

# KIRMIZI ŞARAP KALİTESİ SİNİFLANDIRMASI İÇİN MAKİNE ÖĞRENMESİ ALGORİTMALARININ KARŞILAŞTIRMALI ANALİZİ

ALPEREN ÖNAL



# Çalışmanın Amacı

.....

- Makine öğrenmesi kullanarak kırmızı şarap kalitesini tahmin etmek.
- Farklı algoritmaların doğruluklarını karşılaştırmak.
- Veri seti için en iyi tahmin modelini belirlemek.

# VERİ SETİ HAKKINDA

## Veri Seti Kaynağı

- Adı: Wine Quality Dataset
- Kaynak: UCI Machine Learning Repository
- Bağlantı: [Wine Quality Dataset - UCI](#)

## Veri Setinin Tanımı

### İçerdiği Şarap Türleri:

- Kırmızı şarap (red wine)
- Beyaz şarap (white wine)

### Hedef Değişken:

- quality (Şarap kalitesi, 3-8 arasında bir değer alır.)

### Bağımsız Değişkenler:

- Şarabın kimyasal ve fiziksel özelliklerini içeren 11 özellik bulunmaktadır.

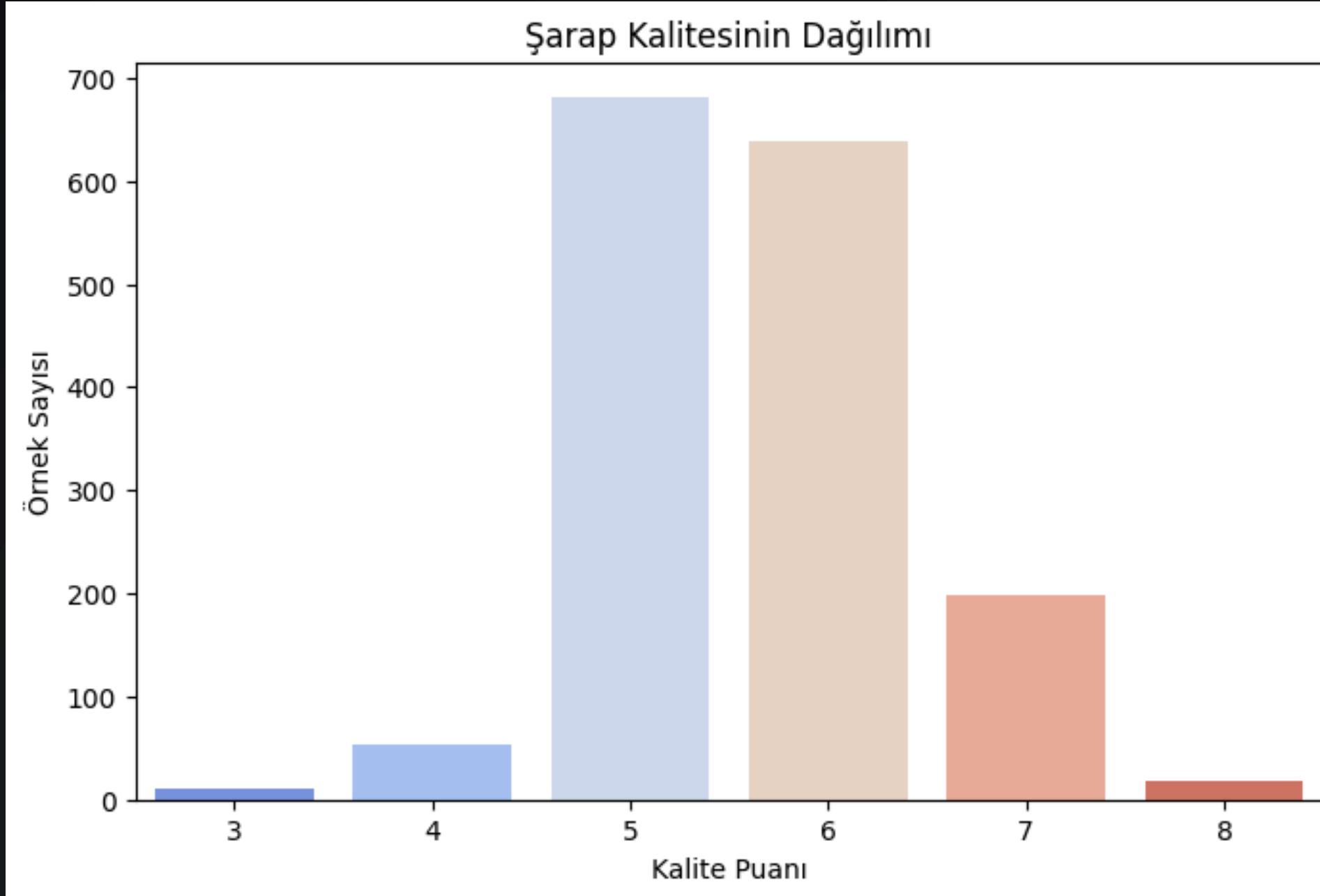
Özellik	Açıklama
fixed acidity	Sabit asitlik (tartarik asit gibi, gram/litre)
volatile acidity	Uçucu asitlik (asetik asit gibi, gram/litre)
citric acid	Sitrik asit oranı (gram/litre)
residual sugar	Artık şeker (gram/litre)
chlorides	Tuz içeriği (sodyum klorür, gram/litre)
free sulfur dioxide	Serbest sülfür dioksit (mg/litre)
total sulfur dioxide	Toplam sülfür dioksit (mg/litre)
density	Yoğunluk (g/cm <sup>3</sup> )
pH	Asidik/Bazik seviyesini gösteren pH değeri
sulphates	Sülfatlar (gram/litre)
alcohol	Alkol oranı (% hacim)
quality (Hedef)	Şarabın kalite skoru (0 ile 10 arasında)

