

T.C.  
İSTANBUL AYDIN ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE KALP HASTALIĞININ  
TEŞHİS EDİLMESİ

YÜKSEK LİSANS TEZİ

Mertcan GÖRGÜN

Bilgisayar Mühendisliği Ana Bilim Dalı

Bilgisayar Mühendisliği Programı

Şubat, 2020

T.C.  
İSTANBUL AYDIN ÜNİVERSİTESİ  
LİSANSÜSTÜ EGİTİM ENSTİTÜSÜ



MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE KALP HASTALIĞININ  
TEŞHİS EDİLMESİ

YÜKSEK LİSANS TEZİ

Mertcan GÖRGÜN

(Y1713.010002)

Bilgisayar Mühendisliği Ana Bilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Dr. Öğr. Üyesi Ahmet GÜRHANLI

Şubat, 2020

T.C.  
İSTANBUL AYDIN ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ MÜDÜRLÜĞÜ



YÜKSEK LİSANS TEZ ONAY FORMU

Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Tezli Yüksek Lisans Programı Y1713.010002 numaralı öğrencisi Mertcan GÖRGÜN'ün “**Makine Öğrenmesi İle Kalp Hastalığı Tahmini**” adlı tez çalışması Enstitümüz Yönetim Kurulunun 14.01.2020 tarihli ve 2020/01 sayılı kararıyla oluşturulan jüri tarafından oybirliğiyle Tezli Yüksek Lisans tezi 17.02.2020 tarihinde kabul edilmiştir.

<u>Unvan</u>	<u>Adı Soyadı</u>	<u>Üniversite</u>	<u>İmza</u>
<b>ASIL ÜYELER</b>			
<b>Danışman</b>	Dr. Öğr. Üyesi	Ahmet GÜRHANLI	İstanbul Aydın Üniversitesi
<b>1. Üye</b>	Prof. Dr.	Ali GÜNEŞ	İstanbul Aydın Üniversitesi
<b>2. Üye</b>	Dr. Öğr. Üyesi	Ali HAMİTOĞLU	İstanbul Sabahattin Zaim Üniversitesi
<b>YEDEK ÜYELER</b>			
<b>1. Üye</b>	Dr. Öğr. Üyesi	Adem ÖZYAVAŞ	İstanbul Aydın Üniversitesi
<b>2. Üye</b>	Dr. Öğr. Üyesi	Ahmet Feyzi ATEŞ	İstinye Üniversitesi

ONAY

Prof. Dr. Ragıp Kutay KARACA  
Enstitü Müdürü

## **YEMİN METNİ**

Yüksek Lisans tezi olarak sunduğum “Makine Öğrenmesi Yöntemleri ile Kalp Hastalığının Teşhis Edilmesi” adlı çalışmanın, tarafımdan, bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın yazıldığını ve yararlandığım eserlerin Bibliyografya’da gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.(25/11/2019)

Mertcan Görgün

## **ÖNSÖZ**

Öncelikle tez konusunu seçerken isteklerim doğrultusunda bana yardımcı olan, planlanmasında, araştırılmasında ve oluşumunda desteğini esirgemeyen, tecrübe ve bilgilerinden yararlandığım sayın tez danışmanım Dr. Öğr. Üyesi Ahmet GÜRHANLI'ya teşekkürü bir borç bilirim. Ayrıca beni bu günlere getiren sevgi ve saygıyı kelimenin tam anlamıyla öğreten desteğini hiçbir zaman esirgemeyen aileme ve her zaman sevgisiyle bana kuvvet veren eşime sonsuz teşekkür ediyorum.

Şubat 2020

Mertcan Görgün

## İÇİNDEKİLER

	<u>Sayfa</u>
ÖNSÖZ.....	iv
İÇİNDEKİLER.....	v
KISALTMALAR.....	viii
ÇİZELGE LİSTESİ.....	ix
ŞEKİL LİSTESİ.....	x
ÖZET.....	xi
ABSTRACT.....	xii
1.GİRİŞ.....	1
1.1 Tezin Amacı.....	2
1.2 Literatür Araştırması.....	3
1.3 Konu ile ilgili Kavramlar.....	3
1.3.1 Kalp Hastalıkları.....	3
1.3.2 Nedenleri.....	4
1.3.2.1 Çevresel Etkiler.....	4
1.3.2.2 Genetik Miras.....	4
1.3.3 Risk Faktörleri.....	4
1.3.3.1 Kolesterol.....	4
1.3.3.2 Tansiyon.....	5
1.3.3.3 Tütün Ütünü Kullanımı.....	6
1.3.3.4 Diyabet.....	6
1.3.3.5 Kilo.....	7
1.3.4 Kalp Hastalığı Belirtileri.....	7
1.3.4.1 Göğüs Ağrısı.....	7
1.3.4.2 Nefes Darlığı.....	7
1.3.4.3 Çarpıntı.....	8
1.3.4.4 Senko(Bayılma).....	8
1.3.5 Kalp Hastalığı Türleri.....	8

1.3.5.1 Koroner Hastalıklar.....	8
1.3.5.2 Kalp Kapağı Hastalıkları.....	9
1.3.5.3 Aort Anevrizması.....	9
1.3.5.4 Ritim Bozuklukları.....	9
1.3.5.5 Konjenital Kalp Hastalıkları.....	9
1.4 Hipotez.....	10
<b>2.YÖNTEMLER.....</b>	<b>11</b>
2.1 Yapay Zeka.....	11
2.2 Makine Öğrenmesi ve Derin Öğrenme.....	11
2.2.1 Öğrenme Teknikleri.....	13
2.2.1.1 Gözetimli Öğrenme.....	13
2.2.1.2 Yarı Gözetimli Öğrenme.....	14
2.2.1.3 Gözetimsiz Öğrenme.....	15
2.2.1.4 Takviyeli Öğrenme.....	16
2.2.1.5 Aşırı Öğrenme.....	16
2.2.2 Test Adımları.....	16
2.2.3 Karışıklık Matrisi.....	16
2.2.3.1 Tahmin Hatası.....	17
2.2.3.2 Doğruluk Oranı.....	18
2.2.3.3 Geri Çağırma (Recall).....	18
2.2.3.4 Hassasiyet (Precision).....	18
2.2.3.5 F Skoru.....	18
2.3 Makine Öğrenmesinin Tıp Alanında Kullanımı ve Önemi.....	19
2.4 Kullanılan Makine Öğrenmesi Yöntemleri.....	19
2.4.1 K-En Yakın Komşu.....	19
2.4.2 Lojistik Regresyon.....	20
2.4.3 Naive Bayes.....	21
2.4.4 Destek Vektör Makineleri.....	22
2.4.4.1 Doğrusal Olmayan DVM.....	23

2.4.4.2 Doğrusal DVM.....	24
2.4.5 Karar Ağacı.....	25
2.4.6 Rastgele Orman.....	26
2.4.7 LightGBM.....	28
2.4.8 XGBoost.....	29
2.4.9 Bagging.....	29
2.4.10 Ridge.....	30
<b>3.PROJE MODELİ.....</b>	<b>31</b>
2.5 Veri Seti.....	32
2.6 Öznitelik Kullanımı ve Grafikleri.....	34
2.6.1 Yaş.....	34
2.6.2 Cinsiyet.....	34
2.6.3 Cinsiyete Göre Hasta Olup Olmama Durumu.....	35
2.6.4 Maksimum Kalp Atış Hızı, Yaş ve Hasta Olup Olmama Durumu.....	35
2.6.5 Pik Egzersizde ST Segmenti Eğimi.....	36
2.6.6 Kandaki Şeker Oranı > 120 mg/dl Olan Kişilerin Hasta Durumu.....	37
2.6.7 Göğüs Ağrısı Tipine Göre Kalp Hastalığı Sıklığı.....	37
<b>4.DEĞERLENDİRME.....</b>	<b>38</b>
<b>5.SONUÇLAR.....</b>	<b>42</b>
<b>KAYNAKLAR.....</b>	<b>44</b>
<b>ÖZGEÇMİŞ.....</b>	<b>47</b>



## KISALTMALAR

<b>KH:</b>	Kalp Hastalığı
<b>RO:</b>	Rastgele Orman
<b>DP:</b>	Doğru Pozitif
<b>DN:</b>	Doğru Negatif
<b>YP:</b>	Yanlış Pozitif
<b>YN:</b>	Yanlış Negatif
<b>DVM:</b>	Destek Vektör Makineleri
<b>NB:</b>	Naive Bayes
<b>LR:</b>	Lojistik Regresyon
<b>BOH:</b>	Bulaşıcı Olmayan Hastalıklar
<b>DVM:</b>	Destek Vektör Makineleri
<b>KDH:</b>	Kalp Damar Hastalıkları
<b>LDL:</b>	Düşük Yoğunluklu Lipoprotein
<b>DVM:</b>	Destek Vektör Makineleri
<b>LightGBM:</b>	Light Gradient Boosting Machine
<b>XGBOOST:</b>	Extreme Gradient Boosting
<b>RO:</b>	Rastgele Orman

## ÇİZELGE LİSTESİ

	<u>Sayfa</u>
<b>Çizelge 2.1:</b> Karışıklık Matrisi .....	17
<b>Çizelge 3.1:</b> Öznitelikler.....	32
<b>Çizelge 4.1:</b> Rastgele Orman Parametreleri.....	40

## ŞEKİL LİSTESİ

	<u>Sayfa</u>
Şekil 1.1: Kalp Hastalığı.....	1
Şekil 1.2: Kalp Hastası Olup Olmama Durumu .....	2
Şekil 1.3: Damardaki Kolesterol Oluşumu.....	5
Şekil 1.4: Koroner Kalp Hastalıklarının Oluşumu.....	8
Şekil 2.1: Sınıflandırma.....	14
Şekil 2.2: Regresyon Doğrusu.....	14
Şekil 2.3: Kümeleme.....	15
Şekil 2.4: Lojistik Regresyon Eğrisi.....	21
Şekil 2.5: Destek Vektör Makineleri Sınır Düzlemi.....	23
Şekil 2.6: Destek Vektör Makineleri Çekirdek Fonksiyonları.....	24
Şekil 2.7: Doğrusal Ayrımda Oluşan Hiper Düzlem ve Destek Vektörleri.....	25
Şekil 2.8: Rastgele Orman Algoritması ve Karar Ağaçları İlişkisi.....	27
Şekil 2.9: LightGBM Mimarisi .....	28
Şekil 2.10: XGBoosting Mimarisi.....	29
Şekil 3.1: EKG Dalgaları ve Segmentleri.....	33
Şekil 3.2: Veri Setindeki Hastaların Cinsiyet Grafiği .....	34
Şekil 3.3: Veri Setinde Kullanılan Hastaların Yaş Grafiği.....	34
Şekil 3.4: Cinsiyete Göre Hasta Olup Olmama Durumu.....	35
Şekil 3.5: Maksimum Kalp Atış Hızı, Yaş ve Hasta Olup Olmama Durumu.....	35
Şekil 3.6: Pik Egzersizde ST Segmentine Göre Kalp Hastalığı Durumu .....	36
Şekil 3.7: Veri Setinde Kullanılan Hastaların Yaş Grafiği.....	36
Şekil 3.8: Göğüs Ağrısı Tipine göre Kalp Hastalığı Sıklığı.....	37
Şekil 4.1: Algoritma Oranları Sonuç Grafiği.....	38
Şekil 4.2: Lojistik Regresyonda İleri Yayılma Yöntemi.....	38
Şekil 5.1: Karışıklık Matrisi.....	43

# MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE KALP HASTALIĞININ TEŞHİS EDİLMESİ

## ÖZET

Kalp hastalıkları, kalbe giden damarlarda tıkanıklık, kalp kapakçığında açıklık ya da kalbin beklenmedik herhangi bir yerinde oluşan bir sorun ile ortaya çıkar. Kalp Hastalığı riskini arttıran iki önemli faktör vardır. Bunlar Genetik Faktörler ve Çevresel Faktörlerdir. Kalp hastalıklarında en yaygın görünen damarlardaki plaklanmadan dolayı olan kalp krizidir. Kalp hastalıklarının en önemli belirtisi göğüs ağrısıdır ve araştırmalar bu ağrılar ile karşılaşan insanların yaş ortalamasının günden güne düştüğünü göstermektedir. Dünya Sağlık Örgütü'nün verilerine göre dolaşım sistemi hastalıkları adı altında Dünya'da her yıl 17,9 milyon insan kalp hastalıklarından, Türkiye'de ise Türkiye İstatistik Kurumu verilerine göre ortalama 168 bin insan hayatını kaybetmektedir. Kalp Hastalığı Göğüs Ağrısı, nefes darlığı, bayılma, yorulma ve halsizlik gibi etkileri olan ve günlük yaşantınızı olumsuz yönde etkileyecek bir hastalıktır. Erken teşhis konulduğunda hastalığın ileri safhalara geçmemesi ve tedavinin başlaması hastanın hayatını kurtarır. Bu çalışmada 165'i Kalp hastası olan 303 denekten oluşan Heart Disease UCI veri seti üzerinde cinsiyet, diyabet, yaş, kolesterol, göğüs ağrısı türleri gibi özelliklerle çeşitli makine öğrenmesi yöntemleri uygulanmıştır. Lojistik Regresyon, K-En Yakın Komşu, Destek Vektör Makineleri, Naive Bayes, Karar Ağacı, Rastgele Orman, LightGBM Model, XGBoost Model, Ridge Model ve Bagging Model algoritmaları karşılaştırılmış, çıkan sonuçlar değerlendirilmiş ve farklı parametreler kullanılarak Rastgele Orman Algoritması ile %90,16 oranında doğruluk değeri elde edilmiştir.

**Anahtar Kelimeler:** *Kalp Hastalığı Tahmini, Makine Öğrenmesi, Rastgele Orman Algoritması*

## **DIAGNOSING HEART DESEASE BY MEANS OF MACHINE LEARNING METHODS**

### **ABSTRACT**

Heart diseases emerge as a result of a complication presenting itself in an unexpected part of the heart, as an infarction in the veins leading to the heart or an orifice in the cardiac valve. There are two important factors increasing the risk of heart disease. These are the genetic factors and environmental factors. The most common among the heart diseases is cardiac arrest, which occurs due to the plaque build-up in the veins. The most important symptom of heart disease is chest pain, and the researchers suggest that the average age of people experiencing these pains go down day by day. According to World Health Organization data, 17,9 million people die due to heart disease every year under the name of circulatory system diseases, and an average of 168 thousand people pass away in Turkey according to the Turkish Statistical Institute data. Heart diseases cause chest pain, shortness of breath, fainting, fatigue and exhaustion symptoms and affect peoples' life negatively. When diagnosed at an early stage, the patient's life can be saved by starting treatment and preventing the disease from advancing. In this research, various machine learning methods were applied on the Heart Disease UCI dataset comprised of 303 subjects including 165 cardiac patients, with factors such as sex, diabetes, age, cholesterol, and types of chest pain. Logistic Regression, K-Nearest Neighbor, Support Vector Machines, Naive Bayes, Decision Tree, Random Forest, LightGBM Model, XGBoost Model, Ridge Model and Bagging Model algorithms were compared, the results were evaluated, and 90,16% truth value was obtained with Random Forest Algorithm using different parameters.

