MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE MÜŞTERİ DAVRANIŞ ANALİZİ

Rabia Gülizar TUNCER ,Doç. Dr. Serkan ONAR
Yıldız Teknik Üniversitesi, Makine öğrenmesi, rgulizartuncer@gmail.com, 0009-0002-1816-361X
Yıldız Teknik Üniversitesi, Makine öğrenmesi ,serkan10ar@gmail.com, 000-0003-3084-7694

Özet

Bu çalışmada, e-ticaret platformlarında müşteri davranışlarını anlamak, tahmin etmek ve aksiyona dönüştürmek amacıyla makine öğrenmesi ve zaman serisi analizleri entegre bir şekilde uygulanmıştır. RetailRocket veri seti kullanılarak kullanıcıların geçmiş etkileşimleri temel alınmış ve üç aşamalı bir proje yaklaşımı benimsen- miştir. Davranış tabanlı sınıflandırma aşamasında, kullanıcı bazında davranış öznitelikleri çıkarılmış, SMOTE yöntemiyle dengesiz veri sorunu giderilmiş ve çeşitli sınıflandırma algoritmaları (Random Forest, Logistic Regression, Naive Bayes, XGBoost) ile "alışveriş yapar mı?" sorusu tahmin edilmiştir. Model performansları F1, ROC AUC gibi metriklerle değerlendirilmiş, açıklanabilir yapay zeka araçlarıyla modellerin içgörüleri analiz edilmiştir.Zaman serileri ile davranışsal analiz aşamasında, kullanıcıların zamana bağlı eylem sıraları Prophet, ARIMA, Holt-Winters ve LSTM gibi zaman serisi modelleriyle analiz edilmiştir. Ayrıca, event log verisi üzerinden Markov geçiş modelleri oluşturularak kullanıcıların etkileşim dizileri arasındaki olasılık temelli geçiş yapıları incelenmiştir. Segmentasyon ve öneri sistemi geliştirme aşamasında, kullanıcı kümeleri KMeans algoritması ile belirlenmiş, segment bazlı davranış istatistikleri çıkarılmış ve Apriori algoritması ile ürün öneri kuralları oluşturulmuştur. Tüm bu analizler, FastAPI çatısı altında geliştirilen bir servis aracılığıyla gerçek zamanlı olarak kullanıcılara tahmin ve öneri sunabilecek şekilde entegre edilmiştir. Sonuç olarak, bu çalışma müşteri davranışlarının çok boyutlu olarak analiz edildiği ve bu analizlerin karar destek sistemlerine dönüştürüldüğü bütüncül bir veri bilimi yaklaşımı sunmaktadır.

Anahtar Kelimeler: Machine learning, customer behavior analysis, time series, segmentation, recommendation system

CUSTOMER BEHAVIOR ANALYSIS WITH MACHINE LEARNING METHODS

Rabia Gülizar TUNCER , Assoc. Prof. Dr. Serkan ONAR
Yıldız Teknik Üniversitesi, Machine learning, rgulizartuncer@gmail.com, 0009-0002-1816-361X
Yıldız Teknik Üniversitesi, Machine learning,serkan10ar@gmail.com, 000-0003-3084-7694

Abstract

In this study, an integrated approach combining machine learning and time series analysis was implemented to understand, predict, and act upon customer behavior in e-commerce platforms. Based on the Retail- Rocket dataset, which includes users' past interactions, a three-phase project methodology was adopted.In the behavior-based classification phase, user-level behavioral features were engineered, the class imbalance problem was addressed using the SMOTE technique, and various classification algorithms (Random Forest, Logistic Regression, Naive Bayes, XGBoost) were employed to predict whether a user would make a pur- chase. Model performances were evaluated using metrics such as F1 and ROC AUC, and model insights were interpreted through explainable AI tools.In the time series and behavioral analysis phase, users' time- dependent action sequences were modeled using time series forecasting methods such as Prophet, ARIMA, Holt-Winters, and LSTM. Additionally, Markov transition models were constructed based on event log data to examine the probabilistic structures of interaction transitions.In the segmentation and recommendation sys- tem development phase, user clusters were identified using the KMeans algorithm, segment-based behavioral statistics were generated, and product recommendation rules were derived using the Apriori algorithm. All these analyses were integrated into a real-time service built with FastAPI, enabling personalized prediction and recommendation for users.As a result, this study presents a holistic data science approach that enables multidimensional analysis of customer behavior and its transformation into actionable decision support systems.

Keywords: Makine Öğrenmesi, müşteri davranışı analizi, zaman serileri, segmentasyon, öneri sistemi

GİRİŞ

Günümüzde sosyal ve resmi işlemler büyük oranda dijital ortamlara taşınmış ve bu platformlar üzerinden yürütülmektedir. Dolayısıyla e-ticaret platformlarında müşteri davranışlarını anlamak, platformlar için stratejik bir avantaj sağlamanın yanı sıra artık bir ihtiyaç halini almıştır.Bu çalışma, kullanıcıların geçmiş etkileşimlerinden yola çıkarak alışveriş yapma eğilimlerini tahmin etmeyi, kullanıcıları gruplamayı ve müşterilerin ihtiyaçlarına uygun, kişiselleştirilmiş ürün önerileri sunmayı hedeflemektedir.Çalışma, makine öğrenmesi ve zaman serileri ile modellemenin bir arada uygulandığı üç katmanlı bir yapıda yürütülmüş olup, müşterilerin davranışlarını baz alan bir öneri sistemi sunmaktadır

