**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.[6]

## 

## **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 makineleri (SVM), denetimli öğrenme modelleri kullanan bir makine öğrenimi algoritmasıdır. Bu algoritma, önceden tanımlanmış sınıflar, etiketler veya çıktılara göre veri noktaları arasındaki sınırları belirlemek için optimal veri dönüşümleri gerçekleştirerek karmaşık sınıflandırma, regresyon ve aykırı değer tespiti problemlerini çözer. Eğitim verilerinin yer aldığı düzlemde iki sınıfın üyelerinden en uzak olacak şekilde bir karar sınırının çizilmesini sağlar .[9]

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

Karar ağacı, hem sınıflandırma hem de regresyon görevlerinde kullanılan, parametrik olmayan bir denetimli öğrenme algoritmasıdır. Hiyerarşik bir ağaç yapısına sahiptir ve bu yapı; kök düğüm, dallar, iç düğümler ve yaprak düğümlerden oluşur.[10] 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**

Thomas Bayes tarafından bulunan koşullu olasılık hesaplama formülüdür. Bu yöntem rastgele değişken için olasılık dağılımı içinde koşullu olasılıklar ile marjinal olasılıklar arasındaki ilişkiyi ifade eder.

P(A/B) = (P(B/A).P(A))/P(B)

Naive Bayes sınıflandırıcısının temeli Bayes teoremine dayanır. Algoritma bir eleman için her durumun olasılığını hesaplar ve olasılık değeri en yüksek olana göre sınıflandırır. Az bir eğitim verisiyle çok başarılı işler çıkartabilir. Test kümesindeki bir değerin eğitim kümesinde gözlemlenemeyen bir değeri varsa olasılık değeri olarak 0 verir yani tahmin yapamaz.[11]

**2.2.4. Rastgele Orman**

Veriyi rastgele alt örneklemlerle alarak, her bir karar ağacını bu farklı veri setleri üzerinde eğitmeyi ifade eder. Daha sonra bu ağaçların tahminleri birleştirilerek (örneğin, sınıflandırmada çoğunluk oyu ile), genel tahmin yapılır [12]. Bu yöntem, modelin varyansını azaltarak doğruluğunu artırmaya yardımcı olur.

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.

**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, sonuç, olay veya gözlemin olasılığını tahmin ederek ikili sınıflandırma görevlerini yerine getiren denetimli bir makine öğrenmesi algoritmasıdır. Model, yalnızca iki olası sonuçtan (evet/hayır, 0/1 veya doğru/yanlış) biriyle sınırlı olan ikili bir çıktı üretir.

Lojistik regresyon, bir veya daha fazla bağımsız değişken arasındaki ilişkiyi analiz eder ve verileri ayrık sınıflara sınıflandırır. Bu yöntem, bir örneğin belirli bir kategoriye ait olup olmadığını matematiksel olarak tahmin eden kestirimsel modelleme uygulamalarında yaygın olarak kullanılır.[13]

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

Yapay sinir ağları (YSA), insan beyninin özelliklerinden olan öğrenme yolu ile yeni bilgiler türetebilme, yeni bilgiler oluşturabilme ve keşfedebilme gibi yetenekleri, herhangi bir yardım almadan otomatik olarak gerçekleştirebilmek amacı ile geliştirilen bilgisayar sistemleridir.[14]

Yapay sinir ağları insan beyni örnek alınarak, öğrenme sürecinin matematiksel olarak modellenmesi sonucu ortaya çıkmıştır. Beyindeki biyolojik sinir ağlarının yapısını, öğrenme, hatırlama ve genelleme kabiliyetlerini taklit eder. Yapay sinir ağlarında öğrenme işlemi örnekler kullanılarak gerçekleştirilir. Öğrenme esnasında giriş çıkış bilgileri verilerek, kurallar koyulur.

