## Automobile İnformation DataSed

metin, ekran görüntüsü, yazılım, multimedya yazılımı içeren bir resim

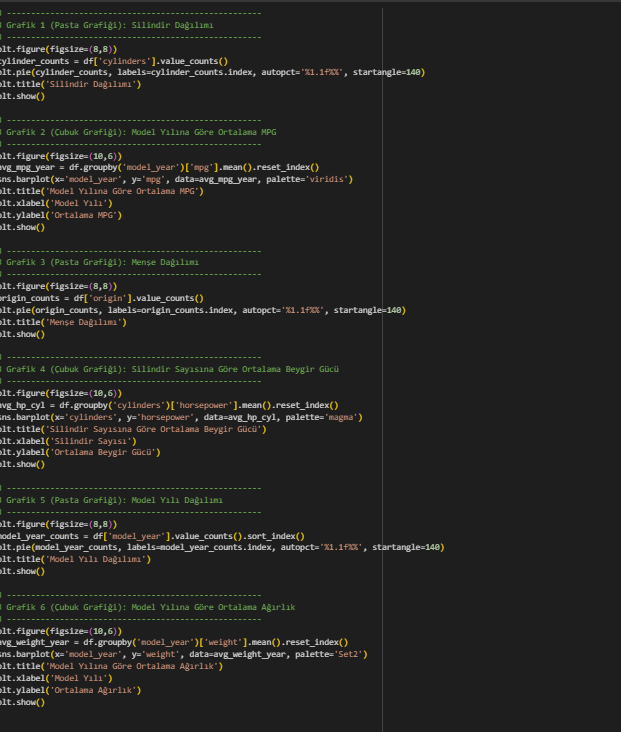
Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Analize başlamadan önce gerekli Python kütüphaneleri projeye dahil edildi. Ardından Kaggle'dan alınan "Automobile.csv" adlı veri seti pandas kütüphanesi yardımıyla yüklendi. İlk 10 satır yazdırılarak veri yapısına genel bir bakış yapıldı. Bu aşama, hangi sütunların bulunduğu, eksik ya da hatalı veri olup olmadığını görmek açısından önemlidir.

metin, ekran görüntüsü, yazılım, multimedya yazılımı içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Veri setinin sütun isimleri incelenmiş ve tüm değişkenler listelenmiştir. Bu adım, analiz sürecinde hangi değişkenlerin kullanılabileceğini belirlemek için yapılmıştır. Ardından, her sütun için minimum ve maksimum değerler hesaplanmıştır. Bu sayede veri setindeki değişkenlerin değer aralıkları görülmüş, uç noktalar belirlenmiş ve olası aykırı değerlerin varlığı ön değerlendirilmiştir.



diyagram, daire, ekran görüntüsü, metin içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, ekran görüntüsü, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, renklilik içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

diyagram, metin, ekran görüntüsü, daire içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, ekran görüntüsü, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, daire, diyagram, renklilik içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, ekran görüntüsü, renklilik, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Veri setindeki çeşitli kategorik ve sayısal değişkenler kullanılarak toplamda 6 adet grafik oluşturulmuştur.  
Bu grafiklerle birlikte:

* Silindir sayısı, menşe ve model yılı gibi kategorik değişkenlerin dağılımları **pasta grafikleri** ile,
* Model yılına göre ortalama MPG, silindir sayısına göre ortalama beygir gücü ve model yılına göre ortalama ağırlık gibi ilişkiler ise **çubuk grafikleri** ile görselleştirilmiştir.

Bu görseller, verinin genel yapısını daha iyi anlamaya ve değişkenler arası farkları sezgisel olarak değerlendirmeye olanak sağlamıştır.

ekran görüntüsü, metin, kare, yazılım içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Veri setindeki değişkenler arasındaki doğrusal ilişkileri incelemek için korelasyon analizi yapılmıştır. Bu analizde yalnızca sayısal veriler dikkate alınmıştır. Amaç, değişkenler arasında anlamlı pozitif veya negatif ilişkiler olup olmadığını belirlemektir. Korelasyon katsayıları, değişkenlerin birbirleriyle olan ilişkisinin gücünü ve yönünü gösterir. Bu sayede, örneğin bir değişkendeki artışın başka bir değişkende artışa mı yoksa azalmaya mı yol açtığı anlaşılabilir.

