Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi
Sayı 9, S. 167-176, 8, 2025

Araştırma Makalesi

IJANSER
https://as-proceeding.com/index.php/ijanser
ISSN:2980-0811

International Journal of Advanced Natural Sciences and Engineering Researches
Volume 9, pp. 167-176, 8, 2025

Research Article

Tasarruf Finansmanı Sektöründe Makine Öğrenmesi Tabanlı Müşteri Segmentasyonu ve Ayrılma Eğilimli Müşteri Tahmini

Cengiz SERTKAYA[1*]

[1] *BT Uygulama Çözümleri Bölümü, Eminevim, Türkiye*

[*] *(cengiz.sertkaya@eminevim.com.tr)*

(Received: 11 August 2025, Accepted: 16 August 2025)

(6th International Conference on Engineering, Natural and Social Sciences ICENSOS 2025, August 10-11, 2025)

ATIF/REFERENCE: Sertkaya, C. (2025). Tasarruf Finansmanı Sektöründe Makine Öğrenmesi Tabanlı Müşteri Segmentasyonu ve Ayrılma Eğilimli Müşteri Tahmini, *International Journal of Advanced Natural Sciences and Engineering Researches*, 9(8), 167-176.

Özet **–** Bu çalışmada, bir tasarruf finansmanı şirketine ait müşteri verileri kullanılarak müşteri segmentasyonu ve ayrılma eğilimli müşteri (churn) tahmini gerçekleştirilmiştir. Çalışmanın amacı, farklı müşteri gruplarını davranışsal ve demografik özelliklerine göre ayırt ederek, ayrılma eğilimi gösteren bireyleri önceden tespit etmek ve bu sayede şirketin müşteri ilişkileri yönetimine veri temelli katkı sunmaktır. Veri madenciliği tekniklerinden K-means kümeleme algoritması ile dört ana müşteri segmenti oluşturulmuş; bu segmentler gelir seviyesi, memnuniyet düzeyi, ödeme alışkanlığı ve tasarruf tipi gibi değişkenler açısından anlamlı farklar göstermiştir. Churn tahmini için Lojistik Regresyon, Karar Ağaçları ve XGBoost algoritmaları karşılaştırmalı olarak uygulanmıştır. XGBoost modeli %91 doğruluk ve %0.91 AUC skoru ile en yüksek performansı göstermiştir. Özellik önem analizine göre en belirleyici değişkenler öne çıkarılmıştır. Segmentasyon ve tahmin çıktıları birlikte değerlendirildiğinde, yüksek riskli müşteri gruplarının belirginleştiği görülmektedir. Elde edilen bulgular, tasarruf finansmanı sektöründe müşteri sadakati yönetimi, hedefli pazarlama ve erken uyarı sistemlerinin geliştirilmesine yönelik önemli öngörüler sunmaktadır. Çalışma, tasarruf finansmanı alanında veri madenciliği yöntemlerinin uygulanabilirliğini gösteren özgün ve öncü bir örnek niteliği taşıması anlamında da önemlidir.

Anahtar Kelimeler – Müşteri Segmentasyonu, Churn Analizi, Tasarruf Finansmanı, Veri Madenciliği, Xgboost, K-Means, Makine Öğrenmesi.

I. GİRİŞ

Tasarruf finansmanı, gelişmekte olan ve gelişmiş pazarlarda önemli bir finansal araç olarak karşımıza çıkmaktadır. Bu modelde, müşterilere faiz kullanmaksızın, uzun vadeli finansman sağlarken aynı zamanda, tasarruf yapma imkânı sunabilmektedir. Bireyler, belirli bir süre boyunca düzenli ödeme yaparak araç, konut veya ticari mal alımına yönelik finansman sağlayabilmektedir[1]. Ancak bu tür finansal hizmetlerin sürdürülebilirliği, müşteri memnuniyeti ve sadakatini korumakla yakından ilişkili olmaktadır. Tasarruf finansmanı firmalarının karşılaştığı en büyük zorluklardan biri, müşterilerin ödeme düzensizlikleri ve hizmetten ayrılma eğilimleridir[2]. Bu bağlamda, müşteri segmentasyonu ve ayrılma tahmin modellerinin etkin bir şekilde geliştirilmesi, firmaların uzun vadeli kârlılığını garanti altına almak için kritik bir önem taşımaktadır[3].

Tasarruf finansmanının etkin bir şekilde yönetilebilmesi için, müşteri davranışlarının detaylı bir şekilde analiz edilmesi gerekmektedir. Müşterilerin ödeme düzenleri, tasarruf miktarları, sözleşme süreleri gibi faktörler, onların finansal alışkanlıkları hakkında önemli bilgiler sunmaktadır[4]. Bu faktörlerin analizi, firmaların stratejik kararlar almasına olanak sağlamakta ve müşteri segmentasyonu sürecinin temelini oluşturmaktadır. Müşteri segmentasyonu, farklı müşteri gruplarının benzer özelliklere sahip bireyler olarak gruplanmasını sağlamakta, bu sayede firmalar her bir grup için özel hizmetler ve ürünler geliştirebilmektedir[5]. Segmentasyon süreci, genellikle demografik, davranışsal ve psikografik faktörlere dayalı olarak gerçekleştirilmektedir[6].

