

**SATIŞ TAHMİNLEMESİNDE HİBRİT BİR YAKLAŞIM:
PESTEL, RFM, GRADIENT BOOSTING**

YÜKSEK LİSANS TEZİ

Esra AKCA

Anabilim Dalı: İstatistik

Programı: İstatistik

Tez Danışmanı: Prof. Dr. Semra ERPOLAT TAŞABAT

OCAK 2022

MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**SATIŞ TAHMİNLEMESİNDE HİBRİT BİR YAKLAŞIM:
PESTEL, RFM, GRADIENT BOOSTING**

YÜKSEK LİSANS TEZİ

Esra AKCA

İstatistik Anabilim Dalı

İstatistik Programı

Tez Danışmanı: Prof. Dr. Semra ERPOLAT TAŞABAT

OCAK 2022

..... tarafından hazırlanan adlı bu tezin tezi olarak uygun olduğunu onaylarım.

.....
Tez Yöneticisi

Bu çalışma, jürimiz tarafından Anabilim Dalında tezi olarak kabul edilmiştir.

Başkan : _____

Üye : _____

Üye : _____

Üye : _____

Üye : _____

Bu tez, Mimar Sinan Güzel Sanatlar Üniversitesi Enstitüsü tez yazım kurallarına uygundur.



Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü tez yazım klavuzuna uygun olarak hazırladığım bu tez çalışmada;

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel etik kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- ücret karşılığı başka kişilere yazdırmadığımı (dikte etme dışında), uygulamalarımı yaptırmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

....

ÖNSÖZ

Tez danışmanlığımı üstlenerek, bilgi ve deneyimleriyle çalışmalarına destek ve yön veren değerli hocam Prof. Dr. Semra ERPOLAT TAŞABAT'a ve BORUSAN Makina ve Güç Sistemleri San. ve Tic. A.Ş. değerli yöneticilerine verdikleri desteklerden dolayı teşekkür ederim...

Ocak 2022

Esra AKCA

SATIŞ TAHMİNLEMESİNDE HİBRİT BİR YAKLAŞIM: PESTEL, RFM, GRADIENT BOOSTING

ÖZET

Teknolojinin ilerlemesi ve internet kullanımının yaygın hale gelmesi ile birlikte adını sıkça duymaya başladığımız büyük veri kavramı ortaya çıkmıştır. Kısaca yapısal olmayan veri yığını olarak ifade edilen büyük verinin anlamlı bilgilere dönüştürülmesi, gizli kalmış örüntülerin ortaya çıkarılması farklı yöntemlerle gerçekleştirilebilir. Veri madenciliği olarak adlandırılan bu yöntem ve algoritmalar topluluğu istatistik ile birleştirilerek daha anlaşılabilir ve anlamlı çözüm yollarına dönüşmüşlerdir. Bu yöntemlerden bir tanesi de RFM analizidir. RFM analizi **R**ecency **F**requency ve **M**onetary kelimelerinin kısaltması olup davranışa dayalı müşteri segmentasyonunu gerçekleştiren etkili ve pratik bir pazarlama modelidir. Pazarlama stratejilerinin geliştirilmesine yardımcı olan RFM analizinin temel dayanak noktası yakın zamanda alışveriş eden, sık alışveriş eden ve alışverişlerinde yüksek miktarda getiri sağlayan müşterilerin gelecekteki pazarlama kampanyalarına olumlu dönüş yapabilecek potansiyel müşteriler olacağı görüşüdür.

Günümüz şartlarında aynı alanda hizmet üreten veya ürün satan bir çok firma bulunmaktadır. Firmalar oldukça rekabetçi koşullarda yarışmaktadır. Hizmet üreten veya ürün satan firmalar için en önemli durum doğru müşteriyle, doğru zamanda iletişim kurmaktır. Bu sayede firmalar kaynaklarını optimum şekilde tüketmektedir. Aynı zamanda müşteriler için doğru zamanda gelen teklifler önemliyken, ihtiyacının olmadığı zamanda gelecek teklifler, ürün ve hizmet satın alan müşteriler için gereksiz zaman kaybına, müşteri şikayetine sebep olabilmektedir.

Karmaşık veri kümelerinden anlamlı sonuçlar elde etmek ve içerdiği verinin belirli özelliklerine göre veri setinin sınıflara ayrılması için, veri madenciliğinin farklı bir yöntemi olan sınıflandırma algoritmaları geliştirilmiştir. Sınıflandırma problemlerinde yapay zeka alanının alt kümesi olan makine öğrenmesi algoritmaları yüksek düzeyde doğru tahmin ediciler olarak kullanılmaktadır.

Bu çalışmada RFM analizinin ardından; elde edilen analiz çıktısı, şirketlerin çevresel olarak etkilendiği; PESTEL (Political, Economic, Social, Technological, Legal and Environmental) analizinin etkileyici faktörleriyle birlikte satışlar üzerindeki sınıflandırma süreci ele alınacaktır. Sınıflandırma algoritmalarından, Gradient Boosting makine öğrenmesi algoritması uygulanacaktır. Modellerin tahmin değerlerinin doğruluğuyla ilgili performans ölçüm sonucu ve model açıklanabilirliği değerlendirilecektir. Söz konusu model ile verimlilik süreçlerinin daha kaliteli hale getirilmesi ve müşteri ilişkilerini daha doğru yönetecek bir sınıflandırma yönteminin geliştirilmesi hedeflenmiştir.

Anahtar Kelimeler : RFM Analizi, Veri Madenciliği, Gradient Boosting Algoritması, PESTEL Analizi, Müşteri İlişkileri Yönetimi

A HYBRIT APPROACH TO SALES PREDICTION: PESTEL, RFM, GRADIENT BOOSTING

ABSTRACT

With the advancement of technology and the widespread use of the internet, the concept of big data, which we hear frequently, has emerged. In short, the transformation of big data, which is expressed as an unstructured data stack, into meaningful information, and revealing hidden patterns can be realized by different methods. This collection of methods and algorithms, called data mining, has been combined with statistics and turned into more understandable and meaningful solutions. One of these methods is RFM analysis. RFM analysis is the abbreviation of Recency Frequency and Monetary, and it is an effective and practical marketing model that performs behavior-based customer segmentation. The basic premise of RFM analysis, which helps to develop marketing strategies, is the view that customers who shop recently, shop frequently, and generate high returns on their purchases will be potential customers who can make a positive return to future marketing campaigns.

In today's conditions, many companies produce services or sell products in the same field. Firms compete in highly competitive conditions. The most important situation for companies that produce services or sell products is to communicate with the right customer at the right time. In this way, companies consume their resources optimally. At the same time, while offers that come at the right time are important for customers, offers that come when they do not need them can cause unnecessary time loss and customer complaints about customers who purchase products and services.

Classification algorithms, a different method of data mining, have been developed to obtain meaningful results from complex data sets and to classify the data set according to certain characteristics of the data it contains. Machine learning algorithms, which are a subset of artificial intelligence, are used as highly accurate estimators in classification problems.

After RFM analysis in this study; With the data envelopment method of the analysis output obtained, companies are affected environmentally; The classification process on sales will be discussed along with the influencing factors of the PESTEL (Political, Economic, Social, Technological, Legal and Environmental) analysis. Among the classification algorithms, the Gradient Boosting machine learning algorithm will be applied. The performance measurement result and model explainability related to the accuracy of the predicted values of the models will be evaluated. This model, it is aimed to improve production processes and develop a classification method that will manage customer relations more accurately.

Keywords : *RFM Analysis, Data Mining, Gradient Boosting Algorithm, PESTEL Analysis, Customer Relations Management*



İÇİNDEKİLER

Sayfa

ÖNSÖZ	Hata! Yer işareti tanımlanmamış.
İÇİNDEKİLER.....	xii
KISALTMALAR.....	xiii
SEMBOLLER.....	xiv
ÇİZELGE LİSTESİ	xvi
ŞEKİL LİSTESİ	xviii
ÖZET	xx
SUMMARY.....	xxi
1. GİRİŞ	1
1.1 Tezin Amacı	2
1.2 Literatür Araştırması	3
1.3 RFM Analizi	4
1.4 PESTEL Analizi	6
1.5 RFM ve PESTEL Analizi Birlikte Kullanımı	7
1.6 Gradient Boosting.....	8
2. ÖNERİLEN TAHMİN MODELİ.....	9
2.1 Veri Seti.....	10
2.2 RFM Analizine İlişkin Sonuçlar.....	14
2.3 Gradient Boosting Algoritması Uygulama.....	16
2.3.1 Model başarı metrikleri	17
2.3.1.1 ROC eğrisi ve AUC değeri.....	17
2.3.1.2 Hata matrisi	17
2.3.1.3 Matthews korelasyonu.....	19
2.3.2 Değişken ve parametre seçimi.....	19
2.3.2.1 Model değişkenleri.....	19
2.3.2.2 Çapraz validasyon ile parametre optimizasyonu.....	21
2.3.3 Gradient Boosting algoritması sonuçları.....	23
2.3.3.1 Model açıklanabilirliği.....	25
3. SONUÇ	27
3.1 Öneriler.....	28
KAYNAKLAR.....	29

KISALTMALAR

ABD	: Amerika Birleşik Devletleri
AHP	: Analytic Hierarchy Process
ARGE	: Araştırma Geliştirme
AUC	: Area Under Curve
ID	: Identify Definition
Light GBM	: Microsoft Gradient Boosting Algorithm
MCC	: Matthew Correlation Coefficient
MSE	: Mean Square Error
PESTEL	: Political, Economic, Social, Technological, Legal and Environmental
RFM	: Recency, Frequency, Monetary
ROC	: Receiver Operating Characteristic
SHAP	: SHAPley Additive ExPlanations
VUCA	: Volatility, Uncertainty, Complexity and Ambiguity

SEMBOLLER

- Σ : Sayı dizisinin toplamını gösteren matematik gösterimidir.
 α : İstatistiksel sabit terim gösterimi.
 δ : İstatistiksel standart sapma





