

Kırmızı Şarap Kalite Tahminleme Sistemi

Proje Açıklaması:

*Bu projede, kırmızı şarapların kalitelerinin tahmin edilmesi amacıyla bir makine öğrenmesi uygulaması gerçekleştirilmiştir. Kullanılan veri seti kırmızı şarapların kalitesine ilişkin çeşitli özellikleri içermektedir. Projenin amacı, bu çeşitli özelliklerden faydalanarak kırmızı şarapların kalitesini tahmin edebilmektir. Proje kapsamında aşağıdaki adımlar sırasıyla gerçekleştirilmiştir:

1. Veri Toplama ve Veri Ön İşleme (Preprocessing):

Veri Seti: wine.csv

(Kaggle platformundan projede kullanılmak üzere bir veri seti seçilmiştir)

*Veri Ön İşleme aşamasında veri seti gözden geçirilip sayısal olmayan verilerin varlığı **info()** fonksiyonu yardımıyla kontrol edilip herhangi bir sayısal olmayan veriye rastlanmamıştır. Bu yüzden herhangi bir veri dönüşümü gerçekleştirilmemiştir. Veri setinde eksik değerlerin varlığı kontrol edilmiştir ve herhangi bir eksik veri bulunamamıştır.

Veri Seti, 1599 farklı şarap örneğinden oluşmakta ve “sabit asitlik”, “uçucu asitlik”, “sitrik asitlik”, “artık şeker”, “klorürler”, “serbest kükürt dioksit”, “toplam kükürt dioksit”, “yoğunluk”, “pH”, “sülfatlar”, “alkol”, “kalite” olmak üzere 12 farklı özellik içermektedir

2. Veri Görselleştirme:

*Korelasyon matrisi ile veri setindeki değişkenler arasındaki ilişkilerin gücü ortaya koyulmuştur. Örneğin, “Hangi değişkenler arasında güçlü ilişki var?”, “İlişkinin yönü nedir? (pozitif mi negatif mi)”. Hangi özelliklerin hedef değişkenle daha güçlü ilişkisi olduğu öğrenilmiştir.

3. Veriyi Eğitim ve Test Olarak Ayırma:

*Bu projede veri seti, %70 training set ve %30 test set olmak üzere `train_test_split` fonksiyonu ile ayrılmıştır. Buradaki amaç modelin aşırı öğrenmesini (overfitting) engellemektir.

4. Model Seçimi:

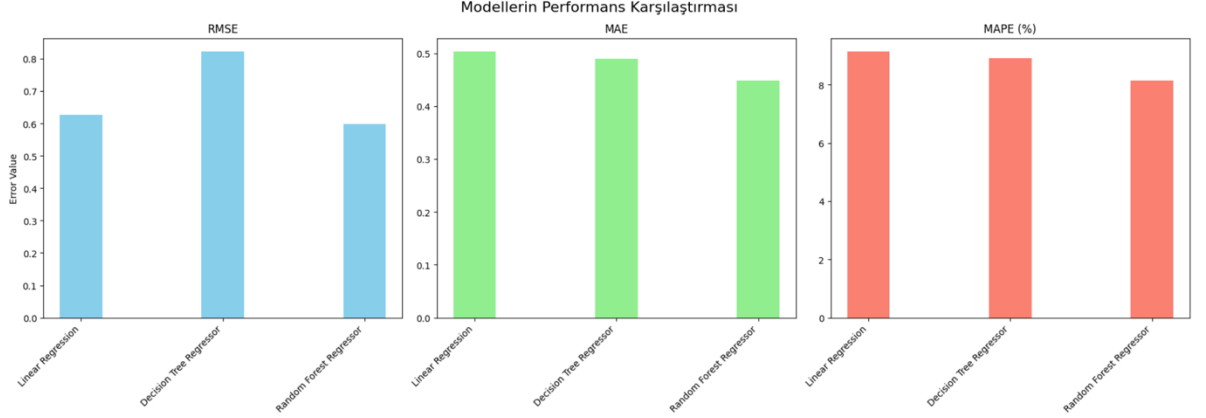
*Projede 3 farklı model kullanılmıştır: Linear Regression, Decision Tree Regressor ve Random Forest Regressor. Her bir model eğitim seti üzerinde ayrı ayrı eğitilmiştir.

5. Performans Değerlendirilmesi:

*Her bir model MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), RMSE (Root Mean Squared Error) performans metrikleri kullanılarak değerlendirme yapılmıştır.

*Random Forest Regressor modeli, düşük MAE ve MAPE (sapma değerleriyle) değerleriyle en dengeli ve başarılı model olmuştur.

6. Model Karşılaştırılması:



7. Yöntem Seçimi:

*Performans metrikleri değerlendirilmesine göre en düşük hatalı model **Random Forest Regressor** olarak belirlenmiştir. Bu model MAE ve MAPE performans metrikleri açısından diğer iki modele göre daha iyi sonuçlar vermiştir. Özellikle MAPE (sapma değerlerinin) değerlerinin daha düşük olması sebebiyle tahminlerin daha güvenilir olduğunu göstermektedir.

8. Sonuç ve Geliştirme Önerileri:

*Bu proje, kırmızı şarapların kalitelerini tahmin etmek amacıyla üç farklı regresyon modelinin karşılaştırılmalı analizini içermektedir. Veri setindeki kalite puanlarının dengesiz dağılımı (genellikle 5 ve 6 arasında yoğunlaşmıştır), modelin nadir kalite puanlarını (3,7,8) doğru tahmin etmesini zorlaştırmaktadır. Ayrıca veri setinin küçüklüğü, modelin genelleme yeteneğini sınırlamaktadır. Elde edilen model, mevcut haliyle bile kırmızı şarapların kalitelerinin yaklaşık tahminini sunabilir ama gerçek dünya uygulamaları için sürekli iyileştirilmesi gerekebilir.