

HİTİT SOSYAL BİLİMLER DERGİSİ

Hitit Journal of Social Sciences

e-ISSN: 2757-7949

Cilt | Volume: 18 • Sayı | Number: 2

Ağustos | August 2025

E-Ticarete Kullanıcılarının Davranışsal Yaklaşımlarının Makine Öğrenmesi Yöntemleri ile Sınıflandırılması ve Müşteri Segmentasyonu

Classification of Users' Behavioral Approaches in E-Commerce and Customer Segmentation Using Machine Learning Methods

Serkan METİN

Corresponding Author | Sorumlu Yazar

Dr. Öğr. Üyesi | Asst. Prof.

Turgut Özal Üniversitesi, Sosyal ve Beşerî Bilimler Fakültesi,

Yönetim Bilişim Sistemleri Bölümü, Malatya, Türkiye

Turgut Özal University, Faculty of Social and Human Sciences,

Department of Management Information Systems, Malatya, Türkiye

serkan.metin@ozal.edu.tr

<https://orcid.org/0000-0003-1765-7474>

<https://ror.org/03r7b1f79>

Makale Bilgisi | Article Information

Makale Türü | Article Type: Araştırma Makalesi | Research Article

Geliş Tarihi | Received: 11.02.2025

Kabul Tarihi | Accepted: 15.08.2025

Yayın Tarihi | Published: 31.08.2025

Atıf | Cite As

Metin, S. (2025). E-Ticarete kullanıcılarının davranışsal yaklaşımlarının makine öğrenmesi yöntemleri ile sınıflandırılması ve müşteri segmentasyonu. *Hitit Sosyal Bilimler Dergisi*, 18(2), 314-329. <https://doi.org/10.17218/hititsbd.1637810>

Değerlendirme: Bu makalenin ön incelemesi iki iç hakem (editörler - yayın kurulu üyeleri) içerik incelemesi ise iki dış hakem tarafından çift taraflı kör hakemlik modeliyle incelendi. Benzerlik taraması yapılarak (Turnitin) intihal içermediği teyit edildi.

Etik Beyan: Bu çalışmanın hazırlanma sürecinde bilimsel ve etik ilkelere uyulduğu ve yararlanılan tüm çalışmaların kaynakçada belirtildiği beyan olunur.

Etik Bildirim: husbededitor@hitit.edu.tr
<https://dergipark.org.tr/tr/pub/hititsbd>

Çıkar Çatışması: Çıkar çatışması beyan edilmemiştir.

Finansman: Bu araştırmayı desteklemek için dış fon kullanılmamıştır.

Telif Hakkı & Lisans: Yazarlar dergide yayınlanan çalışmalarının telif hakkına sahiptirler ve çalışmaları CC BY-NC 4.0 lisansı altında yayımlanmaktadır.

Review: Single anonymized - Two Internal (Editorial board members) and Double anonymized - Two External Double-blind Peer Review

It was confirmed that it did not contain plagiarism by similarity scanning (Turnitin).

Ethical Statement: It is declared that scientific and ethical principles have been followed while conducting and writing this study and that all the sources used have been properly cited.

Complaints: husbededitor@hitit.edu.tr
<https://dergipark.org.tr/tr/pub/hititsbd>

Conflicts of Interest: The author(s) has no conflict of interest to declare.

Grant Support: The author(s) acknowledge that they received no external funding to support this research.

Copyright & License: Authors publishing with the journal retain the copyright to their work licensed under the CC BY-NC 4.0.

Classification of Users' Behavioral Approaches in E-Commerce and Customer Segmentation Using Machine Learning Methods

Abstract

The rapid growth of e-commerce and increasing consumer expectations have intensified the need for efficient customer segmentation methods. Traditional segmentation techniques, which often rely on manual and rule-based approaches, fail to address the complexity and scalability required in modern e-commerce. Machine learning algorithms provide a powerful alternative, offering data-driven insights that enable personalized marketing strategies and targeted customer engagement. This study aims to analyze and compare the effectiveness of K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest algorithms in the classification of customer segments in e-commerce. Existing literature suggests various approaches for customer segmentation, but a direct comparison of these machine learning models in an e-commerce context remains limited. This research seeks to fill this gap by evaluating the classification performance of these models and determining their suitability for segmenting online shoppers. The dataset utilized in this study was obtained from Kaggle and contains consumer behavior-related features, including shopping frequency, promotional engagement, spending habits, and security concerns. To determine the optimal number of clusters, Elbow and Silhouette methods were employed. Once the clustering structure was established, the K-Means algorithm was applied to segment customers into distinct groups based on their behavioral patterns. These clusters were then classified using KNN, SVM, and Random Forest, and their performance was assessed using standard classification metrics such as Accuracy, Precision, Recall, and F1-Score. The results revealed that SVM achieved the highest accuracy (95%), making it the most reliable model for customer segmentation. Random Forest closely followed with an accuracy of 93%, demonstrating strong performance while offering higher scalability and lower computational costs. KNN, on the other hand, showed the lowest accuracy (78%), indicating that it may not be the most suitable choice for high-dimensional or large datasets. In addition, SVM and Random Forest consistently outperformed KNN in terms of Precision, Recall, and F1-Score, confirming their superior classification capabilities. In conclusion, this study highlights the importance of selecting the right machine learning model for customer segmentation in e-commerce. SVM is recommended for applications requiring the highest classification accuracy, while Random Forest is better suited for large datasets with high scalability demands. KNN, despite its simplicity, is more appropriate for small-scale segmentation tasks. The findings provide valuable insights for e-commerce businesses seeking to optimize their segmentation strategies through machine learning, ultimately enhancing customer targeting, engagement, and retention.

Keywords: Customer Segmentation, E-Commerce, SVM, Random Forest, KNN

E-Ticarette Kullanıcılarının Davranışsal Yaklaşımlarının Makine Öğrenmesi Yöntemleri ile Sınıflandırılması ve Müşteri Segmentasyonu

Öz

E-ticaretin hızla büyümesi ve artan tüketici beklentileri, etkili müşteri segmentasyonu yöntemlerine olan ihtiyacı artırmıştır. Geleneksel segmentasyon teknikleri genellikle manuel ve kural tabanlı yaklaşımlara dayandığından, modern e-ticaretin gerektirdiği karmaşıklık ve ölçeklenebilirliği sağlamada yetersiz kalmaktadır. Makine öğrenmesi algoritmaları, veriye dayalı içgörüler sunarak kişiselleştirilmiş pazarlama stratejileri ve hedefe yönelik müşteri etkileşimleri oluşturmayı mümkün kılan güçlü bir alternatif sunmaktadır. Bu çalışma, K-En Yakın Komşu (KNN), Destek Vektör Makineleri (SVM) ve Rastgele Orman (Random Forest) algoritmalarını kullanarak müşteri segmentasyonu sınıflandırma performanslarını analiz etmeyi ve karşılaştırmayı amaçlamaktadır. Mevcut literatürde müşteri segmentasyonu için çeşitli yaklaşımlar önerilmiştir, ancak bu makine öğrenmesi modellerinin e-ticaret bağlamında doğrudan karşılaştırılması sınırlı kalmıştır. Bu araştırma, bu boşluğu doldurmayı hedefleyerek farklı modellerin sınıflandırma performanslarını değerlendirip çevrimiçi alışveriş yapan müşterileri segmente etmede en uygun yöntemi belirlemeyi amaçlamaktadır. Bu çalışmada kullanılan veri seti Kaggle platformundan elde edilmiş olup, alışveriş sıklığı, promosyon katılımı, harcama alışkanlıkları ve güvenlik endişeleri gibi tüketici davranışlarını yansıtan özellikleri içermektedir. En uygun küme sayısını belirlemek için Elbow ve Silhouette yöntemleri kullanılmıştır. Kümeleme yapısı belirlendikten sonra K-Means algoritması uygulanarak müşteriler davranışsal özelliklerine göre farklı