ÇİZELGE LİSTESİ

Sayfa

Çizelge 1.1 : RFM skor skalası.....	6
Çizelge 2.1 : Orijinal veri seti.....	11
Çizelge 2.2 : PESTEL analizi etkileyici verileri.....	12
Çizelge 2.3 : Düzenlenmiş veri seti.	14
Çizelge 2.4 : RFM analizi örnek müşteri skorları.....	16
Çizelge 2.5 : Model değişkenleri.	20
Çizelge 2.6 : Parametre değerleri.....	22
Çizelge 2.7 : Hata matrisi	23
Çizelge 2.8 : Değişkenlere ait önem düzeyleri	24





ŞEKİL LİSTESİ

	<u>Sayfa</u>
Şekil 1.1 : Hedeflenen iş akış süreci.	2
Şekil 1.2 : PESTEL analizi değişkenleri.....	7
Şekil 2.1 : IBM Watson Studio karar akışı.	15
Şekil 2.2 : Hata matrisi.	18
Şekil 2.3 : Çapraz validasyon.	22
Şekil 2.4 : AUC-ROC eğrisi.	24
Şekil 2.5 : Model açıklanabilirlik grafiği.....	25





SATIŞ TAHMİNLEMESİNDE HİBRİT BİR YAKLAŞIM: PESTEL, RFM, GRADIENT BOOSTING

ÖZET

Teknolojinin ilerlemesi ve internet kullanımının yaygın hale gelmesi ile birlikte adını sıkça duymaya başladığımız büyük veri kavramı ortaya çıkmıştır. Kısaca yapısal olmayan veri yığını olarak ifade edilen büyük verinin anlamlı bilgilere dönüştürülmesi, gizli kalmış örüntülerin ortaya çıkarılması farklı yöntemlerle gerçekleştirilebilir. Veri madenciliği olarak adlandırılan bu yöntem ve algoritmalar topluluğu istatistik ile birleştirilerek daha anlaşılabilir ve anlamlı çözüm yollarına dönüşmüşlerdir. Bu yöntemlerden bir tanesi de RFM analizidir. RFM analizi **R**ecency **F**requency ve **M**onetary kelimelerinin kısaltması olup davranışa dayalı müşteri segmentasyonunu gerçekleştiren etkili ve pratik bir pazarlama modelidir. Pazarlama stratejilerinin geliştirilmesine yardımcı olan RFM analizinin temel dayanak noktası yakın zamanda alışveriş eden, sık alışveriş eden ve alışverişlerinde yüksek miktarda getiri sağlayan müşterilerin gelecekteki pazarlama kampanyalarına olumlu dönüş yapabilecek potansiyel müşteriler olacağı görüşüdür.

Günümüz şartlarında aynı alanda hizmet üreten veya ürün satan bir çok firma bulunmaktadır. Firmalar oldukça rekabetçi koşullarda yarışmaktadır. Hizmet üreten veya ürün satan firmalar için en önemli durum doğru müşteriyle, doğru zamanda iletişim kurmaktır. Bu sayede firmalar kaynaklarını optimum şekilde tüketmektedir. Aynı zamanda müşteriler için doğru zamanda gelen teklifler önemliken, ihtiyacının olmadığı zamanda gelecek teklifler, ürün ve hizmet satın alan müşteriler için gereksiz zaman kaybına, müşteri şikayetine sebep olabilmektedir.

Karmaşık veri kümelerinden anlamlı sonuçlar elde etmek ve içerdiği verinin belirli özelliklerine göre veri setinin sınıflara ayrılması için, veri madenciliğinin farklı bir yöntemi olan sınıflandırma algoritmaları geliştirilmiştir. Sınıflandırma problemlerinde yapay zeka alanının alt kümesi olan makine öğrenmesi algoritmaları yüksek düzeyde doğru tahmin ediciler olarak kullanılmaktadır.

Bu çalışmada RFM analizinin ardından; elde edilen analiz çıktısı, şirketlerin çevresel olarak etkilendiği; PESTEL (Political, Economic, Social, Technological, Legal and Environmental) analizinin etkileyici faktörleriyle birlikte satışlar üzerindeki sınıflandırma süreci ele alınacaktır. Sınıflandırma algoritmalarından, Gradient Boosting makine öğrenmesi algoritması uygulanacaktır. Modellerin tahmin değerlerinin doğruluğuyla ilgili performans ölçüm sonucu ve model açıklanabilirliği değerlendirilecektir. Söz konusu model ile verimlilik süreçlerinin daha kaliteli hale getirilmesi ve müşteri ilişkilerini daha doğru yönetecek bir sınıflandırma yönteminin geliştirilmesi hedeflenmiştir.

Anahtar Kelimeler : RFM Analizi, Veri Madenciliği, Gradient Boosting Algoritması, PESTEL Analizi, Müşteri İlişkileri Yönetimi

A HYBRIT APPROACH TO SALES PREDICTION: PESTEL, RFM, GRADIENT BOOSTING

SUMMARY

With the advancement of technology and the widespread use of the internet, the concept of big data, which we hear frequently, has emerged. In short, the transformation of big data, which is expressed as an unstructured data stack, into meaningful information, and revealing hidden patterns can be realized by different methods. This collection of methods and algorithms, called data mining, has been combined with statistics and turned into more understandable and meaningful solutions. One of these methods is RFM analysis. RFM analysis is the abbreviation of Recency Frequency and Monetary, and it is an effective and practical marketing model that performs behavior-based customer segmentation. The basic premise of RFM analysis, which helps to develop marketing strategies, is the view that customers who shop recently, shop frequently, and generate high returns on their purchases will be potential customers who can make a positive return to future marketing campaigns.

In today's conditions, many companies produce services or sell products in the same field. Firms compete in highly competitive conditions. The most important situation for companies that produce services or sell products is to communicate with the right customer at the right time. In this way, companies consume their resources optimally. At the same time, while offers that come at the right time are important for customers, offers that come when they do not need them can cause unnecessary time loss and customer complaints about customers who purchase products and services.

Classification algorithms, a different method of data mining, have been developed to obtain meaningful results from complex data sets and to classify the data set according to certain characteristics of the data it contains. Machine learning algorithms, which are a subset of artificial intelligence, are used as highly accurate estimators in classification problems.

After RFM analysis in this study; With the analysis output obtained, companies are affected environmentally; The classification process on sales will be discussed along with the influencing factors of the PESTEL (Political, Economic, Social, Technological, Legal and Environmental) analysis. Among the classification algorithms, the Gradient Boosting machine learning algorithm will be applied. The performance measurement result and model explainability related to the accuracy of the predicted values of the models will be evaluated. This model, it is aimed to improve production processes and develop a classification method that will manage customer relations more accurately.

Keywords : *RFM Analysis, Data Mining, Gradient Boosting Algorithm, PESTEL Analysis, Customer Relations Management*



1. GİRİŞ

Global seviyede üretilen ve depolanan verinin miktarı hayal edilemeyecek kadar çoktur ve her geçen gün daha da büyümektedir. Bu veri yığınlarının anlamlı verilere dönüştürülerek yararlı kararların alınmasında yardımcı olabilmelerini sağlamak ancak bilimsel analiz teknikler ile temizlenmesi, düzenlenmesi, modellenmesi ve yorumlanması ile mümkündür. Kısaca veri madenciliği olarak nitelendirilebilecek bu sürecin model aşaması veri analizi ile gerçekleştirilebilir. Veri analizi, biriken verinin amaca uygun modeller ile mantıklı, faydalı ve etkin sonuçlara dönüşmesini sağlayan yöntemler bütünüdür. Söz konusu modeller ve yöntemler ele alınan karar problemine göre farklı olabilmektedir. Bu yöntemlerden bir tanesi de RFM'dir.

Karar destek sisteminin bir parçası olan RFM müşterileri segmente ederek, amaca uygun kitleyi en doğru şekilde bulmak için yaklaşık yarım asırdır kullanılan faydalı, basit ve güçlü bir tüketici, Müşteri İlişkileri Yönetimi (CRM) uygulama modelidir. RFM analizi, davranışa dayalı müşteri segmentasyonu için geçerli bir pazarlama modelidir. Diğer taraftan PESTEL olarak adlandırılan bir diğer analiz de stratejik planlamada kullanılan etkili bir yöntemdir. Makine öğrenmesi sınıflandırma algoritmaları ise, müşterilerin geçmişteki verilerindeki örüntülerden hareketle geleceğe dair tahminlerde bulunan güçlü istatistiksel yöntemlerdir.

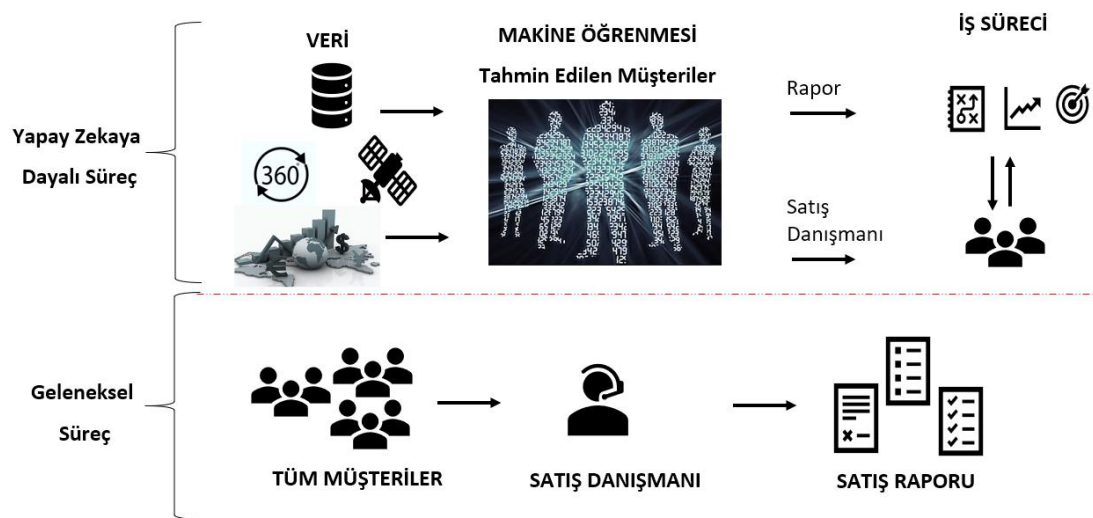
Çalışmada önerilen model ile, RFM analizi çıktılarıyla birlikte PESTEL analizindeki etkileyici değişkenleri kullanılarak, sınıflayıcı makine öğrenmesi modeli geliştirilmiştir. Bu yeni model sayesinde müşterilerin işlem geçmişlerine göre "ne kadar yakın", "ne sıklıkta", "ne kadara satın almışlar" ve "sosyo-ekonomik hassasiyetleri nedir" şeklinde üretilen değişkenler ile tahminleme sağlanmıştır. Böylelikle, müşterilerin firmaya olan eğilimleri tespit edilerek kampanya ve hizmete yanıt verme olasılıkları daha yüksek olan müşteri profillerinin belirlenmesine imkân verilmiştir. Satış tahminleme modeli ile müşteri kayıplarının önüne geçmek, ürün ve hizmet bazında tüketicinin gelecekteki davranışları üzerine tahminlerde bulunmak hedeflenmektedir. Bu sayede Pazarlama alanında şirketlere kolaylık sağlanarak özellikle maliyet ve zamandan tasarruf etmelerine yol gösterilmektedir. Büyük verinin yönetilmesi ve veriden anlamlı bilgilerin çıkarılması şirketlerin hayatlarını sürdürebilmeleri için oldukça önemlidir. Özellikle müşteriyi tanıma, müşteri sürdürülebilirliği ve bağlılığını arttırmak bu sayede mümkündür.

