

T.C.
BİLECİK ŞEYH EDEBALI ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

**AİLEVİ AKDENİZ ATEŞİ'NDE KOLŞİSİN TEDAVİSİNİN MAKİNE ÖĞRENMESİ
ANALİZİ**

YÜKSEK LİSANS TEZİ
MUHAMMET İKBAL YILDIZ

TEZ DANIŞMANI
DR. ÖĞR. ÜYESİ SALİM CEYHAN

BİLECİK, 2024

10619278

T.C.
BİLECİK ŞEYH EDEBALI ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

**AİLEVİ AKDENİZ ATEŞİ'NDE KOLŞİSİN TEDAVİSİNİN MAKİNE ÖĞRENMESİ
ANALİZİ**

YÜKSEK LİSANS TEZİ
MUHAMMET İKBAL YILDIZ

TEZ DANIŞMANI
DR. ÖĞR. ÜYESİ SALİM CEYHAN

BİLECİK, 2024

10619278

BEYAN

Ailevi Akdeniz Ateşi'nde Kolşisin Tedavisinin Makine Öğrenmesi Analizi adlı yüksek lisans tezi hazırlık ve yazımı sırasında bilimsel araştırma ve etik kurallarına uyduğumu, başkalarının eserlerinden yararlandığım bölümlerde bilimsel kurallara uygun olarak atıfta bulunduğumu, kullandığım verilerde herhangi bir tahrifat yapmadığımı, tezin herhangi bir kısmının Bilecik Şeyh Edebali Üniversitesi veya başka bir üniversitede başka bir tez çalışması olarak sunulmadığını, aksinin tespit edileceği muhtemel durumlarda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Bu çalışmanın,			
Bilimsel Araştırma Projeleri (BAP), TÜBİTAK veya benzeri kuruluşlarca desteklenmesi durumunda; projenin ve destekleyen kurumun adı proje numarası ile birlikte, ETİK KURUL onayı alınması durumunda ise ETİK KURUL tarih karar ve sayı bilgilerinin beyan edilmesi gerekmektedir.			
DESTEK ALINMIŞTIR		DESTEK ALINMAMIŞTIR	X
Destek alındı ise;			
Destekleyen kurum;			
Desteğin Türü		Proje Numarası	
BAP (Bilimsel Araştırma Projesi)			
Diğer;.....			
ETİK KURUL onayı var ise;			
ETİK KURUL karar tarih/sayı: 1310920		20.01.2022	

Muhammet İkbāl Yıldız

.....

Tarih

.....

İmza

ÖN SÖZ

Bu yüksek lisans tezi çalışmasının yazılmasında, çalışmamı sahiplenerek takip eden danışmanım sayın Dr. Öğr. Üyesi Salim Ceyhan’a değerli katkı ve emekleri için teşekkürlerimi ve saygılarımı sunarım.

Savunma sınavı sunumu sırasında çalışmamın son haline gelmesindeki değerli katkıları adına değerli jüri üyelerine ayrı ayrı teşekkürlerimi sunarım.

Son olarak bugünlere ulaşmamda büyük katkı ve destekleri olan aileme teşekkür ederim.

Muhammet İkbāl YILDIZ

2024



ÖZET

AİLEVİ AKDENİZ ATEŞİ'NDE KOLŞİSİN TEDAVİSİNİN MAKİNE ÖĞRENMESİ ANALİZİ

Bu çalışmada Ailevi Akdeniz Ateşi hastalığı tespiti oldukça zor işlemler ve süreçler sonrası elde edilebilmektedir. Birçok hasta uzun bir süre boyunca hastalığın belirtileri taşımasına rağmen hastalık teşhisi yapılamamaktadır. Bunun asıl sebebi de birçok Dahiliye doktoru, hastalık belirtileri ile gelen hastalardan farklı hastalıklardan şüphe duymasıdır. Çünkü Ailevi Akdeniz Ateşi hastalığı günümüze kadar sık rastlanan hastalıklar arasında yer almamıştır. Bu çalışmada Ailevi Akdeniz Ateşi (FMF) hastalığının tespit edilmesi için yapılabilecek çalışmalar ve bu çalışmaların uygulama yöntemleri ile ilgili çalışmalar yer almaktadır.

Makine Öğrenmesi yöntemleri ile ilk olarak hastalıkta yer alan önemli kriterlerinin belirlenmesi ve ardından kişinin Ailevi Akdeniz Ateşi hastası olup olmama ihtimali hakkında doktora bilgi vermesi sağlanacağı düşünülmüştür. Fakat veri setinin böyle bir sınıflandırma problemini ele almak için yetersiz olması nedeniyle çalışma hastalık tedavisinde kullanılan en önemli ilaçlardan biri olan kolşisin tedavisine hastaların yanıt vermesi veya vermemesi durumlarını incelemek olarak araştırmaya yeni bir yön verilmiştir. Bu bağlamda, Ailevi Akdeniz Ateşi (AAA) hastalığına ait demografik bilgiler, klinik semptomlar ve genetik varyantlar gibi çeşitli özellikler içeren özelleştirilmiş bir veri seti üzerinde çalışma gerçekleştirilmiştir. Özellik seçimi sürecinde, χ^2 -değerine dayalı SelectKBest algoritması kullanılarak 8 kritik özellik belirlenmiştir.

Bu özellikler üzerinde gerçekleştirilen analizde, Logistic Regression modeli dikkate değer sonuçlar elde etmiştir. Bu model, ortalama doğruluk oranı %85.42, ortalama hassasiyet %85.81, ortalama geri çağırma %99.17 ve ortalama F1 skoru %92.00 olarak belirlenmiştir. Bu sonuçlar, Ailevi Akdeniz Ateşi hastalığının teşhisinde yüksek doğruluk ve hassasiyetle çalışan bir modelin geliştirilmesine olanak tanımaktadır. Bu bulgular, Ailevi Akdeniz Ateşi hastalığının erken teşhisinde ve kişiselleştirilmiş tedavi stratejilerinin geliştirilmesinde önemli bir katkı sağlayabilir. Çalışma, bu alandaki daha geniş kapsamlı araştırmalar için sağlam bir temel oluşturmaktadır ve alanında önemli bir adım olarak değerlendirilebilir.

Anahtar Kelimeler: Makine Öğrenmesi, SelectKBest, Ailevi Akdeniz Ateşi, Kolşisin Tedavisi, Sağlıkta Keşifsel Veri Analizi, Öznitelik Seçimi

ABSTRACT

MACHINE LEARNING ANALYSIS OF COLCHICINE TREATMENT IN FAMILIAL MEDITERRANEAN FEVER

In this study, the detection of Familial Mediterranean Fever (FMF) disease is challenging and can be achieved after complex processes and procedures. Many patients, despite exhibiting symptoms of the disease for an extended period, cannot receive a diagnosis. The main reason for this is that many internal medicine doctors harbor suspicions of different diseases when presented with patients showing symptoms of the disease. This is because Familial Mediterranean Fever has not been commonly encountered among diseases until today. This study encompasses research on the methods and application techniques that can be employed for the detection of Familial Mediterranean Fever disease.

The learning method is utilized, and machine learning methods are employed initially to identify the crucial criteria present in the disease. Subsequently, it is envisioned that the patient will provide information to the doctor about the likelihood of having Familial Mediterranean Fever. However, due to the inadequacy of the dataset, the study has been revised to investigate the response or lack of response of patients to colchicine, one of the most important drugs used in the treatment of the disease. In this context, a customized dataset containing demographic information, clinical symptoms, and genetic variants related to Familial Mediterranean Fever was analyzed. In the feature selection process, 8 critical features were identified using the SelectKBest algorithm based on ANOVA F-value.

In the analysis conducted on these features, the Logistic Regression model yielded noteworthy results. This model determined an average accuracy rate of 85.42%, average precision of 85.81%, average recall of 99.17%, and an average F1 score of 92.00. These results provide the opportunity to develop a model that operates with high accuracy and precision in the diagnosis of Familial Mediterranean Fever. These findings can contribute significantly to the early diagnosis of FMF and the development of personalized treatment strategies. The study establishes a robust foundation for broader research in this field and can be considered a significant step in the field.

Keywords: Machine Learning, SelectKBest, Familial Mediterranean Fever, Colchicine Treatment, Exploratory Data Analysis in Healthcare, Feature Selection

İÇİNDEKİLER

Sayfa

ÖN SÖZ.....	i
ÖZET	ii
ABSTRACT	iii
İÇİNDEKİLER.....	iv
TABLolar LİSTESİ.....	vii
ŞEKİLLER LİSTESİ.....	viii
KISALTMALAR VE SİMGELER LİSTESİ	ix
1. GİRİŞ.....	1
1.1. Tezin Amacı.....	1
1.2. Literatür Araştırması.....	2
2. FMF HASTALIĞI BULGULARI	3
2.1. FMF Ataklarının Ortak Özellikleri	3
2.2. FMF Ataklarının Ortak Özellikleri	4
2.3. Abdominal Ataklar	4
2.4. Eklem Atakları.....	5
2.5. Göğüs Ağrısı.....	5
2.6. Cilt Atakları	6
2.7. Kas Bulguları	6
2.8. Amiloid	7
2.9. Klinik Tanı	7
2.10. Tedavi	8
3. MATERYAL VE YÖNTEMLER.....	10
3.1. Veri Doldurma (Data Imputation).....	11
3.1.1. Veri Doldurma Yöntemleri	11
3.1.1.1. KNNImputer Yöntemi	12

3.1.2. Veri Doldurmanın Zorlukları ve Dikkat Edilmesi Gereken Hususlar	13
3.2. Keşifsel Veri Analizi (Exploratory Data Analysis)	13
3.2.1. EDA'nın Makine Öğrenmesindeki Amaçları	14
3.2.2. EDA Teknikleri ve Uygulamaları	14
3.2.3. EDA'nın Zorlukları ve Dikkat Edilmesi Gereken Hususlar	14
3.3. Özellik Seçimi ve Makine Öğrenmesi Modellerinin Kapsamlı Değerlendirilmesi	15
3.3.1. SelectKBest Yöntemi	15
3.3.2. K-Means Kümeleme Yöntemi	17
3.3.3 Kruskal-Wallis Testi	20
3.3.4 Post-hoc (Dunn's) Testi	20
3.4. Lojistik Regresyon	20
3.5. Random Forest	21
3.5.1. Ağacın Yapısı:	21
3.6. Destek Vektör Makineleri (Support Vector Machine SVM)	22
3.7. En Yakın Komşuluk (K-Nearest Neighbors)	23
3.8. Gradient Boost	23
3.9. Karar Ağaçları	24
4. BULGULAR	26
4.1. Veri Seti	26
4.2. Eksik Veriler	29
4.3. Veri Ön İşleme ve Değişken Seçimi	30
4.4. Demografik ve Coğrafi Özelliklerin İncelenmesi	31
4.5. Kolşisin Yanıtının Demografik ve Klinik Gözlemlerle İlişkisi	31
4.5.1. Kolşisin Yanıtsız Hastaların Özellikleri	34
4.5.1.1. Demografik ve Coğrafi Özellikler	34
4.5.1.2. Klinik Özellikler	34
4.5.1.3. Aile Öyküsü ve Genetik Faktörler	34

5. AİLEVİ AKDENİZ ATEŞİ HASTALARININ ANALİZİ.....	35
5.1. Veri Ön İşleme ve Makine Öğrenmesi Modelleri	35
5.1. Veri Setinin Düzenlenmesi ve One-Hot Encoding İşlemi.....	35
5.3. One-Hot Encoding Yöntemi	35
5.4. Veri Setinin Yapısı.....	35
5.5. Makine Öğrenmesi Modelleri, Özellik Seçimi, Çapraz Doğrulama ve Performans Analizi	36
5.6. Seçilen Özellikler	36
5.7. Makine Öğrenmesi Modelleri ve Performansları.....	36
5.8. Model Değerlendirmesi ve Anlamları.....	37
5.8.1. Özelliklerin İstatistiksel Analizi: Kruskal-Wallis ve Post-hoc Testler.....	37
5.9. Kruskal-Wallis Testi.....	37
5.9.1. Post-hoc (Dunn's) Test Sonuçları: Sayısal Analizler ve Değerlendirmeler	38
6. SONUÇLAR VE ÖNERİLER.....	41
6.1. Sonuçlar	41
6.2. Öneriler.....	41
6.3. Sonuç.....	42
KAYNAKÇA	43

TABLÖLER LİSTESİ

Sayfa

Tablo 1.1. Performans metriklerini hesaplama parametreleri	16
Tablo 1.2. Veri setindeki kategorik değişkenlerin özellikleri, açıklamaları, mümkün değerleri ve her değerın görülme oranları	27
Tablo 1.3. Veri setindeki sayısal değişkenlere ait tanımlayıcı istatistikler	29
Tablo 2.1. Anlamlı özellikler için model performansları	37
Tablo 2.2. Kruskal-Wallis testi sonuçları	38
Tablo 2.3. atksur Özelliđi (Atak Süresi)	39
Tablo 2.4. atksklg Özelliđi (Atak Sıklıđı)	39
Tablo 2.5. ates_yok Özelliđi (Ateş Yokluđu)	39
Tablo 2.6. atkaktfazyk_yok Özelliđi (TİT Protein Varlıđının Yokluđu)	40
Tablo 2.7. titprteinvar_yok Özelliđi (Atak Süresi)	40

ŞEKİLLER LİSTESİ

Sayfa

Şekil 1.1. Veri ön işlem adımları, özellik seçim süreci ve makine öğrenmesi adımları	10
Şekil 1.2. Küme Yapısı	18
Şekil 1.3. Random Forest için oluşturan tek ağaçların yapısı	22
Şekil 2.1. Veri setindeki eksik verilerin yüzdelik değerleri	30
Şekil 2.2.. Coğrafi bölgelerin dağılımı.....	31
Şekil 2.3. AAA hastalarında kolşisin tedavisine verilen yanıtların ateş, karın ağrısı atağı ve atak fazı sıklığı semptomlarının varlığına ve yokluğuna göre sayısal dağılımları	32
Şekil 2.4. Kolsişin yanıtına göre Atak Süresi, TİT Protein ve Ateş değişkenlerinin dağılımları	33

KISALTMALAR VE SİMGELER LİSTESİ

FMF	: Ailevi Akdeniz Ateşi (Familial Mediterranean Fever)
EDA	: Keşifsel Veri Analizi
KNN	: k-nearest neighbors
DYER	: Doğum Yeri
ailefmf	: Aile AAA Öyküsü
krnagrsatk	: Karın Ağrısı Atak
klcadzykblkagrs	: Kalça, Diz ve Ayak Bilek Ağrısı
ateş	: Ateş
kolsisinynt	: Kolşisin Yanıtı
ebe	: Deri Lezyonu
atkaktfazyk	: Atak Akut Faz Yüksekliği Sıklığı
titrteinvar	: TİT’de protein varlığı
yygnmutvar	: MEFV mutasyon varlığı
MEFV	: MEFV geninin mutasyonları
Atksur	: Atak Süresi
Atksklg	: Atak Sıklığı
LRA	: Lojistik Regresyon
RF	: Random Forest

1. GİRİŞ

Ailevi Akdeniz Ateşi (FMF), ateşle birlikte gelen ve seröz zarların ağırlı, enfeksiyon dışı iltihabi ataklarıyla tanınan bir hastalıktır. Bu hastalık, zamanla amiloidoz gibi komplikasyonlara yol açabilir. FMF, otozomal resesif bir kalıtım modeliyle geçer ve bazı etnik gruplarda daha yaygındır. Bu gruplar arasında Yahudiler, Ermeniler, Araplar ve özellikle Türkler bulunmaktadır. Dünya genelinde de nadiren görülen bu hastalık, Türkiye'de oldukça yaygındır. FMF'nin erken teşhisi, hastalığın yönetimi ve olası komplikasyonların önlenmesi açısından büyük önem taşımaktadır.

Bu tez çalışmasında; günümüzde genetik hastalıklar sık rastlanır olmaya başladı. Birçok kişi genetik rahatsızlık yaşıyor veya dolaylı yoldan kalıtsal olarak taşıyıcı olarak gözüküyor. Bu tez çalışmasında yapılan makine öğrenmesi projesi ile Ailevi Akdeniz Ateşi (FMF) hastalığının erken teşhisi üzerine çalışılmaktadır.

1.1. Tezin Amacı

Bu tez, özellikle Doğu Akdeniz bölgesinde yaygın olarak görülen genetik bir inflamatuvar hastalık olan Ailevi Akdeniz Ateşi (AAA)'nin (Tunca, 2005:81) incelenmesini hedeflemektedir. Bu hastalık, Mansueto (Mansueto, 2022:14) tarafından belirtildiği üzere, yüksek ateş, karın ağrısı ve akciğer zarı iltihabı gibi semptomlarla karakterize edilir. Kolşisin, 1970'lerden bu yana AAA tedavisinde önemli bir rol oynamakta olup, bu ilacın etkinliği Zemer ve Livneh (Zemer ve Livneh, 1991:34) tarafından vurgulanmıştır. Ancak, kolşisine dirençli vakaların varlığı, tedavi yaklaşımlarının ve uygulamalarının geliştirilmesinin gerekliliğini ortaya koymaktadır.

Bu bağlamda, makine öğrenimi ve veri analizi metodolojileri, Adato (Adato, 2022:12) 'nın da belirttiği üzere, hastalığın detaylı analizine katkıda bulunabilir. Bu teknikler, yaş, cinsiyet, semptomlar ve genetik özellikler gibi çeşitli parametrelerin hastalık üzerindeki etkilerinin incelenmesinde önemli bir rol oynayabilir. Özellik seçimi, modelin genelleştirici gücünü artırma ve daha derinlemesine analiz yapma açısından kritik bir öneme sahiptir.

