1. **Uygulama Adımları**
2. **Gerekli Kütüphaneleri Yüklenmesi**
3. **Veri Setinin Oluşturulması veya Herhangi Bir Veri Setinin**

**Yüklenmesi**

1. **Model Oluşturma ve Eğitme**
2. **Tahminler Yapma**
3. **Sonuçları Görselleştirme**
4. **Sonuçların Yorumlanması**

**Modelin Eğitimi: Model eğitim süreçlerini, uygulanan yöntem ve adımları açıklayın.**

**Veri Seti:**

**Bilgilerinden oluşmaktadır**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Brand** | **Price** | **Body** | **Mileage** | **EngineV** | **Engine Type** | **Registration** | **Year** | **Model** |
| **BMW** | **4200** | **sedan** | **277** | **2** | **Petrol** | **yes** | **1991** | **320** |
| **Mercedes-Benz** | **7900** | **van** | **427** | **02.9** | **Diesel** | **yes** | **1999** | **Sprinter 212** |
| **Mercedes-Benz** | **13300** | **sedan** | **358** | **5** | **Gas** | **yes** | **2003** | **S 500** |
| **Audi** | **23000** | **crossover** | **240** | **04.2** | **Petrol** | **yes** | **2007** | **Q7** |
| **Toyota** | **18300** | **crossover** | **120** | **2** | **Petrol** | **yes** | **2011** | **Rav 4** |
| **Mercedes-Benz** | **199999** | **crossover** | **0** | **05.** | **Petrol** | **yes** | **2016** | **GLS 63** |
| **BMW** | **6100** | **sedan** | **438** | **2** | **Gas** | **yes** | **1997** | **320** |
| **Audi** | **14200** | **vagon** | **200** | **02.7** | **Diesel** | **yes** | **2006** | **A6** |
| **Renault** | **10799** | **vagon** | **193** | **01.5** | **Diesel** | **yes** | **2012** | **Megane** |
| **Volkswagen** | **1400** | **other** | **212** | **01.8** | **Gas** | **no** | **1999** | **Golf IV** |
| **Renault** | **11950** | **vagon** | **177** | **01.5** | **Diesel** | **yes** | **2011** | **Megane** |
| **Renault** | **2500** | **sedan** | **260** | **1.79** | **Petrol** | **yes** | **1994** | **19** |
| **Audi** | **9500** | **vagon** | **165** | **02.7** | **Gas** | **yes** | **2003** | **A6 Allroad** |
| **Volkswagen** | **10500** | **sedan** | **100** | **01.8** | **Petrol** | **yes** | **2008** | **Passat B6** |

**1. Veri Yükleme ve Temizleme:**

* **Pandas ve numpy gibi kütüphaneler, veri manipülasyonu ve sayısal işlemler için içe aktarılır.**
* **Kod, kullanıcının ortamına bağlı olarak araç verilerini içeren bir CSV dosyasını okumaya çalışır . C:\Users\saim\Desktop/TEST\KURS\1.04. Real-life example.csv).**
* **Yüklenen veri (data) daha sonra içeriğine genel bir bakış elde etmek için yazdırılır.**
* **Veri Temizleme Adımları:** 
  1. **Eksik değerleri kontrol etmek için data.isnull().sum() kullanılır.**
  2. **Eksik değerlere sahip satırlar data\_no\_null = data.dropna(axis=0) kullanılarak veri setinden çıkarılır.**
  3. **'Year', 'Mileage' ve 'EngineV' gibi sayısal özelliklerin dağılımı, saçılma grafiği kullanılarak analiz edilir.**
  4. **'Price' için aykırı değerler (99. yüzdelik dilimin üzerindeki değerler), .quantile(.99) ve koşullu filtreleme kullanılarak tanımlanır ve çıkarılır.**
  5. **(sns.displot) kullanılarak analiz edilir.**
  6. **'Year' sütunundaki potansiyel aykırı değerler (1. yüzdelik dilimin altındaki değerler) benzer mantık kullanılarak ele alınır.**
  7. **Satırlar düşürüldükten sonra, uygun indeksleme sağlamak için indeks sıfırlanır (data\_cleaned = data\_recent\_year.reset\_index(drop=True))**