İzleyen bölümlerde ilk önce RFM analizi, PESTEL analizi ve Gradient Boosting makine öğrenmesi algoritması hakkında bilgi verilecek ardından RFM modeline PESTEL analizinin etkileyici değişkenleri entegre edilerek geliştirilen satış tahminleme modeli; model açıklanabilirliği ile birlikte detaylandırılacaktır. Geliştirilen bu model

BORUSAN Makina ve Güç Sistemleri şirketinin 01.2014 ile 09.2019 tarihleri arasındaki aylık verileri üzerinde sınanacaktır.

1.1 Tezin Amacı

Firmalarda çalışan müşteri temsilcilerinin ziyarette bulunacağı müşterilerin belirlenmesi önem arz etmekte olup, rutin olarak tüm müşterilerle temasa geçmek her zaman mümkün değildir. Bu durum müşteri temsilcileri için zaman ve operasyonel maliyete sebep olmakla birlikte, ihtiyacı bulunan müşteriyle temasa geçilememesi aksine ihtiyacı bulunmayan müşteri ile ilgilenilmesi gibi müşteri şikayetlerine hatta müşteri kaybına neden olmaktadır. Farklı bir bakış açısıyla yeni bir müşteri kazanma maliyeti; mevcut müşteriye hizmet vermenin maliyetinden daha fazladır. Zamanında gerçekleşmeyen bir müşteri ziyareti müşterinin o anki ihtiyacını karşılayamamasına sebep olmaktadır. Bunun yanında müşterinin başka firmalara yönelmesine olanak sağlamaktadır. Müşteri temsilcisinin zamanında yapmadığı bu iletişimden dolayı firmanın pazarda gücünü kaybetmesiyle sonuçlanmaktadır. Başarısızlığın bir kısmı müşterileri görmezden gelmekten kaynaklanırken, diğer bir kısmı ise bir iş modelinin olmamasından kaynaklanmaktadır. Aşağıdaki şekil 1.1’de geleneksel süreçle işleyen pazarlama stratejisi ve önerilen yapay zeka modeli ile işleyen sürece ait karşılaştırma görseline yer verilmiştir. Geleneksel süreçte; satış danışmanlarının uygun gördüğü müşterileri ziyaret etmesiyle satış süreci yönetilmektedir. Kısıtlı kaynak ve zamandan dolayı portföydeki tüm müşterilerle etkileşim kurulamamaktadır. Önerilen yapay zeka modeli ile satın alım ihtimali yüksek olan müşteriler tespit edilerek daha kısa zamanda daha odaklı müşteriler satış danışmanlarına yönlendirilmektedir.



Şekil 1.1: Hedeflenen iş akış süreci

Tarihsel olarak birikmiş müşteri verisinden hareketle, gelecekteki ihtiyaçlarını tespit etmek hem müşteri temsilcisinin eforunu ilgili müşteriye harcamasını sağlayacağı gibi müşterinin de memnuniyetini arttıracaktır.

Pazar karakteristiği ve sektörel incelemelerde müşterilerin alımlarının; sezonsal durumlara, ekonomik göstergelere ve gelişim endekslerine bağlı olarak gerçekleştiği görülmektedir. Bu tez ile müşterilerin söz konusu alışkanlıkları RFM analizi değişkenleri ve PESTEL analizi değişkenlerinin ortak kullanımı ile veriye dayalı güçlü makine öğrenimi yöntemi olan Gradient Boosting ile tahmin edilmiştir.

1.2 Literatür Araştırması

Literatürde müşteri ilişkileri yönetimi alanında RFM analizi ile özdeşleşen önemli çalışmalardan biri ABD'nin önemli perakendecilerinden Target Company'nin "Hamilelik Tahmin Puanı" uygulamasıdır. Uygulamada öncelikle her müşteriye bir kimlik numarası (ID) atanıyor. Böylelikle yapılan tüm alışverişler bu müşteri ID'si ile ilişkilendirilerek kayıt altına alınıyor. Oluşan tarihçe incelenerek hamilelik evresine göre satın alınan 25 farklı ürün belirleniyor. Müşteri bu ürünlerden aldıkça ilişkilendirme ve anlamlandırma devam ediyor. Hedef puana ulaşıldığında ise ilgili müşteriye ürün tanıtım ve indirim kuponları gönderiliyor. Target Company bu uygulama sayesinde 2002-2010 yılları arasında gelirini 23 milyar dolar arttırmayı başarmıştır (Murat, 2017).

RFM'in literatürde farklı uygulamalarına da rastlamak mümkündür. Bunlardan bir tanesi de RFM ile kümeleme işlemini gerçekleştiren uygulamalardır. (Hossaini et al. ,2010) müşteri ilişkileri yönetimini (CRM) geliştirmek adına ağırlıklı RFM modelini K-Means ile birleştirmişlerdir.

Farklı sektörel alanlarda kullanılmış olan RFM modelinin bir örneği de, geri dönüşüm alanında olmuştur. Metal ürün şirketi müşterilerinin hesaplanan RFM skorlarına göre geri dönüşüme en fazla katkıda bulunabilecek olanlar seçilerek, geri dönüşüm sektöründe daha yüksek fayda getirisi hedeflenmiştir (Semra Erpolat Taşabat & Esra Akca, 2020).

Chuang, H. ve Shen ise yaptıkları çalışmada Tayvan'da bir işverenin uzun vadeli kârlı müşterilerinin sadakatini pekiştirmek amacıyla geliştirdikleri müşteri değer analizinde RFM modeli ve K-Means yöntemini kullanmışlardır. Yöntemde öncelikle R, F, M değişkenlerinin önem ağırlıkları Analitik Hiyerarşi Prosesi (AHP) ile belirlenmiş hesaplanan müşteri yaşam değerleri ile müşteriler değerlendirilmiştir. Daha sonra kümeleme analiziyle benzer müşteri segmentleri oluşturmuşlardır (Chuang, H. & Shen, C.,2008).

RFM ile sıralama yapan çeşitli çalışmalar da gerçekleştirilmiştir. Bunlardan bir tanesi (Olson et al., 2009) tarafından gerçekleştirilen çalışma olup bu çalışmada müşterilerin belirli bir ürün tanıtımına yanıt olasılıklarını analiz etmek için sıralama ve RFM modellerinden yararlanılmıştır. Bir diğeri ise Cheng ve Chen tarafından gerçekleştirilen çalışmadır. Bu çalışmada ise üç veri madenciliği tekniği olan lojistik regresyon, karar ağaçları ve sinir ağları algoritmaları arasında müşteri segmentasyonunun göreceli değişimi ele alınmıştır (Cheng ve Chen, 2009). Ha ise bir sonraki müşterilerin RFM değerlerini mevcut değerlerden tahmin etmek, RFM değerlerinin zaman içindeki değişimlerini ve geçiş yollarını görmek için sınıflandırma karar ağacı tekniğini kullanmıştır (Ha, 2007).

Birant yaptığı çalışmada veri madenciliğinde entegre bir model önerisinde bulunmuştur (Birant, 2011). Önerilen model beş ana bölümden oluşmaktadır; veri ön işleme, RFM analizi, müşteri değerlendirme süreçleri ile segmentasyon ve tahmindir. Yaklaşımın her bir kısmı birbiri ardına uygulanır. Her parçanın çıkışı bir sonraki bölüm(ler) in girişidir.

1.3 RFM Analizi

İlk defa Bult ve Wansbeek tarafından literatüre kazandırılmış olan RFM analizi günümüzde artık veri madenciliğinde önemli bir yere sahip modeller arasındadır (Bult and Wansbeek, 1995). Özellikle müşteri ilişkilerinde başarılı sonuçların ortaya çıkarılmasında oldukça yararlı bir model haline gelmiş olup genellikle e-ticaret ve pazarlamada indirim ve kampanya stratejilerinin geliştirilmesinde, bankacılıkta kredi ve hisse senedi satışlarında kullanılmaktadır.

RFM, Recency-Frequency-Monetary değişkenlerinin baş harflerinin birleşiminden oluşturulmuştur. Recency müşterinin en yakın zamandaki alışverişini, frequency

müşterinin alışveriş sıklığını ve monetary ise müşterinin toplam alışveriş tutarını ifade etmektedir.

RFM modelinin merkezinde Pareto Prensibi bulunmaktadır. Pareto prensibine göre; sonuçların % 80'i nedenlerin % 20'sinden gelmektedir. Benzer şekilde % 20 müşterinin toplam gelirin % 80' ine katkıda bulunduğu görüşüne dayanmaktadır. Bu durum kritik müşteri segmentlerine odaklanmanın daha yüksek yatırım getirisi sağlayabileceğine işaret eder.

