****

**TC**

**MANİSA CELAL BAYAR ÜNİVERSİTESİ**

**HASAN FERDİ TURGUTLU TEKNOLOJİ FAKÜLTESİ**

**YAZILIM MÜHENDİSLİĞİ**

**Veri Madenciliği Dersi**

**(Final Ödevi)**

**HAZIRLAYANLAR: MESUT UNUTUR – 202803053**

**HAZIRLAYANLAR: MUSTAFA ARSLAN – 202803075**

**Makine Öğrenmesi Algoritmaları İle Ayakkabı Fiyat Tahmini**

**Predicting Price of Shoe With Machine Learning Algorithms**

**Özet:**

Bu projede makine öğrenmesi yöntemiyle ayakkabı fiyatlarının tahmin edilmesi amaçlanmıştır. Veri seti hazırlanmasında Trendyol sitesinden yararlanıldı. Trendyol sitesinden çekilen veri seti ham haliyle 4116 adettir. Ön işlemeden geçirilen veri seti üzerinde makine öğrenmesi yöntemleri uygulanmış ve istenen başarım oranı elde edilememiştir.

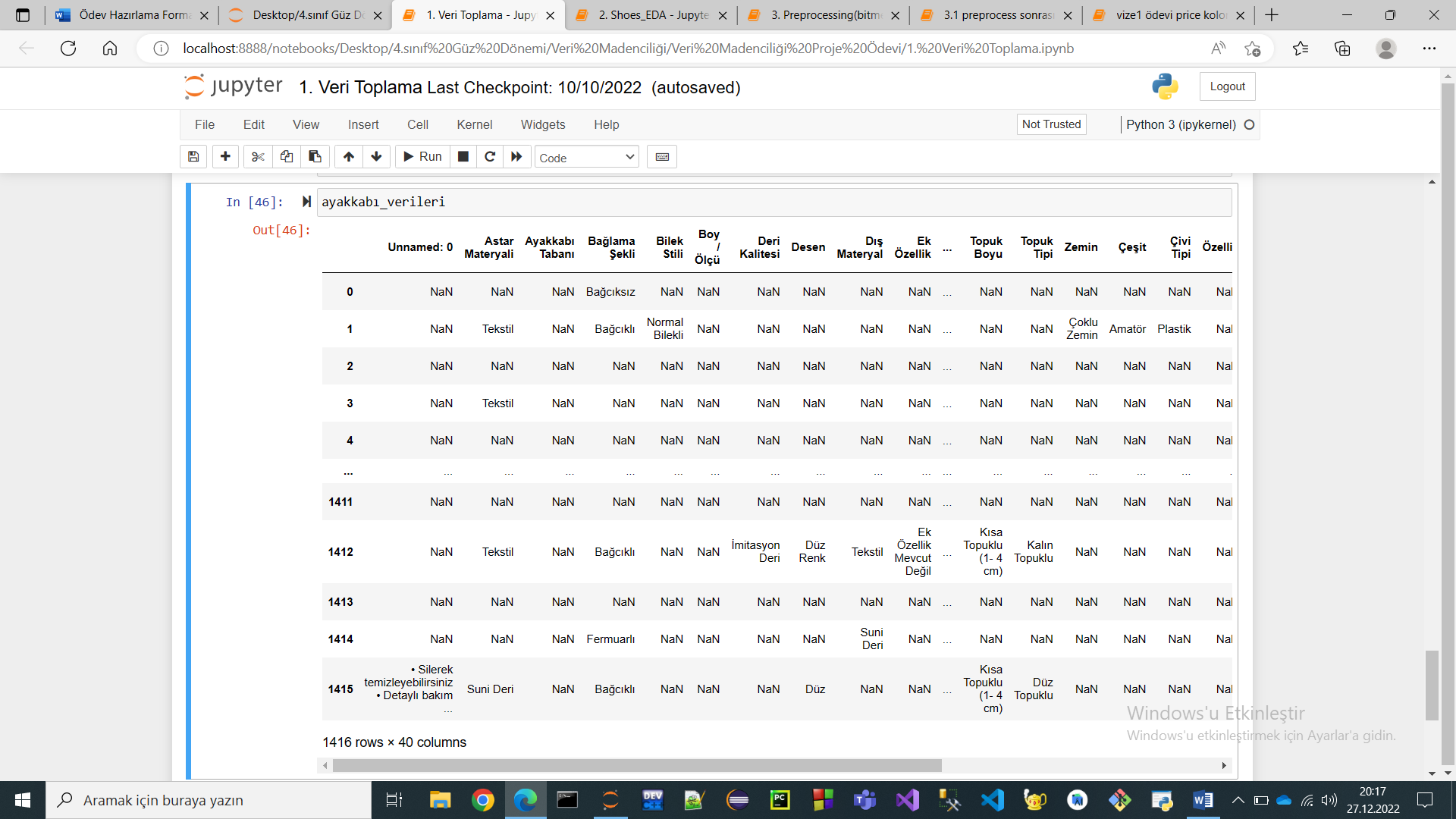
**Anahtar Kelimeler:** Alışveriş, Ayakkabı, Fiyat Tahmini, Makine Öğrenmesi, Trendyol