1. LİTERATÜR ÖZETİ

Singla, Randhawa ve Jain (2017) tarafından yapılan çalışmada, e-ticaret platformlarındaki müşteri ürün yorumlarının duygu analizi, makine öğrenmesi teknikleriyle incelenmiştir. Amazon.com'dan alınan 400.000'den fazla mobil telefon yorumu, Naive Bayes, SVM ve Karar Ağacı algoritmaları ile sınıflandırılmış; en iyi sonuç SVM modeli ile elde edilmiştir. Bu yaklaşım, müşteri davranışlarının analizi ve tahmininde büyük veri analitiğinin gücünü ortaya koymuştur.Singh ve arkadaşları (2019), müşteri segmentasyonu ve sınıflandırması için makine öğrenmesi tekniklerini kullanarak demografik özelliklere göre (yaş, gelir, harcama puanı) segmentasyon ve sınıflandırma yapmış; MLP algoritmasının en iyi performansı sergilediğini göstermiştir. CRM odaklı bu yaklaşım, işletmeler için müşteri davranışlarını anlamada ve karar destek sistemlerinin geliştirilmesinde katkı sunmaktadır. Bhade ve arkadaşları (2018), K-means kümeleme algoritmasını kullanarak müşteri segmentasyonu yapmış ve segment bazlı pazarlama stratejilerinin geliştirilmesine olanak sağlamıştır. Çalışma, müşteri memnuniyetini ve şirket kârlılığını artırma amacı taşımaktadır. Zhou (2025) tarafından gerçekleştirilen çalışmada, RetailRocket veri seti kullanılarak tüketici satın alma davranışlarını tahmin eden bir model geliştirilmiş; Genetik Algoritma (GA) ve XGBoost kombinasyonu ile performans artırılmıştır. Ayrıca, K-means ile veri dengesizliği sorunu çözülerek tahmin doğruluğu iyileştirilmiştir.Jannach ve arkadaşları (2023), e-ticaret platformlarında oturum bazlı öneri sistemlerini incelemiş; kullanıcı davranışlarını analiz ederek trendlerin, promosyonların ve hatırlatmaların karar verme süreçlerine etkisini vurgulamıştır.

2. YÖNTEM

Bu projede kullanılan algoritmalar ve veri işleme adımları, yapay zekâ destekli gelişmiş özellikler sunan **Cursor** adlı kod editörü üzerinde, Python programlama dili kullanılarak gerçekleştirilmiştir. Bu editör; otomatik tamamlama, hata tespiti ve kod önerileri gibi işlevlerle geliştirme sürecini kolaylaştırmıştır.Çalışmada, 4.5 aylık bir dönemdeki davranış verilerini içeren; tıklamalar, sepete eklemeler ve satın almalar gibi etkileşimleri kapsayan **RetailRocket** açık kaynak veri seti kullanılmıştır. Veri seti üzerinde yürütülen analiz süreci üç aşamada planlanmıştır:

Davranış Tabanlı Sınıflandırma:

Kullanıcı bazlı davranış özellikleri üretilmiş, **SMOTE** yöntemiyle veri dengesizliği çözülmüş ve çeşitli algoritmalar kullanılarak "alışveriş yapar mı?" sorusu tahmin edilmiştir. Model performansları **F1, ROC AUC** ve **accuracy** gibi metriklerle değerlendirilmiştir. Ayrıca **SHAP** ile model açıklamaları yapılmıştır.

Zaman Serisi Analizi:
Kullanıcıların alışveriş yolculuğu; **Prophet**, **ARIMA**, **Holt-Winters** ve **LSTM** modelleriyle analiz edilmiş, **Markov geçiş olasılıkları** hesaplanmış ve kullanıcıların davranış sıraları incelenmiştir.

Segmentasyon ve Öneri Sistemi:
K-Means algoritmasıyla kullanıcılar segmente edilmiş, her segmentin davranış istatistikleri analiz edilmiştir.

Apriori algoritmasıyla ürün önerileri belirlenmiş ve **FastAPI** çatısı altında tahmin ve öneri sistemi geliştirilmiştir. Sistemin gerçek zamanlı çalışması lokal ortamda test edilmiştir.

3. BULGULAR

Bu bölümde, gerçekleştirilen analiz ve uygulamalar sonucunda elde edilen bulgular sunulmaktadır. Çalışmada, üç ana aşama kapsamında farklı veri analizi ve modelleme teknikleri uygulanmış ve ayrıntılı şekilde değerlendirilmiştir.Çalışmada, gerçek dünya e-ticaret verisi olan **RetailRocket** veri seti kullanılmıştır. Veri seti, kullanıcıların platform üzerindeki geçmiş etkileşimlerini (örneğin: "görüntüleme", "sepete ekleme", "satın alma" gibi olay türleriyle) ve ilgili zaman damgası (**timestamp**) ile ürün bilgilerini içeren detaylı **event log** kayıtlarından oluşmaktadır.Bu veri seti temel alınarak analizler üç ana aşamada gerçekleştirilmiştir. İlk aşama olan **Davranış Tabanlı Sınıflandırma**'da, kullanıcılara ait özgün davranışsal öznitelikler (örneğin: görüntüleme ve sepete ekleme sayıları, oranları) event log verilerinden türetilerek kullanıcı bazlı bir veri setine dönüştürülmüştür.Daha sonra bu RetailRocket veri setiyle, makine öğrenmesi algoritmalarından **Random Forest**, **Logistic Regression**, **Naive Bayes** ve **XGBoost** modelleri, tek bir **train-test split** (test = 0.2, random_state = 42) üzerinde eğitilmiştir.Eğitilen modeller; **Accuracy**, **F1 Skoru** ve **ROC AUC** gibi standart metriklerle değerlendirilmiştir. Modellerin değerlendirme sonuçları, **Şekil 1**'de verilen grafikte görülmektedir.

