

# WINE QUALITY VERİ SETİ KEŞİFÇİ VERİ ANALİZİ VE MAKİNE ÖĞRENMESİ UYGULAMASI

Ders: Veri analizi

Öğrenci No: 21703919

Ad Soyad: Shirin Dzhorobaeva

Tarih: 01/02/2026

## 1.Giriş ve Problem Tanımı

Bu çalışmada Wine Quality veri seti kullanılarak şarap kalitesinin tahmin edilmesi amaçlanmıştır. Şarap kalitesi, birçok fiziksel ve kimyasal özelliğe bağlı olarak değişmektedir.

Bu proje kapsamında öncelikle keşifçi veri analizi (EDA) yapılmış, değişkenler arasındaki ilişkiler incelenmiştir. Ardından lineer regresyon ile iki sayısal değişken arasındaki ilişki analiz edilmiştir. Son olarak, farklı sınıflandırma algoritmaları kullanılarak şarap kalitesinin tahmini gerçekleştirılmıştır.

## 2.Veri Seti

Bu çalışmada kullanılan Wine Quality veri seti, UCI Machine Learning Repository üzerinden temin edilmiştir. Veri seti, kırmızı şaraplara ait çeşitli kimyasal özelliklerini içermektedir.

Veri setinde toplam 1599 gözlem ve 12 değişken bulunmaktadır. Bağımlı değişken şarap kalitesidir (quality). Bu değişken, analizlerde ikili sınıfa dönüştürülmüş ve kalite değeri 6 ve üzeri olan şaraplar "iyi", diğerleri ise "kötü" olarak etiketlenmiştir.

## 3. Yöntem

Çalışmanın ilk aşamasında keşifçi veri analizi (EDA) gerçekleştirilmiştir. Eksik veri kontrolü yapılmış ve veri setinde eksik gözleme rastlanmamıştır. Aykırı değerler boxplot grafikleri ile incelenmiştir.

İkinci aşamada, alkol oranı ile yoğunluk (density) arasındaki ilişki lineer regresyon modeli kullanılarak analiz edilmiştir. Modelin başarımı  $R^2$  metriği ile değerlendirilmiştir.

Son aşamada sınıflandırma problemi ele alınmış ve Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machines (SVM) ve Decision Tree algoritmaları

uygulanmıştır. Modeller Accuracy, Precision, Recall, F1-Score ve Confusion Matrix metrikleri ile değerlendirilmiştir.

#### **4. Bulgular**

Keşifçi veri analizi sonuçlarına göre bazı değişkenler arasında anlamlı korelasyonlar gözlemlenmiştir. Özellikle alkol oranı ile kalite arasında pozitif bir ilişki olduğu görülmüştür.

Lineer regresyon analizi sonucunda elde edilen  $R^2$  değeri, alkol oranının yoğunluğu belirli bir oranda açıklayabildiğini göstermektedir.

Sınıflandırma modelleri karşılaştırıldığında Logistic Regression ve SVM algoritmalarının diğer modellere göre daha dengeli performans sergilediği gözlemlenmiştir.

#### **5. Sonuç**

Bu çalışma kapsamında Wine Quality veri seti üzerinde farklı makine öğrenmesi teknikleri uygulanmıştır. Elde edilen sonuçlar, şarap kalitesinin fizikal ve kimyasal özellikler kullanılarak tahmin edilebileceğini göstermektedir.

Gelecek çalışmalarda veri seti genişletilebilir ve model performansı hiperparametre optimizasyonu ile artırılabilir.

<https://github.com/KEKTAR/wine-quality-ml-project>