**1.Giriş**

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 40 tane feauture’dan yani 40 tane özellikten oluşmaktadır. Bu 40 özellikten bazıları renk, desen, astar materyali gibi özelliklerdir. Bu 40 özellik önce preprocessing işlemlerinden geçirilmiş ve ardından makine öğrenmesi algoritmaları ile ayakkabının fiyatı tahmin edilmeye çalışılmıştır.

**2.Verilerin Anlatımı**

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 requests kütüphaneleri yardımı ile Trendyol sitesinden çekilmiştir.



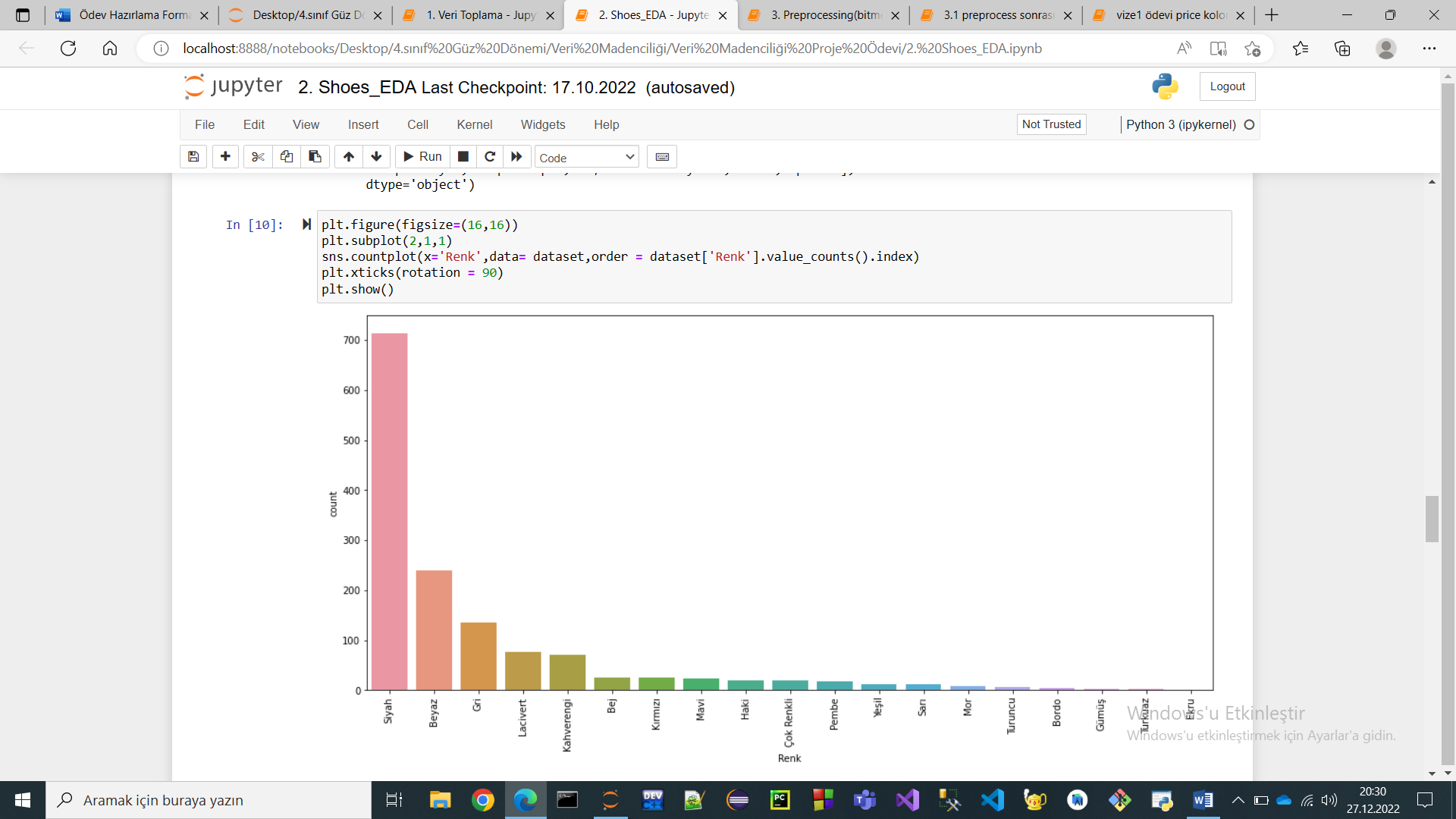
Görsel – 1 Trendyoldan Çekilen Veri Setinin DataFrame’e Dönüştürülmüş hali

