

SİGORTACILIK SEKTÖRÜNDE MAKİNE ÖĞRENMESİ İLE MÜŞTERİ KAYBI ANALİZİ

CUSTOMER CHURN ANALYSIS IN THE INSURANCE SECTOR USING MACHINE LEARNING

Hande Esin AKYİĞİT *, Tuğrul TAŞCI **

ÖZET

Hızlı büyüyen ve rekabet gücünün arttığı günümüzde, yeni müşteri edinme çabası ve maliyeti var olan müşteriye kaybetmeme çabası ve maliyetinden fazla olması, firmaları var olan müşterilerin kaybedilmemesi gerektiği düşüncesine itmiştir. Güçlü rakiplerin olduğu sektörde tüketicilerin bir ürün veya hizmet sebebiyle bir şirkete tercih edip ihtiyaçlarını devamlı olarak bu şirket üzerinden karşılayan sadık müşteri portföyünü arttırmak amacıyla müşterinin tercih ettiği bir ürün veya hizmeti bırakma ihtimali üzerine analitik çalışmalar yapılmıştır. Mevcut müşterilerin profilleri ve davranışları incelenerek şirkete bırakma ihtimali olan müşterileri bulma, bu müşterilerin memnuniyetlerini arttırmayı hedefleyen müşteri kaybı analizi, stratejik karar verme ve planlama sürecinin en önemli aşamalarından biri olmaktadır. Bu çalışmada bankacılık, telekomünikasyon, e-ticaret gibi müşteri sayısı ile gelir miktarının doğru orantılı olduğu sigortacılık sektöründe var olan şirketin verileri kullanılarak bir müşteriye ait içerisinde yaş, cinsiyet, doğum yeri gibi sosyodemografik bilgilerin yanı sıra kullanılan araç marka, model bilgilerinin de bulunduğu öznitelikler belirlenmiştir. Belirlenen öznitelikler makine öğrenmesi algoritmalarından Karar Ağacı (Decision Tree) Algoritması, Rastgele Orman (Random Forest) Algoritması ve K-En Yakın Komşu (K Nearest Neighborhood) Algoritmaları ile terk eden müşterilerin profilleri analiz edilip, terk etme ihtimali olan müşteriler tahmin edilmiştir. Çalışmada en başarılı sonucu veren Rastgele Orman (Random Forest) Algoritması ile bu çalışma bir sınıfa dâhil edilip son kullanıcı tarafından sürekli yapılmasına olanak sağlanmıştır.

Geliş Tarihi/Received: 03.11.2021
Kabul Tarihi/Accepted: 09.02.2022

Araştırma Makalesi/Research Article

*
Sakarya Üniversitesi, Fen Bilimleri Enstitüsü,
Bilişim Sistemleri Mühendisliği Programı,
Sakarya, Türkiye

Department of Information Systems
Engineering,
Sakarya University, Sakarya / Turkey

ORCID: 0000-0001-9225-0776

handesinzeytin.bsmt@gmail.com

**
Sakarya Üniversitesi, Bilgisayar ve Bilişim
Bilimleri Fakültesi, Sakarya, Türkiye

Faculty of Computer and Information
Sciences, Sakarya University, Sakarya /
Turkey

ORCID: 0000-0003-3820-6453

Anahtar Kelimeler: Müşteri Kaybı Analizi, Makine Öğrenmesi, Rastgele Orman, Karar Ağacı, K-En Yakın Komşu

ABSTRACT

In today's fast-growing and increasing competitiveness, the effort to acquire new customers and the cost of not losing the existing customer, and the fact that it is more than the cost, have led companies to think that existing customers should not be lost. In the sector with strong competitors, analytical studies have been carried out on the probability of the customer leaving a product or service preferred by consumers in order to increase the portfolio of loyal customers who prefer a company for a product or service and meet their needs continuously through this company. Customer loss analysis, which aims to find customers who are likely to leave the company by examining the profiles and behaviors of existing customers, and to increase the satisfaction of these customers, is one of the most important stages of the strategic decision-making and planning process. In this study, using the data of the company in the insurance sector, where the number of customers and the amount of income are directly proportional, such as banking, telecommunications, e-commerce, the attributes of a customer, including sociodemographic information such as age, gender, place of birth, as well as vehicle brand and model information, were determined. The profiles of the customers who left were analyzed with the Decision Tree Algorithm, Random Forest Algorithm and K-Nearest Neighborhood Algorithms, which are machine learning algorithms, and the customers who are likely to leave were estimated. With the Random Forest Algorithm, which gave the most successful results in the study, this study was included in a class and allowed to be carried out continuously by the end user.

Keywords: Customer Churn Analysis, Machine Learning, Random Forest, Decision Tree, K-Nearest Neighbor

1. GİRİŞ

Son zamanlarda telekom, sigortacılık, hizmet gibi çeşitli sektörlerde faaliyet gösteren birçok şirket küreselleşme ve artan rekabet koşullarında varlığını korumak ve devam ettirebilmek amacıyla çeşitli stratejiler geliştirmek zorunda kalmışlardır. Artan rekabet koşullarında müşteri davranışları, profilleri ve beklentileri oldukça değişmiştir. Firmalar bu rekabet ortamında üstünlük kazanmak için çeşitli yöntem arayışı içine girmişlerdir. Aranılan yöntemlerin ihtiyaca cevap verebilmesi için firmaların var olan müşterilerinin profillerini analiz edebilmesi, satın alma alışkanlıklarını izleyebilmesi, işletmenin amaç ve hedefleri

doğrultusunda müşteri ilişkileri yönetimini faydalı bir biçimde kullanabilmesi ile mümkün olacaktır. Müşteriyi elde tutmak, müşteri merkezci yöntemler ile müşteriyi değerli kılmak, bağlılık arttırmak firmaların en büyük hedefleri arasında yer almaktadır.

Müşteri kaybı tanımı rekabet koşullarında firmaların müşterilerin tutumlarında meydana geldiği gözlemlenen değişimler sonucunda bulunduğu firmayı bırakmalarının artmasıyla daha fazla gündeme girmiştir. Müşteri kayıplarının fazla olmasının sebepleri arasında, rakip firmalarda memnuniyetsizlik yaşayan müşterileri farklı özellikler vurgulayarak ve çeşitli kampanyalar oluşturarak bulunduğu firmadan farklı firmaya çekme yöntemi oldukça etkili olmaktadır.

Müşteriyi düşünüp faydalarını gözeten çalışan kadrosu, daha düşük ücretlendirme, değişim maliyeti, yapılan tanıtımların etkisi, konum ve çeşitli hizmet teklifleri sebebiyle rakip firmalara geçme yatkınlığı göstermiştir (Farquad, Ravi ve Raju, 2014). Geçiş eğilimi gösteren müşterileri önceden tespit edebilmek ve o müşteriler üzerinden farklı kampanyalar uygulayarak müşterinin tatminsizliğini azaltıp yeniden firmaya bağlamak günümüz işletmeleri için çok önemli bir yer edinmiştir.

Paralelinde bilişim sektöründe yürütülen gelişmeler sayesinde işletmeler, verileri saklayabilir, kolay erişebilir, işleyebilir ve işlenen veriler ile anlamlı bilgiler elde etme imkânına sahip olabilmektedirler. Çeşitli sektörlerde bulunan işletmeler, müşterilerinin satın alma davranışlarının yanı sıra müşteriye ait farklı özellikleri de depolamaktadırlar. Depolanan veri yığınlarından makine öğrenmesi yöntemleri kullanılarak anlamlı bilgiler, isabetli tahminler elde edilebilmektedir. Elde edilen sonuçlar, müşteri odaklı uygulamaların geliştirilebilmesine ve müşteri kaybını engellemek için önlemler alınmasına fayda sağlamaktadır.

