

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

**BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI  
YÜKSEK LİSANS TEZİ**

**DİYABET HASTALIĞININ MAKİNE ÖĞRENMESİ ALGORİTMALARI  
İLE EN İYİ DOĞRU TAHMİNİNİN ELDE EDİLMESİ**

**GİZEM OĞUZTÜRK**

**HAZİRAN 2018**

**Bilgisayar Mühendisliği Anabilim Dalında** Gizem OĞUZTÜRK tarafından hazırlanan **DİYABET HASTALIĞININ MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE EN İYİ DOĞRU TAHMİNİNİN ELDE EDİLMESİ** adlı Yüksek Lisans Tezinin Anabilim Dalı standartlarına uygun olduğunu onaylarım.

Prof. Dr. Hasan ERBAY  
Anabilim Dalı Başkanı

Bu tezi okuduğumu ve tezin **Yüksek Lisans Tezi** olarak bütün gereklilikleri yerine getirdiğini onaylarım.

Prof. Dr. Hasan ERBAY  
Danışman

Jüri Üyeleri

Başkan : Dr. Öğr. Üyesi Hakan KOR \_\_\_\_\_  
Üye (Danışman) : Prof. Dr. Hasan ERBAY \_\_\_\_\_  
Üye : Dr. Öğr. Üyesi Bülent Gürsel EMİROĞLU \_\_\_\_\_

12/06/2018

Bu tez ile Kırıkkale Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu Yüksek Lisans derecesini onaylamıştır.

Prof. Dr. Mustafa YİĞİTOĞLU  
Fen Bilimleri Enstitüsü Müdürü

## ÖZET

### DIYABET HASTALIĞININ MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE EN İYİ DOĞRU TAHMİNİNİN ELDE EDİLMESİ

OĞUZTÜRK, Gizem

Kırıkkale Üniversitesi

Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı, Yüksek Lisans Tezi

Danışman: Prof. Dr. Hasan ERBAY

Haziran 2018, 88 sayfa

Diyabet oranı dünya genelinde hızla artmaktadır. Türkiye Diyabet Vakfı Başkanı'nın yaptığı açıklamaya göre; dünyada 450 milyon kişi diyabetle mücadele etmektedir. Türkiye'de ise bu rakam 10 milyon kişi üzerinde olup, diyabetli nüfus açısından bu rakam dünya ortalamasının yaklaşık 2 katı kadar bir değere tekabül etmektedir. Bunun yanı sıra, Türkiye Avrupa'da diyabetin en hızlı artış gösterdiği ülkedir. 2015 yılı verilerine göre, Türkiye'de her 12 kişiden biri diyabetlidir. Diyabetin erken tespiti sayesinde yaşam tarzı değişikliklerinin başlatılması ve uygun koruyucu önlemler alınması, diyabetin başlangıcını önlemeye veya ertelemeye yardımcı olmaktadır. Geçmişten günümüze kadar yapılan bilimsel çalışmalar sonucunda, pre-diyabet ve tip 2 diyabetin erken teşhis problemi olduğu kanıtlanmıştır. Bu bağlamda diyabet teşhisi için kolay, hızlı ve doğru tanı koyma araçlarına ihtiyaç duyulmaktadır. Erken diyabet teşhisi için makine öğrenimi algoritmalarına dayalı kolay, hızlı ve hassas bir tahmin aracı geliştirmek gerekmektedir. Bu çalışmada kullanılan veri seti, Türkiye'deki diyabetli ve diyabetik olmayan hastaların sağlık profillerinden oluşmaktadır. Hastaların 10 farklı özelliği giriş değişkeni olarak seçilmiş olup, sonuç değişkeni olarak da hasta olup olmadığına ait değerler kullanılmıştır. Deneklerin diyabetik durumun tahmin edilebilmesi için elde edilen veriler, 7 farklı makine öğrenmesi algoritması kullanılarak işleme tabi tutulmuştur. Toplam 2657 adet denekten 1860 adedi algoritmanın eğitimi için kullanılmış olup, kalan 797 veri adedi ise algoritmanın test edilmesi için ayrılmıştır. Diyabet tahmin modelinin

geliştirilebilmesi için açık kaynak kodlu Orange programı kullanılmıştır. Algoritmanın doğruluğunu optimize edebilmek için farklı kombinasyonlar, gizli düğüm sayısı ve beklenti maksimizasyonu (EM) iterasyonları uygulanmıştır. Yapay sinir ağı algoritmasının, %97.2'lik doğru tahmin başarısıyla en iyi başarıyı elde ettiği tespit edilmiştir. İkinci ve üçüncü en iyi başarı ise %96.9 ve %96'lık doğru tahmin başarı oranlarıyla lojistik regresyon ve random forest ile elde edilmiştir. Bu çalışma, yüksek doğrulukta diyabetik ve diyabetik olmayan bireyleri tanımlamak için son derece hassas bir makine öğrenme tahmin aracını tanımlamaktadır. Bu yöntem, hastanelerde veya diyabet önleme programlarında geniş çaplı tarama için kullanılabilir.

**Anahtar Kelimeler:** Diyabet Hastalığı, Veri Madenciliği,  
Makine Öğrenmesi Algoritmaları, Yapay Sinir Ağı,  
K-En Yakın Komşu Algoritması (k-NN),  
Lojistik Regresyon, Hastalık Tahmini

## ABSTRACT

### OBTAINING THE BEST PREDICTION WITH THE MACHINE LEARNING ALGORITHMS OF THE DIABETES

OĞUZTÜRK, Gizem

Kırıkkale University

Graduate School of Natural and Applied Sciences

Department of Computer Engineering, Master Thesis

Supervisor: Prof. Hasan ERBAY

June 2018, 88 pages

The rate of diabetes is increasing worldwide rapidly. According to the statement made by the President of Turkey Diabetes Foundation, 450 million people in the world are fighting against diabetes. This number is over 10 million people in Turkey and this number is about twice that of the world average for diabetes population. Furthermore, Turkey is the country that showed the most rapid increase of diabetes in Europe. According to the data 2015 year, one in every 12 people is diabetes in Turkey. Thanks to early detection of diabetes, various prevention are taken by people like a lifestyle change. In this way, the beginning of diabetes is prevented or delayed by doctors. As a result of scientific research, pre-diabetes and Type II diabetes have been proven to be an early detection problem. Therefore; simple, quick, correct diabetes diagnostic tools are required for the diagnosis of diabetes. Machine learning algorithms are used in prediction of diagnostic tools. The data set of this study have data of diabetic and non-diabetic patient. 10 different characteristics of patients are selected as input variables. Diabetic or non-diabetic patients data used as output variables. Data required to predict diabetic condition been tried to with 7 different learning algorithms. 1860 data are used for algorithm training from the total of 2657 data and the remaining 797 data is reserved for testing. Orange programming are used for the development of the diabetes prediction model. Different combinations, number of hidden nodes, expectation maximization iterations are used for optimize the accuracy of the algorithm. Artificial neural networks algorithm is the most

successful of all with successful prediction of 97.2%. Logistic regression and random forest are the second and third best success with successful prediction of %96.9 and %96. This study defines machine learning prediction tool for diabetic and non-diabetic patients. This method can be used for wide screening in hospital or diabetes prevention programs.

**Keywords:** Diabetes, Data Mining, Machine Learning Algorithms,  
Artificial Neural Network, K Nearest Neighborhood (K-NN),  
Logistic Regression, Diseased Estimation



## TEŐEKKÖR

Tez alıőmam boyunca beni motive eden, ynlendiren ve yardımlarını esirgemeyen danıőman hocam Sayın Prof. Dr. Hasan ERBAY'a teőekkr ederim. alıőmam sırasında akademik bilgilerini paylaőan ve tezime katkı saęlayan Dr. Öęr. Üyesi Hakan KÖR, Öęr. Gör. Ömer Faruk AKMEŐE ve Prof. Dr. őenay ARIKAN'a ise teőekkr bor bilirim. Son olarak da birok konuda olduęu gibi tezimi hazırlamam esnasında da yardımlarını, sabırlarını ve desteklerini benden esirgemeyen aileme teőekkr ederim.



# İÇİNDEKİLER DİZİNİ

Sayfa

<b>ÖZET</b> .....	<b>ii</b>
<b>ABSTRACT</b> .....	<b>iv</b>
<b>TEŞEKKÜR</b> .....	<b>vi</b>
<b>İÇİNDEKİLER DİZİNİ</b> .....	<b>vii</b>
<b>ŞEKİLLER DİZİNİ</b> .....	<b>x</b>
<b>ÇİZELGELER DİZİNİ</b> .....	<b>xi</b>
<b>KISALTMALAR DİZİNİ</b> .....	<b>xii</b>
<b>1. GİRİŞ</b> .....	<b>1</b>
<b>2. LİTERATÜR ÖZETİ: BENZER ÇALIŞMALAR</b> .....	<b>3</b>
<b>3. KURAMSAL TEMELLER</b> .....	<b>4</b>
3.1.Diyabetin Tanımı .....	4
3.2.Diyabetin Önemi .....	4
3.3.Diyabetin Sınıflandırılması .....	7
3.3.1.Tip I Diyabet.....	7
3.3.2.Tip II Diyabet .....	8
3.3.3.Gestasyonel Diyabet (Gestasyonel Diabetes Mellitus (GDM)) .....	9
3.3.4.Diğer Spesifik Tipler .....	10
3.4.Diyabetin Tanı Kriterleri.....	10
3.4.1.Diyabetin Klinik Bulguları .....	11
3.4.2.Diyabetin Laboratuvar Bulguları .....	11
3.5.Kronik (Dejeneratif) Komplikasyonlar .....	15
3.5.1.Makrovasküler Komplikasyonlar .....	15
3.5.2.Mikrovasküler Komplikasyonlar .....	16
3.6.Diyabetin Risk Faktörleri ve Korunması .....	18
<b>4. VERİ MADENCİLİĞİ</b> .....	<b>20</b>
4.1.Verit Madenciligi Tanımı ve Temel Kavramlar.....	20
4.2.Verit Madenciligi Süreci .....	22
4.3.Verit Madenciligi Modelleri .....	24
4.3.1.Tanımlayıcı (Descriptive) Modeller .....	25



4.3.1.1. Kümeleme (Clustering) Analizi .....	25
4.3.1.2. Özetleme (Summarization) Analizi.....	27
4.3.1.3. Birliktelik (Association Rules) Analizi.....	27
4.3.1.4. Sıra Örüntüleri (Sequence Discovery) Analizi .....	28
4.3.2.Tahmin Edici (Predictive) Modeller.....	29
4.3.2.1. Sınıflama (Classification) Analizi.....	29
4.3.2.1.1. Karar Ağacı (Decision Tree) Yöntemi .....	29
4.3.2.1.2. Bayes Sınıflandırması (Bayesian Classification) .	35
4.3.2.1.3. Yapay Sinir Ağları (Neural Networks) .....	36
4.3.2.2. Regresyon (Regression) Analizi .....	36
4.3.2.3. Zaman Serisi (Time Series) Analizi.....	37
4.3.2.4. Kestirim (Prediction).....	37
4.4. Veri Madenciliği Uygulama Alanları .....	38
4.5. Veri Madenciliği ile İlişkili Diğer Bilimler .....	38
4.6. Makine Öğrenmesi Algoritmaları .....	39
4.6.1. Lojistik Regresyon (Logistic Regression) Analizi .....	39
4.6.2. CN2 Rule Induction.....	41
4.6.3. Navie Bayes.....	41
4.6.4. C&R (Classification and Regression) Tree .....	42
4.6.5. Random Forest .....	42
4.6.6. K-En Yakın Komşu Algoritması (k-NN) .....	45
4.6.6.1. K-NN Parametreleri .....	45
4.6.7. Yapay Zeka (Artificial Intelligence) ve Yapay Sinir Ağları (Neural Networks).....	46
4.6.7.1. Yapay Sinir Ağlarının Yapısı.....	48
4.6.7.2. Yapay Sinir Ağları Çalışma Şekli.....	52
4.6.7.3. Öğrenme Algoritmalarına Göre Yapay Sinir Ağlarının Sınıflandırılması .....	53
4.6.7.4. YSA Çeşitleri .....	56
4.6.7.5. YSA'nın Tercih Edilmesinin Nedenleri.....	58
<b>5. YÖNTEM.....</b>	<b>59</b>
5.1. Bulgular.....	65
<b>6. SONUÇLAR .....</b>	<b>70</b>



## ŞEKİLLER DİZİNİ

<u>ŞEKİL</u>	<u>Sayfa</u>
4.1. CRISP-DM İşleyiş Modeli.....	22
4.2. Veri Madenciliği Metotları .....	25
4.3. Kümeleme Örneği .....	26
4.4. Karar Ağacı Yapısı.....	30
4.5. Oluşan Karar Ağacı Yapısı-I.....	34
4.6. Oluşan Karar Ağacı Yapısı-II .....	35
4.7. Veri Madenciliğinin Birçok Disiplinle Birleşimi.....	39
4.8. Random Forest Yönetime Göre Ağaç Yapısı.....	44
4.9. Doğrusal Aktivasyon Fonksiyonu.....	50
4.10. Sigmoid Aktivasyon Fonksiyonu.....	50
4.11. Tanjant Hiperbolik Aktivasyon Fonksiyonu.....	51
4.12. Yapay Nöronun Detaylı Yapısı.....	52
4.13. Öğrenme Yöntemleri.....	54
4.14. Danışmalı Öğrenme Yapısı.....	54
4.15. Danışmansız Öğrenme Modeli.....	55
4.16. İleri Beslemeli Çok Katmanlı Yapay Sinir Ağı .....	57
4.17. İleri Beslemeli ve Geri Beslemeli Ağ Yapıları .....	58
5.1.CRISP-DM İşleyiş Modeli.....	59
5. 2. Önerilen Sistemin Mimarisi .....	61
5. 3. Data Sampler Modülü .....	62
5.4. Artificial Neural Networks.....	66
5.5. Results of Artificial Neural Network Analysis .....	67
5.6. Test Score Modülü .....	68
5.7. Confusion Matris.....	69

## ÇİZELGELER DİZİNİ

<u>ÇİZELGE</u>	<u>Sayfa</u>
3.1. 2015 – 2040 yılları arasındaki IDF Diyabet Atlası Küresel Tahminleri.....	6
3.2. Tip I, Tip II Diyabet Arasındaki Benzerlik ve Farklılıklar .....	9
3.3. Diyabet ve Glikoz Metabolizmasının Diğer Bozukluklarda Konulmuş Tanı Kriterleri .....	12
3.4. Tip 2 Diyabet İçin Metabolik Kontrol Kriterleri.....	15
4.1. Yeni Sanayi Devrimi: Akıllı Üretim Sistemleri Teknoloji Yol Haritası [65].....	21
4.2. Örnek Veri Seti T .....	31
4.3. Sunny'nin Bilgi Kazancının Hesaplanması İçin Gerekli Olan Nitelikler .....	34
5.1. Dataset Description.....	63
5.2. Samples of Dataset.....	64
5.3. Veri Seti Kullanılan Modellerin Doğruluk Değerleri .....	65
6.1. Diyabet Komplikasyonları .....	71
6.2. Kullanılan Yöntemlerin Doğruluk Değerleri .....	72

## KISALTMALAR DİZİNİ

APG	Açlık Plazma Glukozu
BAG	Bozulmuş Açlık Glukozu
BGT	Bozulmuş Glikoz Toleransı
BKİ	Beden Kitle İndeksi
CART	Classification and Regression Tree
CRM	Customer Relationship Management
DM	Diabetes Mellitus
EM	Beklenti Maksimizasyonu
GDM	Gestasyonel Diabetes Mellitus
GFR	Glomerüler Filtrasyon Hızı
HbA1	Hemoglobin A1c Testi
HDL	High-Density Lipoprotein
IDF	Uluslararası Diyabet Federasyonu
k-NN	k-En Yakın Komşu Algoritması
LDL	Low-Density Lipoprotein
MIT	Massachusetts Institute of Technology University
OECD	Organisation for Economic Co-operation and Development
OGTT	Oral Glukoz Tolerans Testi
VLDL	Very Low-Density Lipoprotein
WHO	Dünya Sağlık Örgütü
YSA	Yapay Sinir Ağları

## 1. GİRİŞ

Diyabet, pankreas salgı bezinin yeterli düzeyde insülin hormonu üretmemesi veya ürettiği insülin hormonunun etkili olarak kullanılmaması durumunda oluşan, ömür boyu devam eden bir hastalıktır. Bunu sonucunda kişi, aldığı besinlerden kana geçen şekeri yani glikozu verimli bir şekilde kullanamaz ve kan şekeri yükselir [1]. Kan şekeri seviyesinin normal değerlerde kalması için düzenli olarak müdahale yapılmaması birçok soruna neden olabilir [2].

Dünya nüfusundaki artışla beraber, diyabetli hasta sayısı ciddi oranlarda artış göstermektedir. Diyabet artışının başlıca nedenleri; dengesiz beslenme, aşırı kilo, yaşlanma, ulaşımın kolaylaşması, hareketsizlik, bilgisayarların hayatın her alanına girmesi, internet, akıllı telefonlar, tablet kullanımı ve iş hayatının e-mail trafiğiyle oluşan sürekli strese altında olma durumu şeklinde sıralanabilir [3,4]. Dünyada 450 milyon kişi diyabetle yüz yüze kalmış durumda olup, Türkiye’de ise 10 milyonun üzerinde diyabet hastası bulunmaktadır. Türkiye diyabetli nüfusu açısından dünya ortalamasının hemen hemen 2 katı kadar olmasının yanı sıra, Avrupa’da da diyabetin en hızlı artış gösterdiği ülkedir. Yapılan bilimsel araştırmalar, Türkiye’de 2015 yılı verilerine göre her 6 kişiden birinin diyabetli olduğunu göstermektedir [4]. Bu şekilde geniş bir yelpazede görülen diyabet hastalığı tedavisi için büyük miktarda harcamalar yapılmaktadır. Ayrıca, diyabet hastalığı tedavi sürecinde de ciddi bakım ve iş gücü gerekmektedir. Öte yandan, günümüzde bilgisayarların hesaplama ve veri işleme yeteneğinin çok hızlı gelişmesiyle beraber veri madenciliği teknikleri tüm alanlarda tercih edilmektedir. Özellikle sağlık alanı hesaplama, teşhis ve tedavi sürecinde bilişim desteğini en çok kullanan alanlardan biridir. Veri madenciliği yöntemleri bu süreçte önemli katkılar sağlamaktadır.

Veri Madenciliği, güçlü veri işleme tekniklerini kullanarak büyük miktarda veriden bilgi örüntülerinin çıkarılmasına yönelik analiz sürecidir. Veri madenciliği, aynı zamanda, bilinmeyen önemli bilgileri keşfetmek için farklı alanlarda kullanılabilen bir disiplindir [5].

Bu alıřma, yksek doęrulukta diyabetik ve diyabetik olmayan bireyleri tanımlamak iin son derece hassas bir makine ęrenme tahmin aracını tanımlamaktadır. Bu yntem, hastanelerde veya diyabet nleme programlarında geniř aplı tarama iin kullanılabilir.



## 2. LİTERATÜR ÖZETİ: BENZER ÇALIŞMALAR

Literatürde diyabet hastalığı üzerinde etkili olan değişkenler veya diğer hastalıklarla olan ilişkisini inceleyen çalışmalar incelediğimizde, Kondiloğlu yüksek lisans çalışmasında [6], diyabet hastalarının kan şekeri tahminlerinin genellenebilir özelliğinin olduğu fakat başarı oranının yeterince yüksek olmadığı, hastaya özel tahminlerde ise başarı oranının daha yüksek olduğu ifade etmektedir. Başka bir çalışmada [7] kan şekeri kontrol altında olan ve olmayan hastaların psikiyatrik belirtilerinin kan şekeri kontrolü ile ilişkisi araştırılmıştır. 42 hasta üzerinde yapılan araştırmada kan şekeri kontrolü olan hastalardan 13'ünde ve kan şekeri kontrol altında olmayan hastalardan da 5'inde psikososyal stres tespit edilmiştir. Mohamed, "diyabet tahmini için sınıflandırma algoritmalarının karşılaştırılması" başlıklı yüksek lisans çalışmasında, sınıflandırma algoritmalarını kullanarak diyabet verilerini analiz etmektedir. Böylece, diyabet hastalığının erken teşhisinde ve diyabet hastalığına sebep olabilecek risk faktörlerini kontrol etmede faydalı olacağını ileri sürmektedir [1]. Yapılan bir diğer çalışmada da [8] besinlerin kan şekeri üzerindeki etkileri araştırılmıştır. Besinlerin tüketildikten sonraki kan şekerini artırıcı etkileri göz önüne alınarak pirinç yerine bulgurun tercih edilmesi gerektiği, beyaz ekmek yerine tam buğday ekmeğinin tercih edilmesi gerektiği, meyve suları yerine meyvelerin kendilerinin tüketilmesinin gerektiği gibi bilgiler verilmektedir. Pala ve Yücedağ, 1990 yılında, National Enstitute of Diabetes and Digestive and Kidney Diseases tarafından oluşturulan ve içerisinde 768 hastadan tip 2 diyabet hastasına ait önemli bilgilerin bulunduğu veri setini kullanmış olup, tedavi sürecinde ise doktorlara yardımcı olması amacıyla diyabet tahmin modeli geliştirmeyi hedeflemişlerdir. Yapılan çalışmada, C 5.0 karar ağacı algoritması uygulanarak 27 kural çıkarılmıştır [9].



### 3. KURAMSAL TEMELLER

Çalışmanın bu kısmında diyabet hastalığının genel tanımına ve diyabete etki eden kan ölçüm değerlerinin alt ve üst sınırlarına yer verilmiştir.

#### 3.1. Diyabetin Tanımı

Yunanca dia+betes ve mellitus kelimelerinden türeyen ve “tatlı idrar yapma” anlamına gelen Diabetes mellitus (DM), geçmişten günümüze devam etmektedir [10]. Bir tanıma göre; DM insülin salgılanması, taşınması ve depolanmasındaki defektler sebebiyle karbonhidrat, yağ ve protein metabolizmasındaki bozukluk sonucu oluşmakta olup, hiperglisemi ile karakterize kronik bir metabolizma hastalığıdır [11]. Ayrıca DM, pankreas insülin sekresyonunun mutlak ve rölatif yetersizliği, insülin yetersizliği, insülinin etkisizliği veya insülin moleküllerindeki yapısal bozukluklar sonucunda ortaya çıkmaktadır. Bir çeşit sendrom olan diyabetin etiyojisi, genetik ve klinik tablosu ile heterojen özelliktedir [12]. Başka bir tanıma göre de DM, pankreasın beta hücrelerinden salgılanan insülin hormonunun mutlak veya kısmi eksikliği veya periferik etkisizliği sonucu ortaya çıkmakta, kronik hiperglisemi, karbonhidrat, protein ve yağ metabolizmasında bozukluklar, kapiller membran değişiklikleri ve hızlanmış ateroskleroz ile seyreden, makrovasküler ve mikrovasküler komplikasyonların geliştiği, akut komplikasyonları önlemek ve kronik komplikasyonları azaltmak için sürekli destek, tıbbi bakım ve diyabetlinin öz bakım eğitimlerini gerektiren, sağlık bakım harcamaları içinde en az %10'luk yer tutan kronik seyirli endokrin ve metabolik bir hastalıktır [13-18].

#### 3.2. Diyabetin Önemi

Diyabet hayat boyu devam eden, insanlarda oldukça sık görülen ve tedavi maliyeti çok yüksek olan bir hastalıktır. Kalp-damar hastalıkları, diyabetli olanlarda en önemli ölüm nedeni olup, kalp krizi ve felç geçirme riski diyabetli olmayanlara göre

2 kat daha fazladır. Ayrıca, gelişmiş ülkelerdeki, böbrek yetmezliğinin ve buna bağlı olarak büyük oranlardaki diyaliz harcamalarının en önemli nedeni de diyabettir. Tüm diyabetlilerin %10-20'si böbrek yetmezliğinden dolayı hayatını kaybetmektedir. Yine gelişmiş ülkelerde erişkinlerdeki görme kaybının da en önemli nedeni diyabetik retinopati (diyabete bağlı göz hasarı)'dir [19].

Diyabet oldukça sinsi ilerleyen bir hastalık olması nedeniyle, hastalığın ortaya çıkışının belirlenmesinde oldukça güçlük çekilmektedir. Hastalığın ilk zamanlarda genel olarak semptomsuz olması nedeniyle gelişmiş ülkelerde ülkeler de dahi diyabetiklerin bilinmeyen diyabetlilere oranı 2/1' dir. WHO (Dünya Sağlık Örgütü)'nun tarafından yapılan çalışmalar neticesinde, yaklaşık 171 milyon olan diyabetli sayısının önümüzdeki on yılın sonunda 366 milyona ulaşması kaçınılmazdır [20-21]. Dünyanın her yerinde teşhisi koyulan diyabet hastalarının % 90-95'ni yetişkin ve şişman kişilerde görülen Tip II DM, % 5-10'nu da çocuklarda ve gençlerde görülen Tip I DM ve % 2-3'ünü ise diğer diyabet formları oluşturmaktadır. Gelişmiş ülkelerde toplumun yaklaşık %5-10'unda Tip II diyabet görülmektedir. Diyabetlilerin %30-50'sinin ise henüz teşhis konulmamış vakalar olduğu düşünülmektedir [22].

2040 yılında, Uluslararası Diyabet Federasyonu (IDF) tahminlerine göre,10 yetiştikten 1'i (642 milyon) diyabet hastası olacaktır. Diyabet ile alakalı hastalıklara yapılan sağlık harcamaları ise 802 milyar ABD Dolarını aşacaktır. Tablo 3.1'de, 2015-2040 yılları arasındaki IDF Diyabet Atlası Küresel Tahminleri gösterilmektedir [23].