Genel olarak RFM analizi: Müşteriniz kimdir? Hangi müşterileriniz en karlıdır? Hangi müşterileriniz yakın zamanda sizi ziyaret etti? Hangi müşterileriniz en sadıktır? Hangi kampanyalara hangi müşteriler nasıl tepki verdi? Hangi müşterileriniz kayıp? Hangi müşterilerinize kaybetmeye yüz tutmuş? vb. gibi sorulara cevap bularak uygun kararların verilmesini amaçlar. RFM, optimum hedefleme nedeniyle pazarlama maliyetlerini düşürür, kontrollü hedefleme sayesinde ise müşterilerden gelen olumsuz tepkileri azaltmaktadır.

RFM'in klasik hesaplama yaklaşımı formül (1.1)'de verilmiştir.

$$(RecencyScore \times RecencyWeight) + (FrequencyScore \times FrequencyWeight) + (MonetaryScore \times MonetaryWeight) \quad (1.1)$$

Formül (1.1)'de bulunan 'Recency Score, Frequency Score, Monetary Score'; tarih, frekans ve tutarların her birinin kendi içinde %20'lik şekilde 5 aralığa ayrılarak, en yüksek %20'lik kısma 5, en yüksek 2. kısma 4, 3. kısma 3, 4. kısma 2, olarak ve son en düşük ya da en geride kalmış %20'lik parçaya 1 puan verilmek suretiyle hesaplanmaktadır.

Şöyleki Formül (1.1)'de verilen formüldeki her bir değişkene ilişkin ağırlık değerleri kişilere ve teorilere göre farklılık gösterebilmektedir. Formül (1.2)'de Recency, Frequency ve Monetary değişkenleri için farklı ağırlık değerlerinin kabul edildiği modele yer verilmiştir. Formül (1.2)'ye göre RFM analizinde en önemli değişken Recency ardından Frequency ve son olarak Monetary olarak ele alınmıştır. Bu model kapsamında minimum skor 111, maximum skor 555 olarak hesaplanmaktadır (Yi Fan, 2016).

$$(RecencyScore \times 100) + (FrequencyScore \times 10) + (MonetaryScore) \quad (1.2)$$

RFM analizinde sık olarak kullanılan ve müşteri değer ölçütü olarak adlandırılan 1'den 5'e kadar sayılardan oluşan bir skala mevcuttur. Bu skala değerleri: 1: En zayıf müşteri, 2: Zayıf müşteri, 3: Orta müşteri, 4: Önemli müşteri ve 5: En önemli müşteri şeklindedir. Müşterilere atanan bu skala değerleri analizin gerçekleştirileceği şirketin uzman görüşü doğrultusunda atanır ve analizlere dahil edilir. Çünkü hangi aralıkların yenilik, sıklık ve parasal değerler için ideal olduğuna şirketler karar vererek gerekli gördükleri noktada değişikliğe gidebilirler.

Ancak günümüzde RFM analizini yapan paket programlar birimlere skala değeri atama işlemini uzman görüşüne gereksinim duymadan gerçekleştirebilmektedir. Söz konusu bu programlar birimlere skala değeri atarken %100 lük değeri %20'lik alt parçalara ayırmak suretiyle gerçekleştirmektedir. Skor atama işlemi olarak adlandırılan bu işlemin ayrıntılarına çizelge 1.1'de yer verilmiştir (Cheng ve Chen, 2009). Recency, Frequency ve Monetary değerlerinin %20'lik alt parçaları ve karşılık gelen skorlarına yer verilmiştir.

Çizelge 1.1: RFM skor skalası

Skor	R - Recency (%)	F - Frequency (%)	M - Monetary (%)
5 Skor	0-20	0-20	0-20
4 Skor	20-40	20-40	20-40
3 Skor	40-60	40-60	40-60
2 Skor	60-80	60-80	60-80
1 Skor	80-100	80-100	80-100

1.4 PESTEL Analizi

Political, Economic, Social, Technological, Environmental ve Legal kelimelerinin baş harflerinden oluşan PESTEL kelimesi, bir stratejinin başarılı olup olamama ihtimallerini belirlemede çevresel faktörlerden yararlanarak bir çerçeve oluşturmak için kullanılan analize verilen addır. Şekil 1.2'de PESTEL analizi oluşturan, etkileyici başlıklar gösterilmiştir.



Şekil 1.2: PESTEL analizi değişkenleri

PESTEL kelimesini oluşturan bu terimleri kısaca şu şekilde açıklanmaktadır: Political terimi, devletlerin iş çevresine olan etkileri hakkında bilgi vermektedir. Economic terimi, kurumların doğrudan ve uzun vadede hangi oranlarda etkilendiği üzerine odaklanmaktadır. İlgili ülkedeki ekonomik duruma bağlı olarak, şirketlerin performans artış veya düşüş miktarları analiz edilir. Ekonomik büyüme, döviz kuru, enflasyon oranları, faiz oranları, işletmenin ve tüketicinin harcanabilir gelir miktarı Economic terimi çatısı altında incelenebilir. Social terimi, kültürel etkilerden ve demografiden bahsetmektedir. Nüfus artışı, yaş dağılımı, cinsiyet oranları, davranış biçimleri, kariyer tutumları, sağlık durumları sosyo kültürel faktörlere örnek gösterilebilir. Technological terimi ile teknolojinin etkileri incelenmektedir. Bir işletmenin ilgilendiği sektöre gelişen teknoloji, yapılan inovasyonlar ve ARGE faaliyetleri teknolojik faktörlere dahil edilmektedir. Environmental terimi, genellikle 'yeşil' konular üzerine yoğunlaşmaktadır. Çevre kirliliği veya atıklar, hammaddenin ne kadar miktarda olduğu Environmental faktörü kapsamında ele alınabilir. Legal terimi ile yasal kısıtlamalar veya değişiklikler ele alınmaktadır. Sağlık, güvenlik yasaları, çevre kanunu, iş kanunu, reklam kanunu, ticaret hukuku, vergi hukuku, isim hakkının alınması gibi faktörler legal terimi kapsamındadır.

1.5 RFM ve PESTEL Analizi Birlikte Kullanımı

Şirketler değişkenleri ülke ve etki seviyelerine göre değerlendirmenin ardından ilgili analiz ve modellerde kullanarak gelecekle ilgili dış analizle desteklenmiş daha etkili

ve stratejik sonuçlara ulaşabilmektedirler. Değişimden yararlanarak, değişime karşı koyan faaliyetlerde başarılı olma şansı fazlalaşmaktadır. PESTEL analizinin iyi kullanımı; VUCA (Volatility, Uncertainty, Complexity and Ambiguity) ortamında, kontrol dışında gerçekleşen sebeplerden dolayı başarısızlığa mahkûm edecek hareketlerde bulunmanızdan kaçınmanıza yardımcı olmaktadır. Bu çalışmada; RFM değişkenlerine ek, PESTEL analizi faktörleriyle müşterileri çevresel etkenlerle müşteri hassasiyetlerini ölçerek; fırsatlardan yararlanma, tehlikeleri azaltma amacıyla doğru zamanlamayla karar vermeye olanak sağlamaktadır.

1.6 Gradient Boosting

Denetimli makine öğrenmesi, bilgisayarların açık bir şekilde programlanmadan etiketli eğitim verilerinden öğrenmesini sağlar. Bu denetimli öğrenmenin görevleri regresyon veya sınıflandırmadır. Veri madenciliğinin üç önemli bileşeni olan istatistik, makine öğrenmesi ve veri tabanı teknolojilerini temsil edecek şekilde gradient boosting algoritması, uygulamada kullanılacak sınıflandırma tekniği olarak belirlenmiştir.

Gradient Boosting, regresyon veya sınıflama modelleri için kullanılabilen bir öngörü algoritmasıdır. Genel olarak model hatalarının sıralı modeller ve zayıf tahminciler ile tekrar tahmin edilmesi ve oluşan modellerin belirli bir ağırlıkla toplanması sonucunda varyansın azaltılmasına dayanır (Chen, 2014). Çalışma kapsamında geliştirilen sistemde sıralı modeller kullanılıyor olması ve tahmin edicilerin oldukça küçük olması bu algoritmanın kullanılmasının tercih nedeni olmuştur. Aşağıdaki formül akışında, karar ağaçlarında genelleştirilmiş gradient boosting algoritması gösterilmiştir (Friedman, 1999). Gradient boosting, regresyon ve sınıflandırma problemleri için bir makine öğrenmesi tekniğidir. Gradient boosting algoritmaları karar ağacının büyütülmesi sırasında hem XGBoost hem de LightGBM yaprak bazlı büyüme stratejisini kullanır (A. Swalin, 2018).

Bu algoritma, zayıf tahmin modellerinin bir araya gelmesiyle tipik olarak karar ağaçlarının oluşturduğu güçlü bir model oluşturur. Denetimli makine öğrenmesi algoritmalarının amacı, bir kayıp fonksiyonu tanımlamak ve en aza indirmektir. LightGBM algoritması yaprakları bölerek tahmin gücünü arttırmaktadır (X. Team, 2019). Gradient Boosting algoritması için matematiğin nasıl çalıştığı formül(1.3),

formül(1.4), formül(1.5)'de gösterilmiştir. Aşağıdaki gibi tanımlanan karesel hataya (MSE) sahip olduğu varsayalım:

$$Loss = MSE = \sum (y_i - y_i^p)^2 \quad (1.3)$$

Burada y_i = hedef değeri, y_i^p = tahmin değeri, $L(y_i, y_i^p) = Loss$ fonksiyonu

Tahminler, kayıp fonksiyon (MSE) minimum olacak şekilde istenmektedir. Gradyan iniş kullanarak ve tahminleri öğrenme oranına göre güncelleyerek, MSE'nin minimum olduğu değerler bulunabilmektedir.

$$y_i^p = y_i^p + \alpha * \delta \sum_{i=1}^n (y_i - y_i^p)^2 / \delta y_i^p \quad (1.4)$$

$$y_i^p = y_i^p - \alpha * 2 * \sum_{i=1}^n (y_i - y_i^p) \quad (1.5)$$

Dolayısıyla, temel olarak tahminlerin, artıkların toplamının 0'a yakın (veya minimum) olduğu ve tahmin edilen değerlerin gerçek değerlere yeterince yakın olacağı şekilde güncellenmesi gerekmektedir.