Makine öğrenmesi, büyük veriler üzerinden anlamlı çıkarımlar yapabilen ve yapısal olarak öğrenebilen algoritmalara verilmiş isimdir. Günümüz verilerinin insan gücü ile işlenemeyecek boyutlara gelmiş olması sebebiyle makine öğrenmesi konusu oldukça önemli bir hâle gelmektedir. Firmalar tuttukları müşteri bilgilerini makine öğrenmesi yöntemleri sayesinde anlaşılabilir, sınıflandırılmış ve ilişki kurulmuş anlamlı bilgiler olarak elde etmektedirler.

Günümüz sigortacılığında da diğer sektörler gibi rekabet ortamının artması ve rakip firmaların müşterileri elde etme çabası sonucu firmalar müşterilerini kaybetmemek için çalışmalar yapmaya başlamıştır. Yapılan literatür incelemeleri sonucunda sigortacılık sektöründe müşteri kaybı analizinin diğer sektörlerle oranla daha az yapıldığının fakat bu çalışmaya müsait veri kapasitelerinin olduğu tespit edilmiştir. Bu sebeple bu çalışmada artan rekabet koşullarında bulunduğu konumu koruyabilmek için ve sahip oldukları veri kapasitelerinin müşteri kayıp analizine uygun olması sebebiyle, bir sigortacılık sektöründe özel bir şirkete ait müşteri verileri üzerinde makine öğrenmesi yöntemleri ile müşteri kayıplarına yönelik modeller geliştirilerek tahminler yapılmış ve bu modellerin çıktılarını kullanıcıların her zaman görebilmek istemesi sebebiyle bir arayüz oluşturulmuştur. Bu çıktılar ile birimlerin kendi müşterilerini kaybetmemesi için gerekli aksiyonların alınmasına imkân sağlamaktadır.

2. MAKİNE ÖĞRENMESİ (MACHINE LEARNING)

Yapay zeka branşlarından olan makine öğrenmesi, yapısal olarak sürekli öğrenebilen, büyük ve karmaşık veriler üzerinden anlamlı çıkarımlar yapabilen algoritmalara verilen isimdir. Dijital veri üretiminin kurumsal ve bireysel düzeyde hızla artış gösterdiği günümüz dünyasında, verilerinin insan gücü ile işlenemeyecek boyuta gelmiş olması, klasik algoritmalar ile verilerin sınıflandırılmasının mümkün olmaması ve ileriye dönük tahminlemenin yapılamaması sebebiyle makine öğrenmesi ve yöntemleri önemli bir hâle gelmiştir. Arthur Samuel (1959) makine öğrenmesini "Açıkça programlanmadığı halde makinelerle öğrenme yeteneği kazandıran disiplin." olarak tanımlamıştır.

Makine öğrenmesi, Gözetimli (Supervised), Gözetimsiz (Unsupervised) ve Takviyeli (Reinforcement) olmak üzere üç temel gruba ayrılmaktadır.

Hedef değişkeni belli olan öğrenme yöntemi gözetimli öğrenme(Supervised Learning) olarak adlandırılmaktadır. Tahmin edilmek istenen sınıflar bellidir, buradaki amaç girdi değerleri ile hedef değişken arasında bir bağlam öğrenerek yeni gelen değerlerde bu bağlamdan yola çıkarak tahminler yapmaktır. Veri setindeki hedef değişkeni kategorik ise sınıflandırma (Classification), nümerik ise regresyon (Regression) algoritmaları kullanılmaktadır.

Gözetimsiz öğrenme (Unsupervised Learning) yöntemi şeklinde hedef değişkeni bulunmamakta, sadece girdi bulunmaktadır. Amaç girdiler arasındaki yakınlıklar, düzenlilikleri bulmaktır. Gözetimsiz öğrenme ile kümeleme ve kestirim yapılabilmektedir.

Gözetimli ve gözetimsiz öğrenmeden daha farklı olan takviyeli öğrenme (Reinforcement Learning) yönteminde ödül-ceza olarak adlandırılan sistem bulunmaktadır. Burada makinenin ana hedefi elde etmek istenilen eyleme ulaşırken kullanılan en doğru yolu bulmaktır. Doğru yolu bulmaya çalışırken yaptığı hatalardan çıkarımlar yaparak belli bir ödül-ceza sistem temeli üzerinde çalışmakta ve çıkarımlardan en optimize yol ile doğru eylem bulunmaya çalışılmaktadır.

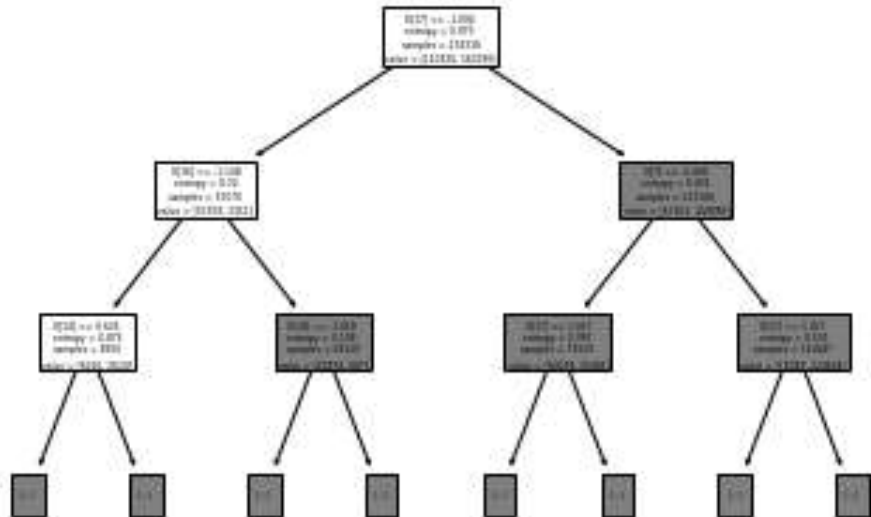
Bu çalışmada, kullanılan veri setinde elde edilen hedef değişkeninin kategorik olması sebebiyle Gözetimli (Supervised) öğrenme yöntemlerinden K-En Yakın Komşu (K-Nearest Neighbors) algoritması, Karar Ağacı (Decision Tree) Algoritması ve bu karar ağaçlarının birleşmesi ile oluşturulan Rastgele Orman (Random Forest) algoritması kullanılmıştır.

2.1. Karar Ağaçları

Parçalardan bütüne doğru ulaşan akıl yürütme metodu kullanılarak verilerin sınıflandırılması veya sonuç tahmini yapılmasına olanak sağlayan karar ağaçları, hedef ve özelliklere göre karar ve yaprak düğümlerinden oluşan ağaç yapısı şekline benzerlik göstererek model oluşturan makine öğrenmesi sınıflandırma yöntemlerindedir. Karar ağaçları kök düğüm, düğüm ve dallardan meydana gelmektedir. Kurallar doğrultusunda sorular sorulup bu kuralların cevapları bir araya getirilerek yeni kurallar oluşturulmaktadır.

Peşi sıra sorulan soruların cevabına göre belirlenen değişken ağaç modelinin birinci seviyesi olan kök düğümünü oluşturmaktadır. Bağımlı değişkeni temsil eden kök ile başlanan ağacın yapısında bütünden parçalara doğru gidildikçe veri kümeleri daha küçük parçalara ayrılarak dalları ve yeni düğümleri oluşturmaktadır. Her bir soruyu bir düğüm temsil etmekte ve verilen cevap sayısı kadar dar olmaktadır. Ağaçta bulunan dallar, 'eğer-ise' kurallarını oluşturmakta ve seçilen adıma göre bir sonraki düğüme aktarılmaktadır.

Farklı bir soru meydana gelene kadar ağaç modeli bu şekilde meydana gelmektedir. Son olarak bir sınıfı temsil eden Şekil 1.'de örnek olarak gösterilen son düğüme ulaşılmaktadır.



