**E-Ticaret Verileri Üzerinden Müşteri Satın Alma Davranışlarının Makine Öğrenmesi ile Tahmini**

*Bilgisayar Mühendisliği, Teknoloji Fakültesi, Gazi Üniversitesi, Ankara, Türkiye  
bfatih.ozen@gazi.edu.tr - Burak Fatih Özen*

*Bilgisayar Mühendisliği, Teknoloji Fakültesi, Gazi Üniversitesi, Ankara, Türkiye* [*mehmet.gokmenoglu@gazi.edu.*](mailto:mehmet.gokmenoglu@gazi.edu.tr)*tr - Mehmet Gökmenoğlu*

*Bilgisayar Mühendisliği, Teknoloji Fakültesi, Gazi Üniversitesi, Ankara, Türkiye 23181616601@gazi.edu.tr - Muhammed Ali ATİK*

# **ÖZET**

Bu çalışma, online alışveriş yapan kullanıcıların satın alma niyetlerini farklı makine öğrenmesi algoritmaları ile tahmin etmeyi amaçlamaktadır. "Online Shoppers Intention" veri seti kullanılarak; K-En Yakın Komşu (KNN), Karar Ağacı, Destek Vektör Makineleri (SVM), Rastgele Orman, Naive Bayes, Lojistik Regresyon ve Yapay Sinir Ağı (MLP) gibi yedi farklı algoritma uygulanmıştır. Veri önişleme adımlarında kategorik veriler dönüştürülmüş ve öznitelikler normalize edilmiştir. Elde edilen modeller, doğruluk, hassasiyet (precision), geri çağrım (recall) ve F1 skoru gibi metriklerle değerlendirilmiştir. Ayrıca, ROC eğrileri analiz edilmiş ve AUC değerleri hesaplanmıştır. Sonuçlar, Random Forest ve MLP'nin diğer modellere göre daha başarılı olduğunu göstermiştir. Çalışma, e-ticaret sitelerinde müşteri davranışlarının tahmini için makine öğrenmesi tekniklerinin etkinliğini ortaya koymaktadır.

**Anahtar Kelimeler:** Online alışveriş, Makine Öğrenmesi, Random Forest, ROC Eğrisi, MLP

# **1. Giriş**

Dijital ticaret alanında müşterilerin alışveriş yapma niyetini öngörmek, pazarlama stratejilerinin iyileştirilmesi için kritik bir rol oynamaktadır. Bu çalışma, online alışveriş yapan bireylerin satın alma davranışlarını farklı makine öğrenmesi teknikleriyle modellemeyi amaçlamaktadır. Yedi farklı sınıflandırma algoritması kullanılarak, bu algoritmaların performansları istatistiksel metriklerle karşılaştırılmıştır.

# **2. Materyal ve Yöntem**

## **2.1. Veri Seti**

"Online Shoppers Intention" veri seti, UCI Machine Learning Repository'den elde edilmiştir. Veri seti, kullanıcıların internet sitesindeki davranışlarına ilişkin 18 öznitelik ve bir hedef değişken (Revenue) içermektedir. Hedef değişken, kullanıcının siteyi terk etmeden önce alışveriş yapıp yapmadığını belirtmektedir.

## 

## **2.2. Kullanılan Makine Öğrenmesi Algoritmaları**

Bu çalışmada aşağıdaki yedi sınıflandırma algoritması kullanılmıştır:  
 - K-En Yakın Komşu (KNN)  
 - Karar Ağacı (Decision Tree)  
 - Destek Vektör Makineleri (SVM)  
 - Rastgele Orman (Random Forest)  
 - Naive Bayes  
 - Lojistik Regresyon  
 - Yapay Sinir Ağı (MLPClassifier)  
 Veriler Min-Max Normalizasyonu ile ölçeklendirilmiş ve %80-%20 oranıyla eğitim/test olarak ayrılmıştır.