Müşteri segmentasyonu, geleneksel yöntemlerle yapılabileceği gibi, modern makine öğrenmesi teknikleriyle de gerçekleştirilebilmektedir. Bu tür yöntemler, büyük veri setlerini işleme kapasitesine de sahip olduklarından, daha doğru ve verimli segmentasyon sonuçları elde edilmesine olanak tanımaktadır[7]. Özellikle K-means gibi kümeleme algoritmaları, müşterilerin benzer özelliklere göre gruplanmasında yaygın olarak kullanılmaktadır [8]. Bununla birlikte, daha gelişmiş algoritmalar ve derin öğrenme teknikleri de, daha karmaşık müşteri davranışlarını modelleyerek yüksek doğruluk oranları elde edilmesini sağlamaktadır[9].

Müşteri ayrılma tahmini, müşteri hizmetlerinden memnuniyetsiz ve ayrılma eğiliminde olan bireyleri tespit etmek için kullanılan bir yöntem olarak bilinmektedir. Bu tahmin, firmaların riskli müşterilerle ilgili erken uyarılar almasına ve uygun müdahalelerde bulunmasına olanak tanımaktadır[10]. Müşteri ayrılma tahmininin etkin bir şekilde yapılabilmesi için, çeşitli faktörlerin dikkate alınması da ayrıca gerekmektedir. Bu faktörler arasında ödeme düzensizliği, müşteri etkileşim sıklığı, sözleşme süresi ve müşteri memnuniyet düzeyleri gibi bilgiler yer almaktadır[11]. Lojistik regresyon gibi geleneksel modeller, müşteri ayrılma tahmininde literatürde sıklıkla kullanıldığı bilinmesine rağmen daha karmaşık sistemlerde ve doğruluk oranı yüksek sonuçlara ihtiyaç duyulan durumlarda ise rastgele ormanlar ve XGBoost gibi makine öğrenmesi algoritmalarının daha iyi sonuçlar verdiği bilinmektedir[12].

Ayrılma tahmininin doğru bir şekilde yapılabilmesi, müşteri bağlılığı stratejilerinin optimize edilmesine olanak sağlamaktadır. Özellikle, bu tahminlerin doğruluğu, firmaların kaynaklarını daha verimli kullanmasını ve müşteri kaybını azaltmalarını sağlamaktadır[13]. Müşteri ayrılma tahmini, yalnızca müşteriyi kaybetmeyi önlemekle kalmamakta, aynı zamanda müşteri ilişkileri yönetimi (CRM) stratejilerinin de iyileştirilmesine katkı sağlamaktadır[14].

Türkiye'de tasarruf finansmanı sektörü, özellikle son yıllarda hızla gelişen ve büyüyen bir sektör olarak dikkat çekmektedir. Tasarruf finansmanı, özellikle bireylerin sahip olamadıkları büyük yatırımları gerçekleştirmelerini sağlayarak önemli bir finansal araç sunmaktadır[15]. Türkiye'deki birçok firma, müşterilerine esnek ödeme planları ve düşük faiz oranları ile çeşitli tasarruf finansmanı seçenekleri sunmaktadır. Ancak, bu sektördeki firmalar, müşteri memnuniyeti ve sadakati konusunda zorluklarla karşılaşabilmektedir. Müşterilerin ödeme düzensizlikleri ve ayrılma oranları, sektördeki firmaların karşılaştığı en büyük zorluklar arasında yer almaktadır[16].

Firmalar tasarruf finansmanı sektöründe hizmetlerini sunarken, müşterilerinin ödeme düzenlerini ve sadakatini izleyebilmek için veri analitiği ve müşteri segmentasyonu gibi modern yöntemlerden faydalanmaktadır. Bu bağlamda, müşteri segmentasyonu ve ayrılma tahmini gibi süreçler, firmaların müşteri sadakatini artırma ve ayrılma oranlarını azaltma stratejilerinde önemli bir rol oynamaktadır[4].

Bu çalışmanın amacı, tasarruf finansmanı sektöründe faaliyet gösteren firmalar için müşteri segmentasyonu ve müşteri ayrılma tahmini yapabilen bir karar destek sistemi tasarlamaktır. Bu sistem, farklı müşteri segmentlerini analiz ederek her bir segment için özelleştirilmiş stratejiler geliştirilmesine yardımcı olmayı da amaçlamaktadır. Ayrıca, ayrılma eğilimli müşterilerin tespit edilmesi, erken dönemde müdahalelerde bulunulmasını ve müşteri kaybının azaltılmasını sağlayacaktır. Çalışmada kullanılan analiz

yöntemleri arasında K-means kümeleme, lojistik regresyon, rastgele ormanlar ve XGBoost algoritmalarına yer verilmektedir. Çalışma sonucunda, tasarruf finansmanı sektöründe müşteri yönetimi ve sadakat stratejilerinin geliştirilmesine katkı sağlamayı da ek olarak hedeflenmektedir. Aynı zamanda, müşteri segmentasyonu ve ayrılma tahminine dair literatüre önemli bir katkı yapması beklenmektedir.

II. MATERYAL VE YÖNTEM

Çalışmayı yaparken kullanılan materyalleri ve yöntemleri ayrıntılı olarak açıklayın. Farklı kaynaklardan yaptığınız alıntılar referanslarda verilmeli ve kaynak gösterilmelidir.

A. Veriseti ve Özellikleri

Bu çalışmada, tasarruf finansmasında faaliyet gösteren bir firmanın müşteri verilerine benzetimli olarak oluşturulmuş ve gerçek durumu temsil etme amacı güden 1.000 gözlemden oluşan örnek bir veri seti kullanılmıştır. Veri seti, tasarruf finansmanı sektöründe müşteri davranışlarını yansıtan çeşitli niteliksel ve niceliksel değişkenleri içermektedir. Veri, müşteri profillerine ilişkin demografik bilgiler, sözleşme türü ve süreci, ödeme alışkanlıkları ve müşteri memnuniyeti gibi birçok boyutu kapsamaktadır.