Şekil 1. Örnek Bir Karar Ağacı (Decision Tree) Modeli

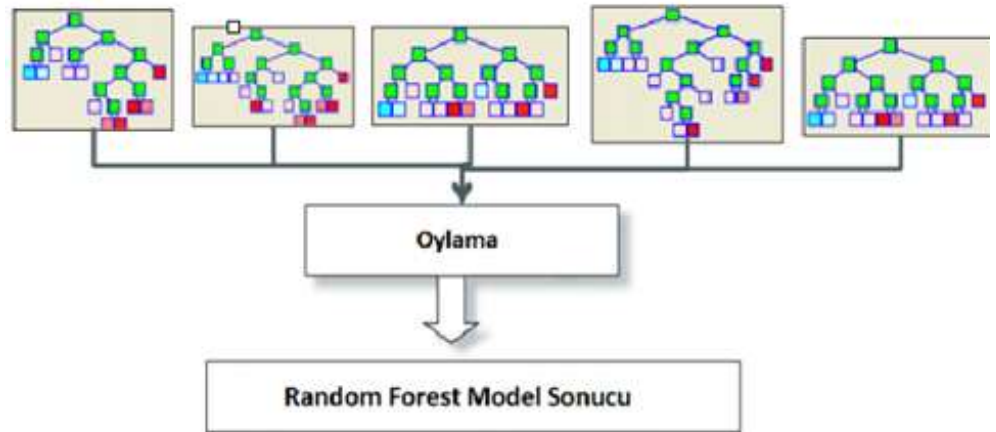
Karar ağaçları, diğer istatistiksel çözümlere göre oluşturulması daha kolay, ağaç şeklinde sınıflandırıcılardır. Modelin ağaca benzeyen yapısı, bağımlı ve bağımsız değişkenler arası ilişkiyi göstermektedir. Ağaçta meydana gelen yaprak düğümleri hedef niteliğinin değeri olmaktadır. Karar düğümü ise nitelikte uygulanan test değerini göstermektedir. Karar ağacı verilen örneği bütünden en küçük parçaya kadar inceleyip sınıflandırmaktadır. Karar ağaçlarının öğrenme algoritmaları, bir hipotezi göstermek için karar ağacı kullanılmaktadır. Öğrenme kümesinde, ham veri bütün olarak incelenerek olabildiğince en iyi şekilde sınıflandırılmaktadır (Tosun, 2006). Algoritma bu işlemi yinelemeli olarak tekrar edip son ortaya çıkardığı karar ağacı modeli en son hipotezi oluşturmaktadır. İdeal olarak kabul edilen karar ağacı, öğrenme kümesi dışındaki verilerde de aynı kuralları oluşturur ya da az hata payıyla aynı hipotez sonuçlarını ortaya çıkarmaktadır

Karar ağaçları hem kategorik hem de sayısal verileri işleyip anlamlandırabilmektedir aynı zamanda bağımsız ve bağımlı değişkenler sebebiyle eksik veya kayıp değerlerden etkilenmemektedirler.

Bazı durumlarda ağaç modeli oluşturma ve ağaç budama karmaşıklığı çok olabilmekte veya az girdi ile oluşturulan karar ağacı modeli anlamlı bilgi yansıtamayabilmektedir.

2.2. Rastgele Orman (Random Forest) Algoritması

Karar ağacı modellerinin en büyük problemleri arasında veriyi ezberleme – aşırı öğrenme bulunmaktadır. Rastgele orman modeli bu problemi önlemek amacıyla veri setinden rassal olarak 100'lerce farklı alt ağaçlar seçip bunları eğitmektedir. Bu yöntem sebebiyle 100'lerce karar ağacı modeli oluşturulup oluşturulan karar ağaçları bireysel olarak tahminde kullanılmaktadır. Gün sonunda ise problem regresyonsa karar ağaçlarının tahminlerinin ortalaması, problem sınıflandırmaysa tahminler arasında en çok oy alan seçilmektedir. Bu yöntem en büyük problem olan aşırı öğrenme ve ezberlemenin de önüne geçip farklı veri setlerinin meydana getirebildiği aykırı veri (Outlier) problemini de minimum seviyeye indirmektedir.



Şekil 2. Örnek Bir Rastgele Orman (Random Forest) Modeli

Birden fazla karar ağacı modeli çıktıları sonucunda oranlama yapılarak oluşturulan rastgele orman algoritması, karar ağaçlarında görülmekte olan aşırı öğrenmenin ve ezber yapmanın önüne geçmiş olmaktadır. Her bir karar ağacı için kullanılan farklı veri setleri sebebiyle aykırı veri (Outlier) problemini de minimum seviyeye indirmektedir. Rastgele Orman (Random Forest) bir torbalama yöntemi olmaktadır. Orijinal kümeden çoklu alt kümeler oluşturan torbalama yönteminde çeşitli yöntemler paralel olarak farklı veri seti kümeleri ile eğitilip tüm modellerin oluşturduğu sonuç oylamaya tabi tutularak nihai sonuç ortaya çıkarılmaktadır. Basit rastgele orman algoritması modeli Şekil 2.'de gösterilmektedir.

2.3. KNN (K-Nearest Neighbors) Algoritması

KNN algoritması, içerisinde tahmin edilecek değerin bağımsız değişkenlerinin oluşturduğu vektörün en yakın komşularının hangi sınıfta yoğun olduğu bilgisi üzerinden sınıfı tahmin etmeye dayanmaktadır.

KNN (K-Nearest Neighbors) Algoritması iki temel değer üzerinden tahmin yapmaktadır ve uzaklık hesaplamaları sırasında kullanılması amacıyla Oklid (Euclidean), Manhattan ve Minkowski Uzaklık Fonksiyonları gibi farklı fonksiyonlar belirlenmektedir. Bu fonksiyonların arasında en yaygın olarak kullanılan Öklid Uzaklık Fonksiyonu olmaktadır. Fonksiyon ile p boyutlu bir uzayda i ve j noktaları arasında bulunan uzaklık Denklem 1'de verilen formül ile elde edilmektedir:

$$d(i, j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (1)$$

Veri setinde bulunan değişkenlerin sayısının ikiden fazla olduğu durumlarda Standardize Edilmiş Öklid Uzaklık Fonksiyonu kullanılmaktadır. Her bir değişken kendi içerisinde z dönüşümü uygulanarak standardize edilip eşitlik formüle yerleştirilmektedir. Böylelikle değişkenler arasındaki ölçüm farklılıkları ortadan kalkmış olmaktadır.

Yapılan çalışmada en yaygın kullanılan yöntemlerden biri olan makine öğrenmesi kullanılarak müşteri kaybı analizi yapıp bir sonraki bölümde adımlarından bahsedilecektir.

3. SİGORTACILIK SEKTÖRÜNDE MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE MÜŞTERİ KAYBI ANALİZİ

Sınıflandırma ve tahminleme algoritmalarında kullanılan verinin kalitesi ve boyutu yapılan tahminleme sonucunda başarı oranı ile doğru orantılıdır. Bu sebeple modeli eğitmek için kullanılmakta olan veri setinin hazırlanması sürecin en önemli aşamalarından kabul edilmektedir (Kunt, 2019).

Bu çalışmada model oluşturma, oluşturulan modeli eğitme ve bu modeli kullanma aşamalarından önce modeli oluştururken ve tahminleme yaparken kullanılacak veri seti özellikleri çıkarılıp uygun formata getirilmiştir.

Bu çalışma bir sigorta şirketine ait anonim veriler ile yapılmıştır. İlgili sigorta şirketinin arabalarına kasko poliçesi yaptıran müşterileri arasında sadece veri işleme izni olan müşterilerine ait veriler anonimleştirilip KVKK kapsamında kullanıma uygun hâle getirilmiştir.

