T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

TEDARİK ZİNCİRİ İÇİN MAKİNE ÖĞRENMESİ ÇERÇEVESİNDE İSTATİSTİKSEL ANALİZ

Kübra KARTALOĞLU

YÜKSEK LİSANS TEZİ

İstatistik Anabilim Dalı

İstatistik Programı

Danışman

Prof. Dr. Ali Hakan BÜYÜKLÜ

Eylül, 2022

T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

TEDARİK ZİNCİRİ İÇİN MAKİNE ÖĞRENMESİ ÇERÇEVESİNDE İSTATİSTİKSEL ANALİZ

Kübra KARTALOĞLU tarafından hazırlanan tez çalışması 20.09.2022 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, İstatistik Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Ali Hakan BÜYÜKLÜ
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Prof. Dr. Ali Hakan BÜYÜKLÜ, Danışman
Yıldız Teknik Üniversitesi _____

Doç. Dr. Öyküm Esra YİĞİT, Üye
Yıldız Teknik Üniversitesi _____

Prof. Dr. Eylem DENİZ HOWE , Üye
Mimar Sinan Güzel Sanatlar Üniversitesi _____

Danışmanım Prof. Dr. Ali Hakan BÜYÜKLÜ sorumluluğunda tarafımca hazırlanan Tedarik Zinciri İçin Makine Öğrenmesi Çerçevesinde İstatistiksel Analiz başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Kübra KARTALOĞLU

İmza

Aileme

TEŞEKKÜR

Bu çalışmanın hazırlanma aşamasında beni destekleyen, bilgi ve tecrübelerini benimle paylaşan tez danışmanım Prof. Dr. Ali Hakan Büyüklü'ye ve hayatımın her aşamasında yanımda olan biricik aileme sonsuz teşekkürlerimi sunarım.

Kübra KARTALOĞLU

İÇİNDEKİLER

SİMGE LİSTESİ vii

KISALTMA LİSTESİ viii

ŞEKİL LİSTESİ ix

TABLO LİSTESİ x

ÖZET xi

ABSTRACT xii

1 GİRİŞ 1

1.1 Literatür Özeti ... 1

1.2 Tezin Amacı ... 2

1.3 Hipotez ... 3

2 TEDARİK ZİNCİRİ YÖNETİMİ 4

2.1.Tedarik Zinciri Yönetimi ... 4

2.2 Tedarik Zinciri Yönetim Süreçleri ... 5

2.2.1 Müşteri İlişkileri Yönetimi ... 5

2.2.2 Müşteri Hizmet Yönetimi ... 5

2.2.3 Talep Yönetimi ... 6

2.2.4 Sipariş Yönetimi ... 6

2.2.5 Üretim (İmalat) Yönetimi ... 6

2.2.6 Satın Alma (Tedarikçi İlişkileri) Yönetimi ... 6

2.2.7 Ürün Geliştirme ... 7

2.2.8 İade Yönetimi ... 7

2.3 Tedarik Zinciri Yönetimi Faydaları ... 7

2.4 Tedarik Zinciri Risk Faktörü ... 8

3 MAKİNE ÖĞRENMESİ 11

3.1 Makine Öğrenmesi 11
3.2 Makine Öğrenmesi Algoritma Yöntemleri 11
3.2.1 Denetimli Öğrenme 12
3.2.2 Denetimsiz Öğrenme 12
3.2.3 Yarı Denetimli Öğrenme 12
3.3 Denetimli Öğrenme Yöntemleri 12
3.3.1 Karar Ağacı Sınıflandırması 13
3.3.2 Topluluk Öğrenme Yöntemleri 15
3.3.3 Lojistik Regresyon 22
3.4 XGBoost ve Lojistik Regresyon 28
3.5 Modelleri Test Etme ve Doğrulama 30
3.6 Modellerin Değerlendirilmesi 30
3.6.1 Karışıklık Matrisi 30
3.6.2 ROC Eğrisi (Receiver Operating Characteristic Curve) 32
4 UYGULAMA **33**
4.1 Veri Düzenleme ve Değişken Açıklama 33
4.2 Verilerin Görselleştirilmesi 36
4.3 Lojistik Regresyon Modeli 41
4.4 XGBoost Modeli 43
4.5 Modellerin Performanslarının Karşılaştırılması 46
5 SONUÇ VE ÖNERİLER **47**
6 KAYNAKÇA **48**
7 TEZDEN ÜRETİLMİŞ YAYINLAR **53**

SİMGE LİSTESİ

H	Entropi
V	Orjinal veri kümesinin alt kümesi
S	Orjinal veri kümesi
D	Bölünmüş veri kümesi
f_t	Zayıf öğrenici
$h(x_i)$	Zayıf öğrenici çıktı varsayımı
α_t	Eğitim hatası
$F_{t-1}(x)$	Güçlendirilmiş sınıflandırıcı
$(w_{i,t})$	Ağırlık
$L(y_i, y_i^p)$	Kayıp fonksiyonu
$\sum(y_i - y_i^p)$	Artıkların toplamı
$\emptyset$	Boş küme
Ω	Düzenlileştirme
$y = x'\beta + e$	OLS model
$\Lambda(x'\beta)$	Lojistik modelin kümülatif dağılım fonksiyonu
$\Phi(x'\beta)$	Standart normal dağılımın cdf'si
β_j	Regresyon modelinin marjinal etkileri
$x'\beta$	Odds oranı

KISALTMA LİSTESİ

KNN	K-nearest Neighborhood
CHAID	ChiSquare Automatic Interaction Detection
CART	Classification and Regression Tree
Bagging	Bootstrap Aggregation
AdaBoost	Adaptive Boosting
XGBoost	Extreme Gradient Boosting
CatBoost	Categorical Boosting
GPU	Graphics Processing Unit
SVM	Support Vector Machines
ROC	Receiver Operating Characteristics
OLS	Ordinary Least Squares

ŞEKİL LİSTESİ

Şekil 2.1 Tedarik zinciri yönetim süreçleri 5
Şekil 3.1 Torbalama (Bagging) Algoritması 16
Şekil 3.2 Güçlendirme algoritmaları tarihçesi 21
Şekil 3.3 Sigmoid fonksiyonu 28
Şekil 3.4 XGBoost akış şeması 28
Şekil 3.5 Lojistik regresyon akış şeması 29
Şekil 3.6 ROC eğrisi 32
Şekil 4.1 Yapılan ödeme türlerinin departmanlara göre grafiği 37
Şekil 4.2 Geç teslimat riskinin departmanlara göre grafiği 37
Şekil 4.3 Sipariş durumlarının departmanlara göre grafiği 38
Şekil 4.4 Müşteri türlerinin departmanlara göre grafiği 38
Şekil 4.5 Teslimat durumlarının departmanlara göre grafiği 39
Şekil 4.6 Siparişlerin gönderim tarzlarının departmanlara göre grafiği 39
Şekil 4.7 Sipariş tarihlerinin departmanlara göre grafiği 40
Şekil 4.8 Eğitim seti karışıklık matrisi 42
Şekil 4.9 Test set karışıklık matrisi 42
Şekil 4.10 ROC eğrisi 43
Şekil 4.11 Karışıklık matrisi 44
Şekil 4.12 Önem matrisinin grafiği 45

TABLO LİSTESİ

Tablo 3.1 Lojistik regresyonun uygulandığı durumlar 22
Tablo 3.2 Tahmin edilen değerler 27
Tablo 3.3 XGBoost ve lojistik regresyon özellikleri 29
Tablo 4.1 Değişken açıklamaları..... 34
Tablo 4.2 LDR level..... 41
Tablo 4.3 Test verisi..... 41
Tablo 4.4 Lojistik regresyon modeli 41
Tablo 4.5 Minimum kayıp değerleri 44
Tablo 4.6 İstatistikler ve doğruluk oranları..... 45

ÖZET

Tedarik Zinciri İçin Makine Öğrenmesi Çerçevesinde İstatistiksel Analiz

Kübra KARTALOĞLU

İstatistik Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Prof. Dr. Ali Hakan BÜYÜKLÜ

Lojistik ağı adıyla da bilinen tedarik zinciri, hizmet ya da ürünlerin tedarikçiden müşteriye olan hareketlerini kapsayan süreçtir. Geniş kapsamlı olarak değerlendirildiğinde tedarik zinciri kavramı, hammadde temini, üretimi, depolanması, depo takibi, stok kontrolü, ürün siparişlerinin bilgilerinin değerlendirilmesi ve aktarılması, dağıtımı ve son olarak tüketiciye ulaşımından sonra memnuniyetinin oluşması açısından birçok değeri kapsamaktadır. Stok takibi ve satış oranlarına göre yeni stok siparişini ayarlamak için optimum süre bulunabilir. Bununla beraber tedarik süreci optimize edilerek depodan çıkan ürün adedini artırarak kapasite genişletilebilir. Başarılı bir tedarik zinciri yönetimi ile firmaların rekabet gücü daha da artarken, maliyetler azalacak ve firmalar rakiplerine karşı daha avantajlı hale gelecektir. Tedarik zincirinde süreci hızlı ve verimli yönetmek kalite artışını da sağlamaktadır. Tedarik zinciri yönetiminde temel iş süreçlerinin entegrasyonu sağlanarak müşteri memnuniyetini artıracak iş modelleri ve stratejilerin oluşturulması bu yüzden çok önemlidir.

Anahtar Kelimeler: Tedarik zinciri yönetimi, geç teslimat riski, sevkiyat

ABSTRACT

Statistical Analysis in the Framework of Machine Learning for the Supply Chain

Kübra KARTALOĞLU

Department of Statistics

Master of Science Thesis

Supervisor: Prof. Dr. Ali Hakan BÜYÜKLÜ

The supply chain, also known as the logistics network, is the process that covers the movement of services or products from the supplier to the customer. When evaluated comprehensively, the concept of supply chain includes many values in terms of raw material supply, production, storage, warehouse monitoring, stock control, evaluation and transfer of information on product orders, distribution and, finally, satisfaction after reaching the consumer. Optimum time can be found to adjust the new stock order according to stock tracking and sales rates. In addition, the capacity can be expanded by optimizing the supply process and increasing the number of products leaving the warehouse. With a successful supply chain management, the competitiveness of companies will increase, costs will decrease and companies will become more advantageous against their competitors. Managing the process quickly and efficiently in the supply chain also provides an increase in quality. Therefore, it is very important to establish business models and strategies that will increase customer satisfaction by integrating basic business processes in supply chain management.

Keywords: Supply chain management, late delivery risk, shipment

1
GİRİŞ

1.1 Literatür Özeti

Gelişen teknoloji ile oluşan yeni ekonomi işletmelerin yönetim biçimlerini değiştirmiştir. İşletmeler, müşterilerin istek ve beklentilerini anlamak, müşterilerin taleplerini doğru şekilde, doğru zamanda ve doğru yerde yerine getirmek için yönetim biçimlerini aktif bir şekilde sürdürmeleri gerekmektedir.

Gelişen teknolojiyle ve artan rekabet koşullarıyla müşteri bulmak zor hale gelmişken olan müşterileri de elde tutmak oldukça zorlaşmıştır. Bu yüzden tüm işletmeler müşteri kaybını önlemek için tedarikçiden müşteriye kadar olan tedarik zinciri yönetim sürecini en iyilemek durumunda kalmışlardır. Böylece rekabet artık işletmeler arasında değil tedarik zinciri yönetimi arasında olmaktadır. Şirketler kar oranlarını artırabilmek ve gelişebilmek için tedarik zincirlerini en verimli şekilde modellemektedirler. Tedarik zincirinde oluşabilecek risklerin yönetilmesinde tedarik zinciri üyelerinin sorumluluğu her geçen gün artmaktadır. Tedarik zinciri yönetimi, risklerin belirlenmesi, değerlendirilmesi, düzenli aralıklarla ve doğru yöntemlerle gözden geçirilmesinde tüm sürecin dahil edildiği stratejik bir yönetim biçimidir.

Müşterilerin talep ettikleri ürünleri tedarikçiden en uygun maliyetle ve zamanında temin etmek, ürünleri zamanında teslim etmek işletmelere avantaj sağlayacaktır. Gelişen ve devamlılığı süren işletmelerin kendine göre bir stratejisi vardır. Tedarik zinciri yönetimi stratejilerini bütün koşullara uygun olarak planlayan işletmeler rakiplerine göre avantajlı ve başarılı işletmelerdir. Tedarik zincirinde oluşabilecek her türlü risk işletmeyi olumsuz etkilemektedir. Ürün veya hizmetlerin doğru zamanda ve doğru şekilde müşteriye sunulmaması durumunda işletmeye duyulan güven sarsılacaktır. Talep edilenlerin teslimatının geç yapılması müşteri kaybına yol açacaktır. Sürekli kontroller sağlanarak oluşabilecek riskler azaltılabileceği gibi işletmeyi olumsuz etkileyecek riskler minimize edilebilecektir.

1.2 Tezin Amacı

Siparişlerin herhangi bir sorun çıkmadan zamanında teslim edilebilmesi için işletmelerin faaliyetlerinde organizasyonun sağlanması gereklidir. Müşteriden gelen ve gelecek olan siparişlerin öngörülmesiyle üretim planlanması yapılmalıdır. Üretim faaliyetleri için kullanılacak olan tüm hizmetler sorunsuz, hızlı ve beklenen şekilde temin edilmelidir. Planlama sürecinde belirli miktarda ürünün stok seviyesinin belirlenmesiyle stok yönetimi yapılmalıdır. Üretimi gerçekleşmiş olan ürünlerin sevkiyat için hazır hale getirilmesi, depolanması ve maliyetlerin çıkartılması da depo yönetimi ile gerçekleşir. Hazır olan ürünlerin depodan alınarak tüketici veya satıcıya ulaştırılması sevkiyat aşamasında gerçekleşir. Sevkiyat hızı, kapasitesi ve maliyetleri sürekli takip edilmelidir. Bununla birlikte başarılı bir tedarik zinciri yönetimi, işletmelerin kısa vadede de uzun vadede de verimlilik artışı ve karlılık sağlamasında önemli rol oynar. İşletmelerin sevkiyat tarihlerini sipariş tarihlerine göre ayarlamaları müşteri memnuniyeti açısından çok önemlidir. Planlanan teslimat günleri ile erken teslimat veya zamanında teslimat olması durumunda tüketicilerin işletmeye olan güveni artacak ve paralel olarak tüketici tercihleri, erken teslimat veya zamanında teslimat yapan işletmeye daha fazla yönelecektir.

Çalışmanın temel amacı, tedarik zinciri yönetimi risklerinden biri olan geç teslimat riskinin işletmeler ve tüketici açısından önemini vurgulamaktır. Tedarik zinciri yönetimi riskleri çerçevesinde oluşabilecek risklerin tanımlanması ve risklerin sınıflandırılması öncelik taşımaktadır. Bu bakış açısından yola çıkarak hazırlanan çalışmada kullanılan tedarik zinciri verisinde, geç teslimat riskinin etki düzeyi incelenmiştir. Hizmet sektöründe faaliyet gösteren işletmelerde sektöre özgü risk sınıflandırmasının yapılması bu çalışmanın esas hedefini oluşturmaktadır.

Değerlendirmenin amacı, belirlenen risk faktörünün hangi etkenlere karşı daha duyarlı olduğunu anlamaya çalışmaktır. Gerekli risk yönetim stratejisinin geliştirilmesi için bu çalışmanın etkili bir başlangıç noktası olduğu düşünülmektedir.

1.3 Hipotez

Tedarik zinciri yönetiminin başarılı bir şekilde yönetilebilmesi işletmeler ve tüketici açısından önemli rol oynamaktadır. Buna bağlı olarak tedarik zinciri yönetiminin başarılı bir şekilde sürdürülememesiyle belirli aksamalar sonucu çeşitli riskler oluşmaktadır. Çalışmada, tedarik zincirinde kullanılan veri setindeki departmanların geç teslimat riskine göre makine öğrenmesi teknikleriyle sınıflandırılıp sınıflandırılamayacağı ve geç teslimat riskinin oluşup oluşmayacağı üzerine hipotezler kurulmuştur.

Çalışmanın birinci bölümünde araştırma yapılan konu hakkında giriş yapılmıştır. Çalışmanın amacı ve araştırma yöntemleri hakkında bilgiler verilmiştir.

Tezin ikinci bölümünde tedarik zinciri yönetimi üzerine literatür taraması yapılmış olup, tedarik zinciri yönetimi, tedarik zinciri yönetim süreçleri, tedarik zinciri yönetiminin faydaları, tedarik zincirinde oluşabilecek riskler ile tez konusuna yönelik bilgiler verilmiştir.

Tezin üçüncü bölümünde makine öğrenmesi ve makine öğrenmesinin algoritmaları hakkında bilgiler verilmiştir. Tezde kullanılacak olan makine öğrenmesi tekniklerinden olan Lojistik Regresyon ve Extreme Gradient Boosting hakkında daha detaylı bilgiler verilmiştir.

Tezin dördüncü bölümünde tedarik zinciri yönetiminde siparişlerin teslimat durumları, gönderim tarzları, sipariş ve sevkiyat tarihlerine göre geç teslimat riskleri, satın alınan ürünlerin planlanan teslimat günleri ve gerçek teslimat günleri karşılaştırılıp R Studio programlama dilinde görsellerle incelenmiştir. Ayrıca makine öğrenmesi tekniklerinden Lojistik Regresyon ve XGBoost modelleriyle geç teslimat riski için sınıflandırma yapılmıştır.

Tezin son bölümünde ise tedarik zincirindeki başarıyı etkileyen nedenlerden biri olan teslimat aşamasında oluşan riskler göz önünde bulundurularak yapılan uygulamada kullanılan makine öğrenmesi tekniklerinin sonuçları karşılaştırılmıştır. Buna yönelik yapılan çalışmada geç teslimat riskinin önemi vurgulanmıştır.

