

**KOCAELİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**ENDÜSTRİ MÜHENDİSLİĞİ
ANABİLİM DALI**

YÜKSEK LİSANS TEZİ

**VERİ MADENCİLİĞİ VE YAPAY SİNİR AĞLARI YÖNTEMLERİ
İLE HAZIR GİYİM SEKTÖRÜNDE TALEP TAHMİNİ**

ESRA ÇETİN

KOCAELİ 2023

KOCAELİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ENDÜSTRİ MÜHENDİSLİĞİ
ANABİLİM DALI

YÜKSEK LİSANS TEZİ

VERİ MADENCİLİĞİ VE YAPAY SİNİR AĞLARI YÖNTEMLERİ
İLE HAZIR GİYİM SEKTÖRÜNDE TALEP TAHMİNİ

ESRA ÇETİN

Dr. Öğr. Üyesi Atakan ALKAN
Danışman, Kocaeli Üniversitesi

Doç.Dr. Gülşen AKMAN
Jüri Üyesi, Kocaeli Üniversitesi

Dr. Öğr. Üyesi İrem DÜZDAR ARGUN
Jüri Üyesi, Düzce Üniversitesi

Tezin Savunulduğu Tarih: 17.01.2023

ETİK BEYAN VE ARAŞTIRMA FONU DESTEĞİ

Kocaeli Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez/proje çalışmada,

- Bu tezin/projenin bana ait, özgün bir çalışma olduğunu,
- Çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı,
- Bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi,
- Bu çalışmanın Kocaeli Üniversitesi'nin abone olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü'nün belirlemiş olduğu ölçütlere uygun olduğunu,
- Kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- Tezin/Projenin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez/proje çalışması olarak sunmadığımı,

beyan ederim.

Bu tez/proje çalışmasının herhangi bir aşaması hiçbir kurum/kuruluş tarafından maddi/alt yapı desteği ile desteklenmemiştir.

Bu tez/proje çalışması kapsamında üretilen veri ve bilgiler tarafından no'lu proje kapsamında maddi/alt yapı desteği alınarak gerçekleştirilmiştir.

Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

Esra ÇETİN

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI

Fen Bilimleri Enstitüsü tarafından onaylanan lisansüstü tezimin/projemin tamamını veya herhangi bir kısmını, basılı ve elektronik formatta arşivleme ve aşağıda belirtilen koşullarla kullanıma açma izninin Kocaeli Üniversitesi'ne verdiğimi beyan ederim. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin/projemin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanımı bana ait olacaktır.

Tezin/projenin kendi özgün çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin/projenin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim kurulu tarafından yayımlanan **“Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge”** kapsamında tezim aşağıda belirtilen koşullar haricinde YÖK Ulusal Tez Merkezi/ Kocaeli Üniversitesi Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü yönetim kurulu kararı ile tezimin/projemin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir.
- Enstitü yönetim kurulu gerekçeli kararı ile tezimin/projemin erişime açılması mezuniyet tarihinden itibaren 6 ay ertelenmiştir.
- Tezim/projem ile ilgili gizlilik kararı verilmemiştir.

ESRA ÇETİN

ÖNSÖZ VE TEŞEKKÜR

Talep tahmini, günümüzün deęişen zorlu rekabet koşullarında, gelecek dönem kapsamında daha karlı ve bilinçli yatırımlar ortaya çıkması adına şirketler için daha önemli bir basamak haline gelmiştir. Bu çalışma şirketler için yakın gelecekte çok daha önemli hale gelen, satışa etki eden anomalilerin analiz edilmesi, satışı etkileyen geçmiş dönemli verilerle birlikte anomalilerin tahminin yapılabirliğinin kontrolü ve anomalilerin satış tahminindeki etkisini veri madencilięi ve yapay sinir aęları ile inceleyerek literatüre katkı sağlamak amacıyla yapılmıştır.

Lisans ve Yüksek lisans öğrenimimde göstermiş olduęu tüm katkı ve destekleri için çok değerli hocam Dr. Öğr. Üyesi Atakan Alkan'a teşekkür ederim. Çalışmalarım süresince beni destekleyen, aldığım kararlarda her zaman yanımda olan, biricik aileme maddi ve manevi tüm emekleri için teşekkür ederim.

Ocak – 2023

Esra ÇETİN

İÇİNDEKİLER

ETİK BEYAN VE ARAŞTIRMA FONU DESTEĞİ.....	i
YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI.....	ii
ÖNSÖZ VE TEŞEKKÜR.....	iii
İÇİNDEKİLER.....	iv
ŞEKİLLER DİZİNİ.....	v
TABLolar DİZİNİ.....	vi
SİMGELER VE KISALTMALAR DİZİNİ.....	vii
ÖZET.....	viii
ABSTRACT.....	ix
1. GİRİŞ.....	1
2. VERİ MADENCİLİĞİ.....	3
2.1. Veri Ve Veri Tabanı Nedir?.....	4
2.2. Verilere Ulaşım Süreci.....	4
2.3. Veri Madenciliğinin Tarihçesi.....	5
2.4. Veri Madenciliğinin Diğer Bilimler İle İlişkisi Ve Kullanım Alanları.....	6
2.5. Veri Madenciliği Avantaj ve Dezavantajları.....	8
2.6. Veri Madenciliği Modelleri.....	9
2.6.1. Sınıflama ve Regresyon (Tahmin Edici).....	9
2.6.2. Kümeleme (Tanımlayıcı, Sınıflayıcı).....	9
2.6.3. Ardışık Zamanlı Örüntüler ve Birliktelik Kuralları (Tanımlayıcı).....	10
3. VERİ MADENCİLİĞİNDE TAHMİN EDİCİ MODELLER.....	11
3.1. Naive Bayes Algoritması.....	11
3.2. Destek Vektör Makineleri.....	13
3.3. K-En Yakın Komşu (K-NN) Algoritması.....	16
4. YAPAY SİNİR AĞLARI.....	18
4.1. Yapay Sinir Ağları Tanımı.....	18
4.2. Yapay Sinir Ağları Tarihçesi Ve Kullanım alanları.....	20
4.3. Yapay Sinir Ağları Avantaj Dezavantajları.....	21
4.4. Yapay Sinir Ağlarında Öğrenme.....	23
4.5. Yapay Sinir ağlarının Sınıflandırılması.....	25
4.5.1. Bağlantı Şekillerine Göre Ysa.....	25
4.5.2. Öğrenme Biçimlerine Göre Ysa.....	26
5. LİTERATÜRDE VERİ MADENCİLİĞİ VE YAPAY SİNİR AĞLARI.....	29
6. UYGULAMA.....	36
6.1. Anomali Tahmini.....	39
6.1.1. Weka’da Naive Bayes Algoritması ile Anomali Tahmini.....	46
6.1.2. Weka’da Destek Vektör Makineleri ile Anomali Tahmini.....	48
6.1.3. Weka’da K-En Yakın Komşu Algoritması ile Anomali Tahmini.....	49
6.2. Yapay Sinir Ağları İle Satış Tahmini.....	51
7. SONUÇ VE ÖNERİLER.....	57
KAYNAKLAR.....	60
KİŞİSEL YAYIN VE ESERLER.....	68
ÖZGEÇMİŞ.....	69

ŞEKİLLER DİZİNİ

Şekil 2.1.	Bilgi Keşfi Süreci	5
Şekil 2.2.	Veri Madenciliği Tarihi Süreci	6
Şekil 2.3.	Disiplinler Arası Veri Madenciliği	7
Şekil 3.1.	DVM'de Sınır Kavramları	14
Şekil 3.2.	DVM'in İşleyişi Modeli	16
Şekil 3.3.	K-NN Algoritma Yapısı	17
Şekil 4.1.	Biyolojik Nöron Yapısı	18
Şekil 4.2.	Yapay Sinir Ağı Hücresi	19
Şekil 4.3.	YSA'ların Türleri	25
Şekil 4.4.	İleri Beslemeli Ysa	26
Şekil 4.5.	Geri Beslemeli Ysa.....	26
Şekil 4.6.	YSA' da Danışmalı Öğrenme.....	27
Şekil 4.7.	Ysa'da Danışmasız Öğrenme.....	28
Şekil 4.8.	Ysa'da Takviyeli Öğrenme	28
Şekil 6.1.	Weka'da ARFF Dosya Görünümü.....	39
Şekil 6.2.	Stok Miktarının Weka'da Gösterimi.....	40
Şekil 6.3.	Fiyat Miktarının Weka'da Gösterimi.....	41
Şekil 6.4.	Satış Miktarının Weka'da Gösterimi	42
Şekil 6.5.	İndirim Oranının Weka'da Gösterimi	43
Şekil 6.6.	Anomalilerin Weka'da Gösterimi.....	44
Şekil 6.7.	Naive Bayes Percentage Split WEKA Sonuç Matrisi	47
Şekil 6.8.	DVM'de Percentage Split (%90-10) WEKA Sonuç Matrisi	49
Şekil 6.9.	KNN'de Percentage Split (%90-10) WEKA Sonuç Matrisi	51
Şekil 6.10.	YSA Katman Yapısı (Anomali Dahil)	52
Şekil 6.11.	Matlap'ta YSA Arayüzü Ve Eğitim Sonuçları	53
Şekil 6.12.	Eğitim, Validasyon ve Test Verileri Eğitim Performansı	53
Şekil 6.13.	Regresyon Sonuç Grafikleri (Anomali Dahil).....	54
Şekil 6.14.	YSA Katman Yapısı (Anomali Hariç)	54
Şekil 6.15.	Anomali Hariç YSA Arayüzü Ve Eğitim Sonuçları	55
Şekil 6.16.	Eğitim, Validasyon ve Test Verileri Eğitim Performansı	55
Şekil 6.17.	Regresyon Sonuç Grafikleri (Anomali Hariç).....	56

TABLolar DİZİNİ

Tablo 2.1.	VM Kullanım Alanları	8
Tablo 6.1.	Veri Setinden Örnek Gösterim	38
Tablo 6.2.	Tüketici Güven Endeksi	44
Tablo 6.3.	Perakende(Tekstil-giyim) Satış Hacim Endeksi Tablosu.....	45
Tablo 6.4.	Naive Bayes Cross Validation Analiz Sonuçları	46
Tablo 6.5.	Naive Bayes Percentage Split Analiz Sonuçları	47
Tablo 6.6.	DVM'de Cross Validation Analiz Sonuçları	48
Tablo 6.7.	DVM'de Percentage Split Analiz Sonuçları	48
Tablo 6.8.	KNN'de Cross Validation Analiz Sonuçları.....	49
Tablo 6.9.	KNN'de Percentage Split Analiz Sonuçları.....	50



SİMGELER VE KISALTMALAR DİZİNİ

Kısaltmalar

ANN	:Artificial Neural Network (Yapay Sinir Ağları)
ARFF	:Attribute Relationship File Format (Nitelik İlişkisi Dosya Biçimi)
DVM	:Destek Vektör Makineleri
KNN	:K-Nearest Neighbors (K-En Yakın Komşu)
MSE	:Ortalama Karesel Hata
PISA	:Uluslararası Öğrenci Değerlendirme Programı
RFM	:Recency, Frequency, Monetary Value (Yenilik, Sıklık, Parasal Değer)
RO	:Rastgele Orman
SS	:Standart Sapma
TUIK	:Türkiye İstatistik Kurumu
TÜFE	:Tüketici Fiyat Endeksi
VM	:Veri Madenciliği
WEKA	:Waikato Environment for Knowledge Analysis (Bilgi Analizi için Waikato Ortamı)
YSA	:Yapay Sinir Ağları

VERİ MADENCİLİĞİ VE YAPAY SİNİR AĞLARI YÖNTEMLERİ İLE HAZIR GİYİM SEKTÖRÜNDE TALEP TAHMİNİ

ÖZET

Günümüz güçlü rekabet koşullarında şirketlerin varlıklarını güvenle sürdürebilmeleri için gelecek tahminleri kritik önem taşıyan bir faktördür. Bu doğrultuda talep tahmin çalışmaları şirketlerin gelecekte güvenilir yol almalarına önemli katkı sağlamaktadır. Şirketler varoluş süreçlerini etkileyecek bu yolda birçok yöntem kullanmaktadır. Literatürde de sıklıkla karşılaşılan yapay sinir ağları ve veri madenciliği yöntemleri bu yöntemlerden bazılarıdır.

Çalışma kapsamında YSA ve veri madenciliği yöntemleri açıklanarak perakende sektöründe lider bir firmanın 4 yıllık veri seti kullanılarak satış tahmin çalışması için stok miktarı, fiyat, indirim oranı, hava koşulları, tüketici güven endeksi ve satışa etki eden anomaliler incellenmiş olup tüm bu veriler ışığında MATLAB ve WEKA programında tahminleme çalışmaları yapılmıştır. İlk olarak satışa etki eden anomalilerin (Range, fiyat, stok ve pandemi) YSA ile yapılacak olan talep tahmin çalışmasına girdi sağlayabilmesi için K-En yakın Komşu, Naive Bayes ve Destek vektör makineleri yöntemleri ile tahmin çalışması yapılmıştır. DVM ile yapılmış olan çözüm %78 doğruluk oranı ile diğer yöntemlere göre daha doğru bir tahmin sonucuna ulaşmıştır. Uygulamanın ikinci bölümünde bu veriler kullanılarak anomaliler dahil ve hariç olmak üzere iki şekilde de YSA ile talep tahmin çalışması yapılmıştır. Anomaliler dahil olarak yapılan YSA ile tahmin çalışması anomali hariç olan tahmin çalışmasından %4 daha doğru tahmin sonucuna ulaşılmıştır. Yapılan bu çalışma ile YSA ile satış tahmini uygulamalarında anomali değişkenlerinin önemini ve sınıflandırma tahmin algoritmalarından da destek vektör makinelerinin diğer algoritmalara göre daha başarılı sonuçlar elde ettiği gösterilmiş olup literatüre ve şirket maliyetlerini azaltma yönünde katkı sağlanmıştır.

Anahtar Kelimeler : Talep Tahmini, Veri Madenciliği, Yapay Sinir Ağları (YSA).

DEMAND FORECASTING IN THE APPAREL INDUSTRY WITH DATA MINING METHODS AND ARTIFICIAL NEURAL NETWORKS

ABSTRACT

In today's fierce competitive conditions, future forecasts are a critical factor for companies to continue their existence with confidence. In this direction, demand forecasting studies make a significant contribution to companies to take a reliable path in the future. Companies use many methods in this way that will affect their existence processes. Artificial neural networks and data mining methods, which are frequently encountered in the literature, are some of these methods.

Within the scope of the study, ANN and data mining methods were explained, and the stock amount, price, discount rate, weather conditions, consumer confidence index and anomalies affecting the sales were examined for sales forecasting using a 4-year data set of a leading company in the retail sector, and in the light of all these data, MATLAB and WEKA program prediction studies have been made. First of all, estimation study was carried out with K-Nearest Neighbor, Naive Bayes and Support vector machine methods so that anomalies (Range, price, stock and pandemic) affecting sales can provide input for demand estimation work to be done with ANN. The solution made with SVM has reached a more accurate estimation result with an accuracy rate of 78% compared to other methods. In the second part of the application, using these data, demand estimation studies were carried out with ANN in two ways, including and excluding anomalies. The estimation study with ANN including anomalies was 4% more accurate than the estimation study excluding anomalies. With this study, it has been shown that the importance of anomaly variables in sales forecasting applications, artificial neural networks and classification estimation algorithms, and that support vector machines achieve more successful results than other algorithms, contributing to the literature and reducing company costs.

Keywords : Demand Forecast, Data Mining, Artificial Neural Networks (ANN).

1. GİRİŞ

Gelişen dijital çağ ile beraber büyük ölçekli verilerin depolanması ve erişimi artık çok daha kolay hale gelmiştir. Ancak işletmeler, üzerinde varlıklarını devam ettirebilmek ve gelişimlerini sürdürebilmek için müthiş öneme sahip bu denli yığın verilerin içerisinden faydalı veriyi elde ederek analiz etmek ise gittikçe zorlaşmaktadır. Bu durumda bu yığın verilerin analizi ve bu analizlerin sonuçları arasındaki veriler arasındaki saklı olan bilgileri ortaya çıkararak, yapıcı, kullanışlı ve kıymetli bilgileri sunmak için ortaya çıkan veri madenciliği kavramı çok daha önemli bir hal almıştır. Veri madenciliği; geniş kapsamlı bir veri setinden, çözülmesi zor, yeni ve kıymetli olan bir bilgi sürecidir. (Kantardzic, 2011)

Veri madenciliği, analiz ve bilgiyi elde edinme işlemlerini gerçekleştirirken farklı bilim dallarından da yararlanmaktadır. Makine öğrenmesi, istatistik ve matematik gibi bilimlerden yararlanan disiplinler arası bir alandır. Veri madenciliği, elde edilmiş olan veri setinden anlamlı çıkarımlarda bulunmak, veri içeriğindeki saklanmış olan bir takım dizgeleri ve yönelimleri ortaya çıkarmak ve değişenler arasındaki ilişkileri tespit ederek karar vermeyi kolaylaştırmak amacı ile uygulanan bir yöntemdir. Veri madenciliğinde birçok farklı disiplinler arası alanlarda kullanılan çok sayıda teknik ve yöntem geliştirilmiştir. (Rygielski ve ark., 2002). Verileri ulaşım ve analizinin yanı sıra bu verileri doğru alanlarda kullanmak şirketlerin varlıklarını sürdürebilmesi için hayati önem taşımaktadır. Perakende sektöründe faaliyet gösteren, müşterilerin giyim alışkanlıklarının moda ve çevresel etkilerle çok sık değişikliğe uğradığı bir firmada talep tahmini stratejik ve operasyonel planlar üzerinde ciddi etkiye sahiptir.

Talebin doğru tahmin edilmesi, geleceğe yönelik insan, malzeme, makine – ekipman ve tesis alanı gibi üretim sektörünün temel kaynaklarının doğru planlanmasını sağlamaktadır. Aksi durumda ise kaynakların verimsiz kullanıldığı ve ciddi maliyetlerin ortaya çıktığı durumlar ile karşılaşabilmektedir. Talep tahmini konusu her sektörde karışık ve zor problemlerin başında gelmektedir. Bunun sebebi gerçek hayat problemlerinin, birçok girdisinin olması ve bu girdilerinde tahmin edilebilirliğinin düşük olmasıdır. Bu sebep ile ciddi avantajları olmasına ve çokça problemde etkin sonuç alınmasına rağmen, geleneksel yöntemler ile her zaman doğru tahminler yapılamamaktadır. (Arslan, 2020) Bu çalışmalarda daha doğru veri ve tahmin sonuçları

için veri madenciliğinin yanında yapay zeka teknikleri kullanılması çalışmaları daha da güçlendirmektedir. Yapay zeka tekniklerinden biri olan literatürde de sıklıkla kullanılan yapay sinir ağları talep tahmin çalışmaları için vazgeçilmez bir yöntemdir.

Yapay sinir ağları, nesnel yöntemlere göre istatistiki veri analizlerini kullanmasıyla, öznel yöntemlere ise arasındaki bağlantı ve derecelerini öğrenmesi sebebiyle benzerlik göstermektedir ve son yıllarda etkin ve verimli bir tahminleme aracı olarak klasik olmayan yöntemler arasında yer almaktadır.(Adıyaman, 2007)

Bu tez çalışmasında ise Türkiye’de perakende sektöründe lider bir şirketin veri setinden yola çıkılarak talep tahmin çalışması yapılmaya çalışılmıştır.

Çalışmanın birinci bölümünde; Veri madenciliği ile ilgili kavramlar ve yöntemleri hakkında bilgi verilmiştir. İkinci ve üçüncü bölümünde araştırma için kullanılan yapay sinir ağları ve veri madenciliği tahmin yöntemlerinden olan Naive Bayes, Destek vektör makineleri, K-En yakın komşu yöntemleri hakkında detaylı bilgi verilerek, dördüncü bölümde ilgili alanda yapılan literatür çalışmalarına yer verilerek araştırmalardan bahsedilmektedir. Son olarak beşinci bölümünde ise; talep tahmini uygulaması kapsamında yapılan tüm çalışmalar ele alınmıştır. Yapay sinir ağları yöntemi ile diğer talep tahmin yöntemlerinin kullanılarak tahminlenen anomali değerleri dahil edilerek ve anomalilerin dahil olmadığı talep tahmin çalışmaları yapılarak sonuçlar karşılaştırılmıştır.

Bu tez çalışmasının amacı şirket yönetiminin stratejik kararlar almasında ve mevcut durum analizlerinde kullanabileceği etkin bir talep tahmin yöntemi ile optimum karar destek yönteminin oluşturulması ve halihazırda talep tahmini için kullanmış olduğu girdilerin analizi yapılarak daha doğru veri toplama çalışmasının yapılmasının sağlanması ve daha önce talep tahmini uygulamalarında sıklıkla kullanılan veri madenciliği ve yapay sinir ağları yöntemlerinin talep tahmini yöntemlerinden önemli bir yöntem olduğu gösterilmeye çalışılarak literatüre katkı sağlamaktır.

2. VERİ MADENCİLİĞİ

Gelişen teknoloji ile beraber günümüzde hayatın birçok alanında bilişim sisteminin kullanıldığı, bilgisayar ve buna bağlı birbirine entegre akıllı cihazlar etkin bir şekilde yer almaktadır. Sürekli olarak gelişmekte olan bu cihazlar ile birlikte hayatımızın her noktasında birden çok veri dahil olmuştur. Ancak bu yeni sistemler ile beraber işletmeler, alışveriş kaynakları, sağlık alanından tutup da ev yaşantımıza kadar tüm alanlarda bu verileri saklamak ve güvenli bir şekilde depolanması gittikçe zorlaşmaktadır. Bu verilerin çoğalmasıyla beraber içeriğindeki anlamlı ve işe yarar bilgilerin kullanılabilirliği ve ortaya çıkarılması da zorlaştığı için günümüzdeki tüm işletmeler bu süreçlerde kendilerini karmaşık durumlarda bulabilmektedirler. Sektörlerdeki canlılığa aynı zamanda piyasa şartları ve yakın zamanda yaşadığımız pandemi gibi olağanüstü zor koşulların da etkisi ile verilerin normalleşmiş versiyonlarına ulaşmak çok daha zorlaşmıştır.

Bu ihtiyaçlar sebebiyle yazılım çeşitleri zaman içerisinde gelişmiş ve gelişmeye de devam etmektedir. Gelişen yazılımlar ile ortaya çıkan veri tabanları içerisinde anlamlı analizler ile doğru bilginin ortaya çıkması ve kullanılabilmesi için bilgiye ulaşım kavramı önemlidir. Tam da bu noktada veri madenciliği yönteminin önemi ortaya çıkmaktadır. Veri madenciliği, büyük hacimli data setlerindeki var olan karmaşık ilişkileri analiz ederek yeni yöntemlerle data sahiplerine daha anlamlı ve verimli olarak faydalanabilecekleri bir veri özetleme yaklaşımıdır. (Hand ve diğerleri, 2001)

VM'nin amacı verilerdeki anlamlı ilişkileri çözerek elde edilen sonuçları harekete geçirecek kararlar için kullanmaktır. Buradaki önemli nokta, değişkenlerin içerdiği ilişkilerden ziyade, geleceğe yönelik etkin tahmin ve karar sistemi kurulması işlemidir. Bu anlamda veri madenciliği, temel bilginin bulunması yolunda bir kara kutu keşfetme yöntemi olarak kabul edilmektedir.

Bu sebeple VM, sadece keşfedilen veri analizi yöntemlerini değil, daha kabul edilebilir öngörülerde bulunmak ve tahmin edilen değişkenler arasındaki bağlantıların elde edilmesi olanaklı olduğu için aynı zamanda sinir ağı tekniklerinde de yararlanmaktadır. (Taşdemir, 2012).

2.1. Veri Ve Veri Tabanı Nedir?

Veri, işlenmemiş ham gerçeklerdir. İşlenmiş ve düzenlenmiş veri bilgiye dönüşmektedir. Veri işleme, kök veriyi düzenleyebilmek ve bu düzenlenmiş verilerden faydalanarak dizgeler, karmaşık öngörüler ya da istatistiksel modellerden faydalanarak sonuçlar ortaya koymaktır. Bilgi, asıl anlamda kaynağa ihtiyaç duymaktadır. Doğru tahmin ve karar vermenin kilit taşları kesin, birbiri ile ilişkili ve doğru zamanda kullanılan bilgidir. (Özdemir, 2010)

Bilgi, belirli bir hedef ışığında işlenmesi ile elde edilen işlenmiş ham verilere denir. Bilgiye ulaşabilmek için, toplanmış olan tüm veriler analiz edilerek anlaşılır bir hale dönüştürülmesine ihtiyaç vardır. Veri analizi ise verinin bilgi haline getirilmesi sürecidir. (Alpaydın 2018)

Verilere ulaşım, analizler doğrultusunda doğru bilgiye ulaşım için büyük önem taşırsa da tüm bu toplanan verilerin kayıplar yaşamadan güvenilir ve en doğru şekli ile gizlenmesi de en az verilere ulaşım kadar önem arz etmektedir.