Müşteri kaybı analizi için poliçesini iptal eden ve poliçesini yenileyen olmak üzere iki farklı tipte müşteri profiline ihtiyaç duyulup bu formatta veri seti hazırlanmıştır. Müşterilerin poliçesini yenileyip yenilemediğini belirleyen bu öznitelik yanında, kayıp olan müşteri profillerini belirlemek amacı ile sosyodemografik bilgiler, kullanılan araç, marka, model bilgileri, poliçesini daha önce kaç kez yenilediği bilgileri de yer almaktadır.

3.1. Veri Setinin Oluşturulması

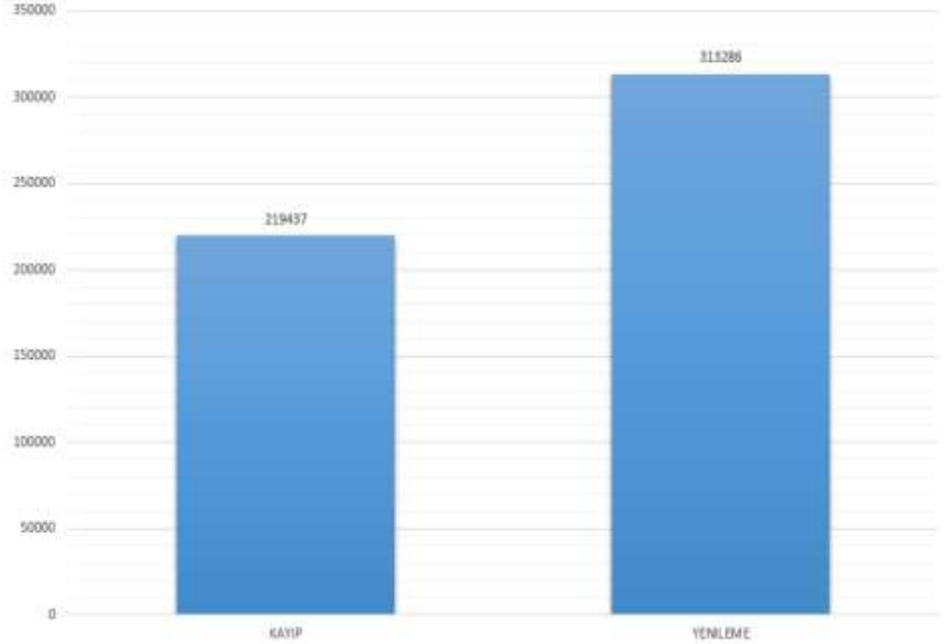
Tahminleme için kullanacağımız model içerisindeki veri setinde, kasko poliçesini yenileyen ve yenilemeyen müşterilerin sosyodemografik bilgileri, kullandığı aracın öznitelikleri, poliçe özellikleri gruplanarak tutulmaktadır.

Müşterilere ait Tablo 1.'deki maddelerde yer alan öznitelikler çıkarılmış ve veri seti oluşturulmuştur.

ÖZİNİTELİK	AÇIKLAMA
YAŞ	Müşterinin yaş bilgisi
CİNSİYET	Müşterinin cinsiyeti
MEDENİ HÂL	Müşterinin medeni hali
ÇALIŞMA DURUMU	Müşterinin çalışma durumu
MESLEK	Müşterinin mesleği
YAŞADIĞI İL	Müşterinin yaşadığı il
EĞİTİM SEVİYESİ	Müşterinin eğitim seviyesi
MÜŞTERİNİN ÖDEDİĞİ TOPLAM PRİM	Müşterinin tüm poliçeleri için ödediği toplam prim
PLAKA SAYISI	Müşterinin şimdiki zamana yaptığı araç/araçları plakalarının sayısı
MARKA	Müşterinin güncel kasko poliçesi yaptırdığı aracın marka bilgisi
MARKA SAYISI	Müşterinin şimdiki zamana kadar tüm poliçeleri dâhilinde yaptığı araç/araçları markalarının sayısı
MODEL	Müşterinin güncel kasko poliçesi yaptırdığı aracın model bilgisi
MODEL SAYISI	Müşterinin şimdiki zamana kadar tüm poliçeleri dahilinde yaptığı araç/araçları modellerinin sayısı
KULLANIM TARZI	Aracın kullanım tarzı
MODEL YILI	Müşterinin güncel kasko poliçesi yaptırdığı aracın model yılı bilgisi
PLAKA İL KODU	Müşterinin güncel kasko poliçesi yaptırdığı aracın plaka il kodu bilgisi
ORTALAMA HASARSIZLIK KADEMESİ	Müşterinin şimdiki zamana kadar tüm poliçeleri dâhilinde belirlenen hasarsızlık kademesi
HASARSIZLIK KADEME	Müşterinin güncel kasko poliçesi için belirlenen hasarsızlık kademesi
ARAÇ YAŞI	Müşterinin güncel kasko poliçesi yaptırdığı aracın yaş bilgisi
SATIŞ KANALI	Müşterinin güncel kasko poliçesini nereden yaptırdığı bilgisi
UNSUR TİPİ	Müşterinin güncel kasko poliçesindeki aracın unsur tipi
YENİLEME	Müşterinin poliçesinin yenilenip yenilenmediğinin bilgisi

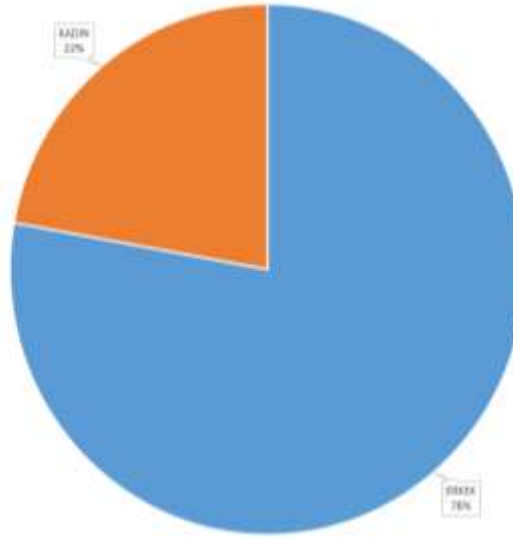
Tablo 1. Özgirişimlerin Veri Setinde Kullanılan Öznitelikler ve Açıklamaları

Veri setinde incelenmek üzere alınan 532.723 müşteri ve her bir müşteri için 22 öznitelik bulunmaktadır. Var olan müşteriler arasında 219.437 müşteri poliçesini yenilememiş, 313.286 müşterinin ise poliçesini yenilemiş olduğu gözlemlenmektedir. Şekil 3.'de çalışmada kullanılacak veri setine ait sınıfların müşteri kayıp adetleri gösterilmiştir. Sınıflar arası farka bakıldığında veri setinin dengeli olduğu gözlemlenmektedir.



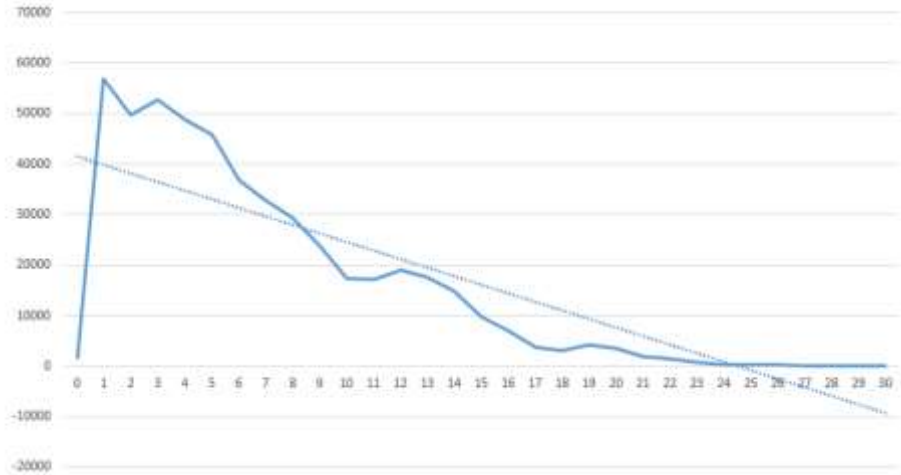
Şekil 3. Kişisel Faktörler Arasındaki Doğrudan İlişkiler

Şekil 4'de veri setinde bulunan ve poliçesini yenileyen müşterilerin cinsiyet dağılımı verilmektedir. Grafiğe göre poliçesini yenileyen müşterilerin %78'inin erkek olduğu gözlemlenmektedir.