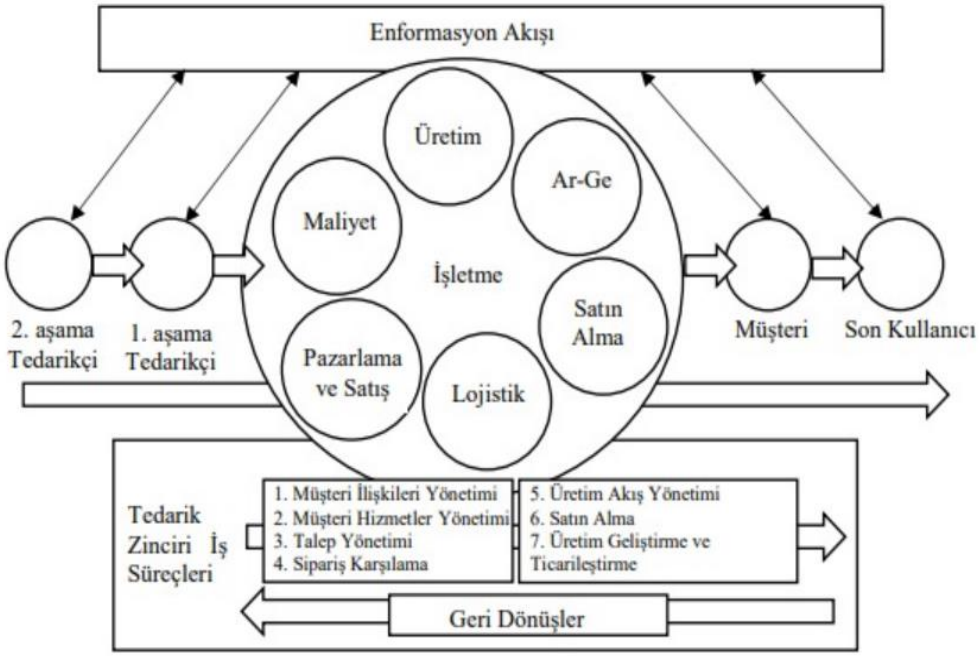
2
TEDARİK ZİNCİRİ YÖNETİMİ

2.1.Tedarik Zinciri Yönetimi

Teknolojinin gelişmesiyle işletmelerin ulusal ve uluslararası alanda daha fazla müşteriye ulaşması sağlanmıştır. Bununla beraber müşterinin talep ve ihtiyaçlarının doğru zamanda doğru bir şekilde karşılanması, müşteri memnuniyetinin kazanılması ve üretimi karlı bir şekilde gerçekleştirmek işletmeler açısından büyük önem taşımaktadır. Alınan mal ve hizmetlere düşük maliyet ile erişim, yönetim giderlerini azaltmak, ürün kalitesini düşürmemek, rekabet gücünü artırmak, güvenilir firmalar ile anlaşmak, tedarikçi firmaların sürekliliğini sağlamak, kesintisiz malzeme ve bilgi akışı sağlamak, stok kayıplarını minimum düzeyde tutmak işletmelerin genel itibariyle tedarik zinciri yönetimi ile hedeflediği amaçlardır. Dolayısıyla tedarik zinciri yönetimi işletmeler için önemli olan bu amaçlara yardım eden bir hizmettir. Tedarik zinciri yönetimi üreticiye gerekli olan hammaddenin üretiminden, tüketiciye ulaşmasını sağlayana kadar birbirini takip eden aşamalarda gerçekleşir.

Tedarik zinciri yönetimi üreticiden tüketiciye kadar bulunan aşamalarda olan kişilerin tümüne avantaj sağlar. Tedarik süreci hızlandırılarak, talep edilen ürünlerin tüketiciye daha hızlı ulaşmasına yardımcı olur. İşletmelerin sipariş karşılama oranlarının artmasını sağlar. Kaynaklar daha doğru kullanıldığından verimlilik artar. İşletmeler bakımından maliyet azalır. Zamanında ve sorunsuz gerçekleşen teslimatlarla müşteri memnuniyeti kazanılır. Tedarik zinciri yönetiminde önemli rol oynayan stok seviyeleri, maliyet ve varlık yönetimi, arz ve talep dengesi, lojistik gibi faaliyetlerin doğru planlanması gerekmektedir. Bu nedenle işletmeler kendilerine uygun tedarik zinciri yönetimi stratejileri ile çalışmalıdırlar. Tedarik zinciri yönetiminin en önemli unsurlarından biri de lojistiktir. Ürün veya hizmetlerin tüketiciye başarılı bir şekilde ulaşabilmesi için işletmelerin iyi bir lojistik hizmetine sahip olmaları gerekmektedir.

2.2 Tedarik Zinciri Yönetim Süreçleri



Şekil 2.1 Tedarik zinciri yönetim süreçleri [24]

Tedarik zinciri yönetiminde tedarikçilerden tüketicilere kadar bir ilişkinin varlığından söz edilir. Tedarik zincirini oluşturan bağlantılar arasındaki ilişki sistemlidir. Üretimden başlayarak ilerleyen süreçte stok kontrolü, ekipman temini, dağıtım, satış, müşteri ilişkileri yönetimi gibi süreçler tedarik zinciri yönetiminin gelişmesini sağlamaktadır.

2.2.1 Müşteri İlişkileri Yönetimi

Tedarik zinciri yönetimi süreçlerinin en başında müşteri ilişkileri yönetimi gelmektedir. Artan rekabetin etkisiyle müşteriler, işletmeler açısından büyük önem arz etmektedir. Artan rekabetle müşteriyi elde tutmak ne kadar zorsa yeni müşteri bulmak da o kadar zor olacağından, işletmeler gelişebilmek adına müşteri talep ve ihtiyaçlarını en iyi şekilde karşılamak durumundadırlar. Bu süreç müşterilerle ilişkilerin iyileştirilip, bu iyileşmenin de sürdürülmesi ile ilgilenir. Müşterilerin talep ve ihtiyaçlarını karşılamak üzere ürün ve hizmet anlaşmaları hazırlanır. Ayrıca bu süreçte işletmenin, müşteriler üzerindeki finansal etkilerini ölçmek için gerekli raporlar da hazırlanır.

2.2.2 Müşteri Hizmet Yönetimi

Bu süreçte, talep edilen ürünün ulaşım uygunluğu, yükleme zamanı ve sipariş durumu hakkında müşterilere gerçek ve tam zamanlı bilgiler verilir. Müşteriler ile yapılan ürün

ve hizmet anlaşmalarının yönetimi ile bu süreç ilgilenir. Tedarik ile ilgili problemlerde danışma imkanının sağlanması, işletmenin satış ve servislerinde kalitenin sağlanması, siparişin yenilenmesi halinde esneklik sağlanması gibi görevleri mevcuttur.

2.2.3 Talep Yönetimi

Tedarik zinciri yönetiminde müşteri talepleri ile işletmenin talepleri karşılayabilmesi için tedarik olanakları dengelenmelidir. Bu sürecin başarılı ve doğru bir şekilde yürütülmesi ile arz ve talep dengesi sağlanmış olacaktır. Başarılı bir tedarik zinciri ile ürün ve hizmet akışının sağlanması verimli bir şekilde gerçekleştirilecektir. Bu süreç, talep tahmini, satın alma ve dağıtımın birbiri ile uyumlu bir şekilde ilerlemesini amaçlar. Ayrıca geç teslimat durumunun oluşması gibi beklenmedik gelişmelere karşı alternatif planlar geliştirerek bu planların yürütülmesi ile ilgilenir.

2.2.4 Sipariş Yönetimi

Diğer yönetim süreçlerinde olduğu gibi sipariş yönetiminde de temel amaç müşterilerin ihtiyaçlarının karşılanmasıdır. Siparişler yerine getirilirken zincirdeki tüm süreçlerin birbiriyle bütün olması gerekmektedir. İşletme, toplam maliyeti azaltabilmek için tedarik zincirinde yer alan ortaklarla ilişkilerini geliştirmelidir. Müşteri ihtiyaçları bununla beraber verimli bir şekilde karşılanabilir.

2.2.5 Üretim (İmalat) Yönetimi

Ürünlerin üretiminden sonra pazara en iyi şekilde hizmet edilmesi adına imalat esnekliğinin sağlanması ile ilgilenilen süreçtir. Ürün çeşitleri ve miktarları istenilen zamanda değiştirildiğinde sabit üretimden dolayı stok maliyetleri de azaltılabilir. Fakat fazla miktarda istenilen ürünlerin karşılanabilmesi için işletme ek ekipmana ihtiyaç duymak zorundadır. Talebin yoğun olmadığı dönemlerde ek ekipmanların maliyeti de işletmeyi ilgilendirmektedir. Bu yüzden müşteri ihtiyaçları karşılanırken daha düşük maliyetle yapılması açısından malzeme ihtiyaç planlamasının da doğru yapılması gerekmektedir.

2.2.6 Satın Alma (Tedarikçi İlişkileri) Yönetimi

Bir işletmecinin ihtiyaçlarını makul fiyat ve kaliteden temin etmesine satın alma denilmektedir. İşletme ve tedarikçiler arasındaki ilişkileri kapsayan süreçtir. İşletmenin tedarikçiler ile nasıl iletişim kurması gerektiği ile ilgilenir. Nasıl ki işletmenin sürdürebilirliği adına müşterilerle ilişkilerin geliştirmesi gerekiyorsa tedarikçilerle de

geliştirilmesi gerekmektedir. Ayrı ayrı bütün tedarikçiler ile hizmet anlaşmaları yapılmalıdır ve bu anlaşmaya uyulması gerekmektedir. Bu süreç, tedarikçilerle yapılan ürün ve hizmet anlaşmalarının yürütülmesinden sorumludur.

2.2.7 Ürün Geliştirme

İşletmeler müşterilerin değişen ihtiyaçlarını karşılayabilmek için müşteriye ürün alternatifi sunmalıdır. Yeni ürün geliştirirken hem kendi pazarlama fonksiyonlarını hem tedarikçilerle olan ilişkilerini entegre etmelidir. Bu süreçle ilgilenen yöneticilerin mevcut olan müşterilerin veya henüz talepte bulunmamış olan müşterilerin ihtiyaçlarını saptamalı ve buna göre bir yol izlemelidir. Yeni ve farklı ürünleri hedef olan pazara ulaşım süresini düşürmek amacıyla müşteriler ve tedarikçiler de bu sürece dahil edilmelidir.

2.2.8 İade Yönetimi

Lojistik kavramı üreticiden tüketiciye olabildiği gibi tüketiciden üreticiye de olabilir. Ürünlerin satılmaması, kalitesini yitirmesi, müşterilerin fikir değiştirmesi gibi nedenlerden dolayı ürünler geri dönebilir. Ürün iade yönetiminin başarılı bir şekilde gerçekleşmesiyle rekabetçi işletmelere karşı avantaj sağlanacaktır.

2.3 Tedarik Zinciri Yönetimi Faydaları

Tedarik zinciri yönetiminin doğru ve sorunsuz bir şekilde yürütülmesi sayısız katkı sağlamaktadır. Bu katkılardan bazıları şu şekilde ifade edilmiştir;

1)Maliyetlerde tasarruf sağlanır.

2)Kaynakların doğru kullanımıyla verimliliği ve kapasiteyi artırır.

3)En iyi seviyede siparişlerin karşılanması sağlanır.

4)Müşterilerin beklenti ve ihtiyaçları başta olmak üzere öngörülmemiş taleplere de hızlı cevap verme imkânı sunar.

5)Stoklar azalır.

6)Sorunsuz ve zamanında gerçekleşen teslimatlarla müşteri memnuniyeti kazanılır.

7) Ani değişimlere karşı işletmenin daha hazırlıklı olmasını sağlar.

8) İşletmenin tüm bilgi ve para akışı yönetilebilir hale gelir.

Bu yararlar ve daha fazlası aynı zincirde yer alan firmaların (tedarikçi, üretici, dağıtıcı, perakendeci vb.) arasında iletişimin tam olarak kurulması, zincir boyunca faaliyetlerin birlikte koordinasyonu ve kontrolü sayesinde ortak amaç olarak belirlenen zincirin bütününde maliyetlerin azaltılması, verimliliğin artırılması, karlılık ve müşteri tatmini gibi amaçlara ulaşmak üzere elde edilebilir[1]. Aynı tedarik zincirinde yer alan firmalar arasındaki iletişimin tam olarak kurulmasıyla bu katkılar ve daha fazlasıyla ortak amaçlar gerçekleştirilir. Başarılı bir tedarik zinciri gelişmeye yardımcı olabileceği alanları belirlemek için aksayan ve durma noktasına gelmiş olan süreçleri gözden geçirir.

2.4 Tedarik Zinciri Risk Faktörü

Tedarik zincirinde meydana gelen çeşitli sebeplerden dolayı oluşan risklerin firmaya olumsuz etkileri bulunmaktadır. Maliyetlerdeki artış, kalite düşüşü, müşteriyi riske sokan güvenlik sorunlarının oluşması, finansal kayıplar, ürün teslimatlarındaki gecikmeler gibi birçok neden bulunmaktadır. İşletmelerin bu olumsuz sonuçlardan etkilenmemeleri için oluşabilecek bu riskleri tanımlayabilmeleri ve bu riskleri bütünüyle değerlendirmeleri gerekmektedir. Böylece işletmeler, tedarik zincirinde meydana gelebilecek her türlü riske hazırlıklı olabilir ve korunabilirler. İşletmelerin istenen tedarik zinciri performansına ulaşamama nedeni tedarik zincirinde oluşan başarısızlığın artmasıyla meydana gelmektedir. Oluşabilecek her türlü riskten kaçınmak mümkün olmayacağı için bu risklerin yönetilmesi adına tedarikçilerin, üreticilerin, dağıtıcıların yani bu süreçte rol oynayan her bir birimin birbiriyle bütün olarak çalışması gerekmektedir. Risk yönetim süreci dörde ayrılmaktadır. İlk olarak oluşabilecek tedarik zinciri risklerinin tespit edilmesi ve sıralanmasıdır. İkinci olarak risklerin ortaya çıkış sebeplerinin değerlendirilmesidir. Üçüncü olarak risklerin büyüklüklerine göre öncelik vermek ve işletmeye gelebilecek zararların önlenebilmesi adına stratejiler oluşturmaktır. Son olarak ise artan veya azalan gelişmelere bağlı olarak risklerin incelenmesidir.

Tedarikçinin iflas etmesi de işletme açısından bir risk oluşturmaktadır. Oluşabilecek bu riski sıfırlamak mümkün olmasa da en aza indirgemek için güvenilir tedarikçiler ile anlaşma yapmak hatta ürünleri birden çok yerden almak doğru olacaktır. Herhangi bir tedarikçinin veya servis veren firmanın yurt dışında olması da risk barındırır. Kültürel farklar veya yasaların farklı olması gibi etkenler tedarik zincirinde risk oluşturmaktadır.

Bu riski en aza indirgemenin yollarından biri ise kültür farklılıklarına ve yasalara hakim olmaktır. Deprem, sel, doğal afetler, yangın gibi etkenler de tedarik zinciri açısından hem öngörülmesi en zor hem de oluşabilecek zarar en büyük olan risklerdendir. Bu riski minimize etmek için de işletmeler, yaptığı işe dair tuttuğu kayıtları sürekli olarak yedeklemelidir. Öngörülemeyen ve aniden değişebilen talepler de risk oluşturmaktadır. Hangi sektörde olursa olsun tüketicinin okuduğu kitapların, tatile gittiği yerlerin, dinlediği müziklerin yakından takip edilmesiyle tüketicilerin beklentilerini öngörmek mümkün olacaktır. Çünkü bu gelişmeler tüketicilerin beklentilerini şekillendirmektedir. Bundan yola çıkarak bir sonraki ayın veya yılın talep oranlarını tahmin edebilmek kolaylaşacaktır. Müşteri beklentilerini sürekli olarak takip eden işletmeler diğer işletmelere göre daha avantajlı olacaktır. Kullanılan cihazlarda oluşan bozulmalar veya nakliye sırasında yolların kapanması gibi oluşabilecek riskler göz önüne alındığında kullanılan cihazların bakımının yapılması ve güvenilir nakliye firmaları ile anlaşma yapılmasıyla oluşan krizler önlenebilir. İşçilik maliyetleri, fiyat dalgalanmaları, talep artması veya azalması gibi ekonomik koşullar gereği riskler oluşabilmektedir. Bu bağlamda oluşabilecek riskleri minimize etmenin yollarından biri ise ekonomiye hakim olmaktır. Bu tür gelişmelerin, yapılan işleri ne tür etkileyeceği konusunda öngörülerde bulunmak oluşan riskleri azaltmaktadır. Ürünlerin sayısının denetlenmemesi veya kalite denetiminin sağlanmaması, tedarikçi seçiminde gereken incelemeyi yapmamak, tüketicinin değişen beklentileri üzerine hala aynı ürünleri sunmak tedarik zinciri açısından risk oluşturmaktadır. Sürekli takip ve kontrollerin sağlanmasıyla oluşabilecek bu riskler azaltılabilmektedir. Taşıma kapsamında; eski ve güvenilmez teknoloji kullanımı, gümrükteki gecikmeler, sınırlı kapasite, yüksek taşıma maliyetleri, çalışma sürelerinin uzun tutulması, ulaştırma alternatiflerinin olmaması, liman kapasitesi ve liman karayolu sıkışıklıkları, ürünlerin kaybolması, taşıma esnasında ürünlerin zarar görmesi gibi risklerin oluşması tedarik zincirinde karmaşıklığa yol açabilir. Tedarik zincirinde sözleşme üzerindeki anlaşmazlıklar, teslimat planının aksaması, maliyetlerdeki ani artışlar, tedarikçinin stok yetersizliği ile oluşan yanlış tedarikçi seçimi, sipariş karşılama hataları ve ara tedarikçi sayının az olması da tedarik zinciri üzerinde belirli riskler oluşturmaktadır. Sistemdeki arızalar, bilgi hataları ve bilgi güvenliğinin ihlali, haberleşme kesintileri, veri kaybı gibi risklerin oluşabilmesi etkin tedarik zinciri yönetiminde bütünleşme adına önemlidir. Tedarik riskleri, tedarikçilerden ya da arz pazarından kaynaklanan sorunların meydana gelme olasılıklarıyla ilgilidir (Manuj & Mentzer, 2008: 138). Tedarik riskleri içerisinde yer alan tedarik edilen malzemelerde

fiyat artışı riski görüşülen yöneticilerin çoğu tarafından işletmeyi tehdit edebilecek bir risk faktörü olarak değerlendirilmemiştir. Hatta tedarikçilerle yapılan anlaşmalar gereği aynı parçayı sağlayan tedarikçinin zamanla uzmanlaşacağı ve maliyetlerini azaltması gerektiği düşüncesiyle tedarikçilerden fiyat indirimlerinin istendiği tedarik zinciri yöneticilerinin ortak ifadelerinden biri olmuştur[3]. Genel olarak işletmelerin, gerçek hayat problemlerinden dolayı yaşanabilecek riskleri kendi tedbirleriyle ortadan kaldırmaları dışsal faktörler içerdiğinden mümkün olmadığı için hafifletici stratejiler benimseyerek oluşabilecek riskleri en aza indirgemeyi planlayarak olası risklerin etkilerini hafifletmeye çalışabilirler.

