T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI
YÖNETİM BİLİŞİM SİSTEMLERİ PROGRAMI
YÜKSEK LİSANS TEZİ

KÜÇÜK VE ORTA ÖLÇEKLİ FİRMALAR İÇİN MAKİNE ÖĞRENMESİ DESTEKLİ KARAR DESTEK SİSTEMİ

Kemal ÇAM

Danışman
Doç. Dr. Can AYDIN

İZMİR - 2021

TEZ ONAY SAYFASI

YEMİN METNİ

Yüksek Lisans Tezi olarak sunduğum “Küçük ve Orta Ölçekli Firmalar için Makine Öğrenmesi Destekli Karar Destek Sistemi” adlı çalışmanın, tarafımdan, akademik kurallara ve etik değerlere uygun olarak yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

..../..../2021

Kemal ÇAM

ÖZET

Yüksek Lisans Tezi

Küçük ve Orta Ölçekli Firmalar İçin

Makine Öğrenmesi Destekli Karar Destek Sistemi

Kemal ÇAM

Dokuz Eylül Üniversitesi

Sosyal Bilimler Enstitüsü

Yönetim Bilişim Sistemleri Anabilim Dalı

Yönetim Bilişim Sistemleri Programı

Günümüzde bilgi teknolojilerinde meydana gelen gelişmeler ile birlikte, hızla artan verinin yönetilmesi ve iş süreçlerine dâhil edilmesi işletmeler için büyük önem arz etmektedir. Bu yüzden verilerin işlenerek analiz etmeye hazır bir duruma getirilmesi, şirketlerin ellerindeki kaynakları etkin bir şekilde kullanmasını, daha iyi planlama yapılabilmesini sağlamaktadır. Aynı zamanda veriler, karar verme sürecine yardımcı olması, karmaşık süreçlerin anlaşılması, problemlerin tanımlanması ve çözüme kavuşturulması noktasında da şirketlere destek olmaktadır.

Bu çalışma kapsamında küçük ve orta ölçekli firmaların karar verme sürecini desteklemek amacıyla, elindeki verileri daha etkin kullanabilmelerini sağlamak ve verilerden bilgi çıkarımı yapabilecekleri bir sistem tasarlanmıştır. Hazırlanan çalışmada, en başta ham verinin veri önişleme, son aşamada tahminlenmesine kadar tüm süreç planlanmıştır ve buna uygun şekilde uygulama tasarlanmıştır. Uygulama eldeki verilerden karar vericilere anlık ve geleceğe yönelik tahminlemeler yapmasına yardımcı olmayı amaçlamaktadır. Uygulamada test için 200 bin verilik bir veri seti kullanılmıştır. Bu veriler en başta veri düzenleme ekranında temizlenmiş ve kullanıma hazır hale getirilmiştir. Temelde dört ana başlık altında bu verilerden rapor elde edilebilmektedir. Bunlar sırasıyla raporlardan verilerin analizini yapabileceğimiz açıklayıcı raporlar, makine öğrenmesine yöntemlerine dayalı tahminleme raporları, zaman içerisinde verilerin alabileceği yeni değerleri

analiz edebilen zaman serisi analizi ve son olarak verileri benzerliklerine göre gruplama yapabildiğimiz kümeleme analizi raporlarıdır. Tüm bu analizler ve tahminlemeler çeşitli grafiklere ve tablolara dökülmüş karar vericiler için raporlanmıştır.

Anahtar Kelimeler: Karar Destek Sistemleri, Veri Analizi, Makine Öğrenmesi, Zaman Serisi Analizi, Kümeleme Analizi, Yönetim Bilişim Sistemleri

ABSTRACT

Master's Thesis

Machine Learning Supported Decision

Support System for Small and Medium Sized Companies

Kemal ÇAM

Dokuz Eylul University

Graduate School of Social Sciences

Department of Management Information Systems

Management Information Systems Program

Nowadays, with improvements occurring in information technology, management of rapidly growing data and incorporate into business processes it is of great importance for businesses. Therefore, processing the data and making it ready for analysis enables companies to use their resources effectively and to make better planning. At the same time, data supports companies in helping the decision-making process, understanding complex processes, identifying and solving problems.

Within the scope of this study, a system has been designed to support the decision-making process of Small and Medium Sized companies, to enable them to use the data they have more effectively and to make predictions from the data. In the prepared study, the all process from organizing the raw data to its final estimation has been planned and the application has been designed properly. The application aims to help decision makers make instant and future predictions from the available data. In practice, a data set of 200 thousand data was used for testing. This data was first cleared on the data editing screen and made ready for use. Basically, reports can be obtained from these data under four main headings. These are descriptive reports where we can analyze instant data from the reports, forecasting reports based on machine learning methods, time series analysis that can analyze the new values that the data can receive over time, and finally cluster analysis reports where we can group the data

according to their similarities. All these analyzes and estimates have been reported for decision makers in various graphs and tables.

Keywords: Decision Support System, Data Analysis, Machine Learning, Time Series Analysis, Cluster Analysis, Management Information Systems

KÜÇÜK VE ORTA ÖLÇEKLİ FİRMALAR İÇİN MAKİNE ÖĞRENMESİ DESTEKLİ KARAR DESTEK SİSTEMİ

İÇİNDEKİLER

İKİNCİ BÖLÜM
TEORİK ÇERÇEVE

ÜÇÜNCÜ BÖLÜM
ARAŞTIRMA YÖNTEMİ VE KÜÇÜK VE ORTA ÖLÇEKLİ FİRMALAR İÇİN KARAR DESTEK SİSTEMİ GELİŞTİRİLMESİ

KISALTMALAR

ALARM	A Logical Alarm Reduction Mechanism (Mantıksal Alarm Azaltma Mekanizması)
AR	Otoregresif
ARIMA	Birleştirilmiş Otoregresif Hareketli Ortalama
ARMA	Otoregresif Hareketli Ortalama
BT	Bilişim Teknolojileri
CHAID	Chi-squared Automatic Interaction Detector (Ki-kare otomatik etkileşim algılama)
COBWEB	Örümcek Ağı
CSV	Comma-separated Value (Virgülle Ayrılmış Değerler)
DBSCAN	Density-based spatial clustering of applications with noise (Gürültü uygulamaların Yoğunluk tabanlı uzaysal küme)
DVM	Destek Vektör Makineleri
EKK	En Küçük Kareler Yöntemi
KDS	Karar Destek Sistemleri
KNN	K-Nearest Neighbor (K En Yakın Komşu)
KOBİ	Küçük ve Orta Ölçekli İşletmeler
MA	Hareketli ortalamalar
MKDS	Mekansal Karar Destek Sistemleri
OPTICS	Ordering points to identify the clustering structure (Kümeleme yapısını tanımlamak için noktaları sıralama)
PCA	Principal component analysis (Temel Bileşen Analizi)
s.	Sayfa
SVR	Support Vektor Regression (Destek Vektör Makinaları)
t.y.	Tarih yok
vd	Ve Diğerleri
YSA	Yapay Sinir Ağları

ŞEKİLLER LİSTESİ

GİRİŞ

Küreselleşen günümüz dünyasında bilgi ve teknolojinin toplumdaki yeri ve önemi yadsınamaz bir gerçektir. Bulunduğumuz çağın bilgi çağı, toplumumuzun bilgi toplumu, insanlarımızın ise bilgi çalışanları olarak nitelendirildiğini düşündüğümüzde, bilgiyi, işletmeler için en az sermaye kadar etkili ve gerekli olan yeni üretim faktörü olarak, teknolojiyi ise onun vazgeçilmez bir ögesi olarak tanımlamak mümkündür (Özdemirci ve Aydın, 2007). Bu sebeple de bilginin yönetimi, kurumsal yönetimin en önemli ögelerinden biridir.

Günümüzde, verilerin toplanması, toplanan verilerin analiz edilerek bilginin üretilmesi, üretilen bilgilerin kullanılması ve geliştirilmesi, bunların belirli görevleri yürüten kişilere ve dolayısıyla iş süreçlerine aktarılmasını ve katma değer yaratılmasının KOBİ'ler tarafından sağlanması işletmeler için büyük önem arz etmektedir (Özdemirci ve Aydın, 2007).

Bu nedenle, en başta veriyi iyi anlamak ve bu veriden anlamlı bilgi üretme aşamalarını iyi kavramak gerekmektedir. Veri, ölçüm, deney, gözlem vb. gibi yöntemlerle elde edilen, enformasyonun işlenmemiş en temel yapısıdır. Veriler biçimlerine göre sayısal, kategorik ve metin olarak sınıflandırabilir. Bilgiye göre daha dar bir anlam ifade eden ve bir hedefe yönelik olan enformasyon ise, herhangi bir konu çerçevesinde düzenlenmiş bilgi parçası olarak tanımlamak mümkündür. Bilgi ise tüm bu süreçler sonunda meydana gelen, enformasyona dönüştürülmüş verilerin biçimlendirilmiş ve anlamlandırılmış halidir. Karar verme sürecinde bilgi kullanılır (Durna ve Demirel, 2008).

Karar veren kişinin farklı pek çok seçenekle karşı karşıya geldiği karar verme işlemi, kişinin kendisi tarafından belirlenmiş ölçütlere ve kendi amaçlarına uygun olanı seçme sürecidir (Tekin,1996).

Karar verme fonksiyonunun yerine getirilmesi için sağlam ve güvenilir bilgilere ihtiyaç duyulmaktadır. Bu da ancak ve ancak tüm seçeneklerin bir arada görülebilmesi ile mümkündür. Bunun yanı sıra bilgi elde etmede zamanla yarışılarak, karar verenlere hızlı bir şekilde iletilmesi sağlanmalıdır (Özata ve Aslan, 2004). Tam da bu noktada karar destek sistemleri karar vericilere, problem çözme sürecinde alternatif çözümleri test etme, bir arada görme, daha hızlı ve etkili kararlar alabilme

ve verileri yeniden gözden geçirme, olanağı sunar (Çil, 2002). Karar destek sitemleri, şirket işleyişini daha iyi anlamak, bu noktada daha etkili ve hızlı kararlar alınmasını sağlamak, maliyetleri en aza indirgemek, mevcut durumu analiz ederek oluşabilecek sorunlara karşı daha hızlı çözümler üretmek gibi pek çok konuda fayda sağlamaktadır. Bunun yanı sıra karar vermede yetkili kişilere, daha etkili ve daha hızlı kararlar verebilme noktasında sağladığı katkıları da göz ardı etmek mümkün değildir. Bu kapsamda şirketler, daha etkin kararlar alabilmek için karar destek sistemlerine ihtiyaç duymaktadır (Sönmez, t.y.).

Verilerden daha anlamlı sonuçlar çıkarmak tek başına yetersizdir. Sadece bugünü anlamak, bir şirketin uzun vadede kararlar alabilmesi için yeterli değildir. Yarının neler getirebileceğini öngörebilmek ve buna uygun hızlı çözümler bulabilmek özellikle küçük ve orta ölçekli şirketler için çok büyük önem arz etmektedir. Bu yüzden uzun vadeli kararlar alabilmek için mevcut karar destek sistemlerinden daha gelişmiş sistemlere ihtiyaç vardır. Günümüzde verilerden öngörü yapmak makine öğrenmesi sayesinde daha kolay yapılabilmektedir.

Matematik, istatistik, olasılık, veri madenciliği gibi pek çok alan ile yakından ilişkili olan makine öğrenmesinin üzerinde durduğu nokta, bilgisayarlara karmaşık örüntüleri algılama ve veriye dayalı mantıklı kararlar verebilme becerisi kazandırmaktır (Makine Öğrenimi, t.y.).

Bu çalışmanın amacı, küçük ve orta ölçekli firmaların yalnızca bugünü değil geleceği de öngörebilecekleri bir makine öğrenmesi destekli karar destek sistemi yapmaktır. Bu uygulama sayesinde, şirketler ellerinde bulunan verileri grafiklere dönüştürebilir. Makine öğrenmesi sayesinde tahminleme yapabilir. Ve bunları grafikler üzerinde görüp yorumlayabilir. Bu sayede yöneticiler daha etkin ve hızlı kararlar alabilir. Bunun yanında geçmiş veriler sayesinde ileride olası durumlar hakkında bilgi sahibi olabilirler.

Bu çerçevede yazılmış olan bu tez üç bölümden oluşmaktadır. Tezin ilk bölümünü küçük ve orta ölçekli firmalar için bilgi yönetimi, problemin tanımı, araştırmanın amacı, sorusu, kısıtları ve literatür taraması oluşturmaktadır. İkinci bölümde ise programın temelini oluşturan kds, makine öğrenimi, zaman serisi analizi gibi teorik bilgilere değinilmiştir. Son bölümde ise, tez kapsamında hazırlanmış olan bu uygulamanın tüm aşamaları detaylı bir şekilde açıklanmış, uygulama kapsamında

neler yapılabileceğine ve bir sonraki çalışmalarda neler yapılabileceğine değinilmiştir.

BİRİNCİ BÖLÜM
KÜÇÜK VE ORTA ÖLÇEKLİ FİRMALARDA BİLGİ YÖNETİMİ

Bilgi teknolojileri; bilginin toplanması, işlenmesi, depolanması ve iletilmesini sağlayan bilgisayar ve iletişim teknolojilerinin bütünü olarak tanımlanabilir. Bilgi teknolojileri işletmelere; ürün geliştirme, üretim, dağıtım, yönetim ve tüm bu faaliyetlerin muhasebeleştirilmesi konularında zaman ve fon tasarrufu sağlamaya yardımcı olduğu gibi aynı zamanda da işletme verilerinin değerlendirilmesi ve yöneticilere raporlanmasında önemli bir rol oynar. Bilgi teknolojilerinin yukarıda ifade edilen tüm fonksiyonlarına kıyasla daha yeni bir uygulama alanı da bilgisayar teknolojilerinin mali ve operasyonel denetimlerde kullanılmasıdır. Cankar (2006)'ın da belirttiği gibi bilgisayar teknolojilerinin kullanımı sayesinde örnekleme yapmaya ihtiyaç olmaksızın verilerin tamamı, klasik yöntemlere göre çok daha az zamanda analiz edilebilmektedir (Çalış, Keleş ve Engin, 2014).

BT'leri; KOBİ'lere işlem maliyetlerinde azalma, rakiplerle mücadele ve işbirliği olanağı, verimliliğinin artması, stok maliyetlerinin azaltılması, üretim ve hizmet kalitesini artırma, kaynakların kontrol ve planlaması, müşterilere daha kolay ulaşabilme vb. faydalar sağlamaktadır (Legris, Ingham ve Collerette, 2003). Bu amaçla kurum ve kuruluşlar bilişim teknolojileri (BT) sahiplik ve kullanımı konusunda çaba göstermektedir **(**Tüzün, Gülmez, Gökbudak ve Yılmaz, 2016).

KOBİ'lerin doğru ve güncel verilere ulaşabilmeleri, bilgi sistemleri altyapısını kurmalarına bağlıdır. KOBİ'lerin uygulayabilecekleri bazı bilgi sistemleri; Yönetim Bilgi Sistemleri, Ofis Otomasyon Bilgi Sistemleri, Veri Tabanı Yönetim Sistemleri ve İletişim Sistemleri'dir. KOBİ'lerin bu bilgi sistemlerini oluşturmak ve yararlanmak için öncelikle bilgi teknolojilerine yatırım yapmaları ve bilgi teknolojilerini doğru yönde kullanmaları gerekmektedir (Ayhan, 2011). Bu bağlamda; KOBİ'lerde bilgi teknolojilerinin kullanım düzeyi giderek artmakta, bu teknolojiler genellikle donanım, yazılım, prosedürler ve insan kaynaklarını içermektedir (Tüzün, Gülmez, Gökbudak ve Yılmaz, 2016).

Bilgi, günümüzde firmalara rekabetçi avantaj sağlayan en önemli kaynaklardandır. Bilgi işletmeler için tarih boyunca da hayati önem taşıyan bir unsur

olmuş ve belirsiz ortamlarda, zamanlı ve doğru bilgi, olayların gelişimini çarpıcı bir şekilde değiştirebilmiştir (Dibrell ve Miller, 2002 akt; Turan, 2009).

Bilişim Teknolojileri (BT) artık işletmelerin her faaliyetinde çok önemli roller oynamaktadır. Günümüz işletmeleri BT'leri kullanarak, sürekli olarak yeniden yapılanmakta ve teknolojinin değişiminin hızı ve yeni ortaya çıkan elektronik ticaret uygulamaları ve bunlara ek olarak gittikçe artan küreselleşme, işletmelerde bilgi ve teknoloji temelli bir yapılanmaya gitme sürecini tetikleyebilmektedir. Yeni teknolojiler işletmeleri sürekli gelişen ve yeni bilgiler unsurları sunan örgütler haline getirmiştir (Gordon ve Gordon, 2002 akt; Turan, 2009). Genelde küçük ve basit yapıya sahip KOBİ'ler, BT temelli bir yapılanma ile büyük, çok bölümlü ve küresel tabanda faaliyet gösteren idari yapılanmaya sahip işletmeler haline gelmektedirler (Gordon ve Gordon, 2002 akt; Turan, 2009).

Günümüzde otomasyon, bilgisayar teknolojileri ve işletmelerin diğer bilgi ve haberleşme teknolojilerini kullanması önemli ölçüde artmış ve gerekli hale gelmiştir. Teknolojiye uyum sağlamak günümüz modern toplumlarında, işletmelerin ve bireylerin en önemli sorunudur (Oktav vd., 1990 akt; Akdede ve Turan, 2008).

Ancak, günümüzde hızla artan otomasyon ve daha bilgili iş görenlere olan gereksinim, KOBİ'leri yeni teknolojilere uyum konusunda zorlamakta ve önemli bir problem olmaktadır. KOBİ'ler kaynak yetersizliğinden, yeni teknolojileri yakından izleyememekte ve kolaylıkla uyum sağlayamamaktadır (Koçak, 1996 akt; Akdede ve Turan, 2008).

Hem teknolojik araç ve gereç elde etmede, hem de yeni teknolojileri kullanacak yetişmiş teknik eleman bulmada ve bunları istihdam etmede güçlük çekmektedirler. Nitelikli elemanlar genellikle büyük işletmelerin iş güvencesi, prestiji ve göreceli yüksek maaşlarını tercih etmektedir. KOBİ'ler, teknolojiyi kullanmadaki avantajlarına rağmen, teknolojileri ve yenilikleri izlemekte ve araştırma ve geliştirme faaliyetlerinde bulunmada geri kalmaktadırlar (Oktav, 1990 akt; Akdede ve Turan, 2008).

KOBİ'lerin teknolojiyi etkin, verimli kullanamamaları ve kendilerini yeterince adapte edememeleri, bilgiye yeterince ulaşamamaları ve karar verme süreçlerinde etkin bir şekilde bilgiyi kullanamamaları sonucunu doğurmaktadır. "Bilgi Çağı" olarak adlandırılan bu dönemde, bilgi, rekabetin en temel unsuru

olmuştur. Bilgi günümüzde önemli bir üretim girdisi olmuş, yetersiz bilgi KOBİ'ler için pazar adaptasyonu sorunlarına yol açmıştır (Koçak, 1996 akt; Akdede ve Turan, 2008). Ancak, KOBİ'ler bilgiyi gereksiz bir masraf olarak görmektedirler. KOBİ yöneticileri bilgi sağlamaya ve kullanmaya gerekli önem ve önceliği vermemekte, sistemli ve etkin bir bilgi akışı sağlayacak sistemleri işletmelerinde kurmamakta ve bilgi eksikliği, ülkemizdeki KOBİ'lerin temel başarısızlık nedenlerinden biri olmaktadır (Oktav, 1990 akt; Akdede ve Turan, 2008). KOBİ'lerin dış pazarlara açılmaları, bu pazarlarda kalıcı olmaları bilgi konusunda ciddi çalışmalar yapmalarını gerektirmektedir (Özaytekin, 2002). Son yıllarda önem kazanan müşteri tatmini, toplam kalite bilinci, teknoloji alanındaki eğilimleri takip etmeleri ve değişen piyasa koşullarına uyum sağlamaları KOBİ'lerin bilgiye yatırım yapmaları ile mümkün olacaktır (Özaytekin, 2002; Akdede ve Turan, 2008).

1.1. PROBLEMIN TANIMI

Günümüzde, küçük ve orta ölçekli firmaların, piyasada bulunan diğer büyük şirketler gibi rekabet edebilecek, gereksinimlerini anlayabilecek, yorumlar yapabilecekleri büyük programları, bilgisayarları, sunucuları bulunmamaktadır. Birçok işletmede, raporlamalarının yapıldığı bölümlerin kod yazmadan yapılamaması, aynı zamanda verinin düzenlemesi için bir yapının olmayışı sağlıklı raporlamalar elde edilmesine engel olmaktadır. İşletmeler için piyasada bulunan birçok alternatif program bulunmaktadır. Fakat bu programların hiçbiri kod yazmadan yapılamayacağından dolayı şirketler için çok fazla zaman kaybı ve iş yükü barındırmaktadır. Bu yüzden bu sorunu ortadan kaldırmak amacıyla geliştirilen bu programda şirketler ellerinde bulunan verileri düzenledikten sonra, diğer programlarda olduğu gibi her bir sorun için ayrı ayrı satırlar dolusu iş yapmak ve nitelikli eleman çalıştırmak yerine kendileri bu bilgileri sistem üzerinden çekebileceklerdir.

Bununla birlikte mevcut sistemlerin kendi içerisinde, tahminleme, zaman serisi ve kümeleme analizi gibi üst düzey işlemleri bir arada yapabilecek bölümlerinin bulunmamasından dolayı, tüm bu özellikleri içerisinde barından tek bir uygulama hazırlanması düşünülmüştür. Şirketlerin ihtiyaçları doğrultusunda karar

vermesini kolaylaştırmak, yöneticilere yardımcı olabilmek ve yönetim süreçlerine katkı sağlaması amaçlanmıştır. Aynı zamanda tüm sektörlere uygulanabilir bir bütünleşik yapıda bulunan uygulamanın bulunmamasından dolayı bu uygulamanın yapılmasına gereksinim duyulmuştur. Buna göre bu programın yapılmasının temel nedenleri;

1-Yöneticilere istediği zaman istenilen raporların kolaylıkla yapılabilmesi,

2- Kullanıcı etkileşimi olan raporlar ve grafikler hazırlanabilmesi

3-Veri düzenleme işlemlerinin kolaylıkla yapılabilmesi,

4-Rapor yapımında sağladığı esneklikler,

5-Yalnızca bugünü değil geleceği planlayabilmesi,

6-En iyi sonucu veren sistemi analiz etmesi ve sadece tahmin değil zaman serisi ve kümeleme gibi analizlerin de kolaylıkla yapılabilmesi gibi sıralanabilir.

1.2. ARAŞTIRMA SORUSU

Bu çalışma için hazırlanan uygulama, belirlenen araştırma sorularına cevap verebilmek amacıyla geliştirilmiştir. Bu sorular:

1- Veri setleri ile yapılabilecek tüm veri ön işleme işlemlerini bu uygulama üzerinden sağlamak mümkün mü?

2- Geliştirilen uygulamaya verilen veri setlerinden oluşturulabilecek grafikler ve şekiller şirketler için fayda sağlayabilir mi?

3- Makine öğrenmesi algoritmalarını kullanarak temizlenen veriden doğru tahminleme yapabilmek mümkün mü?

Araştırmada belirlenen ilk soruda işletme içerisinde tutulan tüm verilerden daha anlamlı grafikler oluşturulabilmesi için, csv ya da excel formatına aktarılan tüm veri setinin temizleme işlemlerinin tümünü yapabileceği öngörülmektedir. İkinci soruda, bu veri setleri tarafından oluşturulan raporların daha etkili ve açık olması, yöneticinin doğru karar almasına yardımcı olması, eksik ve yanlış verilerden arındırılmış doğru grafiklerin oluşması, aynı zamanda oluşabilecek hataların önüne geçmesi öngörülmektedir. Son soruda ise makine öğrenmesi algoritmaları ele alındığında metinsel ifadelerden, boş kolonlardan, eksik satırlardan, makine öğrenmesi algoritmalarının doğruluk oranını doğrudan etkileyebilecek tüm

etkenlerden uzak bir şekilde temizlenmiş bu verilerden yüksek doğruluk oranında, geleceğe dair tahminlemeler yapabilen şirketlerin daha etkin ve hızlı kararlar almasına yardımcı olabileceği öngörülmektedir.

1.3. ARAŞTIRMANIN AMACI

Günümüzde veri, işletmeler için büyük önem arz etmektedir. Sadece bugünü değil geleceği tahmin edebilmek için gerekli olan en büyük kaynak verilerdir. Bu çalışma kapsamında geliştirilen uygulama, her sektöre uygun olarak tasarlanmış olmasından dolayı veri seti, "Kaagle" üzerinden bulunmuş ve denemeleri bu veri seti üzerinden yapılmıştır. Mevcut karar destek sistemlerini incelediğimizde, bütün karar destek sistemleri verinin temizlenmiş ve kullanıma uygun hale gelmiş olması üzerine kuruludur. Ama unutulmamalıdır ki insan faktörünün olduğu her yerde hata ve eksiklikler olacaktır. Bu yüzden gelen verinin temizlenmiş ve sorunsuz olarak sistem tarafından işlemeye, grafikler, raporlar yapılmasına hazır duruma gelmesi güç bir durum olduğu öngörülmüştür. Bu noktada tüm veri ön işleme işlemlerinin yapıldığı, bütünleşik bir sistem tasarımı hazırlanması düşünülmüştür. Örnekle açıklamak gerekirse; boş değerlerin temizlenmesi, cinsiyet (kadın, erkek) gibi verileri binary çevirme, sütun isimleri değiştirme, boş satıları temizleme, karakterleri istenilen değerlerle değiştirme vb. birçok özelliği içinde barındırması planlanmıştır. Tüm bu aşamalardan sonra temizlenmiş olan verilerden, yöneticiler için daha sağlıklı raporlar elde edilmesi için bir gösterge paneli hazırlanması amaçlanmıştır. Bu panel sayesinde yöneticiler için istedikleri zaman istedikleri değişiklikleri yapabilecekleri, çeşitli filtreleri koyabilecekleri ve istedikleri grafikleri hazırlayabilecekleri bütünleşik bir sistem tasarımı öngörülmüştür. Mevcut sistemlerin birçoğunda yalnızca basit grafikler hazırlanabilen karar destek sistemleri bulunmaktadır. Bu yüzden yalnızca bugünü değil geleceği de öngörebilen, tahminler yapabilen sistemlerin hazırlanması ve uygulamaya konulması gerektiği düşünülmüştür. Bu çalışma kapsamında 15 makine öğrenmesi algoritması kullanılmış ve bu algoritmalardan hangisi daha doğru bir tahminleme yapabilecekse bu kullanıcıya önerilmiştir. Bu kapsamda şirketlerin verilerini daha doğru tahminleyebilir duruma gelmesi öngörülmüştür. Zaman serisi analizlerini yapabilecekleri grafikleri ve kümeleme analizi yapabilecekleri kısımlar

da bu sisteme dâhil edilmiştir. Bu sayede yalnızca tahminleme değil, zaman çizgisi üzerinde ne olabileceğine dair çıkarımlar ya da ürünlere göre gruplama yapabilecekleri ve bunların raporlarını alabilecekleri sistem tasarımı yapılması öngörülmüştür. Tüm bunları bir arada yapan makine öğrenmesi tabanlı bir karar destek sistemi yapılması tasarlanmış ve bu sayede karar vericilere daha hızlı ve daha etkin kararlar almasında katkı sağlayacağı öngörülmüştür.

1.4. ARAŞTIRMANIN KISITLARI

Bu çalışmada kullanılan verileri bulmakta oldukça güçlük çekilmiştir. Verilerin her alanda kullanabilir olduğunu bir örnek üzerinde açıklamak için internet kaynaklarına başvurulmuştur. İnternette veri paylaşımı yapan siteler üzerinde yapılan araştırmalar sonucunda, bir bilgisayar firmasının satış verileri elde edilmiştir. İnternet üzerinde çok fazla kaynak bulunmasına rağmen elde edilen kaynakların çoğu küçük veri setlerinden oluşmaktadır. Küçük veri setleri makine öğrenmesi algoritmaları kullanılırken aşırı öğrenmeye sebep olduğundan dolayı bu veriler kullanılamamıştır. Ayrıca bu veriler makine öğrenmesi algoritmalarını kullanılırken veri seti küçük olmasından dolayı çok doğru sonuçlar elde edilememiştir. Bu yüzden, bu çalışmada iki yüz bin satırın üzerinde bir veri seti kullanılmıştır. Bunun yanı sıra bu sitelerden bulunan verilerin tahminlemeye uygun veri barındırmaması veya zaman serisine uygun tarihsel veri bulundurmaması gibi birçok sebepten dolayı tüm bu satırları içerisinde barındıran veri seti bulmak uzun zaman almıştır. Bulunan bu veri setinin temizlenebilmesi içinde bir arayüz geliştirilmiş ve eldeki veri tüm boş değerlerden, metinsel ifadelerden, gereksiz ve yanlış verilerden arındırılmıştır.

1.5. LİTERATÜR TARAMASI

Tıbbi karar destek sistemlerinin, hastaların tedavisinde sağlık personelinin en uygun yöntemi seçmesine yardımcı olabileceği üzerine çalışma yapmıştır. ALARM (“hasta izleme için bir alarm mesajı sistemi sağlamak üzere tasarlanmış sistem”) ağındaki olasılıkları doğru yansıtabilmek için farklı veri setine sahip, 10, 100, 1000 ve 2000’lik sentetik veriler üretilmiştir. Oluşturulan bu kayıtlar 11 algoritma

üzerinde tıbbi karar destek sistemi mekanizması için farklı veri setleri için karşılaştırılmıştır. Bu 11 makine öğrenme yönteminin tıbbi karar destek sistemi çıkarım mekanizması için doğruluğu çeşitli veri setleri üzerinde karşılaştırılmış ve Karar ağacı 44 veri seti için iki testte en iyi algoritma olarak tespit edilmiştir. Veri seti büyüdükçe, diğerlerine göre daha iyi tahmin edebilmektedir (Bal, Amasyalı, Sever, Köse ve Demirhan, 2014).

