

YAPAY ZEKA VE MAKİNE ÖĞRENMESİ EĞİTİMİ PROJE ÖDEVİ

Yunus Korkmaz

Projenin Amacı

- Bu projenin amacı, bir şarap kalite veri setini kullanarak çeşitli regresyon modelleri ile şarap kalitesini tahmin etmek ve bu modellerin performansını karşılaştırmaktır. Özellikle **KNN (K-Nearest Neighbors)**, **Destek Vektör Regresyonu (SVR)**, **Karar Ağacı (Decision Tree)**, **Rastgele Orman (Random Forest)** gibi modeller kullanılarak, **doğruluk ve hata metrikleri** değerlendirilmiştir. Ayrıca, değişken önem düzeyi analiz edilerek, hangi özelliklerin şarap kalitesini en çok etkilediği belirlenmiştir.

Veri Seti Hakkında Bilgi

- Veri seti, şarapların kimyasal özellikleri ve kalitesini içerir. Bu veri setinde, her satır bir şarap örneğini temsil eder ve şu sütunlardan oluşur:
- Kimyasal özellikler (pH, alkol, asitlik vb.)
- **Kalite (quality):** Hedef değişkendir ve şarapların kalitesini temsil eder (0-10 arasında puanlanmıştır).
- Eksik veri kontrolü yapılmıştır ve veri setinde eksik değer bulunmadığı tespit edilmiştir. Bu tür kontroller, modelin doğruluğunu sağlamak için önemlidir.

Kullanılan Modeller ve Yöntemler

1. KNN (K-Nearest Neighbors)

- KNN algoritması, komşularına bakarak tahmin yapar.
- Komşu sayısını optimize etmek için GridSearchCV yöntemi kullanılmıştır.
- En iyi komşu sayısı belirlendikten sonra model yeniden eğitilmiştir.

2. Destek Vektör Regresyonu (SVR)

SVR modeli, veriler arasındaki doğrusal olmayan ilişkileri tahmin etmek için kullanılmıştır.

C parametresi optimize edilerek model performansı artırılmıştır.

3. Karar Ağacı (Decision Tree)

Karar ağacı, özelliklerin bölünmesiyle tahmin yapar.

max_depth ve min_samples_split parametreleri optimize edilmiştir.

4. Rastgele Orman (Random Forest)

- Birden fazla karar ağacından oluşan topluluk modelidir.
- GridSearchCV kullanılarak hiperparametreler optimize edilmiştir.

Deney Sonuçları

- KNN modelinde komşu sayısının artırılması RMSE ve MAE değerlerini düşürmüştür.
- SVR modeli C parametresi optimize edildikten sonra daha iyi sonuçlar vermiştir.
- Rastgele Orman modeli, Karar Ağacı'ndan daha iyi performans göstermiştir. Özellikle, max_depth ve n_estimators gibi parametrelerin optimize edilmesi tahmin doğruluğunu artırmıştır.

Genel Yorum

- Modeller arasında Rastgele Orman, en düşük RMSE ve MAE değerlerini vererek şarap kalitesini tahmin etmede en başarılı model olmuştur.
- KNN modeli, basit ve yorumlanabilir olmasına rağmen daha fazla veriyle daha iyi sonuç verebilir.
- SVR ve Karar Ağacı modelleri iyi performans sergilemiştir, ancak Rastgele Orman kadar etkili değildir.