3
MAKİNE ÖĞRENMESİ

3.1 Makine Öğrenmesi

Makine öğrenmesi, bireylerin öğrenme biçimlerini benzetmek için veri ve algoritmaların kullanımına odaklanarak, doğruluğunu paralel olarak yükselten bilgisayar bilimi dalıdır. İstatistiksel yöntemler kullanılarak, algoritmalar; sınıflandırmalar veya tahminler yapmak üzere eğitilir. Makine öğrenmesinde yer alan süreçler, veri madenciliği ve tahminleme modellerine benzerdir. Makine öğrenmesi teknikleri genel olarak tahmin etmek ya da sınıflandırmak için kullanılmaktadır. Algoritmada etiketli ya da etiketsiz olarak verilerdeki belirli kalıplarla ilgili tahmin üretilecektir. Tahminlerin değerlendirilmesi amacıyla hata işlevi kullanılır ve bilindik örneklerle modelin doğruluğunu değerlendirmek için karşılaştırma yapılabilir. Model, eğitim kümesine daha iyi uyum sağlıyorsa bilinen örnek ve tahmin arasında düzenleme yapılır. Algoritma da bu değerlendirmeyi tekrarlayıp gelişen süreci eniyileyerek doğruluk eşiğine ulaşana kadar güncellenir.

3.2 Makine Öğrenmesi Algoritma Yöntemleri

Makine öğrenmesi algoritmaları, bir veri setini modele dönüştüren algoritmalardır. Denetlenen , denetlenemeyen, sınıflandırma, regresyon vb. hangi tür algoritmanın en iyi şekilde çalıştığı ve verilerin niteliklerine göre değişim göstermektedir. Makine öğrenmesi algoritmalarının sürekli biçimde yeni verilerle karşı karşıya bırakılmasıyla, sınıflandırma, veri analizi ve tahminlemeye dair modelleme konusunda büyük görevlere olanak tanır. Makine öğrenmesi belli bir düzen veya zaman içinde tekrarlanan verilerdeki kalıpları bulmaya çalışmak için gözlemlerin seçilmiş olan algoritmaya girilmesiyle başlamaktadır. Algoritmanın çalıştığını doğrulamak için yeni veriler algoritmaya eklenir. Daha sonra da tahmin ve doğrular kontrol edilir. Makinenin daha fazla veriyle beslenmesi elde edilen sonuçların iyileştirilmesi anlamına gelmektedir. Oluşan tahmin beklenen gibi olmadığında, beklenen çıktı bulunana kadar birçok kez yeniden eğitilir.

Algoritmanın zamanla doğruluğunun artmasıyla en uygun cevabın bulunmasını sağlamaktadır. Makine öğrenmesi algoritmaları, her birinin farklı bir amacı olan kategorilere ayrılmaktadır.

3.2.1 Denetimli Öğrenme

Bağımsız değişkenlerin etki ettiği bağımlı değişkenin varlığına ve bağımsız değişkenleri eğiterek çıkan sonuca referans veren bir yöntemdir. Girdi (x), çıktı (y) değişkenlerine eşlemek için etiketli eğitim verileri kullanılmaktadır. Girdi ve çıktı değişkenleri arasında her zaman bir ilişki bulunur ve makine öğrenmesinin amacı bu ilişkiyi ölçmektir. Denetimli öğrenme problemleri sınıflandırma ve regresyon olmak üzere iki tür olabilir.

3.2.2 Denetimsiz Öğrenme

Denetimsiz öğrenme tekniğinde, algoritma verilerin bizzat kendisinin yararlanıldığı anlamlı sonuçlar elde etmeye çalışmaktadır. Aynı sınıf altında benzer verileri gruplara ayırmaktadır. Eğitim için kullanılan bilgilerin gruplandırılmadığı durumlarda veya etiketlenmediği zamanlarda kullanılır. Doğru çıktı bulunmadığında verileri araştırır ve veri kümelerinden tahminlemeler yapar.

3.2.3 Yarı Denetimli Öğrenme

Yarı denetimli makine öğrenmesi , hem denetlenen hem de denetlenemeyen algoritmaların özelliklerini birleştiren algoritmalardır. Eğitmek amacıyla etiketli ve etiketlenmemiş verileri kullanmaktadır. Ancak eğitim verileri etiketlenmemiş ya da algoritma başlangıcında bütün kurallar elde edilmemiş olabilir.

3.3 Denetimli Öğrenme Yöntemleri

Makine öğrenmesi, eldeki veriden daha anlamlı bir veri elde etmek için kullanılır. Veri setinin içinde sonuçlar da varsa ve bu verilere göre oluşturduğumuz modelde yeni veriler tahmin ediyorsak bu denetimli öğrenmedir.

Denetimli öğrenme problemleri iki şekildedir;

1.Regresyon

2.Sınıflandırma

Regresyonda sayısal çözümleme uygulanır ve sayısal değerler arasındaki ilişkiye göre fonksiyon oluşturulmaktadır. Oluşan verilerle sonuç tahmin edilir. Veriler görselleştirildiğinde aralarındaki ilişkinin eğrisini çizmek olarak da düşünülebilir.

Regresyondaki algoritmalar sırasıyla;

•Doğrusal Regresyon (Linear Regression)

•Çoklu Doğrusal Regresyon (Multilinear Regression)

•Polinomsal Regresyon (Polynominal Regression)

•Lojistik Regresyon (Logistic Regression)

•Karar Ağacı Regresyonu (Decision Tree Regression)

•Rastgele Ormanlar Regresyonu (Random Forest Regression)

•Destek Vektör Regresyonu (Support Vector Regression)

Sınıflandırma, verilerin kategorilere ayrılmasıdır. Veriler etiketlenebilir, kategorilere ayrılabilirse sınıflandırma tekniği kullanılabilir.

Sınıflandırma algoritmaları;

•K - En Yakın Komşu (KNN/K-nearest Neighborhood)

•Naif Bayes Sınıflandırması (Naive Bayes Classifer)

•Karar Ağacı Sınıflandırması (Decision Tree Classifer)

3.3.1 Karar Ağacı Sınıflandırması

Ağaç tabanlı makine öğrenmesi algoritmaları, en çok tercih edilen denetimli öğrenme algoritmalarındandır. Genel itibariyle bütün sınıflandırma ve regresyon problemlerinin çözümüne uyarlanabilirler. Karar ağaçları, gradyen güçlendirme (gradient boosting) ve rasgele ormanlar gibi yöntemler, her çeşit veri bilimi problemlerinde sıklıkla kullanılmaktadır.

Karar Ağaçları, tıpkı ağaç şeklindeki grafikler gibidir. Karar ağacı düğümler ve dallardan oluşur. Düğümler yine iki tiptedir. Yani şans düğümü veya karar düğümleri. Şans düğümleri, karar vericinin elinde hangi alternatiflerin olduğunu gösterirken, karar düğümünde karar vericilerin bazı kararlar alması gerekir. Şans düğümlerinden çıkan dallar çeşitli doğa durumlarını gösterir ve olasılıklar şans dallarıyla ilişkilendirilirken, karar düğümlerinden çıkan dallar eldeki çeşitli alternatifleri gösterir. Karar Ağaçlarının

iki ana türü vardır; sınıflandırma ağaçları ve regresyon ağaçları. Sınıflandırma ağacında, değişken doğası gereği kategoriktir, regresyonda değişken doğası gereği süreklidir.[7]

Karar ağacı yapısı birçok veri kümesinin birtakım kurallar uygulanarak daha küçük kümelere bölünmesi amacıyla kullanılmaktadır. Önceden tanımlanmış olan hedef değişkene sahiptirler. Yapıları itibariyle en yukarıdan en aşağıya inen stratejileri bulunmaktadır. Bir karar ağacı, çok sayıda kayıt içeren bir veri kümesini, birtakım karar kurallarına göre daha küçük kümelere bölme amacıyla kullanılır. Karar ağaçlarında ağacın doğruluğunu etkileyen faktörlerden bir tanesi de bölünmenin nasıl gerçekleşeceğidir. Bütün regresyon ve sınıflandırma problemleri için bölünme kriterleri farklıdır. Karar ağacı algoritması, bir düğümü iki ya da daha fazla alt düğümde bölünmesine karar vermek için birden fazla algoritma kullanılır. Alt düğümlerin oluşturulması, oluşan bu alt düğümlerin homojenliğini artırmaktadır. Karar ağacı tekniğinde en yaygın kullanılan algoritmalar; sürekli değişkenler için En Küçük Kareler Yöntemi (EKK), kategorik değişkenler için Entropi, Gini ve Sınıflandırma Hatasıdır. Entropi, var olan verilerle alakalı belirsizliğin ölçüsüdür. Veri kümesinin sadece bir etiketi varsa, daha düşük bir entropiye sahiptir diyebiliriz. Bu sebeple var olan verilerin, entropiyi minimuma indirgeyecek şekilde bölünmesi gerekir. Bölünmelerin iyi olması, paralel olarak tahminlerin de iyi olması anlamına gelmektedir.

Entropi denklemi eşitlik (3.1)'de gösterilmiştir.[20]

$$H = -\sum p(x) \log p(x) \tag{3.1}$$

p(x), belirli bir sınıfın grup yüzdesini belirtirken, H ise entropiyi ifade etmektedir. Karar ağacının entropi değerini minimuma indirgeyen bölünmelerin yapılması gerekli olduğundan en iyi bölünmenin belirlenebilmesi için bilgi kazancı kullanılmaktadır ve eşitlik (3.2) ile ifade edilmektedir. [20]

$$\mathrm{Gain(S, D)} = \mathrm{H(S)} - \sum_{\mathrm{V}\epsilon\,\mathrm{D}} \frac{|\mathrm{V}|}{|\mathrm{S}|} \mathrm{H(V)} \tag{3.2}$$

Bilgi kazancı denkleminde S orijinal veri kümesiyken D, kümenin bölünmüş bir parçasıdır. Her V, S'nin bir alt kümesidir. V'nin tümü ayrıktır ve S'yi oluşturmaktadır. Dolayısıyla bilgi kazancı, her bir niteliğin entropisi ile bölünme yapılmadan önceki verinin entropi değeri arasındaki fark olarak ifade edilmektedir. Karar ağacı yönteminin

kullanıldığı sınıflandırma problemlerinde kullanılan veri seti, eğitim ve test olarak iki parçaya ayrılmalıdır. Karar ağacı algoritması, eğitim verilerini kullanarak bir model oluşturmaktadır. Modelin problem çözümündeki başarısının hesaplanabilmesi için de oluşturulan model test veri seti üzerinde uygulanır.

En popüler karar ağacı algoritmaları ; ID3, C4.5**,** CHAID**,** CART

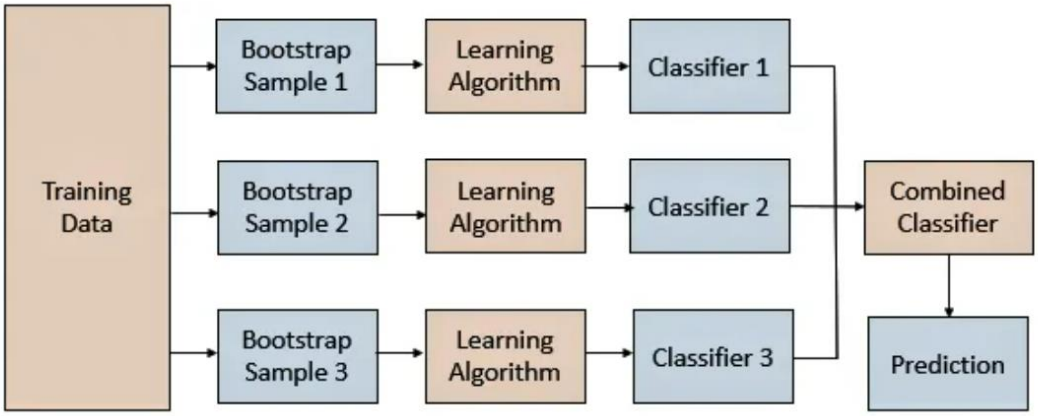
3.3.2 Topluluk Öğrenme Yöntemleri

Kümeleme, sınıflandırma ve regresyon problemlerinde bir öğrenme algoritmasının tahmin doğruluğunu ve kararlılığını artırabilmek için kullanılır. Topluluk öğrenmesi yöntemleriyle, öğrenme algoritmalarının hatalarını azaltılmasıyla ve doğru sınıflandırma oranını artırılmasıyla daha yüksek performansa sahip tahmin yapabilen modellerin oluşturulması amaçlanmaktadır.

3.3.2.1 Torbalama Algoritması (Bootstrap Aggregation-Bagging)

Torbalama algoritmasında, öğreniciler eğitim setinden rasgele seçilmiş olan farklı alt kümelerle eğitilmektedir. Öncelikle veri seti , eğitim seti ve test seti olarak ikiye ayrılır ve bu aşamadan sonra eğitim için ayrılan veriden rasgele seçim yapılır. Örneğin torbadan çekilmiş topun torbaya geri atılması gibi seçilenler tekrar seçilebilecek biçimde eğitim kümesinde kalır. Sayı genelde %60 olup eğitim için ayrılan sayıdan fazla olmamaktadır. Model, farklı eğitim setiyle oluşunca karar farklılıkları oluşacağından başarıyı yükseltmek amaçlandığı için farklı eğitim setleri seçilmektedir. Kararlar ağırlıklı oylama ile birleştirilir. Tahmincilerin birleştirilmesi basamağında regresyon problemlerinde ortalama alınır. Sınıflandırma problemlerinde de sonuçlar oylamayla belirlenmektedir.

Torbalama, tutarsız olan tahminci değişkenin tahminleme işlerliğini de artırabilir. Düşük yanlılığa sahip ancak yüksek varyanslı olan değişkenlerle bu değişkenleri daha kullanılabilir hale getirir. Neticede bu yöntem, tekil ağaçlara oranla daha iyi sonuçlar vermektedir.



Şekil 3.1 Torbalama (Bagging) Algoritması [25]

•Orjinal veri kümesinden çoklu alt kümeler oluşturulur.

•Bu alt grupların her birinde bir zayıf model yani temel model oluşturulmuştur.

•Modeller paralel olarak çalışır ve birbirinden bağımsızdır.

•Nihai tahminler, tüm modellerden gelen tahminler birleştirilerek belirlenir.

3.3.2.2 Rasgele Ormanlar (Random Forests) Algoritması

Rasgele Ormanlar Algoritması, denetimli öğrenme algoritmasıdır. Birden fazla karar ağacının farklı gözlem örnekleri üstünde eğitilerek, modeller oluşturulmasıyla sınıflandırma yapılmasıdır. Bu algoritma adından da anlaşılacağı üzere rasgele orman yaratmaktadır.

Yöntemdeki ağaç sayısı ve bulunan sonuç arasında bir ilişki bulunmaktadır. Sonucun kesinleşmesiyle ağaç sayısının artması doğru orantılıdır. Sonuçları olumsuz etkileyen Aşırı uyum (Overfitting) sorununda, yeteri kadar ağacın ormanda yer alması bu durumla karşı karşıya kalınma oranını azalmaktadır. Hem sınıflandırma hem regresyon görevlerinde kullanılmasıyla beraber bu algoritma, sınıflandırıcı eksik olan değerleri ele alabilir ve sınıflandırıcı kategorik değerler için modellenebilir.

Rasgele Orman Yönteminde iki basamak bulunmaktadır. Bunlardan birincisi Rasgele Orman oluşturulmasıyken ikincisi de ilk basamakta oluşturulan Rasgele Orman sınıflandırıcısından tahminleme yapmaktır.

3.3.2.3 Güçlendirme Algoritması (Boosting)

Güçlendirme Algoritması (Boosting), Topluluk Öğrenmesi (Bagging) ve Rasgele Ormanlar (Random Forests) gibi tahmin performansını artırmaya yönelik bir yaklaşımdır. Bunlardan farklı olarak bu algoritma her adımında bir önceki adımdaki performansı geliştirmeye çalışır. Bagging ve Random Forests'ta her bir bootstrap training seti üzerine kurulabilecek en büyük ağaçlar kurulur, tahminlerin yanlılığı azaltılır ve ortalama alınarak tahminlerin varyansı indirgenmiş olur. Güçlendirme Algoritması (Boosting)'de ise her adımda daha küçük bir ağaç kurulur ve o ağaçtan elde edilen bilgi bir sonraki adıma aktarılarak öğrenme yavaş gerçekleşir. Böylece Overfitting (Aşırı Öğrenme) oluşmaz. Güçlendirme (Boosting) algoritmasındaki amaç, veri setine farklı ağırlıklar verilmesi sonucunda oluşan ağaçlar topluluğundan çıkarımlar üretmektir. Başta bütün gözlemler eşit ağırlığa sahiptir. Farklı ağırlıkların verilmesiyle oluşan ağaç topluluğu büyümeye başladığında, problem bilgisiyle ilişkili ağırlıklandırmalar düzenlenmektedir. Yanlış sınıflandırılmış olan gözlemlerin ağırlığı artırılır. Nadiren yanlış sınıflandırılmış olan gözlemlerin ağırlığı da azaltılır. Böylece ağaçlar güç bir durumla karşı karşıya kaldığında kendini düzenleyebilme becerisi kazanır.

