

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENME Sİ ALGORİTMALARI KULLANILARAK
SAĞLIK SİSTEMİNDE TAHMİN ÇALIŞMASI

YÜKSEK LİSANS TEZİ

Gülsüm SALTAN YAŞLI

Endüstri Mühendisliği Anabilim Dalı

OCAK 2024

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENME Sİ ALGORİTMALARI KULLANILARAK
SAĞLIK SİSTEMİNDE TAHMİN ÇALIŞMASI

YÜKSEK LİSANS TEZİ

Gülsüm SALTAN YAŞLI

Endüstri Mühendisliği Anabilim Dalı

Tez Danışmanı: Doç. Dr. Seher ARSLANKAYA

OCAK 2024

Gölsüm SALTAN YAŞLI tarafından hazırlanan “MAKİNE ÖĞRENMESİ ALGORİTMALARI KULLANILARAK SAĞLIK SİSTEMİNDE TAHMİN ÇALIŞMASI” adlı tez çalışması 15.01.2024 tarihinde aşağıdaki jüri tarafından oy birliği ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Tez Jürisi

Jüri Başkanı : **Doç. Dr. Seher ARSLANKAYA** (Danışman)
Sakarya Üniversitesi

Jüri Üyesi : **Doç. Dr. Tülay K. POLAT**
Sakarya Üniversitesi

Jüri Üyesi : **Dr. Öğr. Üyesi Pelin A. GEZGİNCİ**
Okan Üniversitesi



ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum “MAKİNE ÖĞRENMESİ ALGORİTMALARI KULLANILARAK SAĞLIK SİSTEMİNDE TAHMİN ÇALIŞMASI” başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığını, etik kurul onay belgesi aldığımı, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

(...../...../20.....).

(imza)

Gülsüm Saltan Yaşlı





Canim aileme



TEŞEKKÜR

Yüksek lisans eğitim sürecinde bilgi ve tecrübelerinden faydalandığım, desteğini esirgemeyen değerli danışman hocam Doç. Dr. Seher ARSLANKAYA'ya teşekkürlerimi sunarım.

Bu çalışmamda; gerçek verilerin toplanması, değerlendirilmesi sürecinde tüm bilgi ve tecrübelerini paylaşan, desteğini esirgemeyen Kartal Dr. Lütfi Kırdar Şehir Hastanesi Acil Tıp Kliniği'nde görev yapmakta olan Uzm. Dr. Bengü MUTLU SARIÇİÇEK'e ve Doç. Dr. Avni Uygar SEYHAN'a; yüksek lisans eğitimim boyunca anlayışla beni destekledikleri için Mutlu Akü'deki yöneticilerime teşekkür ederim.

Ayrıca hayatım ve yüksek lisans eğitimim boyunca daima yanımda olan ve daima destekçim olan aileme sonsuz teşekkür ederim.

Gülsüm Saltan Yaşlı



İÇİNDEKİLER

Sayfa

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ	v
TEŞEKKÜR	ix
İÇİNDEKİLER	xi
KISALTMALAR	xv
SİMGELER	xvii
TABLO LİSTESİ	xix
ŞEKİL LİSTESİ	xxi
ÖZET	xxiii
SUMMARY	xxv
1. GİRİŞ	1
1.1. Tezin Kapsamı	2
1.2. Çalışmanın Konusu	3
1.3. Çalışmanın Amacı	3
1.4. Çalışmanın Önemi	3
1.5. Çalışmanın Yöntemi	4
2. LİTERATÜR ARAŞTIRMASI	7
3. MATERYAL VE YÖNTEM	15
3.1. Acil Tıp ve Triyaj	15
3.1.1. Triyaj türleri	18
3.1.1.1. Avustralya triyaj skalası	20
3.1.1.2. Acil durum ciddiyet indeksi	21
3.1.1.3. Kanada triyaj ve aciliyet skalası	23
3.1.1.4. Manchester triyaj skalası	24
3.1.1.5. Güney Afrika triyaj skalası	25
3.1.1.6. Fransa acil hemşireliği hastane ölçeği sınıflandırması	26
3.1.1.7. Acil tıp Japon birliği	26
3.1.1.8. Türkiye triyaj sistemi	26
3.1.2. Yaşamsal bulgular	27
3.1.2.1. Kalp hızı	27
3.1.2.2. Kan basıncı	27
3.1.2.3. Vücut sıcaklığı	27
3.1.2.4. Solunum sayısı	28
3.1.2.5. Oksijen satürasyonu	28
3.1.2.6. Kan şekeri	28
3.1.2.7. Glasgow koma skalası	28
3.2. Makine Öğrenmesi	28
3.2.1. Makine öğrenmesi türleri	31
3.2.1.1. Denetimli öğrenme	31
3.2.1.2. Denetimsiz öğrenme	33
3.2.1.3. Yarı denetimli öğrenme	34
3.2.1.4. Pekiştirmeli öğrenme	34

3.2.2. Makine öğrenmesi sınıflandırma algoritmaları	34
3.2.2.1. Karar ağacı algoritması	35
3.2.2.2. K-en yakın komşu algoritması	39
3.2.2.3. Naive bayes algoritması	43
3.2.2.4. Rastgele orman algoritması	44
3.2.2.5. Doğrusal destek vektör makineleri algoritması	47
3.2.2.6. Doğrusal olmayan destek vektör makineleri	50
3.2.2.7. Lojistik regresyon	53
3.2.3. Makine öğrenmesi sınıflandırma algoritmaları performans ölçümü	56
3.2.4. Çapraz doğrulama	59
3.2.5. Python programı	61
3.3. Veri Seti	63
3.3.1. Veri toplama süreci	63
3.3.2. Veri seti tanıtımı	64
3.3.3. Veri setinin istatistiksel ve grafiksel analizi	70
4. ARAŞTIRMA BULGULARI	77
4.1. Uygulamanın Amacı ve Veri Seti Hazırlık	77
4.2. Triyaj Alan Tahmini Çalışması	78
4.2.1. Karar ağaçları algoritması	79
4.2.1.1. Karar ağacı-gini algoritması	79
4.2.1.2. Karar ağacı-entropi algoritması	80
4.2.2. Doğrusal destek vektör makineleri algoritması	81
4.2.3. Doğrusal olmayan destek vektör makineleri algoritması	82
4.2.3.1. Doğrusal olmayan destek vektör algoritması-RBF	82
4.2.3.2. Doğrusal olmayan destek vektör algoritması-poly	83
4.2.3.3. Doğrusal olmayan destek vektör algoritması-sigmoid	84
4.2.4. Naive bayes algoritması	85
4.2.5. Lojistik regresyon algoritması	86
4.2.6. K-En yakın komşu algoritması	87
4.2.6.1. K-En yakın komşu algoritması-manhattan	87
4.2.6.2. K-En yakın komşu algoritması-minkowski	88
4.2.7. Rassal orman algoritması	89
4.2.7.1. Rassal orman algoritması-entropi	89
4.2.7.2. Rassal orman algoritması-gini	90
4.2.8. AUCROC değer hesabı	91
4.2.9. Algoritma performansı değerlendirme	93
4.2.10. Triyaj alan tahmini için çapraz doğrulama	94
4.3. Yatış-Taburcu Durumu Tahmini Çalışması	97
4.3.1. Karar ağaçları algoritması	97
4.3.1.1. Karar ağacı-gini algoritması	97
4.3.1.2. Karar ağacı-entropi algoritması	98
4.3.2. Doğrusal destek vektör makineleri algoritması	99
4.3.3. Doğrusal olmayan destek vektör makineleri algoritması	100
4.3.3.1. Doğrusal olmayan destek vektör algoritması-RBF	100
4.3.3.2. Doğrusal olmayan destek vektör algoritması-poly	101
4.3.3.3. Doğrusal olmayan destek vektör algoritması-sigmoid	102
4.3.4. Naive bayes algoritması	103
4.3.5. Lojistik regresyon algoritması	104
4.3.6. K-En yakın komşu algoritması	105
4.3.6.1. K-En yakın komşu algoritması-manhattan	105

4.3.6.2. K-En yakın komşu algoritması-minkowski	106
4.3.7. Rassal orman algoritması	107
4.3.7.1. Rassal orman algoritması-entropi	107
4.3.7.2. Rassal orman algoritması-gini	108
4.3.8. AUCROC değer hesabı	109
4.3.9. Algoritma performansı değerlendirme	110
4.3.10. Yatış-taburcu durum tahmini için çapraz doğrulama	112
4.4. Tahmin Çalışmalarının Sınıflandırma Performanslarının Kıyaslanması	115
5. SONUÇ VE ÖNERİLER.....	119
KAYNAKLAR	125
EKLER	133
ÖZGEÇMİŞ.....	143





KISALTMALAR

ABD	: Amerika Birleşik Devletleri
ATS	: Avustralya triyaj skalası
AUC	: ROC eğrisi altında kalan alan
CART	: Regresyon ağaçlarıyla sınıflandırma
CHAID	: Otomatik ki-kare etkileşim belirleme analizi
CO	: Karbonmonoksit
CTAS	: Kanada triyaj ve kesinlik skalası
DILCA	: Kategorik Nitelikte Uzaktan Eğitim uzaklığı
DM	: Diyabet hastalığı
DSVM	: Doğrusal destek vektör makinesi algoritması
ESI	: Acil durum ciddiyet indeksi
FN	: Yanlış negatif sonuç sayısı
FP	: Yanlış pozitif sonuç sayısı
FRENCH	: Fransa acil hemşireliği hastane ölçeği sınıflandırması
GI	: Gini indeksi
GKS	: Glasgow koma skalası
GUI	: Grafiksel kullanıcı arayüzü
H	: Entropi
HT	: Hipertansiyon
JAAM	: Acil tıp Japon birliği
KA	: Karar ağacı algoritması
KAH	: Koroner arter hastalığı
KBY	: Kronik Böbrek Yetmezliği
KKY	: Konjestif kalp yetmezliği
KNN	: K-En yakın komşu algoritması
KOAH	: Kronik Obstrüktif Akciğer Hastalığı
LOGR	: Lojistik regresyon algoritması
MAP	: Ortalama arteriyel kan basıncı
MSA	: Ortalama mutlak hatası
MSE	: Ortalama kare hatası

MTS	: Manchester triyaj skalası
NB	: Naive bayes algoritması
QUEST	: Hızlı, yansız, verimli istatistiksel ağaç
RBF	: Gauss çekirdek fonksiyonu- Radyal temelli fonksiyon
RF	: Rassal orman algoritması
RMSE	: Kök ortalama kare hatası
ROC	: İşlem karakteristiği eğrisi
SATS	: Güney Afrika triyaj skalası
SpO2	: Oksijen satürasyonu
SS	: Solunum sayısı
SVM	: Destek vektör makinesi algoritması
SVO	: Serebro vasküler hastalık
TN	: Doğru negatif sonuç sayısı
TP	: Doğru pozitif sonuç sayısı
XGBoost	: Aşırı gradyan artırma algoritması
Vd.	: Ve diğerleri

SİMGELER

mmHg	: Milimetre cıva [Birim]
°C	: Santigrat derece [Birim]
mg/dl	: Bir desilitre başına mikrogram [Birim]
C	: Destek vektör algoritması hiper parametresi
μ	: Popülasyon ortalaması
k	: En yakın komşu sayısı
ξ_i	: Doğrusal destek vektör makinelerinde aylak değişken
F	: Model performans ölçütü
F_β	: Model performans ölçütü



TABLO LİSTESİ

Sayfa

Tablo 3.1. Dünyada acil servislerde kullanılan triyaj sınıflandırma seviyeleri.	19
Tablo 3.2. Dünyada tanınmış triyaj sistemleri ve uygulandıkları ülkeler.	19
Tablo 3.3. ATS yetişkin fizyolojik belirtilerin özeti.	21
Tablo 3.4. İkili karışıklık matrisi.	56
Tablo 3.5. Veri setinin istatistiksel analizi.	70
Tablo 4.1. Karar ağacı-gini algoritması sınıflandırma raporu.	79
Tablo 4.2. Karar ağacı-entropi algoritması sınıflandırma raporu.	80
Tablo 4.3. Doğrusal destek vektör algoritması sınıflandırma raporu.	81
Tablo 4.4. SVM-RBF sınıflandırma raporu.	83
Tablo 4.5. SVM-Poly sınıflandırma raporu.	84
Tablo 4.6. SVM-Sigmoid sınıflandırma raporu.	85
Tablo 4.7. Naive bayes algoritması sınıflandırma raporu.	86
Tablo 4.8. Lojistik regresyon algoritması sınıflandırma raporu.	87
Tablo 4.9. KNN-Manhattan algoritması sınıflandırma raporu.	88
Tablo 4.10. KNN-Minkowski algoritması sınıflandırma raporu.	89
Tablo 4.11. Rassal orman-entropi algoritması sınıflandırma raporu.	90
Tablo 4.12. Rassal orman-gini algoritması sınıflandırma raporu.	91
Tablo 4.13. Triyaj alan tahmin modeli AUROC değerleri.	92
Tablo 4.14. Triyaj alan tahmin çalışması genel sınıflandırma raporları.	93
Tablo 4.15. Triyaj alan için çapraz doğrulama doğruluk oranları (%).	95
Tablo 4.16. Triyaj alan için çapraz doğrulama sınıflandırma raporu.	96
Tablo 4.17. Karar ağacı-gini algoritması sınıflandırma raporu.	98
Tablo 4.18. Karar ağacı-entropi algoritması sınıflandırma raporu.	99
Tablo 4.19. Doğrusal destek vektör algoritması sınıflandırma raporu.	100
Tablo 4.20. SVM-RBF sınıflandırma raporu.	101
Tablo 4.21. SVM -Poly sınıflandırma raporu.	102
Tablo 4.22. SVM-Sigmoid sınıflandırma raporu.	103
Tablo 4.23. Naive bayes algoritması sınıflandırma raporu.	104
Tablo 4.24. Lojistik regresyon algoritması sınıflandırma raporu.	105
Tablo 4.25. KNN-Manhattan algoritması sınıflandırma raporu.	106
Tablo 4.26. KNN-Minkowski algoritması sınıflandırma raporu.	107
Tablo 4.27. Rassal orman-entropi algoritması sınıflandırma raporu.	108
Tablo 4.28. Rassal orman-gini algoritması sınıflandırma raporu.	109
Tablo 4.29. Yatış-taburcu durumu tahmin modeli AUROC değerleri.	110
Tablo 4.30. Yatış-taburcu durumu tahmin çalışması genel sınıflandırma raporları.	111
Tablo 4.31. Yatış-taburcu için çapraz doğrulama doğruluk oranları (%).	112
Tablo 4.32. Yatış-taburcu için çapraz doğrulama sonuçları.	113
Tablo 4.33. İki tahmin çalışmasının AUCROC değerleri.	115
Tablo 4.34. Çalışmaların sınıflandırma raporlarının kıyaslanması.	116
Tablo 4.35. Çalışmaların çapraz doğrulama sonuçları (%).	117
Tablo 4.36. Çalışmaların çapraz doğrulama doğruluk değerleri.	118



ŞEKİL LİSTESİ

Sayfa

Şekil 3.1. Avustralya triyaj skalası.	20
Şekil 3.2. Acil servis ciddiye indeks akış şeması.	22
Şekil 3.3. CTAS sınıflandırmaları ve maksimum bekleme süreleri.	23
Şekil 3.4. MTS sınıflandırmaları ve maksimum bekleme süreleri.	25
Şekil 3.5. Makine öğrenimi şeması.	30
Şekil 3.6. Makine öğrenmesi türleri.	31
Şekil 3.7. Karar ağacı yapısı.	35
Şekil 3.8. Karar ağacı örneği.	36
Şekil 3.9. K-En yakın komşu algoritması örneği.	39
Şekil 3.10. K-En yakın komşu algoritması örneği.	40
Şekil 3.11. K-en yakın komşu algoritması işlem adımları.	43
Şekil 3.12. Rastgele orman algoritması mimarisi.	47
Şekil 3.13. Doğrusal destek vektör makineleri.	49
Şekil 3.14. Doğrusal destek vektör makinesi örneği.	49
Şekil 3.15. C parametresinin etkisi.	50
Şekil 3.16. Birbirinden doğru ile ayrılamayan veriler.	51
Şekil 3.17. Çekirdek fonksiyonu sınıflandırma örnekleri.	52
Şekil 3.18. K-katmanlı çapraz doğrulama süreci.	61
Şekil 3.19. Yaş grubu bazında hasta sayıları.	70
Şekil 3.20. Korelasyon ısı haritası.	72
Şekil 3.21. Cinsiyet bazında pasta diyagramı.	72
Şekil 3.22. Hastaların acil servise başvururken sahip olduğu şikayet sayısı durumu.	73
Şekil 3.23. Hastaların şikayet grupları ve adetleri grafiği.	73
Şekil 3.24. Özgeçmiş hastalık sayıları grafiği.	74
Şekil 3.25. Hastaların şikayet grupları ve adetleri grafiği.	74
Şekil 3.26. Acile varış şekli pasta diyagramı.	75
Şekil 3.27. Triage alan sınıflandırması durumu.	75
Şekil 3.28. Yatış yapma ve taburcu olma durumları.	75
Şekil 4.1. Karar ağacı-gini algoritması karışıklık matrisi.	79
Şekil 4.2. Karar ağacı-entropi algoritması karışıklık matrisi.	80
Şekil 4.3. Doğrusal destek vektör makinesi algoritması karışıklık matrisi.	81
Şekil 4.4. Doğrusal olmayan destek vektör makinesi-RBF karışıklık matrisi.	82
Şekil 4.5. Doğrusal olmayan destek vektör makinesi-poly karışıklık matrisi.	83
Şekil 4.6. Doğrusal olmayan destek vektör makinesi-sigmoid karışıklık matrisi.	84
Şekil 4.7. Naive bayes algoritması karışıklık matrisi.	85
Şekil 4.8. Lojistik regresyon algoritması karışıklık matrisi.	86
Şekil 4.9. KNN-Manhattan algoritması karışıklık matrisi.	87
Şekil 4.10. KNN-Minkowski algoritması karışıklık matrisi.	88
Şekil 4.11. Rassal orman algoritması-entropi karışıklık matrisi.	90
Şekil 4.12. Rassal orman algoritması-gini karışıklık matrisi.	91
Şekil 4.13. Karar ağacı-gini algoritması karışıklık matrisi.	97

Şekil 4.14. Karar ağacı-entropi algoritması karışıklık matrisi.	98
Şekil 4.15. Doğrusal destek vektör makinesi algoritması karışıklık matrisi.....	99
Şekil 4.16. Doğrusal olmayan destek vektör makinesi-RBF karışıklık matrisi.	100
Şekil 4.17. Doğrusal olmayan destek vektör makinesi-poly karışıklık matrisi.	101
Şekil 4.18. Doğrusal olmayan destek vektör makinesi-sigmoid karışıklık matrisi..	102
Şekil 4.19. Naive bayes algoritması karışıklık matrisi.....	103
Şekil 4.20. Lojistik regresyon algoritması karışıklık matrisi.	104
Şekil 4.21. KNN-Manhattan algoritması karışıklık matrisi.	105
Şekil 4.22. KNN-Minkowski algoritması karışıklık matrisi.	106
Şekil 4.23. Rassal orman algoritması-entropi karışıklık matrisi.....	108
Şekil 4.24. Rassal orman algoritması-gini karışıklık matrisi.	109



MAKİNE ÖĞRENMESİ ALGORİTMALARI KULLANILARAK SAĞLIK SİSTEMİNDE TAHMİN ÇALIŞMASI

ÖZET

Ülkemizde hastanelerin acil sağlık hizmeti verdiği acil servislerin hasta yoğunluğu, artan nüfusla ve artan salgın hastalık türleriyle beraber gün geçtikçe artmaktadır. Acil servislere başvuran hastaların şikayetlerine ve doktor yorumlarına bakıldığında; acil servise sadece durumu acil nitelikte hastalar başvurmamaktadır. Polikliniklere gidip teşhis ve tedavi süreci görmesi gerekirken, acil servislere başvuran hastalarda olmaktadır. Tüm bu durumlar acil servislerdeki hasta yoğunluğunu, dolayısıyla kalabalıklığı artırmaktadır. Acil servislere başvuran hastaların hepsinin acil bir şekilde, aynı anda sağlık hizmeti alabilmesi pek mümkün olamamaktadır. Durumu kritik olan hastaların hemen hizmet görmesi gerekirken, hayati tehlikesi bulunmayan hastaların farklı sınıflandırılması gerekmektedir. Bu sebeple hastaların sağlık durumları göz önüne alınarak bir önceliklendirme ve sınıflandırma yapılması gerekmektedir. Bu sınıflandırmaya ise tıp dilinde triyaj denilmektedir. Triyajdaki amaç; acil servise başvuran hastaların sağlık durumları esas alınarak triyaj alanlarına karar verilip, gerekli yönlendirmeler yapılmasıdır. Bu şekilde acil sağlık hizmetinin verimliliği ve süreçlerin verimliliği artmaktadır. Aynı zamanda hastanenin kaynakları etkin kullanılıp; bekleme ve hizmet süreleri kısaltılmaktadır. Triyaj kavramı ortaya çıktığı zamanlarda kullanılan tek bir standart varken, zaman içerisinde triyaj standartları sürekli gelişerek sayısı da artmıştır. Dünya geneline bakıldığında üç seviyeli, dört seviyeli ve beş seviyeli farklı triyaj sistemlerinin olduğu görülmektedir. Ülkemizde ise kırmızı, sarı, yeşil ana alan olmak üzere, kendi içerisinde bölümlere ayrılan bir triyaj sistemi mevcuttur. Çalışmada amaç; gerçek hayattan alınmış gerçek veriler ile makine öğrenmesi sınıflandırma algoritmalarını eğitip, triyaj alan tahmin modelini ve yatış-taburcu tahmin modelini oluşturup, bu modelleri test verileriyle test ettikten sonra elde edilen tahmin sonuçları göz önüne alınarak kullanılan sınıflandırma algoritmalarının performansını kıyaslamaktır. Bununla birlikte elde edilen makine öğrenmesi modelleri ile gerçek hayattaki problemlere çözüm ışığı olabilmektir. Bunun için, Türkiye'deki bir hastanenin acil servisinden elde edilen gerçek triyaj verileriyle çalışılmak istenmiş ve Kartal Doktor Lütfü Kırdar Şehir Hastanesi Acil Tıp Kliniği ile görüşülmüştür. Çalışmaya başlamadan önce Sakarya Üniversitesi'nden etik kurul izni alınmıştır. Bu çalışmada; Kartal Doktor Lütfü Kırdar Şehir Hastanesi Acil Tıp Kliniği'ne başvuran hastaların yaş, cinsiyet, yaşamsal bulgular, hastalık şikayetleri, önceden sahip olmuş ya da şu an sahip olduğu hastalıkları, acil servise varış şekli, Glasgow koma skalaları gibi verileri toplanmış olup, veriler manuel olarak toplanmıştır. Veri seti 2023 yılı Ocak-Nisan ayları arasında hastaneye başvuran 18 yaşından büyük hastaları kapsamaktadır. Direkt gebelik ile ilgili vakalar veri setine dahil edilmemiştir. Sözel verilerin karşılıklarını elde etmek uzun sürmesi sebebiyle veri toplama süreci 2023 yılı Ocak-Eylül aylarını kapsamaktadır. Çalışmada toplam 3.000 hastanın 56 çeşit girdi bilgisi kullanılmış olup 2 farklı tahmin çalışması yapılmıştır. İlk olarak hastaların triyaj alan sınıflandırmaları yeşil, sarı, kırmızı olarak tahmin edilmiş; ikinci çalışma olarak ise yatış-taburcu durumu yani hastaların sağlık hizmeti aldıktan sonra direkt taburcu

mu olacağı yoksa yatış mı yapacağı tahmin edilmiştir. Tahmin çalışmaları için toplanan bu veriler, % 80 eğitim verisi ve %20 test verisi olacak şekilde ikiye ayrılmıştır. Eğitim verileriyle makine öğrenmesi sınıflandırma algoritmalarıyla 12'şer tane eğitilmiş model elde edilmiştir. Çalışmada; gini ve entropi kriteriyle karar ağacı algoritması, doğrusal destek vektör makinesi algoritması, RBF-poly-sigmoid çekirdekler ile doğrusal olmayan destek vektör makinesi algoritması, naive bayes algoritması, lojistik regresyon algoritması, manhattan ve minkowski uzaklığıyla k-en yakın komşu algoritması, gini ve entropi kriterleriyle rassal orman algoritması kullanılmıştır. Test verileriyle eğitilmiş modeller ile tahmin çalışması yapıp, karışıklık matrisleri elde edilmiştir. Karışıklık matrisleri ile performans ölçütleri olan doğruluk, kesinlik, duyarlılık, F skor, AUCROC değerleri hesaplanmıştır. Ardından her iki tahmin çalışması için de çapraz doğrulama işlemi yapılmıştır. Tüm bu elde edilen sonuçlar, öncelikle kendi tahmin çalışması içerisinde kıyaslanmış ardından iki tahmin çalışmasından elde edilen sonuçlar birlikte değerlendirilmiştir. Değerlendirmeler sonucunda triyaj alan tahmini çalışmasında; %97,79 en yüksek AUROC değeri karar ağacı-entropi algoritmasıyla, %96 en yüksek doğruluk oranı rassal orman gini ve entropi algoritmalarıyla elde edilmiştir. Trijaj alan sınıflandırmasında çapraz doğrulama işlemiyle % 97,05 en yüksek doğruluk oranı karar ağacı-gini algoritmasıyla elde edilmiştir. Yatış-taburcu olma durumu tahmin çalışmasında ise; %68,26 en yüksek AUROC değeri rassal orman-gini algoritmasıyla ve yine %90,8 en yüksek doğruluk oranı rassal orman gini algoritmalarıyla elde edilmiştir. Yatış-taburcu olma durumu sınıflandırmasında çapraz doğrulama işlemiyle % 89,12 en yüksek doğruluk oranı rassal orman-gini algoritmasıyla elde edilmiştir. Karar ağacı algoritmasıyla triyaj alan sınıflandırmasında yüksek performans oranları elde edilebildiği, rassal orman-gini algoritmasıyla yatış-taburcu durum tahmininde en yüksek değerler elde edildiği görülmüştür. Sonuç olarak, triyaj sürecinde makine öğrenmesi algoritmalarının kullanılması durumunda yüksek performans oranlarıyla verimliliği artıracabilecek başarılı triyajlar yapılabilmektedir. Makine öğrenmesi algoritmalarının triyaj süreçlerinde kullanılmasıyla; daha kaliteli triyaj sınıflandırması yapılabileceği, acil servis süreçlerinin daha verimli yürütülebileceği, hastane kaynaklarının daha verimli ve planlı kullanılabileceği, hastaların bekleme sürelerinin azaltılabileceği, hasta memnuniyetinin artırılabilceği öngörülmektedir.

PREDICTION STUDY USING MACHINE LEARNING ALGORITHMS IN HEALTH SYSTEM

SUMMARY

In our country, the patient density of emergency departments, where hospitals provide emergency health services, is increasing day by day with the increasing population and increasing epidemic disease types. Looking at the complaints and doctor's comments of patients applying to emergency services; Only patients with urgent conditions do not apply to the emergency department. Patients who need to go to outpatient clinics for diagnosis and treatment are also admitted to emergency departments. All these situations increase the patient density and therefore crowding in emergency departments. It is not possible for all patients applying to emergency departments to receive emergency health care at the same time. While patients in critical condition should receive service immediately, patients whose condition is not life-threatening should be classified differently. For this reason, patients need to be prioritized and classified according to their health status. This classification is called triage in medical language. Triage practices in emergency departments are an integral part of hospital health services. The purpose of triage is the aim is to decide on triage areas and provide necessary guidance based on the health status of patients applying to the emergency department. In this way, the efficiency of emergency health care and the efficiency of the processes increase. At the same time, the resources of the hospital are used effectively; waiting and service times are shortened. Triage is also an important practice to determine the care priority of patients. It aims to have the right patient in the right place at the right time. Due to the high number of patients applying to the emergency department, many critically ill patients in the emergency department are at risk of experiencing poor health outcomes due to the delay in triage decisions. For this reason, it is very important to decide which patient should be seen first. Triage systems are used to decide on the order of treatment and determine urgency situations when a large number of patients apply. Inconsistent triage decisions or misclassifications lead to an increase in the occurrence and incidence of patient deaths. High triage accuracy leads to better quality emergency services. While there was only one standard used when the concept of triage emerged, the number of triage standards has constantly improved and increased over time. When we look around the world, it can be seen that there are different triage systems with three levels, four levels and five levels. In our country, there is a triage system divided into sections, including red, yellow and green main areas. The aim of the study is to train machine learning classification algorithms with real data from real life, create a triage area prediction model and hospitalization-discharge prediction model, and compare the performance of the classification algorithms used by taking into account the prediction results obtained after testing these models with test data. In addition, with the machine learning models obtained, it is aimed to be a solution light for real-life problems. In this study, data such as age, gender, vital signs, disease complaints, previous or current diseases, mode of arrival to the emergency department, Glasgow coma scales were used, and data were collected manually from patients who applied to the Emergency Medicine Clinic of Kartal Dr.

Lütfü Kırdar City Hospital. In this study, we wanted to work with real triage data obtained from the emergency departments of hospitals in Turkey. For this purpose, the Emergency Medicine Clinic of Kartal Dr. Lütfü Kırdar City Hospital was contacted. Before starting the study, ethics committee approval was obtained from Sakarya University. The dataset includes patients older than 18 years of age admitted to hospital between January and April 2023. Cases directly related to pregnancy are not included in the dataset. The data collection period covers January-September 2023. In the study, 56 types of input information from a total of 3,000 patients were used and 2 different prediction studies were conducted. Firstly, the triage area classifications of the patients were predicted as green, yellow, red; and secondly, the hospitalization-discharge status, that is, whether the patients will be directly discharged or hospitalized after receiving health services, was predicted. These data collected for prediction studies were divided into two parts, 80% training data and 20% test data. With the training data, 12 trained models were obtained with machine learning classification algorithms. In the study, decision tree algorithm with gini and entropy criteria, linear support vector machine algorithm, nonlinear support vector machine algorithm with RBF-poly-sigmoid kernels, naive bayes algorithm, logistic regression algorithm, k-nearest neighbor algorithm with manhattan and minkowski distance, random forest algorithm with gini and entropy criteria were used. Confusion matrices were obtained by performing prediction studies with models trained with test data. Confusion matrices and performance measures of accuracy, precision, recall, F score and AUCROC values were calculated. Then, k-fold cross-validation was performed for both prediction studies. Cross-validated model building enables the development of machine learning applications that work with high accuracy or performance. Different cross-validation techniques make it possible to predict the performance of a model without compromising the test split, eliminating problems that can be caused by an unbalanced data split. In this way, classification processes provide more consistent results. All these results are firstly compared within their own prediction study and then the results obtained from the two prediction studies are evaluated together. As a result of the evaluations, in the triage area prediction study; The highest AUROC value of 97.79% was obtained with the decision tree-entropy algorithm, and the highest accuracy rate of 96% was obtained with the random forest and entropy algorithms. In triage area classification, the highest accuracy rate of 97.05% was obtained with the decision tree-gini algorithm through the k-fold cross-validation process. In the hospitalization-discharge prediction study; The highest AUROC value of 68.26% was obtained with the random forest-gini algorithm and the highest accuracy rate of 90.8% was obtained with the random forest-gini algorithms. In the classification of hospitalization-discharge status, the highest accuracy rate of 89.12% was obtained with the random forest-gini algorithm with k-fold cross-validation. It was observed that high performance rates could be achieved in triage area classification with the decision tree algorithm, and the highest values were obtained in hospitalization-discharge status prediction with the random forest-gini algorithm. As a result, if machine learning algorithms are used in the triage process, successful triages can be performed that can increase efficiency with high performance rates. By using machine learning algorithms in triage processes; it is predicted that better quality triage classification can be made, emergency service processes can be carried out more efficiently, hospital resources can be used more efficiently and planned, waiting times of patients can be reduced, and patient satisfaction can be increased. With the hospitalization prediction study, it is predicted that the doctors will know whether the patient who applies to the emergency service will be hospitalized or discharged before the patient is admitted to

the emergency service, and according to this density, it is predicted that the planning of beds, medicines, health equipment, all the resources and doctors will be more accurate and more efficient.





1. GİRİŞ

Triyaj kelimesi Fransızca kökenli bir kelime olup sınıflandırma anlamına gelmektedir. Triyaj terimi tarih boyunca farklı amaçlarda kullanılmıştır. Bu çalışmada ise sağlık alanında yapılan triyaj ile ilgilenilmiştir. Sağlık triyajı uygulaması, savaşlarda insanların yaralanması ve müdahalelerde zorlanması sonunda bir gereklilik olarak ortaya çıkmıştır. Askerler verimli bir acil sağlık bakım hizmeti alamadı için fazla can kayıpları yaşanmış ve 18.yüzyıl ile düzenli acil sağlık bakım hizmeti alınabilmesi için sistematik olan ilk triyaj kuralları oluşturulmuştur (Almalahy, 2022).

Triyaj; hastane öncesi, afet alanlarında, savaşlarda ve acil servisler gibi alanlarda uygulanan bir standarttır. Acil servislerde uygulanan triyaj özetle; acil servise başvuran hastaların doğru zamanda, doğru sırada ve doğru şekilde sağlık bakım hizmeti alabilmeleri için yapılan sınıflandırma işlemidir.

Triyajın yapılması hem acil servise başvuran hastalar açısından hem de hastane ve çalışan personel açısından önemlidir. Acil servise başvuran hasta için önemine bakıldığında triyajın yapılmasındaki amaçlar (Şimşek, 2018);

- Hastanın hayatını korumak,
- Acil sağlık hizmeti sonrasında gerçekleşme ihtimali olan kayıpları veya ortaya çıkabilecek hastalıkların önüne geçmek,
- Kritik durumda olan hastalar ile kritik durumda olmayan hastaları sınıflandırarak bekleme sürelerini azaltarak, doğru hastaya doğru zamanda hizmet verebilmek şeklinde örneklendirilebilir.

Acil sağlık hizmeti veren hastane için önemine bakıldığında triyajın yapılmasındaki amaçlar (Şimşek, 2018);

- Hastanenin kısıtlı olan kaynaklarını verimli kullanmak,
- Acil servis yoğunluğunu azaltmak,
- Hastane personelinin verimli bir şekilde çalışacağı ortam oluşturmak şeklinde örneklenebilir.

Triyajın amaçları göz önüne alındığında triyaj uygulamaları acil servislerin en önemli parçalarından biridir. Acil servislerde her hastanın müdahalesi en hızlı şekilde yapılmalıdır. Hasta yoğunluğundan, kaynak ve personel yetersizliğinden, triyajın yapılmamasından ya da verimli bir sınıflandırma işlemi yapılmamasından dolayı bu süreler uzayabilmektedir.

Hasta yoğunluğu yani acil servislere başvuran hastaların fazla olması acil servislerin en önemli problemlerinden biridir. Bu problem insan popülasyonunun artmasıyla birlikte gün geçtikçe artarak devam etmektedir. Acil servislere başvuran hasta sayıları fazla olduğunda yaşanabilecek yoğunlukta triyajın yapılmadığı ya da verimli yapılamadığı durumlarda; kritik hastalara doğru zamanda müdahale edilemeyebilir, hayati tehlikesi olan hastalara geç müdahale ile can kayıplı sonuçlar yaşanabilir, zaman ve personel verimli kullanılamayarak çalışanlar stres altında kalabilir vb. olumsuz sonuçlarla karşı karşıya kalınabilir. Bu ve benzeri sebeplerden dolayı kimler acil servisteki hizmete daha acil kimlerin beklemeye daha toleranslı olabileceğine verimli şekilde karar verilebiliyor olmalıdır.

Özetle; hasta yoğunluğunun olduğu zamanlarda acil servislerdeki ilk iş olarak yapılan hasta triyajının verimli yapılabilmesi çok büyük bir önem arz etmektedir. Bu sebeple triyajın hızlı, doğru ve en verimli şekilde yapılıyor olması gerekmektedir.

1.1. Tezin Kapsamı

Bu çalışma, Kartal Doktor Lütfü Kırdar Şehir Hastanesi Acil Tıp Kliniğinden manuel olarak toplanan triyaj verileri kullanılarak makine öğrenmesi algoritmalarıyla öğrenilmiş model oluşturulmasını, sonrasında tahmin çalışmasının yapılmasını ve algoritmaların performanslarının kıyaslanması sürecini kapsamaktadır. Çalışmada kullanılan veriler; 2023 yılının Ocak – Nisan ayları aralığında hastaneyi ziyaret eden, 18 yaşından büyük hastaların verilerini içermektedir. Veri toplama süreci ise daha uzun süreyi kapsamakla birlikte; 2023 yılının Ocak ayında başlamış olup, 2023 yılının Eylül ayında tamamlanmıştır.

Çalışma, toplam 3.000 hastanın verisi içermektedir. Hastaların yaşamsal bulguları, sözlü şikayetleri, özgeçmişlerindeki rahatsızlıkları, acil servise varış şekli gibi birçok bilgi toplanmıştır. Toplam girdi adedi 56 adettir. Toplanan verilerle ilgili ayrıntılı bilgiler çalışmanın içerisinde detaylandırılmıştır.

1.2. Çalışmanın Konusu

Triyaj verilerinin makine öğrenmesi algoritmalarıyla kullanılarak eğitilmiş model elde edilip, tahmin çalışması ve makine öğrenmelerinin tahmin etmedeki performans kriterleri hesaplanarak kıyaslanma çalışması yapılmıştır.

1.3. Çalışmanın Amacı

Acil servis, hastanenin tanısal müdahalelerin hızlı ve verimli gerçekleştirilmesi gereken bir bölümdür. Acil servis bölümüne başvuran kişi sayısı arttıkça, süreçlerin iyileştirilmesine ihtiyaç daha çok artmaktadır. Triyaj tarihine bakıldığında triyaj sürecinin iyileştirilmesiyle ilgili birçok çalışma yapılmıştır. İlk başlarda tek bir triyaj standardı olup sonrasında geliştirilmeler gerekli görülüp, bu süreçler iyileştirilerek birden fazla triyaj standartları oluşturulmuştur. Ülkelerin kullandığı triyaj sistemleri de farklılık göstermektedir. Ülkeler triyaj sistemlerini yeterli görmeyip, geliştirmeler yapmış ve yeni triyaj türleri ortaya çıkmıştır. Bu yeni triyaj türlerinin ortaya çıkmasındaki amaç triyaj süreçlerinin yetersizliği ve iyileştirilme gereği duyulmasıdır.

Tıp alanında yapılan iyileştirmelerinin yanı sıra günümüzde gelişmekte olan yapay zeka kümeleri kullanılarak bu süreçlerin geliştirilebileceği düşünülmektedir.

Bu çalışmadaki amaç; acil servislerde yaşanan hasta yoğunluğu durumunu iyileştirmek, hastaların daha iyi bir hizmet almalarını sağlamak, acil servisteki bekleme sürelerini azaltmak, hastalara doğru zamanda ve doğru yerde hizmet verebilme durumunu yani acil sağlık hizmetinin verimliliğini artırmak, hastanenin kısıtlı kaynaklarının verimli kullanılmasını sağlamak amacıyla triyaj sürecinin iyileştirilmesine, geliştirilmesine hizmet edebilmektir.

Acil servise gelen triyajının yapılıp, hangi alanda hizmet göreceğinin tahminin yapılmasını; acil servise başvuran hastanın hizmet sonucunda yatış alıp almayacağının tahmin edilmesini amaçlanmaktadır.

1.4. Çalışmanın Önemi

Triyaj işlemi, acil servisteki birçok süreci etkilemektedir. Doğru ve verimli şekilde yapılan triyaj işlemleri sonuçları aşağıdaki gibi örneklendirilebilir.

- Hastanedeki hasta yoğunluğu iyi bir şekilde yönetilmiş olur.
- Hastaların bekleme ve müdahale süreleri azalmış olur.

- Hastalar aciliyet durumuna göre sınıflandırılmış olacağından hastalar doğru zamanda tedavi edilmiş olur.
- Kritik durumda olan hastaya zamanında müdahale edilmiş olur.
- Hastanenin kaynakları verimli kullanılmış olur.
- Hasta yoğunluğu verimli şekilde yönetilebildiği için çalışan personelin performansı ve çalışma ortamı iyileştirilmiş olur.

Triyaj işlemi bu denli önemliken, tarih boyunca iyi bir triyaj yapılabilmesi için triyaj sistemleri geliştirilmeye devam ederken, triyaj sürecinde makine öğrenmesi algoritmalarının kullanılarak tahmin çalışması yapılması da süreci dijitalleştirmekle beraber verimliliği de artırması çok büyük önem arz etmektedir. Makine öğrenmesi algoritmaları kullanmak; acil tıp süreçlerinde insan hatalarını, harcamaları, bekleme sürelerini azaltarak, hizmet sağlama hızını artırarak verimliliği artıracaktır.

1.5. Çalışmanın Yöntemi

Triyajdaki alan tahmini ve yatış tahmini yapılırken makine öğrenmesi algoritmalarından faydalanılmıştır.

Makine öğrenmesi, yapay zeka kavramının bir alt kümesidir. Her ne kadar yapay zeka ve makine öğrenmesi kavramları karıştırılsa da aynı anlama gelmemektedirler. Makine öğrenmesi, algoritmaların veriler ile eğitilerek öğrenme yeteneği kazanması olarak tanımlanabilir.

Makine öğrenmesi birçok alanda kullanılabilir. Bu çalışmada sağlık alanındaki triyaj verileri üzerinde kullanılmıştır. Çalışmada makine öğrenmesi sınıflandırma algoritmaları kullanılmıştır. Kullanılan sınıflandırma algoritmaları ise aşağıdaki gibidir.

- Karar ağacı algoritması
- Doğrusal destek vektör makinesi algoritması
- Doğrusal olmayan destek vektör makinesi algoritmaları
- Naive bayes algoritması
- Lojistik regresyon algoritması
- K-En yakın komşu algoritması
- Rassal orman algoritması

Bu algoritmalar 3.000 adet hasta, 56 adet sütun girdisi olan gerçek triyaj verileriyle eğitildikten sonra öğrenilmiş modeller elde edilir. Bu modeller kullanılarak test verileri ile tahmin çalışması yapılmıştır. Elde edilen tahmin verileriyle model performansları hesaplanmış ve algoritmaların performansları kıyaslanmıştır.

Özetle bu çalışma; makine öğrenmesi algoritmalarını gerçek hayattan alınmış triyaj verileri ve hastaların sözlü şikayetleriyle birlikte kullanarak, triyaj verimliliğini artırıp, acil servis süreçlerindeki hizmet kalitesinin artırılmasına katkı sağlaması öngörülmektedir.



2. LİTERATÜR ARAŞTIRMASI

Triyaj ve makine öğrenmesi sınıflandırma algoritmalarıyla ilgili literatür araştırması çalışması; Yükseköğretim Kurulu, Science Direct, Sakarya Üniversitesi e-kütüphane, Google Scholar vd. veri tabanları sorgulanmıştır. Aşağıdaki literatür araştırması bu kaynaklar aracılığı ile gerçekleştirilmiştir.

Mueller ve ark. (2022), acil tıpta yapay zeka kullanımlarını derlemek adına bir çalışma yapmışlardır. Çalışmalarında; karar ağaçları, rastgele orman algoritması, topluluk öğrenme, naive bayes, destek vektör (SVM), k-en yakın komşu algoritmalarından bahsedilmiştir. Acil serviste makine öğrenmesi modelleri kullanılarak daha fazla doğruluk elde edilebileceği, triyajı kolaylaştırabileceğinden bahsedilmiştir. Tıbbi durumların aciliyetinin tahmin edilmesinin yanı sıra hastalıklara yönelik tarama araçları geliştirmek için de kullanılabileceğinden bahsedilmiştir. Bu çalışmaya elde edilebilecek başarılar şöyledir: derin öğrenme modelleri, tıbbi görüntüleme alanında kullanılarak yüksek hassasiyetler ile hayati tehlikeyi hızlı bir şekilde tespit etmesine yardımcı olabilecektir. Acil servislerde makine öğrenmesinin kullanılmasıyla birlikte hasta ihtiyaçları daha iyi karşılanabilecek, maliyet azaltılabilecek, hasta sonuçlarının iyileştirilebilecektir. Acil servislerde aşırı kalabalığın giderek artan bir sorun olduğu için bu durumu iyileştirebileceği, hasta bakım kalitesinin artırabilecektir. Son olarak çalışmada yapay zeka tahminleri ile uzman değerlendirmeleri arasındaki farklar ve hataların da üzerinde durulmuştur (Mueller ve ark., 2022).

Bayrak ve ark. (2022), tarafından yapılan çalışmada, Kaggle platformundan elde ettikleri göğüs kanseri verileri üzerinde k-en yakın komşu, destek vektör makinaları, naive bayes, karar ağaçları ve yapay sinir ağları olmak üzere 5 farklı algoritma ile çalışılmıştır. Birinci veri setinde %98,2 ve ikinci veri setinde %93,8 doğruluk oranıyla yapay sinir ağları algoritmasında en yüksek doğruluk oranını elde etmişler (Bayrak ve ark., 2022).