**Çizelge 3.1.** 2015 – 2040 yılları arasındaki IDF Diyabet Atlası Küresel Tahminleri

	<b>2015</b>	<b>2040</b>
Toplam Dünya Nüfusu	7.3 milyar	9.0 milyar
Yetişkin Nüfus (20-79 aralığı)	4.72 milyar	6.16 milyar
Çocuk Nüfusu (0-14 aralığı)	1.92 milyar	-
<b>Şeker Hastalığı (20-79 aralığı)</b>		
Küresel Prevelans	%8.8 (%7.2-11.4)	%10.4 (%8.5-13.5)
Diyabetli Birey Sayısı	415 milyon (340-536 milyon arası)	642 milyon (521-829 milyon arası)
Diyabete Bağlı Hayatını Kaybedenlerin Sayısı	5.0 milyon	-
<b>Diyabete Bağlı Sağlık Harcamaları (20-79 arası)</b>		
Toplam Sağlık Harcamaları, R=2* 2015 ABD Doları	673 milyar	802 milyar
<b>Gebelikte Hiperglisemi (20-49 aralığı)</b>		
Etkilenen Canlı Doğumların Oranı	%16.2	
Etkilenen Canlı Doğumların Sayısı	20.9 milyon	
<b>Bozulmuş Glukoz Toleransı (20-79 arası)</b>		
Küresel Prevelans	%6.7	%7.8
Bozulmuş Glukoz Toleransı Olan Kişi Sayısı	318 milyon	481 milyon
<b>Tip I Diyabetliler (0-14 aralığı)</b>		
Tip I Diyabetli Çocuk Sayısı	542000	
Her Yıl Yeni Teşhis Konulanların Sayısı	86000	

### 3.3. Diyabetin Sınıflandırılması

Diyabet, klinik olarak 4 grupta incelenmekte olup sırasıyla;

1. Tip I Diyabet,
2. Tip II Diyabet,
3. Gestasyonel Diyabet (Gestasyonel Diabetes Mellitus (GDM)),
4. Diğer spesifik tipler olan Sekonder (başka hastalık nedeniyle, o hastalığa bağlı oluşan) diyabet'dir [24].

#### 3.3.1. Tip I Diyabet

Pankreasta insülin üretiminden sorumlu olan Beta hücrelerinin tahribatı sonucu, mutlak insülin yetersizliği ile meydana gelen, insüline bağımlı diyabet tipidir. Diğer bir ifadeyle, açıklaması olmayan bir şekilde vücut, kendisine ait bir hücreyi, organı, ya da bir hormonu yabancı olarak algılayıp, onun yok edilmesi için iltihabi bir reaksiyon başlatmaktadır. Bu reaksiyon sonucunda insülin üretiminden sorumlu beta hücrelerinin sayısında ve buna bağlı olarak da insülin salgısında azalma meydana gelmektedir. Böylece Tip I diyabet ortaya çıkmaktadır [24,25]. Ömür boyu insülini dışarıdan almak zorunda olduğu için *İnsüline Bağımlı Diyabet* olarak da adlandırılmaktadır [1].

Tip I diyabet, genellikle 35 yaş altı olup, en sık görüldüğü ise 10-15 yaşları arasındadır. Tip I diyabetliler tüm diyabetlilerin yaklaşık olarak %5-10'unu oluşturmaktadır. Bazı virüs enfeksiyonları (kabakulak, konjenital rubella gibi), stres, toksinler ve beslenme şekilleri hastalığın oluşmasını hızlandıran faktörlerdir [15-17, 26].

Tip I diyabetin tedavisinin değişmez kuralı ise insülin enjeksiyonudur. Bunun yanı sıra hastalığın bilincinde olmak, egzersiz yapmak ve sağlıklı beslenmek tedavinin diğer önemli noktalarındandır. Kişinin bu belirtilenleri hayat biçimi haline getirmesi gerekmektedir [1].

### 3.3.2. Tip II Diyabet

Dünyada en sık rastlanan ve en baskın diyabet tipi olup, tüm diyabet hastalarının yaklaşık %90'ında Tip II diyabet görülmektedir [24,27]. Pankreasın ürettiği insülinde meydana gelen yetersizlik sonucu, vücudun bunu gerektiği gibi kullanamaması Tip II diyabetin oluşmasına neden olmaktadır. İnsülin üretiminin yetersiz oluşundan ziyade asıl neden ise üretilen insülinin gerektiği gibi kullanılamamasıdır [28].

Tip II diyabet genellikle erişkin tipi diyabet olarak kabul edilerek 40 yaş ve üstü hastalarda görülmekte olup, son yıllarda ise yaşam tarzının getirdiği etkenlere bağlı olarak genç yaşlarda hatta çocuklarda da sıklıkla rastlanmaktadır. Tip II diyabetin oluşmasını hızlandıran faktörler arasında stres, sedanter yaşam (hareketsiz yaşam), düzensiz ve dengesiz beslenme, aileden genetik olarak gelme, kilo fazlası, egzersiz azlığı yer almaktadır [15-17, 29, 30].

Tip II diyabet hastalarının %80'den fazlası ise obezitedir [15-17, 29, 30]. Norveç'te yapılan prospektif bir çalışmanın, on yıllık izlem sonuçlarına göre zayıf ya da normal vücut ağırlığında olan insanlarda Tip II diyabet görülme sıklığı %0 iken, obezlerde bu olasılık % 13'e kadar ulaşmaktadır [18].

Tip II diyabetin tedavisi için, düzenli egzersiz yapılması, beslenme alışkanlıklarının düzenlenmesi, yaşam tarzının değiştirilmesi gerekir eğer bunlar yapılmasına rağmen kan şekeri normal değerlere ulaşmazsa şeker düşürücü ilaçlar ya da uygun dozda insülin enjeksiyonları tedaviye eklenmelidir [31]. Tip I, Tip II diyabetin benzerlik ve farklılıkları aşağıda yer alan Tablo 3.2'de belirtilmektedir [32].

**Çizelge 3.2.** Tip I, Tip II Diyabet Arasındaki Benzerlik ve Farklılıklar

<b>Tip I</b>	<b>Tip II</b>
Başlangıç yaşı 30 yaşının altındadır.	Başlangıç yaşı 35-40 yaş ve üzeridir.
Ailede diyabet öyküsü mevcut olup yaygın değildir.	Ailede diyabet öyküsü yaygındır.
Daha çok sonbahar-kış aylarında görülmeye başlanmıştır.	Mevsimlerle ilişkisi yoktur.
Kilo kaybı olduğu için genellikle kişiler zayıftır.	Kişiler genellikle obezdir.
Belirtileri ani başlar hızlı ilerler.	Yavaş başlayıp, yavaş ilerler.
Dört kardinal bulgu vardır.	Belirtilerin derecesi ağırdır.
Asemptomatik olabilir.	Belirtilerin derecesi hafiftir.
İdrarda glukoz ve aseton vardır.	İdrarda glukoz vardır.
Ketoasidoza eğilim fazladır.	Ketoasidoza eğilim azdır.
Otoimmün kökenlidir.	Otoimmün kökenli değildir.
Serum insülin düzeyi düşüktür ya da yoktur.	Genellikle insülin düzeyi düşüktür fakat yükselebilir.
Tedavide insülin kullanımı şarttır.	Genellikle diyet ve OAD ile kontrol edilebilmektedir.
Teşhisi kolay konulmaktadır.	Tam koymak daha zordur.
İnsüline direnç vardır.	İnsüline duyarlıdır.
Kronik komplikasyonlar görülebilmektedir.	Kronik komplikasyonlar görülebilmektedir.

### **3.3.3. Gestasyonel Diyabet (Gestasyonel Diabetes Mellitus (GDM))**

Gestasyonel diyabet, ilk kez gebelikte ortaya çıkan, değişik derecelerde glikoz tolerans bozukluğudur. Gestasyonel diyabet zamanla ince bağırsak hormonlarında artış, anormal iştah, kan lipidlerinde değişim, karaciğer metabolizmasında bozukluk, kas ve adipoz doku metabolizmalarında farklılaşma, sinir sisteminde, kan

damarlarında, gözlerde ve bazen de böbreklerde hasara neden olmaktadır. Bu tip diyabette, hamilelik sonrasında ise kan basıncı normale dönmektedir [33].

Uzun dönemde, Gestasyonel diyabetli kadınların Tip II diyabet olma riski oldukça yüksek olup, 10-20 yıl içerisinde Tip II diyabetin daha da gelişme riski %30-40 arasında kalmaktadır [16,34].

### **3.3.4. Diğer Spesifik Tipler**

Diğer spesifik tipler, pankreası etkileyen birçok nedenle ortaya çıkan kan şekeri yüksekliğini tanımlar. İnsülin fonksiyonunda genetik bozukluk, endokrinopatiler, enfeksiyonlar, beta hücre fonksiyonundaki genetik bozukluk, pankreasın ekzokrin hastalıkları, ilaç ve kimyasal ajanlar, diyabetle ilişkili diğer genetik sendromları bu grup içerisine dâhil edilebilmektedir [16].

### **3.4. Diyabetin Tanı Kriterleri**

Şeker hastalığı (diyabet), pankreasın yeterli miktarda insülin üretememesi ya da üretilen insülinin, vücut tarafından etkili bir şekilde kullanılamaması sonucunda ortaya çıkan bir hastalıktır. İnsülin sayesinde şeker, hücrenin içerisine glikojen olarak depolanmaktadır. Bu olay diyabet hastasında farklı olmakta ve kana geçen şekerin yani glikozun etkili şekilde kullanılamaması sonucu kan şekeri seviyesi yükselmektedir (Hperglisemi). Bu durumda, birçok doku ve organda iyileştirilmesi mümkün olmayan hasarlara neden olmaktadır. Hayat boyu süren ve vücutta birçok organın hasar görmesine neden olan diyabet tanısı doğru yapılan çeşitli laboratuvar testleri ve değerlendirmelerden sonra erkenden konulabilir ancak, bir takım yanlış inanışlar (Örn; Şekerli besinler tüketmediğim için diyabet hastası olma ihtimalim yok vs.) bu tanının gecikmesine hatta yıllar boyunca konmamasına neden olup, hastalığın seyrini olumsuz yönde değiştirebilmektedir [35]. Ayrıca diyabetin tanı kriterleri, ilerleyen bölümde klinik ve laboratuvar bulguları olarak iki farklı şekilde incelenmiştir.

### 3.4.1. Diyabetin Klinik Bulguları

Diyabet hastalığının klinik bulguları;

- Ağız kuruluğu ve buna bağlı çok su içmek (polidipsi),
- Sık idrara çıkmak (poliüri),
- Gece çok idrara kalkmak (noktüri),
- Çok acıkmak (polifaji),
- Halsizlik,
- İştahsızlık,
- Açıklanamayan kilo kaybı,
- Baş ağrısı,
- Bulanık görme,
- Kaşıntı,
- Cilt yaraları veya kesiklerinin yavaş iyileşmesi olarak belirtilmektedir [35].

### 3.4.2. Diyabetin Laboratuvar Bulguları

Diyabet hastalığının laboratuvar bulguları;

**Glukoz;** Karbonhidratlı besinlerden elde edilen vücudun temel enerji kaynağıdır. Halk arasında “kan şekeri testi” olarak bilinen glukoz testi, bir tür şeker olan glukozun kandaki seviyesini ölçüp, kan dolaşımındaki şeker miktarını tespit etmek için yapılmaktadır. Açlık kan şekeri ölçülmesi, kanın glukoz miktarını belirleyen mekanizmalar hakkında bilgi vermektedir [35-37]. Aynı zamanda açlık kan şekerinin fazla çıkması diyabet tanısı konulmasında en belirli olanıdır. Normal kişiler aç kalınca insülin dışındaki hormonlar kan şekerini arttırıp insülin ise aşağıya çekilirken, diyabet hastası olanlarda insülin hormonu iyi çalışmadığından kan şekeri yüksek çıkar. Açlık kan şekeri (glukoz) testi 8-12 saat açlıktan sonra ölçülüp eğer;

- 110 mg/dL çıkarsa normal,
- 110-120 mg/dL arası çıkarsa bozulmuş açlık glukozu,



- 126 mg/dL üzerinde çıkarsa da geçici diyabet tanısı konulmaktadır [33, 35-38].

Ayrıca, vücudun glukozu kullanma ya da kandan atma durumunu gösteren aynı zamanda “şeker yüklemesi testi” olarak da bilenen “Oral Glukoz Tolerans Testi (OGTT)” bulunmaktadır. Bu testteki amaç; diyabet (şeker hastalığı) riskini araştırıp, pre-diyabet (gizli şeker) ya da gebeliğe bağlı şeker hastalığı varsa tanısını koymaktır. 10-16 saat aç kalan hastaya açlık kan örneği alındıktan 5 dakika sonra 250-300 mL suda 75 gr glukoz içirilmeli ve bu kan örnekleri, yüklemeden iki saat sonra alınmalıdır. Alınan kan örneklerinde iki saat sonra venöz kanda glukoz düzeyinin 200 mg/dL’den fazla olması diyabet tanısı konması için yeterlidir. Tokluk kan şekeri testi ise yemeklerden iki saat sonra yapıp, ölçülen kan şekeri değerinin 140mg/dL’den yüksek olması durumunda diyabet tanısı konulmaktadır [33, 35-38]. Çizelge 3.3’de diyabet ve glukoz metabolizmasının diğer bozukluklarda konulmuş tanı kriterleri gösterilmektedir [22].

**Çizelge 3.3.** Diyabet ve Glikoz Metabolizmasının Diğer Bozukluklarda Konulmuş Tanı Kriterleri

<b>Diabetes Mellitus</b>	
Rasgele Diabet (+ diabet semptomları)	≥200
AKG (en az 8 saatlik açlığı takiben)	≥126
OGTT'de 2. Saat KG	≥200
<b>Bozulmuş Glukoz Toleransı (IGT)</b>	
OGTT'de 2. Saat	140-199
<b>Bozulmuş Açlık Glikozu (IFG)</b>	
AK (en az 8 saatlik açlığı takiben)	100-125

**HbA1c;** HbA1c testi, glukozun kırmızı kan hücrelerinin içinde bulunan hemoglobinin isimli proteine bağlanması sonucunda oluşup, şeker hastalığını kontrol altında tutmak ve şeker hastası olup olmadığını belirlemek için yapılan önemli bir kan testidir. HbA1c, son üç aylık ya da 100 günlük kan şekerinin yüksekliği hakkında bilgi

vermektedir. Eđer son üç ayda, ortalama kan řekeri yüksek ise HbA1c yüksek çıkar. Hastaların HbA1c değeri;

- % 6.5'den küçük ise iyi kontrol değerde,
- % 6.5-7.5 arasında ise sınırdaki kontrol değerde,
- % 7.5'den büyük ise kötü kontrol değindedir [35, 39].

**Total, HDL ve LDL Kolesterol;** Kolesterol, karaciđer tarafından üretilen ve tüm hücrelerimizde bulunan yağ benzeri bir maddedir. Ortaya çıkmasının en önemli nedeni ise beslenme şekli olup et ve süt gibi hayvansal besinlerin ağırlıklı olarak tüketilmesidir. Kolesterol, kanda kolesterol taşıyıcı proteinlere bağlanmış olarak taşınıp, kendi başına bulunmaz. Taşıyıcı protein ve kolesterolün birleşimi lipoprotein olarak adlandırılmakta olup, üç çeşit lipoprotein, bulunmaktadır:

- LDL: Low-Density Lipoprotein (kötü kolesterol)
- HDL: High-Density Lipoprotein (iyi kolesterol)
- VLDL: Very Low-Density Lipoprotein (trigliserit taşıyan lipoprotein)

Genellikle iyi kolesterol olarak bilinen yüksek yoğunluklu lipoprotein olan HDL, kötü kolesterolü çekerek, damarların içerisinde birikmesini önlemeye yardımcı olmaktadır. Düşük yoğunluklu lipoprotein olan LDL kolesterol ise “kötü” kolesterol olarak adlandırılır ve damar çeperinde kolesterol birikimine yol açarak damar tıkanıklığına neden olur.

HDL kolesterolün fazla olması kolesterolü, karaciđer tarafından emilen yer olan damar duvarlarından uzaklaştırdığı için tercih edilir. Ergenlik çağına kadar HDL kolesterol seviyeleri bakımından, çocuklar arasında cinsiyet farklılığı görülmemekte olup, ergenlikten sonra kadınların HDL kolesterol seviyeleri erkeklerden daha yüksektir [33]. HDL kolesterolünün 35 mg/dL altında olması pozitif risk faktörü, 60 mg/dL'nin üzerinde olması ise negatif risk faktörü olarak tanımlanmaktadır [38, 40].

Beslenme alışkanlıkları ya da genetik dolayı LDL düzeyinin yüksek olması ise damarlarda, sert ve yağlı bir kolesterol birikimine neden olacağından, damar çeperini

daraltarak tıkeyip, diyabet hastalarında kalp krizi riskini 4'e katlamaktadır. Bu nedenle diyabet ve tansiyon kontrol altında tutulmalı ve LDL seviyesi 130 mg/dL'nin altına çekilmelidir. Bu şekilde hayati riskler en aza indirilebilir.

**Trigliserid;** Trigliserid vücudumuzda besin ve enerjinin depo şeklidir. Bu maddeler vücuda alınan ancak yakılamayan besinlerin fazlarından, organların etrafında ve deri altında biriktirilerek oluşturulmaktadır. Oluşma yerleri ise karaciğer gibi metabolizma organlarıdır. Kısaca Trigliserid bağırsaktan emilen sindirilmiş besin maddelerinin yağlaşması sonucu ortaya çıkmaktadır.

Egzersiz yapmamak, alkol ve sigara tüketimi, düzensiz beslenmek Trigliseridin daha da yükselmesine neden olmaktadır. Trigliserid konsantrasyonu yaşa ve cinsine bağlı olarak değişiklik gösterse bile, vücutta olması gereken normal Trigliserid değeri 150 mg/dL altında olmalıdır. 200 mg/dL'ye kadar olanlar hafif yüksek, bunun üzerindeki ise yüksek risk sınıfına girmektedir. Serum Trigliserid konsantrasyonunun yükselmesi de aterosklerozis (arterlerde arteriosklerotik plaklar ve trombozlarla karakterize ileri yaş hastalığı) oluşumunda rol oynamaktadır [38, 41].

**Kan Basıncı Düzeyi;** Kan basıncı düzeyi, diyabetik hastalarda kardiyovasküler morbidite ve mortaliteyi belirleyen en önemli faktörlerden biridir. Diyabetik hastalarda kardiyovasküler hastalıklardan yıllık ölüm riski 2-3 kat daha fazla olduğun için diyabet; hipertansiyon, şişmanlık, hiperlipidemi gibi ateroskleroz gelişmesi için risk faktördür. Diyabet hastalarında insülin yetersizliği kan basıncının yükselmesinde önemli bir risktir [38, 42-43].

Kan basıncı kontrolünün sağlanabilmesi için birçok hastada, yaşam tarzı değişikliğinin yanı sıra en az iki üç antihipertansif ilacın yer aldığı farmakolojik kombinasyon tedavisi gerekir [44].

Genel olarak örnek metabolik kontrol kriterleri, Tip II diyabet için Çizelge 3.4'de gösterilmektedir [33, 38, 45].

**Çizelge 3.4.** Tip 2 Diyabet İçin Metabolik Kontrol Kriterleri

	<b>Normal</b>	<b>Kabul Edilebilir</b>	<b>Kötü</b>
Açlık Plazma Glukozu (mg/dL)	115	115-139	≥140
Postprveiyal Plazma Glukozu (mg/dL)	<140	140-199	≥200
HbA1c (%)	<6.5	<7.5	≥7.5
Total Kolestrol (mg/dL)	<200	200-239	≥240
LDL Kolestrol (mg/dL)	<130	130-159	≥160
HDL Kolestrol (mg/dL)	>45	>35	≤35
Trigliserit (mg/dL)	<200	200-250	>250
Kan Basıncı (mmHG)	≤140/90	160/95	>160/95

### **3.5. Kronik (Dejeneratif) Komplikasyonlar**

Kan şekerinin uzun süre yüksek olması, büyük ve küçük damarlarda aynı zamanda sinirlerde hasarlara yol açmaktadır. Tahribatın farklı organ ve sistemlerde oluşturduğu değişiklikler, kronik komplikasyonlarını meydana getirmektedir. Bu komplikasyonlar arasında mikrovasküler (retinopati ve nefropati), nöropatik değişimler ve makrovasküler hastalık yer almaktadır. Erken tanı ve tedavi, komplikasyonları engelleyeceğinden, erken tanı yöntemlerine başvurularak, diyabet tedavisinde kronik komplikasyonların önlenmesi başlıca hedeflerden biri haline gelmiştir [46, 47].

#### **3.5.1. Makrovasküler Komplikasyonlar**

Makrovasküler komplikasyonları, Tip II diyabette diyabetin var olduğu ancak henüz dışarı çıkmadığı bozulmuş glukoz toleransı döneminde başlamaktadır. Genel olarak, diyabete özgü faktörler, yaşam şekli ile ilgili faktörler ve genetik faktörler olmak üzere sıralanmaktadır [48].

### 3.5.2. Mikrovasküler Komplikasyonlar

Mikrovasküler Komplikasyonlar genel olarak aşağıdaki şekilde sıralanmaktadır [49];

- Diyabetik Retinopati
- Diyabetik Noropati
- Diyabetik Nefropati

**Diyabetik Nefropati;** Diyabet hastalarında hasar gören organlardan biriside böbreklerdir. Böbrekler kandaki zararlı maddeleri süzerek bunların idrar ile atılmasını sağlamaktadır. Normal bir kişinin idrarında protein bulunmaz. Eğer böbreklerde hasar oluşursa ilk önce dışarı idrarla protein atılmaya başlar ve bu atılan protein *ALBUMİN* şeklindedir. *İDRARDA ALBUMİN* bulunması böbreklerin normal fonksiyonunu yapamadığının göstergesidir. İdrarda Albumin'in değerleri;

- 30 mg'a kadar normaldir.
- 30-300 mg/24 saat arasındaki sonuçlar "*mikroalbuminüri*" adını alır ve erken dönem böbrek hastalığını gösterir.
- 300 mg/24 saat üzerindeki düzeyler "*makroalbuminüri*" olarak isimlendirilir ve ilerlemiş böbrek hastalığının göstergesidir.

Mikrovasküler komplikasyonları grubunun içinde en ciddi olanı "*Diyabetik Nefropatidir*". Diyabetik Nefropati aynı zamanda kronik böbrek yetmezliğinin en önemli nedenidir. Diyabetli hastaların %10-20'si böbrek yetersizliği nedeniyle kaybedilmektedir. Tip I diyabetiklerin %30-40'ında, Tip II diyabetiklerin %5-10'unda son dönem böbrek yetmezliği gelişmektedir. Tip II diyabet, Tip I diyabetten çok daha yaygın olduğu için Tip II diyabete bağlı nefropati prevalansı daha fazladır [50].

Diyabetik nefropatinin gelişmesinde hipertansiyon, hiperglisemi, sigara, insülin direnci, ileri yaş, yüksek proteinli beslenme ve genetik risk faktörleri etkili olup, diyabetin süresi uzadıkça, diyabetik nefropatinin sıklığı da giderek artmaktadır [51-54]. Günümüzde böbreklerdeki süzme potansiyelinin hangi evrede olduğunu ve şeker

hastalığının böbreklere vermiş olduğu zararın ne olduğunu en iyi gösteren testlerden birisi de *GFR* ya da *eGFR* olarak bilinen *Glomerüler Filtrasyon Hızı* testidir. *GFR* değerlerine göre böbrek hastalığı evreleri şu şekildedir:

- **Evre I:**  $GFR \geq 90$  mL/dk/1.73 m<sup>2</sup> (vücut yüzey alanı için) ise normal/yüksek *GFR* ile birlikte böbrek hasarı,
- **Evre II:**  $GFR$  60-89 mL/dk/1.73 m<sup>2</sup> ise hafif derecede azalmış *GFR* ve böbrek hasarı,
- **Evre III:**  $GFR$  30-59 mL/dk/1.73 m<sup>2</sup> ise orta derecede azalmış *GFR* ve böbrek hasarı,
- **Evre IV:**  $GFR$  15-29 mL/dk/1.73 m<sup>2</sup> ise ileri derecede azalmış *GFR* ve böbrek hasarı,
- **Evre V:**  $GFR < 15$  mL/dk/1.73 m<sup>2</sup> veya diyaliz uygulanıyorsa son dönem böbrek yetersizliği demektir [49,52].

*GFR*'nin hesaplanması hastanın profiline (yaş, boy ölçüsü, etnik köken kilo, cinsiyet) ve kandaki kreatinin seviyesine bağlı olarak değişmektedir. Böbrek hasarı ilerleyip, böbrek artık iyi çalışmaz duruma gelince kandaki *KREATİN* ve *ÜRE* seviyeleri de yükselmeye başlamaktadır. Çünkü bu maddelerin atılımı böbrekten olmaktadır. Kas hastalığı, şeker hastalığı, böbrek hastalığı, yüksek tansiyon, tiroid hastalığı, yorucu egzersiz yapma, uzun süre protein oranı yüksek gıdalar ile beslenmek gibi birçok faktör kreatinin seviyesinin artmasına neden olmaktadır. Kreatinin oranı 1,5 mg/dL'nin üzerine çıkarsa böbrekler artık sağlıklı çalışmıyor demektir.

Aşırı susuzluk, fazla protein tüketimi, böbrek hastalığı, kalp yetmezliği, sindirim sistemi sorunları, aşırı kas yıkımı üre yüksekliğine neden olmaktadır. 5 mg/mL (1.79 mmol/L) değerinden küçük olan veya 50 mg/dL (17.85 mmol/L) değerinden yüksek olan üre değeri, bir hastalık belirtisi olabilmektedir. Kırıkkale Üniversitesi Tıp Fakültesi'nden alınan değerlere göre kanda normal değerler üre için, 17-43 mg/dL, kreatinin içinse 0,50-0,90 mg/dL'dir. Bu değerler laboratuvarlara göre küçük farklılıklar gösterebilir. Ayrıca, diyabetik böbrek hastalarının sodyum, potasyum alımlarına da dikkat etmeleri gerekmektedir.

