T.C.
MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE FAİZSİZ FİNANSMAN SEKTÖRÜNDE MÜŞTERİ KAYIP TAHMİNİ: CHURN ANALİZİ

YÜKSEK LİSANS TEZİ

Ayşegül KABA

İstatistik Anabilim Dalı
İstatistik Programı

Tez Danışmanı: Prof. Dr. Barış AŞIKGİL

HAZİRAN 2021

MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE FAİZSİZ FİNANSMAN SEKTÖRÜNDE MÜŞTERİ KAYIP TAHMİNİ: CHURN ANALİZİ

YÜKSEK LİSANS TEZİ

Ayşegül KABA

İstatistik Anabilim Dalı

İstatistik Programı

Tez Danışmanı: Prof. Dr. Barış AŞIKGİL

HAZİRAN 2021

Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü tez yazım kılavuzuna uygun olarak hazırladığım bu tez çalışmasında;

- 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.

Yüksek lisans tez çalışmamda ilgi ve desteğini esirgemeyen tez danışmanım sayın Prof. Dr. Barış AŞIKGİL'e, görüş ve önerilerinden dolayı Doç. Dr. Ayça ÇAKMAK PEHLİVANLI'ya sonsuz şükranlarımı sunarım.

Kariyer hayatımda, ihtiyacım olduğu her an yanımda olan sevgili müdürüm Mehmet ÇAKOĞLU'na, maddi ve manevi olarak desteklerini esirgemeyen aileme ve bu çalışmamda motive ederek yardımcı olan değerli dostlarıma sonsuz teşekkürlerimi sunarım.

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE FAİZSİZ FİNANSMAN SEKTÖRÜNDE MÜŞTERİ KAYIP TAHMİNİ: CHURN ANALİZİ

ÖZET

Avrupa'nın birçok ülkesinde uygulamada olan faizsiz finans sisteminin tercih edilmesinde, ihtiyaçların hızlı ve kolay karşılanabilmesi en önemli sebepler arasındadır. Gayrimenkul, taşıt ve iş yeri satışını kolaylaştırmak için oluşturulan faizsiz finans sistemi dünyada olduğu gibi Türkiye'de de giderek yaygınlaşmaktadır ve rekabet her gün artmaktadır.

Günümüz iş dünyasında faizsiz finans sistemine öncülük yapan ve sektörde kendine yer edinmeye çalışan firmaların sürekliliği sağlayabilmesi için sistemden ayrılacak müşterilerin tahmini (Churn Analizi) oldukça önemlidir. Makine öğrenme uygulamaları da bu konuda aktif bir şekilde kullanılmaktadır. Sektörün hızlı bir gelişme sürecinde olması ve firmalar arası rekabetin büyüklüğü nedeniyle ayrılacak müşterilerin analiz ve tahmini faizsiz finans sektöründe yoğun bir şekilde yapılmaktadır.

Bu çalışmanın amacı, faizsiz finans sisteminde yaşanılan kayıpları incelemek ve en iyi kayıp tahminini veren modeli oluşturmaktır. Çalışma, faizsiz finans sektöründeki öncü firmanın 2020 yılına ait verilerini içermektedir. Çalışmada kullanılan veri kümesi 18507 müşteriye ait olup, 14 etkin özellik içeren değişkenlerden oluşmaktadır. Makine öğrenmesi yöntemleri ile modeli kurmadan önce müşteri kaybına sebep olabileceği düşünülen değişkenler, keşifsel veri analizi ile incelenmiştir. Müşteri kaybına sebep olabileceği düşünülen veriler ve Churn değişkeni ile model oluşturulmadan önce veri kümesi %75-%25 oranında bölünerek Lojistik regresyon (LR), K en yakın komşu (KNN) ve Destek vektör kümeleri (SVM) ile en iyi performans veren model incelenmiştir.

Anahtar Kelimeler: Müşteri kayıp analizi, makine öğrenimi, sınıflandırma, faizsiz finans sektörü

CUSTOMER LOSS FORECAST IN THE INTEREST FREE FINANCE SECTOR WITH MACHINE LEARNING METHODS: CHURN ANALYSIS

ABSTRACT

One of the most important reasons for the preference of interest-free financial systems, which are in practice in many European countries, is that the needs can be met quickly and easily. The interest-free finance system, which was created to facilitate the sale of real estate, vehicles and workplaces, is becoming increasingly widespread in Turkey as well as in the world, and competition is increasing day by day.

In order for companies that are leading interest-free financial systems in today's business world and trying to gain a place in the sector to ensure continuity, the forecast of customers who will leave the system (churn analysis) is very important. Machine learning applications are also actively used in this regard. Due to the fact that the sector is in a rapid development process and the size of competition between companies, the analysis and prediction of customers who will leave is carried out intensively in the interest-free financial sector.

The aim of this study is to examine the losses experienced in the interest-free finance system and to create the model that gives the best loss estimation. The study includes the data of the leading company in the interest-free finance sector for 2020. The dataset used in the study belongs to 18507 customers and consists of variables containing 14 active features. Before building the model with machine learning methods, the variables that are thought to cause loss of customers were examined with exploratory data analysis. Before creating the model with the data that is thought to cause customer loss and the Churn variable, the data set was divided by 75%-25% and the best performing model was examined with Logistic regression (LR), K nearest neighbor (KNN) and Support vector sets (SVM).

Keywords: Customer loss Analysis, Machine Learning, Classification, interest-free finance sector

İÇİNDEKİLER

Sayfa

KISALTMALAR

SVM	:	Destek Vektör Kümesi (Support Vector Machine)
AUC	:	Eğri Altındaki Alan (Area Under Curve)
ROC	:	Alıcı Karakteristik İşleci (Receiver Operating Characteristic)
KNN	:	K En Yakın Komşu (K Nearest Neighbour)
LR	:	Lojistik Regresyon (Logistic Regression)
ADASYN	:	Uyarlanabilir Sentetik Örnekleme Yaklaşımı (Adaptiv Synthetic Sampling Approach)
SMOTE	:	Sentetik Azınlık Aşırı Örnekleme Tekniği (Synthetic Minority Oversampling Technique)
CV	:	Çapraz Doğrulama (Cross Validation)
TDFFS	:	Tasarrufa Dayalı Faizsiz Finans Sistemi
GLM	:	Genelleştirilmiş Doğrusal Model (Generalized Linear Model)

ÇİZELGE LİSTESİ

Sayfa

ŞEKİL LİSTESİ

Sayfa

1.GİRİŞ

Tasarrufa dayalı sistem, modern finans sistem ile birlikte ortaya çıkan yeni düzenin sebep olduğu karışıklığı ortadan kaldırmayı amaçlayan alternatif bir yaklaşımdır. Belirlenen sabit faizi yasaklayan ve kazanılan gelirden belli bir ölçüde hisse almayı uygun bulan sistem, kâra ortaklık esasına dayanmaktadır. Tasarrufa dayalı sistem, faizsiz kazanç ve öz sermaye temel ilkelerine dayanır. Bu sistem geçmişte de benzer işlevleri yürüten kurumlar ile ayakta kalmaya çalışmıştır. Modern çağa ayak uyduramayan bu kurumlar maalesef tutunamamışlardır. Ancak; yirminci yüz yılın başlarından itibaren Pakistan bölgesinde, Ortadoğu'da ve diğer islam ülkelerinde sistem kendine uygulama alanı bularak faaliyetlerini geliştirmeye başlamıştır. Artık bugün hem islam ülkelerinde hem de dünyanın birçok ülkesinde tasarrufa dayalı sisteme uygun çalışan bağımlı ve bağımsız birçok kurum vardır ve halen dünya finans sisteminin bir seçeneği olma yönünde ilerlemektedir ve ciddi ölçüde rekabet ortamı yaratmaktadır. Rekabetin temel nedeni olan müşterilerin Churn olma olasılıklarının incelendiği ve makine öğrenimi ile gerçekleştirilecek analizde, 100 ayrı şubeden alınan veriler kullanılarak en iyi performans gösteren model oluşturulacaktır.

1.1 Literatür Taraması

Literatür incelendiğinde makine öğreniminin, son yıllarda artan bir araştırma alanı olduğu görülmektedir. Ancak müşteri ilişkileri yönetiminin önemli bir konusu olan müşteri kaybı analizi (Churn Analizi) hakkında makine öğrenimine dair sınırlı sayıda araştırma bulunmaktadır. Bu araştırmalarda birçok farklı öğrenme yöntemi incelenmiştir. Bunların başında regresyon modelleri, karar ağaçları, rastgele orman (random forest), destek vektör makinesi (support vector machine), Bayes ağları (bayesian network), k-en yakın komşu (k-nearest neighbors), lojistik regresyon gibi yöntemler vardır. Birçok araştırmada hibrit yöntemler de kullanılmıştır. Ancak en popüler teknik, yapay sinir ağları, komite modeller ve hibrit çalışmaların olduğu görülmektedir.

Son çalışmalardan biri olan Coussement ve arkadaşlarının (2017) yapmış olduğu Churn analizi öncesindeki süreç olan, veri kümesinde yapılan işlemlere dair incelemeler tamamlanmıştır. Müşteri Churn tahmininin analistler için çeşitli karar noktaları içerdiğini ve bu karmaşık yapının, işi anlama, veri kümesini anlama, veri ön işleme, veri kümesini modelleme, veriyi değerlendirme ve geliştirme olmak üzere 6

aşamaya ayrıldığı belirtilmiştir. Çalışmalarında bu aşamalardan, veri ön işlemeye odaklanmışlardır. Veri setlerindeki ayrık ve devamlı değişkenlerin formatını uygun hale getirerek ve müşterilere dair doğru öznitelikler seçerek analiz performansının artığını vurgulamışlardır. Çalışmada veri ön işleme işlemleri, veri indirgeme ve veri hazırlama aşaması olarak iki kısma ayrılmıştır. Veri indirgeme yöntemlerinin amacı analizi hangi özelliklerin etkilediğini belirleyip etkisi olmayan özelliklerin modelden çıkarılmasını sağlamak böylelikle veri boyutunu azaltmaktır. Veri hazırlama yöntemleri ise değişkenlerin uygun formatlara dönüştürülmesi için gereklidir. Veri ön işlemenin, değer dönüşümü ve sunumu olarak iki adımı vardır. Çalışmada bu iki adım üzerine analizler yapılmıştır. Kullanılan veri kümesinde 30,104 müşteri vardır. Değişkenlerin 156'sı kategorik 800'ü sürekli değişkendir. Çalışmada eğitim, seçim ve test olmak üzere sırayla veri kümesinin %50, %20 ve %30'luk kısımları alınarak kullanılmıştır. Veri ön işleme işlemlerinin etkisiyle birlikte Lojistik Regresyon modeli için Churn tahmini başarısı ölçülmüştür. Veri ön işleme işlemlerinin tahmin başarısını %34'e kadar çıkarabildiği görülmüştür. Ayrıca çalışma sonuçlarından bir diğeri de lojistik regresyon algoritmasının yapay sinir ağları ve destek vektör makinesi gibi yöntemlere göre daha hızlı olduğu görülmüştür (Coussement, Lessmann ve Verstraeten, 2020).

Amin ve arkadaşlarının (2017) çalışması, veri tabanı üzerinde uygulanan Kapsamlı Algoritma (Exhaustive Algorithm), Genetik Algoritma, Kaplama Algoritması (Covering Algorithm) ve LEM2 Algoritması olan kural üretme yöntemlerini kullanarak müşteri davranışını tahmin etmeyi amaçlamaktadır. Yöntemlerin ölçümü için Kaba Set Sınıflandırması (Rough Set Classification) uygulanmıştır. Bu işlem sonucunda, Genetik Algoritma en iyi müşteri kaybı olasılığı oranını vermiştir.

Kaynar ve arkadaşlarının (2017) çalışmasında, destek vektör makineleri, Naif Bayes ve çok katmanlı yapay sinir ağları kullanılarak 3 model elde edilmiştir. Çalışmalarında 4667 adet müşteriden alınan bilgiler kullanılmıştır. 21 tane öznitelik çıkarılmıştır. Veri kümesinde hem kaybedilmiş müşteriler hem sadık müşteriler vardır. Veri kümesinin rastgele seçilmiş %75'i öğrenme verisi, kalan %25'i test verisi olarak kullanılmıştır. Çalışmada uygulanmış 3 yöntem içerisinden modelleme tahmin başarısı en yüksek olan yöntem %92,35 ile yapay sinir ağları olmuştur. İkinci yöntem %87,15 başarı ile Naif Bayes yöntemidir. Üçüncü ise %77,89 başarı ile SVM yöntemidir. Ayrıca Naif Bayes yöntemi hassasiyet açısından en iyi sonucu vermiştir. Yapay Sinir Ağları ve Naif Bayes yöntemleri çalışma için beklenen başarılı sonuçları verirken, Destek

Vektör Makineleri beklenenden düşük performans göstermiştir. Destek vektör makinesi yönteminin daha düşük başarı göstermesinin nedenleri olarak veri kümesindeki bazı öznitelikler ve örnek sayısının yetersizliği ön görülmüştür.

Vafeiadis ve arkadaşları (2015) YSA, destek vektör makineleri, karar ağaçları, Naif Bayes ve lojistik regresyon gibi sık kullanılan Churn tahmin teknikleri ve telekomünikasyon endüstrisindeki performans sınıflandırıcılarını kullanarak bu algoritmaların performanslarının değerlendirilmesi üzerine çalışmışlardır. Karşılaştırmalı sonuçlar, telekomünikasyon endüstrisindeki kayıp tahmini için Boosted SVM olarak adlandırılan yöntemin en iyi sonucu verdiğini göstermiştir.

Abbasimehr ve arkadaşlarının (2014) çalışması komite öğrenmesi uygulamalarının temel öğreniciler için üç performans göstergesi adına, yani AUC, hassasiyet ve özgüllük açısından önemli bir gelişme getirdiğini göstermektedir. Boosting, diğer tüm yöntemler arasında en iyi sonuçları vermiştir. Bu sonuçlar, komite öğrenmesi yöntemlerinin müşteri kayıp tahmini için en iyi yöntem olabileceğini göstermektedir.

Kim ve arkadaşlarının (2014) çalışmasında müşteri kişisel verilerini ve CDR verisini içeren telekomünikasyon firmasından alınan veri kümesi kullanılmıştır. Kullandıkları yöntem lojistik regresyon ve çok katmanlı algılayıcılardır. Veri kümesi %9,7'si Churn etmiş müşterilerden oluşan 89.412 adet örnek müşteridir. Ağ analizinde önceki çalışmaların aksine, ağ değişkeninin bir yayılım süreci olan SPA' dan oluşturulmuştur ve modeli eğitmek için geleneksel kişisel değişkenlerle birleştirilmiştir. Bu şekilde etkin bir yaklaşım geliştirmişlerdir.

Keramati ve arkadaşları (2014) çalışmalarında performanslarını karşılaştırmak için karar ağaçları, yapay sinir ağları, k-en yakın komşu ve destek vektör makinesi gibi veri madenciliği sınıflandırma tekniklerini kullanmışlardır. İranlı bir mobil operatör şirketinin verilerini kullanarak, bu teknikleri birbirleriyle kıyaslayarak önde gelen farklı veri madenciliği yazılımları arasında bir paralellik yakalamışlardır. Tekniklerin davranışlarını incelemek ve özelliklerini bilmek için bazı değerlendirme ölçütlerinin değerinde önemli iyileştirmeler yapan bir hibrit yöntem önermektedirler. Önerilen yöntem sonuçları, geri çağırma ve duyarlık için %95'in üzerinde elde edilebildiğini göstermiştir. Bunun dışında, veri kümesindeki etkili özniteliklerin çıkarılması için yeni bir yöntem tanıtılmış ve deneyimlenmiştir. Ek olarak, en etkili öznitelik kümesini çıkarmak için yeni bir boyutsallık azaltma yöntemi tanıtmışlardır. Kullanım sıklığı, toplam şikayet sayısı ve kullanım sürelerinin etkili öznitelikler olduğu gösterilmiştir.

Ayrıca, aynı veri kümesi üzerinde yapılan önceki çalışmanın aksine, SMS sıklığının, ücret tutarının ve hizmet türünün en az etkili öznitelikler olduğunu göstermişlerdir.

Verbeke ve arkadaşlarının (2012) çalışmasında müşteri kaybı tahmini problemleri için komite modellerinin veri madenciliğinde yaygın bir kullanımı olduğunu belirtilmiştir. KDD 2009 veri kümesi de dahil olmak üzere telekomünikasyon servis sağlayıcılarından toplanan bir dizi örnek derlemişlerdir. Müşteri kaybı tahmin analizi için veri kümesinde hem tek hem de komite algoritmalarını uygulamışlardır. En iyi performans gösteren sınıflayıcıyı seçmek için kâr temelli bir değerlendirme fonksiyonu önermişlerdir. Az sayıda değişken kullanmışlar ve sonuçların klasik değerlendirme yöntemlerinden daha iyi olduğunu bildirmişlerdir.

Kisioğlu ve Topçu'nun (2011) çalışma sonuçlarına göre, Telekom endüstrisinde etkin bir müşteri kaybı yönetimi için, ortalama MoU (Minutes of Usage), ortalama fatura ödemesi, ara bağlantı çağrılarının sayısı ve tarife türü müşteri kayıp oranını analiz etmek için en önemli faktörlerdir.