Algoritma şu şekilde çalışır;

•Asıl veri kümesinden alt veri kümesi oluşturulur.

•İlk olarak tüm veri noktalarına eşit ağırlık verilir.

•Oluşturulan alt veri kümesinde temel model oluşturulur.

•Veri kümesinin tümünde çıkarsamalar yapmak için oluşturulan model kullanılır.

•Hata, gerçek değerlerle öngörülen değerlerin kullanılmasıyla hesaplanır.

•Hatalı öngörülen değerlere daha fazla ağırlık verilir.

•Başka model oluşturularak veri setinde tahminlemeler yapılır. (Bu model, önceki modeldeki hataları düzeltmeye çalışır)

•Her biri bir önceki modeldeki hataları azaltan çoklu modeller oluşturulur.

•Güçlü öğrenen yani son model, tüm modellerin ağırlıklı ortalamasıdır.

Güçlendirme Algoritmaları; Ada Boost, Gradient Boost ve Extreme Gradient Boost ve CatBoost.

•AdaBoost (Adaptive Boosting)

Adaboost algoritması, Freund ve Schapire (1997) tarafından önerilmiş olup, çeşitli zayıf algoritmaların bir araya getirilerek karma ve daha güçlü bir eğitim modeli sunan bir makine öğrenmesinin gerçekleştirilmesi amaçlanmaktadır. İçerisinde kendisinden daha önce geliştirilmiş olan farklı makine öğrenmesi algoritmalarını birleştirerek ya da karşılaştırarak öğrenme konusunda en iyi sonucun elde edilmesi hedeflenmektedir (Freud ve Schapire, 1997)[14]. Eğitim aşaması, sadece modelin tahminler üretme gücünün iyileştirildiği özellikleri seçer, boyutu azaltır ve potansiyel olarak yürütme süresini iyileştirir, çünkü alakasız özelliklerin hesaplanması gerekmemektedir.

$$F_t(x) = \sum_{t=1}^{T} f_t\,(x) \tag{3.3}$$

Eşitlik (3.3)'te gösterilen her biri f_t bir nesneyi alan zayıf bir öğreniciyken x, nesnenin sınıfını gösteren girdi olarak bir değer gösterir.[23] İki sınıflı problemlerde, zayıf öğrenici çıktısının işareti tahminlenen nesne sınıfını tanımlar. Aynı zamanda T. sınıflandırıcı, pozitif bir sınıftaysa pozitif, değilse negatiftir. $h(x_i)$, eğitim setindeki her örnek için zayıf öğreniciler çıktı varsayımı üretirler. Tüm tekrar aşamalarında t, zayıf bir öğrenici ve katsayı seçilir. α_t öyle ki toplam eğitim hatası E_t sonuçtan t-stage boost sınıflandırıcı minimize edilmiştir ve eşitlik (3.4) ile ifade edilmektedir.[23]

$$E_t = \sum_{i} E[F_{t-1}\,(x_i) + \alpha_t h(x_i)] \tag{3.4}$$

F_{t-1}(x) eğitimin önceki aşamasına kadar oluşturulmuş güçlendirilmiş sınıflandırıcıdır, E(F) bazı hata fonksiyonudur ve $f_t(x) = \alpha_t h(x)$ son sınıflandırıcıya eklenmesi göz önünde bulundurulan zayıf öğrenicidir.

Ağırlıklandırma;

$E(F_{t-1}(x_i))$ o örnek üzerindeki eğitim aşamasının her tekrarında, ağırlık ($w_{i,t}$) eğitim setindeki her örneğe halihazırdaki hataya eşit olarak seçilmektedir. Seçilen ağırlıklar, zayıf öğrenicinin eğitimini bilgilendirmek için kullanılabilir. Söz gelimi yüksek ağırlıklı numune kümelerini ayırmayı tercih eden karar ağaçları yetiştirilebilir.

•Gradient Boosting (Gradyan Güçlendirme)

Gradyan güçlendirme, sınıflandırma ve regresyon problemlerinde kullanılır. Zayıf tahmin modellerinin bir araya gelmesiyle karar ağaçlarından oluşan bir model meydana getirir. Gözden geçirilen bir öğrenme algoritması, kayıp fonksiyonu tanımlama ve en aza indirme amacıyla çalışır.

Gradyan güçlendirme modeli için uygulama adımları aşağıdaki gibi sıralanmıştır;

• Regresyon ağacı oluşturulur.

• Her bir ağacın yaprağı için hata oranı hesaplanır.

• İkinci adımda hesaplanan hata oranları yeni gözlem verileri olarak kullanılır.

• İkinci ağaç, ikinci adımdaki hata oranlarını en aza indirgemeyi hedefler.

• İkinci ve dördüncü adımlar tekrar edilir.

Eşitlik (3.5)'teki karesel hataya (MSE) sahip olunduğu farzedildiğinde:

$$Loss = MSE = \sum (y_i - y_i^p)^2 \tag{3.5}$$

Burada y_i hedef değer, y_i^p tahmin, $L(y_i, y_i^p)$ kayıp fonksiyonudur.[22]

Kayıp fonksiyonunun minimum düzeyde olmasıyla tahminler istenmektedir. Tahminlerin öğrenme oranına bağlı olarak güncellenmesiyle, MSE'nin istenen minimum düzeyde olduğu değerleri eşitlik (3.6) ve (3.7)'deki gibi bulunabilir.[22]

$$y_i^p = y_i^p + \alpha * \frac{\delta \sum (y_i - y_i^p)^2}{\delta y_i^p} \tag{3.6}$$

$$y_i^p = y_i^p - \alpha * 2 * \sum (y_i - y_i^p) \tag{3.7}$$

α öğrenme oranıdır ve $\sum (y_i - y_i^p)$ artıkların toplamıdır.

Tahminler, artıklar toplamının minimum ya da 0'a yakın olacağı şekilde güncellenir. Ayrıca tahmin edilen değerler gerçek değerlere yakın olmalıdır.

•Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting, karar ağacı (decison-tree) temeline dayanan ve eğim artırmalı (gradient-boosting) makine öğrenmesi sistemidir. Güçlü tahmin öngörüsüne sahip ve daha hızlıdır. Performansı iyileştiren ve aşırı uyumu (Overfitting) küçülten bir dizi

düzenlileştirme (regularization) içermektedir. Gradient Boosting ve XGBoost arasındaki temel fark uygulama detaylarına dayanmaktadır. XGBoost, farklı düzenlileştirme teknikleriyle ağaçların karmaşıklığını kontrol ederek daha güçlü performansa sahip olmaktadır.

XGBoost algoritmasında, modelin performansını ölçmek ve eğitim setinde en iyi parametreleri bulmak için amaç fonksiyonu tanımlanmaktadır ve eşitlik (3.8) ile ifade edilir.[20]

$$\mathrm{Obj}(\emptyset) = \mathrm{L}(\mathrm{W}) + \Omega(\emptyset) \tag{3.8}$$

L, eğitimin kayıp fonksiyonudur ve modelin eğitim verilerindeki öngörüsünü kapsamaktadır. Ayrıca kayıp fonksiyonunun en aza indirilmesi gerekmektedir. Ω, düzenlileştirme (regularization) olarak ifade edilmektedir. Aşırı öğrenme veya aşırı uyumu azaltır ve modelin karmaşıklığını kontrol eder.

Eğitim kaybının ortalama kare hatası eşitlik (3.9)'daki gibi ifade edilmektedir.[20]

$$\mathrm{L}(\mathrm{W}) = \sum_{\mathrm{i}} (\mathrm{y_i} - \bar{\mathrm{y}}_{\mathrm{i}})^2 \tag{3.9}$$

Lojistik regresyon için kayıp fonksiyonu eşitlik (3.10)'deki gibidir.[21]

$$\mathrm{L}(\mathrm{W}) = \sum_{\mathrm{i}} [\mathrm{y_i} \ln(1 + \mathrm{e}^{-\bar{\mathrm{y}}_{\mathrm{i}}}) + (1 - \mathrm{y_i}) \ln(1 + \mathrm{e}^{\bar{\mathrm{y}}_{\mathrm{i}}})] \tag{3.10}$$

Parametre ayarı, XGBoost modeli eğitilirken en iyi sonuçlar elde etmek için önemli bir faktördür. Model iki adımdan oluşmaktadır. Birincisi XGBoost modelinin ne kadar iyi performans gösterdiğini gözlemlemek için temel bir model geliştirir. İkincisi de sonuçları oluşturulan bu temel modelle karşılaştırıp parametre ayarlama kullanarak eğitilmesidir. Parametre ayarlamadaki yaklaşım, eğitimi hızlandırmak için büyük ölçüde öğrenme oranı kullanır. Sadece ağaca mahsus parametreleri kullanılır ve öğrenme oranı düşürülür. En uygun sınıflandırıcının değerini bulmak için eğitilir ve işlem son bulur.

XGBoost, düzenlileştirmeye (regularization) sahip olduğundan aşırı uyumun indirgenmesini sağlar. Yöntem, budama işlemi yaparken ayarlanan max_depth değerine kadar bölme işlemini gerçekleştirir. Sonra geriye doğru budamaya başlar. Pozitif katkısı bulunmayan bölümleri yok eder. XGBoost modelinde artırma işleminin her tekrarında çapraz doğrulamaya izin verir. Bundan dolayı tek çalıştırma biçiminde uygun artırma tekrar sayısı daha kolay bulunmaktadır.

XGBoost için kullanılan parametreler;

•Eta: Öğrenme büyüklüğünü kontrol etmektedir. Değeri 0 ve 1 arasında yer almaktadır. 0.01 ile 0.3 arasında daha iyi sonuçlar verir.

•Max_depth: Ağacın derinliğini kontrol etmektedir.

•Nround: Gereken maximum iterasyon sayısını kontrol etmektedir.

•Ntread: Paralel hesaplamayı çalışır hale getirmektedir.

•Gamma: Aşırı uyumu engellemektedir.

•Subsample : Her ağacın eğitimi için kullanılan eğitim örneklerinden rasgele belirlenen kısımdır.

•Colsample_bytree: Her ağacın eğitimi için rasgele belirlenen özellikler kısmıdır. Her ağaç oluşturulurken sütunların alt örnek oranını kapsamaktadır. Oluşturulan her ağaç için yalnızca bir defa yapılmaktadır. Bir ağaç için belirlenen değişkenlerin miktarını kontrol etmektedir. 0.5-0.9 arasında daha iyi sonuçlar vermektedir.

•Colsample_bynode: Her düğümdeki sütunların alt örnek büyüklüğünü kapsamaktadır. Her yeni bölünmede yeniden değerlendirilmektedir.

•CatBoost (Categorical Boosting)

Gradient Bossting'in performansını artırmak amacıyla ve XGBoost'a başka bir seçenek olarak 2017'de "CatBoost: unbiased boosting with categorical features" makalesiyle tanıtılmıştır. "**Cat**egory" ve "**Boost**ing" sözcüklerinin birleşimiyle ortaya çıkmıştır.



Şekil 3.2 Güçlendirme algoritmaları tarihçesi [19]

Güçlü öğrenme hızı, sayısal, ketegorik ve metin verileriyle çalışılabilmesi, GPU (Grafik işlem birimi) yardımıyla görselleştirme alternatifleri sunması diğer yöntemlerden ayırt edilen nitelikleridir. CatBoost, veri hazırlığı sürecinin kısaltılması önemli yetkiye sahiptir. Anlamsız verilerle baş edebilir, kategorik yapılı verilere de kodlama (encoding) yapar. Kategorik verilerle aşırı verimli çalışabilmesinin sebebi kendine özgü kodlama

yöntemi olmasıdır. Veri hazırlığı yapılma aşamasında ekstra kodlama işlemi lazım değildir. Hem öğrenme hızı hem de sonuçların kalitesi etkileneceği için kodlama işleminin olmaması özellikle önerilir. İlaveten simetrik ağaçlar kurmasıyla çok derin ağaçlar kurmadan yüksek öngörü büyüklüğünü yakalar ve aşırı uyum sorununu aşmaktadır. Şayet aşırı uyum durumu ortaya çıkarsa yöntem parametrelerde belirtilen niteliklere ulaşmadan önce öğrenmeyi durdurur. Bu seçenek başlangıç parametrelerinden ayarlanmaktadır.

3.3.3 Lojistik Regresyon

Lojistik regresyon, bağımlı değişkenin kategorik yapıda olduğu regresyon problemi gibidir. Doğrusal sınıflandırmada kullanılmaktadır. Sonucu belirleyen bir veya daha fazla bağımsız değişken bulunan veri setini çözümlemek için tercih edilen bir tekniktir. Yalnızca iki olası sonuç vardır ve sonuç ikili bir değişkenle ölçülür. Lojistik regresyon, yanıt veya sonuç değişkeniyle alakalı öngörücü veya açıklayıcı değişken arasındaki ilişkiyi ifade etmek amacıyla en uygun modeli bulmaya yardımcı olmaktadır. Lojistik regresyon analizi, bağımlı değişkenin yapısına göre üç tiptir. Bunlardan birincisi kategorik bağımlı değişkenin ikili olduğu durumlarda kullanılan İkili Lojistik Regresyon'dur. Bir diğeri ise kategorik bağımlı değişkenin çok kategorili olduğu durumlarda kullanılan Multinominal Lojistik Regresyon'dur. Son olarak çok kategorili ve sıralı yapıda olduğunda Ordinal (Sıralı) Lojistik Regresyon kullanılmaktadır. Lojistik regresyon bağımsız değişken sayısına bağlı olarak da sınıflandırılabilir.

Tablo 3.1 Lojistik regresyonun uygulandığı durumlar[32]

Bağımlı Değişken Kategori Sayısı	Bağımsız Değişken Sayısı	Bağımsız Değişkenin Kategori Sayısı	Uygulanacak Yöntem
2	1	2	Binominal Lojistik Regresyon
2	1	2+	Binominal Lojistik Regresyon
2	2+	Çeşitli	Çok Değişkenli Lojistik Regresyon
2 + sırasız	Tek/Çok	Çeşitli	Multinominal Lojistik Regresyon
2 + sıralı	Tek/Çok	Çeşitli	Ordinal Lojistik Regresyon

Tek bağımsız değişken olduğu durumlarda Tek Değişkenli Lojistik Regresyon kullanılırken iki veya daha fazla değişken olması durumunda Çoklu Lojistik Regresyon kullanılmaktadır.

Lojistik regresyon analizinde doğrusal regresyon analizindeki varsayımların aranmaması sebebiyle daha fazla tercih edilmektedir. Ancak lojistik regresyon analizinin uygulanabilmesi için bazı varsayım ve gereksinimler bulunmaktadır.

• Bazı değişkenlerin modele dahil edilmemesi model yetersizliğine ve hata teriminin büyümesine yol açmaktadır bu yüzden uygun tüm bağımsız değişkenler modele dahil edilmelidir.

• Bazı bağımsız değişkenlerin bağımlı değişken üzerinde etkisi olmamasına rağmen modele dahil edilmesi modelin yorumlanmasını zorlaştırabilir. Bu nedenle

uygun olmayan tüm bağımsız değişkenler modelden çıkarılmalıdır.

• Eksik (kayıp) veri olmamalıdır. Aynı zamanda hatalar katsayı tahminlerinde yanlılığa ve model yetersizliğine sebep olabileceğinden bağımsız değişkenlerde ölçüm hataları küçük olmalıdır.

• Bağımsız değişkenler birbirleriyle ilişkisiz yani bağımsız değişkenler arasında Çoklu Bağlantı (Multicollinearity) olmamalıdır.

• Doğrusal regresyonda olduğu gibi Lojistik Regresyonda da aşırı değerler sonucu etkileyebileceğinden modelde olmamalıdır.

• Örneklem büyüklüğü yeterli olmalıdır.

• Örneklemin rastgele yöntemle seçilmemiş olması gibi durumlarda bağımlı değişkenin beklenen varyansı ile gözlenen varyansı arasında önemli derecede fark varsa modelin yetersiz olduğu ve yeniden tanımlanması gerekmektedir.

Model oluştururken öncelikli amaç modeldeki değişken veya faktörlerin belirlenmesi olmalıdır. Çalışmanın amacına bağlı olarak model seçim stratejileri düzenlenebilmektedir. Bağımsız değişken sayısının fazla olması durumunda model sayısı da fazla olacaktır. Yani değişken seçimi iki veya daha fazla olduğu durumlarda modele önem kazandırmaktadır. Lojistik regresyon modeli stepwise (adımsal) ve standart olmak üzere iki çeşit yöntemle kullanılacak yöntem belirlenmelidir. Stepwise yöntemler de ileriye doğru ve geriye doğru olmak üzere ikiye ayrılmaktadır. Bir bağımsız değişkenin

diğer bağımsız değişkenden daha önemli olduğu düşünülmediği takdirde standart yöntem kullanılabilmektedir.

İkili (binary) bağımlı değişken; 0 ve 1 olmak üzere iki değer alır yani bağımlı değişken ikili bir yanıttır. İkili sonuç modelleri uygulamalı ekonomide en çok kullanılanlar arasında yer almaktadır.

OLS model eşitlik (3.11)'deki gibidir.

$$y = x'\beta + e \tag{3.11}$$

İkili sonuç modelleri, bağımsız değişkenlerin bir fonksiyonu olarak y=1 olasılığını tahmin etmektedir ve eşitlik (3.12)'de gösterilmiştir.[8]

$$p = pr[y = 1|x] = F(x'\beta) \tag{3.12}$$

$F(x'\beta)$ fonksiyonel formuna bağlı olarak üç farklı modeli vardır.