**3. Projede Kullanılan Yöntemler**

**3.1 Preprocessing İşlemleri**

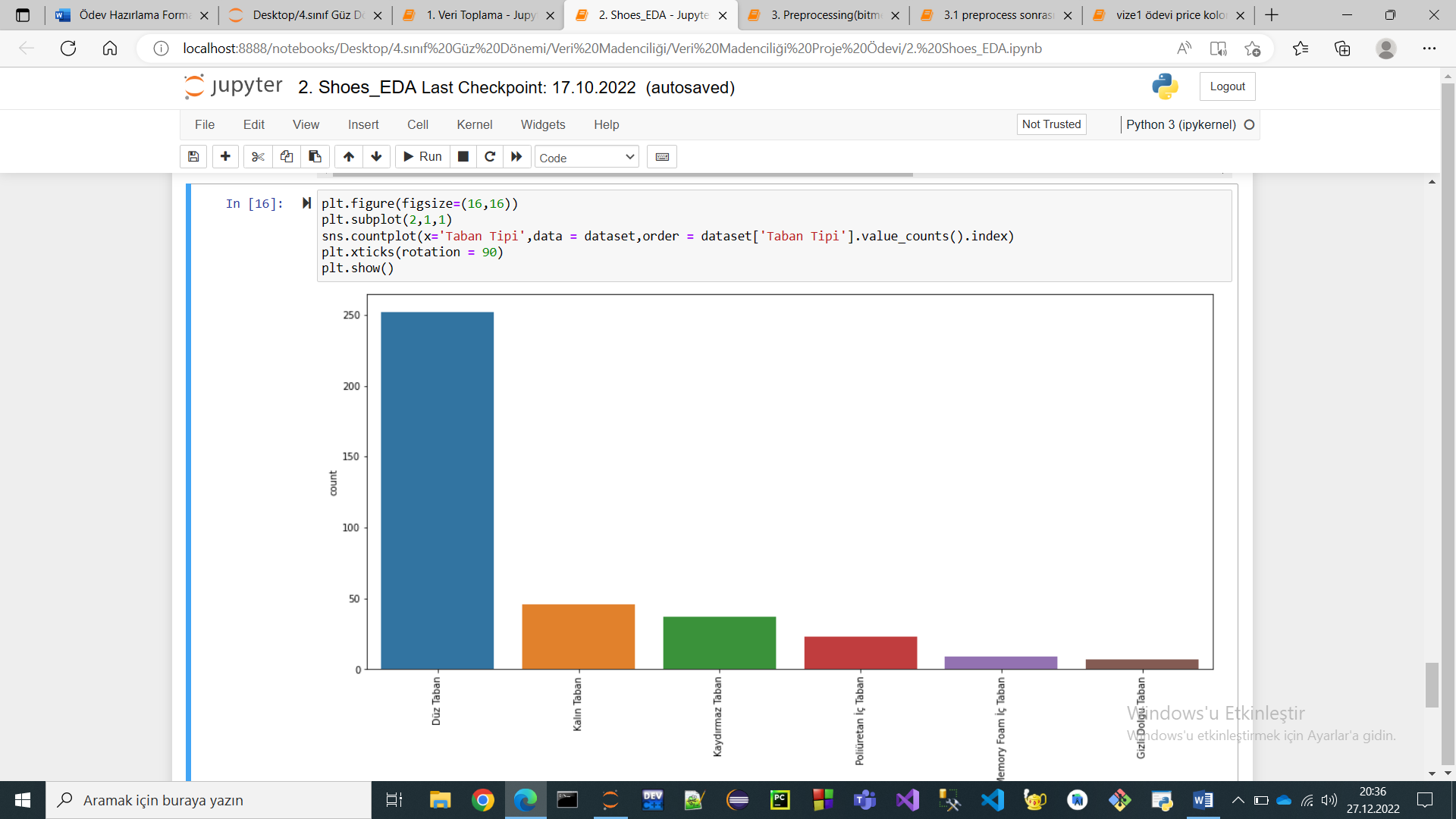
Veri setinin 1416 veriden ve 40 kolondan oluştuğu söylenmişti. Trendyol sitesinden ayakkabı özelliklerine ait veriler çekilirken bazı ayakkabıların featureları diğer ayakkabılarda rastlanmadı. Bu yüzden bütün ayakkabılar için belirlenen 40 feauture ayakkabıların hepsinde bulunmamaktadır. Bazı ayakkabıların 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 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.



