanabilir (García ve diğ. 2015).

- **Minimum – Maksimum Normalizasyon**

Sıklıkla [0,1] veya [1,-1] normalizasyon teknikleri kullanılsa da Minimum maksimum normalizasyon, sayısal bir nitelik olan A'nın tüm sayısal değerlerini, v [yeni - min_A, yeni - max_A] ile gösterilen belirli bir aralığa ölçeklendirmeyi amaçlamaktadır. 2.1'deki denklemde görüleceği üzere, böylelikle, yeni v değerini elde etmek için denklem 2.1'e uygulanarak dönüştürülmüş bir değer elde edilmektedir:

$$V' = \frac{v - \min A}{\max A - \min A} (yeni - \max A - yeni - \min A) + yeni - \min A \quad (2.1)$$

- **Z-Score Normalizasyon**

Min-maks normalizasyonun kullanışlı olmadığı, uygulanmadığı durumlarda veya ilgili özniteliğinin minimum veya maksimum değerleri bilinmediğinde Z-score normalizasyon kullanılabilir. Bunun için 2.2 formülünden yararlanılabilir.

$$\vartheta' = \frac{v - \bar{A}}{\sigma_A} \quad (2.2)$$

$\bar{A}$, A özniteliğinin değerlerinin ortalamasıysa ve σ_A standart sapmaysa, A'nın orijinal v değeri, ϑ' kullanılarak normalleştirilir.

- **Ondalık Ölçekleme**

Sayısal bir özniteliğin mutlak değerlerini azaltmanın başka bir yolu da dönüşümden sonra maksimum mutlak değer her zaman 1'den düşük olacak şekilde ondalık ölçekleme yöntemi kullanarak normalizasyon işlemini gerçekleştirmektir. Bu dönüşüm için 2.3 numaralı formül kullanılmaktadır.

$$\vartheta' = \frac{v}{10^j} \quad (2.3)$$

(j, yeni – $\max_A < 1$ olacak şekilde en küçük tam sayıdır.)

2.4.2.5 Veri İndirgeme

Genel olarak veri indirgeme başlığı altında toplanan yöntemler, veri küpleri birleştirimi (“*data cube aggregation*”), öznitelik/boyut sayısının azaltılması (“*dimensionality reduction*”) amacıyla nitelik seçimi (“*feature selection*”) yapılması, veri sıkıştırma (“*data compression*”), nesne / gözlem sayısının azaltılması (“*numerosity reduction*”) ve ayrıklaştırma (“*discretization*”) olarak özetlenebilir (Han ve diğ. 2012). Tez araştırmasında kullanılan veri indirgeme tekniklerinden; **nitelik seçimi ve ayrıklaştırma** metodları kullanıldığı için, bu metodlar aşağıda detaylı olarak açıklanmıştır:

- **Öznitelik/ Boyut Sayısının Azaltılması**

Birçok potansiyel tahmin değişkeni içeren büyük veri setlerinde veri madenciliği süreci büyük bir sorundur ve bu sorun literatürde boyutsallığın laneti (“*the curse of dimensionality*”) olarak tanımlanmaktadır. Bu sorunu çözmek için doğrusal yöntemlerden PCA (“*Principal Component*”) kullanılır.

Analysis"); doğrusal olmayan modellerden “*Locally Linear Embedding*” ve “*ISOMAP*” geliştirilen yöntemler arasındadır. Belirli bir kritere göre en uygun özelliklerin alt kümesini seçen bir süreç olarak tanımlanan nitelik seçimi prosedürü de boyut sayısını etkili bir şekilde azaltmaya yarayan stratejiler arasındadır (García ve diğ. 2015).

Veri ön işleme basamaklarından biri olan nitelik seçiminin, çeşitli veri madenciliği ve makine öğrenmesi problemleri için verilerin (özellikle yüksek boyutlu verilerin) hazırlanmasında etkili ve verimli olduğu ifade edilmektedir. Nitelik seçiminin amaçları arasında daha basit ve daha anlaşılır modeller oluşturma, veri madenciliği performansını iyileştirme ve temiz, anlaşılır veriler hazırlama yer almaktadır (Li ve diğ. 2018).

- **Temel Bileşenler Analizi**

Temel Bileşenler Analizi (“*Principal Component Analysis – PCA*”) özellikle çok boyutlu veri setlerinde bilinen en eski ve popüler boyut azaltma metodlardan biri olarak kabul edilmektedir (Ersungur ve diğ. 2007). Bu teknik Hotelling (1933) tarafından geliştirilmiş olup, ilk olarak Pearson (1901) tarafından kullanılmıştır. Nümerik niteliklerden oluşan veri setlerine uygulanabilen bu analiz, veri temizleme ve boyut sayısını azaltma yöntemi olarak veri madenciliğinden önce sıklıkla kullanılmaktadır (Witten ve diğ. 2011).

Bu analizin temel amacı, göreceli olarak daha az sayıda değişken kullanarak varyansın çoğunu tanımlayabilen orijinal değişkenlerin bir dizi doğrusal dönüşümünü bulmak olup, özellikle çok fazla bağımsız değişken bulunan ve aralarında yüksek korelasyon mevcut olan veri setlerinde kullanışlı olduğu vurgulanmıştır (García ve diğ. 2015).

X ile temsil ettiğimiz gözlem veri setinin tüm öznitelikleri z -score yöntemine göre standart değerleri içeren yeni bir S veri kümesine dönüştürüldüğünü varsayalım. S veri kümesinin kovaryans matrisinin $\lambda_1, \lambda_2, \dots, \lambda_m$ özniteliklerini içeren E öz vektörü $\det(S - \lambda I) = 0$ eşitliği çözülerek elde edilir ve böylece ana bileşenler $Y = e_1^T X$ şeklinde ortaya konmuş olur. Birinci ana bileşen ise matematiksel olarak:

$$y_1 = e_1^T X = e_{1.1}x_1 + e_{1.2}x_2 + \dots e_{1m}x_m \quad (2.4)$$

- **Ayrıklaştırma**

Bazı sınıflandırma ve kümeleme algoritmaları yalnızca nominal yani kategorik değişkenlerle işlem yapabilmekte, nümerik tipteki niteliklerle çalışmamaktadır. Bu nedenle bu değişkenleri de yapılacak analizlere dahil etmek için, sayısal öznitelikler belirli sayıda farklı aralıkta temsil edilerek "**ayrıklaştırılmalıdır**". Ayrıklaştırma için kullanılabilen oldukça fazla sayıda metod bulunmaktadır. Bunlar arasında; entropi bazlı ayrıklaştırma ("*entropy-based discretization*"), denetimli öğrenme temelli ayrıklaştırma ("*supervised discretization*") ve hata bazlı ayrıklaştırma ("*error-based discretization*") sayılabilir (Witten ve diğ. 2011).

- **Nominal Değişkenlerin Nümerik Değişkenlere Dönüştürülmesi**

Veri indirgeme tekniklerinden biri olan ayrıklaştırma prosedürünün tam tersi bir işlem daha bulunmaktadır. Ayrıklaştırma konusu ele alınırken, nominal değişkenlerin nümerik değişkenlere dönüştürülmesi adımının da es geçilmemesi gerekmektedir. Özellikle gerçek hayattaki veri setleri hem kategorik hem de nümerik tipteki niteliklerden oluşmakta olup, doğrusal regresyon, destek vektör regresyonu, yapay sinir ağları tabanlı birçok algoritma da sayısal değişkenlerin hesaplanmasını desteklemek için tanımlanmış ve doğrulanmıştır. Dolayısı ile bu tipteki makine öğrenmesi algoritmaları da doğal olarak yalnızca sayısal olan öznitelikleri işleyebilmektedir. Bu nedenle de kategorik değişkenlerin veri setinde nümerik olarak temsil edilmesi gerekliliği ortaya çıkmaktadır. Kategorik değişkenlerin sayısal niteliklere dönüşüm yöntemlerinden biri de "*dummy-coding*" olarak bilinen ve kukla değişken olarak adlandırılan kodlama yöntemidir (Hardy, 1993; Witten ve diğ. 2011; Kim ve Hong, 2017).

Kukla değişkenler kümesi, regresyon denklemindeki sabit terim için bir kukla değişken ile bağımsız değişken(ler)in kategorilerinin her olası kombinasyonu için bir kukla değişkenlerinden oluşmaktadır (Miller ve Erickson, 1974).

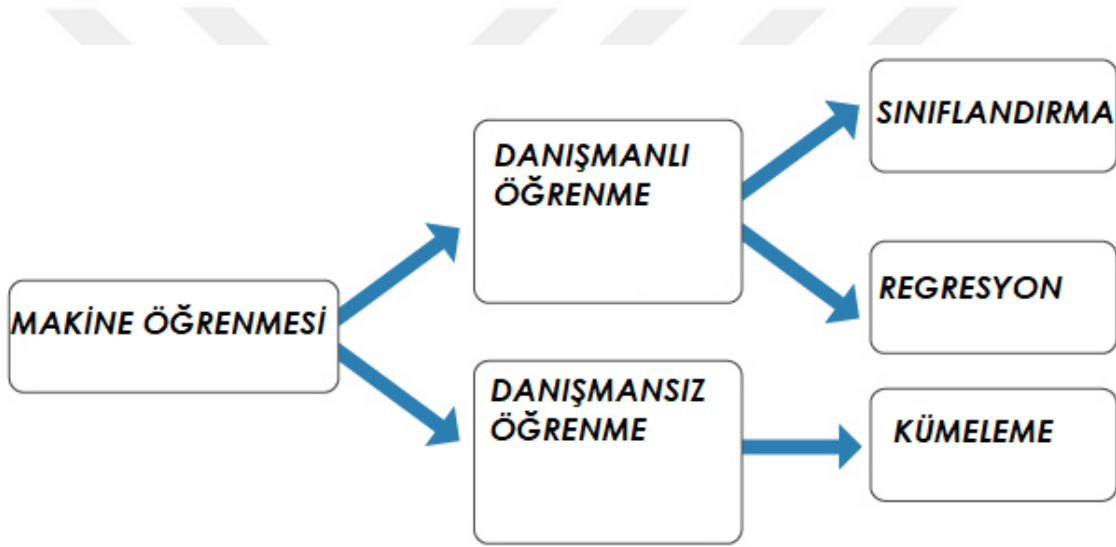
2.4.2.6 Veri Madenciliği Sürecinde Yapay Zekâ- Makine Öğrenmesi Algoritmalarının Kullanılması

Günümüzde veri boyutu gitgide artmakta ve veriler arasındaki ilişkilerin geleneksel yöntemlerle tespit edilmesi günden güne zorlaşmaktadır. Bu noktada veriler arasındaki keşfedilmemiş ilişkilerin ortaya çıkması veri madenciliği sayesinde gerçekleşmekte, bu ilişkiler üzerinden karar verme sürecinin iyileşmesi, ticarete rekabet avantajı gibi çok çeşitli faydalar ortaya çıkabilmektedir. Veri madenciliğinde veriler elektronik olarak depolanmakta ve

bilgisayar tarafından otomatikleştirilmekte olup veriler üzerinden kalıplar başka bir deyişle ilişkiler kurulması amaçlanmaktadır. Makine öğrenmesi algoritmaları ise veri madenciliği sürecinde modelleme aşamasında kullanılmaktadır (Witten ve diğ. 2011).

2.5 MAKİNE ÖĞRENMESİ ALGORİTMALARI

Makine öğrenmesi algoritmaları temel olarak üç ana grupta toplanmakta olup bunlar, **danışmanlı öğrenme**, **danışmansız öğrenme** ve **pekiştirmeli öğrenme** olarak sınıflandırılabilir (Akpınar, 2000). Bununla birlikte bu üç stratejiden birkaçını birlikte kullanarak öğrenen ağlar da bulunmakta olup bunlara; yarı-gözetimli (“*semi-supervised*”), kısmen ise gözetimsiz olarak öğrenme yapan ağlar örnek olarak verilebilir.



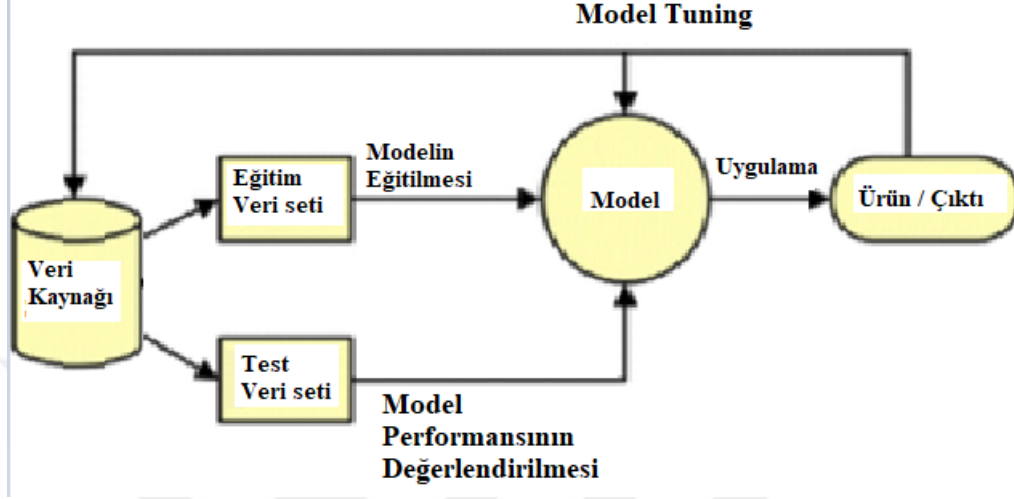
Şekil 2. 7 : Makine Öğrenmesi Yöntemleri- Kagiya ve diğ. (2019)

2.5.1 Danışmanlı Öğrenme

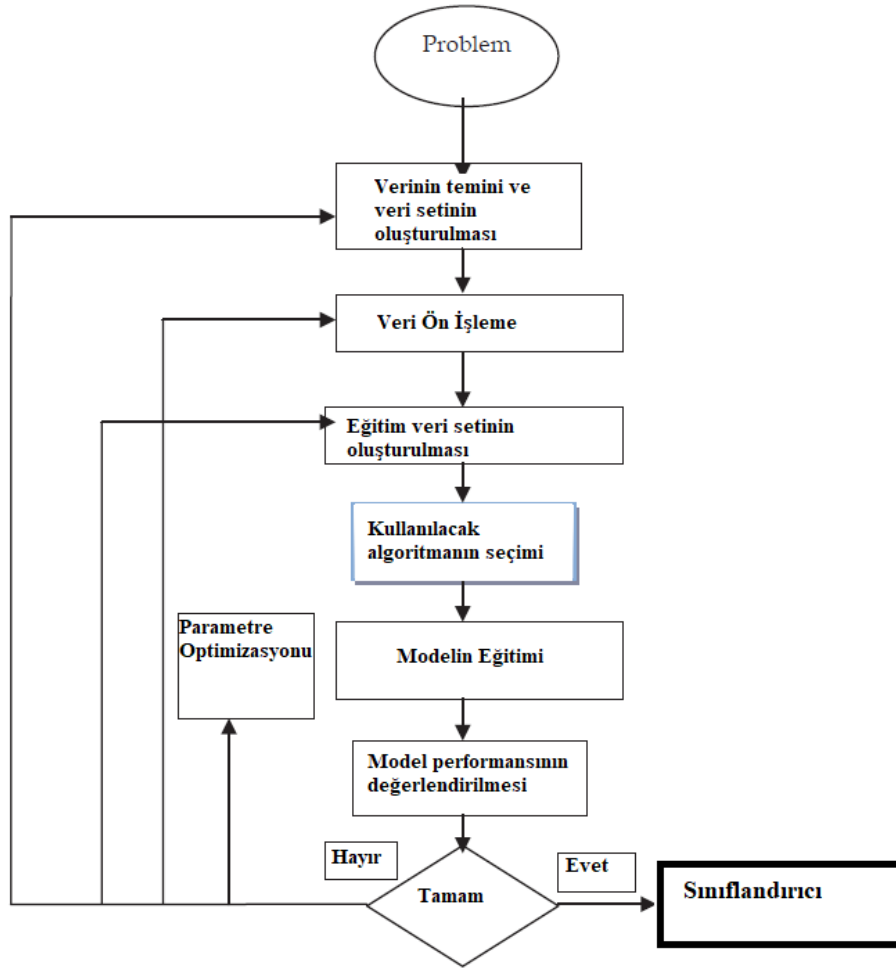
Danışmanlı öğrenme ya da başka bir ifade ile denetimli öğrenmede sınıflandırma ve regresyon olmak üzere iki ana yöntem bulunmaktadır.

Danışmanlı öğrenmede, girdi değerlerine dayanarak çıktının tahmin edilmesi amaçlanmaktadır (Hastie ve diğ. 2001). Eğer her girdi vektörüne sonlu sayıdaki ayrık kategoriden birini atama yapılacak ise bu yönteme sınıflandırma, istenen çıktı bir ya da daha fazla sürekli değerden oluşmakta ise de bu yönteme regresyon denilmektedir (Bishop, 2006).

Sınıflandırma sürecinde, başlangıçtaki veri seti, oranları değişebilen eğitim ve test veri setlerine bölünmekte olup, eğitim veri seti, tahmin edilmesi veya sınıflandırılması gereken çıktı değişkenine sahiptir (Mahesh, 2020). Şekil 2.8 ve Şekil 2.9’da danışmanlı öğrenmede gerçekleştirilmesi gereken iş akış süreçleri gösterilmiştir.



Şekil 2. 8: Danışmanlı öğrenmede iş akış süreci Mahesh (2020)



Şekil 2. 9: Danışmanlı öğrenmede iş akış süreci (Ayodele, 2010)

• Sınıflandırma

Bu tez çalışması kapsamında ele alınan problem bir sınıflandırma problemi olduğu için sınıflandırma süreci ve kullanılan algoritmalar detaylı olarak açıklanmıştır.

Veri madenciliği teknikleri içerisinde en sık kullanılan makine öğrenmesi sınıflandırma algoritmaları, elde bulunan verilerden yola çıkarak istenen nominal yani kategorik tipteki sonucun tahmin edilmesinde kullanılmaktadır (Akpınar, 2000).

Matematiksel olarak sınıflandırma Denklem 2.5'te gösterilmiştir.

$A = \{t_1, t_2, \dots, t_n\}$ bir veritabanı ve t_i de veritabanındaki bir kaydı temsil etmek üzere,

$B = \{B_1, B_2, \dots, B_k\}$ ise k adet sınıftan oluşan sınıflar kümesini temsil etmek üzere,

$f: C \longrightarrow B$; her bir t_i bir sınıfa dahil olmalı ve;

B_j daki her bir eleman ayrı bir sınıf ve her bir sınıf da kendisine ait verileri içerir. Bu doğrultuda

$$B_j = \{ t_i \mid f(t_i) = B_j, 1 \leq i \leq n \text{ ve } t_i \in C \} \quad (2.5)$$

olarak tanımlanmaktadır (Silahtaroglu, 2016).

Sınıflandırmada her bir kaydın önceden tanımlanmış kendisine ait bir sınıfı bulunmakta olup, tahmin modeli oluşturmak için rastgele atanmış eğitim örneklerini içeren eğitim veri seti kullanılır. Eğitim veri setindeki her bir örneğin sınıf etiketi bulunduğundan dolayı bu yöntemle aynı zamanda danışmanlı veya denetimli öğrenme adı verilmektedir (Han ve diğ. 2012).

Sınıflandırma için kullanılabilen çok çeşitli algoritmalar arasında, K En Yakın Komşu Algoritması, Sade Bayes Sınıflandırıcı, Karar Ağaçları, Lojistik Regresyon, Yapay Sinir Ağları, Destek Vektör Makineleri, Aşırı Öğrenme Makineleri, Topluluk ("*Ensemble*") öğrenme algoritmaları sayılabilir (Akpınar, 2000; Zhang and Ma, 2012).

2.5.2 K en Yakın Komşu Algoritması

Bu algoritmanın temeli, Fix ve Hodges'in, 1951 yılında o zamandan beri k-en yakın komşu kuralı olarak bilinen örüntü tanıma için parametrik olmayan bir yöntem geliştirip tanıtmasına dayanmaktadır (Fix ve Hodges, 1951; Peterson, 2009). K en yakın komşu (kNN) algoritması girdi uzayında benzer nesnelerin, çıktı uzayında da benzer olacağını varsayan, eğitim veri setini depolayarak işlem yapan tembel ("*lazy*") bir öğrenme yaklaşımı olarak da bilinen, sınıflandırma dışında kümeleme işlemlerinde de kullanılabilen bir algoritmadır Zhou, 2012; Silahtaroglu, 2016).

Algoritmanın çalışma ilkesi Silahtaroglu (2016) tarafından şu şekilde özetlenmiştir:

- Öncelikle mesafe ölçüm uzayı belirlenir (Mesafe ölçümü için kullanılacak mesafe ölçüleri aşağıda detaylı olarak ele alınmıştır).
- Birbirine en yakın k adet nokta saptanır.
- Belirlenen grubun en çok rastlandığı sınıf tespit edilir.
- Nihai olarak bu gruba belirlenen sınıfın ismi atanır.

Matematiksel olarak ifade edilecek olursa:

Eğitim verisinin verildiğini varsayarak,

$$X = \{x_1, x_2, \dots, x_m\}, x_i \in R^N \text{ ve}$$

$y = \{y_1, y_2, \dots, y_m\}, y_i \in \{\omega_1, \omega_2, \dots, \omega_c\}$ sınıf etiketlerine karşılık geldiğini farz edelim.

Bilinmeyen bir x örneği verildiğinde, k en yakın komşu algoritması ilk olarak m eğitim örneklerinin en yakın k komşusunu tanımlamakta, x_i sınıfına ait k_i örneklerinin sayısını saymakta ve ardından x 'i maksimum k_j örnekleriyle x_j sınıfına atamaktadır (Zuo ve diğ. 2008).

kNN ele alındığında mesafe ölçümü için kullanılabilecek mesafe ölçüleri de ele alınmalıdır:

Hu ve diğ. (2016) kNN de kullanılan bazı mesafe ölçülerini özetlemişlerdir:

Öklid uzaklığı

Bir test örneği ile belirtilen eğitim örnekleri arasındaki mesafenin ölçümüne dayanan kNN algoritmasında en çok kullanılan mesafe ölçülerinden biridir (Peterson, 2009).

$A = (x_1, x_2, \dots, x_m)$ ve $B = (y_1, y_2, \dots, y_m)$ olmak üzere,

$$dis(A, B) = \frac{\sqrt{\sum_{i=1}^m (x_i - y_i)^2}}{m} \quad (2.6)$$

Kosinüs Benzerliği

$$sim(A, B) = \frac{\vec{A} \cdot \vec{B}}{(|\vec{A}| \cdot |\vec{B}|)} \quad (2.7)$$

Chomboon ve diğ. (2015) farklı mesafe ölçülerinin kNN algoritmasının performansına olan etkilerini 11 farklı mesafe ölçüsü kullanarak incelemişlerdir:

$m \times n$ ($1'e n$) satır vektörleri $x_1, x_2, \dots, x_{mx}$ olarak kabul edilen bir $m \times n$ veri matrisi matrix X olarak,

m_y ($1'e n$) satır vektörleri $y_1, y_2, \dots, y_{m_y}$ olarak kabul edilen bir $m_y \times n$ veri matrisi matrix Y olarak verildiğinde;

x_s ve y_t vektörleri arasındaki çeşitli mesafeler aşağıdaki formüllere göre hesaplanabilmektedir:

Minkowski Uzaklığı

Minkowski mesafesinin farklı p değerleri için:

$p = 1$, Minkowski metriği “*city block distance*” ya da bilinen diğer adıyla “*Manhattan*” mesafe ölçüsünü (Craw, 2010),

$p = 2$, Minkowski metriği Öklid mesafe ölçüsünü ve

$p = \infty$, Minkowski metriği “*Chebychev*” mesafe ölçüsünü vermektedir.

$$d_{st} = \sqrt[p]{\sum_{j=1}^n |x_{sj} - y_{tj}|^p} \quad (2.8)$$

Yukarıda detaylı olarak ele alınan mesafe ölçüleri dışında, “*Mahalanobis*”, “*Correlation*”, “*Spearman*”, “*Hamming*”, “*Jaccard*” başta olmak üzere çeşitli sayıda mesafe ölçüleri bulunmaktadır.

2.5.3 Karar Ağaçları

Karar ağaçları, istatistik, veri madenciliği ve makine öğrenmesi çalışma alanlarında en çok kullanılan sınıflandırma algoritmalarından biridir. Günümüzde bir çok karar ağacı algoritması makine öğrenmesi başlığı altında geliştirilmekte olup bunlar arasında ID3, C4.5 ve C5.0 gibi algoritmalar entropiye dayalı bölümlmeyi kullanmakta, Sınıflandırma ve Regresyon Ağaçları olarak bilinen “*Classification and Regression Trees - CART*” algoritması ise “*Gini*” indeksini kullanmaktadır (Loh, 2011).

Karar ağaçlarının çalışma prensibini özetlemek gerekirse, bir dizi nitelik-değerler örneği listesinden oluşan eğitim veri setinde, her örnek belirli bir sınıfa aittir. Algoritma, her bir niteliği sırayla incelemekte ve niteliğin sınıflar arasında ne kadar iyi ayrım yaptığına dair bir bölünme bilgisi (“*goodness of split*”) ölçüsünü hesaplamaktadır. Bu doğrultuda en iyi öznitelik seçilmekte ve veriler bu öznitelik değerlerine göre alt kümelere bölünmektedir. Bu süreç, tüm örnekler doğru bir şekilde sınıflandırılana kadar her bir alt kümeye yinelemeli olarak uygulanmaktadır. Bu sürecin sonunda, düğümlerin öznitelikleri temsil ettiği ve dalların olası öznitelik değerlerini veya değer aralıklarını temsil ettiği bir ağaç yapısı oluşturulmaktadır.

Ağacın uç düğümleri yani yaprakları, hepsi aynı sınıfta olan örnek kümelerine karşılık gelmektedir (Wu ve diğ. 2008).

- **ID3 (“Iterative Dichotomiser 3”) Algoritması**

En temel karar ağaçları algoritmalarından biri olan ve Quinlan tarafından geliştirilen entropi temelli bir algoritma olan ID3 (“Iterative Dichotomiser 3”) yöntem olarak karar ağacında hangi spesifik niteliğe göre dallanmanın yapılacağını belirlemek için entropi kavramına başvurulmaktadır (Quinlan, 1986).

Belirsizlik ölçüsünün bir sistemdeki tanımı olan entropi Shannon tarafından ortaya atılmıştır (Shannon, 1948).

D bir kaynak olmak üzere bu kaynağın $\{n_1, n_2, n_3, \dots, n_n\}$ olmak üzere n tane mesaj ürettiğini varsayarsak, tüm mesajlar birbirinden bağımsız olarak üretilmekte ve n_i mesajlarının üretilme olasılıkları da p_i olarak tanımlanmaktadır. $P = \{p_1, p_2, \dots, p_n\}$ olasılık dağılımına sahip mesajları üreten D kaynağının entropisi $H(D)$ şu şekilde tanımlanmaktadır (Shannon, 1948):

$$H(D) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (2.9)$$

ID3 algoritmasının çalışma prensibini özetlemek gerekirse öncelikle hedef nitelik için 2.9 denkleminde verilen entropi hesaplanır. Sonrasında ise her öznitelik için ağırlıklı ortalamalar hesaplanmaktadır. Ağırlıklı ortalamaları için 2.10’da verilen denklemden faydalanılır.

$$H(X, D) = \sum_{i=1}^n \frac{|D_i|}{|D|} H(D_i) \quad (2.10)$$

Bundan sonraki adım ise, her bir öznitelik için **Bilgi Kazancı** değerinin hesaplanmasıdır. Bilgi kazancının nasıl hesaplanacağı ise denklem 2.11’de verilmiştir.

$$Bilgi\ Kazancı(X, D) = H(D) - H(X, D) \quad (2.11)$$

Böylelikle en büyük bilgi kazancını sağlayan öznitelik seçilmekte ve bu öznitelik de ağacın bir düğümünü temsil etmektedir. Bu adımdan sonra ise en son düğümden çıkan herbir dal için veri

seti yeniden düzenlenmekte ve entropinin hesaplanmasına tekrar başlanarak iteratif bir şekilde tekrar edilme süreci işlenecek veri (satır) bitene kadar devam etmektedir (Özkan, 2016).

- **C4.5 Algoritması**

ID3 algoritmasının dezavantajlarını gidermek ve algoritmanın işlevini geliştirmek için yine Quinlan tarafından geliştirilen C4.5 algoritması sınıflandırma amacıyla kullanılan ve karar ağaçları sınıfından olan bir algoritmadır (Quinlan, 1993). Temelinde ID3 algoritmasında olduğu gibi Entropi ve Bölme Bilgisi olan C4.5 algoritması **Kazanç Oranı** (“Gain Ratio”) adı verilen bir orandan yararlanmaktadır.

$$\text{Kazanç Oranı}(p, T) = \frac{\text{Bilgi Kazancı}(p, T)}{\text{Bölme Bilgisi}(p, T)} \quad (2.12)$$

2.12’de belirtilen Bölme Bilgisi ise 2.13’teki formül ile hesaplanmaktadır.

$$\text{Bölme Bilgisi}(p, T) = - \sum_{j=1}^n P' \left(\frac{j}{p} \right) \times \log \left(P' \left(\frac{j}{p} \right) \right) \quad (2.13)$$

C4.5 algoritması ile oluşturulan karar ağaçları, ID3’teki gibi bir dizi eğitim veri seti kullanılarak oluşturulmaktadır. Bu algoritma kullanılarak ağacın her düğümünde, örnek kümesini bir sınıfa veya alt kümelerine en etkin şekilde bölen verilere ait bir öznitelik seçilmektedir. Bu işlemin kriteri ise verileri bölmek için bir özniteliğin seçilmesiyle ortaya çıkan normalleştirilmiş bilgi kazancı yani entropi farkıdır. Sonuç olarak karar vermek için en yüksek normalleştirilmiş bilgi kazancına sahip öznitelik seçilmektedir (Hssina ve diğ. 2014).

- **C5.0 Algoritması**

C5.0 algoritması da yine Quinlan tarafından C4.5 algoritmasının eksikliklerini gidermek amacıyla C4.5 algoritmasının devamı niteliğinde geliştirilen karar ağacı algoritmalarından biridir (Quinlan, 2002). Algoritma, nihai bir sınıflandırma sonucu vermek üzere oylanacak olan bir sınıflandırıcılar topluluğu oluşturan “boosting” tekniğinin bir çeşidini kullanmaktadır (Wu ve diğ. 2008).

C5.0 algoritmasındaki en önemli değişiklikler artmış verimlilikle birlikte boosting tekniğinin de eklenmiş olması olarak belirtilmektedir. Boosting algoritması, her örnek için örneğin önemini belirten ağırlık belirlemekte, ağırlık ne kadar yüksekse, karar ağacı üzerindeki örneğin etkisi o kadar fazla olmaktadır. Başlangıçta, her örnek aynı ağırlığa sahiptir. Her denemede yeni bir karar ağacı oluşturulur. Her örneğin ağırlığı, öğrenen son denemede oluşturulan karar ağacı tarafından yanlış sınıflandırılan örnekler odaklanacak şekilde ayarlanır, bu da bu örneklerin daha yüksek ağırlığa sahip olacağı anlamına gelir (Quinlan, 1996; Freund ve Schapire, 1997; Pang ve Gong, 2009).

- **CART (“Sınıflandırma ve Regresyon Ağaçları”) Algoritması**

Sınıflandırma ve Regresyon Ağaçları (“*Classification and Regression Trees*”) hem kategorik hem de sürekli niteliklerle çalışabilen, veri setindeki eksik niteliklere sahip verilerle de başa çıkabilen ve ID3, C4.5 algoritmalarından farklı olarak ikili ağaçlar üreterek çalışan, ilk olarak Breiman (1984) tarafından tanıtılan ve teorik temeli aynı zamanda Hunt’s algoritmasına da dayanan ve bu özelliklerinden dolayı da günümüzde oldukça popüler makine öğrenmesi algoritmalarından biridir. CART algoritması karar ağacı oluşturma sürecinde “*Gini İndex*” adı verilen ve aynı zamanda nitelik seçimi yöntemlerinden biri olan bir indeksi kullanmakta olup, doğruluk yani “*accuracy*” değerini arttırmak ve ağaçtaki gereksiz dallanmaları budamak için de maliyet karmaşıklığı “*cost-complexity*” budaması yöntemini kullanmaktadır (Lavanya ve Rani, 2011).

- **Gini İndeks:**

Bu indeks, belirli bir özneliliğin bölündükten sonraki belirli bir sınıfın saflığını belirler. En iyi bölme, bölmeden sonraki kümelerin saflığını- “*purity*” arttırmaktadır. Ekonomide bir popülasyonun çeşitliliğini ölçmek için de kullanılan Gini Katsayısı, belirli bir nitelik için dallanma kararının bir sonucu olarak ilgili sınıfın saflığını saptamak için de kullanılmaktadır (Turban ve diğ. 2015; Tangirala, 2020).

Eğer bir veri seti D, n adet sınıftan oluşan örnekler içeriyorsa, Gini indeksi Denklem 2.14’teki gibi tanımlanır:

$$gini(D) = 1 - \sum_{j=1}^n p_j^2 \quad (2.14)$$

burada p_j , D veri setinde bulunan j sınıfının göreceli frekansı olarak tanımlanmaktadır. Veri seti, sırasıyla N_1 ve N_2 boyutlarıyla D_1 ve D_2 olmak üzere iki alt kümeye ayrılırsa, bölünmüş verilerin Gini indeksi n sınıfından örnekler içerir ve Gini indeksi Denklem 2.15'teki gibi tanımlanır:

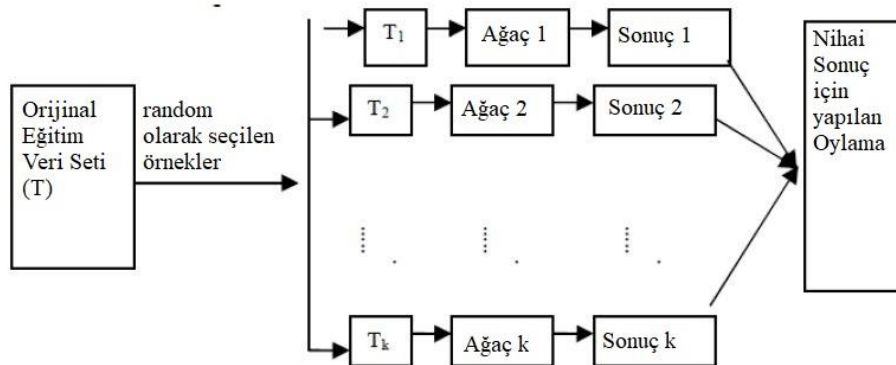
$$gini_{split}(D) = \frac{N_1}{N} gini(D_1) + \frac{N_2}{N} gini(D_2) \quad (2.15)$$

Düğümün dallanmasını saptamak en küçük $gini_{split}(D)$ 'i sağlayan “attribute/split” kombinasyonu seçilir. Böyle bir işlemde, her bir nitelik için tüm olası bölünme noktaları numaralandırılması gerektiği ifade edilmektedir (Turban ve diğ. 2015)

2.5.4 Rastgele Orman Algoritması

Leo Breiman tarafından 2001 yılında önerilen Rastgele Orman Algoritması “*Random Forest Algorithm*” (RF) topluluk “*ensemble*” öğrenme yöntemlerinden biridir. Algoritma; ormandaki her ağacın bağımsız olarak örneklenen rastgele bir vektörün değerlerine bağlı olduğu ve ormandaki tüm ağaçların aynı dağılıma sahip olduğu çok sayıda ağaçtan oluşan tahmin edicilerinin bir kombinasyonu olarak özetlenebilir (Breiman, 2001). Her ağaç bağımsız ve rastgele olarak örneklenmekte ve sonuçları çoğunluk oyu “*majority vote*” kullanılarak birleştirilmektedir. RF algoritması, ağaçlarının her birine yeni gelen veri noktalarını gönderir ve en çok ağaç tarafından sınıflandırılan sınıfı seçmektedir (Salo ve diğ. 2019).

RF algoritmasında, bir ağaç, çeşitli niteliklerin bir kombinasyonu kullanılarak oluşturulan yeni eğitim verileri üzerinde maksimum derinliğe ulaştığında diğer karar ağacı yöntemlerinden farklı olarak budama işlemi yapılmamaktadır ve bu özellik RF algoritmasının en büyük avantajlarından biri olarak ifade edilmektedir (Pal, 2005).



Şekil 2.10: Rastgele Orman Algoritması'nın şematik gösterimi (Liu ve diğ., 2012)

RF algoritması aslında en bilindik topluluk öğrenme yöntemlerinden biri olup literatürde en iyi 10 sınıflandırma algoritmasından biri olarak tanınmaktadır (Fernandez-Delgado ve diğ. 2014). RF algoritması ile birlikte benzer ağaç mimarisinde olan “*Extra-Trees*” (Geurts ve diğ. 2006) ve “*Random Patches*” algoritmaları da literatürde öne çıkan topluluk öğrenme algoritmaları arasındadır (Louppe ve Geurts, 2012).

2.5.5 Basit-Sade Bayes Algoritması

Temelinde Bayes teoremi bulunan bu algoritma basit fakat güçlü bir sınıflandırma algoritması olarak tanımlanmaktadır (Lewis, 1998). Basit ya da Sade “*Naive*” Bayes Algoritması hipotezler için manipölasyon olmaksızın olasılıkları hesaplamaktadır. Bu algoritmanın belirli tipteki makine öğrenmesi problemleri için en pratik algoritmalar arasında olduğu belirtilmiştir (Mitchell, 1997).

Bayesian öğrenme metodlarının mihenk taşı Bayes teorimidir. En sade ve akılda kalıcı hâli ile **Bayes teoremi** denklem 2.16’te görüldüğü gibidir (Bayes, 1763):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.16)$$

Bir veri kümesinin (x_i , $i=1, \dots, n$) p değişkenden oluşan n örnek (veya durum) içerdiğini varsayarsak (Berrar, 2019):

$$x_i = (x_{i1}, x_{i2}, \dots, x_{ip}).$$

Bu durumda her örneğin yalnızca bir sınıfa ait olduğu varsayılmaktadır: $y \in \{y_1, y_2, \dots, y_c\}$.

Sade Bayes yoluyla öğrenme, bir örneğe “*posterior*” bir olasılık atayan Bayes olasılıksal modelinin oluşturulmasını ifade etmektedir: $P(Y = y_j | X = x_i)$.

Sade Bayes Sınıflandırıcısı da bir sınıfa örnek atamak için bu olasılıkları kullanmaktadır. Denklem 2.16’te verilen Bayes teoremine uygulandığında ise Denklem 2.17 elde edilir.

$$P(y_j|x_i) = \frac{P(x_i|y_j)P(y_j)}{P(x_i)} \quad (2.17)$$

x_i değerlerinin birbirinden bağımsız olduğunu varsayarsak- ki bunun çoğu pratik uygulamada ihlal edildiği ve algoritmanın isminin bu nedenle “*Naive*” yani Sade olduğu ifade edilmektedir- bu varsayım sonucunda ise Denklem 2.18’de görülmektedir.

$$P(x_1|x_2, x_3, \dots, x_p, y_j) = P(x_1|y_j) \quad (2.18)$$

Böylece, x ve y_j ’nin ortak olasılığı Denklem 2.19 ve Denklem 2.20’de gösterilmiştir.

$$P(x|y_j) P(y_j) = P(x_1|y_j) \cdot P(x_2|y_j) \cdot \dots \cdot P(x_p|y_j) P(y_j) \quad (2.19)$$

=

$$\prod_{k=1}^p P(x_k|y_j) P(y_j) \quad (2.20)$$

Denklem 2.17’da yerine konulduğunda ise Denklem 2.21 elde edilmektedir. Denklem 2.21’de $P(x)$ paydası sınıfa bağlı olmayıp ölçeklendirme faktörü görevi görmektedir.

$$P(y_j|x) = \frac{\prod_{k=1}^p P(x_k|y_j) P(y_j)}{P(x)} \quad (2.21)$$

Sınıflandırma kuralı ile ilgilenecek olursak, yani bir kural her örneği bir sınıfa atayıp sonra her sınıf için payın değerini hesaplayıp ve bu değer maksimum olduğu sınıfı seçtiğinde bu kurala maksimum posterior kuralı (“*maximum posterior rule*”) denir. Böylelikle ortaya çıkan “kazanan” sınıfı aynı zamanda “*maximum a posteriori*” sınıfı olarak da bilinir ve $\hat{y}$ için Denklem 2.22’deki gibi hesaplanır (Berrar, 2019):

$$\hat{y} = \underset{y_j}{\operatorname{argmax}} \prod_{k=1}^p P(x_k|y_j) P(y_j) \quad (2.22)$$

Denklem 2.22’in uygulanması sonucunda ortaya çıkan model **Sade Bayes Sınıflandırıcısı**’dır. Sade Bayes modellerinin arasında temel olarak “*Bernouilli Naive Bayes*”, “*Gaussian Naive Bayes*”, “*Multinomial Naive Bayes*” ve “*Complement Naive Bayes*” bulunmaktadır. (Sulzmann ve diğ. 2007).

- **Gaussian Naive Bayes**

Naive Bayes sınıflandırmasında sürekli öznitelikleri ele almanın tipik bir yolu, sınıflar üzerinde koşullandırılmış özniteliklerin olasılıklarını temsil etmek için Gaussian dağılımlarını kullanmaktır (Mitchell, 1997). Bu yöntemde, verilen öznitelik değerlerinin sınıf etiketi üzerinde Gaussian dağılım olduğu varsayılmaktadır. Böylece her bir nitelik bir Gaussian olasılık yoğunluk fonksiyonu tarafından Denklem 2.23'deki gibi tanımlanır (Bustamante ve diğ. 2006).

$$X_i \sim N(\mu, \sigma^2) \quad (2.23)$$

Gaussian Naive Bayes modelinde ise Gaussian olasılık yoğunluk fonksiyonu kullanılarak 2.24'deki denklem elde edilmiş olur (Jahromi ve Taheri, 2017).

$$p(x_i|c) = \frac{1}{\sqrt{2\pi\sigma_{c,i}^2}} \exp\left(-\frac{(x_i - \mu_{c,i})^2}{2\sigma_{c,i}^2}\right) \quad (2.24)$$

2.5.6 Lojistik Regresyon

“*Logistik Regresyon*” ya da dilimizde Lojistik Regresyon adıyla sıklıkla kullanılmakta olan algoritma, özellikle ikili “*binary*” hedef nitelik içeren verileri modellemek için kullanılan en yaygın yöntemlerden biridir. Hedef nitelik iki kategoriden oluştuğunda tipik olarak 1/0 biçimini almakta; 1 genellikle başarıyı, 0 ise başarısızlığı temsil etmektedir (Hilbe, 2011). Bu analizin en önemli özelliği, doğrusal regresyon analizinden farklı olarak, hedef niteliğin kategorik olmasıdır (Dayton, 1992).

Bir modelin verilere iyi derecede uyumu için, tahmin edicilerin birbiriyle korelasyon göstermediği yani ilişkisiz olduğu fakat hedef nitelikle önemli ölçüde ilişkili oldukları varsayılır. Bununla birlikte modeldeki gözlemlerin veya verilerin de birbirleri arasında korelasyon göstermemesi de beklenmektedir (Hilbe, 2015).

Doğrusal regresyon modeli matematiksel olarak Denklem 2.25'de gösterildiği gibi ifade edilmektedir (Hilbe, 2015).

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots \beta_p x_{ip} \quad (2.25)$$

Burada $\hat{y}$ regresyon denkleminde bulunan terimlerin toplamı ve tahmin edilmeye çalışılan değer olarak tanımlanmaktadır. Modelin tahmin edilen ile Denklem 2.25'in sağ tarafındaki terimler arasında lineer bir ilişki olduğu görülmektedir fakat Lojistik Regresyon için durum böyle değildir.

Lojistik modelin doğrusal tahmincisi Denklem 2.26'te gösterilmiştir.

$$x_i b = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots \beta_p x_{ip} \quad (2.26)$$

Önemle belirtmek gerekir ki, Lojistik modelin tahmin edilen değeri $\log\left(\frac{\mu}{1-\mu}\right)$ fonksiyonuna bağlıdır. Tahmin edilen değerin, μ ve doğrusal tahmin edicinin doğrusal bir ilişkisi Denklem 2.27'da gösterilmiştir. Denklemdaki her β , ayrı bir katsayıyı veya eğimi göstermektedir. Bir prediktör yorumlanırken, hedef nitelikli ilişkili katsayısı veya değişim oranı cinsinden yorumlanmalı ve her katsayı yorumlandığında diğer tahmin edicilerin sabit tutulduğunu varsayılmalıdır (Hilbe, 2015).

$$\frac{\ln \mu_i}{1-\mu_i} = \eta_i = x_i b = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots \beta_p x_{ip} \quad (2.27)$$

Denklem 2.27'da gösterilen μ , y hedef niteliğinin 1'e eşit olma olasılığı olarak düşünülebileceği; aynı zamanda, bir özelliğin varlığının veya meydana gelme olasılığı olarak da tanımlanabileceği, $1 - p$ 'nin ise bu özelliğin olmama olasılığı olarak düşünülebileceği ifade edilmektedir. Başka bir deyişle; bir olayın meydana gelme olasılığı, başarılı ya da var olma olasılığının, başarısız olma ya da yok olma olasılığına bölümüdür ($\mu/(1 - \mu)$ veya $p/(1 - p)$). Yani; bir olayın meydana gelme olasılığı ile meydana gelmeme olasılığının toplamı her zaman 1'e eşit olmaktadır. Olasılık değerlerinin logaritması "*log of the odds*" istatistikçiler tarafından logit fonksiyonu olarak adlandırılmıştır ve Lojistik Regresyon teriminin temelini oluşturmaktadır (Hilbe, 2015).

μ için logit fonksiyonu (μ lineer tahmin edici xb temelinde belirlendiğinde, indisler gösterilmeden) Denklem 2.28'de gösterilmiştir.

$$\mu = \frac{\exp(xb)}{1 + \exp(xb)} = \frac{1}{1 + \exp(-xb)} \quad (2.28)$$

Regresyon denklemindeki her terim veya $x'\beta$; modeldeki diğer terimlerin sabit tutulması koşuluyla, prediktif x değerindeki bir birimlik değişiklik için, yanıtın log-oranlarının β katsayısı ile değiştiğini gösterir. Denklem 2.27’de verilen Lojistik Regresyon ifadesi üstelleştirildiğinde, yorum log-oranlardan ziyade bir olasılık oranı cinsinden yapılmaktadır (Hilbe, 2015). Sonuç olarak burada bir olayın gerçekleşme olasılığının, gerçekleşmeme olasılığına oranı üstünlük (*odds*) olarak yorumlanmaktadır (Kartal, 2015).

$$\frac{\mu_i}{1 - \mu_i} = e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}} = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \quad (2.29)$$

Maksimum olabilirlik tahmini, istatistikçiler tarafından bir Lojistik Regresyon modelinin parametrelerini tahmin etmek için başka bir deyişle model kestiriminde kullanılan standart yöntemdir (Hilbe, 2015). Katsayıların en çok olabilirlik kestirimini bulmak amacıyla bu yöntem kullanılmakta olup, sapmaların karesini (en küçük kareler) minimum yapmak yerine, bir olayın olma olasılığını maksimum yapılmaya çalışılmaktadır (Hair ve diğ. 2009; Çokluk, 2010).

Lojistik Regresyon analizinde değerlendirilmesi gereken noktalardan biri de modelin uyum iyiliği “*goodness of fit*” değeridir. Uyum iyiliğinin değerlendirilmesi sonucunda; model tarafından tahmin edilen değerlerin gözlenen değerlere ne kadar yakın olduğunun tespit edilmiş olunur. Uyum iyiliğinin değerlendirilmesinde χ^2 uyum iyiliği (herhangi bir kategoride beklenen sayının 5’ten az olmaması koşuluyla) ile birlikte özellikle “*Hosmer–Lemeshow*” testi sıklıkla kullanılmaktadır (Hosmer ve Lemeshow, 2000; Bewick ve diğ. 2005). χ^2 uyum iyiliği Denklem 2.30’daki formül kullanılarak hesaplanmaktadır.

$$\chi^2 = \sum \frac{(\text{gözlenen} - \text{beklenen})^2}{\text{beklenen}} \quad (2.30)$$

Modelin uyum iyiliği dışında, kullanışlılığını ölçmek için geliştirilen “*Cox & Snell*” ve “*Nagelkerke-R²*” gibi istatistikler de mevcuttur. Nagelkerke R², Cox & Snell R²’nin revize edilmiş bir versiyonudur ve 0 ile 1 arasındaki tüm değerleri kapsadığından dolayı sıklıkla tercih

edilmektedir. R^2 istatistikleri, modelin uyum iyiliğini ölçmez, ancak bağımsız değişkenlerin bağımlı değişkenini tahmin etmede ne kadar başarılı olduğunu gösterir ve etki büyüklüğü ölçüleri olarak isimlendirilmişlerdir (Bewick ve diğ. 2005).

İkili “*binary*” hedef nitelikten oluşan Lojistik Regresyon analizi dışında çok değişkenli “multinomial Lojistik Regresyon”, sıralı “ordinal Lojistik Regresyon” ve tek değişkenli Lojistik Regresyon çeşitleri de bulunmaktadır (Stephenson ve diğ. 2001; Çokluk, 2010).

- **Çok değişkenli “*Multinomial*” Lojistik Regresyon**

Makine öğrenmesi kullanılarak oluşturulmak istenen tahmin modellerinde, tahmin edilmeye çalışılan hedef nitelik yani bağımlı değişken her zaman iki kategorili yani “*binary*” olmayıp ikiden fazla değişken de içerebilmektedir. Çok değişkenli “*multinomial*” Lojistik Regresyon, kategorik tipteki bağımlı değişken sayısının ikiden fazla olma durumu olup, ikili “*binary*” Lojistik Regresyonun devamı niteliğindeki bir algoritmadır. Örneğin, bir hastanın sadece exitus ya da hayatta kalma durumunu tahmin etmek yerine; exitus, takipten çıkarılmış ve sağ olmak üzere üç grubumuz olabilir bu durumda ise tahmin prosedürü çok değişkenli Lojistik Regresyon analizi yapılarak gerçekleştirilmelidir (Chan, 2005).

Çok değişkenli Lojistik Regresyon modeline ilişkin matematiksel model Denklem 2.31’da gösterilmiştir:

$$\pi_i = \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}} \quad (2.31)$$

Çok değişkenli Lojistik Regresyonun amacı, gözlemlenen oranlar/boyutlar için hangi açıklayıcı değişken kombinasyonunun en iyi uyumu sağladığını keşfetmek olup, bu süreç açıklayıcı değişkenler arasında var olabilecek korelasyonlar nedeniyle karmaşık olduğu belirtilmektedir. Bu nedenle ileri doğru seçim yöntemi “*forward selection method*”, geriye doğru eleme yöntemi “*backward selection method*”, ileri doğru ve geriye doğru eleme yöntemlerinin kombinasyonundan oluşan adımsal seçim yöntemi “*stepwise selection procedure*” yöntemleri en iyi modeli bulabilmek için geliştirilen yöntemler arasındadır fakat tek başına hiçbir yöntem en iyi modeli bulabilmenin garantisini veremediğinden dolayı bu yöntemlerin hepsinin denenmesi ile tüm olası modellerin oluşturulabileceği ve aday modellerin de detaylı olarak incelenebileceğinin altı çizilmiştir (Stephenson ve diğ. 2001).

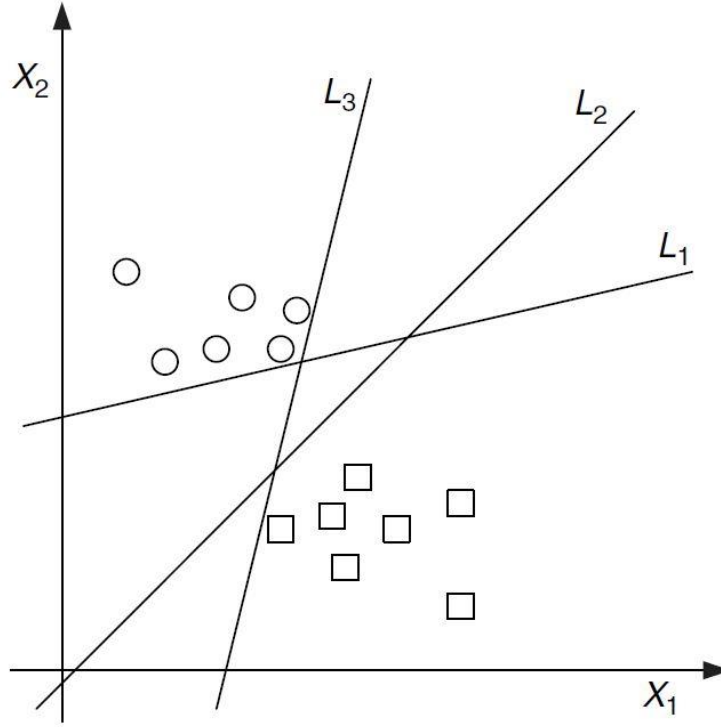
2.5.7 Destek Vektör Makineleri

Destek Vektör Makineleri (DVM), 1995 yılında Cortes ve Vapnik tarafından geliştirilmiştir (Cortes ve Vapnik, 1995). DVM hem sınıflandırma hem de regresyon ("*Support Vector Regression*") amacıyla kullanılabilen popüler makine öğrenmesi algoritmalarından biridir. (Sánchez, 2003). Bu algoritmanın nihai hedefi verileri birbirinden ayırmak için en uygun fonksiyonun tahmin edilmesi olup bu yöntemde sınıflandırma doğrusal ya da doğrusal olmayan fonksiyonun kullanılması söz konusudur (Özkan, 2016).

DVM sınıflandırma problemleri için kullanıldığında, girdi verilerini dönüştürmek için doğrusal olmayan kernel fonksiyonları (oldukça karmaşık, doğal olarak doğrusal olmayan ilişkileri temsil eder), girdi verilerinin doğrusal olarak ayrılabilir hâle geldiği yüksek boyutlu bir özellik uzayına dönüştürmek için sıklıkla kullanılmaktadır. Sınıflandırma problemlerinde, birçok doğrusal sınıflandırıcı (hiperdüzlem- "*hyperplanes*"), verileri; her biri sınıflardan birini temsil eden birden çok alt bölüme ayırabilir. Bununla birlikte, yalnızca bir hiperdüzlem, sınıflar arasındaki maksimum ayrımı elde etmektedir (Turban ve diğ. 2015).

- **Verilerin doğrusal olarak ayrılabilme durumu**

Verilerin doğrusal olarak ayrılabilmesi, gerçek hayatta pek de sık karşılaşılmayan bu durumlarda Sert Marjın DVM ("*Hard Margin SVM*") algoritmasının kullanılması uygun olacaktır (Akşehirli ve diğ. 2013). Şekil 2.11'de verilerin birden fazla doğrusal sınıflandırıcı tarafından ayrılması gösterilmiştir.



Şekil 2. 11: Birden fazla doğrusal sınıflandırıcı -hiperdüzlem “*hyperplanes*” verileri ayırabilir (Olson ve Delen, 2008).

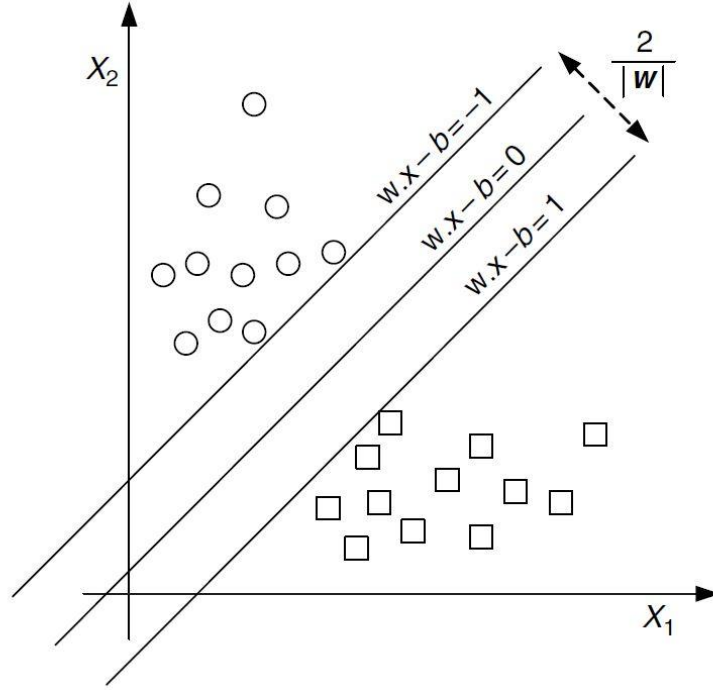
Olson ve Delen (2008) aktarımıyla; Denklem 2.32’de gösterilen eğitim veri setindeki veri kümeleri göz önünde bulundurulduğunda,

$$\{(x_1, c_1), (x_2, c_2), \dots, (x_n, c_n)\}, \quad (2.32)$$

c_i sabiti veri noktasının ait olduğu sınıfı belirtmek ve 1 ("evet") veya 0 ("hayır") olduğu durumlarda - her veri noktası, genellikle ölçeklenmiş $[0, 1]$ veya $[-1, 1]$ değerlerinden oluşan n boyutlu bir gerçek vektör olmak üzere;

$$w \bullet x - b = 0. \quad (2.33)$$

Denklem 2.33’te görüleceği üzere, w vektörü, ayırma hiperdüzlemine dik noktaları göstermekte, offset parametresi b 'yi eklemek, marjın artırılmasına izin vermekte olup yokluğunda, hiperdüzlem orijinden geçmeye zorlanır ve çözümü kısıtlamaktadır. Şekil 2.12’de hiperdüzlemlerin maksimum ayırımı grafiksel olarak gösterilmiştir.



Şekil 2. 12: Hiperdüzlemlerin maksimum ayrımı (Olson ve Delen, 2008).

$$w \bullet x - b = 1, \quad (2.34)$$

$$w \bullet x - b = -1. \quad (2.35)$$

Hiperdüzlemlerin maksimum ayrımı Denklem 2.34 ve Denklem 2.35'te gösterilen formülizasyona göre gerçekleştirilmektedir. Eğitim verileri doğrusal olarak ayrılabilirse, bu hiperdüzlemler aralarında veri kümesi olmayacak şekilde seçilebilir ve ardından mesafeler maksimize edilebilmektedir. Hiperdüzlemler arasındaki mesafe $2/|w|$ olup, sonuç olarak $|w|$ değeri minimize edilmelidir. Denklem 2.36 ve Denklem 2.37'de görüleceği üzere veri noktalarını hariç tutmak için, tüm i değerleri için de bu koşulların sağlanması gerekmektedir.

$$w \bullet x_i - b \geq 1 \text{ veya} \quad (2.36)$$

$$w \bullet x_i - b \leq -1. \quad (2.37)$$

Başka bir gösterim ile de Denklem 2.38 elde edilmiş olunur.

$$c_i(w \bullet x_i - b) \geq 1, \quad 1 \leq i \leq n. \quad (2.38)$$

- **Primal Çözüm**

$$\text{Amaç fonksiyonu : } \min(1/2) ||w||^2 \quad (2.39)$$

$$\text{Kısıt: } c_i(w \bullet x_i - b) \geq 1, \quad 1 \leq i \leq n. \quad (2.40)$$

Amaç fonksiyonu Denklem 2.39’da kısıt ise Denklem 2.40’ta gösterilmiştir. Sonuç olarak problem $|w|$ ’nin belirtilen kısıtlar içinde minimize edilmesidir. Bu da karesel programlama “*quadratic programming (QP)*” optimizasyon problemidir (Cortes ve Vapnik, 1995).

Optimum hiperdüzlemin var ve tek olduğu yukarıda verilen optimizasyon probleminin çözümü ile kanıtlanmış olup bu sayede negatif örnekler, hiperdüzlemin bir tarafında pozitif örnekler ise diğer tarafta bulunur (Vapnik, 1998; Öz, 2019).

- **Dual Çözüm**

Lagrange çarpanları kullanılarak ve *Karush-Kuhn-Tucker* koşullarından yararlanılarak problem, a_i (Lagrange çarpanları) ile ilgili olarak sadece maksimizasyon gerektiren dual probleme dönüştürülmüş olur. Lagrange fonksiyonunun w ve b ’ye göre kısmi türevlerinin alınıp sıfıra eşitlenmesi ile elde edilen denklemler, Lagrange fonksiyonundaki denklemde yerine konulduğunda kuadratik programlama probleminin duali elde edilir ve a_i için belirlenen kısıtlara bağlı olarak da çözülür ve sonuç olarak pozitif a_i değerlerine karşılık gelen x_i örnekleri destek vektörler olarak adlandırılır (Alpaydın, 2013; Öz, 2019).

Sınıflandırma kuralının dual biçimde yazılması ile sınıflandırmanın yalnızca destek vektörlerinin, yani kenar boşluğunda bulunan eğitim verilerinin bir fonksiyonu olduğunu ortaya çıkarmaktadır. Dual formda DVM Denklem 2.41 ve Denklem 2.42’de gösterilmiştir (Turban ve diğ. 2015).

$$\max \sum_{i=1}^n a_i - \sum_{i,j} a_i a_j c_i c_j x_i^T x_j \quad (2.41)$$

α terimleri ağırlık vektörü için eğitim seti açısından dual temsili içermek üzere;

$$w = \sum_i a_i c_i x_i \quad (2.42)$$

- **Verilerin doğrusal olarak ayrılama durumu**

Verilerin doğrusal olarak ayrılama durumunda ise Denklem 2.40'ta verilen kısıtlar geçerli olmayabileceğinden ve bazı veri örneklerinin sınırın içinde veya sınırın yanlış tarafında yer alabileceğinden dolayı her bir örnek için “*slack*” değişkenleri eklenmek suretiyle kısıt denklemi genelleştirilerek yeni optimizasyon problemi sınır maksimizasyonu ve hata minimizasyonu amaçlarının birleşimi olarak yeniden tanımlanır. Başka bir deyişle sınıfları ayırabilecek bir hiperdüzlem bulunmadığında, “*Soft Margin*” Yöntemi olarak bilinen metodla, örnekleri mümkün olan en iyi derecede ayıran ve sınırlar arası mesafeyi maksimize eden bir hiperdüzlem seçilmektedir (Olson ve Delen, 2008; Alpaydın, 2013; Öz, 2019). Bu denklem de 2.43'te gösterilmiştir.

$$c_i(w \bullet x_i - b) \geq 1 - \xi_i \quad 1 \leq i \leq n \quad (2.43)$$

Amaç fonksiyonu daha sonra sıfır olmayan ξ_i 'yi cezalandıran bir ceza fonksiyonu tarafından artırılır ve optimizasyon, geniş bir “margin” ile küçük bir hata cezası arasında bir değiş tokuş problemine dönüşür. Ceza fonksiyonu lineer ise, amaç fonksiyonu Denklem 2.44'teki gibi gösterilebilmektedir.

$$\min ||w||^2 + C \sum_i \xi_i \quad \text{öyle ki} \quad c_i(w \bullet x_i - b) \geq 1 - \xi_i \quad 1 \leq i \leq n. \quad (2.44)$$

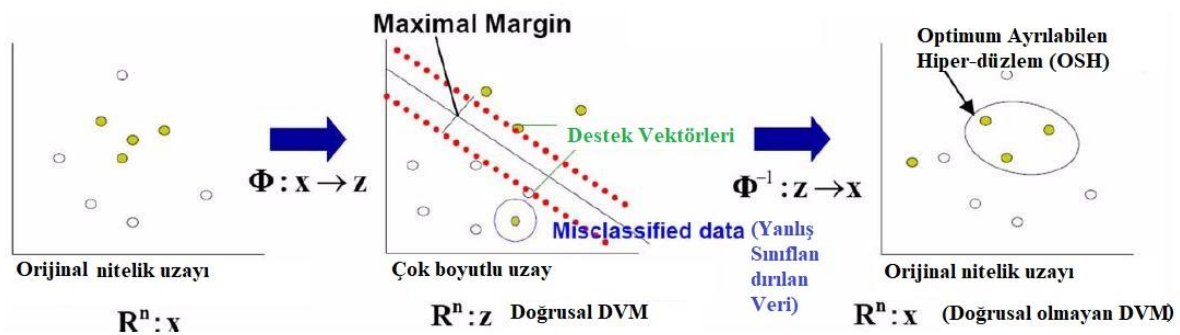
Bu kısıtlama da, $|w|$ değerinin minimize edilmesi ile birlikte Lagrange çarpanları kullanılarak çözülebilir (Turban ve diğ. 2015). C parametresi destek vektörlerinin sayısı ile ayrılama noktasının sayısı arasındaki ağırlığı için kullanılır yani daha iyi bir genellemeye ulaşabilmek için sınır bandı içerisindeki örnekler yanlış sınıflandırılan örneklerle birlikte değerlendirilmiş olur (Alpaydın, 2013; Öz, 2019).

➤ **Doğrusal Olmayan DVM ile Sınıflandırma**

Gerçek hayat problemlerinde verilerin doğrusal olarak ayrılabilmesi nadir olarak gerçekleştiğinden bu problemler doğrusal DVM ile çözülememektedir. Doğrusal olmayan DVM'de veriler doğrusal olmayan çekirdek fonksiyonları yardımıyla doğrusal olarak ayrılabilir

duruma dönüştürülmektedir. Bu tipteki problemler DVM tarafından orijinal verilerden sınıflandırma özellikleri, doğrusal olmayan haritalama yapılarak, verilerin n boyutlu orijinal girdi uzayından daha yüksek boyuta sahip belirleyici nitelik uzayına taşınarak ele alınmaktadır. Esas amaç yeni boyutta, verilerin dönüştürüldüğü sınır bandı en yüksek olan ayırıcı hiperdüzlemi bulmak olup; DVM, belirleyici nitelik uzayında doğrusal sınıflandırma kuralını, maksimum marjini bulmak amacıyla öğrenmektedir (Shihong ve diğ. 2003; Akşehirli ve diğ. 2013; Öz, 2019).

Şekil 2.13'te doğrusal olmayan DVM için haritalama ile iki boyutlu uzaydan, üç boyutlu uzaya verilerin dönüştürülme işlemi temsil edilmiştir.



Şekil 2.13: Doğrusal olmayan DVM için haritalama (Lee ve Kim, 2010).

Doğrusal biçimde ayırım sağlanamayan biçimdeki verilerin doğrusal olarak ayrılabilir forma dönüştürülmesi hedefiyle doğrusal olmayan bir fonksiyondan faydalanılarak $\Phi: \mathbb{R}^n \rightarrow H$ biçiminde haritalanmak suretiyle Hilbert uzayı olarak bilinen daha yüksek boyutlu bir alana yerleştirilip maksimum sınır hiperdüzlemi, doğrusal olmayan DVM kullanılarak bir özellik uzayına uyum sağlaması yoluyla doğrusal bir sınıflandırma problemi elde edilmiş olur (Martínez-Ramón ve Christodoulou, 2006; Öz, 2019).

Dual optimizasyon probleminin Lagrange'ına işaret eden denklemler de Denklem 2.45'e dönüştürülmüş olur:

$$L_d = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j K(x_i^T x_j)$$

(2.45)

▪ Çekirdek Fonksiyonlar ve Kernel Hilesi

Doğrusal olarak verinin ayırt edilmesinin mümkün olmadığı durumlarda veriyi daha büyük boyutlu uzaylara taşıyarak çözümlemek için ϕ fonksiyonlarını kullanılabilmektedir. Sonuç olarak giriş vektörlerini içeren çekirdek fonksiyonu Denklem 2.46’de gösterildiği gibi elde edilmiş olunur. Bununla birlikte çekirdek fonksiyonlarının belirlenmesinin temelinde Mercer teoreminin bulunduğu ve Denklem 2.46’in yazılabildiğini sağlayan bir ϕ eşlemesi varsa pozitif tanımlı ve simetrik $K(x, z)$ çekirdek fonksiyonun da bulunduğunun altı çizilmiştir (Özkan, 2016):

$$K(x_i^T x_j) = \phi(x_i)^T \phi(x_j) \quad (2.46)$$

Yüksek boyutlu veri kümelerinde $\phi(x_i)^T \phi(x_j)$ hesaplanmasının zor olabileceği hatta uygulanacak haritalama fonksiyonu bilinse dahi, kurulan optimizasyon probleminin yüksek boyutlu belirleyici nitelik uzayında çözümünün oldukça karmaşık ve zor hesaplamalar gerektirebileceği belirtilmiş olup, bu sorunu önlemek amacıyla çekirdek düzenlemesi olarak adlandırılan kernel-hilesi (“*kernel-trick*”) yöntemi önerilmiştir (Akşehirli ve diğ. 2013; Çatak, 2018) Böylelikle bir çekirdek fonksiyonu kullanılarak ϕ nitelik haritalamasının yapılabileceği belirtilmiştir. $K(x, z)$ hesaplamasının, $\phi(x)$ hesaplamasına göre daha kolay çözüme kavuşturulacağı ifade edilmekte olup Denklem 2.47’de verilmiştir (Çatak, 2018).

$$K(x, z) = \phi(x)^T \phi(z) \quad (2.47)$$

Tablo 2.3: Literatürde sık kullanılan ve en bilinen γ, r, τ ve d çekirdek parametrelerini içeren çekirdek fonksiyonları (Scholkopf ve Smola, 2001; Öz, 2019).

Çekirdek	$K(x_i^T x_j)$
Doğrusal	$x_i^T x_j$
Radyal tabanlı fonksiyon	$\exp(-\gamma \ x_i - x_j\ ^2), \gamma > 0$
Polinomial	$(\gamma x_i^T x_j + r)^d, \gamma > 0$

Sigmoid	$\tanh(\gamma x_i^T x_j + r)^d, \gamma > 0$
Çoklu Kuadratik	$(\ x_i - x_j\ + \tau)^{\frac{1}{4}}$
Ters Çoklu Kuadratik	$(\ x_i - x_j\ + \tau)^{-\frac{1}{4}}$

- **DVM algoritmasının optimize edilmesi gereken parametreleri**

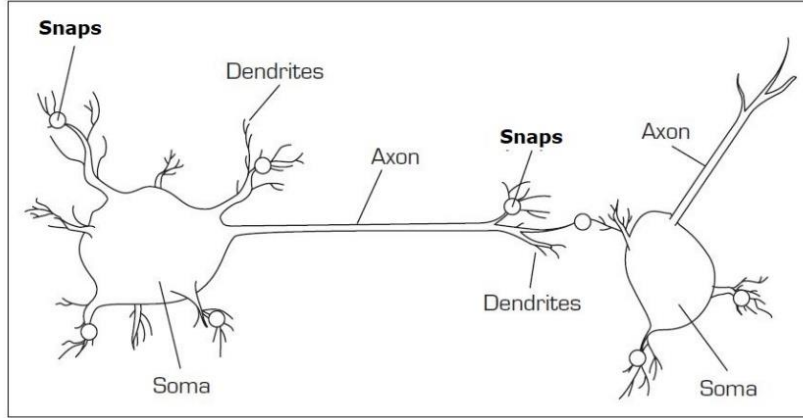
C, tüm DVM modellerinin önemli bir hiper parametresi olan ceza parametresidir. Bununla birlikte DVM modellerinde çekirdek parametresinin seçimi de önemli rol oynamaktadır. Ayrıca çekirdek fonksiyon denklemlerinde gösterildiği gibi, bir çekirdek tipi seçildikten sonra birkaç farklı hiper parametrenin de ayarlanması gerekebilmektedir (Yang ve Shami, 2020).

2.5.8 Yapay Sinir Ağları

Yapay sinir ağları (YSA) - “*Artificial Neural Networks*”, insan beyninin çalışma özelliklerinden olan öğrenme yeteneğini taklit ederek olayları öğrenebilen, çevreden gelen olaylara karşı potansiyel tepkileri belirleyebilen bilgisayar bilimleri - makine öğrenmesi disiplini içinde sınıflandırma, kümeleme, tanıma, eşleştirme ve tahmin gibi çok çeşitli uygulama alanları bulunan ve tüm bu özelliklerinden dolayı da oldukça popüler olan yapay zekâ – makine öğrenmesi algoritmalarından biridir (Öztemel, 2016; Alp, 2019).

YSA temelinde insanın biyolojik sinir sisteminin çalışma prensibi olduğundan dolayı öncelikle biyolojik sinir hücresi ve ilişkili temel kavramların ele alınması gerekmektedir.

Biyolojik sinir sistemi; aksonlar, sinapslar ve dendritlerden oluşan sinirsel iletim yoluyla elektrik ve kimyasal sinyaller - “*neurotransmitter*” göndererek iletişim kuran birbirine oldukça bağlı bir nöron ağı olarak tanımlanabilir (Krogh, 2008).

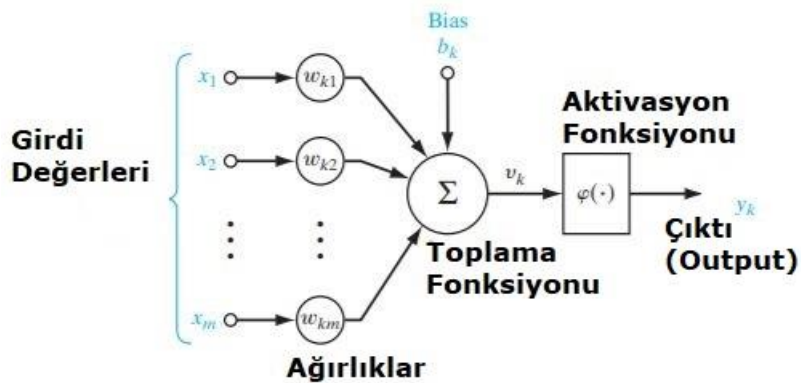


Şekil 2. 14: Biyolojik bir sinir hücresinin yapısı Olson ve Delen (2008).

Şekil 2.14'te de görüleceği üzere; sinapslar bir hücreden diğerine elektrik sinyallerinin geçmesini sağlayan boşluklar olarak tanımlanabilirken somalar ise sinapslardan gelen sinyalleri işleme tutmakla görevlidir. Sinir hücresi ise kendi elektrik sinyalini oluşturur ve akson tarafından dendritlere gönderilir. Bununla birlikte iki hücre arasındaki iletişim sinaptik bağlantılarda nörotransmitter'lar vesilesiyle olmaktadır (Öztemel, 2016).

- **Yapay Sinir Hücresi**

Verilen özelliklerde milyarlarca sinir hücresi bir araya gelerek sinir sistemini oluşturmakta olup, YSA ağları da bu biyolojik sinir sisteminden esinlenerek oluşturulmuştur. Şekil 2.15'te yapay bir sinir hücresi ve elemanları gösterilmiştir.



Şekil 2. 15: Yapay bir sinir hücresi ve elemanları (Haykin, 2009).

Öztemel (2016) yapay sinir hücresinin girdiler, ağırlıklar, toplama fonksiyonu, aktivasyon fonksiyonu ve hücrenin çıktısı olmak üzere beş temel elemanı olduğunu altını çizmiştir.

Yapay bir sinir hücresine dış dünyadan gelen bilgiler girdiler olarak tanımlanmakta, ağırlıklar ise hücreye gelen bilginin önemini ve hücre üzerindeki etkisini göstermekte, toplama fonksiyonu bir hücreye gelen net girdiyi hesaplamakta olup sonraki adımda ise aktivasyon fonksiyonu hücreye gelen net girdiyi işleyerek hücrenin bu girdiye karşılık üreteceği çıktıyı belirlerken, aktivasyon fonksiyonu tarafından belirlenen çıktı değeri ise hücrenin çıktısı olarak ifade edilmektedir.

YSA'nın çalışması ağırlık eğitimi ve test edilmesi olmak üzere iki aşamadan oluşurken öncelikle eğitim veri seti ile algoritmanın öğrenmesi sağlanmakta sonrasında ise öğrenen modele test veri seti gösterilerek algoritmanın çıktıları yorumlanmakta olup, performans değerlendirme işlemleri ile de çıktıların performansı değerlendirilmektedir (Özen ve Gülseçen, 2016).

Toplama Fonksiyonu

Bir sinir hücresine gelen toplam girdiyi hesaplama işlemini gerçekleştirmeye yarayan toplama fonksiyonu; bir sinir hücresine yönelik bağlantıların ağırlıkları ile girdi değerlerinin çarpılması sonucu elde edilen değerlerin toplamı olarak tanımlanmıştır. Toplama fonksiyonu matematiksel olarak Denklem 2.48'de verilmiştir (Özkan, 2021).

$$top_{gm} = w_1x_1 + w_2x_2 + \dots + w_nx_m + b_1 = \sum_{i=0}^m w_ix_i + b_1$$

(2.48)

Literatürde belirtilen toplama fonksiyonları arasında, toplama, çarpma, maksimum, minimum, işaret ve ekleme fonksiyonları bulunmaktadır (Alp, 2019).

Tablo 2. 4: Birleştirme (Toplama) Fonksiyonları

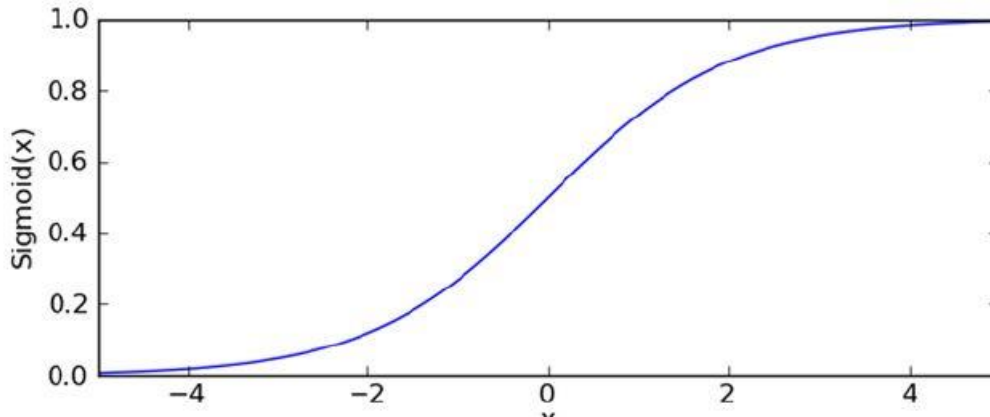
Fonksiyon	Formül
Toplama	$net = \sum_{i=1}^n x_i w_i$
Çarpma	$net = \prod_{i=1}^n x_i w_i$
Maksimum	$net = \max (x_i w_i)$
Minimum	$net = \min (x_i w_i)$

İşaret	$net = sgm(x_i w_i)$
Ekleme	$net = net(eski) + \sum_{i=1}^n x_i w_i$

- **Aktivasyon Fonksiyonları**

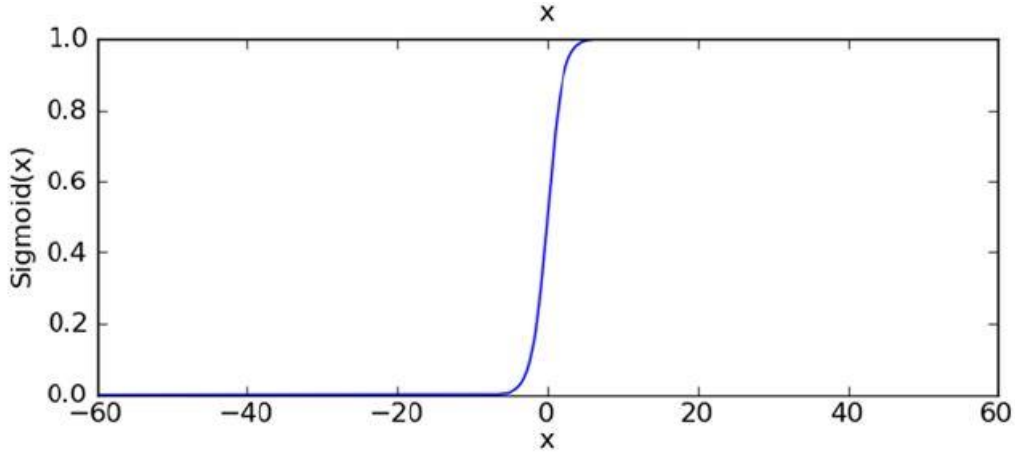
Y nöronunun ateşlenebilmesi başka bir deyişle aktive olması yani etkin hâle gelebilmesi için y-girdi değerinin bir fonksiyon eşliğinde belirli bir eşik değerine ulaşması [$y = f(y - girdi)$] gerekmektedir. Aktivasyon fonksiyonlarından en sık kullanılanı çok katmanlı algılayıcılarda (“*multi layer perceptron*”) da kullanılan S- şeklinde olan ve kendisine gelen değeri 0 ile 1 arasında bir değere dönüştüren lojistik-sigmoid fonksiyonudur. Yumuşak geçişli sigmoid aktivasyon fonksiyonu Şekil 2.16’da ve sert geçişli “*step-funtion*” ise Şekil 2.17’de gösterilmiştir (Harrington, 2012; Silahtaroglu, 2016). Şekil 2.18’de ise hiperbolik tanjant aktivasyon fonksiyonu grafiksel olarak gösterilmiştir.

Yumuşak geçişli Sigmoid fonksiyonu: $f(x) = \frac{1}{1+e^{-x}}$ (2.49)



Şekil 2. 16: Yumuşak geçişli Sigmoid aktivasyon fonksiyonu (Harrington, 2012).

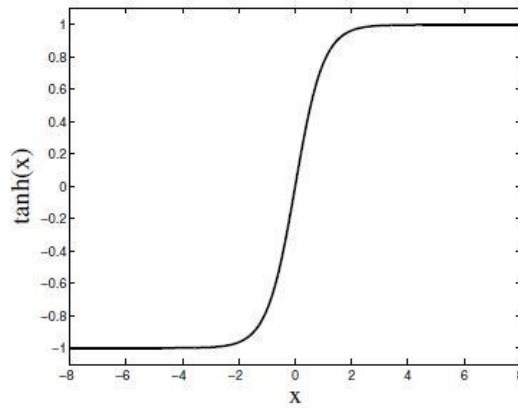
Sert geçişli Adım (Step) function aktivasyon fonksiyonu $f(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$ (2.50)



Şekil 2. 17: Sert geçişli “Step function” aktivasyon fonksiyonu (Harrington, 2012).

Denklem 2.51’de gösterilen diğer bir ateşleme fonksiyonu olarak bilinen hiperbolik tanjant tipi aktivasyon fonksiyonu ise $[-1,1]$ arasında değerler almaktadır (Namin ve diğ. 2009; Silahtaroglu, 2016).

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2. 51)$$



Şekil 2. 18: Hiperbolik tanjant aktivasyon fonksiyonu (Namin ve diğ., 2009).

Literatürde bulunan diğer bazı aktivasyon fonksiyonları arasında doğrusal, doğrusal pozitif, eşik, rampa ve Gauss ve sinüs aktivasyon fonksiyonları bulunmaktadır (Alp, 2019; Eğrioglu ve diğ. 2020).