*SODYUM*, vücudumuzdaki suyun dengesine katkıda bulunmaktadır. Kan basıncının belirli düzeylerde tutulmasını sağlar ve elektriksel özellikleri ile değişik organların görev yapmasında hayati rol oynar. Böbrek hastalıklarında yeterli miktarda atılım gerçekleşmediğinde vücudumuzda birikir. Biriken sodyum, suyun da birikmesine yol açıp; tansiyonu yükseltir ve kalp yetersizliğine sebep olabilir. O nedenle, böbrek hastalarında çoğu kez tuz kısıtlaması yapmak gerekli olur. Kırıkkale Üniversitesi Tıp Fakültesi'nden alınan laboratuvar sonuçlarına göre kandaki normal sodyum aralığı 136-146 mmol/L'dir.

*POTASYUM*, vücudumuzda bulunan özel bir tuzdur. Kaslarımızın kasılmasında çok önemli rolü vardır. Potasyum böbrekler ile atıldığından, böbrek yetersizliğinde kandaki seviyesi yükselir ve bu durumda halsizlik, iştahsızlık, adale kuvvetsizliği ve felçler görülür. Potasyumun daha fazla yükselmesi halinde ise kalp aniden durabilir. Kırıkkale Üniversitesi Tıp Fakültesi'nden alınan laboratuvar sonuçlarına göre kandaki normal potasyum aralığı 3.5-5.5 mmol/L'dir.

### **3.6. Diyabetin Risk Faktörleri ve Korunması**

Diyabetin risk faktörleri belirlenemeyip, hastalığın ilerlemesi önlenemediğinde hasta sayısı artmakta olup beraberinde de birçok hastalığı getirmektedir. Diyabetin risk faktörleri; değiştirilemez ve değiştirilebilir/kontrol edilebilir olarak sınıflandırılabilir. Değiştirilemez risk faktörleri; genetik faktörler, yaş ve cinsiyettir. Değiştirilebilir/kontrol edilebilir risk faktörleri ise BAG (Bozulmuş Açlık Glukozu), BGT (Bozulmuş Glikoz Toleransı), dislipidemi gibi metabolik bozukluklar, gestasyonel diyabet, obezite, sigara, hipertansiyon, sağlıksız beslenme ve aktivite azlığı olarak sayılabilmektedir [14,17-18,24].

Modern yaşam tarzı, insanların hareketliliğini azaltmakta olup, beslenme alışkanlıklarını ise hızla değiştirmektedir. Özellikle doymuş yağlardan zengin olup posadan fakir, kalorisi yüksek ve hızlı hazırlanan beslenme tarzının insanlar tarafından benimsenmesi diyabet prevalansında hızlı bir artışa yol açmıştır [55-61]. İlerleyen yaşla birlikte diyabet riski de hızla artmakta olup, BAG (Bozulmuş Açlık

Glikozu), BGT (Bozulmuş Glikoz Toleransı), gestasyonel diyabet, obezite, sigara, hipertansiyon, sağlıksız beslenme ve aktivite azlığı olan bireylerde APG (Açlık Plazma Glukozu) normal sınırlarda olsa bile, Oral Glukoz Tolerans Testi (OGTT) yapılması ve daha genç yaşlardan itibaren daha sık aralıklarla araştırılması önerilmektedir [14,17]. Diyabet riski altında olan bireyler için APG (Açlık Plazma Glukozu) veya BGT (Bozulmuş Glukoz Toleransı) fazında tespit edildiğinde erkenden alınacaklar tedbirler diyabet ve komplikasyonlarını önleyebilir [24].

Obezite de diyabet riskini oldukça arttırmakta olup, yaşam tarzı değişiklikleri hastalığın önlenmesinde olması gereken bir yaklaşımdır. Özellikle tüm yaş gruplarında obezitenin engellenmesi ile Tip II diyabet riskinin %80 oranında azaltılacağı belirtilmiştir. Yapılan çeşitli klinik çalışmalarda, yüksek riskli yetişkinlerde kilo kaybının diyabetin başlangıcını geciktirdiği, beden kitle indeksinin (BKİ) 31'in üzerine çıktığında, diyabet riskinde de bir artış olduğu gösterilmiştir. Düzenli beslenme ve spor ile vücut ağırlığının ortalama %7'lik bir kaybında, farmakolojik tedavi gibi yan etkiler ortaya çıkarmadan diyabet gelişiminin 3-4 yıl içinde %50 oranında azaltılabileceği belirtilmiştir [16-17,62-63]. Diyabetin önlenmesine ilişkin yapılan çalışmalarda BGT (bozulmuş glikoz toleransı) saptanan kişilerde Tip II diyabet gelişme riskinin 10 yıl içinde yaklaşık olarak %50 olduğu, yaşam tarzında yapılan değişiklikleri uygulanmasıyla bu riskin %58'e varan oranlarda önlenebileceği veya geciktirilebileceği belirtilmiştir [64].

Diyabet yaşam boyu tedavi ve takip gerektiren bir hastalık olup kişiye özel uygulanmış diyet ve egzersiz tedavinin olmazsa olmazıdır. Özellikle Tip II diyabetik hastalar, egzersizsiz bir yaşam biçimi haline getirip en azından haftada üç kez 20-30 dakika süreli yürüyüş yapması, bu tedavi sürecinin ilk adımı olarak ele alınabilir. Bunun yanında Tip I diyabet hastalarında mutlaka insülin tedavisi uygulanırken, Tip II diyabet hastalarında ise diğer diyabet ilaçları ve gerektiğinde insülin kullanılır. Diyabet tedavisinde eğitiminde önemli bir rolü olmaktadır. Koşulların sürekli değişmesinden dolayı, bireylerin meydana gelebilecek durumlara uyum sağlamsı için eğitimin sürekli devam ettirilmesi gerekmektedir [24].



## 4. VERİ MADENCİLİĞİ

Çalışmanın bu bölümünde, veri madenciliğinin tanımına ve temel kavramlarına yer verilmiştir. Ayrıca tez çalışması sırasında test edilmiş olan veri madenciliği algoritmalarının işleyişi detaylandırılarak açıklanmıştır.

### 4.1. Veri Madenciliği Tanımı ve Temel Kavramlar

Sanayi devriminin aşamaları incelendiğinde, birinci aşamada su ve buhar enerjili sistemler ön planda iken 2. 3. ve 4. sanayi devrimlerinde sırasıyla elektrik enerjisi kullanımı, otomasyon sistemler ve veri işleme dayalı üretimin olduğu görülmektedir. Sanayi ve üretimde mevcut bilgiyi işlemenin önemi, kuşkusuz veri kaynaklarının katlanarak artmasından kaynaklanmaktadır. Bu veri kaynakları, kredi kartıyla alışverişlerimiz, mobil telefon operatörleri tarafından saklanan veriler, köprü geçişleri, internet alışverişleri ve sosyal medya kullanıcılarından elde edilen veriler sayılabilir. Küresel tahminler, yeni sanayi devrimiyle bağlantılı teknolojilerin daha çok uygulama alanı bulacağını ve sürekli gelişme eğilim göstereceğini ifade etmektedir.

**Çizelge 4.1.** Yeni Sanayi Devrimi: Akıllı Üretim Sistemleri Teknoloji Yol Haritası  
[65]

YIL	ÖNGÖRÜ
2018	Sanayide kullanılacak robot sayısı yaklaşık 3 milyon olacak. Birbirine bağlı cihaz sayısı 13 milyardan 29 milyara çıkacak.
2020	Nesnelerin interneti pazarının büyüklüğü 656 Milyar USD'den 1.7 Trilyon USD'ye çıkacak.
2025	Endüstriyel robotların yaratacağı ekonomik etki yıllık 0.6-1.2 Trilyon \$ Gelişmiş ülkelerdeki imalat süreçlerinin %15-25 oranında otomasyona dayalı olacak OECD (Organisation for Economic Co-operation and Development) ekonomilerindeki yenilik aracılığıyla, GHYİH artışı verimlilik artışına bağlı hale gelecek.
2030	Dijital teknolojilerin verimlilik, gelir dağılımı ve çevre üzerine güçlü etkileri olacak. Küresel ticaret hacminin yarısı akıllı nesnelerin etkileşimini kullanacak.

Çizelge 4.1'de yer alan TÜBİTAK verilerine göre, veri oluşturacak kaynakların ve verinin önümüzdeki süreçte katlanarak arttığı gösterilmektedir. Artan verilerin bilgiye dönüşmesi yani değerli hale gelmesi için işlenmesi gerekmektedir. Bu noktada veri madenciliği ile büyük hacimli veri toplulukları içerisinde daha önce bulunmamış çok net olmayan fakat potansiyel olarak kullanışlı ve anlamlı bilgilerin, çıkarılması sağlanmaktadır. Bu süreç, büyük veri yığınlarının işlenmesi ile başlayarak, analiz edildikten sonra uzman görüşünün yorumlanması ile son bulmaktadır [66].

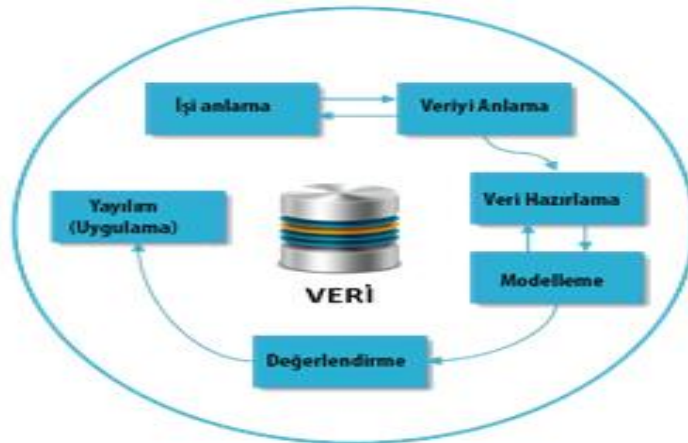
Veri madenciliği hakkında farklı bilim insanlarının tanımlamalarına bakıldığında, evrensel bir fikir birliği olmadığı söylenebilir. En temel ifadeyle, istatistik analiz yöntemleri ve yapay zekâ algoritmalarının birlikte kullanılarak, ham verilerin

işlenmesi ve nitelikli bilgiye dönüştürülmesi işlemine veri madenciliği denilmektedir. Veri madenciliğinin öne çıkan tanımları ise şu şekildedir;

- Veri madenciliği, matematiksel ve istatistiksel teknikler sayesinde veri ambarlarında depolanan büyük miktarlardaki verinin örüntü tanıma teknolojileriyle birlikte incelenmesi yardımıyla anlamlı yeni ilişkiler, eğilimler ve örüntüler bulunması sürecidir [122].
- Veri madenciliği, karar verme sürecini daha iyi sağlamak amacıyla önceden bilinmeyen ilişkileri, veri düzenlerini ve kümeleri keşfetmek ve görüntülemek için verileri inceleme sürecidir [123].

#### 4.2. Veri Madenciliği Süreci

Veri Madenciliği Süreci, veri madenciliği ile ilgili projelerin, daha hızlı, verimli ve daha az maliyetli yapılmasına olanak sağlamaktadır. Doğrusal bir süreç değildir. Veri madenciliği üzerinde araştırma yapan veya bu işi meslek edinmiş kişilerin süreci belli bir kural dâhilinde yürütmesi gerekmektedir. Şekil 4.1'de veri madenciliği sürecinin genel işleyiş modeli verilmiştir.



Şekil 4.1. CRISP-DM İşleyiş Modeli

Veri madenciliği alanında yapılan çalışmalar sonucunda çapraz endüstri standardı süreci olan CRISP-DM oluşturulmuştur. Şekil 4.1’de gösterildiği gibi bu standartlaştırma ile veri madenciliği süreci doğrusal olmayıp, veriden bağımsız şekilde 6 basamaktan oluşmaktadır. Bu basamaklar incelenecek olunursa [66];

**1- İş Anlayış Safhası:** İşin anlaşılması yani veri madenciliği işlemiyle “Ne yapılmak isteniyor?” sorusuna cevap aranmaktadır. Ayrıca çalışma için gerekli kaynaklar, tahmini maliyet, mevcut kısıtlar, olası riskler, amaç ve hedefler değerlendirilerek formülize edilmektedir.

**2- Veri Anlayış Safhası:** Verinin toplanması, tanımlanması, uygunluğunun değerlendirilmesi, model için gerekli başka veri ihtiyaçlarının belirlenmesi, veri yeterlilik ve kalitesinin değerlendirilmesi gerçekleştirilmektedir.

**3- Verinin Hazırlanması Safhası:** Başlangıç verilerinin çalışmalara temel oluşturulacak sonuç verilerine dönüştürülmesi aşamasıdır. En uzun safha olup, modelin oluşturulmasında ortaya çıkan sorunlar nedeniyle bu aşamaya sık sık dönülmekte ve verinin model için hazır duruma getirilmesi sağlanmaktadır. Bu süreçte izlenen yol şu şekildedir;

- Öncelikle, modelin kurulması ve tanımlanan soru için gerekli olduğu düşünülen veri seti belirlenmektedir.
- Daha sonra yapılan veri temizlemesi ile modeli bozacak, kurulmasına engel olacak veriler çıkarılmakta ya da eksik veri mevcutsa tamamlanmakta olup daha doğru modelleme yapılmasına olanak sağlanmaktadır.
- Model için daha kullanılır değişken setleri oluşturmak amacıyla, mevcut değişkenlerde modifikasyonlar yapılmaktadır.
- Veri birleştirme yöntemi ile farklı kaynaklardan toplanan ve uyumsuzluğa neden olan mevcut verilerin tek bir veri tabanında toplanması amaçlanmakta olup,

toplanan verilerin de ne kadar uyumlu olduđu bu aşamada incelenerek değerlendirilmektedir.

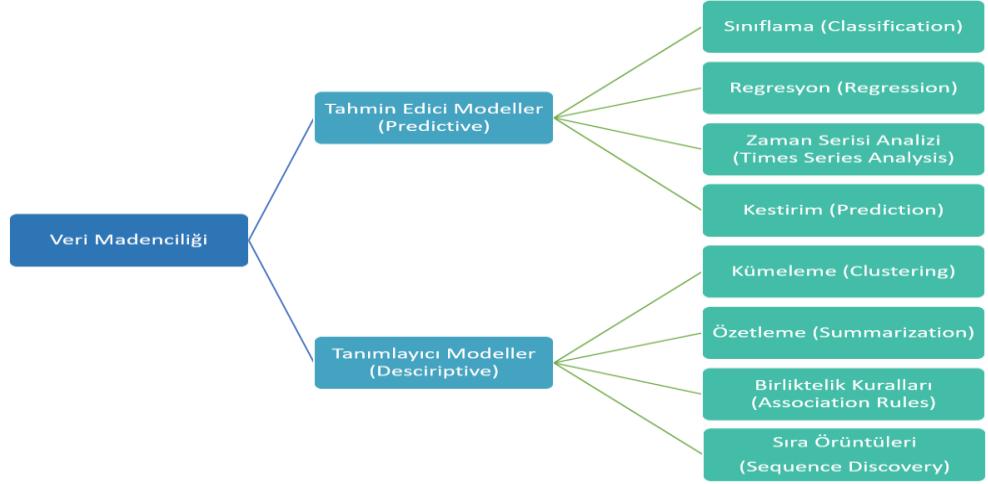
**4- Modelin Oluşturulması Safhası:** Bu aşamada öncelikle uygun modelleme tekniğinin seçilip, daha sonra sonuçları elde etmeye başlamadan önce modelin kalitesi ve geçerliliği test edilmektedir. Eğer model uymadıysa tekrar başka bir modele geçiş yapılmaktadır.

**5- Değerlendirme Safhası:** Bu safhada artık kurulmuş bir model ortaya çıkmakta olup, değerlendirilerek hedefe ulaşip ulaşılmadığı kontrol edilmektedir. Buna bağlı olarak karşılaşılan problemler açıklanmakta, süreç incelenmekte ve gelinen noktanın yeterli olup olmadığı bunun yanında ek bir çalışma gerekip gerekmediği yorumlanmaktadır.

**6- Yayılma (Uygulama) Safhası:** Modelin sonuçlarına göre gözlem ve bakım işleri planlandıktan sonra rapor hazırlanıp, piyasaya sürülmektedir [66].

### **4.3. Veri Madenciliği Modelleri**

Şekil 4.2’de gösterildiği gibi veri madenciliğinde kullanılan modeller, “*Tanımlayıcı (Descriptive)*” ve “*Tahmin Edici (Predictive)*” olmak üzere iki temel başlık altında toplanmaktadır.



Şekil 4.2. Veri Madenciliği Metotları

### 4.3.1. Tanımlayıcı (Descriptive) Modeller

Tanımlayıcı modellerde amaç, belli bir hedefi tahmin etmekten ziyade veri setindeki veriler arasında olan ilişkileri, bağlantıları ve davranışları bulmaktır. Böylece mevcut veriler yorumlanarak, tekrarlanan bir durumda ya da tanımlanmış yeni bir verinin katılmasında neler yapılacağına dair karar verilebilmektedir. Tanımlayıcı modeller Kümeleme (Clustering), Özetleme (Summarization), Birliklik Kuralları (Association Rules) ve Sıra Örüntüleri (Sequence Discovery) olarak sınıflandırılmaktadır [66].

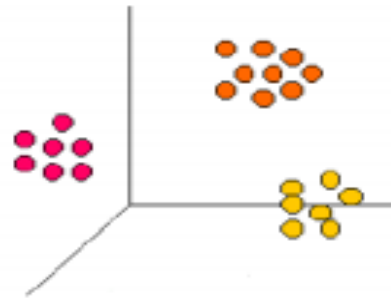
#### 4.3.1.1. Kümeleme (Clustering) Analizi

Grup sayısı bilinmeyen ve gruplandırılmamış verilerin kendi aralarında, benzerlikler göz önüne alınarak sınıflandırılması işlemine kümeleme adı verilmektedir. Diğer bir ifadeyle kümeleme analizi, verilerin birimlere veya değişkenlere göre birbirlerine benzerlikleri bakımından ayrı kümelerde toplanmasını sağlayan bir tekniktir. Kümeleme analizi, birbirlerine benzer bireylerin aynı grupta toplanmasını amaçlaması ile “diskriminant analizine”, birbirine benzer değişkenlerin aynı

gruplarda toplanmasını amaçlaması nedeniyle de “faktör analizine” benzerlik göstermekte olup aynı zamanda veri indirgeme özelliği de bulunmaktadır [67,68].

Kümeleme algoritmaları; küme içinde benzerliğin maksimize (küme içi uzaklıkların minimize edilmesi) edilmesi, kümeler arası benzerliğin ise minimize (kümeler arası uzaklıkların maksimize edilmesi) edilmesi kavramına dayanmaktadır. Sonuçta ise elde edilen farklı kümelere ait elemanlar arasında benzerliklerin az olduğu görülmektedir [66].

Genel olarak birimler arası uzaklıklar için Standardize, Öklit, Manhattan Kareli, Mahalanobis, Minkowski veya Canberra ölçüleri kullanılmaktadır. Kümeleme modellerinde amaç Şekil 4.3’de gösterildiği gibi kümede yer alan üyelerinin birbirlerine oldukça benzediği, fakat özellikleri birbirlerinden çok farklı olan kümelerin bulunması ve veri tabanındaki kayıtların bu farklı kümelere bölünmesidir [69,70].



**Şekil 4.3.** Kümeleme Örneği

Kümeleme yöntemi, satış hareketleri ya da çağrı merkezi kayıtları gibi çok fazla parametre bulunduran çok büyük miktarlardaki verileri analiz etmek için kullanılan en uygun yöntemlerden biridir [71]. Örneğin kümeleme yöntemi, müşteri verisi ile yapılacak bir analizde kullanıldığında zaman öncelikle ortak özelliklere sahip müşteriler aynı grupta toplanmaktadır. Gruplar kendi içinde çok farklı açılardan benzer

özellikler ve davranışlar gösterdiği için pazarlama faaliyetlerinde de benzer harcama eğilimleri göstereceği düşünülür [72].

Kümeleme yöntemleri hiyerarşik ve hiyerarşik olmayan (bölümlemeli) şekilde iki kümeye, bu kümelerde daha alt kümelere ayrılmaktadır [73]. Hiyerarşik kümeleme yönteminde, işleyişin daha kolay şekilde anlaşılır hale gelmesi için dendogram (ağaç grafiği) kullanılır. İşleme bağlı hiyerarşik yöntemlerden en fazla kullanılan metotlar ise Tek bağlantılı, Tam bağlantılı, Ortalama bağlantılı, Merkezi ve Ward'dır. Hiyerarşik olmayan kümeleme yöntemi ise araştırmacının küme sayısı hakkında önden bir bilgisi olup, anlamlı olacak küme sayısına karar vermiş olması durumunda tercih edilmektedir. Bu yöntemde en çok tercih edilen, Mac Queen tarafından geliştirilmiş olan k-ortalama tekniği ve en çok olabilirlik tekniğidir [70].

Kümeleme analizi ile sınıflandırma arasındaki en büyük fark, kümelemenin girdileri önceden tanımlanmamış olup, veriler kendi içerisinde benzerliklerine göre gruplanmaktadır. Sınıflandırma fonksiyonunda ise tanımlı girdiler ve bunların geçmişte aldıkları değerler temel modeli oluşturmaktadır [66].

#### **4.3.1.2. Özetleme (Summarization) Analizi**

Genelleme veya nitelendirme olarak da adlandırılan özetleme analizinin amacı, veriyi basit açıklamalarla alt kümelere eşleştirerek veri tabanını temsil eden özet bilgiyi çıkarmaktır. Bu işlem, verinin bir takım alt parçalarına ulaşarak gerçekleştirilebilmektedir. Böylece ortaya, veriden özet tipli olacak şekilde bir bilgi çıkmaktadır [74].

#### **4.3.1.3. Birliktelik (Association Rules) Analizi**

Veri madenciliğinde kullanılan ilk tekniklerden biri olan birliktelik kuralında, öncelikle mevcut veri kümesindeki veriler analiz edildikten sonra, yüksek sıklıkta birlikte görülen özellik değerlerine göre ilişkisel kurallar keşfedilmektedir [66].



Diğer bir ifadeyle, öncelikle nitelikler arasındaki bütün kombinasyonlar ortaya çıkarılmakta ve her niteliğin sahip olduğu bütün değerler kombinasyonlar ile örtüştürülerek örüntüler keşfedilmeye çalışılmaktadır. Birliktelik analizi, mağaza ürün yerleşim planı, pazar sepet analizi, katalog tasarımı gibi alanlarda yaygın olarak kullanılmaktadır. Birliktelik analizinde modelin kurulması, büyük veri setleri arasında aranan ilişkilerden dolayı hem zor hem de zaman almaktadır. Büyük veri tabanlarında birliktelik kuralları bulunurken geliştirilmiş algoritma ise, “*Apriori Algoritması*”dır [66].

Örneğin Pazar sepet analizi, müşterilerin sıklıkla aldığı ürünler arasında bir örüntü kurup, müşterilerin satın alma alışkanlıklarını ortaya çıkarmaktır. Örnek model geliştirilecek olunursa, X ürününü satın alan müşteri Y ürününü de satın alıyorsa bu durum  $X \rightarrow Y$ , Y [destek = %2, güven = %60] şeklinde ifade edilmektedir. “*Destek*” kuralın sıklığını, “*Güven*” ise tanımlanan kuralın kabul edilebilirliğini göstermektedir. Destek değeri, tüm alışverişlerden %2’lik oranında X ve Y ürünlerinin beraber satıldığını gösterirken, %60 güven ise X ününü alan müşterinin aynı alışverişte bu oran kadar Y ürününü de satın aldığını belirtmektedir. Burada dikkate alınması gereken, kullanıcı tarafından belirlenen minimum destek eşik değeri ve minimum güven değerini aşan birliktelik kurallarıdır [66].

#### **4.3.1.4. Sıra Örüntüleri (Sequence Discovery) Analizi**

Sıra örüntüleri, yapılan gözlem sonuçlarını zaman ve mekân özelliklerine göre sıralanmış olarak gösteren sayı dizileri olarak adlandırılmaktadır. Bu fonksiyona ile ilişkili olan kayıtlar incelenmekte ve buna göre zaman içerisinde sıkça rastlanan trendler bulunmaktadır [66]. Verimli veri tabanları ortaya çıkarmak ve bu veri tabanlarında tutulan dizilere erişim sağlamak için indeksler üretmek, sıklıkla tekrar eden örüntüleri elde etmek, benzerlik için dizileri karşılaştırmak (DNA’daki gibi) ve dizinin kayıp elemanlarını yeniden oluşturmak gibi bu alanda önemli birçok problem yer almaktadır.

### **4.3.2. Tahmin Edici (Predictive) Modeller**

Tahmin edici modellerin amacı, sonuçları bilinen verilerden yola çıkılarak bir model geliştirilmesi ve geliştirilen bu modelden yararlanılarak sonuçları bilinmeyen veri kümelerinin sonuç değerlerinin tahmin edilmesidir. Tahmin Edici Modeller, Sınıflama, Regresyon, Zaman Serisi Analizi ve Kestirim olarak sınıflandırılmaktadır [75].

#### **4.3.2.1. Sınıflama (Classification) Analizi**

Sınıflama, yeni olan nesnenin niteliklerini gözden geçirip daha sonra bu nesneyi önceden tanımlanmış olan bir sınıfa atamaktır. Burada önemli olan, sınıfların her birinin özelliklerinin öncesinde net bir şekilde belirlenmiş olmasıdır. Örneğin, kredi kartı başvurularını düşük, orta ve yüksek risk grubu olarak ayırmak sınıflama olarak gösterilebilmektedir. Sınıflama analizinde, sınıflar önceden incelenen veriler aracılığı ile oluşturulduğu için denetimli öğrenme şeklinde de adlandırılabilirler. Çalışmamızda, sınıflama analizinde yaygın olarak kullanılan yöntemler, ilerleyen bölümde başlıklar halinde incelenmektedir.

