logo, metin, simge, sembol, kırpıntı çizim içeren bir resim

Açıklama otomatik olarak oluşturuldu

T.C.

ONDOKUZ MAYIS ÜNİVERSİTESİ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ BİLGİSAYAR MÜHENDİSLİĞİ

ÖRÜNTÜ TANIMA-BM605

Şarap Kalitesi Veri Seti ile Analiz Raporu

FIRAT KAAN BİTMEZ - 23281855

SAMSUN, 2024-2025 Eğitim Öğretim Yılı Güz Yarıyılı

**1) Veri Kümesi Ne Zaman ve Hangi Makale/Çalışma Kapsamında Oluşturulmuştur?**

Veri kümesi 6 Ekim 2009 tarihinde Paulo Cortez ve arkadaşları tarafından oluşturulmuştur. Bu veri kümesi, Decision Support Systems dergisinde yayımlanan "Modeling wine preferences by data mining from physicochemical properties" adlı makale kapsamında hazırlanmıştır. Çalışma, Portekiz’in Vinho Verde şaraplarına yönelik fiziksel ve kimyasal testler kullanılarak şarap kalitesini modellemeyi amaçlamaktadır.

**2) Veri Kümesi Kaç Sınıftan Oluşmaktadır?**

Veri setlerinde quality adlı hedef değişken, şarap kalitesini ölçen 0 ile 10 arasında bir sayıdır. Ancak pratikte veri setlerinde genellikle 3 ile 9 arasında kalite değerleri bulunur. Bu sınıfların dağılımını daha sonra görselleştirme ile detaylandıracağız.

Name: count, dtype: int64

|  |  |  |
| --- | --- | --- |
| Quality | Red Wine Sınıf Dağılımı | White Wine Sınıf Dağılımı |
| 3 | 10 | 20 |
| 4 | 53 | 163 |
| 5 | 681 | 1457 |
| 6 | 638 | 2198 |
| 7 | 199 | 880 |
| 8 | 18 | 175 |
| 9 |  | 5 |

**3) Veri Kümesi Dağılımı Nasıl? (Dengeli/Dengesiz)**

Veri setinin dengeli olup olmadığını kontrol etmek için sınıf dağılımlarını inceleyeceğiz. Genellikle bu tür veri setlerinde düşük kaliteli şaraplar daha fazla, yüksek kaliteli şaraplar ise daha az bulunur.

**4) Veri Kümesi İçerisinde Kaç Tane Örnek Yer Almaktadır?**

* Kırmızı şarap (Red Wine) veri seti: 1599 örnek 12 Features
* Beyaz şarap (White Wine) veri seti: 4898 örnek 12 Features

**5) Veri Kümesinde Boş/Eksik/Yanlış Değer Var mı?**

Veri setinde eksik veya boş değer bulunmamaktadır. Bunu daha önce kontrol ettik.

|  |  |  |
| --- | --- | --- |
| Quality | Red Wine Eksik Veri Kontrolü | White Wine Eksik Veri Kontrolü |
| fixed acidity | 0 | 0 |
| volatile acidity | 0 | 0 |
| citric acid | 0 | 0 |
| residual sugar | 0 | 0 |
| chlorides | 0 | 0 |
| free sulfur dioxide | 0 | 0 |
| total sulfur dioxide | 0 | 0 |
| density | 0 | 0 |
| pH | 0 | 0 |
| sulphates | 0 | 0 |
| alcohol | 0 | 0 |
| quality | 0 | 0 |

**6) Veri Kümesinde Kaç Adet Öznitelik Vardır? Bu Öznitelikler Hangi Yapıdadır?**

Her iki veri setinde de 12 adet fiziksel ve kimyasal testlere dayalı öznitelik bulunmaktadır. Bu öznitelikler float64 veri türündedir ve hepsi sayısal verilerdir.

**7) Temel İstatistiksel Özellikler**

Her bir öznitelik için ortalama, standart sapma, minimum ve maksimum değerleri bulup analiz edeceğiz.