Şarabın içerisinde bulunan kimyasallar ve açıklamaları.

# Veri Setindeki Zorluklar

• • • •

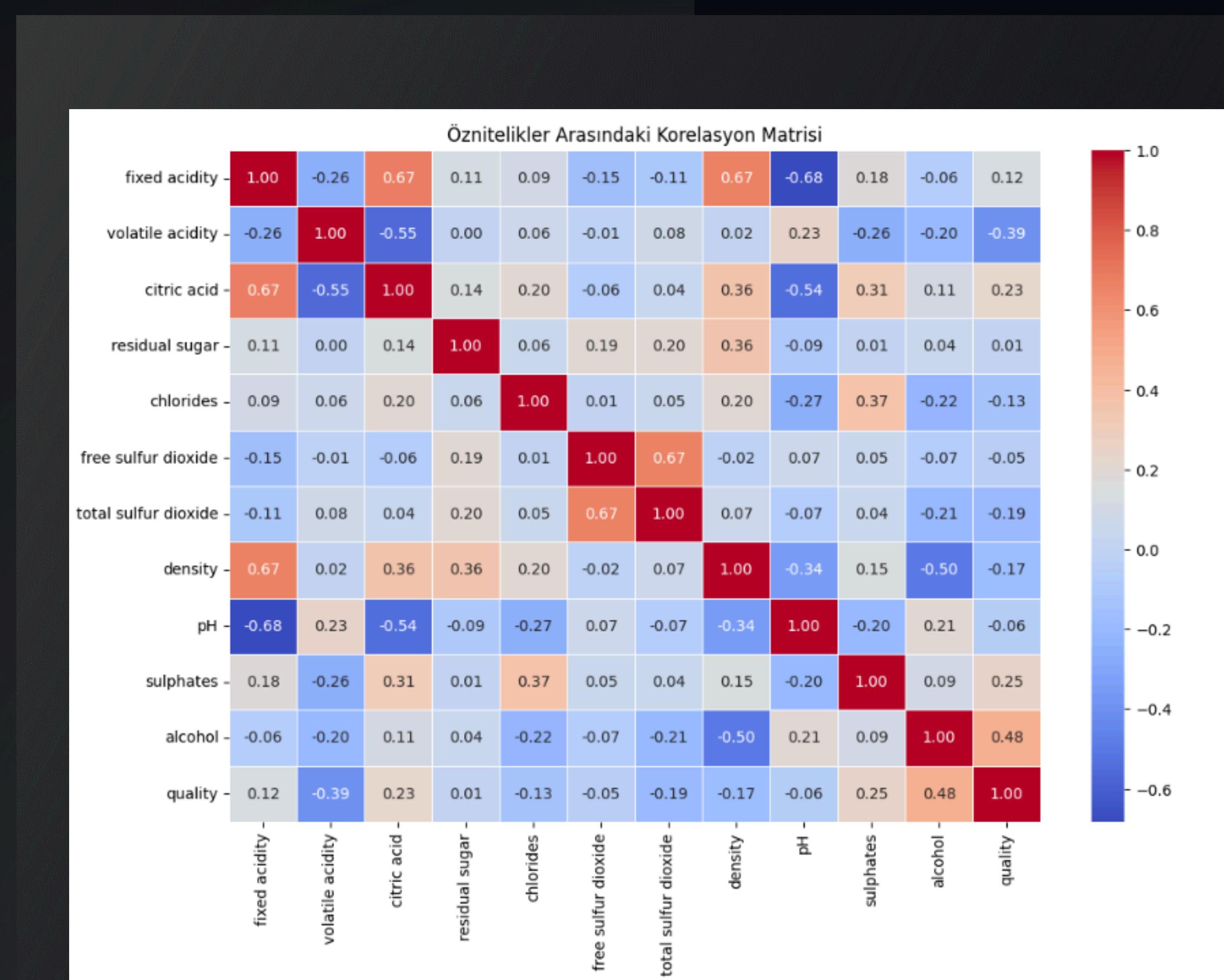
- **Sınıf Dengesizliği:** Çoğu şarap 5-6 kalite aralığında yoğunlaşmıştır, bu da modelin belirli kalite seviyelerinde aşırı öğrenmesine sebep olabilir.
- **Kimyasal Değişkenlerin Etkisi:** Bazı değişkenlerin şarap kalitesi üzerinde düşük korelasyonu mevcuttur, bu yüzden özellik seçimi (feature selection) önemlidir.