2. ÖNERİLEN TAHMİN MODELİ

Bu çalışmada yukarıda sözü edilen RFM modeline yeni bir hibrit model önerisinde bulunulmuştur. Önerilen bu model RFM ve PESTEL analizlerinin birlikte değerlendirilmesiyle Gradient Boosting algoritması kullanılarak satış tahminlemesi üzerinedir. RFM analizi sürekli değişkenlerine PESTEL analizinde yer alan etkileyici olduğu düşünülen faktörler entegre edilmiştir. Bu faktörleri açıklayacak olursak; gelişme ve büyüme ekonomisiyle doğru orantılı olarak gerçekleşmektedir. Özellikle döviz kuru ve faiz oranları üzerinde oluşan spekülasyon ve dalgalanmaların dikkatli biçimde takip edilmesi büyük önem taşımaktadır. Önerilen çalışmada PESTEL analizindeki ekonomik duyarlılık adına etkili olan Döviz kuru ve faiz oranı değişkenleri dahil edilmiştir. Ekonomideki gelişmelerin takip edilmesi suretiyle ekonomik hareketlilik üzerinden müşterilere daha doğru pazarlama seçeneklerinin sunulması mümkündür. Bunlara ek olarak PESTEL analizindeki Politik duyarlılık adı

altında yer alan ve şirketin çalışma alanından dolayı üzerinde etkisi olan inşaat-yapı ihale verileri dahil edilmiştir. Aynı şekilde PESTEL analizindeki Çevresel duyarlılık adı altında yer alan sezonsallık ölçütü dahil edilmiştir. Böylelikle müşterilerin daha iyi tanınması ve dolayısıyla model kalitenin artırılması hedeflenmiştir. Literatürde ilk kez RFM modelinin PESTEL entegrasi sağlanmış olup; RFM skora mantığından farklı olarak, makine öğrenmesi algoritması kullanılmıştır. Aşağıda önerilen bu modelin türetilmesinde uygulanan yaklaşımlar detaylandırılmıştır.

2.1 Veri Seti

Tezde ele alınan veri, BORUSAN Makina ve Güç Sistemleri şirketine ait 01/01/2014 ile 30/09/2019 tarihleri arasındaki orijinal verileri kapsamaktadır.

BORUSAN Makina ve Güç Sistemleri şirketi müşterilerine ait veri yığınının çalışmada kullanılmak üzere seçilen ürün filtreleri: yeni ürün ve ikinci el üründür.

Çalışma kapsamında belirlenen veri setinde eksik veya kayıp, uç veya aykırı değerler için uzman görüşlerine bağlı olarak eldeki verinin tahmin doğruluğunu arttırmak için bazı varsayımlarda bulunulmuştur. Şöyle ki:

Eksik veya Kayıp Değerler: Yapılan süzgeçleme sonucunda elde edilen veri grubunun veri düzeltme ve düzenleme aşamasında uzman görüşü alınmak suretiyle herhangi bir birime (müşteriye) ait minimum tek bir değişkende bile verinin olmaması durumunda bu birimin (müşterinin) veri setinden çıkarılması kararlaştırılmıştır. Her ne kadar kayıp verilerin istatistiksel yöntemler ile doldurulması mümkün olsa da çalışma kapsamında müşteri odaklı bir analiz yapılmak istenmesinden dolayı başka bir ifade ile şirketin tüm müşterileri için genel bir yorum yapmaktansa her müşteri için müşteriye özel bir yorum yapılması amaçlandığından eksik değerleri olan birimler analiz dışı bırakılmıştır. Ayrıca herhangi bir satın alımı alışkanlığı oluşmayan müşterilerin model tarafından eğitilmesi mümkün olmadığından, satın alım geçmişi olmayan veya 1 defa satın alım yapmış müşteriler model dışı bırakılmıştır.

Firmanın kendi dağıtıcıları (dealer), yüksek miktarda alım satım yaptıkları ve her ay düzenli müşteriler olup, modelin eğitimini yanıtacağı için makine alım satımı yapan bu türdeki müşteriler de model dışı bırakılmıştır.

Uç veya Aykırı Değerler: Çalışma kapsamında uç veya aykırı değerler analizden çıkarılmamıştır. Çünkü ele alınan üç değişkenin birbirleriyle bağlantılı olmaması ve

her birinin tek başına bile birim bazında anlam ifade etmesi, aykırı veya uç değer içermelerinin satış verisi olması nedeniyle anlamlı olabileceği ihtimali, herhangi bir değişkendeki uç veya aykırılıkların diğer değişkenlerle birlikte anlam kazanabileceği ihtimali olabileceği göz önünde bulundurulmuştur.

Amaç: BORUSAN Makina ve Güç Sistemleri şirketi tarafından sunulacak yeni bir hizmet, öneri ya da kampanyaya eğilimli, karşılık vermesi muhtemel olan müşterileri belirlemektir.

Veri seti; Toplam satış veri sayısı 11.650 olup müşteri sayısı ise 3.261'dir. Sunucular üzerinde, veritabanında; müşteri bilgileri ve fatura detayları yer almaktadır. Verinin orijinal hali ile RFM, PESTEL analizi değişkenleri önerilen modelin değerlendirilebilmesi için sırasıyla çizelge 2.1, çizelge 2.2 ve çizelge 2.3'de yer verilmiştir. Çizelge 2.1'de müşteri bilgileri ve fatura detaylarına ait veritabanından gelen veriler bulunmaktadır. Çizelge 2.2'de PESTEL analizi şirketin iş bilgisi ve iş süreçleri doğrultusunda etkilendiği değişkenler bulunmaktadır. Bu değişkenler Türkiye Cumhuriyeti Merkez Bankası Elektronik Veri Merkezi sitesi üzerinden elde edilmiştir. Çizelge 2.3'de RFM modeline göre düzenlenmiş halindeki hesaplanan değişkenler yer almaktadır. Çizelge 2.1'de, çizelge 2.2'de ve çizelge 2.3'de yer alan terim detayları:

Müşteri ID: Şirketten alışveriş yapan her bir müşteri.

Bölüm: Ürünün statüsü.

Recency: Her bir müşterinin son alışveriş tarihi.

Frequency: Her bir müşterinin alışveriş sıklığı.

Monetary: Her bir müşterinin alışveriş sonucunda ödediği toplam miktar.

Çizelge 2.1: Orijinal veri seti

Müşteri ID	Bölüm	Gelir	Tarih
ID 1	Yeni Ürün	200,782.66 EUR	07.2017
ID 2	İkinci El	70,122.96 EUR	05.2018
ID 3	İkinci El	146,430.01 EUR	07.2015
ID 4	Yeni Ürün	180,030.00 EUR	05.2019
ID 5	Yeni Ürün	459,312.01 EUR	04.2016

ID 6	İkinci El	150,000.00 EUR	04.2015
ID 7	İkinci El	120,130.00 EUR	03.2016
ID 8	İkinci El	145,220.14 EUR	02.2016
ID 9	Yeni Ürün	100,110.00 EUR	01.2017
ID 10	Yeni Ürün	60,450.01 EUR	05.2017
---	---	---	---
ID 3252	İkinci El	76,198.00 EUR	03.2018
ID 3253	İkinci El	123,986.00 EUR	08.2017
ID 3254	İkinci El	145,000.00 EUR	09.2015
ID 3255	Yeni Ürün	290,000.00 EUR	01.2015
ID 3256	Yeni Ürün	164,000.00 EUR	07.2016
ID 3257	İkinci El	65,489.00 EUR	08.2016
ID 3258	İkinci El	87,650.00 EUR	03.2016
ID 3259	İkinci El	102,560.00 EUR	09.2017
ID 3260	Yeni Ürün	45,750.00 EUR	04.2015
ID 3261	Yeni Ürün	78,500.00 EUR	01.2016

Çizelge 2.2: PESTEL analizi etkileyici verileri

Tarih	Dolar Kuru	Konut Kredisi Faiz Oranı	Aylık İhale Sayıları
01.2014	2,22	11,3	46
02.2014	2,22	13,03	65
03.2014	2,23	13,51	60
04.2014	2,13	13,52	67
05.2014	2,1	12,94	90
06.2014	2,12	12,29	150
07.2014	2,13	11,46	180
08.2014	2,17	11,06	178
09.2014	2,21	10,91	250

10.2014	2,27	10,75	143
---	---	---	---
10.2018	6,39	25,18	349
11.2018	5,88	28,95	260
12.2018	5,39	28,63	143
01.2019	5,32	27,82	80
02.2019	5,39	26,22	49
03.2019	5,28	22,98	38
04.2019	5,46	18,16	278
05.2019	5,75	17,7	348
06.2019	6,07	20,45	645
07.2019	5,83	21,81	560

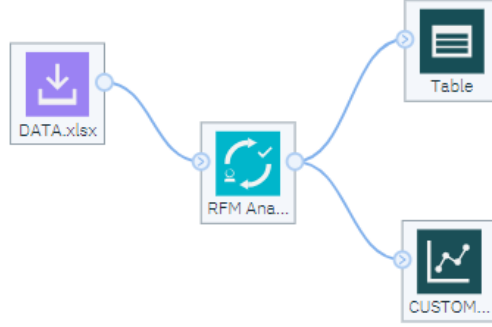
Çizelge 2.3: Düzenlenmiş veri seti

Müşteri ID	Recency	Frequency	Monetary
ID 1	08.2018	22	2,180,424.32 EUR
ID 2	12.2018	35	8,120,111.90 EUR
ID 3	11.2016	2	300,360.76 EUR
ID 4	07.2019	17	1,320,444.11 EUR
ID 5	07.2018	3	2,012,344.67 EUR
ID 6	05.2016	11	500,360.76 EUR
ID 7	04.2017	4	1,590,424.32 EUR
ID 8	05.2016	5	5,321,043.10 EUR
ID 9	01.2018	9	890,980.33 EUR
ID 10	06.2018	5	300,122.00 EUR
---	---		
ID 3252	01.2019	3	291,875.00 EUR
ID 3253	12.2017	2	535,000.00 EUR
ID 3254	11.2017	10	4,832.987,00 EUR
ID 3255	04.2018	11	5,202,765.00 EUR
ID 3256	12.2016	4	300,330.00 EUR
ID 3257	09.2018	12	5,100,900.13 EUR
ID 3258	05.2018	3	500,367.00 EUR
ID 3259	11.2018	2	467,980.00 EUR
ID 3260	12.2017	8	465,976.44 EUR
ID 3261	01.2019	13	1,678,124.22 EUR