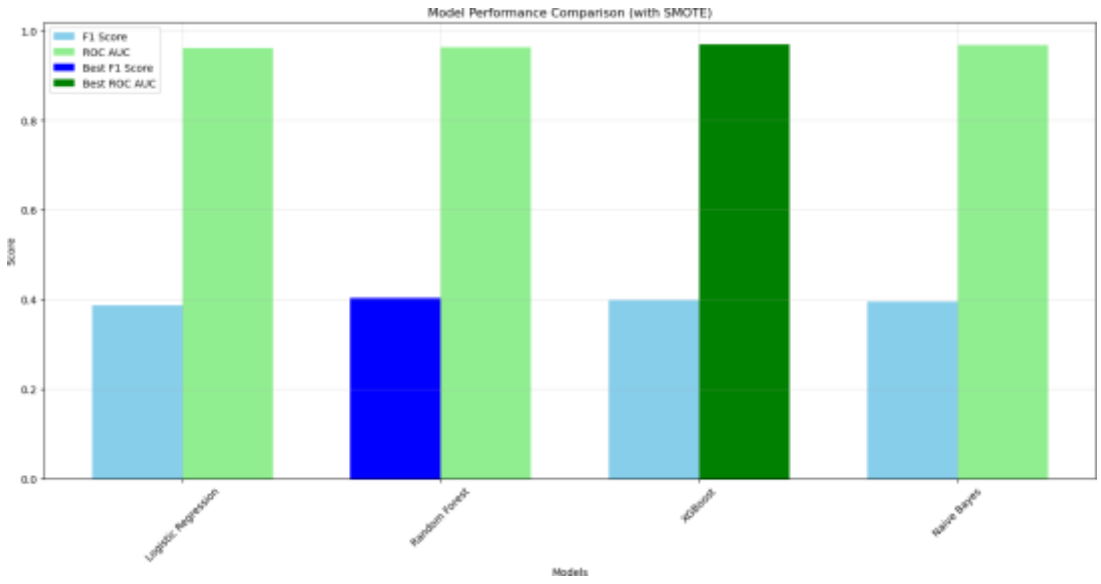


Figure 1: Makine Öğrenmesi Algoritmalarının Başarı Oranları

Figure 1'de sunulan grafik incelendiğinde, spesifik train-test splitinde Random Forest modeli, sayısal olarak çok küçük bir farkla da olsa, XGBoost modelinden daha yüksek bir F1 skoru elde etmiştir. Bu nedenle grafik, "Best F1 Score" olarak Random Forest'ı vurgulamıştır.Ancak projede uygulanan 5-fold cross-validation ortalamasına bakıldığında, XGBoost modelinin F1 skoru 0.5590, Random Forest'ın ise 0.5566 olarak gözlemlenmiştir. Bu durum, XGBoost modelinin ortalama F1 skoru açısından Random Forest modeline kıyasla hafif bir üstünlük sağladığını göstermektedir.Cross-validation sonuçlarının genellikle daha güvenilir olduğu göz önünde bulundurulduğunda, bu durum XGBoost'un da çok güçlü ve hatta genele bakıldığında biraz daha iyi bir model olabileceğine işaret etmektedir.RetailRocket, dengesiz bir veri setidir. Pozitif sınıf (satın alma yapanlar), veri setinin yaklaşık %0.83'ünü oluşturmaktadır. Dengesiz veri setinin etkisini azaltmak için SMOTE yöntemi uygulanmış ve modellerin performansları kıyaslanmıştır.Başarı durumları Figure 2 ile gösterilmiştir

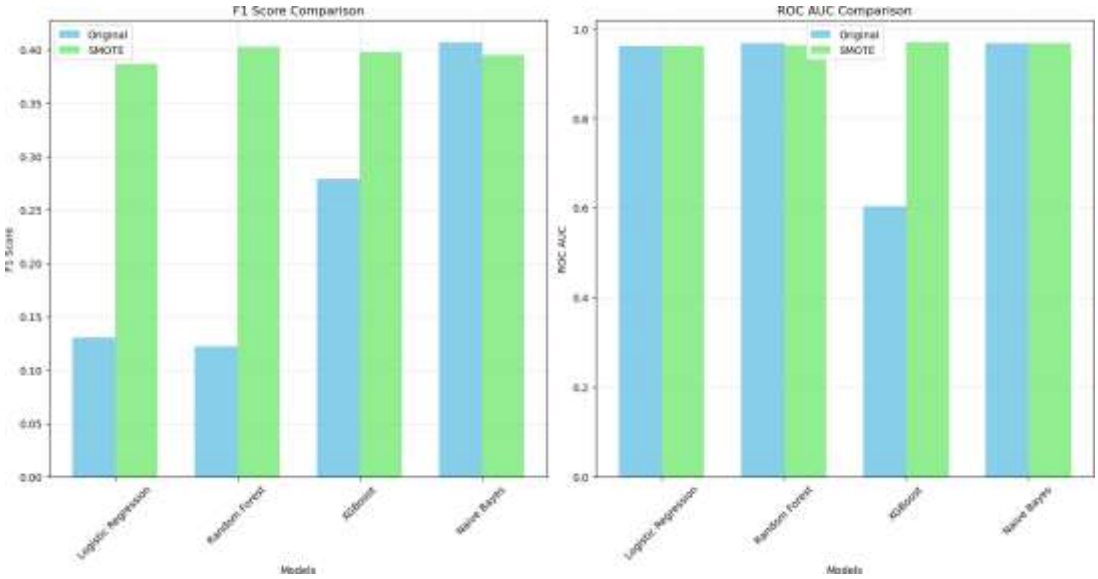


Figure 2: Makine Öğrenmesi Algoritmalarının SMOTE Uygulandıktan Sonraki Başarı Oranları

Figure 2'de açıkça görüldüğü gibi, SMOTE uygulamasıyla birlikte model performansları belirgin biçimde artmış; modeller, azınlık sınıfları daha iyi ayırt etmeye başlamıştır.Modellerin tahmin mekanizmalarını açıklamak ve öne çıkan özellikleri belirlemek amacıyla SHAP analizleri gerçekleştirilmiş, özniteliklerin modele katkısı görselleştirilmiştir.Aşağıda, Random Forest ve XGBoost modellerine ait SHAP grafiklerine yer verilmiştir.

