



Aydın ŞAHİN
19.01.2025

Makine Öğrenimi

2025



Hazırlayan:
Aydın ŞAHİN

İçindekiler



03	Projenin Amacı
04	Materyaller
05	Veri Seti Hakkında Bilgi
08	Deneyde Kullanılan Modeller
09	Deney Sonuçları
10	Tartışma
11	Referanslar

Projenin Amacı



Misyonumuz:

Portekiz "Vinho Verde" şarabının kırmızı çeşitleriyle ilgili 11 parametreden oluşan bir veri seti kullanarak bu 11 metriği analiz edip şarabın kalitesini tahmin eden bir model oluşturmak.



Materiyaller



Kullanılan Modeller:

- Random Forest Classifier (Rastgele Orman)
- Grid Search (Hiperparametre Optimizasyonu)



Veri Seti Hakkında Bilgiler



Veri seti Portekiz "Vinho Verde" şarabının kırmızı ve beyaz çeşitleriyle ilgilidir.

Girdi değişkenleri (fizikokimyasal testlere dayalı):

- 1 - sabit asitlik
- 2 - uçucu asitlik
- 3 - sitrik asit
- 4 - artık şeker
- 5 - klorürler
- 6 - serbest kükürt dioksit
- 7 - toplam kükürt dioksit
- 8 - yoğunluk
- 9 - pH
- 10 - sülfatlar
- 11 - alkol

Çıktı değişkeni (duyusal verilere dayalı):

- 12 - kalite (0 ile 10 arasında puan)

Toplam satır sayısı: 1599

Toplam sütun sayısı: 12

Satırlardaki toplam eksik veri sayısı: 0

Deney Sonuçları



Modelin İlk Doğruluk Oranı:

```
## ekrana yazma  
print("Modelin doğruluk oranı: " , dogruluk_orani) ## Modelin doğruluk oranı: 0.659375  
print("Sınıflandırma Raporu: \n", sinif_raporu)
```



GridSearch Kullandıktan sonra doğruluk oranı

```
## Sonuçları ekrana yazalım  
print("En iyi parametreler: ", gridArama.best_params_)  
print("En yüksek doğruluk: ", gridArama.best_score_) ## En yüksek doğruluk: 0.6849601715686274
```



Sınıf Eşitsizliklerini düzeltince Doğruluk oranı

```
## ekrana yazma  
print("Modelin doğruluk oranı: " , dogruluk_orani) ## Modelin doğruluk oranı: 0.659375  
print("Sınıflandırma Raporu: \n", sinif_raporu)
```

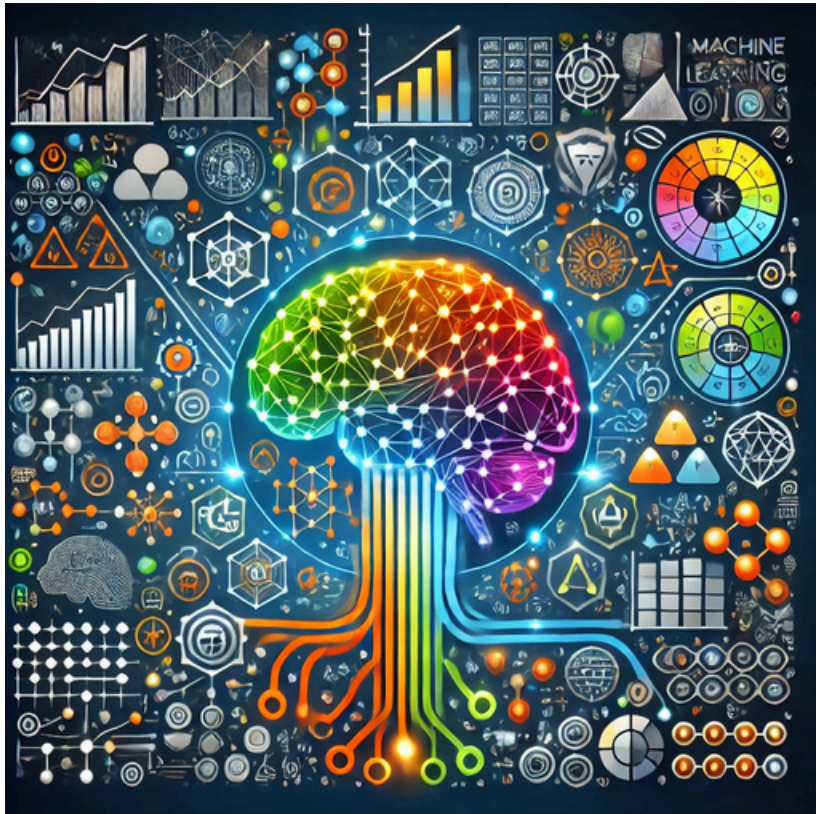

Tartışma



Deneyde önce RandomForestClassifier methodunu kullanarak eğitim ve test yaptık ve 0.659375(yaklaşık %65) doğruluk oranı elde ettik

Sonrasında GridSearch yöntemi kullanarak optimizasyon yaptık ve doğruluk oranı 0.684960(yaklaşık %68) oldu. Buradan yaptığımız çıkarım, GridSearch yöntemini kullanmak modelimizin doğruluk oranını artırdı

Son olarak Sınıf Eşitsizliklerini ortadan kaldırıp modelimizi tekrar eğitip tahmin yaptırдық. Bu işlemin sonucunda modelin doğruluk oranı 0.691204(yaklaşık %69) oldu. Yani bu işlem de modelin doğruluk oranını artırdı.



Referanslar



Gelir Analizi

1	İsmail AÇAR (Şanlıurfa Teknokent Makine Öğrenimi Eğitimi)
2	ChatGPT
3	Kaggle Red Wine Quality veri seti