Huang ve arkadaşlarının (2012) çalışmasında, veri kümesinden yeni öznitelikler oluşturarak ve öznitelikleri bazı özelliklerine göre gruplayarak model başarısını artırmaya çalışmışlardır. Yeni öznitelik oluşturma, veri kümesinin büyüklüğüne dayanır. Öznitelik grupları, doğruluk ve performans açısından en verimli modeli oluşturmak için üç farklı şekilde birleştirilmiştir. Sınıflandırma işleminde lojistik regresyon, doğrusal sınıflandırma, naif bayes, karar ağacı, çok katmanlı algılayıcılar, destek vektör makineleri ve deneysel veri işleme algoritmaları kullanılmıştır. Sonuçları değerlendirmek için ROC (Receive Operating Curves) ölçütü kullanılmıştır. Veri kümesindeki tüm öznitelikleri kullanan destek vektör makineleri ile yapılan sınıflandırma işleminin en iyi sonuçu verdiği görülmüştür.

Bu çalışmanın amacı, faizsiz finans sektöründe sistemden ayrılacak müşterilerin ayrılış sebep tahmininin yapıldığı modeli kurgulamaktır. Firmalar açısından bakıldığında yeni müşteri elde etmektense mevcut müşterilerin ayrılmasını engellemek daha karlı olmaktadır. Mevcut müşterileri koruma hem zaman hem de maliyet açışından firmalar için daha doğru bir stratejidir. Bu çalışmanın da faizsiz finansman sektörünün gelişimini artırmada ve müşteri kayıplarının engellenmesinde faydası olacaktır.

2. FAİZSİZ FİNANSMAN SEKTÖRÜNE GENEL BAKIŞ

2.1 Faizsiz Finansman Sektörüne Genel Bakış ve Tarihçesi

Tasarrufa dayalı sistemi, parasal işlemlerle mal ve hizmet hareketlerinin birbirine bağlı olduğu, yapılan her para hareketinin mutlaka bir hizmete karşılık geldiği, geliri de ortaklık esasına göre paylaşıldığı bir sistem olarak tanımlamak mümkündür (Usulcan, 2013).

Faizsiz finans sistemi ilerleyen günlerde elde edilecek bir para karşılığında şu andaki ticaretten vazgeçmek şeklindeki temel düşünceden yola çıkmıştır. Bu ister varlık karşılığında ister bazı haklar karşılığında olsun aynı sonuca götürmektedir. Faiz odaklı alışılmış yöntemde ihtiyaç sahibi ödünç para almaktadır ve gelecekte faiz ilavesiyle geri ödemektedir (Mabid, 1988). Bu sistem ile tasarruflu para fonuna ihtiyacı olan bireylere bu konuda yardım etmek amaçlanmıştır. Böylece sosyal adaleti ve adilane gelir dağılımını sağlamayı beklemektedir. Aslında sistem her parçasıyla bir bütündür. Bu da maddi tutarlılığı ve güven duygusunu sağlamaktadır. Ayrıca sistemi, aşırı talep ve spekülasyon faaliyetlerinin yol açtığı finansal gerginlikten kaynaklanan potansiyel risklere karşı da korumaktadır (Salahuddin, 2006).

Faizsiz finans sisteminin ilk uygulaması olan ilk İslam bankası 1963 yılında Mısır'da kurulmuştur. Hindistan'da 1923 yılında başka bir faizsiz sistem kurulmuş ve 20 yıl içinde sabit varlık değeri 2.240 ABD dolarına ulaşmıştır. Bu kuruluşlar küçük işletmelere küçük krediler vermiştir ve bu işleyiş 1960 yılına kadar devam etmiştir (Akın, 1986).

1950 yıllarından itibaren özellikle Orta Doğulu araştırmacılar tarafından yeni teknikler geliştirilmiştir. Hatta 1958 yılında bu yeni teknikler ile kooperatif bankası da kurmuşlardır. Bu kooperatif bankasının kurulmasına büyük arazi sahipleri öncülük etmiştirler. 1958 yılına gelindiğinde Müslüman araştırmacılar, sermaye ve vekâlet temelinde tasarruflu bir sistemin kurulabileceği fikrinin değerlendirilmesi üzerine paylaşmışlardır. 1963 yılında da ilk tasarruflu faizsiz sisteme geçiş denemeleri yapılmıştır. 1972'de de Bangladeş'te bir banka daha kurulmuştur. Bu banka küçük çiftçiler ve el sanatları alanında çalışmıştır. (Wilson, 2008; İkbal ve Greuning, 2008).

Yapılan araştırmalara göre faizsiz finans işlemleri orta çağlarda da İslam ülkelerinde uygulanmıştır. Bu uygulamadan elde edilen sonuçlar, bazı Avrupalı finansörler tarafından değerlendirilerek, tasarrufa dayalı faizsiz finans sistem modeli oluşturulup İslami kimliği canlandırma ve kuvvetlendirme çabası ile ortaya çıkarılmıştır.

2.2 Türkiye'de Tasarrufa Dayalı Faizsiz Finans Sistemi (TDFFS)

Türkiye İstatistik Kurumu hanehalkı tüketim harcamalarına göre, Türkiye'de bir ailenin tüketim amaçlı yaptığı harcamalar içinde en yüksek payı yüzde 23,7 ile konut ve kira harcamaları almaktadır (www.tuik.com, 2018). Genç nüfusun ve alınan göçün de etkisi ile konuta olan istek sürekli artan bir ilerleyiş göstermektedir. Müşteri ihtiyaçlarını karşılamakta eksik kalan bankacılık kesimine optimal çözüm olarak ortaya çıkan TDFFS'de konut alım işlemlerine ait detaylı prosedürler vardır. Faizsiz sistemin dışında kalan kuruluşlar müşterilerden yönetmelik gereği çok fazla evrak ve bordro istemektedir. Çoğu müşteri yönetmeliğin gerektirdiği isterleri yerine getiremediği için bankalardan kredi alamamaktadır. En önemlisi de mali olarak düşük gelirli ailelerin bankaya maaş bordrosu gösteremiyor oluşudur. Bazı işçilerin ücretlerinin ise resmiyette ve bordrolarında gerçeklik göstermediği görülmektedir. Bu sebepten aileler ya hiç bordro ibraz edememekte ya da bordroları kredi çekmeye yeterli bulunmamaktadır.

Yaşanan bu sıkıntıdan dolayı faiz hassasiyeti bulunan tasarruf sahiplerinin bankalarda yatırım yapmaktan kaçınarak tasarruflarında döviz ve altını tercih etmesine neden olmaktadır. Kayıt dışı tasarruflar için Türkiye'ye özgü olarak ev hanımlarının düzenli olarak toplandığı ve altın günü olarak adlandırılan bu toplantılarda çekiliş ile çıkan ev sahibine toplu olarak altın verilmesi uygulamasını taklit eden bu sistem, araç ve konut finansman modelinin temel taşlarını atmıştır. Yapılan araştırmada ailelerin TDFFS'yi tercih etmesinin temel sebebinin %94 oranında sistemin faizsiz olmasından kaynaklı olduğu gösterilmiştir (Usulcan, 2013).

Bankacılık sisteminin ihtiyacı olan insanlara kredi vermede 228 uygunsuz kriter kontrolü, bankacılık dışı alternatif finans yöntemlerinden birisi olan faizsiz sistemin cazibesini giderek artırmaktadır (Akpolat, 2018).

3. MÜŞTERİ KAYBI TAHMİNİ

3.1 Müşteri İlişkileri Yönetimine (CRM) Genel Bakış

CRM, açılımı “Customer Relation Management” olan müşterinin sisteme giriş sürecinden başlanılarak sistemi terk ediş sürecine kadar devam eden, yani müşteri yaşam döngüsünü sunan stratejiler ve süreçler bütünüdür (Ling ve Yen, 2001). Bu yaklaşım, müşterilerin nasıl farklılıklar gösterdiğini anlamak ve bu farklılıkların her bir müşteriye göre işletmenin nasıl davranması gerektiği konusunda bir planlama yapmasını gerektirmektedir (Roberts, 2001). CRM, yeni müşteri elde etmek, mevcut müşterileri elinde tutmak, şirket karını artırmak için farklı iletişim kanallarıyla müşteri davranışlarını anlamak ve müşteriyle etkileşime geçmek için kullanılan bir yaklaşımdır (Swift, 2001).

Tüm bu tanımlardan yola çıkarak müşteri ilişkileri yönetiminin organizasyon çıkarı için müşteri etkileşiminde kullanılan tüm strateji, yöntem ve süreçler olduğu söylenebilir. Böylece müşteri algısı değiştirebilir ve şekillendirebilir, müşteri bilgisi daha fazla önem kazanacağı için müşteri ile daha yakın ilişkiler kurulabilir ve müşterinin satın alma davranışları arttırılabilir. Müşteri ilişkileri yönetiminde başarılı olmak firmalar için rekabet ortamı oluşturmaktadır. Ayrıca iyi bir müşteri ilişkileri yönetimi müşteri memnuniyeti ve müşteriyi elde tutma oranlarını artırmak demektir.

CRM uygulamaları ile firmalar, kazanılması çok zor olan müşterilere isabet etmekle beraber, satış hedeflerini ve müşteri kazanımını, müşteri memnuniyetini arttırmakta ve bu müşterilere göre kampanyalarını özelleştirmekte iken, aynı zamanda da satış ve pazarlama maliyetlerini de azaltmaktadır.

İçinde bulunduğumuz dönemde işletmelerin en önemli öz varlığının “müşteri” olduğu gerçeğinden hareketle, bu öz varlığı koruyacak stratejileri geliştirmeleri gerekmektedir. İstatistiki sonuçlar da bu öz varlığı korumanın ne derece önemli olduğuna işaret etmektedir. Reichheld (2001) işletmelerin, 5 yıl içerisinde müşterilerinin yaklaşık %50’sini kaybettiklerini ama buna karşın müşteri elde tutma oranındaki %5’lik bir artışın ise toplam kazanca etkisinin %25-%100 arasında yükseliş sağladığını ifade etmektedir (Reichheld, 2001).

CRM hızla artan rekabet ortamı içerisinde, işletmelerin var olma çabası ile günden güne müşterilerine daha bağlı hale geldiği noktada, müşterilerinin sadakatini sağlamak ve memnuniyeti üst seviyede tutmak zorundadırlar.

CRM' nin amaçlarını özetle şöyle sıralamak mümkündür (Chen ve Popovich, 2003):

- Müşteri ilişkilerini karlı hale getirmek: Müşteri temsilcilerinin müşteriler ile uzun dönemli ilişki kurup, sistemde uzun süre kalmasını sağlamasıdır.

- Farklılaşma sağlamak: Müşteri portföyüne göre ürünlerde kişinin ihtiyacına özel kampanyaları sağlayabilmek, müşterileri ve ihtiyaçlarını birebir tanımak ve onlar için pazarlama yapmaktır.

- Maliyet minimizasyonu sağlamak: Hedef müşterilerin ihtiyacı düşünülerek tasarlanmış CRM projesine ayrılan bütçenin karlılığa dönüşmesi muhtemeldir. Mevcut müşterilerden oluşacak ilave satışlar, müşteriyi sisteme uzun süre dahil etmenin getireceği kazançlar, satış maliyetlerinde sağlanacak tasarruf ve şirket içi iletişim maliyetlerindeki azalmalar göz önünde bulundurulursa ayrılan bütçe çok zaman geçmeden geri kazanılabilecektir.

- İşletmenin verimini artırmak: İşletmeler müşteri memnuniyetini ve rekabet ortamını göz önünde bulundurarak verim arttırma çalışmalarına gitmektedir. Bunun için en iyi çözüm bilişim yatırımlarıdır.

- Uyumlu faaliyetler sağlamak: CRM' in amacı satış faaliyetlerini online hizmetler ile birleştirebilmek ve tüm faaliyetler ile uyumlu olarak hizmet vermeyi sağlamaktır. Bu doğrultuda hem geleneksel satış kanallarından elde edilen bilgiler, hem de diğer alanlardan alınan bilgiler toparlanarak yüksek düzeyde müşteri bilgisi elde edilebilecektir.

- Müşteri taleplerini karşılamak: Müşterilerden alınan geri bildirimler sayesinde hizmeti onların istediği şekilde gerçekleştirmek, hızlı hale getirmek ve memnuniyeti arttırmak mümkündür.

3.2 CRM Bileşenleri

İşletmenin müşterileri ile olan etkileşimini en verimli hale getirmeyi hedefleyen CRM, üç temel bileşenden oluşmaktadır. Bunlar Şekil 3.1'de görüldüğü üzere insan, süreç ve teknoloji bileşenleridir. İnsan faktörü, müşteri beklentilerini iyi anlayıp özümseyen ve bu doğrultuda da gerektiğinde standart süreçlerin ötesinde çözümler oluşturarak müşteri odaklı çalışma modelini anlayan yapının en temel unsurudur. Süreç bileşeni, müşteri talepleri doğrultusunda işletme süreçlerinde gerekli güncellemeler yapılarak iş yapış modelinin müşteri odaklı bir yapıya dönüştürülmesini sağlamaktadır. Teknoloji bileşeni ise müşteri etkileşimi olan tüm noktalarda müşteri bilgilerine erişimi mümkün kılan ve müşteri istekleri doğrultusunda çözüm üretilmesini sağlayan alt yapı unsurudur. CRM; insan, süreçler ve teknoloji bileşenlerinin üstüne kurulan bir iş stratejisi olarak tüm işletme çalışanları tarafından benimsenmesi ve bu doğrultuda iş yapılması halinde başarılı sonuçlar vermektedir.

Şekil 3.1: CRM bileşenleri

CRM felsefesinin hayata geçirilmesinde bileşenlerin önemine ilişkin yapılan araştırma sonuçlarına göre, insan faktörünün en önemli bileşen olduğu açıkça ifade edilmektedir. CRM danışmanlığı veren ISM'nin, 2007'de CRM bileşenlerinin önem derecesini vurgulamak üzere yaptığı araştırma sonucu şu şekildedir (Russell, 2007):

- İnsan : %60
- Süreç : %30
- Teknoloji : %10

CRM Institude Türkiye'nin yaptığı bir araştırma sonucuna göre ise insan faktörü ISM'nin sonucuna paralel şekilde birinci sırada olmakla birlikte, teknoloji ikinci sırada, süreç ise üçüncü sırada yer almaktadır. Türkiye'de, dünyadaki sıralamanın tersine olarak teknolojiye süreçten daha fazla önem verildiği görülmektedir (Odabaşı, 2000):

- İnsan : %45
- Teknoloji : %31
- Süreç : %24

CRM sadece teknolojik bir çözüm olmayıp teknoloji ile desteklenen, sadece mükemmel süreçler sistemi olmayıp müşteri odaklı süreçlerle ile hayata geçen, her şeyden önce de kültürel değişim gerektiren, üst yönetimden en alt seviye çalışana kadar tüm işletme kadrosunun benimsemesi, sahip çıkması gereken bir yönetim bütünüdür. Dolayısıyla bu stratejinin başarısı için de insan faktörü, süreç ve teknoloji faktörüne göre oldukça ağırlıklı öneme sahip olmaktadır.

3.3 CRM Mimarisi

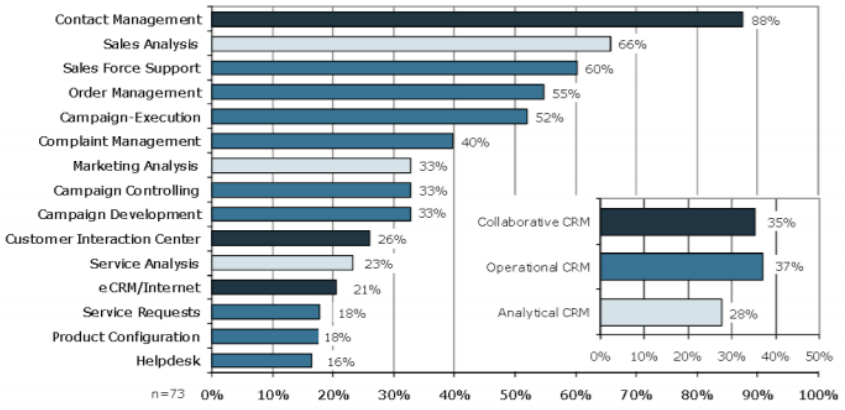
CRM, işletmelerin müşteri stratejilerine dayalı bir yönetimi hedeflediği için her işletmenin kendi iş yapma modeline göre de şekillenmektedir. Bazı işletmeler CRM'i müşterilerinin geçmiş davranış modellerine ve demografik bilgilerine dayanarak onlara yeni öneriler geliştirmekte, bazıları ise müşterinin karlılığını ve müşteriden elde edilen geliri artırmak üzere ve bazıları da müşteri profillerine ve tercihlerine göre internet uygulamalarını şekillendirmek üzere değerlendirmektedir. Bu nedenle farklı işlevi olan CRM uygulamaları söz konusu olmaktadır.

Fayermen (2002) ve Paas ve Kuijlen (2001) CRM' i; Operasyonel CRM, İşbirlikçi CRM ve Analitik CRM olmak üzere üç ana grup altında toplamışlardır. Torggler (2009) da gerçekleştirdiği çalışmasında, mevcut CRM uygulamalarının sahip olduğu fonksiyonel analizi ve çeşitli pazar araştırmaları değerlendirmelerini göz önünde bulundurarak CRM'i üç ana grupta değerlendirmiştir. Torggler'in üç ana grup ve alt kırılımları Çizelge 3.1'de gösterildiği şekilde özetlemektedir.