Doğrusal Aktivasyon Fonksiyonu:

$$f(net) = \alpha \times net \quad (2.52)$$

Doğrusal Pozitif Fonksiyonu:

$$f(net) = \begin{cases} net; & net \geq 0 \\ 0; & net < 0 \end{cases} \quad (2.53)$$

Eşik Aktivasyon Fonksiyonu:

$$f(net) = \begin{cases} \alpha & net > \theta \\ \beta & net \leq \theta \end{cases} \quad (2.54)$$

Rampa Aktivasyon Fonksiyonu:

$$f(net) = \begin{cases} \alpha; & net \geq \alpha \\ \frac{\gamma-\beta}{\alpha-\beta}net + \frac{\alpha\delta+\beta\gamma}{\alpha-\beta}; & \beta \leq net \leq \alpha \\ \delta; & net \leq \beta \end{cases} \quad (2.55)$$

Gauss Aktivasyon Fonksiyonu:

$$f(net) = \exp\left(-\frac{net^2}{\sigma^2}\right) \quad (2.56)$$

Sinüs Aktivasyon Fonksiyonu:

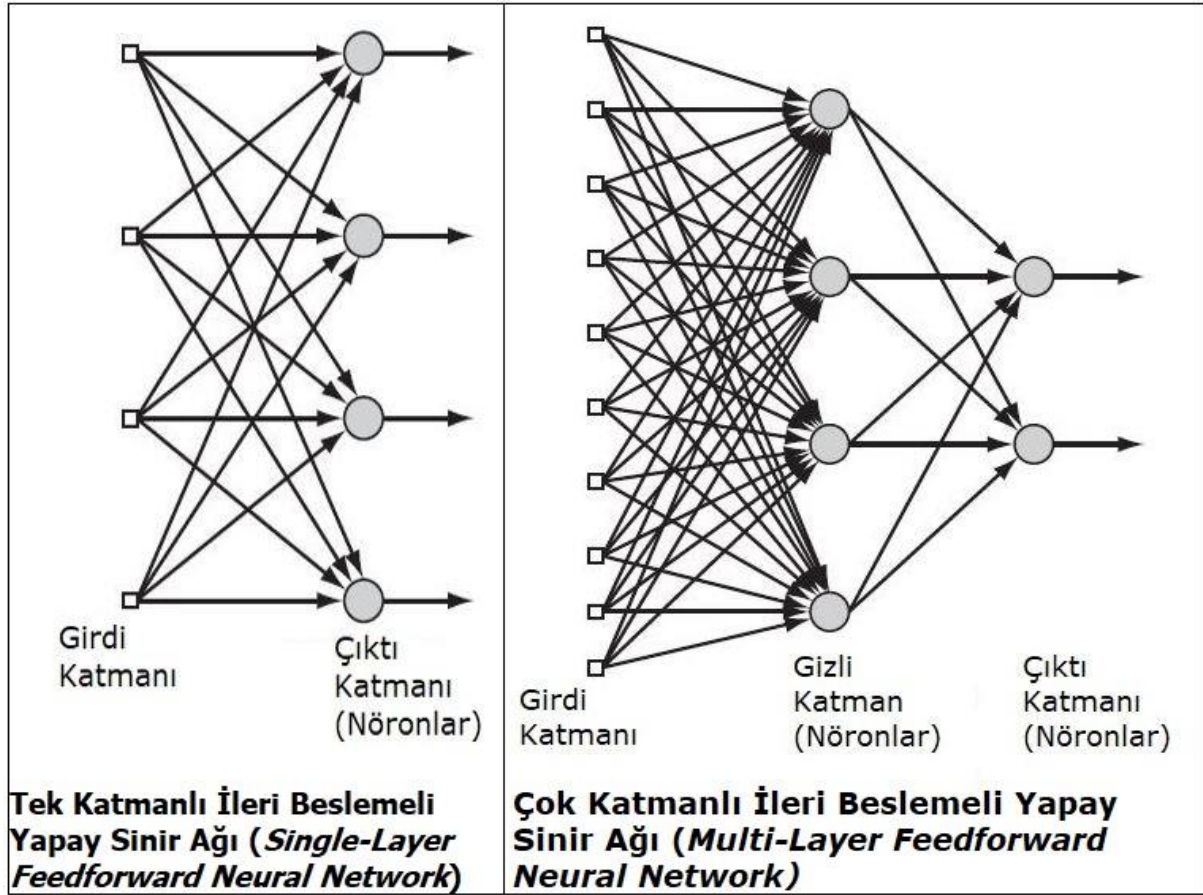
$$f(net) = \sin(net) \quad (2.57)$$

- **Yapay Sinir Ağları'nın Sınıflandırılması**

YSA, Alp (2019) temel olarak katman sayısına göre tek katmanlı YSA (Perceptron, ADALINE ve MADALINE) ve çok katmanlı YSA; ağıın çalışma prensibine göre ise de İleri Beslemeli YSA ve Geri Beslemeli/Yayılmalı YSA olarak sınıflandırılmış olsa da Leijnen ve Veen (2020) literatürde bulunan en detaylı YSA taksonomilerinden birini sunmuşlardır.

- **İleri Beslemeli Yapay Sinir Ağı**

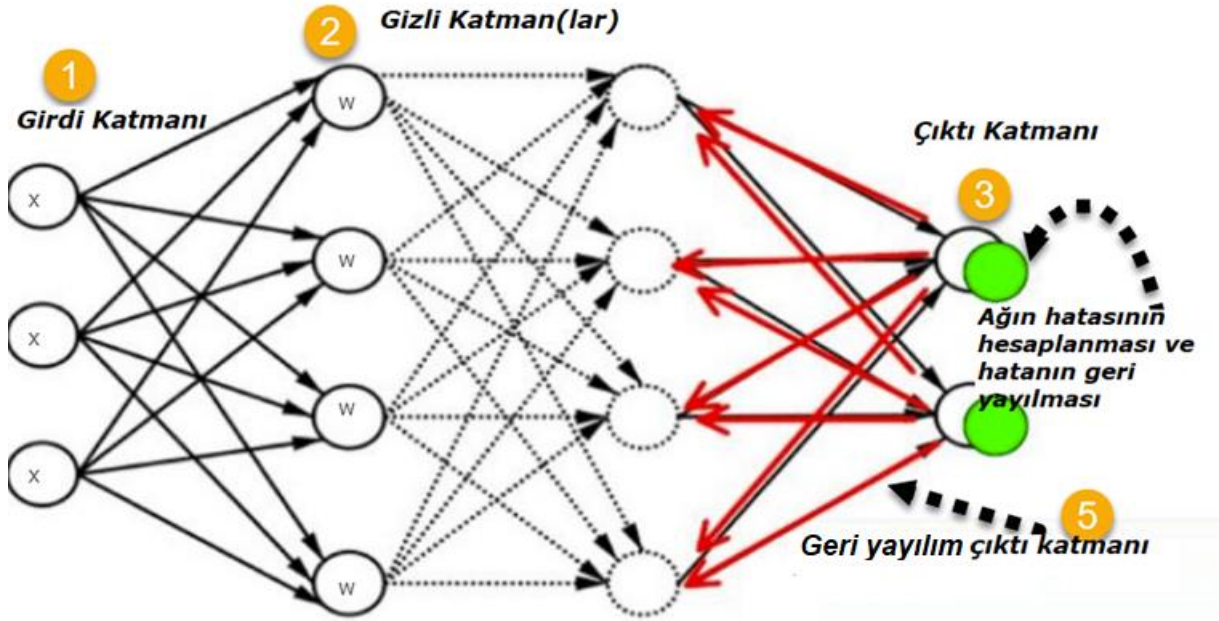
İleri beslemeli YSA bir veya birden fazla katmandan oluşan, her katmanda en az bir nöron bulunan, katmanlardan birinin girdi, birinin çıktı ve diğer katman veya katmanların ise gizli katman olarak adlandırıldığı ve çıktı katmanındaki nöronlarda da dönüştürme fonksiyonunun kullanılabildiği bir ağ olarak tanımlanabilir (Cura, 2008). Şekil 2.19'da tek katmanlı ve çok katmanlı ileri beslemeli YSA mimarileri grafiksel olarak gösterilmiştir.



Şekil 2. 19: Tek ve çok katmanlı ileri beslemeli YSA mimarileri (Haykin, 2009).

➤ Geri Beslemeli Yapay Sinir Ağı

Geri beslemeli yapay sinir ağları (GBYSA) da çok katmanlı ileri beslemeli YSA’nda öğrenme işlemini gerçekleştirmektedir (Han ve diğ. 2012). İlk olarak Hopfield tarafından kullanılan ve bu nedenle literatürde Hopfield ağları olarak da isimlendirilen GBYSA’larda ağdaki tüm nöronlar birbirleri ile bağlantılı olup, ağın proses elamanlarının çıktıları yine ağa belirli bir şekilde geri gönderilerek girdi olarak kullanılmaktadır (Hopfield, 1982; Cura, 2008; Öztemel, 2016). Algoritmanın çalışma prensibi özetlenecek olursa, ağın ürettiği çıktı değeri (gerçek çıktı), ağdan beklenen çıktı ile karşılaştırılır. Sonrasında ise, aralarındaki fark yani hesaplanan hata değeri, ağın ağırlıklarına geriye doğru yayılır ve ağırlıklar ise her bir iterasyonda güncellenir. Başka bir deyişle ağın ağırlık değerlerinin güncellenmesi bu şekilde gerçekleşmekte olup, ağdan optimum performans elde edildiğinde ise ağırlık güncelleme işlemleri durdurularak yeni tahminlerin yapılması bu optimum ağ vesilesiyle yerine getirilmektedir. Sonuç olarak ağın hatası istenilen değerden az olana kadar da bu işlem devam etmektedir (Cinsdikici, 2016; Özen ve Gülseçen, 2016).



Şekil 2. 20: Geri Beslemeli YSA Mimarisi (Johnson, 2022).

Han ve diğ. (2012) aktarımıyla GBYSA'nın oluşturulma ve eğitim süreci şu şekilde özetlenebilir:

Ağırlıkların Belirlenmesi: Ağdaki ağırlıklar, küçük rasgele sayılar olmak üzere rastgele belirlenir. Ayrıca, her birimin kendisiyle ilişkili bir önyargı (bias) değeri bulunmaktadır. Önyargı değerleri de benzer şekilde küçük rastgele sayılar olmak üzere rastgele atanır.

Girdi Katmanındaki Değerlerin İleriye Doğru Yayılması: İlk olarak, eğitim verisi ağın girdi katmanına verilerek beslenir. Sonrasında ise gizli ve çıktı katmanlarındaki her birime gelen net girdi, girdilerinin doğrusal bir kombinasyonu olarak hesaplanmaktadır. Yani birime gelen girdiler aslında bir önceki katmanda ona bağlı olan birimlerin çıktılarıdır. Birime gelen net girdiyi hesaplamak için birime bağlı her bir girdi, karşılık gelen ağırlığı ile çarpılır ve bu toplanır. Böylelikle, gizli ve çıktı katmanlarındaki her bir birimin net girdi ve çıktısı hesaplanmaktadır.

Denklem 2.58'de görüleceği üzere gizli veya çıktı katmanına gelen bir birim j ile temsil edilmek üzere ve j birimine net girdi, I_j olarak gösterilir ve;

$$I_j = \sum_i \omega_{ij} O_i + \theta_j \quad (2.58)$$

ω_{ij} , önceki katmandaki i biriminden j birimine olan bağlantının ağırlığı olup, O_i ise önceki katmandan gelen i biriminin çıktısı ve θ_j birimin etkinliğini değiştiren bir eşik değer olarak görev yapan önyargı (“bias”) değeri olarak tanımlanmaktadır.

Gizli ve çıktı katmanlarındaki her birim net girdisini aldıktan sonra bir aktivasyon fonksiyonu uygulanmaktadır. Fonksiyon, ilgili birim tarafından temsil edilen nöronun aktivasyonunu sembolize etmektedir ve aktivasyon fonksiyonu olarak genellikle Lojistik veya simoid fonksiyonu kullanılmaktadır. j birimi net I_j ile beslendiğinde, j biriminin çıktısı olan O_j Denklem 2.59’da gösterildiği gibi hesaplanır:

$$O_j = \frac{1}{1+e^{-I_j}} \quad (2.59)$$

Hatanın Geri Yayılması: Bu süreçte, ağırlık tahmininin hatasını yansıtmak için ağırlıklar ve önyargı değerleri güncellenerek hata geriye doğru yayılır. Çıktı katmanındaki bir birim j için Err_j hatası ise Denklem 2.60’da gösterildiği gibi hesaplanır:

$$Err_j = O_j(1 - O_j)(T_j - O_j) \quad (2.60)$$

Denklemden temsil edilen O_j j’nin gerçek çıktısı olup T_j ise verilen eğitim örneğinin bilinen sınıf etiketine dayalı olan gerçek çıktı olarak temsil edilmektedir.

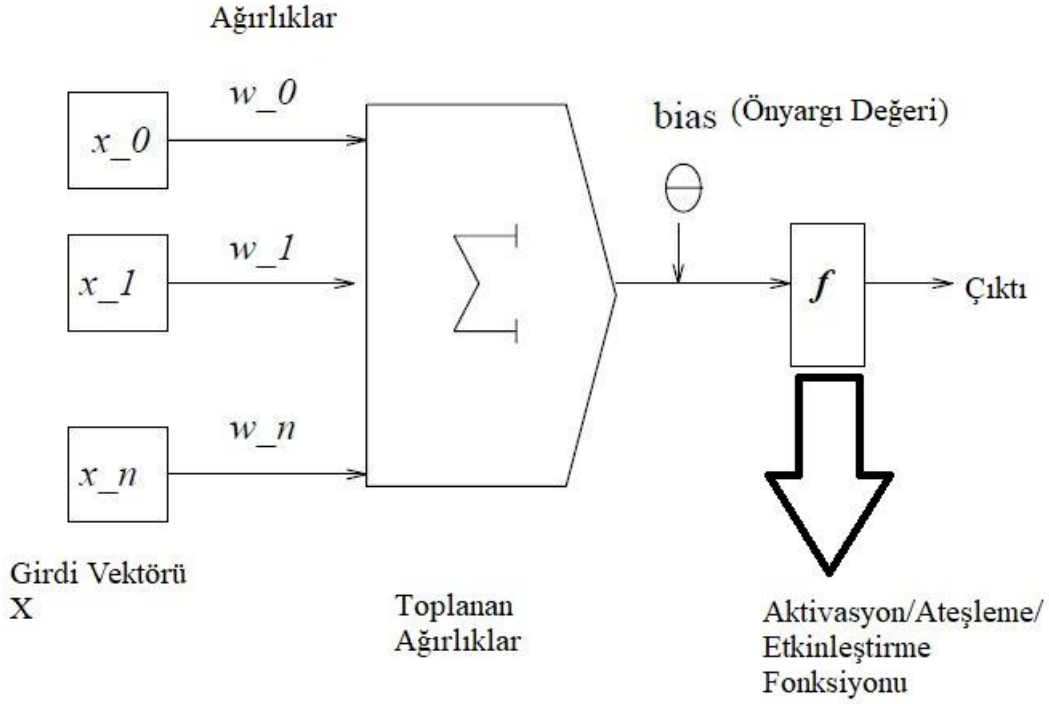
Gizli katman birimi j’nin hatasını hesaplamak için, bir sonraki katmanda j birimine bağlı birimlerin hatalarının ağırlıklı toplamı dikkate alınmakta olup, gizli katman birimi j’nin hatası ise Denklem 2.61’de verildiği gibi hesaplanmaktadır.

$$Err_j = O_j(1 - O_j) \sum_k Err_k \omega_{jk} \quad (2.61)$$

burada ω_{jk} , bir sonraki üst katmandaki j biriminden k birimine bağlantının ağırlığı ve Err_k ise k biriminin hatasıdır. Ağırlıklar ve önyargı değerleri, yayılan hataları yansıtacak şekilde güncellendiğinde Denklem 2.62 ve Denklem 2.63 denklemleri elde edilir. Şekil 2.21’de girdi ve çıktı katman birimleri gösterilmiştir.

$$\Delta\omega_{ij} = (l) Err_j O_i \quad (2.62)$$

$$\omega_{ij} = \omega_{ij} + \Delta\omega_{ij} \quad (2.63)$$



Şekil 2. 21: Gizli veya çıktı katmanı birimi: Ağırlıklı toplam oluşturmak için girdiler karşılık gelen ağırlıklarıyla çarpılır ve net girdiye de doğrusal olmayan bir aktivasyon fonksiyonu uygulanır (Han ve diğ., 2012).

l değişkeni, öğrenme oranı (“*learning rate*”) olarak tanımlanmakta olup tipik olarak 0 ile 1 arasında sabit bir değere sahiptir.

Geri yayılım, ağırlık sınıf tahminleri ile örneklerin gerçek sınıf etiketi arasındaki “*the mean squared distance*” değerini minimize etmek için verilen sınıflandırma problemini modelleyebilen bir ağırlık kümesini ararken “*gradient descent*” yani eğim azaltma yöntemini kullanmayı öğrenir. Öğrenme oranı, karar uzayında (yani, ağırlıkların yakınsadığı, ancak optimum çözüm olmadığı) yerel bir minimumda takılıp kalmaktan kaçınmaya yardımcı olur ve global minimumu bulmayı sağlamaktadır.

Önyargı (bias) değerlerinin güncellenmesi sonucu Denklem 2.64 ve Denklem 2.65 elde edilir.

$$\Delta\theta_j = (l) Err_j \quad (2.64)$$

$$\theta_j = \theta_j + \Delta\theta_j \quad (2.65)$$

Her örneğin tanıtılmasından sonra ağırlıklar ve önyargı (“*bias*”) değerleri güncellenmektedir. Belirlenen kriterler yerine getirildiğinde (önceki $\Delta\omega_{ij}$ değeri veya yanlış sınıflandırılan

örneklerin yüzdesi belirli bir eşik değerin altında kaldığında) öğrenme işlemi durdurulmaktadır (Han ve diğ.2012; Özen ve Gülseçen, 2016).

2.5.9 Aşırı Öğrenme Makineleri

Aşırı öğrenme makineleri ("*ELM*") temelde ileri beslemeli YSA mimarisinde olup bununla birlikte rastgele bağlantılara sahiptir. Ayrıca bu algoritma, hatanın geri yayılması ("*back-propagation*") yöntemini kullanmamaktadır. Bunun yerine, ağı eğitimi rastgele ağırlıklarla başlamakta ve ağırlıklar da en küçük kareler uyumuna göre tek bir adımda eğitilmektedir. Bu da daha az karmaşık bir ağ yapısı ile birlikte geri yayılım yöntemi ile oluşturulan ağlara göre öğrenmenin çok daha hızlı sürede sonuçlanmasını sağlamaktadır (Huang ve diğ. 2006; Leijnen ve Veen, 2020).

Geleneksel olarak, ileri beslemeli YSA'nın tüm parametrelerinin ayarlanması ("*tuning*") gerekmekte ve bu nedenle de farklı parametre katmanları (ağırlıklar ve önyargı- "*bias*" değerleri) arasında bağımlılığın söz konusu olduğunun altı çizilmiştir. Bununla birlikte, gradyan tabanlı "*gradient descent-based learning*" öğrenme yöntemlerinin, yanlış öğrenme adımları nedeniyle genellikle çok yavaş olduğu veya yerel- "*local*" minimumlara kolayca yakınsayabileceği açık bir şekilde görünmektedir. Yapay sinir ağlarının geleneksel yöntemlerle eğitilmesinin birkaç saat, birkaç gün ve hatta daha fazla zaman alabileceğinin şaşırtıcı olmadığı ve bu sürenin de olması gerekenden çok daha fazla olduğu ifade edilmiş ve bu algoritmaların eğitilmesinin yavaşlığına vurgu yapılmış olup bu konseptte dayanarak Huang ve diğ. (2006) yılında daha iyi genelleme performansı ve öğrenme hızı elde edilirken, geri yayılım ("*back-propagation*") algoritması gibi geleneksel ileri beslemeli YSA'lardan binlerce kat daha hızlı olabilen, aşırı öğrenme makinesi ("*ELM*") olarak adlandırılan "*Single hidden layer feedforward networks- (SLFNs)*" algoritmasını tanıtmışlardır. Ayrıca ELM, gizli katmanda eğitim verilerinden bağımsız ve "*randomly*" yani rastgele olarak belirlenen hesaplama düğümleri esaslı çalışmayı önermektedir (Huang ve diğ. 2006). Çatak (2015) ELM'nin en dikkat çekici özelliğinin, gizli katman parametrelerinin rastgele seçilmesi olduğunu ifade etmiştir.

Huang (2014) ELM algoritmasının temel olarak şu öğrenme problemine yanıt aramakta olduğunu ifade etmiştir:

Öğrenme (örneğin: kümeleme, regresyon ve sınıflandırma), gizli nöronlar (biyolojik nöronlar dahil) parametre ayarlaması yani “*tuning*” edilmeden ve bu nöronların çıktı özellikleri ve fonksiyon modellemesi bilinmese bile gerçekleştirilebilir mi?

Tek Gizli Katmanlı Aşırı Öğrenme Makineleri - “*Single hidden layer feedforward networks- (SLFNs)*”

Eğitim veri seti $D = \{(x_i, y_i) \mid i = 1, \dots, n\}$ $x_i \in \mathbb{R}^p$, $y_i \in \{1, 2, \dots, K\}$ bağımsız ve özdeş olarak örneklenmek üzere; ELM'nin amacı $f: X \rightarrow Y$ fonksiyonunu X örnekleri ve Y olası sınıf etiketleri üzerinden öğrenmektir. N adet gizli düğüme sahip tek gizli katmanlı ileri beslemeli sinir ağlarının (SLFN'ler) çıktı etiketi Denklem 2.66'daki gibi tanımlanabilir (Huang ve diğ. 2006).

$$f_N(X) = \sum_{i=1}^N \beta_i G(a_i, b_i, x), x \in \mathbb{R}^n, a_i \in \mathbb{R}^n \quad (2.66)$$

Burada a_i ve b_i , gizli düğümlerin öğrenme parametreleri olup β_i , i . gizli düğümü çıkış düğümüne bağlayan ağırlıktır. Genelleştirilmiş SLFN'ler için ELM'nin çıktı fonksiyonu şu şekilde tanımlanabilir:

$$f_N(X) = \sum_{i=1}^N \beta_i G(a_i, b_i, x) = \beta \times h(x) \quad (2.67)$$

İkili (“*binary*”) sınıflandırma uygulamaları için, ELM'nin karar fonksiyonu Denklem 2.68'deki gibi gösterilmektedir.

$$f_N(X) = \text{sign} \left(\sum_{i=1}^N \beta_i G(a_i, b_i, x) \right) = \text{sign}(\beta \times h(x)) \quad (2.68)$$

Denklem 2.67'in başka bir şekilde gösterimi ise Denklem 2.69'da gösterildiği gibidir.

$$H\beta = T \quad (2.69)$$

H ve T sırasıyla gizli katman matrisi ve çıktı matrisidir.

$$\beta = H^+ T \quad (2.70)$$

H^+ ise H matrisinin *Moore-Penrose* genelleştirilmiş tersi olup, gizli katman matrisi ise Denklem 2.71’de gösterilmiştir.

$$H(\tilde{a}, \tilde{b}, \tilde{x}) = \begin{bmatrix} G(a_1, b_1, x_1) & \cdots & G(a_L, b_L, x_1) \\ \vdots & \ddots & \vdots \\ G(a_1, b_1, x_N) & \cdots & G(a_L, b_L, x_N) \end{bmatrix}_{N \times L} \quad (2.71)$$

Çıktı matrisi ise Denklem 2.72’deki gibi tanımlanmaktadır:

$$T = [t_1, \dots, t_N]^T \quad (2.72)$$

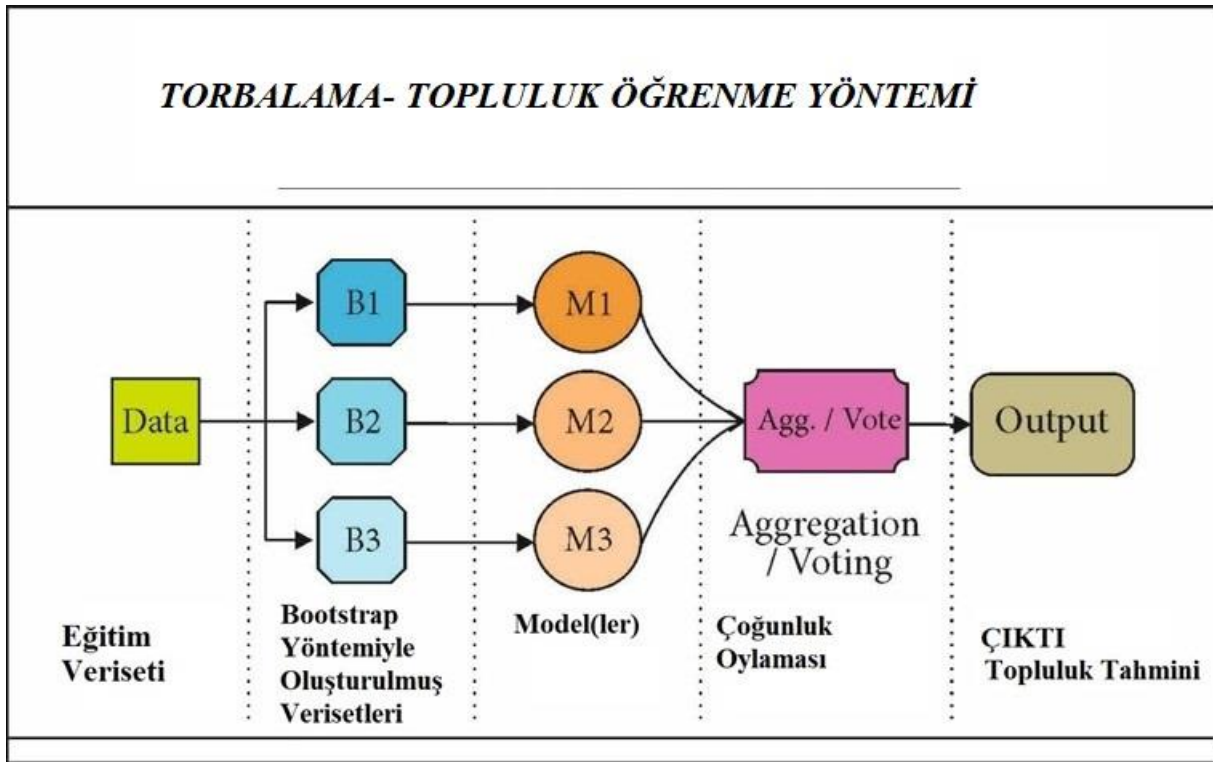
2.5.10 Topluluk Öğrenme Yöntemleri

İlk olarak otomatik karar destek sisteminin varyansını azaltmak ve böylece doğruluğunu artırmak için geliştirilen topluluk öğrenme yöntemleri, o zamandan beri özellik seçimi, güven aralığı tahmini, eksik nitelikler başa çıkmak, artımlı öğrenme, hata düzeltme, dengesiz hedef niteliğe sahip veri setleri gibi çeşitli makine öğrenmesi sorunlarını çözmek için başarıyla kullanılmaktadır. Birden fazla makine öğrenmesi yönteminin uygulanmasının neticesinde oluşan tahminlerin birleştirilerek sonuç olarak daha yüksek performansa sahip bir çıktı üretmeyi amaçlayan yöntemler topluluk öğrenme yöntemleri olarak tanımlanmaktadır. En önemli topluluk öğrenme yöntemleri torbalama - “*Bagging*”, artırma - “*Boosting*”, istifleme “*Stacking*” ve kombinasyon metodları “*combination methods- mixture of experts*” olarak isimlendirilen metodlardır (Polikar, 2012; Zhou, 2012b).

• Torbalama Topluluk Öğrenme Yöntemi

İlk olarak Breiman tarafından tanıtılan ve esasında eğitim veri setini manipüle etme stratejisi ile çalışan böylece her çalıştırmada m adet orijinal eğitim veri setinden rastgele seçilmiş başka bir ifade ile bootstrap örnekleme ile m adet örneğin eğitilmesi prensibine dayanmakta olan torbalama – “*Bagging*” öğrenme algoritması en bilinen topluluk öğrenme yöntemlerinden

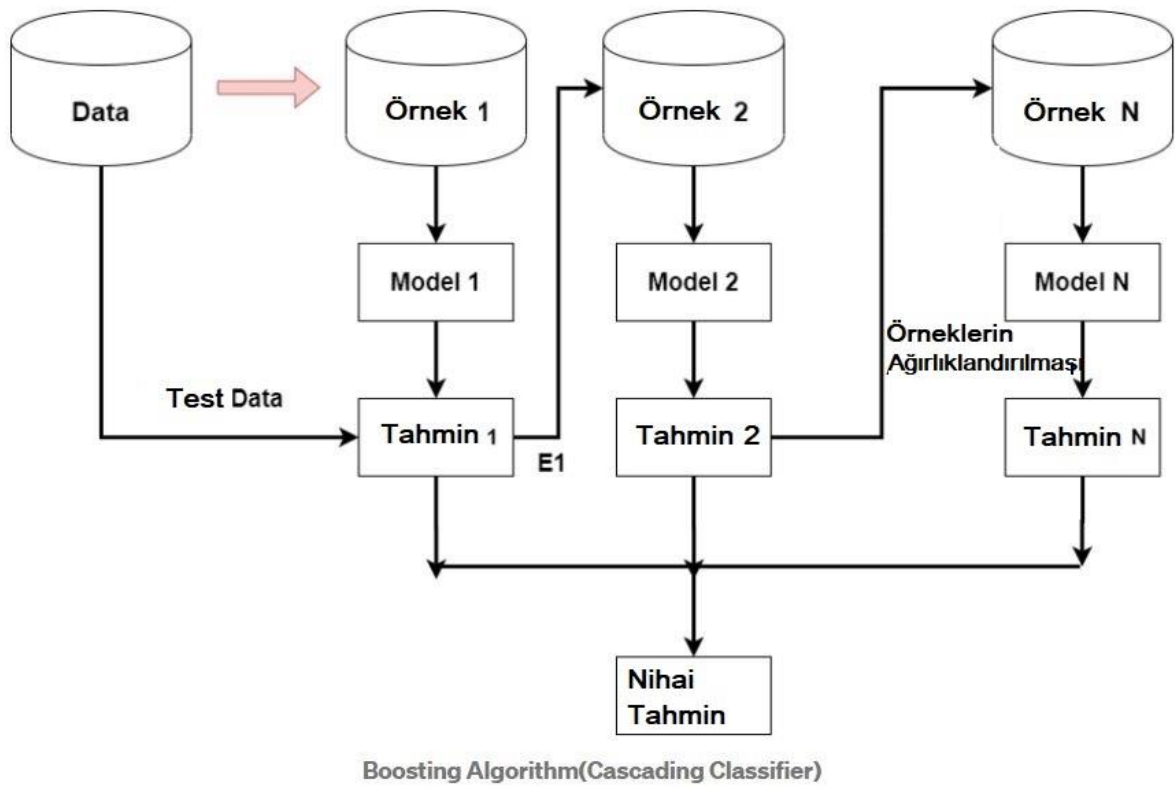
biridir (Breiman, 1996; Dietterich, 2000; Polikar, 2012; Doğaner, 2020). Bu yöntemle oluşturulmuş bir eğitim seti, orijinal eğitim setinin bir bootstrap kopyası- “*bootstrap replicate*” olarak adlandırılır ve orijinal eğitim veri setinin %63.2 sini içerdiği belirtilmiştir (Dietterich, 2000). Bu yöntemle elde edilen eğitim verisetleri ise bir sınıflandırma algoritması ile eğitilir ve sınıflandırma algoritmalarının tahminleri de çoğunluk oylaması- “*majority voting*” tekniğinden yararlanılarak topluluk öğrenme yönteminin nihai sınıflandırma tahmini olarak atanmaktadır. (Polikar, 2012; Doğaner, 2020). Şekil 2.22’de torbalama yönteminin şekilsel gösterimi görülmektedir.



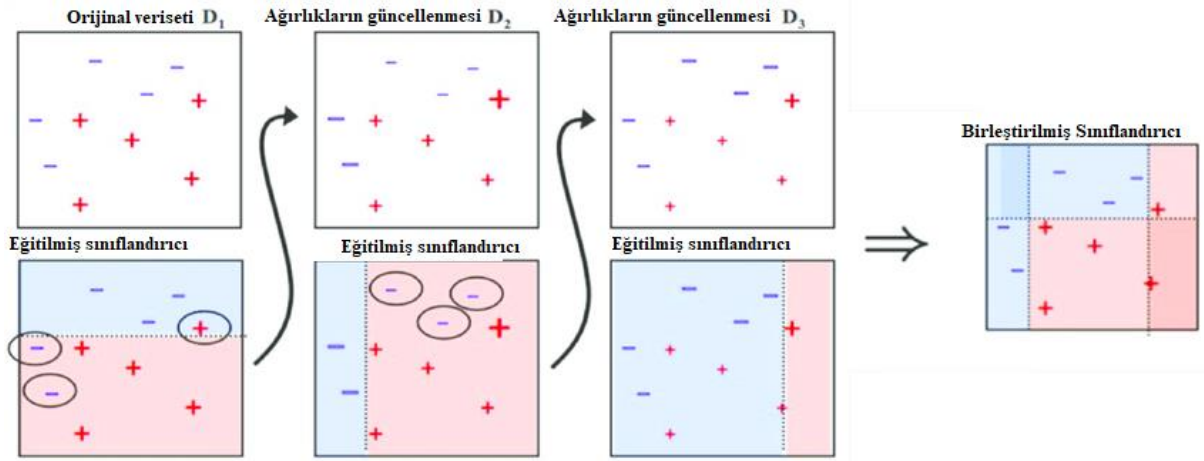
- **Artırma Topluluk Öğrenme Yöntemi**

Artırma- “*Boosting*” olarak bilinen topluluk öğrenme yöntemi, her biri aslında zayıf “*weak*” sınıflandırıcı olarak tanımlanan bir dizi sınıflandırıcı grubunu birleştirerek güçlü bir sınıflandırıcı oluşturan ve bu sayede de düşük eğitim hatası elde edebilen yinelemeli bir yaklaşım olarak tanımlanmaktadır. Artırma yöntemi çoğunluk oylamasını “*majority voting*” kullanarak bir dizi zayıf sınıflandırıcıyı birleştirirken, torbalama olarak bilinen “*bagging*” yönteminden ayrılmaktadır. Bu algorithmada oluşturulan ilk sınıflandırıcı, eğitim veri seti içerisinde rastgele seçilen örnekleri sınıflandırırken ikinci sınıflandırıcı, yarısı birinci sınıflayıcı tarafından yanlış sınıflandırılmış, diğer yarısı doğru sınıflandırılmış örnekler üzerine

eğitilip, son sınıflayıcı da önceki iki sınıflayıcının uyum göstermediği örneklerle eğitilir ve sonuç olarak çoğunluk oyunu alan sınıflandırıcının tahminleri, topluluk algoritmasının nihai tahmini olarak kabul edilir. Torbalamada sınıflandırıcıları eğitmek için seçilen örnekler rastgele seçilmekte ve bu örnekler her alt eğitim kümesinde eşit şansa sahip olup artırmada ise her bir sonraki sınıflandırıcı için eğitim veri seti, daha önce oluşturulmuş sınıflandırıcılar tarafından yanlış sınıflandırılan örneklerle giderek daha fazla odaklanmaktadır. (Polikar, 2012; Zhou, 2012c; Doğaner, 2020). “AdaBoost” ve “LogitBoost” algoritmaları en bilindik boosting algoritmaları arasındadır (González ve diğ. 2020). Şekil 2.23 ve Şekil 2.24’te artırma topluluk öğrenme yönteminin şekilsel olarak gösterimi görülmektedir.



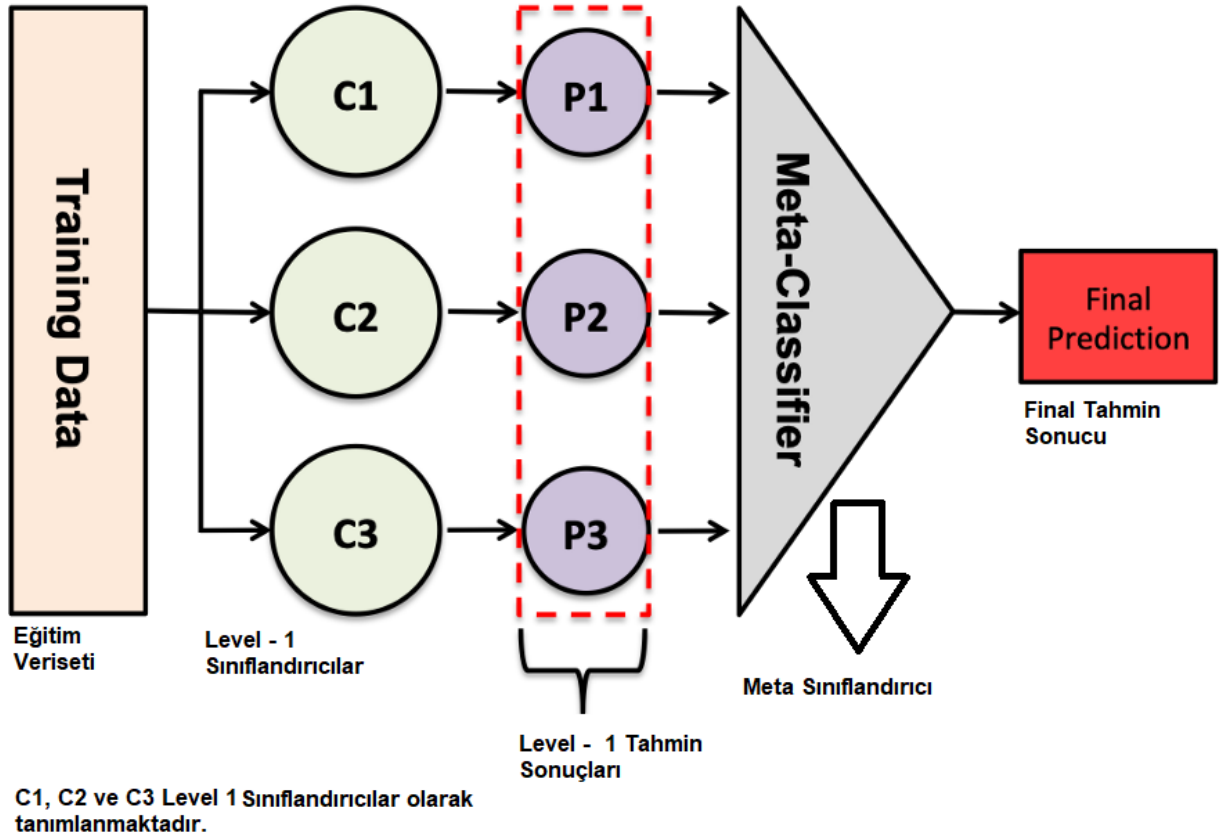
Şekil 2.23: Artırma- “Boosting” Topluluk Öğrenme Yöntemi’nin şekilsel olarak gösterimi (Patel, 2020).



Şekil 2.24: Artırma- “Boosting” Topluluk Öğrenme Yöntemi’nin şekilse olarak gösterimi (Patel, 2020).

• İstifleme Topluluk Öğrenme Yöntemi

İstifleme ya da “*Stacking*” farklı sınıflandırıcıların kombinasyonu sonucu yeni bir sınıflandırıcı elde edilmesi prosedürü olarak tanımlanmakta olup ilk olarak Wolpert (1992) tarafından önerilen bir yöntemdir. Burada, bireysel öğrencilere– “*learners*” ya da sınıflandırıcılara birinci seviye öğrenciler ya da sınıflandırıcılar adı verilirken, bu sınıflandırıcıları kombine eden birleştirici ise ikinci seviye öğrenci veya meta-sınıflandırıcı olarak tanımlanmıştır (Zhou, 2012). İki etaptan oluşan “*stacking*” metodunda farklı sınıflandırma algoritmalarının aynı eğitim veri setine uygulanması sonucu elde edilen tahminler tutulmakta ve ikinci aşamada ise bu tahminler meta sınıflandırıcı tarafından tekrar işlenerek topluluk öğrenme modelinin nihai tahmini bulunmaktadır (Doğaner, 2020). Daha sade bir anlatımla istifleme yönteminde, birden çok birinci düzey sınıflandırıcının tahminlerinin bir meta sınıflandırıcıyı eğitmek için yeni özellikler olarak kullanılmaktadır. Bununla birlikte meta sınıflandırıcının, kullanıcının seçtiği herhangi bir sınıflandırıcı olabileceği ifade edilmiştir (Ceballos, 2020). Şekil 2.25’te istifleme yönteminin şekilsel olarak gösterimi görülmektedir.

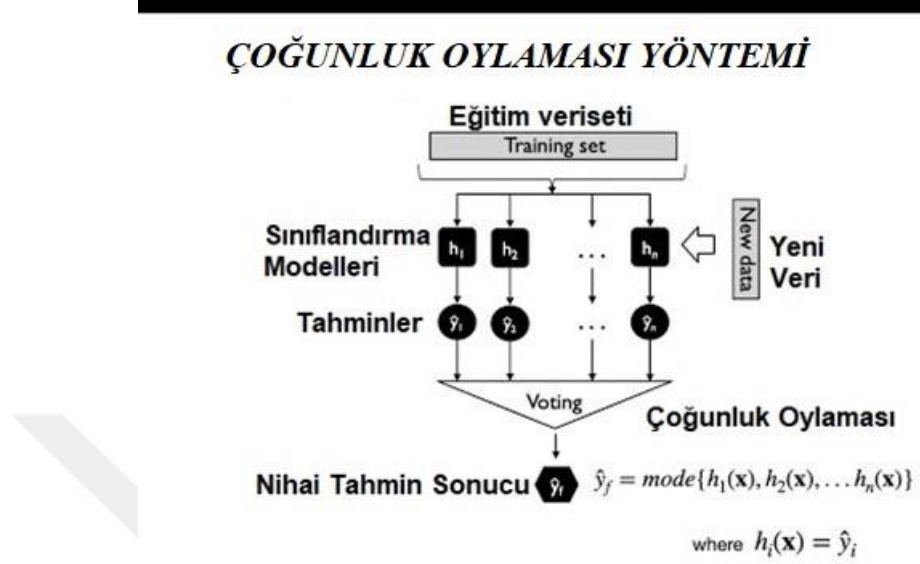


Şekil 2.25: İstifleme – “Stacking” Topluluk Öğrenme Yönteminin şekilsel gösterimi (Ceballos, 2020).

- **Kombinasyon Yöntemler**

Bir dizi makine öğrenmesi modelleri oluşturduktan sonra, en iyi makine öğrenmesi algoritmasını bulmaya çalışmak yerine, güçlü bir genelleme yeteneği elde etmek için kombinasyon yöntemlerine başvurulabilmektedir (Dietterich, 2000). Literatürde çok sayıda birleştirme yöntemi bulunmasına karşın en önemli teknikler sayısal çıktı nitelikleri için ortalama alma – “*simple averaging*”, farklı önem taşıyan farklı ağırlıklara sahip bireysel sınıflandırıcıların tahmin sonuçlarının yani çıktıların ortalamasını alarak birleşik sonucu elde eden ağırlıklı ortalama alma- “*weighted averaging*”, ve nominal çıktılar için en popüler ve temel kombinasyon yöntemi olan oylama – “*voting*” yöntemleri sıralanabilir. Birleştirme yani kombinasyon tekniğini kullanan yöntemlerde oylama – “*voting*” tekniği oldukça popüler yöntemlerden biridir. Oylama yönteminin de kendi içinde farklı sınıfları bulunmaktadır ve bunlar arasında en bilinenlerden ve popüler olan yöntemlerden birisi de çoğunluk oylaması “*majority voting*” tekniğidir (Zhou, 2012). Bu yöntemde kullanılan her sınıflandırma algoritmasının bir tahmin sonucu bulunmakta olup, sonuçlarda son kararın kullanılan sınıflandırıcıların yarısından fazlasından atanması durumunda bu karar nihai karar olarak kabul

edilmektedir. Ayrıca, tahmin sonuçlarından hiçbirini sınıflandırıcıların oyların yarısından fazlasını alamazsa, ret seçeneği verilmekte ve hiçbir tahminde bulunulmamaktadır (Zhou, 2012; Doğaner, 2020). Şekil 2.26’da çoğunluk oylaması yönteminin şekilsel gösterimi verilmiştir.



Şekil 2. 26: Çoğunluk Oylaması – “Majority Voting” şekilsel gösterimi (Raschka, 2020).

2.5.11 Extreme Gradient Boosting (“Xgboost”) Algoritması

Gerçek hayat problemlerinde kullanılan makine öğrenmesi algoritmaları arasında, “*gradient tree boosting*” yönteminin birçok uygulamada öne çıkan tekniklerden biri olduğu ifade edilmektedir (Friedman, 2001). Literatürde öne çıkan bu topluluk öğrenme algoritmaları arasında geleneksel “*gradient boosting machines (GBM)*”, “*XGBoost*” ve bu algoritmanın varyantı “*XGBoost with DART*”, “*LightGBM*” ve bu algoritmanın varyantı “*LightGBM with GOSS*” ve “*CatBoost*” bulunmaktadır (González ve diğ. 2020).

XGBoost algoritmasının ise en ünlü gradyan boosting tabanlı yöntem olduğu ve muhtemelen de en başarılı makine öğrenmesi algoritmalarından biri olduğu, son derece doğru ve ölçeklenebilir bir ağaç boosting sistemi olmakla birlikte öğrenme sürecini maksimuma çıkarmak için sistemin bilgi işlem ve bellek kapasitelerinden tam olarak yararlanmaya odaklandığı ayrıca “*overfitting*” problemini azaltmak ve her tipteki makine öğrenmesi problemine uyarlanılabilecek biçimde geliştirildiğinin altı çizilmiştir (González ve diğ. 2020). Bu nedenle, bu doktora tez çalışmasında topluluk öğrenme yöntemlerinden XGBoost algoritması kullanılmış ve bu nedenle de bu algoritma detaylı olarak ele alınmıştır.

Denetimli öğrenmede, amaç fonksiyonu ve tahmin fonksiyonu kullanılmakta olup, eğitim parametreleri, amaç fonksiyonunu minimize etmek ve ilgili parametreleri öğrenmek için faydalanmakta ve sonrasında ise tahmin fonksiyonu ve elde edilen parametreler, bilinmeyen örneği sınıflandırmak veya sayısal olarak tahmin etmek için kullanılmaktadır. Toplamda iki boyuttan oluşan (n ve m olmak üzere) D veri seti; $D = \{(x_i, y_i)\}, (x_i \in \mathbb{R}^m, y_i \in \mathbb{R}, i = 1, 2, \dots, n)$ olarak tanımlandığında XGBoost modeli Li ve Zhang (2020) aktarımıyla Denklem 2.73’de gösterilmiştir.

$$\hat{y} = \sum_{k=1}^K f_k(x_i), \quad f_k \in F \quad (i = 1, 2, \dots, n) \quad (2.73)$$

ve $F = \{f(x) = w_{q(x)}\} (q: \mathbb{R}^m \rightarrow \{1, 2, \dots, T\}, w \in \mathbb{R}^T$ ise CART karar ağacı yapısı kümesinin bulunduğu q , yaprak düğümlerinin örnek haritasının ağaç yapısı; T , yaprak düğümlerinin sayısı ve w , yaprak düğümlerinin gerçek puanıdır.

XGBoost modeli oluşturulurken optimal bir modelin kurulabilmesi için amaç fonksiyonunun minimize edilmesi ilkesine göre optimal parametrelerin bulunması gerekliliği bulunmaktadır. XGBoost modelinin amaç fonksiyonu, bir hata fonksiyonu terimi L ve bir model karmaşıklık fonksiyonu terimi Ω ’e bölünebilir. Amaç fonksiyonu ise Denklem 2.74’te gösterilmiştir. Amaç fonksiyonuna bağlı olarak Denklem 2.75 ve Denklem 2.76 elde edilir.

$$Obj = L + \Omega \quad (2.74)$$

$$L = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.75)$$

$$\Omega = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (2.76)$$

γT L1’in sabit terimi olup, $\frac{1}{2} \lambda \sum_{j=1}^T w_j^2$ ise L2’nin sabit terimi olarak tanımlanmaktadır. Önemle belirtmek gerekir ki, γ ve λ modelin “overfitting” olmasını önleyen düzeltme parametreleridir.

Eğitim veri seti modeli eğitmek ve parametreleri optimize etmek için kullanılırken, orijinal modeli değişmemesi ve amaç fonksiyonunu mümkün olduğunca minimize etmek için de modele yeni bir f fonksiyonu eklenmektedir. Sonuç olarak 2.77’de yeniden yazılan amaç fonksiyonu ise gösterilmiştir.

$$Obj^{(t)} = \sum_{i=1}^n \left(y_i - \left(\hat{y}_i^{(t-1)} + f_t(x_i) \right) \right)^2 + \Omega \quad (2.77)$$

$\hat{y}_i^{(t-1)}$ değeri (t-1)'inci modelin tahmin edilen değeri; $f_t(x_i)$ değeri de t'inci zamanda elde edilen yeni fonksiyon olarak tanımlanırken; *Obj*, modeli değerlendirmek için kullanılabilecek bir fonksiyondur. *Obj* değeri ne kadar küçük olursa, modele olan etkisi o kadar iyi olacağı belirtilmiştir.

Böylece yukarıda verilmiş olan ağaç oluşturma yöntemini yinelemeli olarak çağırarak, çok sayıda regresyon ağacı yapısı elde edilebilmekte ve *Obj* fonksiyonu da optimal ağaç yapısını aramak ve optimal bir *XGBoost* modeli oluşturmak için mevcut modele koymak için kullanılmaktadır (Li and Zhang, 2020).

2.5.12 Danışmansız Öğrenme

Literatürde temel olarak iki farklı danışmansız öğrenme çeşidi tanımlanmış olup bunlar; kümeleme- “*clustering*” ve “birliktelik kuralları” olarak tanımlanmaktadır. Doktora tez çalışmasında ele alınan problem sınıflandırma problemi olduğundan mütevellit, danışmansız öğrenme ve pekiştirmeli öğrenme stratejilerinin teknik detaylarına girilmemiştir.

2.5.12.1 Kümeleme

Hedef niteliğin olmadığı veya bilinmediği durumlarda kullanılabilen danışmansız öğrenme stratejileri, girdi değerleri esas alınarak bunların gruplanmasını ya da kümelenmesini – “*clustering*” sağlamaktadır (Balaban ve Kartal, 2018). Danışmansız öğrenme stratejilerinden olan kümeleme analizinin, sınıflandırmadan en büyük farkı; sınıflandırma prosedüründe, sınıflar önceden belirli olup, kümeleme analizinde sınıflar belirli değildir (Dunham, 2003; Silahtaroglu, 2016).

Şaşırtıcı olmayan bir şekilde, kümeleme tekniklerinin optimizasyon da içerdiği çünkü kümelemenin amacının, her grup içindeki üyelerin maksimum benzerliğe ve gruplar arasındaki üyelerin minimum benzerliğe sahip olması için gruplar oluşturmak olduğu vurgulanmıştır. İşletmelerin özellikle müşterileri segmentlere ayırmada (müşteri segmentasyonu) kümeleme tekniğinden faydalandıkları belirtilmiş olup ve uygun pazarlama ürünlerini ayrılmış gruplara doğru zamanda, doğru formatta, doğru fiyata yönlendirmede yine bu tekniğin kullanılabileceğinin altı çizilmiştir (Turban ve diğ. 2015).

2.5.12.2 Birliktelik Kuralları

En sade hâli ile veri madenciliği sürecindeki birliktelik kuralları, hangi niteliklerin “*birlikte gittiğini*” bulma işi olarak tanımlanmaktadır (Larose ve Larose, 2015). Başka bir ifade ile birliktelik kuralları ya da ilişki analizi; veritabanındaki bir dizi kaydın diğer kayıtlarla olan bağlantısını açıklayan işlemler dizisi olarak tanımlanmıştır (Silahtaroglu, 2016) .

Bu kurallar pazarlama alanında sıklıkla kullanılmakta olup, “**pazar-sepet analizleri**” yöntemi yardımıyla müşterilerin alışveriş alışkanlıkları belirlenir ve bir müşteri herhangi bir ürünü aldığı anda, sepetine başka hangi ürünleri de koyduğu olasılıksal olarak ortaya konmaktadır. Böylelikle birlikte satılan ürünler de belirlenerek mağaza rafları düzenlenebilmekte, müşterilerin de bu ürünlere daha kolay erişebilmeleri sağlanmaktadır (Özkan, 2016).

2.5.13 Pekiştirmeli Öğrenme

Pekiştirmeli öğrenme, çevresini algılayan ve hareket eden özerk bir ajanın, hedeflerine ulaşmak için en uygun eylemleri seçmeyi nasıl öğrenebileceği sorusunu ele almaktadır. Aslında oldukça genel olan bu problem, bir mobil robotu kontrol etmeyi öğrenmek, fabrikalardaki operasyonları optimize etmeyi öğrenmek veya bir oyunu oynamayı öğrenmek gibi görevleri kapsamaktadır. Bu öğrenme stratejisinin temelinde bir öğretmenin öğrenen sisteme ödül ya da ceza vererek öğrenme sürecine yardımcı olması bulunmaktadır (Mitchell, 1997; Öztemel, 2016).

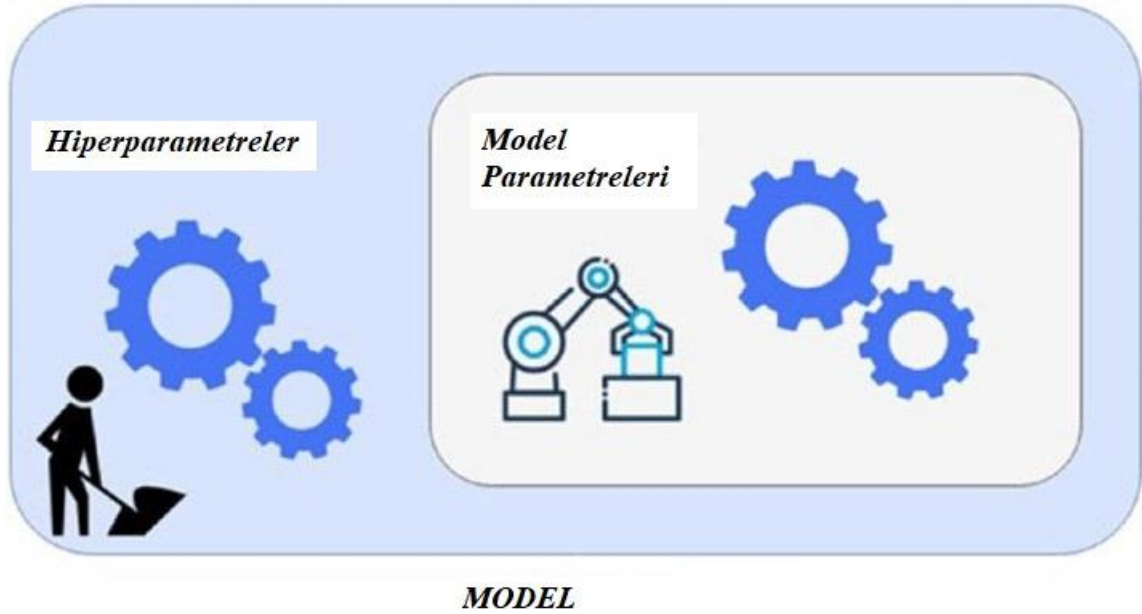
2.6 MAKİNE ÖĞRENMESİNDE HİPERPARAMETRE OPTİMİZASYON TEKNİKLERİ

Temel olarak optimizasyon; sayıların, fonksiyonların veya sistemlerin uç noktalarını (minimum ve maksimum) bulma ile ilgili matematiksel bir disiplin olarak tanımlanmaktadır (Kiranyaz ve diğ. 2014).

Makine öğrenmesi modeli oluştururken doğru hiperparametreleri seçmek, veri bilimi alanında çalışan araştırmacıların karşılaştığı en büyük sorunlardan biri olduğu ifade edilmiştir (Agrawal, 2021). Bu nedenle öncelikle parametre ve hiperparametre ayrımının bilinmesi gerekmektedir.

Parametre(ler): Algoritmanın sağlanan veri setine ayarladığı parametreler olarak tanımlanmakta olup bu ayarda araştırmacının müdahale edebileceği bir durum bulunmamaktadır.

Hiperparametre(ler): Bunlar, eğitime başlamadan önce araştırmacının manuel olarak ayarladığı, veri setinin özellikleri ve algoritmanın öğrenme kapasitesi gibi niteliklere dayanan üst düzey parametreler olarak tanımlanmaktadır (Agrawal, 2021).



Şekil 2.27: Model parametreleri ve hiperparametreler arasındaki farkın şekilsel gösterimi (Agrawal, 2021)

Makine öğrenmesi modelleri oluştururken en iyi hiperparametre yapılandırması/optimizasyonunun gerçekleştirilmesi, modelin performansı üzerinde doğrudan bir etkiye sahip olduğunun altı çizilmiştir. Hiperparametre optimizasyonunun gerçekleştirilebilmesi için de öncelikle makine öğrenmesi algoritmaları, hiperparametreleri ve bu algoritmalara uygun hiperparametre optimizasyon teknikleri ve bu hiperparametre optimizasyon tekniklerinin algoritmanın performansını nasıl etkilediği hakkında derinlemesine araştırma yapılması ve bilgi sahibi olunması gerektiği ifade edilmiştir. Uygulanacak optimizasyon tekniğinin ise farklı problemlere uygulandıklarında ise kendine özgü avantaj ve dezavantajları bulunduğundan dolayı da yine algoritmanın özelliklerine göre seçilmesi gerekmektedir. (Yang ve Shami, 2020; Agrawal, 2021).

Yang ve Shami (2020) makine öğrenmesinde kullanılan hiperparametre optimizasyon tekniklerini temel olarak beş başlık altında toplamıştır. Bunlar Model Kullanmayan Algoritmalar, Gradyan Tabanlı Yöntemler, Bayesian Optimizasyon Teknikleri, Çoklu Doğruluk Dereceli Optimizasyon Algoritmaları (“*Multi Fidelity Optimization Algorithms*”) ve

Metasezgisel Optimizasyon Algoritmaları olarak sınıflandırılmıştır. Doktora tez çalışması kapsamında Model Kullanmayan Algoritmalar arasından “*Grid Search*” ve Metasezgisel Optimizasyon Algoritmalarından Genetik Algoritma ve Parçacık Sürüsü Optimizasyonu kullanıldığı için bu yöntemler detaylı olarak açıklanmıştır.

2.6.1 Algoritmaların Hiperparametre Arama Uzayı

Hiperparametreleri bulunan ve bu nedenle de optimizasyon gerektiren algoritmaların hiperparametreleri, tipleri ve arama uzayları Tablo 2.5’te detaylı olarak ele alınmıştır.

Tablo 2.5: Hiperparametre optimizasyon yapılabilecek makine öğrenmesi algoritmaları, hiperparametre tipi ve arama uzayı (Yang ve Shami, 2020)

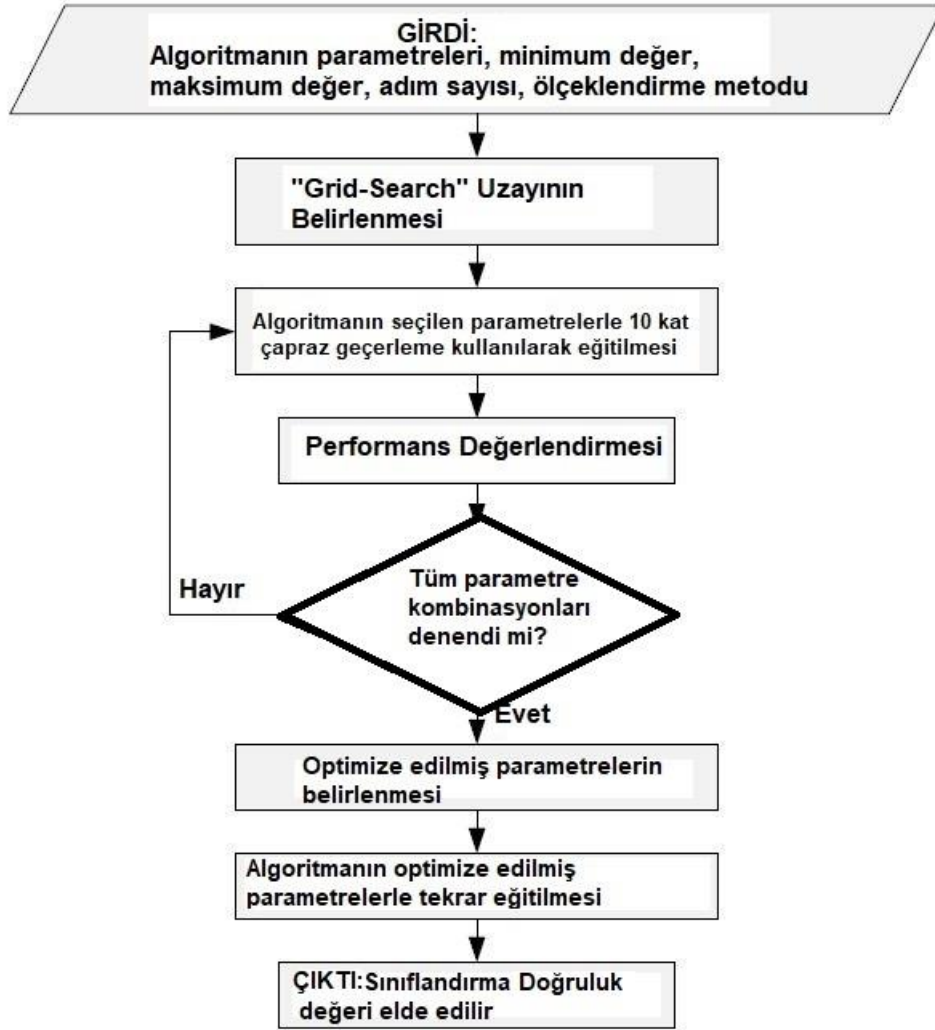
Makine Öğrenmesi Modeli	Hiperparametreler	Hiperparametre Tipi	Arama Uzayı
Rastgele Orman Algoritması	<i>n_estimators</i>	Kesikli	[10,100]
	<i>max_dept</i>	Kesikli	[5,50]
	<i>min_samples_split</i>	Kesikli	[2,11]
	<i>min_samples_leaf</i>	Kesikli	[1,11]
	<i>Criterion</i>	Kategorik	“ <i>gini,entropy</i> ”
	<i>max_features</i>	Kesikli	[1,64]
Destek Vektör Makineleri- “ <i>DVM</i> ” Sınıflandırma Algoritması	<i>C</i>	Sürekli	[20.1,50]
	<i>Kernel</i>	Kategorik	“‘ <i>linear</i> ’, ‘ <i>poly</i> ’, ‘ <i>rbf</i> ’, ‘ <i>sigmoid</i> ’”
K-En Yakın Komşu Algoritması	<i>n_neighbors</i>	Kesikli	[1,20]
Yapay Sinir Ağları – “ <i>Multilayer Perceptron Classifier</i> ”	“ <i>optimizer</i> ”	Kategorik	“‘ <i>adam</i> ’, ‘ <i>rmsprop</i> ’, ‘ <i>sgd</i> ’”
	“ <i>activation</i> ”	Kategorik	“‘ <i>relu</i> ’, ‘ <i>tanh</i> ’”
	“ <i>batch_size</i> ”	Kesikli	[16,64]
	“ <i>neurons</i> ”	Kesikli	[10,100]
	“ <i>epochs</i> ”	Kesikli	[20,50]
	“ <i>patience</i> ”	Kesikli	[3,20]

2.6.2 Model Kullanmayan Algoritmalar

Herhangi bir model kullanmayan algoritmalar arasında ızgara araması – “*grid search*” ve rastgele arama- “*random search*” yöntemleri bulunmaktadır (Yang ve Shami, 2020; Agrawal, 2021).

2.6.2.1 Izgara Araması

Izgara araması –“*grid search*”, hiperparametre uzayının tanımlanan alt kümesine dayanan kapsamlı bir aramadır. Hiperparametreler, minimum değer (alt sınır), maksimum değer (üst sınır) ve adım sayısı kullanılarak belirtilmektedir. Her kombinasyonun performansı, belirlenen performans ölçütleri kullanılarak değerlendirilmektedir (Syarif ve diğ. 2016). Bu yaklaşımın, en iyi hiperparametre kombinasyonunu bulmanın en kesin yolu olduğu, ancak uzun zaman gerektirmesi ve kaynak tüketiminin yüksek olması gibi çeşitli dezavantajları bulunmaktadır (Agrawal, 2021). Belki de en önemlisi, bu yöntemin temel dezavantajı, hiperparametrelerin sayısı arttıkça değerlendirmelerin sayısı da katlanarak arttığından, yüksek boyutlu hiperparametre konfigürasyon uzayı için verimsiz olmasıdır. Literatürde bu üstel büyüme, boyutsallığın laneti “*curse of dimensionality*” olarak adlandırılmaktadır (Claesen ve diğ. 2014; Yang ve Shami, 2020).



Şekil 2.28: Izgara Arama yöntemi kullanılarak hiperparametre optimizasyonunun gerçekleştirilmesi (Syarif ve diğ, 2016)

2.6.3 Meta Sezgisel Optimizasyon Algoritmaları

Meta-sezgiseller kombinatoriyal optimizasyon için nispeten yeni bir yaklaşım olup, zor kombinatoriyal optimizasyon problemlerinin yüksek kaliteli çözümlerini bulmak için kullanılabilen genel bir algoritmik şablon olarak tanımlanmaktadır (Birattari, 2009). Meta-sezgisellerin en büyük avantajlarından birisi probleme veya alana özgü olmamaları ve herhangi bir optimizasyon problemine uygulanabilir olmalarıdır. Meta-sezgisel optimizasyon algoritmaları çok sayıda değişken ve doğrusal olmayan amaç fonksiyonları ile “NP-hard” problemlerini verimli bir şekilde çözmek için kullanılabilirler. (Gogna ve Tayal, 2013). Evrimsel algoritmalar olarak da bilinen meta-sezgisellerin çözüme yaklaşma yeteneklerinin

yüksek olmasının baş nedeni, doğadaki değişim kabiliyetlerini taklit ederek belirli bir uygunluk fonksiyonunu en iyilemeye çalışmaları olarak belirtilmiştir (Özkır, 2019).

Popülasyon tabanlı optimizasyon algoritmaları; genetik algoritma, evrimsel algoritmalar, evrimsel stratejiler ve parçacık sürüsü optimizasyonu dahil olmak üzere büyük bir metasezgisel algoritma türüdür. Popülasyon tabanlı optimizasyon algoritmalarının çalışma prensibinde her nesilde bir popülasyon oluşturarak ve güncelleyerek başlaması; sonrasında ise her nesildeki her bireyin global optimum belirlenene kadar değerlendirilmesi yer almaktadır. Farklı popülasyon tabanlı optimizasyon algoritmaları arasındaki temel farklar, popülasyonları oluşturmak ve seçmek için kullanılan yöntemler olarak belirtilmiştir. Genetik algoritma ve parçacık sürüsü optimizasyonu algoritmalarının, hiperparametre optimizasyon problemleri için yaygın olarak kullanılan iki temel popülasyon tabanlı algoritma olduklarının altı çizilmiştir (Yang ve Shami, 2020; Yao ve diğ. 2019).

2.6.3.1 Genetik Algoritma

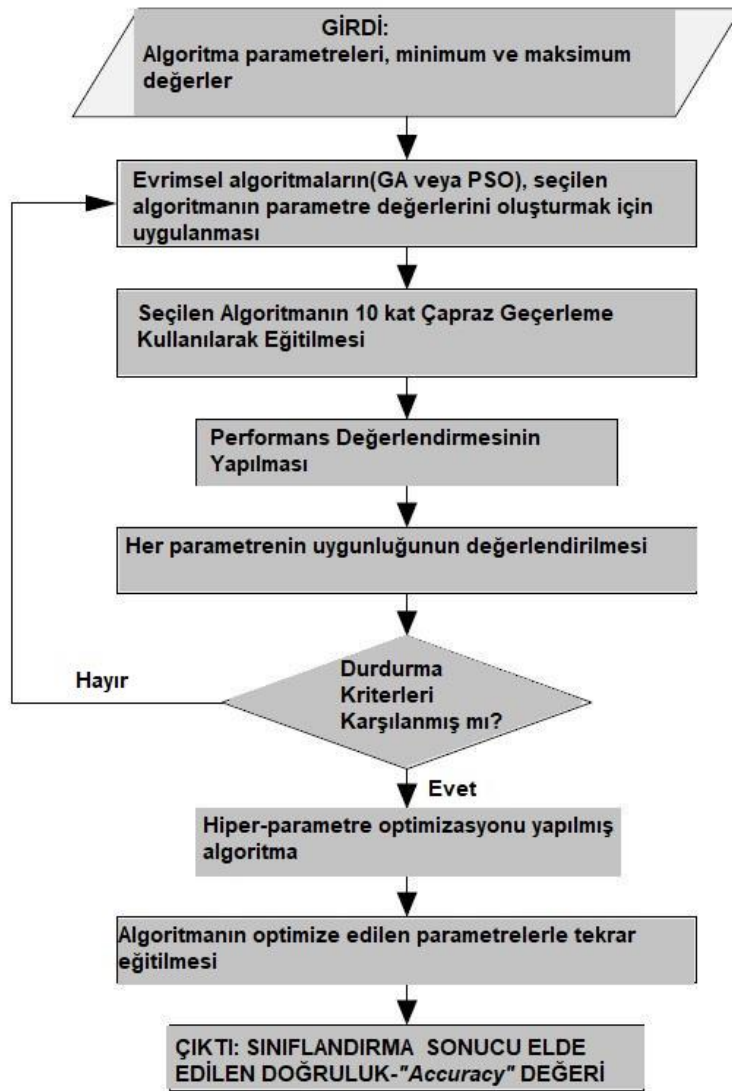
İlk olarak 1973 yılında John Holland tarafından önerilen Genetik Algoritma (GA), biyolojik evrimi yönlendiren süreç olan doğal seçilime dayalı optimizasyon problemlerini çözmek için kullanılan bir yöntemdir (Holland, 1973; Syarif ve diğ. 2016).

Genetik algoritma (GA), doğal seçilimin işlediği süreçleri hesaplamalı olarak taklit etmeye ve bunları işletme ve araştırma problemlerini çözmek için uygulamaya çalışmaktadır. GA eş seçimi, üreme, mutasyon ve genetik bilginin çaprazlanması gibi biyoloji bilim dalından esinlenilmiş faktörlerin etkilerini incelemek için bir çerçeve sağlamaktadır. GA tarafından kullanılan seçme-“*selection*”, çaprazlama-“*crossover*” ve mutasyon-“*mutation*” olmak üzere üç operatör vardır. (Larose ve Larose, 2015).

GA, popülasyon-“*population*” adı verilen potansiyel çözümleri yinelemeli olarak güncelleyerek çalışmaktadır. Popülasyonun her üyesi, her döngüde uygunluk açısından değerlendirilir. Gelecek nesil, ebeveynlerinin özelliklerini de miras alarak daha iyi veya daha kötü bireyleri içerebilmektedir. Böylelikle, daha iyi bireylerin hayatta kalma ve daha yetenekli yavrulara sahip olma olasılıkları daha yüksek olurken, daha kötü bireyler yavaş yavaş ortadan kaybolmaktadır. Başka bir deyişle yeni bir popülasyon belirtilen operatörleri kullanmak suretiyle eski popülasyonun yerini almakta ve üreme veya klonlama için en uygun üyeler seçilmektedir. Birkaç nesil sonra, en iyi uyum yeteneğine sahip birey, global optimum olarak tanımlanmaktadır. Uygunluk fonksiyonu $f(x)$, gen üzerinde değil, kromozom (potansiyel

çözüm) üzerinde çalışan gerçek bir fonksiyondur, böylece $f(x)$ 'deki x , uygunluk değerlendirmesi sırasında kromozom tarafından alınan sayısal değeri ifade etmektedir (Larose ve Larose, 2015; Itano ve diğ. 2018).

GA'nın parametre optimizasyonu için de kullanılabileceği ve algoritmanın çeşitli parametrelerinin en iyi aralıkları ve bağımlılıkları hiç bilinmediğinde uygulamak için çok elverişli olduğu ifade edilmektedir. GA'nın özellikle zaman açısından, ızgara araması-“grid-search” yöntemine göre ızgara aramasının çok fazla parametre kombinasyonu denediğinden dolayı uzun zaman alması nedeniyle daha uygun olabileceği belirtilmektedir (Syarif ve diğ. 2016).



Şekil 2. 29: Evrimsel Algoritma kullanılarak hiper-parametre optimizasyonu süreci
(Syarif ve diğ., 2016)

2.6.3.2 Parçacık Sürü Optimizasyonu Algoritması

Çok sayıda basit kurallarla hareket eden parçacığın karmaşık davranışları oluşturulma amacıyla kullanılması fikri ilk olarak Reeves tarafından 1983’de doğal olayları taklit eden bilgisayar animasyonlarında görüntülerin birleştirilmesi problemini çözmek amacıyla ortaya atılmıştır (Reeves, 1983; Karaboğa, 2018). Parçacık sürü optimizasyonu (PSO) algoritması ise ilk olarak Kennedy ve Eberhart (1995) tarafından popülasyon tabanlı stokastik arama ve optimizasyon süreci olarak tanıtılmıştır. Algoritmanın amacı; bazı çok boyutlu ve muhtemelen doğrusal olmayan fonksiyon veya sistemlerin global optimaline yakınsamaktır. Böylelikle PSO, genetik algoritma (GA), genetik programlama, evrimsel stratejiler ve evrimsel programlama gibi diğer evrimsel algoritmalarla aynı yolu izlemekte ve benzer mantıkla çalışmaktadır (Kiranyaz ve diğ. 2014; Antoniou and Lu, 2021). Evrimsel algoritmaların hepsinin ortak noktasının, nüfus tabanlı olmaları ve doğadan ilham almalarıyla “*local*” optimumda takılı kalmaktan kaçınabilmeleri ve böylelikle optimum çözümleri bulma potansiyellerinin olmasıdır. Bununla birlikte optimum çözümü bulmalarının garantisinin olmadığı vurgulanmıştır (Kiranyaz ve diğ. 2014).

Bu algoritmanın çalışma prensibinin, kuş sürüsündeki her bir kuşun problem uzayında bireysel bir çözüme karşılık geldiği ve bunun da parçacık olarak isimlendirildiği, parçacıkların tümünün optimize edilmiş uygunluk fonksiyonu tarafından uygunluk değerlerine ayrıca parçacıkların uçuşlarını yönlendiren hızlara sahip olduğu belirtilmiş olup sürünün de araştırma uzayı boyunca mevcut olan en iyi parçacığı takip ederek uçuşunu devam ettirmekte olduğu belirtilmiştir (Bağış, 2009).

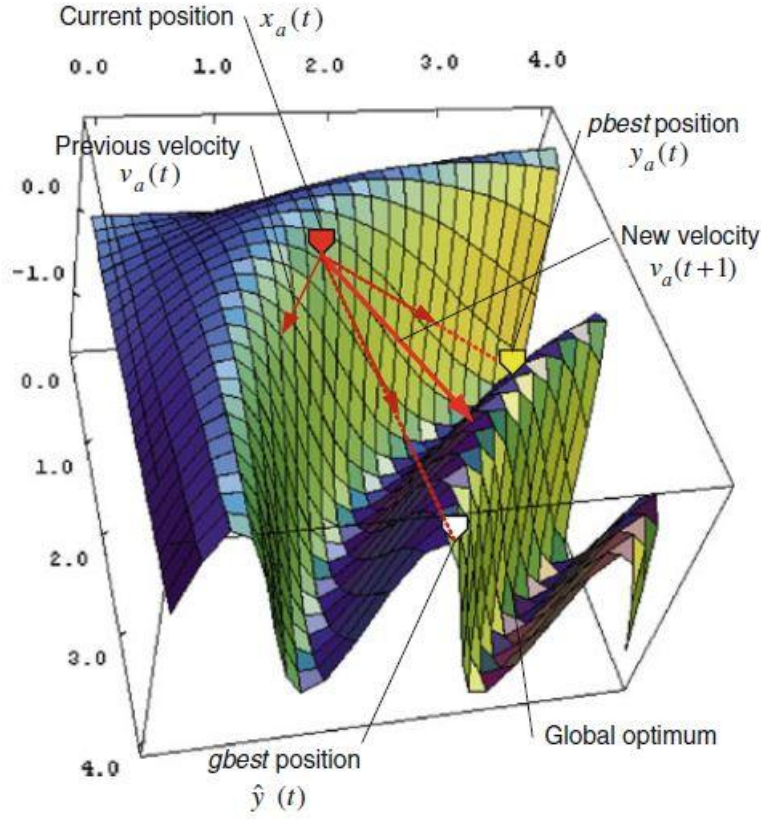
Temel PSO yönteminde (bPSO), bir parçacık sürüsü, her parçacığın konumunun optimizasyon problemine olası bir çözümü temsil ettiği N boyutlu bir arama uzayında uçmaktadır. Sürüdeki her a parçacığı; $\xi_t = \{x_1, x_2, \dots, x_S\}$ aşağıdaki özelliklerle temsil edilir:

$x_{aj}(t)$ = a parçacığının t zamanındaki konumunun j. boyutlu bileşeni

$v_{aj}(t)$ = a parçacığının t zamanındaki hızının j. boyutlu bileşeni

$y_{aj}(t)$ = a parçacığının t anında en iyi (*pbest*) konumunun j. boyutlu bileşeni

$\hat{y}_j(t)$ = t zamanında sürünün “*global*” en iyi konumunun j. boyutlu bileşeni



Şekil 2. 30: PSO'daki temel hız güncelleme mekanizmasının şekilsel gösterimi Kiranyaz ve diğerleri (2014)

Algoritmanın matematiksel olarak gösterimi Kiranyaz ve diğ. (2014) aktarımıyla; f , optimize edilecek uygunluk fonksiyonunu göstermek üzere, genelliği kaybetmeden, amacın N boyutlu uzayda minimum f 'yi bulmak olduğunu varsayalım. Daha sonra a parçacığının kişisel en iyisi, $t + 1$ yinelemesinde Denklem 2.78'de gösterildiği gibi güncellenmektedir.

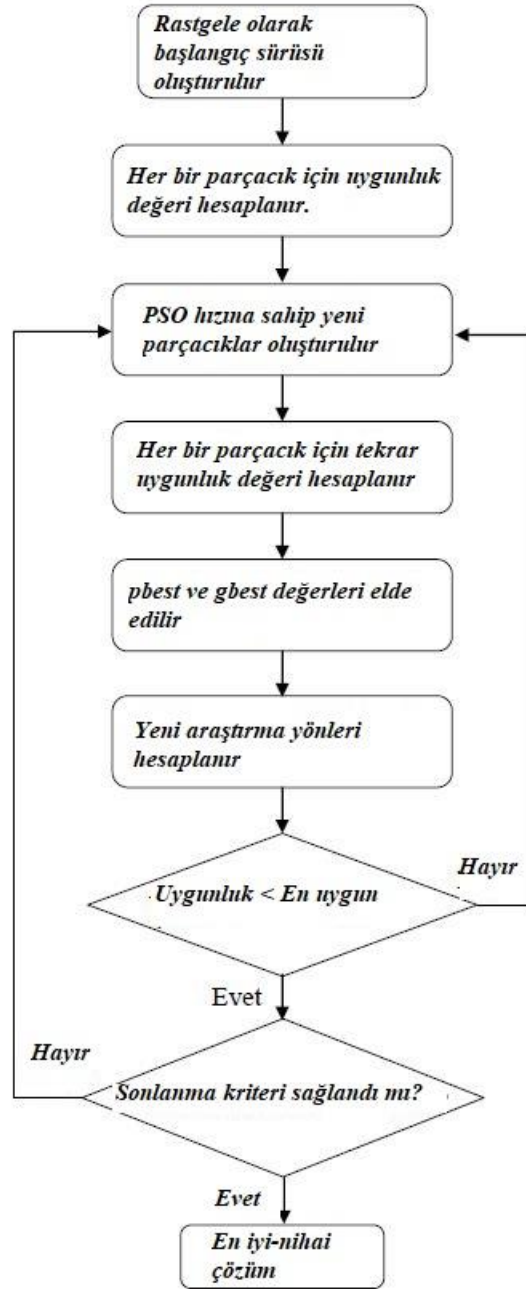
$$y_{a,j}(t+1) = \begin{cases} y_{a,j}(t) & \text{if } f(x_a(t+1)) > f(y_a(t)) \\ x_{a,j}(t+1) & \text{else} \end{cases} \quad \forall j \in [1, N] \quad (2.78)$$

$gbest$, $gbest$ parçacığının indeksi olduğundan, $\hat{y}(t) = \hat{y}_{g(best)}(t) = \min(y_1(t), \dots, y_S(t))$ olarak tanımlanır.

Daha sonra, PSO sürecindeki her yineleme ve her parçacık için konum güncellemeleri, $a \in [1, S]$ ve her bileşen boyutunda $j \in [1, N]$ olmak üzere Denklem 2.79'daki gibi olur.

$$\begin{aligned}
 v_{a,j}(t+1) &= w(t) v_{a,j}(t) + c_1 r_{1,j}(t) (y_{a,j}(t) - x_{a,j}(t)) + c_2 r_{2,j}(t) (\hat{y}_j(t) - x_{a,j}(t)) \\
 x_{a,j}(t+1) &= x_{a,j}(t) + v_{a,j}(t+1)
 \end{aligned}
 \tag{2.79}$$

w atalet ağırlığı ve $c_1; c_2$ hızlanma veya öğrenme sabitleridir. $r_{1j} \sim U(0,1)$ $r_{2j} \sim U(0,1)$ “uniform” dağılıma sahip rastgele değişkenler olarak tanımlanmaktadır. PSO algoritmasının çalışma prensibi Bağış (2009) tarafından Şekil 2.31’deki akış diyagramı olarak paylaşılmıştır.



Şekil 2.31: PSO algoritması için Bağış (2009) tarafından oluşturulmuş akış diyagramı

Genetik algoritma gibi, PSO algoritması da makine öğrenmesinde hiperparametre optimizasyonun gerçekleştirilmesi için dikkat çeken ve literatürde ön çıkan algoritmalar arasındadır.

PSO, algoritmayı kara kutu olarak ele alarak çeşitli problemleri optimize etmek için yaygın olarak kullanılmış olup ayrıca bu özelliği algoritmayı makine öğrenmesinde parametre seçimi probleminde uygun hale getirmekte olduğu vurgulanmıştır. Nihai hedef bu optimizasyon algoritmasının kullanılması suretiyle en yüksek performansa sahip olan makine öğrenmesi modelinin oluşturulmasını sağlayan parametrelerin bir kombinasyonunu bulmaktır (Haidar ve diğ. 2021). Bu amaçla araştırmacılar çeşitli programlama dillerinde paketler geliştirmektedirler. Python programlama dilinde geliştirilen paketlere “*Optunity*” ve “*PSPSO*” ve “*PySwarm*” paketleri öne çıkanlar arasında olup örnek olarak verilebilir (Claesen ve diğ. 2014b; Miranda, 2018; Haidar ve diğ., 2021).

Hiperparametre optimizasyon süreci evrimsel bir algoritma olan GA ile aynı prensipte çalışmakta olup, Şekil 2.31’de akış diyagramı olarak Bağış (2009) tarafından özetlenmiştir. Bu doktora tez çalışmasında PSO algoritmasının makine öğrenmesi sınıflandırma algoritmalarında hiperparametre optimizasyonu probleminde “*Optunity*” kütüphanesi kullanılmıştır (Claesen ve diğ. 2014)

2.7 YORUMLAMA

Yorumlama aşamasında gerçekleştirilen analizler sonucunda elde edilen makine öğrenmesi modellerinin yorumlanması yerine getirilmektedir. Başka bir ifade ile, bu aşamaya gelindiğinde zaten bir veya birden fazla model üretilmiş olup bu modellerin gerçek hayatta kullanılmaya başlanmadan önce kalite ve etkinlik açısından değerlendirilmesi kritik önemdedir. Veri madenciliği uygulamasının özellikle işletmeler için sermaye harcaması ve yatırımı temsil ettiğinden mütevellit söz konusu modellerin geçersiz olması durumunda işletmenin zamanı ve parası boşa harcanmış olacaktır (Larose ve Larose, 2015).

Bu bölümde elde edilen modellerin yorumlanabilmesi için kullanılan model performans değerlendirme yöntemleri ve model performans değerlendirme ölçütleri açıklanmaktadır.

2.7.1 Model performans değerlendirme yöntemleri

Model performans değerlendirme yöntemleri örnekleme için nasıl seçileceği ile ilgilenen yöntemlerdir (Balaban ve Kartal, 2018). Eğitim veri seti kullanılarak model ya da sınıflandırıcı

oluşturulurken algoritmanın eğitimi ve testi için seçilen örneklerin hangi yöntemle seçildiği nihayetinde elde edilen modelin sonuçlarını da etkilemekte olup literatürde modelin performansını değerlendirmek için çeşitli yöntemler bulunmaktadır.