(Al-Asadi, 2018) tarafından yapılan çalışmada, futbolcuların her formasyonda oynayabilir olmaması ve her formasyon için ayrı bir futbolcuya ihtiyaç duyulması, futbolcuların pozisyonları ise ancak ve ancak takım koçları tarafından gözlemle belirlenebiliyor olması sebebiyle, bu tür zorlukları ortadan kaldırmak adına karar destek sistemi oluşturulmuş ve futbol takım yönetimi için makine öğrenmesi yöntemlerinden faydalanılmıştır. Bu amaçla yapılan karar destek sisteminde, her futbolcunun yeteneklerini temel alınmış ve bu doğrultuda en uygun pozisyon belirlenerek takım oluşturulmuştur. Bu çalışmada, 17359 oyuncu içeren FIFA oyunu verileri kullanılmış ve bu veriler analiz ederken, sınıflandırma ve regresyon problemleri için makine öğrenmesi tekniklerinden yararlanmıştır. Ayrıca, veri indirgemek için principal component analysis ve recursive feature elimination algoritmalarından faydalanılmıştır. Faydalanılan bu algoritmalar ile 29 özellik içerisinden 17 tanesini kullanmış ve oyuncular için uygun pozisyonunun belirlenebileceği görülmüştür. Bu algoritmalar ile ikili ve çoklu sınıflandırma yapılmış. % 88.60 ikili sınıflandırma, % 58.53 çoklu sınıflandırma doğruluk oranıyla diğer algoritmalardan daha doğru sonuçlar vermiştir. Algoritmaların performanslarının değerlendirilmesi üç teknik ile yapılmıştır. (Hold-out, Cross Validation and Repeated Random Hold-out). Her oyuncunun uygun pozisyonunu belirledikten sonra istenilen formasyona göre en iyi takım her oyuncunun derecesi dikkate alarak oluşturulmaktadır. Top sürme becerileri için 4 farklı algoritma kullanmış (linear regression, logistic regression, random forest and neural network), bu algoritmalardan en iyi sonucu rastgele orman algoritması % 99.90 doğruluk oranıyla vermiştir.

(Bilgilioğlu, 2018) çalışmasında, MKDS sistemlerinin daha çok uzman görüşü gerektirmesi, bu doğrultuda objektif olamaması ve verinin yapısında ortaya çıkan bir değişimin mekânsal analizleri değiştirebileceğinden dolayı dinamik bir

model oluşturamaması sebebiyle, meydana gelen hataları en aza indirgeyecek envanter bilgilerine dayalı makine öğrenmesi tekniğinin entegre edilmesi ve bu sistemlerin MKDS sistemlerini otomatikleştirmesini amaçlamıştır. Makine öğrenmesinde kullanılan denetimli öğrenme algoritmaları olan Yapay Sinir Ağları, Destek Vektör Makineleri, CHAID ve Karar Ağacı algoritmalarından ID3, C4.5, CART ve Rastgele Orman algoritmaları Aksaray ili taşınmaz değer haritasının üretiminde kullanılmış ve performansları değerlendirilmiştir. Bu çalışma sonucunda Yapay Sinir Ağları, CHAID ve Destek Vektör Makineleri yöntemleri ile oluşturulan modeller sıklıkla çoklu regresyon analizine göre daha gerçekçi sonuçlar vermiştir. Bu tez kapsamında kullanılan tüm yöntemleri kapsayan ve kullanıcıların sonuçları karşılaştırabileceği bir yazılım geliştirilmiştir.

(Monhamady, 2018) çalışmasında**,** sisteme gönderilen bir web sitesinin metninden şirket ismini tahmin edebilen, iki aşamalı bir makine öğrenmesi yöntemi geliştirmiştir. Bu çalışmanın ilk kısmında, kelimeden şirket ismi olup olmadığını tahmin eden bir sınıflandırma gerçekleştirilmiştir. İkinci kısımda ise, geliştirilen kural tabanlı yöntem sayesinde başlıktaki kelimelerini de tarayarak simge benzerlik ölçütlerini kullanarak şirket ismi olmaya aday olan kelimeleri sıralamakta ve en yüksek skorlu kelimeleri şirket ismi olarak tahmin etmektedir. Bulgular sonucunda, ilk aşamada zor olan metinlerdeki isimleri tanımsız olarak ayırmış ve yüksek hassasiyet oranı elde edilmiştir. İkinci aşamada ise oluşturulan kural tabanlı yapıda yüksek doğruluk oranı elde edilmesine rağmen ilk aşamadaki hassasiyet oranından daha düşük bir sonuç elde edilmiştir. İki sınıflandırıcı birleştirildiğinde hem genel doğruluk oranı, hem de hassaslık oranı yüksek olarak elde edilmiştir.

(Dem, 2019) çalışmasında**,** şirketlerin personel istihdam etme aşamasında kişileri seçmekte zorluk çeken insan kaynakları personelinin, görüşmeye çağıracağı kişileri ya da kimi işe alacağını belirlemesi kısmında çok fazla gelen başvuruların değerlendirilmesi, ihtiyaç duyulan adayların başvuru yapan adaylar arasında kaybolmasına ve dikkate alınmaması adına veri madenciliği yöntemlerini kullanarak bir karar destek sistemi oluşturmayı amaçlamıştır. Bu çalışma için 613 kişiden oluşan anket çalışması yapılmış ve karar kriterlerinin belirlenmesi adına bu anketlerden gelen veriler kullanılmıştır. Bu veriler Rapidminer CART, C 4.5, ID3, Naive Bayes ve Random Forrest algoritmaları kullanılarak analiz edilmiştir. Araştırma sonuçlarını

bulmak için Holdout, Çapraz geçerleme ve Bootstrap yöntemleriyle performans testi yapılmıştır. Çalışma sonucunda en iyi sonucu C4.5 algoritması ile %79.29 vermiştir. Bu oran, personel seçmek için veri madenciliğinin yüksek bir doğruluğa ulaşılabildiğini göstermektedir.

(Akkol, 2019) çalışmasında, İzmir İtfaiye Daire Başkanlığı'nda kurumun ihtiyaçları doğrultusunda, üst mercilerle fikir alışverişi yaparak karar vermeyi destekleyen bir sistem hazırlamıştır. Bu doğrultuda yalnızca bugünü değil geleceğe yönelik çıkarımlar yapılmasında da çok önemli ve yararlı bilgiler vermesi amaçlanan bir uygulamadır. Uygulama geliştirilirken kâğıt ve arşiv ortamında tutulan veriler dijitalleştirilmiş, veri tabanına aktarılmıştır. Bu sayede çeşitli sorgular ve analizler yaparak itfaiyenin mevcut durumu ve performansı hakkında geleceğe yönelik tahminler, ileriye dönük bilgiler sunan grafikler, haritalar karar vericilere raporlanmıştır.

İşletmeler için, artan ve akan veri miktarının yönetilebilmesi, yeni teknolojileri iş süreçlerine adapte edilmesi çok önemlidir. Özellikle büyük miktardaki verileri anlamlandırmak ve analiz etmek, makine öğrenmesi tekniklerini operasyonlara dâhil etmek şirketler için büyük önem arz etmektedir. Bu kapsamda (Taşçı, 2019) çalışmasında, emlak sektörünün dinamiklerinden yararlanarak büyük veri sağlayan yapısı sayesinde bu sektörü etkili analiz yapabilmek için yeni yaklaşımlar önermiştir. Yapılan çalışma ile, bankalar, kredi finansman kuruluşları, inşaat şirketleri, aracılar, değerleme uzmanları ve danışmanlara; emlak fiyatları ve endeksleri, değerleme modülü, makro-ekonomik veriler ve trend analizleri gibi bilgiler sağlamak, piyasayı yakından takip etmek, sektör fırsatlarını tespit ederek, operasyonel, taktiksel ve stratejik kararlar alınmasına yönelik bir iş zekası uygulaması geliştirilmesi amaçlanmıştır.

(Akgün, 2019) çalışmasında, son yıllarda çok fazla kullanılan yapay zeka uygulamalarından makine öğrenmesi algoritmalarının kariyer planlamasında kullanılmak üzere üniversiteden yeni mezun olmuş kişilere çalışma sektörü önerisinde bulunacak bir tavsiye sistemi tasarlamıştır. Bu sistem, mezunların eğitim ve iş verilerini kullanarak, özlük bilgilerini dahil etmeden tasarlanmıştır. Bu çalışmada en çok kullanılan yöntemlerden biri olan Cross Industry Standard Process for Data Mining (Çapraz Endüstri Veri Madenciliği Standart Süreci – CRISP-DM)

yöntemi çalışmanın özelliklerine uygun veri madenciliği süreci olarak seçilmiştir. Modelleme kısmında ise makine öğrenmesinde kullanılan (En yakın komşu, Rassal orman, Naive bayes, Karar vektör sınıflandırıcıları, Karar ağacı) algoritmalarından faydalanılmıştır. Yapılan çalışmadan elde edilen bulgular, karmaşıklık matrisinin yardımıyla incelenerek karşılaştırılmıştır. Karşılaştırma sonucunda, gözetimli makine öğrenmesi algoritmalarının doğruluk yüzdelerinde en iyi oranı %67,46 olarak Rassal Orman vermiştir.

(Rizvanche, 2020) çalışmasında, doğrusal olan ya da olmayan örüntüler nedeniyle zaman serileri ile tahminleme yapmanın karmaşıklığından bahsetmiştir. Tekil modellerden daha iyi sonuçlar elde etmek için, hibrit modeller ile tahminlemenin doğruluk açısından daha iyi sonuçlar elde edileceği düşünülmektedir. Bu çalışmada gerçek hayat problemlerini doğrusal ve doğrusal olmayan yaklaşımlar ile birlikte bünyesinde hibrit modelleri de barındıran bir web tabanlı karar destek sistemi yapılmıştır. Uygulamanın araştırma aşamasında talep tahmini yaklaşımlarının sisteme entegre etme sürecine üzerine düşünülmüş ve doğrusal yöntemlerin diğerlerine göre daha iyi sonuçlar elde edilebileceği gözlemlenmiştir. Çalışma sonucunda elde edilen bulgular neticesinde zaman serilerinin birçok faktör barındırması nedeniyle yapılan analizler sonucunda hibrit bir modelin zaman serilerini açıklamak için daha avantajlı olduğu gözlenmiştir.

Zaman serileri tahminleme yaparken kullanılan en temel araçlardan biridir. Literatürde zaman serileri, zaman analizi, makine öğrenmesi, istatistik gibi birçok şekilde kullanılmıştır. (Rizvanche, 2020) yapmış olduğu çalışmada hibrit modellerin daha avantajlı sonuç çıkardığı bulgusuna ulaşmış fakat durumun aksine, (Özdemir**,** 2020) çalışmasında, M4 yarışmasında kullanılmış olan veri setinden alınan ve farklı frekanslara sahip serilerle orantılı olarak seçilen verilerden zaman serilerinin tahmin karşılaştırılmasını incelemiştir. Tahmin performansı karşılaştırılan modellerden en iyi performans veren modellerin makine öğrenmesi modelleri olduğu sonucuna ulaşılmıştır. Bunun yanında verilerin ve zaman frekanslarının uzunluğuna göre tahmin performansı artmaktadır. Bunun yanında kullanılan hibrit modellerin her veri için uygun bir metot olmadığı kanıtlanmıştır.

İKİNCİ BÖLÜM
TEORİK ÇERÇEVE

2.1. KARAR DESTEK SİSTEMİ

Literatürde karar destek sistemleri pek çok kişi tarafından, farklı şekillerde tanımlanmıştır. Bu tanımlar aşağıda verilmiştir:

Kroeber ve Watson, (1987 akt; Çetinyokuş ve Gökçen, 2002) karar destek sistemlerini, yarı-yapısal ve yapısal olmayan kararlar verilirken, karar vericilerin karar vermesini kolaylaştırmak amacıyla yapılan esnek ve etkileşimli bir sistem olarak tanımlarken, Klein ve Grubbstrom, (1998 akt; Çil, 2002) karar vericilerin verdikleri kararların kalitesini ve bu sistemlerin uygulama alanlarını arttırmak amacıyla hazırlanan bilgiyi bir bütün olarak sunan bilişim teknolojisi sistem olarak tanımlanmıştır. Haag, Cummings ve Dawkins (2012) ise karar verme aşamasında verilen kararların yapısal olmadığı durumlarda, karar vericilerin işlerini kolaylaştırmak amacıyla tasarlanmış etkileşimli sistem olarak ifade etmiştir.

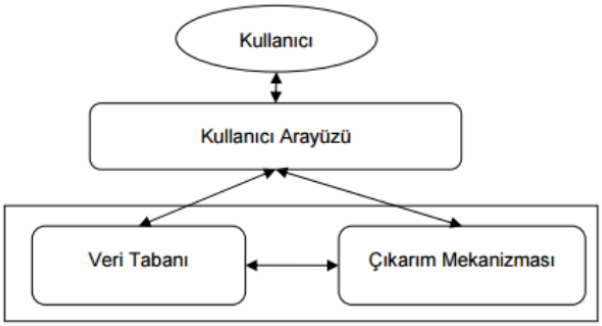
Karar destek sistemlerinde kararlar yapısal, yarı yapısal ve yapısal olmayan olmak üzere üç grupta incelenir:

Yapısal kararlar, kolay bir şekilde yürütülen, sürekli olarak yapılan karar verme durumunun söz konusu olduğu durumları ifade etmektedir. Yarı yapısal kararlar, bir sorunla karşı karşıya kalındığında karar verme yetkisine sahip olan karar vericinin yalnızca sorunun bilinmeyen yönleri hakkında çözüm üretmesi durumudur. Yapısal olmayan kararlar ise kesin bir yargı söz konusu olmamakla birlikte kar verme işlemi gerçekleşmesi için kesin bir yöntem bulunmamaktadır, karar verici geçmiş bilgi ve deneyimlerini kullanır (Turban,1995).

2.1.1. Karar Destek Sistemi Bileşenleri

Karar destek sistemleri kullanıcı, kullanıcı ara yüzü, veri tabanı ve çıkarım mekanizması olmak üzere dört ögeden oluşmaktadır (Akgül, Üstündağ ve Tanrıverdi, 2018).

Şekil 1:KDS bileşenleri (The Components of Decision Support Systems**)**



Kaynak: Akgül, Üstündağ ve Tanrıverdi, 2018.

Karar destek sisteminin bileşenlerinden biri olan kullanıcı ara yüzü, gerekli olan bilgilerin sisteme girildiği veri tabanı, çıkarım mekanizması ve kullanıcı bileşenleri arasındaki etkileşimde rol oynayan ara birimdir. Diğer bir bileşen olan çıkarım mekanizması ise içerisinde tahmin, benzetim, en iyileme ve analiz gibi modellerin bulunduran, karar verme konumunda olan kişileri doğru kararlara ulaştıracak olan bölümdür. Veri tabanı bileşeni, karar destek sistemlerinde kullanılacak olan verilerin ilişkisel olarak tutulduğu ve bu verilerden tecrübe ve deneyimlerine göre anlamlı bilgi çıkarımının yapıldığı alandır (Çebi, 2010).

2.1.2. Karar Destek Sistemi Özellikleri

Karar verme yalnızca bilgisayar alanında karşı karşıya kaldığımız bir durum değildir. Bilgisayar alanı dışında da yaşamımızın hemen her anında karar verme durumunda kalırız. Karşılaşılan durumlar ile ilgili çıkarımlar yapar ve buna göre hareket ederiz. Bu durum göz önünde bulundurulduğunda, alınan kararların şirketler için önemi yadsınmaz bir gerçektir. Alınan kararların doğru olması ve bunun hızlı bir biçimde gerçekleşmesi şirketlerin performansını direkt olarak etkilemektedir. Bu bakımdan düşünüldüğünde karar destek sistemleri çok büyük önem arz etmektedir. Karar destek sistemlerinin özellikleri aşağıdaki gibidir (Gülçe,2010).

- Karar destek sistemleri yarı-yapısal ve yapısal olmayan kararlar alırken, karar vericilere karar verme noktasında yardımcı olmaktadır.
- Karar verme sürecindeki bütün adımları destekler.

- Sadece bireysel değil grup tabanlı olarak da karar verme desteği sağlar.
- Karar destek uygulamalarının kullanımı kolay, esnek ve günceldir.
- Karar destek sistemleri, tüm yönetim aşamaları arasında entegrasyon sağlayabilmektedir.
- Karar vericilerin ihtiyaçları göz önünde bulundurularak tasarlanır.
- Karar destek sistemleri karar vericin karar almasın yardımcı olur, onun yerine kararlar almaz.
- Karar destek sistemleri, karar alma aşamasında alınan kararların kalitesini yükseltmeye odaklanır.
- Karar destek sistemleri, stratejik, operasyonel ve taktiksel kararları alma noktasında karar vericiye fayda sağlar.

  (Aronson, Liang ve Turban, 2005; Gökçen, 2007)

2.1.3. Karar Destek Sistemi Türleri

Karar Destek sistemlerinin beş türü vardır. Bunlar, model, veri, bilgi, grup ve web tabanlı sistemlerdir. Optimizasyon tekniklerini ve istatistik bilimini içerisinde bulunduran karar destek sistemleri Model tabanlı, eşzamanlı olarak çalışan ve büyük verileri içerisinde barındıran karar destek sistemleri veri tabanlı, kullanıcıya hareket bildiren kural tabanlı karar destek sistemleri bilgi tabanlı, ekip çalışmasını arttırmak amacıyla birden fazla kullanıcının etkileşim sağlayabildiği karar destek sistemleri grup tabanlı, internetin gelişimiyle birlikte artan ve grup tabanlı sistemlerin gelişmiş versiyonu olan kullanıcıların bir ağ aracılığıyla çalışmasını sağlayan sistemler ise web tabanlı karar destek sistemi olarak adlandırılır (Turunç, 2006; Çebi, 2010; Akgül, Üstündağ ve Tanrıverdi, 2018).

2.2. MAKİNE ÖĞRENMESİ

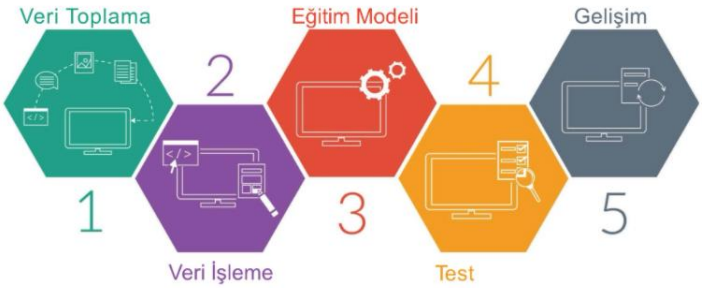
Günümüz bilişim teknolojilerinde meydana gelen ilerlemeler sayesinde, büyük miktardaki veriyi saklama ve her yerden ulaşabilme erişimine sahibiz. Fakat veri miktarı büyüdükçe veriyi anlamak ve bilgiye ulaşmak güçleşmektedir. Verileri tek tek işlemek, bunlar üzerinde analizleri yapmak çok güçtür, bu yüzden geçmiş verilerden yola çıkarak gelecek için tahminlemeler yapmak adına makine öğrenmesi geliştirilmiştir. Makine öğrenmesi yöntemlerinin temel çıkış noktası, nasıl ki insanlar ellerindeki verilere bakarak tecrübe ve bilgi sahibi olabiliyor ise, makinelerde bu işlenmemiş verilerden bilgi ve tecrübe edebilirler (Michalski ve Kodratoff, 1990).

Literatürde birçok tanım yapılmasına rağmen makine öğreniminin açık bir tanımı yoktur (Alpaydın,2010). Makine öğrenimi, örnek veriler ve geçmiş deneyimleri kullanarak tahmine dayalı algoritmaları inceleyen, çeşitli algoritma ve tekniklerin geliştirilmesi için çalışılan bilimsel bir çalışma alanıdır. Makine öğrenimi matematiksel model oluşturmada istatistiksel yöntemleri kullanır. Temel işlevi örneklerden çıkarım elde etmektir. (Alpaydın,2010; Mitchell, 1997) yapmış olduğu tanımda makine öğrenmesi, bilgisayarlardan ve veri tabanlarından elde edilen veri türlerini öğrenme odaklı olmasını sağlayan algoritmaların geliştirme ve tasarım aşamasındaki süreci baz alan bir bilim dalıdır (Mitchell, 1997).

Makine öğrenmesinin; tıp, finans, mekânsal analizler, eğitim planlaması, borsa, dolandırıcılık tespiti vb. gibi çok geniş bir uygulama alanı vardır (Ünsal, 2011). Saymış olduğumuz tüm bu alanlarda geçmiş verileri kullanarak çıkarımlar yapmaya çalışır.

Lantz (2013) yapmış olduğu çalışmasında makine öğrenmesinde kullanılan algoritmaların uygulanması aşamaların beş adımda açıklamıştır. Bu adımlar Şekil 2'de gösterildiği gibidir. Tüm bu adımlar gerçekleşirse model tam anlamıyla kullabilir. Bu aşamada uygulanan adımlar aşağıdaki gibidir:

Şekil 2: Makine Öğrenmesi Modeli Uygulama Aşamaları



Kaynak: Makine Öğrenmesi ile İK Süreçleri Optimizasyonu, t.y.,par. 2.

- Verinin Toplanması

Makine öğrenme modelinin daha doğru sonuçlar verebilmesi için verilerin çeşitliliğin ve büyüklüğün arttırılması gerekmektedir (Ryan, 2018). Veriler problemlere uygun bir şekilde toplanır. Verilerin toplanması çeşitli şirketler, internet gibi birçok yerden probleme uygun verinin toplanması mümkündür. Verinin yeterli olmadığı durumlarda ise araştırmacı kendi verisini toplayabilir. Veriler mülakat, deney, gözlem gibi birçok yolla elde edilebilir. Bu veriler kağıt üzerinde ya da veri tabanlarında da olabilir. Bu yüzden bu verilerin analize uygun bir şekilde elektronik ortama taşınması gerekmektedir (Şahinarslan, 2019).

- Verilerin İşlenmesi

Bu aşamada toplanan verilerde bazen aykırılıklar görülebilir, bu aykırı değerler veri setinden çıkarılmaz ise modelleri kullanırken sapmalara neden olabilir. Aynı zamanda veriler eksik ve kayıp değerler içerebilmektedir, bunların giderilmemesi durumunda algoritmaları kullanırken hatalar meydana gelebilir. Yine makine öğrenmesinin hata yapma olasılığını düşürmek amacıyla bazı veriler tek tipe dönüştürülmesi gerekmektedir (Şahinarslan, 2019).

Çalışma sırasında en çok zaman ve emek harcanan bu aşamada, verilerdeki aykırılıkların ve eksikliklerin giderilmesi, yanlışlıkların düzeltilmesi için veriler analiz edilir ve temizlenir. Bu sayede verilerin daha sağlıklı sonuçlar elde edileceğinden dolayı çalışmanın başarısını doğrudan etkileyecektir (Atan,2020).

- Verilerin Eğitilmesi

Önişleme sürecinde temizlenmiş olan verinin algoritmalar yardımıyla veriler arasındaki ilişkilerin öğrenilmesi sürecidir. Bu süreçte veriler arasında korelasyonların belirlenmesi ve tahminlemelerin yapılması için algoritmalar ile modeller oluşturulur.

- Verilerin Test Edilmesi

Verilerin doğruluğu test edilmesi için eğitim verileri test verileri ile karşılaştırılır. Bu karşılaştırma işlemi algoritmaların eğitim verilerinin test verilerini ne kadar açıkladığını ifade eder. Bu sayede algoritmaların ne kadar sağlıklı çalıştığı anlaşılmış olur.

- Geliştirme

Test edilen verilerin daha iyi çalışması ve daha doğru sonuçlar verebilmesi için çeşitli algoritmalar kullanılması, veri çeşitliliğinin arttırılması, veri miktarının arttırılması işlemlerinin yapıldığı aşamadır.

2.2.1. Makine Öğrenmesi Yöntemleri

Makine öğrenmesinde birçok öğrenme türü vardır. Öğrenme türleri öğrenmenin nasıl gerçekleştiğini ve bunu yaparken hangi algoritmaların kullanıldığını göstermektedir (Öztemel, 2012). Makine öğrenmesinde bulunan öğrenme türleri aşağıdaki gibidir:

• Denetimli Makine Öğrenmesi (SML: Supervised Machine Learning)
• Denetimsiz Makine Öğrenme (UML: Unsupervised Machine Learning)
• Yarı-Denetimli öğrenme (Semi-supervised Learning)
• Pekiştirerek öğrenme (Reinforcement Learning)

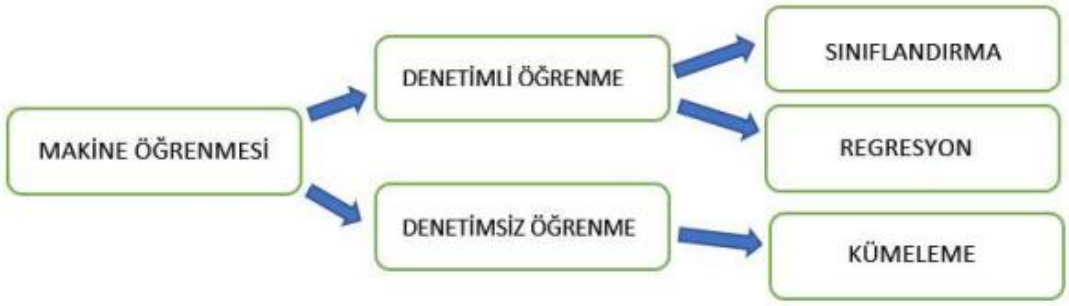
Bu öğrenme türlerinden en çok kullanılan denetimli ve denetimsiz öğrenmedir (Karslı, 2019).

2.2.1.1. Denetimli Makine Öğrenmesi (Gözetimli Öğrenme)

Gözetimli öğrenme de kullanılan algoritmalar bilinen bir veri kümesinden öğrenme gerçekleştirir. Ardından sisteme gönderilen yeni verilerden tahminler yapılması sağlanır. Eğitim için verilen verilerin veriler ne kadar doğru olursa yapılan

tahminler o oranda doğru çıkmaktadır. Tahmin için kullanılan veriler etiketlenmişse genellikle bu gözetimli öğrenme yöntemleri kullanıldığın göstergesidir (Okur, 2020). Gözetimli öğrenmede amaç hangi sınıfa ait olduğu belli olan etiketli verilerin hangi çıktı değerine karşılık geldiğinin belirlenmesidir. Oluşturulan model ile elde edilen sonuç arasındaki fark hata oranını ifade etmektedir. Bu oran her zaman en aza indirgenmeye çalışılır (Kavuncu,2018). Gözetimli öğrenme Şekil 3'de gösterildiği üzere sınıflandırma ve kümeleme olarak 2 başlık altında incelenmektedir.

Şekil 3:Gözetimli Öğrenme Adımları



Kaynak: Evaluatıng machıne learnıng algorıthm's performance ın predıctıng presence of heart disease,2018.

- Sınıflandırma

En çok kullanılan yöntemlerden biri olan sınıflandırma yönteminde, etiketlenmiş olan verilerin değerlerinden faydalanarak, sınıfı belli olmayan verilerin öznitelikleri göz önünde bulundurularak belirli bir sınıfa veya gruba atanması işlemidir (Can, 2020).

Sınıflandırma modeli, veri setinin temin edildiği tahminleme olarak kullanılacak modeli oluşturmak ve oluşturulan bu modelin sınıflandırılmamış veriler üzerinde sınıflandırma algoritmalarını kullanarak test verilerinin tahminlenmesini sağlamak olmak üzere iki aşamadan oluşmaktadır. Aşamalar sonucunda elde edilen tahmin sonuçları incelenir ve bu inceleme sonucunda bilinen sınıf etiketlerine atanır. Atama işleminin ardından modelin doğruluk derecesi belirlenir ve belirlenen bu değer ne kadar yüksekse model o kadar başarılı kabul edilir (Aydın,2020).

Sınıflandırma tekniklerinin nihai amacı verilerin ortak özellikleri göz önünde bulundurularak gruplama işleminin yapılmasıdır. Örneğin; kanser, tümör gibi kötü

hastalıkların tespiti, müşterilerin kredilendirilmesi işlemleri, malzeme dayanıklılık tespiti, hava durumu tespiti, maillerin spam tespiti için sağlık, finans, inşaat vb gibi hemen hemen her alanda sınıflandırma tekniklerinden faydalanılmaktadır (Ünsal,2011).

Makine öğrenmesinin sınıflandırma algoritmaları her veri kümesi için optimum sonucu vermemesi nedeniyle Naive Bayes, K- En Yakın Komşu, Karar Ağaçları, Yapay Sinir Ağları, Rastgele Ormanlar ve Destek Vektör Makinesi Algoritması gibi pek çok algoritma geliştirilmiştir (Göker, 2019).

- Regresyon

Sıklıkla kullanılan bir modelleme tekniği olan regresyon, bağımsız değişkenlerin (tahmin edici), bağımlı değişken (tahmin edilecek) ile arasındaki ilişkiyi ölçmek için kullanılan bir analiz metodudur (Can, 2020). Bu metotta girdiler "bağımsız değişken", elde edilen sonuç ise "bağımlı değişken" olarak da adlandırılmaktadır. Sonuçlar güven aralığı içinde değerlendirilir (Argüden ve Erşahin, 2008).