**2. Özellik Mühendisliği:**

* **Fiyat dağılımındaki potansiyel doğrusallık sorunu çözmek için, orijinal "Price" sütununun doğal logaritması alınarak yeni bir özellik "log\_price" oluşturulur.**
* **"log\_price" özelliğinin dağılımı, benzer grafikleme teknikleri kullanılarak analiz edilir.**

**3. Veri Ön İşleme:**

* **"Price" sütunu artık gerekli olmadığı için (log fiyatını tahmin ediyoruz) düşürülür.**
* **Özellikler arasındaki yüksek korelasyon (çoklu doğrusallık) potansiyeli tanımlanır ve gereksiz bir özellik ('Year') çıkarılabilir.**
* **Kategorik özellikler, her kategorinin içinde ayrı ayrı ikili özellikler oluşturmak için pd.get\_dummies kullanılarak one-hot encoding ile kodlanır. Bu, modelin kategorik ilişkileri daha iyi anlamasına yardımcı olur.**

**4. Eğitim-Test Bölme:**

* **Ön işlenmiş veri (data\_with\_dummies), sklearn.model\_selection.train\_test\_split kullanılarak eğitim ve test kümelerine ayrılır. Bu, modelin eğitim verilerini ezberlemediğinden ve görünmeyen verilere (test kümesi) genelleme yapabilmesini sağlar.** 
  + **Eğitim verileri (x\_train, y\_train) modelin eğitilmesi için kullanılır.**
  + **Test verileri (x\_test, y\_test) modelin görünmeyen veriler üzerindeki performansını değerlendirmek için kullanılır.**

**5. Model Eğitimi ve Değerlendirme:**

* **Fiyatın logaritmasını tahmin etmek için LinearRegression (sklearn.linear\_model'dan) modeli seçilir.**
* **Model, eğitim verileri (x\_train, y\_train) üzerinde eğitilir.**
* **Modelin eğitim verileri üzerindeki performansı, reg.score(x\_train, y\_train) kullanılarak değerlendirilir. Bu, bağımlı değişkenin varyansının model tarafından açıklanan oranını gösteren R-kare değerini sağlar.**
* **Modelin keseni ve katsayıları, reg.intercept\_ ve reg.coef\_ kullanılarak alınır. Bu katsayılar, fiyatın logaritmasını tahmin etmede her bir özelliğin ağırlığını temsil eder.**

**.**

**. Model Özeti ve Test:**

* **Modeldeki özellik adlarını ve bunlara karşılık gelen ağırlıkları gösteren bir özet tablosu (reg\_summary) oluşturulur.**
* **Modelin test verileri üzerindeki performansı, reg.score(x\_test, y\_test) kullanılarak değerlendirilir.**
* **Test kümesi için tahminler (y\_hat\_test), eğitilmiş model kullanılarak elde edilir.**
* **Ayrı bir veri çerçevesi (df\_pf), modelin bireysel test örnekleri üzerindeki performansını analiz etmek için oluşturulur. Bu veri çerçevesi şunları içerir:** 
  + **Tahminler (np.exp(y\_hat\_test) kullanılarak log fiyatından geri dönüştürülür)**
  + **Hedef değerler (np.exp(y\_test) kullanılarak log fiyatından geri dönüştürülür)**

**Tahminler: Tahminleri gösterin ve yorumlayın.**

**Tahmin ve Kalanlar Analizi:**

* **Model, eğitim verileri için fiyatın logaritmasını tahmin etmek için kullanılır (y\_hat).**
* **Gerçek hedef değerler ile tahmin edilen değerler arasındaki ilişki, saçılma grafiği ile görselleştirilir.**

**Kalanlar (gerçek değer ile tahmin edilen değer arasındaki fark), seaborn (sns.displot) ile olasılık yoğunluk fonksiyonu (PDF) grafiği kullanılarak analiz edilir. İdeal olarak, kalanlar sıfır etrafında rastgele dağılmalıdır**