gruplara ayrılmıştır. Ardından, KNN, SVM ve Random Forest algoritmaları bu kümeleri sınıflandırmak için eğitilmiş ve performansları Accuracy (Doğruluk), Precision (Kesinlik), Recall (Duyarlılık) ve F1-Skoru gibi metrikler kullanılarak değerlendirilmiştir. Elde edilen sonuçlara göre, SVM modeli %95 doğruluk oranı ile en yüksek performansı sergilemiştir. Random Forest modeli ise %93 doğruluk oranı ile SVM'ye çok yakın sonuçlar üretmiş olup, daha düşük hesaplama maliyeti ve daha yüksek ölçeklenebilirlik avantajı sunmaktadır. KNN modeli ise %78 doğruluk oranı ile en düşük performansı göstermiştir, bu da yüksek boyutlu veya büyük veri setleri için uygun olmadığını göstermektedir. Ek olarak, Precision, Recall ve F1-Skoru açısından da SVM ve Random Forest'ın KNN'den daha üstün olduğu tespit edilmiştir, bu da onların daha güçlü sınıflandırma yeteneklerine sahip olduğunu doğrulamaktadır. Sonuç olarak, bu çalışma e-ticarette müşteri segmentasyonu için doğru makine öğrenmesi modelinin seçilmesinin önemini vurgulamaktadır. SVM, en yüksek doğruluk oranı gerektiren uygulamalar için önerilirken, Random Forest büyük veri setleri ve ölçeklenebilirlik gereksinimleri için daha uygun bir alternatif sunmaktadır. KNN modeli ise küçük ölçekli segmentasyon görevleri için daha uygun olup, büyük veri setleri için ideal olmadığı belirlenmiştir. Bu bulgular, makine öğrenmesi kullanarak müşteri segmentasyonunu optimize etmek isteyen e-ticaret işletmeleri için yol gösterici olacak ve müşteri hedefleme, etkileşim ve sadakat stratejilerini geliştirmeye katkı sağlayacaktır.

Anahtar Kelimeler: Müşteri Segmentasyonu, E-Ticaret, SVM, Random Forest, KNN

Giriş

İnternet kullanıcılarının sayısı her geçen yıl önemli ölçüde artmaktadır. İnternet, yalnızca bilgi aramak ve iletişim kurmak için değil, aynı zamanda farklı amaçlarla da kullanılmaktadır (Tangkere ve Tumewu, 2016, s.711). 1990'lı yıllardan itibaren internetin gelişimiyle birlikte, veri iletimi, işleme ve depolama teknolojilerinde, bilgi sunum biçimlerinde, ergonomide, erişilebilirlikte, hızda ve güvenilirlikte büyük ilerlemeler kaydedilmiştir. Özellikle iletişim alanında yeni ve geniş imkanlar ortaya çıkmıştır. Bu potansiyeli fark eden bazı şirketler, interneti müşterileriyle etkileşim kurmak için bir araç olarak kullanmaya başlamıştır (Santos ve diğerleri, 2017, s.130). İnternetin hızla büyümesi ve erişilebilirliğinin artmasıyla birlikte e-ticaret kavramı gelişmiştir (Mishra ve diğerleri, 2019, s.142). E-ticaret, tüketicilerin bir satıcıdan internet üzerinden doğrudan mal veya hizmet satın almasına olanak tanıyan bir elektronik ticaret biçimi olarak tanımlanmaktadır (Kasana ve Chaudhary, 2014, s.263). Aynı zamanda E-ticaret sistemi, ürünlerin çevrimiçi olarak müşterilere pazarlanması ve tanıtılması için bir platform sunmaktadır (Tabianan ve diğerleri, 2022, s.1). Teknolojinin hızlı gelişimi, internetin yaygınlaşması ve mobil uygulamaların artan kullanımı, e-ticaretin büyük bir ivme kazanmasına neden olmuştur. Günümüzde e-ticaret, küresel ekonominin en önemli bileşenlerinden biri haline gelmiştir.

E-ticaret alanındaki ilerlemeler, 1970'lerin başında elektronik fon transferi (EFT) sisteminin oluşturulmasıyla başlamıştır. EFT, finansal işlemlerin elektronik ortamda gerçekleştirilmesini sağlayan bilgisayar tabanlı bir sistem olarak tanımlanmaktadır. 1970'lerin sonlarına doğru, EFT sisteminin sınırlıklarını azaltmak amacıyla Elektronik Veri Değişimi (EDI) geliştirilmiştir. EDI, hizmet sağlayıcılar, perakendeciler ve üreticiler gibi daha geniş bir işletme grubunun sisteme katılımını mümkün kılmıştır. 1990'lı yıllardan itibaren e-ticaret; Kurumsal Kaynak Planlama (ERP) sistemleri, veri madenciliği ve veri ambarı uygulamalarını da içerecek şekilde genişlemiştir (Tokse ve Mujmule, 2021, s.1). E-ticaret işlemleri, taraflar arasındaki ilişkilere göre farklı kategorilere ayrılabilir (Santos ve diğerleri, 2017, s.132):

1. B2C (Business to Consumer – İşletmeden Tüketicie): Amazon, Trendyol, Hepsiburada gibi platformlarda işletmelerin doğrudan tüketicilere satış gerçekleştirmesi.
2. B2B (Business to Business – İşletmeden İşletmeye): Şirketlerin diğer şirketlere ürün veya hizmet sunması (örneğin, Alibaba).

3. C2C (Consumer to Consumer – Tüketiciden Tüketicie): Bireylerin birbirine ürün veya hizmet satması (örneğin, eBay, Letgo).

4. C2B (Consumer to Business – Tüketiciden İşletmeye): Bireylerin işletmelere ürün veya hizmet sunması (örneğin, serbest çalışanların Fiverr veya Upwork üzerinden hizmet sunması).

E-ticaret modelinin sağladığı kolaylık ve hız, modern toplumda insanların hızlı yaşam tarzına uyum sağlaması bakımından önemli bir avantaj sunmaktadır. Bu nedenle, giderek daha fazla sayıda e-ticaret şirketi kurulmaktadır (Liu, 2020, s.11761).

Dünya genelinde en çok kazanç sağlayan e-ticaret platformlarının başında Amazon gelmektedir. Amazon, 1994 yılında Jeff Bezos tarafından yalnızca çevrimiçi kitap satışı yapmak amacıyla kurulmuştur. Bezos'un internetin büyüme potansiyelini fark etmesiyle şirket kısa sürede müzik, elektronik ve giyim gibi farklı kategorilere yönelmiştir ve 1997 yılında halka arz edilerek küresel bir teknoloji devi hâline gelmiştir (Denli, 2021). Alibaba, 1999 yılında Jack Ma tarafından Çin'in Hangzhou kentinde, küçük işletmelerin küresel pazarlara erişimini kolaylaştırmak amacıyla kurulmuştur. İlk başta bir çeviri platformu olarak başlayan bu girişim, kısa sürede üreticiler ve alıcıları buluşturan bir B2B e-ticaret platformuna dönüşmüştür. Alibaba, 2014 yılında New York Borsası'nda halka arz edilerek 25 milyar dolarlık rekor yatırım almıştır (Güner, 2020). eBay, 1995 yılında Pierre Omidyar tarafından AuctionWeb adıyla kurulmuştur. Omidyar, bireylerin ürün alıp satabileceği çevrimiçi bir açık artırma sistemi tasarlamıştır. Platformda satılan ilk ürün, bozuk bir lazer işaretçisi olmuştur ve bu işlem, sitenin potansiyelini göstermiştir. eBay, 1998 yılında halka açılmış ve kısa sürede küresel bir pazar yeri hâline gelmiştir (Asar, 2024). Shopify, 2006 yılında Tobias Lütke, Daniel Weinand ve Scott Lake tarafından Kanada'da, başlangıçta bir snowboard mağazası olan Snowdevil için geliştirilmiştir. Bu yazılım, zamanla teknik bilgisi olmayan girişimcilerin çevrimiçi mağaza kurmasını kolaylaştıran bağımsız bir e-ticaret altyapısına dönüşmüştür (Mahadevan, 2025).

E-ticaretin yeni bir pazarlama kanalı olarak yaygınlaşmasıyla birlikte, çevrimiçi kullanıcıların ihtiyaç ve beklentilerinin analiz edilmesi, tüketici odaklı ticaretin etkinleştirilmesi açısından temel bir ön koşul olarak kabul edilmektedir. Bu bağlamda, geleneksel kitlesel pazarlama stratejileri, çevrimiçi kullanıcıların farklılaşan taleplerini karşılamakta yetersiz kalmaktadır. 2004 yılında dijital ağların yaygınlaştığı bir dönemde müşteri segmentasyonu, çevrimiçi pazarlamanın dönüşümüne yanıt olarak, kullanıcı verilerinin anlamlandırılması ve yeni iş fırsatlarına dönüştürülmesi için bir yöntem olarak benimsenmiştir.