Regresyon algoritmaları, gözetimli öğrenme kategorisinde kullanılmasının yanı sıra gözetimsiz öğrenme algoritmalarında normal olmayan durumları tespit etmek için de kullanılabilmektedir (Okur,2020).

Örneğin; müşteri temsilcilerinin yapmış olduğu satışlar üzerinden bir sonraki aylık satış verilerinin tahmini, gelir değişimine göre satılacak ev sayısının tahmini, alınan maaşlar üzerinden kişinin meslek grubuna göre maaş tahmini, atmosferdeki CO_2 oranının tahmin edilmesi, ekipman arıza tespiti gibi birçok alanda kullanılmaktadır. Tahminleme yapılırken doğru sonuçlar elde etmek için birden fazla girdiden yararlanmak büyük önem arz etmektedir (Göker,2019).

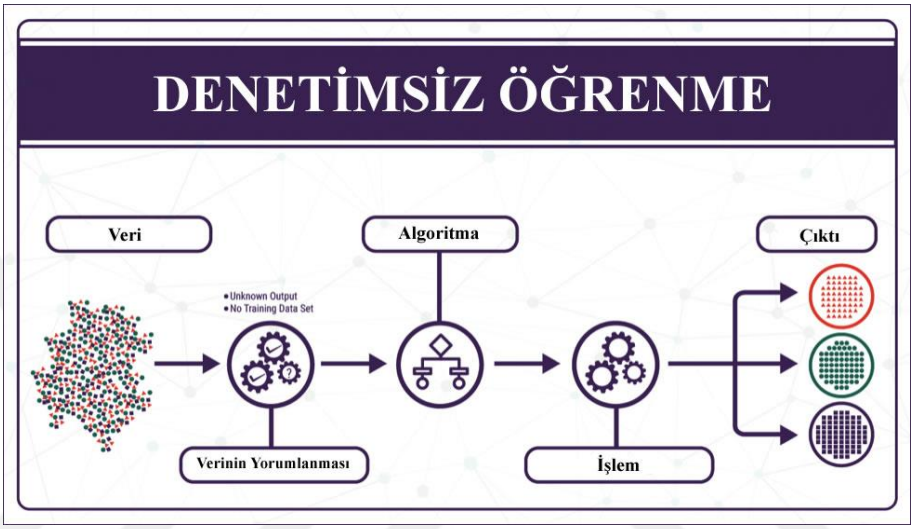
Doğrusal Regresyon, Lojistik Regresyon, Çoklu Doğrusal Regresyon, Polinomal Regresyon, Destek Vektör Makine Regresyonu ve Karar Ağacı Regresyonu en çok kullanılan regresyon algoritmalarıdır (Okur,2020).

2.2.1.2. Denetimsiz Makine Öğrenme (Gözetimsiz Öğrenme)

Yoğunluk tahmini olarak da adlandırılan denetimsiz makine öğrenmesi, sınıfı olmayan verilerle, bu verilere benzer özelliklere sahip verilerin gruplandırılması

amacı ile kullanılan, girdi verileri için uygun bir gösterim biçimi belirleyen bir öğrenme türüdür. Bu öğrenme modelinin, sadece girdi değerlerinden oluşması sebebiyle, girdilere karşılık gelen bir çıktı (sınıf veya etiket) değeri yoktur. Bu yüzden girdilerin hangi sınıfa ait olduğu bilinmemektedir (Fatima ve Pasha, 2017 ; öztemel, 2012) .

Şekil 4:Gözetimsiz Öğrenme Aşamaları



Kaynak: Why unsupervised machine learning is the future of cybersecurit, t,y., par.3.

Bu öğrenme modelinde yer alan veri kümesindeki örneklerin çıktıları bilinmediğinden dolayı amaç sınıflandırma veya tanıma değildir. Gözetimsiz öğrenme modellerine örnek vermek gerekirse; video öneri sistemleri, arkadaş, grup, sayfa öneri sistemleri, ürün öneri sistemleri gibi birçok alanda kullanımı mevcuttur.

Gözetimsiz öğrenme modellerinde etiketlenmiş veri bulunmaz bu yüzden değerlendirme yapmak çok zordur. Gözetimsiz öğrenme algoritmaları, Birliktelik Kuralı, Kümeleme Analizi, Boyut Azaltma olarak 3 gruba ayrılır.

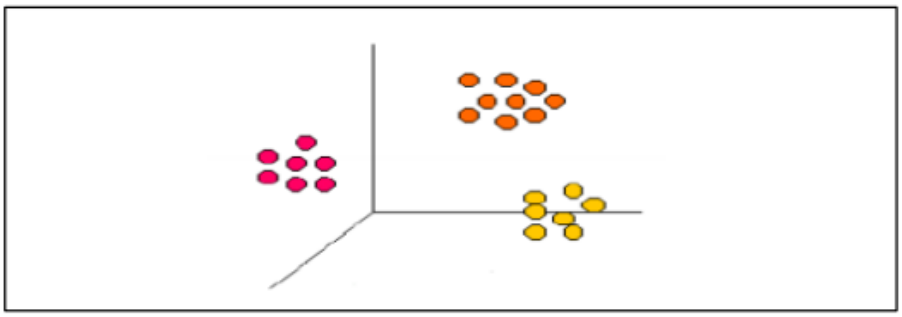
- Kümeleme

Kümeleme; heterojen bir yapıda bulunan verilerin, homojen bir şekilde alt gruplara bölünmesi işlemine denir (Göker,2019). Başka bir deyişle, önceki bilgilerin bilinmediği ya da sınıf sayısı belli olmayan durumlarda verilerin birbirine yakınlık derecesine göre yani benzer olmalarına göre gruplara ayrılmasıdır. Bu tahmin

yönteminde asıl amaç sınıflandırılmamış verileri, anlamlı bir şekilde gruplamaya yardımcı olan yöntemler topluluğudur (Can,2020; Kutlugün, 2017).

Kümeleme yöntemleri 2 başlık altında incelenir. Bu yöntemler; hiyerarşik ve hiyerarşik olmayan yöntemlerdir. Kümeler arasındaki benzerlikler düşünüldüğünde her iki yaklaşımda da ilişkinin düşük olması, kümeler içi benzerlikler baz alındığında ise benzerliklerin yüksek olması amaçlanmaktadır (Alpar, 2013). Kümeleme analizinde benzerlik dışında farklı kriterlere de yer verildiğinde veriler arasındaki uzaklık göz önünde bulundurulmalıdır. Bu uzaklıklar Euclidyen, Standardize Euclidyen, Manhattan Mahalanobis, Kareli Euclidyen, Minkowski veya Canberra ölçüleri kullanılarak hesaplanabilir. Bu hesaplamalar yapıldıktan sonra kümeleme algoritmaları uygulanır (Arslan, 2008). Kümeleme örneği Şekil 2.14'de gibidir:

Şekil 5: Koordinat düzleminde kümeleme örneği



Kaynak: Arslan, 2008.

Kümele analizi yapılırken küme sayısı belirlemek en önemli kriterdir. Çünkü küme sayısı analizin kalitesini doğrudan etkilemektedir. Bu sebeple bu analizin yapılmasında kullanılacak kümelerin sayısını belirlemek amacıyla pek çok yöntem geliştirilmiştir (Kavuncu,2018).

Kümeleme analizi psikoloji, biyoloji, tıp, sosyoloji gibi alanlarda daha sık kullanılsa da hemen hemen her alanda bu yöntemin kullandığı gözlemlenmektedir (Özdamar,2004). Örneğin; okuyucu kitlesi belli olan bir gazetenin, haber başlıklarına göre spor, teknoloji, eğitim gibi kategorilerde çevrimiçi haber önerisi sunması, nesne tanıma, müşteri segmentasyonu, pazar analizi gibi birçok alanda kullanımı mevcuttur.

Kümeleme analizi, yalnızca makine öğrenmesinin değil veri madenciliği, örüntü tanıma gibi pek çok alanda kullanılan bir istatistiksel veri analiz tekniğidir (Uzun, 2007). Bunun yanında, dokümanların ve kullanıcıların denetlenmesi ve sapan verilerin belirlenmesi gibi çok geniş bir kullanım alanına sahiptir. Kümeleme analizi yaparken kullanılan birçok algoritma bulunmaktadır fakat bunlardan en fazla kullanılanlar; K-Means, K-Medoids, OPTICS, DBSCAN, CobWeb algoritmalarıdır (Göker,2019).

- Birliktelik Kuralları

Birliktelik kuralı, veri tabanı içindeki değerlerin kümeleme yöntemlerinin dışında, birbiriyle olan bağlantılarını eş zamanlı olarak bulmaya çalışan bir veri madenciliği yöntemidir (Okur, 2020; Silahtaroğlu, 2008). Bu tekniğin kullanımdaki temel amacı, değerlerin birbiriyle olan ilişki potansiyelini ortaya çıkarmaktır (Ünsal, 2011). Birliktelik kuralının kullanım alanları çok çeşitli olmakla birlikte günümüzde sepet analizi, müşteri alışkanlıklarının tespit analizi, stok kontrolü, firmaların promosyon çalışmaları, mağaza düzeni ve satış stratejisi gibi alanlarında yaygın olarak kullanılmaktadır (Ay ve Çil, 2008).

Örneğin markete giden müşterilerin X ürününü alırken Y ürününü aldığı biliniyorsa ve yine bu müşterilerden birinin X ürününü alırken Y ürününü almadığı biliniyorsa bu müşteri potansiyel bir Y müşterisidir. Fakat sadece bir ürünün satın alınıp alınmadığı durumu söz konusu ise sepet analizi arasındaki bağlantılar destek ve güven ile hesaplanır. Destek ve güven formülleri aşağıdaki gibidir (Alpaydın, 2010).

$$\text{Destek} \quad : \quad P(X \text{ ve } Y) = \frac{\text{X ve Y mallarını satın almış müşteri sayısı}}{\text{Toplam müşteri sayısı}}$$

$$\text{Güven} \quad : P(X|Y) = P(X \text{ ve } Y)/P(Y) = \frac{\text{X ve Y mallarını satın almış müşteri sayısı}}{\text{Y malını satın almış müşteri sayısı}}$$

Birliktelik kuralını uygulayabilmek için sıkça kullanılan algoritmalar; GRI, FP Tree, CHARM ve Apriori ‘dır (Göker,2019).

- Boyut İndirgeme

Makine öğrenmesi yöntemlerini kullanırken büyük miktardaki verilerden kullanılmayan öz nitelikleri tespit edilerek ve eksiltilerek daha doğru sonuçlar verecek şekilde bu büyük miktardaki verinin boyutunu azaltmaya yarayan yönteme boyut azaltma denir (Şirin ve Kutlugün, 2017 akt; Kutlugün, 2017). Bu yöntem verinin doğru tahmin edilmesinde başarıyı attıran en önemli aşamalardan birisidir (Okur,2020). Bu aşamada, verinin doğru tahmin edilmesinde sürecin başarımını, doğruluğunu ve etkinliğini arttırmaya yarayan alt kümeler meydana getirildiğinden daha net sonuçlara ulaşılmaktadır.

Net sonuçlara ulaşma aşamasında kullanılan öznitelik seçme ve öznitelik çıkarma olmak üzere iki tip yaklaşım vardır (Dasgupta, Drineas, Harb, Josifovski ve Mahoney, 2007). Öznitelik seçme yaklaşımında amaç veri seti üzerinde gereksiz görülen özelliklerin kaldırılması iken öznitelik çıkarma yaklaşımında ise amaç ham veriler üzerinde işlem yapabilmek için gerekli olan özelliklere dönüştürme işlemi yapılarak verinin kullanıma hazır hale getirilmesidir.

2.2.2. Makine Öğrenmesi Algoritmaları

2.2.2.1. Çoklu Doğrusal Regresyon Modeli

Doğrusal regresyon literatürde en sık kullanılan modellerden biridir. Tek bir bağımsız değişken ile bağımlı değişken arasındaki ilişkiyi inceler. Temel amacı bağımlı değişkeni tahmin etmek olan çoklu doğrusal regresyon modelleri iki ya da daha fazla bağımsız değişkenin arasındaki ilişkiyi incelemektedir. Çoklu doğrusal regresyon modellerindeki ilişkiyi matematiksel olarak açıklamak gerekirse n tane değişken için: (Alpar,2013).

$Y= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n$

Bu kapsamda yukarıdaki formül incelendiğinde Y bağımlı değişkeni X1,X2,X3… bağımsız değişkenleri, β_0 değeri ise sabit değişkeni ifade etmektedir. Bağımsız değişkenlerde meydana gelen bir birimlik artışın bağımlı değişkende β'nın katsayısı kadar bir artışa ya da azalmaya neden olacağı bilinmektedir. Eğer β değeri negatif bir ifade içeriyorsa bağımsız değişkende meydana gelen bir birimlik artış

bağımlı değişkende β katsayısı kadar azalamaya neden olur. Fakat pozitif bir değer içeriyorsa bağımlı değişken β katsayısı kadar artar (Büyüköztürk, 2020).

2.2.2.2. Lojistik Regresyon

Bağımlı değişkenin kategorik ve sınıflandırma işlemleri için kullanılması yönüyle doğrusal regresyon modellerinden ayrılan makine öğrenmesi sınıflandırma yöntemlerinden birisi olan ve adını logit dönüştürmeden alan lojistik regresyon, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi açıklamak için kullanılan, öngörücü bir analiz çeşididir. Doğrusal regresyonda aranan varsayımların aksine Lojistik Regresyon modellerinde bu varsayımlar aramadığından dolayı daha esnek bir yapıdadır (Yeşilnacar ve Topal,2005).

Lojistik regresyon üç türden oluşmaktadır. Sıralı lojistik regresyon, kategorik sınıflandırmanın düzenli aralıklarla sıralı cevaplar verdiği durumlarda kullanılırken, multinominal lojistik regresyon bir sıraya sahip olmayan minimum üç ve ya üçten fazla kategorik cevabın yer aldığı durumlarda, ikili lojistik regresyon ise kategorik cevabın sadece iki sonuçtan meydana geldiği durumlarda kullanılır (Aydın,2020). Lojistik regresyon formülün başarı ya da başarısızlık oranlarının birbirlerine göre logaritmasının alınmasıyla oluşur. Bu fonksiyona bahis oranı da denmektedir.

$$\text{logit}\,(P) = \log \frac{P}{1 - P}$$

Yukarıda gösterilen bahis oranı formülünde yakınlık durumları 0 ve 1 e göre değerlendirilir. Birden büyük olma durumunda bağımsız değişkenlerin bağımlı değişkeni etkilediği gözlemlenmekte sıfıra yakın olma durumunda ise bağımsız değişkenlerin negatif bir etkisi olduğunu göstermektedir. Lojistik Regresyon modelinde bağımsız değişkenlerin katsayıları β olarak yazıldığında genel bir logit fonksiyonu elde edilir.

$\text{logit}\,(P) = \log P/(1-P) = \beta_0 + \beta_1 X_1 + \ldots + \beta_m X_m$

Fonksiyondaki P değerinin artması durumda logit(P) değeri onunla birlikte artış gösterir

Doğrusal regresyon analizinde yapıldığı gibi lojistik regresyon analizinde de bazı değişken değerlerine bakılarak tahmin yapılmaya çalışılsa da iki analiz yöntemi arasında büyük farklılıklar bulunmaktadır. Bu farklılıkları,

- Tahmin edilecek bağımlı değişkenin doğrusal regresyon analizinde sürekli olması gerekirken lojistik regresyonda ise kesikli bir değer olması
- Doğrusal regresyonda bağımlı değişken tahmin edilmeye çalışılırken, lojistik regresyon analizinde ise bu değişkenin alabileceği değerler arasından herhangi birinin gerçekleşme olasılığı tahmin edilmeye çalışılması
- Lojistik regresyon analizinin uygulanabilmesi için bağımsız değişkenlerinin dağılımına yönelik herhangi bir şart olmamasına rağmen doğrusal regresyon analizinde bu değişkenlerin çoklu normal dağılıma sahip olmasına yönelik bir şartın olması

gibi sıralamak mümkündür (Coşkun, Kartal, Coşkun ve Bircan 2004).

2.2.2.3. Ridge Regresyon

Ridge regresyon, EKK yöntemine alternatif olarak önerilen, temel amacı en küçük varyans değeri ile parametrelerin açıklanmasını sağlayan taraflı bir tahminleme yöntemidir (İmir, 1986). Ridge regresyon bağımsız değişkenleri birbirlerini etkileme oranını en aza indirgemek için kullanılır. Bu yöntemi ilk kez kullanan Hoerl and Kennard (1970), ridge regresyon yöntemini bağımsız değişkenleri standardize etmek, EKK yönteminden farklı olarak regresyon parametreleri için daha küçük varyanslı sonuçlar elde etmek için önermiştir (Üçkardeş, Efe, Narinç, ve Aksoy, 2012).

2.2.2.4. Lasso Regresyon

Lasso regresyon, literatürde yüksek miktardaki verilerin çözümlenmesinde, en küçük kareler kestirme yöntemini kullanarak tahmin başarı oranını arttırmaya yarayan, Tibshrirani, (1996) tarafından ortaya atılan bir makine öğrenmesi

algoritması olarak tanımlanmaktadır. İçinde bulunan ceza fonksiyonu sayesinde regresyon katsayılarını ayarlayarak, yapılan tahminlerin doğruluğunu arttırır.

Lasso regresyonun performans ölçüsü lamda (λ), ceza ölçütü ise "Manhattan Uzaklığı" olarak kullanılmaktadır. Manhattan uzaklığı olarak adlandırılan bu yöntemde mutlak değerlerin toplamı ceza parametresi olarak tanımlanır. Lasso regresyonda kullanılan lamda değeri modeli optimize etmede büyük fayda sağlar. Lamba değerinin yükselmesi durumunda Shrinkage devreye girerek daha iyi tahmin performansı sağlamaktadır (Jaggi, 2014).

2.2.2.5. Enet(Elastic Net) Regresyon

Elastic net regresyonu, ridge ve lasso regreyonun birleşiminden meydana gelmiştir. İçinde Ridge regresyonda olduğu gibi hem ceza ölçütü ve aynı zamanda lasso regresyonda olduğu gibi değişken seçimi yapılabilmektedir. Tahminleme işlemi lamda ile yapılır. λ sıfıra eşit olduğunda ridge regresyon, bire eşit olduğunda ise lasso regresyon kullanılmaktadır. Literatürdeki çalışmalara bakıldığında ise bu parametre genellikle ortalama olarak 0,5 değeri kabul edilir.

2.2.2.6. K-Means (K-Ortalamalar)

K-Ortalama algoritması kümeleme analizinde çokça kullanılan, veri kümesinde bulunan etiketlenmemiş değişkenlerin n kadar verinin k kadar kümeye ayrılması işleminin yapıldığı denetimsiz öğrenme modellerinden biridir**.** Bu kümeleme işlemi sonucunda kümeler içi ilişkinin homojen olması beklenirken, kümeler arası ilişkinin heterojen olması beklenmektedir. (Alzand ve Karacan, 2014)**.** Denetimsiz öğrenme modellerinden biri olan k-ortalama algoritması iki ana bileşenden oluşmaktadır. İlk bileşen için k merkez değeri bir girdi olarak belirlenmeli, ikinci bileşen için ise veri kümesinde yer alan nesnelerin tümünün kendine en yakın uzaklıkta bulunan merkez noktanın içerisindeki gruba dâhil olmalıdır (Li, Shen ve Chen, 2008). Gruba dâhil olduktan sonra, tüm kümlerin ortalamaları tek tek hesaplanır ve bu hesaplanmanın sonucunda yeni bir merkez oluşur. Oluşan bu merkezle nesne arasında yeniden uzaklık hesaplanarak, her nesne

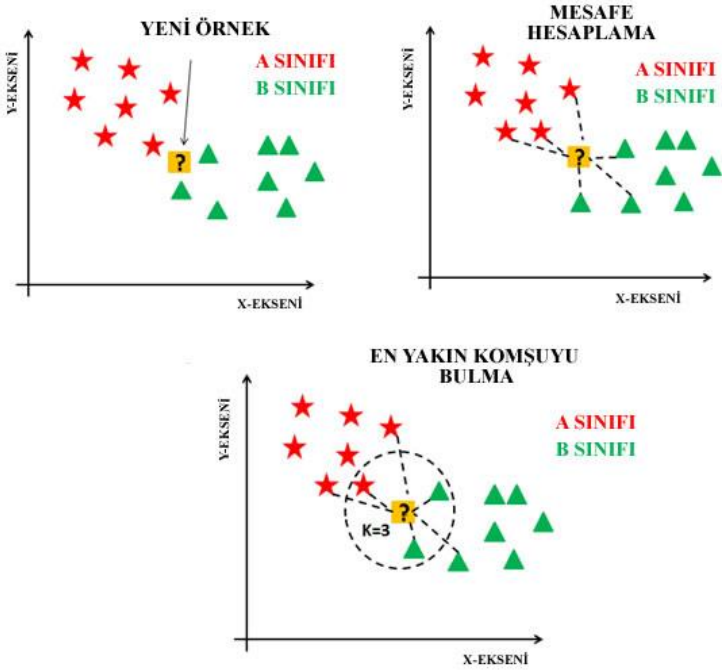
için hangi kümeye ait olacağı k merkezine olan uzaklıkları göz önüne bulundurularak belirlenir. Bu belirlemenin ardından nesnenin hangi merkeze olan uzaklığı en az ise o nesne o merkez kümesinin içinde yer alacak şekilde işlemler devam eder (Jiawei, 2006).

2.2.2.7. K - En Yakın Komşu Algoritması

K-En Yakın Komşu, bir nesnenin hangi sınıfa ait olduğunu belirlemek amacıyla regresyon ve sınıflandırma problemlerinde kullanılan, birbirine olan yakınlık derecelerine göre sınıflandırma işlemi yapan denetimli öğrenme algoritmasıdır (Özkan, 2008).

Görüntü işleme, kontrol sistemleri, veri madenciliği, film tavsiyesi, meteoroloji gibi birçok alanda faydalanılmakta olan KNN algoritmalarının temel amacı, sınıfı belli olmayan her bir nesnenin, o sınıf dışındaki bütün nesneler ile karşılaştırılması ve bu karşılaştırma sonucunda, kendisine en yakın k sınıfındaki örneklerin ortalaması ile birlikte bir gruba atanmasıdır (Özkan 2008). Atama işlemi sırasında, nesnelerin birbirine olan uzaklığını hesaplamak amacıyla jaccard distance, öklid uzaklığı, manhattan distance, simple matching distance formülleri sıklıkla kullanılmaktadır (Özkan, 2008; Lantz, 2013).

Şekil 6: K-En yakın komşu sınıflandırması modeli örnek gösterimi



Kaynak: KNN Classification using Scikit-learn,2018.

K-En Yakın Komşu uygulamalarının sade ve anlaşılır olması, verilerin dağılımlarında varsayımlar yapmaması ve verilerin eğitim sürecini minimuma indirgemesi gibi birçok güçlü yanı bulunmaktadır. Bunun yanı sıra sınıflandırma yaparken yavaş olması, yüksek hafızalı bilgisayarlar gerektirmesi, kayıp verilerin işlem yükünü arttırması gibi pek çok nedenden dolayı da zayıf yönü bulunmaktadır. Bu algoritma kullanılırken ilk olarak eğitim verilerinden öğrenme gerçekleştirilir ve bu öğrenme sonucunda bir k değeri belirlenir. Belirlenen bu k değeri, literatürde sıklıkla eğitim veri setinin karekökü şeklinde bulunur ya da 3 ila 10 arasında bir değer olarak belirlenir. K değeri sistem için uygun kümeleme sayısını ifade eder. Ardından veriler bu değere göre sınıflanır (Lantz, 2013).

2.2.2.8. Destek Vektör Makineleri

Destek vektör makinesi, sınıfları birbirinden ayırt etmek için bir düzlem belirten, gerçekleştirilecek olan işlemlerin açık olarak ifade edilmediği zamanlarda girdilere karşılık gelen çıktıların geçmişi göz önünde bulundurularak iki veya daha

fazla sınıfa ayırmaya ve tahmin etmeye yarayan, sıklıkla kullanılan bir makine öğrenmesi algoritmasıdır (Güner ve Çomak, 2011; Yılmaz, 2013).

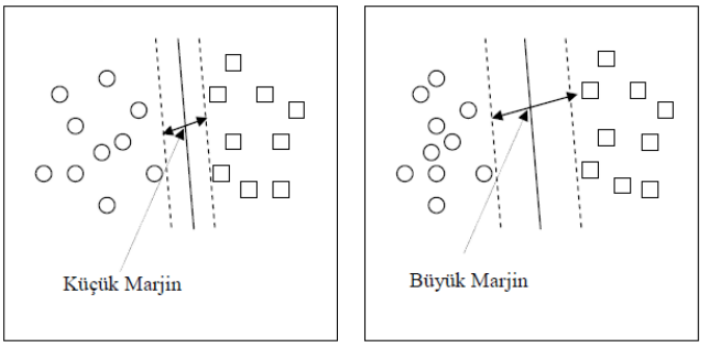
DVM öncelikle veri kümesini lineer olarak ayrılabilme durumuna bakılarak karar verebilmesi noktasında doğru şekilde gruplama yapılmasını sağlamaktır bu noktada DVM algoritmalarının temel amacı hiper düzlemi en iyi ayıran karar fonksiyonunu belirlemektir. Bu işlemi yapısal riski minimize ederek yapmaktadır (Demolli, Ecemiş, Dokuz ve Gökçek, 2019; Yılmaz, 2013).

$$f(x) = \sum_{n=1}^{n} w_i x_i + b$$

Yukarıdaki formülde belirtilen x değeri ayrımı yapılacak olan girdi verisini, b eşik değeri, w değeri ise ağırlık matrisini ifade etmektedir. Eğer fonksiyon değeri sıfırdan büyük ya da eşit ise x vektörü pozitif, sıfırdan küçük ise negatif olarak değerlendirilir (Uzun, 2007).

DVM pek çok alanda kullanılmaktadır bu kullanım alanlarını, astronomi, görüntü işleme, suç bilimi, meteoroloji gibi sıralamak mümkündür (Demolli, Ecemiş, Dokuz ve Gökçek, 2019).

Şekil 7: Maksimal marjin ve esnek marjin



Kaynak: Uzun, 2007.

DVM algoritmaları verinin gürültü içerme durumuna göre en yüksek marjin ve esnek marjin olmak üzere iki türe ayrılmaktadır (Uzun,2007). Bu yöntemlerde iyi bir sınıflandırma yapabilmek için birbirinden farklı grupta yer alan örnekler incelenerek bu örnekler arasındaki uzaklığı en çok arttıran hiper düzlemi belirlemek gerekmektedir (Kutlugün,2017).

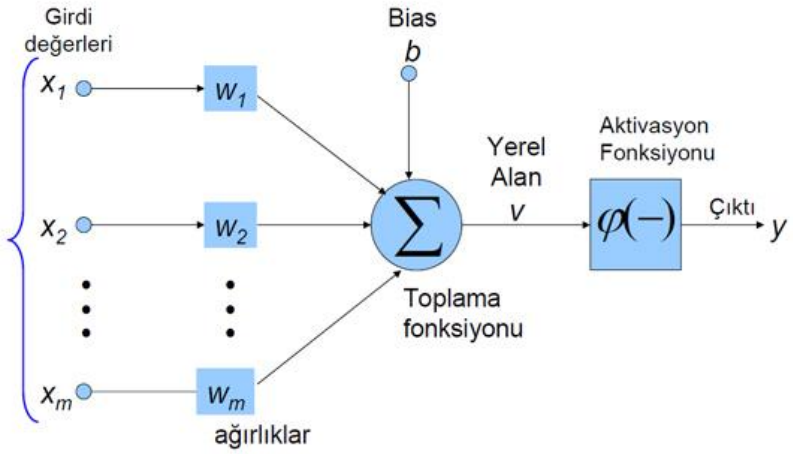
2.2.2.9. Yapay Sinir Ağları

Yapay sinir ağları (YSA) insan beyninin sinir sistemiyle benzer bir şekilde çalışan, biyolojik sinir ağları model alınarak yapılan gözetimli öğrenme algoritmalarından biridir (Kavuncu,2018; Öztemel, 2012). Bu algoritmanın amacı, insana has karar verme, yorum yapma, ilişki kurma, yeni bilgileri keşfetme ve üretme gibi insan beyninde öğrenme yolu ile gerçekleşen bu yetenekleri yardım almaksızın otomatik olarak gerçekleştirmektir **(**Öztemel, 2012; Uğur ve Kınacı, 2006)**.**

Yapay sinir ağları, sinir hücrelerinin birbirleriyle olan farklı birleşimlerinden meydana gelir. Çoğunlukla katmanlar halinde düzenlenir ve donanım şeklinde elektronik devreler ile birlikte tasarlanabilmektedir (Haykin, 1999).

Deneyimlerden yararlanarak öğrenebilme ve bilgiyi saklama yeteneğine sahip olan YSA algoritmasında model yapılandırma süreci, bir kısım verinin yapay sinir ağına iletilmesiyle başlar (Uğur ve Kınacı, 2006). Verinin yapay sinir ağına iletilmesinin ardından yapay sinir ağı çıktı değeri konusunda bir tahminde bulunur. Bulunan tahmini değer ve gerçek olan değer karşılaştırılır. Karşılaştırma sonucunda, ağ herhangi bir başka faaliyet göstermiyorsa bulunan tahmini değer doğrudur. Fakat tahminin kalitesini arttırmak için parametrelerin hangi şekilde, nasıl düzeltileceğini belirlemek için analiz yapılıyorsa yapılan bu tahmin yanlıştır (Marakas, 2003). Yapay sinir ağı hücresinin genel yapısı Şekil 8'teki gibidir.

Şekil 8: Yapay sinir ağı hücresinin genel yapısı



Kaynak: Yapay sinir ağları, 2018.