Şekil 4. Veri Setindeki Kayıp Müşterilerin Cinsiyet Dağılımı

Şekil 5'te veri setinde kasko poliçesi yaptırılan araçların trend dağılımı verilmektedir. Grafiğe göre aracın yaşı arttıkça kasko sigortası yaptıırma oranının da doğrusal bir şekilde düştüğü gözlemlenmektedir.



Şekil 5. Veri Setindeki Poliçe Yaptırılan Araçların Yaşlarının Trendi

Veri seti istatistiksel olarak incelenip kullanılacak veri iyice tanıdıktan sonra birçok ön hazırlık aşamasından geçip model oluşuna uygun hâle getirilmesi gerekmektedir. Eksik verilerin incelenip gerektiğinde uygun koşullarda doldurulması, veri seti içindeki verilerin uyumu için aykırı verilerin (Outliers) tespit edilip düzenlenmesi, veri setindeki alanların diğer alanlarla arasındaki pozitif veya negatif yönlü ilişkinin (Korelasyon) tespiti gibi ön hazırlık aşamaları veri setini modele hazırlayıp modelden maksimum başarı alınmasını sağladığı için veri madenciliği ve makine öğrenmesi yöntemleri süreçlerinde oldukça büyük öneme sahip olmaktadır. Çalışmamızda yapılan veri ön hazırlık çalışmaları bir sonraki başlık olan yöntem altında detaylandırılıp uygulama başlığı altında ise nasıl uygulandığı anlatılmıştır.

3.2. Veri Ön İşleme

Hazırlanan veri setine modelin uygulanabilmesi ve model uygulandığında yüksek başarının elde edilebilmesi için veri seti, veri ön işleme aşamalarından geçmiştir.

3.2.1. Eksik Verilerin Temizlenmesi

Veri setinde eksik değerlerin olup olmadığı kontrol edilmiştir. Makine öğrenmesi algoritmalarında kullanılan veri setlerinde bulunmakta olan eksik kayıtlar, modeli

oluşturma, eğitime ve tahminleme çalışmalarında tutarsız sonuçların oluşmasına sebebiyet vermektedir. Eksik veri olan veri setlerinde genel olarak aşağıdaki adımlar uygulanmaktadır.

Veri setinde eksik değerlerin olup olmadığı kontrol edilmiştir. Makine öğrenmesi algoritmalarında kullanılan veri setlerinde bulunmakta olan eksik kayıtlar, modeli oluşturma, eğitime ve tahminleme çalışmalarında yer alabilecek tutarsız sonuçların meydana gelmesine sebebiyet vermektedir. Eksik veri olan veri setlerinde genel olarak aşağıdaki adımlar uygulanmaktadır.

Veri setinde eksik verilerin bulunduğu sütunun veya satırın silinmesi. Eğer sütun yani özniteliğin veya satırın belirli bir yüzde oranı üzerinde eksik değerler (Missing Value) içeriyorsa veri setinden komple çıkarılması.

Veri setinde bulunan eksik verilerin sabit bir değer ile doldurulması.

Eksik verilerin bulunduğu sütunun ortalaması alınarak eksik verilerin bu ortalama ile doldurulması.

Süreklilik gerektiren sayısal değerler bulunan özniteliklerin kategorikleştirilmesi. Yani çok geniş aralıklarda sayısal verilerin belli aralıklara göre kategorik hâle getirilmesidir. "0'dan küçük", "0-50", "50-100" şeklinde kategorikleştirilebilmektedir.

Veri setinde eksik olan veriyi yüksek ilişkisi (High Correlation) olduğu başka bir özniteliğin değerine göre tahmin edilmesi.

Bu çalışmada veri setindeki boş veri içeren sütunların satır bazında boş adetlerinin genel adede oranı alınmıştır. Bu oran boş verilerin analizini yapmak için kullanılmıştır. Sütun bazında %1'in altında boş olan veriler silinmiş, %1'in üstünde boş olan sütunlarda ise string değerler var ise sütunun modu, nümerik değerler var ise ortalama değerler ile doldurulmuştur.

Veri setinde bulunan değişkenler arasında bulunan ilişki, bu ilişkinin yönü ve bu ilişkinin şiddeti ile bilgileri sağlayan istatistiksel yöntem olan korelasyon analizi uygulanmaktadır. İki veya daha fazla değişkenler arasındaki ilişkinin matematiksel bağıntısı "Regresyon Analizi", ilişkinin yönü ve bu ilişkinin derecesi ise "Korelasyon Analizi" ile incelenmektedir.

3.2.2. Korelasyon Analizi

Korelasyon katsayısı, bağımlı ve bağımsız değişkenler arasında bulunmakta olan ilişkinin gücünü göstermektedir. Örneğin bir insanın kalp krizi geçirme riskiyle sigara kullanımı arasındaki ilişki veya insanlardaki eğitim seviyesi (X) ile içinde bulunduğu coğrafi konum (Y) arasında bulunmakta olan ilişki korelasyon katsayısı ile incelenmektedir. Korelasyon katsayısı değişkenler arasındaki ilişki hakkında bilgi vermektedir.

Korelasyon katsayısı, açıklanan varyans ve açıklanmayan varyans oranı olarak tanımlanmaktadır.

Korelasyon katsayısı değişkenler arasında bulunan doğrusal ilişkinin ölçütü olup incelenen değişkenler birbirinden bağımsız ve $-1 \leq r \leq 1$ arasında bulunmaktadır. Korelasyon katsayısının 0 ile 1 arasında bir değer alması değişkenler arasındaki ilişkinin pozitif yönde bir ilişki olduğunu, -1 ile 0 arasında yer alması ise negatif yönde bir ilişki olduğunu göstermektedir. Tablo 4.2.'de korelasyon aralığı ve ilişki düzeyleri verilmiştir.

3.2.3. Özellik Ölçeklendirme

Bir veri kümesindeki özelliklerin değerlerini, mesafe hesaplamasına orantılı olarak katkıda bulunacak şekilde ölçeklendirme işlemidir. En yaygın olarak kullanılan özellik ölçekleme tekniği Standardizasyon (Veya Z-Skoru Normalleştirme) ve Min-Max ölçeklendirme.

Standardizasyon: Z-Puanı Normalleştirme / Standardizasyon olarak da bilinmektedir. χ özelliklerinin yeniden ölçeklendirilmesi işlemidir, böylece $\mu = 0$ ve $\sigma = 1$ olmaktadır. Teknik olarak, standardizasyon ortalamayı çıkararak ve standart sapmaya bölerek verileri merkezler ve normalleşmektedir. Elde edilen değerlere standart puan (Veya Z-Puanı) denir ve Denklem 2'deki gibi hesaplanabilmektedir.

$$Z = \frac{X - \mu}{\sigma} \quad (2)$$

Formüldeki μ ortalama ve σ ise ortalamadan standart sapmadır. Özelliklerin standart sapması 1 ile 0 civarında ortalananak şekilde standartlaştırılması, farklı birimleri olan ölçümleri karşılaştırdığımızda değil, aynı zamanda birçok makine öğrenme algoritması için genel bir gereklilik olarak da önemli olmaktadır.