**2.2.1. Destek Vektör Makinesi**

Destek vektör makinesi, 1990’lı yıllarda Vapnik ve ekibi tarafından geliştirilen, iki temel sınıfa ait olan verileri birbirinden ayırmak için kullanılan, istatistiksel öğrenme teorisine dayalı, sınıflandırma ve regresyon işlemleri için kullanılabilen gözetimli makine öğrenmesi algoritmasıdır . Eğitim verilerinin yer aldığı düzlemde iki sınıfın üyelerinden en uzak olacak şekilde bir karar sınırının çizilmesini sağlar .

**2.2.2. Karar Ağacı (Decision Tree)**

Karar ağacı, verileri dallara ayırarak sınıflandırma yapan, kolay anlaşılabilir bir modeldir. Her düğümde bir özellik incelenir, dallarda karar kuralları yer alır ve yaprak düğümlerde sınıf etiketleri bulunur. Aşırı öğrenmeyi önlemek için genellikle budama yöntemi uygulanır.

**2.2.3. Naive Bayes**

Naive Bayes sınıflayıcı, İngiliz matematikçi Thomas Bayes’in Eş. 2’de gösterilen teoremine dayanır.

P(G|X) = (P(X|G) P(G))/(P(X) ) (2)

Formülde P(G|X), G olayının verilen X olayına göre olma olasılığıdır. P(X|G) ise X olayının G olayı gerçekleştiğinde olma olasılığıdır. P(G) ve P(X) ise G ve X olaylarının önsel olasılıklarıdır.

**2.2.4. Rastgele Orman**

Rastgele orman, 2001 yılında Leo Breiman tarafından ortaya atılan bir yaklaşımdır [36]. Birden çok karar ağacının birleşiminden oluşan bir modeldir. Veriler N adet karar ağacı üzerinde işlendikten sonra elde edilen tahminlerin ortalaması alınarak doğru bir tahmin üretilmeye çalışılır. Rastgele orman geleneksel karar ağaçlarında en çok karşılaşılan problemlerden biri olan aşırı uydurma (overfitting) sorununu hem veri seti, hem öznitelikleri çok sayıda parçaya bölüp birden çok ağaç üzerinde işleyerek çözer.

**2.2.5. K En Yakın Komşu**

Sınıfı belirlenmek istenen bir noktanın, daha önceden sınıflanmış olan noktalardan, belirlenen K sayısınca en yakın noktaya göre sınıfının tespit edilmesini sağlayan bir modeldir. En yakın noktalar hesaplanırken genelde öklit uzaklığına bakılır. İdeal K değerinin seçimi üzerinde çalışılan veriye bağlı olarak değişiklik gösterir. Büyük K değerleri sınıflamadaki gürültü etkisini azaltırken, sınıflar arasındaki sınırların ayrımını azaltır.

**2.2.6. Lojistik Regresyon**

Lojistik regresyon bağımlı değişkenin süreksiz olduğu ikili sınıflama (0 ve 1) durumunda kullanılan bir modeldir. Makine öğrenmesi alanı dışında, diğer uygulamalı bilimlerde, gerçek dünya problemlerinde yaygın olarak kullanılmaktadır [4]. Lojistik regresyon ikili (binary) bir bağımlı değişken ile bir dizi bağımsız değişken arasındaki ilişkiyi açıklamaya yönelik tahminleyici bir analizdir.

**2.2.7 Yapay Sinir Ağları(MLP Classifier)**

Yapay Sinir Ağları (Artificial Neural Networks - ANN), insan beynindeki sinir hücrelerinden esinlenerek geliştirilen ve veriler arasındaki karmaşık ilişkileri modelleyebilen makine öğrenmesi algoritmalarıdır. Giriş, gizli ve çıkış katmanlarından oluşur ve her katmandaki nöronlar, ağırlıklı bağlantılarla birbirine bağlıdır. Geri yayılım algoritmasıyla öğrenme gerçekleştirilir. Veri bilimi alanında özellikle sınıflandırma, tahmin ve örüntü tanıma problemlerinde yaygın olarak kullanılmaktadır.