# VERİ SETİNDE KORELASYON ANALİZİ

## Sonuçlar

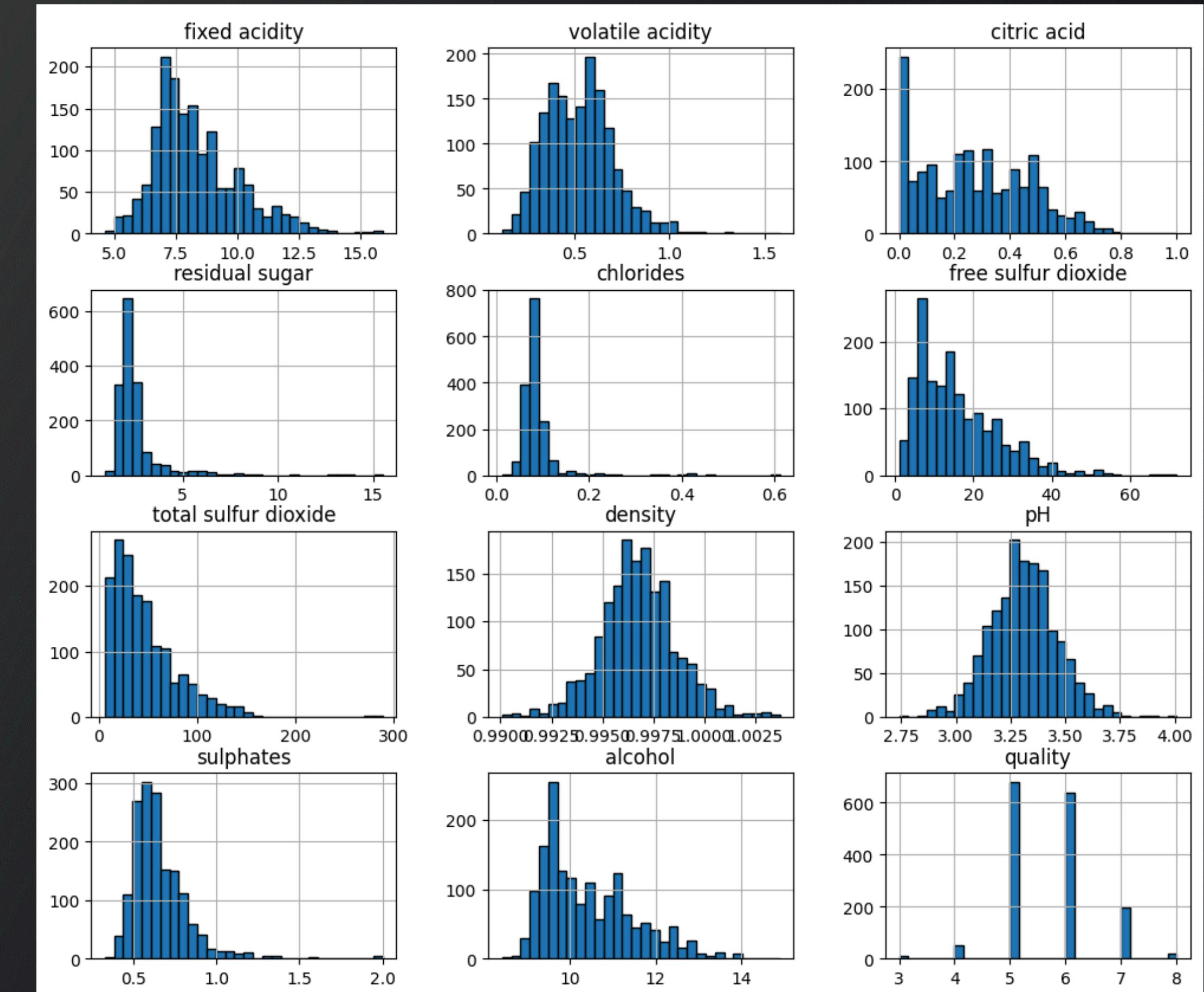
- Alkol ve sülfat seviyesi şarap kalitesini artıran en önemli faktörlerdir.
- Yüksek uçucu asitlik kalitenin düşmesine neden olabilir.
- Bazı değişkenler (örneğin artık şeker) kalite üzerinde düşük etkiye sahip olabilir ve modelden çıkarılabilir.