2.2 RFM Analizine İlişkin Sonuçlar

Çalışma kapsamında analizleri gerçekleştirmek üzere IBM Watson Studio'dan yararlanılmıştır. IBM Watson Studio'da ölçeğe uygun modeller oluşturmak, bunları eğitmek üzere verilerle iş birliğine dayalı ve kolay bir biçimde çalışmalarını için gereken araçlar sağlanmaktadır. Bu şekilde veri biliminin daha kısa sürede hayata geçirilebilmesine imkân tanınmaktadır. Tahmine dayalı analitiği, makine öğrenimi süreçlerini modernleştirmeye ve değer elde etme süresini hızlandırmaya yardımcı olur. Tahmine dayalı analitik, geçici istatistiksel analiz, tahmine dayalı modelleme, veri madenciliği, metin analitiği, optimizasyon, gerçek zamanlı puanlama ve makine öğrenimini kapsayan gelişmiş analitik yeteneklerini kullanır. Bu araçlar, kuruluşların verilerdeki modelleri keşfetmesine ve geçmişte ne olduğunu bilmenin ötesinde, bir sonraki adımda olacakları tahmin etmelerine yardımcı olmaktadır. Çalışmada veriler IBM Watson Studio'ya aktarıldıktan sonra Modeler flow seçilerek analizlere

başlanmıştır. Çizelge 2.3’de programa aktarılmış olan veri setine klasik RFM Analiz modeli entegre edilerek şekil 2.1 ‘deki karar diyagramı oluşturulmuştur. Çizelge 2.4’de, Formül (1.2)’ye göre hesaplanmış olan sonuçlar listelenmiştir. Örneğin ID1 müşterisinin RFM skoru 452’dir.



Şekil 2.1: IBM Watson Studio karar akışı

Çizelge 2.4: RFM analizi örnek müşteri skorları

Customer ID	RFM Skor
ID 1	452
ID 2	125
ID 3	231
ID 4	345
ID 5	111
ID 6	332
ID 7	345
ID 8	131
ID 9	453
ID 10	551
---	---
ID 3252	342
ID 3253	125
ID 3254	325
ID 3255	435
ID 3256	121
ID 3257	345
ID 3258	435
ID 3259	412
ID 3260	534
ID 3261	551

2.3 Gradient Boosting Algoritması Uygulama

Müşteri alışkanlıklarını belirlemede, yaptıkları alışveriş frekanslarıyla PESTEL analizinden kullanılacak olan değişkenler etkin rol oynamıştır. Bu değişkenler Gradient Boosting algoritmasına seçilmeden önce sektör çalışanlarının iş bilgisi ayrıca modele katkısı gözlendikten sonra dahil edilmiştir. Sektöre ait değişkenler ve ülke ekonomisine etki eden değişkenler dahil edilerek Gradient Boosting makine öğrenmesi tahmin modeline RFM analizi sürekli değişkenleriyle model aşamasına geçilmiştir.

2.3.1 Model başarı metrikleri

2.3.1.1 ROC eğrisi ve AUC değeri

Çalışmada model seçim kriteri olarak ROC eğrisi ve buna bağlı olarak AUC değeri kullanılmıştır. AUC değeri binominal bağımlı değişkenlerde oldukça sık kullanılan ve özünde 1. tip ve 2. tip hata oranlarına dayalı bir istatistiktir. Sınıflandırma probleminin performansının değerlendirilmesinde yaygın olarak kullanılan ölçümlerden biridir. ROC farklı sınıflar için bir olasılık eğrisidir. Tipik bir ROC eğrisinde X ekseninde Yanlış Pozitif Oran (FPR) ve Y ekseninde Gerçek Pozitif Orana (TPR) vardır. Eğri altındaki alan (AUC) model becerisinin başka bir deyişle model performansının bir özeti olarak kabul edilebilir. Hesaplanan bu AUC model başarı ölçütü olarak kullanılmıştır.

2.3.1.2 Hata matrisi

Çalışmada bir diğer karşılaştırma ölçütü olarak kesinlik ve hassasiyet değerlerinin harmonik ortalaması alınarak hesaplanan F değeri kullanılmıştır. F değeri formül (2.1)'de gösterildiği gibidir.

$$F-1 \text{ score} = 2(\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (2.1)$$

Formülde kullanılan kesinlik ve hassasiyet ölçümlerinin elde edilişleri ise sırasıyla formül (2.2) ve formül (2.3)'te gösterilmiştir. Şekil 2.2'de verilen hata matrisi ile tahminlerin ve gerçek değerlerin görsel karşılaştırmasına yer verilmiştir. Hata matrisi verideki var olan durum ile sınıflama modelinin doğru ve yanlış tahminlerinin sayısını göstermektedir. Şekil 2.2'de 2x2'lik bir hata matrisi görülmektedir. Bu matris tahmindeki hedef sayısına göre değişmektedir.

$$\text{Precision} = TP / TNF \quad (2.2)$$

$$\text{Recall} = TP / (TP + FN) \quad (2.3)$$

		Var olan Durum	
		Pozitif Durumlar	Negatif Durumlar
Tahmin	Pozitif	TP	FP
	Negatif	FN	TN

Şekil 2.2: Hata matrisi

Formüllerde kullanılan kısaltmaların açıklamaları ise aşağıdaki gibidir.

TP : Doğru tahmin edilen sınıflar

TNF : Yanlış tahmin edilen sınıf sayısı

FN : Hatalı tahminlerin başarısız sınıf sayısı

FP : Yanlış tahmin edilen başarılı sınıf sayısı

TN : Gerçek hata tahmini

RECALL: Kesinlik kavramı genelde p harfi ile gösterilir ve getirilen bilgedeki doğru sonuçların, getirilen bilginin tamamına oranı olarak hesaplanır.

PRECISION: Hassasiyet kavramı da genelde r harfi ile gösterilir ve getirilen doğru sonuçların, getirilmesi gereken doğru sonuçlara oranı ile hesaplanır.

TP ve TNF değerlerinin hesaplanabilmesi için model sonucunda tahmin edilen olasılık değerlerinin hangi kategoriye gireceğini belirlemek gerekmektedir. Verinin negatif ve pozitif değerler içermesi ve veri sayısının dengeli olmamasından dolayı olasılık sınır değeri Odds oranına göre belirlenmiştir. Odds oranı formül (2.4)'teki gibi hesaplanmıştır.

$$Odds\ Ratio = \frac{TP/FP}{FN/TN} \quad (2.4)$$

2.3.1.3 Matthews korelasyonu

Sınıflandırma modellerinde; özellikle dengesiz veri setlerinde formül (2.5)' de hesaplanan Matthews korelasyon katsayısı güvenilir sonuçlar vermektedir. Örneğin; confusion matrix incelemesinde; precision, recall ve F-1 skoruna ayrı ayrı bakmak yerine Matthews korelasyon katsayısına bakılabilir. Yorumlanması pearson correlation katsayısına benzerdir. -1 değeri, varolan gerçek değerler ile sınıflandırıcının

verdiği kararların tamamen birbirine zıt olduğunu gösterirken; 0 değeri rastgele bir sınıflandırma olduğunu, +1 değeri ise sınıflandırmanın tamamen başarılı olduğunu göstermektedir.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (2.5)$$

2.3.2 Değişken ve parametre seçimi

Boosting algoritmalarında genellikle tahmin gücüne dayalı değişken ön seçimi yapılması tavsiye edilmez. Bu seçim işlemi algoritmanın kendisine bırakılır. Buna rağmen değişkenlerden minimum kayıp değer miktarı ya da belirli bir seviyede varyans üretebilme beklentisi vardır.

2.3.2.1 Model değişkenleri

Modele dahil edilmek üzere; RFM analizi sürekli değişkenleri, PESTEL analizi sektörel satışlar üzerinde etkisi olan değişkenler ve ekstra türetilen değişkenlerle birlikte 15 adet değişken kullanılmıştır. RFM analizi sürekli değişkenlerinden yeni türetilen değişkenler olmuştur. Veri setinde yer alan girdi değişkenlere ilişkin açıklamalar çizelge 2.5'deki gibidir. Tahmin edilmek istenen değişken "Hedef" değişken olarak etiketlenmiştir. Bir sonraki ay alışveriş yapma durumu hedef değişkendir. İlgili ay Period değişkeni adı altında Yıl-Ay şeklinde birleştirilmiştir. RFM analizi sürekli değişkenlerinden yararlanılarak elde edilen; Last_Purchase_Amount müşterinin son ayki harcama tutarını ifade etmektedir, Average_Purchase_Amount müşterinin tüm zamanlardaki alışveriş tutarının frekasına bölünmesiyle elde edilmiştir, T3-T6-T9-T12 değişkenleri ise mevsimselliği gözardı etmemek adına; 3 ay, 6 ay, 9 ay ve 12 ay önceki satın alma durumlarını göstermektedir, Periodic_Feature değişkeniyle müşterinin kaç ayda bir satın alma eğilimi gösterdiğini yansıtmaktadır, Tenure değişkeni; müşterinin ilk alışverişinden itibaren geçen süre ile müşterinin ay bazında yaşını ifade etmektedir. PESTEL analizi faktörlerinden yararlanılarak elde edilen; Tender_Count değişkeni Elektronik Kamu Alımları Platformu sitesi üzerinden alınan, yapı ve inşaat sektörü alanındaki aylık ihale sayılarını ifade etmektedir, USD_Ratio; Türkiye Cumhuriyeti Merkez Bankası Elektronik Veri Dağıtım sitesi üzerinden alınan USD kurunun 3 aylık değişim oranıdır, Interest_Rate; Türkiye Cumhuriyeti Merkez Bankası Elektronik Veri Dağıtım sitesi

üzerinden alınan konut kredisi faiz oranının 3 aylık değişim oranıdır. Müşterilerin sahip olduğu ekipmanların çalışma saatleri, müşterilerin çalışma yoğunluğu üzerinde fikir vermektedir. Bu sebeple Work_Hour_Ratio; müşterilerin sahip olduğu ekipmanların aylık ortalama çalışma saati bilgisini ifade etmektedir.