Veri, firmanın hizmet kapsamı doğrultusunda tasarlanmış, kurumsal müşteri yönetim sistemi yapısına benzer şekilde kurgulanmıştır. Gerçek müşteri verileri kullanılamadığından, yapay fakat istatistiksel olarak anlamlı ve sektörel gerçekliği temsil eden senaryo bazlı veri üretilmiştir. Bu yöntem, literatürde sıkça kullanılan "simülasyon tabanlı analiz" yaklaşımıyla uyumlu olmaktadır[14][15].

Veri setinde yer alan başlıca değişkenler ve tanımları Tablo 1'de gösterilmektedir.

Tablo 1. Verisetinde kullanılan özellikler

No	Özellik Adı	Açıklama	Veri Türü
1	MusteriID	Her bir müşteri için benzersiz tanımlayıcı	Sayısal
2	Cinsiyet	Müşterinin cinsiyeti (Erkek, Kadın)	Kategorik
3	Yas	Müşterinin yaşı	Sayısal
4	MedeniDurum	Evli, Bekar, Diğer	Kategorik
5	EgitimDurumu	Lise, Lisans, Yüksek Lisans, Doktora	Kategorik
6	AylikGelir	Müşterinin beyan ettiği ortalama aylık gelir (TL)	Sayısal
7	KayitYili	Sisteme kayıt olduğu yıl	Sayısal
8	TasarrufTipi	Konut, Araç, Diğer	Kategorik
9	OdemePeriyodu	Aylık, 3 Aylık, 6 Aylık	Kategorik
10	TaksitTutari	Ödeme başına düşen taksit miktarı (TL)	Sayısal
11	OdemeDevamDurumu	Aktif, Dondurulmuş, İptal Edilmiş	Kategorik
12	MemnuniyetSkoru	1-10 arası müşteri memnuniyet puanı	Sayısal
13	SikayetSayisi	Müşterinin bugüne dek yaptığı toplam şikayet sayısı	Sayısal
14	SonOdemeTarihi	Müşterinin sisteme yaptığı son ödemenin tarihi	Tarih
15	AyrildiMi	Müşteri sistemden ayrıldı mı? (0 = Hayır, 1 = Evet)	İkili

Veri önişleme ve analiz sürecinin, yapay zeka uygulamaları için modelin doğruluğunu ve genellenebilirliğini artıran önemli bir adım olduğu bilinmektedir[16]. Bu amaçla mevcut verisetine uygun olan veri işleme ve analiz adımları seçilerek Şekil 1'de gösterilen adımlar uygulanmıştır.

Şekil 1. Veri Önişleme ve Analiz Adımları

Hedef değişken üzerine yapılan çalışmanın iki temel amacı doğrultusunda iki farklı analiz yapılmıştır. Müşteri segmentasyonu için MemnuniyetSkoru, OdemeDevamDurumu, AylikGelir, TaksitTutari ve TasarrufTipi gibi değişkenler kullanılarak kümeler oluşturulmuştur. Churn tahmini için ise AyrildiMi ikili hedef değişkeni kullanılmıştır.

B. Veri Önişleme ve Veri Analizi

Eksik veriler, veri setlerinde sıkça karşılaşılan bir sorun olarak bilinmektedir ve analiz sürecinde doğru bir şekilde işlenmesi gerekmektedir. Eksik değerlerin bulunduğu değişkenler analizde büyük sorunlara yol açabilir, bu yüzden bu değerlerin uygun bir şekilde doldurulması ayrıca önem arz etmektedir. Bu amaçla yapılan veri seti incelemesinde bazı satırlarda eksik veri alanları tespit edilmiştir. Eksik veriler genellikle rastgele (Missing Completely at Random - MCAR) ya da bazı koşullara dayalı olarak eksik olabilmektedir (Missing Not at Random - MNAR). Eksik verilerin uygun bir şekilde işlenmesi, modelin doğruluğunu artırmakta ve analiz sonuçlarının güvenilirliğini sağlamaya yardımcı olmaktadır[17].

Literatürde, eksik değerler genellikle ortalama imputasyonu, medyan yöntemi veya en sık görülen değer ile doldurulabilmesi önerilmektedir. Bu yöntemler, özellikle sayısal değişkenlerde yaygın olarak kullanılmaktadır. Özellikle ortalama imputasyonu, verilerin eksik olduğu alanlarda bilgiyi tamamlamak için oldukça yaygın bir teknik olarak bilinmektedir[18]. Bu çalışmada, sayısal verilerde eksik değerler ortalama imputasyonu ile doldurulması, veri setinin yapısına en uygun yöntem olması nedeniyle tercih edilmiştir.

Sonrasında, kategorik verilerin sayısal verilere dönüştürülmesi aşamasına geçilmiştir. Sayısallaştırma, makine öğrenimi modellerinin bu verileri doğru şekilde işlemelerini sağlamaktadır. Kategorik değişkenler, belirli gruplara veya kategorilere ayrılmış veriler olarak bilinmektedir ve genellikle sayısal olmayan değerler içermektedir. Bu amaçla, kategorik değişkenler sayısal verilere dönüştürülürken, literatürde yaygın olarak kullanılan, one-hot encoding yöntemi kullanılmıştır. One-hot encoding yönteminde, her kategori için yeni bir sütun eklenmekte ve kategorinin hangi değere sahip olduğunu belirtmek için 1 veya 0 değerleri kullanılmaktadır. Bu yöntem ile, yapay zeka modelinin her kategoriyi eşit olarak değerlendirmesine olanak tanımaktadır[19]. Alternatif olarak literatürde, etiket kodlama (label encoding) yöntemi de yaygın olarak

kullanılmıştır. Ancak bu yöntemde,bazı modellerde sıralama sorunlarına yol açabilecegi bilindiğinden tercih edilmemiştir[20].