İnsan beyninin yapısından ilham alınarak tasarlanan yapay sinir ağları doğal dil işleme, ses işleme, görüntü işleme, tıp ve haberleşme gibi pek çok alanda yaygın olarak kullanılmaktadır (Kavuncu,2018).

2.2.2.10. Naive Bayes

Bayes teorisine dayanan, sınıflarda bulunan değişkenlerin özelliklerinin birbirinden farklı olduğu varsayılarak sınıflandırma yapan ve veri seti içerisindeki sınıfların yapısını öğrenerek buna uygun bir şekilde yeni bir gözlem geldiğinde hangi sınıfa ait olduğunu belirleyen basit ve etkili bir yöntemdir (Silahtaroğlu,2013).

Bayes teoreminde hesaplamalar koşullu olasılıklar ile bağımlı değişkenler üzerinden yapılır. Koşullu olasılık kavramı gerçekleşmiş bir olay üzerinden yeni bir durum meydana gelme olasılığına dayanır. Örneğin; daha önce yapılmış olan maçlar incelenerek çeşitli etkenler göz önüne alındığında bir maçın galibiyetle ya da mağlubiyetle bitmesi durumu tahmin edilebilir (Kavuncu, 2018).

Naive Bayes çalışma mantığını aşağıdaki formül üzerinden incelemek gerekirse;

$$P(Ci|X) = \frac{P(X|Ci)\ P(Ci)}{P(X)}$$

Bu formülde C değerleri sınıfları, X değerleri ise sınıfı bilinmeyen örnekleri ifade etmektedir. P(Ci |X), bir X değerinin Ci sınıfında olma olasılığını, P (X| Ci) ise C sınıfında yer alan bir X değerinin olma olasılığını ifade etmektedir (Özkan, 2008).

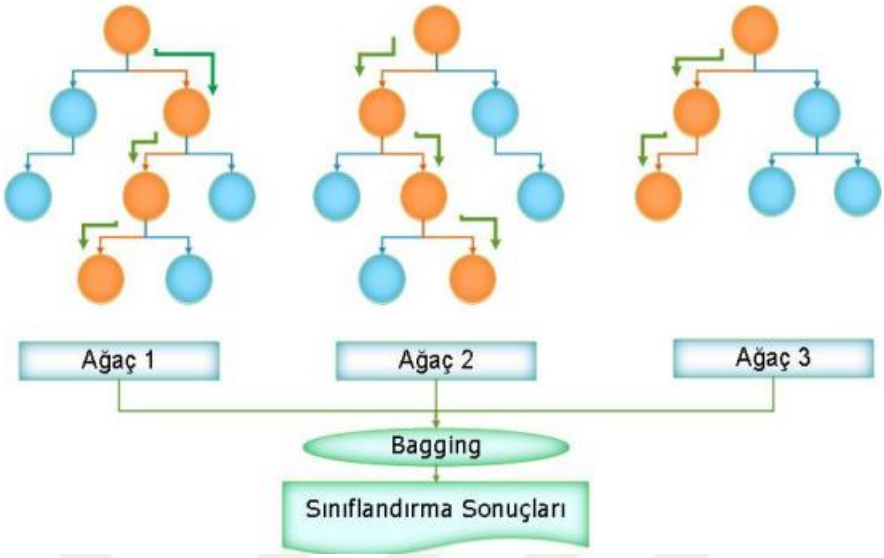
Naive Bayes algoritmaları verileri hızlı bir şekilde eğitilmesi ve yüksek doğruluk oranları vermesi sebebiyle oldukça avantajlıdır. Fakat sınıflandırma problemlerinin karmaşık olması durumda Naive Bayes algoritması yetersiz kalmaktadır (Uzun, 2007).

2.2.2.11. Rastgele Ormanlar

Bu yöntemde, veri seti eğitilirken çok fazla karar ağacı oluşturulur ve tüm bu karar ağaçlarının vermiş olduğu sonuçların ortalamaları alınarak problemin sonucuna ulaşılmaya çalışılır. Rastgele orman yönteminde bootstrap tekniği kullanılır. Bu tekniğin bagging tekniğinden farkı seçilen örnek ile birlikte değişken seçiminde rastgele olmasıdır (Ho, 1998).

Rastgele orman algoritması bir topluluk öğrenme yöntemidir, bu yöntemler aşırı öğrenmeyi engellese de, sapmayı arttırmaktadır. Bu yüzden bu durumu engellemek amacıyla torbalama yöntemi kullanılmaktadır (Kalaycı,2018). Rastgele ormanlarda, kullanılan ağaçların sayısı ve bu ağaçların doğru oylanması karar ağaçlarının yüksek performans göstermesini sağlayacaktır. Aşağıdaki şekilde bir rastgele orman örneği verilmiştir (Kavuncu,2018).

Şekil 9: Rastgele orman ağaç yapısı örneği



Kaynak: Kavuncu,2018.

Rastgele orman yönteminde düğümler tüm değişkenler arasından belirli miktarda rastgele seçilerek, bölünmeyi en iyi yapacak olan değişken belirlenip ağaç dallarına ayrılır. Değişken sicimi için kullanılan random subspace yönteminin kullanılmasındaki temel amaç en optimal dal bölünmesini sağlayan değişkeni bulmaktır. Tüm bu işlemlerin ardından karar ağacı algoritmaları uygulanır ve bu sayede dallara ayırma işleminde hangi değişkenin kullanılacağı karar verilmiş olur. Eğitim ve test veri seti için gini katsayısı belirlenir. Ardından karar ağacı yöntemiyle eğitilen model, yeni bir test verisi geldiğinde, eğitim kısmında oluşan ağaçlardan uygun kısmına yerleştirilir. Her ağaç kendi içerisinde bir tahmin tahminleme yapar, bu tahmin değerleri bir oylamaya sunulur, bu oylama sonucunda en çok oyu alan tahmin değeri ise nihai değer olarak sunulur (Ho, 1998; Breiman, 2001).

2.2.2.12. Karar Ağaçları

Karar ağaçları veri setinde bulunan değişkenleri, arama sezgilerini kullanarak, sınıflandırma, kümeleme ve tahminleme problemlerinin çözümde kullanmak için ağaç modeli oluşturan gözetimli öğrenme yöntemidir (Rocakh,2008; Keskin, 2018).

Karar ağaçları, diğer yöntemler ile karşılaştırıldığında veri setleri içerisindeki karmaşık yapıları basit karar yapılarına dönüştürebilmesi bakımından daha anlaşılır ve yorumlanması daha kolaydır (Keskin,2018).

Karar ağaçları kök ile başlayarak, düğüm, dal ve yaprak olmak üzere kısımlara ayrılır ve sınıflandırma yapmak için kök kısımdan yaprak kısma doğru hareket halindedir (Uzun, 2007; Kutlugün,2017). Karar ağacında düğüm olarak adlandırılan kısım soruları, dal olarak adlandırılan kısım bu sorulara verilen yanıtları, yaprak kısmı ise kararlaştırılan sınıfı temsil etmektedir (Yılmaz, 2013).

Karar ağaçları birbirinden farklı olan veri setlerini, belirli bir hedef değişkeni doğrultusunda, karar verme kuralları ile benzer yani homojen alt gruplara ayırır (Linoff ve Berry, 2004). Karar verme süreci içerisinde algoritmaların temel adımları neredeyse birbiriyle aynıdır. İlk olarak bir kök düğüm oluşturulur. Verilerin tamamı düğümden yani karar verme noktalarından geçer. Düğümde yer alan örnekler benzer ise kök yaprağa dönüşür, örnekler farklı ise karar kuralları doğrultusunda dallara ayrılır. Dallara ayrılmasına bir düğüme düşen örneklerin saflığı göz önünde bulundurularak karar verilir. Saflığı ölçmek için Kategorik Ki-Kare, Gini, Entropi gibi kriterler kullanılır (Murphy,2012). Oluşan hiçbir dal birbirine karışmaz ve ağaç şekline benzeyen bir yapı oluştururlar.

Karar ağaçlarını oluşturabilmek için, tüm düğümlerin hangi kritere göre ve nasıl dallanacağına karar verme noktasında yardımcı olacak birbirinden farklı pek çok sayıda yöntem geliştirilmiştir. Literatürde CHAID, CART, ID3 ve C4.5 bu yöntemler içerisinde en çok kullanılan algoritmalardır (Wu,Chen ve Chain, 2006; Quinlan, 1986; Loh, 2011). Bununla birlikte literatürde, karar ağaçlarını merkeze alarak geliştirilen, daha güçlü tahmin sonuçları veren Artırma (Boosting), Torbalama (Bagging), ve Rastgele Orman (Random Forest) olmak üzere üç tür topluluk algoritmaları da yer almaktadır (James, Witten, Hastie ve Tibshirani, 2013).

Karar ağacı temelli analizler, parametrik model kurulumunda yararlanmak için fazla sayıda değişkenler içerisinden en önemlilerini seçmede, yalnızca belirli bir gruba ait olan ilişkileri tanımlamada, farklı vakaların düşük, orta, yüksek risk grupları olarak kategorize edilmesi vb. gibi pek çok alanda sıklıkla yararlanılmaktadır (Akpınar, 2000).

2.2.2.13. Polinom Regresyon

Polinomal regresyon değişkenler arasındaki ilişkinin linear olmadığı durumda kullanılmaktadır. Bağımlı ve bağımsız değişkenlere göre oluşturulan polinom eşitliği 2.12'de gösterilmiştir.

$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_m x^m + \mathscr{e}$

Yukarıdaki eşitlikte β_0 sabit değeri, $\beta_1 + \beta_2 + \beta_3 + \ldots \beta_m$ değerleri katsayıları ifade etmektedir. m değeri ise polinomun derecesini göstermektedir.

2.2.2.14. Light Gradyan Arttırma (LGBM)

Light Gradyan artırma algoritması, sınıflama ve sıralama için kullanılan, karar ağaçlarından yararlanan, hızlı ve sağlam bir gradyan artırma modelidir. LightGBM'i diğer gradyan artırma algoritmalarından farklı kılan karar ağaçlarının eğitilmesi sırasında kullandığı büyüme yönüdür. Diğer karar ağaçları analiz esnasında dikey şekilde bir büyüme yaparken, LightGBM ise yatay yönlü bir büyüme gerçekleştirmektedir (Ke, Meng, Finley, Wang, Chen, Ma, Ye ve Liu, 2017).

LightGBM daha az bilgisayar hafızasına ihtiyaç duyması, işlem hızını arttırması ve daha büyük verileri işleyebilmesi noktasında diğer gradyan artırma modellerini geride bırakarak, gerek veri bilimi gerekse uzaktan algılama alanında çalışmalar yapan araştırmacıların dikkatini çekmiştir (Ke ve diğer, 2017; Üstüner, Abdikan, Bilgin ve Balık Şanlı, 2020).

LightGBM, diğer gradyan artırma modellerinden ayıran bir diğer özellik ise içerisinde kategorik değişken bulunan verilerin, one-hot encoding gibi işlemlere ihtiyaç duymadan analiz edebilmesi ve büyük veri setleri üzerinde daha kolay uyum yakalayabilmesi sebebiyle daha iyi sonuçlar vermesidir. Literatür bilgileri ışığında, LightGBM modelinin veriyi öğrenme aşamasının diğer modellere kıyasla 20 kat daha hızlı olduğu sonucuna ulaşılmıştır (Ke ve diğer, 2017).

2.2.2.15. Aşırı Gradyan Arttırma (XGBOOST)

Gradyan arttırma modelleri, karar ağaçları temeline dayanan, tahminleri iyileştirmek amacıyla geliştirilmiş yüksek performanslı sınıflandırma yapabilen bir makine öğrenmesi algoritmasıdır (Chen ve Guestrin,2016).

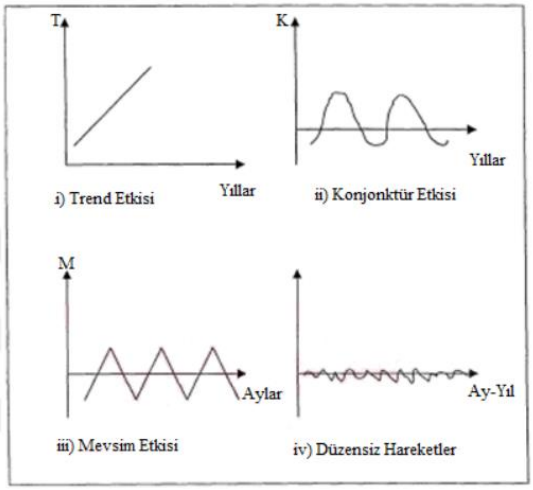
XGBoost algoritması, veriler eğitilirken overfitting (aşırı öğrenme) sorununa çözüm bularak, modelin daha doğru çalışmasını sağlamaktadır (Chen ve Guestrin, 2016; Rumora, Miler ve Medak, 2020). Bunu sağlamak amacıyla, karar ağaçlarındaki gibi ağaç oluştururken, bir önceki ağaçta meydana gelen hataları minimize ederek, veri seti içerisinde daha doğru sonuçlar elde edilmesini sağlamaktadır (Şahinarslan, 2019).

2.3. ZAMAN SERİSİ

Zaman serileri, elde edilen gözlem sonuçlarının, kronolojik bir sıraya göre zaman içerisindeki değişimini gösteren veri dizileridir. Zaman serilerinde bulunan zaman aralıkları saat, gün, ay, yıl gibi periyodik bir döngüde eşit aralıklarla sıralanır (Gujarati, çev. 2017). Veriler birbirini izleyen belirli zaman aralıklarına göre gözlemleniyorsa bu zaman serileri özel olarak isimlendirilir. Örneğin; bu veriler yılda bir defa ise yıllık, yılda iki defa ise altı aylık, yılda dört defa ise mevsimsel, senenin her bir günü içinse günlük vb. gibi (Kuzu,2013). Zaman serisi analizlerinde, zaman aralıklarına göre tahminleme modeli oluşturulmakta ve oluşturulan bu model geçerli hale getirildikten sonra serinin geleceğe yönelik tahminlemesi yapılmaktadır. Zaman serileri analizleri ile yapılan bu tahminler sayesinde doğru bir yol belirlenerek tedbir alma imkânı sağladığından dolayı gelecekle ilgili endişeleri minimum seviyeye indirmektedir (Gujarati, çev. 2017).

Zaman serilerini meydana getiren bileşenlerin oluşması için birbirlerinden farklı zamanlarda yapılan gözlemlerin birbiriyle olan bağımlılıklarının incelenmesi gerekmektedir. Zaman serileri Trend Bileşeni, Konjonktür Bileşeni, Mevsim Bileşeni ve Rassal Bileşen olmak üzere dört bileşenden oluşmaktadır (Çevik,1999; Özek, 2010; Tarı,2010). Bu bileşenlerin örnek gösterimi aşağıdaki gibidir:

Şekil 10: Zaman serisi bileşenleri örnek gösterimi



Kaynak: Çevik,1999.

Trend Bileşeni: Zaman serilerinde uzun süre istikrarlı bir şekilde yükselme, alçalma ve sabit kalma eğilimine sahip olma durumudur.

Mevsimsel Bileşen: Zaman serilerinde belirli dönemlerde farklılık görülür. Bu farklılıklar mevsimsel değişimler olarak adlandırılır.

Konjonktürel Bileşen: Zaman serilerinde sabit uzunluğu olmayan ve sürekli olarak meydana gelen dalgalanmalardır.

Rassal Bileşen: Diğer bileşenler tarafından ifade edilemeyen değişimlerdir (Arda,2020).

2.3.1. Zaman Serisi Analizleri Kullanım Amaçları

Zaman serisi analizleri kullanım amaçlarına göre dörde ayrılır (Şeker,2015).

Şekil 11: Zaman serisi analizi kullanım amaçları

Kaynak: Şeker,2015.

1-Aykırı(Outlier) Verileri Yakalama Aykırı veri; Mevcut verilerden zamana göre bir sıralama yapılır, Yapılan sıralama sonucunda hatayla veya yanlış girilerek ortaya çıkmış verilerin hareketlerinde zamana göre farklılıklar bulunur. Oluşan farklılıklar aykırı verileri gösterir. Bu aykırı değerleri yakalamak için zaman serisi analizi yapılır ve zaman hareketleri üzerindeki farklılıklar gözlemlenir (Şeker, Mert, Al-Naami, Ozalp ve Ayan 2013).

2-Tahmin (Prediction): Mevcut verileri kullanarak tahminler üretme işlemidir. Bu tahminler gerçekleşmesi muhtemel olayları ifade eder (Şeker, Cankir ve Okur 2014).

3-Eksik Verileri Tamamlama(Imputation): Mevcut veriler üzerinde bazen eksik veriler bulunabilir, zaman serisi analizini kullanarak bu eksik veriler tamamlanabilir (Honaker,2010).

4-Hata Düzeltme(Data Scrubbing): Elde bulunan verilerin aykırı değerler içermesi durumunda, diğer verilere yakınlaştırmak için zaman serisi yöntemleri kullanılır (Crosswhite, 2003).

2.3.2. Zaman Serisi Modelleri

Zaman serileri analizinde, verilerin belirli bir sıra üzerinde modellenerek gelecek hakkında tahminleme yapması büyük önem arz etmektedir. T zamandaki mevcut gözlemler bilindiğinden t+1 zamandaki değer tahmin etmeye çalışılır. Zaman serisi

analizlerindeki tahminlerin doğruluğu, tahminlerin güven aralıkları ve limitleri hesaplanarak bulunur (Box, Jenkins, Ljung ve Reinsel, 2016). Zaman serisi modelleri, durağan, durağan olmayan ve mevsimsel model grubu olmak üzere üçe ayrılır. AR(p), MA(q) ve ARMA(p, q) durağan model grubunda, ARIMA(p, d, q) durağan olmayan model grubunda ve ARIMA(p, d, q) (P, D, Q) mevsimsel model grubunda yer alır (Özmen, 1986).

2.3.2.1. Oto Regresif Süreç (AR)

Bu modelde değişkenlerin geçmiş verilere bağlı olduğu düşünülerek tahmin modeli oluşturulur. Bu modeli kullanılırken kısmi oto korelasyon ve oto korelasyon katsayılarının dağılımlarına bakılır. Bu dağılımlar doğrultusunda oto korelasyon katsayısı üstel biçimde sıfıra yakınsa AR modeli olduğu gözlemlenir (Tarı,2010). AR (p) modelinin eşitlik denklemi aşağıdaki gibidir:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + a_t$$

2.3.2.2. Hareketli Ortalamalar (MA)

Geçmiş dönemlerde bulunan belirli miktardaki gözlem değerlerinin hata terimleri ile aynı dönemdeki hata terimlerinin lineer bileşimi olarak ifade edilen zaman serisi modelleridir (Duru,2007). MA (q) modelinin eşitlik denklemi aşağıdaki gibidir:

$$x_t = \theta_0 a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

2.3.2.3. Oto Regresif Hareketli Ortalama (ARMA)

AR ve MA modellerinin birleşimiyle oluşan bu model, durağan zaman serilerinin modellenmesi amacıyla kullanılır. Geçmiş dönemlerde bulunan gözlem değerlerinin hata terimleri ile ondan önceki dönemdeki hata terimlerinin lineer bileşimi olarak ifade edilen zaman serisi modelleridir (Duru,2007). (θ_0 = 1) iken ARMA(p,q) modelinin eşitlik denklemi aşağıdaki gibidir:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \theta_0 a_t - \theta_1 a_{t-1} - \ldots -$$

2.3.2.4. Oto Regresif Entegre Hareketli Ortalama (ARIMA)

Zaman serilerinin neredeyse çoğunda durağanlık yoktur. Bu serilerin durağan olmamasının trend, mevsimsel, konjektural gibi çeşitli nedenleri olabilmektedir. Doğrusal olmayan bu verileri ilk olarak durağan hale getirip ardından işlem yapılması gerekmektedir. Fark alma işlemi yapılarak durağan olmayan serileri, durağan hale getirme işlemine entegre modeller adı verilir. Parametre derecesi p olan oto regresyon, derecesi q olan ise hareketli ortalama ve seriler üzerinde d kez fark alma işlemi yapılıyorsa bu model ARIMA (p,d,q) olarak adlandırılır (Duru,2007). ARIMA (p,d,q) modelinin eşitlik denklemi aşağıdaki gibidir:

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \ldots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \theta_1 a_{t-2} - \ldots - \theta_q a_{t-q}$$

ÜÇÜNCÜ BÖLÜM
ARAŞTIRMA YÖNTEMİ VE KÜÇÜK VE ORTA ÖLÇEKLİ FİRMALAR İÇİN KARAR DESTEK SİSTEMİ GELİŞTİRİLMESİ

3.1. ARAŞTIRMA YÖNTEMİ

Literatürlerdeki mevcut karar destek sistemleri incelendiğinde, karar destek sistemlerinin çoğu sadece geçmiş verileri anlamaya, yorumlamaya çalışmakta ve gelecek hakkında bize çok fazla bir bilgi verememektedir. Mevcut karar destek sistemlerini daha ileriye taşımak amacıyla küçük ve orta ölçekli firmaların kullanabileceği daha iyi bir sistem yapılması düşünülmüş ve uygulamaya dökülmüştür. Geliştirilen bu uygulamanın ne kadar iyi çalıştığı Kaggle sitesinden alınan veri seti üzerinde test edilmiştir. Test edilen bu verilerin başarı oranları göstermektedir ki oluşturulan bu yapı başarılı bir şekilde çalışmaktadır. Bu uygulama geliştirilirken karar vericilere yardımcı olmak amacıyla en iyi model sunulmuş ve bu sayede bulunan algoritmalar içinden en uygun değerin bulunması ve buna göre işlem yapması kolaylaştırılmıştır.

Bu uygulama sayesinde yalnızca makine öğrenmesi algoritmaları değil, aynı zamanda zaman serisi analizleri, kümeleme işlemleri de bu uygulama üzerinden yapılabilmektedir. Bu sayede karar verme konumunda olan kişiler için ellerindeki verileri daha iyi anlayabilmek amacıyla çeşitlilik sağlanmıştır. Yöneticiye bu raporları sistem otomatik olarak oluşturmaktadır. Bu yüzden karar verme aşamasında yöneticiler için uğraş gerektirmemekte ve zamandan tasarruf etmelerini sağlamaktadır. Oluşturulan bu sistemin öngörüleri sayesinde karar verici konumunda bulunan kişilere çıkarımsal çok fazla imkân tanımaktadır.

Bu uygulama içerisinde bulunan ve uygulamanın bütününde kullanılacak olan verilerin düzenlenmesi işlemleri bu uygulama sayesinde daha kolay bir şekilde gerçekleştirilmektedir. Veri düzenleme sayesinde kişiler ellerindeki verileri daha kolay manipüle edebilmektedirler. Bu sayede verilerden raporlama işlemleri yapılırken, tahminleme işlemleri, makine öğrenmesi algoritmaları, zaman serisi analizleri veya kümeleme analizi yapılırken meydana gelebilecek, gözden

kaçabilecek kısımlar burada çözüme kavuşturulmuştur. Bu temizlenmiş veriyle daha doğru işlemler yapılabilmesi sağlanmaktadır.

Yapılan bu uygulamanın geliştirilmesi kısmında Yazılım geliştirme yaşam döngüsünün aşamaları takip edilmiştir.

3.1.1. Yazılım Geliştirme Yaşam Döngüsü

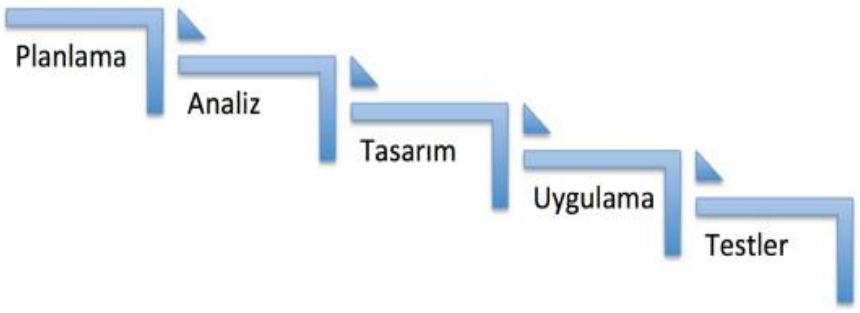
Yazılım bir üründür ve ürün geliştirme işlemleri bir süreçten geçer. Yazılım geliştirme yaşam döngüsü işletmenin ihtiyaçları doğrultusunda, sistem tasarımının yapılması, oluşturulması ve son kullanıcıya teslim edilme sürecidir (Dennis, Wixom ve Roth, 2009).

Bu süreç adımları birbirinden farklı değildir, bir sarmal halinde hareket eder ve uygulama var olduğu sürece tekrar eder. (Şekil 1). (Tecim, t.y.)

Bu çalışma kapsamında yazılım geliştirme yaşam döngüsü öğeleri aşağıdaki gibidir:

1. Planlama
2. Tanımlama
3. Tasarım
4. Geliştirme
5. Testler
6. Bakım

Şekil 12: Sistem Geliştirme Yaşam Döngüsü Örnek Gösterimi



Kaynak: Şeker,2015.

3.1.1.1. Planlama

Planlama aşaması, yazılım geliştirme yaşam döngüsünün ilk aşamasıdır ve bu kısımda iki önemli soruya yanıt aranır; (1) niçin bu sistemin geliştirileceği , (2) Bu sistemin nasıl geliştirileceğidir (Dennis, Wixom ve Roth, 2009).

Planlama aşamasında, günümüz küçük ve orta ölçekli firmaların verilerini analiz etmede yetersiz kaldığı gözlenmektedir. Temel problemlerin başında KOBI'lerin verileri ön işlemede yaşadıkları sıkıntılar vardır. Ellerindeki verileri raporlara dökmemeleri, geleceği analiz edememeleri, verileri analiz edecek nitelikli elemanları çalıştıramamaları, yüksek maliyetli programların ve bu programları yönetecek ekibinin olmayışı gibi birçok nedenden dolayı KOBİ'ler ellerindeki verilerden yeterince yararlanamamaktadırlar. Veriler çoğu işletmede excel gibi paket programları üzerinde veya bilgisayar ortamlarında kaldığından anlamlı sonuçlar çıkarmak için çok fazla argüman bulunmamaktadır. ERP programları kullanan KOBİ'lerde genellikle veri tabanları ve kendi üzerinde raporlama ekranları bulunmasına rağmen çoğu istenileni verememekte ve yetersiz kalmaktadır. Çünkü programlar klasik yapılar üzerine kurulmuş ve sadece istenilen sorguyu çalıştıran ve sorgulama sonuçlarını ekrana döken genel verilerden ibarettir. Bu gibi pek çok nedenden dolayı KOBİ'lerin ellerinde bulunan sistemlerin yetersizliği ve geleceğe dair tahminler ve analizler yapabilen sistemlerin bulunmaması piyasa dinamikleri açısından büyük eksiklik yaratmaktadır. Bununla birlikte raporlama ekranlarından filtrelemeler bulunmayan, kullanıcı etkileşimi sağlayamayan ve çok fazla raporu aynı anda sağlayamayan gösterge panelleri büyük eksiklik yaratmaktadır. Tam bu aşamada bu sorunların giderilmesi, yalnızca geçmişin değil geleceğinde dinamiklerinin yakalanması adına bu geliştirilen programın KOBİ'lere daha fazla yarar sağlayabileceği düşünülmektedir.

Bu yüzden KOBİ'lerin ihtiyaçları ve problemleri göz önüne alınarak yapılan uygulamanın amaçları aşağıdaki gibi sıralanmıştır:

1. KOBİ'lerin standart raporlar yerine kullanıcı etkileşimi yapabilecekleri gösterge panellerinin tasarlanması ve bunların yönetilebilir olması,

2. Veri tabanlarında bulunan verilerin ya da excel tarafında tutulan verilerin sorun çıkarmaması için anlamlı hale getirilip temizlenmesi işlemlerinin yapılması,
3. Karar vericiler için geleceğe yönelik tahminleme yapabilecekleri kısımlar, en uygun sonucu veren yöntemlerin sunulması
4. Zaman serisi analizleri ve kümeleme analizleri ile dinamiklerin ve trendlerin tespiti ve bunların karar vericilere sunulması.

3.1.1.2. Analiz

Analiz safhası, uygulamanın tanımlanmış işlevlerinin, hedeflerinin, ne yaptığının ve ne yapacağının belirlendiği aşamadır. İhtiyaçların analizi, tutarsızlıkların ve eksikliklerin giderildiği aşamadır. Bu aşamada geliştiriciler tarafından öncelikle ihtiyaçlar belirlenir, ardından sistemin dezavantajlarını gidermek amacıyla mevcut sistem incelenir, son olarak da mevcut sistemin eksikliklerine çözümler bulunur (Sistem geliştirme yaşam döngüsü, t.y.).

Bu kapsamda incelendiğinde, KOBİ'lerin kullandığı mevcut sistemlerde tutulan verilerin çok fazla eksik, hatalı veri barındırması sebebiyle elde bulunan tüm veriyi tek tek düzenlemek çok fazla zaman kaybına neden olmaktadır. Bu yüzden elde bulunan veriyle işletmeler uğraşmak istememektedirler. Bu verileri düzenlemek için ilk olarak bir ara yüz hazırlanmasına karar verilmiştir. Bu ara yüz üzerinden eksik veriler düzenlenebilmekte, medyan ile boş satırlar doldurulabilmekte, binary olarak dönüştürülebilmekte ve veriler istenilen değere dönüştürülebilmektedir. Bir sonraki aşamada işletmelerin yalnızca verileri düz grafikler üzerinde görmesinin yeterli olamayacağı düşünülmüş ve kullanıcı etkileşimi sağlayacak bir gösterge paneli tasarımı hazırlanmıştır. Burada karar vericilerin istediklerinde istedikleri grafikleri manipüle edeceği bir alan yaratılması planlanmıştır. Ayrıca tasarlanan sistem üzerinde karar vericilerin daha fazla öngörü yapabilmesini sağlayabilecek bir makine öğrenmesi sistemi yapılması düşünülmüştür. Bu sayede yöneticilerin karar verme ve planlama yapabilmesini verileri yalnızca eldekilerle değil geleceğe dair çıkarımlarla da desteklenmesi öngörülmüştür.