Müşteri davranışlarını analiz etmek isteyen şirketlerin, müşteri verilerini kayıt altına alma kapasitesine, benzer kullanıcı davranışlarını tanımlayacak algoritmalara ve bu bilgiler doğrultusunda strateji geliştirme süreçlerine ihtiyaç duyduğu belirtilmektedir. Müşteri segmentasyonu, gözetimsiz öğrenme sürecine dayalı bir yöntem olarak tanımlanmakta ve benzerlik esasına dayalı olarak farklı kümeleme yaklaşımlarını kullanmaktadır (Gomes ve Meisen, 2023, s.528).

1. Müşteri Segmentasyonu ve Literatür İncelemesi

Müşteri segmentasyonu kavramı, 20. yüzyılın ortalarında Amerikalı pazar bilimci Wendell Smith tarafından ilk kez ortaya atılmıştır (Liu, 2020, s.11761). Bu kavram, müşteri analizlerinde kritik bir araç olarak kabul edilmektedir (Jahan ve Sanam, 2024, s.1). Müşteri segmentasyonu, benzer davranışlara sahip müşterilerin aynı grupta, farklı davranışlara sahip olanların ise ayrı gruplarda sınıflandırılması süreci olarak tanımlanmaktadır (Gomes ve

Meisen, 2023, s.528). Segmentasyonun temel amacı, her müşteri grubuna en uygun pazarlama stratejilerini geliştirerek işletmenin kârlılığını artırmaktır. Bu doğrultuda, doğru müşteriye doğru ürünün tanıtılması, e-ticaret sistemlerinin etkinliğini artıran bir unsur olarak öne çıkmaktadır (Tabianan ve diğerleri, 2022, s.1). Segmentasyon sürecinde dikkate alınan faktörler genel olarak coğrafi, demografik, psikografik ve davranışsal unsurlar olarak sınıflandırılmaktadır (Monil ve diğerleri, 2020, s.2105). Ancak, demografik özellikler bireysel farklılıkları tam olarak yansıtamayabileceğinden, davranışsal segmentasyon müşteriye özgü çözümler geliştirmek açısından daha etkili bir yaklaşım sunmaktadır (Patankar ve diğerleri, 2021, s.239). Demografik segmentasyon, RFM analizi (Yakınlık, Sıklık, Parasal Değer), HVCs analizi (Yüksek Değerli Müşteriler), müşteri statüsü analizi, davranışsal segmentasyon ve psikografik segmentasyon gibi çeşitli türler yaygın olarak kullanılmaktadır (Vijilesh ve diğerleri, 2021, s.823). Geleneksel segmentasyon yöntemleri genellikle manuel ve kural tabanlı yaklaşımlara dayanmakta olup bu yöntemler, iş gücü yoğun, öznel içerikli ve ölçeklenebilirlik açısından sınırlı yöntemler olarak değerlendirilmektedir. Makine öğrenmesi algoritmalarındaki gelişmeler, segmentasyon sürecinin otomatikleştirilmesine ve iyileştirilmesine olanak sağlamaktadır. Denetimli öğrenme algoritmaları (örneğin karar ağaçları, destek vektör makineleri), etiketli verilerle segment tahmini yapabilmektedir (Rajyalaxmi ve diğerleri, 2024, s.1). Denetimsiz öğrenme algoritmaları ise özellikle K-Ortalamlar (K-Means) ve hiyerarşik kümeleme gibi yaklaşımlarla müşteri segmentasyonunda sıklıkla kullanılmaktadır.

Kümeleme algoritmalarında temel ilke; kümeler içindeki örneklerin birbirine benzer, kümeler arasındaki örneklerin ise birbirinden farklı olmasıdır (Narayana ve diğerleri, 2022, s.1280):

1. Küme içi homojenlik: Her bir küme içerisindeki veriler arasında mümkün olduğunca fazla benzerlik bulunmalıdır.
2. Kümeler arası heterojenlik: Farklı kümelerdeki veriler arasında mümkün olduğunca yüksek düzeyde çeşitlilik olmalıdır.

Kümeleme teknikleri kullanılarak müşterilerin alt gruplara veya segmentlere ayrılması, müşteriler ile çevrimiçi alışveriş kanalları arasındaki ilişkilerin analiz edilmesini sağlar. Bu analizlerin sonuçları, müşterilerin çevrimiçi alışveriş yaparken neler beklediğine dair işletmelere değerli bilgiler sunar. Şirketler, bu analizlerden elde edilen verilere dayanarak müşteri ilişkileri yönetimini geliştirebilir ve müşteri memnuniyeti ile bağlılığını artırabilir (Wu ve Chou, 2011, s.332).

Son yıllarda dijital pazarlama stratejilerinin gelişmesiyle birlikte, e-ticaret platformlarında müşteri davranışlarının daha iyi anlaşılması ve yönetilmesi giderek daha fazla önem kazanmaktadır. Bu kapsamda, müşteri segmentasyonu; pazarlama faaliyetlerini kişiselleştirmek, müşteri memnuniyetini artırmak ve işletme kaynaklarını daha etkin kullanmak amacıyla sıkça başvurulan bir analiz yaklaşımı olmuştur. Müşterilerin satın alma alışkanlıklarını, demografik özelliklerini ve davranışsal eğilimlerini dikkate alan bu süreç, veri madenciliği ve makine öğrenmesi yöntemlerinin katkısıyla daha sistematik ve ölçeklenebilir hale gelmiştir. Aşağıda, müşteri segmentasyonu alanında yapılan güncel akademik çalışmaların yöntemsel yaklaşımları, kullanılan algoritmalar ve elde edilen sonuçlar bakımından detaylı bir literatür taraması sunulmuştur.

Namvar ve Gholamian (2010), veri madenciliği araçlarından yararlanarak, RFM (Recency, Frequency, Monetary – Yakınlık, Sıklık, Parasal Değer), demografik veriler ve Müşteri Yaşam Boyu Değeri (LTV) verilerine dayalı iki aşamalı yeni bir müşteri segmentasyon yöntemi geliştirmiştir. İlk aşamada, K-Ortalamlar (K-Means) kümeleme algoritması kullanılarak müşteriler RFM kriterlerine göre farklı segmentlere ayrılmıştır. İkinci aşamada ise her küme, demografik özelliklere göre yeniden alt kümelere bölünmüştür.

Peker ve diğerleri (2017), Türkiye’de faaliyet gösteren bir market zincirinden elde edilen verilerle, market perakendeciliği sektöründe müşterileri sınıflandırmak amacıyla LRFMP (Length, Recency, Frequency, Monetary, Periodicity – Uzunluk, Yakınlık, Sıklık, Parasal Değer, Periyodiklik) modelini önermiştir. Bu çalışmada üç farklı küme doğrulama indeksi, optimum müşteri sayısını belirlemek için kullanılmış, K-Ortalamlar algoritması ise müşterilerin kümeleneğinde uygulanmıştır. Sonuç olarak, beş farklı müşteri grubunun bulunduğu tespit edilmiştir.

Kamthania ve diğerleri (2018), müşteri segmentasyonu sürecinde Temel Bileşen Analizi (PCA) ve ardından k-mod kümeleme algoritmasını kullanmıştır. Ayrıca, önerilen araç seti etkileşimli görselleştirmeler ve haritalar da içermektedir. Aynı yıl, Kansal ve diğerleri (2018) üç farklı kümeleme algoritmasını – K-Means, Agglomerative (hiyerarşik) ve Mean Shift – uygulayarak müşteri segmentasyonu gerçekleştirmiştir.

Patankar ve diğerleri (2021), farklı sektörlerdeki müşterileri harcama ve gelir gibi davranışsal özelliklerine göre sınıflandırmıştır. Bu işlemde K-Ortalamlar algoritması kullanılarak müşteriler, davranışsal niteliklerine göre gruplara ayrılmıştır.

Narayana ve diğerleri (2022), müşteri segmentasyonu amacıyla K-Means, Hiyerarşik Kümeleme (Agglomerative) ve Mean Shift algoritmalarını kullanarak müşterileri cinsiyet, yaş, ilgi alanları ve harcama alışkanlıkları gibi ortak faktörlere göre gruplandırmıştır. Ayrıca, K-Means kümeleme yöntemiyle cinsiyet ve yaş kalıpları incelenmiştir.

Xiahou ve Harada (2022), B2C e-ticaret müşterilerinin alışveriş davranışlarına yönelik boylamsal zaman çizelgeleri ve çok boyutlu veri özelliklerini temel alarak, K-Ortalamlar müşteri segmentasyonu ile destek vektör makinesi (SVM) tahmini kombinasyonuna dayalı bir müşteri kaybı tahmin modeli önermiştir.

