



T.C.

BURSA ULUDAĞ ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

EKONOMETRİ ANABİLİM DALI

İSTATİSTİK BİLİM DALI

**VERİ MADENCİLİĞİ ÇERÇEVESİNDE YAPAY SİNİR AĞLARI
VE BİRLİKTELİK KURALI ANALİZİ ÜZERİNE BORSA
İSTANBUL 30 ENDEKSİNDE BİR UYGULAMA
(YÜKSEK LİSANS TEZİ)**

EDANUR GÜNDOĞDU

BURSA –2023



T.C.

BURSA ULUDAĞ ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

EKONOMETRİ ANABİLİM DALI

İSTATİSTİK BİLİM DALI

**VERİ MADENCİLİĞİ ÇERÇEVESİNDE YAPAY SİNİR AĞLARI
VE BİRLİKTELİK KURALI ANALİZİ ÜZERİNE BORSA
İSTANBUL 30 ENDEKSİNDE BİR UYGULAMA
(YÜKSEK LİSANS TEZİ)**

EDANUR GÜNDOĞDU

Danışman:

Doç.Dr. Z. BERNA AYDIN

BURSA –2023

ÖZET

Yazar Adı ve Soyadı	: Edanur GÜNDOĞDU
Üniversite	: Bursa Uludağ Üniversitesi
Enstitüsü	: Sosyal Bilimler Enstitüsü
Anabilim Dalı	: Ekonometri Anabilim Dalı
Bilim Dalı	: İstatistik Bilim Dalı
Tezin Niteliği	: Yüksek Lisans Tezi
Sayfa Sayısı	: x+105
Mezuniyet Tarihi	:
Tez Danışmanı	: Doç. Dr. Z. Berna AYDIN

VERİ MADENCİLİĞİ ÇERÇEVESİNDE YAPAY SİNİR AĞLARI VE BİRLİKTELİK KURALI ANALİZİ ÜZERİNE BORSA İSTANBUL 30 ENDEKSİNDE BİR UYGULAMA

Veriye ulaşılabilirliğin oldukça kolay olduğu gelişen bilgi ve teknoloji çağında, büyük boyutlu veri yığınlarının işlenebilir hale getirilmesi ile anlamlı ve yararlı bilgiler kullanıcılara sunulmaktadır. Bu bağlamda ülkemizin ekonomik durumu ile doğrudan ilişkili olan ve aynı zamanda önemli derecede yatırımcıya sahip olan Borsa İstanbul Endeks verilerinden, çıkarımlar yapabilmek yatırımcılara yatırım kararı alırken yol gösterici olmaktadır. Bu araştırmanın amacı veri madenciliği tekniklerinden, yapay sinir ağları analizi ile BİST 30 endeksinin 2022 Aralık ayı değerlerini tahmin etmek ve birliktelik kuralı analizi kullanılarak BİST30 endeksinin birlikte hareket ettiği makroekonomik değişkenleri tespit etmektir. Araştırma 2018 Ocak ile 2022 Aralık ayını kapsamakta olup yapay sinir ağı analizinde girdi değişkeni olarak tüketici fiyat endeksi, para arzı, altın ons fiyatı, dolar kuru, euro kuru, mevduat faizi, finansal hizmetler güven endeksi, brent petrol seçilirken çıktı değişkeni olarak Borsa İstanbul 30 Endeksi seçilmiştir.

Yapay sinir ağı analizi için MATLAB (R2021b) programı, birliktelik kuralı analizi için Weka 3.8.5 programı kullanılmıştır.

Araştırma sonucunda yapay sinir ağları analizi ile üretilen tahmin değerleri gerçek değerler ile karşılaştırılmış ve tahmin değerlerinin gerçek değerlere çok yakın olduğu görülmüştür. Birliktelik kuralı analizinde ise makroekonomik değişkenler ve BİST 30 endeksi arasında birliktelikler üretildiği saptanmıştır.

Anahtar Kelimeler: Büyük Veri, Veri Madenciliği, Yapay Sinir Ağları, Birliktelik Kuralı Analizi



ABSTRACT

Name and Surname :Edanur GÜNDOĞDU
University : Bursa Uludag University
Institution : Sosyal Bilimler Enstitüsü
Field : Econometrics
Branch : Statistics
Degree Awarded : Master
Page Number : x+105
Degree Date :
Supervisor/s : Doç. Dr. Z. Berna AYDIN

AN APPLICATION ON ARTIFICIAL NEURAL NETWORKS AND ASSOCIATION RULE ANALYSIS IN THE FRAMEWORK OF DATA MINING ON THE STOCK EXCHANGE ISTANBUL 30 INDEX

In the age of information and technology, where access to data is quite easy, meaningful and useful information is provided to users by making large-sized data masses processable. In this respect, making inferences from the Istanbul Stock Exchange Index data, which is directly related to the economic situation of our country and also has a significant number of investors, guides investors in making their investment decisions. The aim of this research is to predict the December 2022 values of the BIST 30 index using artificial neural network analysis, one of the data mining techniques, and to identify the macroeconomic variables with which the BIST30 index moves together using association rule analysis. The search covers the period between January 2018 and December 2022. In the artificial neural network analysis, consumer price index, money supply, gold ounce price, dollar exchange rate, euro exchange rate, deposit interest rate, financial services confidence index, Brent oil were selected as input variables, while Borsa Istanbul 30 Index was selected as output variable. MATLAB (R2021b) program was used

for neural network analysis and Weka 3.8.5 program was used for association rule analysis.

As a result of the research, the forecast values produced by artificial neural network analysis were compared with the actual values and it was observed that the forecast values were very close to the actual values. In the association rule analysis, it was determined that relationships were produced between macroeconomic variables and the BIST 30 index.

KeyWords: Big Data, Data Mining, Artificial Neural Networks, Association Rule Analysis

ÖNSÖZ

“Veri Madenciliği Çerçevesinde Yapay Sinir Ağları Ve Birliktelik Kuralı Analizi Üzerine Borsa İstanbul 30 Endeksinde Bir Uygulama” adlı tez çalışmamda süreç boyunca görüş, öneri ve tecrübeleriyle çalışmama desteklerini esirgemeyen, fikirlerimi önemseyerek yol gösterici olan kıymetli danışman hocam Doç. Dr. Z. Berna AYDIN’a ve çalışmama görüşleriyle katkı sağlayan Dr. Öğretim Üyesi Zuhal AKBELLEN’e sonsuz teşekkürlerimi sunarım.

Hayatımın her döneminde her şeyden önce iyi insan olmayı öğreten, beni bugünlere getirirken büyük emek veren ve attığım her adımda en büyük destekçim olan sevgili annem ve babama, uzakta olmalarına rağmen varlıklarını hep yanımda hissettiren sevgili ablam ve ağabeyime, henüz dünyaya gelmeden en büyük motivasyon kaynağım olan biricik yeğenime, yürüdüğüm bu yolu güzelleştiren ve her zaman yanımda olan yol arkadaşşıma teşekkür ediyorum.

Tezimi 6 Şubat 2023 tarihinde meydana gelen depremde, etkilenen tüm çocuklara ithaf ediyorum.

Edanur Gündoğdu

Bursa-2023

İÇİNDEKİLER

ÖZET.....	iii
ABSTRACT.....	v
TABLO LİSTESİ.....	xi
ŞEKİL LİSTESİ.....	xii
GİRİŞ	1

BİRİNCİ BÖLÜM VERİ MADENCİLİĞİNDE KAVRAMSAL ÇERÇEVE

1.1. Büyük Veri Nedir?	3
1.2. Veri Madenciliği Tanımları.....	5
1.3. Veri Madenciliği Alanları	5
1.4. Veri Madenciliğinde Bilgi Keşfi.....	7
1.5. Veri Madenciliği Ve İstatistik Kavramları.....	10
1.6. Veri Madenciliğinde Kullanılan Programlar.....	11
1.6.1. R Programlama Dili	11
1.6.2. RapidMiner (YALE)	12
1.6.3. Konstanz Information Miner (KNIME).....	12
1.6.4. WEKA.....	13
1.6.5. MATLAB.....	13
1.7. Veri Madenciliği Kullanım Alanları	13
1.8.1. Tahminleyici Modeller.....	17
1.8.1.1. Sınıflandırma Teknikleri	17
1.8.1.1.1. Karar Ağaçları	17
1.8.1.1.2. Yapay Sinir Ağları	18
1.8.1.1.3. K-En Yakın Komşu	13
1.8.1.2. Regresyon.....	19
1.8.2. Tanımlayıcı Modeller.....	19
1.8.2.1. Kümeleme Teknikleri.....	19
1.8.2.2. Birliktelik Kuralları.....	20

İKİNCİ BÖLÜM

VERİ MADENCİLİĞİNDE YAPAY SİNİR AĞLARI VE BİRLİKTELİK KURALI ANALİZİ

2.1. Yapay Sinir Ağları	21
2.1.1.Yapay Sinir Ağlarının Tanımı.....	21
2.1.2 Yapay Sinir Ağlarının Temel Özellikleri	22
2.1.3 Yapay Sinir Ağının Gelişimi.....	24
2.1.4.Biyolojik Sinir Sistemleri.....	26
2.1.5.Yapay Sinir Ağı ve Öğeleri.....	27
2.1.6.Yapay Sinir Hücrelerinin Mimari Yapısı.....	32
2.1.7. Biyolojik ve Yapay Sinir Ağlarının Benzerlikleri ve Temel Farkları.....	34
2.1.8. Yapay Sinir Ağlarının Sınıflandırılması	35
2.1.8.1.Yapay Sinir Ağlarının Tipine Göre Sınıflandırılması.....	36
2.1.8.2.Yapay Sinir Ağlarının Öğrenme Yöntemine Göre Sınıflandırılması.....	37
2.1.8.3.Yapay Sinir Ağlarının Katmanlarına Göre Sınıflandırılması.....	39
2.1.8.4.Yapay Sinir Ağlarının Yapılarına Göre Sınıflandırılması	51
2.1.9.Yapay Sinir Ağlarında Öğrenme Kuralları	51
2.1.10.Yapay Sinir Ağlarının Görevleri ve Uygulama Alanları	53
2.1.11.Yapay Sinir Ağlarının Avantajlı ve Dezavantajlı Yönleri	55
2.2.Birliktelik Kuralı Analizi	56
2.2.1.Birliktelik Kuralı Analizi Matematiksel Gösterimi ve Kavramları.....	57
2.2.2.Birliktelik Kuralı Analizi Algoritmaları.....	60
2.2.2.1.AIS Algoritması	60
2.2.2.2.SETM Algoritması	60
2.2.2.3.Apriori Algoritması.....	61
2.2.2.4.Apriori-TID Algoritması	63
2.2.2.5.Apriori-Hybrid Algoritması	63
2.2.2.6.FP-Growth Algoritması.....	63
2.2.3.Birliktelik Kuralı Analizi Algoritmalarının Karşılaştırılması	65
2.2.4.Birliktelik Kuralı Analizi Kullanım Alanları	67

ÜÇÜNCÜ BÖLÜM
VERİ MADENCİLİĞİ ÇERÇEVESİNDE YAPAY SİNİR AĞLARI
VE BİRLİKTELİK KURALI ANALİZİ ÜZERİNE BORSA
İSTANBUL 30 ENDEKSİNDE BİR UYGULAMA

3.1. Literatür Taraması.....	76
3.1.1. Ulusal Literatür.....	68
3.1.2. Uluslararası Literatür	74
3.2. Araştırmanın Amacı	76
3.3. Araştırmada Kullanılan Veri Seti.....	77
3.4. Araştırmada Kullanılan Makroekonomik Değişkenler	78
3.5. Araştırmada Kullanılan Yöntemler	79
3.6. Yapay Sinir Ağı Analizi.....	79
3.7. Birliktelik Kuralı Analizi	84
SONUÇ	89
KAYNAKÇA.....	93
EKLER.....	106

TABLO LİSTESİ

Tablo	Sayfa
Tablo 1.1: Veri Madenciliği Tanımları	5
Tablo 1.2: İstatistiksel Analiz İle Veri Madenciliği	11
Tablo 2.1:Toplama Fonksiyonu Örnekleri	29
Tablo 2.2: En Çok Kullanılan Transfer Fonksiyonları Denklemleri	32
Tablo 2.3: Biyolojik Sinir Ağı İle Yapay Sinir Ağı'nın Karşılaştırılması	35
Tablo 2.4: Birliktelik Kuralı Madenciliği Algoritmalarının Karşılaştırılması	66
Tablo 3.1: Araştırmada Kullanılan Değişkenler	77
Tablo 3.2: Yapay Sinir Ağı Analizinde Kullanılan Değişkenler	79
Tablo 3.3: En İyi Performansa Sahip Ağ İçin Yapılan Denemeler	81
Tablo 3.4: YSA Tahmin Sonuçları	83
Tablo 3.5: Birliktelik Kuralı Analizi Sonuçları	85

ŞEKİL LİSTESİ

Şekil	Sayfa
Şekil 1.1: Büyük Verinin Temel Bileşenleri (5V)	4
Şekil 1.2: Veri Madenciliğinin Benimsediği Alanlar	6
Şekil 1.3: Bilgi Keşif Sürecinde Veri Madenciliği	8
Şekil 1.4: Veri Madenciliğine İlişkin Alt Alanlar	15
Şekil 1.5: Veri Madenciliği Modelleri	16
Şekil 1.6: Örnek Karar Ağacı	18
Şekil 2.1: Sinir Sisteminin Blok Şeması Gösterimi	26
Şekil 2.2: Bir Biyolojik Sinir Hücresinin Yapısı	27
Şekil 2.3: Yapay Bir Sinir (Düğüm)	28
Şekil 2.4: En Çok Kullanılan Transfer Fonksiyonları	31
Şekil 2.5: Yapay Sinir Ağı Katmanları	33
Şekil 2.6: Biyolojik Sinir Hücresi Ve Yapay Sinir Hücresi	34
Şekil 2.7: Yapay Sinir Ağlarının Sınıflandırılması	36
Şekil 2.8: İleri Beslemeli Ve Geri Beslemeli Ağ Yapısı	37
Şekil 2.9: Denetimli (Danışmanlı) Öğrenme	38
Şekil 2.10: Denetimsiz (Danışmansız) Öğrenme	39
Şekil 2.11: Tek Katmanlı Yapay Sinir Ağı Modeli	40
Şekil 2.12: Ağırlıkların Ve Sınıf Ayracı Olan Doğrunun Geometrik Gösterimi	41
Şekil 2.13: Basit Algılayıcı Modelinin Yapısı	43
Şekil 2.14: Adaline Ünitesi	44
Şekil 2.15: Çok Katmanlı Algılayıcı Yapay Sinir Ağının Yapısı	46
Şekil 2.16: Geriye Yayılımlı Yapay Sinir Ağları	48
Şekil 2.17: Apriori Algoritması	62
Şekil 2.18: Örnek Veri Seti	64
Şekil 2.19: Örnek Veri Setine Ait FP-Tree(FP-Ağaç)	65
Şekil 3.1: En İyi Performansa Sahip Olan Ağın Mimarisi	82
Şekil 3.2: Eğitim Doğrulama ve Test Verileri İçin MSE (Ortalama Kare Hata)	82
Şekil 3.3: Eğitim Sonucunda Ulaşılan Tahmin Başarıları	83
Şekil 3.4: Birliktelik Kuralı Analizine Ait Veri Setinin Düzenlemeleri	84

GİRİŞ

Veriye ulaşılabilirliğin giderek kolaylaştığı bilgi ve teknoloji çağında verilerden faydalı ve anlamlı bilgiler üreterek çıkarımlar yapmak oldukça önemli hale gelmiştir. Bu doğrultuda çeşitli modelleme teknikleri sunan veri madenciliği kavramından söz etmek yanlış olmayacaktır. Veri madenciliği kavramı, veri içerisinde gizli kalmış ilişkileri tespit ederek bu ilişkiler vasıtasıyla geçerli tahminlerde bulunmaya yarayan süreçleri kapsamaktadır (Koyuncugil,Özgülbaş,2008:4).Veri madenciliğinin sağladığı modelleme teknikleri ile çok büyük boyutlu ve sadece sayılardan ibaret olduğu düşünülen karmaşık veri yığınları işlenerek anlamlı sonuçlara ulaşılabilir. Böylece çok büyük boyutlu veriler amacına uygun bir şekilde işlenerek yalnızca veri tabanlarında yer kaplayan birer yığın olmaktan çıkmaktadır. Ayrıca veri madenciliği tekniklerinin oldukça geniş kapsamlı uygulama alanlarına sahip olması ve klasik istatistiksel yöntemlerin aksine esnek yapısı ile birtakım avantajlar sağlaması gün geçtikçe araştırmacıların ilgi odağı haline gelmiş ve araştırmalara konu olmaya başlamıştır. Farklı araştırmalara konu olan veri madenciliği teknikleri, bir ülkenin ekonomik durumu ile doğrudan ilişkisi olduğu düşünülen ve önemli yatırım araçlarından biri olan borsada da kullanılmaktadır.

Borsa verileri ülkemizde faaliyet gösteren bir kurum olan Borsa İstanbul tarafından açıklanmakta olup yatırımcıların ilgi odağıdır. Nitekim yatırımcılar için yatırıma en uygun sektörü ve şirketi seçebilmek doğru yatırım yapabilmek adına oldukça önemlidir. Öte yandan yatırımcılar için, yalnızca doğru sektörü ve şirketi seçmek yeterli olmayabilir. Dolayısıyla çeşitli spekülasyonlar göz önüne alınarak borsa endeks verilerini etkileyen diğer olası faktörleri de yakından takip etmenin faydası olacağı söylenebilmektedir. Bu bağlamda borsa endeks verilerinin, enflasyon, faiz, euro, dolar, para arzı gibi çeşitli makroekonomik değişkenler ile ilişkisinin takip edilmesi endeksin olası yükseliş ve düşüşlerini saptamaya yardımcı olacaktır.

Araştırmanın amacı, ele alınan veri seti için veri madenciliği modellerinden uygun olan teknikler belirlenerek veriyi işlemek ve anlamlı sonuçlara ulaşarak yatırımcıların yararına bilgiler sunmaktır. Bu amaçla veri madenciliği modellerinden hem tahminleyici hem de tanımlayıcı teknikler kullanılarak anlamlı sonuçlar elde edilmeye çalışılacaktır.

Araştırma üç bölümden oluşmaktadır. Araştırmanın birinci bölümünde veri madenciliği kavramı, veri madenciliğinde kullanılan programlar, veri madenciliği kullanım alanları, veri madenciliği modelleri açıklanmıştır.

Araştırmanın ikinci bölümünde veri madenciliği tahminleyici modellerden yapay sinir ağları analizi ve tanımlayıcı modellerden birliktelik kuralı analizi detaylı bir şekilde açıklanmıştır.

Araştırmanın üçüncü bölümünde önce literatür taramasına yer verilmiştir. Araştırmalar incelendikten sonra araştırma kapsamında kullanılacak değişkenler açıklanarak BİST 30 endeksinin 2022 Aralık ayı değerleri yapay sinir ağları analizi ile tahmin edilmiş ve birliktelik kuralı analizi uygulanarak BİST30 endeksinin birlikte hareket ettiği makroekonomik değişkenler tespit edilmiştir.

BİRİNCİ BÖLÜM

VERİ MADENCİLİĞİNDE KAVRAMSAL ÇEVREÇE

1.1. Büyük Veri Nedir?

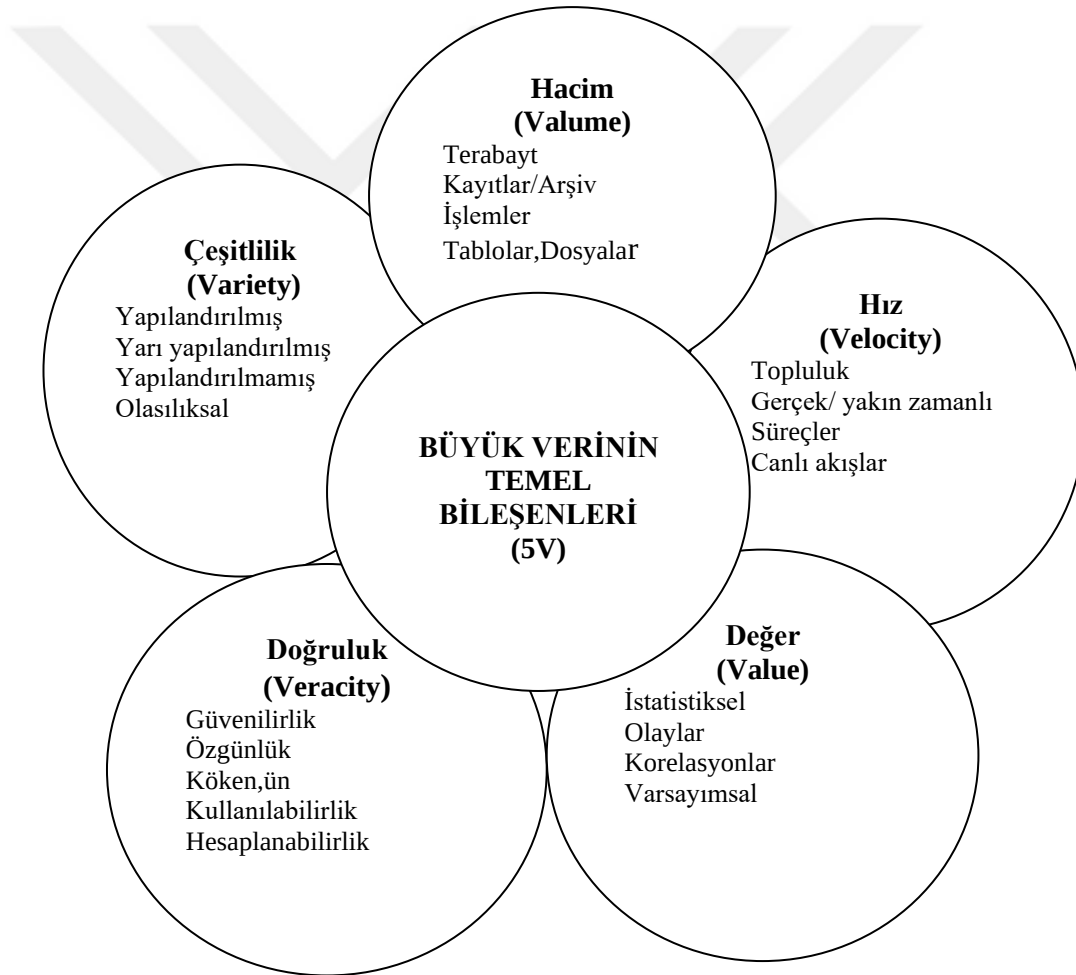
Teknolojinin gelişimi ile birlikte verilere ulaşılabilirlik ve verileri işleyebilirlik artmış, beraberinde büyük veri kavramını ortaya çıkarmıştır (Doğan, Arslantekin, 2016). Büyük veri kavramı yeni olgular, yöntemler ve programlardan söz etmemizi sağlarken bilinen veri, bilim, yazılım gibi kavramlara farklı bakış açıları getirmiştir. Bu bağlamda ‘veribilimci’ ve ‘veribilim’ kavramlarının sıklıkla kullanılmasını sağlayarak aynı zamanda mevcut bilimsel yöntemleri de etkilemiştir (Gürsaka, 2014:20).

Veri madenciliğinde önemli kavramlardan biri olan büyük veri kavramı, verinin sürekli artış gösterdiği göz önüne alındığında, hacimsel olarak büyüklüğünün yanı sıra diğer birtakım özelliklere de daha sahip olması gerektiğini vurgulamaktadır (Köse, 2018:51).

2012 yılında Gartner yapmış olduğu çalışmalar sonucunda, büyük veriyi 3V olarak adlandırmıştır. 3V kavramı büyük miktar veya hacim, büyük hız ve büyük çeşitlilik özelliklerini ifade etmektedir. Ayrıca büyük verinin karar verme yetkisini artırarak, bilginin işlenmesinde yeni biçimlere ihtiyaç duyduğu ifade edilmiştir. Büyük verinin 3V özelliğinin yansıya doğruluk (veracity) özelliğini de göz önünde bulundurması önem arz etmektedir. Çünkü karmaşıklık ve hatalardan arınarak işlenebilmesi için hız ve çeşitlilikle hacimsel olarak devamlı artan verinin, doğruluğunun sağlanması gerekmektedir (Gürsaka, 2014:25-26). Bütün bu özelliklerin yanında büyük veriye ulaşmanın önemli olduğu kadar bu verilerden değer (value) üretebilmekte büyük verinin özelliklerinden birini göstermektedir (Dahdouh vd., 2020).

Büyük verilerin önemli bir özelliği olan hacim, verinin büyüklüğüne göre şekillenir ve giderek artış gösterir. Verinin hacmi arttıkça büyüklüğü birkaç megabayttan birkaç gigabayta yükselmektedir (Cyganek vd., 2016:498). Büyük veri hız özelliği bakımından ifade edildiğinde, üretilme hızının çok hızlı olduğu ve bu hıza bağlı olarak verinin işleme sürecinde hızlı olması gerektiği söylenebilir (Aktan, 2018:4). Büyük verinin

çeşitlilik özelliği ise verinin toplandığı platformların ve formatların farklılığını ifade etmektedir. Örneğin toplanan verinin formatı metinsel olabildiği gibi görsel veya ses formatında olabilmektedir. Yine aynı şekilde toplanan veri sosyal medya platformundan toplanabileceği gibi e-posta vb. aracılığıyla da toplanmış olabilir. Bu nedenle verinin çeşitli olması farklı tekniklere ihtiyaç duyulduğunun bir göstergesi olmuştur (Atan, 2016:148).Büyük verinin bir diğer önemli bileşeni olan değer, her verinin bir değeri olduğunu gösterirken büyük veri setleri için hangi veri değerinin önemli olduğunu ve analize tabi tutulacağını tespit etmek için belirlenmesi oldukça önem teşkil etmektedir (Köse, 2018:54).Aşağıda Şekil 1.1’de büyük verinin temel bileşenleri gösterilmiştir (Dahdouh vd.,2020:19):



Şekil 1.1 : Büyük Verinin Temel Bileşenleri (5V)

Kaynak:(Dahdouh vd.,2020:19)

1.2. Veri Madenciliği Tanımları

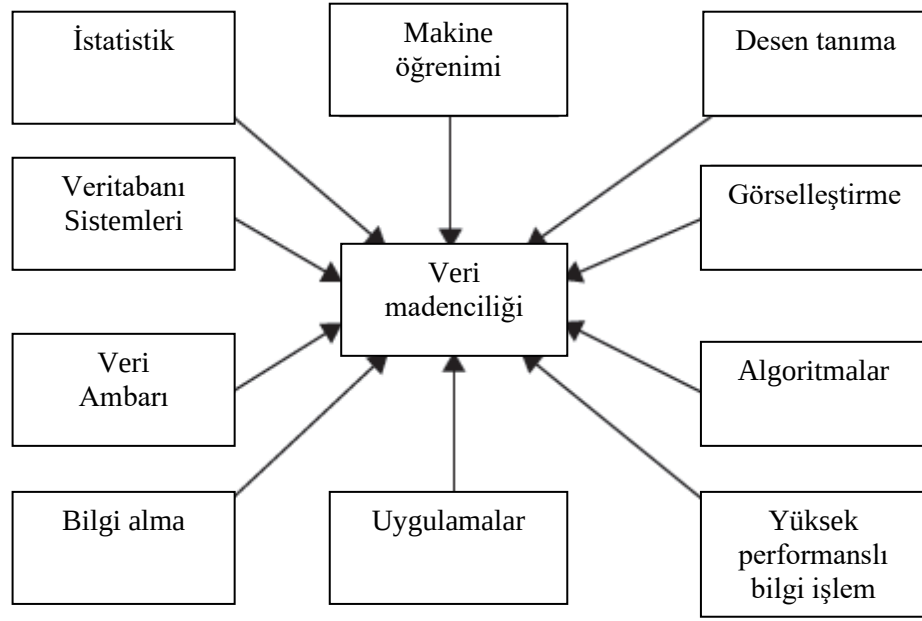
Veriye ulaşılabilirlik önemli olduğu kadar veriden maksimum fayda sağlayarak bilgi üretmekte aynı derecede önemlidir. Veriden bilgi üretme veri madenciliği kavramını ortaya çıkarmıştır (Durmaz,Kocamış, 2008:2).Veri madenciliği gelişmekte ve giderek kullanım alanı artmakta olan bir alan olduğu için tek bir tanımlama yapmak mümkün değildir (Silahtaroglu, 2016:12).Veri madenciliğine ilişkin çeşitli tanımlamalar aşağıda Tablo1.1’de yer almaktadır:

Tablo 1.1:Veri Madenciliği Tanımları

Veri Madenciliği Tanımları
Veri madenciliği veriler üzerinde bir model oluşturarak algoritmaların uygulanmasını sağlayan bilgi keşif sürecinin bir adımıdır (Fayyad vd., 1996:41).
Veri madenciliği veri işlemeye ve büyük boyutlarda veri toplamaya olanak sağlayan bilgisayar teknolojilerinin yardımı ile büyük veri yığınları içerisinde gizli, faydalı, örtük bilgilerin keşfedilmesine izin veren çeşitli alanlarda kullanılan ve kullanıcıya yarar sağlayan bir süreçtir (Şentürk, 2006:3).
Veri madenciliği çok büyük hacimlerde veriye sahip olan veritabanlarından faydalı, iş yarar ve mantıklı sonuçlara varabilmek için, birçok basamaktan oluşan işlemleri yapmaktır (Aktürk,Korukoğlu, 2008:114).
Veri madenciliği verileri girdi olarak alan ve sonuç olarak bilgi üreten bir süreçtir (Weiss,Davison, 2010:1).
Veri madenciliği istatistik bilimine farklı bakış açıları getiren, sınırlarını yeniden çizen yapay zeka ve makine öğrenimi desteğiyle, değerli kullanılabilir bilgilerin büyük veriler içerisinde seçilmesi işlemidir (Koyuncugil, 2007:1-2).

1.3. Veri Madenciliği Alanları

Veri madenciliği teknikleri oldukça geniş kullanım alanına sahip olmasından dolayı birçok alanda kullanılmaktadır. Bu alanlar aşağıda Şekil 1.2’de gösterilmiştir (Han vd., 2012:23):



Şekil 1.2:Veri Madenciliğinin Benimsediği Alanlar

Kaynak: (Han vd., 2012:23)

Yukarıda Şekil 1.2.'de verilen alanlardan en çok istatistik,makine öğrenimi, veri ambarı ve veritabanı sistemleri alanlarında veri madenciliği teknikleri kullanılmaktadır.Verileri toplama,ayırıştırma,analiz etme ve analiz sonuçlarını yorumlama temeline dayanan istatistik ile aynı şekilde bu temele dayanan veri madenciliği arasında doğal olarak bir bağlantı bulunmaktadır.Veritabanı madenciliği çeşitli tekniklerle elde ettiği sonuçları doğrulamada istatistiksel hipotez testlerini kullanırken bunun yanı sıra sınıflandırma gibi veri madenciliği tekniklerinde istatistiksel modeller oluşturabilmektedir (Han vd., 2012: 23-24).

Makine öğrenimi, gelişmiş bilgisayar sistemlerinden yararlanarak verileri işleyebilir ve bilgisayar programlarının karmaşık yapılarını otomatik olarak öğrenebilmektedir. Büyük veri boyutunun karmaşık yapısıyla ilgilenen veri madenciliği makine öğrenimi disipliniyle yüksek oranda ilişkilidir (Han,vd.,2012:24).Makine öğrenimi ilgilenilen veri setinden öğrenmeyi amaçlayan bu amaçla elde edilen çıktıları kümeleyen, sınıflayan, ilişki kuran tekniklerden oluşmaktadır(Balaban,Kartal, 2018:3).

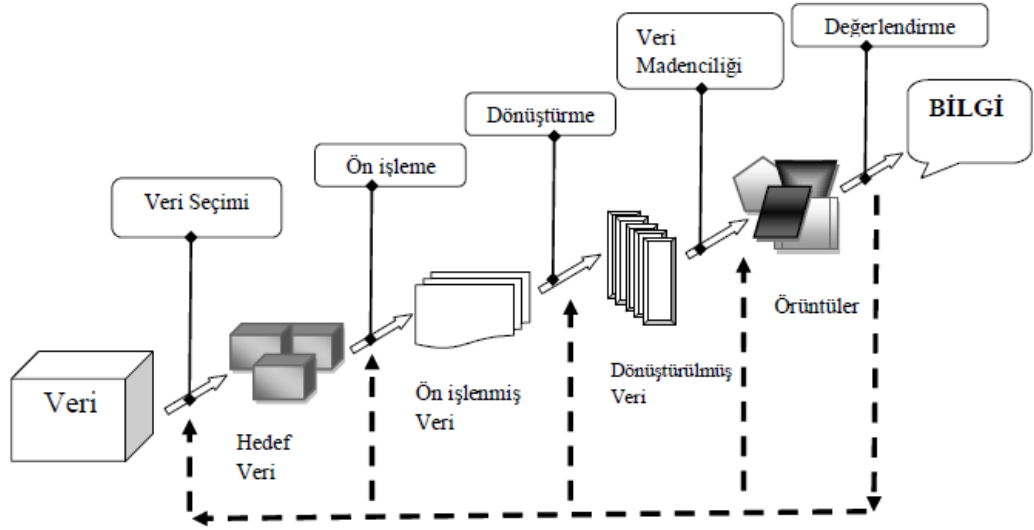
Veritabanı sistemleri yeni veritabanı oluşturmak, yönetmek, kullanıcıların erişimini sağlamak, verileri yedeklemek gibi işlemleri düzenleyen sistemlerden oluşmaktadır (Şentürk,2006:6).Veritabanı sistemleri çok büyük veri kümelerini işlemedeki ölçeklenebilirliği göstermektedir. Bu nedenle veri madenciliği teknikleri büyük veri kümelerini işlemek, yüksek verimlilik ve ölçeklenebilirlik elde etmek amacıyla ölçeklenebilir veritabanı sistemlerinden faydalanırken ileri düzey karmaşık veri analizleri için mevcut veritabanı sistemlerinin kapasitesinin artmasını sağlamaktadır (Han vd., 2012:26).

Veri madenciliği büyük veri yığınları içerisinde anlamlı bilgiler keşfetme sürecini oluştururken aynı zamanda büyük veri yığınlarından keşfedilen bilgiler ışığında var olan mevcut sistemlerin aksaklıklarını eksiklerini tespit etmeyi ve gidermeyi amaçlayarak daha iyi performans gösterecek sistemler oluşturmaktadır. Bu nedenle veri madenciliği birçok alanı benimseyerek geniş kullanım alanı yaratmıştır (Özbay, 2015:264).

1.4. Veri Madenciliğinde Bilgi Keşfi

Veri madenciliğinde esas olan, büyük veri tabanlarını işlenebilir ve yararlı hale getirmektir. Örneğin büyük bir veritabanını ilgilendiğimiz konu doğrultusunda şekillendiremediğimizde ve öncesinde birtakım indirgemeler yapmadığımızda büyük veri tabanı hiçbir şey ifade etmeyecek, binlerce veriyle birlikte bilgide çöp olacaktır. Diğer bir ifadeyle, veri madenciliğinde kullanılacak en uygun teknik seçilmeden önce bir sürece tabii tutulmalıdır(Han vd.,2012:6). Örneğin Savaş vd.(2012) tarafından üzerinde çalışılan konu hakkındaki verilerin özelliklerini bilmeden uygulanacak veri madenciliği tekniklerinin yararlı olmayacağı bu süreçte başarılı olmanın yolunun verilerin özelliklerini bilerek madenciliğe başlamaktan geçtiği belirtilmiştir.

Aşağıda Şekil 1.3’de verinin veri madenciliği aracılığıyla bilgiye dönüşümü gösterilmektedir.



Şekil 1.3: Bilgi Keşfi Sürecinde Veri Madenciliği

Kaynak:(Savaş vd., 2012: 8)

Yukarıdaki şekilden hareketle veri madenciliğinde bilgi keşfi 7 adımda gerçekleşen bir süreçtir.

1. Veri temizleme (veri setine aykırı ve yanlış verilerin tespit edilmesi)
2. Veri birleştirme
3. Veri seçimi
4. Veri dönüştürme
5. Veri madenciliği
6. Örüntü değerlendirmesi
7. Bilgi sunumu

Sürecin ilk 4 adımı verinin ön işlemesinden oluşmaktadır (Han vd., 2012:6-7).