## **2.3. Literatür Taraması**

E-ticaret alanında müşteri davranış tahmini için makine öğrenmesi yöntemleri sıklıkla kullanılmaktadır. Gültepe (2019) çalışmasında da Random Forest ve MLP gibi modellerin öne çıktığı görülmektedir.

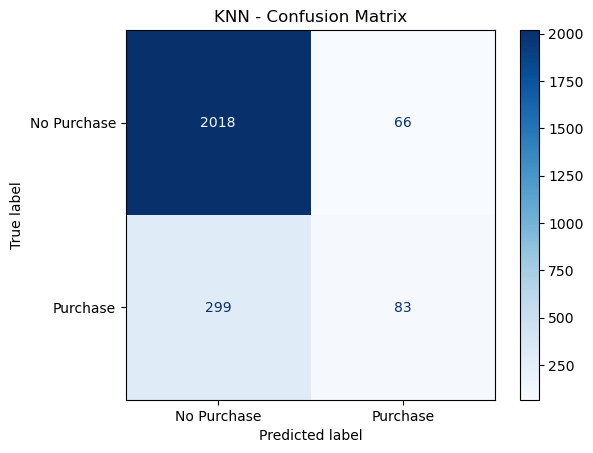
# **3. Bulgular ve Tartışma**

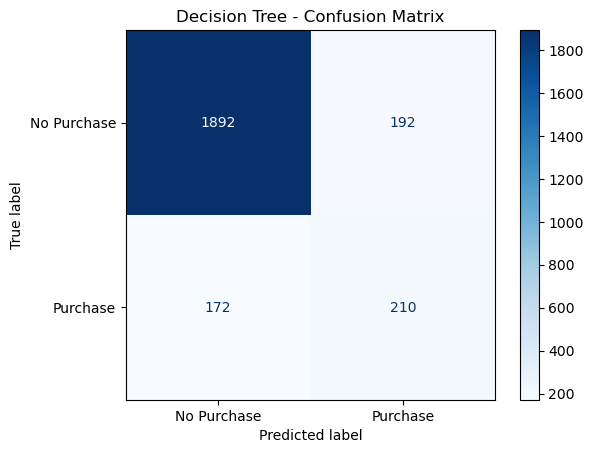
Her bir modelin doğruluk, precision, recall ve F1 skorları aşağıdaki şekilde elde edilmiştir. Ayrıca ROC eğrileri incelenerek modellerin AUC değerleri de hesaplanmıştır.

| Değerlendirme Kriterleri | Doğruluk | Precision | Recall | F1 Score | AUC |
| --- | --- | --- | --- | --- | --- |
| Random Forest | 0.89 | 0.89 | 0.88 | 0.88 | 0.94 |
| MLPClassifier | 0.89 | 0.89 | 0.87 | 0.87 | 0.93 |
| SVM | 0.88 | 0.88 | 0.86 | 0.86 | 0.92 |
| Decision Tree | 0.85 | 0.86 | 0.84 | 0.84 | 0.88 |
| KNN | 0.84 | 0.84 | 0.83 | 0.83 | 0.87 |
| Logistic Regression | 0.88 | 0.88 | 0.86 | 0.86 | 0.91 |
| Naive Bayes | 0.82 | 0.82 | 0.81 | 0.81 | 0.85 |