Bu araştırma, Türkiye'nin çeşitli bölgelerinden seçilen ve Uludağ Üniversitesi Tıp Fakültesi'nde tedavi gören 144 AAA hastasının analizini içermektedir. Bu çalışmada özellikle dikkat çeken nokta, kolşisin tedavisine yanıt vermeyen 22 hastanın ayrıntılı incelemesidir. Bu oran, yaklaşık %15.28 olup, önceki araştırmalarda belirtilen %5-10'luk oranın üstündedir. Bu durumun, üniversite hastanesinde daha yoğun olarak takip edilen dirençli hastaların yüksek oranından kaynaklanabileceği düşünülmektedir.

Bu tezin temel amacı, makine öğrenimi teknikleri ve özellik seçimi kullanılarak, AAA hastalarının tedaviye yanıtını etkileyen özellikleri belirlemektir. Bu bağlamda, "atak süresi," "ateş," ve "TİT protein" özellikleri en bilgi verici üç parametre olarak tespit edilmiştir. Elde edilen bu bilgilerin, hastalığın daha etkin yönetilmesine ve tedavi protokollerinin geliştirilmesine önemli katkılar sağlayabileceği öngörülmektedir.

1.2. Literatür İncelemesi

Ailevi Akdeniz Ateşi (AAA), çeşitli çalışmalarda genetik, çevresel faktörler ve tedavi yöntemleri açısından ele alınmıştır. Örneğin, Türkiye'de yapılan bir araştırma, hastalığı etkileyebilecek genetik ve çevresel etkenleri derinlemesine inceleyerek bu alandaki bilgi birikimine katkıda bulunmuştur (Tunca, 2005). AAA üzerine yapılan başka bir kontrollü, çift-kör çalışma ise kolşisin tedavisinin etkinliğini değerlendirmiştir. Bu çalışmada, kolşisinin atak frekansını önemli derecede azalttığı ve bu sonuçların istatistiksel olarak anlamlı olduğu tespit edilmiştir (Goldstein, 1974). Kolşisin tedavisine yanıt vermeyen hastalar için alternatif bir tedavi olarak Anakinra'nın değerlendirildiği bir çalışmada, Anakinra'nın Interlökin-1'e (IL-1) antagonist olarak etki ederek inflamasyonu kontrol altına alabileceği ve bu durumda etkili bir alternatif olabileceği vurgulanmıştır (Parvaneh, 2015). Ayrıca, ilaç tedavisine yeterli yanıt alınamayan vakalar için yapılan bir meta-analiz, alternatif tedavi yöntemlerinin genel etkinliğini değerlendirmiştir (Goldstein, 1974).

Ozgocmen ve Akgul tarafından gerçekleştirilen bir çalışmada, AAA hastalarında anti-TNF ajanlarının (örneğin infliximab, adalimumab) etkinliği incelenmiş ve özellikle spondilitis ve amiloidoz gibi komplikasyonları olan hastalarda bu ajanların etkili bir tedavi seçeneği olabileceği gözlemlenmiştir (Ozgocmen, 2021). Sarı, Birlik ve Kaşifoğlu'nun gözden geçirme makalesinde ise, colchicine'in AAA tedavisindeki önemli yeri tartışılmış ve colchicine'e dirençli ya da intolerant vakalar için alternatif tedavi seçeneklerinin geliştirilmesinin gerekliliği üzerinde durulmuştur. Bu bağlamda, IL-1 ve TNF inhibitörleri gibi biyolojik ajanların bu tür hastalarda umut verici sonuçlar sunabileceği belirtilmiştir (Sarı, 2014). Ayrıca, Adato'nun çalışması, AAA ile ilişkili MEFV gen varyantlarının etnik ve coğrafi kökenlerini analiz etmek için makine öğrenimi tekniklerini kullanmıştır. Bu çalışma, belirli MEFV varyantlarının Kuzey Afrika, Avrupa ve Batı Asya gibi coğrafi bölgelerle güçlü bir ilişkisi olduğunu ortaya koyarak, hastalığın etiyolojisi ve tedavi yöntemleri hakkında yeni perspektifler sunmaktadır (Adato, 2007).

2. FMF HASTALIĞI BULGULARI

Michigan Ailevi Akdeniz Ateşi, ateşin eşlik ettiği, seröz zarların ağrılı, non-enfeksiyöz enflamasyon nöbetleri ve zaman içinde amiloidoz gelişimi ile belirlenen bir sağlık sorunudur (Koşan, 2003:1). Bu hastalık, otozomal resesif kalıtımla geçer ve özellikle belirli toplumlarda yaygın olarak görülmektedir (Sohar, 1967:53). FMF, özellikle Yahudiler, Ermeniler ve Araplar gibi etnik gruplarda sık rastlanan bir durumdur; bununla birlikte, Türkler de hastalıktan önemli ölçüde etkilenebilir (Yazıcı, 1997:66). Ayrıca, hastalığa dünyanın çeşitli bölgelerinde sporadik olarak rastlanabilmektedir (Müferet, 2006:90).

Ülkemizde yaygın bir sağlık sorunu olan FMF hastalığının erken teşhisi, tedavi süreci ve potansiyel komplikasyonların önlenmesi açısından büyük bir öneme sahiptir. Hastalığın belirtileri arasında ateş, ağrılı seröz zar enflamasyon nöbetleri ve amiloidoz gelişimi bulunmaktadır (Koşan, 2003:6). Bu belirtiler, hastalığın tanısını koymak için kritik ipuçları sunar. Özellikle risk altındaki etnik gruplarda, erken tanı ve tedavi, hastalığın seyrini önemli ölçüde etkileyebilir.

FMF, genetik bir temele dayanır ve hastalığın kalıtım yoluyla geçmesi, risk altındaki popülasyonlarda bu duruma karşı duyarlılığı artırır (Sohar, 1967:53). Bu nedenle, özellikle risk altındaki gruplarda genetik tarama ve danışmanlık önemlidir. Sağlık uzmanları, hastalığın erken aşamalarında müdahale ederek, hastalığın ilerlemesini önleyebilir ve yaşam kalitesini artırabilirler. Bu bağlamda, toplumlar arası farkındalık ve eğitim de büyük bir rol oynamaktadır. (Koşan, 2003:6). FMF'nin etnik gruplar arasında farklı prevalansları vurgulayarak, sağlık politikalarının bu özel durumları dikkate alacak şekilde oluşturulması önemlidir.

2.1. FMF Ataklarının Ortak Özellikleri

FMF, belirgin ateş ve ağrı atakları ile karakterize edilen bir hastalıktır, ve hastalığın klinik tablosu genellikle tekrarlayan ateş ve serozit ataklarından oluşur. Ataklar arasında ise hasta genellikle tamamen sağlıklıdır ve uzun bir süre ataksız dönem geçirebilir. Hastaların %90'ında, ilk atak 20 yaşından önce ortaya çıkar (Kavak, 2003:40). Bununla birlikte, hastaların %50'sinde hastalık belirtileri yaşamın ilk on yılı içinde, %5'inde ise otuz yaşından sonra görülebilir (Paşa vd., 2008:43).

Ataklar genellikle 2-4 gün arasında değişen bir sürede sürer, ancak bazı durumlarda daha uzun veya daha kısa süreli ataklar görülebilir. Tetikleyici faktörler genellikle belirsizdir, ancak enfeksiyonlar veya stresin önemli bir rol oynadığı düşünülmektedir. Ataklar genellikle aniden ortaya çıkar, ancak bazı hastalarda sinirlilik, baş dönmesi, iştah artışı, tat duyusunda

değişiklikler gibi çeşitli duysal ve fiziksel belirtiler prodromal olarak ortaya çıkabilir (Lider ve Livneh, 2007:65).

Ataklar sırasında klinik bulgular çeşitli şekillerde olabilir, ancak en sık görülen nöbet şekli ateş, karın ağrısı ve eklem bulgularının bir arada olduğu nöbettir (Kasapcopur ve Arısoy, 2006:17). Bu belirtiler, hastalığın teşhisinde kritik öneme sahiptir ve hastalığın seyrini anlamak açısından önemli ipuçları sunar. Bu nedenle, hastaların belirtilerini dikkatlice değerlendirmek ve tanıyı doğru bir şekilde koymak, etkili bir tedavi planı oluşturmak için temel bir adımdır (Kasapcopur ve Arısoy, 2006:17).

2.2. FMF Ataklarının Ortak Özellikleri

Ataklar FMF'ye özgü ataklar nadiren sadece ateşle seyredebilir; ancak çoğunlukla diğer klinik bulgularla birlikte görülürler. Ateş, atak boyunca genellikle yüksek seyreder (Kasapcopur ve Arısoy, 2006:17). Pratikte, neredeyse tüm hastaların bir döneminde ateşli atakları vardır. Ancak, az sayıda hastada ateşsiz ataklar da tanımlanmıştır. Ateşin şiddeti, hafiften 40°C'ye kadar değişebilir. Ateş, atak sırasında birkaç saatten dört güne kadar yüksek kalabilir, ancak genellikle 24 saat içinde düşer (Koşan, 2003:6).

Ataklar Bazı durumlarda, hastaların ateşi belirgin bir şekilde yükselmez ve bu nedenle ataklar gözden kaçabilir. Aynı hastanın farklı ataklarında, ateş bazen yüksekken, bazen normal değerlerde ölçülebilir. Eklem atakları sırasında ise sistemik ateş görülmeyebilir (Kasapcopur ve Arısoy, 2006:17). Özellikle kolşisin tedavisi alan hastalarda, ataklar sırasında ateş görülmeyebilir. Bu durum, hastalığın tedavi edilmesiyle ilişkilendirilen özelliklerden biridir ve tedavi planının etkisini değerlendirmek için ateşin tek başına bir gösterge olarak kullanılmasının sınırlamalarını ortaya koymaktadır. Bu nedenle, hastalığın geniş bir klinik spektrumda ortaya çıkabilen çeşitli semptomlarını değerlendirmek, doğru teşhis ve etkili tedavi için önemlidir.

2.3. Abdominal Ataklar

Ataklar FMF hastalarının %95'inde en yaygın belirti, genellikle yatak istirahati gerektirecek kadar şiddetli, peritoniti taklit edecek kadar diffüz veya lokalize karın ağrısıdır. Bu karın ağrısı, ateş başlamadan birkaç saat önce ortaya çıkabilir ve ateş düştükten sonra 1-2 gün boyunca devam edebilir. Akut karın tablosu, bazı hastalarda görülebilir ve bu durum, pek çok hastaya gereksiz yere apendektomi ve laparotomi gibi cerrahi müdahalelerin uygulanmasına neden olabilir (Soher vd., 1967:53).

Ataklar FMF'nin belirgin belirtileri arasında yer alan bu karın ağrısı, hastaların günlük yaşamını önemli ölçüde etkileyebilir (Lider ve Livneh, 2007:65). Bu belirti, hastalığın tanısında kilit bir role sahiptir ve tedavi sürecinin yönlendirilmesinde önemli bir ipucu sağlar. Erken teşhis, hastalığın ilerlemesini önleyebilir ve hastaların yaşam kalitesini artırabilir. Ayrıca, akut karın tablosunun neden olduğu potansiyel cerrahi müdahale riskini değerlendirmek, hastalara daha etkili ve özenli bir sağlık hizmeti sunmada kritik bir faktördür.

2.4. Eklem Atakları

Ataklar Ateş ve karın ağrısından sonra, FMF'de en sık görülen üçüncü klinik belirti eklem tutulumudur ve bu durum hastaların %75'inde ortaya çıkar (Soher vd., 1967:53). Ateş ve karın ağrısı olmaksızın da gelişebilir. Eklem tutulumu, olguların %70'inde artrit ve %30'unda artralji şeklinde gözlenir. Klasik olarak, artrit gezici olmayan, eklemde tahribat yapmayan, sekel bırakmayan, mono veya oligoartiküler bir tarza sahiptir. Genellikle büyük eklemleri ve en sık alt ekstremit eklemlerini etkiler. Eklem atakları genellikle birkaç gün veya 1-2 hafta içinde kendiliğinden kaybolur. Ayak bileğinde oluşan artritlerin %50'sinde ayak sırtında eritem gözlemlenir.

Hastaların %95'inde akut artrit atakları gözlenirken, %5'inde kronik artrit gelişebilir (Örün ve Yalçınkaya, 2003:7). Bu kronik artritler genellikle küçük travmalar veya uzun yürüyüş gibi olaylar tarafından tetiklenebilir. Akut atak sırasında, eklemdeki sinovyal sıvı genellikle steril olup, sinovitin şiddetine bağlı olarak görünümü hafif bulanıktan pürülana kadar değişebilir (Koşan, 2003:6). Birkaç ay veya bir yıla kadar uzayan, spontan ve sekelsiz kaybolan kronik seyirli artritlerin %5 kadarında geriye dönüşümü olmayan, hatta artroplasti gerektirebilecek değişiklikler gözlemlenebilir. Kronik artritler genellikle kalça, diz eklemi etkilerken, ayak bileği, temporomandibular veya sternoklavikular eklemler de etkilenebilir (Lider ve Brik:2007:24).

2.5. Göğüs Ağrısı

FMF'li hastalarda göğüs ağrısı, vakaların %30'unda rapor edilmiştir (Lider ve Brik:2007:24). Bu göğüs ağrıları, genellikle plörezi nedeniyle tek taraflı ve şiddetli olabilir. Fizik muayenede, efüzyona bağlı olarak akciğer seslerinde azalma, bazen frotman veya perküsyonla dolgunluk tespit edilebilir; ancak genellikle belirgin bir fizik muayene bulgusu gözlenmez.

FMF'li hastaların %0-5'inde tekrarlayan perikardit vakaları bildirilmiştir. FMF perikarditi genellikle hastalığın geç aşamalarında görülme eğilimindedir (Ben-Chetrit ve Levy,

1998:351). Akciğer grafilerinde çekilen görüntülerde, hastaların bir kısmında geçici efüzyon, efüzyona bağlı olarak kostafrenik sinüste küntleşme ve nadiren atelektaziler gözlemlenebilir. Bu göğüsle ilgili belirtiler, hastaların klinik seyrini değerlendirmede ve uygun tedavi stratejilerini belirlemede önemli ipuçları sağlar. FMF'nin çeşitli organları etkileyebilen sistemik bir hastalık olduğu göz önüne alındığında, göğüs ağrısı ve ilişkili bulguların doğru değerlendirilmesi, bireyin genel sağlığına yönelik kapsamlı bir yaklaşımın bir parçasıdır.

2.6. Cilt Atakları

FMF'de en karakteristik cilt lezyonu, erizipel benzeri eritemdir ve görülme sıklığı %3-46 olarak bildirilmiştir (Özel vd., 1971:50). Bu lezyonlar genellikle en sık alt ekstremitelerde, bacağın ön yüzünde, ayak bileğinde, ayak sırtında veya malleol üzerinde ortaya çıkar. Tek taraflı, pembe-mor renkli, yaklaşık 10-15 cm çapında, ciltten hafif kabarık eritem şeklindeki döküntülerle karakterizedir. Lezyonun bulunduğu cilt bölgesi gergin, ödemli ve ısı artışı gösterebilir (Haghighat vd., 2006:7). Semptomlar genellikle 2-3 gün içinde kendiliğinden geriler (Yavuz vd., 2011:342).

Erizipel benzeri döküntü dışında, alt ekstremitelerde purpura, cilt altı nodül, makülo-papüler döküntü ve ürtiker gibi farklı cilt lezyonları da görülebilir. Bu çeşitlilik, FMF'nin ciltle ilgili belirtilerini değerlendirirken dikkate alınması gereken önemli bir faktördür. Bu lezyonlar, hastalığın klinik tablosunu tamamlar ve doğru teşhisin konulmasında önemli bir rol oynar. Cilt belirtileri, hastalığın genel seyrini değerlendirmek ve uygun tedavi stratejilerini belirlemek için önemli bir klinik gösterge sunar.

2.7. Kas Bulguları

FMF hastalarının 4'te 1'inde kas ağrıları görülebilmektedir. Bu kas ağrıları, hastada kendiliğinden geçen, egzersize bağlı ya da uzamış febril miyalji sendromu şeklinde klinik bir tablo oluşturabilir (Haghighat vd., 2006:7). Klasik FMF miyaljisi ile karşılaştırıldığında, uzamış febril miyaljiye göre ağrı, hassasiyet ve fonksiyon kaybının daha hafif olduğu belirtilmektedir. Uzamış febril miyalji durumunda genellikle eklem bulguları gözlenmez ve kas enzimleri, elektromiyografik incelemeler ve kas biyopsisi gibi test sonuçları genellikle normal çıkar (Kasapcopur ve Özdoğan, 2007:54).

FMF'de Uzamış febril miyalji, FMF'nin nadir görülen dramatik klinik belirtilerinden biridir ve bu durum, kolşisin tedavisine rağmen ortaya çıkabilir ve tedavi için kortikosteroid gerektirebilir (Haghighat, 2006:7). Bu belirtiler, hastalığın geniş bir semptom yelpazesine işaret eder ve klinik değerlendirme sürecinde önemli bir rol oynar. Kas ağrılarına yönelik detaylı bir

değerlendirme yapılması, hastanın yaşam kalitesini etkileyen faktörleri anlamak ve uygun tedavi stratejilerini belirlemek için önemlidir.