Bu sürece ilişkin açıklamalar aşağıda verilmiştir:

- ✓ *Veri Temizleme*: Bu işlem eksik veya hatalı verilerin tespiti için veriler üzerinde çıkarımlar yapılmadan önce verilerin daha doğru sonuçlar vermesi amacıyla yapılmaktadır. Veri toplama işlemi sırasında veriler hatalı, eksik veya hiç girilmemiş olabilir. Örneğin binary (ikili) değişken tipine ait verilerden oluşan veri setinde 0-1 değerleri yer alması gerekirken hatalı/yanlış veya boş girilmiş bir değer tespit edildiğinde düzeltilebilir. Bu gibi durumlarda ilgili değişkenin

ortalaması alınarak veya ilgili değişken analiz dışı bırakılarak problem çözülebilir (Şentürk, 2006:12).

- ✓ *Veri Birleştirme:* Bu işlem farklı kaynaklardan gelen verilerin tek bir şekle bürünmesidir. Ancak bu işlem farklı tablolardan gelen ve isim olarak farklılık gösterse dahi aynı şeyi temsil ettiği düşünülen değişkenleri tek bir çatı altında topladığı için fazlalıklara neden olacak ve hatalı sonuçları beraberinde getirecektir. Böyle durumlarda aynı şeyi temsil ettiği düşünülen değişkenler arasında korelasyon katsayısına bakarak değişkenlerden biri analizden çıkarılabilir (Şentürk, 2006:14).
- ✓ *Veri Seçimi:* Bu işlemin veri madenciliğinin kilit noktası olduğunu söyleyebiliriz. Çünkü veri madenciliği teknikleri kullanmaya başlamadan önce merak ettiğimiz veya ilgi alanımız olan bir araştırma sorusunun olması ve bu araştırma sorusunun teknik olarak analizini gerçekleştireceğimiz verilere ulaşılabilir olmamız gerekmektedir. Veri seçimi yaparken verilerin seçtiğimiz konuyla bağdaşması anlamlı ve ilgili verileri kapsamı çok önemlidir. Bu nedenle veri seçiminde uzman görüşlerinden yararlanmak veri analisti adına faydalı olacaktır (Köse, 2018:83-85).
- ✓ *Veri Dönüştürme:* Bu işlem daha sağlıklı ve doğru sonuçlara ulaşabilmek verilerin kullanılan algoritmalara uygunluğunu görmek amacıyla yapılmaktadır (Akpınar, 2017:136). Başka bir ifadeyle verileri veri madenciliği sürecine uygun hale getirmek için düzeltme, birleştirme, genelleştirme ve normalleştirme gibi işlemlerin yapılmasıdır (Tüzüntürk, 2010:69).
- ✓ *Veri Madenciliği:* Veri madenciliği çeşitli veri tabanlarında yer alan çok çeşitli ve büyük miktardaki veriler içerisinde önceden keşfedilmemiş yeni bilgiler keşfetme sürecidir. Ayrıca elde edilen bu yeni bilgiler doğrultusunda gelecek tahmininde bulunarak çıkarımlar yapmaktadır (Dener vd., 2009:788).
- ✓ *Örüntü Değerlendirmesi:* Veri madenciliği teknikleri ile keşfedilen bilginin yeni ve katkı sağlayabilirlik açısından değerlendirilmesi ve son olarak bilgiyi sunar hale gelmesidir (Sever, Oğuz, 2002:177).

- ✓ *Bilgi Sunumu:* Veri madenciliği yöntemlerinden uygun olan yöntemin seçilmesi ile veritabanı hakkında keşfedilen yeni bilgiler geçerlilik, yenilik, yararlılık ve basitlik kriterlerine göre değerlendirilmesinin ardından, veri madenciliği kullanıcısına sunulmaktadır (İnan, 2015:17).

1.5. Veri Madenciliği Ve İstatistik Kavramları

Verinin olduğu bir yerde istatistik ve veri madenciliğinden birlikte söz etmek yanlış olmayacaktır. İstatistik veriyi çeşitli analizlerle sınavan ve ispata dayandıran bilim dalıdır. Veri madenciliği ise, temelini istatistikten alan ancak istatistik biliminin sınırlı kaldığı ve yetersiz olduğu durumlarda geniş çaplı analizlerden yararlanan bir disiplindir. İstatistik ve veri madenciliği birbiriyle ortak yönleri sahip olduğu gibi farklı yönleri de sahip olduğu görülmektedir (Emre,Selçukcan Erol, 2017:161).

İstatistik ve veri madenciliği temelini veriden alan, verilerdeki gizli bilgileri ortaya çıkarmaya çalışan, ortak amaca hizmet eden disiplinlerdir. İstatistik bilimi uzun geçmişe sahip olduğu gibi veri madenciliği de zamanla adını sıklıkla duyurmaya başlamıştır. Bu iki disiplinin temeli birbiriyle kesişen analitik yaklaşımlara dayanmaktadır (Zhao,Luan, 2006:8).

Veri madenciliği ve istatistik kavramları birbirlerinden farklı yönleriyle ele alındığında veri madenciliği tümdengelim, istatistik ise tümevarım yöntemini temsil etmektedir. Veri madenciliği anakütlenin tamamını ele alırken, istatistik ise çoğu zaman, analizlerde anakütlenin tamamını değil anakütleden seçilen ve anakütleyi en iyi temsil etme niteliğinde olan örnekleme ele almaktadır. İstatistik teorilere ve hipotezlere dayanarak analizleri gerçekleştirme yetkisine sahip olduğu için doğrulayıcı bir analiz tekniğidir. Veri madenciliği ise teorilerin doğrulanmasını beklemez. Ancak bu veri madenciliğinde gerekli varsayımların olmadığı anlamına gelmemektedir. Nitekim veri madenciliğinde istatistikte kullanılan varsayımlara göre daha sınırlı sayıda da olsa varsayımların varlığından söz edilebilir (Tüzüntürk, 2010:72).

İstatistik ve veri madenciliği disiplinleri arasındaki farklılıklar, kullanılan verinin tipine göre değerlendirildiğinde istatistikte deneysel verilerin varlığı söz konusu iken veri

madenciliğinde verilerin tipik olarak gözlemsel verilerden oluşması söz konusudur (Giudici, 2003:5)

Aşağıda Tablo 1.2 ‘de veri madenciliği ve istatistiksel veri analizinin karşılaştırılması verilmiştir (Ersöz, 2019:24):

Tablo 1.2: İstatistiksel Analiz İle Veri Madenciliği

Veri Madenciliği	İstatistiksel Veri Analizi
Problemleri çözmek için uzman sistemleri kullanır.	Modellerin istatistiksel doğruluğu için testler kullanır.
Veri hakkında daha az varsayıma ihtiyaç duyar.	Hipotez testleri yapar. Önemi doğrulamak için parametrik testler kullanır.
Büyük miktardaki verilerde daha iyi sonuç verir.	Kullanılan yöntemlerden ziyade örneklemeye güvenir.
Veri tabanlarında daha önceden tahmin edilemeyen ilişkilerden ayrıntılı analizler yapmayı mümkün kılar.	Büyük miktardaki veriler için optimize edilir.
Veri ve iş problemlerinin anlaşılmasını gerektirir.	Güçlü bir istatistik bilgi ve becerisi gerektirir.

Kaynak: (Ersöz, 2019:24)

1.6. Veri Madenciliğinde Kullanılan Programlar

Büyük veri tabanlarından önemli bilgiler elde etmek ve veri madenciliği yöntemlerini uygulayabilmek için bilgisayar biliminin yardımı ile çeşitli veri madenciliği programları geliştirilmiştir (Dener vd., 2009:787).Veri madenciliğinde sıklıkla kullanılan programlar R,KNIME, RapidMiner(YALE),WEKA, MATLAB olarak sıralanabilir.

1.6.1. R Programlama Dili

R genel kamu lisansına (General Public License,GPL) sahip ücretsiz bir istatistiksel programlama dilidir (Altunkaynak, 2017:187).Temeli 1976 yılından bu yana istatistiksel programlama dili olarak geliştirilen S diline dayanmaktadır ancak R dili açık kaynak kodlu bir programlama dili olmakla beraber 1990’lı yıllarda Yeni Zelanda’daki Auckland Üniversitesi İstatistik Bölümü’nden Ross Ihaka ve Robert Gentleman

tarafından yazılmıştır (Baydoğan vd., 2014:4).R programının açık kaynak kodlu olması, kullanıcının programın arka yüzünde işleyen kodları görmesini sağlamakta aynı zamanda kullanıcının programın gelişimine katkı sağlamasına olanak tanımaktadır (Doğan,Uluman, 2016:616-617).

R programı çeşitli komutlarla veri madenciliği yöntemlerinden sınıflama, kümeleme ve birliktelik kurallarının algoritmalarının uygulanmasını sağlamaktadır. R programında veri madenciliği algoritmalarını uygulayabilmek için, öncelikle kullanacağımız algoritmalara ait paketlerin yüklenmesi ardından gerekli komutlarla aktif hale getirilmesi gerekmektedir (Altunkaynak, 2017:187-188).

R dilinin artı yönleri; gelişmiş grafik araçlarına sahip olması, ücretsiz olarak kullanılabilir ve amaçlar doğrultusunda farklı paketler eklenerek fonksiyonelliğinin geliştirilmesi iken eksi yönleri ise; öğrenmede güçlük çekilebilecek bir programa dili ve aynı zamanda bu dili kullanabilmek için öncesinde birtakım bilgi birikimine sahip olunmasıdır (Özdemir vd., 2010:296).

1.6.2. RapidMiner (YALE)

RapidMiner Dortmund Teknoloji Üniversitesi Yapay Zeka Biriminde geliştirilmiş bir yazılımdır.2007 yılı itibariyle Java dili kullanılarak geliştirilmiş bir yazılım dili olan Yale, RapidMiner adı altında kullanılmaya başlamıştır. Veri madenciliği ve makine öğrenimini kapsayan Rapidminer programı 22 farklı dosya formatındaki veriyi işleye bilme özelliğine sahip olup 3D görsellerinin yer alması ile kapsamlı bir yazılımdır (Kaya,Özel, 2014:49).

Veri madenciliği tekniklerine ait algoritmaları da içerisinde bulunduran RapidMiner programı 400 den fazla algoritmaya sahiptir (Dener vd., 2009:789).

1.6.3 Konztanz Information Miner (KNIME)

KNIME Almanya'daki Konztanz üniversitesi tarafından geliştirilen açık kaynak kodlu ücretsiz bir programdır. KNIME üniversitede araştırma yapmak amacıyla kullanılmakla birlikte geliştirilen çok sayıda veri analiz tekniklerini programa aktarılmasına da olanak sağlamaktadır. Kullanıcıya görsel veri akışını gösteren, veri hazırlama, temizleme ve madencilik(sinir ağları, birliktelik kuralları, karar ağaçları) gibi olanaklar sağlayan

csv,arff,pmml gibi formatları destekleyen veri keşif platformudur (Tekerek, 2011:166).

1.6.4. WEKA

Weka Yeni Zelanda'daki Waikato Üniversitesi'nde geliştirilen ve adını buradan alan, ücretsiz, açık kaynak kodlu, GNU ve GPL lisansına sahip, java tabanlı, imgesini ise Yeni Zelanda'daki bir kuş türünden alan bir veri madenciliği programıdır. Weka programı veri madenciliği tekniklerinin uygulanmasına olanak tanımakla kalmayıp çeşitli veri ön işleme araçları ve görselleştirmeler içererek kullanıcıya alternatif yollar sunmaktadır. Aynı zamanda analize en uygun olan çıktının alınmasını sağlamaktadır. Program kullanıcıya dört farklı arayüz sunarak analizler için uygun olan arayüzün seçilmesine imkân sağlamaktadır (Aydemir, 2019:57-58).

1.6.5. MATLAB

MATLAB, Cleve Barry Moler tarafından 1970'li yılların sonunda yazılmıştır. Moler'ın MATLAB'ı yazma amacı öğrencilerinin LINPACK ve EISPACK gibi kütüphanelere FORTRAN dilini öğrenmeden erişebilmelerini sağlamaktır. Günümüzde de sanayi ve üniversiteler başta olmak üzere oldukça geniş kapsamlı kullanıma sahip olan MATLAB'ı 1984 yılında Little ve Moler Mathworks şirketi kurularak ticari bir yazılım haline gelmiştir. MATLAB veri analizi, algoritma geliştirme, görselleştirme, sayısal hesaplama gibi binlerce komut ve fonksiyonu gerçekleştirmektedir (Akaslan,Taşkın, 2015:14).

1.7. Veri Madenciliği Kullanım Alanları

Günümüzde veriye ulaşılabilirliğin artmasıyla gerek akademik olarak lisansüstü tezlerde yahut makalelerde gerek sektörsel olarak çeşitli işletme alanlarında veri madenciliğinin kullanımı giderek artış göstermektedir. Nitekim verinin büyük boyutlu olması bir analiz sürecinden geçip işlenerek ortaya yeni ve önemli bilgiler çıkarması bir avantaj olarak değerlendirilebileceğinden veri madenciliği kullanım alanları oldukça kapsamlı ve geniştir (Polat, 2022: 4).

Veri madenciliği kullanım alanları aşağıdaki gibi sıralanabilir (Baykal, 2006:97; Aktürk,Korukoğlu, 2008:114; Silahtaroglu, 2016:15-16):

-Pazarlama

- ✓ Müşterilerin satın alma örüntülerinin belirlenmesi,
- ✓ Müşterilerin demografik özellikleri arasındaki bağlantıların bulunması,
- ✓ Posta kampanyalarında cevap verme oranının artırılması,
- ✓ Mevcut müşterilerin elde tutulması, yeni müşterilerin kazanılması,
- ✓ Pazar sepeti analizi,
- ✓ Müşteri ilişkileri yönetimi,
- ✓ Müşteri değerlendirme,
- ✓ Satış tahmini,

-Bankacılık

- ✓ Farklı finansal göstergeler arasında gizli korelasyonların bulunması,
- ✓ Kredi kartı dolandırıcılıklarının tespiti,
- ✓ Kredi kartı harcamalarına göre müşteri gruplarının belirlenmesi,
- ✓ Kredi taleplerinin değerlendirilmesi,

-Sigortacılık

- ✓ Yeni poliçe talep edecek müşterilerin tahmin edilmesi,
- ✓ Sigorta dolandırıcılıklarının tespiti,
- ✓ Riskli müşteri örüntülerinin belirlenmesi,

-Risk Yönetimi ve Dolandırıcılık Saptama

- ✓ Kredi kartı dolandırıcılığı,
- ✓ İnternet işlemleri,
- ✓ Bilgisayar sistemleri ve bilgisayar ağlarına girme,
- ✓ Üyelik abone dolandırıcılığı,

-Tıp

-Biyoloji

-Spor alanları

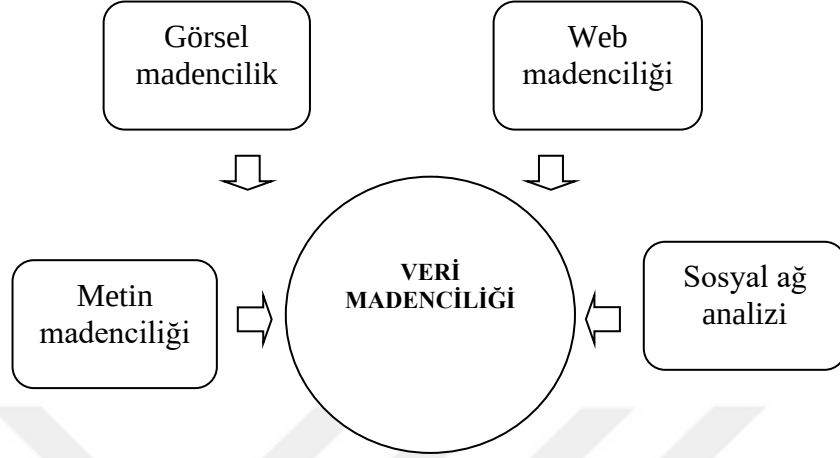
-Astronomi

-Borsa

-Arama motorları

-Web sitesi analizleri

Verinin toplanma şekli, veri madenciliğine ilişkin alt alanları ortaya çıkarmıştır. Bu alt alanlardan bazıları aşağıdaki şekilde verilmiştir (Altunkaynak, 2017:15).



Şekil 1.4:Veri Madenciliğine İlişkin Alt Alanlar

Kaynak: (Altunkaynak,2017,15)

İnternet kullanımının artışıyla birlikte teknolojik cihazlar aracılığıyla bilgiye ve internete erişim hızı bilgilerin web’de toplanmasına olanak vermiştir. Web madenciliği veri madenciliği süreçlerini kapsayan web’deki sitelerden bilgi toplayan veri madenciliğinin alt dalıdır. Web madenciliği web’deki keşfedilmemiş belgeleri bulur çıkarımlar yaparak genellemeler yapar (Etzioni, 1996:65).Web madenciliği kullanıcı hakkında bilgileri toplayarak web sitesi sahibi yararına çıkarımlar yapar. Bu çıkarımlardan yola çıkarak web sitesinde kullanıcı profili baz alınarak yeni düzenlemeler yapılarak kullanıcı memnuniyeti sağlanabilir (Arslan, 2008:38).

Metin madenciliğinde ele alınan veri grubu, veri madenciliğinin ilgilendiği veri grubundan farklıdır. Veriler yapılandırılmış ve yapılandırılmamış veriler olarak gruplandığında yapılanmış veri grubunda düzenlenmiş veriler yer alırken yapılanmamış veri grubunda metinsel veriler (belge, mektup, doküman, kâğıt) yer alır. Metin madenciliği, yapılanmamış veri grubu ile ilgilenerek bu veriler içerisinde veri madenciliğinde olduğu gibi önemli ve gerekli bilgileri ortaya çıkarır (Oğuzlar, 2011:1-2). Metin madenciliğinin amacı kullanılabilecek yeni bilgiler elde etmek ve elde edilen bilgilerle en iyi sonuca ulaşmak için metin analizi yapmaktır (Özyiğit, 2022:644). Metin madenciliği, metin koleksiyonu oluşturma yani ilgilenilen konuya yönelik verileri

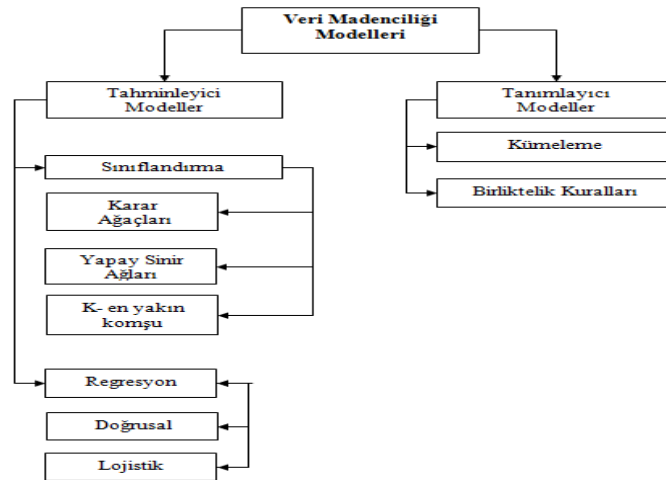
toplama ile başlayarak önışleme ve veri madenciliđi olmak üzere üç aşamadan oluşur ve yorumlanır. Sonrasında araştırmacıya katkı sağlama boyutuna gelerek son bulur (Oğuzlar, 2011:9).

Görsel madencilik, kamera video veya resim görüntülerinin veri madenciliđine uygun veri tipine dönüştürüldükten sonra analiz edilmesidir. Örneđin spor müsabakalarının görsel olarak kayıt altına alınması daha sonra bu müsabakalardan önemli çıkarımlar yapmamızı sağlar. Bir futbolcunun maç esnasında sergilediđi performansa dayanarak gol pozisyonu koşma mesafesi gibi çeşitli kriterler bakımından değerdendirilerek oyuncunun performansı ölçülmekte ve buna bađlı olarak oyuncunun sahadaki etkin olduđu alan belirlenmektedir (Altunkaynak, 2017:16).

Sosyal ađ analizi karmaşık ve düzensiz yapılardan oluşan insan ilişkilerini içinde barındıran veri madenciliđinin alt alanlarından biridir. Sosyal ađ analizinin görsellik açısından zengin olması ađları oluşturan bađlantıları çok karmaşık ađlarda dahi programın sunduđu yakınlştırma olanakları ile kolaylıkla görmemize olanak verir (Gürsakar, 2009:182).

1.8. Veri Madenciliđi Modelleri

Veri madenciliđi modelleri kullanılan algoritmalar ve tekniklere göre tahminleyici ve tanımlayıcı modeller olmak üzere iki gruba ayrılır (Silahtaroglu, 2016:49).



Şekil 1.5: Veri Madenciliđi Modelleri

Kaynak: (Polat, 2022:7)

1.8.1.Tahminleyici Modeller

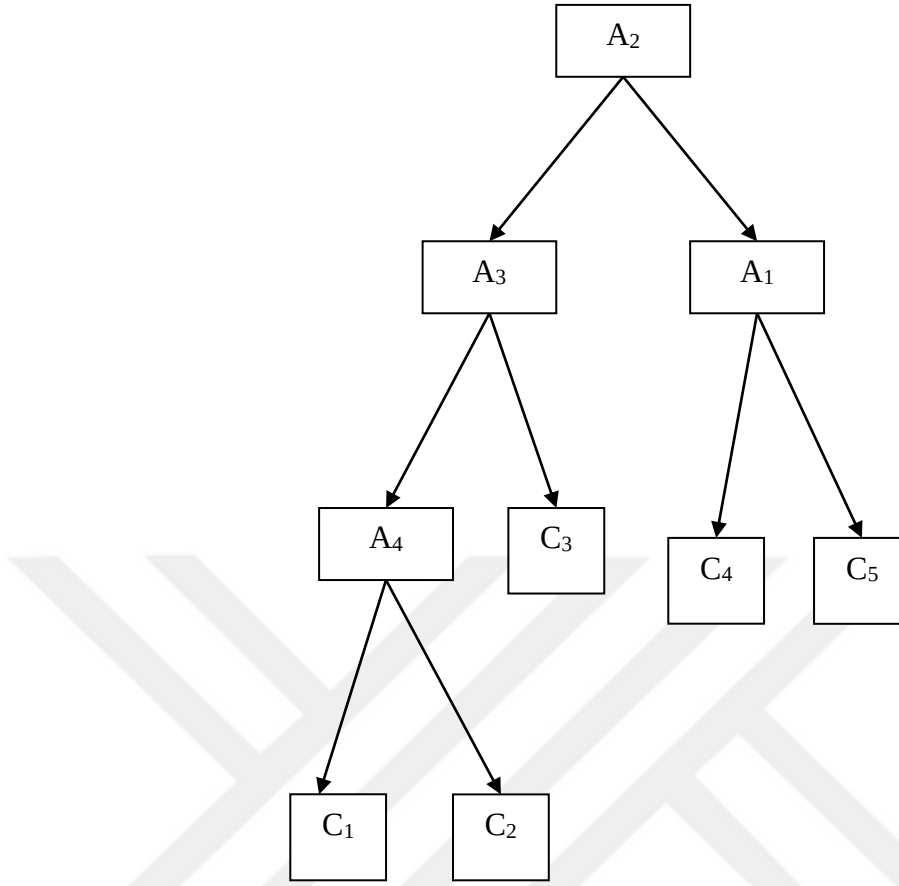
Tahminleyici modeller önceden bilinen veriler için tahmin edici bir model oluşturarak bu modelin sonuçlarının, bilinmeyen araştırma konumuzun tahmini için kullanılmasını sağlamaktadır(Aydın,2007:10).Tahminleyici model bir insan öğrenmesine benzetilebilir çünkü insanlar yaşadığı çevrede, işte, okulda bir şeyleri görerek öğrenir ve öğrendikleriyle bağdaştırmaya çalışarak tahminler yapar. Tahminleyici modelde konuyla ilgili veritabanları incelenerek önceden sahip olunan veritabanı özelliklerine göre tahminler yapılarak bu veritabanlarının birbirlerine benzerlikleri sınıflamaya çalışılır (Silahtaroglu, 2016:50).

1.8.1.1.Sınıflandırma Teknikleri

Sınıflandırma veri madenciliği tekniklerinden en çok bilinen tekniklerden olup belirlenen niteliklere göre girdilerin sınıflara ayrılması işlemine dayanır. Sınıflandırma ele alınan veri setinde verilerin eğitimi ve modelin testi olarak iki aşamada gerçekleştirilir. Eğitim veri setinden çıkarımlar yaparak modelin oluşmasına olanak sağlarken test ise, oluşturulan modelin kesinliğini yani doğruluğunu analiz etmektedir. Böylece test örneklerinin en iyi sınıfı modelin tahmin ettiği sınıf ile karşılaştırılır sonrasında sınıfı bilinmeyen yeni örnekleri hangi sınıfa atayacağını tahmin etmek üzere kullanılır (Emel, Taşkın, 2005:224).

1.8.1.1.1. Karar Ağaçları

Sınıflandırma algoritmalarından olan karar ağaçları, adından da anlaşılacağı üzere bir ağaca benzeyen iki ya da daha fazla dala ayrılabilen, sınıflandırmaya dayalı tahmin için kullanılan karar düğümlerinden ve kararları tutan yaprak düğümlerden oluşur. Bir neticeye kısacası bir karara varabilmek için ağacın en üstündeki kök düğüm ile yaprak düğümler arasındaki yol izlenmelidir(Balaban,Kartal,2018:90).Karar ağaçları bir sınıflandırma aracı olup belirli olasılıkları baz alarak başka kararları etkilediği düşünülen görselleştirilmiş yapılardır (Akpınar, 2017:242).



Şekil 1.6:Örnek Karar Ağacı

Kaynak: (Silahtaroglu,2016:69)

1.8.1.1.2. Yapay Sinir Ağları

Yapay sinir ağları geliştirilmiş bilgisayar sistemlerine dayalı olarak öğrenme yolu ile yeni bilgiler keşfetme, sınıflandırma, genelleme, optimizasyon, ilişkilendirme, özellik belirleme gibi konularda uygulanan elde ettiği bilgiler ile deneyimler kazanıp benzer konularda benzer karar veren bir tekniktir (Öztemel, 2006:29).

1.8.1.1.3. K-En Yakın Komşu

K-en yakın komşu algoritması bir örneğin sınıfını, belirlenen k değerine göre diğer örneklerle olan uzaklığının bulunmasıdır(Balaban,Kartal, 2018:54). K-en yakın komşu algoritması diğer örneklerle olan uzaklığı tespit ettikten sonra küçükten büyüğe sıralama işlemi yaparak en küçük uzaklığa sahip k tane örnekte en çok tekrar eden sınıfa atamasını yapar. Burada önemli olan k değerinin seçimidir. Ele alınan veri setine göre k

değeri küçüklüğü ve büyüklüğü belirlenerek daha duyarlı ve daha iyi sonuçlar verecek şekilde seçilmelidir (Altunkaynak, 2017:110).

1.8.1.2.Regresyon

Regresyon analizi tahminleyici modellerden sınıflandırma teknikleri arasında yer alan bağımlı değişken ile bağımsız değişken arasında ilişkiyi açıklamaya çalışan bir analiz olup yapay sinir ağlarına alternatif olarak kullanılabilir (Akpınar, 2017:278).

Klasik doğrusal regresyon modeli ve lojistik regresyon modelinde bağımsız değişkenler ile,bağımlı değişkenler ile arasındaki ilişki modellenmektedir.Ancak doğrusal regresyon analizinde bağımlı değişken sürekli iken lojistik regresyon analizinde bağımlı değişken kategorik,bağımsız değişken kategorik, sürekli veya kategorik ve sürekli değişkenin bir karması olabilmektedir(Bayram,2015:211).

1.8.2.Tanımlayıcı Modeller

Tanımlayıcı modeller ilgili veri setiyle betimlemeler yaparak veri setindeki özellikleri tespit edip birbiriyle ilişkili örüntüler sunmaktadır (Bozkurt Uzan,Özüçaylıyan, 2020:3470).Bu modeller verinin tanınmasını, benzer özelliklerin kümelenmesini, verilerin birbirleriyle olan ilişkilerini, birlikte hareketini ele almaktadır (Erten, 2015:16-17).

1.8.2.1.Kümeleme Teknikleri

Tanımlayıcı modeller arasında yer alan kümeleme teknikleri içerdiği çeşitli algoritmalar sayesinde ilgilenilen veri setinin benzer nesnelerdeki gruplara bölünmesi işlemidir (Karaatlı,Altıntaş, 2018:873).Kümeleme analizi içerisinde kullanılan farklı algoritmalar ile tahmin eden ve öğrenen bir yöntem olmamakla birlikte veriyi tanımlamaya yardımcı olan bir yöntemdir(Köse, 2018:35).

Kümeleme yönteminde başlangıçta tüm gözlemler ayrı birer küme olarak değerlendirilirken sonraki aşamada kümeler arasındaki uzaklıklar hesaplanarak birbirine yakın olan kümeler birleştirilir ve yeni kümeler oluşturulur (Altunkaynak, 2017:127).

Kümelemede küme sayısı önceden bilinmemekle birlikte analiz sonucunda ortaya çıkmaktadır. Aynı zamanda elde edilen küme ilgilenilen veri setine özel olduğu için başka veriler için kullanılması doğru olmayacaktır (Çelik, 2009:37).

Kümeleme analizinin aşamaları aşağıdaki gibi sıralanabilir (Hasanlı, 2014:24):

- ✓ Veriler arasındaki benzerliğin tespiti için kullanılacak ölçü ve değişkenlerin belirlenmesi,
- ✓ Benzerlikleri belirlenen verilerin kümelenmesi,
- ✓ Oluşturulan kümelerin uygunluğunun tespit edilmesi,
- ✓ Uygunluğu tespit edilen kümelerin istatistiksel olarak geçerliliğinin belirlenmesi işlemlerine dayanmaktadır.

1.8.2.2. Birliktelik Kuralları

Olayların birlikte gerçekleşme olasılıklarını ortaya koyan veri madenciliği modelleri birliktelik kuralları olarak adlandırılır (Özkan, 2020:217). Birliktelik kuralları ilk olarak sık geçen öğeleri bulur sonrasında ise bu öğeler arasındaki birliktelikleri ortaya koyarak belirlenen destek ve güven seviyelerine göre kurallar üretir (Ünsal, 2020:109).

Birliktelik kuralları temelinde pazar sepeti analizine dayanmaktadır. Market satışları göz önüne alındığında müşterilerin hangi malları birlikte satın aldığı tespit edilerek market rafları belirlenen bu birlikteliğe göre düzenlendikten sonra satışların artması beklenebilir. Ancak birliktelik kuralları birbiriyle ilişkili kurallar üretirken buna bağlı olarak ilginç bilgilerde sunmaktadır. Sıklıkla örnek verilen 1980’li yıllara dayanan bebek bezi ve bira satışlarının birbiriyle ilişkili olduğu sonucu ilginç bilgiler sunduğunun göstergesidir (Akpınar, 2017: 202).

Bölüm 2 de tanımlayıcı modellerden olan birliktelik kuralları ve tahmin edici modellerden sınıflandırma teknikleri adı altında olan yapay sinir ağları detaylı olarak anlatılacaktır.

İKİNCİ BÖLÜM

VERİ MADENCİLİĞİNDE YAPAY SİNİR AĞLARI VE BİRLİKTELİK KURALI ANALİZİ

2.1. Yapay Sinir Ağları

Değişen dünya ve gelişen teknoloji ile birlikte insanların talepleri doğrultusunda bilgiyi elde etme, edinilen bilgiyi işleme ve ileriye yönelik tahminlerde bulunma yeteneklerine sahip zeki sistemler geliştirilmiş bu sistemler yapay zeka biliminin de gelişmesine olanak tanımıştır. Matematiksel olarak formüle edilebilen, problemleri çözümü yapabilen geleneksel bilgisayar sistemlerinin yanı sıra özellikle belirli bir formülasyon kullanılarak çözülemeyen problemlerin çözümü için yapay zeka sistemleri geliştirilmiştir (Öztemel, 2012:13-14).

Yapay zeka çalışmaları çerçevesinde ortaya çıkan ve yapay zeka çalışmalarına katkı sağlayan alanlardan biri olan yapay sinir ağları, yapay zekanın bir alt dalını oluşturmaktadır (Yurtoğlu, 2005: 3).

Veri madenciliğinde önemli bir rol oynayan, makine öğrenimi kapsamında geliştirilen yapay sinir ağları (Akpınar, 2017:56)bu bölümde genel olarak açıklanmıştır.

2.1.1.Yapay Sinir Ağlarının Tanımı

Yapay sinir ağları biyolojik sinir sistemlerindeki gibi sinir hücresi (nöronları) ağındaki karar sürecini genel bir şekilde simüle etmeye çalışan ağlardır. Bu ağlar bir sinir ağını simüle ederek geleneksel bilgisayarlara uyarlandığında yeni bir bilgisayar yapısı ve algoritması ortaya çıkarmıştır (Graupe, 2013:1).

Yapay sinir ağları insanların, insan benzeri makineler yaratma isteklerinden yola çıkarak geliştirilen insan beyninin modellenmesi üzerine kurulan sistemlerdir (Warner,Mısra, 1996:284).

Yapay sinir ağıları, sinirler arasındaki bağlantıların öğrenme ve eğitilme yeteneğine sahip girdi-çıkı sinirlerinden başka bir deęişle nöronlarından oluşan sistemdir (Arıkan Kargı, 2015:25). En genel haliyle yapay sinir ağıları beynin belirli bir görevi işleme ve yerine getirme sürecini, elektronik bileşenler aracılığıyla dijital ortama aktarma yapan yani bilgisayarlardaki yazılım sistemlerine uyarlayan bir makinedir (Haykin, 2009:32).

Yapay sinir ağıları kendi kendine öğrenme ile birlikte ezberleme ve edinilen bilgilerin birbirleriyle ilişkilerini bulma yeteneğine sahip, biyolojik sinir ağlarının taklit edilmesine dayanan bilgisayar programlarıdır (Elmas, 2011:23).

Biyolojik sinir sistemlerinden ilham alarak geliştiren yapay sinir ağıları, performansı maksimum kılacak şekilde geliştirilmiş basit hesaplama elemanlarının karmaşık bağlantılardan oluşan bir sistemdir (Akpınar, 2017:306).

Yapay sinir ağıları deneyimsel bilgiyi depolamak ve kullanım durumuna hazırlamak için kullanılan bir işlemci olmak ile birlikte beyine iki açıdan benzemektedir. Bu benzerlikleri şu şekilde sıralayabiliriz (Haykin, 2009: 32):

- ✓ Bilgi ağı tarafından öğrenme süreci yardımıyla ağıın çevresinden edinilir.

Bu ağı karmaşık bir sistemi oluşturan parçaların birbirleriyle etkileşimini göstermektedir (Gürsaka vd.,2008: 2).

- ✓ Çevreden elde edilen bilgilerin sinaptik ağırlıklar olarak da tanımlanan nöronların bağlantı güçleri depolamak için kullanılır.

2.1.2 Yapay Sinir Ağlarının Temel Özellikleri

Yapay sinir ağlarının sahip olduğı temel özellikler aşağıdaki gibi sıralanabilir (Öztürk,Şahin, 2018:29):

- ✓ Doğrusal Olmama
- ✓ Paralel Çalışma
- ✓ Öğrenme
- ✓ Genelleme
- ✓ Hata Toleransı ve Esneklik
- ✓ Eksik Verilerle Çalışma

- ✓ Çok Sayıda Değişken ve Parametre Kullanma
- ✓ Uyarlanabilirlik

Doğrusal Olmama: Yapay sinir ağ fonksiyonları nonlinear(doğrusal olmama) özelliğine sahiptir. Ağ üzerinde dağılmış belli türdeki doğrusal olmayan alt birimlerde, özellikle ağdan istenen eşleştirmenin denetimi veya tanımlamasında karşılaşılan nonlinear yapıyı matematiksel olarak işleyerek işlevin doğru bir şekilde yerine getirilmesini sağlar. Dolayısıyla işlevin doğru gerçekleşmesi için esneklik sunarak ağ parametrelerinin başarısını arttırmakta ve maliyetlerini minimuma indirmektedir(Efe, Kaynak, 2000:2-3).

