

Car Price Prediction - Raporu

Giriş: Bu rapor, Kaggle'da bulunan "Car_Price_Prediction" veri seti üzerinde gerçekleştirilen bir veri madenciliği ve makine öğrenimi projesini detaylı bir şekilde ele almaktadır. Projemizin temel hedefi, otomobil özellikleri kullanılarak bir otomobilin fiyatını tahmin etmektir. Bu amaç doğrultusunda, veri madenciliği sürecimizi aşağıdaki adımlarla sürdürdüm:

1. Veri Keşfi: Veri madenciliği sürecimize başlamadan önce, veri setimizi anlamaya çalıştık. Veri setimiz 6.019 dan fazla otomobil kaydını içeriyor ve 13 farklı özelliği kapsıyor. Bu özellikler arasında marka, model, yıl, kilometre, yakıt türü, vites türü, motor hacmi, güç, tork ve daha birçok detay bulunuyor. Temel istatistikler, özelliklerin dağılımları ve eksik verilerin kontrolü gibi analizleri gerçekleştirdik.

2. Veri Ön İşleme: Veri setinin temizlenmesi ve hazırlanması aşamasında eksik verileri doldurduk, aykırı değerleri ele aldık ve kategorik değişkenleri sayısallaştırdık. Özellik mühendisliği yoluyla yeni özellikler türettik ve veriyi eğitim ve test veri setleri olarak böldük.

3. Model Oluşturma: Makine öğrenimi modelleri oluşturmak için çeşitli algoritmaları değerlendirdik. Başlangıçta, Linear Regression modeli ile başladık, ardından Random Forest, gradient boosting gibi daha karmaşık modelleri denedik. Hiperparametre ayarlaması ve çapraz doğrulama ile modellerimizi geliştirmeye çalıştık.

4. Model Değerlendirmesi: Oluşturduğumuz modelleri eğitim veri setine uyguladık ve test veri seti üzerinde performanslarını değerlendirdik. R-kare metriğiyle modellerimizin performansını kritik bir şekilde değerlendirdik. En iyi modelin, test veri seti üzerinde Random Forest ve 0.8835 R-kare değeri ile başarılı bir şekilde çalıştığını gözlemledik.

5. Sonuçlar ve Öneriler: Bu analiz sonuçları, otomobil fiyatlarını tahmin etmek için makine öğrenimi modellerinin kullanılabileceğini göstermektedir. Ancak, daha fazla veri toplama, daha fazla özellik mühendisliği ve daha fazla model geliştirme çalışması yapılabilir. Ayrıca, modelin gerçek dünya uygulamalarında nasıl performans göstereceğini değerlendirmek için daha fazla test yapılması gerekmektedir. Ayrıca, modelin güncel verilerle sürekli olarak güncellenmesi ve bakımı da önemlidir.

6. Son Düşünceler: Bu proje, veri madenciliği ve makine öğrenimi tekniklerinin otomobil fiyat tahminleri gibi karmaşık görevlerde nasıl kullanılabileceğini göstermektedir. Daha fazla veri ve daha iyi özellik mühendisliği ile modelin performansının daha da artırılabilmesine inanıyorum. Projenin gelecekteki çalışmalarımıza ışık tutmasını umuyorum.