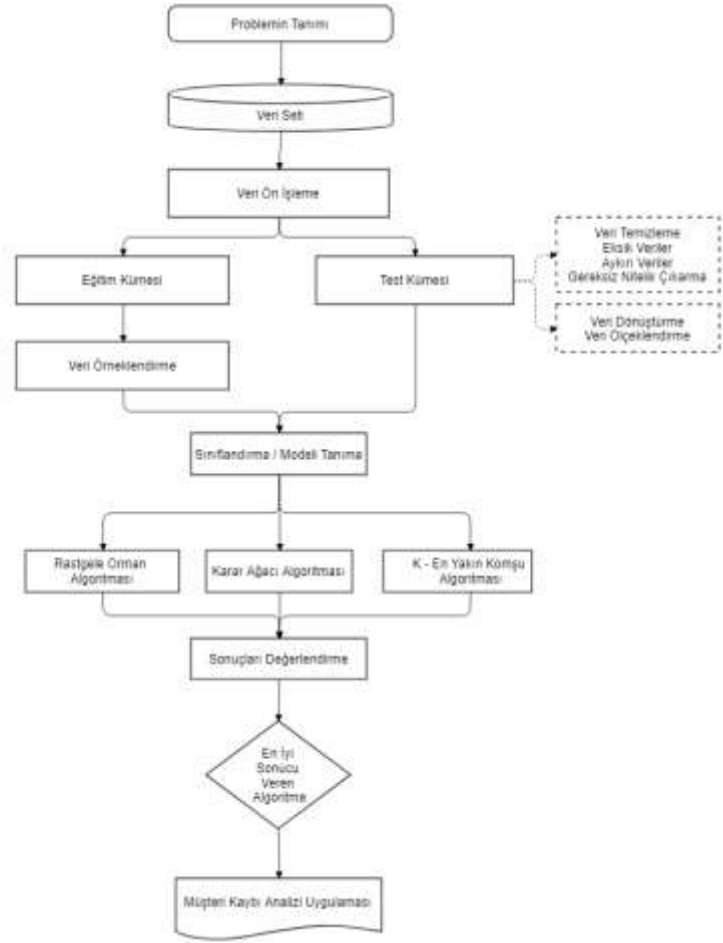
3.2.4. Eğitim Ve Test Kümesi Oluşturma

Uygulamada veri setinin %33'lük dilimi 135.112 olan satır test veri seti olarak ayrılmış, kalan 274.316 satır veri ise eğitim veri seti olarak ayrılmıştır. Veri setimizde bulunan yenilenen müşteri ve kayıp müşteri dağılımı dengesiz olmamaktadır, dengesiz olma durumunda çözümlenmek için az olan sınıfa ait veri örneği artırılabilir, fazla olan sınıfa ait veri örneği azaltılabilir veya farklı performans metriklerine göre başarı ölçümlenebilmektedir.

4. UYGULAMA

Müşteri kaybı, müşterilerin artan rekabet ortamında buldukları firmayı tercih etmekten vazgeçmeleri anlamına gelmektedir. Müşteri İlişkileri Yönetimi (MİY) ölçeklerinden olan müşteri kaybı, müşterinin son alışverişini yaptıktan sonra odağını diğer firmalara kaydırmasından bahsetmektedir. Nitekim firmalar artan rekabet ortamında yerini ve müşterilerini korumak amacıyla çeşitli stratejik kararlar alırken öncelikli olarak sadık olarak adlandırılan müşterilerinden gelecek istikrarlı gelire odaklanmaları sebebiyle var olan müşterilerin elde tutulmasına daha çok önem vermişlerdir. Eldeki müşterileri firmada tutmak, yeni bir müşteri kazanmak ve onu sadık müşteri yapmaktan daha maliyetli olmaktadır. Nitekim müşteri kaybı yönetimi adı verilen süreçte istenen, firmayı tercih etmeyi bırakma ihtimali olan müşterileri tespit etmektir. Bu sebeple müşteri kaybının doğru yönetilmesi ancak firmadan vazgeçme ihtimali bulunan müşterinin doğru şekilde belirlenmesi ile mümkün olmaktadır. Bu noktada temel amaç, müşterileri kayıp olacak ve kayıp olmayacak müşteri şeklinde sınıflandırmaktır.

Uygulamaya ait akış şeması Çizelge 1.' de gösterilmektedir.



Çizelge 1. Çalışmanın Akış Diyagramı

4.1. Veri Seti Hazırlaması

Artan rekabet dünyasında mevcut müşteri portföyünü korumak, rakip firmaların stratejisine yenik düşmeden müşterilerin gitme ihtimalini önceden tespit edip, bu durumun önüne geçmek için gitme ihtimali olan müşterilerin gitme sebeplerini çözümlenip müşterileri memnun etmek gerekmektedir. Bu çalışma belirlenen bu problem üzerine yapılmıştır.

Müşteri kaybı analizi yapılan bu çalışmada özel bir sigortacılık şirketinin verileri kullanılarak 2017 ve 2020 yılları arasında yer almış müşterilere ait içerisinde sosyodemografik bilgilerin yanı sıra kullanılan araç marka, model bilgilerinin de içinde yer aldığı 23 öznitelik belirlenmiştir. Veri seti Oracle veri tabanında bulunan tablolardan PLSQL veri tabanı sorgusu kullanılarak hazırlanıp csv formatında export alınmıştır. Python programlama dilinde bulunan veri okuma, veri ön işleme ve veri temizleme aşamalarının yapılmasına olanak sağlayan pandas kütüphanesi ile hazırlanan veri seti proje içerisine alınmıştır.

4.2. Veri Ön İşleme

Öznitelikler belirlenirken Plaka, TCKN , VKN gibi müşteriye özel olan bilgilerin eğitim modeline faydası olmayacağı gerekçesiyle çalışmaya dâhil edilmemiştir. Bunun yanı sıra müşteri özelinde poliçesinin iptal olup olmadığı bilgisi ve yenileme sayısı gibi öznitelikler de modelde ezbere sebebiyet verdiği için yer almamıştır.

Öznitelikler belirlendikten sonra kullanılan arayüz'e (Spyder) dâhil edilen veri setinde veri ön işleme aşamalarına başlanmıştır. İlk olarak veri setindeki sütunların satır bazında boş adetlerinin genel adede oranı alınmıştır. Bu oran boş verilerin analizini yapmak için kullanılmıştır. Sütun bazında %1'in altında boş olan veriler silinmiş, %1'in üstünde boş olan sütunlarda ise kategorik değerler var ise sütunun modu, nümerik veriler var ise ortalama değerler ile doldurulmuştur.

Veri seti üzerindeki boş olan verileri doldurma yöntemleri ile boş veriler temizlendikten sonra modelin başarı oranını arttırmak için kategorik değişkenleri işlemek için kullanılan kodlama yöntemi olan Label Endocing ile string türündeki sütunlardaki her bir veriye alfabetik sıralamaya göre benzersiz bir tam sayı atanmıştır.

Tüm sütunlardaki verilerin sayısal olduğuna emin olduktan sonra veri seti üzerinde aykırı değerlerin tespiti için z score puanları hesaplanmıştır. Z score'ün -3 ile 3 arasında olması gerekmektedir. -3'ün altındaki ve 3'ün üstündeki değerler aykırı değerler olacağı için veri setinden çıkarılmıştır.