Paralel Çalışma: Yapay sinir hücreleri diğer bir adıyla nöronları birbirleriyle süreç içinde karşılıklı katmanlar üreterek bağlantılar yaratırlar. Yaratılan bu bağlantılar paralel çalışma özelliğine sahiptir ve bu özellik sayesinde karmaşık problemleri çözümleyebilmenin yanında yapay sinir hücrelerinin süreç boyunca bozulma durumuyla karşılaştıklarında aksaklık yaşamadan sürecin sağlıklı bir şekilde ilerlemesini sağlamaktadır (Sert, 2014:20).

Öğrenme: Yapay sinir ağlarında öğrenme aşdaki belirlenen bağlantı ağırlıklarının tekrarına dayalı olarak yinelenmesi ile oluşmaktadır. Öğrenme gerçekleştikten sonra ağ elde ettiği bilgileri bünyesinde barındırır ve dolayısıyla elde edilen bilgiler uzun süre saklanmaktadır. Bu nedenle önceden saklanan bilgilerden yola çıkarak öğrenme yöntemi belli olmayan problemlerin çözümüne de olanak sağlamaktadır (Şen, 2004:5).

Genelleme: Yapay sinir ağlarında önceden öğretilen örnekleri baz alarak ağ, daha önceden görmediği örnekler hakkında bilgiler üreterek genelleme yapar(Öztemel, 2012:32).

Eksik Verilerle Çalışma: Yapay sinir ağları öğrenme becerisini kazanıp eğitildikten sonra eksik verilerle çalışabilir ve eksik bilgilere rağmen sonuç üretebilirler ayrıca eksik veriler ağın performansına etkide bulunmamaktadır(Sevimli Deniz, 2019:247).

Hata Toleransı ve Esneklik: Eksik bilgiler ile de çalışabilme özelliği olan yapay sinir ağları bu özelliğe dayalı olarak hatalara karşı toleranslıdırlar. Ağdaki bazı hücrelerin çalışamaz duruma gelmesi ve bozulması ağın çalışma durumunu etkilemez, ağ çalışmaya devam eder. Ancak bozuk olan hücrelerinin durumuna göre ağın çalışma performansında düşüşler olabilir öte yandan hücrelerin durumuna ağ eğitimi sırasında ağ kendisi karar verdiği için kullanıcı bu bilginin sahibi değildir (Öztemel, 2012:33).

Çok Sayıda Değişken ve Parametre Kullanımı: Yapay sinir ağları çok sayıda değişken ve parametre kullanımına olanak vermektedir.

Uyarlanabilirlik: Yapay sinir ağları sinaptik ağırlıklarını çevredeki değişikliklere uyarlayabilir. Başka ifade ile belirli ortamda çalışması için eğitilmiş bir sinir ağı, ortamdaki küçük değişikliklere karşı yeniden eğitilerek ortama göre uyarlanabilmektedir (Haykin, 2009:33).

2.1.3 Yapay Sinir Ağının Gelişimi

Yapay sinir ağları beynin çalışma yapısının sayısal bilgisayarlarda taklit edilmesi ile öne çıkmış ve ilk olarak biyolojik hücreler ya da nöron adı verilen yapıların matematiksel modellenmesi üzerinde çalışılmıştır(Efe,Kaynak, 2000:1).

1943 yılında ilk yapay sinir ağı modeli Warren McCulloch ve Walter Pitts tarafından geliştirilmiştir. Bu araştırmacılar elektrik devreleri aracılığıyla insan beynini hesaplama ve analiz yeteneğinden esinlenerek ilk olarak yapay sinir ağı hücresini geliştirmiş sonrasında bu hücreler aracılığıyla mantıksal ifadelerinin formüle edilmesinin mümkün olacağını göstererek birden fazla hücrenin birlikte çalışması gerektiği kanısına vardıldıktan sonra paralel çalışma ve öğrenme kurallarını belirlemişlerdir (Keskenler,Keskenler, 2017:11).

1949 da Hebb “Organization of Behavior” isimli kitabında öğrenme ile ilgili temel teoriyi ele aldı ve öğrenmeye dayalı olan aynı zamanda uyum sağlayabilen sinir ağları modeli için esas oluşturacak Hebb Kuralı’nı geliştirmiştir. Bu kural sinir ağının bağlantı sayısının değiştirilebilirliğine bağlı olarak öğrenmeyi gerçekleştireceğini öne sürmekteydi(Elmas, 2011:26).

1951 yılında Dean Edmonds ve Marvin Minsky Hebb kurallarını baz alarak elektro-mekanik bir öğrenme makinesi geliştirmiştir(Akpınar, 2017:312).

1954 yılında Farley ve Clark tarafından rassal ağlar ile adaptif tepki üretme kavramı ortaya atıldı. Ortaya atılan bu kavram 1958 yılında Rosenblatt tarafından gelişme kaydetti. Rosenblatt tarafından geliştirilen algılayıcı model yapay sinir ağları

gelişiminde önemli bir yer bir edinmiş ve daha sonra geliştirilen çok katmanlı algılayıcılara zemin hazırlamasından dolayı oldukça önemli bir yere sahip olmuştur (Öztemel, 2012:37).

1959 yılında yapay sinir ağlarının mühendislik alanındaki uygulamalarına ait ilk çalışmalar Stanford Üniversitesinde Bernand Widrow ve Marcion Hoff tarafından geliştirilen ADALINE (ADaptive LInear NEuron) ve MADALINE adlarını verdikleri yapay sinir ağı modelleri ile karşımıza çıkmaktadır. ADALINE modeli Rosentblatt tarafından geliştirilen algılayıcı modelin, gelişmiş halini yansıtırken kendisinden sonra yapılan çalışmalara temel oluşturmuştur. MADALINE ise gerçek problemlere uyarlanmış ilk sinir ağı olarak telefon hatlarındaki ses yankılarının giderilmesi amacıyla kullanılmıştır ve günümüzde de kullanılmaya devam etmektedir (Keskenler,Keskenler, 2017:12-13).

1969 yılında yayınlanan Perception adlı kitapta Minsky ve Popert tek katmanlı yapay sinir ağlarının XOR(ExclusiveOr) problemi gibi basit problemleri çözemeyeceğini kanıtladıktan sonra yapay sinir ağı gelişiminde duraklama yaşanmıştır(Bayır,2006:4).

1982’de John Hopfield tarafından yayımlanan Yapay Sinir Ağları ve Gelişen Kolektif Hesapsal Yetenekli Fiziksel Sistemler (Neural Networks and Physical Systems with Emergent Collective Computational Abilities) adlı kitabında yapay sinir ağlarının analog devre modelini geliştirerek analog-ikil dönüştürücü tasarımı, Gezgin Satıcı gibi problemleri ele alarak çözümler üretmiştir (Akpınar, 2017:322).

1986 yılında Rumelhart ve arkadaşlarının geliştirildiği çok katmanlı yapay sinir ağları için geriye yayılma algoritması XOR problemine çözüm üretmiştir ve önemli ölçüde kullanılan bir algoritma haline gelmiştir (Arıkan Kargı, 2015:30).

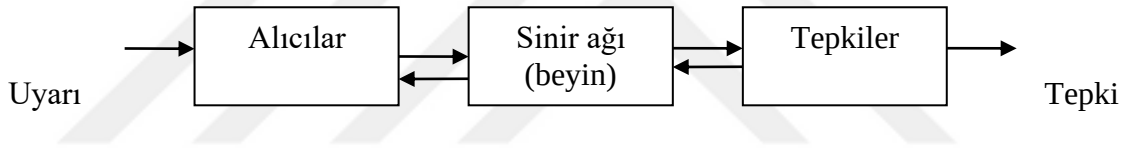
1988’de Broomhead ve LoweRadyal tabanlı fonksiyonlar modelini çok katmanlı algılayıcılara alternatif olarak geliştirdiler. Bu ağların daha gelişmiş şekli olan Probabilistik Ağlar ve Regresyon Ağları Specht tarafından geliştirilmiştir (Öztemel, 2012:40).

1987 yılından itibaren gerek sempozyumlar gerek konferanslar olmak üzere yapay sinir ağıları tartışılarak gelişimine katkıda bulunulmuştur. Bugün ise yapay sinir ağıları günlük hayatta kullanılan ve insanların hayatını kolaylaştıracak, pratik bilgileri insanlığa fayda sağlamak amacıyla üretmektedir (Öztemel, 2012:40).

2.1.4.Biyolojik Sinir Sistemleri

Biyolojik sinir sistemlerinden esinlenerek geliştirilen yapay sinir ağlarının yapısını ve işleyişini kavrayabilmek adına öncelikle biyolojik sinir sistemlerinin yapısını ve işleyişini tanımak faydalı olacaktır.

İnsan beynindeki sinir sistemi sürekli olarak bilgiyi alan, algılayan ve algılanan bilgiler bağlamında karar veren bir sistemdir (Haykin, 2009:36).

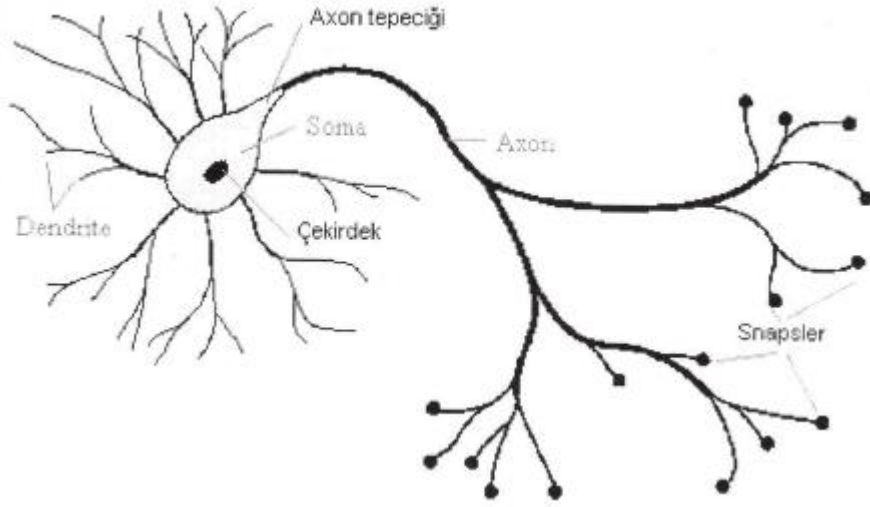


Şekil 2.1: Sinir Sisteminin Blok Şeması Gösterimi

Kaynak: (Haykin, 2009:36)

Çevreden başka bir deyişle dış ortamdan gelen uyarılar elektriksel sinyallere dönüşerek sinir ağına (beyine) alıcılar vasıtasıyla iletilir ardından sinir ağına(beyinde) üretilen elektriksel sinyaller çıktı olarak uygun tepkiler üretir. Sinir ağı alıcılar ve tepkiler arasında ileri ve geri besleme yaparak uygun tepkiler üreten bir merkezdir (Arıkan Kargı, 2015:35-36).

İnsan beyni birbirine bağlı 10 milyardan fazla biyolojik sinir hücresinin daha özel bir ifadeyle nöronunun bir koleksiyondur. Bu nöronlardan her biri bilgi almak, işlemek ve iletmek için biyokimyasal reaksiyonları kullanan hücrelerdir (Abraham, 2005:901).



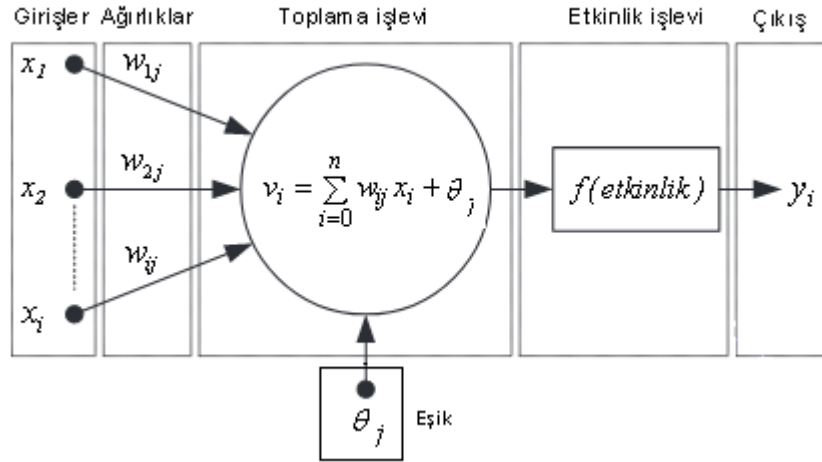
Şekil 2.2: Bir Biyolojik Sinir Hücresinin Yapısı

Kaynak: (Öztemel, 2012:47)

Biyolojik bir sinir hücresi sinapsler, soma, axon ve dentrite olmak üzere dört yapıdan oluşmaktadır. Sinapsler sinir hücreleri arasındaki elektrik sinyallerinin geçmesini sağlayarak sinyalleri somaya ulaştırır. Soma işleme tabi tuttukten sonra sinir hücresi kendi elektrik sinyalini oluşturarak axon aracılığıyla dentrite gönderir. Dentritler ise bu sinyalleri diğer hücrelere göndermektedir (Öztemel, 2012:47).

2.1.5.Yapay Sinir Ağı ve Öğeleri

Yapay sinir ağları genel itibariyle insan beyninden esinlenerek geliştirilmiştir. Böylece yapay sinir ağlarının temel elemanlarını yapay nöronlar olarak da adlandırılabilen yapay sinir ağı hücreleri oluşturmaktadır. Yapay sinir ağı hücreleri aralarında bağlantılar kurarak ve tabakalar şeklinde gruplar oluşturarak yapay sinir ağlarını ortaya çıkartmaktadır (Yurtoğlu, 2005:12).



Şekil 2.3: Yapay Bir Sinir (Düğüm)

Kaynak: (Elmas, 2011:31)

Yapay sinir ağlarının en küçük yapı taşı olan yapay sinir ağı hücresi McCulloch ve Pitts tarafından 1943 yılında geliştirilmiştir. Bir yapay sinir ağı hücresinin matematiksel gösterimi Şekil 2.3’de verilmiştir. Buna göre girdi değerler $x(i)$ ile sembolize edilerek $i = 0, 1, 2, \dots, n$ değerlerini almaktadır. Girdi değerleri tek tek bağlantı ağırlıklarıyla çarpılarak $w(i)$ şeklinde sembolize edilmektedir. Bu çarpımlar toplanarak transfer fonksiyonuna gönderilmekte ve bir sonuç üretilmektedir. Nihayetinde üretilen sonuçlar çıktıya dönüştürülmektedir (Yurtoğlu, 2005:14).

Yukarıdaki verilen bilgiler doğrultusunda bir yapay sinir ağı hücresinin öğeleri aşağıda açıklanmıştır:

Girdiler: Yapay sinir hücresinin öğrenmesi istenilen örnekler göz önüne alınarak dış dünyadan başka hücrelerden veya hücrenin kendisinden bilgiler elde etmektedir ve bu bilgiler girdi olarak adlandırılmaktadır (Öztemel, 2012:49). Girdiler aşağıdaki şekilde gösterilmektedir:

$$i = 0, 1, 2, 3, \dots, n \text{ olmak üzere; } X_i = [x_1, x_2, x_3, \dots, x_n] \quad (2.1)$$

Ağırlıklar: Girdilerin hücreler arasındaki bağlantılarını sağlamakta ve gelen bilginin önemini, pozitif veya negatif etkisini göstermektedir. Ancak ağırlıkların büyük veya

küçük değerlere sahip olması herhangi bir anlam ifade etmez. Aksine ağırlık değerinin sıfır olması durumu ağ için önemli olabilmektedir (Öztemel, 2012:49).

Toplama Fonksiyonu: Girdilerin ağırlıklandırılması ile toplama fonksiyonu oluşmaktadır. Toplama fonksiyonu sinirdeki her bir ağırlığın tek tek ait olduğu girişlerle çarpımının toplanması ile oluşan bu değer eşik değeri ile toplanarak transfer fonksiyonuna gönderilmesini sağlar. Bu fonksiyon toplama işlevine ek olarak minimum, maksimum ve normalleştirme algoritmalarını da uygulamaktadır (Elmas, 2011:31). Aşağıda Tablo 2.1’de toplama fonksiyonunda gerçekleşen uygulama örnekleri gösterilmiştir (Öztemel, 2012:50):

Tablo 2.1: Toplama Fonksiyonu Örnekleri

Net Giriş	Açıklama
Çarpım Net Girdi = $\prod_i x_i w_i$	Girdi değerleri tek tek ağırlık değerleri ile çarpılır ardından elde edilen sonuçlar birbirleri ile çarpılarak net girdi bulunur.
Maksimum Net Girdi = $\text{Max}(x_i w_i), i=1, \dots, N$	$i=1, \dots, N$ ’e kadar göstermek üzere N sayısı kadar girdi içinden ağırlıklar ile çarpılarak en büyük değer net girdi olarak seçilir.
Minimum Net Girdi = $\text{Max}(x_i w_i), i=1, \dots, N$	$i=1, \dots, N$ ’e kadar göstermek üzere N sayısı kadar girdi içinden ağırlıklar ile çarpılarak en küçük değer net girdi olarak seçilir.
Çoğunluk Net Girdi = $\sum_i \text{sgn}(x_i w_i)$	N adet girdi içinden ağırlıklar ile çarpılarak pozitif ve negatif değerlerin sayısı bulunur ve büyük olan değer net girdi olarak seçilir.
Kümülatif Toplam Net Girdi = $\text{Net (eski)} + \sum_i (x_i w_i)$	Hücreye gelen bilgiler ağırlıklı olarak toplanır. Ardından önceden gelen bilgiler ile toplanarak kümülatif toplam elde edilir. Elde edilen bu değer net girdi olarak seçilir.

Kaynak: (Öztemel, 2012:50)

Aktivasyon(Transfer)Fonksiyonu: Toplama fonksiyonundan gelen girdiyi bir süreçten geçirerek hücre çıktısını belirleyen, hücre modellerine ve işlevlerine göre çeşitliliği bulunan doğrusal olmayan fonksiyonlardır. En iyi performansı sağlayan aktivasyon fonksiyonu deneme-yanılma prosedürü kullanılarak belirlenir. Bu seçim önemli ölçüde yapay sinir ağının verilerine ve öğrenmesine bağlı olarak gerçekleşmektedir(Keçe, 2006:33).

Yapay sinir hücresinin davranışını belirleyen aktivasyon fonksiyonu, biyolojik sinir hücresinde olduğu gibi toplamın belli bir değer üzerinde seyrettiği durumlarda hücrenin kısa süreli bir darbe göndermesini elde etmek için çeşitli fonksiyonlardan yararlanmaktadır (Efe & Kaynak, 2000:7).

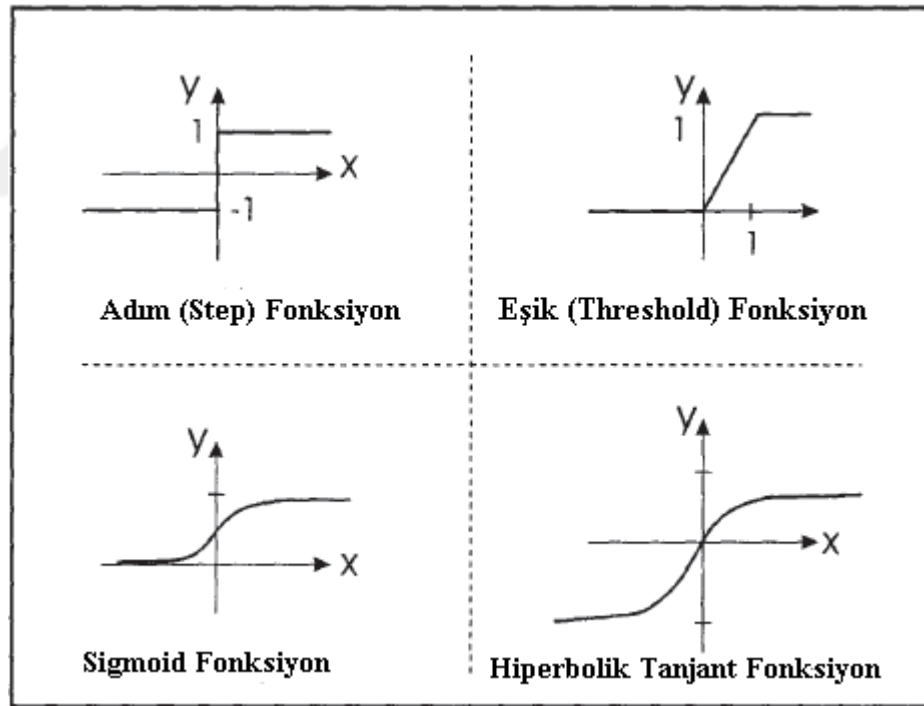
Aktivasyon işlevi, daha karmaşık görevleri öğrenmek için girdiye uygulanan doğrusal olmayan bir dönüşümdür. Bu fonksiyon uygulanmadığında çıkış sinyali basit doğrusal bir fonksiyon haline gelir. Bu durumda yapay sinir ağının doğrusal olmayan durumları öğrenmesi mümkün olmayacaktır. Bu bağlamda karmaşık veri kümelerinde doğrusal olmayan fonksiyonlar aktivasyon fonksiyonun işlevini yerine getirmesi adına kolaylık sağlamaktadır. Aktivasyon fonksiyonunun sürekli ve türevlenebilir olması optimizasyon algoritmasının uygun ağırlıklar bulma işlevinde kolaylık sağlayacaktır (Gonzalez, 2007:5).

Aktivasyon fonksiyonu, gelen hücrelerin her biri için farklı bir ağırlık ile çarpıldıktan sonra toplanarak meydana gelen değerleri göstermektedir. Burada farklı aktivasyon fonksiyonları kullanmak mümkündür ancak genellikle aktivasyon fonksiyonundan çıkışa gönderilecek her hücre için aynı aktivasyon fonksiyonu kullanılmaktadır (Sarle,1994:2).

Yapay sinir hücreleri için yaygın olarak kullanılan aktivasyon fonksiyonları doğrusal, adım(step) fonksiyonu, eşik fonksiyonu, sigmoid fonksiyon, hiperbolik tanjant fonksiyonu olarak sıralanmaktadır (Yurtoğlu,2005:18; Sarle,1994:2; Beale,vd., 2017:1-6; Öztemel, 2012:51). Bu fonksiyonlara ait açıklamalar aşağıda verilmiştir:

- ✓ **Doğrusal Fonksiyon:** Bu fonksiyon gelen girdiler üzerinde hiçbir değişiklik yapmadan olduğu gibi çıktıya gönderilmesini sağlamaktadır.

- ✓ **Adım (Step) Fonksiyonu:** Bu fonksiyona gelen net girdi değerinin eşik değerinin altında veya üstünde olması durumuna göre hücre çıktısı 1 veya -1 değerini almaktadır.
- ✓ **Sigmoid Fonksiyonu:** Bu fonksiyon artı ve eksi sonsuz arasında bir değere sahip olan girişleri alabilirken çıktı olarak 0 ve 1 arasında değer almaktadır.
- ✓ **Eşik Fonksiyonu:** Bu fonksiyon net bilgi değerlerinin 0 veya 1 den büyük ya da küçük olma durumuna göre değer alırken aynı zamanda 0 ile 1 arasında da değer almaktadır. Bunun dışında herhangi bir değer almaları söz konusu değildir.
- ✓ **Hiperbolik Tanjant Fonksiyonu:** Bu fonksiyonda gelen net girdi değerleri, tanjant fonksiyonundan geçirilerek hesaplanmakta ve -1 ile 1 arasında değerler almakta, negatif değerler alabilen bir değişken öngörüldüğünde kullanılması daha iyi sonuçlara ulaşmayı sağlayacaktır (Gonzalez, 2007:6). En çok kullanılan aktivasyon fonksiyon grafiği Şekil 2.4’de gösterilmiştir:



Şekil 2.4: En Çok Kullanılan Transfer Fonksiyonları

Kaynak: (Yurtoğlu, 2005:18)

Yukarıda yaygın olarak kullanılan aktivasyon fonksiyonlarının matematiksel gösterimi aşağıda Tablo 2.2’de gösterilmiştir (Arıkan Kargı, 2015:42-43).

Tablo 2.2: En Çok Kullanılan Transfer Fonksiyonları Denklemleri

Doğrusal Fonksiyon	$F(x) = x$
Adım Fonksiyonu	$f(x) = \begin{cases} 1 & x > 0 \\ -1 & x \leq 0 \end{cases}$
Eşik Fonksiyonu	$f(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$
Sigmoid Fonsiyonu	$f(x) = \frac{1}{1 + e^{-x}}$
Hiperbolik Tanjant Fonksiyonu	$f(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})}$

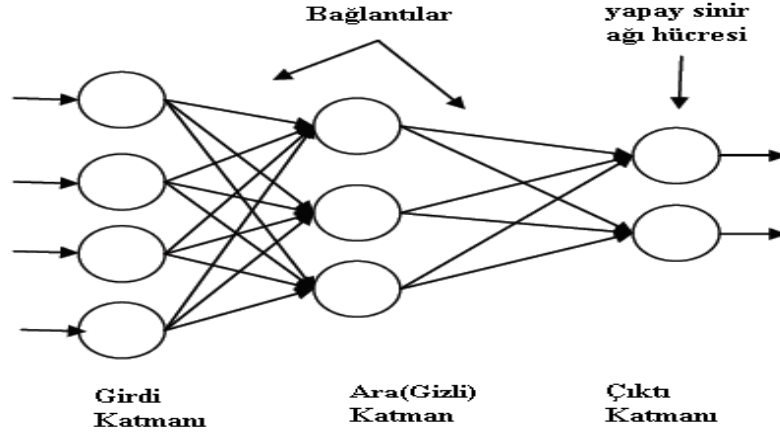
Kaynak: (Arıkan Kargı, 2015: 42-43)

- ✓ **Çıktı:** Yapay sinir hücresine ilk aşamada gelen girdilerin ağırlıklı toplamalarının transfer fonksiyonundan çıkmasıyla elde edilen sonuçlar çıktı olarak adlandırılmaktadır. Elde edilen çıktı sinir ağı kullanıcısına yalnızca ağıın sonuçlarını iletmekle kalmayıp, başka bir hücreye de girdi olabilmektedirler (Telli, 2016:53).

2.1.6.Yapay Sinir Hücrelerinin Mimari Yapısı

Yapay sinir hücrelerinin birbirine bağlanması, katmanlı bir yapıyı ortaya çıkarmakla birlikte yapay sinir ağlarını oluşturmaktadır (Koç vd.,2004:3354).Yapay sinir ağları, yapay sinir hücrelerinin birbirine bağlanması ile oluşturulduğundan ağdaki hücre sırasını belirtmek adına katman terimi kullanılmaktadır (Yıldız,Yezegel,2010:150).

Katman halinde düzenlenmiş yapay sinir hücrelerine sahip sinir ağları; girdi olarak kullanılan bağımsız değişkenlerin sayısından oluşan girdi, çıktı olarak tahmin edilen değişkenlerden oluşan çıktı ve bu iki katman arasında kalan gizli (ara)olmak üzere üç katmandan oluşmaktadır (Detienne,vd., 2003:242). Aşağıda Şekil 2.5’de katmanlar halinde oluşan bir yapay sinir ağı gösterilmiştir:



Şekil 2.5: Yapay Sinir Ağı Katmanları

Kaynak:(Yıldız ,Yezegel, 2010:151)

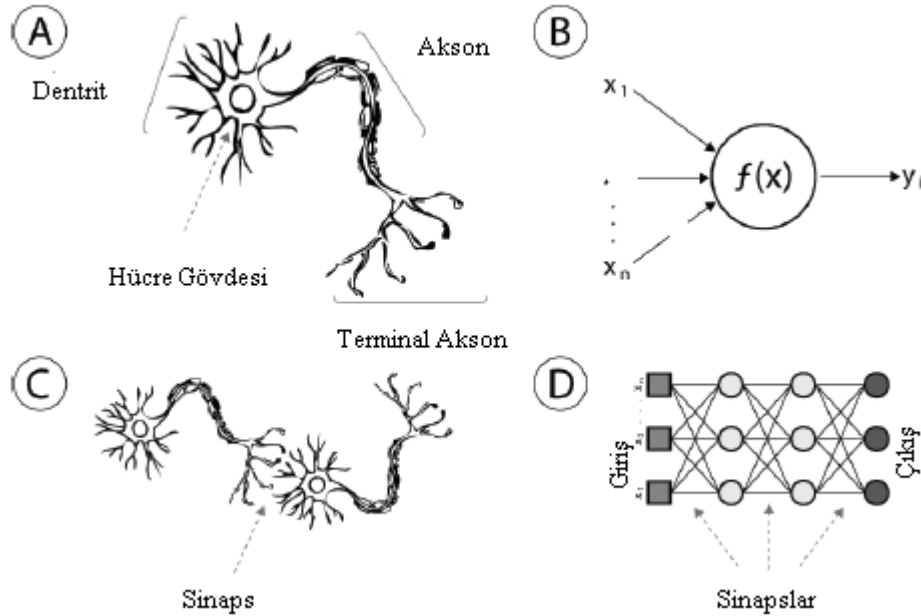
Yapay sinir ağını oluşturan katmanlar aşağıda verilmiştir:

- ✓ **Girdi Katmanı:** Girdi katmanı yalnızca söz konusu girdi değerlerini bir sonraki katmana iletmekle görevlidir. Bu katman dış dünyadan gelen girdiler üzerinde herhangi bir işlem gerçekleştirmemektedir (Öztemel, 2012:53).Girdi değerlerini bir sonraki katmana iletmekle görevli olduğu için bu katman etkisiz görülerek bazı araştırmacılar tarafından yapay sinir ağı katman sayısına dâhil edilmemektedir (Ataseven, 2013:102-103).
- ✓ **Ara (Gizli) Katman:** Çok sayıda yapay sinir ağı hücresi içeren bu katman içerdiği bu hücreler sayesinde ağ içindeki diğer hücreler ile bağlantılar oluşturmaktadır. Ayrıca girdi ve çıktı katmanı tek bir katmandan oluşurken gizli katman olarak da adlandırılan ara katman bir veya birden çok katmandan oluşabilmektedir. Birden çok katmandan oluşan ara katmanın çalışma yapısı girdi katmanından gelen sinyalleri işlemden geçirerek varsa diğer katmanlara göndermek ve oradan çıkış katmanına iletmek şeklindedir. Bu katmandaki yapay sinir ağı hücrelerinin ve katmanlarının sayısı oldukça önemlidir nitekim ağın basit ya da karmaşık yapıda olmasını sağlamaktadır(Çanakçı, 2006:16). Birçok ağ türünde gizli katmandaki bir yapay sinir hücresinin görevi, sadece bir önceki katmandaki tüm hücrelerden sinyal alarak oluşan çıktıyı kendinden sonraki katmanın tüm hücrelerine göndermektedir (Yurtoğlu, 2005:20-21).

- ✓ **Çıktı Katmanı:** Çıkış katmanından elde edilen çıktı ağıdaki tüm hücrelerin toplam etkisinin sonucunu göstermektedir. Nitekim ağda ilk olarak girdi katmanına gelerek süreci başlatan yapay sinir hücresi çıktı katmanı ile süreci tamamlamaktadır(Dombaycı,Gölcü,2009:1159). Ara katmandan gelen sinir hücreleri çıktı katmanında girdi olarak kullanılmaktadır. Yapay sinir ağlarında ağın tamamı bu süreci takip ederek ilerlemektedir. Bu ağlar, tamamen veya kısmen bağlanabildiği gibi çevreden ve veri setinden öğrendikleriyle ağın performansının gelişimine imkan sağlamaktadırlar (Koleyni, 2009:144).

2.1.7. Biyolojik ve Yapay Sinir Ağlarının Benzerlikleri ve Temel Farkları

İnsanda öğrenme biyolojik sistem vasıtasıyla, nöronlar arasındaki sinaptik (synaptic) bağlantıların ayarlanması ile gerçekleşir. Bu nöronlar dentritler aracılığıyla geçirici (tetikleyici) veya yasaklayıcı giriş uyarısı alabilmektedirler. Böylece girişler toplanarak giriş gövdesinde yerini alır. Bu girişler belirli bir eşik değerini geçtiği zaman, hücre diğer hücrelere aksonu vasıtasıyla bir etki iletir. Bu yapay nöron için de aynı şekilde bir model oluşturmaktadır (Ersoy,Karal, 2012:192).Aşağıda Şekil 2.6’ da biyolojik bir sinir hücresi ile biyolojik sinaps, yapay sinir hücresi ile yapay sinapslerin görünimleri gösterilmiştir (Maltarollo vd., 2013: 205).



Şekil 2.6: Biyolojik Sinir Hücresi ve Yapay Sinir Hücresi

Kaynak (Maltarollo vd.,2013:205)

Yapay sinir ağıları ile biyolojik sinir ağıları arasındaki benzerlikler de Tablo 2.3’ de verilmiştir.

Tablo 2.3: Biyolojik Sinir Ağı İle Yapay Sinir Ağının Karşılaştırılması

Biyolojik Sinir Ağı	Yapay Sinir Ağı
Sinir Sistemi	Sinirsel Hesaplama Sistemi
Sinir	Düğüm(Sinir, İşlem elemanı)
Sinaps	Sinirler arası bağlantı ağırlıkları
Dendrit	Toplama İşlemi
Hücre Gövdesi	Etkinlik İşlevi
Akson	Sinir Çıkışı

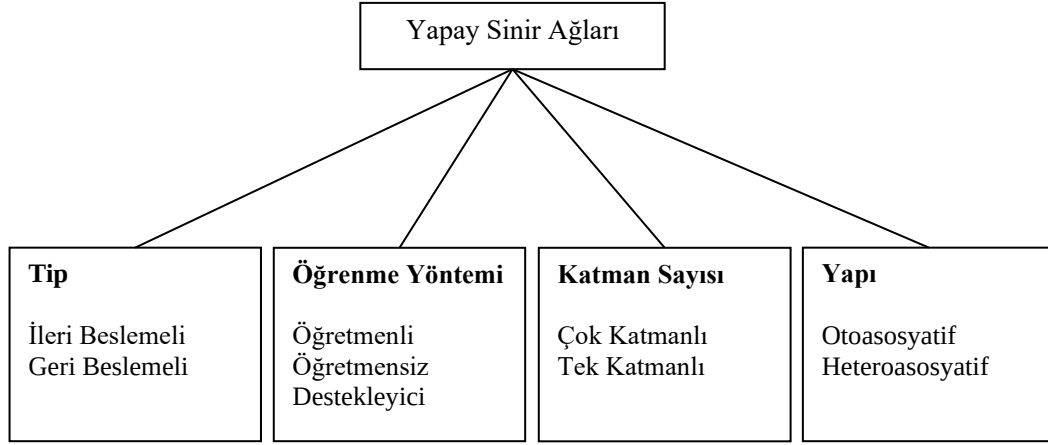
Kaynak:(Elmas, 2011:33)

Yapay sinir ağlarının temeli her ne kadar biyolojik sinir sistemi göz önüne alınarak atılmış olsa da canlı sistemin ileri düzeydeki karmaşıklığı, etkinlik ve kapasitesi sinir ağının tam olarak tasarımına olanak vermemektedir. Canlı sistemlerde hücre ve hücreler arasındaki bağlantılar oldukça karmaşık bir yapı oluşturmaktadır. Dolayısıyla biyolojik sinir sistemlerinde sadece nöronların varlığından söz etmek doğru olmayacaktır çünkü bu sistemlerde sinir hücrelerinin yaklaşık 50 katı kadar sayıya ulaşabilen ara hücrelerin (glia hücreleri) varlığından da söz edilmesi gerekmektedir(Yazıcı vd., 2007:68).

Bütün bunlar göz önüne alındığında, yapay sinir ağıları biyolojik sinir sistemlerinin karmaşık yapısını ve özelliklerinin tümünü yansıtmıyor olsa da kaba bir benzetimi olduğu yorumunu yapmak mümkündür.

2.1.8. Yapay Sinir Ağlarının Sınıflandırılması

Yapay sinir ağlarının sınıflandırılması çeşitli kategorilere ayrılarak yapılmaktadır. Bu kategorileri tip, öğrenme yöntemi, katman sayısı, yapı olarak sıralamak mümkündür. Yapay sinir ağlarının sınıflandırılması şekil 2.7’ deki gibidir (Hamzaçebi, 2021:30):



Şekil 2.7: Yapay Sinir Ağlarının Sınıflandırılması

Kaynak: (Hamzaçebi, 2021:31)

2.1.8.1.Yapay Sinir Ağlarının Tipine Göre Sınıflandırılması

Yapay sinir ağları, yapay sinir hücreleri arasındaki bağlantı sayısına göre ileri beslemeli ağlar ve geri beslemeli ağlar olarak sınıflandırılmaktadır (Hamzaçebi, 2021:31).

