



T.C.
KONYA TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



KALP YETMEZLİĞİNİN MAKİNE
ÖĞRENME ALGORİTMALARIYLA TESPİTİ

ORÇUN BAĞRA

YÜKSEK LİSANS

Bilgisayar Mühendisliği Anabilim Dalı

Haziran - 2024
KONYA
Her Hakkı Saklıdır

TEZ KABUL VE ONAYI

Orçun BAĞRA tarafından hazırlanan “KALP YETMEZLİĞİNİN MAKİNE ÖĞRENME ALGORİTMALARIYLA TESPİTİ” adlı tez çalışması 07/06/2024 tarihinde aşağıdaki jüri tarafından oy birliği / oy çokluğu ile Konya Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı’nda YÜKSEK LİSANS/~~DOKTORA~~ TEZİ olarak kabul edilmiştir.

Jüri Üyeleri

Başkan

Prof. Dr. Mustafa Servet KIRAN

Danışman

Prof. Dr. Halife KODAZ

Üye

Dr. Öğr. Üyesi Eyüp SIRAMKAYA

İmza

.....

.....

.....

Yukarıdaki sonucu onaylarım.

Prof. Dr. Mevlüt UYAN
Enstitü Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Orçun BAĞRA

07/06/2024

ÖZET

YÜKSEK LİSANS

KALP YETMEZLİĞİNİN MAKİNE ÖĞRENME ALGORİTMALARIYLA TESPİTİ

Orçun BAĞRA

Konya Teknik Üniversitesi
Lisansüstü Eğitim Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Halife KODAZ

2024, 88 Sayfa

Jüri

Prof. Dr. Halife KODAZ
Prof. Dr. Mustafa Servet KIRAN
Dr. Öğr. Üyesi Eyüp SIRAMKAYA

Verinin önemli ve anlamlı hale geldiği günümüzde sağlık alanındaki verilerin de gelişen teknoloji karşısında bilgiye dönüşmesi bir zorunluluk haline gelmiştir. Kardiyovasküler hastalıklar, vücudun ihtiyacı olan kan miktarının yeterli miktarda pompalanamaması, kalbin yapısal ve işlevsel bozukluğa sahip olması, oksijen yetersizliği, damar tıkanıklığı gibi birçok sebeplerden dolayı baş gösteren bir hastalıktır. Dünya genelinde yılda yaklaşık olarak 18 milyon kişinin kardiyovasküler hastalıklardan kaynaklı yaşamını yitirdiği bilinmektedir. Hastane personelinin karar verme sürecinde yardımcı olması ve olası beşerî hataların en aza indirgenmesi amacıyla birçok alanda olduğu gibi tıp alanında da yapay zekâ algoritmaları kullanılmaya başlanmıştır. Bu tez çalışmasında etik kurul onaylı framingham kriterleri ve uzman görüşüyle oluşturulan ölçek ile Nevşehir Devlet Hastanesi Kardiyoloji Bölümünden 202 adet veri toplanarak veri seti elde edilmiştir. Bu veri setinin yanı sıra Kaliforniya Üniversitesinin açık erişime sunduğu “Heart Failure Prediction” isimli kalp yetmezliği için oluşturulmuş 299 adet veriye sahip veri seti de çalışmaya dahil edilmiştir. Çalışmada toplamda 501 adet örnek iki veri seti üzerinde, Yapay Sinir Ağları, Karar Ağacı, Rastgele Orman, Lojistik Regresyon, K en Yakın Komşu, Naive Bayes makine öğrenme algoritmalarıyla oluşturulan modellere beşli çapraz doğrulama uygulanmıştır. Toplanan verilerin ön işleme sürecinden sonra karmaşıklık matrisi ile doğruluk, duyarlılık, f1 skor ve kesinlik kriterleriyle oluşturulan modellerin performans sonuçları hesaplanmıştır. Oluşturulan veri setinin doğruluk değerleri Karar Ağacı (%95), Rastgele Orman (%93), Lojistik Regresyon (%90), K en Yakın Komşu (%81), Naive Bayes (%80), Yapay Sinir Ağları (%76) oranlarında başarı elde edilmiştir. Heart Failure Prediction isimli hazır veri seti için ise sıralama Rastgele Orman (%97), Karar Ağacı (%94), Lojistik Regresyon (%84), Naive Bayes (%83), Yapay Sinir Ağları (%79), K en Yakın Komşu (%70) oranlarında başarı elde edilmiştir. Ayrıca WEKA programı ile kalp hastalığını etkileyen en önemli öznitelikler Ejeksiyon Fraksiyon, kreatinin, anemi, nefes darlığı, diyabet, Atriyal Fibrilasyon olarak tespit edilmiştir.

Anahtar Kelimeler: Kalp Yetmezliği, Makine Öğrenmesi, Sınıflandırma, Veri Seti Oluşturma

ABSTRACT**MS/PhD THESIS****DETECTION OF HEART FAILURE WITH MACHINE LEARNING
ALGORITHMS****Orçun BAĞRA****Konya Technical University
Institute of Graduate Studies
Department of Computer Engineering****Advisor: Prof. Dr. Halife KODAZ****2024, 88 Pages****Jury
Prof. Dr. Halife KODAZ
Prof. Dr. Mustafa Servet KIRAN
Dr. Öğr. Üyesi Eyüp SIRAMKAYA**

Nowadays, where data has become important and meaningful, it has become inevitable for data in the field of health to transform into information in the face of developing technology. Cardiovascular disease is a disease that occurs due to many reasons such as the body's inability to pump the amount of blood it needs, structural and functional disorders of the heart, oxygen deficiency, and vascular occlusion. It is known that approximately 18 million people throughout the world die from cardiovascular diseases annually. Artificial intelligence algorithms have begun to be used in the field of medicine, as in many other fields, in order to assist hospital staff in the decision-making process and to minimize possible human errors. In this thesis study, a data set has been obtained by collecting 202 data from the Cardiology Department of Nevşehir State Hospital with the scale created by the ethical committee-approved Framingham criteria and specialist opinion. In addition to this data set, a data set containing 299 data created for heart failure called "Heart Failure Prediction", which is open access by the University of California, was also included in the study. In the study, five-way cross-validation has been applied to models created with Artificial Neural Networks, Decision Tree, Random Forest, Logistic Regression, K Nearest Neighbor, Naive Bayes machine learning algorithms on two data sets with a total of 501 samples. After the pre-processing process of the collected data, the performance results of the models created with the complexity matrix and accuracy, sensitivity, f1 score and precision criteria have been calculated. The accuracy values of the created data set have been achieved in Decision Tree (95%), Random Forest (93%), Logistic Regression (90%), K Nearest Neighbor (81%), Naive Bayes (80%) and Artificial Neural Networks (76%). Besides, with the WEKA program, the most important attributes affecting heart disease have been determined as Ejection Fraction, creatinine, anemia, shortness of breath, diabetes and Atrial Fibrillation.

Keywords: Classification, Dataset Creation, Heart Failure, Machine Learning.

ÖNSÖZ

Tez çalışmamın her aşamasında destek veren danışmanım Prof. Dr. Halife KODAZ hocama, veri setinin parametrelerinin hazırlanmasında uzman görüşü aldığım Medipol Üniversitesi Kardiyoloji Bölümü Prof. Dr. Dursun DUMAN hocama, tez yazım sürecinde bana sonsuz desteği ile güç veren değerli eşim Akgül BAĞRA'ya sonsuz teşekkürlerimi bir borç bilirim.

Orçun BAĞRA
KONYA-2024



İÇİNDEKİLER

ÖZET	1
ABSTRACT.....	2
ÖNSÖZ	3
İÇİNDEKİLER	4
SİMGELER VE KISALTMALAR	6
1. GİRİŞ	8
1.1. Tezin Amacı ve Organizasyonu.....	9
2. KAYNAK ARAŞTIRMASI	11
2.1. Oluşturulan Veri Seti	11
2.2. Heart Failure Prediction Veri Seti	12
2.3. Kalp Yetmezliğine Yönelik Literatürdeki Çalışmalar	13
3. MATERYAL VE YÖNTEM.....	18
3.1. Kalp Yetmezliği ve Teşhis Yöntemleri	18
3.1.1. Göğüs bölgesi röntgeni	18
3.1.2. Ekokardiyografi	20
3.1.3. Kan testi	23
3.1.4. New York efor testi.....	24
3.1.5. Amerikan kalp derneği efor testi	25
3.2. Makine Öğrenmesi.....	26
3.2.1. K En Yakın Komşu algoritması.....	29
3.2.2. Karar Ağacı algoritması.....	30
3.2.3. Lojistik Regresyon algoritması.....	31
3.2.4. Naive Bayes algoritması	32
3.2.5. Rastgele Orman algoritması.....	33
3.2.6. Yapay Sinir Ağları algoritması	34
3.3. Veri Ön İşleme.....	36
3.3.1. Eksik verilerin bulunması ve temizlenmesi	39
3.3.2. Verilerin birleştirilmesi.....	40
3.3.3. Standardizasyon	40
3.3.4. Normalizasyon	41
3.3.5. Özellik seçimi	42
3.3.6. Heart failure prediction veri seti ön işleme.....	43
3.3.7. Çapraz doğrulama (cross validation)	48
3.4. Değerlendirme Ölçütleri	49
3.4.1. Karmaşıklık matrisi.....	49
3.5. Önerilen Yöntem.....	51
3.5.1. Verinin toplanması ve ön işlenmesi.....	52
3.5.2. Verilerin bölünmesi	52

3.5.3. Modelin eğitimi.....	53
3.5.4. Modelin değerlendirilmesi.....	54
4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA.....	55
4.1. Oluşturulan Veri Setinin Sonuçları.....	55
4.1.1. K En Yakın Komşu algoritması sonuçları	55
4.1.2. Karar Ağacı algoritması sonuçları	58
4.1.3. Lojistik Regresyon algoritması sonuçları	60
4.1.4. Naive Bayes algoritması sonuçları	61
4.1.5. Rastgele Orman algoritması sonuçları	63
4.1.6. Yapay Sinir Ağları algoritması sonuçları	65
4.2. Hazır Veri Setinin Sonuçları.....	67
4.2.1. K En Yakın Komşu algoritması sonuçları	67
4.2.2. Karar Ağacı algoritması sonuçları	68
4.2.3. Lojistik Regresyon algoritması sonuçları	70
4.2.4. Naive Bayes algoritması sonuçları	71
4.2.5. Rastgele Orman algoritması sonuçları	73
4.2.6. Yapay Sinir Ağları algoritması sonuçları	74
4.3. Tartışma	76
5. SONUÇLAR VE ÖNERİLER.....	80
5.1 Sonuçlar	80
5.2 Öneriler	81
KAYNAKLAR	82
EKLER	85

SİMGELER VE KISALTMALAR

Kısaltmalar

ACC	:	Avrupa Kardiyoloji Cemiyeti
AF	:	Atriyal Fibrilasyon
AHA	:	Amerikan Kalp Birliği
ANN	:	Artificial Neural Network
ASC	:	Attribute Selected Classifier
AUC	:	Area Under Curve
CART	:	Classification and Regression Tree
CNN	:	Evrişimsel Sinir Ağları
DL	:	Derin Öğrenme
DT	:	Karar Ağacı
EF	:	Ejeksiyon Fraksiyon
EKG	:	Elektrokardiyografi
FN	:	Yanlış Negatif
FP	:	Yanlış Pozitif
FPR	:	Yanlış Pozitif Değer
GB	:	Gradyan Arttırma
GSB	:	Gradyan Hayatta Kalmayı Arttırma
HT	:	Yüksek Tansiyon
KNN	:	K En Yakın Komşu
KY	:	Kalp Yetmezliği
LR	:	Lojistik Regresyon
LVEF	:	Sol Ventrikül Ejeksiyon Fraksiyon
LVH	:	Kalp Duvar Kalınlığı
METS	:	Metabolik Equivalents
ML	:	Makine Öğrenme
MRG	:	Kardiyak Manyetik Rezonans Görüntüleme
MSSQL	:	Microsoft SQL
NB	:	Naive Bayes
NYHA	:	New York Kalp Cemiyeti
PAAG	:	Ön Arka Akciğer Grafisi
PET	:	Pozitron Emisyon Tomografi
RF	:	Rastgele Orman
ROC	:	Receiver Operating Characteristic
RSF	:	Rastgele Hayatta Kalma Ormanı
RT	:	Random Tree
SVM	:	Destek Vektör Makineleri
TN	:	Gerçek Negatif
TP	:	Gerçek Pozitif

TPR	:	Gerçek Pozitif Değer
VTYS	:	Veri Tabanı Yönetim Sistemi
WEKA	:	Waikato Environment For Knowledge Analysis
YSA	:	Yapay Sinir Ağı



1. GİRİŞ

Yaşam kalitesini ve hayatı tehdit eden kalp yetmezliği (KY), kalbin işlevsel veya yapısal bozukluk sonrasında kalbin ihtiyacı olan oksijen miktarının yeterli gelmemesi, vücudun kan ihtiyacına yetersiz kalma, kalbin kasılma sonucu kanın dolması ve boşalma özelliğinin yitilmesi, akut hastalıklar, genetik gibi durumlarda oluşmaktadır. Avrupa Kardiyoloji Cemiyeti (ACC) tarafından en son 2016 yılında hazırlanan kardiyovasküler durum bildirgesi kılavuzuna göre sol ventrikül ejeksiyon fraksiyon (LVEF) değerini 3 kategoriye göre ayrılmıştır (Taner, 2020).

Kalbin çalışmasındaki yapısal bozulma ve/veya hasarlara bağlı değişiklik ile bireyin semptomlarını dikkate alarak Amerikan Kalp Birliği (AHA) ve (ACC) tarafından oluşturulan kalp yetmezliği evrelerini geliştirmişlerdir. Ejeksiyon Fraksiyon (EF) değeri %40 altında olması durumunda düşük ejeksiyon fraksiyonlu KY, %40 – 49 dahil değerlerinde olması orta düzey ejeksiyon fraksiyonlu KY, %50 ve üzeri değerlerinde korunmuş ejeksiyon fraksiyonlu KY olarak sınıflandırılmıştır. AHA tarafından geliştirilen kardiyovasküler evreleri “Evre A, Evre B, Evre C, Evre D” olarak kalp yetmezliğinin başlangıç evresinden en ileri düzeye kadarki 4 aşamasını temsil etmektedir. Evre A’da hastada semptom ve yapısal kalp hastalığı olmadığı gibi Evre D’de en ileri kalp yetmezliği hastaları sınıflandırılmıştır. New York Kalp Cemiyeti (NTHA) tarafından geliştirilen sınıflandırmaya göre ise evreler “Sınıf I, Sınıf II, Sınıf III, Sınıf IV” olarak. Sınıf I günlük yapılan aktivitelerde fiziksel olarak bir çarpıntı, nefes darlığı gibi şikâyet olmadığını, Sınıf IV ise hastada günlük fiziksel aktivite de bile semptom göstermektedir (İlhan, 2021).

Hastaya KY tanısı konulabilmesi için uygulanan tetkikleri şu şekilde ifade edebiliriz: Laboratuvar analizleri, Elektrokardiyografi (EKG), Ön arka akciğer grafisi (PAAG), Kardiyak manyetik rezonans görüntüleme (MRG), 6 dakika yürüme testi, Torasik ultrasonografi, Pozitron emisyon tomografi (PET) ve Miyokard perfüzyon sintigrafisi. Bu tetkiklerin yanı sıra kalp yetmezliği teşhisinde framingham kriterleri temel alınarak tanı konulmaktadır (İlhan, 2021).

Framingham kriterleri şu şekildedir:

Majör Kriterler: Kardiyomegali, akciğerde raller, s3 galo, akut pulmoner ödem, hepatojuguler reflü, boyun vnöz dolgunluğu, pnd veya ortopne, 5 gün içinde 4.5 kg zayıflama (Taner, 2020).

Minör Kriterler: Taşikardi, plevral sıvı, hepatomegali, efor dispnesi, geceleri kuru öksürük olması, bilateral ayak bileği ödemi, maksimum vital kapasite.

Kalp yetmezliği teşhisi için tıp dünyasında kabul görmüş 2 majör veya 1 majör ile 2 minör kriter olması yeterli gelecektir (Taner, 2020).

Bu çalışmanın amacı kalp yetmezliğine ait framingham kriterleri temel alınarak geliştirilen ölçek doğrultusunda gerçek dünya verileri toplanarak Makine Öğrenme (ML) algoritmaları ile geliştirilen modeller ile kalp yetmezliği tahmin edecek bir sistem geliştirmektedir. Geliştirilen sistemin sağlık çalışanlarının iş yükünü hafifletmesiyle birlikte oluşabilecek beşerî hataları en aza indirmek ayrıca öğrencilere kalp yetmezliğine dair eğitim sürecinde yardımcı kaynak sunulabilmesi amaçlanmaktadır.

Çalışmada Makine Öğrenme algoritmalarından Yapay Sinir Ağları (ANN), Karar Ağacı (DT), Rastgele Orman (RF), Lojistik Regresyon (LR), Naive Bayes (NB), K En Yakın Komşu (KNN) algoritmaları beşli çapraz doğrulama ve değerlendirme ölçütü olarak da karmaşıklık matrisi kullanılmıştır.

Yapılan tez çalışmasının literatüre katkısı KY için bir ölçek geliştirilmesi ve hastane verilerini içeren bir veri seti oluşturulmasıdır. Diğer katkısı Kaliforniya Üniversitesinden edinilen veri seti üzerinde detaylı bir analiz süreci gerçekleşmiştir. Bir diğer katkısı ise sağlık alanındaki öğrencilere eğitim sürecinde ve sağlık çalışanlarına iş yükünün hafifletilmesi sürecinde fayda sağlamaktır.

1.1. Tezin Amacı ve Organizasyonu

Tez çalışması beş ana bölüm olacak şekilde organize edilmiştir. Birinci bölümde KY hastalığı ve teşhisinin önemi, dünyada görülme sıklığı ve tezin hedef ve amaçlarından bahsedilmiştir.

İkinci bölümde KY veri seti için kullanılacak parametrelerin framingham kriterleri temel alınarak oluşturulması, oluşturulan veri seti için verilerin toplanması, örnek sayısı, veri ön işleme adımlarından bahsedilmiştir. Bu tezde ölçeği geliştirilip

örneklerinin gerçek dünya verilerinden toplanarak oluşturulan veri seti yanı sıra Kaliforniya Üniversitesinin “<https://archive.ics.uci.edu>” adresinde yer alan hazır veri seti hakkında bilgiler sunulmuştur. Ayrıca KY’ne yönelik literatürde yer alan çalışmalara yer verilmiştir.

Üçüncü bölümde göğüs bölgesi röntgeni genel bilgisi, EKG ve kan testlerinin genel yorumlanmasından bahsedilmiştir. KY teşhisinde dünyaca kabul görmüş New York Efor Testi ile Amerikan Kalp Derneği Efor Testi için uygulanan adımlar anlatılmıştır. Oluşturulan ve hazır olarak el alınan veri setlerinin analizinde kullanılmak üzere YSA, KNN, DT, LR, NB, RF ML algoritmalarının çalışma mantıkları hakkında genel bilgi verilmiştir. Çalışmada kullanılan veri setlerine uygulanan veri ön işleme adımları eksik verilerin bulunması ve temizlenmesi, verilerin birleştirilmesi, standardizasyon, normalizasyon, özellik seçimi başlıklarıyla oluşturulan veri seti üzerinde uygulaması sunulmuştur. Analiz sonrasında uygulanacak değerlendirme ölçütleri olan karmaşıklık matrisi süreci anlatılmıştır.

Dördüncü bölümde oluşturulan ve heart failure prediction veri seti için ayrı ayrı olarak uygulanan beş ML algoritması (YSA, KNN, DT, LR, NB, RF) için Python Jupyter platformu ile analiz sonuçları açıklanmıştır.

Beşinci ve son bölümde ise yapılan çalışmadan elde edilen sonuçların değerlendirilmesi ve karşılaştırılması ele alınmıştır. Sonrasında araştırma izni için alınan etik kurul raporu da sunulmuştur.

2. KAYNAK ARAŞTIRMASI

Kaynak araştırması üç bölümde oluşturulmuştur. Birinci kısımda tez için kullanılacak framingham kriterleri ve uzman görüşüyle hazırlanan veri seti parametreleri ve bu parametrelerin neyi ifade ettiği açıklanmıştır. İkinci kısımda ise Kaliforniya Üniversitesinin “<https://archive.ics.uci.edu>” adresinden edinilen hazır veri seti üzerinde yer alan örneklem sayısı ve genel bilgilendirme yapılmıştır. Üçüncü bölümde ise kalp yetmezliğine yönelik yapılan çalışmaların özetlenmesi yapılmıştır.

2.1. Oluşturulan Veri Seti

Framingham kriterleri temel alınıp uzman görüşüne başvurularak oluşturulan veri seti ölçeğinde 25 adet parametre bulunmaktadır. Nevşehir Devlet Hastanesi Kardiyoloji bölümünde Eylül 2022’den Mart 2023 tarihleri arasında E-26171210-929 sayılı etik kurul onayı ile oluşturulan ölçek doğrultusunda toplamda 202 adet veri toplanmıştır. Python Jupyter platformunda verilerin analiz edilebilmesi için dosya formatı csv uzantılı hale getirilmiştir. Oluşturulan ölçeğe ait parametreler ve bu parametrelerin veri tipleri ile temsil ettiği verinin açıklamaları Çizelge 2.1’de gösterilmiştir.

Çizelge 2.1. Oluşturulan veri seti parametreleri ve parametre açıklamaları

#	Öznitelik Adı	Veri Tipi	Açıklama
1	Yaş	int64	Hastaya ait yaş verisidir.
2	Cinsiyet	int64	Hastaya ait cinsiyet verisidir. Kadın {0}, Erkek {1} olarak saklanır.
3	Sigara Kullanım Durumu	int64	Sigara içmeme durumunu {0}, içme durumunu ise {1} olarak saklanmaktadır.
4	Alkol Kullanım Durumu	int64	Alkol içmeme durumunu {0}, içme durumunu ise {1} olarak saklanmaktadır.
5	Aile Öyküsü	int64	1. derece yakın akrabalarında kalp rahatsızlığı olup olmadığını tutmaktadır.
6	Atriyal_Fibrilasyon	int64	Hastada ritim bozukluğunun olup olmadığını temsil etmektedir.
7	Daha Önce Felç Geçirme Durumu	int64	Hastanın daha öncesinde felç geçirip geçirmediğini tutmaktadır.
8	Diyabet	int64	Şeker hastalığının olup olmadığını temsil

			eder.
9	Yüksek Tansiyon	int64	Yüksek tansiyon olup olmadığı temsil eder.
10	Kalp Büyümesi Var mı	int64	Hastada kalp büyümesi olup olmadığını temsil eder.
11	Kalp Duvar Kalınlığı Var mı	int64	Kalp duvar kalınlığının olup olmadığı tutulur.
12	Karaciğerde Sıvı Birikmesi Var mı	int64	Karaciğerde sıvı birikmesinin durumu tutulur.
13	Nefes Darlığı Var mı	int64	Hastanın nefes darlığı yaşayıp yaşamadığı bilgisi tutulur.
14	Bacaklarda Şişme Var mı	int64	Ayak ve bacaklarda şişmenin olup olmadığı bilgisi tutulur.
15	Damar Hastalığı Var mı	int64	Hastanın genel olarak damar hastalığı olup olmadığı bilgisi saklanır.
16	Böbrek Yetmezliği Var mı	int64	Hastanın böbrek yetmezliği olup olmadığı tutulur.
17	Anemi Var mı	int64	Hastada kansızlığın olup olmadığı tutulur.
18	LDL	float64	Laboratuvar değerine bakılır.
19	HDL	float64	Laboratuvar değerine bakılır.
20	Sodyum	float64	Laboratuvar değerine bakılır.
21	Glikoz	float64	Laboratuvar değerine bakılır.
22	TSH	float64	Laboratuvar değerine bakılır.
23	Kreatinin	float64	Laboratuvar değerine bakılır.
24	Ejeksiyon Fraksiyon	int64	Kalbin her atışında kanın pompalanan miktarını gösteren bilgi saklanmaktadır.
25	Tanı	int64	Hastanın kalp yetmezliği olup olmadığı tutulur.

2.2. Heart Failure Prediction Veri Seti

Çalışmada oluşturulan veri setinin yanı sıra kaggle web sayfasında “*Heart Failure Prediction*” başlıklı hazır veri seti de kullanılmıştır. Sağlık alanında oluşturulan çok değişkenli bu veri setinde 13 özellik bulunmaktadır. Özellik türleri tam sayı olarak oluşturulmuştur. İlgili veri setinde sınıflandırma, kümeleme ve regresyon analizlerine uygun hazırlanmıştır. Veri setinde 299 adet veri bulunmaktadır. Veri setine kaggle sayfasından ulaşılabildiği gibi Kaliforniya Üniversitesinin “<https://archive.ics.uci.edu>” adresinden de ulaşılabilmektedir. Hazır olarak indirilen bu veri setine ait parametreler ve bu parametrelerin veri tipleri ile temsil ettiği verinin açıklamaları Çizelge 2.2’de gösterilmiştir.

Çizelge 2.2. Heart Failure Prediction veri seti parametreleri ve parametre açıklamaları

#	Öznitelik Adı	Veri Tipi	Açıklama
1	Yaş	int64	Hastaya ait yaş verisidir.
2	Anemi	int64	Hastada kansızlığın olup olmadığı tutulur.
3	Sigara Kullanım Durumu	int64	Sigara içmeme durumunu {0}, içme durumunu ise {1} olarak saklanmaktadır.
4	Ejeksiyon Fraksiyon	int64	Kalbin her atışında kanın pompalanan miktarını gösteren bilgi saklanmaktadır.
5	Kreatinin Fosfokinaz	int64	Laboratuvar değerine bakılır.
6	Cinsiyet	int64	Hastaya ait cinsiyet verisidir. Kadın {0}, Erkek {1} olarak saklanır.
7	Trombosit	int64	Laboratuvar değerine bakılır.
8	Diyabet	int64	Şeker hastalığının olup olmadığını temsil eder.
9	Yüksek Tansiyon	int64	Yüksek tansiyon olup olmadığını temsil eder.
10	Serum Sodyum	int64	Laboratuvar değerine bakılır.
11	Serum Kreatinin	int64	Laboratuvar değerine bakılır.
12	Zaman	int64	Hastanın ne kadar süredir takip edildiği.
13	Ölme Olayı	int64	Tedavi sürecinde hastanın yaşamı yitirmesi olayı.

2.3. Kalp Yetmezliğine Yönelik Literatürdeki Çalışmalar

Altun çalışmasında uyku bozukluğunun nedenlerini tespit edebilmek için uyku evrelerini belirleyebilmek için ML ve Derin Öğrenme (DL) algoritmalarıyla sınıflandırma gerçekleştirmiştir. Physionet veri seti erişim sayfasından 994 adet bireyin polisomnografik sinyal içeren veri seti üzerinde çalışmıştır. Çalışmasında veri ön işleme adımında Seyrek Filtreleme ve Yeniden Yapılandırmadan Bağımsız Bileşen Analiz algoritmaları kullanmıştır. Sınıflandırma algoritması olarak Destek Vektör Makineleri, KNN, RF ve DL algoritmalarını tercih etmiştir. Sınıflandırma metriği olarak karmaşıklık matrisi sonucunda RF algoritmasıyla %97,3 doğruluk oranına ulaşmıştır (Altun, 2023).

Karaj, yüksek lisans çalışmasında geliştirmiş olduğu web sayfasında kullanıcılara dosyadan veri aktarabileceği sistem geliştirmiştir. Web ortamına aktarılan veriler üzerinde anlık değişiklik ve düzenleme fırsatı veren Karaj kullanmış olduğu 7

farklı sınıflandırma ve kümeleme algoritmalarıyla ML destekli olarak programı kullananlara karar vermesinde destek bir uygulama geliştirmiştir. Çalışmasında farklı veri setleri kullanan Karaj ML algoritmalarıyla sonuç tahminleme, çıkan değerlerin karşılaştırılması ve verinin görselleştirilmesini seçilen algoritmalara göre gerçekleştirmiştir. Plotly Dash geliştirme ortamını baz alan Karaj Python platformundaki Pandas, NumPy, Keras, TensorFlow, Scikit-learn kütüphanelerini kullanmıştır. Veri kümesindeki dağılımı eğitim için %70, Test için ise %30 oranında bölmüştür. Uygulamasında MLPClassifier Yapay Sinir Ağı modeli (YSA), SVM, KNN, DT, NB ve CatBoost algoritmalarını kullanmıştır. En iyi sonucu % 97 oran ile KNN algoritmasında hesaplamıştır (Karaj, 2023).