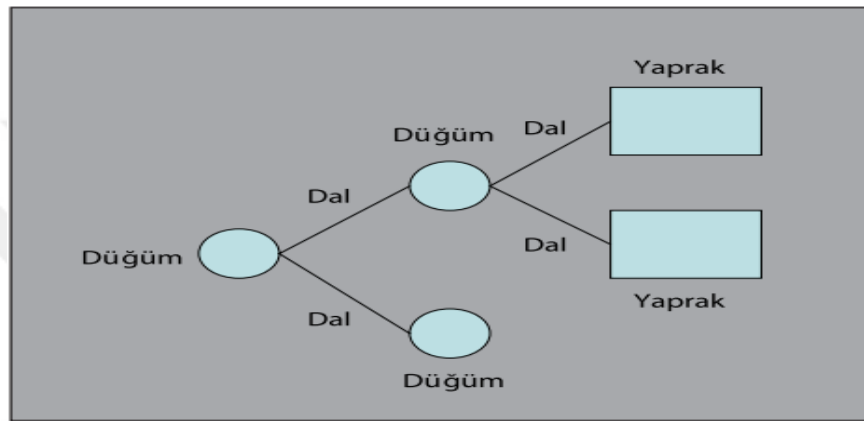
##### **4.3.2.1.1. Karar Ağacı (Decision Tree) Yöntemi**

Karar ağacı yöntemi, yorumlamasının kolay olması, veri tabanı sistemleri ile rahatça entegre edilmesi, kuruluşunun ucuz olması ve güvenilirliğinin daha iyi olması açısından en çok kullanılan sınıflama tekniklerinden biridir. Karar ağaçlarının amacı bağımlı değişkendeki farklılıkları maksimize seviyeye getirecek şekilde veriyi sıralı şekilde parçalarına (farklı gruplara) ayırmaktır [66].

Uygulanan istatistiksel yöntemlerde ya da yapay sinir ağlarında, veriden bir fonksiyon öğrenildikten sonra bu fonksiyonun anlaşılabilir bir kural olarak yorumlanması zordur. Karar ağacı yöntemi ile ağaç oluşturulduktan sonra, kökten yaprağa doğru inilerek kurallar çıkarılabilmektedir. Daha sonra bu kurallar, karar

verici bir uzmana gösterilerek sonucun anlamlı olup olmadığı denetlenebilmektedir. Karar verici daha sonra başka teknik kullanacak olsa bile karar ağacı yöntemi sayesinde önemli değişkenler ve kurallar konusundan bilgi edinilmiş olmuştur [66].

Karar ağaçları Şekil 4.4'deki gibi ağaç görünümünde olup, yaprak, dal ve karar düğümlerinden meydana gelmektedir [66, 76].



Şekil 4.4. Karar Ağacı Yapısı

Karar düğümü, veriye uygulanacak testi belirlemektedir. Testin sonucu, ağacın veri kaybetmeden dallara ayrılmasına neden olmaktadır. Her düğümde test ve dallara ayrılma işlemleri, ardışık olarak gerçekleşmekte ve bu ayrılma işlemi üst seviyedeki ayrımlara bağlı olmaktadır. Ağacın her bir dalı tanımlanacak sınıfın belirlenmesine adaydır. Bir dalın sonucunda sınıflama işlemi gerçekleşmiyorsa, tekrar bir karar düğümü oluşur. Eğer, dalın sonunda belirli bir sınıf elde edilebiliyorsa, o dalın sonunda yaprak var demektir. Bu yaprak, verileri kullanarak elde edilmek istenen sınıflarından birini tanımlamaktadır. Karar ağacı işlemi kök düğümünden başlar ve yukarıdan aşağıya doğru yaprağa ulaşana kadar ardışık düğümleri takip ederek gerçekleşmektedir [75].

Karar ağacı yöntemleri, gelecekte olabilecek olaylar için tahmin kuralları oluşturmada, risk grupları kategorileri oluşturmada, bilinmeyen yeni bir örneğin sınıflandırılmasında, kategori birleştirilmesinde uygulanmaktadır [66].

Bu çalışmamada da karar ağaçları sınıflandırma yöntemlerinden biri olan C&R Tree algoritması kullanılıp, makine öğrenmesi başlığı altında detaylandırılmaktadır.

**Çizelge 4.2. Örnek Veri Seti T**

<b>Haftasonu</b> <b>(Weekend)</b>	<b>Hava</b> <b>(Weather)</b>	<b>Ebeveyn</b> <b>(Parents)</b>	<b>Para</b> <b>(Money)</b>	<b>Karar</b> <b>(Decision)</b>
<b>Örnek</b> <b>(Example)</b>				<b>Kategori</b> <b>(Category)</b>
H1	Güneşli	Evet	Zengin	Sinema
H2	Güneşli	Hayır	Zengin	Tenis
H3	Rüzgarlı	Evet	Zengin	Sinema
H4	Yağmurlu	Evet	Fakir	Sinema
H5	Yağmurlu	Hayır	Zengin	Evde Kalmak
H6	Yağmurlu	Evet	Fakir	Sinema
H7	Rüzgarlı	Hayır	Fakir	Sinema
H8	Rüzgarlı	Hayır	Zengin	Alışveriş
H9	Rüzgarlı	Evet	Zengin	Sinema
H10	Güneşli	Hayır	Zengin	Tenis

Kullanılan karar ağaç yönteminin gerçekleştirilecek adımlarını daha detaylı incelemek için Çizelge 4.2'deki örnek veri seti göz önünde bulundurulursa [77];

- Öncelikle, mevcut verilerimizden, hava (weather), ebeveyn (parents) ve para (money) durumuna göre ne yapılacağına karar verilip, ağacın kökü olarak

adlandırılan veri setindeki en ayırt edici özellik belirlenir. Bu özelliğin nasıl belirlendiği alt kısımda anlatılmaktadır.

- Daha sonra alt kümelerine inilerek, ağacın çocuk düğümü olan A düğümüne ait alt veri kümesi belirlenir.
- Her belirlenen alt veri kümesi için bilgi kazancı hesaplanarak en ayırt edici özellik bulunur.
- ID3, C4.5 gibi karar ağacı metotlarında en ayırt edici niteliği belirlemede her nitelik için bilgi kazancı ölçülmektedir.
- Bilgi kazancı ölçümünde, rastgeleliği, belirsizliği ve beklenmeyen durumun ortaya çıkma olasılığını gösteren Entropi kullanılmaktadır.

Aşağıda toplamları 1 olan  $(p_1, p_2, \dots, p_s)$  olasılıkların üzerinden yapılan entropi ve bilgi kazancı hesaplamaları şu şekildedir;

Entropi hesaplaması;

$$H(p_1, p_2, \dots, p_s) = -\sum_{i=1}^s p_i \log(p_i) \quad (1)$$

A niteliğinin S veri kümesindeki bilgi kazancı (gain) hesaplaması;

$$Kazanç(S, A) = Entropi(S) - \sum P_v Entropi(P_v) \quad (2)$$

Örnek veri kümesindeki (T kümesi) 10 örnekten yola çıkılırsa ve bunlardan;

- 6 örnek için karar sinema,
- 2 örnek için karar tenis oynama,
- 1 örnek için karar evde kalma,
- 1 örnek içinse karar alışverişe gitmek olursa sistemin entropisi aşağıdaki şekilde hesaplanmaktadır.

Entropi;

$$H(T) = -(6/10) \log_2(6/10) - (2/10) \log_2(2/10) - (1/10) \log_2(1/10) - (1/10) \log_2(1/10)$$

$$H(T) = 1,571$$

Sistemimin entropisini bulduktan sonra, kök özelliğinin hangisi olacağına karar verebilmek için hava (weather), ebeveyn (parent) ve para (money) özelliklerinin bilgi kazançlarını hesaplayıp en yüksek kazançta sahip olan özelliği, kök özellik olarak konumlandırmak gerekmektedir.

Örneğin hava (weather) için bilgi kazancı hesaplanacak olunursa öncelikle, entropi değeri bulunur ve daha sonra bu değere bağlı olarak bilgi kazancı hesaplanır.

Güneşli (Sunny)=3 (1'i sinema, 2'si tenis)

Rüzgârlı (Windy)=4 (3'ü sinema, 1'i alışveriş)

Yağmurlu (Rainy) =3 (2'si sinema, 1'i evde kalmak)

Entropi değerleri;

$$E(T_{\text{sunny}}) = - (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0,918$$

$$E(T_{\text{windy}}) = - (3/4) \log_2 (3/4) - (1/4) \log_2 (1/4) = 0,811$$

$$E(T_{\text{rainy}}) = - (2/3) \log_2 (2/3) - (1/3) \log_2 (1/3) = 0,918$$

Hava durumu Bilgi Kazancı;

$$\text{Gain}(T, \text{weather}) = 1,571 - ((3/10) * 0,918 + (4/10) * 0,811 + (3/10) * 0,918)$$

$$\text{Gain}(T, \text{weather}) = 0,70$$

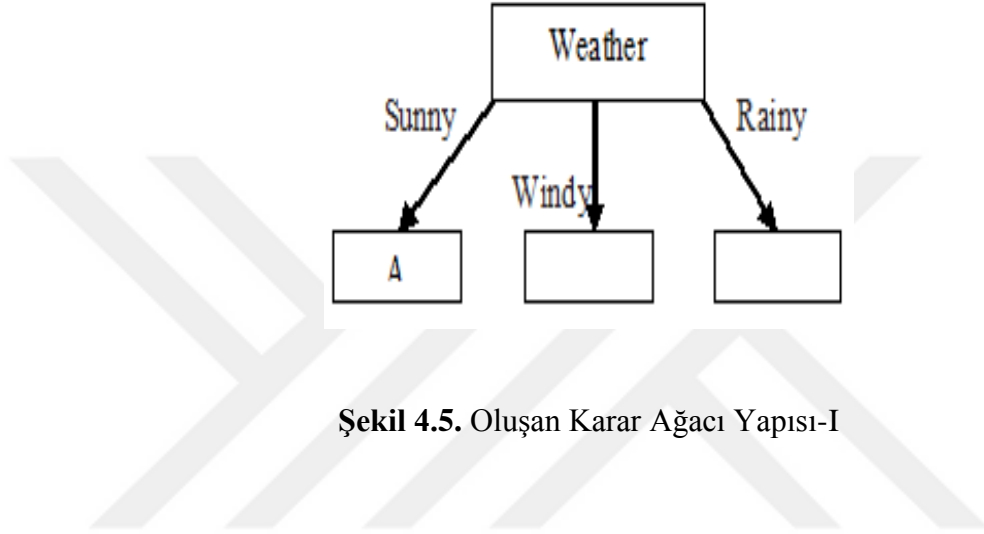
Hava durumuna ait bilgi kazancından sonra aynı şekilde ebeveyn (parent) bilgi kazancı, para (money) durumu bilgi kazancı da hesaplanıp, sonuç değerleri aşağıdaki gibi çıkmaktadır;

$$\text{Gain}(T, \text{weather}) = 0,70$$

Gain (T,parents)=0,61

Gain (T,money)=0,2816

En büyük bilgi kazancına hava (weather) özelliği sahip olduğu için, bu özellik en ayırt edici olarak bulunmuş olup, ağacın kökünde yer alacak nitelik olarak seçilir. Şekil 4.5’de bu süreç sonucunda oluşan karar ağacı yer almaktadır.



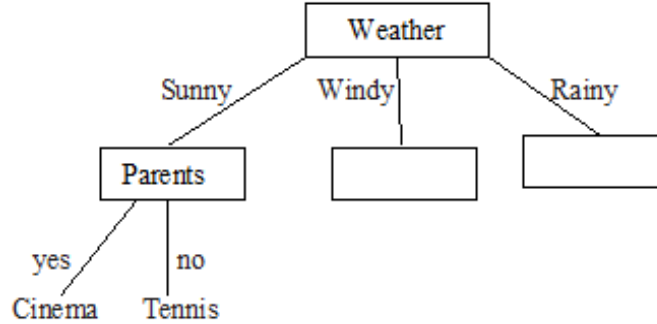
Şekil 4.5. Oluşan Karar Ağacı Yapısı-I

Daha sonraki süreçlerde, oluşan karar ağacı daha da dallandırılmak istenirse sunny, windy ve rainy tek tek ele alınıp, her birinin bir önceki gibi bilgi kazançları hesaplanır. Örneğin, Sunny’nin bilgi kazancının hesaplanması için gerekli olan nitelikler Çizelge 4.3’de yer almaktadır.

Çizelge 4.3. Sunny’nin Bilgi Kazancının Hesaplanması İçin Gerekli Olan Nitelikler

Haftasonu (Örnek)	Hava	Ebeveyn	Para	Karar (Kategori)
H1	Güneşli	Evet	Zengin	Sinema
H2	Güneşli	Hayır	Zengin	Tenis
H10	Güneşli	Hayır	Zengin	Tenis

Yine yapılan hesaplamalar sonucu modelimizde en ayırt edici ebeveyn (parents) olarak belirlenmiş olup, oluşan karar ağacı Şekil 4.6’da gösterilmektedir.



Şekil 4.6. Oluşan Karar Ağacı Yapısı-II

- Bu dallanma, örneklerin hepsi aynı sınıfa ait, örnekleri bölecek özellikler kalmamış ve kalan özelliklerin değerini taşıyan örnek olamayana kadar devam etmektedir [77].

#### 4.3.2.1.2. Bayes Sınıflandırması (Bayesian Classification)

İstatiksel bir sınıflandırma olup mevcut verilerin belirlenmiş sınıflara ait olma olasılıklarını tahmin etmektedir. Bayes sınıflandırmasının gerçekleştirilmesi kolay olup, belirsizlik taşıyan durumlar için karar verme konusunda çok kullanışlıdır. Ancak değişkenler arası ilişkinin modellenmiyor olması ve değişkenlerin birbirinden tamamen bağımsız olduğu varsayımı dezavantajıdır. Özellikle, bu çalışmada da kullanılan Naive Bayesian Sınıflandırıcı (Simple Bayesian Classifier) oldukça başarılı olup makine öğrenmesi başlığı altında detaylandırılmaktadır [66].

Makine öğrenmesi, veri madenciliği gibi çok çeşitli alanlarda kullanılan k-en yakın komşu algoritması (k-NN); uygulanmasının kolay, öğrenme sürecinin güçlü olması nedeniyle sınıflandırma yöntemlerinde yaygın olarak tercih edilmektedir. Bu



çalışmada da makine öğrenmesi başlığı altında, k-NN algoritması detaylandırılıp, parametrelerin sınıflandırılması üzerindeki etkisi incelenmiştir [78].

#### **4.3.2.1.3. Yapay Sinir Ağları (Neural Networks)**

Yapay sinir ağları, temelinde insan beyninin işlemesi mantığını bulundurarak nöronların matematiksel şekilde modellenmesidir. Öğrenme faaliyeti ile modeli geliştiren bu tekniğin amacı öğrenme süreci boyunca davranış biçimlerini anlayıp, hatayı en aza indirmektir. Bu çalışmada %97.2'lik doğru tahmin başarısıyla en iyi başarıyı elde eden yapay sinir ağları makine öğrenimi başlığı altında detaylandırılmaktadır [66].

#### **4.3.2.2. Regresyon (Regression) Analizi**

Regresyon, bir kriter değişkeni ile bir ya da daha fazla tahmin değişkenleri arasındaki ilişkiyi matematiksel hale dönüştürmede kullanılan istatistiksel bir yöntemdir. Regresyon analizinin asıl amacı, değişkenler arasındaki ilişkinin niteliğini belirlemektir. Regresyonda verilerin bir fonksiyon olarak tanımlanması gerekmektedir. Regresyon analizi sırasında kurulan matematiksel modelde bulunan değişkenler bir bağımlı değişken ve bir ya da birden çok bağımsız değişkenden meydana gelmektedir. Buradaki değişkenler hem sayılabilir hem de ölçülebilir özelliktedir. Tek değişkenden oluşan modeller basit doğrusal regresyon, birden fazla bağımsız değişkenlerden oluşan modeller ise çoklu regresyon olarak adlandırılmaktadır. Amaç her bir tahmin değişkeninin, kriter değişkenindeki toplam değişmeye olan katkısının saptanması ve böylece tahmin değişkenlerinin doğrusal kombinasyonunun değerinden hareketle, kriter değerinin tahminidir [66,79].

İki değişken arasındaki ilişkinin gücünü belirlemek, değişkenler arasındaki ilişkinin türünü belirlemek, geleceğe yönelik değerleri tahminde bulunmak gibi çeşitli konularda kullanılmakta olup en fazla araştırma, matematik, finans, tıp gibi bilim

alanlarında tercih edilmektedir. Bu çalışmada ise lojistik regresyon yöntemi kullanılmış olup, makine öğrenimi başlığı altında detaylandırılmaktadır [66,79].

#### **4.3.2.3. Zaman Serisi (Time Series) Analizi**

Zaman serileri; ekonomi, istatistik, meteoroloji, jeofizik, tıp, tarım ve biyoloji alanlarında ortaya çıkan, zamana bağlı verilerin analizinde tercih edilen bir yöntem olup, yapılan gözlem sonuçlarının kronolojik sırayla sıralanmış halidir. Genel olarak zaman serisinde yer alan veriler günlük, haftalık, aylık, üç aylık, altı aylık, yıllık ve daha uzun dönemli aralıklarla düzenlenmekte ve toplanmakta olup; zaman serilerinde yer alan verilerin davranışları trend ve döngüsel olmak üzere ikiye ayrılmaktadır. “trend” serideki verilerin ortalama değerinde oluşan farklılaşmayı tanımlamak için kullanılırken, “döngü” veride tekrarlanan davranışı tanımlamak için kullanılmaktadır. Zaman serisi analizlerinde hisse değerleri, ürün talebi, hava durumu, ekonomik değerler gibi gelecek değerlerin tahmin edilmesi çalışmaları gerçekleştirilmektedir [66].

#### **4.3.2.4. Kestirim (Prediction)**

Birçok veri madenciliğine ait uygulamalarda, geçmiş ya da mevcut verilerden yola çıkılarak gelecekteki veri değerleri tahmin edilmeye çalışılmaktadır. Kestirim modelleri bir sınıflama modeli şeklinde olup, onu sınıflama modellerinden ayıran en önemli özellik ise gelecekteki verilerin tahmin edilmesidir. Yani kestirim modelleri teknik bir özellik olmayıp, uygulamanın bir özelliği olarak da tanımlanabilmektedir. Kestirim modeli nelere dikkat edilmesi gerektiğini ya da hangi değişkenlerin önemli olduğunu ortaya çıkararak, birtakım değişkenlerin elenmesine yardımcı olup bu sayede ileri seviyedeki araştırmacıların işini biraz daha kolaylaştırmaktadır [66].

#### **4.4. Veri Madenciliği Uygulama Alanları**

Veri madenciliğinin uygulama alanları bilimsel, iş dünyası, eğitim, sağlık ve biyomedikal olmak üzere sınıflandırılabilir. Bilimsel çalışmalarda çok büyük miktarlardaki verilerin işlenmesi için, mevcut tekniklerin yetersiz kalmasından ötürü veri madenciliği ile uygun bir platform sağlanmaktadır. İş dünyasının veri madenciliği adımlarını kullanmasının nedeni ise müşteriye tanımak ve müşteri memnuniyeti sağlayıp, hızlı ve doğru kararlar verebilmektir. İş çevresinin her sektöründe veri madenciliğini kullanmak mümkündür, fakat en fazla ihtiyaç duyulan alanlar Finans (Bankacılık, Sigortacılık), Pazarlama, Telekomünikasyon ve Perakendedir. Bunun yanı sıra veri madenciliği eğitim alanında da eğitim kalitesini ve performansını arttırmak için kullanılmaktadır. Sağlık ve tıp alanında ise çok fazla yaygın olmasının nedeni, özellikle tarama testlerinden elde edilen veriler ile önceliklerin ve risklerin tespit edilip buna göre teşhis ve kararlar verilebilmesidir [66].

#### **4.5. Veri Madenciliği ile İlişkili Diğer Bilimler**

Günümüzde, karar verme süreci için gerekli olan birçok alanda veri madenciliği yaygın olarak kullanılmaktadır ayrıca veri madenciliği birçok bilim dalı ile ilişkili olup disiplinler arasında bir çalışmadır. Bu dallar Şekil 4.7’de gösterildiği gibi veri tabanı teknolojileri, istatistik, makine öğrenme, görselleştirme, bilişim ve diğer disiplinler şeklindedir. Veri madenciliği ile bahsi geçen bu disiplinler arasında bir sınır çizmek zordur. Veri madencisinin tüm bu disiplinlerden yararlanıp hangi tekniği veya teknik kombinasyonunu kullanacağı, çalışmanın amacı ile bağlantılıdır [80].



Şekil 4.7. Veri Madenciliğinin Birçok Disiplinle Birleşimi

Özellikle büyük veri yapılarının değerlendirilmesi için veri madenciliği, istatistik ve makine öğrenme tekniklerinden yararlanmaktadır. Buna istinaden veri madenciliği, makine öğrenmesi ve istatistik bilimi arasında yakından bir ilişki söz konusudur [80].

#### 4.6. Makine Öğrenmesi Algoritmaları

Makine öğrenmesi, bilgisayar programları aracılığıyla mevcut verilerden elde edilen deneyimlerin, gelecekteki olayları tahmin etmesine ve modelleme yapmasına olanak sağlayan bir yapay zekâ alanıdır. Veri madenciliği gibi birçok disiplin ile ilişkili bir disiplindir. Veri madenciliğinde kullanılan algoritmaların bir kısmı makine öğrenmesi alanındaki çalışmalar sonucu üretilmektedir. Yapay zekânın bir alt üyesi olan makine öğrenmesinin, geçmişten bugüne gelişimi hızla artmış ve bu alanda çok sayıda algoritma geliştirilmiştir. Bu bölümde, araştırmada kullanılan makine öğrenmesi algoritmalarının genel işleyişi ve yapısı incelenmektedir.

##### 4.6.1. Lojistik Regresyon (Logistic Regression) Analizi

Lojistik regresyon, açıklayıcı değişkenlere göre, cevap değişkeninin beklenen değerlerinin olasılık olarak elde edildiği sınıflama ve atama yapmaya yardımcı olan bir regresyon yöntemidir. Bu yöntem ile açıklayıcı değişkenlerin etkileri, bağımlı değişken üzerinde olasılık olarak hesaplanarak, risk faktörlerinin tahmini şekilde

belirlenmesi sağlanmaktadır. Lojistik regresyon aşağıda gösterildiği gibi formülüne edilmektedir [81,82].

$$P = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}} \quad (3)$$

P: İncelenen olayın gözlenme olasılığını,

$\beta_0$ : Bağımsız değişkenler sıfır değerini aldığı anda bağımlı değişkenin değerini diğer bir ifadeyle sabiti,

$\beta_1, \beta_2, \dots, \beta_k$ : Bağımsız değişkenlerin regresyon katsayılarını,

$X_1, X_2, \dots, X_k$ : Bağımsız değişkenleri,

k: Bağımsız değişken sayısını,

e: Euler sayısını göstermektedir. Yaklaşık değeri 2,71

Bağımsız değişkenlerin en düşük düzeylerinde olasılık 0'a yaklaşmakta, fakat hiçbir zaman 0'a eşitlenmemektedir. Ters durumda ise eğim giderek azalmaya başlamakta ve sonuçta 1'e yaklaşmasına rağmen, hiçbir zaman 1'e eşitlenmemektedir [83].

Ayırıcı bir analiz tekniği olma özelliği taşıyan lojistik regresyon analizini, doğrusal regresyon analizinden ayıran birtakım farklılıklar bulunmaktadır. Tahmin edilecek bağımlı değişken, doğrusal regresyon analizinde sürekli iken, lojistik regresyon analizinde kategoriktir ve kesikli bir değer olmalıdır. Doğrusal regresyon analizinde bağımlı değişkenin değeri, lojistik regresyonda ise bağımlı değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı kestirilir. Bağımsız değişkenlerin ise doğrusal regresyon analizinde çoklu normal dağılım göstermesi şartı aranırken, lojistik regresyonda ise bağımsız değişkenlerin dağılımı ile ilgili hiçbir ön koşul olmayıp daha esneklerdir. Lojistik regresyon genel olarak üçe ayrılmaktadır. Bunlar; İkili (Binary) Lojistik Regresyon, Sıralı (Ordinal) Lojistik Regresyon ve Çoklu (Multinomial) Lojistik Regresyon'dur [84,85].

Daha öncesinde, lojistik regresyon analizi ile yapılan bir çalışmada, çocuklarda doğum ağırlığına etki eden önemli risk faktörlerini ortaya çıkarmak için Tıp verileri kullanılmıştır. Lojistik modele eklenecek bağımsız değişkenler, tek değişkenli lojistik regresyon analiziyle belirlendikten sonra, çok değişkenli modele eklenen her bir değişkenin önemliliği gösterilmiş olup Tek değişkenli modelde gerekli olduğu halde, çok değişkenli modelde önemsiz olan değişkenler modele dâhil edilmeyip dışarda bırakılmıştır. Böylece istenilen model elde edilmiştir. Hem biyolojik olarak kabul edilebilir hem de doğru sınıflama oranının iyi olmasından dolayı, varılan sonuç, final modelin risk faktörlerini belirleme de kullanılabileceği yönündedir [85].

#### **4.6.2. CN2 Rule Induction**

CN2, üzerinde çalışılacak veri setinde çeşitli kurallar oluşturmaya yönelik bir öğrenme algoritması olup veri kümesindeki gürültülü örneklerden kaynaklanan problemlerle ilgilenmektedir. Gürültülü örneklerden kasıt, ölçülen bir değerdeki hata, veri girişi ya da iletimi sırasında oluşan hatalar, hatalı veri toplama gereçleri ve çeşitli teknolojik kısıtlar sonucu oluşan yanlış nitelik değerleridir. Bu durum için CN2, bir dizi kural oluşturup aynı zamanda karar ağacında budama yapabilecek çeşitli istatistiksel teknikler sunmaktadır. CN2 rule induction, sadece sınıflandırma için çalışmakta olup kural tabanlı bir algoritmadır [86].