Inokuchi ve ark. (2022), tarafından yapılan çalışmada; hastane öncesi telefon triyajındaki verimliliği artırmak amacıyla makine öğrenmesi modelleri kullanılarak, Japonya'daki mesai sonrası ev çağrısı hizmeti veri seti ile retrospektif bir kohort çalışması yapılmıştır. Destek vektör makinesi, lojistik regresyon, rastgele orman,

gradyan destekli karar ağacı ve derin sinir ağına dayalı 5 tahmin modeli geliştirilmiştir. Model girdileri; telefon triyaj seviyesi, yaş, cinsiyet, 80 başlıca şikayet kategorisi ve 10 komorbite dahil olmak üzere mevcut telefon tabanlı veriler kullanılmıştır. Çalışma sonucunda, rastgele orman ve gradyan destekli karar ağacı algoritmalarının daha iyi performans gösterdiği elde edilmiştir. Böylece süreci iyileştirmeyi ve erken müdahaleyi kolaylaştırabileceği gösterilmiştir (Inokuchi ve ark., 2022).

Jiang ve ark. (2021); acil servis triyajında hastaların önceliklendirilmesi, yüksek riskli hastaların belirlenmesi ve sınırlı kaynakların verimli bir şekilde kullanabilmesi amacıyla makine öğrenmesi algoritmalarını kullanarak model geliştirmişlerdir. Çalışmada; acil servise başvuran kardiyovasküler hastalık şüphesi olan hastalardan elde edilen veriler (demografik bilgiler, yaşamsal bulgular, kan şekeri ve diğer mevcut triyaj skorları) ile triyaj seviyelerine karar vermek amacıyla makine öğrenmesi algoritmaları kullanılmıştır. Kullanılan algoritmalar; çoklu lojistik regresyon, rastgele orman, XGBoost algoritmaları ve gradyan destekli karar ağacı algoritması. Sonrasında da algoritmaların performansları karşılaştırılmıştır. Çalışma sonucunda dört makine öğrenimi modelinin de triyaj için iyi bir ayırt edici yeteneğe sahip olduğu ve XGBoost algoritmasının diğer modellere göre hafif bir avantaj gösterdiği sonucuna varılmıştır. Sınırlı kaynakların verimliliğini artırmak için bir strateji olarak düşük riskli hastaların acil servis triyajı için kullanılabileceği belirtilmiştir (Jiang ve ark., 2021).

Coşar ve Deniz (2021), çalışmalarında lojistik regresyon, k-en yakın komşu ve rastgele orman algoritmalarını kullanarak kalp hastalıklarının tespit edilmesini amaçlamışlardır. Bunun için, Kaggle platformundaki hazır kalp yetmezliği veri seti kullanılmıştır. Veri setlerinde 11 farklı girdi bulunmaktadır. Yapılan çalışmada, Rastgele orman algoritmasının %85 doğruluk sonucu ile en iyi performansı verdiği sonucuna ulaşılmıştır (Coşar ve Deniz, 2021).

Başer ve ark. (2021), makine öğrenmesi algoritmalarını kullanarak diyabet hastalığının sınıflandırılması konusunda çalışmışlardır. Çalışmada, ABD'de bulunan 130 hastanedeki 70.000 kayıttan elde edilen veri seti ile çalışarak bireylerin diyabet durumlarının sınıflandırılması ve sonrasında algoritmaların performanslarının kıyaslanması amaçlanmıştır. Kullanılan algoritmalar; karar ağaçları, k-en yakın komşu, lojistik regresyon, naive bayes ve rastgele orman algoritmalarıdır. Algoritmaların performansları kıyaslandığında rastgele orman algoritmasının daha doğru sınıflandırma yaptığı sonucuna varılmıştır (Başer ve ark., 2021).

Saygın ve Baykara (2021), karaciğer yetmezliği teşhisinde özellik seçimi kullanmıştır. Makine öğrenmesi algoritmalarının performanslarını incelemiştir. Çalışmasında özellik seçimi için ileri yönde seçim yöntemini kullanmıştır. Bu şekilde veri setinde kullanacağı öz nitelikleri tespit etmiştir. SVM, karar ağaçları, lojistik regresyon ve hafif gradyan güçlendirme makinesi gibi algoritmaları kullanılmıştır. Hafif gradyan güçlendirme makinesi sınıflandırıcısı algoritması %82,12 en yüksek başarıya ulaşmıştır (Saygın ve Baykara, 2021).

Ivanov ve ark. (2021); triyaj kaynak ihtiyacının belirlenmesi, triyaj yoğunluğunun etkisinin azaltılması için retrospektif bir çalışma yapmışlardır. Bu çalışmada, doğru ESI modelleri üretmek için geçmiş verileriyle doğal dil işleme ve XGBoost algoritmaları kullanıp kullanılamayacağını belirlemeyi amaçlamışlardır. Model, Acil Durum Ciddiyet Endeksi (ESI) standardını kılavuz olarak kullanarak çalışma klinisyenleri tarafından keskinliği doğru şekilde atanan rastgele bir karşılaşma örneğine karşı test edilmiştir. Model, %75,7 oranla ESI sınıflandırmalarını doğru tahmin etmiştir. Elde edilen modelin, hemşire triyajına göre %26,9 oranla daha doğru atamalar sağladığı görülmüştür (Ivanov ve ark., 2021).

Pak ve ark. (2021), triyajdan sonra bekleme odasına atanan düşük akut hastaların bekleme süresini tahmin etmek amacıyla modeller oluşturmayı amaçlamıştır. Çalışmada, bekleme süresi tahminlerini önemli ölçüde iyileştiren makine öğrenimi algoritmalarının kullanılması önerilmektedir. Lasso regresyon, ridge regresyon, rastgele orman algoritmaları kullanılmıştır. Çalışmada lasso regresyonun büyük ölçüde daha iyi doğruluk oranları verdiği görülmüştür (Pak ve ark., 2021).

Mišić ve ark. (2020) tarafından yapılan çalışmada, acil servisteki ameliyat sonrası hastaların tekrar hastaneye kabullerini doğru bir şekilde tahmin edebilesi için makine öğrenimi yaklaşımları kullanılmıştır. Yeniden kabul ihtimalinin, hastaneden taburcu olmadan önce doğru bir şekilde tahmin edilebileceğini öne sürülmüştür. Çalışmada, ameliyattan sonraki 30 gün içerisinde acil servisten kaynaklanan gelecekte hastaneye yeniden yatış olup olmama durumunu çalışılmıştır. Çalışma sonucunda iyi performansı %86 AUC ile gradyanla güçlendirilmiş karar ağacı algoritmasıyla elde edilmiştir. Ameliyat sonrası yeniden kabulü tahmin etmeye yönelik bir makine öğrenimi yaklaşımı, acil servis aracılığıyla 30 günlük yeniden kabulleri doğru bir şekilde tahmin etmek için hastaneye özel modeller üretebileceği sonucuna varılmıştır (Mišić ve ark., 2020).

Shafaf ve Malek (2019); acil tıp triyajında makine öğrenmesi algoritmaları kullanarak hastaların hastalığının, hastaneye kabul durumunun, taburcu olma ve ölüm oranının tahmin edilebilmesi üzerine çalışmışlardır. ESI model triyajda tahmin sonuçlarının daha kolay analiz edilebilmesi için yüksek riskli hastayı tahmin edebilen ve hastaları uygun şekilde ayırabilen; makine öğrenimi tekniklerini kullanan elektronik triyaja dayalı bir değerlendirme yöntemi önermektedirler. Yaptıkları makine öğrenme tabanlı çalışma sonucunda kullandıkları algoritmaları kıyaslamışlardır. Özetle; yapay zeka ve makine öğrenmesi tekniklerinin yetenekleri acil tıpta; hastalık tahmini, yatış veya taburcu tahmini, hasta triyajı gibi bazı önemli konularda kullanılabileceğini belirtmişlerdir. Yüksek riskli hastalıklar ve influenza gibi bulaşıcı hastalıkların erken öngörüsü ve tanısı ile acil serviste çoklu hastalık ilerlemesi komplikasyonlarını önlemek için gerekli müdahaleler daha hızlı gerçekleştirileceğinden bahsedilmiştir. Bu bağlamda, genellikle %73'ten %92'ye kadar yüksek doğruluk gösteren lojistik regresyon, bayes ağı, derin öğrenme vb. gibi farklı makine öğrenme algoritmaları ile çalışılmıştır. Ek olarak, bu algoritmalar ve diğer yöntemler hasta kabulünün tahmin edilmesinde ve hasta triyajının iyileştirilmesinde yardımcı olabileceği sonucuna varılmıştır (Shafaf ve Malek, 2019).

Dilwani (2019), çalışmasında makine öğrenmesi algoritmaları ile acil triyaj çalışması ve yatış tahmin çalışması yapmıştır. Hazır veri setinde 60.735 hastanın MTS veri setini kullanarak ilk bölümde hastaneye yatış tahmini yapmış, ikinci bölümde triyaj seviyelerini belirlemiştir. Girdi olarak; yaş, hayati fonksiyonlar, hastaneye yatış yapma durumu, önemli hastalık şikayetleri ve triyaj sınıf bilgilerini kullanmıştır. Tahmin çalışmasında kullanılan algoritmalar; lojistik regresyon, rassal orman, XGBoost, KNN, otomatik kodlayıcı ve derin sinir ağı algoritmaları. Makine öğrenme algoritmalarının acilde hasta yatış ve acil triyaj durumlarını önemli ölçüde tahmin edebildiği sonucuna ulaşmıştır. (Dilwani, 2019).

Raita ve ark. (2019), makine öğrenmesi algoritmalarını kullanarak kritik hastaları stabil hastalardan doğru bir şekilde ayıran ve önceliklendiren acil servis modeli kurmayı amaçlamıştır. 2007'den 2015'e kadar Ulusal Hastane ve Ambulatuvar Tıbbi Bakım Araştırması acil servisteki yetişkin hastaların verileri kullanılmıştır. Değişkenlere örnek olarak; demografik bilgiler, triyaj hayati belirtileri, ana şikayetler, komorbiditeler ve beş seviyeli triyaj modeli olan ESI kullanılmıştır. Veri seti; %70'e %30 eğitim ve test verisi olacak şekilde ayrılarak; kement regresyonu, rastgele orman,

gradyan destekli karar ağacı ve derin sinir ağı algoritmaları kullanılmıştır. Çalışmada, kritik bakım ve hastaneye yatış tahmini yapılmıştır. 135.470 uygun acil servis ziyaretinin %2,1'inde kritik bakım sonucu ve %16,2'sinde hastaneye yatış sonucu elde edilmiştir. Geleneksel yaklaşımla karşılaştırıldığında, makine öğrenimi modelleri kritik bakım ve hastaneye yatış sonuçlarını tahmin etmede üstün bir performans göstermiştir. Modern makine öğrenimi modellerinin uygulanması, klinisyenlerin triyaj karar verme sürecini geliştirebilir ve böylece daha iyi klinik bakım ve optimum kaynak kullanımı elde edilebileceği sonucuna ulaşılmıştır. Kritik bakım sonucu tahmininde, dört makine öğrenimi modeli de referans modelden daha iyi performans göstermiş ve ESI triyaj seviyesi 3 ila 5'te daha az yetersiz triyajlı hasta tespit edilmiştir. Benzer şekilde; hastaneye yatış sonucu tahmininde tüm makine öğrenimi modelleri, ESI triyaj seviyeleri 1 ila 3'te daha az aşırı triyaj ile referans modelden daha iyi performans göstermiştir (Raita ve ark., 2019).

Parker ve ark. (2019), triyaj sırasında bir hastanın hastaneye yatış ihtiyacını tahmin edebilecek bir model oluşturarak acil servislerin kalabalıklığının ve hastaneye yatış için bekleme sürelerinin azaltılmasını amaçlamıştır. Çalışmada, Singapur'daki şehir hastanesine 10 yıl boyunca yapılan acil servis ziyaretlerinin elektronik kayıtları kullanılmıştır. Hastaların hastaneye yatış ihtimalini tahmin eden bir lojistik regresyon modeli geliştirilmiş ve model performansı değerlendirilmiştir. Model, 0,825 AUC ile iyi bir performans göstermiştir (Parker ve ark., 2019).

Kim ve ark. (2020), septik şok için bir önceliklendirme tarama aracı olarak makine öğrenmesini kullanılmıştır. Organ yetmezliği değerlendirmesi yapılmıştır. Çalışmaya, enfeksiyon şüphesiyle acil servise başvuran yetişkin 20 yaşından büyük yetişkin 49.560 hastanın verileri dahil edilmiştir. Çalışmada kullanılan veriler; demografik özellikler, hayati belirtiler, bilinç düzeyi, ana şikayetler ve ilk kan testi sonuçları olarak sıralanmıştır. Destek vektör makinesi, gradyan artırma, rastgele orman, ridge regresyon, lasso regresyon, çok değişkenli adaptif regresyon gibi topluluk öğrenme algoritmaları kullanılmıştır. Tüm topluluk makine öğrenmesi algoritmaları 0,883 ile 0,929 arasında değişen AUC değerleriyle yüksek performans göstermiştir (Kim ve ark., 2020).

Patel ve ark. (2018), pediatrik astımın tahmin edilmesiyle ilgili çalışmışlardır. Çalışmalarındaki amaç, pediatrik astımda hastane düzeyinde bakım ihtiyacının erken tahmini için triyaj sırasında mevcut olan klinik verileri, hava durumu, sosyoekonomik

özellikleri hakkındaki bilgilerle birleştirerek dört yaygın makine öğrenimi yaklaşımının performansını karşılaştırmaktır. Çalışmada 29.392 hastanın verisi kullanılmış olup bu verilerin %80'e %20 olacak şekilde eğitim ve test verisi olarak kullanılmıştır. Gradyan güçlendirme, karar ağaçları, rastgele orman ve lasso regresyon algoritmaları kullanılmıştır. Çalışma sonucunda; karar ağacı %72, lojistik regresyon %83, rastgele orman %82, gradyan güçlendirme ise %84 doğruluk AUC performans değerleri elde edilmiştir. Gradyan güçlendirme algoritması en yüksek oranla daha iyi performans göstermiştir (Patel ve ark., 2018).

Lee ve ark. (2020), acil serviste bekleme süresi değerinin yüksek ve düşük akut hastalar arasında farklılık gösterdiği durumlarda hasta bekleme sürelerini en aza indirmek için optimal bir planlama politikası geliştirmeyi amaçlamıştır. Acil servis verimliliğine ulaşmak için çeşitli faktörler dikkate alındığında hastanın planlaması karmaşık hale gelmesi sorununu çözmek için acil servis hastalarının planlama sürecinde pekiştirmeli öğrenme algoritmaları kullanılarak model oluşturulmuştur. Bu çalışma, özellikle acil servisin dinamik ortamında karar vericilere yardımcı olmak amacıyla acil servis uygulamaları için pekiştirmeli öğrenmenin başarılı bir şekilde uygulandığını göstermektedir (Lee ve ark., 2020).

Shen ve Lee (2020); triyajla ilgili çalışmasında planlama, uygulama, kontrol etme, önlem alma yani PUKÖ döngüsü kullanılarak süreçte hastaların çok beklemesi gibi durumların iyileştirilmesi için çalışma yapmıştır. Bu çalışmada 18 dakikalık triyaj bekleme süresini 13 dakikaya düşürerek triyaj sürecinde iyileştirme elde etmişlerdir (Shen ve Lee, 2020).

Hong ve ark. (2018), çalışmalarında triyaj sırasında toplanan bilgilere ek olarak hasta geçmişini kullanarak acil serviste hastaneye yatışı öngörmeyi amaçlamışlardır. 972 değişken kullanılmış olup veriler %80'e %10 olacak şekilde eğitim ve doğrulama, %10'si test için kullanılmıştır. Lojistik regresyon, XGBoost ve derin sinir ağları algoritmaları kullanılmıştır. Değişkenlerin tamamı üzerinde eğitilen modeller lojistik regresyon için 0,91, XGBoost için 0,92 ve derin sinir ağları için 0,92 doğruluk oranı vermiştir. En başarılı performans lojistik regresyon algoritmasıyla elde edilmiştir (Hong ve ark., 2018).

Chen ve ark. (2020); triyajdaki kalabalıklaşma sorununu ve hastaların durumu hakkında etkili klinik karar verilebilmesi için yaptıkları çalışmalarında, bir acil

servisten 2018-2018 yılları arasında travma dışı yetişkinlerin 104.083 acil servis ziyaretinin tamamının retrospektif bir kohortu oluşturulmuştur. 18.308 hastanın verisi kayıt eksikliği gibi çeşitli sebeplerden dolayı veri setinden çıkarılmıştır. Bu veriler; demografik bilgiler, hayati belirtiler, triyaj bilgileri gibi özetlenmektedir. Triage'da toplanan değişkenleri ve klinik notları kullanarak eğilimi tahmin etmek için lojistik regresyon, derin öğrenme yöntemleri kullanılmıştır. Hekimlerin klinik anlatımlarının ve doğal dil işlemeye dayalı yapılandırılmış verilerin kullanıldığı önerilen derin öğrenme modeli, F1 puanı açısından yaygın olarak kullanılanlardan daha iyi performansı göstermiştir. Ayrıca acil serviste klinik karar vermede hayati belirleyiciler olarak hizmet eden klinik anlatımların öznel bölümünün önemini de kanıtlanmıştır (Chen ve ark., 2020).

Jilani ve ark. (2019), insan ve fiziksel kaynakların optimal planlanmasını desteklemek için acil servislere katılımı doğru bir şekilde tahmin eden, kaynak tahsisine yardımcı olabilecek ve yoğun hastaneler üzerindeki baskıyı azaltabilecek acil servis bakımlarını tahmin etmek için sezgisel tabanlı bir bulanık mantık modeli oluşturmuşlardır. Dört hastanenin Ocak-2011-Aralık-2015 arasındaki geçmiş ziyaret verileri modeli eğitmek ve test etmek için kullanılmıştır. Bu çalışmada, acil servise başvuruyu tahmin elde etmek için sezgisel tabanlı bir zaman serisi modeli geliştirerek aşırı kalabalıklaşmanın giderilmesi amaçlanmıştır. Günlük kısa vadeli, aylık ve uzun vadeli (dört ay sonrası) olacak şekilde zaman serileri oluşturulmuştur. Acil servise başvuruların, daha sonra hastaneye başvuruların tamamen rastgele bir süreç olmadığını ve bu nedenle kabul edilebilir bir doğrulukla tahmin edilebileceğini ortaya çıkarılmıştır. Tahmin zaman aralıkları genişledikçe (günlükten aya kadar) tahmin doğruluğu arttığı sonucuna ulaşılmıştır. Önerilen bulanık zaman serisi modelinin Harvey, Leybourne ve Newbold testi kullanılarak istatistiksel olarak anlamlı bir performansa sahip olduğu kanıtlanmıştır (Jilani ve ark., 2019).

Pineda ve ark. (2015), influenza virüsünün tespiti için yedi makine öğrenimi sınıflandırma algoritmasıyla çalışmışlardır. 2008-2011 yılları arasında dört hastanenin acil servis raporlarından eğitim ve test verileri elde ederek, modeller oluşturulmuştur. Yapılan çalışma sonucunda en iyi sonuçlar naive bayes algoritmasıyla elde edilmiştir (Pineda ve ark., 2015).

3. MATERYAL VE YÖNTEM

3.1. Acil Tıp ve Triyaj

İlk yardım kelimesi ilk defa 1881’de savaş yaralıları için çıkarılan yönetmelikte kullanılmıştır. 18. Yüzyılda acil ve ilk yardım sistemlerinin kurucusu olarak bilinen cerrah Esmarsch Schleswig ilk yardımın önemini anlatan bir eser yazmıştır. İlk yardımın tarihsel boyunca değişimine bakıldığında büyük savaşlarda yaralıların cepheden taşınarak kurtarılma çalışmalarında hızlandığı görülmektedir (Dumlu, 2020).

2.Dünya savaşı yıllarında acil sağlık çalışmalarında düzenlemeler yapılmıştır ve büyük hastanelerde kaza odası adı altında acil servis odalarına gereksinim artmıştır. Amerikalılar savaşlardan büyük deneyimler kazanmıştır ve ardından acil bakım hizmetlerinin sistematikleştirilmesi için 1960’lı yıllarda sağlık la ilgili yasalar çıkarmıştır (Dumlu, 2020). ABD’de hastane öncesi acil bakım birimlerin oluşturulması ve çalışması üzere personel yetiştirilmesi için çalışmalar başlatılarak 1970’te Cincinnati Üniversitesi’nde acil tıp anabilim dalı kurulmuştur (Kuşku Kıyak, 2013).

Acil tıp, hayati tehlikesi olan durumların en hızlı şekilde tanımlanıp ve ardından müdahalesinin yapılmasını amaçlayan bir tıp dalıdır (Kuşku Kıyak, 2013). Hastaların yaşı, cinsiyeti, ödeme gücüne bakılmadan acil bir durumun tanısı, tedavisi ve gerekli birimlere yönlendirilmesiyle beraber acil durumların önlenmesi için çalışılmaktadır (Aydın, 2011).

Acil tıp bilinci dünyada ikiye ayrılmaktadır. Bunlar; Anglo-Amerikan sistemi, Franko-Germen sistemidir (Değirmenci, 2020).

Anglo-Amerikan sisteminde acil tıp birimi ayrı bir tıpta uzmanlık alanı olarak benimsenmiştir. Hasta, sağlık personelleri tarafından sahada hastane öncesi bakımı yapar ve sağlık hizmeti alabilmesi için hızlıca hastanenin acil servis birimine taşınır. Türkiye’nin de aralarında bulunduğu ABD, Kanada, İngiltere, Avustralya gibi ülkelerde bu sistem uygulanmaktadır (Değirmenci, 2020).

Franko-Germen sisteminde hastaya sađlık hizmeti verilebilmesi iin acil hekimleri ve tıbbi imkanlar olay yerine tařınır ve anestezi uzmanları hastane ncesi acil sađlık hizmetini gerekleřtirirler. Hastalar sahadan sonra direkt olarak hastanede servise yatırılır. Acil servis gibi ayrı bir birim bulunmamaktadır. Acil servis, ayrı bir tıpta uzmanlık alanı deđildir. Bu sistem; Avusturya, Fransa, Almanya, Rusya, gibi lkelerde zetle Avrupa’nın %70’inde uygulanmaktadır (Deđirmenci, 2020). Bu lkelere ek olarak; Estonya, Slovenya, Polonya, İskoya, İtalya, Macaristan ve Hırvatistan gibi lkelerde bu gruba aittir (Aydın, 2011).

Acil tıp; ncelikli olarak acil servislerdeki hizmetleri kapsarken bununla birlikte hastane ncesi sađlık hizmetlerini ve afet ncesi planlamalarıyla da ilgilidir. Hastalık ve yaralanmalarda acil tıp uzmanları; hastanın yařam bulgularının dzenlenmesini, deđerlendirilmesini ve tedavisini sađlamayı amalamaktadır (Kuřku Kıyak, 2013).

Acil servisler; yataklı sađlık kuruluřları olup, kesintisiz sađlık hizmeti hizmet vermektedir. İhtiya duyulduđuunda hastalar karřılanarak ilk tıbbi mdahale yapılmakta olup, kesin tedavi srecine kadar tıbbi bakımı kesintisiz yapılmaktadır (Dumlu, 2020). Aynı zamanda acil yardım; travma, ilk yardım, triyaj ve hastane ncesi bakımı kapsamaktadır (Kuřku Kıyak, 2013).

Trijaj kavramı Fransızca “Trier” fiilinden tretilmiř olup ayıklama ve nceliklendirme anlamına gelmektedir (Almalahy, 2022). Bu kelime ilk olarak tarımsal rnlerin kalitesine ve fiziksel zelliklerine gre ayıklanıp, tanımlanması iin kullanılmıřtır (Dumlu, 2020). Bazı kaynaklarda ise Japonya’daki kahve ekirdeklerinin tasnifi iin kullanıldığını belirtirken, bazı kaynaklar ise İngiltere’deki kalitesiz yn rnlerinin ayrıştırılması iin kullanıldığını belirtmektedir (Bostancıođu, 2021).

Trijaj uygulaması, savařlar sırasında yařanılan sıkıntılar ve acil mdahale zorlukları sonucunda bir gereklilik olarak ortaya ıkmıřtır. 17. Yzyıla kadar savařlarda askerler sađlık bakım hizmeti alamamaları sonucunda 18.yzyılda dzenli sađlık hizmeti alabilmeleri adına sistematik triyaj kuralları ortaya ıkmıřtır (Almalahy, 2022).

Oxford İngilizce szlđuinde triyaj; tedavi sırasına karar vermek iin hastalıklara veya yaralılara aciliyet derecelerinin atanması olarak tanımlanmıřtır (Elbaih, 2017).

Trijaj; ok sayıda hastanın olduđu ortamlarda ncelikli olarak mdahale ve sađlık kuruluřuna sevk edilmesi gerekenleri sınıflandırmak amacıyla hem olay yerinde hem de ynlendirilen sađlık kuruluřunda uygulanan sınıflandırma iřlemidir (Dumlu, 2020).

Acil servislerde triyaj, optimal sonuçlar elde etmek amacıyla acil ilgilenilmesi gereken hastaları tanımlamak için kullanılan bir karar verme sürecidir (Bostancıoğlu, 2021).

Triyaj, hastaların acil servise ilk başvurdukları zaman uygulanan bir karar verme sürecidir. Bu karar verme sürecinde, hastaların durumlarına göre tıbbi açıdan önceliklendirmenin yapılması işlemidir (Almalahy, 2022). Hastaların bakım önceliğini tespit edebilmek adına da önemli bir uygulamadır. Hastanın doğru zamanda, doğru yerde acil sağlık hizmeti alması amaçlamaktadır.

Triyajın; hastanın hayatını korumak, sonrasında gerçekleşebilecek kayıpları veya ileri yaralanmaları önlemek, kısıtlı kaynakları koruma hastanelere yapılan gereksiz başvuruları önlemek, kritik durumda olmayan hastanın bakımını kolaylaştırmak, personelin verimli çalışmasını sağlamak, hasta akışını ve işlem hareketlerini kolaylaştırmak ve bekleme sürelerini optimize etmek için uygulanmaktadır (Şimşek, 2018).

Acil servislerde triyaj uygulamaları, hastane sağlık hizmetlerinin ayrılmaz bir parçası durumundadır (Değirmenci, 2020). Acil servise başvuran hasta sayılarının fazla olmasından dolayı yaşanan yoğunlukta acil serviste bulunan birçok kritik hasta triyaj kararındaki gecikmeden dolayı kötü sağlık sonuçlarına maruz kalma riskiyle karşı karşıyadır. Bu sebeple kimlerin acil tedaviye ihtiyacı olduğunu belirlemek, hangi hastanın daha önce bakılması gerektiğine karar vermek amacıyla triyaj sistemlerine ihtiyaç doğmuştur. Çok sayıda hastanın başvurusunda tedavi sırasına karar vermek ve aciliyet durumlarının belirlenmesi için triyaj sistemleri kullanılmaktadır (Çetin, 2021).

Triyaj kararlarındaki tutarsız durumlar ya da yanlış sınıflandırmalar hastalarda ölüm vakalarının görünmesine ve bu vakaların artışına yol açmaktadır (Çetin, 2021). Yüksek triyaj doğruluğu, daha kaliteli bir acil servis hizmetini sağlamaktadır (Kamali ve ark., 2019). Bu sebeple, triyajın hızlı ve doğru bir şekilde yapılması gerekmektedir. Triyaj; hastaların bilgilerinin toplanması, odaklanmış verimli ve doğru bir değerlendirmenin yapılmasını, aciliyet seviyelerinin belirlenmesini, hastaların ve hastaların ihtiyaçlarının önceliklendirilmesi süreçlerini içermektedir (Çetin, 2021).

Triyajın amacı; acil servislerdeki hastaların durumlarının hızlı tanımlanması ve böylece hayati tehlikeleri için girişimlerinin hızlıca yapılabilmesini, personelin doğru yerde ve doğru şekilde etkin kullanımını sağlamak, hastaların önceliklerini belirlemek, ilk tanılama işlemlerini verimli yapıp hasta akışını hızlandırmak ve servis

yoğunluğunu azaltmaktır. Triyajın uygulanması acil servisteki düzeni iyileştirir, işleyiş hızını artırır. Hasta ve hasta yakınlarının korku ve endişelerini azaltır, dolaylı bir şekilde de bakım ekibi için sakin bir çalışma ortamı sunmaktadır. Triyajın en yaygın çeşitleri aşağıdaki gibidir (Almalahy, 2022).

- Savaş alanında triyaj
- Afet alanında triyaj
- Olay yerinde triyaj
- Hasta yatışlarında triyaj
- Acil servislerde triyaj

3.1.1. Triyaj türleri

Acil servislere başvuran hastaların mümkün olduğunca hızlı bir şekilde sağlık hizmeti alması amaçlanmıştır. Ancak hasta yoğunluğu, kaynak yetersizliği, personel eksikliği, gibi sebeplerden bu süreler uzayabilmektedir. Özetle, sürelerin uzaması hastaların sağlıkları açısından tehdit oluşturmaktadır. Bu nedenle triyaj sistemlerinin kullanımı meydana gelmiştir (Aydın, 2011). Triyaj sistemlerinin hedefleri; kritik zamana duyarlı bakım ihtiyacı olan hastaları belirlemek ve önceliklendirmektir (Çetin, 2021).

Trijajın kullanıldığı sağlık sistemlerinde süreçlerin standart hale getirilebilmesi ve sürecin iyileştirilmesi için triyaj skalaları geliştirilmiştir. Ancak evrensel bir skala oluşturulmamıştır. Bazı ülkeler kendi triyaj skalalarını oluşturmuşlardır. Bu triyaj skalaları seviyelerden oluşturulmuştur. Her seviye için acil durumların tanımları oluşturulup, hastanın şikayetlerine uygun triyaj kategorisi belirlenmektedir (Almalahy, 2022).

Uluslararası alanlarda geliştirilmiş çeşitli triyaj sınıflama sistemleri bulunmaktadır. Bunlar; üç, bazıları dört ve bazıları ise beş seviyeli sistemler olarak örneklendirilebilir (Çetin, 2021). İlk kullanılan sistemler üç seviyeliyken, yetersiz kaldığı düşünülüp üçlü sistem geliştirilmiş ve beş seviyeli sistemler kullanılmaya başlamıştır. Bazı acil servislerde üç seviyeli sistemde daha iyi sınıflandırma yapabilmek için bir seviye daha ekleyerek dört seviyeli olarak kullanılmaktadır (Almalahy, 2022). Dünyada acil servislerde kullanılan triyaj seviyeleri Tablo 3.1’de gösterilmiştir (Almalahy, 2022).

Tablo 3.1. Dünyada acil servislerde kullanılan triyaj sınıflandırma seviyeleri.

2'li sınıflandırma	3'lü sınıflandırma	4'lü sınıflandırma	5'li sınıflandırma
Acil	Çok acil	Hayati tehlike	Hemen bakım
Acil değil	Acil	Çok acil	Çok acil
	Acil değil	Acil	Acil
		Acil değil	Acil değil
			Sevk

Ülkeler bazında kullanılan triyaj sistemleri Tablo 3.2'deki gibidir (Çetin, 2021).

- Avustralya triyaj skalası (ATS)
- Acil durum ciddiyet indeksi (ESI)
- Kanada triyaj ve kesinlik skalası (CTAS)
- Manchester triyaj skalası (MTS)
- Güney Afrika triyaj skalası (SATS)
- Fransa acil hemşireliği hastane ölçeği sınıflandırması (FRENCH)
- Acil tıp Japon birliği (JAAM)
- Türkiye triyaj sistemi

Tablo 3.2. Dünyada tanınmış triyaj sistemleri ve uygulandıkları ülkeler.

Trijaj sistemi	Kullanılan ülkeler
Manchester triyaj sistemi	İngiltere, İrlanda, Portekiz, Hollanda
Avustralya triyaj sistemi	Avustralya, Yeni Zelanda
Kanada triyaj ve aciliyet sistemi	Kanada
Aciliyet ciddiyet indeksi	Amerika
Güney Afrika triyaj sistemi	Güney Afrika
Fransa triyaj sistemi	Fransa

3.1.1.1. Avustralya triyaj skalası

Hastaları sağlık durumlarını aciliyete göre sıralamak için günümüzde Avustralya ve Yeni Zelanda'nın tüm acil servislerinde kullanılan beş seviyeli bir ulusal sistemdir (Almalahy, 2022).

Avustralya acil servislerinde 1994 yılında uygulanmaya başlayan Avustralya triyaj sistemi beş seviyelidir. Bir hastanın tıbbi değerlendirmesi ve tedavisi için en fazla bekleme süresini tespit etmek için kullanılan bir araçtır (Dumlu, 2020).

ATS ile belirlenen tedavi süresi, o kategoride yer alan hastaların değerlendirilme ve tedavisi için beklemesi gereken maksimum süreyi temsil etmektedir (Çetin, 2021). Avustralya triyaj skalasına ait kategoriler, maksimum bekleme süreleri ve başarı oranları Şekil 3.1'de verilmiştir (Dumlu, 2020).

	ATS KATEGORİSİ	MAKSİMUM BEKLEME SÜRESİ	BAŞARI ORANI %
1	Acele-Hayati tehlike var	Hemen	100
2	Daha az acele-hayati tehlike var	10 dakika	80
3	Hayati tehlike riski var	30 dakika	75
4	Hayati tehlike potansiyeli	60 dakika	70
5	Az acil	120 dakika	70

Şekil 3.1. Avustralya triyaj skalası.

2015 yılında yapılan bir analizde ATS'nin kabul edilebilir sınırlar içerisinde güvenilir olduğu ancak mükemmel sonuçlar için geliştirilmesi gerektiği tespit edilmiştir (Bostancıoğlu, 2021).

Avustralya'da acil servislerde 1993 yılında uygulanmaya başlanan Avustralya Triage Skalası, 1990'lı yılların sonlarında çeşitli geliştirmelere uğramış ve Avustralya Triage Skalası olarak tekrar adlandırılmıştır (Bostancıoğlu, 2021).

Avustralya triyaj sisteminde yetişkinler için fizyolojik belirtilerin özeti Tablo 3.3'te özetlenmiştir (Bostancıoğlu, 2021).

Tablo 3.3. ATS yetişkin fizyolojik belirtilerin özeti.

	Kategori 1 Hemen	Kategori 2 10 dakika	Kategori 3 30 dakika	Kategori 4 60 dakika	Kategori 5 120 dakika
Havayolu	Tam veya kısmi tıkalı havayolu	Açık	Açık	Açık	Açık
Solunum	Ağır solunumsal sıkıntı Solunum yokluğu Hipoventilasyon	Orta seviye Solunumsal sıkıntı	Hafif solunumsal sıkıntı	Solunum sıkıntısı yok	Solunum sıkıntısı yok
Dolaşım	Ağır hemodinamik problem Dolaşım yokluğu Kontrol edilemeyen kanama	Orta seviye Hemodinamik problem	Hafif hemodinamik problem	Hemodinamik problem yok	Hemodinamik problem yok
GKS	GKS<9	GKS 9-12	GKS>12	Normal GKS	Normal GKS

Avustralya Triyaj Skalası, Kanada ve Büyük Britanya için bir başlangıç noktası olup ATS'yi referans alarak yapılan çalışmalar ile geliştirilmiştir. Bu çalışmaların sonucunda Kanada Triyaj Sistemi (CTAS) ve Manchester Triyaj Skalası (MTS) oluşturulmuştur (Almalahy, 2022).

3.1.1.2. Acil durum ciddiyet indeksi

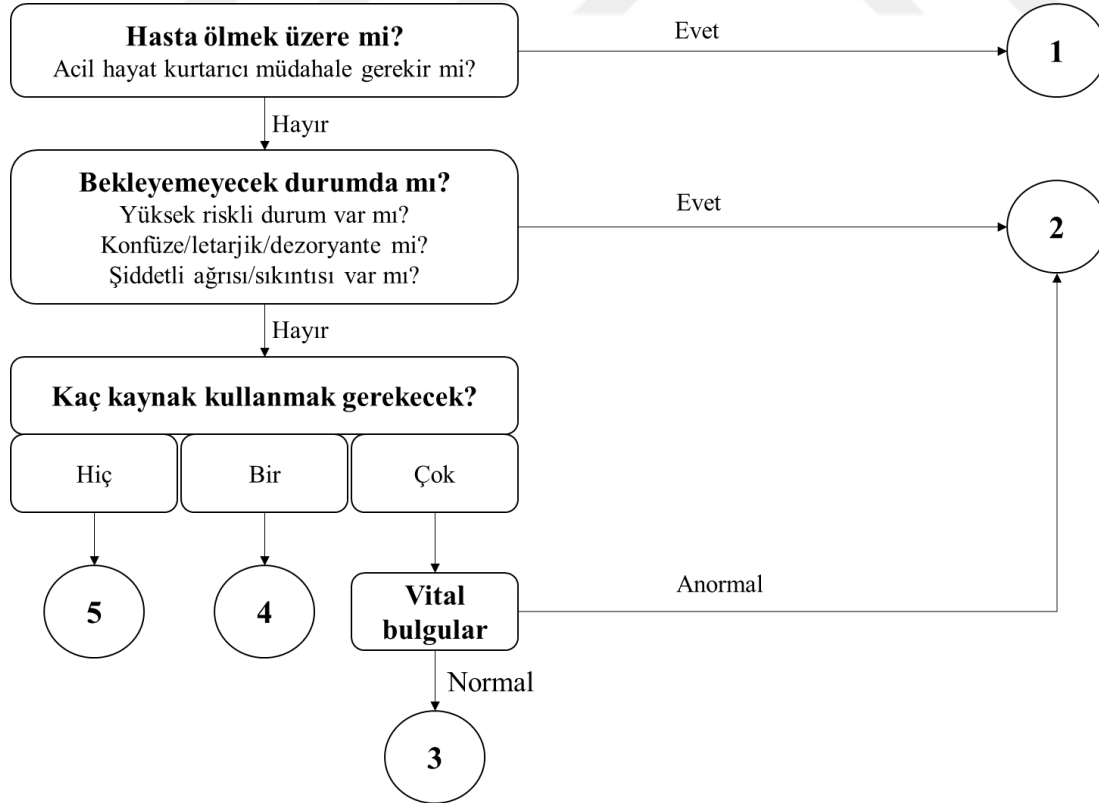
Acil durum ciddiyet indeksi (ESI), 1990'lı yıllarda Richard Wuerz ve David Eitel tarafından geliştirilmiş olan bir triyaj sistemidir (Almalahy, 2022). Amerika Birleşik Devletleri'nde geliştirilmiştir (Christ ve ark., 2009). Amerika Birleşik Devleti'nin bir bölümü ve birçok Avrupa ülkesinde acil servislerde kullanılmaktadır (Almalahy, 2022).

Acil Servis Ciddiyet İndeksi, beş seviyeli bir triyaj sistemidir. İlk olarak triyaj hemşiresi gelen hastalar içinden “Önce kim görülmeli?” sorusunu sorarak öncelik sıralamasını yapar. Aynı anda birden fazla öncelikli hasta geldiği durumlarda ise “Her hasta güvenli şekilde ne kadar süre bekleyebilir?” sorusu önemli hale gelmektedir (Yazıcı, 2022).

İlk olarak, triyaj hemşiresi kesinlik seviyesini tespit eder. Bir hasta yüksek kesinlik seviyesi kriterlerini karşılamıyorsa (seviye1 ve seviye2) triyaj hemşiresi bir triyaj seviyesi (seviye 3-4-5) belirlemeye yardımcı olmak için beklenen kaynak ihtiyacını tespit eder (Yazıcı, 2022). ESI triyajında çalışacak hemşirelerin deneyimli, kapsamlı bir triyaj eğitimi almış olması gerekmektedir (Yazıcı, 2022).

Hastanın tedavi önceliği tüketilecek kaynaklara göre belirlenmesi ve her hasta için hayati bulgular ile zaman kaybedilmemesi sebebiyle kısa sürede uygulanabilmesini sağlamaktadır. Hastanın hayati bulgularına göre bazı 3.kategori hastaları yeniden sınıflandırıp, eğer gerekli ise 2.kategoriye aktarmak için kullanılmaktadır. Aynı zamanda doktorların yapacağı değerlendirme ve tedavi için bekleme zaman aralığı tanımlanmamaktadır (Dumlu, 2020). ESI'nin diğer triyaj uygulamalarından farkı; sadece başvuru şikayetlerini dikkate almayıp, hastalar için tüketilecek kaynakları da dikkate alarak bu verileri de kategori belirlemede kullanmasıdır (Değirmenci, 2020).

1. Kategoriye ait hastaları daha fazla kaynağa ihtiyacı var olarak değerlendirirken, 5. Kategorideki hastaların herhangi bir test, tedavi vb. ihtiyaca gerekli olmadığı şekilde değerlendirmektedir (Dumlu, 2020). Acil durum ciddiyet indeksinin akış şeması Şekil 3.2'de verilmiştir (Yazıcı, 2022).



Şekil 3.2. Acil servis ciddiyet indeksi akış şeması.

ESI'nin faydası acil müdahaleye ihtiyacı olan hastaların hızlı bir şekilde tespit edilmesidir. Odak noktası ise kısıtlı kaynak ortamında hastaları hızlı şekilde sıralayabiliyor olmasıdır (Yazıcı, 2022).

ESI, oluşturulduğu ilk günden günümüze kadar daha iyi sınıflandırma ve değerlendirme performansı elde edebilmek için 4 defa güncellenmiştir (Dumlu, 2020).

3.1.1.3. Kanada triyaj ve aciliyet skalası

Kanada triyaj ve aciliyet skalası (CTAS), 1990'lı yıllarda New Brunswick eyaletinde görev yapan hekim grubu tarafından geliştirilmiştir. Hastaların aciliyet durumlarına göre ne kadar sürede tedavi edilmeleri gerektiğini belirleme, acil servis verimliliğine katkıda bulunma, iş yükünü hafifletme amaçlarıyla geliştirilmiştir (Dumlu, 2020). CTAS sınıflandırmaları ve maksimum bekleme süreleri Şekil 3.3'te verilmiştir (Dumlu, 2020).

CTAS KATEGORİSİ	MAKSİMUM BEKLEME SÜRESİ	Renk
1 Diriltme	Hemen	Mavi
2 Çok acil	15 dakika	Kırmızı
3 Acil	30 dakika	Sarı
4 Az acil	60 dakika	Yeşil
5 Acil değil	120 dakika	Beyaz

Şekil 3.3. CTAS sınıflandırmaları ve maksimum bekleme süreleri.

Avustralya Triage Skalası esas alınarak geliştirilen CTAS'da da beş seviyeli skala bulunmaktadır. Ölçekte, doktor değerlendirilmeleri olduğu gibi her seviye klinik tanıları da içermektedir (Dumlu, 2020). Hastanın semptomları değişirse triyaj uygulaması tekrarlanır ve hastanın semptomlarında değişiklik olup olmadığı kontrol edilmektedir (Almalahy, 2022).

CTAS, acil durumlarda hastanın triyaj sınıflandırmasını hızlı şekilde yapmaktadır ve uygulanabilirliği yüksek bir skaladır (Dumlu, 2020). ATS ile arasındaki en büyük fark, CTAS’da ciddiyet seviyesi aynı olan hastalar için ağrı skalası, triyaj zamanı ve hemşirelik değerlendirmelerinin tanımlanmış olmasıdır (Almalahy, 2022). CTAS, 2004 yılında triyajda bekleyen hastaların triyaj kategorilerine uygun değerlendirilme süreleri güncellendiği için yenilenmiştir (Değirmenci, 2020).

3.1.1.4. Manchester triyaj skalası

Manchester Triage skalası, Büyük Britanya’da 1994 yılında hastaların risk tahminini yapan ve önceliğini belirleyen beş seviyeli bir şikayet tabanlı bir sınıflandırma sistemidir (Almalahy, 2022). Avrupa ülkelerinde yaygın olarak kullanılmaktadır. İngiltere acil servislerinde kullanımı ise zorunludur (Değirmenci, 2020). Hollanda, İsveç, Portekiz ve Avustralya gibi ülkelerde yapılan araştırmalarda geçerliliği ve doğruluğunun orta derecede olduğu görülmüştür. Aynı zamanda mortaliteyi tahmin etmede iyi derece olduğu belirtilmektedir (Almalahy, 2022). MTS; güvenilebilirlik, duyarlılık, özgüllük ve tekrarlanabilirlik yönünden birbirlerinden ayrı olarak iyi sonuçlar göstermektedir (Bostancıoğlu, 2021).

Acil doktorları ve hemşireleri arasında bir standart oluşturmak ve ortak bir triyaj sistemi oluşturmak amacıyla geliştirilmiştir (Dumlu, 2020). Hastaları acile geliş sırasıyla değil de hastanın klinik durumlarının ciddiyet durumuna göre standardize etmeyi, hızlı karar vermeyi amaçlamaktadır (Değirmenci, 2020).

Aciliyet seviyeleri ve bekleme süreleri farklı renklerle tanımlanmıştır. Başvuru bilgilerini temel alan 52 adet akış şemasını içermektedir (Değirmenci, 2020). Her akış şeması altı temel ayırmacıya dayanmaktadır (Almalahy, 2022). Akış şemaları; şiddetli ağrı, dirençli kusma, hava yolu tıkanıklığı gibi öncelik belirten semptomlar içermektedir. Triage görevlisi hastaya en uygun akış şemasını uygulayarak triyaj kategorisini seçer (Değirmenci, 2020).