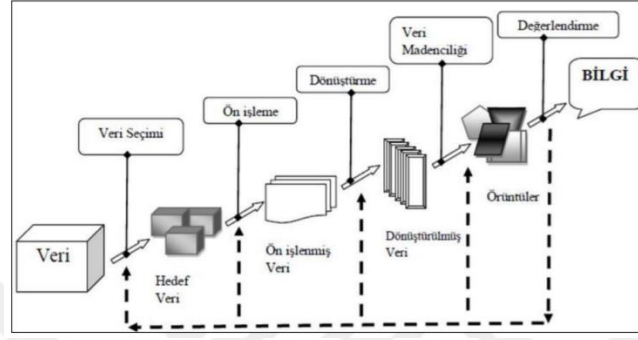
Veri tabanı, ilişkisel verilerin, kullanım yerine göre elverişli olarak düzenlenmiş olduğu, kontrol edilebilir bilgi depolarına verilen addır. (Erken, 2017) Veri tabanı yalnızca verilerin toplandığı alan olmamakla birlikte birbiri ile ilgili verileri amacına uygun olarak düzenli ve sıralı bir şekilde de bulundurarak ihtiyaç duyulduğunda bilgi kayıpları yaşamadan kolay erişim imkanı sunmaktadır.

2.2. Verilere Ulaşım Süreci

Birçok çalışmacı veri madenciliğini doğru bilgiye ulaşma sürecindeki bir basamak olarak ele almaktadır. Ancak bilgi keşfi süreci aşağıdaki maddelerin tekrarlamalı bir dizilimi ile oluşur: veri tabanından ilişkili verilerin alınmasıdır. Şekil 2.1’de de görüldüğü üzere bilgi keşfi süreci;

1. Veri seçimi: veri tabanından ilişkili verilerin alınmasıdır.
2. Veri temizleme: Karmaşık ve tutarsız veriler yok edilir. Kayıp veri bölümleri için strateji belirlenir.
3. Veri dönüşümü: Veriler bu adımda uygun olan şekillere birleştirilir ya da dönüştürülür. Kullanılan değişken sayısı bu aşamada azaltılabilir.

4. Veri madenciliği: Uygulamanın amacı ile uyumlu olacak şekilde veri madenciliği tekniği belirlenir. Tahmin edici, sınıflayıcı gibi yöntemler uygulanır.
5. Desen değerlendirme: Veri madenciliği ile elde edilen sonuçlar değerlendirilir.
6. Bilgi sunumu: Anlaşılan bilgi, bir sonraki çalışma için raporlanır. Bu hali ile önceden yapılmış olan çalışmalar ile karşılaştırılarak, doğruluğu analiz edilebilir. (Güner, 2018)



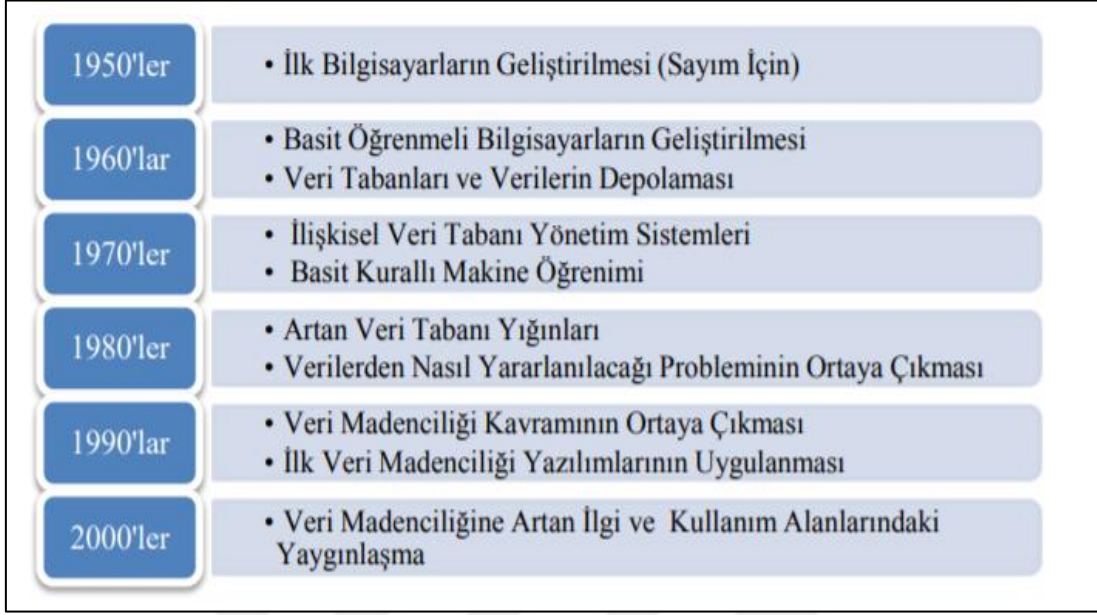
Şekil 2.1. Bilgi Keşfi Süreci (Savaş ve diğerleri, 2012)

Geçmişten günümüze kadar faydalı deseni bulmak ve anlamak kavramı; bilgi hasadı, bilginin keşfedilmesi gibi çeşitli adlar ile tanımlanmıştır. Veri madenciliğini isim olarak çoğunlukla matematikçiler, analistler, ve yönetim bilgi sistemlerini kullanan toplumlar kullanmayı tercih etmekteydi. Şekilde de görüldüğü üzere veriler içerisindeki faydalı bilgiyi elde etmek için gerekli olan işlemlerin hepsi bilgi keşfi olarak tanımlanmaktadır. Veri madenciliği ise bu işlemin özel bir adımıdır. Ayrıca veri madenciliği belirli bir algoritma kullanılarak veri setinden desen çıkarma işlemidir. (Sebik, 2018)

2.3. Veri Madenciliğinin Tarihçesi

Veriler, 1950'li yıllar itibari ile ilk bilgisayarlar sayımlar amacıyla kullanılmaya başlanmasıyla gelişmeye başlamıştır. Veri tabanı yönetim sistemleri ise 1980'lerde kullanımı yayılmış ve bilimsel alanlarda kullanılmaya başlamıştır. Bilginin keşfi çalışmaları için grup toplantısı 1989'da gerçekleşmiştir. Ardından 1991 yılında "Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop" başlıklı makalesinin yayınlanmasıyla bilgi keşfi ve VM ile ilgili ilk tasvir ve görüşleri gün yüzüne çıkarması ile bu akış daha da hızlanmış ve sonunda da 1992'de VM için ilk yazılım geliştirilmiştir. (Savaş vd., 2012) 2000'lerde veri madenciliği devamlı bir şekilde gelişmiş ve neredeyse tüm bilimsel alanlarda kullanılmaya başlamıştır. Elde edilen sonuçların

faydaları fark edildikçe, veri madenciliği sürecine ilgi artmıştır (Savaş vd., 2012) Aşağıdaki şekil 2.2’ de veri madenciliğinin yıllara göre süreci yer almaktadır;



Şekil 2.2. Veri Madenciliği Tarihi Süreci (Armutlu, 2018)

2.4. Veri Madenciliğinin Diğer Bilimler İle İlişkisi Ve Kullanım Alanları

Veri madenciliği çok yönlü oluşu ve veri içeren tüm alanlarda kullanılabilmesi ile disiplinler arası bir yöntemdir. Veri madenciliğinin temelini oluşturan bu disiplinler, istatistik, yapay zeka, veri tabanı teknolojisi (Demiral vd., 2017), makine öğrenmesi, görselleştirme, bilgi bilimi ve diğer disiplinlerdir (Alagöz vd., 2014).

Özellikle makine öğrenimi, veri tabanları ve istatistik veri madenciliğini en çok besleyen bilim alanlarındandır. Hatta aynı hedefe odaklanıldığında istatistik ve veri madenciliği aynı disiplin olarak kabul edilebilmektedir. Ancak en önemli fark istatistik daha çok teoriye dayalı ve bir hipotez üzerinden yol alırken veri madenciliği sezgisel algoritmalar ile yürütülür. Veri madenciliğinde bir hipotez kurulmaz ve nereye ulaşılacağı çalışma bittiğinde yani sonuç verdiğinde bulunmuş olur. (Han, 2012 ve Gencer ,2019)

Aşağıdaki Şekil 2.3’de veri madenciliğinin diğer disiplinler ile ilişkisi gösterilmiştir. Birçok farklı bilimin merkezinde olması ile beraber birbirleri arasında süreklilik olan bir bilgi akışı olmaktadır. Bu sayede de veri madenciliği gelişmeye devam etmekte ve farklı daha da farklı disiplinlerle beraber kullanılarak günümüzde vazgeçilmez bir yöntemdir.



Şekil 2.3. Disiplinler Arası Veri Madenciliği (Savaş vd., 2012)

Son yıllarda yapılan çalışmalarda artan rekabet koşulları, gelişen bilim ve teknoloji ile beraber artan veri kaynağı sebebiyle birçok farklı yöntem ve bu yöntemlerin hibrit kullanımı yaygınlaşmıştır. Veri madenciliği yöntemleri de içerdiği birçok bilim ve değişen teknolojiye başarılı bir şekilde entegre edilmesi, veriyi ayıklama ve anlamlı analizler çıkartılabilmesi sebebiyle bu yöntemlerin başında gelmekte ve birçok alanda kullanılmaktadır.

Veri madenciliği kurumsal hayattan, günlük ve bilimsel yaşama kadar her alanda en küçük kısımlar dahi dahil olmak üzere kullanılabilmekte ve doğru sonuçlara ulaşılabilir. Bu alanlarda bazıları;

- Pazar Yönetimi, İnternet Ve Güvenlik,
- Sigortacılık, Borsa, Bankacılık,
- Spor, Eğitim, Perakendecilik, Endüstri
- Genetik, Biyoloji, Tıp
- Ürün Geliştirme ve Test Sonuçlarının Analizleri (Şık, 2014).

Veri madenciliği çalışmalarını Gorunescu yapmış olduğu çalışmada aşağıdaki Tablo 2.1 'de özetlemiştir. (Gorunescu, 2011)

Tablo 2.1. VM Kullanım Alanları (Gorunescu, 2011)

Kullanım Alanları	Kullanım Oranları (%)
Müşteri Analitiği	32.8
Bankacılık	24.4
Direkt Pazarlama	16.1
Kredi Puanlama	15.6
Telekominikasyon	14.4
Dolandırıcılık Tespiti	13.9
Satın Alma	11.7
Sağlık	11.7
Finans	11,1
Bilim	10.6
Reklamcılık	10.6
E-Ticaret	10
Sigortacılık	10
Web Madenciliği	8.3
Sosyal Ağlar	7.8
İlaç	7.8
Biyoteknoloji	7.8

2.5. Veri Madenciliği Avantaj ve Dezavantajları

Veri madenciliği, yöntemlerinin kullanımının kolay ve uyarlanabilir oluşuyla bu kadar yaygın olması özellikle son dönemlerde teknik olarak da başarılı sonuçlar vermesi kaynaklı akademik ve bilimsel çalışmalarda da sıklıkla yer almaktadır. Veri madenciliği yönteminin avantaj ve dezavantajlarını aşağıdaki şekilde değerlendirebiliriz. Veri madenciliği avantajları;

- VM yöntemleri büyük çaplı problemlerin çözümü için etkin bir yöntem olarak kullanılabilir.
- VM yöntemleri karışık durum içeren problemlerin çözümü için hızlı ve kolay uygulanabilir.
- VM yöntemlerinin hem sayısal hem de kategorik değişkenlerin çözümlerinde kullanılabilir.

Veri madenciliği dezavantajları ise şunlardır;

- VM yöntemlerinde 0-1 arasında giriş verisi olması gerekmektedir.
- VM yöntemleri ile elde edilen sonuçların açıklaması sunulamamaktadır.

- VM yöntemlerinin sonuçları her zaman en doğru sonuca ulaşabileceğinin garantisi yoktur. Elde edilen sonuçları farklı yöntemler ile tekrarlayarak en iyi sonuç seçilebilir. (Polat,2022)

2.6. Veri Madenciliği Modelleri

Veri madenciliği modellerini incelediğimizde temel olarak tahmin etme ve tanımlayıcı yöntemler olmak üzere ikiye bölünmektedir. Tahmin edici modeller, eldeki var olan gerçekleşmiş veri setinden faydalanılarak model hazırlanması ve bu model ile analizler yapılarak gelecekteki sonuçları belirli olmayan veri setleri için sonuçların tahmin edilmesi hedeflenmektedir. Tanımlayıcı modeller ise eldeki verilerin karar vermeye destek olması için içerisindeki örüntülerin ve veriler arasındaki ilişkilerin anlaşılmasının sağlanmaktadır. (Özekes, 2003)

2.6.1. Sınıflama ve Regresyon (Tahmin Edici)

Regresyon ve sınıflama, önemli veri gruplarını ortaya çıkaran ya da veri setindeki gelecek eğilimleri tahmin eden formları kurabilen veri analizi modelleridir. Kategorik veri setleri ile tahmin edici modeller oluşturulurken sınıflama kullanılırken, süreklilik içeren veri setlerinin tahmin edilmesinde ise regresyon kullanılmaktadır. Kullanılan başlıca tahmin edici modeller aşağıdaki gibidir; (Özekes, 2003).

- Karar Ağaçları
- Destek Vektör Makineleri
- Genetik Algoritmalar
- K-En Yakın Komşu
- Bellek Temelli Nedenleme
- Naive-Bayes
- Rastgele Orman

2.6.2. Kümeleme (Tanımlayıcı, Sınıflayıcı)

Kümeleme modellerinde ise karar vermeye yardımcı olacak mevcut verilerin örüntülerinin tanımlanması sağlanmaktadır. Kümeleme çalışmasında küme içerisindeki verilerin benzerlik oranı fazla, kümeler arasındaki benzerliği az olmalıdır. Tanımlayıcı

analizi bir veri setindeki bilgileri belirli yakınlık eşiklerine göre gruplara bölme işlemidir. Kullanılan başlıca yöntemler aşağıdaki gibidir;

- Bölümlemeli Yöntemler,
- Yoğunluk Temelli Modeller
- Hiyerarşik Yöntemler
- Grid (Izgara) Tabanlı Algoritmalar (Akpınar, 2000)

2.6.3. Ardışık Zamanlı Örüntüler ve Birliktelik Kuralları (Tanımlayıcı)

Birliktelik kuralı, geçmiş zamanlı eldeki veri setinin analiz edilip bu veriler içerisindeki davranışların incelenmesi ile tespit edilen birliktelik içeren davranışların, geleceğe yönelik çalışmalarda kullanılmasını destekleyen bir yöntemdir. Bu yöntem, veriler arasındaki ilişkiyi veya korelasyonları ortaya çıkarmakta kullanılmaktadır. Böylece veri setindeki saklı olan ilk etapta öngörülemeyen ilişkiler elde edilebilmektedir. Elde edilen ilişki bilgisi genellikle pazarlamasyon, satış ve karar verici analizlerin yanında yönetimsel araştırma alanlarında da kullanılabilir. (Akpınar, 2000)

Kullanılan başlıca yöntemler aşağıdaki gibidir;

- Apriori Algoritması,
- Carma Algoritması,
- Sequence Algoritması,
- GRI Algoritması,
- Eclat Algoritması,
- FP-Growth Algoritması (Akpınar, 2000)

3. VERİ MADENCİLİĞİNDE TAHMİN EDİCİ MODELLER

Tahmin, geçmişte yaşamış olduğumuz tecrübelerden ya da veri setlerinden çıkarılan sonuçların bilgilerin gelecekteki öngörüsünü elde edebilme olarak denilmektedir. Karar verme durumlarında tahmin edici yöntemler daha doğru sonuçlara ulaşabilmek için önemlidir. Tahmin etmeye yarayan bu yöntemlerden elde edilen verileri kullanarak bir model geliştirmesi amaçlanır. Ortaya çıkan bu model uygulanarak sonuçları belirsiz olan veri setleri içerisinde gerçek ve etkili sonuçlar bulması amaçlanmaktadır. (Polat,2022)

Perakende sektöründe yapılmış olan satış tahmini uygulama çalışması için kullanılan veri madenciliği tekniklerinden tahmin edici modeller aşağıdaki gibidir. Tezin bu bölümünde tahmin edici modellerden uygulamada kullanılan yöntemler detaylı olarak açıklanmıştır. Uygulamada kullanılan Yapay sinir ağları yöntemi ise ayrı bir başlık altında incelenecektir.

Uygulamada kullanılan yöntemler aşağıdaki gibidir;

- 3.1. Naive-Bayes
- 3.2. Destek Vektör Makineleri
- 3.3. K-En Yakın Komşu

3.1. Naive Bayes Algoritması

Naive Bayes sınıflandırıcıları, belli bir sınıftaki bir özellik değerinin etkisinin diğer özellik değerleri ile bağlı olmadığını varsaymaktadır. Bu varsayıma, sınıf koşullu bağımsızlık denir. Bu yaklaşım ilgili hesaplamaları çalışmalarda basitleştirdiği için bu anlamda "naif" olarak kabul edilmektedir. (Kalemci, 2018)

Naive Bayes modelinin çalışma şekli, aynı ölçütlerin sonuç üzerindeki hareketlerinin istatistiksel olarak hesaplanması temellidir. Bu ana sebep sayesinde birçok yazılım uygulamasında kullanılabilir. Naive Bayes modelinde sıralı data kullanımı istenmemektedir. Bu sebepler veri içerisindeki sıralı değeri olan bağımlı veya bağımsız değişkenlerin kategorik verilere dönüştürülmesine ihtiyaç vardır. Bu duruma örnek olarak veri setindeki bağımsız değişkenlerden birini yaş kabul edersek, kategorik verilere aşağıdaki gibi dönüştürebiliriz;

- yaş ≤ 25

- $26 \leq \text{yaş} \leq 35$
- $36 \leq \text{yaş} \leq 45$

Naive bayes yönteminde eğitim aşamasındayken, her sonucun eğitim kümesindeki tekrar sayısı hesaplanmaktadır. Ortaya çıkan sonuç öncelikli olasılık olarak tanımlanmaktadır.

Bütün olasılıklarının toplamda değeri ise 1'dir. Bu olasılıklar içerisindeki ortaya çıkan maksimum değere o çıktının sınıfını vermektedir. (Akbulut, 2006).

Teori aşağıdaki şekilde ifade edilebilir;

İlişkili olan X ve Y olaylarını ele alalım ; Y olayının yaşanabilme ihtimali X olayının yaşanabilmesine bağlı olmaktadır. Bu koşullu olasılık ise Denklem (3.1)'deki gibi belirtilmektedir;

$$P(Y|X) = \frac{P(X \cap Y)}{P(X)} \quad (3.1)$$

Denklem (3.1)'den yararlanılarak iki olayın kesişim olasılığına Denklem (3.2)'deki şekilde ulaşılmaktadır;

$$P(X \cap Y) = P(Y|X) \cdot P(X) \quad (3.2)$$

Denklem (3.2)'nin elde edilme mantığı ile X olayının olasılığı ve sonucunda kesişim olasılığı da Denklem (3.3) ve Denklem (3.4) 'deki gibi ortaya çıkmaktadır;

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)} \quad (3.3)$$

$$P(X \cap Y) = P(X|Y) \cdot P(Y) \quad (3.4)$$

Böylelikle Y'nin koşullu olasılığı Denklem (3.5) ile aşağıdaki şekilde hesaplanmaktadır;

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)} \quad (3.5)$$

Bayes yaklaşımı, İstatistikte önemli bir yere sahip olan, koşullu olasılıkların hesaplanması için uygulanan bir teoremdir. Bundan faydalanarak sınıflandırma çalışmaları uygulanabilmektedir. Teoriye göre koşullu olasılıkların hesaplanması aşağıdaki gibidir;

$P(X)$: X olayı meydana gelme olasılığı

$P(Y)$: Y olayı gerçekleşme olasılığı

$P(X|Y)$: Y olayından sonra X olayı gerçekleşme olasılığı

$P(Y|X)$: X olayından sonra Y olayı gerçekleşme olasılığı

X olayının önsel gerçekleşme ihtimali, tüm ayrık parçalarının birleşmiş halidir. Denklem (3.6)'deki gösterilmektedir; (Akkoç, 2012)

$$P(X) = P(X \cap Y) + P(X \cap \bar{Y}) \quad (3.6)$$

$$P(Y|X) = \frac{P(X \cap Y)}{P(X)} \quad (3.7)$$

Denklem (3.7)'deki $P(Y|X)$ eşitliği, koşullu olasılık denklemi kullanılarak açılımından sonra bayes formülü bulunur. Ortaya çıkan bayes formülüzasyonu, 2 ve 2'den fazla olay baz alındığında Denklem (3.8), (3.9) ve (3.10)'deki gibi ortaklaştırılabilir; (Erken, 2017)

$$P(Y|X) = \frac{P(X \cap Y)}{P(X \cap Y) + P(X \cap \bar{Y})} \quad (3.8)$$

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X|Y) \cdot P(Y) + P(X|\bar{Y}) \cdot P(\bar{Y})} \quad (3.9)$$

$$P(Y_i|X) = \frac{P(X|Y_i) \cdot P(Y_i)}{\sum_{j=1}^n P(X|Y_j) \cdot P(Y_j)} \quad (3.10)$$

Naive bayes sınıflandırma modeli, istatistiksel ve olasılıksal olarak çok önemli olmakla birlikte veri madenciliği çalışmalarında da etkin bir şekilde kullanılmaktadır. Bu kapsamda olan sıklıkla metin madenciliği şeklinde isimlendirilen çalışmalarda da oldukça etkin kullanım alanını mevcuttur. (Erken, 2017)

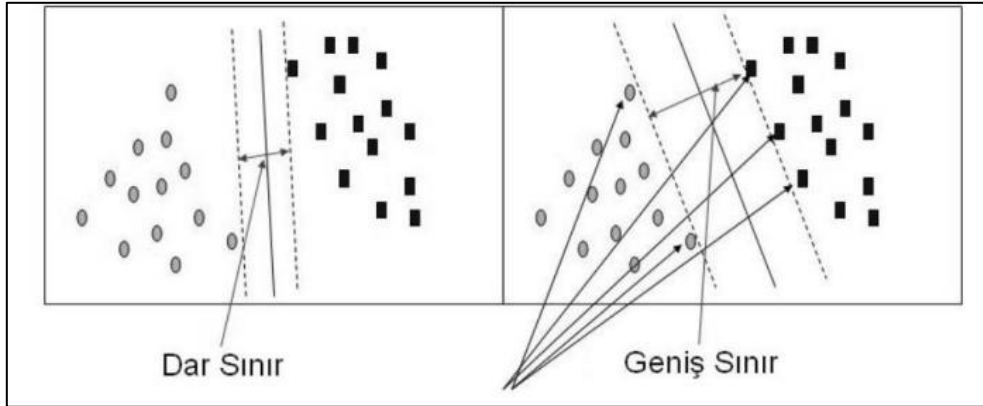
3.2. Destek Vektör Makineleri

Literatürde DVM yöntemi, sınıflandırma ve regresyon analizlerinin ikisinde de uygulanmasına rağmen genellikle sınıflandırma uygulamalarında daha sık tercih edilmektedir. DVM yöntemi denetimli öğrenme modelini temel almaktadır. Algoritma çalışma mantığında verilerin çeşitlerine bağlı olarak çekirdek fonksiyonlar da uygulanabilmektedir. Böylelikle doğrusal ve doğrusal olmayan sınıflandırma işlemlerinin

ikisini de işletebilmektedir. Eğer sınıflandırma uygulamasında, tam ayrıştırılabilen veri kullanımı olursa genel olarak tüm veriler bir hiper düzlemle sınıflandırılabilir. Fakat, eğer tam ayrıştırılmayan veriler kullanılırsa, çoğunlukla aynı boyutta tek bir düzlem ile sınıflandırılmamaktadır. Bu nedenle de farklı çekirdek fonksiyonları kullanılmaktadır. (Metlek & Kayaalp, 2020).

Destek vektör makineleri, kullanılmaya başlandığı ilk zamanlarda hiper bir düzlemi iki ayrı sınıfa ayırabilmek için geliştirilmiştir ancak kullanım amacını çok gruplu ve doğrusal olarak ayırımı gerçekleştirilemeyen örnekler için genişletilerek kullanımını alanını da genişletmiştir. Var olan sınıf sayısı iki sınıftan fazla olduğunda, datada kullanılan yöntem bire karşı bir (one versus one) ya da bire karşı hepsi (one versus all) şeklindedir. Böylelikle çoklu sınıflarda da veri setleri sınıflandırılabilir. (James ve diğerleri, 2013). DVM bu özelliği ile çok boyutlu veri setleri ile başa çıkabilmektedir bunun yanında esnek olması ve farklı alanlardaki problemlere de uyarlanabilmesi ile disiplinler arası çok sık kullanılan bir yöntemdir. (Kaşıkçı, 2019)

DVM analiz edilirken vurgulanması gereken önemli kavram, sınır farklı bir söyleyişle marj kavramıdır. DVM, geniş sınır kuralını yaratmaya çalışan bir yöntemdir. Aşağıdaki Şekil 3.1’de sınır gösterimleri geniş ve dar sınırlara örnek göstermektedir. (Kaya, 2016)



Şekil 3.1. DVM'de Sınır Kavramları (Erdal,2011)

DVM modelinin süreç adımları aşağıdaki gibidir;(Hazım, 2018):

1. Veri setinin okunması.
2. İşlem sınıfı, işlem zamanı ve ardışık işlemler içerisindeki farklılıklar kontrol edilerek verilerin yeniden sıralanması.