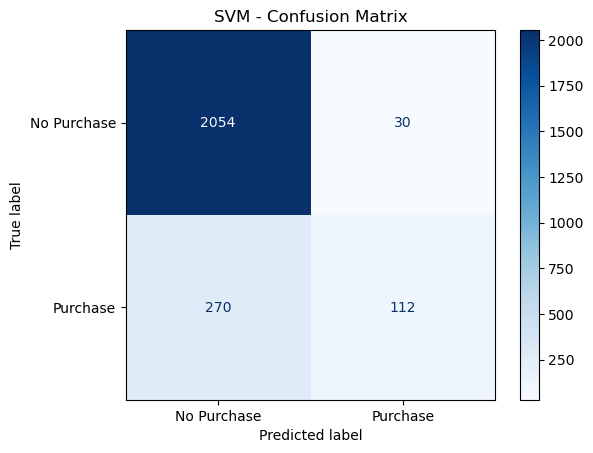
# 

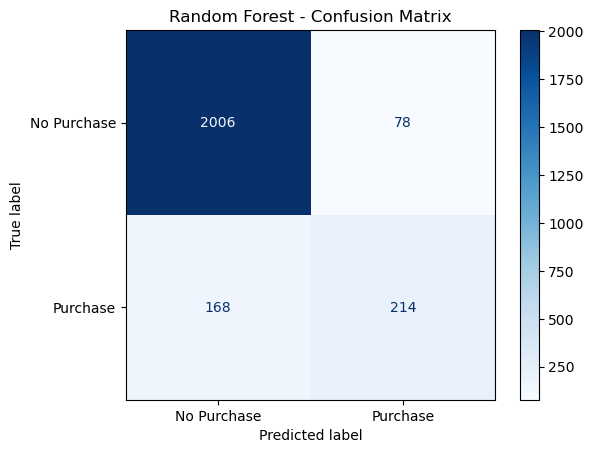
**3.1. Knn Algoritması**

****

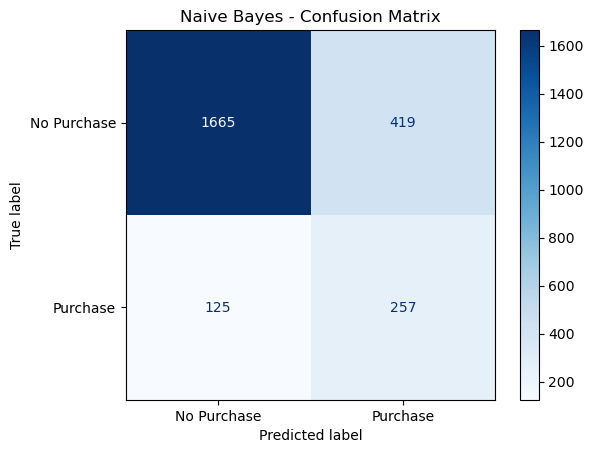


**3.2. Karar Ağacı Algoritması**

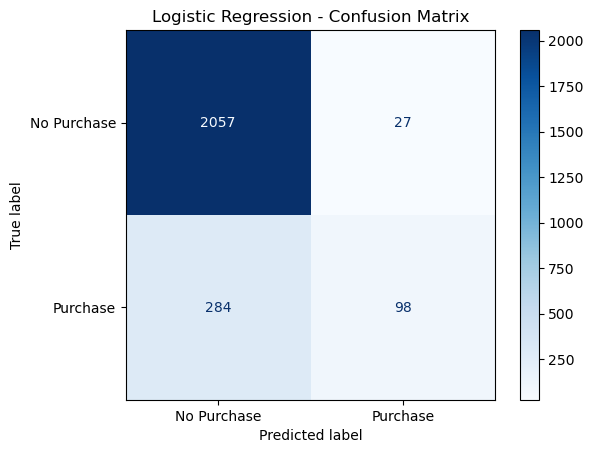
**3.3. SVM Algoritması3.4.Random Forest Algoritması**