3.1.1.3. Tasarım

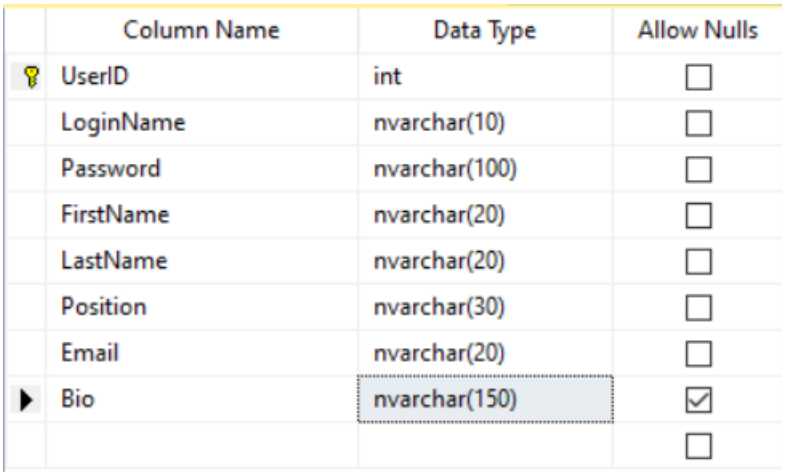
Tasarım aşaması, sistemin yapısının oluşturulduğu ve bütün bileşenlerinin hazırlandığı kısımdır. Genellikle bu aşamada sistemin hedefi ve etkileşimde bulunduğu bileşenlerin uyumu gözetilir. Ardından bu mantıksal tasarım fiziksel tasarıma çevrilir. Bu evrede kullanılacak olan tüm sistem parçaları, programlama dili gibi tüm evrelerin tasarımı hazırlanır (Şeker,2014).

Bu aşamada KOBİ'ler için geliştirilen programın ara yüz ve veri tabanı tasarımı yapılmıştır. Ara yüz hazırlanırken C# ve Python birlikte kullanılmıştır. C# üzerinde giriş ekranı tasarımları ve programın ara yüz tasarımı hazırlanmıştır. Ara yüz tasarımında kullanıcının giriş yapabilmesini sağlayan bir ekran hazırlanmıştır. Buna göre kullanıcı ilk olarak verilmiş olan kullanıcı adı ve parola ile giriş yapacak ardından verilmiş olan yetki kapsamında menülere erişim sağlayacaktır. Bu aşamada rapor hazırlayan ya da veri düzenleme işlemini yapacak kişi farklı olacağından menülerle yetkilerle aktif olamayacak şekilde ayarlanmıştır. Bununla birlikte verilerin düzenlenmesi işlemlerinin yapıldığı kısımlar da c# winform aracılığıyla yapılmıştır. Veri düzenleyecek ya da rapor hazırlayacak kişilerin nasıl yapacaklarını bilmediklerinden dolayı bir yardım ekranı hazırlanmış ve bu kısımda olabildiğince her menünün ne işe yaradığı tek tek anlatılmıştır. Bununla birlikte raporlama kısmı içinde Devexpress Kütüphanesi kullanılmıştır. Bu sayede rapor kısımları hazırlanmıştır.

Düzenlenen veriler kaydedildikten sonra makine öğrenmesi, zaman serisi ve kümeleme analizi yapabilmek için C#'tan python ekranına geçiş yapmaktadır. Bu kısımda C# ve python birbiriyle entegre bir şekilde çalışmaktadır. Python ara yüzü yazarken tkinter kütüphanesi kullanılmıştır. Bu sayede C# tarafında düzenlenen veriler python tarafında otomatik olarak içeriye aktarılmaktadır.

Hazırlanmış olan uygulamanın veri tabanı MsSQL üzerinde oluşturulmuştur. Bu kısımda yalnızca kullanıcıların bilgileri tutulmaktadır. Buna göre giriş çıkış işlemleri yapılmaktadır. Hazırlanmış olan veri tabanı tasarımı Şekil 13'de gösterildiği gibidir.

Şekil 13: Veri Tabanı Tablo İçeriği



Column Name	Data Type	Allow Nulls
UserID	int	☐
LoginName	nvarchar(10)	☐
Password	nvarchar(100)	☐
FirstName	nvarchar(20)	☐
LastName	nvarchar(20)	☐
Position	nvarchar(30)	☐
Email	nvarchar(20)	☐
Bio	nvarchar(150)	☑
		☐

Kaynak: Yazar tarafından derlenmiştir.

Oluşturulan sistem verilerin düzenlenmesi, makine öğrenmesi ile tahminleme ve raporlama kısımları olarak 2 katmandan oluşmaktadır. İlk katmanda verilerin düzenleme işlemi yapılır. İkinci katman ise, düzenlenen verinin tahmin ve raporlar yapılmak üzere hazır hale getirildiği kısımdır. Düzenlenen verilerden yapılabilecek olan raporlar; zaman serisi, makine öğrenmesi, kümeleme analizi ve açıklayıcı raporlamadır. Aynı zamanda 15 farklı tahminleme algoritmasını kullanarak tahminleme işlemi yapılmaktadır.

Şekil 14: Sistem Katmanları

Kaynak: Yazar tarafından derlenmiştir.

3.1.1.4. Uygulama

Daha önceki aşamalarda düşünülen ve kâğıda dökülen tüm süreçlerin gerçekleştiği ve bir yazılım haline getirildiği aşamadır**.** Tüm testlerin, kodlama, kod kontrol ve yönteminin yapıldığı aşamadır (Şeker,2014).

Geliştirilen bu uygulama kullanıcı etkileşimli rapor yapısı, bir yöneticin sorunlarına cevap verebilecek bir geliştiricinin ise verilerini kolayca yapılandırabileceği yapıda tasarlanmıştır. Ayrıca bu uygulama çeşitli ihtiyaçlara göre yeniden tasarlanabilecek, kullanıcı isteklerine göre şekillendirilebilecek yapıda olmasından dolayı dinamik bir uygulamadır. Yapılan bu çalışma c# winform ve python ile tasarlanmıştır.

Uygulama ara yüzü c# winform yapısı ile hazırlanmıştır. Winform yapısının python'a göre daha gelişmiş olmasından dolayı bu yapı tercih edilmiştir. Geliştirilen uygulama ara yüzünün daha modern gözükmesi için ise Bunifu kütüphanesi kullanılmıştır. Bu sayede klasik program ara yüzlerinden farklı bir tasarım elde edilmeye çalışılmıştır. Uygulama temelini oluşturan veri düzenleme kısmında ise veriler daha kolay işlenebilmesi adına excel ya da csv olarak gelen verilerin düzenlenip, tekrar pythona aktarılması aşamasında büyük önem arz eden kısım olmaktadır. C# ile python arasında bağlantı kurulmuş c# tan csv olarak kaydedilen veri python tarafında karşılanmaktadır. Raporlama kısmında Devexpress kütüphanelerinden yararlanılmıştır. Bu sayede csv, sql, excel fark etmeksizin gelen verinin raporlara dönüştürülmesi kolay hale getirilmiştir.

Tahminleme işlemlerinin yapılabilmesi ve raporların hazırlanabilmesi için makine öğrenmesi, zaman serisi veya kümeleme analizi gibi işlemlerin yapılabilmesi adına python programlama dili kullanılmış ve tüm grafiksel ara yüz tasarımları tkinter kütüphanesi kullanılarak tasarlanmıştır. Makine öğrenmesiyle tahminle ve raporlama yapabilmek amacıyla 15 farklı algoritma kullanmıştır. Verilerin düzenlenme aşamasından sonra elde edilen veriler kullanıcılara belirli bir kaydetme yeri ve adı sunmuştur ki bu aşamada kullanıcılar kaydettikleri veriler üzerinden direk işlem yapabilmesi için uygulamanın açılımı sağlanmıştır. Tahminleme açıldıktan sonra kolon isimleri otomatik olarak sistem tarafından doldurulmakta ve kullanıcıya yardımcı olmaktadır. Bu sayede kullanıcı hangi kolona değer girmesi gerektiğini

bilmektedir. Ayrıca en iyi tahminleme yöntemini kullanıcı bilemeyeceğinden dolayı burada kullanıcıya yardımcı olmak amacıyla hangi algoritmanın kullanılacağı sistem tarafından otomatik olarak kullanıcıya iletilmiştir. Verilerin sistem tarafından okunması dataframe oluşturmak amacıyla pandas kütüphanesi, hesaplama işlemlerin yapılması için numpy, makine öğrenmesi algoritmalarını kullanabilmek amacıyla ise scikit-learn kütüphanesi kullanılmıştır. Şekil 15'de örnek bir makine öğrenmesi algoritması kullanmak için yazılan kod parçacığı yer almaktadır.

Şekil 15: Tahminleme Modeli Örnek Kod Parçacığı

```
def KMeans(calculate):
    from sklearn.cluster import KMeans
    global kmeans_r2
    global kmeans_pr
    kmeans_model=KMeans().fit(X_train,y_train)
    kmeans_pred= kmeans_model.predict(X_test)
    kmeans_r2=r2_score(y_test, kmeans_pred)
    print('KMeans Score:',np.sqrt(mean_squared_error(y_test, kmeans_pred)))
    print('KMeans Score:',kmeans_r2)
    if r2_score(y_test, kmeans_pred)<0.99:
        #Model Tuning
        kmeans_params = {
            'n_clusters': np.arange(1,10,1),
            'n_init': np.arange(1,20,1),
            'max_iter':np.arange(100,1000,100)
        }
        kmeans_cv_model = GridSearchCV(kmeans_model,param_grid = kmeans_params,cv=3,n_jobs = -1,verbose = 2).fit(X_train, y_train)
        kmeans_tuned = KMeans(n_clusters=kmeans_cv_model.best_params_['n_clusters']
                              ,n_init=kmeans_cv_model.best_params_['n_init']
                              ,max_iter=kmeans_cv_model.best_params_['max_iter']).fit(X_train,y_train)
        kmeans_tuned_pred = kmeans_tuned.predict(X_test)
        if len(calculate)>1:
            print('KMeans MSE:',np.sqrt(mean_squared_error(y_test, kmeans_tuned_pred)))
            kmeans_r2=r2_score(y_test, kmeans_tuned_pred)
            print('KMeans Score:',kmeans_r2)
            print("-"*34)
        else:
            if r2_score(y_test, kmeans_pred)<r2_score(y_test, kmeans_tuned_pred):
                kmeans_pr=kmeans_tuned.predict(calculate)
                print(kmeans_pr)
            else:
                kmeans_pr=kmeans_model.predict(calculate)
                print(kmeans_pr)
    else:
        kmeans_r2
        kmeans_pr=kmeans_model.predict(calculate)
```

Kaynak: Yazar tarafından derlenmiştir.

Zaman serisi analizi yaparken ARIMA tahminleme modeli kullanılmıştır. ARIMA modeli diğer AR, MA ve ARMA gibi modeller ile ARIMA arasında en başarılı sonucun belirlenebilmesi için hata kareler toplamına bakılmıştır. ARIMA modeli diğerlerine göre daha doğru tahminler üretmesinden dolayı tercih edilmiştir.

Kümele analizi yapabilmek için ise Mixture of Gaussian algoritması kullanılmıştır. Kümeleme analizi Python ile yapılmış olup scikit-learn kütüphanesi kullanılmıştır. Gaussian Mixture algoritması, KMeans algoritmasından daha iyi sonuç verdiğinden dolayı kümeleme analizi kısmında bu algoritma kullanılmıştır.

Python ile raporlama aşamasında grafikler oluşturulurken matplotlib kütüphanesi kullanılmıştır. Bu sayede makine ve öğrenmesi, zaman serisi analizi ve

kümeleme analizinde içinde bulunduğu bir sistem hazırlanmıştır. Bu sistem sayesinde karar vericilerin tahminlemelerden alacağı sonuçlar raporlanmıştır. Örnek, rapor hazırlama kod parçacığı Şekil 16'da ki gibidir.

Şekil 16: Örnek Grafik Hazırlama Kod Parçacığı

```
def plotMovingAverage(series, window, plot_intervals=False, scale=1.96, plot_anomalies=False):
    rolling_mean = series.rolling(window=window).mean()

    plt.figure(figsize=(19,11))
    plt.title("Moving average\n window size = {}".format(window))
    plt.plot(rolling_mean, "g", label="Hareketli Ortalam Trend")

    # Plot confidence intervals for smoothed values
    if plot_intervals:
        mae = mean_absolute_error(series[window:], rolling_mean[window:])
        deviation = np.std(series[window:] - rolling_mean[window:])
        lower_bond = rolling_mean - (mae + scale * deviation)
        upper_bond = rolling_mean + (mae + scale * deviation)
        plt.plot(upper_bond, "r--", label="Üst Band / Alt Band")
        plt.plot(lower_bond, "r--")

        # Having the intervals, find abnormal values
        if plot_anomalies:
            anomalies = pd.DataFrame(index=series.index, columns=series.columns)
            anomalies[series<lower_bond] = series[series<lower_bond]
            anomalies[series>upper_bond] = series[series>upper_bond]
            plt.plot(anomalies, "ro", markersize=10)

    plt.plot(series[window:], label="Gerçek Değerler")
    plt.legend(loc="upper left")
    plt.grid(True)

plotMovingAverage(ads, 50)
plt.xlabel(df.columns[0])
plt.ylabel(df.columns[1])
plt.savefig("../UserLogin/Raporlar/Time/myplot3.png")
plt.close()
```

Kaynak: Yazar tarafından derlenmiştir.

3.1.1.5. Test

Bilgi sisteminin ömrü boyunca devam eden, kullanıcıya sunulmadan önce sistemin değerlendirildiği aşamadır. Veriler öncelikle örnek veri üzerinde gerçekleşir ve ardından gerçek veriler kullanılır. Oluşan problemler kullanıcıya sunulmadan önce tespit edilmesi maliyeti düşürmektedir (Tecim, t.y.).

Uygulama geliştirildikten sonra, test etmek için Kaggle üzerinden alınan birkaç test verisi üzerinde kullanmıştır. Bu verilerde boş olan değerler ya da yanlış değerler çok fazla bulunduğundan bu verilerdeki eksiklikler programın veri düzenleme kısmının geliştirilmesinde oldukça yardımcı olmuştur. Bu aşamada ihtiyaçlar belirlenmiş ve uygulamaya eklenmiştir. Ayrıca veri düzenleme ekranında

gereksiz yerler kaldırılmış çok fazla soruna ve anlam karmaşasına sebebiyet vermemesi için işlemler sade ve basit bir şekilde hazırlanmıştır.

Test aşamasında makine öğrenmesi tahminleme algoritmalarını kullanırken farklı veri setleri üzerinde denendiğinde meydana gelen hatalar değerlendirilmiş ve oluşan hatalar giderilmiştir. Aynı zamanda raporlama aşamasında elde edilen raporların, eksik veya hatalı gelmemesi adına, veri düzenleme kısmından elde edilen çıktıdan yapılacağından dolayı bu kısımda kullanıcıların programa uyması gerekmekte ve kolon isimlerini sistem otomatik olarak çekmektedir. Bu kısımda çok fazla hata meydana gelmesinden dolayı uygulamanın en çok zaman alan kısımlarından biridir. Tüm bu aşamaların ardından elde edilen sonuçlar derlendiğinde meydana gelen sorunların giderildiği tespit edilmiştir.

3.1.1.6. Bakım

Uygulama adımları ve testlerinin ardından bu uygulama ikiyüz binlik büyük bir veri seti üzerinde denenmiş ve test edilmiştir. Sistem üzerinden elde edilen raporlar, tahminlemeler takip edilmiş ve oluşabilecek sorunlarla ilgili denetim gerçekleşmiştir. Tüm bu aşamaların sonucunda elde edilen bulgularda sistemin sorunsuz çalıştığı ve ihtiyaçlara karşılık verebileceği tespit edilmiş ve kullanıma hazır hale getirilmiştir.

3.2. UYGULAMA ARAYÜZÜ

Geliştirilen uygulama veri düzenleme, tahminleme aşaması ve yöneticinin karar vermesine yardımcı raporlar olmak üzere 3 bölümden oluşmaktadır. Ara yüz tasarlama kısmında; kullanıcı ihtiyacı gözetilmiş, sade ve anlaşılır bir yapı kullanılmış, kullanıcı dostu bir uygulama ortaya çıkarılmaya çalışılmıştır.

Hazırlan bu uygulama küçük ve orta ölçekli firmaların ihtiyaçları gözetilerek hazırlanmış olup birçok uygulama alanına uygun şekilde tasarlanmıştır. Uygulama yapısı itibariyle karar verme konumunda olan kişiler için veri düzenleme, raporlama, tahminleme, gibi kategorilere ayrılmıştır. Bu kategoriler kullanıcıların kolay bir şekilde istediği yere ulaşımını sağlayabilmek için anlaşılır bir şekilde tasarlanmıştır.

Oluşturulan bu sistem üç temel amaç üzerine hazırlanmıştır. Bunlar sırasıyla;

1.Küçük ve orta ölçekli firmalar için hazırlanmış olan karar destek sisteminde verilerin temizlenmesi aşamasında kolaylık sağlayabilmek,

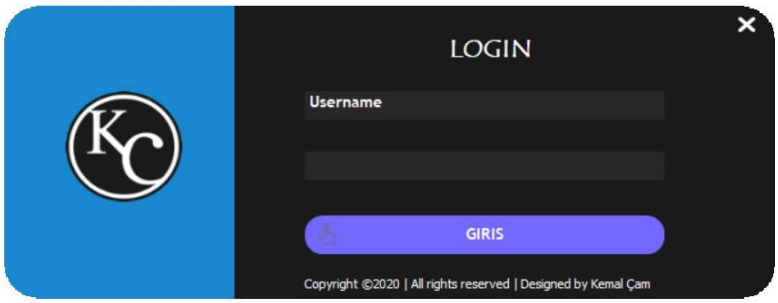
2. Temizlenen verilerden, makine öğrenmesi algoritmalarını kullanarak tahminleme işlemlerinin yapılması, bu tahminleme işlemleri sayesinde yöneticiye karar almasında yardımcı olmayı sağlayabilmek,

3. Zaman serisi, kümeleme analizi gibi birçok dinamiği içerisinde bulunduran geçmiş verilerden geleceğe yönelik çıkarımlar yapabilen, bu verileri karar verme konumunda olan kişilere raporlayabilen bütünleşik bir sistem sağlayabilmek amaçlanmıştır.

3.2.1. Kullanıcı Giriş Ekranı

Sisteme her kullanıcı için ayrı ayrı kullanıcı adı ve şifre tanımlanır. Sistemde eğer ki kullanıcının adı ve şifresi veri tabanında kayıtlı ise ilk olarak bilgiler kontrol edilir ve doğru olması durumunda, sistem giriş ekranına yönlendirir. Kullanıcıların giriş yapmış olduğu giriş ekranı görüntüsü Şekil 17'de ki gibi gösterilmiştir.

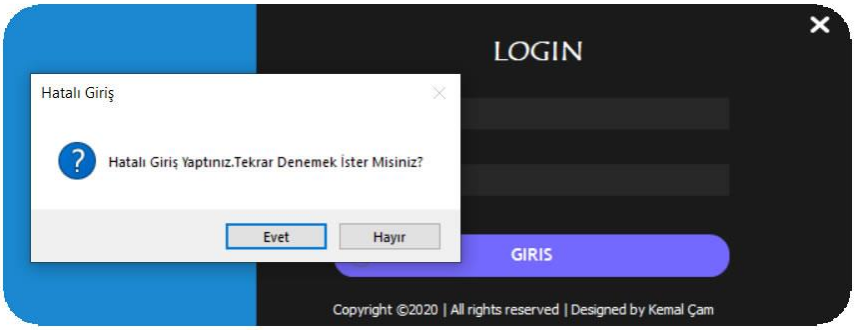
Şekil 17: Kullanıcı Giriş Ekranı



Kaynak: Yazar tarafından derlenmiştir.

Eğer ki, kullanıcı, yanlış bir kullanıcı adı ve şifre kullanmış ise sistem kullanıcının sisteme girmesine izin vermemektedir. Sistem tarafından kullanıcıya gönderilen hata mesajı Şekil 18'de ki gibidir.

Şekil 18: Kullanıcı Giriş Ekranı Hata Mesajı



Kaynak: Yazar tarafından derlenmiştir.

3.2.2. Veri Düzenleme Ara Yüzü

Hazırlanan veri düzenleme ara yüzü Şekil 19'de göründüğü gibi 3 bölümden oluşmaktadır. Bu bölümler; veri ön işleme, veri eğit ve veri görüntülemedir. Verilerin tahminlenmesi aşamasında ilk kullanacak bölüm bu kısımdadır. Verilerin tahminlenmesi için içerideki metinsel ifadelerden arındırılması, kolonların sıralanması, boş kolonların ortadan kaldırılması gibi birçok adımdan meydana gelen tahminlemenin temelini oluşturan en önemli kısımdır.

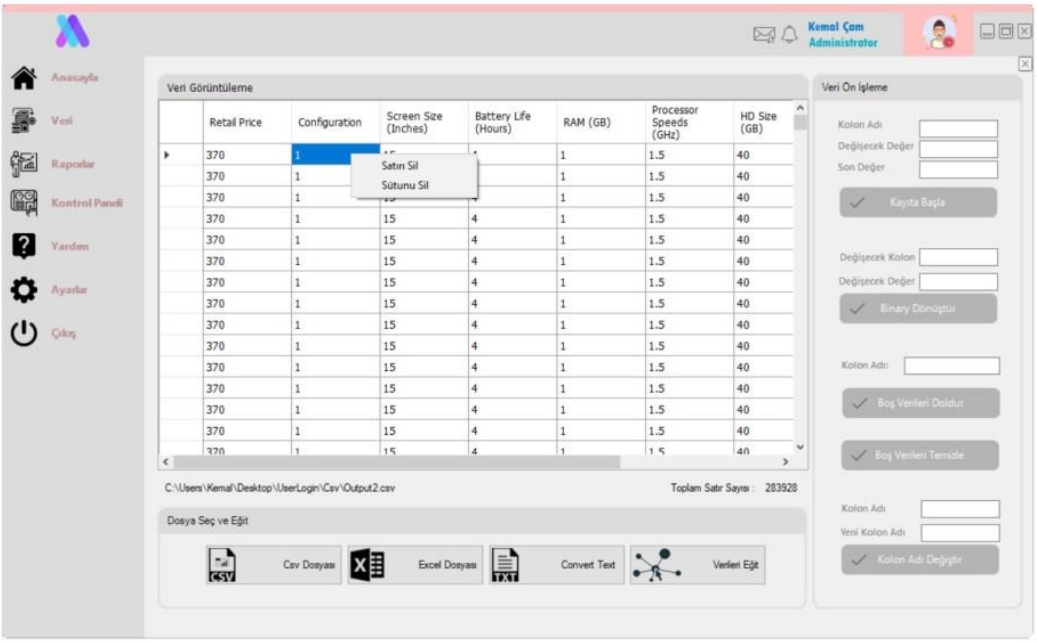
Şekil 19:KDS Ara yüzü

Kaynak: Yazar tarafından derlenmiştir.

Bu kısımda elde bulunan veriler dosya seçme ve eğitme butonuna tıklanarak, kullanılmak istenilen csv veya excel dosyası seçilir ve bu dosya içerisindeki veriler veri görüntüleme kısmında listelenir. Her verinin üst kısmında listelenen kolonun ismi yazmaktadır. Bu kolonlardan gereksinim duyulmayanlar içerisinden Şekil 20'de gösterildiği üzere sağ tık yapılarak istenilmeyen kolonlar ya da satırlar sistemden atılmaktadır. Bu sayede sadece ihtiyaç duyulan, gerekli olan kolonlar işleme koyulmaktadır.

Bunun yanı sıra, yine veri görüntüleme ekranında bulunan kolonlar çift tıklanarak istenildiği kolonu bir sola kolaylıkla taşıyabilmektedir. Bu sayede kolonların yerleşimleri de kolaylıkla ayarlanabilmektedir.

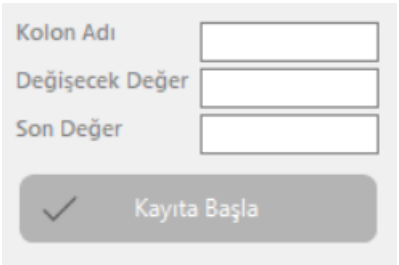
Şekil 20: Kolon Çıkarma İşlemi



Kaynak: Yazar tarafından derlenmiştir.

Bununla birlikte veri ön işleme bölümünün ilk kısımda yer alan kolonlar içerisinde bulunan metinsel ifadelerin değiştirilmesi aşamasında Şekil 21'de gösterildiği üzere ilk olarak ilgili kolonun ismi sisteme girilir. Girilen kolon isminde değişmesi istenen değer yazılır. Ardından ilgili değer hangi değer ile değişmesi isteniyorsa o değer, son değer kısmına girilir. Kayıta başla butonuna tıklandıktan sonra sistem bütün satılar içerisinde bu ifadeyi aramaya başlar ve tüm satırlardaki değerleri değiştirir. Eğer ki sütunda girilen değere karşılık gelen bir ifade yoksa sistem otomatik hata mesajı verir.

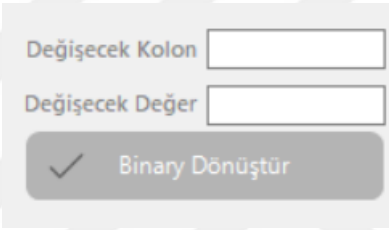
Şekil 21: Metinsel İfade Değişimi



Kaynak: Yazar tarafından derlenmiştir.

Yine veri ön işleme aşamalarından biri olan binary dönüştür kısmında ise veriler bir ve sıfıra dönüştürülür. Bu kısımda amaç cinsiyet gibi ifadeleri tanımlayabilmek için bu ifadeleri bir ve sıfıra dönüştürmektedir. Yine bu aşamada değişecek kolon ismi yazılır ve değişmesi istenen değer yazılır. Değişmesi istenen değer 1 geriye kalan değerler ise 0 olarak tanımlanmıştır. Binary dönüştür butonuna tıklandığında ise otomatik olarak tüm değerler değişmekte ve yine sistemde olmayan bir değer olduğunda otomatik olarak sistemden atılmaktadır. Binary dönüştür aşaması Şekil 22'de gösterildiği gibidir.

Şekil 22: Binary Dönüştür



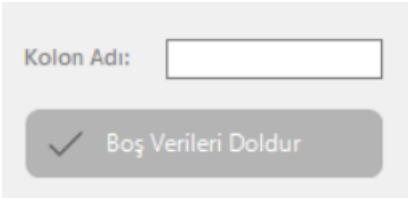
Kaynak: Yazar tarafından derlenmiştir.

Sistem tarafına gelen excel ve csv dosyasındaki değerler her zaman satırları dolu tam anlamıyla düzgün verilerden meydana gelmez. Genellikle unutulan değerler ya da yanlış yazılan değerler çok fazla bulunmaktadır. Oluşabilecek tüm bu hataların, diğer aşamalarda sıkıntılar meydana getirmesini önlemek için bu değerlerin, excel ya da csv dosyasından atılması gerekmektedir. Uygulamada, veri ön işleme aşamasında boş verileri temizle butonuna basıldıktan sonra sistem, tüm satırlar arasında dolaşır ve boş değerleri ve daha önce sisteme tanıtılmış olmaması istenen değerleri gelen csv dosyasının içinden temizler.

Satırlarda meydana gelen boş değerler bazen bir hayli fazla olduğunda ya da diğer sütunlarda bir sıkıntı yok iken tek bir kolondaki değerlerin boş olmasından dolayı bunları sistemden atmak tahminleme aşamasında sıkıntılar meydana getirebilmektedir. Bunların önlenmesi adına Şekil 23'de gösterilen boş verileri doldur kısmında kolon adı girilerek bu kolondaki tüm değerler taranır. Bu değerlerin mod'u alınır ve boş satırlara bu değerler gönderilir. Sistem tasarımı yapılırken eldeki

verilere yanlış değerler girilmesi, ortalama alındığında sıkıntı yaratacağından dolayı en sık girilen değerin (mode) sistem tarafından hesaplanıp satırlara doldurulmasına karar verilmiştir

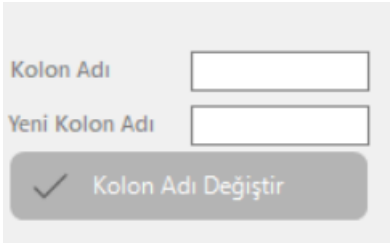
Şekil 23: Boş Veri Doldur



Kaynak: Yazar tarafından derlenmiştir.

Son olarak veri görüntüleme kısmında, görüntülen kolon adları bazen çok kısa ya da gereksiz uzun olabilmektedir. Bu kolon adlarının anlam karmaşasına neden olmaması adına sistem tasarımı yapılırken kolon adlarının yeniden yazılabilmesi için Şekil 24'de gösterildiği üzere bir alan açılmıştır. Değiştirilmek istenen kolonun adı girilerek kolon adı değiştir butonuna tıklayarak sistem üzerinde bulunan kolonların isim değiştirme işlemleri yapılır.

Şekil 24: Kolon Adı Değiştirme



Kaynak: Yazar tarafından derlenmiştir.

Tüm bu aşamaların ardından oluşturulan yeni dosya eğitilmek için hazır hale getirilmiş demektir. Bu aşamadan sonra dosya seç ve eğit sekmesinde bulun verilerimi eğit butonuna tıklayarak dosya kaydetme işlemi gerçekleşir. Ön tanımlı bir isim atanır ve klasör yolu belirtilir. Bu kısıma kaydedilen veri dosyaları tahminleme

ara yüzü tarafında otomatik olarak okunur. Bu aşamadan sonra tahminleme ara yüzü açılmaktadır.