Grafik 1 – Renklere göre ayakkabı sayıları

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



Grafik 2 – Taban tipine göre ayakkabı sayıları

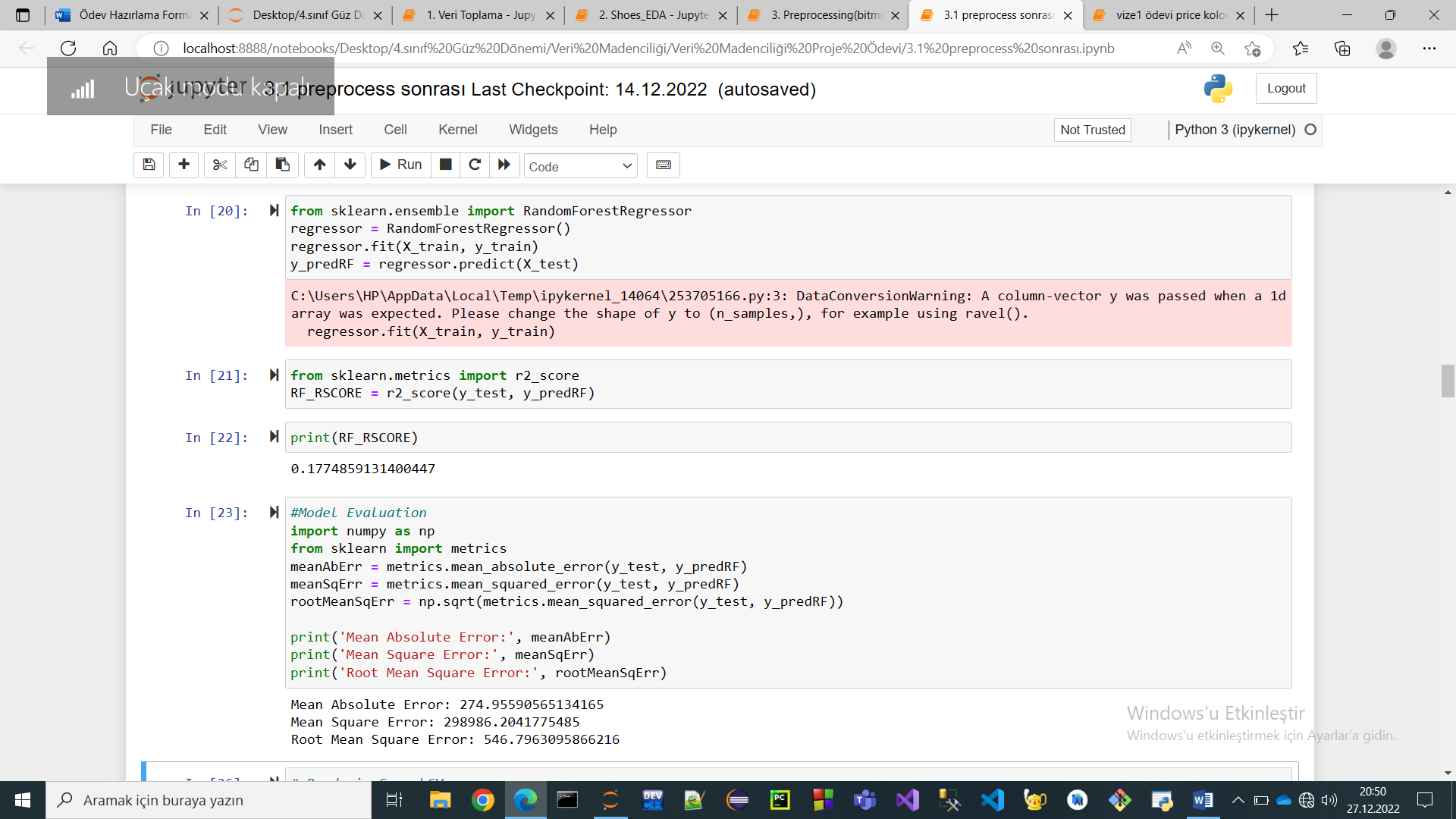
Projenin ön işleme aşamasında son olarak veri setindeki kolonlar her biri kendi içinde sayısal hale getirilmiştir. Label encoder yapılmıştır.