Tabianan ve diğerleri (2022), müşteri segmentasyonunda özellikle davranışsal faktörlere odaklanan bir çalışma gerçekleştirmiştir. Toplanan verilerin işlenmesi ve segmentlere ayrılması sürecinde K-Ortalamlar algoritması tercih edilmiştir. Çalışmada, üç küme arasındaki ilişki – etkinlik türü, ürünler ve kategoriler – analiz edilmiştir. Bu yaklaşım sayesinde, satıcıların en kârlı müşteri segmentinden en az kârlı segmente kadar odaklanması hedeflenmiştir.

Wulansari ve Heikal (2024), Endonezya’daki en çok ziyaret edilen üç e-ticaret platformu – Shopee, Tokopedia ve Lazada – üzerindeki müşteri davranışlarını analiz etmek için RFM modeli ile K-Means kümeleme tekniklerini birlikte kullanmıştır.

Kumar ve diğerleri (2025), standart bir e-ticaret müşteri veri kümesi üzerinde K-Means ve Hiyerarşik Kümeleme algoritmalarını uygulamıştır. Araştırma, özellikle sık aranan markalara odaklanmış ve her iki tekniğin küme ayrımı performansları silüet skorları ve görselleştirme yöntemleriyle karşılaştırılmıştır. Elde edilen sonuçlara göre, K-Means algoritması Hiyerarşik

Kümelemeye kıyasla daha iyi ayrılmış müşteri segmentleri oluşturmuştur. Bu segmentasyon işlemi, müşteriler marka arama desenlerine göre gruplara ayrılmıştır.

2. Yöntem

2.1. Amaç ve Önem

Bu çalışmanın amacı, KNN, SVM ve Random Forest makine öğrenmesi algoritmalarının e-ticaret müşterilerinin davranışsal segmentasyonundaki sınıflandırma başarımını karşılaştırmaktır. Bu bağlamda çalışmanın özgün yönü, sınıflandırma algoritmalarının aynı veri seti üzerinde karşılaştırmalı analizinin gerçekleştirilmesi ve optimum modelin önerilmesidir. Ayrıca, çalışmada kullanılan veri setinin davranışsal faktörlere odaklanması, klasik RFM temelli segmentasyon yaklaşımlarına alternatif sağlamaktadır.

2.2. Veri Seti

Çalışmada kullanılan veri seti açık dağıtım yapan Kaggle dataset sitesinden alındı. Bu veri seti, tüketici davranışları ve çevrimiçi alışveriş tercihleri hakkında kapsamlı içgörüler sunmaktadır. Veri seti anket yanıtlarına dayanır ve alışveriş sıklığı, tercih edilen ödeme yöntemleri, çekici faktörler, güvenlik endişeleri ve sıklıkla satın alınan ürün kategorileri gibi temel hususlara odaklanmaktadır. Veri setinde her bir alan için toplam 200 adet bilgi bulunmaktadır. Türkçeye çevrilmiş veri setine ait temel özellikler Tablo 1’de verilmiştir.

Tablo 1. Verilere Ait Tanımlayıcı Değerler

Alan Adı	Veri Türü
Cinsiyet	Erkek, Kadın
Çevrimiçi Alışveriş Sıklığı	Nadiren, Ayda bir, Haftada bir, Haftada birkaç
Çevrimiçi Alışveriş Oranı	%0-20, %21-40, %41-60, %61-80, %81-100
İnceleme Kontrol Sıklığı	Her zaman, Nadiren, Sık sık, Ara sıra
Çekici Faktörler	Kolaylık, İndirimler, Çeşitlilik, Zaman Tasarrufu, Kullanıcı dostu arayüz
Perakendeci Seçim Faktörleri	Marka güvenilirliği, Ürün yorumları, Fiyatlar, Müşteri hizmetleri
Tercih Edilen Ödeme Yöntemi	Kredi/Banka Kartı, Kapıda Ödeme, Dijital Cüzdanlar, Banka havalesi
Yerel ve Uluslararası Satıcılar	Tercih yok, Yerel tercih, Uluslararası tercih
Tercih Edilen Alışveriş Platformu	Amazon, Daraz, AliExpress, Others
Güvenlik Endişesi Seviyesi	Çok Endişeli, Nötr, Biraz Endişeli, Hiç Endişeli Değil
Promosyon Katılım Durumu	Ara sıra, Nadiren, 2 Sık sık, Her zaman
Fiyat Hassasiyeti	Orta, Az, Çok, Hiç
Rahat Harcama Aralığı	<1k, 1k-5k, 5k-10k, = >10k
Sık Satın Alınan Ürünler	Giyim, Kitaplar, Elektronik, Oyunlar, Gıda, vb.
En Büyük Dezavantaj	Ürün kalitesi, Hiç sorun yok, İade sorunları, Gecikmeli teslimat
Ürün Orijinalliği Endişesi	Nötr, Her zaman güveniyor, Endişeli, Ara sıra endişeli
İstenen Geliştirmeler	Daha iyi ürün açıklamaları, Daha iyi müşteri yorumları, vb.

2.3. Verilerin Analizi ve Bulgular

Makine öğrenmesi algoritmaları, geniş hacimdeki işlem, davranışsal ve demografik verileri analiz ederek, geleneksel yöntemlerin gözden kaçırabileceği gizli kalıpları, ilişkileri ve eğilimleri ortaya çıkarabilmektedir (Rajyalaxmi ve diğerleri, 2024, s.1) Bu özellikleri sayesinde E-ticarette müşteri segmentasyonu için daha doğru sonuçlar verebilmektedir. Etkili bir müşteri segmentasyonu elde etmek için en iyi K değerini belirlemek kritik öneme sahiptir (Kumar ve diğerleri, 2025, s.121). Etiketlenmemiş bir veri kümesi verildiğinde, küme içindeki örneklerin birbirine en yakın olacak şekilde dağıtılması ve kümeler arasındaki mesafenin

maksimum seviyeye çıkarılması gerekmektedir (Xiahou ve Harada, 2022, s.463). Elbow ve Silhouette yöntemleri, dışsal bir bilgiye ihtiyaç duyulmadan bir kümenin ne kadar iyi oluşturulduğunu belirlemek için kullanılır (Saputra ve diğerleri, 2019, s.343).

2.3.1. Elbow (Dirsek) Yöntemi

Elbow (Dirsek) Yöntemi, kümeleme analizinde en uygun küme sayısını (k) belirlemek için kullanılan bir yöntemdir. Kümeleme algoritmalarında, doğru küme sayısını belirlemek kritik bir adımdır. K-Means gibi algoritmalarda, kümelerin kalitesini ölçmek için WCSS (Within-Cluster Sum of Squares - Küme İçi Kareler Toplamı) kullanılır. Dirsek yöntemi, nispeten küçük k değerleri için uygundur. Dirsek yöntemi, farklı k değerlerinin kare farkını hesaplar. K değeri arttıkça, ortalama bozulma derecesi (distortion degree) küçülür. Her kategoride bulunan örnek sayısı azalır ve örnekler ağırlık merkezine daha yakın hale gelir. K değeri arttıkça, bozulma derecesindeki iyileşme etkisinin en çok azaldığı konum, dirseğin karşılık geldiği k değeridir (Cui, 2020 s.7). WCSS için Dirsek Yöntemi değerleri eşitlik 1’de verilen formülle hesaplandı.

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (1)$$

k= Küme sayısı

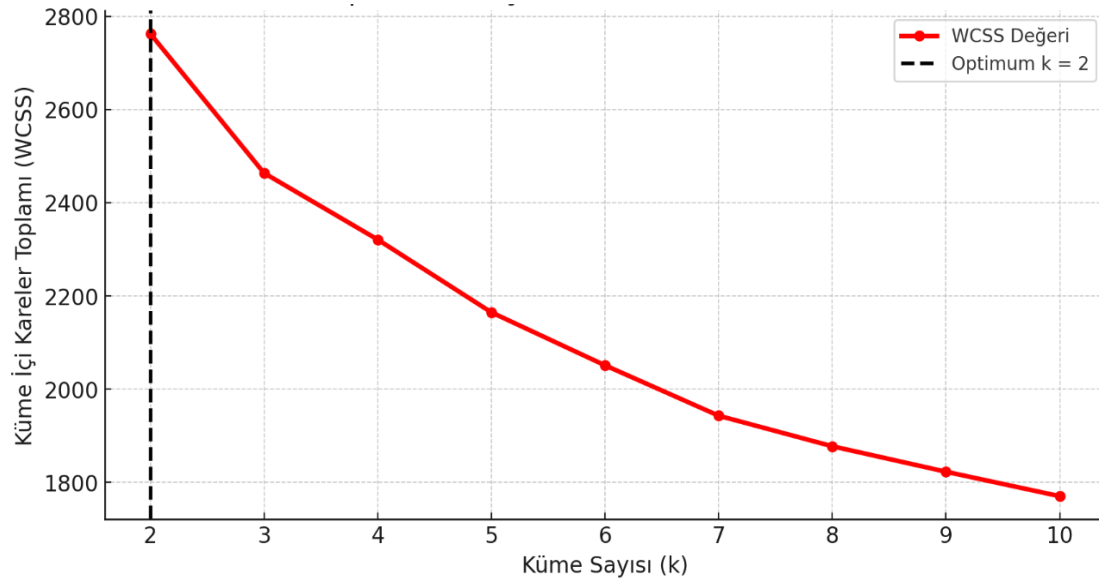
C_i = i.küme

x= Kümedeki veri noktası

μ_i =i.kümenin merkez noktası

Elbow yöntemi ile elde edilen k değeri Şekil 1’de verilmiştir.