# ÖZNİTELİK DAĞILIMI

## Çıkarım

Ciddi aykırı değer bulunduran değişkenlerin (residual sugar, total sulfur dioxide, chlorides) aykırı değerlerinin temizlenmesi veya normalize edilmesi model performansını artıracaktır.



Veri setindeki özniteliklerin dağılım grafikleri.

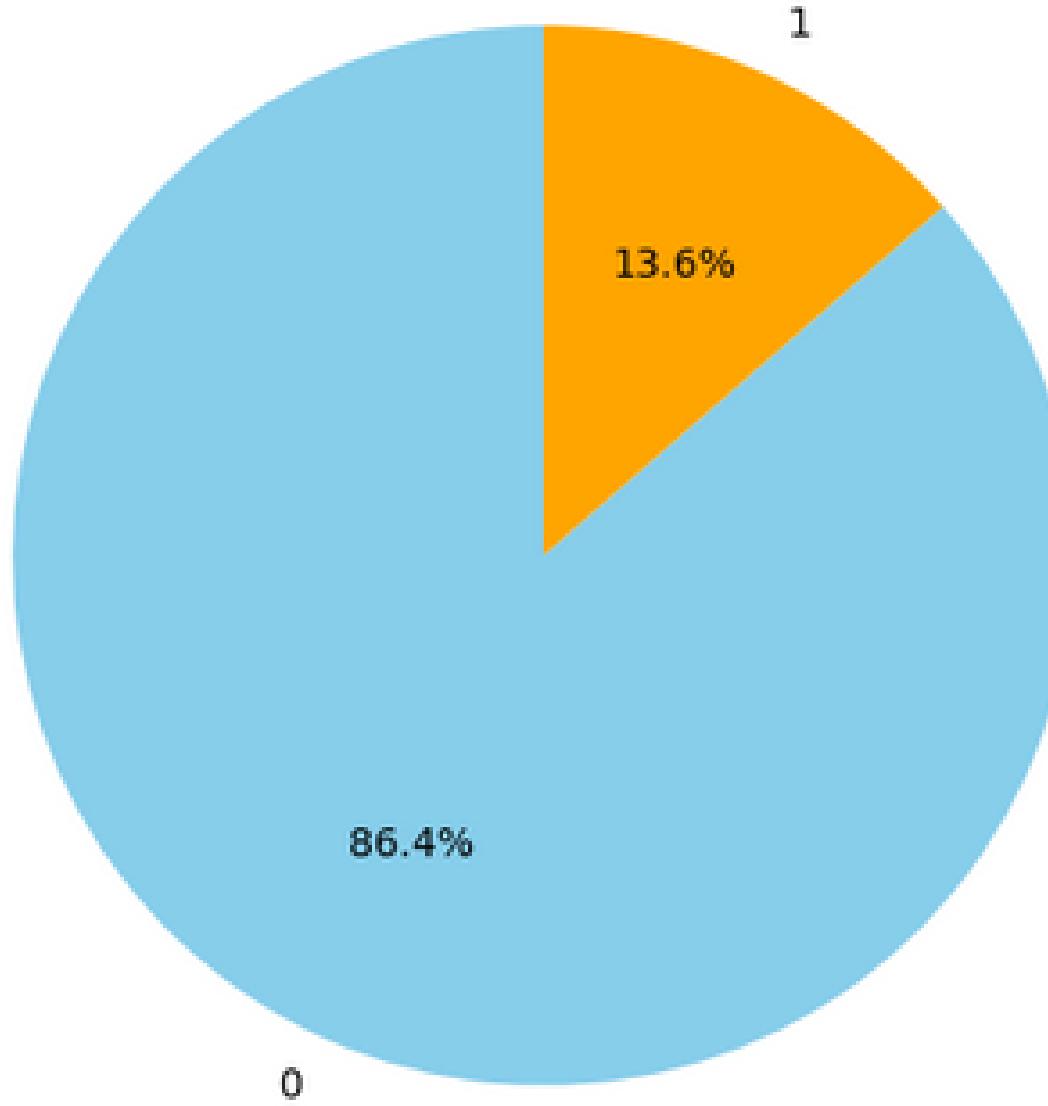
# SINIFLANDIRMANIN İKİLİ HALE GETİRİLMESİ

Veri setinde bulunan 3 ile 8 arasındaki kalite puanlarını daha net analiz edebilmek için ikili (binary) sınıflandırma yapıldı. 7 ve 8 puan alan şarapları "premium (1)", diğerleri "normal (0)" olarak belirlendi.

## Faydalar

- Sınıf dengesini sağlamak,
- Modelin daha iyi genelleme yapmasını sağlamak,
- Şarabın iyi olup olmadığını belirlemek için basit ve etkili bir yaklaşım sunmak.

Premium vs Non-Premium Şarap Yüzdelik Dağılımı



Sınıflandırmanın ikili hale getirilmesinin ardından sınıfların veri setindeki yüzdesel dağılımları.



# Kullanılan Makine Öğrenmesi Modelleri

.....

- Lojistik Regresyon (Basit, doğrusal modeller için iyi)
- Karar Ağacı (Kurallar üzerinden karar verme, overfitting riski var)
- Rastgele Orman (Birden çok karar ağacı ile daha sağlam modelleme)
- Destek Vektör Makineleri (SVM) (Verileri daha iyi ayıran sınıflandırıcı)
- XGBoost (En yüksek doğruluk oranına sahip boosting algoritması)