#### **4.6.3. Naive Bayes**

Matematiksel ve istatistiksel çalışmalarda önemli bir yer tutan bu teorem; herhangi bir durumun modelini oluşturmada evrensel doğruları ve gözlemleri kullanarak, sonuçlar oluşturmayı hedeflemektedir. Bu yaklaşımı klasik istatistiksel yöntemlerden ayıran en önemli özellik; kesin olmayan bir bilginin tahmininde, gözlemleri ve öznel görüşleri kullanmasıdır [87]. Kısaca, koşullu olasılıklar ile sıra dışı olasılıklar arasındaki ilişkiyi göstermektedir. Naive Bayes yönteminde neyin sınıflandırıldığından çok nasıl sınıflandırıldığı göz önünde bulundurulmalıdır. Başka bir ifadeyle öğretilen veri tipinin ne olduğundan ziyade (text, binary), bu veriler

arasında ne şekilde oransal bir ilişki kurulduğu önemlidir. Naive Bayes sınıflandırıcı Bayes teoremine dayanmaktadır.

#### 4.6.4. C&R (Classification and Regression) Tree

CART ya da C&R adıyla anılan teknik, ID3 algoritmasında olduğu gibi dallara ayırma kriterlerini seçmek için entropiden yararlanmaktadır [88]. Kök veya uç olmayan her bir düğümde iki adet dal bulunmaktadır. Ağacın karmaşıklık ölçüsü budama işlemlerini etkilemektedir. 1984 tarihinde Breiman tarafından önerilen hem sayısal hem de nominal veri türlerini, girdi ve kestirimsel değişken olarak kabul edebilen CART algoritmasının sınıflandırma ve regresyonu destekleyici bir yapısı vardır. Sürekli hedef değişkenleriyle çalışmaktadır. İşleyiş öncesinde verinin hazırlanmasına ihtiyaç duymaktadır. Herhangi bir t düğümündeki s dallara ayrılma kriteri şu şekilde gösterilir:

$$CART(s/t) = 2PLPR \sum E(a_i) = \sum_{i=1}^k \sum_{j=1}^n Z_{ij} \|x_j - a_i\| \quad (4)$$

CART algoritması, finasta, kalite kontrolünde, kredi risk tahminlerinde, elektrik mühendisliğinde, pazarlamada, kimya, biyoloji ve sağlık alanında yaygın şekilde kullanılmıştır. Aynı zamanda görüntü sıkıştırma teknolojisindeki ağaçsal vektörel niceleme yaklaşımına da çok fazla katkıda bulunmuştur. CART algoritması ile sağlık alanında biyopsi sonrası agresif prostat kanseri tahmini, tarım alanında ise kuraklık koşullarında soya fasulyesi rekoltesinin modellenmesi ve bu duruma nelerin etki ettiği yapılan çalışmalardan bazılarıdır.

#### 4.6.5. Random Forest

Sınıflandırmada çok başarılı performans gösteren Random Forest algoritması hem hızlı hem de doğruluk oranının yüksek olması nedeniyle, hızlandırma ve torbalama olarak bilinen toplu öğrenmenin çok başarılı iki yöntemine göre daha fazla avantaj

sağlamaktadır. Aslında rastgelelik özelliği sayesinde torbalama yönteminin biraz daha gelişmiş halidir [89-91].

Random Forest, rastgele değişkenler arasından en iyi dalı kullanarak her bir düğümü dallara ayırmak yerine, her bir düğümde rastgele olarak seçilen değişkenler arasından en iyisini kullanarak her bir düğümü dallara ayırır. Her bir veri seti orijinal veri setinden yer değiştirmeli olarak üretilir. Sonra rastgele özellik seçimi kullanılarak ağaçlar geliştirilir. Geliştirilen ağaçlar budanmamaktadır [90,92]. Bu durum Random Forest tahminlerinin doğruluğunu önemli kılmaktadır. Budama işleminin olmaması Random Forest yöntemini, diğer karar ağaçları yöntemlerine göre daha üstün hale getirmektedir. Ayrıca hızlı olması, mümkün olduğunca birbirinden farklı ve istenildiği kadar ağaçla çalışılması tercih edilme nedenleri arasındadır [93].

Genellikle, biyomedikal, sağlık, moleküler biyoloji, fizik, astronomi, finans analizi, kontrol sistemleri, yazılım geliştirme alanlarında yaygın olarak kullanılmaktadır. Random Forest algoritması kullanıcı tarafından 2 parametre almaktadır;

*m parametresi:* En iyi bölünmeyi belirlemek için her bir düğümde kullanılacak değişken sayısı

*n parametresi:* Geliştirilecek ağaç sayısı'dır.

İlk olarak eğitim veri setinin üçte ikisi kadarı önyükleme örnekleri, kalan üçte birlik kısmı da hataları test etmek için kullanılmaktadır. Daha sonra budama olmadan her bir önyüklemeli örnekten ağaç geliştirilmektedir.  $M$  değişkenleri her bir düğümde  $m$  tüm değişkenler arasından rastgele olarak seçilir ve bu değişkenler arasından en iyi dal belirlenir. Burada önemli olan, yapılan öngörü ile düşük korelasyon sağlayan değişken sayısının seçimidir.  $M$  değişken sayısının,  $M$  tane değişken sayısının kareköküne eşit olarak alınması optimum sonuca en yakın sonuçlar vermektedir. Random Forest algoritmasından genelleştirilmiş hata, değişken önemi (variable importance) ve yakınlık analizi (proximity analysis) olmak üzere üç tür parametre elde edilmektedir. Random Forest, budama yapılmadan en büyük boyutta ağaç üretmek için CART (Classification and Regression Tree) algoritmasını kullanmaktadır [90,94-96]. Random Forest, sınıfların homojenliğini ölçmek için



GINI indeksini kullanmaktadır. Eğer GINI indeksi düşükse bu sınıfın o kadar homojen olduğunu gösterir. Ayrıca Bir alt düğümün GINI indeksi bir üst düğümün GINI indeksinden daha fazla olursa o dal başarılıdır. GINI indeksi sıfır olduğu zaman da ağaç dallanma işlemi bitmiş olur. Aşağıda GINI indeksinin formülüne edilmiş şekli yer almaktadır.

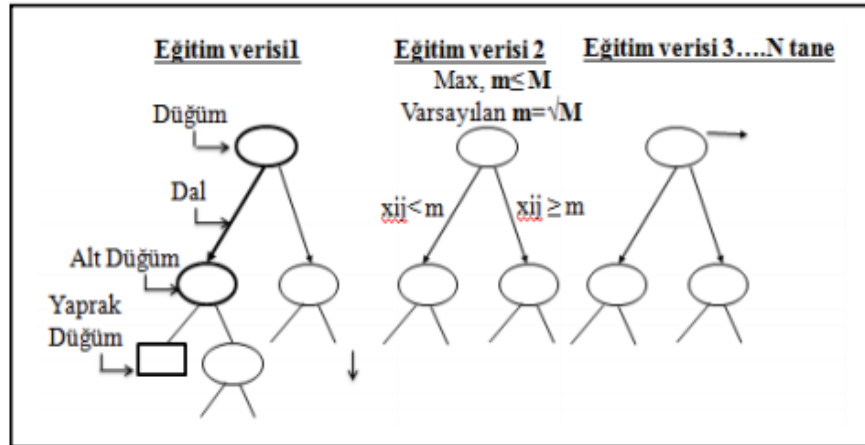
$$\sum \sum_{j \neq i} (f(C_i, T) / |T|) (f(C_j, T) / |T|) \quad (5)$$

$T$  : Tüm veri seti

$C_i$  : Rastgele seçilen bir pikselin ait olduğu sınıf

$f(C_i, T)$  : Seçilen örneğin sınıfına ait olma olasılığını gösterir [93, 97]

Kısacası belirlenen eğitim verileri kullanılması sonucu oluşan bölünme ölçütlerine göre düğümler dallara ayrılıp, ağaç yapıları oluşmaktadır. Şekil 4.8’de bölünme şekillerine göre oluşturulan ağaç yapıları gösterilmektedir [98, 99].



**Şekil 4.8.** Random Forest Yönetime Göre Ağaç Yapısı

#### 4.6.6. K-En Yakın Komşu Algoritması (k-NN)

Bu algoritma, makine öğrenme algoritmaları arasında popüler olarak kullanılmakta olup, analitik olarak izlenebilir olmakla beraber kolay ve kullanışlı, örtülü sınıflandırma yöntemlerinden birisidir. Avantajı olduğu kadar büyük miktarda bellek alanına ihtiyaç duyması, veri seti boyutu arttıkça buna bağlı maliyetin ve işlem yükünün de önemli derecede artması, performansın k komşu, öz nitelik, uzaklık ölçütü sayısı gibi parametre ve özelliklere bağlı olarak etkilenmesi de beraberinde bir takım dezavantajı getirmektedir [100, 101].

Öğrenme algoritmalarında, gerçekleşen öğrenme işlemi eğitim setinde tutulan verilere dayanmaktadır. Yeni gelen örnek de bu eğitim setinde tutulan örnekler ile karşılaştırılıp, aralarındaki benzerliğe göre sınıflandırılır. K-NN algoritmasında eğitim setinde yer alan örnekler n boyutlu sayısal nitelikler ile gösterilip, her örneğin n boyutlu uzayda bir noktayı temsil edeceği biçimde, tüm eğitim örnekleri n boyutlu bir uzayda saklanmaktadır. Daha sonra bilinmeyen bir örnek devreye girdiğinde algoritma, n boyutlu uzayı araştırarak bilinmeyen örnekleme en yakın olan k örnekleme bulmaktadır [102].

##### 4.6.6.1. K-NN Parametreleri

K-NN algoritması basit yapıya sahip olmasından dolayı az parametre gerektirmektedir. Komşu sayısı (k), uzaklık ölçütü ve ağırlıklandırma yöntemi, k-NN algoritmasının performansında önemli olan parametreleridir.

- **Komşu Sayısı (k):** Sınıflandırma boyunca  $k=1$  için, sadece en yakın komşunun bulunduğu sınıfa atanırken, k sayısı örnek sayısına (N) yaklaştıkça veri setinde yer alana tüm veriler dikkate alınmakta ve oylamaya göre seçim yapılmaktadır.
- **Ağırlıklandırma:** Komşular için ağırlık değerleri atanması ile sınıflandırılmakta olan örneğe daha yakın olan komşu örneklerin, çoğunluk oylamasına daha fazla katkı eklemesi hedeflenir. En çok kullanılan ağırlık değeri atama yöntemleri, her

bir komşunun ağırlığının,  $d$ , komşular arası uzaklık olmak üzere,  $1/d$  ya da  $1/d^2$  şeklinde alınmasıdır. K-NN algoritmasının performansında diğer önemli kritik nokta sise örnekler arası yakınlığın nasıl ölçümleneceğidir. Yakınlık, Öklid veya Manhattan, Minkowski, Chebyshev gibi uzaklık ölçütleri kullanılarak hesaplanabilmektedir.

- **Öklit Uzaklığı Formülü:** İki nokta arasındaki doğrusal uzaklık olup, P ve Q arasındaki Öklid uzaklığı  $P = (x_1, x_2, \dots, x_n)$  ve  $Q = (y_1, y_2, \dots, y_n)$  olmak üzere aşağıdaki gibi formülüne edilmektedir.

$$d = \left( \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \right) \quad (6)$$

Temel k-NN algoritmasında, sınıf etiketinin çoğunluk oylamasına dayalı olarak belirlenmesi, simetrik olmayan dağılıma sahip veri setlerinde sıklıkla görülen sınıfların, yeni örneklerin sınıf etiketlerinin belirlenmesinde daha baskın bir role sahip olmalarına neden olmaktadır. Bu sebeple, k-NN algoritmasının uzaklık ölçütünün etki değerine farklı biçimde ağırlık değeri atayan yöntemler bulunmaktadır [103].

#### 4.6.7. Yapay Zeka (Artificial Intelligence) ve Yapay Sinir Ağları (Neural Networks)

Literatürde "Artificial Intelligence" olarak da adlandırılan Yapay Zekâ biliminin amacı, makinelerin; planlama, öğrenme, algılama, iletişim kurma, nesnelere oynatabilme, yer değiştirme, geçmiş bilgilerden faydalanma yeteneğine sahip olmasını sağlamaktır. O halde yapay zekâ teknolojisi makinelerin, insanoğlunun düşünme fonksiyonlarının bazı yönlerini basitleştirip, elle yapılan karmaşık işlemleri otomatik şekilde yapmasına imkân sağlar şeklinde yorumlanabilir.

Yapay zekânın da veri madenciliği ve makina öğrenmesi gibi çerçevesi net bir tanımı olmayıp, özellikle son dönemlerde, yapay zekânın yeniden tanımlanması ile ilgili her geçen gün yeni bir bilgi ortaya konulurken, kimi araştırmacılara göre de bu bilgilerden yola çıkarak yapay zekâ tanımı giderek muğlak bir hal almaktadır. Çünkü yapılan çalışmalar cevaptan çok yeni sorular ortaya çıkarmaktadır. Yapay zeka, insanın düşünme yapısını anlamak ve bunun benzerini ortaya çıkaracak bilgisayar işlemlerini geliştirmeye çalışmak olarak tanımlanır [104].

Günümüzde yapay zekâ alanında önemli adımlar atılmış olmakla birlikte, araştırmacılar, akıllı bir şekilde hareket eden, robotlara makinelere ulaşma gayretinde olup, yapay zekânın geliştirilmesine ilişkin çalışmalarına devam etmektedirler. Her geçen gün gelişen yapay zekâ uygulamaları birçok şirket ve yatırımcı tarafından yakından takip edilmektedir. Bu uygulamalardan bazıları şu şekildedir [105];

- Konuşma Tanıma (Speech Recognition)
- Doğal Dil Üretimi (Natural Language Generation)
- Metin Analizi ve NLP (Text Analytics and NLP)
- Doğal dil işleme (NLP)
- Makine Öğrenme Platformları (Machine Learning Platforms)
- Sanal Ajanlar (Virtual Agents)
- Karar Yönetimi (Decision Management)
- Yapay Zekâ Optimize Donanım (AI-optimized Hardware)
- Derin Öğrenme Platformları (Deep Learning Platforms)
- Robotik Süreç Otomasyonu (Robotic Process Automation)
- Biyometri (Biometrics) [105].

Yapılan çalışmalardan da anlaşılacağı gibi yapay zekânın en önemli özelliği gerçek dünyanın ihtiyaçlarına göre uyarlanmasıdır. Buna istinaden insanlık adına faydası oldukça büyüktür. Aslında farkında olmadan insanlar hayatın her alanında rutin işlerini yürütürken yapay zekâ teknolojisi ile iç içedir. Her ne kadar kendinden daha üstün bir varlığın ortaya çıkma düşüncesi ilk başta insanoğlundaki endişe yaratsa da önemli olan insanoğlunun yapay zekâ teknolojisini hangi yönde kullanacağı sorusudur; İnsan hayatını kolaylaştırmak adına mı yoksa kötü amaçlar doğrultusunda

kullanılmak adına mı? Bana göre makineden ziyade tartışılması gereken, işin bu boyutudur. Yapay Zekâ insanlığın yararına kullanıldığı sürece, gelişen teknoloji ile birlikte uzun vadede herkesin sağlığını, yaşam şeklini, üretkenliğini, en önemlisi zaman kavramını çok daha verimli hale getirecektir [106,107].

1950'li yıllarda ortaya çıkmalarına rağmen, ancak 1980'li yılların ortalarında çok amaçlı kullanım için yeterli seviyeye ulaşmış sayısal tahmin yöntemlerinden birisi olan Yapay Sinir Ağları (YSA)'da bir Yapay Zekâ biliminin kolu olup, bilgisayarların öğrenmesine yönelik, araştırmacıların çok yoğun ilgi gösterdikleri bir araştırma alanıdır [106,107].

Birçok probleme çözüm üretebilecek yeteneğe sahip Yapay Sinir Ağları teknolojisi, insanlığın doğayı araştırma ve taklit etme çabalarının en son ürünlerinden bir tanesidir. Genel olarak YSA'lar, insan beyninin en temel özelliği olan öğrenme işlemini örnekler yardımı ile gerçekleştiren ve elde ettikleri bilgiler doğrultusunda kendi deneyimlerini oluşturup daha sonra benzer konularda çevreden gelen olaylara karşı nasıl tepkiler üreteceğini belirleyebilen bilgisayar sistemleridir [108].

#### **4.6.7.1. Yapay Sinir Ağlarının Yapısı**

Biyolojik sinir sisteminden esinlenerek geliştirilmiş yapay sinir ağı, birbirine hiyerarşik olarak bağlı ve paralel olarak çalışabilen yapay sinir hücrelerinin bir araya gelmesi sonucu oluşmaktadır. Yapay nöron ya da proses elemanları olarak da adlandırılan bu hücrelerin birbirine bağlandıkları ve her bağlantının bir değeri olduğu kabul edilmektedir. Öğrenme yolu ile elde edilen bilgi, proses elemanlarının bağlantı değerlerinde saklanmaktadır. Bu değerlere ise ağırlık adı verilmektedir. Gelen bilginin önemini ve hücre üzerindeki etkisini ağırlıklar göstermektedir. Proses elemanları birbirlerine paralel şekilde üç katman halinde bağlanarak yapay sinir ağlarını oluşturur [108].

Bu katmanlar;

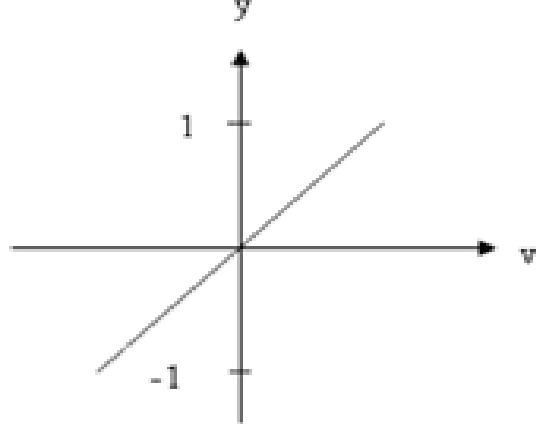
- Girdi Katmanı
- Ara Katmanlar
- Çıktı Katmanı

Ağın öğrenmesini isteyen örnekler tarafından belirlenen ve dışardan öğrenme yolu ile gelen bilgiler “Girdi Katmanı” ile ağa iletilir. Daha sonra, gizli katman olarak da adlandırılan Ara Katmanda ise; girdi katmanından iletilen bilgiler, sinir hücreleri tarafından bir “toplama fonksiyonu” ile toplanıp, ağa gelen net girdi hesaplanır. Hesaplanma, her gelen girdi değerinin kendi ağırlığı ile çarpılıp toplanması şeklindedir. Toplama fonksiyonu olarak tanımlanan ve aşağıdaki gibi formülize edilen bu fonksiyon, genellikle deneme yanılma yolu ile belirlenmekte ve her problem karşısında farklı formüller oluşturularak kullanılmaktadır [108, 109,115].

$$NET = \sum_i^n G_i A_i$$

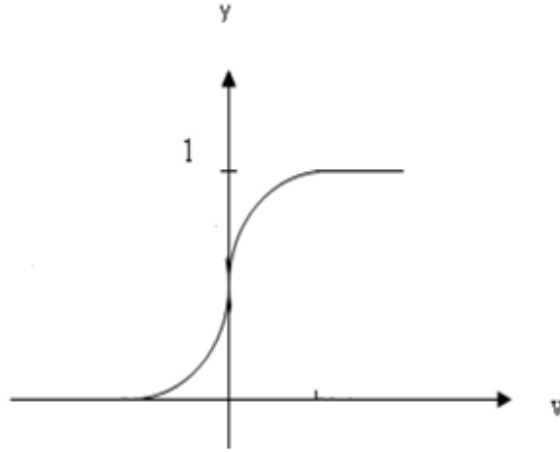
Ağa gelen net girdi hesaplandıktan sonra “aktivasyon fonksiyonu” tarafından işlenir. Aktivasyon fonksiyonu genellikle doğrusal olmayan bir fonksiyondur. Aktivasyon fonksiyonu seçilirken dikkat edilmesi gereken en önemli nokta ise hesaplamının yavaşlamaması için fonksiyonunun türevinin kolay hesaplanabilir olmasıdır. Aktivasyon fonksiyonları, bir yapay sinir ağlarında nöronun çıkış genliğini genellikle [0,1] veya [-1,1] değerleri arasında sınırlamaktadırlar. Bu fonksiyonun seçimi girdilere ve ağın neyi öğrenmesini istediğine göre değişmektedir. Günümüzde en çok “çok katmanlı algılayıcı modelinde” kullanılan “sigmoid fonksiyonudur”. Hücre modellerinde sıkça kullanılan aktivasyon fonksiyonlarından kısaca bahsedilecek olunursa [104, 108-110];

**1. Doğrusal (Lineer) Aktivasyon Fonksiyonu:** Doğrusal problemleri çözmek için seçilen doğrusal fonksiyonu hücrenin net girdisini doğrudan hücre çıkışı olarak verebilmektedir. Diğer bir deyişle toplama fonksiyonundan çıkan sonuç, belli bir katsayı ile çarpılıp hücrenin çıktısı olarak verilmektedir ( $y = A * V$  ( $A$  sabit sayı)). Şekil 4.9’da gösterilmiştir [109-110].



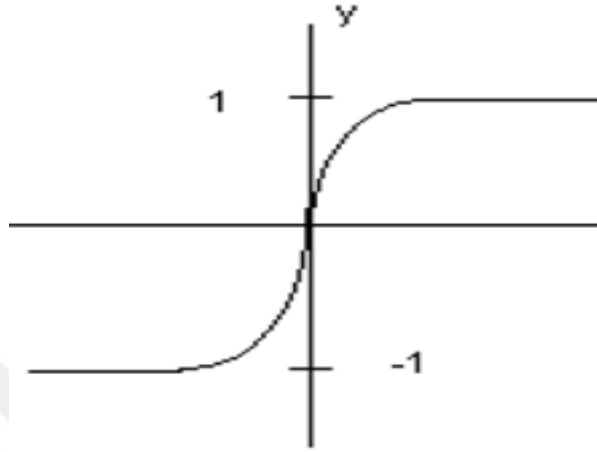
Şekil 4.9. Doğrusal Aktivasyon Fonksiyonu

**2. Sigmoid Aktivasyon Fonksiyonu:** Yapay sinir ağı uygulamalarında doğrusal olmayışı nedeni ile en çok kullanılan fonksiyondur ( $y = \frac{1}{1+e^{-v}}$ ). Sigmoid aktivasyon fonksiyonlarının türevi alınabilir ve süreklidir. Bu fonksiyon girdi değerinin her biri için 0 ile 1 arasında değer üretir [109-110].



Şekil 4.10. Sigmoid Aktivasyon Fonksiyonu

**3. Tanjant Hiperbolik Aktivasyon Fonksiyonu:** Sigmoid fonksiyonuna benzer bir fonksiyon olup, tek farkı çıkış değerleri -1 ile 1 arasında değişmektedir ( $y = \frac{1 - e^{-2v}}{1 + e^{2v}}$ ) [109-110].



**Şekil 4.11.** Tanjant Hiperbolik Aktivasyon Fonksiyonu

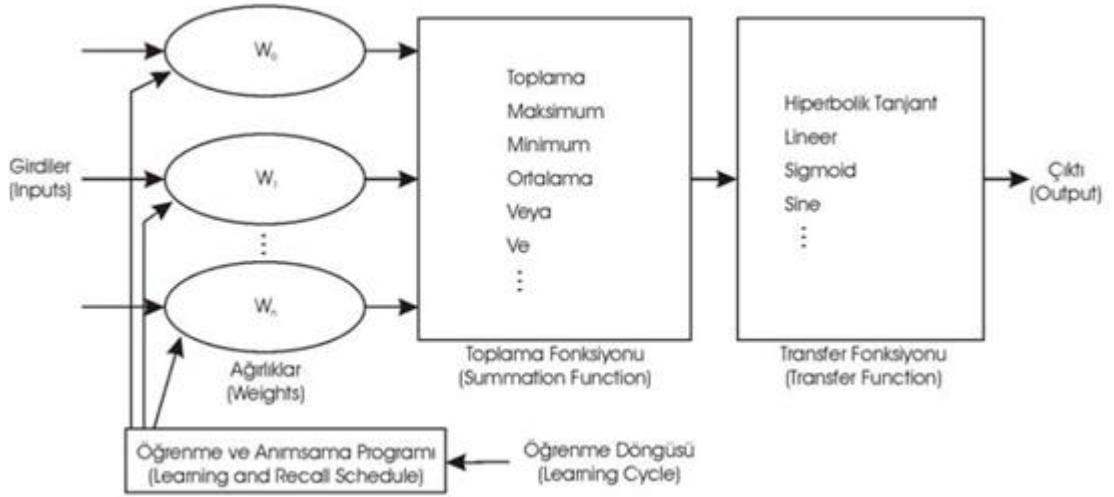
**4. Step (Adım) Aktivasyon Fonksiyonu:** Gelen net girdinin belirlenen eşik değerinin altında ve üstünde olmasına göre çıktı 1 ya da 0 alınır [109-110].

**5. Eşik Değer Aktivasyon Fonksiyonu:** Gelen bilgiler 0'dan küçük eşit ise 0 çıktısı, 1'den büyük eşit ise 1 çıktısı, 0 ile 1 arasında olduğunda yine kendini veren çıktılar üretilebilir [109-110].

**6. Sinüs Aktivasyon Fonksiyonu:** Öğrenilmesi tahmin edilen olayların sinüs fonksiyonuna uygun dağılım gösterdiği zaman kullanılır [109-110].