3.2.3. Raporlama ve Menü Yapısı

Uygulama ara yüzündeki menü temelde 7 ana başlıktan oluşmaktadır. Bunlar sırasıyla, Ana sayfa, Veri, Rapor, Kontrol Paneli, Yardım, Ayarlar ve Çıkış tır. Oluşturulan başlıklardan raporlama bölümü kendi içerisinde 4 başlık altında toplanmıştır. Oluşturulan menü yapısı ve bu başlıklar Şekil 25'de ki gibidir.

Şekil 25: Menü Yapısı Gösterimi

Kaynak: Yazar tarafından derlenmiştir.

Menü başlıkları ve açıklamaları aşağıdaki gibidir:

- Anasayfa: Bu başlık şirketlere göre düzenlenmesi için bilgi amaçlı olarak hazırlanmıştır. Bu bölüme şirketin logosu bilgileri konulacaktır. Bu kısmın tasarımı yeniden hazırlanacağından dolayı ilk etapta boş bir şekilde görüntülenecektir.
- Veri: Bu kısım verilerin düzenleme işlemlerinin yapıldığı en temel aşamadır. Gelen verinin tüm işlemleri bu kısımda yapılır. Bu aşamada veriler temizlenir

ve kullanıma hazır hale getirilir. Eğer ki veri temiz ise işlem yapılmaz fakat veride eksilikler, hatalar varsa öncelikle bu bölümde o sorunlar giderilir.

- Rapor: Oluşturulan rapor yapısı kendi içerisinde 4 bölümden oluşmaktadır. İlk bölüm olan betimsel(açıklayıcı) raporlar, veriler için hazırlanmış olan kullanıcı etkileşimine izin veren, içerisinde birçok yapıyı aynı anda bulunduran ve istenilen raporun yapılabildiği kısımdır. Bu kısım verinin anlaşılması için hazırlanır. Verinin geçmişi ile yani "ne olduğu" ile ilgilenir ve bu sorunun cevabının arandığı kısımdır. İkinci bölüm zaman serisi analizi, bu kısım verinin bir zaman dilimi içerisinde nasıl hareket ettiğini gösteren kısımdır. Bu kısımda yalnızca verinin belirli bir dönem içerisindeki hareketi değil geleceğe yönelik hareketi de tahminlenir. Bu tahminlemeler zaman serisi raporları altında toplanmaktadır. Üçünü bölüm olan kümeleme analizinde ise amaç verileri benzerliklerine göre gruplama işleminin yapılmasıdır. Bu kısımda verilerin yoğunluk dereceleri gözlemlenmiş olur. Elimizdeki veriyi alt gruplara böldüğümüzde nerede ne kadar bir yoğunluk olduğu gözlemlenebilir. Son bölümde ise tahminleme raporları, verinin geleceği ve eldeki verilerin çalışma doğruluk oranlarını bize göstermektedir. İçerisinde çeşitli istatistiki verileri de barındıran, bunun yanında verilerin makine öğrenmesi algoritmalarını kullanarak ne kadar doğrulukla hesapladığına dair tüm raporları da içinde barındıran bir yapıda oluşturulmuştur.
- Kontrol Paneli: Bu kısımda, betimsel(açıklayıcı) analiz kısmına giden raporların tasarımları yapılır. Raporlar veriye, şirkete göre değişkenlik gösterdiğinden dolayı tüm sektörlere uygulanabilmesi adına hazırlanmış olan bu bölümde, raporlar kullanıcı etkileşimli olarak yapılabilmektedir. Bu bölümde eldeki veriler grafiklere dökülür.
- Yardım: Bu bölüm, oluşturulan programın hangi yapıda ve neler barındırdığını anlatan kısımdır. İnsanların programı hemen anlama ya da bilme şansı yoktur. Bu yüzden uygulamanın bu aşamasında programın kullanımı hakkında bilgiler verilir ve bu bilgiler sayesinde kullanıcılara programı nasıl kullanacaklarını öğrenebilme imkanı tanınır. Bu bölümün

avantajı kullanıcıların takıldıkları yerde argüman olarak kullanabilecekleri bir alan yaratmaktır.

- Ayarlar: Bu bölüm ise programda bulunan raporların oluşturulduğu ve aynı zamanda şirket bilgilerinin tanımlandığı aşamadır. Bu bölümde ana hatlarıyla elde bulunan veriler değişkenlik gösterdiğinde grafikler de değişime uğrayacaktır. Bu değişimi sağlayabilmek ve oluşturulan grafikleri temel program ara yüzünde görüntüleyebilmek adına bu kısım oluşturulmuştur. Bu kısım, temelde betimsel(açıklayıcı) analiz raporları ve zaman serisi, kümeleme ve tahminleme raporlarının içinde bulunduğu tüm raporlar olmak üzere iki ana başlıktan oluşmaktadır. Betimsel(açıklayıcı) raporlamaların mantığı Kontrol paneliyle aynıdır. Fakat tüm raporlar, diğerlerinden ayrılmaktadır. Tüm raporlarda raporlama bölümü altındaki tüm grafikler değiştirilebilir yapıda olduğu için tekrardan eğitim yapılır ve dosyalar oluşturulur. Bu dosyalar da raporlama bölümüne göre belirli yerlere aktarılır. Aynı zamanda, ana sayfada görüntülenecek olan şirket bilgilerinin değiştirildiği kısımdır. Bu kısımda değiştirilen tüm değerler ana sayfada değişime uğrayacak ve bilgiler güncellenecektir.
- Çıkış: Programda tüm işlemlerin sonlandırıldığı bölümdür. Bu bölüm programdan güvenli bir şekilde çıkış izni sağlar.

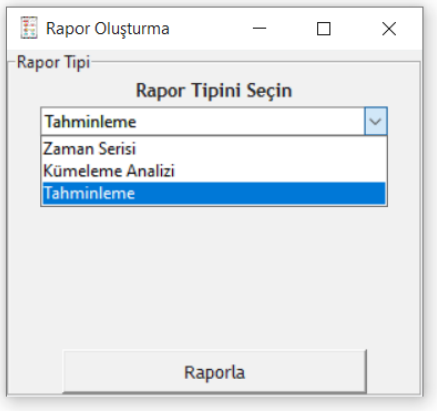
3.2.3.1. Kullanıcı Raporları

Bu kısım temelde 4 ana rapor üzerine kurulmuştur. Bu kısımda Tahminleme, Kümeleme ve Zaman serisi, Betimsel(açıklayıcı) analizlerin raporlar oluşturulmaktadır.

3.2.3.1.1. Tahminleme Raporları

Tahminleme raporlarını oluşturmak için öncelikle ilgili seçim kısmından tahminleme seçeneği seçilir ve ardından sistem üzerinden daha önceden tahminlenmek istenen dosya seçilir ve tahminleme işlemine başlanır. Tahminleme seçim ekranı Şekil 26'da gösterildiği gibidir.

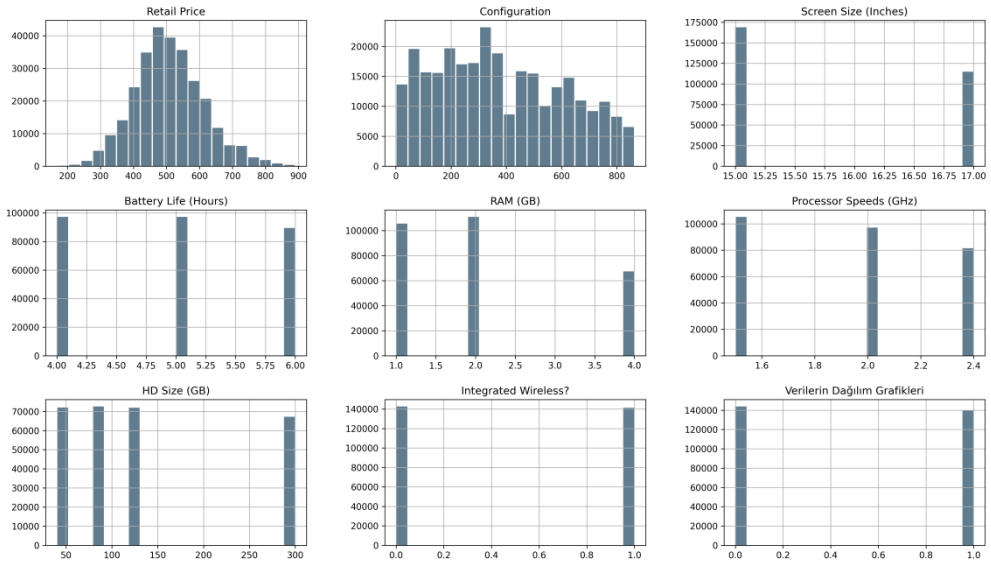
Şekil 26: Tahminleme Seçim Ekranı



Kaynak: Yazar tarafından derlenmiştir.

İlk olarak sistem elimizdeki veriler hakkında bilgi sahibi olmak için bize Şekil 27'de ki grafikte gösterildiği üzere, verilerin dağılımlarını içeren histogram grafiklerini oluşturmaktadır. Bu sayede yönetici ilk olarak eldeki verilerin neler olduğu hakkında bilgi sahibi olmuş olur. Örneğin; aşağıdaki grafikte bulunan Screen Size grafiği üzerindeki verileri inceleyecek olursak verilerin "En fazla bulunduğu bölge 15 bölgesi ve en fazla ekran boyutu 15 inç olan ekranların satışı yapılmıştır." bilgisine ulaşılabilir. Verilerin dağılımları yalnızca bu gibi verileri değil artık verileri de göstermektedir. Eğer ki dağılımların dışında değerler olursa bu değerler yanlışlıkla girilmiş ya da hata yapılmış olabilme ihtimaline karşı daha iyi tahminlemeler yapmak adına verilerden atılır.

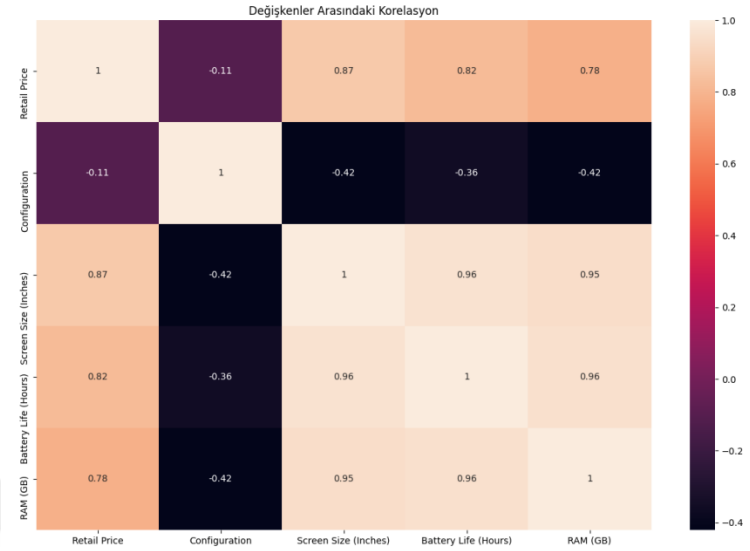
Şekil 27: Verilerin Dağılımlarını Gösteren Örnek Grafiği



Kaynak: Yazar tarafından derlenmiştir.

Veriler arasındaki ilişkiler, tahminleme başarısını etkileyen en önemli faktörlerin başında gelmektedir. Veriler arasındaki korelasyonlar verilerin birbirini açıklama oranlarını ifade eder. Sıfıra yakın olma durumunda birbiriyle bir ilişkisi bulunmadığı bire yakın olma durumunda ise birbiriyle ilişkili olduğu gözlemlenir. Aşağıdaki Şekil 28'de ki grafiğe bakıldığında Retail Price değişkeninin Screen Size değişkeni ile pozitif ilişkili olduğu sonucu çıkarılabilir. Bu yüzden değişkenlerin birbiriyle ilişkili olma durumu ve yönünü belirlemek için bu matris, analizler için oldukça önemlidir.

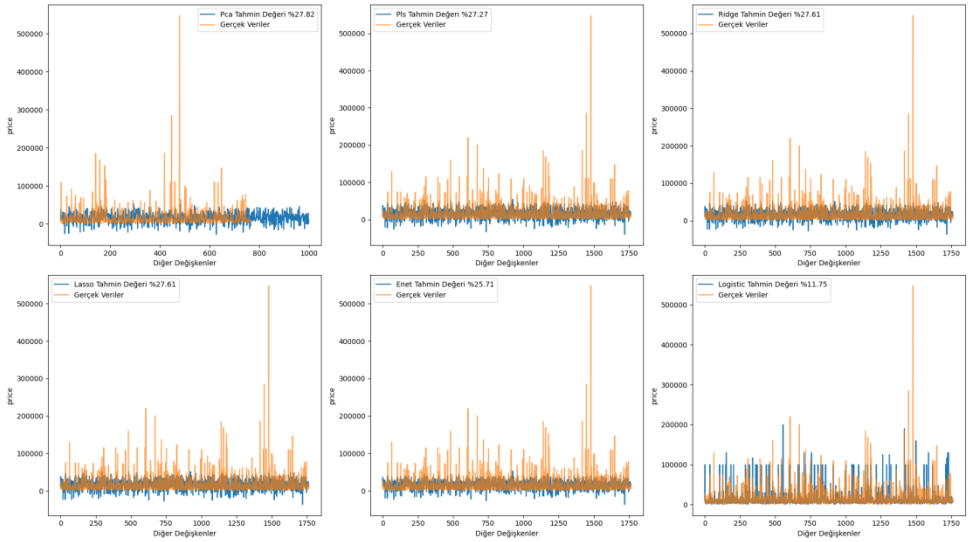
Şekil 28: Korelasyon Grafiği



Kaynak: Yazar tarafından derlenmiştir.

Bir sonraki aşamada uygulamada kullanılan makine öğrenmesi algoritmalarının Şekil 29'da gösterildiği üzere mevcut veriler ile tahminlenen veriler arasında modelin ne kadar doğru sonuçlar verdiği karşılaştırılmaktadır. En doğru sonuçları veren yöntemlerle tahminleme işlemi yapmayı sağlamaktadır.

Şekil 29: Mevcut veriler ile Algoritmaların Karşılaştırılması



Kaynak: Yazar tarafından derlenmiştir.

Yukarıdaki görselde 6 algoritma sonuçları gösterilmiştir. Mavi çizgiler modeldeki tahmin edilen verileri, yeşil çizgiler ise model üzerindeki gerçek verileri göstermektedir. Grafik üzerinde modelin tahminleme oranları da gösterilmiştir. Bu oranlar ilgili algoritmanın verileri yüzde kaç doğrulukla açıkladığını göstermektedir. Tüm bunların sonucunda, elde edilen değerlerin sonunda aşağıdaki grafikte gösterildiği üzere bir sonuç çıkarımı yapılmaktadır. Bu kısımda algoritma için en doğru değeri veren grafik sistem üzerinde gösterilmektedir.

3.2.3.1.2. Kümeleme Raporları

Bu raporda amaç, verileri benzerliklerine göre gruplama işleminin yapılmasıdır. Veriler belirli işlemlerden geçer ve sonunda veriler kendi içerisinde homojen olacak şekilde dağılımları gösterilir. Kümeleme raporunda ilk olarak Şekil 30'da ki grafikte gösterildiği üzere istatistiksel çıkarımlar verilir. Burada verilerin sayısı, ortalaması, standart sapmaları, minimum, maximum değerleri gibi birçok ayrıntı yer almaktadır. Bu kısımda veri hakkında bilgiler bulunmaktadır.

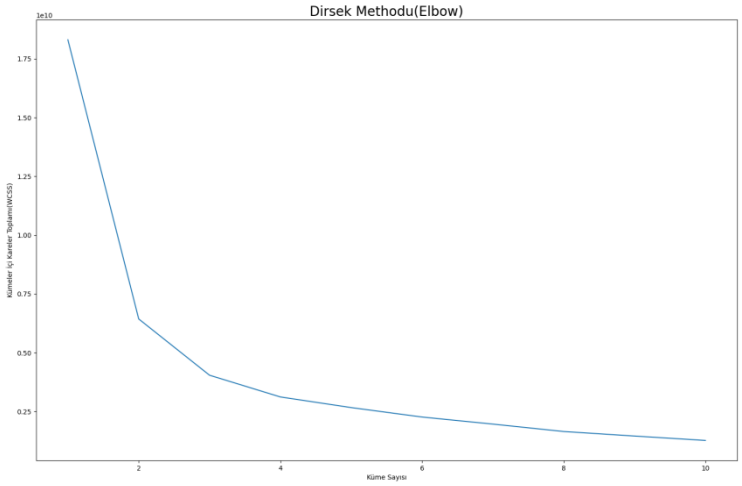
Şekil 30: Kümeleme Raporu İstatistiksel Çıkarım Ekranı

	Retail Price	Configuration	Screen Size (Inches)	Battery Life (Hours)	RAM (GB)	Processor Speed	Integrated Wireless	HD Size (GB)	Bundled Applications?
count	283786.0	283786.0	283786.0	283786.0	283786.0	283786.0	283786.0	283786.0	283786.0
mean	508.12	379.62	15.81	4.97	2.1	1.93	0.5	132.09	0.49
std	104.61	231.44	0.98	0.81	1.14	0.37	0.5	97.8	0.5
min	168.0	1.0	15.0	4.0	1.0	1.5	0.0	40.0	0.0
25%	440.0	193.0	15.0	4.0	1.0	1.5	0.0	40.0	0.0
50%	500.0	347.0	15.0	5.0	2.0	2.0	0.0	80.0	0.0
75%	575.0	576.0	17.0	6.0	2.0	2.4	1.0	120.0	1.0
max	890.0	864.0	17.0	6.0	4.0	2.4	1.0	300.0	1.0

Kaynak: Yazar tarafından derlenmiştir.

İkinci kısımda dirsek metodu adını verdiğimiz, küme sayısını belirlemek için kullanılan yöntem kullanılır. Dirsek metodunun örnek gösterimi Şekil 31'de gösterildiği gibidir. Dirsek metodu, noktaları küme merkezine uygun şekilde atar.

Şekil 31: Dirsek Metodu Grafiği

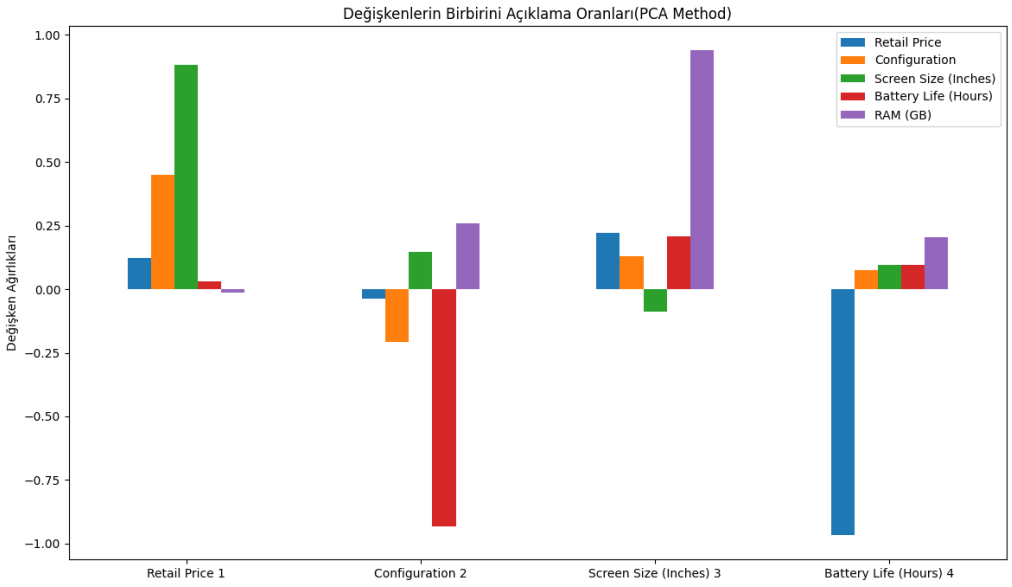


Kaynak: Yazar tarafından derlenmiştir.

Küme sayısı belirleme işleminden sonra Pca algoritması ile indirgeme işlemi yapılır. İndirgeme işlemi verilerin kullanılabilirliğini arttırmak ve daha iyi sonuçlar almak için önemlidir. İndirgeme işlemi yapıldıktan sonra verilerin birbiriyle ilişkileri kontrol edilir. Kontrol işlemi sonrasında grafiğe dökülür ve veriler daha iyi bir şekilde bu Şekil 32'de ki grafik üzerinden gözlemlenebilmektedir. Bu boyut

indirgemede kullanılan bileşenlerin sayısı dirsek metodundan gelmektedir. Bu aşama kümelemenin gerçekleşmesi için gereken son aşamadır.

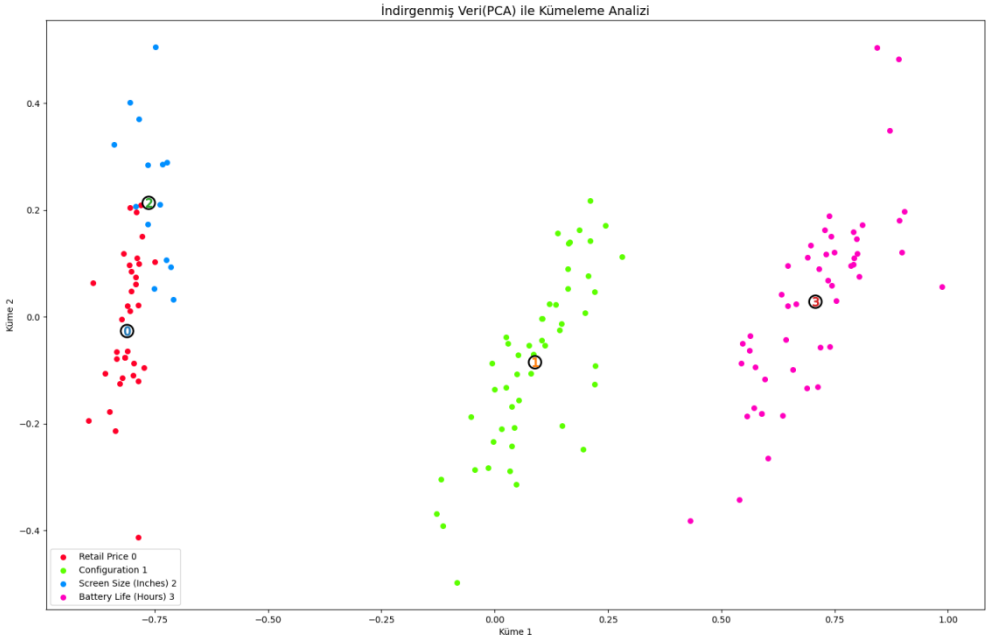
Şekil 32: PCA ile Boyut İndirgeme Grafiği



Kaynak: Yazar tarafından derlenmiştir.

Son aşamada ise, bu işlemler daha önceki aşamalarda belirlenen yöntemlerin sonucunda Gaussian Mixture algoritması kullanılarak kümeleme işlemi yapılmaktadır. Kümeleme işlemi Şekil 33'de gösterildiği üzere renkler ile ilgili kümeler belirtilmiş ve veriler kümelenmiş bir şekilde grafiğe dökülmüştür.

Şekil 33: Kümeleme Analizi Grafiği

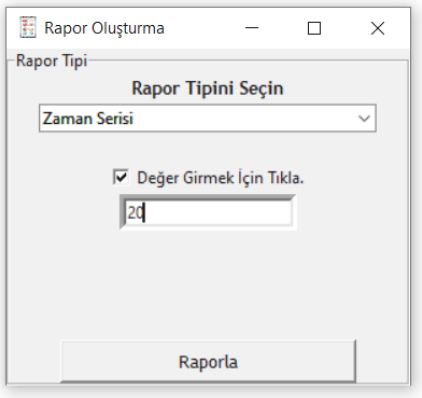


Kaynak: Yazar tarafından derlenmiştir.

3.2.3.1.3. Zaman Serisi Raporları

Zaman serisi analizleri, verilerin zaman içerisinde almış olduğu değerlerden ileride alabileceği değerlerin tahminlenmesi işlemidir (Erturan,2017). Zaman serisi analizlerinde amaç geleceği tahminleyip buna yönelik kararlar alabilmektir. Oluşturulan sistem üzerinde, kişi ya da şirketler ellerinde bulunan verileri sistem üzerinden seçmektedir. Seçtikleri verilerin ne kadar sürelik zaman dilimini tahminlemesini istediğini sisteme girerek tahminleme işlemi yapılmaktadır. Tahminleme işleminin sonucunda elde edilen tüm raporlar sistem tarafına gönderilmekte ve yönetici değerlendirmesi için seçenek sunmaktadır. Tahminleme yapılırken ARIMA algoritması kullanılmıştır. Bu algoritma kullanılmadan önce eldeki veriler Dicky Fuller testi yapılarak durağanlaştırılmıştır. Zaman serisi analizinin yapıldığı ekran görüntüsü Şekil 34'de gösterildiği gibidir.

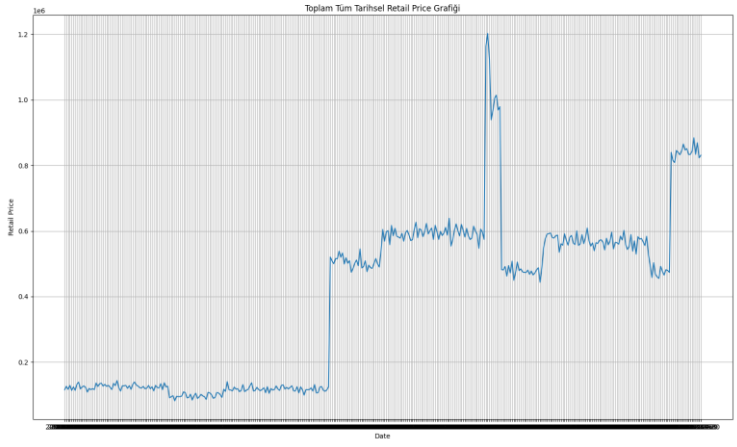
Şekil 34: Zaman Serisi Analizi Ekranı



Kaynak: Yazar tarafından derlenmiştir.

Zaman Serisi raporla ekranında ilk olarak verinin belirli bir zaman düzlemindeki hareketi verilir. Böylece ilk olarak verinin zaman içerisinde nasıl hareket ettiği gözlemlenmiş olur. Verinin zaman içerisindeki dalgalanmaları Şekil 35'de gösterildiği gibidir.

Şekil 35: Zaman Serisi Grafiği

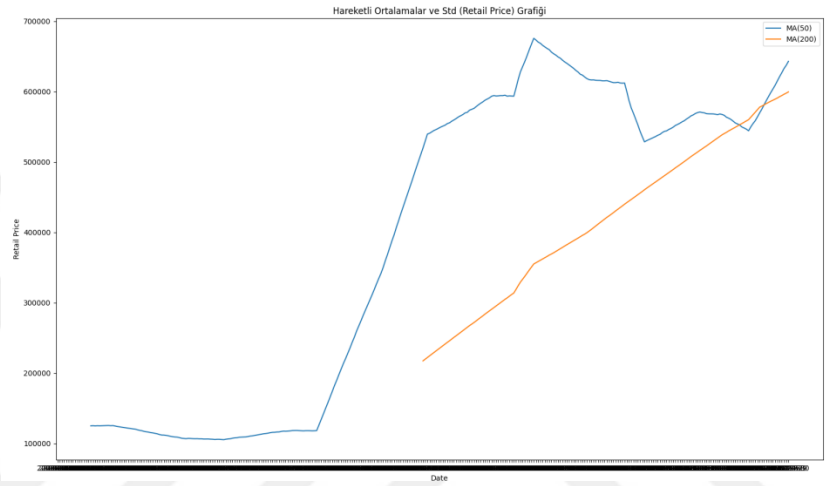


Kaynak: Yazar tarafından derlenmiştir.

Hareketli ortalamalar, birden fazla fiyatın ortalamasının alındığı ve buna göre hesaplanan grafikteki en önemli göstergelerdendir. Fiyatların hareketi hakkında bilgi sağlar. Hareketli ortalamalarda en çok kullanılan 50 günlük ve 200 günlük

ortalamalardır. Bu çalışmada da verilerin açıklanmasında kullanılmıştır. Bu ortalamaların kesişmesi bize fiyatın aşağı ya da yukarı yönlü hareketini göstermektedir. Eğer 50 günlük ortalama 200 günlük ortalamanın üstünde ise fiyatlar yukarı yönlü, aksi durumda ise aşağı yönlü hareket eder. Hareketli ortalamaların gösterildiği grafikler Şekil 36'da gösterildiği gibidir.

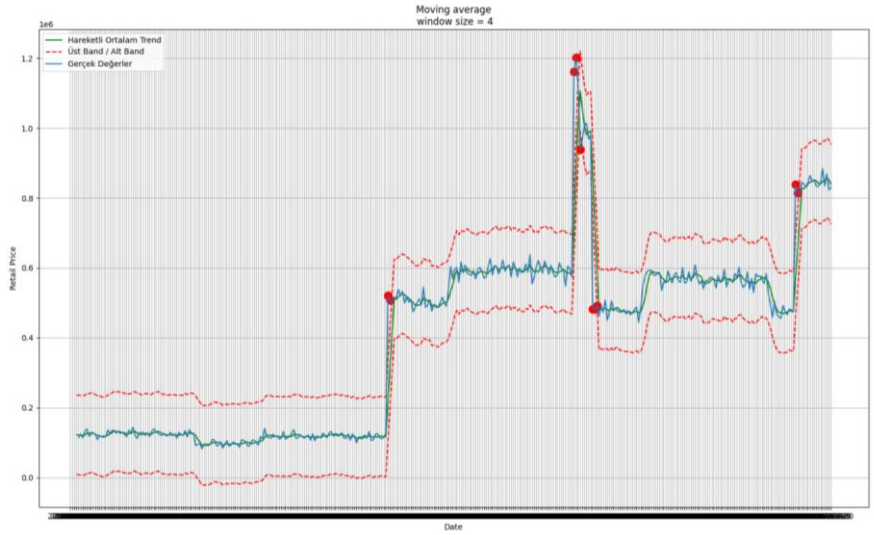
Şekil 36: Hareketli Ortalamalar Grafiği



Kaynak: Yazar tarafından derlenmiştir.