Veri işleme adımlarına devam ederken veri setinde bulunan sütunların birbirleri ile aralarındaki ilişkilerin yönü ve derecesini incelemek için Korelasyon Analizi uygulanmıştır. Bu analiz sonucu belirlenen korelasyon katsayısı sütunlar arasındaki ilişkinin bağımlı değişken mi yoksa bağımsız değişken mi olduğunu göstermektedir. Korelasyon katsayısı -1 ile 0 arasında çıkan sütunlar arasında bağımsız değişken, 0 ile 1 arasında çıkan sütunlar arasında ise bağımlı değişken ilişkisinin olduğu bilgisine erişilmiştir. Bu analiz sonucunda korelasyon katsayısı 0 olan verilerin ilişkileri kurulamadığı, modelin öğrenimine bir fayda sağlamayacağı gerekçesiyle temizlenmelidir.

Tüm veri setinin kesirli sayılara (Float) çevrildiğinden emin olduktan sonra verilerin dağılımı ve özellikler arasındaki ölçek farklılıkları sebebiyle veri setine standart scaler uygulanıp veri seti veri temizleme aşamasını tamamlamıştır.

4.3. Veri Setini Eğitim ve Test Olarak Ayırma

Veri temizleme aşaması tamamlanan veri setinden rastgele seçilen %33'lük dilimi 135.112 olan satır test veri seti olarak ayrılmış, kalan 274.316 satır veri ise eğitim veri seti olarak ayrılmıştır. Modellerin karşılaştırılması sırasında veri setinden doğacak üstünlüklerin engellenmesi sebebiyle test ve eğitim veri setlerinin bölme işlemi bir kez yapılmış ve her model için aynı eğitim ve test verileri kullanılmıştır.

4.4. Veri Setine Algoritmaların Uygulanması

Veri setinden rastgele seçilen %67'lik kısım, 274.316 satır eğitim verisi, çalışmalar dâhilinde belirlenen algoritmalar ile modele dahil edilip, sonuç test için ayrılan veri seti ile karşılaştırılıp değerlendirilmiştir.

4.4.1. Karar Ağacı (Desicion Tree) Algoritmasının Uygulanması ve Sonuçları

Eğitim olarak ayrılan 274.316 satır veriye karar ağacı algoritması uygulanıp müşterilerin kayıp olup olmayacağı tahmin edilmiştir, test veri kümesiyle ise bu algoritmanın doğruluğu ölçülmüştür.

	Öngörülen	
	S=0	S=1
Gerçek Sınıf	S=0 43629	11048
	S=1 11025	69410

Tablo 2. Karar Ağacı (Desicion Tree) Algoritması İle Oluşturulan Modelin Karmaşıklık Matrisi

Tablo 2.'de gösterilen, model sonucu oluşturulan karmaşıklık matrisi incelendiğinde tabloda DN=43629 DP=69410, YN= 11048 YP=11025 olarak görülmektedir. Doğru olarak sınıflanan ayrılmayan müşteri sayısı 43.629 ayrılan müşteri sayısı 69.410'dur. Toplamda 113.039 müşteri doğru sınıflandırılmıştır. 11.048 müşterinin gerçekte ayrılmış olup sınıflandırma sonucunda ayrılmamış olarak ve 11.025 müşterinin gerçekte ayrılmamış olup sınıflandırma sonucunda ayrılmış olarak etiketlendiğini görmekteyiz. Bu modelin genel başarı oranı %84'tür.

4.4.2. Veri Setine Rastgele Orman(Random Forest) Algoritmasının Uygulanması

Eğitim olarak ayrılan 274.316 satır veriye Rastgele Orman (Random Forest) algoritması uygulanıp müşterilerin kayıp olup olmayacağı tahmin edilmiştir, test veri kümesiyle ise bu algoritmanın doğruluğu ölçülmüştür.

Aktör analizi değerlendirmelerine katılan yedi uzmanın tamamı lisans ve üzeri eğitime sahip olup %71,42'si 18-25 yaş arasındaki genç girişimcilerdir. Katılımcıların %85,71'inin özgirişim tecrübesi bulunmaktadır.

Tablo 3. Karar Ağacı Algoritması İle Oluşturulan Modelin Karmaşıklık Matrisi

		Öngörülen	
		S=0	S=1
Gerçek Sınıf	S=0	45839	8838
	S=1	8665	71770

Tablo 3'te gösterilen, model sonucu oluşturulan karmaşıklık matrisi incelendiğinde tabloda DN=45839 DP=71770, YN= 8838 YP=8665 olarak görülmektedir. Doğru olarak sınıflanan ayrılmayan müşteri sayısı 45.839 ayrılan müşteri sayısı 71.770'dur. Toplamda 117.609 müşteri doğru sınıflandırılmıştır. 8.838 müşterinin gerçekte ayrılmış olup sınıflandırma sonucunda ayrılmamış olarak ve 8.665 müşterinin gerçekte ayrılmamış olup sınıflandırma sonucunda ayrılmış olarak etiketlendiğini görmekteyiz. Bu modelin genel başarı oranı %87'dir.

4.4.3. Veri Setine K-En Yakın Komşu (K-Neighbors) Algoritmasının Uygulanması

Eğitim olarak ayrılan 274.316 satır veriye k-en yakın komşu algoritması uygulanıp müşterilerin kayıp olup olmayacağı tahmin edilmiştir, test veri kümesiyle ise bu algoritmanın doğruluğu ölçülmüştür.

Tablo 4. Karar Ağacı Algoritması İle Oluşturulan Modelin Karmaşıklık Matrisi

		Öngörülen	
		S=0	S=1
Gerçek Sınıf	S=0	38325	16352
	S=1	16959	63476

Tablo 4'te gösterilen, model sonucu oluşturulan karmaşıklık matrisi incelendiğinde tabloda DN=38.325 DP=63.476, YN= 16.352 YP=16.959 olarak görülmektedir. Doğru olarak sınıflanan ayrılmayan müşteri sayısı 38.325 ayrılan müşteri sayısı 63.476'dır. Toplamda 101.801 müşteri doğru sınıflandırılmıştır. 16.352 müşterinin gerçekte ayrılmış olup sınıflandırma sonucunda ayrılmamış olarak ve 16.959 müşterinin gerçekte ayrılmamış olup sınıflandırma sonucunda ayrılmış olarak etiketlendiğini görmekteyiz. Bu modelin genel başarı oranı %75'tir.

4.5. Müşteri Kayıp Analizi Uygulaması

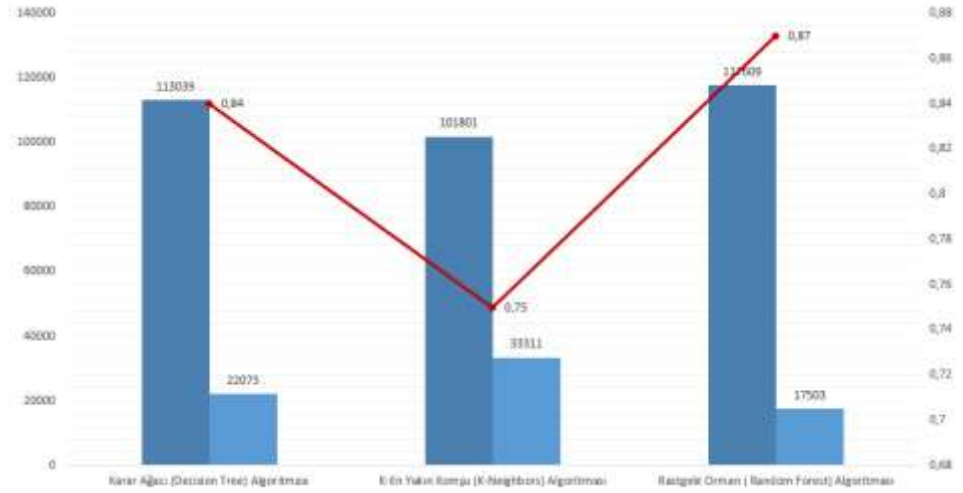
Müşteri bilgilerini ve davranışlarını inceleyerek, bu müşteriler arasından ayrılma ihtimali olan müşterilere ait özellikleri çıkarıp terk etme olasılığı yüksek olan müşterileri önceden tahmin etme amacını devam ettirmek için tüm bu adımlar bir sınıfa dâhil edilmiştir. Müşteri Kaybı Analizi Uygulaması önyüzündeki hesapla butonuna basıldığında bu sınıf çağrılarak içerisindeki en yüksek başarı oranını veren karar ağacı algoritması ile müşterilerin terk edip etmeyeceği analiz edilip terk etme oranı yansıtılmaktadır. Bu oran doğrultusunda belirlenen kural çerçevesinde müşterinin aranıp aranmama durum bilgisi ekrana yansıtılmaktadır.