**Keywords:** *Heart Disease Estimation, Machine Learning, Random Forest Algorithm*

## 1. GİRİŞ

İnsanların varoluşundan beri süre gelen zamanda yiyecek, barınma gibi ihtiyaçları olmuştur. Bu ihtiyaçlar karşılanamadığında, yaşam süresi boyunca genetik ya da çevresel etkenlerden dolayı çeşitli hastalıklara yakalanmışlardır. Bu hastalıkların bir türü de kalp rahatsızlıklarıdır. Kalp hastalıkları kalbin ana damarları ya da taşıma damarında oluşan tıkanıklık, kalp kapakçıklarında ya da kalp odacıklarında oluşan bir problemden dolayı ortaya çıkan erken teşhiste hayat kurtarılabilen ve tedavisi mümkün olan hastalıklardır. Kalp hastalıkları Göğüs Ağrısı, bayılma, nefes alamama, çabuk yorulma gibi etkileri olan ve hayat standartlarınızı büyük ölçüde etkileyecek hastalıklardır. Hastanın Ağrı çeşidine, cinsiyetine, yaşına, kandaki yağ oranına, kandaki şeker oranına vb. değerlere göre teşhis konulabilir. Ancak hastalık ilerlemişse hayat boyu kullanılacak ilaçlar tedavi sürecine dahil edilir, bu durum hastanın günlük yaşantısını da etkiler. Dolayısıyla teşhis hastalığın ileri safhalarına gelmeden konulmalıdır.

BOH'lar içerisinde, kalp ve damar hastalıkları tüm ölüm nedenleri arasında ilk sırada yer almakta, özellikle iskemik kalp hastalıkları ve serebrovasküler hastalıklar ilk iki ölüm nedenini oluşturmaktadır. Kalp ve damar hastalıklarının küresel ölçekte uzun bir süre daha bir numaralı ölüm sebebi olmaya devam edeceği tahmin edilmektedir(Tc Sağlık Bakanlığı, 2015).



**Şekil 1.1: Kalp Hastalığı**

**Kaynak:** (Google.com, 2019)

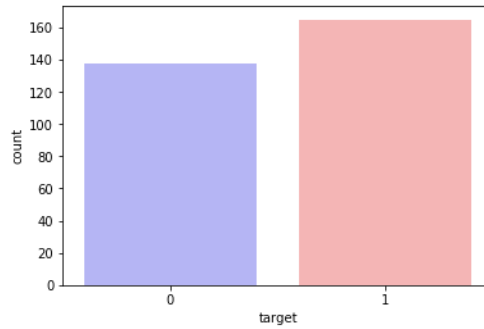
İçerisinde kalp hastası olan ve olmayan belirli bir gruptan alınan Cinsiyet, yaş, hasta olup olmama durumu, kolesterol, diyabet vb. değerlere sahip olan verilere makine öğrenmesi teknikleri uygulandığında erken teşhis yapılabilmektedir.

Kalp hastalığının tespit edilmesinde makine öğrenmesi ile ilgili kavramlar anlatılmaktadır ve kullanılmaktadır.

Tezin birinci bölümünde Kalp hastalıklarının çeşitleri, nedenleri, belirtileri, erken teşhis edilmesinin önemi gibi bilgiler verilmiştir. İkinci bölümünde tezde kullanılan yöntemler, makine öğrenmesi tarihi, algoritmaları, tıp alanında kullanımı ve öneminden bahsedilmiştir. Üçüncü bölümünde veri seti, teşhisi sağlayan en etkili algoritma ve modelin detayları ele alınmıştır. Dördüncü bölümünde tezin değerlendirmesi yapılmıştır. Beşinci ve son bölümde ise sonuçlar ve etkisi konu alınmıştır.

### 1.1 Tezin Amacı

Bu çalışmada içerisinde 303 veri bulunan Heart Disease UCI veri setinden yararlanılmıştır. Veri seti üzerinde makine öğrenmesi yöntemleri kullanılarak bir kişinin kalp hastası olup olmama durumu tespit edilmeye çalışılmıştır. Veri setinde bulunan 303 veriden 165'i kalp hastasıdır. Kişilerin cinsiyeti, yaşı, kalp hastası olup olmama durumu, göğüs ağrısının türü, kolesterol oranı, diyabet, Kalp atış hızı, elektrokardiyografik ölçüm değerleri gibi kişilere özgü farklı değerler içermektedir. Veri setinde Plot grafiklerinde de belirtildiği gibi cinsiyet, kan şekerinin (<120) olması, kişide kalp hastalığının olup olmama durumu boolean değerler (0 ve 1) ile gösterilmektedir. Çalışmada kalp hastalığı olan kişiler 1 olmayan kişiler 0 ile gösterilmektedir.



Şekil 1.2: Kalp Hastası Olup Olmama Durumu

Kalp hastalarının tespit edilebilmesi için toplam 303 kişinin 13 öznitelik bilgisi kullanılmış, 10 makine öğrenmesi algoritması özniteliklere uygulanmış, algoritmalarından gelen sonuçlar karşılaştırılmış ve en iyi sonucu veren algoritma ile kalp hastalığı teşhisi konulması amaçlanmıştır.

## **1.2 Literatür Araştırması**

Tıp alanında birçok çalışma yapılmıştır. Ancak makine öğrenmesi algoritmalarının kalp hastalıklarının teşhisinde kullanıldığı az sayıda çalışmaya rastlanılmıştır. Bunlardan biri 2016 yılında Faruk BULUT tarafından Adaboost kullanılarak yapılan çalışmada yüksek tansiyonun kalp krizi tespitinde %87,89 oranında etkisi olduğu saptanmıştır. Murat KARAKOYUN ve Mehmet HACIBEYOĞLU tarafından yürütülen biyomedikal veri setleri ile makine öğrenmesi algoritmaları karşılaştırılması yapılarak DVM sınıflandırması kullanılarak kalp hastalıkları teşhis etmede %59,72 oranında başarı elde edilmiştir. Elif KARTAL ve Mehmet Erdal BALABAN tarafından 2018 yılında yayınlanan çalışmada makine öğrenmesi yöntemleriyle kardiyak risk analizi yapılmış ve en yüksek oran karar ağacı ile %98,2 olarak tespit edilmiştir. Ömer Faruk BOYRAZ, Volkan SEYMEN, Mehmet Recep BOZKURT ve Özdemir ÇETİN tarafından yapay sinir ağı yöntemi kullanılarak yapılan araştırmada %90 oranında sonuç elde edilmiştir.

## **1.3 Konu ile İlgili Kavramlar**

Makine öğrenmesi uygulanan kalp hastalığı ile ilgili temel kavramlardan bahsedilmektedir. Kalp hastalığının tanımı, belirtileri, nedenleri, ağrı türleri konu alınmıştır.

### **1.3.1 Kalp Hastalıkları**

Ömür boyunca sürekli çalışan ve tüm vücuda kan dağıtan insan anatomisinin önemli organlarının başında gelen kalp bazen çeşitli engellerden dolayı görevini aksatabilmektedir. Kalp ortalama olarak dakikada 70, günde 105.000 ve yılda 37 milyon kez kasılarak, içindeki kanı vücuda pompalıyor. Kalp krizi, ritim bozukluğu, damar tıkanıklıkları, doğumsal problemler, kapakçık hastalıkları gibi rahatsızlıklar kalbin düzgün çalışmasına, kanı vücuda pompalamasına engel olur ve insan yaşamını büyük ölçüde etkileyip yaşam kalitesinin düşürür. Bu hastalıklar bazen etkili bir biçimde hastaya kendini hissettiriyor ve bir uzmana danışıldığı zaman teşhis



konulabiliyor. Bazen de hastalık hastanın hissetmeyeceği, hastayı rahatsız etmeyen gizli bir şekilde ilerliyor ve hasta belirli bir rahatsızlık hissetmediği için bir uzmana gitmiyor. Sonuç olarak hastalığın teşhisinde geç kalınmış olunabiliyor.

Dünya Sağlık Örgütü verilerine göre 2020 yılında dünya üzerindeki tüm ölümlerin %36'sı kalp damar hastalıklarına bağlı olarak gerçekleşecektir. Türkiye'de ise önümüzdeki 10 yılda koroner kalp hastası sayısının iki kat artış göstererek 5,6 milyona ulaşması bekleniyor. Hal böyle olunca gerek bilim adamları gerekse endüstri kalp damar hastalıklarına (KDH) büyük ilgi duymakta ve her gün bu hastalıkların tedavisine yönelik yeni gelişmeler kaydedilmektedir (Güleç, 2009)

### **1.3.2 Nedenler**

Kalp hastalığının nedenlerini çevresel ve genetik nedenler olarak ayırabiliriz. Çevresel nedenler insanların yaşadığı çevrenin etkileri, genetik nedenler ise insanların aileden gelen rahatsızlıklarının nesilden nesile aktarılmasıyla hastalık formlarının kişilerde oluşmasıdır.

#### **1.3.2.1 Çevresel Nedenler**

Çevresel nedenler, insanların yaşadıkları çevreden kaynaklı sağlıklarının etkilenip sonraki evrelerde sağlık problemlerine yol açan nedenlerdir. Yaş, cinsiyet, tütün ürünü kullanımı, diyabet, kilo bu nedenler arasında sayılabilir ve incelenebilir.

#### **1.3.2.2 Genetik Nedenler**

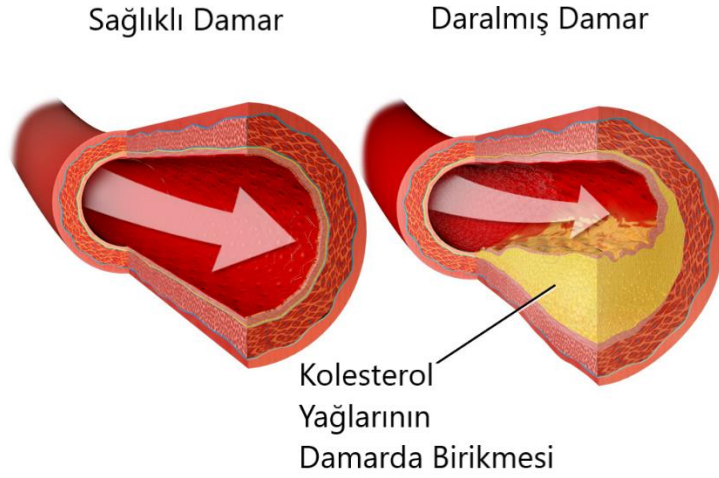
Genetik nedenler, katılımsal olarak aile bireylerinden gen haritasının oluşmasından dolayı yeni doğan bir bireye aktarılan genlerin içerisinde barındırdığı hastalıklardır. Araştırmalar kalp hastalıklarına etki eden kolesterol ve tansiyon özniteliklerinin çevresel nedenlerden daha çok genetik nedenlere bağlı olduğu üzerinde durulmuştur.

### **1.3.3 Risk Faktörleri**

#### **1.3.3.1 Kolesterol**

Kolesterol oranı, canlı vücudunda dolaşımı sağlayan ve tüm hücrelerde bulunan kanın, içerisindeki yağlanma oranına denir. Kalp rahatsızlıklarını oluşturan risk faktörlerinden en önemlisi kabul edilir, fakat ilaçlar ile tedavisi mümkündür. Vücutta

fazla olan kolesterolü karaciğere taşıyan iyi huylu kolesterol ve fazla kolesterolü vücutta taşıyan ve damar duvarlarında yağ birikimine yol açan kötü kolesterol vardır.



**Şekil 1.3:** Damardaki Kolesterol Oluşumu

Kaynak: (Türk Kardiyoloji Vakfı, 2019)

Kişinin kolesterol düzeyi ne kadar yüksekse, kalp hastası olma ihtimali de o kadar yüksektir. Türkiye’de birinci sırada gelen ölüm nedeni kalp-damar hastalığıdır (William, 1959).

Kolesterol yüksekliği ile Kardiyovasküler hastalık arasındaki ilişkinin en dramatik kanıtı ailevi hiperkolesterolemide kendini göstermektedir. Homozigot ailevi hiperkolesterolemide çok yüksek olan LDL-kolesterol düzeyleri daha 2. dekatta yaygın ateroskleroz ve ölümcül klinik sonuçlara yol açmaktadır (Kayıkçoğlu, 2014).

### 1.3.3.2 Tansiyon

Vücuda kan dolaşımı sırasında atardamarlardaki kan basıncına tansiyon denmektedir. Kan basıncının artması durumuna yüksek tansiyon, azalması durumuna ise düşük tansiyon denir. Yaşın ilerlemesi, aşırı şeker ve tuz tüketimi, Alkol tüketimi, stres gibi faktörler tansiyon riskini arttırmaktadır.

Yapılan alıřmalar tansiyon hastalarında kolesterol dūřürücü tedavinin kan basıncını kontrol etmek kadar yararlı olabileceğini göstermektedir (Güle, 2009).

### **1.3.3.3 Tütün Ürünü Kullanımı**

Tütün ürünleri kullanımı her toplumu ilgilendiren ve hayati bir sorundur. Tütün sadece kalbe deęil etki ettięi her organa zarar veren bir maddedir.

Bugün yařayan 500 milyon insanın tütün kullanımının sonuçlarından ölmesi beklenmektedir. 21. yüzyılda 1 milyar insanın tütünden öleceęi tahmin edilmektedir. Ölümlerin yarısı 35-69 yař insanlarda olmaktadır. 1995 yılında dünyada 1.1 milyon kişinin sigaraya baęlı kanser nedeniyle öldüęü, bu ölümlerin 765.000'den fazlasının geliřmiř ülkelerde gerekleřtięi, bu sayının da 500.000'ini akcięer kanserinin oluřturduęu bildirilmektedir (Karlıkaya, 2006).

Yapılan arařtırmaya göre, tütün kullanımı tüm kronik akcięer hastalıklarının %80'inden, kalp hastalıęı ve kansere baęlı ölümlerin de üçte birinden sorumlu bulunmuřtur (US Department of Health and Human Services, 1982).

### **1.3.3.4 Diyabet**

Kanda bulunan řeker oranına kan řekeri denmektedir. Diyabetin en ok etki ettięi bölge damarlardır. ünkü damarlarda dolařan kan řekerle birlikte yoęunlařarak plaklařmaya neden olmaktadır. Bu durumda da damar tıkanıklıkları gözlemlenmektedir.

Bir arařtırmada, Türkiye'nin 59 yerleřim biriminde 20 yař ve üzerindeki 3687 kiřide yapılan taramada, diyabet prevalansı incelendi. Diyabet tanısı kapsamına kendilerinde bu tanının konduęunu öne sürenler ile taramada alık kan řekeri 130 mg/dl veya 2 saat postprandiyal deęeri 170 mg/dl'in üzerinde bulunan kiřiler girdi. 83 eriřkine (% 2.25) diyabet varlıęını bildirdi; geri kalanlarda % 1.6 oranında yeni diyabet keřfedildi. 35-64 yařlarını ilgilendiren yař standardizasyonu ile diyabet prevalansı kadınlarda % 6.3, erkeklerde % 4.6 bulundu. Anılan oranlar uluslararası kıyaslamada erkeklerde orta düzeyi, kadınlarda oldukça yüksek düzeyi temsil etti. Kırsal kesime göre řehirlerde tüm eriřkinlerde anlamlı bir prevalans farkı yoktu, ancak erkeklerde diyabete řehirde

daha sık ( $p<0.03$ ) rastlandı (Onat ve diğeri, 1991). Bu durumda diyabetin varlığının kalp hastalıklarına azımsanmayacak derecede etkisi vardır.

### **1.3.3.5 Kilo**

VKI insan sağlığı için çok önemlidir. Risk faktörleri birbirlerini tetikler ve kilo vücuttaki şeker ve kolesterolü etkileyen en büyük etkidir. Dolayısıyla VKI yüksek olan insanlarda kolesterol ve şekerin fazla olmasının da etki ettiğini göz önüne alındığında kalp hastası olma ihtimali daha yüksektir(Onat ve diğeri, 1991).

Obesite'nin çeşitli kardiyovasküler risk faktörlerini arttırıcı etkisi taramalarda doğrulanmıştır ve Diyabetin de obesiteyle bağıntılı biçimde sıklığı bu tarafında desteklenmiştir.

### **1.3.4 Kalp Hastalığı Belirtileri**

Kalp hastalığının, hastaları hekimlere yönlendiren ve teşhis konmasını kolaylaştıran çeşitli belirtileri vardır. Bunlar; göğüs ağrısı, nefes darlığı, çarpıntı ve bayılmadır.