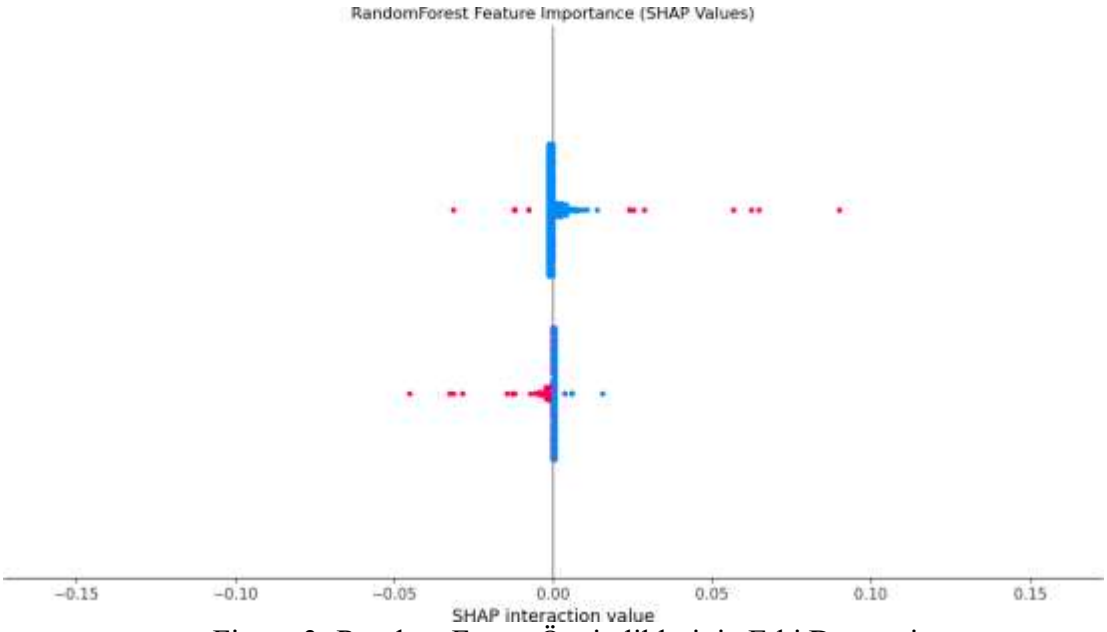


Figure 3: Random Forest Özniteliklerinin Etki Derecesi

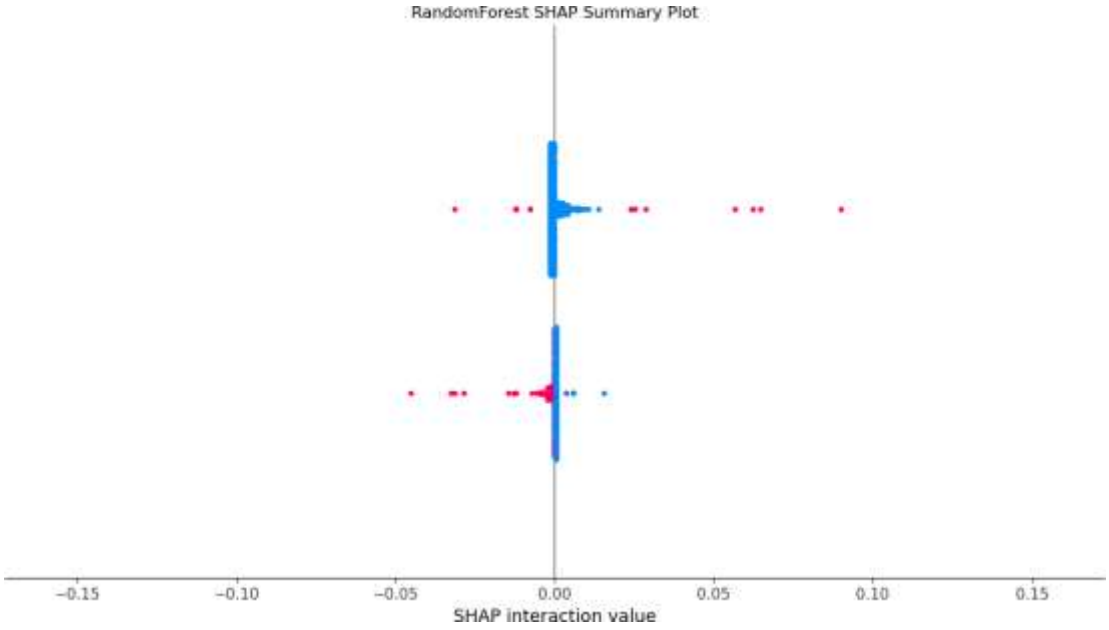


Figure 4: Random Forest Öznitelik Etki Dağılımı

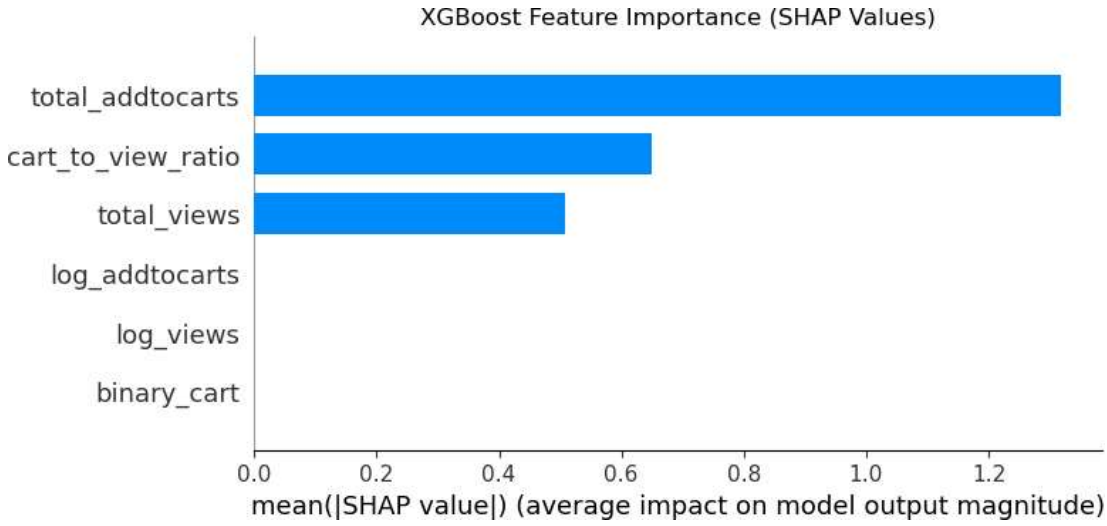


Figure 5: XGBOOST Özniteliklerinin Etki Derecesi

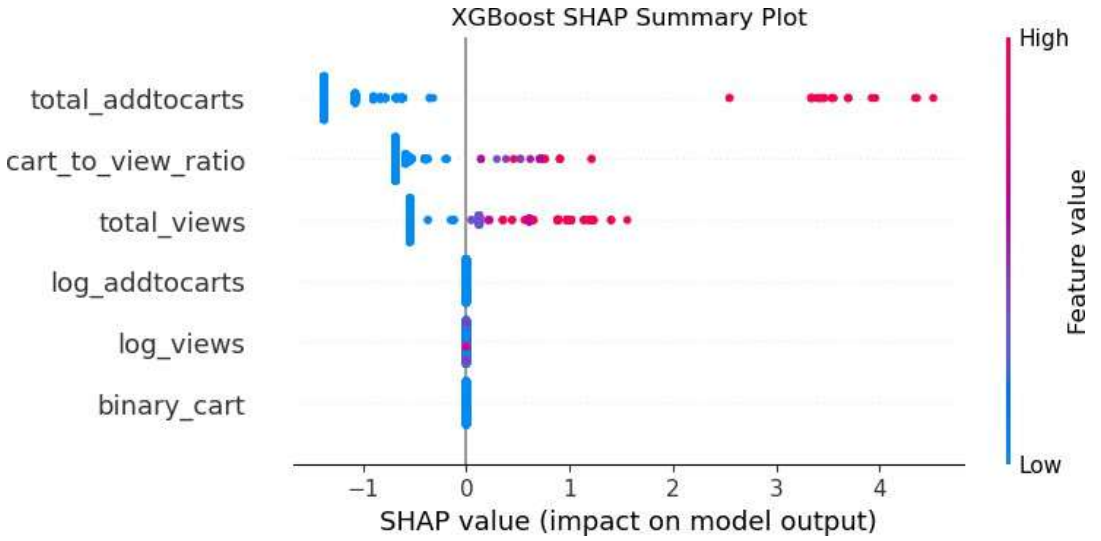


Figure 6: XGBOOST Öznitelik Etki Dağılımı

SHAP analizi sonucunda, Figure 3, Figure 4, Figure 5 ve Figure 6'da yer alan **cart_to_view_ratio**, **binary_cart** ve **log_addtocarts** değişkenlerinin, hedef değişken üzerinde önemli etkiler oluşturduğu görülmektedir.Ayrıca segment bazlı analizler, düşük