2.8. Amiloid

FMF hastalığının seyri üzerinde belirleyici bir rol oynayan temel özellik, amiloidozun varlığıdır. Bu hastalıkta, enfeksiyon, malignite, doku hasarı, FMF atağı ve diğer inflamatuvar olaylar sırasında ortaya çıkan bir akut faz reaktanının parçalanma ürünü olan Serum amiloid A (SAA) adlı proteinin kritik bir rol oynadığı düşünülmektedir (Kivity, 2011:13). Amiloidozla ilişkilendirilen başlıca klinik belirti genellikle proteinüri olarak kendini gösterir (Kivity, 2011:13). Zaman içinde proteinürinin nefrotik sendrom, üremi ve nihayetinde böbrek nakli veya son dönem böbrek yetmezliği gibi durumları tetikleyebileceği bilinmektedir.

Hastalar genellikle normotansif ve non-hematüriktir. Kolşisin tedavisi, FMF hastalarının sadece küçük bir bölümünde amiloidoz gelişimine neden olmaktadır. Yapılan çalışmalar, etnik kökenin, kalıtımın ve çevresel faktörlerin amiloidoza yakalanma riskini etkileyebileceğini göstermiştir (Shohat M vd., 1999:92).

Amiloid, yavaşça çeşitli organ ve dokularda birikebilir, ancak bu birikim özellikle böbreklerde organ disfonksiyonunun belirgin bir belirtisi olarak ortaya çıkar. Eğer tanı anında amiloidoz gelişmemişse, kolşisin tedavisiyle FMF hastalarının ortalama yaşam beklentisi genellikle sağlıklı bireylerinkine benzerdir (Kavak, 2003:40). İnflamatuvar olayın şiddeti ile amiloidoz sıklığı arasında belirgin bir ilişki tespit edilememiştir. Amiloidoz gelişmiş olan FMF hastalarının bir kısmında geçmişe ait herhangi bir atak öyküsü bulunmamakta ve bu durum, subklinik inflamasyonun varlığını ve akut atak olmadan amiloidoz ve nefrotik sendrom gelişimini açıklamaktadır (Ben-Chetrit ve Levy, 1998:351).

Amiloid tanısı, tutulan organın biyopsisinde amiloid birikiminin gösterilmesi ile konulur. Bu amaçla en sık renal ve rektal biyopsi kullanılmaktadır (Sarıman, 2011:192). Kalıcı proteinüri, diyaliz tedavisi veya böbrek nakli gerektiren son dönem böbrek hastalığı olan FMF hastalarında, kolşisin tedavisi en az 2 mg/gün dozunda devam etmelidir (Örün E ve Yalçınkaya, 2003:7). FMF hastalığının seyri ve amiloidozun gelişimi üzerindeki etkileşimler hala detaylı bir şekilde anlaşılamamış olup, bu konuda yapılan ileri araştırmalar, hastalığın daha etkili yönetimi için kritik öneme sahiptir.

2.9. Klinik Tanı

FMF hastalığı için spesifik bir kesin tanı koydurucu muayene bulgusu veya özgün bir laboratuvar testi mevcut değildir. FMF tanısı, klinik bulgular, aile öyküsü, biyokimyasal ve

genetik laboratuvar verileri, tedaviye verilen yanıt ve diğer ailesel periyodik ateş sendromlarının dışlanması temelinde konulmaktadır. Genetik analiz, FMF hastalığı tanısına destek sağlayabilir; ancak bu, kesin tanı koymak için yeterli olmayan bir kriterdir.

Eğer hastanın atak sırasında inceleniyorsa, atağa eşlik eden inflamatuvar bulguların varlığı (lökositoz, sedimantasyon hızı (ESR), C-reaktif protein (CRP) ve fibrinojen yükselmesi) ve bu testlerin atak sona erdikten sonra normale dönmesi, tanıya katkı sağlayabilir. Ancak, bu test sonuçlarının pozitif bulunmasının FMF'ye özgü olmadığı ve sadece vücutta inflamasyonun varlığına işaret ettiği unutulmamalıdır.

Klinik pratikte, hastalığın tanısını koymak için belirli kabul edilmiş tanı kriterleri kullanılmaktadır (Akkuş vd., 2002:20). Bu kriterler, hastanın klinik özellikleri, aile öyküsü ve laboratuvar bulgularını içerir, ancak FMF tanısının konulmasında mutlak bir güvence sağlamaz. Hekimler, bu kriterlere dayanarak klinik değerlendirme yapar ve benzer periyodik ateş sendromlarını dışlamak için gerekli incelemeleri gerçekleştirir. FMF tanısının konulması sürecinde multidisipliner bir yaklaşım benimsenmeli ve klinik bulguların, genetik verilerin ve laboratuvar test sonuçlarının birleştirilmesiyle sağlam bir teşhis elde edilmeye çalışılmalıdır.

2.10. Tedavi

Fenantren türevi olan kolşisin, bitkisel kökenli bir bileşik olarak bilinir ve mitozla metafazda etkileyerek hücre bölünmesini durdurma özelliğine sahiptir. Aynı zamanda monosit ile nötrofil kemotaksisini azaltır. Lökosit cAMP düzeyini arttırarak lizozomal degranülasyonu inhibe eder ve hücre duvarını sağlamlaştırır. Serum amiloid A (SAA) düzeyini azaltma yeteneği de bulunmaktadır (Schwade vd., 1977:23).

Yapılan araştırmalar, düzenli kolşisin tedavisinin Familial Akdeniz Ateşi (FMF) hastalarının çoğunda atak sayısını ve şiddetini azaltmanın yanı sıra, tüm hastalarda amiloid gelişimini önlediğini göstermiştir (Örün ve Yalçınkaya, 2003:7).

Kolşisin tedavisi genellikle 1 mg/gün dozunda başlar, yaş veya vücut ağırlığına bakılmaksızın. Remisyon elde edilene kadar bu doz, 0,5 mg/gün artışlarla kademeli olarak 1,5 mg veya 2 mg'a kadar çıkarılabilir. 1,5 mg ve daha yüksek dozlar, tedaviyi gün içinde iki doza bölerek verilmelidir; ancak bazen uyum için tek doz 1,5 mg/gün olarak da verilebilir. Dozların vücut ağırlığı veya yüzey alanına göre ayarlanmasını öneren yaklaşımlar da mevcuttur (Padeh, 2005:577).

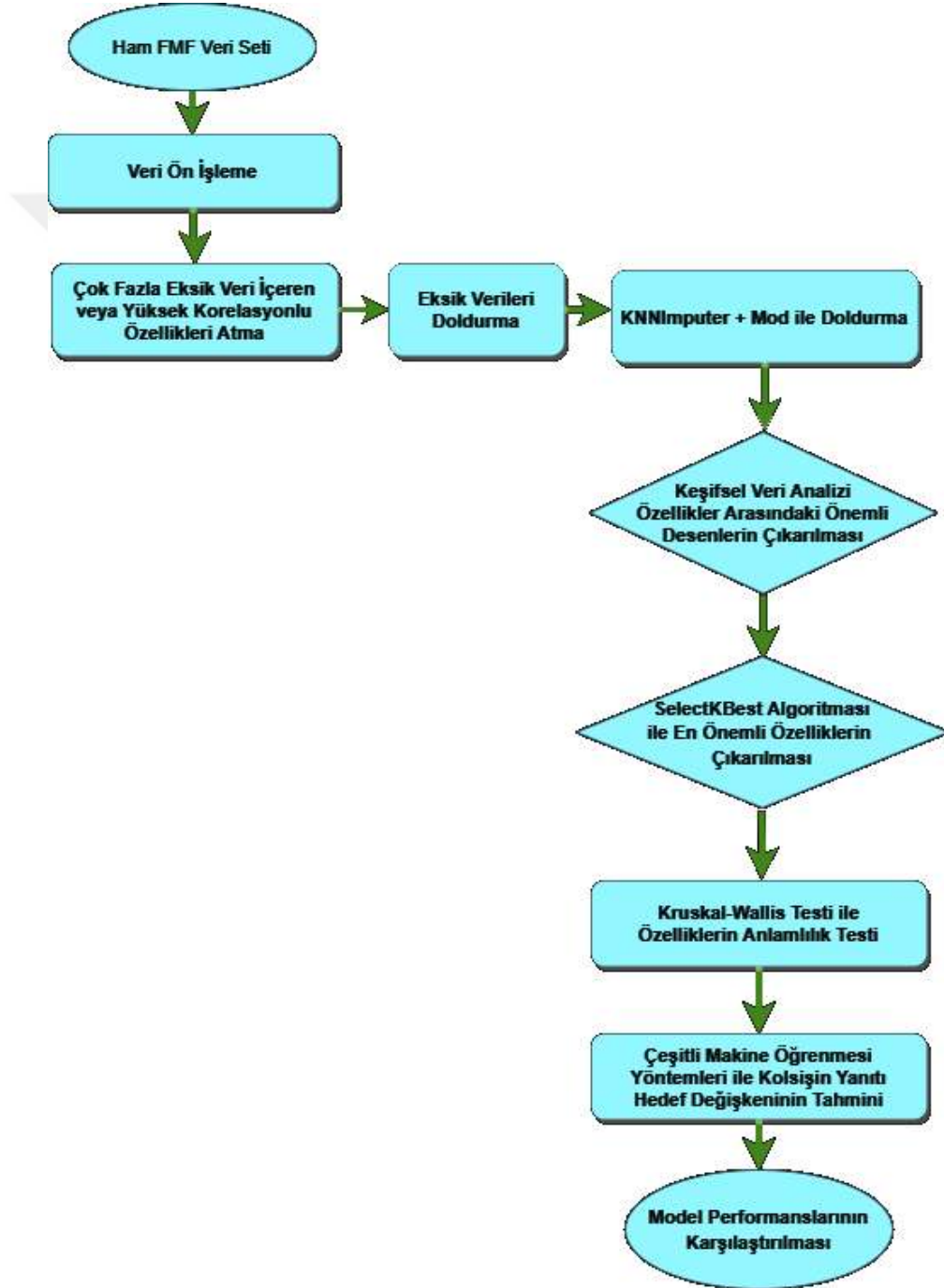
Çocuklarda kullanılabilecek kolşisin dozu konusunda belirgin bir fikir birliği bulunmamaktadır. Vücut yüzey alanı 1 m²'den fazla olan çocuklar için erişkin dozu bölünmüş

dozlarda kullanılabilir. Çocuklarda etkili doz genellikle 0,02-0,03 mg/kg/gün (maksimum 2 mg/gün) olarak önerilmektedir.



3. MATERYAL VE YÖNTEMLER

Bu tez çalışmasının "Materyal ve Yöntem" bölümü, araştırmanın metodolojik çerçevesini ve kullandığı analitik teknikleri ayrıntılı bir şekilde tanımlar. Çalışma, genetik bilgiler, laboratuvar sonuçları ve kişisel sağlık verileri gibi çeşitli veri kaynaklarından elde edilen zengin veri setlerini kullanır. Şekil 1’de verilen akış diyagramı, tez çalışmasında veri setinin hazırlanmasından başlayarak incelenen modellerin performanslarının karşılaştırılmasına kadar olan süreci detaylı bir şekilde göstermektedir.



Şekil 1.1. Veri ön işlem adımları, özellik seçim süreci ve makine öğrenmesi adımları

Akış diyagramından görülebileceği gibi, ilk olarak, ham FMF veri seti üzerinde gerekli ön işleme adımları gerçekleştirildi. Bu adımda, genetik bilgiler, laboratuvar sonuçları ve kişisel sağlık verileri birleştirilerek tek bir veri dosyası oluşturuldu. Daha sonra, çok fazla eksik veri içeren veya yüksek korelasyon gösteren özellikler atıldı ve eksik verileri doldurma aşamasında, KNNImputer ve mod ile doldurma yöntemleri kullanıldı. Bu yöntemler sayesinde, eksik verilerin modelin performansını olumsuz etkilemesinin önüne geçilmiş oldu.

Bir sonraki adım, keşifsel veri analizi (EDA) ile özellikler arasındaki önemli desenlerin çıkarılması işlemidir. Bu aşamanın ardından, SelectKBest algoritması kullanılarak en önemli özelliklerin seçilmesini sağlandı. Bu seçim ile özelliklerin model üzerindeki etkisini anlaşılması ve gereksiz özelliklerin modelleme sürecinden çıkarması sağlandı. Seçilen özelliklerin anlamlılık testi, Kruskal-Wallis testi ile gerçekleştirildi. Bu istatistiksel test, seçilen özelliklerin hedef değişken üzerinde anlamlı bir etkiye sahip olup olmadığını değerlendirmeye olanak tanımıştır.

Son olarak, tezde de bahsedilen çeşitli makine öğrenimi yöntemleri kullanarak hedef değişkenin tahmini yapıldı ve modellerin performansları karşılaştırmalı olarak değerlendirildi. Bu süreçte, en uygun modelin seçilmesi ve projenin başarısını maksimize etmek amacıyla kritik kararlar alındı ve sonuçlar tablo olarak verildi. Bu işlemlerin ayrıntıları takip eden bölümlerde sırasıyla verilmiştir.

3.1. Veri Doldurma (Data Imputation)

Veri doldurma süreci, veri setlerinde sıklıkla karşılaşılan eksik veri problemlerini ele alır. Bu süreç, eksik veri noktalarını tahmin ederek doldurmayı içerir ve veri setinin bütünlüğünü artırarak, daha güvenilir ve doğru veri analizi ve makine öğrenmesi modellerinin geliştirilmesini sağlar. Eksik veriler, özellikle büyük ve karmaşık veri setlerinde, analiz sonuçlarını yanıltıcı hale getirebilir ve modellerin performansını olumsuz etkileyebilir. Eksik verilerin etkili bir şekilde doldurulması, analiz için daha temsil edici ve kapsamlı bir veri seti oluşturur ve makine öğrenmesi modellerinin gerçek dünya durumlarını daha doğru bir şekilde yansıtmasına olanak tanır. Bu süreç, veri bilimi ve makine öğrenmesi projelerinde kritik bir öneme sahiptir ve eksik veri problemlerinin çözümünde temel bir rol oynar.

3.1.1. Veri Doldurma Yöntemleri

Veri doldurma yöntemleri, basit yöntemlerden karmaşık istatistiksel tekniklere kadar değişkenlik gösterir. Basit yöntemler arasında ortalama, medyan veya mod kullanarak eksik

verilerin doldurulması yer alır. Bu yöntemler, veri setinin genel dağılımını temsil eden bir değerle eksik verileri doldurmayı amaçlar. Daha karmaşık yöntemler ise, çoklu atama (multiple imputation), k-nearest neighbors (KNN) ve regresyon modellerini içerir. Bu yöntemler, eksik verilerin doldurulmasında veri setindeki diğer özniteliklerin bilgilerini kullanır. Özellikle, çoklu atama yöntemi, eksik verilerin birden fazla kez doldurulmasını ve sonuçların birleştirilmesini içerir, böylece eksik veri doldurma sürecinin belirsizliğini azaltır. KNN ve regresyon gibi yöntemler, eksik değerlerin tahmin edilmesinde veri setindeki benzer örnekleri veya öznitelik ilişkilerini kullanır.

3.1.1.1. KNNImputer Yöntemi

KNNImputer, eksik veri doldurma sürecinde kullanılan etkili bir yöntemdir ve K-En Yakın Komşu (K-Nearest Neighbors - KNN) algoritmasına dayanır. Bu yöntem, eksik veri içeren bir özelliği (feature), veri setindeki benzer örneklerin (komşuların) değerlerine dayanarak doldurur. KNNImputer'ın temel prensibi, her bir eksik veri noktası için, veri setindeki diğer tam verilere göre bir benzerlik ölçütü kullanarak en yakın 'k' komşuyu bulmaktır. Daha sonra, bu komşuların ilgili özelliği için ortalama veya medyan gibi istatistiksel bir ölçü kullanılarak eksik değer hesaplanır.

KNNImputer yöntemi, özellikle veri setindeki örnekler arasındaki ilişkilerin ve benzerliklerin önemli olduğu durumlarda etkilidir. Bu yöntem, veri setinin yapısını ve içsel ilişkilerini koruyarak, eksik verileri mantıklı ve gerçekçi bir şekilde doldurabilir. Ancak, KNNImputer'ın doğru çalışabilmesi için, veri setinin iyi ölçeklendirilmiş ve normalleştirilmiş olması önemlidir. Ayrıca, 'k' değerinin seçimi de kritik bir öneme sahiptir; çok düşük bir 'k' değeri, gürültülü tahminlere yol açabilirken, çok yüksek bir 'k' değeri, veri setinin genel özelliklerini yansıtmayan tahminlere neden olabilir.

KNNImputer kullanımının bir diğer önemli noktası, hesaplama maliyetidir. Özellikle büyük veri setlerinde ve yüksek 'k' değerleri için, KNNImputer oldukça zaman alıcı olabilir. Bu nedenle, veri setinin boyutu ve eksik veri oranı dikkate alınarak, KNNImputer'ın kullanımı planlanmalıdır. Ayrıca, bu yöntemin uygulanabilirliği ve etkinliği, veri setinin özelliklerine ve eksik verilerin doğasına bağlı olarak değişkenlik gösterebilir. Son olarak, KNNImputer sonrasında, veri setinin yeniden analizi ve eksik veri doldurma işleminin etkilerinin değerlendirilmesi önemlidir. KNNImputer, eksik veri doldurma sürecinde dikkatli bir şekilde ele alınması gereken ve doğru uygulandığında veri setinin bütünlüğünü ve analiz sonuçlarının güvenilirliğini artıracak bir yöntemdir.