Bilekyiğit yüksek lisans tez çalışmasında kaggle ve UCI sayfalarında elde ettiği 4 farklı veri seti üzerinde sınıflandırma yapmıştır. Çalışmasında doğruluk skorunu arttırmak için beşli çapraz doğrulama (cross validation) kullanmayı tercih etmiştir. Normalizasyon ve standardizasyon yöntemleriyle veri ön işleme sürecini tamamlamıştır. SVM, RF, DT, KNN algoritması, LR ve NB yöntemlerini kullanmıştır. Veri setlerindeki niteliklerin sonuca katkısı üzerinde daha anlamlı olan niteliklerin seçimi için LIME öznelik seçme algoritmasını kullanmıştır. Bilekyiğit çalışmasında veri setlerinin ön işleme ve uygulanan sınıflandırma algoritmaları sonucunda en iyi skoru tüm veri setleri için RF algoritması olduğunu belirlemiştir (Bilekyiğit 2022).

Erdaş çalışmasında sağlık sektöründeki çeşitli branşlardaki veri setleri üzerinde SVM, NB, RF, DT, LR sınıflandırma algoritmaları kullanarak bu makine öğrenme algoritmaları üzerinde alıcı işlem karakteristik (Receiver Operating Characteristic – ROC) eğrisi, doğruluk, geçerlilik F1 skor ve güvenilirlik sonuçları üzerinde karşılaştırma yapmıştır. Oluşturulan ML algoritmalarıyla analizi yapılmadan önce veri türü nesne ve string olan niteliklerin sayısal veriye dönüştürdükten sonra 5 farklı veri setini 5 farklı algoritma modeli sonrasında ortalama doğruluk değeri en yüksek olarak RF algoritması sonuç vermiştir. Ayrıca veri setleri arasında KY veri kümesi denge oran sonucuna göre en başarılı sonucu yakalamıştır (Erdaş, 2022).

Badik çalışmasında Pakistan'ın Faisalabad Tıp Üniversitesi Allied Hastanesinden elde edilen 299 örnekten oluşan veri setini kullanarak sağlık personellerinin kullanabileceği ve beşerî hataların aza indirecek, tedavi sürecinde yardımcı bir destek sistemi tasarlamayı amaçlamıştır. Veri setinde toplamda 12 bağımsız değişken bulunmaktadır. Badik çalışmasında JupyterLab programını kullanarak sınıflandırma niteliğini 1 değerine karşılık vefat eder, 0 verisine karşılık

hayatta kalır olarak kabul etmiştir. Standart Scaler ile verileri ölçeklendirme işlemi sonrasında kategorik nitelikleri veri setine eklemiştir. Çalışmasında veri setini üçte bir oranda test, üçte iki oranda eğitim olarak belirlemiştir. İki farklı yol izleyerek 12 bağımsız niteliklerin 6 niteliğe düşürerek 10 ayrı sınıflandırıcı modellerini oluşturmuştur. Çıkan iki sonucu karşılaştırıldığında CatBoost algoritması daha iyi sonuç verdiğini bulmuştur (Badik, 2022).

Özoğur ve Orman sağlık sektöründeki verilerin değerlendirmesinde verilerin ön işleme tabi tutulmasının makine öğrenme algoritmalarına olan etkisi üzerine etkisini araştırmışlardır. Çalışmalarında diyabet veri seti (PIMA) kullanmışlardır. Veri ön işleme sürecinin önemini ortaya koymak için ilk olarak verileri işlemeden ham olarak incelemişlerdir. İkinci adımda veriler üzerindeki sınıf dengesizliğini gidererek incelemişlerdir. Üçüncü adımda ise eksik verilerin giderilmesi ve sınıfların dengelenmesiyle oluşturulan veri seti üzerinde çalışmışlardır. 5 farklı sınıflandırma algoritması ve 5 katmanlı çapraz doğrulama yöntemini kullanmışlardır. Veri ön işleme sürecine tabi tutulan ve tutulmadan analiz sonuçları karşılaştırılmış olup veri ön işleme süreci tamamlanan kümenin hemen hemen tüm sınıflandırıcı algoritmalarda daha başarılı olduğunu hesaplamışlardır (Özoğur ve Orman, 2022).

Aydın çalışmasında 299 kalp hastalığı olan kişilere ait veriler üzerinden ve Rastgele Hayatta Kalma Ormanı (Random Survival Forest – RSF), Gradyan Arttırma (Gradient Boosting Survival – GSB) ve Cox algoritmaları kullanılarak sağ kalım modelleri geliştirmiştir. Oluşturulan modelde %75 eğitim, %25 test verileri rastgele bölünmüş ve test verilerine göre tahminde bulunmuştur. Çalışmada kritik parametre belirlemek için her bir nitelik ayrı ayrı oluşturulan modellere tabi tutularak niteliklerin tahmin performansına etkisini değerlendirmiştir. Diğer niteliklerden daha yüksek performans gösteren nitelikler daha önemli kabul edilmiştir. Diğer bir kritik parametre belirleme yöntemi olarak verilerin üçerli, beşerli ve tamamını gruplayıp modellere göre sonuçların ayrı ayrı değerlendirilmesidir. Bu değerlendirme sonrasında RSF ve GSB modelleri kullanılarak en iyi tahmini yapan beşerli gruplanan nitelik grubu verdiğini tespit etmiştir. Sonuç olarak çalışmasında sağ kalım modeli ve kritik parametre seçimi ile yaş, Enjeksiyon Fraksiyon, Serum Kreatinin, Trombosit, Yüksek Tansiyon niteliklerinin önemli rol oynadığı tespit etmiştir (Aydın, 2021).

Sayın ve arkadaşları çalışmalarında veri seti paylaşım sayfası kaggle.com adresinden yaklaşık 130000 elektrokardiyografik sinyallerini Evrişimsel Sinir Ağları (Convolutional Neural Networks – CNN) algoritmasını kullanarak sınıflandırma

yapmışlardır. Beş farklı kategoriye göre sınıflandırmışlardır. Çalışmalarında kullandıkları diğer veri setinde ise yaklaşık 15000 adet 125 Herz frekansa sahip elektrokardiyografik sinyallerden oluşmaktadır. Değerlendirme ölçeği olarak karmaşıklık matrisi kullanılmış olup %94,4 doğruluk sonucuna ulaşmışlardır (Sayın ve ark., 2021).

Potur ve Erginel çalışmalarında KY bulunan hastalara ait veri seti üzerinde LR, SVM, NB, J48 karar ağacı ve çok katmanlı algılayıcı algoritmaları kullanılarak sınıflandırma yapmışlardır. WEKA (Waikato Environment for Knowledge Analysis) yazılımında bulunan CfsSubsetEval, ReliefFAttributeEval ve InfoGainAttributeEval özellik seçim yöntemleriyle de niteliklerin önem sıralamasını tespit etmişlerdir. Veri setinde gerekli veri ön işleme adımlarını ve standardizasyonları yaparak %20 test, %80 eğitim verisini ayırıp toplam beş farklı sınıflandırma algoritmalarıyla ve nitelik seçme yöntemleriyle ayrı ayrı değerlendirmişlerdir. Değerlendirme kriterleri olarak; f-ölçütü, doğru sınıflandırma oranı ve kappa istatistiği tercih etmişlerdir. Çalışmanın sonunda Çok Katmanlı Algılayıcı %90 doğru sınıflandırma oranı elde etmişlerdir (Potur ve Erginel, 2021).

Çoşar ve Deniz çalışmalarında üç adet ML kullanılarak kalp hastalıklarının tespitini yapmışlardır. Kaggle web sayfasından elde edilen on bir özniteliğe sahip farklı veri setinin birleşimi ile 918 örneklemden oluşan Kalp Yetmezliği Tahmin Veri Seti elde etmişlerdir. Çalışmalarında ML algoritmalarından KNN, LR ve RF algoritmaları kullanmışlardır. En iyi sonucu %88 doğruluk oranı ile RF sınıflandırma algoritması sonuç verdiğini bulmuşlardır (Çoşar ve Deniz ,2021).

Bektaş ve Babur çalışmalarında veri seti olarak 24482 niteliğe sahip 97 meme kanseri hastasına ait Kent Ridge 2 mikro dizi kullanmışlardır. Özniteliklerin çok fazla olması ve daha doğru karar verebilmek adına çalışmalarında ilgisiz özniteliklerin çıkarılması adına WEKA programında yer alan Correlation -based Feature Subset Selection ve bestFirst yöntemleri tercih ederek sınıflandırmaya etkisi yüksek 139 tane öznitelik belirlemişlerdir. Sınıflandırma algoritmaları olarak SVM, RF, K-Star algoritması ve seçimli algılayıcı Sinir Ağı Yöntemi kullanılmıştır. Sınıflandırma sonucunda oluşturulan modellerin performans değerlendirmelerini ise Hata Matrisi (Confusion Matrix) ve Alıcı İşletim Karakteristiği (Receiver Operating Characteristics) yöntemlerini tercih etmişlerdir. Oluşturulan modellerdeki performans metrik değerlerinin doğruluk ve geçerliliğini arttırmak adına çapraz doğrulama tekniği kullanmışlardır. Sonuç olarak çalışmalarında 24482 öznitelik ve 97 örneklem için en iyi

doğruluk oranı SVM algoritmasında %67.01 iken, öznelik seçimi sonrasında 139 adete düşen veri seti üzerindeki en iyi doğruluk oranı RF algoritması ile %90.72 olarak hesaplamışlardır (Bektaş ve Babur, 2016).

Al-Jada yüksek lisans tez çalışmasında EKG sinyallerini kullanarak kalp ataklarının tespitini gerçekleştirmiştir. Bu atakların hasta birey, sağlıklı veya hastalığı tespit edilmemiş bireyler üzerinde 3 ayrı sınıflandırma oluşturmuştur. Verilerin tespitini Oto Korelasyon ve Korelasyon yöntemlerini tercih etmiştir. Verilerin simülasyonu için zar atma yöntemi kullanılmış ve bu zarların birden fazla atılarak topluluk ortalaması ile zaman ortalaması elde etmiştir. Sonuç olarak oluşturulan modeller ile %98 başarı elde etmiştir (Al-Jada, 2018).



3. MATERYAL VE YÖNTEM

Bu başlık altında kalp yetmezliğinin tıp alanında teşhisi için kullanılan; röntgen, EKG ve çeşitleri, kan testi, New York efor testi ve sınıflandırmaları, Amerikan Kalp Derneği efor testi ve evreleri hakkında bilgi verilmiştir. Oluşturulan ve hazır veri setleri üzerinde analiz yapımında kullanılan makine öğrenme algoritmalarından; KNN, DT, LR, NB, RF ve YSA algoritmaları işlenmiştir. Bu algoritmalarda kullanılmak üzere oluşturulan veri seti; eksik verilerin bulunması ve temizlenmesi, verilerin birleştirilmesi, verilerin standardizasyon ve normalizasyon edilmesi, özellik seçilmesi gibi veri ön işleme adımlarına tabi tutulmuştur. Çapraz doğrulama uygulamanın önemi ve değerlendirme ölçütlerinin açıklanması ile bölüm sonlandırılmıştır.

3.1. Kalp Yetmezliği ve Teşhis Yöntemleri

Teknolojinin çok hızlı gelişmesiyle birlikte etki ettiği alanda doğru orantılı olarak genişlemektedir. Bu durum insanlığın hizmetinde olan her bilim dalını yakından ilgilendirmiş ve alanlarında teknolojiyi yakalamak adına çaba sarf etmelerine vesile olmuştur kısaca insandan daha hızlı analiz ve sentez yapan bilgisayar yazılımlarının avantajlarından yararlanmaya itmiştir. Bu bilimlerden biri de sağlık alanıdır. Sağlık alanının hemen hemen her biriminde teknolojinin son noktası insan sağlığı için hizmet vermektedir.

3.1.1. Göğüs bölgesi röntgeni

Röntgen, radyasyon adı verilen X ışınlarının bir cismin içinden geçmesiyle oluşturduğu görüntüleme yöntemidir. Sanayide kullanımı yanında yaygın olarak sağlık hizmetinde kullanılan röntgen, insan ya da hayvanların organ, kemik hastalıkları veya araştırılan bölge hakkında bilgi edinmeye olanak sağlamaktadır.

Hücrelerin kontrolsüz şekilde bölünmesi ile oluşturdukları kitlelerin iyi veya kötü huylu olarak nitelendirilmesi için görüntüleme yapılması gerekmektedir. Kitlenin belli sınırlarının olması ve pürüzsüzlüğü iyi huylu olduğuna işaret ederken tam tersi durumda kötü huylu olduğunu göstermektedir. Hastanın tedavi edilme koşullarına doğudan etki eden bu faktör yine görüntüleme ile sağlanmaktadır.

Hastalıkların birçoğunun belirtileri birbirine benzemektedir. En doğru veriyi elde etmek yapılacak olan analizin doğruluk payını artıracığından görüntüleme bu alanda en güçlü veri setini oluşturmaktadır. Şekil 3.1 için görüntülenen akciğerin üzerinde bulunan anormalliklerin yarıçap, doku, pürüzlülük, yoğunluk fraktal boyut ve yoğunluk analizleri sonucunda kuvvetli ihtimal olarak kanser teşhisi konulmaktadır.



Şekil 3.1. Kanserli akciğer röntgeni (Delen ve ark., 2005)

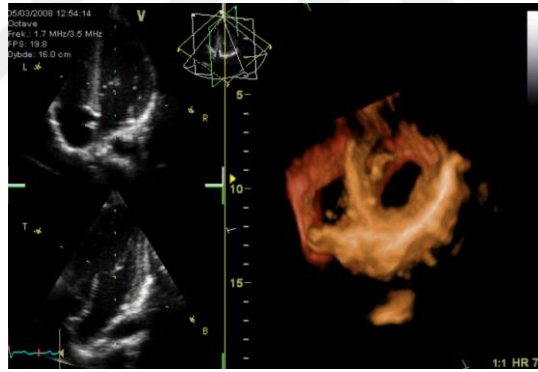
Kanser tüm dünyanın gerçeği olduğu için bu alanda oldukça fazla çalışma yapılmaktadır. Kanserle mücadelede en önemli materyallerden biri erken teşhis edilebilmesidir. Bu nedenle kanserli bireylerin röntgen verileri ortak bir alanda saklanıp veri işleme yapılarak yeni röntgen çekilen bireylerde kanser oranının ne olduğu belirlenmesi erken teşhis için önemli bir adım olmaktadır.

200 000'den fazla meme kanseri teşhisi için toplanan veri seti üzerinde karar ağaçları, yapay sinir ağları teknikleri ve lojistik regresyon kullanarak tahmin edilmiş ve bu analizlerde sırasıyla % 93.6, % 91.6 ve % 89.2 doğruluk sağlanmıştır (Delen ve ark., 2005). Başka bir meme kanseri çalışmasında ise çapraz doğrulama kullanılarak destek vektörleri makinası algoritmasıyla bu oran % 98.5 gibi yüksek bir doğruluğa çıkarılabilmektedir (Polat ve Güneş, 2007).

3.1.2. Ekokardiyografi

EKG yüksek frekansta ses dalgası üretebilen bir cihazdır. Bu cihazda yer alan ve prop adı verilen elektronik uçlarından çıkan ses dalgalarını insanın göğüs kafesinden geçtikten sonra kalbe ulaşmaktadır. Bu ses dalgası kalbin duvarlarından, kalp kapakçıklarından ve kaslarından geri dönerek diğer proba girmektedir. Bu yöntemle kalbin faaliyetleri izlenebilmektedir.

Normal olan kalbin ses dalgasına vereceği tepki bilindiğinden normal olmayan durumlarda ses dalgası başka şekilde yansımaktadır. Bu yansımadan faydalanılarak cihazda oluşan kalbin görüntüsünün normal olmadığı izlenebilmektedir. Kalp boşlukları, damar çapları, duvarların hareketleri, basıncı ve ejeksiyon fraksiyonunu büyük oranda araştırılabilmektedir. Şekil 3.2’de iki ve üç boyutlu EKG görüntüleri gösterilmiştir.



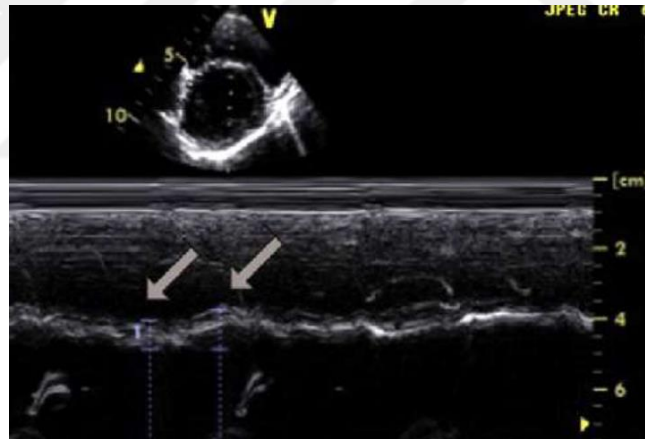
Şekil 3.2. Sol tarafta 2 boyutlu, sağ tarafta 3 boyutlu ekokardiyografi görüntüleri
(John Hopkins Medicine, 2022)

EKG cihazı yalnızca kalp organı değil böbrek, mide karaciğer gibi hayatı önem taşıyan organları da görüntülemektedir. Bu görüntüleme sonucunda şüphe barındıran teşhislerde yemek borusunda açılan delikten indirilen cihaz kalbi daha yakından inceleyerek en doğru sonuca gitmeyi sağlamaktadır. Bununla beraber sağlık alanının dışında endüstride basınçlı kap kontrolü, kapalı cihaz çalışma anormalliklerinin arıza analizi için kullanılabilir. Teknolojinin her geçen gün gelişmesiyle ekokardiyografi çeşitlenmiş ve görüntü işleme teknikleri ile net tanıları konulmasını kolaylaştırmıştır.

Kontrast EKG: Bu yöntem, kalp kapakçıklarındaki anormalliklerin görselleştirilebildiği bir yöntemdir. EKG esnasında damardan verilen periferik hava kabarcıklarının dinlenme ve öksürük anında basınç farkı kaynaklı hareketleri sayesinde görüntülenen kalbin anormalliklerini ortaya koyulabilmektedir.

Bu yöntem kalp anormalliklerinin yanında hipertansiyon gibi diğer etkenlerinde eşlik ettiği tanılarda güvenilirliği azalmaktadır. Aynı şekilde gebelerde de görüntüleme yöntemi olarak kullanılmamaktadır (Türkyılmaz, 2023).

M Mode EKG: Kalbe enjekte edilen hava kabarcıklarının basınç sebebiyle paradoksik hareketleri izlenerek oluşturulan görüntüleme yöntemidir. Kalp duvar kalınlığı, boşlukların boyutu gibi önemli parametreler hakkında bilgi edinilmektedir. Zaman ve hareket grafiği olarak düşünülen bu görüntüleme tekniğinde doğru sonuç elde edebilmek için kalbin sol tarafı, mitral kapakçık ve aort tarafından kesitler elde edilmelidir. Şekil 3.3’de kalbin M-Mode EKG görüntüsü gösterilmiştir.



Şekil 3.3. Kalbin M-mode ekokardiyografi görüntüsü
(Türkyılmaz, 2023)

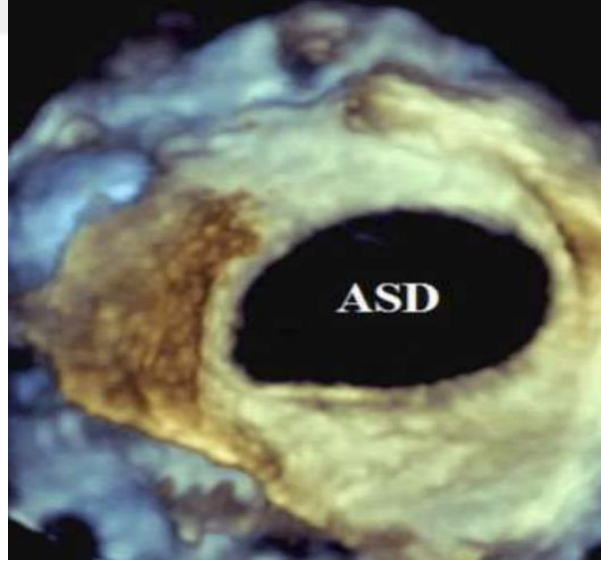
Doppler EKG: Doppler EKG yöntemi ASD’ nin (Atriyal Septal Defekt) yanında, boyut ,akış hızı ve yönü ayrıca hasara karşı uygulanacak tedavi yöntemiyle ilgili önemli detayları da göstermektedir. Önceki çalışmalarda renkli doppler yönteminin gerçeğe en yakın ASD çapını belirlediği bilinmektedir (Morimoto ve ark., 1987). Hastalığı tespit etmenin yanı sıra cerrahi onarım sonrasında ameliyatın başarısını ortaya koymak için de kullanılmaktadır. Şekil 3.4’de renkli akım doppler kullanımı gösterilmiştir.



Şekil 3.4. Renkli akım doppler kullanımı
(Türkyılmaz, 2023)

3 Boyutlu EKG:

Teknoloji gelişmesiyle doğru orantılı büyüyen görüntüleme teknolojisinin önemli adımlarından biri de şüphesiz 3 boyutlu görüntülemedir. Bu görüntüleme yöntemi hastalığın ve hasarı en net şekilde ortaya koyan yöntemlerinden biridir. Diğer yöntemlerde şüphe olan noktaların ortadan kaldırılması için 3 boyutlu ekokardiyografiye başvurması da bunu ispatlamaktadır (Acar, 2004). Şekil 3.5’de kalp kapakçıklarındaki açıklığın üç boyutlu ekokardiyografisiyle belirlenmesi gösterilmiştir.



Şekil 3.5. Kalp kapakçıklarındaki açıklığın 3 boyutlu ekokardiyografisiyle belirlenmesi
(Türkyılmaz, 2023)

ASD, kalp kapakçıklarında oluşan açıklığı ifade etmektedir. Bu açıklık oksijen bakımından daha zengin olan temiz kan ile kirli kanın kalpte karışması ve nispeten kirli kanın vücutta dolaşma durumudur. Bu durum halk arasında delik kalp olarak anılmakla birlikte genellikle cerrahi müdahale ile tedavi edilmektedir.

3.1.3. Kan testi

İnsanlar varoluşundan bu yana sürekli bir arayış ve araştırma davranışlarında bulunmuşlardır. Bir insanın hayatı ve yaşadıkları diğer insanlara örnek olarak akıllarda ya da yazılı kaynaklarda saklanarak bir sonraki nesillere aktarımı yapılmış ve her geçen gün geliştirilmiştir. Bu gelişmenin en büyük adımlarından biri de kan testi uygulamalarıdır. Elektronik cihaz teknolojisinin gelişmesiyle birlikte sağlık sektöründeki gelişme göz ardı edilemeyecek kadar büyüktür. Bu teknolojilerden biri de kan testleri cihazlarıdır. Kanın içerisindeki bileşiklerin değişiminden kaynaklanan rahatsızlığı saptamada önemli rol oynamaktadır. Hemen hemen her alanda kan testine başvurulmasıyla ilgili uzman, hastalık adına net saptamalar yapabilmektedir. Bu yüzden kan testlerinin sınıflara ayrılması bu konuda işlemleri kolaylaştırmaktadır.

Özellikle lösemi ve anemi hastalıklarının belirlenmesinde hematolojik tahliller kullanılmaktadır. Bu kan testi, kanın içerisinde bulunan hücre sayısını ve oranını tespit edebilmektedir. Kırmızı ve beyaz kan hücresi sayısı, hemoglobin, trombosit sayısı ve büyüklüğü analizi bu hastalıklar hakkında bilgi verebilmektedir. Bunların yanında genel sağlık durumu belirlenmesinde de kan sayımı yapılması veya takip edilmesi erken teşhis durumlarını da beraberinde getirmektedir.

İnsan vücudunda oluşan aksaklıklar beraberinde bazı belirtileri meydana getirmektedir. Bu belirtiler doğrultusunda yapılacak testlerde hedef küçültmesi yapılarak en net sonuca götürebilmektedir. Kan pıhtılaşması sorunu yaşayan bir hastada fibrinojen kan testi istenmekte ve bu teste fibrinojen seviyesindeki bir artış kalp hastalığı ve inme riskinin tahmininde bulundurmaktadır. Fibrinojen seviyesini normal koşullara indirilmesi için spor, sağlıklı beslenme ve düzenli yaşam şartları gerekmektedir.

Kanda serbest dolaşan kolesterol ve yağları saptamak için lipid kan testi yapılmaktadır. Bu kan testi özellikle kalp rahatsızlığı olan hastalara mutlaka yapılmalıdır. Test sonucu normal olmayan bireylerde kronik kalp rahatsızlığını tetiklediğinden belirli aralıklarla bu testi yaptırması gerekmektedir.

Kan, organlara oksijen taşıyan ve organların da çalışması sonucu ortaya çıkardığı atıkları toplayıp temizlemeye yardımcı olan bir vücut sıvısıdır. Organ veya organların oluşturduğu insan sisteminde meydana gelen aksamaların en belirgin belirtileri kanda mevcut olmaktadır. Bu yüzden kan testi oldukça fazla sınıfa ayrılmıştır. Hipofiz bezi salgıları, böbrek üstü bezi salgıları, cinsel bulaşıcı hastalıkların belirtisi,

protein sentezi sorunları, hamilelik durumu, enzimlerin seviyelerinin belirlenmesi, vitamin ya da mineral eksiklikleri ve enfeksiyon gibi oldukça fazla alanda kendini etkin bir şekilde göstermektedir.

3.1.4. New York efor testi

Kalp yetersizliği, kalbin yerine getirmesi gereken görevleri yapamaması durumudur. Bu durumda fonksiyonel potansiyel ve kapasiteyi değerlendirmek için New York Kalp Birliği sınıflandırması yöntemi kullanılmaktadır. Hastanın öyküsünün kullanıldığı bu sınıflandırma türünde semptom oluşabilmesi için efor seviyesine göre sınıflandırılarak incelenmektedir.

Koç 2006 yılında yaptığı çalışmada New York Kalp Birliği tarafından ortaya konulan spesifik etkinlik ölçülerini ve fonksiyonel sınıflandırma kategorisini (Sınıf I, Sınıf II, Sınıf III, Sınıf IV) aşağıdaki gibi özetlemiştir (Koç, 2006).

Sınıf I: Klinik değerlere göre hasta olup gerçekte fiziksel olarak aktivitelerde bir kısıtlama veya noksanlık hissedilmemesi, nefes nefese kalma, hemen yorulma durumları gözlenmez. Metabolik Equivalents (METS) değeri 7 ye eşit ve daha düşük enerji gerektiren etkinliklerde hastalar rahatlıkla yapabilirler. Örneğin; futbol, basketbol, kayak, 5 mph hızında yürüme gibi.

Sınıf II: Basit bir fiziksel etkinliklerde yorulma ve nefes darlığı oluşması gözlenen kalp yetmezliği hastalarında fiziksel etkinliklerde hafif kısıtlı olarak kategoridendirler. METS değeri 5 ve altındaki enerji ihtiyacı etkinlikleri (Bahçe işleri, dans, cinsel aktivite, 4 mph hızında yürüme) yapabilirler. Bu sınıfa ait hastalar 7'ye eşit ve 7'den büyük METS enerjiye ihtiyaç gerektiren etkinlikleri yapamazlar.

Sınıf III: Belirgin derecede fiziksel etkinlikleri yapmada zorlanma yaşayan kalp hastalarıdır Dinlenme sırasında herhangi bir belirti göstermeyen ama hastalığa sahip kişilerdir. Bu durumdaki hastalara asemptomatik tanımı yapılmaktadır. Sıradan bir etkinlikte bile yorulma ve nefes darlığı şikayetleri bulunmaktadır. 2 ve altındaki METS enerjisi gerektiren etkinlikleri (Basit ev işleri, Duş alma, 2.5 mph hızında yürüme gibi) rahat yapabilirler. 5 ve üstü METS enerjiye ihtiyaç gereken etkinlikleri yapamazlar.

Sınıf IV: Hastanın dinlenme zamanında bile kalp yetmezliği semptomları görülmektedir. Temel yaşam fonksiyonlarını bile yaparken rahatsızlık baş göstermektedir. 2 ve üstü METS enerji gerektiren hiçbir etkinliği kesinlikle yapamazlar.

Amerikan Kardiyoloji Koleji/Amerikan Kalp Derneği Uygulama Kılavuzları Çalışma Grubunun 1997 yılındaki Egzersiz Testi Yönergelerine göre efor testi hastanın fiziksel özellikleri göz önünde bulundurulması gerekmektedir, Sağlıklı insanlarda Bruce efor özellikleri ile test yapılmalıdır, Harcanan eforun şiddetini 3 dakikada bir artırılması ile hastanın en yüksek performansı elde edilmesi gerekmektedir.

Efor testinden kalpteki kan çıkışının basıncın çok fazla değişmeden miktarın artışı beklenmektedir. Kardiyak kan çıkış hızını dinamik eforla artırılması mümkün olmakta ve bunun için bisiklet veya 6 dakikalık yürüyüş yapılması istenilen testi sağlamaktadır (Bernard R.Chaitman, 2005).