MTS’de 52 akış şeması ve 5 renk mevcuttur. Klinik önceliği aşağıdaki 3 temel maddeye göre belirlenmektedir (Dumlu, 2020).

- Hem acil servis hem de hasta için yaklaşımı geliştirmek
- Hasta durumu ilerlemeden muayene edilmesini ve bekleme süresinin kısaltılmasını sağlamak
- Klinik önceliğinin triyajda belirlemek.

Manchester triyaj skalasının sınıflandırmaları ve maksimum bekleme süreleri Şekil 3.4'te verilmiştir (Almalahy, 2022).

MTS KATEGORİSİ	MAKSİMUM BEKLEME SÜRESİ	Renk
1 Resüsitasyon	Hemen	Kırmızı
2 Çok acil	10 dakika	Turuncu
3 Acil	60 dakika	Sarı
4 Standart	120 dakika	Yeşil
5 Acil değil	240 dakika	Mavi

Şekil 3.4. MTS sınıflandırmaları ve maksimum bekleme süreleri.

3.1.1.5. Güney Afrika triyaj skalası

2004 yılında Güney Afrika Triage Grubu (SATS); Cape Town ve Stellenbosch Üniversitelerinin Acil Tıp Bölümü sponsorluğunda toplanarak, Güney Afrika’da bir triyaj ölçeği oluşturmayı amaçlayarak bir triyaj skalası oluşturmuşlardır. Güney Afrika’da kullanılan ilk ulusal triyaj sistemidir. Oluşturulan bu triyaj sistemi sayesinde bekleme süreleri ciddi oranda azaltmıştır (Elbailh, 2017). Bu triyaj sisteminde beş renk kodlu beş kategoriden oluşmaktadır. Renk kodlarının tanımları aşağıdaki gibidir (Torun, 2016).

- Kırmızı – resüsitasyon / fizyolojik olarak istikrarsız hastalar,
- Turuncu – potansiyel olarak stabil olmayan fizyolojisi olan ciddi vakalar veya potansiyel olarak hayatı tehdit eden patolojiler,
- Sarı – fizyolojik olarak stabil, ciddi medikal veya travma sorunlarına neden olabilecek vakalar,
- Yeşil – minör yaralanmalar
- Mavi – Vefat etmiş.

SATS triyaj skalası; Botswana, Malawi, Brezilya, Polonya ile Yeni Zelanda'da uygulanmaktadır ve Sudan, Gana, Ruwanda, Tanzanya, Suudi Arabistan'da kullanılması için çalışılmaktadır (Torun, 2016).

3.1.1.6. Fransa acil hemşireliği hastane ölçeği sınıflandırması

Fransa Acil Hemşireliği Hastane Ölçeği Sınıflandırılması (FRENCH), 1990'lı yıllarda hastaların şikayetleri esas alınarak geliştirilmiştir. Beş seviyeli bir skaladır. Sonrasında 2006 yılında bekleme sürelerinin azaltılması, kaynak ihtiyacının da değerlendirilmesi amacıyla yenilenmiştir (Değirmenci, 2020).

3.1.1.7. Acil tıp Japon birliği

Japonya'da uygulanan triyaj skalası Avrupa triyaj skalalarından farklıdır. 1960'ta Japonya acil sağlık hizmetleri kurulmuştur. Ancak acil tıp alanında eğitilmiş hekimler olmadan hizmet vermeye başlamıştır. 1973 yılında "Acil tıp Japon birliği" kurulmuştur. İlk zamanlarda travma ve yanık üzerine yoğunlaşmıştır. 1977 yılında devlet acil sağlık hizmetleri kurulmuştur ve üç seviyeli olarak hizmet vermiştir. Acil servislerde hastalara acil tıp doktorları tarafından bakılmaktadır (Kuşku Kıyak, 2013).

3.1.1.8. Türkiye triyaj sistemi

Ülkemizde ilk defa 2009 yılında yataklı sağlık tesislerinde acil servis hizmetlerinin uygulama usul ve esasları hakkında tebliğde triyaj uygulamasının nasıl olması gerektiği konusuna yer verilmiştir (Çetin, 2021). 2009 yılında uygulanmaya başlanmıştır. Hastaneye başvuru yapılırken hemşireler, acil tıp teknisyenleri, doktorlar, sağlık memurları gibi sağlık personeli tarafından uygulanmaktadır (Almalahy, 2022). Sağlık bakanlığında triyaj; acil servislere gelen hastaların, görevli sağlık çalışanı tarafından hastalıkları ile ilgili şikâyetleri ile şikayetlerinin şiddeti ve sağlık durumlarının aciliyeti göz önüne alınarak, sağlık önceliklerinin ve yönlendirilecekleri tedavi alanının belirlenmesi işlemidir. Sınıflandırma olarak kırmızı, sarı, yeşil renkleri kullanılmaktadır. Triage sırasında hastaların yaşamsal bulgularının tıbbi cihazlarla ölçümü yapılmaktadır (Almalahy, 2022).

Türkiye'de uygulanan triyaj sınıflandırması ve kullanılan renk tanımlamalarının ayrıntıları ekler bölümünde Tablo A.1'de verilmiştir (Resmi gazete, 2023).

3.1.2. Yaşamsal bulgular

Yaşamsal bulgu diğer adıyla vitaller; hastanın hayati durumunun, yaşamı sürdürmesi için gerekli fonksiyonlarının durumunu gösteren bulgulardır. Bu bulgular hastanın sağlığının değerlendirilmesinde yardımcı olur ve hastalıklara dair ipuçları vermektedir.

Çalışma kapsamına alınan yaşamsal bulgular; kalp hızı- nabız, kan basıncı- tansiyon, vücut sıcaklığı-ateş, solunum sayısı, oksijen satürasyonu, kan şekeri, glasgow koma skalasıdır.

3.1.2.1. Kalp hızı

Kalp hızı yani diğer adıyla nabız, kalbin dakikada kaç kere attığını ifade etmektedir. Yaşa bağlı olarak normal kalp hızı değişebilmektedir. Kalp atım hızı için kabul görmüş değerler 60 atım/dakika ile 100 atım/dakika arasındadır (Yazıcı, 2022). Yetişkinlerde dakikada 100 atım ve üzeri olması durumu taşikardi; 60 atım ve altı olması durumu bradikardi olarak adlandırılmaktadır.

3.1.2.2. Kan basıncı

Kişinin tansiyonunu temsil etmektedir. Sistolik basınç ve diyastolik basınç olarak ikiye ayrılmaktadır. Sistolik basınç; sol ventrikül kasılması sırasında arterler üzerindeki maksimum basınçtır (Yazıcı, 2022). Diğer bir deyişle; kalbin kasılması sırasında ulaşılan en yüksek kan basıncıdır. Halk arasında büyük tansiyon değeri olarak da bilinmektedir ve 120 mmHg – 130 mmHg arasında olması beklenmektedir. Diyastolik basınç; her bir kalp kasılması arasındaki arterler üzerindeki dinlenme basıncıdır (Yazıcı, 2022). Diğer bir deyişle; kalp atımları arasında, kalp aktif şekilde kan pompalamazken, damarlar içinde oluşan kan basıncı değeridir. Halk arasında küçük tansiyon olarak bilinmektedir ve 70 mmHg-90 mmHg arasında olması beklenmektedir. MAP değerinin; 70 mmHg-100 mmHg arasında olması beklenmektedir.

3.1.2.3. Vücut sıcaklığı

Vücut sıcaklığı diğer bir deyişle ateş değeri birçok kaynaktan etkilenen bir durumdur (Yazıcı, 2022). Sağlıklı bir bireyin alnından ölçülen vücut sıcaklığı 35.7 °C ile 37.5 °C arasında değişmektedir. Yapılan çalışmalarda triyaj sırasında yüksek ateş sınır değeri olarak 38.3 °C olarak kabul görmüştür. Ateşin hastane içi mortalite ile ilişkisini gösteren bir çalışma bulunmamaktadır (Değirmenci, 2020).

3.1.2.4. Solunum sayısı

Solunum sayısı, bir kişinin bir dakikadaki nefes sayısıdır. Aynı zamanda solunum sayısı, hastalık şiddetinin en önemli göstergelerinden biridir. Normal solunum sayısı, 12 solunum/dakika – 20 solunum/dakika arasındadır. Solunum sayısı, 20 solunum/dakika değerinden fazla olduğu durumda problemlili bir durum olduğu düşünülmektedir (Değirmenci, 2020).

3.1.2.5. Oksijen satürasyonu

Satürasyon, kandaki oksijenle doymuş hemoglobin moleküllerinin yüzdesidir. Bu değer, parmak ucundan pulse oksimetre cihazı ile ölçülerek elde edilebilir (Değirmenci, 2020). Kandaki oksijen değerinin belirli bir düzeyde olması gerekmektedir. Sağlıklı bir bireyde kandaki oksijen değeri %92-%100 arasında olması gerekmektedir.

3.1.2.6. Kan şekeri

Karbonhidratların sindirilmesi sonucunda ortaya çıkan glikozun kan dolaşımındaki mg/dl cinsinden düzeyine kan şekeri denilmektedir. Birimi miligram/desilitre (mg/dl) ile ifade edilir. Kan şekerinin sağlıklı kişilerde olması gereken değeri, 70-120 mg/dl'dir.

3.1.2.7. Glasgow koma skalası

Glasgow koma skalası (GKS), 1974 yılında Glasgow Üniversitesi'ndeki profesörler tarafından yayınlanmıştır. Hastanın nörolojik değerlendirmesini yapmak için geliştirilmiştir. Hastanın şuur seviyesi belirlenir. Glasgow koma skalası, göz yanıtı skala değeri, sözel yanıtı skala değeri, motor yanıtı skala değeri olmak üzere üç skala değerinin toplanmasıyla bulunur. En iyi göz yanıtı 4 puan, en iyi sözel yanıtı 5 puan, en iyi motor yanıtı 6 puandır. En iyi Glasgow koma skala değeri ise 15 puandır.

3.2. Makine Öğrenmesi

Yapay zeka kavramı ilk olarak 1950'li yıllarda Alan Turing tarafından ortaya çıkarılmıştır. Yapay zekanın öncülerinden biri olan Alan Turing, yapay zekanın sayısal olarak ölçülemeyeceğini; sayısal olarak ölçmenin yerine davranışlarının değerlendirilmesi gerektiğini belirtmiştir (Batal, 2016).

Yapay zekanın davranışlarının değerlendirilmesi için Turing testini önermiştir. Turing testi; bir hakem, bir insan ve bir makine ile gerçekleştirilmektedir. Hakem

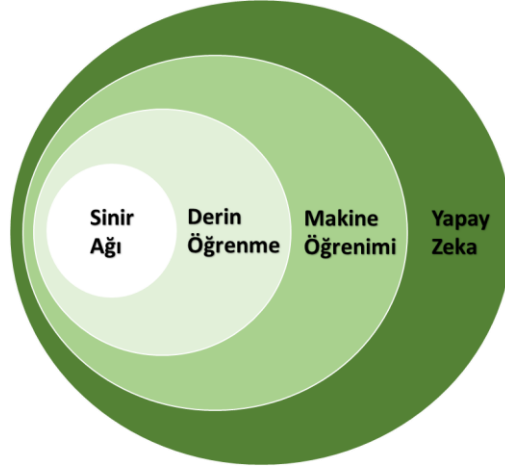
görülmeyecek şekilde yarışmacılara sorular sorar, iki yarışmacı da soruları cevaplamaktadır. Hakem olarak katılanların belirli bir oranı, insan ile makineyi birbirinden ayırt edemezse; o makine yeterince insansı olarak değerlendirilmektedir. Amaç, hangisinin insan, hangisinin bilgisayar olduğunu bulmaktır. Eğer deneyi yapan kişi güvenilir bir şekilde bunu söyleyemez ise, o zaman bilgisayar Turing testini geçer ve insanlar kadar kavrama yeteneğinin olduğu varsayılmaktadır (Batal, 2016).

Yapay zeka, son yıllarda makine öğreniminin ve derin öğrenmenin ortaya çıkmasıyla birlikte hızlı bir evrim geçirmiştir. Bu teknikler; büyük veri kümelerini analiz etme, tahminlerde bulunma, daha önceden elde edilmesi imkansız olduğu düşünülen iç görüler sağlama yetenekleriyle giderek daha popüler hale gelmiştir (Sharifani ve Amini, 2023).

Makine öğrenimi için literatürde birçok tanımlamalar yapılmıştır. Arthur Samuel tarafından çalışmasında; açık olarak programlamadan bilgisayarlara öğrenme yeteneği veren bir çalışma alanı olarak tanımlanmıştır. Örnek verilerin veya geçmiş deneyimlerin kullanılarak, bir performans kriterini optimize edecek şekilde bilgisayarların programlanmasıdır.

Makine öğrenmesi, bir problemi o probleme ait veri setine göre modelleyen algoritmalarının genel adıdır (Atalay ve Çelik, 2017). Makine öğrenimi; yapay zekanın bir alt katmanında olup, derin öğrenmenin ise bir üst kümesinde yer almaktadır. Kümelerin içeriden dışarı doğru sıralaması ise şu şekildedir: Sinir ağları – derin öğrenme – makine öğrenmesi – yapay zeka (Gökalp, 2022). Yani yapay zeka ve makine öğrenimi aslında aynı şey olmamakla birlikte makine öğrenimi yapay zekanın sadece bir alt kümesidir.

Makine öğrenmesi, verideki ilişkileri otomatik biçimde tespit edebilen, tespit edilen bu ilişkileri geleceği öngörmeye veya belirsizlik altında karar vermede kullanan bir dizi yöntemdir (Kartal, 2021). Gelecekle ilgili tahminde bulunmak, çıkarım yapmak, karar vermek ya da strateji belirlemek gibi durumlarda mevcut veriyi dikkate alarak gelecek için değer yaratacak önemli örüntülerin makine öğrenmesi teknikleri sayesinde veriden çıkarılabilir (Kartal, 2021). Makine öğrenimi şeması Şekil 3.5’te verilmiştir (Gökalp, 2022).



Şekil 3.5. Makine öğrenimi şeması.

Makine öğreniminde algoritmalar, büyük veri kümelerindeki ilişkileri bulmak, en iyi kararları vermek, tahminleri yapmak için eğitilirler. Veri setleriyle eğitilen algoritmalar ile tahmin çalışmaları yapılmaktadır. Makine öğrenimi uygulamaları, kullanımlarıyla birlikte gelişir ve ne kadar fazla veriyle eğitilirse o kadar doğru sonuçlar üretmektedir (Gökalp, 2022).

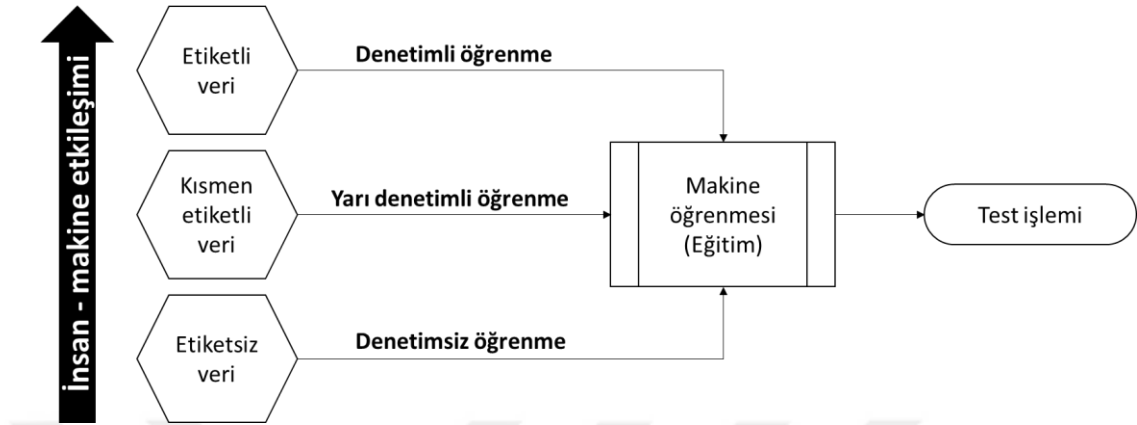
Makine öğreniminin gelişimiyle araştırma ve inovasyon için yeni yollar açarak sağlık, finans, ulaşım ve birçok alanda önemli roller almaya başlamıştır. Çok miktarda veriyi analiz etme ve işleme becerisine sahip olan bu yöntemler, karmaşık sistemleri anlayabilmeyi, bilinçli kararlar verebilmeyi ve gerçek dünya sorunlarına daha verimli bakış açıları kazandırılmasını sağlamaktadır (Sharifani ve Amini, 2023).

Makine öğrenmesinin günümüzde kullanım alanlarına bakıldığında; üretim, tıp, sağlık, finans, vb. Makine öğrenimi; gelişmeyi destekleyerek, yeni gelir akışlarını açığa çıkararak ve zor sorunları çözerek işletmelere yardımcı olmaktadır. Makine öğrenmesi sayesinde veri ile yapılan çalışmalarda insanların gözden kaçırabileceği noktalar yakalanabilir. Modeller insan müdahalesi olmadan çalışabilir. Modeller daha fazla veriyle eğitildikçe ve kullanıldıkça doğruluk oranları zamanla artabilir.

Makine öğreniminin dezavantajlarından bahsedilecek olursa; eğitilmiş model elde etmek için yüksek miktarda veri ile çalışmak daha başarılı olacağından ilk eğitim maliyetli olabilir. Verileri elde etme süreci uzun ve zor bir süreç olabilir. Yeterli veri olmadığında sağlıklı ve başarılı sonuçlar elde edilemeyebilir. Uzman yardımı olmadan elde edilen modellerin yorumlanması, belirsizliklerin ortadan kaldırılması ve test edilmesi zor olabilmektedir.

3.2.1. Makine öğrenmesi türleri

Makine öğrenmesi algoritmaları verilerin etiket durumuna göre Şekil 3.6'daki gibi gruplara ayrılır (Luo ve El Naqa, 2019).



Şekil 3.6. Makine öğrenmesi türleri.

3.2.1.1. Denetimli öğrenme

Makine öğreniminde en çok kullanılan ve en popüler yöntemdir. Her eğitim etiketi bir çıktı etiketi üretir. Etiketli verilerle sistem eğitilir. Eğitimle sistemin sorunları öğrenmesi sağlanır. Verilen eğitimler manuel etiketlerle beslenmesi sağlanır (Gökalp, 2022).

Tahminler yapılırken, regresyon ya da sınıflandırma yöntemi kullanılmaktadır. Gelen mailin spam olduğunu saptamak için bu yöntem kullanılmak istendiğinde, spam olarak sınıflandırılan e-postalar sınıflandırılarak spam olarak değerlendirilecek önemli örnekler makine öğrenimi algoritması beslenmesi gerekmektedir. Kan değerleri sonucunda bir kişinin hasta olup olmadığının tahmin edilmesi için eğitim setinde kan değerleri sonucunda kişinin hasta olup olmadığı etiketlenmiş olmalıdır. Bu etiketler ve eğitim veri seti doğrultusunda algoritmalar ile modeller eğitilir (Gökalp, 2022). Başka bir örnek olarak; etiketli ses kayıtlarından oluşan veri setinde konuşan kişinin kim olduğunu tahmin eden bir ses tanıma sisteminde de denetimli öğrenmeden bahsedilebilmektedir. Bir veri seti ile eğitilerek, bir görselin hangi canlıya ya da hangi insana ait olduğunu tanımlanması da buna benzer bir örnek olarak verilebilmektedir. Geçmiş deneyimlerden elde edilen etiketli verilerden oluşan bir veri setine dayanarak bir arabanın, bir uçağın ya da üretim makinesinin bozulup bozulmayacağını ya da tahmine dayalı ne zaman bozulacağını tahmin çalışmaları da etiketli veriler içermesinden dolayı denetimli öğrenme modelleridir (Somogyi, 2021).

Verileri sınıflandırmak veya sonuçları tahmin etmek için algoritmalar eğitilir. Algoritmalar eğitilirken etiketlenmiş veriler kullanılmaktadır. Denetimli öğrenme, istenen çıktıyı öğretme amacıyla algoritmaların etiketli veri seti ile eğitilmesidir. Bu eğitim veri seti, modelin öğrenmesini sağlayan girdileri ve doğru çıktıları içerir.

Denetimli makine öğrenimi algoritmaları, dışarıdan yardıma ihtiyaç duyan algoritmalarlardır. Giriş veri kümesi eğitim ve test veri kümesine bölünmüştür (Mahesh, 2020). Problem için toplanan veri seti; tahmin edilecek bir bağımlı değişkeni ya da sınıf etiketini içermektedir. Veri seti, eğitim ve test verisi olmak üzere ikiye ayrılmaktadır. Eğitim verisi ile algoritma modelleri eğitilmektedir. Eğitilen tahmin modelinin başarısı ise test verisi ile değerlendirilmektedir.

Tüm algoritmalar, eğitim veri setinden model öğrenir tahmin ya da sınıflandırma işlemi için test veri setine uygular (Mahesh, 2020). Makine öğrenme problemi tanımlanır. Problemin ne olduğu ve çözülmesi için neler gerektiği araştırılır (Kartal, 2021).

- Veri seti makine öğrenmesi sürecine hazırlanır. Ana veri seti; eğitim ve test veri seti olarak ikiye bölünür. Veriler, kullanılacak algoritmalara uygun şekilde düzenlenir.
- Eksik veriler, aykırı değerler, kayıp değerler tespit edilir ve veriler eğitim için hazırlanır. Veri seti dengeli hale getirilir.
- İlgili algoritmalar eğitilerek öğrenmiş modeller oluşturulur.
- Öğrenmiş model ile test verileri test edilir ve elde edilen sonuçlar üzerinden modellerin performansları ölçülür.
- Elde edilen performansların geliştirilmesi için çalışmalar yapılabilir. Örneğin; algoritmaya ait parametrelerin ayarlanması, bagging, boosting, blending gibi yöntemlerden yararlanma.
- Modelleme sonuçlarının yorumlanması.

Denetimli öğrenmenin en önemli avantajları basit ve tasarımının kolay olmasıdır. Ancak, milyonlarca veri kümesinin etiketlenmesi zor olabilir. Sınıflandırma algoritmalarının yaygın kullanım durumları şunlardır (Panesar, 2019):

- Sağlık verilerine göre hastanın hastalık teşhisi
- Bir hasta için en uygun tedavi yolunun önerilmesi
- Bir klinikte gün içinde en yoğun kabullerin olduğu dönemin tanımlanması

- Hastaların hastaneye yeniden yatış durumunun belirlenmesi
- Hastalığı tanımlamak için görüntüleri kullanarak sınıflandırmak. Örneğin, bir fotoğraftan bir hastanın aşırı kilolu olup olmadığını belirlemek.

Denetimli bir diğer adıyla gözetimli öğrenmede kullanılan sınıflandırma algoritmalarından bazıları aşağıdaki gibidir.

- K- En yakın komşu
- Lojistik regresyon
- Doğrusal destek vektör makineleri
- Doğrusal olmayan destek vektör makineleri
- Karar ağaçları
- Rastgele ormanlar
- Naive bayes

3.2.1.2. Denetimsiz öğrenme

Denetimsiz öğrenme, etiketlenmemiş veri seti kullanarak verileri kümelemek için makine öğrenimi algoritmalarını kullanmaktadır. Bu algoritmalar, herhangi bir müdahaleye veya veri etiketine ihtiyaç duymamaktadır (Petekçi, 2021).

Modelin eğitimi için etiketsiz veriler kullanılır. Algoritmalar, etiketsiz verilerden bunları anlamaya, tanımaya ve keşfetmeye çalışmaktadır. Sistemden istenilen sonuçlar bu nedenle bilinmemektedir. Eğitim ve gözetim dışı sonuçlar üretmektedir. Bu modelle doğru cevap ile eğitilmemiştir. Kendi başına kalıpları bulması gerekmektedir (Gökalp, 2022). Eğitim verisi, etiketleme gerektirmediği için kurulum süreci kolaydır.

Denetimsiz öğrenmede ise tahmin edilecek bir sınıf etiketi ya da bağımlı değişken bulunmamaktadır. Problem için mevcut verinin tamamı eğitim verisi olmaktadır. Algoritmaların amacı; ölçümlerin ve gözlemlerin kümelere ayrılmasıdır.

Denetimsiz öğrenmede algoritmalar metinleri konularına göre kategoriler altında gruptandırabilmektedir. Bir yazıdaki anlamı ve duyguyu anlamak için doğal dil işleme araçlarını kullanırlar. Market zincirlerinde yapılan denetimsiz makine öğrenimi çalışmaları sonucunda “bir müşteri kola satın alıyorsa büyük ihtimalle cips de satın alacaktır.” gibi yorumlar elde edebilmektedir.

Online olarak alışveriş yapan insanların geçmiş satın almaları göz önüne alınarak ve diğer özelliklerine göre gruptandırarak, online alışveriş sayfalarında önerilerin

sunulması, sosyal medyada birbirini tanıyor olabilecek insanların tespit edilmesi ve kullanıcılara bu kişilerin öneri olarak sunulması, dolandırıcılık ya da kötüye kullanım tespit edilme durumları denetimsiz öğrenme örnekleri olarak verilebilir (Somogyi, 2021). Denetimsiz bir diğer adıyla gözetimsiz öğrenmede kullanılan algoritmalarından bazıları aşağıdaki gibidir.

- Öbekleme: K- Merkezli öbekleme (k-Means), Beklenti büyütme (Expectation Maximization).
- Boyut azaltma: Temel Bileşen Çözümlemesi (Principal Component Analysis).
- Birliktelik kuralları: Apriori, Eclat.

3.2.1.3. Yarı denetimli öğrenme

Yarı denetimli öğrenme, algoritmaları eğitmek için az miktarda etiketlenmiş veri seti ile büyük miktarda etiketlenmemiş verinin birlikteliğini kullanan makine öğrenimidir. Özetle; etiketli eğitim verilerini kullanan denetimli makine öğrenimi ile etiketsiz verileri kullanan denetimsiz öğreniminin birleşimidir (Petekçi, 2021). İlk olarak etiketli veriler kullanılarak algoritma kısmen eğitilmektedir. Kısmen eğitilen model ile etiketsiz veriler etiketlenmektedir. Daha sonra ortaya çıkan yeni veri seti ile model tekrar eğitilmektedir. Resim, metin, biyoinformatik alanlarında yani etiketlenmemiş verilerin çok olduğu alanlarda kullanımının arttığı görülmektedir (Petekçi, 2021).

3.2.1.4. Pekiştirmeli öğrenme

Pekiştirmeli öğrenme diğer adıyla takviyeli öğrenme, karmaşık bir hedefe nasıl ulaşılması gerektiğini ya da birden fazla adımda belirli bir boyut doğrultusunda nasıl maksimize edileceğini öğrenen, hedefe yönelik algoritmalar (Petekçi, 2021). En fazla ödül elde etmek için ortamdaki en uygun davranışı öğrenmeyi amaçlar. Algoritmanın amacı, mümkün olan en fazla ödül puanı biriktirerek hedefe varmaktır. Video oyunları bu algoritmalara örnek olarak gösterilebilmektedir. Pekiştirmeli öğrenme, canlıların öğrenme yapısına benzemektedir. Örnek olarak; bir bebek yürümeyi öğrenirken defalarca düşer ve her düşüşünde daha fazla tecrübe kazanmaktadır ve tecrübelerin sonunda yürümeyi öğrenir. Bu model; denetimli, denetimsiz ve yarı denetimli öğrenme modellerinden farklıdır (Doğan, 2023).

3.2.2. Makine öğrenmesi sınıflandırma algoritmaları

Bu çalışmada makine öğrenmesi sınıflandırma algoritmaları kullanılmıştır. Kullanılan sınıflandırma algoritmaları hakkında bilgi verilmiştir.

3.2.2.1. Karar ağacı algoritması

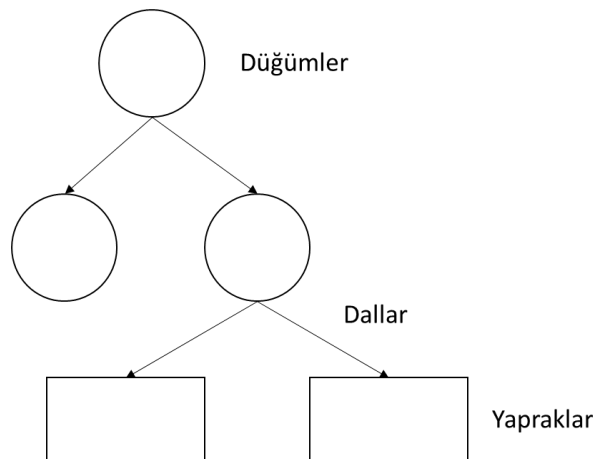
Karar ağacı algoritması, denetimli öğrenme algoritmasıdır. Karar ağaçları, özellik ve hedefe göre karar ve yaprak düğümlerinden oluşan ağaç yapısında model oluşturan bir sınıflandırma yöntemidir (Başer ve ark., 2021). Karar Ağacı, seçimleri ve bu seçimlerin sonuçlarını bir ağaç şeklinde temsil eden bir grafiktir (Mahesh, 2020).

Bir karar ağacı; çok fazla veri içeren bir veri setini, bir dizi karar verme kuralları ile daha küçük kümeler bölerek kullanılan bir algoritmadır. Karar ağacı; görselleştirilebildiği için anlaması ve yorumlaması kolay olan bir algoritma olmakla birlikte kayıp değerleri desteklememektedir. Eğer veri setinde kayıp veriler varsa veri setinden çıkartılmalı ya da algoritma kayıp verilerle çalışabilecek şekilde düzenlenmelidir. Aynı zamanda sayısal ve kategorik verilerin ikisi de işlenebilmektedir. Zaman zaman veriyi net açıklayamayan uzun ve karmaşık ağaçlar üretilebilir. Bu dallar takip edilemeyebilir. Bir karar ağacı algoritması kullanıldığında ezber öğrenme problemi yaşanabilmektedir. Ezber öğrenme problemi yaşanmaması için kısıtlama ve budama yöntemleri kullanılabilir.

Karar ağacının temel yapısını aşağıdaki 3 bölüm oluşturmaktadır (Bilgin, 2021).

- Düğümler
- Dallar
- Yapraklar

Karar ağacının yapısı Şekil 3.7’de gösterilmiştir (Singh ve ark., 2022).



Şekil 3.7. Karar ağacı yapısı.

Her ağaç düğümlerden ve dallardan oluşur. Her düğüm, sınıflandırılacak bir gruptaki nitelikleri temsil eder ve her dal, düğümün alabileceği bir değeri temsil eder (Singh ve ark., 2022).

Karar ağaçları, verileri kök düğümden başlayarak son yaprak düğüme kadar yukarıdan aşağıya bir yaklaşımla sıralayarak verilerin sınıflandırılmasını gerçekleştirir (Rastogi ve ark., 2020).

Karar ağacı algoritmasında; düğümlerde ilgili soruları sorup, alınan cevaplara göre hangi dallar üzerinden hangi düğüme gidileceğine karar verilerek, ulaşılabacak en son yaprağa en kısa yoldan ve en kısa sürede ulaşmak amaçlanır. Karar ağacında verilen karar, en son gidilen yaprağa göre verilmektedir. Karar ağacı yapısına ait örnek Şekil 3.8’de verilmiştir (Mahesh, 2020).



Şekil 3.8. Karar ağacı örneği.

Her öz nitelik ağaç yapısı oluşturulurken bir düğüm ile ilişkilendirilip, kök ve düğümler arasındaki yapı ise dallardır. Bir düğümden dallar ile diğer düğüme geçilir (Sindayigaya ve Dey, 2022). Karar ağacı oluşturmada, temel algoritmanın özellikleri aşağıdaki gibi sıralanabilir (Mahesh, 2020):

- Ağaç yukarıdan aşağıya doğru, tekrar eden bir şekilde böl ve çöz metodu ile oluşturulur.
- Başlangıçta, tüm eğitim örnekleri tek bir köktedir.

- Ağaç, tüm verileri içeren bir düğüm ile başlar.
- Nitelikler kategoriseldir (Eğer sürekli veriler var ise ayrıklaştırılır.)
- Eğer örneklerinin hepsi aynı sınıfa aitse düğüm yaprak olarak sonlanır ve sınıf etiketini alır.
- Eğer değilse, örnekleri sınıflara en iyi ayıracak olan nitelik seçilir.
- Nitelikler; sezgiseller, istatistiksel ölçütler temelinde seçilir (Örnek: Entropi)

Karar ağacı ile sınıflandırma algoritmalarında durma koşulları ise aşağıdaki gibi sıralanabilir:

- Bir düğüm için tüm örnekler aynı sınıfa ait ise.
- Örnekleri bölecek nitelik kalmamış ise.
- Kalan niteliklerin değerini taşıyan örnek yok ise.

Safsızlığı ölçmenin en yaygın iki yolu Gini Endeksi ve Entropi Ölçümüdür. Entropiye dayalı algoritmalarda, karar ağacının dallanması en yüksek bilgi kazancına (en düşük entropiye) sahip nitelikten başlanarak gerçekleştirilir. Entropi; rassallığı ve belirsizliği ve tahmin edilmeyen durumun ortaya çıkma ihtimalini göstermektedir. Bir sistemin düzensizliğini ve belirsizliğini ifade etmektedir (Aytekin ve ark., 2018). Bir süreç için tüm örnekler tarafından içerilen enformasyonun beklenen değeridir. Enformasyon ise rastgele bir olayın gerçekleşmesiyle ilişkili bir ölçüttür.

Entropi Ölçüsü; 0 (tüm gözlemler aynı sınıfa ait olduğunda) ile $\log_2(m)$ (tüm m grupları eşit oranda temsil edildiğinde) arasında değişir (Kushwah ve ark., 2022).

Bir olayın gerçekleşme olasılığı $P(A)$ ise, A olayını tanımlayan entropi $-\log_x P(A)$ şeklinde tanımlanmaktadır (Aytekin ve ark., 2018).

Bir olay ile ilgili ortaya çıkabilecek olan birden fazla durum varsa, gerekli enformasyon miktarı tüm olasılıklara bağlı olup entropi denklem 3.1'deki gibi ifade edilmektedir (Luo ve El Naqa, 2019).

$$Entropi H = - \sum_{k=1}^m P(s_k) \log_2 P(s_k) \quad (3.1)$$

Entropi, bir olaydaki kesinsizliğin yani bir olayın sonucunun kesin olarak belirlenmesi için gerekli olan enformasyon miktarını ifade ederken, bilgi kazanımı terimi ise entropinin tersi olacak şekilde formüle edilmektedir (Maimon ve Rokach, 2014). Bilgi

kazanımı (information gain), 0 ile 1 arasında bir değer almakla birlikte iki entropi arasındaki fark olarak tanımlanmaktadır. Bilgi kazanımı yüksek olan öznitelikler, modeldeki en kıymetli öznitelikler olarak yorumlanmaktadır. Bilgi kazanımları büyükten küçüğe doğru sıralanmaktadır. Buna göre seçilerek adımlar yürütülmektedir. Bilgi kazanımı formülü ise denklem 3.2'deki gibidir (Aytekin ve ark., 2018).

$$\text{Bilgi kazanımı}(\text{Öznitelik}) = \text{Entropi } H(\text{Veri seti}) - \text{Entropi } H(\text{Öznitelik}) \quad (3.2)$$

Safsızlığı ölçmenin bir diğer yolu ise Gini indeksidir. p_k , j bölümündeki k sınıfının gerçekleşme olasılığıdır. j bölümündeki tüm gözlemler aynı sınıfa aitse gini indeksi 0 olur. Bu durum ise mükemmel saflığı göstermektedir. Ayrıca, j sınıfında tüm m grupları eşit oranda temsil edildiğinde $(m - 1)/m$ gini indeksi maksimum değere ulaşır. Gini indeksinin hesaplama yöntemi denklem 3.3'teki gibidir (Maimon ve Rokach, 2014).

$$\text{Gini indeksi } GI = 1 - \sum_{k=1}^m p_k^2 \quad (3.3)$$

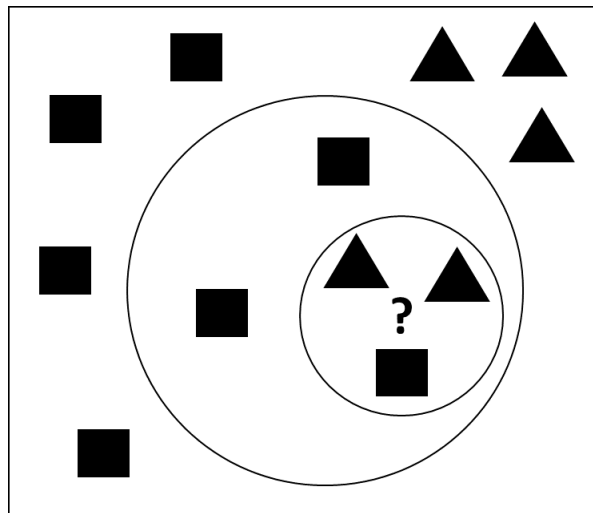
Karar ağaçları oluşturmak için, entropiye dayalı algoritmalar, sınıflandırma ve regresyon ağaçları, bellek tabanlı sınıflandırma yöntemleri olmak üzere makine öğrenmesi başlığı altında birçok yöntem geliştirilmiştir. En çok kullanılan karar ağacı algoritmaları aşağıdaki gibidir (Çalış ve ark., 2014).

- ID3 algoritması. Entropiye dayalı bir karar ağacı algoritmasıdır. Bu algoritmada, veri kümesinde tüm değerlerin nominal olması ve bilinmeyen değerlerin olmaması gerekmektedir. Sadece kategorik veri ile çalışan bir yöntemdir.
- C4.5 ve C5.0 algoritması: ID3 karar ağacı algoritmasının geliştirilmiş, nümerik değer içeren veriyle uygulanabilen halidir. Nümerik verilerin kategorik hale getirilmesini sağlayan bir yöntemi bulunmaktadır. C5.0 algoritması C4.5 algoritmasına göre daha hızlıdır ve hafızayı daha verimli kullanmaktadır.
- CART: Hem sayısal hem de nominal değerdeki veri setleriyle çalışmaktadır. Sınıflandırma ve regresyon problemleri için bir çözüm yöntemidir. İkili olarak özyinelemeli biçimde bölünen bir yapıdadır. Algoritma çalışırken gini indeksinden yararlanmaktadır.

- CHAID: Ki-kare testlerini kullanarak bölme işlemi gerçekleştirmektedir. Dalların sayısı iki ile tahmin edici kategori sayısı arasında değişmektedir.
- QUEST: İkili karar ağacı yapısı kullanan bir sınıflandırma algoritmasıdır. İkili ağaç kullanmasının en önemli sebebi, ikili ağaçlarda doğrudan durma ve budama gibi tekniklerin kullanılmasıdır. Ağacın dallanması sırasındaki önyargılı seçimin genelleştirilmesi ve hesaplama maliyetinin düşürülmesi amacıyla geliştirilmiştir.

3.2.2.2. K-en yakın komşu algoritması

K-En yakın komşular algoritması, ilk olarak 1950'lerin başında ortaya atılmıştır (Dilki ve Başar, 2020). Denetimli ve denetimsiz öğrenme sağlayan en basit makine öğrenmesi algoritmalarından biridir. KNN, sınıfı bilinen eğitim veri setindeki örneklerin uzaklık ölçüsüne dayalı olan denetimli öğrenme tekniklerinden biridir (Kemalbay ve Alkış, 2020). K-En yakın komşular algoritmasının fikri şudur: eğer özellik uzayındaki bir örneğin K en yakın örneğinin özellik uzayındaki en yakın komşuları belirli bir kategoriye aitse, o zaman örnek de bu kategoriye aittir (Xing ve Bei, 2019). KNN, eğitim kümesindeki örneklerin her birinin sınıflandırılacak örnek değerine olan uzaklıklarının hesaplanması ve en küçük uzaklığa sahip k sayıda gözlemin seçilmesi prensibine dayanmaktadır (Xing ve Bei, 2019). K-en yakın komşu algoritmasının görselleştirilmiş örneği Şekil 3.9'da gösterilmiştir (Artificial Intelligence Technology, 2021).



Şekil 3.9. K-En yakın komşu algoritması örneği.

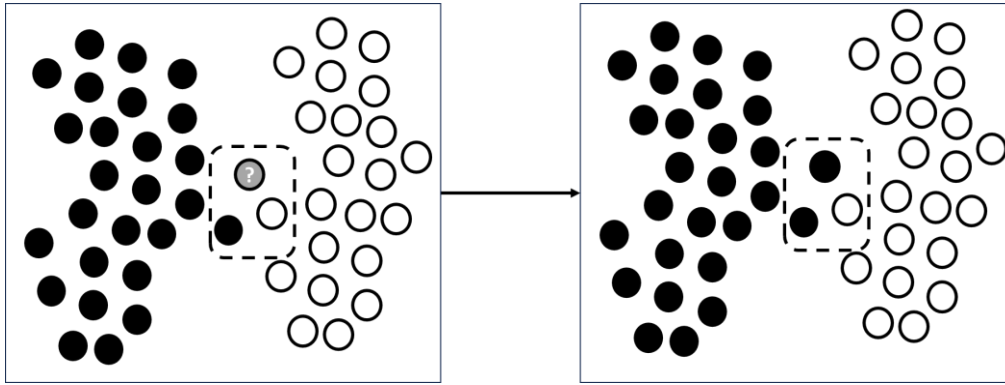
KNN algoritmasının temel konsepti, kısa bir mantıkla "kırmızı mürekkebin etrafındakiler kırmızıya, siyah mürekkebin etrafındakiler siyaha döner" şeklindedir (Artificial Intelligence Technology, 2021). KNN algoritmasının özetle işleyişi;

- İlk olarak k sayısı belirlenmektedir.
- Belirlenen değişkenler üzerinde yeni gözleme mesafeye göre en yakın k kayıt incelenmektedir. Kullanılan mesafe ölçütü veri türüne bağlı değişiklik gösterebilmektedir. Kullanılabilecek uzaklık ölçütleri aşağıda açıklanmıştır.

K-komşu sayısı: K, bilinmeyen noktanın en yakın komşularının sayısıdır. Modelin doğruluk oranı k değerine göre değişkenlik göstermektedir. K değerine göre modelin sınıflandırma verileri de şekillenmektedir. Bu sebeple k parametresi sınıflandırma çalışması için oldukça önemlidir (Chakravarthy ve ark., 2023).

K değeri ne kadar büyük olursa, sınırlar bulanık olur. Bulanık olması nedeniyle yetersiz uyum yaratma olasılığı oluşur. K değeri ne kadar küçükse, karar sınırları o kadar keskin olur ve aşırı uyumun oluşma olasılığı oluşur (Artificial Intelligence Technology, 2021).

Yeni bir örnek görüldüğü durumda, eğitim verilerinden söz konusu örneğe en yakın k tane örnek tespit edilerek yeni örneğin sınıfı, k en yakın komşusunun sınıf etiketlerinin çoğunluk oylamasına göre belirlenmektedir (Taşcı ve Onan, 2016). KNN örneği Şekil 3.10'da gösterilmiştir (Esmer ve ark., 2020).



Şekil 3.10. K-En yakın komşu algoritması örneği.

KNN algoritmasında verilerin tiplerine göre uzaklık çeşidi tercih edilir (Taşcı ve Onan, 2016). KNN algoritmasında kullanılan uzaklık türleri aşağıdaki gibi özetlenmiştir.

$x_i = (x_1, x_2, x_3, \dots, x_k)$, $y_i = (y_1, y_2, y_3, \dots, y_k)$ iki boyutlu veri nesnesi olmak üzere uzaklık ölçüleri aşağıdaki gibidir.