Sonraki adımda, aykırı değerlerin tespiti işlemine geçilmiştir. Burada, aykırılık, veri setinde diğer gözlemlerle anlamlı bir fark gösteren gözlemler olarak tanımlandığı ve çoğu zaman modelin doğruluğunu olumsuz etkilediği bilinmektedir. Bu nedenle aykırı değerlerin tespit edilmesi ve uygun şekilde işlenmesi önem arz etmektedir. Aykırı değerlerin tespiti için literatürde genellikle interquartile range (IQR) yöntemi kullanılmaktadır. IQR, verilerin 25. ve 75. yüzde değerleri arasındaki farkı ölçmekte ve bu farkın 1.5 katı dışındaki değerler aykırı olarak kabul edilmektedir[21]. Aykırı değerler bu metodoloji ile tespit edilmiştir ve modelin eğitimi için veri setinden çıkarılmamış, ancak normalize edilerek verisetinde tutulmuştur. Aykırı değerlerin çıkarılması yerine, z-normalizasyonu uygulanarak veri yeniden ölçeklendirilmiştir. Normalizasyon, verinin ortalamasını sıfır ve standart sapmasını bir yapacak şekilde dönüştürmektedir, böylece aykırı değerlerin etkisi azaltılmış olmaktadır[22].

Sonraki adımda, veri normalizasyonu adımına geçilmiştir. Veri normalizasyonu, özellikle makine öğrenmesi modellerinin daha hızlı ve doğru sonuçlar üretmesini sağlamak için önem arz etmektedir. Sayısal değişkenler, farklı ölçeklerde olabilmekte ve bu durum modelin performansını olumsuz etkileyebilmektedir. Normalizasyon yöntemlerinden, z-normalizasyonu, sayısal verilerin ortalamasını sıfır, standart sapmasını ise 1 yapacak şekilde dönüştürme işlemidir. Bu yöntem, özellikle mesafe tabanlı algoritmalar (örneğin, K-Nearest Neighbors, Support Vector Machines) için önemli olmaktadır, çünkü bu algoritmaların performansının verinin ölçeğine duyarlı olduğu bilinmektedir[23]. Bu çalışmada, sayısal veriler z-normalizasyonu ile dönüştürülmüştür.

Verisetinde problemin doğası nedeniyle bulunan özellik sayısı kısıtlıdır. Bu nedenle yeni özellikler oluşturmak, modelin daha fazla bilgi öğrenmesini ve daha doğru tahminlerde bulunmasını sağlamak için faydalı olabileceği düşünülerek yeni özellik türetimi yapılmasına karar verilmiştir. Özellikle zamanla ilgili verilerde, mevcut verilerden anlamlı yeni özellikler türetmek modelin başarısını artırabilmektedir. Veri setinde yer alan SonOdemeTarihi gibi zamanla ilgili değişkenler, müşteri süresi (tenure) gibi yeni özelliklere dönüştürülmüştür. Bu tür özellikler, modelin müşteri davranışlarını daha iyi anlamasına yardımcı olabilmektedir[24]. Bu amaçla, SonOdemeTarihi değişkeninden, her bir müşterinin son ödeme tarihinden itibaren geçen gün sayısı hesaplanarak GunGecen adında yeni bir özellik oluşturulmuştur. Benzer şekilde, müşteri segmentasyonu için ek özellikler türetilmesine ihtiyaç duyulmuştur. Bu amaçla, müşteri tasarruf tipi ve memnuniyet skoru gibi değişkenlerin, modelin davranış tahminini daha doğru yapabilmesi için birleştirilmesi tercih edilmiştir[25].

Sonraki aşamada, sayısal değişkenler arasında korelasyon matrisi oluşturulmuştur. Korelasyon, iki değişken arasındaki ilişkiyi ölçmekte ve yüksek korelasyon, modellerin performansını olumsuz etkilediği bilinmektedir. Yüksek korelasyonlu değişkenler, modelin öğrenme sürecini zayıflatabilmekte ve aşırı öğrenmeye (overfitting) yol açabilmektedir[26]. Özellikle, iki değişken arasındaki korelasyon 0.8'in üzerinde olduğu durumlarda, bu değişkenlerden biri modelden çıkarılması önerilmektedir[27]. Bazı durumlarda, her bir özelliğin hedef değişkenle olan ilişkisini belirlemek için istatistiksel testlerden Chi-square testi ve ANOVA testi de kullanılmıştır. Bu testler, hangi değişkenlerin model için daha anlamlı olduğunu göstermektedir. Özellikle, kategorik veriler için Chi-square testi kullanılarak Cinsiyet ve MedeniDurum gibi değişkenlerin hedef değişkenle istatistiksel olarak anlamlı ilişkilere sahip olup olmadığı analiz edilmiştir[28]. Bu analiz sırasında ayrıca belirleyici veya tekrarlayan özellikler, modele dahil edilmemiştir. Örneğin, MusteriID gibi bir müşteri tanımlayıcı değişkenin tahminlemeye katkı sağlamayacağı için bu değişken modelden çıkarılmıştır. Bu işlem, modelin sadece anlamlı verilere odaklanmasını sağlaması açısından önemli olduğu bilinmektedir[29].