3.1.5. Amerikan kalp derneği efor testi

Kliniğe başvuran hastaların özellikleri ve doktor tedavi yöntemlerini analiz edebilme ve yönlendirebilmektedir (Micheal R. Bristow, 2001).

Koç 2006 yılında yaptığı çalışmasında Amerikan Kalp Derneği tarafından ortaya konulan sınıflandırma (Evre A, Evre B, Evre C, Evre D) aşağıdaki gibi özetlemiştir.

Evre A: Kalp yetmezliğinin baş gösterme olasılığı yüksek olmasına rağmen kalbin yapısında bir bozukluk tespit edilemeyen hastalardır. Örnek vermek gerekirse kronik arter hastalığı, kardiyotoksik, sistemik yüksek tansiyon.

Evre B: Kalbin yapısal olarak bozulmaların başlaması yanı sıra KY için gerekli bulgu ve semptomların oluşmaması durumudur. Yüksek risk ile kalp yetmezliğinin meydana gelmesi beklenir. Örneğin; eski miyokard infarktüsü, kardiyak dilatasyon, asemptomatik kapak hastalığı.

Evre C: Yapısal bozukluk içeren kalp hastalığı ile birlikte öncesinden veya şimdi dahi semptom görülen kalp yetmezliği durumlarıdır. Hasta olup belirti göstermeyip semptomlar için tedavi süreci başlamış hastalar, nefes darlığı ve çabuk yorulma yaşayan hastalar bu sınıfa dahildirler.

Evre D: Kalp yetmezlik seviyesi ileri düzeye ulaşmış ve yapılan tüm tedavilere rağmen istirahat zamanında dahi KY semptomlarını yaşayan hastalardır. Sık olarak hastane yatışı yapan hastalar, kalbi destekleyici cihaz kullanan hastalar örnek verilebilir.

KY rahatsızlığına sahip bireylerin büyük çoğunluğunda sol ventrikül dilatasyonu görülürken EF değerinin de %40 dan düşük olduğu belirlenmiştir. Bunun yanında hastaların %60'ında New York Kalp Derneği sınıflandırması durumunda ikinci ve üçüncü kategorisindeki semptomlar görülmektedir. Tedavilerin uygulanmasına rağmen kalp yetersizliği olan hastaların yaklaşık beşte biri hastalık belirtileri devam edebilmektedir.

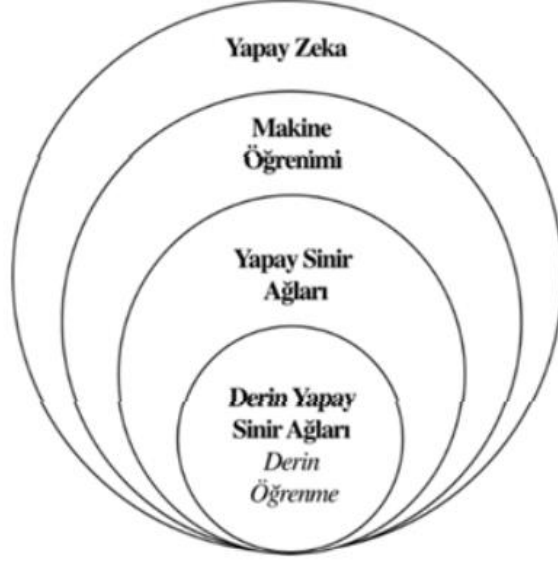
3.2. Makine Öğrenmesi

1950 yılında Alan Turing tarafından “Makineler düşünebilir mi?” sorusunun “Bilgisayar Makineleri ve Zekâ” isimli makalesiyle ortaya atılmıştır. Ortaya atılan bu düşünce ile birçok araştırmacı makine öğrenmesinin mümkün olabileceği üzerine çalışmalar yapmıştır. Arthur Samuel tarafından makine öğrenmesi kullanılarak geliştirilen dama oyunu ilk uygulama olarak kabul görmüştür (Doğançay, 2023). Arthur Samuel makine öğrenmesini 1959 yılında bilgisayar teknolojilerinin doğrudan kodlanmadan otonom olarak öğrenme olayını gerçekleştiren bir alandır şeklinde tanımlamıştır (Bilekyiğit, 2022).

Yapay zekâ kavramının bir alt dalı olan makine öğrenmesi temelde bir veri setinin ön işleme adımlarıyla hazırlanması, birleştirilmesi, dönüştürülmesi işlemlerinden sonra çeşitli algoritmalar ile veri setindeki datalar üzerinde istatistiksel ve matematiksel işlemler sonrasında sonuçların tahmin edilmesi ve karar verme noktasında destek sistemi oluşturulması süreci olarak tanımlanabilir (Erdaş, 2022).

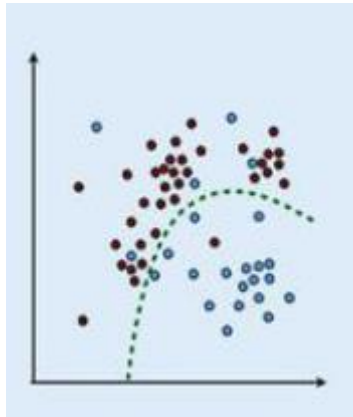
Makine öğrenmesi sayesinde günümüzde bir arabanın hasar durumu, kilometre değeri, boya değişen, marka model, üretim yılı gibi bilgilerin sisteme girilmesiyle aracın otomatik fiyatlandırması yapılabilmektedir. Yine satılık bir evin oda sayısı, metre kare bilgisi, il ilçe mevki bilgileri, bina yaşı, cephe bilgisi vs. bilgileri ışığında kiralık ya da satılık fiyatlarının belirlenmesi yapılabilmektedir. Örnekleri çoğaltmak gerekirse; kredi not hesaplama, spam mail tespiti, metinsel bir ifadeden karakter analizi, dolandırıcılık tespiti, sağlık alanlarında erken teşhis, satış tahmini şeklinde örnekler verilebilir. Şekil

3.6’da yapay zekadan derin öğrenmeye kadar birbirleri arası ilişkisi kümeler halinde gösterilmiştir.



Şekil 3.6. Yapay zekâ ve alt kolları (Bingöl ve ark., 2020)

Denetimli öğrenme, veri seti içinde yer alan niteliklerden bir tanesi o verinin etiketini oluşturmaktadır. Bu etiket bazen numerik bazense alfa numerik olabilmektedir. Örneğin; ölme olayı, hasta olup olmama durumu, satılması beklenen tahmini dondurma sayısı. Denetimli öğrenme ile veri setindeki etiketli verilerden belirli oranda kullanılarak bir model oluşturulur. Oluşturulan model sonrasında hiç kullanılmayan ve etiketleri bilinen diğer veriler ile test işlemi yapılır. Doğru tahmin oranı hesaplanarak doğruluk, güvenilirlik gibi istatistiki değerlendirme ölçütleri hesaplanır. Şekil 3.7’de denetimli öğrenme ML algoritması gösterilmiştir.



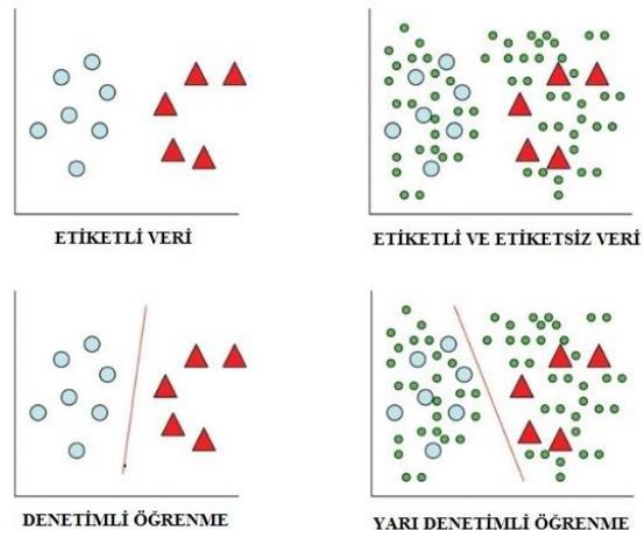
Şekil 3.7. Denetimli öğrenme (Bilekyiğit, 2022)

Denetimsiz öğrenme, verilerin bir sınıflandırma çıktısı veya etiketi olmayan öğrenme yöntemidir. Bu yöntemde daha çok benzer verilerin kendi içlerinde kümeleme ve birbirleri arası ilişkileri analiz edilir. Denetimsiz öğrenmede model oluşturmadan kümeleme yöntemiyle çözüm üretilebilmektedir. Örneğin; bir alışveriş merkezinin yapmış olduğu ankete göre müşteri profillerine özel kampanyalar sunması. Şekil 3.8’de denetimsiz öğrenme ve kümeleme örneği verilmiştir.



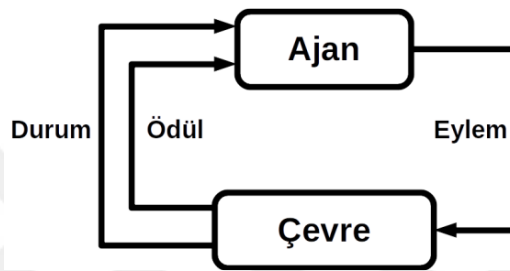
Şekil 3.8. Denetimsiz öğrenme - kümeleme örneği (Sarıman, 2011)

Yarı denetimli öğrenme yöntemi, verilerin bir bölümünde çıktı etiketi bulunmazken bir bölümündeki verilerde hem girdi hem de çıktı verileri bulunan makine öğrenme yöntemidir. Denetimli öğrenme ile denetimsiz öğrenme algoritmalarının birlikte kullanıldığı algoritmalarlardır. Bazı sağlık verilerinin tutulduğu veri setlerinde her niteliğe ait verinin ve çıktının olmaması örnek verilebilir. Şekil 3.9’da yarı denetimli öğrenme verilmiştir.



Şekil 3.9. Yarı denetimli öğrenme (Bilekyiğit, 2022)

Pekiştirmeli/Takviyeli öğrenme yöntemi diğer öğrenme yöntemlerine göre daha yavaş çalışmaktadır. Bunun gerekçesi hazır bir veri seti ile model oluşturulmazken, algoritma dışarıdan gelen dönütlerle öğrenmeyi gerçekleştirir. Buna en iyi örnek olarak labirent içerisinde çıkmaya çalışan robot örneği verilebilir. Düzenek içerisindeki robot alternatif yolları deneyerek labirentten çıkmaya çalışır. Çıkmaz yola girdiğini algılayıp tekrar geri dönerek gidilen yolun yanlış olduğunu öğrenir ve diğer yolları deneyerek sonuca ulaşmayı dener. Bu sebeple diğer algoritmalara göre daha yavaştır. Şekil 3.10'da pekiştirmeli ve takviyeli öğrenme gösterilmiştir.



Şekil 3.10. Pekiştirmeli/takviyeli öğrenme (Bilekyiğit, 2022)

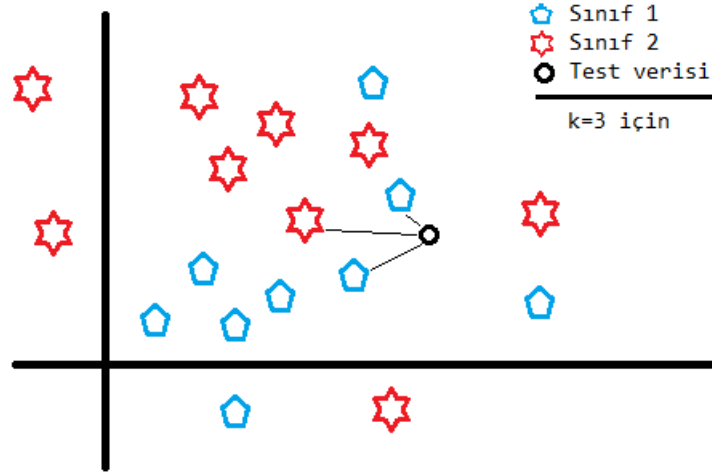
3.2.1. K En Yakın Komşu algoritması

Denetimli öğrenme türlerinden olan KNN algoritması temelde verilerin birbirine olan uzaklığını esas alarak sınıf belirlemede kullanılır. Uzaklık belirleme çeşitleri en yaygın olanı Öklid olmakla birlikte Manhattan, Minkowski mesafe hesaplama yöntemleridir.

1967 yılında T.M. Cover ve P.E. Hart çalışmalarında KNN algoritmasını kullanmışlardır. Verilerin birbirine olan uzaklıklarıyla en çok hangi sınıfa komşuysa o verinin ilgili sınıfa ait olduğunu söylemişlerdir (Gärtner ve ark, 2004).

Veri setine ait örneklerin her biri diğerlerine olan uzaklığı hesaplandıktan sonra seçilen k değerine göre komşuları belirlenir. Sınıflandırılacak verinin en çok hangi sınıftan komşusu fazlaysa bu veri de o sınıfa dahil edilir. K değeri çift olması durumunda sınıflandırmadaki komşuluk sayılarının eşit olması durumunda sınıflandırma gerçekleşmemesinden dolayı komşuluk sayısı tek seçilmelidir. Tek olarak verilen komşuluk değerlerine göre oluşturulan modelin doğruluk sonuçları karşılaştırılarak en iyi sonucu veren k değerine göre bu algoritmayı modellemek en

doğrusu olacaktır. Komşuluk değeri 3 için örnek bir veri setinde test verisinin sınıflandırılması Şekil 3.11’de gösterilmiştir.



Şekil 3.11. Knn algoritması sınıflandırma gösterimi

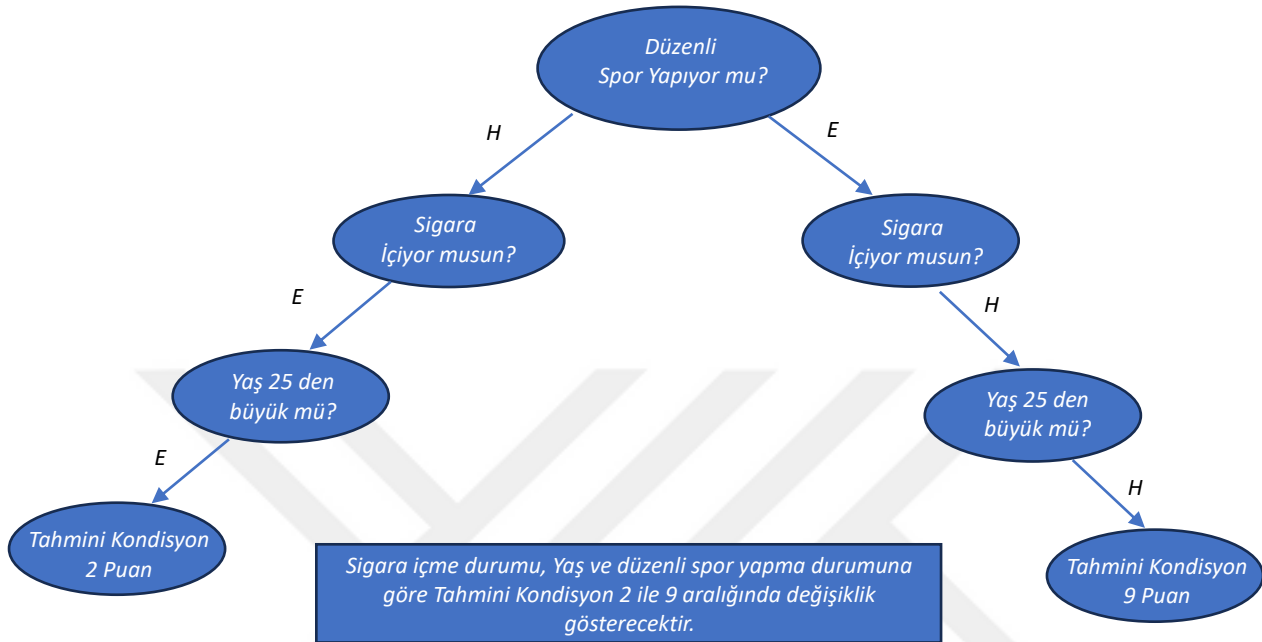
KNN algoritmasının uygulaması ve veri setinin eğitimi kolay olması avantajları arasındayken bu algoritma için büyük verilerde her birinin uzaklığının hesaplanması sistemin yavaş çalışmasını da dezavantajı olarak karşımıza gelmektedir.

3.2.2. Karar Ağacı algoritması

Karar ağaçları (DT) kök, düğüm, yapraklardan oluşan ağaç görünümündeki karar yapısıyla sınıflandırma yapmamızı sağlayan veri seti ile kolay uyum sağlayan, güvenilir ve oluşturma maliyeti az olan, günümüzde hemen hemen en güvenilir sonuç veren denetimli makine öğrenme algoritmasıdır. Örneğin; hava durumunun güneşli, yağmurlu, rüzgârlı olma durumlarına göre futbol oynanıp oynanamama durumunun tespiti verilebilir. Derse devamsızlık sayısı, derse katılım düzeyi, ödevleri yapma sayısı gibi bilgilerden yola çıkarak sınav notunun tahmini. Örneklerden de anlaşıldığı gibi bağımsız değişken olan havanın durumu (güneşli, rüzgârlı vs.), devamsızlık sayısı gibi verilerden yola çıkarak bağımlı değişken olan top oynama ya da oynamama gibi niteliğin sınıflandırılması üzerine kurulmuştur. Karar ağaçlarında bağımlı değişkenler başlangıç(kök) ve bitiş(yaprak) noktalarında kullanılırken, dallanmanın olduğu noktalarda ise bağımsız değişkenler bulunur.

DT algoritmasının tarihsel gelişimine bakıldığında Charles J. Stone ve arkadaşlarının 1987 senesinde yazmış oldukları “Sınıflandırma ve Regresyon Ağaçları”

kitabında açıklamışlardır. 1986 yılında ise J.R. Quinlan karar ağacı yapısına ID3 ve 1993 yılında ise “Makine Öğrenimine Yönelik Programlar” isimli kitabında C4.5 yöntemlerini kitaplarında yer vermiştir (Lee ve ark., 2001). Şekil 3.12’de örnek karar ağacı algoritması verilmiştir.

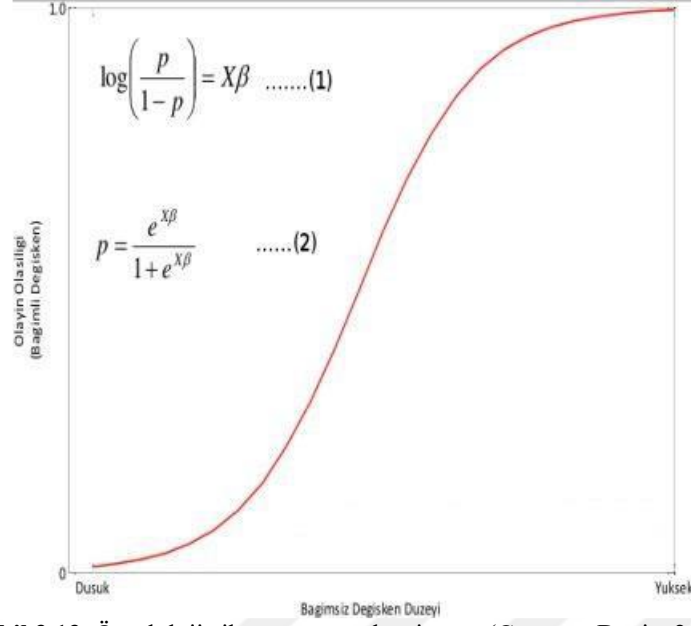


Şekil 3.12. Örnek karar ağacı algoritması

3.2.3. Lojistik Regresyon algoritması

Lojistik Regresyon (LR) algoritması veri setinde yer alan bağımsız değişkenlerden oluşturulan model ile bağımlı olan sonuç değişkenini tahminlemede kullanılan ve ayrıca temelde “doğru – yanlış”, “ölü – sağ”, “hastalıklı – sağlıklı”, “pozitif – negatif” gibi ikili sınıflandırmalarda kullanılmaktadır. Tahmin edilen değer kategorik olabildiği gibi sayısal değerlerde olabilir. Oluşturulan LR modeli her niteliklerin ağırlıklarına denk gelen regresyon katsayısına göre hesaplanarak eğitilir.

LR analizinde sadece sınıflandırma yapılmakla birlikte isminden de anlaşıldığı üzere regresyon analizi de yapılabilmektedir. Bağımlı değişkenlerin her birine Şekil 3.13’de verilen sigmoid formüllerin sonucunda 1 ve 0 aralığındaki değerler hesaplanır. Hesaplanan bu sonuca göre sınıflandırma yapılır (Görmez, 2021). Şekil 3.13’de örnek lojistik regresyon algoritması verilmiştir.



Şekil 3.13. Örnek lojistik regresyon algoritması (Çoşar ve Deniz, 2021)

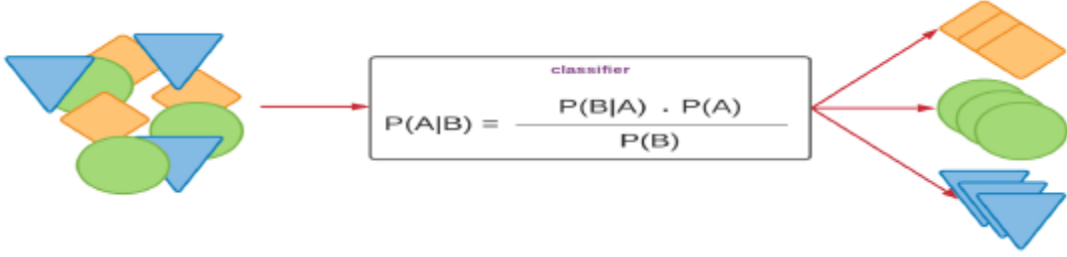
Bu algoritmada dikkat edilmesi ve unutulmaması gereken şeylerden biri bağımsız değişkenlerin birbirinden bağımsız olması gerekir yani bağımsız değişkenlerin birbirleri arasında yüksek seviyede korelasyonun olması eş doğrusallık oluşturacaktır.

3.2.4. Naive Bayes algoritması

Bu algoritma olasılık temelli çalışan Thomas Bayes tarafından 18. Yüzyıllarda ortaya atılan bir ML modelidir. Naive Bayes (NB) özniteliklerin her birini olasılık değerlerine göre bir sınıfa dahil eder. Veri setindeki örneklem sayılarının çok ve karmaşık olduğu durumlarda problem çözümü için yaygın olarak tercih edilmektedir.

NB modelini oluşturmak kolay ve modelin öğrenmesi hızlıdır. Örneklem sayısının az olabildiği veri setlerinde dahi iyi bir performans gösterebilmektedir. Verilerin fazlaca ve karmaşık olduğu veri setlerinde de özellik bağımsızlık varsayımı ile birçok algoritmaya göre iyi sonuçlar verebilmektedir. Bu avantajların yanı sıra sınırlılıkları da bulunmaktadır. Örnekler içerisinde eksik değere sahip verileri göz ardı eder. Sonucu olumsuz yönde etkileyebilecek aykırı değerleri işleme alabilir. Bu sebeple veri ön işleme adımlarını iyi tektik etmek gerekir.

Gelen bir elektronik postanın gereksiz mail olup olmadığı, bir müşterinin belirli bir kredi tutarını ödeme olasılığı, kredi kartı harcamalarından yola çıkarak kart sahibinin cinsiyet sınıflandırılması gibi örnekler alışıla gelmiş NB model örnekleridir. Şekil 3.14'de NB sınıflandırmasının formülü ve Şekil 3.15'de grafiği verilmiştir.



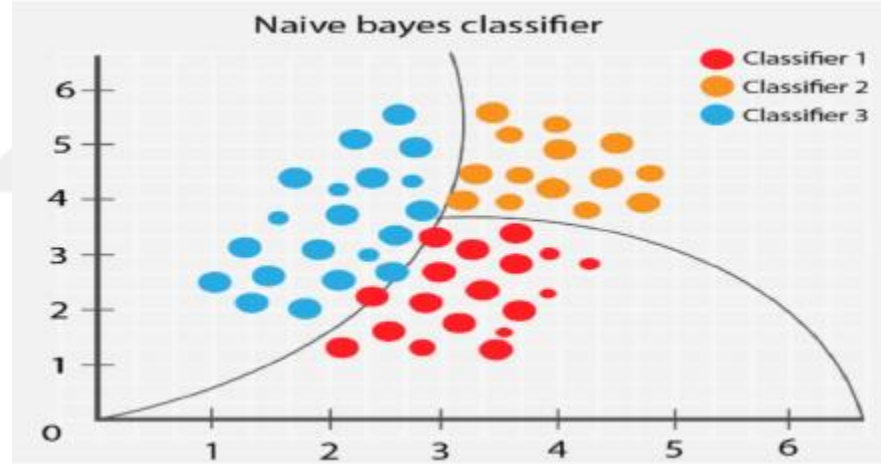
Şekil 3.14. Naive bayes sınıflandırmasının formülü (Karaj, 2023)

$P(M)$: M olayının gerçekleşme olasılığı

$P(K)$: K olayının gerçekleşme olasılığı

$P(M|K)$: K olayı gerçekleştiğinde M olayının gerçekleşme olasılığı

$P(K|M)$: M olayı gerçekleştiğinde K olayının gerçekleşme olasılığı



Şekil 3.15. Örnek naive bayes algoritması (Karaj, 2023)

3.2.5. Rastgele Orman algoritması

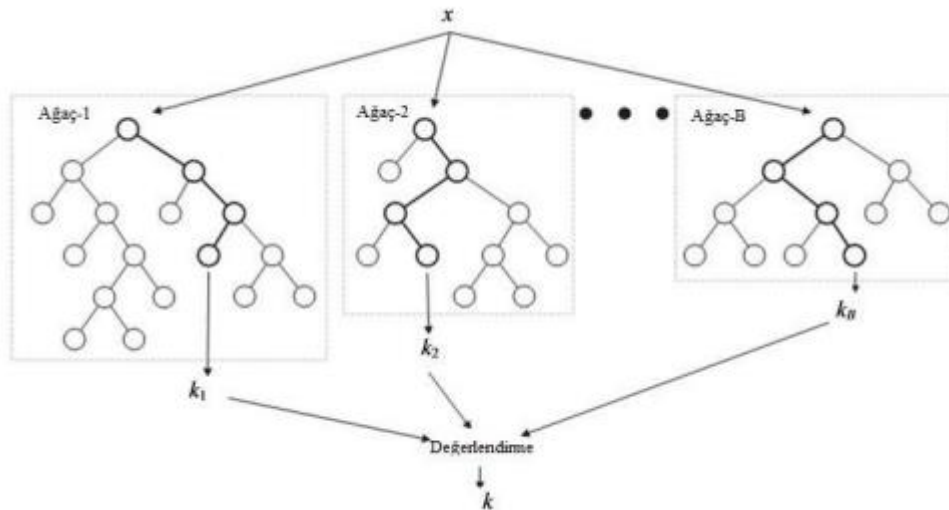
Karar ağaçlarının bir araya gelerek oluşturduğu, büyük verilerin analizinde çokça tercih edilen, yüksek oranda doğrulukla tahminde bulunan güçlü bir algoritmadır. Bu algoritma ile sınıflandırma yapılabildiği gibi regresyon analizlerinde de ML problemlerinde tercih edilmektedir.

Yukarıdan (Kök) aşağıya (düğüm ve yapraklar) doğru karar verme yöntemine göre yönlenen bu denetimli öğrenme algoritmasında Entropy, en küçük kareler ve Gini

yöntemleri kullanılmaktadır. Veri setinde yer alan özniteliklerin çıkarımları doğrultusunda oluşturulan karar yapılarının öğrenilmesiyle çıktıda yer alan hedef özniteliğinin tahmin edilmesini amaçlanmaktadır (Karaduman, 2023).

Veri setinde yer alan örneklerin sınıf sayısının fazla olup örneklem sayısının buna oranla yetersiz olması durumunda oluşturulan karar ağacı modelinde oldukça fazla dallanma yaşanabilmektedir. Küçük veri setlerinde dallanmanın fazla olması durumlarında ise modelde aşırı öğrenme olabileceği için karar ağacında budamaya gidilebilmektedir. Modelin ezbere gitmesinin önlenmek için ağacın dalları budanmaktadır (Görmez, 2021).

RF algoritmasını kullanarak oluşturulan modeller kuralları net, dallanmaları açık olması nedeniyle çok kararlı çalışmaktadırlar. Modeli oluşturma maliyeti düşük, kolay anlaşılabilen hem sınıflandırma hem de regresyon analizinde kullanılabilme gibi avantajlara sahiptir. Veri seti üzerinde örneklerin az olması sınıflandırmanın çok olması durumlarında aşırı karmaşık ağaçlar üretilmesi sonucunda komplike yapılar karşımıza dezavantaj olarak gelmektedir. Şekil 3.16'da örnek rastgele orman algoritmasının akış diyagramı ile gösterilmiştir.