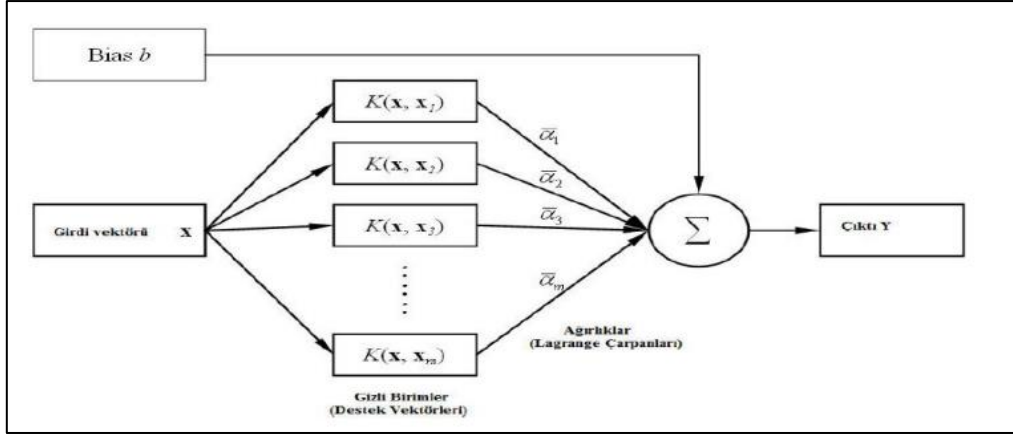
3. Her bir işlemin iki alanın vektörü olabilecek veri şekline dönüştürülmesi.
4. Pozitif ve negatif işlem grupları olarak tanımlanan iki ayrı veri seti oluşturulması
5. Çekirdek fonksiyonunun seçilmesi.
6. DVM'nin eğitilmesi.
7. Uygulama sonrasında sınıflandırmanın performansının kaydedilmesi.
8. Mevcuttaki işlemlerin değerlendirilmesi.
9. Uygulama geçerli işlem verileri için 1'den 3'e kadar olan adımlar için yeniden başlatılması.
10. Kaydedilen sınıflandırıcı ile mevcutta sınıflandırıcı tarafından üretilmiş olan vektörün yer değiştirmesi.
11. Üretilen kararın kabul edilmesi.

Destek vektör makineleri için belirlenmiş olan öğrenim problemi; yüksek seviyeli boyut içeren giriş vektörü x ile çıkış vektörü y aralarında bazı bilinmeyen ve doğrusal içerik içermeyen bağlar vardır (haritalama ile fonksiyon $y=f(x)$). Bu fonksiyonların asıl sebebi olan olasılık fonksiyonları ile ilgili bir bilgi yoktur. Bu sebeple ihtiyaç olan tek bir serbest öğrenme dağılımının gerçekleştirilmesidir.

Bir eğitim data kümesindeki elde olan tek bilgi $D = \{(x_i, y_i) \in X \times Y\}, i=1, l$ eğitim veri çiftlerinin sayısını ifade eder ve bu sebeple eğitilen data büyüklüğü D ye eşit olmaktadır. Genelde y_i ile d_i ile de ifade edilir, d istenen değer (hedef)dir. (Elmas,2012)

Destek vektör makineleri algoritması, Vapnik'in örüntülerin tanımlanması bunun yanında da sınıflandırma ile ilgili konu olan sorunların çözümlenebilmesi için uygulanmaya başlamıştır. Destek vektör makineleri algoritmasının en temelinde istatistiki öğrenmeye dayalı diğer bir deyişle Vapnik-Chervonenkis (VC) teorisini baz almaktadır. (Li, et al., 2009).

Aşağıdaki Şekil 3.2'de yer alan $K(x_i, x_j)$, girdilerin içi çarpımlarının hesaplanmasında kullanılan kernel fonksiyonlarını, \bar{a}_m 'ler langrange çarpanlarını yani ağırlıkları ve Y çıktı değerleri ise kernel fonksiyonu ile ortaya çıkan iç çarpımlar ile ağırlık değerlerinin kombinasyonlarından elde edilen toplamları ifade etmektedir. Girdi vektörü, gizli birimlere aktarılarak ardından ağırlıklandırılır ve bu işlemler sonrasında çıktı değerine ulaşılmaktadır. (Ayhan, 2013).



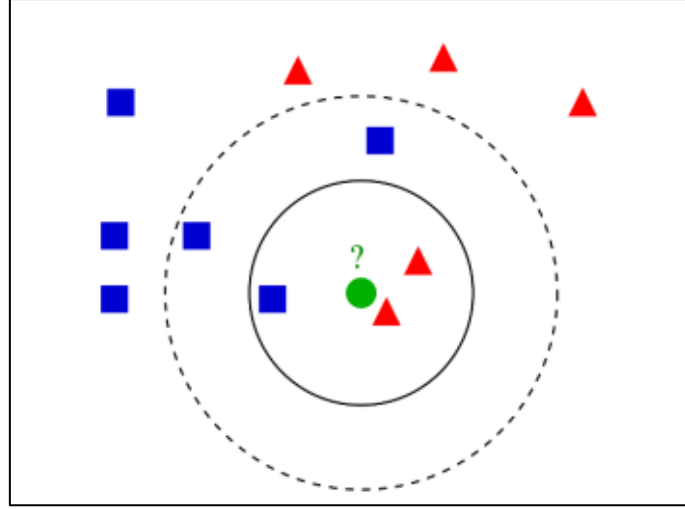
Şekil 3.2. DVM'in İşleyişi Modeli (Ayhan,2013)

3.3. K-En Yakın Komşu (K-NN) Algoritması

K-NN modelinin kullanımı çok eskiye dayandığı bilinen basit bir algoritmadır. Algoritmanın genel çalışma şekli, yeni gelen örnek bir veri ile K adet komşu arasındaki uzaklığın ölçülmesi ile oluşmaktadır. K-NN denetimli bir algoritmadır. Bu sebeple algoritma kullanılırken veriler ilk eğitim ve test olacak şekilde iki parçaya ayrılmaktadır. Sürecin sonrasında ise;

- Parametre K belirlenmelidir. K parametresi verilmiş olan bir noktaya en yakın komşu sayılarını temsil etmektedir. K değeri, komşu sınıflar arasındaki uzaklık ve sınıflandırma performansını doğrudan etkilemektedir.
- Dataya katılacak yeni veri için mevcuttaki diğer veriler ile uzaklıkları her bir veri için tek tek hesaplanır.
- Uzaklıkların hesaplanması için farklı fonksiyonlar kullanılabilir. Literatürde ise yaygın olarak Minkowski, Öklid, Manhattan ve Chebyshev fonksiyonları kullanılmaktadır.
- Hesaplanan uzaklıklar içerisinde en yakını olan k komşusu belirlenir. Özgün niteliklerine göre K. Komşunun grubuna tayin olur.
- Atanılan sınıf, tahmin edilmek istenen sınıf değeri olarak kabul edilir. Böylelikle yeni verinin sınıfı bulunur. (Metlek & Kayaalp, 2020).

Aşağıdaki Şekil 3.3'de görülen algoritma yapısında önceden belirlenmiş mesafe ölçütü metodu ile sınıfı belirsiz verinin sınıfını tahmin edebiliriz.



Şekil 3.3. K-NN Algoritma Yapısı (Deniz, 2021)

Literatürde genel olarak öklid (Euclidean) ölçüm metodu kullanır. Öklid ölçüm metodu Denklem (3.11)'de belirtilmiştir;

$$d_{öklid}(x_1, x_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad (3.11)$$

Diğer mesafe ölçüm metotları ise, Minkowski için Denklem (3.12), Manhattan için Denklem (3.13) ve Chebyshev için ise Denklem (3.14)'deki gibi uzaklık ölçüm metotları kullanılabilir.

$$d_{Minkowski}(x_1, x_2) = \sqrt[n]{\sum_{i=1}^n |x_{1i} - x_{2i}|} \quad (3.12)$$

$$d_{Manhattan}(x_1, x_2) = \sum_{i=1}^n |x_{1i} - x_{2i}| \quad (3.13)$$

$$d_{Chebyshev}(x_1, x_2) = \max_i |x_{1i} - x_{2i}| \quad (3.14)$$

K-NN algoritması içerisinde farklı ölçütlerde kullanılarak doğru sınıflandırma tahminleri yapabildiğini gösteren çalışmalarda mevcuttur. (Weinberger ve ark., 2006)

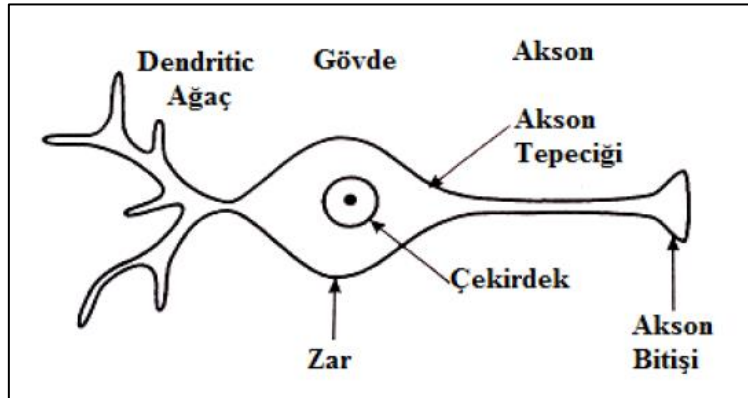
4. YAPAY SİNİR AĞLARI

4.1. Yapay Sinir Ağları Tanımı

YSA insan beyindeki nöronların öğrenme sürecini simüle etmek için tasarlanmış bilgisayar tabanlı sistemlerdir (Kwon, 2011). Literatürdeki birçok tanımı yapılmış olan yapay sinir ağları beyin çalışma prensiplerinin matematiksel olarak bilgisayarlarda taklit edilebilmesi düşüncesi ile oluşmuştur.(Efe ve Kaynak, 2000).

YSA birbiri ile hiyerarşik bağı olan ve paralel şekilde de çalışabilen yapay sinir hücrelerinden oluşmaktadır. Literatürde süreç elemanları olarak da kabul edilen bu hücrelerin birbirlerine bağlandığı ve oluşan her bir bağlantının değeri olduğu genel kabul olarak görülmektedir. Ortak noktalardan biri , bilginin öğrenme yöntemi ile ortaya çıktığı ve süreç elemanlarının bağlantı değerlerinde gizlendiği, dolayısı ile dağıtık bir hafızanın da söz konusu olduğudur. Yapay sinir ağı, proses elemanlarının birbirleriyle bağlanmaları sonucunda meydana gelen ağa verilen addır.Bir ağın yaratılması, biyolojik sinir sistemi ve sinir hücreleri üzerindeki keşiflere dayanmaktadır (Öztemel, 2003).

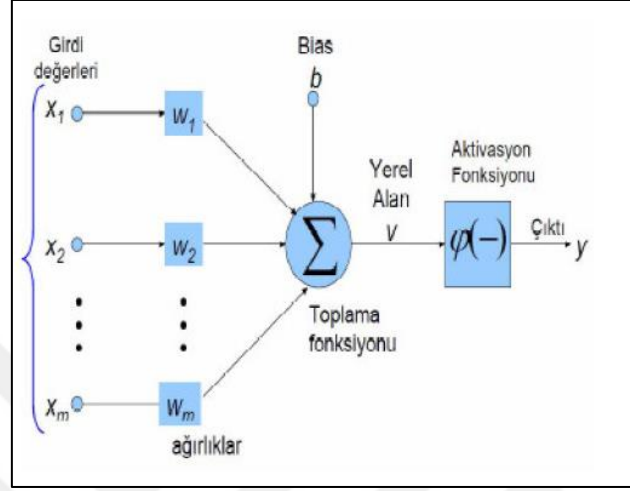
Biyolojik sinir sistemleri, özel sinir hücreleri nöronlar tarafından kontrol edilir. Oluşturulan yapay nöronun tasarımsal ve işlevsel benzerlikleri Şekil 4.1 ve Şekil 4.2 karşılaştırıldığında doğrulanabilmektedir. Biyolojik nöron incelendiğinde bilgi dentrit (Dendrite) yoluyla nörona gelir, soma işler ve akson (Axon) yoluyla iletilirler (Şekil 4.1).



Şekil 4.1. Biyolojik Nöron Yapısı (Kartalopoulos, 1996)

Benzer şekilde Yapay Sinir Ağları'nda da bilgi (girdi), ağırlıklardan giriş yaparak toplanır ve transfer fonksiyonuyla işlenir. İşlenen bilgiler çıktı üzerinden iletilir. (Oruç, 2022)

Problemleri çözmek amacıyla kullanılan yapay sinir ağı hücreleri, bir ağı şeklinde birbirlerine bağlanmışlardır. Aşağıdaki şekilde hücrelerin arasındaki meydana gelen bilgi geçişi, bağlantı değerleri ve ilişkileri gösterilmektedir. Sistemin öğrenme becerisi ve akılcı davranışları, bağlantı değerlerinin aracılığıyla gösterilmiştir. (Tekin, 2009)



Şekil 4.2. Yapay Sinir Ağı Hücresi (Adıyaman 2007)

YSA'da girdi ve çıktı katmanlar birer katmandan oluşurken aradaki katmanların sayısı değişiklik gösterebilir. (Aydın ve Süsler, 2022)

Girdi Katmanı: Bu katman ysa ağı için verileri getiren proses unsurlarının var olduğu alandır. Dışarıdan alınan girdi verilerini ara katmanlara ulaştırmaktadır. Bu katman, bir veya daha fazla ara katmanla iletişim kurabilir. Bu katmanındaki her giriş nöronu sinir ağının çıktılarında etkili olan bazı bağımsız değişkenlerinin yerini almaktadır. Girdi değerlerinde işlem yapılmamaktadır. Bu sebeple girdi katmanına pasif katman denilebilmektedir.

Ara (Gizli) Katman: Girdi ve çıktı katmanları arasındaki iletişimi sağlamaktadır. Problemin zorluğuna ve karışıklığına göre birden fazla ara katman var olabilmektedir. Bu katmadan girdi katmanından gelen değerlerin ağırlık değerleri çarpılıp toplamları alındıktan sonra transfer fonksiyonu ile dönüşebilmektedir. İşlemin sonucu varsa diğer bir ara katmana, yoksa direkt çıktı katmanına gönderilmektedir.

Çıktı Katmanı: Ara katmanın bağlandığı katmana denir. Ara katmandan gelmiş olan veriye bir takım işlemler uygulayarak çıktıya dönüştürmekte ardından orada bulunan

nöronlar ile çıktıyı iletmektedir. Bu katman, ağı görevlerini öğrenme şeklini cevaplayan birimleri ile ağ dışına bilgiyi aktarabilmektedir.

Son yıllarda sinir ağı teknolojisinde büyük adımlar atıldı. Bu atılım, çok çeşitli bilimsel uygulamalar üzerinde artan araştırmalarla kendini göstermektedir. Yapay sinir ağlarına ilgi, bu konuda çalışan bilim insanlarının sayısına, konuya ilişkin çalışmalar için ayrılan finansman miktarlarına, bu konuya ilişkin düzenlenen büyük konferansların sayısına ve sinir ağlarıyla ilişkili makale içeren dergi sayısına yansımaktadır. Yapay sinir ağları, örüntü tanıma, tanımlama, sınıflandırma, konuşma, görme ve kontrol sistemleri gibi çeşitli uygulama alanlarında karmaşık işlevler yerine getirecek şekilde eğitilmiştir. (Öztürk, 2020)

4.2. Yapay Sinir Ağları Tarihçesi Ve Kullanım alanları

Bilim insanları beyin ve matematiksel çalışmalara ilgi duyarak nörobiyolojik konularda çalışmalar yapması ile YSA çalışmalarına temel hazırlamışlardır. YSA çalışmalarının tarihsel süreci aşağıdaki gibidir.(Hamzaçebi, 2008).

- 1943 yılında Warren McCulloch/Walter YSA hücrelerini temel alan aritmatiksel metotlar teorisini ortaya atarak Pitts “Sinir Aktivitesinde Düşüncelere Ait Bir Mantıksal Hesap” modelini geliştirmiştir.
- 1948 yılında Wiener Sibernetik kitabıyla kontrol, iletişim ve istatistiki sinyal işleme modelleri üzerine çalışmalar yapmıştır.
- 1949’da Hebb kuralı ile öğrenebilecek ve uyum sağlayabilecek sinir ağlarının bağlantı sayısında gerçekleşen değişiklikler ile ağı öğrenbilme görüşleri ortaya çıkmıştır.
- 1956-1962 yılları arasında Adaline ve Windrow öğrenme modeli ile birlikte Tek Katmanlı Algılayıcı (Perceptron) geliştirilmiştir.
- 1965’de ilk makine öğrenme kitabı yazılmıştır.
- 1967’de Grosberg öğrenme modeli ile bazı olgunlaşmış öğrenme modelleri geliştirilmiştir.
- 1969’da Tek katmanlı algılayıcılar ile problemlerin çözümüne katkı sağlayıp sağlayamayacağı yapılan çalışmalar ile gösterilmiştir. 1972’de Kohonen ve Anderson çalışmaları ile beraber doğrusal ilişkilendiricilerle Korelasyon-matriks belleğinin olgunlaştırılması aynı zamanda da Çağrışımlı Bellek (Associative memory) çalışmaları yürütülmüştür.

- 1974’de Çok katmanlı algılayıcılar ile ilk çalışmalar geliştirilmiştir. (Geri yayılım algoritması)
- Öğretmensiz öğrenme çalışmalarının yürütülmesi:
- 1978’de Adaptif rezonans teorisi geliştirilmiştir.
- 1982’de ise Öz örgütlemeli Harita modeli ile birlikte Kohonen Öğrenmesi algoritmaları geliştirilmiştir.
- 1982’de Hopfield Ağlarının ve Çok Katmanlı algılayıcıların çalışmaları geliştirilmiştir.
- 1984: Boltzman makine çalışmaları geliştirilmiştir.
- 1985’de çok katmanlı algılayıcılar ile yapılan çalışmalar geliştirilerek Delta öğrenme kuralı ile geliştirilmiştir.
- 1988’de Radyal Tabanlı Fonksiyon ve Olasılıksal Sinir Ağlarının modelleri geliştirilmiştir.
- 1991’de ise Genelleştirilmiş Regresyon algoritması geliştirilmesi tamamlanmıştır.

YSA yıllardır kendini geliştirmeye devam etmekte ve gelişen teknoloji ile birçok yöntem ile birlikte kullanılabilmesi sebebiyle, kontrol ve sistem tanımlama konularından; tıp, haberleşme, elektronik, otomasyon, probabilistik fonksiyon kestirimleri yaklaşımları, uzay ve enerji alanlarına kadar etkin rol almaya başlamıştır. Bunların yanı sıra gelecekteki eğilimler hakkında fikir yürütmekte tahmin aracı olarak kullanılması oldukça yaygınlaşmıştır. Bunlara örnek olarak hava tahmini ve ekonomik kriz döngüleri verilebilir. Çeşitli veriler arasındaki ilişkileri keşfetmek ve izlemek amacıyla oluşturulmuş bu hesaplama sistemi, tahmin edilen ve gerçek sonuçları karşılaştırarak kendini eğiterek bu alandaki kullanıma uygunluğunu kanıtlamıştır. (Oruç, 2022)

4.3. Yapay Sinir Ağları Avantaj Dezavantajları

YSA’ları kullanılarak yapılan bazı avantajlar şu şekilde sıralanabilir (Ballı, 2014).

- YSA’ların diğer algoritmalar ile karşılaştırıldığında bazı seçkin nitelikleri vardır. Bunlar; doğrusal olmayışı, dağıtılmış hafıza yapısının olması, örneklerden öğrenme yeteneğinin bulunması, paralel işlem ve genelleme yapabilmesi, uyarlanabilirliği, hata temizleyebilirliği, hızlı hesaplama yeteneği, kolay kurgu yapısı, tasarım ve analizlerinin kolay oluşuyla birden fazla hazır paket programlarına erişilmesi şeklindedir.

- YSA'lar matematiksel ifade edilemeyen karışık problemleri kolayca çözebilmektedir. Uygunsuz ya da belirsiz ve eksik bilgilerle işlem yapabilmesi aynı zamanda sinaptik ağırlıkların desteği ile öğrenilen bilgiyi dağınık hafızasında koruma yeteneği bulunmaktadır.
- YSA, girişte verilen bilgiler ile kendi kendine ilişkiler oluşturabilmektedir. Yani
- teorik yöntem oluşuyla denklem içermemektedir. Buna ek olarak model içinde ne olduğunun bilgisi net değildir. Bu karmaşık özellikleriyle, birtakım kişiler YSA'yı "kara kutu" olarak tanımlamaktadır.
- YSA'daki sayısal nöronlar ölemez bu sebeple hafızadaki bilgiler asla kaybolmamaktadır.
- YSA'ya sunulan giriş verilerinin fazla oluşuyla orantılı olarak güvenilir sonuçlara ulaşabilmektedir. Diğer yöntemler ile kıyaslandığında hareketli ve kesikli veri setlerinde dahi daha güvenilir ve tutarlı sonuçları üretebilmektedir.
- Maliyeti ucuz ve kullanımı kolaydır. Zamandan tasarruf sağlarlar.
- YSA, yeni bilgiler meydana geldiğinde ve ortamda birtakım farklılıkların olması halinde tekrardan eğitilebilmektedir.

Özetle akademik çalışmalarda sıklıkla tercih edilmesiyle en güçlü avantajlı yanları kolay, esnek, tasarruflu, veri işleme, eksik bilgiyi tamamlayabilme, hata toleransı, güncellenebilir ve doğrusal ilişki zorunluluğu olmaması sayılabilir.

YSA'nın yukarıda maddelenen faydalarının yanında bir takım dezavantajları da bulunmaktadır ve bunlar aşağıdaki gibi özetlenebilir; (Öztemel, 2016)

- YSA ile etkin bir çalışma oluşturulabilmesi için paralel işleyen işlemciler gerektirir ve bu şekilde donanıma bağlı olarak çalışma şekli olması dezavantaj olarak karşımıza çıkabilmektedir. Kullanılan bilgisayarın çok çekirdekli işlemcisinin olması ve yazılımının yeterliliği bu vakit kaybının önüne geçebilmektedir.
- Problemlerin her çeşitinde YSA modeli kullanılması uyuşmayabilir.
- Çözüme götürecek ağ yapısının saptanması çoğunlukla deneme yanılma yöntemi ile yapılmaktadır bu da ciddi bir sorundur. Çünkü eğer uygun bir model bulunamazsa problem çözülemez ya da düşük performanslı bir çözüme ulaşılmış olunur. Yani YSA her şartta çözüm üretir ancak en doğru sonucu garanti etmemektedir.

- Kurulan her ağda parametre değerini belirleyebilecek kurallar yoktur bu da belirleme çalışmasında objektifliği engellemektedir. Çözümde kullanılacak parametrelerin (sınır sayısı ,gizli katman vb) belirlenmesi kullanıcı tecrübesine bağlıdır. Bu sebeple değişkenler tek tek değerlendirilmesi gerekir bu da ciddi bir problem teşkil etmektedir.
- YSA 'da ağa problemi tanıtmak ciddi bir problemdir. Çünkü YSA sadece nümerik veri setleri ile çalışmaktadır. Tüm değişkenler nümerik olarak dönüştürülmelidir ancak her kullanıcının tecrübesi bu durumu çözebilmek için yeterli olmayabilir.
- Çalışmadaki ağın eğitiminin tamamlanma süresi belli değildir. Buna ek olarak bitirme süresine karar verecek bir yöntem de geliştirilmemiştir. Eğitimin yeterli görüldüğü doğruluk seviyesine gelmesi ile eğitim tamamlanabilir. Ancak böylelikle ağın ezberlemesi mümkündür bu da gerektiğinden fazla uyum sorunlarını meydana getirmektedir. Ağın çözümüne olan güven sarsılmaktadır.
- Son olarak sayabileceğimiz YSA'nın aleyhine olan durum ise ağın öğrenme süresi uzun zaman alabilmektedir. Deneme yanılma ile tahminleme çalışması yapıldığı için süresinin kısaltılması çok önemlidir.

4.4. Yapay Sinir Ağlarında Öğrenme

Yapay sinir ağları taşıdıkları adaptif özelliği sayesinde zamanla öğrenebilirler. Yani, daha önce gerçekleşmiş olan deneyimleri temel alarak problemleri çok daha hızlı çözebilirler. YSA'daki bu olaya "Öğrenme" denilmektedir. Öğrenme olayı bağlantıdaki ağırlıkların belirlenip istenilen çıktılarının en doğru şekilde elde edilebilmesidir.(Adıyaman, 2007)

YSA çevre ile ilişkilidir ve buradan aldığı bilgileri ağdaki ağırlıkları kullanarak belirleyebilmektedir. YSA toplamış olduğu bu bilgileri geliştirerek bilgi birikimi artışı sağlamaktadır böylelikle etkin bir başarıya ulaşabilmek için algoritma üzerindeki ağırlıklarda değişiklikler yapmaktadır. YSA ile elde edilmek istenen sonucu yakalayabilmek için ağdaki ağırlıklar probleme göre şekillenerek öğrenme fonksiyonunun başlaması sağlanmaktadır. Örnekler üzerinde çalışılarak öğrenme gerçekleşmiş olmaktadır. Örnekler, problemlerdeki geçmiş tecrübeyi kullanarak, problem içerisindeki girdilerin ve çıktılarının aralarındaki ilişkileri ortaya çıkartıp gelecekteki oluşabilecek yeni çıktının tespitine olanak sağlayan en büyük etkendir. (Kartalopoulos, 1996).