5. SONUÇ

Günümüz şartlarında şirketler arası rekabet ortamının artması, pazarlama stratejilerinin gelişmesi ve müşterilerin daha bilinçli hâle gelmesi ile müşteri sadakati önem kazanmıştır. Bir şirketin sektörde varlığını devam ettirebilmesi için mevcut müşterilerini elde tutması ve yeni müşteriler edinmesi oldukça önemlidir. Sigortacılık sektöründe yeni bir müşteri

edinmek, mevcuttaki müşterinin ayrılmasını önlemekten çok fazla maliyetlidir. Mevcut müşterilerin profilleri ve davranışları incelenerek şirketi bırakma ihtimali olan riskli müşterileri bulma işlemine müşteri kaybı analizi denir. Bu sayede şirketler rekabet açısından üstünlük kazanır ve çeşitli stratejiler geliştirerek bu kayıpların önüne geçmeye çalışır.

Bu çalışmada sigortacılık sektöründe, aldığı sigorta poliçesini iptal etme olasılığı olan müşterileri, makine öğrenmesi yöntemlerinden olan 3 farklı sınıflandırma yöntemi ile tespit edilmeye çalışılmıştır. Bu 3 farklı sınıflandırma yöntemlerinden en yüksek başarı oranını %87 oranla veren Rastgele Orman (Random Forest) Algoritması, peşinde %84 başarı oranı veren Karar Ağacı (Decision Tree) Algoritması ve %80 başarı oranı veren K-En Yakın Komşu (K Neighbors) Algoritması olduğu gözlemlenmiştir. Şekil 6.'da bu üç algoritmanın doğru tahmin ettiği müşteri sayıları ve bu tahminlerin başarı oranları verilmektedir.



Şekil 6. Uygulanan Algoritmaların Karşılaştırması

%87 doğruluk oranını veren Rastgele Orman (Random Forest) Algoritması ile aynı makine öğrenmesi modeli bir sınıfa dâhil edilmiştir. Python programlama dili kullanılarak yapılan önyüze tahmin edilecek müşterilerin tutulduğu dosya dizini ismi girilerek tahmin et butonuna basılmasının ardından yapılan makine öğrenmesi sınıfı tetiklenerek, excelde bulunan müşteriler sınıfta öğretilen model yardımıyla kayıp olma ihtimali hesaplanıp, bu oran doğrultusunda müşterinin aranıp aranmayacağı bilgisi ekrana yazdırılmaktadır.

Yapılan çalışmamızın en kritik ve önemli süreci modellerde kullanılacak verinin kaynak sistemlerden üretilip bu üretilen verinin veri işleme yöntemleri ile temizlenerek modele uygun hâle getirilmesidir. Çalışmadaki tüm süreçler tamamen ücretsiz Spyder programı üzerinde Python programlama dili ile Sklearn kütüphanesi kullanılarak yapılmıştır. Makine öğrenmesi sürecinin ardından oluşturulan arayüz de yine python programlama dili ile oluşturulmuştur. Python programlama dili ve kütüphanelerinin makine öğrenmesi uygulamalarında başarılı sonuç verdiği gözlemlenmiştir.

Çalışmada en başarılı olan algoritmanın hesaplanması gibi manuel kalmış tüm süreçler otomatikleştirilip son kullanıcının tetiklediği sınıfa o başarı oranı ile devam edilmesi bir sonraki çalışmalara örnek olarak gösterilmektedir. Ek olarak uygulamanın arayüzünün tasarımı kullanıcıların veya şirketlerin tercihine göre geliştirilebilmektedir.

KAYNAKLAR

- Aydoğan E., G. A. (2009). Veri Madenciliği Teknikleri İle Bir Kozmetik Markanın Ayrılan Müşteri Analizi Ve Müşteri Bölümlemesi. *Mühendislik ve Fen Bilimleri Dergisi*.
- Arthur L. Samuel (1959). Some Studies in Machine Learning Using the Game of Checkers. II—Recent Progress, pp 366-400
- Baykal, A., & Coşkun, C. (2011). Veri madenciliğinde sınıflandırma algoritmalarının bir örnek üzerinde karşılaştırılması. *Akademik Bilişim Konferansı*.
- Bagheri F, T. M. (2015). Customer behavior mining based on RFM model. *Journal of Industrial Engineering and Management Studies*.
- Burez, J. V. (2009). Handling class imbalance in customer churn. *Elsevier*, 4626-4636.
- Buttle, F. M. (2015). Customer Relationship Management. *Concepts and Technologies*.
- Huigevoort, C. (2015). Customer churn prediction for an insurance company. *Eindhoven Teknoloji Üniversitesi*.
- Karaağaç, Ş. (2015). Churn Analysis And Churn Prediction In A Private Bank. *Marmara Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi*.
- Kaynar, O., Tuna, M., Görmez, Y., & Deveci, M. (2017). Makine Öğrenmesi Yöntemleriyle Müşteri Kaybı Analizi. *C. Ü. İktisadi ve İdari Bilimler Dergisi*.
- Kim, K., Jun, C., & Lee, J. (2014). Improved churn prediction in telecommunication industry by analyzing a large network. *Elsevier*.
- Koçtürk, Y. (2010). Veri Madenciliğine Bağlılık. *İstanbul Teknik Üniversitesi, Yüksek Lisans Tezi*.
- Kotler, P., & Keller, K. (2015). *Marketing Management*. 5. Baskı.
- Kunt, M. (2019). Telekomünikasyon Sektöründe Müşteri Kaybı Analizi. *Ankara Üniversitesi Fen Bilimleri Enstitüsü*.
- Nettleton, D. (2014). *Commercial Data Mining. Analysis and Modeling for Predictive Analytics Projects*.
- Özmen, M. (2006). Churn Modelling In Telecommunications Sector. *İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü*.
- Ravi, V., Raju, S., & Farquad, M. (2014). Churn prediction using Comprehensible support vector machine: An analytical CRM application. *Applied Soft Computing*, 31-40.
- Scheel, I., Aldrin, M., Glad, I., Sorum, R., Lyng, H., & Frigessi, A. (2005). The influence of missing value imputation on detection of differentially expressed genes from microarray data. *Bionformatics*, 4272-4279.
- Sharma, A., & Panigrahi, D. (2011). A Neural Network Based Approach for Predicting Customer Churn in Cellular Network Services. *International Journal of Computer Applications*.
- Tosun, T. (2006). Veri Madenciliği Teknikleriyle Kredi Kartlarında Müşteri Kaybetme Analizi. *İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi*.
- Tsiptsis, K., & Chorianopoulos, A. (2009). *Data Mining Techniques in CRM: Inside Customer Segmentation*. John Wiley & Sons.
- Vafeiadis, T., Diamantaras, K., Sarigiannidis, G., & Chatzisavvas, K. (2015). A comparison of machine learning techniques for customer churn prediction. *Elsevier*.
- Yılmaz, H. (2014). Random Forests Yönteminde Kayıp Veri Probleminin İncelenmesi ve Sağlık Alanında Bir Uygulama.