#### **1.3.4.1 Göğüs Ağrısı**

Göğüs ağrısı kalp ile ilgili olan şikayetlerin en önemli belirtilerindendir. Tanım olarak; soğuk hava, sık egzersiz ve stres gibi faktörler ile birlikte göğüs kafesinin ortasında baskı, sıkışma, yanma gibi etkiler olarak bilinir. Bu çalışmada da özneliklerden birisi göğüs ağrısı şiddetini belirtmektedir ve dört tip göğüs ağrısı tipi vardır.

#### **1.3.4.2 Nefes Darlığı**

Egzersiz yapıldığı sırada ya da günlük olarak yapılan yürüyüşlerde ortaya çıkabilen rahat nefes alamama durumu olarak belirtilir. Ayrıca nefes darlığı egzersizler haricinde stabil dinleme durumunda da görülebilir ve kalp rahatsızlığının belirtisi olabilir.

### 1.3.4.3 Çarpıntı

Belirtiler arasında bulunan çarpıntı, kalp atış hızının aniden çok yükselmesi ya da çok düşmesi ile oluşan düzensiz kalp atışlarına bağlı olarak hissedilen bir duygudur. Çarpıntıya etki eden faktörler kalp hastalıklarına etki eden faktörler ile aynıdır. Kalp çarpıntısı taşikardi olarak ta bilinir.

### 1.3.4.4 Senko (Bayılma)

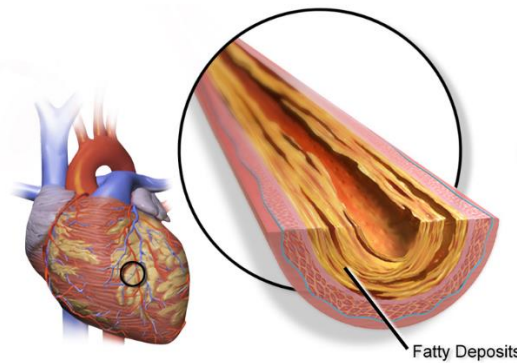
Kişinin genellikle kan basıncındaki ani değişimler sebebiyle bilincinin kapanması durumudur. Bayılma ciddi bir kalp hastalıklarının belirtisi olabileceği gibi birçok ciddi hastalığında habercisi olabilir.

### 1.3.5 Kalp Hastalığı Türleri

Kalp hastalıklarının genetik ve çevresel faktörlerin etki ederek rahatsızlığa yol açan farklı türleri vardır. Bu türler; koroner hastalıklar, kapakçık hastalıkları, anevrizma, ritim bozuklukları ve konjenital hastalıklar olarak belirtilebilir.

#### 1.3.5.1 Koroner Hastalıklar

Koroner Arter Hastalığı, koroner arterlerin duvarlarında oluşan plaklardan ötürü ortaya çıkan bir hastalıktır. Koroner kalp hastalığı veya kısaca CHD (Coronary Heart Disease) olarak da adlandırılır(Bridget, 2010). Bu hastalıklar genellikle yaşlı yetişkinlerde görülse de özellikle ateroskleroz, (damar tıkanıklığı) geçmişi çocukluk çağına dayandığından birincil korunma çabaları yaşamın erken dönemlerinde başlamalıdır (McGill ve diğerleri, 2008).



Şekil 1.4: Koroner Kalp Hastalıklarının Oluşumu

Kaynak: (wikipedia.org, 2013)

### **1.3.5.2 Kalp Kapağı Hastalıkları**

Kalp toplamda 4 odadan ve 4 kapaklarından oluşur. Kalbin her kasılmasında kapaklar açılıp kapanır ve bu yaşam boyu devam eder. Ancak kapaklar yeteri kadar açılıp kapanmadığında bazı sağlık problemleri ile karşılaşılır. Belirtisini göstermeyebilir. Mekanik arızalar olduğu için ilaçla tedavisi mümkün değildir.

### **1.3.5.3 Aort Anevrizması**

Anevrizma bir damarın normal çapının 1,5 kat veya daha fazla genişlemesi olarak tanımlanmaktadır. Anevrizmalar insan vücudunda en sık aort da olmak üzere herhangi bir arterde oluşabilir. Aort anevrizmalarının sıklığı, tanı yöntemlerindeki gelişmeler ve yaşlı popülasyon oranındaki artışa paralel olarak son 20 yılda artış göstermiştir (Nalbant ve diğerleri, 2019). Aort anevrizması belirtisini gösteren bir hastalık olmadığı için tanı hemen konulamayabilir.

Aort anevrizmaları Amerika Birleşik Devletleri'nde ölüm nedenleri arasında on üçüncü sırada yer almaktadır (Sakalihan ve diğerleri, 2005).

### **1.3.5.4 Ritim Bozuklukları**

Kalbin normal atış hızının azalması ya da artması ile belirtilerini gösteren aritmi olarak ta bilinen olağan dışı farklılıklardır. Kalbin çalışması sırasında yaşanan atım bozuklukları ve duraksamalar da ritim bozukluğu olarak adlandırılır. Her yaşta görülebilir.

### **1.3.5.5 Konjenital Kalp Hastalıkları**

Konjenital kalp rahatsızlıkları doğuştan gelen kalp rahatsızlıkları olarak tasvir edilebilir. Genellikle yeni doğanlarda belirtisiz olabileceği gibi kalp üfürmesi ya da huzursuzluk vb. çeşitli şikayetlerle de hekimlere başvurulabilir. Yetişkin bireylerde ise belirtilerini gösterir ve doğuştan gelen bu hastalık ansızın çıkabilir. Hastalığın erken tahmin etme durumunda tedavi çok daha kolay sağlanabileceği öngörülmektedir.

#### **1.4 Hipotez**

Arařtırmalar sonucunda bugüne kadar kalp hastalıkları için kullandığımız Heart Disease UCI veri seti ile ilgili akademik alıřmalarda “Rastgele Orman” yntemi kullanılmıř ancak bařarı oranı bu alıřmadaki bařarı oranından daha ok daha dřk bulunmuřtur. Rastgele orman algoritması eęiteęi veriyi rastgele seeeęi iin karıřık durumdaki veri setleri iin olduka elveriřli olduęu ve parametre olarak aęacın daha fazla dallanmasının, budama iřleminin kısaltılmasının sonuca byk lde etkisi olacaęı ngrlmř ve kullanılmıřtır.

## **2.YÖNTEMLER**

### **2.1 Yapay Zekâ**

Yapay zekâyı kısaca tanımlamak gerekirse makinelerin insanlar gibi düşünüp zekasını kullanarak insanlar gibi becerikli bir şekilde bazı kararları verdikleri mekanizmalardır. 1950 yıllarında ortaya çıkartılmıştır. İnsanlar tarafından oluşturulan yapay zekalar zayıf ya da güçlü olabilir. Zayıf yapay zekalar sadece insanların programladıkları kadar düşünebilir ve karar verebilirler. Güçlü yapay zekalar ise kendi kendine öğrenebilen algoritma kullanılarak hesaplarda bulunurken kendi programını geliştirebilen ve hata yaptığında bu hataları tekrar etmemek için çalışan, hatalardan ders alan bir sistemdir. Yapay zekâ çalışmaları zaman ilerledikçe hızlanmış ve farklı bilgiler de açığa çıkmıştır. Çalışmalar sonucunda makine öğrenmesi ile derin öğrenme bulunmuş yapay zekâ kullanımını artmış ve daha da güçlenmiştir.

### **2.2 Makine Öğrenmesi ve Derin Öğrenme**

Makine Öğrenmesi bilgisayar sisteminde verilen veriler ile işlem yapan algoritmik ve istatistiksel yöntemlerin yardımıyla öğrenme ve öğrendiklerini teste tabi tutarak sonuç çıkartma işlemidir. Makine öğrenmesi yapay zekâ ile birlikte çalışır, yapay zekanın alt kümelerinden birisi olan makine öğrenmesi eğitim verisini algoritma içerisine alarak karar verir ve matematiksel oluşturulan model karara destek olur (Bishop, 2006). Tahmin edilmesi için oluşturulan bu model gelecek zamana yönelik tahmin edici görevini üstlenir. Sonuç olarak veriler eğitim verisi ve test verisi olarak ayrılır. Algoritmaya dahil edilen verilerden yüzdeler olarak çıkarım sağlanır.

Makine öğrenmesinin tarihinde bilim insanı Alan Turing, kendi soyadını verdiği test olan Turing testi ile yapay zekâ çalışmalarının başlangıcı sayılabilecek bir adım atmıştır. İkinci Dünya Savaşında Almanya'nın ordu haberleşme aleti olan Enigma'nın şifrelerinin kırılmasında ve savaşın kısılmasında en büyük etkenin Alan Turing' te olduğu bilinmektedir.



1956 yılında Darmouth Kolejinde düzenlenen bir yaz okulunda Stanford Üniversitesi'nden McCarthy ilk kez yapay zekâ terimini kullanmıştır. Yapay zekâ terimi kullanılmadan önce Turing'in kullandığı terim olan makine zekası terimi kullanılmaktaydı. 1959 yılında Arthur Samuel tarafından oluşturulan dama programında yapay zekanın en fazla kullanılan alanlarından biri olan makine öğrenmesi ismi ilk kez kullanılmıştır. Bu tarihten itibaren makine öğrenmesi üzerine çalışmalar sürdürülmüş ancak devrim niteliğinde bir ilerleme kaydedilmemiştir ta ki teknolojilerin oyunlar üzerine yoğunlaşarak 'Farklı ne yapılabilir?' Sorusuna cevap arayana kadar. 1990'lı yıllardan sonra oyunlarında yardımıyla günümüzde oynanan oyunlar, görüntü işleyiciler, dil işleyiciler, veri madenciliği, robotik kodlama gibi pek çok alanda yapay zekâ ve makine öğrenmesi kullanılmaktadır (Topal, 2017).

Turing ve McCarthy çalışmalarında her zaman insan beyninin düşünebilen bir makine olma olasılığından bahsetmiştir. McCarthy' ye göre belirli bir düzen içerisinde bir bebeğin beyni sağduyulu ve eğitim alabilecek bir makine olarak değerlendirilmektedir. Turing'e göre ise insanların nasıl düşündüğünü araştırmayı hedeflemiştir ve düşünen bir makine yapmak insanların düşünce sistemini çözmek için yardımcı olabilir. Yani ona göre 'Bilgisayar Mekanizması ve Zekâ' adlı makalesinde dediği gibi akıllı davranabilen bir bilgisayar olma olasılığı vardır. Turing'e göre zayıf ve güçlü yapay zekalar arasında net bir farklılık vardır. İnsan düşüncesini taklit eden bir yapay zekâ ise zayıf bir zekaya sahiptir. Ancak iyi programlanmış bir yapay zekâ adeta bir zihindir ve güçlüdür (Topal, 2017).

Derin öğrenme, makine öğrenmesine göre daha az dokunuş gerektiren ama daha çok veriye ihtiyacı olan makine öğrenmesinin alt dalıdır. Derin öğrenme 2010'lu yıllarda kullanımı yaygınlaşmıştır ve büyük veri içerisinde rolünü almıştır. Derin öğrenme ile birlikte makine öğrenmesinde birçok katmanda işlenen veriyi tek bir seferde işleyen, yine makine öğrenmesinin kütüphanelerini kullanan ve içerisinde tanımlanan parametreleri de kendisi bulan, hangi parametrenin en iyi değeri vereceği, en iyi oranı çıkartacağını hesaplayan ve işleyen çalışmalarda ve araştırmalarda daha homojen ve sonuçlar bulan bir yöntem hayatımıza girmiştir.

## 2.2.1 Öğrenme Teknikleri

Yüksek boyutlarda karmaşık bir veri setinin içerisinde anlamlı ve saf veri çıkarma işlemine veri madenciliği demektir. Veri madenciliğinin keşfedilmesi en fazla makine öğrenmesine katkı sağlamıştır. Bu sayede derin öğrenme ile çalışılmaya başlanmıştır. Makine öğrenmesini keşfedildiği tarihten bu yana yöntemler sürekli gelişmeye devam etmektedir. Gelişen bu yöntemler sayesinde birbirinden farklı metodlar bilime katkı sağlamıştır. Katkı sağlayan bu yöntemleri algoritmaları göz önüne alınarak şu şekilde sıralayabiliriz.

- Yarı Gözetimli Öğrenme
- Gözetimli Öğrenme
- Gözetimsiz Öğrenme
- Takviyeli Öğrenme

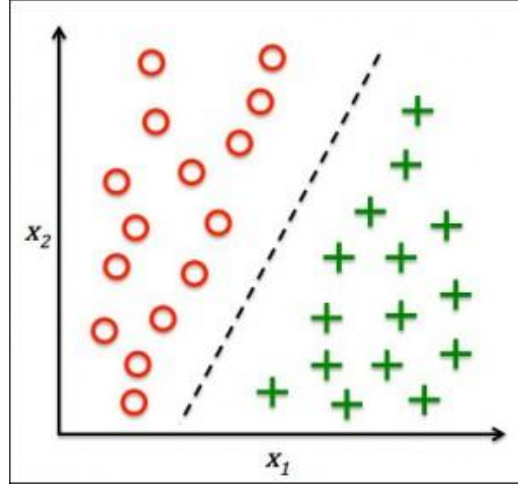
Algoritmaları farklı konulara ve veri setlerinden doğru sonuçlar alabilmek için veri setine uygun yöntemler kullanılması gerekir. Bu yöntemler Sınıflandırma, kümeleme, regresyon, öznitelik ve veri ilişkisi belirleme olarak bilinir ve veri setine etki ederek daha sağlıklı sonuçlar verir.

### 2.2.1.1 Gözetimli Öğrenme

Gözetimli öğrenmede sistemin etiketli veriler kullanan bir veri setinin kullanıcının istediği giriş ve çıkışları barındıran ve algoritmaların bu giriş ve çıkışlara nasıl ulaşabileceğini bulan bir yöntem geliştirmektedir. Verinin ayrıştığı sınıflar bilinir ve bu sınıflara göre hareket edilir. Bu öğrenme şeklinde algoritma tahminde bulunur makine ise bunu düzenler ve bu durum yüksek oran bulana kadar devam eder.

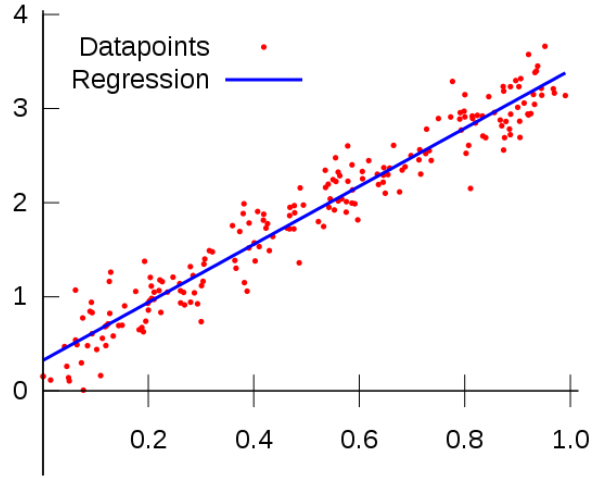
Gözetimli öğrenme tekniklerinde belirleyici olarak 2 görev bulunur: Sınıflandırma ve Regresyon

Sınıflandırma: Veri setindeki verilerin kendi özelliklerine göre ayrılmasıdır. Makine öğrenmesiyle gözlemlenen verilerden çıkan sonuçların kategorize edilmesidir.



Şekil 2.1: Sınıflandırma

Regresyon: Regresyon tekniğinde makine öğrenmesi algoritmaları değişkenler arasındaki ilişkileri anlamlı ve tahmin etmelidir. Bağımlı bir değişkene ve diğer bağımsız bir grup değişkene odaklanır ve eğilimleri tahmin etmek analiz etmek için kullanılır.