**2.3. Literatür Taraması**

E-ticaret sitelerinde müşteri davranışlarını tahmin etmek için kullanılan makine öğrenmesi yöntemleri, geniş bir araştırma alanıdır. Gültepe (2019), Random Forest ve MLP algoritmalarının yüksek doğruluk sağladığını belirtmiştir. SVM, karar sınırlarını optimize ederek özellikle doğrusal olmayan verilerde başarılı sonuçlar vermektedir [9]. Karar ağaçları, anlaşılır ve açıklanabilir sonuçlar sağlarken, IBM (2025) karar ağaçlarının işlevselliği hakkında kapsamlı bilgiler sunmaktadır [10]. Naive Bayes, özellikle küçük veri setlerinde etkili olurken, Hatipoğlu (2025) bu algoritmanın basit ama güçlü yapısına vurgu yapmıştır [11]. Rastgele Orman, çoklu karar ağaçlarının birleşimiyle daha doğru tahminler yapmaktadır ve Google Developers (2025) bu algoritmanın gücünü ortaya koymaktadır [12]. Yapay Sinir Ağları (YSA) ise biyolojik sinir ağlarının çalışma prensiplerine dayanarak öğrenme ve genelleme yapabilen güçlü bir modeldir [13].

.

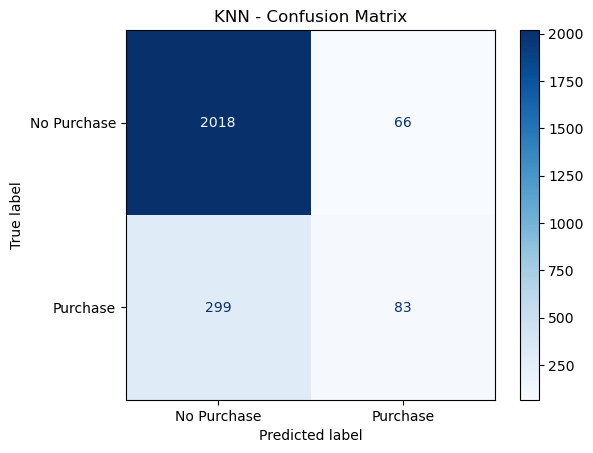
# **3. Bulgular ve Tartışma**

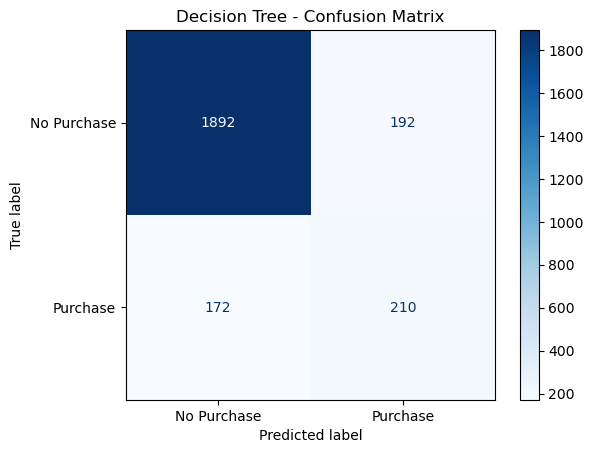
Her bir modelin doğruluk, precision, recall ve F1 skorları aşağıdaki şekilde elde edilmiştir. Ayrıca ROC eğrileri incelenerek modellerin AUC değerleri de hesaplanmıştır.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Değerlendirme Kriterleri | Doğruluk | Precision | Recall | F1 Score | AUC |
| Random Forest | 0.89 | 0.89 | 0.88 | 0.88 | 0.92 |
| MLPClassifier | 0.89 | 0.89 | 0.87 | 0.87 | 0.89 |
| SVM | 0.88 | 0.88 | 0.86 | 0.86 | 0.86 |
| Decision Tree | 0.85 | 0.86 | 0.84 | 0.84 | 0.73 |
| KNN | 0.84 | 0.84 | 0.83 | 0.83 | 0.69 |
| Logistic Regression | 0.88 | 0.88 | 0.86 | 0.86 | 0.85 |
| Naive Bayes | 0.82 | 0.82 | 0.81 | 0.81 | 0.80 |