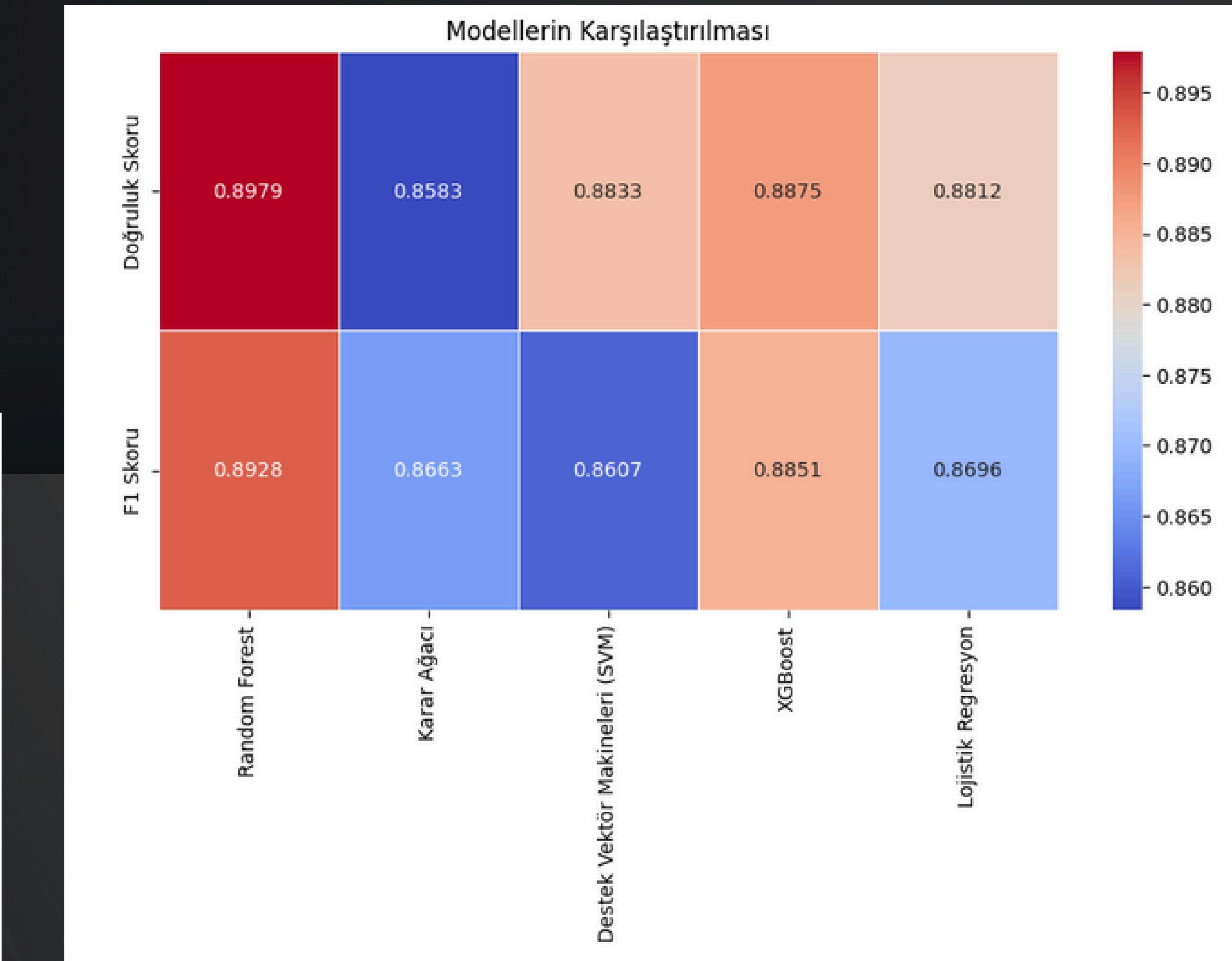