Şekil 3.16. Örnek rastgele orman algoritması (Sel, 2020)

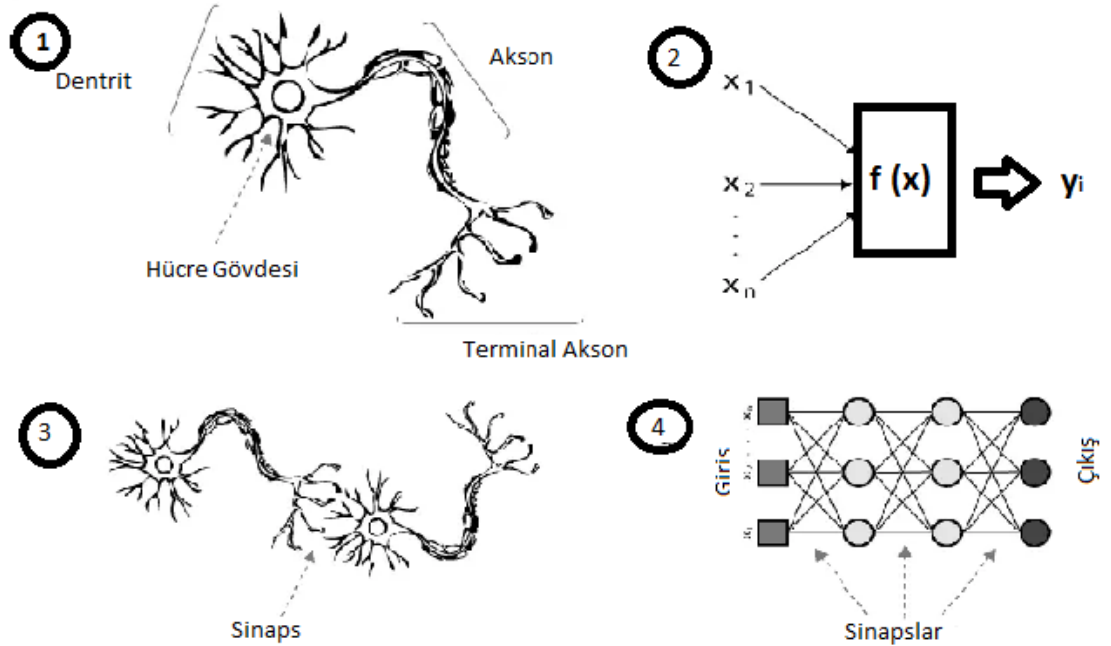
3.2.6. Yapay Sinir Ağları algoritması

İnsan beyninin daha öncesinden öğrendiği ve daha sonrasında hatırladığı, tümevarım ile genelleme yapabildiği ya da tümdengelim ile kestirim yapabildiği, öğrenme algoritmasından esinlenerek matematiksel model oluşturularak ortaya çıkan yapılara yapay sinir ağları denilmektedir (Öztürk, Ergin Şahin, 2018).

Yapay sinir ağıları 1943 yılında Pitts ve MCCulloch hücre modelini yapay sinir tanımını yaparak geliştirmişlerdir. Çalışmalarında nöronlar sabit değere sahip lojik değerler halinde modellenmiştir. Makine öğrenmenin ve yapay zekâ alanında çalışmaların yoğunlaşmaya başladığı 1949 yıllarında Hebb sinir ağlarının öğrenme sürecini ait kuralları ortaya koymuştur (Şen, 2004).

Veri setindeki girdi değerleri çıktı değerine birer ağırlık oranı bulunmaktadır. Bu oranlar girdi değerlerinin çıktı değerine ulaşmada etkin rol üstlenmektedirler. Yapay sinir ağlarının genel kullanım türleri: sınıflandırma, kümeleme, vektör sayısallaştırma, desen uygunluğu, fonksiyon yaklaşımı, tahmin yapmak, kontrol sorunları, en iyileme (optimizasyon), arama çalışmaları vs. olarak karşımıza çıkmaktadır (Şen, 2004).

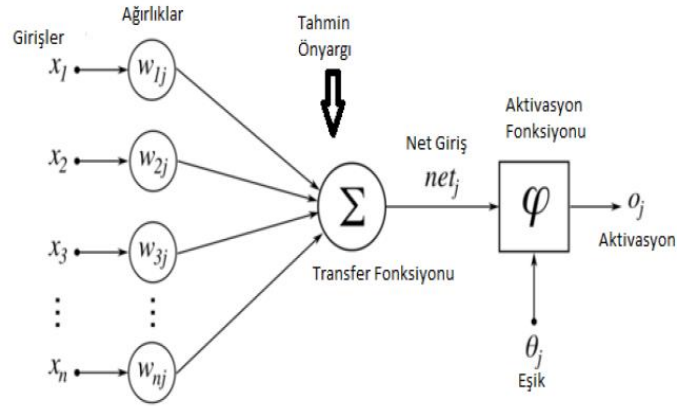
Yapay sinir ağları ile biyolojik sinir ağlarından esinlenerek ortaya atılan yapay sinir ağlarının eleman karşılıkları şu şekildedir. Toplama fonksiyonu olarak dentrit, nöron hücreleri işlemci elemanı, aksonlar yapay nöron çıkışı, transfer fonksiyonu hücre gövdesi, sinapslar ağırlıklar olarak karşılıkları eşleştirilmiştir. Şekil 17’de biyolojik ve yapay sinir ağları sunulmuştur (Öztürk, Ergin Şahin, 2018).



Şekil 3.17. Biyolojik ve yapay sinir ağları (Öztürk, Ergin Şahin, 2018)

Yapay sinir ağları eksik verilerle çalışma, doğrusal olmama, esneklik ve hata toleransı, genelleme, paralel çalışma, uyarlanabilirlik, çok sayıda değişken ve parametre kullanma gibi birçok özelliklere sahiptir. Şekil 18’de n adet verinin ağırlıkları çarpılarak

tüm verilerin değerleri toplanmaktadır. Transfer fonksiyonunda toplanan değerler net giriş fonksiyonu ve aktivasyon fonksiyonlarından geçerek sonuç elde edilir (Öztürk, Ergin Şahin, 2018).



Şekil 3.18. Yapay sinir hücresi (Öztürk, Ergin Şahin, 2018)

3.3. Veri Ön İşleme

Günümüzde elde edilen verilerin çokluğu ve çeşitliliği bu verilerin analiz edilerek bilgi formatına dönüştürülmesi gerektiği yadsınamaz bir gerçektir. Kâğıt ve kalem ile bu verilerin analiz edilmesi ise neredeyse imkansızdır. Verilerin büyüklüğü ve yüksek sayıda öz niteliğe sahip olması nedeniyle bilgisayar analizlerine ve makine öğrenme, yapay zekâ ve derin öğrenme gibi algoritmalara ihtiyaç duyulmaktadır. Bu algoritmaların daha verimli sonuçlar ortaya koyabilmesi için verilerin analiz öncesi ön işleme sürecine tabi tutulması gerekmektedir. Verilerin kaydı sırasında iletim veya yanlış kodlama gibi algoritmik sorunlar ile oluşan hatalar sonucu veri tabanına eksik veri ile kaydedilmesi, kayıtlar arasında aykırı değerlerin olması, farklı niteliklerdeki verilerin farklı değer aralıkları olabileceği gibi bunların ölçeklendirilmesi gerekebilir (Polatgil, 2023).

Çalışmada kullanılacak verilerde eksiklik olması, tutarsız veya tamamlanmamış verileri içermesi, anlam ifade etmeyen yanlış verilerin olması başarı skorlarını olumsuz etkileyecektir. Verinin anlamlandırılmasında ve makine öğrenme algoritmalarıyla analiz sürecinde başarıyı artırmak için veri ön işleme adımlarının uygulanması önem arz etmektedir.

Bu çalışmada 30 Eylül 2022 tarihli Nevşehir Devlet Hastanesi Kardiyoloji bölümünden veri toplamak üzere E-26171210-929 sayılı etik kurul onayı sonrasında veriler toplanmaya başlanmıştır. Eylül 2022'den Mart 2023 tarihine kadarki süre

içerisinde veriler retrospektif olarak toplanmıştır ve ekler bölümünde etik kurul onayı sunulmuştur. Toplanan veriler daha sonra “.csv” uzantılı formata dönüştürülerek veri analizinde kullanılmak üzere oluşturulmuştur. Çalışmada analiz sürecinde Python Dili Jupyter platformu kullanılmıştır.

Kalp yetmezliğinin makine öğrenme algoritmalarıyla tespiti için framingham faktörleri ve uzman görüşü alınarak oluşturulan veri seti öznitelikleri toplamda 25 adet sütundan oluşmaktadır. Bu öznitelikler kişinin laboratuvar sonuçlarını, sigara, alkol tüketim durumlarını, ailesinde kalp rahatsızlığı olup olmadığı, kalp rahatsızlığını temsil edebilecek parametreleri ve en önemlisi kalp yetmezliği tanısının konulup konulmadığı bilgilerini içermektedir. Veri setinde kayıtlı tüm öznitelikleri ve bu özniteliklerin veri tipleri Çizelge 3.1’de gösterilmiştir.

Çizelge 3.1. Oluşturulan veri seti öznitelikleri

#	Öznitelik Adı	Veri Tipi	Açıklama
1	Yaş	int64	Hastaya ait yaş verisidir.
2	Cinsiyet	int64	Hastaya ait cinsiyet verisidir. Kadın {0}, Erkek {1} olarak saklanır.
3	Sigara Kullanım Durumu	int64	Sigara içmeme durumunu {0}, içme durumunu ise {1} olarak saklanmaktadır.
4	Alkol Kullanım Durumu	int64	Alkol içmeme durumunu {0}, içme durumunu ise {1} olarak saklanmaktadır.
5	Aile Öyküsü	int64	1. derece yakın akrabalarında kalp rahatsızlığı olup olmadığını tutmaktadır.
6	Atrial Fibrilasyon	int64	Hastada ritim bozukluğunun olup olmadığını temsil etmektedir.
7	Daha Önce Felç Geçirme Durumu	int64	Hastanın daha öncesinde felç geçirip geçirmediğini tutmaktadır.
8	Diyabet	int64	Şeker hastalığının olup olmadığını temsil eder.
9	Yüksek Tansiyon	int64	Yüksek tansiyon olup olmadığını temsil eder.
10	Kalp Büyümesi Var mı	int64	Hastada kalp büyümesi olup olmadığını temsil eder.
11	Kalp Duvar Kalınlığı Var mı	int64	Kalp duvar kalınlığının olup olmadığını tutulur.
12	Karaciğerde Sıvı Birikmesi Var mı	int64	Karaciğerde sıvı birikmesinin durumu tutulur.
13	Nefes Darlığı Var mı	int64	Hastanın nefes darlığı yaşayıp yaşamadığı bilgisi tutulur.
14	Bacaklarda Şişme Var mı	int64	Ayak ve bacaklarda şişmenin olup olmadığını bilgisi tutulur.
15	Damar Hastalığı Var mı	int64	Hastanın genel olarak damar hastalığı olup

			olmadığı bilgisi saklanır.
16	Böbrek Yetmezliği Var mı	int64	Hastanın böbrek yetmezliği olup olmadığı tutulur.
17	Anemi Var mı	int64	Hastada kansızlığın olup olmadığı tutulur.
18	LDL	float64	Laboratuvar değerine bakılır.
19	HDL	float64	Laboratuvar değerine bakılır.
20	Sodyum	float64	Laboratuvar değerine bakılır.
21	Glikoz	float64	Laboratuvar değerine bakılır.
22	TSH	float64	Laboratuvar değerine bakılır.
23	Kreatinin	float64	Laboratuvar değerine bakılır.
24	Ejeksiyon Fraksiyon	int64	Kalbin her atışında kanın pompalanan miktarını gösteren bilgi saklanmaktadır.
25	Tanı	int64	Hastanın kalp yetmezliği olup olmadığı tutulur.

Yapay Zekâ, ML ve DL gibi uygulamalarda daha iyi sonuçlar alabilmek için işlenmemiş verilerin üzerinde eksik verilerin tamamlanması temizlenmesi, aykırı verilerin ayıklanması, ortak bilgiyi temsil eden verilerin birleştirilmesi veya farklı bilgiyi temsil eden verilerin bölümlenmesi, özellik seçimi veya özellik çıkarılması gibi adımlara tabi tutularak verilerin ön işleme adımlarının uygulanması daha yüksek doğruluk ve güvenilirlik açısından önem arz etmektedir. Oluşturulan veri setine ait örnek verileri kolonlarıyla birlikte Çizelge 3.2’de verilmiştir.

Çizelge 3.2. Oluşturulan veri setinin örnek verileri

	Yas	Cinsiyet	Siğara	Alkol	Aile_Oykusu	Atrial_Fibrilasyon	Daha_Once_Felc_Gecirdimi	Diyabet	Yuksek_Tansiyon	Kalp_Buyumesi_Varmi	...	Bobrek_Yetmezligi_Varmi	Anemi_Varmi	LDL	HDL	Sodyum	Glikoz	TSH_Degeri	Kreatinin	EF	Tani	
1	72	1	0	0	0	1	0	0	1	0	...	0	0	113.5	46.0	148.7	119.6	0,80	1,20	62	0	
2	46	1	0	0	1	1	0	0	0	0	...	0	0	129.7	52.4	138.5	148.5	1,70	1,30	62	0	
3	53	0	0	0	0	1	0	0	0	0	...	0	0	149.6	62.5	138.7	121.7	2,20	1,20	62	0	
4	81	1	0	0	0	1	0	1	1	0	...	0	0	73.7	44.3	138.8	127.0	2,80	1,40	62	0	
5	74	0	0	0	0	1	0	0	1	0	...	0	0	127.1	53.4	135.0	141.7	1,60	0,70	55	1	
...
198	58	0	0	0	0	1	0	1	0	0	...	0	0	108.8	47.8	138.9	133.7	1,10	0,70	62	0	
199	72	0	0	0	0	0	0	0	1	0	...	0	0	57.3	58.4	134.3	109.2	3,20	1,10	50	0	
200	80	0	0	0	0	1	1	0	0	0	...	0	0	159.0	54.9	132.0	90.8	2,10	1,60	40	0	
201	60	1	1	0	0	1	0	0	0	0	...	0	0	56.2	39.7	141.0	100.4	1,70	0,70	50	0	
202	76	0	0	0	0	1	0	1	0	0	...	0	0	113.9	34.7	141.9	107.5	1,70	0,90	30	1	

3.3.1. Eksik verilerin bulunması ve temizlenmesi

Veri setinde yer alan özniteliklere ait eksik, hatalı, veri grubunun çok altında veya çok üstünde gibi aykırı verilerin tespiti, analize için gereksiz verilerin çıkarılma işlemlerine veri temizleme olarak karşımıza gelmektedir. MedData yazılımı ile hasta bilgilerinin tutulduğu ve Oracle Veri Tabanı Yönetim Sisteminde (VTYS) verilerin saklandığı yazılımdan edinilen bilgiler doğrultusunda Çizelge 3.2’de gösterilen özniteliklere ait veriler toplanmıştır. Örneğin; hasta şikayetlerinin ne olduğu, sigara alkol kullanma durumu, kullandığı ilaçlar, geçirdiği operasyonlar, teşhisi konmuş diğer hastalıkları, ailesinde kalp rahatsızlığı olup olmadığı gibi bilgilerin saklandığı hasta hikayesi başlığı altında toplanan bilgiler ilgili tabloda tutulmaktadır. Tablodaki bu bilgilerin incelenmesi sonucu ilgili kolonda örneğin ‘sigara’ içermesi durumunda tüm bilgilerin toplandığı tablonun sigara kolonuna lojik 1 değeri ile güncellenmektedir. Çizelge 3.3’de eksik verilerin tespiti ve doğrusuyla güncellenmesi için kullanılan örnek MSSQL komutları sunulmuştur.

Çizelge 3.3. Eksik verilerin tespiti ve doğrusuyla güncellenmesi

```

update Hasta_Hikayesi set Anemi_Varmi=1 where Dosya_No in (
select distinct(Dosya_No) from Hasta_Hikayesi where hikaye like '%kansızlık%')
update Hasta_Hikayesi set Anemi_Varmi=1 where Dosya_No in (
select distinct(Dosya_No) from Hasta_Hikayesi where hikaye like '%anemi%')
update Hasta_Hikayesi set Sigara=1 where Dosya_No in (
select distinct(Dosya_No) from Hasta_Hikayesi where hikaye like '%sigara%')
|
update Hasta_Hikayesi set Alkol=1 where Dosya_No in (
select distinct(Dosya_No) from Hasta_Hikayesi where hikaye like '%alkol%')

```

Bu işlem hasta hikayesi içerisinde yer alan LVH içeren veriler için *kalp duvar kalınlığı* kolonuna, HT için *yüksek tansiyon* kolonuna, felç geçirmiş ifadesi için *daha önce felç geçirdiği mi* kolonuna, kalp büyüme içeren verilerde *kalp büyümesi var mı* kolonuna şeklinde 15 adet öznitelik lojik 1 verisinin güncellenmesiyle eksik verilerin tespiti ve doğrusuyla güncellenmesi yapılmıştır. Örnek MSSQL komutları Çizelge 3.3’de verilmiştir. Hasta hikayesi içerisinde yer almayan bu 15 özniteliğin tablodaki karşılıklarına da lojik 0 olarak veri eklenmiştir. Örneğin, sigara içmeyen bireye lojik 0 verisinin atamasıdır. Çizelge 3.4’de boş veriye sahip öznitelikle 0 değerini atayan komutlar verilmiştir.

Çizelge 3.4. Boş verilerin güncellenmesi

```

SQLQuery2.sql - DE...85FLZ(Lenovo (D4)) * x SQLQuery1.sql - DE...85FLZ(Lenovo (S2)) *
update Hasta_Hikayesi set Yuksek_Tansiyon = 0 where Yuksek_Tansiyon is null
update Hasta_Hikayesi set Atriyal_Fibrilasyon = 0 where Atriyal_Fibrilasyon is null

update Hasta_Hikayesi set nefes_darligi = 0 where nefes_darligi is null

update Hasta_Hikayesi set diyabet = 0 where diyabet is null

```

3.3.2. Verilerin birleştirilmesi

Veri setinde farklı noktalarda yer alan aynı bilgiye karşılık gelen verileri belirleyerek bu verilerin birleştirilmesi önem arz etmektedir. Örneğin; hasta şikayetleri bilgisini içeren alanda kansızlık verisini anemi olarak da kaydedilebilmesi, Atriyal Fibrilasyon (AF) durumunu kısaca sinüs olarak kaydedilmesi ve ön işleme adımında bunların birleştirilmemesi veri kaybına yol açacaktır. Bu durumun yaşanmaması için ise aynı anlama gelen bu tarz veriler için birleştirilmeye gidilmesi gerekmektedir. Bu işlemin örneği Çizelge 3.5’de verilmiştir.

Çizelge 3.5. Veri birleştirme

```

update Hasta_Hikayesi set Atriyal_Fibrilasyon=1 where Dosya_No in (
select distinct(Dosya_No) from Hasta_Hikayesi where hikaye like '%AF%')

update Hasta_Hikayesi set Atriyal_Fibrilasyon=1 where Dosya_No in (
select distinct(Dosya_No) from Hasta_Hikayesi where hikaye like '%Sinüs%')

update Hasta_Hikayesi set Anemi_Varmi=1 where Dosya_No in (
select distinct(Dosya_No) from Hasta_Hikayesi where hikaye like '%kansızlık%')

update Hasta_Hikayesi set Anemi_Varmi=1 where Dosya_No in (
select distinct(Dosya_No) from Hasta_Hikayesi where hikaye like '%anemi%')

```

3.3.3. Standardizasyon

Analizde verilerin sonuca etkisinin eşit oranda katkı vermesi için aynı birimlerde olması beklenir. Standardizasyon yapılmak istenen her bir verinin değeri, o değerlerin aritmetik ortalamasında çıkarılarak standart sapma değerine bölünmesiyle ilgili veriyi standart bir ölçüme getirilmiş olacaktır. ML algoritmalarında verilerin standardize edilmesi sonucu doğrudan etkilemektedir. Standardizasyon hesaplama formülü denklem 3.1’de gibidir (Gemici, 2012). Çizelge 3.6’da yapılan standardizasyon uygulaması verilmiştir.

$$Z = \frac{x - \mu}{\sigma} \quad (3.1)$$

Z: Ölçüm Sonucu, X: Dönüştürülecek Değer, μ : Aritmetik Ortalama, σ : Standart Sapma.

Çizelge 3.6. Standardizasyon uygulaması

	Yaş	Cinsiyet	Sigara	Alkol	Aile Öyküsü	AF	Daha Öce Felç Geçirdi mi	Diyabet	Yüksek Tansiyon	Kalpте Büyüme	...	Böbrek Yetmezliği	Anemi	LDL	HDL	Sodyum	Glukoz	TSH	Kreatinin	EF	Tanı
0	0,40	1,0	0,0	0,0	0,0	1,0	0,0	0,0	1,0	0,0	...	0,0	0,0	0,09	0,48	3,19	0,25	0,52	0,43	0,74	0,0
1	1,63	1,0	0,0	0,0	1,0	1,0	0,0	0,0	0,0	0,0	...	0,0	0,0	0,09	0,01	0,25	0,19	0,12	0,74	0,74	0,0
2	1,08	0,0	0,0	0,0	0,0	1,0	0,0	0,0	0,0	0,0	...	0,0	0,0	0,08	0,73	0,31	0,22	0,09	0,43	0,74	0,0
3	1,11	1,0	0,0	0,0	0,0	1,0	0,0	1,0	1,0	0,0	...	0,0	0,0	0,10	0,60	0,34	0,13	0,36	1,05	0,74	0,0
4	0,56	0,0	0,0	0,0	0,0	1,0	0,0	0,0	1,0	0,0	...	0,0	0,0	0,09	0,06	0,75	0,09	0,17	1,12	0,07	1,0

3.3.4. Normalizasyon

Veri setinde tekrarlı verileri ortadan kaldırmak, verilerin daha tutarlı olmasını sağlamak ve aykırı değerlerin önüne geçmek için kullanılır. Özniteliklere ait değerlerin o özneliğe ait en küçük değerden çıkarılıp, en büyük değerle en küçük değerden çıkan değere bölümü ile hesaplanmasına denir (Polatgil, 2023). Denklem 3.2’de normalizasyon formülü verilmiştir.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3.2)$$

X_{norm} = Değişkenin normalizasyon sonucu, X = Özneliğe ait değer.

X_{min} = Özneliğe ait en küçük değer, X_{max} = Özneliğe ait en büyük değer.

Veri setinin normalizasyon işlemi sonrası değerleri Çizelge 3.7’de verilmiştir.

Çizelge 3.7. Normalizasyon uygulaması

	Yaş	Cinsiyet	Siğara	Alkol	Aile Öyküsü	AF	Daha Önce Felç Geçirdi mi	Diyabet	Yüksek Tansiyon	Kalpde Büyüme	...	Böbrek Yetmezliği	Anemi	LDL	HDL	Sodyum	Glikoz	TSH	Kreatinin	EF	Tanı
0	0.650794	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	...	0.0	0.0	0.001714	0.307403	1.000.000	0.121343	0.027682	0.333333	1.000.000	0.0
1	0.238095	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.002074	0.387704	0.507246	0.194861	0.058824	0.388889	1.000.000	0.0
2	0.349206	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.002516	0.514429	0.516908	0.126685	0.076125	0.333333	1000000	0.0
3	0.793651	1.0	0.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	...	0.0	0.0	0.000830	0.286073	0.521739	0.140168	0.096886	0.444444	1.000.000	0.0
4	0.682540	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	...	0.0	0.0	0.002016	0.400251	0.338164	0.177563	0.055363	0.055556	0.833333	1.0

3.3.5. Özellik seçimi

Veri setinde yer alan özniteliklerin, analize ne oranda katkı sağladığını tespit etmek için her bir özneliğin tek tek değerlendirilme sürecine özellik seçimi denilmektedir. Özellik seçiminin amacı kullanılacak algoritmalarda daha başarılı ve hızlı sonuç vermesi için kullanılmaktadır. 25 öznelikten oluşan veri setinin WEKA programı ile özellik seçimi için “ARFF” uzantılı dosya formatına uygun şekilde hazırlanmıştır. Hazırlanan dosya Şekil 3.19’da verilmiştir.

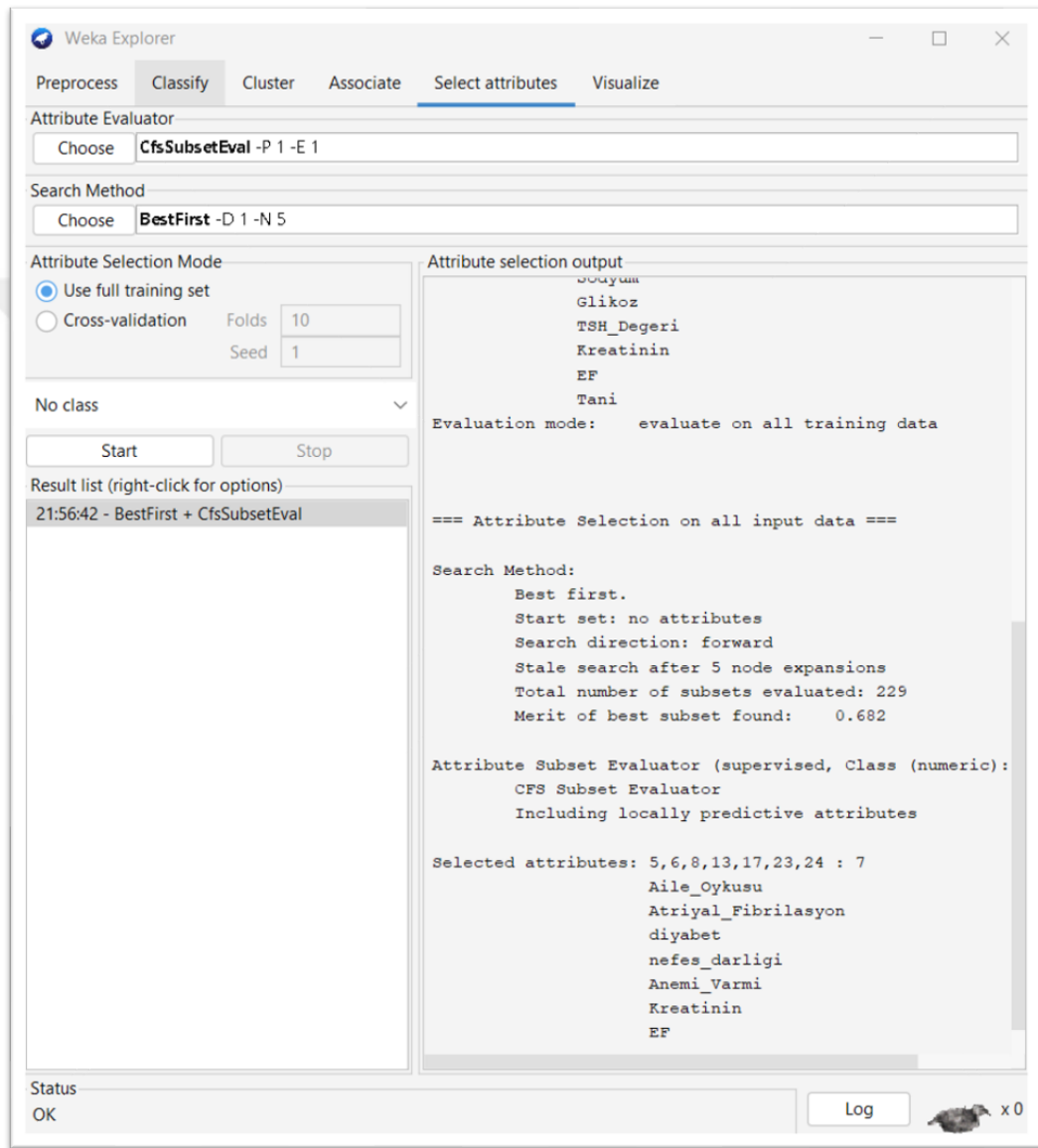
```

1 relation eksiksiz_veriseti
2
3 @attribute Yas numeric
4 @attribute Cinsiyet numeric
5 @attribute sigara numeric
6 @attribute Alkol numeric
7 @attribute Aile_Oykusu numeric
8 @attribute Atrial_Fibrilasyon numeric
9 @attribute daha_once_felc_gecirdimi numeric
10 @attribute diyabet numeric
11 @attribute Yuksek_Tansiyon numeric
12 @attribute Kalp_Buyumesi_Varmi numeric
13 @attribute Kalp_Duvar_Kalin numeric
14 @attribute Karacigerde_Sivi_Birikmesi numeric
15 @attribute nefes_darligi numeric
16 @attribute Bacaklarda_Sisme numeric
17 @attribute Damar_Hastaligi_Varmi numeric
18 @attribute Bobrek_Yetmezligi_Varmi numeric
19 @attribute Anemi_Varmi numeric
20 @attribute LDL numeric
21 @attribute HDL numeric
22 @attribute Sodyum numeric
23 @attribute Glikoz numeric
24 @attribute TSH_Degeri numeric
25 @attribute Kreatinin numeric
26 @attribute EF numeric
27 @attribute Tani numeric
28
29 @data
30 72,1,0,0,0,1,0,0,1,0,1,0,1,1,0,0,0,113.5,46,148.7,119.6,0.8,1.2,62,0
31 46,1,0,0,1,1,0,0,0,0,0,0,1,0,0,0,0,129.7,52.4,138.5,148.5,1.7,1.3,62,0
32 53,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,149.6,62.5,138.7,121.7,2.2,1.2,62,0
33 81,1,0,0,0,1,0,1,1,0,1,0,1,0,0,0,0,73.7,44.3,138.8,127.2,0.8,1.4,62,0
34 74,0,0,0,0,1,0,0,1,0,0,0,1,0,0,0,0,127.1,53.4,135,141.7,1.6,0.7,55,1
35 68,1,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,93.4,49.4,142.6,175,1.2,1.2,55,1
36 74,1,0,0,0,1,0,0,1,0,0,0,0,0,0,0,0,55.2,40.6,133,111.9,1,1.3,62,0
37 82,0,0,0,0,1,0,0,1,0,0,0,0,0,0,0,0,90.8,52.1,134.1,136.6,2.6,0.6,50,1
38 69,0,0,0,0,1,0,0,0,0,1,0,0,0,0,0,0,112.5,65.3,146.5,107.8,2.2,0.7,62,0
39 53,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,190.1,98.8,139.5,133.5,2,0.8,62,0
40 76,0,0,0,0,1,0,0,1,0,1,0,1,0,0,0,0,148.2,53.3,134.8,150.6,1.6,1,62,0
41 63,1,0,0,0,1,0,0,1,0,1,0,1,0,0,0,0,81.6,54.8,139.4,108.5,0.9,0.8,62,0

```

Şekil 3.19. Veri setinin arff formatı

Verilerin özellik seçimi için çeşitli algoritmalar kullanıldığı gibi bu çalışmada WEKA programıyla CfsSubsetEval algoritması ile özellik seçimi yapılmıştır. Özellik seçimi sonrasında EF, Kreatinin, Anemi, Nefes Darlığı, Diyabet, AF, Aile Öyküsü özneliklerinin sonuca katkısının daha yüksek olduğu tespit edilmiştir. Bu sonuçları gösteren WEKA programının ekran görüntüsü Şekil 3.20’de gösterilmiştir.