3.1.2. Veri Doldurmanın Zorlukları ve Dikkat Edilmesi Gereken Hususlar

Veri doldurma süreci, bazı zorlukları ve dikkat edilmesi gereken hususları içerir. İlk olarak, veri doldurma yöntemi seçimi, veri setinin yapısı ve eksik verilerin doğası göz önünde bulundurularak yapılmalıdır. Yanlış bir veri doldurma yönteminin seçilmesi, analiz sonuçlarını ve modelin doğruluğunu olumsuz etkileyebilir. İkinci olarak, veri doldurma sürecinde oluşabilecek yapay eğilimler ve varyansların azalması, analiz sonuçlarını çarpıtabilir. Bu nedenle, doldurma işlemi sonrasında veri setinin yeniden analiz edilmesi ve doldurma etkilerinin değerlendirilmesi önemlidir. Son olarak, veri doldurma işleminin uygulanabilirliği ve etkinliği, veri setinin boyutu ve eksik veri oranına bağlı olarak değişebilir. Yüksek oranda eksik verilere sahip veri setlerinde, veri doldurma yöntemlerinin etkinliği sınırlı olabilir ve bu durumda eksik verilerin doldurulması yerine alternatif yöntemlerin (örneğin, veri kümesinin yeniden örneklenmesi) değerlendirilmesi gerekebilir. Veri doldurma, veri analizi ve makine öğrenmesi projelerinde dikkatli bir şekilde ele alınması gereken bir süreçtir ve bu sürecin doğru bir şekilde yönetilmesi, projenin başarısını önemli ölçüde artırabilir.

3.2. Keşifsel Veri Analizi (Exploratory Data Analysis)

Keşifsel Veri Analizi (EDA), veri setlerinin temel yapısını ve içerdiği bilgileri anlamak için yapılan ilk ve en önemli adımlardan biridir. Bu aşama, verilerin ham formundan başlayarak derinlemesine bir keşif ve analiz sürecini kapsar. EDA'nın amacı, veri setindeki temel eğilimleri, varyansları, desenleri ve potansiyel sorunları (örneğin, aykırı değerler veya eksik veriler) tespit etmektir. Bu süreç, daha sonra uygulanacak olan makine öğrenmesi modellerinin geliştirilmesine sağlam bir zemin hazırlar ve modelleme sürecindeki potansiyel zorlukların önlenmesine yardımcı olur.

EDA, veri setinin kapsamlı bir şekilde incelenmesini sağlayarak, analistlere veri hakkında değerli içgörüler sunar. Bu içgörüler, modelin daha doğru ve etkili bir şekilde geliştirilmesi için gereklidir. Görselleştirme teknikleri ve istatistiksel özetler, EDA sürecinin temel araçlarıdır. Bu araçlar, verinin karmaşıklığını basitleştirerek, analistlerin veri setini daha kolay anlamasını ve yorumlamasını sağlar.

EDA, veri setinin anlaşılmasını ve modelleme sürecinin başarısını doğrudan etkileyen bir süreçtir. Bu nedenle, EDA'nın doğru ve etkili bir şekilde yapılması, makine öğrenmesi projelerinin başarısında kritik bir role sahiptir. Veri setinin doğru bir şekilde anlaşılması, modelin veri setinin temsil ettiği gerçek dünya durumlarına daha iyi uyum sağlamasına ve daha doğru tahminler yapmasına olanak tanır.

3.2.1. EDA'nın Makine Öğrenmesindeki Amaçları

EDA'nın makine öğrenmesindeki temel amacı, veri setinin kapsamlı bir anlayışını sağlamak ve veriye dayalı modellerin geliştirilmesine yön vermektedir. Veri setinin temel karakteristiklerinin, dağılımlarının ve öznitelikler arasındaki ilişkilerin analizi, modelin performansını ve güvenilirliğini doğrudan etkileyebilir. EDA, veri setindeki gizli yapıları, desenleri ve anomalileri keşfetmeye yardımcı olur. Bu süreç, veri kalitesi problemlerini (eksik veya hatalı veriler, aykırı değerler) erken aşamada tespit etmeye ve bu problemlerin model üzerindeki etkilerini azaltmaya yardımcı olur. Ayrıca, öznitelik seçimi ve dönüşümü için de temel teşkil eder; böylece makine öğrenmesi modellerinin daha doğru ve etkili olmasını sağlar.

3.2.2. EDA Teknikleri ve Uygulamaları

EDA teknikleri, veri setinin görselleştirilmesi ve istatistiksel analizlerini içerir. Görselleştirme araçları (örneğin, histogramlar, kutu grafikleri, saçılım grafikleri), verinin dağılımını ve öznitelikler arası ilişkileri görsel olarak ifade eder, böylece karmaşık veri setleri daha anlaşılır hale gelir. İstatistiksel özetler, verinin merkezi eğilimini ve dağılımını belirlemek için kullanılır; bu, modelin hangi özelliklere odaklanması gerektiğini belirlemeye yardımcı olur. Çok boyutlu analiz teknikleri, özellikle yüksek boyutlu veri setlerinde, verinin karmaşıklığını azaltarak analizini kolaylaştırır. Bu teknikler, veri setinin daha iyi anlaşılmasını sağlar ve makine öğrenmesi modelleme sürecinin etkinliğini artırır. EDA, veri setinin temel özelliklerini ortaya çıkarmak ve makine öğrenmesi modellerinin daha doğru ve güvenilir olmasını sağlamak için kritik bir rol oynar.

3.2.3. EDA'nın Zorlukları ve Dikkat Edilmesi Gereken Hususlar

Keşifsel Veri Analizi (EDA), veri bilimi ve makine öğrenmesi projelerinde temel bir aşama olmakla birlikte, bazı zorlukları ve dikkat edilmesi gereken önemli noktaları içerir. İlk olarak, EDA süreci genellikle zaman alıcıdır ve büyük veri setleri ile çalışırken bu süreç daha da karmaşık hale gelebilir. Veri setinin boyutu ve karmaşıklığı arttıkça, uygun analiz yöntemlerini ve görselleştirme tekniklerini seçmek zorlaşabilir. İkinci olarak, EDA sürecinde subjektif yorumlar ve yanlılık riski bulunmaktadır. Analistin önyargıları ve tecrübesi, analiz sonuçlarını etkileyebilir. Bu nedenle, veri analizinde objektiflik sağlamak ve çok yönlü bir bakış açısı benimsemek önemlidir. Üçüncü olarak, aykırı değerlerin ve eksik verilerin yönetimi, EDA'nın önemli bir parçasıdır. Aykırı değerlerin ve eksik verilerin yanlış yönetimi, analiz sonuçlarını ve sonraki modelleme aşamalarını olumsuz etkileyebilir. Bu durumlar için dikkatli ve bilinçli bir yaklaşım gerekir. Son olarak, EDA sırasında elde edilen bulguların yorumlanması

ve bu bulguların model seçimi veya hipotez geliştirme sürecine nasıl entegre edileceği, kritik öneme sahiptir. EDA'nın sonuçları, modelin yapısını ve parametrelerini belirlemede etkili olabilir; bu nedenle, elde edilen içgörülerin doğru bir şekilde yorumlanması ve kullanılması gerekmektedir. EDA sürecinin bu zorlukları, veri bilimcilerin ve makine öğrenmesi uzmanlarının dikkatli bir şekilde ele alması gereken konular arasındadır. Bu zorlukların üstesinden gelmek, projenin başarısını ve modelin doğruluğunu önemli ölçüde artırabilir.

3.3. Özellik Seçimi ve Makine Öğrenmesi Modellerinin Kapsamlı Değerlendirilmesi

Makine öğrenmesi projelerinde özellik seçimi, modelin performansını, genellemesini ve verimliliğini önemli ölçüde etkileyen bir süreçtir. Özellik seçimi, veri setindeki en bilgilendirici ve önemli özellikleri (features) belirlemeye yöneliktir. Bu süreç, özellikle yüksek boyutlu veri setlerinde, modelin karmaşıklığını azaltır ve eğitim süresini kısaltır. Özellik seçimi, hem gereksiz bilgilerin çıkarılmasını hem de modelin temelde odaklanması gereken anahtar bilgilerin belirlenmesini içerir.

Özellik seçiminin başarılı bir şekilde gerçekleştirilmesi, modelin doğruluğunu artırabilir. Doğru özelliklerin seçilmesi, modelin gereksiz gürültüden etkilenmesini azaltır ve daha temiz, daha anlamlı veri setleri üzerinde eğitilmesini sağlar. Bu, özellikle karmaşık ve çok boyutlu veri setlerinde önemlidir, çünkü bu tür veri setleri genellikle fazla sayıda gereksiz veya az bilgilendirici özellik içerebilir.

Ayrıca, özellik seçimi, modelin genellemesini iyileştirmek için de kritik öneme sahiptir. Modelin, eğitim veri setinde gördüğü özellikler üzerinden genellemeler yapabilmesi için, bu özelliklerin veri setinin temel karakteristiklerini yansıtmaları gerekir. Özellik seçimi, modelin yalnızca eğitim veri setindeki desenlere değil, genel veri yapısına odaklanmasını sağlar.

Özellik seçimi süreci, modelin anlaşılabilir ve yorumlanabilir olmasına katkı sağlar. Daha az sayıda ancak daha etkili özelliklerle çalışmak, modelin karar verme süreçlerini izlemeyi ve anlamayı kolaylaştırır. Bu, özellikle karmaşık modellerde, modelin kararlarının arkasındaki mantığı anlamak ve açıklamak için önemlidir.

3.3.1. SelectKBest Yöntemi

SelectKBest, özellik seçimi için kullanılan popüler bir yöntemdir. Bu yöntem, her bir özelliğin hedef değişkenle olan ilişkisini ölçerek, en bilgilendirici 'k' sayıda özelliği seçer. SelectKBest, özelliklerin hedef değişkenle ilişkisini değerlendirmek için çeşitli istatistiksel testler

kullanabilir. Örneğin, sınıflandırma problemleri için chi-kare testi veya ANOVA F-testi, sürekli hedef değişkenler için ise korelasyon katsayısı gibi testler tercih edilebilir.

SelectKBest yöntemi, özellikle karmaşık ve yüksek boyutlu veri setlerinde, makine öğrenmesi modellerinin performansını iyileştirmede kritik bir rol oynar. Bu yöntem, veri setindeki en bilgilendirici özellikleri (features) seçerek, modelin eğitim sürecine odaklanmasını sağlar. Seçilen özellikler, modelin doğruluğunu ve genellemesini artırmak için temel bir unsur haline gelir.

Bu süreçte, belirlenen bir sayıda en yüksek puan alan özellikler seçilir. Bu özelliklerin seçimi, her bir özelliğin hedef değişkenle olan ilişkisinin istatistiksel bir değerlendirilmesiyle gerçekleşir.

Bu yöntemin etkisini değerlendirmek için, seçilen özelliklerle çeşitli makine öğrenmesi modellerinin performansı değerlendirilebilir. Bu, çapraz doğrulama yöntemiyle gerçekleştirilebilir. Çapraz doğrulama, modelin farklı veri alt kümeleri üzerindeki performansını test eder ve böylece modelin genel performansına dair güvenilir bir değerlendirme sağlar.

Her bir modelin performansı, doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1 skoru gibi metrikler kullanılarak ölçülür. Bu metrikler, modelin sınıflandırma yeteneğinin farklı yönlerini değerlendirir ve modelin gerçek dünya verileri üzerindeki etkinliğine dair kapsamlı bir bakış sunar. Örneğin, kesinlik ve duyarlılık, modelin pozitif sınıflamalarının ne kadar doğru olduğunu ve tüm pozitif durumları ne kadar iyi tespit ettiğini gösterirken, F1 skoru bu iki metriğin dengeli bir birleşimidir. Bu performans metriklerini hesaplamada kullanılan parametreler Tablo 1.1’de ve hesaplama formülleri devamında verilmiştir.

Tablo 1.1. Performans metriklerini hesaplama parametreleri

Doğru Pozitif (TP) Pozitif değer Model tarafından pozitif tahmin edilmesi	Yanlış Pozitif (FP) Negatif değer model tarafından pozitif tahmin edilmesi
Yanlış Negatif (FN) Pozitif değer Model tarafından negatif tahmin edilmesi	Doğru Negatif (TN) Negatif değer Model tarafından negatif tahmin edilmesi

Doğruluk (Accuracy): Sistemde doğru yapılan tahminlerin tüm tahminlere oranı olarak hesaplanır. Bu karşılaştırma kriteri, özellikle dengesiz veri setlerinde tek başına güvenilir değildir.

$$Doğruluk = \frac{(TP + TN)}{(TP + TN + FN + FP)} \quad (1)$$

Duyarlılık (Recall): Pozitif olarak tahmin edilmesi gereken işlemlerin ne kadarının pozitif tahmin edildiğini gösteren ölçümdür.

$$Duyarlılık = \frac{TP}{(TP + FN)} \quad (2)$$

Kesinlik (Precision): Pozitif olarak tahmin edilen bir durumdaki başarıyı göstermektedir.

$$Kesinlik = \frac{TP}{(TP + FP)} \quad (3)$$

F1-Skor: Recall ve Precision'ın harmonik ortalamasıdır ve her iki metriği dengeli bir şekilde değerlendirir.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

Bu yöntemle yapılan özellik seçiminin, farklı modellerin performansına olan etkisi, özellikle karmaşık veya dengesiz veri setlerinde hangi özelliklerin modelin tahmin kabiliyetine en çok katkıda bulunduğunu belirlemek için önemlidir. Bu süreç, modelin doğruluğunu ve güvenilirliğini artırarak, veri setinin doğru bir şekilde kullanılmasını ve yorumlanmasını sağlar.

3.3.2. K-Means Kümeleme Yöntemi

Kümeleme, benzer özellik gösteren verilerin bir araya getirilerek gruplara ayrılması olarak tanımlanabilir. Bu analiz türünde temel hedef, gruplar arasında heterojenlik ve gruplar içinde homojenlik sağlamaktır. Benzer bireylerin aynı kümeye toplanması, küme içi homojenliği destekleyen bir yöntemdir. Kümeleme problemi bir optimizasyon sorunudur ve küme elemanlarının küme ortalamasından uzaklıklarının toplamını en aza indirgeyerek optimal kümeleme sağlanır. Bireyler arasındaki benzerlik, uzaydaki konumlarına bağlıdır. Uzaydaki konumları itibarıyla birbirine daha yakın olan bireyler aynı kümede gruplanır.

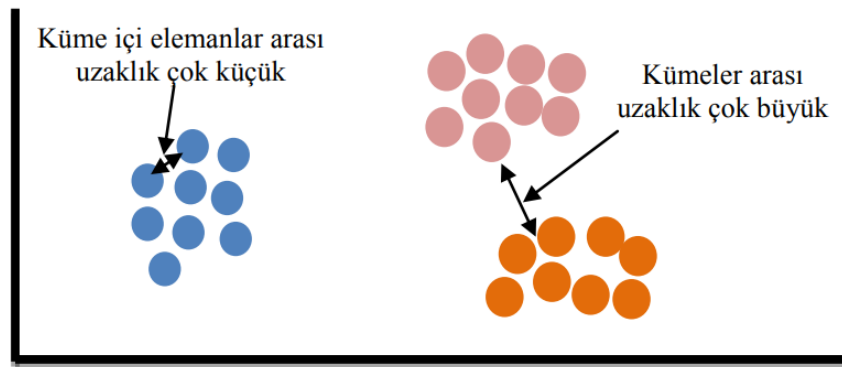
Literatürde çeşitli kümeleme algoritmaları bulunmaktadır ve kullanılacak algoritmanın seçimi, belirli bir amaç ve veri tipine bağlıdır. Bölme yöntemleri, hiyerarşik yöntemler,

yoğunluk tabanlı yöntemler, ızgara tabanlı yöntemler ve model tabanlı yöntemler gibi başlıca kümeleme yöntemleri genel olarak sınıflandırılabilir. Her bir yöntem, belirli bir problem veya veri yapısı için daha uygun olabilir, bu nedenle kümeleme analizi yapılırken dikkatlice seçilmelidir (Özekes S, 2003).

Kümeleme işleminde, birçok durumda küme içindeki nesneler arasındaki uzaklık oldukça küçükken, kümeler arası uzaklık ise önemli ölçüde büyük olmaktadır (Ekim U, vd., 2011). Bu bağlamda, K-means algoritması, gözetimsiz öğrenme yöntemleri arasında yaygın olarak kullanılan ve keskin bir kümeleme sağlayan bir algoritmadır (Sarıman G, 2011).

K-means algoritması, n adet veri nesnesinden oluşan bir veri setini, belirli bir k sayısı kadar kümeye bölme mantığına dayanır. Temel amacı, gerçekleştirilen bölümlenme işlemi sonucunda elde edilen kümelerin içindeki nesneler arasındaki benzerliğin maksimum, kümeler arası benzerliğin ise minimum olmasını sağlamaktır. Bu algoritma, sürekli olarak kümelerin güncellendiği ve en uygun çözüme ulaşana kadar devam eden döngüsel bir yapıyla öne çıkar (Demiralay M ve Çamurcu A.Y., 2005, Silahtaroglu G., 2008).

K-means algoritması, veri setini belirli bir sayıda kümeye bölerek, küme içi homojenliği ve kümeler arası heterojenliği optimize etmeye çalışır. Bu süreç, küme içi elemanlar arasındaki uzaklığı minimize ederek ve kümeler arası uzaklığı maksimize ederek gerçekleşir. Çalışma prensibi, veri setindeki nesnelerin iteratif olarak kümelenmesi ve kümelerin ortalamalarının güncellenmesi üzerine kuruludur. Bu şekilde, K-means algoritması sürekli olarak iyileştirilen kümelerin elde edilmesini sağlar (Çalışkan S, Soğukpınar İ, 2008).



Şekil 1.2. Küme Yapısı

Kümeleme işlemi sırasında, genellikle küme içindeki nesneler arasındaki uzaklık oldukça küçükken, kümeler arasındaki uzaklık ise belirgin bir şekilde büyük olabilir (Yavuz Ü, vd.,

2011). Bu durumu düzeltmek ve net bir kümeleme elde etmek için kullanılan en popüler gözetimsiz öğrenme yöntemlerinden biri, K-means algoritmasıdır (Sarıman G, 2011).