Minkowski uzaklığı: Öklid uzayında tanımlı bir dizidir. Formülü denklem 3.4'teki gibidir (Taheri ve ark., 2020).

$$Minkowski\ uzaklığı = \left(\sum_{i=1}^k (|x_i - y_i|)^q \right)^{1/q} \quad (3.4)$$

Öklid uzaklığı: Sınıflandırma algoritmalarında en çok kullanılan uzaklık ölçütüdür. Minkowski uzaklığında; $q=2$ olduğunda Minkowski uzaklığı, Öklid uzaklığı halini alır. Öklid uzaklığı, Minkowski uzaklığının özel bir durumudur. Minkowski uzaklığının formülü denklem 3.5'teki gibidir (Zhao ve ark., 2022).

$$Öklid\ uzaklığı = \left(\sum_{i=1}^k (|x_i - y_i|)^2 \right)^{1/2} = \sqrt{\sum_{i=1}^k (|x_i - y_i|)^2} \quad (3.5)$$

Manhattan uzaklığı: N boyutlu iki nokta arasındaki farkların mutlak değerlerinin toplamıdır. Minkowski uzaklığında; $q=1$ olduğunda Minkowski uzaklığı, Manhattan uzaklığı halini alır. Manhattan uzaklığı, Minkowski uzaklığının özel bir durumudur. Kategorik veriler için kullanışlı bir uzaklıktır (Şanlı, 2023). Manhattan uzaklığının formülü denklem 3.6'daki gibidir (Ehsani ve Drabløs, 2020).

$$Manhattan\ uzaklığı = \sum_{i=1}^k (|x_i - y_i|) \quad (3.6)$$

Chebyshev uzaklığı: Minkowski uzaklığının $n \rightarrow \infty$ olduğu özel durum olup, iki nokta arasındaki farkların mutlak değerlerinin en fazla olanı olarak tanımlanmaktadır. Minkowski uzaklığında; $q \rightarrow \infty$ olduğu durumda Chebyshev uzaklığı elde edilir. Chebyshev Mesafesi, uzaydaki bir boyut boyunca en uzun mesafeyi oluşturan uzaklığı belirtir. Başka bir ifade ile iki N boyutlu vektörün bir boyuttaki en uzun mesafesidir. Chebyshev uzaklığı formülü denklem 3.7'deki gibidir (Ehsani ve Drabløs, 2020).

$$Chebyshev\ uzaklığı = \left(\sum_{i=1}^k (|x_i - y_i|)^\infty \right)^{1/\infty} = \max_i (|x_i - y_i|) \quad (3.7)$$

Mahalanobis uzaklığı: Hintli istatistikçi Mahalanobis tarafından önerilen bir uzaklık ölçütüdür. Bu uzaklık kullanılacağı zaman tüm veri setinin de sayısal değerler

içerdiğinden emin olunmalıdır (Atcılı, 2023). Değer vektörü ile ortalama farkının, kovaryans matris tersi ve değer vektörü ile ortalama farkının transpozunun çarpımlarının karekökü olarak hesaplanmaktadır. Diğer uzaklık ölçütlerinden en büyük farkı, mesafe ayırımını elipsel bir düzlem üzerinde yapmasıdır. Mahalanobis uzaklığının formülü denklem 3.9'deki gibidir. Formüle bakıldığında; x örnek veri, μ örnek popülasyonun ortalaması, S^{-1} ters kovaryans matrisidir. kovaryans matrisi bir birim matris olduğunda Mahalanobis mesafesi Öklid mesafesine eşdeğerdir. Kovaryans matrisi köşegen bir matris olduğunda Mahalanobis mesafesi standartlaştırılmış Öklid mesafesine eşdeğerdir (Tang ve ark., 2023).

$$\text{Öklid uzaklığı} = \sqrt{(x - \mu)^T (x - \mu)} \quad (3.8)$$

$$\text{Mahalanobis uzaklığı} = \sqrt{(x - \mu)^T S^{-1} (x - \mu)} \quad (3.9)$$

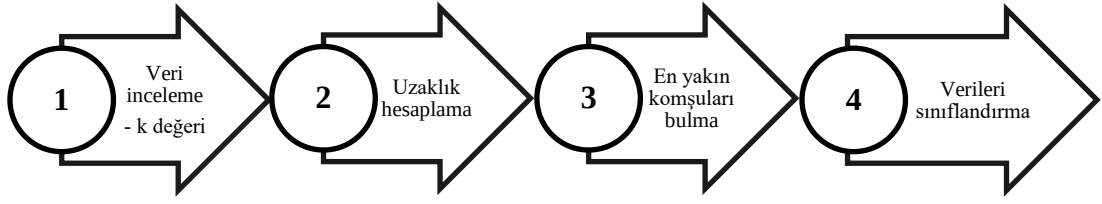
Dilca uzaklığı: Dilca uzaklığı, kategorik değişkenler arasındaki uzaklığı ölçmek için kullanılan iki aşamalı bir ölçüttür. (Kılınç ve ark., 2016).

Hamming uzaklığı: İki sekans arasındaki düzenleme mesafesini ölçmek için kullanılan birkaç dizi metriğinden biridir. Adını Amerikalı matematikçi Richard Hamming'den almıştır. İki ikili veri dizisinin karşılaştırılması için kullanılmaktadır. Hamming uzaklığı, tüm verileri gözden geçirip ve kaç özelliğin farklı olduğunun sonucunu vermektedir (Ehsani ve Drabløs, 2020).

K komşu değeri dikkate alınarak k adet en yakın uzaklığa sahip noktalar seçilir. K adet sınıf kategorisi öğrenilen eğitim veri noktaları içerisinde yeni sınıfı belirleyebilmek için ağırlıklı oylama yöntemi tercih edilir. Eğitim verilerindeki k adet en yakın komşunun oyu yeni veri setinin noktası x_z olan uzaklık ile ters orantılıdır ve denklem 3.19'da verilmiştir (Taşcı ve Onan, 2016).

$$oy(x_m) = \begin{cases} \infty, & d(x_m, x_z) = 0 \text{ ise} \\ \frac{1}{d(x_m, x_z)}, & \text{diğer durumlarda} \end{cases} \quad (3.10)$$

Denklemde yer alan tüm değerler toplanır ve yeni veri setinin kategorisi, en yüksek ağırlıklı puanlamaya sahip olan sınıfa atanır. Elde edilecek doğruluk oranına göre en uygun k değeri deneyerek bulunabilir (Xing ve Bei, 2019). KNN akışı ise Şekil 3.11'deki gibi özetlenebilir (Dilwani, 2019).



Şekil 3.11. K-en yakın komşu algoritması işlem adımları.

3.2.2.3. Naive bayes algoritması

Naive Bayes algoritması; bir olayın meydana gelme olasılığını hesaplamak için Bayes teoremini kullanır. Tüm değişkenlerin birbirinden bağımsız olduğunu varsayması sebebiyle algoritma saf olarak kabul edilmektedir. Belirli bir eşik verildiğinde, bu yöntem bir vektörün bir sınıfta mı yoksa başka bir sınıfta mı olduğunu olasılıksal olarak sınıflandırmak için kullanılabilir (Panesar, 2019).

Naive Bayes sınıflandırıcısı, bir sınıfta belirli bir özelliğin varlığının başka herhangi bir özelliğin varlığıyla ilişkili olmadığını varsaymaktadır (Mahesh, 2020).

Yani değişkenlerin birbirinden bağımsız olduğu varsayımını yapar ve yine de iyi bir sınıflandırıcı olduğunu kanıtlar. Naive Bayes, metnin sınıflandırılmasına odaklanır ve öncelikle kümeleme ve sınıflandırma gerçekleştirmek için kullanılır. Olasılıksal bir sınıflandırıcıdır ve belirli bir nesneyi sınıflandırma olasılığını kullanmaktadır (Rastogi ve ark., 2020).

Naive Bayes'in pozitif yönleri; basit bir şekilde uygulanması ve doğru sonuçlar elde edilmesini sağlarken, negatif yönlerinden biri ise sınıf ve nitelik bakımından bağımsızlık kabulüne ihtiyaç duymasıdır. Olayların birbirinden bağımsız olduğu kabul edilmektedir (Başer ve ark., 2021).

Naive Bayes, Bayes teoremi esaslı olup bu algoritma özelliklerin birbirinden bağımsız olduğunu kabul eder. Sınıflandırma problemi, maximum $P(C = c|X)$ elde etmeyi amaçlamaktadır. D eğitim kümesi ve C sınıfı denklem 3.11'de verilmiştir. D eğitim kümesinde, X örnek özelliği göz önüne alındığında örneğin c sınıfına ait olma olasılığı denklem 3.12'de verilmiştir (Artificial Intelligence Technology, 2021).

$$D = \{X = (x_1, x_2, x_3, \dots, x_n), \quad C = (C_1, C_2, C_3, \dots, C_m)\} \quad (3.11)$$

$$P(C = c|X) = \frac{P(X|C = c)P(C = c)}{P(X)} \quad (3.12)$$

$P(C = c|X)$: Son olasılık. X verileri göz önüne alındığında, C hipotezinin doğru olma olasılığı. C hipotezinin meydana gelen X olayı üzerindeki olasılığı.

$P(C = c)$: Hedefin önsel olasılığı, veri setindeki c kategorisine sahip örneklerin oranı, X ile hiçbir ilgisi yoktur. Tahmin edici öncelikli olasılık.

$P(X)$: Özelliğin önsel olasılığı. Sınıf öncelikli olasılık. Normalde $P(X)$ 'i dikkate alınmayabilir çünkü $P(X)$ sınıflandırma sırasında sabit bir değer olarak görülebilir.

Naive bayes sınıflandırıcısında, niteliklerin hepsi eşit derecede öneme sahip ve birbirinden bağımsızdır. Nitelik koşullu bağımsızlık varsayımı, bir koşul olarak örnek sınıflandırma göz önüne alındığında, her bir nitelik değerinin dağılımının diğer nitelik değerlerinin dağılımından bağımsız olduğunu belirtir. Naive Bayes'in saf olmasının nedeni tam olarak modelinde kullanılan nitelik bağımsızlığı varsayımından kaynaklanmaktadır. Bu varsayımın yapılması, hesaplamayı etkili bir şekilde basitleştirir ve Bayes sınıflandırıcısına büyük veri tabanlarında daha yüksek doğruluk ve eğitim hızı sağlar (Artificial Intelligence Technology, 2021).

Naive Bayes modelinin performansı genellikle özellik bağımsızlığı hipotezinin ne ölçüde karşılandığına bağlıdır. Örnek olarak boy ve kilo gibi iki özellik tamamen bağımsız değildir. Bu korelasyon kaçınılmaz olarak modelin doğruluğunu etkileyecektir ancak korelasyon büyük olmadığı sürece Naive Bayes modeli kullanılmaya devam edilebilir. Pratik uygulamalarda farklı özellikler nadiren tamamen bağımsızdır (Artificial Intelligence Technology, 2021).

Naive bayes sınıflandırıcısı; uygulanması kolay ve çoğu durumda iyi sonuçlar elde ettiği için literatürde avantajlı görülmektedir. Naive Bayes'in son derece karmaşık sınıflandırma yöntemlerinden bile daha iyi performans gösterdiği bilinmektedir (Rawal ve Lal, 2023). Değişkenler arasında bağımsızlık varsayımı, doğruluk kaybına neden olmaktadır. Bazı durumlarda değişkenler arasında bağımlılık durumu olabilir ancak bu bağımlılık Naive bayes modellenemez.

3.2.2.4. Rastgele orman algoritması

Rastgele orman algoritması, 2001 yılında Leo Breiman ve Adele Cutler tarafından geliştirilen, birden fazla karar ağacının çalışması mantığına dayanan bir topluluk öğrenme algoritmasıdır (Cihan, 2018).

Birden fazla karar ağacı çalıştırıldıktan sonra, karar ağaçlarından elde edilen sonuçların ortalaması alınarak bir sınıflandırma bir tahmin elde edilir. Birden fazla karar ağacı çalıştırıldığı için karar ağaçları algoritmasında karşımıza çıkan aşırı öğrenme(overfitting) problemi en aza indirmektedir (Sevli, 2019).

Karar ağaçları algoritmasına yapılan en önemli eleştirilerden biri bir soruya cevap verip, eğitim seti bölündükten sonra bu sorunun tekrar değerlendirilmemesidir. Örnek olarak sorulan soru cinsiyet durumu ise karar ağacı kadınlar ve erkekler olarak ayrılmaktadır. Algoritma sonrasında ya hep kadınlar ya da hep erkekler üzerinden ilerlemektedir. Cinsiyetten bağımsız durumları değerlendirmemektedir. Ancak rastgele orman algoritması birden fazla karar ağacı algoritması çalıştırdığı için her adımda rastgelelik oluşturmayı amaçlamaktadır.

Gerçek hayattan örneklendirilecek olunursa; bir kişi, bir seyahate çıkmayı planlamakta ve gideceği yerdeki en iyi yemek restoranlarını listelemek istemektedir. Bunun için oraya daha önce gitmiş bir arkadaşına en iyi restoranlarla ilgili rastgele sorular sormuştur. Sonrasında arkadaşı tavsiyede bulunmuştur. Kişi, buraya kadar karar ağacı algoritmasını çalıştırmıştır. Ancak almış olduğu tavsiye ile en sağlıklı sonuca ulaşamayacağını düşünmüş ve birden fazla arkadaşına aynı konuyla ilgili rastgele sorular sormuştur ve böylece birden fazla karar ağacı algoritması çalıştırarak rastgele orman algoritmasını çalıştırmıştır. Aldığı tüm tavsiyelerin en sık tekrar eden sonucunu analiz edip kararını vermiştir.

Rastgele orman algoritması, girdi değişken sayısının fazla olduğu verilerde iyi performans göstermesiyle birlikte, eksik verilerin olduğu veri setlerinde de iyi performansla tahmin yapmasından dolayı kolaylıklar sağlamaktadır (Başer ve ark., 2021). Rastgele orman algoritması, modeli oluştururken örneklem boyutunun parametresini belirleyerek her sınıftan kaçar adet örneklem yapacağını yerine koyma yöntemi ile belirlemektedir. Bu parametre, rastgelelik etkisi sağlayarak ormandaki her ağaç verisinin farklı yönlerini görmektedir. Bu şekilde dengesizlik problemini aşmayı ve modelin performansının artırılması amaçlanmaktadır (Başer ve ark., 2021).

Rastgele orman algoritmasının avantajları ve güçlü yönleri (Pandimurugan ve ark., 2022);

- Diğer sınıflandırma algoritmalarına göre daha iyi performans göstermektedir.
- Eksik ya da gürültülü veri setlerinde iyi sonuçlar elde edilebilmektedir.

- Yalnızca en önemli özellikleri seçer.
- Girdi değişken sayısı fazlayken de iyi bir performans elde edilebilmektedir.
- Aşırı öğrenme, ezberlemeye karşı duyarlıdır.
- Model oluştuktan sonra ağaçta budama işlemi yapmaya gerek duymamaktadır.
- Veri sayısı büyük ve küçük olan modellerde iyi sonuçlar elde edilebilmektedir.
- Dengeli olmayan veri setlerinde de kullanılabilir.

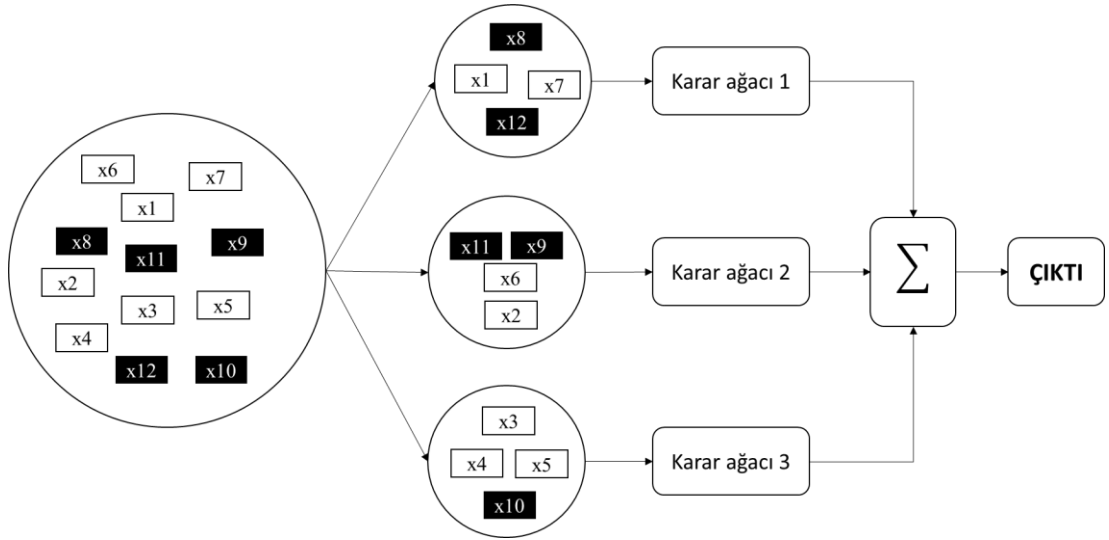
Rastgele orman algoritmasının dezavantajları ve zayıf yönleri (Pandimurugan ve ark., 2022);

- Birden fazla karar ağacı içerdiği için model basit ve yavaş olmayan bir şekilde yorumlanamayabilir.
- Modeli veriye göre ayarlamak gerekmektedir bu nedenle ön işleme adımı vardır.
- Model sonunda güven aralığı elde edilememektedir.
- Hafızasında birden fazla karar ağacının sonucunu tuttuğu için bellek gereksinimi artmaktadır.

Rastgele orman algoritması uygulama aşamaları aşağıdaki gibidir (Radhika ve ark., 2022);

- Orijinal veri setinden N adet örneklen alınmaktadır. Bu örneklem bootstrap yöntemi ile alınmaktadır. Bootstrap yöntemi; parametrik olmayan koşulların olduğu durumda kullanılan tekrarlı öğrenme yöntemlerinden biridir. Bu yöntem sayesinde örneklerin birden fazla seçilmesine izin verilmektedir.
- Her bir örneklemin üçte biri eğitim, üçte biri test verisi olarak planlanmaktadır.
- Her düğüm için en iyi dallara bölünecek sınıf değişkenin rastgele seçilmesi
- En iyi dallanma kriteri, seçilen sınıf değişkeni için hesaplanır. Bulunan bilgiye göre veri seti her düğümde iki alt dala ayrılmaktadır. Yaprak düğüm elde edilene kadar tekrar edilmektedir.
- Oluşturulan ağaçların tahminleri birleştirilerek yeni bir tahmin oluşturulur. Sınıflama ağaçları için en fazla oyu alan, regresyon ağaçları için ortalama alınarak nihai tahmin sonucu elde edilir.

Rastgele orman algoritmasının mimarisi Şekil 3.12’de verilmiştir (Radhika ve ark., 2022).



Şekil 3.12. Rastgele orman algoritması mimarisi.

3.2.2.5. Doğrusal destek vektör makineleri algoritması

Destek Vektör Makineleri, sınıflandırma çalışmalarında gözetimli öğrenme tekniklerinden biridir. SVM hem sınıflandırma hem de regresyon çalışmalarında kullanılmaktadır ancak en çok sınıflandırma yöntemi olarak kullanılan bir makine öğrenmesi algoritmasıdır.

Hem doğrusal sınıflandırma yapabilmekte hem de çekirdek fonksiyonlarının da kullanılmasıyla doğrusal olmayan sınıflandırma da yapabilmektedir. Bu sebeple destek vektör makineleri doğrusal ve doğrusal olmayan olmak üzere ikiye ayrılır (Panesar, 2019).

SVM; tahmin çalışmaları, duygu analizi, metin analizi, görüntü tanıma, el yazısı tanıma çalışmalarında kullanıldığı görülmektedir (Rad ve ark., 2023).

Doğrusal destek vektör makineleri bir düzlemdeki noktaları sınıflandırmak için doğru çizerek, oluşan sınıfların noktalarının bu doğruya en uzakta olmasını amaçlamaktadır (Bonaccorso, 2017). Yanlış sınıflandırma ihtimalini azaltmak için bu doğru ile iki sınır düzlem arasındaki mesafe maksimuma çıkarılması sağlanır (Osisanwo ve ark., 2017).

Destek vektör makineleri, belirlenen iki sınıf arasındaki hiper düzlemi, doğruyu bulmaktadır. Veri noktalarından hiper düzeleme ya da doğruya olan mesafeye marjin adı verilmektedir. Marjin ne kadar geniş ise sınıflar o kadar iyi ayrıştırılmaktadır. En iyi uyum çizgisi, marjini maksimuma çıkarmayı amaçlayan bir optimizasyon problemiyle elde edilmektedir (Panesar, 2019).

SVM sınıflandırma algoritmasının en önemli avantajlarından biri, genelleme hatasını azaltarak bu modelin aşırı uyumlamaya karşı dirençli hale getirmesidir (Zhou, 2021).

Yalnızca iki boyutta değil daha yüksek veya sonsuz boyutlu uzaylarda da hiper düzlemler oluşturulabilmektedir (Garreta ve ark., 2017).

X ve Y , R^d 'nin alt kümeleri ve d özellik sayısı olmak üzere, X ve Y 'nin elemanlarını farklı taraflara ayırabilen bir hiper düzlem mevcut ise, X ve Y birbirinden doğrusal olarak ayrılabilir.

Doğrusal olarak ayrılabilen veriler için Hard-Marjin ve Soft-Marjin olmak üzere iki tür destek vektör makinesi bulunmaktadır (Suykens ve ark., 2014).

Bazen örnekler marjin bölgesine girebilir. Buna Soft Marjin denir. Hard Marjin, verimiz doğrusal olarak ayrılabiliriyorsa çalışır ve aykırı değerlere karşı çok duyarlıdır. Bu yüzden bazı durumlarda Soft Marjini tercih etmemiz gerekebilir (Khemchandani ve Chandra, 2016).

Hard-marjin destek vektör makineleri

Eğitim verileri $\{x_i, y_i\}, i = 1, 2, 3, \dots, L, y_i \in \{-1, 1\}, x_i \in R^d$ olacak şekilde hedef değişken 2 seçenekli girildiği varsayıldığında, pozitif ve negatif eğitim verileri birbirinden doğrusal olarak ayırabilen hiper düzlemin denklemi denklem 3.13'teki gibidir (Kuzu ve Yakut, 2020).

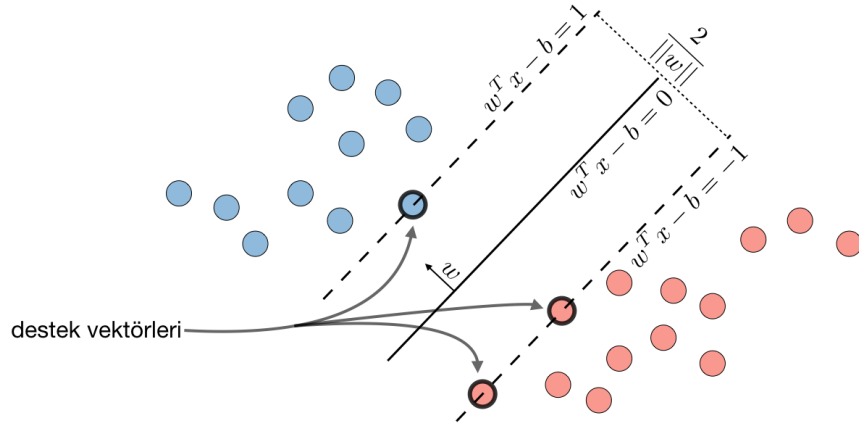
$$w \cdot x + b = 0 \quad (3.13)$$

SVM'nin amacı; farklı sınıflara ait verileri bölen hiper düzlemi, tüm sınıfların birbirine en yakın noktalarına en uzak biçimde tutmaktır. Bu sebeple w ve b seçimi çok önemlidir. Denklem 3.14 ve denklem 3.15'teki gibi özetlenebilir (Kuzu ve Yakut, 2020).

$$x_i \cdot w + b = \begin{cases} \geq 1, & y_i = 1 \text{ için} \\ \leq -1, & y_i = -1 \text{ için} \end{cases} \quad (3.14)$$

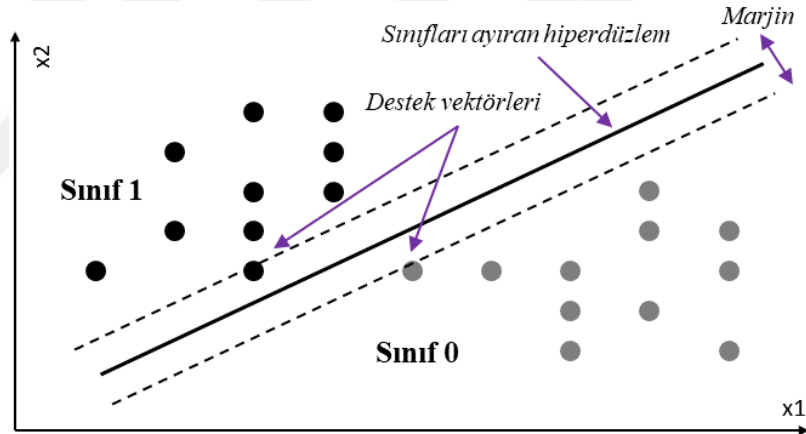
$$\forall i \text{ için } y_i(x_i \cdot w + b) - 1 \geq 0 \quad (3.15)$$

Doğrusal destek vektör makineleri düzlemsel olarak Şekil 3.13'teki gibi gösterilebilir. w hiper düzlemin normali; $|b|$, hiper düzlemden orijine dik uzaklık; $\|w\|$, w 'nin Öklid normudur (Amidi, 2023).



Şekil 3.13. Doğrusal destek vektör makinelere.

Şekil 3.14'te siyahlar ve griler olmak üzere iki sınıf bulunmaktadır. Siyahlar sınıf-1'i, griler ise sınıf-0'ı oluşturmaktadır. Yeni bir değer için çıkan sonuç 0'dan az ise, gri noktalara daha yakın olacaktır. Tam tersi, çıkan sonuç 0'a eşit veya fazla ise bu durumda siyah noktalara daha yakın olacaktır.



Şekil 3.14. Doğrusal destek vektör makinesi örneği.

Ayırıcı düzlem ise destek düzlemlerin, destek vektörlerin tam ortasından geçmektedir. Amaç, ayırıcı düzlemin destek vektörlerinden uzakta olması olduğu için marjinin maksimum olması gerekir. Marjin, $\frac{1}{\|w\|}$ olarak gösterilir. Marjinin maksimum olabilmesi için matematiksel olarak $\|w\|$ değerinin minimum olması gerekir. $\forall i$ için $a_i > 0$ olmak üzere Lagrange Çarpanları yöntemi sonucunda denklem 3.16 denklem elde edilir (Kuzu ve Yakut, 2020).

$$b = \frac{1}{N_S} - \sum_{s \in S} (a_m y_m x_m \cdot x_s) \quad (3.16)$$

Soft-marjın destek vektör makineleri

Soft-Marjın SVM, eğitim verilerinin hatasız olarak sınıflandırılmadığında kullanılmaktadır (Kuzu ve Yakut, 2020).

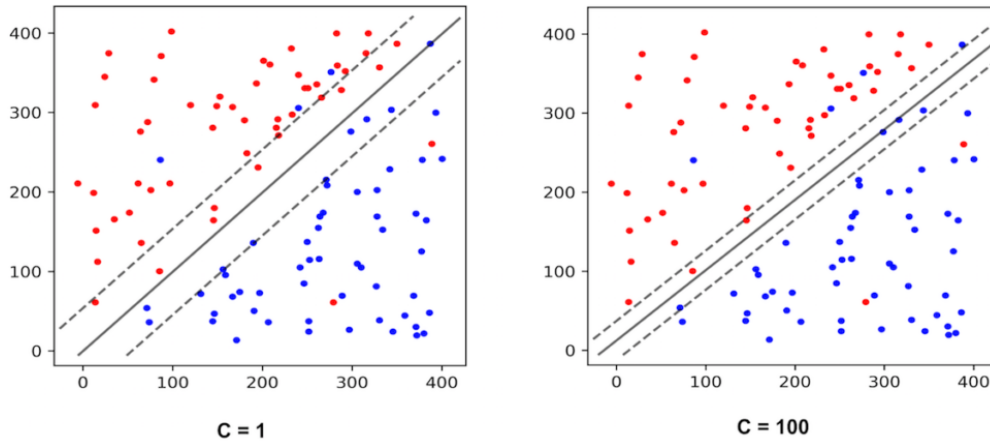
Soft-Marjın destek vektör makinelerinde $\forall i$ için denklem 3.15'i ifade edebilmek için negatif olmayan bir aylak değişken tanımlanmaktadır. Aylak değişken tanımlandığında denklem 3.17 elde edilmektedir.

$$\forall i \text{ için } y_i(x_i \cdot w + b) - 1 + \xi_i \geq 0 \quad (3.17)$$

ξ_i 'nin seçimi birçok durumu bir arada düşünmeyi zorunlu kılar ve seçimler marjini etkiler. ξ_i ile marjın arasındaki dengeyi sağlamak gerekmektedir ve bunun için C parametresi kullanılmaktadır. Bu parametre büyük seçilirse model performansı yüksek olur.

İkisi arasındaki dengeyi SVM içerisindeki C hiper parametresi ile kontrol edilebilmektedir. C ne kadar büyükse marjın o kadar dar olmaktadır. Ayrıca model aşırı uyum sağlamış (overfit) olursa C 'nin azaltılması gerekmektedir.

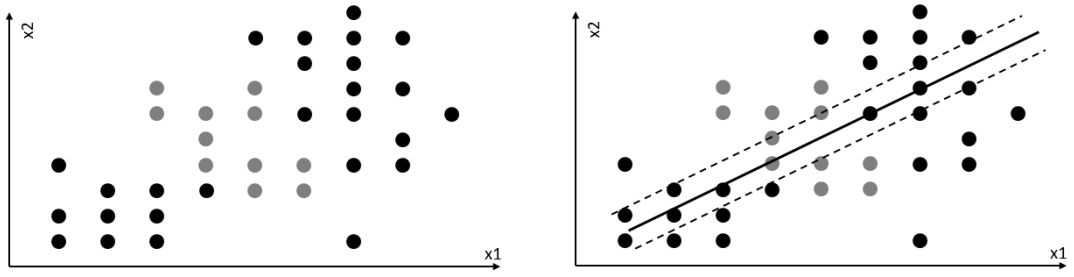
C parametresinin önemi Şekil 3.15'te görülmektedir (Akca, 2023).



Şekil 3.15. C parametresinin etkisi.

3.2.2.6. Doğrusal olmayan destek vektör makineleri

Destek vektör makinesinin verileri doğrusal olarak sınıflandırmadığı durumlarda bu yöntem kullanılır. Veri boyutunda fazladan bir boyut daha eklenir ardından SVM veri uzayında bir düzlem oluşturarak, ayrılmış olur. Doğru ile birbirinden ayrılamayan verilerin düzlemdeki örneği Şekil 3.16'deki gibidir (Metlek ve Kayaalp, 2020).



Şekil 3.16. Birbirinden doğru ile ayrılamayan veriler.

Doğrusal olmayan destek vektör algoritmalarında amaç; sınıfları ayıran en iyi çizgiyi bulmaktır. Çekirdek yöntemi, doğrusal olmayan SVM’de performansı artırmaktadır. Doğrusal olmayan destek vektör algoritmalarında çeşitli çekirdek hileleri kullanılarak yüksek başarı oranları elde edilebilmektedir (Metlek ve Kayaalp, 2020). Doğru ile ayrılamayan sınıfların boyutu artırılarak bir düzlem ile sınıflarının ayrılması sağlanabilmektedir (Murty Raghava, 2016). Bunun için kullanılan çekirdek yöntemlerinden bazıları:

Doğrusal çekirdek fonksiyonu: En basit çekirdek fonksiyonudur. Formülü denklem 3.18’deki gibidir (Kung, 2014).

$$K(x_i, x_j) = x_i^T x_j \quad (3.18)$$

Polinom (Poly) çekirdek fonksiyonu: Durağan olmayan bir çekirdek fonksiyonudur. Formülü denklem 3.19’deki gibidir.

$$K(x_i, x_j) = (1 + x_i^T x_j)^d \quad (3.19)$$

Gauss çekirdek fonksiyonu (RBF): Radyal tabanlı bir çekirdek fonksiyonudur. Aşırı tahmin olduğunda üstel ifade yaklaşık olarak doğrusallaşır ve doğrusal olmayan yapısını kaybeder. Formülü denklem 3.20’teki gibidir (Somogyi, 2021).

$$K(x_i, x_j) = (e)^{-\gamma \|x_i - x_j\|^2}, \gamma > 0 \quad (3.20)$$

Alternatif formül ise denklem 3.21’teki gibi ifade edilmektedir (Martínez ve Christodoulou, 2022).

$$K(x_i, x_j) = (e)^{\left(\frac{\|x_i - x_j\|^2}{2\sigma^2} \right)} \quad (3.21)$$

Laplace çekirdek fonksiyonu: Üstel çekirdek fonksiyonudur. Formülü denklem 3.22’teki gibidir (Metlek ve Kayaalp, 2020).

$$K(x_i, x_j) = (e)^{\left(-\frac{\|x_i - x_j\|}{\sigma}\right)} \quad (3.22)$$

Anova çekirdek fonksiyonu: Radyal tabanlı bir çekirdek fonksiyonudur. Formülü denklem 3.23’deki gibidir (Metlek ve Kayaalp, 2020).

$$K(x_i, x_j) = \sum_{i=1}^n \exp\left(-\sigma(x_i^i - x_j^i)^2\right)^d \quad (3.23)$$

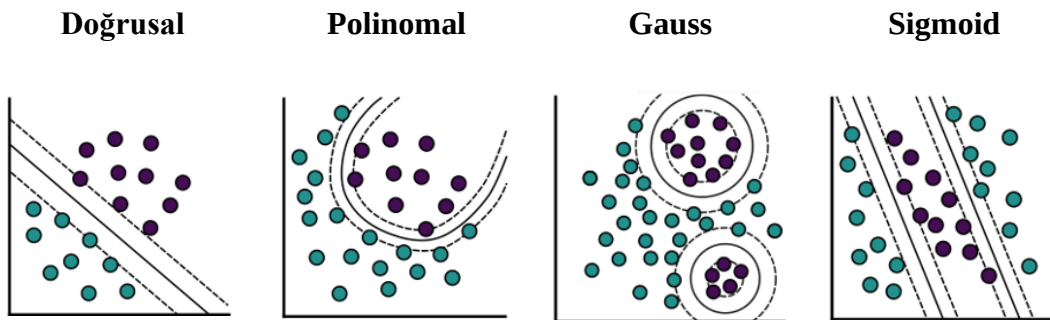
Sigmoid çekirdek fonksiyonu: Sigmoid çekirdeğinde α eğim ve c kesme sabiti olmak üzere iki ayarlanabilir parametresi vardır. α için ortak bir değer ($1 / N$) kullanılabilir ve N değeri veri boyutunu ifade etmektedir. Formülü denklem 3.24’teki gibidir (Kutateladze, 2022).

$$K(x_i, x_j) = \tanh(\alpha x_i^T x_j + c) \quad (3.24)$$

Bessel çekirdek fonksiyonu: Bessel çekirdeği düzgün fraksiyonel uzaylar teorisinde iyi bilinen bir çekirdektir. Formülü denklem 3.25’teki gibidir (Metlek ve Kayaalp, 2020).

$$K(x_i, x_j) = \frac{J_{v+1}(\sigma \|x_i - x_j\|)}{\|x_i - x_j\|^{-n(v+1)}} \quad (3.25)$$

Çekirdek fonksiyonlarının bazılarının ait düzlem üzerindeki örnekler Şekil 3.17’deki gibidir (Akca, 2023).



Şekil 3.17. Çekirdek fonksiyonu sınıflandırma örnekleri.

3.2.2.7. Lojistik regresyon

Lojistik regresyon, istatistikçi David Cox tarafından 1958'de geliştirilmiş olan bir sınıflandırma algoritmasıdır (Panesar, 2019).

Lojistik regresyon, bir gözlemin belirli bir gruba ait olma olasılığını, bir dizi bağımsız değişken (orijinal değişkenlerimizin uygun etkileşimleri ve dönüşümleri dahil) kullanarak tahmin eden bir sınıflandırma tekniğidir. Bunun yanı sıra, bağımlı sınıf değişkenin ikiden fazla kategorik değişken olduğu durumlarda bağımsız değişkenlerle olan neden-sonuç ilişkisini belirlemede kullanılmaktadır (Başer ve ark., 2021).

- Müşteri davranışını tahmin etme
- Tıbbi bir durumun olasılığını tahmin etmek
- Hastanın hastaneye kabul edilip edilmeyeceği
- Hastaya verilen ilacın olumlu-olumsuz etkisini tahmin etme
- Ciltteki bir lekenin iyi huylu-kötü huylu tahmini

Tahmin edilecek çıktının özelliklerine göre kullanılacak lojistik regresyon modeli değişmektedir (Pate ve ark., 2023).

İkili lojistik regresyon: Tahmin edilecek çıktının 2 tane olduğu durumlarda kullanılır. Örnek olarak: Evet-hayır, hasta-hasta değil, dolandırıcı-dolandırıcı değil, sağ-ölü, etkili-etkisiz.

Çok terimli lojistik regresyon: Çok terimli lojistik regresyon, ikiden fazla kategoriye sahip kategorik bir sonucun olasılığını tahmin etmek için kullanılır (örneğin, bir müşterinin üç farklı üründen birini seçme olasılığı). Birden çok bağımsız değişken olduğu durumlarda kullanılır.

Sıralı lojistik regresyon: Bir müşterinin bir ürünü 1 ile 5 arasında derecelendirme olasılığı gibi sıralı bir sonucun olasılığını tahmin etmek için kullanılır.

Çoklu lojistik regresyon: Bir olayın ortaya çıkma ihtimalini tahmin etme amacıyla kullanılan birden fazla bağımsız değişken olduğu durumlarda çoklu lojistik regresyon kullanılır. İkiden fazla kategoriye sahip kategorik bir sonucu tahmin etmek için kullanılır.

Örnek olarak, bir sınıfa ait olma olasılığına dayanan bir sınıflandırma yöntemidir. Olasılıkların R'de sürekli olması ve (0, 1) arasında sınırlı olması gerektiğinden, z terimini filtrelemek için bir eşik fonksiyonu eklemek gerekir. Bunun için

sigmoid(lojistik) fonksiyonu kullanılmaktadır. Lojistik regresyon adı da sigmoid yani lojistik fonksiyonunun kullanılmasıdır (Pate ve ark., 2023).

Klasik regresyon yöntemlerinde bağımlı değişkenin 0 ve 1'lerden oluştuğu durumlarda, elde edilen değerlerin 0 ve 1 arasında olması her iş probleminde mümkün olmamaktadır. Sigmoid fonksiyonu, z doğrusal formunun değerlerini 0 ve 1 arası değerlere eşlemektedir. Formüller denklem 3.26 ve denklem 3.27'de verilmiştir (Qutab ve ark., 2022).

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (3.26)$$

$$\sigma(\bar{x}; \bar{w}) = \frac{1}{1 + e^{-\bar{x} \cdot \bar{w}}} \quad (3.27)$$

Bir örneğin bir sınıfa ait olma olasılığı denklem 3.28'deki gibi tanımlanabilir.

$$P(y|\bar{x}) = \sigma(\bar{x}; \bar{w}) \quad (3.28)$$

Bu noktada en uygun parametreleri bulabilmek için log olabilirliğini en büyükmek gerekir. Formül denklem 3.29'deki gibidir.

$$L(\bar{w}; y) = \log P(y|\bar{w}) = \sum_i \log P(y_i|\bar{x}_i, \bar{w}) \quad (3.29)$$

Optimizasyon problemi kayıp fonksiyonun minimizasyonu olarak ifade edilebilir. Formül denklem 3.30'teki gibidir.

$$\begin{aligned} J(\bar{w}) &= - \sum_i \log P(y_i|\bar{x}_i, \bar{w}) \\ &= - \sum_i (y_i \log \sigma(z_i) + (1 - y_i) \log (1 - \sigma(z_i))) \end{aligned} \quad (3.30)$$

Eğer $y=0$ ise, ilk terim boş olur ve ikincisi $\log(1 - x)$ olur. Bu da 0 sınıfının log olasılığıdır. Öte yandan $y=1$ ise ikinci terim 0 olur ve birinci terim x 'in log olasılığını temsil eder. Her iki durum da tek bir ifadeye gömülür. Bir hedef dağılım ile yaklaşık

bir dağılım arasındaki çapraz entropinin en aza indirilmesi anlamına gelir. Formül denklem 3.31'teki gibidir.

$$H(X) = - \sum_{x \in X} p(x) \log_2 q(x) \quad (3.31)$$

$\log_2$, orijinal dağıtımı tahmin edilen dağıtımla kodlamak için talep edilen ekstra bit sayısını ifade eder. $J(\bar{w})=0$ Olduğunda iki dağılım birbirine eşittir. Bu sebeple çapraz entropiyi en aza indirmek, hedef dağılımlar kategorik olduğu durumda tahmin hatasını optimize etmenin bir yoludur.

Çoklu lojistik regresyonda, bağımsız değişkenler kesikli ya da nominal ölçekli değişkenleri çalışmaya dahil etmek için dizayn değişkenleri kullanılması gerekmektedir. Modeldeki tüm bağımsız değişkenlerin her birinin en az aralık ölçekli olduğu varsayıldığı durumda;

$X' = (x_1, x_2, x_3, \dots, x_p)$ vektörü ile gösterilsin. Sonuç değişkeninin mevcut olduğu ($Y = 1$) zamanki koşullu olasılık, $P(Y = 1/x) = \pi(x)$ 'e eşit olacaktır. Çoklu lojistik regresyon modelinin logiti denklem 3.32 ile gösterilir:

$$g(x) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (3.32)$$

Bu durumda denklem 3.33 bulunmuş olur.

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad (3.33)$$

Bağımsız değişkenler kesikli ya da nominal ise, bu değişkenler yerine kukla değişkenler bu değişkenleri temsil etmektedir. Genel olarak nominal değişken k kategoriye sahipse o zaman $k - 1$ kukla değişkenine ihtiyaç olur.

x_j : j . Bağımsız değişken,

k_j : Kategori sayısı,

$k_j - 1$: Kukla değişken sayısı,

D_{ju} : Kukla değişkeni ve katsayıları $u = 1, 2, 3, \dots, k_j - 1$

j . Değişkeni kesikli olan p değişkenli modelin logit fonksiyonu denklem 3.34'teki gibidir.

$$g(x) = \beta_0 + \beta_1 X_1 + \dots + \sum_{u=1}^{k_j-1} \beta_{ju} D_{ju} + \beta_p X_p \quad (3.34)$$

3.2.3. Makine öğrenmesi sınıflandırma algoritmaları performans ölçümü

Karışıklık matrisi; iki veya daha fazla sınıfa sahip modelin ya da algoritmanın performansının hesaplanması için kullanılmaktadır (Dilwani, 2019). Karışıklık matrisi, model tarafından yapılan doğru ve yanlış sınıflandırmaları ayrıştırır ve bunları uygun etikete bağlar (Panesar, 2019).

Eğitim veri seti ile eğitilen algoritmalar test veri seti ile test edilmektedir. Test veri seti ile test edildikten sonra bir karışıklık matrisi elde edilir. Karışıklık matrisi tablosu Tablo 3.4'te verilmiştir (Larner, 2021).

Tablo 3.4. İkili karışıklık matrisi.

Karışıklık matrisi		Tahmin edilen	
		Pozitif	Negatif
Gerçekte olan	Pozitif	Doğru pozitif (TP)	Yanlış negatif (FN)
	Negatif	Yanlış pozitif (FP)	Doğru negatif (TN)

Doğru pozitif (TP): Tahmin edilecek sınıfın gerçekte pozitif olduğu, tahmin edilen sınıfın pozitif olduğu sonuçların sayısı.

Yanlış pozitif (FP): Tahmin edilecek sınıfın gerçekte negatif olduğu, tahmin edilen sınıfın pozitif olduğu sonuçların sayısı.

Doğru negatif (TN): Tahmin edilecek sınıfın gerçekte negatif olduğu, tahmin edilen sınıfın negatif olduğu sonuçların sayısı.

Yanlış negatif (FN): Tahmin edilecek sınıfın gerçekte pozitif olduğu, tahmin edilen sınıfın negatif olduğu sonuçların sayısı.

Algoritmaların performansları bazı ölçütlere göre değerlendirilmektedir. Bu ölçütler ve hesaplama yöntemleri aşağıda açıklanmıştır.

Doğruluk: Doğru tahmin edilen gözlem sayısının toplam gözlem sayısına oranlanması ile elde edilmektedir. Formül denklem 3.35'teki gibidir.

$$Doğruluk = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (3.35)$$

Hata oranı: Yanlış tahmin edilen gözlem sayısının toplam gözlem sayısına oranlanmasıyla elde edilmektedir. Formül denklem 3.36'deki gibidir.

$$Hata\ oranı = \frac{(FP + FN)}{(TP + TN + FP + FN)} \quad (3.36)$$

Kesinlik ölçütü: Doğru tahmin edilen pozitif gözlem sayısının toplam pozitif tahmin edilen gözlem sayısına oranlanması ile elde edilmektedir (Panesar, 2019). Kesinlik ölçütünün alabileceği en iyi değer 1'dir. Formül denklem 3.37'taki gibidir.

$$Kesinlik = \frac{TP}{(TP + FP)} \quad (3.37)$$

Duyarlılık: Kaynaklarda duyarlılık değeri, anma değeri, hatırlama değeri, geri çağırma değeri olarak geçmektedir. Duyarlılık değeri, doğru tahmin edilen pozitif gözlem sayısının toplam doğru pozitif tahmin edilen gözlem sayısına oranlanması ile elde edilmektedir (Panesar, 2019). Duyarlılık ölçütünün alabileceği en iyi değer 1'dir. Formül denklem 3.38'deki gibidir.

$$Duyarlılık = \frac{TP}{(TP + FN)} \quad (3.38)$$

Özgünlük: Modelin yanlış sınıflandırmaları döndürmede ne kadar iyi performans gösterdiğini ifade etmektedir. Doğru tahmin edilen negatif gözlem sayılarının toplam negatif gözlem sayısına oranlanmasıyla elde edilmektedir (Hossin ve Sulaiman, 2015). Formül denklem 3.39'deki gibidir.

$$Özgünlük = \frac{TN}{(TN + FP)} \quad (3.39)$$

F ölçümü: Sınıflandırıcının ne kadar hassas olduğunu, kaç tane örneği doğru şekilde sınıflandırdığını ve ne kadar sağlam olduğunu anlatmaktadır (Dilwani, 2019). Aritmetik ortalamanın ötesine geçerek kesinlik ve geri çağırmanın harmonik

ortalamasını hesaplar. P, kesinliği ve r geri çağırmaı belirtir (Hossin ve Sulaiman, 2015). Formül denklem 3.40 ve denklem 3.41'teki gibidir.

$$F = \frac{2 \times \text{Kesinlik} \times \text{Anma}}{\text{Kesinlik} + \text{Anma}} \quad (3.40)$$

$$F = \frac{1}{\frac{1}{2} \left(\frac{1}{p} + \frac{1}{r} \right)} = \frac{2pr}{p + r} \quad (3.41)$$

F_β ölçümü: Kesinlik ölçütüne, duyarlılık ölçütüne göre β kat daha fazla ağırlık verilmek istendiğinde F_β ölçütü aşağıdaki gibi hesaplanmaktadır (Larner, 2021). Formül denklem 3.42'teki gibidir.

$$F_\beta = \frac{(1 + \beta^2) \times \text{Kesinlik} \times \text{Anma}}{(\beta^2 \times \text{Kesinlik}) + \text{Anma}} \quad (3.42)$$

Doğruluk oranı, duyarlılık değeri, kesinlik değeri ve F ölçütü değeri daha yüksek olan sınıflandırıcı modelleri diğer sınıflandırıcı modellerine göre daha başarılıdır. Çalışmada algoritma performansları bu değerlere göre değerlendirilmiştir.

