İçindekiler

[1. PROJE RAPORU 2](#_Toc57397655)

[1.1 Problem: 2](#_Toc57397656)

[1.2 Verileri Elde Etmek İçin Kullanılan Crawler Yöntemi: 2](#_Toc57397657)

[1.3 Veri Setinin Özellikleri: 2](#_Toc57397658)

[1.4 Ön İşleme Adımında Yapılanlar: 2](#_Toc57397659)

# PROJE RAPORU

## Problem:

Kendi veri setimizi oluşturmak ve oluşturduğumuz bu veri setini makine öğrenmesi algoritmaları tarafından işlenmek üzere hazır hale getirmektir. Bu nedenle veri setini oluşturmak için web ortamından veri elde etmemiz gerekmektedir. Ben arabam.com adresinden 500 adet aracın renk, yıl, km ve fiyat bilgilerini alarak bu verileri kullanacağım.

## Verileri Elde Etmek İçin Kullanılan Crawler Yöntemi:

Ben python kütüphaneleri olan BeautifulSoup4, requests kütüphanelerini kullandım. İlk önce veri çekeceğim siteleri değişkenlere atadım. Ardından veriyi kaydedeceğim excel dosyasını oluşturdum ve veri çekme işlemine başladım. İlk önce sayfa kaynak kodunu değişkene atadım. Ardından gelen veriyi üç kez parçalara ayırdıktan sonra bir döngü içerisinde tekrar parçalayarak istediğim bilgilere ulaştım. Ardından bu bilgileri değişkenlere atadım ve atadığım değişen içerisinde düzenlemeler (noktalama işaretleri, boşluklar vb) yaptım. Ardından aynı döngü içerisinden verileri excel dosyasına aktardım ve excel dosyasını kaydedip kapattım.

## Veri Setinin Özellikleri:

Veri setimizde 500 adet aracın verisi bulunmakta. Aracın rengi, yılı, km si ve fiyatı olmak üzere 4 kolondan oluşuyor. Aracın rengi haricindeki veriler sayısal değerler, aracın rengi ise metin olarak bulunmaktadır.

## Ön İşleme Adımında Yapılanlar:

SimpleImputer kütüphanesini kullanarak “Araç Rengi” kısmındaki boş olan kısımları en çok tekrar eden veriyi yazarak doldurdum.

Ölçekleme yapılarak sayılar küçültüldü.

Train ve test yapıldı.

Ek olarak veri çekme sırasında “TL”, “.”, “,” gibi karakterler “.replace” komutuyla silindi.