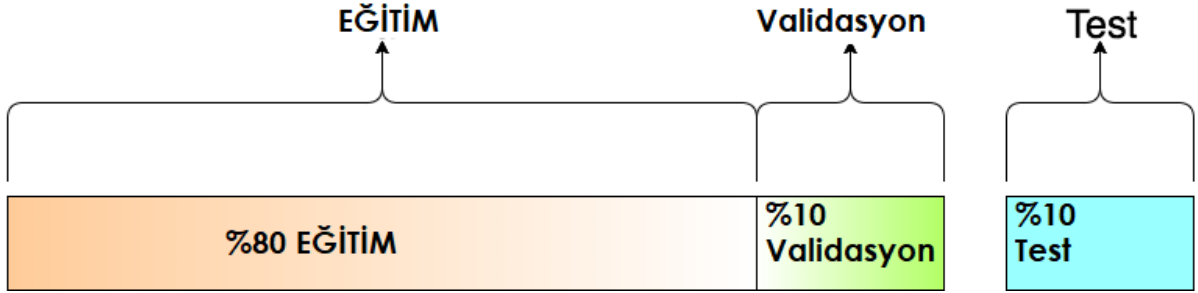
- **Dışarıda Tutma (“Holdout”)**: Dışarıda tutma yöntemi, verileri eğitim seti ve test seti adı verilen birbirinden tamamen farklı iki alt kümeye bölmektedir. Verilerin 2/3'ünün eğitim seti ve kalan 1/3'ünün test seti olarak atanması yaygındır. Eğitim seti, model oluşturmak için kullanılmakta ve elde edilen sınıflandırma modeli daha sonra test setinde test edilmektedir (Olson ve Delen, 2008).

- **Tekrarlı holdout (“Random subsampling or repeated holdout”)**: Dışarıda tutma yönteminin bir diğer varyasyonu olan bu yöntemde, klasik holdout yönteminin k kez tekrarlanması sonucunda genel doğruluk tahmininin, her yinelemeden elde edilen doğrulukların ortalaması olarak alındığı bir yöntemdir (Han ve diğ. 2012).

- **Tabakalı Örnekleme (“Stratified Sampling”)**: Tabakalı örneklemede, örnekler gruplardan veya tabakalardan bölünerek alınmaktadır. Sonuç olarak her tabakadan veya katmandan örnek alınmış olunur. Eğer örnek alma yöntemi rastgele olursa bu yöntemle de rastgele tabakalı örnekleme – “*random stratified sampling*” ismi verilmektedir (Hunt ve Tyrrell, 2004).

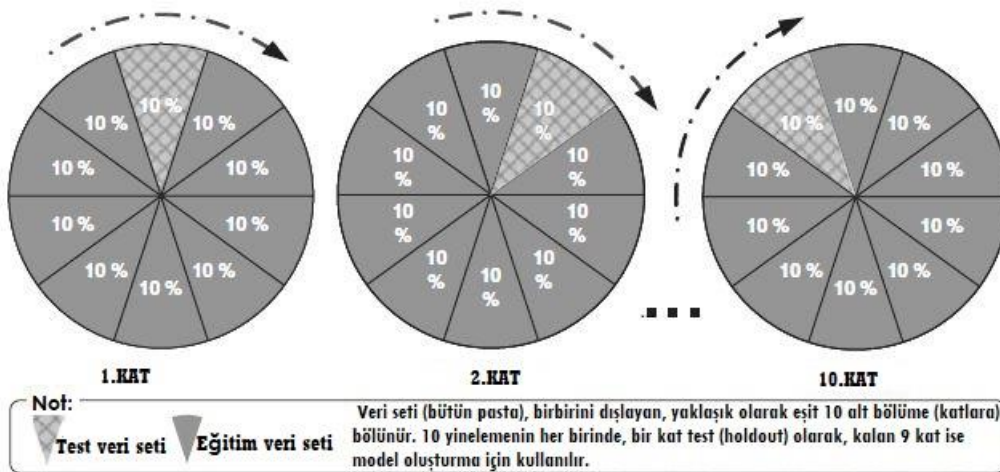
- **Üçlü Ayırma (Three-Way Split)**: Eğitim verilerinden oluşturulan modelin, yeni verileri ne kadar genellediğini tespit etmek temel amaç olduğu için, veri setini eğitim ve test seti olmak üzere iki parçaya ayırmayı esas alan dışarıda tutma yöntemi kullanılmaktadır. Bununla birlikte bu yöntemin diğer bir varyasyonu olan **üçlü ayırma** yönteminin özellikle parametre optimizasyonu gerektiren modeller için kullanılabileceği belirtilmektedir. Başka bir ifade ile, hiperparametre optimizasyonu için dışarıda tutma yani ikili bölme yönteminde küçük bir değişiklik yapılması gerekmekte ve veri setinin **eğitim, doğrulama- “validation” ve test seti** olmak üzere üç parçaya ayırmanın daha doğru sonuçlar vereceği vurgulanmıştır. Parametre optimizasyonu – “*parameter tuning*” aşamasından sonra test seti performansına dayalı bir modelin seçilmesinin daha makul bir yaklaşım olarak görüldüğü ifade edilmektedir. Özellikle ikili ayırma yönteminde, test setinin birden çok kez yeniden kullanılmasının yanlılığa- “*bias*” sebep olabileceği ve bunun da muhtemelen genelleme performansına ilişkin aşırı iyimser tahminlerle sonuçlanabileceği ve bu durumun “*test setinin bilgi sızdırması*” olarak tanımlandığı vurgulanmıştır. Bu sorunu önlemek için, veri setini eğitim, doğrulama ve test veri seti olarak bölen üçlü ayırma yönteminin kullanılabileceği belirtilmiştir. Hiperparametre “*tuning*” ve

model seçimleri için bir eğitim-doğrulama çiftine sahip olmak, model değerlendirmesi için test setini "bağımsız" tutmamıza izin vermektedir (Sebastian Raschka, 2020).



Şekil 2. 32: Üçlü Ayırma – “three way split” yönteminin şekilsel gösterimi

• **Çapraz geçerleme (Cross validation):** Çapraz geçerlemenin kendi içinde farklı türleri bulunmaktadır. Çapraz geçerleme türlerinden biri olan **k-kat çapraz geçerleme (“k-fold cross validation”)** veriler rastgele olarak k adet birbirini dışlayan alt kümeler veya "katlara" bölünmektedir. Böylelikle eğitim ve test prosedürü k kez yapılmaktadır. Doğruluk- “accuracy” k adet yinelemeden sonra elde edilen toplam doğru tahmin edilen sınıflandırma sayısıdır. Diğer bir yöntem ise **birini dışarıda bırak çapraz geçerleme (“leave one out cross validation”)** yöntemidir. Bu yöntemde model n kez eğitilir (n, veri kümesindeki kayıt sayısını temsil eder) ve her seferinde, model oluşturmak için kullanılan kayıtların geri kalanıyla birlikte yalnızca bir kayıt test seti olmaktadır (Olson ve Delen, 2008).



Şekil 2. 33: K- kat çapraz geçerleme (Olson ve Delen, 2008)

•**Bootstrap örnekleme (*bootstrap sampling*)**: Bu yöntem, eğitim için orijinal veri setinden sabit sayıda rastgele örnek değiştirilerek seçilir ve model oluşturmada kullanılmakta olup veri setinin geri kalanı ise test için kullanılmaktadır. Bu işlem istenildiği kadar tekrarlanabilmektedir (Turban ve diğ. 2015).

2.7.2 Model performans değerlendirme ölçütleri

Yapay zekâ- makine öğrenmesi sınıflandırma modelinin performansının değerlendirilmesinde kullanılan ölçütler bu bölümde ele alınmıştır.

Bir sınıflandırma modeli oluşturulduktan sonra bu modelin performansının hem eğitim veri seti hem de test veri seti için performansının ölçülmesi yani modelin başarısının çeşitli performans değerlendirme ölçütleri ile değerlendirilmesi gerekmektedir. Böylelikle, modelin öngörülerde kullanılıp kullanılamayacağına karar verilmektedir (Özkan, 2021).

Sınıflandırma modellerinin performans değerlendirmek için gerçek veri ile öngörülen yani tahmin edilen değerleri karşılaştırmak amacıyla **karışıklık matrisi – “confusion matrix”** adı verilen bir tablodan yararlanılmaktadır (Akpınar, 2017).

Bu tablonun yorumlanması ile elde edilen dört değer (doğru pozitif, doğru negatif, yanlış pozitif ve yanlış negatif değerleri) aracılığı ile doğruluk- “*accuracy*”, hata oranı- “*error rate*”, duyarlılık- “*sensitivity, true positive rate, recall*”, özgüllük ya da belirleyicilik- “*specificity, true negative rate*”, pozitif öngörü değeri-kesinlik- “*precision, positive predictive value*”, negatif öngörü değeri- “*negative predictive value*”, F-ölçütü- “*F-measure*”, yanlış pozitif oranı- “*false positive rate*”, yanlış negatif oranı- “*false negative rate*”, pozitif olabilirlik oranı- “*positive likelihood ratio*”, negatif olabilirlik oranı- “*negative likelihood ratio*”, tanısal üstünlük oranı- “*diagnostic odds ratio*” hesaplanabilmektedir. Alıcı işlem karakteristik eğrileri – “*Receiver Operating Characteristic-ROC*” de performans değerlendirmesinde sıklıkla kullanılmaktadır.

		GERÇEK SINIF	
		POZİTİF	NEGATİF
TAHMİN	POZİTİF	Doğru Pozitif (dp) True Positive (tp)	Yanlış Pozitif (yp) False Positive (fp)
	NEGATİF	Yanlış Negatif (yn) False Negative (fn)	Doğru Negatif (dn) True Negative (tn)

dp(tp): Gerçekte o sınıfa ait olup modelin bulduğu sonuçlara göre o sınıfa ait olduğu bulunan değerlerin sayısı

dn(tn): Gerçekte o sınıfa ait olmayan, modelin sonuçlarına göre o sınıfa ait olmayan değerlerin sayısı

yp (fp): Gerçekte o sınıfa ait olmayan ancak modelin sonuçlarına göre o sınıfa ait olduğu bulunan değerlerin sayısı

yn (fn): Gerçekte o sınıfa ait olmayan ancak modelin sonuçlarına göre o sınıfa ait olmadığı olarak bulunan değerlerin sayısı

Şekil 2. 34: Makine Öğrenmesi Sınıflandırma Algoritmalarının Performansını Değerlendirmek Amacıyla Kullanılan Performans Değerlendirme Ölçütlerinin oluşturulmasında Yararlanılan Karışıklık Matrisi

Elde edilen dp, dn,yp ve fn değerlerine göre doğruluk, hata oranı, belirleyicilik, yanlış pozitif oranı, yanlış negatif oranı, pozitif öngörü değeri (kesinlik), negatif öngörü değeri, F-ölçütü, pozitif olabilirlik oranı, negatif olabilirlik oranı, tanısal üstünlük oranı değerleri aşağıdaki formüllere göre hesaplanmaktadır.

İkili- “binary” sınıflandırmada kullanılan performans değerlendirme ölçütlerinin hesaplanması için aşağıda verilen formüllerden yararlanılmaktadır.

Doğruluk: Makine öğrenmesi sınıflandırma modelinin toplamdaki doğru tahmin sayısının, veri setinde bulunan tüm örneklem sayısına oranı olarak tanımlanmaktadır.

$$\text{Doğruluk (ACC)} = \frac{DP+DN}{DP+DN+FP+FN} \quad (2.80)$$

Hata Oranı (ERR): Hata oranı, 1'den ACC değerinin çıkartılması ile elde edilmektedir.

Makine öğrenmesi sınıflandırma modelinin hata oranı formülü:

$$\text{Hata Oranı (ERR)} = 1 - \text{Doğruluk (ACC)} \quad (2.81)$$

Duyarlılık (Sensitivity, True Positive Rate, Recall-TPR): Sınıflandırıcı tarafından doğru bir şekilde sınıflandırılmış pozitif örneklerin yani DP değerinin veri setinde bulunan toplam pozitif örneklem sayısına oranı olarak tanımlanmaktadır.

$$\text{Duyarlılık (TPR)} = \frac{DP}{POZ} = \frac{DP}{DP+YN} \quad (2.82)$$

Özgüllük (Specificity, True Negative Rate –SPC veya TNR): Sınıflandırıcı tarafından doğru bir şekilde sınıflandırılmış negatif örneklerin yani DN değerinin veri setinde bulunan toplam negatif örneklem sayısına oranı olarak tanımlanmaktadır.

$$\text{Belirleyicilik (SPC)} = \frac{DN}{DN+YP} \quad (2.83)$$

Yanlış Pozitif Oranı (False Positive Rate- FPR): Gerçekte negatif sınıfta bulunan fakat sınıflandırıcı tarafından pozitif şeklinde tahmin edilmiş sınıfların, toplam negatif örneklem sayısına oranı olarak tanımlanmaktadır.

$$\text{Yanlış Pozitif Oranı (FPR)} = 1 - \text{SPC} = \frac{YP}{NEG} = \frac{YP}{YP+DN} \quad (2.84)$$

Yanlış Negatif Oranı (False Negative Rate- FNR): Gerçekte pozitif sınıfta bulunan fakat sınıflandırıcı tarafından negatif şeklinde tahmin edilmiş sınıfların, toplam pozitif örneklem sayısına oranı olarak tanımlanmaktadır.

$$\text{Yanlış Negatif Oranı (FNR)} = 1 - \text{TPR} = \frac{YN}{POZ} = \frac{YN}{YN+DP} \quad (2.85)$$

Kesinlik- Pozitif Öngörü Değeri (*Precision / Positive Predictive Value – PPV*): Sınıflandırıcı tarafından doğru tahmin edilmiş pozitif örneklerin, sınıflandırıcı tarafından doğru tahmin edilmiş pozitif örneklem sayısına oranı olarak tanımlanmaktadır.

$$\text{Pozitif Öngörü Değeri (PPV)} = \frac{DP}{TPOZ} = \frac{DP}{DP+YP} \quad (2.86)$$

Negatif Öngörü Değeri (*Negative Predictive Value- NPV*): Sınıflandırıcı tarafından doğru tahmin edilmiş negatif örneklerin, sınıflandırıcı tarafından doğru tahmin edilmiş negatif örneklem sayısına oranı olarak tanımlanmaktadır.

$$\text{Negatif Öngörü Değeri (NPV)} = \frac{DN}{TNEG} = \frac{DN}{DN+YN} \quad (2.87)$$

F- ölçütü (*F-measure*): Pozitif Öngörü Değeri (Kesinlik) ile Doğru Pozitif Oranı (Duyarlılık) değerini birlikte değerlendirme imkânı sunan ve bunların harmanik ortalaması olarak formülize edilmiş bir değerdir.

$$\text{F- Ölçütü (F)} = \frac{2 \cdot \text{PPV} \cdot \text{TPR}}{\text{PPV} + \text{TPR}} \quad (2.88)$$

Pozitif Olabilirlik Oranı (*Positive Likelihood Ratio- LR+*): Yanlış pozitiflik oranına bölünen gerçek pozitiflik oranıdır. Başka bir ifade ile bir örnek verecek olursak, bir hastada pozitif bir testin beklenme olasılığının, hastalığı olmayan bir hastada pozitif bir testin beklenme olasılığına bölümü olarak da tanımlanabilir (Shreffler ve Huecker, 2021).

$$\text{Pozitif Olabilirlik Oranı (LP+)} = \frac{\text{TPR}}{\text{FPR}} = \frac{\text{TPR}}{1-\text{SPC}} \quad (2.89)$$

Negatif Olabilirlik Oranı (*Negative Likelihood Ratio – LR-*): Doğru negatif oranına bölünen yanlış negatif oranıdır. Başka bir ifade ile bir örnek verecek olursak, negatif olabilirlik oranı veya LR-; “hastalığı olan bir hastanın test sonucunda negatif olarak test edilme olasılığının, hastalığı olmayan bir hastanın test sonucunda negatif olarak test edilme olasılığına bölümü olarak tanımlanmaktadır (Shreffler ve Huecker, 2021).

$$\text{Negatif Olabilirlik Oranı (LP-)} = \frac{\text{FNR}}{\text{TNR}} = \frac{1-\text{TPR}}{\text{SPC}} \quad (2.90)$$

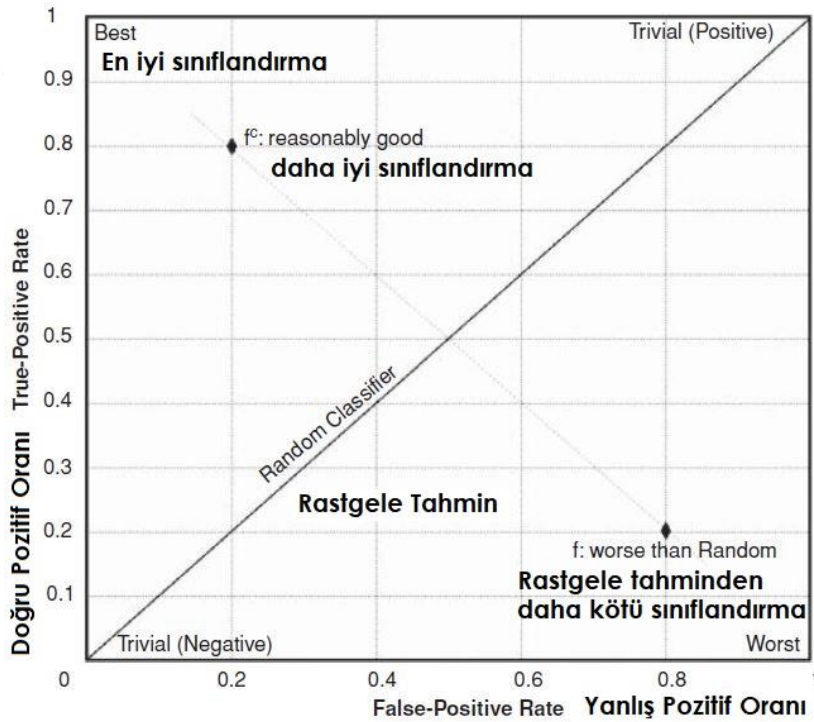
Tanısal Üstünlük Oranı (*Diagnostic Odds Ratio – DOR*): Pozitif Olabilirlik Oranı’nın Negatif Olabilirlik Oranı’na bölünmesi sonucu elde edilen değer olarak tanımlanmaktadır (Shreffler ve Huecker, 2021).

$$\text{Tanısal üstünlük oranı (DOR)} = \frac{\text{LR}+}{\text{LR}-} \quad (2.91)$$

Alıcı İşlem Karakteristik Eğrileri- “Receiving Operationg Characteristic-ROC Curves”

Bir ROC eğrisi, yatay eksenin (x eksen) yanlış pozitif oranını (FPR) ve dikey ekseninde (y eksen) bir sınıflandırıcının doğru pozitif oranını (TPR) gösterdiği bir çizim olarak tanımlanmaktadır. ROC, tipik olarak bir ikili sınıflandırma senaryosunda, değişen karar kriterleri üzerinden bir öğrenme algoritmasının performansını görselleştirmek için çok güçlü bir grafiksel araçtır. ROC analizi, sınıflandırıcının duyarlılığı ve özgüllüğü arasındaki ilişkiyi incelemektedir. (Japkowicz ve Shah, 2011).

ROC eğrisinin altındaki alan, bir sınıflandırıcının doğruluk ölçüsünü belirlemektedir. 1 değeri mükemmel bir sınıflandırıcıyı belirtirken, 0,5 değeri ise rastgele şanstıan daha iyi olmadığını göstermektedir. İyi bir sınıflandırma modelinin performansının grafiğin son üst köşesine yaklaşması beklenmektedir (Alpaydın, 2013; Turban ve diğ., 2015; Balaban ve Kartal, 2018).



Şekil 2. 35: ROC uzayı (Japkowicz ve Shah, 2011)

2.8 DİYABETİK AYAĞIN AMPÜTASYON RİSKİNİN ÖNGÖRÜLMESİNE YÖNELİK LİTERATÜR TARAMASI

Uluslararası çalışmalara erişmek için Web of Science veri tabanında (“predict”+“amputation” +“diabetic”+“foot”) anahtar kelimeleri kullanılarak, başlık, özet ve anahtar kelimelerde 1975-2021 yılları arasında yayınlanmış, uluslararası alan indekslerinde (SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI) taranan İngilizce dilindeki makalelere erişilmiştir. Makaleler en çok atıf alandan en az atıf alana doğru sıralanmıştır. Bu çalışma kapsamında toplamda ulaşılan 261 makaleden, doktora tez araştırmasının konusu ile ilişkili **141** makale tespit edilmiştir. **Bu makalelerden “60” adeti “Yapay Zekâ” ve ilişkili sınıfa girdiği saptanmış ve bu nedenle de detaylı olarak doktora tez çalışmasında yorumlanmıştır.** Yorumlanan çalışmalar 1996-2021 tarihleri arasında olup kronolojik sıra ile doktora tez çalışmasında ele alınmıştır. Bu makaleler dışında kalan 81 makale incelenmiş olup, doktora tez çalışmasının temeli olan yapay zekâ ile doğrudan ilişkili olmadıkları için tezin bu bölümünde ayrıntılı olarak verilmemiştir. Bununla birlikte Yükseköğretim Kurulu Tez Merkezi'nde bulunan tezleri taramak amacıyla, gelişmiş tarama seçeneği kullanılarak “diyabetik ayak” anahtar kelimesiyle tarama yapılmış olup, ulaşılan 106 tezin tamamı “Sağlık Bilimleri – örn: Tıpta Uzmanlık, Hemşirelik vb.” alanında yapılan tez çalışmaları olduğu saptanmış, “Sağlık Bilişimi” alanında daha önce bu kapsamda yapılmış hiçbir tez çalışmasına (yüksek lisans ve doktora dahil) ulaşılamamıştır.

Mayfield ve diğ. (1996) diyabetik hastalarda çeşitli risk faktörlerinin amputasyon riskine etkisini ölçmek ve klinik verilere dayalı bir ayak riski puanlama sistemi geliştirmek amacıyla 25-85 yaş arasındaki diyabetik hastalardan oluşan bir popülasyon bazında bir vaka kontrol çalışması gerçekleştirmişlerdir. Vaka hastaları (61) 1983 ile 1992 yılları arasında alt ekstremite amputasyonu operasyonu geçirirken, kontrol denekleri (183) 1992 yılına kadar herhangi bir amputasyon işlemi geçirmemişlerdir. Araştırmada, amputasyon işlemi geçiren 61 kişi tespit edilmiş ve 183 kontrol deneği ile karşılaştırılmıştır. Lojistik Regresyon testi sonuçlarına göre, erkeklerin amputasyona geçirme olasılığı kadınlardan daha yüksek olarak bulunurken (olasılık oranı [OR] 6,5,%95 CI 2,6-15), retinopati böbrek veya kardiyovasküler hastalığı olanların, olmayanlara göre amputasyon geçirme olasılığı daha yüksek olarak tespit edilmiştir (OR: 4,6; %95 CI 1,7-12). Amputasyon riski, periferik nöropati, periferik vasküler hastalık, kemik deformiteleri ve ayak ülseri öyküsü risk faktörleriyle hemen hemen eşit derecede ilişkili olarak saptanmıştır. Demografik farklılıklar ve diyabet şiddeti değişkenleri kontrol edildikten sonra, amputasyon için OR'ler hesaplanmış ve bir ayak risk faktörü için 2,1 (%95 CI 1,4-3,3), iki risk

faktörü için 4,5 (%95 CI 2,9-6,9) ve üç veya dört risk faktörü için ise 9,7 (%95 CI 6,3-14,8) olarak saptanmıştır. Sonuç olarak erkek cinsiyeti, göz, kalp ve böbreğin organ komplikasyonları ve zayıf glukoz kontrolü, daha yüksek bir amputasyon riski ile ilişkilendirilmiştir. Periferik nöropati, periferik vasküler hastalık, kemik deformitesi ve önceden geçirilmiş ülser öyküsü ise artmış bir alt ekstremité amputasyonu riski ile eşit şekilde ilişkili olarak tespit edilmiştir. Araştırmada önerilen risk sınıflandırma sistemi, önceden yayınlanmış risk sınıflandırma sistemlerinden altı önemli şekilde farklılık gösterdiği ifade edilmiştir. Bunlar sıralanacak olursa: 1) klinik izlenimlerden ziyade verilere dayanmakta olduğu, 2) ayağa ilişkin spesifik faktörlerin hastaların sonuçları ile ilişkilendirildiği, 3) sevk bazlı olmaktan ziyade popülasyon temelli bir araştırma olduğu, 4) vasküler durumu içerdiği, 5) aynı anda dört risk faktörünü de içerdiği ve 6) sonuçların yorumlanmasını zorlaştıran demografik faktörleri ve diyabet şiddetini kontrol ettiği vurgulanmıştır.

Edelman ve diğ. (1997) diyabetik hastalarda ayak ülserinin başarılı iyileşmesini öngörmede öykü, fizik muayene ve manyetik rezonans görüntülemenin (MRI) değerini belirlemek amacıyla yaptıkları araştırmada, 64 ardışık diyabetik hastadaki 78 ülseri prospektif olarak takip etmişlerdir. Tek değişkenli analiz, ayrık bağımsız değişkenleri değerlendirirken χ^2 analizi ile veya sürekli bağımsız değişkenler değerlendirilirken Wilcoxon Rank-Sum Testi ile gerçekleştirilmiştir. İyileşme olasılığını öngörmek için değişkenleri modellemek amacıyla öncelikle iyileşmenin tek değişkenli öngörücülerini belirlenmiş ve bunlar Lojistik Regresyon modelinde (tek değişkenli öngörücüler Lojistik Regresyon modeline girilmiş ve bağımsız öngörücüler ise aşamalı regresyon ile) tanımlanmıştır. Hastaların 6 aydaki iyileşmelerinin tek değişkenli öngörücüler arasında 65 yaş altında olma, son 15 yıl içinde diyabet teşhisi, ağrısız ülser, palpe edilebilen periferik nabızlar, 0,5'ten büyük ABI indeksi ve doktorun osteomyelit olasılığına ilişkin genel değerlendirmesi bulunmaktadır. Bununla birlikte, çok değişkenli Lojistik Regresyon modelinde iyileşmenin öngörücüler; Doppler muayenesinde işitilebilir bir nabızın varlığı ($p = 0,01$) ve ağrısız bir ülser ($p = 0,04$) olarak saptanmıştır. MRI görüntülemesinde osteomyelit tanısı bu hastalarda iyileşmeyi öngörmemiştir. Sonuç olarak klinik muayenenin vasküler bileşenlerinin, diyabetik ayak ülseri olan hastalarda iyileşmenin en iyi öngörücüler olduğu vurgulanmıştır.

Coppini ve diğ. (1998) diyabetik ayak hastalığı için risk faktörlerini tanımlamak amacıyla yaptıkları kohortta, ilk kez diyabetik ayak ülseri atağı geçiren 405 diyabetik hastanın kayıtlarını (yapılandırılmış ayak muayenesi ve Biyoteziyometre(VPT) ile değerlendirilen ayak titreşimi algılama eşikleri dahil olmak üzere) retrospektif olarak gözden geçirmişlerdir. Tüm analizler,

Lojistik Regresyon analizi ile yapılmış her risk faktörü için olasılık oranları (OR) hesaplanmıştır. Duyarlılık ve özgüllük değerleri, önceki olasılıkların gözlemlenen vaka sıklığına eşleştirilmesi ile elde edilmiştir. İstatistiksel olarak anlamlı olasılık oranları (OR) : anormal ayak VPT (OR 4,38; CI 1,11–17,26; $p = 0,01$); anormal klinik muayene (en az 1 anormallik- OR: 2,3, CI 1,00–5,20; $p < 0,01$); ve HbA1c (“glikozillenmiş hemoglobin” - sonradan alt ekstremitte komplikasyonları gelişen hastalarda OR 1,30, CI 1,01–1,66; $p < 0,02$) olarak tespit edilmiştir. Uzun dönem komplikasyonları öngörmede VPT'nin duyarlılığı (%70) klinik testten (%55) daha iyi olarak saptanmış, ancak tüm testler benzer özgüllük göstermiştir (%70-72). Olay riski yani diyabetik ayak ülseri geçirme riski de diyabet süresinin her 10 yılı için iki katına çıkmıştır (OR: 2,10, CI 1,11–4,30; $p = 0,02$). Objektif ve uygulaması basit olan yaşa göre düzeltilmiş VPT ölçümlerinin, diyabet kliniklerindeki yarı kantitatif testlere göre gelecekteki diyabetik ayak komplikasyonlarını daha iyi öngördüğü sonucuna ulaşılmıştır.

Ahroni ve diğ. (1999) diyabet özellikleri, tıbbi öykü, ayak deformitesi, duysal nöropati ve plantar ayak basıncı arasındaki ilişkiyi değerlendirmek için diyabetik hastalardan oluşan 517 kişi ve 1017 ayağı prospektif olarak takip etmişlerdir. Tek değişkenli analizlerde, en az bir plantar basınç ölçümü ile vücut kütlesi, cinsiyet, ırk, yaş, insülin kullanımı, bazı ayak deformiteleri, plantar nasır ve duysal nöropati arasında önemli ilişkiler tespit edilmiştir. Diyabet süresi, HbA1c ve ayak ülseri veya amputasyon öyküsü, plantar basınçla ilişkili olarak tespit edilememiştir. Çoklu regresyon analizlerinde, vücut kütlesi, insülin kullanımı, beyaz ırk, erkek cinsiyet, plantar nasır ve diyabet süresi belirli bölgelerdeki basınçlarla anlamlı olarak ilişkili olarak saptanmıştır. Ayak deformiteleri öncelikle ön ayak basınçlarıyla ilişkili olarak bulunmuştur. Çalışmada, geriye doğru eliminasyon yöntemiyle gerçekleştirilen Lojistik Regresyon analizinde, sonucun öngörücüsü olarak tanımlanan iki veya daha fazla bölgede yüksek basıncın bulunması, yalnızca vücut kütlesi değişkeni için istatistiksel olarak anlamlı kalmaya devam etmiştir. Sonuç olarak diyabetik deneklerde yüksek ayakkabı içi plantar basıncın, kısmen kolayca bulunabilen klinik özelliklerden tahmin edilebileceği ifade edilmiştir.

Pittet ve diğ. (1999) hastaneye kabul anında diyabetik ayak ülserinin konservatif tedavisinin başarısızlığının öngörücü kriterlerini belirlemek amacıyla büyük bir eğitim kurumunda diyabetik ayak lezyonu nedeniyle kabul edilen tüm diyabetik hastalarının 5 yıllık ileriye dönük uzun vadeli takibini gerçekleştirmişler, toplamda 120 diyabetik ayak hastasının ayak lezyonlarının tanı ve sınıflandırılmasında Wagner sınıflandırması kullanmışlar ve çalışma değişkenleri olarak hasta demografikleri, enfeksiyon ve diyabetle ilgili klinik parametreleri kullanmışlardır. Çalışmada konservatif tedavinin başarısızlığı ana sonuç ölçütü olarak

belirlenirken, Lojistik Regresyon analizi ile bağımsız tahmin değişkenleri seçilmiştir. Konservatif tedavinin başarısızlığını öngören bağımsız faktörler ateşin varlığı (olasılık oranı [OR] = Santigrat derece başına 1,1; %95 güven aralığı [CI], 1,0-1,2) ve artmış serum kreatinin düzeyi (OR = 1,002 / litre başına mikromol; %95 güven aralığı [CI], 1,0020-1,0021), daha önce diyabetik ayak lezyonu nedeniyle hastaneye yatış öyküsü (OR = 1,4; % 95 güven aralığı [CI], 1,2-1,6) ve gangren durumu (Wagner Grade 4 veya 5; OR = 1,8; %95 güven aralığı [CI], 1,5-2,2) olarak tespit edilmiştir. Sonuç olarak, kültür testi ile birlikte, uzun süreli parenteral ve oral antibiyotikleri içeren konservatif tedavi yaklaşımının, diyabetik ayak ülseri veya şüpheli osteomiyeliti nedeniyle kabul edilen diyabetik ayak hastalarının büyük bir kısmında amputasyon olmaksızın başarılı olduğu ifade edilmiştir.

Kaletta ve diğ. (2001) 1 yıllık süre içinde osteomiyelit veya selülit tanısı ile başvuran 29 diyabetik hastanın retrospektif bir dosya incelemesini belirli bir eritrosit sedimantasyon hızını osteomiyelit ile ilişkilendirmek amacıyla gerçekleştirmişlerdir. İki grup arasındaki değişkenlerdeki farklılıklar ve her değişkenin eritrosit sedimantasyon hızı üzerindeki etkisini belirlemek için Student T testi, Lojistik Regresyon analizi gerçekleştirilmiştir. Karşılaştırılan çeşitli laboratuvar değerleri ve demografik faktörlerden eritrosit sedimantasyon hızı, iki grup arasında önemli ölçüde farklılık gösteren tek ölçü olarak tespit edilmiştir. Osteomiyelitin mevcut olduğu bir seviye olan 70 mm / sa'lik optimum kesme değeri için ROC eğrileri kullanılmış ve bu kesme değerinin üstündeki seviyelerde osteomiyelit en yüksek hassasiyet (%89,5) en yüksek özgüllük (%100), en yüksek pozitif öngörü değeri (%100) ve negatif öngörü değeri (%83) ile saptanmıştır. Bu çalışma, diyabetik ayak infeksiyonlarında klinik şüpheyile birlikte, eritrosit sedimantasyon hızının osteomiyelit için oldukça öngörücü olduğunu ve 70 mm / saat değerinin, kemik infeksiyonunun varlığını veya yokluğunu doğru bir şekilde tahmin etmek için en uygun sınır olduğunu göstermektedir.

Perkins ve diğ. (2001), distal simetrik polinöropati tanısında standartlaştırılmış elektrofizyolojik testlerle karşılaştırıldığında dört basit duyusal tarama testlerinin (10 g Semmes-Weinstein monofilaman incelemesi [SWME], yüzeysel ağrı hissi, açma-kapama yöntemiyle titreşim testi ve zamanlanmış yöntemle titreşim testi) alıcı işlem karakteristiğini değerlendirmek amacıyla gerçekleştirdikleri araştırmada, 478 denek ile her test için alıcı-işlem karakteristiği (ROC) eğrilerini, pozitif ve negatif olasılık oranlarını ile birlikte duyarlılık ve özgüllük değerlerini sunmuşlardır. Kohort, nöropati durumu bilinmeyen bir diyabet kliniğine giden seçilmemiş hastalar, şüpheli nöropati nedeniyle Diyabetik Nöropati Araştırma Kliniğine sevk edilen hastalar, toplulukta diyabetli hastalar için reklamlar aracılığıyla seçilen nöropati

durumu bilinmeyen kişiler ve referans hastalar olmak üzere dört farklı kaynaktan belirlenmiştir. Her denek, klinikte 4-5 saatlik bir kalış süresince en az yedi farklı denetçi tarafından değerlendirilirken, üçlü tarama testleri dokuz farklı denetçi kullanılarak deneklerin bir alt kümesinde gerçekleştirilmiştir. Tanımlanan klinik tabakalar arasında aşağıdaki gibi önemli farklılıklar gözlenmiştir: hastaların yaşı (referans grup ile daha genç olmak), diyabet süresi (daha şiddetli nöropati için daha uzun, $p < 0,0001$) ve nöropati süresi (şiddetli nöropati için daha fazla süre, $p < 0,0001$). Kategorik değişkenler arasındaki ilişkileri tanımlamak için kullanılan Ki Kare (X^2) testi, nöropatisi olmayan diyabetik deneklere kıyasla klinik nöropatili deneklerde bir komplikasyonun varlığını temsil ederken, p değerleri, nöropatinin komplikasyonu ve varlığı arasında önemli bir ilişki olduğunu göstermektedir ($p = 0,001$). Kontenjan tablosu analizleri, nöropati evresi ile birlikte ayak ülseri, retinopati, nefropati ve erektil disfonksiyon öyküsünün önemli ölçüde artan prevalansını ortaya koymaktadır. Araştırma bulgularında, dört basit tarama testi, benzer alıcı-işlem karakteristiği sergilemiştir. ROC eğrisi analizlerine göre kesme noktaları, pozitif veya anormal bir testin, uygulanan sekiz uyarıcının beş yanlış tepkisi ile temsil edildiğini ortaya koymaktadır. Negatif veya normal bir test ise bir veya uygulanan sekiz uyarının daha az yanlış yanıtı ile sonuçlanmıştır. Bu kriterlere göre, açma-kapama yöntemiyle titreşim testi, zamanlanmış yöntemle titreşim testi, SWME ve yüzeysel ağrı hissi testi için Lojistik Regresyon modeli sonuçlarına göre pozitif olasılık oranlarının nokta tahminleri sırasıyla 26,6, 18,5, 10,2 ve 9,2 olarak tespit edilmiştir. Negatif olabilirlik oranlarının nokta tahminleri sırasıyla 0,33, 0,51, 0,34 ve 0,50'dir. Ayrıca, Kruskal-Wallis Ki-Kare test yaklaşımı ile gerçekleştirilen tarama testleri, karşılaştırılabilir duyarlılık ve özgüllük sonuçları göstermiştir. 10-g SWME, yüzeysel ağrı testi ve açma-kapama yöntemiyle yapılan titreşim testinin hızlı olduğu ve her birinin uygulanmasının yaklaşık 60 saniye gerektirdiği, bununla birlikte zamanlanmış titreşim testinin daha uzun sürdüğü ve yorumlamasının daha karmaşık olduğu ifade edilmiştir. Ayrıca, iki basit testin kombinasyonunun (örneğin, 10-g SWME ve açma-kapama yöntemiyle titreşim testi kombinasyonu) her bir ayrı tarama testine değer katmayacağının altı çizilmiştir. Sonuç olarak, diyabetik nöropati için yıllık tarama, yüzeysel ağrı hissi testi, SWME veya açma-kapama yöntemiyle titreşim testi kullanılarak yapılmasının gerekliliği vurgulanmıştır. Bu üç basit testten herhangi birinin hem diyabet hem de birinci basamak kliniklerinde diyabetik nöropatinin yıllık taraması için güvenle kullanılabileceği ifade edilmiştir. Her duyuşal modalite için bildirilen alıcı-işlem karakteristiğinin, nöropati olasılığını tahmin etmek için bireysel hastaların fiziksel muayenesindeki pozitif bulgulara da uygulanabileceği ifade edilmiştir.

Mayfield ve diğ. (2002) medial arteriyel kalsinoz (MAC)'ın bir risk faktöründen çok otonom nöropatinin bir belirteci olduğunu ve sonuçların otonom nöropatiye bağlı olduğunu test etmek için yaptıkları araştırmada, ayak radyografisi bulunan ve devam eden prospektif başka bir çalışmadaki tüm denekleri ($n = 473$) çalışmalarına dahil etmişlerdir. Otonom nöropati, zamanlanmış solunumla kalp hızı değişkenliği veya postural hipotansiyonun bulunması ile ölçülmüştür. Lojistik Regresyon modeline göre, otonom nöropatinin çalışmada değerlendirilen her iki ölçümü de başlangıçtaki MAC'nin bağımsız öngörücüleri olarak tespit edilmiştir. MAC bulunması, ülserasyon ,ampütasyon ve mortalite için artmış risk ile ilişkili olarak saptanmıştır. Sonuç olarak, MAC bulunan hastalardaki artmış mortalite, ampütasyon ve ülserasyonun postural hipotansiyon veya kalp hızı değişkenliği olarak ölçülen otonomik nöropati ile açıklanabileceği vurgulanmıştır.

Treece ve diğ. (2004) ülser özelliklerinin derecelendirilmesine dayanan daha önce yayınlanmış S (AD) SAD sisteminin geçerliliğini göstermek amacıyla multidisipliner bir kliniğe yeni sevk edilen 300 diyabetik ayak hastasını prospektif olarak takip etmişlerdir. Ki kare test sonuçlarına göre, hastaların akıbeti ile alan ($\chi^2 = 25,9$, $p < 0.001$), derinlik ($\chi^2 = 33,8$; $p < 0.001$), sepsis ($\chi^2 = 13,5$, $p = 0,004$) ve arteriyopatiye ($\chi^2 = 33,7$, $p < 0,001$) göre önemli farklılıklar tespit edilmiş ancak denervasyon ($\chi^2 = 5,1$; $p = 0,16$) ve hastaların sonlanımı arasında önemli bir fark saptanamamıştır. Bu ilişkilerin gücü Somers d kullanılarak doğrulanmıştır. Alan, derinlik, sepsis ve arteriopatide bağımsız değişkenleri kullanılarak Lojistik Regresyon analizi yapılmıştır. Bu değişkenler arasında modele önemli katkı sağlayanlar alan ($p = 0,01$), derinlik ($p < 0,001$) ve arteriopatide ($p < 0,001$) olarak tespit edilmiştir.

Volaco ve diğ. (2004) kritik ayak iskemisi (CFI) olan diyabetik hastalarda ayak ülserlerinin sonucunu tahmin eden risk faktörlerini belirlemek amacıyla yaptıkları araştırmada, iskemik ayak ülserleri (CFI) bulunan 98 diyabetik hastayı retrospektif olarak takip etmişlerdir. Araştırmada muhtemel sonlanımın (ölüm , majör ampütasyon veya iyileşmeyen lezyon) tahmini için çoklu Lojistik Regresyon analizi gerçekleştirilmiştir. Öngörücü değişkenlerin seçiminde önceki araştırma sonuçları dikkate alınmış ve sonuç değişkeni ile iki değişkenli analizlerdeki önemli ilişkilerin bir kombinasyonu ile belirlenmiş olup, ek olarak aşamalı çoklu Lojistik Regresyon tekniklerinin uygulanmasıyla doğrulanmıştır. Kötü sonuç riski, son dönem böbrek hastalığı nedeniyle diyalize giren hastalarda 8,9 kat, cerrahi komplikasyonlar varsa 7,0 kat ve C-reaktif protein (CRP) ikinci beşte birlik kısmın üzerinde 5,4 kez (kesme değeri 8 mg / dl) arttığı tespit edilmiştir. Sonuç olarak, yapılan araştırmada diyaliz tedavisi, yüksek CRP

seviyeleri ve cerrahi komplikasyonların, iyileşmeyen lezyonlar, majör amputasyon ve ölüm için güçlü öngörücüleri olduğu bulunmuştur.

Faglia ve diğ. (2007) TransAtlantic InterSociety Consensus kriterlerine göre kritik ekstremité iskemisi (CLI) tanısı almış diyabetik hastalarda ayak bileği üstü amputasyonu tahmin edebilen transkutanöz oksijen (TcPO₂) değerlerini belirlemek amacıyla bir uzuvda CLI nedeniyle tedavi gören 564 diyabetik hastayı retrospektif olarak değerlendirmişlerdir. Tedavi sonrası TcPO₂ değerleri tüm hastalarda ayak sırtında ölçülmüştür. Çalışma bulguları olarak, ROC eğrisi ile değerlendirilen tedavi sonrası TcPO₂ değerinin, revaskülarizasyon ihtiyacını belirlemek için en iyi eşik olarak 34 mmHg olduğu belirlenmiş ve ROC eğrisi altında kalan alan 0.89 (%95 CI 0,85 – 0,94) olarak tespit edilmiştir. Lojistik Regresyon analizi sonuçlarına göre, bu eşik için ayak bileği üstü amputasyon olasılığı %9,7 olarak tespit edilmiş ve eğer TcPO₂ > 40 mmHg üstü değeri için %3'e düşeceği öngörülmüştür. Sonuç olarak, TcPO₂ seviyeleri <34 mmHg değeri altında olması, revaskülarizasyon ihtiyacını gösterirken, ≥34 <40 mmHg değerleri için bu ihtiyaç daha az acil görünmekte olup, yine de önemli bir amputasyon riskinin olduğu ifade edilmiştir. 40 mmHg'nin üzerindeki TcPO₂ seviyelerinin ise, revaskülarizasyon ihtiyacının, doku kaybının ciddiyetine ve prosedürün neden olduğu olası morbiditeye bağlı yeniden değerlendirilmesi gerekliliğinin altı çizilmiştir.

Lavery ve diğ. (2008) kısmi ayak amputasyonunu takiben geniş, kronik ve iskemik olmayan diyabetik ayak yaraları olan bireylerde 1. ve 4. haftalarda yara alanı küçülme (PWAR) yüzdesine göre yara iyileşme olasılığını değerlendirmek amacıyla yaptıkları araştırmada, 162 hastadan oluşan 16 haftalık bir randomize klinik araştırmadan elde ettikleri verileri, negatif basınçlı yara tedavisi (NPWT) (n = 77) ile standart nemli yara bakım tedavisi (MWT) (n = 85) ile ilişkili sonuçları karşılaştırmak için analiz edilmiştir. Bu çalışmada, yara alanı azalma yüzdesi ile nihai iyileşme arasındaki ilişkiyi değerlendirmek için bayesci Lojistik Regresyon teknikleri kullanılmıştır. Modelleme ilk olarak PWAR için 1. haftada ve sonra tekrar 4. haftada yapılmıştır. Çalışma sonucunda, PWAR'daki erken değişiklikler, 16. haftada son iyileşmenin habercisi olarak bulunmuştur. Spesifik olarak, 1 haftada ≥ %15 PWAR'a veya 4 haftada ≥ %60 PWAR'a ulaşan yaraların iyileşme olasılığı %68 ve 77 olarak tespit edilirken (sırasıyla), eğer bu yara alanı azalmaları elde edilemezse %31 ve 30 olasılığa sahip olacağı öngörülmüştür. NPWT alan hastaların MWT alanlara kıyasla hem 1 haftada %15 PWAR hem de 1 ayda %60 yara alanı azalması elde etme olasılıkları 2,5 kat daha yüksek olarak saptanmıştır (olasılık oranları sırasıyla 2,51 ve 2,49). Bu çalışmanın sonuçları, klinisyenlerin 16 haftada iyileşme

olasılığını tahmin etmek için, uygulanan tedavide 1 hafta kadar erken bir sürede, yaranın PWAR değerini hesaplayabileceklerini göstermektedir.

Parisi ve diğ. (2008), diyabetik ayak ülserlerinin sonucunun belirleyicileri olarak Wagner, Texas Üniversitesi (UT) ve S(AD)SAD yara sınıflandırma sistemlerini karşılaştırmak amacıyla yaptıkları araştırmada, 94 denekten veri toplamışlardır. Araştırmada sonuç değişkeni iyileşme insidansı olarak belirlenmiştir. Lojistik Regresyon analizi, temel değişkenler ile ülser iyileşmesi arasındaki ilişkiyi analiz etmek amacıyla ve çok değişkenli analiz, seçilen değişkenlerin adım adım modele dahil edilmesi suretiyle gerçekleştirilmiştir. Ülser süresi, Kaplan-Meier sağkalım analizi kullanılarak değerlendirilmiş ve sonucun tahmininin doğruluğu log-rank testi kullanılarak üç sınıflandırma sistemi ile karşılaştırılmıştır. Araştırma sonuçlarına göre, S (AD) SAD sisteminde yer alan derinlik ($p = 0,002$), enfeksiyon ($p = 0,006$) ve denervasyon ($p = 0,002$), UT derecesi ($p = 0,002$) ve aşaması ($p = 0,032$) ve Wagner derecesi ($p = 0,002$) için iyileşme açısından önemli farklılıklar tespit edilmiştir. Sonuç olarak, üç sınıflandırma sistemi eşit derecede iyi performans göstermekle birlikte, çoklu Lojistik Regresyon analizi sonuçlarına göre, S (AD) SAD sınıflandırmasına göre sınıflandırılan beş faktör arasında enfeksiyonun yokluğu değişkeninin en önemli faktör olduğunu ve iyileşme olasılığını 4,3 kat artırdığını gösterilmiştir. S (AD) SAD skoru ≤ 9 (toplam olası maksimum skor 15) olan ülserlerin iyileşme olasılığı, ≥ 10 puan olanlardan 7,6 kat daha yüksek olarak saptanmıştır ($p < 0,001$, Doğruluk = 0,70).

Seuc ve diğ. (2009) periferik arter hastalığı tanısı olan hastaların, yalnızca hastaneye yatışta ölçülen klinik değişkenlere dayanarak kısa süreli majör amputasyonu tahmin etmek amacıyla yaptıkları araştırmada, toplamda 463 hastadan oluşan bir hasta popülasyonunda boylamsal, retrospektif bir kohort çalışması gerçekleştirmişlerdir. Amputasyonun önemli öngörücülerini saptamak için Lojistik Regresyon analizi gerçekleştirilmiş olup; başvuru sırasında ölçülen on iki klinik değişken potansiyel öngörücü olarak kabul edilmiştir. Lojistik Regresyon analizi ile belirlenen önemli majör amputasyon öngörücüleri olarak, lezyon tipi, başlangıç tanısı, plantar bölge lezyonu, diyabet öyküsü, etkilenen ayak parmaklarının sayısı ve etkilenen bacak alanı değişkenleri saptanmıştır. Hastaların %80'inden fazlası doğru şekilde sınıflandırılmış olup, duyarlılık %42 ve özgüllük %96 olarak tespit edilmiştir. Araştırmada sonuç olarak, vasküler hastalık tanısı olan hastalardaki majör amputasyon riskinin, hastaneye yatışta ölçülen klinik değişkenler öngörücü olarak kullanılarak yüksek performansla tahmin edilebileceği vurgulanmıştır.

Widatalla ve diğ. (2009) diyabetik ayak ülseri olan hastalarda majör ve minör amputasyonlar için risk faktörlerini tespit etmek amacıyla gerçekleştirdikleri çalışmada 2321 hastayı prospektif olarak takip etmişler ve Uluslararası Diyabetik Ayak Konsensüsü tarafından geliştirilen yara sınıflandırması sistemini kullanarak, hastalar iskemi, nöropati, yara çaplarının doğrusal ölçümü, yara derinliği ve enfeksiyona ilaveten son dönem böbrek yetmezliği açısından değerlendirmişlerdir. Araştırmadaki ülseler arasında en yaygın olanı plantar ülseler olarak saptanırken (%42,6) ve onu ayak başparmağı ülseleri (%39) takip etmiştir. Araştırmadaki hastaların geçirdikleri amputasyon dereceleri (%28,5) sınıflandırıldığında, %10 hastada majör amputasyon prosedürü geçirdikleri saptanırken bunların %8,7'si diz altı amputasyon, geri kalanının ise diz üstü amputasyon işlemi olduğu ve %18,5 hasta ise minör (ayak parmağı) amputasyon işlemi geçirdiği saptanmıştır. Bununla birlikte en sık ampute edilen ayak parmağının %9,9'luk oranla ayak baş parmağı olduğu tespit edilmiştir. Araştırmada majör amputasyon için Grade-3 bacak iskemisinin olması en önemli risk faktörü olarak saptanmıştır ($p = 0,000$, olasılık oranı = 5,08; CI = 2,56–10,07). Araştırmada, periferik duyuşal nöropatinin major amputasyon ($p = 0,027$, olasılık oranı = 2,43; CI = 1,08-5,45) ve küçük parmak amputasyonu ($p = 0,002$, olasılık oranı = 2,16; CI = 1,32-3,5) ile anlamlı derecede ilişkili olarak tespit edilmiştir. Araştırmada, deri altı fasyaya, kaslara, tendonlara, kemiklere ve eklemlere kadar olan derin ülselerin, ayak parmağı amputasyonu ile anlamlı şekilde ilişkili olduğu saptanmıştır ($p < 0,005$, olasılık oranı = 3,45; CI = 2,02–5,88). Derece 3 ve 4 enfeksiyonun ayak parmağı amputasyonu ile önemli ölçüde ilişkili olduğu saptanmıştır ($p < 0,005$, olasılık oranı = 2,4; CI = 1,55-3,7). Ayrıca son dönem böbrek hastalığının (diyalizdeki hastalar) majör amputasyon ile önemli ölçüde ilişkili olduğu bulunmuştur ($p = 0,003$, olasılık oranı = 4,39; CI = 1,53-12,61).

Yesil ve diğ. (2009) yaptıkları gözlemsel çalışmada, verileri Ocak 1999 ile 2008 arasında toplanan 670 diyabet ayak hastasından diyabetik ayak ülseri ataklarında prospektif olarak toplanmışlar ve amputasyonun potansiyel prediktörlerini retrospektif olarak değerlendirmek için kullanmışlardır. En az altı ay boyunca takip için hastaneye gelmeyen hastaların dışlanmasından sonra, 574 ayak ülseri hastasının verileri değerlendirilmiştir. Sınıflandırma sistemi olarak Meggitt-Wagner kullanılmıştır. Seçilen her değişkenin sonuç üzerindeki bağımsız etkisini tahmin etmek için Lojistik Regresyon kullanılmıştır. Wagner sınıflandırma sistemi tarafından belirlenen bacak iskemisi, osteomyelit, kangren ve ülser derinliği, genel ve majör amputasyonların başlıca bağımsız öngörücüleri olarak tespit edilmiştir. İleri yaş, koroner arter hastalığının varlığı, sigara kullanımı ve ülser boyutu değişkenleri ise genel veya majör

ampütasyonlarla ilişkili olduğu tespit edilmiştir. Bununla birlikte, akut faz reaktanlarının başlangıç seviyeleri (beyaz kan hücresi sayısı, polimorfonükleer lökosit sayısı, trombosit sayısı, eritrosit sedimentasyon hızı (ESR), serum C-reaktif protein (CRP) ve albümin) ve azalmış hemoglobin seviyeleri ampütasyon riski ile ilişkilendirilmiştir. Ayrıca çok değişkenli analiz, başlangıç CRP ve ESR seviyelerindeki bir standart sapma artışının sırasıyla genel ve majör ampütasyonların bağımsız prediktörleri olduğunu göstermiştir. Sonuç olarak ekstremitte iskemisi, osteomyelit, lokal ve yaygın kangren varlığı ve ülser derinliği ampütasyonun bağımsız öngörücüleri olarak tespit edilmiştir. Çalışmada, hastaların başlangıçtaki ESR ve CRP değerlerinin, klinisyenlere ampütasyon öngörüsünde bulunmaları için yardımcı olabileceği vurgusu yapılmıştır.

Barberan ve diğ. (2010) akut diyabetik ayak infeksiyonları nedeniyle hastaneye yatırılan 78 ardışık hastayı retrospektif olarak analiz etmişler ve ampütasyona ihtiyaç duyan ve duymayan hastaların verilerini karşılaştırarak, hastaneye yatış sırasında kısa süreli ampütasyonu tahmin etmek için öngörücü faktörleri tespit etmek amacıyla bir araştırma gerçekleştirmişler ve bu amaçla hastaların klinik / epidemiyolojik, laboratuvar, görüntüleme ve tedavi değişkenlerini kayıt altına alarak karşılaştırmalı olarak analiz etmişlerdir. Bağımlı değişken olarak hastaların ampütasyon statüsü değişkeni belirlenmiş ve bağımsız değişkenler ise iki değişkenli analiz sonucuna göre, en önemli faktörler seçilerek, çok değişkenli Lojistik Regresyon analizi gerçekleştirilmiştir. Çok değişkenli analizde, ampütasyon riski; Wagner derece 4 veya 5 (20 kat daha yüksek), damar tıkanıklığı (12,5 kat daha yüksek) ve yüksek sedimentasyon oranı (birim başına %6 daha yüksek) ile artmış olduğu saptanmıştır. Lojistik Regresyon analizi sonucunda ampütasyon operasyonu geçiren hastaların %76,9'unda ve geçirmeyenlerin %92,3'ünde doğru tahmin sonucuna ulaşılmıştır.

Uccioli ve diğ. (2010) kritik bacak iskemisi (CLI) olan 510 diyabetik hastanın uzun vadeli sonuçlarını değerlendirmek amacıyla yaptıkları çalışmada, iyileşme, iyileşmeme, majör ampütasyon ve mortalite olmak üzere 4 farklı sonuç ölçütleri kullanmışlar ve hastaların ülserlerinin tamamı University of Texas (TWC) sınıflandırmasına göre sınıflandırılmıştır. Araştırma bulgularında, ampütasyon geçiren hastaların daha sık diyalize girdiği ve yaklaşık yarısının iskemik kalp hastalığına sahip olduğu tespit edilmiştir. Aynı zamanda, ampütasyon geçiren hastaların daha düşük bazal transkütan oksijen TcPO₂ değerleri ve daha yüksek transkütan karbondioksit TcPCO₂ değerlerine sahip olduğu saptanmıştır. Bununla birlikte ampütasyon geçiren hastaların sistolik kan basıncı önemli ölçüde daha düşük olduğu bulunmuştur. Ek olarak ampütasyon geçiren hastaların daha sık olarak, acil servise başvurduğu

görülmüştür. Tüm potansiyel öngörücü değişkenler için tek değişkenli Lojistik Regresyon analizleri gerçekleştirilmiş ve risk oranları (HR) ve ilgili %95 CI değerleri sunulmuştur. Tüm potansiyel öngörücüler ile aynı anda çok değişkenli bir regresyon analizi gerçekleştirilmiş ve sonucu en iyi tahmin eden değişkenler belirlenmiştir. Monovariate analiz sonuçlarına göre, ülser alanı $> 5 \text{ cm}^2$, enfeksiyon, TWC D3, minör amputasyon sayısı, iskemik kalp hastalığı, diyaliz ve yaş majör amputasyon için risk faktörleri olarak bulunmuştur. Çok değişkenli regresyon analizinde; iyileşme için tahmin değişkenler: ülser boyutu, enfeksiyon ve iskemik kalp hastalığı, majör amputasyon için: ülser boyutu, geçirilmiş minör amputasyon sayısı ve yaş, perifer arter hastaları için ek tahmin değişkenleri iyileşme için: transkutan oksijen TcPO₂ değeri değişimi, majör amputasyon için: bazal TcPO₂ değeri, bazal A1C, TcPO₂ değeri değişimi ve perkutan anjiyoplasti prosedürünün başarısızlığı olarak saptanmıştır. Araştırmada özellikle 5 cm^2 den büyük ülser boyutunun, hastaların iyileşme şansını azalttığının ve majör amputasyon riskinin arttırdığının altı çizilmiştir. Bu doğrultuda tüm ülselerin perkutan anjiyoplasti dahil agresif tedaviyi gerektirdiğini ve küçük iskemik ülseler için bile bu tedavinin düşünülmesi gerektiği ifade edilmiştir.

Akinci ve diğ. (2011) diyabetik ayak enfeksiyonu olan hastaların akut faz reaktanlarının, başlangıç ve tedavi sonrası düzeylerinin, hastaların sonlanımı hakkında öngörücü değerini değerlendirmeyi amaçladıkları çalışmalarında prospektif olarak takip edilen hastalardan toplanan 165 diyabetik ayak ülser epizodundanki veriyi analiz etmişlerdir. Lojistik Regresyon analizi, seçilen her değişkenin sonuç üzerindeki bağımsız etkisini tahmin etmek ve değerlendirmek için gerçekleştirilmiştir. Tek değişkenli analiz sonuçları, başlangıç ve tedavi sonrası CRP seviyelerinde, eritrosit sedimentasyon oranları ve beyaz kan hücresi sayımlarında 1-SD artışın ve tedavi sonrası albümin seviyelerinde 1-SD düşüşün artmış amputasyon riski ile önemli ölçüde ilişkili olduğunu göstermiştir. Lojistik Regresyon analiz sonuçlarına göre ise de iskemi, osteomyelit, daha yüksek Wagner derecesi, daha geniş yara alanı, tedavi sonrası CRP seviyeleri, tedavi öncesi ve sonrası WBC düzeyleri, amputasyonu tahmin etmede anlamlı olarak saptanmıştır. Bununla birlikte amputasyon operasyonu geçiren hasta grubunda, başlangıç ve tedavi sonrası CRP seviyeleri, ESR oranları, beyaz kan hücresi sayıları ve trombosit sayıları önemli ölçüde artarken, albümin oranlarının önemli ölçüde baskılandığı saptanmıştır. ROC eğrisi analizine göre tedavi sonrası CRP seviyesi de amputasyon riskiyle güçlü bir şekilde ilişkili bulunmuştur. Sonuç olarak, akut faz reaktanlarının dolaşımdaki seviyelerinin, diyabetik ayak enfeksiyonlarında amputasyon riski ile ilişkili bulunmuştur.

Fleischer ve diğ. (2011), tedavi sonrası laboratuvar değerlerinin ileri diyabetik ayak infeksiyonu olan hastalarda klinik yanıtı tahmin etmeye yardımcı olup olmayacağını belirlemek amacıyla gerçekleştirdikleri araştırmada, orta veya şiddetli diyabetik ayak infeksiyonları nedeniyle hastaneye yatırılan ardışık 38 hasta belirlemişler ve bu hastaların kayıtlarını retrospektif olarak değerlendirmişlerdir. Kesin tedaviye başladıktan kısa bir süre sonra (yani tedavi sonrası) ölçülen periferik beyaz kan hücresi sayısı, nötrofil sayısı, ESR ve CRP seviyeleri, Lojistik Regresyon modelleri kullanılarak klinik yanıtla ilişkileri açısından incelenmiştir. Gerçekleştirilen hem tek değişkenli hem de çok değişkenli analizler sonucunda, beyaz kan hücresi sayısının $\geq 11,000$ hücre / mL olarak tanımlanan lökositozun, tedaviden ortalama $3 \pm 1,4$ gün sonra, orta veya şiddetli diyabetik ayak infeksiyonları bulunan hastalarda kötü klinik yanıtı tahmin etmek için en önemli ve tek öngörücü olduğu bulunmuştur. Sonuç olarak bu çalışma sonucunda, kesin tedaviye başladıktan kısa bir süre sonra gözlenen lökositozun, 90 gün içinde istenmeyen bir klinik yanıtı öngördüğü sonucuna varılmıştır.

Redlich ve diğ. (2011) Ampütasyon planlanan kritik uzuv iskemisi (CLI) olan diyabetik hastalarda anjiyoplasti (PTA) sonrası muhtemel sonlanımı tahmin etmek amacıyla anjiyogramları ABI değerini ve transkutanöz oksijen değerlerini ($TcPO_2$) incelemişlerdir. PTA, CLI bulunan 28 diyabetik hastaya uygulanmıştır. Verilerin analizinde, parametrik testlerden varyans analizi ve Student t testi, kategorik verilerin grup karşılaştırmaları için parametrik olmayan testler (Wilcoxon ve Mann-Whitney U testi) kullanılmış olup, Lojistik Regresyon analizi, anjiyografik ve $TcPO_2$ parametrelerini belirlemek amacıyla majör ampütasyondan kaçınma açısından klinik başarının prediktörleri olarak gerçekleştirilmiştir. Ekstremitte kurtarma oranları Kaplan – Meier analizi ile hesaplanmış ve log-rank testi ile karşılaştırılmıştır. Araştırmada, en az bir tibial arter açıklığı sağlandığında %88,2 ($n = 15$) hastada ampütasyon işleminin gereksiz olduğu bulgusuna ulaşılmıştır. ABI değerinin, uzuv kurtarma için hiçbir öngörü değerine sahip olmadığı bulunmuştur. Ayrıca, $TcPO_2$ değerleri sadece ampütasyon gerektirmeyen hastalarda anlamlı olarak artmıştır ($p = 0,015$). İşlem sonrası anjiyogramlarda ayağı besleyen sadece bir tibial arter veya sadece peroneal arter açık olan hastalarda $TcPO_2$, sonlanımı güvenilir bir şekilde tahmin edebilmiştir. Sonuç olarak, tedavinin bir parçası olarak diz altı PTA prosedürünün, CLI'li diyabetik hastaların yarısından fazlasında majör ampütasyonu önlemede etkili olduğu tespit edilmiş ve $TcPO_2$ değerinin de anjiyografik sonlanım kriterleri sağlanmadığında bile, geçerli bir öngörücü olduğu saptanmıştır.

Yekta ve diğ. (2011) diyabetik ayak hastalarında ampütasyonu ve ilişkili demografik, davranışsal ve klinik faktörleri belirlemek amacıyla yaptıkları araştırmada 94 diyabetik ayak

hastasını takip etmişlerdir. Çalışmada amputasyon operasyonu geçirenlerin daha yaşlı ve aynı zamanda daha az eğitilmiş ve daha uzun diyabet süresine sahip olduğu ve yüksek riskli yaralara ek olarak zayıf glisemik kontrolü olduğu (kan glukoz düzeyi ve HbA1c değerleri) tespit edilmiştir. Diyabet komplikasyonları hastaların amputasyon durumuna göre karşılaştırıldığında, amputasyon operasyonu geçiren hastalarda retinopati, nefropati ve hipertansiyonun daha sık olduğu sonucuna ulaşılmıştır. Lojistik Regresyon analizinde Wagner sınıflandırması ≥ 3 ve HbA1c değişkenleri amputasyonla ilişkili bulunmuştur. ROC eğrisinde, %9,75 HbA1c ve 305 g/dL glukoz kesme noktası, artan amputasyon riskini tahmin etmek için ayırt edici olmuştur.

Erdman ve diğ. (2012), yeni bir standartlaştırılmış hibrit görüntü tabanlı puanlama sisteminin (CSI), diyabetik ayak infeksiyonlarında prognostik değeri olup olmadığını belirlemek amacıyla yaptıkları çalışmada, osteomyelit şüphesi bulunan 77 hasta ve 101 ayak ^{99m}Tc -WBC tek foton emisyonlu bilgisayarlı tomografi CT (SPECT) / CT görüntü yorumlama ve hasta grafiklerinin bağımsız olarak incelenmesi yoluyla retrospektif olarak değerlendirmişlerdir. CSI'nin başarılı klinik sonucun prediktörü olarak rolünü belirlemek için Lojistik Regresyon analizi kullanılmıştır. Çalışmada, bu modele dayalı olarak, sonucu tahmin etmede CSI'nin optimal kesim değerini belirlemek için ROC eğrileri kullanılmış ve doğruluk, ROC eğrisi altındaki alanla (AUC) ve ayrıca pozitif öngörü değeri (PPV) ile değerlendirilmiştir. Olumlu sonucu tahmin etmek için yapılan ROC analizinde doğruluk 0,79 olarak tespit edilmiştir. Görüntüye dayalı osteomyelit değerlendirmesi, sonucu tahmin etmede CSI'dan daha az doğru sonuç verdiği bulunmuştur (AUC = 0,66, p = 0,016). Orta şiddette (CSI sınıflandırma: 3-6) hastalarda, antibiyotik süresi 42 güne uzatıldığında tedavi başarısızlığının %68'den % 36'ya düşmesi de araştırmanın çarpıcı bulguları arasındadır (p = 0,026). Sonuç olarak, standart bir skorlama sistemi olan CSI'ya dahil edilen hibrid görüntülerden türetilmiş infeksiyon parametreleri diyabetik ayak infeksiyonlarında prognostik değere sahip olduğu vurgulanmıştır.

Griffin ve diğ. (2012) geçirilmiş başparmak amputasyonunun gelecekteki uzuv kaybının öngörücüsü olup olmadığını belirlemek amacıyla yaptıkları çalışmada, ayak parmağı amputasyonu yapılan 63 hasta belirlemişler ve hastaların klinik notlarını retrospektif olarak incelemişlerdir. İstatistiksel analiz için öncelikle tek değişkenli analiz Ki-kare testi (nominal veriler) ve Mann-Whitney U testi (sürekli veriler) ile gerçekleştirilmiş; çok değişkenli analiz ise, tek değişkenli analizdeki en önemli faktörleri içeren Lojistik Regresyon modeli kullanılarak gerçekleştirilmiştir (revaskülarizasyon, diyabet, yaş, sigara). Sonuç olarak, diyabeti olmayan hastalar ile gelecekteki ekstremité kaybı arasında anlamlı bir korelasyon tespit edilmekle

birlikte (Ki-kare = 4,31, $p = 0,038$), tanımlanmış başka hiçbir risk faktörünün majör amputasyon riskini öngörmediği tespit edilmiştir.

Sun ve diğ. (2012) farklı Wagner dereceleri olan diyabetik ayak hastalığı olan 789 hastada alt ekstremitte amputasyonu (LEA) için risk faktörlerini açığa çıkarmak amacıyla bir araştırma yapmışlardır. Bu çalışma Tayvan'da multidisipliner diyabetik ayak bakım merkezinde gerçekleştirilmiştir. Çok değişkenli kademeli ordinal Lojistik Regresyon analizi gerçekleştirilmiştir. Sonuç olarak, yüksek Wagner derecesi, düşük ABI serum albümin ve hemoglobin ve artmış beyaz kan hücresi (WBC) sayısı, artmış amputasyon riski ile önemli ölçüde ilişkili olarak bulunmuştur. Hastaların yaraları Wagner sınıflandırmasına göre katmanlandırıldığında, yukarıdaki öngörücülerin çoğu ve tahmini glomerüler filtrasyon hızı (eGFR) yalnızca Wagner 3. derecede tespit edilmiştir. Wagner derece 2. ve 4. Sınıflarda ise , lökosit (WBC) sayısı artan LEA riski ile pozitif olarak ilişkili birincil öngörücüler olarak tanımlanmıştır. Sonuç olarak Wagner sınıflandırması, LEA için potansiyel risk faktörlerini önemli ölçüde açığa çıkarmış ve farklı derecelerde farklı öngörücüler göstermiştir. Düşük ABI, albümin veya eGFR gibi diyabetik ayak amputasyonu için geleneksel olarak tanınan öngörücüler, neredeyse yalnızca Wagner grade 3 olan hastalarda bulunmuştur.

Moura Neto ve diğ. (2013), diyabetik ayağın gelişimini tahmin edebilecek kolay değerlendirilebilir değişkenleri belirlemek amacıyla yaptıkları çalışmada, 496 hastayı retrospektif olarak değerlendirmişlerdir. Diyabetik ayak ile ilişkili risk faktörlerinin belirlenmesi, tek değişkenli ve çok değişkenli Lojistik Regresyon analizi ile yapılmıştır. Sadece erkek cinsiyet ülserasyon için anlamlı bir prediktör olarak tespit edilmiştir ($p < 0,001$). Amputasyonun belirleyicileri erkek cinsiyet ($p < 0,001$; OR = 3,44, %95 CI= 1,81–6,56) ve nöro-iskemik diyabetik ayak kategorisi ($p < 0,049$; OR = 4,6; %95 CI = 1.01–20.9) olarak saptanmıştır. Araştırmada ayrıca, nöropati ve periferik vasküler hastalık kombinasyonunun, diyabetik ayak hastalarında amputasyon riski ile ilişkili olduğu bulunmuştur. Özellikle kombine risk faktörleri bulunan erkek hastaların potansiyel olarak daha kötü sonuçları nedeniyle özel ilgi gösterilmesi gereken bir hasta grubu olduğunun altı çizilmiştir.

Tsai ve diğ. (2013) diyabetik ayak ülseri olan hastalarda böbrek fonksiyonuna göre alt ekstremitte amputasyonu için öngörücülerini değerlendirmeyi amaçlamışlar ve diyaliz hastalarının özellikle ABI ve Doppler dalga formları ile hastaların vaskülatürünü değerlendirmişlerdir. Diyabetik Ayak Bakım Merkezine kabul edilen 658 diyabetik hastanın tüm klinik değişkenleri analiz edilmiş, bununla birlikte periferik arter hastalığını

değerlendirmek için bir uzman, diyalizdeki 78 hastanın Doppler görüntülerini retrospektif olarak incelemiştir. Majör amputasyonun bağımsız risk faktörlerini belirlemek amacıyla çok değişkenli Lojistik Regresyon analizi gerçekleştirilmiştir. Wagner yara sınıflandırmaları, azalmış serum albümin seviyeleri ve düşük ABI değerleri, diyaliz yapılmayan gruptaki hastalar arasında majör amputasyon için önemli öngörücüler olarak tespit edilmiştir. Bununla birlikte, bu göstergeler diyalizdeki diyabetik hastalarda amputasyon riskinin öngörücüleri olarak saptanamamıştır. Dorsalis pedis arterinde veya posterior tibial arterde zayıf monofazik dalga formlarının varlığı, diyaliz hastaları arasında majör amputasyon için bağımsız öngörücüleri olarak tespit edilmiştir (olasılık oranı: 7,61; $p = 0,008$). Araştırmada sonuç olarak, diyabetik ayak ülseri tedavisi gören diyaliz hastalarında yaygın olarak bulunan, diz altı arterlerin zayıf monofazik Doppler dalga formlarının, majör amputasyon için bağımsız öngörücü olarak tanımlanabileceği vurgusu yapılmıştır.

Utsunomiya ve diğ. (2014) endovasküler tedavi (EVT) sonrası yara iyileşmesi için deri perfüzyon basıncının (SPP) tahmini değerini belirlemek amacıyla yaptıkları araştırmada kritik ayak iskemisi bulunan (CLI, 123 ekstremit) ve balon anjiyoplasti ve/veya stent uygulanan 113 hastanın SPP değerlerini operasyon öncesi ve operasyondan 48 saat sonra kaydetmişlerdir. İkili Lojistik Regresyon analizi, anjiyografik ve SPP parametrelerinin yara iyileşmesinin öngörücüleri olarak değerlendirilmesi amacıyla gerçekleştirilmiş ve CLI'li hastalarda yara iyileşmesini tahmin etmek amacıyla bir eşik SPP değerini belirlemek için ROC eğrileri kullanılmıştır. Yara iyileşmesini tahmin etmek için yapılan ROC eğrisi sonuçlarına göre EVT'den sonra SPP ölçümünün 0.81 değeri ile eğri altında alana sahip olduğu saptanmıştır (%95 CI 0,723 ile 0.899, $p < 0,001$). Lojistik Regresyon analizi, EVT sonrası SPP ölçümünün yara iyileşmesinin bağımsız bir prediktörü olduğunu göstermiştir ($p < 0.001$). SPP değerleri > 30 mmHg, 40 mmHg ve 50 mmHg değerleri için yara iyileşme olasılıkları sırasıyla %69,8, %86,3 ve %94,5 olarak hesaplanmıştır. EVT sonrası SPP ölçümünün, CLI'li hastalarda yara iyileşmesinin bağımsız bir öngörücüsü olduğu tespit edilmiştir. Çalışmada, 30 mmHg'lik bir SPP değerinin EVT sonrası yara iyileşmesini öngörmeye en iyi eşik değer olduğu gösterilmiştir.

Choi ve diğ. (2015) diyabetik ayak ülseri nedeniyle amputasyon operasyonu geçiren hastalarda kardiyak komorbiditeler yüksek mortalite oranları ile ilişkili olduğunu ifade etmişler ve spinal anestezi altında majör amputasyon operasyonu geçiren tip II diyabetli, diyabetik ayak hastalarında preanestetik ekokardiyografik ve laboratuvar değerlerinin hastane mortalitesini tahmin edip edemeyeceğini belirlemek amacıyla gerçekleştirdikleri araştırmada, 215 hastanın arşivlenmiş tıbbi kayıtlarını retrospektif olarak incelemişlerdir. Hastaların demografik verileri

ve ameliyat öncesi laboratuvar ve transtorasik ekokardiyografik değerleri; mortalite ile bağımsız olarak ilişkili faktörleri belirlemek amacıyla, çok değişkenli Lojistik Regresyon yöntemi kullanılarak analiz edilmiştir. Araştırma sonucunda, preoperatif hemoglobin düzeyi (risk oranı: 0,082; %95 CI, 0,013-0,509; $p = 0,007$) ve sol ventrikül ejeksiyon fraksiyonu (risk oranı: 0,874; %95 CI, 0,779-0,981; $p = 0,023$) hastane içi mortalitenin önemli belirleyicileri olarak saptanmıştır.

Chuan ve diğ. (2015) diyabetik ayak için kullanılan sınıflandırma sistemlerinden biri olan PEDIS sınıflandırma sistemini doğrulamak ve klinik uygulamayı kolaylaştırmak için bir skor sistemi oluşturmuşlar ve bu sistemi daha önce yayınlanmış iki ortak skor sistemi ile karşılaştırmışlar ve bu amaçla 364 diyabetik ayak hastasını retrospektif olarak değerlendirmişlerdir. Araştırmada, sonuç değişkenleri, iyileşme, iyileşmeme, amputasyon (minör ve majör amputasyonlar dahil) veya ölüm olarak belirlenmiştir. Tek değişkenli analize göre anlamlı değişkenler ($p < 0,05$) kullanılarak hangi özelliklerin ve klinik değişkenlerin bağımsız olarak sonlanımla ilişkili olduğunu değerlendirmek için çok değişkenli bir Lojistik Regresyon analizi gerçekleştirilmiştir. Araştırma sonucunda, PEDIS sınıflandırma sistemi ile değerlendirildiğinde, belirlenen diyabetik ayak sonlanımının, her bir alt kategorinin artan ciddiyeti ile kötüleştiği bulunmuştur. Ayrıca, daha uzun ülser süresi öyküsü (OR: 1,024, 95% CI, 1,010–1,039, $p = 0,001$), alt ekstremitte perfüzyonunun ciddiyeti (OR: 8,098, 95% CI, 3,658–17,929, $p < 0,001$), daha büyük ülser boyutu (OR: 2,461, 95% CI, 1,373–4,412, $p = 0,002$), daha derin yara (OR: 12,494, 95% CI, 4,076–38,297, $p < 0,001$), daha şiddetli infeksiyon (OR: 7,202, 95% CI, 3,407–15,224, $p < 0,001$), ve duyu kaybı (OR: 9,545, 95% CI, 3,184–28,611, $p < 0,001$), olumsuz sonlanımın bağımsız öngörücüleri olarak bulunmuştur. Daha da önemlisi, ROC eğrileri ile performansı değerlendirilen yeni önerilen PEDIS skrolama sisteminin, özellikle SINBAD ve Wagner skor sistemleri ile karşılaştırıldığında yüksek doğruluğa sahip olduğu saptanmıştır (**0,95**, 0,88 ve 0,86).

Pickwell ve diğ. (2015) enfekte diyabetik ayak ülseri olan 575 bireyden oluşan bir kohortta, alt ekstremitte amputasyonunun bağımsız prediktörlerini ve Uluslararası Diyabetik Ayak Çalışma Grubu (IWGDF) sınıflandırma sistemini'nin amputasyonu öngörücü değeri ile birlikte amputasyonu tahmin etmek için bir risk skoru geliştirmeyi amaçlamışlardır. Farklı değişkenlerin amputasyonla ilişkisi çok değişkenli Cox orantılı risk regresyon modelleri kullanılarak analiz edilmiştir. İlk olarak, bütün modele karşılık gelen skor, hastaların tüm demografik özellikleri, ayak, ülser ve infeksiyon karakteristiklerini içeren 1 yıllık amputasyon insidansına ilişkin Lojistik Regresyon modelinin doğrusal terimi olarak tanımlanmıştır. Daha

sonra, bu bütün model, en az anlamlı öngörücünün iteratif olarak (geriye elimine edilerek tüm öngörücüler anlamlı olana kadar azaltılmıştır ($p < 0,05$). Nihai risk puanları, kullanılabilirliği artırmak için parametrelerin yuvarlandığı indirgenmiş modellerin bir versiyonu olarak belirlenmiştir. Bütün modeller ve risk puanları, IWGDF / IDSA sınıflandırması ile birlikte alıcı işlem karakteristiği (ROC) eğrileri ile değerlendirilmiştir. Araştırma sonucunda amputasyon için bağımsız risk faktörleri olarak yara çevresindeki ödem, kötü koku, pürülan olmayan eksüda, derin ülser, probdan kemiğe pozitif test sonucu, pretibial ödem, ateş (>38.0 C) ve yüksek (üst sınırın üç katı) C-reaktif protein olarak tespit edilmiştir. Artan IWGDF enfeksiyon şiddeti de bağımsız olarak amputasyonu öngörmüştür. Çalışmada ayrıca herhangi bir amputasyon ve küçük ayak parmakları hariç amputasyon için (cinsiyet, palpasyonda ağrı, yara çevresi ödem, ülser boyutu, ülser derinliği ve periferik arter hastalığı değişkenleri dahil olmak üzere) bir risk skoru geliştirilmiştir. Geliştirilen risk skoru amputasyonu IWGDF sisteminden daha iyi tahmin etmiştir (sırasıyla ROC eğrilerinin altındaki alan 0,80, 0,78 ve 0,67). Araştırma bulgularının diğer kohortlarda doğrulanması gerektiğinin altı çizilmiştir.

Schoen, Glance ve Thompson (2015) ayak bileği tepe sistolik hızının (APVS) diyabetik ayak lezyonlarında iyileşmesini tahmin edip edemeyeceğini belirlemek amacıyla yaptıkları araştırmada, 100 diyabetik hastayı prospektif olarak takip etmişler ve alt ekstremitte arterlerinin taraması için sevk edilen diyabetik hastalar, eğer; ülser, gangren veya doku nekrozu gibi ayak lezyonları varsa ve pedal nabızları palpabl değilse çalışmaya dahil edilmiştir. İstatistiksel analiz, iki grubu karşılaştırmak için Mann-Whitney U testi ve normal dağılmayan sürekli değişkenler için Wilcoxon testi kullanılarak, kesikli değişkenler için ki-kare testi gerçekleştirilmek suretiyle yapılmış olup, yaranın iyileşmemesinin bağımsız öngörücülerini belirlemek için Lojistik Regresyon analizi uygulanmıştır. Araştırmada, APVS'nin maksimum duyarlılık ve özgüllük ile optimum kesme değerini belirlemek için ROC eğrisi kullanılmıştır. APSV, iyileşmiş veya iyileşen lezyonları olan uzuvlarda, iyileşmemiş lezyonlu uzuvlara kıyasla anlamlı olarak daha yüksek olarak tespit edilirken (53,0 cm/s (41,8–81,6) kıyasla 19,2 cm/s (12,4–26,5), $p < 0,0001$). APSV değerinin, 35 cm / s'lik bir kesme noktasında, diyabetik ayak lezyonlarının iyileşmemesini öngörmeye, %92,9'luk bir duyarlılık, %92,9'luk bir pozitif öngörü değeri ve %90,6'lık negatif öngörü değeri göstermiştir. Revaskülarizasyon öncesi ve sonrası APSV değerleri arasında anlamlı farklılık saptanmıştır (20,4 cm / s (12,4–26,3) ve 48,8 cm / s (36,1–80,8), $p < 0,0001$). Araştırmada sonuç olarak APSV değerinin, bu hasta grubunda yüksek doğrulukla diyabetik ayak lezyonlarının iyileşmeyeceğini tahmin edebileceği bulunmuştur.

Tardivo vd. (2015) tardivo algoritması olarak adlandırılan diyabetik ayağı değerlendirmek ve amputasyonu tahmin etmek için basit bir prognostik skor oluşturmak amacıyla bir çalışma gerçekleştirmişlerdir. Bu çalışmada diyabetik ayak amputasyon risk skorunun hesaplanması için 62 diyabetik ayak hastasından alınan verilerle birlikte üç ana değişken kullanılmış olup bunlar; ülserin Wagner sınıflandırması, Periferik Arter Hastalığı Sınıflandırması kullanılarak değerlendirilen Periferik Arter Hastalığı (PAH) belirtileri ve ülserin yeri olarak belirlenmiştir. Çalışmada değerlendirilen hastaların %85,5'i iyileşmiş ve uzuvları amputasyondan kurtulmuştur. Gerçekleştirilen Lojistik Regresyon analizine göre hesaplanan olasılık oranına göre (CI 95%, 12,2–1886,5), 12 veya daha yüksek puana sahip hastaların (nihai skor yukarıda belirtilen değişkenlere göre bu üç puanın çarpımından oluşup 1'den 32'ye kadar bir puan aralığındadır) amputasyon olasılığının anlamlı derecede yüksek olduğunu göstermiştir. Tardivo algoritmasının, diyabetik ayak için doktorlar tarafından kolayca erişilebilen basit bir prognostik ölçü olduğu belirtilmiştir. Bu algoritma ile ister konservatif ister cerrahi tedavi olsun amputasyon riskinin ve en iyi tedavi rejiminin belirlenmesine yardımcı olacağı ifade edilmiştir.