Logaritmik kayıp: Sınıf etiketi yerine sürekli olasılığın tahmin edildiği problemler için kullanılır. Log kaybı, doğruluğun güvenilirliğine ilişkin olasılıksal bir ölçüm sağlar ve gerçek etiketlerin ve tahminlerin dağılımı arasındaki entropiyi dikkate alır. p_i : i . veri noktasının bir sınıfa ve y_i 'nin gerçek etikete (0 veya 1) ait olma olasılığıdır. İkili sınıflandırma problemi için logaritmik kayıp hesabı denklem 3.43'teki gibidir (Panesar, 2019).

$$\text{Logaritmik kayıp} = -\frac{1}{N} \sum_{i=1}^N y_i \log p_i + (1 - y_i) \log(1 - p_i) \quad (3.43)$$

ROC Eğrisi: Sınıflandırma modellerinin performansının değerlendirilmesinde kullanılan bir diğer ölçüt ise ROC eğrisinin altında kalan alandır. ROC eğrisi, gerçek pozitiflerin oranını yanlış pozitiflerin oranına göre çizmektedir. ROC, sınıflandırıcının duyarlılığının ve özgüllüğünün görselleştirilmesini sağlar. Yanlış pozitiflere izin vererek kaç tane doğru pozitif sınıflandırmanın elde edilebileceğini vurgular. Düşey y ekseninde doğru pozitif oranı ve yatay x ekseninde yanlış pozitif oranı yer almaktadır.

ROC eğrisi, eğrinin altında kalan alana göre yorumlanmaktadır. Eğrinin altında kalan alan büyükse yani yüksek bir ROC varsa modelin başarısı yüksek olarak yorumlanmakta ve istenen bir durumdur. Eğrinin altında daha küçük bir alan veya eğrinin altında daha az alan istenmemektedir. Grafikte çapraz bir çizgi de bulunmaktadır. Çapraz çizgiye yakınlık daha az doğru model olduğunu göstermektedir. Çalışma da modellerin ROC eğrisi altında kalan alan AUC hesaplanmıştır. Karışıklık matrisi ölçütlerinin yanı sıra AUCROC değerlerine göre de değerlendirilmiştir. Formüller denklem 3.44 ve denklem 3.45'teki gibidir.

$$\text{Doğru pozitif oranı} = \frac{TP}{(TP + FN)} \quad (3.44)$$

$$\text{Yanlış pozitif oranı} = \frac{FP}{(TN + FP)} \quad (3.45)$$

MAE: Regresyon modellerinde ortalama mutlak hatayı temsil etmektedir. Formül denklem 3.46'daki gibidir (Artificial Intelligence Technology, 2021).

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (3.46)$$

MSE: Regresyon modellerinde ortalama kare hatayı temsil etmektedir. Formül denklem 3.47'deki gibidir (Kushwah ve ark., 2022).

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (3.47)$$

RMSE: MSE'nin kareköküdür. Formül denklem 3.48'deki gibidir (Glackin ve Adivar, 2023).

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (3.48)$$

3.2.4. Çapraz doğrulama

Çapraz doğrulama yapılmış model oluşturmak, yüksek performansla çalışan makine öğrenimi modellerinin geliştirilmesini sağlamaktadır. Birden fazla çapraz doğrulama

teknikleri, test bölünmesindeki farklılıklardan fırsat vermeden bir algoritmanın performansını tahmin edilmesini sağlamakta ve düzgün dağılmamış bir veri bölünmesinin sebep olabileceği problemleri de ortadan kaldırmaktadır. Bu şekilde, sınıflandırma işlemlerinin daha tutarlı sonuç vermesi sağlanmaktadır. Çapraz doğrulama yöntemlerini kullanmak, test bölünmesinden fırsat vermeden oluşturulan modelin performansını tahmin edilmesini sağlamakta ve aynı zamanda dengesiz veri bölünmesinden kaynaklı sorunları ortadan kaldırmaktadır (Pal ve ark., 2020).

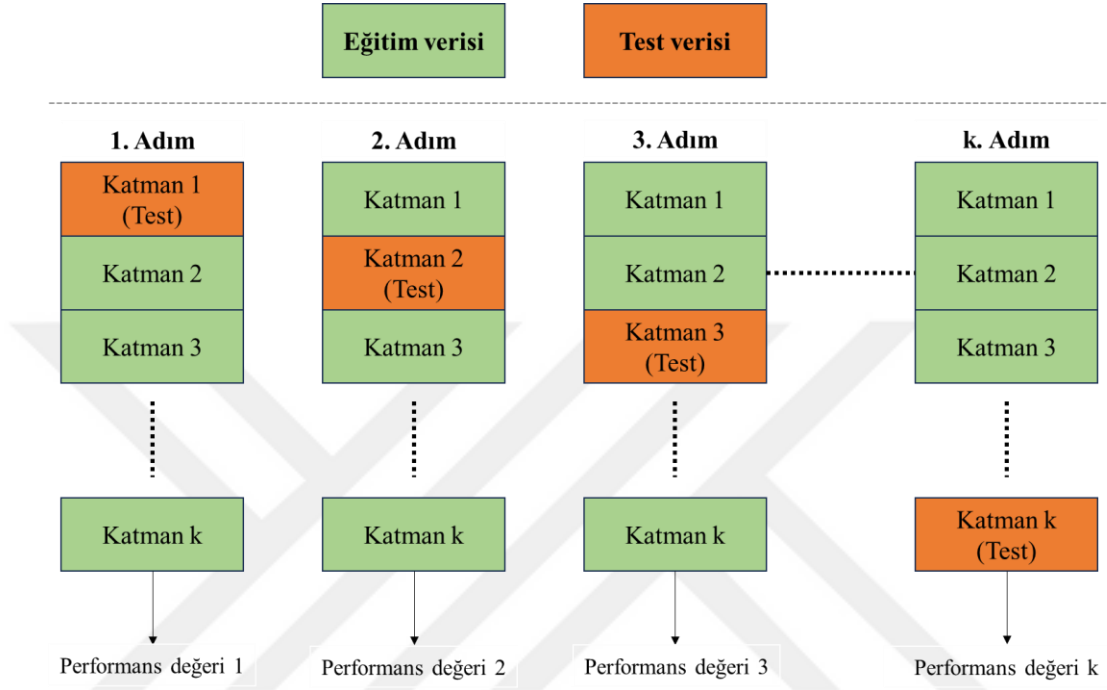
Sınıflandırma sürecinde; veri seti, test ve eğitim olarak ikiye ayrılır ve eğitim setiyle makine öğrenmesi sınıflandırma algoritmaları eğitilir. Elde edilen model test verileri ile test edilir ve karışıklık matrisi ile modelin performansı anlaşılmaya çalışılır. Ancak bu süreç, veri setinin belirli bir kısmının eğitim seti olduğu, belirli bir kısmının test veri seti olduğu durumu yansıtır. Eğer veri setinin başka bir bölümünü eğitim seti olarak kullanılsa ya da eğitim seti-veri seti ayrımındaki oran farklı olarak kullanılsa farklı performans değerleri elde edilebilir. Eğitim seti olarak kullanılan veri setinde, verinin dağılımından kaynaklı durumlar yaşanabilir. Örnek olarak, eğitim seti olarak ayrılan kısımda tahmin edilecek aynı alana yoğunlaşmış veriler yer alabilir. Bu ve benzeri durumlar model performansını etkileyebilmektedir. Bu nedenle, çapraz doğrulama yapılmadan modelin performansına yorum yapılması doğru olmayacaktır.

Birçok çapraz doğrulama tekniği bulunmaktadır. Aşağıda en sık kullanılan iki tanesi özetlenmiştir. Bunlar, monte carlo çapraz doğrulama ve k-katmanlı çapraz doğrulama.

Monte Carlo çapraz doğrulama: Bu çapraz doğrulama veri setini rastgele bölmekte ve her doğrulamada veri setinin farklı bölümlerini kullanarak modeli eğitir ve doğrulamasını yapar (Öztürk, 2023). Bu yöntem, tekrarlanan rastgele alt örnekleme çapraz doğrulaması olarak da bilinmektedir (Patro, 2023) Bu yöntem için Scikit-Learn kütüphanesinde ShuffleSplit() fonksiyonu kullanılmaktadır (Öztürk, 2023).

K-katmanlı çapraz doğrulama: K-katmanlı çapraz doğrulama literatürde “k-fold cross validation” olarak bilinmektedir. K-katmanlı çapraz doğrulama, makine öğrenmesi modellerini değerlendirmek ve verdikleri sonuçların performansını ölçmek için kullanılan bir tekniktir (Xiong ve ark., 2020). K-katmanlı çapraz doğrulama yöntemi için önce bir k parametresi belirlenmektedir. Bu k parametresi veri setinin kaç parçaya bölüneceğini belirtmektedir (Tamilarasi ve ark., 2020) Veri k adet parçaya bölünmekte ve k adet sınıflandırma işlemi yapılmaktadır. Her sınıflandırma adımında bölünen

parçalarından başka bir tanesi test verisi olarak kullanılmakta, geriye kalan $k-1$ tane parça ise modeli eğitmek için kullanılmaktadır. Bu işlem k kez tekrarlanır (Pal ve ark., 2020). Bu şekilde oluşturulan her parçanın hem eğitim verisi hem test verisi olarak kullanılması sağlanmaktadır. K-katmanlı çapraz doğrulama süreci Şekil 3.18’de özetlenmiştir (Nti ve ark., 2021).



Şekil 3.18. K-katmanlı çapraz doğrulama süreci.

Bu çalışmada; veri setindeki verinin dağılımından kaynaklanabilecek hataları minimize edebilmek adına, çapraz doğrulama sırasında veri setinin rastgele bölünmesi sağlanarak monte carlo çapraz doğrulama işlemi yapılmıştır.

3.2.5. Python programı

Algoritma sözcüğü, Ebu Abdullah Muhammed İbn Musa El Harezmi adındaki Müslüman Türk aliminden almıştır. Arapça adı Al-Khwarizmi’dir. Bu sözcük; Latinceye “Algoritmi”, Fransızcaya “Algorithme”, İngilizceye “Algortihm” şeklinde geçmiştir. Algoritma; bir problemin çözümü için izlenmesi gereken yol anlamına gelmektedir. Sorunu sonlu bir zamanda, açıkça, yürütülebilir, sıralı, basit ve gerektiği zaman tekrar eden sıralı adımlar dizisidir (Çobanoğlu, 2020). Algoritmada yer alan her adıma ise komut adı verilmektedir.

Programlama dili, kişinin bilgisayara ne yapması gerektiğini anlatan komutlar bütününe programlama dili denilmektedir (Çobanoğlu, 2020).

Program; bir algoritma, bilgisayar ortamında programlama dillerine uygulandığında program adını almaktadır (Çobanoğlu, 2020).

Python nesne tabanlı bir programlama dilidir. 1991 yılında Hollandalı bilgisayar mühendisi Guido van Rossum tarafından geliştirilmeye başlanmıştır. Python ismi, Monty Python adında İngiliz komedi grubunun “Monty Python’s Flying Circus” isimli gösterisinden almıştır (Çobanoğlu, 2020).

Python; modüler, 100% nesne yönelimli, açık kaynak kullanımlı yani ücretsiz şekilde kullanılabilen, yorumlayıcı kullanan hızlı ve dinamik bir programlama dilidir. Aynı zamanda Java programlama dili gibi hem derleyici hem de yorumlayıcı özelliği bulunmaktadır.

Python programlama dili ile, konsol uygulamaları, görsel masaüstü uygulamaları yani GUI uygulamaları, web tabanlı uygulamalar gerçekleştirilebilir (Çobanoğlu, 2020).

Python programlama dilini kullanabilmek için Python editörlerinden birine ihtiyaç vardır. Python editörü; kod yazmak, derlemek, çalıştırmak ve hataları ayıklamak için kullanılan araçlara verilen isimdir. Kısaca IDE olarak tabir edilmektedir. IDE’nin açılımı Integrated Development Environment yani Türkçe tabiri ise Entegre Yazılım Geliştirme Ortamıdır. Python için birden fazla IDE bulunmaktadır. IDLE, PyCharm, Spyder gibi entegre yazılım geliştirme ortamları örnek olarak verilebilir.

Spyder, birçok insanın kullandığı açık kaynak kullanımlı bir entegre yazılım geliştirme ortamıdır. Veri analizi, hata ayıklama ve veri görselleştirme için geniş bir geliştirme ortamı sağlamaktadır (AWS Amazon, 2023).

Kitaplık; yazılacak Python programlarına dahil edilebilecek, sıfırdan kod yazmamak için en sık kullanılan kodlardan oluşan koleksiyonlara Python kitaplıkları denir. Python kitaplıkları sayesinde en sık kullanılan kodlar sıfırdan yazılmaz direkt olarak kullanılabilmekte ve yeni programın içerisine dahil edilebilmektedir. Örnek olarak, veri bilimi ve makine öğrenmesi için Python kitaplıkları bulunmaktadır. En sık kullanılan Python kitaplıkları; Matplotlib, Pandas (Veri bilimi, makine öğrenmesi), NumPy, Keras vb. gibi örneklendirilebilir (AWS Amazon, 2023).

Python programlama dilinin en önemli özelliklerinden birisi derlemeye gerek kalmadan kullanılabilmesidir. Diğer programlama dillerinde olan derleme işlemi Python’da olmadığı için Python programlama dilinde hızlı bir şekilde program geliştirilebilmektedir (Özgül, 2013).

Python programlama dili ile yapılabilecekler aşağıda örneklendirilmiştir.

- Konsol uygulamaları, GUI uygulamaları programlanabilir.
- Makine öğrenmesi konularında algoritmalar ile modeller kurularak tahmin çalışmaları yapılabilir.
- Robotik kodlama çalışmaları yapılabilir.
- Veri bilimi alanında çalışarak veri analizi gibi çalışmalar yapılabilir.
- Mobil uygulamalar tasarlanabilir.

Son zamanlarda Python programlama dili makine öğrenimi alanında giderek daha önemli bir yer almaya başlamıştır. Python programlama dili, makine öğrenimi çalışmalarını sağlamak için kitaplık hizmeti sağlamaktadır. Doğal dil işleme, metin ve duygu analizi gibi problemleri için birçok çözüm yolu sağlamaktadır. Python'daki makine öğrenimi için kullanılan en meşhur kütüphanelerden biri de scikit-learn kütüphanesidir. Scikit-Learn kullanılarak makine öğrenimi alanında çalışmalar yapılabilir. Veri setlerinde çeşitli özellikleri arama, ön işleme algoritmaları uygulanabilir. Eğitim verileri ile algoritmalar eğitilebilir, test verileri ile tahmin çalışmaları yapılabilir (Sultonov, 2023).

3.3. Veri Seti

3.3.1. Veri toplama süreci

Çalışmada kullanılan veriler Kartal Doktor Lütfü Kırdar Şehir Hastanesi Acil Tıp Kliniğinden elde edilmiştir. Çalışma; prospektif bir çalışma olup, veri toplamaya başlanmadan önce Sakarya Üniversitesi Etik kurulundan etik kurul izni ve Kartal Doktor Lütfü Kırdar Şehir Hastanesi'nden baş hekim onayı alınarak başlanmıştır. Çalışmadaki veriler, genel olarak Ocak 2023-Nisan 2023 ayları arasında acil servise başvuran 18 yaş üstü hastaların verilerini içermektedir. Toplanan veriler rutinde kayıt altına alınan standart veriler olup, çalışmaya özgü özel nitelikte veri toplanmamıştır.

Çalışma için verileri alınan katılımcılar; acil servise başvuru saatleri rastgele seçilerek, acil servise başvuran kişiler arasından rastgele seçilmiştir. Özel bir şikayete ya da özgeçmişe sahip hastalar özellikli olarak seçilmemiş, tüm hastalar rastgele seçilmiştir. Çalışmaya; gebe olan ve gebelikle ilgili şikayetleri olan kadın doğum bölümü hastaları direkt kadın doğum acile gittiği için bu hastalar çalışmaya dahil edilmemiştir. Bununla beraber vitalleri stabil olan, majör fraktür ve amputasyon olmayan travma hastaları travma için özel alana alındığı için bu hastalar çalışmaya dahil edilmemiştir.

Çalışma için; yaklaşık olarak her hafta, haftada en az tam bir gün hastane ziyaret edilerek, hastaların triyaj kağıtlarından ilgili triyaj bilgilerine ulaşılmıştır. Bunun yanı sıra hastaların triyaj kağıtlarından ulaşılan protokol numaraları ile hastane bilgi yönetimi sistemindeki verilere de erişilip; veriler kontrol edilip, doktor notları, önceki ziyaretlerinde kaydedilmiş özgeçmiş hastalıkları, acil servis hizmet sonuçları gibi bilgilere erişilmiştir. Bu bilgiler sistemde metinsel şekilde bulunmaktadır. Metinsel alanlar tek tek okunarak gerekli bilgilere ulaşılmıştır. Toplanan veriler hastane bilgi yönetim sisteminden direkt rapor alınarak erişebilecek bilgiler olmamakla birlikte çalışma boyunca veriler manuel olarak toplanmıştır. Bu sebeple veri toplama süreci daha geniş bir zaman aralığını kapsamıştır.

3.3.2. Veri seti tanıtımı

Acil servisten hastaların sözlü sağlık şikayetleri, hastaların iletmiş olduğu ya da önceki ziyaretlerinden bilgi yönetim sistemine kaydolmuş sahip olunan özgeçmiş hastalıkları, acil servise varış şekli, Glasgow koma skalası, yaş, cinsiyet, yaşamsal bulguları içermektedir.

Veri seti; acil servise başvuran, rastgele seçilmiş 18 yaşından büyük hastaların verilerini içermektedir. Toplam 3.000 hastanın verisi ile çalışılmıştır. Her satır, bir hastanın triyaj verilerini içerir. Her hastaya ait 56 adet triyaj verisi bulunmaktadır. Tahmin çalışması için toplam 56 adet girdi yani 56 adet sütun bulunmaktadır. Tahmin edilebilecek sonuç yani çıktı olarak kullanılabilecek 2 adet sütun bulunmaktadır. Bu sütunlar acil servis hizmeti sonucunu göstermektedir. Bunlardan biri triyaj alan sınıflandırma sonucu diğeri ise taburcu durumudur. Veri setindeki kategorik girdiler işlenmeden önce 0-1 olarak dönüştürülmüştür.

Çalışmada kullanılan veri setinin bilgilerini içeren tablolar ek olarak paylaşılmıştır. Çalışma için kullanılan veri setinin özellikleri ekler bölümündeki Tablo B.1, Tablo B.2, Tablo B.3, Tablo B.4, Tablo B.5 ve Tablo B.6'daki gibidir. Bu tablolarda veri setinin özellikleri, nicel, kategorik veri olma durumları, girdi, çıktı olma durumları gibi bilgiler yer almaktadır. Veri setinde toplanan bilgiler aşağıda ayrıntılı olarak açıklanmıştır.

Genel bilgiler, hastanın nüfus kaydında geçen, makine öğrenmesi modeline katkı sağlayabileceği düşünülmüş genel bilgileri içermektedir. Veri setinde toplam 2 tane genel bilgi girdisi bulunmaktadır.

- Cinsiyet: Hastanın cinsiyetini ifade etmektedir. Kadın ise 0, erkek ise 1 olarak ifade edilmiştir. Kategorik bir veridir.
- Yaş: 18 yaş üstündeki hastaların yaşını belirtir. Nicel bir veridir.

Yaşamsal bilgiler (vital bulgular); hastanın hayati durumunun, yaşamı sürdürmesi için gerekli fonksiyonlarının durumunu gösteren bulgulardır. Bu bulgular hastanın sağlığının değerlendirilmesinde yardımcı olup ve hastalıklara dair ipuçları vermektedir. Veri setinde toplam 9 tane vital bilgi girdisi bulunmaktadır.

- Sistolik kan basıncı: Kalbin kasılması sırasında ulaşılan en yüksek kan basıncı değeridir. Diğer bir deyişle büyük tansiyon değeridir. “mmHg” olarak ifade edilir. Nicel bir veridir.
- Diyastolik kan basıncı: Kalp atımları arasında, kalp aktif şekilde kan pompalamazken, damarlar içinde oluşan kan basıncı değeridir. Diğer bir deyişle küçük tansiyon değeridir. “mmHg” olarak ifade edilir. Nicel bir veridir.
- Ortalama arter basıncı/MAP: Belirli bir anda hastadaki sistolik ve diyastolik boyunca ortalama arter basıncı değeridir. Nicel bir veridir.
- Kalp atış hızı/Nabız: Dakikadaki kalp atım sayısıdır. Nicel bir veridir.
- Solunum sayısı: Dakikadaki solunum sayısıdır. Nicel bir veridir.
- Vücut sıcaklığı/Ateş: Hastanın alnından ölçülen vücut sıcaklık değeridir. °C olarak ifade edilmektedir. Nicel bir veridir.
- Oksijen satürasyonu: Kandaki oksijenin yoğunluk yüzdesidir. % ile ifade edilir. Nicel bir veridir.
- Kan şekeri: Kandaki glikoz miktarını göstermektedir. “mg/dl” olarak ifade edilmektedir. Nicel bir veridir.
- Glasgow koma skalası (GKS): Acil servise başvuran hastanın bilinç durumunu başlangıçta ve sonraki değerlendirmelerde güvenilir ve objektif olarak kaydetmeyi amaçlayan bir yöntemdir. 3 ile 15 arasında bir değer alabilmektedir. Nicel bir veridir.

Şikayet bilgileri; hastanın acil servise başvurduğu zaman sağlık personeline ilettiği şikayetlerdir. Çalışma boyunca toplanan veriler doğrultusunda şikayetler aşağıdaki gibi gruplandırılmıştır. Eğer hasta acil servise başvurduğunda bir ya da birkaç şikayetini belirtiyorsa o hasta için ilettiği şikayetler 1 olarak, iletmiyorsa o şikayetler 0 olarak kaydedilmiştir. Veri setinde toplam 32 tane şikayet girdisi bulunmaktadır.

- Arrest durumu: Kalbin kan pompalama işlevinin anlık durması ya da bitmesi. 0 ise arrest durumu yaşanmamış, 1 ise arrest durumu yaşanmıştır. Kategorik bir veridir.
- Boğaz ağrısı, öksürük, kulak ağrısı durumu: Hastada boğaz ağrısı, öksürük ve/veya kulak ağrısı şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Balgam varlığı, balgam niteliğinde değişiklik durumu: Hastada balgam şikayeti olma durumu ya da balgamda durum değişikliğinin yaşanmasıdır. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Yan ağrısı durumu: Hastada yan ağrısı şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Göğüs ağrısı durumu: Hastada göğüs ağrısı şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Nefes darlığı durumu: Hastada nefes darlığı şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Çarpıntı durumu: Hastada çarpıntı şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Yüksek ateş durumu: Hastada vücut sıcaklığının yüksek olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Karın ağrısı durumu: Hastada karın ağrısı şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Kan kusma, siyah dışkılama durumu: Hastada kan kusma, siyah dışkılama şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Baş ağrısı durumu: Hastada baş ağrısı şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Yaygın vücut ağrısı, halsizlik, beslenme bozukluğu durumu: Hastada yaygın vücut ağrısı ve/veya halsizlik şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Yüksek kan şekeri durumu: Hastada yüksek kan şekeri şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Bulantı, kusma durumu: Hastada bulantı ve/veya kusma şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.

- Baş dönmesi durumu: Hastada baş dönmesi şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- İshal durumu: Hastada ishal şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Bayılma durumu: Hastada bayılma şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Bayılayazma durumu: Hastada bayılayazma şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Mekanik ağrı durumu: Hastada mekanik ağrı şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Travma durumu: Hastada travma şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- İdrar yakınmaları durumu: Hastada idrar yakınması şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Sara nöbeti durumu: Hastada sara nöbeti şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Burun kanaması durumu: Hastada burun kanaması şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Bilinç bulanıklığı durumu: Hastada bilinç bulanıklığı şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Nörolojik defisit durumu: Hastada nörolojik defisit şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Psikiyatrik rahatsızlıklar durumu: Hastada psikiyatrik rahatsızlık şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Kendine veya çevresine zarar verme durumu: Hastada kendine verme şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Kaşıntı durumu: Hastada kaşıntı şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Yanık ya da yangına maruziyet durumu: Hastada yanık olması durumu ve/veya yangına maruz kalınması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Vajinal kanama durumu: Hastada vajinal kanama şikayeti olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.

- Alkol-madde kullanımı, CO intoks durumu: Hastada alkol ya da madde kullanımı ve/veya karbonmonoksit maruziyet olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Gözden şikayet durumu: Hastada gözden şikayet olması durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.

Özgeçmiş bilgiler; acil servise başvuran hastanın güncelde sahip olduğu ve sağlık görevlilerine ilettiği hastalıklar ya da geçmişte başlamış etkisi halen devam eden hastalıklardır. Hasta acil servise başvurduğunda bu hastalıkları ilgililere kendisi iletmış olabilir ya da hastanenin hastane bilgi yönetim sisteminde önceki ziyaretlerinde, yani önceden kayıt altına alınmış olabilir. Çalışma boyunca toplanan veriler doğrultusunda özgeçmiş bilgileri aşağıdaki gibi gruplandırılmıştır. Eğer acil servise başvuran hasta bir ya da birkaç hastalığa sahipse o hasta için var olan özgeçmiş hastalık bilgileri 1 olarak, var olmayan özgeçmiş hastalık bilgileri 0 olarak kaydedilmiştir. Bazı hastaların özgeçmişinde herhangi bir hastalık olmayabilir. Bu sebeple bütün özgeçmiş verileri 0 olabilmektedir. Veri setinde toplam 12 tane özgeçmiş hastalık girdisi bulunmaktadır.

- Şeker hastalığı durumu (DM): Diyabet hastalığına sahip olma durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Hipertansiyon hastalığı durumu (HT): Hipertansiyon hastalığına sahip olma durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Koroner arter hastalığı durumu (KAH): Koroner arter hastalığına sahip olma durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Konjestif kalp yetmezliği hastalığı durumu (KKY): Konjestif kalp yetmezliği hastalığına sahip olma durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Ritim bozukluğu: Ritim bozukluğu hastalığına sahip olma durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Serebro vasküler hastalık durumu (SVO): Serebro vasküler hastalığına sahip olma durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- KOAH, astım, akciğer hastalıkları, pulmoner emboli: Koah, astım, pulmoner emboli, akciğer hastalığına sahip olma durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.

- Kronik böbrek yetmezliği (KBY), tek böbrek, böbrek nakli durumu: Kronik böbrek yetmezliği hastalığına sahip olma, tek böbrek ve/veya böbrek nakli olma durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Karaciğer yetmezliği, siroz, karaciğer nakli durumu: Karaciğer yetmezliği, siroz hastalığına sahip olma durumu, böbrek nakli olma durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Malignite ve komplikasyonları durumu: Malignite komplikasyonları yani kanser hastalığı olma durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Gebelik ve komplikasyonları durumu: Gebelik ve/veya komplikasyonları olma durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.
- Diğer nörolojik hastalıkların durumu: Diğer nörolojik hastalıklara sahip olma durumudur. 0 ise şikayet yok, 1 ise şikayet vardır. Kategorik bir veridir.

Durum bilgisi, hastanın acil servise geliş durumuyla ilgili bilgidir.

- Acile varış şekli: Hastanın acil servise kendi imkanları ile gelerek, ayakta başvurduğunu ya da 112 ambulans ile geldiğini belirtmektedir. 0 ise ayakta yani kendi imkanlarıyla, 1 ise 112 ambulans ile geldiğini belirtmektedir. Kategorik bir veridir.

Çıktı bilgileri haricindeki tüm bilgi başlıklarının hepsi girdiler kümesini oluşturmaktadır. Çünkü çıktı bilgileri birer sonuçtur. Acil servise başvuruya başlayan süreç çıktı bilgilerinden bir ya da birkaçı ile sonuçlanabilmektedir. Bunlar makine öğrenmesi modelinde tahmin edilecek alanı oluşturmaktadırlar.

Veri setinde 2 adet çıktı bilgisi sütunu bulunmaktadır. Bunlar; taburculuk durumu ve triyaj alan sınıflandırmasıdır.

- Taburculuk durumu: Hastanın acil servise başvurduktan sonra yatış mı yaptığını yoksa yatış yapmadan taburcu mu olduğunu belirtmektedir. Kategorik bir veridir.
- Triage alan sınıflandırması: Hastanın triyaj sınıflandırması yapıldıktan sonra hangi alanda acil sağlık hizmeti alacağını sonucudur. Kırmızı, sarı, yeşil alan sınıflandırmaları yapılmaktadır. Kategorik bir veridir.

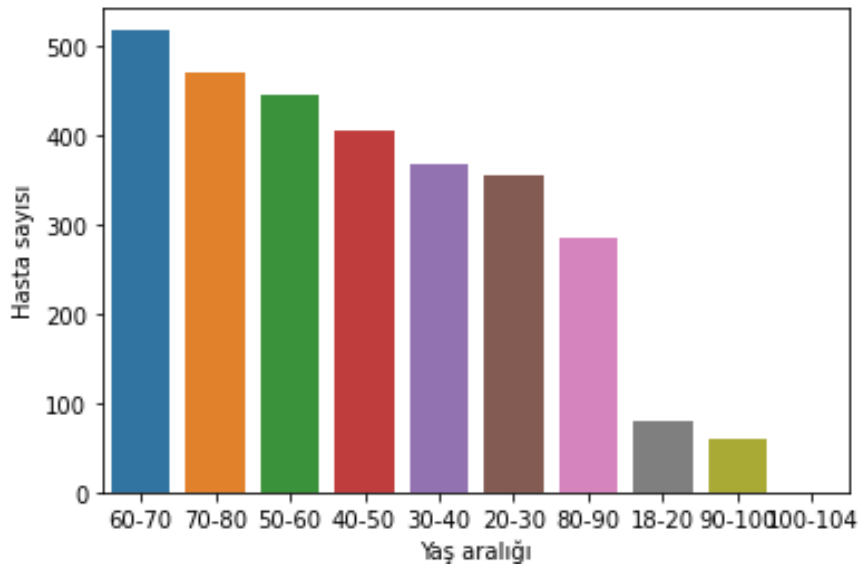
3.3.3. Veri setinin istatistiksel ve grafiksel analizi

Veri toplama sürecinde eksik veri olmayacak şekilde veriler toplanmıştır. Veri setinde eksik veri bulunmamaktadır. Nicel verilerin istatistiksel analizi Tablo 3.5'teki gibidir.

Tablo 3.5. Veri setinin istatistiksel analizi.

İndeks	yas	sistolik	diastolik	map	nabiz	ss	ates	spo2	kansekeri	gks
Adet	3.000	3.000	3.000	3.000	3.000	3.000	3.000	3.000	3.000	3.000
Ortalama	55,6	136	80,4	99,1	84,4	18,7	36,6	95,9	128	14,8
Std. Sapma	20,1	28,2	16	18,6	17,7	4,25	0,521	4,52	62,9	1,45
Min.	18	0	0	0	0	0	33,7	49	33	3
%25	39	119	70	87	72	16	36,3	95	97	15
%50	57	133	79	97	82	18	36,5	97	109	15
%75	72	152	89	109	94	20	36,7	99	137	15
Max.	104	281	186	212	202	45	41	100	1196	15

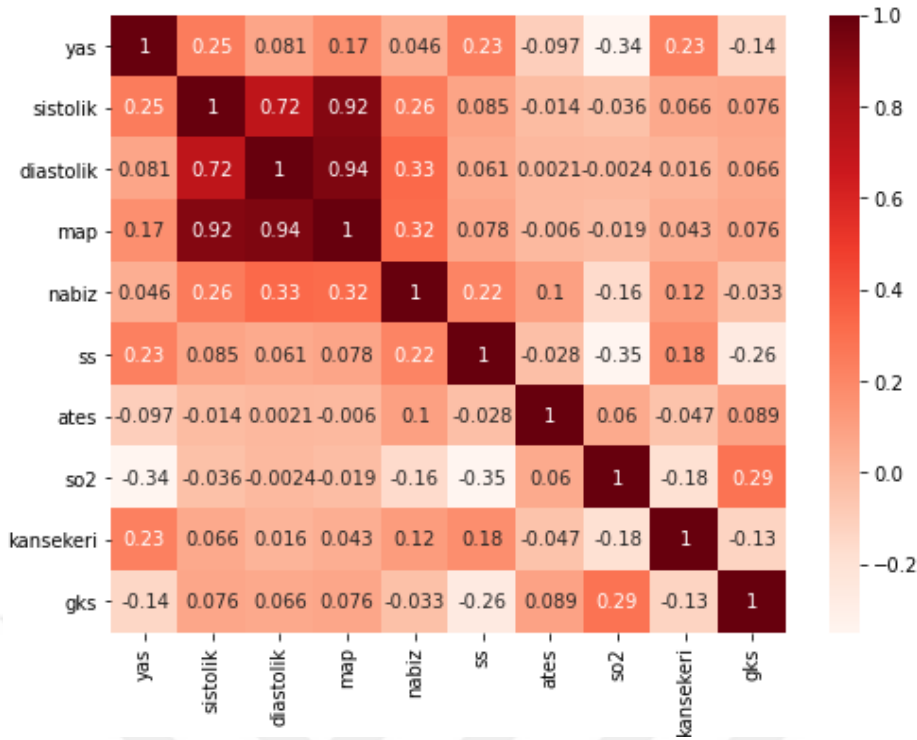
Hastaların yaş ortalaması 55.6, standart sapması 20.1'dir. En genç hasta 18 yaşında, en yaşlı hasta ise 104 yaşındadır. En çok 60-70 yaş aralığında, ikinci olarak 70-80 yaş aralığındaki hastaların yer aldığı ve 100 yaş ve üzeri hastaların verilerinin en az olduğu görülmektedir. Yaş gruplarıyla ilgili grafik Şekil 3.19'deki gibidir.



Şekil 3.19. Yaş grubu bazında hasta sayıları.

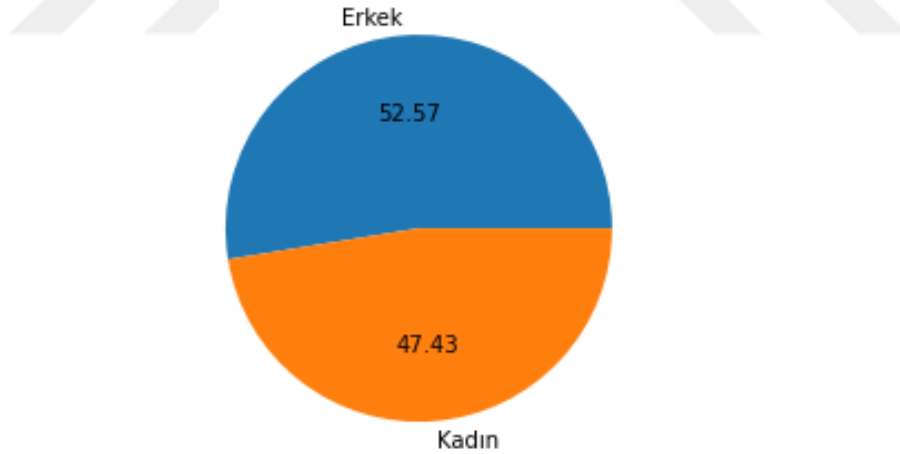
- Sistolik kan basıncı ortalaması 136 mmHg'dir. Standart sapması, 28.2 mmHg'dir. En düşük sistolik kan basıncı değeri 0, en yüksek kan basıncı değeri ise 281 mmHg'dir.
- Diyastolik kan basıncı ortalaması 80.4 mmHg'dir. En düşük sistolik kan basıncı değeri 0, en yüksek kan basıncı değeri ise 186 mmHg'dir. Standart sapması, 16 mmHg'dir.
- Ortalama arter basıncı (MAP), sistolik ve diyastolik kan basıncı değerlerine bağlı olmakla birlikte ortalama 99.1 mmHg'dir. Standart sapması, 18.6 mmHg'dir. En düşük MAP değeri 0, en yüksek MAP değeri 212 mmHg'dir.
- Hastaların nabız değeri ortalaması 84.4 atım/dakikadır. Standart sapması, 17.7 atım/dakikadır. En düşük nabız değeri 0, en yüksek nabız değeri 202 atım/dakikadır.
- Solunum sayısı ortalama 18.7 adet/dakika, standart sapması 4.25 adet/dakikadır. En düşük solunum sayısı 0, en yüksek solunum sayısı 45 adet/dakikadır.
- Vücut sıcaklığı ortalama değeri 36.6 C°, standart sapması 0.521 C°'dir. En düşük vücut sıcaklığı 33.7 C°, en yüksek vücut sıcaklığı 41 C°'dir.
- Satürasyon değeri ortalaması 95.9%, standart sapması 4.52 %'dir. En düşük satürasyon değeri 49%, en yüksek satürasyon değeri 100%'dür.
- Kan şekeri ortalama değeri 128 mg/dl, standart sapması 62,9 mg/dl'dir. En düşük kan şekeri değeri 33 mg/dl, en yüksek kan şekeri değeri 1196 mg/dl'dir.
- Veri setindeki hastaların Glasgow koma skalaları (GKS) incelendiğinde en düşük 3, en yüksek 15 değerini almışlardır.

Nicel değişkenler arasındaki ilişkinin yönü ve büyüklüğünü belirten katsayıyı bulmak için korelasyon değerleri hesaplanmıştır. Korelasyon, değişkenler arasındaki doğrusal ilişkinin yönünü ve gücünü belirlemeye yaramaktadır. Korelasyon değerleri ile bir ısı haritası oluşturulmuştur. Hesaplanan korelasyon değerleri ve ısı haritası Şekil 3.20'deki gibidir.



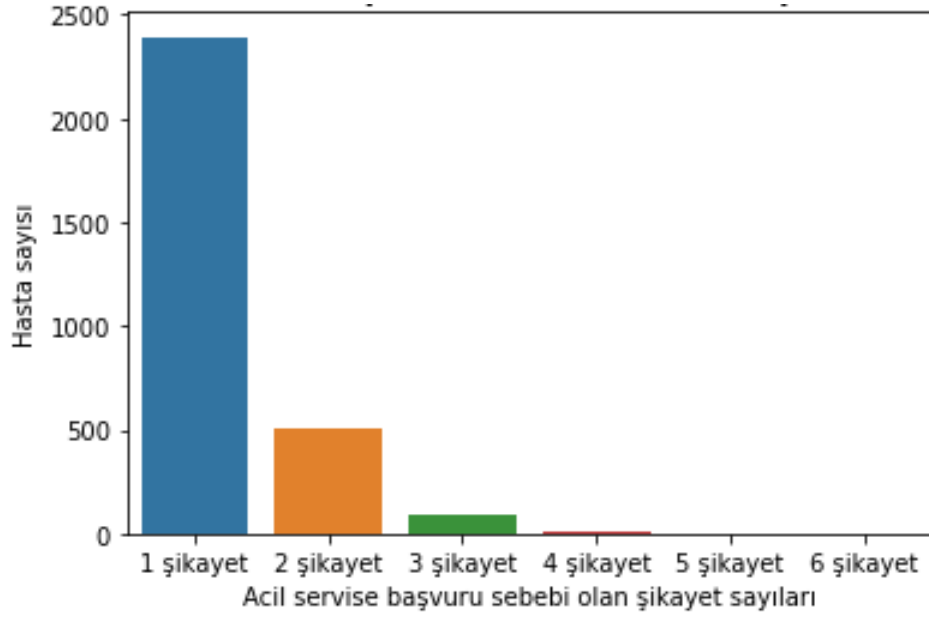
Şekil 3.20. Korelasyon ısı haritası.

Kategorik verilerin analizi yapıldığında; veri setinin %52,57'si erkek, %47,43'ü ise kadın hastalar oluşturmaktadır. Cinsiyet bazındaki durum Şekil 3.21'deki gibidir.



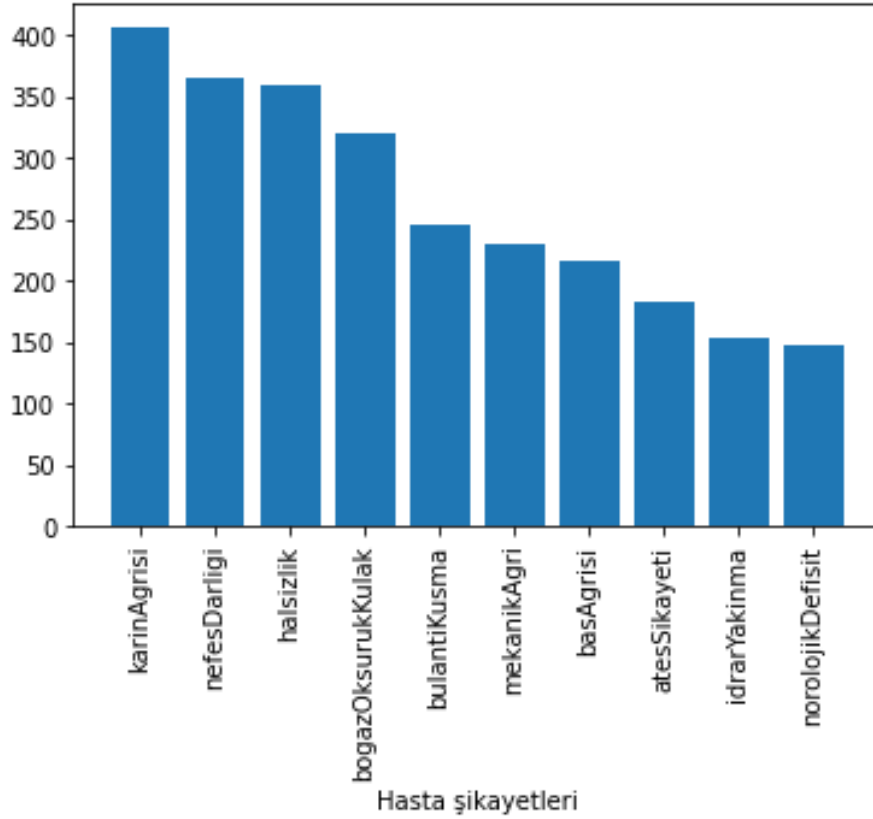
Şekil 3.21. Cinsiyet bazında pasta diyagramı.

Hastaların hastaneye başvururken sahip oldukları şikayet sayıları aşağıdaki Şekil 22'de özetlenmiştir. Bu grafiğe göre; hastaneye başvuran yaklaşık 2.393 hastanın acil servise başvururken 1 grup şikayete sahip olduğu, 503 hastanın 2 grup şikayete sahip olduğu, 90 hastanın 3 grup şikayete sahip olduğu, 11 hastanın 4 grup şikayete sahip olduğu, 2 hastanın 5 grup şikayete sahip olduğu, 1 hastanın da 6 grup şikayete sahip olduğu görülmektedir.



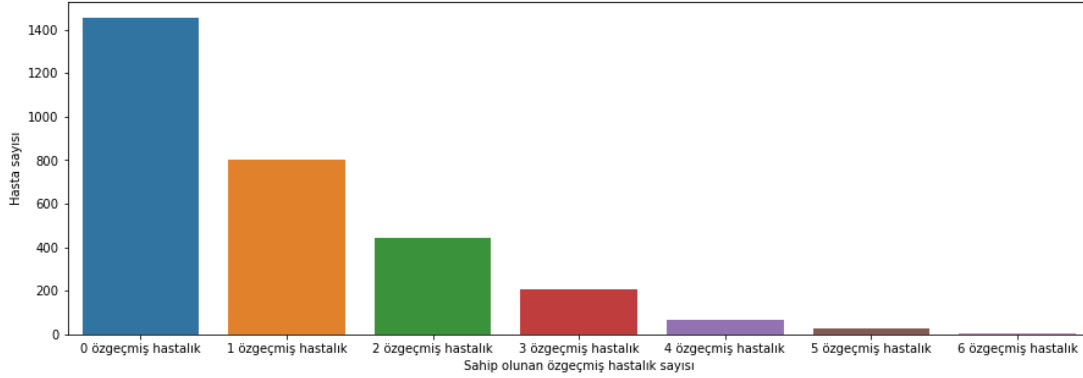
Şekil 3.22. Hastaların acil servise başvururken sahip olduğu şikayet sayısı durumu.

