**Ayakkabı Fiyatı Tahmini**

**Shoes Price Prediction**

Mehmet Fırat Kahvecioğlu, 202805004

[202805004@ogr.cbu.edu.tr](mailto:202805001@ogr.cbu.edu.tr)

***Özet—* Bu projede makine öğrenmesi yöntemiyle ayakkabı fiyatlarının tahmin edilmesi amaçlanmıştır. Veri seti hazırlanmasında Trendyol sitesinden yararlanıldı. Trendyol sitesinden çekilen veri seti ham haliyle 1416 adettir. Ön işlemeden geçirilen veri seti üzerinde makine öğrenmesi yöntemleri uygulanmıştır.**

***Anahtar Kelimeler—* Alışveriş, Ayakkabı, Fiyat Tahmini, Makine Öğrenmesi, Trendyol**

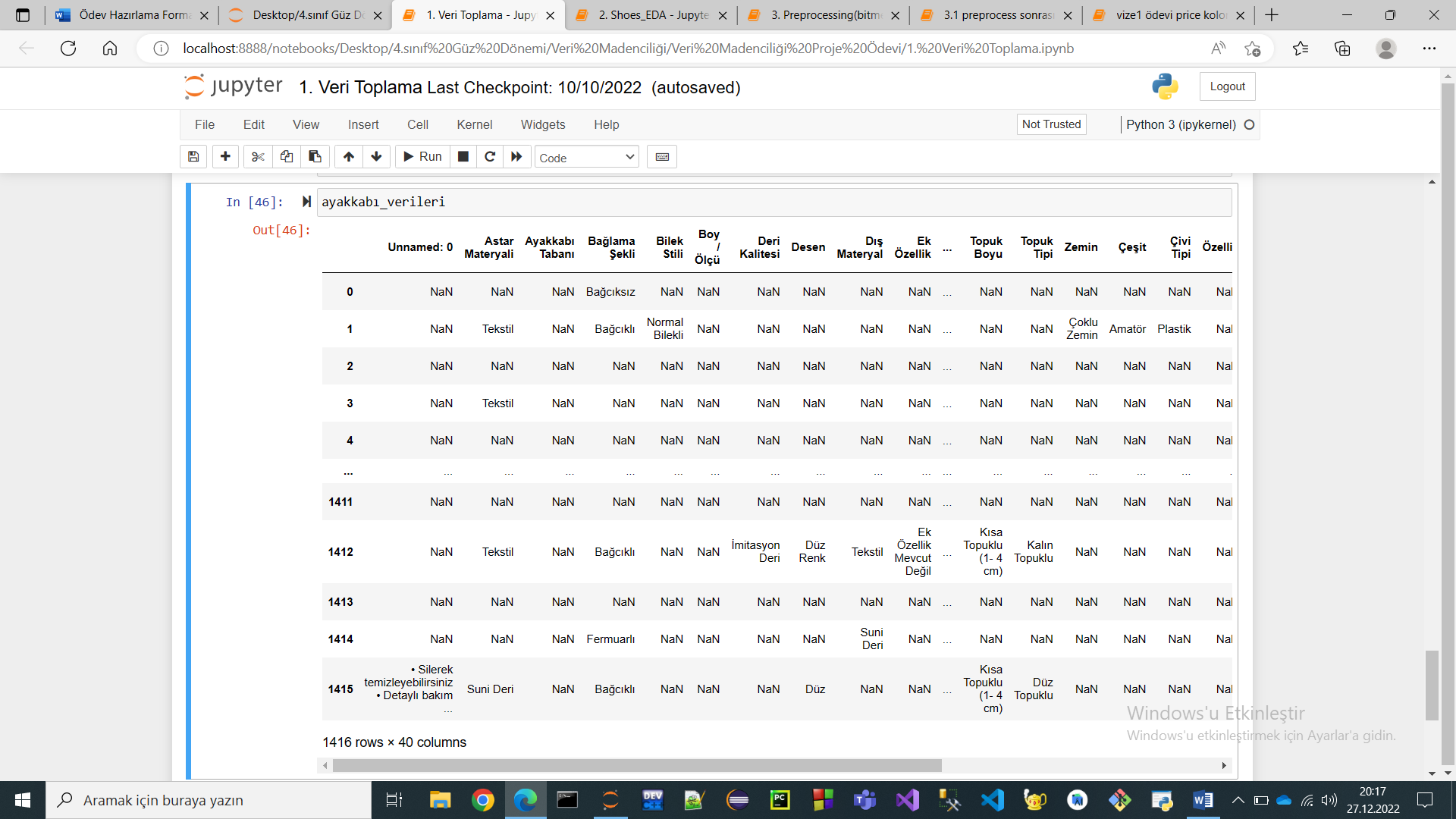
1. GİRİŞ

Günlük hayatta alışveriş yaparken alacağımız ürünün özelliklerine bakarak o ürünün fiyatını anlamaya çalışırız. Bu projede ise ayakkabı özelliklerine bağlı olarak bir ayakkabının fiyatı tahmin edilmeye çalışılmıştır. Daha önce başka ürünlerin özelliklerine bakarak fiyat tahmini yapan projelerin olma ihtimali çok yüksek ancak bu konuda detaylı bir literatür taraması yapılmamıştır. Bu projede veri seti ham hali ile 4 tane feauture’dan yani 4 tane özellikten oluşmaktadır. Bu 4 özellik Renk, Kullanım Alanı, Materyal ,Topuk Boyudur. Bu özellikler 40 özellik arasından önce preprocessing işlemlerinden geçirilmiş ve ardından makine öğrenmesi algoritmaları ile ayakkabının fiyatı tahmin edilmeye çalışılmıştır.