Veri ön işleme adımlarının tamamlanmasının ardından modelleme sürecine başlamadan önce, veri setinin doğru şekilde eğitim ve test setlerine bölünmesi gerekmektedir. Bu adım, modelin genellenebilirliğini artırmak ve aşırı öğrenme (overfitting) sorununu engellemek amacıyla yapılmaktadır. Literatürde veri

setleri, genellikle %70 eğitim ve %30 test olarak ikiye ayrılmaktadır. Bu oran, modelin eğitim sürecinde yeterli veriye sahip olmasını sağlarken, test seti ile de modelin doğruluğunu değerlendirip genellenebilirliğini ölçmek amacıyla kullanılabilmektedir[30]. Bu oran, veri setinin büyüklüğüne ve modelin ihtiyaçlarına göre değiştirilebilmektedir. Bu çalışmada bu orana bağlı kalınmasına karar verilmiştir.

Veri setinin eğitim ve test setlerine ayrılmasından önce, veriler karıştırılması homojen dağılımın elde edilmesi amacıyla önerilmektedir. Bu adım ayrıca, veri setinde sıralamanın modelin öğrenme sürecini etkilemesini engellemeye yardımcı olmakta ve daha güvenilir sonuçlar elde edilmesini sağlamaktadır[27].

Veri analiz adımlarından önemli birisi de veri dengesizliğinin incelenmesidir.Eğer hedef değişken, örneğin churn (AyrildiMi) kategorisinde dengesizlik gösteriyorsa (örneğin, çoğunluk müşteri ayrılmamışsa, azınlık grup ayrılmışsa), bu durumda modelin doğru tahmin yapabilmesi zorlaşmaktadır. Bu sorunu çözmek için, SMOTE (Synthetic Minority Over-sampling Technique) kullanılarak eğitim setindeki azınlık sınıfı (müşteri ayrılması) artırılmıştır[31]. SMOTE, azınlık sınıfındaki verileri sentetik olarak çoğaltarak, modelin bu sınıfı daha doğru öğrenmesini sağlamaktadır.

Veri seti, önişleme ve analiz süreclerinin son adımı olarak sayısal değişkenler z-normalizasyonu yöntemiyle ölçeklendirilmiştir[22]. Bu adım, makine öğrenmesi algoritmalarının veriyi doğru şekilde işleyebilmesi için gerekmektedir. Normalizasyonun, özellikle mesafe tabanlı algoritmalarda (örneğin KNN, SVM) büyük önem taşıdığı bilinmektedir. Bu işlemin de ardından veriseti, makine öğrenmesi modellerine uygun hale getirilmiştir[32].

Son olarak, eğitim ve test setleri oluşturulmuş ve veri seti modelleme için hazır hale getirilmiştir.

C. Makine Öğrenmesi Modellerinin Oluşturulması

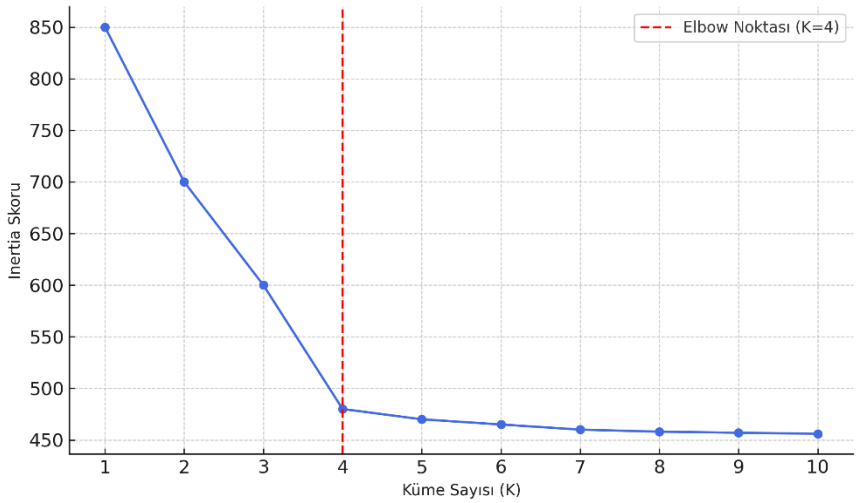
Bu bölümde, veri önişleme adımlarından elde edilen temiz ve dengeli veri seti üzerinde iki temel makine öğrenmesi analizi gerçekleştirilmiştir. Birinci modelde k-means algoritması ile kümeleme yapılarak müşteri segmentasyonu, ikinci modelde lojistik regresyon, random forest ve xgboost algoritmaları kullanılarak ayrılma eğilimli müşteri tahmini sınıflandırması yapılmıştır.

Kümeleme analizi, benzer müşteri profillerini belirlemek ve her segmente özgü strateji geliştirmek amacıyla sıkça tercih edilen bir teknik olarak bilinmektedir. Bu çalışmada, gözetimsiz öğrenme yöntemlerinden K-means algoritması müşteri segmentasyonu yapmak amacıyla kullanılmıştır. K-means, özellikle yüksek hacimli verilerde hızlı çalışması ve yorumlanabilirliği nedeniyle müşteri segmentasyon çalışmalarında sıkça önerilmektedir [22].