Ai ve diğ. (2016) serum sistatin C düzeyleri ile diyabetik ayak ülserlerininprognozu arasındaki ilişkiyi saptamak amacıyla gerçekleştirdikleri araştırmada, tip 2 diyabetli 1018 hastayı içeren popülasyon temelli bir kohort çalışması yapmışlar ve hastaları diyabetik olmayan ayak ülseri grubu (NDF, n = 865,% 85,5) ve diyabetik ayak ülseri grubu (DFU, n = 147,% 14,5) olmak üzere 2 gruba ayırarak, 1 yıllık takipten sonra DFU'ları klinik prognoza göre iyileşen (n = 110,% 74,8) ve iyileşmeyen (n = 37,% 25,2) olmak üzere tekrar 2 gruba ayırmışlardır. Araştırmada,sürekli değişkenler için gruplar arasındaki farklılıklar Student's t testi, Kruskal-Wallis testi veya ANOVA testi kullanılarak değerlendirilmiş olup, kategorik değişkenler ki-kare testi kullanılarak analiz edilmiştir. Spearman korelasyon analizi, farklı klinik özellikler arasındaki karşılıklı ilişkiyi keşfetmek için gerçekleştirilmiş olup, DFU sonlanımının tahmininde farklı klinik faktörlerin öngörücü değerini çoklu Lojistik Regresyon analizi gerçekleştirilmek suretiyle belirlenmiştir. DFU'nin muhtemel sonlanımı için sistatin C'nin optimal sınırını saptamak amacıyla ROC eğrisi analizi yapılmıştır. Araştırmada, iyileşme grubu ile karşılaştırıldığında, iyileşmeyen grubun daha yaşlı, daha uzun diyabet süresine sahip olduğu ve DFU'da önemli ölçüde artmış serum sistatin C konsantrasyonlarına sahip olduğu saptanmıştır (p <0.01). Çoklu Lojistik Regresyon analizine göre, sistatin C'nin istenmeyen DFU sonucu riskinde artışla ilişkili olarak saptanmıştır (OR: 7.279,%95 CI: 1,299-40,784, p <0,05). Sistatin C, düzeylerine göre çeyreklere ayrıldığında, çeyrek 4'ün iyileşme oranı diğer gruplara göre anlamlı olarak daha düşük olarak tespit edilmiştir (%57,9) (p <0.01). Araştırmada, serum

sistatin C ile diyabetik ayak ülseri prognozu arasında güçlü ve bağımsız bir ilişki olduğu sonucuna varılmış olup, sistatin C düzeylerinin $> 1,35$ mg / L kesme değerinin, tedavi edilemez ayak ülseri riskini altı kattan fazla artırdığını öngördüğü sonucuna varılmıştır.

Beaney ve diğ. (2016) multidisipliner bir diyabetik ayak kliniğinde tedavi gören hastalarda minör veya majör amputasyon gerekliliğini hangi faktörlerin öngördüğünü belirlemek amacı ile yaptıkları araştırmada, 165 hastaya ilişkin verileri retrospektif olarak analiz etmişlerdir. Amputasyon, Lojistik Regresyon modeli kullanılarak %5 anlamlılık düzeyinde ileri aşamalı model seçimi ile modellenmiştir. Başlangıçtaki zayıf glisemik kontrol ile diğer faktörlere göre düzeltildiğinde amputasyon riski arasında bir ilişki olduğu bulunurken ($HbA1c \leq 58$ mmol / mol (%7,5) daha az amputasyon riski ile ilişkili bulunmuştur [0,14 olasılık oranı ile (OR: 0,04-0,53; $p = 0,0036$)]. Amputasyonu öngören diğer istatistiksel olarak anlamlı faktörler ise; klinik randevuları kaçırma ($p = 0,0079$); yüksek Charlson indeksi ($p = 0,03314$) ve hipertansiyon ($p = 0,0216$) olarak tespit edilmiştir. Önceden geçirilmiş revaskülarizasyon amputasyona karşı koruyucu olarak saptanmamıştır ($p = 0,0035$). Bu kohortta zayıf glisemik kontrol, klinik randevuları takip etmeme, önceki revaskülarizasyon ve hipertansiyon daha yüksek amputasyon riski ile ilişkilendirilmiştir. Ayrıca, diyabetik ayak kliniğinde uzun süreli takip sonucu elde edilen glisemik kontrolün faydaları da bu çalışmada gösterilmiş olup, bu sonucun daha uzun vadede diyabetle ilgili sağlık faydalarına dönüşebileceği vurgulanmıştır.

Bravo-Molina ve diğ. (2016) diyabetik ayak sendromunda kullanılan yara sınıflandırma sistemlerinden TEXAS, PEDIS ve Wagner yara sınıflandırmalarının amputasyonu tahmin edip etmediğini belirlemek amacıyla Lojistik Regresyon analizi kullanarak yaptıkları araştırmada, 250 diyabetik ayak hastasını prospektif olarak takip etmişler ve yara sınıflandırma sistemlerindeki derecelerini kaydetmişlerdir. Araştırmada kullanılan üç yara sınıflandırma sistemi, minör amputasyonla ilişkili olmasına rağmen amputasyonu tahmin edememiştir. Yapılan Lojistik Regresyon analizine göre minör amputasyonun öngörücüleri olarak; osteomyelit varlığı, yaranın ayaktaki lokasyonu ve majör amputasyonun öngörücüsü olarak da yüksek CRP seviyeleri değişkenleri tespit edilirken, ABI seviyeleri, takipte majör amputasyonu öngörmüştür.

Fife ve diğ. (2016) diyabetik ayak ülseri (DFU) olan hastalar için uygulamada, araştırmalarda ve klinik çalışmalarda kullanılmak üzere bir iyileşme indeksi geliştirmek amacıyla yaptıkları çalışmada, Amerika Birleşik Devletleri Yara Evrak Dairesi verilerinden 6440 veri retrospektif olarak incelenmiş ve iki değişkenli analizlerle önemli değişkenler belirlenmiş, önemli faktörlere

($p < 0,05$) dayalı olarak çok değişkenli bir Lojistik Regresyon modeli oluşturulmuş ve bir veri örneği üzerinde test edilmiştir. Lojistik Regresyon modelinde iyileşmeyi önemli ölçüde öngören değişkenler; ülser süresi, ülser boyutu, herhangi bir etiyolojiye sahip eş zamanlı ülser sayısı, infeksiyon, yaş, Wagner derecesi, ayak tedavi görmeme durumu, diyaliz, böbrek nakli, periferik vasküler hastalık ve herhangi bir nedenle hastanın hastaneye kaldırılması olarak saptanmıştır. Çalışmada, daha önce Yara İyileşme İndeksi (WHI) olarak tanımlanan, DFU'lu hastaların iyileşme olasılığını öngören, hastaya ve yaraya özgü değişkenleri içeren, doğrulanmış bir tabakalandırma sistemi sunulmuştur. Sonuç olarak çalışmada önerilen modelin, diyabetik ayak ülserlerinin iyileşmesini tahmin etmek için kullanılabilecek kapsamlı, kullanıcı dostu, doğrulanmış bir öngörücü model olduğu bulgusuna ulaşılmıştır.

Beyaz ve diğ. (2017) diyabetik ayak sendromu nedeniyle diz altı amputasyon geçiren diyabetik hastalarda sağkalımı tahmin etmede kullanılabilecek parametreleri belirlemek amacıyla yaptıkları araştırmada, diz altı amputasyon uygulanan 470 hastayı operasyondan ölüm zamanına kadar geçen süre bazında retrospektif olarak değerlendirmişlerdir. Çalışmadaki verilerin normalliği Shapiro Wilk testi ile analiz edilmiş, normal dağılım gösteren değişkenlerin iki bağımsız grubunu karşılaştırmak için Student t testi kullanılırken, normal olmayan veriler için Mann-Whitney U testi kullanılmıştır. Kategorik değişkenler arasındaki ilişkileri araştırmak için Pearson ki-kare testi kullanılmış olup, tahmini olasılık oranları ve %95 güven aralıkları (CI) Lojistik Regresyon yöntemi ile hesaplanmıştır. Araştırmada sağkalım olasılıkları Kaplan Meier analizi kullanılarak tahmin edilmiştir. Ayrıca, genel sağkalım olasılıklarını etkileyen önemli faktörleri belirlemek için Cox regresyon analizi gerçekleştirilmiştir. Araştırma sonucunda, hastaların sağkalım oranlarının ilk yıldan sonra %64 olduğu ve ameliyat sonrası ortalama yaşam beklentisinin 930 ± 106 gün olduğu saptanmıştır. Hemodiyaliz tedavisi ($p = 0,001$), endovasküler girişim ($p = 0,04$), cinsiyet ($p = 0,004$), yaş ($p = 0,001$), BUN düzeyi ($p = 0,001$) ve insülin kullanım süresi ($p = 0,003$) değişkenlerinin mortalite için anlamlı öngörücüler olduğu saptanmıştır.

Kasbekar ve diğ. (2017) diyabetik ayak hastalarında amputasyon riskini tahmin etmek için kanıta dayalı bir araç oluşturmak amacıyla yaptıkları araştırmada, 301 diyabetik ayak hastasının (diabetes mellitus öyküsü ve alt ekstremitte ülseri olan) klinik kayıtlarını retrospektif olarak incelemişler, yapay zekâ karar ağacı algoritmalarından “C5.0” ve “*boosted ensemble C5.0*” algoritmaları ile tahmin modelleri oluşturmuşlardır. Yapılan yapay zekâ-makine öğrenmesi analizleri sonucunda amputasyon kararı için en etkili belirleyicilerin Doppler ölçümleri ve ülserasyonun Wagner derecelendirmesi olduğu tespit edilmiştir. İlk analiz olan C5.0

sınıflandırıcısı ile, %94 doğruluk değeri elde edilirken, ikinci algoritma olan “*boosted ensemble C5.0*” sınıflandırıcısı ile %100 ve %96 doğruluk değerleri elde edilmiştir. Araştırmada sonuç olarak günümüzün kişiye özel önleyici ve hassas tıp çağında, geliştirilen tahmin modellerinin (bu iki sınıflandırıcının) diyabetik ayaklı hastalarda uzvun prognozu için doğru bir kılavuz görevi görebileceğinin ve gelecekteki amputasyon riskini tahmin edebileceğinin altı çizilmiştir.

Lanting ve diğ. (2017) diyabetli hastalarda ayak komplikasyonu öyküsü (ülserasyon veya amputasyon) ile noninvaziv vasküler değerlendirmeler arasındaki ilişkiyi araştırmak amacıyla yaptıkları araştırmada, 127 hastanın bilateral ABI indeksi, “*Toe Brachial Index*”-TBI indeksi ve sürekli Doppler (CWD) ölçümlerini gerçekleştirmişlerdir. Diyabetik ayak komplikasyonu için bilinen risk faktörleri ve hasta öyküsü arasında Spearman korelasyon analizi gerçekleştirilmiş ve TBI’ nin önceki bir diyabetik ayak komplikasyonu olasılığı üzerindeki etkisini saptamak için Lojistik Regresyon analizi gerçekleştirilmiştir. Lojistik Regresyona analizine göre, TBI <0,6 (OR : 7,74, p = 0,001), ve daha uzun diyabet süresi olan hastalarda (OR : 1.06, p = 0,05) diyabetik ayak komplikasyon riskini en yüksek olarak saptanmıştır. HbA1c, diyabetik ayağa bağlı komplikasyon riskini bağımsız olarak öngörmemiştir (OR : 1,10, p = 0,356). Sonuç olarak, TBI indeksini, alt ekstremit vasküler hastalığının bir ölçüsü olarak önceliklendirmenin, diyabetik ayağa bağlı komplikasyon riski taşıyan hastaları prospektif olarak saptamak için yararlı olabileceğinin altı çizilmiştir.

Wu ve diğ. (2017) periferik arter hastalığı ile ilişkili diz altı amputasyon sonrasında hasta özelliklerini ve sonuçlarını gözden geçirmek ve diz üstü amputasyonu öngören risk faktörlerini belirlemek amacıyla yaptıkları araştırmada, 210 diz altı amputasyon operasyonu geçiren hastayı retrospektif olarak takip etmişlerdir. Öncelikle, tek değişkenli analiz, kategorik değişkenler için Fisher testi ve sürekli değişkenler için Student t-testi kullanılarak gerçekleştirilmiştir. Diz altı amputasyon operasyonunun başarısızlığı ile ilişkili bağımsız risk faktörlerini tanımlamak amacıyla $p \leq 0,10$ olan değişkenler, çoklu Lojistik Regresyon analizi gerçekleştirilmek üzere seçilerek olasılık oranları (ORs) ve %95 güven aralıkları (CI) hesaplanmıştır. Gerçekleştirilen çok değişkenli analiz sonuçlarına göre, son dönem böbrek yetmezliği (Olasılık Oranı [OR] = 3,85; p = 0,01) ve preoperatif ambulatuvar olmayan durumun (OR = 5,58; p = 0,01) sonraki diz üstü amputasyon riskini tahmin etmede bağımsız risk faktörleri olarak tespit edilmiştir.

Ferreira ve diğ. (2018) diyabetik ayak ülseri (DFU) olan hastalarda alt ekstremitte amputasyonunun (“LEA”) kısa vadeli öngörücülerini belirlemek amacıyla gerçekleştirdikleri araştırmada, 479 hastanın klinik ve sosyodemografik verilerini retrospektif olarak incelemişler ve 180 gün içinde LEA oluşumunu Lojistik Regresyon modeli ile değerlendirmişlerdir. Gerçekleştirilen analiz sonucunda, kısa vadeli LEA'nın bağımsız öngörücüler olarak; orta ila şiddetli infeksiyon [OR: 5,23 (CI 95% 2,51-10,80; p <0,001)], geçmiş LEA öyküsü [OR: 3,93 (IC 95% 1,81–8,53; p = 0,001)], periferik arter hastalığı [OR: 3,51 (CI 95% 1,29–9,58; p = 0,014)] ve herhangi bir yürüme engelinin varlığı [OR: 3.35 (CI 95% 1,58–7,13; p = 0,002)] değişkenleri saptanmıştır.

Su ve diğ. (2018) diyabetik ayak ülseri komorbiditeleri ile ilişkili önceden var olan öngörücü faktörleri saptamak amacıyla yaptıkları araştırmada, King diyabetik ayak sınıflandırmalarından, King sınıflandırması derece III'de bulunan 100 hastayı prospektif olarak en az 1 yıl boyunca takip etmişlerdir. Araştırmada, hastalar ülserin tedavi yanıtına göre başarı grubu (ülserin tamamen kapanması veya küçülmesi), stabil grup ve tedavinin başarısızlığı grubu (amputasyon veya infeksiyonun tekrarlaması) olmak üzere 3 gruba ayrılmış olup, değişkenler için 3 gruplu karşılaştırmada χ^2 testi, Fisher testi ve tek yönlü varyans analizi kullanılmış olup, 2 gruplu karşılaştırmada Student t testi uygulanmıştır. Varyans analizi sonucunda p değeri 0,1'den küçük olan öngörücü faktörler ise, tek değişkenli Lojistik Regresyon modeli kullanılarak detaylı olarak analiz edilmiştir. Perkütan transluminal anjiyoplasti (PTA) öyküsü, tedavi başarısızlığı ile güçlü bir şekilde ilişkili olarak saptanmıştır ([OR]: 14,33; % 95 güven aralığı [CI], 1,71-120,32; p = 0,014), oysa retinopati bulunması (OR: 0,21;% 95 CI, 0,07-0,65; p = 0,006) tedavi başarısı için en önemli negatif öngörücü olarak saptanmıştır. Önceki debridman öyküsünün, tedavi başarısızlığını öngörmek için sınırdan önem derecesinde olarak bulunmuştur (OR: 0,09; %95 CI, 0,01-1,01; p = 0,051). Cinsiyet, yaş, dislipidemi, hipertansiyon, koroner arter hastalığı, serebrovasküler olay, kronik böbrek hastalığı ve son dönem böbrek hastalığı ve yara durumu değişkenleri ise istatistiksel olarak anlamlı olarak saptanamamıştır.

Altay ve diğ. (2019) nötrofil-lenfosit oranının (NLO) diyabetik ayak infeksiyonunda erken evrede risk değerlendirmesi yapıp yapmayacağı ve takibinde kullanılma durumunu saptamak amacıyla gerçekleştirdikleri araştırmada, toplamda 101 hastanın hastaneye kabul ve tedavinin 14. gününde NLR değerlerini, laboratuvar ve klinik verileriyle eşleştirdikleri bir kohort çalışmasında hastaları 5 yıl boyunca prospektif olarak takip etmişlerdir. Verilerin normal dağılımı Kolmogorov-Smirnov testi ile değerlendirilmiş karşılaştırmak için, iki grup sayısal

değişken, Student T testi veya Mann-Whitney U testi kullanılırken, kategorik değişkenler için ki kare testi veya Fischer testi kullanılmıştır. İlk ve 14. gün verilerinin karşılaştırması, eşleştirilmiş T-testi ve Wilcoxon işaretli sıra testi ile gerçekleştirilmiş olup, sayısal değişkenler arasındaki ilişkiyi analiz etmek için Spearman'ın korelasyon katsayısı kullanılmıştır. Bağımlı değişkenler için olası risk faktörleri, p değeri < 0.25 olan değişkenler olarak seçilmiştir. Bağımsız öngörücüler, çok değişkenli Lojistik Regresyon analizi yapılarak, geriye dönük yöntemlerle değerlendirilmiştir. Başvuru anındaki NLO değerinin hem Wagner hem de PEDIS sınıflandırma sistemlerine göre sınıflandırılan ülserlerde, ağır vakalarda daha yüksek olduğu saptanmıştır (artmış Wagner sınıflandırması için: NLR 6,7'ye karşı 4,2; $p = 0,04$; PEDIS sınıflandırması için NLR 6,3'e karşı 3,6; $p = 0,03$). Vasküler girişim yapılan hastalarda diğerleri ile eşleştirildiğinde (12,6'ya karşı 4,6; $p = 0,02$); amputasyon riski daha yüksek olduğu bulunurken, (9,2'ye karşı 4,1; $p = 0,005$) bu hastaların daha sonra iyileştiği bulunmuştur (6,9'a karşı 4,3; $p < 0,001$). NLO değerinin hem IV antibiyotik tedavisi süresi ($r = 0,374$; $p = 0,005$) hem de hastanede kalış süresi ($r = 0,337$; $p = 0,02$) ile pozitif korelasyon gösterdiği bulunmuştur. 14. gündeki NLO değerinin, vasküler girişim uygulanan hastalarda (5,1'e karşı 2,9; $p = 0,007$) diğer hastalar ile eşleştirildiğinde daha yüksek olduğu saptanmıştır. Araştırmada sonuç olarak başvuru anında daha yüksek NLR değerleri olan hastalarda daha şiddetli diyabetik ayak enfeksiyonu, daha yüksek amputasyon riski, uzun süreli hastanede yatış ihtiyacı ve agresif tedavi ihtiyacının yoğun olduğu saptanırken, yapılan tedaviler sonucu iyileşme şanslarının da daha fazla olduğu bulunmuştur.

Berli ve diğ. (2019) periferik arter hastalarında majör alt ekstremitte amputasyonlarının revizyon cerrahisinin potansiyel öngörücü faktörlerini belirlemeyi amaçladıkları çalışmalarında, 180 hastayı retrospektif olarak değerlendirmişler ve revizyon operasyonu yapılan ve yapılmayan hastalar arasındaki kategorik değişkenlerdeki farklılıklar ki-kare testi kullanarak, sürekli değişkenlerdeki farklılıkları Lojistik Regresyon yöntemi ile analiz etmişlerdir. Araştırmada, periferik arter hastalığının evresi revizyon amputasyon seviyesiyle ilişkili olarak saptanamamıştır ($p = 0,4369$). Daha önce balon anjiyoplasti geçiren önemli ölçüde daha fazla hastanın revizyon cerrahisi ihtiyacı olduğu bulunmuştur (%66,7'ye karşı %45,2; $p = 0,009$).

Chaudhary ve diğ. (2019) charcot nöroartropatisi (CN) olan veya olmayan diyabetli hastalarda mortaliteyi incelemek amacıyla gerçekleştirdikleri araştırmada, CN bulunan hastalar (Kohort 1) ile ayak komplikasyonu olmayan hastaları (Kohort 2) diyabet süresi, mikrovasküler ve makrovasküler komplikasyonlar, ayak ülseri ile ilgili veriler, amputasyon ve mortalite sonuçları ile ilgili detaylar prospektif olarak kaydedilerek çok değişkenli Lojistik Regresyon analizi

kullanılarak karşılaştırılmıştır. Araştırmada, 260 CN'li diyabet hastası ve CN'siz 520 hastaya ait veriler analiz edilmiştir. CN'li kişilerin daha genç ($55 \pm 9,1$ 'e karşı $59,9 \pm 8,1$ yıl, $p < 0,001$) ve mikrovasküler komplikasyon prevalansının daha yüksek olduğu saptanmıştır. CN'li hastaların, 2,7 kat (OR: 2,72,%95 CI 1,4–5,2, $p = 0,003$) artmış mortalite riskine sahip olduğu bulunurken, yaygın koroner arter hastalığı semptomlarının bulunması ve düşük eGFR, CN'li kişilerde daha yüksek mortalite riski ile ilişkili olarak bulunmuştur.

Lee ve diğ. (2019) deri hidrasyon düzeyinin diyabetik yara iyileşmesi üzerindeki etkisini değerlendirmek ve diyabetik yara iyileşmesini öngörmeye en güvenilir parametre olarak kabul edilen doku oksijenasyon düzeyiyle karşılaştırmak amacıyla gerçekleştirdikleri araştırmada, diyabetik ayak ülserli 263 hastayı retrospektif olarak değerlendirmişlerdir. Araştırmada, perkütan transluminal anjiyoplasti öncesi ve sonrası toplanan deri hidrasyonu ve transkutanöz oksijen basıncı verileri analiz edilmiş olup, deri hidrasyonu ve doku oksijenasyonu zayıf, orta veya kabul edilebilir olarak derecelendirilmiş, yara iyileşmesinin sonlanımı ise, amputasyon olmasının iyileşme, minör amputasyon veya majör amputasyon olarak 3 gruba ayrılarak derecelendirilmiştir. Araştırmada ülserlerin iyileşme sonlanımları ise; başlangıçtaki deri hidrasyonu, başlangıçta deri altı oksijen basıncı, perkütan translüminal anjiyoplasti sonrası deri hidrasyonu ve perkütan translüminal anjiyoplasti sonrası deri altı oksijen basıncı olmak üzere dört parametre kullanılarak karşılaştırılmıştır. Araştırmada deri hidrasyonu ve deri altı oksijen basıncı alt gruplarını sonuç değişkenleriyle karşılaştırmak için Pearson ki-kare testi kullanılmış, Wilcoxon işaretli sıra testi, perkütan transluminal anjiyoplasti sonrası ölçümler arasındaki transkutanöz oksijen basıncı / cilt hidrasyon değerlerinin istatistiksel analizi amacıyla gerçekleştirilmiştir. Mantel-Haenszel ki-kare testi, deri hidrasyonu ve deri altı oksijen basıncı alt gruplarındaki doğrusal eğilimleri sonuç değişkenleriyle karşılaştırmak için kullanılmıştır. Başlangıçtaki ve prosedür sonrası perkütan transluminal anjiyoplastideki öngörücü değişkenlerin majör amputasyon riski üzerindeki etkileri, çok değişkenli Lojistik Regresyon analizi kullanılarak analiz edilmiştir. Araştırma sonucunda, belirlenen dört parametrenin her birinin, yara iyileşmesinin muhtemel sonlanımı açısından istatistiksel olarak önemli korelasyonlar sergilemiştir. Hem cilt hidrasyonu hem de deri altı oksijen basıncının eş zamanlı analizinde, başlangıçtaki deri hidrasyonu önemli bir parametre ($p = 0,0018$) iken, deri altı oksijen basıncı perkütan translüminal anjiyoplastiyi takiben istatistiksel olarak anlamlı bir parametre olarak saptanmıştır ($p < 0,0001$). Araştırma sonucunda, deri hidrasyon seviyelerinin, diyabetik ayak hastalarının yara iyileşmesini öngörmesi açısından yararlı bir öngörücü olabileceği vurgulanmış olup, özellikle revaskülarizasyondan önceki deri hidrasyon

seviyesinin, yara iyileşmesini öngörmeye transkutanöz oksijen basıncından daha üstün olduğu saptanmıştır.

Mineoka ve diğ. (2019) tip 2 diyabetli hastalarda, diyabetik ayak sendromu riski veya diyabetik ayak ülseri bulunan hastalar ile trombosit lenfosit oranı (PLR) arasındaki ilişkiyi saptamak amacıyla yaptıkları kesitsel araştırmada, 453 hastanın klinik ve demografik verilerini değerlendirmişlerdir. Araştırmada, gruplar arasındaki farklılıkların tespitinde, Student t-testi, varyans analizi, Pearson χ^2 testi kullanılmıştır. Değişkenler arasındaki ilişkileri değerlendirmek amacıyla Pearson'un korelasyon katsayısı kullanılmıştır. Diyabetik ayak ülseri için PLR'nin kesme değerini belirlemek ve duyarlılık, özgüllük ve eğri altındaki alanı hesaplamak amacıyla ROC eğrileri kullanılmıştır. Bununla birlikte, diyabetik ayak ülseri açısından yüksek riskli veya diyabetik ayak ülseri olan hastalar ile PLR arasındaki öngörücü faktörleri saptamak için tek değişkenli ve çok değişkenli Lojistik Regresyon analizleri gerçekleştirilmiştir. Diyabetik ayak sendromu açısından yüksek riskli veya diyabetik ayak ülseri olan hastalarda PLR değerinin daha yüksek olduğu saptanmıştır. (117 ± 40 'a karşı 107 ± 31 , $p = 0,003$ ve 148 ± 65 'e karşı 113 ± 56 , $p < 0,001$). ROC eğrisi analizine göre, PLR'nin 130,6'lık kesme değerinin %85 duyarlılık ve %70 özgüllük ile diyabetik ayak ülseri için optimum kesme değerini olduğu gösterilmiştir. Çok değişkenli Lojistik Regresyon analizine göre, PLR'nin yaygın diyabetik ayak ülseri (OR: 1,02; %95 CI 1,01-1,04, $p = 0,003$) ile pozitif korelasyon gösterdiği tespit edilmiştir. Araştırmada sonuç olarak, PLR'nin tip 2 diyabetli hastalarda diyabetik ayak sendromu açısından yüksek riskli ve diyabetik ayak hastaları için bir anlamlı bir öngörücü olabileceği gösterilmiştir.

Orozco-Covarrubias ve diğ. (2019) tip II diyabetli hastalardaki amputasyon riski için prognostik faktörleri belirlemek amacıyla gerçekleştirdikleri araştırmada, 134 tip II diyabetli ve amputasyon öyküsü bulunan vaka ile 134 kontrol grubunu (tip II diyabetli ve amputasyon öyküsü bulunmayan) 1: 1 oranında eşleştirerek olmak üzere toplam 268 diyabetik hasta ile bir vaka-kontrol çalışması gerçekleştirmişlerdir. Risk faktörleri ile amputasyon arasındaki ilişkinin analizi, olasılık oranları (OR) hesaplanarak ve amputasyon riski için belirlenen önemli faktörlerle Lojistik bir tahmin modeli geliştirilmek suretiyle tabakalı tesadüfi örnekleme, yöntemi kullanılarak yapılmıştır. Araştırma sonucunda, HbA1c'nin %8'den büyük olmasının amputasyona yol açan patolojilerin (alt ekstremitelerde ülser varlığı (OR: 2,8, $p = 0,00$), nöropati varlığı (OR: 3,6, $p = 0,01$) ve periferik arter hastalığı (OR: 2,5, $p = 0,05$) gelişiminde ana prognostik faktör olduğu bulunmuştur.

Sarwar ve diğ. (2019) diyabetik ayak hastalarının laboratuvar ve klinik parametrelerini değerlendirerek amputasyon riskini tespit etmek amacıyla yaptıkları gözlemsel araştırmada, 574 hastayı 6 ay boyunca takip etmişler ve elde ettikleri verilerin istatistiksel analizini Lojistik Regresyon analizi ve tek yönlü ANOVA testi kullanmak suretiyle gerçekleştirmişlerdir. Araştırmada, Wagner Sınıflandırma sistemi ile değerlendirilen osteomyelit, iskemi, gangren ve ülser derinliğinin ve ileri yaş, koroner arter hastalığı, ülser boyutu ve sigara kullanımı değişkenlerinin hem genel amputasyonun hem de majör amputasyonun başlıca öngörücüleri olduğu saptanmıştır. Ayrıca, başlangıçtaki akut faz reaktan seviyeleri (polimorfonükleer lökosit sayısı, beyaz kan hücresi sayısı, eritrosit sedimentasyon hızı (ESR), trombosit sayısı, albümin ve serum C-reaktif protein (CRP)) ve azalmış hemoglobin seviyeleri de amputasyon riski ile ilişkilendirilmiştir. ESR ve CRP'nin başlangıç düzeylerindeki standart sapmadaki artışın hem genel amputasyonları hem de majör amputasyonları öngördüğü de gerçekleştirilen çok değişkenli analiz yoluyla açığa çıkarılmıştır.

Sayiner ve diğ. (2019) hastalar arasında diyabetik ayak amputasyonu ihtiyacının belirleyicileri olabilecek klinik özellikleri tanımlamak amacıyla yaptıkları, retrospektif gözlemsel çalışmada, 18 yaş üstü ve diyabetik ayak ülseri olan toplam 400 tip 2 diyabetli hastayı incelemişler ve bu hastaların yaş, cinsiyet, diyabet süresi, sigara öyküsü, geçirilmiş diyabetik ayak amputasyonu, hipertansiyon varlığı (HT), koroner arter hastalığı (KAH), perifer arter hastalığı, gibi klinik, radyolojik ve sintigrafik profilleri, serebrovasküler hastalık, kardiyovasküler hastalık, retinopati, nöropati ve nefropati öyküleri analiz edilmiştir. Çalışmada elde edilen verilerin analizinde, verilerin normal dağılım gösterip göstermediğini belirlemek için Kolmogorov Smirnov testi kullanılmış, klinik verilerin amputasyon sonucu ile ilişkisini incelemek için ki-kare testi ve bazı klinik ve laboratuvar verilerinin amputasyon ile ilişkisini belirlemek için Mann-Whitney U testi kullanılmış olup, değişkenlerin olasılık oranını hesaplamak için tek değişkenli ikili Lojistik Regresyon analizi gerçekleştirilmiştir. Çok değişkenli ikili Lojistik Regresyon analizi, tek değişkenli analize göre istatistiksel olarak anlamlı değişkenlerden sadece eşdoğrusallığı olmayan değişkenler modele dahil edilmek suretiyle gerçekleştirilmiştir. Ampüte hastalarda proteinüri sıklığı, ampüte olmayanlara göre anlamlı olarak daha yüksek olarak saptanmıştır ($p < 0,05$). Ampüteler ayrıca anlamlı olarak daha uzun sigara içme öyküleri sergilerlerken ($p < 0,001$) geçmiş amputasyon öyküsüne sahip olanlarda yeniden amputasyon sıklığı anlamlı olarak daha yüksek olarak bulunmuştur ($p = 0,038$). Çok değişkenli Lojistik Regresyon analizinden sonra, perifer arter hastalığı varlığı ve geçmiş diyabetik ülser öyküsü amputasyon kararını belirleyen önemli faktörler olarak saptanmıştır. Erkek cinsiyet, koroner

arter hastalığı öyküsü, perifer arter hastalığı, hipertansiyon öyküsü, proteinüri, Wagner derece 4-5 olan ülserler, sigara içme öyküleri, geçmiş diyabetik ülser öyküsü ve geçmiş amputasyon öyküsü değişkenlerinin amputasyon ihtiyacını önceden belirlemede etkili olduğu belirlenmiştir. Amputasyona ihtiyaç duymayan hastalarda modelin bu değişkenler ile özgüllüğü %86 olarak belirlenmiştir. Çok değişkenli Lojistik Regresyon analizi sonuçlarına göre anlamlı öngörücüler; periferik arter hastalığı varlığı ve diyabetik ayak ülseri öyküsü olarak tespit edilmiştir.

Ugwu ve diğ. (2019) diyabetik ayak ülserlerini değerlendirmek amacıyla, altı üçüncü basamak sağlık kurumunda diyabetik ayak infeksiyonu nedeniyle hastaneye yatırılan ve taburcu oluncaya veya ölüme kadar multidisipliner bir ekip tarafından süreçleri takip edilen 336 diyabetik hastayı prospektif olarak takip etmişlerdir. Kategorik değişkenler için Ki-Kare testi ve sürekli değişkenler için t-testini kullanılarak demografik, klinik ve laboratuvar değişkenleri ile amputasyon (majör veya minör) arasındaki ilişkiler yapılan alt analiz aracılığı ile belirlenmiştir. Amputasyonun bağımsız prediktörlerini belirlemek amacıyla, ilk olarak, bağımlı değişken olarak amputasyon ile her değişken için tek değişkenli Lojistik Regresyon analizi gerçekleştirilmiş ve değişkenler için olasılık oranları (OR'ler) ve %95 güven aralıkları (CI) hesaplanmıştır. Sonrasında bu tek değişkenli analiz sonucunda, önemli öngörücüler olarak ortaya çıkan tüm değişkenler eşzamanlı olarak geriye doğru bir seçim yöntemi kullanılarak indirgenen çok değişkenli bir regresyon modeline girilmek suretiyle ikinci bir analiz işlemi gerçekleştirilmiştir. Amputasyonun tek değişkenli prediktörleri olarak; hastaneye yatmadan önce 1 aydan daha uzun ülser süresi, perifer arter hastalığı bulunması, Wagner derecesi ≥ 4 üzerinde olması, infeksiyon, proteinüri, lökositoz ve osteomyelit olarak tespit edilen, gerçekleştirilen çok değişkenli regresyon analizi sonuçlarına göre, amputasyonun anlamlı bağımsız prediktörleri olarak sadece üç değişken ortaya çıkmıştır ve bunlar da; 1 aydan uzun ülser süresi, perifer arter hastalığı ve osteomyelit varlığı olarak saptanmıştır. Çalışılan popülasyonda yaş, cinsiyet, diyabet türü ve süresi, nöropati, glisemik kontrol ve anemi amputasyonu öngörmemiştir. Bu risk faktörlerine dikkat edilmesi ile, bu hastalar arasındaki amputasyon oranının düşebileceği ifade edilmiştir.

Adeleye ve diğ. (2020) Nijerya'da diyabetik ayak ülseri (DFU) nedeniyle hastaneye yatırılan hastalarda mortaliteyi öngören faktörleri belirlemek amacıyla ve Nijerya'nın 6 jeopolitik bölgesinde altı üçüncü basamak sağlık kurumunda yürütülen gözlemsel bir çalışma olarak tasarlanan Nijerya'daki Diyabetik Ayak Ülserinin Çok Merkezli Değerlendirmesi çalışmasının bir parçası olarak gerçekleştirdikleri araştırmada, diyabetik ayak ülseri tedavisi için hastaneye yatırılan ve araştırmaya katılmayı kabul eden ardışık tip 2 diyabetli toplamda 323 hastayı

taburculuk veya ölüme kadar gözlemlemişler ve hastalar ile ilgili klinik, biyokimyasal ve radyolojik değerlendirmeleri kayıt altına alarak, hastaları ise multidisipliner bakıma tabi tutmuşlardır. Araştırmada mortalite riski ile ilişkili öngörücü faktörleri tespit etmek amacıyla tek değişkenli Lojistik Regresyon analizi gerçekleştirilmiş olup, değişkenler için olasılık oranları ve %95 güven aralıkları hesaplanmıştır. Bu tek değişkenli analiz düzeyinde mortaliteyi önemli ölçüde öngördüğü gösterilen değişkenler, daha sonra model güvenilirliği için Hosmer-Lemeshow uyum iyiliği testine dayanarak eş zamanlı olarak kademeli geriye doğru çok değişkenli bir Lojistik Regresyon analizine tabi tutulmuştur. Gerçekleştirilen analiz sonuçlarına göre, mortalite riski ile ilişkili değişkenler; 120 aydan fazla diyabet süresi ($p < 0,005$), ülser süresi > 1 ay ($p = 0,020$), Wagner derece 3 ve üstü ülser şiddeti ($p = 0,001$), periferik arter hastalığı ($p = 0,005$), proteinüri ($p < 0,001$), pozitif kan kültür örneği ($p < 0,001$), düşük HDL ($p < 0,001$), başvuru anındaki şok tablosu ($p < 0,001$), kalp yetmezliği ($p = 0,027$) ve böbrek yetmezliği ($p < 0,001$) olarak saptanmıştır. Çok değişkenli regresyon analizinde ise, bakteremi varlığı (OR: 5,053;%95 CI 2,572–9,428) ve böbrek yetmezliği (OR: 2,838;%95 CI 1,349–5,971), diğer değişkenlerden bağımsız olarak mortaliteyi önemli ölçüde öngörmüştür. Sonuç olarak bu araştırmada, DFU hastalarında yüksek hastane içi mortalite riskinin bulunduğunu göstermiş olup, ölümlerin çoğunun ileri ülser, bakteremi, kalp yetmezliği ve böbrek yetmezliği olanlarda meydana gelmiştir. Araştırmada tespit edilen bu faktörlere diyabetik ayak ülserli hasta popülasyonunda ivedi olarak dikkat edilmesi gerekliliği vurgulanmıştır.

Goyal ve diğ. (2020) DFU'da enfeksiyon ve iskemi varlığını tanımlamak için yeni bir veri seti üzerinde görüntü işleme teknikleri kullanmışlardır. Toplamda 1459 DFU fotoğrafından oluşan ve iskemi ve enfeksiyon vakalarının sınıf etiketlerine sahip bu veri seti, ilk kez bu araştırma amacıyla kullanılmıştır. NHS'nin etik kurul kararı ile alınan mevcut veri setinin herhangi bir kayıt veya meta veri(veriye ilişkin veri,bilgi) içermediği ifade edilmiştir. İlk kez bu araştırma için kullanılan DFU görüntülerinden oluşan bu yeni veri setinin halka açık olmadığı için iskemi ve enfeksiyonun sınıflandırılması görüntülere dayanarak tek tek uzman gözetiminde gerçekleştirildiği ifade edilmiştir. Her DFU'nun Texas sınıflandırma sistemini takip eden uzman gözetimindeki etiketleme işleminin, bilgisayarın öğrenme aşamasında oldukça kritik bir öneme sahip olduğu ifade edilmiştir. DFU görüntülerinin sınıflandırma süreci, iki sağlık uzmanı (diyabetik ayak konusunda uzman danışman doktorlar) tarafından gerçekleştirilmiştir. Sınıflandırma konusunda anlaşmazlık olduğunda, nihai karar daha kıdemli doktor tarafından verilmiştir. Bu sınıflandırma sonuçları, enfeksiyonun ve iskeminin ikili sınıflandırması için kullanılmıştır. Çalışmada, bilgisayarla görme tekniği ile DFU'daki iskemi ve enfeksiyonun

tanımlanması ve görüntülerden renk ROI'sini çıkarmak için yeni bir Superpixel Renk Tanımlayıcıları (SPCD) önerilmiştir. Bu sınıflandırma problemleri için, bir dizi sınıflandırma algoritmaları da denenmiştir. Bunlar arasında, BayesNet, Random Forest ve Çok Katmanlı Perceptron bulunmaktadır ve bu algoritmalar diğer makine öğrenmesi sınıflandırma algoritmaları arasında en yüksek doğruluğa ulaştığı saptanmıştır. Sonrasında ise iskemi ve enfeksiyonun daha etkili tespit edilebilmesi için Ensemble Convolutional Neural Network (CNN) modeli kullanılmıştır. Ensemble CNN modeli, birden fazla CNN modelindeki (Inception-V3, ResNet 50 ve Inception ResNetV2) özellikleri birleştirerek Destek Vektör Makineleri sınıflandırıcısını kullanmaktadır. Bununla birlikte ayak görüntülerinde ülserli bölgeyi tanımlayan ve bu alanda mevcut göze çarpan özellikleri bulmaya odaklanan bir veri büyütme yöntemi de kullanılmıştır. Nihai olarak, önerilen tekniklerin ikili sınıflandırma (iskemi ve enfeksiyon) performansı değerlendirilmiştir. Hem enfeksiyon hem de iskemi veri setleri %70 eğitim, %10 validasyon ve %20 test setlerine ayrılmış ve 5 kat çapraz geçerleme tekniğini kullanılmıştır. Performans değerlendirme ölçütleri olarak doğruluk, duyarlılık, kesinlik, özgüllük, F-Ölçütü ve Matthew Korelasyon Katsayısı ve ROC eğrisi (AUC) altında kalan alan kullanılmıştır. Sonuç olarak, önerilen yöntem ile iskemi sınıflamasında enfeksiyondan daha iyi performans göstermiştir. Önerilen Ensemble CNN derin öğrenme algoritması, diğer makine öğrenme algoritmalarına kıyasla her iki sınıflandırma görevi için daha iyi performans göstermiş, iskemi sınıflamasında %90 doğruluk ve enfeksiyon sınıflamasında %73 doğruluk saptanmıştır. Daha dengeli bir veri kümesi ile DFU'nun yapay zekâ teknikleri ile analizinin, kullanılan yöntemlerin performansının gelecekte geliştirilebileceği ifade edilmiştir. Ayrıca hem derin öğrenme hem de geleneksel makine öğrenme yöntemlerinin optimizasyon teknikleri birleştirilmesi algoritmaların performansını da artırabileceğinin altı çizilmiştir.

Henig ve diğ. (2020) çoklu ilaca dirençli organizmaların (MDRO'lar), diyabetik ayak enfeksiyonlarının (DFI) önemli patojenleri olduğunu belirtmişler ve MDRO patojenleri ile ilişkili diyabetik ayak enfeksiyonlarının klinik sonuçlar üzerindeki etkisini değerlendirmek amacıyla gerçekleştirdikleri araştırmada, 648 hastadan oluşan belirlenen tarih aralığındaki gerçekleştirdikleri kohorta, diyabetik ayak enfeksiyonu nedeni ile tıp merkezine kabul edilen ve kültür sonucu pozitif hastaları dahil etmişlerdir. Çalışmada belirlenen muhtemel sonlanım, diyabetik ayak enfeksiyonlarından sonraki bir yıl içinde diz üstü ve diz altı amputasyonu, yeniden hastaneye yatışlar ve mortalite olarak belirlenmiştir. Sonlanım ile DFI-MDRO arasındaki ilişkiler (metisiline dirençli *Staphylococcus aureus* [MRSA], vankomisine dirençli enterokok, üçüncü nesil sefalosporine dirençli *Enterobacteriaceae* [3GCR-EC],

Acinetobacterumannii ve *Pseudomonas aeruginosa*'yı içeren tek bir grup olarak değerlendirilmiş) analiz edilmiştir. Araştırmada, DFI-MDRO'ya sahip olma olasılığını tahmin eden Lojistik Regresyon modeli, DFI-MDRO hastaları ile MDRO olmayan patojenlere sahip DFI hastaları karşılaştırılarak hesaplanmıştır. Yapılan analiz sonuçlarına göre, DFI-MDRO bulunmasının, 1 yıllık diz üstü, diz altı amputasyon riski ile veya yeniden hastaneye yatış riski ile ilişkili olarak tespit edilememiştir, ancak tekrarlayan diyabetik ayak infeksiyonu epizodları ile ilişkili olduğu saptanmıştır (OR: 2,1;%95 CI, 1,38–3,21). Araştırmada sonuç olarak, DFI-MDRO olmayan hastalar ile karşılaştırıldığında DFI-MDRO olan hasta grubunun, tekrarlayan diyabetik ayak infeksiyonları riskinin yaklaşık 2 kat artmasıyla ilişkilendirilmiştir.

Hüsers ve diğ. (2020) şeffaf karar kuralları ürettikleri, belirsizliği sezgisel olarak ölçtüğü ve önceki mevcut bilimsel bilgileri kabul ettiği için Bayes tahmin modelleri geliştirmek amacıyla yaptıkları araştırmada, 237 diyabetik ayak hastasını prospektif olarak takip ettikleri boylamsal çalışmada, PEDIS sınıflandırmasına göre sınıflandırılan ülserlerin, altı aylık herhangi bir amputasyon ve majör amputasyon olmak üzere iki tip amputasyon riskini hesaplamak için Lojistik Regresyon yöntemini kullanmışlardır. Araştırma sonucunda, perfüzyon, ülser boyutu ve derinliği risk faktörlerinin muhtemel sonlanım için öngörücü olduğu bulunurken, infeksiyon durumu ve nöropatinin teşhisinde kullanılan hissiyat değişkenlerinin öngörücü değeri bulunamamıştır. Önceki bilgileri kullanan majör amputasyon modelinin, önceki bilgileri kullanmayan muadil modelden daha iyi performans gösterdiği saptanmıştır (AUC 0,765'e karşı AUC 0,790, Cohen'in d 2,21). Aksine, genel amputasyonu tahmin eden modellerin birbirine benzer şekilde performans gösterdiği tespit edilmiştir (0,793'e karşı 0,790, Cohen'in d 0,22).

Lim ve diğ. (2020) diyabetli ve periferik arter hastalığı bulunan hastalarda uzuv kurtarmada, negatif basınçlı yara tedavisi (NPWT) kullanımının, yara bölgesini nemli tutarak, eksüdayı azaltarak, infeksiyonu kontrol ederek ve granülasyonu uyarak yara iyileşmesini iyileştirdiğini belirtmişler ve diyabetli ve periferik arter hastalığı bulunan hastalarda ev tipi NPWT kullanımının uzuv kurtarma ve amputasyondan korunma açısından sonuçlarını incelemek amacıyla yaptıkları araştırmada, yatan hastalarda periferik arter hastalığı tedavisi için revaskülarizasyon, debridmanı veya minör amputasyon operasyonundan sonra evde NPWT tedavisi ile taburcu edilen 118 hastayı retrospektif olarak değerlendirmişlerdir. Araştırmada, kategorik gruplar arasındaki karşılaştırma, ki-kare testi veya Fisher testi kullanılarak yapılmış olup, sürekli değişkenler için Student t testi kullanılmıştır. Ev tipi NPWT tedavisinin başarısızlığı ile ilişkili faktörleri tanımlamak için tek değişkenli analiz gerçekleştirilmiş olup burada p değeri <.1 olan değişkenler, ev tipi NPWT tedavisinin başarısızlığı ile ilişkili bağımsız

risk faktörlerini tanımlamak için binomial regresyon modeli kullanılarak çok değişkenli analize dahil edilmiştir. Ayrıca, ev tipi NPWT kullanımı sonucundaki tahmini iyileşme oranını ve sepsis kontrolü için farklı tedaviler sonucundaki öngörülen iyileşme oranlarını karşılaştırmak ve belirlemek için Kaplan-Meier sağkalım analizi gerçekleştirilmiştir. Araştırma sonucunda 1 yıllık genel sağkalım %89 olarak saptanmıştır. Ev tipi NPWT'nin başarısızlığını öngören risk faktörlerinin başlangıçtaki son dönem böbrek yetmezliği ve kangren öyküsü olarak tespit edilmiştir. Araştırmada sonuç olarak, ev tipi NPWT kullanımının, diyabetli ve periferik arter hastalığı bulunan hasta popülasyonunda uzuv kurtarma ve amputasyondan korunmada, yararlı bir yardımcı tedavi olduğu vurgusu yapılmıştır.

Rastogi ve diğ. (2020) ilk nöropatik diyabetik ayak ülserinden sonra, uzuv amputasyonu ve mortaliteyi değerlendirmek amacıyla yaptıkları araştırmada, nöropatik diyabetik ayak ülseri bulunan toplam 2880 hasta ve benzer sayıda diyabetik ayak ülseri bulunmayan ve diyabeti bulunan hastaları, yaş ve diyabet süresi açısından eşleştirerek, 14 yıldan fazla beş sevk merkezinde prospektif olarak değerlendirmişler ve sonlanım olarak ölüm değişkenini kullanarak, çeşitli diyabetik komorbiditeler ve amputasyon, mortalite öngörücüleri olarak değerlendirilmiştir. Verilerin normalliği Kolmogorov-Smirnov testi kullanılarak, iki kohort arasındaki farklılıklar kategorik değişkenler için ki-kare testi ve sürekli değişkenler için Mann-Whitney U testi ile analiz edilmiş olup, başvuru sırasındaki yaş, diyabet süresi, ülser süresi, nöropati, nefropati, eGFR, periferik arter hastalığı, amputasyon (majör veya minör), koroner arter hastalığı ve serebrovasküler hastalık dahil olmak üzere diyabetik ayak ülserinin muhtemel sonlanımını ve mortalite sonucunu etkileyebilecek değişkenler bağımsız değişkenler olarak aşamalı Lojistik Regresyon modeline dahil edilmiştir. Sağkalım analizi ise, tabakalı Kaplan-Meier eğrileri kullanılarak grafiksel olarak analiz edilmiş, bu eğrilerdeki farklılıklar Log rank testi kullanılarak hesaplanmıştır. Araştırma sonucunda, mortalite riskinin glisemik kontrolden bağımsız olduğu tespit edilirken [OR: 1,03 (0,80–1,32; p = 0,83)], diyabet süresinin 10 yıldan uzun süredir olması [OR: 1,31(1,02–1,70, p = 0,035)], nefropati [OR: 1,47 (1,04–2,09, p < 0,030)], minör amputasyon [OR : 1,85 (1,40–2,44; p < 0,001) veya majör amputasyon [OR: 2,96 (2,01–4,34, p < 0,001)] mortalite riskini tahmin ettiği saptanmıştır.

Schmidt ve diğ. (2020) osteomiyelit nedeni olan organizmalar ile cerrahi tedaviyi takiben 1 yıllık diyabetik ayak sonlanımı (tekrarlayan ülser, amputasyon ve ölüm) arasındaki ilişkiyi belirlemek için yaptıkları 223 hastadan oluşan gözlemsel ve kohorta dayalı araştırmada, hastaların mevcut demografik, klinik ve laboratuvar verilerini, tek değişkenli ve çok değişkenli analizler yapmak suretiyle incelemişlerdir. Araştırmada ayrıca, 1 yıllık muhtemel sonlanımın,

mikrobiyolojik öngörücülerini tanımlamak amacıyla Rastgele Orman Algoritması kullanılmıştır. Cerrahi olarak tedavi edilen osteomiyeliteli hastalar, ayaktan tedavi gören diyabetli hastalara kıyasla daha ağırlı periferik nöropati semptomları sergilemiştir ($p < 0,0001$). Osteomiyeliti bulunan hastaların proksimal bölgelerinden alınan kültür sonuçları incelendiğinde, zaman zaman polimikrobiyal olmak üzere çeşitli olduğu saptanırken kohortun %13,9'unda çoklu ilaca dirençli organizmaları tespit edilmiştir. Hastaların proksimal bölgelerinden alınan kemik kültürlerinde, çoklu ilaca dirençli organizmaların varlığının, tek değişkenli modellemede ($p < 0,001$) ve rastgele orman algoritmasında Gini değerlerine göre hesaplanan önem sıralamasında diyabetik ayak ülseri nüksünün önemli bir öngörücüsü olduğu saptanmıştır. Sonuç olarak bu araştırmanın, çoklu ilaca dirençli organizmalar ile diyabetik ayak ülserlerinin nüksü arasındaki ilişkiyi belirlemek amacıyla oluşturulan tahmin modellerini kullanan ilk çalışma olduğu vurgulanmıştır.

Gazzaruso ve diğ. (2021) diyabetik ayak ülserlerinin muhtemel sonlanımının öngörücülerini belirlemek amacıyla yaptıkları araştırmada, multidisipliner bir ekip tarafından takip edilen 583 hastanın klinik ve demografik verilerini retrospektif olarak değerlendirmişlerdir. İkidenden fazla grup arasındaki farklılıkları bulmak için ANOVA veya uygun olduğunda Kruskal-Wallis testi kullanılmış olup, normal dağılım gösteren parametrelerde iki grup arasında anlamlı farklılıkları bulmak için Student t testi kullanılarak veya normal dağılım göstermeyen değişkenlerde ise Mann-Whitney U testi kullanılmıştır. Araştırmada kullanılan kategorik değişkenlerin karşılaştırmaları için Pearson ki-kare testi kullanılmış olup, tüm değişkenleri taramak için tek değişkenli analiz yapılmış olup analiz sonucundaki anlamlı değişkenler derecelendirilerek aşamalı bir yaklaşımla çoklu Lojistik Regresyon analizinde diyabetik ayak ülserlerinin muhtemel sonlanımının potansiyel öngörücülerini belirlemek amacıyla kullanılmıştır. Uyum iyiliği Hosmer-Lemeshow ki kare testi kullanılarak verilerin daha küçük rastgele alt örnekleri üzerinde değerlendirilmiştir. Araştırma sonucunda, bozulmuş böbrek fonksiyonu, diyabetik ayak ülserlerinin iyileşmemesi, amputasyon ve mortalite ile ilişkilendirilmiş olup, önceki kardiyovasküler hastalık yine diyabetik ayak ülserlerinin iyileşmemesi, ülserlerin nüksmesini ve mortaliteyi öngörmüştür. Daha düşük vücut kitle indeksi(BMI), diyabetik ayak ülserlerinin iyileşmemesini ve mortalite riski ile ilişkili bulunurken, osteomiyelit, amputasyon ve ölümün belirleyicisi olarak saptanmıştır. Periferik arter hastalığının belirteçleri, minör amputasyon ve diyabetik ayak ülserlerinin nüksünü öngörmüştür. Araştırmada sonuç olarak diyabetik ayak ülserlerine bütüncül ve konservatif bir yaklaşımın, hastalık prognozunu iyileştirebileceği vurgusu yapılmıştır.

3 MALZEME VE YÖNTEM

3.1 PROBLEM VE TEZ ÇALIŞMASININ AMACI

Bu tez çalışmasında esas olarak “diyabetik ayak hastalarının amputasyon riskinin yapay zekâ teknikleri ile tahmin edilmesi” problemi ele alınmıştır. Diyabetik ayak hastalarının amputasyon riskinin yapay zekâ teknikleriyle tahmin edilmesi amacıyla yapay zekânın alt disiplinlerinden biri olan - makine öğrenmesi sınıflandırma algoritmaları kullanılarak tahmin modeli/modelleri geliştirmek ve yapay zekâ temelli bir karar destek sistemi oluşturmak hedeflenmiştir.

Bu amaçla çalışma kapsamında Fayyad ve diğ. (1996) tarafından oluşturulan Veri Tabanlarında Bilgi Keşfi (VTBK) süreci kullanılmıştır. Bilgi keşfinin bu adımlara göre araştırma kapsamında gerçekleştirilen işlemlerin açıklamaları Bölüm 3.3, 3.4, 3.5 ve 3.6’da detaylı olarak açıklanmıştır (Fayyad ve diğ.1996; Han ve diğ. 2012).

3.2 DOKTORA TEZ ÇALIŞMASININ ÖZGÜNLÜĞÜ

Bu doktora tez çalışmasının veri setinin orijinal olması, algoritmaların çalışma prensibi ve amaca yönelik özgün veri setinin çeşitli alt kırılım senaryoları altında uygulanan oldukça geniş yelpazede makine öğrenmesi sınıflandırma algoritmalarının kullanılmış olması ve bununla birlikte hiperparametre optimizasyonu gerektiren makine öğrenmesi algoritmalarının ızgara arama yöntemi ve çeşitli meta sezgisel optimizasyon algoritmalarının kullanılarak parametre optimizasyonunun gerçekleştirilmesi başta olmak üzere bir çok açıdan literatüre önemli katkılar sağlayacak, orijinal bir çalışma olduğu düşünülmektedir.

Bu çalışma Türkiye’de bu kapsamda (ilgili problemin geniş yelpazede yapay zekâ makine öğrenmesi sınıflandırma algoritmaları ve parametre/hiperparametre optimizasyon teknikleri kullanılarak çözülmesi) yapılan ilk çalışma ve Dünya’da ise gerçekleştirilen sınırlı sayıdaki çalışmadan üçüncü çalışma (yapısal veri ile yapılmış ikinci çalışma) olma özelliğini taşımaktadır.

Araştırmada diyabetik ayak hastalarının ülserlerini kategorize etmek amacıyla uzman hekim/hekimler gözetiminde, dört farklı (Meggit Wagner- University of Texas, PEDİS ve S(AD) SAD) tıbbi sınıflandırma sisteminin kullanılmış olması da çalışmanın özgün yönlerindendir.

Dahası, R-Shiny karar destek sistemi bu probleme yönelik geliştirilmiş olan ilk karar destek sistemidir.

3.3 DİYABETİK AYAK VERİ SETİNİN OLUŞTURULMASI VE VERİNİN TOPLANMASI

Araştırma için veri toplamaya başlamadan önce, İstanbul Üniversitesi Tıp Fakültesi Etik Kurul Onayı (23.05.2019 – Sayı: 713) alınmıştır. Etik Kurul Onay Belgesi EK-1’de sunulmuştur.

- **Veri setinin Temini için Seçilen Kliniğin Özellikleri**

Tez araştırmasında elde edilen veriler, İstanbul Üniversitesi, İstanbul Tıp Fakültesi, Sualtı Hekimliği ve Hiperbarik Tıp Anabilim Dalı’daki multidisipliner bir ekip tarafından yönetilen, Türkiye’nin en eski ve en büyük diyabetik ayak kliniğinden temin edilmiştir. Diyabetik ayak hastaları genellikle uzun süre hastane yatış süreleri ile karakterizedir. Diyabetik Ayak Yarası ve İnfeksiyonunun Tanısı, Tedavisi ve Önlenmesi: Ulusal Uzlaş Raporu’na göre diyabetik ayak hastalarının taburcu edilmeden önce klinik olarak stabil olması, herhangi bir acil cerrahi müdahale gereksiniminin olmaması, glisemik kontrolünün sağlanmış olması, kullanılacak antibiyotiklerin planlanmış olması, ayağın yükten kurtarılması (“off-loading”) ve özgül yara bakımının uygulanması gerekir. Bu durumda taburcu edilen hastanın ayaktan izlenmesine de devam edilmektedir. Bu nedenle araştırmadaki 407 hasta verisinin edinilmesi zor, kıymetli bir hazine olduğu konunun uzmanlarınca ifade edilmiştir. Ayrıca kliniğe başvuran hastalar arasında, genellikle Türkiye’nin çeşitli illerinden gelen, önceki tedavilerde iyileşememiş hastalar da bulunmaktadır. Klinikte her perşembe günü çeşitli uzmanların bir araya gelip, diyabetik ayak hastalarının tedavi sürecine ilişkin görüş beyan ettiği **Kronik Yara Konseyi yapılmaktadır**. Bu kliniğin, Türkiye’de ve Avrupa’daki en ideal yara bakım kliniklerinden biri olduğu bu konuda uzman çeşitli hekimler tarafından ifade edilmektedir. Dolayısı ile bu araştırmada elde edilen verilerden ortaya çıkan faydalı modelin ideal/ideale en yakın bir model önerisi olduğu açıktır.

Bu çalışmada İstanbul Üniversitesi, İstanbul Tıp Fakültesi, Sualtı Hekimliği ve Hiperbarik Tıp Anabilimdalı’nda diyabetik ayak enfeksiyonu tanısıyla Ocak 2009 ile Eylül 2019 tarihleri arasında tedavi görmüş ve tedavi süreci sonlanmış hastalar retrospektif olarak değerlendirilmiş, hastaların başvuru anındaki demografik ve klinik özellikleri ile takip ve tedavi sonucu sağlanan fayda ve amputasyon ilişkisini belirlemek amaçlanmıştır. **Çalışmada, 9824 diyabetik ayak**

hastasından sadece takip sonucu yani sonlanımı bilinen 407 diyabetik ayak hastası uzman hekim gözetiminde seçilmiş ve çalışmaya dahil edilmiştir.

3.3.1 Orijinal Veri Seti

Veri toplama sürecinde, iki Sualtı Hekimliği ve Hiperbarik Tıp Uzmanı ve bir Enfeksiyon Hastalıkları ve Klinik Mikrobiyoloji Uzmanı olmak üzere üç hekimin görüşleri alınarak ve literatür taraması yapılarak oluşturulmuş formdan yararlanılmıştır. Kullanılan formun örneği EK-2’de sunulmuştur. Veri toplama sürecinde öncelikle bu formlara kaydedilen bilgiler daha sonra dijital ortama aktarılmıştır. Diyabetik ayak hastalarının amputasyon riskini öngörmek için veri setinde kullanılmış nitelikler Tablo 3.1, Tablo 3.2 ve Tablo 3.3’te sunulmuştur.

Tablo 3. 1: Orijinal veri setinde bulunan nitelikler

CİNSİYET	BINARY: ERKEK: 1 KADIN: 2
YAŞ	NÜMERİK
DİYABET SÜRESİ	NÜMERİK
BOY	NÜMERİK
KİLO (AĞIRLIK)	NÜMERİK
VÜCUT KİTLE İNDEKSİ (VKİ)¹	NÜMERİK
HİPERTANSİYON	BINARY: VAR:1 ; YOK: 0
SİGARA ÖYKÜSÜ	BINARY: VAR:1 ; YOK: 0
SİGARA (PAKET/yıl)	NÜMERİK
DİYABETİK AYAK İNFEKSİYONU (DFU) SÜRESİ (GÜN)	NÜMERİK
İNSÜLİN	BINARY: VAR:1 YOK:0
ORAL ANTİDİYABETİK AJAN	BINARY: VAR:1 YOK:0
TEKRARLAYAN DAE ÖYKÜSÜ	BINARY: VAR:1 YOK:0
GEÇMİŞTEKİ AMPÜTASYON ÖYKÜSÜ	BINARY: VAR:1 YOK:0
GEÇMİŞTEKİ AMPÜTASYON ÖYKÜSÜ VARSA, DÜZEYİ	KATEGORİK: MİNÖR AMPÜTASYON: 1, MAJÖR AMPÜTASYON:2, AMPÜTASYON YOK: 0
BÖBREK (RENAL) YETMEZLİK	BINARY: VAR:1 YOK:0
GEÇMİŞTE VASKÜLER CERRAHİ ÖYKÜSÜ	BINARY: VAR:1 YOK:0
DİYALİZ	BINARY: VAR:1 YOK:0
CHARCOT AYAK	BINARY:VAR:1 YOK:0
PEDİS SINIFLANDIRMA SİSTEMİNE GÖRE SINIFLANDIRILMIŞ- PERİFERİK ARTER HASTALIĞI (PAH) EVRESİ:	ORDINAL: EVRE 1: 1 ; EVRE 2: 2 ; EVRE 3: 3
YARA GENİŞLİĞİ (cm²)	NÜMERİK

¹ VKİ = Ağırlık (Kg) / boyun metre cinsinden karesi

Tablo 3. 2: Orijinal veri setinde bulunan nitelikler (devam)

PEDİS SINIFLANDIRMA SİSTEMİNE GÖRE SINIFLANDIRILMIŞ: YARA DERİNLİĞİ	ORDINAL: EVRE 1: 1 EVRE 2: 2 EVRE 3: 3
PEDİS SINIFLANDIRMA SİSTEMİNE GÖRE SINIFLANDIRILMIŞ: İNFEKSİYON	ORDINAL: EVRE 1: 1 ; EVRE 2: 2 ; EVRE 3: 3 ; EVRE 4: 4
OSTEOMİYELİT	BINARY: VAR:1 ; YOK: 0
LÖKOSİT (WBC) (10³/µl)	NÜMERİK
C-Reaktif Protein (CRP) - (mg/L)	NÜMERİK
Sedimentasyon (ESR) (mm/saat)	NÜMERİK
Hemoglobin(g/dL)	NÜMERİK
Bilirubin (mg/dL)	NÜMERİK
Albümin (g/dL)	NÜMERİK
Glikozillenmiş Hemoglobin (HbA1C) - (%)	NÜMERİK
Glukoz (mg/dL)	NÜMERİK
Nötrofil (PNL) (10³/µl)	NÜMERİK
Trombosit (PLT) (10³/µl)	NÜMERİK
Alanin aminotransferaz (ALT) (U/L)	NÜMERİK
Aspartat Aminotransferaz testi AST(U/L)	NÜMERİK
Kreatinin(mg/dL)	NÜMERİK
Total kolesterol(mg/dL)	NÜMERİK
Yüksek Yoğunluklu Lipoprotein (HDL) kolesterol(mg/dL)	NÜMERİK
Düşük Yoğunluklu Lipoprotein (LDL) kolesterol(mg/dL)	NÜMERİK
Trigliserit(mg/dL)	NÜMERİK
RETİNOPATİ	BINARY: VAR:1 YOK:0
KORONER ARTER HASTALIĞI (KAH)	BINARY: VAR:1 YOK:0
ASTİM/KOAH	BINARY: VAR:1 YOK:0
ACE İNHİBİTÖRÜ KULLANMA DURUMU	BINARY: VAR:1 YOK:0
STATİN KULLANMA DURUMU	BINARY: VAR:1 YOK:0
CANAGLİFOZİN KULLANMA DURUMU	BINARY: VAR:1 YOK:0
MEGGİT-WAGNER SINIFLANDIRILMA SİSTEMİ KULLANILARAK DERECELENDİRİLMİŞ ÜLSER	ORDINAL: GRADE 0 : 0 GRADE 1: 1 GRADE 2: 2 GRADE 3: 3 GRADE 4: 4 GRADE 5: 5
HİPERBARİK OKSİJEN TEDAVİSİ (HBOT) TEDAVİSİ DURUMU	BINARY: VAR:1 YOK:0
HBOT TEDAVİSİ /SEANS SAYISI	NÜMERİK

Tablo 3. 3: Orijinal veri setinde bulunan nitelikler (devam)

VASKÜLER GİRİŞİM (ANJİOPLASTİ, REVASKÜLARİZASYON, BYPASS) TEDAVİSİ	BINARY: VAR: 1 YOK:0
HASTANIN İYİLEŞME DURUMU	BINARY: VAR:1 YOK:0
HEDEF NİTELİK: HASTANIN GENEL AMPÜTASYON DURUMU (İKİLİ KATEGORİK, “BINARY” SINIFLANDIRMA PROBLEMİ	BINARY: AMPÜTASYON VAR: 1 AMPÜTASYON YOK: 0
HEDEF NİTELİK: HASTANIN MAJÖR VE MİNÖR AMPÜTASYON DURUMU (ÜÇLÜ KATEGORİK, “MULTI CLASS” SINIFLANDIRMA PROBLEMİ	KATEGORİK: MİNÖR AMPÜTASYON VAR:1 MAJÖR AMPÜTASYON VAR: 2 AMPÜTASYON YOK: 0

Oluşturulan bu form ile belirtilen nitelik alanları retrospektif olarak kayıt altına alınarak veri toplanmıştır. Veri toplama sürecinde hastane bilgi sistemi üzerinden hastaların laboratuvar sonuçları kayıt altına alınmış, hasta kartları ve hasta dosyaları taranmıştır. **Toplamda 18 ay süren ve gece nöbetlerinde de çalışmak suretiyle devam eden veri toplama sürecinden sonra toplam 407 hastanın verisi elde edilmiştir.**

Diyabetik ayak hastalarının, hastaneye başvuru anındaki yaralarının sınıflandırılması ise çalışmadaki en zahmetli süreci oluşturmuştur. Hastaların yaralarının sınıflandırılma işlemi iki sağlık uzmanı (diyabetik ayak konusunda uzman danışman uzman hekimler) tarafından gerçekleştirilmiştir. Sınıflandırma konusunda anlaşmazlık olduğunda, nihai karar daha kıdemli doktor tarafından verilmiştir. Toplamda 407 diyabetik ayak hastasının yaralarının sınıflandırılması sürecinde Meggit-Wagner, University of Texas, PEDİS ve SAD sınıflandırma sistemlerinin tümü kullanılmıştır. Her bir hastanın yarasının kategorizasyonu bu şekilde gerçekleştirilmiştir. Yapay Zekâ- makine öğrenmesi modeli geliştirilmek amacıyla Meggit Wagner ve PEDİS sınıflandırma sistemi esas alınmıştır. University of Texas ve SAD sınıflandırma sistemlerinin fazla sayıda kategoriden oluşması ve veri setinin boyutsallığını arttırması nedeniyle tahmin modeli oluşturulma sürecine dahil edilmemiştir.

3.4 VERİ ÖN İŞLEME

Veri setinin veri ön işleme aşamasında, ilgili yapay zekâ-makine öğrenmesi analizlerinin gerçekleştirilmesi ve tahmin modellerinin oluşturulması süreci, R programlama dili ve RStudio editörü kullanılmak suretiyle gerçekleştirilmiştir (R Development Core Team, 2021; R Studio Team, 2020).

Oluşturulan orijinal veri seti, yapay zekâ-makine öğrenmesi sınıflandırma algoritmaları kullanmak ve analiz için gözden geçirilmiştir. Bu kapsamda, bazı nitelik alanları veri setinden çıkarılmış, bazı nitelik alanları ise uzman görüşü alınarak birleştirilmiştir. Bu kapsamda “Form No, Metformin, Gliptin/Glp Agonist, Glinid, Alfa Glukozidaz, Glitazon, Sulfonilüre, Geçmişte Başparmak Ampütasyonu, Geçmişte Diğer Parmakların Ampütasyonu, Geçmişte Metatars Ampütasyonu, Geçmişte Bilek Ampütasyonu, Geçmişte Diz Altı Ampütasyonu, Geçmişte Diz Üstü Ampütasyonu, Hastaların Kullandıkları Antibiyotiklere İlişkin Bilgiler, Antibiyotik Grupları, Hastaların Hastanede Yatış Süreleri, Hastaların Kliniğe Başvuruları Sırasında Geçirdikleri Ampütasyon Operasyonunun Düzeyleri (Başparmak, Diğer Parmaklar, Metatars, Bilek, Dizaltı ve Dizüstü, Hastaların University of Texas ve SAD Sınıflandırma Sistemlerine Göre Sınıflandırılmış Ülser Dereceleri, Hastaların Büyüme Faktörü Tedavisi Alıp Almama Durumlarına İlişkin Sütun, Hastaların Cerrahi Girişim Geçirip Geçirmeme Durumlarına İlişkin Sütun, Hastaların By-Pass ve Revaskülarizasyon Geçirip Geçirmeme Durumları (*Birleştirilerek Vasküler Girişim – By-Pass, Anjioplasti Sütunu Eklenmiştir*) ve Exitus” sütunları silinmiştir. Bu sütunlara ilişkin bilgilerin istatistiki analiz olarak paylaşılması kararı alınmıştır. Veri setinin orijinal hâlinin, kategorik veriler ile birlikte, nümerik değer alan nitelik alanlarının minimum, maksimum, ortalama ve medyan değerlerinden oluşan özete Şekil 3.1 ve Şekil 3.2 ve Şekil 3.3’te yer verilmiştir.

YAS	CINSİYET	DM SÜRESİ/ YIL	BOY
Min. :32.00	1:286	Min. : 0.00	Min. :150.0
1st Qu.:54.00	2:121	1st Qu.:10.00	1st Qu.:165.0
Median :60.00		Median :15.00	Median :170.0
Mean :60.53		Mean :15.63	Mean :169.5
3rd Qu.:67.00		3rd Qu.:20.00	3rd Qu.:175.0
Max. :92.00		Max. :50.00	Max. :190.0
		NA's :8	NA's :49
KİLO	HİPERTANSİYON	SİGARA	SİGARA PAKET /YIL
Min. : 40.00	0:125	0 :250	Min. : 0.00
1st Qu.: 70.00	1:282	1 :153	1st Qu.: 0.00
Median : 80.00		NA's: 4	Median : 0.00
Mean : 81.42			Mean : 14.63
3rd Qu.: 90.00			3rd Qu.: 20.00
Max. :140.00			Max. :150.00
NA's :49			NA's :34
DAE SÜRESİ /GÜN	İNSÜLİN	ORAL A.D.	TEKRARLAYAN DAE
365 : 25	0: 61	0:226	0:166
120 : 15	1:346	1:181	1:241
60 : 13			
210 : 13			
150 : 11			
(Other):298			
NA's : 32			
GEÇMİŞTE AMP.MINÖR,MAJÖR	BÖBREK YETMEZLİĞİ		
0:250	0:330		
1:138	1: 77		
2: 19			
GEÇMİŞTE VASKÜLER CERRAHİ	DİYALİZ	CHARCOT	AYAK PAH EVRE
0:216	0:364	0:346	1:112
1:191	1: 43	1: 61	2:186
			3:109
YARA GENİŞLİĞİ(cm2)	YARA DERİNLİĞİ	ENFEKSİYON	OSTEOMİYELİT
Min. : 0.04	1: 87	1: 4	0:175
1st Qu.: 6.00	2:105	2: 78	1:232
Median : 12.00	3:215	3:260	
Mean : 31.99		4: 65	
3rd Qu.: 40.00			
Max. :225.00			
NA's :2			

Şekil 3. 1: Orijinal Diyabetik Ayak Veri Setinin R Proglama Dilinde Alınmış Özeti

LÖKOSİT	CRP	SEDİM
Min. : 3.70	Min. : 0.18	Min. : 2.00
1st Qu.: 7.75	1st Qu.: 10.00	1st Qu.: 46.00
Median :10.00	Median : 38.00	Median : 84.00
Mean :10.97	Mean : 72.15	Mean : 78.32
3rd Qu.:12.80	3rd Qu.:104.00	3rd Qu.:110.00
Max. :36.00	Max. :507.00	Max. :176.00
NA's :16	NA's :12	NA's :22
HEMOGLOBİN	BİLİRUBİN	ALBÜMİN
Min. : 5.30	Min. :0.0200	Min. :1.760
1st Qu.: 9.60	1st Qu.:0.2400	1st Qu.:3.290
Median :11.00	Median :0.3600	Median :3.790
Mean :11.09	Mean :0.4142	Mean :3.749
3rd Qu.:12.50	3rd Qu.:0.4935	3rd Qu.:4.300
Max. :16.80	Max. :1.7800	Max. :5.370
NA's :30	NA's :224	NA's :162
HbA1C	GLUKOZ	INR
Min. : 5.100	Min. : 48.0	Min. :0.810
1st Qu.: 7.200	1st Qu.:120.5	1st Qu.:0.975
Median : 8.700	Median :171.0	Median :1.050
Mean : 9.027	Mean :191.8	Mean :1.171
3rd Qu.:10.400	3rd Qu.:241.0	3rd Qu.:1.140
Max. :19.600	Max. :589.0	Max. :4.200
NA's :81	NA's :44	NA's :232
PNL	PLT	ALT
Min. : 2.600	Min. : 95.0	Min. : 3.00
1st Qu.: 4.775	1st Qu.: 238.0	1st Qu.: 11.00
Median : 6.300	Median : 304.0	Median : 15.00
Mean : 7.345	Mean : 322.9	Mean : 19.37
3rd Qu.: 8.525	3rd Qu.: 381.5	3rd Qu.: 23.00
Max. :33.800	Max. :1138.0	Max. :287.00
NA's :99	NA's :56	NA's :47
AST	KREATİNİN	TOTAL KOLESTEROL
Min. : 7.00	Min. : 0.300	Min. : 65.0
1st Qu.: 14.00	1st Qu.: 0.800	1st Qu.:137.5
Median : 16.40	Median : 1.000	Median :170.5
Mean : 20.37	Mean : 1.514	Mean :176.2
3rd Qu.: 22.00	3rd Qu.: 1.375	3rd Qu.:208.2
Max. :199.00	Max. :10.200	Max. :379.0
NA's :47	NA's :24	NA's :203

Şekil 3. 2: Orijinal Diyabetik Ayak Veri Setinin R Proglama Dilinde Alınmış Özeti (Devamı)

```

|  TRİGLİSERİD      LDL      HDL      NÖROPATİ
Min. : 50.0      Min. : 22.0      Min. : 8.00      0: 56
1st Qu.:106.3    1st Qu.: 79.0      1st Qu.: 25.00    1:351
Median :151.0    Median :103.0      Median : 33.00
Mean :178.0      Mean :105.9        Mean : 34.48
3rd Qu.:195.0    3rd Qu.:132.0      3rd Qu.: 41.00
Max. :897.0      Max. :228.0        Max. :115.00
NA's :212        NA's :208          NA's :214
GANGREN RETİNOPATİ KORONER ARTER HASTALIĞI ASTİM/KOAH
0:297  0:325      0:207      0:358
1:110  1: 82      1:200      1: 49

ACE İNH. KULLANIMI STATİN KULLANIMI CANAGLİFOZİNE KULLANIMI
0:319      0:337      0:400
1: 88      1: 70      1: 7

WAGNER GRADE HBO DURUMU/VAR YOK HBO TEDAVİ VAC TEDAVİ
1: 58      0:174      Min. : 0.00      0:212
2: 73      1:233      1st Qu.: 0.00    1:195
3:179      Median : 15.50
4: 88      Mean : 22.42
5: 9        3rd Qu.: 32.75
Max. :150.00
NA's :25

DEBRİDMAN VASKÜLER GİRİŞİM(BYPASS,ANJİOPLASTİ) İYİLEŞME
0:160      0:330      0: 57
1:247      1: 77      1:350

2'Lİ KARAR 3'LÜ KARAR(MİNÖR,MAJÖR,YOK)
0:228      0:228
1:179      1:148
              2: 31

```

Şekil 3. 3: Orijinal Diyabetik Ayak Veri Setinin R Proglama Dilinde Alınmış Özeti (Devamı)

Veri setinde bulunan toplam 407 hastanın ampütasyon statüleri incelendiğinde 179 hastanın ampütasyon operasyonu geçirdiği (majör ampütasyon + minör ampütasyon), 228 hastanın ise ampütasyon operasyonu geçirmediği görülmektedir.

Dolayısı ile veri setinin genel ampütasyonu tahmin etmek açısından dengeli bir veri seti olduğu söylenilebilir.

Bununla birlikte hedef nitelik olan “ÜÇLÜ KARAR” niteliği incelendiğinde; majör ampütasyon, minör ampütasyon ve ampütasyon olmaması açısından çoklu sınıflandırma problemi olarak ele alınması gerekliliği doğmaktadır ve bu durumda veri setinin dengesiz bir veri seti olduğu görülmektedir (Majör Ampütasyon: 31, Minör Ampütasyon: 148, Ampütasyon Yok: 228).

3.4.1 Orijinal Veri setinin Özellikleri

Orijinal veri seti hem nümerik hem de kategorik niteliklerden oluşmaktadır. Bununla birlikte bazı makine öğrenmesi sınıflandırma algoritmaları yalnızca nümerik niteliklerle çalışmaktadır. Bu nedenle her tipteki veri setine makine öğrenmesi sınıflandırma algoritmalarının uygulanması ve çeşitli kombinasyonları denemek için veri setinin yalnızca nümerik ve yalnızca

kategorik olmak üzere iki farklı alt veri seti daha oluşturulmuştur. Ayrıca nitelik seçimi – “*feature selection*” prosedürünün gerçekleştirilmesi sonucu toplamda nitelik sayısı yirmi iki (22) niteliğe düşmüş olup tüm analizler ayrıca nitelik seçimi işleminden sonra önem derecesi yüksek nitelikler için tekrarlanmıştır.

Uzman görüşü alınmak suretiyle retrospektif olarak elde edilen hasta veriler öncelikle forma kaydedilmiş olup, formdan da dijital ortama manuel olarak tek tek aktarılmış olduğundan veri setinde tekrar eden veri bulunmamaktadır.

- **EKSİK VERİNİN “*MISSING VALUE*” DOLDURULMASI**

Bununla birlikte veri setinde eksik niteliklere ait veri bulunmaktadır. Bu nedenle bu eksik veri R programlama dili kullanılarak tamamlanmıştır. Veri setinin paylaşılan özetinde görüleceği üzere ilgili nitelik alanları için boş olan hücrelerdeki veri hasta dosyaları, epikriz, arşiv veya hastane bilgi sistemi kayıtlarında bulunmamakta olduğundan ve retrospektif olarak veri seti oluşturulduğundan dolayı veri revize edilememekte, veriye ilişkin bilgi de teyit edilememektedir. Bu nedenle, ilgili sütunlarda boş olan hücreler, verinin olmadığı hücreler (*missing value*) olarak kabul edilmiştir.

Veri setindeki ilgili sütunlarda boş olan hücreler, Torgo (2011) tarafından oluşturulan “DMwR” paketi içerisindeki “knnImputation” fonksiyonu kullanılarak eksik verilerin tamamlanması yoluna gidilmiştir. Paket içerisinde bulunan “knnImputation” fonksiyonu her “NA” değerini, k adet en yakın komşularına bakarak varsayılan olarak k adet en yakın komşusunun ağırlıklı ortalamasını almak suretiyle doldurur. Bu fonksiyon kullanılırken varsayılan k değeri ($k=10$) kullanılmak suretiyle eksik değerler doldurulmuştur. Veri setindeki verinin olmadığı hücrelerin “knnImputation” fonksiyonu kullanılarak doldurulduktan sonraki özeti Şekil 3.4, ve Şekil 3.5’te paylaşılmıştır.

YAŞ	CİNSİYET	BMI	diyabet_suresi	HİPERTANSİYON	SİGARA	sigara_paketyıl
Min. :32.00	1:286	Min. :14.69	Min. : 0.00	0:125	0:254	Min. : 0.00
1st Qu.:54.00	2:121	1st Qu.:24.65	1st Qu.:10.00	1:282	1:153	1st Qu.: 0.00
Median :60.00		Median :27.68	Median :15.00			Median : 0.00
Mean :60.53		Mean :28.44	Mean :15.65			Mean : 14.96
3rd Qu.:67.00		3rd Qu.:31.14	3rd Qu.:20.00			3rd Qu.: 25.00
Max. :92.00		Max. :51.95	Max. :50.00			Max. :150.00
daesuresi_gun	İNSÜLİN	oral_antidiyabetik	tekrarlayan_dae	gecmisteki_amputasyon	böbrek_yetmezligi	
Min. : 0.0	0: 61	0:226	0:166	0:250	0:330	
1st Qu.: 90.5	1:346	1:181	1:241	1:138	1: 77	
Median : 155.0				2: 19		
Mean : 218.2						
3rd Qu.: 260.3						
Max. :2920.0						
gecmiste_vaskuler_cerrahi	DIYALİZ	charcot_ayak	pah_evre	yara_genisligi	yara_derinligi	ENFEKSİYON
0:216	0:364	0:346	1:112	Min. : 0.04	1: 87	1: 4
1:191	1: 43	1: 61	2:186	1st Qu.: 6.00	2:105	2: 78
			3:109	Median : 12.00	3:215	3:260
				Mean : 32.01		4: 65
				3rd Qu.: 41.00		
				Max. :225.00		
OSTEOMİYELİT	LÖKOSİT	CRP	SEDİM	HEMOGLOBİN	BİLİRUBİN	
0:175	Min. : 3.7	Min. : 0.18	Min. : 2.00	Min. : 5.30	Min. :0.0200	
1:232	1st Qu.: 7.8	1st Qu.: 10.20	1st Qu.: 47.00	1st Qu.: 9.70	1st Qu.:0.3127	
	Median :10.0	Median : 40.00	Median : 83.00	Median :11.00	Median :0.3717	
	Mean :11.0	Mean : 72.69	Mean : 78.14	Mean :11.09	Mean :0.4016	
	3rd Qu.:12.8	3rd Qu.:105.00	3rd Qu.:109.00	3rd Qu.:12.40	3rd Qu.:0.4481	
	Max. :36.0	Max. :507.00	Max. :176.00	Max. :16.80	Max. :1.7800	

Şekil 3. 4: “knnimputation” fonksiyonu kullanılarak eksik verilerin doldurulmasından sonra veri setinin R dilindeki özeti

ALBÜMİN	HbA1c	GLUKOZ	INR	PNL	PLT	
Min. :1.760	Min. : 5.100	Min. : 48.0	Min. :0.8100	Min. : 2.600	Min. : 95.0	
1st Qu.:3.397	1st Qu.: 7.600	1st Qu.:125.0	1st Qu.:0.9893	1st Qu.: 5.200	1st Qu.: 248.5	
Median :3.753	Median : 8.700	Median :172.0	Median :1.0400	Median : 6.735	Median : 316.6	
Mean :3.729	Mean : 8.976	Mean :190.1	Mean :1.1322	Mean : 7.631	Mean : 325.7	
3rd Qu.:4.081	3rd Qu.:10.050	3rd Qu.:231.5	3rd Qu.:1.1655	3rd Qu.: 8.800	3rd Qu.: 382.0	
Max. :5.370	Max. :19.600	Max. :589.0	Max. :4.2000	Max. :33.800	Max. :1138.0	
ALT	AST	KREATİNİN	total_kolesterol	TRİGLİSERİD	LDL	
Min. : 3.00	Min. : 7.0	Min. : 0.300	Min. : 65.0	Min. : 50.0	Min. : 22.0	
1st Qu.: 11.93	1st Qu.: 14.0	1st Qu.: 0.800	1st Qu.:151.0	1st Qu.:137.7	1st Qu.: 86.0	
Median : 15.00	Median : 17.0	Median : 1.000	Median :174.0	Median :168.0	Median :100.3	
Mean : 18.96	Mean : 20.1	Mean : 1.509	Mean :175.4	Mean :175.5	Mean :102.0	
3rd Qu.: 21.00	3rd Qu.: 21.0	3rd Qu.: 1.400	3rd Qu.:196.1	3rd Qu.:190.6	3rd Qu.:113.9	
Max. :287.00	Max. :199.0	Max. :10.200	Max. :379.0	Max. :897.0	Max. :228.0	
HDL	NÖROPATİ	GANGREN	RETİNOPATİ	koroner_arter_hastaligi	astim_koah	ace_inh_kullanimi
Min. : 8.00	0: 56	0:297	0:325	0:207	0:358	0:319
1st Qu.: 30.49	1:351	1:110	1: 82	1:200	1: 49	1: 88
Median : 35.00						
Mean : 35.01						
3rd Qu.: 39.39						
Max. :115.00						
statin_kullanimi	canaglifozine_kullanimi	wagner_grade	hbo_durumu	hbo_seans	vac_durumu	DEBRİDMAN
0:337	0:400	1: 58	0:174	Min. : 0.00	0:212	0:160
1: 70	1: 7	2: 73	1:233	1st Qu.: 0.00	1:195	1:247
		3:179		Median : 20.00		
		4: 88		Mean : 23.17		
		5: 9		3rd Qu.: 34.50		
				Max. :150.00		
vasküler_girisim_durumu	İYİLEŞME	İKİLİ_KARAR				
0:330	0: 57	Amputasyon Yok:228				
1: 77	1:350	Amputasyon Var:179				

Şekil 3. 5 “knnimputation” fonksiyonu kullanılarak eksik verilerin doldurulmasından sonra veri setinin R dilindeki özeti (devam)

- **Veri Setindeki Aykırı Değerlerin İncelenmesi**

Veri seti incelendiğinde, aykırı değerlerin veri setinin doğasından kaynaklandığı tespit edilmiştir. Bununla birlikte, amputasyon operasyonu geçiren hastalar ile ilgili nitelik alanlarının aldığı değerlerin aykırılık içerebileceği ve buna bağlı olarak da tahmin edilmeye çalışılan amputasyon operasyonu geçiren hastaların aykırı değere sahip hastalar olabileceğinden ve veri setindeki hasta sayısının kısıtlı olması da göz önüne alındığında, aykırı değerlere sahip olan nitelik alanlarının veri setinden çıkarılmamasına karar verilmiştir.