Hebb, öğrenme kavramının sinir ağlarındaki davranışını aşağıdaki hipotez ile açıklamaktadır;

“Sinir hücrelerinden biri olan j siniri hücresi diğer bir sinir i hücresini sürekli olarak uyardığında, bu hücreler arasında, bazı metabolik değişimlerle büyüme süreçleri gözlenmektedir. Bu da i hücresini j hücresi tarafından uyarılma performansını arttırmış olmaktadır. (Hebb,1949) Bu hipotez sayısal olarak aşağıdaki Denklem (4.1)’deki gibi açıklanmaktadır;

X_i : j hücresinin i hücresinden gelen girdisi,

Y_j : j hücresinin çıktısı,

bağlantı ağırlık değeri,

w_{ij} : i ile j hücreleri arasındaki bağlantı ağırlık değeri

$\alpha > 0$: Öğrenme hızı parametresi, olmak üzere;

$$W_{ij}^{yeni} = W_{ij}^{eski} + \alpha X_i Y_j \quad (4.1)$$

Yapay sinir ağlarının ilk kullanıldığı günden bugünüme kadar birçok öğrenme kuralı geliştirilmiştir. Başlıca kullanılan öğrenme kuralları şunlardır; (Rençber, 2018)

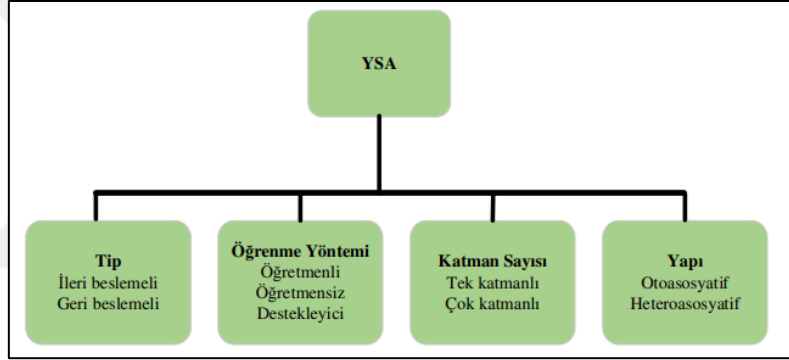
- Hebb Öğrenme Kuralı: Donald Hebb 1949’da bu kuralı öne sürmüştür. Bilinen en eski öğrenme kuralıdır. Aynı zamanda diğer öğrenme kuralları için de altyapı niteliğindedir. Bu kural algoritmadaki girdi ve çıktı değerlerinin işaretlerinin aynı olması halinde nöronlar arası ağırlığın artırılması, aksi bir durumda ise azaltılması yönünde bir prensiple çalışmaktadır.
- Hopfield Öğrenme Kuralı: Hebb kuralından farklı olarak Hopfield öğrenme kuralında kullanıcı tarafından 0-1 arasında tanımlanan bir öğrenme katsayısı bulunmaktadır. Bu öğrenme kuralında ağırlıklar öğrenme katsayısı kadar artırılır veya azaltılır.
- Delta Öğrenme Kuralı: Bu kural Widrow-Hoff ve en küçük ortalamalar karesi kuralı olarak da tanımlanmaktadır. Gerçek çıktı ile ağ tarafından üretilen çıktı arasındaki hatayı en küçük değere getirmeyi amaçlar, hata karelerinin minimum olduğu değerleri bulmak temeline dayanmaktadır.

- Kohonen Öğrenme Kuralı: Fonksiyonlar yarış halindedir, bu yarışın kazananı olan ağırlık yeni ağırlık olarak atanır ve çıktı değeri üretilir.
- Gradyan İniş Öğrenme Kuralı: Fonksiyonun türevi alınarak gerçek ile tahmin değeri arasındaki fark minimumda tutulmaya çalışılır.

4.5. Yapay Sinir ağlarının Sınıflandırılması

YSA Şekil 4.3 'deki gibi farklı karakteristik özelliklerine (Tip, öğrenme, katman sayısı, yapı) göre analiz edilirken YSA aşağıdaki gibi 3 sınıfta sınıflandırılabilir.

- Bağlantı şekillerine göre
- Katman sayılarına göre
- Öğrenme biçimlerine göre



Şekil 4.3. YSA'ların Türleri (Es, 2013)

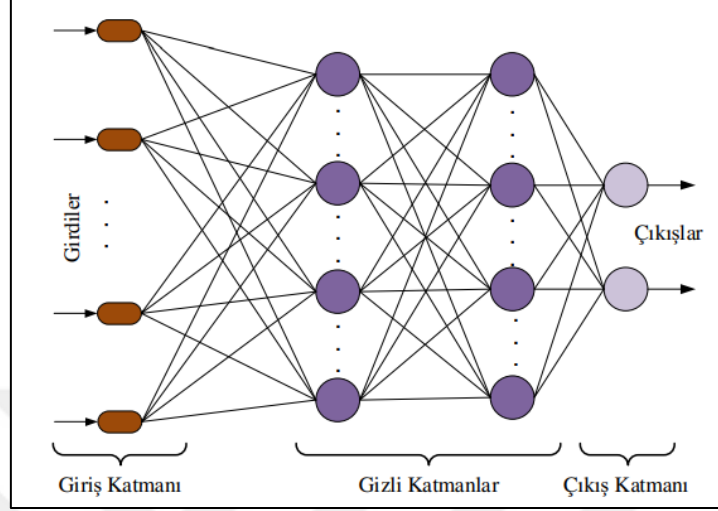
4.5.1. Bağlantı Şekillerine Göre Ysa

YSA'yı oluşturan sinir hücrelerinin birbirine bağlantı şekline göre ileri beslemeli ve geri beslemeli olarak ikiye ayrılır. (Uçarsu, 2022) İki bağlantı şekline göre gösterimlerin detaylı açıklamaları ve örnek gösterim şekilleri aşağıdaki gibidir. Bunlardan ilki olan ileri beslemeli YSA literatürde sıklıkla kullanıldığı üzere bu tez çalışmasının uygulama aşmasında da bu bağlantı şekline göre problem çözülmüştür.

4.5.1.1. İleri Beslemeli Ysa

İleri beslemeli bir sinir ağı; en basit yapay sinir ağı mimarisidir. Bilgi, giriş düğümlerinden, gizli düğümler (varsa) boyunca ve çıkış düğümlerine doğru, yalnızca bir

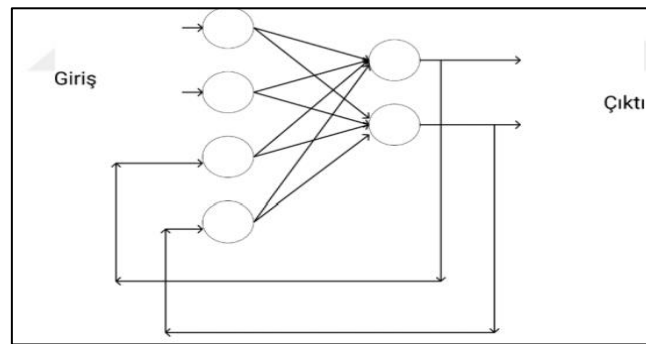
yönde hareket eder. Aşağıdaki Şekil 4.4 'de de görüldüğü üzere ağda döngü yoktur. Böyle bir ağın en yaygın eğitim yöntemi, düğümler arasındaki bağlantıların ağırlıklarını ayarlamak için çıktı sonucunu geriye doğru kullanmaktır (Gültepe, 2019)



Şekil 4.4. İleri Beslemeli YSA (Ballı, 2014)

4.5.1.2. Geri Beslemeli YSA

Geri beslemeli YSA'da; veri kendinden önceki katmana da iletilebilir. Ara katmanların herhangi birinde üretilen çıktı bir önceki katmana veya aynı katmanda bulunan başka bir nörona iletilebilir. Geri beslemeli YSA ağ yapısı Şekil 4.5'te verilmektedir. (Uçarsu, 2022)



Şekil 4.5. Geri Beslemeli Ysa (Uçarsu, 2022)

4.5.2 Öğrenme Biçimlerine Göre Ysa

Öğrenme biçimine göre yapay sinir ağları 3 başlıkta incelenmektedir;

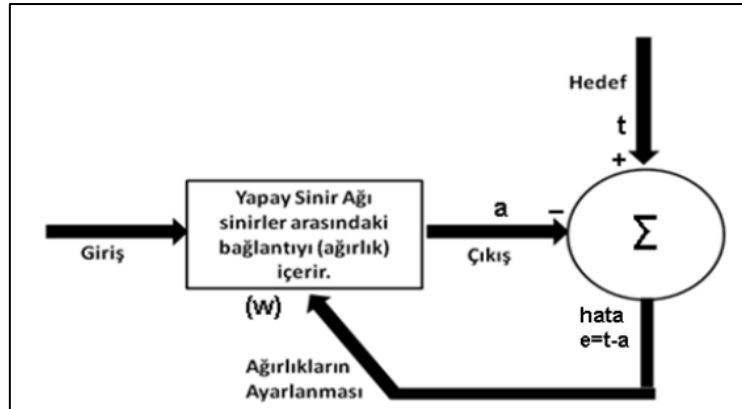
- Danışmanlı Öğrenme

- Danışmansız Öğrenme
- Takviyeli Öğrenme

4.5.2.1. Danışmanlı Öğrenme

Yapay sinir ağları yöntemlerin en sık kullanılan öğrenme yöntemi danışmalı ve danışmasız olan öğrenme yöntemleridir. Danışmalı öğrenmenin en belirgin özelliği eğitim aşamasıdır. Danışmalı öğrenmede ağın kullanılmadan önce eğitilmesi gerekmektedir. Ağı eğitmek için giriş ve çıkış verilerini ağa sunmak gerekmektedir. (Çetin, 2003)

Yapay Sinir Ağının ağırlık değerleri saklanmışsa verilen giriş-çıkış çiftleri ile hedefe ulaşılmıştır. Öğrenme ise; ağırlıkların sürekli yenilenerek ulaşılacak istenilen sonuca kadar geçen süreci ifade etmektedir. Yapay sinir ağı öğrenme işlemini tamamladıktan sonra ağın tanımadığı yeni veriler giriş yapılarak, gerçek çıkışı ile sinir ağının sonucu incelenip değerlendirilir. (Atmaca, 2020). Ağ süreci şekil 4.6 da gösterilmektedir. Verilen yeni örneklerde gerçek çıkışa yakın sonuçlar veriyorsa işlem tamamlanmış olup ağ modeli işin öğrenmesini gerçekleştirmiştir. Ama modele giriş yapılan örnekler uygunluk değerinden fazla ise iş öğrenilmemiş, model tarafından ezberlenmiştir. Genel olarak örnek girişlerinin %80'ni ile ağ eğitilip, geriye kalmış olan %20'si ile test edilerek ağ tepkisi analiz edilir. (McCulloch ve Pitts 1943)

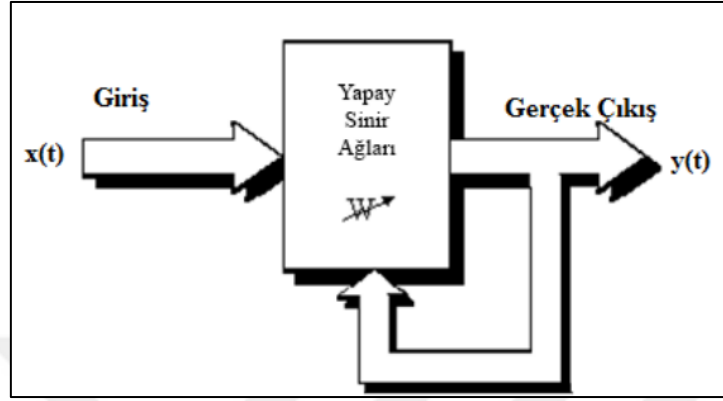


Şekil 4.6. YSA' da Danışmanlı Öğrenme (Karaođlan ,2012)

4.5.2.2. Danışmansız Öğrenme

Danışmasız öğrenme ve danışmalı öğrenme farklıdır. Şekil 4.7'de de görüldüğü üzere bu öğrenme türünde, danışmalı öğrenmedeki gibi istenilen y çıkışları bilgisi yoktur. Bu

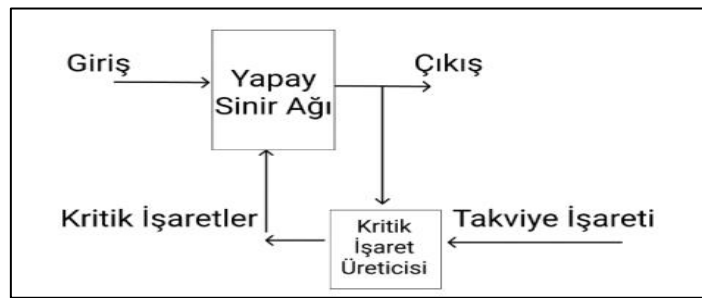
nedenle kesin bir hata bilgisi ağıın davranışının deęiřtirilmesi için kullanılamamaktadır. Cevabın doęru ya da yanlıř olup olmadıęı bilinmedięi için öğrenme, girişlerin verdięi cevaplar gözlenerek istenilen sonuçlar elde edilebilir. Bu yaklaşım daha çok sınıflandırma problemlerinin çözümünde kullanılmaktadır (McCullogh ve Pitts 1943)



Şekil 4.7. Ysa'da Danışmasız Öğrenme (Erdem, 2007)

4.5.2.3. Takviyeli Öğrenme

Danışmanlı ve danışmansız öğrenme türleri bir araya gelerek destekleyici öğrenme türünü oluşturmaktadır. Destekleyici öğrenme türünde sisteme bir öğretici yardım eder. Öğretici, YSA'da gerçek çıkış verisini kullanmak yerine, sistemin ürettięi çıkışa bakarak sonucun doęruluęunun ya da yanlıřlıęının ifadesini içeren sinyali gönderir. Şekil 4.8'de de görüldüğü üzere sistem gelen bu sinyale bakarak öğrenme aşamasını sürdürür. (Güler ve Tuna 2019)



Şekil 4.8. Ysa'da Takviyeli Öğrenme (Güler ve Tuna, 2019)

5. LİTERATÜRDE VERİ MADENCİLİĞİ VE YAPAY SİNİR AĞLARI

Yeni bir çalışmaya başlarken daha önce yapılan çalışmaların incelenmesi ve yapılacak yeni çalışmaların bilimsel tekniklere uygun olarak literatüre katkı sağlaması, yapılmış olan diğer çalışmaları da desteklemesi önem göstermektedir.

Uygulamada yapılacak olan talep tahmini çalışması için literatürdeki talep tahmini uygulamaları ve özellikle perakende sektöründe yapılmış olan çalışmalar da incelenerek veri madenciliği yöntemleri ve yapay sinir ağı yönteminin sıklıkla kullanıldığı tespit edilmiştir.

Çalışmanın doğru bir şekilde yapılabilmesi için literatür taraması yapılmıştır.

Kılıç ve Önal 2022 yılında yaptıkları çalışmada 2012-2019 yılları arasında Borsa İstanbul'da faaliyet gösteren 144 firmanın verileri analiz edilerek elde edilen finansal hilelerinin tespitinde kabul gören oranları YSA ile tahminleme çalışması yapılmıştır. Çalışma, finansal tablolardaki hile riskini %88,89 oranında doğru tahmin ederek başarılı bir sonuç ortaya çıkarmıştır.

Büyükatak 2022 yılında yaptıkları çalışmada 6890 öğrenci anketi kullanılarak öğrencilerin başarı durumları ve okuma becerileri yeterlilik düzeylerinin sınıflandırma doğruluğunu PISA 2018 Türkiye örneğine dayalı olarak Yapay Sinir Ağları, karar ağaçları, K-En Yakın komşu ve Naive Bayes yöntemleri ile incelemiştir. Çalışma sonucunda veri setleri tüm teknikler kullanılarak sınıflandırılmaya uygun bulunmuştur.

Ata ve Erbudak 2022 yılında yaptıkları çalışmada T.C Merkez Bankası verileri kullanılarak karar destek ağaçları, doğrusal regresyon, DVM, gauss regresyon yöntemleri kullanılarak sınıflandırma çalışmaları yapılmıştır. Yöntemler daha önceki çalışmalar ile karşılaştırılarak derin öğrenme metodlarının daha iyi sonuçlar elde edildiği sonucuna varılmıştır.

Gökçe ve arkadaşları 2022 yılında yaptıkları çalışmada yapay Sinir Ağları (YSA) ve Holt Winters yöntemi ile 2015 – 2020 yılları arasındaki aylık veriler ile iler beslemeli YSA'yı kullanarak 2019-2020 yıllarındaki gerçekleşmiş olan satış adetlerini tahmin çalışması yapmışlardır. YSA ile elde edilen sonuç diğer bir yöntem olan Holt Winters yöntemine göre daha başarılı sonuca ulaşmıştır.

Mohideen ve Raj 2022 yılında yaptıkları çalışmada Pima Indians diyabet veri setini analiz ederek Optimize Edilmiş Gaussian Naive Bayes ve regresyon yöntemleri ile tahmin çalışması yapılmıştır. Analiz sonuçlarında Optimize Edilmiş Gauss Naive Bayes'in diğer yöntemlere göre %81,85 sınıflandırıcı doğruluğuna ulaşılmıştır.

Tan, Mu, Fu, Yuan, Sun, Liang ve Sun 2022 yılında yaptıkları çalışmada, Naive Bayes Sınıflandırıcısı (NBC) ve Olasılıksal Yapay Sinir Ağı yöntemleri ile öncelikli olarak tahmin çalışması yapılmıştır. Çalışma sonuçlarının doğruluğu sırasıyla %85 ve %95'e ulaşmıştır.

Iliç, Srdjeviç, ve Srdjeviç 2022 yılında yaptıkları çalışmada Sırbistan'ın Voyvodina Eyaletindeki beş lokasyondan 2013-2019 dönemini kapsayan verileri kullanılarak Naive bayes yöntemi ile tahmin çalışması yapılmıştır. Sonuç olarak 68 vakanın 64'ünde su numunesinin sınıfını doğru bir şekilde elde edilmiştir.

Polat 2022 yılında yapmış olduğu çalışmada çalışmada beyaz eşya sektöründeki bir firmanın satış verilerini inceleyerek YSA, Destek vektör makinesi ve regresyon yöntemleri tahminleyerek YSA modelinden daha iyi bir sonuç elde edilmiştir.

Canbulut 2022 yılında yapmış olduğu çalışmada, pandeminin sürmesi halinde taşınacak yolcu miktarının tahmin edilebilmesi için ülkemize göre daha erken vakaların görüldüğü ülkeler ile benzerlikler analiz edilerek WEKA yazılımı kullanılarak Lineer Regresyon ve Yapay Sinir Ağları yöntemleri ile tahmin çalışması yapılmıştır. Yapay Sinir Ağları algoritması daha doğru tahminlere ulaşmıştır.

Acı ve Doğansoy 2022 yılında yaptıkları çalışmada yerel bir süpermarketten elde edilen iki yıllık e-ticaret verilerinin TÜFE, yıllık işsizlik oranı gibi faktörler de göz önüne alınarak altı farklı yöntem ile; Yapay Sinir Ağları, Gauss Süreç Regresyonu, Regresyon Ağacı, Destek Vektör Regresyonu ve Ağaç Toplulukları talep tahmini çalışması yapılmıştır. Sonuçlar, korelasyon katsayısı (R), Kök Ortalama Kare Hata (KOKH) ve Ortalama Mutlak Hata (OMH) kriterleri ile değerlendirildiğinde en başarılı performans Uzun ve Kısa Vadeli Hafıza Ağları yöntemlerinden elde edilmiştir.

Büyükkeçeçi ve Okur 2022 yılında yaptıkları çalışmada işletmeyi terk edebilecek müşterilerin tahmini için toplamda 21 sınıflandırma yöntemi ve telekomünikasyon,

bankacılık ve e-ticaret sektörlerine ait veri setleri analiz edilerek tahmin çalışmaları yapılmıştır. Ek olarak RFM (Recency, Frequency, Monetary Value) bölümlenmesi ve Ki-Kare Testleri kullanılarak boyut indirgeme çalışmaları yapılmış ve elde edilen sonuçlar kıyaslanmıştır.

Maden 2022 yılında yapmış olduğu çalışmada Knime veri analiz platformunda tıp öğrencilerinin sınav sonuçları veri seti kullanılarak karar Ağaçları, Yapay Sinir Ağları, Rastgele Orman, Naive Bayes ve K-En Yakın komşu algoritmaları ile tahmin çalışmaları yapılmış ve sonuçların başarı değerleri kıyaslanmıştır. Doğruluk oranları kıyaslandığında farklı kriterlerde değerlendirilen algortimalarda en iyi sonuç veren yöntemler farklılaşmakla beraber Knime platformunun veri madenciliği kullanımı için uygun olduğu sonucuna varılmıştır.

Gençosman ve Şanlı 2021 yılında yaptıkları çalışmada içerisinde 12 PAH türü olan farklı 4 arıtma çamuru üzerinde UV çalışmaları gerçekleştirmiştir. Bunlar üzerinde çok katmanlı algılayıcı ağı, k-NN, C4.5 karar ağacı, rastgele orman ve torbalama modelleriyle sınıflandırma tahmin çalışmaları yapılmıştır. RO (%95), k-NN (%95) ve ÇKA (%91) ile sırasıyla en etkili sonuç üreten yöntemlerdir.

Avcı 2021 yılında yapmış olduğu çalışmada tıbbi görüntü işleme teknikleri kullanılarak DVM, RF ve YSA yöntemleri ile ulaşılan sonuçların diğer algoritmalara göre daha iyi performans gösterdiği görülmüştür.

Çörekçioğlu, Ercan ve Elibüyük 2021 yılında yaptıkları çalışmada yapay sinir ağı yöntemlerinin tekstil sektöründe ki kullanımına yönelik literatür çalışmaları araştırılmıştır. Çalışma sonucunda Tekstil sektöründe dokuma , terbiye ve kumaş konfor parametrelerinin de dahil olduğu birçok alanda YSA yöntemi başarılı sonuçlar ile tercih edilebileceği gösterilmiştir.

Aghalarova ve Keser 2021 yılında yaptıkları çalışmada ortaokul öğrencilerinin Matematik ve Portekizce derslerindeki performanslarının tahmini için yeni bir yapay sinir ağı algoritması kullanılmıştır. Uygulanan rastgele arama algoritması ve ysa modelinin hiper-parametreleri optimize edilmiştir. Uygulama sonuçları Matematik dersi için %97.0 ve %92.3 Portekizce dersi için ise %97.6 ve %87.9 olarak elde edilmiştir.

Özgür 2021 yılında yapmış olduğu çalışmada Elektrik piyasasındaki Türkiye için yapılmış çalışma örnekleri ve Türkiye elektrik piyasasında fiyat tahmini uygulaması yapmış olup yapay sinir ağı temelli modellerin etkinlikle kullanılabileceği ancak geleceğe yönelik fiyat tahmini yapmak üzere tasarlanan YSA temelli modelin eğitilmesinde zaman serilerini içeren verilerin büyük önemi olduğu sonucuna ulaşmıştır.

Peromingo ve ark. 2020 yılında yaptıkları çalışmada kuru kürlenmiş jambon olgunlaşması için AFB1 üretiminin veri madenciliği teknikleri kullanılarak tahmin edilme çalışması yapılmıştır. Elde edilen tahmin denklemlerinin korelasyon katsayıları ≥ 0.74 ve tahmin hataları ≤ 0.38 göstermiştir.

Van Nguyen ve ark., 2020 yılında yaptıkları çalışmada yeniden üretilmiş ürünlerin yüksek doğrulukta, sağlam bir talep tahmin modeli sağlamayı ve müşteri talebinin yordayıcıları olarak çevrimiçi pazar faktörlerinin doğrusal olmayan etkisini veri madenciliği ile incelemiştir.

Montella ve ark. 2020 yılında yapmış oldukları çalışmada 2011-2013 döneminde İspanya'da meydana gelen elektrikli 2 tekerlekli araçlı (PTW) kazalara katkıda bulunan çarpışma özellikleri arasındaki kalıpların hangileri olduğunu 78.611 kazadan elde edilen veriler ile analiz etmişlerdir.

Chen ve ark. 2020 yılında yapmış oldukları çalışmada Veri madenciliği yöntemlerini kullanarak kömür yakıtlı güç üniteleri için enerji verimliliği endekslerini karşılaştırma çalışması yapılmıştır.

Tomasevic ve ark. 2020 yılında yaptıkları çalışmada veri madenciliği ve yapay sinir ağları kullanılarak öğrenci sınavı performans tahmini görevini çözmek için uygulanan en son teknolojiye sahip makine öğrenim tekniklerinin kapsamlı bir analizini ve karşılaştırılmasını yaparak gelecekteki başarılarını tahmin etmeye çalışılmıştır.

Güven 2020 yılında yapmış olduğu çalışmada ürün çeşitliliği renkleriyle birlikte dikkate alınarak 2014-2018 yılları arası satışlarını YSA, DVM ve rastgele ormanlar yöntemleri ile satış tahmini yapılmış olup Kurulan 14 veri seti sonuçlarına göre 8'inde YSA renk detaysız modellerin sonuçları diğer yöntemlerden daha iyi sonuçlanmıştır. YSA renksiz

çalışmalarda 14 modelden 11'inde daha etkili sonuç verirken , DVM ise 13 modelde renkli verilere göre daha başarılı sonucuna ulaşmıştır.