**3.2 Modelin Eğitilmesi**

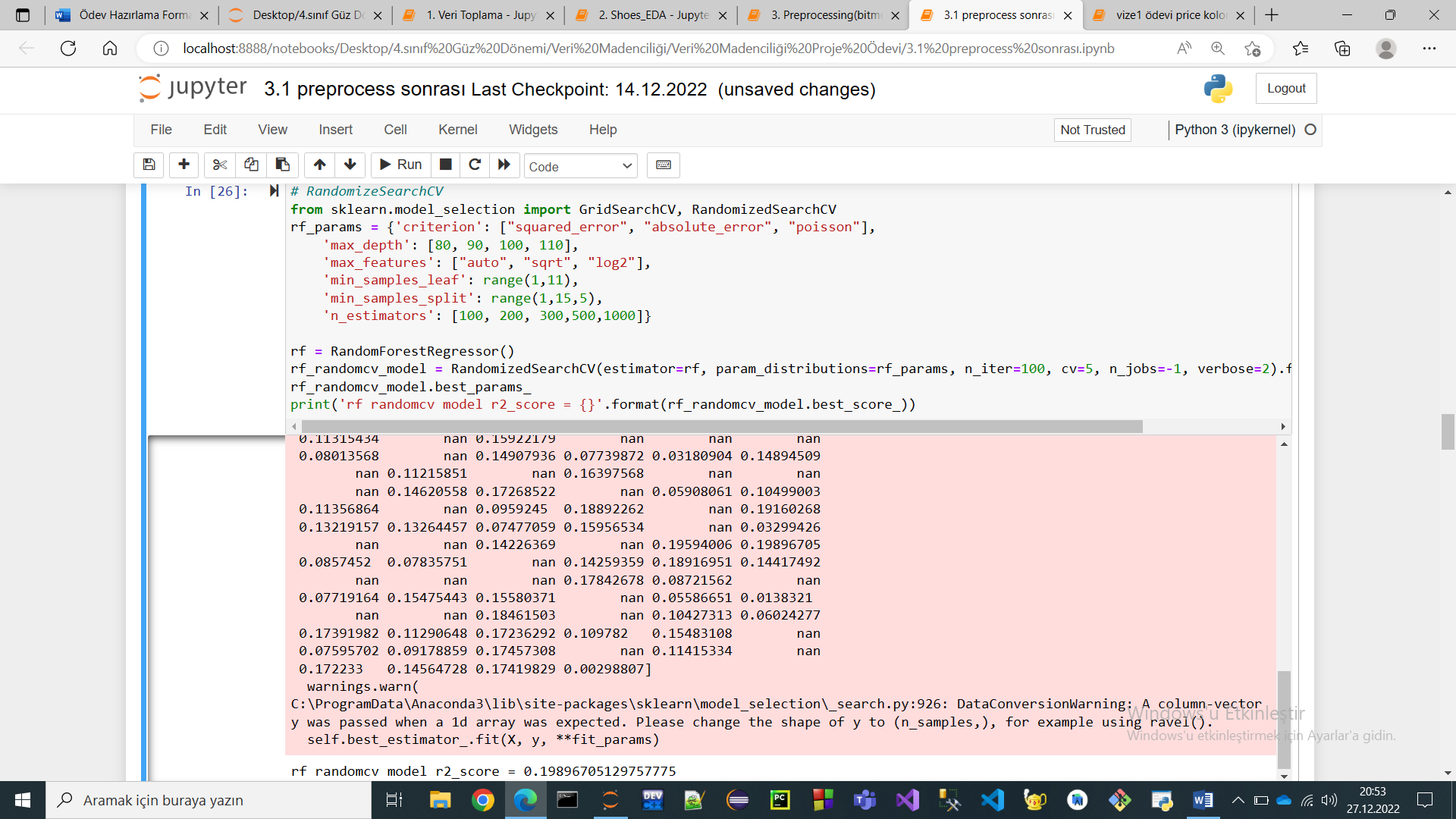
Model eğitime geçmeden önce veri setindeki değerler z-score ile bir daha düzenlenmiştir. Ve bunun ardından veri seti train\_test\_split metodu ile train ve test veri seti olarak ayrılmıştır. Veri setinin %20’si test veri seti olarak kullanılmıştır. Veri seti Random Forest Regressor algoritması ile eğitilmiştir. r2 \_score 0.17749 olarak çıkmıştır. Daha sonra başarımı artırmak için RandomizedSearchCV metodu denenmiş veri seti 500 kere fit edilmiş ve bu sefer başarım 0.19897 olarak çıkmıştır. Ve bu değer makine öğrenmesi yöntemleri ile ayakkabı tahmini yapmak için yetersizdir.

**4.Deneysel Sonuçlar ve Tartışma**

Makine öğrenmesi algoritmaları ile ayakkabı fiyat tahmini projesi çıktıları şöyledir.



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



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 %19.9’ 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 Mesut Unutur tarafından yapıldı. Veri ön işleme, modelin eğitilmesi, çıkan sonucun değerlendirilmesi ise Mesut UNUTUR ve Mustafa ARSLAN tarafından yapılmıştır.