- **Min-Max Normalizasyon Prosedürünün Gerçekleştirilmesi**

Veri setindeki tüm nümerik nitelikler min-max normalizasyon yöntemi ile normalize edilmiştir. Min-max normalizasyon prosedüründen sonra veri setinin R dilindeki özeti Şekil 3.6 , Şekil 3.7 ve Şekil 3.8’de gösterilmiştir.

YAŞ	CİNSİYET	BMI	diyabet_suresi	HİPERTANSİYON	SİGARA	sigara_paketyıl
Min. :0.0000	1:286	Min. :0.0000	Min. :0.0000	0:125	0:254	Min. :0.00000
1st Qu.:0.3667	2:121	1st Qu.:0.2674	1st Qu.:0.2000	1:282	1:153	1st Qu.:0.00000
Median :0.4667		Median :0.3486	Median :0.3000			Median :0.00000
Mean :0.4756		Mean :0.3691	Mean :0.3129			Mean :0.09976
3rd Qu.:0.5833		3rd Qu.:0.4415	3rd Qu.:0.4000			3rd Qu.:0.16667
Max. :1.0000		Max. :1.0000	Max. :1.0000			Max. :1.00000
daesuresi_gun	İNSÜLİN	oral_antidiyabetik	tekrarlayan_dae	gecmisteki_amputasyon	böbrek_yetmezligi	
Min. :0.00000	0: 61	0:226	0:166	0:250	0:330	
1st Qu.:0.03099	1:346	1:181	1:241	1:138	1: 77	
Median :0.05308				2: 19		
Mean :0.07473						
3rd Qu.:0.08915						
Max. :1.00000						
gecmiste_vaskuler_cerrahi	DİYALİZ	charcot_ayak	pah_evre	yara_genisligi	yara_derinligi	ENFEKSİYON
0:216		0:364	1:112	Min. :0.00000	1: 87	1: 4
1:191		1: 43	2:186	1st Qu.:0.02649	2:105	2: 78
			3:109	Median :0.05316	3:215	3:260
				Mean :0.14213		4: 65
				3rd Qu.:0.18208		
				Max. :1.00000		
OSTEOMİYELİT	LÖKOSİT	CRP	SEDİM	HEMOGLOBİN	BİLİRUBİN	
0:175	Min. :0.0000	Min. :0.00000	Min. :0.0000	Min. :0.0000	Min. :0.0000	
1:232	1st Qu.:0.1269	1st Qu.:0.01977	1st Qu.:0.2586	1st Qu.:0.3826	1st Qu.:0.1663	
	Median :0.1950	Median :0.07857	Median :0.4655	Median :0.4957	Median :0.1998	
	Mean :0.2261	Mean :0.14306	Mean :0.4376	Mean :0.5035	Mean :0.2168	
	3rd Qu.:0.2817	3rd Qu.:0.20682	3rd Qu.:0.6149	3rd Qu.:0.6174	3rd Qu.:0.2432	
	Max. :1.0000	Max. :1.00000	Max. :1.0000	Max. :1.0000	Max. :1.0000	

Şekil 3.6: Tüm nümerik niteliklerin min-max normalizasyon yöntemi ile normalizasyon işleminden sonra veri setinin R dilindeki özeti

ALBÜMİN		HbA1C		GLUKOZ		INR		PNL					
Min.	:0.0000	Min.	:0.0000	Min.	:0.0000	Min.	:0.00000	Min.	:0.00000				
1st Qu.:	:0.4534	1st Qu.:	:0.1724	1st Qu.:	:0.1423	1st Qu.:	:0.05290	1st Qu.:	:0.08333				
Median	:0.5522	Median	:0.2483	Median	:0.2292	Median	:0.06785	Median	:0.13252				
Mean	:0.5454	Mean	:0.2673	Mean	:0.2626	Mean	:0.09504	Mean	:0.16123				
3rd Qu.:	:0.6430	3rd Qu.:	:0.3414	3rd Qu.:	:0.3392	3rd Qu.:	:0.10486	3rd Qu.:	:0.19872				
Max.	:1.0000	Max.	:1.0000	Max.	:1.0000	Max.	:1.00000	Max.	:1.00000				
PLT		ALT		AST		KREATİNİN		total_kolesterol					
Min.	:0.0000	Min.	:0.00000	Min.	:0.00000	Min.	:0.00000	Min.	:0.0000				
1st Qu.:	:0.1472	1st Qu.:	:0.03143	1st Qu.:	:0.03646	1st Qu.:	:0.05051	1st Qu.:	:0.2739				
Median	:0.2124	Median	:0.04225	Median	:0.05208	Median	:0.07071	Median	:0.3471				
Mean	:0.2212	Mean	:0.05621	Mean	:0.06820	Mean	:0.12210	Mean	:0.3518				
3rd Qu.:	:0.2752	3rd Qu.:	:0.06338	3rd Qu.:	:0.07292	3rd Qu.:	:0.11111	3rd Qu.:	:0.4175				
Max.	:1.0000	Max.	:1.00000	Max.	:1.00000	Max.	:1.00000	Max.	:1.0000				
TRİGLİSERİD		LDL		HDL		NÖROPATİ		GANGREN		RETİNOPATİ		koroner_arter_hastaligi	
Min.	:0.0000	Min.	:0.0000	Min.	:0.0000	0:	56	0:	297	0:	325	0:	207
1st Qu.:	:0.1036	1st Qu.:	:0.3107	1st Qu.:	:0.2101	1:	351	1:	110	1:	82	1:	200
Median	:0.1393	Median	:0.3803	Median	:0.2523								
Mean	:0.1481	Mean	:0.3883	Mean	:0.2525								
3rd Qu.:	:0.1660	3rd Qu.:	:0.4462	3rd Qu.:	:0.2934								
Max.	:1.0000	Max.	:1.0000	Max.	:1.0000								
astim_koah		ace_inh_kullanimi		statin_kullanimi		canaglifozine_kullanimi		wagner_grade		hbo_durumu			
0:	358	0:	319	0:	337	0:	400	1:	58	0:	174		
1:	49	1:	88	1:	70	1:	7	2:	73	1:	233		
								3:	179				
								4:	88				
								5:	9				

Şekil 3.7: Tüm nümerik niteliklerin min-max normalizasyon yöntemi ile normalizasyon işleminden sonra veri setinin R dilindeki özeti (devam)

hbo_seans	vac_durumu	DEBRİDMAN	vasküler_girisim_durumu	İYİLEŞME	İKİLİ_KARAR
Min. :0.0000	0:212	0:160	0:330	0: 57	Amputasyon Yok:228
1st Qu.:0.0000	1:195	1:247	1: 77	1:350	Amputasyon Var:179
Median :0.1333					
Mean :0.1545					
3rd Qu.:0.2300					
Max. :1.0000					

Şekil 3.8: Tüm nümerik niteliklerin min-max normalizasyon yöntemi ile normalizasyon işleminden sonra veri setinin R dilindeki özeti (devam)

3.4.2 Veri Dönüştürme

3.4.2.1 Nümerik Veri Seti

Veri setindeki en önemli niteliklerin tespit edilmesi amacıyla gerçekleştirilen “Temel Bileşenler Analizi – *Principal Component Analysis*” nümerik tipteki nitelik alanları ile çalışmaktadır. Bununla birlikte, bazı yapay zekâ- makine öğrenmesi algoritmaları (yapay sinir ağları tabanlı algoritmalar, aşırı öğrenme makineleri ve destek vektör makineleri) de yalnızca nümerik tipteki nitelik alanları ile sonuç üretmektedir. Bu nedenle, orijinal hâli nümerik ve kategorik değerlerden oluşan veri setinin tamamen nümerik gösteriminin (alt kırılımının) oluşturulması gerekliliği doğmuştur. Bu nedenle, veri setinde bulunan kategorik nitelikler sahte veya kukla “*dummy*” değişkenler kullanılarak, veri setinde temsil edilmiştir.

YAŞ	CİNSİYET	BMI	diyabet_suresi	HİPERTANSİYON	SİGARA
Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000
1st Qu.:0.3667	1st Qu.:0.0000	1st Qu.:0.2674	1st Qu.:0.2000	1st Qu.:0.0000	1st Qu.:0.0000
Median :0.4667	Median :1.0000	Median :0.3474	Median :0.3000	Median :1.0000	Median :0.0000
Mean :0.4756	Mean :0.7027	Mean :0.3682	Mean :0.3129	Mean :0.6929	Mean :0.3782
3rd Qu.:0.5833	3rd Qu.:1.0000	3rd Qu.:0.4415	3rd Qu.:0.4000	3rd Qu.:1.0000	3rd Qu.:1.0000
Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000
sigara_paketyıl	daesuresi_gun	İNSÜLİN	oral_antidiyabetik	tekrarlayan_dae	
Min. :0.0000	Min. :0.00000	Min. :0.0000	Min. :0.0000	Min. :0.0000	
1st Qu.:0.0000	1st Qu.:0.03099	1st Qu.:1.0000	1st Qu.:0.0000	1st Qu.:0.0000	
Median :0.0000	Median :0.05308	Median :1.0000	Median :0.0000	Median :1.0000	
Mean :0.1001	Mean :0.07484	Mean :0.8501	Mean :0.4447	Mean :0.5921	
3rd Qu.:0.1560	3rd Qu.:0.09160	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	
Max. :1.0000	Max. :1.00000	Max. :1.0000	Max. :1.0000	Max. :1.0000	
böbrek_yetmezligi	gecmiste_vaskuler_cerrahi	DIYALİZ	charcot_ayak	yara_genisligi	
Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.00000	
1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.02649	
Median :0.0000	Median :0.0000	Median :0.0000	Median :0.0000	Median :0.05316	
Mean :0.1892	Mean :0.4693	Mean :0.1057	Mean :0.1499	Mean :0.14215	
3rd Qu.:0.0000	3rd Qu.:1.0000	3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:0.18208	
Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.00000	
OSTEOMİYELIT	LÖKOSİT	CRP	SEDİM	HEMOGLOBİN	BİLİRUBİN
Min. :0.00	Min. :0.0000	Min. :0.00000	Min. :0.0000	Min. :0.0000	Min. :0.0000
1st Qu.:0.00	1st Qu.:0.1269	1st Qu.:0.01977	1st Qu.:0.2586	1st Qu.:0.3826	1st Qu.:0.1731
Median :1.00	Median :0.1950	Median :0.07857	Median :0.4598	Median :0.4957	Median :0.2159
Mean :0.57	Mean :0.2243	Mean :0.14217	Mean :0.4352	Mean :0.5056	Mean :0.2252
3rd Qu.:1.00	3rd Qu.:0.2817	3rd Qu.:0.19498	3rd Qu.:0.6149	3rd Qu.:0.6174	3rd Qu.:0.2549
Max. :1.00	Max. :1.0000	Max. :1.00000	Max. :1.0000	Max. :1.0000	Max. :1.0000

Şekil 3.9: Tüm niteliklerin nümerik olduğu veri setinin R dilindeki özeti

ALBÜMİN	HbA1C	GLUKOZ	INR	PNL	
Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.00000	Min. :0.00000	
1st Qu.:0.4573	1st Qu.:0.1728	1st Qu.:0.1423	1st Qu.:0.05600	1st Qu.:0.08494	
Median :0.5567	Median :0.2483	Median :0.2237	Median :0.06554	Median :0.13367	
Mean :0.5474	Mean :0.2669	Mean :0.2602	Mean :0.09256	Mean :0.15970	
3rd Qu.:0.6410	3rd Qu.:0.3384	3rd Qu.:0.3364	3rd Qu.:0.09399	3rd Qu.:0.19977	
Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.00000	Max. :1.00000	
PLT	ALT	AST	KREATİNİN	total_kolestero1	
Min. :0.0000	Min. :0.00000	Min. :0.00000	Min. :0.00000	Min. :0.0000	
1st Qu.:0.1467	1st Qu.:0.03142	1st Qu.:0.03646	1st Qu.:0.05051	1st Qu.:0.2780	
Median :0.2100	Median :0.04225	Median :0.05208	Median :0.07071	Median :0.3399	
Mean :0.2201	Mean :0.05620	Mean :0.06809	Mean :0.12155	Mean :0.3475	
3rd Qu.:0.2745	3rd Qu.:0.06338	3rd Qu.:0.07292	3rd Qu.:0.11069	3rd Qu.:0.3993	
Max. :1.0000	Max. :1.00000	Max. :1.00000	Max. :1.00000	Max. :1.0000	
TRİGLİSERİD	LDL	HDL	NÖROPATİ	GANGREN	RETİNOPATİ
Min. :0.00000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000
1st Qu.:0.09961	1st Qu.:0.3156	1st Qu.:0.2015	1st Qu.:1.0000	1st Qu.:0.0000	1st Qu.:0.0000
Median :0.13216	Median :0.3761	Median :0.2414	Median :1.0000	Median :0.0000	Median :0.0000
Mean :0.14609	Mean :0.3895	Mean :0.2446	Mean :0.8624	Mean :0.2703	Mean :0.2015
3rd Qu.:0.16293	3rd Qu.:0.4475	3rd Qu.:0.2811	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:0.0000
Max. :1.00000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000
koroner_arter_hastaligi	astim_koah	ace_inh_kullanimi	statin_kullanimi	canaglifozine_kullanimi	
Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.000	Min. :0.0000	
1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.000	1st Qu.:0.0000	
Median :0.0000	Median :0.0000	Median :0.0000	Median :0.000	Median :0.0000	
Mean :0.4914	Mean :0.1204	Mean :0.2162	Mean :0.172	Mean :0.0172	
3rd Qu.:1.0000	3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:0.000	3rd Qu.:0.0000	
Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.000	Max. :1.0000	

Şekil 3.10: Tüm niteliklerin nümerik olduğu veri setinin R dilindeki özeti (devam)

hbo_durumu	hbo_seans	vac_durumu	DEBRİDMAN	vasküler_girisim_durumu	
Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	
1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	
Median :1.0000	Median :0.1267	Median :0.0000	Median :1.0000	Median :0.0000	
Mean :0.5725	Mean :0.1554	Mean :0.4791	Mean :0.6069	Mean :0.1892	
3rd Qu.:1.0000	3rd Qu.:0.2498	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:0.0000	
Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	
İYİLEŞME	İKİLİ_KARAR	gecmisteki_amputasyon_0	gecmisteki_amputasyon_1		
Min. :0.00	Min. :0.0000	Min. :0.0000	Min. :0.0000		
1st Qu.:1.00	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000		
Median :1.00	Median :0.0000	Median :1.0000	Median :0.0000		
Mean :0.86	Mean :0.4398	Mean :0.6143	Mean :0.3391		
3rd Qu.:1.00	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000		
Max. :1.00	Max. :1.0000	Max. :1.0000	Max. :1.0000		
gecmisteki_amputasyon_2	pah_evre_1	pah_evre_2	pah_evre_3	yara_derinligi_1	
Min. :0.00000	Min. :0.0000	Min. :0.000	Min. :0.0000	Min. :0.0000	
1st Qu.:0.00000	1st Qu.:0.0000	1st Qu.:0.000	1st Qu.:0.0000	1st Qu.:0.0000	
Median :0.00000	Median :0.0000	Median :0.000	Median :0.0000	Median :0.0000	
Mean :0.04668	Mean :0.2752	Mean :0.457	Mean :0.2678	Mean :0.2138	
3rd Qu.:0.00000	3rd Qu.:1.0000	3rd Qu.:1.000	3rd Qu.:1.0000	3rd Qu.:0.0000	
Max. :1.00000	Max. :1.0000	Max. :1.000	Max. :1.0000	Max. :1.0000	
yara_derinligi_2	yara_derinligi_3	ENFEKSİYON_1	ENFEKSİYON_2	ENFEKSİYON_3	
Min. :0.000	Min. :0.0000	Min. :0.000000	Min. :0.0000	Min. :0.0000	
1st Qu.:0.000	1st Qu.:0.0000	1st Qu.:0.000000	1st Qu.:0.0000	1st Qu.:0.0000	
Median :0.000	Median :1.0000	Median :0.000000	Median :0.0000	Median :1.0000	
Mean :0.258	Mean :0.5283	Mean :0.009828	Mean :0.1916	Mean :0.6388	
3rd Qu.:1.000	3rd Qu.:1.0000	3rd Qu.:0.000000	3rd Qu.:0.0000	3rd Qu.:1.0000	
Max. :1.000	Max. :1.0000	Max. :1.000000	Max. :1.0000	Max. :1.0000	
ENFEKSİYON_4	wagner_grade_1	wagner_grade_2	wagner_grade_3	wagner_grade_4	wagner_grade_5
Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.00000
1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.00000
Median :0.0000	Median :0.0000	Median :0.0000	Median :0.0000	Median :0.0000	Median :0.00000
Mean :0.1597	Mean :0.1425	Mean :0.1794	Mean :0.4398	Mean :0.2162	Mean :0.02211
3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:1.0000	3rd Qu.:0.0000	3rd Qu.:0.00000
Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.00000

Şekil 3. 11: Tüm niteliklerin nümerik olduğu veri setinin R dilindeki özeti (devam)

ENFEKSİYON_4	wagner_grade_1	wagner_grade_2	wagner_grade_3	wagner_grade_4	wagner_grade_5
Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.00000
1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.00000
Median :0.0000	Median :0.0000	Median :0.0000	Median :0.0000	Median :0.0000	Median :0.00000
Mean :0.1597	Mean :0.1425	Mean :0.1794	Mean :0.4398	Mean :0.2162	Mean :0.02211
3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:1.0000	3rd Qu.:0.0000	3rd Qu.:0.00000
Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.00000

Şekil 3. 12: Tüm niteliklerin nümerik olduğu veri setinin R dilindeki özeti (devam)

3.4.2.2 Temel Bileşenler Analizi Sonuçları

Veri ön işleminin önemli adımlarından biri olan nitelik seçimi – “*feature selection*” temel bileşenler analizi “*principal component analysis*”(PCA) gerçekleştirilmek suretiyle yerine getirilmiştir.

PCA, R programlama dilinde Kassambara ve Mundt (2020) tarafından oluşturulan “*factoextra*” ve Husson ve diğ. (2020) tarafından oluşturulan “*factoMineR*” paketleri kullanılarak gerçekleştirilmiştir.

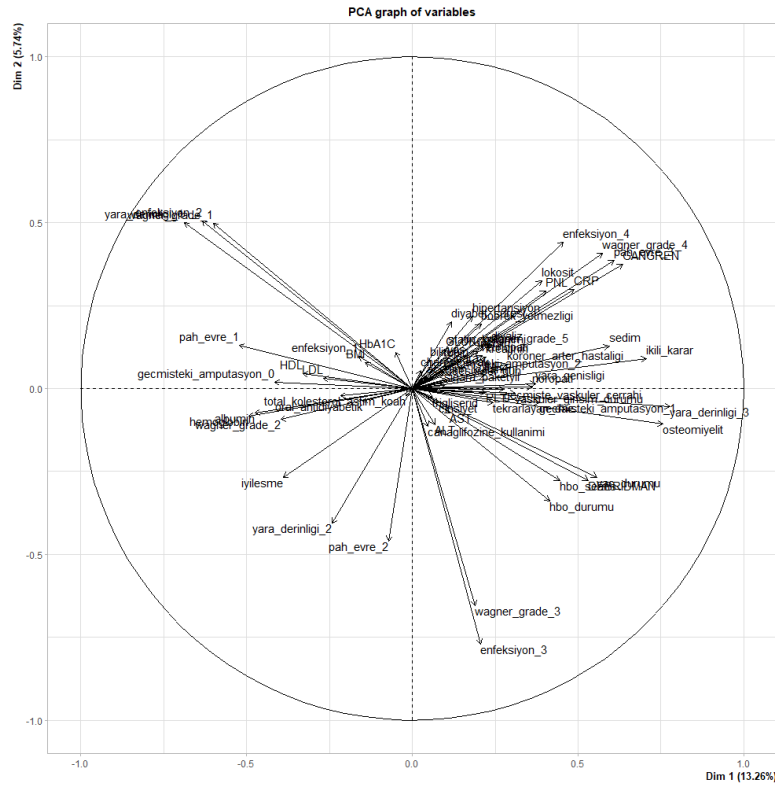
- Temsilin kalitesini tahmin etmek için $\cos^2$ değerleri kullanılmaktadır.

- Bir değişken korelasyon çemberine ne kadar yakınsa, faktör haritasında o kadar iyi temsil edilmektedir.

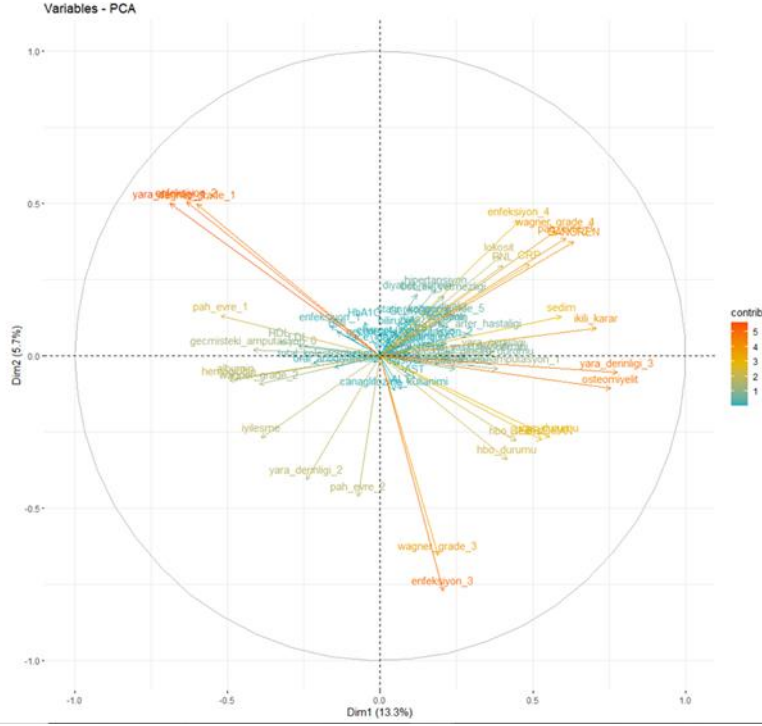
- Yüksek $\cos^2$ değeri, temel bileşen üzerindeki değişkenin iyi bir temsilini göstermektedir. Bu durumda değişken, korelasyon çemberinin çevresine yakın konumlandırılmaktadır. Düşük $\cos^2$ değeri ise, değişkenin temel bileşenler tarafından mükemmel bir şekilde temsil edilmediğini gösterir. Bu durumda değişken dairenin merkezine yakın konumlandırılmaktadır.

- Grafiğin merkezine yakın olan değişkenler diğer bileşenlere göre daha az önemli olarak kabul edilmektedir.

`fviz_pca_var()` fonksiyonu; PCA sonucunda oluşan korelasyon çemberinin grafiksel gösteriminin elde edilmesi için kullanılmıştır.

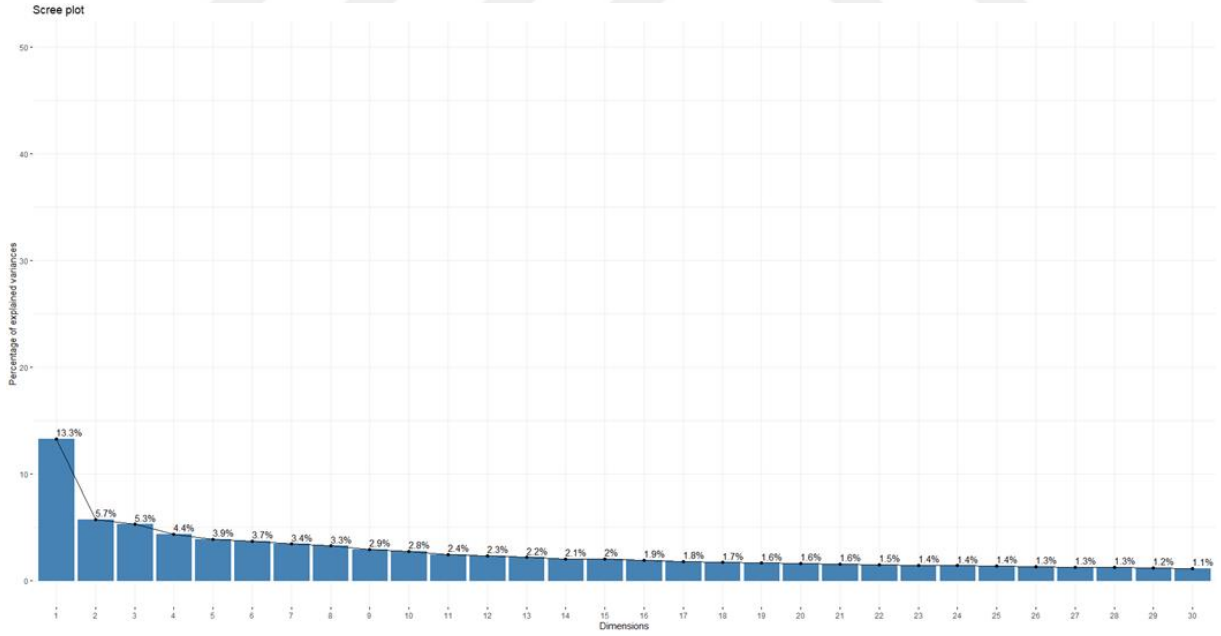


Şekil 3. 13: Değişkenlerin PCA sonucunda şekilsel gösterimi

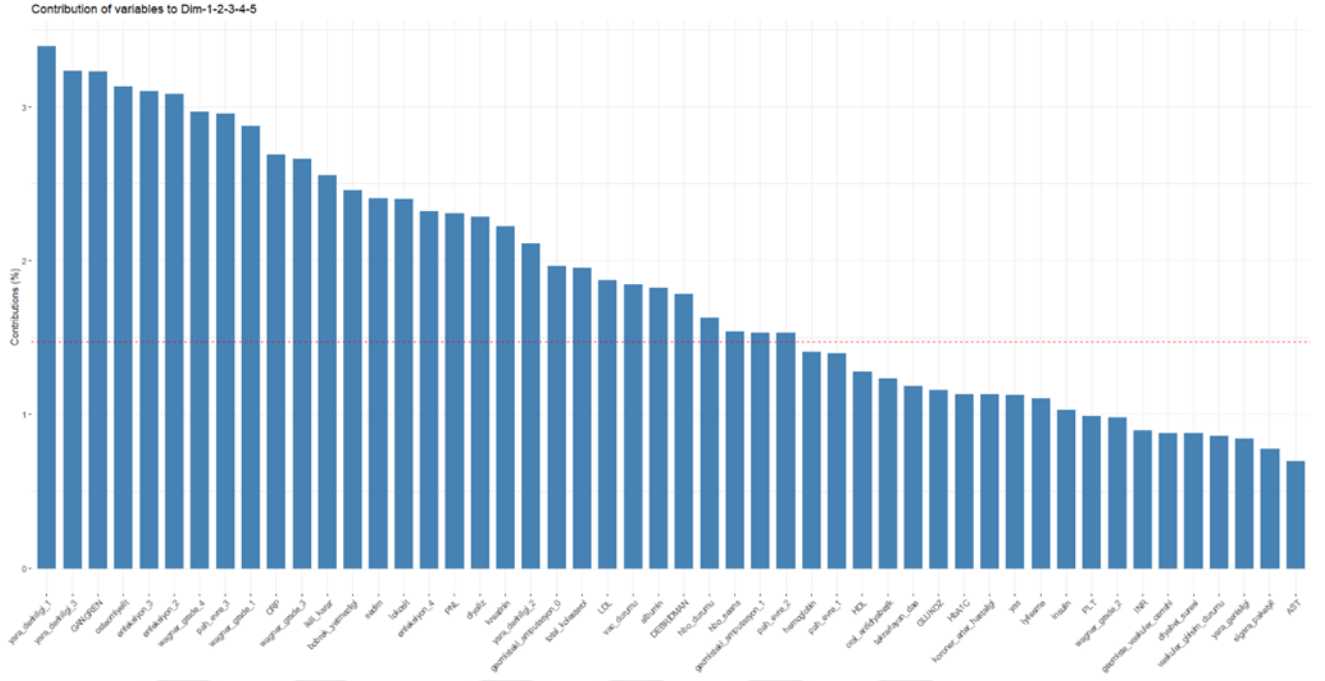


Şekil 3. 14: Değişkenlerin PCA sonucunda şekilsel gösterimi (renklendirilmiş versiyon)

“Scree plot” grafiği, *fviz_eig()* fonksiyonu kullanılarak oluşturulmuştur.



Şekil 3. 15: PCA sonucunda ortaya çıkan “scree-plot” grafiği

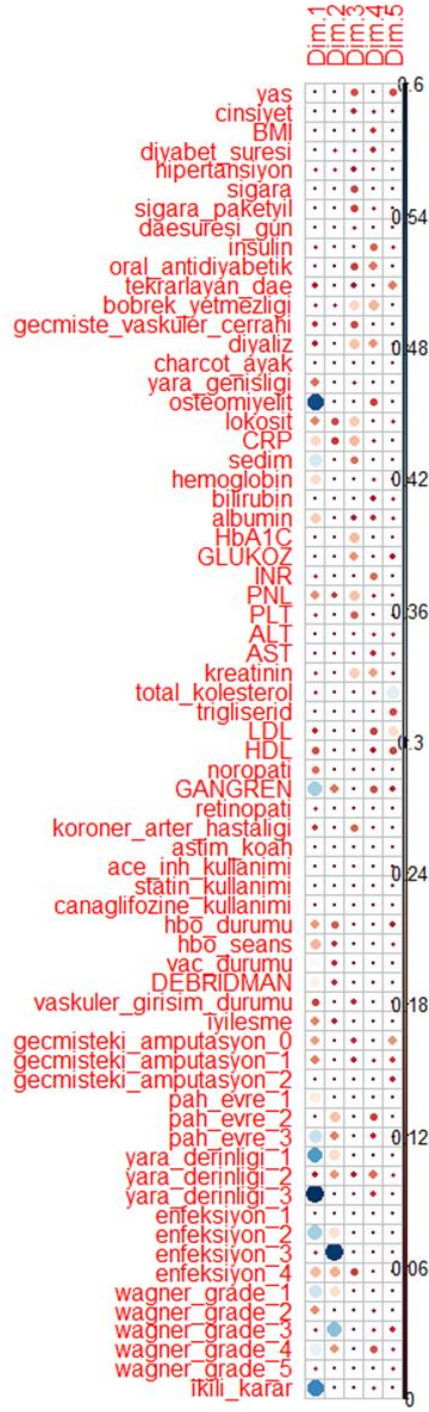


Şekil 3. 16: PCA sonucunda ortaya çıkan değişkenlerin ortalama katkısını gösteren grafik

Yukarıdaki kırmızı çizgi, beklenen ortalama katkıyı gösterir. Belirli bir bileşen için, bu kesimden - “*cut-off*” değerinden daha büyük bir katkısı olan bir değişken, bileşene katkıda bulunmada önemli olarak kabul edilmektedir.

Sonuç olarak nitelik seçimi için bu kesim değeri kullanılmış olup, bu çizginin üzerindeki değerler daha önemli nitelikler olarak kabul edilmiştir.

Tüm boyutlarda (toplamda 5 boyut bulunmaktadır) değişkenlerin $\cos^2$ değerlerinin görselleştirilmesi Wei ve Simko (2021) tarafından geliştirilen “*Corrplot*” paketi kullanarak gerçekleştirilmiştir.



Şekil 3. 17: Tüm boyutlarda değişkenler cos2 değerlerinin görselleştirilmesi

3.4.2.3 Kategorik Veri Seti

Doktora tez çalışması kapsamında orijinal veri setinin bir alt kırılımı da, nümerik tipteki niteliklerin kategorik değişkenlere dönüşmesini sağlayan ayrıklaştırma – “discretization” prosedürünün gerçekleştirilmesi sonucu oluşturulmuş kategorik veri setidir. Nümerik tipteki

değişkenlerin histogram grafiklerine bakılarak ve alanda uzman hekim görüşü ve danışman Öğretim Üyelerinin görüşü de alınarak oluşturulmuş kategorik veri setinin R dilinde alınmış özeti aşağıdaki şekillerde sunulmuştur (Şekil 3.18, Şekil 3.19, Şekil 3.20, Şekil 3.21).

```

cinsiyet      BMI      hipertansiyon  sigara  insulin
1:286   normal   :111    0:125      0:254    0: 61
2:121   obese    :126    1:282      1:153    1:346
          overweight :166
          underweight: 4

oral_antidiyabetik  tekrarlayan_dae  gecmisteki_amputasyon
0:226              0:166              0:250
1:181              1:241              1:138
                              2: 19

bobrek_yetmezligi  gecmiste_vaskuler_cerrahi  diyaliz
0:330              0:216              0:364
1: 77              1:191              1: 43

charcot_ayak  pah_evre  yara_derinligi  enfeksiyon  osteomiyelit
0:346         1:112    1: 87          1: 4         0:175
1: 61         2:186    2:105          2: 78          1:232
                  3:109    3:215          3:260
                              4: 65

hemogloblin  bilirubin  albumin  INR
Low   :357   High   : 2   Low   : 60   High   : 43
Normal: 50   Low    : 18  Normal:347 Low    : 3
                  Normal:387
                              Normal:361

```

Şekil 3. 18: Kategorik veri setinin R dilinde alınmış özeti-1

```

      PLT      kreatinin      HDL      noropati GANGREN
High :107    High : 89    Low :310    0: 56    0:297
Low  : 17    Low  : 48    Normal: 97  1:351    1:110
Normal:283   Normal:270

retinopati koroner_arter_hastaligi astim_koah
0:325      0:207      0:358
1: 82      1:200      1: 49

ace_inh_kullanimi statin_kullanimi canaglifozine_kullanimi
0:319      0:337      0:400
1: 88      1: 70      1: 7

wagner_grade hbo_durumu vac_durumu debridman
1: 58      0:174      0:212      0:160
2: 73      1:233      1:195      1:247
3:179
4: 88
5: 9

vaskuler_girisim_durumu iyilesme kategorik_diyabet_suresi
0:330      0: 57      (-Inf,10]:136
1: 77      1:350      (10,20] :189
                        (20, Inf]: 82

```

Şekil 3. 19: Kategorik veri setinin R dilinde alınmış özeti-2 (devam)

```

kategorik_sigara_paket_yil kategorik_daesuresi_gun
(-Inf,0] :250      (-Inf,50] :43
(0,25] : 56      (50,100] :78
(25,50] : 69      (100,150] :77
(50, Inf]: 32      (150,200] :59
                        (200,300] :70
                        (300, Inf]:80

kategorik_yara_genisligi kategorik_hbo_seans kategorik_CRP
(-Inf,10] :172      (-Inf,0] :173      (0,5] : 63
(10,50] :162      (0,30] :119      (5,30] :117
(50,75] : 22      (30, Inf]:115      (30,70] : 79
(75,100] : 25      (70,Inf]:148
(100, Inf]: 26

      kategorik_WBC      kategorik_sedim      kategorik_GLUKOZ
(0,10.3] :219      (0,20] : 29      (0,100] : 50
(10.3,20]:171      (20,50] : 90      (100,120]: 41
(20,Inf] : 17      (50,100] :150      (120,200]:175
                        (100,Inf]:138      (200,300]: 96
                        (300,Inf]: 45

kategorik_HbA1C kategorik_total_kolesterol kategorik_LDL
(0,6.4] : 39      (0,130] : 47      (0,100] :202
(6.4,8] : 97      (130,150]: 50      (100,130]:153
(8,10] :169      (150,200]:223      (130,Inf]: 52
(10,Inf]:102      (200,Inf]: 87

      kategorik_YAS      kategorik_trigliserid      kategorik_ALT
(0,50] : 55      (0,150] :137      (-Inf,45]:390
(50,60] :156      (150,200]:195      (45, Inf]: 17
(60,70] :138      (200,Inf]: 75
(70,Inf]: 58

```

Şekil 3. 20: Kategorik veri setinin R dilinde alınmış özeti-3 (Devam)

kategorik_AST	kategorik_PNL	ikili_karar
(-Inf,42]:387	(-Inf,11]:353	0:228
(42, Inf]: 20	(11,15] : 39	1:179
	(15, Inf]: 15	

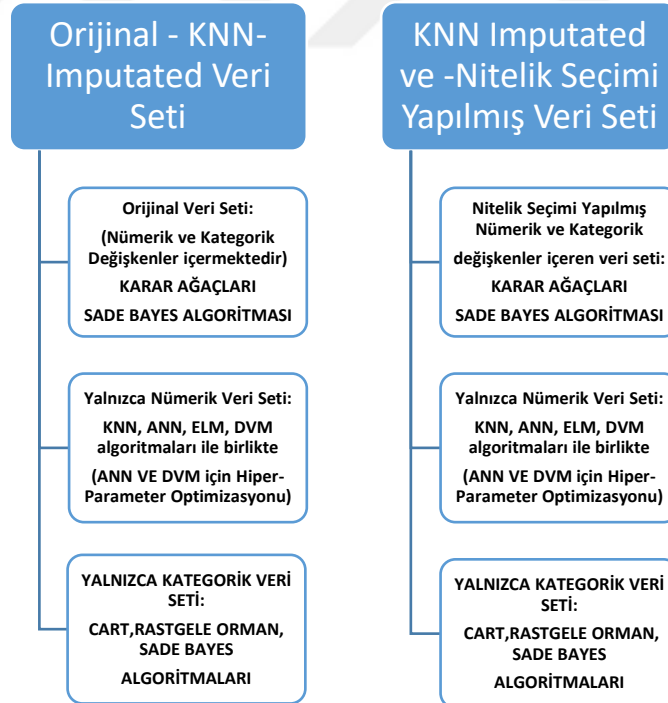
Şekil 3. 21: Kategorik veri setinin R dilinde alınmış özeti-4 (Devam)

3.5 VERİ MADENCİLİĞİ- YAPAY ZEKÂ – MAKİNE ÖĞRENMESİ MODELLERİNİN KURULMASI

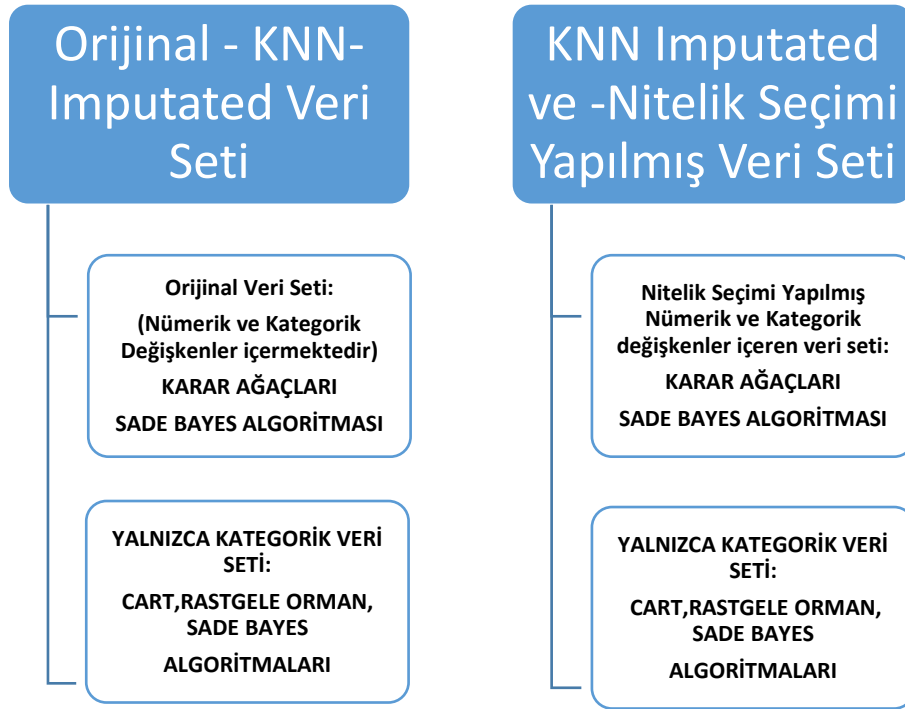
3.5.1 Orijinal Veri Setinin Alt Kırımları ile Birlikte Elde Edilen Alt Kırılım Veri Setleri ve Uygulanan Algoritmalar

Doktora tez çalışmasında literatürde danışmanlı öğrenme adı verilen ve tezin “GENEL BİLGİLER” kısmında detaylı olarak açıklanan makine öğrenmesi sınıflandırma algoritmalarından faydalanılmıştır.

Oluşturulan veri setleri ve kullanılan algoritmalar Şekil 3.22 ve Şekil 3.23’te özetlenmiştir.



Şekil 3. 22: Hedef niteliğin ikili “binary” olduğu durum için oluşturulan veri setleri ve kullanılan algoritmalar



Şekil 3. 23: Hedef niteliğin hastaların majör ve minör ampütasyon statülerini de içeren “multi-class” veri seti ve kullanılan algoritmalar

Şekil 3.22 ve Şekil 3.23’te özetlendiği ve açıkça görülebileceği gibi, bu tez çalışmasında hedef nitelik ampütasyon olup olmaması veya minör ampütasyon olması, majör ampütasyon olması ve ampütasyon olmaması olarak belirlenmiştir.

3.6 MODEL PERFORMANS DEĞERLENDİRME YÖNTEMLERİ & ÖLÇÜTLERİ

Bu tez çalışmasında diyabetik ayak hastalarının hem genel – “overall” hem de majör ve minör ampütasyon- “multi class” riskinin öngörülmesi tahmini için çeşitli modeller oluşturulmuştur. Bu modellerin kurulmasında, Bölüm 2.5 ve 2.6’da detayları verilen k-En Yakın Komşu algoritması, Sade (Naive) Bayes Sınıflandırıcı, Sınıflandırma ve Regresyon Ağaçları algoritması (CART), Rastgele Orman algoritması, Yapay Sinir Ağları, Aşırı Öğrenme Makineleri, Lojistik Regresyon, Destek Vektör Makineleri ve XGBoost algoritmalarından yararlanılmıştır. Hiperparametre optimizasyonu için ayrıca Genetik Algoritma ve Parçacık Sürü Optimizasyonu algoritmaları da kullanılmıştır.

Algoritmaların karşılaştırılması için Bölüm 2.7’de belirtilen Makine Öğrenmesi Model Performans Değerlendirme Yöntemleri ve Ölçütleri kullanılmıştır.

Bu tez çalışmasında sınıflandırma yaparken model performans değerlendirme ölçütleri hesaplanırken;

- Pozitif sınıf: Diyabetik ayak hastalarının amputasyon geçirme durumu
- Negatif sınıf: Diyabetik ayak hastalarının amputasyon geçirmeme durumu

biçiminde ele alındığından kontenjans tablosu oluşturulurken faydalanılan dp, yn, yp ve dn aşağıdaki gibidir:

- **dp:** Gerçekte amputasyon operasyonu geçiren diyabetik ayak hastalarının, algoritma/model tarafından da doğru tahmin edilmesi
- **yn:** Gerçekte amputasyon operasyonu geçiren diyabetik ayak hastalarının, algoritma/model tarafından amputasyon operasyonu geçirmemiş olarak yanlış tahmin edilmesi,
- **yp:** Gerçekte amputasyon operasyonu geçirmeyen diyabetik ayak hastalarının, algoritma/model tarafından amputasyon operasyonu geçirmiş olarak yanlış tahmin edilmesi,
- **dn:** Gerçekte amputasyon prosedürü geçirmeyen diyabetik ayak hastalarının, algoritma/model tarafından da amputasyon operasyonu geçirmemiş olarak doğru tahmin edilmesi,

Model performans değerlendirme yöntemleri olarak;

- 5,10,15 ve 20 kat Çapraz Geçerleme yöntemi
- İkili ayırma – “*Hold-out*” yöntemi, %90 eğitim %10 test, %80 eğitim %20 test, %70 eğitim %30 test ve %60 eğitim %40 test olarak uygulanmıştır.
- Bootstrap metodu 50, 100, 200 örneğin eğitim veri setinde kullanılması sureti ile gerçekleştirilmiştir.
- Üçlü ayırma metodu %80 eğitim, %10 validasyon ve %10 test ve tabakalı çapraz geçerleme yöntemi kullanılarak uygulanmıştır. Veri setinin %80’ini oluşturan eğitim veri setinde model kurulumu gerçekleştirilmiş olup, %10’luk validasyon veri setinde optimizasyon gerektiren algoritmalarda hiperparametre optimizasyonu gerçekleştirilmiş, kalan %10’luk test veri setinde ise modellerin daha önce hiç karşılaşmadığı ya da görmediği bu veri seti üzerinde gerçeğe en yakın sonuçların elde

edilmesi sağlanmıştır. Böylelikle “*data leakage*” ya da başka bir ifade ile veri sızması önlenmiştir. Dahası; hiperparametre optimizasyonu gereken algoritmalarda ızgara arama- “*grid search*”, genetik algoritma ve parçacık sürü optimizasyonu gerçekleştirilmek suretiyle hiperparametre optimizasyonu da gerçekleştirilmiştir.

3.7 KULLANILAN ARAÇLAR

Doktora tezinin bu kısmında çalışmada kullanılan araçlar tanıtılmıştır.

- Retrospektif olarak elde edilen hasta veriler öncelikle basılı formlara kaydedilmiş olup, formdan da Microsoft Excel 2016 kullanılmak suretiyle dijital ortama doktora tezinin sahibi tarafından manuel olarak tek tek aktarılmıştır.
- İstatistiksel veri analizinde istatistiksel programlardan olan SPSS v23.0 kullanılmıştır.
- Doktora tez çalışmasında gerçekleştirilen analizler, hem R programlama dili ile R Studio ortamında hem de çeşitli algoritmaların hiperparametre optimizasyonu Scikit-Learn (sklearn) ve Keras kütüphaneleri kullanılarak Genetik Algoritma için; Fortin ve diğ. (2012) tarafından geliştirilen “*DAEP*”, Parçacık Sürü Optimizasyonu için ise Claesen ve diğ.(2014b) tarafından geliştirilen “*Optunity*” paketi kullanılarak, Python programlama dili Anaconda Inc.(2020) kullanılmak suretiyle gerçekleştirilmiştir (Pedregosa ve diğ, 2011; Fortin ve diğ, 2012; Claesen ve diğ. 2014b; Chollet, 2015; Anaconda Inc, 2020; R Development Core Team, 2021; Rose, 2021).

Tablo 3. 4: Doktora tez çalışmasında R programlama dilinde kullanılan paket ve kütüphaneler

Kullanılan Paket	Kullanılan kütüphane	Kullanım Amacı
readxl (Wickham ve Bryan, 2019)	readxl	Veri seti/setlerinin okunması
plyr (Wickham, 2020)	plyr	Multi işlevli/ Veri ön işleme
DMwR2 (Torgo, 2016)	DMwR2	Veri ön işleme (knn imputation)
caret (Kuhn, 2022)	caret	Multi işlevli/ Veri ön işleme Modelleme
fastDummies (Kaplan, 2020)	fastDummies	Veri ön işleme Kukla Değişkenler Oluşturma
ROSE (Lunardon ve diğ, 2021)	rose	Veri ön işleme Dengesiz veri setinin dengeli hâle getirilmesi
factominer (Husson ve diğ. 2020)	factominer	Veri ön işleme

		Temel Bileşenler Analizinin Gerçekleştirilmesi
factoextra Kassambara ve Mundt (2020)	factoextra	Veri ön işleme Temel Bileşenler Analizinin Gerçekleştirilmesi
tunePareto (Müssel ve diğ., 2021)	tunePareto	Veri setinin çapraz geçerleme için analistin belirlediği sayıda parçalara ayrılması
splittols (Mayer, 2022)	splittols	Veri setinin parçalara ayrılması (örneğin: eğitim, validasyon ve test)
corrplot (Wei ve Simko, 2021)	corrplot	Korelasyon Tablosunun Oluşturulması
rpart (Therneau ve Atkinson, 2022)	rpart	Sınıflandırma ve Regresyon Ağaçları Algoritmalarının Uygulanması
randomforest (Liaw ve Wiener, 2022)	randomforest	Rastgele Orman Algoritmasının Uygulanması
stats glm-2(Marschner ve Donoghoe, 2018)	glm	Lojistik Regresyon Algoritmasının Uygulanması
ROCR (Sing ve diğ., 2020)	ROCR	Roc eğrilerinin çizdirilmesi
e1071 (Meyer ve diğ., 2021)	e1071	Sade Bayes Algoritmasının Uygulanması
class (Ripley ve Venables, 2022)	class	K-En Yakın Komşu Algoritmasının Uygulanması
nnet (Ripley ve Venables, 2022b)	nnet	Tek Katmanlı Yapay Sinir Ağır Modelinin Kurulması
kernlab (Karatzoglou ve diğ., 2019)	kernlab	Destek Vektör Makineleri Modelinin Kurulması
elmNNRcpp (Mouselimis ve Gosso, 2022)	elmNNRcpp	Aşırı Öğrenme Makineleri Modelinin Kurulması
xgboost (Chen ve diğ., 2022)	xgboost	"Extreme Gradient Boosting" Algoritmasının Uygulanması
shiny (Chang ve diğ., 2021a)	shiny	Makine öğrenmesi modellerinin R-Shiny kullanıcı arayüzüne entegre edilmesi ve yeni tahminlerin yapılmasının sağlanması

4 BULGULAR

Doktora tezinin bu bölümünde öncelikle veriyi anlamak için gerçekleştirilen istatistiksel analizler sunulmuştur. Devamında ise, “**MATERYAL VE METOD**” kısmında belirtilen sınıflandırma algoritmalarının uygulanması sonucu tahmin edilmeye çalışılan diyabetik ayak hastalarının amputasyon riskinin öngörülmesine ilişkin oluşturulan modellerin çeşitli performans değerlendirme yöntemleri ile değerlendirilmesi sonucu elde edilen makine öğrenmesi performans bulguları paylaşılmıştır.

4.1 VERİYİ ANLAMA VE GERÇEKLEŞTİRİLEN İSTATİSTİKSEL ANALİZLER

Veri setinde kayıp değerlerin doldurulması prosedüründen önce temel istatistiksel analizlerin gerçekleştirilmesi gerekmektedir.

Kullanılan testler:

- Normallik testi (ShapiroWilk ve Kolmogorov-Smirnov testlerinin her ikisinin de uygulanması yoluyla gerçekleştirilmiştir)
- Tanımlayıcı istatistikler (Frekans analizi, Betimsel istatistikler doktora tez çalışmasında sunulmuştur.)
- Bağımsız grup karşılaştırması (Bağımsız örneklem T testi, One-Way ANOVA testleri uygulanmış ve sonuçları paylaşılmıştır)
- Korelasyon Analizi (Spearman rho (ρ) sıralama korelasyon katsayısı kullanılmak suretiyle gerçekleştirilmiş ve doktora tez çalışmasında sunulmuştur)

Verilerin çözümlenmesinde frekans, yüzde, aritmetik ortalama, standart sapma, minimum, maksimum gibi tanımlayıcı istatistiklerden faydalanılmıştır. Bağımsız iki grubun ortalamaları karşılaştırılmasında parametrik testlerden bağımsız örneklem T testi, ikiden fazla bağımsız grubun karşılaştırılmasında tek yönlü varyans analizi (One-way ANOVA) testi kullanılmıştır. ANOVA testinde grup farklılıkları için varyans homojenliği Levene homojenlik testi ile değerlendirilmiş olup homojenlik sağlanıyorsa Tukey HSD, sağlanmıyorsa Tamhane çoklu grup karşılaştırma testi kullanılmıştır. Tüm test sonuçları 0.05 anlamlılık düzeyinde değerlendirilmiştir.

Tablo 4. 1: Diyabetik Ayak Hastalarına ait Sosyo-Demografik ve Genel Özelliklere İlişkin Bilgiler

Değişkenler		Hasta/Kişi sayısı (n)	%
Cinsiyet	Erkek	286	70,3
	Kadın	121	29,7
Yaş	65 yaş altı	271	66,6
	65 ve üzeri	136	33,4
Hipertansiyon	Yok	125	30,7
	Var	282	69,3
Sigara	Yok	250	62,0
	Var	153	38,0
İnsülin	Yok	61	15,0
	Var	346	85,0
Oral Antidiyabetik Ajan	Yok	226	55,5
	Var	181	44,5
Tekrarlayan DFU Öyküsü	Yok	166	40,8
	Var	241	59,2
Geçmişteki Ampütasyon Öyküsü (Yok, Minör, Majör)	Yok	250	61,4
	Baş parmak	138	33,9
	Diğer parmaklar	19	4,7
Böbrek Yetmezliği	Yok	330	81,1
	Var	77	18,9
Geçmişte Vasküler Cerrahi Öyküsü	Yok	216	53,1
	Var	191	46,9
Diyaliz	Yok	364	89,4
	Var	43	10,6
Charcot Ayak	Yok	346	85,0
	Var	61	15,0
PEDIS Sınıflandırma Sistemine Göre Sınıflandırılmış- Periferik Arter Hastalığı (P.A.H.) Evresi:	Evre 1	112	27,5
	Evre 2	186	45,7
	Evre 3	109	26,8
PEDIS Sınıflandırma Sistemine Göre Sınıflandırılmış: Yara Derinliği	Evre 1	87	21,4
	Evre 2	105	25,8
	Evre 3	215	52,8
PEDIS Sınıflandırma Sistemine Göre Sınıflandırılmış: Enfeksiyon	Evre 1	4	1,0
	Evre 2	78	19,2
	Evre 3	260	63,9
	Evre 4	65	16,0

Tablo 4.1’de görüleceği üzere; araştırmada bulunan diyabetik ayak hastalarının yaklaşık %70,3’ü erkek, %29,7’si kadın; %66,6’sı 65 yaş altı ve %33,4’ü 65 yaş ve üzerindedir.

Kişilerin %69,3’ünde hipertansiyon mevcut olup, %38’inin sigara kullandığı saptanmıştır. Hastaların %85’inde insülin kullanımı, %44,5’inde oral antidiyabetik ajan kullanımı saptanmıştır. Bununla birlikte hastaların %59,2’sinde tekrarlayan DFU olduğu; %33,9’unda

geçmişte başparmak amputasyonu, %4,7'sinde geçmişte diğer parmaklarında amputasyonu operasyonu geçirdiği tespit edilmiştir. Ayrıca hastaların %18,9'unda böbrek yetmezliği mevcut olup, %46,9'unda geçmişte vasküler cerrahi öyküsü bulunduğu, %10,6'sının ise böbrek yetmezliği nedeniyle diyalize girdiği, %15'inde ise charcot ayak olduğu bulunmuştur.

Hastaların diyabetik ayak ülserleri PEDIS-Periferik Arterial Hastalık sınıflandırmasına göre kategorize edildiğinde; hastaların %27,5'inin 1.evrede, %45,7'sinin 2.evrede ve %26,8'inin ise 3.evrede olduğu ortaya çıkmıştır. Hastaların yara derinliği PEDIS sınıflandırma sistemi kullanılarak kategorize edildiğinde; %21,4'ünün 1.evrede, %25,8'inin 2.evrede ve %52,8'inin ise 3.evrede olduğu; yine PEDIS sınıflandırma sistemi kullanılarak infeksiyon evreleri sınıflandırıldığında, %1'inin 1.evrede, %19,2'sinin 2.evrede, %63,9'unun 3.evrede ve %16'sının ise 4.evrede olduğu bulunmuştur.

Tablo 4. 2: Hastalara Ait Sosyo-Demografik ve Genel Özelliklere Ait Dağılım

		Hasta/Kişi sayısı (n)	%
Osteomyelit Durumu	Yok	175	43,0
	Var	232	57,0
Nöropati Durumu	Yok	56	13,8
	Var	351	86,2
Gangren Durumu	Yok	297	73,0
	Var	110	27,0
Retinopati Durumu	Yok	325	79,9
	Var	82	20,1
Koroner Arter Hastalığı	Yok	207	50,9
	Var	200	49,1
Astım/Koah Durumu	Yok	358	88,0
	Var	49	12,0
A.C.E. İnhibitörü Kullanım Durumu	Yok	319	78,4
	Var	88	21,6
Statin Kullanma Durumu	Yok	337	82,8
	Var	70	17,2
Canaglifozine Kullanma Durumu	Yok	400	98,3
	Var	7	1,7
Meggitt-Wagner Sınıflandırılma Sistemi Kullanılarak Derecelendirilmiş Ülser	Grade 1	58	14,3
	Grade 2	73	17,9
	Grade 3	179	44,0
	Grade 4	88	21,6
	Grade 5	9	2,2
Hiperbarik Oksijen Tedavisi (HBOT) Tedavisi Durumu	Yok	174	42,8
	Var	233	57,2
Negatif Basıncı Yara Örtüsü Tedavisi Durumu	Yok	212	52,1
	Var	195	47,9
Yumuşak Doku ile Sınırlı Debridman Tedavisi	Yok	160	39,3
	Var	247	60,7
Vasküler Girişim (Anjioplasti, Revaskülarizasyon, Bypass) Tedavisi	Yok	330	81,1
	Var	77	18,9

Tablo 4.2’de özetlendiği üzere; hastaların %57’sinde osteomyelit, %86,2’sinde nöropati, %27’sinde gangren, %20,1’inde retinopati, %49,1’inde koroner arter hastalığı, %12’sinde astım/koah saptanmıştır.

Hastaların %21,6’sı Anjiyotensin Dönüştürücü Enzim (A.C.E) inhibitörü kullanmakta olup, %17,2’si statin, %1,7’si canaglifozine kullanmaktadır.

Diyabetik ayak hastalarının ülserleri Meggit-Wagner sınıflandırma sistemi kullanılarak evrelendirildiğinde, hastaların %14,3’ü grade 1, %17,9’u grade 2, %44’ü grade 3, %21,6’sı grade 4 ve %2,2’si ise grade 5 olarak sınıflandırılmıştır.

Hastaların %57,2’sinin Hiperbarik Oksijen Tedavisi (HBOT) aldığı, %47,9’u ise Negatif Basıncılı Yara Örtüsü (NPWT) tedavisi gördüğü, %60,7’sinde ise yumuşak doku ile sınırlı debridman prosedürünün gerçekleştirildiği ve %18,9’unda ise vasküler girişim (Anjioplasti, Revaskülarizasyon, Bypass) prosedürlerinin gerçekleştirildiği tespit edilmiştir.

Tablo 4. 3: Hastalara Ait Sosyo-Demografik ve Genel Özelliklere Ait Dağılım

		Hasta/Kişi sayısı (n)	%
Hastanın İyileşme Durumu	Yok	57	14,0
	Var	350	86,0
Hedef Nitelik: Hastanın Genel “Overall” Ampütasyon Durumu (İkili Kategorik, “Binary” Sınıflandırma Problemi)	Yok	228	56,0
	Var	179	44,0
Hedef Nitelik: Hastanın Majör ve Minör Ampütasyon Durumu (Üçlü Kategorik, “Multi Class” Sınıflandırma Problemi)	Yok	228	56,0
	Minör	148	36,4
	Majör	31	7,6

Tablo 4.3’te yapılan tedaviler sonucunda hastaların sonlanımları (“outcome”) incelendiğinde ise hastaların %86’sında iyileşme gerçekleştiği görülmektedir. Bununla birlikte genel amputasyon oranı %44 olarak saptanmıştır. Detaylı olarak incelendiğinde ve amputasyon operasyonu geçiren hastaların amputasyon derecelerine bakıldığında, hastaların %36,4’ünde minör, %7,6’sında ise majör amputasyon operasyonu geçirdiği bulunmuştur.

Tablo 4. 4: Hastaların Genel Özelliklerine ait Değerlerin Tanımlayıcı İstatistiksel Analizi

	N	$\bar{X}$	Std. Sapma	Minimum	Maksimum
Yaş (Yıl)	407	60,53	10,08	32	92
Dm Süresi/ Yıl	399	15,63	9,17	0	50
Boy (Cm)	358	169,55	8,57	150	190
Kilogram (Kg)	358	81,42	16,10	40	140
VKI (Kg/M ²)	358	28,38	5,59	14,69	51,95
Sigara Paket /Yıl	123	44,36	28,98	5	150
Diyabetik Ayak İnfeksiyonu Süresi /Gün	375	217,43	285,93	0	2920
Yara Genişliği (Cm ²)	405	31,99	44,18	0,04	225,00
Lökosit (10 ³ /µl)	391	10,97	4,79	3,70	36,00
Crp (mg/L)	395	72,15	86,14	0,18	507,00
Sedimentasyon (mm/saat)	385	78,32	39,40	2,0	176,0
Hemoglobin (g/dL)	377	11,09	2,12	5,30	16,80
Bilirubin (mg/dL)	183	0,41	,26	0,020	1,780
Albümin (g/dL)	245	3,75	0,67	1,76	5,37
Hba1c (%)	326	9,02	2,30	5,10	19,60
Glukoz (mg/dL)	363	191,82	93,95	48,0	589,0
INR	175	1,17	0,51	0,81	4,20
PNL (10 ³ /µl)	308	7,35	3,98	2,60	33,80
PLT (10 ³ /µl)	351	322,88	127,22	95,0	1138,0
ALT (U/L)	360	19,37	19,06	3,0	287,0
AST (U/L)	360	20,37	14,72	7,0	199,0
Kreatinin (mg/dL)	383	1,51	1,67	0,30	10,20
Total Kolesterol (mg/dL)	204	176,24	54,97	65,0	379,0
Trigliserid (mg/dL),	195	178,05	122,53	50,0	897,0
LDL(mg/dL)	199	105,93	39,69	22,0	228,0
HDL (mg/dL)	193	34,48	13,93	8,0	115,0
HBOT Tedavisi /Seans Sayısı	209	40,98	24,11	3,0	150,0

Tablo 4.4'te verilen ve araştırmada hastaların antropometrik ve diğer özelliklere ait tanımlayıcı istatistikleri değerlendirildiğinde;

Hastaların ortalama yaşı 60,5 yıl, ortalama “*Diabetes Mellitus*” süresi 15,63 yıl, ortalama boyu 169,55 cm, ortalama kilosu 81,42 kg, ortalama VKI değeri 28,38 kg/m², yılda ortalama sigara kullanımı 44,36 paket/yıl, ortalama DFU süresi 217,43 gün, ortalama yara genişliği 31,99 cm², ortalama lökosit değeri 10,97 (10³/µl), ortalama CRP 72.15 (mg/L), ortalama sedimentasyon 78,32 (mm/saat), ortalama hemoglobin 11,09 (g/dL), ortalama bilirubin 0,41 (g/dL), ortalama albümin 3,75 (g/dL), ortalama HbA1C 9,02 (%), ortalama glukoz 191,82(mg/dL), ortalama INR 1,17, ortalama PNL 7,35 (10³/µl), ortalama PLT 322,88(10³/µl), ortalama ALT 19,37 (U/L), ortalama AST 20,37 (U/L), ortalama kreatinin 1,51 (mg/dL), ortalama total kolesterol 176,24 (mg/dL), ortalama trigliserid 178,05 (mg/dL), ortalama LDL 105,93(mg/dL), ortalama HDL 34,48 (mg/dL) ve ortalama HBOT tedavi seans sayısı ise 40,98 seans sayısı olarak saptanmıştır.

Tablo 4. 5: Hastaların Genel Özelliklerine Ait Normallik Testi Analizleri

	Normallik Testleri					
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	Df	p	Statistic	df	p
Yaş	0,044	407	0,058	0,995	407	0,233
Dm Süresi/ Yıl	0,114	399	0,000	0,964	399	0,000
Vkı	0,094	358	0,000	0,949	358	0,000
Sigara Paket /Yıl	0,163	123	0,000	0,901	123	0,000
DFU Süresi /Gün	0,234	375	0,000	0,503	375	0,000
YARA GENİŞLİĞİ (cm ²)	0,235	405	0,000	0,692	405	0,000
Lökosit (10 ³ /µl)	0,120	391	0,000	0,859	391	0,000
CRP (mg/L)	0,202	395	0,000	0,787	395	0,000
Sedimentasyon (mm/saat)	0,085	385	0,000	0,966	385	0,000
Hemoglobin(g/dL)	0,047	377	0,041	0,994	377	0,179
Bilirubin (mg/dL)	0,165	183	0,000	0,834	183	0,000
Albümin(g/dL),	0,080	245	0,001	0,980	245	0,002
HbA1C(%)	0,074	326	0,000	0,946	326	0,000
Glukoz(mg/dL)	0,104	363	0,000	0,916	363	0,000
INR	0,322	175	0,000	0,452	175	0,000
PNL (10 ³ /µl)	0,149	308	0,000	0,800	308	0,000
PLT(10 ³ /µl)	0,085	351	0,000	0,905	351	0,000
ALT(U/L)	0,222	360	0,000	0,473	360	0,000
AST(U/L)	0,216	360	0,000	0,563	360	0,000
Kreatinin(mg/dL)	0,312	383	0,000	0,526	383	0,000
Total Kolesterol(mg/dL)	0,062	204	0,056	0,965	204	0,000
Trigliserid(mg/dL),	0,197	195	0,000	0,707	195	0,000
LDL(mg/dL),	0,052	199	0,200	0,987	199	0,065
HDL(mg/dL),	0,092	193	0,000	0,915	193	0,000
HBOT Tedavisi /Seans Sayısı	0,207	209	0,000	0,868	209	0,000

Tablo 4.5'te görüleceği üzere, doktora tez çalışmasında, hastaların genel özelliklerine ait değerlerin normallik dağılımı analizinde **Kolmogorov Smirnov** ve **Shapiro-Wilk** testleri ile değerlendirildiğinde;

Kolmogorov-Smirnov testi sonuçlarına göre yaş, total kolesterol (mg/dL) ve LDL (mg/dL) değeri normal dağılıma uygun olup ($p \geq 0.05$) bununla birlikte diğer değerler normal dağılım göstermemiştir. Shapiro-Wilk testi sonuçlarına göre ise yaş, hemoglobin (g/dL) ve LDL (mg/dL) değerleri normal dağılıma uygun olup ($p \geq 0.05$) bununla birlikte diğer değerler normal dağılım göstermemiştir.

Tablo 4. 6: Hastaların Genel Özelliklerin Logaritmik Dönüştürülmüş Değerlerine Ait Normallik Dağılım Analizleri

	Normallik Testleri					
	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	p	Statistic	df	p
LOGDMSÜRESİ	0,167	383	0,000	0,935	383	0,000
LOGBKİ	0,058	358	0,006	0,986	358	0,002
LOGSİGARAPAKET	0,145	123	0,000	0,957	123	0,001
LOGDAESÜRESİ	0,049	374	0,033	0,990	374	0,012
LOGYARAGENİŞLİĞİ	0,092	405	0,000	0,972	405	0,000
LOGLÖKOSİT	0,041	391	0,135	0,989	391	0,005
LOGCRP	0,072	395	0,000	0,965	395	0,000
LOGSEDİM	0,154	385	0,000	0,879	385	0,000
LOGHEMOGLOBİN	0,049	377	0,031	0,991	377	0,019
LOGBİLİRUBİN	0,070	183	0,030	0,966	183	0,000
LOGALBÜMİN	0,098	245	0,000	0,940	245	0,000
LOGHbA1C	0,052	326	0,034	0,989	326	0,011
LOGGLUKOZ	0,040	363	0,200	0,992	363	0,060
LOGINR	0,227	175	0,000	0,628	175	0,000
LOGPNL	0,068	308	0,001	0,978	308	0,000
LOGPLT	0,039	351	0,200	0,992	351	0,049
LOGALT	0,094	360	0,000	0,966	360	0,000
LOGAST	0,106	360	0,000	0,945	360	0,000
LOGKREATİNİN	0,177	383	0,000	0,854	383	0,000
LOGTRIGLİSERİD	0,073	195	0,013	0,974	195	0,001
LOGHDL	0,106	193	0,000	0,968	193	0,000
LOGHBOTEDAVİ	0,209	209	0,000	0,924	209	0,000

Tablo 4.6’da verildiği üzere doktora tez çalışmasında, hastaların genel özelliklerine ait değerleri logaritmik dönüşüm sonucunda tekrar normallik dağılım analizleri açısından Kolmogorov Smirnov ve Shapiro-Wilks testleri ile tekrar değerlendirildiğinde;

Kolmogorov-Smirnov testi sonuçlarına göre lökosit ($10^3/\mu\text{l}$), glukoz (mg/dL) ve PLT ($10^3/\mu\text{l}$) değerlerinin normal dağılıma uygun olduğu ($p \geq 0.05$); Shapiro-Wilk testi sonuçlarına göre ise glukoz (mg/dL) değerinin normal dağılıma uygun olduğu görülmektedir ($p \geq 0.05$).

Her ne kadar normallik dağılım analizleri sonucunda hastaların genel özelliklerinden lökosit ($10^3/\mu\text{l}$), glukoz (mg/dL) ve PLT ($10^3/\mu\text{l}$) değerleri dışındaki değerler normal dağılım göstermemiş olsa da, çalışmanın örneklem büyüklüğü 30’un üzerinde olduğundan ve özellikle “Merkezi Limit Teoremi”ne göre örneklem sayısı 100’ün üzerinde olduğundan toplamalar ve ortalamalar her zaman normal dağıldığı kabul edilebileceğinden dolayı, genel hasta özelliklerine ilişkin değerlerin normal olarak dağıldığı kabul edilmiştir (George ve Mallery, c2010.; Jolliffe, 1995). Bu nedenle çalışmada parametrik testler gerçekleştirilmiş olup istatistiksel analizler bu kapsamda ele alınmıştır.

Tablo 4. 7: Hastaların Genel Özellikleri (Kategorik Değişkenler) ile Karar Aşaması Arasındaki İlişki Analizi

		GENEL ("OVERALL") AMPÜTASYON STATÜSÜ				χ² (p)
		Yok		Var		
		N	%	n	%	
Cinsiyet	Erkek	153	67,1	133	74,3	3,091
	Kadın	75	32,9	46	25,7	(0,213)
Yaş	65 yaş altı	158	69,3	113	63,1	2,486
	65 ve üzeri	70	30,7	66	36,9	(0,115)
Hipertansiyon	Yok	76	33,3	49	27,4	1,673
	Var	152	66,7	130	72,6	(0,196)
Sigara	Yok	148	64,9	102	58,3	1,846
	Var	80	35,1	73	41,7	(0,174)
İnsülin	Yok	42	18,4	19	10,6	4,796
	Var	186	81,6	160	89,4	(0,029*)
Oral Antidiyabetik Ajan	Yok	123	53,9	103	57,5	0,525
	Var	105	46,1	76	42,5	(0,469)
Tekrarlayan DFU	Yok	97	42,5	69	38,5	0,663
	Var	131	57,5	110	61,5	(0,415)
Geçmişteki Ampütasyon Öyküsü (Yok, Minör, Majör)	Yok	157	68,9	93	52	14,915 (0,001*)
	Minör	59	25,9	79	44,1	
	Majör	12	5,3	7	3,9	
Böbrek Yetmezliği	Yok	194	85,1	136	76	5,425
	Var	34	14,9	43	24	(0,020*)
Geçmişteki Vasküler Cerrahi Operasyon Öyküsü	Yok	131	57,5	85	47,5	4,002
	Var	97	42,5	94	52,5	(0,045*)
Diyaliz	Yok	210	92,1	154	86	3,912
	Var	18	7,9	25	14	(0,048*)
Charcot Ayak	Yok	196	86	150	83,8	0,369
	Var	32	14	29	16,2	(0,543)
PEDIS Sınıflandırma Sistemine Göre Sınıflandırılmış- Periferik Arter Hastalığı (P.A.H) Evresi	Evre 1	95	41,7	17	9,5	113,043 (0,000*)
	Evre 2	116	50,9	70	39,1	
	Evre 3	17	7,5	92	51,4	
PEDIS Sınıflandırma Sistemine Göre Sınıflandırılmış Yara Derinliği	Evre 1	84	36,8	3	1,7	160,175 (0,000*)
	Evre 2	86	37,7	19	10,6	
	Evre 3	58	25,4	157	87,7	
PEDIS Sınıflandırma Sistemine Göre Sınıflandırılmış İnfeksiyon	Evre 1	4	1,8	0	0	96,114 (0,000*)
	Evre 2	76	33,3	2	1,1	
	Evre 3	136	59,6	124	69,3	
	Evre 4	12	5,3	53	29,6	

Tablo 4.7’de detaylı olarak gösterildiği üzere; hastaların genel özellikleri ile amputasyon statüleri arasındaki ilişki analizi ki-kare testi ile değerlendirildiğinde, insülin, geçmişteki

ampütasyon statüsü, böbrek yetmezliği, geçmişte vasküler cerrahi operasyonu durumu, diyaliz, PEDİS sınıflandırma sistemine göre sınıflandırılmış- periferik arter hastalığı (P.A.H) evresi: PEDİS sınıflandırma sistemine göre sınıflandırılmış yara derinliği ve PEDİS sınıflandırma sistemine göre sınıflandırılmış enfeksiyon evreleri ile ampütasyon statüsü arasında anlamlı bir ilişki saptanmıştır ($p \leq 0,05$).

Tablo 4. 8: Hastaların Genel Özellikleri (Kategorik Değişkenler) ile Ampütasyon Statüsü arasındaki İlişki Analizi

		GENEL (“OVERALL”) AMPÜTASYON STATÜSÜ				X ² (p)
		Yok		Var		
		n	%	n	%	
Osteomiyelit Durumu	Yok	161	70,6	14	7,8	161,315 (0,000*)
	Var	67	29,4	165	92,2	
Nöropati Durumu	Yok	51	22,4	5	2,8	32,381 (0,000*)
	Var	177	77,6	174	97,2	
Gangren Durumu	Yok	219	96,1	78	43,6	140,015 (0,000*)
	Var	9	3,9	101	56,4	
Retinopati Durumu	Yok	196	86	129	72,1	12,039 (0,001*)
	Var	32	14	50	27,9	
Koroner Arter Hastalığı Durumu	Yok	129	56,6	78	43,6	6,784 (0,009*)
	Var	99	43,4	101	56,4	
Astım/Koah Durumu	Yok	201	88,2	157	87,7	0,019 (0,890)
	Var	27	11,8	22	12,3	
Ace İnh. Kullanımı Durumu	Yok	180	78,9	139	77,7	0,099 (0,753)
	Var	48	21,1	40	22,3	
Statin Kullanım Durumu	Yok	195	85,5	142	79,3	2,704 (0,100)
	Var	33	14,5	37	20,7	
Canaglifozine Kullanım Durumu	Yok	225	98,7	175	97,8	0,501 (0,479)
	Var	3	1,3	4	2,2	
Wagner Grade	Grade 1	58	25,4	0	0	185,513 (0,000*)
	Grade 2	67	29,4	6	3,4	
	Grade 3	98	43	81	45,3	
	Grade 4	5	2,2	83	46,4	
	Grade 5	0	0	9	5	
Hbot Tedavisi Alıp Almama Durumu	Yok	120	52,6	54	30,2	20,675 (0,000*)
	Var	108	47,4	125	69,8	
Negatif Basınçlı Yara Tedavisi (NPWT) Durumu	Yok	153	67,1	59	33	46,844 (0,000*)
	Var	75	32,9	120	67	

Tablo 4.8: Hastaların Genel Özellikleri (Kategorik Değişkenler) ile Ampütasyon Statüsü arasındaki İlişki Analizi (devam)

		GENEL (“OVERALL”) AMPÜTASYON STATÜSÜ				χ² (p)
		Yok		Var		
		n	%	n	%	
Yumuşak Doku ile Sınırlı Debridman Prosedürü Durumu	Yok	122	53,5	38	21,2	43,795 (0,000*)
	Var	106	46,5	141	78,8	
Vasküler Girişim (Bypass, Anjioplasti) Operasyonu Durumu	Yok	201	88,2	129	72,1	16,925 (0,000*)
	Var	27	11,8	50	27,9	
Hastanın İyileşme Durumu	Yok	14	6,1	43	24	26,624 (0,000*)
	Var	214	93,9	136	76	

Tablo 4.8’de görülebileceği üzere, doktora tez çalışmasında hastaların genel özellikleri (kategorik değişkenler) ile genel-“overall” statüleri arasındaki ilişki analizi **ki-kare testi ile** değerlendirildiğinde, osteomyelit, nöropati, gangren, retinopati, koroner arter hastalığı, Wagner dereceleri, HBOT durumu, NPWT durumu, yumuşak doku ile sınırlı debridman prosedürü durumu, vasküler girişim ve iyileşme durumu ile genel-“overall” amputasyon statüleri arasında anlamlı bir ilişki tespit edilmiştir ($p \leq 0,05$).

Tablo 4. 9: Hastaların Ampütasyon Statülerine göre Genel Özelliklerinin (Nümerik Değişkenler) Karşılaştırılması

	GENEL (“OVERALL”) AMPÜTASYON STATÜSÜ ¹				t	p
	Yok		Var			
	$\bar{X}$	SS.	$\bar{X}$	SS.		
Yaş	59,92	9,66	61,31	10,57	1,384	0,167
D.M. Süresi/ Yıl	15,46	9,44	15,85	8,84	0,430	0,668
VKI	29,33	5,66	27,02	5,21	-3,918	0,000*
Sigara Paket /Yıl	41,28	26,75	48,16	31,35	1,314	0,191
DFU Süresi /Gün	212,35	276,25	224,04	298,78	0,392	0,695
Yara Genişliği (Cm²)	24,46	40,35	41,68	47,03	3,885	0,000*
Lökosit(10³/µl)	10,19	4,16	11,96	5,34	3,579	0,000*
CRP(Mg/L)	48,8	71,33	101,21	94,00	6,115	0,000*
Sedimentasyon(Mm/Saat)	64,36	37,32	95,98	34,68	8,591	0,000*
Hemogloblin(G/Dl)	11,60	2,05	10,45	2,03	-5,435	0,000*
Bilirubin(Mg/Dl)	0,39	0,2	0,44	0,3	1,376	0,171
Albümin(G/Dl)	3,91	0,66	3,54	0,63	-4,451	0,000*
Hba1c (%)	9,07	2,22	8,97	2,4	-0,379	0,705
Glukoz(Mg/Dl)	184,78	82,28	200,37	106,04	1,576	0,116

¹Bağımsız Örneklem T Testi, *P≤0,05

Tablo 4. 9: Hastaların Ampütasyon Statülerine göre Genel Özelliklerinin (Nümerik Değişkenler) Karşılaştırılması (devam)

	GENEL (“ <i>OVERALL</i> ”) AMPÜTASYON STATÜSÜ ¹				t	p
	Yok		Var			
	$\bar{X}$	SS.	$\bar{X}$	SS.		
INR	1,2	0,56	1,14	0,46	-0,764	0,446
PNL(10 ³ /µl)	6,89	3,72	7,99	4,26	2,361	0,019*
PLT(10 ³ /µl)	307,91	122,87	343,89	130,63	2,634	0,009*
ALT(U/L)	18,14	12,91	20,91	24,64	1,287	0,199
AST(U/L)	18,76	9,03	22,37	19,49	2,165	0,032*
Kreatinin(Mg/Dl)	1,45	1,62	1,6	1,72	0,897	0,370
Total Kolesterol(Mg/Dl)	175,9	53,49	176,62	56,9	0,093	0,926
Trigliserid(Mg/Dl)	162,94	106,18	196,05	138	1,892	0,060
LDL(Mg/Dl)	105,99	40,73	105,87	38,64	-0,022	0,983
HDL(Mg/Dl)	36,84	14,72	31,54	12,33	-2,669	0,008*
HBOT Tedavisi /Seans Sayısı	36,80	20,92	44,52	26,09	2,373	0,019*

¹Bağımsız Örneklem T Testi, *P≤0,05

Tablo 4.9’da görüleceği üzere; araştırmada ampütasyon operasyonu geçiren hastaların yara genişliği (cm²) ($\bar{X}$: 42.68, SS: 47.03), Lökosit (10³/µl) ($\bar{X}$: 11.96, SS: 5.34), CRP (mg/L) ($\bar{X}$: 101.21, SS: 94.00), sedimantasyon (mm/saat) ($\bar{X}$: 95.98, SS: 34.68), PNL (10³/µl) ($\bar{X}$: 7.99, SS: 4.26), PLT (10³/µl) ($\bar{X}$: 343.89, SS: 130.63), AST (U/L) ($\bar{X}$: 22.37, SS: 19.49) ve HBOT tedavisi /seans sayısı ($\bar{X}$: 44.52, SS: 26.09) değerleri anlamlı olarak ampütasyon operasyonu geçirmeyen hastaların ortalama değerinden yüksek olarak saptanmıştır.

Bununla birlikte, ampütasyon olmayan hastaların VKI ($\bar{X}$: 29.33, SS: 5.66), hemoglobin (g/dL) ($\bar{X}$: 11.60, SS: 2.05), albümin (g/dL) ($\bar{X}$: 3.91, SS: 0.66) ve HDL (mg/dL) ($\bar{X}$: 36.80 SS: 20.92) değerleri ampütasyon operasyonu geçiren hastaların ortalama değerinden yüksek olduğu tespit edilmiştir (p≤0,05).