# Gerekli kütüphanelerin yüklenmesi

import pandas as pd

# Red ve White wine veri setlerinin yolları

red\_wine\_path = r'/mnt/data/winequality-red.csv'

white\_wine\_path = r'/mnt/data/winequality-white.csv'

# Veri setlerinin okunması

red\_wine = pd.read\_csv(red\_wine\_path, sep=';')

white\_wine = pd.read\_csv(white\_wine\_path, sep=';')

# Red Wine için istatistiksel özet

print("\nRed Wine Dataset - Statistical Summary:")

print(red\_wine.describe())

# White Wine için istatistiksel özet

print("\nWhite Wine Dataset - Statistical Summary:")

print(white\_wine.describe())

**8) Box Plot Grafikleri ile Özniteliklerin Yorumlanması**

Box plot ile her bir özniteliğin dağılımını ve uç değerlerini gözlemleyeceğiz.

import seaborn as sns

import matplotlib.pyplot as plt

# Red wine için box plot

plt.figure(figsize=(10, 6))

sns.boxplot(data=red\_wine)

plt.title('Red Wine Özniteliklerinin Box Plot Grafiği')

plt.xticks(rotation=90)

plt.show()

# White wine için box plot

plt.figure(figsize=(10, 6))

sns.boxplot(data=white\_wine)

plt.title('White Wine Özniteliklerinin Box Plot Grafiği')

plt.xticks(rotation=90)

plt.show()

**9) Özniteliklerin Histogram Grafikleri**

Her bir özniteliğin dağılımını görmek için histogram grafikleri çizeceğiz.

# Red wine histogram

red\_wine.hist(bins=15, figsize=(10, 8))

plt.suptitle('Red Wine Özniteliklerinin Histogramı')

plt.show()

# White wine histogram

white\_wine.hist(bins=15, figsize=(10, 8))

plt.suptitle('White Wine Özniteliklerinin Histogramı')

plt.show()

**10) Scatter Plot ile Veri Kümesinin Görselleştirilmesi**

Scatter plot ile farklı özniteliklerin birbirleriyle olan ilişkilerini inceleyeceğiz.

# Red Wine Scatter Plot

sns.pairplot(red\_wine, diag\_kind="kde")

plt.suptitle('Red Wine Scatter Plot Grafikleri')

plt.show()

# White Wine Scatter Plot

sns.pairplot(white\_wine, diag\_kind="kde")

plt.suptitle('White Wine Scatter Plot Grafikleri')

plt.show()

**11) Scatter Plot ile Özniteliklerin Sınıf Ayrımını Gösterme**

Scatter plot ile farklı öznitelikler arasındaki sınıf ayrımını inceleyeceğiz. Örneğin, alcohol ile quality arasındaki ilişkiyi gözlemleyeceğiz.

sns.scatterplot(x='alcohol', y='quality', data=red\_wine)

plt.title('Red Wine - Alcohol ve Quality İlişkisi')

plt.show()

sns.scatterplot(x='alcohol', y='quality', data=white\_wine)

plt.title('White Wine - Alcohol ve Quality İlişkisi')

plt.show()

**12) Violin Plot ile Sınıf Yoğunluklarının Görselleştirilmesi**

Violin plot ile her bir sınıfın yoğunluğunu göstereceğiz.

# Red wine violin plot

plt.figure(figsize=(10, 6))

sns.violinplot(x='quality', y='alcohol', data=red\_wine)

plt.title('Red Wine - Quality ve Alcohol Violin Plot')

plt.show()

# White wine violin plot

plt.figure(figsize=(10, 6))

sns.violinplot(x='quality', y='alcohol', data=white\_wine)

plt.title('White Wine - Quality ve Alcohol Violin Plot')

plt.show()

**13) Sınıflandırma Algoritmaları ile Deneyler**

Veri kümesinde sınıflandırma algoritmaları kullanarak sonuçları değerlendireceğiz. KNN, SVM ve Naive Bayes algoritmaları ile sınıflandırma deneyleri yapacağız ve başarı oranlarını inceleyeceğiz.

from sklearn.model\_selection import train\_test\_split

from sklearn.neighbors import KNeighborsClassifier

from sklearn.svm import SVC

from sklearn.naive\_bayes import GaussianNB

from sklearn.metrics import accuracy\_score

# Red Wine için KNN

X = red\_wine.drop('quality', axis=1)

y = red\_wine['quality']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.3, random\_state=42)

knn = KNeighborsClassifier()

knn.fit(X\_train, y\_train)

y\_pred\_knn = knn.predict(X\_test)

print(f"KNN Accuracy for Red Wine: {accuracy\_score(y\_test, y\_pred\_knn)}")

# SVM

svm = SVC()

svm.fit(X\_train, y\_train)

y\_pred\_svm = svm.predict(X\_test)

print(f"SVM Accuracy for Red Wine: {accuracy\_score(y\_test, y\_pred\_svm)}")

# Naive Bayes

nb = GaussianNB()

nb.fit(X\_train, y\_train)

y\_pred\_nb = nb.predict(X\_test)

print(f"Naive Bayes Accuracy for Red Wine: {accuracy\_score(y\_test, y\_pred\_nb)}")