Şekil 3.20. WEKA programı ile özellik seçimi

3.3.6. Heart failure prediction veri seti ön işleme

Kaliforniya Üniversitesi <https://archive.ics.uci.edu/datasets> adresinde sağlık, tıp, ilaç, gıda, canlı, fen bilimleri, işletme, hukuk, sosyal bilimler gibi alanlarda sınıflandırma, regresyon, kümelere vb. analiz türlerinde sayısal, kategorik veya

karmaşık özellik türleri barındıran, hazır veri setlerini ücretsiz olarak erişime sunduğu 600’den fazla veri seti içermektedir.

İlgili veri seti sayfasından “Heart Failure Prediction” başlıklı hazır veri oluşturulan veri seti yanı sıra da analiz edilmiştir. Veri setinde 299 adet örneklem bulunmaktadır. Çizelge 3.8’de hazır veri setinin özniteliklerinin adları, veri tipleri ve ne anlama geldikleri gösterilmiştir.

Çizelge 3.8. Heart failure prediction veri seti öznitelikleri

#	Öznitelik Adı	Veri Tipi	Açıklama
1	Yaş	float64	Bireyin Yaş Verisini
2	Anemi	int64	Kansızlık olup olmadığı
3	Kreatinin Fosfokinaz	int64	Laboratuvar değeri
4	Diyabet	int64	Şeker hastası olup olmadığı
5	Ejeksiyon Fraksiyon	int64	Kanın Pompalama Miktarı (Her Kalp Atıştaki)
6	Yüksek Tansiyon	int64	Yüksek Tansiyon olup olmadığı
7	Trombosit	float64	Laboratuvar değeri
8	Serum Kreatinin	float64	Laboratuvar değeri
9	Serum Sodyum	int64	Laboratuvar değeri
10	Cinsiyet	int64	Bireyin Cinsiyet Bilgisi
11	Sigara İçme	int64	Sigara İçip İçmeme Durumu
12	Zaman	int64	Hastanın Gözlem Süresi
13	Ölme Olayı	int64	Gözlem Sürecinde Ölme Durumu

KY ait Heart failure prediction veri setinde yer alan 10 adet örnek veriler Çizelge 3.9’da verilmiştir.

Çizelge 3.9. Heart failure prediction veri setine ait örnek veriler

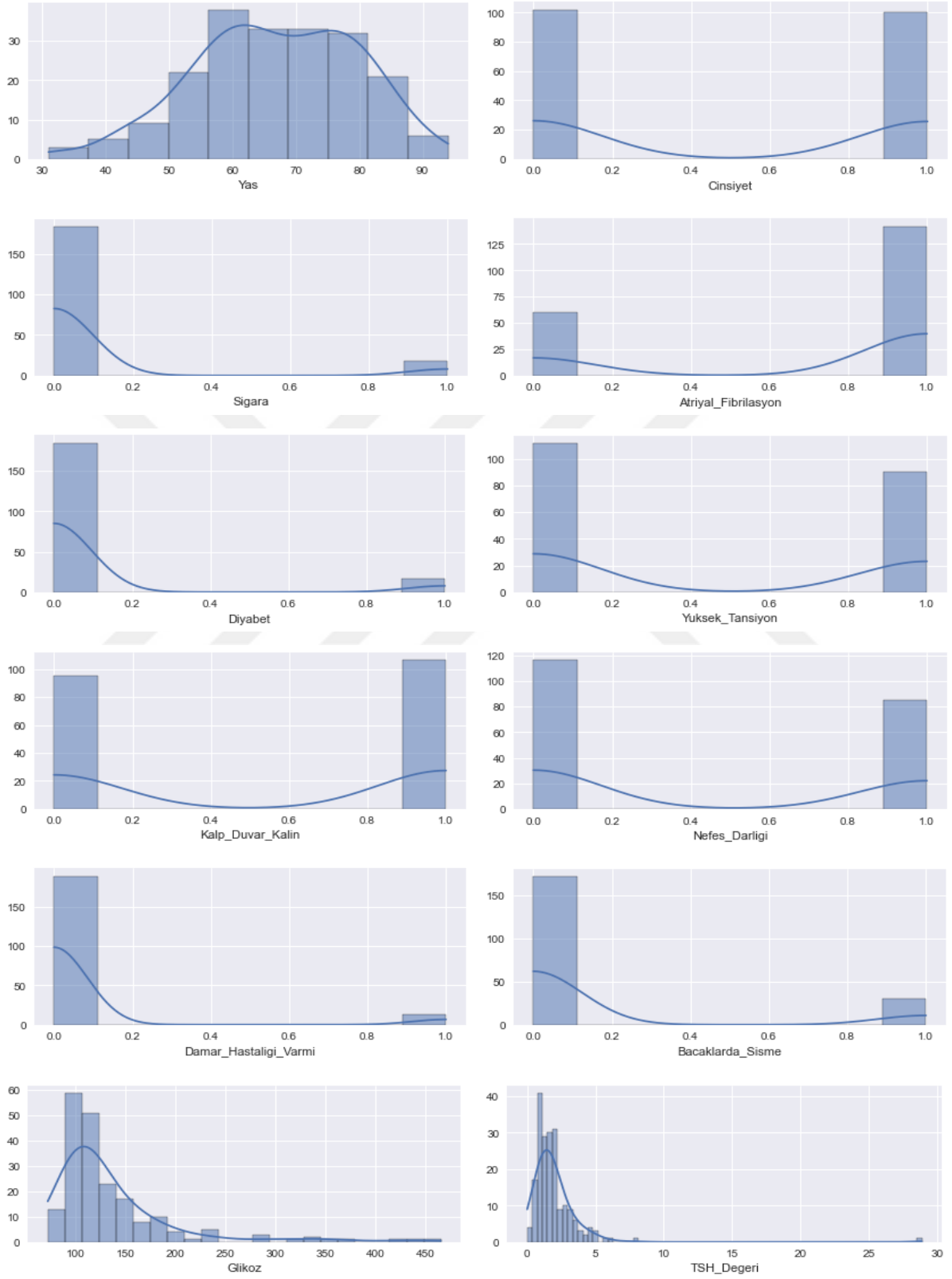
#	Yaş	Anemi	Kreatinin Fosfokinaz	Diyabet	Ejeksiyon Fraksiyon	Yüksek Tansiyon	Trombosit	Serum Kreatinin	Serum Sodyum	Cinsiyet	Sigara İçme	Zaman	Ölme Olayı
1	75	0	582	0	20	1	265000	1,9	130	1	0	4	1
2	55	0	7861	0	38	0	263358	1,1	136	1	0	6	1
3	65	0	146	0	20	0	162000	1,3	129	1	1	7	1
4	50	1	111	0	20	0	210000	1,9	137	1	0	7	1
5	65	1	160	1	20	0	327000	2,7	116	0	0	8	1
6	90	1	47	0	40	1	204000	2,1	132	1	1	8	1
7	75	1	246	0	15	0	127000	1,2	137	1	0	10	1
8	60	1	315	1	60	0	454000	1,1	131	1	1	10	1
9	65	0	157	0	65	0	263358	1,5	138	0	0	10	1
10	80	1	123	0	35	1	388000	9,4	133	1	1	10	1

KY ait 299 adet veri ve 13 ayrı niteliklerin ortalama değeri, standart sapması, minimum ve maksimum değerleri ile çeşitli yüzdesel değerlerin istatistikî sonuçları Çizelge 3.10’da verilmiştir.

Çizelge 3.10. Heart failure prediction veri seti istatistik bilgisi

	Yaş	Anemi	Kreatinin Fosfokinaz	Diyabet	Ejeksiyon Fraksiyon	Yüksek Tansiyon	Trombosit	Serum Kreatinin	Serum Sodyum	Cinsiyet	Sigara İçme	Zaman	Ölme Olayı
count	299	299	299	299	299	299	299	299	299	299	299	299	299
mean	60,83	0,43	581,83	0,41	38,08	0,35	263358,02	1,39	136,62	0,64	0,2	130,26	0,32
std	11,89	0,49	970,28	0,49	11,83	0,47	97804,23	1,03	4,41	0,47	0,46	77,61	0,46
min	40	0	23	0	14	0	25100	0,5	113	0	0	4	0
25%	51	0	116,5	0	30	0	212500	0,9	134	0	0	73	0
50%	60	0	250	0	35	0	262000	1,1	137	1	0	115	0
75%	70	1	582	1	45	1	303500	1,4	140	1	1	203	1
max	95	1	7861	1	80	1	850000	9,4	148	1	1	285	1

KY için oluşturulan veri seti için gözlem niteliğindeki bazı verilerin grafiksel gösterimi Şekil 3.21’de verilmiştir.



Şekil 3.21. Oluşturulan veri setinin gözlem özellikleri grafikleri

“Heart Failure Prediction” isimli veri setini kullanarak yazılmış makaleler incelenmiş olup bu incelemede kullanılan veri ön işleme adımları, uygulanan yöntemler, çapraz doğrulama durumları ve değerlendirme ölçütleri ele alınmıştır.

Gürgen ve Serttaş çalışmalarında KY bulunan hastaların erken teşhisi için kullandığı makine öğrenimi algoritmalarının verimliliklerini karşılaştırmışlardır. ML algoritması olarak KNN, AdaBoot, LightGBM, XGBoost, RF, Classification and Regression Tree (CART), LR, CatBoost kullanılmıştır. %99 ile en yüksek doğruluk değerini RF algoritması ile yakalamıştır. Çalışmalarında kadınların erkeklere göre KY yakalanma riskinin daha düşük olduğu gözlemlenmiştir (Gürgen ve Serttaş, 2023).

Bozkurt Keser ve Keskin sınıflandırmaya dayalı ML algoritmasıyla kalp yetmezliğinin sağ kalım tahminini gerçekleştiren çalışmasında RF, YSA ve GSB algoritması kullanmışlardır. Çalışmalarında kullanılan algoritmalarından YSA %86,67 oranla diğer algoritmalarından daha iyi sonuç verdiğini bulmuşlardır (Keser ve Keskin, 2023).

Altunbey Özbay ve Özbay KY hastalığında ML algoritmalarını kullanarak hastaların hayatta kalma durumlarının analiz ederek performans karşılaştırması yapmışlardır. Çalışmalarında Attribute Selected Classifier (ASC), Random Tree (RT), PART, NB, RF, CvR, J48, AdaBoost, OneR, JRip algoritmalarını kullanmışlardır. Her algoritmayı 5’li ve 10’lu çapraz doğrulama yaparak sonuçları ayrı ayrı değerlendirmiştir. RF algoritmasının oluşturulan diğer modellere göre daha doğru sonuç verdiğini tespit etmişlerdir (Altunbey Özbay ve Özbay, 2022).

Kalp yetmezliğinde sağ kalım sınıflandırması üzerine çalışan Potur ve Erginel kullanılan algoritmalarla doğruluk değerini arttırmak için 3 farklı öznelik seçim yöntemi kullanmıştır. Bu öznelik seçme yöntemleri InfoGainAttributeEval, CfsSubsetEval, ReliefFAttributeEval olarak karşımıza gelmektedir. Çalışmalarında ML algoritmalarından çok katmanlı algılayıcı, LR, NB, J48, SVM yöntemleri tercih edilmiştir. Diğer çalışmalar gibi değerlendirme ölçütü olarak karmaşıklık matrisi sonucuna göre en iyi değeri çok katmanlı algılayıcı %90 doğruluk oranına ulaşmışlardır (Potur ve Erginel, 2021).

Ahmad ve arkadaşları KY hastaların hayatta kalma analizlerini bir vaka çalışması olarak yaptıkları araştırmada kullanılan söz konusu veri setinde Cox regresyon algoritması ve istatistiksel yöntem olan Kaplan Meier kullanılmıştır. Özneliklerin değerlendirmek için ise Martingale yöntemi tercih etmişlerdir.

Çalışmalarında sonuç olarak ilerleyen yaş, yüksek tansiyon (HP) olma durumu, ejeksiyon fraksiyon değerinin düşük olması hastalığın olma durumunu arttırmakta ve ölme oranını destekleyen faktörler olarak belirtmişlerdir (Ahmad ve ark., 2017).

Chicco ve Jurman çalışmalarında Lineer Regresyon, RF, DT, Artificial Neural Network ANN, SVM, KNN, NB ve son olarak Gradient Boosting (GB) ML algoritmalarını ve YSA kullanmışlardır. Bu yöntemlerin yanı sıra biyoistatistiksel yöntemleri de alternatif olarak analize dahil etmişler çıkan iki sonucu karşılaştırmışlardır. Özellik seçimine göre sonuca en etkin öznelik olarak serum kreatinin ile EF olduğunu bulmuşlardır (Chicco ve Jurman, 2020).

3.3.7. Çapraz doğrulama (cross validation)

Çapraz doğrulama, oluşturulan bir modelin doğruluğunu hesaplamak için kullanılan değerlendirme ölçekleri için veri setinin belirli oranda ayrılması gerekmektedir. Değerlendirmenin hatasız yapılabilmesi için veri setinde yer alan tüm verilerin belirli parçalarda hepsinin test edilmesi gerekmektedir.

Geliştirilen model için veri setinin belirli oranda eğitim ve test için ayrılma işlemine k-katlı çapraz doğrulama denilmektedir. Veri setinin k sayıda parçaya bölünerek 1 parçanın test, k-1 parçanın ise eğitim verileri olarak ayrılması sürecinin tüm bölümlerin sırayla test edilir. Bu test edilme işlemi sonrasında çıkan tüm değerlerin ortalaması alınarak nihai sonuç hesaplanmış olur. Çizelge 3.11'de çapraz doğrulama sunulmuştur.

Çizelge 3.11. Çapraz Doğrulama (Cross Validation)

Bölüm 1	Test	Eğitim	Eğitim	Eğitim	Eğitim	Hata Oranı 1
Bölüm 2	Eğitim	Test	Eğitim	Eğitim	Eğitim	Hata Oranı 2
Bölüm 3	Eğitim	Eğitim	Test	Eğitim	Eğitim	Hata Oranı 3
Bölüm 4	Eğitim	Eğitim	Eğitim	Test	Eğitim	Hata Oranı 4
Bölüm 5	Eğitim	Eğitim	Eğitim	Eğitim	Test	Hata Oranı 5
						Ortalama Hata Oranı

3.4. Değerlendirme Ölçütleri

ML algoritmalarıyla elde edilen veri setine ait özelliklerin tahminleri ile gerçek dünyadaki verilerin karşılaştırılması sonucu çıkan istatistiki değerleri ölçer sürecidir. Bir değerlendirme yöntemi belirli bir veri setinde yüksek doğrulukta sonuç verirken farklı veri setlerinde ve algoritmalarda düşük doğrulukta sonuç üretebilmektedir. Bu çalışma için oluşturulan veri seti ve kaggle web sayfasından kalp yetmezliği için Kaliforniya Üniversitesi tarafından oluşturulmuş hazır veri setleri değerlendirilmiştir.

3.4.1. Karmaşıklık matrisi

Sınıflandırma algoritmalarının hesapladığı sonuçlara göre çıktı özniteliğine ait değerlerin gerçek sonuçla karşılaştırılma işlemine karmaşıklık matrisi denilmektedir. Karmaşıklık matrisi (Confusion Matrix) temel olarak veri setine ait gerçek sonuçlar ile geliştirilen modelin oluşturduğu tahmin sonuçlarının doğru ve/veya yanlış tahmin durumlarının karşılaştırılmasıdır. Bu karşılaştırmalar sonrasında Doğruluk (Accuracy), Kesinlik (Precision), Duyarlılık (Recall), F1-Skor vb. değerlerin hesaplanması bulunmaktadır. Gerçek değerlerle tahmin edilen değerleri karşılaştıran karmaşıklık matrisi Çizelge 3.12’de gösterilmiştir.

Çizelge 3.12. Karmaşıklık matrisi (confusion matrix)

		Gerçek Değerler	
		Gerçek	Yanlış
Tahmin Edilen Değer	Pozitif	Gerçek Pozitif (TP)	Yanlış Pozitif (FP)
	Negatif	Yanlış Negatif (FN)	Gerçek Negatif (TN)

Gerçek değerler ile geliştirilen modelin tahmin sonuçlarının karşılaştırıldığı bu ölçüm matrisinde karşımıza aşağıdaki kavramlar gelmektedir.

Gerçek Pozitif (TP): Gerçek dünya verisinde **doğru** olan ve geliştirilen model tarafından da **doğru** olarak tahmin edilenlerin sayısını temsil eder.

Yanlış Pozitif (FP): Gerçek dünya verisinde **yanlış** olan ve geliştirilen model tarafından da **doğru** olarak tahmin edilenlerin sayısını temsil eder.

Gerçek Negatif (TN): Gerçek dünya verisinde **doğru** olan ve geliştirilen model tarafından da **yanlış** olarak tahmin edilenlerin sayısını temsil eder.

Yanlış Negatif (FN): Gerçek dünya verisinde **yanlış** olan ve geliştirilen model tarafından da **yanlış** olarak tahmin edilenlerin sayısını temsil eder.

Doğruluk (Accuracy): Geliştirilen model sonrasında karmaşıklık matrisinde yer alan ve doğru tespit edilen (Gerçek Pozitif, Gerçek Pozitif) sonuçlarının toplamı bütün küme parametrelerinin (Gerçek Pozitif, Gerçek Negatif, Yanlış Pozitif, Yanlış Pozitif) toplamına bölümü sonucu çıkan değerdir. Doğruluk formülü Denklem 3.3'de verilmiştir.

$$Doğruluk = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.3)$$

Kesinlik (Precision): Gerçekteki doğru olan değer doğru olarak tespit edilenlerin sayısının pozitif olarak tahmin edilen tüm değerlerin sayısının toplamına oranı ile bulunan değerdir. Kesinlik formülü Denklem 3.4'de verilmiştir.

$$Kesinlik = \frac{TP}{TP + FP} \quad (3.4)$$

Duyarlılık (Recall): Gerçekteki doğru olan değer doğru olarak tespit edilmesinin gerçek pozitif ve yanlış pozitif değerlerinin toplamına bölümü ile hesaplanmaktadır. Kalp yetmezliği olan hastanın sağlıklı olarak tahminini temsil etmektedir. Duyarlılık formülü Denklem 3.5'de verilmiştir.

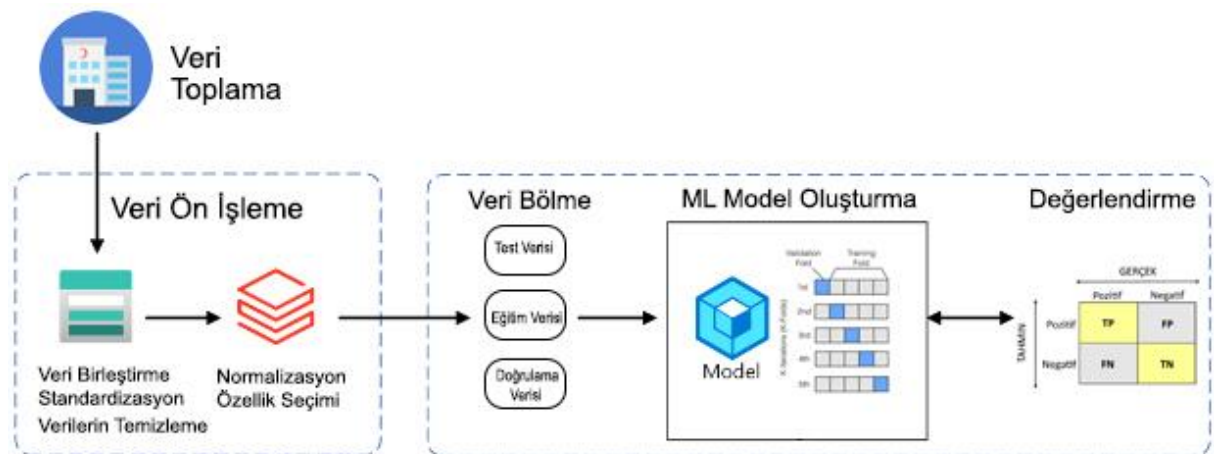
$$Duyarluluk = \frac{TP}{TP + FN} \quad (3.5)$$

F1 Skor Değeri (F1_Score): Hesaplanan hassasiyet ile kesinlik değerlerinin harmonik ortalamasıyla bulunan değerdir. F1 skor formülü Denklem 3.6'da verilmiştir.

$$F1\ Skor = \frac{2 * Kesinlik * Hassasiyet}{Kesinlik + Hassasiyet} \quad (3.6)$$

3.5. Önerilen Yöntem

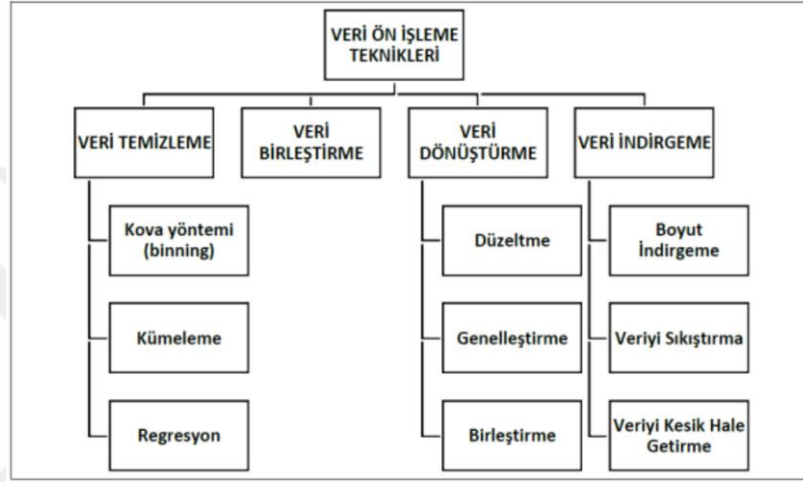
Nevşehir Devlet Hastanesi Eylül 2022'den Mart 2023 tarihleri arasında E-26171210-929 sayılı etik kurul onayı ile toplanan 150 adet kalp rahatsızlığı bulunmayan, 52 adet kalp rahatsızlığı bulunan toplamda 202 veri üzerinde çalışılmıştır. Bu çalışmada YSA, RF, NB, LR, DT, KNN algoritmaları kullanılmıştır. Veri analizinde güvenilirlik ve doğruluk değerlerini artırmak için bu algoritmaların hepsine verilerin %20 oranda bölünüp çapraz doğrulama (Cross Validation) yapılmıştır. Çapraz doğrulama sonrasında çıkan her sonuç değerlerinin ortalaması alınmıştır. Değerlendirme ölçütleri olarak temelinde istatistik oranlara dayalı olan karmaşıklık matrisi (Confusion Matrix) ile doğruluk, kesinlik, duyarlılık ve F1 skor değerleri hesaplanmıştır. En etkili sonuç veren makine öğrenme algoritma yöntemi tespit edilmiştir. Önerilen yöntemin işlem basamakları Şekil 3.22'de verilmiştir.



Şekil 3.22 – Önerilen yöntemin işlem basamakları

3.5.1. Verinin toplanması ve ön işlenmesi

Veri ön işleme süreçleri Anaconda Navigatör IDE üzerinde çalıştırılan Notebook Jupyter platformu ile geliştirilen komutlarla gerçekleştirilmiştir. Bu süreçte eksik verilerin tespiti ve temizlenmesi, farklı noktalardaki ortak verilerin birleştirilmesi, verilerin eşit oranda katkı sağlaması için normalize edilmesi, tekrarlı verileri ortadan kaldırılması adımları gerçekleştirilmiştir. Veri ön işleme sürecinin adımları Şekil 3.23’de sunulmuştur.

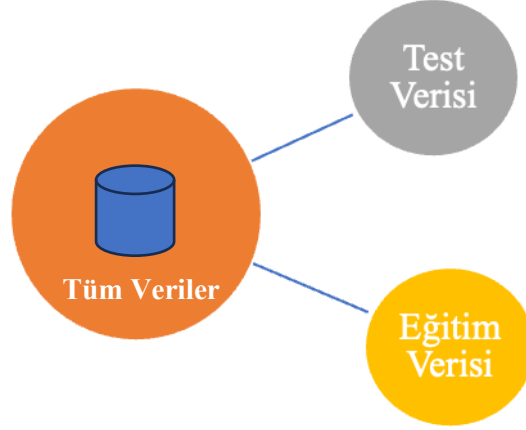


Şekil 3.23 – Veri ön işleme adımları (Toprak ve Zaim, 2021)

Notebook Jupyter üzerinde Numpy, Pandas, Seaborn, Sklearn kütüphaneleri eklenerek ve veri setinin ekleme, eklenecek verilerin okunması ve hedef değişkeninin silinmesi gerçekleştirilmiştir.

3.5.2. Verilerin bölünmesi

ML algoritmaları temelde bağımsız değişkenlerden ve aralarındaki matematiksel fonksiyonlarla oluşan bağlantı ile bağımlı değişken olan hedef/sonuç değişkeninin hesaplanmasını amaçlanmaktadır. Bu amaç doğrultusunda verilerin bir bölümü eğitim bir bölümü de test için ayrılmalıdır. Bölünen eğitim verilerinden oluşturulan model ile eğitim sürecine dahil edilmeye ve sistemin hiç bilmediği test verileriyle model test edilmektedir. Cevapları bilinen test verileri ile tahminde bulunulan sonuçlar karşılaştırılarak modelin performans değerlendirilmesi yapılmaktadır. Veri seti Python komutları ile eğitim ve test olarak bölünmüştür. Şekil 3.24’de verilerin bölünmesi verilmiştir.