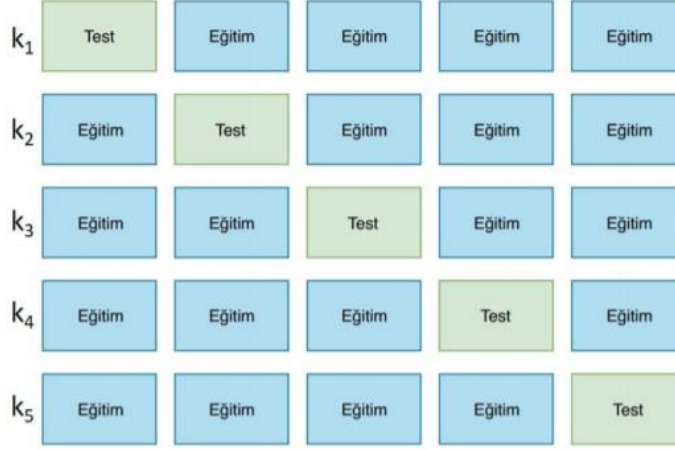
Çizelge 2.5: Model değişkenleri

Değişken adı	Açıklama
Hedef	Bir sonraki ay alışveriş yaptı/yapmadı (1/0)
Recency_Month	Son alışverişinden sonraki geçen ay sayısı
Frequency	Toplam alışveriş sayısı
Monetary	Müşteriden elde edilen toplam gelir
Last_Purchase_Amount	Müşterinin son harcama tutarı
Average_Purchase_Amount	Müşterinin ortalama harcama tutarı
T3	Son alışverişinden 3 ay önce alışveriş yapmış mı (0-1)
T6	Son alışverişinden 6 ay önce alışveriş yapmış mı (0-1)
T9	Son alışverişinden 9 ay önce alışveriş yapmış mı (0-1)
T12	Son alışverişinden 12 ay önce alışveriş yapmış mı (0-1)

Periodic_Feature	Müşterinin ilk alışverişi ve son alışverişi arasındaki geçen ay sayısının toplam alışveriş sayısına oranı.
Period	Yıl+Ay
Work_Hour_Ratio	Müşterinin sahip olduğu ekipmanların aylık ortalama çalışma saati bilgisi
Tender_Count	İnşaat-Yapı alanındaki ihale sayısı
Tenure	Müşterinin ilk alışverişinden itibaren geçen ay sayısı
USD_Ratio	USD alış kurunun 3 aylık değişim oranı
Interest_Rate	Konut kredisi faiz oranının 3 aylık değişim oranı

2.3.2.2 Çapraz validasyon ile parametre optimizasyonu

Çalışmada gradient boosting algoritmasının eğitilmesi, test edilmesi ve parametre optimizasyonu için çapraz validasyon yöntemi kullanılmıştır. Çapraz validasyon ile veri seti alt kümelere bölünmüştür. Verilen k sayısı kadar işlem k kadar tekrarlanır. Her tekrarda, k alt kümesi bölünür, test ve k-1 eğitime dönüştürülür bu şekilde hipotez test edilir. Varsayılan doğruluk değeri, verileri rastgele biçimde bölerek her çapraz doğrulama katmanında bulunur. İşlemin sonunda, varsayılan doğruluk tüm k denemelerinin ortalaması alınarak hesaplanır (Kohavi, 1995). Bu çalışmada, yanlılığın düşük olması ve varyans, çapraz doğrulama 5 kat çapraz doğrulama kullanılmıştır (Han, Kamber ve Pei, 2014). Geliştirilen bu metodoloji şekil 2.3’de gösterilmiştir.



Şekil 2.3: Çapraz validasyon

Model parametrelerinin optimizasyonunu yapmak için eğitim seti k katlamalı çapraz doğrulama ile 5 eşit parçaya ayrılmış olup, her seferinde farklı bir parça dışarıda tutularak aynı parametrelerle model eğitilmiş ve dışarıdaki parça tahmin edilmiştir. Her bir satırda yer alan 4 eğitim veri kümesi için 1 test veri kümesi kullanılmaktadır. Bu şekilde test kümesinin yer değiştirmesi sağlanmış olur. Bu şekilde eğitim seti dışında tutulan ve tahmin edilen parçalar bir araya getirildiğinde eğitim setindeki tüm gözlemleri kapsamış ve bir validasyon başarı istatistiği hesaplanmıştır. Bu yöntemle yanlış bir kümenin seçilmesinin önüne geçilmiş olarak parametre optimizasyonları sağlanmıştır.

Parametre optimizasyonu sonrası Python programlama dili üzerinden elde edilen optimum parametreler aşağıdaki çizelge 2.6'daki listelenmiştir.

Çizelge 2.6: Parametre değerleri

Parametre	Değer
Boosting_type	gbdt
Colsample_bytree	1.0
Importance_type	split
Lambda_11	1.5
Lambda_12	1
Learning_rate	0.1
Max_depth	-1
Min_data_in_leaf	300
Min_child_weight	0.001
Min_split_gain	0.0

2.3.3 Gradient Boosting algoritması sonuçları

Microsoft LightGBM algoritmasının Python kütüphanesinde yapılmış olan denemeler sonucunda en güçlü tahminleri veren model oluşturulmuştur. Modele ait yukarıda listelenen metriklerin çıktıları çizelge 2.7’ de listelendiği gibidir.

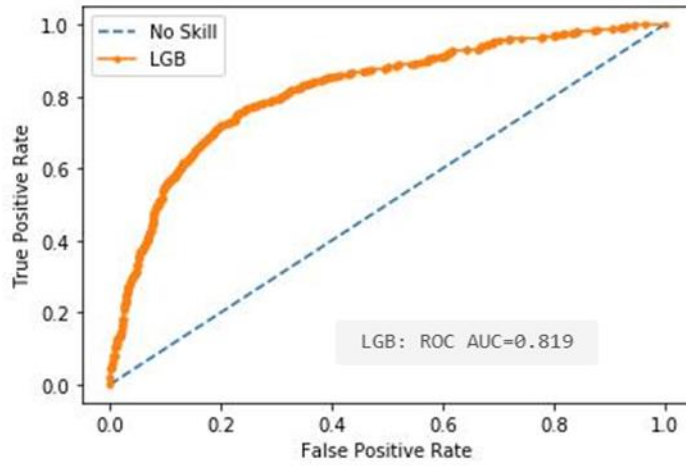
Precision değeri incelendiğinde; 0 değeri yani müşterilerin alım yapmama durumunun ifade edildiği değeri 0.94 kesinlikle doğru bilindiğini ve 1 değeri yani müşterilerin alım yapma durumunun ifade edildiği değeri 0.40 ihtimalle doğru bilindiği yani tahmin edilen 10 kişiden 4’ünün satın alım yaptığı bilinmektedir. Bu durum Confusion Matrix de 1 (Pozitif) durumu olan sonuçlar için düşük gözükmeyle birlikte Recall alanı incelendiğinde satın alan müşterilerin 0.75’inin doğru tahmin edildiği sonucuna ulaşılmaktadır. Recall başarısı olarak yadsınamaz şekilde Pazarın satın alım durumunun tahmin edildiği söylenebilir.

Çizelge 2.7: Hata matrisi

	Precision	Recall	F-1 Skor
0	0.94	0.77	0.81
1	0.40	0.75	0.52

Precision ve Recall değerlerinin ortak değerlendirilmesi olan şekil 2.4’de yer alan ROC grafiği incelendiğinde 0.819 ile yeterli düzeyde olduğu tespit edilmiştir.

F-1 Skor olarak 0.52 tespit edilmiştir.



Şekil 2.4: AUC-ROC eğrisi

Mathew korelasyon hesaplaması yapıldığında; 0.316 değerine ulaşılmış olup, bulunan değer 0 ile 1 arasında yer aldığı için sınıflandırmanın başarılı olduğu söylenebilir.

Modele dahil edilen 18 bağımsız değişkenin önem seviyeleri incelendiğinde ise aşağıdaki çizelge 2.8'deki gibi sıralaması elde edilmiştir. Buradaki önem düzeyleri; tahminde kullanılan değişkenlerin geçtiği ağaç sayısını ifade etmektedir. Daha çok ağaçtan geçen değişkenin; genel tahminlerde daha belirleyici olduğu söylenebilir.