Çizelge 3.1: CRM kategorizasyonu (Torggler, 2009).

İşbirlikçi CRM	Kontak Yönetimi		eCRM/Internet		Müşteri Etkileşim Merkezi		
Operasyonel CRM	Pazarlama Otomasyonu	Kampanya Geliştirme	Satış Otomasyonu	Sipariş Yönetimi	Servis Otomasyonu	Yardım Masası	
		Kampanya Devreye Alma		Satış Gücü Desteği		Şikayet Yönetimi	
		Kampanya Kontrol		Ürün Konfigürasyonu		Servis Çağrısı	
Analitik CRM	Pazarlama Analizi		Satış Analizi		Servis Analizi		

Torggler (2009) yaptığı araştırmada, Avusturya'da CRM kullanımının dağılımını analiz ederek Şekil 3.2'deki grafiği hazırlamıştır. Bu sonuçlara göre en yaygın olarak %37 oranında Operasyonel CRM kullanılmakta olup, devamında ise %35 oranında İşbirlikçi CRM ve %28 oranında ise Analitik CRM kullanılmaktadır. Alt kırılımlara bakıldığında ise kontak yönetiminin %88 ile en yüksek oranda kullanıldığı görülmektedir. Araştırmaya katılan firmalarda, Analitik CRM en az kullanılan kategori olmakla birlikte alt kategori alanında bakıldığında ise satış analizi alt kategorisinin %66 ile ikinci en fazla kullanılan kategori olduğu görülmektedir.



Şekil 3.2: CRM fonksiyonlarının kullanım grafiği (Torggler, 2009)

Firmalar başlangıçta operasyonel ve işbirlikçi CRM' i daha yaygın olarak kullanmış olsalar da zaman içinde müşteri seçme, müşteri edinme müşteri koruma ve müşteri derinleştirmeye ilişkin stratejik kararlarda analitik araçların ne kadar önemli olduğunun farkına vararak, Analitik CRM üzerinde daha fazla yoğunlaşmaya başlamışlardır.

4. MAKİNE ÖĞRENME TEKNİKLERİ

4.1 Makine Öğrenimine Genel Bakış

Makine öğrenme hem istatistik biliminin hem de bilgisayar biliminin konusudur. Bu alan çok yakın dönemde duyulmaya başlansa da istatistik biliminin bu alandaki çalışmaları 1950'li yıllara dayanmaktadır. O yıllarda alanın sadece akademiyle sınırlı kalmasının sebebi geliştirilen algoritmaların etkin ve hızlı bir şekilde çalıştırılabileceği bilgisayar yazılım ve donanımlarının bulunmamasıydı. Dolayısıyla, makine öğrenmesinin bu alandaki algoritmalarının faydalı olabilmesi için veri üzerinde çalışmalar yapılması ile birlikte güçlü donanım ve yazılımlar gerektirmektedir. 1980'den sonra bilgisayar bilimi alanındaki gelişmeler bu algoritmaların pratikteki kullanımlarını kolaylaştırmış ve yaygınlaştırmış ve bu alanı günümüzde popüler bir araştırma ve uygulama mecrası haline getirmiştir.

Makine öğrenmesi, yüz tanımada, internetten alışveriş yapmada, sosyal medyanın kullanımında, sahtekarlık tespitinde ya da bankalarla iletişime geçmede kullanılır. Makine öğrenmesi algoritmaları iki ana kola ayrılır. Denetimli makine öğrenmesi ve denetimsiz makine öğrenmesidir.

- **Denetimli Makine Öğrenmesi**

Denetimli makine öğrenmesi etiketlenmiş veri kümeleri ile hangi sonuçlara ulaşılması gerektiğini öğretir. Denetimsiz makine öğrenmesi algoritmalarına göre daha çok tercih edilir. Etiketlenmiş veri, gözlemler sonrası algoritmaya etiket yaptırır. Her gözlemde öğrendiklerini gerçek tahminler yapmada kullanır.

- **Denetimsiz Makine Öğrenmesi**

Denetimsiz makine öğrenmesi algoritmaları tanımlanmış çıktısı olmayan, isimlendirilmemiş verileri kullanarak öğrenir. Denetimsiz makine öğrenimi kategorizasyon için yararlı olabilecek özelliklerin bulunmasında yardımcı olur.

4.2 Makine Öğrenmesinin İş Hayatındaki Yararları

Makine öğrenmesinden günümüzde pek çok alanda yararlanılmaktadır. İş hayatında ise farklı amaçlar için optimum sonuca ulaşmayı amaç edinenlerin sayısı her gün çoğalmaktadır. Makine Öğrenmesinden iş dünyasının faydalandığı alanlar:

- Müşterileri anlama ve elde tutma
- Müşteri kayıplarını belirlemede
- CRM sistemlerinde öncülü belirlemede
- Dinamik fiyatlandırma yapmada
- Müşteri Sınıflandırmada
- İnsan Kaynaklarında
- Öneride bulunma
- Verileri organize etme ve iş birliğinde arttırma
- Tahminde bulunma

4.3 Keşifsel Veri Analizi (EDA)

Keşifsel veri analizi, verileri analiz etmek ve temel özelliklerini anlayarak özetlemek için kullanılan bir yaklaşımdır. Keşifsel veri analizi, istatistiksel testler ve grafikler kullanarak verileri görselleştirmede ve daha fazlasını elde etmede kullanılır.

Keşifsel veri analizinin adımları aşağıdaki gibidir.

- İstatistiksel test; korelasyon elde etmek için Pearson Korelasyonu, Spearman Korelasyonu, Kendall testi gibi bazı istatistiksel testler yapılır.
- Niceliksel test; sayısal özelliklerin yayılımını ve kategorik özelliklerin sayısını bulmak için testler kullanılır.
- Görselleştirme; görselleştirme verilerin anlaşılması için çok önemlidir. Kategorik özelliklerin anlaşılması için çubuk grafikler, pasta grafikler gibi grafik teknikleri kullanılırken, sayısal özellikler için dağılım grafikleri ve histogram kullanılır.

4.4 Lojistik Regresyon (LR)

Lojistik regresyon, doğrusal regresyonun geniş bir kümesi olan genelleştirilmiş doğrusal modeller (GLM) ailesine üye tahminleme modelidir. Adından da anlaşılacağı gibi lojistik regresyon tıpkı doğrusal regresyon modeli gibi parametrik ve açıklayıcı değişkenlerle bağımlı veriler arasında ilişkiyi tahminleyen bir model biçimidir. Çoklu sınıflandırma (binary classification) problemlerini tahmin etmede kullanılır. Problemin konusu olan olayın olma ihtimali p ile tanımlanacak olursa lojistik regresyon aşağıdaki gibi lineer bir model çizer.

$$\log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \quad (1)$$

Burada k, bağımsız değişkenlerin toplam sayısını ifade etmektedir. β model katsayısını belirlemektedir. Belirlenen model katsayıları ile hesaplanan p değeri olasılık değeri olup, 0 ile 1 arasında olmak zorundadır. p aşağıdaki gibi ifade edilir.

$$p = \frac{1}{1+\exp[-(\beta_0+\beta_1 x_1+\cdots+\beta_k x_k)]} \quad (2)$$

Bu fonksiyona literatürde sigmoid fonksiyonu adı verilmektedir.

Lojistik regresyon, açıklayıcı değişkenler üssel temel fonksiyonlar ile dönüştürülemediği zaman doğrusal karar sınırları oluşturur. Bu durum modelin tahminleme performansını düşürdüğü için modelin en büyük dezavantajıdır. Çünkü gerçek dünyada birçok problem yapısı gereği doğrusal olmayan karar sınırlarını gerektirmektedir. Bunun yanında açıklayıcı değişkenlerin eğim parametrelerinin istatistiksel olarak anlamlı olup olmadığını anlamak için fırsat tanıdığından istatistiksel çıkarım problemlerine uygundur. Tahminleme başarısı lojistik regresyondan çok daha iyi olan birçok modelde istatistiksel çıkarım yapmak mümkün değildir (Kuhn ve Johnson, 2013).

Veri kümesini test etmek, eğitmek ve lojistik regresyon modelinin performansını sadece test kümesinde değil, tüm veri kümesinde değerlendirmek için cross validation (çapraz doğrulama) da kullanılır. Lojistik regresyonda kullanacağımız diğer parametrelerden bir diğeri de grid search hiper-parametresidir.

Grid search ile hiper-parametre seçim işleminde, belirlenen aralıkta bulunan tüm verilerin kombinasyonlarının sonuçları gözlenir ve duruma göre en iyi kombinasyon hiper-parametre grubu olarak seçilir (Brownlee, 2016)

4.5 K En Yakın Komşu Algoritması (KNN)

Denetimli öğrenme yöntemlerinden biri olan K En Yakın Komşu algoritması hem sınıflama hem de regresyon ayağında kullanılabilen çok amaçlı bir algoritmadır. KNN ile temelde yeni noktaya en yakın noktalar aranır. K, bilinmeyen noktanın en yakın komşularının miktarını temsil eder. Sonuçları tahmin etmek için algoritmanın k miktarı seçilir. Bilinmeyen veri, eğitim kümesindeki diğer veriler ile karşılaştırılarak bir uzaklık ölçümü yapılır. Hesaplanan uzaklığa göre bir sınıfa atanamamış veriye en optimal sınıf bulunur.

KNN algoritmalarında temel yaklaşım, benzer nokta ya da değişken gruplarının yüksek ihtimal ile aynı sınıfa ait olmasıdır. Bu noktada, seçilmiş bir mesafe ölçütü kullanılarak sınıfı bilinmeyen verinin yakınlığı bulunur. Mesafe hesaplamada en çok kullanılan uzaklık ölçüsü, ôklid uzaklığıdır (Hu, 2016).

Ôklid uzaklığı, iki nokta arasında, $x_1 = (x_{11}, x_{12} \dots x_n)$ ve $x_2 = (x_{21}, x_{22} \dots x_{2n})$ olmak üzere, eşitlik 3'te verildiği gibidir.

$$dist_{\ddot{o}klid}(x_1 x_2) = \sqrt{\sum_{i=1}^{n}(x_{1i} - x_{2i})^2} \quad (3)$$

Ôklid uzaklığı dışında Manhattan, Minkowski, Chebyshev gibi farklı uzaklık hesaplama ölçütleri de kullanılabilir (Prasath, 2019). Söz konusu ölçütler aşağıdaki gibidir.

$$dist_{minkowski}(x_1 x_2) = \sqrt[r]{\sum_{i=1}^{n}(x_{1i} - x_{2i})^r} \quad (4)$$

$$dist_{manhattan}(x_1 x_2) = \sum_{i=1}^{n}|x_{1i} - x_{2i}| \quad (5)$$

$$dist_{chebyshev}(x_1 x_2) = \max_i |x_{1i} - x_{2i}| \quad (6)$$

Farklı uzaklık ölçütleri kullanılarak KNN algoritmasının kendi içinde daha doğru sınıflandırmalar yapılabildiğini belirten çalışmalar mevcuttur (Weinberger, 2006).

4.6 Destek Vektör Makinesi (SVM)

Destek vektör makineleri (SVM), orijinal adı olarak Support Vector Machine olarak da bilinmektedir. İstatistiksel öğrenme kuramı ve yapısal risk olarak tanıtılmıştır.

Destek vektör makineleri istatistik ve makine öğrenmesinde çeşitli uygulama içermektedir. Destek vektör makineleri sınıflandırma problemlerinde en çok kullanılan yöntemlerden birisidir. Yüksek seviyede doğru sonuç vermesi, doğrusal ve doğrusal olmayan verileri modelleyebilmesi, birbirinden bağımsız çok sayıda değişken ile çalışabilmesi, iyi bir sınıflandırma yapabilmesi tercih edilme gerekçelerindendir. SVM yöntemi VC'de kullanılan diğer algoritmalar ile karşılaştırıldığında daha az karmaşık oluşu ile diğer yöntemlerden farklılık göstermektedir (Osowski, Siwekand, ve Markiewicz, 2004). Bundan dolayı büyük verilerin sınıflandırılmasında diğer yöntemlerden daha uygundur.

Destek Vektör Makinesi yaklaşımının temel avantajları şunlardır (Osowski, Siwekand, ve Markiewicz, 2004):

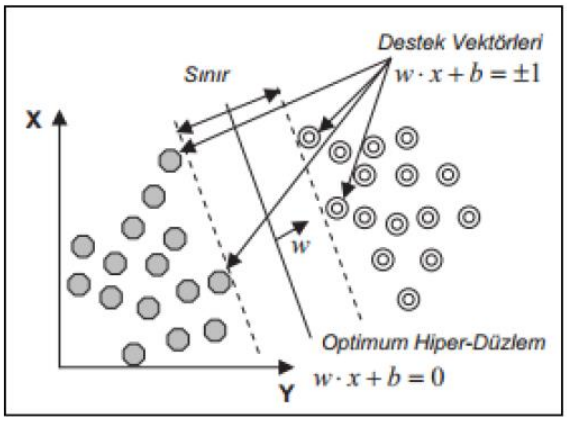
- Ampirik riskin minimizasyonu ve aşırı uyumun önlenmesi arasında bir orta yol bulmaya çalışmaktadırlar.
- Konveks bir kuadratik programlama problemidir.
- Öğrenme tekniği; eğitim kümesindeki çok az veriyle bile iyi geneller ve genelleme hatası üzerindeki sınırlar, doğrudan eğitim verilerinde tahmin edilebilir.

4.6.1 Doğrusal Ayrılabilen Veriler için Destek Vektör Makineleri

Destek vektör makineleri ile gruplandırma yapılırken doğrusal olarak ayrılabilen veriler için, geliştirilme aşamasında olan bilgi ile elde edilen karar fonksiyonu kullanılarak birbirlerinden ayrılan iki sınıf oluşturabilmek amaçlanmaktadır. SVM girdi olarak seçilen veri kümesi içerisindeki fonksiyonlara göre bu fonksiyondan sağlanan çıktılardan iki sınıfa ait olma şartı ile sınıflandırmaktadır (Meyrueıs, Soubarı, Guessoum, Namane, 2014).

Elde edilen sınıfa ait aralarındaki mesafenin en az olduğu iki mesafe en yükseğe çekilerek hiper düzlem elde edilir. Bu sayede görünmeyen verilerde en iyi şekilde yayılması sağlanır. Hiper düzlemin uzaklığı, verilerin en geniş alanda ve en iyi şekilde

sınıflandırılmasıdır. Aralarındaki mesafenin en az olduğu düğümlerin birbirlerine olan mesafesi en yükseğe çıkarılarak en yüksek seviyedeki uzaklığı veren hiper düzlem tercih edilmektedir buna da optimum hiper düzlem denilmektedir. Oluşan ilgili gurubun çizgileri hiper düzleme paralel bir düzlem hattında bulunmaktadır (Burges, 1998).

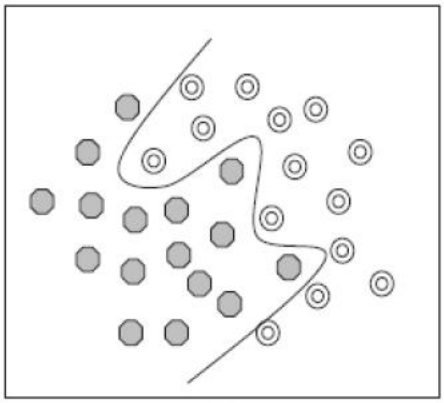


Şekil 4.1: Doğrusal olarak ayrılabilen veriler için optimum hiper düzlemin tayin edilmesi (Vapnik, 1995)

İki boyutlu bir sınıflandırma problemi için doğrusal SVM'nin geometrik gösterimi yukarıdaki Şekil 4.1'de gösterilmiştir. Destek vektörleri ayırma hiper düzlemine en yakın olan iki sınıfa ait örnekler olarak ifade edilmektedir (Burges, 1998).

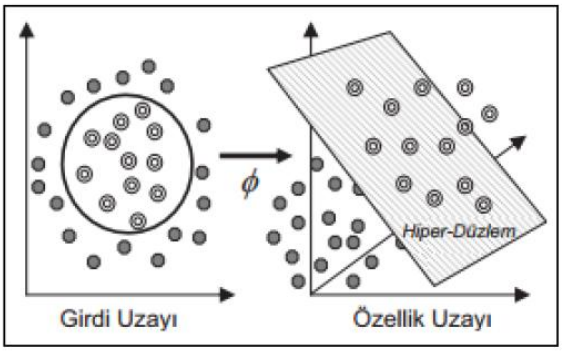
4.6.2 Doğrusal Olarak Ayrılamayan Veriler İçin Destek Vektör Makineleri

Doğrusal olmayan SVM, veri kümesinin doğrusal bir fonksiyonla belirli bir hata ile ayrılamaması durumunda kullanılan algoritmalardır. Gerçek yaşam problemlerinde bir veri kümesinin hiper düzlem ile doğrusal olarak ayrılması genellikle mümkün değildir. Dolayısıyla sınıfların ayrılma işlemi, ayırma eğrisinin keşif edilmesiyle mümkün olmaktadır. Veri kümesinin doğrusal ayrılamama durumunun geometrik gösterimi Şekil 4.2'de verilmiştir. Bu durumda p boyutlu girdi vektörü x'in P boyutlu özellik vektörü Φ' ye dönüştürülmesi gerekmektedir (Cortes ve Vapnik, 1995).