Tablo 4. 10: Hastaların Minör ve Majör Ampütasyon (Üçlü Kategorik - “*Multi Class*”) Statülerine Göre Genel Özelliklerinin (Nümerik Değişkenler) Karşılaştırılması

	Minör ve Majör (Üçlü Kategorik - “Multi Class”) Ampütasyon Statüsü ¹						F	p
	Yok		Minör		Major			
	$\bar{X}$	SS	$\bar{X}$	SS	$\bar{X}$	SS		
Yaş	59,92	9,66	60,87	10,53	63,42	10,69	1,781	0,170
Dm Süresi/ Yıl	15,46	9,44	15,1	8,23	19,42	10,71	2,962	0,053
VKI	29,33	5,66	26,78	5,08	28,04	5,72	8,284	0,000*
Sigara Paket /Yıl	41,28	26,75	49,12	32,43	45,36	28,89	0,945	0,391
DFU Süresi /Gün	212,35	276,25	230,30	324,7	193,86	106	0,264	0,768
Yara Genişliği (cm²)	24,46	40,35	35,10	41,88	72,68	57,42	18,257	0,000*
Lökosit(10³/µl)	10,19	4,16	11,75	5,42	13,00	4,87	7,653	0,001*
CRP (mg/L)	48,80	71,33	97,80	97,46	118,49	72,94	20,625	0,000*
Sedimentasyon (mm/saat)	64,36	37,32	93,31	34,90	110,77	29,89	39,260	0,000*
Hemoglobin (g/dL)	11,60	2,05	10,60	1,99	9,73	2,12	17,088	0,000*
Bilirubin (mg/dL)	0,39	0,20	0,43	0,27	0,5	0,45	1,485	0,229
Albümin (g/dL)	3,91	0,66	3,60	0,62	3,26	0,64	12,102	0,000*
Hba1c (%)	9,07	2,22	9,06	2,43	8,41	2,19	0,780	0,459
Glukoz(mg/dL)	184,78	82,28	201,42	108,72	194,52	91,39	1,296	0,275
INR	1,20	0,56	1,11	0,35	1,32	0,79	1,461	0,235
PNL(10³/µl)	6,89	3,72	7,81	4,19	9,18	4,65	3,809	0,023*
PLT(10³/µl)	307,91	122,87	348,93	134,61	313,9	101,39	4,165	0,016*
ALT(U/L)	18,14	12,91	20,79	25,73	21,52	18,39	0,954	0,386
AST(U/L)	18,76	9,03	22,27	19,54	22,88	19,6	2,718	0,067
Kreatinin(mg/dL)	1,45	1,62	1,46	1,56	2,32	2,30	3,424	0,034*
Total Kolesterol(mg/dL)	175,9	53,49	182,34	55,99	146,13	53,51	2,793	0,064
Trigliserid(mg/dL)	162,94	106,18	197,77	139,26	186,82	135,64	1,829	0,163
LDL(mg/dL)	105,99	40,73	109,62	37,05	85,25	42,08	2,261	0,107
HDL(mg/dL)	36,84	14,72	32,85	12,01	24,15	11,86	5,891	0,003*
HBOT Tedavisi /Seans Sayısı	36,8	20,92	44,92	26,79	41,93	21,56	2,809	0,063

¹Tek Yönlü Varyans (One-Way ANOVA) Analizi, *p≤0.05

Tablo 4.10’da görüleceği üzere; doktora tez çalışmasında, hastaların genel özelliklerine ait değerlerin (nümerik değişkenler) minör ve majör Ampütasyon (Üçlü Kategorik - “*Multi Class*”) statülerine göre karşılaştırılması tek yönlü varyans analizi ile değerlendirildiğinde;

Hastaların VKI, yara genişliği(cm²), lökosit (10³/µl), CRP (mg/L), sedimentasyon (mm/saat), hemoglobin (g/dL), albümin(g/dL), PNL (10³/µl), PLT (10³/µl), kreatinin (mg/dL) ve HDL (mg/dL) değerleri ile amputasyon statüleri arasında anlamlı bir farklılık saptanmıştır (p≤0.05).

Varyansların homojenliğini test etmek için **Levene homojenlik testi** kullanılmıştır ve **varyans homojenliği sağlanıyorsa ($p \geq 0.05$) Tukey HSD, homojenlik sağlanmıyorsa ($p < 0.05$) Tamhane testi kullanılmıştır.**

VKI değerleri için gruplar arasındaki farklılıklar Tukey HSD çoklu karşılaştırma testi ile değerlendirildiğinde, amputasyon operasyonu geçirmeyen hastaların ortalama VKI değeri ($\bar{X}$: 29.33, SS: 5.66) anlamlı olarak minör amputasyon operasyonu geçiren hastaların ortalama VKI değerinden ($\bar{X}$: 26.78, SS: 5.08) yüksek olduğu tespit edilmiştir.

Yara genişliği (cm^2) için gruplar arasındaki farklılıklar Tamhane çoklu karşılaştırma testi ile değerlendirildiğinde, majör amputasyon operasyonu geçiren hastaların ortalama yara genişliği (cm^2) ($\bar{X}$: 72.68, SS: 57.42) anlamlı olarak minör amputasyon operasyonu geçiren hastaların ortalama yara genişliği (cm^2) ($\bar{X}$: 35.10, SS: 41.88) ve amputasyon operasyonu geçirmeyen hastaların ortalama yara genişliği (cm^2) ($\bar{X}$: 24.46, SS: 40.35) değerinden yüksek olduğu saptanmıştır.

Lökosit ($10^3/\mu\text{l}$) için gruplar arasındaki farklılıklar Tamhane çoklu karşılaştırma testi ile değerlendirildiğinde, majör amputasyon operasyonu geçiren hastaların ortalama lökosit ($10^3/\mu\text{l}$) ($\bar{X}$: 13.00, SS: 4.87) ve minör amputasyon operasyonu geçiren hastaların ortalama lökosit ($10^3/\mu\text{l}$) ($\bar{X}$: 11.75, SS: 5.42) değerleri; anlamlı olarak amputasyon operasyonu geçirmeyen hastaların ortalama lökosit ($10^3/\mu\text{l}$) ($\bar{X}$: 10.19, SS: 4.16) değerinden yüksek olduğu saptanmıştır.

CRP (mg/L) için gruplar arasındaki farklılıklar Tamhane çoklu karşılaştırma testi ile değerlendirildiğinde, majör amputasyon operasyonu geçiren hastaların ortalama CRP (mg/L) ($\bar{X}$: 118.49, SS: 72.94) ve minör amputasyon geçiren hastaların ortalama CRP (mg/L) ($\bar{X}$: 97.80, SS: 97.46) değerleri; anlamlı olarak amputasyon operasyonu geçirmeyen hastaların ortalama CRP (mg/L) ($\bar{X}$: 48.80, SS: 71.33) değerinden yüksek olduğu bulunmuştur.

Sedimentasyon (mm/saat) için gruplar arasındaki farklılıklar Tamhane çoklu karşılaştırma testi ile değerlendirildiğinde, majör amputasyon operasyonu geçiren hastaların ortalama sedimentasyon (mm/saat) ($\bar{X}$: 110.77, SS: 29.89) değerleri anlamlı olarak minör amputasyon operasyonu geçiren hastaların ortalama sedimentasyon (mm/saat) ($\bar{X}$: 93.31, SS: 34.90) ve amputasyon operasyonu geçirmeyen hastaların ortalama sedimentasyon (mm/saat) ($\bar{X}$: 64.36, SS: 37.32) değerinden yüksek olduğu görülmektedir.

Hemoglobin (g/dL) değerleri için gruplar arasındaki farklılıklar Tukey HSD çoklu karşılaştırma testi ile değerlendirildiğinde, amputasyon operasyonu geçirmeyen hastaların ortalama hemoglobin (g/dL) değeri ($\bar{X}$: 11.60, SS: 2.05) anlamlı olarak minör amputasyon operasyonu geçiren hastaların ortalama hemoglobin (g/dL) ($\bar{X}$: 10.60, SS: 1.99) ve majör amputasyon operasyonu geçiren hastaların ortalama hemoglobin (g/dL) değerinden ($\bar{X}$: 9.73, SS: 2.12) yüksek olduğu görülmektedir. Albümin (g/dL) değerleri için gruplar arasındaki farklılıklar Tukey HSD çoklu karşılaştırma testi ile değerlendirildiğinde, amputasyon operasyonu geçirmeyen hastaların ortalama albümin (g/dL) değeri ($\bar{X}$: 3.91, SS: 0.66) anlamlı olarak minör amputasyon operasyonu geçiren hastaların ortalama albümin (g/dL) ($\bar{X}$: 3.60, SS: 0.62) ve majör amputasyon operasyonu geçiren hastaların ortalama albümin (g/dL) değerinden ($\bar{X}$: 3.26, SS: 0.64) yüksek olduğu tespit edilmiştir.

PNL ($10^3/\mu\text{l}$) değerleri için gruplar arasındaki farklılıklar Tukey HSD çoklu karşılaştırma testi ile değerlendirildiğinde, major amputasyon operasyonu geçiren hastaların ortalama PNL ($10^3/\mu\text{l}$) değeri ($\bar{X}$: 9.18, SS: 4.65) anlamlı olarak amputasyon operasyonu geçirmeyen hastaların ortalama PNL ($10^3/\mu\text{l}$) değerinden ($\bar{X}$: 6.89, SS: 3.72) yüksek olduğu saptanmıştır. PLT ($10^3/\mu\text{l}$) değerleri için gruplar arasındaki farklılıklar Tukey HSD çoklu karşılaştırma testi ile değerlendirildiğinde, minör amputasyon operasyonu geçiren hastaların ortalama PLT ($10^3/\mu\text{l}$) değeri ($\bar{X}$: 348.93, SS: 134.61) anlamlı olarak amputasyon operasyonu geçirmeyen hastaların ortalama PLT ($10^3/\mu\text{l}$) değerinden ($\bar{X}$: 307.91, SS: 122.87) yüksek olduğu tespit edilmiştir. Kreatinin (mg/dL) değerleri için gruplara arasındaki farklılıklar Tamhane çoklu karşılaştırma testi ile değerlendirildiğinde, majör amputasyon operasyonu geçiren hastaların ortalama kreatinin (mg/dL) ($\bar{X}$: 2.32, SS: 2.30) anlamlı olarak minör amputasyon operasyonu geçiren hastaların ortalama kreatinin (mg/dL) ($\bar{X}$: 1.46, SS: 1.56) ve amputasyon operasyonu geçirmeyen hastaların ortalama kreatinin (mg/dL) ($\bar{X}$: 1.45, SS: 1.62) değerinden yüksek olduğu bulunmuştur. HDL (mg/dL) değerleri için gruplar arasındaki farklılıklar Tukey HSD çoklu karşılaştırma testi ile değerlendirildiğinde, amputasyon operasyonu geçirmeyen hastaların ortalama HDL (mg/dL) değeri ($\bar{X}$: 36.84, SS: 14.72) anlamlı olarak majör amputasyon operasyonu geçiren hastaların ortalama HDL (mg/dL) değerinden ($\bar{X}$: 24.15, SS: 11.86) yüksek olduğu görülmektedir.

Bununla birlikte hastaların yaş, diabetes mellitus süresi, yıllık sigara kullanımı (paket), DFU süresi, bilirubin (mg/dL), HbA1C (%), glukoz(mg/dL) , INR, ALT (U/L), AST (U/L), total

kolesterol (mg/dL), trigliserid (mg/dL) ve HBOT Tedavisi /Seans Sayısı değerleri ile hastaların amputasyon statüleri arasında anlamlı farklılık saptanamamıştır ($p>0.05$).

4.3 SPEARMAN RHO KORELASYON ANALİZİ TESTİ SONUÇLARI

Tablo 4. 11: Hastaların Genel Özellikleri ile Genel “Overall” Amputasyon Statüsü Arasındaki İlişkinin Değerlendirmesinde Spearman Rho Korelasyon Analizi

	Genel “Overall” Amputasyon Statüsü	
	r	p
Kilo (N=358)	-0,186	0,000**
İnsülin Kullanımı (N=407)	0,109	0,029*
Geçmişteki Amputasyon Statüsü (Yok, Minör, Majör) (N=407)	0,157	0,001**
Böbrek Yetmezliği (N=407)	0,115	0,020*
Geçmişteki Vasküler Cerrahi Öyküsü (N=407)	0,099	0,046*
Diyaliz Durumu (N=407)	0,098	0,048*
PEDİS-P.A.H. Evresi (N=407)	0,512	0,000**
Yara Genişliği (Cm ²) (N=405)	0,307	0,000**
PEDİS-Yara Derinliği Evresi (N=407)	0,620	0,000**
PEDİS- İnfeksiyon Evresi (N=407)	0,484	0,000**
Osteomiyelit Durumu (N=407)	0,630	0,000**
Lökosit (10 ³ /µl) (N=391)	0,186	0,000**
CRP (Mg/L) (N=395)	0,361	0,000**
Sedimantasyon (Mm/Saat) (N=385)	0,403	0,000**
Hemoglobin (G/Dl) (N=377)	-0,275	0,000**
Albümin (G/Dl) (N=245)	-0,282	0,000**
PNL (10 ³ /µl) (N=308)	0,140	0,014**
PLT (10 ³ /µl) (N=351)	0,158	0,003**
Trigliserid (Mg/Dl) (N=195)	0,154	0,032*
HDL (Mg/Dl) (N=193)	-0,183	0,011*
Nöropati Durumu (N=407)	0,282	0,000**
Gangren Durumu (N=407)	0,587	0,000**
Retinopati Durumu (N=407)	0,172	0,000**
Koroner Arter Hastalığı Durumu (N=407)	0,129	0,009*
Wagner Grade (N=407)	0,668	0,000**
Hbot Durumu/Var Yok (N=407)	0,225	0,000**
Hbot Tedavisi/ Seans Sayısı (N=209)	0,149	0,031*
Npwt Tedavi Durumu (N=407)	0,339	0,000**
Yumuşak Doku ile Sınırlı Debridman Prosedürü Statüsü (N=407)	0,328	0,000**
Vasküler Girişim Operasyonu Durumu (Bypass, Anjioplasti) (N=407)	0,204	0,000**
İyileşme (N=407)	-0,256	0,000**

Spearman Rho, *P≤0.05, **P≤0.01

Tablo 4.11’de görüleceği üzere Araştırmada hastaların genel özellikleri ile **genel “overall” ampütasyon statüsü** arasındaki ilişki durumu Spearman Rho korelasyon testi ile değerlendirildiğinde;

Hastaların kilo ($r:-0.186$, $p\leq 0.01$), hemoglobin (g/dL) ($r:-0.275$, $p\leq 0.01$), albümin (g/dL) ($r:-0.282$, $p\leq 0.01$), HDL (mg/dL) ($r:-0.183$, $p\leq 0.05$) ve iyileşme ($r:-0.256$, $p\leq 0.01$) nitelikleri ile genel – “overall” ampütasyon statüleri arasında negatif yönlü, zayıf düzeyde ve anlamlı bir ilişki saptanmıştır.

Hastaların insülin kullanımı ($r:0.109$, $p\leq 0.05$), geçmişteki ampütasyon operasyonu geçirme öyküsü ($r:0.157$, $p\leq 0.01$), böbrek yetmezliği durumu ($r:0.115$, $p\leq 0.05$), geçmişteki vasküler cerrahi operasyonu öyküsü ($r:0.099$, $p\leq 0.05$), diyaliz ($r:0.098$, $p\leq 0.05$), yara genişliği (cm²) ($r:0.307$, $p\leq 0.01$), lökosit (10³/µl) ($r:0.186$, $p\leq 0.01$), CRP (mg/L) ($r:0.361$, $p\leq 0.01$), PNL (10³/µl) ($r:0.140$, $p\leq 0.05$), PLT (10³/µl) ($r:0.158$, $p\leq 0.01$), trigliserid (mg/dL), ($r:0.154$, $p\leq 0.05$), nöropati durumu ($r:0.282$, $p\leq 0.01$), retinopati durumu ($r:0.172$, $p\leq 0.01$), koroner arter hastalığı durumu ($r:0.129$, $p\leq 0.01$), HBOT tedavisi alıp almama durumu ($r:0.225$, $p\leq 0.01$), HBOT tedavisi/seans sayısı ($r:0.149$, $p\leq 0.05$), NPWT tedavisi alıp almama durumu ($r:0.339$, $p\leq 0.01$), yumuşak doku ile sınırlı debridman tedavisi alıp almama durumu ($r:0.328$, $p\leq 0.01$) ve vasküler cerrahi girişim statüsü ($r:0.204$, $p\leq 0.01$) ile genel – “overall” ampütasyon statüsü arasında pozitif yönde, zayıf düzeyde ve anlamlı bir ilişki tespit edilmiştir. Hastaların PEDIS-P.A.H. evresi ($r:0.512$, $p\leq 0.01$), PEDIS-infeksiyon evresi ($r:0.484$, $p\leq 0.01$), sedimantasyon (mm/saat) ($r:0.403$, $p\leq 0.01$), osteomyelit durumu ($r:0.630$, $p\leq 0.01$), gangren durumu ($r:0.587$, $p\leq 0.01$) ile genel – “overall” ampütasyon statüsü arasında pozitif yönde, normal düzeyde ve ayrıca yara derinliği ($r:0.620$, $p\leq 0.01$) ve Wagner grade değişkeni ($r:0.668$, $p\leq 0.01$) ile genel – “overall” ampütasyon statüsü arasında pozitif yönde, güçlü düzeyde ve anlamlı bir ilişki bulunmuştur.

4.4 VERİ MADENCİLİĞİ SÜRECİNDE MAKİNE ÖĞRENMESİ SINIFLANDIRMA ALGORİTMALARININ UYGULANMASI SONUCU ELDE EDİLEN BULGULAR – İKİLİ SINIFLANDIRMA

Bu bölümde diyabetik ayak hastalarının ampütasyon riskini öngörmek için kullanılan sınıflandırma algoritmaları olan k-En Yakın Komşu algoritması, Sade-“Naive” Bayes

Sınıflandırıcı, Sınıflandırma ve Regresyon Ağaçları algoritması(CART), Rastgele Orman algoritması, Yapay Sinir Ağları, Destek Vektör Makineleri ve XGBoost algoritması ile birlikte hiperparametre optimizasyonu olarak seçilen ızgara arama yöntemi ve genetik algoritma ile parçacık sürü optimizasyonu algoritmalarından elde edilen bulgular paylaşılmıştır. Öncelikli amaç genel ampütasyonu tahmin etmek olduğundan ve genel ampütasyon veri setinin hedef niteliği dengeli bir dağılıma sahip olduğundan, ilk olarak ikili- “*binary*” sınıflandırma sonuçları paylaşılmıştır.

4.4.1 K- En Yakın Komşu Algoritmasından Elde Edilen Bulgular (İkili Sınıflandırma)

4.4.1.1 Orijinal Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular

Tablo 4. 12: Orijinal Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular

K-En Yakın Komşu Algoritması – Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test K=default	0,7949	0,8235	0,7727	0,7368	0,85	0,777741204
%80 Eğitim %20 Test K=default	0,85	0,8286	0,8667	0,8286	0,8667	0,8286
%70 Eğitim %30 Test K=default	0,7769	0,6415	0,8824	0,8095	0,7595	0,715774294
%60 Eğitim %40 Test K=default	0,821	0,7465	0,8791	0,8281	0,8163	0,785185634
BOOTSTRAP (1:50) K=default	0,7533	0,8	0,7189	0,6772	0,8298	0,733495803
BOOTSTRAP (1:100) K=default	0,7414	0,7535	0,733	0,6605	0,8118	0,703941655
BOOTSTRAP (1:200) K=default	0,7446	0,5932	0,8562	0,7527	0,7405	0,663498982
ÜÇLÜ AYIRMA YÖNTEMİ ¹	0,9268	0,8889	0,9565	0,9412	0,9167	0,914302694
ÜÇLÜ AYIRMA YÖNTEMİ ²	0,6829	0,5556	0,7826	0,6667	0,6923	0,606100826

1-%80 EĞİTİM (Izgara arama-“grid search” yöntemi ile optimum komşu sayısı K=15 olarak tespit edilmiştir) %10 Geçerleme (Validasyon)

%10 TEST/ Validasyon Veri Seti Performans Sonuçları

2-%80 EĞİTİM, %10 Validasyon (Geçerleme), %10 TEST/ TEST Veri Seti Performans Sonuçları

Tablo 4.12’de detaylı performans sonuçlarının verildiği orijinal veri setinin nümerik alt kırılım veri setinde, doğruluk değeri esas alınarak karşılaştırılan k-En Yakın Komşu algoritması modelleri ikili ayırma (“*hold-out*”, k değeri default), bootstrap sampling ve üçlü ayırma yöntemi kullanılarak karşılaştırılmıştır. Optimum k parametresini elde edebilmek için 20-kat çapraz geçirme (“*cross validation*”) kullanılmıştır. Algoritmanın validasyon ve test veri seti performansları, doğruluk temel alınarak karşılaştırılmıştır. Izgara arama yöntemi- “*grid search*” kullanılarak gerçekleştirilen hiperparametre optimizasyonu sonucunda en yüksek doğruluk değeri k komşu sayısı 15 ile 0,9268 olarak elde edilmiştir. Bu nedenle k-En Yakın Komşu algoritmasının parametre ayarlamasında modelin k parametresi 15 olarak belirlenmiştir. Ardından k=15 ile oluşturulan model, test performansının bulunması için kullanılmıştır. Sonuç olarak test performansı 0,6829 doğruluk değerine sahiptir.

4.4.1.2 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular

Tablo 4. 13: K-En Yakın Komşu Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular

K-En Yakın Komşu Algoritması – Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,7436	0,7647	0,7273	0,6842	0,8000	0,722213735
%80 Eğitim %20 Test	0,7625	0,7714	0,7556	0,7105	0,8095	0,73969863
%70 Eğitim %30 Test	0,7603	0,6981	0,8088	0,7400	0,7746	0,718439608
%60 Eğitim %40 Test	0,7716	0,8028	0,7473	0,7125	0,8293	0,754959414
BOOTSTRAP SAMPLING (1:50)	0,784	0,7812	0,7860	0,7310	0,8284	0,755266764
BOOTSTRAP SAMPLING (1:100)	0,6912	0,6966	0,6872	0,6235	0,7528	0,658026059
BOOTSTRAP SAMPLING (1:200)	0,8029	0,8462	0,7716	0,7279	0,8741	0,782604638
ÜÇLÜ AYIRMA YÖNTEMİ ¹	0,9268	0,8889	0,9565	0,9412	0,9167	0,914302694
ÜÇLÜ AYIRMA YÖNTEMİ ²	0,6341	0,5556	0,6957	0,5882	0,6667	0,571435426

1-%80 Eğitim (Izgara arama-“grid search” yöntemi ile optimum komşu sayısı K=15) %10 Geçerleme (Validasyon) Veri Seti Performans Sonuçları

2-%80 EĞİTİM , %10 Validasyon (Geçerleme), %10 TEST/ TEST Veri Seti Performans Sonuçları

Tablo 4.13'te görüleceği üzere, ızgara arama yöntemi-“*grid search*” kullanılarak gerçekleştirilen hiperparametre optimizasyonu sonucunda en yüksek doğruluk değeri k komşu sayısı 15 ile 0,9268 olarak elde edilmiştir. Bu nedenle k-En Yakın Komşu algoritmasının parametre ayarlamasında modelin k parametresi 15 olarak belirlenmiştir. Ardından k=15 değeri ile parametre ayarlaması sonucunda oluşturulan model, test performansının bulunması için kullanılmıştır. Sonuç olarak test performansı 0,6341 doğruluk değerine sahiptir.

4.4.2 Sade Bayes Sınıflandırıcısından Elde Edilen Bulgular (İkili Sınıflandırma)

4.4.2.1 Orijinal Veri Setinden Edilen Bulgular

Tablo 4. 14: Sade Bayes Algoritması için Orijinal Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Sade Bayes Algoritması-Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,8462	0,7647	0,9091	0,8667	0,8333	0,812511328
%80 Eğitim %20 Test	0,8125	0,8286	0,8	0,7632	0,8571	0,794556502
%70 Eğitim %30 Test	0,8182	0,8302	0,8088	0,7719	0,8594	0,799989239
%60 Eğitim %40 Test	0,8086	0,7887	0,8242	0,7778	0,8333	0,783212078
BOOTSTRAP SAMPLING (1:50)	0,7745	0,7342	0,8037	0,7296	0,8073	0,731892772
BOOTSTRAP SAMPLING (1:100)	0,7787	0,731	0,8128	0,7361	0,8088	0,733541136
BOOTSTRAP SAMPLING (1:200)	0,81	0,7863	0,8272	0,7667	0,8428	0,776376317
ÜÇLÜ AYIRMA YÖNTEMİ ¹	0,8537	0,8889	0,8261	0,8	0,9048	0,842110249

1-%80 Eğitim, %10 Geçerleme (Validasyon), %10 Test- Validasyon Veri Seti Performans Sonuçları

Tablo 4.14'te görüleceği üzere; Sade Bayes algoritması için veri setinin bu alt kırılımındaki en yüksek performans üçlü ayırma metodundaki validasyon veri setinde elde edilmiş olup; 0,8537 doğruluk ve 0.8421 F-Ölçütü değerleri elde edilmiştir.

Tablo 4. 15: Sade Bayes Algoritması için Orijinal Veri Setinde Çapraz Geçerleme Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Sade Bayes Algoritması– Çapraz Geçerleme Performans Değerlendirme Yöntemi- K Değerleri	Ortalama Doğruluk	Ortalama Hata
K=5	0,8085818	0,1914182
K=10	0,8112195	0,1887805
K=15	0,8160494	0,1839506
K=20	0,8108333	0,1891667

Tablo 4.15’te görüleceği üzere çeşitli K değerleri ile çapraz geçerleme performans değerlendirme yöntemi kullanılarak elde edilmiş en yüksek ortalama doğruluk 0,8160494 değeri ile “K=15” katsayısı kullanılarak elde edilmiştir.

4.4.2.2 Orijinal Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinden Elde Edilen Bulgular

Tablo 4. 16: Sade Bayes Algoritması için Orijinal Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Sade Bayes Algoritması– Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,7692	0,8824	0,6818	0,6818	0,8824	0,76923708
%80 Eğitim %20 Test	0,8375	0,7429	0,9111	0,8667	0,82	0,800039053
%70 Eğitim %30 Test	0,8264	0,8491	0,8088	0,7759	0,873	0,810851311
%60 Eğitim %40 Test	0,8704	0,8873	0,8571	0,8289	0,907	0,857106363
BOOTSTRAP (1:50)	0,7447	0,8917	0,6393	0,6393	0,8917	0,744694722
BOOTSTRAP (1:100)	0,7924	0,8129	0,7783	0,7152	0,8587	0,760926746
BOOTSTRAP (1:200)	0,8185	0,8205	0,8171	0,7619	0,8645	0,790114952
ÜÇLÜ AYIRMA YÖNTEMİ	0,7805	0,7222	0,8261	0,7647	0,7917	0,742842612

1-%80 EĞİTİM , %10 Validasyon (Geçerleme), %10 Test, Validasyon Veri Seti Performans Sonuçları

Tablo 4.16’da görüleceği üzere %60 eğitim veri seti kullanılarak oluşturulmuş modelin test performansı 0,8704 ile en yüksek doğruluk değerine ulaşmıştır. Buna karşılık üçlü ayırma performans değerlendirme yöntemi sonucunda algoritmanın validasyon veri setindeki doğruluğu 0.7805 değerinde kalmıştır.

Tablo 4. 17: Sade Bayes Algoritması için Orijinal Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinde Çapraz Geçerleme Performans Değerlendirme Yöntemi Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Sade Bayes Algoritması- Çapraz Geçerleme Performans Değerlendirme Yöntemi- K Değerleri	Ortalama Doğruluk	Ortalama Hata
K=5	0,8255044	0,1744956
K=10	0,8205488	0,1794512
K=15	0,8279541	0,1720459
K=20	0,8233333	0,1766667

Tablo 4.17’de görüleceği üzere çeşitli K değerleri ile veri setinde çapraz geçerleme performans değerlendirme yöntemi kullanılarak elde edilmiş en yüksek ortalama doğruluk 0,8279541 değeri ile “K=15” katsayısı kullanılarak elde edilmiştir.

4.4.2.3 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılımından Elde Edilen Bulgular

Tablo 4. 18: Sade Bayes Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Değerlerden Oluşan Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Sade Bayes Algoritması- Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,8205	0,8824	0,7727	0,75	0,8947	0,810830679
%80 Eğitim %20 Test	0,825	0,8571	0,8	0,7692	0,878	0,810774543
%70 Eğitim %30 Test	0,8678	0,8868	0,8529	0,8246	0,9062	0,854569686
%60 Eğitim %40 Test	0,7716	0,831	0,7253	0,7024	0,8462	0,761307421
BOOTSTRAP (1:50)	0,742	0,6899	0,7798	0,6943	0,7763	0,692093007
BOOTSTRAP (1:100)	0,8012	0,7517	0,8366	0,7676	0,8244	0,759566801
BOOTSTRAP (1:200)	0,8239	0,8049	0,8385	0,792	0,8491	0,798397896
ÜÇLÜ AYIRMA YÖNTEMİ ¹	0,8537	0,8889	0,8261	0,8	0,9048	0,842110249

1-%80 EĞİTİM, %10 Validasyon (Geçerleme), %10 TEST/ Validasyon Veri Seti Performans Sonuçları

Tablo 4.18’de görüleceği üzere, %70 eğitim veri seti kullanılarak oluşturulmuş modelin test performansı 0,8678 ile en yüksek doğruluk değerine ulaşmıştır. Buna karşılık üçlü ayırma performans değerlendirme yöntemi sonucunda algoritmanın validasyon veri setindeki doğruluk 0.8537 değerinde kalmıştır.

Tablo 4. 19: Sade Bayes Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Değerlerden Oluşan Alt Kırılım Veri Setinde Çapraz Geçerleme Performans Değerlendirme Yöntemi Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Sade- Bayes Algoritması– Çapraz Geçerleme Performans Değerlendirme Yöntemi- K Değerleri	Ortalama Doğruluk	Ortalama Hata
K=5	0,8182	0,1818
K=10	0,8206	0,1794
K=15	0,8157	0,1843
K=20	0,8108	0,1892

Tablo 4.19’da görüleceği üzere çeşitli K değerleri ile veri setinde çapraz geçerleme performans değerlendirme yöntemi kullanılarak elde edilmiş en yüksek ortalama doğruluk 0.8206 değeri ile “K=10” katsayısı kullanılarak elde edilmiştir.

4.4.2.4 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinden Elde Edilen Bulgular

Tablo 4. 20: Sade Bayes Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinden Elde Edilen Bulgular

Sade Bayes Algoritması– Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
ÜÇLÜ AYIRMA YÖNTEMİ %80 EĞİTİM %10 GEÇERLEME (Validasyon) %10 TEST Validasyon Performans Sonuçları	0,8049	0,7778	0,8261	0,7778	0,8261	0,7778

Tablo 4.20’de görüleceği üzere nitelik seçimi yapılarak boyut sayısı azaltılmış veri setinin ayırıklaştırma işlemi yapılarak kategorik hâle getirilmiş alt kırılım veri setinin üçlü ayırma

metodu kullanılarak elde edilen validasyon veri setindeki performansına ilişkin en önemli bulgular 0,8049 doğruluk ve 0,7778 F-Ölçütü değerlerinin elde edilmiş olmasıdır.

4.4.3 Sınıflandırma ve Regresyon Ağaçları (“CART”) Algoritmasından Elde Edilen Bulgular (İkili Sınıflandırma)

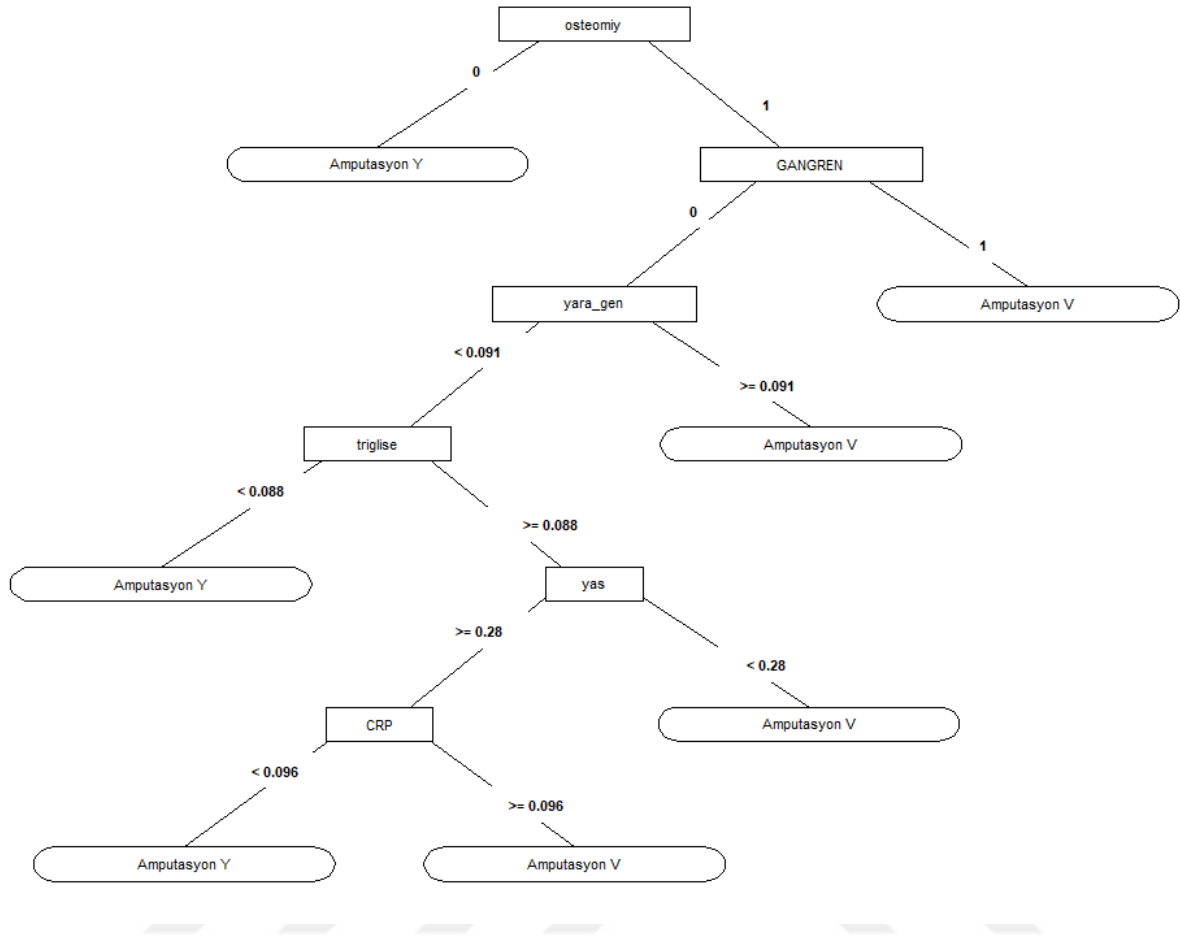
4.4.3.1 Orijinal Veri Setinden Edilen Bulgular

Tablo 4. 21: Sınıflandırma ve Regresyon Ağaçları (“CART”) Algoritması için Orijinal Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

CART Algoritması– Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,8718	0,7647	0,9545	0,9286	0,84	0,838717794
%80 Eğitim %20 Test	0,775	0,6857	0,8444	0,7742	0,7755	0,727267539
%70 Eğitim %30 Test	0,8017	0,7736	0,8235	0,7736	0,8235	0,7736
%60 Eğitim %40 Test	0,7346	0,7042	0,7582	0,6944	0,7667	0,699265666
BOOTSTRAP (1:50)	0,5544	0,6139	0,5114	0,4755	0,6474	0,535908665
BOOTSTRAP (1:100)	0,7241	0,6966	0,7438	0,6601	0,7744	0,677859011
BOOTSTRAP (1:200)	0,7419	0,7778	0,716	0,6642	0,8169	0,716525326
ÜÇLÜ AYIRMA YÖNTEMİ ¹	0,8049	0,6111	0,9565	0,9167	0,7586	0,733335999
ÜÇLÜ AYIRMA YÖNTEMİ²	0,8537	0,8889	0,8261	0,8	0,9048	0,842110249

1-%80 EĞİTİM, %10 Validasyon (Geçerleme), %10 Test/ “Budama” yapılmamış Validasyon Veri Seti Performans Sonuçları

2-%80 EĞİTİM, %10 Validasyon (Geçerleme), %10 Test/ “Budama” Yapılmış Validasyon Veri Seti Performans Sonuçları



Şekil 4. 3: Nümerik ve Kategorik Değişkenlerden Oluşan Orijinal Veri Setinde Sınıflandırma ve Regresyon Ağaçları Algoritmasından Elde Edilen “Budama Yapılmış” Karar Ağacı Yapısı

Tablo 4.21’de performans sonuçları verilen Sınıflandırma ve Regresyon Ağaçları (“CART”) algoritması için veri setinin bu alt kırılımındaki en yüksek performans ikili ayırma metodunda %90 eğitim veri seti kullanılarak oluşturulan modelin geriye kalan %10’luk test veri setinde test edilmesi sonucunda 0,8718 doğruluk değeri ile elde edilmiş olup, üçlü ayırma metodunda ise en yüksek performans elde edilen ağacın budanması sonucu validasyon veri setinde elde edilmiş olup; 0,8537 doğruluk ve 0.8421 F-Ölçütü değeri elde edilmiştir. Şekil 4.3’te ise nümerik ve kategorik değişkenlerden oluşan orijinal veri setinde Sınıflandırma ve Regresyon Ağaçları algoritmasından elde edilen “budama yapılmış” karar ağacı yapısı sunulmuştur.

Tablo 4. 22. Sınıflandırma ve Regresyon Ağaçları (“CART”) için Orijinal Veri Setinde Çapraz Geçerleme Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

CART Algoritması Geçerleme Değerlendirme Değerleri	Çapraz Performans Yöntemi- K	Ortalama Doğruluk	Ortalama Hata
K=5		0,8158988	0,8158988
K=10		0,8157317	0,8157317
K=15		0,8380071	0,8380071
K=20		0,8428571	0,8428571

Tablo 4.22’de görüleceği üzere çeşitli K değerleri ile veri setinde çapraz geçerleme performans değerlendirme yöntemi kullanılarak elde edilmiş en yüksek ortalama doğruluk 0.842871 değeri ile “K=10” katsayısı kullanılarak elde edilmiştir.

4.4.3.2 Orijinal Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinden Elde Edilen Bulgular

Tablo 4. 23: Sınıflandırma ve Regresyon Ağaçları (“CART”) Algoritması için Orijinal Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinin Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

CART Algoritması- Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,8462	0,8235	0,8636	0,8235	0,8636	0,8235
%80 Eğitim %20 Test	0,8125	0,7714	0,8444	0,7941	0,8261	0,782585423
%70 Eğitim %30 Test	0,7603	0,6981	0,8088	0,74	0,7746	0,718439608
%60 Eğitim %40 Test	0,821	0,7746	0,8571	0,8088	0,8288	0,791330656
BOOTSTRAP (1:50)	0,7947	0,8408	0,7615	0,7174	0,8691	0,774213734
BOOTSTRAP (1:100)	0,814	0,7552	0,8557	0,7883	0,8309	0,771395089
BOOTSTRAP (1:200)	0,8239	0,843	0,8098	0,7669	0,8742	0,803151376
ÜÇLÜ AYIRMA YÖNTEMİ ¹	0,7805	0,7222	0,8261	0,7647	0,7917	0,742842612
ÜÇLÜ AYIRMA YÖNTEMİ ²	0,8537	0,7778	0,913	0,875	0,84	0,823541868
ÜÇLÜ AYIRMA YÖNTEMİ³ Sonuçları	0,7805	0,7222	0,8261	0,7647	0,7917	0,742842612

1-%80 EĞİTİM, %10 Validasyon (Geçerleme), %10 Test, “Budama” yapılmamış Model- Validasyon Veri Seti Performans Sonuçları

2-%80 EĞİTİM, %10 Validasyon (Geçerleme), %10 Test, “Budama” Yapılmış Model -Validasyon Veri Seti Performans Sonuçları

3-%80 EĞİTİM, %10 Validasyon (Geçerleme), %10 Test, “Budama” Yapılmış Test Veri Seti Performans Sonuçları

Tablo 4.23'te görüleceği üzere üçlü ayırma performans değerlendirme metodu kullanılarak performansı değerlendirilen modelin validasyon veri setindeki performansı- doğruluk değeri budama yapılarak oluşturulmuş modelde, budama yapılmadan oluşturulmuş modele göre daha yüksektir (0,8537 vs 0,7805). Algoritmanın üçlü ayırma performans değerlendirme yöntemi kullanarak değerlendirilen test performansı sonucunda 0,7805 doğruluk değeri elde edilmiştir.

Tablo 4. 24: Sınıflandırma ve Regresyon Ağaçları (“CART”) için Orijinal Veri Setinde Çapraz Geçerleme Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

CART Algoritması Çapraz Geçerleme Performans Değerlendirme Yöntemi- K Değerleri	Ortalama Doğruluk	Ortalama Hata
K=5	0,8011442	0,1988558
K=10	0,8057927	0,1942073
K=15	0,7984127	0,2015873
K=20	0,8109524	0,1890476

Tablo 4.24'te görüleceği üzere çeşitli K değerleri ile veri setinde çapraz geçerleme performans değerlendirme yöntemi kullanılarak elde edilmiş en yüksek ortalama doğruluk 0.8109524 değeri ile “K=20” katsayısı kullanılarak elde edilmiştir.

4.4.3.3 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılımından Elde Edilen Bulgular

Tablo 4. 25: Sınıflandırma ve Regresyon Ağaçları (“CART”) Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılımının Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

CART Algoritması Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,8205	0,8235	0,8182	0,7778	0,8571	0,799997877
%80 Eğitim %20 Test	0,8	0,8286	0,7778	0,7436	0,8537	0,783802264
%70 Eğitim %30 Test	0,843	0,7925	0,8824	0,84	0,8451	0,815558959
%60 Eğitim %40 Test	0,7778	0,7183	0,8242	0,7612	0,7895	0,73912803
BOOTSTRAP (1:50)	0,6037	0,5063	0,6743	0,5298	0,6533	0,517783496
BOOTSTRAP (1:100)	0,536	0,8207	0,3317	0,4685	0,7204	0,596490769
BOOTSTRAP (1:200)	0,7817	0,6179	0,9068	0,8352	0,7565	0,710302223

Tablo 4.25: Sınıflandırma ve Regresyon Ağaçları (“CART”) Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılımının Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri (devam)

CART Algoritması Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
ÜÇLÜ AYIRMA YÖNTEMİ ¹	0,8537	0,7778	0,913	0,875	0,84	0,823541868

1-%80 EĞİTİM, %10 Validasyon (Geçerleme), %10 Test/ Budama Yapılmış Model Validasyon Veri Seti Performans Sonuçları

Tablo 4.25’te görüleceği üzere, Sınıflandırma ve Regresyon Ağaçları (“CART”) algoritması için veri setinin bu alt kırılımındaki en yüksek performansı üçlü ayırma metodunda %80 eğitim veri seti kullanılarak oluşturulan modelin geriye kalan %10’luk validasyon veri setinde geçerleme yapılması sonucunda elde edilmiş olup; 0,8537 doğruluk değeri elde edilmiştir.

Tablo 4 26: Sınıflandırma ve Regresyon Ağaçları (“CART”) Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılımının Çapraz Geçerleme Performans Değerlendirme Yöntemi Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

CART Algoritması Çapraz Geçerleme Performans Değerlendirme Yöntemi- K Değerleri	Ortalama Doğruluk	Ortalama Hata
K=5	0,8157	0,1843
K=10	0,8059	0,1941
K=15	0,8034	0,1966
K=20	0,8084	0,1916

Tablo 4.26’da görüleceği üzere çeşitli K değerleri ile veri setinde çapraz geçerleme performans değerlendirme yöntemi kullanılarak elde edilmiş en yüksek ortalama doğruluk 0.8157 değeri ile “K=5” katsayısı kullanılarak elde edilmiştir.

4.4.3.4 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinden Elde Edilen Bulgular

Tablo 4. 27: Sınıflandırma ve Regresyon Ağaçları (“CART”) Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış ve Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinin Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

CART Performans Değerlendirme Yöntemleri	Algoritması–Değerlendirme	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
ÜÇLÜ AYIRMA YÖNTEMİ ¹		0,8537	0,7778	0,913	0,875	0,84	0,823541868
ÜÇLÜ AYIRMA YÖNTEMİ ²		0,878	0,7778	0,9565	0,9333	0,8462	0,848484297
ÜÇLÜ AYIRMA YÖNTEMİ ³		0,7317	0,6111	0,8261	0,7333	0,7308	0,666646281

1-%80 EĞİTİM, %10 Validasyon (Geçerleme), %10 Test, Budanmamış Model Validasyon Veri Seti Performans Sonuçları

2-%80 EĞİTİM, %10 Validasyon (Geçerleme), %10 Test, Budanmış Model Validasyon Veri Seti Performans Sonuçları

3-%80 EĞİTİM, %10 Validasyon (Geçerleme), %10 Test, Budanmış Model Test Veri Seti Performans Sonuçları

Tablo 4.27’de görüleceği üzere, nitelik seçimi yapılarak boyut sayısı azaltılmış veri setinin ayırıklaştırma işlemi yapılarak kategorik hâle getirilmiş alt kırılım veri setinin üçlü ayırma metodu kullanılarak elde edilen ağacın budanmış olduğu validasyon veri setindeki performansına ilişkin en önemli bulgular 0,878 doğruluk ve 0,848484297 F-Ölçütü değerlerinin elde edilmiş olmasıdır. Daha da önemlisi modelin daha önce hiç karşılaşma olasılığının olmadığı ve ayrı olarak tutulan test verisetine ilişkin performansı 0,7317 doğruluk ve 0,666646281 F-Ölçütü değerleridir.

4.4.4 Rastgele Orman Algoritması Elde Edilen Bulgular (İkili Sınıflandırma)

4.4.4.1 Orijinal Veri Setinden Elde Edilen Bulgular

Tablo 4. 28: Rastgele Orman Algoritması için Orijinal Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Rastgele Orman-“Random Forest” Algoritması Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,8205	0,7059	0,9091	0,8571	0,8	0,774186679
%80 Eğitim %20 Test	0,85	0,8286	0,8667	0,8286	0,8667	0,8286
%70 Eğitim %30 Test	0,8595	0,8302	0,8824	0,8462	0,8696	0,838123646

Tablo 4.28: Rastgele Orman Algoritması için Orijinal Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri (devam)

Rastgele Orman- “Random Forest” Algoritması Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%60 Eğitim %40 Test	0,8333	0,7887	0,8681	0,8235	0,8404	0,805724414
BOOTSTRAP SAMPLING (1:50)	0,7003	0,8734	0,5753	0,5974	0,863	0,709503889
BOOTSTRAP SAMPLING (1:100)	0,7816	0,8	0,7685	0,7117	0,8432	0,753271152
BOOTSTRAP SAMPLING (1:200)	0,8387	0,8205	0,8519	0,8	0,8679	0,810120333
ÜÇLÜ AYIRMA YÖNTEMİ¹ VALİDASYON	0,8537	0,8333	0,8696	0,8333	0,8696	0,8333
ÜÇLÜ AYIRMA YÖNTEMİ² TEST	0,7073	0,6111	0,7826	0,6875	0,72	0,647052595

1-%80 EĞİTİM , %10 Validasyon (Geçerleme), %10 Test/ Izgara Arama-“Grid Search” yöntemi ile parametre optimizasyonu/ “mtry = 2” Validasyon Veri Seti Performans Sonuçları

2-%80 Eğitim %10 Validasyon (Geçerleme), %10 Test/ TEST Veri Seti Performans Sonuçları

Rastgele Orman Algoritmasının performansını etkileyen en önemli parametrelerden biri “mtry” olarak tanımlanan her bölmede aday olarak rastgele örneklenen değişkenlerin sayısı olarak tanımlanan parametredir.

Tablo 4.28’de görüleceği üzere, Rastgele Orman Algoritması için veri setinin bu alt kırımındaki en yüksek performans ikili ayırma metodunda %70 eğitim veri seti kullanılarak oluşturulan modelin geriye kalan %30’luk test veri setinde test edilmesi sonucunda 0,8595 doğruluk değeri ile elde edilmiştir. İkili ayırma performans değerlendirme yönteminde “mtry” için parametre optimizasyonu yapılmamıştır.

Üçlü ayırma metodunda ise ızgara arama yönteminde optimize edilmesi gereken “mtry” değeri “2” olarak tespit edilmiş ve algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,8537 doğruluk elde edilirken, test veri setindeki doğruluk 0, 7805 değeri olmuştur.

Tablo 4 29: Rastgele Orman Algoritması için Orijinal Veri Setinde Çapraz Geçerleme Performans Değerlendirme Yöntemi Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Rastgele Orman Algoritması-Çapraz Geçerleme Performans Değerlendirme Yöntemi- K Değerleri	Ortalama Doğruluk	Ortalama Hata
K=5	0,8280939	0,1719061
K=10	0,8406098	0,1593902
K=15	0,818254	0,181746
K=20	0,8286905	0,1713095

Tablo 4.29’da görüleceği üzere çeşitli K değerleri ile veri setinde çapraz geçerleme performans değerlendirme yöntemi kullanılarak ve “mtry” değeri optimize edilmeden elde edilmiş en yüksek ortalama doğruluk 0,8406098 değeri ile “K=10” katsayısı kullanılarak elde edilmiştir.

4.4.4.2 Orijinal Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinden Elde Edilen Bulgular

Tablo 4. 30: Rastgele Orman Algoritması için Orijinal Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Rastgele Orman Algoritması Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,7949	0,8824	0,7273	0,7143	0,8889	0,789501246
%80 Eğitim %20 Test	0,8375	0,8286	0,8444	0,8056	0,8636	0,816938147
%70 Eğitim %30 Test	0,8264	0,8491	0,8088	0,7759	0,873	0,810851311
%60 Eğitim %40 Test	0,8333	0,831	0,8352	0,7973	0,8636	0,813801265
BOOTSTRAP SAMPLING (1:50)	0,5904	0,9809	0,3105	0,5049	0,9577	0,66665286
BOOTSTRAP SAMPLING (1:100)	0,8118	0,7958	0,8232	0,7635	0,849	0,779315462
BOOTSTRAP SAMPLING (1:200)	0,8169	0,6585	0,9379	0,8901	0,7824	0,756981596
ÜÇLÜ AYIRMA YÖNTEMİ¹	0,8293	0,7778	0,8696	0,8235	0,8333	0,799997877

1-%80 EĞİTİM, %10 Validasyon, %10 Test/ Izgara Arama-“Grid Search” yöntemi ile parametre optimizasyonu “mtry = 4”- Validasyon Veri Seti Performans Sonuçları

Tablo 4.30’da görüleceği üzere, Rastgele Orman Algoritması için veri setinin bu alt kırılımındaki en yüksek performans ikili ayırma metodunda %80 eğitim veri seti kullanılarak

oluşturulan modelin geriye kalan %20’lik test veri setinde test edilmesi sonucunda 0,8375 doğruluk değeri ile elde edilmiştir. İkili ayırma performans değerlendirme yönteminde “*mtry*” için parametre optimizasyonu yapılmamıştır.

Üçlü ayırma metodunda ise ızgara arama yönteminde optimize edilmesi gereken “*mtry*” değeri “4” olarak tespit edilmiş ve algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,8293 doğruluk elde edilmiştir.

Tablo 4. 31: Rastgele Orman Algoritması için Orijinal Veri Setinin Ayrıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinde Çapraz Geçerleme Performans Değerlendirme Yöntemi Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Rastgele Orman Algoritması- Çapraz Geçerleme Performans Değerlendirme Yöntemi- K Değerleri	Ortalama Doğruluk	Ortalama Hata
K=5	0,8256850	0,1743150
K=10	0,8282317	0,1717683
K=15	0,8209877	0,1790123
K=20	0,8245238	0,1754762

Tablo 4.31’de görüleceği üzere çeşitli K değerleri ile veri setinde çapraz geçerleme performans değerlendirme yöntemi kullanılarak ve “*mtry*” değeri optimize edilmeden elde edilmiş en yüksek ortalama doğruluk 0,8282317 değeri ile “K=10” katsayısı kullanılarak elde edilmiştir.

4.4.4.3 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılımından Elde Edilen Bulgular

Tablo 4. 32: Rastgele Orman Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Rastgele Orman Algoritması Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,8718	0,8824	0,8636	0,8333	0,9048	0,857147427
%80 Eğitim %20 Test	0,85	0,8	0,8889	0,8485	0,8511	0,823536548
%70 Eğitim %30 Test	0,8347	0,8302	0,8382	0,8	0,8636	0,814820267
%60 Eğitim %40 Test	0,784	0,7746	0,7912	0,7432	0,8182	0,758575201
BOOTSTRAP (1:50)	0,742	0,5886	0,8532	0,744	0,741	0,657239081
BOOTSTRAP (1:100)	0,7723	0,7931	0,7574	0,7012	0,8361	0,744324058

Tablo 4. 32: Rastgele Orman Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri (devam)

Rastgele Orman Algoritması Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
BOOTSTRAP (1:200)	0,831	0,7154	0,9193	0,8713	0,8087	0,785691082
ÜÇLÜ AYIRMA YÖNTEMİ ¹ VALİDASYON	0,878	0,8333	0,913	0,8824	0,875	0,857147427
ÜÇLÜ AYIRMA YÖNTEMİ ² TEST	0,7805	0,6667	0,8696	0,8	0,7692	0,727292562

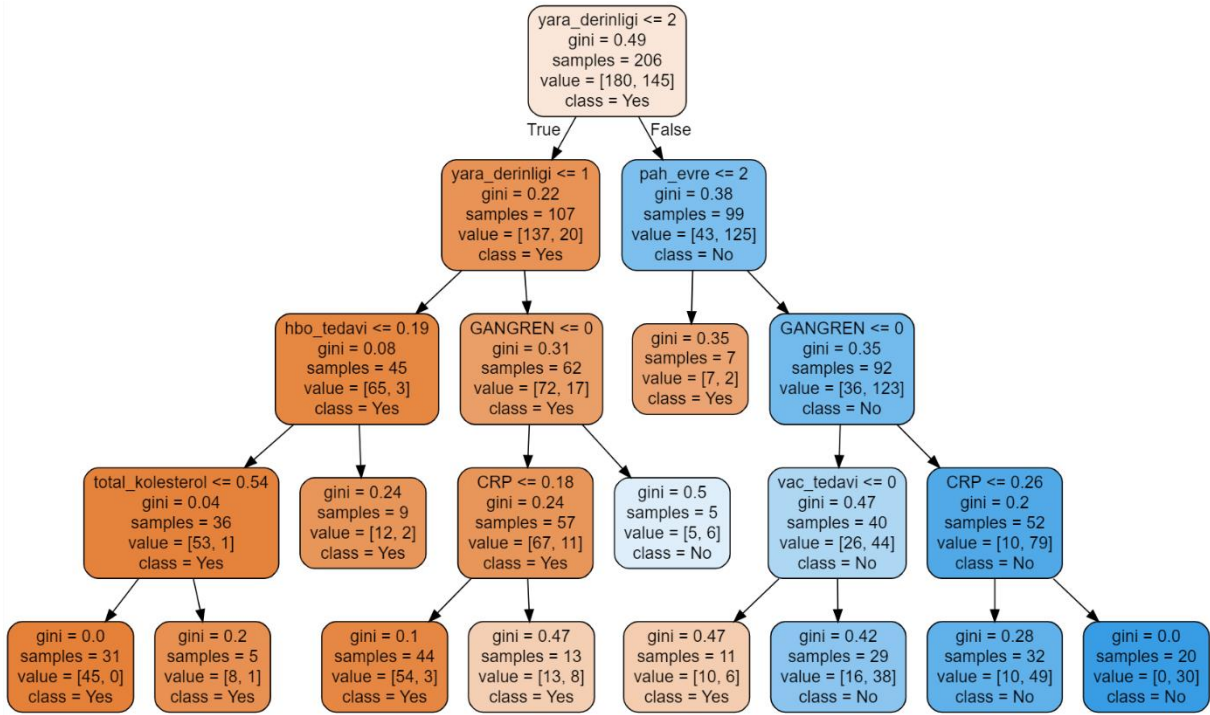
1-%80 EĞİTİM , %10 Validasyon(Geçerleme) ve İzgara Arama-“Grid Search” yöntemi ile parametre optimizasyonu sonucunda tespit edilen “mtry = 3” değeri; Validasyon Veri Seti Performans Sonuçları

2-%80 Eğitim, %10 Validasyon (Geçerleme) ve %10 Test Veri Seti Performans Sonuçları

Tablo 4.32’de görüleceği üzere, Rastgele Orman Algoritması için veri setinin bu alt kırılımındaki en yüksek performans ikili ayırma metodunda %90 eğitim veri seti kullanılarak oluşturulan modelin geriye kalan %10’luk test veri setinde test edilmesi sonucunda 0,8718 doğruluk değeri ile elde edilmiştir. İkili ayırma performans değerlendirme yönteminde “mtry” için parametre optimizasyonu yapılmamıştır.

Üçlü ayırma metodunda ise ızgara arama yönteminde optimize edilmesi gereken “mtry” değeri “3” olarak tespit edilmiş ve algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,878 doğruluk elde edilmiştir. Bununla birlikte test veri setindeki doğruluk ise 0,7805 değeri olmuştur.

Elde edilen “Rastgele Orman” ağaçları topluluğundan seçilen bir adet ağacın grafiksel olarak gösterimi Şekil 4.4’te sunulmuştur.



Şekil 4. 4: Rastgele Orman Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılım Veri Setinin %80'i Kullanılarak Oluşturulmuş Modelin Grafikselsel Gösterimi Çıktısı

Tablo 4. 33: Rastgele Orman Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılım Veri Setinde Çapraz Geçerleme Performans Değerlendirme Yöntemi Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Rastgele Orman Algoritması- Çapraz Geçerleme Performans Değerlendirme Yöntemi- K Değerleri	Ortalama Doğruluk	Ortalama Hata
K=5	0,8378	0,1622
K=10	0,8256	0,1744
K=15	0,8206	0,1794
K=20	0,8231	0,1769

Tablo 4.33'te görüleceği üzere çeşitli K değerleri ile veri setinde çapraz geçerleme performans değerlendirme yöntemi kullanılarak ve “mtry” değeri optimize edilmeden elde edilmiş en yüksek ortalama doğruluk 0,8378 değeri ile “K=5” katsayısı kullanılarak elde edilmiştir.

4.4.4.4 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinden Elde Edilen Bulgular

Tablo 4. 34: Rastgele Orman Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Ayırıklaştırma İşlemi Yapılarak Kategorik Hâle Getirilmiş Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Rastgele Orman Algoritması Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
ÜÇLÜ AYIRMA YÖNTEMİ ¹ Validasyon Veri Seti Performans Sonuçları	0,8293	0,7778	0,8696	0,8235	0,8333	0,799997877
ÜÇLÜ AYIRMA YÖNTEMİ ² Validasyon Veri Seti Performans Sonuçları	0,8537	0,7778	0,913	0,875	0,84	0,823541868

1-%80 EĞİTİM , %10 Validasyon (Geçerleme), %10 Test/ Izgara Arama – “Grid Search” Yöntemi ile parametre optimizasyonu gerçekleştirilmeden Validasyon Veri Seti Performans Sonuçları

2-%80 EĞİTİM , %10 Validasyon (Geçerleme), %10 Test/ Izgara Arama-“Grid Search” yöntemi ile parametre optimizasyonu gerçekleştirildikten sonra elde edilen “mtry = 4” değeri için Validasyon Veri Seti Performans Sonuçları

Tablo 4.34’te performans sonuçları verilen, üçlü ayırma metodu kullanılarak performansı değerlendirilen algoritmanın öncelikle “mtry” değeri optimize edilmeden validasyon performansına bakıldığında doğruluk 0,8293 olarak saptanmıştır. Sonrasında ise ızgara arama yöntemi kullanılarak optimize edilmesi gereken “mtry” değeri “4” olarak tespit edilmiş ve algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,8537 doğruluk değeri elde edilmiştir.

4.4.5 XGBOOST Algoritmasından Elde Edilen Bulgular (İkili Sınıflandırma)

4.4.5.1 Orijinal Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular

Tablo 4. 35: XGBOOST Algoritması için Orijinal Veri Setinin Nümerik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Extreme Gradient Boosting-“XGBOOST” Algoritması– Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,85	0,6923	0,9259	0,8182	0,8621	0,750003125
%80 Eğitim %20 Test	0,8272	0,7059	0,9149	0,8571	0,8113	0,774186679
%70 Eğitim %30 Test	0,8197	0,6591	0,9103	0,8056	0,8256	0,7250235
%60 Eğitim %40 Test	0,7963	0,6486	0,9205	0,8727	0,757	0,744144114
BOOTSTRAP SAMPLING (1:50)	0,7147	0,8671	0,6037	0,6143	0,8618	0,719129918
BOOTSTRAP SAMPLING (1:100)	0,7355	0,5278	0,885	0,7677	0,7224	0,625537723
BOOTSTRAP SAMPLING (1:200)	0,7122	0,5378	0,8428	0,7191	0,709	0,615374302
ÜÇLÜ AYIRMA YÖNTEMİ ¹	0,7561	0,5	0,9565	0,9	0,7097	0,642857143

1-%80 EĞİTİM, %10 Validasyon (Geçerleme), %10 Test / Validasyon Veri Seti Performans Sonuçları

Bu algoritmanın uygulanması R programlama dilinde R Studio editörü kullanılarak “*xgboost*” paketindeki “*xgboost*” kütüphanesi kullanılmak suretiyle gerçekleştirilmiştir (Chen ve diğ. 2022).

Tablo 4.35’te görüleceği üzere, XGBOOST Algoritması için veri setinin bu alt kırılımındaki en yüksek performans ikili ayırma metodunda %90 eğitim veri seti kullanılarak oluşturulan modelin geriye kalan %10’luk test veri setinde test edilmesi sonucunda 0,85 doğruluk değeri ile elde edilmiştir. İkili ayırma performans değerlendirme yönteminde parametre optimizasyonu yapılmamıştır.

Üçlü ayırma metodunda ise ızgara arama yönteminde optimize edilmesi gereken değerler [*max.dept* = 2, *eta* = 1, *nthread* = 2, *nrounds* = 2, *objective* = “*binary:logistic*”] şeklinde tespit edilmiş ve algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,7561

doğruluk elde edilmiştir. Literatürde opsiyonel olarak optimize edilebileceği belirtilen “*min_child_weight*”, “*gamma*”, “*alpha*”, “*lambda*” gibi parametreleri de bulunan algoritmanın optimize edilmesi analistin seçimine bırakıldığı için optimize edilmemiştir (Yang ve Shami, 2020).

4.4.5.2 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular

Tablo 4. 36: XGBOOST Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Extreme Gradient Boosting- “XGBOOST” Algoritması- Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,725	0,5	0,9091	0,8182	0,725	0,5
%80 Eğitim %20 Test	0,7284	0,641	0,8095	0,7576	0,7284	0,641
%70 Eğitim %30 Test	0,7705	0,6066	0,9344	0,9024	0,7705	0,6066
%60 Eğitim %40 Test	0,821	0,7	0,913	0,8596	0,821	0,7
BOOTSTRAP SAMPLING (1:50)	0,8128	0,8535	0,7834	0,7403	0,8128	0,8535
BOOTSTRAP SAMPLING (1:100)	0,7566	0,6454	0,835	0,7339	0,7566	0,6454
BOOTSTRAP SAMPLING (1:200)	0,7924	0,6218	0,9118	0,8315	0,7924	0,6218
ÜÇLÜ AYIRMA YÖNTEMİ ¹	0,7805	0,6111	0,913	0,8462	0,75	0,709686159

1-%80 EĞİTİM, %10 Validasyon (Geçerleme), %10 Test / Validasyon Veri Seti Performans Sonuçları

Tablo 4.36’da görüleceği üzere, XGBOOST Algoritması için veri setinin bu alt kırılımındaki en yüksek performans bootstrap sampling ayırma metodunda [1:50] kullanılarak oluşturulan modelin test veri setinde test edilmesi sonucunda 0,8128 doğruluk değeri ile elde edilmiştir. Bootstrap sampling performans değerlendirme yönteminde parametre optimizasyonu yapılmamıştır.

Üçlü ayırma metodunda ise ızgara arama yönteminde optimize edilmesi gereken değerler [*max.dept* = 3, *eta* = 1, *nthread* = 2, *nrounds* = 3, *objective* = “binary:logistic”] şeklinde tespit

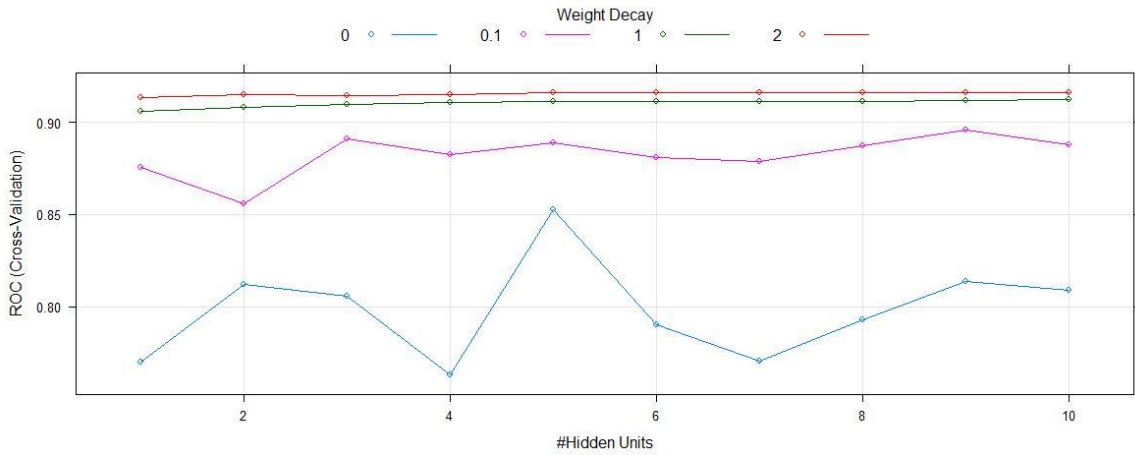
edilmiş ve algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,7805 doğruluk değeri elde edilmiştir. Literatürde opsiyonel olarak optimize edilebileceği belirtilen “min_child_weight”, “gamma”, “alpha”, “lambda” gibi parametreleri de bulunan algoritmanın optimize edilmesi analistin seçimine bırakıldığı için optimize edilmemiştir (Yang ve Shami, 2020).

4.4.6 Yapay Sinir Ağları Algoritmasından Elde Edilen Bulgular (İkili Sınıflandırma)

4.4.6.1 Orijinal Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular

R programlama dili ve R Studio ve “nnet” paketi kullanılarak tek katmanlı yapay sinir ağı kullanılmıştır (Ripley ve Venables, 2022). Katmandaki gizli nöron sayısını ifade eden “size” değeri ‘3’ ve aşırı öğrenme durumunun otomatik optimizasyonu için kullanılan “decay” değeri ‘0.1’ default değerler olarak belirlenmiştir. Üçlü ayırma yönteminde ise size ve decay parametrelerinin optimizasyonu ızgara arama yöntemi ile gerçekleştirilmiştir.

Izgara arama yönteminde en büyük değeri kullanarak optimal modeli seçmek için ROC değerleri kullanılmıştır. Model için kullanılan son değerler, “size = 9” ve “decay = 2” olarak tespit edilmiştir.



Şekil 4. 5: Izgara arama ile gerçekleştirilen parametre optimizasyonunda ROC değerlerine göre değişen “decay” ve “hidden units” değerleri

Tablo 4. 37: Yapay Sinir Ağları Algoritması için Orijinal Veri Setinin Nümerik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Yapay Sinir Ağları Algoritması– Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,7692	0,8235	0,7273	0,7	0,8421	0,756744339
%80 Eğitim %20 Test	0,8125	0,8286	0,8	0,7632	0,8571	0,794556502
%70 Eğitim %30 Test	0,7603	0,6792	0,8235	0,75	0,7671	0,712846348
%60 Eğitim %40 Test	0,7531	0,6901	0,8022	0,7313	0,7684	0,710102899
BOOTSTRAP SAMPLING (1:50)	0,8091	0,775	0,8349	0,7799	0,831	0,777442279
BOOTSTRAP SAMPLING (1:100)	0,8065	0,7589	0,84	0,7698	0,8317	0,76431114
BOOTSTRAP SAMPLING (1:200)	0,7653	0,6923	0,8187	0,7364	0,7844	0,713669378
ÜÇLÜ AYIRMA YÖNTEMİ ¹	0,8049	0,6667	0,913	0,8571	0,7778	0,750004686

1-%80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/ Validasyon Veri Seti Performans Sonuçları

Tablo 4.37’de görüldüğü üzere ikili ayırma yönteminde (*default: decay = 0.1 size= 3*) parametreleriyle %80 eğitim veri seti ile oluşturulan eğitim modeli, geriye kalan %20’lik test veri setinde test edildiğinde 0,8125 doğruluk değeri elde edilirken, üçlü ayırma yönteminde ızgara arama yöntemi ile parametre optimizasyonu gerçekleştirildiğinde “*size = 9 ve decay =2*” değerleri ile algoritmanın validasyon veri setindeki performans ölçütlerine bakıldığında doğruluk değerinin 0,8049 olduğu görülmektedir.

4.4.6.2 Çok Katmanlı Yapay Sinir Ağları Parametrelerinin Genetik Algoritma Kullanılarak Optimize Edilmesi

Python (Anaconda) Jupiter Lab – “*Scikit Learn*” kütüphanesi kullanılarak çok katmanlı perceptron modeli oluşturmak için: “*MLP-Classifer*” fonksiyonu kullanılmıştır.

Genetik Algoritma ile Parametre Optimizasyonu için: “*DEAP*” kütüphanesinden “*EvolutionaryAlgorithmSearchCV*” fonksiyonu kullanılarak gerçekleştirilmiştir (Fortin ve diğ. 2012).

Parametre optimizasyonu sonucunda;

$['alpha': 0.02, 'hidden_layer_sizes': (100, 100, 100), 'solver': 'sgd', 'activation': 'identity', 'learning_rate': 'adaptive', 'learning_rate_init': 0.1, 'max_iter': 1000]$ değerlerine ulaşılmıştır. Algoritma belirtilen değerler ile optimum sonucu üretmiş olup hiperparametre optimizasyonu sonucunda elde edilen performans değerleri Tablo 4.38’de paylaşılmıştır. Sonuç olarak genetik algoritma ile optimize edilen modelin test veri setindeki doğruluğu 0,80 olarak saptanmıştır.

Tablo 4. 38: Çok Katmanlı Yapay Sinir Ağları Parametrelerinin Genetik Algoritma Kullanılarak Optimize Edilmesi Sonucunda Elde Edilen Performans Değerlendirme Ölçütleri

Yapay Sinir Ağları “MLP” Algoritması–Performans Değerlendirme Yöntemi	Doğruluk	Duyarlılık	Kesinlik	F-Ölçütü
%80 Eğitim ve 3 Kat Çapraz Geçerleme Yapısı +Genetik Algoritma ile hiperparametre optimizasyonu %20 TEST PERFORMANS DEĞERLERİ	0.80	0.72	0.85	0.78

4.4.6.3. Çok Katmanlı Yapay Sinir Ağları Parametrelerinin Parçacık Sürü Optimizasyonu Kullanılarak Optimize Edilmesi

Python (Anaconda) Jupiter Lab IDE kullanılarak “Keras” kütüphanesi kullanılarak Yapay Sinir Ağı modeli oluşturulmuş ve Parçacık Sürü Optimizasyonu ile de parametre optimizasyonu gerçekleştirilmiştir. PSO ile Parametre Optimizasyonu için “Optunity” kütüphanesi aracılığı ile gerçekleştirilmiştir (Chollet, 2015; Claesen ve diğ. 2014).

Tablo 4. 39: Çok Katmanlı Yapay Sinir Ağları Parametrelerinin Parçacık Sürü Optimizasyonu Kullanılarak Optimize Edilmesi Sonucunda Elde Edilen Performans Değerlendirme Ölçütleri

Yapay Sinir Ağları-“Artificial Neural Networks” MLP Algoritması– MLP Algoritması– Performans Değerlendirme Yöntemi	Doğruluk	Hata
%80 Eğitim ve 3 Kat Çapraz Geçerleme Yapısı +Parçacık Sürü Optimizasyonu Algoritması ile hiperparametre optimizasyonu gerçekleştirilmesi sonucunda elde edilen %20 TEST PERFORMANS DEĞERLERİ	0,792798	0,207202

Tablo 4.39’da çok katmanlı yapay sinir ağları parametrelerinin parçacık sürü optimizasyonu kullanılarak optimize edilmesi sonucunda elde edilen performans değerlendirme ölçütleri paylaşılmıştır. Parametre optimizasyonu sonucunda; {'optimizer': 1.53955078125 (adam), 'activation': 1.4169921875, 'batch_size': 1.2158203125, 'neurons': 85.4541015625, 'epochs': 32.6123046875, 'patience': 8.56982421875} optimum değerleri üretmiş olup 0,79 doğruluk değerine ulaşılmıştır. Sonuç olarak parçacık sürü optimizasyonu ile hiperparametre optimizasyonu gerçekleştirilen modelin test veri setindeki doğruluğu 0,792798 olarak saptanmıştır.

4.4.6.4 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular

Tablo 4. 40: Yapay Sinir Ağları Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Yapay Sinir Ağları- Algoritması- Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,7949	0,8235	0,7727	0,7368	0,85	0,777741204
%80 Eğitim %20 Test	0,85	0,8571	0,8444	0,8108	0,8837	0,833307369
%70 Eğitim %30 Test	0,8099	0,7736	0,8382	0,7885	0,8261	0,780978939
%60 Eğitim %40 Test	0,8395	0,831	0,8462	0,8082	0,8652	0,819441435
BOOTSTRAP SAMPLING (1:50)	0,7093	0,775	0,6605	0,6294	0,7978	0,694652521
BOOTSTRAP SAMPLING (1:100)	0,7324	0,5931	0,8359	0,7288	0,7342	0,65398484
BOOTSTRAP SAMPLING (1:200)	0,8014	0,7295	0,8562	0,7946	0,8059	0,760659668
ÜÇLÜ AYIRMA YÖNTEMİ¹	0,8293	0,7778	0,8696	0,8235	0,8333	0,799997877

1- %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/ Izgara Arama yöntemi ile parametre optimizasyonu sonucunda elde edilen optimum değerler ("size = 10 ve decay = 2")- Validasyon Veri Seti Performans Sonuçları

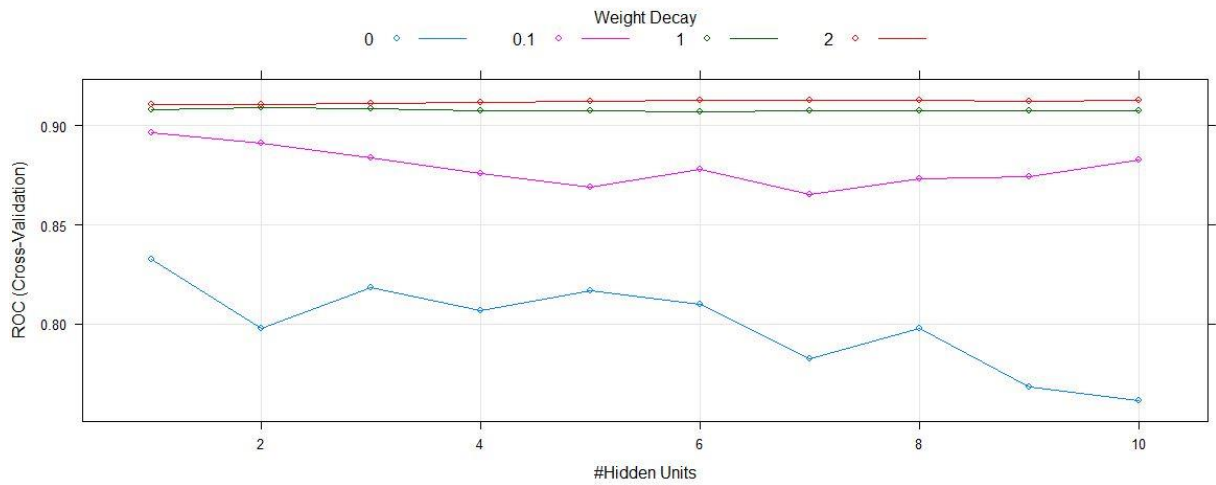
Tablo 4.40’ta performans sonuçları incelenebilecek olan ve R programlama dili ve R Studio ve “nnet” paketi kullanılarak tek katmanlı yapay sinir ağı kullanılmıştır. Katmandaki gizli nöron

sayısını ifade eden “size” değeri ‘3’ ve aşırı öğrenme durumunun engellemesi için otomatik optimizasyonun gerçekleştirilmesinde kullanılan “decay” değeri ‘0.1’ olarak belirlenmiştir. Üçlü ayırma yönteminde ise size ve decay parametrelerinin optimizasyonu ızgara arama- “grid search” yöntemi ile gerçekleştirilmiştir.

Izgara arama yönteminde en büyük değeri kullanarak optimal modeli seçmek için ROC değerleri kullanılmıştır. Model için kullanılan son değerler, “size = 10” ve “decay = 2” olarak tespit edilmiştir.

Algoritmanın performans sonuçları incelendiğinde ise, ikili ayırma %80 eğitim veri seti ile oluşturulan modelin geriye kalan %20’lik veri setinde test edilmesi sonucunda ve “decay=0.1” ve “size=3” parametreleri için en iyi performansı gösterdiği saptanmış olup 0,85 doğruluk oranına erişilmiştir.

Üçlü ayırma yöntemi kullanılarak gerçekleştirilen ve ızgara arama yöntemi kullanılarak parametre optimizasyonunun gerçekleştirilmesi suretiyle algoritmanın validasyon veri setindeki performansı 0,8293 doğruluk değerine erişmiştir.



Şekil 4. 6: Izgara arama ile gerçekleştirilen parametre optimizasyonunda ROC değerlerine göre değişen “decay” ve “hidden units” değerleri

4.4.6.4 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılım Veri Setinde Çok Katmanlı Yapay Sinir Ağları Parametrelerinin Genetik Algoritma Kullanılarak Optimize Edilmesi

Python (Anaconda) Jupiter Lab Scikit-Learn kütüphanesi kullanılarak çok katmanlı perceptron modeli oluşturmak için “*MLP-Classifier*” fonksiyonu kullanılmıştır.

Genetik Algoritma ile hiperparametre optimizasyonu için: “*EvolutionaryAlgorithmSearchCV*” fonksiyonu kullanılarak gerçekleştirilmiştir. Tablo 4.41’de çok katmanlı yapay sinir ağları parametrelerinin genetik algoritma kullanılarak optimize edilmesi sonucunda elde edilen performans değerlendirme ölçütleri paylaşılmıştır.

Tablo 4. 41: Çok Katmanlı Yapay Sinir Ağları Parametrelerinin Genetik Algoritma Kullanılarak Optimize Edilmesi Sonucunda Elde Edilen Performans Değerlendirme Ölçütleri

Yapay Sinir Ağları-MLP Algoritması- Performans Değerlendirme Yöntemi	Doğruluk	Duyarlılık	Kesinlik	F-Ölçütü
%80 Eğitim ve 3 Kat Çapraz Geçerleme Yapısı +Genetik Algoritma ile hiperparametre optimizasyonu %20 TEST VERİ SETİ PERFORMANS DEĞERLERİ	0,76	0,72	0,80	0,76

Hiperparametre optimizasyonu sonucunda;

Best individual is: {'alpha': 0.005, 'hidden_layer_sizes': (5, 3), 'solver': 'adam', 'activation': 'identity', 'learning_rate': 'invscaling', 'learning_rate_init': 0.02, 'max_iter': 100} with fitness: 0.8307692307692308

değerleri optimum sonucu üretmiş olup sonuç olarak genetik algoritma kullanılarak hiperparametre optimizasyonu gerçekleştirilen modelin test veri setindeki doğruluğu 0,76 olarak saptanmıştır.

4.4.6.5 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılım Veri Setinde Çok Katmanlı Yapay Sinir Ağları Parametrelerinin Parçacık Sürü Optimizasyonu Kullanılarak Optimize Edilmesi

Python (Anaconda) Jupiter Lab IDE kullanılarak Keras kullanılarak Yapay Sinir Ağı modeli oluşturulmuş ve Parçacık Sürü Optimizasyonu ile de hiperparametre optimizasyonu gerçekleştirilmiştir. Parçacık Sürü Optimizasyonu ile hiperparametre optimizasyonu için “Optunity” kütüphanesi aracılığı ile gerçekleştirilmiştir (Claesen ve diğ. 2014).

Tablo 4.42’de nitelik seçimi yapılarak boyut sayısı azaltılmış veri setinin nümerik alt kırılım veri setinde çok katmanlı yapay sinir ağları parametrelerinin parçacık sürü optimizasyonu kullanılarak optimize edilmesi sonucu elde edilen doğruluk ve hata değerleri paylaşılmıştır.

Tablo 4. 42: Çok Katmanlı Yapay Sinir Ağları Parametrelerinin Parçacık Sürü Optimizasyonu Kullanılarak Optimize Edilmesi Sonucunda Elde Edilen Performans Değerlendirme Ölçütleri

Yapay Sinir Ağları-MLP Algoritması-Performans Değerlendirme Yöntemi	Doğruluk	Hata
%80 Eğitim ve 3 Kat Çapraz Geçerleme Yapısı +Parçacık Sürü Optimizasyonu Algoritması ile hiperparametre optimizasyonu gerçekleştirilmesi sonucunda elde edilen %20 TEST PERFORMANS DEĞERLERİ	0,8035584604212055	0,196441539578945

Parçacık sürü optimizasyonu kullanılarak gerçekleştirilen hiperparametre optimizasyonu sonucunda;

```
{'optimizer': 0.9026367187500002, 'activation': 1.4978515625, 'batch_size': 0.9244009270521549, 'neurons': 58.1552734375, 'epochs': 43.6103515625, 'patience': 9.17744140625}
```

değerleri optimum sonucu üretmiş ve Tablo 4.42’de görüleceği üzere 0,8035584604212055 doğruluk değerine ulaşılmıştır.

4.4.7 Aşırı Öğrenme Makineleri Algoritmasından Elde Edilen Bulgular (İkili Sınıflandırma)

4.4.7.1 Orijinal Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular

Aşırı Öğrenme Makineleri –“*Extreme Learning Machines (ELM)*” Algoritmasının uygulanması için R Programlam Dili ve R Studio Editörü tercih edilmiştir.

Bu kapsamda, “*elmNNRcpp*” paketi ve kütüphanesi ile çalışılmıştır (Mouselimis ve Gosso, 2022).

İlgili kütüphanede;

Gizli katman sayısı-“*nhid*” : 200

Aktivasyon Fonksiyonu: “*purelin*”

Girdi ağırlıklarının ve önyargı-“*bias*” ağırlıklarının başlatılacağı dağılımı belirten karakter dizisi – “*init_weights*”: “*uniform_negative*”

Önyargı-“*bias*” ağırlıkları (Eğer “*TRUE*” olarak kodlanır ise, gizli katmana sapma ağırlıkları da eklenmektedir) : “*TRUE*”

Eşik değeri: “0.7” değeri olarak belirlenmiştir. Bu eşik değerin dayanağı ise Koçoğlu (2017) tarafından gerçekleştirilen doktora tez çalışmasıdır. Bu çalışmada ELM algoritmasının optimum sonucu “0.7” eşik değerinde ürettiği tespit edilmiştir.

Bu parametreler kullanılarak algoritmanın performans değerlendirme yöntemleri ve sonuçları Tablo 4.43’te sunulmuştur.

Tablo 4. 43: Aşırı Öğrenme Makineleri Algoritması için Orijinal Veri Setinin Nümerik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Aşırı Öğrenme Makineleri Algoritması– Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,875	0,8889	0,8636	0,8421	0,9048	0,864867348
%80 Eğitim %20 Test	0,7407	0,3	1	1	0,7083	0,461538462
%70 Eğitim %30 Test	0,6967	0,3208	0,9855	0,9444	0,6538	0,478917989
%60 Eğitim %40 Test	0,6235	0,10294	1	1	0,60645	0,186664732
BOOTSTRAP SAMPLING (1:50)	0,6944	0,271	0,9954	0,9767	0,6576	0,424277791
BOOTSTRAP SAMPLING (1:100)	0,5447	0,06944	1	0,29412	0,57188	0,112353905
BOOTSTRAP SAMPLING (1:200)	0,6655	0,2562	0,975	0,8857	0,6341	0,397436448
ÜÇLÜ AYIRMA YÖNTEMİ ¹	0,8293	0,6667	0,9565	0,9231	0,7857	0,774224141

1-%80 Eğitim, %10 Validasyon (Geçerleme), %10 Test, Validasyon Veri Seti Performans Sonuçları

Tablo 4.43'te görüleceği üzere belirtilen parametrelerin seçildiği deney koşullarında en yüksek performansı ikili ayırma yönteminde %90 eğitim veri setinin kullanılması suretiyle oluşturulan modelin %10'luk test veri setindeki performansının değerlendirilmesi neticesinde 0,875 doğruluk değerine erişilmiştir. Üçlü ayırma yönteminde ise algoritmanın validasyon performansı 0,8293 doğruluk değerine ulaşmıştır.