Vural, Küçükaşçı ve Ayvaz 2019 yılında yapmış oldukları çalışmada elektronik perakende şirketinin 'notebook' ürünleri için önceki yıllara ait satış datası ile yapay sinir ağı yöntemi kullanılarak gelecek zamanlar için satış tahmini çalışılmıştır. Kurulan modellemenin tahmin hataları ve sonuçları incelenmiştir.

Zazzaro ve ark. 2021 yılında yapmış oldukları çalışmada epilepsi hastalarının nöbet geçirme sinyalinin veri madenciliği ile tahmin etme çalışması yapmışlardır. Kamuya açık EEG veri setleri üzerinde yapılan testlerde % 99'un üzerinde bir kesinliğe ulaşan mükemmel sonuçlar göstermiştir.

Ercan, 2021 yılında yapmış olduğu çalışmada 2019 yılına ait toplam 11.521 hane halkından elde edilen verileri çalışmanın analizinde kullanarak ev dışı gıda tüketim sınıfının belirlenmesinde Yapay Sinir Ağları yöntemi kullanılmıştır. Modelleme sonucu, doğru tahmin edilme oranı eğitim veri setinde %77.28 iken test veri setinde %80.39 olarak elde edilmiştir. Model tarafından ev dışı gıda tüketimi yapmayan haneler eğitim veri setinde %68.98, test veri setinde ise %66.70 ile doğru tahmin edilmiştir.

Karaatlı Ve Altıntaş 2018 yılında yapmış oldukları çalışmalarında Borsa İstanbul'daki işlenmiş hisse senetlerinin artma, azalma, sabit kalış değişimlerine göre veri seti oluşturmuşlardır. Sektör ve işletme açısından anlamlı bir kümeleme oluşturması için analiz edilmiştir. Çalışma Weka programında Beklenti maksimizasyon algoritmasını kullanarak 134 işletmenin, hisse senetlerinin artışı oranlarını 3 seviye ayırmışlardır.

Şengül ve ark. 2019 yılında yapmış oldukları çalışmada Weka yazılımı kullanılarak lojistik regresyon analiziyle yükseköğretim kurumlarında öğretim üyesi ders performansının öğrencinin ders başarısı üzerindeki etkisi araştırılmıştır. 5820 adet anket verisinden hareketle öğrencilere öğretim üyesi değerlendirmesi ile ilgili yöneltilen 16 soru sorulmuş ve analiz edilmiştir. Öğretim üyesi tutumunun öğrencinin ders başarısını etkilediği sonucuna varılmıştır.

Türkoğlu ve ark. 2019 yılında yapmış oldukları çalışmada endüstriyel bir şirketteki soğuk dövme makinelerinde, kümeleme, sınıflandırma ve karar ağacı gibi veri madenciliği

yöntemlerini kullanarak gerçek hayattaki veri setlerindeki makine duraksamalarını tahmin edilmiştir. Weka ve Rapidminer programları kullanılarak eğitilmiş ve test edilmiştir. Weka'da kullanılan J48 Decision Tree algoritmasının %77 doğruluk ve 0.4 hata değeri ile en verimli sonucu verdiği gözlenmiştir. Rapidminer programı kullanılarak modellenen karar ağacında ise, %71.8'lik doğruluk sonucuna ulaşılmıştır.

Alkaragole ve Kurnaz 2019 yılında yapmış oldukları çalışmada diyabet hastalığı teşhisinde kullanılmak amacıyla hibrit veri madenciliği teknikleriyle önerilen framework'ü geliştirmişlerdir. Sonuçlar, önerilen framework hibrit sınıflandırmasının diğer sınıflandırıcıları% 94'lük bir doğruluk oranı ile geride bıraktığını göstermiştir.

Akkuş ve Zontul 2019 yılında yapmış oldukları çalışmada daha öncesinde belirlenmiş oldukları parametreleri göz önünde bulundurarak ülkelerin gelişmişlik seviyelerine göre durumlarını incelemişlerdir. 214 ülke ve 2015 veriyi kullanarak Self Organizing Map ve K-Means kümeleme modellerini uygulayıp, sonuçlarında bulunan kümeler incelenmiştir. Ek olarak Türkiye'nin kümeler arasındaki durumu ayrıca analiz edilmiştir.

Altay 2019 yılında yapmış olduğu çalışmada üniversitelerin kütüphanelerindeki VM çalışmalarının genel konumunun çıkartılabilmesi ve faydalarından yararlanabilmek için YORDAM programı kullanılarak 2010-2018 yıllarındaki çalışmaları Ayırıcı Hiyerarşik Kümeleme yöntemi ile incelemiştir. Çalışma sonucunda mevcuttaki durum incelenerek gelecek için stratejik önerilerde bulunulmuştur

Ayan ve ark. 2019 yılında yapmış oldukları çalışmada 2017 yılında kozmetik, elektronik ve pazar sektörlerinde faaliyet gösteren rakipler tarafından paylaşılan Twitter verilerini, Sosyal Medya Madenciliği ile analiz etmişlerdir. Kümelenede daha fazla etkileşime giren kümeler Pareto İlkesi'nin yardımıyla incelenmiştir.

Xiaoyu Qi ve ark. 2019 yılında yapmış oldukları çalışmada 2017 -2018 yılı arasında Pekin'deki hava kirletici konsantrasyonları ve meteorolojik koşullar arasındaki ilişkileri veri madenciliği ile analiz etmişlerdir. Elde edilen sonuçlar kentsel yönetim performansını daha da iyileştirirken, meteorolojik koşullara göre hava kalitesini tahmin etmede yardımcı olabilmektedir.

Ajak ve ark. 2018 yılında yapmış oldukları çalışmada madencilik programında problemleri cevherle karşılaşma olasılığını tahmin etmek için gerçek bir maden işletmesine ön görüşel veri madenciliği algoritmaları uygulamışlardır. En uygun veri madenciliği algoritması, kıl oluşumunu % 78,6 hassasiyetle öngören sınıflandırma ağacı seçilmiştir.

Köseoğlu ve Demirci 2018 yılında yapmış oldukları çalışmada teknoloji temelli şehirlerde var olan geniş verilerin yenilikçiliklerinin analizi için, nesne interneti, yığın veriler, VM, yapay zeka teknikleri gibi akıllı teknoloji ve şehirlerin uygulamalarındaki ilişkileri incelemişleridir.

Bilgiç ve Esen 2018 yılında yapmış oldukları çalışmalarında özellikle VM ve yapay zeka teknikleri ile ilgili 2016'dan bu yana yapılmış pazarlama sektöründeki çalışmaları analiz etmişleridir. Çalışma sonucunda bilgisayar alanında uzman kişilerce geliştirilen farklı modellerin hızlı bir şekilde bu alandaki çalışmalara uyarlanıp kullanıldığı tespit edilmiştir.

Nagy ve ark. 2017 yılında yapmış oldukları çalışmada kümeleme ve sınıflandırma yöntemleri ile toplu taşıma araçlarında farklı mesafe ölçüm tekniklerinin kullanılabilirliğini kontrol edilebilirliğini incelemiştir. Kümelene ile Manhattan uzaklık ölçümü % 78 hassasiyet ve % 63 sınıflandırma ile sonuçlandı. İncelenen formdaki sınıflandırmanın trafik bölgelerinin özelliklerini keşfetmek için uygun bir araç olmadığını göstermektedir.

Eren ve Sataoğlu 2017 yılında yapmış oldukları çalışmada giyim üzerine faaliyette olan firmanın bir mağazasının satışları birbirinden farklı iki Yapay Sinir Ağı Modeli ile tahmin edilmiştir. İki model tahmini incelendiğinde ANN modelinin, klasik ANN modeline kıyaslanarak sezonsal etkiye göre daha doğru tahmin sonucu elde edilmiştir.

Adıyaman 2007 yılında yapmış oldukları çalışmada yapay sinir ağları, regresyon ve eğri uydurma yöntemlerini kullanarak altın ürün satışlarını tahmin etme çalışması yapılmıştır. Yapay sinir ağı yönteminin diğer yöntemlere kıyasla daha etkin bir yöntem olduğu sonucuna ulaşılmıştır.

6. UYGULAMA

Çalışmada perakende sektöründe İstanbul’da faaliyetlerine devam eden özellikle giyim markası olarak üretimde ciddi bir pazar payına sahip olan firmanın üretilen ürünün cinsine göre 10 grupta (denim, dokuma, elbise vb.) toplanan 2019-2022 yılları arasındaki 4 yıllık satış miktarları bağımlı değişken olarak ele alınmıştır. Bağımsız değişkenlerde ise stok miktarı, fiyat, indirim oranı, hava durumu ve Türkiye istatistik kurumu (TÜİK)’undan temin edilen tüketici güven endeksi ve Perakende(Tekstil-giyim) satış Hacim endeksi verileri aylık olarak toplanarak veri seti hazırlanmıştır.

Bu değişkenlerin yanında şirketin hali hazırda da bir sonraki sezonun bütçe tahmini için kontrol ettiği ve satışa olumlu olumsuz etki katsayısı olarak girdiği, anomali değişkeni de eklenmiş olup çalışmada öncelikle satış miktarını etkileyen anomalilerin de satışa etkisi tespit edilmeye çalışılmıştır. Geçmiş yıllarda toplanan veriler ışığında 4 Anomali değişkeni (Range, fiyat, stok ve pandemi anomalisi) dataya eklenmiştir. Kullanılan değişkenleri tanımları aşağıdaki gibidir;

Stok miktarı; Hedeflenen “t” anında doğru zamanda, doğru ürün karışımının mağazada bulundurulduğu adettir.

Satış adet; T anında ilgili ürünün bedelinin ödenmesiyle satıcı tarafından satın alınan miktardır.

Fiyat: Ürünün belirlenen müşteri tarafından kabul görmesi beklenen ilk bedelidir. Aynı dönemde birden fazla ürün satışa sunulduğu için bu çalışmada ay bazında ortalama fiyat kullanılmıştır.

İndirim Oranı: İndirim, ilgili ürünün fiyatındaki azalmadır. Belirli bir ürünün satış adedini artırmak amacıyla birçok ticari stratejide kullanılan bir araçtır. Aynı dönemde x ürününün indirimli olduğu bölge ve mağazalar farklılık göstermekte olduğu için bu çalışmada ay bazında ortalama indirim oranı kullanılmıştır.

Tüketici Güven Endeksi; Tüketici mali ve ekonomik durumlarının güncel durum değerlendirmelerinin ve gelecekteki beklentileriyle yakın gelecekte harcama veya tasarruf eğilimlerini ölçmek için kullanılan bir endekstir. (Özdemir, 2013)

Perakende Satış Hacim Endeksi; fiyat değişimlerinden arındırılmış bir hacim endeksi özelliği taşımanın yanında perakende ticaret sektörü içerisinde farklı büyüklükte olan firmaların satış miktarlarını ay bazında ölçmeye yarayan endeks çeşididir. Bu endeks, hane halkı isteklerinin ve tüketicilerin güvenini kısa dönemli göstermektedir. (Kaya ve ark., 2018)

Hava durumu; Satışa sunulan ürün çeşitliliğinin çok fazla olması, hava şartlarına karşı müşteri ihtiyacının değişmesi sebebiyle bir sonraki dönem satış planları yapılırken hava durumu büyük önem taşımaktadır. Örneğin sıcaklıklar mevsim normallerine göre sıcak gitmesi durumunda dış giyimleri ya da daha kalın kıyafetlerin satışları düşük seyretmektedir.

Stok anomalisi; Satış potansiyelinin altında stok bulundurulması durumunda stok eksikliği, potansiyelin üzerinde bulundurulması durumunda ise stok fazlası durumları oluşmaktadır. Bu durumlardan birinin yaşanması durumunda ilgili ayda stok anomalisi meydana gelmektedir.

Fiyat anomalisi: Müşterilerin ilgili ürün için ödemesi beklenen fiyatın üstünde ya da altında bulundurulması durumudur. Müşteri gözünde, ilgili ürünün ederinden daha fazla fiyatta bulundurulması durumunda satış beklendiği gibi gerçekleşmemektedir.

Range anomalisi; ilgili ay ve grupta dönemin şartlarına ya da müşterinin alım davranışlarına uygun olamayan ürün bulundurmak anlamına gelmektedir. Örneğin; bayram döneminde günlük okazyonlu ürün buldurmak, Yaz ayında kalın pantolon buldurmak vb.

Pandemi anomalisi ise ; Geçtiğimiz dönemlerde yaşamış olduğumuz mağazaların kapanması ve dışarı çıkılmasının yasaklanması müşterilerin alış-veriş davranışlarını değişmesi halidir.

Yukarıda belirtilmiş olan anomaliler ilgili ay ve ürün grubundaki mağazalardan ve bölgelerden gelen notlarla seçilmiş olan en yüksek satış kaybına sebep olan anomaliler olarak data setine giriş yapılmıştır. (Herhangi bir ay da birden fazla anomali gerçekleşmiş olabilir ancak en çok karşılaşılan anomali seçilmektedir.)

Literatür taramasında incelenen gerek veri madenciliği açısından kaynak niteliğinde olan Han ve arkadaşlarının ‘Data Mining Concepts and Techniques’ isimli kitabı gerekse benzer veri setleri ile tahmin yapılan diğer bilimsel araştırmaların ışığında doğrusal ve doğrusal olmayan verilerden yapılan tahmin çalışmalarında etkin sonuçlar elde edilebilen destek vektör makinaları ve yapay sinir ağları yöntemlerinin kullanılmasına karar verilmiştir.(Arslan,2020). Bu yöntemlere ek olarak anomalinin tahmin edilmesinde sınıflandırma tahmin algoritmalarından literatürde sıklıkla kullanılan Naive bayes ve K-En yakın komşu algoritmaları ile de anomali tahmini yapılmış olup yöntemler kıyaslanmıştır. Oluşturulan modellerin çözümü WEKA (Waikato Environment for Knowledge Analysis) ve MATLAB programları ile gerçekleştirilmiştir.

Çalışma iki aşamalı gerçekleşecektir. İlk olarak halihazırda şirketin de kullandığı anomalilerin tahmin edilmesi ve sonuçların değerlendirilmesi çalışması yapılacaktır. İlk çalışmada anomali değişkeni bağımlı değişkenken satış adet bağımsız değişken olarak ele alınmaktadır. 3 veri madenciliği ve sınıflandırma algoritması ile satış tahmini yapılması istenen dönemler için anomali tahminlemesi yapılarak çalışmanın ikinci aşamasına girildiği düşünülmektedir. Ardından çalışmanın ikinci aşamasında tahmin edilen bu anomali değişkenleri ile beraber ve aynı zamanda da anomali değişkeni göz ardı edilerek 2 şekilde de Yapay sinir ağları ile satış tahmini yapılacaktır. Bu analizlerle beraber sonuçların değerlendirilmesi ve yöntemlerin kıyaslanması ile çalışma tamamlanacaktır. Örnek veri seti Tablo 6.1’de gösterilmektedir.

Tablo 6.1. Veri Setinden Örnek Gösterim

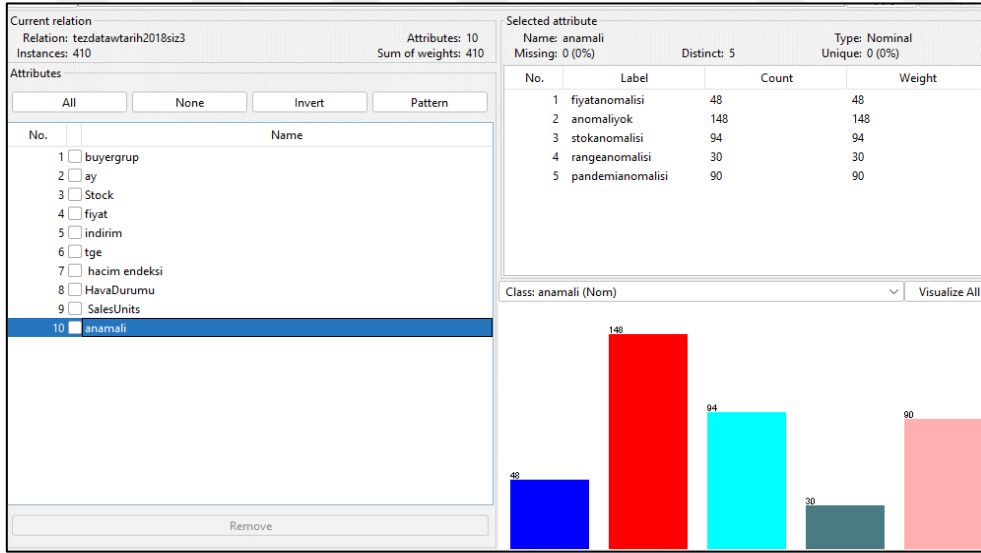
ÜRÜN GRUP	DENİM	DIŞ GİYİM	ÖRME	ÖRME	BASIC	ELBISE
AY	202001	202012	202002	202108	202202	202005
STOK	100.000	92.345	94.698	146.817	83.309	391.719
FİYAT	40,5	150,4	36,7	44,1	67,4	44,4
İNDİRİM ORANI	0,3	0,1	0,3	0,4	0,3	0,1
TGE	58,8	80,1	57,3	78,2	71,2	59,5

Tablo 6.1. (Devam) Veri Setinden Örnek Gösterim

PERAKEDE SATIŞ HACİM ENDEKSİ	135,1	125,9	133,5	191,1	148,1	59,9
HAVA DURUMU	5,9	10,3	7,4	27,7	7,6	21,2
SATIŞ ADET	120.000	43.484	72.219	96.157	63.550	300.000
ANOMALİ	STOK	PANDEMİ	STOK	STOK	RANGE	PANDEMİ

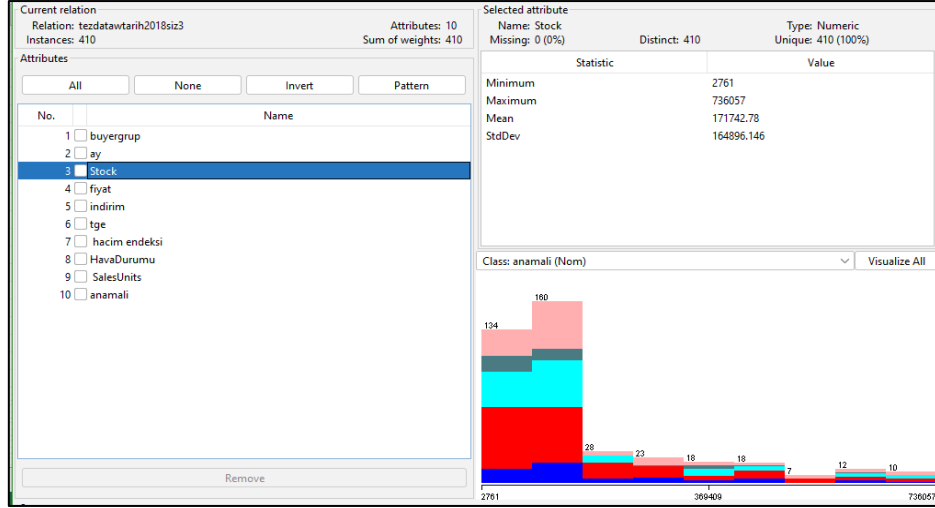
6.1. Anomali Tahmini

Veri madenciliği süreçleri uygulanan veri setinde ilk olarak anomalilerin tahmin edilmesi ve bir sonraki satış tahmini çalışmasına girdi sağlaması için veriler analiz edilmiş ve normalizasyonu işlemi için ARFF uzantılı dosyaya dönüştürülerek WEKA programında çalışır hale gelmiştir. Bu ilk çalışmada satış miktarı bağımsız değişkenken anomali değişkeni ise bağımlı değişkendir. Aşağıdaki Şekil 6.1’de Wekada data setinin ARFF dosya görünümü gösterilmiştir.



Şekil 6.1. Weka'da ARFF Dosya Görünümü

Toplamda 10 değişkenli 410 adetlik (41 ay x 10 grup) bir data analizlere konu edilmiş olup 10 ürün grubunda çalışmalar yapılmıştır. Yukarıda örnek kesiti verilen datanın değişken bazlı detaylı olarak incelenmesi aşağıdaki gibi belirtilmiştir.



Şekil 6.2. Stok Miktarının Weka'da Gösterimi

Tahminleme çalışması için gerekli olan stok miktarının Weka'da gösterimi Şekil 6.2'te gösterilmiştir. İstatistik durum olarak;

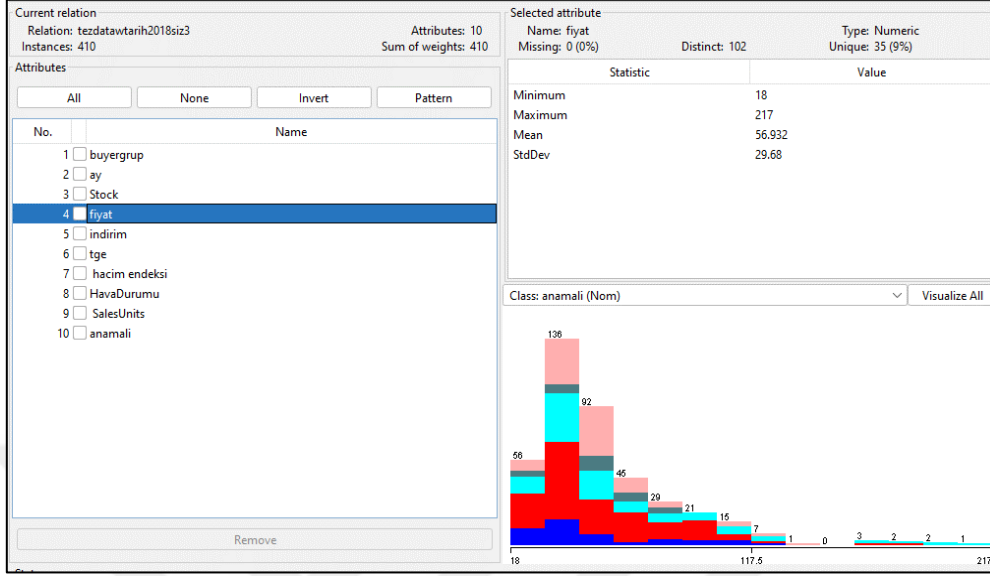
- Minimum stok miktarı 2761 adettir. (Data seti içerisindeki 4 yıllık ürün grubundan mağazada en düşük bulundurduğu stok miktarıdır.)
- Maximum stok miktarı 736.057 adettir. (Data seti içerisindeki 4 yıllık ürün grubundan mağazada en yüksek bulundurduğu stok miktarıdır.)
- 410 adet veri setindeki 4 yıllık verinin stok miktarı ortalaması 171742 adetken standart sapması 164.896 adettir.

Yapılan analizlere göre satış potansiyelinin altında stok bulundurulması durumunda stok eksikliği potansiyelin üzerinde bulundurulması durumunda ise stok fazlası durumları oluşmaktadır. Bu durumlardan birinin yaşanması durumunda ilgili aydan stok anomalisi girişi yetkililerce girilmektedir.

Tahminleme çalışması için gerekli olan fiyat miktarının Weka'da gösterimi Aşağıda Şekil 6.3'te gösterilmiştir. İstatistik durum olarak;

- Minimum fiyat 18 TL'dir. (Data seti içerisindeki 4 yıllık ürün grubundan en düşük fiyata sahip olan ürün 18 TL'dir.)
- Maximum fiyat 217 TL'dir. (Data seti içerisindeki 4 yıllık ürün grubundan en düşük fiyata sahip olan ürün 217 TL'dir.)

- 410 adet veri setindeki 4 yıllık verinin fiyat ortalaması 56.93 TL'iken standart sapması 29.68 TL'dir.

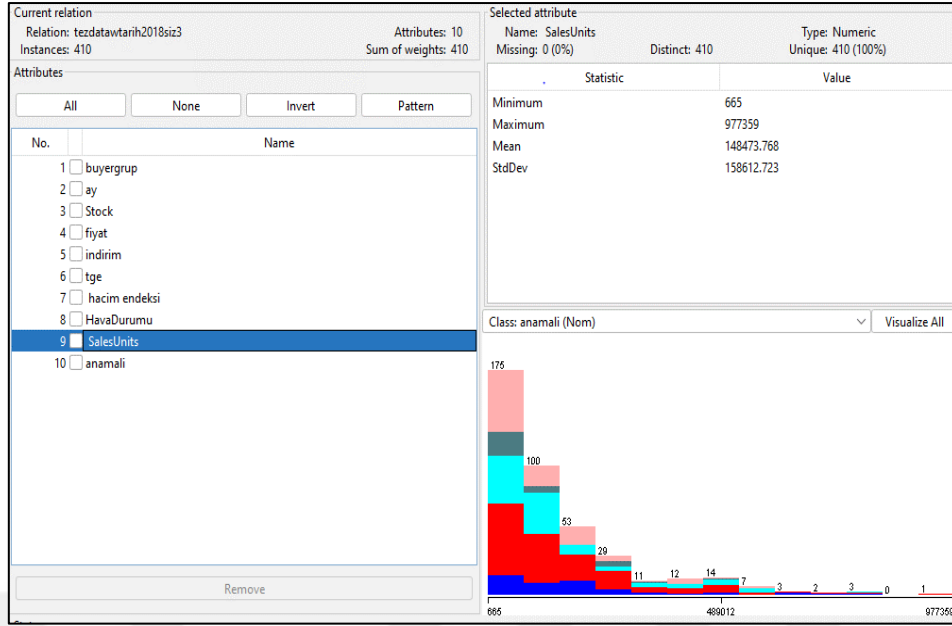


Şekil 6.3. Fiyat Miktarının Weka'da Gösterimi

Yapılan analizlere göre satış potansiyelinin altında fiyatta bulundurulması durumunda müşteriler uygun fiyata yöneldiği için ilgili üründe satış adet karşılanamamaktadır. Potansiyelin üzerinde fiyatta bulundurulması durumunda ise ilgili fiyata müşteriler tarafından fazla bulunduğu için rakibe ya da farklı buyer gruplara yönelim olmaktadır. Bu durumlardan birinin yaşanması durumunda ilgili ayda fiyat anomali girişi yetkililerce girilmektedir.