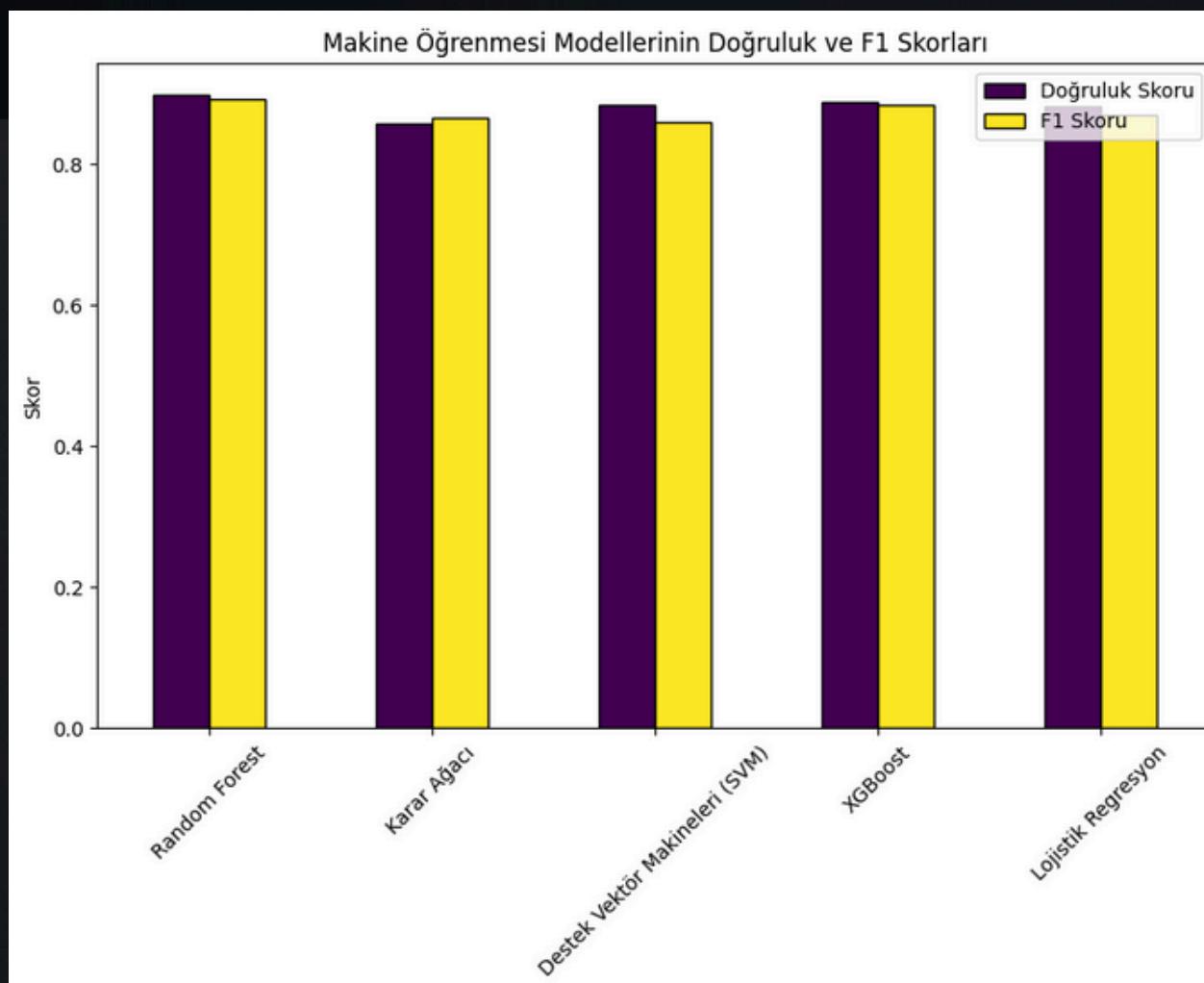
# MODEL PERFORMANS KARŞILAŞTIRMASI