etkileşimli ve yüksek etkileşimli kullanıcılar arasında belirgin davranış farklılıkları olduğunu ortaya koymuştur.

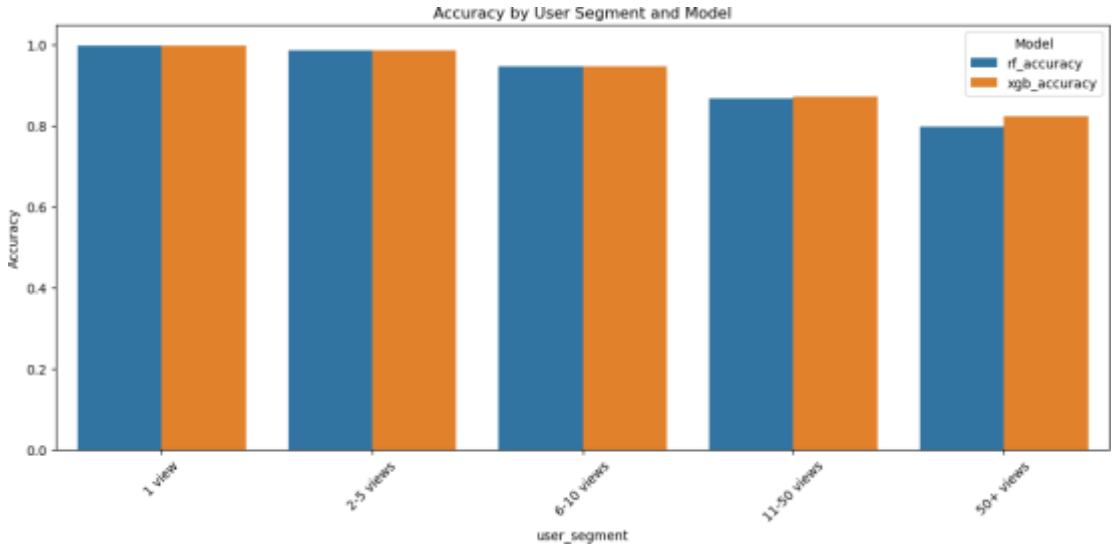


Figure 7: Segment Performans Grafiği

Çalışmanın ikinci adımı olan **Zaman Serisi ve Davranışsal Analiz** aşamasında, kullanıcı davranışlarının zamana bağlı değişimlerini incelemek amacıyla zaman serisi modelleme yöntemleri uygulanmıştır.**Prophet**,

ARIMA ve **Holt-Winters** gibi klasik zaman serisi modelleriyle birlikte, **LSTM (Long Short-Term Memory)** modeli de test edilmiştir.Modellerin hata metriklerine göre performans durumları **Figure 8**'de sunulmuştur.