Şekil 4.2: Doğrusal olarak ayrılmayan destek vektör

Bu amacı gerçekleştirmek için doğrusal olmayan haritalama yaklaşımından yararlanılır (Busuttil, 2003).



Şekil 4.3: Verinin daha yüksek bir boyuta dönüştürülmesi (Kavzaoğlu ve Çölkesen, 2010)

Doğrusal olmayan haritalama, orijinal girdi uzayı x'in bir Hilbert uzayı olan daha yüksek boyutlu F özellik uzayına dönüştürülerek doğrusal ayrımının gerçekleştirilmesi için kullanılan bir yaklaşımdır (Suykens, 2002). Hilbert uzayı, pozitif skaler çarpıma sahip ve öğeleri fonksiyonlardan oluşan tam iç çarpım uzayları olarak ifade edilmektedir (Cortes ve Vapnik, 1995). Doğrusal olmayan haritalama yaklaşımı ile iki boyutlu veri seti üç boyutlu özellik uzayına taşınarak veri setinin doğrusal ayrımı sağlanılabilmektedir.

4.7 Performans Değerlendirme Yöntemi

Churn tahmini gibi karmaşık veri madenciliği problemlerinde herhangi bir sınıflandırma yöntemi her probleme uygulanabilir değildir. Yapay öğrenme algoritmaları farklı modelleri oluşturmak için farklı değerlerle ve ayarlamalarla çalıştırılmaktadır. Modeller oluşturulduktan sonra, iyi bir tahmin yapmak ve en iyi modeli seçmek için modellerin karşılaştırılmaları gerekmektedir. Algoritmanın modellerinin ne kadar iyi tahmin edilip karşılaştırabildiğini bize söyleyen skorlama ölçütüne ihtiyaç duyulmaktadır.

Test veri kümesine yeni bir örnek geldiğinde, model bu veriyi kendi algoritmasına göre tahmin etmektedir. Çizelge 4.1'de belirtildiği gibi, tahmin doğruysa (test veri kümesindeki ile aynı değerde), doğru olarak sayılmaktadır. Tahmin yanlış ise, yanlış olarak sayılmaktadır. Eğer tahmin negatif (Churn olmamış) ve doğruysa tahmin **doğru negatif**, eğer tahmin pozitif (Churn olan) ve doğruysa tahmin **doğru pozitif** olarak sayılmaktadır. Tahmin negatif ancak gerçek sınıf pozitif ise **yanlış negatif**, tahmin pozitif ancak gerçek sınıf negatifse de **yanlış pozitif** olarak adlandırılmaktadır. Hata matrisi genellikle birçok ölçümün temelini oluşturmaktadır.

Çizelge 4.1: Hata Matrisi

	Gerçek Kayıp Müşteriler	**Gerçek Sadık Müşteriler**
Tahmin edilen Kayıp Müşteriler	Doğru pozitif	Yanlış Pozitif
Tahmin edilen Sadık Müşteriler	Yanlış negatif	Doğru negatif

Hata matrisine dayanan performans ölçümlerinin tanımları aşağıdaki gibidir.

$$yprate = Yanlış\ pozitif\ oran = \frac{YP}{N} \quad (7)$$

$$dprate = Doğru\ pozitif\ oran = \frac{DP}{N} \quad (8)$$

$$Duyarlık\ (Sensitivity) = \frac{DP}{(DP+YN)} \quad (9)$$

$$Kesinlik\ (Precision) = \frac{DP}{(DP+YP)} \quad (10)$$

$$\text{Özgüllük } (Specificity) = \frac{DN}{(DN+YP)} \quad (11)$$

$$Hatasızlık\ (Accuracy) = \frac{(DP+DN)x100}{DP+DN+YP+YN} \quad (12)$$

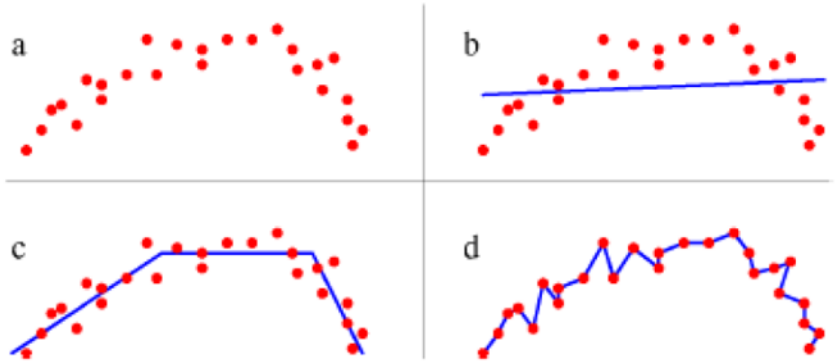
$$F = \frac{2\ x\ Kesinlik * Duyarlılık}{Kesinlik+Duyarlılık} \quad (13)$$

P = Gerçek pozitif sayısı
N = Gerçek negatif sayısı
DP = Doğru pozitifler
DN = Doğru negatifler
YP = Yanlış pozitifler
YN = Yanlış negatifler' dir.

Farklı sınıflayıcıları değerlendirmek için kullanılabilecek bazı ölçütlere örnek olarak hatasızlık, F-ölçüsü, kaldırma tablosu (lift-chart), ROC alanı verilebilmektedir. Her bir ölçüm performansı değerlendirmede farklı bir yaklaşımı ifade etmektedir (Duda, Hart ve Stork, 1999).

Değerlendirilmesi gereken diğer iki konu ise; aşırı öğrenme (over-fitting) ve eksik öğrenmedir (under fitting). Aşırı öğrenme ve eksik öğrenme ile yalnızca Churn tahmininde değil tüm sınıflandırma problemlerinde karşılaşılabilmektedir. Genellikle sınıflandırma modelinin veya yapısının aşırı karmaşık olması neticesinde meydana gelir ve verilerde bulunmayan kalıplar keşfedilebilir.

Şekil 4.4'de veri kümesinin veri noktaları (a), eksik öğrenme modeli (b), fit model (c) ve aşırı öğrenme modeli (d) gösterilmektedir.



Şekil 4.4: Eksik öğrenme, fit ve aşırı öğrenme modeli (Lee,Ivrissimtzis ve Seide, 2006)

Doğru pozitif oran, Churn olarak sınıflandırılan Churn sayısı (doğru pozitifler) ile gerçek Churn sayısı (Pozitif = DP + YN) arasındaki orandır. Yanlış pozitif oran, gerçekte Churn olmayan ancak tahminde Churn kabul edilen kişi sayısı (yanlış pozitif) ile gerçek Churn olmayanların sayısı (Negatif = YP + DN) arasındaki orandır.

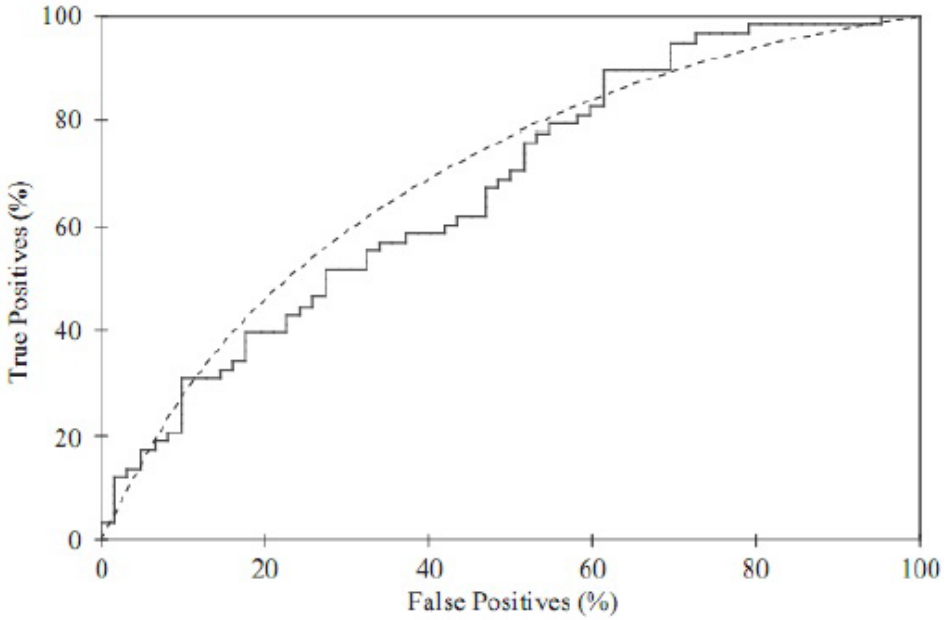
Churn tahmini alanı ve veri kümemiz için en uygun ölçek, bir sonraki alt bölümde detaylandırılacak olan ROC alanıdır.

4.7.1. ROC/ AUC Eğrisi

ROC (Receiver Operating Characteristic) farklı sınıflar için bir olasılık eğrisidir. ROC eğrileri, sınıf dağılımı veya hata değerleri dikkate alınmaksızın bir sınıflandırıcının performansını göstermektedir. Dikey eksendeki doğru pozitif oranına karşılık (hassasiyet) yatay eksen üzerinde yanlış pozitif oranı (özgüllük) çizilmektedir.

Grafikteki her adım, eşiği değiştirerek oluşan bir noktadır. ROC eğrisindeki her nokta, bir karar eşiğine karşılık gelen bir hassasiyet ve özgürlük çiftini temsil etmektedir. Bu grafiği kullanarak, hassasiyet ve özgüllük arasındaki en uygun denge noktası tespit edilebilmektedir. ROC analizi ayrıca, herhangi bir eşikten farklı olarak bir sınıflandırıcının öngörme kabiliyetini değerlendirmeyi sağlamaktadır. AUC (area under curve) denilen ROC eğrisinin altındaki alan, çeşitli sınıflandırıcıların doğruluğunu karşılaştırmak için ortak bir önlemdir. ROC, bir yöntemin örnekleri doğru

şekilde sınıflandırma yeteneğini değerlendirmektedir. Bu yaklaşıma göre, en büyük AUC' ye sahip sınıflandırıcı daha iyi kabul edilecektir. Bir sınıflandırıcının AUC' si 1'e yakınsa, doğruluğu daha yüksek demektir.



Şekil 4.5: ROC eğrisi (Witten, Frank ve Hall, 2011)

5.UYGULAMA

5.1 Veri Kümesine Genel Bakış

Uygulamada faizsiz finans firmasından alınan bir senelik müşteri verisi kullanılmıştır. Veriler 2020 yılına ait olup, çalışmanın konusu olan churn analizi de yıllık olarak incelenmiştir. Müşterilerin bir yıllık güncel durumu göz önünde bulundurularak ayrılan müşteriler churn olarak kabul edilmektedir. Veriler üzerinde churn analizine geçilmeden önce, kullanılan veri kümesini daha iyi anlamak adına keşifsel veri analizi incelemesi yapılmıştır. RF ile değişkenlerin modele katkıları, önem düzeyleri incelenmiş, sınıf bilgisi ile doğrudan ilişkili ve aşırı öğrenmeye neden olabilecek ayrılma nedeni değişkeni veri kümesinden çıkarılmıştır. Çalışmada uygun hale getirilen veriler ile Lojistik regresyon, k en yakın komşu ve destek vektör algoritmaları kullanılarak müşteri kaybı analizi tahmininin nasıl yapıldığı gösterilmektedir.

Müşteri kaybının inceleneceği veri kümesi 18507 müşteri ve 14 değişkenden oluşmaktadır. Müşteri kayıp analizinin inceleneceği değişkenler Çizelge 5.1'de açıklanmıştır.

Çizelge 5.1: Değişkenler ve tanımları

Medeni_durum	Müşterilerin evli veya bekar olup olmadığını gösteriyor.
Meslek_grubu	Sisteme giren müşterilen hangi sektör grubunda olduğunu gösteriyor.
Gor_orjin	Sisteme dahil olan müşterilere veya müşteri olması istenilen kişilere ulaşılan alanları gösterir.
Gorusme_sayi	Firmanın müşterisi olan veya müşterisi olması için yaptığı görüşme sayılarını gösteriyor.
Gor_mecra	Müşterilerin sisteme ilk temas ettikleri alanları gösteriyor.
Kampanya_cins	Müşterilerin kampanya modellerinde hangi sisteme (Ev, Otomotiv) girdiklerini gösteriyor.
Kampanya_tip	Firmanın müşterilerine sunduğu çekilişli ve çekilişsiz kampanya gruplarını gösteriyor.
Kampanya_bedeli	Müşterilerin sisteme dahil oldukları bedelleri gösteriyor.

Vade	Müşterilerin dahil olduğu çekiliş grubunu ifade ediyor.
Engeçteslim_ay	Müşteriye çekilişle çıkmaz ise en son kaçıncı ayda teslimatını alabileceğini ifade ediyor.
Sube_tip	Müşterilerin sisteme girdikleri yeri gösteriyor.
Meslek_kod	Sisteme dahil olan müşterilerin meslek grubunu gösteriyor.
Bölge	Müşterilerin sisteme dahil oldukları bölgeleri ifade etmektedir.
Ayrılma_nedeni	Sisteme dahil olan müşterilerin neden ayrılmak istediğini gösteriyor.

Veri kümesi Python uygulaması ile Numpy ve Pandas kütüphanelerinde incelenmiştir. Uygulamanın geliştirilebilmesi için belirlenen paketler import komutu ile uygulamaya dahil edilmiştir.

Pandas ileri düzey veri yapıları araçlarını barındırmaktadır. Python'da daha hızlı analiz yapmak için geliştirilmiştir. Pandas, açık kaynaklı bir kütüphanedir. Söz konusu durum programlamada Python dilinin yoğun tercih edilmesine zemin hazırlamıştır. Verileri okumak için read_csv() fonksiyonu kullanılmıştır. Söz konusu ilk parametre verinin bulunduğu csv dosyasıdır. İkinci parametrede ise oluşturulacak dizilerin hangi ayırıcı karakter ile belirleneceği bilgisini içermektedir. Csv formatında import edilen veri kümesini anlamak için önce veriye ait değişkenleri incelemek gerekmektedir. İncelenen veri kümesi Çizelge 5.2'de gösterilmiştir.

Çizelge 5.2: Örnek veri kümesi

Cinsiyet	Medeni_durum	Meslek_grubu	Gor_orjin	Gorusme_Sayi	Gor_mecra	Kampanya_cins	Kampanya_tip	Kampanya_bed	Vade	Engecteslim_ay	Sube_tip	Meslek_kod	Bölge	Churn
Kadın	Evli	OZEL_SEKTOR	Eski_Musteri	17	DİĞER	EV	EV-CEKILIS	500000	240	73	Sube	İŞÇİ	MARMARA 3	0
Erkek	Evli	OZEL_SEKTOR	Giden_Arama	9	DİĞER	OTOMOTIV	OTOMOTIV-CEK	30500	61	33	Sube	İŞÇİ	KONYA	1
Erkek	Bekar	OZEL_SEKTOR	Giden_Arama	42	NULL	EV	EV-CEKILIS	160000	160	52	Sube	İŞÇİ	İÇ ANADOLU	0
Erkek	Evli	OZEL_SEKTOR	Gelen_Arama	29	DİĞER	EV	EV-CEKILIS	81000	160	52	Sube	EMEKLİ	GÜNEYDOĞU	0
Erkek	Evli	OZEL_SEKTOR	Eski_Musteri	39	DİĞER	OTOMOTIV	OTOMOTIV-CEK	60000	61	33	Sube	TEKSTİLCİ	İSTANBUL ANAD	1
Erkek	Evli	OZEL_SEKTOR	Gelen_Arama	38	DİĞER	EV	EV-CEKILIS	300000	240	73	Sube	İŞÇİ	MARMARA 1	1
Erkek	Evli	OZEL_SEKTOR	Gelen_Arama	38	DİĞER	EV	EV-CEKILIS	400000	140	36	Sube	İŞÇİ	MARMARA 1	0
Kadın	Evli	OZEL_SEKTOR	Eski_Musteri	12	DİĞER	OTOMOTIV	OTOMOTIV-CEK	28000	61	33	Sube	MÜŞTERİ TEMSİ	MARMARA 4	0
Kadın	Evli	OZEL_SEKTOR	Eski_Musteri	36	DİĞER	OTOMOTIV	OTOMOTIV-CEK	40000	37	37	Sube	İŞÇİ	MARMARA 4	0
Kadın	Evli	OZEL_SEKTOR	Eski_Musteri	36	DİĞER	OTOMOTIV	OTOMOTIV-CEK	40000	31	15	Sube	İŞÇİ	MARMARA 4	1
Erkek	Evli	OZEL_SEKTOR	Eski_Musteri	26	NULL	EV	EV-OZELIHTIYAC	150000	39	9	Sube	İŞÇİ	KONYA	0
Erkek	Evli	OZEL_SEKTOR	Eski_Musteri	29	DİĞER	OTOMOTIV	OTOMOTIV-OZE	50000	15	4	Sube	KAYNAKÇI	İSTANBUL ANAD	0
Erkek	Evli	OZEL_SEKTOR	FMT_Araması	34	NULL	EV	EV-CEKILIS	300000	120	41	Sube	DİĞERLERİ	MARMARA1	0
Erkek	Evli	OZEL_SEKTOR	Eski_Musteri	22	DİĞER	EV	EV-OZELIHTIYAC	300000	60	9	Sube	KUYUMCU	MARMARA 3	1
Erkek	Evli	OZEL_SEKTOR	Eski_Musteri	59	DİĞER	OTOMOTIV	OTOMOTIV-CEK	50000	37	17	Sube	TEKNİSYEN	KARADENİZ	0
Erkek	Evli	OZEL_SEKTOR	Eski_Musteri	59	DİĞER	OTOMOTIV	OTOMOTIV-CEK	25000	37	37	Sube	TEKNİSYEN	KARADENİZ	0
Erkek	Evli	OZEL_SEKTOR	Eski_Musteri	43	DİĞER	EV	EV-CEKILIS	200000	140	56	Sube	OTO TAMİRCİSİ	İÇ ANADOLU	0
Erkek	Evli	OZEL_SEKTOR	Gelen_Arama	13	DİĞER	OTOMOTIV	OTOMOTIV-OZE	75000	34	11	Sube	İŞÇİ	GÜNEYDOĞU	0
Erkek	Evli	OZEL_SEKTOR	Gelen_Arama	13	DİĞER	EV	EV-CEKILIS	350000	180	92	Sube	İŞÇİ	GÜNEYDOĞU	0
Erkek	Evli	OZEL_SEKTOR	Gelen_Arama	13	DİĞER	OTOMOTIV	OTOMOTIV-OZE	70000	32	10	Sube	İŞÇİ	GÜNEYDOĞU	0
Erkek	Evli	OZEL_SEKTOR	Gelen_Arama	10	İNTERNET	OTOMOTIV	OTOMOTIV-CEK	25000	61	61	Sube	KUAFÖR	KARADENİZ	1
Erkek	Evli	OZEL_SEKTOR	Eski_Musteri	21	NULL	EV	EV-CEKILIS	150000	160	52	Sube	İŞÇİ	CUKUROVA	0
Kadın	Evli	OZEL_SEKTOR	Eski_Musteri	28	DİĞER	OTOMOTIV	OTOMOTIV-CEK	40000	37	20	Sube	DEPO SORUMLU	MARMARA 4	0
Erkek	Evli	OZEL_SEKTOR	FMT_Araması	21	İNTERNET	OTOMOTIV	OTOMOTIV-CEK	30000	49	27	Sube	MAKİNA OPERA	İÇ ANADOLU	1
Erkek	Evli	OZEL_SEKTOR	Eski_Musteri	16	DİĞER	OTOMOTIV	OTOMOTIV-CEK	35000	31	17	Sube	İŞÇİ	KARADENİZ	1
Erkek	Evli	OZEL_SEKTOR	FMT_Araması	30	DİĞER	EV	EV-CEKILIS	300000	160	52	Sube	SAĞLIK PERSON	CUKUROVA	0
Erkek	Evli	OZEL_SEKTOR	Eski_Musteri	40	DİĞER	OTOMOTIV	OTOMOTIV-CEK	35000	81	38	Sube	OPERATÖR	KONYA	0
Kadın	Evli	OZEL_SEKTOR	Eski_Musteri	54	DİĞER	EV	EV-CEKILIS	200000	100	36	Sube	PASTACI	EGE	0
Kadın	Evli	OZEL_SEKTOR	Eski_Musteri	54	DİĞER	EV	EV-CEKILIS	200000	100	36	Sube	PASTACI	EGE	0
Erkek	Evli	OZEL_SEKTOR	Gelen_Arama	36	DİĞER	OTOMOTIV	OTOMOTIV-CEK	50000	49	27	Sube	İŞÇİ	CUKUROVA	0