İleri beslemeli ağlarda veriler yalnızca girdi katmanından çıktı katmanına doğru ileri doğru aktarılmakta olup bir katmandaki yapay sinir hücrelerinin çıkışları bir sonraki katmana ağırlıklandırılarak giriş verileri görevi görmektedir. Aynı katmanda yer alan yapay sinir hücreleri ve bir önceki katmanda yer alan yapay sinir hücreleri arasında bir geri besleme davranışı göstermemektedir (Asilkan,Irmak, 2009:380).

İleri beslemeli ağlar çevreden ve veri setinden öğrenilen bilgilerden hareketle çalışma prensibi geliştirmişlerdir. Bu ağlarda iyileştirme başka bir değişle ağın performansını arttırma önceden belirlenen kriterler esas alınarak yapılmaktadır (Koleyni, 2009: 144).

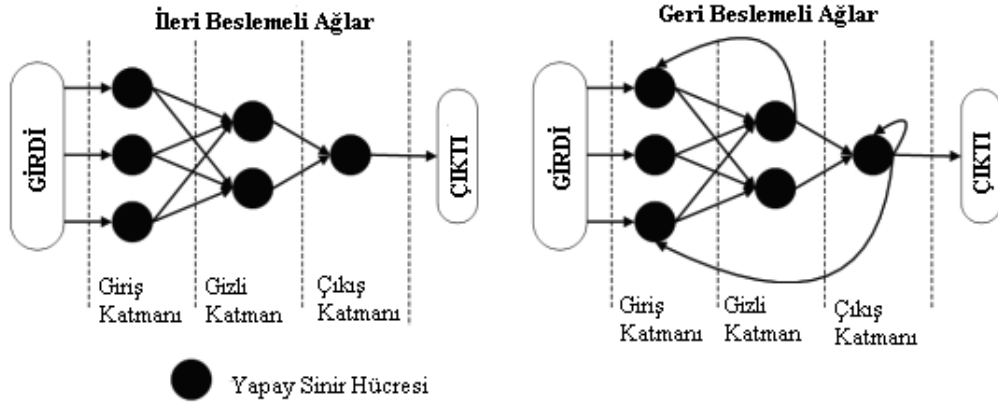
İleri beslemeli ağlarda, geri yayılım öğrenme kuralı ağın çıkış verisi için mevcut olan hataya göre her bir katmanın ağırlıklarını yinelemek için kullanılmakta ve bir katmandaki yapay sinir hücresi başka bir katmandaki yapay sinir hücresine girdi sağlamaktadır (Arslan,İnce, 1996:250).

İleri beslemeli ağlarda giriş, gizli ve çıkış katmanı yer almaktadır. Gizli katman birden çok olabilmektedir. Üstelik giriş ve çıkış katmanındaki hücre sayısı ilgilenilen probleme

göre belirlenirken gizli katman ve bu katmanlarda yer alan hücrelerin sayısı deneme yanılma yöntemiyle belirlenmektedir(Efe & Kaynak, 2000:13).

Geri beslemeli yapay sinir ağları, ileri beslemeli yapay sinir ağlarında olduğu gibi veri akışının yalnızca ileriye değil, aynı zamanda geriye doğru da veri akışını sağlayan ağlar olarak tanımlanmaktadır. Bu ağlarda minimum bir adet geri besleme çevrimi bulunmaktadır. Söz konusu olan bu geri besleme giriş katmanını, gizli katman ve çıkış katmanının kendi içindeki yapay sinir hücreleri arasında gerçekleşeceği gibi farklı katmanlardaki yapay sinir hücreleri arasında da gerçekleşmektedir (Asilkan,Irmak, 2009:380).

Geri beslemeli ağların bağlantılar oluşturmamasından dolayı ağın işleyişi ileri beslemeli ağlara göre daha yavaş gerçekleşmektedir. Nitekim bağlantılar aracılığıyla oluşan döngüler, girdilerin karşılığını daha yavaş bir şekilde gerçekleştirirken bu ağların eğitimin ise daha uzun sürede gerçekleşmesini sağlamaktadır (Yurtoğlu, 2005:28).



Şekil 2.8: İleri Beslemeli ve Geri Beslemeli Ağ Yapısı

Kaynak (Krenker vd., 2011:6)

2.1.8.2.Yapay Sinir Ağlarının Öğrenme Yöntemine Göre Sınıflandırılması

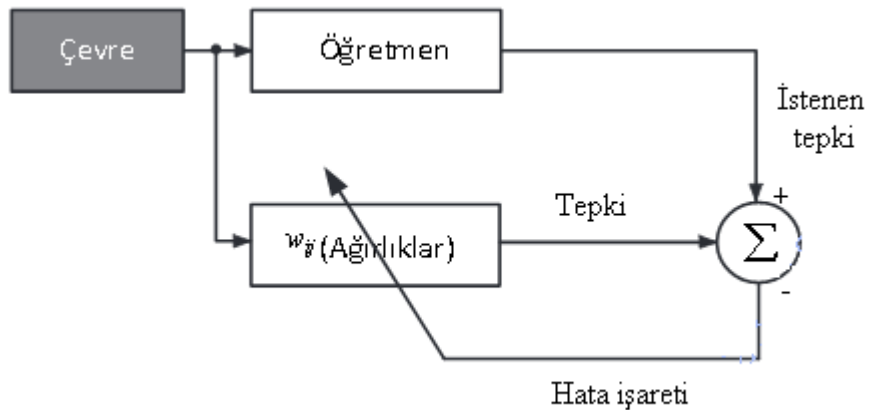
Yapay bir sinir hücresinin ağırlığı onunla çarpılan girdi kadar güçlüdür. Bu bağlamda yapay bir sinir hücresindeki girdiler için ağırlıklar ayarlanarak istenilen çıktıya ulaşılmaktadır. Ağda, ağırlıklar algoritmalar aracılığıyla ayarlanmaktadır. Bu ağırlık

ayarlama işlemine ağı eğitilmesi veya ağı öğrenmesi adı verilmektedir (Gershenson, 2003: 3).

Yapay sinir ağlarının önemli bir özelliği olan öğrenme esnek bir sisteme sahip olmakla birlikte programlaması da kolay olmaktadır dolayısıyla oluşan yeni problemlere cevap vermesi adına ağı eğitilmesini mümkün kılmaktadır (Dastres, Soori, 2021:15).

Yapay sinir ağlarında, ağırlıklardaki değişimler yapay sinir hücrelerine aktararak ağı eğitilmesi amaçlanırken aynı zaman da öğrenmesi sağlanmaktadır. Bir sinir ağında 3 ana öğrenme türü vardır. Bunlar aşağıdaki gibidir (Fyfe, 2000:10):

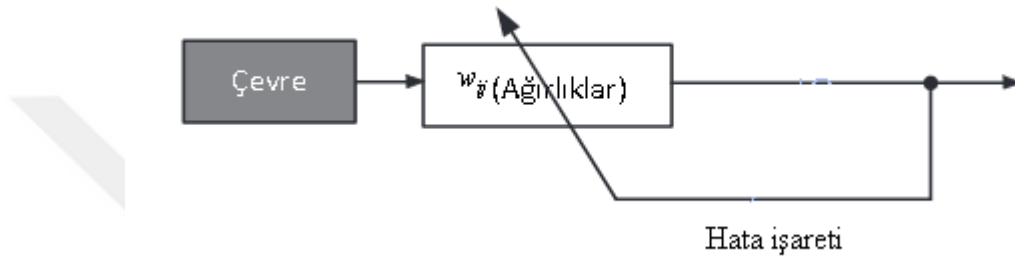
- ✓ **Denetimli Öğrenme:** Öğretmenli öğrenme olarak da adlandırılan bu öğrenme türü ile ağı tanımlanan girdi verileriyle istenen çıktıya ulaşılması sağlanmaktadır. Giriş verileri, aktivasyon fonksiyonu çıkış hücrelerine ulaşana kadar ağı boyunca ileri doğru yayılım göstermektedir. Ardından ağı hesapladığı çıktı değerlerini, istenen çıktı değerleri ile karşılaştırarak eğer çıktı değerleri istenen düzeyde ise ağıda bir değişiklik yapılmamaktadır. Ancak ağı verdiği çıktı değerleri, istenen çıktı değerlerinden farklı ise ağı tanımlanan ağırlıklar değiştirilerek ağı gelecekte aynı yanıtla tekrar sunulması durumunda doğru çıktıyı verme olasılığının daha yüksek olması amaçlanmaktadır (Fyfe, 2000:10). Denetimli öğrenme aşağıda Şekil 2.9’da gösterilmiştir.



Şekil 2.9:Denetimli (Danışmanlı)Öğrenme

Kaynak: (Elmas, 2011:89)

- ✓ **Denetimsiz Öğrenme:** Öğretmensiz öğrenme olarak da adlandırılan bu öğrenme türü çıktıların istenen değerleri ağı tanıtmadan gerçekleşmesidir. Burada amaç denetimsiz öğrenme algoritmaları aracılığıyla geçmiş veri kümesinin içinde gizli kalmış bilgilerin elde edilmesidir. Denetimsiz öğrenmede girdiler aynı zamanda çıktı olarak da kullanılarak veriler arasında ilişkiler kurulmakta ve en iyi sonuca ulaşılabilme adına ağı eğitimi gerçekleştirilmektedir (Hamzaçebi,2021:33). Aşağıda Şekil 2.10’da denetimsiz öğrenme gösterilmiştir:



Şekil 2.10: Denetimsiz (Danışmansız) Öğrenme

Kaynak: (Elmas, 2011:137)

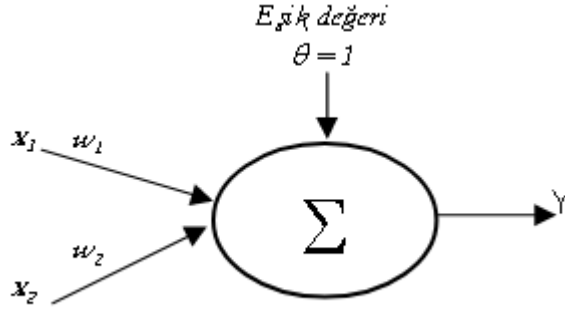
- ✓ **Takviyeli Öğrenme:** Bu öğrenme süreci genel itibariyle danışmanlı öğrenme yöntemine benzemekle birlikte bir danışman yardımı gerektirmektedir. Ancak takviyeli öğrenme yönteminde sisteme yalnızca girdi değerleri verilirken istenen çıktı değerleri verilmemektedir. Burada ağı çıktı değerlerini kendisinin üretmesi beklenir ve üretilen çıktı değerleri göz önüne alınarak danışman yardımıyla bir sinyal bildirilerek takviye yapılmaktadır. Sonuç olarak ağı danışmandan gelen bu sinyalle bağlantıların ağırlıklarını değiştirerek süreci ilerletmektedir (Arıkan Kargı, 2015: 67).

2.1.8.3.Yapay Sinir Ağlarının Katmanlarına Göre Sınıflandırılması

Yapay sinir ağları katmanlarına göre tek katmanlı ve çok katmanlı yapay sinir ağları olarak ikiye ayrılmaktadır.

Tek katmanlı yapay sinir ağları yalnızca girdi ve çıktı katmanından oluşmaktadır. Bir veya birden fazla girdi ve çıktı içeren ağda çıktılar girdilere bağlanarak oluşturulan her

bir bağlantı için ağırlık değeri verilmektedir (Öztemel,2012:59). İlk model örneklerinden biri olan tek katmanlı yapay sinir ağlarında, çıkış fonksiyonun doğrusal olması nedeniyle sınıfları birbirinden ayırma işlevini gören ve ötelemeyi ifade eden eşik değeri (θ) yer almaktadır (Nabiyev, 2021:588).Aşağıda Şekil 2.11’de tek katmanlı yapay sinir ağı modeli yer almaktadır:



Şekil 2.11: Tek Katmanlı Yapay Sinir Ağı Modeli

Kaynak: (Nabiyev, 2021:588)

Yukarıdaki şekilden hareketle tek katmanlı bir yapay sinir ağında, Y ile gösterilen çıktı değeri ağırlıklandırılmış giriş değerlerinin θ eşik değeri ile toplanmasının ardından aktivasyon fonksiyonundan geçirilmesi ile hesaplanmaktadır (Nabiyev, 2021: 588).

$$y = f \sum_i^n w_i x_i + \theta \quad (2.2)$$

θ ile gösterilen eşik değeri ağıta tanıtılan örneklerin iki sınıfa ayrılmasını sağlayan doğruyu bulmakla görevlidir. Eşik değeri kullanılarak iki sınıfa ayrılan ağın çıktısı 1 veya -1 değerini almaktadır. Bu değerler sınıfları temsil etmekle beraber ağın çıktısı 1 olduğunda birinci sınıfta,-1 olduğunda ise ikinci sınıfta yer aldığını göstermektedir (Öztemel, 2012:60).

$$f(x) = \begin{cases} 1, & y > 0 \text{ ise} \\ -1, & y < 0 \text{ ise} \end{cases} \quad (2.3)$$

Tek katmanlı ağlarda sınıflara ayırmayı sağlayan doğru denklemi aşağıdaki gibidir:

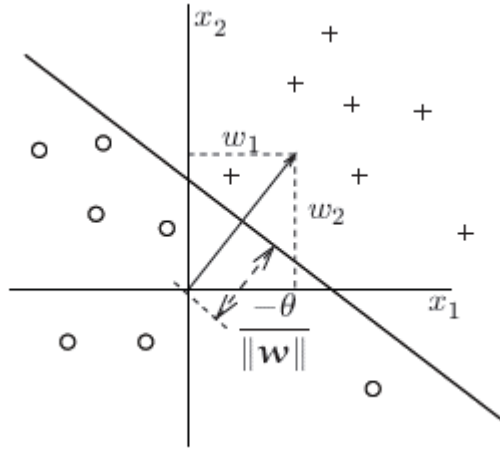
$$w_1x_1 + w_2x_2 + \theta = 0 \quad (2.4)$$

Buradan hareketle;

$$x_1 = -\left(\frac{w_2}{w_1}\right) \cdot x_2 - \frac{\theta}{w_1} \quad (2.5)$$

$$x_2 = -\left(\frac{w_1}{w_2}\right) \cdot x_1 - \frac{\theta}{w_2} \quad (2.6)$$

olarak bulunmaktadır. Yukarıdaki denklemler kullanılarak iki sınıfa ayırmayı sağlayan doğrunun geometrik gösterimi Şekil 2. 12’de verilmiştir:



Şekil 2.12: Ağırlıkların ve Sınıf Ayırıcı Olan Doğrunun Geometrik Gösterimi

Kaynak: (Kröse, Smagt, 1996:24)

Tek katmanlı ağın çalışma yapısı ele alındığında ağın öğrenmeyi gerçekleştirebilmesi ve en iyi sonuca ulaşabilmesi adına ağırlık değerlerinin değiştirilmesi gerekmektedir. Ağırlık değerlerinin değişmesi durumunda doğrunun eğimi de değişmektedir. Matematiksel olarak ifade edildiğinde t zaman biriminde ağırlık değerlerinin Δw kadar değiştirilmesi

$$w_i(t+1) = w_i(t) + \Delta w_i(t) \quad (2.7)$$

olmaktadır ve öğrenme esnasında en iyi sonuca ulaşmak için her iterasyonda gerçekleşmektedir. Ayrıca en iyi sonuca ulaşmak adına θ ile ifade edilen eşik değeri de her iterasyonda $\Delta\theta$ kadar değişmekte ve aşağıdaki şekilde formüle edilmektedir.

$$\theta(t + 1) = \theta(t) + \Delta \theta(t) \quad (2.8)$$

Eşik değerinin değiştirilmesi doğrunun sınıflar arasında kaymasını sağlarken aynı zamanda transfer fonksiyonunun yerini belirlemektedir (Arıkan Kargı, 2015:51).

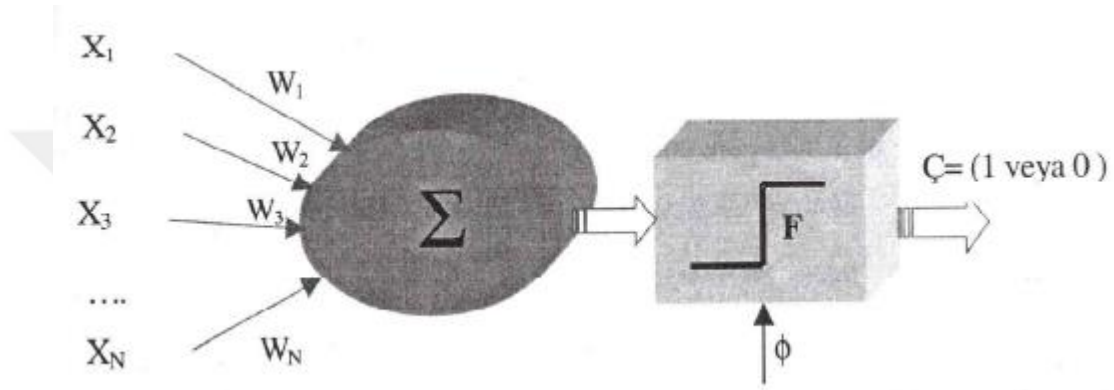
Tek katmanlı yapay sinir ağlarında kullanılan modellere ilişkin açıklamalar aşağıda verilmiştir:

Basit Algılayıcı (Perceptron) Modeli: Rosenblatt tarafından geliştirilmiş olan basit algılayıcı modeli yapay sinir ağlarının tarihsel gelişiminde önemli bir yere sahiptir ve algoritmik olarak tanımlanan ilk sinir ağıdır (Haykin, 2009:77). Basit algılayıcı veya perceptron olarak adlandırılan bu model genel olarak akıllı sistemlerin temel özelliklerini baz alarak varsayımsal bir sinir sistemi veya makine için geliştirilmiştir (Rosenblatt, 1958:387).

Basit Algılayıcı Modeli bir sinir hücresinin birden fazla girdi alarak yalnızca 0 veya 1 değerlerini alabilen çıktılar üretmesine dayanmaktadır. Çıktı değeri oluşturulurken eşik değer fonksiyonu kullanılmaktadır. Bu modelde girdiler ağa tanıtılırken, girdi değerlerine karşılık 0 ve 1 değerlerini alan çıktılarda ağa tanıtılmaktadır ve sonrasında öğrenme kuralı yardımıyla ağın çıktı değeri bulunmaktadır. Eğer bulunan çıktı değeri istenilen değerden farklı ise ağırlık değerleri ve çıktı değerleri değiştirilmektedir (Öztemel, 2012:61-62). Denetimli öğrenme algoritmasını benimsemiş olan basit algılayıcı modelde belirlenen girdi değerlerinin ağa tanıtılmasına karşılık, istenilen çıktı değerlerine ulaşabilmek için ağın uygun ağırlıkları bulması istenmektedir (Hamzaçebi, 2021:43). Basit Algılayıcı modelinde çıktı hesaplanırken belirlenen eşik değeri ve girdi değerleri karşılaştırılarak çıktı değerinin 1 veya 0 olacağına karar verilmektedir (Arıkan Kargı, 2015:53).

$$y = \begin{cases} 1 & \text{Eğer Net} > \theta \text{ ise} \\ 0 & \text{Eğer Net} \leq \theta \text{ ise} \end{cases} \quad (2.9)$$

Net girdi değeri θ ile gösterilen eşik değerinden büyük olduğu durumda çıktı değeri 1, net girdi değeri θ eşik değerinden küçük veya eşit olması durumunda çıktı değeri 0 değerini almaktadır. Aşağıda Şekil 2.13’de basit algılayıcı model yapısı gösterilmiştir.



Şekil 2.13. Basit Algılayıcı Modeli Yapısı

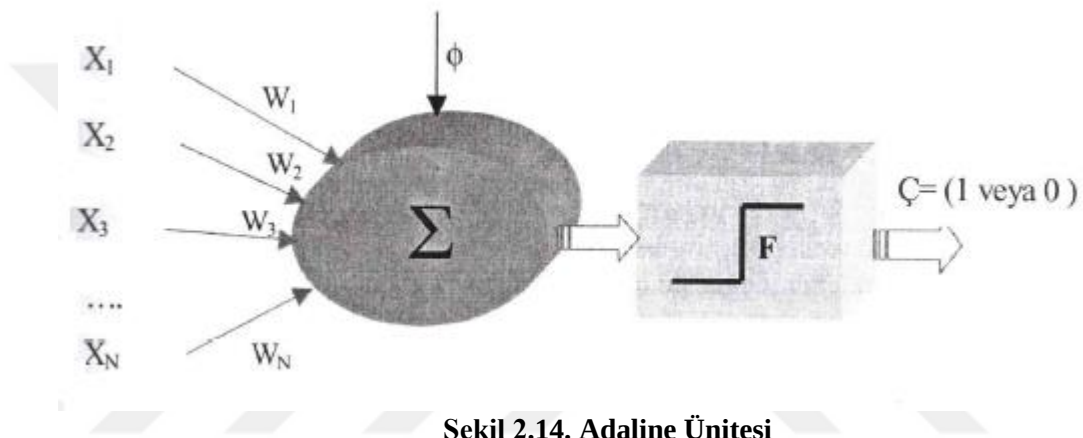
Kaynak: (Öztemel, 2012: 61)

Basit algılayıcı modelde gerçekleşen çıktı değerleri ile istenilen çıktı değerleri birbirinden farklı olduğunda iki durum söz konusu olacaktır (Öztemel, 2012:62-63).

- ✓ Ağın beklenen çıktısının 0 değerini alması ancak gerçekleşen çıktının 1 değerinde olması net girdi değerinin belirlenen eşik değerinin üstünde olduğunu göstermekte ve ağırlık değerlerinin azaltılması gerektiği vurgulanmaktadır.
- ✓ Ağın beklenen çıktısının 1 değerini alması ancak gerçekleşen çıktının 0 değerinde olması net girdi değerinin belirlenen eşik değerinin altında olduğunu göstermekte ve ağırlık değerlerinin artırılması gerektiği vurgulanmaktadır.

ADALINE/MADALINE Modeli: Adaline, Adaptive Linear Element (uyum gösteren lineer eleman) kısaltması olarak Windrow ve öğrencisi Hoff tarafından 1959 yılında geliştirilen bir model olmakla birlikte negatif geri besleme yapısını kullanarak basit

algılayıcı modelinin yetersiz kaldığı problemlerde başarılı sonuçlar elde etmiştir(Akpınar, 2017:320). Adaline modelinde aktivasyon fonksiyonu olarak doğrusal fonksiyon kullanılırken, perceptron olarak da adlandırılan basit algılayıcı modelinde aktivasyon fonksiyonu olarak eşik fonksiyonu kullanılmaktadır. Bunun dışında iki model arasında farklılık yoktur ve iki modelde doğrusal problemleri çözümleyebilmektedir(Arı,Berberler,2017:59).Adaline modeli aşağıda Şekil 2.14’de verilmiştir:



Şekil 2.14. Adaline Ünitesi

Kaynak: (Öztemel, 2012: 69)

Tek katmanlı ve tek bir yapay sinir hücresinden oluşan Adaline modeli basit bir ağ olmasına karşılık Delta kuralı öğrenme algoritmasını geliştirerek yapay sinir ağlarının ilk endüstriyel uygulamalarında kullanılmasına katkı sağlamıştır (Silva vd., 2017:41).

Adaline modeli, delta kuralı veya en küçük kareler yöntemi olarak adlandırılan bu yöntem aracılığıyla ağın gerçekleşen çıktısının istenilen çıktı değerlerine göre hatayı minimum kılacak şekilde ağırlıkların değiştirilmesi işlemine dayanmaktadır (Öztemel, 2012:69).

Delta öğrenme kuralının hatayı minimize etmesi bütün eğitim kümesinde girdi vektörleri bitene kadar devam etmektedir. Eğitim tamamlanması durumunda çıktı değeri 1 veya 0 olarak seçilmektedir (Hamzaçebi, 2021:50).

Gerçekleşen ve istenilen çıktı değerleri arasında fark olması durumunda Delta kuralı yalnızca hatayı düzeltmekle sınırlı kalmayıp aynı zamanda hatayı yok etmeye çalışmaktadır. Bu kuralın işleyiş aşamaları aşağıdaki gibidir(Elmas, 2003:53):

- ✓ Çıkış düğümlerindeki gerçekleşen ve istenilen değerler arasındaki farklılıkları bularak hatayı tespit etmektedir.
- ✓ Düzeltme faktörü bulunmaktadır. Bu faktör aşağıdaki şekilde ifade edilmektedir.

$$(E+\varepsilon)d \quad (2.10)$$

Formülde E hata değerini, ε hataya ilave edilen miktarı, d ise sabiti ifade etmektedir. Burada ε ve d değerleri istenilen çözüme ulaşabilmek için değiştirilebilir niteliğe sahiptir.

- ✓ Son olarak tüm bu işlemler ardından ağırlıklar yeniden düzenlenmektedir. Eğer gerçekleşen çıktı değeri 0, beklenen çıktı 1 ise ağırlıklar arttırılmaktadır. Eğer gerçekleşen çıktı 1, beklenen çıktı 0 ise ağırlıklar azaltılmaktadır.

MADALINE modeli eğitim sürecinde herhangi bir değişikliğin yapılmadığı ADALINE modellerinin birleşiminden oluşmaktadır (Fausett, 1994:88).MADALINE ağırları genel itibariyle iki katmandan oluşmakta ve her katmanında farklı bir ADALINE modeli yer almaktadır (Öztemel, 2012:73).

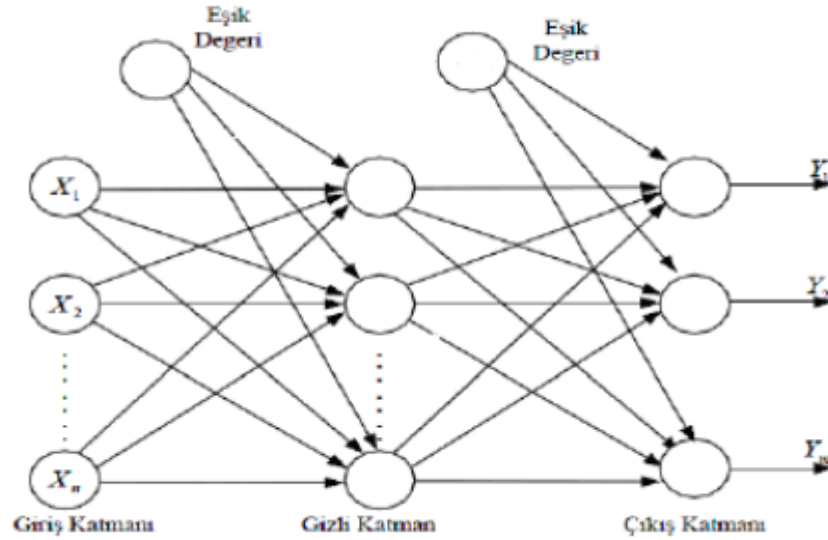
Basit Algılayıcı modeller yalnızca doğrusal problemleri çözme yeteneğine sahip olmasından dolayı sınırlı kalmışlardır. Böylece çok katmanlı ağ yapısı, basit algılayıcıların sınırlı kaldığı modelleri (XOR) başka bir değişle doğrusal olmayan modelleri çözmek için geliştirilmiştir (Arıkan Kargı, 2015:81).

Çok katmanlı ağların temel özelliklerini aşağıdaki şekilde sıralamak mümkündür (Haykin, 2009:153):

- ✓ Ağdaki her yapay sinir hücresinin modeli, türevlenebilen doğrusal olmayan bir aktivasyon fonksiyonu içerir.
- ✓ Ağ hem giriş düğümlerinde hem de çıkış düğümlerinde gizlenen bir veya birden fazla katman içerir.

- ✓ Ağıın kapsamı, ağıın sinaptik ağırlığı aracılığıyla saptanan yüksek derecede bağlanabilirlik sergiler.

Çok katmanlı yapay sinir ağıının yapısı aşağıda Şekil 2.15’de verilmiştir:



Şekil 2.15:Çok Katmanlı Algılayıcı Yapay Sinir Ağıının Yapısı

Kaynak: (Arıkan Kargı, 2015: 83)

Çok katmanlı yapay sinir ağıları giriş ve çıkış katmanı arasında en az bir adet gizli katman bulunduran ağlardır. Bu ağlarının en temel özelliği tabakalı bir yapıdan oluşması ve ayrıca gizli tabakadaki hücre çıkışlarında doğrusal olmayan işlemcilerin kullanılabilir olmasıdır(Şen, 2004:46-47).

Çok katmanlı ağların çalışmasında gizli katmandaki yapay sinir hücreleri önemli bir rol oynamaktadır. Bu hücreler ağ boyunca eğitim verilerini karakterize edip özelliklerini keşfederek giderek ilerleme göstermektedir (Haykin, 2009:156).

Doğrusal olmayan problemlere çözüm üretebilen çok katmanlı yapay sinir ağlarında bilgi akışı ileri doğru gerçekleştiği için bu ağlar ileri beslemeli ağ yapısına sahiptir ve bu ağlar eğitim sürecinde birçok öğrenme algoritmasını kullanmaya olanak verdiği için yaygın olarak kullanılmaktadır (Ataseven, 2013:104).

İleri beslemeli ağ yapısına sahip olan çok katmanlı ağlar, öğrenme sürecinde geriye yayılım algoritmasını kullanmaktadır. Geriye yayılım algoritması birçok farklı uygulamalarda kullanılabilir olması ve karmaşık, belirsiz problemlere doğrusal olmayan çözümlerde üretebilmesi açısından oldukça avantaj sağlamaktadır (Anderson,McNeill, 1992:32).

İleri besleme gösteren ve ağın eğitimi için geri yayılım algoritmasını kullanan ağlarda katman sayısı ve her bir katmandaki işlem elemanı sayısı oldukça önemlidir. Katman sayısı ve eleman sayısını belirlemede en iyi sonucu veren bir uygulama bulunmamaktadır ancak araştırmacılar tarafından belirlenen genel kurallar vardır. Bu kurallar aşağıda verilmiştir (Anderson,McNeill, 1992:33):

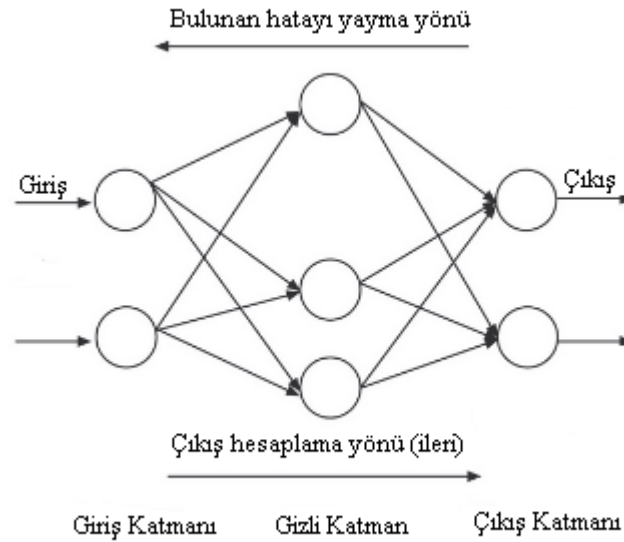
Kural 1: Girdi değerleri ile istenilen çıktı değerleri arasındaki ilişkinin karmaşıklığına bağlı olarak gizli katmandaki işlem elemanlarının sayısı artmalıdır.

Kural 2: Modellenen süreç birden çok aşamaya ayrıldığında ek gizli katmana ihtiyaç duyulabilir. Buna karşılık işlem aşamalara ayrılmazsa ek katmanlar ezberlemeyi kolayca gerçekleştirir ve bu gerçek bir genel çözüm olmayabilir.

Kural 3:Modelde ağın eğitimi için kullanılacak olan veri miktarı, gizli katmanlardaki işleme elemanları sayısının bulunmasında bir üst sınır olmaktadır. Bu üst sınır eğitim için oluşturulan veri setindeki girdi-çıktı çifti örneklerinin sayısı ağdaki girdi ve çıktı işleme elemanlarının toplam sayısına bölünerek bulunmaktadır. Ölçeklendirme faktörü olarak adlandırılan bu değer beş ile on arasında yer alırken daha gürültülü veriler için büyük ölçeklendirme faktörleri kullanılmalıdır. Son derece fazla gürültülü veriler, yirmi ile elli arasında ölçeklendirme faktörü gerektirirken veri girişlerinin doğru yapıldığı durumda ölçeklendirme faktörü ikiye düşürülebilmektedir. Ayrıca veri eğilimlerinde genellemeyi engellemesiyle yeni oluşturulacak kümelere etki etmeyeceğinden dolayı gizli katmanda işlem sayısının az olması istenmektedir.

Geriye Yayılma Algoritması: Geriye yayılım algoritması, çok katmanlı ağlarda ağın eğitilme sürecinde kullanılan delta kuralının genelleştirilmiş hali olarak da bilinmektedir(Elmas, 2003:124).

Geriye yayılım algoritmasının öğrenme aşamasında ilk olarak giriş değerleri ağı belirli bir sırada verilmektedir. Her bir değer çıktı değeri bulununcaya kadar ileriye yayılmaktadır. İşlemler sonucunda hesaplanan çıktı değeri ile istenilen çıktı değeri karşılaştırılarak ve aralarındaki fark saptanarak hata tespit edilmektedir. Bulunan hata değeri ağıdaki katmanlarda sinaptik ayarlamaların yapıldığı geri besleme bağlantılarında girdi olarak kullanılmaktadır. Geriye yayılım algoritmasında amaç, ağıdaki katmanların her birinde ağırlıkları düzenleyerek geriye doğru yayılım aracılığıyla toplam hatayı minimuma indirmektedir. Burada toplam hata değeri minimum olana kadar süreç tekrarlanmaktadır(Elmas, 2003:125). Şekil 2.16’da geriye yayımlı yapay sinir ağı gösterilmiştir.



Şekil 2.16: Geriye Yayımlı Yapay Sinir Ağları

Kaynak:(Çalışkan Çavdar , Aydın, 2018:100)

Geriye yayılma algoritması, her iterasyonda ileri yayılım olarak adlandırılan yapay sinir ağına tanıtılan girdi değerlerine karşı çıktı değerlerinin oluşması ve geri yayılma olarak adlandırılan çıktı değerlerinden elde edilen hataları temel alarak ağı için ağırlıkların yeniden düzenlenmesini içeren iki aşamalı yapıdan oluşmaktadır(Özgen, 2007:70). Bu iki aşamalı yapı ileriye doğru hesaplama ve geriye doğru hesaplama olarak adlandırılmaktadır. Aşağıda bu aşamalar açıklanmıştır (Nabiyev, 2021:599-600):

İleriye Doğru Hesaplama: Net girişlerin hesaplanması ile ara katmanda yer alan tüm proses elemanlarının ve çıktı katmanında yer alan proses elemanlarının çıkışı, türevi alınabilir bir aktivasyon fonksiyondan geçirilerek belirlenmektedir. Çıkış değerleri bulununca süreç tamamlanmaktadır. Bu sürecin denklemi aşağıda verilmiştir:

$$Y_k^i = G_k \quad (2.11)$$

G_k = Giriş değerlerini Y_k^i = Çıktı değerlerini göstermek üzere; Net değerler aşağıdaki gibi bulunmaktadır:

$$NET_j = \sum_{k=1}^n w_{kj} Y_k \quad (2.12)$$

Net değerlere karşılık elde edilen çıktı değerleri ise aşağıdaki şekilde ifade edilmektedir:

$$X_j = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-(NET_j + \beta_j)}} \quad (2.13)$$

Denklemden β_j parametresi sabit değeri ifade ederken j. elemana bağlanan eşik elemanının ağırlığını göstermektedir. Eşik değeri çıktıları daima 1 olmaktadır.