K-means algoritması, her veriyi sadece bir kümeye ait olabileceği keskin bir şekilde belirleyen bir kümeleme algoritmasıdır. Bu algoritma, veri noktalarını belirli sayıda kümeye bölerek, her veri noktasının hangi kümeye ait olduğunu belirler. Her veri noktası, kümesinin merkezine olan uzaklığına göre en yakın kümeye atanır. Bu şekilde, benzer özelliklere sahip olan veri noktaları aynı kümeye ait olur ve bu, küme içindeki uzaklığı minimize ederken, kümeler arasındaki uzaklığı maksimize etmeye çalışır. K-means algoritması, basit ancak etkili bir kümeleme yöntemi olarak bilinir ve geniş bir uygulama alanına sahiptir (Sarıman G, 2011).

Son yıllarda, küme sayısının belirlenmesi konusunda yoğun çabalar harcanmıştır. Küme sayısının belirlenmesinde en pratik yöntem, genellikle (Koltan Ş, Patır S., 2011) numaralı denklemle ifade edilir (Doğan İ, 2022). Ancak, veri sayısı çok büyük olduğunda, bu yöntem pratik olmayabilir.

K-means algoritmasının işleyiş adımları aşağıdaki gibidir:

1. Küme Merkezlerinin Belirlenmesi:

- Başlangıçta, kümeleme için gerekli olan k sayıda merkez nokta iki şekilde belirlenir:
- **Rastgele Yöntem:** Rasgele seçilmiş k adet nesne, başlangıç merkezleri olarak kullanılır.
- **Ortalama Yöntem:** Tüm nesnelerin ortalaması alınarak merkez noktalar hesaplanır.

2. Nesnelerin Kümelenmesi:

- Her bir nesne, belirlenen merkez noktalara olan uzaklığına göre incelenir.
- Her nesne, kendisine en yakın olan merkez noktaya sahip kümeye atanır.

3. Küme Merkezlerinin Güncellenmesi:

- Her bir kümenin merkez noktası, o kümede yer alan nesnelerin ortalama konumuyla güncellenir.

4. İterasyon ve İstikrarın Sağlanması:

- 2. ve 3. adımlar, küme merkezleri değişiklik göstermediği ve sabit kaldığı noktaya kadar tekrar edilir.
- Bu süreç, kümelerin merkezleri istikrarlı bir duruma gelene kadar devam eder.

Bu dört adım, K-means algoritmasının temelini oluşturur ve veri setindeki nesnelerin etkili bir şekilde kümelenmesini sağlar. Algoritma, merkez noktalar sabitlenene ve kümeler arasındaki toplam varyans minimuma inene kadar çalışmaya devam eder.

3.3.3 Kruskal-Wallis Testi

Kruskal-Wallis testi, non-parametrik bir istatistiksel yöntem olarak tanımlanır ve özellikle, normal dağılım varsayımını karşılamayan veriler üzerinde kullanılır. Bu test, farklı grupların medyanlarını karşılaştırmak için etkili bir yöntemdir ve genellikle üç veya daha fazla bağımsız grubun bağımlı değişken üzerindeki etkilerini değerlendirmek için tercih edilir (Daniel, 1990).

Testin uygulanmasında, öncelikle tüm veri noktaları sıralanır ve bu sıralamaya dayalı olarak ranklar atanır. Daha sonra, her gruptaki rankların toplamı hesaplanır ve bu toplamlar üzerinden bir test istatistiği hesaplanır. Kruskal-Wallis testinin H istatistiği, gruplar arasındaki farklılığın büyüklüğünü ölçer ve bu istatistik, belirli bir dağılımı takip eder. Bu istatistik ve ilgili p-değer, gruplar arasında anlamlı bir farklılık olup olmadığını belirlemek için kullanılır.

3.3.4 Post-hoc (Dunn's) Testi

Kruskal-Wallis testi sonrasında anlamlı farklılıklar tespit edildiğinde, hangi grupların arasında bu farkların olduğunu belirlemek için Post-hoc (Dunn's) testi kullanılır. Bu test, çoklu karşılaştırmalar için uygundur ve her bir grup çifti arasındaki farkları ayrı ayrı değerlendirir (Dunn, 1964:241).

Dunn's testi, her bir grup çifti için ayrı ayrı p-değerleri hesaplar ve bu p-değerleri, gruplar arasındaki anlamlı farklılıkların belirlenmesinde kullanılır. Çoklu karşılaştırmalarda hata oranını kontrol etmek için genellikle Bonferroni düzeltmesi gibi yöntemler kullanılır. Bu test, Kruskal-Wallis testinin bulgularını daha detaylı bir şekilde analiz etmekte ve belirli gruplar arasındaki farklılıkları belirlemekte önemli bir rol oynar (Zar, 2005).

Bu yöntemler, istatistiksel analizlerde önemli bir yer tutmakta ve araştırmacılara farklı gruplar arasındaki ilişkileri daha detaylı bir şekilde inceleme olanağı sağlamaktadır.

3.4. Lojistik Regresyon

Lojistik regresyon analizinde (LRA), logit dönüşümüne tabi tutulan bağımlı (yordayıcı) değişkenin yapısı, analizin sınıflandırılmasında belirleyici bir rol oynamaktadır. Bu analiz, bağımlı değişkenin özelliklerine göre üç farklı modele ayrılır. İlk olarak, 'İkili Lojistik

Regresyon Modeli', kategorik bağımlı değişkenin ikili durumlarını (örneğin cinsiyet; kadın-erkek) değerlendirmek için kullanılır. 'Multinomial Lojistik Regresyon Modeli', çok kategorili bağımlı değişkenlerin (örneğin medeni durum; evli-bekar-boşanmış) analizinde tercih edilirken; çok kategorili ve sıralı bir yapı söz konusu ise (örneğin Likert tipi ölçekler, az-orta-çok), 'Sıralı (Ordinal) Lojistik Regresyon Modeli' uygulanır (Barak vd, 2005).

LRA'nın bağımsız (yordanan) değişken sayısına göre sınıflandırıldığı diğer bir durum da mevcuttur. Bu sınıflandırmaya göre, LRA, tek bağımsız değişken içerdiğinde 'Tek Değişkenli Lojistik Regresyon', iki veya daha fazla bağımsız değişken içerdiğinde ise 'Çoklu Lojistik Regresyon Analizi' olarak adlandırılır (Cook vd, 2001).

3.5. Random Forest

Random Forest (RF) yöntemi, geleneksel sınıflandırma ve regresyon ağaçlarından farklı olarak istenilen sayıda ağaç oluşturabilme ve bu ağaçları her bir düğümdeki kestirim değerleri temel alarak genelleştirme yeteneği sunma özelliğine sahiptir. (Süel, 2014).

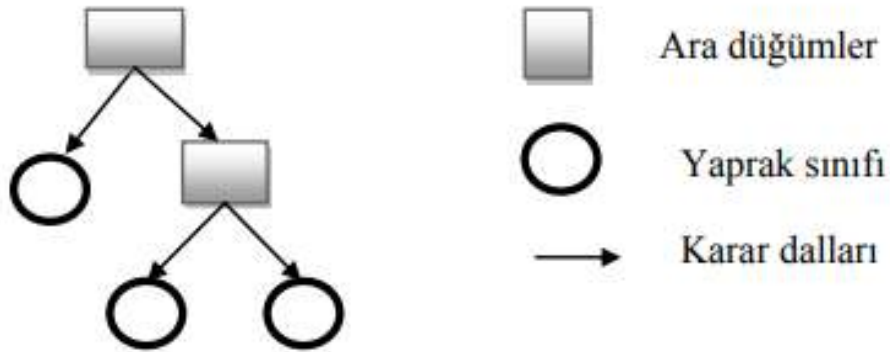
Farklı veri grupları üzerinde sınıflandırıcıların performansındaki değişkenlik, bazen tek bir sınıflandırıcının sonuçlarının yeterince etkili olmadığı durumları ortaya çıkarabilir. Bu durumda, başarı oranını ve doğruluğunu artırmak amacıyla sınıflandırıcılar birleştirilerek oylama yöntemi kullanılır. Oylama, sayısal tahminleri veya olasılık değerlerini kullanarak sınıflandırıcıları birleştirme stratejisidir. Her sınıf için olasılık değerleri veya sayısal tahminler toplanır ve en yüksek toplam değere sahip olan sınıf kazanan olarak belirlenir (Coşgun, 2009).

Random Forest (RF), Leo Breiman ve Adele Cutler tarafından geliştirilen, içinde oylama yöntemini barındıran bir sınıflama algoritmasıdır. RF, birçok karar ağacının birleşmesiyle oluşur ve bireysel ağaçlar tarafından oylanarak kazanan sınıf belirlenir. Karar ağaçları bağımsızdır ve veri setinden bootstrap tekniği ile çekilen örneklerden oluşturulur. "Bagging" yönteminden farklı olarak, RF'ta dallara ayırıcı özellikteki değişkenin tüm değişkenler arasından rastgele seçilen m tane değişken içinden belirlenmesi öne çıkar. Her ağaç için m sayısı sabittir ve genellikle $(p, \text{değişken sayısını ifade eder})$ olarak alınır.

3.5.1. Ağacın Yapısı:

Ağaç, dallar ve yapraklardan oluşur. Her bir nitelik bir düğüm tarafından temsil edilir. En son yapı "yaprak" (terminal), en üst yapı "kök" ve bunların arasındaki yapılar "dal" olarak adlandırılır. RF yönteminde, ağaç bütün verinin oluşturduğu tek bir düğümle başlar. Eğer örneklerin tamamı aynı sınıfa aitse, düğüm yaprak olarak sonlanır ve sınıf etiketi verilir. Eğer

örnekler aynı sınıfa ait değilse, en iyi bölme özelliklerine göre sınıflandırma yapılır (Breiman, 2001).



Şekil 1.3. Random Forest için oluşturan tek ağaçların yapısı

3.6. Destek Vektör Makineleri (Support Vector Machine)

Destek Vektör Makineleri (SVM), Vapnik tarafından geliştirilen bir öğrenme yöntemidir ve istatistiksel öğrenme teorisi ile yapısal riski en aza indirme ilkesine dayanmaktadır (Vapnik 1995, 1998). Bu yöntem, sınıflandırma ve regresyon problemlerini çözmeyi amaçlamaktadır. SVM, herhangi bir sınıflandırma veya regresyon problemiyle karşılaştığında, bu problemi yerel çözümlere takılmadan çözebilmek için bir karesel programlama problemine dönüştürür. Yerel çözümlere takılmama yeteneği, SVM'nin diğer tekniklere göre avantajlı olmasını sağlar. Ayrıca, SVM yüksek genelleme yeteneğine sahiptir ve çeşitli alanlarda uygulanmıştır, örneğin el yazısı tanıma, yüz tanıma, 3-boyutlu nesne tanıma, ses tanıma, konuşmacı tanıma, metin sınıflandırma gibi (Cristianini vd. 2000, Schölkopf vd. 1999).

SVM sınıflandırıcıları, optimal bir ayırıcı aşırıdüzlem oluşturarak margin'i (aralığı) maksimum yapan bir prensibe dayanır. Bu bağlamda, margin kavramı, ayırıcı aşırıdüzlemden en yakın veri noktasına olan minimum uzaklığı ifade eder.

Destek Vektör Makinesi sınıflandırıcılarında iki temel durumla karşılaşılabılır: verilerin doğrusal olarak ayrılabilir olması ve doğrusal olarak ayrılamayan bir yapıya sahip olması. Ancak, gerçek yaşam problemleri genellikle birçok bileşeni içeren karmaşık yapılar gösterir, bu nedenle doğrusal olarak ayrılamayan verilerin sınıflandırılması için bu veriler önce farklı bir

uzaya aktarılmalıdır. SVM sınıflandırıcıları, doğrusal SVM sınıflandırıcıları ve doğrusal olmayan SVM sınıflandırıcıları olarak iki temel kategoride incelenebilir.

3.7. En Yakın Komşuluk (K-Nearest Neighbors)

K-NN algoritması, T. M. Cover ve P. E. Hart tarafından önerilen bir sınıflandırma yöntemidir, bu yöntemde örnek veri noktasının bulunduğu sınıf ve en yakın komşunun belirlenmesi k değerine göre gerçekleştirilir (Cover vd, 1967). Bu algoritma, eski, basit ve etkili örüntü sınıflandırma yöntemlerinden biridir ve makine öğrenme alanında popüler olarak kullanılmaktadır. Nesnelerin sınıflandırılması, birçok farklı alanda, örneğin örüntü tanıma, veri madenciliği, yapay zeka, istatistik, bilişsel psikoloji, tıp, biyoinformatik gibi, önemli bir araştırma konusudur (Keller vd., 1985).

K-NN algoritması, eğitim gerektirmemesi, kolay gerçekleştirilebilir olması, analitik olarak izlenebilir olması, yerel bilgilere adapte edilebilmesi, paralel gerçekleştirime uygun olması ve gürültülü eğitim verilerine karşı dirençli olması gibi avantajlara sahiptir (Qiu vd, 2008). Ancak, yüksek bellek alanı gereksinimi, veri seti ve öznitelik boyutu arttıkça işlem yükünün ve maliyetin önemli ölçüde artması gibi dezavantajları da vardır. Ayrıca, performansı, k komşu sayısı, uzaklık ölçütü ve öznitelik sayısı gibi parametreler ve özelliklerle etkileşimli olarak değişebilir (Liu vd, 2012).

3.8. Gradient Boost

Gradient Boosting Machine (GBM), zayıf öğrenicileri güçlü öğrenicilere dönüştürme prensibi üzerine çalışan bir yöntemdir. GBM'nin çalışma mantığı genel olarak dört adımla açıklanır. İlk adımda, Denklem 5'te belirtilen gibi model katsayılarının veriye ne kadar iyi uydurulduğunu değerlendirerek kayıp fonksiyonu tanımlanır:

$$L(y_i, F(x)) \quad (5)$$

Bu noktada, y_i gözlemlenen değeri ve $F(x)$ ise tahmin edilen değeri ifade eder. İkinci adımda, sabit değişken belirlenir ve tüm gözlemler için kayıp fonksiyonunun değerinin minimum olduğu durumu bulmak amaçlanır (Denklem 6).

$$F_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum L(y_i, \gamma) \quad (6)$$

Bu formülde, $F_0(x)$, kayıp fonksiyonunu minimize edecek ilk tahmin ediciyi (γ) belirlemeyi amaçlar. Burada γ , tüm gözlemler için kayıp fonksiyonunun değerini minimize edecek sabit değeri ifade eder. Bu sabit, kayıp fonksiyonunun minimum değerini bulmak için kullanılır ve

genellikle bir optimizasyon algoritması aracılığıyla hesaplanır. y_i , gözlemlenen gerçek değerleri temsil ederken, γ bu ilk tahmini temsil eder ve kayıp fonksiyonu bu iki değer arasındaki farka göre hesaplanır.

Üçüncü adımda, önceki tahminlerden kaynaklanan hatalar hesaplanır. Daha sonra, bu hatalar üzerinden karar ağaçları oluşturularak her ağacın yapısı için değerler hesaplanır.

Son adımda, her bir gözlem için tahmin değerleri üretilir. (Friedman, 2001).

3.9. Karar Ağaçları

Karar ağacı yöntemi, sınıflandırma ve tahmin etme süreçlerinde kullanılan önemli veri madenciliği tekniklerinden biridir. Karar ağacı, girdisi olmayan bir kök düğüm ve her biri birer girdi alan iç düğümlerden oluşan yönlü bir ağaç yapısına sahiptir. Ağaçtaki çıktıları, bir başka düğüm tarafından girdi olarak alınan iç düğümler veya test düğümleri ile, girdi olmayan düğümler veya yaprak düğümleri olarak adlandırılmaktadır. Her iç düğüm, örnek uzayını öznitelik değerlerinin belirli bir fonksiyona tabi tutulmasıyla iki veya daha fazla parçaya ayırmaktadır. İç düğümler, öznitelikler üzerinde gerçekleştirilen testleri, dalların test sonuçlarını ve her bir yaprak düğümün sınıf etiketini temsil etmektedir.

Karar ağaçları sınıflandırmada kullanıldığında, basit yapısı sayesinde oluşturulan modelin anlaşılabilir olması, parametrik olmaması nedeniyle bilgi keşfi için uygun bir yapı sunması ve diğer sınıflandırma yöntemlerine kıyasla daha hızlı bir şekilde oluşturulması gibi avantajlara sahiptir. Ayrıca, karar ağaçlarından elde edilen kuralların elde edilmesi de oldukça kolaydır. Karar ağaçları, kategorik ve nümerik verilerin sınıflandırılmasında kullanılabilir. Ancak, bu üstün özelliklere rağmen, karar ağaçları bazı zorluklarla karşılaşmaktadır. Bu zorluklar arasında birden fazla öznitelik içeren çıktıları mümkün kılamamaları, kısmen değişken sonuçlar üretmeleri, test verisindeki küçük değişikliklere karşı duyarlı olmaları ve nümerik veri setleri için karmaşık bir ağaç yapısı oluşturmaları yer almaktadır.

Karar ağaçları, eğitim verisinden bir model oluşturulması, bu modelin test verisi kullanılarak değerlendirilmesi ve ilgili modelin gelecekteki değerlerin tahmin edilmesi gibi aşamalardan geçmektedir. Veri setlerinden otomatik olarak karar ağacı yapısını oluşturmak amacıyla birçok karar ağacı algoritması geliştirilmiştir. Bu algoritmalar genellikle genelleştirme hatasını minimize eden en uygun karar ağacı yapısını oluşturmayı hedefler, ancak düğüm sayısı, ortalama derinlik veya başka amaç fonksiyonlarını minimize etmeyi de amaçlayabilirler. Karar ağacı algoritmalarının genellikle küçük boyutlu ve az derinlikli ağaçlar

oluşturması hedeflenir. Büyük ve karmaşık ağaçlar, düşük genelleştirme başarımına sahip olabilir. Bu nedenle, küçük boyutlu ağaçlar oluşturmak için birçok yaklaşım geliştirilmiştir.