Çizelge 5.2'de, oluşturulan veri kümesinden örnek bir bölüm gösterilmiştir. Örnek veri kümesindeki Churn değişkeni incelendiğinde 1 değeri Churn olan müşterileri gösterirken, 0 değeri Churn olmayan müşterileri göstermektedir. Makine öğrenmesinde modelin çalışabilmesi için kategorik verilerin sayısal anlamda dönüştürülmesi gerekmektedir. Verilerin kategorik olup olmadığını anlamak için değişken tiplerini incelemek gerekmektedir. Değişkenlerin tipleri Çizelge 5.3'de gösterilmiştir.

Çizelge 5.3: Değişken tipleri

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18507 entries, 0 to 18506
Data columns (total 15 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Cinsiyet         18507 non-null  object
 1   Medeni_durum     18507 non-null  object
 2   Meslek_grubu     18507 non-null  object
 3   Gor_orjin        18507 non-null  object
 4   Gorusme_Sayi     18507 non-null  float64
 5   Gor_mecra        18507 non-null  object
 6   Kampanya_cins    18507 non-null  object
 7   Kampanya_tip     18507 non-null  object
 8   Kampanya_bedeli  18507 non-null  int64
 9   Vade             18507 non-null  int64
 10  Engecteslim_ay   18507 non-null  int64
 11  Sube_tip         18507 non-null  object
 12  Meslek_kod       18507 non-null  object
 13  Bölge            18507 non-null  object
 14  Churn            18507 non-null  int64
dtypes: float64(1), int64(4), object(10)
memory usage: 2.1+ MB
```

Veri kümesi incelendiğinde 1 float, 4 int, 10 object alandan oluşmaktadır. Veride numerik ve kategorik veriler bulunmaktadır. Kategorik verilerin sayısal forma dönüştürülmesi için en çok kullanılan One Hot Encoder dönüşümü uygulanmıştır. One Hot Encoder dönüşümü uygulanan değişkenler; Cinsiyet, Medeni_durum, Meslek_grubu, Gor_orjin, Gor_mecra, Kampanya_cins, Kampanya_tip, Sube_tip, Meslek_kod ve bölge değişkenleridir.

Sayısal verilerin dağılımları hakkında bilgi veren tanımlayıcı istatistik değerleri Çizelge 5.4'de verilmiştir.

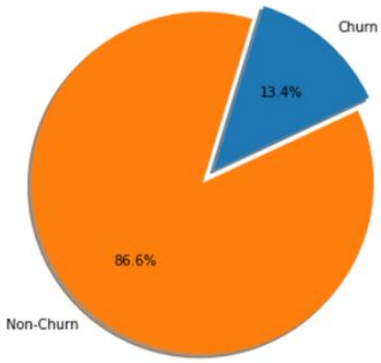
Çizelge 5.4: Veri kümesi için tanımlayıcı istatistikler

	count	mean	std	min	5%	25%	50%	75%	90%	95%	99%	max
Gorusme_Sayi	18507.0	20.836.772	13.415.528	1.0	5.0	11.0	18.0	27.0	38.0	46.0	63.0	197.0
Kampanya_bedeli	18507.0	159.905.373.102	106.419.627.721	15000.0	35000.0	60000.0	150000.0	230000.0	300000.0	350000.0	500000.0	1000000.0
Vade	18507.0	116.991.138	77.597.450	7.0	21.0	49.0	100.0	200.0	240.0	240.0	240.0	240.0
Engecteslim_ay	18507.0	43.129.519	23.073.516	1.0	11.0	27.0	38.0	63.0	73.0	73.0	96.0	122.0
Churn	18507.0	0.133625	0.340258	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0

Tanımlayıcı istatistik verileri incelendiğinde görüşme sayısı, kampanya bedeli ve vade değişkeninin ortalama etrafında dağıldığı gözlemlenirken, en geç teslim ay ve churn değişkeninin ortalamadan uzak dağıldığı gözlenmiştir.

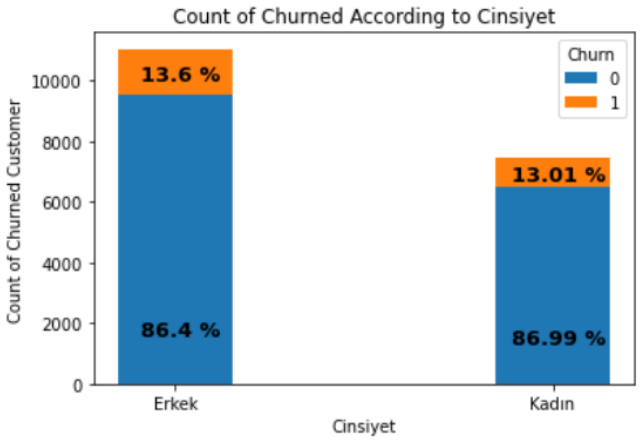
5.2 Keşifsel Veri Analizi

Verileri keşifsel veri analizi tekniği ile incelememizin amacı, müşteriyi kaybetmeye sebep olan değişkenlerin, hedef değişkenimiz ile olan ilişkilerini anlamaya çalışmaktır. Veri kümesinde toplamda 18507 adet gözlem bulunmaktadır ve bu verilerin 2473 adeti ayrılan müşterilere aittir. Müşterilen bir yıl içerisinde sistemden ayrılma oranı Şekil 5.1'de belirtildiği gibi %13.4'tür.



Şekil 5.1: Müşteri Churn oranı

Churn analizine ilk olarak veri setindeki değişkenleri inceleme ile başlanılmıştır. İlk olarak cinsiyet değişkeni incelenmiştir ve cinsiyet değişkenine ait churn oranları Şekil 5.2'de gösterilmiştir.




```
Cinsiyet Counter({'Erkek': 11049, 'Kadın': 7458})
```

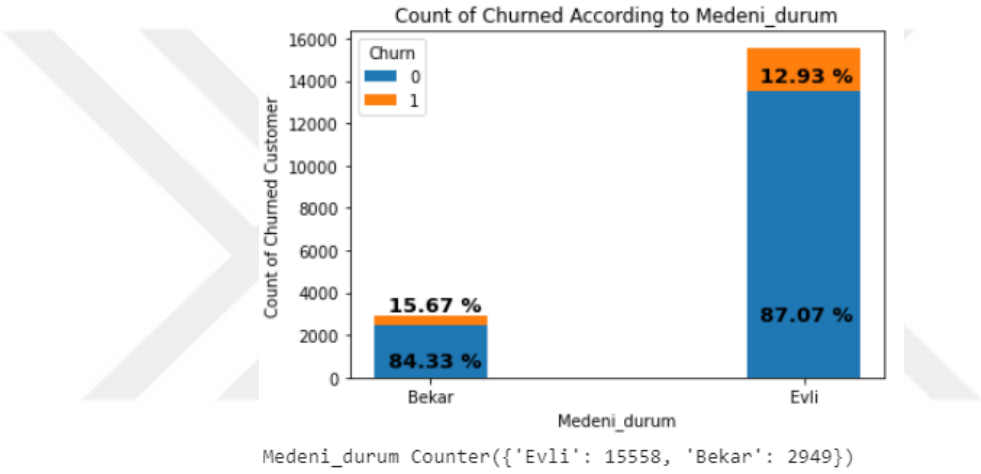
Şekil 5.2: Cinsiyet sütununda Churn grafiği

Cinsiyet, ayrılma ihtimalini tahmin etmede iyi bir değişkendir. Veri setinde toplamda 11049 erkek müşteri, 7458 kadın müşteri bulunmaktadır. Bir yıl içerisinde sistemden 1503 erkek müşteri, 970 kadın müşteri ayrılmıştır. Erkek ve kadın müşterilerin ayrılma oranlarına bakıldığında, aralarında çok az fark olsada erkek müşterilerin sistemden kadın müşterilere göre daha fazla ayrıldığı gözlemlenmiştir.

```
Analiz['Churn_Rate']=Analiz['Churn'].replace("No", 0).replace("Yes", 1)
grp=Analiz.groupby(['Churn_Rate','Cinsiyet'])["Cinsiyet"].count()
grp.head()
```

```
Churn_Rate  Cinsiyet
0           Erkek       9546
            Kadın       6488
1           Erkek       1503
            Kadın        970
Name: Cinsiyet, dtype: int64
```

'Medeni_durum' değişkenin müşteri kaybı üzerindeki etkisi Şekil 5.3'de incelenmiştir.

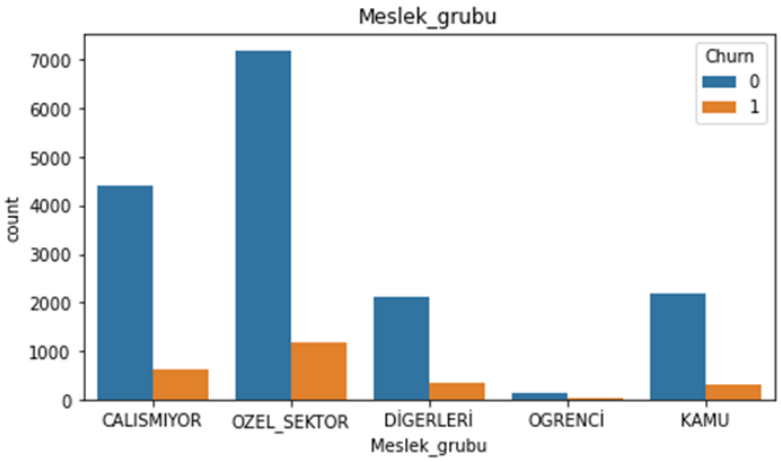


Şekil 5.3: Medeni durum sütununda Churn grafiği

Müşteriler Churn olma durumuna göre kendi içinde evli ve bekar olarak incelenmiştir. Veri setinde toplamda 15558 evli müşteri, 2949 bekar müşteri bulunmaktadır. Bekar müşterilerin sistemden ayrılma oranları %15,67 iken, evli müşterilerin sistemden ayrılma oranları %12,93'tür. Müşteri kaybı evli ve bekar olarak bakılmadan sadece medeni durum içerisinde ayrılan müşteri yüzdesi olarak incelendiğinde ise tüm veri kümesinin içerisinde medeni durumu değişkeni üzerinde Churn olanların oranı %15,50'dir. Bekar olan müşterilerimiz bu oranın üzerinde ayrılma gösterirken, evli olan müşterilerin sistemden ayrılmada genel oranın altında kaldığı gözlemlenmiştir.