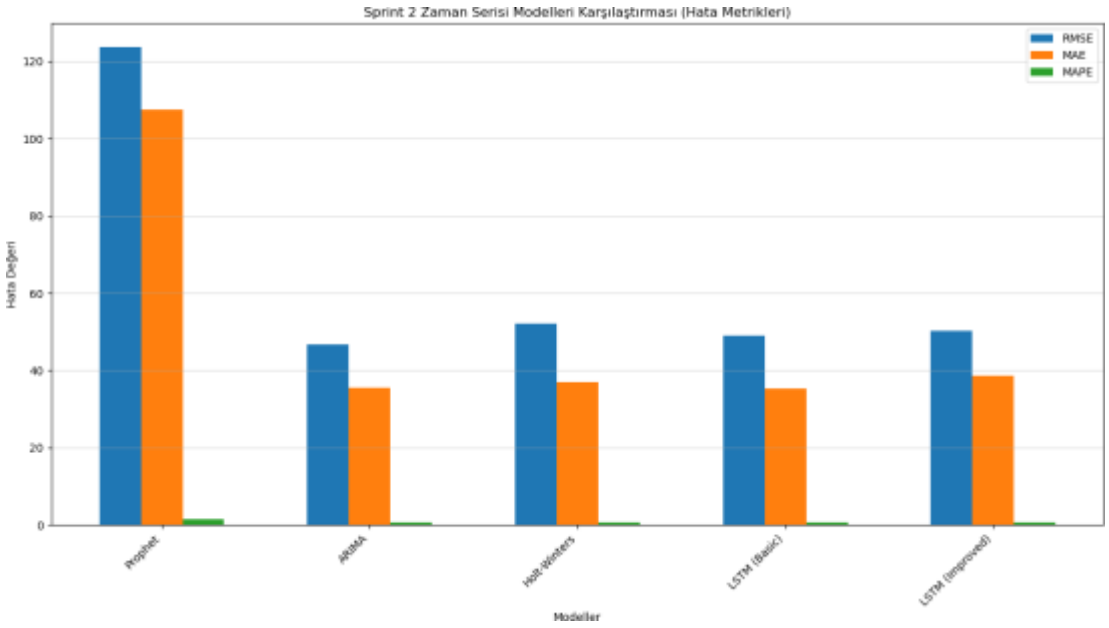


Figure 8: Zaman Serisi Modellerinin Hata Metriklerine Göre Karşılaştırılması

Figure 8'de görüldüğü gibi, **Prophet** modeli diğer modellere kıyasla daha yüksek hata metrikleri verirken, **ARIMA** modeli en düşük **RMSE**, **MAE** ve **MAPE** değerleriyle öne çıkmıştır.Hata metriklerinden de anlaşılacağı üzere, en yüksek performansı **ARIMA** modeli göstermektedir. **LSTM** modeline ise hiperparametre optimizasyonu uygulanmıştır. Bu uygulama sonucunda **MAPE** değeri 61'den 46'ya düşürülerek modelin performansı iyileştirilmiştir.Bu durum **Figure 9** ile görselleştirilmiştir

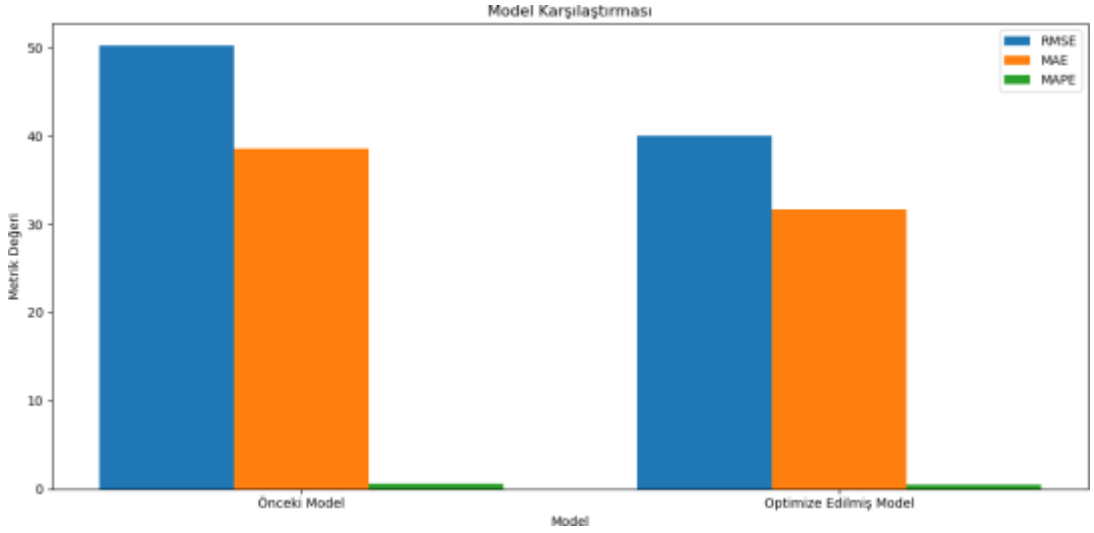


Figure 9: Hiperparametre Optimizasyonun LSTM Üzerindeki Etkisi

Bu işlemlerden sonra **Markov geçiş analizleri** gerçekleştirilmiştir. Analiz sonucunda, kullanıcıların **view**, **addtocart** ve **transaction** olayları arasındaki geçişlerini incelemek amacıyla Markov geçiş matrisleri oluşturulmuş; geçiş olasılıkları **Figure 10**'da görselleştirilmiştir.