Karar ağacı oluşturma sürecinde kullanılan yaklaşımlardan biri düğüm ayırma ölçütlerinin kullanılmasıdır. Bilgi kazancı, ki-kare istatistiği, GINI indeksi gibi ölçütler, başlıca kullanılan düğüm ayırma ölçütleri arasında yer almaktadır. Ayrıca, genelleştirme performansını artırmak amacıyla kullanılan budama yöntemleri de bulunmaktadır. Budama yöntemleri, ağacın düşük istatistiksel geçerliliğe sahip alt ağaçlarını ortadan kaldırarak, daha küçük boyutlu bir ağaç elde edilmesini sağlar. Bu şekilde genelleştirme doğruluk oranı iyileştirilir. Budama yöntemleri, düğümlerin yukarıdan aşağıya veya aşağıdan yukarıya doğru taranmasıyla gerçekleştirilir. Budama yöntemlerinden bazıları maliyet-karmaşıklık budama yöntemi, azaltılmış hata budamasıdır. Bu yöntemler, düğümleri belirli bir ölçüte göre budarlar ve böylece ağacın boyutunu kontrol ederler. Ancak, herhangi bir budama yönteminin her durumda en uygun sonucu vermediği gözlenmiştir.

4. BULGULAR

Bu çalışma, Ailevi Akdeniz Ateşi (AAA) tanısı konmuş hastalar üzerinde, Tel-Hashomer Tanı Kriterlerine göre hazırlanan veri seti üzerinden kapsamlı bir analiz gerçekleştirmiştir. Özellikle kolşisin tedavi yanıtlarını etkileyen kritik faktörlerin belirlenmesine odaklanılmıştır. Bu faktörlerin kolşisin tedavi yanıtları üzerindeki etkileri, Kruskal-Wallis istatistiksel testi kullanılarak değerlendirilmiştir ve bu değişkenlerin tedaviye verilen yanıtlar açısından anlamlı farklılıkları ortaya koyulmuştur.

4.1. Veri Seti

Araştırmanın amacı, AAA hastalığına sahip 155 bireyin klinik, demografik ve genetik özelliklerini derinlemesine incelemektir. Veri, Uludağ Üniversitesi Tıp Fakültesi Dahili Bilimler Bölümü, İç Hastalıkları Anabilim Dalı Romatoloji Bilim Dalı Başkanı Prof. Dr. Yavuz Pehlivan tarafından sağlanmıştır. Toplam 21 değişken içeren ilk veri seti, demografik, klinik ve genetik olmak üzere üç kategoriye ayrılmıştır. Bu veri seti, hem sayısal hem de kategorik değişkenlerin birleşiminden oluşmaktadır.

Ön analiz aşamasında, bazı değişkenlerde yüksek düzeyde eksik veri bulunması ve belirli değişkenler arasında önemli korelasyonlar tespit edilmiştir. Bu korelasyonlar, bazı değişkenlerin tekrarlayan bilgiler içerdiğini göstermektedir. Dolayısıyla, eksik veya yüksek korelasyonlu değişkenler çıkarılmış ve diğer değişkenlerin model üzerindeki etkisi daha açık hale getirilmiştir. Rafine edilen son veri seti, 3 demografik, 8 klinik ve 4 genetik özelliği içermekte olup, bu özelliklerin detaylı analizi ve dağılımları Tablo 1.2’de sunulmuştur.

Bu geniş kapsamlı veri seti analizi, AAA hastalığıyla ilgili daha derinlemesine anlayış sağlamak adına önemli bir adımdır. Araştırmanın ilerleyen aşamalarında, elde edilen bulguların AAA hastalığının teşhis, tedavi ve genetik temelleri üzerindeki etkilerini daha iyi anlamak için ileri düzey istatistiksel yöntemlerin uygulanması planlanmaktadır.

Tablo 4.2. Veri setindeki kategorik deęişkenlerin özellikleri, açıklamaları, mümkün deęerleri ve her deęerin görölme oranları

Sütun Adı	Açıklama	Deęer-Oranlar
cinsiyet	Hastanın Cinsiyeti	Erkek %55.56 / Kadın %44.44
dyer	Doęum Yeri	Marmara %59.03 / Karadeniz %40.97
ailefmf	Aile AAA Öyküsü	Var %61.81 / Yok %38.19
krnagrsatk	Karın Ağrısı Atak	Var %95.83 / Yok %4.17
klcadzykblkagrs	Kalça, Diz ve Ayak Bilek Ağrısı	Var %29.17 / Yok %70.83
ateş	Ateş	Var %85.42 / Yok %14.58
kolsisinynt	Kolşisin Yanıtı	Var %84.72 / Yok %15.28
ebe	Deri Lezyonu	Var %96.53 / Yok %3.47
atkaktfazyk	Atak Akut Faz Yükseklięi Sıklıęı	Var %86.11 / Yok %13.89
titprteinvar	TİT’de protein varlıęı	Var %16.67 / Yok %83.33
yygnmutvar	MEFV mutasyon varlıęı	Var %69.44 / Yok %30.56
MEFV	MEFV geninin mutasyonları	Homozigot %41.0 / Heterozigot %59.0

Bu çalışmada, hastaların demografik ve klinik özelliklerini deęerlendirmek amacıyla belirlenmiş olan çeşitli sütunlardan elde edilen verilere odaklanılmıştır. İlk olarak, hastaların cinsiyet dağılımına bakıldığında, hastaların %55.56'sının erkek ve %44.44'ünün kadın olduęu belirlenmiştir.

Doęum yeri (dyer) sütununda, hastaların %59.03'ünün Marmara bölgesinde, %40.97'sinin ise Karadeniz bölgesinde doğduęu saptanmıştır. Ailevi Akdeniz Ateşi (AAA) öyküsü (ailefmf) sütununda, hastaların %61.81'inin ailelerinde AAA öyküsü bulunduęu, %38.19'unun ise bu öyküye sahip olmadığı tespit edilmiştir.

Klinik belirtiler arasında yer alan karın ağrısı atakları (krnagrsatk) sütununda, hastaların %95.83'ünün bu ataklara maruz kaldığı, %4.17'sinin ise bu ataklardan etkilenmediği görülmüştür. Kalça, diz ve ayak bilek ağrısı (klcadzykblkagrs) sütununda ise hastaların %29.17'sinin bu tür ağrılara sahip olduğu, %70.83'ünün ise bu ağrılardan muzdarip olmadığı belirlenmiştir.

Ayrıca, hastalarda ateş (ateş) olup olmadığına dair sütunda %85.42'sinin ateş belirtileri gösterdiği, %14.58'inin ise ateşli olmadığı saptanmıştır. Kolşisin tedavisine yanıt (kolsisinynt) sütununda, hastaların %84.72'sinin bu tedaviye yanıt verdiği, %15.28'inin ise yanıt vermediği tespit edilmiştir.

Dermatolojik lezyonların varlığı (ebe) sütununda hastaların %96.53'ünde deri lezyonları bulunduğu, %3.47'sinde ise deri lezyonu olmadığı belirlenmiştir. Atak akut faz yüksekliği sıklığı (atkaktfazyk) sütununda, hastaların %86.11'inin atak akut faz yüksekliği yaşadığı, %13.89'unun ise bu durumdan etkilenmediği görülmüştür.

Genetik özelliklere odaklanıldığında, TİT (Tümör Nekroz Faktörü Receptor-associated Periodic Syndrome) hastalığında protein varlığı (titprteinvar) sütununda %16.67'sinin protein içerdiği, %83.33'ünün ise bu özelliğe sahip olmadığı belirlenmiştir. MEFV mutasyon varlığı (yygnmutvar) sütununda, hastaların %69.44'ünde MEFV mutasyonlarına rastlandığı, %30.56'sında ise bu mutasyonlara rastlanmadığı saptanmıştır.

MEFV geninin mutasyonlarına (MEFV) dair sütunda homozigot hastaların %41.0'ı ve heterozigot hastaların %59.0'ı belirlenmiştir. Bu geniş veri seti, hastalığın çeşitli yönlerini anlamak ve tedavi stratejilerini geliştirmek açısından önemli bir kaynak sağlamaktadır.

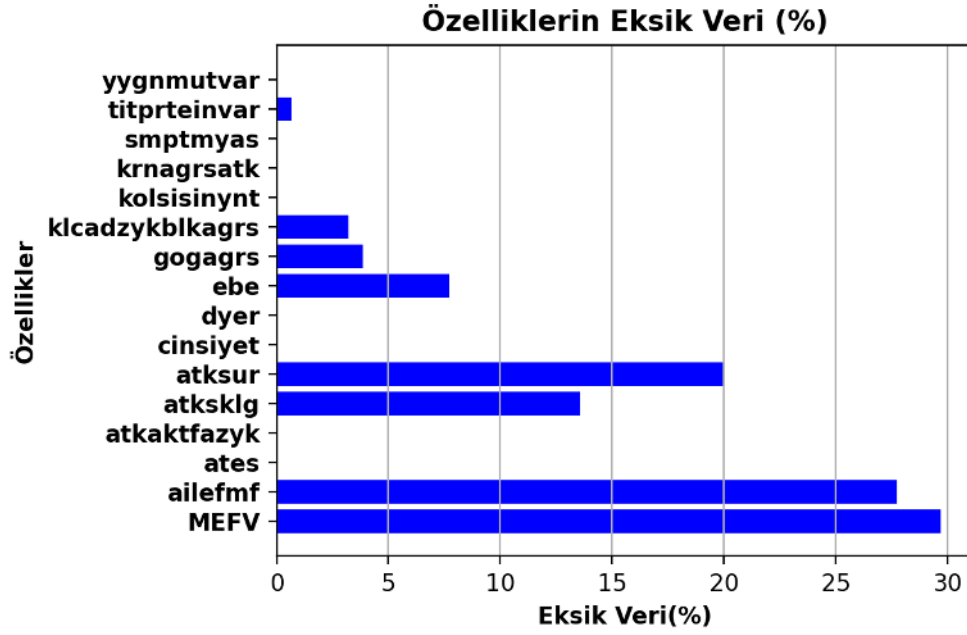
Tablo 1.3. Veri setindeki sayısal değişkenlere ait tanımlayıcı istatistikler

	smptmyas (yıl)	atksur (gün)	atksklg (gün)
Ortalama	18.92	2.94	61.28
Standart Sapma	10.44	1.21	42.43
Medyan	17.0	3.0	60.0
IQR	16.0	2.0	50.0
Min	2.0	1.0	1.0
Max	45.0	7.0	180.0
Shapiro-Wilk p-değeri	0.000451	0.0	0.0

* Shapiro-Wilk p-değeri sayısal verilerin normal dağılıma sahip olup olmadığını belirler.

4.2. Eksik Veriler

Proje için toplanan hastalarda verilerin bir kısmı eksik bir şekilde elimize ulaştı. Veri Setinde ki eksik verilerin kısmı yukarı da yer alan Şekil 2.1’de eksiklik yüzde değerleri ve isimleri yer almaktadır.



Şekil 2.1. Veri setindeki eksik verilerin yüzdelik değerleri

Görölmüş olduđu gibi özellikle genetik veriler MEFV oranları yüzde 30’a yakın oranda eksiklik içermektedir. Ailelerde yer alan fmf varlığı verisinin eksikliği de en çok eksik olan veridir. Bu verilerin eksik olmasının en temel nedeni hastalar tarafından verilen kişisel verileri hakkında bilgiye sahip olmamalarıdır.

4.3. Veri Ön İşleme Süreci ve Özelliklerin Seçimi

Bu çalışmada, veri hazırlama sürecinde, eksik veri yönetimi ve aykırı değerlerin ele alınması için bir dizi metodolojik adım uygulanmıştır. Bu süreçte, sayısal değişkenlerde eksik verilerin doldurulması için K-En Yakın Komşu İmputasyon (KNNImputer) tekniğı tercih edilmiştir. Bu yöntem, eksik veri yapılarının ve değişkenler arasındaki ilişkilerin bütünlüğünü muhafaza etmeyi hedeflemektedir.

Özellikle, 'atksur' ve 'atksklg' gibi değişkenlerde saptanan aykırı değerler, model performansı üzerinde olası olumsuz etkileri göz önünde bulundurularak veri setinden çıkarılmıştır. Bu yaklaşım, modelin genel veriye uyumunu artırma ve aşırı uyum riskini azaltma amacı gütmektedir.

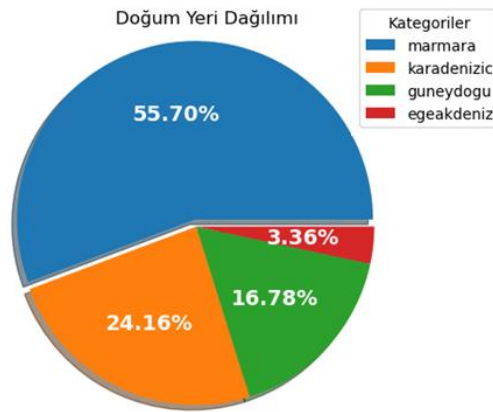
Kategorik değişkenlerin ilişkileri, çapraz tablo grafikleri kullanılarak detaylıca incelenmiştir. Bu inceleme sonucunda, ilişkili kategorik değişkenlerdeki eksik veriler, gruplandırma ve mod değerlerine dayanarak doldurulmuştur. Bu strateji, veri setinin eksikliklerinin, değişkenler arasındaki doğal ilişkileri bozmadan tamamlanmasını sağlamıştır.

Bu veri ön işleme süreci, analizde kullanılan veri setinin güvenilirliğini artırmayı ve modelin daha sağlam ve güvenilir sonuçlar üretmesini sağlamak üzere özenle tasarlanmıştır. Bu yöntemler, veri setinin içeriğini daha anlamlı ve kapsamlı hale getirerek analitik çalışmalara daha sağlıklı bir zemin sunmayı amaçlamaktadır.

4.4. Demografik ve Coğrafi Özelliklerin İncelenmesi

Bu çalışma kapsamında incelenen veri setinde bulunan "dyer" (doğum yeri) değişkeni, coğrafi bölgelere göre daha spesifik bir sınıflandırmaya tabi tutulmuştur. İlk başta şehir bazında tanımlanan bu değişken, daha geniş bir perspektif sunması amacıyla 'Güney Doğu ve Doğu Anadolu', 'Marmara', 'Karadeniz ve İç Anadolu', ve 'Ege ve Akdeniz' olmak üzere dört ana kategoriye ayrılarak yeniden düzenlenmiştir. Bu yeni kategorizasyon, hastaların kolşisin tedavisine olan yanıt oranları, klinik belirtiler ve genetik değişkenler açısından coğrafi dağılımlarının daha ayrıntılı bir şekilde incelenmesine olanak sağlamaktadır.

Şekil 3, coğrafi bölgelerin dağılımını daha görsel bir şekilde ifade etmek amacıyla kullanılan bir pasta grafiği üzerinden bu sınıflandırmanın nasıl yapıldığını göstermektedir. Yeniden yapılan bu coğrafi sınıflandırma, analizin daha derinlemesine ve bölgesel bir perspektifle yapılmasına imkan tanıyarak, hastaların belirli coğrafi bölgelerde nasıl farklılık gösterdiğini anlamamıza yardımcı olmaktadır. Bu şekilde elde edilen bilgiler, tedavi stratejilerinin ve genetik özelliklerin coğrafi faktörlere bağlı olarak nasıl değişebileceği konusunda daha kapsamlı bir anlayış sunmaktadır.

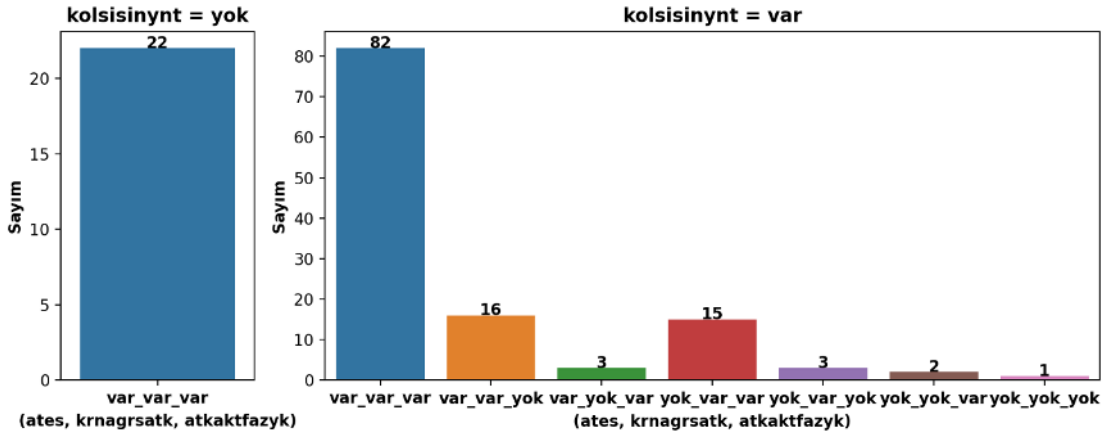


Şekil 2.2.. Coğrafi bölgelerin dağılımı

4.5. Kolşisin Yanıtının Demografik ve Klinik Gözlemlerle İlişkisi

Bu bölümde, eksik veri içermeyen toplam 144 hasta verisi üzerinde bir çalışma sunulmaktadır. Bu hastalar, belirli kriterlere uyan ve "ateş", "karın ağrısı atağı" ve "atak akut faz yüksekliği" semptomlarına sahip olanlar olarak seçilmiştir. Seçilen bu kriterlere uymayan

ve "kolşisin yanıtı" olmayan hasta sayısı 22'dir; bu da toplam hasta popülasyonunun yaklaşık %15.28'ini oluşturmaktadır.