Son olarak beklenen (istenilen) çıkış değerlerini gösteren B_i 'den ağa tanıtılan girdi değerlerine karşılık ürettiği son çıktı değerini gösteren ζ_i çıkartılarak hata değeri bulunmaktadır.

$$H_i = B_i - \zeta_i \quad (2.14)$$

Geriye Doğru Hesaplama: Bu hesaplamada beklenen çıktı değeri ile ağa tanıtılan girdi değerlerine karşılık üretilen son çıktı değerinin farklarının alınması sonucunda hata değerine ulaşılır. H_i bir süreç elemanından elde edilen hata olmakla birlikte ağı çıktısı katmanında meydana gelen toplam hata aşağıdaki gibi hesaplanır:

$$E = \sqrt{\sum_i H_i^2} \quad (2.15)$$

Ağın eğitimi sürecinde bu toplam hata değerinin azaltılmasını amaçladığı için katmanlardaki ağırlıklar değiştirilerek süreç elemanlarına dağıtılmaktadır. Ara (gizli) katman ile çıktı katmanı arasındaki ağırlıkların k. iterasyondaki değerleri, değişim miktarlarının eski değerlerine eklenmesiyle elde edilmektedir:

$$w_{jm}(k) = w_{jm}(k-1) + \Delta w_{jm}(k) \quad (2.16)$$

Yukarıdaki denkleme w_{jm} değeri, ara katmandaki j. elemanın çıktı katmanındaki m. elemanla olan bağlantısının ağırlığını göstermektedir. Ağırlıklardaki değişimi Δw_i ile gösterilerek aşağıdaki gibi formülize edilmektedir:

$$\Delta w_{jm}(k) = \lambda \delta_m X_j + \mu \Delta w_{jm}(k-1) \quad (2.17)$$

Yukarıda öğrenme katsayısını gösteren λ ağırlıkların değişimini, momentum katsayısı μ , ağırlık değişiminin belirli bir oran ile bir sonraki değişime eklenmesini, δ_m ise m. çıkıştaki hatasını, $f'(NET)$ ise aktivasyon fonksiyonunun türevini göstermektedir. Sigmoid fonksiyonu kullanılması durumunda ise denklem şu şekilde formülize edilmektedir:

$$\delta_m = f'(NET).E_m = X_m(1 - X_m).E_m \quad (2.18)$$

Çıktı katmanında yer alan süreç elemanların eşik değerini ifade eden β^c değişim miktarı, çıkışın 1 olmasını temel alarak aşağıdaki gibi formülize edilmektedir:

$$\Delta \beta_m^c(k) = \lambda \delta_m + \mu \Delta \beta_m^c(k-1) \quad (2.19)$$

Eşik değerinin k. iterasyondaki ağırlığının yeni değeri de aşağıdaki gibi formülize edilmektedir:

$$\beta_m^c(k) = \beta_m^c(k-1) + \Delta\beta_m^c(k) \quad (2.20)$$

Ara katman ve çıktı katmanı arasındaki ağırlıkların yinelenmesinde yukarıdaki hesaplamalarla benzerlik göstermektedir. Yalnızca en son ara katmana gelen bilgiler girdi katmanından veya önceki ara katmandan geldiği için, hataların oluşmasında girdi ve ara katman arasındaki ağırlıkların payı bulunmaktadır. Bu nedenle girdi katmanı ile ara katman arasındaki ağırlıkların değişiminde, çıktı katmanındaki süreç elemanlarının hepsini ele almak gerekmektedir (Nabiyev, 2021:601).

$$\delta_j = f'(NET) \cdot \sum_m \delta_m w_{jm} = X_j (1 - X_j) \cdot \sum_m \delta_m w_{jm} \quad (2.21)$$

Son olarak, hata değerinin minimum olması amaçlanarak geri yayılım algoritmasında ileri ve geri doğru hesaplamalar tamamlanarak ağırlıklar değiştirilir ve hata değeri minimum seviyeye gelene kadar süreç ilerletilirken minimum hatanın bulunmasıyla süreç sonlandırılmaktadır.

2.1.8.4.Yapay Sinir Ağlarının Yapılarına Göre Sınıflandırılması

Yapay sinir ağları girdi ve çıktı hücrelerinin yapılarına göre otoasosyatif ve heteroasosyatif ağlar olarak gruplandırılmaktadır. Girdi olarak gelen yapay sinir hücrelerinin aynı zaman da çıktı hücreleri olarak kullanıldığı ağlar otoasosyatif olarak adlandırılırken farklı girdi ve çıktı hücrelerinin kullanıldığı ağlar heteroasosyatif ağlar olarak adlandırılmaktadır. Bu ağlara örnek vermek gerekirse Hopfield ağları otoasosyatif ağlar grubuna dahil olurken Kohonen ağı ve çok katmanlı algılayıcılar heteroasosyatif ağlar grubunda yer almaktadır (Hamzaçebi, 2021:33).

2.1.9.Yapay Sinir Ağlarında Öğrenme Kuralları

Yapay sinir ağlarında kullanılan pek çok öğrenme kuralı bulunmaktadır. Temel öğrenme kuralları Hebb, Hopfield, Delta ve Kohonen olarak sıralanmaktadır. Ancak bu öğrenme kuralları çoğunlukla Hebb kuralı çerçevesinde geliştirilmiştir (Öztemel, 2012: 26). Öğrenme kuralları aşağıda açıklanmıştır:

- ✓ **Hebb Öğrenme Kuralı:** Bu kural 1949 yılında Donald Hebb tarafından geliştirilmiştir (Hebb, 1949).Kurala göre yapay bir sinir hücresinin başka bir yapay sinir hücresi ile matematiksel olarak aynı işarete sahip olması durumunda aktif olduğu kabul edilmektedir. Aktif olan yapay bir sinir hücresinin başka bir sinir hücresinden aynı zamanda bilgi aldığı durumlarda hücreler arasındaki bağlantı kuvvetlendirilmelidir. Bu bağlamda yapay sinir hücresi aktif ise bağlı olduğu diğer hücre aktif, pasif olması durumunda ise bağlı olduğu diğer hücrenin pasif olduğu ifade edilmektedir (Öztemel, 2012:26).
- ✓ **Hopfield Öğrenme Kuralı:** Öğrenme katsayısının program kullanıcısı tarafından sabit olarak belirlendiği bu kural Hebb kuralına benzemektedir. Burada bağlantıların kuvvetlendirilmesi veya zayıflatılması amaçlanmaktadır. Bu amaç doğrultusunda beklenen çıktı ve girdi değerleri aktif ise sabit olan öğrenme katsayısı kuvvetlendirilirken, beklenen çıktı ve girdi değerleri pasif ise sabit olan öğrenme katsayısı zayıflatılmaktadır (Bayır, 2006:30).
- ✓ **Delta Öğrenme Kuralı:** Windrow ve Ted Hoff tarafından 1960 yılında geliştirilen bu kural en küçük kareler olarak da bilinir. Kurala göre girişler ağı eğretilerek çıkış değeri elde edilmektedir. Elde edilen çıkış değeri ile istenilen çıkış değeri arasında karşılaştırma yapılarak hata değeri bulunmaktadır. Bulunan hata değerinin sıfır olması durumunda ağırlıklarda bir değişiklik yapılmadan öğrenme işlemi tamamlanmaktadır. Fakat hata değeri sıfırdan daha büyük bir değer alması durumunda hata minimize edilene dek ağırlık değerlerinde değişiklik yapılması gerekmektedir (Elmas, 2011:99).
- ✓ **Kohonen Öğrenme Kuralı:** Biyolojik öğrenmeyi temel alarak geliştirilmiş bir kural olan kohonen öğrenme kuralında yapay sinir hücreleri öğrenmeyi ve ağırlıkları yenilemeyi kendi yapılarında gerçekleştirmektedir. Öğrenme işlemini sonucunda en büyük çıktıya ulaşan hücre kazanan hücre olarak adlandırılmaktadır. Kazanan hücre komşu hücrelerin ağırlıklarını yenileyerek sürece devam etmektedir. Bu kural çıkış verisi gerektirmediği için öğretmensiz (danışmansız) eğitimle öğrenmeyi gerçekleştirmektedir (Şen, 2004:66).

2.1.10.Yapay Sinir Ağlarının Görevleri ve Uygulama Alanları

Yapay sinir ağları problemleri çözmek ve alınan kararlara yardımcı olmak için kullanılan araçlardır. Bu amaçlar doğrultusunda yapay sinir ağlarının gerçekleştirdiği görevler aşağıdaki şekilde sıralanabilir (Öztemel, 2012:204):

- ✓ *Tahmin:* Yapay sinir ağları ile ulaşılan bilgiler aracılığıyla ileriye dönük tahminler yapılabilir bunlara örnek olarak hava durumu, hisse senedi fiyat tahmini verilebilir.
- ✓ *Sınıflandırma:* Yapay sinir ağları sunulan bilgileri kategorize ederek sınıflandırma yapabilir.
- ✓ *Veri İlişkilendirme:* Yapay sinir ağları veri ilişkilendirme ile ağa sunulan verilerin hatalı ve eksik olup olmadığını test ederek eksik verileri öğrenilen bilgiler ile tamamlayabilir.
- ✓ *Veri Filtreleme:* Bu amaçla eğitilen ağlar, birçok veri arasından uygun veriyi filtreleyebilir.
- ✓ *Tanıma ve Eşleştirme:* Eğitilen ağ değişik şekillerdeki verileri tanıyabilir aynı zamanda eksik ve karmaşık bilgileri eşleştirme yapabilir.
- ✓ *Teşhis:* Bu amaç doğrultusunda geliştirilen ağlar ağdaki sorunların, ağın işleyiş yapısında ki aksaklıkların teşhisini yapabilir.
- ✓ *Yorumlama:* Yapay sinir ağları eğitilen ağın işleyişinden elde edilen bulgular ve araştırılan konu üzerinde toplanan örnekleri kullanarak yeni olayları yorumlayabilir.

Teknolojinin giderek hayatımızda önemli bir yer edinmesi ile birlikte, evlerimizde kullandığımız teknolojik aletlerden bilgisayarlara ve cep telefonlarına kadar birçok yapay sinir ağı uygulamaları görmektedir (Öztemel, 2012:203).Yapay sinir ağları, gerçek hayatta karşılaştığımız problemlere uygun çözümler bularak geniş çaplı bir uygulama alanı yaratmıştır. Üstelik uygulama alanları sınırsız olup giderek artmaya devam etmektedir (Yurtoğlu, 2005:9). Yapay sinir ağlarının çeşitli kullanım alanları aşağıda verilmiştir:

- ✓ *Bankacılık*: Kredi başvuru değerlendirme, kredi riski ölçme, nakit tahmini, firma sınıflandırma, çek ve diğer belge okuyucuları, döviz kuru tahmini gibi uygulamalarda yapay sinir ağları kullanılmaktadır (Hagan vd., 2014:1-6).
- ✓ *Üretim Sanayii*: Ürün tasarımı ve analizi, proses ve makine teşhis etme, üretim proses kontrolü, kalite kontrol sistemleri, makine bakım analizi, planlama ve yönetim kimyasal süreç sistemlerinin dinamik modellenmesi gibi uygulamalarda yapay sinir ağları kullanılmaktadır (Hagan vd., 2014:1-6).
- ✓ *Savunma Sanayii*: Yüz tanıma, yeni tür sensörler, silah yönlendirme, hedef izleme, radar ve görüntü sinyali, veri sıkıştırma, sinyal/görüntü tanımlama, gürültü bastırma gibi uygulamalarda yapay sinir ağları kullanılmaktadır (Hagan vd., 2014:1-6).
- ✓ *Sigortacılık*: Sigorta sektörün de dolandırıcılığın tespit edilmesi, olağan dışı olayların tespiti, poliçe sahiplerinin tutumlarına göre ayrılması ve etkin prim fiyatlandırmanın belirlenmesi gibi fayda sağlayan uygulamalar için yapay sinir ağı kullanılmaktadır (Smith,Gupta, 2000:1037).
- ✓ *Pazarlama*: Pazar bölümlenme, yeni ürün analizi, müşteri özelliklerinin belirlenmesi, tüketici harcama modelinin piyasaya etkilerinin sınıflandırılması, pazara yönelimin belirlenmesi gibi uygulamalar için yapay sinir ağları kullanılmaktadır (Hakimpoor vd.,2011:1010-1011).
- ✓ *Borsa ve Para Yönetimi*: Borsanın karmaşık yapısı ve volatilitelerini dikkate alarak şirketlere ait verileri, ekonomik puan, politik puan, anapara ve anaparadaki değişim, spekülasyonlar gibi veriler ışığında yapay sinir ağı eğitilerek yatırım hareketleri üretilebilir(Efe,Kaynak,2000:110-111).
- ✓ *Tıp ve Sağlık*: Yapay sinir ağları RNA ve DNA'da aminoasit incelemelerinde, hastaların ilaçlara karşı gösterdiği reaksiyonların incelenmesinde, kanser teşhisinde, kazalardan hasar almadan korunmada, ilaç sanayisinde zaman ve maliyet tasarrufu etmede oldukça sık kullanılmaktadır (Elmas, 2011:164).
- ✓ *Ulaştırma ve Havacılık*: Otomobiller de fren denetimi, mühendislik hataları, uçaklarda motor sorunlarının giderilmesi adına titreşim seviyeleri ve sesin görüntülenmesiyle birlikte erken uyarı sistemi geliştirilmesi, havalimanlarında bagaj geçişlerinde bomba gibi tehlikeli maddelerin ortaya çıkarılması amacıyla yapay sinir ağları kullanılmaktadır (Elmas, 2011:163).

- ✓ *Telekomünikasyon:* Ürünlerin satışı ile birlikte müşterilerin ürünlere bağımlı hale gelmesi ve buna bağlı olarak müşteri devamlılığının sağlanması, bir müşteri kaybı yaşandığında nedenlerinin belirlenmesi, müşterinin rakip firmaya katılmasının zamanının ve nedeninin tahmin edilmesi ve bununla birlikte birçok mühendislik uygulamaları da dahil olmak üzere telekomünikasyon endüstrisinde yapay sinir ağları kullanılmaktadır (Smith,Gupta, 2000:37).

2.1.11.Yapay Sinir Ağlarının Avantajlı ve Dezavantajlı Yönleri

Yapay sinir ağlarının yukarıda açıklanan tüm özellikleri doğrultusunda avantajlı yönleri aşağıdaki gibi sıralanabilir (Yurtoğlu, 2005:35-36):

- ✓ Yapay sinir ağları insan beynini taklit eden yapısı aracılığıyla öğrenme yöntemi ile veriyi eğitebilir böylece geleneksel teknikler için oldukça karmaşık olarak nitelendirilen problemlere kolaylıkla çözüm üretebilmektedir.
- ✓ Tek bir merkezi işlem elemanının sırasıyla gerçekleştirdiği harekete rağmen yapay sinir ağları esneklik özelliği sayesinde çok sayıda basit işlem elemanından oluşan bağlantı ağırlıklarının ayarlanmasına olanak veren her biri büyük bir problemin parçası ile ilgilenen bağlantılardan oluşur. Buna bağlı olarak ağı bir bölümünde meydana gelen sorun modelde, performans düşüklüğüne neden olsa da modelin çalışmasını tamamen etkilememektedir.
- ✓ Paralel olarak çalışabilen yapay sinir ağları gerçek zamanlı işlemleri gerçekleştirebilmektedir.
- ✓ Sınırsız sayıda değişken ve parametre ile çalışan yapay sinir ağları oldukça güçlü çıkarımlarda bulunabilecek sonuçlar üretmektedir.

Yapay sinir ağlarının dezavantajlı yönleri aşağıdaki gibi sıralanabilir (Gonzalez, 2007:29;Araç & Gürhanlı, 2020:42;Öztemel, 2012:35):

- ✓ Yapay sinir ağlarında ağ mimarisi deneme yanılma yoluyla tasarlanmaktadır. Bu sebeple zaman açısından süreç, yazılım paketlerini programlayarak her ne kadar kısaltılsa da birkaç ağı tasarlamak ve uygun bir şekilde değerlendirmek oldukça zaman almaktadır.
- ✓ Yapay sinir ağları sadece sayısal bilgiler ile çalışmaktadır bu nedenle problemlerin yapay sinir ağlarıyla çözümüne başlanmadan önce sayısal değerlere çevrilmesi gerekmektedir burada kullanıcının performansına bağlı

olarak belirlenecek görüntüleme mekanizması ağın performansına doğrudan etki edecektir.

- ✓ Yapay sinir ağları deneme yanılma ve kullanıcının bilgi ve tecrübesine göre sondalama çözüm ürettiğinden ağın davranışı açıklanamaz neden ve nasıl olduğu bilinmez ve buna bağlı olarak en iyi çözüm olduğunun bir kesinliği olmaması ağa olan güveni azaltmaktadır.
- ✓ Ağın eğitiminin ne zaman bitirileceğine karar veren bir kural bulunmamaktadır. Ağın hatasının belirli bir değerin altına indirilmesi ağın eğitiminin tamamlanması için yeterlidir ancak bu yeterlilik optimum sonuca ulaşabilen bir mekanizma bulunmadığını göstermektedir.

2.2.Birliktelik Kuralı Analizi

Genelleştirilmiş öngörüşel yöntemler ve sepet analizi olarak adlandırılan veri madenciliği yöntemlerinden biri olan birliktelik kuralı analizi belirli kurallar çerçevesinde sonuçlar üreterek ilişki kurmayı amaçlayan bir analizdir (Ersöz, 2019:85). Birliktelik kuralı analizi ele alınan veri setinde sıkça tekrar eden nitelik-değerli koşulları gösteren kuralların bulunması işlemine dayanmaktadır (Oğuzlar, 2004b:317).

Birliktelik kuralı analizinde amaç ele alınan veri setindeki, birlikte artan veya birlikte azalan değerleri tespit ederek birbiriyle ilişkilendirmektir (Özkan,Boran,2014:65). Sınıflandırma analizi ile benzerlik gösteren birliktelik kuralları analizi sınıflandırma analizinde olduğu gibi yalnızca sınıfı belirlemek ile kalmayıp, kurallar doğrultusunda verilerin özellikleri ile ilgili de çıkarımlar yapmaktadır. Üstelik birliktelik kurallarında küme olarak birlikte kullanılması söz konusu olmamaktadır. Birliktelik kuralları farklı kurallar üreterek veri kümesinin altında yatan farklı düzenlilikleri ifade etmekte ve çok küçük veri kümesinden dahi birçok birliktelik kuralı üretilebilmektedir (Witten vd., 2011:72).

Birliktelik kuralı yöntemi birden çok seçim yapılabilen gözlemlerin, bu seçimleri arasındaki birliktelik yapılarının oluşturulmasını sağlayan yöntemlerdir (Altunkaynak, 2017:18).

Veri ögeleri veya model için seçilen olaylar arasında önemli bağımlılıklar veya başka bir değişle ilişkiler tanımlayan birliktelik kuralı analizi, bağımlılık analizi olarak da adlandırılmaktadır. Bulunan ilişkiler ilgilenilen veri ögeleri veya olaylar hakkında bilgi vererek tahmine dayalı modeller arasında yer almaktadır (Shearer, 2000: 21).

Birliktelik kuralı madenciliği iki adımlı bir süreçten oluşmaktadır(Oğuzlar, 2004a:49):

1. Problemin çözümünde kullanılacak olan veri setindeki bütün sıkça tekrar eden kümeler bulunmaktadır.
2. Tekrar edilen kümelerden yola çıkarak güçlü çıkarımlar elde edilmesi amacıyla birliktelik kuralları üretilmektedir.

Birliktelik kuralı analizinde kuralların türetilmesi ile sol ve sağ olmak üzere iki kısım oluşmaktadır. Sol taraf öncül veya başka bir deyişle koşulu ifade ederken sağ taraf ise bu koşula bağlı olduğunu gösteren bağlı koşulu ifade etmektedir. Sol taraf ve sağ taraf birbirine bağlı hareket ettiği için sol taraf doğruyu gösterirken sağ tarafta aynı şekilde doğruyu göstermektedir. Buna göre birliktelik kuralının genel formu aşağıda gibidir (Şentürk, 2006: 19):

$$\text{IF } \langle \text{ön koşul} \rangle \text{ THEN } \langle \text{bağlı koşul} \rangle \quad (2.22)$$

Burada ön koşul bir ya da birden fazla girdi verisinin değerini veya değer aralığını gösterirken bağlı koşul ise, bir ya da birden fazla çıktı verisinin değerini göstermektedir (Şentürk, 2006:19).

2.2.1.Birliktelik Kuralı Analizi Matematiksel Gösterimi ve Kavramları

Veri madenciliğinde tanımlayıcı modeller arasında yer alan birliktelik kuralı analizinin matematiksel gösterimi 1993 yılında Agrawal, Imielinski ve Swami tarafından ortaya konulmuştur (Agrawal vd.,1993).Modele göre $I = \{i_1, i_2, \dots, i_m\}$ kümesi öge kümesini gösterirken her bir i değeri ögeyi başka bir deyişle ürünü ifade etmektedir.D tüm hareketleri,T ise i 'deki ürünlerin her bir hareketini gösterirken,her hareket tanımlayıcı TID ile ifade edilmektedir.Buna göre X ve Y bir kümenin ögeleri olması durumunda T

işlemler kümesinin yalnızca X, T 'nin alt kümesi olması durumunda A 'yı kapsadığı görülmektedir (Necati, Duykuluoğlu, 2020:200).

Bir birliktelik kuralı aşağıdaki şekilde ifade edilmektedir (Silahtaroğlu, 2016:140):

$X \Rightarrow I_j, X, I$ nın bir alt kümesidir. Burada I_j I içindeki bir elemanı göstermektedir ve bu eleman X içinde bulunmamaktadır. $X \Rightarrow I_j$, kuralının T için uygun olup olmadığının belirlenmesinde güven seviyesinden yararlanılmaktadır. Güven seviyesi c ile gösterilirken T içindeki tüm X 'lerin ne kadarının I_k 'yı sağladığını göstermektedir. Destek seviyesi ise T içindeki işlemlerin ne kadarının X 'i sağladığını göstermektedir.

Birliktelik kuralının belirlenmesinde önemli kavramlar rol oynamaktadır. Bu kavramlar destek ve güven olarak adlandırılırken destek değeri öncül ve sonuç kısımlarında yer alan verileri içeren işlem hacminin veri tabanındaki toplam işlem sayısına oranıyla elde edilmekte iken, güven değeri öncül ve sonuç kısımlarında yer alan verileri içeren işlem hacminin öncül işlem sayısına oranlanmasıyla bulunmaktadır (Tüzüntürk, 2010: 82).

Birliktelik kuralları analizinde kurallar hakkında çıkarımlar yapabilmek için aşağıdaki değerlere ihtiyaç duyulmaktadır (Altunkaynak, 2017:120):

Destek(Support):İlgilenilen seçeneğin ya da seçenekler kümesinin klasik olasılığını gösteren destek değeri, ilgili seçeneğin ya da seçenekler kümesinin bütün gözlemler içindeki oranını göstermekte ve aşağıdaki gibi hesaplanmaktadır:

n : gözlem sayısı

$N(X)$ = X seçeneğinin görülme sayısı, $N(X, Y)$ = X, Y seçenek kümesinin birlikte görülme sayısını gösterirken; X seçeneği için destek değeri

$$P(x) = \frac{N(x)}{n} \quad (2.23)$$

X ve Y seçenek kümesi için destek değeri

$$P(X, Y) = \frac{N(X, Y)}{n} \quad (2.24)$$

olarak bulunmaktadır. Birliktelik kuralını oluşturabilmek için D ile sembolize edilen minimum destek seviyesinin önceden belirlenmesi gerekmektedir ve minimum destek seviyesi $0 \leq D \leq 1$ arasında değer almaktadır. Araştırmacılar tarafından uygulamalarda genellikle 0.30 alınan minimum destek değeri 0' a yaklaştığında daha fazla birliktelik kuralı üretmekte ve birlikteliklerin sayısı artmakta, 1'e yaklaştığında ise daha az birliktelik kuralı üretmekte ve birliktelik bulma azalmaktadır.

Güven (Confidence): İlgilenilen veri kümesinde X seçeneğini seçenlerin aynı zamanda Y seçeneğini de seçme olasılığını gösteren koşullu olasılık değeri olarak tanımlanmaktadır. Bu değer aşağıdaki gibi hesaplanmaktadır:

$$P(Y|X) = \frac{N(X,Y)}{N(X)} \quad (2.25)$$

veya

$$P(Y|X) = \frac{P(X,Y)}{P(X)} \quad (2.26)$$

şeklinde ifade edilmektedir. G ile sembolize edilen güven değeri $0 \leq G \leq 1$ arasında değer almaktadır. Birliktelik kuralı analiz sonuçlarını yorumlamak için güven değeri kullanılan oldukça önemli bir değerdir. Güven değeri olarak araştırmacılar, uygulamalarda genellikle 0.50 veya 0.50'den daha büyük değerler seçmektedirler.

Kaldıraç (Lift) : İstenilen bir X seçeneğinin B yerine A içinde aranmasının yani A'nın B'nin alt kümesi $A(A \subset B)$ olması durumunda o özellik ile karşılaşma olasılığının kaç defa arttırılacağını göstermektedir. K ile sembolize edilen kaldıraç değeri ilgilenilen veri setinden elde edilen destek değerinin güven değerine bölünmesi ile bulunmaktadır ve aşağıdaki gibi hesaplanmaktadır:

$$K(X) = \frac{P(X|Y)}{P(X)} \quad (2.27)$$

Yukarıda yer alan değerlendirme ölçütleri temel alındığında, birliktelik kuralı analizinde istenen destek ve güven değerlerinin yüksek olduğu sonuçlar elde etmektir (Köse, 2018:188).

2.2.2.Birliktelik Kuralı Analizi Algoritmaları

Birbirleri ile belirli sıklıkta hareket eden verilerin analizi ve bu analiz sonuçları doğrultusunda üretilen birliktelik kuralları çeşitli algoritmalar yardımıyla hesaplanmaktadır(Akpınar, 2017:209).

2.2.2.1.AIS Algoritması

AIS Algoritması 1993 yılında geliştirilen ilgili veritabanındaki nesnelerin isimlerini A'dan Z'ye sıralama temeline dayanan geniş nesne kümeleri üretmek amacıyla geliştirilmiş birliktelik kuralı analizinde kullanılan algoritmalarından biridir (Silahtaroglu, 2016:141).

Veritabanlarını birden fazla tarama yoluyla büyük öge kümeleri oluşturmayı amaçlayan AIS algoritması, ilk tarama esnasında tek tek öğelerin desteğini sayarak büyük veya sık olduğunu belirlemektedir. Her geçişte büyük öge kümeleri, aday öge kümeleri oluşturmak için genişletilirken işlem taramasının ardından, bir önceki taramada tespit edilen büyük ürün kümeleri ile üzerinde çalışılan işlemin öğeleri arasındaki ortak öge kümelerini belirlemektedir. Bulunan bu ortak öge kümeleri, yeni aday öge kümeleri oluşturmak için diğer öğelerle genişletilmektedir. Büyük bir, l nesne kümesi sadece büyük olan ve alfabetik sıralamasında l'deki öğelerin herhangi birinden sonra meydana gelen işlemde yer alan öğelerle genişletilmektedir (Dunham vd.,2000:6).

2.2.2.2.SETM Algoritması

AIS algoritmasına benzer olarak SETM algoritması da veritabanını tarayarak sayma işlemi yapmaktadır. Fakat SETM algoritması SQL için oluşturulmuş olup ilişkisel işlemleri kullanmaktadır (Girotra, vd.,2013:19).Bu algoritma büyük öge kümelerinin tespiti için, tanımlanan nesnenin ismi ve nesneye ait olan özellik numarası olarak adlandırılan TID (Tansaction Identification) numarasını da ele almaktadır (Silahtaroglu, 2016:141).

SETM algoritması büyük öge kümelerini bulmak için aday öge kümelerini veritabanının taranması sırasında oluşturmakta ve geçişin sonunda sayma işlemini yapmaktadır. Oluşturulan işlemin TID'si ile birlikte aday olan öge kümesinin bir kopyasını sıralı olarak kaydederken geçişin sonunda aday öge kümelerinin destek sayısını belirleme işlemini ise bu sıralı yapının toplanması ile yapmaktadır (Khurana,Sharma, 2013:2).

2.2.2.3.Apriori Algoritması

Apriori Algoritması, 1994 yılında Agrawal ve Srikant tarafından geliştirilmiş olan analiz öncesi önsel bilgiler aracılığıyla destek veya güven seviyelerinden en az bir tanesi dikkate alınarak yeni bilgiler keşfetme sürecinden oluşmaktadır (Altunkaynak, 2017:120).

Apriori Algoritması, veritabanında k adet nesneyi bünyesinde barındıran bir kümenin k-1 adet nesneye sahip olan geniş nesne kümelerini bularak birleştirmesi ve ardından alt kümeleri geniş olmayanları silerek analiz dışı bırakması işlemlerine dayanan bir süreçten oluşmaktadır (Karaibrahimoğlu, 2014:66).

Apriori algoritması, veritabanından ilk geçişte büyük 1 öge kümelerini tespit etmek için öge oluşumlarını sayarak işleme başlamaktadır. Ardından bir sonraki geçişi gösteren k geçişi ile apriori-gen fonksiyonu aracılığıyla önce, k-1'inci geçişte bulunan büyük L_{k-1} öge kümelerini, aday nesne kümeleri C_k 'ı oluşturmak için kullanmaktadır. Bu işlemi takiben ikinci olarak veritabanı tarama işlemi yapılarak C_k 'daki aday nesne kümelerinin destek değeri azaltılmaktadır. Son olarak sayım işleminin hızlı yapılması adına belirli bir t işleminde yer alan C_k 'daki adayların doğru belirlenmesi gerekmektedir (Agrawal & Srikant, 1994:489).

Aşağıda Şekil 2.17.'de Apriori algoritmasının kodları verilmiştir:

```

1)  $L_1\{ \text{Geniş 1-öge kümeleri} \}$  ;
2) for( $k = 2$  ;  $L_{k-1} \neq 0$ ;  $k++$ ) do begin
3)  $C_k = \text{apriori-gen}(L_{k-1})$ ; // Yeni adaylar
4) forall işlemler  $t \in D$  do begin
5)  $C_t = \text{altküme}(C_k; t)$  ; //  $t$  de bulunan adaylar
6) forall adaylar  $c \in C_t$  do
7)  $c.\text{count}++$ ;
8) end
9)  $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$ 
10) end
11)  $\text{Answer} = \bigcup_k L_k$  ;

```

Şekil 2.17: Apriori algoritması

Kaynak: (Agrawal ,Srikant, 1994:489)

En yaygın kullanıma sahip Apriori algoritması aşağıdaki aşamaları içermektedir (Özkan ,2020:219):

- ✓ Analize başlamadan önce ilk olarak birliktelik kuralında destek ve güven seviyeleri ile karşılaştırma yapmak amacıyla eşik değeri belirlenmektedir. Belirlenen eşik değeri analiz tamamlandıktan sonra çıkan sonuçlara eşit ya da büyük olması istenmektedir.
- ✓ Analiz için kullanılacak olan veritabanlarındaki ürünler için tekrar sayıları olarak adlandırılan destek sayıları hesaplanmaktadır. Hesaplanan destek sayıları belirlenen eşik değerden küçük olması durumunda analiz dışı bırakılır ancak koşula uygun sonuçlar analizde kalmaya devam etmektedir.
- ✓ Bir önceki adımda seçilen ürünler ikili gruplar halinde sınıflandırmaya tabii tutulduğunda aynı şekilde destek ve eşik değeri karşılaştırılarak, eşik değerinden küçük değerler analiz dışında bırakılmaktadır.
- ✓ Oluşturulan ikili gruplardan sonra ürünler üçerli, dörderli veya daha fazla gruplandırmalara tabii tutularak tekrar eşik değeri temel alınarak karşılaştırılmakta ve eşik değerine uygunluğuna göre işleme devam edilmektedir.
- ✓ Son olarak belirlenen ürün grubunun ardından destek ölçütüne göre kurallar üretilmekte ve üretilen kurallar için güven seviyeleri hesaplanmaktadır.

2.2.2.4.Apriori-TID Algoritması

Apriori Algoritmasında olduğu gibi Apriori TID algoritmasında da aday öge kümeleri oluşturma işlevi gerçekleşmektedir. Ancak Apriori TID algoritmasında veritabanından ilk geçiş sonrası sayım desteğine başvurulmamaktadır. Algoritma $k > 1$ için, bir dizi aday öge kümesi kullanmaktadır. Bir işlemin k -öge kümesi bulunmadığında, aday öge kümelerinin o işlem için bir girişi olmamaktadır. Bu sebeple veritabanına kıyasla aday öge kümelerini içeren işlem sayısında düşüşler meydana gelmektedir. Böylece k değerinin artması ile birlikte işlemlerdeki aday sayısı azalmaya devam etmektedir ve dolayısıyla her giriş, karşılık gelen işlemlerden daha küçük olarak bulunmaktadır. İlk geçişlerde Apriori-TID algoritmasına göre daha iyi performans gösteren Apriori algoritması, daha fazla geçiş vermesi nedeniyle Apriori-TID algoritmasının gerisinde kalmaktadır (Girotra vd.,2013:20).

2.2.2.5.Apriori-Hybrid Algoritması

Apriori-Hybrid algoritması, Apriori ve Apriori-TID algoritmalarının birleşiminden oluşmaktadır (Silahtaroglu, 2016:146).

Veritabanı taranırken tüm geçişlerde aynı algoritma kullanılması zorunlu değildir. Apriori algoritması veri tabanındaki her işlemi incelerken Apriori-TID veritabanında Ck'nın destek sayılarını bulmak için tarayarak veritabanından daha küçük hale getirmektedir. Bu algoritmalar göz önünde bulundurularak Apriori-Hybrid algoritması geliştirilmiştir (Khurana,Sharma,2013:3).Apriori-Hybrid algoritması ilk geçişte Apriori algoritmasını kullanarak geçişin sonunda aday öge setlerinin bellekte olmasını beklediğinden Apriori-TID algoritmasına geçmektedir (Girotra vd.,2013:20).

2.2.2.6.FP-Growth Algoritması

Apriori algoritmasının çok sayıda aday seti oluşturmaları gerektirmesi ve tüm veritabanlarının tekrar tekrar taramasını yaparak aday öge kümelerinin desteğini bulmak amacıyla veritabanındaki her işlemin üzerinden geçmesi algoritmanın önemli dezavantajlarındandır. Apriori algoritmasının verdiği önemli dezavantajları giderebilmek için daha avantajlı bir algoritma olan FP-Growth kullanılmaktadır. FP-Growth algoritması böl ve yönet stratejisini benimseyerek adından da anlaşıldığı üzere sık model büyümesi yani FP büyümesi gerçekleştirmektedir. Algoritma ilk olarak sık öğeleri temsil eden veritabanını, öge kümesi ilişkilendirme bilgilerini içeren FP ağacına

bir başka deyişle sık model ağacına sıkıştırılmaktadır. Ardından sıkıştırılmış veritabanını, her biri sık kullanılan öge ile ilişkili bir dizi koşullu veritabanına bölerek her veritabanını ayrı ayrı incelemektedir. Her bir öge için yalnızca ilişkili veri setlerini inceleyen bu algoritma modellerin büyümesi ile birlikte aranacak veri setlerinin boyutunu da önemli ölçüde azaltmaktadır (Han vd.,2012:257).

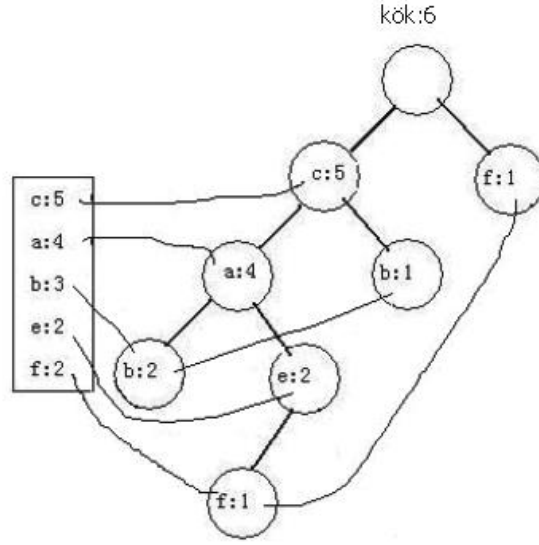
F-P Growth algoritması, özellik-değer çiftlerinin eşleşmesini sayarak başlamaktadır. Bu ilk geçiş işleminden sonra, ikinci geçişte başlangıçta boş olan ancak geçişten sonra oluşan bir ağaç yapısı görüntüsü meydana gelmektedir. Algoritma büyük öge kümelerini bularak hızlı bir şekilde ağacı oluşturmak için her durumdaki öğeleri eklemekten önce ilk geçişte kaydedilmiş olan veri kümesindeki oluşma sıklıklarını temel alarak azalan düzeyde sıralamaktadır. Böylece minimum destek eşiği şartını sağlamayan her örnekteki ayrı öğeler ağaca dahil edilmeyerek veri kümesinden kaldırılmaktadır (Witten vd.,2011:216).