Alt ve üst band kullanımı (Bollinger Bantları), bu bantların sıkıştığı bölgeden sonra yapılacak olan bir sonraki hareketin yönü, fiyatın yönünü aşağı ya da yukarı yönlü kırmasını ifade eder. Bollinger bantları, bazı durumlarda da tepenin teyit edilmesinde kullanılır. Bollinger bantlarında eğer ki fiyat üst bandı kırarsa fiyatlarda artış meydana gelir. Fakat yönü aşağı olursa fiyatların düşüşe geçtiği gözlemlenir. Bollinger bantları, gösterildiği grafikler Şekil 37'de gösterildiği gibidir.

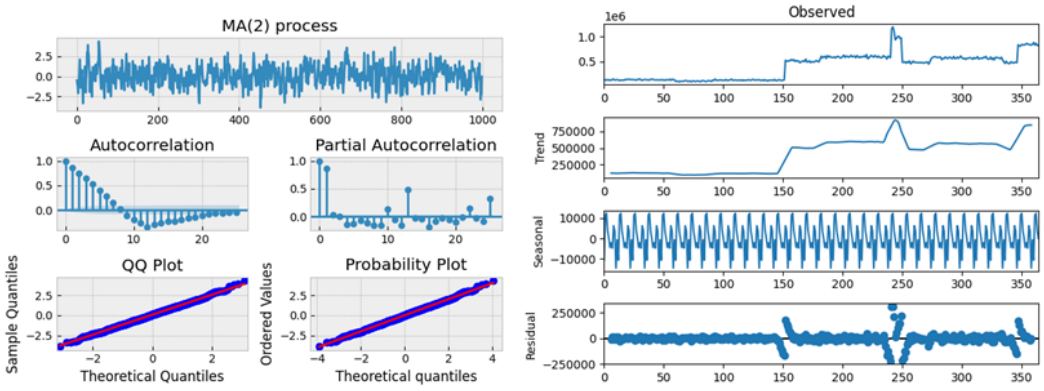
Şekil 37: Bollinger bantları (Alt ve Üst Band) Grafiği



Kaynak: Yazar tarafından derlenmiştir.

Gerçekleşen, trend ve mevsimsel verilerin gösterildiği grafiklerin tamamını, hareketli ortalama sürecini ve bu süreçte hesaplanan oto korelasyonları gösteren oto korelasyonlar ve verilerin dağılımları grafiği Şekil 38'de gösterildiği gibidir.

Şekil 38: Oto korelasyonlar ve Verilerin Dağılımları

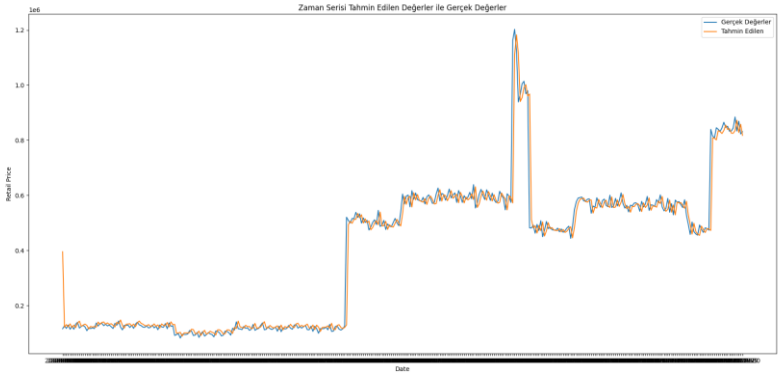


Kaynak: Yazar tarafından derlenmiştir.

Elimizdeki verilerin ne kadar doğru tahminlenmiş olduğu çok önemlidir. Bu yüzden verinin ilk olarak ne kadar doğrulukla tahmin edildiğini grafik üzerinde gösterimi Şekil 39'da ki gibidir. Bu kısımda, veri ARIMA algoritması kullanılarak

tahminlenmiş ve grafik ortamına iki veri karşılaştırılmak üzere dökülmüştür. Elde edilen grafikten de anlaşılmaktadır ki bu algoritma eldeki verileri açıklama da oldukça yeterlidir.

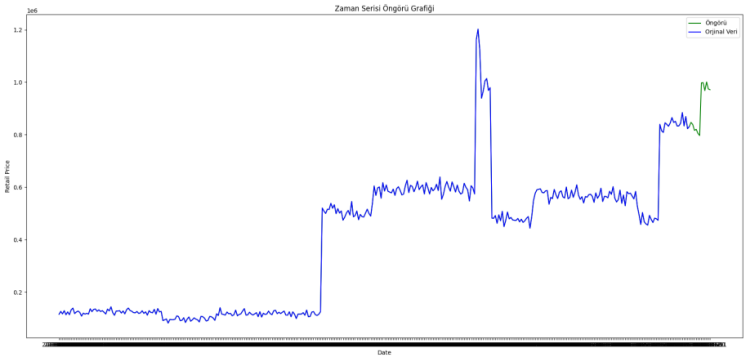
Şekil 39: Gerçekleşen ve Tahminlenen Zaman Serisi



Kaynak: Yazar tarafından derlenmiştir.

Tüm bu aşamalardan sonra veriler tahminleme için ARIMA algoritmasıyla tahminlenir ve grafiğe dökülür. Elde edilen öngörü grafiği Şekil 40'da ki gibidir. Bu grafikte önceki veriler mavi ile yeni veriler ise turuncu ile gösterilmiştir. Elde edilen grafik 20 aylık veriler için tahminlenmiştir. Bu grafik, işletmelerin bir sonraki adımı öngörebilmesi ve gelecek hakkında bilgi edinebilmesi için önemli bir kaynaktır.

Şekil 40: Zaman Serisi Öngörü Grafiği

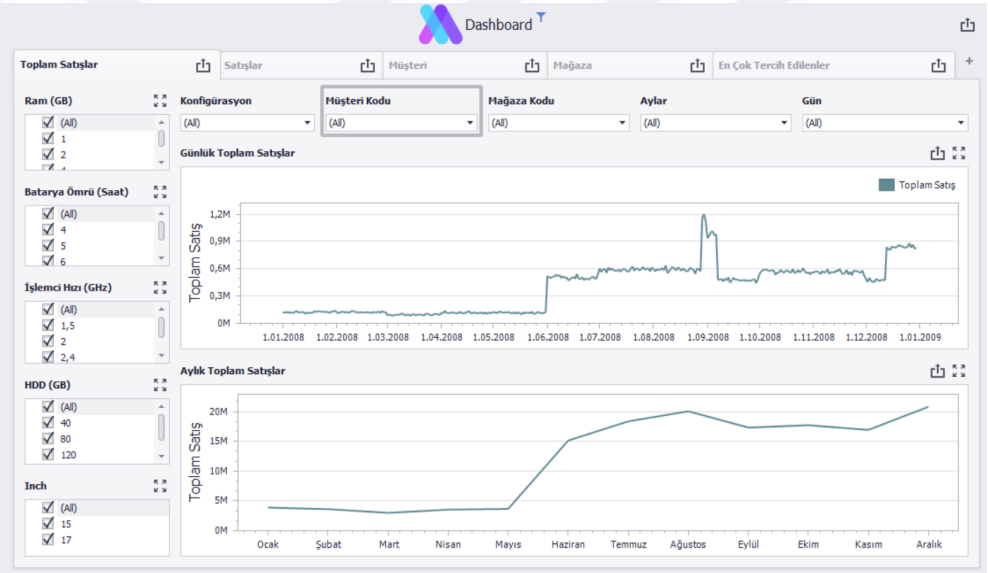


Kaynak: Yazar tarafından derlenmiştir.

3.2.3.1.4. Betimsel (Açıklayıcı) Raporlar

Bu raporlar veriler hakkında çok fazla bilgi sağlamaktadır. Bu kısımda veriler kullanıcı etkileşimine açık bir şekilde, kullanıcıların istedikleri her şeyi sağlayabilecekleri kısımdır. Eldeki veriler bu kısımda Şekil 41'de gösterildiği üzere 5 başlık altında toplanmıştır. Bu başlıklar; toplam satışlar, satışlar, müşteri, mağaza, en çok tercih edilenlerdir. Bu bölüm kişiye veya şirketin eldeki verisine göre düzenlenebilir yapıdadır. Bu yüzden veri ne olursa olsun istenilen rapor istenilen şekilde yapılabilir.

Şekil 41: Betimsel(Açıklayıcı) Raporlama Ekranı



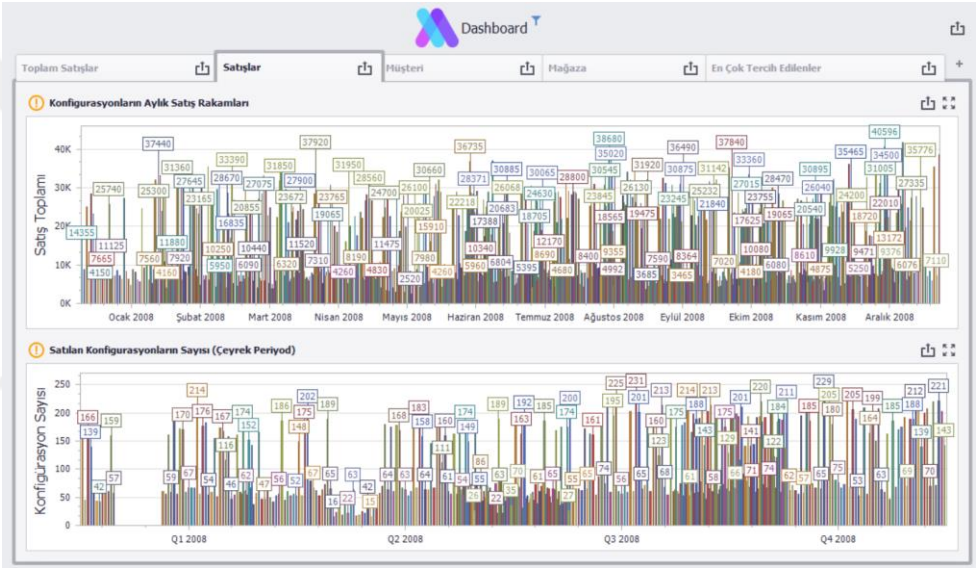
Kaynak: Yazar tarafından derlenmiştir.

Betimsel(açıklayıcı) raporların ilk kısmı olan satışlarda, kullanıcı etkileşimi sağlayacak ve veriye göre filtreleme yapılabilecek eldeki verinin tüm detayları bulunmaktadır. Örneğin; bütün müşteriler yerine yalnızca bir müşteri için bir seçim yapılacaksa müşteri kodu kısmından ilgili müşteri seçilir ve filtreleme işlemi yapılır. Bağlı olan tüm grafiklerde buna göre değişir. Bu alanda veriler bir bilgisayar firmasına ait olduğu için, ekranda bulunan filtreler ram, batarya ömrü, işlemci, hdd, müşteri kodu, mağaza kodu, gün, ay vb. birçok filtreyi içerisinde bulundurmaktadır.

Ayrıca bu kısımda yukarıdaki Şekil 41'de görüldüğü üzere aylık ve günlük bazda toplam satış grafikleri bulunmaktadır.

Bir diğer kısım olan satışlar kısmında, konfigürasyon olarak adlandırılan, yani müşterilerin aldıkları ürünlerin özelliklerine göre tanımlanan numaraların aylık ve çeyrek bazda satış grafikleri görünmektedir. Veriler büyüdükçe görüntülenecek çok fazla değer bulunduğundan grafikler görünmez hal alabilmektedir. Bu yüzden, verilerin filtreleme ekranında filtrelenerek daha doğru bilgiler elde edilmektedir. Satışlar kısmı örnek grafiği Şekil 42'de gösterildiği gibidir.

Şekil 42: Satışlar Başlığı

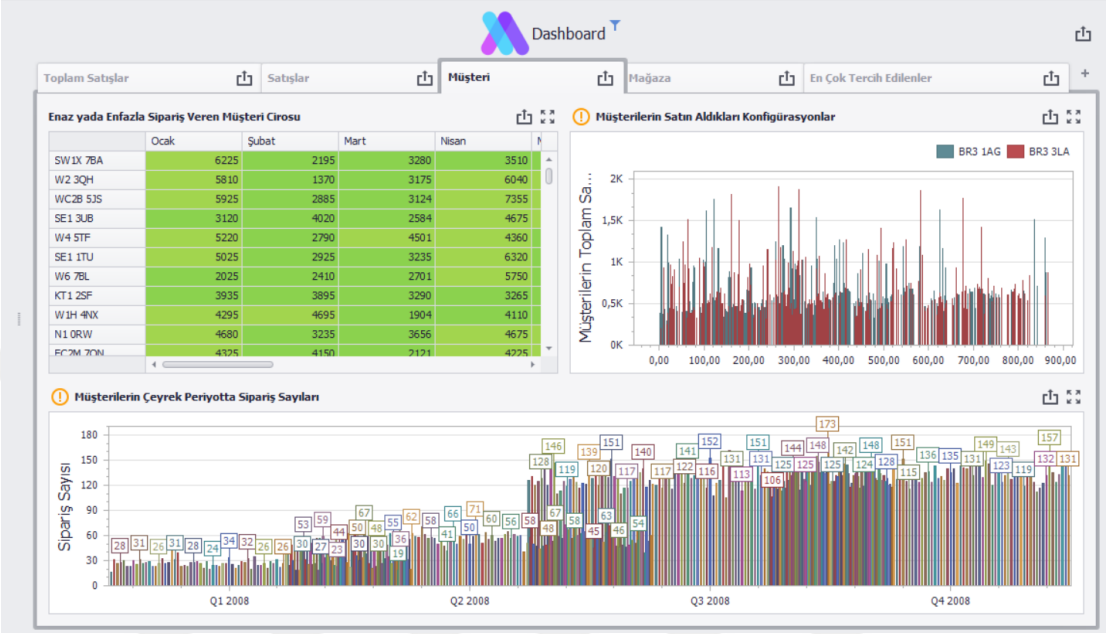


Kaynak: Yazar tarafından derlenmiştir.

Müşterilerin ne satın aldığı ve ne zaman satın aldıkları çok önemlidir. Bu yüzden ilk olarak hangi müşterinin daha fazla eşya satın aldığı ve hangi aylarda satın aldığı belirlenmesi adına bir liste hazırlanmıştır. Bu liste üzerinde bulunan bilgiler sayesinde kişiye özel indirimler, fırsatlar vb. birçok şey yapılarak mağaza satışlarının arttırılması öngörülebilir. Aynı zamanda en çok hangi modellerden satın aldıkları bilgisi de mağazada stok bulundurmak adına önemli bir bilgidir. Bu bilgiler de, bu kısımda histogram grafiği üzerinde gösterilmektedir. Son olarak çeyrek dilimlerde müşterilerin satın alma sayıları gösterilmiş, bu sayede de hangi aylarda daha yoğun olacağı ve hangi ürünleri stok tutmaları gerektiği bilgisi karar vericilerin karar

almasında fayda sağlayacağı düşünülmüştür. Müşteriler kısmı örnek grafiği Şekil 43'de gösterildiği gibidir.

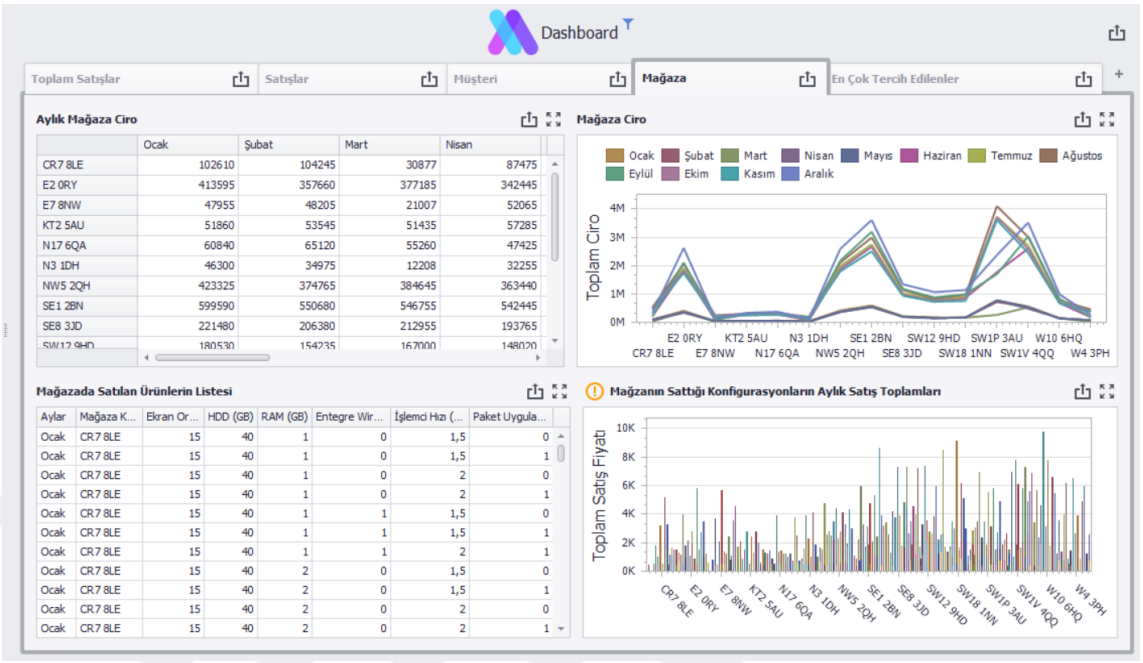
Şekil 43: Müşteriler Başlığı



Kaynak: Yazar tarafından derlenmiştir.

Bir sonraki kısım olan mağaza kısmında, Şekil 44'de gösterildiği üzere mağazanın aylık ciroları gösterilmektedir. Aylık cirolara göre en fazla ya da en az iş yapan mağazalar belirlenmektedir. Bunun yanında da hangi aylarda daha az iş yapıldığı ve satışların azaldığı bilgisi de listelenmektedir. Mağaza ürünlerinden hangi ürünlerin daha çok satıldığı, hangi mağaza da daha çok satıldığı gibi bilgiler de bu bölümde cevap kazanmaktadır.

Şekil 44: Mağaza Başlığı



Kaynak: Yazar tarafından derlenmiştir.

Son olarak en çok tercih edilenler kısmında, mağazada satılan ürünlerin satış sayıları gösterilmektedir. Bu sayılar aylık bazda gösterilmiştir. İçerisinde hdd, Ram, Ekran, İşlemci gibi bütün ürünlerin bulunduğu 4 farklı grafikten oluşmaktadır. Bu bölümde yine diğer bölümler gibi yönetici açısından büyük önem arz etmektedir. Çünkü hangi ürünün daha çok satıldığını bilmek, hangi üründen, hangi ayda, ne kadar stok tutması gerektiğini planlamak yöneticiye büyük kolaylık sağlamaktadır. Tüm bu grafiklerin bulunduğu ekran Şekil 45'de gösterildiği gibidir.

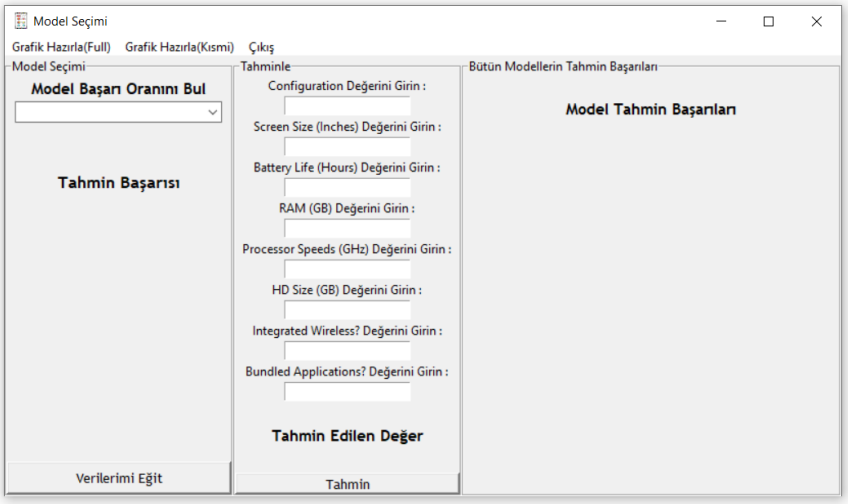
Şekil 45: En çok tercih edilenler başlığı

Kaynak: Yazar tarafından derlenmiştir.

3.2.4. Tahminleme Arayüzü

Veri düzenleme aşamasının ardından kaydedilen ve hazırlanan veriler sistem tarafından okunur, ardından veri tahminleme ara yüzü açılır. Açılan bu ara yüz üzerindeki menüler grafik hazırla(full), grafik hazırla(kısmi) ve çıkış işlemi olarak tanımlanmıştır. Full raporlar, verinin tüm istatistiksel ve yönetimsel raporları içerirken, kısmi raporlar yalnızca yönetimsel raporları içerir. Fakat kısmi raporların, full raporlardan farkı elde edilen grafiklerin daha ayrıntılı gösterilmesidir. Bu sayede karar verme konumundaki kişilere, grafik okuma işlemini daha etkin kullanabilme şansı sağlar. Tahminleme ara yüzü kendi içerisinde 3 bölüme ayrılmaktadır. Model Seçimi, Tahminle ve Bütün Modellerin Başarı oranları, Tahminleme ara yüzü Şekil 46'da gösterildiği gibidir.

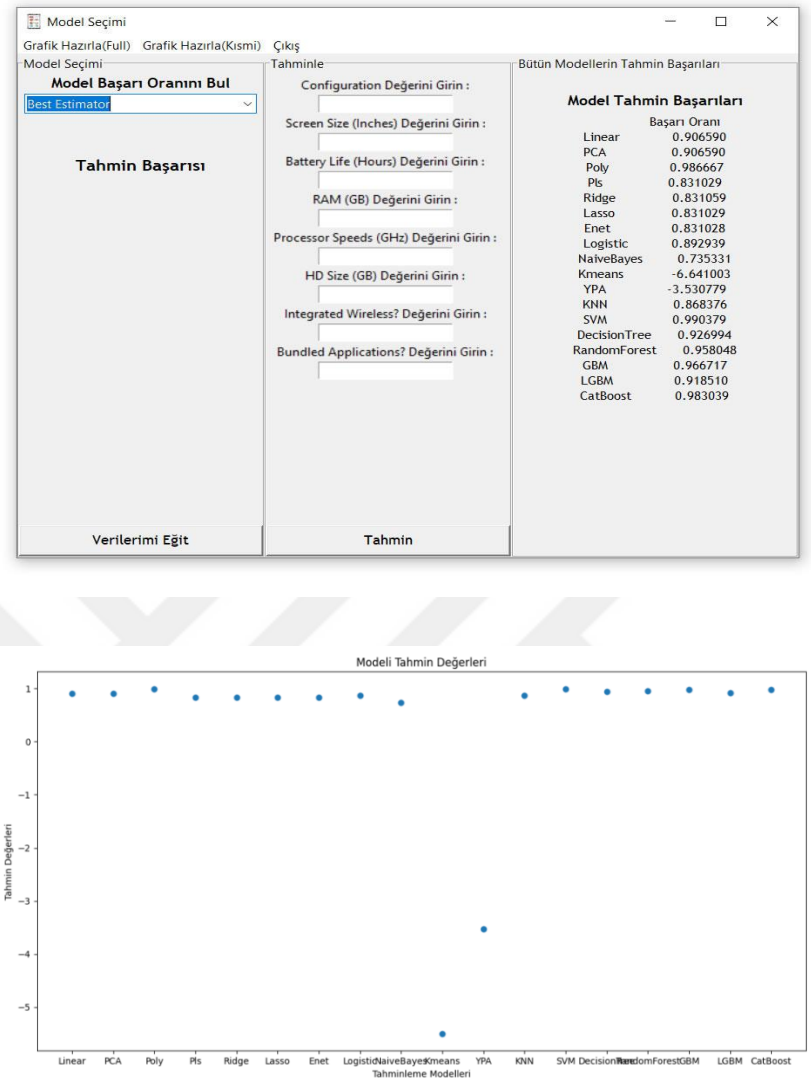
Şekil 46: Tahminleme Ara yüzü



Kaynak: Yazar tarafından derlenmiştir.

Model seçimi bölümünde, kullanıcıların ihtiyacı olan tahminleme modelini bilemeyeceğinden ve tek tek bakmaması adına, en iyi tahminleyen modeli seçmek için seçenekler arasına Best Estimator seçeneği eklenmiştir. Bu sayede modellerin tümü işleme sokulur, aralarından en iyi sonucu veren değer ekranına basılır ve grafiği çizdirilir. Bütün algoritmalar bu bölümde karşılaştırılır. Tahmin Başarı sonuçları Model Tahmin Başarısı bölümünde listelenir. Örnek gösterim Şekil 47'de ki gibidir. Elimizdeki verilerden yola çıkarak ekrana basılan değerler incelendiğinde bize kullanılacak en uygun algoritmanın SVM algoritması olduğunu göstermektedir.

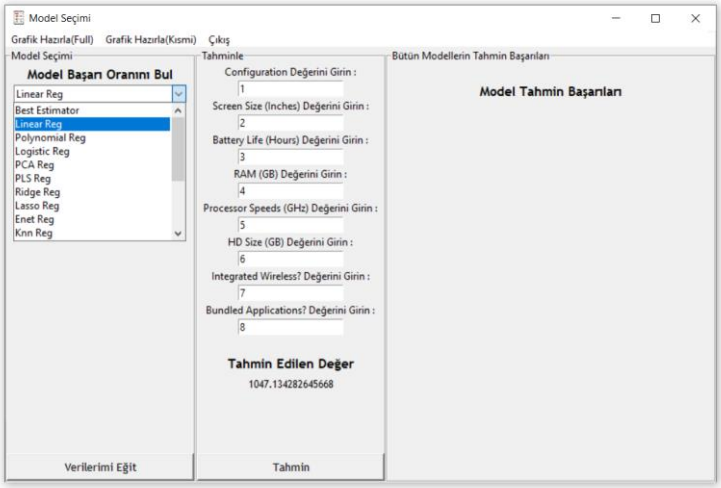
Şekil 47: Model Tahmin Başarısı Örnek Grafiği ve Çıktı Ekranı



Kaynak: Yazar tarafından derlenmiştir.

İlk olarak model seçme bölümünde bir algoritma seçilir. Bu seçim iki yolla olabilir. Kullanıcı istediği algoritma varsa ilk onu kullanabilir ya da ne kullanacağını bilmiyorsa yukarıda anlatıldığı üzere gerçekleşen işlemi yapıp en uygun algoritma sistem tarafından kullanıcıya sunulabilir. Algoritma seçimi yapıldıktan sonra verilerimi eğit butonuna tıklanır, tahmin başarısı altında ne kadar doğrulukla çalıştığını ekrana basar. Bu görüntülenen değer, bu algoritmanın eğitilen verilerin test edilen veriler üzerinde ne kadar doğrulukla tahminleme yapabildiğini göstermektedir. Bu değerler Şekil 48'de ki gibi ekrana basılmaktadır.

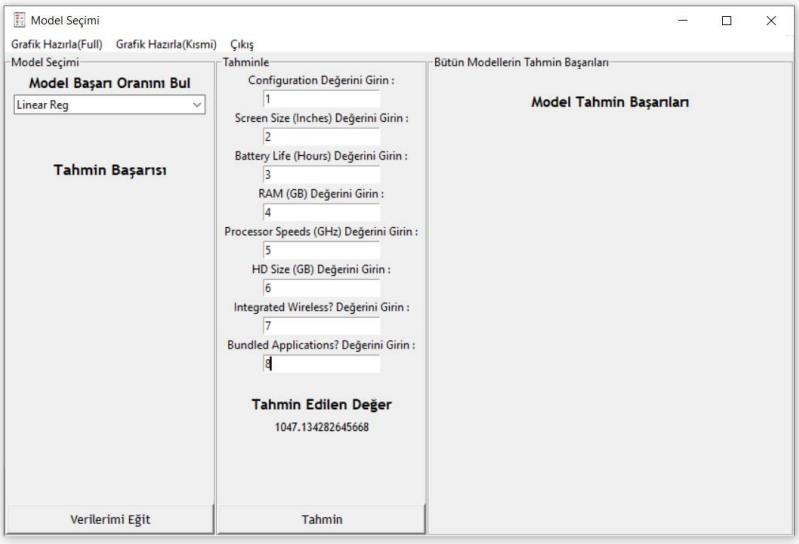
Şekil 48: Model Algoritmaları Gösterimi



Kaynak: Yazar tarafından derlenmiştir.

Tüm bu işlemlerin ardından tahminleme seçeneği seçilir ve ilgili kolon adı verilen kutucuklara tahminlemek istenilen değerler girilir. Ardından tahmin seçeneğine tıklanarak kullanıcı tarafından işlem tamamlanmış olur. Tahminlenen değerin gösterimi aşağıdaki gibidir.

Şekil 49: Tahmin Ekranı



Kaynak: Yazar tarafından derlenmiştir.