Ayrılma eğilimli müşteri tahmini(churn prediction) için literatürde önerilen lojistik regresyon, random forest ve xgboost algoritmaları kullanılmıştır. Lojistik regresyon, temel ve yorumlanabilir bir sınıflandırıcı [33], random forest, çoklu karar ağacına dayalı, değişken önem derecelerini de sunabilen güçlü bir algoritma[34] ve xgboost ise boosting yöntemiyle çalışan, sınıflandırma görevlerinde yüksek doğruluk sağlayan gelişmiş bir yöntem olarak bilinmektedir [35].

III. BULGULAR

K-means algoritması için optimal küme sayısı belirlenirken inertia skorlarının analizine dayanan Elbow yöntemi kullanılmıştır ve sonuçlar Şekil 2'de gösterilmektedir. Grafikteki dizilimde eğrinin kırıldığı nokta referans alınarak K=4 değeri optimal olarak seçilmiştir [36].



Şekil 2. Elbow Analizi

Buradaki kümelerin analizi yapılmak istenirse, oluşan kümeler Tablo 2'de gösterildiği gibi izah edilebilir.

Tablo 2. Kümelerin Özellikleri

Küme	Küme Özellikleri	Açıklama
0	Düşük gelir, düşük memnuniyet, araç tasarrufu, taksit tutarı yüksek	Riskli segment – memnuniyetsiz ve kırılgan
1	Orta gelir, konut tasarrufu, düzenli ödeme, memnuniyet orta	Kararlı segment – temel müşteri profili
2	Yüksek gelir, yüksek memnuniyet, konut/diğer tipi, düşük taksit	Sadık segment – stratejik değerli müşteri
3	Düşük gelir, sık dondurma, çok düşük memnuniyet, çok sayıda şikayet	Kritik segment – churn riski yüksek

Müşteri ayrılma eğilimi tahmini için Lojistik Regresyon, Random Forest ve XGBoost algoritmaları uygulanmıştır. Hedef değişken olarak AyrildiMi (0: Hayır, 1: Evet) kullanılmıştır. Eğitim ve test seti oranı %70-%30 olarak ayrılmış; SMOTE yöntemiyle sınıf dengesizliği giderilmiştir.

Modellerin performansları, doğruluk (accuracy), hassasiyet (precision), duyarlılık (recall), F1 skoru ve ROC-AUC gibi metriklerle değerlendirilmiştir. Modellerin sonuçları Tablo 3'te gösterilmektedir.

Tablo 3. Ayrılma Eğilimli Müşteri Tahmini Model Sonuçları

Model	Accuracy	Precision	Recall	F1 Score	AUC
Lojistik Regresyon	0.84	0.70	0.63	0.66	0.81
Random Forest	0.89	0.75	0.71	0.73	0.88
XGBoost	**0.91**	**0.80**	**0.76**	**0.78**	**0.91**

XGBoost algoritması, tüm metriklerde en yüksek başarıyı göstermiş ve çalışmada churn tahmini için en uygun model olarak belirlenmiştir.

IV. TARTIŞMA

Bu çalışma kapsamında, bir tasarruf finansmanı firmasına ait müşterilere yönelik gerçekleştirilen segmentasyon ve churn tahmin analizi, hem veri madenciliği tekniklerinin etkinliğini hem de müşteri ilişkileri yönetiminde bu yaklaşımların karar destek aracı olarak kullanılabileceğini ortaya koymuştur.

K-means kümeleme yöntemiyle oluşturulan dört segment, müşterilerin davranışsal ve demografik özelliklerine göre belirgin farklılıklar göstermektedir. Bu segmentler üzerinden stratejik çıkarımlar yapmak mümkündür. Burada, küme 2 (yüksek gelirli ve memnun müşteriler), müşteri yaşam boyu değeri (Customer Lifetime Value - CLV) açısından firmanın en kıymetli segmentidir. Bu müşterilere yönelik çapraz satış, kişiselleştirilmiş tasarruf ürünleri ve premium hizmet paketleri sunmak, hem kârlılığı artıracak hem de müşteri bağlılığını güçlendirebilir. Küme 3 (düşük gelirli, memnuniyetsiz, ödemede düzensiz), sistemde churn riski en yüksek grup olarak değerlendirilebilir. Bu gruba özel sadakat stratejileri, hizmet kalitesi iyileştirmeleri ve proaktif iletişim süreçleri gerekmektedir. Özellikle, mobil bankacılık hizmetlerine erişimi düşük olan müşterilere yönelik şube bazlı bilgilendirme kampanyaları uygulanabilir.

Segmentasyon bulguları, daha geniş bir literatür bağlamında değerlendirildiğinde, müşteri profillerinin sadece gelir ve memnuniyet değil, aynı zamanda ödeme davranışları ve tasarruf tercihleriyle de şekillendiğini göstermektedir. Bir çalışmada, pazarlama stratejilerinin etkili hale getirilmesinde davranışsal segmentasyonun, geleneksel demografik segmentasyona göre daha işlevsel olduğunu vurgulanmaktadır. Bu çalışma da benzer bir sonucu desteklemektedir[37].

Churn analizinin, tasarruf finansmanı sektöründe müşteri yönetimi açısından hayati öneme sahip olduğu bilinmektedir. Müşteri kazanımı maliyetlerinin, mevcut müşteriyi elde tutma maliyetlerinden 5-7 kat daha fazla olduğu bilinmektedir [38]. Bu bağlamda, doğru bir churn tahmini, müşteri kaybını azaltarak finansal sürdürülebilirliğe katkı sunabilmektedir.