Çizelge 2.8: Değişkenlere ait önem düzeyleri

Değişkenler	Önem Düzeyi
Last_Purchase_Amount	457
Monetary	432
USD_Ratio	324
Average_Purchase_Amount	280
Periodic_Feature	251
Work_Hour_Ratio	238
Recency_Month	233
Tender_Count	182
Frequency	153

2.3.3.1 Model açıklanabilirliği

Açıklanabilirlik ve yorumlanabilirlik; Bir makinenin veya derin öğrenme sisteminin iç mekaniğinin insan terimleriyle ne ölçüde açıklanabileceğini ifade etmektedir. Tahmin modelinin neden belirli bir çıktı ürettiğini anlamak zor olduğu için genellikle kara kutular olarak eleştirilir. Model açıklanabilirliği ile, karmaşıklığın önüne geçmek ve makine öğrenmesi algoritmalarının çıktısını hangi değişkenlerin değiştirdiğini açıklamak amaçlanmıştır. Lundberg ve Lee tarafından geliştirilen SHAP (SHAPley Additive ExPlanations), bireysel tahminleri açıklamaya yönelik bir yöntemdir. SHAP, oyunun teorik olarak optimal Shapley Değerlerine dayanmaktadır. Shapley yalnızca gözlemlenen veri noktalarını kullanır. (Lundberg, Scott M., Gabriel G. Erion, & Su-In Lee, (2018)). İnsan sezgisi doğrultusunda güvenilir açıklamalar yapar. SHAP, oyun teorisinde sağlam bir teorik temele sahiptir. Farklı bir deyişle; makine öğrenimi algoritmalarının çıktısını açıklamak bir oyun teorisi yaklaşımıdır. Tahmin, özellik değerleri arasında oldukça dağıtılır. Lundberg et al. Karar ağaçları, rastgele ormanlar ve gradyan destekli ağaçlar gibi ağaç tabanlı makine öğrenimi modelleri için bir SHAP çeşidi olan TreeSHAP'ı önermişlerdir. Değerlerin önemi Shapley değerleri ile yapılır. Örneğin, bir bireyin şeker hastası olması için hangi faktörlerin daha önemli olduğu. Burada faktörlerin olumlu ve olumsuz etkileri olmuş olabilir. Önemli olan etkilemesidir. Shapley değerleri gibi özellik nitelikleri "kuvvetler" olarak görselleştirilebilir. Her öznetelik değeri, tahmini artıran veya azaltan bir güçtür. Tahmin, taban çizgisinden başlar. Shapley değerlerinin temeli, tüm tahminlerin ortalamasıdır. Çizimde, her Shapley değeri, tahmini yukarı (pozitif değer) veya aşağı (negatif değer) iten bir oktur. Bu kuvvetler, veri örneğinin gerçek tahmininde birbirini dengelemektedir.

Uygulamada satış potansiyeli bulunan müşterilerin olasılığının tahmin edilmesinin yanı sıra, bu satın alma potansiyeline neden olacak SHAP açıklamaları belirlenmiştir. Aşağıda Şekil 2.5'de örnek bir müşterinin güç grafiği sonucu gösterilmiştir.



Şekil 2.5: Model açıklanabilirlik grafiği

Şekil 2.5’de; Satın alma potansiyeli bulunan müşteri olasılığı 0.93’tür. Müşteri olasılığı için öncelikli güçlendirici etken; Periodic Feature, ikincil etken; Monetary olmuştur. Belirleyici etkenler olarak müşteri verilerinin yanında, PESTEL analizi değişkenlerinden olan USD Ratio’nun da belirleyici etken olduğunu görmekteyiz.

SHAP, ağaç tabanlı modeller için hızlı bir uygulamaya sahiptir. SHAP, LightGBM ağaç güçlendirme çerçevelerine entegre edilerek, uygulamanın açıklanabilirliği sağlanmaktadır.



3. SONUÇ

Veri madenciliği, verinin büyüklüğünden ziyade verinin doğru, etkin, amaca yönelik kullanımını ve bilgiyle ne yapılacağı yönünde önemli noktaları içinde barındırmaktadır. Akıllı kararlar verebilmemiz ve geleceğe dönük kararlar alabilmemiz için verinin analiz edilmesi oldukça önemlidir. Şirket amacına uygun yöntemlerle verinin analiz edilmesi şirketlere oldukça getiri sağlayacaktır. Günümüzde veri analizi, Endüstri 4.0 gelişimi ve 'kişiselleşmiş ürüne' yönelim ile önem ve değer kazanmıştır. Makine öğrenmesinin etkin olarak kullanıldığı bu çalışmada modelde, farklı başarı kriterleri dikkate alındığında iyi sonuç üretebildiği görülmektedir. Burada seçilen kriterin ölçülmek istenen performans mantığına hizmet edip etmediği önemlidir. Örneğin AUC değeri tüm örnekleme yanlış tahminlerin sıralaması ile ilgilendiğinden dağılımın bütününe önemsemekte, F-1 değeri ise karar matrisi odaklı olduğundan sadece belirlenen sınır değer üzerindeki ya da altındaki yanlış tahminlere odaklanmaktadır.

Kullanılan model, satış yapılan müşterilerden elde edilen paternleri öğrenerek gelecekte satın alımı muhtemel müşterileri tahmin etmektedir. RFM analizinin, makine öğrenmesi tekniklerine dahil edildiği bu çalışmada; müşterilerin satın alma alışkanlıklarına dayanarak müşteriler, RFM skalasında alım değerlerine göre segmente edilmiştir. Ardından PESTEL analizi etkileyici faktörleriyle birlikte Gradient Boosting algoritmasına dahil edilmiştir. Bunun sonucunda kampanya ve hizmetlere en fazla katkı sağlayabilecek potansiyeldeki müşteriler belirlenmiştir. Model açıklanabilirliği ile müşterilerin spesifik olarak belirleyici faktörleri de bulunmuştur. Bu yeni yaklaşım, müşterilerin sosyo-ekonomik hassasiyet etkileriyle müşterilerin daha doğru ve isabetli belirlenmesine yardımcı olmaktadır. Satış alanındaki sektörlerde kar artışı, müşteri memnuniyet artışı ve gereksiz giderlerin azaltılması avantajlarını sağlayacağı gibi; maksimum kazancı minimum süre ve eforla sunabilme yetisine sahip olmaktadır.

Özetle, farklı değişken yaklaşımlarının birlikte kullanılmasının sonuçları iyileştirmede faydalı olacağı kesindir. Genel anlamda, makine öğrenmesi uygulamalarında temel amaç; veriden örüntüler öğrenmek, bu örüntüleri kullanarak değer oluşturmaktır. Bu amaç doğrultusunda büyük veri içerisinde değerli, kullanılabilir bilgiler açığa çıkarılarak; karar destek sistemi değerlendirilmiştir.

3.1 Öneriler

Yapılan çalışma; satış tahminlerinin müşterilerin geçmiş işlem hareketlerini veren RFM (recency, frequency, monetary) değişkenlerine ilaveten, PESTEL analizi değişkenlerini de ilave ederek Gradient Boosting makine öğrenmesi algoritmasıyla başarılı düzeyde tahminler elde edilebileceğini gösteriyor.

Modele giren değişkenler incelendiğinde ülkenin genel ekonomik ve sektörel durum göstergelerinin modeli oldukça etkilediği gözükmektedir. Dolayısıyla modele PESTEL değişkenlerinden yeni değişkenler türetilip beslenirse başarı oranının artacağı gözükmektedir. Örneğin, yapı-inşaat sektörü, madencilik sektörüne ait göstergelerin satın alma durumunu etkileyeceği düşünülmektedir, modele dahil edilmesiyle AUC-ROC değerleri, F1-score değerleri karşılaştırılıp değerlendirilebilir.

Makine öğrenmesi modellerinin başarısında en büyük etken, modele giren değişkenlerin çeşitliliği ve kalitesi olduğu için, müşterilerin satın alma geçmişine veya müşterilerin birbirlerinden ayırt edici özellikleri tespit edilip model beslenmesi gelecekteki satın alma davranışlarını tahmin etmekte başarıyı arttıracaktır.

KAYNAKLAR

- A. Swalin.** “Catboost vs. Lightgbm vs. Xgboost”. (2018), [Online]. Available: <https://towardsdatascience.com/catboost-vs-light-gbm-vs-xgboost-5f93620723db>. Access date: 19.03.2019.
- Bult, J. R. & Wansbeek, T..** “Optimal selection for direct mail”. Marketing Science. Vol. 14. No. 4. (1995) 378-394. ISSN: 0732-2399.
- Cheng, CH. & Chen, YS.** (2009). Classifying the segmentation of customer value via RFM model and RS theory, Cilt. 36, No. 3, (April 2009).4176-4184, ISSN: 0957-4174.
- Chuang, H. & Shen, C.** (2008). A study on the applications of data mining techniques to enhance customer lifetime value – based on the department store industry, Proceedings of the 7th International Conference on Machine Learning and Cybernetics, pp. 168-173, ISBN:978-1424420964, Kunming, China, July 2008, IEEE.
- Derya Birant,** (2011). Data Mining Using RFM Analysis, Knowledge-Oriented Applications in Data Mining, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-154-1, InTech.
- Ha, S.H.** (2007). Applying knowledge engineering techniques to Customer analysis in the service industry, Advanced Engineering Informatics, Vol. 21, No.3, (July 2007) 293-301, ISSN:1474-0346.
- Han, J., Kamber, M., & Pei, J.** (2014). Data mining: Data mining concepts and techniques. Proceedings, 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA. 2013. <https://doi.org/10.1109/ICMIRA.2013.45>
- Hosseini, S.M.; Maleki, A. Gholamian, M.R.** (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty, Expert Systems with Applications, Vol. 37, No. 7, (July 2010) 5259-5264, ISSN:0957-4174.

internet: Necip Murat, (2017). Hamileden Önce Hamileliği Bilen Marka Target, <https://pazarlamasyon.com/hamileden-once-hamileligi-bilen-marka-target/>.

J.H. Friedman, (February 24, 1999). Greedy Function Approximation a Gradient Boosting Machine.

Kohavi, R., (1995). The Power of Decision Tables. ECML.

Lundberg, Scott M., Gabriel G. Erion, & Su-In Lee, (2018). “Consistent individualized feature attribution for tree ensembles.” arXiv preprint arXiv:1802.03888.

Semra Erpolat Taşabat; Esra Akca, (2020). Recycle Project With RFM Analysis in Industrial Material Sector, Sigma Journal of Engineering and Natural Sciences, ISSN: 1681-1692.

Tianqi CHEN, (Oct. 22 2014). Introduction to Boosted Trees, University of Washington.

Yi Fan, (2016). Rfm Analysis in R Math 3201 Datamining Foundation.

X. Team. (2019) “Introduction to boosted trees”. Available: <https://xgboost.readthedocs.io/en/latest/tutorials/model.html>. Access date: 19.03.2019.