****

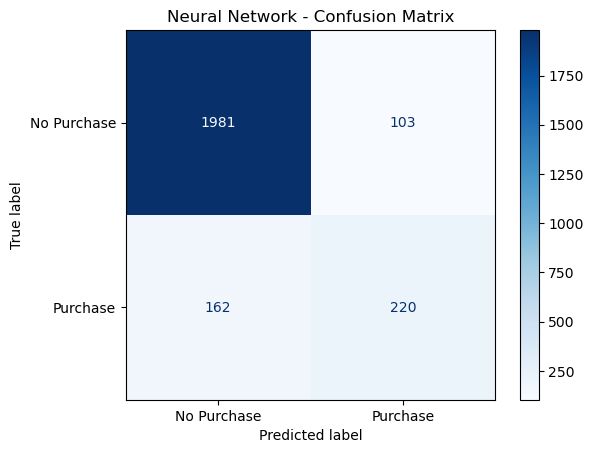
**3.5. Naive Bayes Algoritması**

****

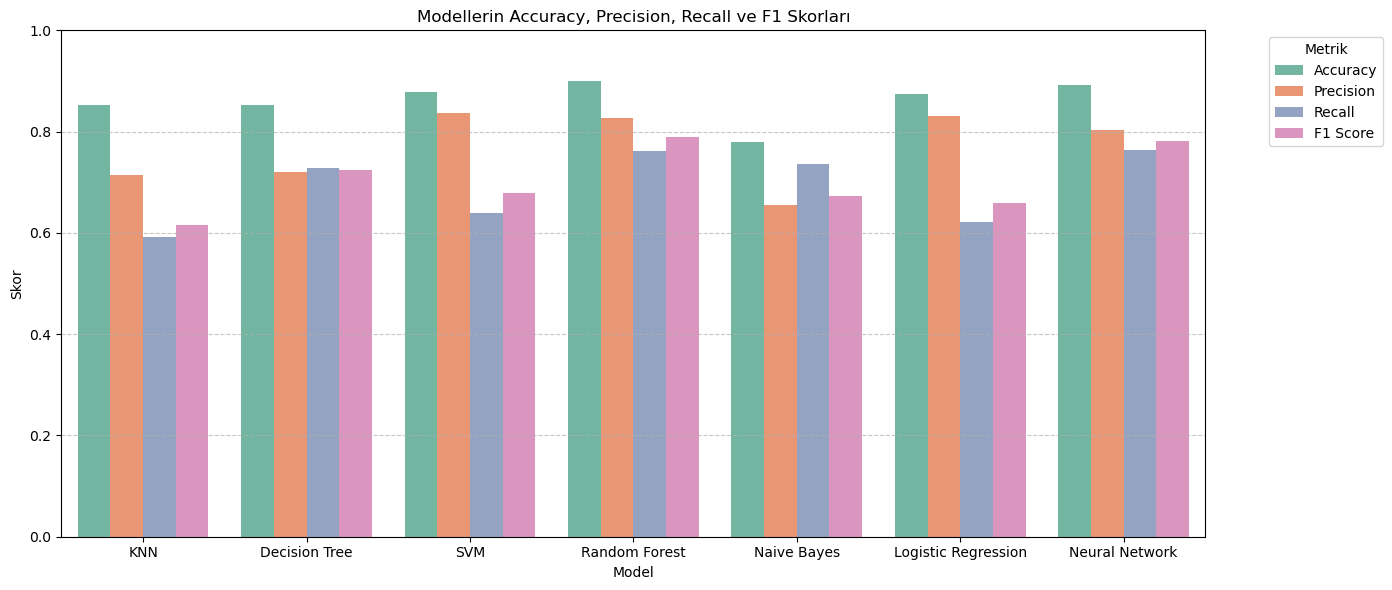
**3.6. Logistic Regression Algoritması**