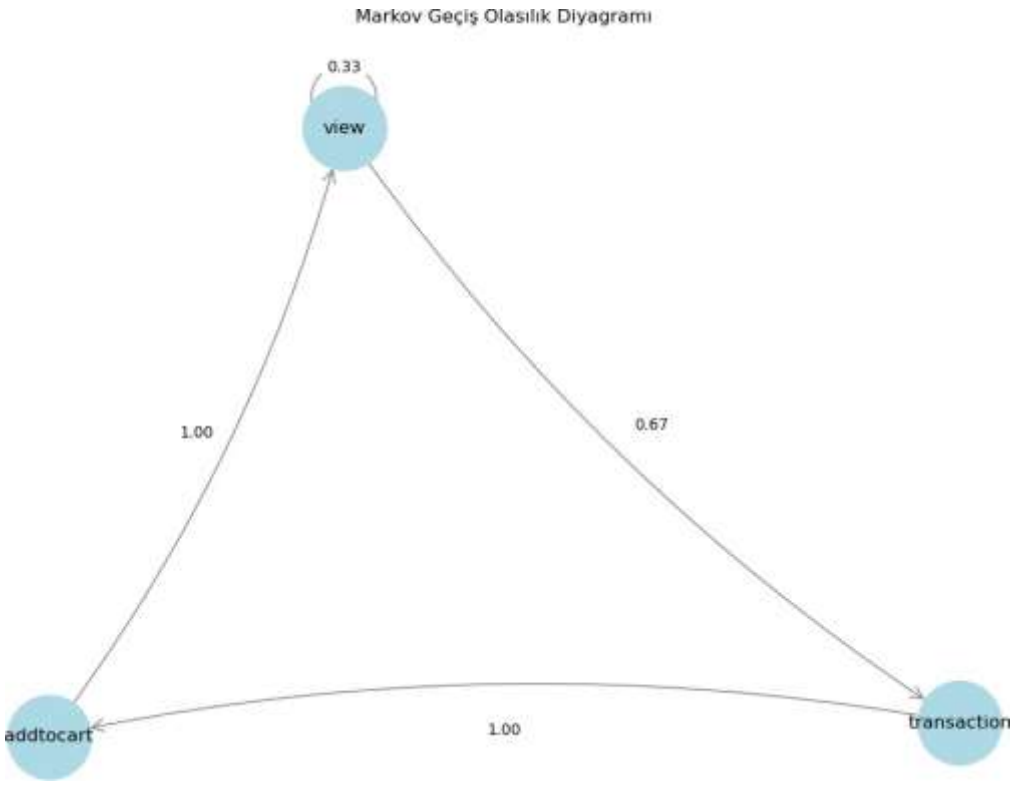


Figure 10: Markov Geçiş Diyagramı

Figure 10 incelendiğinde, diyagramda en yüksek olasılıkla gerçekleşen geçişin **addtocart → transaction** (olasılık = 1.00) olduğu gözlemlenmiştir. Bu durum, sepete ekleme işlemi gerçekleştiren kullanıcıların alışverişi tamamlama eğilimlerinin yüksek olduğunu göstermektedir.Öte yandan, **view → view** (olasılık = 0.33) geçişi, kullanıcıların önemli bir kısmının yalnızca ürünleri incelediğini ve herhangi bir aksiyon almadığını ortaya koymaktadır. **View → addtocart** geçişinin (olasılık = 1.00), diğer geçişlerle karşılaştırıldığında daha düşük sayıda gerçekleşmesi, kullanıcıların ürünü sepete ekleme motivasyonunun platform genelinde görece zayıf olduğunu işaret etmektedir.Bu analiz, platformda alışverişe geçişi artıracak stratejilerin geliştirilmesi gerektiğine işaret etmektedir.

Segmentasyon ve Öneri Sistemi Bulguları

Üçüncü aşamada, kullanıcı segmentasyonu **K-Means** algoritması ile gerçekleştirilmiş ve her segmentin davranışsal istatistikleri çıkarılmıştır. Bu bulgular **Figure 11**'de sunulmuştur.

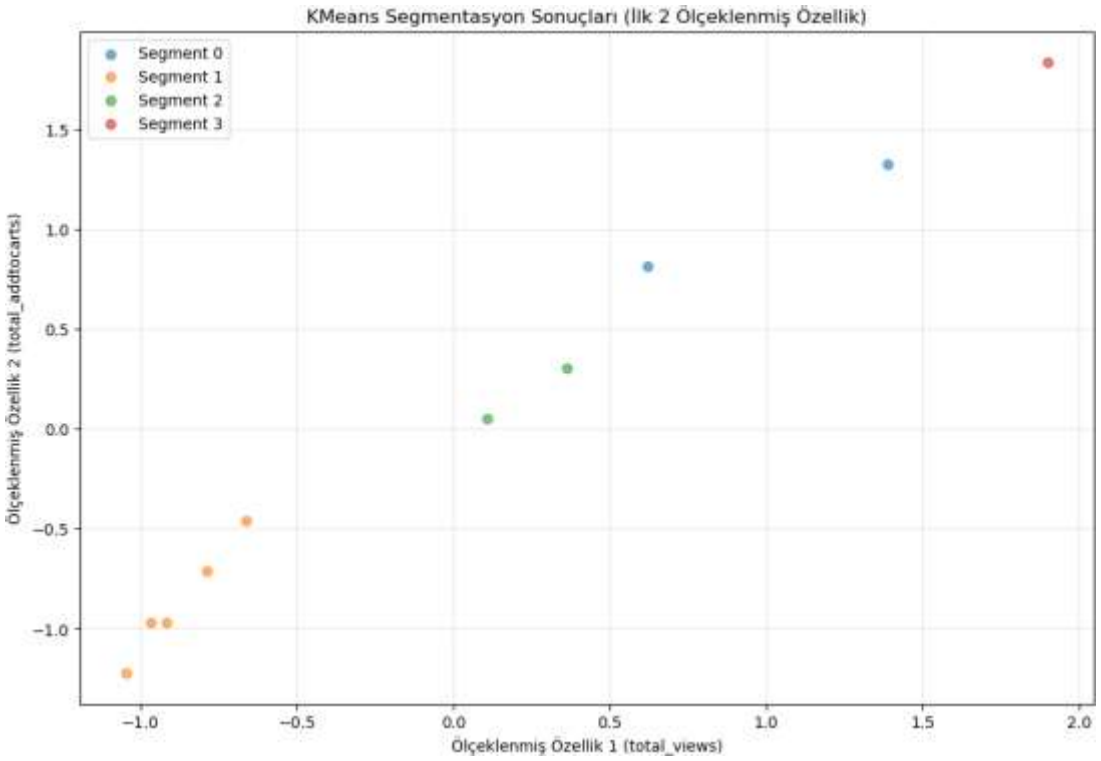


Figure 11: Müşteri Segmentleri

Figure 11 incelendiğinde, müşterilerin davranışlarına göre kümeler oluşturduğu görülmektedir.Grafiğe bakıldığında, **Segment 1** içerisindeki müşterilerin pasif olduğu, **Segment 2** içerisinde yer alanların ise potansiyel alıcılar olduğu yorumu yapılabilir.Ürün öneri sistemi, **Apriori algoritması** kullanılarak geliştirilmiş ve **support-confidence-lift** metrikleri dikkate alınarak güçlü öneri kuralları oluşturulmuştur.En sık birlikte satın alınan ürünler ve bu ürünlere ait öneri kuralları belirlenmiş ve ilgili `.csv` dosyalarına kaydedilmiştir.**FastAPI** çatısı altında geliştirilen tahmin ve öneri sistemi lokal ortamda çalıştırılmış ve segment bazlı önerilerin doğruluğu test edilmiştir.Modellerin başarı durumlarını gösteren tablolar aşağıda sunulmuştur.