Şekil 1. Optimum k Değeri İçin Elbow Yöntemi



2.3.2. Silhouette Yöntemi

Silhouette Yöntemi, hem ayrışmayı (separation) hem de iç tutarlılığı (cohesion) birleştiren bir silhouette katsayısı kullanır (Saputra ve diğerleri, 2019, s.343). Silhouette katsayısı -1 ile 1 arasında değişir. 1’e yakın değerler, kümeleme modelinin daha tutarlı kümeler oluşturduğunu gösterir. 0’a yakın değerler, örneğin iki komşu küme arasındaki karar sınırına çok yakın

olduğunu gösterir. Negatif değerler, örneğin yanlış kümeye atanmış olabileceğini gösterir. Silhouette katsayısı eşitlik 2'deki formülle hesaplandı.

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

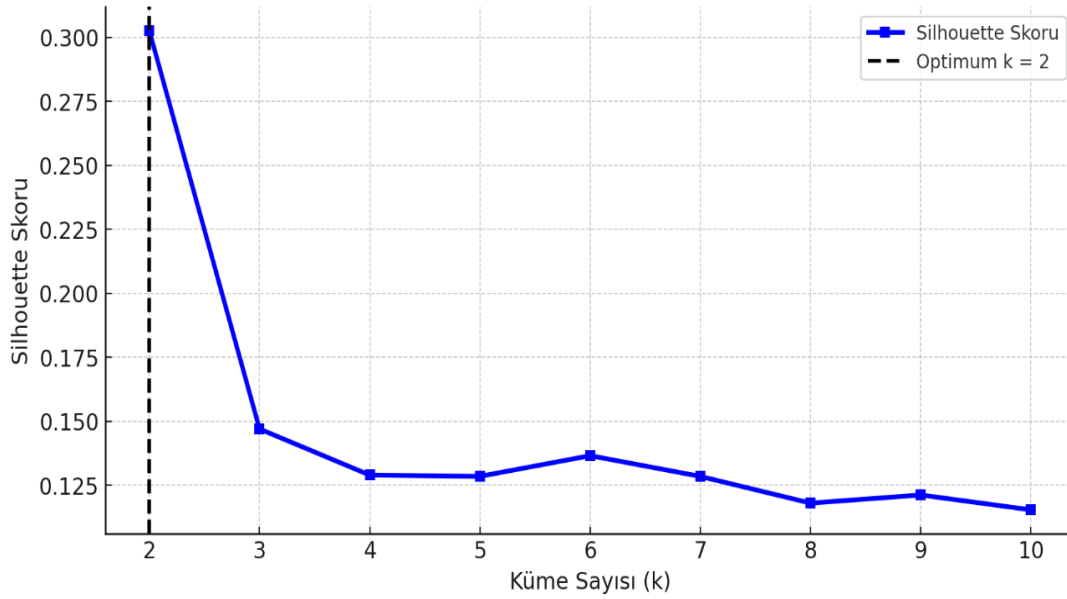
$S(i)$: Silhouette katsayısı

$a(i)$: Küme bütünleşmesi (aynı küme içindeki ortalama mesafe)

$b(i)$: Küme ayrışması (en yakın diğer kümedeki ortalama mesafe)

Silhouette yöntemi ile elde edilen k değeri Şekil 2'de verilmiştir.

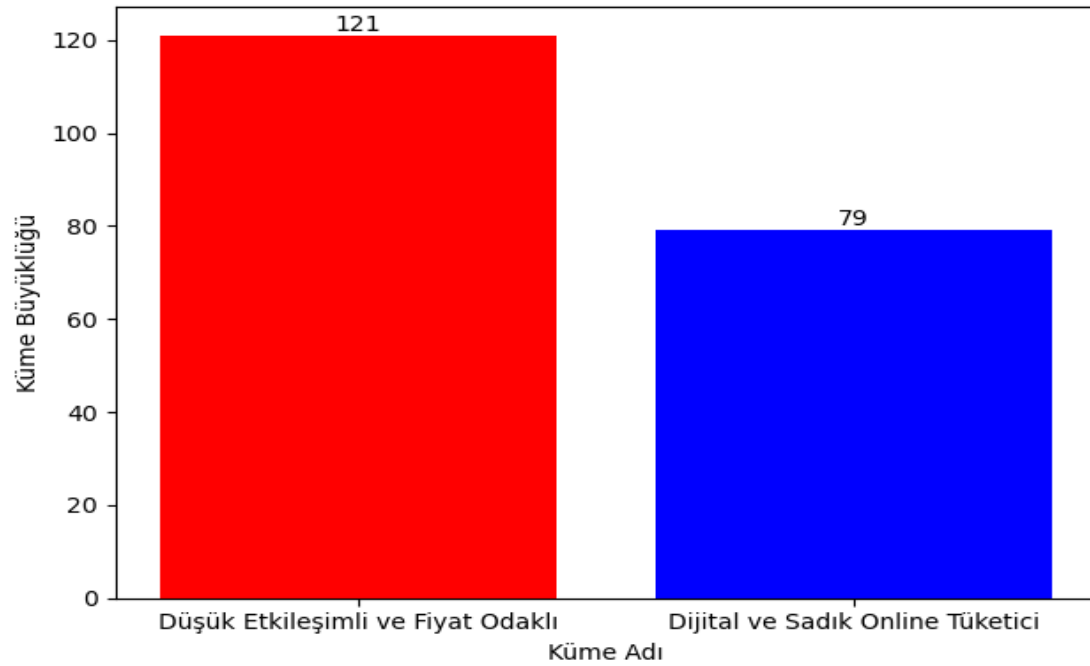
Şekil 2. Optimum k Değeri İçin Silhouette Yöntemi



Bu veri setinde en iyi küme sayısını belirlemek için her iki yöntem de kullanıldı. Sonuçlar, $k = 2$ 'nin en uygun küme sayısı olduğunu gösterdi. Böylece veri setindeki örüntüler daha anlamlı ve doğru bir şekilde gruplandırılmış oldu.

2.3.3. K-Means Kümeleme

K-Means algoritması, bölme tabanlı bir kümeleme yöntemi olup basitliği, verimliliği ve hızı ile avantajlıdır. Ancak, bu algoritma başlangıç noktalarına oldukça bağımlıdır ve ilk örneklerin seçimi sonuçların farklı olmasına neden olabilir (Ahmed ve diğerleri, 2020, s.4). Algoritma, iki ayrı aşamadan oluşur. İlk aşamada, her bir küme için bir merkez nokta tanımlanır. Bu, toplamda k adet merkez noktası belirlenmesi anlamına gelir. İkinci aşamada ise veri setindeki her bir nokta, en yakın merkez noktası ile ilişkilendirilir. Noktalar ile merkez noktaları arasındaki mesafe genellikle Öklid mesafesi (Euclidean distance) kullanılarak hesaplanır (Nazeer ve Sebastian, 2009, s.1). Kümeleme sonucunda elde edilen kümeler ve üye sayıları Şekil 3'te verilmiştir.

Şekil 3. Küme Adı ve Büyüklük Değerleri

Şekil 3'teki K-Means kümeleme sonuçları incelendiğinde optimum küme sayısı kullanıcıların alışveriş davranışlarına göre iki farklı gruba ayrıldığını gösteriyor. Gruplara ait genel karşılaştırmalar Tablo 2'de verilmiştir.