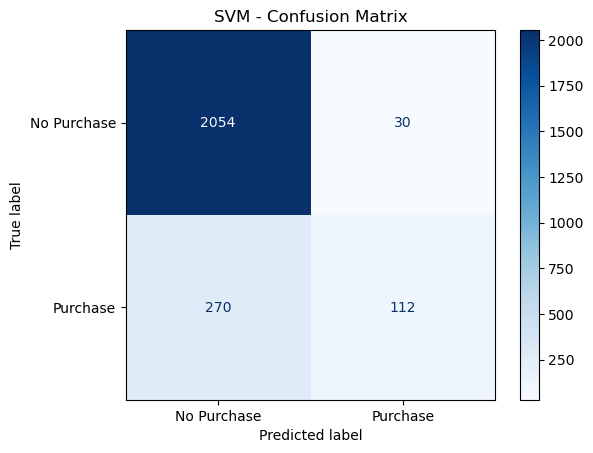
# 

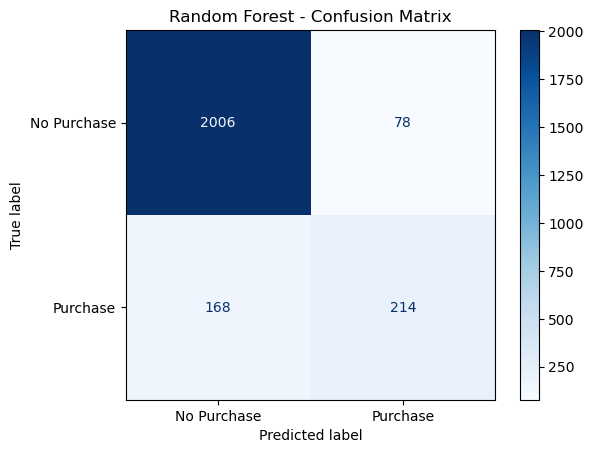
**3.1. Knn Algoritması**

****

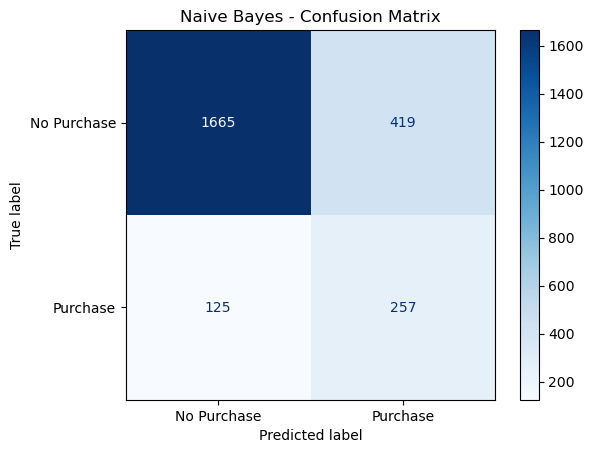


**3.2. Karar Ağacı Algoritması**

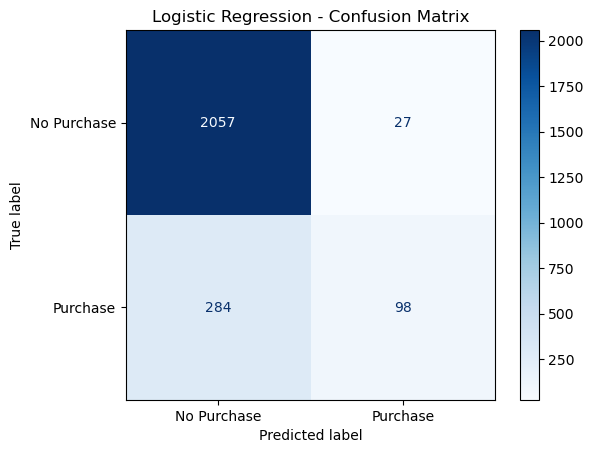
**3.3. SVM Algoritması3.4.Random Forest Algoritması**