Doğrusal olasılık modelinde, $F(x'\beta) = x'\beta$

$$p = pr[y = 1|x] = x'\beta \tag{3.13}$$

Regresyon modeliyle ilgili bir sorun, tahmin edilen olasılıkların 0 ile 1 arasında sınırlı olmamasıdır.

Logit model için, $F(x'\beta)$ lojistik modelin kümülatif dağılım fonksiyonudur (cdf) ve eşitlik (3.14) ile hesaplanmaktadır.[4]

$$F(x'\beta) = \Lambda(x'\beta) = \frac{e^{x'\beta}}{1 + e^{x'\beta}} = \frac{exp(x'\beta)}{1 + exp(x'\beta)} \tag{3.14}$$

Tahmin edilen olasılıklar 0 ile 1 arasında sınırlıdır.

Probit model için, $F(x'\beta)$ standart normal dağılımın cdf'sidir ve eşitlik (3.15)'deki gibi ifade edilmektedir.[8]

$$F(x'\beta) = \Phi(x'\beta) = \int_{-\infty}^{x'\beta} \Phi(z)dz \tag{3.15}$$

Tahmin edilen olasılıklar 0 ile 1 arasında sınırlıdır.

Probit ve logit modelleri maksimum olabilirlik yöntemi kullanılarak tahmin edilmektedir.

Katsayıların yorumlanmasında;

• x'teki bir artış , y=1 olma olasılığını artırır/azaltır. (Bu sonucu daha fazla/daha az olası kılar). Başka bir deyişle x'teki bir artış, 1'in sonucunu daha fazla veya daha az olası kılar.

• Katsayıların büyüklüğünü değil işaretini yorumluyoruz. Farklı modellerin farklı katsayıları ölçekleri olduğundan, büyüklük katsayı kullanarak yorumlanmaz.

F fonksiyonunun fonksiyonel formu nedeniyle katsayılar modeller arasında farklılık gösterir. Farklı modeller arasında katsayıların büyüklüğünü karşılaştırmamalıyız.

Probit ve logit modeller tahmin edilirken, katsayıların raporlanmasından sonra marjinal etkiler raporlanır. Marjinal etkiler, bağımsız bir x değişkeninde 1 birimlik bir değişiklik verildiğinde, y=1 olasılığındaki değişikliği yansıtır.

Regresyon modeli için marjinal etkiler eşitlik (3.16)'da gösterilmiştir.[8]

OLS regresyon modeli için marjinal etkiler katsayılardır ve x'e bağlı değildir.

$$\frac{\partial p}{\partial x_j} = \beta_j \tag{3.16}$$

Eşitlik (3.16)'da j indisi j. bağımsız değişkeni ifade eder. İ indeksini kullandığımızda, i. gözlemi ifade eder.

İkili modeller için marjinal etkiler (probit ve logit);

Logit ve probit modeller için marjinal etkiler eşitlik (3.17) ile ifade edilmektedir.[8]

$$\frac{\partial p}{\partial x_j} = F'(x'\beta)\beta_j \tag{3.17}$$

Marjinal etkiler x'e bağlıdır bu nedenle marjinal etkileri belirli bir x değerinde (tipik olarak ortalamalar) tahmin etmemiz gerekir. $F'(x'\beta) > 0$ olduğu için katsayılar ve marjinal etkiler aynı işaretlere sahiptir.

Logit model için marjinal etkiler eşitlik (3.18)'de gösterilmiştir.[4]

$$\frac{\partial p}{\partial x_j} = \Lambda(x'\beta)[1 - \Lambda(x'\beta)]\beta_j = \frac{e^{x'\beta}}{(1 + e^{x'\beta})^2}\beta_j \tag{3.18}$$

Probit model için marjinal etkiler de eşitlik (3.19)'da gösterilmiştir.[8]

$$\frac{\partial p}{\partial x_j} = \Phi(x'\beta)\beta_j \tag{3.19}$$

•Ortalamada marjinal etkiler;

Eşitlik (3.20)'de görüldüğü üzere marjinal etkiler, örneklemdeki ortalama kişi için tahmin edilmektedir.[8]

$$\frac{\partial p}{\partial x_j} = F'(\overline{\boldsymbol{x}}'\beta)\beta_j \tag{3.20}$$

Çoğu makale marjinal etkileri ortalamada rapor etmektedir.

• Ortalama marjinal etkiler;

Marjinal etkiler, bireysel marjinal etkilerin ortalaması olarak tahmin edilir ve eşitlik (3.21) ile hesaplanmaktadır.[31]

$$\frac{\partial p}{\partial x_j} = \frac{\sum F'(\boldsymbol{x}'\beta)}{n}\beta_j \tag{3.21}$$

Bu, marjinal etkileri tahmin etmek için daha iyi bir yaklaşımdır ancak makaleler hala bir önceki yaklaşımı kullanmaktadır. Uygulamada, marjinal etkileri tahmin etmenin iki yolu çoğu zaman hemen hemen aynı sonuçları vermektedir.

• Ayrık değişkenler için kısmi etkiler;

Bir değişkenin iki ayrık değeri için olasılıkları tahmin edilir ve farkı alınır ve eşitlik (3.22) ile ifade edilmektedir.[4]

$$F = (\hat{\beta}_0 + \hat{\beta}_1\, x_1 + \hat{\beta}_2(k+1)) - F(\hat{\beta}_0 + \hat{\beta}_1\, x_1 + \hat{\beta}_2(k)) \tag{3.22}$$

• Marjinal etkilerin yorumlanması;

· x'teki bir artış, yüzde olarak ifade edilen marjinal etki ile y=1 olma olasılığını artırır(azaltır).

Kukla bağımsız değişkenler için marjinal etki, temel kategoriye (x=0) göre ifade edilir. Sürekli bağımsız değişkenler için marjinal etki, x'teki bir birimlik değişim için ifade edilir. Marjinal etkilerin hem işaretini hem büyüklüğünü yorumlarız. Probit modeller ve logit modeller hemen hemen aynı marjinal etkiler üretmektedir.

Odds oranı veya bağıl risk $\mathrm{p}/(1-\mathrm{p})$'dir ve y=0 olasılığına göre y=1 olma olasılığını ölçer ve eşitlik (3.23) ile hesaplanmaktadır. [4]

$$p = \frac{\exp(\boldsymbol{x}'\beta)}{1 + \exp(\boldsymbol{x}'\beta)}$$
$$\frac{p}{1-p} = \exp(\boldsymbol{x}'\beta) \tag{3.23}$$
$$ln\frac{p}{1-p} = \boldsymbol{x}'\beta$$

2'lik bir olasılık oranı , y=1 sonucunun y=0 sonucunun iki katı olduğu anlamına gelmektedir. Odds oranları lojistik model ile tahmin edilmektedir ve olasılık oranları yerine marjinal etkilerin raporlanması ekonomide daha yaygındır.

Modelleri tahmin ettikten sonra, her gözlem için y=1 olma olasılığını eşitlik (3.24) ile tahmin edebiliriz.[8]

$$\hat{p} = pr[y = 1|x] = F\left(\boldsymbol{x}'\hat{\beta}\right) \tag{3.24}$$

Regresyon modeli için tahmin edilen olasılıklar 0 ile 1 arasında sınırlı değildir. Logit ve probit modelleri için tahmin edilen olasılıklar 0 ile 1 arasında sınırlandırılmıştır. Öngörülen olasılık y=1 olasılığını gösterir. Tahmin edilen olasılık 0.5'ten büyükse y=1 , aksi takdirde y=0 olacağını tahmin edebiliriz.

Tablo 3.2 Tahmin edilen değerler [8]

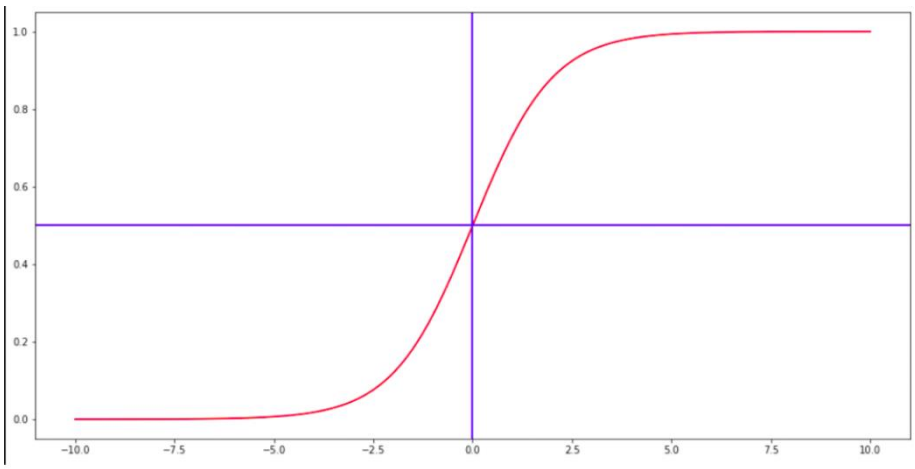
	Gerçek y=1	Gerçek y=0
Tahmin edilen =1	Doğru	Yanlış
Tahmin edilen =0	Yanlış	Doğru

Logit ve probit model arasında seçim;

Seçim bilinmeyen veri üretme sürecine bağlıdır. Modeller hemen hemen aynı sonuçlar üretir.(farklı katsayılar ancak benzer marjinal etkiler). Bağımlı değişkenler kodlanırken 0 ve 1 kategorilerini tersine çevirirsek, katsayıların işaretleri tersine çevrilir ancak büyüklükleri aynıdır.

Lojistik regresyon, örnek değerlerin gözlem büyüklüklerini maximum düzeye çıkaran parametreleri seçmektedir. Lojistik regresyon, bağımlı değişkenin binary olması durumunda kullanılmaktadır. Veriler değerlendirildiğinde sonucun 0 veya 1 olmasını tahmin etmeye sınıflandırma adı verilmektedir. Yapılan tahminlemede olasılık değeri 0.5'ten küçükse model sonucu 0, olasılık değeri 0.5'ten büyük veya eşitse model sonucu 1 olmaktadır. Bu kısımda Sigmoid fonksiyonu devreye girmektedir.

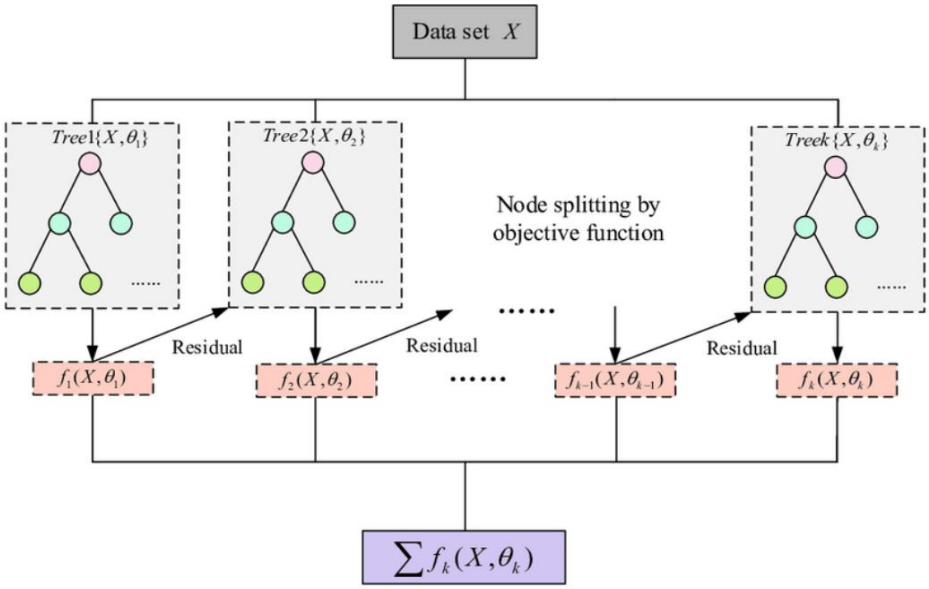
Şekil 3.3'te Z değerinin artı sonsuza gitmesi durumunda y 1 değerine, eksi sonsuza gitmesi durumunda 0 değerine yaklaşmaktadır. Değer tahmin olarak 0.5'ten büyük veya eşit olduğu durumda 1, küçük olduğu durumda 0 sonucunu almaktadır.



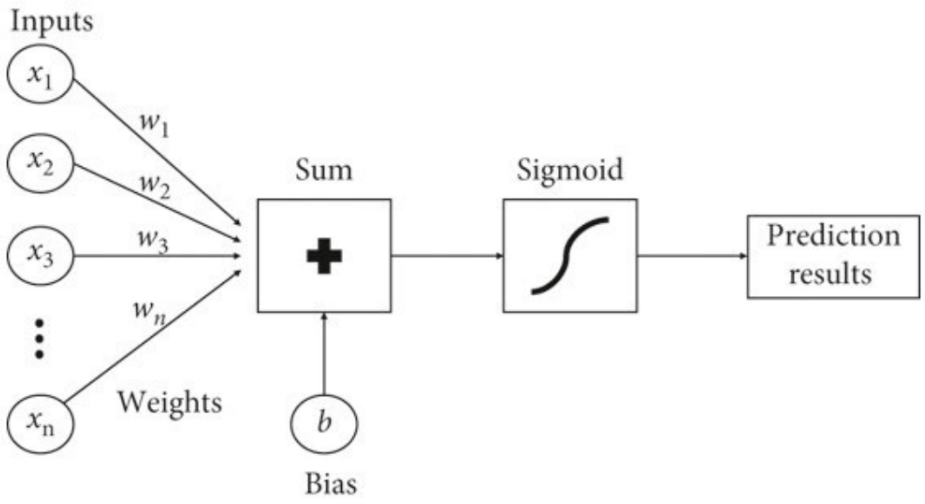
Şekil 3.3 Sigmoid fonksiyonu [28]

3.4 XGBoost ve Lojistik Regresyon

XGBoost, optimize edilmiş bir Gradient Boosting makine öğrenimi algoritmasıdır. Temel olarak Gradient Boost, sınıflandırma ve tahmin için tümevarım tabanlı bir yaklaşımdır. Buna karşılık Lojistik Regresyon, genelleştirilmiş bir doğrusal denklem kullanan doğrusal bir analiz yaklaşımıdır.



Şekil 3.4 XGBoost akış şeması [19]



Şekil 3.5 Lojistik regresyon akış şeması [17]

Her iki algoritmanın da kendine has özellikleri vardır. Her iki algoritma özellikleri Tablo 3.3'te özetlenmiştir.

Tablo 3.3 XGBoost ve lojistik regresyon özellikleri[27]

	XGBoost	**Lojistik Regresyon**
Yorumlanabilirlik	İyi	Uygun
Doğrusal ilişkiye sahip bir veri kümesindeki performans	Uygun	İyi
Çok sayıda özellikte performans	İyi	Kötü
Küçük bir eğitim setinde performans	İyi	Kötü
Çok sayıda aykırı değer içeren bir veri kümesindeki performans	İyi	Kötü (veya aykırı değeri veri kümesinden kaldırın)
Çarpık bir veri kümesindeki performans	İyi	Kötü (veya minör ağırlığı yeniden dengelemek)
Sürekli bir veri kümesindeki performans	Kötü	İyi
Veri kümesinde işleme eksik değerler	İyi	Kötü (veya eksik değerleri veri kümesinden kaldırın/düzeltin)

3.5 Modelleri Test Etme ve Doğrulama

Bir modelin yeni durumlara ne kadar iyi genelleneceğini bilebilmek için veriler test seti ve eğitim seti olarak iki gruba ayrılır. Hata oranı yeni durumlarda örneklem dışı hata (genelleme hatası) olarak adlandırılmaktadır. Model test kümesinde değerlendirilerek genelleme hatasının tahmini alınır. Bulunan bu değer modelin daha önce hiç görmediği durumlarda ne kadar iyi performans gösterdiği konusunda bilgi verir. Veriler genel olarak 2/3'ü eğitim seti, 1/3'ü test seti olarak ikiye bölünür.

3.6 Modellerin Değerlendirilmesi

Öğrenme algoritmalarında kurulan modellerin başarısının değerlendirilmesi için doğruluk (accuracy), karışıklık matrisi (confusion matrix) ve ROC eğrisi gibi yöntemler kullanılmaktadır.

3.6.1 Karışıklık Matrisi

Karışıklık matrisi, makine öğrenmesindeki sınıflandırma modellerinin performansını değerlendirmek amacıyla tahminlerin ve gerçek değerlerin karşılaştırılması için kullanılmaktadır. Karışıklık matrisi, veri setindeki durum ile sınıflandırma modelindeki doğru veya yanlış tahminlerin sayısını tablo olarak biçimlendirmektedir. Performans, modelin genel doğruluğu açısından ölçülür.

•Gerçek Pozitif (True Positive - TP) : Pozitif tahmin edilmesi ve bunun doğru olması

•Gerçek Negatif (True Negative -TN) :Negatif tahmin edilmesi ve bunun doğru olması

•Yanlış Pozitif (False Positive -FP) : Pozitif tahmin edilmesi ve bunun yanlış olması

•Yanlış Negatif (False Negative -FN) : Negatif tahmin edilmesi ve bunun yanlış olması

Gerçek Pozitifler = TP+FN

Gerçek Negatifler = TN+FP

Toplam = TP+TN+FP+FN

Karışıklık matrisiyle doğruluk, kesinlik, duyarlılık, F1 ve Kappa değeri bulunabilir.