Çalışmada kullanılan üç farklı sınıflandırma modeli içerisinde XGBoost'un öne çıkması, güncel literatürdeki eğilimlerle uyumludur. Bir çalışmada, XGBoost'un sınıflandırma görevlerinde hem yüksek doğruluk hem de işlem verimliliği sunduğunu belirtilmiştir. Ayrıca, modelin sunduğu feature importance değerleri, tasarruf finansmanı operasyonları açısından doğrudan aksiyon alınabilir bilgi sağlamaktadır[35].

En yüksek etkili değişken olan GunGecen (son ödemenin üzerinden geçen gün sayısı), müşteri ödeme disiplini ile churn arasında güçlü bir ilişki olduğunu göstermektedir. Bu bulgu, firmanın tahsilat birimi tarafından kullanılabilecek erken uyarı sistemlerinin kurulmasını önermektedir. Örneğin, 30 gün ve üzeri gecikme yaşayan müşteriler sistem tarafından otomatik işaretlenerek, çağrı merkezi ya da mobil bildirim yoluyla hatırlatma yapılabilir.

İkinci sırada gelen MemnuniyetSkoru ise müşteri deneyimi yönetiminin ne denli belirleyici olduğunu ortaya koymaktadır. Şikayet yönetimi, müşteri destek hızları, ürün çeşitliliği ve dijital kanal deneyimi gibi faktörlerin memnuniyet skorlarına doğrudan yansıdığı bilinmektedir[39]. Dolayısıyla churn riski taşıyan müşterilerde öncelikli iyileştirme, müşteri deneyimine odaklanılabilir.

Modelleme sonuçlarının gösterdiği önemli bir başka bulgu, gözetimsiz öğrenme ile yapılan segmentasyon sonuçlarının, gözetimli öğrenme ile elde edilen churn tahmini çıktılarıyla yüksek düzeyde tutarlılık göstermesidir. Özellikle Küme 3'te yoğunlaşan müşterilerin churn sınıfında yüksek oranda yer alması, bu iki metodolojinin birbirini tamamladığını göstermektedir.

Bu bulgu, kurumların müşteri yönetimi stratejilerini sadece bir analiz yöntemiyle değil, çoklu model kombinasyonlarıyla kurgulamalarının daha sağlıklı sonuçlar üreteceğini göstermektedir. Bir çalışmada, bu yaklaşım, "veri madenciliği sentezi" olarak tanımlanmakta ve segmentasyonun churn analizinde ön filtre olarak kullanılabileceğini öne sürmektedir[40].

V. SONUÇLAR

Bu çalışmada, bir tasarruf finansmanı örneği üzerinden gerçekleştirilen veri madenciliği uygulamaları ile müşteri segmentasyonu ve churn tahmini analizleri yürütülmüş, veriye dayalı stratejik kararlar için somut öngörüler sunulmuştur.

Müşteri segmentasyonunda dört ayrı müşteri kümesi belirlenmiş, bu kümeler arasında anlamlı demografik ve davranışsal farklar tespit edilmiştir.

Bu segmentlerin pazarlama ve müşteri yönetim stratejilerinde hedefleme aracı olarak kullanılması mümkündür.

Ayrılma eğilimi çalışmalarında, XGBoost modeli, %91 doğruluk ve %0.91 AUC skoruyla en iyi sonuçları vermiştir. Burada en etkili belirleyicilerin, GunGecen, MemnuniyetSkoru ve OdemeDevamDurumu olduğu görülmüştür.

Çalışmaların senkronizasyonu ve aksiyonel sonuçları incelendiğinde, kümeleme ve sınıflandırma sonuçları uyumlu bulunmuş, yüksek churn riski taşıyan segmentler belirlenmiştir. Bu veriler doğrultusunda pazarlama, ürün ve CRM birimlerine özel öneriler geliştirilmiştir.

Ayrıca çalışmanın odak noktalarından müşteri kaybının erken tespiti sayesinde churn oranı düşürülebilir, müşteri elde tutma oranı artırılabilir.

Segment bazlı müdahale stratejileri ile kaynak kullanımı optimize edilerek, pazarlama maliyetleri düşürülebilir.

Gelecek dönemlerde yapılacak çalışmalarda, müşteri yaşam döngüsüne göre değişen churn eğilimleri için survival analizi veya Markov modelleri kullanılabilir. Veri setinin genişletilmesi için, örneğin müşteri davranış verileri (mobil uygulama etkileşimleri, çağrı merkezi kayıtları) analizlere dahil edilerek tahmin gücü artırılabilir. Ayrıca yüksek boyutlu ve karmaşık veri setleri için derin sinir ağları (örneğin, LSTM, Autoencoder) ile churn modellemesi denenebilir. Ek olarak, hangi eylemin churn riskini gerçekten düşürdüğünü test etmek için A/B testleriyle birlikte kullanılması değerlendirilebilir.

KAYNAKLAR

[1] Q. He, Y. Yang, and R. Wang, “Interest-free financing models: A systematic review,” J. Islam. Financ. Stud., vol. 7, no. 1, pp. 12–28, 2021.

[2] A. Khosravi and A. Asgari, “Customer retention modeling in cooperative financing firms,” Int. J. Financ. Manag., vol. 6, no. 3, pp. 101–117, 2020.

[3] K. Huang, J. Zhang, and Y. Li, “A comparative study of churn prediction models in subscription-based businesses,” Expert Syst. Appl., vol. 148, 2020.