```
Churn_Rate  Medeni_durum
0           Bekar           2487
            Evli           13547
1           Bekar            462
            Evli            2011
Name: Medeni_durum, dtype: int64
```

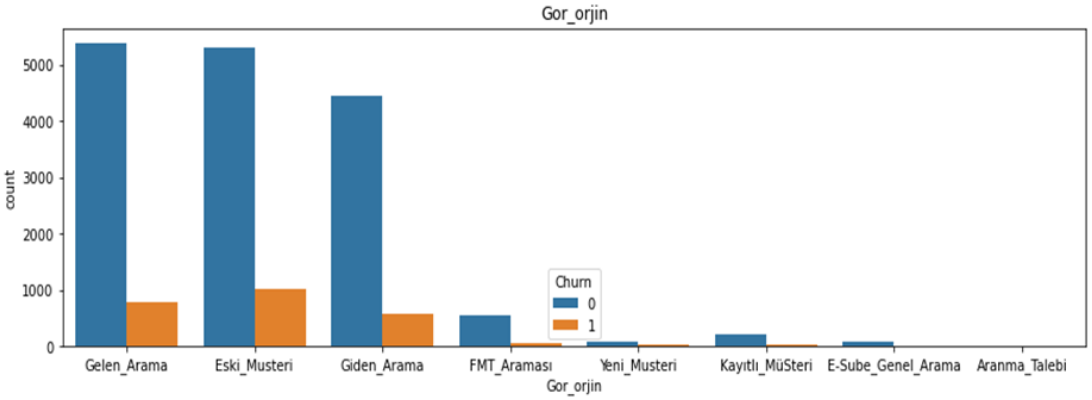
‘Meslek_grubu’ değişkenleri incelendiğinde Şekil 5.4’ de belirtilen sonuç çıkmıştır.



Şekil 5.4: Meslek grubu sütununda Churn grafiği

Özel Sektörde çalışan ve hiç çalışmayan müşterilerin Churn olması diğer meslek grubu ve kamuda görev yapan diğer müşterilere göre daha fazladır. Şekil 5.9 ve Şekil 5.10’da kampanya_cinsine göre meslek gruplarının Churn olma durumları incelenmiştir.

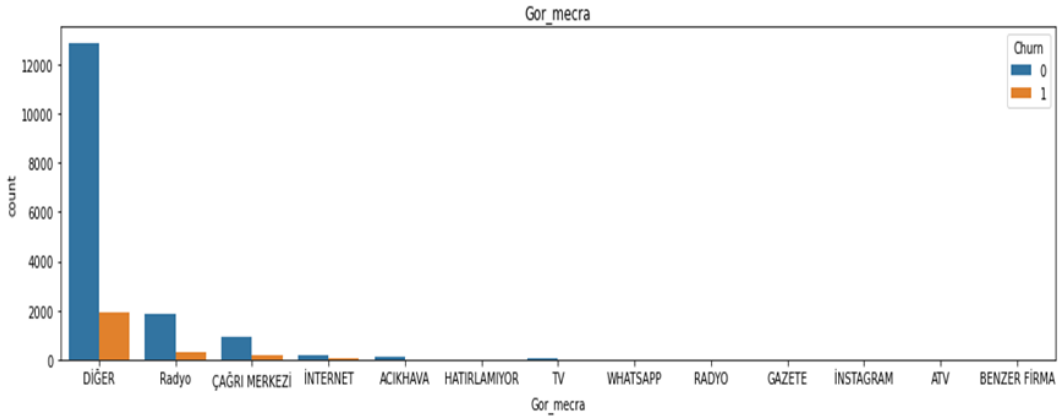
‘Gor_orjin’ değişkenlerine ait veriler incelendiğinde eski müşteriler ile gerçekleşen görüşmelerde müşteri kaybına rastlanmaktadır.



Şekil 5.5: Gor_orjin sütununda Churn grafiği

Müşteri kayıpları Şekil 5.5’de görüleceği üzere eski müşteriler ile yapılan görüşmelerin ardından, gelen arama müşterilerinde yüksek olduğu gözlenmektedir. Bu durumu takip eden giden arama ve FMT aramalarıdır.

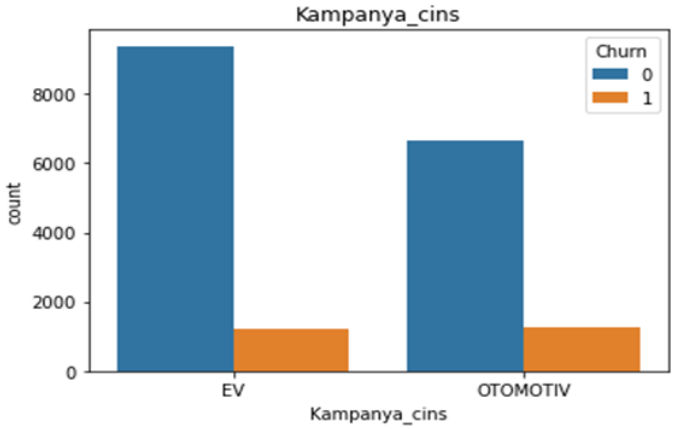
‘Gor_mecra’ değişkeni faizsiz finans sisteminde reklam oluşturabilmek adına çok önemli olsa da Şekil 5.6’da da görüldüğü üzere verinin doğru alınmayışı veya hiç alınmaması bu alanda yeterli bilgiye ulaşmamızı engellemektedir.



Şekil 5.6: Gor_mecra sütununda Churn grafiği

Müşterilerimizle ilk kontak kurulan alanlarda churn olma durumu incelendiğinde ilk diğer alanından sonra Radyo üzerinden gelen müşterilerimizde kayıp yaşandığı görülmektedir.

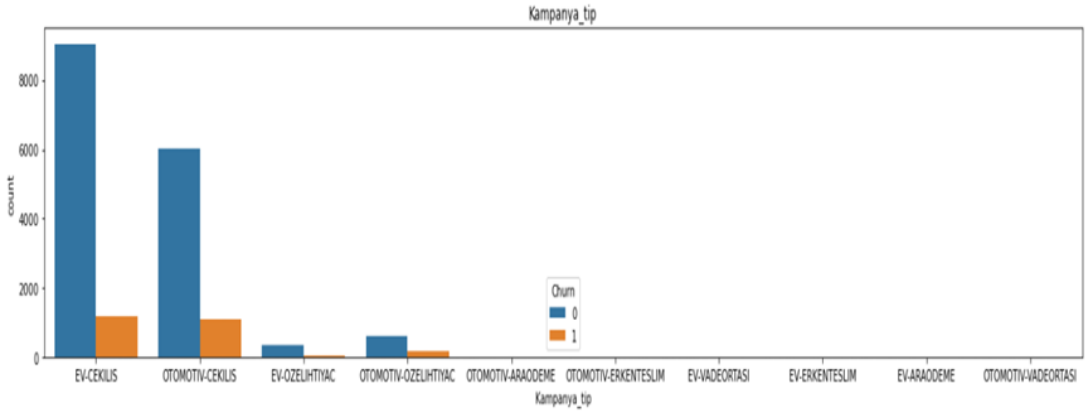
‘Kampanya_cins’ değişkeni aslında müşterilerin en çok hangi kampanyayı tercih ettiğini göstermektedir. Şekil 5.7’de görüldüğü üzere ev kampanyası daha çok tercih edilmektedir.



Şekil 5.7: Kampanya_cins sütununda Churn grafiği

Kampanya cinsine göre Churn durumunu incelediğimizde ev kampanyasına giren müşteriler otomotiv kampanyasına giren müşterilere göre çok olsa da ayrılmalar her iki kampanyada da hemen hemen eşit seviyelerdedir.

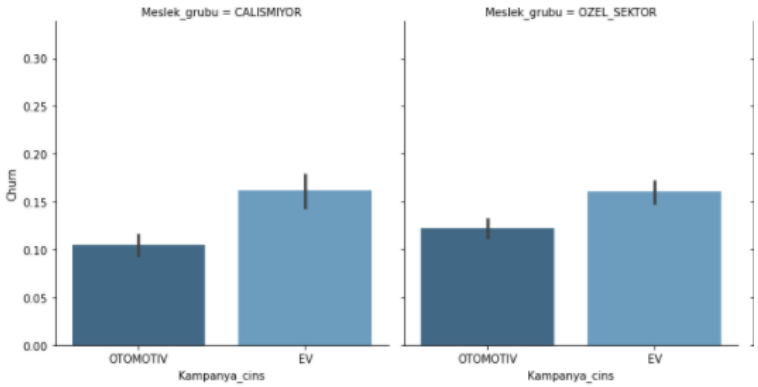
‘Kampanya_tip’ değişkenine ait Churn grafiği incelendiğinde şekil 5.8’de değinilen ev sistemini tercih eden müşteri sayısı çoğunluktadır.



Şekil 5.8: Kampanya_ tip sütununda Churn grafiği

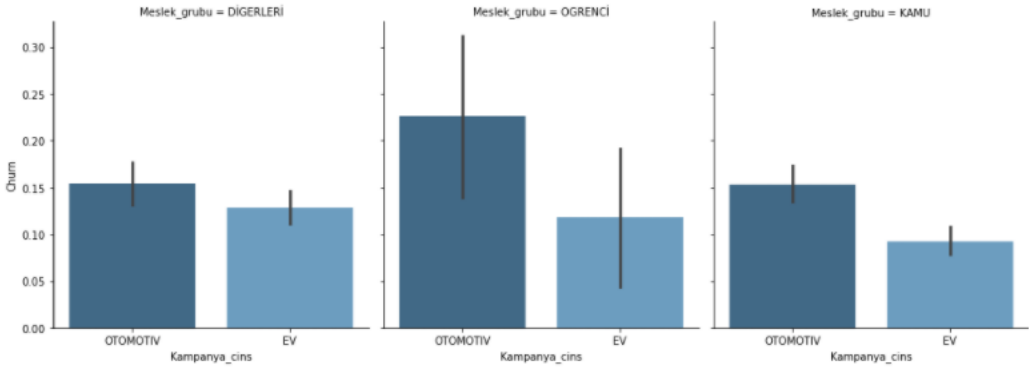
Kampanya tipine göre inceleme yapıldığında ev-çekiliş sistemine giren müşterilerin Churn olması diğer sistemlere göre daha fazladır. Ev-çekiliş grubundan sonra müşterilerin en çok ayrıldığı sistem otomotiv- çekiliş sistemi olarak görülmektedir.

‘Meslek_grubu’ değişkeni incelendiğinde meslek grubunun çalışmayan müşteriler, özel sektörde çalışan müşteriler, öğrenci olan müşteriler, kamuda çalışan müşteriler ve diğer olarak gruplandığı görülmektedir.



Şekil 5.9: Meslek_grubu sütununda çalışmayan ve özel sektörde çalışan müşterilere ait Churn grafiği

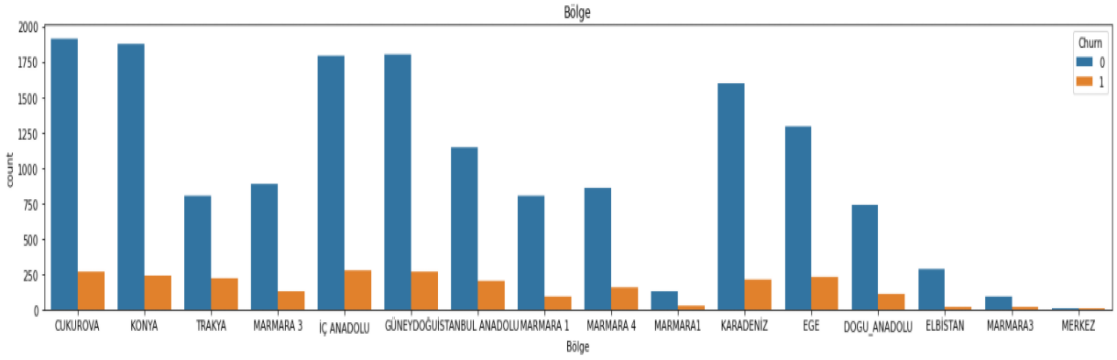
Şekil 5.9 grafiğinde çalışmayan ve özel sektörde çalışan müşteriler gözlenmektedir. Çalışmayan ve özel sektörde çalışan müşterilerin en çok ev sisteminden ayrıldığı gözlenmiştir.



Şekil 5.10: Meslek_grubu sütununda öğrenci, kamu personeli ve diğer meslek grubundaki müşterilere ait Churn grafiği

Şekil 5.10 incelendiğinde ise öğrenci olan müşterilerin, kamu personeli olan müşterilerin ve diğer sektörde görev yapan müşterilerin otomotiv sisteminden daha fazla ayrıldığı gözlenmiştir.

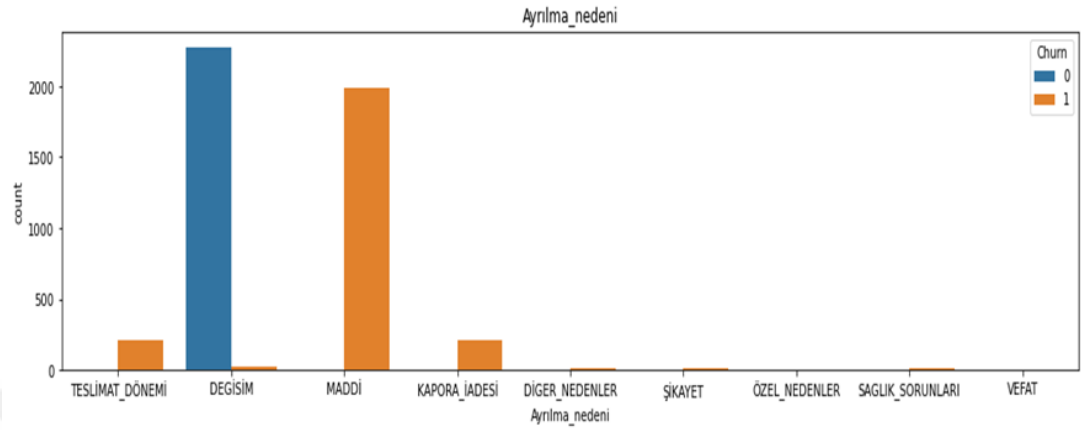
'Bölge' değişkeninin ayrılan müşterileri incelemede tercih edilmesinin sebebi, şube bazlı müşteri ayrılmalarına göre daha genel sonuç vermesidir.



Şekil 5.11: Bölge sütununda Churn grafiği

Müşteri kayıplarının bölge değişkeni içerisindeki dağılımı Şekil 5.11'de görüldüğü üzere en çok İç Anadolu ve Güneydoğu Bölgesi'nde yaşanmaktadır. Müşteri kaybında bu bölgeleri takip eden bölgeler Çukurova ve Ege Bölgeleri'dir.

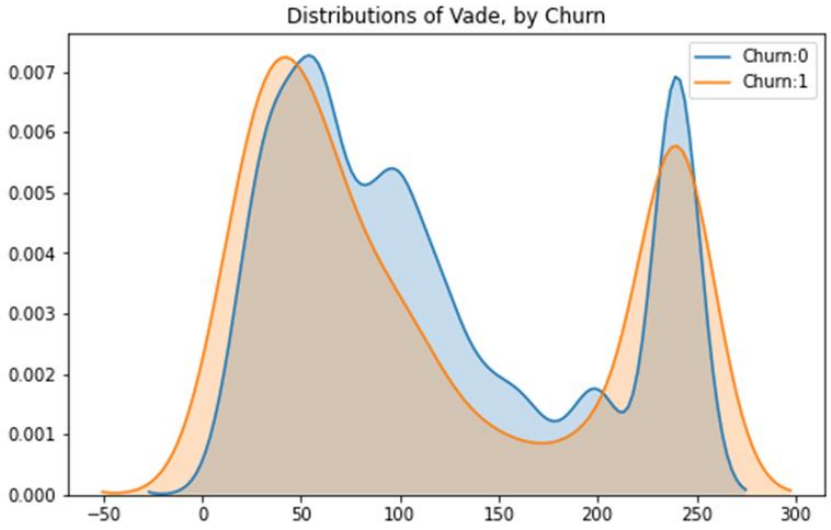
‘Ayrılma_nedeni’ değişkeni incelendiğinde Şekil 5.12’de gözlenen sonuçlar elde edilmiştir.



Şekil 5.12: Ayrılma nedeni sütununda Churn grafiği

Müşterilerin ayrılma nedeni incelendiğinde büyük çoğunluğunun maddi sıkıntıdan kaynaklı sistemden ayrıldıkları gözlenmektedir. Ayrılma nedenleri arasındaki teslimat nedeni ayrılmaları ise teslimatını hak eden müşterilerin evrak toplama sürecindeki sıkıntılardan veya istediği ev yada otomobili bulamadığından kaynaklanmaktadır. Kapora iadesindeki ayrılışların sebebi, sisteme girmeyi düşünüp belli bir kapora ödeyen müşterilerin sisteme girmekten vazgeçmeleri neticesinde sistemden ayrılırken firma tarafından müşteriye yapılan kaporanın iadesidir. Şekil 5.12’de belirtilen diğer nedenler, şikayet, özel nedenler, sağlık sorunları ve vefat seçenekleri müşterinin ayrılmasına neden olan etkenler olsa da diğer seçeneklere göre daha az etkilidir.

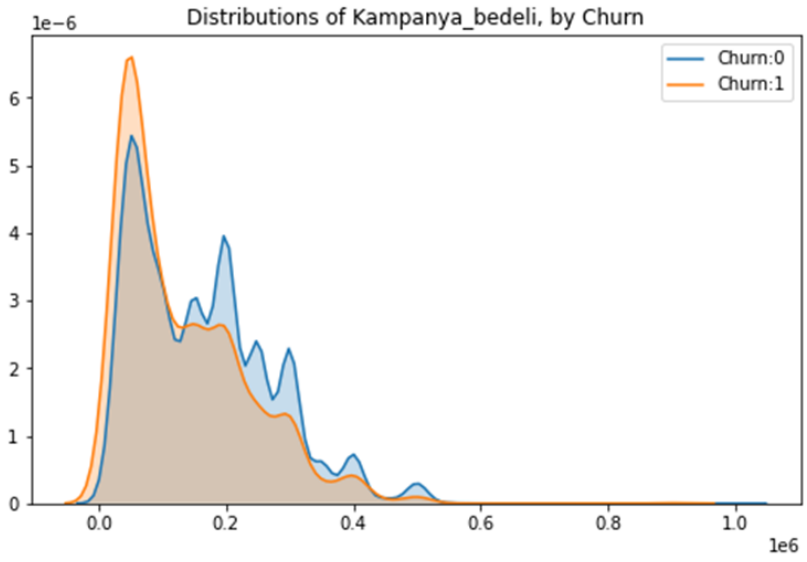
‘Vade’ değişkenine göre müşterinin sistemde kalma süresini gösteren Churn grafiği Şekil 5.13’de verildiği gibidir.



Şekil 5.13: Vade sütununda Churn grafiği

Vade değişkenine göre veri kümesi incelendiğinde 100 ay vadeye kadar Churn olan ve Churn olmayan müşterilerin hemen hemen birbirine yakın olduğu gözlenmektedir. Teslimatını almış ve firmanın tahsilatını gerçekleştirdiği müşterilerde müşteri kaybı azalmaktadır.

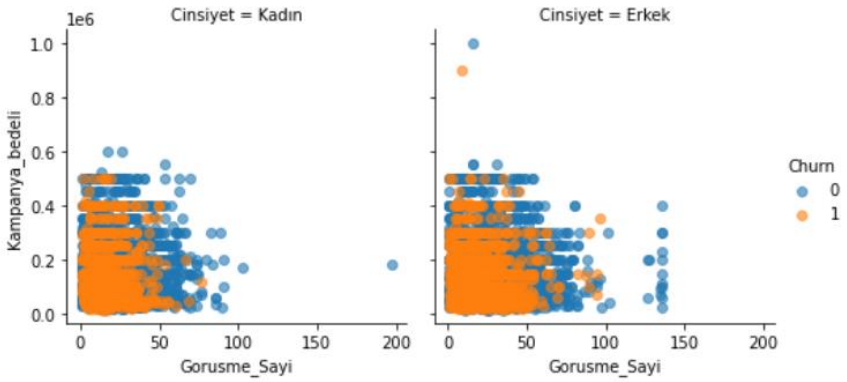
‘Kampanya_bedeli’ değişkeni Şekil 5.14’de incelenmiş olup, müşterilerin sisteme kayıt oldukları bedel üzerinden ayrılma durumları gösterilmiştir.



Şekil 5.14: Kampanya bedeli sütununda Churn grafiği

Kampanya bedeli düşük olan müşterilerde ayrılma oranı yüksek iken, kampanya bedelindeki yükseliş arttıkça müşterilerin ayrılma oranıda düşmektedir.

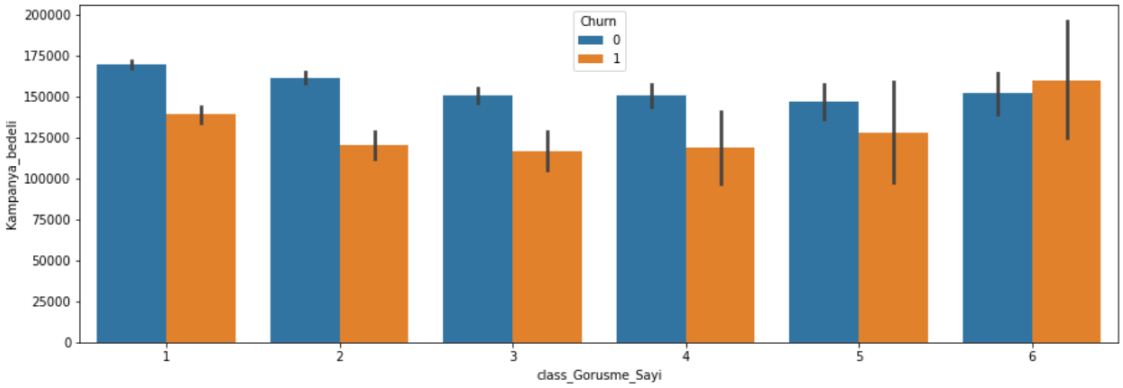
'Kampanya_bedeli, Cinsiyet, Görüşme_Sayıları' değişkenlerinin saçılım grafiği üzerindeki Churn durumu Şekil 5.15'de verildiği gibidir.



Şekil 5.15: Kampanya bedeli, Cinsiyet, Görüşme Sayılarına göre saçılım Churn grafiği

Saçılım grafiği incelendiğinde kampanya bedelinin dağılımında cinsiyet değişkenine göre farklılık gözlenmemektedir, ancak; müşterilerle yapılan görüşme sayılarına cinsiyet değişkeni üzerinden bakıldığında erkek müşteri ile gerçekleştirilen görüşmelerin kadın müşterilere göre daha fazla olduğu ve buna rağmen erkek müşterilerde ayrılmaların daha fazla yaşandığı gözlenmektedir.

'Görüşme_Sayıları, Kampanya_bedeli' değişkenleri gruplandırılarak müşteri ayrılma durumları Şekil 5.16' da incelenmiştir.



Şekil 5.16: Görüşme Sayıları, Kampanya bedeli gruplanmalarına göre Churn grafiği

Görüşme sayıları değişken ortalaması baz alınarak sınıflara bölündü ve sınıf aralıkları içerisinde müşteri kayıp durumları incelendi.