Hastaların şikayetleri incelendiğinde en çok karın ağrısı şikayeti ile sonrasında sırayla nefes darlığı ve halsizlik şikayetleriyle başvurulduğu Şekil 3.23'te görülmektedir.



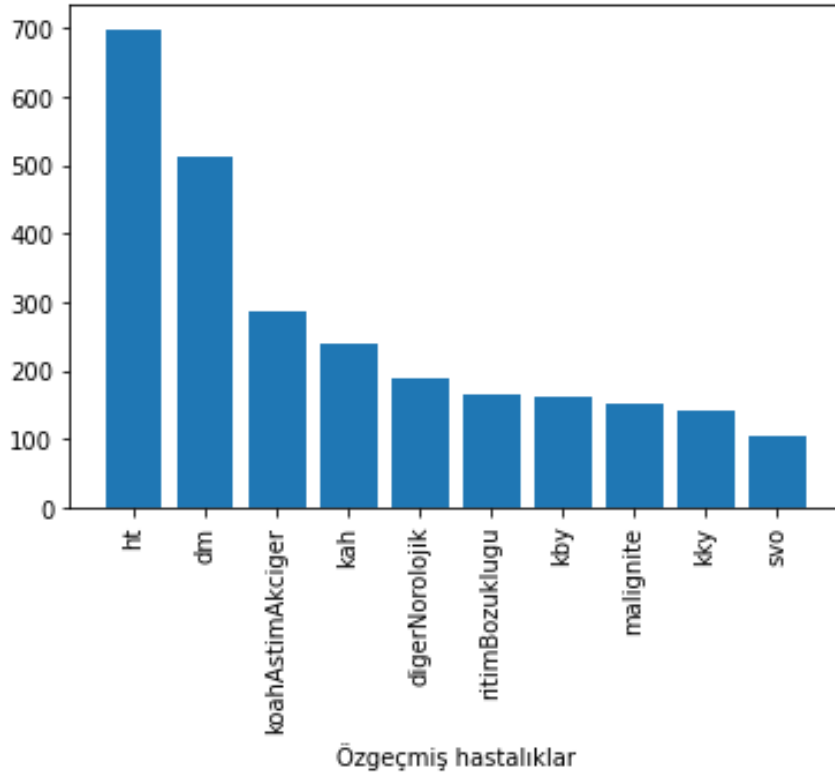
Şekil 3.23. Hastaların şikayet grupları ve adetleri grafiği.

Şekil 3.24'teki grafiğe göre; hastaneye başvuran yaklaşık 1.454 hastanın acil servise başvururken hiç özgeçmiş hastalığı olmadığı, 801 hastanın 1 grup özgeçmiş hastalığa, 442 hastanın 2 özgeçmiş hastalığa, 206 hastanın 3 özgeçmiş hastalığa, 68 hastanın 4 özgeçmiş hastalığa, 25 hastanın 5 özgeçmiş hastalığa ve 4 hastanın da 6 özgeçmiş hastalığa sahip olduğu görülmektedir.



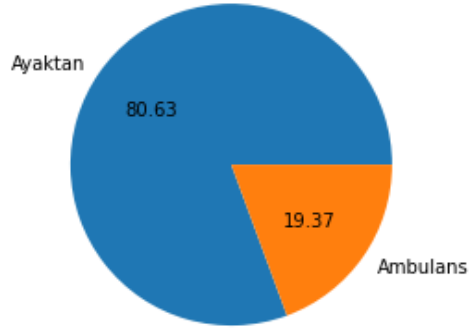
Şekil 3.24. Özgeçmiş hastalık sayıları grafiği.

Şekil 3.25'te özgeçmiş hastalık bilgileri incelendiğinde hastaların en sık sahip olduğu özgeçmiş hastalığın yüksek tansiyon olduğu görülmektedir. Sonrasında sırayla; diyabet, koah-astım-pulmoner emboli-akciğer hastalıkları olduğu görülmektedir.



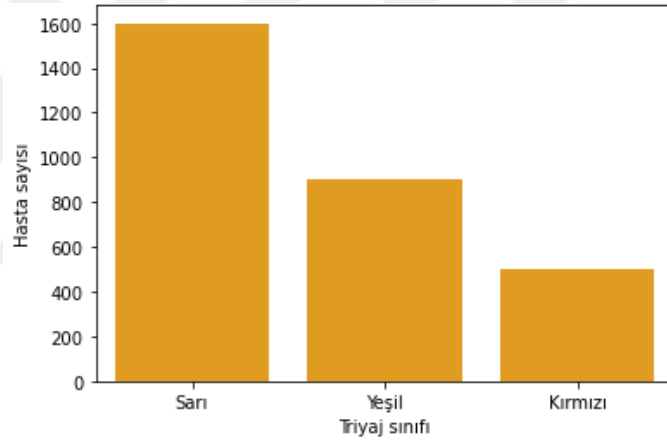
Şekil 3.25. Hastaların şikayet grupları ve adetleri grafiği.

Veri setindeki hastaların %80,63'ü acil servise kendi imkanlarıyla ayakta, %19,37'si ambulans ile acil servise gelerek sağlık hizmeti almıştır. Şekil 3.26'daki gibidir.



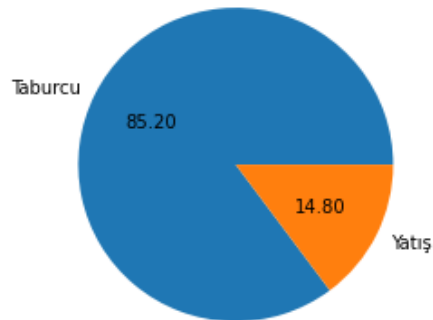
Şekil 3.26. Acile varış şekli pasta diyagramı.

Acil servise başvuran hastaların triyaj sınıflandırması; %53,33'ü sarı, %30'u yeşil, %16,67'si ise kırmızı alana yönlendirildiği görülmektedir. Şekil 3.27'deki gibidir.



Şekil 3.27. Triyaj alan sınıflandırması durumu.

Acil servise başvuran hastaların yatış-taburcu olma durumları incelendiğinde Şekil 3.28 deki gibi hastaların %85,20'si taburcu olduğu, %14,80'i yatış yaptığı görülmektedir.



Şekil 3.28. Yatış yapma ve taburcu olma durumları.

4. ARAŞTIRMA BULGULARI

4.1. Uygulamanın Amacı ve Veri Seti Hazırlık

Bu çalışmadaki amaç; acil servislerde yaşanan hasta yoğunluğu durumunu iyileştirmek, hastalara doğru zamanda, doğru alanda daha iyi bir hizmet verilmesini sağlamak, acil servisteki bekleme sürelerini azaltmak, hastane kaynaklarının daha verimli ve planlı kullanılmasını sağlamak amacıyla toplanan triyaj verileriyle sınıflandırma çalışması yapılarak makine öğrenmesi modelleri oluşturmak ve elde edilen bu modeller ile triyaj süreçlerinde makine öğrenmesi algoritmalarıyla tahmin çalışması yapmaktır. Acil servis triyaj sürecinden toplanan triyaj veri seti, hastaların triyaj verileri ile verileri alınan hastaların triyaj sınıflandırma sonuçları ile yatış-taburcu olma sonuçlarını içermektedir. Bu çalışmada toplanan triyaj verileri ile iki sonuç tahmin edilmiştir.

- Triyaj alan sınıflandırması
- Yatış-taburcu olma durumu

Trijaj alan tahmini; acil servise gelen hastanın yaşamsal bulguları belirlenip, şikayet, özgeçmiş hastalık bilgileri ve benzeri bilgileri alınıp; acil sağlık hizmetini kırmızı, sarı, yeşil alanlardan hangisinde alması gerektiğinin tahmin edilmesidir. Yatış-taburcu durumu tahmini ise; acil servise başvuran hastaların triyaj alan sınıflandırmasındaki ile aynı bilgilerinin alınıp, doktora muayenesinden sonra hastaneye yatış mı yapacağının yoksa taburcu mu olacağının tahmin edilmesidir.

Veri seti 3.000 adet satır içermektedir. Her satır bir hastanın triyaj bulgularını içerdiği için 3.000 hastayı içermektedir. Makine öğrenmesi algoritmaları kullanılarak öncelikle eğitilmiş bir model oluşturulacak ve oluşturulan model test edilecektir. Modeli eğitmek için eğitim verileri ve modeli test etmek için test verileri oluşturulmuştur. Bunun için veri setinin %80'i eğitim için, %20'si test verisi olarak ayrılmıştır. Bu veriler Python programında rastgele olacak şekilde ayrılmıştır. Bu durumda 2.400 hastanın verisi modeli eğitmek için, 600 hastanın verisi modeli test etmek için kullanılmıştır.

Sınıflandırma algoritmalarıyla tahmin çalışması için aşağıdaki algoritmalar ile çalışılmıştır. Algoritmalarda bazı parametre değişikliği yapılarak aynı algoritmalar ile hem triyaj alan hem de yatış-taburculuk durumu için çalışılmıştır.

- Karar ağaçları algoritması (Gini, Entropi)
- Doğrusal destek vektör makineleri algoritması
- Doğrusal olmayan destek vektör makineleri algoritması (RBF, Poly, Sigmoid)
- Naive bayes algoritması
- Lojistik regresyon algoritması
- K-En yakın komşu algoritması (Manhattan, Minkowski)
- Rassal orman algoritması (Gini, Entropi)

Eğitim veri setiyle algoritmalar eğitildikten sonra test verileri ile tahmin çalışması yapılmıştır. Yapılan tahmin çalışmasının performansını ölçmek için karışıklık matrisleri elde edilmiştir.

Karışıklık matrisinde triyaj alan tahmini için:

- 0, kırmızı alanı
- 1, sarı alanı
- 2, yeşil alanı temsil etmektedir.

Karışıklık matrisinde yatış-taburculuk durumu tahmini için;

- 0, taburcu olma durumunu
- 1, hastaneye yatış yapma durumunu temsil etmektedir.

Karışıklık matrislerinden yola çıkarak sınıflandırma raporları oluşturulmuştur. Sınıflandırma raporlarında algoritmaların tahmin performanslarına dair ölçütler elde edilmiştir. Daha sonrasında AUCROC değerleri hesaplanmıştır. Her iki tahmin çalışmasında kullanılan tüm algoritmalar için çapraz doğrulama yapılmıştır. Son olarak iki tahmin çalışmasında da elde edilen tüm performans sonuçlarının kıyaslanması yapılmıştır.

4.2. Trijaj Alan Tahmini Çalışması

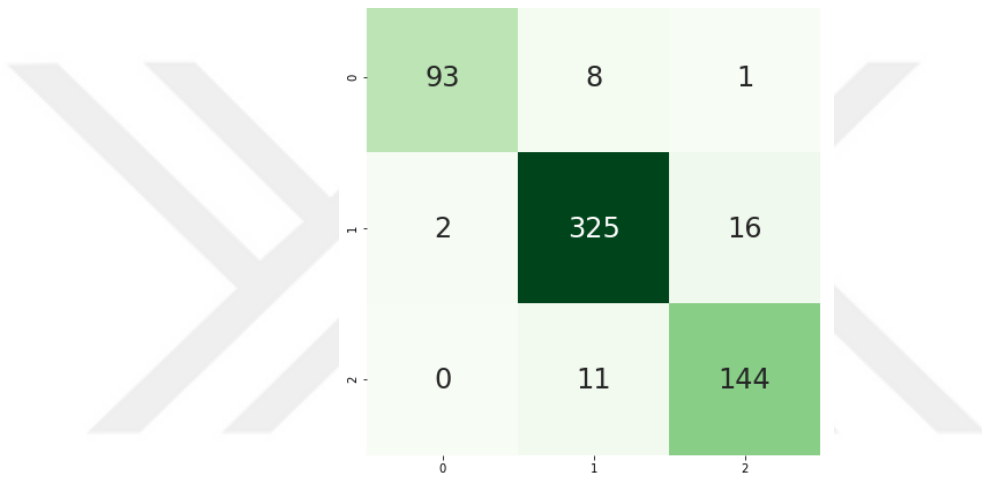
Trijaj alan tahmini çalışmasında acil servise başvuran hastaların triyaj sınıflandırma sonuçları tahmin edilmesi amaçlanmıştır.

4.2.1. Karar ağaçları algoritması

Triyaj alan tahmini yapılırken kullanılan sınıflandırma algoritmalarından biri karar ağaçları algoritmasıdır. Karar ağaçları algoritmasında kriter olarak “gini” ve “entropi” kullanılarak iki ayrı karar ağacı sınıflandırma modeli oluşturulmuştur. İki ayrı tahmin sonucu elde edilmiştir.

4.2.1.1. Karar ağacı-gini algoritması

Karar ağacı algoritmasında kriter gini kullanılarak model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.1’deki gibidir.



Şekil 4.1. Karar ağacı-gini algoritması karışıklık matrisi.

Karar ağacı-Gini algoritmasıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.1’deki gibidir.

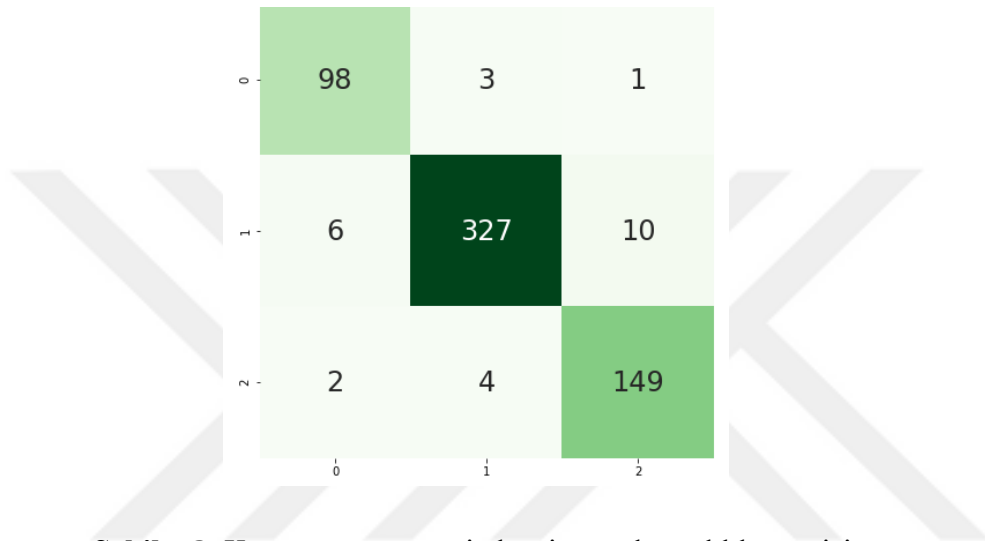
Tablo 4.1. Karar ağacı-gini algoritması sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Kırmızı (0)	%98	%91	%94	102
Sarı (1)	%94	%95	%95	343
Yeşil (2)	%89	%93	%91	155
Ortalama	%94	%93	%93	600

Karar ağacı-Gini algoritmasıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %93.6, ortalama kesinlik oranı %94, ortalama duyarlılık oranı %93 ve ortalama F1 skor oranı %93'tür.

4.2.1.2. Karar ağacı-entropi algoritması

Karar ağacı algoritmasında kriter entropi kullanılarak model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.2'deki gibidir.



Şekil 4.2. Karar ağacı-entropi algoritması karışıklık matrisi.

Karar ağacı-Entropi algoritmasıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.2'deki gibidir.

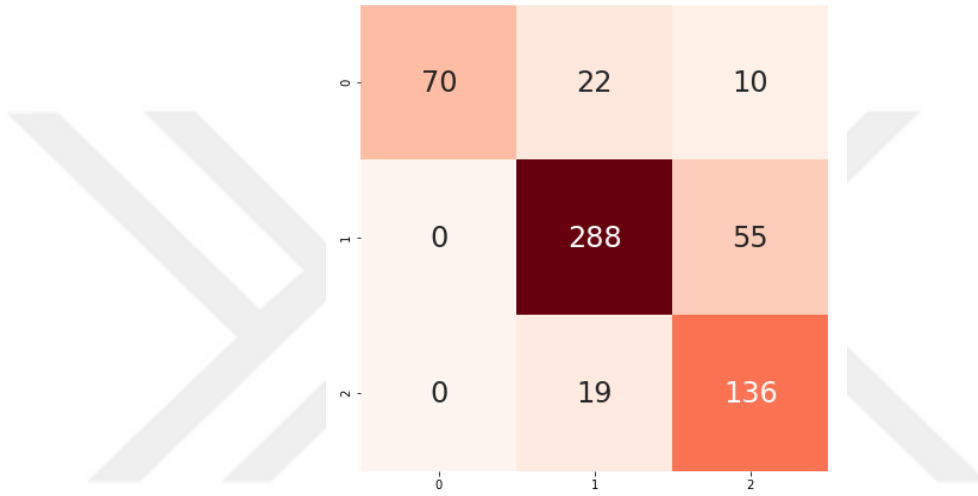
Tablo 4.2. Karar ağacı-entropi algoritması sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Kırmızı (0)	%92	%96	%94	102
Sarı (1)	%98	%95	%97	343
Yeşil (2)	%93	%96	%95	155
Ortalama	%94	%96	%95	600

Karar ağacı-Entropi algoritmasıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %95.6, ortalama kesinlik oranı %94, ortalama duyarlılık oranı %96 ve ortalama F1 skor oranı %95'tir.

4.2.2. Doğrusal destek vektör makineleri algoritması

Triyaj alan tahmini yapılırken kullanılan sınıflandırma algoritmalarından biri doğrusal destek vektör algoritmasıdır. Doğrusal destek vektör makinesi algoritmasıyla model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.3'teki gibidir.



Şekil 4.3. Doğrusal destek vektör makinesi algoritması karışıklık matrisi.

Doğrusal destek vektör algoritmasıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.3'teki gibidir.

Tablo 4.3. Doğrusal destek vektör algoritması sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Kırmızı (0)	%100	%69	%81	102
Sarı (1)	%88	%84	%86	343
Yeşil (2)	%68	%88	%76	155
Ortalama	%85	%80	%81	600

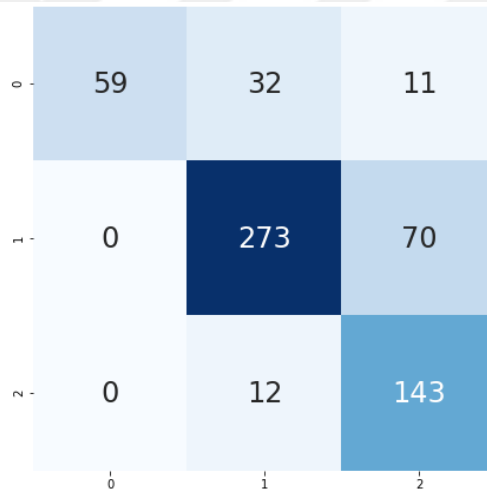
Doğrusal destek vektör algoritmasıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %82.3, ortalama kesinlik oranı %85, ortalama duyarlılık oranı %80 ve ortalama F1 skor oranı %81'dir.

4.2.3. Doğrusal olmayan destek vektör makineleri algoritması

Triyaj alan tahmini yapılırken kullanılan sınıflandırma algoritmalarından biri doğrusal olmayan destek vektör algoritmasıdır. Destek vektör algoritmasıyla model eğitilirken bu çalışmada 3 çeşit çekirdek kullanılmıştır. Bu çekirdekler: Radyal tabanlı çekirdek yani RBF, poly çekirdek ve sigmoid çekirdek. Üç çekirdeğin hepsiyle sınıflandırma çalışması yapılmıştır. Üç ayrı tahmin sonucu elde edilmiştir.

4.2.3.1. Doğrusal olmayan destek vektör algoritması-RBF

Radyal tabanlı RBF çekirdeği kullanılarak destek vektör makinesi ile model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.4'teki gibidir.



0	59	32	11
1	0	273	70
2	0	12	143
	0	1	2

Şekil 4.4. Doğrusal olmayan destek vektör makinesi-RBF karışıklık matrisi.

Doğrusal olmayan destek vektör algoritması RBF çekirdeğiyle yapılan tahmin çalışmasında modelin; doğruluk oranı %79.1, ortalama kesinlik oranı %83, ortalama duyarlılık oranı %77 ve ortalama F1 skor oranı %77'dir.

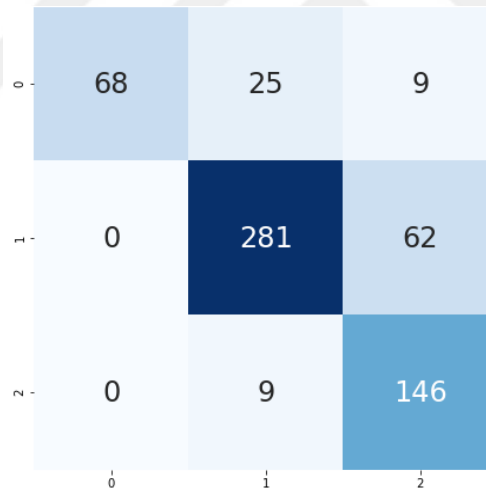
RBF çekirdeği ile yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.4'teki gibidir.

Tablo 4.4. SVM-RBF sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Kırmızı (0)	%100	%58	%73	102
Sarı (1)	%86	%80	%83	343
Yeşil (2)	%64	%92	%75	155
Ortalama	%83	%77	%77	600

4.2.3.2. Doğrusal olmayan destek vektör algoritması-poly

Poly çekirdeği kullanılarak destek vektör makinesi ile model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.5'teki gibidir.

**Şekil 4.5.** Doğrusal olmayan destek vektör makinesi-poly karışıklık matrisi.

Doğrusal olmayan destek vektör algoritması poly çekirdeğiyle yapılan tahmin çalışmasında modelin; doğruluk oranı %82.5, ortalama kesinlik oranı %85, ortalama duyarlılık oranı %81 ve ortalama F1 skor oranı %81'dir.

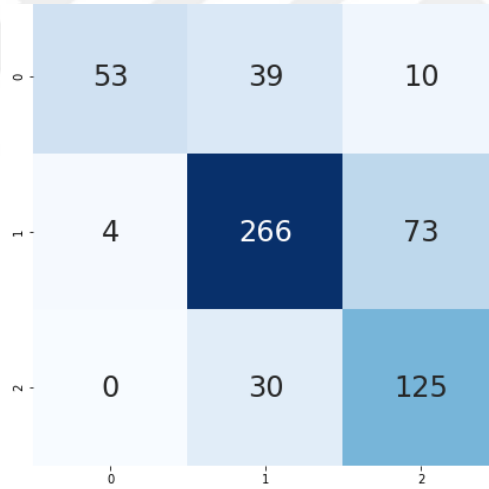
Poly çekirdeği ile yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.5'teki aşağıdaki gibidir.

Tablo 4.5. SVM-Poly sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Kırmızı (0)	%100	%67	%80	102
Sarı (1)	%89	%82	%85	343
Yeşil (2)	%67	%94	%78	155
Ortalama	%85	%81	%81	600

4.2.3.3. Doğrusal olmayan destek vektör algoritması-sigmoid

Sigmoid çekirdeği kullanılarak destek vektör makinesi ile model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.6'daki gibidir.

**Şekil 4.6.** Doğrusal olmayan destek vektör makinesi-sigmoid karışıklık matrisi.

Doğrusal olmayan destek vektör algoritması sigmoid çekirdeğiyle yapılan tahmin çalışmasında modelin; doğruluk oranı %74, ortalama kesinlik oranı %77, ortalama duyarlılık oranı %70 ve ortalama F1 skor oranı %71'dir.

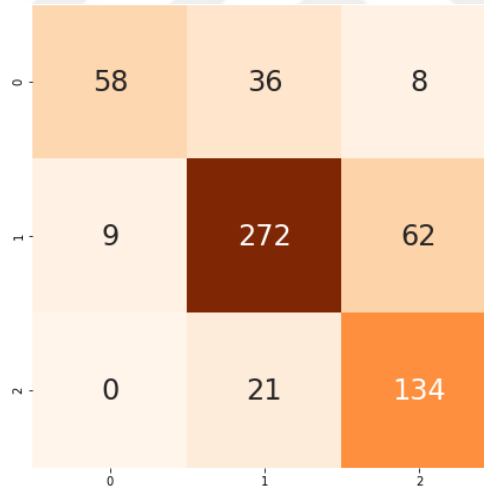
Sigmoid çekirdeği ile yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.6'daki gibidir.

Tablo 4.6. SVM-Sigmoid sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Kırmızı (0)	%93	%52	%67	102
Sarı (1)	%79	%78	%78	343
Yeşil (2)	%60	%81	%69	155
Ortalama	%77	%70	%71	600

4.2.4. Naive bayes algoritması

Triyaj alan tahmini yapılırken kullanılan sınıflandırma algoritmalarından biri naive bayes algoritmasıdır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.7'deki gibidir.

**Şekil 4.7.** Naive bayes algoritması karışıklık matrisi.

Naive bayes algoritmasıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %77.3, ortalama kesinlik oranı %78, ortalama duyarlılık oranı %74 ve ortalama F1 skor oranı %77'dir.

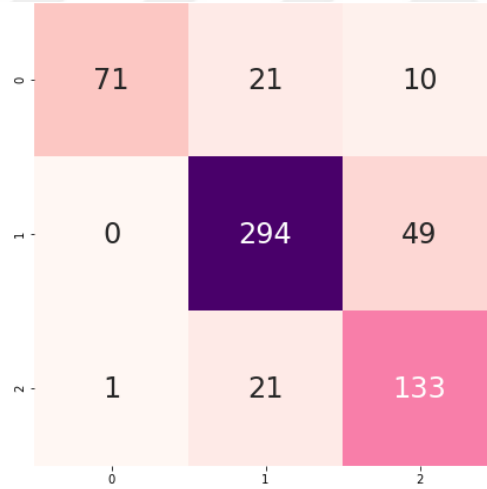
Naive bayes algoritmasıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.7'deki gibidir.

Tablo 4.7. Naive bayes algoritması sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Kırmızı (0)	%87	%57	%69	102
Sarı (1)	%83	%79	%81	343
Yeşil (2)	%66	%86	%75	155
Ortalama	%78	%74	%77	600

4.2.5. Lojistik regresyon algoritması

Triyaj alan tahmini yapılırken kullanılan sınıflandırma algoritmalarından biri lojistik regresyon algoritmasıdır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.8'deki gibidir.



Şekil 4.8. Lojistik regresyon algoritması karışıklık matrisi.

Lojistik regresyon algoritmasıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %83, ortalama kesinlik oranı %85, ortalama duyarlılık oranı %80 ve ortalama F1 skor oranı %82'dir.

Lojistik regresyon algoritmasıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.8'deki gibidir.

Tablo 4.8. Lojistik regresyon algoritması sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Kırmızı (0)	%99	%70	%82	102
Sarı (1)	%88	%86	%87	343
Yeşil (2)	%69	%86	%77	155
Ortalama	%85	%80	%82	600

4.2.6. K-En yakın komşu algoritması

Triyaj alan tahmini yapılırken kullanılan sınıflandırma algoritmalarından biri k-en yakın komşu algoritmasıdır. 2 ayrı uzaklık ölçütü kullanılarak 2 ayrı sınıflandırma modeli geliştirilmiştir. Bu uzaklıklar: Manhattan ve Minkowski uzaklıklarıdır. Her şıkş model eğitilmeden önce en iyi doğruluk oranı veren k değerinin tespiti için algoritma 1-50 arasında k değeriyle ayrı ayrı çalıştırılmış ve ayrı ayrı k değerleri bulunmuştur.

4.2.6.1. K-En yakın komşu algoritması-manhattan

K-En yakın komşu algoritmasında manhattan uzaklığı kullanılarak ve k-komşu sayısı=4 olarak kullanılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.9'deki gibidir.

0	65	26	11
1	11	268	64
2	0	10	145
	0	1	2

Şekil 4.9. KNN-Manhattan algoritması karışıklık matrisi.

K-En yakın komşu algoritması manhattan uzaklığıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.9'deki gibidir.

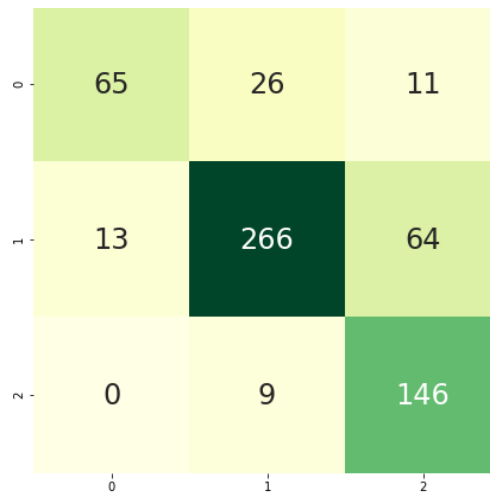
Tablo 4.9. KNN-Manhattan algoritması sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Kırmızı (0)	%86	%64	%73	102
Sarı (1)	%88	%78	%83	343
Yeşil (2)	%66	%94	%77	155
Ortalama	%80	%78	%78	600

K-En yakın komşu algoritması manhattan uzaklığıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %79.6, ortalama kesinlik oranı %80, ortalama duyarlılık oranı %78 ve ortalama F1 skor oranı %78'dir.

4.2.6.2. K-En yakın komşu algoritması-minkowski

K-En yakın komşu algoritmasında minkowski uzaklığı ve k-komşu sayısı=4 olarak alınmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.10'deki gibidir.



Şekil 4.10. KNN-Minkowski algoritması karışıklık matrisi.

K-En yakın komşu algoritması minkowski uzaklığıyla yapılan sınıflandırmada modelin sınıf bazında performans ölçütleri Tablo 4.10'daki gibidir.

Tablo 4.10. KNN-Minkowski algoritması sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Kırmızı (0)	%83	%64	%72	102
Sarı (1)	%88	%78	%83	343
Yeşil (2)	%66	%94	%78	155
Ortalama	%79	%78	%77	600

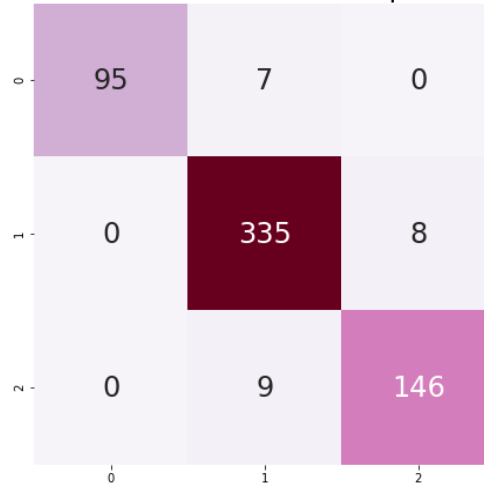
K-En yakın komşu algoritması minkowski uzaklığıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %79.5, ortalama kesinlik oranı %79, ortalama duyarlılık oranı %78 ve ortalama F1 skor oranı %77'dir.

4.2.7. Rassal orman algoritması

Triyaj alan tahmini yapılırken kullanılan sınıflandırma algoritmalarından biri rassal orman algoritmasıdır. Rassal orman algoritmasında kriter olarak “gini” ve “entropi” kullanılarak iki ayrı sınıflandırma modeli oluşturulmuştur. Elde edilen sınıflandırma modelleri, test verileriyle test edilmiş ve ayrı ayrı iki tane karışıklık matrisi elde edilmiştir. Karışıklık matrisleri doğrultusunda modellerin sınıflandırma raporları oluşturulmuştur.

4.2.7.1. Rassal orman algoritması-entropi

Rassal orman – Entropi algoritması kullanılarak model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.11'deki gibidir.



Şekil 4.11. Rassal orman algoritması-entropi karışıklık matrisi.

Rassal orman – Entropi algoritmasıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.11’deki gibidir.

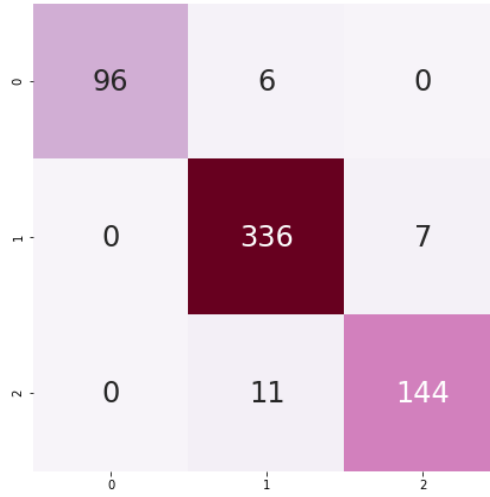
Tablo 4.11. Rassal orman-entropi algoritması sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Kırmızı (0)	%100	%93	%96	102
Sarı (1)	%95	%98	%97	343
Yeşil (2)	%95	%94	%94	155
Ortalama	%97	%95	%96	600

Rassal orman – Entropi algoritmasıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %96, ortalama kesinlik oranı %97, ortalama duyarlılık oranı %95 ve ortalama F1 skor oranı %96’dır.

4.2.7.2. Rassal orman algoritması-gini

Rassal orman – Gini algoritması kullanılarak model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.12’deki gibidir.



Şekil 4.12. Rassal orman algoritması-gini karışıklık matrisi.

Rassal orman – Gini algoritmasıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.12’deki gibidir.

Tablo 4.12. Rassal orman-gini algoritması sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Kırmızı (0)	%100	%94	%97	102
Sarı (1)	%95	%98	%97	343
Yeşil (2)	%95	%93	%94	155
Ortalama	%97	%95	%96	600

Rassal orman – Gini algoritmasıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %96, ortalama kesinlik oranı %97, ortalama duyarlılık oranı %95 ve ortalama F1 skor oranı %96’dır.

4.2.8. AUCROC değer hesabı

ROC eğrisi makine öğrenmesi modelinin performansının değerlendirilmesi için kullanılan bir ölçüttür. ROC eğrisi, eğrinin altında kalan alana göre yorumlanmaktadır. Eğrinin altında kalan AUC değerini vermektedir. Eğrinin altında kalan alan büyükse yani yüksek bir ROC varsa modelin başarısı yüksek, eğrinin altında daha küçükse

modelin başarısı düşük olarak yorumlanmaktadır. Modellerin performanslarının AUC değerleri Tablo 4.13’te özetlenmiştir.

Tablo 4.13. Triyaj alan tahmin modeli AUROC değerleri.

Algoritma	AUCROC değeri
Karar ağacı-Entropi algoritması (KA-Entropi)	96,79%
Rassal orman algoritması-Entropi (RF-Entropi)	96,16%
Rassal orman algoritması-Gini (RF-Gini)	96,13%
Karar ağacı-Gini algoritması (KA-Gini)	94,53%
Doğrusal olmayan destek vektör algoritması-Poly (SVM-Poly)	85,59%
Lojistik regresyon algoritması (LOGR)	85,22%
Doğrusal destek vektör algoritması (DSVM)	84,96%
K-En küçük komşu algoritması-Minkowski (KNN-Minkowski)	83,73%
K-En küçük komşu algoritması-Manhattan (KNN-Manhattan)	83,72%
Doğrusal olmayan destek vektör algoritması-RBF (SVM-RBF)	82,39%
Naive Bayes algoritması (NB)	80,48%
Doğrusal olmayan destek vektör algoritması-Sigmoid (SVM – Sigmoid)	77,30%

Tablo 4.13’e bakıldığında en yüksek AUCROC değeri %96,79 elde edilen algoritmanın karar ağacı-entropi olduğu görülmektedir.

4.2.9. Algoritma performansı değerlendirme

Triyaj sınıflandırmasında, algoritma ve kullanılan farklı kriterler dahil 12 farklı model elde edilmiştir. Karışıklık matrislerinden yola çıkılarak hesaplanan performans ölçütleri Tablo 4.14'teki gibidir.

Tablo 4.14. Triyaj alan tahmin çalışması genel sınıflandırma raporları.

Algoritma	Kesinlik değeri	Duyarlılık değeri	F Skor	Doğruluk değeri	AUCROC değeri
Karar ağacı-Entropi algoritması	94,00%	96,00%	95,00%	95,60%	96,79%
Rassal orman algoritması-Entropi	97,00%	95,00%	96,00%	96,00%	96,16%
Rassal orman algoritması-Gini	97,00%	95,00%	96,00%	96,00%	96,13%
Karar ağacı-Gini algoritması	94,00%	93,00%	93,00%	93,60%	94,53%
Doğrusal olmayan destek vektör algoritması-Poly	85,00%	81,00%	81,00%	82,50%	85,59%
Lojistik regresyon algoritması	85,00%	80,00%	82,00%	83,00%	85,22%
Doğrusal destek vektör algoritması	85,00%	80,00%	81,00%	82,30%	84,96%
K-En küçük komşu algoritması-Minkowski	79,00%	78,00%	77,00%	79,50%	83,73%
K-En küçük komşu algoritması-Manhattan	80,00%	78,00%	78,00%	79,60%	83,72%
Doğrusal olmayan destek vektör algoritması-RBF	83,00%	77,00%	77,00%	79,10%	82,39%
Naive Bayes algoritması	78,00%	74,00%	77,00%	77,30%	80,48%
Doğrusal olmayan destek vektör algoritması-Sigmoid	77,00%	70,00%	71,00%	74,00%	77,30%

Tablo 4.14'e göre;

- Modellerin test verileriyle test edilmesi sonucunda en yüksek doğruluk oranına sahip algoritma rassal orman algoritmasıdır. Rassal orman algoritmasında

entropi ve gini olmak üzere iki kriterle çalışılmıştır. İki model ile elde edilen doğruluk oranı %96'dır.

- Ortalama kesinlik değerlerine bakıldığında; en yüksek kesinlik oranına sahip algoritma ise rassal orman algoritmasıdır. Hem entropi hem de gini kriterleriyle elde edilen modellerde %97 ile en yüksek kesinlik değerine sahiptir.
- Ortalama duyarlılık değerlerine bakıldığında; en yüksek duyarlılık oranına sahip algoritma karar ağacı- entropi algoritmasıdır. Entropi kriteriyle elde edilen modelde %96 ile en yüksek duyarlılık değeri elde edilmiştir.
- Ortalama F skoruna bakıldığında; en yüksek F skoruna sahip algoritma ise rassal orman algoritmasıdır. Hem entropi hem de gini kriterleriyle elde edilen modellerde %96 ile en yüksek F skoruna sahiptir.
- AUCROC değerlerine bakıldığında en yüksek AUCROC %96,79 değerine sahip algoritma karar ağacı-entropi algoritmasıdır.

4.2.10. Triyaj alan tahmini için çapraz doğrulama

Tablo 4.14'te sonuçlar veri setinin %80'i eğitim için %20'si test için kullanılarak bir kere eğitim ve test sonucu elde edilen sonuçlardır.

Trijaj sınıflandırması için veri setiyle çapraz doğrulama yapılması için veri seti rassal parçalara bölünür ve bölünen her bir parçanın hem eğitim setinde hem de test setinde yer alması sağlanmıştır. Böylelikle verinin, veri setindeki dağılımından kaynaklanan sapma ve hataların en aza indirilmesi hedeflenmiştir. Çalışmada, çapraz doğrulma yapılırken verinin sırayla değil rassal olarak bölünmesi ve işlemin 10 kere tekrar etmesi sağlanmıştır.

Modeller, farklı veri versiyonlarıyla 10 kere eğitilmiş ve test edilmiştir. Her adımda elde edilen doğruluk oranları hesaplanmıştır. Ardından çapraz doğrulama doğruluk oranı, her adımda bulunan doğruluk oranlarının ortalaması hesaplanarak elde edilmiştir.

Bu çalışmayla elde edilen doğruluk değeri sonuçları Tablo 4.15'teki gibidir.

Tablo 4.15. Triyaj alan için çapraz doğrulama doğruluk oranları (%).

Algoritma	1.	2.	3.	4.	5.	6.	7.	8.	9.	10	Ort. Doğruluk
KA-Gini	98,3	97,8	98,5	97,5	97,5	93,8	95,2	95,8	98,3	97,7	97,0
KA- Entropi	97,2	97,7	96,3	95,7	96,0	94,3	97,8	97,7	97,0	97,7	96,7
RF-Entropi	93,8	94,7	95,5	93,8	96,0	93,5	94,3	95,7	95,0	95,2	94,8
RF-Gini	94,3	93,2	93,2	94,8	94,5	95,0	95,0	94,5	94,8	93,7	94,3
DSVM	84,5	83,2	82,5	85,2	82,5	82,5	83,8	86,3	82,2	83,3	83,6
LOGR	79,8	83,2	81,5	84,3	83,2	85,5	83,7	82,3	85,2	82,3	83,1
SVM-Poly	84,0	82,5	85,3	81,8	82,7	80,0	81,7	82,2	82,5	84,5	82,7
SVM-RBF	80,0	80,0	81,0	80,8	78,7	80,2	81,3	80,7	79,2	78,5	80,0
KNN- Manhattan	79,3	81,2	78,2	81,3	79,2	80,0	78,7	80,7	79,8	76,5	79,5
KNN - Minkowski	79,7	79,2	78,0	75,7	76,2	80,5	79,0	79,7	81,2	83,2	79,2
NB	76,2	76,3	73,7	76,5	72,3	77,2	76,3	78,2	75,3	74,8	75,7
SVM- Sigmoid	71,5	71,5	76,5	74,0	70,3	69,3	79,2	73,0	73,5	73,2	73,2

Tablo 4.15'e bakıldığında aynı veri setinin farklı parçalarıyla 10 kere eğitilme ve 10 kere test edilmesi sonucunda elde edilen doğruluk oranları görülmektedir. Bu doğruluk oranlarının ortalamalarına bakıldığında en yüksek doğruluk oranını karar ağacı-gini algoritmasıyla elde edilmiştir. İkinci olarak ortalama en yüksek doğruluk oranı karar ağacı-entropi algoritmasıyla elde edilmiştir.

Çapraz doğrulama çalışmasında her parçanın eğitim süresi, test süresi, kesinlik, duyarlılık değerleri ve F skorları hesaplanmıştır. 10 doğrulama için hesaplanan bu verilerin ortalaması Tablo 4.16'da verilmiştir.

Tablo 4.16. Triyaj alan için çapraz doğrulama sınıflandırma raporu.

Algoritma	Ortalama eğitim süresi	Ortalama test süresi	Ortalama kesinlik	Ortalama duyarlılık	Ortalama F skor
KA-Gini	0,0201	0,0067	96,42%	96,34%	96,36%
KA-Entropi	0,0218	0,0066	97,52%	97,01%	97,25%
RF-Entropi	0,1302	0,0133	95,56%	94,30%	94,87%
RF-Gini	0,1168	0,0127	95,25%	93,76%	94,45%
DSVM	0,1361	0,0276	87,21%	80,38%	82,26%
LOGR	0,2415	0,0109	86,66%	80,04%	82,08%
SVM-Poly	0,1550	0,0349	85,34%	79,91%	81,11%
SVM-RBF	0,2082	0,0995	85,40%	77,03%	78,83%
KNN-Manhattan	0,0037	0,1357	81,00%	75,35%	76,73%
KNN -Minkowski	0,0032	0,0647	79,56%	75,27%	76,39%
NB	0,0073	0,0089	78,11%	71,22%	72,91%
SVM-Sigmoid	0,2566	0,0573	76,27%	70,78%	72,28%

Çapraz doğrulama yapılmadan, sadece veri setinin %80 eğitim, %20 test verisi olarak kullanılarak elde edilen sonuçlarda en yüksek doğruluk oranı rassal orman-gini ve rassal orman-entropi algoritması ile %96 oranla elde edilirken; çapraz doğrulama yapıldığında en yüksek doğruluk oranı karar ağacı-gini algoritmasıyla %97 oranla elde edilmiştir. Bu sebeple çapraz doğrulama yapmak sınıflandırma modellerinde büyük öneme sahiptir.

Tablo 4.16’da;

- Ortalama kesinlik değerine, ortalama duyarlılık değerlerine ve ortalama F skoruna bakıldığında; üç ölçütte de en yüksek değerler elde edilen algoritma karar ağacı-entropi algoritmasıdır.

- Ortalama eğitim süresi en kısa olan k-en yakın komşu-minkowski algoritmasıdır. Ortalama test süresi en kısa olan karar ağacı-entropi algoritmasıdır.

4.3. Yatış-Taburcu Durumu Tahmini Çalışması

Yatış-taburcu durumu tahmini çalışmasında acil servise başvuran hastaların hastaneye yatış mı yaptığı yoksa taburcu mu olduğu sonuçları tahmin edilmesi amaçlanmıştır. Sonrasında algoritmaların performanslarını ölçebilmek için karışıklık matrisleri ve sınıflandırma raporları elde edilip, AUCROC değerleri hesaplanmıştır ve son olarak çapraz doğrulama işlemi yapılmıştır.

4.3.1. Karar ağaçları algoritması

Yatış-taburcu durumu tahmini yapılırken kullanılan sınıflandırma algoritmalarından biri karar ağaçları algoritmasıdır. Karar ağaçları algoritmasında kriter olarak “gini” ve “entropi” kullanılarak iki ayrı karar ağacı sınıflandırma modeli oluşturulmuştur. İki ayrı tahmin sonucu elde edilmiştir.

4.3.1.1. Karar ağacı-gini algoritması

Karar ağacı algoritmasında kriter gini kullanılarak model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.13'teki gibidir.

0	515	5
1	52	28
	0	1

Şekil 4.13. Karar ağacı-gini algoritması karışıklık matrisi.

Karar ağacı-Gini algoritmasıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.17'deki gibidir.