Tablo 2. Segment 1: Düşük Etkileşimli ve Segment 2: Dijital Sadık Müşteri Özellikleri

Özellik	Düşük Etkileşimli ve Fiyat Odaklı	Dijital ve Sadık Online Tüketici
Alışveriş Sıklığı	Nadiren veya Ayda Bir Kez	Haftada Birkaç Kez
Alışveriş Oranı	%0-20 veya %21-40	%41-60 veya daha fazla
Fiyat Hassasiyeti	Orta-Yüksek	Düşük
Güvenlik Endişesi	Yüksek	Düşük
Promosyon Katılımı	Ara sıra veya nadiren	Sık sık veya her zaman
Harcama Aralığı	1.000- 5.000 veya daha az	5.000 veya daha fazla
Ödeme Yöntemi Tercihi	Kapıda ödeme veya banka kartı	Dijital cüzdanlar ve kredi kartları
Platform Kullanımı	Yerel perakendecileri daha çok tercih ediyorlar	Uluslararası platformları da aktif kullanıyorlar

Tablo 2'deki verileri incelendiğinde Düşük Etkileşimli ve Fiyat Odaklı Tüketiciler daha ihtiyatlı bir alışveriş tarzları var, öncelikleri güvenlik ve fiyat politikaları olduğu görülmektedir. E-ticaret platformlarını kullanmaktan çekiniyorlar veya fiziksel mağazaları daha çok tercih ediyorlar. Bu grup için e-ticaret firmaları, güven artırıcı kampanyalar ve fiyat odaklı indirimler sunabilir. Dijital ve Sadık Online Tüketiciler e-ticaret platformlarını aktif kullanıyorlar. Ödeme yöntemleri konusunda daha esnekler, kredi kartı ve dijital cüzdanları rahatça kullanıyorlar. Marka itibarı ve müşteri hizmetleri bu grupta daha önemli, fiyat yerine hizmet kalitesine daha fazla değer veriyorlar. E-ticaret firmaları bu gruba kişiselleştirilmiş kampanyalar sunarak müşteri sadakatini artırabilir.

2.3.4. K-En Yakın Komşu (K-Nearest Neighbor – KNN) Algoritması

Sınıflandırma, giriş verilerini analiz etmeyi ve veri sınıfını tanımlayan bir model oluşturmayı içerir. KNN, gözetimli öğrenme algoritmasını kullanan (Putra ve diğerleri, 2024, s.227). KNN algoritması, model performansına dayalı olarak en uygun komşu sayısının (k) seçilmesine odaklanarak sınıflandırma yapar (Yacoub ve diğerleri, 2024, s.2). KNN sınıflandırıcısında, örnekler arasındaki mesafe, Öklid formülü kullanılarak hesaplandı (Zulkarnain ve diğerleri, 2024, s.18). Öklid formülü eşitlik 3'te verilmiştir.

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (3)$$

$d(p, q)$ = Noktaları arasındaki Öklid mesafesini temsil eder.

p = test verisi

q = gerçek veri

N = veri noktalarının boyutudur.

2.3.5. Destek Vektör Makineleri (Support Vector Machine – SVM)

Destek Vektör Makineleri (SVM), farklı disiplinlerden çeşitli veri kategorilerini sınıflandırmak için eğitilen bir gözetimli öğrenme tekniğidir. İkili sınıflandırma (two-class classification) problemlerinde yaygın olarak kullanılır. Hem doğrusal (linear) hem de doğrusal olmayan (non-linear) veri sınıflandırma görevlerine uygulanabilir (Ahmad ve diğerleri, 2018, s.33791). SVM'nin en önemli güçlü yönlerinden biri, Doğrusal (Linear), Polinom (Polynomial), Gauss (RBF) ve Sigmoid çekirdek fonksiyonları kullanarak doğrusal olmayan karar sınırlarını yönetebilme yeteneğidir. Bu çekirdek fonksiyonları kullanarak SVM, verilerdeki karmaşık ilişkileri ve desenleri yakalayabilir (Khan ve diğerleri, 2024, s.72).

2.3.6. Random Forest (Rastgele Orman, RF)

Rastgele Orman algoritması, Denetimli Öğrenme (Supervised Learning) grubuna ait bir yöntemdir. Bu yöntem, en doğru sınıflandırma algoritmalarından biri olup, aşırı öğrenme (overfitting) olmadan büyük miktarda değişken girdisini işleyebilir ve karar ağaçları arasındaki korelasyonu ortadan kaldırmaya yardımcı olur (Noviyanti ve Alamsyahs, 2024, s.44). Rastgele Orman (Random Forest) Algoritmasının uygulanması (Chunduru ve diğerleri, 2024, s.7):

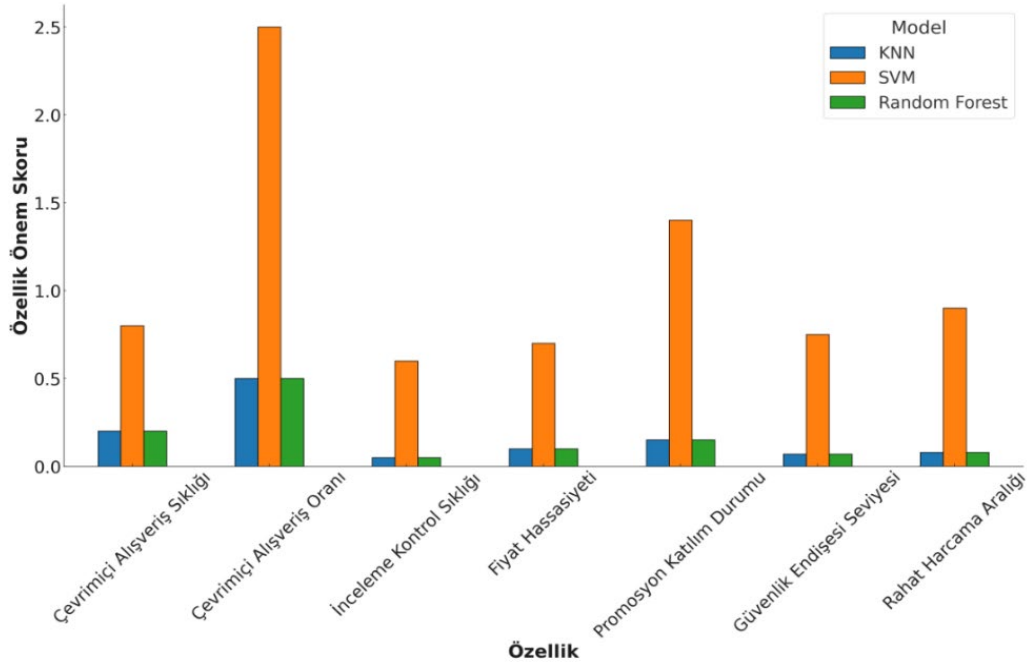
Bu algoritma temel olarak dört aşamada gerçekleştirilir:

- i. Sağlanan veri setinden rastgele örnekler seçilir.
- ii. Her örnek için bir karar ağacı oluşturulur ve her birinin sonucu tahmin edilir.
- iii. Her tahmin edilen sonuç için bir oy kullanılır.
- iv. En fazla oyu alan tahmin sonucu, nihai tahmin olarak seçilir.

K-En Yakın Komşu (KNN), Destek Vektör Makineleri (SVM) ve Rastgele Orman (Random Forest) gibi makine öğrenimi algoritmaları ile sınıflandırma işlemi gerçekleştirilmeden önce özellik önem sıralamasını çıkarmak, modelin başarısını artırmak ve yorumlanabilirliğini sağlamak açısından kritik bir adımdır. Öncelikle, gereksiz veya düşük etkili özellikleri belirlemek, modelin hesaplama maliyetini düşürür ve genel performansını artırır. Gereksiz özelliklerin dahil edilmesi, modelin aşırı öğrenmesine (overfitting) neden olabilir ve sınıflandırma doğruluğunu olumsuz etkileyebilir. Özellikle KNN gibi mesafeye dayalı algoritmalarda, fazla özelliklerin varlığı gürültüyü artırabilir ve modelin tahmin yeteneğini

düşürebilir. Özellik seçimi ayrıca modelin açıklanabilirliğini geliştirir. SVM ve Random Forest gibi algoritmalarda, hangi özelliklerin daha etkili olduğunu anlamak, işlemeye alınan verinin içgörülerini keşfetmeye yardımcı olur. Yöntemlere ait özellik önem sıraları Şekil 4’de verilmiştir.