• **Doğruluk (Accuracy) değeri**

Olumlu (pozitif) veya olumsuz (negatif) sonuçların gözlem sayısına oranıdır. Accuracy (doğruluk) değeri, sistemdeki doğru tahminlerin tüm tahminlere oranıdır ve (3.25) eşitliği ile hesaplanmaktadır. [18]

$$\text{Doğruluk değeri} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \tag{3.25}$$

• **Kappa değeri**

İki veya daha fazla değerlendiricinin sonuçları arasındaki uyumun güvenilirliğini ölçen bir yöntemdir ve (3.26) eşitliği ile ifade edilir[20]. Kappa değeri -1 ile 1 arasında yer almaktadır. Kappa değeri 1'e yaklaştıkça değerlendirici sonuçları birbiriyle uyumluyken, 0'a yaklaştıkça uyum rassal ve -1'e yaklaştığında sonuçlar birbiriyle ters değerlendirilir.

$$\text{Kappa değeri} = \frac{\text{Pr}(\alpha) - \text{Pr}(\text{e})}{1 - \text{Pr}(\text{e})} \tag{3.26}$$

α gerçek değer, e temsil edilen değerdir.

• **Kesinlik**

Gerçek pozitif olanların, tahmin değeri pozitif olanların tümüne oranıdır ve (3.27) eşitliği ile hesaplanmaktadır.[18]

$$\text{Kesinlik} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.27}$$

• **Duyarlılık**

Gerçek değeri pozitif olanların, gerçek değeri pozitif olan ve tahmini değeri de pozitif ve negatif olan gözlem sayılarının toplamlarının oranıdır ve eşitlik (3.28) ile gösterilmiştir.[18]

$$\text{Duyarlılık} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.28}$$

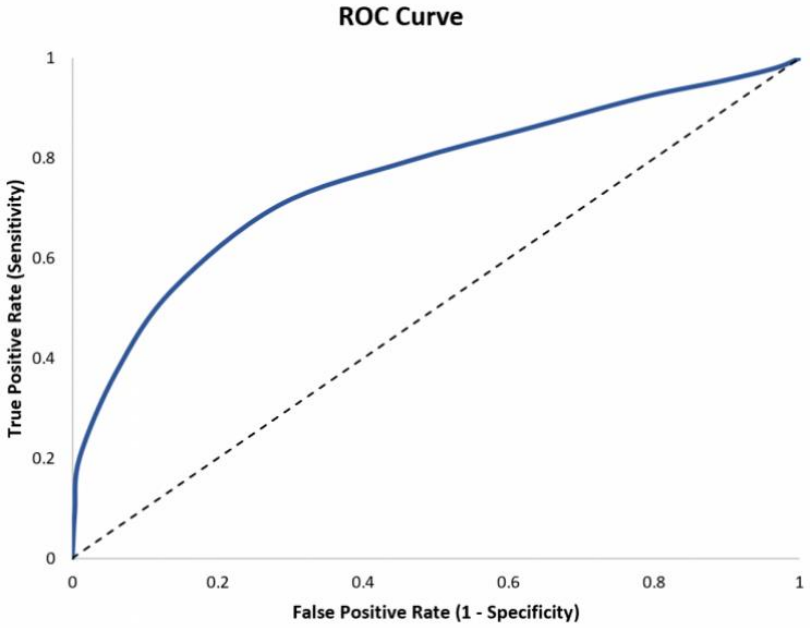
• **F1 değeri**

F1 değeri (3.29) eşitliğindeki gibi kesinlik ve duyarlılık değerlerinin harmonik ortalamasıyla bulunur. F1 değeri 0-1 arasında değer almaktadır ve 1'e yaklaştıkça skor iyileşmektedir.[20]

$$\text{F1 değeri} = 2 * \frac{\text{Kesinlik} - \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \tag{3.29}$$

3.6.2 ROC Eğrisi (Receiver Operating Characteristic Curve)

ROC analizi, bir testin teşhis performansını (en yüksek özgüllük ve duyarlılık değerlerini) belirlemek, farklı kesim noktalarına ulaşmak, duyarlılık ve özgüllük değerlerini kıyaslamak, belirli duyarlılık değerlerine denk gelen özgüllük değerlerini ya da bazı özgüllük değerlerine denk gelen duyarlılık değerlerini belirlemek, iki ya da daha fazla testin teşhis performansını kıyaslamak için kullanılır.



Şekil 3.6 ROC eğrisi [27]

Duyarlılık (Sensitivity): Sonuç gerçekten olumlu olduğunda, modelin bir gözlem için olumlu bir sonuç öngörme olasılığı.

Özgüllük (Specificity): Sonuç gerçekten olumsuz olduğunda, modelin bir gözlem için olumsuz bir sonuç öngörme olasılığı.

ROC analizi sonuçları; yanlış pozitiflik (false positive), doğru pozitiflik (true positive), Youden İndeksi (J), eğri altındaki alan (area under curve) gibi ölçütler yardımıyla yorumlanmaktadır. Kesim noktası değiştikçe tanı testinin duyarlılık ve özgüllük değeri değişeceğinden, başarısının değerlendirmesi için belirlenecek kesim noktası da farklı olacaktır. Olası tüm kesim noktalarında tanı testinin her yanlış pozitif sonuç için ne kadar doğru pozitif sonuç vereceğini gösteren bir grafiğin de tanı testini değerlendirmede kullanılabileceği düşünülmüştür. Böylelikle farklı kesim noktalarında testin yanlış pozitif (1 – specificity) oranının x-ekseninde, doğru pozitif oranının (sensitivity) y-ekseninde olduğu ROC eğrisi (Receiver Operating Characteristic) bulunur.

4
UYGULAMA

Çalışmada, DataCo Global şirketi tarafından Makine Öğrenmesi Algoritmalarının ve R Studio Programlama Dili'nin kullanılmasına izin veren tedarik zincirleri veri kümesi kullanılmıştır. Şirket, dünya genelinde petrol ve gaz sektöründeki şirketlere doküman kontrolü, taksonomiler, veri entegrasyonu, fiziksel veri yönetimi ve uyumu içeren çözümler sunmaktadır. Çalışmanın amacı, tedarik zinciri yönetiminde departmanlara göre geç teslimat riskinin sınıflandırmaya dayalı makine öğrenmesi teknikleriyle belirlenebilmesidir. Yapılan uygulamada, R Studio'da bağımsız değişkenlerin her biri bağımlı değişken olan departmanlara göre görsellerle ayrıştırılmıştır.

4.1 Veri Düzenleme ve Değişken Açıklama

Uygulamada 180519 gözlem ve 18 değişken vardır. Veride normalde 53 değişken vardır fakat konu itibariyle bazı değişkenler, faktör dönüşümünde sorun yaratmıştır ve birbiriyle aynı olan değişkenler de var olduğundan dolayı çıkarılmıştır. Ek olarak kıyafet ve ayakkabı departmanı, kitap ve cd departmanı, klima ve teknoloji departmanı, spor malzemeleri ve golf malzemeleri departmanı birleştirilmiş olup DNN değişkenine atanmıştır. Lojistik regresyon ve XGBoost modelleri de yeni oluşturulan DNN değişkeni ile sınıflandırılmıştır. Bağımlı değişken geç teslimat riski olan LDR değişkeni olarak belirlenmiştir. Veride eksik değer tespiti var mı diye kontrol edilmiş ve eksik değerlerin çıkarılması için işlem yapılmıştır ve veride eksik gözlem bulunmadığı sonucuna ulaşılmıştır.

Değişken düzeyleri hesaplanmış ve her bir değişkendeki düzeyler hakkında Tablo 4.1'de görüldüğü üzere açıklamalarına yer verilmiştir.

Tablo 4.1 Değişken açıklamaları

Değişken Adı	Açıklaması
TYPE	Yapılan ödeme türünü belirtir (borç,havale,eft..)
DFSR	Satın alınan ürünün gerçek nakliye günleri
DFSS	Satın alınan ürünün planlanan teslimat günleri
DS	Siparişin teslimat durumu (zamanında,erken,geç..)
LDR	Geç teslimat riski (0,1)
CN	Kategori ismi
CC	Müşterinin satın alma işlemini yaptığı ülke (E.E, Puerto Rico)
CS	Müşteri türleri (alıcı, kurumsal…)
DNN	Departmanlar (kıyafet,kitap,teknoloji…)
OC	Siparişin hedef ülkesi (Avusturalya,Çin…)
OD	Siparişin verildiği tarih
OIQ	Sipariş başına ürün sayısı
OR	Siparişin teslim edildiği bölge (Doğu Afrika,Doğu Asya…)
OS	Sipariş durumu (tamamlandı, iptal, beklemede…)
PN	Ürün adı
SD	Sevkiyat tarihi
SM	Gönderim tarzı (1.sınıf, 2.sınıf…)

TYPE: 4 düzeylidir. Yapılan ödeme türünü belirtir.

Debit: Borç

Transfer: EFT/Havale

Cash: Peşin

Payment: Kredi kartı

DFSR: 7 düzeylidir. Satın alınan ürünün gerçek nakliye günlerini belirtir.

DFSS: 4 düzeylidir. Satın alınan ürünün planlanan teslimat günlerini belirtir.

DS: 4 düzeylidir. Siparişin teslimat durumunu belirtir.

Advance shipping: Erken sevkiyat

Late delivery: Geç teslimat

Shipping canceled: Sevkiyat iptal edildi

Shipping on time: Zamanında sevkiyat

LDR: 2 düzeylidir. Geç teslimat riskini belirtir.

Erken teslimat:0

Geç teslimat:1

CN: 50 düzeylidir. Kategori ismini belirtir.

CC: 2 düzeylidir. Müşterinin satın alma işlemini yaptığı ülkeyi belirtir.

E.E

Puerto Rico

CS: 3 düzeylidir. Müşteri türlerini belirtir.

Consumer: Alıcı

Corporate: Kurumsal

Home Office: Ev-Ofis

DNN: 7 düzeylidir. Mağaza departmanlarının birleşimini belirtir.

Apparel: Ayakkabı ve kıyafet

Book: Kitap ve cd

Health-Beauty: Sağlık ve güzellik

Outdoors: Dış mekan malzemeleri

Petshop: Evcil hayvan malzemeleri

Sport: Spor ve golf malzemeleri

Technology: Teknoloji ve klima

OC: 164 düzeylidir. Siparişin hedef ülkesini belirtir.

OD: 1127 düzeylidir. Siparişin verildiği tarihi belirtir. (gün-ay-yıl)

OIQ: 5 düzeylidir. Sipariş başına ürün sayısını belirtir.

OR: 23 düzeylidir. Siparişin teslim edildiği bölgeyi belirtir.

OS: 9 düzeylidir. Sipariş durumunu belirtir.

Complete: Tamamlandı

Pending: Beklemede(telefonda)

Closed: Kapalı

Pending Payment: Ödeme beklemede

Canceled: İptal edildi

Processing: İşleniyor

Suspected Fraud: Şüpheli işlem

On Hold: Beklemede

Payment Review: Ödeme incelemede

PN: 118 düzeylidir. Ürün adını belirtir.

SD: 1131 düzeylidir. Sevkiyat tarihini belirtir. (gün-ay-yıl)

SM: 4 düzeylidir. Gönderim tarzını belirtir.

Standard Class: Standart sınıf

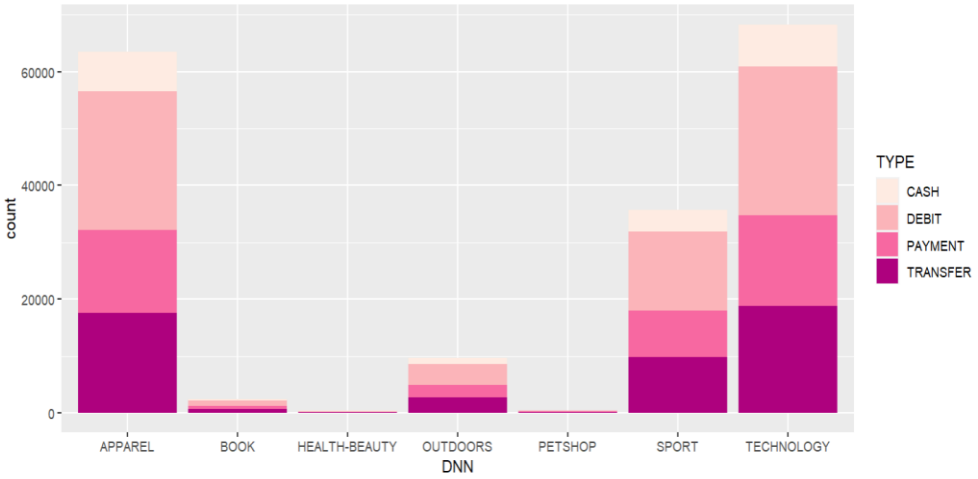
First Class: Birinci sınıf

Second Class: İkinci sınıf

Same Day: Aynı gün

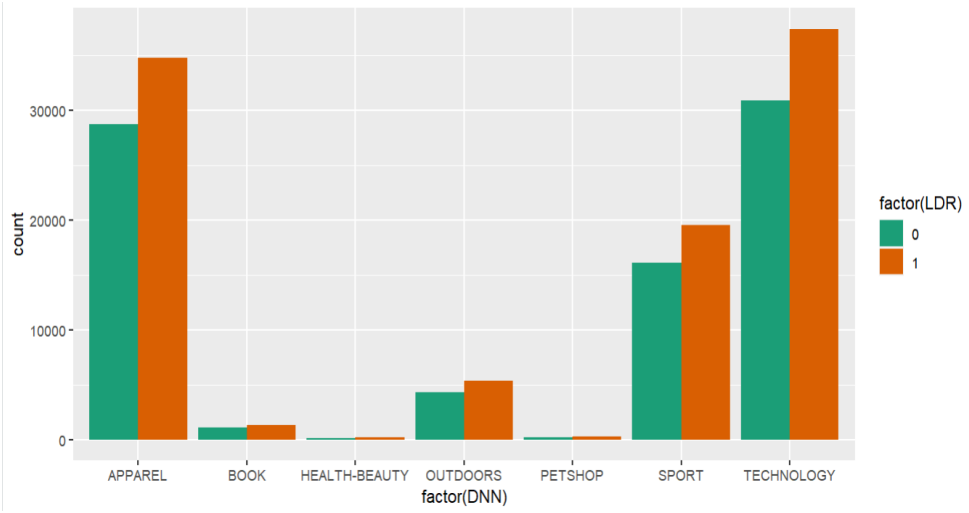
4.2 Verilerin Görselleştirilmesi

Şekil 4.1'de Kıyafet, Spor ve Teknoloji deprtmanlarının frekansları fazla olduğundan yapılan ödeme türlerine göre daha kolay analiz edilebilmektedir. Genel olarak alışverişlerin borç olarak yapılması diğer ödeme türlerine göre en fazla olarak görülmektedir. Ödemeleri nakit olarak yapılan alışverişlerin ise en az olduğu görülmektedir. Finansman maliyetlerini azaltmak için işletmenin hem kendi tedarik zinciri hem de müşteriyle yaptığı sözleşmeleri gözden geçirmesi gerekmektedir. Eğer şirketin kendi ödemelerinin vadesi alacaklarının ödemesinden az ise iş yapılan şirketlerle ve müşterilerle konuşulup alacakların ve ödemelerin eşleştirilmesi gerekmektedir. Ödeme vadelerini, tahsilatların bir miktar gecikeceği varsayılarak yapmak en doğru yoldur.



Şekil 4.1 Yapılan ödeme türlerinin departmanlara göre grafiği

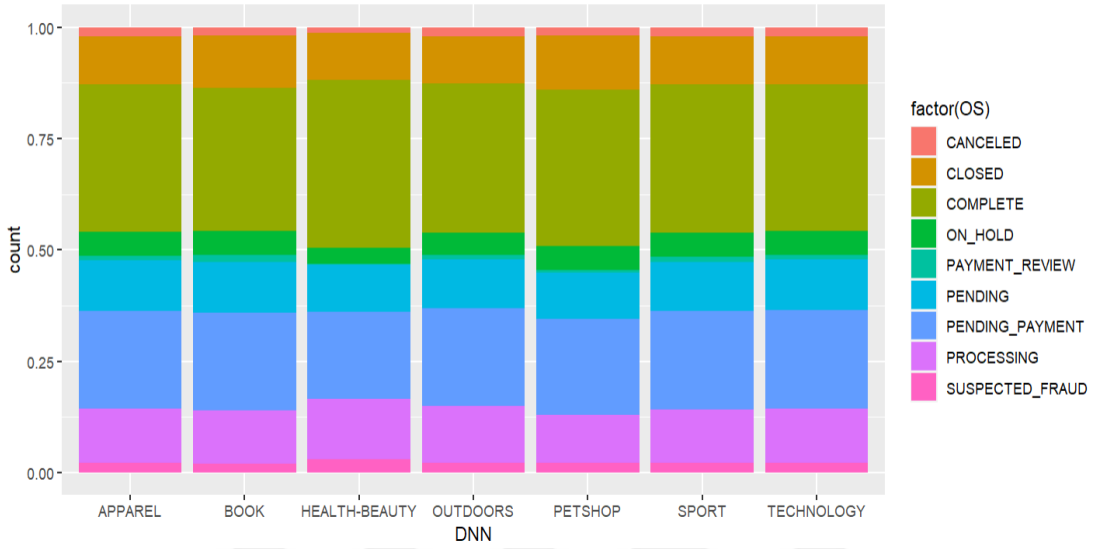
Şekil 4.2'de görüldüğü üzere bütün departmanlarda erken teslim edilen ve geç teslim edilen ürünler birbirine yakın düzeydedir. Sağlık-güzellik departmanı ve evcil hayvan malzemeleri departmanı dışındaki 5 departmanda geç teslimat riski erken teslimat riskine göre daha fazladır. Müşterinin işletmeyi tercih etme açısından geç gönderim büyük bir risktir. Müşterilerin gönderilen siparişlerin durumunu çevrimiçi bir şekilde görebilmeleri için siparişlerin kargoya verildiğini, tahmini kargoya verilişi tarihine kadar onaylamak gerekmektedir. Kargoya geç verilen siparişler, müşteri deneyimini olumsuz yönde etkileyebilir. Şikayetlerin ve olumsuz değerlendirmelerin oluşmasına ayrıca müşteri ile iletişimin artmasına sebep olabilir. Ayrıca siparişlerin geç gönderilmesi müşteri kaybına da sebep olabilir.