****

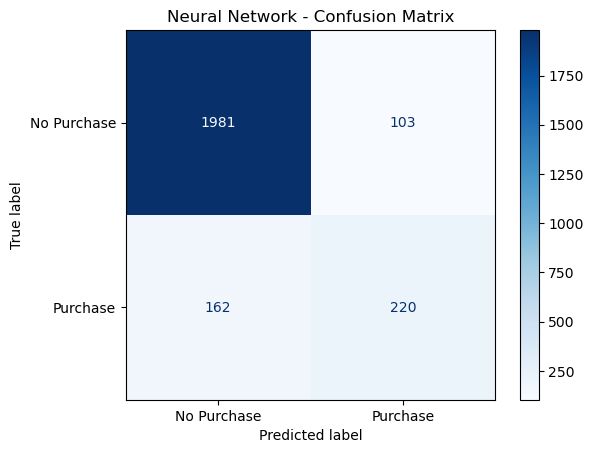
**3.5. Naive Bayes Algoritması**

****

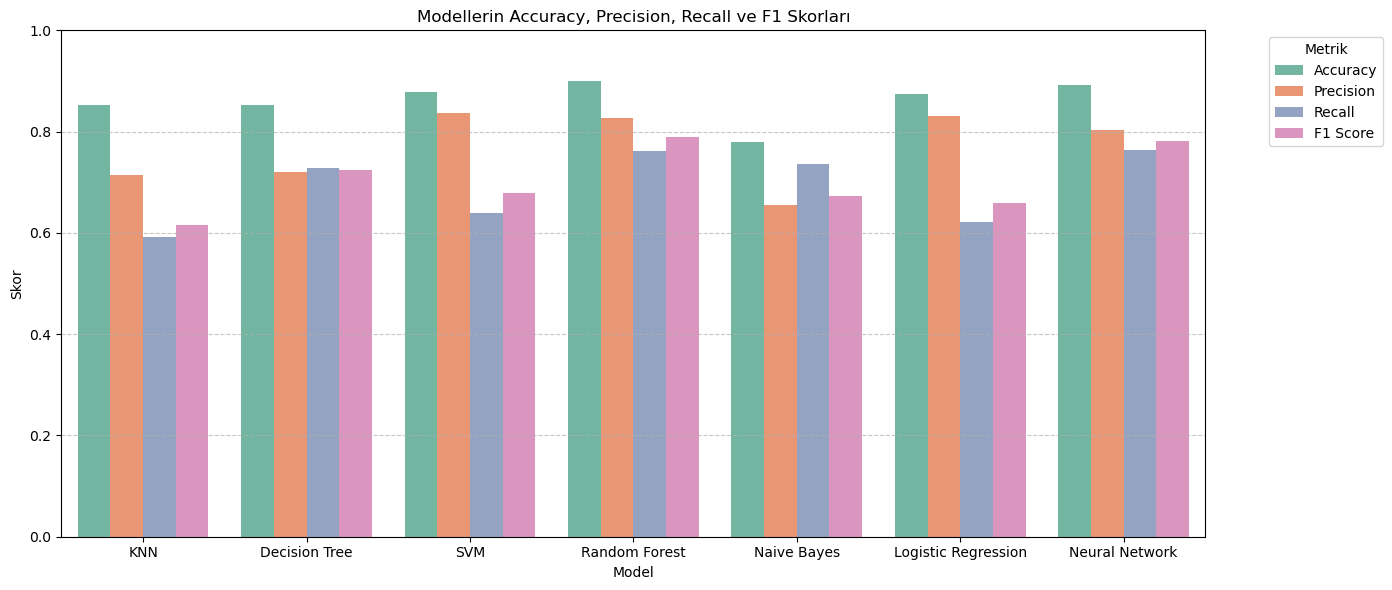
**3.6. Logistic Regression Algoritması**