```
def class_Gorusme_Sayi(Analiz):
    if Analiz['Gorusme_Sayi']<=20 :
        return 1
    elif (Analiz['Gorusme_Sayi']>20) & (Analiz['Gorusme_Sayi']<=30):
        return 2
    elif (Analiz['Gorusme_Sayi']>30) & (Analiz['Gorusme_Sayi']<=40):
        return 3
    elif (Analiz['Gorusme_Sayi']>40) & (Analiz['Gorusme_Sayi']<=50):
        return 4
    elif (Analiz['Gorusme_Sayi']>50) & (Analiz['Gorusme_Sayi']<=60):
        return 5
    elif Analiz['Gorusme_Sayi']>60:
        return 6
Analiz['class_Gorusme_Sayi']=Analiz.apply(lambda Analiz: class_Gorusme_Sayi(Analiz),axis=1)
```

Yapılan sınıflandırmada:

1. grup müşteri ile yapılan görüşme sayısının 20 ve daha az olduğu sınıfı göstermektedir.
2. grup müşteri ile yapılan görüşmenin 20 ile 30 arasında olan sınıfı göstermektedir.
3. grup müşteri ile yapılan görüşmenin 30 ile 40 arasında olan sınıfı göstermektedir.
4. grup müşteri ile yapılan görüşmenin 40 ile 50 arasında olan sınıfı göstermektedir.
5. grup müşteri ile yapılan görüşmenin 50 ile 60 arasında olan sınıfı göstermektedir.
6. grup ise 60 da daha fazla görüşme yapılan müşterileri ifade etmektedir.

Gruplar incelendiğinde 6. grup müşterilerde yani en fazla aranan ve sistemin anlatıldığı müşterilerde daha fazla müşteri kaybı yaşandığı gözlenmektedir.

5.3 Modellerin Kurulması

Bu çalışmadaki amaç en iyi tahmini yapan modeli oluşturmaktır. Öncelikle RF ile değişkenlerin modele katkıları, önem düzeyleri incelenmiş, sınıf bilgisi ile doğrudan ilişkili ve aşırı öğrenmeye neden olabilecek ayrılma nedeni değişkeni veri kümesinden çıkarılmıştır. Veri kümesi ile gerçekleştirilen analizlerde churn olan müşterilere ait recall oranının sıfıra çok yakın olduğu tespit edilmiştir. Recall oranından elde edilen bilgiye dayanarak modellerin churn olan müşterileri modelleyemediği tespit edilmiştir.

Churn olmayan müşterilere ait sınıf, rastgele azaltarak her iki sınıf sayıları eşit hale getirilmiştir. Veri kümesi 18507'den 4946 veriye düşürülmüştür. Modeli oluştururken kullanılan veri kümesi, test ve eğitim veri kümesi olarak %75-%25 oranında ayrılarak sınıflandırılmıştır.

Öncelikli olarak veri kümesinde var olan kategorik değişkenleri "One Hot Encoder" yöntemi kullanarak 1 ve 0 değeri atandı.

```
Analiz.Churn=Analiz.Churn.replace({1, 0})
Analiz=Analiz.drop(columns=['Churn_Rate'],axis=1)
```

```
Analiz2=pd.concat([Analiz,pd.get_dummies(Analiz[categorical])],axis=1).drop(Analiz[categorical],axis=1)
```

Dönüşümü yapılan veri kümesinin sınıf dengesizliği olup olmadığının kontrolü gerçekleştirilmiştir.

```
print("Percentage of Churned(1)     Customer:%",round(Analiz.Churn.value_counts()[1]/Analiz.shape[0]*100,2))
print("Percentage of Not Churned(0)  Customer:%",round(Analiz.Churn.value_counts()[0]/Analiz.shape[0]*100,2))
```

```
Percentage of Churned(1)     Customer:% 50.0
Percentage of Not Churned(0)  Customer:% 50.0
```

Elde edilen oranlar doğrultusunda hedef değişkenin dengesiz olmadığı tespit edilmiştir.

5.3.1 Standart Ölçek (Standart Scale)

Genel olarak, veri kümelerinin çoğu çok farklı ölçeklere sahiptir ve bazen çok büyük aykırı değerler içermektedir. Bu iki özellik, verilerin görselleştirilmesinde zorluklara yol açmakta ve daha da önemlisi, birçok makine öğrenimi algoritmasının tahmin performansını düşürebilmektedir. Bu durumu göz önünde bulundurarak ve oluşacak modellerde en iyi skoru yakalayabilmek için veri kümesine standart scale uygulanmıştır.

```
from sklearn.preprocessing import StandardScaler
scaler=StandardScaler()
X_scl=scaler.fit_transform(X)
```

5.3.2 Grid Search ile Lojistik Regresyon

Lojistik regresyonun amacı, sınıflandırma problemi için bağımlı ve bağımsız değişkenler arasındaki ilişkiyi tanımlayan doğrusal bir model kurmaktır. Grid Search belirlenen aralıkta en doğru tahminler ile sonuçlanan bir modelin optimal hiper parametrelerini bulmak için kullanılmaktadır. Hiper paremetreleri ayarlamak için birkaç strateji vardır. Bunlardan ikisi rastgele arama ve grid aramadır. Grid arama ile her hiper parametre için önce model kurulmuştur sonra hiper paremetre seçenekleri oluşturulmuştur.

```
from sklearn.linear_model import LogisticRegression
logistic=LogisticRegression()
parameters = {"C": [10 ** x for x in range (-5, 5, 1)],
              "penalty": ['l1', 'l2'],'solver': ('linear', 'lbfgs', 'liblinear')}
```

```
from sklearn.model_selection import GridSearchCV
grid_cv = GridSearchCV(estimator=logistic,
                       param_grid = parameters,
                       cv = 10)
grid_cv.fit(X, y)
print("The best parameters : ", grid_cv.best_params_)
print("The best score      : ", grid_cv.best_score_)
```

```
The best parameters :  {'C': 0.1, 'penalty': 'l1', 'solver': 'liblinear'}
The best score      :  0.7223003312476997
```

Müşteri kayıplarının modellenmesinde en iyi model ve en iyi skor çalışmanın önceliğidir. Churn olmayan müşterilere ait sınıf, rastgele azaltılarak churn olan ve churn olmayan her iki sınıfında sayıları eşit hale getirilip veri kümesi 18507'den 4946 veriye düşürülmüştür. Dengeli hale getirilen, test ve eğitim olarak ayrılan veri kümesi modelinin Lojistik regresyon analizi ile doğruluk ve kesinlik performansı Çizelge 5.5'de görüldüğü gibidir.

Çizelge 5.5: Grid Search ile elde edilen modelin performansı

	Accuracy_Test	Accuracy_Train	F1 Score_Test	F1 Score_Train	Precision_Test
Logistic_Regr.	0.751011	0.802642	0.706667	0.775735	0.858796

Çizelge 5.5' de görüldüğü gibi Lojistik regresyon ile modeli tahmin etmede elde edilen performans değeri 0.75'dir. Daha iyi sonuç alınıp alınamayacağı değerlendirilmek üzere çapraz doğrulama, SMOTE ve ADASYN algoritmaları ile lojistik regresyon analizi yapılmıştır.

5.3.3 Çapraz Doğrulama (Cross Validation) ile Lojistik Regresyon

Çapraz doğrulama (CV), bir makine öğrenimi modelinin etkinliğini test etmek için kullanılan tekniklerden biridir. Genelleştirilmiş performansı değerlendirmede çok dayanıklı bir yöntemdir. Modeli değerlendirmek için kullanılan örnekleme prosedürüdür. Modelin uygulanması sonrası, test ve eğitim veri kümesindeki değişkenlerin birbiriyle olan ilişkileri ve ortalaması aşağıdaki gibidir.

```
lrm = LogisticRegression(solver='liblinear',penalty='l1',C=0.1)
cv = cross_validate(estimator=lrm,
                    X=X_scl,
                    y=y,
                    cv=10,return_train_score=True
                   )
print('Test Scores            : ', cv['test_score'], sep = '\n')
print("-"*50)
print('Train Scores           : ', cv['train_score'], sep = '\n')
```

```
Test Scores            :
[0.52121212 0.57373737 0.56969697 0.55353535 0.69090909 0.92121212
 0.90890688 0.89068826 0.67206478 0.8805668 ]
--------------------------------------------------
Train Scores           :
[0.82453381 0.82228713 0.82498315 0.81824309 0.8038643  0.78072343
 0.78301887 0.78481581 0.78481581 0.78908356]
```

```
print('Mean of Test Sets : ', cv['test_score'].mean())
print('Mean of Train Sets : ', cv['train_score'].mean())
```

```
Mean of Test Sets :  0.7182529750950803
Mean of Train Sets :  0.8016368965613994
```

Puanlamanın varsayılan değeri, test ve eğitim veri kümesinin doğruluk puanını hesaplamaktadır. Test ve eğitim veri kümesindeki puanlamalar incelendiğinde değişkenlerin birbirleri ile olan ilişkilerinin ortalamalarının test veri seti için 0.71, eğitim veri seti için 0.80 olduğu gözlenmiştir.

5.3.4 SMOTE ile Lojistik Regresyon

Lojistik regresyon analizinde elde edilen performans puanından daha iyi sonuç elde edebilmek adına SMOTE algoritması ile yeniden performans değerlendirilmesine gidilmiştir. 18507 müşteriden oluşan veri kümesinin 2473'ü churn olan müşterileri, geri kalan 16034 müşteri ise churn olmayan müşterileri temsil etmektedir. Doğru değerlendirme yapabilmek adına churn olan müşteri sayısı sentetik veri üretme yöntemi ile churn olmayan müşteri sayısı ile eşit hale getirilmiştir. 32068 müşteriden oluşan veri kümesine SMOTE algoritması ile Lojistik regresyon analizi yapılmıştır.

Ancak; sentetik veri ile arttırılan veri kümesinden elde edilen performans puanı Çizelge 5.6'da belirtildiği üzere 0.64 olarak elde edilmiştir.

Çizelge 5.6: SMOTE ile elde edilen modelin performansı

	AUC Score	Accuracy_Test	Accuracy_Train	F1 Score_Test	F1 Score_Train	Precision_Test	Precision_Train	Recall_Test	Recall_Train
Logistic_Regr.	0.629707	0.866436	0.867507	0.015924	0.024403	0.500000	0.766667	0.008091	0.012399
Resampled_Logistic	0.664711	0.615941	0.620141	0.624191	0.625297	0.610992	0.616947	0.637974	0.633877
L.Regression_with_SMOTE	0.641007	0.866436	0.866427	0.003226	0.004296	0.500000	0.571429	0.001618	0.002156

Churn olan veri kümesinin sentetik veri ile arttırılarak churn olmayan veri kümesi ile eşit hale getirilmesi SMOTE analizinde istenilen performansı vermemiştir. SMOTE analizinden daha iyi bir sonuç alınabilmesi için churn olan veri kümesine göre dengeleme yapılması gerekmektedir. Dengeleme için churn olan sınıfı, churn olmayan sınıfa yaklaştırmak için rastgele sıralama yapılmıştır ve her iki sınıf için 2473 gözlem toplamda 4946 gözlem ile uygulama tekrarlanmıştır.

Çizelge 5.7: Dengelenmiş veride SMOTE ile elde edilen modelin performansı

	AUC Score	Accuracy_Test	Accuracy_Train	F1 Score_Test	F1 Score_Train	Precision_Test	Precision_Train	Recall_Test	Recall_Train
Logistic_Regr.	0.841492	0.751011	0.802642	0.706667	0.775735	0.858796	0.898510	0.600324	0.682480
Resampled_Logistic	0.854744	0.788197	0.782960	0.752830	0.747570	0.902715	0.893553	0.645631	0.642588
L.Regression_with_SMOTE	0.851574	0.762328	0.791049	0.709486	0.755443	0.911168	0.910959	0.580906	0.645283

Çizelge 5.7'de belirtildiği üzere dengeli hale getirilen veri kümesinde SMOTE algoritması ile elde edilen performans puanı 0.85'dir. Lojistik regresyondan elde edilen puandan daha yükek olduğu görülmektedir.

5.3.5 ADASYN ile Lojistik Regresyon

ADASYN ile üretilen örnekler, SMOTE algoritmasının veri ile en yakın komşusu arasında oluşturduğu doğrunun çok yakınında üretilmekte ve böylece daha gerçekçi olması sağlanmaktadır. Sentetik veri üretme yöntemi ile churn olan müşteri sayısı churn olmayan müşteri sayısı ile eşit hale getirilmiştir. 32068 müşteriden oluşan veri kümesi ile ADASYN algoritmasından elde edilen performans puanı Çizelge 5.8'de belirtildiği üzere 0.64 olarak elde edilmiştir.

Çizelge 5.8: ADASYN ile elde edilen modellerin performansı

	AUC Score	Accuracy_Test	Accuracy_Train	F1 Score_Test	F1 Score_Train	Precision_Test	Precision_Train	Recall_Test	Recall_Train
Logistic_Regr.	0.629703	0.866436	0.867507	0.015924	0.024403	0.500000	0.766667	0.008091	0.012399
Resampled_Logistic	0.664708	0.616690	0.620847	0.625107	0.626347	0.611602	0.617417	0.639222	0.635540
L.Regression_with_SMOTE	0.641603	0.866436	0.866210	0.003226	0.001076	0.500000	0.250000	0.001618	0.000539
L.Regression_with_ADASYN	0.641156	0.866220	0.866427	0.000000	0.004296	0.000000	0.571429	0.000000	0.002156

Churn olan veri kümesinin sentetik veri ile arttırılarak churn olmayan veri kümesi ile eşit hale getirilmesi ADASYN analizinde istenilen performansı vermemiştir. ADASYN analizinden daha iyi bir sonuç alınabilmesi için churn olan veri kümesine göre dengeleme yapılması gerekmektedir. Dengeleme için churn olan sınıfı, churn olmayan sınıfa yaklaştırmak için rastgele sıralama yapılmıştır ve her iki sınıf için 2473 gözlem toplamda 4946 gözlem ile uygulama tekrarlanmıştır.

Çizelge 5.9: Dengelenmiş veride ADASYN ile elde edilen modellerin performansı

	AUC Score	Accuracy_Test	Accuracy_Train	F1 Score_Test	F1 Score_Train	Precision_Test	Precision_Train	Recall_Test	Recall_Train
Logistic_Regr.	0.841492	0.751011	0.802642	0.706667	0.775735	0.858796	0.898510	0.600324	0.682480
Resampled_Logistic	0.854744	0.788197	0.782960	0.752830	0.747570	0.902715	0.893553	0.645631	0.642588
L.Regression_with_SMOTE	0.851574	0.762328	0.791049	0.709486	0.755443	0.911168	0.910959	0.580906	0.645283
L.Regression_with_ADASYN	0.851747	0.761520	0.790779	0.708786	0.755051	0.908861	0.910891	0.580906	0.644744

Çizelge 5.9'da ADASYN ile kurulan modelin performansı incelendiğinde SMOTE ile kurulan modelle hemen hemen aynı performasn sonuçlarına ulaşıldığı gözlenmiştir. Lojistik regresyon analizi yukarıda incelenen algoritmalar ile gerçekleştirilmiştir ve en iyi AUC skorunu verenin ADASYN algoritması ile lojistik regresyon analizi olduğu tespit edilmiştir.

5.4 K En Yakın Komşu Algoritması

Müşteri kayıp analizinde daha iyi sonuç veren modeli bulmak adına dengeli hale getirilen veri kümesi ile KNN algoritmasıyla yeni bir veri noktası için tahmin yapılmıştır ve yeni verinin eğitim kümesi içerisindeki en yakın komşusu ya da komşuları bulunmuştur. Bu komşuların sınıflarına göre bir sınıflandırma yapılmıştır.