Şekil 4.2 Geç teslimat riskinin departmanlara göre grafiği

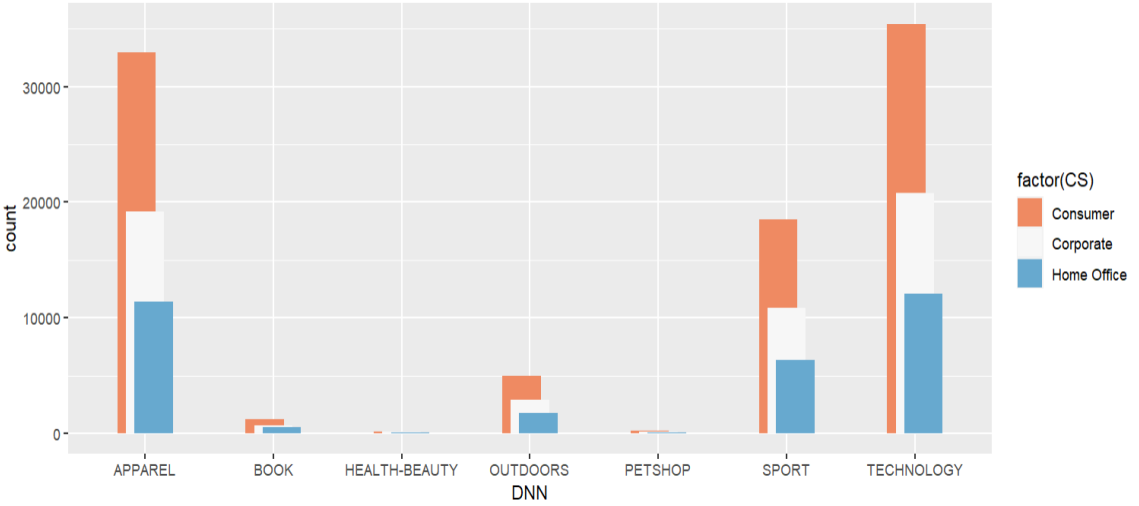
0: Erken teslimat 1: geç teslimat olarak belirtilmektedir.

Şekil 4.3'te görüldüğü üzere yapılan incelemelere göre, siparişi iptal edilenler genel olarak diğer işlemlere göre en azdır. Sipariş durumu tamamlananların diğer işlemlere göre en fazla olduğu gözlemlenmektedir. Ödemesi beklemede olanlar tamamlanan duruma göre daha az olsa da diğer işlemlere göre daha fazladır. Şüpheli durum, dolandırıcılık riski de az da olsa vardır diyebiliriz.



Şekil 4.3 Sipariş durumlarının departmanlara göre grafiği

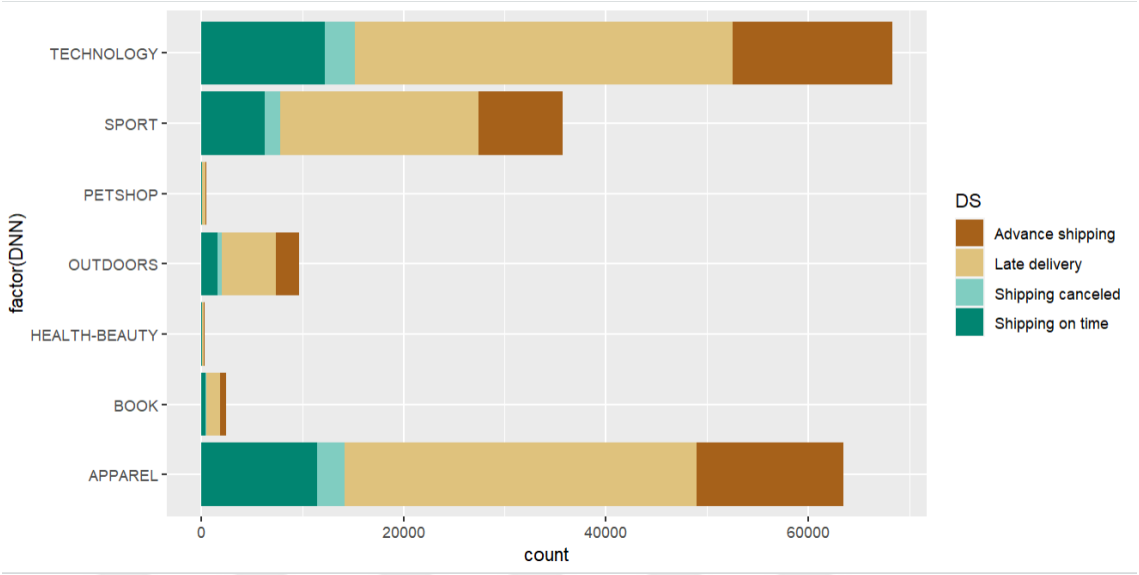
Şekil 4.4'te departmanlara göre müşteri türleri gösterilmektedir. Müşteri türlerinden biri olan Home Office olarak sipariş verenler tüm departmanlarda en azken, müşteri türlerinden biri olan Consumer olarak sipariş verenlerin tüm departmanlarda en fazla olduğu görülmektedir.



Şekil 4.4 Müşteri türlerinin departmanlara göre grafiği

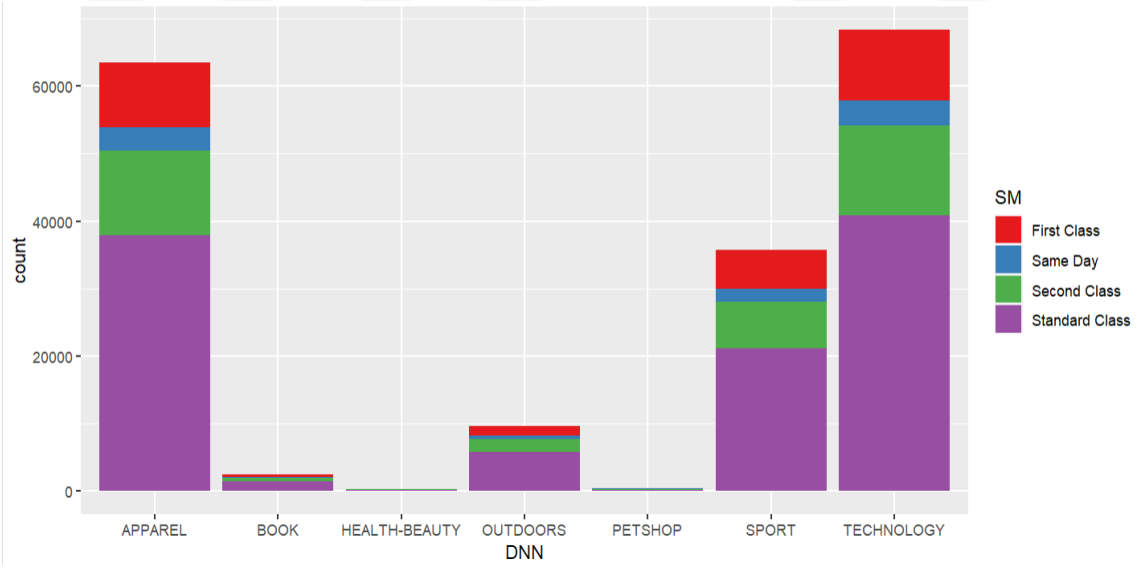
Şekil 4.5'te bütün departmanlarda geç teslimatların olduğu açıkça görülmektedir. İptal

edilen siparişler ise en azdır. Erken teslimat da zamanında yapılan teslimatlara göre daha fazladır diyebiliriz.

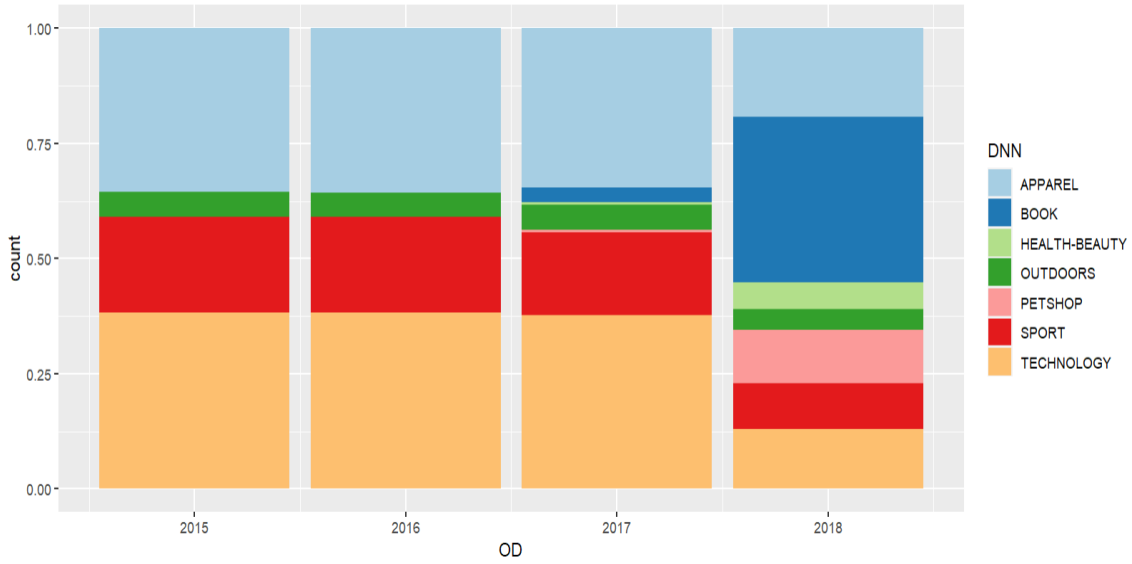


Şekil 4.5 Teslimat durumlarının departmanlara göre grafiği

Şekil 4.6'da görüldüğü üzere departmanlar siparişlerin gönderim tarzına göre sınıflandırılmıştır. Standart sınıf olarak gönderimler en fazlayken aynı gün içinde yapılan gönderimler bütün departmanlarda en azdır diyebiliriz.



Şekil 4.6 Siparişlerin gönderim tarzlarının departmanlara göre grafiği



Şekil 4.7 Sipariş tarihlerinin departmanlara göre grafiği

Satın alma davranışı, ürün kalitesinin yanısıra hedef tüketicinin ekonomik, psikolojik ve kültürel özellikler tarafından da etkilenmektedir. Ayrıca markalar son yıllarda satın alma davranışları bakımından son derece belirleyici olmaya başlamıştır. Markalaşma, işletmelerin satış ve pazarlama stratejilerini belirlerken tüketicilerin alışveriş eğilimlerini de etkilemektedir.

Şekil 4.7'de kıyafet departmanında 2018 yılında diğer yıllara göre verilen siparişler daha azdır. Kitap departmanında 2015, 2016 yıllarında sipariş görülmezken 2018 yılında verilen siparişler artış göstermiştir. Bunun yanı sıra 2017 yılında bu departmanda verilen siparişler az da olsa bulunmaktadır.

Sağlık ve güzellik departmanında 2015 ve 2016 yıllarında herhangi bir sipariş görülmezken 2017 yılında nadir de olsa siparişler göze çarpmaktadır. 2018 yılında da siparişlerde artış görülmektedir.

Dış mekan malzemeleri departmanında bütün yıllarda birbirine yakın düzeyde sipariş verilmiştir. Evcil hayvan malzemeleri departmanında 2018 yılında artış görülürken 2015 ve 2016 yıllarında sipariş verilmediği görülmektedir. Spor malzemeleri ve teknoloji departmanında ise 2018 yılında diğer yıllara göre siparişlerde azalış görülmektedir.

4.3 Lojistik Regresyon Modeli

Lojistik regresyonda kullanılan veri setinde departmanlara göre geç teslimat riskinin oluşup oluşmadığı kontrol edilmiştir. Öncelikle departmanlar ve geç teslimat riski değişkenleri faktör olarak dönüştürülmüştür. Ardından geç teslimat riski değişkeninin levellarının oranı kontrol edilmiştir.

Tablo 4.2 LDR level

0	1
81542	98977

Daha iyi modeller elde etmek amacı ile gözlemler yaklaşık eşit oranlarda örneklenmiştir ve sırasıyla training ve test verileri oluşturulmuştur.

Tablo 4.3 Test verisi

0	1
81417	98852

Train ve test verilerini ayırdıktan sonra lojistik regresyon modeli kurulmuştur.

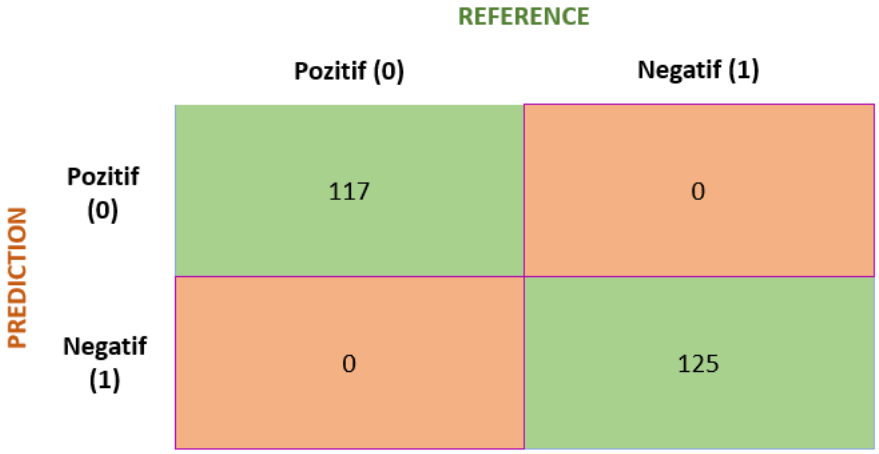
Tablo 4.4 Lojistik regresyon modeli

Coefficients:

	Estimate	**Std. Error**	**Z value**	**Pr (>|z|)**
(Intercept)	-4.99101	1.60624	-3.107	0.00189**
data.DFSR	6.53189	1.23756	5.278	1.31e-07***
data.DFSS	-6.54660	1.24720	-5.249	1.53e-07***
data.DNNBOOK	1.87834	3.30943	0.568	0.57032
data.DNNOUTDOORS	-0.81745	1.32284	-0.618	0.53661
Data.DNNSPORT	0.86120	1.28354	0.671	0.50225
data.DNNTECHNOLOGY	0.07379	0.79731	0.093	0.92626
data.OIQ	0.25457	0.27778	0.916	0.35943

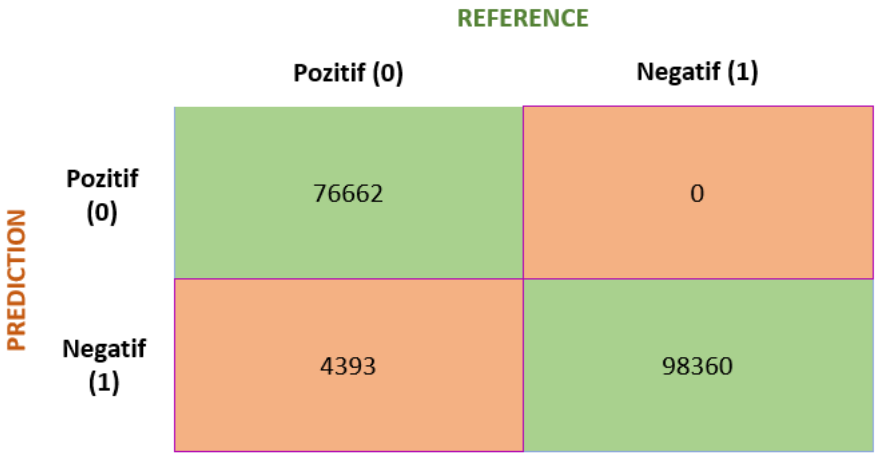
Training set için tahminler oluşturulduktan sonra confusion matrix oluşturulmuştur.

Training set üzerindeki sınıflandırma performansında hata oranı 0.032'dir.



Şekil 4.8 Eğitim seti karışıklık matrisi

Accuracy değeri: (117+125) / (117+125+8+0) = 242/251 = 0.9661 (%96,6) olarak görülmektedir. Bu aşama tamamlandıktan sonra veri test verisi için eğitilmektedir.



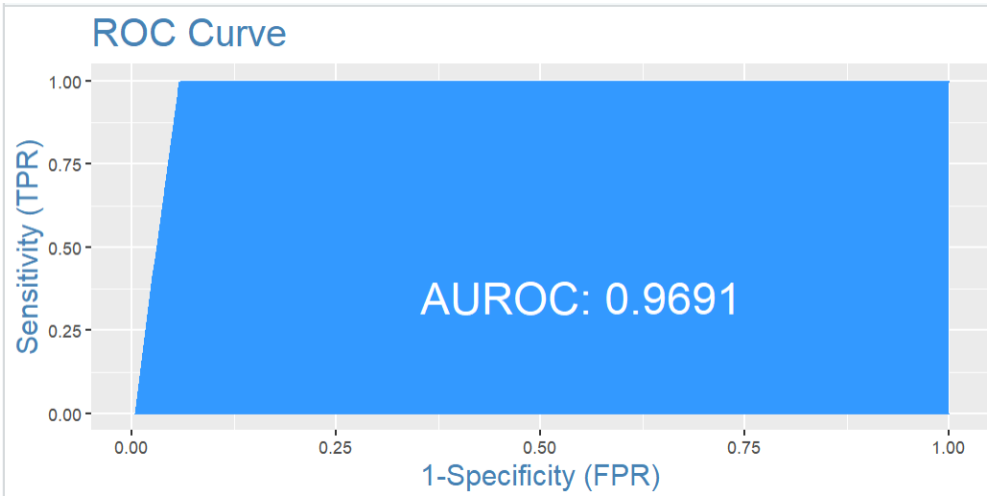
Şekil 4.9 Test set karışıklık matrisi

Hata oranı 0,0244 bulunmuştur.