Şekil 4. Modeller İçin Özellik Önem Sıralaması



SVM modeli, özellik önem skorlarını diğer modellere kıyasla daha yüksek hesaplanmıştır. Çevrimiçi Alışveriş Oranı, tüm modellerde en önemli özellik olarak öne çıkarken, SVM’de bu fark daha belirgin hale gelmiştir. Random Forest ve KNN, özellikleri daha dengeli şekilde değerlendirirken, Çevrimiçi Alışveriş Sıklığı ve Promosyon Katılım Durumu bu iki modelde orta seviyede önem arz etmektedir. Fiyat Hassasiyeti, Güvenlik Endişesi ve İnceleme Kontrol Sıklığı, tüm modellerde daha düşük önem skorlarına sahiptir. Bu, alışveriş davranışlarını tahmin ederken bu özelliklerin daha az etkili olduğunu göstermektedir.

KNN, SVM ve Random Forest gibi sınıflandırma modellerinde accuracy, precision, recall ve F1-skoru, modelin tahmin performansını değerlendirmek için kullanılır. Accuracy (Doğruluk), modelin toplam doğru tahmin oranını ölçer. Dengeli veri setlerinde faydalıdır, ancak dengesiz sınıflarda yanıltıcı olabilir. Precision (Kesinlik), yanlış pozitifleri (False Positive - FP) azaltmayı hedefler. Recall (Duyarlılık), yanlış negatifleri (False Negative - FN) azaltmaya odaklanır. F1-Skoru: Precision ve recall’un dengelenmiş bir versiyonudur. Modelin genel başarısını değerlendirir ve dengesiz veri setlerinde daha iyi bir ölçüm sunar. Değerlere ait hesaplamalar eşitlik 4, 5, 6 ve 7’deki formüller ile yapılır.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{F1 - Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

TP (True Positive): Doğru pozitif tahminler
TN (True Negative): Doğru negatif tahminler
FP (False Positive): Yanlış pozitif tahminler
FN (False Negative): Yanlış negatif tahminler

Tablo 3. Sınıflandırma Modellerinin Performans Metrikleri

Model	Accuracy	Precision	Recall	F1-Score
K-NN	0,78	0,82	0,78	0,77
SVM	0,95	0,95	0,95	0,95
Random Forest	0,93	0,93	0,93	0,92

Destek Vektör Makineleri (SVM), Rastgele Orman (Random Forest) ve K-En Yakın Komşu (KNN) algoritmalarının performans metrikleri incelendiğinde, SVM en iyi sonuçları üretirken, Random Forest ona çok yakın bir performans göstermekte ve KNN en düşük sonuçları sergilemektedir. Accuracy (Doğruluk) açısından bakıldığında, SVM modeli %95 ile en yüksek doğruluk oranına sahiptir. Random Forest %93 doğruluk oranı ile SVM'ye çok yakın bir performans sergilemektedir. KNN ise %78 doğruluk ile en düşük sonucu vermektedir. Bu, KNN modelinin diğer modellere kıyasla daha fazla yanlış tahmin ürettiğini göstermektedir. Precision (Kesinlik) değerleri incelendiğinde, SVM ve Random Forest modellerinin her ikisi de %95 ve %93 kesinlik oranına sahiptir. Bu, yanlış pozitif tahminlerin düşük olduğunu ve modelin doğru pozitifleri daha iyi tahmin ettiğini gösterir. KNN modeli %82 precision oranı ile diğer iki modele kıyasla daha düşük bir kesinlik sergilemektedir. Bu, modelin yanlış pozitif tahminleri biraz daha fazla yaptığını gösterir. Recall (Duyarlılık) açısından SVM yine %95 ile en yüksek sonucu elde etmektedir. Random Forest %93 recall oranı ile ona yakın bir performans göstermektedir. KNN ise %78 recall ile en düşük sonucu üretmektedir. Recall değerinin düşük olması, modelin bazı gerçek pozitifleri kaçırdığı anlamına gelir. Bu nedenle, KNN modeli özellikle duyarlılığın önemli olduğu uygulamalarda yetersiz olabilir. F1-Skoru, Precision ve Recall'un dengelenmiş bir metriği olduğundan, modellerin genel performansını değerlendirmek için kullanışlıdır. SVM modeli %95 F1-Skoru ile en iyi sonucu verirken, Random Forest %92 ile ona çok yakın bir performans göstermektedir. KNN modeli ise %77 ile en düşük F1-Skoruna sahiptir. Bu da KNN modelinin, diğer modellere kıyasla daha fazla yanlış tahmin ürettiğini ve genel performansının daha düşük olduğunu göstermektedir.

Sonuç

Bu çalışma, e-ticarette kullanıcı davranışlarını sınıflandırmak ve müşteri segmentasyonu yapmak amacıyla K-En Yakın Komşu (KNN), Destek Vektör Makineleri (SVM) ve Rastgele Orman (Random Forest) algoritmalarının performanslarını karşılaştırmıştır. Kullanılan veri seti, tüketicilerin alışveriş sıklığı, promosyon katılımı, fiyat hassasiyeti, güvenlik endişesi ve ödeme tercihleri gibi çeşitli değişkenleri içermektedir. İlk olarak, Elbow ve Silhouette yöntemleri ile optimum küme sayısı belirlenmiş, ardından K-Means algoritması kullanılarak müşteriler benzer davranışsal kalıplarına göre kümelenmiştir. Son olarak, belirlenen bu kümeler, KNN, SVM ve Random Forest algoritmaları ile sınıflandırılmış ve performansları Accuracy, Precision, Recall ve F1-Skoru gibi metriklerle ölçülmüştür.

Elde edilen bulgular, SVM algoritmasının %95 doğruluk oranı ile en başarılı sınıflandırma performansını sergilediği görülmüştür. Bu sonuç, Tanır ve Ramazanov (2023) tarafından yürütülen çalışmada elde edilen %79,89 doğrulukla tahmin oranının üzerinde bir başarıya

ulaşmıştır. Random Forest algoritması ise %93 doğruluk oranıyla SVM'ye yakın bir performans göstermiştir. Bu sonuç, Doğanlı (2025) tarafından gelir ve harcama skorlarına dayalı müşteri segmentasyonu çalışmasında elde edilen bulgularla benzerlik göstermektedir. Her iki çalışmada da RF algoritması, öznitelik önem sıralaması açısından etkili bir araç olarak öne çıkmıştır. KNN algoritması ise %78 doğruluk oranı ile en düşük performansı sergilemiştir. Bu bulgu, Akbulut (2006) tarafından gerçekleştirilen kozmetik sektörüne yönelik segmentasyon ve ayrılma tahmini analizleri ile doğrudan örtüşmektedir. Akbulut'un çalışmasında da KNN algoritmasının müşteri ayrılma eğilimlerini belirlemede sınırlı doğruluk sunduğu ortaya konmuştur.

Bu çalışmanın bazı kısıtları bulunmaktadır. Kullanılan veri seti belirli bir tüketici grubu ile sınırlıdır. Daha büyük ve çeşitli veri setleri ile modelin genellenebilirliği artırılabilir. Çalışmada belirli sayıda özellik (alışveriş sıklığı, promosyon katılımı vb.) kullanılmıştır. Ancak, farklı özelliklerin (örneğin, müşteri yorumları, marka sadakati gibi) eklenmesi model performansını etkileyebilir. Bu kapsamda çalışma, e-ticaret sektöründeki müşteri segmentasyonu literatürüne hem teorik hem uygulamalı düzeyde katkı sağlamaktadır. Gelecekte daha geniş veri setleri, derin öğrenme algoritmaları ve dinamik modellerle bu çalışmalar geliştirilebilir. Özellikle farklı sektörlerde benzer karşılaştırmalar yapılarak algoritmaların sektörel geçerlilikleri test edilebilir.