Şekil 2.2: Regresyon Doğrusu

### 2.2.1.2 Yarı Gözetimli Öğrenme

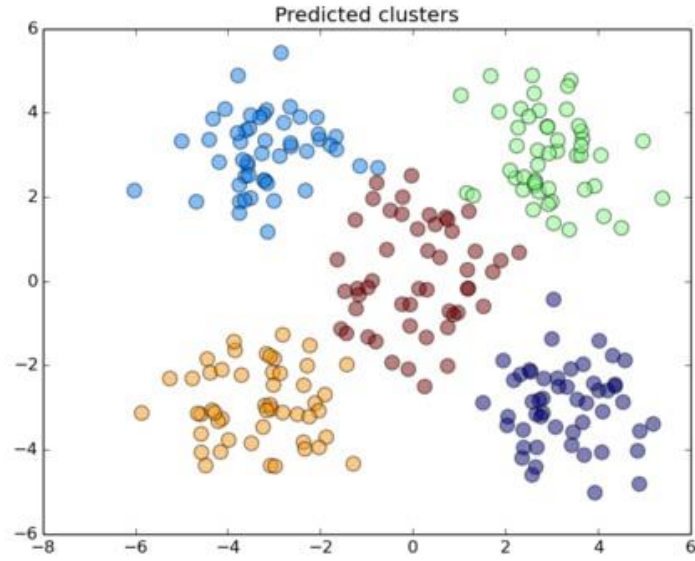
Veri setlerinde genellikle etiketlenmiş veri miktarı etiketlenmemiş veri sayısından daha azdır. Bu etiketlenmemiş veriler diğer tekniklerde fazla kullanılmaz yarı gözetimli öğrenme ise etiketlenmemiş verilere kullanım şansı veren gözetilen bir öğrenme biçimidir. Yapılan araştırmalarda etiketlenmemiş verilerin küçük miktarlarda etiketlenmiş verilerle kullanıldığında daha homojen ve sağlıklı oranlar verdiği

gözenlenmiş ve kanıtlanmıştır. Yarı gözetimli öğrenme gözetimli ve gözetimsiz öğrenmenin tam ortasında kalır.

### 2.2.1.3 Gözetimsiz Öğrenme

Gözetimsiz öğrenmede direktif verebilecek bir insan operatörü yoktur. Sadece makine karar mekanizmasıdır, veri setinden gelen verileri analiz ederek ilişkileri ve korelasyonları belirler. Çıkan sonuçları yorumlama tamamen makine öğrenmesi algoritmasına bırakılmıştır. Ne kadar fazla veri işleme girerse karar mekanizması becerisi o kadar artmaktadır. Yöntem olarak kümeleme yöntemini kullanır.

Kümeleme: Tanımlanmış kriterlere göre benzer verileri farklı gruplara ayırmaktır. Verilerin birkaç küme olacak şekilde ayrılması daha kolay ve şeffaf bir analiz yapılmasını sağlamaktadır.



Şekil 2.3: Kümeleme

#### **2.2.1.4 Takviyeli Öğrenme**

Veri setinden elde edilen bir dizi eylem, sınıf parametreleri ve son değeri içeren düzenli öğrenme süreçleri olan bir öğrenme yöntemidir. Makine öğrenmesi algoritması farklı sonuçları ve seçenekleri değerlendirerek sonuç elde etmeye çalışır. Tavsiyeli makine öğrenmesi hatayı öğretir hatalardan sonuç çıkartarak tekrar edilmemesini sağlar. Eski deneyimlerinden çıkarımlar yaparak en iyi sonucu elde etmeye çalışır.

#### **2.2.1.5 Aşırı Öğrenme (Overlearning)**

Algoritmanın, eğitim verisinde ezberlediği ve bu ezberlerini test verisinde uygulamaya çalışmasından dolayı ortaya eğitim verisi ile test verisinin doğruluk oranlarında büyük farklar çıkartmasına aşırı öğrenme denmektedir. Veri setinin içerisindeki verilerin farklılıkları aşırı öğrenmeyi engellemektedir, bu durumdan dolayı veride çeşitlilik aşırı öğrenmeyi önler.

#### **2.2.2 Test Adımları**

Makine öğrenmesi algoritmalarının yürüttüğü eğitim sona erdiğinde genelde veri setinin ayrılan üçte birlik bölümü teste tabi tutulur ve sonuçlar oransal olarak başarı yüzdesini verir. Bu aşamada verilerin üçte ikisiyle eğitim yapan algoritma daha önce karşılaşmadığı verileri gözden geçirerek algoritmanın model üzerindeki başarı oranı ölçer. Başarıları doğru pozitif, doğru negatif, yanlış pozitif ve yanlış negatif olarak 4 gruba ayırır ve karışıklık matrisini oluşturur

#### **2.2.3 Karışıklık Matrisi**

Algoritmaların çıktılarının doğruluğu test etmek için kullanılan hedefteki sayıya göre boyutu genişleyen kare bir matristir. Çalışmada makine öğrenmesi yapılırken olan birçok algoritma işleme girmektedir. Öğrenmeden sonra teste tabii tutulur ve çıkan sonuçlar kullanıcı tarafından değerlendirilerek konu için en uygun en performanslı algoritma ve parametrelerine karar verilir. Karışıklık matrisinde algoritmanın tahmin ettiği sonuçların doğru ve yanlış tahminlerin sayısı gösterilir. Tahmin sütun ve satırları

aşağıda gösterilmektedir(Çizelge 2.1). Pozitif ve negatif şeklinde edeceği tahminler için aşağıdaki açıklamalar kullanılabilir.

**Çizelge 2.1:** Karışıklık Matrisi

TAHMİN	POZİTİF DURUM	NEGATİF DURUM
POZİTİF	TP	FP
NEGATİF	FN	TN

Burada TP, FP, FN ve TN durum ve tahmin arasındaki bağlantıdan net oranı vermek için kullanılır.

TP: Algoritmanın doğru tahmin ettiği pozitif değerlerin sayısını gösterir.

FP: Algoritmanın yanlış tahmin ettiği pozitif değerlerin sayısını gösterir.

FN: Algoritmanın yanlış tahmin ettiği negatif değerlerin sayısını gösterir.

TN: Algoritmanın doğru tahmin ettiği negatif değerlerin sayısını gösterir.

### 2.2.3.1 Tahmin Hatası

Karışıklık matrisinde algoritmanın doğru ve başarı elde etmek için kullandığı verilerden hatalı tahmin ettiği verilerdir. Bu bilginin elde edilmesinde Eş. 2.1'deki formül uygulanmıştır.

$$\frac{FP + FN}{TP + TN + FP + FN}$$

(2.1)

### 2.2.3.2 Doğruluk Oranı

Toplam olasılık 1 olduğu için tahmin hatası 1'den çıkarıldığı zaman doğruluk oranı elde edilir. Doğruluk oranı algoritmanın doğru tahmin ettiği sonuçların oranıdır. Formülü Eş. 2.2'de gösterildiği gibidir.

$$\frac{TP + TN}{TP + TN + FP + FN} \quad (2.2)$$

### 2.2.3.3 Geri Çağırma (Recall)

Birçok çalışmada anma, isabet oranı gibi isimleri de olan geri çağırma metriği çalışma içerisindeki pozitif sınıfa ait örneklerden olan doğru tahminlerin oranını vermektedir. Formülü Eş. 2.3' te verilmiştir.

$$\frac{TP}{TP + FN} \quad (2.3)$$

### 2.2.3.4 Hassasiyet (Precision)

Algoritmanın çalışmada bulunan verilerden toplam kaç tanesinin doğru bilindiği bu metrikte görüntülenir ve karışıklık matrisine dahil edilir. Formülü Eş. 2.4' teki gibidir.

$$\frac{TP}{TP + FP} \quad (2.4)$$

### 2.2.3.5 F Skoru

F skoru hassasiyet ve geri çağırma değerlerini barındıran istatistik tabanlı bilgisayar biliminde verilerin başarısı, doğruluğu için kullanılan bir ölçüt birimidir. Matematik bilimindeki harmonik ortalamayı kullanır ve tek bir sonuç verdiği için avantajlıdır. F skorunun formülü Eş. 2.5' te belirtilmiştir.

$$2 \frac{\text{Hassasiyet} \cdot \text{İsabet Oranı}}{\text{Hassasiyet} + \text{İsabet Oranı}} \quad (2.5)$$

## 2.3 Makine Öğrenmesinin Tıp Alanında Kullanımı ve Önemi

Makine öğrenmesi birçok alanda kullanılmaktadır. Kullanılan bu alanlardan en belirgin ve etkili olan tıp alanıdır. Bir canlının hayatı çok önemlidir ve hastalıklardan korunmalıdır. Hasta olduklarında ise erkenden bir hekime başvurmalı ve gerekli tetkikler yapıldıktan sonra tedavi sürecine başlanmalıdır. Makine öğrenmesi hekime başvurulmasının ardından devreye girmektedir hekimlerin gerekli tetkikleri yapabilmesi ve erken teşhis konulabilmesi için yardımcı olmaktadır. Teknolojini gelişimiyle birlikte tüm tıbbi cihazlar dijital hale gelmektedir. Bu da algoritmaları oluşturmak ve makine öğrenmesi yapılabilmesi için yeterli veri seti oluşumunu sağlamaktadır. Özellikle hayati önem taşıyan hastalıklar için bu konuda ciddi çalışmalar vardır. Bu çalışmada onlardan bir tanesidir. Tahmin gücü ve başarı oranı yüksek algoritmaların kullanılması erken teşhisin doğruluğunu arttırmaktadır. Bu konuda yapılan çalışmalar ve uygulanan cihazların başarısı makine öğrenmesinin tıp alanındaki önemini ispatlar niteliktedir. Ülkemizde dahil olmak üzere birçok ülkede bu yöntemler kullanılmaktadır. Öğrenmeye dayalı algoritmalar sayesinde hastalıklara teşhis koyulabilmekte ve hayatlar kurtarılabilmektedir.

## 2.4 Kullanılan Makine Öğrenmesi Yöntemleri

Makine öğrenmesinin farklı kütüphanelerinde barındırdığı birçok algoritma vardır. Bu çalışmada kullanılan makine öğrenmesi algoritmaları, Lojistik Regresyon, K-En Yakın Komşu, Destek Vektör Makineleri, Naive Bayes, Karar Ağacı, Rastgele Orman, LightGBM Model, XGBoost Model, Ridge Model ve Bagging Model algoritmasıdır.

### 2.4.1 K-En Yakın Komşu

K-en yakın komşu algoritması en sık kullanılan algoritmalarından biridir ve çok basittir. Ayrıca tembel bir algoritmadır, eğitim verişlerini anlamak yerine ezberler bu da eğitim verilerini öğrenmediği anlamına gelir. Çalışmanın sonucunda bir tahmin verilmesi gerekir ve bu tahmini bulabilmesi için algoritma tüm veri setinin tarar ve ezberlediği komşularını arar. Çalışmada belirlenen bir K değeri vardır ve bu değer algoritmanın veri setinde bakacağı elemanların sayısını belirtir. İstatistik tabanlı bir yöntemdir. Algoritmaya bir değer girdiğinde o değere en yakın olan K tane eleman alınır ve işleme



girer. Bu uzaklık Öklid fonksiyonu ile hesaplanır. Manhattan ve Minkowski fonksiyonları da kullanılmaktadır. Bu fonksiyonlar hesaplanan veriler sıraya dizilir ve değerin uygunluğuna göre sınıflandırılır. Bu aşamada kullanılan formül Eş. 2.6' da ki gibidir.

$$d(i, j) = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2} \quad (2.6)$$

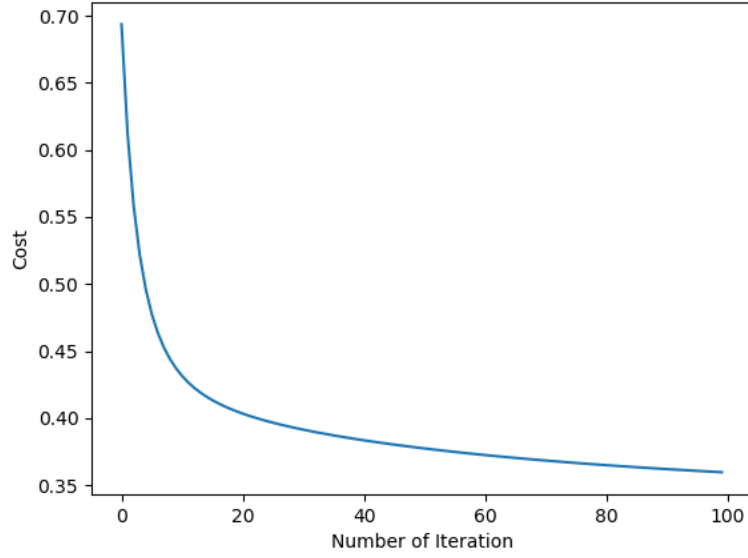
Karmaşık eğitim verilerine dirençli olması sebebiyle bilinirliği en fazla olan makine öğrenmesi algoritmalarından biridir. Verileri anlamayıp ezberlediği için büyük verilerde avantajlı değildir ve fazla hafızaya ihtiyaç duymaktadır.

#### 2.4.2 Lojistik Regresyon

Lojistik fonksiyonu, 19. Yüzyıl artan nüfus ve oto katalitik kimyasal tepkilerin seyrinin tanımlanması için icat edildi. Her iki durumda da  $w(t)$  miktarının zaman yolunu ve büyüme oranını dikkate alınmaktadır (Cramer, 2002). Lojistik fonksiyonunun formülü Eş. 2.7'de verilmiştir.

$$W(t) = \Omega \frac{\exp(\alpha + \beta t)}{1 + \exp(\alpha + \beta t)} \quad (2.7)$$

Eş. 1'de  $\Omega$ ,  $W$ ' nun doyma seviyesinin üst limitini,  $\alpha$  eğrinin x eksenindeki konumunu,  $\beta$  ise eğrinin eğimini temsil eder. Lojistik Regresyon istatistiksel bir metottur ve kategorileri sınıflandırmak için kullanılır. Bu yöntemde olasılık olarak bağımsız değişkenler ile sonuçta regresyondan çıkan sonuç değişkenleri arasındaki ilişkiyi incelenir ve hesaplanır (Baş ve diğerleri, 2018). Lojistik Regresyon eğrisi Şekil 2.4' te gösterilmektedir.



**Şekil 2.4:** Lojistik Regresyon Eğrisi

Çoklu doğrusal regresyon ile lojistik regresyon arasında azımsanmayacak derecede benzerlikler vardır. Katsayıların kestirilmesi, En küçük kareler yönteminin kullanılması gibi benzerlikler mevcuttur. Ayrıca lojistik regresyon matematik ve olasılık ilkelerine dayanır.

### 2.4.3 Naive Bayes

Naive Bayes, adını Thomas Bayes adlı bilim insanından alan bir sınıflandırma algoritmasıdır (Bayes, 1763). Bu sınıflandırma algoritmasının temeli tamamı ile olasılığa dayanır. Olasılık yardımıyla yaptığı hesaplamalar ile veri setindeki verilerin sınıflandırılmasını sağlar. Tembel bir algoritmadır. Ancak dengesiz ve düzensiz veri setlerinde çalışabilmektedir. Veri setindeki her bir eleman için ayrı ayrı tüm olasılıkları hesaplar ve olasılık değeri en yüksek olana göre sınıflandırır. Eğitim verisi ne kadar fazla ise o kadar kesin sonuç alınır ancak az olan veri ile de yüksek başarı oranıyla çalışabilmektedir. Bu algoritma bir teoreme dayanmaktadır ve formülü Eş. 2.8' de verilmiştir.

$$P\left(\frac{C_k}{x}\right) = \frac{P\left(\frac{x}{C_k}\right) \cdot P(C_k)}{P(x)}$$

(2.8)

$P(x)$ : Veri setinden bir herhangi bir verinin  $x$  olması

$P(C_k)$ :  $k$  sınıfının olasılığı

$P\left(\frac{C_k}{x}\right)$ :  $x$  örneğinin  $k$  sınıfından olması

$P\left(\frac{x}{C_k}\right)$ :  $k$  sınıfından bir olayın  $x$  örneğinin gerçekleşmesi

Test kümesinde olan bir verinin eğitim kümesinde ya da eğitim kümesinde olan bir verinin test kümesinde herhangi bir karşılığı yoksa olasılık hesaplayamaz tahmin yapamaz ve sonucu "0" olasılık olarak çıkarır.