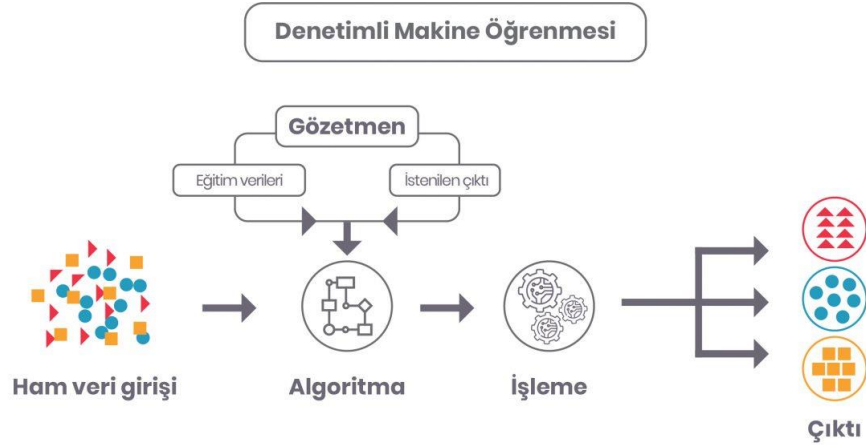
Şekil 3.24- Verilerin bölünmesi

Verileri bölme sürecinde dikkat edilmesi gereken bir diğer husus ise çapraz doğrulama yöntemidir. Çapraz doğrulama verilerin belirli bir oranda ve sırayla tüm verilerin eğitim ve test veri kümelerinde kullanılmasını ve test edilip değerlendirilmesini amaçlamaktadır. Bu sayede test için ayrılan verilerin sonucu kesin doğru tahmin edilebilecekler içerisinde seçilip modelin performans sonucunu yanılmayı engellemektedir.

Bu çalışmada beşli çapraz doğrulama uygulanmıştır. Veriler beş parçaya ayrılarak sırayla eğitim ve test verileri tek tek ayrı gruplara alınmıştır. Her grup sonrasında karmaşıklık matrisi ile doğruluk değeri hesaplanıp en son bu değerlerin ortalaması alınarak nihai doğruluk değeri bulunmuştur.

3.5.3. Modelin eğitimi

Bu çalışmada RF, NB, LR, DT, KNN algoritmaları kullanılmıştır. Veri analizinde doğruluk değerlerini artırmak için her bir modele beşli çapraz doğrulama (Cross Validation) yapılmıştır. Örnek olarak karar ağacına ait %80 eğitim, %20 oranla test veri kümesine bölünmüş Python komutları verilmiştir. İşlenmemiş bir verinin girdi halinden verilerin bölünme işlemi sonrasında kullanılacak algoritma ile işlenerek sınıflandırılma yapılmıştır. Şekil 3.25’de denetimli makine öğrenme model şeması verilmiştir.



Şekil 3.25 – Denetimli ML model şeması

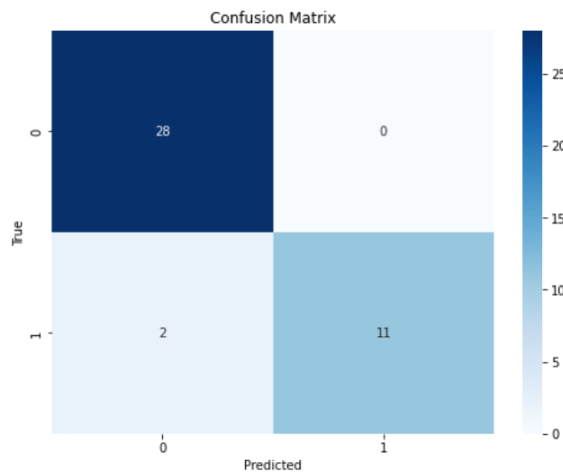
(<https://www.turhost.com/blog/makine-ogrenmesi-machine-learning-nedir/>)

Erişim Tarihi: 21.02.2024

3.5.4. Modelin değerlendirilmesi

Oluşturulan ve çapraz doğrulama uygulanan her bir modelin sonuçlarının değerlendirilmesi için karmaşıklık matrisi kullanılmıştır. Karmaşıklık matrisi ile modelin doğruluk, kesinlik, duyarlılık, F1 skorları ölçülmüştür.

Karmaşıklık matrisine ait gerçekte doğru iken modelin doğru (0,0 koordinatındaki 28 değeri) ve gerçekte yanlışken modelin de yanlış (1,1 koordinatlarındaki 11 değeri) olarak tahmin etme durumlarının yüksek olması beklenir. Şekil 3.26’da örnek bir karmaşıklık matrisi verilmiştir.



Şekil 3.26 – Örnek karmaşıklık matrisi

4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA

Bu çalışmada veri toplama izni alınarak elde edilen veri seti ve Kaliforniya Üniversitesinin “Heart Failure Prediction” isimli veri seti olmak üzere 2 ayrı veri seti üzerinde çalışılmıştır. Nevşehir Devlet Hastanesine KY şikayetiyle gelen hastaların bilgilerinden oluşturulan veri setinde 150 sağlıklı tanısı, 52 hasta tanısı konmuş toplamda 202 veri toplanmıştır. Bu veri setinde 25 öznitelik bulunmaktadır. Hazır olarak edinilen veri setinde ise 299 örnek veri ve 13 öznitelik bulunmaktadır. Bu veri setleri üzerinde YSA, KNN, DT, LR, NB, RF algoritmaları ile analiz gerçekleştirilmiştir. Doğruluk değerinin güvenilirliğini arttırmak için çalışmada kullanılan tüm algoritmalar üzerinde beşli çapraz doğrulama uygulanmıştır. Her algoritma modeline ait ROC eğrisi ve değerlendirme ölçütü sonuçları verilmiştir.

4.1. Oluşturulan Veri Setinin Sonuçları

202 örnek ve 25 öznitelik ile toplanılarak oluşturulan veri seti üzerinde YSA, KNN, DT, LR, NB, RF algoritmaları için analizleri ve değerlendirme ölçütleri bu başlık altında verilmiştir.

4.1.1. K En Yakın Komşu algoritması sonuçları

KNN algoritması için veri seti ilk olarak %80 eğitim ve %20 test verisi olarak bölünmüştür. Bu işlem sonrasında doğruluk değeri 0,81 olarak hesaplanmıştır. Oluşturulan modelin güvenilirliğini arttırmak için beşli çapraz doğrulama uygulanmıştır. Modelin oluşturulması için Sklearn kütüphanesinde yer alan KNeighborsClassifier kütüphanesi eklenmiştir.

KNN algoritması için uygulanan 5 sonuçlu çapraz doğrulama (cross validation) sonuçlarına bakıldığında en yüksek değer olarak sonuç 2,3 ve 4 de 0.83 tüm sonuçların ortalamasına bakıldığında ise 0.81 sonucu hesaplanmıştır. Bu modelin sonuçları Çizelge 4.1’de verilmiştir.

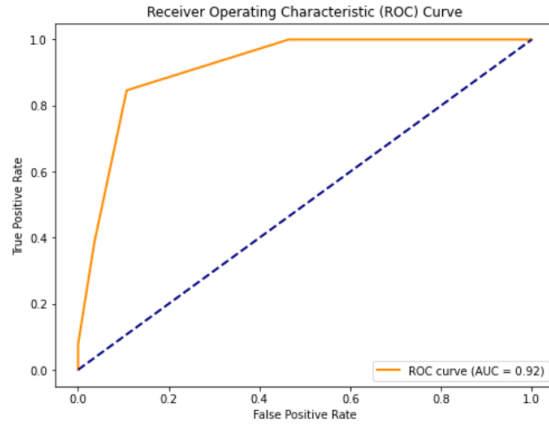
Çizelge 4.1. Knn algoritması çapraz doğrulama (5 fold) uygulama sonuçları
5 li Çapraz Doğrulama Yapılan KNN Algoritmasının Sonuçları

Fold1	Fold2	Fold3	Fold4	Fold5
0.76	0.83	0.83	0.83	0.78
Ortalama Sonuç: 0.81				

Beşli çapraz doğrulama için ayrılan test ve eğitim verilerini temel alarak oluşturulan KNN algoritmasına ait sonuçları ayrı ayrı hesaplanarak dizi içerisine kaydedilmiştir. Döngünün tamamlanmasıyla dizide kayıtlı sonuçların numpy kütüphanesinin mean komutu ile ortalaması hesaplanmıştır.

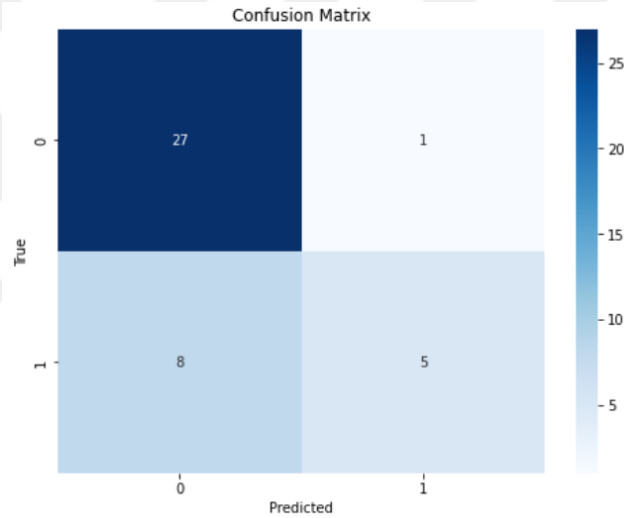
ML algoritmalarıyla oluşturulan sınıflandırma işlemi sonrasında çıkan karmaşıklık matris sonuçlarının gerçek pozitif değer (True Positive Rate - TPR) ve yanlış pozitif değerleri (False Positive Rate - FPR) arasında oluşturulan eğriye Receiver Operating Characteristic (ROC) denilmektedir. Veri setinde etiketli veriler içerisinde 1 olarak sınıflandırılmış olan verilerin model sonucuyla 1 olarak tespit edilmesine gerçek pozitif değer denilmektedir. TPR değeri grafikte y ekseninde gösterilmektedir. Veri setinde etiketlenen verilerin yanlış olarak tespit edilmesi ise yanlış pozitif değer denilmektedir. FPR değeri grafikte x ekseninde gösterilmektedir. FPR değerinin 0, TPR değerinin 1 olması oluşturulan modelin en mükemmel sonucu elde edildiğini göstermektedir. Bu durumu yakalamak oldukça zordur. FPR değerinin 1, TPR değerinin ise 0 olması sınıflandırmanın en kötü sonucudur.

Şekil 4.1'de gösterilen mavi çizgi rastgele sınıflandırma (Random Classifier) çizgisi olarak adlandırılmaktadır. Başarılı bir sınıflandırma modelinin eğrisi bu çizginin üst bölgesinde yani TPR 1 noktasına doğru olması beklenir. ROC eğrilerinin FPR 0 noktasına doğru altında kalan alanları ise Area Under Curve (AUC) olarak isimlendirilir. AUC değeri yüksek olan eğriler daha başarılı sayılır. En yüksek AUC değeri ise 1'dir. Şekil 4.1'de KNN sınıflandırma algoritmasına ait ROC eğrisi verilmiştir.



Şekil 4.1. KNN sınıflandırma algoritmasına ait ROC eğrisi

KNN sınıflandırma algoritmasında kullanılan beşli çapraz doğrulamanın 5. fold değeri (0,78) sonucunun karmaşıklık matrisi Şekil 4.2’de verilmiştir.



Şekil 4.2. KNN sınıflandırma algoritmasına ait karmaşıklık matrisi

Çizelge 4.2’de KNN sınıflandırma algoritmasına ait doğruluk, kesinlik, duyarlılık, makro ve ağırlıklı ortalamalar vb. performans sonuçları sunulmuştur.

Çizelge 4.2. KNN sınıflandırma algoritmasına ait performans sonuçları

	Precision	Recall	F1-Score	Support
1	0.77	0.96	0.86	28
2	0.83	0.38	0.53	13
Accuracy	0.81	0.81	0.81	0.81
Macro AVG	0.80	0.67	0.69	41
Weighted AVG	0.79	0.78	0.75	41

4.1.2. Karar Ağacı algoritması sonuçları

Kök (root), düğüm (node), yaprak (leaf) elemanlarından oluşan DT algoritması için veri seti ilk olarak %80 eğitim ve %20 test verisi olarak bölünmüştür. Bu işlem sonrasında doğruluk değeri 0,95 olarak hesaplanmıştır. Oluşturulan modelin güvenilirliğini arttırmak için beşli çapraz doğrulama uygulanmıştır. Modelin oluşturulması için Sklearn kütüphanesinde yer alan DecisionTreeClassifier kütüphanesi eklenmiştir.

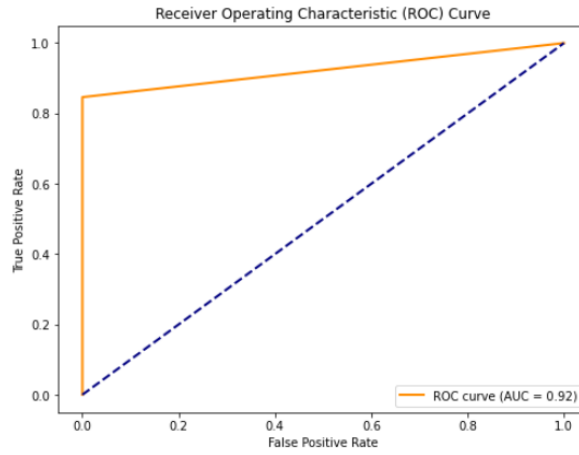
DT algoritması için uygulanan 5 sonuçlu çapraz doğrulama (cross validation) sonuçlarına bakıldığında tüm test verilerinin doğru tahmin edilmesi fold 4 de en yüksek değer olarak 1.0 sonucu bulunmuştur, tüm sonuçların ortalamasına bakıldığında ise 0,95 sonucu hesaplanmıştır. Bu modelin sonuçları Çizelge 4.3’de verilmiştir.

Çizelge 4.3. DT algoritması çapraz doğrulama (5 fold) uygulama sonuçları

5 li Çapraz Doğrulama Yapılan DT Algoritmasının Sonuçları				
Fold1	Fold2	Fold3	Fold4	Fold5
0.97	0.92	0.95	1.0	0.95
Ortalama Sonuç: 0.95				

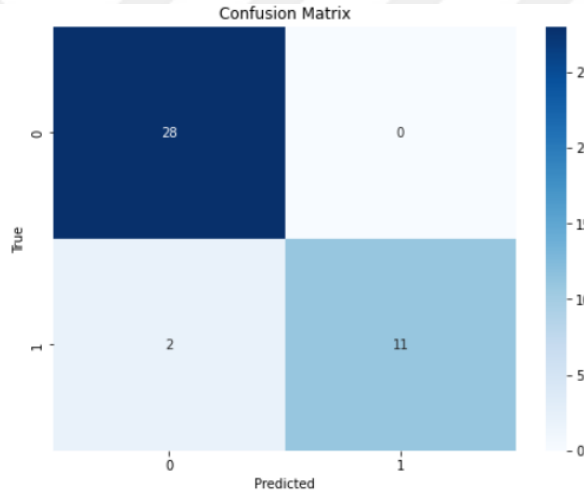
Beşli çapraz doğrulama için ayrılan test ve eğitim verilerini temel alarak oluşturulan DT algoritmasına ait sonuçları ayrı ayrı hesaplanarak dizi içerisine kaydedilmiştir. Döngünün tamamlanmasıyla dizide kayıtlı sonuçların numpy kütüphanesinin mean komutu ile ortalaması hesaplanmıştır.

DT algoritması için TPR ve FPR değerlerine bağlı olarak oluşturulan ROC eğrisi Şekil 4.3’de verilmiştir. Eğrinin altında kalan alan ise 0,92 olarak hesaplanmıştır.



Şekil 4.3. DT sınıflandırma algoritmasına ait ROC eğrisi

DT sınıflandırma algoritmasında kullanılan beşli çapraz doğrulamanın 5. fold değeri (0,95) sonucunun karmaşıklık matrisi Şekil 4.4’de verilmiştir.



Şekil 4.4. DT sınıflandırma algoritmasına ait karmaşıklık matrisi

Çizelge 4.4’de DT sınıflandırma algoritmasına ait doğruluk, kesinlik, duyarlılık, makro ve ağırlıklı ortalamalar vb. performans sonuçları sunulmuştur.

Çizelge 4.4. DT sınıflandırma algoritmasına ait performans sonuçları

	Precision	Recall	F1-Score	Support
1	0.93	1.00	0.97	28
2	1.00	0.85	0.92	13
Accuracy	0.95	0.95	0.95	0.95
Macro AVG	0.97	0.92	0.94	41
Weighted AVG	0.95	0.95	0.95	41

4.1.3. Lojistik Regresyon algoritması sonuçları

LR algoritması için veri seti ilk olarak %80 eğitim ve %20 test verisi olarak bölünmüştür. Bu işlem sonrasında doğruluk değeri 0,90 olarak hesaplanmıştır. Oluşturulan modelin güvenilirliğini arttırmak için beşli çapraz doğrulama uygulanmıştır. Modelin oluşturulması için Sklearn kütüphanesinde yer alan LogisticRegression kütüphanesi eklenmiştir.

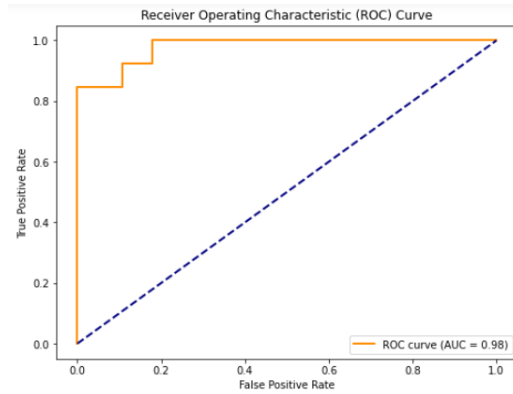
LR algoritması için uygulanan 5 sonuçlu çapraz doğrulama (cross validation) sonuçlarına bakıldığında en yüksek değer olarak sonuç fold 2'de 0.92 olarak bulunurken, tüm sonuçların ortalamasına bakıldığında ise 0.90 sonucu hesaplanmıştır. Bu modelin sonuçları Çizelge 4.5'de verilmiştir.

Çizelge 4.5. LR algoritması çapraz doğrulama (5 fold) uygulama sonuçları

5 li Çapraz Doğrulama Yapılan LR Algoritmasının Sonuçları				
Fold1	Fold2	Fold3	Fold4	Fold5
0.87	0.92	0.90	0.90	0.90
Ortalama Sonuç: 0.90				

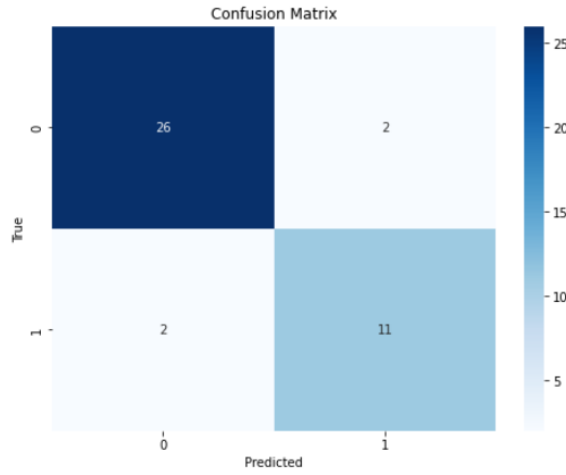
Beşli çapraz doğrulama için ayrılan test ve eğitim verilerini temel alarak oluşturulan LR algoritmasına ait sonuçları ayrı ayrı hesaplanarak dizi içerisine kaydedilmiştir. Döngünün tamamlanmasıyla dizide kayıtlı sonuçların numpy kütüphanesinin mean komutu ile ortalaması hesaplanmıştır.

LR algoritması için TPR ve FPR değerlerine bağlı olarak oluşturulan ROC eğrisi Şekil 4.5'de verilmiştir. Eğrinin altında kalan alan ise 0,98 olarak hesaplanmıştır.



Şekil 4.5. LR sınıflandırma algoritmasına ait ROC eğrisi

LR sınıflandırma algoritmasında kullanılan beşli çapraz doğrulamanın 5. fold değeri (0,90) sonucunun karmaşıklık matrisi Şekil 4.6'da verilmiştir.



Şekil 4.6. LR sınıflandırma algoritmasına ait karmaşıklık matrisi

Çizelge 4.6'da LR sınıflandırma algoritmasına ait doğruluk, kesinlik, duyarlılık, makro ve ağırlıklı ortalamalar vb. performans sonuçları sunulmuştur.

Çizelge 4.6. LR sınıflandırma algoritmasına ait performans sonuçları

	Precision	Recall	F1-Score	Support
1	0.93	0.93	0.93	28
2	0.85	0.85	0.85	13
Accuracy	0.90	0.90	0.90	0.90
Macro AVG	0.89	0.89	0.89	41
Weighted AVG	0.90	0.90	0.90	41

4.1.4. Naive Bayes algoritması sonuçları

NB algoritması için veri seti ilk olarak %80 eğitim ve %20 test verisi olarak bölünmüştür. Bu işlem sonrasında doğruluk değeri 0,80 olarak hesaplanmıştır. Oluşturulan modelin güvenilirliğini arttırmak için beşli çapraz doğrulama uygulanmıştır. Modelin oluşturulması için Sklearn kütüphanesinde yer alan NaiveBayes GaussianNB kütüphanesi eklenmiştir.

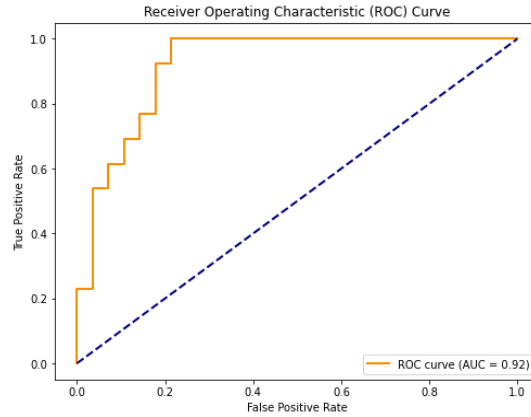
NB algoritması için uygulanan 5 sonuçlu çapraz doğrulama (cross validation) sonuçlarına bakıldığında en yüksek değer olarak sonuç 2’de 0.87 tüm sonuçların ortalamasına bakıldığında ise 0.80 sonucu hesaplanmıştır. Bu modelin sonuçları Çizelge 4.7’de verilmiştir.

Çizelge 4.7. NB algoritması çapraz doğrulama (5 fold) uygulama sonuçları

5 li Çapraz Doğrulama Yapılan NB Algoritmasının Sonuçları				
Fold1	Fold2	Fold3	Fold4	Fold5
0.80	0.87	0.82	0.75	0.80
Ortalama Sonuç: 0.80				

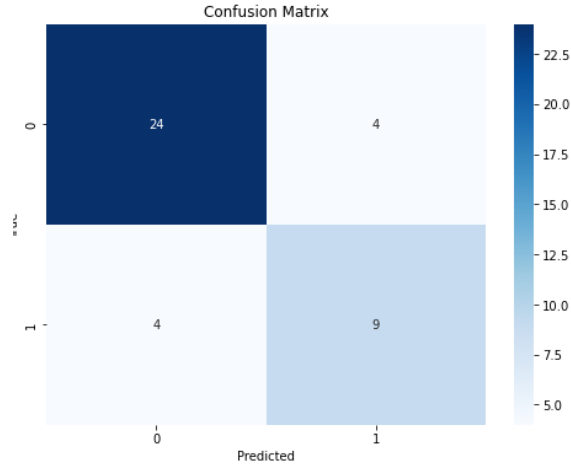
Beşli çapraz doğrulama için ayrılan test ve eğitim verilerini temel alarak oluşturulan NB algoritmasına ait sonuçları ayrı ayrı hesaplanarak dizi içerisine kaydedilmiştir. Döngünün tamamlanmasıyla dizide kayıtlı sonuçların numpy kütüphanesinin mean komutu ile ortalaması hesaplanmıştır.

NB algoritması için TPR ve FPR değerlerine bağlı olarak oluşturulan ROC eğrisi Şekil 4.7’de verilmiştir. Eğrinin altında kalan alan ise 0,92 olarak hesaplanmıştır.



Şekil 4.7. NB sınıflandırma algoritmasına ait ROC eğrisi

NB sınıflandırma algoritmasında kullanılan beşli çapraz doğrulamanın 5. fold değeri (0,80) sonucunun karmaşıklık matrisi Şekil 4.8’de verilmiştir.



Şekil 4.8. NB sınıflandırma algoritmasına ait karmaşıklık matrisi

Çizelge 4.8’de NB sınıflandırma algoritmasına ait doğruluk, kesinlik, duyarlılık, makro ve ağırlıklı ortalamalar vb. performans sonuçları sunulmuştur.

Çizelge 4.8. NB sınıflandırma algoritmasına ait performans sonuçları

	Precision	Recall	F1-Score	Support
1	0.86	0.86	0.86	28
2	0.69	0.69	0.69	13
Accuracy	0.80	0.80	0.80	0.80
Macro AVG	0.77	0.77	0.77	41
Weighted AVG	0.80	0.80	0.80	41

4.1.5. Rastgele Orman algoritması sonuçları

RF algoritması için veri seti ilk olarak %80 eğitim ve %20 test verisi olarak bölünmüştür. Bu işlem sonrasında doğruluk değeri 0,93 olarak hesaplanmıştır. Oluşturulan modelin güvenilirliğini arttırmak için beşli çapraz doğrulama uygulanmıştır. Modelin oluşturulması için Sklearn kütüphanesinde yer alan RandomForestClassifier kütüphanesi eklenmiştir.

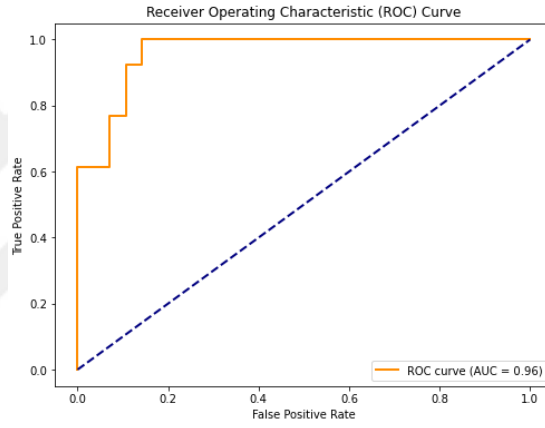
RF algoritması için uygulanan 5 sonuçlu çapraz doğrulama (cross validation) sonuçlarına bakıldığında en yüksek değer olarak sonuç 2’de tüm test verileri doğru tespit ederek 1.0 değeri bulunmuştur, tüm sonuçların ortalamasına bakıldığında ise 0.93 sonucu hesaplanmıştır. Bu modelin sonuçları Çizelge 4.9’da verilmiştir.

Çizelge 4.9. RF algoritması çapraz doğrulama (5 fold) uygulama sonuçları

5 li Çapraz Doğrulama Yapılan RF Algoritmasının Sonuçları				
Fold1	Fold2	Fold3	Fold4	Fold5
0.92	1.0	0.92	0.95	0.85
Ortalama Sonuç: 0.93				

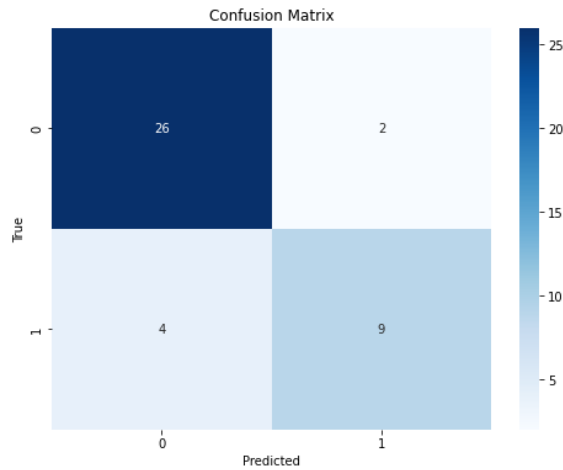
Beşli çapraz doğrulama için ayrılan test ve eğitim verilerini temel olarak oluşturulan RF algoritmasına ait sonuçları ayrı ayrı hesaplanarak dizi içerisine kaydedilmiştir. Döngünün tamamlanmasıyla dizide kayıtlı sonuçların numpy kütüphanesinin mean komutu ile ortalaması hesaplanmıştır.

RF algoritması için TPR ve FPR değerlerine bağlı olarak oluşturulan ROC eğrisi Şekil 4.9'da verilmiştir. Eğrinin altında kalan alan ise 0,96 olarak hesaplanmıştır.



Şekil 4.9. RF sınıflandırma algoritmasına ait ROC eğrisi

RF sınıflandırma algoritmasında kullanılan beşli çapraz doğrulamanın 5. fold değeri (0,85) sonucunun karmaşıklık matrisi Şekil 4.10'da verilmiştir.



Şekil 4.10. RF sınıflandırma algoritmasına ait karmaşıklık matrisi

Çizelge 4.10'da RF sınıflandırma algoritmasına ait doğruluk, kesinlik, duyarlılık, makro ve ağırlıklı ortalamalar vb. performans sonuçları sunulmuştur.