FP-Growth algoritmasında oluşturulan FP ağacında öge setleri aynı örnek yolunu paylaştığı için avantaj sağlamaktadır. Dolayısıyla veri seti bilgileri büyük ölçüde sıkıştırılırken veri kümesi yalnızca iki kez taranmakta ve aday öge kümesine ihtiyaç duymamaktadır (Wu vd.,2008:1100). Aşağıda Şekil 2.18.'de veri kümesi ve Şekil 2.19.'da ilgili veri kümesine ait %30 minimum destek seviyesini gösteren FP ağacı örneği verilmiştir:

İşlemler
a b c
a c e f
d f
a b c
a c e g
b c

Şekil 2.18: Örnek Veri Seti

Kaynak: (Wu vd., 2008: 1100)



Şekil 2.19: Örnek Veri Setine Ait FP-Tree (FP-Ağacı)

Kaynak: (Wu vd., 2008:1100)

2.2.3.Birliktelik Kuralı Analizi Algoritmalarının Karşılaştırılması

AIS algoritması veritabanının tekrar tekrar taranmasına dayanan bir algoritma olması nedeniyle ve küçük çok sayıda aday öge seti oluşturması, buna bağlı olarak da daha fazla alan gerektirmesi açısından dezavantajlı olduğunu göstermektedir. SETM algoritmasının ise AIS algoritmasına göre daha az aday öge seti oluşturması nedeniyle daha avantajlı olduğu görülmektedir (Mishra vd., 2013:156).

Apriori-TID algoritması standart veritabanı işlemlerini kullanmadan apriori-gen kullanarak daha hızlı aday öge üretimi yapan SETM algoritmasının geliştirilmiş bir versiyonudur. Buna göre Apriori-TID algoritması ile Apriori algoritması karşılaştırıldığında aynı sayıda aday öge kümesi üretmelerine rağmen veri sunumunda Apriori-TID algoritması farklılıklar göstermesinden dolayı daha cazip sonuçlar göstermektedir. Bu bağlamda veritabanını yalnızca ilk geçişte taraması bakımından Apriori-TID algoritması Apriori algoritmasına göre daha fazla avantaj sağlamaktadır (Mishra vd.,2013:157).

Veri madenciliğinde kullanılan birliktelik kuralı analizinde diğer algoritmalara göre en hızlı ilişkilendirme yapan algoritma FP-Growth algoritmasıdır (Saxena,Rajpoot, 2020,

s. 5).FP-Growth algoritması Apriori algoritmasına göre daha ölçeklenebilir bir yapıda olması ve daha hızlı çalışması bakımından, Apriori algoritmasına üstünlük sağlamaktadır. Ayrıca minimum destek sayısının küçük olma durumu Apriori algoritmasında aday öge küme sayısının artışına sebep olurken, bu artış FP-Growth algoritmasında görülmemektedir (Erpolat,2012:140).

Aşağıdaki Tablo 2.4’de birliktelik kuralı analizinde kullanılan çeşitli algoritmaların birbirlerine karşı sağladıkları üstün ya da zayıf yönleri çeşitli kriterlere göre karşılaştırılmıştır (Kumbhare,Chobe, 2014:930):

Tablo 2.4: Birliktelik Kuralı Madenciliği Algoritmalarının Karşılaştırılması

Özellikler	AIS	SETM	Apriori	Apriori- TID	Apriori-Hybrid	Fp-growth
Veri Destek	Az	Az	Sınırlı	Genellikle büyük	Çok büyük	Çok büyük
İlk Aşama Hızı	Yavaş	Yavaş	Yüksek	Yavaş	Yüksek	Yüksek
Sonraki Aşama Hızı	Yavaş	Yavaş	Yavaş	Yüksek	Yüksek	Yüksek
Doğruluk	Çok az	Az	Az	Apriori’den daha doğru	AprioriTID’den daha doğru	Daha doğru

Kaynak: (Kumbhare , Chobe, 2014: 930)

Birliktelik kuralı analizinde kullanılan algoritmalar Tablo 2.4’de veri desteği, ilk aşama ve son aşama hızı, doğruluk kriterlerine göre değerlendirilmiştir. Veri destek kriterine göre, Apriori-TID, Apriori-Hybrid ve FP-Growth algoritmalarının daha büyük, ilk aşama hızı kriterinde Apriori, Apriori-Hybrid ve FP-Growth algoritmalarının daha hızlı, sonraki aşama hızı kriterine göre Apriori-TID, Apriori-Hybrid ve FP-Growth algoritmalarının daha hızlı ve doğruluk kriterine göre de en doğru sonucu FP-Growth algoritmasının verdiği görülmektedir. Tüm bu kriterler performans açısından göz önüne alındığında ise, en iyi performansı gerçekleştiren algoritmanın FP-Growth algoritması olduğu sonucuna varılmaktadır (Kumbhare,Chobe, 2014: 930).

2.2.4.Birliktelik Kuralı Analizi Kullanım Alanları

Birliktelik kuralı analizi, veri madenciliğinde yaygın kullanım alanlarına sahiptir. En yaygın olarak bilinen birliktelik kuralı analizi uygulamaları pazarlama alanında gerçekleşmektedir. Örneğin bu alanda müşterilerin alışveriş esnasında satın aldıkları ürünler arasında ilişkiler kurulmakta, birbirleriyle ilişkili ürünler marketlerde yakın reyonlara sıralanarak stratejiler geliştirilmekte ve müşteriler daha fazla ürün satın almaları için yönlendirilmektedir. Böylece satışların ve karın artırılması sağlanmaktadır.

Birliktelik kuralları, tıp finans alanlarında da kullanılmakta olup farklı olayların birbirleriyle olan ilişkilerinin tespit edilmesi ve bu neticede önemli çıkarımlar elde edilmesi amaçlanan problemlerde önemli bir yere sahiptir(Akpınar, 2000:7).

Birliktelik kuralları aşağıdaki örneklerde verildiği gibi eş zamanlı olayların gerçekleşmesinde ilişkilendirmeler yapmaktadır(Akpınar, 2000:7).

- ✓ Müşteriler bira satın aldığı anda,%75 ihtimalle patates cipsi de almaktadırlar,
- ✓ Düşük yağlı peynir ve yağsız yoğurt alan müşteriler,%85 ihtimalle diyet süt de satın almaktadırlar.

Birliktelik kuralları eğitim alanında öğrencilerin ders seçimleri üzerine yapılan araştırmalarda ve öğrencilerin başarılı oldukları dersleri tespit etmede kullanılırken, finans alanında da birlikte tercih edilen finansal ürünleri saptamada kullanılmaktadır. Bunlara ek olarak trafik kazalarına sebep olan davranışların hangilerinin birlikte gerçekleştiğini öngörmek için de birliktelik kuralı analizinden yararlanılmaktadır(Bayram , Dünder, 2021: 42-43).

ÜÇÜNCÜ BÖLÜM

VERİ MADENCİLİĞİ ÇERÇEVESİNDE YAPAY SİNİR AĞLARI VE BİRLİKTELİK KURALI ANALİZİ ÜZERİNE BORSA İSTANBUL 30 ENDEKSİNDE BİR UYGULAMA

3.1 Literatür Taraması

Literatürde yurtiçi ve yurtdışında veri madenciliği yöntemlerinden yapay sinir ağları ve birliktelik kuralı analizi ile yapılan birçok araştırma bulunmaktadır. Bu bölümde literatürde yer alan araştırmaların uygulama alanları, analiz teknikleri ve ne gibi sonuçlara ulaşıldığı hakkında bilgi verilecektir.

3.1.1. Ulusal Literatür

Tektaş ve Karataş (2004) araştırmalarında 2002-2003 yılları arasında İMKB’de işlem gören çimento ve gıda sektörüne ait 7 hisse senedinin fiyatı üzerinde yapay sinir ağları ve regresyon yöntemini kullanarak çıkarımlarda bulunmuşlardır. Çalışmada iki veri seti oluşturularak hem haftalık hem de günlük veriler için korelasyon değerleri karşılaştırılmıştır. Haftalık veri setinde girdi değişkeni olarak aylık enflasyon, ağırlıklandırılmış mevduat faiz oranı, İMKB 100 Endeksi, İMKB Sınai Endeksi ve T.C Merkez Bankası Amerikan Doları satış kuru kullanılırken günlük veri setinde ise faiz ve enflasyon değerleri analizde yer almamıştır. Sonuç olarak, günlük verilerin haftalık verilere göre performansı arttırmada daha etkili olduğu görülmüştür. Ayrıca haftalık veriler ile yapılan yapay sinir ağları ve doğrusal regresyon analizi karşılaştırılmasında yapay sinir ağları modelinin tahmin gücünün daha başarılı olduğu bulunmuştur (Tektaş,Karataş, 2004:337-349).

Toraman (2008) araştırmasında 02.01.2008-30.01.2018 tarihleri arasında 22 işlem gününe ait kapanış fiyatlarını alarak Erdemir A.Ş. ve Kardemir A.Ş(D grubu) şirketlerinin hisse senedi fiyatlarını yapay sinir ağları yöntemi ile tahmin etmeye çalışmıştır. Girdi değişkeni olarak Dünya ve Türkiye ekonomisine ait göstergeler ve analiz için seçilen iki işletmeye ait bilanço ve gelir tablolarından elde edilen oranlar kullanılmıştır. Yapay sinir ağları ile tahmin edilen modelde veri seti eğitim doğrulama

ve test olarak üçe ayrılmıştır. Sonuç olarak, yapay sinir ağları yönteminin oldukça başarılı tahminler gerçekleştirdiği görülmüştür(Toraman, 2008: 44-57).

Kutlu ve Badur (2009) araştırmalarında ileri beslemeli yapay sinir ağları ve hareketli ortalama yöntemi ile İMKB 100 endeksini tahmin etmeye çalışmışlardır. 2001-2006 yıllarını baz alarak yapılan bu çalışmada girdi değişkenleri olarak önceki günün endeks verisi, önceki günün Amerikan doları verisi, önceki günün gecelik faiz verisi, haftanın günlerinin etkisi ve önceki güne ait Fransa, Almanya, İngiltere, NASDAQ, DOW JONES, S&P500,Brezilya, Japonya borsa endeks verilerini kullanmışlardır. Bu veriler ışığında üç adet model kurularak R^2 değerleri yorumlanmıştır. Buna göre M1 modelinin R^2 değeri 0.96 bulunarak en iyi model olarak seçilmiştir. Seçilen modelde değişkenler önceki günün borsa endeksi, Amerikan doları ve gecelik faiz değerleri girdi olarak kullanılmıştır. Araştırma sonucunda hareketli ortalama yöntemi ile tahmini yapılan İMKB 100 endeksinde doğruluk oranı %50,4 olarak bulunurken yapay sinir ağları kullanılarak yapılan tahminlemenin doğruluk oranı %55,1 olarak bulunmuştur. Böylece yapay sinir ağlarının performans olarak daha iyi sonuçlar elde ettiği görülmüştür (Kutlu,Badur, 2009: 25-40).

Akcan ve Kartal (2011) araştırmalarında İMKB’de işlem gören yedi sigorta şirketinin hisse senedi fiyatlarını 15 gün-2 ay aralığında tahmin etmeyi amaçlamışlardır. Araştırma kapsamında girdi değişkeni olarak makro ve mikro ekonomik değişkenlerden yararlanılmışlardır.Makro ekonomik değişken olarak İMKB 100 endeksi günlük kapanış verileri, tüketici fiyat endeksi, dolar efektif satış kuru, günlük cumhuriyet altın fiyatı, kullanılırken mikro ekonomik değişken olarak ise; fiyat/kazanç oranı, piyasa değeri/defter değer oranı, aktif karlılık oranı, alınan net prim büyüme hızı oranı, alınan primler/ toplam aktifler oranı, öz sermaye karlılık oranı, hisse başına kar tutarı, net kar/alınan primler oranı değişkenlerini kullanmışlardır.Araştırmada geriye yayımlı çok katmanlı yapay sinir ağları yöntemi ve performans ölçümü olarak MAPE ve MAE kriterleri ele alınmıştır. Araştırma sonucunda yapay sinir ağı ile iki aya kadar yapılan tahminlerin başarılı sonuçlara ulaştığı kanısına varılmıştır (Akcan,Kartal, 2011:27-40).

Erdoğan ve Özyürek (2012) tarafından yapılan araştırmada İMKB de işlem gören beyaz eşya sektörüne ait olan Arçelik, Alarko, Vestel Beyaz Eşya, Vestel TV, İhlas Ev

Aletleri şirketlerinin 2011-2012 yılındaki haftalık verileri alınarak cuma günü kapanış fiyatlarının tahmini gerçekleştirilmiştir. Girdi değişkeni olarak dolar kuru, İMKB endeksi, pazartesi, salı, çarşamba ve perşembe günleri kapanış fiyatları alınmış olup 6 girdi değişkenine sahip giriş katmanı, bir gizli katman ve bir çıkışa sahip çıkış katmanından oluşan yapay sinir ağı modeli oluşturulmuştur. Modeli yorumlamak adına kriter olarak ortalama yüzde hata kriteri baz alınmıştır. Araştırma sonucunda yapay sinir ağlarının hisse senedi fiyat tahmininde başarılı sonuçlara ulaştığı bulunmuştur(Erdoğan , Özyürek, 2012:85-92).

Gemici (2012) tarafından hazırlanan tez çalışmasında İMKB’de işlem gören 10 hisse senedinin aralarındaki birlikteliklerin belirlenmesi amaçlanmıştır. Bu amaç doğrultusunda 63 günlük işlem günü verileri için minimum güven seviyesi %80 ve minimum destek seviyesi de %35 olarak seçilmiştir. Analizde kullanılan birliktelik kuralı algoritmalarından apriori algoritması sonuçlarına göre belirlenen destek ve güven seviyelerine göre 9 birliktelik kuralı üretilmiştir (Gemici, 2012).

Yakut ve diğ.(2014) tarafından yapılan çalışmada BIST endeksi tahmin edilmeye çalışılmıştır. Veri seti olarak 2005-2012 yılları arasındaki borsa endeksinin bir iki ve üç gün öncesine ait değerleri, Amerikan dolar kuru, gecelik faiz oranı, Japonya Borsası, Brezilya Borsası, İngiltere Borsası, Fransa Borsası, Almanya Borsası değerleri ele alınmıştır. Araştırmada yöntem olarak ileri beslemeli yapay sinir ağları ve destek vektör makineleri olarak sigmoid bir kernel fonksiyonu kullanılmıştır. Tahmin sonuçlarının değerlendirilmesinde R^2 , MPE, MSPE, RMSPE ve MAPE kriterleri kullanılmıştır. Analiz sonucunda yapay sinir ağları ve destek vektör makineleri yönteminin borsa endeksi tahmininde başarılı sonuçlar elde ettiği bulunmuştur (Yakut vd., 2014:139-157).

Çalışkan ve Deniz (2015) araştırmalarında 25.11.2014-01.12.2014 tarihleri arasında BİST30 endeksinde yer alan 27 şirketin değerlerini yapay sinir ağları ile tahmin etmişlerdir. Araştırmada makroekonomik ve yabancı borsa hisse senetleri değişkenleri olmak üzere 32 değişken girdi değişkeni olarak kullanılmıştır. Çıktı değişkeni olarak ise, hisse senetlerinin bir sonraki gün fiyat tahmini alınmıştır. Araştırma sonucunda 27 hisse senedi için, seçilen performans kriterlerinden ortalama mutlak hata 21 kuruş, ortalama mutlak yüzde hata ise %1,80 olarak bulunurken hisse senetleri fiyatlarının

artışı veya azalışı ile ilgili yapılan tahminde de ortalama başarı oranı %58 olarak bulunmuştur(Çalışkan,Deniz, 2015:177-194).

Filiz vd. (2017) araştırmalarında makine öğrenmesi algoritmalarından k en yakın komşu algoritması, Naive Bayes sınıflandırıcısı, C4.5 sınıflandırma algoritması ve yapay sinir ağları yöntemi kullanarak en iyi sınıflandırma yapan algoritmayı tespit etmeye çalışmışlardır. Araştırmada BİST-50 endeksi ve BİST-50 endeksini etkileyen DAX, FTSE, S&P 500, BISTBANK, BISTMALI, BISTSINAI, GOLDINDEX, Euro/Dolar Paritesi, Ham Petrol Fiyatı ve Faiz oranları değişkenleri kullanılmıştır. Analiz sonucunda en iyi performansı gösteren sınıflandırma algoritmasının C4.5 algoritması olduğu görülmüştür(Filiz, vd.,2017:231-241).

Özer ve diğ. (2018) araştırmalarında 2012-2016 yıllarına ait gelişmekte olan ülkelerin haftalık kapanış endeks değerlerini ele alarak bulanık mantık ve yapay sinir ağları ile tahmin gerçekleştirmişlerdir. Analizde iki etkin model oluşturarak tahmin performansları karşılaştırılmıştır. Verilerin %60'ı eğitim %40'ı ise test amaçlı kullanılmıştır. Sonuç olarak bulanık mantık ve yapay sinir ağları yönteminin tahmin sonuçlarının birbirine yakın olduğu saptanmıştır (Özer vd., 2018:100-123).

Sarıkaya (2019) araştırmasında 2005- 2012 yılları arasında BİST-100 endeksinin kapanış fiyatlarını kullanarak yapay sinir ağları ve regresyon yöntemi ile 2013 yılı BİST Ulusal-100 endeksi kapanış değerlerini tahmin etmiştir. Tahmin sonuçları 2013 yılı gerçekleşen BİST Ulusal-100 endeks kapanış fiyatları ile karşılaştırılmıştır. Yapay sinir ağları modeli kübik regresyon modeline göre daha iyi performans gösterdiği sonucuna ulaşılmıştır(Sarıkaya, 2019:325-340).

Çınaroğlu ve Avcı (2020) araştırmalarında 2015-2018 yılları arasında BİST-100 ve BİST Ulaştırma Endeksinde işlem gören Türk Hava Yollarına ait olan hisse senedi fiyatlarını yapay sinir ağlarını kullanarak tahmin etmişlerdir. Araştırmada girdi değişkeni olarak BIST100 endeksi, BIST Ulaştırma Endeksi, dolar kuru ve petrol fiyatları tanımlanırken çıktı değişkeni olarak THY hisse senedi değeri tanımlanmıştır. Araştırma sonucunda en iyi MSE kriteri eğitim verileri için 0.0031 test verileri için 0.0030 olarak bulunmuştur. Ardından model 10 günlük yeni test verisi üzerinde

çalıştırılarak gerçekleşen değer ile tahmin edilen değer arasında 3.36 kuruş tutarında bir sapmanın olduğu bulgusuna ulaşılmıştır (Çınaroğlu, Avcı, 2020:1-20).

Ünsal (2020) araştırmasında 2019 yılına ait BİST endeksi kapanış verilerini kullanarak endekste yer alan şirketlerin birbirleri ile ilişkilerini ortaya çıkarmayı amaçlamıştır. Bu amaç doğrultusunda günlük olarak elde edilen verileri veri madenciliği yöntemlerinden biri olan k-means kümeleme tekniği ile kümelere ayırarak ardından oluşturulan kümelere apriori algoritması uygulayarak çeşitli birliktelik kuralları oluşturmuştur. Araştırma sonucunda 249 işlem günü için 4 adet küme oluşturulmuş, oluşturulan bu kümeler ile güvenilirlik oranı en yüksek 10 birliktelik kuralı üretilmiştir. Böylece günlük fiyat değişimlerini ele alarak birbiriyle etkileşimli bir başka ifade ile en çok hareket eden hisse senetlerine ulaşılmıştır (Ünsal, 2020:106-112).

Karakul (2020) araştırmasında 4 Ocak 2010 - 7 Ocak 2020 tarihleri arasında 2511 iş gününe ait verileri kullanarak BIST-100 endeksini etkileyen değişkenler arasındaki ilişkileri yapay sinir ağları yöntemi ile modellemiştir. Araştırmada veri seti %90'ı eğitim, %10'u test olarak ayrılmıştır. Kurulan modeller içinde en iyi performansı gösteren model seçilerek yapay sinir ağlarının tahmin gücünün finans alanındaki başarısı saptanmıştır (Karakul, 2020:497-509).

Kantar (2020) tarafından yapılan araştırmada 2009-2019 yılları arasındaki BIST 100 aylık verileri ve altın, döviz kuru sepeti, mevduat faizi, emisyon, doğrudan sermaye yatırımları, portföy yatırımları ve sanayi üretim endeksi verileri kullanılmıştır. Araştırmada BIST-100 endeksinin makroekonomik değişkenler aracılığıyla yapay sinir ağları ve ARMA modeli ile tahmini gerçekleştirilmiştir. Analizde öncelikle veriler durağanlaştırarak sonrasında ARMA modeli, çıktı değişkeni olarak tanımlanan BIST-100 endeksi ve girdi değişkeni olarak tanımlanan makroekonomik değişkenler ileri beslemeli iki katmanlı yapay sinir ağı yöntemi ile tahmin edilmiştir. Tahmin sonuçlarının değerlendirilmesinde kriter olarak en yüksek R^2 ve en düşük MSE değeri baz alınmıştır. Sonuç olarak ARMA (1,1) modelinde R^2 kriteri 0.95, MSE kriteri 0.00452, yapay sinir ağı modelinde R^2 değeri 0.99, MSE kriteri ise, 0.001316 şeklinde

bulunarak yapay sinir ağıları modelinin daha iyi bir tahmin yöntemi olduğu sonucuna varılmıştır (Kantar, 2020:121-131).

Kocabıyık ve diğ. (2021) araştırmalarında BİST’ de işlem gören şirketlerin yer aldığı endekslerin birlikte hareketleri incelenmiştir. Araştırmada iki ayrı veri seti kullanılmıştır. İlk olarak 2014-2021 yılları arasında 30 farklı temel ve sektörel endekse ait verilerin birlikteliği tespit edilirken ikinci olarak 5 farklı sektör endekslerinin birbirleri ile hareketliliği tespit edilmiştir. Sonuç olarak ilk veri setinde yer alan temel ve sektörel endeksler arasında 703 birliktelik kuralları tespit edilmiş ve %99 güven düzeyi göz önüne alındığında ise 101 birliktelik kuralı bulunmuştur. İkinci olarak sektörel endekslerin yer aldığı veri setinde ise 45 birliktelik kuralı tespit edilmiştir (Kocabıyık, Dağ, Karaatlı, 2021:659-672).

Karaatlı ve diğ.(2021) araştırmalarında piyasanın yönünü ve alternatif yatırım araçlarını tespit etmek amacıyla BİST-30da işlem gören şirketlerin makroekonomik değişkenler ile ilişkisini ortaya koymayı amaçlamışlardır. 2014-2019 yılları arasını kapsayan bu araştırmada birliktelik kuralı analizinden FP-Growth algoritması kullanılmıştır. Birinci grup veri setinde BİST-30 da yer alan tüm paylara ve makroekonomik değişkenlere yer verilerek 81 birliktelik kuralı bulunmuştur. Bulunan bu kuralların 77’si BIST-30 endeksinde yer alan bankacılık sektörüne ait hisselerini içermektedir. Ardından bankacılık endeksi analiz dışı bırakılarak ikinci bir veri grubu ile analiz yapılmış olup 76 birliktelik kuralı bulunmuştur. Bulunan bu kuralların 71’i altın değişkenini içermektedir. Sonuç olarak bankacılık endeksi analiz dışında tutulduğunda aynı sektörde yer alan payların ortak hareketi söz konusu olduğu ve makroekonomik değişkenlerden altın değişkeninin BİST-30 endeksi ile büyük ölçüde birlikte değişim gösterdiği bulunmuştur (Karaatlı vd., 2021:548-571).

İnce ve Alan (2022) yatırımcıların karar verme süreçlerinde etkili olabilecek bir araştırma ortaya koymuşlardır. Araştırmada Türkiye’de faaliyet gösteren 102 holding seçilerek yatırım alanları sektöre göre sınıflandırılmıştır. Ardından veri madenciliği tekniklerinden birliktelik kuralı analizinde yer alan apriori algoritması kullanılarak %50

güven seviyesinde 35 adet kural üretilmiştir. Üretilen kurallar doğrultusunda yatırımcı şirketlere yatırım alanları hakkında bilgiler sunulmuştur(İnce, Alan, 2022:1-15).

3.1.2.Uluslararası Literatür

Lee ve Chiu (2002) araştırmalarında SGX-DT Nikkei vadeli fiyatlarını nakit dış ticaret döneminde yapay sinir ağları yöntemini kullanarak tahmin etmeye çalışmışlardır. Yapay sinir ağı modelinde geri yayılım algoritması seçilmiştir. Analiz sonucunda kurulan modelin iyi performans gösterdiği kanısına varılmıştır (Lee ,Chiu:229-237).

Chen vd.(2003) araştırmalarında yapay sinir ağları yöntemini kullanarak Tayvan hisse senedinin piyasa getiri endeksinin yönünü modellemeye çalışmışlardır. Araştırmada veri seti olarak Ocak 1982 ile Ağustos 1987 dönemi seçilmiştir. Analiz sonucunda yapay sinir ağları yönteminin güçlü bir tahminci olduğu saptanmıştır (Chen,vd.,2003:901-923).

Ting ve diğ. (2006) araştırmalarında altı aylık günlük stok zaman serilerini kullanarak, stok içi ve stoklar arası birliktelik kurallarını bulmayı amaçlamışlardır. Verilerin %70'i eğitim amaçlı %30'u ise test için ayrılmıştır. Eğitim verileri birliktelik kurallarını bulmak ve sonrasında stok fiyat hareketlerini sınıflandırmak için kullanılmıştır. Stok içi ve stoklar arası ilişkisel sınıflandırma sonuçlarına göre destek sayısının oluşturulması gerektiğine varılmıştır (Ting,vd.,2006:29-36).

Lee vd. (2007) araştırmalarında Kore Bileşik Hisse Senedi Fiyat Endeksini(KOSPI) geri yayımlı sinir ağı (BPNN) modeli ve zaman serisi mevsimsel otomatik gerileyen entegre hareketli ortalama (SARIMA) modeli ile tahmin ederek analiz sonuçlarını karşılaştırmayı amaçlamışlardır. Analiz sonucunda orta vadeli tahminlerde SARIMA modelinin genellikle KOSPI değeri için BPNN modelinden daha doğru tahminler ürettiği ve BPNN modelinin genellikle KOSPI getirilerini tahmin eden SARIMA modelinden daha iyi performans gösterdiği bulunmuştur. Ancak kurulan iki modelin tahmin doğruluklarındaki fark istatistiksel olarak anlamlı bulunmamıştır (Lee,vd.,2007:372-378).

Bahrammirzaee (2010) araştırmasında finans uygulamalarında geleneksel yöntemlerin aksine yapay zekâ tekniklerine olan ilginin artmasına değinmiştir. Bu amaçla yapay sinir ağıları, uzman sistemler ve hibrit zeka sistemlerinin finansal piyasalarda kullanımını araştırarak kredi değerlendirme, portföy yönetim, finansal tahmin ve planlama olarak üç bölümde kategorilere ayırmıştır. Sonuç olarak finansal problemlerde doğrusal olmayan modeller açısından yapay zekâ yöntemlerinin geleneksel istatistiksel yöntemlere göre daha üstün olduğu bulunmuştur (Bahrammirzaee, 2010:1165-1195).

Na ve Sohn (2011) araştırmalarında Kore Bileşik Hisse Senedi Fiyat Endeksinde (KOSPI) meydana gelen değişimleri tahmin etmeyi amaçlamışlardır. Bu amaçla çeşitli dünya borsa endekslerini de ele alarak birliktelik kuralı analizi gerçekleştirerek çıkarımlar da bulunmuşlardır. Birliktelik kuralı analizi için minimum destek seviyesi %5, minimum güven seviyesi ise %70 olarak seçilmiştir. Araştırma sonucunda KOSPI endeksi küresel borsa endekslerinden ABD ve Avrupa'daki endeksler ile aynı yönde hareket ederken, Doğu Asya ülkelerinde ise ters yönde hareket ettiği bulunmuştur (Na, Sohn, 2011:9046–9049).

Argiddi ve Apte (2012) araştırmalarında birliktelik kuralı analizini kullanarak, altı adet şirket payı verilerini First Intra Then Inter (FITI) algoritmasıyla Hindistan Borsasında gelecek tahminini yapmak adına kullanmışlardır (Argiddi ,Apte, 2012:30-34).

Liao ve Chou (2013) çalışmalarında Tayvan ve Çin (Hong Kong) arasında 29 Haziran 2010'da imzalanan Ekonomik İşbirliği Çerçeve Anlaşması (ECFA) kapsamında Tayvan ve Çin borsalarının ortak hareketlerinin incelenmesini amaçlamışlardır. Çalışmada veri madenciliği tekniklerinden birliktelik kuralı analizi ve kümeleme analizi kullanılmıştır. Anlaşma kapsamında ele alınan 30 adet endeks arasında sektörel olarak birlikte değişimler tespit edilmiştir (Liao,Chou, 2013:1542-1554).

Arafah ve Mukhlash (2015) arařtırmalarında Endonezya borsasında iřlem g ren hisse senetleri arasında birliktelik kurallarını ortaya koymayı amalamıřlardır. Analizde 5 yıllık veri seti ile alıřılarak Bulanık Apriori algoritması kullanılmıřtır. Analiz sonucunda Endonezya borsasında iřlem g ren hisse senetlerinin birlikte hareketi tespit edilmiřtir (Arafah,Mukhlash, 2015:235-243).

Jalpa ve Rustom (2017) arařtırmalarında Hindistan borsasındaki farklı hisse senetleri arasındaki ortak hareketleri tespit etmek ve yatırımcılara fayda saėlamayı amalamıřlardır. Analizde birliktelik kuralı algoritmalarından Apriori ve FP-Growth algoritmasını kullanmıřlardır. Analiz sonucunda FP-Growth algoritmasının daha iyi sonulara ulařtıėı kanısına varılmıřtır (Jalpa,Rustom, 2017:6-15).

Islamiyah vd., (2019) arařtırmalarında birliktelik kuralı algoritmalarından Apriori ve FP-Growth algoritmasını kullanarak t keticinin satın alma davranıřını belirlemeyi amalamıřlardır. Analiz sonucunda Apriori ve FP-Growth algoritmasını karřılařtırarak FP-Growth algoritmasının daha hızlı sonulara ulařtıėını ortaya koymuřlardır (Islamiyah, vd., 2019:320-323).

Sahoo ve Mohanty (2020) arařtırmalarında 25 Aėustos 2004 ile 24 Ekim 2018 tarihleri arasındaki Bombay borsası verilerini YSA-GWO ve hibrit YSA-GWO y ntemi ile tahmin etmiřlerdir. Arařtırmada Tahmin edilen modelin performansını belirlemek adına ortalama kare hata, ortalama karesel hatanın k k , ortalama mutlak hata ve medyan ortalama hata kriterleri kullanılmıřtır. Analiz sonunda, hibrit YSA-GWO y nteminin daha iyi sonular verdiėi saptanmıřtır (Sahoo,Mohanty, 2020:77-87).

3.2. Arařtırmanın Amacı

 lkemizin ekonomik durumu ile doėrudan iliřkili olan Borsa İstanbul Endeksi, yatırımcılar iin  nem arz etmektedir. Dolayısıyla endeks verilerinin tahminin gerekleřtirilmesi ve endeks ile birlikte hareket eden deėiřkenlerin tespit edilmesi yatırımcılara avantaj saėlayacaktır.

Bu çalışmanın amacı veri madenciliği tekniklerinden, yapay sinir ağları analizi ile BİST 30 endeks değerlerini tahmin etmek ve birliktelik kuralı analizi kullanılarak BİST30 endeksinin birlikte hareket ettiği makroekonomik değişkenleri tespit etmektir. Araştırma kapsamında 2022 Aralık ayına ait BİST30 endeksinin günlük değerlerinin tahminini yapılmıştır. Tahmin için 2022 Aralık ayının seçilmesi ele alınan veri setinin dönemleri arasındaki en yakın tarih olmasıdır. Ardından tahmin edilen değerler ile gerçekleşen değerler karşılaştırılarak yapay sinir ağlarının performansı gösterilmiştir. Ayrıca yapılan tahminler de girdi değişkenleri olarak kullanılan makroekonomik değişkenler ile BİST30 endeksinin ortak hareketi birliktelik kuralı analizi ile incelenmiştir

3.3. Araştırmada Kullanılan Veri Seti

Araştırmada Ocak 2018 ve Aralık 2022 tarihleri arasında gerçekleşen 1253 adet günlük veri seti ile çalışılmıştır. Araştırma kapsamında veri seti olarak Borsa İstanbul 30 Endeksi, Altın Ons Fiyatı, Dolar Kuru, Euro Kuru, Brent Petrol, Tüketici Fiyat Endeksi, Para Arzı, Mevduat Faizi, Finansal Hizmetler Güven Endeksi ele alınmıştır. Ayrıca aylık olarak yayınlanan veriler ilgili ayın günlük verisi olarak değerlendirilmiştir. Aşağıdaki Tablo 3.1 'de araştırmada kullanılan veri seti ve kaynaklarına yer verilmiştir:

Tablo 3.1:Araştırmada Kullanılan Değişkenler

Sıra	Değişkenler	Kısaltması	Kaynak
1	Borsa İstanbul 30 Endeksi	BIST30	investing.com
2	Tüketici Fiyat Endeksi	TUFE	TCMB
3	Para Arzı	M1	TCMB
4	Altın Ons Fiyatı	ALTIN	investing.com
5	Dolar Kuru	DOLAR	investing.com
6	Euro Kuru	EURO	investing.com
7	Mevduat Faizi	FAIZ	TCMB
8	Finansal Hizmetler Güven Endeksi	FHGE	TCMB
9	Brent Petrol	BRENT	investing.com

3.4. Araştırmada Kullanılan Makroekonomik Değişkenler

Bu bölümde araştırmada kullanılan makroekonomik değişkenlere ait tanımlamalara yer verilmiştir.

- ✓ **Tüketici Fiyat Endeksi:** Bir ülkenin ekonomik yapısını önemli düzeyde etkileyen enflasyon, fiyatlar genel seviyesindeki sürekli artışı göstermektedir. Enflasyona bağlı olarak üreticinin karında azalma meydana gelirken buna bağlı olarak tüketicinin de refah düzeyinde azalma meydana gelmektedir (Akgün, 2015:168-168). Tüketiciler tarafından satın alınan mal ve hizmetlerin fiyatında meydana gelen değişimleri gösteren Tüketici Fiyat Endeksi enflasyon denilince akıllara gelen endekstir (Tunalı,Özkan, 2016:55).Bu çalışmada da enflasyon değerini temsilen Tüketici Fiyat Endeksi seçilmiştir.
- ✓ **M1 Para Arzı:**M1 para arzı, dolaşımdaki para ile vadesiz mevduatların toplamını göstermektedir. Vadesiz mevduatlar ödemelerde doğrudan çek aracılığıyla kullanılmaktadır. Dolayısıyla tıpkı nakit para gibidir (Aktop, 2010:31).
- ✓ **Altın Ons Fiyatı:** Küresel anlamda önemli bir yere sahip olan altın değerli bir madendir. Altının hem yatırım aracı olarak kullanılabilmesi hem de endüstriyel olarak kullanılabilmesi altına olan talebi arttırmaktadır (Aygün Alıcı,Köseoğlu, 2021:257).Ons ağırlık birimini ifade etmekle birlikte ons altın dünya piyasaları tarafından kabul görmüş altının değerini dolar bazında gösteren bir birimdir.
- ✓ **Dolar Kuru:** Amerikan dolarının Türk lirası karşısındaki değerini göstermektedir.
- ✓ **Euro Kuru:**Euoru'nun Türk lirası karşısındaki değerini göstermektedir.
- ✓ **Mevduat Faizi:** Bankalardaki Türk lirası mevduata verilen faiz oranını göstermektedir (Güler, 2016:397).
- ✓ **Finansal Hizmetler Güven Endeksi:** Finansal hizmetler anket sorularına verilen yanıtlardan yola çıkarak üretilen Finansal Hizmetler Güven Endeksi, TCMB tarafından 2012 yılı Mayıs ayı itibariyle yayınlanmaktadır. Anket soruları finansal yöneticilerin geçmiş değerlendirmeleri ve gelecek beklentileri üzerine hazırlanmakta ve böylece katılımcıların ankete verdiği yanıtlar temel alınarak finansal hizmet sektörünün gelişmesine katkı sağlayacak göstergelerin belirlenmesi amaçlanmaktadır (Süsay vd.,2020: 85).