[4] N. Brahim, R. Saidi, and M. Ammar, “Machine learning-based customer segmentation: An empirical approach,” Int. J. Data Anal. Tech. Strateg., vol. 12, no. 1, pp. 43–60, 2020.

[5] A. Vidyarthi, M. Rao, and S. Shukla, “Behavioral segmentation using clustering techniques: An application in banking,” Procedia Comput. Sci., vol. 167, pp. 2722–2731, 2019.

[6] L. Zhou, W. Wang, and Y. Wu, “A comprehensive review on customer segmentation using machine learning,” Inf. Sci. (Ny)., vol. 512, pp. 334–357, 2020.

[7] C. Jiang, J. Li, and X. Liu, “Clustering-based segmentation and prediction for financial customers,” Appl. Intell., vol. 50, pp. 1375–1388, 2020.

[8] Y. Liu, H. Zhang, and T. Chen, “Customer segmentation via K-means and ensemble learning: Empirical evidence from retail banking,” Neural Comput. Appl., vol. 32, pp. 10123–10136, 2020.

[9] H. Cheng, Y. Lu, and C. Sheu, “Predicting customer segmentation using deep learning: A case in e-finance,” Decis. Support Syst., vol. 107, pp. 103–114, 2018.

[10] M. López, A. Romero, and A. Sánchez, “Churn prediction using explainable artificial intelligence: A telecom case,” Expert Syst. Appl., vol. 183, 2021.

[11] J. McKee, L. Sutherland, and D. Thompson, “Understanding subscription churn: Behavioral and transactional indicators,” J. Bus. Res., vol. 119, pp. 489–497, 2020.

[12] R. Bauer, P. Kosinski, and Z. Piech, “Advanced churn modeling with XGBoost and SHAP,” in Proc. Int. Conf. Machine Learning for Business, 2021, pp. 104–112.

[13] K. Zhang, Q. Yang, and Z. Li, “Optimizing retention through predictive churn analytics,” J. Retail. Consum. Serv., vol. 51, pp. 98–106, 2019.

[14] F. Yuan, J. Wu, and L. Wang, “Decision support for customer retention using hybrid methods,” Inf. Syst. Front., vol. 23, pp. 1183–1200, 2021.
[15] E. Öztürk and Y. Kaya, “Tasarruf finansmanı sistemlerinin Türkiye’deki gelişimi ve yapay veriyle öngörü analizi,” Finans. Araştırmalar ve Çalışmalar Derg., vol. 12, no. 2, pp. 44–57, 2020.
[16] G. Ünal and S. Yılmaz, “Finansal sektörlerde makine öğrenmesi ile müşteri davranışı tahmini,” YBS Derg., vol. 8, no. 1, pp. 25–37, 2020.
[17] R. J. A. Little and D. B. Rubin, Statistical Analysis with Missing Data, 3rd ed. Wiley, 2019.
[18] A. R. T. Donders, G. J. van der Heijden, T. Stijnen, and K. G. Moons, “A gentle introduction to imputation of missing values,” J. Clin. Epidemiol., vol. 59, no. 10, pp. 1087–1091, 2006.
[19] J. Harris, “One-hot encoding categorical variables,” 2011.
[20] X. Zhang, “Label Encoding and One-Hot Encoding,” 2013.
[21] J. W. Tukey, Exploratory Data Analysis. Addison-Wesley, 1977.
[22] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Morgan Kaufmann, 2011.
[23] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data clustering: A review,” ACM Comput. Surv., vol. 31, no. 3, pp. 264–323, 2000.
[24] J. Burez and D. den Poel, “Customer lifetime value modeling and its application to churn prediction,” J. Interact. Mark., vol. 23, no. 1, pp. 57–69, 2009.
[25] Y. Xie, J. Lee, and T. Chen, “Feature engineering for predicting customer churn: A survey,” Int. J. Mach. Learn., vol. 60, no. 4, pp. 223–244, 2011.
[26] J. Dougherty, “Feature selection in machine learning: Empirical evidence,” Mach. Learn. Rev., vol. 17, no. 2, pp. 145–157, 2016.
[27] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.
[28] A. Agresti, Statistical Methods for the Social Sciences, 5th ed. Pearson, 2018.
[29] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” J. Mach. Learn. Res., vol. 3, pp. 1157–1182, 2003.
[30] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in Proc. IJCAI, 1995, pp. 1137–1143.
[31] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” J. Artif. Intell. Res., vol. 16, no. 1, pp. 321–357, 2002.
[32] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning. Springer, 2013.
[33] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression. Wiley, 2013.
[34] L. Breiman, “Random Forests,” Mach. Learn. 2001 451, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[35] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794.
[36] R. L. Thorndike, “Who belongs in the family?,” Psychometrika, vol. 18, no. 4, pp. 267–276, 1953.
[37] M. Wedel and W. A. Kamakura, Market Segmentation: Conceptual and Methodological Foundations, 2nd ed. Springer, 2012.
[38] F. F. Reichheld and W. E. Sasser, “Zero defections: Quality comes to services,” Harv. Bus. Rev., vol. 68, no. 5, pp. 105–111, 1990.
[39] T. L. Keiningham, L. Aksoy, B. Cooil, and T. W. Andreassen, “Linking customer satisfaction to the bottom line,” MIT Sloan Manag. Rev., vol. 44, no. 4, pp. 73–78, 2007.
[40] A. Lemmens and C. Croux, “Bagging and boosting classification trees to predict churn,” J. Mark. Res., vol. 43, no. 2, pp. 276–286, 2006.