Öneri sistemleri, anket analizleri, mail spam tespit etme ve metin kategorize edilmesi gibi birçok alanda kullanılmaktadır. Özellikle tahmini gerçek zamanlı olan uygulamalarda başarı oranı olarak çok daha yüksek sonuçlar vermektedir (Bayes, 1763).

#### 2.4.4 Destek Vektör Makineleri

Destek vektör makineleri 1963 yılında Alexey Chervonenkis ve Vladimir Vapnik tarafından keşfedilen istatistiksel öğrenme teorisine dayanan bir öğrenme algoritmasıdır. DVM 1995 yılında Vapnik ve ekibi tarafından daha da geliştirip bugünkü haline evrilmiştir. Uygulanması bir çizgi yardımıyla yapılmaktadır. İki bölüme ayrılmış verileri uygun olan şekilde birbirinden ayrılması için kullanılır. Diğer makine öğrenmesi algoritmalarına karşı birçok avantajı vardır.

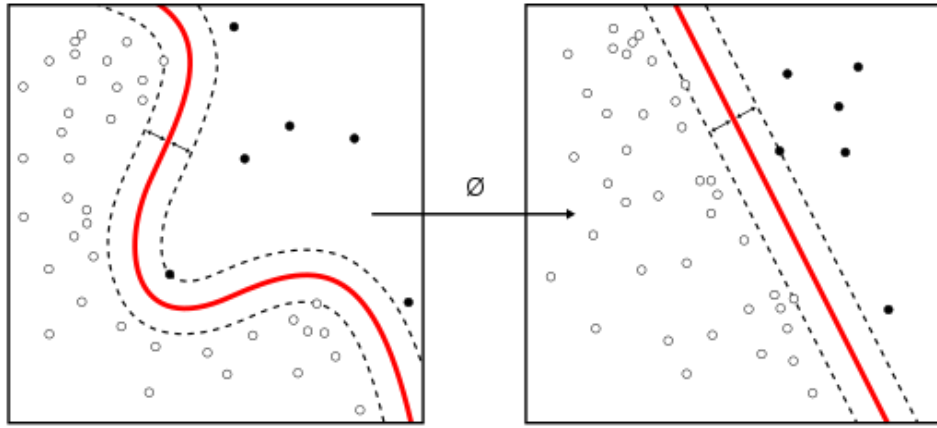
- Çalışmadaki örneklem sayısının, boyut sayısından az olduğu durumlarda etkilidirler.
- Farklı çekirdek fonksiyonlarını karar mekanizmasında kullanır.
- Büyük verilerde daha etkili ve başarılı sonuçlar verir.
- Çok sayıda bağımsız değişkenle çalışılabilmektedir.

Avantajları olduğu gibi dezavantajları da mevcuttur. Örneğin boolean tahminler, olasılıksal tahminler üretemez. DVM veri setindeki verilerin doğrusal olup olmamasına bağlı olarak değişkenlik gösterir ve iki grupta incelenir.

#### 2.4.4.1 Doğrusal Olmayan Destek Vektör Makineleri

Doğrusal olmayan DVM, veri setini doğrusal olarak ayrılmayarak hatanın belirli belirsiz olduğu durumlarda belirli bir hata almadan ayrılması durumunda oluşan öğrenme algoritmasıdır. Normal yaşantıdaki olayların normal düzlem üzerinde doğrusal şekilde ayrılması pek mümkün değildir. Bundan dolayı ayırma eğrisi sınıflamadan en fazla öneme sahip özelliktir. Sınır düzlemindeki ayrılma eğrisi Şekil 2.5’ te gösterilmiştir. Verilere çekirdek fonksiyonu adı altında fonksiyonlar uygulanır ve özelliklerine göre sınıflandırılır. Uygulanan çekirdek fonksiyonlarından doğrusal çekirdek fonksiyonu Eş. 2.9’ da gösterilmiştir.

Çekirdek fonksiyonları ayrılanamayacak olan problemleri ayırabilmeyi sağlamaktadır ve veriyi haritalama yaparak daha iyi bir gösterim sunmak için kullanılır.



Şekil 2.5: Destek Vektör Makineleri Sınır Düzlemi

Lineer Çekirdek Fonksiyonu

$$K(x_i, x_j) = (x_i^T \cdot x_j)$$

(2.9)

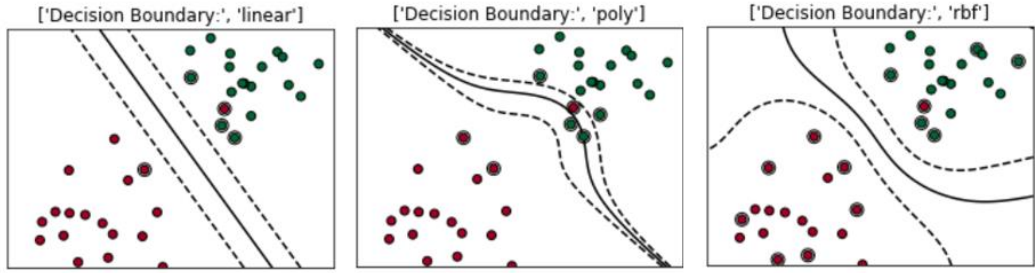
Polinomsal Çekirdek Fonksiyonu Eş. 2.10'de gösterilmiştir.

$$K(x_i, x_j) = (x_i^T \cdot x_j)^d \quad (2.10)$$

Gaussian Radyal Çekirdek Fonksiyonu Eş. 2.11'de gösterilmektedir.

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (2.11)$$

DVM çekirdek fonksiyonlarının ayrıntılı olarak sınır düzlemleri ve verilerin dağılımları ile sınıflandırması Şekil 2.6'da gösterilmiştir.



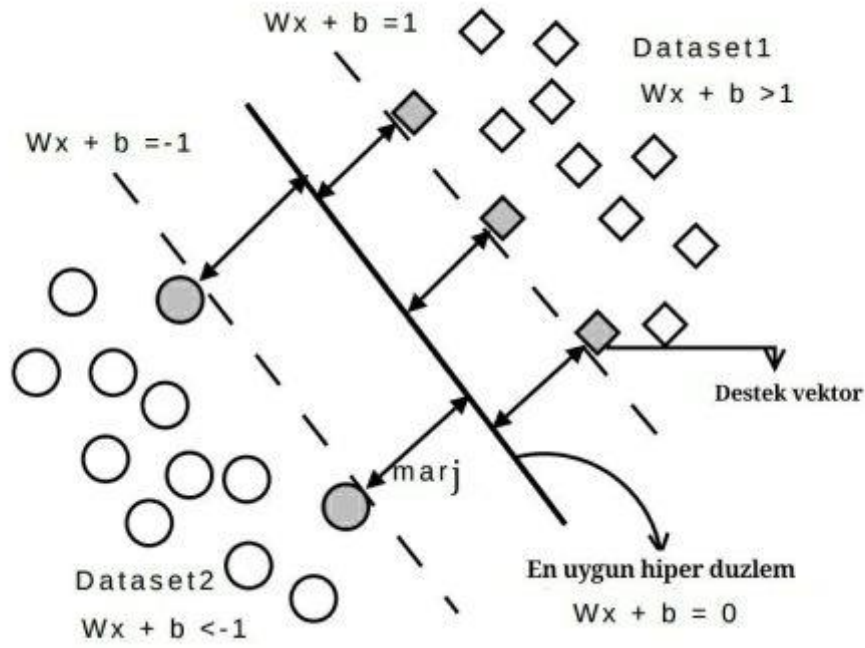
Şekil 2.6: Destek Vektör Makineleri Çekirdek Fonksiyonları

#### 2.4.4.2 Doğrusal Destek Vektör Makineleri

Doğrusal olarak ayrılabilen veriler sayesinde iki sınıfın birbirinden ayrıştırılabildiği sınır düzleminin doğrusal olduğu Doğrusal DVM Eş. 2.12'de gösterilmiştir.

$$F(x) = w^T \cdot x + b = \sum_{i=1}^n w_i \cdot x_i + b \quad (2.12)$$

Doğrusal DVM' de ayrılabilme durumunda veri setlerinin karşılaştırıldığı, marj aralığının belirtildiği destek vektörleri ve oluşan en uygun hiper düzlemi şekil 2.7' de verilmiştir.



**Şekil 2.7:** Doğrusal Ayrımda Oluşan Hiper Düzlem ve Destek Vektörleri

#### 2.4.5 Karar Ağacı

Veri madenciliğinde en çok kullanılan yöntemlerden biri olan karar ağaçları diğer sınıflandırma algoritmalarında olduğu gibi karar ağaçlarında da asıl olan amaç sınıflandırmaktır ve karar ağaçları rastgele orman algoritması gibi algoritmalara da ilham olmuştur. Veri setinden gelen öznelilikler düğümlerdir ve bu düğümler doğru-yanlış, evet-hayır gibi soruları cevaplar ve bu şekilde her bir düğümden veri ikiye bölünür. Bölünen veriler özneliliklere etki eden özellik vektörleri incelenir ve başarı oranı ve bilgisi en yüksek olan düğüm dallanma yapmak için algoritmaya girer. Dallandıktan sonra algoritma tüm verileri sınıflandırmaya kadar son bulmaz ve devam eder.

Karar ağacı yapısı olasılık ve istatistik temelli bir algoritmadır. Özellikle istatistik biliminde karmaşıklık değeri olarak ta bilinen entropi çok önemlidir. Entropi beklenmeyen durumların olma olasılığına denir. Ya da algoritmada olabilecek sapmaları öngörmek için gereklidir. Entropinin formülü Eş. 2.13' te verilmiştir.

$$H(x) = - \sum S(x) \log S(x) \quad (2.13)$$

Entropi formülünde  $S(x)$  belirli bir sınıfa ait grubun oranını vermektedir ve logaritmik çarpımının ters integrali alınarak  $H(x)$  bulunur. Burada  $H(x)$  ise o grubun Entropi'sinin kendisidir.

$$Kazanc(S, D) = H(S) - \sum_{V \in D} \frac{|V|}{|D|} H(V) \quad (2.14)$$

Karar ağaçlarında bir diğer önemli belirleyici ise bilgi kazancıdır. Bilgi kazancı oluşturulan alt grubun entropi değeri ile tüm veri setinin entropi değeri arasındaki farktır. Formülü Eş 2.14'te verilmiştir. Burada  $S$  tüm veri setini,  $D$ ,  $S$ 'nin parçalanmış bir alt bölümüdür.  $V$  ise  $D$ 'nin altında kalan bir karar mekanizmasıdır.

#### 2.4.6 Rastgele Orman

2001 senesinde Leo Breiman rastgele orman algoritmasını keşfedilen bir yaklaşımdır (Breiman, 2001). Rastgele orman algoritması, karar ağaç algoritmasının  $n$  defa uygulanmasıyla tahminlerin gücünü arttırmak için kullanılan bir metottür. Parametreleri yardımıyla arzu edilen sayıda ağaç oluşturulabilmektedir. Ayrıca en uygun özneliği belirlemede kullanılan en yaygın algoritmadır ve hem regresyon hem de sınıflandırma görevleri için kullanılabilir. Bundan dolayı diğer sınıflandırma ve regresyon algoritmalarının içerisinde en avantajlısıdır.

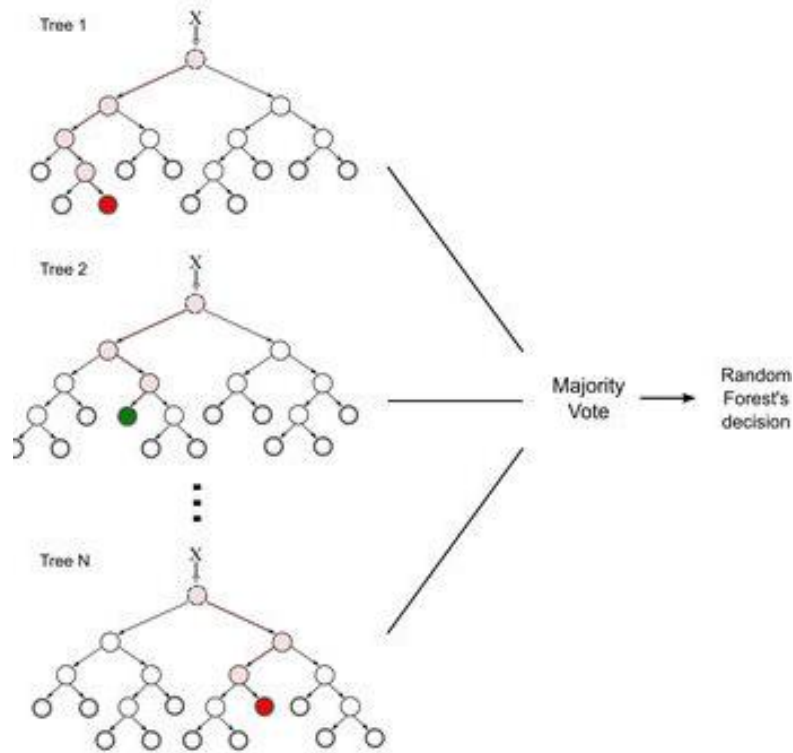
Rastgele orman algoritmasında iki adet parametre kullanılır. Bunlar Ağaç sayısını belirlemek için ( $N$ ) ve düğümlerdeki değişken sayısını belirleyen ( $m$ ) parametreleridir. Dallara ayrılmak için seçilen değişkenin uygun test kriteri (cut-off değeri) "gini katsayısı" ile belirlenir (Akman ve diğerleri, 2011). GINI katsayısının formülü Eş. 2.14'teki gibidir.

$$GINI(T) = 1 - \sum_{k=1}^n (S_k)^2. \quad (2.14)$$

GINI indeksi her düğümde hesaplanır. GINI indeksinin sıfır olması durumunda dallanma son bulur ve düğümün homojenliği kanıtlanır. En düşük hata oranına sahip ağaç en yüksek ağırlığı, en yüksek hata oranına sahip ağaç ise en düşük ağırlığı alır (Akman ve diğerleri, 2011). Sonrasında ağaç ağırlıklarına göre sınıflar oy kullanır ve oylar toplanır. En yüksek oya sahip ağaç yapısı belirlenerek karar verilir.

Rastgele orman algoritmasının kalite ve algoritmik olarak önemini Sklearn kütüphanesi pekiştirmiştir. Sklearn oluşturulan düğümler için ormandaki ağaçların kirliliği azalma miktarına bakarak, özelliklerden birini ön planda tutarak algortimanın daha kolay ve etkili bir sonuç elde etmesine yardımcı olur.

Rastgele orman algoritmasını elde etmek için hem değişken bazında hem de gözlem bazında CART algoritmasını bağ edilmesi gerekir. Yani değişken ve gözlem bazında rassallık sağlanırsa hem aşırı öğrenme düşer hem de tahmin oranı artar.