Tahminleme çalışması için gerekli olan satış miktarının Weka'da gösterimi Şekil 6.4'de gösterilmiştir. Satış miktarı 10 değişken arasından problemin ilk aşamasında bağımsız değişken olarak alınırken problem çözümünün ikinci aşamasında bağımlı değişken olarak kabul edilerek çözüme dahil edilmiştir. İstatistik durumları aşağıdaki şekilde açıklanmaktadır;

- Minimum satış miktarı 665 adettir. (Data seti içerisindeki 4 yıllık ürün grubundan en düşük satış adetidir.)
- Maximum satış miktarı 977359 adettir. (Data seti içerisindeki 4 yıllık ürün grubundan en yüksek satış adetidir.)
- 410 adet veri setindeki 4 yıllık verinin satış miktarı ortalaması 148473 adetken standart sapması 158612 adettir.

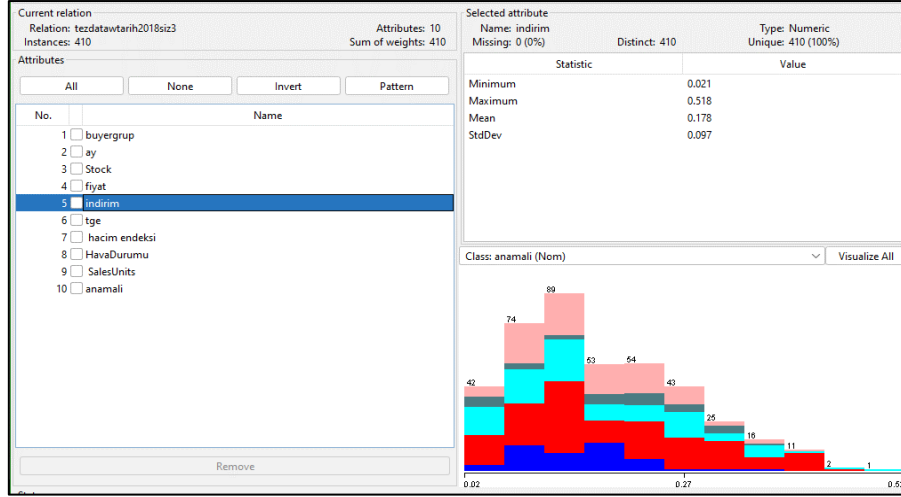


Şekil 6.4. Satış Miktarının Weka'da Gösterimi

Yapılan analizlere göre planlanan satış miktarının altında satış gerçekleştirdiğinde satış kaybı, planlanan miktardan daha fazla satış gerçekleştirilmesi durumunda ise potansiyelinin üzerinde satış gerçekleştirmektedir. Bu durumlardan birinin yaşanması durumunda ilgili ay detaylı analiz edilerek satış kanallarından da gelen bilgiler ışığında belirtilen anomalilerden ilgili olanı girilmektedir.

Tahminleme çalışması için gerekli olan indirim oranının Weka'da gösterimi aşağıdaki Şekil 6.5'de gösterilmiştir. İstatistik durum olarak diğer değişkenler ile aynı özelliklerini açıkladığımızda ise;

- Minimum indirim oranı 0,002'dir. (Data seti içerisindeki 4 yıllık ürün grubundan en düşük indirim oranıdır. Ürün satış ömrünün ilk dönemlerinde indirimsiz satış görmektedir.)
- Maximum indirim oranı 0,52'dir. (Data seti içerisindeki 4 yıllık ürün grubundan en yüksek indirim oranıdır. Özellikle sezon sonu dönemlerde indirim oranı artmaktadır. Belli bir ürün bir mağazada %70 indirimli iken başka bir mağazada indirimsiz olabilmektedir. Datadaki indirim oranı ay bazındaki ortalama indirim oranıdır.)
- 410 adet veri setindeki 4 yıllık verinin indirim oranı ortalaması 0,18 iken standart sapması 0,09 dir.

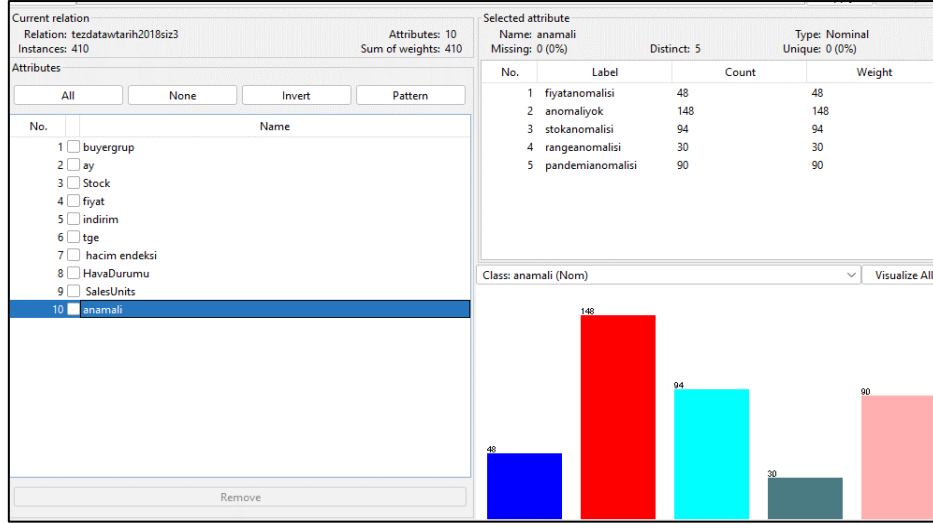


Şekil 6.5. İndirim Oranının Weka'da Gösterimi

Yapılan analizlere göre planlanan indirim oranının altında indirim gerçekleştirdiğinde yüksek beğeni almayan ve indirim bekleyen müşterilerde satış kaybı yaşanırken, planlanan miktardan daha fazla indirim gerçekleşmesi durumunda potansiyelinin üzerinde satış gerçekleştirmekte ve indirimli ürünlere daha fazla rağbet gösterildiği için satabilecek diğer ürünlerin potansiyeli de düşmektedir. Bu durumlardan birinin yaşanması durumunda ilgili ay detaylı analiz edilerek satış kanallarından da gelen bilgiler ışığında fiyat anomalisi olarak ilgili aya girilmektedir.

Tahminleme çalışması için gerekli olan uzman kişilerce ve ilgili satış kanallarından toplanan anomalilerin Weka'da gösterimi aşağıdaki Şekil 6.6'de gösterilmiştir. Anomaliler her bir mağazada sorumlu olan uzman tarafından toplanmakta olup ardından bölge bazlı en yüksek anomaliler belirlenerek datada tutulmaktadır. Merkez ofisin de teyidi ile ilgili ürün grubundaki o ayki gerçekleşen anomali belirlenmiş olur. Belirlenmiş anomali aynı zamanda ülke sorumlusu çalışanlar tarafından da kontrol edilerek anomalilerin üründe gerçekleşmiş olduğu herhangi bir problemten kaynaklanıp kaynaklanmadığı tespit edilerek seçilmektedir. Örneğin sık bir ürün range anomalisi altında dataya girilmeden önce üründe herhangi bir kusurun olup olmadığı mağazacılık ve tedarik ekiplerinde kontrol edilmektedir.

İstatistik durum olarak toplamda 410 veri setinin 48'inde fiyat anomalisi, 94'ünde stok anomalisi, 30'unda range anomalisi, 90'ında pandemi anomalisi ve 148'inde de anomali görülmemiştir.



Şekil 6.6. Anomalilerin Weka'da Gösterimi

Yukarıdaki açıklamalara ek olarak Range anomalisini açıklamamız gerekirse; ilgili ay ve grupta dönemin şartlarına ya da müşterinin alım davranışlarına uygun olamayan ürün buldurmak anlamına gelmektedir. Örneğin; bayram döneminde günlük okazyonlu ürün buldurmak, Yaz ayında kalın pantolon buldurmak vb.

Pandemi anomali ise ; Geçtiğimiz dönemlerde yaşamış olduğumuz mağazaların kapanması ve dışarı çıkılmasının yasaklanması müşterilerin alış-veriş davranışlarını değiştirmiştir. Normalde çok satımlı olan ürünler müşterilerde gerçekleşen gelecek kaygısı ve nerede giydireceğini bilememesi durumlarından kaynaklı satış görememiştir. Pandemi döneminin başlamasıyla yaşanan bu değişikliklerin görüldüğü ve mağazaların kapalı olduğu dönemler analiz edilerek pandemi anomali uzman kişiler ve ilgili ekiplerle kontrol edilerek girilmiştir.

Tablo 6.2. Tüketici Güven Endeksi (TUIK, 2022)

TGE	2019	2020	2021	2022
Ocak	58,48	58,8	83,3	73,2
Şubat	57,92	57,3	84,5	71,2
Mart	59,48	58,2	86,7	72,5
Nisan	64,45	54,9	80,2	67,3
Mayıs	55,51	59,5	77,3	67,6
Haziran	57,74	62,6	81,7	
Temmuz	56,73	60,9	79,5	

Tablo 6.2. (Devam)Tüketici Güven Endeksi (TUIK, 2022)

Ağustos	58,5	59,6	78,2	
Eylül	55,14	82	79,7	
Ekim	56,6	81,9	76,8	
Kasım	59,42	80,1	71,1	
Aralık	58,23	80,1	68,9	

Tablo 6.3. Perakende(Tekstil-giyim) Satış Hacim Endeksi Tablosu (TUIK, 2022)

Perakende(Tekstil-giyim) Satış Hacim Endeksi	2019	2020	2021	2022
Ocak	128,5	135,05	106,61	145,43
Şubat	129,49	133,54	113,29	148,07
Mart	131,59	92,89	152,04	175,47
Nisan	135,72	31,16	118,89	205,04
Mayıs	137,99	59,91	112,33	179,6
Haziran	127,67	119,63	185,47	
Temmuz	131,88	147,28	204,96	
Ağustos	143,04	125,57	191,08	
Eylül	146,87	137,12	212,75	
Ekim	146,89	153	219,42	
Kasım	153,5	145,38	190,8	
Aralık	156,12	125,86	195,41	

Yukarıdaki verilere ek olarak Tüketici güven endeksi ve Perakende satış hacim endeksi verileri de tüketicilerin ekonomik durumlarının mevcut ülke durumları ile ilişkin durum değerlendirmelerinde ve yakın gelecekteki harcama eğilimlerinde doğrudan etkilendiği için bu çalışmadaki talep tahmini analizi için veri olarak girdi sağlanmıştır. Veriler Tablo 6.2 ve Tablo 6.3’de gösterilmiştir.

6.1.1. Weka’da Naive Bayes Algoritması ile Anomali Tahmini

Naive Bayes algoritması olasılıksal ilkeler göz önüne alınarak hazırlanmış olan birçok hesap yöntemleriyle, sistem tarafına verilen dataların sınıflarını yani kategorilerini bulabilmek için kullanılmaktadır.

Bu modelde sınıflandırılmak istenen belli bir oran ile öğretilmiş data seti sistem tarafına aktarılır. Öğrenimde kullanılabilmesi için aktarılan dataların kesinlikler bir sınıfa ya da kategoriye ait olmaları gerekmektedir. Öğretilen verilerdeki yapılmış olan olasılık işlemleriyle, sistem tarafına test için verilen yeni verilerin iştilerek verilmiş olan testlerin hangi sınıfta oldukları bulunmaya çalışılmaktadır. (<https://kodedu.com/>)

Bu tez çalışmasında satış miktarını etkileyen anomalilerin Weka programı içerisinde yer alan Naive Bayes algoritması kullanılarak analizlerin yapılması sağlanmıştır. Analiz sonuçlarına aşağıdaki tablolarda yer verilmiştir.

Uygulanacak olan 3 yöntemde de aynı şartlar altında doğruluk değerleri kıyaslanmış olup 2 grupta incelenmiştir. Bunlardan ilki Cross validation olup (Çapraz geçerlilik); veri setini ilgili parça sayısına bölerek böldüğü grup sayısından birini seçer ve onu test seti yaparak analiz eder ardından kalan gruplar içinde aynı işlemi yaparak bölünen parça sayısı kadar model oluşturmuş olur.

İkincisi ise data setinden belirtilen oranını test seti olarak ayırır ve eğitilmiş veriyi test ederek sonuca ulaşmış olur.

Tablo 6.4. Naive Bayes Cross Validation Analiz Sonuçları

Naive Bayes Algoritması (Cross -Validation)	Parça Sayısı	Doğruluk (%)
1	3	62
2	5	60
3	10	62
4	20	62
Ortalama/SS	-	62

Tablo 6.5. Naive Bayes Percentage Split Analiz Sonuçları

Naive Bayes Algoritması (Percentage split)	Eğitim(%)	Test(%)	Doğruluk (%)
1	90	10	66
2	80	20	68
3	70	30	62
4	50	50	53
Ortalama/SS			62

Tablo 6.4 ve 6.5 de Naive Bayes algoritması kullanılarak eğitim ve testler oranları belirlenerek doğruluk oranı ortalama ikisinde de %62 olarak hesaplanmıştır. En yüksek doğruluk oranı veri setinin %20 u test olarak ayırılarak %68 elde edilmiştir. Weka’da elde edilen sonuç matrisi ve değerleri aşağıdaki Şekil 6.7’de gösterilmektedir.

The screenshot shows the WEKA software interface for a Naive Bayes classifier. The 'Test options' section is set to 'Percentage split' at 80%. The 'Classifier output' section displays the following summary:

```

Time taken to test model on test split: 0 seconds

=== Summary ===
Correctly Classified Instances      56          68.2927 %
Kappa statistic                    0.5738
Mean absolute error                0.1786
Root mean squared error            0.3023
Relative absolute error            60.0286 %
Root relative squared error        79.0614 %
Total Number of Instances         82
  
```

The 'Detailed Accuracy By Class' section shows the following metrics:

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,333	0,068	0,375	0,333	0,353	0,279	0,712	0,254	fiyatanom.
	0,656	0,140	0,750	0,656	0,700	0,531	0,888	0,817	anomaliyo.
	0,667	0,115	0,667	0,667	0,667	0,552	0,809	0,676	stokanom.
	0,333	0,089	0,125	0,333	0,182	0,155	0,861	0,427	rangeanom.
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	pandemian.
Weighted Avg.	0,683	0,095	0,716	0,683	0,697	0,592	0,870	0,743	

The 'Confusion Matrix' section shows the following matrix:

```

=== Confusion Matrix ===
 a  b  c  d  e  <-- classified as
3  2  2  0  0  | a = fiyatanomalisi
3  21  5  3  0  | b = anomaliyok
2  3  14  2  0  | c = stokanomalisi
0  2  0  1  0  | d = rangeanomalisi
0  0  0  0  17  | e = pandemianomalisi
  
```

Şekil 6.7. Naive Bayes Percentage Split WEKA Sonuç Matrisi

6.1.2. Weka’da Destek Vektör Makineleri ile Anomali Tahmini

Yapılan literatür araştırmasında görüldüğü üzere DVM algoritması kullanımı kolay ve hızlı olduğu aynı zamanda diğer algoritmalara göre daha doğru sonuçlar elde ettiği için regresyon ve sınıflandırma problemlerinde kullanılan en yaygın yöntemdir.

Naive bayes algoritmasında kullanılan data ile aynı şartlarda Weka’da DVM ile tahmin çalışmaları yapılmıştır. Sonuçlar aşağıdaki Tablo 6.6 ve Tablo 6.7’deki gibidir;

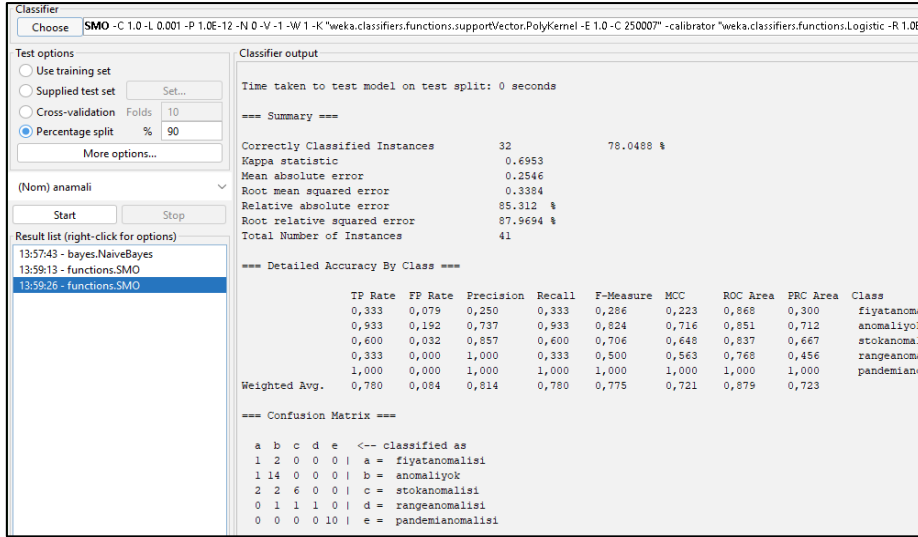
Tablo 6.6. DVM'de Cross Validation Analiz Sonuçları

Destek Vektör Makineleri (Cross Validation)	Parça Sayısı	Doğruluk (%)
1	3	70
2	5	68.5
3	10	67
4	20	66
Ortalama/SS		68

Tablo 6.7. DVM'de Percentage Split Analiz Sonuçları

Destek Vektör Makineleri (Percentage split)	Eğitim(%)	Test(%)	Doğruluk (%)
1	90	10	78
2	80	20	75,6
3	70	30	69
4	50	50	62,4
Ortalama/SS			71

Tablo 6.6 ve 6.7 de Weka’da Destek vektör makineleri algoritması kullanılarak eğitim ve testler oranları belirlenerek yapılmış olan çalışmalarda, doğruluk oranı ortalama sırasıyla %68 ve %71 olarak hesaplanmıştır. En yüksek doğruluk oranı veri setinin %10 u test olarak ayrılarak elde edilen çalışmada %78 olarak elde edilmiştir. Weka’da elde edilen sonuç matrisi ve değerleri aşağıdaki Şekil 6.8’de gösterildiği gibidir.



Şekil 6.8. DVM’de Percentage Split (%90-10) WEKA Sonuç Matrisi

6.1.3. Weka’da K-En Yakın Komşu Algoritması ile Anomali Tahmini

K- En Yakın Komşu Modeli, denetimli öğrenme tekniğine dayalı en basit makine öğrenmesi algoritmalarından biridir. Bu algoritma, yeni durum/veriler ile mevcut durumlar arasındaki benzerliği varsayarak analiz yapmaktadır. Yeni durumu mevcut kategorilere en çok benzeyen kategoriye yerleştirmektedir.(Sandıkçı,2022)

Yukarıdaki çalışmalara ek olarak Weka’da son kullanılan algoritma K-en yakın komşu algoritmasıdır. KNN algoritması ile anomali tahmin çalışmaları yapılmıştır. Sonuçlar aşağıdaki gibidir;

Tablo 6.8 KNN'de Cross Validation Analiz Sonuçları

K-En Yakın Komşu (Cross - Validation)	K Sayısı	Parça Sayısı	Doğruluk (%)
1	k=1	3	61
2		5	62
3		10	62
4		20	62
1	k=2	3	60
2		5	60
3		10	61
4		20	62

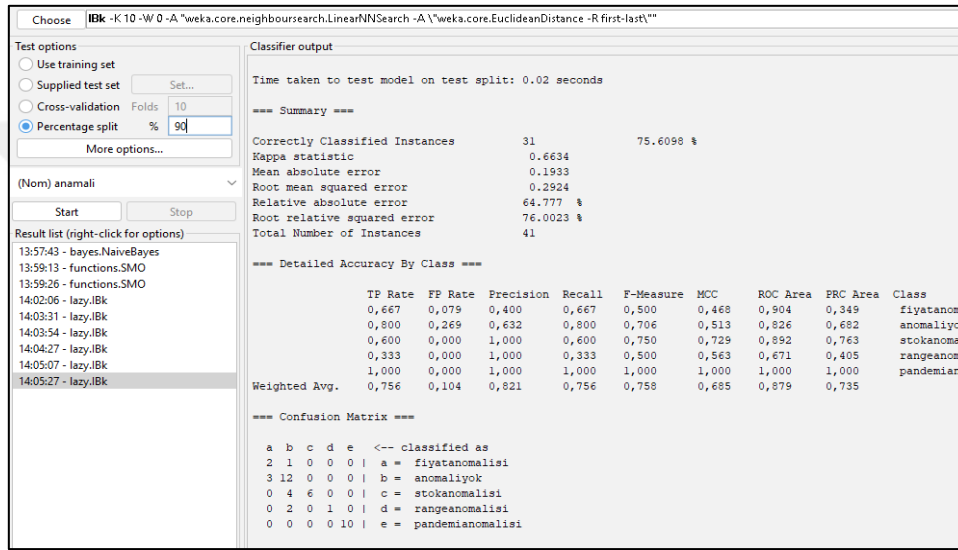
Tablo 6.8. (Devam) KNN'de Cross Validation Analiz Sonuçları

1	k=5	3	66
2		5	65
3		10	66
4		20	68
1	k=8	3	66
2		5	64,6
3		10	64,8
4		20	65
1	k=10	3	65,8
2		5	67,3
3		10	64,4
4		20	65,6
Ortalama/SS			64

Tablo 6.9. KNN'de Percentage Split Analiz Sonuçları

K-En Yakın Komşu (Percentage split)	Eğitim(%)	Test(%)	K Sayısı	Doğruluk (%)
1	90	10	1	63
	90	10	2	56
	90	10	5	75
	90	10	8	75
	90	10	10	75
2	80	20	1	65
	80	20	2	61
	80	20	5	74
	80	20	8	72
	80	20	10	69
3	70	30	1	66
	70	30	2	62
	70	30	5	65
	70	30	8	64
	70	30	10	65
4	50	50	1	57
	50	50	2	55
	50	50	5	62
	50	50	8	61
	50	50	10	60
Ortalama/SS				65.1

KNN yönteminden diğer yöntemlerden farklı olarak k komşu sayısında farklılık yaparak analiz sonuçlarını en iyi sonuç için değerlendirebildiğimiz için literatürde de sıklıkla kullanılan k değerinin 1 ve 10 arasında değerler verilerek sonuçlar elde edilmiştir. Tablo 6.8 ve 6.9 da Weka’da KNN algoritması kullanılarak eğitim ve testler oranları belirlenerek yapılmış olan çalışmalarda, doğruluk oranı ortalama sırasıyla %64 ve %65 olarak hesaplanmıştır. En yüksek doğruluk oranı veri setinin %10 u test olarak ayrılarak elde edilen çalışmada k=(5,8 ve 10) değerlerini aldığıında %75 olarak elde edilmiştir. Weka’da elde edilen sonuç matrisi ve değerleri aşağıdaki Şekil 6.9’daki gibidir.



Şekil 6.9. KNN’de Percentage Split (%90-10) WEKA Sonuç Matrisi

6.2. Yapay Sinir Ağları İle Satış Tahmini

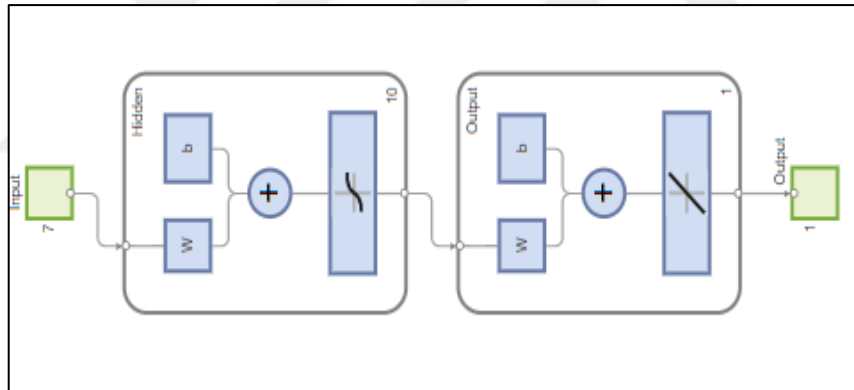
Yukarıdaki çalışmalar ile anomalilerin tahmin çalışması yapılmıştır. Çalışmanın bu aşamasında ise aynı veri seti kullanılarak devam edilecektir. Ancak satış miktarı bağımsız değişkenken kabul edip tahmini kontrol edilecekken anomali değişkeni ise bağımlı değişkendir. Anomalilerin bağımsız değişken olduğu ve göz ardı edildiği iki türlü de tahmin çalışması yapılacak olup sonuçlar karşılaştırılacaktır.