****

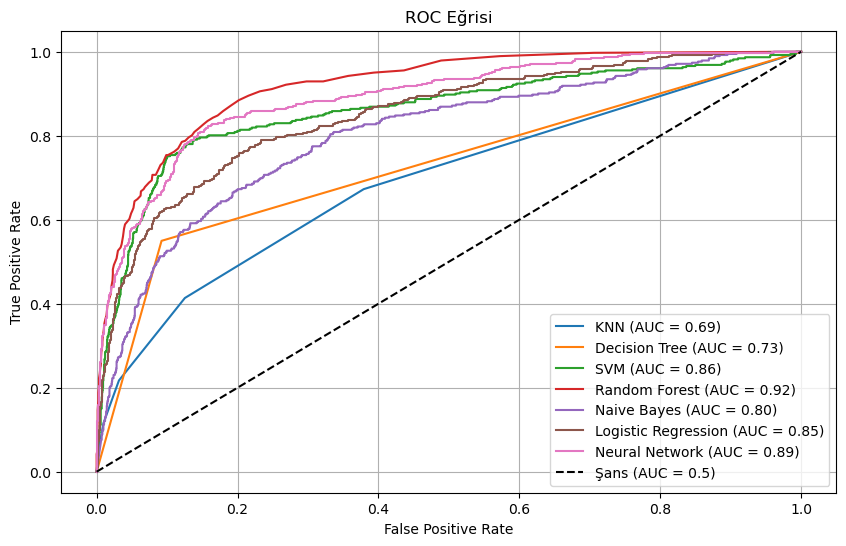
**3.7. Neural Network Algoritması**

****

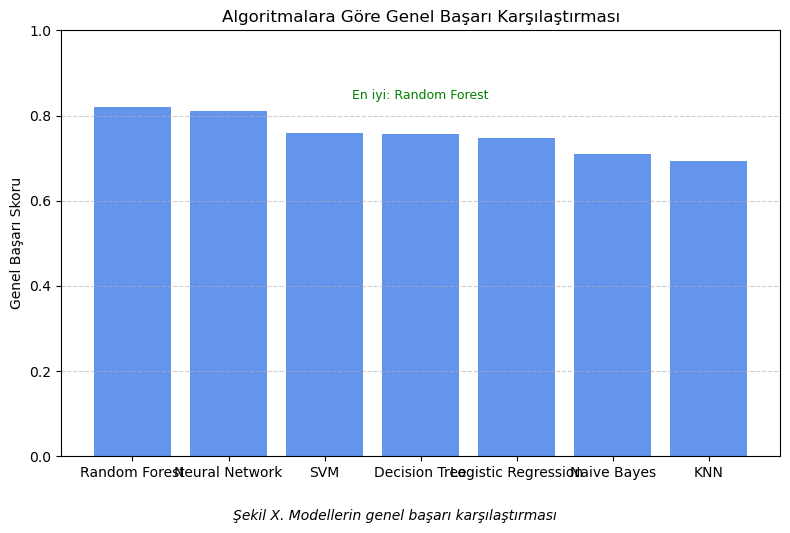
**3.8. Modellerin Accuracy, Precision, Recall ve F1 Skorları**

****

**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>

[8] <https://github.com/VeriBilimiProjesi/VeriBilimiProje/blob/main/VeriBilimi/VeriBilimi.py>

[9] *"What is Support Vector Machine (SVM)?,"* Spiceworks. [Çevrimiçi]. Ulaşılabilir: <https://www.spiceworks.com/tech/big-data/articles/what-is-support-vector-machine/> [Erişim: 4 Mayıs 2025].

[10] "Decision Trees," *IBM*. [Çevrimiçi]. Ulaşılabilir: <https://www.ibm.com/think/topics/decision-trees> [Erişim: 4 Mayıs 2025].

[11] E. Hatipoğlu, "Machine Learning Classification - Naive Bayes (Part-11)," *Medium*. [Çevrimiçi]. Ulaşılabilir: <https://medium.com/@ekrem.hatipoglu/machine-learning-classification-naive-bayes-part-11-4a10cd3452b4> [Erişim: 4 Mayıs 2025].

[12] G. Developers, "Random forests," Google Developers, 2025. [Çevrimiçi]. Mevcut: https://developers.google.com/machine-learning/decision-forests/random-forests. [Erişim: 4 Mayıs 2025].

[13] M. Sharma, "What is Logistic Regression?", Spiceworks, 24 Mayıs 2023. [Çevrimiçi]. Mevcut: https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/. [Erişim: 5 Mayıs 2025].

[14] E. Yıldırım, "Yapay Sinir Ağı (Artificial Neural Network) Nedir?", Veri Bilimi Okulu, 2 Mayıs 2020. [Çevrimiçi]. Mevcut: https://www.veribilimiokulu.com/yapay-sinir-agiartificial-neural-network-nedir/. [Erişim: 4 Mayıs 2025].