Random Forest Regression, makine öğrenimi alanında sıkça kullanılır ve karmaşık ilişkileri modeller. Birden fazla kara ağacı bir araya gelerek oluşturulur ve orman yapısı üzerinden tahminlerde bulunulur. [1] Ayakkabı fiyatlarını etkileyen çeşitli faktörler lineer olmayan bir ilişkide olabildiği için lineer olmayan ilişkileri daha iyi modelleyebilen Random Forest Regression yöntemi kullanılmıştır.

2. VERİLERİN ANLATIMI

Makine öğrenmesi algoritmaları ile yapılan ayakkabı fiyat tahmini projesinde veri setinin ham halinde 1416 adet ayakkabı verisi bulunmaktadır ve 40 adet kolondan yani featuredan oluşmaktadır. Bu veriler BeautifulSoup ve Selenium ve Chrome Web Driver kullanarak yapıldı. Selenium, web tarayıcılarını otomatize eder ve HTML üzerinden veri kazıma işlemi yapar. [4]. BeautifulSoup requests kütüphaneleri yardımı ile Trendyol [2] sitesinden çekilmiştir.



Görsel – 1 Trendyoldan Çekilen Veri Setinin DataFrame’e Dönüştürülmüş hali

3. ÖNERİLEN YÖNTEMLER

3.1. Preprocessing Süreci

Veri setinin 1416 veriden ve 40 kolondan oluştuğu söylenmişti. Trendyol sitesinden ayakkabı özelliklerine ait veriler çekilirken bazı ayakkabıların featureları diğer ayakkabılarda rastlanmadı. Bu yüzden bütün ayakkabılar için belirlenen 40 feauture ayakkabıların hepsinde bulunmamaktadır. Bazı ayakkabıların NaN olmayan feature sayıları 3 ya da 4 ü geçmemektedir. Bu yüzden veri ön işleme adımında ilk olarak datasette NaN sayısı çok fazla olan kolonlar özellik seçimi yapılarak [3] veri setinden drop edilmiştir.

Veri setinde ayakkabının rengine ait olan feature’da NaN özellik sayısı 0’dır. Ayakkabılarının renklerine göre ayrıldığı histogram grafik aşağıdaki şekilde gösterilmiştir.

A screen shot of a computer

Description automatically generated

Grafik 1 – Renklere göre ayakkabı sayıları

Bu işlemden sonra veri setindeki sayısal verilerin yanında bulunan değer ifade eden kategorik veriler o feature’dan atıldı. Örneğin Price kolonundaki “114 TL” olan ayakkabı fiyatı “114” olarak güncellenerek makine öğrenmesi algoritmalarının anlayabileceği hale getirildi.

LabelEncoder kategorik verileri sayısal değerlere dönüştürmek için kullanılır. Kategorik veriler sayısal olmayan veri türleri olduğu için bazı makine öğrenmesi algoritmalarıyla çalışırken metin tabanlı verileri sayısal verilere çevirmemiz gerekebilir. [5]. Projenin ön işleme aşamasında son olarak veri setindeki kolonlar her biri kendi içinde sayısal hale getirilmiştir. Label encoder yapılmıştır.

3.2 Modelin Eğitilmesi

Veri setinin %20’si test veri seti olarak kullanılmıştır. Veri seti Random Forest Regressor algoritması ile eğitilmiştir. r2 \_score 0.2553315159057056 olarak çıkmıştır..

4. DENEYSEL SONUÇLAR VE TARTIŞMA

Makine öğrenmesi algoritmaları ile ayakkabı fiyat tahmini projesi çıktıları şöyledir.

A screenshot of a computer

Description automatically generated

Görsel 2 – Parameter tuning öncesi Random Forest Regressor r2\_score ve MAE, MSE, RMSE değerleri

A screenshot of a computer

Description automatically generated

Görsel 3 – RandomizedSearchCV sonucunda çıkan r2\_score değeri

5. SONUÇ

Bu projede makine öğrenmesi algoritmalarından RandomForestRegressor ile ayakkabı fiyat tahmini yapılmıştır. Ön işlemeden geçirildikten sonra eğitilen veri setinde RandomizedSearchCV’ünde kullanılmasıyla başarım %26’ a yükselmiştir. Ancak bu başarım oranı ayakkabı fiyatı tahmini için yetersizdir. Bu yüzden tutarlı bir şekilde ayakkabı fiyatı tahmin edilememiştir.

**Yazarların Katkıları:**

Veri toplama kısmı ,veri ön işleme, modelin eğitilmesi, çıkan sonucun değerlendirilmesi ve tüm proje Mehmet Fırat KAHVECİOĞLU tarafından yapılmıştır.

# Kaynakça

|  |  |
| --- | --- |
| [1] | L. XinHai, "Using" random forest" for classification and regression," *Chinese Journal of Applied Entomology,* vol. 50, no. 4, pp. 1190-1197, 2013. |
| [2] | "Trendyol - Alışveriş Noktası," 29 Aralık 2023. [Online]. Available: https://www.trendyol.com/ |
| [3] | J. Brownlee, Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python., Machine Learning Mastery., 2020. |
| [4] | P. Ramya, V. Sindhura and P. V. Sagar, "Testing using selenium web driver," in *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Coimbatore, India, 2017. |
| [5] | B.-B. Jia and M.-L. Zhang, "Multi-Dimensional Classification via Decomposed Label Encoding," *IEEE Transactions on Knowledge and Data Engineering,* vol. 35, no. 2, pp. 1844-1856, 2023. |