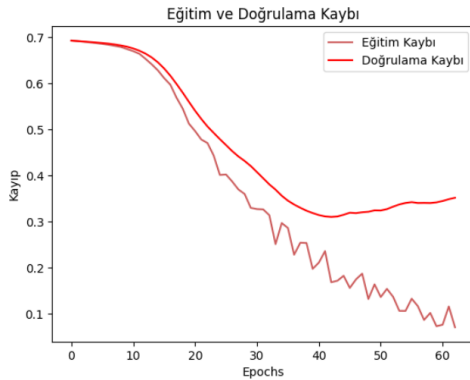
Çizelge 4.10. RF sınıflandırma algoritmasına ait performans sonuçları

	Precision	Recall	F1-Score	Support
1	0.87	0.93	0.90	28
2	0.82	0.69	0.75	13
Accuracy	0.85	0.85	0.85	0.85
Macro AVG	0.84	0.81	0.82	41
Weighted AVG	0.85	0.85	0.85	41

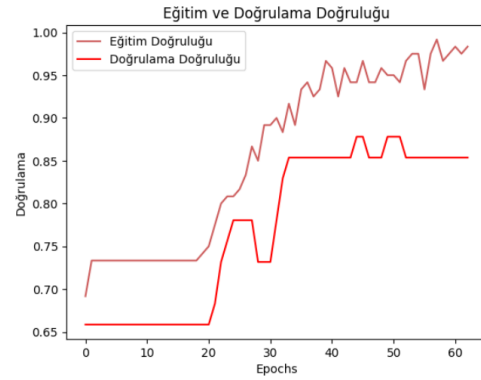
4.1.6. Yapay Sinir Ağları algoritması sonuçları

İnsan beyninin bilgi işleme tekniğini temel alan yapılar olan YSA, biyolojik nöron hücrelerinin kendi aralarında kurduğu ağı dijital olarak modellenmesidir. Katmanlar halinde kurulmuş nöral networkler giriş katmanı, gizli katman(lar) ve çıkış katmanlardan oluşmaktadır. Çalışmada Jupyter Notebook bünyesinde Keras ve Tensorflow kütüphaneleri kullanılmıştır.

Keras modeline ait sequential komutu ve aktivasyon komutlarından relu metodu ile 1 giriş katmanı, 6 adet nöron içeren 2 adet gizli katman ve bir çıkış katman içeren yapay sinir ağı oluşturulmuştur. 0 ile 1 değerleri arasında değer alan relu aktivasyon fonksiyonu, giriş katmanındaki değer ile relu fonksiyonundaki kat sayının çarpımı sonucu 0.5'in altında olması halinde çıkışı 0, eşit ve üstünde olması halinde ise 1 olarak hesaplamaktadır. Çıkış katmanı 0 ve 1 değerlerini içereceği için bu katmanı sigmoid olarak belirlenmiştir. Şekil 4.11'de eğitim ve doğrulama kaybını gösteren eğri ve Şekil 4.12'de eğitim ve doğruluk eğri verilmiştir.

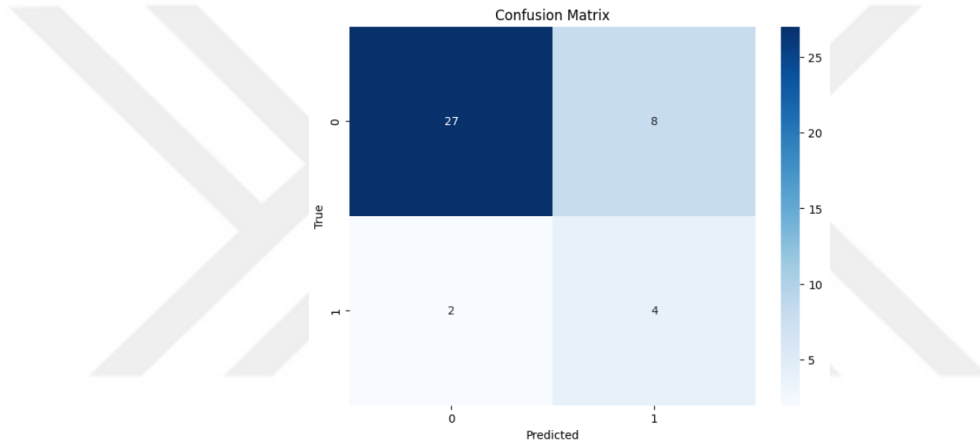


Şekil 4.11. Eğitim ve Doğrulama Kaybı



Şekil 4.12. Eğitim ve Doğrulama Eğrisi

Yerel networkün ‘adam’ modeli ile optimize edilirken; sonucun 0 ile 1 arasında olması için loss parametresi ‘binary_crossentropy’, doğruluk değerini hesaplamak için de metrik parametresi ‘accuracy’ olarak model tanımlanmıştır. Oluşturulan bu modele veri setine ait eğitim ve test olarak ayrılmış verilerle doldurulmuştur. YSA öğrenme sürecinin ve hataların en aza indirilmesi için gerekli tekrar komutu ‘epoch’ parametresi 100 olarak belirlenmiştir. Bu süreçte normal değer ile tahmin edilen değer 100 kez karşılaştırılıyor ve en iyi tahminleme modeli oluşturuluyor. Test verilerinin tahmin etme işlemiyle hesaplanan sonuçların 0.5 değerinden düşük olması durumunda çıkışın 0, eşit ve yüksek olması durumunda ise 1 olarak eklenmiştir. Karmaşıklık matrisi kullanılarak doğruluk değerleri hesaplanmıştır. YSA algoritması sonucunda hesaplanan karmaşıklık matrisi Şekil 4.13’de verilmiştir.



Şekil 4.13. YSA algoritmasına ait karmaşıklık matrisi

Çizelge 4.11’de YSA algoritmasına ait doğruluk, kesinlik, duyarlılık, makro ve ağırlıklı ortalamalar vb. performans sonuçları sunulmuştur.

Çizelge 4.11. YSA algoritmasına ait performans sonuçları

	Precision	Recall	F1-Score	Support
1	0.93	0.77	0.84	35
2	0.69	0.69	0.69	6
Accuracy	0.76	0.76	0.76	41
Macro AVG	0.63	0.72	0.64	41
Weighted AVG	0.84	0.76	0.79	41

4.2. Hazır Veri Setinin Sonuçları

Bu bölümde Kaliforniya Üniversitesinin “Heart Failure Prediction” isimli veri seti üzerinde YSA, KNN, DT, LR, NB, RF algoritmalarını kullanarak analiz yapılmıştır. Güvenilirliği arttırmak için beşli çapraz doğrulama uygulanmıştır. Her algoritma modeline ait ROC eğrisi ve değerlendirme ölçütü sonuçları verilmiştir. 299 örnek ve 13 öznitelikten oluşan veri seti üzerinde YSA, KNN, DT, LR, NB, RF algoritmaları için analizleri ve değerlendirme ölçütleri bu başlık altında verilmiştir. Nevşehir Devlet Hastanesi hastalarından tarafınca toplanarak oluşturulan veri seti analiz başlıklarında ilgili algoritmaların örnek komutlarıyla aynı algoritma kullanıldığı için örnek komutlar tekrar verilmemiştir.

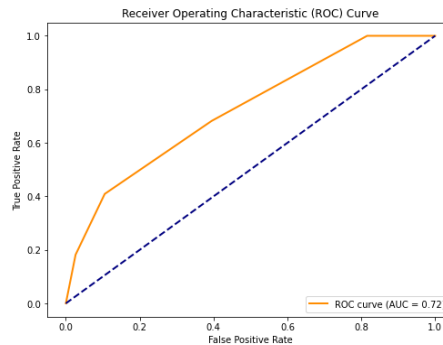
4.2.1. K En Yakın Komşu algoritması sonuçları

KNN algoritması için uygulanan 5 sonuçlu çapraz doğrulama (cross validation) sonuçlarına bakıldığında en yüksek değer olarak sonuç 1’de 0.73 tüm sonuçların ortalamasına bakıldığında ise 0.70 sonucu hesaplanmıştır. Bu modelin sonuçları Çizelge 4.12’de verilmiştir.

Çizelge 4.12. KNN algoritması çapraz doğrulama (5 fold) uygulama sonuçları

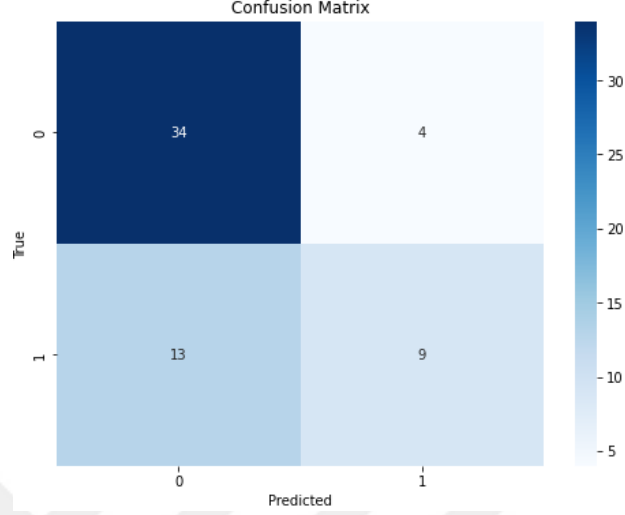
5’li Çapraz Doğrulama Yapılan KNN Algoritmasının Sonuçları				
Fold1	Fold2	Fold3	Fold4	Fold5
0.73	0.68	0.71	0.65	0.71
Ortalama Sonuç: 0.70				

KNN algoritması için TPR ve FPR değerlerine bağlı olarak oluşturulan ROC eğrisi Şekil 4.14’de verilmiştir. Eğrinin altında kalan alan ise 0,72 olarak hesaplanmıştır.



Şekil 4.14. KNN sınıflandırma algoritmasına ait ROC eğrisi

KNN sınıflandırma algoritmasında kullanılan beşli çapraz doğrulamanın 5. fold değeri (0,70) sonucunun karmaşıklık matrisi Şekil 4.15’de verilmiştir.



Şekil 4.15. KNN sınıflandırma algoritmasına ait karmaşıklık matrisi

Çizelge 4.13’de KNN sınıflandırma algoritmasına ait doğruluk, kesinlik, duyarlılık, makro ve ağırlıklı ortalamalar vb. performans sonuçları sunulmuştur.

Çizelge 4.13. KNN sınıflandırma algoritmasına ait performans sonuçları

	Precision	Recall	F1-Score	Support
1	0.72	0.89	0.80	38
2	0.69	0.41	0.51	22
Accuracy	0.71	0.71	0.71	0.71
Macro AVG	0.71	0.65	0.66	60
Weighted AVG	0.71	0.72	0.70	60

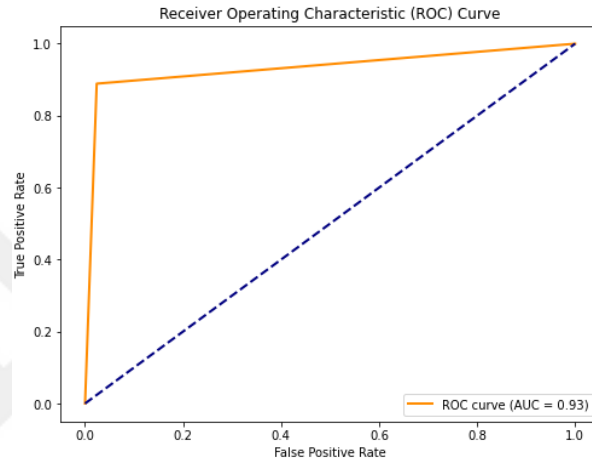
4.2.2. Karar Ağacı algoritması sonuçları

DT algoritması için uygulanan 5 sonuçlu çapraz doğrulama (cross validation) sonuçlarına bakıldığında en yüksek değer olarak sonuç 1, 3, 4, ve 5’de 0.95 tüm sonuçların ortalamasına bakıldığında ise 0.94 sonucu hesaplanmıştır. Bu modelin sonuçları Çizelge 4.14’de verilmiştir.

Çizelge 4.14. DT algoritması çapraz doğrulama (5 fold) uygulama sonuçları

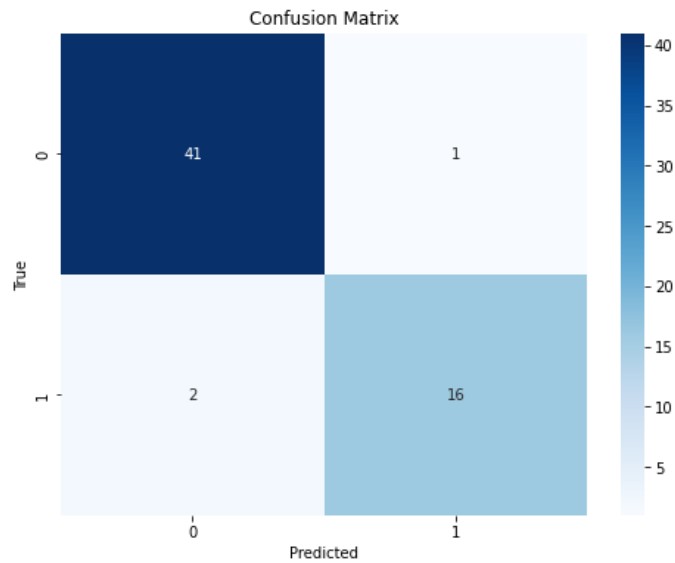
5'li Çapraz Doğrulama Yapılan DT Algoritmasının Sonuçları				
Fold1	Fold2	Fold3	Fold4	Fold5
0.95	0.90	0.95	0.95	0.95
Ortalama Sonuç: 0.94				

DT algoritması için TPR ve FPR değerlerine bağlı olarak oluşturulan ROC eğrisi Şekil 4.16'da verilmiştir. Eğrinin altında kalan alan ise 0,93 olarak hesaplanmıştır.



Şekil 4.16. DT sınıflandırma algoritmasına ait ROC eğrisi

DT sınıflandırma algoritmasında kullanılan beşli çapraz doğrulamanın 5. fold değeri (0.94) sonucunun karmaşıklık matrisi Şekil 4.17'de verilmiştir.



Şekil 4.17. DT sınıflandırma algoritmasına ait karmaşıklık matrisi

Çizelge 4.15’de DT sınıflandırma algoritmasına ait doğruluk, kesinlik, duyarlılık, makro ve ağırlıklı ortalamalar vb. performans sonuçları sunulmuştur.

Çizelge 4.15. DT sınıflandırma algoritmasına ait performans sonuçları

	Precision	Recall	F1-Score	Support
1	0.95	0.98	0.96	42
2	0.94	0.89	0.91	18
Accuracy	0.95	0.95	0.95	0.95
Macro AVG	0.95	0.93	0.94	60
Weighted AVG	0.95	0.95	0.95	60

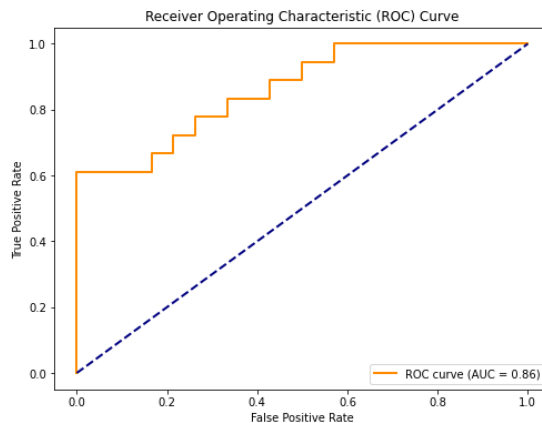
4.2.3. Lojistik Regresyon algoritması sonuçları

LR algoritması için uygulanan 5 sonuçlu çapraz doğrulama (cross validation) sonuçlarına bakıldığında en yüksek değer olarak sonuç 2, 4 ve 5’de 0.85 tüm sonuçların ortalamasına bakıldığında ise 0.84 sonucu hesaplanmıştır. Bu modelin sonuçları Çizelge 4.16’da verilmiştir.

Çizelge 4.16. LR algoritması çapraz doğrulama (5 fold) uygulama sonuçları

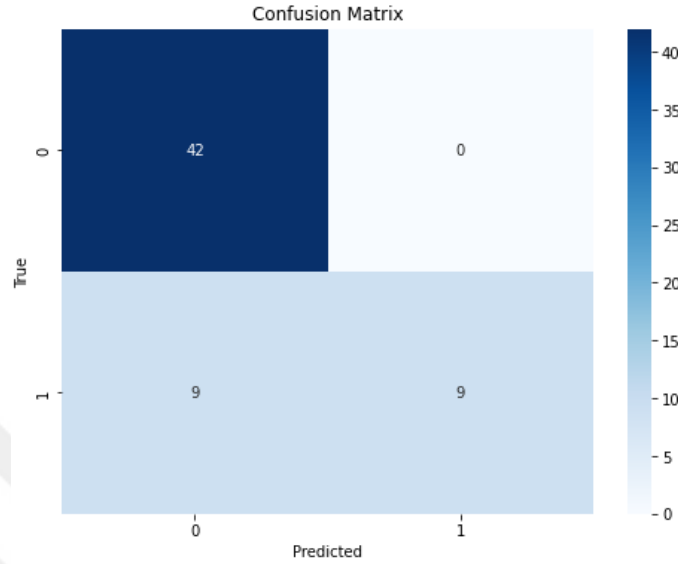
5’li Çapraz Doğrulama Yapılan LR Algoritmasının Sonuçları				
Fold1	Fold2	Fold3	Fold4	Fold5
0.83	0.85	0.81	0.85	0.85
Ortalama Sonuç: 0.84				

LR algoritması için TPR ve FPR değerlerine bağlı olarak oluşturulan ROC eğrisi Şekil 4.18’de verilmiştir. Eğrinin altında kalan alan ise 0,86 olarak hesaplanmıştır.



Şekil 4.18. LR sınıflandırma algoritmasına ait ROC eğrisi

LR sınıflandırma algoritmasında kullanılan beşli çapraz doğrulamanın 5. fold değeri (0.84) sonucunun karmaşıklık matrisi Şekil 4.19’de verilmiştir.



Şekil 4.19. LR sınıflandırma algoritmasına ait karmaşıklık matrisi

Çizelge 4.17’de LR sınıflandırma algoritmasına ait doğruluk, kesinlik, duyarlılık, makro ve ağırlıklı ortalamalar vb. performans sonuçları sunulmuştur.

Çizelge 4.17. LR sınıflandırma algoritmasına ait performans sonuçları

	Precision	Recall	F1-Score	Support
1	0.82	1.0	0.9	42
2	1.0	0.5	0.67	18
Accuracy	0.84	0.84	0.84	0.84
Macro AVG	0.91	0.75	0.78	60
Weighted AVG	0.88	0.85	0.83	60

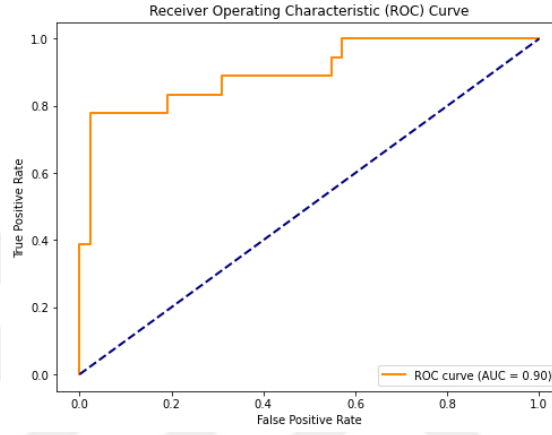
4.2.4. Naive Bayes algoritması sonuçları

NB algoritması için uygulanan 5 sonuçlu çapraz doğrulama (cross validation) sonuçlarına bakıldığında en yüksek değer olarak sonuç 2’de 0.85 tüm sonuçların ortalamasına bakıldığında ise 0.83 sonucu hesaplanmıştır. Bu modelin sonuçları Çizelge 4.18’de verilmiştir.

Çizelge 4.18. NB algoritması çapraz doğrulama (5 fold) uygulama sonuçları

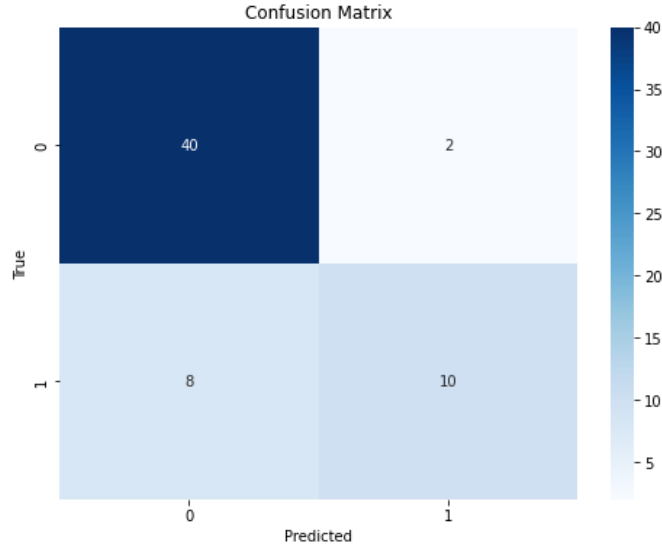
5'li Çapraz Doğrulama Yapılan NB Algoritmasının Sonuçları				
Fold1	Fold2	Fold3	Fold4	Fold5
0.83	0.85	0.83	0.83	0.83
Ortalama Sonuç: 0.83				

NB algoritması için TPR ve FPR değerlerine bağlı olarak oluşturulan ROC eğrisi Şekil 4.20'de verilmiştir. Eğrinin altında kalan alan ise 0,90 olarak hesaplanmıştır.



Şekil 4.20. NB sınıflandırma algoritmasına ait ROC eğrisi

NB sınıflandırma algoritmasında kullanılan beşli çapraz doğrulamanın 5. fold değeri (0,83) sonucunun karmaşıklık matrisi Şekil 4.21'de verilmiştir.



Şekil 4.21. NB sınıflandırma algoritmasına ait karmaşıklık matrisi

Çizelge 4.19’da NB sınıflandırma algoritmasına ait doğruluk, kesinlik, duyarlılık, makro ve ağırlıklı ortalamalar vb. performans sonuçları sunulmuştur.

Çizelge 4.19. NB sınıflandırma algoritmasına ait performans sonuçları

	Precision	Recall	F1-Score	Support
1	0.83	0.95	0.89	42
2	0.83	0.56	0.67	18
Accuracy	0.84	0.84	0.84	0.84
Macro AVG	0.83	0.75	0.78	60
Weighted AVG	0.83	0.83	0.82	60

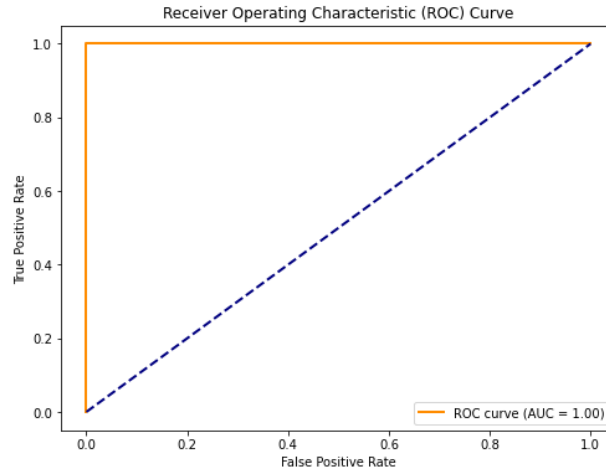
4.2.5. Rastgele Orman algoritması sonuçları

RF algoritması için uygulanan 5 sonuçlu çapraz doğrulama (cross validation) sonuçlarına bakıldığında en yüksek değer olarak sonuç 1 ve 5’de 1.0 tüm sonuçların ortalamasına bakıldığında ise 0.97 sonucu hesaplanmıştır. Bu modelin sonuçları Çizelge 4.20’de verilmiştir.

Çizelge 4.20. RF algoritması çapraz doğrulama (5 fold) uygulama sonuçları

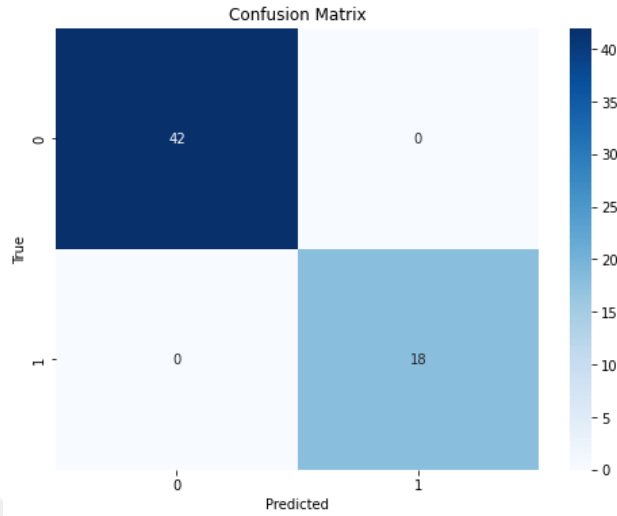
5’li Çapraz Doğrulama Yapılan RF Algoritmasının Sonuçları				
Fold1	Fold2	Fold3	Fold4	Fold5
1.0	0.93	0.96	0.96	1.0
Ortalama Sonuç: 0.97				

RF algoritması için TPR ve FPR değerlerine bağlı olarak oluşturulan ROC eğrisi Şekil 4.22’de verilmiştir. Eğrinin altında kalan alan ise 1.0 olarak hesaplanmıştır.



Şekil 4.22. RF sınıflandırma algoritmasına ait ROC eğrisi

RF sınıflandırma algoritmasında kullanılan beşli çapraz doğrulamanın 5. fold değeri (1.0) sonucunun karmaşıklık matrisi Şekil 4.23’de verilmiştir.



Şekil 4.23. RF sınıflandırma algoritmasına ait karmaşıklık matrisi

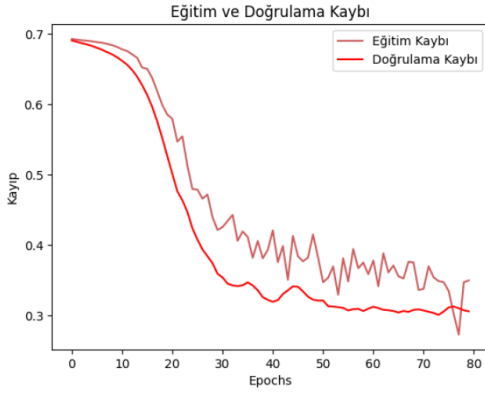
Çizelge 4.21’de RF sınıflandırma algoritmasına ait doğruluk, kesinlik, duyarlılık, makro ve ağırlıklı ortalamalar vb. performans sonuçları sunulmuştur.

Çizelge 4.21. RF sınıflandırma algoritmasına ait performans sonuçları

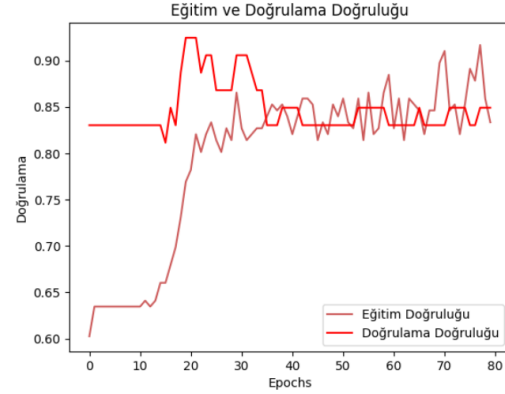
	Precision	Recall	F1-Score	Support
1	1.0	1.0	1.0	42
2	1.0	1.0	1.0	18
Accuracy	1.0	1.0	1.0	1.0
Macro AVG	1.0	1.0	1.0	60
Weighted AVG	1.0	1.0	1.0	60

4.2.6. Yapay Sinir Ağları algoritması sonuçları

Bir giriş katmanı iki gizli katman ve bir çıkış katmandan oluşan YSA modeli relu aktivasyon komutu ile oluşturulmuştur. Giriş katmanındaki değer ile relu fonksiyonundaki kat sayının çarpımı sonucu 0.5’in altında olması halinde çıkışı 0, eşit ve üstünde olması halinde ise 1 olarak hesaplamaktadır. Çıkış katmanı 0 ve 1 değerlerini içereceği için bu katmanı sigmoid olarak belirlenmiştir. Şekil 4.24’de eğitim ve doğrulama kaybını gösteren eğri ve Şekil 4.25’de eğitim ve doğruluk eğri verilmiştir.