Girdi setine karşılık gelecek şekilde üretilen çıktı, başka bir hücreye, çıktı katmanına girerek dış dünyaya yâda kendisine girdi olarak gönderilir. Yapay sinir ağlarında bir sinir hücresinin görünüşte birden çok çıktısı olmasına rağmen aslında tek bir çıktı değeri bulunmaktadır. Şekil 4.12'de bir yapay nöronun yapısı gösterilmiştir [111].





Şekil 4.12. Yapay Nöronun Detaylı Yapısı

#### 4.6.7.2. Yapay Sinir Ağları Çalışma Şekli

Teknik olarak da bir Yapay Sinir Ağı'nın temel görevi ağı gelen girdi setine karşılık doğru çıktıyı üretebilmektir. Buna istinaden ağırlık değerlerinin doğru ayarlanması gerekmektedir. Bu işlemede “ağın eğitilmesi” adı verilir. Eğitim sırasında, ağırlık değerleri başlangıçta rastgele atanmaktadır. Daha sonra en doğru değerler bulununcaya kadar her örnek, ağı sunularak ağırlıklar öğrenme kuralına göre değiştirilir. Ağırlıkları kullanarak ağ, girdiler hakkındaki kararı verdiği için, ağın zekâsının bu ağırlıklarda saklandığı söylenebilir. Ağın sahip olduğu ağırlık değerleri ne kadar doğru ise ağın performansı da o kadar yüksek olur [108]. Her ağı eğitmede gerekli olan örnekler ise, birincisi ağı eğitmek (eğitim seti) için, ikincisi ağın performansını sınamak için (test seti) kullanılır. Ağ, test setindeki örneklere doğru cevaplar verirse eğitilmiş olarak kabul edilir. Ağın eğitimi tamamlandıktan sonra öğrenip öğrenmediğini (performansını) ölçmek için yapılan denemelere ise ağın test edilmesi denmektedir. Ağın hiç görmediği test setindeki örnekler ağı gösterilerek, ağ tarafından verilen cevaplara bakılır. Eğer doğru cevaplar verebiliyorsa bu ağın performansı iyi olarak kabul edilir. Eğer cevaplar yetersiz olursa tekrar eğitime tabi tutulur ve bu işlem eğitim setindeki örneklerin tamamı için doğru cevaplar bulununcaya kadar tekrarlanır [112].

Ağın bir olayı öğrenmesi YSA'ların yapısı için oldukça gereklidir. Bu durumda, o olay için en doğru yapay sinir ağı modelini seçmesi ile mümkündür. Şu zamana kadar çok fazla yapay sinir ağı modeli geliştirilmiştir. Bir yapay sinir ağının modelini sembolize eden bilgiler [108];

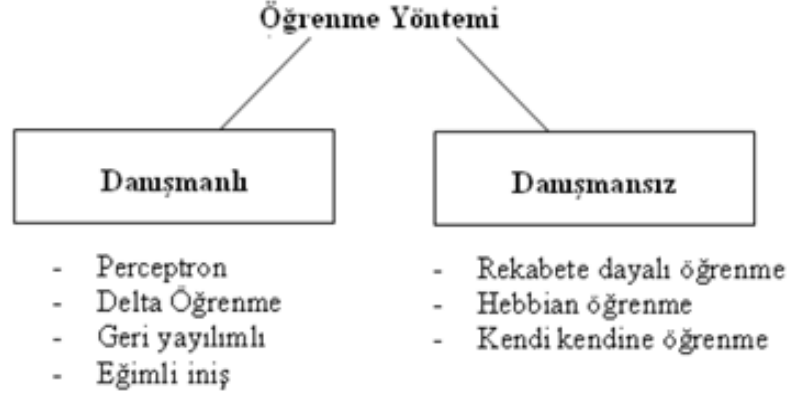
- Ağın topolojisi,
- Kullanılan toplama fonksiyonu,
- Kullanılan aktivasyon fonksiyonu,
- Öğrenme stratejisi,
- Öğrenme kuralı'dır.

Yapay sinir ağlarında en çok kullanılan modeller ise:

- Tek ve çok katmanlı algılayıcılar,
- Vektör kuantizasyon modelleri,
- Kendi kendini organize eden model,
- Adaptif rezonans teorisi modelleri,
- Hopfield ağları,
- Counterpropagation ağı,
- Neocognitron ağı,
- Boltzman makinesi
- Probabilistic ağlar,
- Elman ağı
- Radyal temelli ağlardır [108].

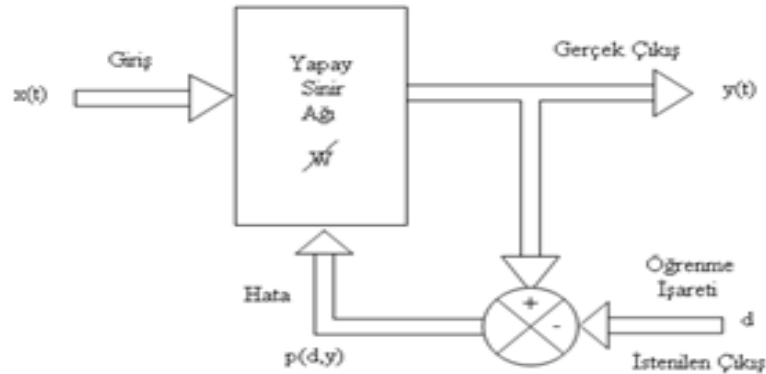
#### **4.6.7.3. Öğrenme Algoritmalarına Göre Yapay Sinir Ağlarının Sınıflandırılması**

YSA'da, Şekil 4.13'de de özet olarak gösterildiği gibi, “danışmalı (supervised)” ve “danışmasız (unsupervised)” olmak üzere iki tür öğrenme yöntemi bulunmaktadır [113].



**Şekil 4.13.** Öğrenme Yöntemleri

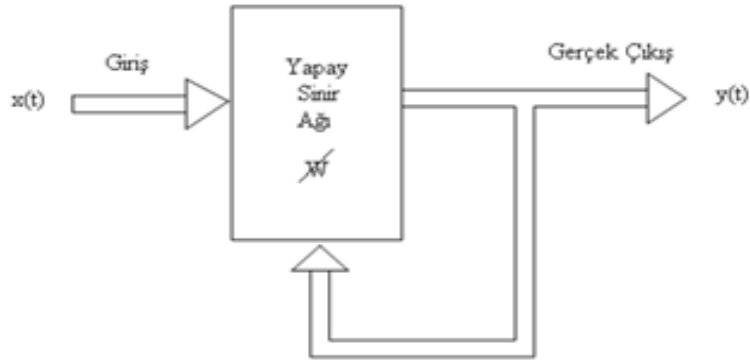
Danışmanlı öğrenmede, ağın eğitilmesi sırasında öncelikle ağa giriş ve çıkış bilgileri sunulur. Her bir girişe karşılık uygun çıktılar ağa sağlanmalıdır. Kısacası, ağın girdiye uygun çıktı değerleri verebilmesi için, girdi-çıkı ilişkisini elde edebilmesini sağlayacak bir mekanizma olmalıdır. Bağlantı ağırlıkları belirleninceye kadar aynı veri seti defalarca işlemde geçirilip, tekrarlanır. Danışmanlı öğrenmenin performans kriterini ise “hata terimi” belirler. Öğrenme işlemi gerçekleştikçe bu hata terimi azalmaktadır ve minimum olduğu değer ise iterasyonun durdurulması gereken andır [114]. Şekil 4.14’de danışmanlı öğrenme modeli gösterilmiştir [110]. Genellikle, kullanılan ağların büyük çoğunluğu danışmanlı öğrenmeyi tercih etmektedirler.



**Şekil 4.14.** Danışmanlı Öğrenme Yapısı

Eğer girdi verileri çıktıya uygun bilgi taşıyorsa, ağlar öğrenme işlemini hiç yapamayabilir. İdeal olan geniş bir veri seti alıp, veri setinin bir bölümü ile eğitimin, geri kalanıyla da gerekli testlerin yapılmasıdır. Eğer ilgili problem ağ tarafından çözülemezse ilk olarak ağın yapısını ya da kullanıcının seçimine ve yaratıcılığına bağlı olan eğitime kurallarını gözden geçirmek gerekir [109]. Eğitime esnasında ağırlıkların ayarlanabilmesi için gerekli adaptif geri beslemeyi sağlayacak çok fazla farklı eğitime kuralı (algoritma) vardır. Geriye doğru hata beslemesi ya da diğer ismiyle geri yayılma en yaygın tekniktir [111]. Eğitime işlemi, gerçekleştirildikten sonra ise ağırlık sabitlenir.

Danışmansız öğrenme ise herhangi bir eğitici olmaksızın, ağ için sadece giriş bilgileri yeterli olup eğitici bilgiye ihtiyaç duyulmadan, girdilerin ağ tarafından analiz edilmesi ve bu analiz sonucunda bağlantıların oluşturulmasıdır. Kısacası sistem, veriye göre hangi özellikleri kullanacağına kendi kendine karar vermektedir. Şekil 4.15’de danışmansız öğrenme modeli yer almaktadır [104,109-110,113].



Şekil 4.15. Danışmansız Öğrenme Modeli

Yapay Sinir Ağları’nda ayrıca ağın öğrenme sürecinde, ağın yapısı için büyük önem taşıyan birtakım elemanlar bulunmaktadır. Bu elemanlar;

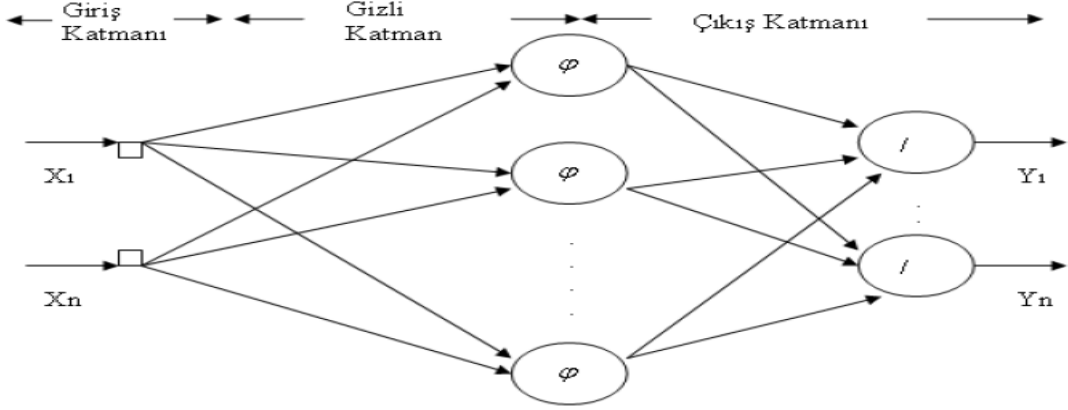
- **Öğrenme Fonksiyonu:** Bu fonksiyon girdi değerlerine karşı, o girdilere ait değişken bağlantı ağırlıklarını ayarlamaktadır.

- **Hata Fonksiyonu:** Bu fonksiyon o anki çıktı ile istenilen çıktı arasındaki farkı hesaplayıp, gerekirse transformasyon uygular. Bu hata hali (geri yayılma değeri) önceki tabakaya geri yayılarak, bir sonraki öğrenme esnasında, öğrenme fonksiyonunun bağlantıları ayarlaması için, gerekli ise kullanılmaktadır.
- **Öğrenme Oranı:** YSA'nın öğrenme gücü ile hızı ters orantılıdır. Diğer bir ifade ile daha fazla öğrenme daha az hız ve buna bağlı daha fazla zaman demektir. Bir ağın ne kadar eğitileceğini öğrenme oranı belirlemektedir. Öğrenme oranı ne kadar küçük olursa öğrenme süreci de o kadar yavaş gerçekleşecektir. Bu durumda beraberinde maksimum doğruluk derecesine yakınsamayı getirir [111].

#### 4.6.7.4. YSA Çeşitleri

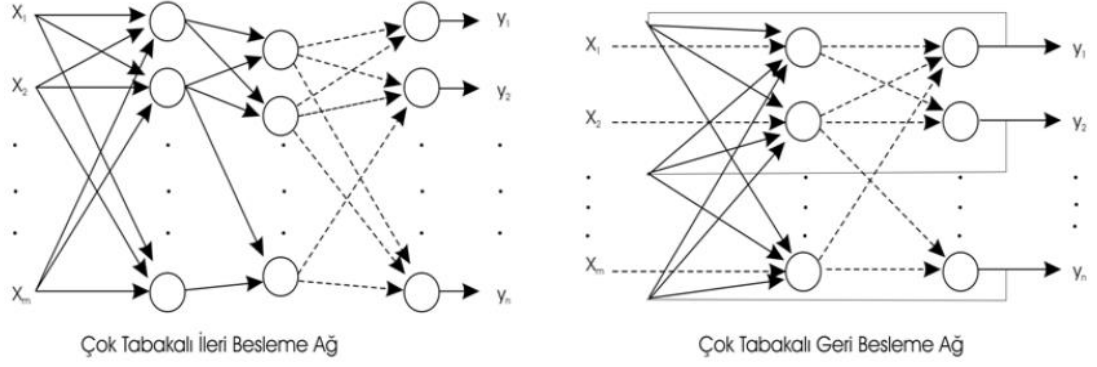
Genel olarak; mimarisi, öğrenme yöntemi, bağlantı yapılarından dolayı YSA'lar üç şekilde sınıflandırılır. Bunlardan ilk sınıflandırma ağın eğitime kuralına göre, danışmanlı ve danışmansız olarak adlandırılan öğrenme algoritmalarıdır. İkincisi ise kalitatif ve kantitatif olmak üzere ağın kullandığı veriye göre yapılan sınıflandırmadır. İster danışmanlı, ister danışmansız eğitime kullansın, kalitatif veriler ile çalışan ağlar sınıflandırma ağları; danışmalı eğitime kullanan kantitatif veriler ile çalışan ağlar ise regresyon olarak adlandırılmaktadır. Son Sınıflandırma ise ağın yapısına göre olan sınıflandırmadır. Bunlar ileri besleme (feedforward) ve geri besleme (feedback) şeklindedir [111]. İleri besleme, tek katman ileri besleme ve çok katman ileri besleme olarak iki şekildedir. İleri beslemede giriş katmanındaki her sinir hücresi, sonucu etkilemesi istenen bir değişkene hızlı şekilde karşılık gelmektedir. Döngüsel bir durum olmayıp işaretler girdi katmanından çıktı katmanına kadar tek yönlü bağlantılar ile iletilir. Çok Katmanlı Algılayıcılar (Multilayered Perceptrons) ve Öğrenme Vektör Nicelendirmesi (Learning Vector Quantization) ağları ileri beslemeli ağlara örnek olarak verilebilir. Özellikle çok katmanlı ağ yapısında giriş ve çıkışların dışında arada gizli katmanlar yer almaktadır. Ağda yer alan girişler öncelikle, ilk gizli katmana bağlanır, gizli katmanın çıkışları ise kendisinden sonra gelen diğer gizli katmanın girişlerini oluşturmaktadır. Sistem

ağ çıkışına kadar bu şekilde devam eder. Şekil 4.16'da ileri beslemeli çok katmanlı yapay sinir ağı gösterilmiştir [104-109].



Şekil 4.16. İleri Beslemeli Çok Katmanlı Yapay Sinir Ağı

Geri beslemede ise genellikle danışmansız öğrenme kurallarının uygulandığı ağlarda kullanılıp, bağlantılar döngü içermekte ve her seferinde yeni veri setini kullanmakta bu yüzden girdiye daha yavaş karşılık alınmaktadır. Geri besleme hem bir katmandaki hücreler arasında hem de katmanlar arasındaki hücreler arasında olabilir. Bu yapı ile doğrusal olmayan dinamik bir davranış gösterir. En az bir hücrenin çıkışı kendisine ya da diğer hücreye giriş olarak verilmektedir. Bir andaki çıkış hem o an hem de önceki girişi yansıtabildiği için özellikle önceden tahmin uygulamaları için kullanılır. Hopfield, Elman ve Jordan ağları geri beslemeli ağlara örnek olarak verilebilmektedir [109]. İleri ve geri ağ yapıları Şekil 4.17'de gösterilmiştir [111].



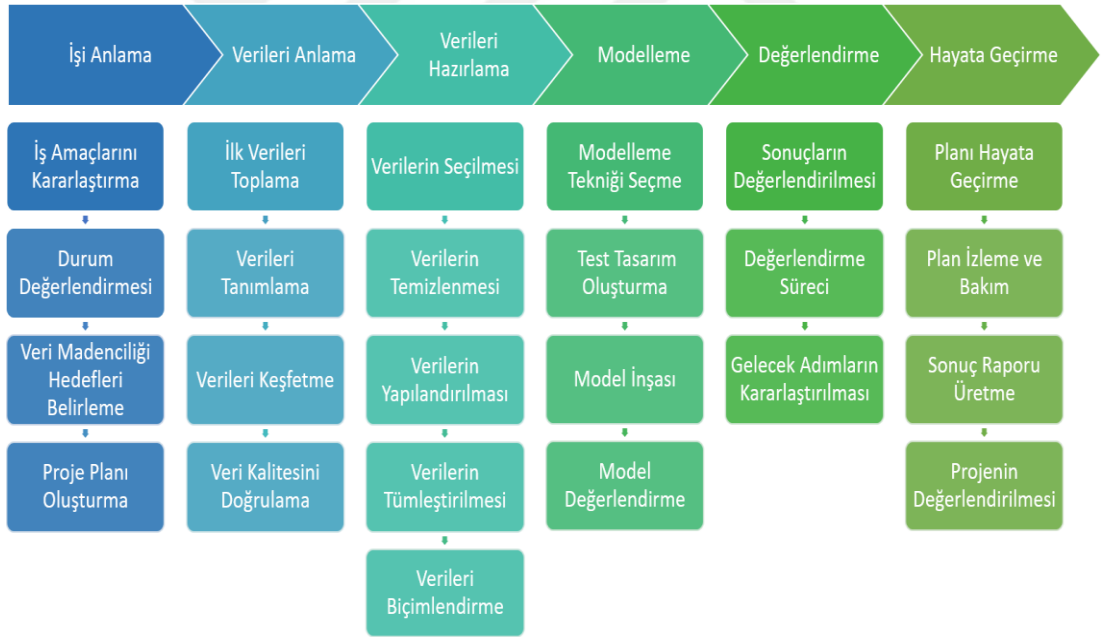
**Şekil 4.17.** İleri Beslemeli ve Geri Beslemeli Ağ Yapıları

#### 4.6.7.5. YSA'nın Tercih Edilmesinin Nedenleri

YSA'lar çeşitli özellikleri sayesinde çok fazla avantajlar sunmaktadır. YSA'ları doğrusal olmayan yapıları göz önünde bulundurması en önemli özelliklerinden birisidir. Böylece veri seti doğrusal olmayan yapı içerse de analizler gerçekleştirilebilmektedir. Diğer bir avantajı ise YSA'nın öğrenme özelliğine sahip olmasıdır. Böylece karmaşık problemlere çözüm sağlayabilmekte, benzer olaylar karşısında genelleme yapabilmektedirler. Ayrıca bağlantıdaki ağırlıklarda bilgi saklayabildikleri için bu bilgilere göre kendi ilişkilerini oluşturup herhangi bir denklem içermemektedirler. Ayrıca çok fazla değişken ve parametre ile çalıştıkları için genellemeler yaparak iyi bir öngörüde bulunabilirler. Belirsiz ve karmaşık bilgileri bile işleyip, analiz ettiği bilgi kümesi için uzman olarak düşünebilirler [111].

## 5. YÖNTEM

Veri madenciliği yöntemleri birçok araştırmacı tarafından tahmin amaçlı kullanılmaktadır. Veri madenciliği teknikleri altındaki sınıflandırma ve değerlendirme, eğitim verilerinin oluşturulmasına, tahmin modelinin sınıflandırılmasına ve ayrıca sınıflandırma verimliliğinin test edilmesine yardımcı olur [116]. Kurulan modellerin gelecekte elde edilen veri setlerine uyarlanabilmesini sağlanabilir. Çalışmada veri madenciliği için CRISP metodolojisi uygulanmıştır. Sürecin güvenilir ve standart olabilmesi için veri madenciliği CRISP metodolojisi adım adım prosedürler önermektedir. Bu metodoloji; işi anlama, veriyi anlama, veri hazırlığı, modelleme, değerlendirme ve uygulamayı içeren döngüsel bir süreci izlemektedir [117]. Şekil 5.1’de Crisp-Dm işleyiş modeli gösterilmiştir [65].



Şekil 5.1. CRIPS-DM İşleyiş Modeli

Bu çalışmada veri seti olarak Kırıkkale Üniversitesi Eğitim ve Araştırma Hastanesi’nde toplanan 2657 klinik kaydı kullanılmıştır. Bu verilerde, diyabet

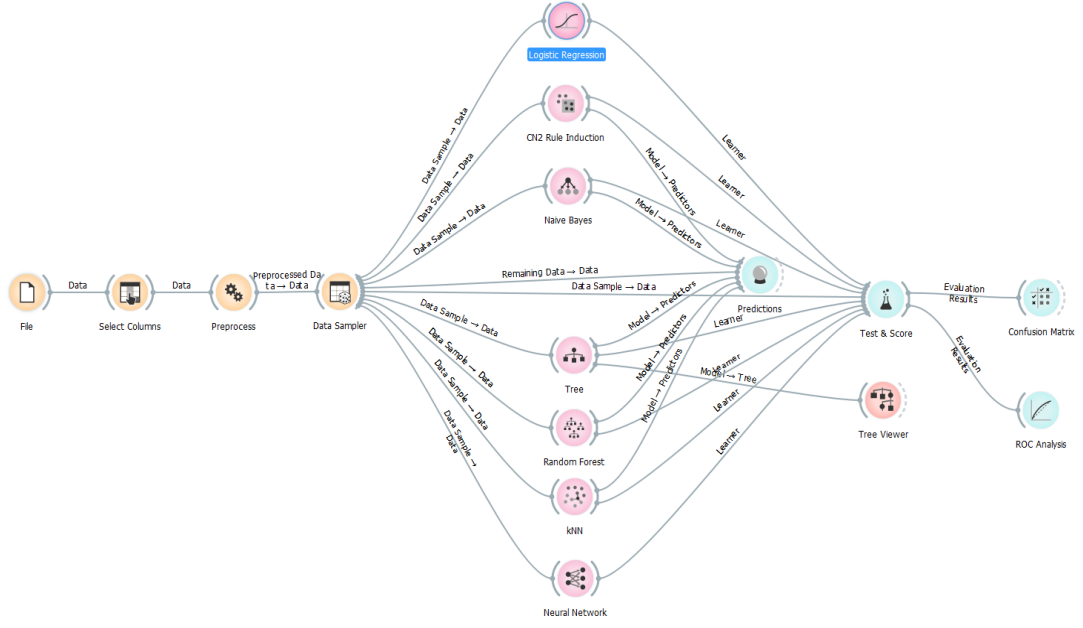


şüphesiyle gelen hastaların cinsiyetleri, yaşları ve diyabet hastalığıyla ilgili olan laboratuvar göstergelerinin yer aldığı kayıtlar bulunmaktadır. Elde edilen bu verilerden yola çıkarak hastanın diyabet olup olmadığı tahmin edilmesi amaçlanmıştır.

Daha öncede bahsedildiği gibi veri madenciliğinde modeller, tahmin edici ve tanımlayıcı olmak üzere iki ana başlık altında toplanmaktadır. Tahmin edici modellerde, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmekte ve bu modelden yararlanılarak sonuçları bilinmeyen veriler için sonuç değerlerin tahmin edilmesi amaçlanmaktadır [118]. Tahmin edici modellerde amaç geleceği öngörecek çıkarımların yapılmasıdır.

Tanımlayıcı modellerde ise karar vermeye rehberlik etmede kullanılacak mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır [119].

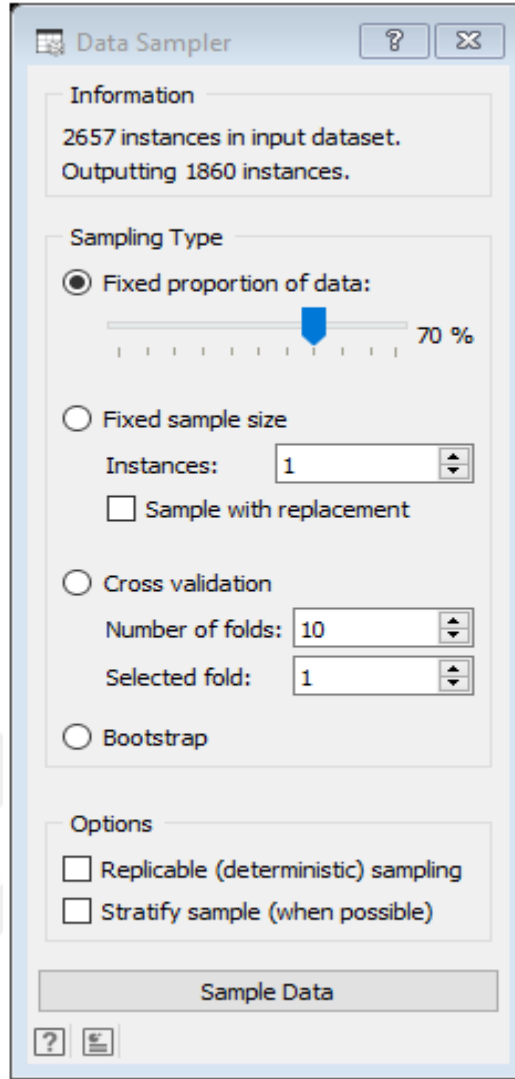
Yapılan çalışmada, Orange programı kullanılarak k-NN, Karar ağacı, Random Forest, Yapay sinir ağları, Naive Bayes, Logistic Regression ve CN2 rule inducer makine öğrenmesi algoritmalarıyla tahmini yapılmış ve söz konusu yöntemlerin doğruluğu karşılaştırılmıştır. Oluşturulan modellerde verinin bir kısmı eğitim verisi olarak kullanılırken bir kısmı da modelin doğruluğunu test etmek için kullanılmıştır. Önerilen mimari Şekil 5.2' de görülmektedir.