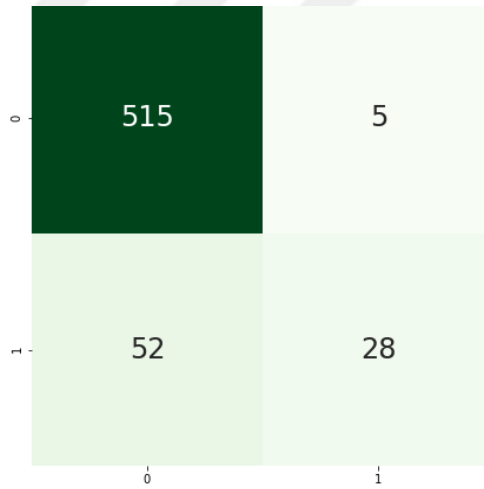
Tablo 4.17. Karar ağacı-gini algoritması sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Taburcu (0)	%91	%99	%95	520
Yatış (1)	%85	%35	%50	80
Ortalama	%88	%67	%72	600

Karar ağacı-Gini algoritmasıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %90.5, ortalama kesinlik oranı %88, ortalama duyarlılık oranı %67 ve ortalama F1 skor oranı %72'dir.

4.3.1.2. Karar ağacı-entropi algoritması

Karar ağacı algoritmasında kriter entropi kullanılarak model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.14'teki gibidir.



Şekil 4.14. Karar ağacı-entropi algoritması karışıklık matrisi.

Karar ağacı-Entropi algoritmasıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.18'deki gibidir.

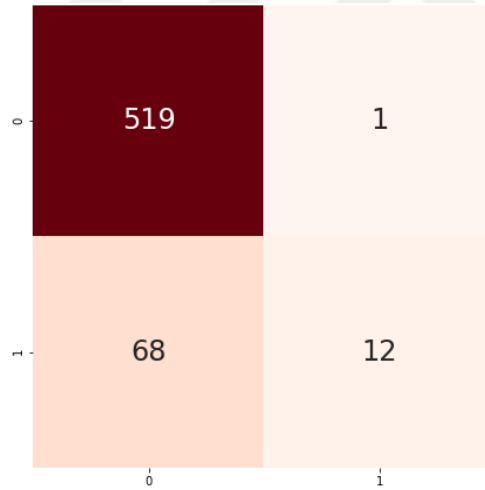
Karar ağacı-Entropi algoritmasıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %90.5, ortalama kesinlik oranı %88, ortalama duyarlılık oranı %67 ve ortalama F1 skor oranı %72'dir.

Tablo 4.18. Karar ağacı-entropi algoritması sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Taburcu (0)	%91	%99	%95	520
Yatış (1)	%85	%35	%50	80
Ortalama	%88	%67	%72	600

4.3.2. Doğrusal destek vektör makineleri algoritması

Yatış-taburcu durumu tahmini yapılırken kullanılan sınıflandırma algoritmalarından biri doğrusal destek vektör algoritmasıdır. Doğrusal destek vektör makinesi algoritmasıyla model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.15'teki gibidir.

**Şekil 4.15.** Doğrusal destek vektör makinesi algoritması karışıklık matrisi.

Doğrusal destek vektör algoritmasıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.19'daki gibidir.

Doğrusal destek vektör algoritmasıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %88.5, ortalama kesinlik oranı %90, ortalama duyarlılık oranı %57 ve ortalama F1 skor oranı %60'tır.

Tablo 4.19. Doğrusal destek vektör algoritması sınıflandırma raporu.

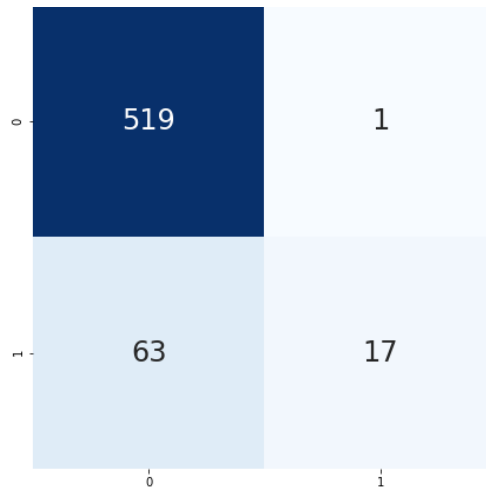
Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Taburcu (0)	%88	%100	%94	520
Yatış (1)	%92	%15	%26	80
Ortalama	%90	%57	%60	600

4.3.3. Doğrusal olmayan destek vektör makineleri algoritması

Yatış-taburcu durumu tahmini yapılırken kullanılan sınıflandırma algoritmalarından biri doğrusal olmayan destek vektör algoritmasıdır. Bu algoritma çekirdek hilesi olarak adlandırılmaktadır. Destek vektör algoritmasıyla model eğitilirken bu çalışmada 3 çeşit çekirdek kullanılmıştır. Bu çekirdekler: Radyal tabanlı çekirdek yani RBF, poly çekirdek ve sigmoid çekirdek. Üç çekirdeğin hepsiyle sınıflandırma çalışması yapılmıştır. Üç ayrı tahmin sonucu elde edilmiştir.

4.3.3.1. Doğrusal olmayan destek vektör algoritması-RBF

Radyal tabanlı RBF çekirdeği kullanılarak destek vektör makinesi ile model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.16'daki gibidir.



Şekil 4.16. Doğrusal olmayan destek vektör makinesi-RBF karışıklık matrisi.

RBF çekirdeği ile yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.20'deki gibidir.

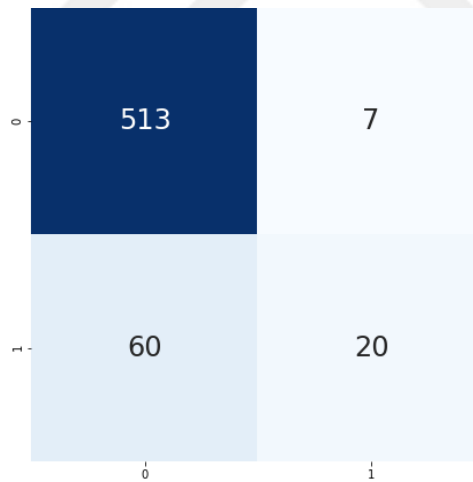
Tablo 4.20. SVM-RBF sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Taburcu (0)	%89	%100	%94	520
Yatış (1)	%94	%21	%35	80
Ortalama	%92	%61	%64	600

Doğrusal olmayan destek vektör algoritması RBF çekirdeğiyle yapılan tahmin çalışmasında modelin; doğruluk oranı %89.3, ortalama kesinlik oranı %92, ortalama duyarlılık oranı %61 ve ortalama F1 skor oranı %64'tür.

4.3.3.2. Doğrusal olmayan destek vektör algoritması-poly

Poly çekirdeği kullanılarak destek vektör makinesi ile model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.17'teki gibidir.



Şekil 4.17. Doğrusal olmayan destek vektör makinesi-poly karışıklık matrisi.

Poly çekirdeği ile yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.21'deki aşağıdaki gibidir.

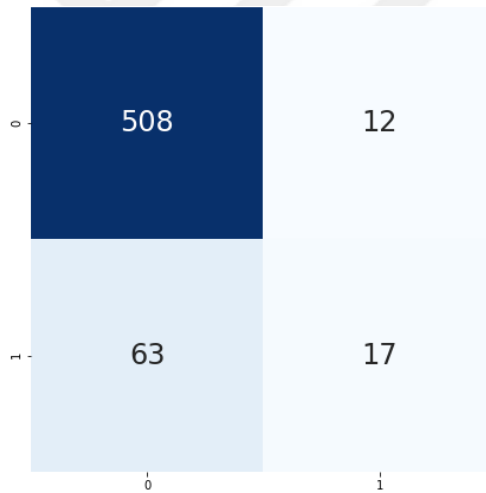
Doğrusal olmayan destek vektör algoritması poly çekirdeğiyle yapılan tahmin çalışmasında modelin; doğruluk oranı %88.8, ortalama kesinlik oranı %82, ortalama duyarlılık oranı %62 ve ortalama F1 skor oranı %66'dır.

Tablo 4.21. SVM -Poly sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Taburcu (0)	%90	%99	%94	520
Yatış (1)	%74	%25	%37	80
Ortalama	%82	%62	%66	600

4.3.3.3. Doğrusal olmayan destek vektör algoritması-sigmoid

Sigmoid çekirdeği kullanılarak destek vektör makinesi ile model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.18’teki gibidir.

**Şekil 4.18.** Doğrusal olmayan destek vektör makinesi-sigmoid karışıklık matrisi.

Doğrusal olmayan destek vektör algoritması sigmoid çekirdeğiyle yapılan tahmin çalışmasında modelin; doğruluk oranı %87,5, ortalama kesinlik oranı %74, ortalama duyarlılık oranı %59 ve ortalama F1 skor oranı %62’dir.

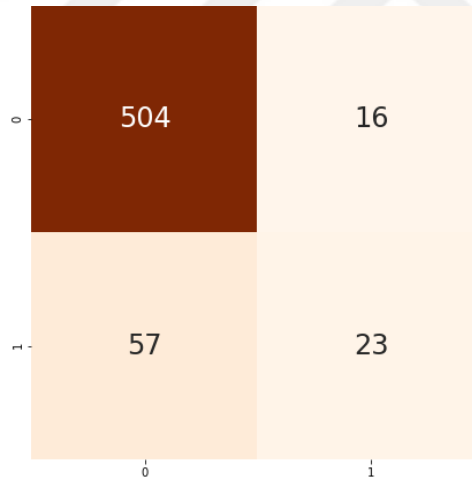
Sigmoid çekirdeği ile yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.22’deki gibidir.

Tablo 4.22. SVM-Sigmoid sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Taburcu (0)	%89	%98	%93	520
Yatış (1)	%59	%21	%31	80
Ortalama	%74	%59	%62	600

4.3.4. Naive bayes algoritması

Yatış-taburcu durumu tahmini yapılırken kullanılan sınıflandırma algoritmalarından biri naive bayes algoritmasıdır. Naive bayes algoritması ile model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.19’teki gibidir.

**Şekil 4.19.** Naive bayes algoritması karışıklık matrisi.

Naive bayes algoritmasıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %87.8, ortalama kesinlik oranı %74, ortalama duyarlılık oranı %63 ve ortalama F1 skor oranı %66’dır.

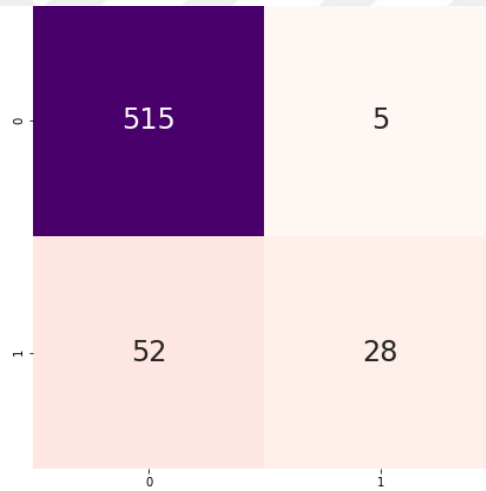
Naive bayes algoritmasıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.23’teki gibidir.

Tablo 4.23. Naive bayes algoritması sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Taburcu (0)	%90	%97	%93	520
Yatış (1)	%59	%29	%39	80
Ortalama	%74	%63	%66	600

4.3.5. Lojistik regresyon algoritması

Yatış-taburcu durumu tahmini yapılırken kullanılan sınıflandırma algoritmalarından biri lojistik regresyon algoritmasıdır. Lojistik regresyon algoritması ile model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.20'deki gibidir.

**Şekil 4.20.** Lojistik regresyon algoritması karışıklık matrisi.

Lojistik regresyon algoritmasıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %90,5, ortalama kesinlik oranı %88, ortalama duyarlılık oranı %67 ve ortalama F1 skor oranı %72'dir.

Lojistik regresyon algoritmasıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.24'teki gibidir.

Tablo 4.24. Lojistik regresyon algoritması sınıflandırma raporu.

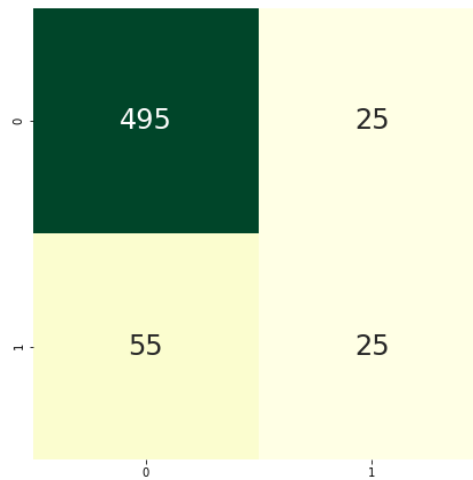
Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Taburcu (0)	%91	%99	%95	520
Yatış (1)	%85	%35	%50	80
Ortalama	%88	%67	%72	600

4.3.6. K-En yakın komşu algoritması

Yatış-taburcu durumu tahmini yapılırken kullanılan sınıflandırma algoritmalarından biri k-en yakın komşu algoritmasıdır. K-En yakın komşu algoritmasıyla model eğitilirken bu çalışmada 2 ayrı uzaklık ölçütü kullanılarak 2 ayrı sınıflandırma modeli geliştirilmiştir. Bu uzaklıklar: Manhattan ve Minkowski uzaklıklarıdır.

4.3.6.1. K-En yakın komşu algoritması-manhattan

K-En yakın komşu algoritmasında manhattan uzaklığı kullanılarak model eğitilmeden önce en iyi doğruluk oranı veren k-komşu değerinin tespiti için algoritma 1-50 arasında k değeriyle ayrı ayrı çalıştırılmıştır. En iyi doğruluk oranını veren ilk k değeri ve manhattan uzaklığı ile model eğitilmiş ve tahmin çalışması yapılmıştır. Modelde manhattan uzaklığı ve k-komşu sayısı=3 olarak alınmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.21'deki gibidir.

**Şekil 4.21.** KNN-Manhattan algoritması karışıklık matrisi.

K-En yakın komşu algoritması manhattan uzaklığıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.25'teki gibidir.

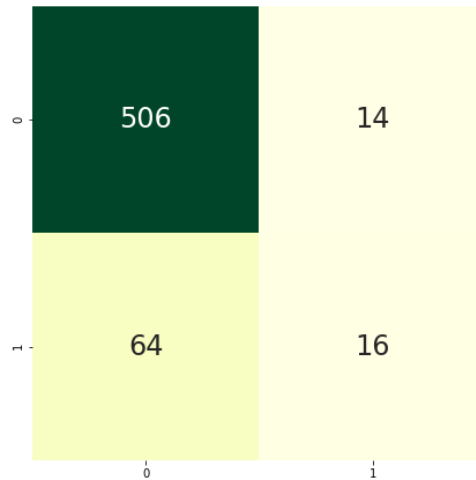
Tablo 4.25. KNN-Manhattan algoritması sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Taburcu (0)	%90	%95	%93	520
Yatış (1)	%50	%31	%38	80
Ortalama	%70	%63	%65	600

K-En yakın komşu algoritması manhattan uzaklığıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %86,6, ortalama kesinlik oranı %70, ortalama duyarlılık oranı %63 ve ortalama F1 skor oranı %65'tir.

4.3.6.2. K-En yakın komşu algoritması-minkowski

K-En yakın komşu algoritmasında minkowski uzaklığı ile model eğitilmeden önce en iyi doğruluk oranı veren k-komşu değerinin tespiti için algoritma 50 tane (1-50) k değeriyle ayrı ayrı çalıştırılmıştır. En iyi doğruluk oranını veren ilk k değeri ve minkowski uzaklığı ile model eğitilmiş, tahmin çalışması yapılmıştır. Modelde k-komşu sayısı=5 olarak alınmıştır. Karışıklık matrisi Şekil 4.22'deki gibidir.



Şekil 4.22. KNN-Minkowski algoritması karışıklık matrisi.

K-En yakın komşu algoritması minkowski uzaklığıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.26'daki gibidir.

Tablo 4.26. KNN-Minkowski algoritması sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Taburcu (0)	%89	%97	%93	520
Yatış (1)	%53	%20	%29	80
Ortalama	%71	%59	%61	600

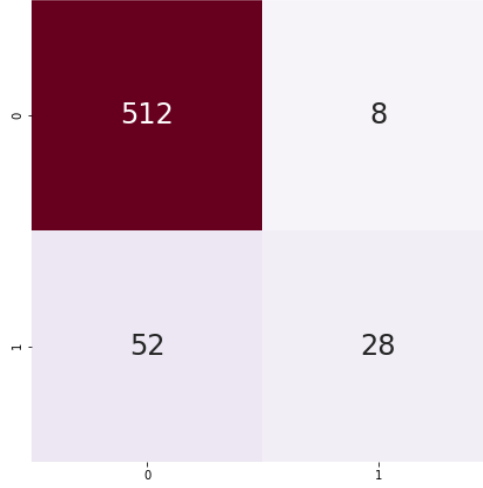
K-En yakın komşu algoritması minkowski uzaklığıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %87, ortalama kesinlik oranı %71, ortalama duyarlılık oranı %59 ve ortalama F1 skor oranı %61'dir.

4.3.7. Rassal orman algoritması

Yatış-taburcu durumu tahmini yapılırken kullanılan sınıflandırma algoritmalarından biri rassal orman algoritmasıdır. Rassal orman algoritmasıyla beraber iki ayrı kriter kullanılmıştır. Bu kriterler “gini” ve “entropi” kriterleridir. Rassal orman algoritmasında kriter olarak “gini” ve “entropi” kullanılarak iki ayrı sınıflandırma modeli oluşturulmuştur. İki ayrı tahmin sonucu elde edilmiştir.

4.3.7.1. Rassal orman algoritması-entropi

Rassal orman – Entropi algoritması kullanılarak model eğitilmiş ve tahmin çalışması yapılmıştır. Yapılan tahmin çalışması sonucunda karışıklık matrisi elde edilmiştir. Elde edilen karışıklık matrisi Şekil 4.23'teki gibidir.



Şekil 4.23. Rassal orman algoritması-entropi karışıklık matrisi.

Rassal orman – Entropi algoritmasıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.27’deki gibidir.

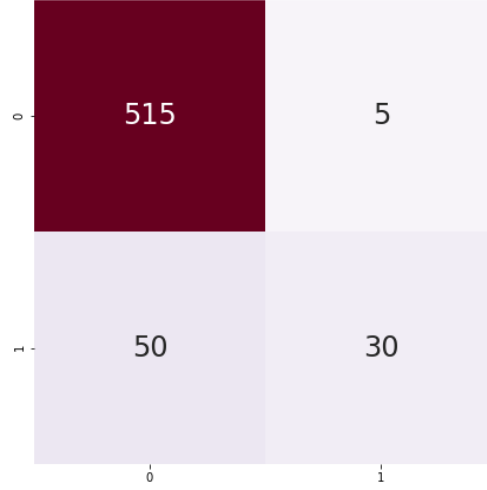
Tablo 4.27. Rassal orman-entropi algoritması sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Taburcu (0)	%91	%98	%94	520
Yatış (1)	%78	%35	%48	80
Ortalama	%84	%67	%71	600

Rassal orman – Entropi algoritmasıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %90,8, ortalama kesinlik oranı %84, ortalama duyarlılık oranı %67 ve ortalama F1 skor oranı %71’dir.

4.3.7.2. Rassal orman algoritması-gini

Rassal orman – Gini algoritması kullanılarak model eğitilmiş ve tahmin çalışması yapılmıştır. Sınıflandırma işlemi sonucunda oluşturulan karışıklık matrisi Şekil 4.24’teki gibidir.



Şekil 4.24. Rassal orman algoritması-gini karışıklık matrisi.

Rassal orman – Gini algoritmasıyla yapılan sınıflandırmada performans raporuna bakıldığında modelin sınıf bazında performans ölçütleri Tablo 4.28’deki gibidir.

Tablo 4.28. Rassal orman-gini algoritması sınıflandırma raporu.

Sınıf	Kesinlik (Precision)	Duyarlılık (Recall)	F1 Skoru	Veri adedi
Taburcu (0)	%91	%99	%95	520
Yatış (1)	%86	%38	%52	80
Ortalama	%88	%68	%74	600

Rassal orman – Gini algoritmasıyla yapılan tahmin çalışmasında modelin; doğruluk oranı %90, ortalama kesinlik oranı %88, ortalama duyarlılık oranı %68 ve ortalama F1 skor oranı %74’tür.

4.3.8. AUCROC değer hesabı

ROC eğrisi makine öğrenmesi modelinin performansının değerlendirilmesi için kullanılan bir ölçüttür. ROC eğrisi, eğrinin altında kalan alana göre yorumlanmaktadır. Eğrinin altında kalan AUC değerini vermektedir. Eğrinin altında kalan alan büyükse yani yüksek bir ROC varsa modelin başarısı yüksek, eğrinin altında daha küçükse modelin başarısı düşük olarak yorumlanmaktadır. Makine öğrenmesi sınıflandırma algoritmalarıyla elde edilen modellerin performanslarının AUC değerleri Tablo 4.29’da özetlenmiştir.

Tablo 4.29. Yatış-taburcu durumu tahmin modeli AUROC değerleri.

Algoritma	AUCROC değeri
Rassal orman algoritması-Gini	68,26%
Karar ağacı-Gini algoritması	67,01%
Karar ağacı-Entropi algoritması	67,01%
Lojistik regresyon algoritması	67,01%
Rassal orman algoritması-Entropi	66,73%
K-En küçük komşu algoritması-Manhattan	63,22%
Naive Bayes algoritması	62,83%
Doğrusal olmayan destek vektör algoritması-Poly	61,82%
Doğrusal olmayan destek vektör algoritması-RBF	60,50%
Doğrusal olmayan destek vektör algoritması-Sigmoid	59,47%
K-En küçük komşu algoritması-Minkowski	58,65%
Doğrusal destek vektör algoritması	57,40%

Tablo 4.29'a bakıldığında en yüksek AUCROC değeri %68,26 elde edilen algoritmanın rassal orman-gini algoritması olduğu görülmektedir.

4.3.9. Algoritma performansı değerlendirme

Yatış-taburcu durumu sınıflandırmasıyla ilgili veri seti %80'i algoritmaları eğitmek, %20'si ise elde edilen modeli test etmek için kullanılmıştır. 12 farklı algoritma ile çalışılmış olup, 12 farklı model elde edilmiştir. Modeller test verileri ile test edilmiş ve karışıklık matrisleri oluşturulmuştur. Bu karışıklık matrislerinden yola çıkılarak hesaplanan performans ölçütleri Tablo 4.30'daki gibidir.

Tablo 4.30. Yatış-taburcu durumu tahmin çalışması genel sınıflandırma raporları.

Algoritma	Kesinlik değeri	Duyarlılık değeri	F Skor	Doğruluk değeri	AUCROC değeri
Rassal orman algoritması-Gini	88,0%	68,0%	74,0%	90,8%	68,26%
Lojistik regresyon algoritması	88,0%	67,0%	72,0%	90,5%	67,01%
Karar ağacı-Entropi algoritması	88,0%	67,0%	72,0%	90,5%	67,01%
Karar ağacı-Gini algoritması	88,0%	67,0%	72,0%	90,5%	67,01%
Rassal orman algoritması- Entropi	84,0%	67,0%	71,0%	90,0%	66,73%
K-En küçük komşu algoritması-Manhattan	70,0%	63,0%	65,0%	87,0%	63,22%
Naive Bayes algoritması	74,0%	63,0%	66,0%	87,8%	62,83%
Doğrusal olmayan destek vektör algoritması-Poly	82,0%	62,0%	66,0%	88,8%	61,82%
Doğrusal olmayan destek vektör algoritması-RBF	92,0%	61,0%	64,0%	89,3%	60,50%
Doğrusal olmayan destek vektör algoritması-Sigmoid	74,0%	59,0%	62,0%	87,5%	59,47%
K-En küçük komşu algoritması-Minkowski	71,0%	59,0%	61,0%	86,6%	58,65%
Doğrusal destek vektör algoritması	90,0%	57,0%	60,0%	88,5%	57,40%

Tablo 4.30’a göre;

- Modellerin test verileriyle test edilmesi sonucunda en yüksek doğruluk oranına sahip algoritma rassal orman-gini algoritmasıdır. Elde edilen doğruluk oranı %90,8’dir.

- Ortalama kesinlik değerlerine bakıldığında; en yüksek kesinlik oranına sahip algoritma radyal temelli doğrusal olmayan destek vektör algoritmasıdır. Elde edilen kesinlik oranı %92'dir.
- Ortalama duyarlılık değerlerine bakıldığında; en yüksek duyarlılık oranına sahip algoritma rassal orman-gini algoritmasıdır. Gini kriteriyle elde edilen modelde %68 ile en yüksek duyarlılık değeri elde edilmiştir.
- Ortalama F skoruna bakıldığında; en yüksek F skoruna sahip algoritma ise rassal orman-gini algoritmasıdır. Elde edilen F skoru %74'tür.
- AUCROC değerlerine bakıldığında en yüksek AUCROC %90,8 değerine sahip algoritma rassal orman-gini algoritmasıdır.

4.3.10. Yatış-taburcu durum tahmini için çapraz doğrulama

Tablo 4.30'da sonuçlar veri setinin %80'i eğitim için %20'si test için kullanılarak bir kere eğitim ve test sonucu elde edilen sonuçlardır. Yatış-taburcu veri setiyle çapraz doğrulama yapılması için veri seti rassal parçalara bölünür ve bölünen her bir parçanın hem eğitim setinde hem de test setinde yer alması sağlanmıştır. Böylelikle verinin, veri setindeki dağılımından kaynaklanan sapma ve hataların en aza indirilmesi hedeflenmiştir.

Çalışmada, çapraz doğrulama yapılırken verinin sırayla değil rassal olarak bölünmesi ve işlemin 10 kere tekrar etmesi sağlanmıştır. Modeller, farklı veri versiyonlarıyla 10 kere eğitilmiş ve test edilmiştir. Her adımda elde edilen doğruluk oranları hesaplanmıştır. Ardından çapraz doğrulama doğruluk oranı, her adımda bulunan doğruluk oranlarının ortalaması hesaplanarak elde edilmiştir. Bu çalışmayla elde edilen doğruluk değeri sonuçları Tablo 4.31'deki gibidir.

Tablo 4.31. Yatış-taburcu için çapraz doğrulama doğruluk oranları (%).

Algoritma	1	2	3	4	5	6	7	8	9	10	Ortalama doğruluk
RF-Gini	88,5	88,3	89,9	89,0	88,3	91,2	88,3	87,7	89,5	90,5	89,1
LOGR	88,0	90,5	87,5	89,8	87,2	89,2	88,8	90,00	87,8	86,8	88,6
RF-Entropi	88,5	90,2	87,3	89,8	85,5	88,8	88,8	88,2	88,7	89,7	88,6
SVM-Poly	87,8	88,8	88,2	88,8	88,7	88,2	87,5	89,2	88,5	89,2	88,5

Tablo 4.31. (Devamı): Yatış-taburcu için çapraz doğrulama doğruluk oranları (%).

Algoritma	1	2	3	4	5	6	7	8	9	10	Ortalama doğruluk
DSVM	88,8	85,3	88,7	87,5	89,7	87,7	89,2	90,2	84,5	87,7	87,9
SVM-RBF	87,0	86,3	87,7	87,2	88,2	88,8	88,5	89,0	87,5	86,7	87,7
NB	85,7	86,5	86,7	87,2	85,8	88,7	87,3	86,7	88,2	85,8	86,9
KNN- Manhattan	87,0	85,0	87,7	85,7	89,0	86,2	85,8	87,7	84,3	87,3	86,6
SVM- Sigmoid	86,1	84,8	85,0	86,8	84,7	84,0	84,8	86,3	85,8	86,8	85,5
KNN - Minkowski	84,3	86,0	85,8	84,3	84,8	86,5	83,2	85,3	85,2	85,8	85,1
KA-Gini	81,8	84,0	82,5	81,0	82,2	80,2	84,0	83,8	81,2	82,7	82,3
KA- Entropi	81,2	81,8	81,2	83,1	82,5	82,5	82,3	82,8	82,0	82,8	82,2

Tablo 4.31'e bakıldığında aynı veri setinin farklı parçalarıyla 10 kere eğitilme ve 10 kere test edilmesi sonucunda elde edilen doğruluk oranları görülmektedir. Bu doğruluk oranlarının ortalamalarına bakıldığında en yüksek doğruluk oranını rassal orman-gini algoritmasıyla elde edilmiştir. İkinci olarak ortalama en yüksek doğruluk oranı lojistik regresyon algoritmasıyla elde edilmiştir. Çapraz doğrulama çalışmasında her parçanın eğitim süresi, test süresi, kesinlik, duyarlılık değerleri ve F skorları hesaplanmıştır. 10 parça için hesaplanan bu verilerin ortalaması aşağıda Tablo 4.32'de verilmiştir.

Tablo 4.32. Yatış-taburcu için çapraz doğrulama sonuçları.

Algoritma	Ortalama eğitim süresi	Ortalama test süresi	Ortalama kesinlik	Ortalama duyarlılık	Ortalama F skor
RF-Gini	0,1117	0,0137	84,37%	65,58%	69,87%
LOGR	0,0905	0,0096	84,75%	66,63%	71,08%

Tablo 4.32. (Devamı): Yatış-taburcu için çapraz doğrulama sonuçları.

Algoritma	Ortalama eğitim süresi	Ortalama test süresi	Ortalama kesinlik	Ortalama duyarlılık	Ortalama F skor
RF-Entropi	0,1168	0,0145	86,54%	64,43%	68,82%
SVM-Poly	0,1026	0,0234	82,40%	62,70%	66,42%
DSVM	0,0970	0,0198	89,51%	59,58%	62,66%
SVM-RBF	0,1223	0,0603	86,71%	60,58%	63,95%
NB	0,0048	0,0083	75,37%	63,52%	66,64%
KNN-Manhattan	0,0028	0,1114	76,91%	55,65%	56,76%
SVM-Sigmoid	0,0994	0,0269	70,00%	57,40%	58,92%
KNN -Minkowski	0,0031	0,0589	73,23%	54,71%	54,98%
KA-Gini	0,0232	0,0081	65,28%	65,92%	65,48%
KA-Entropi	0,0208	0,0077	65,47%	66,04%	65,73%

Tablo 4.32’de;

- En yüksek ortalama kesinlik değeri doğrusal destek vektör makinesi algoritmasıyla, en yüksek duyarlılık değeri ve en yüksek F skoru ise lojistik regresyon algoritmasıyla elde edilmiştir.
- Ortalama eğitim süresi en kısa olan k-en yakın komşu-manhattan algoritmasıdır. Ortalama test süresi en kısa olan karar ağacı-entropi algoritmasıdır.

Çapraz doğrulama yapılmadan, sadece veri setinin %80 eğitim, %20 test verisi olarak kullanılarak elde edilen sonuçlarda en yüksek doğruluk oranı rassal orman-gini algoritması ile %90,8 oranla elde edilirken; çapraz doğrulama yapıldığında en yüksek doğruluk oranı yine rassal orman-gini algoritmasıyla ancak %89,12 oranıyla elde edilmiştir. Bu sebeple çapraz doğrulama yapmak sınıflandırma modellerinde büyük öneme sahiptir.

4.4. Tahmin Çalışmalarının Sınıflandırma Performanslarının Kıyaslanması

Triyaj sınıflandırması ve yatış-taburcu durumu için ayrı ayrı 12 algoritma ile sınıflandırma modeli oluşturulmuş ve tahmin çalışması yapılmıştır. Aşağıda Tablo 4.33'te iki veri seti için hesaplanan ROCAUC değerleri bulunmaktadır.

Tablo 4.33. İki tahmin çalışmasının AUCROC değerleri.

Algoritma	Yatış-Taburcu AUCROC değeri	Triyaj Alan AUCROC değeri
Karar ağacı-Entropi algoritması	67,01%	96,79%
Rassal orman algoritması-Entropi	66,73%	96,16%
Rassal orman algoritması-Gini	68,26%	96,13%
Karar ağacı-Gini algoritması	67,01%	94,53%
Doğrusal olmayan destek vektör algoritması- Poly	61,82%	85,59%
Lojistik regresyon algoritması	67,01%	85,22%
Doğrusal destek vektör algoritması	57,40%	84,96%
K-En küçük komşu algoritması-Minkowski	58,65%	83,73%
K-En küçük komşu algoritması-Manhattan	63,22%	83,72%
Doğrusal olmayan destek vektör algoritması- RBF	60,50%	82,39%
Naive Bayes algoritması	62,83%	80,48%
Doğrusal olmayan destek vektör algoritması- Sigmoid	59,47%	77,30%

Tablo 4.33'e göre; yatış-taburcu durumu için en yüksek AUCROC değeri %68,26 rassal orman-gini algoritması ile elde edilirken, triyaj alan sınıflandırması için en yüksek AUCROC değeri %96,79 karar ağacı-entropi algoritmasıyla elde edilmiştir.

Aşağıda Tablo 4.34’te iki veri seti için çaprazlama sonucunda elde edilen ortalama eğitim ve test süreleri yer almaktadır.

Tablo 4.34. Çalışmaların sınıflandırma raporlarının kıyaslanması.

Algoritma	Yatış-taburcu ortalama eğitim süresi	Triyaj alan ortalama eğitim süresi	Yatış-taburcu ortalama test süresi	Triyaj alan ortalama test süresi
KA-Entropi	0,0208	0,0218	0,0077	0,0066
KA-Gini	0,0232	0,0201	0,0081	0,0067
DSVM	0,0970	0,1361	0,0198	0,0276
SVM-RBF	0,1223	0,2082	0,0603	0,0995
SVM-Poly	0,1026	0,1550	0,0234	0,0349
SVM- Sigmoid	0,0994	0,2566	0,0269	0,0573
NB	0,0048	0,0073	0,0083	0,0089
LOGR	0,0905	0,2415	0,0096	0,0109
KNN - Minkowski	0,0031	0,0032	0,0589	0,0647
KNN- Manhattan	0,0028	0,0037	0,1114	0,1357
RF-Gini	0,1117	0,1168	0,0137	0,0127
RF-Entropi	0,1168	0,1302	0,0145	0,0133

Tablo 4.34’e göre; yatış-taburcu durumu için en kısa eğitim süresi k-en yakın komşu-manhattan algoritması ile elde edilirken, triyaj alan sınıflandırması için en kısa eğitim süresi k-en yakın komşu -minkowski algoritmasıyla elde edilmiştir. Aynı zamanda, yatış-taburcu sınıflandırması ve triyaj alan sınıflandırması için en kısa test süresi karar ağacı-entropi algoritmasıyla elde edilmiştir.

Aşağıda Tablo 4.35’te iki veri seti için çaprazlama sonucunda elde edilen ortalama kesinlik, ortalama duyarlılık ve ortalama F skoru yer almaktadır.

Tablo 4.35. Çalışmaların çapraz doğrulama sonuçları (%).

Algoritma	Yatış- taburcu ortalama kesinlik	Triyaj alan ortalama kesinlik	Yatış- taburcu ortalama duyarlılık	Triyaj alan ortalama duyarlılık	Yatış- taburcu ortalama F skor	Triyaj alan ortalama F skor
KA- Entropi	65,47	97,52	66,04	97,01	65,73	97,25
KA-Gini	65,28	96,42	65,92	96,34	65,48	96,36
DSVM	89,51	87,21	59,58	80,38	62,66	82,26
SVM-RBF	86,71	85,40	60,58	77,03	63,95	78,83
SVM-Poly	82,40	85,34	62,70	79,91	66,42	81,11
SVM- Sigmoid	70,00	76,27	57,40	70,78	58,92	72,28
NB	75,37	78,11	63,52	71,22	66,64	72,91
LOGR	84,75	86,66	66,63	80,04	71,08	82,08
KNN - Minkowski	73,23	79,56	54,71	75,27	54,98	76,39
KNN- Manhattan	76,91	81,00	55,65	75,35	56,76	76,73
RF-Gini	84,37	95,25	65,58	93,76	69,87	94,45
RF-Entropi	86,54	95,56	64,43	94,30	68,82	94,87

Tablo 4.35’e göre;

- Yatış-taburcu durumu için en yüksek ortalama kesinlik değeri %89,51 doğrusal destek vektör algoritması ile elde edilirken, triyaj alan sınıflandırması için en yüksek ortalama kesinlik değeri %97,52 karar ağacı-entropi algoritmasıyla elde edilmiştir.

- Yatış-taburcu durumu için en yüksek ortalama duyarlılık değeri %66,63 lojistik regresyon algoritması ile elde edilirken, triyaj alan sınıflandırması için en yüksek ortalama duyarlılık değeri %97,01 karar ağacı-entropi algoritmasıyla elde edilmiştir.
- Yatış-taburcu durumu için en yüksek ortalama F skoru %71,08 lojistik regresyon algoritması ile elde edilirken, triyaj alan sınıflandırması için en yüksek ortalama F skoru %97,25 karar ağacı-entropi algoritmasıyla elde edilmiştir.

Tablo 4.36’da iki veri seti için çapraz doğrulama sonucunda elde edilen ortalama doğruluk oranları yer almaktadır. Yatış-taburcu durumu için en yüksek ortalama doğruluk değeri %89,12 rassal orman-gini, triyaj alan sınıflandırması için en yüksek ortalama doğruluk değeri %97,05 karar ağacı-gini algoritmasıyla elde edilmiştir.

Tablo 4.36. Çalışmaların çapraz doğrulama doğruluk değerleri.

Algoritma	Yatış-taburcu ortalama doğruluk değeri	Trijaj alan ortalama doğruluk değeri
KA-Entropi	82,23%	96,73%
KA-Gini	82,33%	97,05%
DSVM	87,87%	83,60%
SVM-RBF	87,68%	80,03%
SVM-Poly	88,47%	82,72%
SVM-Sigmoid	85,53%	73,20%
NB	86,85%	75,68%
LOGR	88,57%	83,10%
KNN -Minkowski	85,13%	79,22%
KNN-Manhattan	86,57%	79,48%
RF-Gini	89,12%	94,30%
RF-Entropi	88,55%	94,75%

5. SONUÇ VE ÖNERİLER

Bu çalışmada, makine öğrenmesi sınıflandırma algoritmalarıyla sağlık sistemi üzerinde iki ayrı konuda tahmin çalışması yapılmıştır. Çalışmada, Kartal Lütfü Kırdar Eğitim Araştırma Hastanesi Acil Tıp biriminden toplanan gerçek triyaj verileriyle çalışılmıştır. Veri toplama sürecine Sakarya Üniversitesi etik kurulundan etik kurulu izni ve Kartal Lütfü Kırdar Eğitim Araştırma Hastanesi'nden baş hekim onayı alınarak başlanmıştır. Veriler; yaklaşık 9 ay boyunca haftalık hastane ziyaretleri boyunca, çalışmaya destek olan uzman doktor desteğiyle, manuel olarak toplanmıştır. Toplanan veriler; cinsiyet, yaş, yaşamsal bulgular, hastaların sözlü şikayetleri, hastaların belirttiği ya da sistemde önceden kayıt altına alınmış özgeçmiş hastalıklar, acil servise varış şekli, triyaj sonucu olarak triyaj alan sınıflandırması ve acil sağlık hizmeti sonucu olarak yatış-taburcu olma durumu gibi toplam 56 adet girdi, 2 adet çıktı olmak üzere 58 çeşit veri toplanmıştır. Çalışmaya 3.000 hastanın verisi dahil edilmiştir. Veri toplama aşamasında eksik veri olmamasına özen gösterilmiştir.

Çalışmada amaç; gerçek hayattan alınmış gerçek veriler ile makine öğrenmesi sınıflandırma algoritmalarını eğitip, triyaj alan tahmin modelini ve yatış-taburcu tahmin modelini oluşturup, bu modelleri test verileriyle test ettikten sonra elde edilen tahmin sonuçları göz önüne alınarak kullanılan sınıflandırma algoritmalarının performansını kıyaslamaktır. Bununla birlikte elde edilen makine öğrenmesi modelleri ile gerçek hayattaki problemlere çözüm ışığı olabilmektir.

Çalışmada kullanılan veri setinin 2 adet çıktı sütunu bulunmaktadır. Bunlardan biri triyaj alan sınıfı, diğeri ise hastaneye yatış yapma ya da taburcu olma durumudur. İki çalışmada da kullanılan girdiler aynıdır. Öncelikli olarak çalışmada veri setinin istatistiksel ve grafiksel analizi yapılmıştır. Asıl çalışma olarak, iki çalışma için de aynı makine öğrenmesi sınıflandırma algoritmaları kullanılmıştır. Veri setinin %80'i eğitim veri seti ve %20'si test veri seti olacak şekilde ikiye ayrılmıştır. Sınıflandırma algoritmaları ve alt kriterleri olmak üzere 12 sınıflandırma algoritması, eğitim veri seti ile eğitilerek modeller elde edilmiştir. Bu modeller, test veri seti ile test edilmiştir. Modellerin performanslarının hesaplanabilmesi için her model için karışıklık matrisleri ve sınıflandırma raporları elde edilmiştir. Daha sonrasında her model için

AUCROC deęerleri hesaplanmış ve doęrulama sırasında verilerin rassal olarak bölünmesi saęlanarak monte carlo apraz doęrulama iřlemi yapılmıřtır. En son olarak; triyaj alan tahmini ve yatıř-taburcu tahmini iin ayrı ayrı yrtlen bu iřlemler sonucunda elde edilen makine ęrenmesi sınıflandırma algoritmalarının performansları kıyaslanmıřtır.

Bu alıřmada, saęlık sisteminin bir parası olan triyaj srelerinde sınıflandırma algoritmalarının performansları incelenerek, hangi algorithmadan elde edilen modellerin triyaj srelerinde daha yksek bařarı gstereceęi belirlenmeye alıřılmıřtır. Modellerin tahmin performansları incelendięinde; apraz doęrulama iřlemleri sonucunda triyaj alan sınıflandırması iřleminde karar aęacı algoritmasında gini kriteri kullanılarak elde edilen modelin %97,05 oranıyla doęru tahmin ettięi grlmektedir. Yani bu modelin acil servise gelecek 100 kiřiden yaklařık 97 kiřiyi doęru sınıflandırması ngrlmektedir. apraz doęrulama iřlemleri sonucunda yatıř-taburcu durumu tahmininde ise rassal orman algoritmasında gini kriteri kullanılarak elde edilen modelin %89,12 oranıyla doęru tahmin ettięi grlmektedir.

Trijaj alan tahmini alıřmasında apraz doęrulama iřlemi yapılmadan nce; rassal orman-entropi ve rassal orman-entropi algoritmalarıyla en yksek doęruluk oranı elde edilirken, apraz doęrulama iřlemiyle en yksek doęruluk oranı karar aęacı-gini algoritmasıyla elde edilmiřtir. Aynı zamanda triyaj alan tahmin alıřmasının genelinde rassal orman-entropi ve rassal orman-gini algoritmaları arasında performans anlamında bir fark bulunmamaktadır. Modelin en yksek kesinlik deęeri ve en yksek F skoru deęeri, rassal orman-gini ve rassal orman-entropi algoritmalarıyla elde edilmiřtir. Bařarı oranları aynıdır. Karar aęacı-entropi algoritması; en yksek AUCROC ve duyarlılık deęerinde ve apraz doęrulama iřleminin ortalama kesinlik, ortalama duyarlılık, ortalama F skor, ortalama test sresi gibi deęerlerinde en yksek bařarıya sahip algoritma olmuřtur. Aynı zamanda karar aęacı-entropi algoritması, yatıř-taburcu tahmin alıřmasının apraz doęrulama test srelerinde de en yksek bařarıya, en dřk test sresine sahip algoritma olmuřtur. İki tahmin alıřmasının da apraz doęrulama iřlemlerinin eęitim srelerinde en yksek bařarıya, en kısa eęitim sresine sahip algoritma ise k-en yakın komřu algoritması olmuřtur. Triyaj alan tahmininde minkowski uzaklıęı ile en kısa eęitim sresi elde edilirken, yatıř-taburcu tahmininde en kısa eęitim sresi manhattan uzaklıęı ile elde edilmiřtir.

Yatış-taburcu durum tahmini çalışmasında çapraz doğrulama işlemi yapılmadan önce; rassal orman-gini algoritması, en yüksek AUCROC değerinde, duyarlılık, F skor, doğruluk değerlerinde ve çapraz doğrulama işleminin ortalama doğruluk değerinde en yüksek başarıya sahip algoritma olmuştur. Triyaj alan tahmini çalışmasında en yüksek kesinlik değerine sahip algoritma rassal orman-gini ve rassal orman-entropi algoritmalarıyken, yatış-taburcu tahmin çalışmasında en yüksek kesinlik değerine sahip algoritma radyal tabanlı doğrusal olmayan destek vektör algoritması olmuştur. Aynı zamanda yatış-taburcu tahmin çalışmasının çapraz doğrulama işleminde en yüksek ortalama kesinlik değeri doğrusal destek vektör algoritmasıyla elde edilmekle birlikte en yüksek ortalama kesinlik değeri ve en yüksek ortalama duyarlılık değeri lojistik regresyon algoritmasıyla elde edilmiştir.

Literatür araştırması kısmında; triyaj alan sınıflandırması, yatış-taburcu olma durumları ile ilgili makine öğrenmesi sınıflandırma algoritmaları kullanılarak yapılan çalışmalardan bahsedilmiştir. Yapılan bu literatür araştırmasında ele alınan çalışmalarda, Türkiye’de sayısal ve sözel veri içeren gerçek triyaj verileri ile makine öğrenmesi sınıflandırma algoritmalarının kullanılmadığı görülmüştür.