Şekil 2.3. AAA hastalarında kolşisin tedavisine verilen yanıtların ateş, karın ağrısı atağı ve atak fazı sıklığı semptomlarının varlığına ve yokluğuna göre sayısal dağılımları

Belirlenen kriterlere uyan ve kolşisin tedavisine olumlu yanıt veren toplam 122 hasta bulunmaktadır. Şekil 6'nın incelenmesi, en az bir kriterin 'yok' değerini aldığı durumlarda bile hastaların kolşisin tedavisine olumlu yanıt verdiklerini göstermektedir. Bu durum, belirlenen kriterlerin hastaların tedavi yanıtlarını etkileyen önemli göstergeler olduğunu düşündürmektedir. Bu bulgular, tedavi stratejilerinin belirlenmesi ve hastaların spesifik semptomlara yanıt verme durumlarının daha iyi anlaşılması için değerli bir perspektif sunmaktadır.

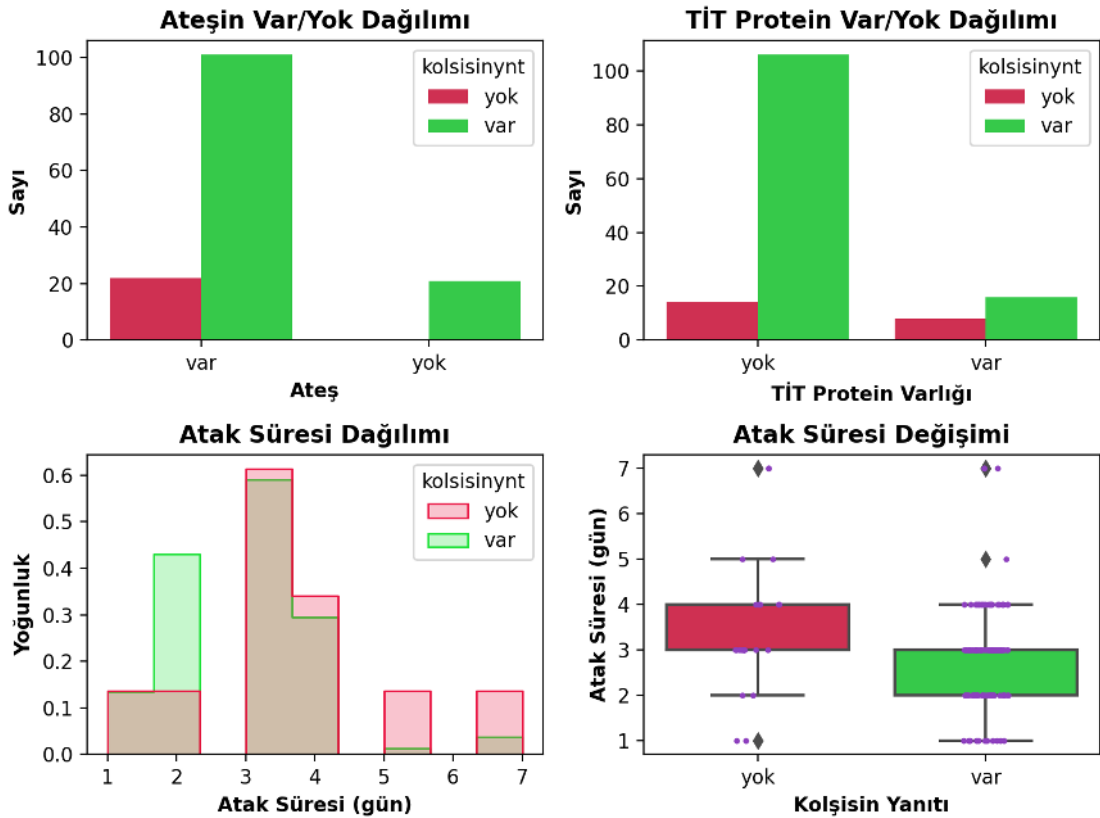
AAA hastalığına yönelik kolşisin tedavisi, ateş, TİT protein varlığı ve atak süresi gibi değişkenlerin dağılımı üzerindeki etkileri bakımından incelenmiştir. Elde edilen bulgular şu şekildedir:

- 1. Ateşin Varlığı/Yokluğu:** Kolşisin tedavisine yanıt veren hastaların büyük bir kısmında ateş görülürken, yanıt vermeyenlerde bu durum daha az rastlanmıştır. Bu durum, kolşisinin ateşli semptomları etkileyebileceğini ima etmektedir. İlginç olarak, ateşin bulunmadığı vakalarda, kolşisin tedavisinden olumlu yanıt alınması sıklıkla gözlemlenmiştir.
- 2. TİT Proteininin Varlığı/Yokluğu:** Araştırma sonuçları, TİT proteininin yokluğunun, kolşisin tedavisine yanıt veren hastalarda daha fazla görüldüğünü ortaya koymuştur. Bu, TİT protein eksikliğinin, kolşisin tedavisine verilen yanıtta bir gösterge olabileceğini düşündürmektedir.

3. Atak Süresi Dağılımı: Atak süresinin, kolşisin tedavisine yanıt veren ve vermeyen hastalar arasında farklılık gösterdiği belirlenmiştir. Kolşisin tedavisine yanıt veren gruplarda, atak süresi genellikle daha kısadır, buna karşın yanıt vermeyen gruplarda bu süre uzundur.

4. Atak Süresi Değişimi (Boxplot Analizi): Boxplot analizi, kolşisin tedavisine yanıt vermeyen hastalarda atak süresinin daha yüksek varyans gösterdiğini işaret etmektedir. Bu bulgu, yanıt vermeyen hastalarda atak sürelerinin daha değişken olduğunu ve tedaviye farklı tepkiler sergilenebileceğini göstermektedir.

Bu gözlemler, kolşisin tedavisinin etkinliğinin değerlendirilmesinde kullanılabilecek potansiyel biyolojik ve klinik göstergeler ile tedaviye yanıtı etkileyebilecek faktörler hakkında kritik bilgiler sağlamaktadır. Bu bulgular, daha geniş klinik çalışmalar ve araştırmalar için temel oluşturabilir ve AAA hastalarının tedavisinde önemli bir yol gösterici olabilir.



Şekil 2.4. Kolşisin yanıtına göre Atak Süresi, TİT Protein ve Ateş değişkenlerinin dağılımları

Bu çalışmada, toplam 144 Ailevi Akdeniz Ateşi (AAA) hastası üzerinde yapılan veri seti analizi ele alınmıştır. İncelenen hasta grubu içinde, ateş, karın ağrısı ve atak sırasındaki akut faz reaksiyonları gibi belirtileri gösteren, ancak kolşisin tedavisine yanıt vermeyen 22

hasta belirlenmiştir. Bu özel alt grup, incelenen toplam hasta popülasyonunun yaklaşık %15.28'lik bir kısmını temsil etmektedir. Bu oran, AAA hastaları arasında kolşisin tedavisine dirençli bir alt grubun varlığını göstermektedir ve bu grup, hastalığın tedavi yöntemlerinin geliştirilmesi açısından önemli bir odak noktası oluşturabilir. Bu bulgular, AAA'nın tedavisinde alternatif yaklaşımların araştırılmasını gerektirebilir ve bu alt grubun özellikleri, hastalığın daha iyi anlaşılması ve yönetilmesi için kritik öneme sahip olabilir.

4.5.1. Kolşisin Yanıtsız Hastaların Özellikleri

4.5.1.1. Demografik ve Coğrafi Özellikler

Cinsiyet Dağılımı: Alt grubu oluşturan hastaların 12'si erkek, 10'u kadındır.
Coğrafi Dağılım: Bu hastalardan 14'ü Marmara grubundan, 8'i ise Karadeniz grubundandır.

Yaş: Semptomların başlangıç yaşı ortalaması 19.09, standart sapması ise 10.91'dir.

4.5.1.2. Klinik Özellikler

Atak süresi: Atak süresi ortalaması 3.50 gün, standart sapması 1.54 gündür.

Atak sıklığı: Atak sıklığı ortalaması 45.45 gün, standart sapması 33.70 gündür.

Deri Lezyonu (Ebe): Tüm hastalarda var olarak gözlemlenmiştir.

Eklem Ağrıları: Kalça, diz ve ayak bileği ağrısı 17 hastada yok, 5 hastada var olarak kaydedilmiştir.

4.5.1.3. Aile Öyküsü ve Genetik Faktörler

Aile öyküsü: 17 hastanın ailesinde AAA öyküsü olduğu, 5 hastanın ise ailesinde AAA öyküsü olmadığı tespit edilmiştir.

-Tit Protein Varlığı: 14 hastada var, 8 hastada yoktur.

-Yaygın mutasyon: 12 hastada var, 10 hastada yoktur.

-Genetik Mutasyonlar: MEFV geninin heterozigot ve homozigot sayısı eşittir.

Bu bulgular, kolşisin tedavisine yanıtsız hastaların özelliklerinin çok yönlü bir şekilde incelenmesi gerektiğini göstermektedir. Özellikle coğrafi ve demografik faktörlerin, yanı sıra klinik ve genetik varyasyonların bu yanıtsızlıkta rol oynayabileceği anlaşılmaktadır.

5. AİLEVİ AKDENİZ ATEŞİ HASTALARININ ANALİZİ

5.1. Veri Ön İşleme ve Makine Öğrenmesi Modelleri

Bu bölüm, Ailevi Akdeniz Ateşi (AAA) hastalarının kapsamlı veri seti üzerinde yapılan analizleri sunmaktadır. Çalışma, hastaların demografik, klinik ve genetik özelliklerini içeren veri seti üzerinde detaylı bir ön işleme ve analiz süreci gerçekleştirmiştir.

5.1. Veri Setinin Düzenlenmesi ve One-Hot Encoding İşlemi

Bu çalışmada, Ailevi Akdeniz Ateşi (AAA) hastalarına ait eksiksiz ve önceden temizlenmiş bir veri seti kullanılmıştır. Veri seti, hastaların klinik, demografik ve genetik özelliklerini içermekte olup, analiz için hazır bir durumda sunulmuştur. Ana odak, veri集中的i kategorik değişkenlerin sayısal formata dönüştürülmesi ve analiz için uygun hale getirilmesidir.

5.3. One-Hot Encoding Yöntemi

Makine öğrenmesi modellerinin kategorik verileri etkin bir biçimde işleyebilmesi için bu verilerin sayısal formata dönüştürülmesi kritik bir öneme sahiptir. Bu araştırmada, kategorik değişkenlerin sayısallaştırılmasında One-Hot Encoding tekniği tercih edilmiştir. Bu yöntem, her kategorik değişkeni temsil eden ayrı sütunlar oluşturarak, ilgili kategorinin mevcut olup olmadığını 0 ve 1 değerleriyle ifade eder. Bu dönüşüm, modelin kategorik verileri daha hassas bir şekilde analiz etmesine ve dolayısıyla daha doğru tahminlerde bulunmasına yardımcı olur. Bu yöntemin uygulanması, modelin performansını artırarak, veri集中的i kategorik özelliklerin etkilerini daha ayrıntılı bir şekilde anlamasına olanak sağlar.

5.4. Veri Setinin Yapısı

Veri seti, AAA hastalığının çeşitli yönlerini yansıtan zengin bir bilgi seti sunmaktadır. Veriler, hastaların yaş, cinsiyet, genetik özellikler, klinik semptomlar ve tedavi yanıtları gibi çeşitli özellikleri içermektedir. One-Hot Encoding işlemi sonrasında, bu kategorik değişkenler sayısal formata dönüştürülmüş ve analiz için hazır hale getirilmiştir.

Bu bölüm, veri setinin nasıl düzenlendiğini ve kategorik verilerin One-Hot Encoding yöntemiyle nasıl işlendiğini açıklamaktadır. Bu işlemler, veri setinin makine öğrenmesi modelleri tarafından daha etkili bir şekilde kullanılmasını sağlar ve analiz sürecinin başarısını artırır. Özellikle, AAA hastalığına dair geniş kapsamlı bir veri setinin analizi, hastalığın daha iyi anlaşılması ve tedavi stratejilerinin geliştirilmesi için önemli bir adımı temsil eder.

5.5. Makine Öğrenmesi Modelleri, Özellik Seçimi, Çapraz Doğrulama ve Performans Analizi

Bu bölüm, Ailevi Akdeniz Ateşi (AAA) hastalığının tedavi ve seyrine ilişkin derinlemesine anlayış kazanmak amacıyla, ilgili hasta veri seti üzerinde yürütülen kapsamlı makine öğrenmesi analizlerini, özellik seçimi metodolojilerini ve çapraz doğrulama tekniklerini detaylı bir biçimde tartışmaktadır. Bu analizler, hastalığın tedavi süreçlerine ve gelişimine dair önemli bilgiler sunmayı amaçlamakta ve bu sayede AAA hastalığının daha etkin bir şekilde yönetilmesine katkıda bulunmayı hedeflemektedir. İlgili çalışmalar, makine öğrenmesi yaklaşımlarının sağlık verileri üzerindeki uygulamalarına dair değerli örnekler teşkil etmekte, böylece hastalığın tanı ve tedavi süreçlerinin optimizasyonuna yönelik stratejilerin geliştirilmesine olanak sağlamaktadır. Bu bölümde ayrıca, analiz sürecinde kullanılan özellik seçimi ve model doğrulama yöntemlerinin detaylarına ve bu yöntemlerin analiz sonuçları üzerindeki etkilerine de yer verilmektedir.

5.6. Seçilen Özellikler

Analiz sürecinde SelectKBest yöntemi kullanılarak seçilen özellikler şunlardır:

- Atak Süresi (atksur)
- Atak Sıklığı (atksklg)
- Ailede AAA Yokluğu (ailefmf_yok)
- Karın Ağrısı Atağının Yokluğu (krnagrsatk_yok)
- Ateşin Yokluğu (ates_yok)
- Atak Akut Faz Yüksekliğinin Yokluğu (atkaktfazyk_yok)
- TİT Protein Varlığının Yokluğu (titprteinvar_yok)
- MEFV Mutasyonunun Yokluğu (yygnmutvar_yok)

Bu özellikler, AAA hastalarının kolsişin yanıtlarını etkileyen en önemli faktörlerdir.

5.7. Makine Öğrenmesi Modelleri ve Performansları

Çalışmada kullanılan modeller ve performansları Tablo 2.1’de verilmiştir:

Tablo 2.1. Anlamlı özellikler için model performansları

Model	Doğruluk (%)	Hassasiyet (%)	F1 Skoru (%)
Logistic Regression	85.42	85.81	92.00
Random Forest	81.23	86.99	89.08
Support Vector Machine	84.73	84.73	91.72
K-Nearest Neighbors	84.04	85.67	91.17
Gradient Boost	80.52	88.04	88.40
Decision Tree	75.71	88.87	84.81

5.8. Model Değerlendirmesi ve Anlamları

Bu modellerin performans değerlendirmesi, seçilen özelliklerin AAA hastalığının tedavi yanıtlarını tahmin etmede ne kadar etkili olduğunu göstermektedir. Logistic Regression ve SVM modellerinin özellikle yüksek doğruluk ve F1 skorları, bu özelliklerin tedaviye yanıtları etkileyen önemli faktörler olduğunu doğrulamaktadır. Çapraz doğrulama ile elde edilen bu sonuçlar, modelin genellemesini ve güvenilirliğini artıran değerli bir yöntem olarak öne çıkmaktadır. Bu analizler, AAA hastalığının daha derinlemesine anlaşılması ve tedavi stratejilerinin optimize edilmesi için önemli katkılar sunar.

5.8.1. Özelliklerin İstatistiksel Analizi: Kruskal-Wallis ve Post-hoc Testler

Bu bölümde, Ailevi Akdeniz Ateşi (AAA) hastalığına ilişkin veri setinde belirlenen anlamlı özelliklerin istatistiksel analizleri ele alınmaktadır. Analiz, Kruskal-Wallis testi ve bunu takiben yapılan Post-hoc (Dunn's) testleri içermektedir. Bu testler, seçilen özelliklerin hastaların kolşisin tedavisine yanıtını nasıl etkilediğini anlamak için kullanılmıştır.

5.9. Kruskal-Wallis Testi

Kruskal-Wallis testi, kategorik gruplar arasında sayısal değişkenlerin dağılımını karşılaştırmak için kullanılan bir non-parametrik testtir. Bu test, farklı grupların aynı

popülasyondan gelip gelmediğini anlamak için kullanılır. AAA hastalığı veri setinde, 'kolsisinynt' değişkeni grup olarak kullanılmış ve diğer sayısal özellikler bu gruplar arasında test edilmiştir.

Tablo 2.2. Kruskal-Wallis testi sonuçları

Özellik	H-istatistiği	p-değeri
TİT Protein Varlığının Yokluğu (titprteinvar_yok)	7.203279	0.007277
Atak Sıklığı (atksklg)	4.953040	0.026045
Atak Süresi (atksur)	4.479078	0.034312
Ateşin Yokluğu (ates_yok)	4.402639	0.035883
Atak Akut Faz Yüksekliği Yokluğu (atkaktfazyk_yok)	4.159175	0.041410

Tablo 2.2, 'kolsisinynt' ile istatistiksel olarak anlamlı ilişkisi bulunan özelliklerin H-istatistik ve p-değerlerini göstermektedir. Tablo 4'deki her bir özellik, AAA hastalarının kolşisin tedavisine yanıtlarını etkileyen önemli faktörler olarak saptanmış ve bu özelliklerin gruplar arasında istatistiksel olarak anlamlı farklılıklara sahip olduğunu göstermektedir. Bu sonuçlar, tedaviye yanıtı etkileyebilecek önemli biyolojik ve klinik faktörlerin belirlenmesinde yardımcı olabilir ve AAA hastalığının tedavisine yönelik daha etkili stratejilerin geliştirilmesine katkı sağlayabilir.