4.4.7.2 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular

Aşırı Öğrenme Makineleri –“*Extreme Learning Machines (ELM)*” Algoritmasının uygulanması için R Programlam Dili ve R Studio Editörü tercih edilmiştir. Bu kapsamda, “*elmNNRcpp*” paketi ve kütüphanesi ile çalışılmıştır (Mouselimis ve Gosso, 2022).

İlgili kütüphanede

Gizli katman sayısı- “*nhid*” : 200

Aktivasyon Fonksiyonu: “*purelin*”

Girdi ağırlıklarının ve önyargı-“bias” ağırlıklarının başlatılacağı dağılımı belirten karakter dizisi – “*init_weights*”: “*uniform_negative*”

Önyargı-“*bias*” ağırlıkları (Eğer “*TRUE*” olarak kodlanır ise, gizli katmana sapma ağırlıkları da eklenmektedir) : “*TRUE*”

Eşik değeri: “0.7” değeri olarak belirlenmiştir. Bu eşik değerin dayanağı ise Koçoğlu (2017) tarafından gerçekleştirilen doktora tezidir. Bu çalışmada ELM algoritmasının optimum sonucu “0.7” eşik değerinde ürettiği tespit edilmiştir.

Tablo 4. 44: Aşırı Öğrenme Makineleri Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Aşırı Öğrenme Makineleri Algoritması– Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,825	0,6111	1	1	0,7586	0,758612128
%80 Eğitim %20 Test	0,775	0,5	0,9767	0,95	0,6885	0,655172414
%70 Eğitim %30 Test	0,6885	0,3793	0,9688	0,9167	0,6327	0,536580725
%60 Eğitim %40 Test	0,6975	0,4051	0,9759	0,9412	0,6328	0,566411825
BOOTSTRAP SAMPLING (1:50)	0,6233	0,13924	0,9726	0,78571	0,61032	0,236558215
BOOTSTRAP SAMPLING (1:100)	0,6991	0,3285	0,9505	0,8182	0,6761	0,468786431
BOOTSTRAP SAMPLING (1:200)	0,7367	0,418	0,9811	0,9444	0,6872	0,579505578
ÜÇLÜ AYIRMA YÖNTEMİ ¹	0,7805	0,5	1	1	0,7188	0,666666667

1-%80 Eğitim, %10 Validasyon (Geçerleme), %10 Test / Validasyon Veri Seti Performans Sonuçları

Tablo 4.44’te görüleceği üzere belirtilen parametrelerin seçildiği deney koşullarında en yüksek performansı ikili ayırma yönteminde %90 eğitim veri setinin kullanılması suretiyle oluşturulan modelin %10’luk test veri setindeki performansı 0,825 doğruluk değerine erişmiştir. Üçlü ayırma yönteminde ise algoritmanın validasyon performansı 0,7805 doğruluk değerine ulaşmıştır.

4.4.8 Destek Vektör Makineleri Algoritmasından Elde Edilen Bulgular (İkili Sınıflandırma)

4.4.8.1 Orijinal Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular

R Programlama dili ve R Studio editörü kullanılarak gerçekleştirilen analizlerde “*e1071*” paketindeki “*svm*” fonksiyonu kullanılmış olup öncelikle algoritmanın “*default*” olarak tanımlı parametreleri ile analizler gerçekleştirilmiştir.

Üçlü ayırma yönteminde ise ızgara arama-“*grid-search*” yöntemi ile “*C*” değeri için hiperparametre optimizasyonu gerçekleştirilmiştir. En uygun modeli seçmek için doğruluk parametresi esas alınmış olup en yüksek doğruluk değerinin elde edilmesini sağlayan *C* değeri ‘0,5’ olarak saptanmıştır. Yine optimize edilmesi gereken kernel “*Radial*” olarak tespit edilmiştir.

C – “*Cost*” sabiti: Lagrange formülasyonundaki sabit terim olan ‘*C*’-sabitidir.

Tablo 4. 45: Destek Vektör Makineleri Algoritması için Orijinal Veri Setinin Nümerik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Destek Vektör Makineleri Algoritması– Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,8462	0,8824	0,8182	0,7895	0,9	0,833368981
%80 Eğitim %20 Test	0,8125	0,8571	0,7778	0,75	0,875	0,799981333
%70 Eğitim %30 Test	0,8182	0,9057	0,75	0,7385	0,9107	0,81359865
%60 Eğitim %40 Test	0,7469	0,8451	0,6703	0,6667	0,8472	0,745373952
BOOTSTRAP SAMPLING (1:50)	0,8043	0,8805	0,7477	0,7216	0,8939	0,793169964
BOOTSTRAP SAMPLING (1:100)	0,8592	0,7786	0,9154	0,8651	0,8558	0,819573961
BOOTSTRAP SAMPLING (1:200)	0,8014	0,887	0,7407	0,7083	0,9023	0,787641321
ÜÇLÜ AYIRMA YÖNTEMİ ¹	0,8293	0,7778	0,8696	0,8235	0,8333	0,799997877

1-%80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/ Izgara Arama – “Grid Search” yöntemi ile parametre optimizasyonu- Validasyon Veri Seti Performans Sonuçları

Tablo 4.45’te görüleceği üzere belirtilen parametrelerin seçildiği deney koşullarında en yüksek performans bootstrap sampling [1:100] kullanılması suretiyle 0,85925 doğruluk değerine erişmiştir. Üçlü ayırma yönteminde ise algoritmanın ızgara arama metodu ile gerçekleştirilen hiperparametre optimizasyonu sonucunda validasyon performansı 0,8293 doğruluk değerine ulaşmıştır.

4.4.8.2 DVM Algoritmasının Hiperparametrelerinin Genetik Algoritma Kullanılarak Optimize Edilmesi

Python (Anaconda) Jupiter Lab Scikit-Learn kullanılarak model oluşturmak için: “*SVC()*” fonksiyonu kullanılmış olup

Genetik Algoritma ile Parametre Optimizasyonu için: “*EvolutionaryAlgorithmSearchCV*” fonksiyonu kullanılarak gerçekleştirilmiştir.

Parametre optimizasyonu sonucunda;

“*Best individual is: {'C': 29.265739247500335, 'kernel': 'linear'} with fitness: 0.7876923076923077*”

Değerleri optimum sonucu üretmiş olup hiperparametre optimizasyonu sonucunda elde edilen performans değerleri Tablo 4.46’da paylaşılmıştır. Tablo 4.46’da performans değerlendirme yöntem detayları verilen ve karma bir yapı olan eğitim veri setinde genetik algoritma ve ‘3’ kat çapraz geçiş yönteminin kullanılması suretiyle hiperparametre optimizasyonu gerçekleştirilen DVM algoritmasının doğruluk değeri 0,72 olarak bulunmuştur.

Tablo 4. 46: Destek Vektör Makineleri Algoritmasının Genetik Algoritma Kullanılarak Optimize Edilmesi Sonucunda Elde Edilen Performans Değerlendirme Ölçütleri

Destek Vektör Makineleri – “DVM Performans Değerlendirme Yöntemi	Doğruluk	Duyarlılık	Kesinlik	F-Ölçütü
%80 Eğitim ve 3 Kat Çapraz Geçerleme Yapısı +Genetik Algoritma ile hiperparametre optimizasyonu %20 Test TEST/ Performans Sonuçları	0,72	0,62	0,75	0,68

4.4.8.3 DVM Algoritmasının Hiperparametrelerinin Parçacık Sürü Optimizasyonu Kullanılarak Optimize Edilmesi

Python (Anaconda) Jupiter Lab Scikit-Learn kütüphanesi kullanılarak DVM sınıflandırma modeli oluşturulmuş ve Parçacık Sürü Optimizasyonu ile de parametre optimizasyonu gerçekleştirilmiştir. PSO ile Parametre Optimizasyonu için “Optunity” kütüphanesi aracılığı ile gerçekleştirilmiştir.

Parametre optimizasyonu sonucunda;

{‘C’: 2.5390625, ‘kernel’: 2.921875 (‘poly’)} değerlerinin optimum değerler olduğu tespit edilmiştir.

Tablo 4. 47: Destek Vektör Makineleri Algoritmasının Parçacık Sürü Optimizasyonu Kullanılarak Optimize Edilmesi Sonucunda Elde Edilen Performans Değerlendirme Ölçütleri

Destek Vektör Makineleri Algoritması- Performans Değerlendirme Yöntemi	Doğruluk	Hata
%80 Eğitim – 3 kat Çapraz Geçerleme + PSO ile hiperparametre optimizasyonu %20 TEST / Test Performans Sonuçları	0,7813725490196078	0,2186274509803922

Tablo 4.47’de performans değerlendirme yöntem detayları verilen ve karma bir yapı ile performansı değerlendirilen, eğitim veri setinde parçacık sürü optimizasyonu algoritması ve ‘3’ kat çapraz geçerleme metodolojisi kullanılmak suretiyle hiperparametre optimizasyonu gerçekleştirilen PSO algoritmasının %20 test veri setindeki doğruluk değeri 0,7813 olarak bulunmuştur.

4.4.8.4 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılımından Elde Edilen Bulgular

R Programlama dili ve R Studio editörü kullanılarak gerçekleştirilen analizlerde “e1071” paketindeki “svm” fonksiyonu kullanılmış olup öncelikle algoritmanın “default” olarak tanımlı parametreleri ile analizler gerçekleştirilmiştir.

Üçlü ayırma yönteminde ise ızgara arama-“grid-search” yöntemi ile “C” değeri için hiperparametre optimizasyonu gerçekleştirilmiştir. En uygun modeli seçmek için doğruluk parametresi esas alınmış olup en yüksek doğruluk değerinin elde edilmesini sağlayan C değeri ‘0,25’ olarak saptanmıştır.

C – “Cost” sabiti: Lagrange formülasyonundaki sabit terim olan 'C'-sabitidir.

Tablo 4. 48: DVM Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Destek Vektör Makineleri Algoritması– Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,7692	0,9412	0,6364	0,6667	0,9333	0,780518739
%80 Eğitim %20 Test	0,8524	0,9714	0,7333	0,7391	0,9706	0,839475873
%70 Eğitim %30 Test	0,8182	0,8302	0,8088	0,7719	0,8594	0,799989239
%60 Eğitim %40 Test	0,8457	0,7183	0,9451	0,9107	0,8113	0,803137888
BOOTSTRAP (1:50)	0,7893	0,9125	0,6977	0,6919	0,9146	0,787034094
BOOTSTRAP (1:100)	0,7824	0,5172	0,9795	0,9494	0,7318	0,669616364
BOOTSTRAP (1:200)	0,8208	0,8718	0,784	0,7445	0,8944	0,803136918
ÜÇLÜ AYIRMA YÖNTEMİ¹ Validasyon Performans Sonuçları	0,8537	0,8333	0,8696	0,8333	0,8696	0,8333

1-%80 Eğitim, İzgara Arama yöntemi ile parametre optimizasyonu ve %10 Validasyon (Geçerleme) Veri Seti Performans Sonuçları

Tablo 4.48’de görüleceği üzere belirtilen parametrelerin seçildiği deney koşullarında en yüksek performans ikili ayırma yönteminde %60 eğitim veri setiyle oluşturulan modelin geriye kalan %40’lık test veri setindeki test performansında elde edilmiş ve en yüksek sonucu üretmiş olup detaylı olarak incelendiğinde doğruluk değeri 0,8457 olarak saptanmıştır. Buna karşılık üçlü ayırma metodunda ızgara arama yönteminin kullanılması sonucunda validasyon veri setindeki performansı ise 0,8537 doğruluk değerine erişmiştir.

4.4.8.5 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılımında DVM Algoritmasının Hiperparametrelerinin Genetik Algoritma Kullanılarak Optimize Edilmesi

Python (Anaconda) Jupiter Lab Scikit-Learn kütüphanesi kullanılarak model oluşturmak için: “SVC()” fonksiyonu kullanılmış olup

Genetik Algoritma ile Parametre Optimizasyonu için: “*EvolutionaryAlgorithmSearchCV*” fonksiyonu kullanılarak gerçekleştirilmiştir.

Parametre optimizasyonu sonucunda;

Best individual is: {'alpha': 0.005, 'hidden_layer_sizes': (5, 3), 'solver': 'adam', 'activation': 'identity', 'learning_rate': 'invscaling', 'learning_rate_init': 0.02, 'max_iter': 100} with fitness: 0.8307692307692308

değerleri optimum sonucu üretmiş olup hiperparametre optimizasyonu sonucunda elde edilen performans değerleri Tablo 4.49’da paylaşılmıştır. Tablo 4.49’da görüleceği üzere karma bir yapı metodolojisi ile performansı değerlendirilen ve eğitim veri setinde genetik algoritmanın ‘3’ kat çapraz geçirme metodolojisi ile hiperparametre optimizasyonu gerçekleştirilen DVM algoritmasının doğruluk değeri 0,78 olarak bulunmuştur.

Tablo 4. 49: Destek Vektör Makineleri Algoritmasının, Genetik Algoritma Kullanılarak Optimize Edilmesi Sonucunda Elde Edilen Performans Değerlendirme Ölçütleri

Destek Vektör Makineleri Algoritması Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Kesinlik	F-Ölçütü
%80 Eğitim + 3 Kat Çapraz Geçerleme Genetik Algoritma ile hiperparametre optimizasyonu %20 Test/ TEST Performans Sonuçları	0,78	0,72	0,80	0,76

4.4.8.6 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik Alt Kırılımında DVM Algoritmasının Hiperparametrelerinin Parçacık Sürü Optimizasyonu Kullanılarak Optimize Edilmesi

Python (Anaconda) Jupiter Lab *Scikit-Learn* kütüphanesi “*SVC()*” fonksiyonu kullanılarak DVM sınıflandırma modeli oluşturulmuş ve Parçacık Sürü Optimizasyonu ile de parametre optimizasyonu gerçekleştirilmiştir. PSO ile parametre optimizasyonu için “*Optunity*” kütüphanesi aracılığı ile gerçekleştirilmiştir.

Parametre optimizasyonu sonucunda;

{'C': 41.259765625, 'kernel': 0.58984375 (linear)}

değerlerinin optimum değerler olduğu tespit edilmiştir. Tablo 4.50’de görüleceği üzere karma bir yapı ile performansı değerlendirilen ve eğitim veri setinde parçacık sürü algoritmasının ‘3’ kat çapraz geçişleme ile kullanılması suretiyle hiperparametre optimizasyonu gerçekleştirilen DVM algoritmasının doğruluk değeri 0,80 olarak bulunmuştur.

Tablo 4. 50: Destek Vektör Makineleri Algoritmasının, Parçacık Sürü Optimizasyonu Kullanılarak Optimize Edilmesi Sonucunda Elde Edilen Performans Değerlendirme Ölçütleri

Destek Vektör Makineleri Algoritması- Performans Değerlendirme Yöntemi	Doğruluk	Hata
%80 Eğitim + 3 Kat Çapraz Geçişleme PSO ile hiperparametre optimizasyonu %20 Test / TEST Veri Seti Performans Sonuçları	0,8059549745824256	0,1940450254175744

4.4.9 Lojistik Regresyon Analizinden Elde Edilen Bulgular (İkili –“Binary” Sınıflandırma)

4.4.9.1 Orijinal Veri Setinden Edilen Bulgular

Tablo 4.51: Lojistik Regresyon için Orijinal Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Lojistik Regresyon Algoritması– Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Keskinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,8205	0,7647	0,8636	0,8125	0,8261	0,787875666
%80 Eğitim %20 Test	0,8	0,6857	0,8889	0,8276	0,7843	0,749997119
%70 Eğitim %30 Test	0,7934	0,8868	0,7206	0,7121	0,8909	0,78990591
%60 Eğitim %40 Test	0,8457	0,8451	0,8462	0,8108	0,875	0,827594758
BOOTSTRAP SAMPLING (1:50)	0,8683	0,8846	0,8565	0,8166	0,9113	0,849240959
BOOTSTRAP SAMPLING (1:100)	1	1	1	1	1	1
BOOTSTRAP SAMPLING (1:200)	1	1	1	1	1	1
ÜÇLÜ AYIRMA YÖNTEMİ¹	0,8293	0,7778	0,8696	0,8235	0,8333	0,799997877

1-%80 EĞİTİM, %10 Validasyon (Geçerleme), %10 Test / Validasyon Veri Seti Performans Sonuçları

Tablo 4.51’de görüleceği üzere Lojistik Regresyon algoritması için orijinal veri setinde çeşitli performans değerlendirme yöntemleri kullanılarak elde edilen performans değerlendirme ölçütleri sunulmuştur. Bu kapsamda en yüksek doğruluk performansı 0,8683 değeri ile Bootstrap Sampling (1:50) oranı için elde edilirken, veri setinin 1:100 ve 1:200 oranlarındaki ayrılmalarda ise “*overfitting*” yani modelin veri setini ezberleme problemi ortaya çıkmıştır.

Tablo 4. 52: Lojistik Regresyon Algoritması için Orijinal Veri Setinde Çapraz Geçerleme Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Lojistik Regresyon Analizi Çapraz Geçerleme Değerlendirme Yöntemi- K Değerleri	Ortalama Doğruluk	Ortalama Hata
K=5	0,7985	0,2015
K=10	0,8059	0,1941
K=15	0,7715	0,2285
K=20	0,7862	0,2138

Tablo 4.52’de Lojistik Regresyon algoritması için orijinal veri setinde çapraz geçerleme performans değerlendirme yöntemleri kullanılarak elde edilen performans değerlendirme ölçütleri sunulmuştur. Çapraz geçerleme yöntemi kullanılarak elde edilen en yüksek performans, aynı zamanda tipik olarak kullanılan 10 kat çapraz geçerleme yönteminde 0,8059 doğruluk değeri ile elde edilmiştir.

4.4.9.2 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılımından Elde Edilen Bulgular

Tablo 4. 53: Lojistik Regresyon Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Lojistik Regresyon Analizi Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
%90 Eğitim %10 Test	0,9	0,9412	0,8636	0,8421	0,95	0,88889645
%80 Eğitim %20 Test	0,825	0,7714	0,8667	0,8182	0,8298	0,794111072
%70 Eğitim %30 Test	0,8264	0,8491	0,8088	0,7759	0,873	0,810851311
%60 Eğitim %40 Test	0,7901	0,7324	0,8352	0,7761	0,8	0,753617024
BOOTSTRAP SAMPLING (1:50)	1	1	1	1	1	1

Tablo 4. 53: Lojistik Regresyon Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılım Veri Setinde Çeşitli Performans Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri (devam)

Lojistik Regresyon Analizi Performans Değerlendirme Yöntemleri	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
BOOTSTRAP SAMPLING (1:100)	1	1	1	1	1	1
BOOTSTRAP SAMPLING (1:200)	1	1	1	1	1	1
ÜÇLÜ AYIRMA YÖNTEMİ	0,8293	0,7778	0,8696	0,8235	0,8333	0,799997877

1-%80 EĞİTİM, %10 Validasyon (Geçerleme), %10 Test / Validasyon Veri Seti Performans Sonuçları

Tablo 4.53'te Lojistik Regresyon Algoritması için nitelik seçimi yapılarak boyut sayısı azaltılmış veri setinin nümerik ve kategorik alt kırılım veri setinde çeşitli performans değerlendirme yöntemleri kullanılarak elde edilen performans değerlendirme ölçütleri sunulmuştur. Tablo 4.53'te görüldüğü üzere en yüksek test performansı 0,90 doğruluk değeri ile veri setinin %90'ından oluşan eğitim veri seti kullanılarak oluşturulmuş makine öğrenmesi modeli ile elde edilmiştir. Bununla birlikte Bootstrap Sampling performans değerlendirme yönteminde yine “*overfitting*” yani modelin veri setini ezberleme probleminin ortaya çıkması dikkat çekicidir.

```

Call:
glm(formula = ikili_karar ~ ., family = "binomial", data = training)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.57627 -0.00001  0.00000  0.00000  2.05224

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -31.02897   48196.14604  -0.001  0.9995
gecmisteki_amputasyon1 -2.94768     1.47423  -1.999  0.0456 *
gecmisteki_amputasyon2 -14.99258  158248.39253   0.000  0.9999
bobrek_yetmezligi1    12.84533   5831.75659   0.002  0.9982
diyaliz1        -14.11803   5831.75696  -0.002  0.9981
pah_evre2        -1.54651     1.50095  -1.030  0.3028
pah_evre3         0.83847     1.63115   0.514  0.6072
yara_derinligi2    -33.92016   4426.13174  -0.008  0.9939
yara_derinligi3     -5.75675     6.02158  -0.956  0.3391
enfeksiyon2       -12.92345   48987.96803   0.000  0.9998
enfeksiyon3         7.20677   50029.30461   0.000  0.9999
enfeksiyon4        -7.24987   50156.42602   0.000  0.9999
osteomiyelit1      3.32082  157418.54639   0.000  1.0000
lökosit           0.52248     0.32464   1.609  0.1075
CRP              -0.03470     0.01765  -1.966  0.0493 *
sedim             0.08524     0.03415   2.496  0.0126 *
albumin           1.25624     1.01005   1.244  0.2136
PNL              -0.60598     0.37373  -1.621  0.1049
kreatinin        -0.06707     0.73360  -0.091  0.9272
total_kolesterol   0.03010     0.03024   0.995  0.3196
LDL              -0.04590     0.03747  -1.225  0.2206
GANGREN1         30.91029   4426.12754   0.007  0.9944
wagner_grade2      8.93602   13418.75045   0.001  0.9995
wagner_grade3     18.28763  158002.09423   0.000  0.9999
wagner_grade4     35.82011  158095.90996   0.000  0.9998
wagner_grade5     32.97573  159421.39233   0.000  0.9998
hbo_durumu1       -5.28730     2.40497  -2.198  0.0279 *
hbo_tedavi        0.07232     0.03203   2.258  0.0239 *
vac_tedavi1       4.35483     1.73790   2.506  0.0122 *
debridman1       -2.21941     1.51143  -1.468  0.1420
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 185.003  on 134  degrees of freedom
Residual deviance: 36.263  on 105  degrees of freedom
(233 observations deleted due to missingness)
AIC: 96.263

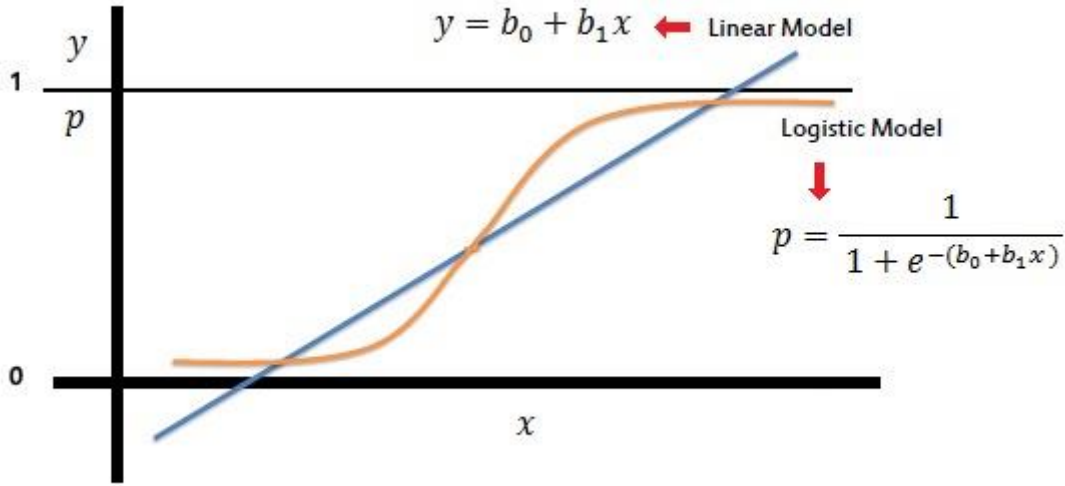
Number of Fisher Scoring iterations: 21

```

Şekil 4. 7: Lojistik Regresyon Analizi Modelinin R Programlama Dilinde Alınmış Ekran Görüntüsü

Şekil 4.7’de görüldüğü üzere; geçmişteki minör amputasyon geçirme durumu, CRP(mg/L)sedimentasyon, HBOT tedavisi alıp almama durumu, HBOT tedavisi seans sayısı ve NPWT tedavisi görüp görmeme durumu niteliklerinin β katsayıları istatistiksel olarak anlamlıdır.

Şekil 4.8’de tekrar hatırlatma yapılacak olursa; Lojistik modelin linear regresyondan en büyük farkı hedef değişkenin kategorik olması (linear regresyonda sayısal bir değer tahmin edilmeye çalışılır) ve bunun için de sigmoid fonksiyonu ile dönüşüm yapılması idi.



Şekil 4. 8: Linear Model vs. Lojistik Model

Veri setlerinin çeşitli alt kırılım senoryolarının uygun makine öğrenmesi algoritmalarının denenmesi sonucunda, nitelik seçimi sonucunda boyut sayısı azalmış, nümerik ve kategorik değerlerden oluşan veri setinin alt kırılımında en iyi modellerden biri olarak görünen Lojistik Regresyon algoritmasına ait Şekil 4.7'in yardımıyla yazılmış ve yeni bir örneğin sınıfının tahmin edilmesinde kullanılabilecek olan ikili Lojistik Regresyon denklemi şu şekilde olmaktadır:

Bütün niteliklerin yer aldığı Lojistik Regresyon fonksiyonu:

$$\begin{aligned} \text{logit } \pi(x) = g(x) = & -31.02897 - 2.94768\beta_{\text{gecmisteki amputasyon1}} + \\ & -14.99258\beta_{\text{gecmisteki amputasyon2}} + 12.84533\beta_{\text{bobrek yetmezligi}} + -14.11803\beta_{\text{diyaliz1}} + \\ & -1.54651\beta_{\text{pah evre2}} + 0.83847\beta_{\text{pah evre3}} + -33.92016\beta_{\text{yara derinligiz}} + \\ & -5.75675\beta_{\text{yara derinligiz3}} + -12.92345\beta_{\text{enfeksiyon2}} + 7.20677\beta_{\text{enfeksiyon3}} + \\ & -7.24987\beta_{\text{enfeksiyon4}} + 3.32082\beta_{\text{osteomiyelit}} + 0.52248\beta_{\text{lökosit}} + -0.03470\beta_{\text{CRP}} + \\ & 0.08524\beta_{\text{sedim(antasyon)}} + 1.25624\beta_{\text{albumin}} + -0.60598\beta_{\text{PNL}} + -0.06707\beta_{\text{kreatinin}} + \\ & 0.03010\beta_{\text{total kolesterol}} + -0.04590\beta_{\text{LDL}} + 30.91029\beta_{\text{GANGREN}} + 8.93602\beta_{\text{wagnergrade2}} + \\ & 18.28763\beta_{\text{wagnergrade3}} + 35.82011\beta_{\text{wagnergrade4}} + 32.97573\beta_{\text{wagnergrade5}} + \\ & -5.28730\beta_{\text{hbo durumu}} + 0.07232\beta_{\text{hbo tedavi}} + 4.35483\beta_{\text{NPWT}} + -2.21941\beta_{\text{debridman}} \end{aligned}$$

Tablo 4. 54: Lojistik Regresyon analizi sonucunda elde edilmiş olan Nagelkerke R Square değeri

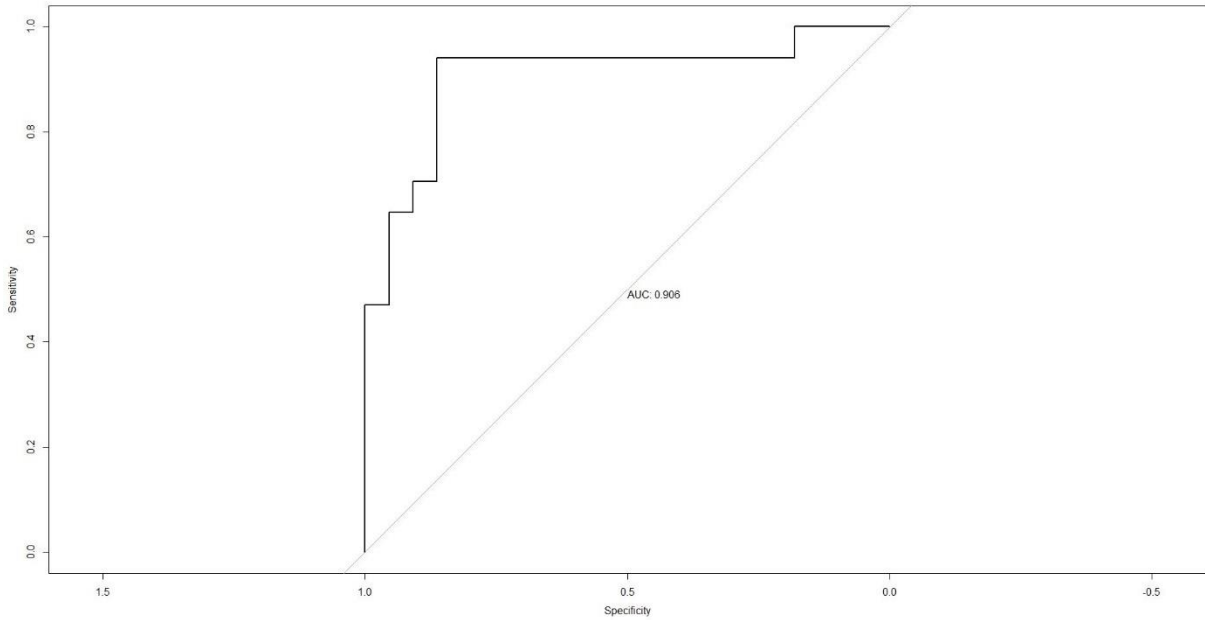
Step	-2 Log likelihood	Nagelkerke R Square
1	261,335	,694

Bağımlı değişkendeki değişimin % olarak tahmin etmek için model özetindeki Nagelkerke R^2 değeri üzerinden değerlendirilmektedir. Tablo 4.54'te detayları verilen ve elde edilen bu regresyon modeli Nagelkerke R^2 değerine göre ise oldukça iyi bir oran olan %69,4 oranında açıklanmaktadır.

Tablo 4. 55: Lojistik Regresyon Analizi -Elde Edilmiş olan Hosmer and Lemeshow Test Sonucu

Step	Chi-square	df	p
1	3,154	8	0,924

Araştırma için kurulan Lojistik model uygunluğu değerlendirildiğinde, Tablo 4.55'te detayları verilen ve uygulanan Hosmer-Lemeshow testi sonucunda uygun bir model olduğu saptanmıştır ($\chi^2:3.154$, $p \geq 0,05$).

**Şekil 4. 9:** Lojistik Regresyon Modelinden Elde Edilen ROC Eğrisi

Şekil 4.9'de görüleceği üzere Lojistik Regresyon modelinden elde edilen ROC eğrisi altında kalan alan AUC değeri 0.906 olarak hesaplanmıştır.

Tablo 4. 56: Lojistik Regresyon Algoritması için Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılım Veri Setinde Çapraz Geçerleme Değerlendirme Yöntemleri Kullanılarak Elde Edilen Performans Değerlendirme Ölçütleri

Lojistik Regresyon Analizi Çapraz Geçerleme Performans Değerlendirme Yöntemi- K Değerleri	Ortalama Doğruluk	Ortalama Hata
K=5	0,8059	0,1941
K=10	0,7961	0,2039
K=15	0,7912	0,2088
K=20	0,801	0,199

Tablo 4.56’da Lojistik Regresyon algoritması için nitelik seçimi yapılarak boyut sayısı azaltılmış veri setinin nümerik ve kategorik alt kırılım veri setinde çapraz geçerleme değerlendirme yöntemleri kullanılarak elde edilen performans değerlendirme ölçütleri sunulmuştur. Çapraz geçerleme yöntemi kullanılarak elde edilen en yüksek performans, 5 kat çapraz geçerleme yönteminde 0,8059 doğruluk değeri ile elde edilmiştir.

4.4.10 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılım Veri Setinde En iyi Performansı Gösteren İkili Ayırma ve Üçlü Ayırma Performans Değerlendirme Yöntemleri ile Elde Edilen Performans Değerlendirme Ölçütleri (İkili Sınıflandırma)

Bu bölümde, diyabetik ayak hastalarının genel – “overall” amputasyon riskini tahmin etmek için kullanılmış olan nitelik seçimi yapılarak boyut sayısı azaltılmış veri setinin nümerik ve kategorik alt kırılım veri setinde en iyi performansı gösteren ikili ayırma yönteminde (%90 Eğitim, %10 Test) ve üçlü ayırma yönteminde (%80 Eğitim, %10 Validasyon(Geçerleme), %10 Test) deney koşullarındaki performans değerlendirme ölçütlerinden **“Doğruluk”** ve **“F-Ölçütü”** esas alınarak kullanılan makine öğrenmesi sınıflandırma algoritmaları karşılaştırmalı olarak sunulmuştur.

Tablo 4 57: Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Alt Kırılım Veri Setinde En iyi Performansı Gösteren İkili Ayırma ve Üçlü Ayırma Performans Değerlendirme Yöntemleri ile Elde Edilen Doğruluk ve F-Ölçütü Performans Değerlendirme Ölçütleri

İkili Ayırma (“Hold-Out”) %90 Eğitim, %10 Test Performans Değerlendirme Yöntemi	Doğruluk	F-Ölçütü
Lojistik Regresyon	0,90	0,88889645
Sınıflandırma ve Regresyon Ağaçları (CART)	0,8205	0,799997877
Sade Bayes Algoritması	0,8205	0,810830679
Rastgele Orman Algoritması	0,7949	0,789501246
ÜÇLÜ AYIRMA YÖNTEMİ %80 Eğitim %10 Validasyon (Geçerleme) %10 Test / Validasyon Performans Sonuçları	Doğruluk	F-Ölçütü
Lojistik Regresyon	0,8293	0,799997877
Sınıflandırma ve Regresyon Ağaçları (CART)	0,8537	0,823541868
Sade Bayes Algoritması	0,8537	0,842110249
Rastgele Orman Algoritması Izgara Arama-“Grid Search” yöntemi ile parametre optimizasyonu sonucunda tespit edilen “ <i>mtry</i> = 3” değeri	0,878	0,857147427
ÜÇLÜ AYIRMA YÖNTEMİ %80 Eğitim %10 Validasyon (Geçerleme) %10 Test / TEST Performans Sonuçları	Doğruluk	F-Ölçütü
Rastgele Orman Algoritması	0,7805	0,727292562

Tablo 4.57’de görüleceği üzere ikili ayırma yönteminde en iyi performansın elde edildiği %90 eğitim ile oluşturulmuş makine öğrenmesi modellerinin geriye kalan %10 veri seti ile performansının değerlendirildiği deney koşullarında en iyi performansı 0.90 doğruluk oranı ile Lojistik Regresyon algoritması göstermiştir. Bununla birlikte üçlü ayırma yönteminde %80 eğitim veri seti kullanılarak oluşturulmuş makine öğrenmesi modellerinin geriye kalan %10 validasyon (geçerleme) veri setindeki en iyi performansı 0.8780 doğruluk oranı ile “Rastgele Orman Algoritması” göstermiş olup, bu algoritmanın da geriye kalan %10 test veri setindeki performansı ise 0.7805 doğruluk değerine ulaşmıştır.

4.5 VERİ MADENCİLİĞİ SÜRECİNDE MAKİNE ÖĞRENMESİ SINIFLANDIRMA ALGORİTMALARININ UYGULANMASI SONUCU ELDE EDİLEN BULGULAR – ÇOK KATEGORİLİ SINIFLANDIRMA

Çalışmada elde edilen diyabetik ayak hastalarının ampütasyon riskini öngörmek için elde edilen veri setinin hedef değişkeninde yer alan “**minör ampütasyon**” ve “**majör ampütasyon**” kategorilerindeki örnek sayıları arasında dengesizlik bulunmakta olup bu durum algoritmaların da performans sonuçlarını etkilemektedir. Dolayısı ile bu dengesizliğin de giderilmesi için literatür detaylı olarak incelenmiş ve çeşitli yöntemlerin bulunduğu tespit edilmiştir. Bunlardan bazıları “*oversampling*” ve “*undersampling*” adı verilen yöntemlerdir. Çalışmada dengesiz veri setini dengeli hâle getirmek için “*oversampling*” yöntemi kullanılmıştır. Bu kapsamda “*caret*” paketindeki, “*caret*” kütüphanesindeki “*up.sample*” fonksiyonu ve “*ROSE*” paketinde ve “*rose*” kütüphanesindeki “*ovun.sample*” fonksiyonu kullanılarak giderilmiştir. İlgili paket, kütüphane ve fonksiyonların hangi alt kırılım veri setlerinde kullanıldığı ise ilgili bölümlerde açıklanmıştır. Sonrasında ise dengesiz bir sınıf dağılımına sahip olan “**minör ampütasyon, majör ampütasyon ve ampütasyon yok**” etiketlerini içeren veri setinde hem dengeleme işlemi yapılarak hem de yapılmayarak elde edilen performans sonuçları paylaşılmıştır. Çok kategorili (“*multi-class*”) sınıflandırma için veri ön işleme adımında detaylı olarak ele alınan prosedürler gerçekleştirildikten orijinal yani nümerik ve kategorik niteliklerin ve de veri setinin ayrıklaştırma prosedürü sonucunda tamamen kategorik niteliklerin yer aldığı veri setlerinde yalnızca üçlü ayırma (eğitim, validasyon[geçerleme] ve test) performans değerlendirme yaklaşımı kullanılmıştır.

4.5.1 Orijinal Veri Setinden Elde Edilen Bulgular

4.5.1.1 Orijinal Veri Setinin %80’ini Oluşturan “Eğitim” Veri Setinde Veri Seti Dengeleme İşlemi Gerçekleştirilmeden Elde Edilen Bulgular

Bu veri setinde, veri setinin %80’ini oluşturan eğitim veri setinde **veri seti dengeleme işlemi gerçekleştirilmeden** uygulanan çeşitli yapay zekâ makine öğrenmesi sınıflandırma algoritmalarının performans sonuçlarından elde edilen bulgular karşılaştırılmalı olarak paylaşılmıştır.

Tablo 4. 58: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Minör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/ Validasyon Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmamış Model	0,5667	0,4667	0,6667	0,4375	0,6923	0,451628511
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0,6852	0,6667	0,7037	0,5556	0,7917	0,606100826
Rastgele Orman Algoritması “mtry = 10”	0,7741	0,7333	0,8148	0,6875	0,8462	0,70966181
Sade Bayes Algoritması	0,6741	0,5333	0,8148	0,6154	0,7586	0,57141607
Lojistik Regresyon Algoritması	0,7111	0,5333	0,8889	0,7273	0,7742	0,615372188
Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/ TEST Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Rastgele Orman Algoritması “mtry = 10”	0,8489	0,9286	0,7692	0,6842	0,9524	0,787882093

Tablo 4.58’de görüleceği üzere üçlü ayırma metodu kullanılarak gerçekleştirilen algoritmaların performans değerlendirmesinde, Rastgele Orman Algoritması diğer algoritmalarla karşılaştırıldığında daha iyi performans göstermiş ve algoritma için ızgara arama yönteminde

optimize edilmesi gereken “*mtry*” değeri “10” olarak tespit edilmiştir. Algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,7741 doğruluk değeri elde edilirken, test veri setindeki performansı ise 0,8489 değeri olmuştur.

Tablo 4. 59: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/ Validasyon Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmamış Model	0,57237	0,25	0,89474	0,2	0,91892	0,22222222
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0,48684	0	0,97368	0	0,90244	NaN
Rastgele Orman Algoritması “ <i>mtry</i> = 10”	0,61184	0,25	0,97368	0,5	0,925	0,33333333
Sade Bayes Algoritması	0,69737	0,5	0,89474	0,33333	0,94444	0,3999976
Lojistik Regresyon Algoritması	0,96053	1	0,92105	0,57143	1	0,727273884
Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test, TEST Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Lojistik Regresyon Algoritması	0,68421	0,5	0,86842	0,28571	0,94286	0,363632893

Tablo 4.59’da görüleceği üzere üçlü ayırma metodu kullanılarak gerçekleştirilen algoritmaların performans değerlendirmesinde, Lojistik Regresyon Algoritması diğer algoritmalarla karşılaştırıldığında daha iyi performans göstermiş ve bu kapsamda algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,96053 doğruluk elde edilirken, test veri setindeki performansı ise 0,68421 değeri olmuştur.

4.5.1.2 Orijinal Veri Setinin %80’ini Oluşturan “Eğitim” Veri Setinde Veri Setinin “Oversampling” Metodu ile Dengeleme İşlemi Gerçekleştirdikten Sonra Elde Edilen Bulgular

Tablo 4. 60: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Minör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), Validasyon Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmamış Model	0,8444	0,8	0,8889	0,8	0,8889	0,8
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0,8778	0,8667	0,8889	0,8125	0,9231	0,838725286
Rastgele Orman Algoritması “mtry = 4”	1	1	1	1	1	1
Sade Bayes Algoritması	0,7444	0,6	0,8889	0,75	0,8	0,666666667
Lojistik Regresyon Algoritması	0,863	0,8	0,9259	0,8571	0,8929	0,82756623

Tablo 4. 60: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Minör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları (devam)

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test, Test Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0.8379	0.7143	0.9615	0.9091	0.8621	0.8379

Dengesiz bir veri seti olan veri setinin bu alt kırılımında, yalnızca “**eğitim**” veri setinde dengeleme işlemi “*oversampling*” yöntemi kullanılarak gerçekleştirilmiştir. Bu kapsamda “*caret*” paketindeki, “*caret*” kütüphanesindeki “*up.sample*” fonksiyonu kullanılmıştır. Tablo 4.60’ta görüleceği üzere üçlü ayırma metodu kullanılarak gerçekleştirilen algoritmaların performans değerlendirmesinde, Sınıflandırma ve Regresyon Ağaçları (CART) [Budama Yapılmış] diğer algoritmalarla karşılaştırıldığında daha iyi performans göstermiş ve bu kapsamda algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,8778 doğruluk elde edilirken, test veri setindeki performansı ise 0.8379 değeri olmuştur. Rastgele Orman Algoritması ise tüm performans değerlendirme ölçütlerinde % 100 performans göstermesine rağmen bu durum algoritmanın ezberleme yapması nedeniyle yani “*overfitting*” durumunun olabileceği düşünülerek algoritmanın test veri setindeki performansına bakılmamıştır.

Tablo 4. 61: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test /Validasyon Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmamış Model	1	1	1	1	1	1
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	1	1	1	1	1	1
Rastgele Orman Algoritması “mtry = 4”	1	1	1	1	1	1
Sade Bayes Algoritması	0,94737	1	0,89474	0,5	1	0,666666667
Lojistik Regresyon Algoritması	1	1	1	1	1	1
Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/ TEST Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sade Bayes Algoritması	0,7387	0,6667	0,8108	0,2222	0,9677	0,333312499

Dengesiz bir veri seti olan veri setinin bu alt kırılımında, yalnızca “**eğitim**” veri setinde dengeleme işlemi “*oversampling*” yöntemi kullanılarak gerçekleştirilmiştir. Bu kapsamda

“caret” paketindeki, “caret” kütüphanesindeki “*up.sample*” fonksiyonu kullanılmıştır. Tablo 4.61’de görüleceği üzere üçlü ayırma metodu kullanılarak gerçekleştirilen algoritmaların performans değerlendirmesinde, “Sade Bayes Algoritması” diğer algoritmalarla karşılaştırıldığında daha iyi performans göstermiş ve bu kapsamda algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,94737 doğruluk elde edilirken, test veri setindeki performansı ise değeri 0,7387 olmuştur. Sınıflandırma ve Regresyon Ağaçları, Rastgele Orman Algoritması ve Lojistik Regresyon algoritmaları ise tüm performans değerlendirme ölçütlerinde %100 performans göstermesine rağmen bu durum algoritmanın ezberleme yapması nedeniyle yani “*overfitting*” durumunun olabileceği düşünülerek algoritmanın test veri setindeki performansına bakılmamıştır.

4.5.2 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Niteliklerden Oluşan Alt Kırılımından Elde Edilen Bulgular

4.5.2.1 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Niteliklerden Oluşan Alt Kırılımının %80’ini Oluşturan “Eğitim” Veri Setinde, Veri Seti Dengeleme İşlemi Gerçekleştirilmeden Elde Edilen Bulgular

Veri setinin bu alt kırılımında, veri setinin %80’ini oluşturan eğitim veri setinde veri seti dengeleme işlemi gerçekleştirilmeden uygulanan çeşitli yapay zekâ makine öğrenmesi sınıflandırma algoritmalarının performans sonuçlarından elde edilen bulgular karşılaştırılmalı olarak paylaşılmıştır.

Tablo 4. 62: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Minör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/ Validasyon Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmamış Model	0,7111	0,5333	0,8889	0,7273	0,7742	0,615372188
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0,7593	0,6667	0,8519	0,7143	0,8214	0,689679667
Rastgele Orman- “Random Forest” Algoritması “mtry = 2”	0,7074	0,6	0,8148	0,6429	0,7857	0,620709631
Sade Bayes Algoritması	0,6556	0,5333	0,7778	0,5714	0,75	0,551692985
Lojistik Regresyon Algoritması	0,7593	0,6667	0,8519	0,7143	0,8214	0,689679667
Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/ TEST Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Lojistik Regresyon Algoritması	0,7747	0,8571	0,6923	0,6	0,9	0,70586782

Tablo 4.62’de görüleceği üzere üçlü ayırma metodu kullanılarak gerçekleştirilen algoritmaların performans değerlendirmesinde, Lojistik Regresyon Algoritması diğer algoritmalarla

karşılaştırıldığında daha iyi performans göstermiştir. Algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,7593 doğruluk değeri elde edilirken, test veri setindeki performansı ise 0,7747 değeri olmuştur.

Tablo 4. 63: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), Validasyon Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmamış Model	0,72368	0,5	0,94737	0,5	0,94737	0,5
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0,5	0	1	NaN	0,90476	NaN
Rastgele Orman- “Random Forest” Algoritması “mtry = 2”	0,75	0,5	1	1	0,95	0,666666667
Sade Bayes Algoritması	0,68421	0,5	0,86842	0,28571	0,94286	0,363632893
Lojistik Regresyon Algoritması	0,84868	0,75	0,94737	0,6	0,97297	0,666666667
Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/ TEST Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Lojistik Regresyon Algoritması	0,6667	0,3333	1	1	0,9487	0,499962499

Tablo 4.63'te görüleceği üzere üçlü ayırma metodu kullanılarak gerçekleştirilen algoritmaların performans değerlendirmesinde, Lojistik Regresyon Algoritması diğer algoritmalarla karşılaştırıldığında daha iyi performans göstermiştir. Algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,84868 doğruluk değeri elde edilirken, test veri setindeki performansı ise 0,6667 değeri olmuştur.

4.5.2.2 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Nümerik ve Kategorik Niteliklerden Oluşan Alt Kırılımının %80'ini Oluşturan “Eğitim” Veri Setinde, Veri Seti Dengeleme İşlemi Gerçekleştirildikten Sonra Elde Edilen Bulgular

Tablo 4. 64: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Minör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), Validasyon Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmamış Model	0,8296	0,7333	0,9259	0,8462	0,8621	0,8296
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0,8296	0,7333	0,9259	0,8462	0,8621	0,8296
Rastgele Orman Algoritması “mtry = 2”	1	1	1	1	1	1
Sade Bayes Algoritması	0,7111	1	0,81579	0,36364	1	0,7111
Lojistik Regresyon Algoritması	0,6963	0,4667	0,9259	0,7778	0,7576	0,6963

Tablo 4. 65: Hedef Niteliğin 3 Kategoriden Oluştığı (Multi-Class) Diyabetik Ayak Hastalarının “Minör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları (devam)

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test, Test Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmamış Model	0,7940	0,8571	0,7308	0,6316	0,9048	0,7940
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0,8324	0,8571	0,8077	0,7059	0,9130	0,8324

Dengesiz bir veri seti olan veri setinin bu alt kırılımında, yalnızca “**eğitim**” veri setinde dengeleme işlemi “*oversampling*” yöntemi kullanılarak gerçekleştirilmiştir. Bu kapsamda “*caret*” paketindeki, “*caret*” kütüphanesindeki “*up.sample*” fonksiyonu kullanılmıştır. Tablo 4.64’te görüleceği üzere üçlü ayırma metodu kullanılarak gerçekleştirilen algoritmaların performans değerlendirmesinde, “Sınıflandırma ve Regresyon Ağaçları (CART)” diğer algoritmalarla karşılaştırıldığında daha iyi performans göstermiş ve bu kapsamda algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; budama yapılmamış model için validasyon veri setinde 0,8296 doğruluk değeri elde edilirken, test veri setindeki doğruluk değeri ise 0,7940 olmuştur. Sınıflandırma ve Regresyon Ağaçlarının budama yapılmış modeli için validasyon veri setinde 0,8296 doğruluk değeri elde edilirken, test veri setindeki doğruluk değeri ise 0,8324 olarak saptanmıştır. Rastgele Orman Algoritması ise tüm performans değerlendirme ölçütlerinde %100 performans göstermesine rağmen bu durum algoritmanın ezberleme yapması nedeniyle yani “*overfitting*” durumunun olabileceği düşünülerek algoritmanın test veri setindeki performansına bakılmamıştır.

Tablo 4. 66: Hedef Niteliğin 3 Kategoriden Oluştugu (Multi-Class) Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (Three-Way Split) Performans Sonuçları

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/Validasyon Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmamış Model	0.97368	1	0.94737	0.66667	1	0.97368
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0.98684	1	0.97368	0.80000	1	0.98684
Rastgele Orman Algoritması “mtry = 2”	1	1	1	1	1	1
Sade Bayes Algoritması	0.90789	1	0.81579	0.36364	1	0.90789
Lojistik Regresyon Algoritması	0.92105	1	0.84211	0.40000	1	0.92105
Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test, TEST Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış	0.9730	1	0.9459	0.6000	1	0.9730

Dengesiz bir veri seti olan veri setinin bu alt kırılımında, yalnızca “**eğitim**” veri setinde dengeleme işlemi “*oversampling*” yöntemi kullanılarak gerçekleştirilmiştir. Bu kapsamda

“caret” paketindeki, “caret” kütüphanesindeki “up.sample” fonksiyonu kullanılmıştır. Tablo 4.63’te görüleceği üzere üçlü ayırma metodu kullanılarak gerçekleştirilen algoritmaların performans değerlendirmesinde, “Sınıflandırma ve Regresyon Ağaçları (CART)” diğer algoritmalarla karşılaştırıldığında daha iyi performans göstermiş ve bu kapsamda algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; budama yapılmamış model için validasyon veri setinde 0.97368 doğruluk değeri elde edilmiştir. Sınıflandırma ve Regresyon Ağaçlarının budama yapılmış modeli için validasyon veri setinde 0.98684 doğruluk değeri elde edilirken, test veri setindeki doğruluk değeri ise 0.9730 olarak saptanmıştır. Rastgele Orman Algoritması ise tüm performans değerlendirme ölçütlerinde %100 performans göstermesine rağmen bu durum algoritmanın ezberleme yapması nedeniyle yani “*overfitting*” durumunun olabileceği düşünülerek algoritmanın test veri setindeki performansına bakılmamıştır.

4.5.3 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Ayrıklaştırma Prosedürü Sonucunda Tamamen Kategorik Niteliklerden Oluşan Alt Kırılımından Elde Edilen Bulgular

4.5.3.1 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Ayrıklaştırma Prosedürü Sonucunda Tamamen Kategorik Niteliklerden Oluşan Alt Kırılımının %80’ini Oluşturan “Eğitim” Veri Setinde, Veri Seti Dengeleme İşlemi Gerçekleştirilmeden Elde Edilen Bulgular

Bu veri setinde, veri setinin %80’ini oluşturan eğitim veri setinde veri seti dengeleme işlemi gerçekleştirilmeden uygulanan çeşitli yapay zekâ makine öğrenmesi sınıflandırma algoritmalarının performans sonuçlarından elde edilen bulgular karşılaştırılmalı olarak paylaşılmıştır.

Tablo 4. 67: Hedef Niteliğin 3 Kategoriden Oluştugu (“Multi-Class”) Diyabetik Ayak Hastalarının “Minör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (“Three-Way Split”) Performans Sonuçları

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/Validasyon Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmamış Model	0,8407	0,8667	0,8148	0,7222	0,9167	0,787879338
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0,8407	0,8667	0,8148	0,7222	0,9167	0,787879338
Rastgele Orman- “Random Forest” Algoritması “mtry = 4”	0,7926	0,7333	0,8519	0,7333	0,8519	0,7333
Sade Bayes Algoritması	0,7222	0,6667	0,7778	0,625	0,8077	0,645176899
Lojistik Regresyon Algoritması	0,7556	0,7333	0,7778	0,6471	0,84	0,687508592
Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/TEST Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Lojistik Regresyon Algoritması	0,6841	0,7143	0,6538	0,5263	0,8095	0,60605528

Tablo 4.66’da görüleceği üzere üçlü ayırma metodu kullanılarak gerçekleştirilen algoritmaların performans değerlendirmesinde, Lojistik Regresyon Algoritması diğer algoritmalarla karşılaştırıldığında daha iyi performans göstermiştir. Algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,7556 doğruluk değeri elde edilirken, test veri setindeki performansı ise 0,6841 değeri olmuştur.

Tablo 4. 68: Hedef Niteliğin 3 Kategoriden Oluştugu (“Multi-Class”) Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (“Three-Way Split”) Performans Sonuçları

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), Validasyon Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmamış Model	0,5	0	1	NaN	0,90476	NaN
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0,5	0	1	NaN	0,90476	NaN
Rastgele Orman- “Random Forest” Algoritması “mtry = 4”	0,5	0,5	1	1	0,95	0,666666667
Sade Bayes Algoritması	0,69737	0,5	0,89474	0,33333	0,94444	0,3999976
Lojistik Regresyon Algoritması	0,83553	0,75	0,92105	0,5	0,97222	0,6

Tablo 4. 67: Hedef Niteliğin 3 Kategoriden Oluştığı (“Multi-Class”) Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (“Three-Way Split”) Performans Sonuçları (devam)

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test, TEST Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Lojistik Regresyon Algoritması	0,6396	0,3333	0,9459	0,3333	0,9459	0,3333

Tablo 4.67’de görüleceği üzere üçlü ayırma metodu kullanılarak gerçekleştirilen algoritmaların performans değerlendirmesinde, Lojistik Regresyon Algoritması diğer algoritmalarla karşılaştırıldığında daha iyi performans göstermiştir. Algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,83553 doğruluk değeri elde edilirken, test veri setindeki performansı ise 0,6396 değeri olmuştur.

4.5.3.2 Nitelik Seçimi Yapılarak Boyut Sayısı Azaltılmış Veri Setinin Ayırıklaştırma Prosedürü Sonucunda Tamamen Kategorik Niteliklerden Oluşan Alt Kırılıminin %80’ini Oluşturan “Eğitim” Veri Setinde, Veri Seti Dengeleme Prosedürü Gerçekleştirildikten Sonra Elde Edilen Bulgular

Bu veri setinde, veri setinin %80’ini oluşturan eğitim veri setinde veri seti dengeleme işlemi gerçekleştirildikten sonra uygulanan çeşitli yapay zekâ makine öğrenmesi sınıflandırma algoritmalarının performans sonuçlarından elde edilen bulgular karşılaştırılmalı olarak paylaşılmıştır.

Tablo 4. 69: Hedef Niteliğin 3 Kategoriden Oluştugu (“Multi-Class”) Diyabetik Ayak Hastalarının “Minör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (“Three-Way Split”) Performans Sonuçları

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/ Validasyon Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmamış Model	0,7111	0,5333	0,8889	0,7273	0,7742	0,615372188
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0,7111	0,5333	0,8889	0,7273	0,7742	0,615372188
Rastgele Orman Algoritması “mtry = 3”	1	1	1	1	1	1
Sade Bayes Algoritması	0,7444	0,6	0,8889	0,75	0,8	0,666666667
Lojistik Regresyon Algoritması	0,7593	0,6667	0,8519	0,7143	0,8214	0,689679667
Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test, TEST Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Lojistik Regresyon Algoritması	0,7775	0,7857	0,7692	0,6471	0,8696	0,709696357

Dengesiz bir veri seti olan veri setinin bu alt kırılımında, yalnızca “**eğitim**” veri setinde dengeleme işlemi “*oversampling*” yöntemi kullanılarak gerçekleştirilmiştir. Bu kapsamda “*caret*” paketindeki, “*caret*” kütüphanesindeki “*up.sample*” fonksiyonu kullanılmıştır. Tablo

4.68’de görüleceği üzere üçlü ayırma metodu kullanılarak gerçekleştirilen algoritmaların performans değerlendirmesinde, Lojistik Regresyon Algoritması diğer algoritmalarla karşılaştırıldığında daha iyi performans göstermiştir. Algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,7593 doğruluk değeri elde edilirken, test veri setindeki performansı ise 0,7775 değeri olmuştur.

Tablo 4. 70: Hedef Niteliğin 3 Kategoriden Oluştugu (“Multi-Class”) Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (“Three-Way Split”) Performans Sonuçları

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/Validasyon Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmamış Model	0,92105	1	0,84211	0,4	1	0,571428571
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0,92105	1	0,84211	0,4	1	0,571428571
Rastgele Orman Algoritması “mtry = 3”	1	1	1	1	1	1
Sade Bayes Algoritması	0,92105	1	0,84211	0,4	1	0,571428571
Lojistik Regresyon Algoritması	0,94737	1	0,89474	0,5	1	0,666666667

Tablo 4. 69: Hedef Niteliğin 3 Kategoriden Oluştugu (“*Multi-Class*”) Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskini Tahmin Etmek için Gerçekleştirilen Makine Öğrenmesi Sınıflandırma Algoritmalarının Üçlü Ayırma (“*Three-Way Split*”) Performans Sonuçları (devam)

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/ TEST Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Lojistik Regresyon Algoritması	0,8063	0,6667	0,9459	0,5	0,9722	0,571440816

Dengesiz bir veri seti olan veri setinin bu alt kırılımında, yalnızca “**eğitim**” veri setinde dengeleme işlemi “*oversampling*” yöntemi kullanılarak gerçekleştirilmiştir. Bu kapsamda “*caret*” paketindeki, “*caret*” kütüphanesindeki “*up.sample*” fonksiyonu kullanılmıştır. Tablo 4.69’da görüleceği üzere üçlü ayırma metodu kullanılarak gerçekleştirilen algoritmaların performans değerlendirmesinde, Lojistik Regresyon Algoritması diğer algoritmalarla karşılaştırıldığında daha iyi performans göstermiştir. Algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,94737 doğruluk değeri elde edilirken, test veri setindeki performansı ise 0,8063 değeri olmuştur.

4.6 DİYABETİK AYAK HASTALARININ MAJÖR AMPÜTASYON RİSKİNİN “*ONE VERSUS REST*” YAKLAŞIMI KULLANILARAK ÖNGÖRÜLMESİ

Bu doktora tez çalışması kapsamında öncelikli hedef olarak diyabetik ayak hastalarının genel – “*overall*” ampütasyon riskinin yapay zekâ teknikleri ile öngörülmesi olup, bu problem öncelikli olarak ele alınmıştır. Bununla birlikte veri setinde majör ampütasyon sınıfında toplamda **31 hasta** bulunmaktadır. Deneysel olarak nümerik ve kategorik niteliklerden oluşan orijinal veri setinde bu kapsamda “*one versus rest*” yaklaşımı kullanılarak 31 hastadan oluşan majör ampütasyon sınıfı “**pozitif sınıf**” olarak belirlenmiş, geriye kalan toplamda 376 hasta ise “**negatif sınıf**” olarak belirlenmiştir. Bu durumda ele alınan problem tekrar ikili “*binary*” sınıflandırma problemine dönüşmektedir. Ayrıca, veri seti dengesiz veri seti olarak tanımlanmakta ve bu dengesizliğin de yine giderilmesi gerekmektedir. Bununla birlikte

yalnızca model eğitiminde kullanılan “**eğitim**” veri setinde veri seti dengeleme işlemi yapılmalıdır.

Eğitim veri setindeki dengesizlik, “*ROSE*” paketinde ve “*rose*” kütüphanesindeki “*ovun.sample*” fonksiyonu kullanılarak giderilmiştir. Bu yöntemde, azınlık durumunda bulunan örnekler rastgele olarak “*oversampling*” yöntemi ile, çoğunluk durumunda bulunan örnekler ise “*undersampling*” metodu ile veya “*undersampling*” ve “*oversampling*” yöntemlerinin kombinasyonu yolu ile dengeli örnekler oluşturulmaktadır. Performans değerlendirme yöntemi olarak ise de yine üçlü ayırma (eğitim, validasyon ve test) yaklaşımı tercih edilmiştir. Böylece veri setinin modeli eğitmek için ayrılan “**eğitim**” veri seti, tüm veri setinin %80’i, %70’i ve %60’ı kullanılarak elde edilmiştir. Böylelikle farklı eğitim veri seti senaryolarının oluşturulması sağlanmıştır.

Tablo 4. 71: Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskinin “*One-versus-All*” metodolojisi, Eğitim Veri Setinde Veri Seti Dengeleme İşlemi Yapılarak ve Üçlü Ayırma (%80 Eğitim, %10 Validasyon,%10 Test) Performans Değerlendirme Yöntemi Kullanılarak Öngörülmesi

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test/ Validasyon Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmamış	0,7857	0,5	0,81579	0,22222	0,93939	0,307690178
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış	0,8333	0,5	0,86842	0,28571	0,94286	0,363632893
Rastgele Orman Algoritması “mtry = 3”	0,9048	0,5	0,94737	0,5	0,94737	0,5
Sade Bayes Algoritması	0,8095	0,5	0,84211	0,25	0,94118	0,333333333

Tablo 4. 70: Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskinin “*One-versus-All*” metodolojisi, Eğitim Veri Setinde Veri Seti Dengeleme İşlemi Yapılarak ve Üçlü Ayırma (%80 Eğitim, %10 Validasyon,%10 Test) Performans Değerlendirme Yöntemi Kullanılarak Öngörülmesi (devam)

Model Performans Değerlendirme Yöntemi: %80 Eğitim, %10 Validasyon (Geçerleme), %10 Test, TEST Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Rastgele Orman Algoritması	0,9024	0,5	0,94595	0,5	0,94595	0,5

Dengesiz bir veri seti olan veri setinin bu alt kırımında, yalnızca “**eğitim**” veri setinde dengeleme işlemi “*ROSE*” paketindeki, “*rose*” kütüphanesindeki “*ovun.sample*” fonksiyonu kullanılmıştır. Tablo 4.70’te görüleceği üzere üçlü ayırma metodu kullanılarak gerçekleştirilen algoritmaların performans değerlendirmesinde, Rastgele Orman Algoritması, diğer algoritmalarla karşılaştırıldığında daha iyi performans göstermiştir. Rastgele Orman Algoritmasında optimize edilmesi gereken “*mtry*” değeri ızgara arama metodunun gerçekleştirilmesi suretiyle optimize edilmiş olup optimum değer “3” olduğu saptanmıştır. Algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,9048 doğruluk değeri elde edilirken, test veri setindeki performansı ise 0,9024 doğruluk değerine ulaşmıştır.

Tablo 4. 72: Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskinin “*One-versus-All*” metodolojisi, Eğitim Veri Setinde Veri Seti Dengeleme İşlemi Yapılarak ve Üçlü Ayırma (%70 Eğitim, %20 Validasyon,%10 Test) Performans Değerlendirme Yöntemi Kullanılarak Öngörülmesi

Model Performans Değerlendirme Yöntemi: %70 Eğitim, %20 Validasyon (Geçerleme), %10 Test/ Validasyon Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmamış Model	0,8333	0,25	0,89474	0,2	0,91892	0,222222222
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0,881	0,5	0,92105	0,4	0,94595	0,444444444
Rastgele Orman- “ <i>Random Forest</i> ” Algoritması	0,9048	0,25	0,97368	0,5	0,925	0,333333333
Sade Bayes Algoritması	0,9762	1	0,97368	0,8	1	0,888888889
Model Performans Değerlendirme Yöntemi: %70 Eğitim, %20 Validasyon (Geçerleme), %10 Test, TEST Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sade Bayes Algoritması	0,7805	0,71429	0,78667	0,2381	0,96721	0,35714875

Dengesiz bir veri seti olan veri setinin bu alt kırılımında, yalnızca “**eğitim**” veri setinde dengeleme işlemi “*ROSE*” paketindeki, “*rose*” kütüphanesindeki “*ovun.sample*” fonksiyonu kullanılmıştır. Tablo 4.71’de görüleceği üzere üçlü ayırma metodu kullanılarak gerçekleştirilen algoritmaların performans değerlendirmesinde, Sade Bayes Algoritması, diğer algoritmalarla

karşılaştırıldığında daha iyi performans göstermiştir. Algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,9762 doğruluk değeri elde edilirken, test veri setindeki performansı ise 0,7805 doğruluk değeri olmuştur. Ayrıca özellikle “Kesinlik-Pozitif Öngörü Değeri” nin 0,2381 olarak saptanması ve oldukça düşük performans göstermesi dikkat çekicidir.

Tablo 4. 73: Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskinin “One-versus-All” metodolojisi, Eğitim Veri Setinde Veri Seti Dengeleme İşlemi Yapılarak ve Üçlü Ayırma (%60 Eğitim,%20 Validasyon,%20 Test) Performans Değerlendirme Yöntemi Kullanılarak Öngörülmesi

Model Performans Değerlendirme Yöntemi: %60 Eğitim, %20 Validasyon (Geçerleme), %20 Test/ Validasyon Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0,8171	0,33333	0,85526	0,15385	0,94203	0,210529252
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0,9146	0,33333	0,96053	0,4	0,94805	0,36363438
Rastgele Orman- “Random Forest” Algoritması “mtry = 10”	0,9024	0,16667	0,96053	0,25	0,9359	0,2000024
Sade Bayes Algoritması	0,8293	0,83333	0,82895	0,27778	0,98438	0,41666875

Tablo 4. 72: Diyabetik Ayak Hastalarının “Majör Ampütasyon” Riskinin “One-versus-All” metodolojisi, Eğitim Veri Setinde Veri Seti Dengeleme İşlemi Yapılarak ve Üçlü Ayırma (%60 Eğitim,%20 Validasyon,%20 Test) Performans Değerlendirme Yöntemi Kullanılarak Öngörülmesi (devam)

Model Performans Değerlendirme Yöntemi: %60 Eğitim, %20 Validasyon (Geçerleme), %20 Test, TEST Veri Seti Performans Sonuçları	Doğruluk	Duyarlılık	Belirleyicilik	Kesinlik	Negatif Öngörü Değeri	F-Ölçütü
Sınıflandırma ve Regresyon Ağaçları (CART) Budama Yapılmış Model	0,825	0,8333	0,8243	0,2778	0,9839	0,416687499

Dengesiz bir veri seti olan veri setinin bu alt kırımında, yalnızca “**eğitim**” veri setinde dengeleme işlemi “*ROSE*” paketindeki, “*rose*” kütüphanesindeki “*ovun.sample*” fonksiyonu kullanılmıştır. Tablo 4.72’de görüleceği üzere üçlü ayırma metodu kullanılarak gerçekleştirilen algoritmaların performans değerlendirmesinde, Sınıflandırma ve Regresyon Ağaçları modelinin budama yapılmış versiyonu, diğer algoritmalarla karşılaştırıldığında daha iyi performans göstermiştir. Algoritmanın validasyon veri setindeki performansı değerlendirildiğinde; 0,9146 doğruluk değeri elde edilirken, test veri setindeki performansı ise 0,825 değeri olmuştur. Bununla birlikte özellikle “Kesinlik-Pozitif Öngörü Değeri” nin 0,2778 değeri gibi oldukça düşük performans göstermesi dikkat çekicidir.

4.7 GELİŞTİRİLEN MODELLERİN R SHINY KULLANICI ARA YÜZÜNE ENTEĞRE EDİLMESİ

Diyabetik ayak hastalığının genel – “*overall*” ampütasyon riskinin tahmin edilmesinde kullanılan ve en yüksek performansı gösteren Lojistik Regresyon ve Rastgele Orman Ağaçları modelleri seçilerek alternatif iki model seçilmiş, oluşturulmuş ve Shiny kütüphanesinden yararlanılarak web uygulaması olarak kullanıma sunulmuştur. Geliştirilen bu uygulamanın diyabetik ayak tedavisi ile uğraşan hekim, klinik veya kullanıcıların hastaların ampütasyon riskini öngörmesinde yardımcı olabileceği düşünülmektedir. Ücretsiz olarak hesap açılabilen bununla birlikte premium versionları da bulunan Shinyapps.io üzerinden geliştirilen bu uygulama sayesinde geliştirilen modellerin gerçek hayatta kullanılması hedeflenmiştir (Chang ve diğ., 2021).

Diabetic Foot Amputation Risk Prediction Project

Model About

☐ Show Data

Model Inputs

Select Variables

Previous Amputation History: Minor Amputation

Renal Failure Status: No

Dialysis Status: No

PEDIS Peripheral Arterial Disease: Grade 2

PEDIS Wound Dept Grades: Grade 2

PEDIS Infection Grades: Grade 2

Osteomyelitis: No

WBC($10^3/\mu\text{l}$): 15

Input CRP(mg/L): 20

Sedimentation(mm/hour): 20

Albumin(g/dL): 3,8

PNL($10^3/\mu\text{l}$): 4,4

Creatinine(mg/dL): 1,1

Total Cholesterol(mg/dL): 200

Low-density Lipoprotein(mg/dL): 150

GANGRENE Status: No

Wagner Grades: Grade 2

Hyperbaric Oxygen Therapy/HBOT Treatment Session: 0

Negative Pressure Wound Therapy Status: No

Debridement Status: No

Hyperbaric Oxygen Therapy/HBOT Treatment Status: No

Select Model

☒ Random Forest ☐ Logistic Regression

Predict

Value: No

Levels: No Yes

Şekil 4. 10: Diyabetik Ayak Hastalarının Genel – “Overall” Ampütasyon Riskinin Tahmini için Geliştirilen R-Shiny Uygulaması ve Örnek Değerlerin Girilmesi ile Verilen Tahmin Çıktısı

5 TARTIŞMA VE SONUÇ

Diyabet ülkemizde ve dünyada oldukça yaygın olan bir hastalıktır. Diyabetin en ciddi komplikasyonlarından biri olan diyabetik ayak hastalığı, dünyada her 30 saniyede bir ayak kaybına neden olan tedavisi güç bir hastalık olmakla birlikte diyabetin en çok hastanede yatmaya neden olan komplikasyonudur (Mutluoğlu, 2013). Ayrıca, amputasyon operasyonu hastalar için fiziksel ve psikososyal bir travmaya neden olur (Saltoglu ve diğ., 2016). Diyabetik ayak, doğru tedavi edilmediği takdirde hayati sonuçlar doğurabilecek multifaktöriyel bir hastalık olduğundan, amputasyon riski ile ilişkili faktörlerin önceden belirlenmesi hastalığın prognozu açısından kritiktir.

Literatür incelendiğinde diyabet hastalığının yapay zekâ – makine öğrenmesi sınıflandırma algoritmaları ve bunların optimum parametrelerinin parametre optimizasyon teknikleri ile optimize edildiği çalışmalar saptanmış olup bu durum diyabetik ayak hastalığının amputasyon riskinin yapay zekâ teknikleri ile öngörülmesi için geçerli değildir (Pekel ve Özcan, 2018; Özmen Pekel ve Özcan, 2020). Dolayısı ile diyabetik ayak hastalarının amputasyon riskinin yapay zekâ- makine öğrenmesi sınıflandırma algoritmaları kullanılarak öngörülmesine yönelik bir gerekliliğin olduğu saptanmıştır. Böylelikle bu tez araştırması gerçekleştirilmiş ve tez araştırması sonucunda, farklı yapay zekâ – makine öğrenmesi sınıflandırma algoritmalarından yararlanılarak elde edilen modeller sonucunda farklı performans değerleri bulunmuştur. Araştırma sonucunda diyabetik ayak hastalığının özellikle genel amputasyon riskinin makine öğrenmesi sınıflandırma algoritmaları ile öngörülebileceği sonucuna varılmıştır.

Yükseköğretim Kurulu Tez Merkezi'nde bulunan tezler incelendiğinde “Sağlık Bilişimi” alanında daha önce bu kapsamda yapılmış hiçbir tez çalışmasına (yüksek lisans ve doktora dahil) ulaşılamamıştır. Sistemik literatür taramasının gerçekleştirilmesi suretiyle geçmiş çalışmalar incelendiğinde amputasyon riski için yaygın olarak istatistiksel metodların kullanıldığını ve lojistik regresyon analizinin uygulanması yoluyla çeşitli risk ve/veya önleyici faktörlerin ortaya çıkarıldığı saptanmıştır.

Bu çalışmada elde edilen bulgular geçmiş çalışmalarda elde edilen bulguları doğrulamakla birlikte bu araştırmanın gerçekleştirilmesi ile ilk kez ortaya çıkarılan faktörler de keşfedilmiştir. Şekil 4.4'te görülebileceği üzere Rastgele Orman Algoritmasının en önemli niteliği yani dallanmanın başladığı nitelik yara derinliği değişkenidir. Bu değişkeni ise, ilgili ağaç yapısında

periferik arter hastalığının şiddeti, hastanın HBOT tedavisi seans sayısı değişkeni, gangren, CRP, hastanın NPWT tedavisi statüsü ve total kolesterol değişkenleri takip etmektedir. Şekil 4.3'te sunulan nümerik ve kategorik değişkenlerden oluşan orijinal veri setinde sınıflandırma ve regresyon ağaçları algoritmasından elde edilen “budama yapılmış” karar ağacı yapısına bakıldığında ise dallanmanın osteomiyelit (kemik enfeksiyonu) değişkeninden başladığı bu değişkeni ise gangren, yara genişliği, trigliserid, yaş ve CRP değişkenlerinin takip ettiği görülmektedir. Ayrıca Şekil 3.16'da verilen Temel Bileşenler Analizi sonuçlarına göre de veri setindeki en önemli ilk beş değişkenin yara derinliği, osteomiyelit, gangren, enfeksiyon derecesinin şiddeti ve periferik arter hastalığının şiddeti değişkenleri olduğu saptanmıştır. Lojistik Regresyon modelindeki istatistiksel olarak anlamlı değişkenlere bakıldığında ise, CRP, ESR, hastanın geçmişteki amputasyon statüsü, hastanın HBOT tedavisi alıp almama durumu, HBOT tedavisi seans sayısı ve NPWT tedavisi statüsünü belirten değişkenler öne çıkmaktadır.

Bu sonuçlar incelendiğinde ise, elde edilen modellerin sonuçlarının **tıbbi literatürle ve birbirleri ile de uyumlu olduğu görülmektedir**. Yara derinliği, enfeksiyonun şiddeti, periferik arter hastalığının şiddeti, gangren, akut faz reaktanlarından olan CRP ve ESR ile birlikte enfeksiyon belirteçlerinden olan lökosit (WBC) değerlerinin yüksekliği, yara genişliği Wagner sınıflandırmasına göre evre 4 ve evre 5 nitelikleri geçmiş çalışmalarda da amputasyon riskini öngören bir nitelik olarak öne çıkmaktadır (Widatalla ve diğ., 2009; Yesil ve diğ., 2009; Barberan ve diğ., 2010; Uccioli ve diğ., 2010; Akinci ve diğ., 2011; Yekta ve diğ. 2011; Sun ve diğ. 2012; Bravo-Molina ve diğ., 2016; Ferreira ve diğ., 2018; Sarwar ve diğ., 2019; Sayiner ve diğ., 2019). Bununla birlikte, hastanın HBOT tedavisi alıp almama durumu ve HBOT tedavisi seans sayısı, trigliserid ve total kolesterol değişkenleri sistematik literatür taraması sonucunda geçmişte gerçekleştirilen lojistik regresyon modellerinde anlamlı olarak tespit edilen değişkenlerden biri olmayıp bu araştırma sonucunda makine öğrenmesi modellerinde kullanılan değişkenlerdir. HBOT'un diyabetik ayak enfeksiyonlarında yardımcı bir tedavi olarak kullanılmasının sağladığı yararları vurgulayan randomize çalışmalar bulunmakta olup HBOT, diyabetik ayak enfeksiyonlarında; iyileşme oranlarını arttırmakta, iyileşme süresini kısaltmakta, amputasyon oranlarını düşürmekte olduğunun Aktaş (2015) tarafından altı çizilmiştir. Bu araştırma sonucunda da diyabetik ayak enfeksiyonlarında HBOT tedavisinin rolü yapay zekâ makine öğrenmesi tekniklerinin kullanılması yoluyla tekrar açığa çıkarılmıştır. Araştırma sonucunda ortaya çıkarmış olduğumuz bu niteliklerin diyabetik ayak hastalarının amputasyon riski ile ilgili yapılacak olan gelecek çalışmalara ışık tutacağına inanıyoruz.

Sistematiik literatür taramasına ilişkin elde edilen 60 çalışmanın karşılaştırmalı tasnifinin yer aldığı Bölüm 2.8’de görüleceğı üzere bu doktora tez çalışmasının Türkiye’de ve Dünyada geniş yelpazedeki makine öğrenmesi sınıflandırma algoritmalarının kullanılması suretiyle yapılan ilk çalışmalardan biri olduğu bulgusuna ulaşılmıştır. Ayrıca, geçmiş çalışmalar incelendiğinde, ilgili çalışmaların tamamının diyabetik ayak hastalarının amputasyon riskini öngörmek için yapılmış çalışmalar olmadığı görülecektir. Geçmişte yapılmış çalışmaların içerisinde diyabetik ayak hastalarının ülserlerinin iyileşme şansını, ülser oluşma riskini veya mortalite riskini tahmin etmeye yönelik yapılmış farklı çalışmalar da bulunmuştur. Literatür incelemesinde tespit edilen uluslararası çalışmalarda multifaktöriyel bir hastalık olan diyabetik ayak ile ilişkili amputasyonların önlenmesine/saptanmasına yönelik yapılan çalışmaların büyük çoğunluğunda istatistiksel yöntemlerin ve sadece Lojistik Regresyon algoritmasının (makine öğrenmesi sınıflandırma algoritmaları arasında da bulunan) kullanıldığı görülmektedir. Bu algoritma dışında günümüzde sıklıkla kullanılmakta olan diğer makine öğrenmesi algoritmalarının kullanıldığı ve araştırma konusu ile doğrudan ilgili yalnızca iki çalışma saptanmış olup bunların birisi de görüntü işleme çalışmasıdır (Goyal ve diğ. 2020). Dolayısı ile yapısal veri ile gerçekleştirilen ve bu doktora tez çalışması ile karşılaştırılabilecek tek çalışmada ise sınırlı sayıda makine öğrenmesi sınıflandırma algoritması kullanıldığı görülmektedir (C5.0 ve boosted ensemble C5.0 algoritmaları; Kasbekar ve diğ. 2017). Bu çalışma sonucunda elde edilmiş olan en iyi modellerin performanslarının doğruluk açısından bakıldığında %100 ve %96 doğruluk oranına sahip modeller olduğu belirtilmiştir. Bununla birlikte makine öğrenmesi problemlerinde %100 doğruluk oranı her zaman istenen bir durum değildir. %100 doğruluk değerlerine sahip makine öğrenmesi modellerinin eğitim verilerini ezberlemesi sonucu ortaya çıkan aşırı öğrenme - “*overfitting*” problemi literatürde önemle üzerinde durulan bir konudur (Powell ve diğ. 2016; Vabalas ve diğ. 2019). İlgili çalışmada aşırı öğrenme durumunun olup olmadığı test edilmemiş olup bu problemin oluşmasının önlenmesi için hangi önlemlerin alındığı da paylaşılmamıştır. Dahası bu çalışmada doğruluk değeri dışındaki diğer performans değerlendirme ölçütleri de paylaşılmamıştır. Bu açılardan bakıldığında ilgili çalışmanın metodolojik olarak eksiklikleri olduğu düşünülmektedir.

Diyabetik ayak hastalarının muhtemel sonlanımların (iyileşme, amputasyon-minör amputasyon veya majör amputasyon, ölüm) uzun süre takip gerektirmesi ve hastane yatışlarının da buna paralel olarak oldukça uzun olması nedeniyle bu hastalığa yönelik veri seti oluşturmak araştırmaların gerçekleştirilmesindeki en büyük problem olarak karşımıza çıkmaktadır. Doktora tez çalışması için veri toplanan kliniğin özellikleri ve yoğun çalışmalar sonucunda

elde edilmiş veri setinin özellikleri ve çalışmanın özgün yönleri Bölüm 3.2 ve 3.3’ de paylaşılmıştır.

Doktora tez çalışması için orijinal veri setinin çeşitli alt kırılım veri seti senaryoları oluşturulmuş (nümerik veri seti, kategorik veri seti, nitelik seçimi yapılarak boyut sayısı azaltılmış alt kırılım orijinal, nümerik ve kategorik veri setleri), hedef niteliğimiz olan diyabetik ayak hastalarının hem genel amputasyon hem de majör ve minör amputasyon risklerini tahmin etmek için bu veri setlerine uygun makine öğrenmesi sınıflandırma algoritmaları kullanılmıştır. Sonuç olarak en iyi performansı gösteren modeller, nitelik seçimi yapılarak oluşturulmuş ve kategorik ve nümerik değerlerden oluşan alt kırılım veri setinde, temelinde topluluk öğrenme metodolojisi bulunan “*ensemble*” bir yöntem olan “Rastgele Orman Algoritması” ve “Lojistik Regresyon” algoritmaları olup, seçilen deney koşullarında diğer algoritmalara kıyasla daha iyi performans göstermiştir. Üçlü ayırma (%80 eğitim, %10 validasyon, %10 test) performans değerlendirme yönteminin kullanıldığı deney koşullarında Rastgele Orman Algoritması en yüksek doğruluğa (Tablo 4.32: validasyon veri setindeki doğruluk: 0,8780; test veri setindeki doğruluk: 0,7805) değerleri ile ulaşmıştır. İkili ayırma performans değerlendirme yönteminin kullanıldığı deney koşullarında (%90 eğitim, %10 test) Lojistik Regresyon algoritması en yüksek doğruluğa 0,90 değeri ile ulaşmıştır (Tablo 4.53). Ayrıca bu regresyon modeli Nagelkerke R² değerine göre ise oldukça iyi bir oran olan %69,4 oranında açıklanmaktadır (Tablo 4.54). Dahası, araştırma için kurulan lojistik model uygunluğu değerlendirildiğinde, Hoshmer-Lemeshow testi sonucunda uygun bir model olduğu saptanmıştır (Tablo 4.55). Önemle belirtmek gerekir ki bu sonuçlar yalnızca bu veri seti için ve seçilen deney koşullarında geçerli olup, veri setine başka kaynaklardan veri eklendiğinde, veri ön işleme tekniklerinde değişiklik yapıldığında, hiperparametresi bulunan algoritmalar için farklı hiperparametre optimizasyon teknikleri kullanıldığında veya geliştirilmekte olan yeni makine öğrenmesi sınıflandırma algoritmaları kullanıldığında da değişebilir.

Bu doktora tez çalışmasında aynı zamanda hedef niteliğin üç kategoriden oluştuğu (minör amputasyon, majör amputasyon ve amputasyon yok) nitelikler de makine öğrenmesi sınıflandırma algoritmaları kullanılarak tahmin edilmeye çalışılmıştır. Çok *kategorili* (“*multi-class*”) sınıflandırma için veri ön işleme adımında detaylı olarak ele alınan prosedürler gerçekleştirildikten (veri seti dengeleme işlemi yapılarak ve de yapılmadan) orijinal yani nümerik ve kategorik niteliklerin olduğu, veri setinin ayırıklaştırma prosedürü sonucunda tamamen kategorik niteliklerin yer aldığı ve “minör amputasyon” sınıfının “amputasyon yok”

sınıfına eklenerek “majör ampütasyon” tahmin probleminin tekrar ikili- “*binary*” sınıflandırma problemine dönüştürüldüğü “*one-versus-rest*” yaklaşımı ile oluşturulmuş veri setlerinde yalnızca üçlü ayırma (eğitim, validasyon[geçerleme] ve test) performans değerlendirme yaklaşımı kullanılmıştır. Sonuç olarak makine öğrenmesi algoritmalarının “minör ampütasyon” riskini tahmin etmede daha başarılı olduğu fakat “majör ampütasyon” riskini tahmin etmede dengeli performans sonuçları üretemediği ve kesinlik- pozitif öngörü değeri’nin çoğunlukla oldukça düşük değerde kaldığı saptanmıştır. Bu durumun “majör ampütasyon” operasyonu geçiren hasta sayısının yalnızca “31” hasta ile sınırlı olmasından kaynaklandığı düşünülmektedir. Böylelikle, araştırma sonucunda ortaya çıkarılan diğer bir bulgu da majör ampütasyon operasyonu gibi bir diyabetik ayak hastasının hayatını kökten değiştirebilecek bir cerrahi işlem ile ilişkili faktörlerin saptanması ve bu durumun önceden tahmin edilmesi için gelecek çalışmalara olan ihtiyacın yoğun olduğudur.