- ✓ **Brent Petrol:** Ekonomik faaliyetleri etkilemede önemli rol oynayan petrol fiyatları hem petrol ihracatı yapan ülkeler için hem de petrol ithal eden ülkeler için önem arz etmektedir (Timur,Günay,2019:5205).Piyasalarda birçok petrol türü bulunmaktadır. Ancak kuzey denizinden çıkarılan brent petrol, içerisinde barındırdığı düşük sülfür miktarından dolayı ilgi görmektedir. Ayrıca denizden çıkarılması nakliye maliyetlerini minimum seviyeye indirirken, brent petrolden benzin ve motorinin üretilmesi önemli avantaj sağlamaktadır (Kurt, 2021:41).

3.5.Araştırmada Kullanılan Yöntemler

Araştırmada yapay sinir ağları analizi ve birliktelik kuralı analizi kullanılmıştır. Yapay sinir ağı analizi için Matlab(R2021b) programı kullanılırken birliktelik kuralı analizi için Weka 3.8.5 programı kullanılmıştır. Analizlere başlamadan önce gerekli düzenlemeler Ms Excel aracılığıyla gerçekleştirilmiştir.

3.6.Yapay Sinir Ağı Analizi

Araştırmada BİST30 endeks verileri yapay sinir ağları ile tahmin edilmeye çalışılmıştır. Analizde Matlab (R2021b) programı kullanılarak girdi değişkeni olarak 8 makroekonomik değişken seçilirken çıktı değişkeni olarak da BİST30 endeks verileri seçilmiştir. Aşağıdaki tablo 3.2’de verilere ilişkin bilgiler yer almaktadır:

Tablo 3.2:Yapay Sinir Ağı Analizinde Kullanılan Değişkenler

Girdi Değişkenleri	Çıktı Değişkeni
Tüketici Fiyat Endeksi	Borsa İstanbul30 Endeksi
Para Arzı	
Altın Ons Fiyatı	
Dolar Kuru	
Euro Kuru	
Mevduat Faizi	
Finansal Hizmetler Güven Endeksi	
Brent Petrol	

Yapay sinir ağı analizine başlamadan önce veri setimiz 0-1 aralığında normalleştirme işlemine tabii tutulmuştur. Normalleştirme işlemi sonrasında veriler Matlab programına tanıtılarak veri setinin %70'i eğitim, %15'i doğrulama,%15'i ise test için kullanılmıştır. Analizde denetimli öğrenme türü kullanılarak iterasyon sayısı 1000 olarak seçilmiştir. Ayrıca analizde geri yayılım gösteren Levenberg-Marquard (LM) algoritması, kullanılırken performans kriteri olarak MSE (Mean Squared Error) ve R^2 (çoklu determinasyon katsayısı)kullanılmıştır.

MSE kriteri istenilen çıktı değerleri ile ağıın ürettiği çıktı değerleri arasındaki farkın kareleri toplamının ortalamasından oluşmakta olup, y_i istenilen çıktı değerlerini, $\hat{y}_i$ ağıın ürettiği çıktı değerlerini, n veri sayısını göstermek üzere aşağıdaki gibi hesaplanmaktadır (Arıkan Kargı, 2015:79):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.1)$$

R^2 değeri çoklu determinasyon katsayısını göstermek üzere 0 ile 1 arasında değer almaktadır ve aşağıdaki gibi hesaplanmaktadır (Arıkan Kargı, 2015:18):

$$R^2 = \hat{\beta}_1 \sum yx_1 + \hat{\beta}_2 \sum yx_2 + \dots \hat{\beta}_n \sum yx_n / \sum y^2 \quad (3.2)$$

R^2 değerinin 0'a yaklaşması bağımsız değişkenlerin bağımlı değişken üzerindeki değişmeyi yeterince açıklayamadığını gösterirken, 1'e yaklaşması ise bağımsız değişkenin bağımlı değişken üzerindeki değişmeyi iyi derecede açıkladığını göstermektedir ve dolayısıyla R^2 değerinin 1' e yaklaşması istenmektedir (Arıkan Kargı, 2015:18).

Yapay sinir ağları analizinde en iyi sonuca ulaşmak adına deneme-yanılma yolu ile program birçok defa çalıştırılmıştır. Aşağıda Tablo 3.3'de bu denemelerden bazılarına yer verilmiştir:

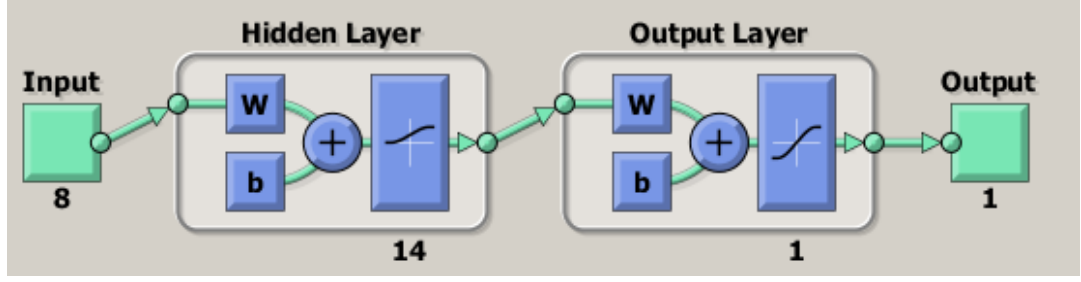
Tablo 3.3: En İyi Performansa Sahip Ağ İçin Yapılan Denemeler

Ağ Yapısı	Gizli Katman	Eğitim Fonksiyonu	Öğrenme Fonksiyonu	Transfer Fonksiyonu	MSE	R	R ²
8-5-1	1	TRAİNLM	LEARNGDM	purelin-logsig	0,14679	0,78	0,60
8-7-1	1	TRAİNGDA	LEARNGDM	tansig-purelin	0,0012396	0,98	0,96
8-2-4-1	2	TRAİNLM	LEARNGDM	tansig-purelin-logsig	0,15565	0,79	0,62
8-10-1	1	TRAİNLM	LEARNGDM	logsig-purelin	0,00017027	0,99	0,98
8-17-1	1	TRAİNGDM	LEARNGDM	purelin-tansig	0,0022677	0,96	0,92
8-5-8-1	2	TRAİNGDA	LEARNGDM	tansig-tansig-purelin	0,00062542	0,98	0,96
8-10-10-1	2	TRAİNGDA	LEARNGDM	tansig-logsig-purelin	0,00054473	0,99	0,98
8-14-1	1	TRAİNLM	LEARNGDM	logsig-tansig	0,00013557	0,99	0,98
8-8-10-1	2	TRAİNGDX	LEARNGDM	logsig-tansig-purelin	0,00057777	0,99	0,98

Yapılan denemelerde en iyi modeli bulabilmek için performans kriteri olarak kullanılan MSE (ortalama kare hata) değerinin en küçük ve R² (determinasyon katsayısının) en büyük olduğu ağ yapısı seçilmiştir. Böylece MSE= 0,00013557 olan 8-14-1 ağ yapısına sahip MLP (çok katmanlı algılayıcı) modeli en iyi model olarak seçilmiştir. Ayrıca MSE değerinin en küçük olduğu en iyi modelde R²=0,98 olarak bulunmuştur.

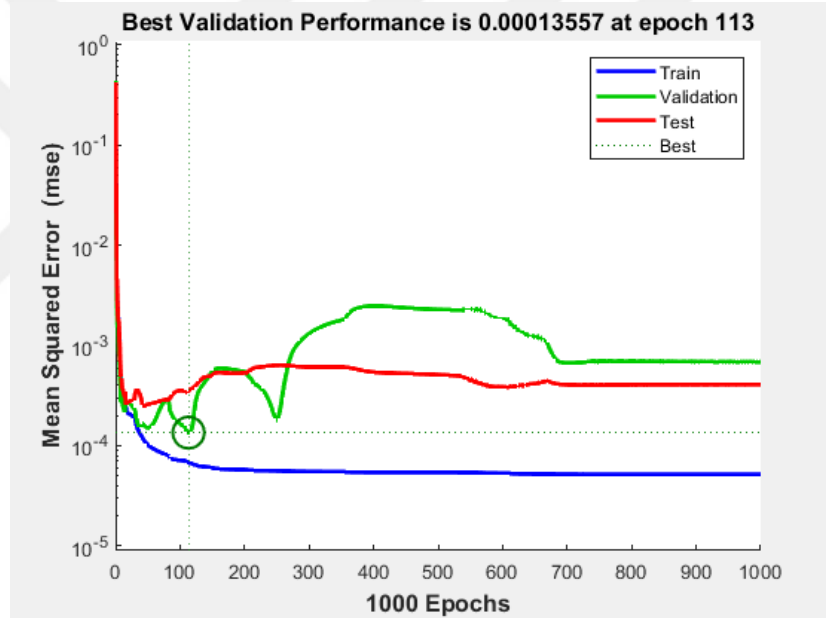
En iyi model olarak seçilen ağ tek gizli katmana sahip olmakla birlikte 14 nöronlu oluşmaktadır. Modelde girdi katmanı ve gizli katman arasında transfer fonksiyonu olarak logaritmik sigmoid (logsig) fonksiyonu kullanılırken, gizli katman ve çıkış katmanı arasında hiperbolik tanjant sigmoid (tansig) kullanılmıştır.

Aşağıda en iyi performansı sergileyen yapay sinir ağının mimari yapısı gösterilmiştir:



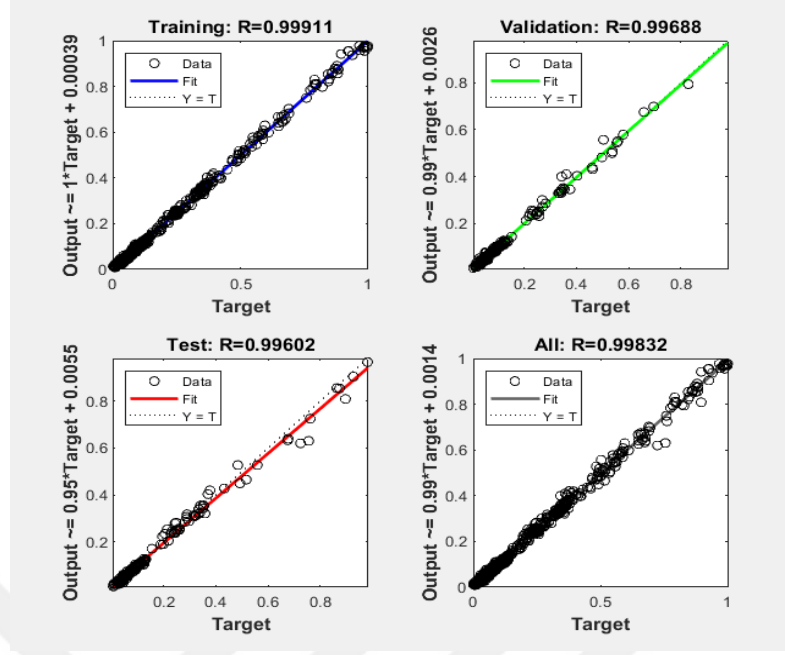
Şekil 3.1:En İyi Performansa Sahip Olan Ağın Mimarisi

Aşağıda Şekil 2.7’de eğitim, doğrulama ve test kümelerine ilişkin MSE (ortalama kare hata) değerlerinin görüntüsü verilmiştir. Buna göre ağ 1000.iterasyonda optimum sonuca ulaşmıştır.



Şekil 3.2:Eğitim Doğrulama ve Test Verileri İçin MSE(Ortalama Kare Hata)

En iyi model için R değerleri çıktısı aşağıda Şekil 3.3 de verilmiştir. Gerçek değerler ile tahmin edilen değerleri gösteren bu çıktıda kurulan modelin tahmin gücünün iyi seviyede olduğu görülmektedir.



Şekil 3.3: Eğitim Sonucunda Ulaşılan Tahmin Başarıları

Aşağıda Tablo 3.4’de 2022 yılı Aralık ayı günlük değerleri ile yapay sinir ağı analizi sonucunda elde edilen tahmin değerleri yer almaktadır. Sonuçlara bakıldığında tahmin değerlerinin gerçek değerlere oldukça yakın olduğu görülmektedir. Dolayısıyla yapay sinir ağlarının tahmin gücünün yüksek olduğu söylenebilir.

Tablo 3.4:Yapay Sinir Ağı Analizi Tahmin Sonuçları

Tarih	2022 Aralık Gerçekleşen	YSA Tahmin
01.12.2022	5.433,58	5.004,50
02.12.2022	5.313,58	5.348,93
05.12.2022	5.318,85	5.349,68
06.12.2022	5.366,43	5.336,35
07.12.2022	5.180,48	5.224,23
08.12.2022	5.200,78	5.121,33
09.12.2022	5.355,84	5.406,48
12.12.2022	5.582,63	5.471,01
13.12.2022	5.648,80	5.634,52
14.12.2022	5.434,83	5.659,52
15.12.2022	5.570,59	5.718,56

16.12.2022	5.582,67	5.724,42
19.12.2022	5.804,74	5.732,97
20.12.2022	5.849,77	5.773,76
21.12.2022	5.860,95	5.793,26
22.12.2022	5.883,16	5.724,14
23.12.2022	5.885,22	5.808,28
26.12.2022	5.940,73	5.797,11
27.12.2022	5.876,73	5.815,77
28.12.2022	5.789,35	5.837,30
29.12.2022	5.915,15	5.819,02
30.12.2022	5.944,83	5.834,05

3.7.Birliktelik Kuralı Analizi

Araştırmada yapay sinir ağları analizinde girdi değerleri olarak kullanılan makroekonomik değişkenlerin BİST30 endeksi ile ortak hareketi tespit edilmeye çalışılmıştır. Analiz 1253 adet veri seti ele alınarak, veri madenciliği yöntemlerini uygulamada kullanılan Weka 3.8.5 paket programında gerçekleştirilmiştir.

Birliktelik kuralı analizine başlamadan önce veri seti bir önceki günün kapanış fiyatları göz önünde bulundurularak, ilgili değerın bir önceki güne göre artış göstermesi durumunda 1 azalış göstermesi durumunda ise 0 değeri verilerek kategorik hale dönüştürülmüştür. Bu düzenlemeler MS Excel de gerçekleştirilmiştir. Veri setine ait düzenlemeleri gösteren bir kesit aşağıda Şekil 3.1’de yer almaktadır:

Şekil 3.4: Birliktelik Kuralı Analizine Ait Veri Setinin Düzenlemeleri

	A	B	C	D	E	F	G	H	I
1	TUFE	BRENT	ALTIN	FHGE	M1	FAİZ	DOLAR	EURO	BIST30
2	1	0	1	1	0	0	0	0	1
3	1	1	1	1	0	0	0	1	0
4	1	1	1	1	0	0	0	0	0
5	1	0	1	1	0	0	0	1	1
6	1	1	0	1	0	0	0	0	0
7	1	1	0	1	0	0	0	0	0
8	1	1	1	1	0	0	1	0	0
9	1	1	1	1	0	0	1	1	1
10	1	1	1	1	0	0	1	1	0
11	1	1	1	1	0	0	0	1	0
12	1	0	1	1	0	0	0	1	1
13	1	1	0	1	0	0	1	1	1
14	1	0	0	1	0	0	0	0	1
15	1	0	1	1	0	0	0	0	0
16	1	1	0	1	0	0	0	0	1
17	1	1	1	1	0	0	1	1	1

Weka 3.8.5 programına düzenlenen verilerin tanıtılmasının ardından birliktelik kuralı analizi için güven aralığı %95 ve FP-Growth algoritması seçilerek kurallar güven değerine göre sıralanmıştır. Analiz sonucunda 38 kural üretilmiştir. Aşağıda Tablo 3’de bu kurallar yer almaktadır.

Tablo 3.5: Birliktelik Kuralı Analizi Sonuçları

	Birliktelik Kuralları	Conf	Lift.	Lev.	Conv.
1	FAİZ=1: 671 ==> TUF=1: 671	1	1.04	0.02	28.16
2	M1=1,FAİZ=1: 484 ==>TUF=1: 484	1	1.04	0.02	20.31
3	M1=1,FHGE=1: 474 ==>TUF=1: 474	1	1.04	0.02	19.89
4	DOLAR=1,FAİZ=1: 392 ==>TUF=1: 392	1	1.04	0.01	16.45
5	BRENT=1, FAİZ=1: 363 ==> TUF=1: 363	1	1.04	0.01	15.23
6	BIST30=1, FAİZ=1: 362 ==> TUF=1: 362	1	1.04	0.01	15.19
7	EURO=1, FAİZ=1: 357 ==> TUF=1: 357	1	1.04	0.01	14.98
8	FAİZ=1, ALTIN=1: 347 ==>TUF=1: 347	1	1.04	0.01	14.56
9	DOLAR=1, BRENT=1, EURO=1]: 325 ==>TUF=1: 321	0.99	1.03	0.01	2.73
10	M1=1, DOLAR=1, BIST30=1: 324 ==> TUF=1: 320	0.99	1.03	0.01	2.72
11	M1=1, BIST30=1: 520 ==> TUF=1: 512	0.98	1.03	0.01	2.42
12	M1=1, DOLAR=1, BRENT=1: 325 ==> TUF=1: 320	0.98	1.03	0.01	2.27
13	BIST30=1, EURO=1: 384 ==> TUF=1: 378	0.98	1.03	0.01	2.3
14	BRENT=1, EURO=1: 381 ==> TUF=1: 375	0.98	1.03	0.01	2.28
15	M1=1, DOLAR=1, EURO=1: 454 ==> TUF=1: 446	0.98	1.03	0.01	2.12
16	M1=1, EURO=1: 551 ==>TUF=1: 541	0.98	1.02	0.01	2.1
17	M1=1, DOLAR=1: 601 ==> TUF=1: 590	0.98	1.02	0.01	2.1
18	M1=1, BRENT=1: 523 ==> TUF=1: 513	0.98	1.02	0.01	2
19	DOLAR=1, BRENT=1: 401 ==> TUF=1: 393	0.98	1.02	0.01	1.87
20	DOLAR=1, BIST30=1: 397 ==> TUF=1: 389	0.98	1.02	0.01	1.85
21	M1=1: 959 ==> TUF=1: 938	0.98	1.02	0.02	1.83
22	DOLAR=1, EURO=1: 564 ==> TUF=1: 551	0.98	1.02	0.01	1.69
23	DOLAR=1, FHGE=1: 330 ==> TUF=1: 322	0.98	1.02	0	1.54
24	BRENT=1, FHGE=1: 330 ==>TUF=1	0.98	1.02	0	1.54
25	EURO=1: 679 ==> TUF=1: 662	0.97	1.02	0.01	1.58
26	EURO=1, ALTIN=1: 358 ==> TUF=1: 349	0.97	1.02	0	1.5
27	BRENT=1: 694 ==> TUF=1: 676	0.97	1.02	0.01	1.53

28	DOLAR=1: 726 ==>TUF=1: 707	0.97	1.02	0.01	1.52
29	M1=1, ALTIN=1: 499 ==> TUF=1: 485	0.97	1.01	0.01	1.4
30	ALTIN=1, FHGE=1: 331 ==> TUF=1: 321	0.97	1.01	0	1.26
31	BRENT=1, BIST30=1: 396 ==> TUF=1:384	0.97	1.01	0	1.28
32	BIST30=1: 692 ==> TUF=1: 670	0.97	1.01	0.01	1.26
33	DOLAR=1, ALTIN=1: 366 ==>TUF=1: 354	0.97	1.01	0	1.18
34	BRENT=1, ALTIN=1: 393 ==> TUF=1: 380	0.97	1.01	0	1.18
35	BIST30=1, ALTIN=1: 388 ==> TUF=1: 374	0.96	1.01	0	1.09
36	ALTIN=1: 658 ==>TUF=1: 634	0.96	1.01	0	1.1
37	FHGE=1: 600 ==> TUF=1: 578	0.96	1.01	0	1.09
38	BIST30=1, FHGE=1: 330 ==> TUF=1: 316	0.96	1	-0	0.92

Tablo 3.5’de Conf.güven düzeyini, Lift. birlikte hareket etme katsayısını, Lev. kaldıraç etkisini ve Conv. ise ilişki düzeyini göstermektedir (Karaatlı vd., 2021:564).

Analizde makroekonomik değişkenlerin kendi aralarında ortak hareketi söz konusu olmasına rağmen üretilen birliktelik kurallarından yalnızca 9 kuralda BİST30 endeksinin makroekonomik değişkenler ile ortak hareketinin söz konusu olduğu görülmektedir. Bu kurallara ait yorumlamalara aşağıda yer verilmiştir:

Kural 6: 1253 işlem günü içerisinde BİST30 ve FAİZ değişkenleri 362 kez aynı yönde birlikte hareket etmişlerdir. TUF değişkeni de bu iki değişkenin beraber yer aldığı kümelerde 362 kez yer alarak birliktelik hareketine katılmıştır. Bu kuralın güven ölçütü 362/362 işlemi sonucunda 1 olarak hesaplanmıştır.

Kural 10: 1253 işlem günü içerisinde M1, DOLAR,BİST30 değişkenleri 324 kez aynı yönde birlikte hareket etmişlerdir. Birliktelik hareketine bu değişkenler ile birlikte 320 kez değişim gösteren TUF değişkeni de katılım göstermiştir. Bu kuralın güven ölçütü 320/324 işlemi sonucunda %99 olarak hesaplanmıştır

Kural11:1253 işlem günü içerisinde M1 değişkeni ve BİST30 değişkeni 520 kez aynı yönde birlikte hareket etmişlerdir. TUF değişkeni de bu iki değişkenin beraber yer

aldığı kümelerde 512 kere yer alarak birliktelik hareketine katılmıştır. Bu kuralın güven ölçütü 512/520 işlemi sonucunda % 98 olarak hesaplanmıştır.

Kural13:1253 işlem günü içerisinde BİST30 değişkeni ve EURO değişkeni 384 kez aynı yönde birlikte hareket etmişlerdir. TUFE değişkeni de bu iki değişkenin beraber yer aldığı kümelerde 378 kez yer almıştır. Bu kuralın güven ölçütü 378/384 işlemi sonucunda %98 olarak hesaplanmıştır

Kural20:1253 işlem günü içerisinde DOLAR değişkeni ve BİST30 değişkeni 397 kez aynı yönde birlikte hareket etmişlerdir. TUFE değişkeni de bu iki değişkenin beraber yer aldığı kümelerde 389 kez yer almıştır. Bu kuralın güven ölçütü 389/397 işlemi sonucunda %98 olarak hesaplanmıştır.

Kural31:1253 işlem günü içerisinde BRENT değişkeni ve BİST30 değişkeni 396 kez aynı yönde birlikte hareket etmişlerdir. TUFE değişkeni de bu iki değişkenin beraber yer aldığı kümelerde 384 kez yer almıştır. Bu kuralın güven ölçütü 384/396 işlemi sonucunda %97 olarak hesaplanmıştır.

Kural32: 1253 işlem günü içerisinde BİST30 değişkeni 692 kez hareket ederken TUFE değişkeni 670 kez bu harekete katılım sağlamıştır. Bu kuralın güven ölçütü 670/692 işlemi sonucunda %97 olarak hesaplanmıştır.

Kural35:1253 işlem günü içerisinde BİST30 değişkeni ve ALTIN değişkeni 388 kez aynı yönde birlikte hareket etmişlerdir. TUFE değişkeni de bu iki değişkenin beraber yer aldığı kümelerde 374 kez yer almıştır. Bu kuralın güven ölçütü 374/388 işlemi sonucunda %96 olarak hesaplanmıştır.

Kural 38: 1253 işlem günü içerisinde BİST30 değişkeni ve FHGE değişkeni 330 kez aynı yönde birlikte hareket etmişlerdir. TUFE değişkeni de bu iki değişkenin beraber yer aldığı kümelerde 316 kez yer almıştır. Bu kuralın güven ölçütü 316/330 işlemi sonucunda %96 olarak hesaplanmıştır.

Elde edilen birliktelik kurallarına göre BİST30 deęiřkeninin yapay sinir aęı analizinde girdi deęiřkeni olarak tanımlanan TUFEM1,ALTIN, DOLAR, EURO, FAİZ, FHGE, BRENT deęiřkenlerinin tamamıyla birlikte hareket ettięi görölmüřtür. Ancak oluřan kurallardan güven düzeyi en yüksek olan kural 6'dır. Kural 6'da BİST30 deęiřkeni ile FAİZ deęiřkeninin maksimum güven düzeyinde birlikte hareket ettięi ve ortak hareketlerine TUFEM1 deęiřkeninin de katıldığı görölmektedir.

Dolayısıyla BİST30 endeksini yakından takip eden yatırımcıların endeksin yukarı veya ařaęı yönde hareketinin tahmini için, endeks ile birlikte hareket eden deęiřkenlerin yol gösterici olabileceęi söylenebilir.



SONUÇ

Veri madenciliği hızla artış gösteren veri yığınlarından araştırmacının amacına uygun ve faydalı çıkarımlar yapmak için kullanılırken büyük boyutlu verilerin analizi için farklı modelleme yöntemleri sunmaktadır. Tahminleyici ve tanımlayıcı modeller olarak iki gruba ayrılan bu yöntemler araştırmacının, araştırmak istediği konuya ve kullanılacak olan verinin özelliğine, niteliğine göre seçilmektedir. Tahminleyici modellerden biri olan yapay sinir ağları yöntemi, biyolojik sinir ağı yapısını temel alarak geliştirilmiş olup geleneksel yöntemlerin aksine varsayımlara dayanmamaktadır. Bu sebeple yapay sinir ağları yönteminin kullanımı geleneksel yöntemlere göre esneklik sağlamaktadır. Tanımlayıcı modellerden biri olan birliktelik kuralı analizi ise veri yığınları içerisinde ortak hareket eden veri gruplarını tespit etmek için kullanılmaktadır.

Araştırma kapsamında ele alınan Borsa İstanbul 30 endeksi, en iyi 30 şirketi temsil ederken bu şirketler imalat, teknoloji, mali kuruluşlar, ulaştırma, madencilik gibi sektörlerde faaliyet göstermektedir. Çeşitli sektörlerde faaliyeti gerçekleştiren şirketleri, temsilen açıklanan bu endeks verileri hem yatırımcılar hem de ülkemiz adına oldukça önemlidir. Dolayısıyla endeks verilerinin ülkemizin ekonomik durumuyla olan ilişkisi ve yatırımcıların, yatırım kararları almasındaki etkisi yadsınamaz.

Araştırmanın amacı yapay sinir ağları yöntemi ile Borsa İstanbul 30 Endeksini makroekonomik değişkenler aracılığıyla modelleyerek tahmin etmek ve Borsa İstanbul 30 Endeksinin makroekonomik değişkenler ile ortak hareketini saptayarak yatırımcıların yararına bilgiler sunmaktır. Ancak unutulmamalıdır ki yatırım kararı alacak olan yatırımcılar yalnızca tespit edilen kurallar doğrultusunda değil, çeşitli teknik analiz desteğini ve diğer ekstrem noktaları da göz önüne almalıdır. Araştırmada yapay sinir ağları yönteminde girdi değişkenleri olarak kullanılan makroekonomik değişkenler, literatür taraması sonucunda analize uygun olarak seçilmiştir. Seçilen değişkenler ışığında tahmin edilen Borsa İstanbul 30 Endeksinin makroekonomik değişkenler ile birlikte değişimi birliktelik kuralı analizi ile incelenmiştir. Dolayısıyla araştırmada veri madenciliği yöntemlerinden hem tahminleyici hem de tanımlayıcı modeller ile çalışılmıştır. Araştırma sürecinde aşağıdaki adımlar izlenmiştir:

- ✓ Veriye ulaşılabilirliğin ve veri toplamanın oldukça kolay olduğu dijitalleşen teknoloji dünyasında, gün geçtikçe daha çok artan veri yığınlarını işleyebilir duruma getiren veri madenciliği yöntemleri tanıtılmıştır.
- ✓ Tanıtılan modeller kapsamında yapay sinir ağları yöntemi ile Borsa İstanbul 30 Endeksini modelleyerek en iyi performansı gösteren model seçilmiştir.
- ✓ Seçilen en iyi model aracılığıyla Borsa İstanbul 30 Endeksi 2022 Aralık ayı değerleri tahmin edilmiştir.
- ✓ Tahmin edilen modele ait girdi değişkenleri ile Borsa İstanbul 30 Endeksinin ortak hareketi birliktelik kuralı analizi aracılığıyla tanımlanmıştır.
- ✓ Analiz sonucunda elde edilen bulgular değerlendirilmiştir.

Yapay sinir ağları analizinde 8 makroekonomik değişken girdi değişkeni olarak seçilmiş olup buna karşılık çıktı değişkeni olarak Borsa İstanbul 30 Endeksi seçilmiştir. Analize başlamadan önce veri setinin çok büyük ve çok küçük değerler içerdiği göz önüne alınarak veri setine normalleştirme işlemi yapılmıştır. 1253 adet günlük veri ile çalışılmış ve veri setinin %70'i eğitim, %15'i doğrulama, %15'i ise test verisi olarak alınmıştır. Performans kriteri olarak MSE ve R^2 seçilmiştir. Deneme-yanılma yöntemiyle en küçük MSE değerini bulana kadar program tekrar tekrar çalıştırılmıştır. En iyi performansa sahip olan model simüle edilerek Borsa İstanbul 30 Endeksi 2022 Aralık ayı günlük değerleri tahmin edilmiştir. Analiz sonucunda aşağıdaki bulgulara ulaşılmıştır:

- ✓ MSE değeri en küçük ve R^2 değeri en büyük olan, dolayısıyla en iyi performans gösteren ağın mimarisi 8-14-1 olarak bulunmuştur. 1 gizli katmandan oluşan ağ ileri beslemeli ve çok katmanlı algılayıcıdır.
- ✓ En iyi performans gösteren ağın simüle edilerek Borsa İstanbul 30 Endeksine ait 2022 Aralık ayı günlük değerleri tahmin edilmiştir. Tahmin sonuçları ile gerçek değerler karşılaştırılarak yapay sinir ağı analizinin gerçek değerlere yakın değerler ürettiği sonucuna varılmıştır.

Son olarak araştırmada 1253 günlük veri setine Birliktelik kuralı analizi uygulanmıştır. Birliktelik kuralı analizine başlamadan önce veri seti bir önceki günün kapanış fiyatları göz önünde bulundurularak, ilgili değerlerin bir önceki güne göre artış göstermesi durumunda 1 azalış göstermesi durumunda ise; 0 değeri verilerek Excel de

dönüştürülmüştür. Analizde üretilecek olan birliktelik kurallarında güven düzeyi %95 olarak seçilerek kuralların güven düzeyine göre sıralanması istenmiştir. Birliktelik kuralı analiziyle ulaşılan sonuçlar aşağıdaki gibidir:

- ✓ Analiz sonucunda 38 birliktelik kuralına ulaşılmıştır.
- ✓ Ulaşılan bu kurallardan 9 tanesi Borsa İstanbul 30 Endeksi ile araştırmada kullanılan makroekonomik değişkenler arasında gerçekleşmiştir.
- ✓ Maksimum güven düzeyine sahip olan kural Borsa İstanbul 30 Endeksi, Mevduat Faizi ve Tüketici Fiyat Endeksi arasında gerçekleşmiştir.

Araştırma bulgularının literatür ile uyumu değerlendirildiğinde, bu araştırmanın aşağıdaki araştırmalar ile uyumlu olduğu görülmüştür:

- ✓ Kantar (2020) araştırmasında BİST 100 endeksini ARMA modeli ve yapay sinir ağları modeli ile tahmin ederek yapay sinir ağlarının yüksek tahmin gücüne sahip olduğu bulgusuna ulaşmıştır.
- ✓ Karakul (2020) araştırmasında BİST 100 endeksi ile gecelik faiz oranları ve dolar kuru arasındaki ilişkiyi yapay sinir ağları ile tahmin ederek yapay sinir ağları modelinin finans alanında uygulanabilir olduğu ve güçlü tahminler ürettiği sonucuna ulaşmıştır.
- ✓ Tektaş ve Karataş (2004), Yakut ve diğ. (2014), Chen (2003), Bahrammirzae (2010) araştırmalarında finans alanında yapay sinir ağları modeli kullanmışlardır. Analiz sonucunda yapay sinir ağı modelinin güçlü tahminler ürettiğini sonucuna varmışlardır.
- ✓ Karaatlı ve diğ. (2021) araştırmalarında BİST-30'da işlem gören şirketler ile çeşitli makroekonomik değişkenler arasındaki ilişkiyi ortaya koymuşlardır. Birliktelik kuralı analizinden FP-Growth algoritmasını kullanarak 81 birliktelik kuralı tespit etmişlerdir. BİST-30 endeksinde yer alan bankacılık sektörüne ait hisselerinin birlikte hareket ettiği sonucuna varmışlardır. Bankacılık endeksi analiz dışında bırakıldığında ise 76 birliktelik kuralı tespit edilmiş ve bu kuralların 71'nin altın değişkenini içerdiği sonucuna ulaşmışlardır.

- ✓ Gemici (2012) araştırmasında İMKB’ de işlem gören 10 hisse senedinin aralarındaki birlikte değişimi saptamayı amaçlamış ve analiz sonucunda hisse senetleri arasında 9 birliktelik kuralının üretildiğini saptamıştır.
- ✓ Na ve Sohn (2011), Lia ve Chou (2013),Arafah ve Mukhlash (2015) araştırmalarında yabancı borsada yer alan endeksleri baz alarak veri madenciliği tekniklerinden birliktelik kuralı analizini kullanmışlardır. Araştırma sonucunda birliktelik kuralları üretilerek endekslerin birlikte değişimi saptanmıştır.

Araştırmanın kısıtları çerçevesinde, Borsa İstanbul 30 Endeksi ve çeşitli makroekonomik değişkenler ele alınmıştır. Bu bağlamda gelecek çalışmalarda farklı makroekonomik değişkenler ele alınarak Borsa İstanbul 50 Endeksi veya Borsa İstanbul 100 Endeksi yapay sinir ağları analizi ile modellenerek en iyi performansa sahip model belirlenebileceği gibi farklı analizler ile yapay sinir ağı analizi karşılaştırılarak en iyi model seçilebilir. Ayrıca araştırmada birliktelik kuralı analizi için kullanılan FP-Growth algoritması yerine farklı bir algoritma ve güven düzeyi seçilerek değişkenlerin ortak hareketi incelenebilir.

KAYNAKÇA

Abraham, A. Abraham, A. (2005). Artificial Neural Networks. *Artificial Neural Networks Handbook of Measuring System Design* (s. 901-908) içinde. New York: John Wiley & Sons.

Agrawal, R., & Srikant, R. (1994). Fast Algorithms for Mining Association Rules . *Proceedings of the 20th VLDB Conference*, (s. 487-499). Santiago, Chile.

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining Association Rules between Sets of Items in Large Databases. *ACM SIGMOD Conference on Management of Data*, (s. 207-216). Washington.

Akaslan, D., & Taşkın, S. (2015). *MATLAB ile Teknik Programlama*. İstanbul: Papatya Yayıncılık Eğitim.

Akcan, A., & Kartal, C. (2011). İMKB Sigorta Endeksini Oluşturan Şirketlerin Hisse Senedi Fiyatlarının Yapay Sinir Ağları İle Tahmini. *Muhasebe ve Finansman Dergisi* , 27-40.

Akgün, E. (2015). *Krizi Anlamak Makroekonomik Göstergeler Işığında Ulusal ve Uluslararası Krizler Karşılaştırmalı Türkiye Örneği*. Bursa: Ekin Basım Yayın Dağıtım.

Akpınar, H. (2000). Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. *İ.Ü İşletme Fakültesi Dergisi* , 29 (1), 1-22.

Akpınar, H. (2017). *Data Veri Madenciliği Veri Analizi*. İstanbul: Papatya Yayıncılık Eğitim.

Aktan, E. (2018). Büyük Veri: Uygulama Alanları, Analitiği ve. *Bilgi Yönetim Dergisi* , 1 (1), 1-22.

Aktop, V. S. (2010). *Türkiye'de Para Arzının İçselliği: Ekonometrik Bir Uygulama*. Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, İktisat Anabilim Dalı Genel İktisat Programı, İzmir.

Aktürk, H., & Korukoğlu, S. (2008). Veri Madenciliği Teknolojisi Kullanarak Fiyat Değişimlerinde Paralellik Gösteren Hisse Senetlerinin Bulunması ve Risk Azaltılması. *Akademik Bilişim 2008* (s. 113-119). Çanakkale: Çanakkale Onsekiz Mart Üniversitesi.