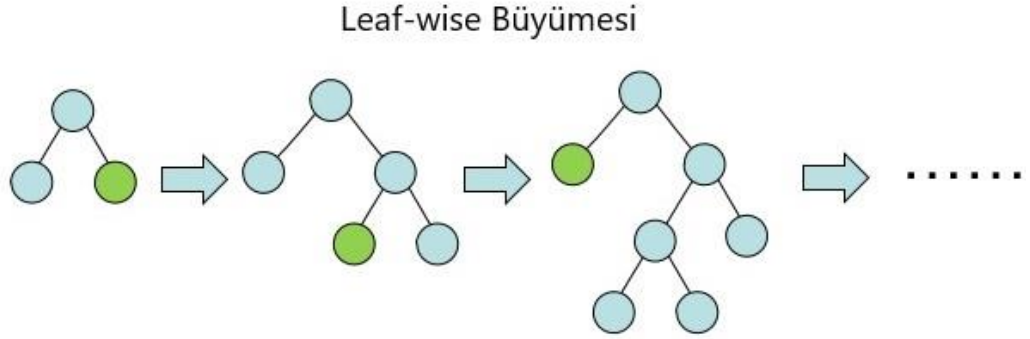


**Şekil 2.8:** Rastgele Orman Algoritması ve Karar Ağaçları İlişkisi



## 2.4.7 LightGBM

Light Gradient Boosting Machine algoritması karar ağacı alt yapısını kullanan hızlı ve yüksek performanslı gradient boosting framework' ü olarak tanımlanabilir. Diğer karar ağacı alt yapısı olan algoritmalarından farklı olarak LightGBM algoritmasında ağaç dikey büyürken diğer tüm algoritmalarda yatay büyümektedir. Özellikle sınıflandırma ve sıralama için kullanılabilir bir yapıdadır. Aslında temelde yatan ve geliştirilen fikir XGBoost yöntemidir. LightGBM, XGBoost algoritmasının eğitim süresinin performansını arttırmak için Microsoft'un gradient boosting altyapısı ile geliştirdiği bir yöntemdir. Level-wise yerine Leaf-wise büyüme planını kullanır.

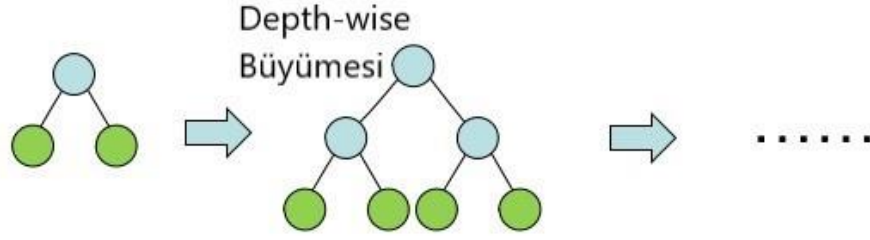


**Şekil 2.9:** LightGBM Mimarisi

Son zamanlarda veri sayısının artmasıyla birlikte, LightGBM hızı konusunda dünden bugüne gelen veri bilimi algoritmalarının önüne geçmiştir ve sınıflandırma amacıyla kullanılmaktadır. Büyük veri setlerinde kullanılmaya uygundur, küçük veri setlerinde kullanılır ancak efektif bir sonuç veremeyebilir. Uygulanması kolay olan LightGBM için parametre ayarlaması zordur. Çünkü toplamda 100'e yakın parametresi mevcuttur. Parametreler arasında en yüksek başarı oranını elde etmek için hangi parametrenin kullanılacağı ve değerinin kaç olacağı gibi tahminler yapılması gerekir. Bu da farklı kombinasyonlar ile hesaplandığında çok fazla bir zamana tekabül etmektedir.

## 2.4.8 Extreme Gradient Boosting

XGBoost son derece ölçülebilir, esnek ve çok yönlü bir araçtır, kaynaklardan doğru bir şekilde faydalanmak ve önceki gradient boosting kısıtlamalarını aşmak için oluşturulmuştur. XGBoost ile LightGBM ve CATBoost arasındaki temel fark, aşırı uyumu kontrol etmek için yeni bir düzeltme tekniği kullanmasıdır. Dolayısıyla, model uyumu esnasında daha hızlı ve güçlüdür (Daoud, 2019).



Şekil 2.10: XGBoosting Mimarisi

Extreme Gradient Boosting algoritması yaygın kullanımı olan yüksek verimli, esnek olacak şekilde tasarlanmış ve optimize edilmiş gradient boosting kütüphanesidir. Makine uygulaması algoritmalarını uygularken gradient boosting kütüphanesini esas alır.

## 2.4.9 Bagging

Bagging yöntemi de Gradient Boosting metodunu kullanır. Bu yöntemde modele giren veri setinden çıkartılan eğitim verisi kullanılarak yeni eğitim verileri üretilir ve tekrar tekrar eğitim yapılır. Tekrarlanan eğitim kümesinden rastgele seçimler ile yeni bir eğitim verisi türetilir. Eğitim veri setinin içerisinde alınan örnekler tekrar eğitim verisine geri iade edilir.

Bagging yöntemi aşırı öğrenme ve model performansını arttırmaya yönelik ortaya çıkmış bir yaklaşımdır. Bagging “Bootstrap aggregaiton” ifadesinin kısaltılmış haldir. Ağaç yapısında olduğu için topluluk öğrenme yöntemleri arasındadır. Bagging yöntemi elde edilmesi kolay bir algoritmadır. Cart algoritması bag edildiğinde Bagging yöntemi ortaya çıkar (Breiman, 1994).

Bagging ile boosting arasında en büyük fark ağaçların birbiri ile bağımlı olup olmama durumu ve hata optimizasyonu bir önceki ağaçlardan birikmiş olarak toplanıp toplanmamasıdır.

#### 2.4.10 Ridge

Ridge regresyonu, çoklu eşdoğrusallık problemi olan çoklu regresyon verilerinin analiz edilmesi için kullanılan bir yöntemdir. Çoklu eşdoğrusallık mevcut olduğunda, en küçük kareler yönteminin tahminleri tarafsız olur ancak varyansları çok daha büyük olduğu için bu tahminler gerçek değerden çok uzak olabilir (Göktaş, 2016). Ridge içinde bulunduğumuz çağda keşfedilen bir algoritmadır. Yeni kullanılmaya başlanmıştır regresyon uygulanacak olan problemlere genelde lojistik regresyon uygulanmaktadır. Lojistik regresyon Ridge regresyona göre çok daha performanslı çalışmaktadır. Özellikle çağımızda Boosting ve Bagging kütüphanelerinin ve çalışmalarının olduğu varsayılırsa Ridge bu yöntemlerin biraz gerisinde kalmıştır.

$$\sum_i = 1N\{y_i - \sum_{j=0}^M \beta_j x_{ij}\}^2 + \lambda \sum_{j=0}^M W_j^2$$

(2.15)

Ridge regresyon, verilerin aşırı öğrenmesini ve katsayıların büyüklüğünü ölçmek için kullanılır. Ridge regresyonunun hesaplama formülü Eş. 2.15'te verilmiştir.

### 3. PROJE MODELİ

Bu çalışmanın 165'i kalp hastası olan bir topluluktan alınan 303 insanın çeşitli özelliklerini barındıran Heart Disease UCI veri seti kullanılmış ve Python dilinde derleyici olarak IDLE kullanılarak yazılmıştır. Çalışma 16 GB Ram, 4 GB Ekran Kartı i7 8. Nesil işlemcisi olan bir dizüstü bilgisayarda yapılmıştır. Programın stabil çalışabilmesi için toplam 84 kez çalıştırılmıştır. Açık kaynak kodlu derin öğrenme destekli yapay zekâ uygulamalarını geliştirilmesini sağlayan Google şirketinin ürettiği Tensorflow derin öğrenme kütüphanesini içerisinde barındırdığı bilimsel hesaplamalar ve matematik konusunda önemli bir kütüphane olan Numpy, .txt gibi dosyalardan veri okunmasına yardımcı olan ve veri analiz araçlarını içerisinde barındıran Pandas, veri görselleştirmeleri için kullanılan Matplotlib ve Seaborn kütüphaneleri kullanılmıştır.

Ayrıca çalışmada kullanılan algoritmalar için ise Scikit-Learn kütüphanesi kullanılmıştır.

Normalize edilen veri seti, makine öğrenmesi algoritmalarıyla tek tek işlem görmüştür. 303 verinin 4/5'si(%80'i) eğitime girmiştir, kalan 1/5(%20) veri ise test için kullanılmıştır. Kalp hastalıkları ile ilgili Python dilinde makine öğrenmesiyle daha önce yapılan çalışmalarda başarı oranı olarak maksimum %88,56 değerine Rastgele orman ve KNN Algoritmalarıyla ulaşılmıştır. Mevcut çalışmada bu değeri arttırmak ve farklı algoritmalar kullanılarak doğruluk oranını arttırmak amaçlanmıştır.

### 3.1 Veri Seti

Macaristan Kardiyoloji Enstitüsü'nden Andras JANOSI, Zürih Üniversite Hastanesi'nden William STEINBRUNN, Basel Üniversite Hastanesi'nden Matthias PFISTERER, Cleveland Klinik şirketinden Ph. D. Robert Detrano toplamda 303 hasta ile çalışmış ve tüm verileri California Üniversitesi'nde birleştirerek bu veri setini oluşturmuşlardır. Veri setinde toplam 13 öznitelik konu alınmış ve bu özniteliklerin başlıca kullanılanların anlamı ve kullanıldığı yerler neden tercih edildikleri açıklanmıştır. Öznitelikler aşağıda gösterilmiştir(Çizelge 3.1).

**Çizelge 3.1: Öznitelikler**

Öznitelik No	Öznitelik Bilgisi
1	Yaş
2	Cinsiyet
3	Göğüs Ağrısının Şiddeti
4	Tansiyon(Dinlenme)
5	Kandaki Serum Kolesterol Oranı
6	Açlık Kan şekeri > 120 mg/dl
7	Elektrokardiyografi sonuçları (0,1,2)
8	Maksimum Kalp Atış Hızı
9	Egzersiz Nedenli Anjin
10	Egzersiz Nedenli S ve T Dalga Depresyonu
11	S ve T Dalga Segmenti tepe eğimi
12	Floroskopi ile Renklendirilmiş Büyük Damar sayısı (0-3)
13	Kalp Hastalığı Tanısı (Anjiyografik Hastalık Durumu) (0,1)

Göğüs Ağrısının Şiddeti: Kalp hastalıklarında en belirgin olan ve belirtilerin başında gelen göğüs ağrısının veri setinde dört tipte incelenmiştir. Hiç ağrı olmaması 0, az ağrı 1, orta şiddette 2 ve çok şiddetli ağrı 3 olarak belirtilmiştir.

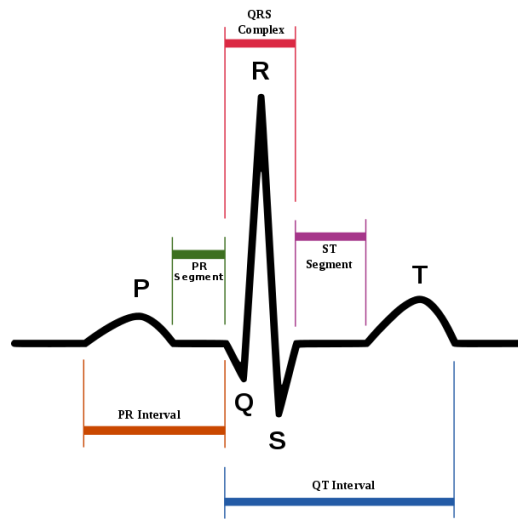
Kandaki Serum Kolesterol Oranı: Kanda bulunan maddelerden özellikle kolesterolün yol açtığı damar tıkanıklıkları miktarı da kalp hastalıkları için önemlidir.

Açlık Kan şekeri > 120 mg/dl: Şeker de aynı kolesterol gibi damarları tıkaçıcı ve katmanlaştırıcı etkiye sahiptir o yüzden kalp hastalıkları tahmininde şeker oranının 120 den az ya da fazla olma durumu ele alınmıştır.

EKG: Elektrokardiyografi çekilen hastalarda bulunan semptomlar bu veri setine dahil edilmiştir. Üç alanda incelenmektedir.

- Değerin 0 olduğu durumda normal olduğu,
- Değerin 1 olduğu durumda S ve T dalga anormalliği olduğu atış hızında bir farklılık olduğu,
- Değerin 2 ise hipertansiyon ve kandaki basınç ile birlikte ventrikül hipertrofiye olduğu belirtilmektedir.

ST Segmenti Tepe Eğimi: Pik egzersizde ST segmentinin kalp atış dalgalarının top noktasını ifade etmektedir. ST segmenti Kalp atışının durağanlaştığı noktalarda gözlemlenmektedir. Grafiği Eş. 3.1’de verilmiştir.



**Şekil 3.1:** EKG Dalgaları ve Segmentleri

Floroskopi ile Renklendirilmiş Büyük Damar Sayısı: Hastaların floroskopi yapılarak belirlenen büyüme gösteren damar sayıları 0 ile 3 arasında değer verilmektedir.

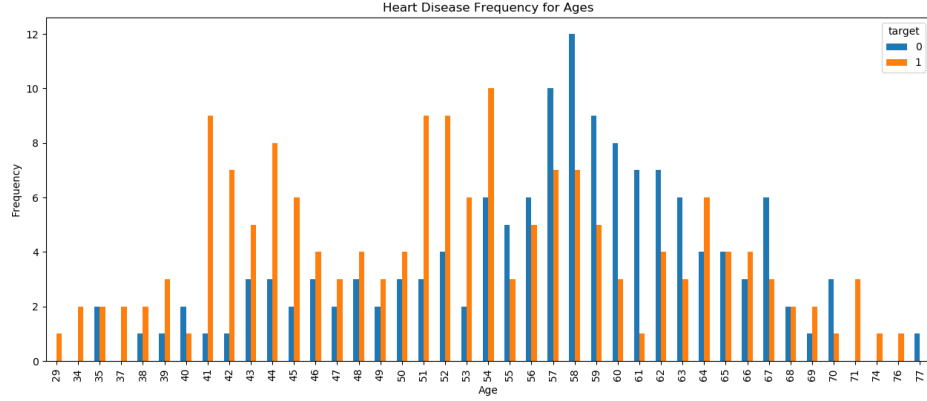
Anjiyografik Hastalık Durumu: Kalp hastalığı tanısı koyulmuş anjiyo yapılan ve çıkan sonuçlarının değerlendirildiği hastaların damar çaplarının %50'nin üzerinde ya da altında olması durumuna göre 0 ve 1 olarak derecelendirilmiştir.

## 3.2 Öznitelik Kullanımı ve Grafikleri

### 3.2.1 Yaş

İnsanlar yaşlandıkça vücutta çeşitli problemler ile karşılaşmaktadırlar. Bu yüzden kalp problemleriyle karşılaşılabilenler için yaş öznitelik olarak eklenmiştir.

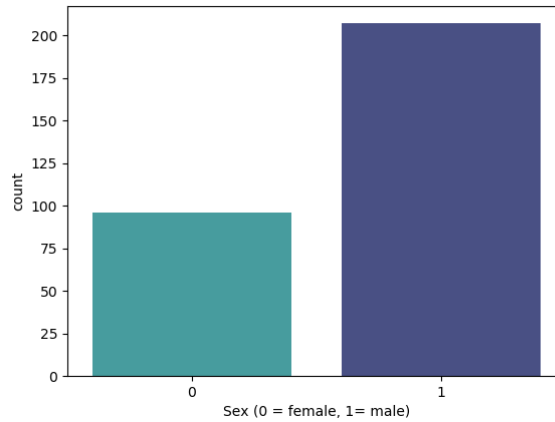
Çalışmada kullanılan yaş grafiği Eş.2 3.2’de verilmiştir.