**Şekil 5. 2. Önerilen Sistemin Mimarisi**

Şekil 5.2’de görülen mimariye göre toplanan veriler ön işleme tabi tutulmuştur. Verilerin kalitesi büyük ölçüde tahminin sonucunu etkiler. Bu durum, verileri ön işlemenin modelde önemli bir rol oynadığı anlamına gelmektedir [120].

Ön işleme aşamasında eksik veriler temizlenmiştir. Bağımlı değişkenin belirlenmesinin ardından veriler eğitim ve test verileri olmak üzere ikiye ayrılmıştır. Verilerin %70’i eğitim verileri olarak kullanılırken %30’u modeli test etmek için kullanılmıştır. Bunun için Şekil 5.3’te gösterilen, Data Sampler modülünden yararlanılmıştır.



Şekil 5. 3. Data Sampler Modülü

Her kayıt için verinin ismi, veri tipi ve tanımı Çizelge 5.1’de, gösterilmektedir.

**Çizelge 5.1. Dataset Description**

<b>Name</b>	<b>Type</b>	<b>Description</b>	<b>Role</b>
<b>RESULT</b>	Categorical	0(no diabetes) / 1(diabetes)	Target
<b>GENDER</b>	Categorical	1(Female) / 2(Male)	Input
<b>AGE</b>	Numerical	Age: Range is between 3 and 91 years.	Input
<b>CREATININE</b>	Numerical	A compound that is produced by metabolism of creatine and excreted in the urine.	Input
<b>PROXIMAL UREA</b>	Numerical		Input
<b>eGFR</b>	Numerical		Input
<b>LDL</b>	Numerical	The form of lipoprotein in which cholesterol is transported in the blood.	Input
<b>TRIGLYCERIDE</b>	Numerical	An ester formed from glycerol and three fatty acid groups. Triglycerides are the main constituents of natural fats and oils, and high concentrations in the blood indicate an elevated risk of stroke.	Input
<b>SODIUM</b>	Numerical		Input
<b>POTASSIUM</b>	Numerical		Input
<b>ALBUMIN</b>	Numerical	a simple form of protein that is soluble in water and coagulable by heat, such as that found in egg white, milk, and (in particular) blood serum.	Input

Veri setinden bir kesit ise Çizelge 5.2’de gösterilmiştir. RESULT değişkeninde 1 diyabet olanları, 0 diyabet olmayanları ifade etmektedir. Diğer verilerde kan numunelerindeki değerlerdir.

**Çizelge 5.2.** Samples of Dataset

RESULT	GENDER	AGE	CREATININE	PROXIMAL UREA	eGFR	LDL	TRIGLYCERIDE	SODIUM	POTASSIUM	ALBUMIN
1	1	58	0,7	21	87,2	117	117	141	4,96	4,44
1	1	68	0,5	22	124,22	122	142	139	3,96	9,41
1	1	46	0,56	23	118,72	101	656	140	4,48	0.68
1	2	66	2,04	85	33,15	141	138	147	6,11	53.86
1	2	80	0,9	45	82,04	132	205	147	4,63	1,09
1	1	52	0,51	37	128,71	101	83	136	4,24	1,45
0	2	49	0,9	37	96,13	88,6	182	143	4,98	4,82
0	2	49	0,77	33	115,1	78,4	193	143	4,81	5,1
0	2	49	0,82	21	107,04	51,8	191	141	4,6	4,8
0	2	23	0,79	22	131,6	100	125	141	4,6	4,93

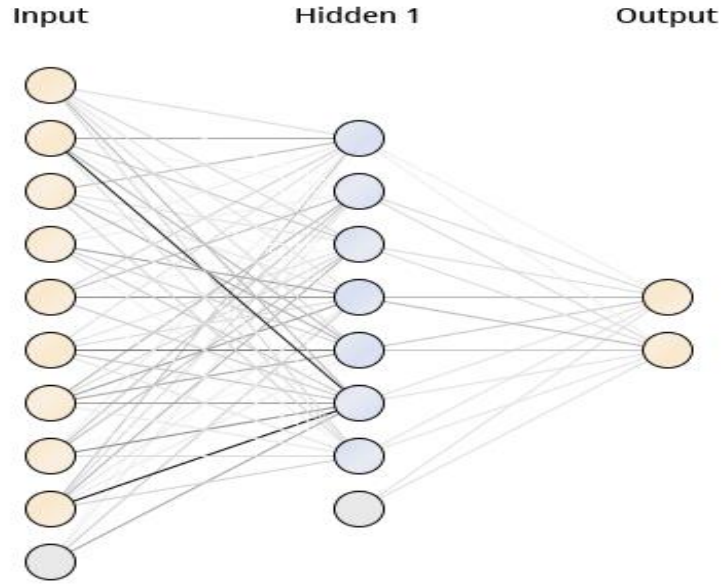
## 5.1. Bulgular

Çizelge 5.3'e göre seçilen algoritmalar içinde yapay sinir ağı diğer tahmin yöntemlerine göre en doğru sonucu vermektedir.

**Çizelge 5.3.** Veri Seti Kullanılan Modellerin Doğruluk Değerleri

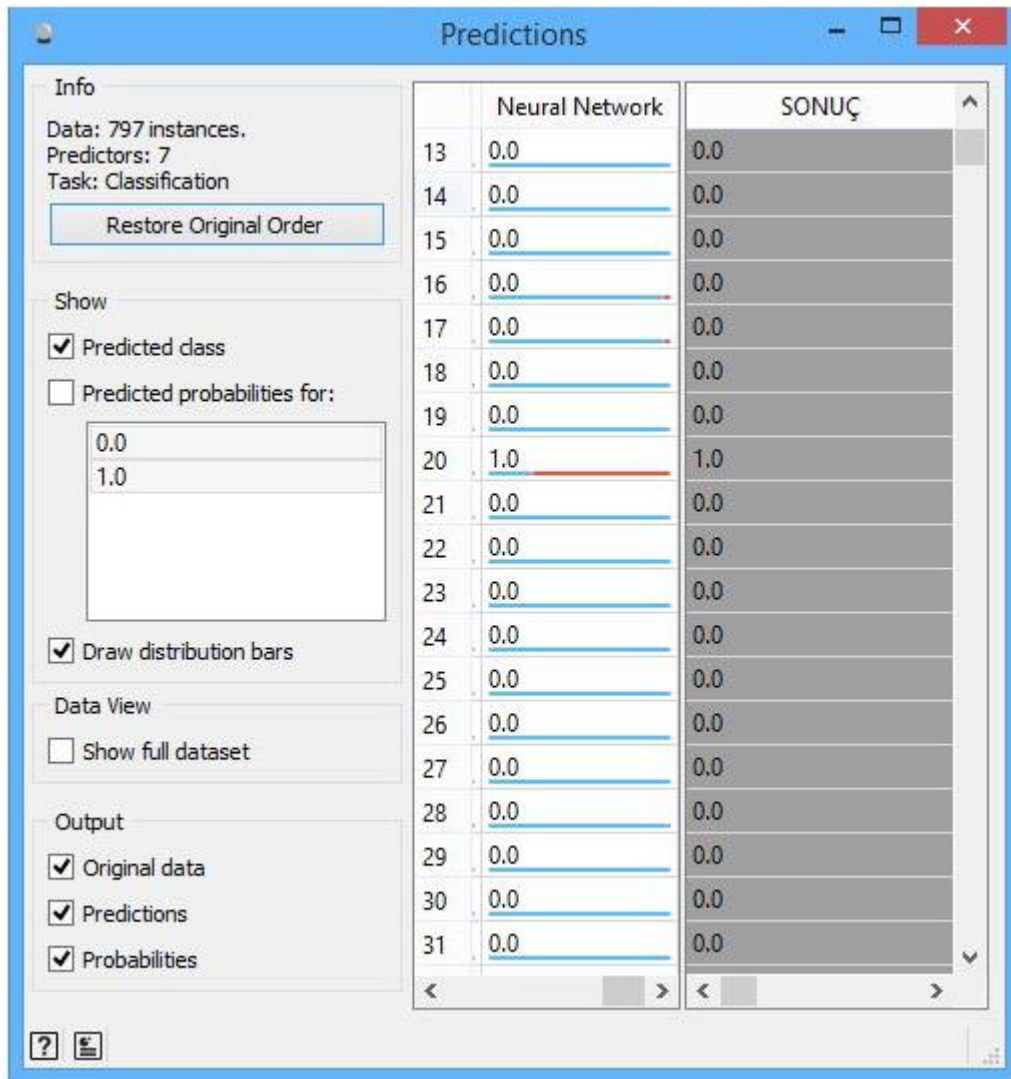
<b>Method</b>	<b>Accuracy (%)</b>
<b>kNN</b>	93.3
<b>Tree</b>	95.2
<b>Random Forest</b>	96
<b>Neural Network</b>	97.2
<b>Navie Bayes</b>	95.1
<b>Logistic Regression</b>	96.9
<b>CN2 rule inducer</b>	94.1

Önceki bölümlerde detaylı şekilde anlatıldığı gibi, Yapay Sinir Ağları (YSA), yapay zekâ ve insan beyninin biyolojik tepkisi kavramlarına dayanan öğrenme ilkesine dayanan matematiksel modellerdir. Ayrıca, YSA bilgi sınıflandırmak, modellemek ve tahmin etmek için yapı sistemleri süreçlerine katılır. YSA, bir öğrenme ve genelleme süreci ile ayarlanabilen veya eğitilebilen sistemin birimleri olan işleme elemanları veya nöronlara sahiptir. Bu ağda, bağlantılar sadece bir yönde ilerler, bu döngüleri ve tekrarlayan bağlantıları engeller [121]. Verideki değişkenlerin her birinin bir nöronu temsil ettiği durum Şekil 5.4'te görülmektedir.



**Şekil 5.4.** Artificial Neural Networks

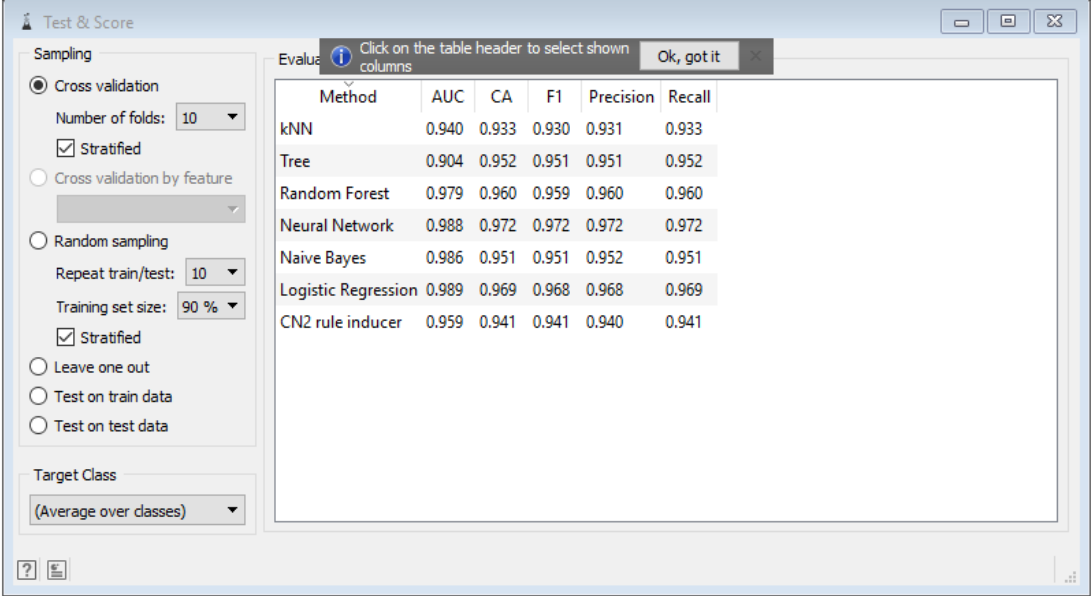
Şekil 5.5'te yapay sinir ağı yöntemi sonuçları yer almaktadır. Sonuç sütunundaki 0 ve 1 değerleri hastaya ait gerçek sonuçlar gösterirken, yapay sinir ağı sütunundaki değerler program tarafından tahmin edilen sonuçları göstermektedir. Burada “0” değeri hastalığın olmadığını gösterirken “1” değeri hastalığın var olduğunu yansıtmaktadır.



Şekil 5.5. Results of Artificial Neural Network Analysis



Şekil 5.6’da yer alan Test&Score Modülüyle de sonuçlar gösterilecek olunursa, modelin başarısı buradan da çıkarılabilmektedir.

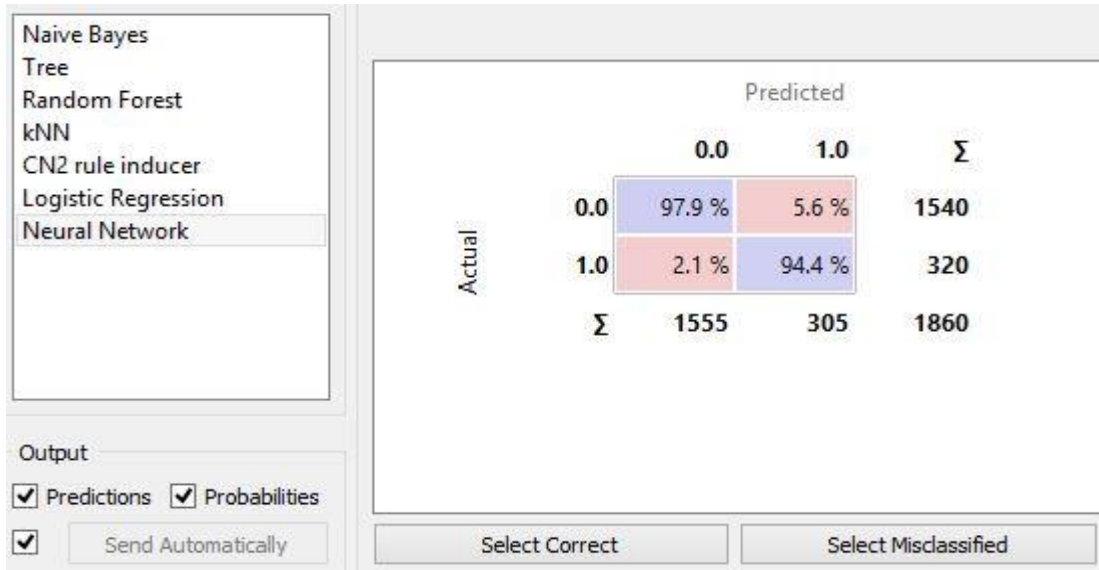


The screenshot shows the 'Test & Score' module interface. On the left, there are settings for 'Sampling' and 'Target Class'. The 'Sampling' section includes options for 'Cross validation' (selected), 'Number of folds' (10), 'Stratified' (checked), 'Cross validation by feature', 'Random sampling', 'Repeat train/test' (10), 'Training set size' (90%), and 'Stratified' (checked). The 'Target Class' is set to '(Average over classes)'. The main area displays a table of evaluation metrics for various methods. A tooltip above the table reads 'Click on the table header to select shown columns' and 'Ok, got it'.

Method	AUC	CA	F1	Precision	Recall
kNN	0.940	0.933	0.930	0.931	0.933
Tree	0.904	0.952	0.951	0.951	0.952
Random Forest	0.979	0.960	0.959	0.960	0.960
Neural Network	0.988	0.972	0.972	0.972	0.972
Naive Bayes	0.986	0.951	0.951	0.952	0.951
Logistic Regression	0.989	0.969	0.968	0.968	0.969
CN2 rule inducer	0.959	0.941	0.941	0.940	0.941

Şekil 5.6. Test Score Modülü

Yine Şekil 5.7’de gösterilen, her bir modelin başarısını değerlendirebilen tablo olan karmaşık matrisinin verildiği kısımda, sonuçları daha ayrıntılı yorumlamak mümkün olabilmektedir.








Şekil 5.7. Confusion Matris

## 6. SONUÇLAR

Uluslararası diyabet federasyonu 2015 verilerine göre dünya üzerinde yaklaşık her 12 kişiden birinde diyabet olduğu, diyabeti olan 2 kişiden birinin bunun farkında olmadığı, her 7 saniyede 1 kişinin diyabet nedeniyle hayatını kaybettiği, sağlık harcamalarının 1/9'unun diyabet için yapıldığı, şu an dünyada % 46,3'ü teşhis edilmemiş 387 milyon diyabetli olduğu, bu oranın dünyada % 8,3, Türkiye'de %14,8 olduğu, 2035 yılına kadar 205 milyon civarında diyabet hastası artışı olacağı ifade edilmektedir [124].



**Çizelge 6.1.** Diyabet Komplikasyonları

<b>Komplikasyonlar</b>	<b>Bölge</b>	<b>Açıklama</b>
Makrovasküler		Yeni tedavi edilen tip 2 diyabet hastalarındaki inme riski, genel popülasyondaki riskin iki katından fazladır[125].
		Diyabeti olan kişilerde kardiyovasküler hastalık gelişmesi, diyabeti olmayan kişilere kıyasla iki ila altı kat daha olasıdır[126].
Mikrovasküler		Böbreklerin filtreleme sistemlerinde diyabetin neden olduğu hasar (diyabetik nefropati), böbrek yetmezliğinin önde gelen nedenlerinden biridir[127].
		Retinada diyabetin neden olduğu mikrovasküler hasar (diyabetik retinopati), körlüğün önde gelen nedenlerinden biridir[128].
		Diyabetin sinirlere verdiği hasar (diyabetik nöropati), sıklıkla ayak ve bacak amputasyonuna neden olan ayak yaraları ve ülserlerinin önde gelen nedenlerinden biridir[129].

Diyabetin dünyada ve ülkemizde artış oranı, ekonomik maliyeti ve insan vücuduna verdiği zararlar dikkate alındığında diyabet üzerinde yapılan çalışmaların değeri daha iyi anlaşılmaktadır.

Bu tez çalışmasında, yüksek doğrulukta diyabetik ve diyabetik olmayan bireyleri tanımlamak için son derece hassas bir makine öğrenme tahmin aracı tanımlanmaktadır. Bu yöntem, hastanelerde veya diyabet önleme programlarında geniş çaplı tarama için kullanılabilir. Çizelge 6.2’de araştırmada kullanılan yöntemlerin doğruluk değerleri yer almaktadır. Test edilen makine öğrenmesi algoritmalarından Neural Network algoritmasının diğerleri ile karşılaştırıldığında %97.2’lik oranla en iyi sınıflandırma doğruluğuna sahip olduğu tespit edilmiştir. Logistic Regression algoritması %96,9’luk oranla ikinci yüksek doğruluk oranına sahiptir. %96 doğru tahmin oranı ile Random Forest algoritması 3. Sırada yer almaktadır.

**Çizelge 6.2.** Kullanılan Yöntemlerin Doğruluk Değerleri

<b>Metot</b>	<b>Doğruluk (%)</b>
<b>Neural Network</b>	97.2
<b>Logistic Regression</b>	96.9
<b>Random Forest</b>	96

Çalışma sonucu elde edilen %97,2’lik doğru tahmin yüzdesi oldukça yüksek bir değere sahiptir. Yapay sinir ağı yöntemi, sınıflama, kümeleme, tahmin, optimizasyon ve arama çalışmaları gibi farklı amaçlarla otomotiv, bankacılık, uzay, elektrik, sağlık, finans, sigorta, askeri ve endüstri gibi birçok alanda kullanılmaktadır.

Bu çalışmada kullanılan deęişken ve yöntemlerin diyabet araştırma ve tedavi merkezlerinde teşhis yöntemi olarak kullanılması bu alanda çalışan kişilere büyük kolaylık sağlayacaktır.

Sonraki çalışmalarda farklı deęişkenler ve farklı algoritmalar ile doğru tahmin başarı yüzdesi artırılabilir. Ayrıca, bu çalışmada uygulanan modelin masaüstü, web tabanlı veya mobil cihazlara uygun olmak üzere uygulamaları hazırlanabilir. Özellikle web tabanlı uygulamalar ile farklı bölgelerde bulunan sağlık çalışanları aynı platform üzerinden elde ettikleri yararlı sonuçları paylaşabilmektedir.



## KAYNAKLAR

- [1] Anonim, Diyabet Hakkında Herşey. Türkiye Diyabet Vakfı.  
<http://www.turkdiab.org/diyabet-hakkinda-hersey.asp?lang=TR>  
(Erişim Tarihi: 20.04.2018)
- [2] Turhan, H., Tip 2 Diabetes Mellitus' lu Hastalarda Tedavi Şekline ve Hastalık Süresine Göre Depresyon ve Anksiyete. Yayımlanmış Uzmanlık Tezi. Taksim Eğitim ve Araştırma Hastanesi, İstanbul, 2007.
- [3] Anonim, Diyabet Oranı 10 Yılda Yüzde 100 Arttı. Türkiye Diyabet Vakfı.  
[http://www.turkdiab.org/admin/PICS/webfiles/TURKIYE\\_DIYABET\\_VAKFI\\_BASIN\\_BULTENI\\_2017.pdf](http://www.turkdiab.org/admin/PICS/webfiles/TURKIYE_DIYABET_VAKFI_BASIN_BULTENI_2017.pdf) (Erişim Tarihi: 20.04.2018)
- [4] Mohamed, M. Y., Diyabet Tahmini İçin Sınıflandırma Algoritmalarının Karşılaştırılması. Yüksek Lisans Tezi. Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Kayseri, Şubat 2015.
- [5] Sullivan, R., Introduction to Data Mining For the Life Sciences. Humana Press (Springer), ISBN 978-1-58829-942-0, Ohio, 600 pp, 2012.
- [6] Kondiloğlu, A., Diyabet Akıllı Destek Sistemi: Bilgi Erişim Teknikleri ile Kan Şekeri Seviyesinin Tahmini. Yüksek Lisans Tezi. Türk Hava Kurumu Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, Mart 2015.
- [7] Kuloğlu, M., Karaoğlu, A., Atmaca, M., Özkan, Y., Tezcan, A. E., Tip II Diyabetik Hastalarda Psikiyatrik Belirtiler ve Kan Şekeri Kontrolü. Düşünen Adam, 13 (1), 19-23, 2000.
- [8] Çiftçi, H., Akbulut, G., Yıldız, E., Mercanlıgil, S. M., Kan Şekerini Etkileyen Besinler. Hacettepe Üniversitesi, Sağlık Bilimleri Fakültesi Beslenme ve Diyetetik Bölümü, Ankara, Şubat 2008.  
<http://sbu.saglik.gov.tr/Ekutuphane/kitaplar/t68.pdf> (Erişim Tarihi: 21.04.2018)

- [9] Pala, T. ve Yücedağ, İ., Veri Madenciliği Tekniklerinden Sınıflandırma Kullanılarak Tip 2 Diyabet Tanısı. International Artificial Intelligence and Data Processing Symposium, Eylül 2016, İnönü Üniversitesi, Malatya, 2016.
- [10] Sürmeli, K., Diyabet Tedavisi Gören Hastaların Hastalık ve Tedavileri Hakkındaki Bilgi Düzeylerinin Değerlendirilmesi. Bitirme Ödevi. Erciyes Üniversitesi, Eczacılık Fakültesi, Kayseri, Mayıs 2012.
- [11] Anonim, Who Consultation. Definition, Diagnosis and Classification of Diabetes Mellitus and Its Complications. Part 1: Diagnosis and Classification of Diabetes Mellitus. Who, Report No: 99.2, Geneva, 1999.
- [12] Erdoğan G., Koloğlu Endokrinoloji: Temel ve Klinik (2. Baskı), s. 342-343, MN Medikal-Nobel, İstanbul, 2005.
- [13] Bennet P. H., Knowler W. C., Definition, Diagnosis and Classification of Diabetes Mellitus and Glucose Homeostasis. In "Joslin Diabetes Mellitus". (eds) Kahn C. R., Weir G. C., King G. L., Jacobson A. M., Moses A. C., Smith R. J., 14th ed, Philadelphia, Lippincott Williams &Wilkins, p: 331-339, 2005.
- [14] Anonim, Diabetes Mellitus ve Komplikasyonlarının Tanı Tedavi ve İzlem Klavuzu. Türkiye Endokrinoloji ve Metabolizma Derneği (TEMED): Diabetes Mellitus Çalışma ve Eğitim Grupları 2009. 4. Baskı. İstanbul, 2009.
- [15] Anonim, What Is Diabetes? International Diabetes Federation: Diabetes Atlas, 4th Edition, 2009. <http://www.diabetesatlas.org> (Erişim Tarihi: 25.04.2018)
- [16] Anonim, Diabetes Care. American Diabetes Association: Standards of Medical Care in Diabetes-2010, 33(1):11-61, 2010.
- [17] Anonim, Diabetes Care. American Diabetes Association: Standards of Medical Care in Diabetes-2011, 34 (1):11-61, 2011.