Altunkaynak, B. (2017). *Veri Madenciliği Yöntemleri ve R Uygulamaları*. Ankara: Seçkin Yayıncılık.

Anderson, D., & McNeill, G. (1992). Artificial Neural Networks Technology. *New York: Kaman Sciences Corporation* , 1-83.

Araç, Y. E., & Gürhanlı, A. (2020). Yapay Sinir Ağını Kullanarak Müşteri Memnuniyeti Analizi. *DÜMF Mühendislik Dergisi* (11:1), 39-55.

Arafah, A. A., & Mukhlash, I. (2015). The Application of Fuzzy Association Rule on Co-Movement Analyze of Indonesian Stock Price. *Procedia Computer Science* 59 , 235-243.

Argiddi, R., & Apte, S. (2012). Future Trend Prediction of Indian IT Stock Market Using Association Rule Mining Of Transaction Data. *International Journal of Computer Applications* , 39 (10), 30-34.

Arı, A., & Berberler, M. E. (2017). Yapay Sinir Ağları İle Tahmin ve Sınıflandırma Problemlerinin Çözümü İçin Arayüz Tasarımı. *Acta Infologica* , 1 (2), 55-73.

Arıkan Kargı, V. S. (2015). *Yapay Sinir Ağ Modelleri ve Bir Tekstil Firmasında Uygulama*. Bursa: Ekin Basım Yayın Dağıtım.

Arslan, A., & İnce, R. (1996). The Neural Network Approximation To The Size Effect In Fracture Of Cementitious Materials. *Engineering Fracture Mechanics* , 54 (2), 249-261.

Arslan, H. (2008). Sakarya Üniversitesi Web Sitesi Erişim Kayıtlarının Web Madenciliği İle Analizi. *Sakarya Üniversitesi Fen Bilimleri Enstitüsü* .

Asilkan, Ö., & Irmak, S. (2009). İkinci El Otomobillerin Gelecekteki Fiyatlarının Yapay Sinir Ağları İle Tahmin Edilmesi. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* , 14 (2), 375-391.

Atan, S. (2016). Veri, Büyük Veri Ve İşletmecilik. *Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* , 19 (35), 137-153.

Ataseven, B. (2013). Yapay sinir ağları ile öngörü modellemesi. *Marmara Üniversitesi Sosyal Bilimler Enstitüsü Öneri Dergisi* , 10 (39), 101-115.

Aydemir, E. (2019). *WEKA ile Yapay Zeka*. Ankara: Seçkin Yayıncılık San. ve Tic A.Ş.

Aydın, S. (2007). *Veri Madenciliği ve Anadolu Üniversitesi Uzaktan Eğitim Sisteminde Bir Uygulama*. Doktora Tezi, Anadolu Üniversitesi Sosyal Bilimler Enstitüsü, Eskişehir.

Aygün Alıcı, V., & Köseoğlu, M. (2021). Türkiye'de Altın Fiyatlarını Etkileyen Faktörlerin Ekonometrik Analizi. *Uluslararası Ekonomi, İşletme ve Politika Dergisi* , 5 (2), 254-275.

Bahrammirzaee, A. (2010). A Comparative Survey Of Artificial Intelligence Applications In Finance: Artificial Neural Networks, Expert System And Hybrid Intelligent Systems. *Neural Comput & Applic.* , 19 (8), 1165-1195.

- Balaban, M., & Kartal, E. (2018). *Veri Madenciliği ve Makine Öğrenmesi Temel Algoritmaları ve R Dili İle Uygulamaları*. İstanbul: Çağlayan Kitapevi ve Eğitim Çözümleri.
- Baydoğan, M. G., Orbay, B., & Çetin, U. (2014). XIX. Türkiye'de İnternet Konferansı. *XIX. Türkiye'de İnternet Konferansı*, (s. 1-42). İzmir.
- Bayır, F. (2006). *Yapay Sinir Ağları ve Tahmin Modellemesi Üzerine Bir Uygulama*. Yüksek Lisans Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı Sayısal Yöntemler Bilim Dalı, İstanbul.
- Baykal, A. (2006). Veri Madenciliği Uygulama Alanları. *D.Ü Ziya Gökalp Eğitim Fakültesi Dergisi* (7), 95-107.
- Bayram, N. (2015). *Sosyal Bilimlerde SPSS ile Veri Analizi* (Gözden Geçirilmiş Genişletilmiş 5.Baskı b.). Bursa: Ezgi Kitapevi.
- Bayram, S. S., & DüNDAR, S. (2021). Türkiye'de Banka Şubesi Lokasyonunun Veri Madenciliği İle Analizi. *Uluslararası Bankacılık Ekonomik ve Yönetim Araştırmaları Dergisi* , 4 (1), 34-52.
- Beale, M. H., Hagan, M. T., & Demuth, H. B. (2017). *Neural Network Toolbox User's Guide*. The Math Works Inc.
- Bozkurt Uzan, Ş., & Özüçalışlıyan, M. Ö. (2020). Turizm Sektöründe Veri Madenciliği. *Türk Turizm Araştırmaları Dergisi* , 4 (4), 3465-3485.
- Chen, A.-S., Leung, M. T., & Daouk, H. (2003). Application Of Neural Networks To An Emerging Financial Market: Forecasting And Trading The Taiwan Stock Index. *Computers & Operations Research* 30 , 901-923.
- Cyganek, B., Graña, M., Krawczyk, B., Kasprzak, A., Porwik, P., & Woźniak, K. W. (2016). A Survey of Big Data Issues in Electronic Health Record Analysis. *Applied Artificial Intelligence* , 30 (6), 497-520.
- Çalışkan Çavdar, Ş., & Aydın, A. D. (2018). *Finans Alanında Yapay Zeka ve Ekonometrik Uygulamalar: Ekonometrik ve Hibrit Modeller Tahmini*. Ankara: Seçkin Yayıncılık.
- Çalışkan, M. M., & Deniz, D. (2015). Yapay Sinir Ağlarıyla Hisse Senedi Fiyatları ve Yönlerinin Tahmini. *Eskişehir Osmangazi Üniversitesi İİBF Dergisi* , 10 (3), 177-194.
- Çanakçı, A. (2006). *Yapay Sinir Ağlarının Makroekonomik Bir Model Üzerine Uygulanması: Bir Türkiye Örneği*. Yüksek Lisans Tezi, Gazi Üniversitesi Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, Ankara.

Çelik, M. (2009). *Veri Madenciliğinde Kullanılan Sınıflandırma Yöntemleri ve Bir Uygulama*. Yüksek Lisans Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.

Çınaroğlu, E., & Avcı, T. (2020). THY Hisse Senedi Değerinin Yapay Sinir Ağları İle Kestirimi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi* , 34 (1), 1-20.

Dahdouh, K., Dakkak, A., Oughdir, L., & Ibriz, A. (2020). Improving Online Education Using Big Data Technologies. F. Altınay içinde, *The Role Of Technology in Education* (s. 15-34). London, United Kingdom: Intech Open.

Dastres, R., & Soori, M. (2021). Artificial Neural Network Systems. *International Journal of Imaging and Robotics* , 21 (2), 14-25.

Dener, M., Dörterler, M., & Orman, A. (2009). Açık Kaynak Kodlu Veri Madenciliği Programları: WEKA'da Örnek Uygulama. *Akademik Bilişim 09 - XI. Akademik Bilişim Konferansı Bildirileri* (s. 787-796). Şanlıurfa: Harran Üniversitesi.

Detienne, K. B., Detienne, D. H., & Joshi, S. A. (2003). Neural Networks as Statistical Tools for Business Researchers. *Organizational Research Methods* , 6 (2), 236-265.

Doğan, C., & Uluman, M. (2016). İstatistiksel Veri Analizinde R Yazılımı ve Kullanımı. *İlköğretim Online* 15(2) , 615-634.

Doğan, K., & Arslantekin, S. (2016). Büyük Veri:Önemi,Yapısı ve Günümüzdeki Durum. *DTCF Dergisi* 56.1 , 15-36.

Dombaycı, Ö. A., & Gölcü, M. (2009). Daily means ambient temperature prediction using artificial neural network method: A case study of Turkey. *Renewable Energy* (34), 1158-1161.

Dunham, M. H., Xiao, Y., Gruenwald, L., & Hossain, Z. (2000). A Survey of Association Rules. *ACM Survey Journal* , 1-65.

Durmaz, C., & Kocamış, M. (2008). Oracle Data Miner ile Öğrenci Kayıtları Üzerine Bir Veri Madenciliği Uygulaması. *Bitirme Ödevi* . İstanbul: İstanbul Teknik Üniversitesi.

Efe, M., & Kaynak, O. (2000). *Yapay Sinir Ağları ve Uygulamaları*. İstanbul: Boğaziçi Üniversitesi.

Elmas, Ç. (2003). *Yapay Sinir Ağları (Kuram,Mimari,Eğitim,Uygulama)*. Ankara: Seçkin Yayıncılık.

Elmas, Ç. (2011). *Yapay Zeka Uygulamaları*. Ankara: Seçkin Yayıncılık.

- Emel, G. G., & Taşkın, Ç. (2005). Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi* , 6 (2), 222-239.
- Emre, İ. E., & Selçukcan Erol, Ç. (2017). Veri Analizinde İstatistik mi Veri Madenciliği mi ? *Bilişim Teknolojileri Dergisi* , 10 (2), 161-167.
- Erdoğan, E., & Özyürek, H. (2012). Yapay Sinir Ağları İle Fiyat Tahminlemesi. *Sosyal ve Beşeri Bilimler Dergisi* , 4 (1), 85-92.
- Erpolat, S. (2012). Otomobil Yetkili Servislerinde Birliktelik Kurallarının Belirlenmesinde Apriori ve FP-Growth Algoritmalarının Karşılaştırılması. *Anadolu Üniversitesi Sosyal Bilimler Dergisi* , 12 (1), 151-166.
- Ersoy, E., & Karal, Ö. (2012). Yapay Sinir Ağları ve İnsan Beyni. *İnsan ve Toplum Bilimleri Araştırmaları Dergisi* , 1 (2), 188-205.
- Ersöz, F. (2019). *Veri Madenciliği Teknikleri ve Uygulamaları* (Güncellenmiş 3.Baskı b.). Ankara: Seçkin Yayıncılık.
- Erten, H. (2015). *Veri Madenciliği Teknikleri İle Organ Nakli İçin Uygun Donör Oranının Hesaplanması*. Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Etzioni, O. (1996). The World-Wide Web:Quagmire or Gold Mine . *Communications of The Acm* , 39 (11), 65-68.
- Fausett, L. (1994). *Fundamentals Of Neural Network:Architectures, Algorithms and Applications*. London: Prentice Hall.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine Volume 17* (3), 37-54.
- Filiz, E., Karaboğa, H. A., & Akoğul, S. (2017). BİST-50 Endeksi Değişim Değerlerinin Sınıflandırılmasında Makine Öğrenmesi Yöntemleri ve Yapay Sinir Ağları Kullanımı. *Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi* , 26 (1), 231-241.
- Fyfe, C. (2000). *Artificial Neural Networks and Information Theory*. The University of Paisley.
- Gemici, B. (2012). *Veri Madenciliği ve Bir Uygulaması*. Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, Ekonometri Bölümü, İzmir.
- Gershenson, C. (2003) *Artificial Neural Networks for Beginners*. 02.02.2023 tarihinde <https://arxiv.org/abs/cs/0308031> adresinden alındı.

Girotra, M., Nagpal, K., Minocha, S., & Sharma, N. (2013). Comparative Survey on Association Rule Mining Algorithms. *International Journal of Computer Applications* (0975 – 8887) , 84 (10), 18-22.

Giudici, P. (2003). *Applied Data Mining:Statistical Methods for Business and Industry*. England: West Sussex,John Wiley & Sons.

Gonzalez, S. (2007). Neural Networks for Macroeconomic Forecasting: A Complementary Approach to Linear Regression Models. Finance Canada Working Paper.

Graupe, D. (2013). *Principles Of Artifical Neural Networks* (3 rd Edition b.). World Scientific.

Güler, E. (2016). Türk Bankacılık Sisteminde Mevduat Faizi Değişimlerinin Katılma Fonları Üzerindeki Etkisi Ve Likitide Riski Yönetimi. *Finansal Araştırmalar ve Çalışmalar Dergisi* , 8 (15), 385-412.

Gürsakal, N., Aydın, Z. B., Gürsakal, S., & Tüzüntürk, S. (2008). Ağ Bilimi ve İstatistik. *9.Ekonometri ve İstatistik Sempozyumu* . (Dokuz Eylül Üniversitesi.) Kuşadası.

Gürsakal, N. (2009). *Sosyal Ağ Analizi*. Bursa: Dora Yayıncılık.

Gürsakal, N. (2014). *Büyük Veri*. Bursa: Dora Basım-Yayın Dağıtım.

H.Dunham, M. (2003). *Data Mining:Introductory and Advanced Topics*. NJ: Prentice-Hall Upper Saddle River.

Hagan, M. T., Demuth, H. B., Beale, M. H., & Jesus, O. D. (2014). *Neural Network Design* (2 b.). PWS Publishing Co.

Hakimpoor, H., Arshad, K. A., Tat, H. H., & Rahmandoust, K. a. (2011). Artificial Neural Networks' Applications in Management. *World Applied Sciences Journal* , 14 (7), 1008-1019.

Hamzaçebi, C. (2021). *Matlab Uygulamalı Yapay Sinir Ağları* (Güncellenmiş 2.Baskı b.). Ankara: Seçkin Yayıncılık.

Han, J., Kamber, M., & Pei, J. (2012). *Data mining:concepts and techniques*. Waltham: MA:Elsevier.

Hasanlı, H. (2014). *Çok Boyutlu Veri Tabanlarında Veri Madenciliği Yöntemleri Kullanılarak Bilgi Keşfi*. Ege Üniversitesi Fen Bilimleri Enstitüsü, İzmir.

Haykin, S. (2009). *Neural Networks and Learning Machines*. New Jersey: Person International Edition.

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. J.Wiley;Chapman & Hall.

İnan, O. (2015). *Veri Madenciliği Uygulamaları İçin Veri İndirgeme Algoritmalarının Geliştirilmesi ve Resim Madenciliğine Uygulanması*. Doktora Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, Konya.

İnce, A. R., & Alan, M. A. (2022). A Study on The Use of Data Mining in The Planning of Investment Field. *Journal of Research in Economics, Politics & Finance* , 7 (1), 1-15.

Islamiyah, Ginting, P. L., Dengen, N., & Taruk, M. (2019). Comparison of Priori and FPGrowth Algorithms in Determining Association Rules. *International Conference on Electrical Electronics and Information Engineering (ICEEIE)*, (s. 320-23). Denpasar, Bali, Indonesia: IEEE.

Jalpa, P., & Rustom, M. (2017). A Novel Hybrid Method for Generating Association Rules for Stock Market Data. *International Journal of Latest Technology in Engineering, Management & Applied Science 3rd Special Issue on Engineering and Technology* , 6 (7), 6-15.

Kantar, L. (2020). BIST100 Endeksinin Yapay Sinir Ağları ve ARMA Modeli İle Tahmini. *Muhasebe ve Finans İncelemeleri Dergisi* , 3 (2), 121-131.

Karaatlı, M., & Altıntaş, E. (2018). Borsa İstanbul işletmelerinin Veri Madenciliği İle Kümelenmesi. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* , 10 (26), 817-886.

Karaatlı, M., Kocabıyık, T., Yalçınır Çal, D., & Çolak, M. (2021). BIST-30 Endeksinde Yer Alan Payların Ortak Hareketlerinin Veri Madenciliği Kapsamında Birliktelik Kuralı İle İncelenmesi. *Finansal Araştırmalar ve Çalışmalar Dergisi* , 13 (25), 548-571.

Karaibrahimoğlu, A. (2014). *Veri Madenciliğinde Birliktelik Kuralı İle Onkoloji Verilerinin Analiz Edilmesi: Meram Tıp Fakültesi Onkoloji Örneği*. Doktora Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Konya.

Karakul, A. K. (2020). Yapay Sinir Ağları İle BIST-100 Endeks Değerinin Tahmin Edilmesi. *Mehmet Akif Ersoy Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* , 7 (2), 497-509.

Kaya, M., & Özel, S. A. (2014). Açık Kaynak Kodlu Veri Madenciliği Yazılımlarının Karşılaştırılması. *Akademik Bilişim 14 - XVI. Akademik Bilişim Konferansı Bildirileri* (s. 47-53). Mersin: Mersin Üniversitesi.

Keçe, A. (2006). *Yapay Sinir Ağları İle Plastik Enjeksiyon Süreci Başlangıç Parametrelerinin Belirlenmesi*. Yüksek Lisans Tezi, Bursa Uludağ Üniversitesi , Endüstri Mühendisliği Anabilim Dalı, Bursa.

Keskenler, M. F., & Keskenler, E. F. (2017). Geçmişten Günümüze Yapay Sinir Ağları ve Tarihçesi. *Takvim-i Vekayi* , 5 (2), 8-18.

Khurana, K., & Sharma, M. S. (2013). A Comparative Analysis of Association Rules Mining Algorithms. *International Journal of Scientific and Research Publications* , 3 (5), 1-4.

Kocabıyık, T., Dağ, O., & Karaatlı, M. (2021). Borsa İstanbul Endekslerinin Birlikte Hareketi:FP Growth Algoritması İle Bir Uygulama. *Uluslararası İşletme,Ekonomi ve Yönetim Perspektifleri Dergisi (IJBEMP)* , 5 (2), 659-672.

Koç, M., Balas, C. E., & Arslan, A. (2004). Taş Dolgu Dalgakıranların Yapay Sinir Ağları ile Ön Tasarımı. *İMO Teknik Dergi* (225), 3351-3375.

Koleyni, K. (2009). Using Artificial Neural Network For Income Convergence. *Global Journal Of Business Research* , 3 (2), 141-152.

Koyuncugil, A. S. (2007). *Veri Madenciliği ve Sermaye Piyasalarına Uygulanması*. Sermaye Piyasası Kurulu .

Koyuncugil, A. S., & Özgülbaş, N. (2008). İMKB'de İşlem Gören Kobi'lerin Güçlü Ve Zayıf Yönleri:Chaid Karar Ağacı Uygulaması. *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* , 23 (1), 1-21.

Köse, İ. (2018). *Veri Madenciliği Teori Uygulama ve Felsefesi*. İstanbul: Papatya Yayıncılık Eğitim A.Ş.

Krenker, A., Bester, J., & Kos, A. (2011). Introduction to the Artificial Neural Networks. *Artificial Neural Networks - Methodological Advances and Biomedical Applications* , 1-18.

Kröse, B., & Smagt, P. v. (1996). *An Introduction to Neural Networks* (8.Baskı b.). The University of Amsterdam.

Kumbhare, T. A., & Chobe, S. V. (2014). An Overview of Association Rule Mining Algorithms. *International Journal of Computer Science and Information Technologies* , 5 (1), 927-930.

Kurt, A. (2021). *Pandemi Döneminde Brent Petrol ve Kripto Paralar Arasındaki İlişkinin İncelenmesi*. Yüksek Lisans Tezi, Afyon Kocatepe Üniversitesi, Uluslararası Ticaret ve Finansman Anabilim Dalı, Afyonkarahisar.

- Kutlu, B., & Badur, B. (2009). Yapay Sinir Ağları ile Borsa Endeksi Tahmini. *İstanbul Üniversitesi İşletme Fakültesi İşletme İktisadi Enstitüsü Dergisi* (63), 25-40.
- Lee, K., Yoo, S., & Jin, J. J. (2007). Neural Network Model vs.SARIMA Model In Forecasting Korean Stock Price Index (KOSPI). *Issues In Information Systems* , VIII (2), 372-378.
- Lee, T.-S., & Chiu, C.-C. (2002). Neural Network Forecasting of An Opening Cash Price Endex. *International Journal of Systems Science* , 33 (3), 229-237.
- Liao, S.-H., & Chou, S.-Y. (2013). Data Mining Investigation Of Co-Movements On The Taiwan And China Stock Markets For Future İvestment Portfolio. *Expert Systems With Applications* , 40 (5), 1542-1554.
- Maltarollo, V. G., Honório, K. M., & Silva, A. B. (2013). Applications of Artificial Neural Networks in Chemical Problems. K. Suzuki içinde, *Artificial Neural Networks - Architectures and Applications* (s. 203-223). Intech Open.
- McCulloch, W. S., & Pitts, W. (1943). A Logical Calculus Of The Ideas Immanent In Nervous Activity. *Bulletin Of Mathematical Biophysics* (5), 115-133.
- Mishra, V., Mishra, T. K., & Mishra, A. (2013). Algorithms for Association Rule Mining:A General Survey on Benefits And Drawbacks of Algorithms. *International Journal of Advanced Research in Computer Science* , 4 (8), 155-159.
- Na, S. H., & Sohn, S. Y. (2011). Forecasting changes in Korea Composite Stock Price Index (KOSPI) using association rules. *Expert Systems with Applications* , 9046–9049.
- Nabiyev, V. (2021). *Yapay Zeka:İnsan-Bilgisayar Etkileşimi* (6.Baskı b.). Ankara: Seçkin Yayıncılık.
- Necati, C., & Duykuluoğlu, A. (2020). *Sosyal Bilimlerde Veri Madenciliği* (2.Baskı b.). Ankara: Pegem Akademi.
- Oğuzlar, A. (2004a). *Veri Madenciliğine Giriş*. Bursa: Ekin Kitapevi.
- Oğuzlar,A. (2004b). *Veri Madenciliğinde Birliktelik Kuralları*. Öneri Dergisi, 6 (22), 315-321.
- Oğuzlar, A. (2011). *Temel Metin Madenciliği*. Bursa: Dora Basım Yayın Dağıtım.
- Özbay, Ö. (2015). Veri Madenciliği Kavramı ve Eğitimde Veri Madenciliği Uygulamaları. *Uluslararası Eğitim Bilimleri Dergisi* , 2 (5), 262-272.
- Özdemir, A., Yıldıztepe, E., & Binar, M. (2010). İstatistiksel Yazılım Geliştirme Ortamı:R. *Akademik Bilişim 10 - XII. Akademik Bilişim Konferansı Bildirileri*, (s. 293-297). Muğla.

Özer, A., Sarı, S. S., & Başakın, E. E. (2018). Bulanık Mantık ve Yapay Sinir Ağları İle Borsa Endeks Tahmini: Gelişmiş ve Gelişmekte Olan Ülkeler Örneği. *Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* (1), 100-123.

Özgen, D. (2007). *Yapay Sinir Ağları Analizi ve Türk Finans Piyasaları:İMKB 30 Endeksi Uygulaması*. Yüksek Lisans Tezi, Marmara Üniversitesi, Bankacılık ve Sigortacılık Enstitüsü Sermaye Piyasaları ve Borsa Bölümü, İstanbul.

Özkan, M., & Boran, L. (2014). Veri Madenciliğinin Finansal Kararlarda Kullanımı. *Çankırı Karatekin Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* , 4 (1), 59-82.

Özkan, Y. (2020). *Veri Madenciliği Yöntemleri* (4.Basım Güncellenmiş ve Genişletilmiş b.). İstanbul: Papatya Bilim Üniversite Yayıncılık.

Öztemel, E. (2006). *Yapay Sinir Ağları*. İstanbul: Papatya Yayıncılık.

Öztemel, E. (2012). *Yapay Sinir Ağları* (3.Basım b.). İstanbul: Papatya Yayıncılık.

Öztürk, K., & Şahin, M. E. (2018). Yapay Sinir Ağları ve Yapay Zekâ'ya Genel Bir Bakış. *Takvim-i Vekayi* , 6 (2), 25-36.

Özyiğit, H. (2022). Muhasebe Alanına Güncel Yaklaşımlar:Metin Madenciliği. *Muhasebe ve Vergi Uygulamaları Dergisi* , 637-663.

Polat, E. D. (2022). *Beyaz Eşya Sektöründe Satış Tahmini: Bir Veri Madenciliği Uygulaması*. Yüksek Lisans Tezi, Bursa Uludağ Üniversitesi Sosyal Bilimler Enstitüsü, Bursa.

Rosenblatt, F. (1958). The perceptron: A probabilistic Model For Information Storage and Organization In The Brain. *Psychological Review* , 65 (6), 386–408.

Sahoo, S., & Mohanty, M. N. (2020). Stock Market Price Prediction Employing Artificial Neural Network Optimized by GrayWolf Optimization. In: *Patnail S, Ip A, Tavana M, Jain V (eds) New paradigm in decision science and management.* , 77–87.

Sarıkaya, G. (2019). Yapay Sinir Ağları Ve Regresyon Modelleri İle Bist Ulusal -100 Endeksinin Tahmini. *21. Yüzyılda Eğitim ve Toplum* , 325-340.

Sarle, W. S. (1994). Neural Networks and Statistical Models. *Proceedings of the Nineteenth Annual SAS Users Group International Conference*, (s. 1-13).

Savaş, S., Topaloğlu, N., & Yılmaz, M. (2012). Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi* , 1-23.

Saxena, A., & Rajpoot, V. (2020). A Comparative Analysis of Association Rule Mining Algorithms. *IOP Conf. Series: Materials Science and Engineering*, Volume 1099, *International Conference on Applied Scientific Computational Intelligence using Data Science (ASCI 2020)* (s. 1-10). Jaipur India: IOP Publishing.

Sert, F. (2014). *Hava Durumunun Yapay Sinir Ağları İle Kestirimi ve Bulanık Mantıkla Sınıflandırılması*. Yüksek Lisans Tezi, Bursa Uludağ Üniversitesi Sosyal Bilimler Enstitüsü, Bursa.

Sever, H., & Oğuz, B. (2002). Veri Tabanlarında Bilgi Keşfine Formel Bir Yaklaşım Kısım I: Eşleştirme Sorguları ve Algoritmalar. *Bilgi Dünyası* , 3 (2), 173-204.

Sevimli Deniz, S. (2019). Veri Madenciliği Kullanılarak Türkiye'nin Turizm Gelirlerinin Aylara Göre Yapay Sinir Ağları İle Tahminlenmesi. *Sosyal Bilimler Enstitüsü Dergisi* , 241-255.

Shearer, C. (2000). The CRISP-DM Model:The New Blueprint for Data Mining. *Journal of Data Warehousing* , 5 (4), 13-23.

Silahtaroğlu, G. (2016). *Veri Madenciliği Kavram ve Algoritmaları*. İstanbul: Papatya Yayıncılık Eğitim.

Silva, I. N., & vd. (2017). *Artificial Neural Networks A Practical Course*. Switzerland: Springer.

Smith, K. A., & Gupta, J. N. (2000). Neural networks in business: techniques and applications for the operations researcher. *Computers & Operations Research* (27), 1023-1044.

Süsay, A., & Ergin Ünal, A. (2020). Seçilmiş Güven Endeksleri,VIX ve CDS Primlerinin Büyümeye Etkisi:Türkiye Örneği. *Ufuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* , 9 (18), 83-97.

Şen, Z. (2004). *Yapay Sinir Ağları İlkeleri*. İstanbul: Su Vakfı Yayınları.

Şentürk, A. (2006). *Veri Madenciliği Kavram ve Teknikler*. Bursa: Ekin Yayınevi.

Tekerek, A. (2011). Veri Madenciliği Süreçleri ve Açık Kaynak Kodlu Veri Madenciliği Araçları. *Akademik Bilişim'11 - XIII. Akademik Bilişim Konferansı Bildirileri* (s. 161-169). Malatya: İnönü Üniversitesi.

Tektaş, A., & Karataş, A. (2004). Yapay Sinir Ağları ve Finans Alanına Uygulanması:Hisse Senedi Fiyat Tahminlemesi . *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi* , 3 (4), 337-349.

Telli, Ş. (2016). *BIST100 Endeksinin Ekonomik Takvim Etkisi Gözetilerek Yapay Sinir Ağları İle Öngörülmesi*. Yüksek Lisans tezi, Eskişehir Anadolu Üniversitesi Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı Finansman Bilim Dalı, Eskişehir.

Timur, E., & Günay, B. (2019). Petrol Fiyatları ile Pay Senetleri Arasındaki İlişkinin Ekonometrik Analizi. *Social Sciences Studies Journal(SSS journal)* , 5 (44), 5204-5211.

Ting, J., Fu, T.-c., & Chung, F.-l. (2006). Mining of Stock Data: Intra – and Inter-Stock Pattern Associative Classification. *Proceedings of the 2006 International Conference on Data Mining DMIN 2006*, (s. 29-36). Las Vegas, Nevada, USA.

Toraman, C. (2008). Demir-Çelik Sektöründe Yapay Sinir Ağları İle Hisse Senedi Fiyat Tahmini: Erdemir A.Ş. ve Kardemir A.Ş. Üzerine Bir Tahmin Uygulamas. *Muhasebe ve Finansman Dergisi Sayı* (39), 44-57.

Tunalı, H., & Özkan, İ. E. (2016). Türkiye'de Tüketici Güven Endeksi ve Tüketici Fiyat Endeksi Arasındaki İlişkinin Ampirik Analizi. *İktisat Politikası Araştırmaları Dergisi* , 3 (2), 54-67.

Tüzüntürk, S. (2010). Veri Madenciliği ve İstatistik. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* , XXIX (1), 65-90.

Ünsal, Ö. (2020). Veri Madenciliği Teknikleri İle Hisse Senetleri Arasındaki Fiyat Etkileşimlerinin Belirlenmesi. *Mühendislik Bilimleri ve Tasarım Dergisi* , 106-112.

Warner, B., & Mısra, M. (1996). Understanding Neural Networks as Statistical Tools. *The American Statistician* , Vol.50 (4), 284-293.

Weiss, G. M., & Davison, B. D. (2010). Data Mining. *To appear in the Handbook of Technology Management, H. Bidgoli (Ed.), John Wiley and Sons* , 2-17.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers is an imprint of Elsevier.

Wu, B., Zhang, D., Lan, Q., & Zheng, J. (2008). An Efficient Frequent Patterns Mining Algorithm based on Apriori Algorithm and the FP-tree Structure. *Third 2008 International Conference on Convergence and Hybrid Information Technology* (s. 1099-1102). Busan, Korea (South): Institute of Electrical and Electronics Engineers.

Yakut, E., Elmas, B., & Yavuz, S. (2014). Yapay Sinir Ağları ve Destek Vektör Makineleri Yöntemleriyle Borsa Endeksi Tahmini. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* , 19 (1), 139-157.

Yazıcı, A. C., Ögüş, E., Ankaralı, S., Canan, S., Ankaralı, H., & Zeki, A. (2007). Artificial Neural Networks: Review. *Türkiye Klinikleri Journal of Medical Sciences* (27), 65-71.

Yıldız, B., & Yezegel, A. (2010). Fundamental Analysis With Artificial Neural Network. *The International Journal of Business and Finance Research* , 4 (1), 149-157.

Yurtoğlu, H. (2005). *Yapay Sinir Ağları Metodolojisi İle Öngörü Modellemesi (DPT-Uzmanlık Tezleri)*. Ankara: Devlet Planlama Teşkilatı Yayınları.

Zhao, C.-M., & Luan, J. (2006). Data Mining:Going Beyond Traditional Statistics. *New Directions For Institutional Research* , 7-16.



EKLER

ŞİRKET	KOD	ŞİRKETİN SEKTÖRÜ
AKBANK T.A.Ş.	AKBNK	MALİ KURULUŞLAR / BANKALAR
ALARKO HOLDİNG A.Ş.	ALARK	MALİ KURULUŞLAR / HOLDİNGLER VE YATIRIM ŞİRKETLERİ
ARÇELİK A.Ş.	ARCLK	İMALAT / METAL EŞYA MAKİNE ELEKTRİKLİ CİHAZLAR VE ULAŞIM ARAÇLARI
ASELSAN ELEKTRONİK SANAYİ VE TİCARET A.Ş.	ASELS	TEKNOLOJİ / SAVUNMA
ASTOR ENERJİ A.Ş.	ASTOR	İMALAT / METAL EŞYA MAKİNE ELEKTRİKLİ CİHAZLAR VE ULAŞIM ARAÇLARI
BİM BİRLEŞİK MAĞAZALAR A.Ş.	BIMAS	TOPTAN VE PERAKENDE TİCARET / PERAKENDE TİCARET
EMLAK KONUT GAYRİMENKUL YATIRIM ORTAKLIĞI A.Ş.	EKGYO	MALİ KURULUŞLAR / GAYRİMENKUL YATIRIM ORTAKLIKLARI
ENKA İNŞAAT VE SANAYİ A.Ş.	ENKAI	İNŞAAT VE BAYINDIRLIK / İNŞAAT VE BAYINDIRLIK İŞLERİ
EREĞLİ DEMİR VE ÇELİK FABRİKALARI T.A.Ş.	EREGL	İMALAT / ANA METAL SANAYİ
FORD OTOMOTİV SANAYİ A.Ş.	FROTO	İMALAT / METAL EŞYA MAKİNE ELEKTRİKLİ CİHAZLAR VE ULAŞIM ARAÇLARI
GÜBRE FABRİKALARI T.A.Ş.	GUBRF	İMALAT / KİMYA İLAÇ PETROL LASTİK VE PLASTİK ÜRÜNLER
HACI ÖMER SABANCI HOLDİNG A.Ş.	SAHOL	MALİ KURULUŞLAR / HOLDİNGLER VE YATIRIM ŞİRKETLERİ
HEKTAŞ TİCARET T.A.Ş.	HEKTS	İMALAT / KİMYA İLAÇ PETROL LASTİK VE PLASTİK ÜRÜNLER
KARDEMİR KARABÜK DEMİR ÇELİK SANAYİ VE TİCARET A.Ş.	KRDMD	İMALAT / ANA METAL SANAYİ
KOÇ HOLDİNG A.Ş.	KCHOL	MALİ KURULUŞLAR / HOLDİNGLER VE YATIRIM ŞİRKETLERİ
KOZAALTIN İŞLETMELERİ A.Ş.	KOZAL	MADENCİLİK VE TAŞ OCAKÇILIĞI / METAL CEVHERİ MADENCİLİĞİ
KOZAANADOLU METAL MADENCİLİK İŞLETMELERİ A.Ş.	KOZAA	MADENCİLİK VE TAŞ OCAKÇILIĞI / METAL CEVHERİ MADENCİLİĞİ

ODAŞ ELEKTRİK ÜRETİM SANAYİ TİCARET A.Ş.	ODAS	ELEKTRİK GAZ VE SU / ELEKTRİK GAZ VE BUHAR
PEGASUS HAVA TAŞIMACILIĞI A.Ş.	PGSUS	ULAŞTIRMA VE DEPOLAMA
PETKİM PETROKİMYA HOLDİNG A.Ş.	PETKM	İMALAT / KİMYA İLAÇ PETROL LASTİK VE PLASTİK ÜRÜNLER
SASA POLYESTER SANAYİ A.Ş.	SASA	İMALAT / KİMYA İLAÇ PETROL LASTİK VE PLASTİK ÜRÜNLER
TAV HAVALİMANLARI HOLDİNG A.Ş.	TAVHL	MALİ KURULUŞLAR / HOLDİNGLER VE YATIRIM ŞİRKETLERİ
TOFAŞ TÜRK OTOMOBİL FABRİKASI A.Ş.	TOASO	İMALAT / METAL EŞYA MAKİNE ELEKTRİKLİ CİHAZLAR VE ULAŞIM ARAÇLARI
TURKCELL İLETİŞİM HİZMETLERİ A.Ş.	TCELL	BİLGİ VE İLETİŞİM / TELEKOMÜNİKASYON
TÜRPAŞ-TÜRKİYE PETROL RAFİNERİLERİ A.Ş.	TUPRS	İMALAT / KİMYA İLAÇ PETROL LASTİK VE PLASTİK ÜRÜNLER
TÜRK HAVA YOLLARI A.O	THYAO	ULAŞTIRMA VE DEPOLAMA
TÜRKİYE GARANTİ BANKASI A.Ş.	GARAN	MALİ KURULUŞLAR BANKALAR
TÜRKİYE İŞ BANKASI A.Ş.	ISCTR	MALİ KURULUŞLAR / BANKALAR
TÜRKİYE ŞİŞE VE CAM FABRİKALARI A.Ş.	SISE	MALİ KURULUŞLAR / HOLDİNGLER VE YATIRIM ŞİRKETLERİ
YAPI KREDİ BANKASI A.Ş.	YKBNK	MALİ KURULUŞLAR BANKALAR