5.9.1. Post-hoc (Dunn's) Test Sonuçları: Sayısal Analizler ve Değerlendirmeler

Kruskal-Wallis testi sonucunda anlamlı bulunan özellikler için yapılan Post-hoc (Dunn's) test sonuçları, AAA hastalarının kolşisin tedavisine yanıtlarının daha ayrıntılı bir şekilde incelenmesini sağlamıştır. Aşağıda, her bir özellik için elde edilen sayısal sonuçlar ve bu sonuçların anlamları detaylandırılmıştır.

Tablo 2.3. atksur Özelliği (Atak Süresi)

Grup	Grup 0	Grup1
0	1.000	0.0343
1	0.0343	1.000

Yorum: Atak süresi, kolşisin tedavisine yanıt veren (Grup 1) ve vermeyen (Grup 0) hastalar arasında istatistiksel olarak anlamlı bir farklılık göstermiştir. P-değeri (0.0343), bu iki grup arasında atak süresi açısından anlamlı bir fark olduğunu belirtir.

Tablo 2.4. atksklg Özelliği (Atak Sıklığı)

Grup	Grup 0	Grup1
0	1.000	0.0260
1	0.0260	1.000

Yorum: Atak sıklığı da benzer şekilde, tedaviye yanıt veren ve vermeyen gruplar arasında istatistiksel olarak anlamlı bir farklılık göstermiştir. P-değeri (0.0260), bu farkın anlamlı olduğunu göstermektedir.

Tablo 2.5. ates_yok Özelliği (Ateş Yokluğu)

Grup	Grup 0	Grup1
0	1.000	0.0343
1	0.0343	1.000

Yorum: Ateşin yokluğu, kolşisin tedavisine yanıt üzerinde etkili bir faktör olarak ortaya çıkmıştır. P-değeri (0.0358), bu özelliğin tedaviye yanıtta belirgin bir fark yarattığını gösterir.

Tablo 2.6. atkaktfazyk_yok Özelliği (TİT Protein Varlığının Yokluğu)

Grup	Grup 0	Grup1
0	1.000	0.0414
1	0.0414	1.000

Yorum: Atak akut faz yüksekliğinin yokluğu, tedaviye yanıt verme durumunda anlamlı bir farklılık göstermiştir. Bu özelliğin tedaviye yanıtta etkili olduğu anlaşılmaktadır.

Tablo 2.7. titprteinvar_yok Özelliği (Atak Süresi)

Grup	Grup 0	Grup1
0	1.000	0.0072
1	0.0072	1.000

Yorum: TİT protein varlığının yokluğu, en düşük p-değeri (0.0072) ile tedaviye yanıt üzerinde oldukça belirgin bir fark yaratmaktadır. Bu, özellikle bu proteinin tedavi sürecindeki önemini vurgulamaktadır.

Bu sayısal sonuçlar, AAA hastalarının kolşisin tedavisine yanıtlarını etkileyen faktörlerin detaylı bir şekilde incelenmesinde önemli rol oynamaktadır. Post-hoc testler, özellikle belirli özelliklerin tedaviye yanıtta nasıl farklılaştığını belirlemekte ve bu bilgiler, hastalığın tedavisine yönelik stratejilerin geliştirilmesinde önemli içgörüler sunmaktadır. Bu analizler, tedaviye yanıtı etkileyebilecek kritik faktörlerin daha iyi anlaşılması ve bireyselleştirilmiş tedavi yaklaşımlarının geliştirilmesine katkıda bulunabilir.

6. SONUÇLAR VE ÖNERİLER

6.1. Sonuçlar

Bu çalışmada, Ailevi Akdeniz Ateşi (AAA) hastalarına ait bir veri seti üzerinde yapılan analizler sonucunda, hastalığın teşhis ve tedavi süreçlerinde önemli bulgular elde edilmiştir. Makine öğrenmesi yöntemleri ve çeşitli istatistiksel analiz teknikleri kullanılarak, AAA hastalığının çeşitli yönleri detaylı bir şekilde incelenmiştir.

Demografik ve Klinik Gözlemler: AAA hastalarının demografik ve klinik özelliklerinin, kolşisin tedavisine verilen yanıtlar üzerinde önemli etkileri olduğu tespit edilmiştir. Özellikle ateş, atak süresi ve TİT protein varlığı gibi değişkenler, tedaviye yanıt verme durumunu etkileyen önemli faktörler olarak ortaya çıkmıştır.

Makine Öğrenmesi Modellerinin Değerlendirilmesi: Logistic Regression, Random Forest, ve Support Vector Machine gibi çeşitli makine öğrenmesi modellerinin performansları değerlendirilmiş ve bu modellerin AAA hastalığının teşhisinde yüksek doğruluk ve hassasiyetle çalıştığı gözlemlenmiştir.

Özellik Seçimi ve İstatistiksel Analizler: Seçilen özelliklerin istatistiksel analizi, Kruskal-Wallis ve Post-hoc (Dunn's) testleri kullanılarak yapılmış ve bu özelliklerin tedaviye yanıt üzerindeki etkileri belirlenmiştir.

6.2. Öneriler

Bu çalışmanın bulguları, AAA hastalığının teşhis ve tedavi süreçlerinde kullanılabilir ve önemli içgörüler sunmaktadır. Bu bağlamda aşağıdaki öneriler sunulabilir:

1. Kişiselleştirilmiş Tedavi Yaklaşımları: Elde edilen bulgular, farklı demografik ve klinik özelliklere sahip hastaların tedaviye farklı yanıtlar verebileceğini göstermektedir. Bu nedenle, tedavi stratejilerinin kişiselleştirilmesi ve her hastanın özgün ihtiyaçlarına göre uyarlanması önemlidir.

2. Erken Teşhis ve İzleme: AAA hastalığının erken teşhisinde ve hastaların tedavi süreçlerinin izlenmesinde makine öğrenmesi modellerinin kullanımı, daha etkin ve hızlı teşhis olanakları sunabilir.

3. Geniş Kapsamlı Araştırmalar: Elde edilen sonuçlar, daha geniş kapsamlı araştırmalar için sağlam bir temel oluşturur. Bu bulgular, ileri araştırmalarda derinlemesine incelenebilir ve daha geniş hasta popülasyonları üzerinde test edilebilir.

4. Eğitim ve Bilinçlendirme: AAA hastalığı hakkında toplumda ve sağlık profesyonelleri arasında daha fazla bilinçlendirme ve eğitim gerekmektedir. Bu, hastalığın erken teşhisi ve etkin tedavisinde önemli bir rol oynayabilir.

5. Veri Setinin Genişletilmesi: Gelecekteki çalışmalar için, farklı coğrafi bölgelerden ve etnik gruplardan daha fazla sayıda hastayı içeren daha geniş veri setlerinin oluşturulması önerilir. Bu, hastalığın daha kapsamlı anlaşılmasına ve tedavi yaklaşımlarının daha geniş bir yelpazede geliştirilmesine katkı sağlayacaktır.

6.3. Sonuç

Bu tez çalışması, Ailevi Akdeniz Ateşi hastalığının karmaşıklığını ve tedavi süreçlerindeki zorlukları ortaya koymuştur. Makine öğrenmesi ve istatistiksel analiz yöntemlerinin kullanımı, bu hastalığın daha iyi anlaşılması ve tedavi edilmesi yönünde önemli adımlar atılmasını sağlamıştır. Elde edilen bulgular ve öneriler, bu alandaki gelecekteki araştırmalar için değerli bir temel oluşturmakta ve AAA hastalığının daha etkin bir şekilde yönetilmesine katkıda bulunmaktadır.

KAYNAKÇA

- A, Adato. R, Brenner. A, Levy et al.** (2022). Determining the origin of different variants associated with familial mediterranean fever by machine-learning, *Sci Rep*, 12, 15206.
- B, B. A.** (2003). Familial Mediterranean fever. *Pediatr Nephrol*; 18: 853-9.
- C, Çalışkan S. K. Soğukpınar, İ.** (2008). KxKNN: K-Means ve K En Yakın Komşu Yöntemleri ile Ağlarda Nüfuz Tespiti, *EMO Yayınları*, 120-124.
- Ç, Çalışkan S.** (2008). Veri Madenciliğinde Kümeleme Teknikleri Üzerine Bir çalışma: KMeans ve K-Medoids Kümeleme Algoritmalarının Karşılaştırılması, *Journal of Natural & Applied Sciences*, 15-3,192-202.
- D, Daniel, W. W.** (1990). Applied Nonparametric Statistics. Boston: *PWS-Kent*.
- E, Demiralay M., Çamurcu., A. Y.** (2005). Cure, Agnes ve K-Means Algoritmalarındaki Kümeleme Yeteneklerinin Karşılaştırılması, *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 4(8), 1-18.
- F, Dunn. O, J.** (1964). Multiple comparisons using rank sums. *Technometrics*, 6(3), 241-252.
- G, Guyon. I, Elisseeff.** (2003). An Introduction to Variable and Feature Selection, *Journal of Machine Learning Research*. 3 1157-1182.
- H, Haghighat M. Derakhshan A, Karamifar H.** (2006). Familial Mediterranean Fever. *Shiraz E Medical Journal*; 1: 7-18.
- I, Hollander. M. D, A. Wolfe, E.** (2013). Chicken, Nonparametric statistical methods. John Wiley & Sons.
- İ, İ. Sarı, M. Birlik, T. Kaşifoğlu.** (2014). Familial Mediterranean fever: An updated review, *Eur. J. Rheumatol*, 21: 1-33.
- J, Kavak US. Özen S.** (2003). Ailesel Akdeniz Ateşi. *Sted*; 40: 12-137.
- K, Kasapçopur Ö. Arısoy N.** (2006). Ailesel Akdeniz Ateşi ve diğer otoenflamatuar hastalıklar. *Türk Pediatri Ars*; 41: 9-17.
- L, Kasapçopur Ö. Özdoğan H.** (2007). Ailesel Akdeniz Ateşi İ.Ü. Cerrahpaşa Tıp Fak. Sürekli Tıp eğitimi etkinlikleri, Türkiye'de sık karşılaşılan hastalıklar I. Enfeksiyon Hastalıkları, Romatizmal Hastalıklar, Afetlerde Ezilme Yaralanmaları, *Sempozyum Dizisi* 55: 43-54.

- M, Kruskal. W, H. & Wallis, W. A.** (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47 (260), 583-621.
- N, Koltan Ş. Patır, S.** (2011). Kümeleme Analizi ve Pazarlamada Kullanımı, *Akademik Yaklaşımlar Dergisi (Journal of Academic Approach)*, 2(1), 91-113.
- O.,Koşan C.** (2003). Ailevi Akdeniz Ateşine tanısal yaklaşım. *AÜTD*; 35: 1-6.
- Ö, Özekes, S.** (2003). Veri madenciliği modelleri ve uygulama alanları, *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*,3,65-82.
- P, Paşa S. Altınta, A. Devecioğlu, B. Cil, T. Danis, R. Isi, H. et al.** (2008). Familial Mediterranean Fever gene mutations in the Southeastern region of Turkey and their phenotypical features. *Amyloid*; 15: 49-53.
- R, R. C. Goldstein, A.D. Schwabe.** (1974). Prophylactic colchicine therapy in familial Mediterranean fever. A controlled, double-blind study, *Ann Intern Med*, 81: 792-4.
- S. Sarıman G.** (2011). Veri Madenciliğinde Kümeleme Teknikleri Üzerine Bir çalışma: KMeans ve K-Medoids Kümeleme Algoritmalarının Karşılaştırılması, *Journal of Natural & Applied Sciences*, 15-3,192-202.
- Ş, Sefer Üstebay. D, Ü. Üstebay.** (2015). Yunus Yılmaz Ailevi Akdeniz Ateşi Familial Mediterranean Fever [Erişim: 12.12.2023, https://cms.galenos.com.tr/Uploads/Article_35983/Jarem-5-89.Pdf]
- Yavuz, Ü. Ekim, U. ve Köklü, M.** (2011). Üniversite Öğrencilerin Ortak Zorunlu Derslerdeki Başarılarının K-Means Algoritması İle İncelenmesi, *NWSA: Engineering Sciences*, 6(1), 342-347.
- Muhammet, İ, Y.** (2021). Lojistik regresyon modeli nasıl çalışır? 2021-1 [Erişim: 22.01.2024, <https://www.ikbalyildiz.com/lojistik-regresyon-modeli-nasil-calisir/>]
- Cook, D. Dixon, P. Duckworth, W. M, Kaiser M. S, Koehler, K. Meeker, W. Q. & Stephenson W. R.** (2001). Binary response and logistic regression analysis, in Part of the Iowa State University NSF/ILI project. Beyond Traditional Statistical Methods.
- Barak, A. Karahan, S. ve Saraçbaşı, O.** (2005). Ordinal lojistik regresyon modelleri. 4. İstatistik Kongresi'nde sunulan bildiri, Belek-Antalya.
- T, S, Ozgocmen. O, Akgul.** (2021). Anti-TNF agents in familial Mediterranean fever: report of three cases and review of the literature. *Mod Rheumatol* 21: 684-90.

U, Üstebay. S, Üstebay, D. Ü. Üstebay, Y. Yılmaz. (2015). Ailevi Akdeniz Ateşi. *Journal of Academic research in medicine.*

Süel, H. (2014). Isparta-Sütçüler Yöresinde av türlerinin habitat uygunluk modellemesi. *Doktora tezi, Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü, Isparta.*

Ü, Ü, Akkuş Ş. Çalışkan S, Kasapçopur Ö. (2002). Tubular functions in Familial Mediterranean Fever. *Turkish J Pediatr*; 44: 317-20.

V, V. J. Parvaneh, R. Shiari. (2015). Treatment of Colchicine-Resistant Familial Mediterranean Fever With Anakinra. *Arch Pediatr Infect Dis*, 3(3): e21664.

W, W. H, Kruskal, W. A. Wallis. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583-621.

Coşgun, E. Karabulut, E. Karaağaoğlu, E. (2009). Random Forest ve Destek Vektör Makinası Yöntemleri ile Gen Seçimi ve Sınıflaması. *VI. Ulusal İstatistik Kongresi: 29 Mayıs 2009- Antalya, Türkiye.*

Breiman, L. (2001). Random Forest [Elektronik Sürüm]. *Machine Learning*, 45, 5-32.

Breiman, L. Cutlar, A. Random Forest. [Erişim: 20 Mart 2010, University of Stat Barkly Ağ Sitesi. http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm]

Y, Y. Yeşilada E, Savacı S, Yüksel Ş, Gülbay G, Otlı G, Kaygusuzoğlu E.Ailesel. (2005). Akdeniz Ateşi düşünülen olgularda MEFV gen mutasyonu. *İnönü Üniversitesi Tıp Fakültesi Dergisi*; 12:8-235.

Z. Zemer, D. Livneh, A. Livneh, Y. L. Danon, M. Pras, E. Sohar. (1991). Long-term colchicine treatment in children with familial Mediterranean fever, *Arthritis Rheum*, 7: 34-973

Vapnik, V. (1995) *The Nature of Statistical Learning Theory*, Springer-Verlag Publishing, New York, s.187.

Cristianini, N. and Taylor, J. S. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, s.172.

Cover, T.M. and Hart, P.E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, IT13(1):21–27.

Bhatia, N. and Vandana. (2010). “Survey of nearest neighbor techniques”, *International Journal of Computer Science and Information Security*, 8(2):302-305.

- Qiu, X.Y. Kang, K. and Zhang, H.X.** (2008). Selection of kernel parameters for K-NN, *IEEE International Joint Conference on Neural Networks (IJCNN)*, 61-65.
- Batista, G.E.A.P.A. and Silva, D.F.** (2009). How k-nearest neighbor parameters affect its performance, *Simposio Argentino de Inteligencia Artificial (ASAI 2009)*, 95–106.
- Keller, J. M. Gray, M.R. and Givens, J.A.** (1985). “A fuzzy K-nearest neighbor algorithm”, *IEEE Transactions on Systems, Man and Cybernetics*, SMC15(4):580-585.
- Mao, C. Hu, B. Wang, M. and Moore, P.** (2015). Learning from neighborhood for classification with local distribution characteristics, *IEEE International Joint Conference on Neural Networks (IJCNN)*, 1-8.
- Liu, H. and Zhang, S.** (2012). Noisy data elimination using mutual k-nearest neighbor for classification mining, *Journal of Systems and Software*, 85(5):1067-1074
- O, Maimon. L, Rokach.** (2010). Classification Trees. *Data Mining and Knowledge Discovery Handbook*, Editor: O., Maimon, L. Rokach, Springer, New York, A.B.D., 149-175.
- J, Gehrke.** (2003). “Decision Trees”, *The Handbook of Data Mining*, Editor: Nong Ye, Lawrence Erlbaum Associates Publishers, London, 149-175.
- P, J, García-Laencina. J, L.** (2010). Sancho-Gómez, A.R. Figueiras-Vidal, “Pattern Classification with Missing Data: A Review”, *Neural Comput. Appl.*, 19(2), 263-282, .
- D, Birant.** (2011). “Comparison of Decision Tree Algorithms for Predicting Potential Air Pollutant Emissions with Data Mining Models”, *Journal of Environmental Informatics*, 17(1), 46-53.