Şekil 3.2: Veri Setindeki Hastaların Yaş Grafiği

### 3.2.2 Cinsiyet

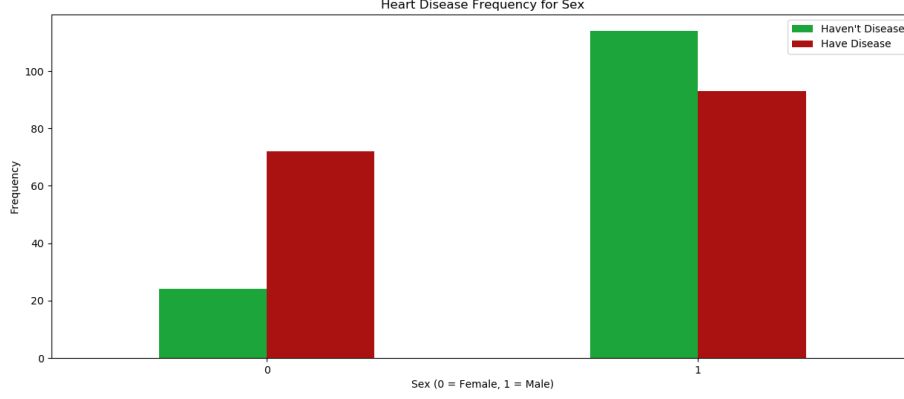
Özellikle kalp hastalıkları kadınlar ve erkeklerde farklı yaşlarda ve farklı sonuçlarla yaşandığı için öznitelikler içerisine eklenmiştir. Toplam 303 kişiden 98’i kadın, 205 ise erkek algoritmalara girmiştir. Cinsiyet Grafiğini Eş. 3.3’te görülebilir.



Şekil 3.3: Veri Setindeki Hastaların Cinsiyet Grafiği

### 3.2.3 Cinsiyete Göre Hasta Olup Olmama Durumu

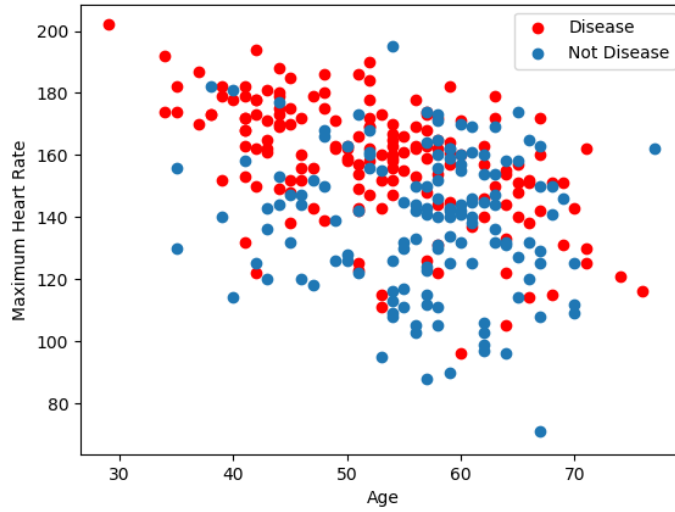
Kalp hastalığı cinsiyetlere göre ayrıldığında erkeklerin kadınlara göre rakam olarak daha fazla, ancak veri setinde oransal farklılıklarda kadınların erkeklere göre çok daha fazla kalp hastası olduğu Eş. 3.4'teki grafikte belirtilmiştir.



Şekil 3.4: Cinsiyete Göre Hasta Olup Olmama Durumu

### 3.2.4 Maksimum Kalp Atış Hızı, Yaş ve Hasta Olup Olmama Durumu

Çalışmaya dahil edilen özneliklerden bir tanesi de kalp atış hızıdır. Kalp atış hızı yaş ve kalp hastası olup olmama durumuna göre farklılık gösterebilir. Veri setindeki kalp atış hızları Eş. 3.5'teki noktasal grafikte verilmiştir.

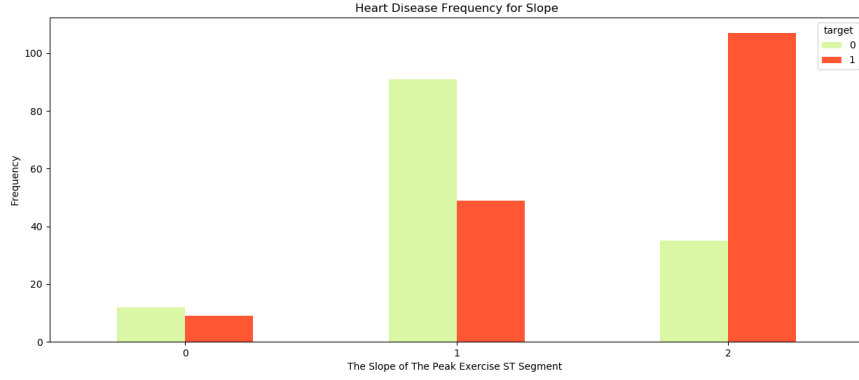


Şekil 3.5: Maksimum Kalp Atış Hızı, Yaş ve Hasta Olup Olmama Durumu



### 3.2.5 Pik Egzersizde ST Segmenti Eğimi

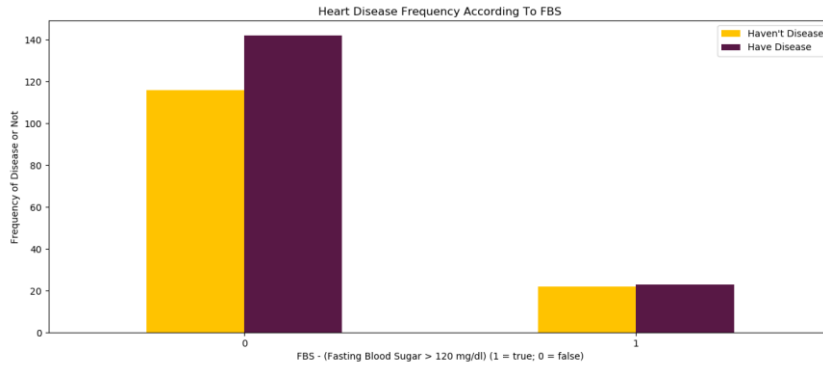
ST segmenti, süresi kalp atış hızıyla ters orantılı olarak gerçekleşen kalp atış hızları arasındaki elektriksel durgunluğu yassı şeklinde gösterir. Eş. 3.6'daki grafikte kalp hastası olanlar 1, olmayanlar 0 ile gösterilmiştir. ST segmentinin olmaması 0, az olması 1, çok olması ise 2 oranları ile tanımlanmıştır.



Şekil 3.6: Pik Egzersizde ST Segmentine Göre Kalp Hastalığı Durumu

### 3.2.6 Kandaki Şeker Oranı > 120 mg/dl Olan Kişilerin Hastalık Durumu

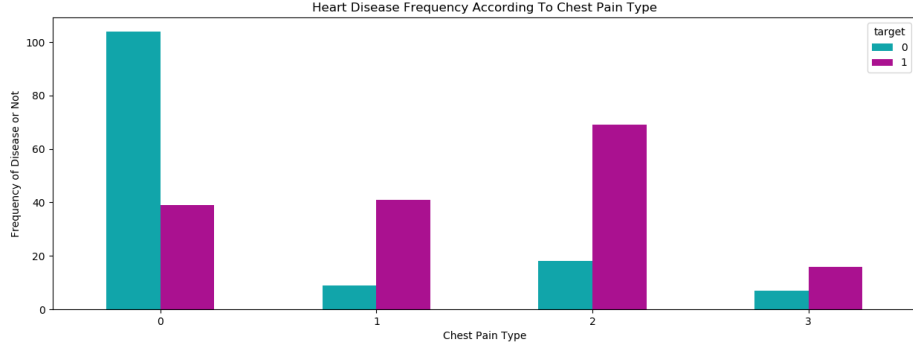
Veri setinde kandaki şeker oranının (kan şekerinin) 120'den büyük olup olmama durumu ile kişilerin hasta olup olmama durumu karşılaştırılmıştır. Kan şekeri 120'den küçük olanlar 0, büyük olanlar ise 1 ile gösterilmiştir.



Şekil 3.7: Veri Setinde Kullanılan Hastaların Yaş Grafiği

### 3.2.7 Göğüs Ağrısı Tipine Göre Kalp Hastalığı Sıklığı

Göğüs ağrısı 4 tip olarak tanımlanmıştır. 0 ile 3 arasındaki değerler verilmiştir ve toplamda hangi ağrı tipinde kaç adet hasta olduğu bulunması amaçlanmıştır. Eş. 3.8 'de grafiği verilmiştir.

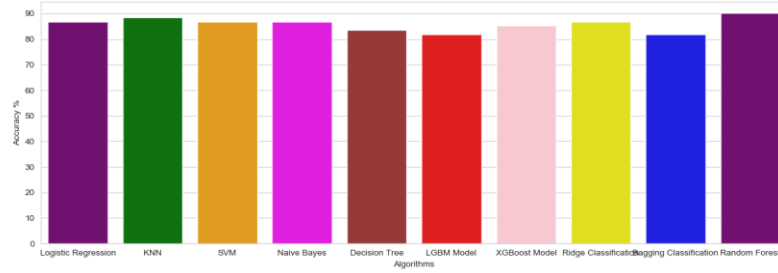


Şekil 3.8: Göğüs Ağrısı Tipine göre Kalp Hastalığı Sıklığı

#### 4. DEĞERLENDİRME

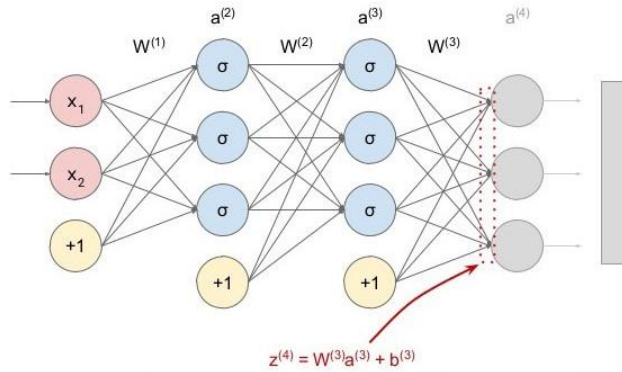
Makine öğrenmesi yöntemlerinden 10 tanesi karşılaştırılmıştır ve en yüksek başarı oranı rastgele orman algoritmasından %90,16'lık doğrulukla elde edilmiştir. Tüm algoritmaların karşılaştırma grafiği Eş.4.1'de verilmiştir.

Lojistik regresyon algoritmasında öncelikle normalize edilen data sklearn kütüphanesini kullanarak tüm algoritmaların kullanacağı gibi veri setini %80'i eğitim verisi olarak %20'si test verisi olarak kullanılmıştır.



Şekil 4.1: Algoritma Oranları Sonuç Grafiği

Lojistik regresyon algoritmasında kullanılan ileriye doğru yayılma yönteminin formül ile belirtilmiş hali ve yayılımı Eş. 4.2' de gösterilmiştir. Grafikte bulunan (W) ağırlık matrislerini, (a) ise vektörleri temsil etmektedir. Hassasiyet parametreleri olan sapma(bias) değeri 0 olarak, ağırlık(weight) değeri ise 0.01 olarak belirtilmiştir. Lojistik regresyon algoritma parametresi olarak farklı bir parametre uygulanmamıştır. Bu haliyle sonuç olarak başarı oranı %86,89 olarak bulunmuştur.



Şekil 4.2: Lojistik Regresyonda İleri Yayılma Yöntemi

K-En Yakın Komşu algoritması yine normalize edilen veriyi kullanarak ve iki farklı parametre kullanarak %85,25'luk bir başarı oranı elde edilmiştir. Kullanılan parametreler; komşu sayısını belirleyen `n_neighbors` ve komşu aramak için paralel olarak çalıştırılacak olan görev sayısını belirleyen `n_jobs` parametresidir. Değer olarak sırasıyla 5 ve 2 verilmiştir. Ancak en iyi K değeri bu değildir. Bu yüzden en iyi K değerini bulmak için en yakın k komşularının ortalamalarını bularak tekrar eğitime girmesini sağlamak daha iyi bir başarı oranı bulmaya yardımcı olmuştur ve bu testin sonucunda çıkan başarı oranı %88,52 olarak belirlenmiştir.

Destek vektör makineleri algoritmasında daha öncesinde yapılan çalışmalarda `random_state`, `probability`, `gamma` gibi çeşitli parametreler kullanılmış ancak kullanılan veri seti için parametresiz şekilde çalışan bu çalışmadan daha yüksek bir sonuç bulunamamıştır. Sonuç olarak bulunan başarı oranı %86,89 olarak belirtilmiştir. Lojistik Regresyon yöntemiyle aynı başarı oranını yakalamıştır.

Naive Bayes kullandığı Gauss modelinde başarı oranı %86,89 olarak bulunmuştur. Yine DVM gibi parametre kullanılmamıştır. DVM ve LR ile aynı başarı oranı sağlanmıştır.

Karar Ağacı algoritması aslında kullanılan veri setine en uygun algoritmalarından biridir. Hatta kullanılan birçok algoritma kök olarak ağaç yapılarından türemiş ve üretilmiştir. Çalışmada karar ağacı algoritması parametreleriyle birlikte kullanılmış ve en iyi sonucu bu şekilde verdiği gözlemlenmiştir. Kullanılan parametrelerden `max_depth=5`, `Random_state = 10` olarak belirlenmiştir. `Max_depth` ağacın maksimum genişleme derinliğidir. Eğer parametre olarak tanımlanmazsa ağacın düğümler tüm yapraklar çözümlenene kadar genişlemeye devam eder. `Random_state` ise ağacın durma noktasıdır rastgele sayı üreticidir. Aslında modelin hızını artırır ve aynı eğitim verisi ve parametreleri olduğunda her zaman aynı veriyi vermesine olanak sağlar. Karar ağacının bu parametreler yardımıyla bulduğu başarı oranı %83,61'dir ve bu oranla şu ana kadar başarı ile tanımlanan tüm algoritmaların gerisinde kalmaktadır.

LGBM Model alt yapısı zaten gradient boosting olduğu için boosting algoritmalarına yakın sonuçlar vermektedir. Çalışmada kullanılan parametresi bulunmamaktadır. Algoritmanın başarı oranı olarak çıkarttığı %81,97 oranı şimdiye kadar ki açıklanan başarı oranlarının en düşük olanı olduğu için sonuca olumlu bir etkisi bulunmamaktadır.

XGBoost Algoritmasının gradient boosting ve sklearn kütüphanesini kullanarak ortaya çıkarttığı başarı oranı %85,25 olarak saptanmıştır. Parametresiz olarak bulunan bu başarı oranıyla birlikte daha önce uygulanan algoritmalarından çıkan sonuçlardan en yüksek başarı oranını geçememiştir.

Ridge Algoritması aslında regresyon verilerini analiz etmek için uygun olan çalışmalarda kullanılsa da burada da kullanılması sayısal veriler olduğu için uygun bulunmuştur. Lojistik regresyon, DVM, Gauss yöntemi gibi kullanılan algoritmalar ile aynı başarı oranı olan %86,89 çıktısını vermiştir.

Bagging sınıflandırma algoritması verdiği sonuç olarak çalışmanın en düşük başarı oranı sonucunu veren algoritma olmuştur. Herhangi bir parametre kullanılmamıştır. Daha önce yapılan çalışmalarda bu veri seti üzerinde Bagging yönteminin kullanıldığı gözlemlenmemiştir. Sonuçta çıkan başarı oranı %80,33 olarak belirtilmiştir.

Bu çalışmada en fazla üzerinde durulan algoritma olan rastgele orman ayrıca çalışmadaki en yüksek başarı oranı değerinin de sahibidir. Rastgele orman algoritması aslında bir çok karar ağacının birleşip karar almasından farklı bir şey değildir. Birçok parametresi vardır. Bu çalışmada kullanılan parametreleri ve değerleri aşağıdaki gibidir.

**Çizelge 4.1:** Rastgele Orman Parametreleri

Parametre	Değer
N_estimators	350
Random_state	1
Max_depth	5

N\_estimators parametresi oluşturulan ormandaki karar verecek ağaç sayısını belirlemeye yardımcı olur. Bu durumda toplamda 350 adet ağacın ormana girişine izin verilmiştir. Random\_state hem ağaçlar oluştururken kullanılan örneklerin önceden yapılan tahminlerin rasgeleliğini hem de her bir düğümdeki en iyi ayrımı ararken göz önünde bulundurulacak özelliklerin örneklenmesini kontrol eder.