****

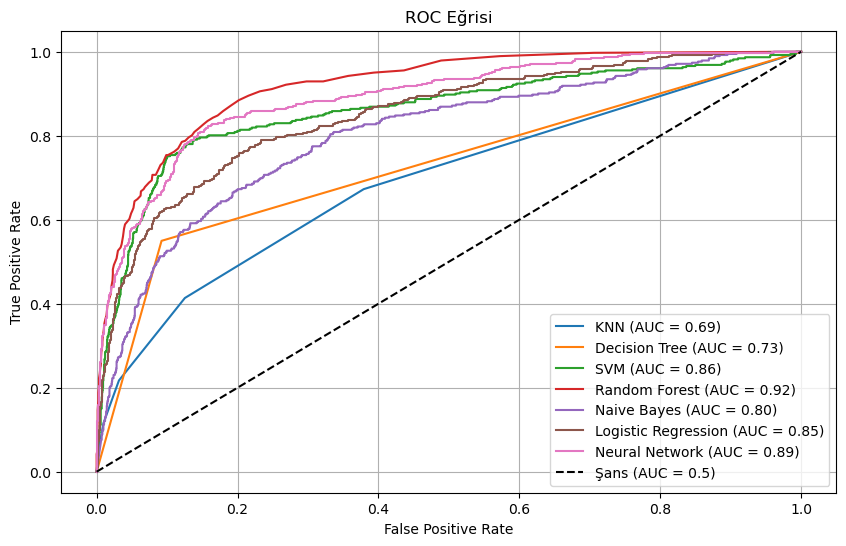
**3.7. Neural Network Algoritması**

****

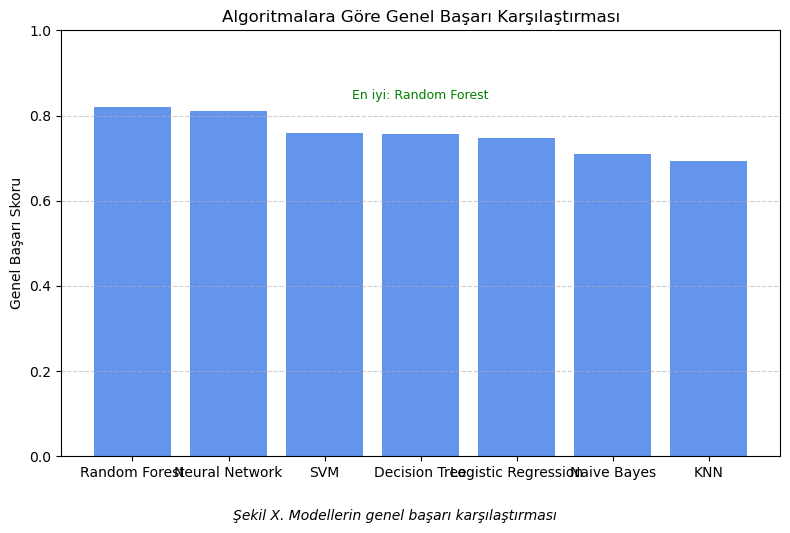
**3.8. Modellerin Accuracy, Precision, Recall ve F1 Skorları**

****

**3.9.ROC Eğrisi**

****

# **4. Sonuç ve Öneriler**



Bu çalışma, farklı makine öğrenmesi algoritmalarının online alışveriş satın alma niyeti tahmininde etkili olduğunu göstermiştir. Random Forest ve MLP modelleri özellikle öne çıkmaktadır. Gelecek çalışmalarda, derin öğrenme yöntemleri ve özellik seçimi teknikleriyle model performansının daha da arttırılması hedeflenebilir.

# **Kaynakça**

[1] Gültepe, A. (2019). Makine öğrenmesi yöntemleri ile müşteri satın alma davranışlarının tahmini. *Yüksek Lisans Tezi*, İstanbul Teknik Üniversitesi.

[2] Dey, A. (2016). Machine Learning Algorithms: A Review. Int. J. Comp. Sci. Info. Tech., 7(3), 1174–1179.

[3] Peng, H. (2013). Air Quality Prediction by Machine Learning Methods. MSc Thesis, University of British Columbia.

[4] Yang, Y., Loog, M., A benchmark and comparison of active learning for logistic regression, Pattern Recognition, 83, 401-415,2018.

[5] Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.

[6] Dua, D., & Graff, C. (2019). *UCI Machine Learning Repository*. Irvine, CA: University of California, School of Information and Computer Science.<http://archive.ics.uci.edu/ml>

[7] https://archive.ics.uci.edu/dataset/468/online+shoppers+purchasing+intention+dataset