Öncelikle oluşturulan yapay sinir ağı algoritmasının katman sayısı, katmanlardaki nöron sayısı, kullanılacak fonksiyonu (aktivasyon) ve öğrenme kuralı ağın tasarımı için önemli girdileri belirlenmiştir. YSA modeli tasarımı ve eğitimi için diğer kullanmış olduğumuz Weka programından farklı olarak Matlab NNtool arayüzü kullanılarak analizler yapılmıştır.

Satış tahmini için tasarlanan algoritmada literatürde de birçok çalışmada kullanılan, doğrusal ve doğrusal olmayan modellerdeki başarılı tahminlerde bulunan aynı zamanda da kullanımı kolay olması sebebiyle sıklıkla tercih edilen ileri beslemeli geri yayılım (feed-forward backprop) algoritması kullanılmıştır.

Çalışmada YSA giriş parametresi olarak stok miktarı, fiyat, indirim oranı, hava koşulları ve Türkiye istatistik kurumu (TÜİK)'undan temin edilen tüketici güven endeksi, Perakende(Tekstil-giyim) satış Hacim endeksi ve anomali değişkenleri olarak 7 girdi ve satış miktarı da bir çıktı parametresi olarak alınmıştır. Veri setinin %70 i eğitim verisi olarak alınırken %15 lik kısmı validasyon %15 lik kısmı da test verisi olarak seçilmiştir.

Oluşturulan YSA yapısı şekil 6.10'daki gibidir. Algoritmanın eğitilmesi için literatürde en çok kullanılan traingdx (Variable Learning Rate Backpropagation) eğitim algoritması, öğrenme fonksiyonu için levenberg- marquardt ve aktivasyon fonksiyonu için de tanjant sigmoid fonksiyonu (tansig) kullanılarak YSA modeli tasarlanmıştır.

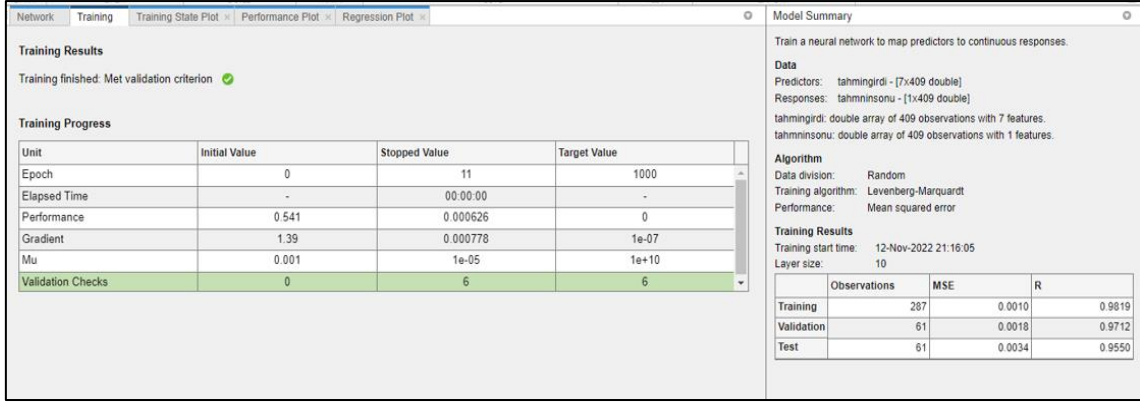


Şekil 6.10. YSA Katman Yapısı (Anomali Dahil)

Matlabta ilgili veri kümeleri girişi yapıldıktan sonra normalize edilmiş veri setinin transpozesi alınmış ve ilgili model yukarıdaki parametreler ile kurumuştur. Oluşturulan modelde gizli katmandaki nöron sayısı da 10 olarak belirlenmiştir.

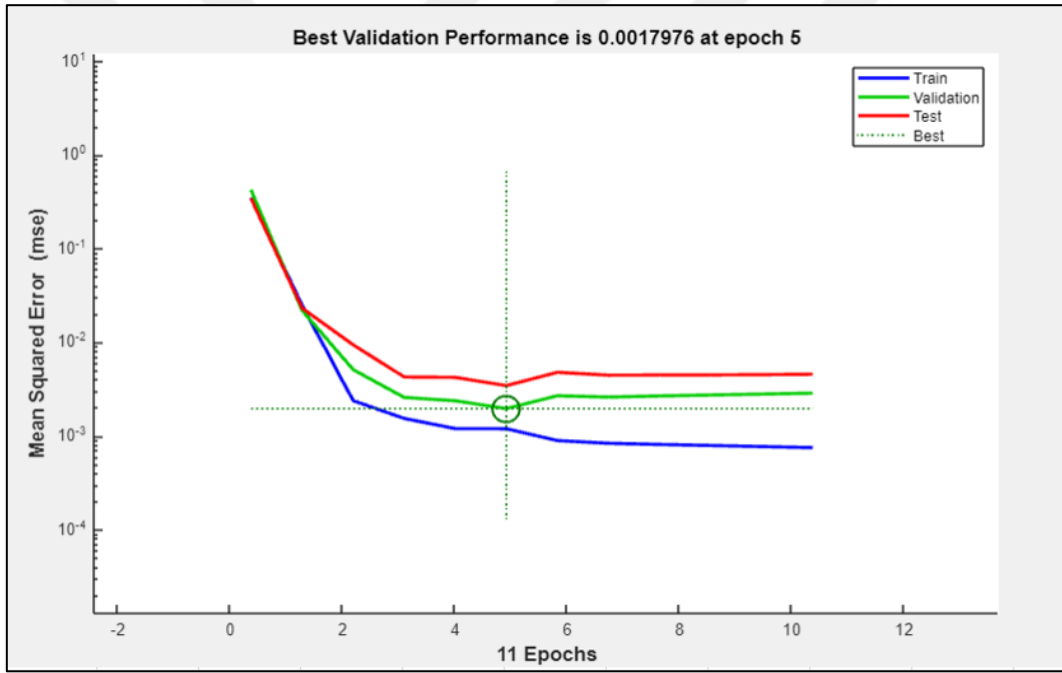
Oluşturulan YSA modeli eğitiminde durma kriteri olarak “1000” iterasyon alınmıştır. Eğitimi 0.0000001 s sürede 11 iterasyonda tamamlanmış olup 5. İterasyonda en doğru sonuca ulaşmıştır.

Kullanılan Matlab YSA arayüzü eğitim sonuçları Şekil 6.11’te gösterilmiştir.



Şekil 6.11. Matlap'ta YSA Arayüzü Ve Eğitim Sonuçları

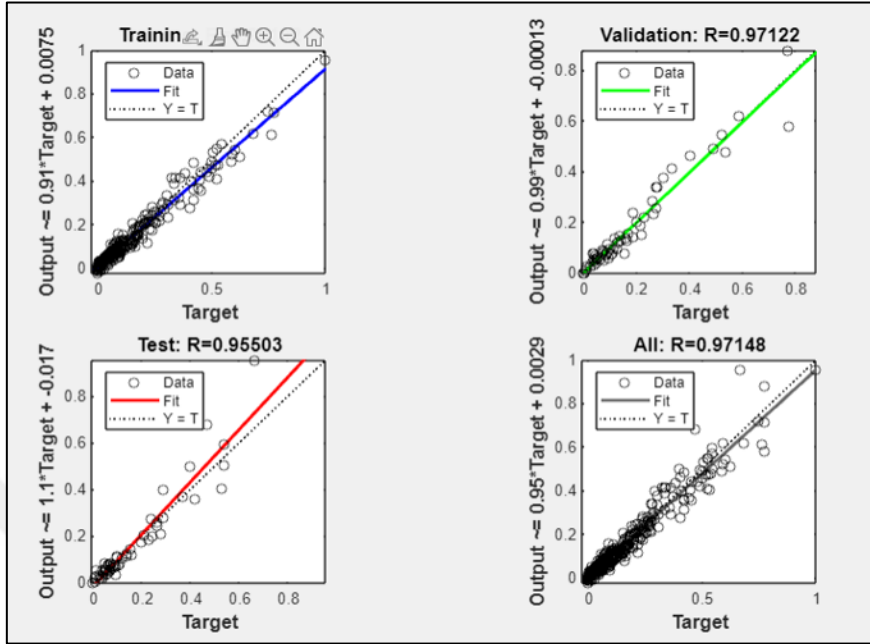
Eğitime ilişkin sonuçlar ve en iyi değerin bulunduğu 5. İterasyonunun iterasyona bağlı hata değişim grafiği aşağıdaki Şekil 6.12'de gösterilmiştir.



Şekil 6.12. Eğitim, Validasyon ve Test Verileri Eğitim Performansı

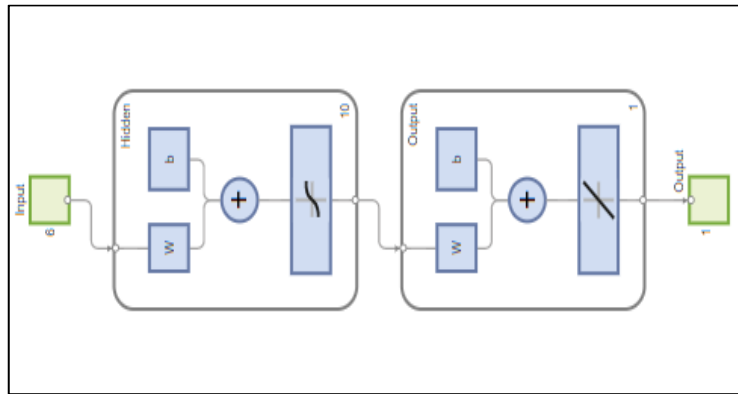
Modelleme performansı incelendiğinde gerçek satış değerleri ile YSA'nın tahminlediği satış değerlerinin karşılaştırılmalı eğitim, validasyon ve test sonuçlarına ait regresyon grafikleri Şekil 6.13 'de gösterilmektedir. Grafikte de görüldüğü üzere korelasyon katsayısı 0 ile 1 arasında değişerek bu değer 1'e yaklaşması ağın başarımını göstermektedir. Grafikler incelendiğinde eğitim regresyon değeri 0.9819, validasyon değeri 0.9712, test değeri ise 0.9550 olarak gerçekleşerek 1'e çok yakın değerler alınmıştır. Tüm bunlar ışığında toplam değeri kıyasladığımızda ise 0.9714 değeri ile

başarılı bir tahmin gerçekleşmiş olup MSE değerleri ise eğitim, validasyon ve test verileri olarak sırasıyla 0.0010, 0.0018 ve 0.0034 olarak ölçülmüştür.



Şekil 6.13. Regresyon Sonuç Grafikleri (Anomali Dahil)

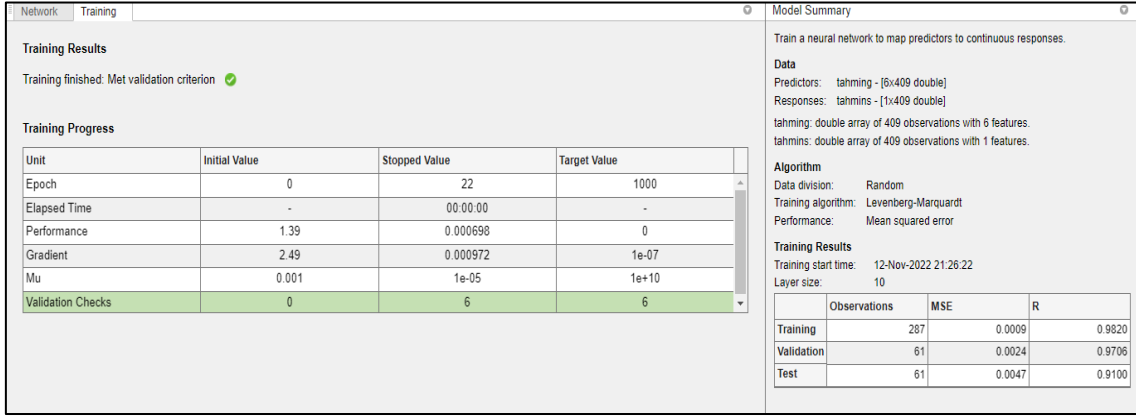
Yukarıdaki çalışmalar incelendiğinde anomali değişkenleri ile tahmin sonuçları %95 in üzerinde gerçekleşmiştir. Çalışmanın son algoritması olarak anomali değişkenleri datadan çıkartılıp yukarıdaki aynı YSA algoritması kurulduğundaki performans sonuçları ise aşağıdaki gibidir. İlk çalışmadan farklı olarak 6 girdi ile devam edildiğinde YSA yapısı aşağıdaki Şekil 6.14'deki gibidir.



Şekil 6.14. YSA Katman Yapısı (Anomali Hariç)

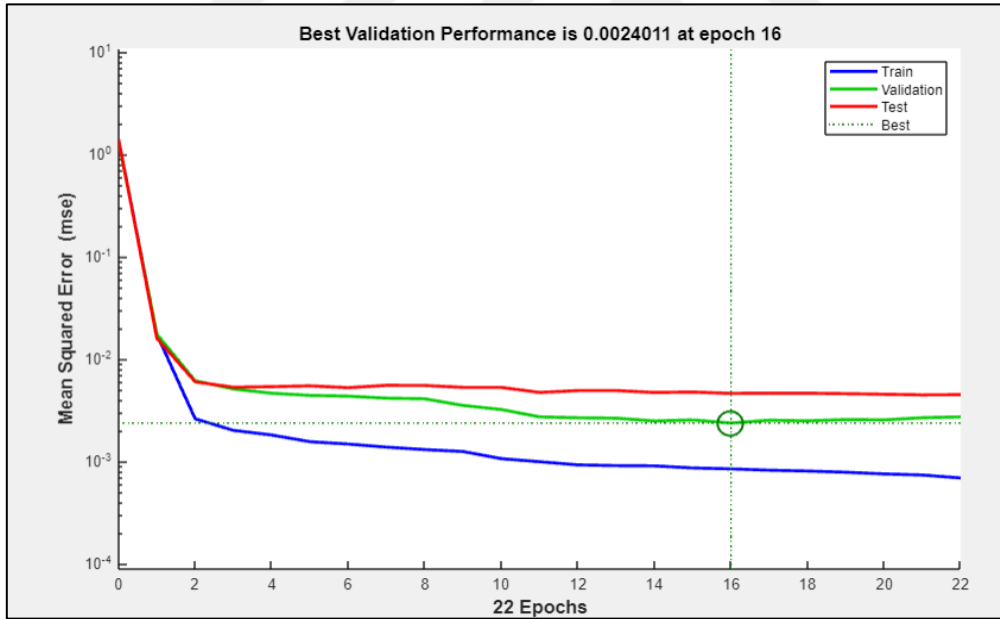
İkinci olarak oluşturulan bu YSA modelinde de ilk model ile aynı şartlar altında model kurulmuş olup eğitiminde durma kriteri olarak "1000" iterasyon alınmıştır. Eğitimi

0.0000001 s sürede 22 iterasyonda tamamlanmış olup 16. İterasyonda en doğru sonuca ulaşmıştır. Kullanılan Matlab YSA arayüzü eğitim sonuçları Şekil 6.15'te gösterilmiştir.



Şekil 6.15. Anomali Hariç YSA Arayüzü Ve Eğitim Sonuçları

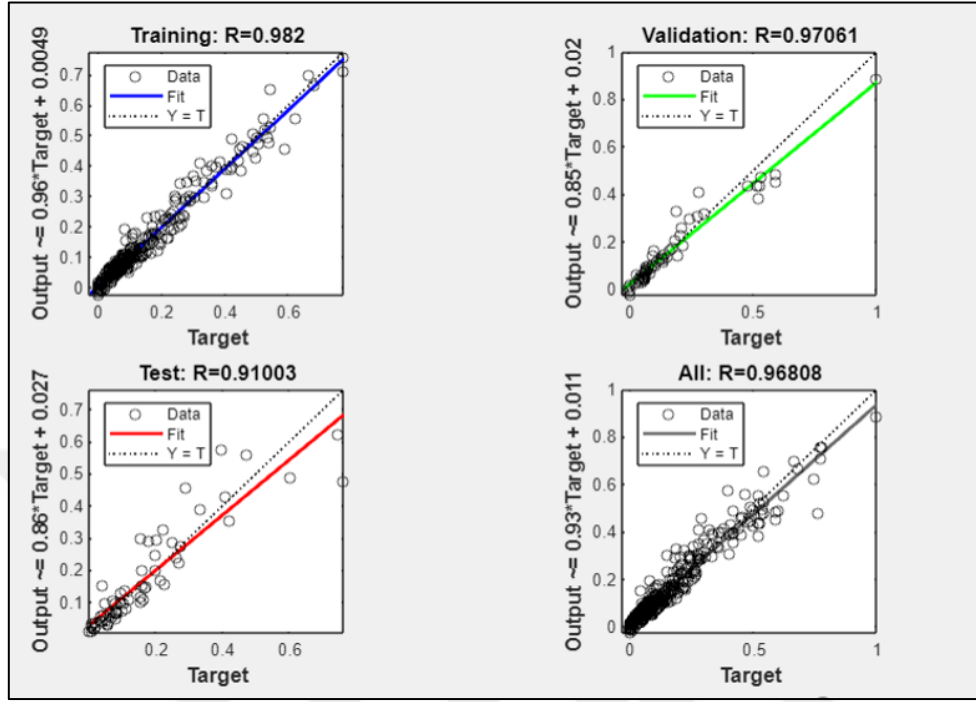
Eğitime ilişkin sonuçlar ve en iyi değer bulunduğ 16. İterasyonunun iterasyona bağlı hata değişim grafiğ i aşağıdaki Şekil 6.16'da gösterilmiştir



Şekil 6.16. Eğitim, Validasyon ve Test Verileri Eğitim Performansı

Modelleme performansı incelendiğ inde gerçek satış değerleri ile YSA' nın tahminlediğ i satış değerlerinin karşılaştırılmalı eğitim, validasyon ve test sonuçlarına ait regresyon grafikleri Şekil 6.17 'de gösterilmektedir. Grafikler incelendiğ inde eğitim regresyon değeri 0.982, validasyon değeri 0.970, test değeri ise 0.910 olarak gerçekleşerek 1'e çok yakın değerler alınmıştır. Tüm bunlar ışığında toplam değeri kıyasladığımız da ise 0.968

değeri ile başarılı bir tahmin gerçekleşmiş olup MSE değerleri ise eğitim, validasyon ve test verileri olarak sırasıyla 0.0009, 0.0024 ve 0.0047 olarak ölçülmüştür.



Şekil 6.17. Regresyon Sonuç Grafikleri (Anomali Hariç)

Anomali dahil ve hariç YSA algoritmalarının sonuçları incelendiğinde ikisi de totalde %96 'ın üzerinde doğruluk sonucu göstererek başarılı sayılmaktadır. Ancak anomalilerin değişken olarak alındığı modelde test sonuçlarının doğruluğu %95 iken anomali hariç modelde %91 olarak gerçekleşmiştir.

7. SONUÇ VE ÖNERİLER

Günümüz koşullarında djitalleşmenin getirdiği artan rekabet koşulları, bilgiye ulaşmanın ve bu bilgileri kıyaslanmanın çok daha kolay olduğu, bunların yanında çevresel koşulların ve farkındalıkların da artmasıyla şirketlerin yaşamsal faaliyetlerine devam etmesi gittikçe daha da zorlaşmaktadır. Daha önce örneği görülmemiş olan yakın zamanda yaşadığımız pandemi döneminde olduğu gibi birçok şirket bu ihtiyacı önceliğine almış olup geleceği tahmin etmek ve geçmişi de en doğru şekilde analiz etmek istemektedir. Bu sebeple veri madenciliği yöntemleri ve yapay zeka teknikleri gün geçtikçe daha da önemli bir hale gelmektedir.

Bu çalışma perakende sektöründe satış tahmini yapılan şirketler için yakın gelecekte çok daha önemli hale gelen, satışa etki eden anomalilerin analiz edilmesi, satışı etkileyen geçmiş dönemli verilerle birlikte anomalilerin tahminin yapılabilirliğinin kontrolü ve anomalilerin satış tahminindeki etkisini veri madenciliği ve yapay sinir ağları ile inceleyerek literatüre katkı sağlamak amacıyla yapılmıştır. Sektörde önde gelen bir giyim markasının 10 grupta toplanan 4 yıllık veri datası kullanılarak çalışmanın ilk kısmında anomali tahmini ve sonrasında da bu anomalilerle beraber talep tahmini yapılmaya çalışılmıştır.

İncelenmiş olan literatür taramasında talep tahmin uygulamalarında sıklıkla kullanılan yapay sinir ağları ve veri madenciliği yöntemlerinin kullanılması çalışma için uygun bulunmuş olup tahmin modellemesi için ana hatlarıyla incelenmiştir. Çalışma iki aşamalı gerçekleşmektedir. 2019-2022 yılları arasındaki 4 yıllık satış verileri aylar ve grup bazında toplandıktan sonra ilk olarak WEKA paket programı kullanılarak veri madenciliği sınıflandırma tahmin yöntemlerinden olan destek vektör makineleri, k-en yakın komşu ve naive bayes yöntemleri ile tahmin çalışması yapılmış olup ikinci olarak da bu anomaliler kullanılarak ve göz ardı edilerek Matlab paket programında yapay sinir ağları yöntemi ile satış tahmini çalışması yapılmıştır. Seçilen bu yöntemler doğrusal ve doğrusal olmayan veri setlerinde sıklıkla kullanılıp etkili sonuçlara ulaşabildiklerinden ilgili çalışmada çözüm yöntemi olarak tercih edilmişlerdir. Dört yöntem için de stok, satış, fiyat, indirim oranı ve anomali değerlerine ek olarak satışlar üzerinde doğrudan etkisi bilinen ‘tüketici güven endeksi’, ‘hava durumu’ ve ‘perakende satış hacim endeksi’

değişkenleri de veri olarak eklenmiştir. Yöntemlerde seçilen parametrelerin sonuçları etkilediği bilindiğinden her yöntemin farklı kriterlerde kombinasyonları denenmiştir.

Uygulamanın ilk kısmında 3 yöntem ile de aynı kombinasyon ve test aşamaları gerçekleştirilerek tahmin çalışması yapılmıştır. Naive bayes algoritması ile anomali tahminlemede en yüksek doğruluk oranı %80 i eğitim %20 si test olarak ayrılmış olan uygulamada gerçekleştiren olup %68 olarak sonuçlanmıştır. Destek vektör makineleri ile analiz edilen tahminleme çalışmasında ise en başarılı doğruluk %90 i eğitim %10 si test olarak ayrılmış olan uygulamada gerçekleştiren olup %78 oran olarak sonuçlanmıştır. Son olarak K-NN ile analiz edilen çalışmada en yüksek doğruluk oranı %90 i eğitim %10 si test olarak ayrılmış olan uygulamada $k=5$ iken gerçekleştiren olup %75 olarak sonuçlanmıştır. 3 algoritmanın sonuçlarını kıyasladığımızda en yüksek doğruluk oranı %78 ile DVM yöntemi ile elde edilmiştir. Literatürdeki çalışmalar incelendiğinde de DVM makineleri ile olan çözüm yöntemleri diğer yöntemlere göre başarılı performans göstermektedir. Literatürde %75 üzerinde doğruluk değeri olan talep tahminleri başarılı sayılmaktadır. Bunu göz önüne aldığımızda elde edilen sonuç literatürde karşılaşılan uygulamalar ile örtüşmektedir. Ancak doğruluk oranı %90'ın altında bir seviyede olduğu için yapay sinir ağları ile satış tahmini yaparken anomali dahil ve ayrı olmak üzere 2 çalışma yapılmış olup anomali değerlerin satışlar üzerinde anlamlı etkisinin var olup olmadığı anlaşılmaya çalışılmıştır.

Uygulamanın ikinci kısmında anomalilerin bağımsız değişken olarak alındığı YSA algoritması ile satış tahmini çalışmasında toplam eğitim regresyon değeri %97 test değeri doğruluk yüzdesi ise %95 olarak gerçekleşmiştir. YSA ile anomaliler göz ardı edilerek yapılan tahmin çalışmasında ise toplam eğitim regresyon değeri %96 test değeri doğruluk yüzdesi ise %91 olarak gerçekleşmiştir. Eğitim ve doğrulama sonuçları her iki uygulamada da %95 üzerinde gerçekleşmesine rağmen test doğruluk değeri anomalilerin de katkısı ile %4 daha iyi performans göstermiş olup test MSE değerleri ise anomaliler dahil olan çalışmada 0.0034 iken anomaliler hariç olan uygulamada 0.0047 olarak ölçülmüştür. %4 'lük bir iyileşme bütçe çalışmalarında yaklaşık 1,5 milyon TL'lik bir kaybın önüne geçmekte olup aynı zamanda yanlış adette ürün üretimi ve stok tutma maliyetlerini de azaltmaktadır. Şirketin zaman ve kar yönetimi olarak daha iyi stratejik kararlar vermesine olanak sağlamaktadır.