- [18] Olgun, N., Eti Aslan, F., Coşansu, G., Çelik S., Diabetes Mellitus. Dahili ve Cerrahi Hastalıklarda Bakım, s: 829-864. Ed: by Karadakovan, A., Aslan, F. E., Nobel Tıp Kitabevi, Adana, 2010.
- [19] Anonim, Diyabetin Önemi, Ege Diyabetliler Derneği.  
<http://www.egediyabet.org/1197,diyabetin-onemi> (Erişim Tarihi: 25.04.2018)
- [20] King, H., Rewers, M., and WHO ad hoc Diabetes Reporting Grup, Global Estimates for Prevalence of Diabetes Mellitus and Impaired Glucose Tolerance in Adults. Diabetes Care, 16:157-177, 1993.
- [21] King, H., Aubert, Rf., Herman, Wh., Global Burden of Diabetes 1995-2025. Diabetes Care, 21:1414-1431, 1998.
- [22] Satman, İ., Yılmaz, C., İmamolu, S., Diabetes Mellitus ve Komplikasyonlarının Tanı, Tedavi ve İzlem Kılavuzu. Türkiye Endokrinoloji ve Metabolizma Derneği, 2007.  
[http://temd.org.tr/admin/uploads/tbl\\_kilavuz/DIYABET2017\\_web.pdf](http://temd.org.tr/admin/uploads/tbl_kilavuz/DIYABET2017_web.pdf) (Erişim Tarihi: 29.04.2018)
- [23] Anonim, Diyabet İstatistikleri, Türk Diyabet Cemiyeti.  
<http://www.diabetcemiyeti.org/c/diyabet-istatistikleri>  
(Erişim Tarihi: 29.04.2018)
- [24] Olgun, N., Yalın H., Gülyüz Demir H., Diyabetle Mücadelede Diyabet Risklerinin Belirlenmesi ve Tanılama. Family Physician 2(2): 36-44, 2011.
- [25] İzgi, C., Tip II Diyabetli Hastalarda Hastalık Algısının Diyabet Yönetimine Etkisi ve Etkileyen Faktörlerin İncelenmesi. Yüksek Lisans Tezi. Süleyman Demirel Üniversitesi, Sağlık Bilimleri Enstitüsü, Isparta, 2014.
- [26] Yılmaz, T., Tip 1 Diabetes Mellitus. Diabetes Mellitus, s:38-51, Ed: by İmamoğlu, Ş., Ersoy, C., 3. Baskı, Deomed Medikal Yayıncılık, İstanbul, 2009.

- [27] Woodrow, P., Intensive Care Nursing. 2nd Ed., London and Newyork: Roudledge Taylor &Francis Group, p. 481-487, 2006.
- [28] Malek, M., Tip 2 Diyabetli Hastalara Verilen Beslenme Eğitiminin Beslenme Bilgi ve Alışkanlıkları Üzerine Etkisinin İncelenmesi. Doktora Tezi. Fen Bilimleri Enstütüsü, Ankara, 2010.
- [29] Kabalak, T., Çetinkalp, Ş., Tip 2 Diabetes Mellitus. Diabetes Mellitus, s:54-72, Ed: by İmamoğlu, Ş., Ersoy, C., 3. Baskı, İstanbul, Deomed Medikal Yayıncılık, 2009.
- [30] Satman, I., Diabetes Mellitus Epidemiyolojisi. Diabetes Mellitus, s:11-35, Ed: by İmamoğlu, Ş., Ersoy, C., 3. Baskı. İstanbul, Deomed Medikal Yayıncılık, 2009.
- [31] Anonim, Tip 2 Diyabet Nedir? Metabolik Cerrahi.  
<http://www.metabolikcerrahi.com/tip2-diyabet-nedir>  
(Erişim Tarihi: 01.05.2018)
- [32] Beyazıt, E., Diyabetes Mellitus Tanısı Olan Bireylerde Verilen Planlı Eğitimin Metabolik Kontrol Üzerindeki Etkisinin İncelenmesi. Yüksek Lisans Tezi (basılmamış). Cumhuriyet Üniversitesi, Sivas, 2005.
- [33] Bozkurt, N., Yıldız, E., Diabetes Mellitusda Beslenme Tedavisi. Diyet El Kitabı, s: 257-288, Yenilenmiş 5. Baskı. Hatibogğu Yayınları: 116, Kaynak Dizisi 36, Ankara, 2008.
- [34] Gogas, D., Deyneli, O., Aydın, H., Tarçın, Ö., Diyabet ve Gebelik, Diabetes Mellitus 2009, s: 531-536, Ed: by İmamoğlu, Ş., Ersoy, C., 3. Baskı, İstanbul, Deomed Medikal Yayıncılık, 2009.

- [35] Kartal, A., Diyabetli Hastalarda Planlı Eğitim Programının Sağlık İnancına ve Diyabet Yönetimine Etkisinin İncelenmesi. Doktora Tezi (basılmamış). Ege Üniversitesi, İzmir, 2006.
- [36] Peters, A. N., Schriger, D. L., The New Diagnostic Criteria for Diabetes: The Impact on Management of Diabetes ve Macrovascular Risk Factors. The American Journal of Medicine, 105 (1): 15-19, 1998.
- [37] Sencer, E., Metabolizma ve Beslenme Hastalıkları. Geriatri İlavesi ile Nobel Tıp Kitabevi, Tayf Ofset, İstanbul, 2001.
- [38] Homafar, A., Tip II Diyabetik Bireylerde Balık Yağının (Omega 3) Açlık Kan Şekeri, Kan Basıncı, Serum Lipid Profili, İnsülin Düzeyi, Duyarlılığı ve Direnci Üzerindeki Etkileri. Doktora Tezi (basılmamış). Gazi Üniversitesi, Ankara, 2008.
- [39] Olgun, N., Gedik, S., Diyabet Tedavisinde Evde Glisemi ve Glukozüri Takibi. Diyabet Forumu, 8(2): 25-29, 2003.
- [40] Unutmaz, N., Çiftçi, H. ve Bambul, N., Hiperkolesterolemili Tip II Diyabetes Mellitus Hastalarında Simvastatin ve Lovastatin Etkinliğinin Araştırılması. Göztepe Tıp Dergisi, 13: 5-12, 1998.
- [41] Egwim, P. O., Plasma ve Lipoprotein Cholesterol in Weaning ve Post Weaning Rast Fed Casein ve Soya Protein Diets. Nutrition Researrch, 5(11): 1237-1245, 1985.
- [42] Arslan, P., Enerji Hesaplama Yöntemleri. Şişmanlık Çesitli Hastalıklarla Etkileşimi ve Diyet Tedavisinde Bilimsel Uygulamalar. Türkiye Diyetisyenler Derneği Yayını: 4, Volkan Matbaacılık, s: 35-50, Ankara, 1992.
- [43] Karcier, S. M., Hipertansiyon, Obesite, Dislipidemi, Glukoz İntoleransı ve İnsülin Direnci. Türk Kardioloji Derneği Arşivi, 21: 325-328, 1993.

- [44] Bitik, B., Yıldız O. B., Diyabetik Hastalarda Hipertansiyon Tedavisi. Türkiye Klinikleri J Nephrol-Special Topics 2009, 2(3):34-9, 2009.
- [45] Collazo-Clavell, M., Diyabet tedavisi. Mayo Clinic, Günes Kitabevi. (Çeviri: Alper Gürlek), s: 194, Ankara, 2004.
- [46] UK Prospective Diabetes Study (UKPDS) Group, Intensive Blood-Glucose Control with Sulphonylureas or İnsülin Compared With Conventional Treatment and Risk of Complications in Patients with Type 2 Diabetes (UKPDS 33). Lancet 1998; 352:837-53, 1998.
- [47] Behrman, Re., Kliegmen, Rm., Nelson Essentials of Pediatrics. Cev. Ed: by Tuzcu, M.: Endokrin Hastalıkları. Nobel Tıp Kitabevleri, İstanbul, 679-693, 2001.
- [48] Yazıcı, H., Hamuryudan, V., Sonsuz, A., Cerrahpaşa İç Hastalıkları. İstanbul Medikal Yayıncılık Ltd. Şti., İstanbul, 2007.
- [49] Dağdelen, A., Z., Aydın Merkezde Diyabet Bilinci: Diyabetle İlgili Bilgi Düzeyi ve İlişkili Faktörler. Uzmanlık Tezi. Adnan Menderes Üniversitesi Tıp Fakültesi, Aydın, 2012.
- [50] Anonim, Diyabetik Nefropati, Medicana.  
<http://www.medicana.com.tr/saglik-rehberi-detay/3578/diyabetik-nefropati>  
(Erişim Tarihi: 05.05.2018)
- [51] Kurt, M., Atmaca, A., Gürlek, A., Diyabetik Nefropati. Hacettepe Tıp Dergisi, 3(1): 12-17, 2004.
- [52] Tuğrul, A., Diyabetik Nefropati. Trakya Üniversitesi Tıp Fakültesi Dergisi, 19(2): 113-121, 2002.

- [53] Gross, JI., De Azevedo, Mj., Silveiro, Sp., Canani, Lh., Caramori, Ml., Zelmanovitz, T., Diabetic Nephropathy: Diagnosis, Prevention and Treatment. *Diabetes Care*, 28: 164-176, 2005.
- [54] Bogdanovic, R.: Diabetic Nephropathy in Children. *Nephrol Dial Transplant*. 16 (Suppl 6): 120-122, 2001.
- [55] Lindstrom, J., Neumann, A., Sheppard, K., E., Gilis-Januszewska, A., Greaves C., J., Handke U et al. Take Action to Prevent Diabetes-The IMAGE Toolkit for the Prevention of Type 2 Diabetes in Europe. *Horm Metab Res.*, 42 (1):37-55, 2010.
- [56] The DPP Study Group, The Diabetes Prevention Program: Baseline Characteristics of The Randomized Cohort. *Diabetes Care*, 23(11):1619-29, 2000.
- [57] Absetz, P., Oldenburg, B., Hankonen, N., Valve, R., Heinonen, H., Nissinen, A., et al. Type 2 Diabetes Prevention in the Real World: Three-Year Results of the GOAL Lifestyle Implementation Trial. *Diabetes Care* 2009, 32(8):1418-1420, 2009.
- [58] Balkau, B., Lange, C., Fezeu, L., Tichet, J., Lauzon-Guillain, B., Czernichow, S., et al. Predicting Diabetes: Clinical, Biological, and Genetic Approaches: Data From the Epidemiological Study on the Insulin Resistance Syndrome (DESIR). *Diabetes Care* 2008, 31(10): 2056-61, 2008.
- [59] Schwarz, P. E., Lindstrom, J., Kissimova-Scarbeck, K., Szybinski, Z., Barengo N. C., Peltonen, M., et al. The European Perspective of Type 2 Diabetes Prevention: Diabetes in Europe-Prevention Using Lifestyle, Physical Activity and Nutritional Intervention (DE-PLAN) Project, *Exp Clin Endocrinol Diabetes* 2008;116:167-172, 2008.

- [60] Schwarz, P. E., Li, J., Lindstrom, J., Tuomilehto, J., Tools for Predicting the Risk of Type 2 Diabetes in Daily Practice, *Horm Metab Res* 2009, 41:86-97, 2009.
- [61] Waugh, N., Scotland, G., McNamee, P., Gillett, M., Brennan, A., Goyder, E., et al. Screening for Type 2 Diabetes: Literature Review and Economic Modelling, *Health Technol Assess* 2007, 11:1-125, 2007.
- [62] Monnier, L., Lapinski, H., Colette, C., Contributions of Fasting and Postprandial Plasma Glucose Increments to the Overall Diurnal Hyperglycemia of Type 2 Diabetic Patients. *Diabetes Care*, 200, 26:881-885, 2003.
- [63] Holman, R. R., Paul, S. K., Bethel M. A., Neil H. A. W., Matthews D. R., Long-Term Follow-Up After Tight Control of Blood Pressure in Type 2 Diabetes. *N Engl J Med* 2008, 359: 1565-1576, 2008.
- [64] Anonim, Studies, International Diabetes Federation. <http://www.idf.org/studies> (Eriřim Tarihi: 01.02.2018).
- [65] Anonim, Yeni Sanayi Devrimi Akıllı Üretim Sistemleri Teknoloji Yol Haritası, TÜBİTAK Bilim, Teknoloji ve Yenilik Politikaları Daire Başkanlığı, 2016. [http://www.tubitak.gov.tr/sites/default/files/akilli\\_uretim\\_sistemleri\\_tyh\\_v27aralik2016.pdf](http://www.tubitak.gov.tr/sites/default/files/akilli_uretim_sistemleri_tyh_v27aralik2016.pdf) (Eriřim Tarihi: 01.02.2018).
- [66] Argüden, Y., Erřahin, B., Veri Madencilięi: Veriden Bilgiye, Masraftan Deęere. ARGE Danıřmanlık Yayınları, İstanbul, 2008.
- [67] Çakmak, Z., Kümeleme Analizinde Geçerlilik Problemi ve Kümeleme Sonuçlarının Deęerlendirilmesi, *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 3:187-205, 1999.

- [68] akmak, Z., Uzgören, N., Keek, G., Kemeleme Analizi Teknikleri ile İllerin Kltrel Yapılarına Gre Sınıflandırılması ve Deęiřimlerinin İncelenmesi. Dumlupınar niversitesi Sosyal Bilimler Dergisi, Sayı: 12.  
[http://birimler.dpu.edu.tr/app/views/panel/ckfinder/userfiles/17/files/DERG\\_/12/15-36.pdf](http://birimler.dpu.edu.tr/app/views/panel/ckfinder/userfiles/17/files/DERG_/12/15-36.pdf) (Eriřim Tarihi: 01.04.2018).
- [69] Arslan, H., Sakarya niversitesi Web Sitesi Eriřim Kayıtlarının Web Madencilięi ile Analizi. Yksek Lisans Tezi, Sakarya niversitesi, Fen Bilimleri Enstits, Sakarya, 2008.
- [70] Sarıman, G., Veri Madencilięinde Kemeleme Teknikleri zerine Bir alıřma: K-Means ve K-Medoids Kemeleme Algoritmalarının Karřılařtırılması. Sleyman Demirel niversitesi Fen Bilimleri Enstits Dergisi, 15(3): 192-202, 2011.
- [71] Iřık, M., Blnmeli Kemeleme Yntemleri ile Veri Madencilięi Uygulamaları. Yksek Lisans Tezi, Fen Bilimleri Enstits, İstanbul, 2006.
- [72] zcan, C., Veri Madencilięinin Gvenlik Uygulama Alanları ve Veri Madencilięi ile Sahtekarlık Analizi. Yksek Lisans, İstanbul Bilgi niversitesi Sosyal Bilimler Enstits Biliřim Teknoloji Hukuku Programı, İstanbul, 2014.
- [73] Berkhin, P., Survey of Clustering Data Mining Techniques.  
<http://citeseer.nj.nec.com/berkhin02survey.html> (Eriřim Tarihi: 07.04.2004).
- [74] lloęlu Glen, ., Veri Madencilięi Teknikleri ile stn Yetenekli ęrencilerin İlgili Alanlarının Analizi. Doktora Tezi, Gazi niversitesi Biliřim Enstits, Ankara, 2014.
- [75] Ayık, Y.Z., zdemir, A., Yavuz, U., Lise Tr ve Lise Mezuniyet Bařarısının, Kazanılan Faklte ile İliřkisinin Veri Madencilięi Teknięi ile Analizi, Atatrk niversitesi, Sosyal Bilimler Enstits Dergisi 10 (2): 441-454, 2007.

- [76] Han, J., Kamber M., Data Mining Concepts and Techniques, Morgan Kaufmann Publishers, 1st Ed., San Francisco, USA, 2000.
- [77] Oğuz, O., Karar Destek Sistemleri, Karar Ağaçları ve Uygulaması.  
<https://www.slideshare.net/oguzzhanoguzz/karar-aalar-ve-entropi-kds>  
(Erişim Tarihi: 27.05.2018).
- [78] Taşcı, E., Onan, E., K-En Yakın Komşu Algoritması Parametrelerinin Sınıflandırma Performansı Üzerine Etkisinin İncelenmesi, Adnan Menderes Üniversitesi. <http://ab.org.tr/ab16/bildiri/102.pdf> (Erişim Tarihi: 15.04.2018).
- [79] Gültekin, F., Regrasyon Analizi, Balıkesir Üniversitesi.  
<http://w3.balikesir.edu.tr/~bsentuna/wp-content/uploads/2013/03/Regresyon-Analizi.pdf> (Erişim Tarihi: 15.04.2018).
- [80] Savaş, S., Topaloğlu, N., Yılmaz, M., Veri madenciliği ve Türkiye'deki Uygulama Örnekleri. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi (21): 1-23, 2011.
- [81] Özdamar, K., Paket Programlar ile İstatistiksel Veri Analizi. Cilt 1, 2.Baskı, Kaan Kitabevi, s:475-477, Eskişehir, 2002.
- [82] Lemeshow, S., Hosmer, D., Applied Logistic Regression (Wiley Series in Probability and Statistics). Wiley-Interscience; 2 Sub edition .p:2-4, 2000.
- [83] Girginer, N., Cankuş, B., Tramvay Yolcu Memnuniyetinin Lojistik Regresyon Analiziyle Ölçülmesi: Estram Örneği. Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 15(1): 181-193, 2008.
- [84] Elhan, A.H., Lojistik Regresyon Analizinin İncelenmesi ve Tıpta Bir Uygulaması. Biyoistatistik Yüksek Lisans Tezi, Ankara Üniversitesi, s:4-29, Ankara, 1997.



- [85] Bircan, H., Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama, Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 2 : 185-208, 2004.
- [86] Hacibeyoğlu, M., Arslan, A., Kahramanlı, S., Improving Classification Accuracy with Discretization on Datasets Including Continuous Valued Features, World Academy of Science, Engineering and Technology, 5(6): 497-500, 2011.
- [87] Akar, M., Gündoğdu, S., Bayes Teorisinin Su Ürünlerinde Kullanım Olanakları. Journal of Fisheries Sciences, 8 (1): 8-16, 2013.
- [88] Şekeroğlu, S., Hizmet Sektöründe Bir Veri Madenciliği Uygulaması. Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, İstanbul, 2010.
- [89] Freund, Y., Schapire, R.E., Experiments with a new boosting algorithm. In: Machine Learning. Proceedings of the Thirteenth International Conference. Pp: 148-156, 1996.
- [90] Breiman, L., Random Forests, Machine learning, 2001 Kluwer Academic Publishers, 45(1): 5-32, 2001.
- [91] Breiman, L., Bagging Predictors. Machine Learning 26(2): 123-140, 1996.
- [92] Archer K.J., Empirical Characterization of Random Forest Variable Importance Measure, Computational Statistical Data Analysis, Computational Statistics & Data Analysis, 52(4): 2249-2260, 2008.
- [93] Pal, M., Random Forest Classifier For Remote Sensing Classification, International Journal of Remote Sensing, 26(1): 217-222, 2005.

- [94] Horning N., RandomForests: An algorithm for image classification and generation of continuous fields data sets, International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences (GISIDEAS), 9-11 December, Hanoi, Vietnam, 2010.
- [95] Breiman L., Manual on Setting up, Using, and Understanding Random Forests V3.1, [http://oz.berkeley.edu/users/breiman/Using\\_random\\_forests\\_V3.1.pdf](http://oz.berkeley.edu/users/breiman/Using_random_forests_V3.1.pdf) (Eriřim Tarihi: 20.05.2018).
- [96] Akar, Ö., Güngör, O., Rastgele Orman Algoritması Kullanılarak Çok Bantlı Görüntülerin Sınıflandırılması. Jeodezi ve Jeoinformasyon Dergisi, 1(2): 139-146, 2002.
- [97] Watts J. D., Powell S. L., Lawrence R. L., Hilker T., Improved Classification of Conservation Tillage Adoption Using High Temporal and Synthetic Satellite Imagery, Remote Sensing of Environment 115: 66-75, 2011.
- [98] Liaw A., Wiener M., Classification And Regression By Random Forest, R News, Vol.2/3, December, 2002.
- [99] Özdarıcı, A., Akar, Ö., Güngör, O. Rastgele Orman Sınıflandırma Yöntemi Yardımıyla Tarım Alanlarındaki Ürün Çeşitliliğinin Sınıflandırılması. TUFUAB VI. Teknik Sempozyumu, 23-26 Şubat, Antalya, 2011.
- [100] Bhatia, N., Vandana, Survey of Nearest Neighbor Techniques, International Journal of Computer Science and Information Security, 8(2):302-305, 2010.
- [101] Liu, H., Zhang, S., Noisy Data Elimination Using Mutual K-Nearest Neighbor for Classification Mining, Journal of Systems and Software, 85(5):1067-1074, 2012.
- [102] Han, J., Kamber, M., Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, Burlington, 2006.

- [103] Anonim, K-En Yakın Komşu Methodu (K-Nearest Neighborhood), Yazılıma Giriş, <http://yazilimagiris.com/2017/11/k-en-yakin-komsu-methodu-k-nearest-neighborhood/> (Erişim Tarihi: 21.05.2018).
- [104] Saraç T., Yapay Sinir Ağları. Gazi Üniversitesi Endüstri Mühendisliği Bölümü Ana Bilim Dalı Seminer Projesi, Haziran 2004, Ankara, 2004.
- [105] Anonim, Günümüzde Yapay Zeka Teknolojilerinin En İyi Uygulamaları, Elektrik Port. <http://www.elektrikport.com/haber-roportaj/gunumuzde-yapay-zeka-teknolojilerinin-en-iyi-uygulamalari/19013#ad-image> ,%20http://blog.udentify.co/04/2017/en-cok-ilgi-ceken-yapay-zeka-uygulamalari/ (Erişim Tarihi: 22.05.2018).
- [106] Pirim, H., Yapay Zeka, Journal of Yasar University,1(1): 81-93, 2006.
- [107] Demirhan, A., Kılıç, A. Y., Güler, İ., Artificial Intelligence Applications in Medicine, Yoğun Bakım Dergisi, 9(1):31-41, 2010.
- [108] Öztemel E., Yapay Sinir Ağları. Papatya Yayınları, İstanbul, 2006.
- [109] Açmalı, F., Diyabet Hastalığının Yapay Sinir Ağları Kullanılarak Sınıflandırılması. Yüksek Lisans Tezi, Sakarya Üniversitesi, Sakarya, 2010.
- [110] Sağıroğlu, S., Besdok, E., Erler, M., Mühendislikte Yapay Zeka Uygulamaları- I Yapay Sinir Ağları. Ufuk Kitap Kırtasiye-Yayıncılık, 2003.
- [111] Yurtoğlu, H., Yapay Sinir Ağları Metodolojisi ile Öngörü Modellemesi: Bazı Makroekonomik Değişkenler İçin Türkiye Örneği, DPT, EMSAGM Şubat 2005, Rapor No:2683, s. 104, 2005.
- [112] Uysal, A. E., Yapay Zekanın Temelleri ve Bir Yapay Sinir Ağı Uygulaması. Yüksek Lisans Tezi, Marmara Üniversitesi, İstanbul, 2009.

- [113] ELMAS, Ç., Yapay Zeka Uygulamaları. Seçkin Yayıncılık ISBN 978- 975-02-0614-6, Kasım 2007.
- [114] ELMAS, Ç., Yapay Sinir Ağları (Kuram, Mimari, Eğitim, Uygulama). Seçkin Yayıncılık ISBN 975 347 712 4, Nisan 2003.
- [115] TÜRKER, N., RF/Mikrodalga Düzlemsel İletim Hatlarının Yapay Sinir Ağı ile Analiz ve Sentezi. Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 92s, İstanbul, 2004.
- [116] Sinha P., Sinha P., Comparative Study of Chronic Kidney Disease Prediction Using KNN and SVM. International Journal of Engineering Research and Technology. 4(12):608-12, 2015.
- [117] Chapman P., Clinton J., Kerber, R., et al. CRISP-DM 1.0: Step-By-Step Data Mining Guide: NCR Systems Engineering Copenhagen (USA and Denmark), DaimlerChrysler AG (Germany), SPSS Inc (2000) (USA) and OHRA Verzekeringen En Bank Groep B.V (The Netherlands), August, 2000.
- [118]Anonim,[http://kocaeli2007.kocaeli.edu.tr/kocaeli2007/TAM\\_METIN\\_NUMARALISIRALI-PDF/556-560.pdf](http://kocaeli2007.kocaeli.edu.tr/kocaeli2007/TAM_METIN_NUMARALISIRALI-PDF/556-560.pdf), (Erişim Tarihi: 22.05.2007).
- [119] Zahn, E., Informations Technologie und Informations Management, München, 300-357, 1997 (Erişim Tarihi: 08.11.2017).
- [120] M. Karim, M. Kamal Orabi1 Yasser, M. Rabah Thanaa Early predictive system for diabetes mellitus disease ICDM 2016, LNAI 9728 (2016), pp. 420-427, 2016.
- [121] J. Jiang, P. Trundle, J. Ren Medical image Analysis with Artificial Neural Networks Comput. Med. Imaging Graph., 34 (8), pp. 617-631, 2010.

- [122] Anonim, What Is Data Mining?, Oracle(b) 2016,  
[https://docs.oracle.com/cd/B28359\\_01/datamine.111/b28129/process.htm#CHDFGCIJ](https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm#CHDFGCIJ) (Erişim tarih: 10.04.2016).
- [123] Benoit G., Data mining, Annual Review of Information Science and Technology. Elsevier, 36(1): 265-310, 2002.
- [124] Anonim, Diabetes Statistics, International Diabetes Federation,  
[http://www.idf.org/sites/default/files/Atlas-poster-2014\\_EN.pdf](http://www.idf.org/sites/default/files/Atlas-poster-2014_EN.pdf)  
(Erişim 15.07.2015)
- [125] Jeerakathil T, Johnson JA, Simpson SH, Majumdar SR., Short-Term Risk for Stroke is Doubled in Persons with Newly Treated Type 2 Diabetes Compared with Persons without Diabetes: A Population-Based Cohort Study. Stroke; A Journal of Cerebral Circulation, 38:1739–43, 2007.
- [126] Anonim, Cardiovascular Disease & Diabetes, American Heart Association.  
[http://www.heart.org/HEARTORG/Conditions/Diabetes/WhyDiabetes-Matters/Cardiovascular-Disease-Diabetes\\_UCM\\_313865\\_Article.jsp](http://www.heart.org/HEARTORG/Conditions/Diabetes/WhyDiabetes-Matters/Cardiovascular-Disease-Diabetes_UCM_313865_Article.jsp)  
(Erisim 23.07.2015).
- [127] Hovind P., Rossing P., Tarnow L., Smidt UM., Parving HH. Progression of Diabetic Nephropathy. Kidney Int 2001;59:702–9, 2001.
- [128] Fong DS, Aiello LP, Ferris FL, Klein R. Diabetic Retinopathy. Diabetes Care 2004; 27: 2540-53, 2007.
- [129] Dang C.N., Boulton A.J., Int J Low Extrem Wounds. 2003;2:4-12.  
<http://ijl.sagepub.com/content/2/1/4.abstract> (Erişim 23.07.2015).