Tablo 1. Makine öğrenmesi modelleri performans karşılaştırması

Model	F1 Score (Original)	F1 Score (SMOTE)	ROC AUC (Original)	ROC AUC (SMOTE)
Logistic Regression	0.13	0.38	0.96	0.96
Random Forest	0.12	0.41	0.96	0.96
XGBoost	0.27	0.40	0.60	0.96
Naive Bayes	0.41	0.39	0.96	0.96

Table 1 tablosu, Figure 1 ile de grafiklerini gösterdiğimiz makine öğrenmesi modellerinin SMOTE öncesi ve sonrası F1 skoru ve ROC AUC performanslarını göstermektedir. Table 1'de de görüldüğü gibi SMOTE uygulaması, dengesiz veri setine karşı model performansını iyileştirmiştir.

Table 2: Zaman serisi modelleri performans karşılaştırması

Model	RMSE	MAE	MAPE
Prophet	125	108	2.2
ARIMA	47	36	0.8
Holt-Winters	52	37	0.9
LSTM (Basic)	49	35	1.0
LSTM (Improved)	50	38	0.7

Table 2 tablosu, Figure 8 ile grafiklerini gösterdiğimiz zaman serisi modellerinin tahmin performanslarını RMSE, MAE ve MAPE hata metriklerine göre karşılaştırmaktadır. Table 2'de görüldüğü gibi ARIMA modeli genel olarak en düşük hata değerleriyle öne çıkmaktadır.

Table 3: Optimize edilmiş ve optimize edilmemiş LSTM

Model	RMSE	MAE	MAPE
Optimize edilmemiş Model	50	38	0.9
Optimize Edilmis Model	40	32	0.4

Son olarak Table 3 tablosunda, Figure 9 ile de gösterdiğimiz optimize edilmiş LSTM modelinin, hata metrikleri açısından önceki model ile karşılaştırmalı performansı sunulmaktadır. Görüldüğü gibi, optimize edilmiş model, tüm metriklerde daha düşük hata oranları ile üstünlük sağlamıştır.

4. SONUÇ, TARTIŞMA VE ÖNERİLER

Sonuç olarak bu çalışma, e-ticaret platformlarında müşteri davranışlarının hem zamansal hem de davranışsal açıdan çok boyutlu bir yaklaşımla analiz edildiği kapsamlı bir veri bilimi süreci ortaya koymuştur. RetailRocket veri seti temel alınarak gerçekleştirilen analizlerde, kullanıcıların geçmiş etkileşimlerinden türetilen davranışsal

özniteliklerin yanı sıra, olaylar arasındaki geçişlerin ve zaman serilerinin incelenmesi yoluyla müşteri davranışlarının detaylı bir resmine ulaşılmıştır.

Çalışmada, klasik makine öğrenmesi algoritmaları (Random Forest, Logistic Regression, XGBoost, Naive Bayes) ile derin öğrenme yöntemleri (LSTM) ve zaman serisi modelleri (Prophet, ARIMA, Holt-Winters) kullanılarak tahmin performansları karşılaştırılmış, SMOTE gibi veri dengeleme yöntemlerinin modellerin başarımına etkisi değerlendirilmiştir.

Ayrıca, SHAP ve segment bazlı analizler ile modellerin karar verme süreçlerinin açıklanabilirliği sağlanmış, belirli kullanıcı gruplarının davranış kalıpları ortaya konmuştur. Zaman serisi analizleri sayesinde, hem olaylar arasındaki geçişler hem de alışveriş davranışlarının günlük, haftalık ve sezonluk eğilimleri belirlenerek, geleceğe yönelik tahminler güçlendirilmiştir.

Kullanıcı segmentasyonu ve Apriori algoritmasıyla oluşturulan öneri sistemi, hem teorik hem de uygulamalı bir çerçevede değerlendirilerek e-ticaret platformlarının müşteri deneyimini iyileştirme ve satış artırma potansiyelini somut olarak ortaya koymuştur.

Gelecek çalışmalar için, farklı sektörlerden veri setlerinin kullanılması, derin öğrenme tabanlı öneri sistemlerinin geliştirilmesi, gerçek zamanlı tahmin ve öneri entegrasyonu, kullanıcıların davranışsal motiflerini daha ayrıntılı modellemek için gelişmiş zaman serisi ve sıralı derin öğrenme modellerinin (ör. Transformer tabanlı modeller) denenmesi önerilmektedir.

Ayrıca, veri gizliliği ve etik konularının da derinlemesine ele alınması, müşteri odaklı veri analizlerinin sürdürülebilir ve sorumlu bir şekilde gerçekleştirilmesine katkı sağlayacaktır.

KAYNAKLAR

Bhade, P., Waghmare, V., & Wankhede, V. (2018). Customer segmentation using K-means clustering algorithm. *International Journal of Engineering Research and Technology*, *7*(4).

Jannach, D., Zanker, M., & Schnabel, T. (2023). Session-based recommender systems in e-commerce. *ACM Computing Surveys*, *55*(1), 1–36.

Singh, P., Singh, K. K., & Singh, A. (2019). Machine learning-based classification and segmentation techniques for CRM. *International Journal of Information and Decision Sciences*, *11*(1), 103–116.

Singla, Z., Randhawa, S., & Jain, S. (2017). Sentiment analysis of customer product reviews using machine learning. *2017 International Conference on Intelligent Computing and Control (I2C2)*, 530–535. https://doi.org/10.1109/I2C2.2017.8321910

Zhou, W. (2025). RetailRocket veri seti kullanılarak müşteri satın alma davranışlarının tahmini: Genetik algoritmalar ve XGBoost uygulaması. *Journal of Data Analytics*, *5*(1), 12–25.