Elde edilen en iyi performansa sahip makine öğrenmesi modellerinin bir karar destek sistemi oluşturulmak ve bunların bir arayüze/sisteme entegrasyonu suretiyle uygulamaya geçmesi veri madenciliği projelerinin nihai hedeflerindendir. Bu doktora tez çalışmasında da R Shiny kullanıcı arayüzü geliştirilerek bu hedefe ulaşılmıştır. R-Shiny kullanıcı ara yüzüne genel- “overall” ampütasyonu tahmin etmek için oluşturulmuş ve en yüksek performansı gösteren modeller olan Rastgele Orman Algoritması ve Lojistik Regresyon modelleri entegre edilmiştir. Bu sayede yeni bir örnek geldiğinde ilgili algoritmaların tahmin/sınıflandırma yaparak ilgili örneğin hangi sınıfa ait olduğunu gösteren çıktı sonucu üretilmesi sağlanmıştır. Bu uygulamanın özellikle hekim sayısının yetersiz olduğu bölgelerdeki diyabetik ayak tedavisi ile uğraşan özellikle birinci ve ikinci basamak sağlık merkezleri ve klinikler başta olmak üzere çok geniş alanda kullanım imkânı olduğunu düşünülmektedir. Bu uygulama aynı zamanda üçüncü basamak sağlık merkezlerinde de araştırma amacıyla da kullanılabilir. Böylelikle hastaların ampütasyon operasyonu geçirmeden erken teşhis ve tedavi imkanına kavuşması ümit edilmektedir. Elbette bu sistem hiçbir hastaya hekim muayenesi olmadan teşhis koydurmayı amaçlamamakta bundan ziyade, hekim/sağlık personeline yardımcı bir sistem olması, böylelikle de teşhise ve/veya tedaviye destek olması hedeflenerek tasarlanmıştır. Web ortamına aktarılan uygulama, planlanan patent başvurusu nedeniyle halka açık olarak paylaşılmamıştır.

Bu tez çalışmasının limitasyonları sıralanacak olursa; çalışmanın retrospektif ve “*unblinded*” olmayan yapısı ve veri setinde kayıp değerlerin bulunması nedeniyle bazı kısıtlılıklarının bulunduğu söylenilebilir. Ayrıca veri girişi tek bir kullanıcı tarafından yapılmıştır. Yine de

araştırmadaki veri sayısı, uygulamalı makine öğrenmesi çalışmalarına kıyasla kısmen az gibi görünse de diyabetik ayak hastaların hastanede kalış sürelerinin karakteristik olarak oldukça uzun olmasından mütevellit, araştırmada kullanılan veri setinin gerek kayıt sayısının gerekse de nitelik sayısının benzer çalışmalara göre son derece yeterli hatta üstün olduğu düşünülmektedir.

Sonuç olarak, bu araştırma bulgularının diyabetik ayak ile ilişkili amputasyon riskinin önlenmesi için gerçekleştirilecek olan ileriki araştırmalar için yol gösterici olabileceğini düşünüyoruz. Ayrıca bu çalışmanın tıp ve sağlık alanındaki verilerin analizinde yapay zekâ ve makine öğrenmesi metotlarının kullanımının yaygınlaşmasına katkı sağlaması açısından da önemli bir araştırma olduğunu düşünmekteyiz. Doktorlar veya sağlık/tıp alanındaki diğer araştırmacıların yapay zekâ- makine öğrenmesi yöntemleri ile veri analizine ilişkin farkındalıklarının artması ve veri bilimciler ile birlikte çalışması, tıbbi verilerinin daha verimli bir şekilde analiz edilmesini sağlayacak ve elde edilen sonuçların özellikle erken teşhis ve tedavi aşamalarında hastalara faydalı olacağı umulmaktadır. Önemli bir halk sağlığı problemi olan diyabetin ve diyabet ile ilişkili komplikasyonlarının önlenmesi ve özellikle diyabetik ayak ile ilgili amputasyonların önlenmesi halk sağlığı için tartışmasız gereklidir. Böylelikle diyabetik ayak hastalarının amputasyon risklerinin tespiti ve öngörüsü ile önleyici tedbirlerin alınmasına katkı sağlanabilecektir. Dolayısıyla bu çalışmanın, insan sağlığına ve toplumsal refaha olumlu etkilerinin yanı sıra sağlık harcamalarının azaltılmasına ve ilişkili diğer süreçlere (zaman, üretim, verimli kaynak kullanımı, vb.) de pozitif yönde bir etki yapabilecek nitelikte bir doktora tez çalışması olduğu düşünülmektedir.

Bu doktora tez çalışması ile Türkiye’de ilk kez geniş çapta makine öğrenmesi sınıflandırma algoritmaları kullanılarak diyabetik ayak hastalarının amputasyon riskine ilişkin tahmin modelleri oluşturulmuş ve bir web uygulaması olarak paylaşılmıştır. Bu araştırma bildiğimiz ve tespit edebildiğimiz kadarıyla dünyada da belirtilen kapsamda gerçekleştirilen ilk araştırmalardan olma özelliğini taşımaktadır. Araştırma bulgularının öncelikle diyabetik ayak ile ilgili amputasyonların önlenmesine ve diyabetik ayak hastalarına, sağlık personellerine ve de bilgisayar bilimleri/biyoistatistik/enformatik alanında çalışan araştırmacılara faydalı olmasını ümit ediyoruz.

6. KAYNAKLAR

- Adeleye, O.O., Ugwu, E.T., Gezawa, I.D., Okpe, I., Ezeani, I., Enamino, M., 2020. Predictors of intra-hospital mortality in patients with diabetic foot ulcers in Nigeria: data from the MEDFUN study. *BMC Endocr Disord* 20, 134.
- Agrawal, T., 2021. *Hyperparameter Optimization in Machine Learning: Make Your Machine Learning and Deep Learning Models More Efficient*. Apress, Berkeley, CA.
- Ahroni, J.H., Boyko, E.J., Forsberg, R.C., 1999. Clinical correlates of plantar pressure among diabetic veterans. *Diabetes Care* 22, 965–972.
- Ai, L., Hu, Y., Zhang, X., Zeng, H., Zhao, Jun, Zhao, Jungong, Chai, Y., Lu, J., Tang, J., Bao, Y., Liu, F., Jia, W., 2016. High cystatin C levels predict undesirable outcome for diabetic foot ulcerations: High serum Cystatin C tied to unfavorable outcome of foot ulcerations in type 2 diabetes. *Wound Rep and Reg* 24, 560–567.
- Akgül, M., Sönmez, Ö.E., Özcan, T., 2020. Diagnosis of Heart Disease Using an Intelligent Method: A Hybrid ANN – GA Approach, in: Kahraman, C., Cebi, S., Cevik Onar, S., Oztaysi, B., Tolga, A.C., Sari, I.U. (Eds.), *Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making*. Springer International Publishing, Cham, pp. 1250–1257.
- Akinci, B., Yener, S., Yesil, S., Yapar, N., Kucukyavas, Y., Bayraktar, F., 2011. Acute Phase Reactants Predict the Risk of Amputation in Diabetic Foot Infection. *Journal of the American Podiatric Medical Association* 101, 1–6.
- Akpınar, H., 2017. *DATA, Veri Madenciliği, Veri Analizi*, 2. Basım, Eylül 2017. ed. Papatya Yayıncılık Eğitim, İstanbul, Türkiye.
- Akpınar, H., 2014. *Data: Veri Madenciliği Veri Analizi*, 1st ed. Papatya Yayıncılık Eğitim, İstanbul.
- Akpınar, H., 2000. VERİ TABANLARINDA BİLGİ KEŞFİ ve VERİ MADENCİLİĞİ. *İstanbul Üniversitesi İşletme Fakültesi Dergisi* 29, 1–22.
- Aktaş, Ş., 2015. Diyabetik ayakta fiziksel yardımcı tedavi yöntemleri. *TOTBID Dergisi* 14.
- Aktaş, Ş., 2010. *Kronik Yara Tedavisinde Vakum Yardımlı Kapama*, in: Topalan, M., Aktaş, Ş. (Eds.), *Güncel Yönleriyle Kronik Yara*. İstanbul Üniversitesi Tıp Fakültesi Kronik Yara Konseyi, ANKARA.
- Aktas, S., Baktiroglu, S., Demir, L., Kılıçoğlu, Ö., Topalan, M., Güven, E., Mirasoğlu, B., Yanar, F., 2016. INTRALESIONAL APPLICATION OF EPIDERMAL GROWTH FACTOR IN LIMB THREATENING ISCHAEMIC DIABETIC FOOT ULCERS. *ACTA ORTHOPAEDICA et TRAUMATOLOGICA TURCICA* 50, 277–283.
- ALP, S., 2019. *Yapay Sinir Ağları*, in: ALP, S., Öz, E. (Eds.), *Makine Öğrenmesinde Sınıflandırma Yöntemleri ve R Uygulamaları*. NOBEL AKADEMİK YAYINCILIK, ANKARA-Türkiye, p. 216.
- Alpaydın, E., 2013. *Yapay Öğrenme | Introduction to Machine Learning*. Boğaziçi Üniversitesi Yayınevi.
- Altay, F.A., Kuzi, S., Altay, M., Ateş, İ., Gürbüz, Y., Tütüncü, E.E., Şentürk, G.Ç., Altın, N., Şencan, İ., 2019. Predicting diabetic foot ulcer infection using the neutrophil-to-lymphocyte ratio: a prospective study. *J Wound Care* 28, 601–607.
- American Diabetes Association, 2014. Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 37, S81–S90.
- Anaconda Inc., 2020. *Anaconda Software Distribution*.
- Antoniou, A., Lu, W.-S., 2021. *Practical Optimization: Algorithms and Engineering Applications*, Texts in Computer Science. Springer US, New York, NY.

- BAĞIŞ, S., 2009. *YAPAY ZEKA ALGORİTMALARI KULLANILARAK SİSTEM MODELLEME* (Yüksek Lisans Tezi). Erciyes Üniversitesi, YÖK, Erciyes Üniversitesi.
- Balaban, M.E., Kartal, E., 2018. *Veri Madenciliği ve Makine Öğrenmesi*, İkinci Baskı. ed. Çağlayan Kitabevi, İSTANBUL.
- Barberan, J., Granizo, J.-J., Aguilar, L., Alguacil, R., Sainz, F., Menendez, M.-A., Gimenez, M.-J., Martínez, D., Prieto, J., 2010. Predictive model of short-term amputation during hospitalization of patients due to acute diabetic foot infections. *Enfermedades Infecciosas y Microbiología Clínica* 28, 680–684.
- Bayes, T., 1763. LII. An essay towards solving a problem in the doctrine of chances. By. *Philosophical transactions of the Royal Society of London* 53, 370–418.
- Beaney, A.J., Nunney, I., Gooday, C., Dhatariya, K., 2016. Factors determining the risk of diabetes foot amputations – A retrospective analysis of a tertiary diabetes foot care service. *Diabetes Research and Clinical Practice* 114, 69–74.
- Berli, M.C., Wanivenhaus, F., Kabelitz, M., Götschi, T., Böni, T., Rancic, Z., Waibel, F.W.A., 2019. Predictors for reoperation after lower limb amputation in patients with peripheral arterial disease. *Vasa* 48, 419–424.
- Berrar, D., 2019. Bayes' Theorem and Naive Bayes Classifier, in: *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier, pp. 403–412.
- Bewick, V., Cheek, L., Ball, J., 2005. Statistics review 14: Logistic regression. *Crit Care* 9, 112.
- Beyaz, S., Güler, Ü.Ö., Bağır, G.Ş., 2017. Factors affecting lifespan following below-knee amputation in diabetic patients. *Acta Orthopaedica et Traumatologica Turcica* 51, 393–397.
- Birattari, M., 2009. *Tuning Metaheuristics, Studies in Computational Intelligence*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Bishop, C.M., 2006. *Pattern Recognition and Machine Learning*. Springer.
- Bravo-Molina, A., Linares-Palomino, J.P., Lozano-Alonso, S., Asensio-García, R., Ros-Díe, E., Hernández-Quero, J., 2016. Influence of wound scores and microbiology on the outcome of the diabetic foot syndrome. *Journal of Diabetes and its Complications* 30, 329–334.
- Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32.
- Breiman, L., 1996. Bagging predictors. *Mach Learn* 24, 123–140.
- Breiman, L. (Ed.), 1984. *Classification and regression trees*, 1. CRC Press repr. ed. Chapman & Hall/CRC, Boca Raton, Fla.
- Brownlee, M., 2005. The Pathobiology of Diabetic Complications: A Unifying Mechanism. *Diabetes* 54, 1615–1625.
- Bulut, M., 2019. ÖNSÖZ, in: Bulut, M., Dilmen, N., Gökhan Bora, E., Gezer, M., Erol Selçukcan, Ç., Şener Türker, L. (Eds.), *Sağlıkta Yapay Zekâ*. Çağlayan Kitabevi, İstanbul, Türkiye, p. 307.
- Bustamante, C., Garrido, L., Soto, R., 2006. Comparing Fuzzy Naive Bayes and Gaussian Naive Bayes for Decision Making in RoboCup 3D, in: Carbonell, E.J.G., Siekmann, J. (Eds.), *MICAI 2006: Advances in Artificial Intelligence-Lecture Notes in Artificial Intelligence*. Presented at the 5th Mexican International Conference on Artificial Intelligence Apizaco, Mexico, November 13-17, 2006 Proceedings, Springer-Verlag, Berlin Heidelberg NewYork, pp. 237–247.
- ÇATAK, F.Ö., 2018. *DESTEK VEKTÖR MAKİNESİ*.
- Çatak, F.Ö., 2015. Secure Multi-party Computation Based Privacy Preserving Extreme Learning Machine Algorithm Over Vertically Distributed Data, in: Arik, S., Huang, T., Lai, W.K., Liu, Q. (Eds.), *Neural Information Processing, Lecture Notes in Computer Science*. Springer International Publishing, Cham, pp. 337–345.

- Ceballos, F., 2020. *Stacking Classifiers for Higher Predictive Performance. TowardsDataScience*. <https://towardsdatascience.com/stacking-classifiers-for-higher-predictive-performance-566f963e4840> (Ziyaret Tarihi: 1.28.22).
- Chan, Y.H., 2005. Biostatistics 305. Multinomial logistic regression. *Singapore Med J, Basic Statistics For Doctors* 46, 11.
- Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B., 2021a. *Package 'shiny.'*
- Chang, W., Cheng, J., Allaire, J., Xie, Y., McPherson, J., 2021b. Shiny: web application framework for R.
- Chaudhary, S., Bhansali, A., Rastogi, A., 2019. Mortality in Asian Indians with Charcot's neuroarthropathy: a nested cohort prospective study. *Acta Diabetol* 56, 1259–1264.
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., Yuan, J., 2022. *Package 'xgboost.'* CRAN.
- Choi, J.J., Lee, J.Y., Lee, K.C., Jo, Y.Y., 2015. Preanesthetic mortality prediction in diabetics undergoing major lower limb amputation at a tertiary referral hospital: Implications of preoperative echocardiographic and laboratory values. *Journal of Critical Care* 30, 1021–1024.
- Chollet, F., 2015. *Keras*.
- Chomboon, K., Chujai, P., Teerarassammee, P., Kerdprasop, K., Kerdprasop, N., 2015. An Empirical Study of Distance Metrics for k-Nearest Neighbor Algorithm, in: *The Proceedings of the 2nd International Conference on Industrial Application Engineering 2015*. Presented at the International Conference on Industrial Application Engineering 2015, The Institute of Industrial Applications Engineers, pp. 280–285.
- Chuan, F., Tang, K., Jiang, P., Zhou, B., He, X., 2015. Reliability and Validity of the Perfusion, Extent, Depth, Infection and Sensation (PEDIS) Classification System and Score in Patients with Diabetic Foot Ulcer. *PLOS ONE* 10, e0124739.
- Cinsdikici, M.G., 2016. *Yapay Sinir Ağları*.
- Claesen, M., Simm, J., Popovic, D., Moreau, Y., De Moor, B., 2014. Easy Hyperparameter Search Using Optunity. arXiv:1412.1114 [cs].
- Çokluk, Ö., 2010. Lojistik Regresyon Analizi: Kavram ve Uygulama. *Kuram ve Uygulamada Eğitim Bilimleri/Educational Sciences: Theory & Practice* 10, 1357–1407.
- Coppini, D.V., Young, P.J., Weng, C., Macleod, A.F., Sönksen, P.H., 1998. Outcome on diabetic foot complications in relation to clinical examination and quantitative sensory testing: a case-control study. *Diabetic Medicine* 15, 765–771.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach Learn* 20, 273–297.
- Coşansu, G., 2009. 21. Yüzyılın Sağlık Krizi: Diyabet. *Florence Nightingale Hemşirelik Dergisi* 17, 115–122.
- Craw, S., 2010. Manhattan Distance, in: Sammut, C., Webb, G.I. (Eds.), *Encyclopedia of Machine Learning*. Springer US, Boston, MA, pp. 639–639.
- Cura, T., 2008. *Modern Sezgisel Teknikler ve Uygulamaları*, 1.Baskı. ed. Papatya Yayıncılık Eğitim, İstanbul.
- Dasgupta, A., Nath, A., 2016. Classification of Machine Learning Algorithms. *International Journal of Innovative Research in Advanced Engineering* 3, 6–11.
- Davazdahemami, B., Delen, D., 2019. The confounding role of common diabetes medications in developing acute renal failure: A data mining approach with emphasis on drug-drug interactions. *Expert Systems with Applications* 123, 168–177.
- Dayton, C.M., 1992. LOGISTIC REGRESSION ANALYSIS. *Lecture Notes* 9.
- Delen, D., 2009. Analysis of cancer data: a data mining approach. *Expert Systems* 26, 100–112.

- Delen, D., Davazdahemami, B., Eryarsoy, E., Tomak, L., Valluru, A., 2019. Using predictive analytics to identify drug-resistant epilepsy patients. *Health Informatics Journal* 146045821983312.
- Dietterich, T.G., 2000. *Ensemble Methods in Machine Learning*, in: Kittler, J., Roli, F. (Eds.), *Multiple Classifier Systems: First International Workshop, MCS 2000, Cagliari, Italy, June 21-23, 2000: Proceedings, Lecture Notes in Computer Science. Presented at the International Workshop on Multiple Classifier Systems*, Springer, Berlin ; New York, pp. 1–15.
- Doğaner, A., 2020. *TOPLULUK ÖĞRENME YÖNTEMLERİ İLE RENAL HÜCRELİ KARSİNOMUN TAHMİN EDİLMESİ* (Doctoral Thesis). İNÖNÜ ÜNİVERSİTESİ/MALATYA/TÜRKİYE, Yükseköğretim Kurulu.
- Dunham, M.H., 2003. *Data Mining: Introductory and Advanced Topics*, 1st Edition. ed. Pearson.
- Edelman, D., Hough, D.M., Glazebrook, K.N., Oddone, E.Ž., 1997. Prognostic value of the clinical examination of the diabetic foot ulcer. *J Gen Intern Med* 12, 537–543.
- Eğrioğlu, E., Yolcu, U., Baş, E., 2020. *YAPAY SİNİR AĞLARI Öngörü ve Tahmin Uygulamaları*, 2.Baskı. ed. NOBEL AKADEMİK YAYINCILIK, ANKARA-Türkiye.
- Emre, İ.E., Erol, N., Ayhan, Y.İ., Özkan, Y., Erol, Ç., 2019. The analysis of the effects of acute rheumatic fever in childhood on cardiac disease with data mining. *International Journal of Medical Informatics* 123, 68–75.
- Erdman, W.A., Bueth, J., Bhore, R., Ghayee, H.K., Thompson, C., Maewal, P., Anderson, J., Klemow, S., Oz, O.K., 2012. Indexing Severity of Diabetic Foot Infection With 99mTc-WBC SPECT/CT Hybrid Imaging. *Diabetes Care* 35, 1826–1831.
- Ersungur, M., Kiziltan, A., Polat, Ö., 2007. TÜRKİYE’DE BÖLGELERİN SOSYO-EKONOMİK GELİŞMİŞLİK SİRALAMASI: TEMEL BİLEŞENLER ANALİZİ. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi* 21, 55–66.
- Ertuğrul, M.B., 2020. Diyabetik Ayak Enfeksiyonunda Osteomyelit Tanısı, in: Mutlu, M., Ertugrul, M.B. (Eds.), *Diyabetik Ayak Problemleri, Enfeksiyonlar ve Charcot Nöroosteoartropatisi. Türkiye Klinikleri*, Ankara, pp. 84–88.
- Faglia, E., Clerici, G., Caminiti, M., Quarantiello, A., Curci, V., Morabito, A., 2007. Predictive Values of Transcutaneous Oxygen Tension for Above-the-ankle Amputation in Diabetic Patients with Critical Limb Ischemia. *European Journal of Vascular and Endovascular Surgery* 33, 731–736.
- Farhangfar, A., Kurgan, L.A., Pedrycz, W., 2007. A Novel Framework for Imputation of Missing Values in Databases. *IEEE Trans. Syst., Man, Cybern. A* 37, 692–709. <https://doi.org/10.1109/TSMCA.2007.902631>
- Fayyad, U., Piatetsky-Shapiro, Smyth, P., 1996a. The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM* 39, 27–34.
- Fayyad, U., Shapiro, G.P., Smyth, P., 1996b. From Data Mining to Knowledge Discovery in Databases. *AI Magazine* 17, 18.
- Fernandez-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research* 15, 3133–3181.
- Ferreira, L., Carvalho, A., Carvalho, R., 2018. Short-term predictors of amputation in patients with diabetic foot ulcers. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 12, 875–879.
- Fife, C.E., Horn, S.D., Smout, R.J., Barrett, R.S., Thomson, B., 2016. A Predictive Model for Diabetic Foot Ulcer Outcome: The Wound Healing Index. *Advances in Wound Care* 5, 279–287.

- Fix, E., Hodges, J.L., 1951. *Discriminatory analysis, nonparametric discrimination: Consistency properties*. (Technical Report 4). USAF School of Aviation Medicine, Randolph Field.
- Fleischer, A.E., Wrobel, J.S., Leonards, A., Berg, S., Evans, D.P., Baron, R.L., Armstrong, D.G., 2011. Post-treatment Leukocytosis Predicts an Unfavorable Clinical Response in Patients with Moderate to Severe Diabetic Foot Infections. *The Journal of Foot and Ankle Surgery* 50, 541–546.
- Fortin, F.-A., De Rainville, F.-M., Gardner, M.-A., Parizeau, M., Gagné, C., 2012. DEAP: Evolutionary Algorithms Made Easy. *Journal of Machine Learning Research* 12, 2171–2175.
- Freund, Y., Schapire, R.E., 1997. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences* 55, 119–139.
- Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. *Ann. Statist.* 29. <https://doi.org/10.1214/aos/1013203451>
- García, S., Luengo, J., Herrera, F., 2015. *Data Preprocessing in Data Mining, Intelligent Systems Reference Library*. Springer International Publishing, Cham.
- Gazzaruso, C., Gallotti, P., Pujia, A., Montalcini, T., Giustina, A., Coppola, A., 2021. Predictors of healing, ulcer recurrence and persistence, amputation and mortality in type 2 diabetic patients with diabetic foot: a 10-year retrospective cohort study. *Endocrine* 71, 59–68.
- George, D., Mallery, P., 2010. *SPSS for Windows step by step : a simple guide and reference, 17.0 update /*, 10th ed. Boston : Allyn & Bacon, Boston :
- Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. *Mach Learn* 63, 3–42.
- Gogna, A., Tayal, A., 2013. Metaheuristics: review and application. *Journal of Experimental & Theoretical Artificial Intelligence* 25, 503–526.
- González, S., García, S., Del Ser, J., Rokach, L., Herrera, F., 2020. A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. *Information Fusion* 64, 205–237.
- Goyal, M., Reeves, N.D., Rajbhandari, S., Ahmad, N., Wang, C., Yap, M.H., 2020. Recognition of ischaemia and infection in diabetic foot ulcers: Dataset and techniques. *Computers in Biology and Medicine* 117.
- Griffin, K.J., Rashid, T.S., Bailey, M.A., Bird, S.A., Bridge, K., Scott, J.D.A., 2012. Toe Amputation: A predictor of future limb loss? *Journal of Diabetes and Its Complications* 251–254.
- Gülseçen, S., Demirkol, D., 2019. *Semiyotik , İnsan Bilgisayar Etkileşimi ve Yapay Zeka, in: Sağlık Bilimlerinde Yapay Zekâ*. Çağlayan Kitabevi, İstanbul, Türkiye, p. 307.
- Haidar, A., Field, M., Sykes, J., Carolan, M., Holloway, L., 2021. PSPSO: A package for parameters selection using particle swarm optimization. *SoftwareX* 15, 100706.
- Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., 2009. *Multivariate Data Analysis*, 7th Edition. ed. Pearson Prentice Hall.
- Han, J., Kamber, M., Pei, J., 2012. *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufman Publishers, MA, USA.
- Hardy, M.A., 1993. *Regression with Dummy Variables, Series: Quantitative Applications in the Social Sciences*. Sage Publications, Newbury Park, CA: Sage.
- Harrington, P., 2012. *Machine learning in action*. Manning Publications Co, Shelter Island, N.Y.
- Hastie, T., Tibshirani, R., Friedman, J., 2001. *The Elements of Statistical Learning Data Mining, Inference, and Prediction*, Second Edition. ed. Springer.
- Haykin, S.S., 2009. *Neural networks and learning machines*, 3rd ed. ed. Prentice Hall, New York.

- Henig, O., Pogue, J.M., Martin, E., Hayat, U., Ja'ara, M., Kilgore, P.E., Cha, R., Dhar, S., Kaye, K.S., 2020. The Impact of Multidrug-Resistant Organisms on Outcomes in Patients With Diabetic Foot Infections. *OFID* 7, ofaa161.
- Hilbe, J.M., 2015. *Practical Guide to Logistic Regression*. CRC Press Taylor & Francis Group, 6000 Broken Sound Parkway NW, Suite 300 Boca Raton.
- Hilbe, J.M., 2011. Logistic Regression.
- Holland, J.H., 1973. Genetic Algorithms and the Optimal Allocation of Trials. *SIAM J. Comput.* 2, 88–105.
- Hopfield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences* 79, 2554–2558.
- Hosmer, D., Lemeshow, S., 2000. *Applied Logistic Regression*, Second Edition. ed. John Wiley and Sons, New York, NY, USA.
- Hotelling, H., 1933. ANALYSIS OF A COMPLEX OF STATISTICAL VARIABLES INTO PRINCIPAL COMPONENTS. *Journal of educational psychology* 24, 417–441. <https://doi.org/doi:10.1037/h0071325>
- Hssina, B., Merbouha, A., Ezzikouri, H., Erritali, M., 2014. A comparative study of decision tree ID3 and C4.5. *SpecialIssue* 4. <https://doi.org/10.14569/SpecialIssue.2014.040203>
- Hu, L.-Y., Huang, M.-W., Ke, S.-W., Tsai, C.-F., 2016. The distance function effect on k-nearest neighbor classification for medical datasets. *SpringerPlus* 5, 1304.
- Huang, G.-B., 2014. An Insight into Extreme Learning Machines: Random Neurons, Random Features and Kernels. *Cogn Comput* 6, 376–390.
- Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006. Extreme learning machine: Theory and applications. *Neurocomputing* 70, 489–501.
- Hunt, N., Tyrrell, S., 2004. *Stratified Sampling* <https://archive.ph/20131013132818/http://nestor.coventry.ac.uk/~nhunt/meths/strati.html#selection-531.0-531.20> (Ziyaret tarihi: 2.4.22).
- Hüßers, J., Hafer, G., Heggemann, J., Wiemeyer, S., John, S.M., Hübner, U., 2020. Predicting the amputation risk for patients with diabetic foot ulceration – a Bayesian decision support tool. *BMC Med Inform Decis Mak* 20, 200.
- Husson, F., Josse, J., Le, S., Mazet, J., 2020. *Package 'FactoMineR.'*
- International Diabetes Federation, 2021. *IDF Diabetes Atlas (No. 10th Edition)*. International Diabetes Federation.
- International Diabetes Federation, 2019. *IDF DIABETES ATLAS (No. Ninth Edition)*. International Diabetes Federation.
- Itano, F., de Abreu de Sousa, M.A., Del-Moral-Hernandez, E., 2018. Extending MLP ANN hyper-parameters Optimization by using Genetic Algorithm, in: 2018 International Joint Conference on Neural Networks (IJCNN). *2018 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Rio de Janeiro, pp. 1–8.
- Jahromi, A.H., Taheri, M., 2017. A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features. *2017 Artificial Intelligence and Signal Processing Conference (AISP)*, IEEE, pp. 209–212.
- James V. Miranda, L., 2018. PySwarms: a research toolkit for Particle Swarm Optimization in Python. *JOSS* 3, 433.
- Japkowicz, N., Shah, M., 2011. *Evaluating Learning Algorithms, A Classification Perspective*, First Edition. ed. Cambridge University Press, 32 Avenue of the Americas, New York, NY 10013-2473, USA.
- Johnson, D., 2022. Back Propagation Neural Network: What is Backpropagation Algorithm in Machine Learning? Guru99. <https://www.guru99.com/backpropagation-neural-network.html> (Ziyaret tarihi: 1.24.22).

- Jolliffe, I.T., 1995. Sample Sizes and the Central Limit Theorem: The Poisson Distribution as an Illustration. *The American Statistician* 49, 269–269.
- Kagiyama, N., Shrestha, S., Farjo, P.D., Sengupta, P.P., 2019. Artificial Intelligence: Practical Primer for Clinical Research in Cardiovascular Disease. *Journal of the American Heart Association* 8, 12.
- Kaleta, J.L., Fleischli, J.W., Reilly, C.H., 2001. The Diagnosis of Osteomyelitis in Diabetes Using Erythrocyte Sedimentation Rate. *Journal of the American Podiatric Medical Association* 91, 6.
- Kaplan, J., 2020. Package ‘fastDummies.’
- Karaboğa, D., 2018. *Yapay Zekâ Optimizasyon Algoritmaları*, 6. Baskı. ed. NOBEL AKADEMİK YAYINCILIK.
- Karatzoglou, A., Smola, A., Hornik, K., 2019. Package ‘kernlab.’
- Kartal, E., 2015. *Sınıflandırmaya Dayalı Makine Öğrenmesi Teknikleri ve Kardiyolojik Risk Değerlendirmesine İlişkin Bir Uygulama (Doctoral Thesis)*. İstanbul Üniversitesi, Enformatik Bölümü, İstanbul.
- Kasbekar, P.U., Goel, P., Jadhav, S.P., 2017. A Decision Tree Analysis of Diabetic Foot Amputation Risk in Indian Patients. *Front. Endocrinol.* 8, 1–6.
- Kassambara, A., Mundt, F., 2020. Package ‘factoextra.’ CRAN.
- Kennedy, J., Eberhart, R., 1995. Particle Swarm Optimization. *Proceedings of IEEE International Conference on Neural Networks*, IEEE, Perth, Australia, p. 7.
- Kim, K., Hong, J., 2017. A hybrid decision tree algorithm for mixed numeric and categorical data in regression analysis. *Pattern Recognition Letters* 98, 39–45. <https://doi.org/10.1016/j.patrec.2017.08.011>
- Kiranyaz, S., Ince, T., Gabbouj, M., 2014. *Multidimensional Particle Swarm Optimization for Machine Learning and Pattern Recognition, Adaptation, Learning, and Optimization*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Kılıçoğlu, Ö.İ., Demirel, M., Aktaş, Ş., 2018. New trends in the orthopaedic management of diabetic foot. *EFORT Open Reviews* 3, 269–277.
- Koçoğlu, F.Ö., 2017. *MÜŞTERİ KAYIP ANALİZİ PROBLEMİNİN ÇÖZÜMÜNDE ANALİTİK YAKLAŞIMLAR (Doctoral Thesis)*. İstanbul Üniversitesi, Enformatik Bölümü, YÖK, İstanbul Üniversitesi.
- Krogh, A., 2008. What are artificial neural networks? *Nat Biotechnol* 26, 195–197.
- Kuhn, M., 2022. Package ‘caret.’
- Lanting, S.M., Twigg, S.M., Johnson, N.A., Baker, M.K., Caterson, I.D., Chuter, V.H., 2017. Non-invasive lower limb small arterial measures co-segregate strongly with foot complications in people with diabetes. *Journal of Diabetes and its Complications* 31, 589–593.
- Larose, D.T., Larose, C., 2015. *Data Mining and Predictive Analytics*, Second Edition. ed. John Wiley & Sons, Inc., Hoboken, New Jersey, New Jersey(USA)& Canada.
- Lavanya, D., Rani, K.U., 2011. Performance Evaluation of Decision Tree Classifiers on Medical Datasets. *IJCA* 26, 1–4.
- Lavery, L.A., Barnes, S.A., Keith, M.S., Seaman, J.W., Armstrong, D.G., 2008. Prediction of Healing for Postoperative Diabetic Foot Wounds Based on Early Wound Area Progression. *Diabetes Care* 31, 26–29.
- Lawrence, A.L., Armstrong, D.G., Lawrence, B.H., 1996. Classification of diabetic foot wounds. *THE JOURNAL OF FOOT AND ANKLE SURGERY* 35, 4.
- Lee, J.J., Kim, Y.S., 2010. Development of advanced pattern recognition model for evaluation of lateral displacement on soft ground using support vector machine. *KSCE J Civ Eng* 14, 173–182.

- Lee, T.-Y., Kim, K.-B., Han, S.-K., Jeong, S.-H., Dhong, E.-S., 2019. Skin Hydration Level as a Predictor for Diabetic Wound Healing: A Retrospective Study. *Plastic and Reconstructive Surgery* 143, 848e–856e.
- Leijnen, S., Veen, F. van, 2020. The Neural Network Zoo. *Proceedings* 47, 9.
- Lewis, D.D., 1998. *Naive (Bayes) at forty: the independence assumption in information retrieval*, in: Nédellec, C., Rouveirol, C. (Eds.), *Machine Learning: ECML-98*. Presented at the 10th European Conference on Machine Learning, Springer, Berlin/Heidelberg, Chemnitz, Germany, pp 4-15.
- Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J., Liu, H., 2018. Feature Selection: A Data Perspective. *ACM Comput. Surv.* 50, 1–45.
- Li, S., Zhang, X., 2020. Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm. *Neural Comput & Applic* 32, 1971–1979.
- Liaw, A., Wiener, M., 2022. *Package ‘randomForest.’*
- Lim, K., Lim, X., Hong, Q., Yong, E., Chandrasekar, S., Tan, G.W.L., Lo, Z.J., 2020. Use of home negative pressure wound therapy in peripheral artery disease and diabetic limb salvage. *Int Wound J* 17, 531–539.
- Lin, K., Hu, Y., Kong, G., 2019. Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model. *International Journal of Medical Informatics* 125, 55–61.
- Liu, Y., Wang, Y., Zhang, J., 2012. *New Machine Learning Algorithm: Random Forest*, in: Liu, B., Ma, M., Chang, J. (Eds.), *Information Computing and Applications, Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 246–252.
- Loh, W., 2011. Classification and regression trees. *WIREs Data Mining Knowl Discov* 1, 14–23.
- Louppe, G., Geurts, P., 2012. *Ensembles on Random Patches*, in: Flach, P.A., De Bie, T., Cristianini, N. (Eds.), *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2012, Bristol, UK, September 24-28, 2012. Proceedings, Part I, Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg
- Lunardon, N., Menardi, G., Torelli, N., Lunardon, N., 2021. ROSE: Random Over-Sampling Examples.
- Macfarlane, R.M., Jeffcoate, W.J., 1999. Classification of diabetic foot ulcers: The S(AD) SAD System. *The Diabetic Foot*, Vol 2, No 4
- Mahesh, B., 2020, Machine Learning Algorithms - A Review, *International Journal of Science and Research*, 9, 7.
- Marschner, I., Donoghoe, M., 2018. *Package ‘glm2.’*
- Martínez-Ramón, M., Christodoulou, C., 2006. Support Vector Machines for Antenna Array Processing and Electromagnetics. *Synthesis Lectures on Computational Electromagnetics* 1, 1–120.
- Mayer, M., 2022. *Package ‘splitTools.’*
- Mayfield, J.A., Caps, M.T., Boyko, E.J., Ahroni, J.H., Smith, D.G., 2002. Relationship of medial arterial calcinosis to autonomic neuropathy and adverse outcomes in a diabetic veteran population. *Journal of Diabetes and its Complications* 16, 165–171.
- Mayfield, J.A., Reiber, G.E., Nelson, R.G., Greene, T., 1996. A Foot Risk Classification System to Predict Diabetic Amputation in Pima Indians. *Diabetes Care* 19, 704–709.
- Meggitt, B., 1976. Surgical Management of Diabetic Foot. *Br J Hosp Med* 227–332.
- Memiş, A., 2015. *DIYABETİK AYAK HASTALARINDA SINIFLANDIRMA VERİLERİNİN YATIŞ SÜRESİ VE PROGNOZLA İLİŞKİLERİNİN ARAŞTIRILMASI*. Gülhane Askeri Tıp Akademisi (GATA) HAYDARPAŞA EĞİTİM HASTANESİ-SUALTI HEKİMLİĞİ VE HİBERBARİK TIP SERVİSİ, İSTANBUL.

- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., 2021. *Package 'e1071.'*
- Miller, J.L.L., Erickson, M.L., 1974. On Dummy Variable Regression Analysis: A Description and Illustration of the Method. *Sociological Methods & Research* 2, 409–430.
- Mineoka, Y., Ishii, M., Hashimoto, Y., Yamashita, A., Nakamura, N., Fukui, M., 2019. Platelet to lymphocyte ratio correlates with diabetic foot risk and foot ulcer in patients with type 2 diabetes. *Endocr J* 66, 905–913.
- Mitchell, T.M., 1997. *Machine Learning*. McGraw-Hill Science/Engineering/Math.
- Moura Neto, A., Zantut-Wittmann, D.E., Fernandes, T.D., Nery, M., Parisi, M.C.R., 2013. Risk factors for ulceration and amputation in diabetic foot: study in a cohort of 496 patients. *Endocrine* 44, 119–124.
- Mouselimis, L., Gosso, A., 2022. *The Extreme Learning Machine Algorithm- R package*.
- Müssel, C., Lausser, L., Kestler, H., 2021. *R Package 'TunePareto.'*
- Mustoe, T.A., Pierce, G.F., Morishima, C., Deuel, T.F., 1991. Growth factor-induced acceleration of tissue repair through direct and inductive activities in a rabbit dermal ulcer model. *Journal of Clinical Investigation* 87, 694–703.
- Mutluoğlu, M., 2013. *Kronik Yara Sınıflandırmaları*, in: Baktıroğlu, S., Aktaş, Ş. (Eds.), *Kronik Yarada Güncel Yaklaşımlar*. İstanbul Üniversitesi Tıp Fakültesi Kronik Yara Konseyi, İstanbul, p. 230.
- Nabiyev, V.V., 2012. *Yapay Zeka*, 4.Baskı. ed. Seçkin Yayıncılık.
- NAGAURI, M.R., 2020. *Guide To Ensemble Methods: Bagging vs Boosting* analyticsindiamag. URL <https://analyticsindiamag.com/guide-to-ensemble-methods-bagging-vs-boosting/> (Ziyaret Tarihi: 1.27.22).
- Namin, A.H., Leboeuf, K., Muscedere, R., Wu, H., Ahmadi, M., 2009. *Efficient hardware implementation of the hyperbolic tangent sigmoid function*, in: 2009 IEEE International Symposium on Circuits and Systems. 2009 IEEE International Symposium on Circuits and Systems - ISCAS 2009, IEEE, Taipei, Taiwan, pp. 2117–2120.
- National Diabetes Data Group, 1979. Classification and Diagnosis of Diabetes Mellitus and Other Categories of Glucose Intolerance. *Diabetes* 28, 1039–1057.
- Oğuzlar, A., 2003. VERİ ÖN İŞLEME. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* 67–76.
- Oladipupo Ayodele, T., 2010. *Types of Machine Learning Algorithms*, in: Zhang, Y. (Ed.), *New Advances in Machine Learning*. InTech.
- Olson, D.L., Delen, D., 2008. *Advanced data mining techniques*. Springer, Berlin Heidelberg.
- Onat, A., Çakır, H., Karadeniz, Y., Dönmez, İ., Karagöz, A., Yüksel, M., Can, G., 2014. Turkish Adult Risk Factor survey 2013: rapid rise in the prevalence of diabetes. *Türk Kardiyoloji Dernegi Arsivi-Archives of the Turkish Society of Cardiology* 42, 511–516.
- Orozco-Covarrubias, V.H., Cabrera-Pivaral, C.E., Aguilera-Velasco, Á., Alfaro-Alfaro, N., González-Gálvez, G., Sánchez-Michel, L., 2019. Prognostic Factors for the Development of Amputations in People with Type 2 Diabetes Mellitus. *J Res Med Dent Sci* 7, 219–224.
- Öz, E., 2019. *DESTEK VEKTÖR MAKİNELERİ*, in: Alp, S., Ersoy, Ö. (Eds.), *MAKİNE ÖĞRENMESİNDE SINIFLANDIRMA YÖNTEMLERİ VE UYGULAMALARI*. NOBEL AKADEMİK YAYINCILIK, ANKARA, p. 216.
- Özen, Z., Gülseçen, S., 2016. *R ile Geri Yayılım Yapay Sinir Ağı Algoritması Kullanarak Bir Sınıflandırma Uygulaması*, in: *R ile Veri Madenciliği Uygulamaları*. Çağlayan Kitabevi, İSTANBUL.
- Özkan, Y., 2021. *Uygulamalı Derin Öğrenme*, 1.Basım. ed. Papatya Yayıncılık Eğitim, İSTANBUL.
- Özkan, Y., 2016. *Veri Madenciliği Yöntemleri*, 3. Baskı. ed. Papatya Yayıncılık Eğitim, İSTANBUL.

- Özkır, V.Ç., 2019. *Genetik Algoritma*, in: Alp, S., Öz, E. (Eds.), *Makine Öğrenmesinde Sınıflandırma Yöntemleri ve R Uygulamaları*. NOBEL AKADEMİK YAYINCILIK, ANKARA, pp. 93–111.
- Özmen, E.P., Özcan, T., 2020. Diagnosis of diabetes mellitus using artificial neural network and classification and regression tree optimized with genetic algorithm. *Journal of Forecasting* 39, 661–670.
- Öztemel, E., 2016. *Yapay Sinir Ağları*, 4. Baskı/Özkaracan Matbaacılık. ed. Papatya Bilim Kitabevi.
- Pal, M., 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing* 26, 217–222.
- Pang, S., Gong, J., 2009. C5.0 Classification Algorithm and Application on Individual Credit Evaluation of Banks. *Systems Engineering - Theory & Practice* 29, 94–104.
- Parisi, M.C.R., Zantut-Wittmann, D.E., Pavin, E.J., Machado, H., Nery, M., Jeffcoate, W.J., 2008. Comparison of three systems of classification in predicting the outcome of diabetic foot ulcers in a Brazilian population. *European Journal of Endocrinology* 159, 417–422.
- Patel, A., 2020. *Boosting- Ensemble meta Algorithm for Reducing bias*. Medium. <https://medium.com/ml-research-lab/boosting-ensemble-meta-algorithm-for-reducing-bias-5b8bfdce281> (Ziyaret Tarihi: 1.27.22).
- Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin *Philosophical Magazine and Journal of Science* 2, 559–572.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., 2011. Scikit-learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON* 12, 6.
- Pekel, E., Özcan, T., 2018. DIAGNOSIS OF DIABETES MELLITUS USING STATISTICAL METHODS AND MACHINE LEARNING ALGORITHMS. *Sigma Journal of Engineering and Natural Sciences* 36, 1265–1282.
- Pekel Özmen, E., Özcan, T., 2022. *Diagnosis of Breast Cancer Using Novel Hybrid Approaches with Genetic Algorithm*, in: Kahraman, C., Cebi, S., Cevik Onar, S., Oztaysi, B., Tolga, A.C., Sari, I.U. (Eds.), *Intelligent and Fuzzy Techniques for Emerging Conditions and Digital Transformation*. Springer International Publishing, Cham, pp. 589–595.
- Perkins, B.A., Olaleye, D., Zinman, B., Bril, V., 2001. Simple Screening Tests for Peripheral Neuropathy in the Diabetes Clinic. *DIABETES CARE* 24, 7.
- Peterson, L.E., 2009. K-nearest neighbor. *Scholarpedia*, http://scholarpedia.org/article/K-nearest_neighbor (Ziyaret tarihi: 25.03.2022).
- Pickwell, K., Siersma, V., Kars, M., Apelqvist, J., Bakker, K., Edmonds, M., Holstein, P., Jirkovská, A., Jude, E., Mauricio, D., Piaggese, A., Ragnarson Tennvall, G., Reike, H., Spraul, M., Uccioli, L., Urbancic, V., van Acker, K., van Baal, J., Schaper, N., 2015. Predictors of Lower-Extremity Amputation in Patients with an Infected Diabetic Foot Ulcer. *Diabetes Care* 38, 852–857.
- Pittet, D., Wyssa, B., Herter-Clavel, C., Kursteiner, K., Vaucher, J., Lew, P.D., 1999. Outcome of Diabetic Foot Infections Treated Conservatively: A Retrospective Cohort Study With Long-term Follow-up. *Archives of Internal Medicine* 159, 851.
- Polikar, R., 2012. *Ensemble Learning*, in: Zhang, C., Ma, Y. (Eds.), *Ensemble Machine Learning*. Springer US, Boston, MA.

- Powell, M., Hosseini, M., Collins, J., Callahan-Flintoft, C., Jones, W., Bowman, H., Wyble, B., 2016. I TRIED A BUNCH OF THINGS: THE DANGERS OF UNEXPECTED OVERFITTING IN CLASSIFICATION (preprint). *Neuroscience*.
- Quinlan, J.R., 1996. Bagging Boosting, and C4.5. *AAAI-96 Proceedings* 725–730.
- Quinlan, J.R., 1993. *C4.5: Programs for machine learning*-. Morgan Kaufman Publishers, San Mateo, CA, USA.
- Quinlan, J.R., 1986. *Induction of decision trees*. *Mach Learn* 1, 81–106.
- Quinlan, R., 2002. *Data Mining Tools See5 and C5.0*.
- R Development Core Team, 2021. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- R Studio Team, 2020. R Studio Team, *RStudio: Integrated Development for R*, RStudio,. 250 Northern Ave, Boston, MA 02210.
- Raschka, S., 2020. *Majority Voting* (L07: Ensemble Methods). <https://www.youtube.com/watch?v=EFk21H6Q1ew> (Ziyaret tarihi: 25.03.2022)
- Raschka, Sebastian, 2020. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. *arXiv:1811.12808 [cs, stat]*.
- Rastogi, A., Goyal, G., Kesavan, R., Bal, A., Kumar, H., Mangalanadanam, Kamath, P., Jude, E.B., Armstrong, D.G., Bhansali, A., 2020. Long term outcomes after incident diabetic foot ulcer: Multicenter large cohort prospective study (EDI-FOCUS investigators) epidemiology of diabetic foot complications study. *Diabetes Research and Clinical Practice* 162, 1–8.
- Reddy, B.K., Delen, D., Agrawal, R.K., 2019. Predicting and explaining inflammation in Crohn's disease patients using predictive analytics methods and electronic medical record data. *Health Informatics Journal* 25, 1201–1218. <https://doi.org/10.1177/1460458217751015>
- Redlich, U., Xiong, Y.Y., Pech, M., Tautenhahn, J., Halloul, Z., Lobmann, R., Adolf, D., Ricke, J., Dudeck, O., 2011. Superiority of Transcutaneous Oxygen Tension Measurements in Predicting Limb Salvage After Below-the-Knee Angioplasty: A Prospective Trial in Diabetic Patients With Critical Limb Ischemia. *CardioVascular and Interventional Radiology* 34, 271–279.
- Reeves, W.T., 1983. Particle Systems A Technique for Modeling a Class of Fuzzy Objects. *Computer Graphics* 2, 17.
- Ripley, B., Venables, W., 2022a. *Package 'class.'*
- Ripley, B., Venables, W., 2022b. *Package 'nnet.'*
- Rose, J., 2021. RStudio.
- Salo, F., Injadat, M., Moubayed, A., Nassif, A.B., Essex, A., 2019. *Clustering Enabled Classification using Ensemble Feature Selection for Intrusion Detection*, in: 2019 International Conference on Computing, Networking and Communications (ICNC). Presented at the 2019 International Conference on Computing, Networking and Communications (ICNC), IEEE, Honolulu, HI, USA, pp. 276–281.
- Saltoglu, N., Kilicoglu, O., Baktiroglu, S., Osar-Siva, Z., Aktas, S., Altindas, M., Arslan, C., Aslan, T., Celik, S., Engin, A., Eraksoy, H., Ergonul, O., Ertugrul, B., Guler, S., Kadanali, A., Mulazimoglu, L., Olgun, N., Oncul, O., Oznur, A., Satman, I., Sencan, I., Tanriover, O., Turhan, O., Tuygun, A.K., Tuzun, H., Cinar Yasti, A., Yilmaz, T., 2016. Diagnosis, Treatment and Prevention of Diabetic Foot Wounds and Infections: Turkish Consensus Report. *Klimik Dergisi/Klimik Journal* 28, 2–34.
- Sánchez A, V.D., 2003. Advanced support vector machines and kernel methods. *Neurocomputing* 55, 5–20.
- Sarwar, A., Mudassar, M., Khalid, R., 2019. AN OBSERVATIONAL RESEARCH TO EVALUATE PREDICTING POSSIBILITIES OF DIABETIC FOOT ULCERS

- THROUGH LABORATORY AND CLINICAL BASELINE PARAMETERS. *Indo Am. J. P. Sci (IAJPS)* 06, 12604–12611.
- Satman, I., Omer, B., Tutuncu, Y., Kalaca, S., Gedik, S., Dinccag, N., Karsidag, K., Genc, S., Telci, A., Canbaz, B., Turker, F., Yilmaz, T., Cakir, B., Tuomilehto, J., 2013. Twelve-year trends in the prevalence and risk factors of diabetes and prediabetes in Turkish adults. *European Journal of Epidemiology* 28, 169–180.
- Satman, I., Yilmaz, T., Sengul, A., Salman, S., Salman, F., Uygur, S., Bastar, I., Tutuncu, Y., Sargin, M., Dinccag, N., Karsidag, K., Kalaca, S., Ozcan, C., King, H., 2002. Population-Based Study of Diabetes and Risk Characteristics in Turkey: Results of the Turkish Diabetes Epidemiology Study (TURDEP). *Diabetes Care* 25, 1551–1556.
- Sayiner, Z.A., Can, F.I., Akarsu, E., 2019. Patients' clinical characteristics and predictors for diabetic foot amputation. *Primary Care Diabetes* 13, 247–251.
- Schaper, N.C., 2004. Diabetic foot ulcer classification system for research purposes: a progress report on criteria for including patients in research studies. *Diabetes/Metabolism Research and Reviews* 20, S90–S95.
- Schmidt, B.M., Ye, W., Zhou, S., 2020. Multidrug resistant organism predicts ulcer recurrence following surgical management of diabetic foot osteomyelitis. *Int Wound J* 17, 1–8.
- Schoen, D.E., Glance, D.G., Thompson, S.C., 2015. Clinical decision support software for diabetic foot risk stratification: development and formative evaluation. *J Foot Ankle Res* 8, 73.
- Scholkopf, B., Smola, A.J., 2001. *Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond*, 1st Edition. ed, Adaptive Computation and Machine Learning. The MIT Press, Cambridge, Massachusetts London, England.
- Seuc, A.H., López, M., Rodríguez, L., Montequín, J.F., 2009. Short-term prediction of major lower-limb amputation based on clinical indicators on admission: a single institutional experience in a developing country. *Int Angiol* 28, 38–43.
- Shannon, C.E., 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal* 27, 55.
- Shihong, Y., Ping, L., Peiyi, H., 2003. SVM classification: Its contents and challenges. *Appl. Math. J. Chinese Univ. Ser. B* 18, 11.
- Shreffler, J., Huecker, M.R., 2021. *Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios*. NCBI. URL https://www.ncbi.nlm.nih.gov/books/NBK557491/#_ncbi_dlg_citbx_NBK557491 (Ziyaret Tarihi: 2.7.22).
- Silahtaroglu, G., 2016. *Veri Madenciliği Kavram ve Algoritmaları*. Papatya Yayıncılık Eğitim.
- Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., Unterthiner, T., Ernst, G.M.F., 2020. *Package 'ROCR'*.
- Stephenson, B., Cook, D., Dixon, P., Duckworth, W., Kaiser, M., Koehler, K., Meeker, W., 2001. CHAPTER 3. *BINARY RESPONSE AND LOGISTIC REGRESSION ANALYSIS*, in: *Beyond Traditional Statistical Methods*. Iowa State University.
- Steward, T.A., 1998. *Intellectual capital: The new wealth of organizations*, in: Ruckdeschel, C. (Ed.), . Doubleday/ Currency, Performance Improvement, p. 56.
- Su, C.-L., Chang, C.-C., Peng, Y.-S., Chen, M.-Y., 2018. The Predictive Factors Associated With Comorbidities for Treatment Response in Outpatients With King Classification III Diabetes Foot Ulcers. *Ann Plast Surg* 81, S39–S43.
- Süleymanlar, G., Utaş, C., Arinsoy, T., Ates, K., Altun, B., Altıparmak, M.R., Ecder, T., Yılmaz, M.E., Camsarı, T., Basci, A., Serdengeçti, K., 2011. A population-based survey of Chronic RENal Disease In Turkey--the CREDIT study. *Nephrology Dialysis Transplantation* 26, 1862–1871.

- Sulzmann, J.-N., Fürnkranz, J., Hüllermeier, E., 2007. *On Pairwise Naive Bayes Classifiers*, in: Kok, J.N., Koronacki, J., Mantaras, R.L. de, Matwin, S., Mladenić, D., Skowron, A. (Eds.), *Machine Learning: ECML 2007, Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 371–381.
- Sun, J.-H., Tsai, J.-S., Huang, C.-H., Lin, C.-H., Yang, H.-M., Chan, Y.-S., Hsieh, S.-H., Hsu, B.R.-S., Huang, Y.-Y., 2012. Risk factors for lower extremity amputation in diabetic foot disease categorized by Wagner classification. *Diabetes Research and Clinical Practice* 95, 358–363.
- Syarif, I., Prugel-Bennett, A., Wills, G., 2016. SVM Parameter Optimization using Grid Search and Genetic Algorithm to Improve Classification Performance. *TELKOMNIKA* 14, 1502.
- Tangirala, S., 2020. Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm*. *IJACSA* 11.
- Tardivo, J.P., Baptista, M.S., Correa, J.A., Adami, F., Pinhal, M.A.S., 2015. Development of the Tardivo Algorithm to Predict Amputation Risk of Diabetic Foot. *PLoS ONE* 10, e0135707. <https://doi.org/10.1371/journal.pone.0135707>
- TEMĐ Diabetes Mellitus Çalışma ve Eğitim Grubu, 2018. *TEMĐ DİABETES MELLİTUS VE KOMPLİKASYONLARININ TANI, TEDAVİ VE İZLEM KILAVUZU*, Türkiye Endokrinoloji ve Metabolizma Derneği (TEMĐ). ed. Türkiye Endokrinoloji ve Metabolizma Derneği (TEMĐ).
- Therneau, T., Atkinson, B., 2022. *Package ‘rpart.’*
- Tonta, Y., Küçük, M.E., 2005. Sanayi Toplumundan Bilgi Topluma Geçiş Sürecinde Temel Dinamikler, *Türk Kütüphaneciliği*, 19, 4.
- Torgo, L., 2016. *Package ‘DMwR2.’*
- Torgo, L., 2011. *Data mining with R: learning with case studies*, Chapman & Hall/CRC data mining and knowledge discovery series. Chapman & Hall/CRC, Boca Raton.
- Treece, K.A., Macfarlane, R.M., Pound, N., Game, F.L., Jeffcoate, W.J., 2004. Validation of a system of foot ulcer classification in diabetes mellitus: Original article. *Diabetic Medicine* 21, 987–991.
- Tsai, C.-Y., Chu, S.-Y., Wen, Y.-W., Hsu, L.-A., Chen, C.-C., Peng, S.-H., Huang, C.-H., Sun, J.-H., Huang, Y.-Y., 2013. The value of Doppler waveform analysis in predicting major lower extremity amputation among dialysis patients treated for diabetic foot ulcers. *Diabetes Research and Clinical Practice* 100, 181–188.
- Tsang, M.W., Wong, W.K.R., Hung, C.S., Lai, K.-M., Tang, W., Cheung, E.Y.N., Kam, G., Leung, L., Chan, C.W., Chu, C.M., Lam, E.K.H., 2003. Human Epidermal Growth Factor Enhances Healing of Diabetic Foot Ulcers. *Diabetes Care* 26, 1856–1861.
- Turban, E., Delen, D., Turban, E., 2015. *Business intelligence and analytics: systems for decision support*, Tenth edition. ed. Pearson, Boston-USA.
- Türkiye Diyabet Programı*, 2015. T.C. Sağlık Bakanlığı - Türkiye Halk Sağlığı Kurumu.
- Uccioli, L., Gandini, R., Giurato, L., Fabiano, S., Pampana, E., Spallone, V., Vainieri, E., Simonetti, G., 2010. Long-Term Outcomes of Diabetic Patients With Critical Limb Ischemia Followed in a Tertiary Referral Diabetic Foot Clinic. *Diabetes Care* 33, 977–982.
- Ugwu, E., Adeleye, O., Gezawa, I., Okpe, I., Enamino, M., Ezeani, I., 2019. Predictors of lower extremity amputation in patients with diabetic foot ulcer: findings from MEDFUN, a multi-center observational study. *J Foot Ankle Res* 12, 34.
- Uludağ, M.O., 2010. Diyabete Bağlı İkincil Hastalıklar (Komplikasyonlar). *Mised* 24, 39–44.
- Ulusal Diyabet Konsensus Grubu, 2019. *Diyabet Tanı ve Tedavi Rehberi*, Armoni Nüans Baskı Sanatları A.Ş. ed. Şişli - İstanbul.

- Undersea & Hyperbaric Medical Society, 2014. *Hyperbaric Oxygen Therapy Indications: : The Hyperbaric Oxygen Therapy Committee Report*. Best Publishing Company.
- Utsunomiya, M., Nakamura, M., Nagashima, Y., Sugi, K., 2014. Predictive Value of Skin Perfusion Pressure After Endovascular Therapy for Wound Healing in Critical Limb Ischemia. *Journal of Endovascular Therapy* 21, 662–670.
- Vabalas, A., Gowen, E., Poliakoff, E., Casson, A.J., 2019. Machine learning algorithm validation with a limited sample size. *PLoS ONE* 14, e0224365.
- Vapnik, V.N., 1998. *Statistical Learning Theory*, 1st Edition. ed, Adaptive and Learning Systems for Signal Processing, Communications and Control. Wiley Publishing Inc., New York, NY, USA.
- Volaco, Chantelau, Richter, Luther, 2004. Outcome of critical foot ischaemia in longstanding diabetic patients: a retrospective cohort study in a specialised tertiary care centre. *Vasa* 33, 36–41.
- Wagner, F., 1981. The dysvascular foot: a system for diagnosis and treatment. *Foot Ankle* 64–122.
- Wei, T., Simko, V., 2021. *Package ‘corrplot.’* CRAN.
- WHO, 2010. *Diabetes Mellitus*. World Health Organization. URL <https://www.who.int/mediacentre/factsheets/fs138/en/> (Ziyaret Tarihi: 2.20.19).
- Wickham, H., 2020. *Package ‘plyr.’*
- Wickham, H., Bryan, J., 2019. *Package ‘readxl.’*
- Widatalla, A., Mahadi, S.E., Shawer, M., Elsayem, H., Ahmed, M., 2009. Implementation of diabetic foot ulcer classification system for research purposes to predict lower extremity amputation. *International Journal of Diabetes in Developing Countries* 29, 1.
- Witten, I.H., Frank, E., Hall, Mark.A., J.Pall, C., 2011. *Data Mining: Practical Machine Learning Tools and Techniques*, Fourth Edition. ed. Elsevier, 50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States.
- WOLPERT, D.H., 1992. Stacked Generalization. *Neural Networks*, 5, 241–25.
- Wu, J.T., Wong, M., Lo, Z.J., Wong, W.-E., Narayanan, S., Tan, G.W.L., Chandrasekar, S., 2017. A Series of 210 Peripheral Arterial Disease Below-Knee Amputations and Predictors for Subsequent Above-Knee Amputations. *Ann Vasc Dis* 10, 217–222.
- Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.-H., Steinbach, M., Hand, D.J., Steinberg, D., 2008. Top 10 algorithms in data mining. *Knowl Inf Syst* 14, 1–37.
- Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* 415, 295–316.
- Yao, Q., Wang, M., Chen, Y., Dai, W., Li, Y.-F., Tu, W.-W., Yang, Q., Yu, Y., 2019. Taking Human out of Learning Applications: A Survey on Automated Machine Learning. *arXiv:1810.13306 [cs, stat]*.
- Yekta, Z., Pourali, Ghasemi-rad, M., Ravanyar, Nezhadrahim, 2011. Clinical and behavioral factors associated with management outcome in hospitalized patients with diabetic foot ulcer. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy* 371.
- Yesil, S., Akinci, B., Yener, S., Bayraktar, F., Karabay, O., Havitcioglu, H., Yapar, N., Atabey, A., Kucukyavas, Y., Comlekci, A., Eraslan, S., 2009. Predictors of amputation in diabetics with foot ulcer: Single center experience in a large Turkish cohort. *HORMONES* 8, 286–295.
- Yılmaz Akşehirli, Ö., Ankaralı, H., Aydın, D., Saraçlı, Ö., 2013. Tıbbi Tahminde Alternatif Bir Yaklaşım: Destek Vektör Makineleri. *Türkiye Klinikleri J Biostat* 5, 19–28.
- Zhang, C., Ma, Y. (Eds.), 2012. *Ensemble Machine Learning*. Springer US, Boston, MA.
- Zhou, Z.-H., 2012a. *Ensemble Methods: Foundations and Algorithms*, 0 ed. Chapman and Hall/CRC.

Zuo, W., Zhang, D., Wang, K., 2008. On kernel difference-weighted k-nearest neighbor classification. *Pattern Anal Applic* 11, 247–257.



7. EKLER

EK-1: ETİK KURUL ONAY BELGESİ



T.C.
İSTANBUL ÜNİVERSİTESİ
İSTANBUL TIP FAKÜLTESİ
KLİNİK ARAŞTIRMALAR ETİK KURULU



Sayı : 713
Konu: Doç. Dr. Çiğdem EROL hk.

Tarih : 23.05.2019

Sayın Doç. Dr. Çiğdem EROL
Enformatik Bölümü

İlgi: Enformatik Bölümünün 24/04/2019 gün ve 83374 sayı yazısı

Sorumlu araştırmacılığını üstlendiğiniz ve Araş. Gör. Denizhan DEMİRKOL' un yürüteceği 2019/656 dosya numaralı "Diyabetik Ayak Hastalarının Amputasyon Riskinin Yapay Zeka Teknikleri ile Öngörülmesi" başlıklı çalışma kurulumuzun 10/05/2019 tarih ve 09 sayılı toplantısında görüşülerek etik yönden uygun bulunmuş olup, tutanaklar ekte sunulmuştur.

Bilgilerinizi rica ederim.

Prof. Dr. A.Yağız ÜRESİN

İstanbul Tıp Fakültesi Klinik Araştırmalar
Etik Kurul Başkanı

Eki: İstanbul Tıp Fakültesi Klinik Araştırmaları Etik Kurulu Karar Formu

Şekil 7.1: Etik Kurul Onay Belgesinin Ekran Görüntüsü

EK-2: ARAŞTIRMADA KULLANILAN DİYABETİK AYAK İZLEM FORMU

HASTA İLK BAŞVURDUĞUNDA DOLDURULACAK FORM			
DİYABETİK AYAK İZLEM FORMU*		FORM NO¹:	Form doldurma tarihi:/...../.....
Hasta No:	Yaş:	Cinsiyet: Erkek (1) ☐ Kadın (2) ☐	
Bağımlılık Düzeyi:			
Aile ile yaşıyor ☐ Tek yaşıyor ☐			
Diyabet Tipi:			
Tip(1) ☐			
Tip2 (2) ☐	Diyabet süresi: yıl	Ağırlık: kg	Boy: cm
Diyabetik ayak enfeksiyonu süresi: gün		Hipertansiyon ² : Yok (1) Var(2)	Vücut/kitle indeksi (kg/m ²):
Kullandığı antidiyabetik: insülin (1) ☐		Son ölçülen HbA1c düzeyi:	
Oral antidiyabetik (2) ☐		Sigara kullanımı: Yok (1) ☐ Var (2) ☐ paket/yıl	
Kullandığı oral antidiyabetik ilaçlar:			
Tekrarlayan ayak enfeksiyonu öyküsü: Yok (1) ☐ Var (2) ☐ kez			
Geçmişte amputasyon [*] : Yok (1) ☐ Var(2) ☐ Varsa düzeyi: Başparmak (1) ☐ Diğer parmaklar (2) ☐ Metatars (3) ☐			
Geçmişte vasküler cerrahi: Yok (1) ☐ Var(2) ☐			
Böbrek yetmezliği: Yok (1) ☐ Var(2) ☐		Diyaliz: Yok (1) ☐ Var(2) ☐	
Charcot Ayak durumu:Yok(1) ☐ Var(2) ☐			
Periferik Arter Hastalığı (PAH) / Doku beslenmesi (perfüzyon) (etkilenen ayakta)	Evre 1: PAH belirtileri yok, palpabl dorsopedal ve/veya posteriyor tibial arter (1) ☐		
	Evre 2: PAH belirtileri var, ciddi iskemi yok, kladikasyon var (2) ☐		
	Evre 3: Ciddi iskemi varlığı (3) ☐		
Yara genişliği	 cm ² (ülserin dış sınırları ülseri çevreleyen sağlam deriden itibaren ölçülmelidir)		
Yara yeri	Başparmak (1) ☐ Diğer parmaklar (2) ☐ Metatars (3) ☐ Ayağın üstünde arka ayak (4) ☐ Topuk (5) ☐		
Yara derinliği (doku kaybı)	Evre 1: Yüzeysel, deri yüzeyi etkilenmiş, dermisten daha derin dokular etkilenmemiş (1) ☐		
	Evre 2: Derin ülser, subkutan yapılara (kas, fasya, tendon) penetre olmuş (2) ☐		
	Evre 3: Kemik sondajı ⁵ (probing to bone) pozitif veya direkt bakı ile kemik görünüyor (3) ☐		
İnfeksiyon	Evre 1: İnfeksiyon belirtileri yok (1) ☐		
	Evre 2: Cilt ve cilt altı dokuları tutmuş 2 cm ² 'den küçük lokal yara yüzeyi, sistemik bulgular yok (2) ☐		
	Evre 3: 2 cm ² 'den büyük yara yüzeyi, ciltten daha derin yapılar da tutulmuş (apse, septik artritis, fasit, osteomyelit), sistemik inflamatuvar yanıt sendromu (SIRS) ⁷ bulguları yok (3) ☐		
	Evre 4: SIRS ⁷ belirtileri ile beraber herhangi bir ayak enfeksiyonu (4) ☐		
Osteomyelit	Direkt grafi: Negatif (1) ☐ Pozitif (2) ☐ MR: Negatif (1) ☐ Pozitif (2) ☐		
	Kemik sondajı ⁵ : Negatif (1) ☐ Pozitif (2) ☐ İşaretili lökosit sintigrafisi: Negatif (1) ☐ Pozitif (2) ☐		
	Patolojik inceleme: Negatif (1) ☐ Pozitif (2) ☐		
Daha önce geçirdiği ayak enfeksiyonunda osteomyelit: Yok (1) ☐ Var (2) ☐			
Daha önce geçirilmiş debridman: Yok (1) ☐ Var (2) ☐			
Lökosit düzeyi: / (mm ³)	C-reaktif protein düzeyi: (mg/dL)	Eritrosit sedimentasyon hızı: mm/saat	
Hemoglobin düzeyi: / (g/dL)	Bilirubin: (mg/dL)	Albümin: (g/dL)	
Başvuru anındaki Glukoz: (mg/dL)	INR:	PNL(10 ³ /µl): PLT(10 ³ /µl):	
ALT: (U/L)	AST: (U/L)	Kreatinin: (mg/dL)	

Şekil 7. 2: Araştırmada Kullanılan Diyabetik Ayak İzlem Formunun Ekran Görüntüsü – Sayfa 1

Total kolesterol düzeyi: -----mg/dL		
Trigliserid düzeyi: ----- mg/dL	LDL düzeyi: ----- mg/dL	HDL düzeyi: ----- mg/dL
HASTA İZLEMİ		
Önerilen Antibiyotik ⁵		
Yok (1) ☐ Var (2) ☐	Ajan: -----	Süre: -----
	Ajan: -----	Süre: -----
	Ajan: -----	Süre: -----
	Ajan: -----	Süre: -----
Hastanede Yatış: Yok (1) ☐ Var (2) ☐ gün		
Ampütasyon varsa: Başparmak (1) ☐ Diğer parmaklar (2) ☐ Metatars (3) ☐ Bilek (4) ☐ Diz altı (5) ☐ Diz üstü (6) ☐		
Gangren durumu: Yok (1) ☐ Var (2) ☐		
Hastaneye tekrar yatış: Yok (1) ☐ Var (2) ☐ kez, ... gün		Hastayı izlem süresi: gün
Nöropati ³	Evre 1: Etkilenen ayakta duyu kaybı yok (1) ☐ Evre 2: Etkilenen ayakta duyu kaybı var (2) ☐	
Retinopati:		
Koroner Arter Hastalığı:		
ASTIM, KOAH Durumu: Yok (1) ☐ Var (2) ☐	Alkol kullanımı: Yok (1) ☐ Var (2) ☐	
ACE inhibitörü Kullanımı: Yok (1) ☐ Var (2) ☐		
Statin Kullanımı: Yok (1) ☐ Var (2) ☐	Canaglifozine Kullanımı: Yok (1) ☐ Var (2) ☐	

Tablo 1. Diyabetik ayak, Meggit-Wagner sınıflandırması

Grade 0	Sağlam cilt
Grade I	Yüzeyel ülser
Grade II	Genellikle infeksiyon ile seyreden ve tendon gibi cilt altı dokulara uzanım gösteren ülser
Grade III	Abse ya da osteomyelit ile seyreden derin ülser
Grade IV	Sınırlı gangren
Grade V	Yaygın gangren

Tablo 2. Diyabetik ayak, Teksas Üniversitesi Sınıflandırması

	Grade 0	Grade I	Grade II	Grade III
A	Yara yok	Yüzeysel yara	Tendon ya da kapsül tutulumu	Eklemler ya da kemik tutulumu
B	Yara yok, infeksiyon mevcut	Yüzeysel enfekte yara	İnfeksiyon ile seyreden tendon ya da kapsül tutulumu	İnfeksiyon ile seyreden eklemler ya da kemik tutulumu
C	Yara yok, iskemi mevcut	İskemi ile seyreden yüzeysel yara	İskemi ile seyreden tendon ya da kapsül tutulumu	İskemi ile seyreden eklemler ya da kemik tutulumu
D	Yara yok, iskemi ve infeksiyon mevcut	İskemi ve infeksiyon ile seyreden yüzeysel yara	İskemi ve infeksiyon ile seyreden tendon ya da kapsül tutulumu	İskemi ve infeksiyon ile seyreden eklemler ya da kemik tutulumu

Tablo 3. Diyabetik ayak, S (AD) SAD sınıflandırması

Grade	Alan	Derinlik	Sepsis	Arteriel yetmezlik	Denervasyon
0	Sağlam cilt	Sağlam cilt	İnfeksiyon yok	Pedal nabızlar normal	Normal VAE
I	< 10 mm ²	Cilt&Cilt altı	Yüzeysel	Zayıf pedal nabız, ya da tek taraflı yok	VAE artmış
II	10-30 mm ²	Tendon, Eklem kapsülü	Sellülit	Her iki pedal nabız yok	Belirgin nöropati
III	> 30 mm ²	Eklem aralığı, Kemik	Osteomiyelit	Gangren	Charcot ayağı

HASTA DURUMU SONUCLANDIĞINDA DOLDURULMASI GEREKEN ALANLAR:

Uygulanan Diğer Tedaviler :		Hiperbarik Oksijen Tedavisi: Yok (1) ☐ Var(2) ☐ seans	
Yapılan cerrahi:		Cerrahi girişim yok (1) ☐	Yumuşak doku ile sınırlı debridman (2) ☐
		Ampütasyon (3) ☐	Vasküler girişim (By-pass) (4) ☐
Rehabilitasyon	Fizyoterapi:	Yok (1) ☐ Var (2) ☐ varsa toplam maliyet TL	
	Protez uygulaması:	Yok (1) ☐ Var (2) ☐ TL	Özel ayakkabı yapımı: Yok (1) ☐ Var (2) ☐ TL
Yeni görüntüleme	Direkt grafi:	Yok (1) ☐ Var (2) ☐ kez	MR görüntüleme: Yok (1) ☐ Var (2) ☐ kez
	İşaretili lökosit sintigrafisi:	Yok (1) ☐ Var (2) ☐ kez	
Sonuç ⁶ :		İyileşme (1) ☐	Tekrarlayan enfeksiyon (2) ☐ Tekrarlayan cerrahi girişim gereksinimi (3) ☐ Exitus (4) ☐

ACIKLAMALAR

*Çalışmaya alınacak hastalar ilk kez diyabetik ayak patolojisi nedeniyle başvuran veya daha önce diyabetik ayak patolojisi (enfeksiyonu vb.) olup tedavi sonrasında iyileşmiş ve yeni atak sayılan hastalar olacaktır. Daha önce metatars düzeyinin üzerinde amputasyon yapılmış olan hastalar çalışmaya alınmayacaktır.

- Form no kısmına başa merkezin kısaltması (İÜ, DEÜ, PÜ, AÜ gibi) ardından da form no yazılacak. Örneğin, İstanbul Üniversitesi İstanbul Tıp Fakültesi için olgu formu "İÜ – 1". Aynı hasta tedavisini tamamlayıp iyileşmiş olarak kabul edildikten sonra ikinci kez yeni bir atak ile çalışmaya alındığında ilk başvurusu "İÜ – 1 – A" ikinci başvurusu "İÜ – 1 – B" şeklinde kaydedilmelidir.
- Hipertansiyon tanısı için; hastanın antihipertansif ilaç kullanımı veya başvuru sırasında sistolik/diastolik kan basıncının 140/90 mmHg üzerinde olması veya hastaneye yatışı sırasında Kardiyoloji Uzmanı tarafından hipertansiyon tanısı konularak antihipertansif ilaç başlanmış olması gereklidir.
- Nöropati tanısı için monofilaman test sonucu pozitif veya nöroloji konsültasyonu sonrası nöropati tanısının konulması gereklidir.
- Sistemik İnflamatuvar Yanıt Sendromu (SIRS):
 - Vücut ısısının 38.3°C'nin üzerinde veya 36°C'nin altında olması,
 - Kalp hızının 90/dakika veya üzerinde olması,
 - Solunum hızının 20/dakika üzerinde veya PaCO₂ basıncının 32 mmHg'nin altında olması,
 - Lökosit sayısının 12000/mm³ üstünde veya 4000/mm³ altında olması,
 Bulgularından en az ikisinin saptanması.
- Hastanın izlemi sırasındaki tüm antibiyotikleri (antibiyotik değişikliği yapıldıysa bunlar da dahil) yazılmalıdır.
- Hastada yeni enfeksiyon atağı gelişmesi durumunda hasta bu aşamada çalışmadan çıkarılmalıdır. İyileşme ile sonuçlanan hastalar ikinci kez atak geçirdiklerinde yeni hasta olarak çalışmaya alınabilir. Bu durum için 1. açıklama dikkate alınarak numaralandırma yapılacaktır.
- Son olarak her hasta için Meggit- Wagner, University of Texas ve SAD sınıflandırmaları girilecektir. |

8. ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı	DENİZHAN DEMİRKOL
Doğum Yeri	
Doğum Tarihi	
Uyruğu	☒ T.C. ☐ Diğer:
Telefon	
E-Posta Adresi	
Web Adresi	

Eğitim Bilgileri	
Lisans	
Üniversite	YEDİTEPE ÜNİVERSİTESİ
Fakülte	TİCARİ BİLİMLER FAKÜLTESİ
Bölümü	BİLİŞİM SİSTEMLERİ VE TEKNOLOJİLERİ BÖLÜMÜ
Mezuniyet Yılı	12.06.2013

Yüksek Lisans	
Üniversite	YEDİTEPE ÜNİVERSİTESİ
Enstitü Adı	SOSYAL BİLİMLER ENSTİTÜSÜ
Anabilim Dalı	YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI
Programı	YÖNETİM BİLİŞİM SİSTEMLERİ YÜKSEK LİSANS (TEZLİ, İNGİLİZCE) PROGRAMI

Doktora	
Üniversite	İstanbul Üniversitesi
Enstitü Adı	Fen Bilimleri Enstitüsü
Anabilim Dalı	Enformatik Anabilim Dalı
Programı	Enformatik Programı

Makale ve Bildiriler	
Uluslararası hakemli dergilerde yayımlanan makaleler:	
Demirkol, D., Koçoğlu, F., Aktaş, Ş., Erol, Ç., 2022. <i>A BIBLIOMETRIC ANALYSIS OF THE RELATIONSHIP BETWEEN DIABETES AND ARTIFICIAL INTELLIGENCE</i> . Journal of Istanbul Faculty of Medicine-JMED,85-2	
Demirkol, D., Seneler, C., 2019. Evaluation of Student Information System (SIS) In Terms of User Emotion, Performance and Perceived Usability: A Turkish University Case (An Empirical Study). <i>Procedia Computer Science</i> 158, 1033–1051.	
Demirkol, D., Seneler, C., Daim, T., Shaygan, A., 2020. Measuring emotional reactions of university students towards a Student Information System (SIS): A Turkish university case. <i>Technology in Society</i> 63, 101412.	

Demirkol, D., Seneler, C., Daim, T., Shaygan, A., 2020. Measuring perceived usability of university students towards a student information system (SIS): A Turkish university case. *Technology in Society* 62, 101281.

Şerefoğlu Henkoğlu, H., **Demirkol, D.**, 2018. Measurement of Web Usability: Web Site of Aydın Adnan Menderes University Library. *Bilgi Dünyası*, 19,2.

Ulusal hakemli dergilerde yayımlanan makaleler:

Demirkol, D., Kartal, E., Şeneler, Ç., Gülseçen, S., 2019. Bir Öğrenci Bilgi Sisteminin Kullanılabilirliğinin Makine Öğrenmesi Teknikleriyle Tahmin Edilmesi. *Veri Bilimi* 1, 9.

Demirkol, D., Şeneler, Ç., 2018. A TURKISH TRANSLATION OF THE SYSTEM USABILITY SCALE: THE SUS-TR. *Usak University Journal of Social Sciences* 11, 238–253

Demirkol, D., Şeneler, Ç., 2018. A Turkish Translation Of The Emotion Words Prompt List: The EWPL-TR. *Adnan Menderes Üniversitesi, Sosyal Bilimler Enstitüsü Dergisi*, 5, 218–235.

Uluslararası bilimsel toplantılarda sunulan ve bildiri kitaplarında (proceedings) basılan bildiriler :

Demirkol, D., Erdebilli, B., 2017. AHP Approach for Selecting Enterprise Resource Planning (ERP). *New Challenges in Management and Business Conference*

Demirkol, D., Genç, F.N., 2018. LEGAL STATUS OF BITCOIN: EXAMPLES WITH CURRENT ASSESSMENT FROM TURKEY AND THE WORLD. *1st International Congress of Political, Economic and Financial Analysis Conference*

Demirkol, D., Kartal, E., Şeneler, Ç., Gülseçen, S., 2018. Predicting Usability of a Student Information System by Using Machine Learning Techniques. *ICONDATA, International Conference on Data Science and Applications*, pp. 106–107.

Demirkol, D., Koçoğlu, F.Ö., Aktaş, Ş., Erol, Ç., 2020. A BIBLIOMETRIC ANALYSIS ON THE RELATIONSHIP OF DIABETES AND ARTIFICIAL INTELLIGENCE. *7th International Management Information Systems Conference*.

Demirkol, D., Şeneler, Ç., 2019. Evaluation of Student Information System (SIS) In Terms Of User Emotions, Performance And Perceived Usability: A Turkish University Case (Empirical Study). *3rd World Conference on Technology, Innovation and Entrepreneurship Conference*

Demirkol, D., Şeneler, Ç., 2018. EVALUATION OF A STUDENT INFORMATION SYSTEM (SIS) IN TERMS OF USER EMOTIONS, PERFORMANCE AND PERCEIVED USABILITY: A PILOT STUDY. *XV. European Conference on Social and Behavioral Conference*

Yavuzarslan, M., **Demirkol, D.**, Gülseçen, S., 2019. The Role of Virtual Reality Implementations in Rehabilitation of Patients With Stroke. *6th International Management Information Systems Conference*.

Yazılan ulusal/uluslararası kitaplar veya kitaplardaki bölümler:

Demirkol, D., Genç, F.N., 2020. *Bitcoin'in Yasal Statüsü: Türkiye ve Dünyada Örneklerle Güncel Değerlendirme*, in: Ay, S., Keser, H.Y. (Eds.), *TİCARETTE DİJİTAL DÖNÜŞÜM VE E-TİCARET*. TESAM & Değişim Yayınları, BURSA, pp. 139–159.

Demirkol, D., Hancı, A.K., 2020. *TÜRKİYE’DE YENİ FAALİYET GÖSTERMEYE BAŞLAMIŞ BİR E-TİCARET WEBSİTESİNİN KULLANILABİLİRLİĞİNİN DEĞERLENDİRİLMESİ / EVALUATION OF THE USABILITY OF A NEWLY OPERATING E-COMMERCE WEBSITE IN TURKEY*, in: Gülseçen, S., Rızvanoğlu, K., Tosun, N., Akadal, E. (Eds.), *İNSAN BİLGİSAYAR ETKİLEŞİMİ, ARAŞTIRMA VE UYGULAMALAR, HUMAN COMPUTER INTERACTION RESEARCH AND APPLICATIONS*. Istanbul University Press, Istanbul-Turkey, pp. 129–147.

Demirkol, D., Seneler, Ç., 2018. *Evaluation of a Student Information System (SIS) in Terms of User Emotions, Performance And Perceived Usability: A Pilot Study*, in: Dorczak, R., Arslan, H., Musualik, R. (Eds.), *Recent Researches on Social Sciences- Monographs and Studies of Jagiellonian University- Institute of Public Affairs*. Jagiellonian University Institute of Public Affairs Ul. Prof. Łojasiewicza 4, 30-348 Kraków, Krakow.

Gülseçen, S., **Demirkol, D.,** 2019. *Semiyotik , İnsan Bilgisayar Etkileşimi ve Yapay Zekâ*, in: Sağlık Bilimlerinde Yapay Zekâ. Çağlayan Kitabevi, İstanbul, Türkiye, p. 307.

Koçoğlu, F.Ö., **Demirkol, D.,** 2020. *Data in the Context of Industry 4.0*, in: Gülseçen, S., Sharma, S.K., Akadal, E. (Eds.), *Who Runs the World: Data*. Istanbul University Press, İSTANBUL, pp. 71–92.

Ulusal bilimsel toplantılarda sunulan ve bildiri kitaplarında basılan bildiriler:

Demirkol, D., Acarlı, Ö., Aktaş, Ş., Erol, Ç., 2020. *Diyabetik ayak hastalarının retrospektif analizi: Ön çalışma*. 13. Ulusal Sualtı Hekimliği ve Hiperbarik Tıp Kongresi (HİTAM’20), pp. 176–181.