```
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                     metric_params=None, n_jobs=None, n_neighbors=5, p=2,
                     weights='uniform')
```

KNN ile oluşturulan model sonucunda verinin AUC skoru Çizelge 5.10'da gösterildiği üzere 0.72'dir.

Çizelge 5.10: KNN ile elde edilen modelin performansı

Precision Score	0.6860902255639098
Recall Score	0.5906148867313916
Accuracy Score	0.6604688763136621
F1 Score	0.6347826086956522
AUC Score	0.7296153102143033

Model performansı tatmin edici olmadığından GridSearchCV ile optimum k-komşuları bulmaya çalışılmıştır. 24 tane komşu kullanılmış ve elde edilen en iyi puan 0.64'dür.

```
The best parameters: KNeighborsClassifier(n_neighbors=24)
The best score: 0.6499129325974644
```

5.5 Destek Vektör Makinesi (SVM)

Destek Vektör Makinesi ile dengeli hale getirilen veri kümesi test ve eğitim olarak ayrıldıktan sonra kernelin linear algoritması ile modellendirilmiştir. Modelin doğru sınıflandırılması 0.75 olarak elde edilmiştir.

```
svm_model= SVC(kernel = "linear").fit(X_train,y_train)
svm_model
```

```
SVC(kernel='linear')
```

```
accuracy_score(y_test,y_pred)
```

```
0.7550525464834277
```

Test için ayrılan %25'lik veri kümesinin yani 1237 verinin hem gerçekte olan hem de tahmin edilen değerleri karşılaştırılarak Çizelge 5.11'deki hata matrisi elde edilmiştir.

Çizelge 5.11: Hata Matrisi

	Gerçek Kayıp Müşteriler	**Gerçek Sadık Müşteriler**
Tahmin edilen Kayıp müşteriler	570	49
Tahmin edilen Sadık müşteriler	254	364

Çizelge 5.11'deki hata matrisi incelendiğinde churn olan doğru tahminin 934, churn olmayan tahminin ise 303 olduğu gözlenmiştir.

Çizelge 5.12: SVM ile test veri kümesinin tahmin değerleri

	precision	recall	f1-score	support
0	0.69	0.92	0.79	619
1	0.88	0.59	0.71	618
accuracy			0.76	1237
macro avg	0.79	0.75	0.75	1237
weighted avg	0.79	0.76	0.75	1237

Churn olan modelin doğruluğunun tahmin değerleri Çizelge 5.12'de verilmiştir. Tahmin değerinin kesinliği 0.88'dir.

Destek vektör kümesi ile parametre aralığı tanımlanarak müşteri kayıplarının tahminini en iyi şekilde verecek model oluşturuldu.

```
Fitting 10 folds for each of 2 candidates, totalling 20 fits
GridSearchCV(cv=10, estimator=SVC(), n_jobs=-1, param_grid={'C': [1, 10]},
             verbose=2)
```

```
print("En iyi parametreler:"+str(svc_cv_model.best_params_))
```

```
En iyi parametreler:{'C': 10}
```

Çizelge 5.13: SVM ile elde edilen modelin performansı

Doğruluk Oranı / Accuracy Score	0.7550525464834277
Precision Score	0.8813559322033898
Recall Score	0.5889967637540453

Çizelge 5.13'de incelendiğinde modelin performansı 0.75'dır.

Çizelge 5.14: ROC/ AUC performans skoru

Model	ROC/AUC
Lojisitk Regresyon	0.751011
Lojistik Regresyon ADASYN	0.851747
Lojistik Regresyon SMOTE	0.851574
SVM (Destek Vektör Kümesi)	0.755052
KNN (En Yakın Komşu)	0.729615

Çizelge 5.14'de belirtildiği üzere elde edilen performans değerleri yorumlanacak olursa, doğruluk oranı iyi dengelenmiş ve çarpık olmayan veya sınıf dengesizliği olmayan sınıflandırma problemleri için geçerli bir değerlendirme biçimidir. F1 puanı,

yani bir tür sınıf örneğinin diğer tür sınıf örneklerinden daha fazla olduğu durumda kullanılır. Bir sınıflandırma problemi olduğunda AUC/ROC eğrisine güvenilebilir. Sınıflandırma modelinin performansını kontrol etmek için en önemli değerlendirme ölçütlerindendir. Modelin sınıflar arasında ne kadar ayrım yapabildiğini gösterir. Yukarıda belirtilen çizelgeler incelendiğinde en yüksek AUC/ROC skorunu veren ADASYN algoritması ile Lojistik regresyon modelidir.

6.SONUÇ

Bu çalışmada, faizsiz finans sistemi Churn davranışının tahmin edilmesi için mevcuttaki performansı en yüksek sınıflandırma algoritmaları ile karşılaştırılmıştır. Algoritmalar, faizsiz finans kuruluşundan alınan 2020 yılına ait bir yıllık veri üzerinden uygulanmıştır. Çalışılan verinin büyüklüğünün böyle bir analize uygun olduğu test sonucunda ortaya konulmuştur. Bu çalışmada, açıklayıcı değişkenlerin çok önemi vardır. Çünkü faizsiz finans firmaları için amaç sadece Churn olacak müşteriyi tahmin etmek değil aynı zamanda müşteri kayıplarının nedenlerini anlamaktır. Müşteri kayıplarını azaltıcı sistemsel değişikliklere gitmek ve müşteriye uygun önlemler alabilmektir. Bu açıdan bakıldığında çalışmada öne çıkan bazı açıklayıcı değişkenlerin sektöre bu amaçlar yolunda da katkı sağlayacağı düşünülüyor. Daha önce de belirtildiği gibi en optimal parametreler ile çalıştırılan algoritmalar kıyaslandığında en başarılı sonuca ADASYN algoritması ile Lojistik regresyon modelinin ulaştığı gözlenmiştir.

Modelin sonuçlarına göre Churn edecek müşteriyi ayrıştırmada en önemli değişkenler kampanya bedelinin büyüklüğü, ayrılma nedeni ve teslimatın gerçekleşeceği en geç teslim ayının büyüklüğüdür. Kampanya bedeli yüksek olan müşterilerin sistemden ayrılmadığı ve devam ettiği gözlemlenirken, kampanya bedeli düşük olan müşterilerin sistemden daha kolay ayrılabildiği gözlenmiştir. Müşterinin bu davranışı firma için risk ifade edebileceği için müşteri kampanya bedeli üzerinden skorlamak ve önlem almak riski azaltıcı faktör olacaktır. Kampanya bedeli kadar etkisi olan diğer bir faktör müşterilerin ayrılma nedenleridir. Türkiye ekonomik şartları göz önünde bulundurulduğunda maddi yönden ayrılan müşterilerin yoğun olduğu gözlenmiştir. Ayrılma nedenlerinin sebepleri arasında teslimat dönemi de önemli bir yer taşımaktadır. Teslimat döneminde yaşanan kefil sorunları veya evrak süreçleri müşteri memnuniyeti ve devamlılığı açısından dikkat edilmesi gereken alanlardandır. Müşterilerin sistemden ayrılmasında önemli bir paya sahiptir.

Müşteriler ile yapılan görüşmeler sistem açısından önem taşıyan diğer faktörlerden biridir. Görüşmeler ve müşteri ziyaretleri müşteriye sistemi anlatmada, merak uyandırmada çok önemlidir. Görüşme yapılan müşteriler incelendiğinde en çok görüşmelerin bay müşteriler ile yapıldığı ancak yine en çok müşteri kaybının da bay müşterilerde olduğu gözlenmiştir. Faizsiz finans sisteminin hangi meslek grubuna

hitap ettiği, yeni kampanya oluşturulmasında veya reklam kampanyası hazırlanmasında çok önemli bir yere sahiptir. Meslek gurubuna göre inceleme yapıldığında özel sektör çalışanlarının sistemi diğer meslek gruplarına göre daha fazla tercih ettiği ve yine özel sektör grubunda yer alan müşterilerin sistemden daha çok ayrıldığı gözlenmiştir. Meslek grubu alt kırılımlarına bakıldığında öğrencilerin faizsiz finans sisteminde otomotiv sistemine girdikleri gözlenmiştir. Otomotiv sisteminden ayrılan müşteriler incelendiğinde yine öğrencilerin sistemden en çok ayrıldığı tespit edilmiştir.

Özetle çalışma sonucunda Lojistik regresyon algoritmasının faizsiz finans sisteminde Churn davranışını tahmin etmede oldukça başarılı olduğu görülmektedir. Yapılan araştırmalar ve incelemelere göre Churn davranışı sektördeki diğer firmalarla ilgili daha fazla bilgi ile desteklenmediği sürece tam anlamıyla açıklanabilmiş değildir. Bu çalışma tek bir faizsiz finans sisteminin bilgileri kullanılarak gerçekleştirilmiştir. Diğer firmalar ile ilgili bilgiler çalışmalara dahil edilerek geliştirilebilir. Bir başka gelişim alanı da müşteri kaybının zaman içerisinde gelişim göstermesinden kaynaklanan sorunlara odaklanmaktır. Söz konusu çalışmanın konusu olan modeller zaman eksenindeki büyümeyi göz ardı etmiş modeller olduğu için müşteri kaybındaki (Churn) davranışın zaman içerisindeki gelişimini açıklayamamaktadır. Bunu takip eden yeni çalışma ve araştırmalar bu açıdan da düşünülerek geliştirilmelidir.

KAYNAKÇA

Abbasimehr, H. Setak, M. ve Tarokh, M.(2014) A Comparative Assessment of the Performance of Ensemble Learning in Customer Churn Prediction, The International Arab Journal of Information Technology, vol. 11, no. 6.

Akın, C. (1986). Faizsiz Bankacılık ve Kalkınma. İstanbul: Kayıhan Yayınları.

Akpolat, O.M.(2018). Konut satın alma maliyet analizleri: Bankalar-katılım bankaları-elbirliği sistemi karşılaştırması, Tokat Gaziosmanpaşa Üniversitesi Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı Muhasebe Finansman Bilim Dalı (Yüksek Lisans Tezi).

Amin, A. Anwar, S. Adnan, A. Nawaz, M. Alawfi, K. Hussain, A. ve Huang, K. (2017). Customer churn prediction in the telecommunication sector using a rough set approach," Neurocomputing, vol. 237, pp. 242-254.

Brownlee, J. (2016) How to Grid Search Hyperparameters for Deep Learning Models in Python With Keras.

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition, data mining and knowledge discovery, Kluwer AcademicPublishers, 2 (2), 121-167.

Busuttil, S. (2003). Support vector machines. 1st Computer Science Annual Workshop (CSAW'03), Msida. 34-39.

Chen, J. ve Popovich, K. (2003). Understanding customer relationship management (CRM): People, process and technology, Business Process Management Journal, 9 (5), 672-688.

Coussement, K., Lessmann, S. ve Verstraeten G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry," Decision Support Systems, vol. 95, pp. 27-36.

Cortes, C. Vapnik, V. (1995). Support-Vector Network, Machine Learning, 20(3):273–297.

Duda, R. O. Hart,P. E. ve Stork, D. G. (1999). Pattern Classification, Wiley, 2nd edition.

Huang, B. Kechadi M. T. ve Buckley, B. (2012). Customer Churn Prediction in Telecommunications, Expert Systems with Applications, vol. 39, no.1, pp. 1414-1425.

Hu, L.Y. Huang, M.W. Ke, S.W. Tsai, C.F. (2016), The Distance Function Effect on kNearest Neighbor Classification for Medical Datasets, Springer Plus, 5(1), 1-9.

İkbal, Z. Greuning, H. V. (2008) Principles and Development of İslamic Finance, World Bank, No: 424481

Kavzaoğlu, T.Çölkesen, Ġ. (2010). Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının EtkilerininİncelenmesiHarita Dergisi, Sayı 144,73-82.

Kaynar, O. Tuna, M. F. Görmez Y. ve Deveci, M. A. (2017) Makine öğrenmesi yöntemleriyle müşteri kaybı analizi, C.Ü. İktisadi ve İdari Bilimler Dergisi.

Keramati, A. Jafari-Marandi, R. Aliannejadi, M. Ahmadian, I. Mozaffari M. ve Abbasi, U.(2014).Improved churn prediction in telecommunication industry using datamining techniques,Applied Soft Computing, vol. 24, pp. 994–1012.

Kim, K. Jun, C.-H. ve Lee, J. (2014). Improved churn prediction in telecommunication industry by analyzing a large network, Expert Syst. Appl., vol. 41, no. 15, pp. 6575–6584.

Kisioglu P. ve Topcu, Y. I.(2011) Applying Bayesian Belief Network approach to customer churn analysis: a case study on the telecom industry of Turkey, Expert Syst.Appl., vol. 38, pp. 7151–7157.

Kuhn, M. Ve Johnson, K. (2013). Applied Predictive Modeling

Lee, Ivrissimtzis, Seide (2006). Geometric Modeling and Processing 4th International Conference, Pittsburgh, PA, USA, Proceedings.

Ling, R. ve Yen,D (2001). Customer Relationship Management: An Analysis Framework and Implementation Strategies, Journal Of Computer Information Systems, vol. 41, no. 3, pp. 82-97

Mabid, A (1988). Faizsiz Para Ekonomilerinin Nisbî Etkinliği: Kâğıt Para Olayı, çev. Abdullah Yavaş, İslâm İktisadı Araştırmaları I içinde, der. Ahmet Tabakoğlu, İstanbul: Dergah Yayınları, s. 80-113.

Namane, A. Guessoum, A. Soubarı, E. H. ve Meyrueıs, P. (2014). CSM neuralnetwork for degraded printed character optical recognition, Journal ofVisual Communication and Image Representation, 25, 1171-1186.

Odabaşı, Y. (2000). Satışta ve Pazarlamada Müşteri İlişkileri Yönetimi, Sistem Yayıncılık, İstanbul.

Osowski, S. Siwekand, K. ve Markiewicz, T. (2004), MLP and SVM Networks, Proceedings of the 6th Nordic Signal Processing Symposium.(pp.37-40)

Prasath, V.B.S. Alfeilat, H.A.A. Hassanat, A.B.A. Lasassmeh, O. Tarawneh, A.S. Alhasanat, M.B. Salman, H.S.E. (2019), Distance and Similarity Measures Effect on the Performance of K-Nearest Neighbor Classifier—A Review, Big Data, 7(4), 221-248.

Reichheld (2001). A paradigm shift for a successful launch of a locallybased start-up in the food supply chain**.**

Roberts ,G. ve Phelps (2001). Customer Relationship Management: How to Turn a Good Business Into a Great One!, Londra: Hawksmere.

Russell, D.(2007).Building Supply Chain Collaboration: A Typology of Collaborative Approaches. International Journal of Logistics Management, 18, 174-196.

Salahuddin, A. (2006). Islamic Banking Finance and Insurance. Gombak Kuala Lumpur: A.s.Noordeen Publishing, https://dergipark.org.tr/tr/download/article-file/65371

Shanmugam, B. Zahari, Z. (2009). A primer on Islamic finance.

Suykens, A. K. (2002). Least Squares Support Vector Machines.

Swift, R.S. (2001). Accelerating Customer Relationships:Using CRM and Relationship Technologies, New Jersey: Prentice Hall Professional.

Torggler, M. (2009). The Functionality and Usage of CRM Systems

Usulcan, E. (2013). Katılım Bankalarında Bireysel Pazarlama Faaliyetleri ve Tüketicilerin Katılım Bankaları Tarihinde Etkili Olan Faktörler, Yüksek Lisans Tezi, T.C. Atılım Üniversitesi Sosyal Bilimler Enstitüsü İşletme Yönetimi Anabilim Dalı, Ankara

Vafeiadis, T. Diamantaras, K. Sarigiannidis, G. ve Chatzisavvas, K. (2015).A comparison of machine learning techniques for customer churn prediction, Simulation Modelling Practice and Theory, vol. 55, pp.1-9.

Vapnik, V.N. (1995),The Nature of Statistical Learning Theory, Springer-Verlag, New York.

Verbeke, W. Dejaeger, K. Martens, D. Hur ,J. ve Baesens, B. (2012).New Insights into Churn Prediction in the Telecommunication Sector: A Profit Driven Data Mining Approach", European Journal of Operational Research, vol. 218, pp. 211-229.

Weinberger, K.Q. Blitzer, J. Saul, L.K. (2006). Distance Metric Learning for Large Margin Nearest Neighbor Classification.

Wilson, R. (2008). Muhammed Bakır Es-Sadr"ın İslam Ekonomi Düşüncesine Katkısı"", Hakikat Dergisi. 91.

Witten, I.H., Frank, E. and Hall, M.A. (2011). Data Mining: Practical Machine Learning Tools and Techniques. 3rd Edition, Morgan Kaufmann Publishers, Burlington.

Url-1 < https://makersturkiye.com/makine-ogrenmesi-nedir-is-hayatinda-nasil-faydalanir/