Test verisi üzerinde daha düşük bir hata oranı verdi diyebiliriz.

Accuracy değeri: (76662+98360) / (76662+98360+4393+0)

= 175022/179415 = 0.97 (%97) olarak görülmektedir. Test veri seti için doğru sınıflandırma oranı gayet iyidir.



Şekil 4.10 ROC eğrisi

ROC Eğrisi iki grubun ne kadar ayırt edilebildiğinin ölçüsüdür. ROC Eğrisinin altında kalan alan ne kadar büyükse doğru oranı da o kadar yüksektir. Şekil 4.10'da görüldüğü üzere ROC Eğrisinin altında kalan alan %96,9'dur. Bu da iki grubun çok iyi ayrıldığı anlamına gelmektedir. Yani model, geç teslimat riskinin olup olmadığını yüksek doğruluk oranıyla sınıflandırmaktadır.

4.4 XGBoost Modeli

XGBoost modelinde de lojistik regresyon modelindeki gibi departmanlara göre geç teslimat riskinin olup olmadığı analiz edilmiştir. Gerekli paketler yüklendikten sonra faktör değişkenler numerik hale dönüştürülmüştür.

Veri seti, 0.8 eğitim seti ve 0.2 test seti olarak bölünmüştür. Eğitim veri setinde 144415 gözlem bulunurken test veri setinde 36104 gözlem bulunmaktadır. Veri setinde XGBoost modelini uygulayabilmek için bağımlı değişkenin son sütunda olduğu veri seti ayarlanmıştır.

Eğitim ve test veri setleri oluşturulduktan sonra çapraz doğrulama ve XGBoost için parametre ayarları yapılmıştır.

XGBoost modelinde, bağımlı değişken etiketinin numerik olması gerekmektedir. Bu yüzden bağımlı değişkenlere etiket tanımlanmış ve etiketler numerik olarak dönüştürülmüştür. Ayrıca XGBoost modelinde etiketler sıfırdan başlamalıdır.

Çapraz doğrulama ile minimum kayıp tespiti için iterasyon yazılıp en az kayıp veya hatayı veren basamak bulunmuştur. cv.nround iterasyon tur sayısı 150 olarak belirlenmiştir. 150

iterasyon içinde minimum loglossun gerçekleştiği adım bulunarak bu turdaki kayıp değerlere bakılmıştır.

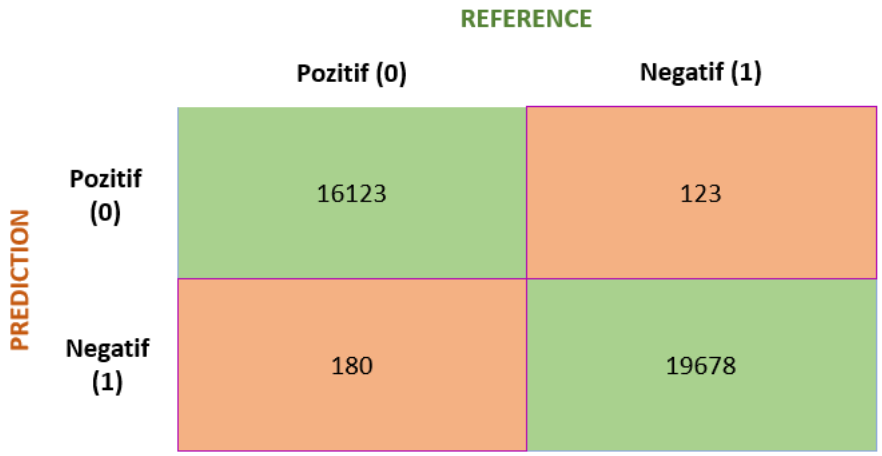
Tablo 4.5'te görüldüğü üzere 150.adımda minimum kayıp değeri bulunmuştur. Bu yüzden XGBoost modeli iterasyon sayısı 150 olarak eğitilmiştir.

Tablo 4.5 Minimum kayıp değerleri

İter.	**train_mlogloss_mean**	**train_mlogloss_std**	**test_mlogloss_mean**	**test_mlogloss_std**
150	0.06478844	0.0007175378	0.0757519	0.001141133

Minimum hatanın bulunduğu adımda iterasyonu durdurmak gerekmektedir. Model eğitildikten sonra test veri setinde tahminleme yapılmıştır.

XGBoost modeli için bağımlı değişkenin etiketinin sıfırdan başlaması gerektiğinden dönüşüm uygulanmıştı. Etiketler yeniden ilk haline dönüştürülmüştür ve ardından karışıklık matrisi ile modelin doğruluk oranları analiz edilmiştir.



Şekil 4.11 Karışıklık matrisi

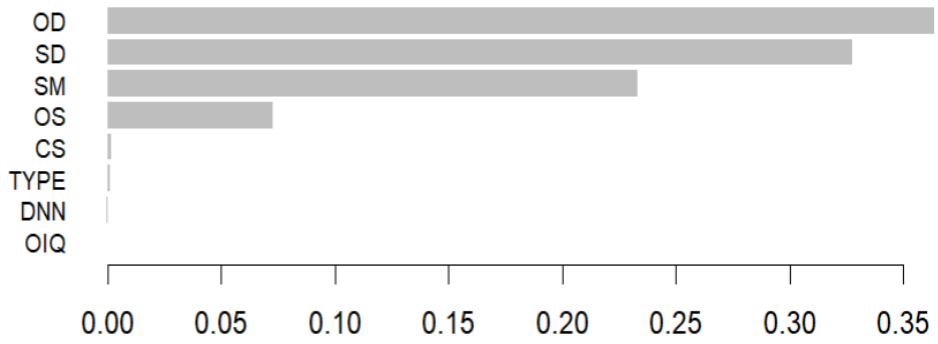
Modelin tahminleme doğruluğu ve istatistiksel sonuçları Tablo 4.6'da gösterilmiştir.

Tablo 4.6 İstatistikler ve doğruluk oranları

Statistics	
Accuracy	0.9916
Kappa	0.9831
Sensitivity	0.9890
Specificity	0.9938
Pos Pred Value	0.9924
Neg Pred Value	0.9909
Prevalance	0.4516
Detection Rate	0.4466
Detection Prevalence	0.4500
Balanced Accuracy	0.9914

Karışıklık matrisinde, modelin tahminleme doğruluğu ve istatistiksel sonuçlarına bakıldığında, modelin çoklu sınıf sınıflandırma doğruluğu olan accuracy değeri 0.9916, kappa değeri de 0.9831 bulunmuştur. Model geç teslimat riskini oldukça iyi sınıflandırmıştır diyebiliriz.

Bağımsız değişkenlerin, bağımlı değişken üzerindeki önemini incelemek için önem matrisi kullanılmıştır. Şekil 4.12'de önem matrisinin grafiği gösterilmiştir. Bağımlı değişken olan geç teslimat riskini etkileyen en önemli değişkenler siparişin verildiği tarih ve siparişin teslim edildiği tarih olarak görülmektedir.



Şekil 4.12 Önem matrisinin grafiği

4.5 Modellerin Performanslarının Karşılaştırılması

Lojistik Regresyon modelinin doğruluk oranı 0.9690, XGBoost modelinin doğruluk oranı 0.9916 ile tahminleme gerçekleşmiştir. İki algoritmada da başarılı sonuçlar alınmıştır fakat bu çalışmada XGBoost modeli daha iyi performansa sahiptir diyebiliriz. Geç teslimat riskini yüksek olasılık değeri ile sınıflandırmıştır.

Makine öğrenmesi modelleri, daha karmaşık ve büyük verileri analiz edebilen, büyük ölçekli verilerde bile daha hızlı ve daha doğru sonuçlar elde edebilmeyi sağlamak için önemli rol oynarlar. Bilinmeyen risklere karşı kaçınma şansını mümkün kılabilmektedir.

Çalışmada kullanılan tedarik zinciri veri seti büyük ölçekli bir veri seti olduğundan seçilen makine öğrenmesi algoritmalarından Lojistik Regresyon modeli ve XGBoost modelinin performanslarının değerlendirilmesi ile her iki algoritmada da geç teslimat riskinin varlığı büyük ölçüde sınıflandırılmıştır. İşletmenin tedarik zincirine yönelik kar ve devamlılığın sağlanabilmesi adına stratejiler geliştirmesi gerektiği vurgulanabilir.

5
SONUÇ VE ÖNERİLER

Bu çalışmada kullanılan tedarik zinciri veri setinde her bir değişkenin geç teslimat riskleri analiz edilmiştir. Kullanılan makine öğrenmesi tekniklerinden olan XGBoost ve Lojistik Regresyon modellerinden hangi modelin veriyi daha iyi açıkladığı araştırılmıştır. Ayrıca RStudio programlama dili ile görseller kullanılarak veri seti incelenmiştir. Karşılaştırılan modellerin doğruluk oranlarına göre XGBoost modeli Lojistik Regresyon modeline göre daha iyi performansa sahiptir. XGBoost sınıflandırma yöntemine ilişkin geç teslimat riskini etkileyen değişkenlerin önemine bakılmıştır ve genel olarak en önemli dört değişken siparişin verildiği tarih, sevkiyat tarihi, gönderim tarzı ve sipariş durumudur. Geç teslimat riskinin belirlenmesinde de sipariş tarihi ile sevkiyat tarihi birbiriyle ilişkili olduğundan önem matrisinde de en önemli değişkenler arasında sipariş tarihi ve sevkiyat tarihi görülmektedir. Yapılan çalışmaya istinaden tedarik zinciri yönetiminin amaçlarından biri olan teslimat durumunun doğru yapılmaması durumunda oluşabilecek geç teslimat riski üzerinde durulmuştur. Sevkiyat planlamasının doğru yapılmaması ile müşteri kaybı oluşması ve işletmeye duyulan güvenin azalması söz konusudur.

Başarılı bir tedarik zinciri yönetiminde bütün süreçlerin birbiriyle uyumlu olması gerekmektedir. Tedarik zinciri yönetiminin koordinasyonu ile amaçlarını tedarik zincirinin bütünü yönünde en iyide tutması, işletmeye daha çok yarar sağlayacağı düşünülmüştür. Planlanan üretim süresinde işin tamamlanması ve zamanında müşteriye ulaştırılması açısından süreçleri iyi kontrol etmelidir.

Ayrıca işletmelerin tedarik zinciri yönetimini başarılı bir şekilde yürütebilmesi adına ödemelerin de zamanında yapılması önemli bir faktördür.

Bu çalışmadan yola çıkarak sevkiyat planlamasının doğru yapılmasıyla geç teslimat risklerinin en aza indirilmesi için aksiyon alınması gerektiği sonucuna varılmıştır.

6
KAYNAKÇA

[1] Özdemir, A. İ. (2004). Tedarik zinciri yönetiminin gelişimi, süreçleri ve yararları. Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, (23).

[2] Yurttaş, A. (2020). Tedarik zinciri yönetiminde risk faktörlerinin, tedarik zinciri bütünleşmesine ve performansına etkisi: Yapısal eşitlik modellemesi ile İTOB OSB örneği.

[3] ÇIKMAK, S., & UNGAN, M. C. (2021). Otomotiv Sektöründeki Tedarik Zinciri Riskleriyle İlgili Bir Araştırma. Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 25(4), 1692-1709.

[4] Gujarathi, D. M. (2022). Gujarati: Basic Econometrics. McGraw-hill.

[5] Küçükoğlu, M. T. (2020). Tedarik zincirinde risk değerlendirme: Risklerin tanımlanması, gruplandırılması ve önceliklendirilmesi üzerine bir çalışma. İşletme Araştırmaları Dergisi, 12(2), 2126-2141.

[6] Erdal, H. (2018). Tedarik zinciri risk yönetimi: Kavramsal çerçeve ve tedarik yönlü bir literatür araştırması. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 24(4), 764-796.

[7] Makkar, S., Devi, G., & Solanki, V. K. (2019, January). Applications of machine learning techniques in supply chain optimization. In International Conference on Intelligent Computing and Communication Technologies (pp. 861-869). Springer, Singapore.

[8] Katchova, A. (2013). Probit and logit models. Econometrics Academy, ed. Website and YouTube Channel.

[9] Pulat, M., & Kocakoç, İ. D. (2021). Türkiye'de Makine Öğrenmesi ve Karar Ağaçları Alanında Yayınlanmış Tezlerin Bibliyometrik Analizi. Yönetim ve Ekonomi Dergisi, 28(2), 287-308.

[10] Bordoloi, M., & Biswas, S. K. (2018). Sentiment Analysis of Product using Machine Learning Technique: A Comparison among NB SVM and MaxEnt. International Journal of Pure and Applied Mathematics, 118(19), 71-83.

[11] Şenel, S., & Alatli, B. (2014). Lojistik regresyon analizinin kullanıldığı makaleler üzerine bir inceleme. Journal of Measurement and Evaluation in Education and Psychology, 5(1), 35-52.

[12] Özden, C., & Çiğdem, A. C. I. (2018). Makine öğrenmesi yöntemleri ile yaralanmalı trafik kazalarının analizi: Adana örneği. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 24(2), 266-275.

[13] Murat, G. Ö. K. (2017). Makine öğrenmesi yöntemleri ile akademik başarının tahmin edilmesi. Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji, 5(3), 139-148.

[14] Bayer, H., & Çoban, T. (2015). WEB İSTATİSTİKLERİNDE MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE KRİTİK PARAMETRE TESPİTİ. Ejovoc (Electronic Journal of Vocational Colleges), 5(4), 23-41.

[15] BİLGİN, G. (2021). Makine öğrenmesi algoritmaları kullanarak erken dönemde diyabet hastalığı riskinin araştırılması. Journal of Intelligent Systems: Theory and Applications, 4(1), 55-64.

[16] Tomak, L., & Yüksel, B. E. K. (2009). İşlem karakteristik eğrisi analizi ve eğri altında kalan alanların karşılaştırılması. Journal of Experimental and Clinical Medicine, 27(2).

[17] Bari Antor, M., Jamil, A. H. M., Mamtaz, M., Monirujjaman Khan, M., Aljahdali, S., Kaur, M., ... & Masud, M. (2021). A comparative analysis of machine learning algorithms to predict alzheimer's disease. Journal of Healthcare Engineering, 2021.

[18] AKSOY, B., HALİS, H. D., & SALMAN, O. K. M. (2020). Elma bitkisindeki hastalıkların yapay zekâ yöntemleri ile tespiti ve yapay zekâ yöntemlerinin performanslarının karşılaştırılması. International Journal of Engineering and Innovative Research, 2(3), 194-210.

[19] Guo, R., Zhao, Z., Wang, T., Liu, G., Zhao, J., & Gao, D. (2020). Degradation state recognition of piston pump based on ICEEMDAN and XGBoost. Applied Sciences, 10(18), 6593.

[20] Yangın, G. (2019). Xgboost ve karar ağacı tabanlı algoritmaların diyabet veri setleri üzerine uygulaması (Master's thesis, Mimar Sinan Güzel Sanatlar Üniversitesi).

[21] Özdemir, M. C. (2021). Arıma ve XGBoost modelleri ile elektrik talep tahmini (Master's thesis, Sakarya Üniversitesi).

[22] WEERADDANA, N., & PREMARATNE, S. (2021). Unique approach for cricket match outcome prediction using xgboost algorithms. Journal of Theoretical and Applied Information Technology, 99(9), 2162-2173.

[23] Barstuğan, M. (2014). Özellik çıkarma ve DVM tabanlı AdaBoost algoritması ile biyomedikal veri sınıflandırma (Master's thesis, Selçuk Üniversitesi Fen Bilimleri Enstitüsü).

[24]https://feaconomy.com/, Tedarik Zinciri Yönetim Süreçleri, https://feaconomy.com/tedarik-zinciri-surecleri/, 31 Mayıs 2020.

[25]https://corporatefinanceinstitute.com/, Torbalama Algoritması, https://corporatefinanceinstitute.com/resources/knowledge/other/bagging-bootstrap-aggregation/, 28 Temmuz 2021.

[26]https://www.veribilimiokulu.com/, CatBoost Nedir, https://www.veribilimiokulu.com/catboost-nedir-diger-boosting-algoritmalarindan-farki-nelerdir/, 6 Ağustos 2020.

[27]https://www.statology.org/, ROC Curve, https://www.statology.org/interpret-roc-curve/, 9 Ağustos 2021.

[28]https://towardsdatascience.com/, Lojistik Regression -Detailed Overview, https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc, 15 Mart 2018.

[29]https://medium.com/, Lojistik Regresyon, https://medium.com/machine-learning-t%C3%BCrkiye/lojistik-regresyon-81d9685f075d, 30 Mart 2021.

[30] https://easonlai888.medium.com/, XGBoost or Logistic Regression model for Diabetes Prediction, https://easonlai888.medium.com/xgboost-or-logistic-regression-model-for-diabetes-prediction-1c3670cbbf6e, 1 Temmuz 2022.

[31] ŞAHİN, S., & ORHUNBİLGE, N. (2011) SOSYAL BİLİMLERDE KALİTATİF BAĞIMLI DEĞİŞKENLERE YAKLAŞIMDA İSTATİSTİK YÖNTEMLER (Doctor of Philosophy Thesis, İstanbul Üniversitesi)

[32]https://avys.omu.edu.tr, Lojistik Regresyon Analizi, https://avys.omu.edu.tr/storage/app/public/yukselt/118305/Lojistik%20regresyon%20analizi.pdf.

7
TEZDEN ÜRETİLMİŞ YAYINLAR

Bildiri

1.Kartaloğlu K., Büyüklü, A.H., (2022). "Ürün Tedarik Yönetiminde İstatistiksel Çalışmalar", 2.Uluslararası Bilimsel Araştırmalar ve Yenilikçi Çalışmalar Sempozyumu, 2-5 Mart 2022, Bandırma.