Kaynakça

- Ahmad, I., Basher, M., Iqbal, M. J., & Rahim, A. (2018). Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. *IEEE ACCESS*, 6, 33789-33795. <https://doi.org/10.1109/ACCESS.2018.2841987>
- Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(1), 1-12. <https://doi.org/10.3390/electronics9081295>
- Asar, Ç. A. (2024). eBay'de Satılan İlk Ürünün Ne Olduğunu Öğrenince 'Şaka Herhalde?' Diyeceksiniz. Erişim adresi: <https://www.webtekno.com/ebay-satilan-ilk-urun-h151745.html>
- Chunduru, A., Kishore, A. R., Sasapu, B. K., & Seepana, K. (2024). Multi chronic disease prediction system using CNN and random forest. *SN Computer Science*, 5(157), 1-13. <https://doi.org/10.1007/s42979-023-02521-6>
- Akbulut S. (2006). Veri Madenciliği Teknikleri ile Bir Kozmetik Markanın Ayrılan Müşteri Analizi ve Müşteri Segmentasyonu, (Yüksek Lisans Tezi), Gazi Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği, Ankara. Erişim adresi: https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=lqu6rrVmt1RZ_Ntn3IVpPw&no=OCDd8HOHVAwmqA4VBI6IGg
- Cui, M. (2020). Introduction to the K-Means clustering algorithm based on the Elbow method. *Accounting, Auditing and Finance*, 1, 5-8. <https://doi.org/10.23977/accaf.2020.010102>
- Denli, İ. (2021). Görevinden Ayrılan Jeff Bezos'ın İlham Verici Amazon'u Kurma Hikâyesi. Erişim adresi: <https://www.webtekno.com/jeff-bezos-amazon-kurulusu-h113225.html>
- Doğanlı B. (2025). Müşteri segmentasyonu ve davranış analizi: Random forest algoritması kullanılarak gelir ve harcama davranışlarının incelenmesi. *Business, Economics and Management Research Journal*, 8(1), 52-66. <https://doi.org/10.58308/bemarej.1646966>
- Gomes, M. A., & Meisen, T. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. *Information Systems and e-Business Management*, 21, 527-570. <https://doi.org/10.1007/s10257-023-00640-4>
- Güner, O. O. (2020). Alibaba.com'un İsmi Neden Alibaba? İşte Jack Ma'nın Akıllamaz Başarı Hikayesi. Erişim adresi: <https://www.webtekno.com/alibaba-basari-hikayesi-jack-ma-h83302.html>
- Jahan, I., & Sanam, T. F. (2024). A comprehensive framework for customer retention in e-commerce using machine learning based on churn prediction, customer segmentation, and recommendation. *Electronic Commerce Research*, 1-44. <https://doi.org/10.1007/s10660-024-09936-0>
- Kamthania, D., Pahwa, A., & Madhavan, S. S. (2018). Market segmentation analysis and visualization using k-mode clustering algorithm for e-commerce business. *Journal of Computing and Information Technology*, 26(1), 57-68. <https://doi.org/10.20532/cit.2018.1003863>
- Kasana, J., & Chaudhary, N. (2014). A comparative study of eBay and Amazon in online shopping. *International Research Journal of Commerce Arts and Science*, 5(2), 263-275. <https://doi.org/10.32804/CASIRJ>
- Kansal, T., Bahuguna, S., Singh, V., & Choudhury, T. (2018). Customer segmentation using k-means clustering. 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), 135-139. <https://doi.org/10.1109/CTEMS.2018.8769171>
- Khan, T. A. Sadiq, R., Shahid, Z., Alam, M. M., & Suud, M. M. (2024). Sentiment analysis using support vector machine and random forest. *Journal of Informatics and Web Engineering*, 3(1), 67-75. <https://doi.org/10.33093/jiwe.2024.3.1.5>
- Kumar, S., Rani, R., Pippal, S. K., & Agrawal, R. (2025). Customer segmentation in e-commerce: K-means vs hierarchical clustering. *Telecommunication Computing Electronics and Control*, 23(1), 119-128. <https://doi.org/10.12928/TELKOMNIKA.v23i1.26384>
- Liu, F. (2020). 3D block matching algorithm in concealed image recognition and e-commerce customer segmentation. *IEEE Sensors Journal*, 20(20), 11761-11769. <https://doi.org/10.1109/JSEN.2019.2936169>
- Mahadevan, A. (2025). Snowboard'lardan 160 Milyar Dolara: Shopify'ın Köken Hikayesi. Erişim adresi: <https://www.doola.com/tr/blog/from-snowboards-to-160-billion-the-shopify-origin-story/>

- Mishra, M., Chopde, J., Shah, M., Parikh, P., Babu, R. C., & Woo, J. (2019). Big data predictive analysis of Amazon product review. *KSII The 14th Asia Pacific International Conference on Information Science and Technology (APIC-IST)*, 141-147. Erişim adresi: <https://www.calstatela.edu/sites/default/files/amazonprodreviewapic-ist2019.pdf>
- Monil, P., Darshan, P., Jecky, R., Vimarsh, C., & Bhatt, B. R. (2020). Customer segmentation using machine learning. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 8(6), 2104-2108. <http://doi.org/10.22214/ijraset.2020.6344>
- Namvar, M., & Gholamian, M. R. (2010). A two-phase clustering method for intelligent customer segmentation. *2010 International Conference on Intelligent Systems, Modelling and Simulation*, 215-219. <https://doi.org/10.1109/ISMS.2010.48>
- Narayana, V. L., Sirisha, S., Divya, G., Pooja, N. L. S., & Nouf, A. (2022). Mall Customer Segmentation Using Machine Learning. *Proceedings of the International Conference on Electronics and Renewable Systems (ICEARS 2022)*, 1280-1288. <https://doi.org/10.1109/ICEARS53579.2022.9752447>
- Nazeer, K. A. A., & Sebastian, M. P. (2009). Improving the accuracy and efficiency of the k-means clustering algorithm. *Proceedings of the World Congress on Engineering*, 1-5. Erişim adresi: https://www.researchgate.net/publication/44260003_Improving_the_Accuracy_and_Efficiency_of_the_k-means_Clustering_Algorithm
- Noviyanti, C. N., & Alamsyah. (2024). Early detection of diabetes using random forest algorithm. *Journal of Information System Exploration and Research*, 2(1), 41-48. <https://doi.org/10.52465/joiser.v2i1.245>
- Patankar, N., Dixit, S., Bhamare, A., Darpel, A., & Raina, R. (2021). Customer segmentation using machine learning. *Recent Trends in Intensive Computing*, 239-244. <https://doi.org/10.3233/APC210200>
- Peker, S., Kocyigit, A., & Eren, P. E. (2017). LRFMP model for customer segmentation in the grocery retail industry: A case study. *Marketing Intelligence & Planning*, 35(4), 544-559. <https://doi.org/10.1108/MIP-11-2016-0210>
- Putra, F., Tahiyat, H. F., Ihsan, R. M., & Efrizoni, L. (2024). Application of K-Nearest Neighbor Algorithm Using Wrapper as Preprocessing for Determination of Human Weight Information. *Indonesian Journal of Machine Learning and Computer Science Journal*, 4, 273-281. <https://doi.org/10.57152/malcom.v4i1.1085>
- Rajyalaxmi, M., Vijai, C., Srivastava, K., Kalyan, N., Pravallika, B., & Dutt, A. (2024). Application of machine learning algorithms for customer segmentation in e-commerce management. *2024 International Conference on Science Technology Engineering and Management (ICSTEM)*, 1-5. <https://doi.org/10.1109/ICSTEM61137.2024.10560944>
- Saputra, D. M., Saputra, D., & Oswari, D. L. (2019). Effect of distance metrics in determining k-value in k-means clustering using Elbow and Silhouette method. *Advances in Intelligent Systems Research*, 172, 341-346. <https://doi.org/10.2991/aisr.k.200424.051>
- Santos, V. F., Sabino, L. R., Morais, G. M., & Gonçalves, C. A. (2017). E-commerce: A short history follow-up on possible trends. *International Journal of Business Administration*, 8(7), 130-138. <https://doi.org/10.5430/ijba.v8n7p130>
- Tabianan, K., Velu, S., & Ravi, V. (2022). K-means clustering approach for intelligent customer segmentation using customer purchase behavior data. *Sustainability*, 14, 7243. <https://doi.org/10.3390/su14127243>
- Tangkere, I. H., & Tumewu, F. J. (2016). The influence of customer perception and customer attitude on customer purchase intention of Zalora online shop in Manado. *Jurnal Berkala Ilmiah Efisiensi*, 16(4), 709-721. Erişim adresi: <https://ejournal.unsrat.ac.id/index.php/jbie/article/view/13708>
- Tanır, D., & Ramazanov, S. (2023). Veri Madenciliği Yöntemleri ile Türkiye’de Fertlerin E-Ticaret Kullanımını Etkileyen Faktörlerin Analizi. *KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi*, 25(44), 46-65. Erişim adresi: <https://dergipark.org.tr/tr/pub/kmusekad/issue/78068/1169823>
- Wu, R. S., & Chou, P. H. (2011). Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. *Electronic Commerce Research and Applications*, 10, 331-341. <https://doi.org/10.1016/j.elerap.2010.11.002>