Yapılan bu çalışma ile beraber gerçekleşen anomalilerin, satış tahmini için toplanan data'lara önemli bir girdi sağladığı kanıtlanmış olmaktadır. Aynı zamanda daha yüksek doğruluk oranlı anomali tahmini çalışmalarının yapılabilmesi için sayısal oranlarla anomali değerleri desteklenerek yapılacak benzer çalışmalara yol gösterici bir nitelik taşıdığına inanılmaktadır. Çalışmada satış tahmini uygulamalarında anomali değişkenlerinin önemini yapay sinir ağları ve sınıflandırma tahmin algoritmalarından da destek vektör makinelerinin diğer algoritmalara göre daha başarılı sonuçlar elde ettiği gösterilmiş olup literatüre katkı sağlanmıştır. Tüm bunlara ek olarak yapılan doğru tahmin sonuçları ile çalışmada kullanılan farklı ürün grupları için ayrı ayrı avantajlar sağlanmakta olup bu çalışmadan yola çıkılarak maliyet analizleri ve bir sonraki dönem için doğru yatırım çalışmaları yapılabilecektir. Bu çalışma 10 ürün grubunda (denim,dokuma vb.) alt toplam satış adeti baz alınarak tamamlanmıştır. Ancak ürün grubu bazında daha detaylı tahmin çalışması yapılması istenmesi durumunda tek bir ürün grubunda da ilgili çalışma yol gösterici niteliğindedir.

Gelişen ve gelişmekte olan bir firma için talep tahminini sürdürülebilir şekilde karşılayabilmesi oldukça önemlidir. Doğru ürünün doğru zamanda, doğru adette ve doğru yerde olmasının sağlanması, yapılan bütçe çalışmalarının gelecekte olası gerçekleşebilecek anomalilerden haberdar olarak minimum maliyet ve optimum kazançla giderilebilmesi için doğru satış ve anomali tahminine ihtiyaç vardır. Tahmin çalışmaları gelecek yatırımların değerlendirilmesinde, kar politikalarının belirlenmesinde, perakende sektörünün işleyişinde oldukça önemli olduğundan şirketlerin farklı tahmin çalışmalarını göz önünde bulundurarak karar vermeleri gerekmektedir. Elde edilen tahmin değerleri uzun dönemli ve karlı satış politikalarının belirlenme sürecinde etkili olacaktır.

KAYNAKLAR

- Acı, M., Doğansoy, G.A. (2022). Makine Öğrenmesi Ve Derin Öğrenme Yöntemleri Kullanılarak E-Perakende Sektörüne Yönelik Talep Tahmini. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 37(3), 1325-1340.
- Adıyaman, F. (2007). Talep Tahmininde Yapay Sinir Ağlarının Kullanılması, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, 223596
- Aghalarova, S., & Keser, S. B. (2021). Önerilen Yapay Sinir Ağı Algoritması ile Ortaokul Öğrencilerin Akademik Performansının Tahmini. *Veri Bilimi*, 4(2), 19-32.
- Ajak, A. D., Lilford, E., & Topal, E. (2018). Application Of Predictive Data Mining To Create Mine Plan Flexibility In The Face Of Geological Uncertainty. *Resources Policy*, 55, 62-79.
- Akbulut, S. (2006). Veri Madenciliği Teknikleri İle Bir Kozmetik Markanın Ayrılan Müşteri Analizi Ve Müşteri Segmentasyonu. Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, Gazi Üniversitesi, Ankara, 180239.
- Akkoç, B. (2012). Sosyal Ağ Analizi İçin Bayes Ağlarının Kullanımı. Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, Selçuk Üniversitesi, Konya.
- Akkuş, B., & Zontul, M. (2019). Veri Madenciliği Yöntemleri İle Ülkeleri Gelişmişlik Ölçütlerine Göre Kümeleme Üzerine Bir Uygulama. *AURUM Journal Of Engineering Systems And Architecture*, 3(1), 51-64.
- Akpınar, H. (2000). Veri Tabanlarında Bilgi Keşfi Ve Veri Madenciliği. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 29(1), 1-22.
- Alagöz, A., Öge, S. Ve Ortakarpuz, M. (2014). Bir Kurumsal Zekâ Teknolojisi Olarak Veri Madenciliği İle Muhasebe Bilgi Sistemi İlişkisi. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 1-21.
- Alkaragole, M. L. Z., & Kurnaz, S. (2019). Comparison Of Data Mining Techniques For Predicting Diabetes Or Prediabetes By Risk Factors. *International Journal Of Computer Science And Mobile Computing*, 8, 61-71.
- Alpaydın, E., 2000, Zeki Veri Madenciliği; Ham Veriden Altın Bilgiye Ulaşma Yöntemleri, [Ziyaret Tarihi: 4 Nisan 2022].
- Altay, A. (2019). Üniversite Kütüphanelerinde Veri Madenciliği Uygulamaları: Kırklareli Üniversitesi Kütüphane Ve Dokümantasyon Daire Başkanlığı Örneği. *Ağrı İbrahim Çeçen Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 5(2), 287-316.
- Armutlu, Ş., (2018), Veri Madenciliği İle Kütüphane Kullanımı Ve Ders Başarısı Arasındaki İlişkinin İncelenmesi, Yüksek Lisans Tezi, Uşak Üniversitesi, Sosyal Bilimler Enstitüsü, Uşak, 531486.

- Arslan, O.M. (2020). Veri Madenciliği İle Talep Tahmini Ve Bir İşletmede Uygulama. Yüksek Lisans Tezi, Gazi Üniversitesi, Bilim Enstitüsü, Ankara, 672695.
- Atmaca, M. (2020). Alüminyum Ve Prinç Alaşımli Malzemelerin Isı İletim Katsayısının Bulanık Yapay Sinir Ağları Yöntemi Kullanılarak Tespit Edilmesi. Yüksek Lisans Tezi, Gaziantep Üniversitesi, Fen Bilimleri Enstitüsü, Gaziantep, 633207.
- Avcı, H. (2021). Tıbbi Görüntü İşleme ile Tanı Koymada Veri Madenciliği ve Derin Öğrenme Yöntemlerinin Performanslarının İncelenmesi.
- Ayan, B., Can, M., & Gürsoy, U. T. (2019). Sosyal Medya Madenciliği İle Firmaların Twitter Verilerinin İncelenmesi. *International Journal Of Economic & Administrative Studies*, (23).
- Aydın, B. Z., & Süssler, B. (2022) Finansal Başarısızlığın Yapay Sinir Ağları Ve Çok Değişkenli İstatistiksel Analiz Teknikleri İle Tahmin Edilmesi: Borsa İstanbul'da Bir Uygulama.
- Ayhan, S (2013). Kaba Küme Ve Destek Vektör Makineleri Kullanılarak Nitelik İndirgeme Ve Sınıflandırma Problemlerinin Çözümü İçin Bütünleşik Bir Yaklaşım. Doktora Tezi, Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü, Eskişehir.
- Ballı, Müzeyyen Tuğba, (2014). "Yapay Sinir Ağları İle Talep Tahmin Ve Gıda Sektöründe Uygulanması", Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Bilgiç, E., & Esen, M. F. (2018). Endüstri 4.0 Işığında Veri Madenciliği Ve Pazarlama: Son Gelişmeler, Yeni Trendler. *İşletme Ekonomi Ve Yönetim Araştırmaları Dergisi*, 1(2), 21-29.
- Büyükkatak, E. (2022). PISA 2018 Türkiye Örneğinde Okuma Okuryazarlık Düzeylerinin Farklı Veri Madenciliği Sınıflandırma Yöntemleri İle İncelenmesi.
- Büyükkeçeci, M., & Okur, M. C. Müşteri Kayıplarının Tahmini Üzerine Bir Veri Madenciliği Uygulaması. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen Ve Mühendislik Dergisi*, 24(72), 887-900.
- Canbulut, G. Pandemi Sürecinde Şehir İçi Yolcu Taşıma Sayılarının Veri Madenciliği Yöntemi İle Belirlenmesi: Kayseri Örneği. *Mühendislik Bilimleri Ve Tasarım Dergisi*, 10(3), 1043-1053.
- Chen, D., Cao, L., & Si, H. (2020). Benchmark Value Determination Of Energy Efficiency Indexes For Coal-Fired Power Units Based On Data Mining Methods. *Advanced Engineering Informatics*, 43, 101029.
- Çetin, E. 2003. Yapay Sinir Ağları (Kuram, Mimari, Eğitim, Uygulama). Seçkin Yayıncılık
- Çörekcioğlu, M., Ercan, E., & Elibüyük, S. A. (2021). Yapay Sinir Ağı Yöntemlerinin Tekstil Sektöründe Kullanım Uygulamaları. *Teknik Bilimler Dergisi*, 11(2), 14-20.

- Demiral, G., Soba, M. Ve Armutlu, G. (2017). Kütüphane Veri Tabanında Veri Madenciliği: Uşak Üniversitesi Örneği. *Bartın Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi*, 8(16), 241-264.
- Deniz, E. (2021). Yapay Sinir Ağları Ve K-En Yakın Komşu Algoritması İle Toprak Çeşitliliğinin Belirlenmesi. Yüksek Lisans Tezi, Lisansüstü Eğitim Enstitüsü, İstanbul Arel Üniversitesi, İstanbul, 719158.
- Efe, Önder Ve Kaynak, Okyay (2000). Yapay Sinir Ağları Ve Uygulamaları. *İstanbul: Boğaziçi Üniversitesi Yayınevi*.
- Elmas, M. (2012). Destek Vektör Makineleri İle Fiyat Tahminleri Ve Kuyumculuk Sektöründe Bir Uygulama. *İstanbul Üniversitesi*.
- Erbudak, A. E. (2022). Veri Madenciliği Ve Makine Öğrenimi İle Döviz Kuru Tahmini Uygulaması, Yüksek Lisans Tezi , Altınbaş Üniversitesi/Lisansüstü Eğitim Enstitüsü, İstanbul, 726738.
- Ercan, U. (2021). Ev Dışı Gıda Tüketim Sınıflarının Yapay Sinir Ağları İle Tahmin Edilmesi. *İşletme Araştırmaları Dergisi*, 13(4), 3265-3277.
- Erdal, H. İ. (2011). Destek Vektör Makineleri İle Tahmine Dayalı Modelleme Ve Bir Uygulama, Doktora Tezi, İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Erdem, M. 2007. Motorlu Taşıtlarda Yavaşlatıcı Etkilerinin Deneysel Analizi Ve Yapay Sinir Ağları İle Modellenmesi. Doktora Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara
- Eren, U., & Satoglu, S. I. (2017). Perakende Giyim Sektöründe Yapay Sinir Ağları İle Talep Tahmini. *Endüstri-İşletme Kurultayı*.
- Erken, Ş., (2017) Veri Madenciliği Yöntemleri Ve Optimizasyona Dayalı Modeller Üzerine Bir Araştırma Ve Bir Uygulaması, Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, Sosyal Bilimler Enstitüsü, İzmir, 475824.
- Es, H. A., (2013). “Yapay Sinir Ağları İle Türkiye Net Enerji Talep Tahmini”, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Ankara.
- Gencer, C. (2019). Veri Ambarlama Ve Veri Madenciliği. 2018-2019 Bahar Yarıyılı Ders Notları. Gazi Üniversitesi, Bilişim Enstitüsü, Ankara.
- Gençosman, B. Ç., & Şanlı, G. E. (2021). Arıtma Çamurlarında Polisiklik Aromatik Hidrokarbonların (PAH'ların) Giderimlerinin Veri Madenciliği Yöntemleri İle Tahmini.
- Gorunescu, F. (2011). Data Mining: Concepts, Models And Techniques (Vol. 12). *Springer Science & Business Media*.

- Gökçe, H., Sönmez, E. F., Selen, A. V. C. I., & Aladağ, Z. Uygun Normalizasyon Tekniği Ve Yapay Sinir Ağları Analizi İle Otomobil Satış Tahminlemesi. *İşletme Ekonomi Ve Yönetim Araştırmaları Dergisi*, 5(1), 19-45.
- Güler, K. (2019). Türkiye'deki Altın Ve Döviz Fiyatlarının Değişiminin Ekonomik Veriler Kullanarak Yapay Zeka İle Tahmini Ve Kriz Öngörüsü. Yüksek Lisans Tezi, Yalova Üniversitesi, Fen Bilimleri Enstitüsü, 588110.
- Gültepe, Y. (2019). Makine Öğrenmesi Algoritmaları İle Hava Kirliliği Tahmini Üzerine Karşılaştırmalı Bir Değerlendirme. *Avrupa Bilim Ve Teknoloji Dergisi*, 16, 8-15.
- Güner, D., (2018). Veri Madenciliği Yöntemi Olarak Bulanık Mantık İle Hava Kalitesi Değerlendirmesi, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 526552.
- Güven, İ. (2020). Perakende Hazır Giyim Endüstrisinde Yapay Zeka Yöntemleri İle Talep Tahmini (Doctoral Dissertation).
- Han, J., Kamber, M., Pei J. (2012). Data Mining Concepts And Techniques. USA: *Morgan Kaufmann Publishers*. 1-5,6-8,1,18-19,83-87, 99, 243, 327-393, 398-400, 408-413, 423, 426, 44-448, 450-457.
- Hand, D., Mannila, H., Smyth, P. (2001). Principles Of Data Mining. Cambridge: *MIT Press*.
- Hamzaçebi, C. (2008). Mevsimsel Zaman Serisi Tahmininde Yapay Sinir Ağlarının Performansının İyileştirilmesi. *Bilgi Bilimleri* , 178 (23), 4550-4559.
- Hazım, L,R. (2018). Four Classification Methods Naïve Bayesian, Support Vector Machine, Knearest Neighbors And Random Forest Are Tested For Credit Card Fraud Detection. *Altınbas University Graduate School Of Sciences Engineering*
- Hebb, D. O., 1949. The Organization Of Behaviour, Wiley, Newyork
- Iliç, M., Srdjević, Z., & Srdjević, B. (2022). Water Quality Prediction Based On Naïve Bayes Algorithm. *Water Science And Technology*, 85(4), 1027-1039.
- James G, Witten D, Hastie T, Tibshirani R. Support Vector Machines. Casella G, Fienberg S, Olkin I, Editors. An Introduction To Statistical Learning With Applications In R. New York: Springer; 2013.
- Kadir, Kaya, Şenel, M. C., & Erdem, K. O. Ç. (2018). Perakende Ticaret Sektörünün Türkiye'deki Genel Durumu. *Akademik Araştırmalar Ve Çalışmalar Dergisi* (Akad), 10(19), 502-515.
- Kalemci, Ö., & Timor, D. M. (2018). Veri Madenciliği Yöntemi İle Prostat Kanseri İçin Erken Uyarı Protokollerinin Geliştirilmesi.

- Kantardzic, M. (2011). *Data Mining: Concepts, Models, Methods, And Algorithms*. John Wiley & Sons.
- Karaatlı, M., & Altıntaş, E. (2018). Borsa İstanbul İşletmelerinin Veri Madenciliği İle Kümelenmesi-Clustering The Companies Listed On Stock Exchange Istanbul By Data Mining. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10(26), 871-886.
- Karaođlan, S. (2012). Bazı Mühendislik Problemlerinde Yapay Sinir Ağları Ve Bulanık Mantığın Uygulanması, Gaziantep Üniversitesi Matematik Bölümü, Yüksek Lisans Tezi, Ankara.
- Kartalopoulos, Stamatios V. (1996). *Understanding Neural Networks And Fuzzy Logic: Basic Concepts And Applications*. New York: IEEE Press.
- Kaşıkcı, M. (2019). Transkriptom Veri Seti Üzerinde Derin Öğrenme Yöntemi İle Klasik Veri Madenciliği Yöntemlerinin Sınıflama Performanslarının Karşılaştırılması.
- Kaya, K. (2016) Destek Vektör Makineleri Yardımıyla Tüketici Kredilerinin Sınıflandırılması (Doctoral Dissertation, Fen Bilimleri Enstitüsü).
- Kılıç, İ., & Önal, S., (2022) Finansal Hilelerin Tespit Edilmesinde Kullanılan Veri Madenciliği Yöntemleri Ve Borsa İstanbul'da Bir Uygulama. *Muhasebe Ve Denetim Bakış*, 22(67), 181-208.
- Köseođlu, Ö., & Demirci, Y. (2018). Akıllı Şehirler Ve Yerel Sorunların Çözümünde Yenilikçi Teknolojilerin Kullanımı. *Uluslararası Politik Araştırmalar Dergisi*, 4(2), 40-57.
- Kwon, S. J. (2011). *Artificial Neural Networks*. Nova Science Publishers, Inc.
- Larose, T.D. (2005). *Discovering Knowledge In Data: An Introduction To Data Mining*. New Jersey : A. John Willey&Sons, Inc.
- Li, S., Li, H., Li, M., Shyr, Y., Xie, L. and Li, Y., 2009, Improved prediction of lysine acetylation by support vector machines, *Protein and peptide letters*, 16, 977-983.
- Maden, E. (2021). Tıp Fakültesi Öğrencilerinin Kurul Sınavı Başarılarının Veri Madenciliği Algoritmaları Kullanılarak İncelenmesi, Yüksek Lisans Tezi, Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, Kocaeli.
- Mcculloch, W.S. Pitts, W. (1943). A Logical Calculus Of Ideas İmminent İn Nervousactivity, *Bull. Math. Biophysics*, 5, 115-33.
- Metlek, S., & Kayaalp, K. (2020). Otistik Spektrum Bozukluğunun Makine Öğrenme Algoritmaları İle Tespiti. *Journal Of Intelligent Systems: Theory And Applications*, 3(2), 60-68.

- Mohideen, D. F. M., Raj, J. S. S., & Raj, R. S. P. (2022). Regression İmputation And Optimized Gaussian Naïve Bayes Algorithm For An Enhanced Diabetes Mellitus Prediction Model. *Brazilian Archives Of Biology And Technology*, 64.
- Montella, A., De Oña, R., Mauriello, F., Riccardi, M. R., & Silvestro, G. (2020). A Data Mining Approach To Investigate Patterns Of Powered Two-Wheeler Crashes İn Spain. *Accident Analysis & Prevention*, 134, 105251.
- Nagy, V., Horváth, B., & Horváth, R. (2017). Land-Use Zone Estimation İn Public Transport Planning With Data Mining. *Transportation Research Procedia*, 27, 1050-1057.
- Oruç, E.N., (2022) Yapay Sinir Ağları Kullanılarak Kısa Süreli Güneş Enerjisi Tahmini, İstanbul Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü, Yüksek Lisans Tezi, İstanbul, 706964.
- Oğuz, A. T. A., & Erbudak, A. E. Veri Madenciliği Ve Makine Öğrenimi İle Döviz Kuru Tahmini Uygulaması. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 34(2), 553-563.
- Özekes, Ö.(2003). Veri Madenciliği Modelleri Ve Uygulama Alanları, *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, Cilt:2, Sayı :3, Sayfa:65-82
- Özdemir, A., Aslay, F.Y., Çam, H., Veritabanında Bilgi Keşfi Süreci: Gümüşhane Devlet Hastanesi Uygulaması, *SÜ İİBF Sosyal Ve Ekonomik Araştırmalar Dergisi*, Sayı 20, 347-365, 2010.
- Özdemir, G. Z. (2013). Tüketici güveninin tüketim harcamaları ile ilişkisi ve öngörü gücü: türkiye örneği. Uzmanlık Yeterlilik Tezi, Türkiye Cumhuriyet Merkez Bankası İletişim ve Dış İlişkiler Genel Müdürlüğü, Ankara,
- Özgür, Y. (2021). Elektrik Piyasasında Fiyat Tahmin Modelleri.
- Öztemel, E., 2003. Yapay Sinir Ağları, Papatya Yayıncılık, İstanbul
- Öztemel, E. (2016). Yapay Sinir Ağları, 1 St, Papatya Bilim, İstanbul, S. 34-35
- Öztürk, S. (2020). Hızlı Moda Sektöründe Makine Öğrenmesi Yöntemleri İle Satış Miktarlarının Tahmin Edilmesi (Doctoral Dissertation, Fen Bilimleri Enstitüsü).
- Peromingo, B., Caballero, D., Rodriguez, A., Caro, A., & Rodriguez, M. (2020). Application of data mining techniques to predict the production of aflatoxin B1 in dry-cured ham. *Food Control*, 108, 106884.
- Polat, E.D., (2022), Beyaz Eşya Sektöründe Satış Tahmini Bir Veri Madenciliği Uygulaması, Yüksek Lisans Tezi, Bursa Uludağ Üniversitesi Sosyal Bilimler Enstitüsü , Bursa, 718986.

- Qi, X., Mei, G., Cuomo, S., Liu, C., & Xu, N. (2021). Data analysis and mining of the correlations between meteorological conditions and air quality: A case study in Beijing. *Internet of Things*, 14, 100127.
- Rençber, Ö. F. (2018). Basamak Korelasyon, Kohonen ve ANFIS Yapay Sinir Ağ Modellerinin Sınıflandırma Performanslarının Karşılaştırılması: Lojistik Performans Endeksi Üzerine Uygulama. *Ege Academic Review*, 18(3), 521-535.
- Rygielski, C., Wang, J.-C. ve Yen, D. C., (2002). Data Mining Techniques for Customer Relationship Management. *Technology in Society*, 24 (4), 483-502.
- Sandıkçı, Y., (2022), Pamuk üretim rekolte tahmini için veri madenciliği yöntemlerinin kullanılması, Yüksek Lisans Tezi, Harran Üniversitesi, Fen Bilimleri Enstitüsü, Şanlıurfa.
- Savaş, S., Topaloğlu, N., Yılmaz, M., Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri", *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, No:21, 5, 2012.
- Sebik, N.B., (2018), VERİ Madenciliği Yöntemi İle Akciğer Kanseri Teşhisine Yönelik Öneri Sistemi Tasarımı, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 493624.
- Şengül, C. A. N., Özdil, T., & Yılmaz, C. (2019). Üniversite Eğitiminde Öğretim Üyesi Tutumunun Ders Başarısına Etkisinin Veri Madenciliği Teknikleriyle Araştırılması. *Yönetim ve Ekonomi Dergisi*, 26(2), 609-623.
- Şık, M.Ş., (2014), Veri Madenciliği Ve Kanseri Erken Teşhisinde Kullanımı, Yüksek Lisans Tezi, İnönü Üniversitesi, Sosyal Bilimler Enstitüsü, Malatya, 388086.
- Tan, Q., Mu, X., Fu, M., Yuan, H., Sun, J., Liang, G., & Sun, L. (2022). A new sensor fault diagnosis method for gas leakage monitoring based on the naive Bayes and probabilistic neural network classifier. *Measurement*, 194, 111037.
- Taşdemir, M. (2012). Öğrenci Başarısına Etki Eden Faktörlerin Regresyon Analizi İle Tespiti. Yüksek Lisans Tezi, Dicle Üniversitesi Sosyal Bilimler Enstitüsü, Diyarbakır.
- Tekin, Mahmut (2009). Üretim Yönetimi. (6. Baskı). Konya: Günay Ofset.
- Tomasevic, N., Gvozdencovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & education*, 143, 103676.
- Tuna, M. (2019). Altın Fiyatlarının Yapay Sinir Ağları ve Arma Modelleri ile Tahminlenmesi. Yüksek Lisans Tezi, Adnan Menderes Üniversitesi, Sosyal Bilimler Enstitüsü, Aydın, 589740.
- TÜİK, <https://www.tuik.gov.tr/>, (Erişim Tarihi: 10.04.2022)

- Türkođlu, B., Komesli, M., & Ünlütürk, M. S. An industrial case study on data mining. In International management information systems conference, *Ankara-Türkiye* (pp. 107-110).
- Uçarsu, O., (2022), Yapay Sinir Ağları Kullanarak Dolar Kuru Gelecek Deđer Tahmini Yapılması, Yüksek Lisans Tezi, Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, Kocaeli.
- URL-1: <https://kodedu.com/>, (Ziyaret tarihi: 22 Kasım 2022)
- Van Nguyen, T., Zhou, L., Chong, A. Y. L., Li, B., & Pu, X. (2020). Predicting Customer Demand For Remanufactured Products: A Data-Mining Approach. *European Journal Of Operational Research*, 281(3), 543-558.
- Vural, D., KÜÇÜKAŞCI, E. Ş., & Ayvaz, B. (2019). Türkiye'de Perakende Sektöründe Elektronik Ürönlere Olan Talebin Yapay Sinir Ađı ile Modellenmesi. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 18(35), 1-16.
- Weinberger, K.Q. ve ark., (2006), Distance Metric Learning for Large Margin Nearest Neighbor Classification, Department of Computer and Information Science, University of Pennsylvania
- Yıldırım, M., (2022) Yapay Sinir Ağları Kullanarak Dolar Kuru Gelecek Deđer Tahmini Yapılması, Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Kocaeli, 716527.
- Zazzaro, G., Cuomo, S., Martone, A., Montaquila, R. V., Toraldo, G., & Pavone, L. (2021). Eeg signal analysis for epileptic seizures detection by applying data mining techniques. *Internet of Things*, 14, 100048.

KİŞİSEL YAYIN VE ESERLER

Çetin E., Avcı S., Alkan A., Aladağ Z., Ashby, Ahp Ve Aksiyomatik Tasarım Yöntemleri İle Kürek Malzemesi Seçimi, *Uluslararası Marmara Fen Ve Sosyal Bilimler Kongresi (IMASCON)*, Kocaeli, 2 Kasım 2019.



ÖZGEÇMİŞ

İlk, orta ve lise öğrenimini Kütahya 'da tamamladı. 2014-2018 yılları arasında Kocaeli Üniversitesi Mühendislik Fakültesi Endüstri Mühendisliği Bölümü'nde lisans eğitimini tamamladı. 2019 yılında Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Endüstri Mühendisliği Ana Bilim Dalı'nda yüksek lisans öğrenimine başladı. 2020 Ocak ayından bu yana perakende sektöründe hizmet veren bir firmada Planlama Mühendisi olarak görevini sürdürmektedir.