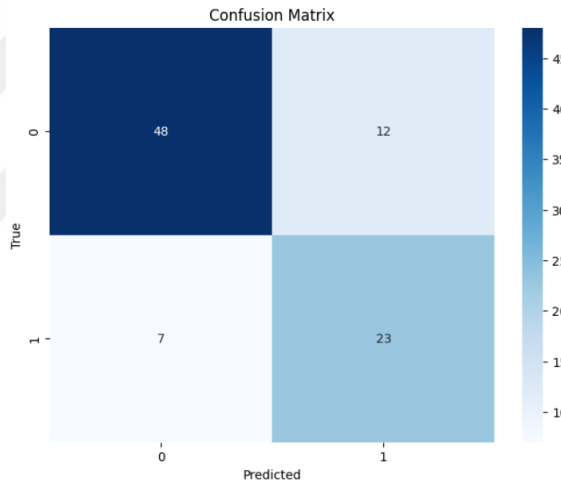


Şekil 4.24. Eğitim ve Doğrulama Kaybı



Şekil 4.25. Eğitim ve Doğrulama Eğrisi

YSA modeli adam algoritması ile optimize edilmiştir. Parametre ve değerleri şu şekildedir: loss - binary_crossentropy, metrik - accuracy, epoch - 100 olarak belirlenmiştir. Karmaşıklık matrisi kullanılarak doğruluk değerleri hesaplanmıştır. YSA algoritması sonucunda hesaplanan karmaşıklık matrisi Şekil 4.26'da verilmiştir.



Şekil 4.26. YSA algoritmasına ait karmaşıklık matrisi

Çizelge 4.22'de YSA algoritmasına ait doğruluk, kesinlik, duyarlılık, makro ve ağırlıklı ortalamalar vb. performans sonuçları sunulmuştur.

Çizelge 4.22. YSA algoritmasına ait performans sonuçları

	Precision	Recall	F1-Score	Support
1	0.87	0.80	0.83	60
2	0.66	0.77	0.71	30
Accuracy	0.79	0.79	0.79	90
Macro AVG	0.76	0.78	0.77	90
Weighted AVG	0.80	0.79	0.79	90

4.3. Tartışma

Bu tez çalışmasında Nevşehir Devlet Hastanesi Kardiyoloji bölümünde Eylül 2022'den Mart 2023 tarihleri arasında kalp rahatsızlığı şikayetiyle gelen hastalardan oluşturulan bu veri seti üzerinde ML algoritmaları çalışılmıştır. Bu algoritmalarından elde edilen sonuçlar değerlendirilirken aynı zamanda algoritmaların doğruluk değerleri karşılaştırılmıştır.

Oluşturulan modellerin doğruluk değerlerindeki güvenilirliği arttırmak adına tez çalışmasında kullanılan YSA, DT, KNN, LR, NB, RF algoritmalarının tamamına beşli çapraz doğrulama uygulanmıştır. Veri seti örneklerin veri ön işleme adımları sonrasında bu beş algoritma için ayrı ayrı beşli çapraz doğrulama ile oluşturulan modellere ait ortalama doğruluk değerleri DT %95, RF %93, LR %90, KNN %81, NB %80, YSA %76 olarak hesaplanmıştır ve yüksekte düşüğe doğru sıralanmıştır. Oluşturulan modellerden elde edilen performans sonuçları Çizelge 4.23'de verilmiştir.

ML algoritmalarından en başarılı doğruluk değerine karar ağacı algoritmasından elde edilse de diğer sınıflandırma algoritmalarının da başarılı sonuç elde edildiği görülmüştür. Literatür taramasından elde edilen bilgiler ışığında çalışmaların RF algoritmasının ön plana çıktığı görülmüştür. RF algoritmasının karmaşık ve çok örnekli veri setlerinde diğer algoritmalara göre daha iyi sonuç verdiğini, veri setinin 202 örnekten oluştuğunu düşünürsek DT algoritmasının ön plana çıkma sebebi olarak yorumlanmıştır.

Tez çalışmasında oluşturulan veri setinin yanı sıra Kaliforniya Üniversitesinin veri seti arşiv web sayfasından elde edilebildiği gibi kaggle web sayfasından da elde edilebilen "*Heart Failure Prediction*" başlıklı hazır veri seti de kullanılmıştır. Bu veri setine ait veri birleştirme, veri çoğaltma, özniteliklere ait eksik veri olmadığı için veri ön işleme sürecine gidilmemiştir. Nevşehir Devlet Hastanesi'nden toplanarak oluşturulan veri setine uygulanan beşli çapraz doğrulama ve ML algoritmalarının tamamı bu veri seti için de uygulanmıştır. Bu algoritmalar için ayrı ayrı beşli çapraz doğrulama ile oluşturulan modellere ait ortalama doğruluk değerleri RF %97, DT %94, LR %84, NB %83, YSA %79, KNN %70 olarak hesaplanmıştır ve yüksekte düşüğe doğru sıralanmıştır. Heart Failure Prediction veri setine ait oluşturulan modellerden elde edilen performans sonuçları Çizelge 4.24'de verilmiştir.

Çizelge 4.23. Oluşturulan veri setine ait tüm performans sonuçları

Algoritma	Değerlendirme Kriterleri	Precision	Recall	F1-Score	Support
DT	1	0.93	1.00	0.97	28
	2	1.00	0.85	0.92	13
	Accuracy	0.95	0.95	0.95	0.95
	Macro AVG	0.97	0.92	0.94	41
	Weighted AVG	0.95	0.95	0.95	41
RF	1	0.87	0.93	0.90	28
	2	0.82	0.69	0.75	13
	Accuracy	0.85	0.85	0.85	0.85
	Macro AVG	0.84	0.81	0.82	41
	Weighted AVG	0.85	0.85	0.85	41
LR	1	0.93	0.93	0.93	28
	2	0.85	0.85	0.85	13
	Accuracy	0.90	0.90	0.90	0.90
	Macro AVG	0.89	0.89	0.89	41
	Weighted AVG	0.90	0.90	0.90	41
KNN	1	0.77	0.96	0.86	28
	2	0.83	0.38	0.53	13
	Accuracy	0.81	0.81	0.81	0.81
	Macro AVG	0.80	0.67	0.69	41
	Weighted AVG	0.79	0.78	0.75	41
NB	1	0.86	0.86	0.86	28
	2	0.69	0.69	0.69	13
	Accuracy	0.80	0.80	0.80	0.80
	Macro AVG	0.77	0.77	0.77	41
	Weighted AVG	0.80	0.80	0.80	41
YSA	1	0.93	0.77	0.84	35
	2	0.69	0.69	0.69	6
	Accuracy	0.76	0.76	0.76	41
	Macro AVG	0.63	0.72	0.64	41
	Weighted AVG	0.84	0.76	0.79	41

Çizelge 4.24. Heart Failure Prediction veri setine ait tüm performans sonuçları

Algoritma	Değerlendirme Kriterleri	Precision	Recall	F1-Score	Support
DT	1	0.95	0.98	0.96	42
	2	0.94	0.89	0.91	18
	Accuracy	0.95	0.95	0.95	0.95
	Macro AVG	0.95	0.93	0.94	60
	Weighted AVG	0.95	0.95	0.95	60
RF	1	1.0	1.0	1.0	42
	2	1.0	1.0	1.0	18
	Accuracy	1.0	1.0	1.0	1.0
	Macro AVG	1.0	1.0	1.0	60
	Weighted AVG	1.0	1.0	1.0	60
LR	1	0.82	1.0	0.9	42
	2	1.0	0.5	0.67	18
	Accuracy	0.84	0.84	0.84	0.84
	Macro AVG	0.91	0.75	0.78	60
	Weighted AVG	0.88	0.85	0.83	60
KNN	1	0.72	0.89	0.80	38
	2	0.69	0.41	0.51	22
	Accuracy	0.71	0.71	0.71	0.71
	Macro AVG	0.71	0.65	0.66	60
	Weighted AVG	0.71	0.72	0.70	60
NB	1	0.83	0.95	0.89	42
	2	0.83	0.56	0.67	18
	Accuracy	0.84	0.84	0.84	0.84
	Macro AVG	0.83	0.75	0.78	60
	Weighted AVG	0.83	0.83	0.82	60
YSA	1	0.87	0.80	0.83	60
	2	0.66	0.77	0.71	30
	Accuracy	0.79	0.79	0.79	90
	Macro AVG	0.76	0.78	0.77	90
	Weighted AVG	0.80	0.79	0.79	90

Heart Failure Prediction isimli hazır veri üzerinde çalışma yapan arařtırmacılar, kullandıkları yöntemler, en iyi sonuç veren yöntem ve karmařıklık matris sonuçları Çizelge 4.25’de verilmiřtir.

Çizelge 4.25. Heart failure prediction veri seti üzerinde çalışma yapan arařtırmacılar ve sonuçları

Veri Ön İşleme	Kullanılan Yöntem	En İyi Sonuç Veren Yöntem	Doğruluk (Accuracy)	Duyarlılık (Recall)	F1-Score	Keskinlik (Precision)	ROC	Referans
-	KNN, AdaBoot, LightGBM, XGBoost, RF, Classification and Regression Tree (CART), LR, CatBoost	RF	0.99	1.00	0.98	1.00	0.99	Gürgen ve Serttaş, 2023
Standardizasyon	RF, YSA ve GSB	YSA	0.86	0.86	0.86	86.9	0.89	Keser ve Keskin, 2023
Normalizasyon ve SMOTE	Attribute Selected Classifier (ASC), Random Tree (RT), PART, NB, RF, CvR, J48, AdaBoost, OneR, JRip	RF	0.87	0.871	0.871	0.873	-	Altunbey Özbay ve Özbay, 2022
Öznitelik Seçimi	LR, NB, J48, SVM	SVM	0.9	-	0.86	-	-	Potur ve Erginel,2021
-	LR, RF, DT, ANN, SVM, KNN, NB,GB, YSA	RF	0.74	-	0.54	0.75	0.80	Chicco ve Jurman, 2020

5. SONUÇLAR VE ÖNERİLER

Bu bölümde kalp hastalıklarından kalp yetmezliği konusu üzerine sınıflandırma çalışmaları yapan araştırmacılara oluşturulan veri seti ve “*Heart Failure Prediction*” başlıklı veri setinde kullanılan yöntemler, elde edilen sonuçlar ve yeni çalışma alanları önerilmiştir.

5.1 Sonuçlar

Kalbin yapısal ve işlevsel bozukluğuna bağlı olarak yeterli oksijen miktarının gelmemesi, vücudun ihtiyacı olan kanın kalp tarafından yeterince pompalanamaması, kalbin kasılma işlemi ile gerçekleşen kanın pompalanma özelliğinde zayıflama ve çeşitli akut hastalıklarına bağlı olarak oluşan kalp yetmezliği yaşam kalitesini olumsuz etkilemektedir. KY teşhisinin yapılabilmesi ve dünyaca kabul gören evre ve sınıf derecelerinin tespiti için laboratuvar sonuçları, görüntüleme sonuçları, kişisel şikâyet bilgileri, aile geçmişinde genetik kalıtlar, bireydeki başka hastalıkların olup olmaması gibi bilgiler ışığında kalp yetmezliğinin teşhis edilmesi ciddi bir iş yükü olduğundan ve olası beşerî hataların en aza indirgemek amacıyla ML modelleri geliştirilmiştir.

Tez çalışmasında kalp yetmezliği sınıflandırma sürecinde kullanılmak üzere iki adet veri seti kullanılmıştır. Nevşehir Devlet Hastanesi Kardiyoloji Bölümünde Eylül 2022’den Mart 2023 tarihleri arasında E-26171210-929 sayılı etik kurul onayı ile kalp rahatsızlığı şikayetiyle gelen hasta kayıtlarından edinilmiştir ve etik kurul evrakları ekte sunulmuştur. Diğer veri seti ise Kaliforniya Üniversitesinin “<https://archive.ics.uci.edu>” adresinden edinilen “*Heart Failure Prediction*” başlıklı kalp yetmezliğine yönelik hazırlanan ve erişime açılan veri setidir. Oluşturulan veri setine veri ön işleme adımları uygulanarak veri seti analiz için hazır hale getirilmiştir. Oluşturulan veri setindeki örneklem sayısı 202, Heart Failure Prediction isimli veri seti örneklem sayısı 299 toplamda ise 501 adet örnek analiz edilmiştir.

Bu tez çalışmasıyla veri setleri üzerinde ML modelleri oluşturulmuştur. ML algoritmalarıyla oluşturulan modellerin doğruluk değerini arttırmak için beşli çapraz doğrulama uygulanmış ve ortalama sonuç modelin doğruluk değeri olarak kabul edilmiştir. Oluşturulan veri seti için kullanılan algoritmalar ve sonuçları; DT ile %95, RF ile %93, LR ile %90, KNN ile %81, NB ile %80, YSA %76 doğruluk değeri hesaplanmıştır. Heart Failure Prediction isimli hazır veri seti için kullanılan algoritmalar

ve sonuçları; RF ile %93, DT ile %94, LR ile %84, NB ile %83, YSA %79, KNN ile %70 doğruluk değeri hesaplanmıştır. Oluşturulan modeller ile başarılı sonuçlar elde edilmiştir.

5.2 Öneriler

KY tespit için önem arz eden farklı özellikler eklenerek analize dahil edilebilir. Verilerin toplanması için çalışmamda 6 ay gibi bir süre ayrılmıştır. Bu sürenin artırılarak büyükşehir devlet hastanesinden ya da daha fazla hasta sirkülasyonu olan şehir hastanelerinden veri toplanması ile örneklem sayısı artırılabilir. Oluşturulan veri seti bir web sayfasında paylaşarak bu alanda çalışma yapacak araştırmacılara materyal olarak sunulabilir. Veri ön işleme sürecinde doğal dil işleme teknikleri kullanılarak daha başarılı sonuçlar elde edilebilir. ML algoritmalarıyla daha iyi sonuçlar elde edebilmek için geliştirilen modellere daha hassas ayarlamalar yapılabilir. Çalışmada 6 farklı ML ve beşli çapraz doğrulama yapılmış olsa da farklı algoritmalar üzerinde de çalışılabilir. EKG, röntgen gibi görsel görseller üzerinde görüntü işleme alanında çalışılabilir.

KAYNAKLAR

- Ahmad, Tanvir, Assia Munir, Sajjad Haider Bhatti, Muhammad Aftab, and Muhammad Ali %J PloS one Raza. "Survival Analysis of Heart Failure Patients: A Case Study." 12, no. 7 (2017): e0181001.
- Ahmet, SEL %J İstatistik ve Uygulamalı Bilimler Dergisi. "Pandemi Sürecinde Altın Fiyatları İle Kripto Para İlişkisinin Makine Öğrenme Metotları İle İncelenmesi." 1, no. 2 (2020): 85-98.
- Al-Jada, Omer Saad Abdulqader. "Spectral Analysis of Ecg Signals for Pre-Diagnosis of Heart Diseases." Fen Bilimleri Enstitüsü, 2018.
- AYDIN, Ahmet %J Çukurova Üniversitesi Mühendislik Fakültesi dergisi. "Critical Parameter Selection and Survival Model Development for Heart Failure Patients." 36, no. 1 (2021): 155-62.
- Badik Şevval Tuğçe, and Mutlu Akar. "Kalp Yetmezliği Hasta Verilerinin Farklı Sınıflandırma Yöntemleriyle Analizi.", 2022.
- BEKTAŞ, BURCU, and Sebahattin BABUR. "Makine Öğrenmesi Teknikleri Kullanılarak Meme Kanseri Teşhisinin Performans Değerlendirmesi." (2016).
- Breiman, L.; Jerome H.F; Richard A.O; Charles J.S (1984), Classification And Regression Trees, Chapman & Hall/CRC Texts in Statistical Science Series
- Bilekyiğit, Sema, and Ayşe Eldem. "Kalp Yetmezliği Riskinin Makine Öğrenmesi Yöntemleri İle Analiz Edilmesi." Karamanoğlu Mehmetbey Üniversitesi, 2022.
- Bingöl, Kaan. "Depreme Dayanıklı Mimari Tasarım Aşamasında Derin Öğrenme ve Görüntü Sınıflama Yöntemi ile Burulma Düzensizliği Tespiti.", 2020.
- Chicco, Davide, Giuseppe %J BMC medical informatics Jurman, and decision making. "Machine Learning Can Predict Survival of Patients with Heart Failure from Serum Creatinine and Ejection Fraction Alone." 20 (2020): 1-16.
- Coşar, Mustafa, and Emre %J Avrupa Bilim ve Teknoloji Dergisi Deniz. "Makine Öğrenimi Algoritmaları Kullanarak Kalp Hastalıklarının Tespit Edilmesi." no. 28 (2021): 1112-16.
- Erdaş Adem, and Mehmet Serdar Güzel. "Medikal Veri Üzerinde Makine Öğrenmesi Yöntemlerinin Karşılaştırılması.", 2022.
- Görmez Buraki and Erkan Bostancı. "Adli bilişimde makine öğrenmesi: Makine öğrenmesi algoritmaları ile terör olaylarının tahmin edilmesi çalışması. " Ankara Üniversitesi, 2021.
- Gürgen, Gülay, Soydan %J Euroasia Journal of Mathematics Serttaş, Engineering, Natural, and Medical Sciences. "Kalp Yetmezliği Hastalığının Erken Teşhisinde

- Makine Öğrenimi Algoritmalarının Performans Karşılaştırması." 10, no. 28 (2023): 165-74.
- Güven İlhan, Pınar. "Düşük Ejeksiyon Fraksiyonlu Kalp Yetmezliği Olan Hastalarda Sarkopeni Parametrelerinin Değerlendirilmesi." (2021).
- Karaj, Amir. "Kardiyovasküler Hastalıklarının Teşhisine Yönelik Makine Öğrenmesi Algoritmaları İle Karar Destek Sistemi Tasarımı." (2023).
- KESER, Sinem BOZKURT, and Kemal %J Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi Keskin. "Kalp Yetmezliği Hastalarının Sağ Kalım Tahmini: Sınıflandırmaya Dayalı Makine Öğrenmesi Algoritmalarının Bir Uygulaması." 23, no. 2 (2023): 362-69.
- Koç, Mevlüt. "Kronik Kalp Yetersizliğinde Fonksiyonel Kapasite ve Prognozu Değerlendirmede Serum Nt-Probnp Düzeyi." (2006).
- ÖZBAY, Feyza ALTUNBEY, and Erdal ÖZBAY. "Makine Öğrenmesi Algoritmalarının Kalp Yetmezliği Hastalarının Hayatta Kalma Tahmini Üzerindeki Performans Karşılaştırılması."
- ÖZOĞUR, Hatice NİZAM, and Zeynep %J Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi ORMAN. "Sağlık Verilerinin Analizinde Veri Ön İşleme Adımlarının Makine Öğrenmesi Yöntemlerinin Performansına Etkisi." 16, no. 1: 23-33.
- Öztürk Başak, and Halife Kodaz. "Yeni Pap Smear Veri Seti Kullanılarak Derin Öğrenme ile Rahim Ağzı Kanseri Teşhisi." Konya Teknik Üniversitesi, 2022.
- Öztürk, Ergin Şahin. "Yapay Sinir Ağları ve Yapay Zekâya Genel Bir Bakış." No. 2 (2018): 25-36.
- Polatgil, MESUT %J Düzce Üniversitesi Bilim ve Teknoloji Dergisi. "Veri Ölçekleme ve Eksik Veri Tamamlama Yöntemlerinin Makine Öğrenmesi Yöntemlerinin Başarısına Etkisinin İncelenmesi." 11, no. 1 (2023): 78-88.
- Potur, Ezgi Aktaş, and Nihal %J Avrupa Bilim ve Teknoloji Dergisi Erginel. "Kalp Yetmezliği Hastalarının Sağ Kalımlarının Sınıflandırma Algoritmaları ile Tahmin Edilmesi." no. 24 (2021): 112-18.
- SAYIN, Hamdi, Osamah Khaled Musleh SALMAN, Bekir AKSOY, and KÖSE %J Düzce Üniversitesi Bilim ve Teknoloji Dergisi Utku. "Ekg Sinyallerini Kullanarak Kalp Ritimlerinin Yapay Zekâ ile Sınıflandırılması." 9, no. 1 (2020): 7-15.
- Sevli, Onur. "Göğüs Kanseri Teşhisinde Farklı Makine Öğrenmesi Tekniklerinin Performans Karşılaştırması." European Journal of Science and Technology (2019): 176-85. <https://doi.org/10.31590/ejosat.553549>.
- Şen, Z (2004), Yapay Sinir Ağları İlkeleri, *Yapay Sinir Ağları İlkeleri*.

Taner, Neda. "Kalp Yetmezliđi Olan Hastalarda Farmasötik Bakımın Etkisi Üzerine Prospektif Bir Çalışma." (2020).

Taşcı, Erdal, and Aytuđ %J Akademik Biliřim Onan. "K-En Yakın Komřu Algoritması Parametrelerinin Sınıflandırma Performansı Üzerine Etkisinin İncelenmesi." 1, no. 1 (2016): 4-18.

Toprak, Ahmet, and Abdül Halim %J Veri Bilimi Zaim. "Dađıtık Mesajlaşma Altyapısı Kullanılarak Büyük Boyutlu Verilerin Gerçek Zamanlı Olarak İşlenmesi." 4, no. 1 (2021): 45-52.

Türkyılmaz Ayře Beril, and Mehmet Ertürk, Arda Güler. "Transtorastik veya Transözofageal Ekokardiyografi Eřliđinde Perkütan Atriyal Septal Defekt Kapatmanın Erken Dönem Sonuçlarının Karşılaştırılması." Sağlık Bilimleri Üniversitesi, 2023.

İnternet Taraması (online). Denetimli Makine Öğrenme Model řeması. <https://www.turhost.com/blog/makine-ogrenmesi-machine-learning-nedir/> [Ziyaret Tarihi: 21 řubat 2024].

EKLER

Ek – Etik Kurul Raporu





T.C.
NEVŞEHİR HACI BEKTAŞ VELİ ÜNİVERSİTESİ
Bilimsel Araştırmalar ve Yayın Etik Kurulu

Sayı : E-39769797-050.05.04-2100123821
Konu : Etik Kurul Kararları

15.08.2022

HACIBEKTAŞ TEKNİK BİLİMLER MESLEK YÜKSEKOKULU MÜDÜRLÜĞÜNE

09.08.2022 tarihinde yapılan Bilimsel Araştırmalar ve Yayın Etik Kuruluna biriminiz tarafından yapılan başvurulara ait kararlar ekte gönderilmiştir.
Bilgilerini ve gereğini rica ederim.

Prof. Dr. Şahlan ÖZTÜRK
Kurul Başkanı

Ek: HALİFE KODAZ-ORÇUN BAĞRA

Bu belge, güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Kodu : 0EVA-6L09-03LD Belge Doğrulama Adresi : <https://ebysorgu.nevsehir.edu.tr>

Adres:
Telefon No :
e-Posta :
Kep Adresi : nevsehiruniversitesi@hs01.kep.tr

Fax No :
İnternet Adresi :
<http://www.nevsehir.edu.tr>

Bilgi İçin :Leyla KARAGEDİK
Memur
Dahili No:10064





T.C.
NEVŞEHİR HACI BEKTAŞ VELİ ÜNİVERSİTESİ
Bilimsel Araştırmalar ve Yayın Etik Kurulu

KURUL KARARI

TOPLANTI SAYISI
08

KARAR SAYISI
275

TOPLANTI TARİHİ
09.08.2022

Üniversitemiz Hacıbektaş Teknik Bilimleri Meslek Yüksekokulunda görev yapmakta olan öğretim elemanı Arş. Gör. aynı zamanda Konya Teknik Üniversitesinde Yüksek Lisans Programı öğrencisi Orçun BAĞRA'nın "Kalp Yetmezliği Teşhisinin Makine Öğrenme Algoritmalarıyla Tespiti" isimli yüksek lisans tezi hakkında alınan 03.08.2022 tarih ve 2100121593 sayılı başvuru dosyasının görüşülmesi.

2022.08.275. Üniversitemiz Hacıbektaş Teknik Bilimleri Meslek Yüksekokulunda görev yapmakta olan öğretim elemanı Arş. Gör. aynı zamanda Konya Teknik Üniversitesinde Yüksek Lisans Programı öğrencisi Orçun BAĞRA'nın "Kalp Yetmezliği Teşhisinin Makine Öğrenme Algoritmalarıyla Tespiti" isimli yüksek lisans tezi hakkında alınan 03.08.2022 tarih ve 2100121593 sayılı başvuru dosyası görüşüldü.

Yapılan görüşmeler sonucunda, aşağıdaki tabloda isimleri belirtilen araştırmacılar tarafından hazırlanan "*Kalp Yetmezliği Teşhisinin Makine Öğrenme Algoritmalarıyla Tespiti*" isimli yüksek lisans tezi ve ilgili belgeler araştırmanın gerekçe, amaç, yaklaşım ve yöntemleri dikkate alınarak incelenmiş olup, projenin gerçekleştirilmesinde etik sakınca bulunmadığına kurumumuz üyeleri tarafından oy birliği ile karar verilmiştir.

YÜRÜTÜCÜ	ARAŞTIRMACI/UZMAN
Prof. Dr. Halife KODAZ	Orçun BAĞRA (Öğrenci)

Prof. Dr. Şahlan ÖZTÜRK
Kurul Başkanı

Prof. Dr. Zübeyde KUMBIÇAK
Üye

Prof. Dr. Suzan ÇOBAN
Üye

Prof. Dr. Lütfi BUYRUK
Üye

Prof. Dr. Ahmet OĞUZ
Üye

*Prof. Dr. Merter Rahmi
TELKENAROĞLU
(İznil)

*Prof. Dr. Muhammet Şevki
AYDIN
(İznil)

*Prof. Dr. Fatih
ÖZDEMİR
(İznil)

*Prof. Dr. Ensar ÇETİN
(İznil)

Bu belge, güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Kodu : 0EBU-T2BP-0OLM Belge Doğrulama Adresi : <https://ebysorgu.nevsehir.edu.tr>

Adres:

Telefon No :

e-Posta :

Kap Adresi : nevsehiruniversitesi@hs01.kep.tr

Fax No :

İnternet Adresi :

<http://www.nevsehir.edu.tr>

Bügi İcin :Leyla KARAGEDİK

Memur

Dahili No:10064





T.C.
NEVŞEHİR VALİLİĞİ
İl Sağlık Müdürlüğü

NEVŞEHİR İL SAĞLIK MÜDÜRLÜĞÜ - NEVŞEHİR
HASTANE HİZMETLERİ BİRİMİ
30/09/2022 09:47 - E-26171210 - 929 - 1215



Sayı : E-26171210-929
Konu : Bilimsel Çalışma Hk.

NEVŞEHİR DEVLET HASTANESİNE

Müdürlüğümüze bağlı Nevşehir Devlet Hastanesinde, Dr.Öğretim Üyesi Dr. Gülden KÜÇÜKAKÇA ÇELİK'in "Kalça Atroplastisi Cerrahisi Uygulanan Hastaların Erken Dönem Konfor Düzeyi Üzerinde Beklentilerin ve Anksiyetenin Etkisi" konulu araştırma talebi, Araştırma Görevlisi Dr. Pelin CALPBİNİCİ'nin "Koçluğa Dayalı Liderlik Ölçeği:Türkçe Geçerlilik ve Güvenirlilik Çalışması" konulu araştırma talebi, Nevşehir Devlet Hastanesi personeli Uzm. Dr.Fırat DURNA'nın "Koronar Arter Bypass Grefti Ameliyatı Sonrası Atriyal Fibrilasyonun Öngörülmesinde Biyobelirteç Olarak Sistemik İmmün-İnflamasyon İndeksi ve Sistemik Enflamasyon Yanıt İndeksinin Rolü" konulu araştırma talebi, Nevşehir Devlet Hastanesi personeli Hemşire Ercan KILIÇ'ın "Ağrı Şikayeti ile Acil Kliniğe Başvuran Bireylerin Ağrı Şiddeti ve Ağrı Korkusunun Ağrının Felaketleştirilmesi Arasındaki İlişkinin Belirlenmesi" konulu araştırma talebi, Nevşehir Devlet Hastanesi personeli Tıbbi Sekreter Ümit ÇOBAN'ın "Sağlık Çalışanlarının Sağlık Teknolojisi Algılarının Ölçülmesi" konulu araştırma talebi, Nevşehir Devlet Hastanesi personeli Uzm.Dr.Esra DEMİR ÜNAL'ın "Alzheimer Hastalık Progresyonunda Periferik Serum Biyomarkerlerinin Kognitif Değerlendirme Ölçekler ile İlişkisi ve Terapötik Ajanlarla İlgisi" konulu araştırma talebi, Öğretim Görevlisi Orçun BAĞRA'nın "Kalp Yetmezliği Teşhisinin Makine Öğrenme Algoritmalarıyla Tespiti" konulu Bilimsel Araştırma çalışmalarının yapılması uygun görülmüştür.

Gereğini rica ederim.

Dr. Tanık AKBAŞ
Başkan

Bu belge, güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Kodu: bfa42192-a7e9-4489-8498-0c5758a3d425 Belge Doğrulama Adresi: <https://www.turkiye.gov.tr/saglik-bakanligi-ebys>

15 Temmuz Mah. Sıhhat Sok. Merkez/NEVŞEHİR

Telefon: Faks No:

e-Posta: dilber.uysal@saglik.gov.tr İnternet Adresi: Dilber.Uysal@saglik.gov.tr

Bilgi için: Dilber UYSAL
ODYOMETRİST

Telefon No: (0 384) 213 11 10