Bu çalışmanın literatürdeki diğer çalışmalardan farklı yönlerinden birkaçı; öncelikle bu çalışmada kullanılan veriler gerçek hayattan alınmış verilerdir ve veriler sadece sayısal bilgiler içermemekle birlikte hastaların sözlü olarak iletmış oldukları hastalık şikayetlerini ve özgeçmiş hastalıklarını da içermektedir. Tüm bu veriler özenle araştırılmış ve toplanmıştır. Sonrasında elde edilen bu veriler, acil tıp uzmanı doktor ile gruplandırılmış ve kategorik değişkenlere dönüştürülmüştür.

Hastaların triyaj kağıtlarından ilgili triyaj verilerine ulaşılmış ve triyaj kağıtlarından ulaşılan protokol numaraları ile hastane bilgi yönetimi sistemindeki verilere de erişilip; veriler kontrol edilip, doktor notları, önceki ziyaretlerinde kaydedilmiş özgeçmiş hastalıkları, acil servis hizmet sonuçları gibi bilgilere erişilmiştir. Bu bilgiler sistemde metinsel şekilde bulunmaktadır. Metinsel alanlar tek tek okunarak gerekli bilgilere ulaşılmıştır. Toplanan veriler hastane bilgi yönetim sisteminden ya da elektronik sağlık kayıtlarından direkt rapor olarak erişilecek bilgiler olmamakla birlikte çalışma boyunca veriler manuel olarak toplanmıştır. Tüm bunlar, çalışmanın özgün yönlerinin oluşmasını sağlamaktadır.

Türkiye’de triyaj ile ilgili yapılan çalışmalar incelendiğinde; makine öğrenmesi modelleri ile triyaj alan tahmini ve yatış-taburcu durumu tahmini yapılan çalışma mevcuttur. Ancak çalışmalarda, yurt dışındaki hastanelerin elektronik sağlık kayıtlarından elde edilmiş verilerle yapıldığı görülmektedir. Aynı zamanda çalışmalarda kullanılan algoritma sayısı daha sınırlıdır. Türkiye’deki triyajla ilgili diğer çalışmalara bakıldığında; triyaj süreçlerinde gerçek veri toplanarak çalışmalar yapılmış ancak makine öğrenmesi sınıflandırma algoritmaları kullanılmadığı, sadece istatistiksel analizler yapıldığı ve tıp alanıyla ilgili çalışmalar yapıldığı görülmektedir. Bu çalışmanın; İstanbul gibi kalabalık bir şehirde bulunan, hasta profili çeşitliliği fazla olan bir hastaneden gerçek sayısal ve sözel veriler toplanarak yapılması, veri setinin 56 çeşit girdi ve 3.000 hastanın verisini içermesi, 12 çeşit makine öğrenmesi sınıflandırma modelinin kullanılması açısından Türkiye’de bir ilk olmaya aday olduğu söylenebilir. Literatür araştırması sırasında buna benzer bir çalışmaya rastlanılmamıştır.

Acil servisler; hasta yoğunluğunun çok olduğu, hastalar için saniyelerin bile önemli olduğu, hastalara doğru zamanda ve doğru yerde hizmet verilmesi gerektiği bölümlerdir. Bu nedenle acil servislerde triyaj sınıflandırması büyük öneme sahiptir.

Geleceğe ait çalışmalarda;

- Hastane bilgi yönetim sisteminin metinsel alanlarından ulaşılabilen bilgilere daha kolay ulaşmayı sağlayacak yazılımsal bir geliştirme sağlanarak verilere ulaşım hızlandırılabilir ya da hastane bilgi yönetim sisteminde triyaja özel bir ekran olursa verilere daha kolay erişebilir.
- Verilere ulaşım manuel olmaktan kurtarılıp, kolaylaştırıldıktan sonra veri setine dahil edilen hasta sayısı artırılarak model geliştirilebilir. Böylece veri seti büyüklüğü artırıldığında triyaj alan sınıflandırmasındaki tahminlerin doğruluk oranının artırılması sağlanabilir.
- Üçlü triyaj üzerinde yapılan triyaj alan sınıflandırma çalışması diğer triyaj sistemleriyle de çalışılabilir ve ürün haline getirilebilir.
- Yatış-taburcu durumunda; taburcu olan hasta veri sayısı artırılarak, modellerin performans değerleri arttırılabilir.
- Yatış-taburcu durum tahmininde; yatış yapan hastaların hangi bölüme yatış yaptığıyla ilgili veri seti örneklem sayısı artırılarak, hastaların hangi bölüme yatış yapacağı tahmin edilebilir.

- Bu çalışma, diğer makine öğrenmesi algoritmalarıyla ve diğer çapraz doğrulama yöntemleri kullanılarak da model performansları değerlendirilebilir.
- Elde edilen modeller veri anlamında geliştirildikten sonra yazılımsal olarak da geliştirilerek ürün haline dönüştürülebilir. Bu fiziksel ürün, acil servislerde triyaj sınıflandırma ve yatış-taburculuk durum tahmini amacıyla kullanılmaya başlanabilir. Acil servise başvuran hastalar bu ürünü triyaj alan sınıflandırması için kullanırken arka planda yeni gelen hastanın yatış mı yapacağı, taburcu mu olacağı tahmin sonucu doktorlara ve ilgililere iletilebilir.
- Bu ürün triyaj sınıflandırması yaparken almış olduğu bilgiler doğrultusunda hastaya gerekli yönlendirmeleri yapabilir. Örnek olarak; göğüs ağrısı şikayeti ile bir hasta geldiğinde triyaj ekranında EKG sırasını otomatik olarak verebilir ve EKG sonucuna göre triyaj sınıflandırması yapabilir.
- Aynı zamanda bu ürün ile hastanın ağzından çıkan cümleler kayıt altına alınabilir hale getirilebilir ve sözel bilgilerin işlenmesi kısmı için de metin madenciliği yöntemleri kullanılabilir. Bu doğrultuda sınıflandırma algoritmaları ve metin madenciliği yöntemleri beraber kullanılarak yeni modeller geliştirilebilir.
- Geliştirilen ürün ve modeller öncelikle pilot hastanelerde uygulanabilir ve sonrasında ulusal ve uluslararası düzeyde yaygınlaştırılabilir.

Bu çalışmada sonucunda elde edilen modellerin geliştirilip hayata geçirildiğinde, yani triyaj alan sınıflandırmasının bir sağlık personeli tarafından değil de makine öğrenmesi sınıflandırma algoritmalarıyla geliştirilmiş modeller aracılığı ile yapılmasıyla; triyaj süreçlerini hızlandıracağı, triyaj sınıflandırma kalitesini artıracığı, acil servis hizmet kalitesini ve verimliliğini artıracığı, triyajı personelin bilgi ve tecrübesine bağlılıktan kurtaracağı, triyaj işlemi için görevlendirilen personel sayısını azaltabileceği öngörülmektedir.

Acil servisler aynı zamanda hastane kaynaklarının en verimli şekilde kullanılması gerektiği, zaman ve kaynak planlamasının çok önemli olduğu bölümlerdir. Bu çalışma sonucunda elde edilen modellerin geliştirilip hayata geçirilmesinin; acil servise başvuran hastanın daha başvuru aşamasındayken yani acil servis hizmeti almadan önce hastanın yatış mı yapacağı yoksa taburcu mu olacağı tahmin sonucu doktorlar tarafından bilinmiş olur ve bu yoğunluğa göre yatak, ilaç, sağlık ekipmanlarının,

ellerindeki tüm kaynakların ve doktor planlanmasının daha doğru ve daha verimli yapılmasını sağlayacağı öngörülmektedir.

Bu çalışmada geçerliliğe yönelik tehditler şu şekilde özetlenebilir;

- Veri seti; hastaların ya da hasta yakınlarının sözlü olarak ilettiği, triyaj kağıtlarında ve hastane bilgi yönetim sisteminde metinsel olarak tutulan bilgileri içermesi sebebiyle veri toplama aşaması zor ve uzun bir süreci kapsamaktadır. Bu nedenle gelecekteki çalışmalarda bu zorluğun ortadan azaltılması daha uygun olacaktır.
- Gerçek hayatta bazı hastalarda tüm girdilerinin karşılığı bulunmayabilmektedir. Bu çalışmaya dahil edilen verilerde, tüm girdilerin karşılığı olmasına özen gösterilmiştir. Ancak gerçek hayatta bu olmayabilir.
- Acil servisler hasta sirkülasyonunun fazla olduğu birimlerdir bu nedenle bu yoğunluk sebebiyle tüm bu veriler düzenli elde edilemeyebilir.
- Çalışmada elde edilen modeller bir ürün haline getirilirse acil servisin yoğunluğu, hastaların hastalık şikayetleri sebebiyle panik hallerinden dolayı bu ürün sağlıklı kullanılamayabilir. Bu nedenle yönlendirici personelin bulunması gerekebilir.
- Kardiyak hastaların büyük çoğunluğu hastanenin komşuluğundaki “Koşuyolu Kalp Hastanesi’ne” gitmekte olduğu için veri seti bu anlamda ayrıntılı veri içermeyebilir. Elde edilen modeller bu ve benzeri hastalarla karşılaştığında verimli bir sınıflandırma yapamayabilir.

KAYNAKLAR

- Akca, M. F. (2023, 5 Aralık). *Makine öğrenmesi serisi-2*. <https://medium.com/deep-learning-turkiye/nedir-bu-destek-vekt%C3%B6r-makineleri-makine-%C3%B6%C4%9Frenmesi-serisi-2-94e576e4223e>
- Almalahy, A. A. (2022), *Samsun kamu hastaneleri acil servislerinde çalışan hemşirelerin triyaj bilgi düzeylerinin değerlendirilmesi*. [Doktora tezi], Ondokuz Mayıs Üniversitesi.
- Amidi, S. (2023, 5 Aralık). *Makine öğrenmesi*. <https://stanford.edu/~shervine/l/tr/teaching/cs-229/cheatsheet-supervised-learning#tree>
- Artificial Intelligence Technology. (2021). *Official Textbooks for Huawei ICT Academy*.
- Atalay, M., & Çelik, E. (2017). Büyük veri analizinde yapay zekâ ve makine öğrenmesi uygulamaları-artificial intelligence and machine learning applications in big data analysis. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(22), 155-172.
- Atcılı, A. (2023, 6 Aralık). *Machine Learning Mesafeleri*. <https://medium.com/machine-learning-t%C3%BCrkiye/machine-learning-mesafeleri-8ac88ca393>
- AWS Amazon (2023, 5 Aralık). *Python nedir?*. <https://aws.amazon.com/tr/what-is/python/>
- Aydin, A. G. (2011). *Acil servis triajında yapay zeka yöntemlerinin güvenilirliği*. [Uzmanlık tezi], Akdeniz Üniversitesi.
- Aytekin, Ç., Sütçü, C. S., & Özfidan, U. (2018). Karar ağacı algoritması ile metin sınıflandırma: *Müşteri yorumları örneği*. *Journal of International Social Research*, 11(55).
- Başer, B. Ö., Yangın, M., & Sarıdaş, E. S. (2021). Makine öğrenmesi teknikleriyle diyabet hastalığının sınıflandırılması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 25(1), 112-120.
- Batal, M. S. (2016). *Yapay zeka uygulamaları ve yapay zekanın geleceği*. Uluslararası Sunhıll Üniversitesi, 1-9.
- Bayrak, E. A., Kırıcı, P., Ensari, T., Seven, E., & Dağtekin, M. (2022). Göğüs Kanseri Verileri Üzerinde Makine Öğrenmesi Yöntemlerinin Uygulanması. *Journal of Intelligent Systems: Theory and Applications*, 5(1), 35-41.
- Bilgin, G. (2021). Makine Öğrenmesi Algoritmaları Kullanarak Erken Dönemde Diyabet Hastalığı Riskinin Araştırılması. *Zeki Sistemler Teori ve Uygulamaları Dergisi*, 4 (1), 55-64.
- Bonaccorso, G. (2017). *Machine learning algorithms*. Packt Publishing Ltd.

- Bostancıoğlu, A. A. (2021), *Göğüs ağrısı şikayeti ile acil servise başvuran hasta yönetiminde manchester triyaj sistemi ve modifiye heart(hear) skorunun etkinliği* [Uzmanlık tezi], Sakarya Üniversitesi.
- Chakravarthy, S. S., Bharanidharan, N., & Rajaguru, H. (2023). Deep Learning-Based Metaheuristic Weighted K-Nearest Neighbor Algorithm for the Severity Classification of Breast Cancer. *IRBM*, 44(3), 100749.
- Chen, C. H., Hsieh, J. G., Cheng, S. L., Lin, Y. L., Lin, P. H., & Jeng, J. H. (2020). Emergency department disposition prediction using a deep neural network with integrated clinical narratives and structured data. *International Journal of Medical Informatics*, 139, 104146.
- Christ, M., Grossmann, F., Winter, D., Bingisser, R., & Platz, E. (2009). Modern Triage in the Emergency. *Emergency*, (2249).
- Cihan, Ş. (2018). *Koroner arter hastalığı riskinin makine öğrenmesi ile analiz edilmesi* [Yüksek Lisans Tezi]. Kırıkkale Üniversitesi.
- Coşar, M., & Deniz, E. (2021). Makine Öğrenimi Algoritmaları Kullanarak Kalp Hastalıklarının Tespit Edilmesi. *Avrupa Bilim ve Teknoloji Dergisi*, (28), 1112-1116.
- Çalış, A., Kayapınar, S., & Çetinyokuş, T. (2014). Veri Madenciliğinde Karar Ağacı Algoritmaları ile Bilgisayar ve İnternet Güvenliği Üzerine Bir Uygulama. *Endüstri Mühendisliği*, 25(3), 2-19.
- Çetin, S. B. (2021), *Bilgisayara dayalı karar destek sisteminin acil servis triyaj yönetimine etkisi* [Doktora tezi], Akdeniz Üniversitesi.
- Çobanoğlu, B. (2020). Herkes için Python. Pusula.
- Değirmenci, S. (2020), *Ankara Üniversitesi Tıp fakültesi Acil Tıp anabilim dalında oluşturulan ulusal triyaj projesi 'acil servis triyaj karar destek sistemi – Ankütriyaj'ın Kanada triyaj aciliyet sistemi ve manchester triyaj sistemi ile karşılaştırılması*, [Uzmanlık tezi], Ankara Üniversitesi.
- Dilki, G., & Başar, Ö. D. (2020). İşletmelerin İflas Tahmininde K-En Yakın Komşu Algoritması Üzerinden Uzaklık Ölçütlerinin Karşılaştırılması. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 19(38), 224-233.
- Dilwani, A. A. R. (2019). *Makine öğrenmesi ile acil triyaj ve hastane yatış tahmini* [Yüksek lisans tezi]. Ankara Üniversitesi
- Doğan, Ö. (2023, 5 Aralık). *Pekiştirmeli öğrenmeye giriş*. <https://medium.com/deep-learning-turkiye/peki%C5%9Ftirmeli-%C3%B6%C4%9Frenmeye-giri%C5%9F-c7c2a8cce50b>
- Dumlu, Ö.F. (2020), *Acil servis başvurularının triyaj sistemine ve acil parametrelerine göre değerlendirilmesi* [Yüksek lisans tezi], Sivas Cumhuriyet Üniversitesi.
- Ehsani, R., & Drabløs, F. (2020). Robust distance measures for k NN classification of cancer data. *Cancer informatics*, 19, 1176935120965542.
- Elbaih, A. H. (2017). Triyaj Türleri. *Arşiv Kaynak Tarama Dergisi*, 26(4), 441-467.

- Esmer, S., Uçar, M. K., İbrahim, Ç. İ. L., & Bozkurt, M. R. (2020). Parkinson hastalığı teşhisi için makine öğrenmesi tabanlı yeni bir yöntem. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 8(3), 1877-1893.
- Garreta, R., Moncecchi, G., Hauck, T., & Hackeling, G. (2017). Scikit-learn: machine learning simplified: implement scikit-learn into every step of the data science pipeline. *Packt Publishing Ltd*.
- Glackin, C. E., & Adivar, M. (2023). Using the power of machine learning in sales research: process and potential. *Journal of Personal selling & sales ManageMent*, 43(3), 178-194.
- GÖKALP, Ö. M.(2022), Makine öğrenmesi. Gazi Üniversitesi, Gazi Bilişim Enstitüsü, Adli Bilişim Bölümü.
- Hong, W. S., Haimovich, A. D., & Taylor, R. A. (2018). Predicting hospital admission at emergency department triage using machine learning. *PloS one*, 13(7), e0201016.
- Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, 5(2), 1.
- Inokuchi, R., Iwagami, M., Sun, Y., Sakamoto, A., & Tamiya, N. (2022). Machine learning models predicting undertriage in telephone triage. *Annals of Medicine*, 54(1), 2989-2996.
- Ivanov, O., Wolf, L., Brecher, D., Lewis, E., Masek, K., Montgomery, K., ... & Reilly, C. (2021). Improving ED emergency severity index acuity assignment using machine learning and clinical natural language processing. *Journal of Emergency Nursing*, 47(2), 265-278.
- Jiang, H., Mao, H., Lu, H., Lin, P., Garry, W., Lu, H., ... & Chen, X. (2021). Machine learning-based models to support decision-making in emergency department triage for patients with suspected cardiovascular disease. *International Journal of Medical Informatics*, 145, 104326.
- Jilani, T., Housley, G., Figueredo, G., Tang, P. S., Hatton, J., & Shaw, D. (2019). Short and Long term predictions of Hospital emergency department attendances. *International journal of medical informatics*, 129, 167-174.
- Kamali, M. F., Tezcan, T., & Yildiz, O. (2019). When to use provider triage in emergency departments. *Management Science*, 65(3), 1003-1019.
- Kartal, E. (2021). Makine öğrenmesi ve yapay zeka. *Tıp Bilişimi*.
- Kemalbay, G., & Alkış, B. N. (2020). Borsa endeks hareket yönünün çoklu lojistik regresyon ve k-en yakın komşu algoritması ile tahmini. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 27(4), 556-569.
- Khemchandani, J. R., Chandra, S. (2016). *Twin Support Vector Machines: Models, Extensions and Applications*. Almanya: Springer International Publishing.
- Kılınç, D., Borandağ, E., Yücalar, F., Tunali, V., Şimşek, M., & Özçift, A. (2016). KNN algoritması ve r dili ile metin madenciliği kullanılarak bilimsel makale tasnifi. *Marmara Fen Bilimleri Dergisi*, 28(3), 89-94.

- Kim, J., Chang, H., Kim, D., Jang, D. H., Park, I., & Kim, K. (2020). Machine learning for prediction of septic shock at initial triage in emergency department. *Journal of critical care*, 55, 163-170.
- Kung, S. Y. (2014). Kernel Methods and Machine Learning. *Birleşik Krallık: Cambridge University Press*.
- Kushwah, J. S., Kumar, A., Patel, S., Soni, R., Gawande, A., & Gupta, S. (2022). Comparative study of regressor and classifier with decision tree using modern tools. *Materials Today: Proceedings*, 56, 3571-3576.
- Kuşku Kıyak, S. (2013)., *Türkiye'de acil tıp alanında yayınlanmış makalelerin dünya tıp literatürüne katkısının analizi*. [Tıpta uzmanlık tezi], İstanbul Üniversitesi.
- Kutateladze, V. (2022). The kernel trick for nonlinear factor modeling. *International Journal of Forecasting*, 38(1), 165-177.
- Kuzu, B. S., & Yakut, S. G. (2020). Destek vektör makineleri yardımıyla imalat sanayisinde mali başarısızlık tahminlerinin teknoloji yoğunluğuna göre incelenmesi. *Osmaniye Korkut Ata Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 4(2), 36-54.
- Larner, A.J. (2021). *The 2X2 Matrix Contingency, Confusion and the Metrics of Binary Classification*. Springer.
- Lee, S., & Lee, Y. H. (2020, March). Improving emergency department efficiency by patient scheduling using deep reinforcement learning. In *Healthcare* (Vol. 8, No. 2, p. 77). MDPI.
- Luo, Y., & El Naqa, I. (2019). Machine learning for radiation oncology. *Big Data in Radiation Oncology*, 41-59.
- Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*. [Internet], 9(1), 381-386.
- Maimon, O. Z., Rokach, L. (2014). *Data Mining With Decision Trees: Theory And Applications (2nd Edition)*. Singapur: World Scientific Publishing Company.
- Martínez-Ramón, M., Christodoulou, C. (2022). *Support Vector Machines for Antenna Array Processing and Electromagnetics*. Polonya: Springer International Publishing.
- Metlek, S., & Kayaalp, K. (2020). *Makine Öğrenmesinde, Teoriden Örnek MATLAB Uygulamalarına Kadar Destek Vektör Makineleri*. İksad Yayınevi.
- Mišić, V. V., Gabel, E., Hofer, I., Rajaram, K., & Mahajan, A. (2020). Machine learning prediction of postoperative emergency department hospital readmission. *Anesthesiology*, 132(5), 968-980.
- Mueller, B., Kinoshita, T., Peebles, A., Graber, M. A., & Lee, S. (2022). Artificial intelligence and machine learning in emergency medicine: a narrative review. *Acute medicine & surgery*, 9(1), e740.
- Murty, M., Raghava, R. (2016). *Support Vector Machines and Perceptrons: Learning, Optimization, Classification, and Application to Social Networks*. Almany: Springer International Publishing.

- Nti, I. K., Nyarko-Boateng, O., & Aning, J. (2021). Performance of machine learning algorithms with different K values in K-fold cross-validation. *J. Inf. Technol. Comput. Sci*, 6, 61-71.
- Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3), 128-138.
- Özgül, F. (2013). Python Kılavuzu.
- Öztürk, M. (2023, 6 Aralık). *Python çapraz doğrulama teknikleri*. <https://miracozturk.com/capraz-dogrulama-teknikleri-cross-validation/>
- Pak, A., Gannon, B., & Staib, A. (2021). Predicting waiting time to treatment for emergency department patients. *International Journal of Medical Informatics*, 145, 104303.
- Pal, K., & Patel, B. V. (2020, March). Data classification with k-fold cross validation and holdout accuracy estimation methods with 5 different machine learning techniques. In 2020 fourth international conference on computing methodologies and communication (ICCMC) (pp. 83-87). IEEE.
- Pandimurugan, V., Usha, D., Guptha, M. N., & Hema, M. S. (2022). Random forest tree classification algorithm for predicating loan. *Materials Today: Proceedings*, 57, 2216-2222.
- Panesar, A. (2019). Machine learning and AI for healthcare (pp. 1-73). Coventry, UK: Apress.
- Parker, C. A., Liu, N., Wu, S. X., Shen, Y., Lam, S. S. W., & Ong, M. E. H. (2019). Predicting hospital admission at the emergency department triage: A novel prediction model. *The American journal of emergency medicine*, 37(8), 1498-1504.
- Pate, A., Riley, R. D., Collins, G. S., Van Smeden, M., Van Calster, B., Ensor, J., & Martin, G. P. (2023). Minimum sample size for developing a multivariable prediction model using multinomial logistic regression. *Statistical Methods in Medical Research*, 32(3), 555-571.
- Patel, S. J., Chamberlain, D. B., & Chamberlain, J. M. (2018). A machine learning approach to predicting need for hospitalization for pediatric asthma exacerbation at the time of emergency department triage. *Academic emergency medicine*, 25(12), 1463-1470.
- Patro, R. (2023, 6 Aralık). *Çapraz Doğrulama: K Fold ve Monte Carlo*. <https://towardsdatascience.com/cross-validation-k-fold-vs-monte-carlo-e54df2fc179b>
- Petekçi, A.R.(2021), *Teorik Olarak Makine Öğrenimi*. Kitapyurdu Doğrudan Yayıncılık.
- Pineda, A. L., Ye, Y., Visweswaran, S., Cooper, G. F., Wagner, M. M., & Tsui, F. R. (2015). Comparison of machine learning classifiers for influenza detection from emergency department free-text reports. *Journal of biomedical informatics*, 58, 60-69.

- Qutab, I., Malik, K. I., & Arooj, H. (2022). Sentiment classification using multinomial logistic regression on roman Urdu text. *Int. J. Innov. Sci. Technol*, 4, 223-335.
- Rad, J. A., Parand, K., Chakraverty, S. (2023), *Learning with Fractional Orthogonal Kernel Classifiers in Support Vector Machines: Theory, Algorithms and Applications*. (2023). Almany: Springer Nature Singapore.
- Radhika, V., Prasad, C. R., & Chakradhar, A. (2022, January). *Smartphone-based human activities recognition system using random forest algorithm*. In 2022 International Conference for Advancement in Technology (ICONAT) (pp. 1-4). IEEE.
- Raita, Y., Goto, T., Faridi, M. K., Brown, D. F., Camargo, C. A., & Hasegawa, K. (2019). Emergency department triage prediction of clinical outcomes using machine learning models. *Critical care*, 23(1), 1-13.
- Rastogi, V., Satija, S., Sharma, P.K., Singh S. (2020). Machine Learning Algorithms: Overview. *International Journal of Advanced Research in Engineering and Technology*, 11(9), pp. 512-517.
- Rawal, A., & Lal, B. (2023). Predictive model for admission uncertainty in high education using Naïve Bayes classifier. *Journal of Indian Business Research*, 15(2), 262-277.
- Resmi gazete. (2023, 5 Aralık). *Yataklı sağlık tesislerinde acil servis hizmetlerinin uygulama usul ve esasları hakkında tebliğ*. <https://www.resmigazete.gov.tr/eskiler/2022/09/20220913-5.htm>
- Saygın, E., & Baykara, M. (2021). Karaciğer Yetmezliği Teşhisinde Özellik Seçimi Kullanarak Makine Öğrenmesi Yöntemlerinin Başarılarının Ölçülmesi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 33(2), 367-377.
- Sevli, O. (2019). Göğüs kanseri teşhisinde farklı makine öğrenmesi tekniklerinin performans karşılaştırması. *Avrupa Bilim ve Teknoloji Dergisi*, (16), 176-185.
- Shafaf, N., & Malek, H. (2019). Applications of machine learning approaches in emergency medicine; a review article. *Archives of academic emergency medicine*, 7(1).
- Sharifani, K., & Amini, M. (2023). Machine Learning and Deep Learning: A Review of Methods and Applications. *World Information Technology and Engineering Journal*, 10(07), 3897-3904.
- Shen, Y., & Lee, L. H. (2020). Improving the wait time to triage at the emergency department. *BMJ Open Quality*, 9(1), e000708.
- Sindayigaya, L., & Dey, A. (2022). Machine Learning Algorithms: A Review. *Information Systems Journal*, vol. ISJ-RA-3392, sp, 6.
- Singh, N., Jena, S., & Panigrahi, C. K. (2022). A novel application of Decision Tree classifier in solar irradiance prediction. *Materials Today: Proceedings*, 58, 316-323.
- Somogyi, Z. (2021). *The Application of Artificial Intelligence*. Springer International Publishing: Berlin/Heidelberg, Germany.
- Sultonov, S. (2023). Importance of python programming language in machine learning. *International Bulletin of Engineering and Technology*, 3(9), 28-30.

- Suykens, J. A., Signoretto, M., & Argyriou, A. (Eds.). (2014). Regularization, optimization, kernels, and support vector machines. *CRC Press*.
- Şanlı, A. (2023, 6 Aralık). *Makine Öğrenmesi — 3: KNN & CART*. <https://medium.com/@sanlian/makine-%C3%B6%C4%9Frenmesi-3-knn-cart-c6d201e7278f>
- Şimşek, D. Ö. (2018). Triaaj Sistemlerine Genel Bakış ve Türkiye’de Acil Servis Başvurularını Etkileyen Faktörlerin Lojistik Regresyon ile Belirlenmesi. *Sosyal Güvençe*, (13), 84-115.
- Taheri, R., Ghahramani, M., Javidan, R., Shojafar, M., Pooranian, Z., & Conti, M. (2020). Similarity-based Android malware detection using Hamming distance of static binary features. *Future Generation Computer Systems*, 105, 230-247.
- Tamilarasi, P., & Rani, R. U. (2020, March). *Diagnosis of crime rate against women using k-fold cross validation through machine learning*. In 2020 fourth international conference on computing methodologies and communication (ICCMC) (pp. 1034-1038). IEEE.
- Tang, Y., Chang, Y., & Li, K. (2023). Applications of K-nearest neighbor algorithm in intelligent diagnosis of wind turbine blades damage. *Renewable Energy*, 212, 855-864.
- Taşcı, E., & Onan, A. (2016). K-en yakın komşu algoritması parametrelerinin sınıflandırma performansı üzerine etkisinin incelenmesi. *Akademik Bilişim*, 1(1), 4-18.
- Torun, G. (2016). *Acil servise başvuran travma hastalarının mortalite tahmininde triyaj erken uyarı skoru (TWES) değerliliği*. [Uzmanlık tezi], Uludağ Üniversitesi.
- Xing, W., & Bei, Y. (2019). *Medical health big data classification based on KNN classification algorithm*. IEEE Access, 8, 28808-28819.
- Xiong, Z., Cui, Y., Liu, Z., Zhao, Y., Hu, M., & Hu, J. (2020). Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation. *Computational Materials Science*, 171, 109203.
- Yazıcı, M. (2022). *Acil servise başvuran erişkin hastaların triyaj vital değerlerinin, hastanın aciliyet durumunu öngörmedeki değeri*. [Tıpta uzmanlık tezi], Ankara Üniversitesi.
- Zhao, H., Wang, R., Lei, Y., Liao, W. H., Cao, H., & Cao, J. (2022). Severity level diagnosis of Parkinson’s disease by ensemble K-nearest neighbor under imbalanced data. *Expert Systems with Applications*, 189, 116113.
- Zhou, Z. (2021). *Machine Learning*. Singapur: Springer Nature Singapore.

EKLER

EK A. Türkiye’de uygulanan triyaj sisteminin ayrıntıları

EK B. Veri seti özelliklerini içeren tablolar



EK A

Tablo A.1. Türkiye triyaj sistemi.

Renk	Alan ve vaka niteliği	Vakanın yönlendirildiği birim kriterleri	Örnek durumlar
Yeşil	Ayaktan başvuran, genel durumu itibariyle stabil olan, akut semptomları sebebiyle beklemesi hayati tehdit, morbidite veya komplikasyon oluşturmayacak basit sağlık sorunları bulunan hastalar	Bu kategoride olan hastalar acil servis içerisinde yeşil alanda muayene edilmelidir.	Genel durumu ve hayati bulguları stabil olan hastada her türlü basit belirti Yüksek risk taşımayan ve hafif derecedeki her türlü ağrı Aktif yakınması olmayan düşük riskli hastalık öyküsü
Sarı	Kategori 1: Hayati tehdit etme olasılığı, uzuv kaybı riski ve önemli morbidite oranı olan durumlar	Bu kategoride değerlendirilen hastalar hekim veya triyaj eğitimi almış sağlık personelinin bilgisi dahilinde doğrudan ilgili tanı/tedavi alanına alınmalıdır.	Diastolik>110 mmHg, sistolik>180 mmHg olan kan basıncı yüksekliği Herhangi bir nedenle orta derecede kan kaybı Yardımcı solunum kaslarının solunuma katılmadığı orta derecede solunum sıkıntısı Nöbet geçirme öyküsü(uyanık) Ateş yüksekliği olan onkoloji hastası veya steroid kullanan hasta İnatçı kusma Amnezi ile kafa travması olan ancak bilinci açık hasta Kardiyak öykü ile uyumlu olmayan göğüs ağrısı 65 yaş üstü karın ağrısı olan hasta Şiddetli karın ağrısı olan hasta Deformite, ciddi lacerasyon ve ezilme yaralanması içeren ekstremitte yaralanması Suistimal riski veya şüphesi olan hasta Stresli ve kendine zarar verme riski olan hasta

Tablo A.2. (Devamı): Türkiye triyaj sistemi.

Renk	Alan ve vaka niteliği	Vakanın yönlendirildiği birim kriterleri	Örnek durumlar
Sarı	Kategori 2: Orta ve uzamış dönem belirtileri olan ve ciddiyet potansiyeli taşıyan durumlar. (Solunum sayısı, nabız, kan basıncı, oksijen satürasyonu, vücut sıcaklığı anormal olan hastalar ile medikal tedavi ihtiyacı olan hastalar ve subjektif ağrı skoru maksimum skorun %80'i olan hastalardır.)	Bu kategoride değerlendirilen hastalar hekim veya triyaj eğitimi almış sağlık personelinin bilgisi dahilinde doğrudan ilgili tanı/tedavi alanına alınmalıdır.	Basit kanamalar
			Göğüs ağrısı ve solunum sıkıntısı olmayan basit göğüs yaralanmaları
			Solunum sıkıntısı olmayan yutma zorluğu
			Bilinç kaybı olmayan minör kafa travmaları
			Dehidratasyon belirtileri olmayan kusma ve ishaller
			Normal görme fonksiyonu olan göz inflamasyonları veya gözde yabancı cisim
			Minör ekstremitte travması (ayak bileği burkulması, muhtemel basit fraktür, komplike olmayan laserasyon) normal vital bulgular
			Şiddetli olmayan karın ağrısı
			Zarar verme riski olmayan davranış bozukluğu olan hastalar
			Kardiyak arrest
Kırmızı	Kategori 1: Hayatı tehdit eden ve hızlı agresif yaklaşım ve acil olarak eş zamanlı değerlendirme ve tedavi gerektiren durumlar. Bu durumlarda hasta hiç bekletilmeden kırmızı alana alınır.	Kırmızı ana renk kodunda değerlendirilen hastalardan bilinci kapalı, hava yolu güvenliği, solunum ve dolaşımı olmayan hastalar derhal yeniden canlandırma(resüsitasyon) odasına alınacaktır.	Solunumsal arrest
			Hava yolu tıkanıklığı riski
			Majör çoklu travma
			Solunum sayısı<10 adet/dakika
			Sadece ağrıya yanıt veren veya yanıtız olan hastalar
			Devam eden veya uzamış nöbet

Tablo A.3. (Devamı): Türkiye triyaj sistemi.

Renk	Alan ve vaka niteliği	Vakanın yönlendirildiği birim kriterleri	Örnek durumlar
			İlaç aşırı alımı olan hastanın yanıtsız veya hipoventilasyonda olması
			Kardiyak ağrıya benzer göğüs ağrısı
			Yardımcı solunum kaslarının solunuma katıldığı veya bakılabilen ise pulseoksimetri değerinin <%90'nın altında olduğu ciddi nefes darlığı olan hastalar
			Şiddetli stridor veya yutkunma güçlüğü ile beraber olan havayolu tıkanıklığı riski
			Dolaşım bozukluğu - Nemli, soğuk deri, perfüzyon bozukluğu - Kalp hızı <50 veya >150 olması - Hemodinamik bulgularla beraber olan hipotansiyon
Kırmızı	Kategori 2: Hayatı tehdit etme olasılığı yüksek olan ve 10 dakika içerisinde değerlendirilip tedavi edilmesi gerekli durumlar	Bilinci kapatabilecek, havayolu güvensiz ve solunumu durabileceği ön görülen hastalar doğrudan kırmızı alana alınacaktır.	Akut hemiparazi/disfazi
			Letarji ile ateş (her yaş)
			İrriyasyon gerektiren asit/alkali ile göz teması
			Majör fraktür veya amputasyon gibi ciddi lokalize travma
			Herhangi bir nedenle olan ciddi ağrı - Önemli sedatif veya diğer toksik maddelerin oral alımı
			Davranışsal/Psikiyatrik - Şiddet içeren agresif davranışlar - Kendine veya diğerlerine zarar veren davranışlar

EK B**Tablo B.1.** Veri seti bilgileri-Genel bilgiler.

Bilgiler	Sınıf	Değişken adı	Değişken açıklaması	Veri tipi	Veri seti sınırları
Genel bilgiler	Girdi	cinsiyet	Cinsiyet	Kategorik	Kadın (0) Erkek (1)
	Girdi	yas	Yaş	Nicel	18-104

Tablo B.2. Veri seti bilgileri-Yaşamsal bilgiler.

Bilgiler	Sınıf	Değişken adı	Değişken açıklaması	Veri tipi	Veri seti sınırları
Yaşamsal bilgiler (Vital bulgular)	Girdi	sistolik	Sistolik tansiyon değeri (mmHg)	Nicel	0-281
	Girdi	diastolik	Diastolik tansiyon değeri (mmHg)	Nicel	0-186
	Girdi	map	MAP değeri (mmHg)	Nicel	0 -212
	Girdi	nabiz	Nabız sayısı (atım/dakika)	Nicel	0-202
	Girdi	ss	Solunum sayısı (adet/dakika)	Nicel	0-45
	Girdi	ates	Ateş değeri (C°)	Nicel	33.7-41
	Girdi	spo2	Oksijen satürasyon değeri-SpO2 (%)	Nicel	49-100
	Girdi	kansekeri	Kan şekeri değeri (mg/dl)	Nicel	33+
	Girdi	gks	Glasgow koma skalası (GKS)	Nicel	3-15

Tablo B.3. Veri seti bilgileri-Hastalık şikayetleri.

Bilgiler	Sınıf	Değişken adı	Değişken açıklaması	Veri tipi	Veri seti sınırları
Hastaların belirttiği şikayet bilgileri	Girdi	arrest	Arrest durumu	Kategorik	Yok (0) Var (1)
	Girdi	BogazOksuruk Kulak	Boğaz ağrısı, öksürük, kulak ağrısı durumu	Kategorik	Yok (0) Var (1)
	Girdi	balgam	Balgam varlığı, niteliğinde değişiklik durumu	Kategorik	Yok (0) Var (1)
	Girdi	yanAgrisi	Yan ağrısı durumu	Kategorik	Yok (0) Var (1)
	Girdi	gogusAgrisi	Göğüs ağrısı durumu	Kategorik	Yok (0) Var (1)
	Girdi	nefesDarligi	Nefes darlığı durumu	Kategorik	Yok (0) Var (1)
	Girdi	carpinti	Çarpıntı durumu	Kategorik	Yok (0) Var (1)
	Girdi	atesSikayeti	Yüksek ateş durumu	Kategorik	Yok (0) Var (1)
	Girdi	karinAgrisi	Karın ağrısı durumu	Kategorik	Yok (0) Var (1)
	Girdi	kanKusma	Kan kusma/siyah dışkılama durumu	Kategorik	Yok (0) Var (1)
	Girdi	basAgrisi	Baş ağrısı durumu	Kategorik	Yok (0) Var (1)
	Girdi	halsizlik	Yaygın vücut ağrısı, halsizlik, beslenme bozukluğu durumu	Kategorik	Yok (0) Var (1)
	Girdi	yuksekSeker	Yüksek kan şekeri durumu	Kategorik	Yok (0) Var (1)
	Girdi	bulantiKusma	Bulantı, kusma durumu	Kategorik	Yok (0) Var (1)

Tablo B.3. (Devamı): Veri seti bilgileri- Hastalık şikayetleri.

Bilgiler	Sınıf	Değişken adı	Değişken açıklaması	Veri tipi	Veri seti sınırları
Hastaların belirttiği şikayet bilgileri	Girdi	basDonmesi	Baş dönmesi durumu	Kategorik	Yok (0) Var (1)
	Girdi	ishal	İshal durumu	Kategorik	Yok (0) Var (1)
	Girdi	bayilma	Bayılma durumu	Kategorik	Yok (0) Var (1)
	Girdi	bayilayazma	Bayılayazma durumu	Kategorik	Yok (0) Var (1)
	Girdi	mekanikAgri	Mekanik ağrı durumu	Kategorik	Yok (0) Var (1)
	Girdi	travma	Travma durumu	Kategorik	Yok (0) Var (1)
	Girdi	idrarYakinma	İdrar yakınmaları durumu	Kategorik	Yok (0) Var (1)
	Girdi	saraNobeti	Sara nöbeti durumu	Kategorik	Yok (0) Var (1)
	Girdi	burunKanama	Burun kanaması durumu	Kategorik	Yok (0) Var (1)
	Girdi	bilincBulanik	Bilinç bulanıklığı durumu	Kategorik	Yok (0) Var (1)
	Girdi	norolojikDefisit	Nörolojik defisit durumu	Kategorik	Yok (0) Var (1)
	Girdi	psikiyatrik	Psikiyatrik rahatsızlıklar durumu	Kategorik	Yok (0) Var (1)
	Girdi	zararVerme	Kendine ya da çevresine zarar verme durumu	Kategorik	Yok (0) Var (1)
	Girdi	kasinti	Kaşıntı durumu	Kategorik	Yok (0) Var (1)

Tablo B.3. (Devamı): Veri seti bilgileri- Hastalık şikayetleri.

Bilgiler	Sınıf	Değişken adı	Değişken açıklaması	Veri tipi	Veri seti sınırları
Hastaların belirttiği şikayet bilgileri	Girdi	yanikYangin	Yanık ya da yangına maruziyet durumu	Kategorik	Yok (0) Var (1)
	Girdi	vajinalKanama	Vajinal kanma durumu	Kategorik	Yok (0) Var (1)
	Girdi	alkol	Alkol-madde alımı, CO intoks durumu	Kategorik	Yok (0) Var (1)
	Girdi	gozSikayet	Gözden şikayet durumu	Kategorik	Yok (0) Var (1)

Tablo B.4. Veri seti bilgileri-Özgeçmiş hastalık bilgileri.

Bilgiler	Sınıf	Değişken adı	Değişken açıklaması	Veri tipi	Veri seti sınırları
Sahip olduğu hastalıklar Özgeçmiş hastalık bilgileri	Girdi	dm	Şeker hastalığı durumu (DM)	Kategorik	Yok (0) Var (1)
	Girdi	ht	Hipertansiyon hastalığı durumu (HT)	Kategorik	Yok (0) Var (1)
	Girdi	kah	Kroner arter hastalığı durumu (KAH)	Kategorik	Yok (0) Var (1)
	Girdi	kky	Konjestif kalp yetmezliği hastalığı durumu (KKY)	Kategorik	Yok (0) Var (1)
	Girdi	ritimBozuklugu	Ritim bozukluğu durumu	Kategorik	Yok (0) Var (1)
	Girdi	svo	Serebro vasküler hastalık durumu (SVO)	Kategorik	Yok (0) Var (1)
	Girdi	koahAstimAkciger	KOAH, astım, akciğer hastalığı, pulmoner emboli durumu	Kategorik	Yok (0) Var (1)
	Girdi	kby	Kronik böbrek yetmezliği (KBY), tek böbrek, böbrek nakli durumu	Kategorik	Yok (0) Var (1)

Tablo B.4. (Devamı): Veri seti bilgileri-Özgeçmiş hastalık bilgileri.

Bilgiler	Sınıf	Değişken adı	Değişken açıklaması	Veri tipi	Veri seti sınırları
Sahip olduğu hastalıklar	Girdi	karaciger	Karaciğer yetmezliği, siroz, karaciğer nakli durumu	Kategorik	Yok (0) Var (1)
	Girdi	malignite	Malignite ve komplikasyonları durumu	Kategorik	Yok (0) Var (1)
Özgeçmiş hastalık bilgileri	Girdi	gebelik	Gebelik ve komplikasyonları durumu	Kategorik	Yok (0) Var (1)
	Girdi	digerNorolojik	Diğer nörolojik hastalıkların durumu	Kategorik	Yok (0) Var (1)

Tablo B.5. Veri seti bilgileri-Durum bilgileri

Bilgiler	Sınıf	Değişken adı	Değişken açıklaması	Veri tipi	Veri seti sınırları
Durum bilgisi	Girdi	acileVaris	Acile varış şekli	Kategorik	Ayaktan (0) Ambulans (1)

Tablo B.6. Veri seti bilgileri-Sonuç bilgileri

Bilgiler	Sınıf	Değişken adı	Değişken açıklaması	Veri tipi	Veri seti sınırları
Sonuç bilgisi	Tahmin	YatisTaburcu	Taburculuk durumu	Kategorik	Yatış
					Taburcu
	Tahmin	TriyajAlan	Triyaj alan sınıflandırması	Kategorik	Kırmızı
					Sarı
					Yeşil

ÖZGEÇMİŞ

Ad-Soyad : Gülsüm SALTAN YAŞLI

ÖĞRENİM DURUMU:

- **Lisans** : 2015, Sakarya Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü
- **Yükseklisans** : 2024, Sakarya Üniversitesi, Endüstri Mühendisliği Anabilim Dalı

MESLEKİ DENEYİM VE ÖDÜLLER:

- 2016-2017 yılları arasında Birgi Mefar İlaç Sanayi'nde Operasyonel Mükemmellik uzmanı olarak çalıştı.
- 2017-2018 yılları arasında Hepsiburada.com'da Süreç Geliştirme uzmanı olarak çalıştı.
- 2018-Halen Mutlu Akü ve Malzemeleri Sanayii A.Ş.'de Süreç Geliştirme yöneticisi olarak çalışmaktadır.

TEZDEN TÜRETİLEN ESERLER:

- Arslankaya S., Saltan Yaşlı G. (2023), Predicting Diabetes Disease Using Data Mining Classification Algorithms And Comparison Of Algorithm Performances, *Zeki Sistemler Teori ve Uygulamaları Dergisi (Hakem sürecinde)*

DİĞER ESERLER:

- Korkusuz Polat T., Saltan Yaşlı G. (2023), Risk Prioritization in A Manufacturing Project with Fuzzy SWARA and Fuzzy MOORA Methods, *Erzincan University, Journal of Science and Technology (Yayım Sürecinde)*