Max\_depth ağacın ne zamana kadar dallanabileceğinin sınırıdır. Ağacın maksimum

derinliđi de denir. Parametre olarak olmama durumunda dđđümler tamamen çözümlenene kadar dallanmaya devam eder ve bu durum hız ve doğruluk olarak net sonuç alınmasını zorlaştırır. Rastgele orman algoritması özellikle sađlık alanındaki çalıřmalarda sıkça kullanılmaktadır. Çünkü tıptaki laboratuvar bileřenlerinin doğru tanımlanması ve tahmin edilmesi, hastalıkların tahmini, hastalık tahmininden sonra hastanın tıbbi geçmiřini analiz ederek uygulanacak tetkikler ve tanı yöntemleri gibi birçok kararın verilmesine yardımcı olur. Rastgele orman yöntemi daha önce bu alanda yapılmıř çalıřmalarda kullanılmıř ancak akademik olarak herhangi bir çalıřmaya rastlanmamıřtır. Özellikle çalıřmada kullanılan veri setine birçok kez rastgele orman algoritması uygulanmıřtır. Alınan sonuçlarda en fazla başarı oranı %88,52 olarak sađlanmıřtır. Yapılan çalıřmada rastgele orman algoritması ve parametrelerinin yardımıyla Heart Disease UCI veri setinde %90,16 başarı oranı elde edilmiřtir.

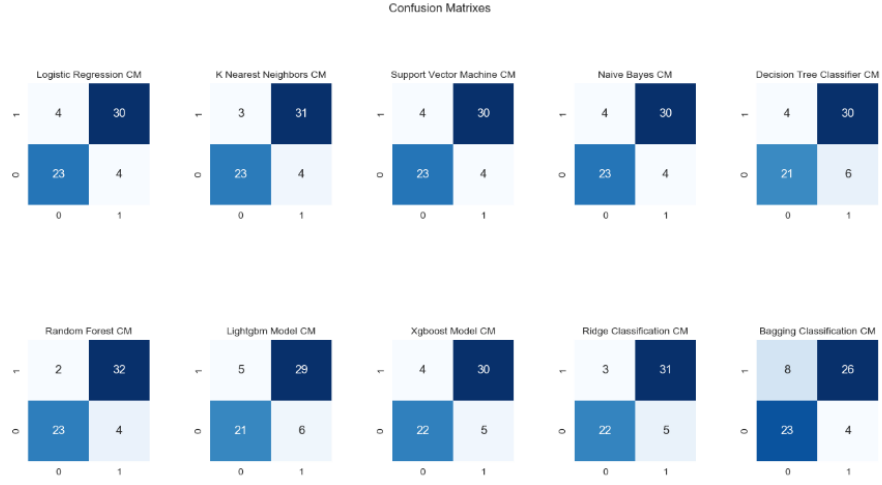
## 5. SONUÇLAR

Kalp hastalıklarının erken teşhis ve azımsanmayacak derecede yüksek bir başarı oranı ile tahmin edilmesi hastaların hayatını kurtarmakta büyük rol oynamaktadır. Bu çalışma 303 kişiden aralarında 165 kalp hastası bulunan 13 öznitelikle birlikte kalp hastalığının tahmin edilmesi konusunda yüksek başarı oranlarında tespit edilebildiği görülmüştür. Veri setini eğitebilmek için en uygun ve yüksek oranı veren algoritma rastgele orman algoritması olmuştur.

Bu konuda yapılan ve akademik olmayan çalışmalarda en yüksek oran %88,65 iken bu çalışmada bulunan oran %90,16 olarak bu algoritmanın kalp hastalıklarını tahmin etmek amacıyla şimdiye kadar kullanılan başarı oranı en yüksek uygulaması olduğu görülmüştür. Karmaşıklık matrisine bakıldığında rastgele orman algoritmasıyla eğitilen ve hatalı veriler olan yanlış pozitif ve yanlış negatif değerleri sırasıyla 4 ve 2 olarak saptanmıştır.

Ayrıca K en yakın komşu algoritması, 2 komşusu olan K en yakın komşu algoritması daha önce yapılan çalışmalarda %77,05 bulunurken bu çalışmada %85,25 olarak bulunmuştur. Ancak rastgele orman algoritmasının başarı oranı çok daha yüksek olduğu için bu algoritma üzerinde durulmuştur. Eğitime entegre edilen öznitelikler ile birlikte işlenen algoritmadaki en yüksek doğruluk oranı bulunmuş ve hastalığın tahmin edildiği gözlemlenmiştir.

Toplam verinin %20' sini oluşturan 61 test verisi her bir algoritma ile işlem görmüş ve Eş. 5.1' deki karışıklık matrisi ortaya çıkmıştır. Bu matrise göre LR için 61 veriden 8'i yanlış tahmin edilebilir, KNN için 7 veri, DVM için 8 veri, NB için 8 veri, Karar ağaçları için 10 veri, RO için 6 veri, LGBM Model için 11 veri, XGBoost için 8 veri, Bagging algoritması için ise 12 veri yanlış tahmin edilebilir şeklinde sonuçlanmıştır. Bu durumda üzerinde çalışılan veri setinde Bagging en fazla hatayı alırken RO en az hatayı alarak hastaneye gelen hastaların %90,16'si için hastalık tahmini doğru yapabilmektedir. Bu durumda ise sapma olarak hastaların %9,84'ünün hasta olma olasılığı yanlış tahmin edilebilmektedir.



**Şekil 5.1:** Karışıklık Matrisi

Sonuç olarak bu araştırma sonuçları kalp hastalıklarını teşhis etmede değil sadece hekimlere değerlendirmeleri ve ileri tetkikler uygulamaları için fikir verme amaçlı kullanıldığında kalp hastası olan insanların hastalığının hayati tehlike boyutuna gelmeden önlenebilir olması için hekimlere uygulanabilir bir kaynak sunmaktadır.



## KAYNAKLAR

- T.C. Sağlık Bakanlığı**, Türkiye Halk Sağlık Kurumu (2015-2020). Türkiye Kalp ve Damar Hastalıkları Önleme ve Kontrol Programı Eylem Planı.
- Altan ONAT, Ender ÖRNEK, Mustafa ŞENOCAK, Yavuz GÖZÜKARA, Günsel ŞURDUMAVCI, Yaşar KARAASLAN, Veyis TAŞKIN, Fehmi TABAK, Mehmet İŞLER, Ufuk ÖZİŞİK, Özden ÖZ, Remzi ÖZCAN.** (1991). "Survey on Prevalence of Cardiac Disease and its Risk Factors in Adults in Turkey: 6. Diabetes and Obesity". *Turk Kardiyol. Dern. Ars.* 1991; 19(3): 178-185
- İŞLER, Y, NARİN, A.** (2012). "WEKA Yazılımında k-Ortalama Algoritması Kullanılarak Konjestif Kalp Yetmezliği Hastalarının Teşhisi". *Teknik Bilimler Dergisi 2* (2012): 21-29
- Levitan, B.** (2004). "Is non-diabetic hyperglycaemia a risk factor for cardiovascular disease?" A metaanalysis of prospective studies. *Archives of Internal Medicine*, 2004, 164(19):2147–2155.
- Türkiye İstatistik Kurumu** (2018). Ölüm Nedeni İstatistikleri, 2018. (E.T. 14.10.2019)
- Cramer, J.S.** (2002). "The Orijins of Logistic Regression". Faculty of Economics and Econometrics, University of Amsterdam, and Tinbergen Institute, TI 2002-119/4
- Fix, E., Hodges, J.L.** (1989). "Contribution to Nonparametric Discriminant Analysis and Density Estimation". *School of Mathematical Sciences, University of Bath*, Vol. 57, No. 3 (Dec., 1989), pp. 233-238
- Vapnik, V., Cortes, C.** (1995). "Support Vector Networks". Kluwer Academic Publishers, Boston. Manufactured in The Netherlands, *Machine Learning*, 20, 273-297
- Bayes, T.** (1763). "An Essay towards solving a Problem in the Doctrine of Chances". By the late . Mr. Bayes, F.R.S communicated by Mr. Price, in a Letter to John Canton, M.F.R.S., Dec. 23, 1763
- William, A. Belson** (1959). "Matching and Prediction on the Principle of Biological Classification". *Journal of the Royal Statistical Society Series C*, Royal Statistical Society, vol. 8(2), pages 65-75, June 1959
- Breiman, L.** (2001). "RANDOM FORESTS". Statistics Department University of California Berkeley, CA 94720, January 2001

- Friedman, Jerome H.** (1999). "Greedy Function Approximation: A Gradient Boosting Machine". *Ann. Statist.* 29 (1999), no. 5, 1189--1232.
- Chen, T., Guestrin, C.** (2014). "XGBoost: A Scalable Tree Boosting System". University of Washington, March 27, 2014
- Tikhonov, A.N.** (1963). "Solution of Incorrectly Formulated Problems and The Regularization Method." *Soviet Math. Dokl.* Vol. 5., 1963
- Breiman, L.** (1994). "Bagging Predictors". Department of Statistics University of California Berkeley, California 94720, Technical Report No. 421, September 1994
- Shen, L., Chen, H., Yu, Z., Kang, W., Zhang, B., Li, H., Liu, D.** (2016). Evolving support vector machines using fruit fly optimization for medical data classification, *Knowledge-Based Systems*, 96, 61-75, 2016.
- Sevli, O.** (2019). "Göğüs Kanseri Teşhisinde Farklı Makine Öğrenmesi Tekniklerinin Performans Karşılaştırması" *Avrupa Bilim ve Teknoloji Dergisi* Sayı 16, S. 176-185, Ağustos 2019
- Baş, S. & Uzun, A.** (2018). "Tedarik Zincirinde Müşteri Siparişlerinin Lojistik Regresyon Analizi İle Değerlendirilmesi", *Ömer Halis Demir Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 11(3), 67-81, 2018
- Domingos, P., Pazzani, M.** (1997). On the Optimality of the Simple Bayesian Classifier under Zero-One Loss, *Machine Learning*, 29, 103-130, 1997
- AKMAN M, GENÇ Y, ANKARALI H** (2011). "Random Forests Yöntemi ve Sağlık Alanında Bir Uygulama." *Türkiye Klinikleri J Biostat.*, 2011
- Daoud, E.A.** (2019). "Comparison between XGBoost, LightGBM and CatBoost Using a Home Credit Dataset". *World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering* Vol:13, No:1, 2019
- Göktaş, A., Sevinç, V.** (2016). Two New Ridge Parametres And A Guide For Selecting An Appropriate Ridge Parameter In Linear Regression. *Gazi University Journal of Science.* 2016; 29(1): 201-211.
- Kayıkçıoğlu, M.** (2014). Homozygous familial hypercholesterolemia. *Türk Kardiyol Dern Ars* 2014;42 Suppl 2:19–31.
- Güleç, S.** (2009). "Kalp Damar Hastalıklarında Global Risk ve Hedefler" *Türk Kardiyol. Dern. Arş. - Arch Turk Soc Cardiol* 2009;37 Suppl 2:1-10
- Karlıkaya C., Öztuna, F. Solak Z.A., Özkan M., Örsel O.** (2006). "Tütün Kontrolü". *Toraks Dergisi* 2006; 7(1): 51-64
- US Department of Health and Human Services** (1982). A Report of the Surgeon General: The health consequences of smoking. Washington (DC), US Department of Health and Human Services. 1982.

**Okcu, Z., Keleş, F.** (2011). Kalp-Damar Hastalıkları ve Antioksidanlar. Atatürk Üniversitesi Ziraat Fakültesi Dergisi. 2011; 40(1): 153-160.

**KARABAY, O., KARABAY, M.** (2017). HIV/AIDS Hastalarında Kalp ve Damar Hastalıkları. OTSBD. 2017; 2(3): 31-46.

**Bridget, B. Kelly** (t.y). Institute of Medicine; Fuster, Valentin (2010). Promoting Cardiovascular Health in the Developing World: A Critical Challenge to Achieve Global Health. Washington, D.C: National Academies Press. ISBN 0-309-14774-3.

**McGill, H.C., McMahan, C.A., Gidding, S.S.** (March 2008). "Preventing heart disease in the 21st century: implications of the Pathobiological Determinants of Atherosclerosis in Youth (PDAY) study". Circulation 117

**EMİRZEOĞLU, L., NALBANT, S., ÇELİK, S., TEREKECİ, M., ÖKTENLİ, Ç.** (2019). Aort Anevrizmalı Hastalarda Anti-Endotel Hücre Antikor Düzeylerinin Sağlıklı ve Romatizmal Hastalığa Sahip Bireylerle Karşılaştırılması. MMJ. 2019; 6(2): 70-74.

**Sakalihan, N., Limet, R.** (2005). Defawe OD. Abdominal aorticaneurysm. Lancet. 2005;365(9470):1577-89.

**Bishop, C. M.** (2006), Pattern Recognition and Machine Learning

**Samuel, Arthur** (1959). "Some Studies in Machine Learning Using the Game of Checkers". IBM Journal of Research and Development. 3 (3): 210–229.

**ŞENER, S.** (t.y). Yapay Zekâ, Makine Öğrenimi ve Derin Öğrenme Arasındaki Farklar, Alındığı Tarih: 25.04.2019 <https://www.endustri40.com/yapay-zeka-makine-ogrenimi-ve-derin-ogrenme-arasindaki-farklar/>

**TOPAL, Ç.** (2017). Alan Turing'in Toplumbilimsel Düşünü: Toplumsal Bir Düşünce Olarak Yapay Zekâ, DTCF Dergisi 57.2 (2017): 1340-1364

## İNTERNET KAYNAKLARI

Tüik, (2019). "Türkiye Ölüm İstatistikleri" [www.tuik.gov.tr](http://www.tuik.gov.tr) (01.12.2019) <<http://www.tuik.gov.tr/UstMenu.do?metod=temelist> >

Who, (2020). "Kardiyovasküler Hastalıklar" [www.who.int](http://www.who.int) (03.11.2019) <[https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab\\_1](https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1)>

Uci, (1988). "Heart Disease Data Set" [www.archive.ics.uci.edu](http://www.archive.ics.uci.edu) (05.11.2019) <<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>>

TKD (2020). "Yüksek Kolesterol" [www.tkd.org.tr](http://www.tkd.org.tr) (08.11.2019) <<https://www.tkd.org.tr/kardiyobil/kalp-hastaliklari/yukse-kolesterol>>

## ÖZGEÇMİŞ

### Mertcan GÖRGÜN

**Adres** : idealtepe Mah. Yasa Sok. No:2 D:2 Maltepe/İSTANBUL

**Gsm** : 05423990357

**E-mail** : [mertcangorgun@gmail.com](mailto:mertcangorgun@gmail.com)

**DOĞUM YERİ** : İstanbul  
**DOĞUM TARİHİ** : 16.10.1993  
**MİLLİYET** : T.C.  
**MEDENİ DURUMU** : Evli

### PROFESYONEL DENEYİM

---

**03.2019-:** Samsung Electronics (Raporlama ve Analiz Uzmanı)

**2017 - 2018:** Samsung Electronics (Satış Eğitim Uzmanı)

**2013 - 2016** : İnditex(Bershka) (Mağaza Sorumlusu)

**2014 - 2015** : TÜİK Türkiye İstatistik Kurumu (Stajyer)

**2012 - 2013** : Bilge Adam (Programcı)

### EĞİTİM BİLGİLERİ

**İSTANBUL AYDIN ÜNİVERSİTESİ – BİLGİSAYAR MÜHENDİSLİĞİ (YÜKSEK LİSANS) / 2017 –**

Not Ortalaması - / 4

**İSTANBUL AYDIN ÜNİVERSİTESİ – İSTATİSTİK BÖLÜMÜ / 2013 – 2016**

Burs : %50 başarı burslu

Not Ortalaması : 2,9 / 4

**PLATO MYO – BİLGİSAYAR PROGRAMCILIĞI / 2011-2013**

Burs : %100 başarı burslu

Not Ortalaması : 3,6 / 4

**SULTANAHMET ATML – BİLGİSAYAR PROGRAMCILIĞI / 2007 – 2011**

Not Ortalaması : 4,3 / 5