Tüm modellerin başarı oranlarını karşılaştırmak için aşağıdaki tablolar oluşturulmuştur.

## Sonuçlar

**En İyi Model:** Random Forest

**En Kötü Model:** SVM (Destek Vektör Makineleri)



Eğitilen modellerinin performans analizi (doğruluk ve F1 skoru).

Eğitilen modellerinin performans analizi.

Özellik	Değer
Fixed Acidity	5.4
Volatile Acidity	0.25
Citric Acid	0.29
Chlorides	0.054
Free Sulfur Dioxide	19.0
Total Sulfur Dioxide	49.0
Density	0.99666
pH	3.40
Sulphates	0.76
Alcohol	11.0

X şarap markasının kimyasal özellikleri.

# Modelin Yeni Bir Örnek Üzerinde Test Edilmesi

.....

## MODELLERİN SINIF TAHMİNİ SONUÇLARI:

◆ Lojistik Regresyon Tahmini: 0

◆ XGBoost Tahmini: 1

◆ Destek Vektör Makineleri Tahmini: 0

◆ Karar Ağacı Tahmini: 1

◆ Random Forest Tahmini: 1

X markası için modellerin tahmin sonuçları.

# Modelin Yeni Bir Örnek Üzerinde Test Edilmesi

.....

Özellik	Değer
Fixed Acidity	6.4
Volatile Acidity	0.55
Citric Acid	0.69
Chlorides	0.054
Free Sulfur Dioxide	19.0
Total Sulfur Dioxide	49.0
Density	0.99666
pH	1.00
Sulphates	0.76
Alcohol	14.0

Y şarap markasının kimyasal özellikleri.

## MODELLERİN SINIF TAHMİNİ SONUÇLARI:

◆ Lojistik Regresyon Tahmini: 1

◆ XGBoost Tahmini: 1

◆ Destek Vektör Makineleri Tahmini: 0

◆ Karar Ağacı Tahmini: 1

◆ Random Forest Tahmini: 1

Y markası için modellerin tahmin sonuçları.

# KAYNAKÇA

- Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009).  
Wine Quality [Dataset]. UCI Machine Learning Repository.  
<https://doi.org/10.24432/C56S3T>.

---

# TEŞEKKÜRLER

---