SONUÇ

Karar destek sistemleri eldeki bilgilerin harmanlanarak, ileride gerçekleşme ihtimali bulunan durumları öngörerek, yeni alınacak olan kararların doğruluk oranını arttıran sistemlerdir. Genel anlamıyla açıklamak gerekirse karar vericilerin, karar vermesine yardımcı olan sistemlerdir (Akgül, Üstündağ ve Tanrıverdi, 2018). Günümüzde birçok şirket büyük miktarda veriye sahip olmasına karşı, teknik anlamda yetersiz olmaları ve ellerindeki verileri kullanmayı bilmemeleri, yeterli düzeyde bilgi sahibi olmamalarından dolayı, verileri karar verme sürecine dâhil edememektedir.

Veriler işlenmemiş, ham bir şekilde bulunan enformasyon parçacıklarına verilen addır. Bu bağlamda enformasyon ise birbiriyle ilişkili olan verilerin, belirli bir amaca hizmet etmek için bir araya getirilerek verilere anlam kazandırılmasına denir. Bilgi ise deneyim ve enformasyonun bir araya getirilerek, anlamlı hale getirilmesine denir. Bu çalışma da amaç yukarıda anlatılan süreçte olduğu gibi işlenmemiş olan ham bir veriden bilgiye giden tüm süreci karar verme sürecine dâhil etmektir (Durna ve Demirel, 2008).

Çalışma kapsamında küçük ve orta ölçekli firmaların ellerinde bulunan verileri işleyebilmeleri ve bu verilerden anlamlı bilgiler çıkarabilmeleri çok önemlidir. Günümüzde KOBI'lerin ne yazık ki bu kapsamda çok yetersiz kaldığı, ellerindeki verileri kullanmakta eksik kaldıkları görülmektedir. Bu kapsamda incelendiğinde, bu uygulama üzerinden firmaların, ellerindeki verileri işleyebilecekleri, veriden bilgiye giden süreçte etkin olarak rol oynayabilecekleri görülmüştür. İlk olarak verinin temizlenmesi ve daha doğru sonuçlar vermesi için verinin düzenlenme işlemi yapılması gerekmektedir. İşletmeler bunların nasıl olacağını bilemediğinden dolayı ya bünyesinde yeni bir kişi barındırmakta ya da bu işlemlerle uğraşmamaktadırlar. Bu uygulama sayesinde verileri temizlemek adına uygulama içerisine Veri adında bir sekme açılmıştır. Bu sekme sayesinde tüm işlemler kolayca yapılabilmektedir.

Yapılan çalışmayı eldeki uygulamalardan ayıran en büyük özellik, diğer karar destek sistemlerinin yanı sıra içerisinde makine öğrenmesi algoritmalarını, zaman

serileri ve kümeleme analizini bir arada yapabilecekleri ve bunların raporlarını alabilecekleri entegre bir sistem olmasıdır. KOBI'ler elinde bulundurdukları verilerin yalnızca bugünü anlamak için değil geleceği planlamak adına da kullanabileceklerdir. Bu kapsamda bu uygulama üzerinde öncelikle ihtiyaç olan tüm ögeler belirlenmiştir ve bu kapsamda çalışmalara başlanmıştır. Verinin ham halinden bilgiye ulaşıncaya kadarki süreç adım adım takip edilmiş ve bu yönde ilerlenmiştir.

Bu kapsamda bu uygulama üç farklı kısımdan oluşmaktadır. İlk aşama, eldeki ham verilerin uygun hale dönüştürülmesi aşamasıdır. Bu aşamada kullanıcılar ellerindeki dosyaları sistemde açarlar ve gereksinim duyulan tüm işlemleri bu aşamada gerçekleştirip ellerindeki verileri tahminleme ve raporlamalar için uygun hale dönüştürürler. Uygun hale dönüştürülen veriler kayıt edildikten sonra ikinci aşamada tahminleme işlemleri yapılır. Kullanıcıların ellerindeki verileri tahminleyebilmesi amacıyla sistem tarafında 15 farklı algoritma kullanılmıştır. Bunlar arasından en iyisini sistem önerebilir ya da kullanıcılar kendi isteğine göre seçebilir. Son aşamada ise tüm bu verilerin raporları oluşturulur. Raporlar kendi içerisinde dörde ayrılır. Betimsel(açıklayıcı) raporlar, mevcut verilerin analizini yapan raporlardır. Tahmin raporlarında yöneticiler elindeki verilerden bir sonraki aşamaları tahminlemek için en uygun yöntemleri bulabilir ve bunların raporlarını alabilir. Bu kısımda 15 farklı makine öğrenmesi algoritması kullanılmıştır. Bir sonraki aşamada Zaman Serisi raporları bulunmaktadır. Bu raporlarda ilerleyen zamanlarda olabilecekleri öngörmek amacıyla tüm algoritmalar içerisinde en iyi sonucu veren ARIMA modeli kullanılmıştır. Kümeleme analizinde ise eldeki verilerin benzerliklerine göre gruplama yapılması amacıyla Gaussian Mixture algoritması kullanılmıştır.

Sonuç olarak, bu çalışmaya başlarken en başta düşünülen tüm hedefler, karar verme konumunda olan kişilerin ihtiyaçları gözetilerek oluşturulmuştur. Eldeki verilerden geleceğe yönelik tahminler yapabilen ve veriden bilgi elde etmeyi sağlayan bir karar destek sistemi geliştirilmiştir. Yönetim bilişim sistemlerinin çalışma alanıyla bağlantılı olarak hazırlanan bu uygulamanın, karar verme konumunda olan kişilere destek olabileceği öngörülmektedir. Hazırlanan bu uygulama sonucunda elde edilen yapılandırılmış verilerden oluşturulan raporların, tüm süreçlere fayda sağlayabileceği düşünülmektedir. Bu raporlardan

betimsel(açıklayıcı) raporlar, stratejik karar verme noktasında ne kadar satış olduğunu, müşterilerin hangi ürünlerden ne kadar aldığı, hangi mağazanın ne kadar ciro yaptığı gibi birçok bilgiyi içerisinde bulundurmaktadır. Bu sayede yöneticinin tüm bu bilgilere dayanarak karar vermesini kolaylaştırmayı amaçlamıştır. Taktiksel karar verme, içerisinde bulunan tahminleme yapısı sayesinde ileriki aylarda neler olabileceğine, hangi satışların yapılabileceğine, ne kadar malzeme alınması gerektiğine karar verme noktasında, birçok planlama yapılabilmesini sağlamaktadır.

İlerleyen çalışmalarda, uygulamayı web tabanlı bir yapıda oluşturmaya ve daha esnek hale getirmeye çalışılacaktır. Çünkü masaüstü uygulamaları kullanıcı etkileşimli raporların yapımında yetersiz kalmaktadır. Web tabanlı sistemlerin esnek yapısının, raporları daha zengin hale getirmeye yardımcı olacağı düşünülmektedir. Program ara yüzü c# ile yazılmış olsa da betimsel(açıklayıcı) raporlar haricindeki tüm tahminleme ve raporlar python ile yazıldığından dolayı esnek yapılar oluşturulmasına engel teşkil etmektedir. Bu yüzden web tarafındaki raporların, masaüstü tarafında elde edilen raporlara göre daha esnek bir yapıda olması sebebiyle daha basit ve kullanışlı olacaktır. Ayrıca büyük verilerin eğitim aşamaları uzun sürdüğünden algoritmaların bu süreyi daha da kısaltmasını sağlayacak iyileştirmeler ile daha hızlı bir sistem yapılması sağlanmak istenmektedir. Bunun yanında ileride bu sistemlere ek olarak, müşterilerden gelen destek taleplerinde ki sorunların anlaşılması adına metin madenciliği yönetimlerini kullanarak yöneticilere raporlar sunulabilmesi yönünde çalışmalar yapılacaktır.

KAYNAKÇA

Akdede, S.H. ve Turan, A.H. (2008). Bilişim sistemlerinin KOBİ'lerin performansına etkileri: kaynak temelli yaklaşım ile Denizli ilinde ampirik bir uygulama. *Ankara Siyasal Bilimler Fakültesi Dergisi,63*(4), 1-28.

Akgül, E., Üstündağ, M.T. ve Tanrıverdi, M.(2018). Perakende sektöründe kampanya yönetimine yönelik iş zekası uygulaması. *Artificial Intelligence Studies, 4*(1), 8-25. doi: 0.30855/AIS.2018.01.01.02

Akgün, M.(2019). *Kariyer planlama için karar destek sistemi* (Yayınlanmamış yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.

Akkol, E. (2019).*Makine öğrenmesi teknikleriyle itfaiye istasyonlarının performansının ölçülmesine yönelik web uygulaması*(Yayınlanmamış Yüksek Lisans Tezi).Dokuz Eylül Üniversitesi, İzmir.

Akpınar, H. (2000). Veri tabanlarında bilgi keşfi ve veri madenciliği. *İstanbul Üniversitesi İşletme Fakültesi Dergisi, 29*(1), 1-22.

Al-Asadi, M.A.M. (2018). *Decısıon support system for a football team management by usıng machıne learnıng techniques* (Yayınlanmamış yüksek lisans tezi). Selçuk Üniversitesi, Konya.

Alpar, R.(2013). *Uygulamalı çok değişkenli istatistiksel yöntemler*(4.baskı).Ankara: Detay Yayıncılık.

Alpaydın, E. (2010). *Introduction to machine learning* (2.baskı). United States of America : MIT Press.

Alzand, H.R.A. ve Karacan H. (2014). Bölümleyici kümeleme algoritmalarının farklı veri yoğunluklarında karşılaştırılması. *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Dergisi,56-62.*

Arda, E. (2020). *Yapay zeka yöntemleri ile finansal zaman serileri öngörüleri* (Yayınlanmamış Doktora Tezi). Başkent Üniversitesi, Ankara.

Argüden, Y. ve Erşahin, B. (2008). *Veri madenciliği, veriden bilgiye, masraftan değere* (1.baskı). İstanbul: ARGE Danışmanlık Yayınları.

Aronson, J.E., Liang, T.P. ve Turban, E. (2005). *Decision support systems and intelligent systems*(4.baskı). N.J.: Prentice Hall.

Arslan, H. (2008). *Web sitesi erişim kayıtlarının web madenciliği ile analizi* (Yayınlanmamış yüksek lisans tezi).Sakarya Üniversitesi, Sakarya.

Atan, S. (2020). Knn, naive bayes ve karar ağacı makine öğrenme algoritmaları, bu algoritmaların sosyal bilimlerde kullanım imkânları. *SocArXiv.* https://doi.org/10.31235/osf.io/8r5pu

Ay, D. ve Çil, İ.(2008). Migros Türk A.Ş.'de birliktelik kurallarının yerleşim düzeni planlamada kullanılması. *Endüstri Mühendisliği Dergisi, 21*(2), 14-29.

Aydın, K.E. (2020). *Web içerik sınıflandırması için makine öğrenmesi* (Yayınlanmamış yüksek lisans tezi).İstanbul Teknik Üniversitesi, İstanbul.

Ayhan, A.(2011). KOBİ'lerde kurumsal yönetime geçiş için bilgi teknolojileri uygulamalarında yapılması gerekenler. *Anahtar Dergisi*, 276.

Bal, M., Amasayali, M.F., Sever, H., Köse, G. ve Demirhan, A.(2014). Performance evaluation of the machine learning algorithms used in ınference mechanism of a medical decision support system, *The Scientific World Journal,* 1-15, doi: https://doi.org/10.1155/2014/137896

Bilgilioğlu, S. S. (2018). *Makine öğrenmesi teknikleri ile mekansal karar destek sistemlerinin geliştirilmesi: Aksaray ili örneği* (Yayınlanmamış yüksek lisans tezi). Aksaray Üniversitesi, Aksaray.

Box, G.E. P., Jenkins G.M., Ljung G.M. ve Reinsel G.C. (2016) *Time Series Analysis Forecasting and Control* (5.baskı). New Jersey: John Wiley & Sons.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

Büyüköztürk, Ş. (2020). *Sosyal bilimler için veri analizi el kitabı*(28.baskı).Ankara: Pegem Yayıncılık.

Can, Ö.Y. (2020). *Makine öğrenmesi teknikleri kullanılarak kredi risk analizi* (Yayınlanmamış yüksek lisans tezi).İstanbul Aydın Üniversitesi, İstanbul.

Cankar, İ. (2006). Denetimin yeni paradigması: sürekli denetim. *Sayıştay Dergisi,*61,69-81.

Chen, T. ve C. Guestrin (2016). Xgboost: A scalable tree boosting system. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,*785-794, doi: https://doi.org/10.1145/2939672.2939785

Coşkun, S., Kartal, M., Coşkun, A. Ve Bircan, H. (2004).Lojistik regresyon analizinin incelenmesi ve diş hekimliğinde bir uygulaması. *Cumhuriyet Üniversitesi Diş Hekimliği Fakültesi Dergisi, 7*(1), 41-50.

Crosswhite, C. E. (2003). Method for determining optimal time series forecasting parameters. *U.S. Patent No. 6,611,726*. Washington, DC: U.S. Patent and Trademark Office.

Çalış, Y.E., Keleş, E. ve Engin, A.(2014). Hilenin ortaya çıkartılmasında bilgi teknolojilerinin önemi ve bir uygulama. *Muhasebe ve Finansman Dergisi,* 93-103.

Çebi, S. (2010). *Aksiyomlarla tasarım esaslı bulanık karar destek sistemi geliştirme ve bir uygulama*(Yayınlanmamış Doktora Tezi).İstanbul Teknik Üniversitesi, İstanbul.

Çetinyokuş, T. ve Gökçen, H. (2002). Borsada göstergelerle teknik analiz için bir karar destek sistemi. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi,17*(1),43-58.

Çevik, O. (1999). *Zaman serileri analizinde Box-Jenkins yöntemi ve turizm verileri üzerine bir uygulama* (Yayınlanmamış Doktora Tezi).Kırıkkale Üniversitesi, Kırıkkale.

Çil, İ.(2002). Bilgi tabanlı imalat karar destek sistemleri ve bir uygulama. *EndüstriMühendisliği,13*(1), 15-27.

Dasgupta, A., Drineas, P., Harb, B., Josifovski,V. ve Mahoney M.W. (2007). Feature selection methods for text classification. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining içinde (230-239,ss). ACM.

Dem, F.(2019). *Data mınıng anwendungen ın entscheıdungsunterstützungssystemen und eın anwendungsbeıspıel beı der personalauswahl* (Yayınlanmamış yüksek lisans tezi). Marmara Üniversitesi, İstanbul.

Demolli, H., Ecemiş, A., Dokuz, A.Ş. ve Gökçek, M. (2019).Makine öğrenmesi algoritmalarıyla güneş enerjisi tahmini: Niğde ili örneği. *International Turkic World Congress on Science and Engineering,*775-783.

Dennis, A., Wixom, B. H., Roth, R. M. (2014). *System Analysis and Design* (5.baskı), USA: John Wiley & Sons, Inc.

Durna, U. ve Demirel, Y. (2008).Bilgi yönteminde bilgiyi anlamak. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi,*30, 129-156.

Duru, Ö. (2007). *Zaman serilerinin analizinde arıma modelleri ve bir uygulama* (Yayınlanmamış Doktora Tezi).İstanbul Üniversitesi, İstanbul.

Evaluatıng machıne learnıng algorıthm's performance ın predıctıng presence of heart disease. (2018, 13 Aralık).*Tonyeiyalla*. Erişim adresi http://tonyeiyalla.com/posts/heartdiseasepredictions/

Fatima, M. ve Pasha, M. (2017). Survey of machine learning Aalgorithms for disease diagnostic. *Journal of intelligent Learning Systems and Applications*, 9,1-16.

George E. P. Box, G. M. (2015). Time Series Analysis: Forecasting and Control. New Jersey: Wiley.

Gökçen, H. (2007). *Yönetim Bilgi Sistemleri* (BASKI). Ankara: Palme Yayıncılık.

Göker, H. (2019). *Makine öğrenmesi teknikleri kullanılarak çocukluk çağı dikkat eksikliği ve hiperaktivite bozukluğunun tahmin edilmesine yönelik uzman sistem tasarımı* (Yayınlanmamış doktora tezi).Gazi Üniversitesi, Ankara.

Gujarati, D. ve Porter, D. C.(2018) *Temel Ekonometri*(5.baskı). (Ü. Şeneşen ve G. Günlük Şeneşen, Çev.).İstanbul: Literatür Yayıncılık (Orijinal çalışma basım tarihi 2017).

Gülçe, G. (2010). *Veri ambarı ve veri madenciliği teknikleri kullanılarak öğrenci karar destek sistemi oluşturma*(Yayınlanmamış Yüksek Lisans Tezi).Pamukkale Üniversitesi, Denizli.

Güner, N. ve Çomak, E (2011). Mühendislik öğrencilerinin matematik I derslerindeki başarısının destek vektör makineleri kullanılarak tahmin edilmesi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 17*(2), 87–96.

Hagg, S., Cummings, M. ve Dawkins, J. (2012). *Managment information systems fort he information age*(9.baskı). Colombus: McGraw-Hill Higher Education.

Haykin, S.(1999). *Neural networks a comprarision foundation* (2.baskı). USA: Pearson Education.

Ho, T.K. (1998). The random subspace method for constructing decision forests. *IEEE Transaction on Pattern Analysis and Machine Intelligence, 20*(8), 832-844. Doi: 10.1109/34.709601

Honaker, J. ve King G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science, 54*(2), 561-581.

İmir, E. (1986). *Çoklu bağıntılı doğrusal modellerde ridge regresyon yöntemiyle parametre kestirimi: Türkiye'de (1963-1983) enflasyon analizi* (Yayınlanmamış yüksek lisans tezi).Anadolu Üniversitesi, Eskişehir.

Jaggi, M. (2014). An equivalence between the lasso and support vector machines. *Arxiv: 1303.1152v2*, 1-19.

James, G., Witten, D., Hastie, T. ve Tibshirani, R. (2013). *An Introduction to Statistical Learning* (7.baskı), New York: Springer.

Jiawei, H.(2006). *Cluster analysis, data mining: concepts and techniques* (3.baskı). USA: Elsevier Inc.

Kalaycı, S. (2018). *Makine öğrenmesi yöntemleri ile kredi risk analizi* (Yayınlanmamış yüksek lisans tezi).İstanbul Teknik Üniversitesi, İstanbul.

Karslı, Ö.B. (2019). *Makine öğrenme yöntemleri ile karaciğer hastalığının teşhisi* (Yayınlanmamış yüksek lisans tezi).Ağrı İbrahim Çeçen Üniversitesi, Ağrı.

Kavuncu, S.K. (2018). *Makine öğrenmesi ve derin öğrenme: nesne tanıma uygulaması* (Yayınlanmamış yüksek lisans tezi).Kırıkkale Üniversitesi, Kırıkkale.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. ve Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *31st Conference on Neural Information Processing Systems*, 1-9.

Keskin, M.V. (2018). *Büyük veride makine öğrenmesi uygulaması* (Yayınlanmamış yüksek lisans tezi).Yıldız Teknik Üniversitesi, İstanbul.

KNN Classification using Scikit-learn. (2018, 2 Ağustos). *Datacamp*. Erişim adresi https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn.

Kutlugün, M.A. (2017). *Gözetimli makine öğrenmesi yoluyla türe göre metinden ses sentezleme* (Yayınlanmamış yüksek lisans tezi).İstanbul Sabahattin Zaim Üniversitesi, İstanbul.

Kuzu, S. (2013). *Yapısal kırılmaları göz önüne alarak türk imalat sanayi ekonomik değişkenleri arasında uzun dönemli ilişkilerin araştırılması* (Yayınlanmamış Yüksek Lisans Tezi).İstanbul Üniversitesi, İstanbul.

Lantz, B. (2013). *Machine Learning with R* (1.baskı). Birmingham: Packt Publishing Ltd.

Legris, P., Ingham, J. ve Collerette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model. *Information and Management,40*(3), 191-204.

Li, D., Shen, J. ve Chen, H.(2008). A fast k-means clustering algorithm based on grid data reduction. 2008 IEEE Aerospace Conference. Doi: 10.1109/AERO.2008.4526465

Linoff, G.S. ve Berry, M.J.A. (2004). *Data mining techniques: for marketing, sales, and customer relationship management* (2.baskı). Indianapolis: Wiley Publishing.

Loh, W. Y. (2011). Classification and regression trees. *WIREs Data Mining and Knowledge Discovery, 1*(1),14–23.

Makine Öğrenimi (t.y.). Makine Öğrenimi wiki içinde.15 Aralık 2020 tarihinde https://tr.wikipedia.org/wiki/Makine_%C3%B6%C4%9Frenimi adresinden erişildi.

Makine Öğrenmesi ile İK Süreçleri Optimizasyonu. (t.y.). *Pmabilisim*. Erişim adresi http://pmabilisim.com/work1.html.

Marakas, G. M. (2003) *Modern Data Warehousing, Mining, and Visualization: Core Concepts.* New Jersey: Prentice Hall.

Michalski, R.S. ve Kodratoff, Y. (1990). Research in machine learning: Recent progress, classification of methods, and future directions. *In Machine Learning*, 3, 3-30. https://doi.org/10.1016/B978-0-08-051055-2.50004-3

Mitchell, T.M.(1997). *Machine learning* (1.baskı). ABD: McGraw-Hill Education.

Monhamady, K.K. (2018). *Developing machine learning methods for business ıntelligence* (Yayınlanmamış yüksek lisans tezi). Abdullah Gül Üniversitesi, Kayseri.

Murphy, K. P. (2012). *Machine Learning A Probabilistic Perspective*. (1.baskı). London: MIT Press.

Okur, H. (2020). *Makine öğrenmesi yöntemleriyle kredi riski tahmini* (Yayınlanmamış yüksek lisans tezi).Gazi Üniversitesi, Ankara.

Özata, M. ve Aslan, Ş. (2004).Klinik karar destek sistemleri ve örnek uygulamalar. *Kocatepe Tıp Dergisi,*5, 11-17.

Özaytekin, S. (2002), *Küçük ve orta boy işletmelerde bilgi ihtiyacı ve bilgi kaynaklarına ulaşmada bilgi sağlayıcının rolü* (Yayınlanmamış yüksek lisans tezi). Marmara Üniversitesi, İstanbul.

Özdamar, K.(2004). *Paket programlar ile istatistiksel veri analizi-2 (Çok Değişkenli Analizler)* (5.baskı). Eskişehir: Kaan Kitapevi.

Özdemir, O.(2020). *Performance comparıson of machıne learnıng methods and tradıtıonal tıme serıes methods for forecastıng* (Yayınlanmamış yüksek lisans tezi). Orta Doğu Teknik Üniversitesi, Ankara.

Özdemirci, F. ve Aydın, C. (2007). Kurumsal bilgi kaynakları ve bilgi yönetimi.*Türk Kütüphaneciliği, 21*(2), 164-185.

Özek, T. (2010). *Zaman serisi modelleri üzerine bir simülasyon çalışması* (Yayınlanmamış Yüksek Lisans Tezi).Selçuk Üniversitesi, Konya.

Özkan, Y. (2008). *Veri madenciliği yöntemler* (3.baskı).İstanbul: Papatya Yayıncılık.

Özmen, A. (1986). *Zaman serisi analizinde box-jenkins yöntemi ve banka mevduat tahmininde uygulama denemesi* (Yayınlanmamış Doktora Tezi).Anadolu Üniversitesi, Eskişehir.

Öztemel, E. (2012). *Yapay sinir ağları* (3.baskı).İstanbul: Papatya Yayıncılık.

Rizvanche, S. (2020). *Talep tahmininde makine öğrenmesi ve zaman serileri temelli bir karar destek sistemi* (Yayınlanmamış yüksek lisans tezi). Eskişehir Teknik Üniversitesi, Eskişehir.

Rokach, L. ve Maimon O. (2008). *Data mining with decision trees: theory and applications* (69.baskı). USA: World Scientific Publishing Co.

Rumora, L., Miler, M. ve Medak, D. (2020). Impact of various tmospheric corrections on sentinel-2 land cover classification accuracy using machine learning classifiers. *ISPRS International Journal of Geo-Information, 9*(4), 277, doi: 10.3390/ijgi9040277.

Ryan, J. (2018). Machine Learning: In Plain English. 01 Ocak 2021 tarihinde https://dzone.com/articles/machine-learning-in-plain-english, adresinden erişildi.

Silahtaroğlu, G.(2008). *Veri madenciliği kavram ve algoritmaları* (2.basım).İstanbul: Papatya Yayıncılık.

Silahtaroğlu, G.(2013). *Kavram ve algoritmalarıyla veri madenciliği kavram ve algoritmaları* (2.baskı).İstanbul: Papatya Yayıncılık.

Sistem geliştirme yaşam döngüsü, (t.y.). Sistem geliştirme yaşam döngüsü wiki içinde. 21 Aralık 2020 tarihinde https://tr.qaz.wiki/wiki/Systems_development_life_cycle adresinden erişildi.

Sönmez, C (t.y.). *Karar Destek Sistemleri*. 15 Aralık 2020 tarihinde https://web.itu.edu.tr/~sonmez/lisans/es/KararDestek.pdf, adresinden erişildi.

Şahinarslan, F.V. (2019). *Makine öğrenmesi algoritmaları ile nüfus tahmini: Türkiye örneği* (Yayınlanmamış yüksek lisans tezi).İstanbul Teknik Üniversitesi, İstanbul.

Şeker, S. E., Mert, C., Al-Naami, K., Ozalp, N. ve Ayan, U. (2013). Correlation between the economy news and stock market in Turkey. *International Journal of Business Intelligence and Review (IJBIR) , 4* (4), 1-21.

Şeker, S. E.(2014). Yazılım geliştirme hayat döngüsü. *YBS Ansiklopedi, 1*(2), 1-5.

Şeker, S. E., Cankir, B., ve Okur, M. E. (2014). Strategic competition of internet interfaces for XU30 quoted companies. *International Journal of Computer and Communication Engineering ,3*(6), 464-468.

Şeker, S. E.(2015). Yazılım geliştirme modelleri ve sistem/yazılım yaşam döngüsü. *YBS Ansiklopedi, 2*(3), 18-39.

Şeker, S.E. (2015). Zaman serisi analizi (time series analysis). *YBS Ansiklopedi, 2*(4), 23-31.

Tarı, R. (2010). *Ekonometri* (13.baskı)**.** Kocaeli; Umuttepe Yayınları.

Taşçı, H. C.(2019). *Emlak sektöründe makine öğrenme teknikleri kullanılarak iş zekası uygulaması geliştirilmesi* (Yayınlanmamış yüksek lisans tezi). Dokuz Eylül Üniversitesi, İzmir.

Tecim, V (t.y.). *Sistem Geliştirme Yaşam Döngüsü*. 2 Ocak 2021 tarihinde http://debis.deu.edu.tr/userweb//vahap.tecim/dosyalar/sgyd.pdf, adresinden erişildi.

Tekin, M. (1999). *Üretim yönetimi* (4.baskı).Konya: Arı Ofset.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, 58*(1), 267–288.

Turan, A.(2009). Küçük ve orta büyüklükteki işletmelerde(kobi) bilişim teknolojileri(bt), örgütsel rekabetçi stratejileri ve başarım ilişkisi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi,23*(3),105-122.

Turban, E.(1995). *Decision support and expert systems: management support systems* (4.baskı). N.J.: Prentice Hall.

Turunç, Ö. (2006). *Bilgi teknolojileri kullanımının işletmelerin örgütsel performansına etkisi* (Yayınlanmamış Doktora Tezi).Süleyman Demirel Üniversitesi, Isparta.

Tüzün Rad, S., Gülmez, Y.S., Gökbudak, F. ve Yılmaz, E.K.(2016). KOBİ'lerde bilişim teknolojileri kullanımı: Erdemli örneği. *Türk ve İslam Dünyası Sosyal Araştırmalar Dergisi, 3*(6),59-69.

Uğur, A. ve Kınacı, A. C. (2006). Yapay zekâ teknikleri ve yapay sinir ağları kullanılarak web sayfalarının sınıflandırılması. *XI. Türkiye'de İnternet Konferansı (inet-tr'06)* içinde (345-349.ss). Türkiye: Ankara.

Uzun, E. (2007). *İnternet tabanlı bilgi erişimi destekli bir otomatik öğrenme sistemi* (Yayınlanmamış doktora tezi).Trakya Üniversitesi, Edirne.

Üçkardeş, F., Efe, E., Narinç, D. ve Aksoy, T. (2012).Japon bıldırcınlarında yumurta ak indeksinin ridge regresyon yöntemiyle tahmin edilmesi. *Akademik Ziraat Dergisi 1*(1), 11-20.

Ünsal, Ö. (2011). *Mesleki alan seçimlerinin makine öğrenmesi algoritması kullanılarak belirlenmesi* (Yayınlanmamış yüksek lisans tezi).Gazi Üniversitesi, Ankara.

Üstüner, M., Abdikan, S., Bilgin, G. ve Balık Şanlı, F. (2020). Hafif gradyan artırma makineleri ile tarımsal ürünlerin sınıflandırılması. *Türk Uzaktan Algılama ve CBS Dergisi, 1*(2), 97-105.

Yeşilnacar, E. ve Topal, T. (2005). Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek Region (Turkey), *Engineering Geology*, 79, 251-266.

Yapay sinir ağları (2018,13 Ekim). *Psikolojik*. Erişim adresi https://www.psikolojik.gen.tr/yapay-sinir-aglari.html

Yılmaz, R.(2013). *Türkçe dokümanların sınıflandırılması* (Yayınlanmamış yüksek lisans tezi).Aydın Adnan Menderes Üniversitesi, Aydın.

Why unsupervised machine learning is the future of cybersecurity. (t.y.). *Technative.* Erişim adresi https://www.technative.io/why-unsupervised-machine-learning-is-the-future-of-cybersecurity/.

Wu, R. C., Chen, R. S. ve Chian, S. S. (2006). Design of a product quality control system based on the use of data mining techniques. *IIE Transactions*, *38*(1), 39-51.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 45*(1), 5-32.