Elde edilen korelasyon değerleri bir matris şeklinde görselleştirilmiş ve ısı haritası (heatmap) yardımıyla sunulmuştur. Bu görsel, hangi değişkenler arasında güçlü ilişkiler olduğunu kolayca fark etmeyi sağlar. Renk tonları sayesinde pozitif ve negatif ilişkiler hızlıca ayırt edilebilir. Bu analiz, özellikle özellik seçimi (feature selection) aşamasında, çok güçlü ilişkili değişkenlerin tespit edilip modellerde gereksiz tekrarların önüne geçilmesi açısından önemlidir.

metin, ekran görüntüsü, yazılım, multimedya yazılımı içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Bu analizde, araçların yakıt verimliliği (mpg) beygir gücü, ağırlık ve ivmelenme gibi teknik özelliklerle tahmin edilmiştir. Model oluşturulmadan önce verideki eksik ve sonsuz değerler temizlenmiştir.

Model sonucunda:

* **R² = 0.706:** Model, mpg değerinin %70’ini açıklayabiliyor.
* **Ağırlık ve beygir gücü:** mpg üzerinde anlamlı ve negatif etkiye sahip.
* **İvmelenme:** mpg üzerinde anlamlı bir etkisi yok.
* **Model genel olarak anlamlıdır**, fakat düşük Durbin-Watson değeri otokorelasyon ihtimalini gösterir.
* **Condition number** çok yüksek, bu da çoklu doğrusal bağlantı (multicollinearity) riski olabileceğini düşündürür.

metin, ekran görüntüsü, yazılım, multimedya yazılımı içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**scikit-learn ile Doğrusal Regresyon**

Bu modelde, LinearRegression sınıfı kullanılarak mpg (yakıt verimliliği) tahmin edilmiştir. Model, belirlenen bağımsız değişkenlerle eğitilmiş ve aynı veriler üzerinde test edilmiştir.

Sonuç olarak:

* **Katsayılar:** Her bir bağımsız değişkenin mpg üzerindeki etkisini gösterir.
* **R² skoru:** Modelin veriye ne kadar iyi uyduğunu gösterir. 1’e ne kadar yakınsa, model o kadar başarılıdır.
* **MSE (Ortalama Kare Hatası):** Tahminlerin ne kadar sapma gösterdiğini ifade eder. Düşük MSE, daha iyi tahmin performansı anlamına gelir.

Bu analiz, modelin genel performansını değerlendirmek ve değişkenlerin etkisini görmek için kullanılmıştır.

metin, ekran görüntüsü, yazılım, multimedya yazılımı içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**Gerçek Değerler vs Tahminler Grafiği**

Bu grafik, modelin tahmin ettiği mpg değerleri ile gerçek mpg değerlerini karşılaştırmak için çizilmiştir. Her bir nokta bir gözlemi temsil eder.

* Noktalar **doğruya ne kadar yakınsa**, model o kadar başarılı tahmin yapmıştır.
* Eğer tüm noktalar **45 derecelik bir çizgi boyunca sıralansaydı**, model mükemmel tahmin yapmış olurdu.

Bu grafik, modelin başarısını görsel olarak değerlendirmek için kullanılır.

**Genel Değerlendirme**

Bu çalışma kapsamında, otomobil veri seti kullanılarak veriye genel bir bakış yapılmış, görselleştirme teknikleriyle değişkenlerin yapısı analiz edilmiş ve ardından istatistiksel yöntemlerle regresyon modelleri oluşturulmuştur. Yapılan analizler sayesinde, yakıt verimliliğini etkileyen faktörler belirlenmiş ve bu faktörlerin etkileri hem istatistiksel hem de görsel olarak değerlendirilmiştir.

Sonuç olarak, oluşturulan modeller, belirli değişkenlerin mpg üzerindeki etkilerini anlamada başarılı sonuçlar vermiştir. Bu tür veri analizleri, yalnızca model kurmakla kalmayıp, aynı zamanda veriyi anlamak, yorumlamak ve karar destek sistemleri oluşturmak açısından da kritik öneme sahiptir. Elde edilen bulgular, ileride yapılacak daha gelişmiş modelleme ve optimizasyon çalışmaları için temel teşkil etmektedir.

Metehan ÜNAL – 426979

Eren AVCI – 426939

Yiğit TANER – 439919

Yakup BİROL – 426931