Assignment 4

Name – Kajal

Reg no – 21BAC10039

Campus – Vit Bhopal

Code:

1. Load the Dataset

import pandas as pd

dataset = pd.read\_csv('winequality-red.csv')

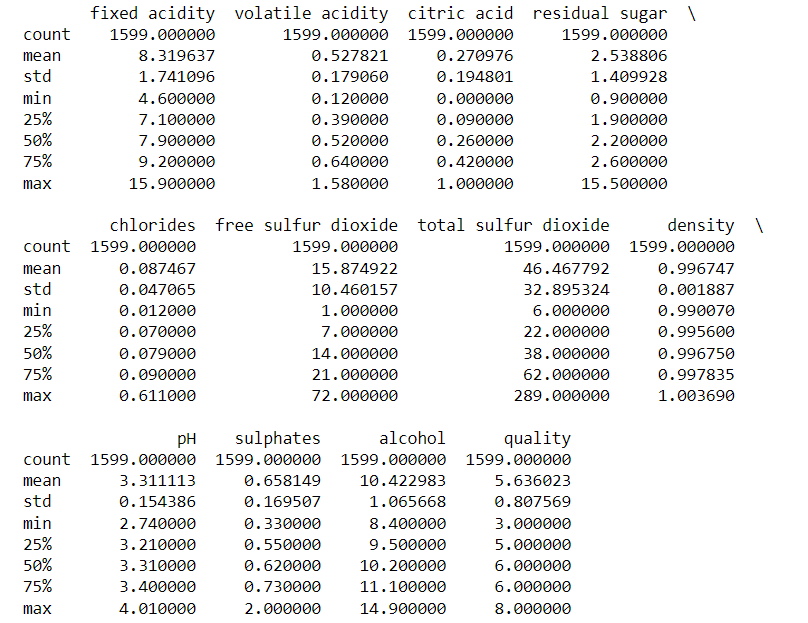
2. Data preprocessing including visualization

# Display summary statistics of the dataset

summary\_stats = dataset.describe()

print(summary\_stats)

Output:



import matplotlib.pyplot as plt

import seaborn as sns

# Countplot of wine quality

plt.figure(figsize=(6, 4))

sns.countplot(data=dataset, x='quality')

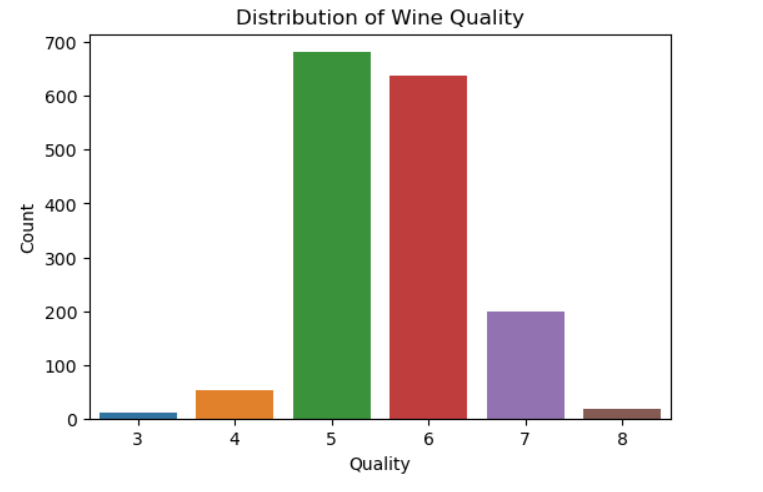
plt.title('Distribution of Wine Quality')

plt.xlabel('Quality')

plt.ylabel('Count')

plt.show()

Output:



3 & 4. Machine Learning Model building and Evaluate the model

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

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy\_score, classification\_report

X = dataset.drop('quality', axis=1)

y = dataset['quality']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

clf = RandomForestClassifier(n\_estimators=100, random\_state=42)

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

y\_pred = clf.predict(X\_test)

# Evaluate the model

accuracy = accuracy\_score(y\_test, y\_pred)

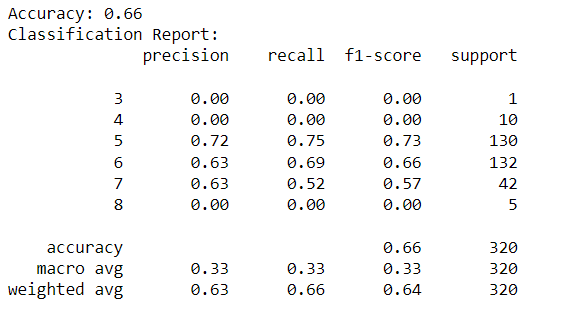
classification\_rep = classification\_report(y\_test, y\_pred)

print(f'Accuracy: {accuracy:.2f}')

print('Classification Report:')

print(classification\_rep)

Output:



5. # Create a random observation

new\_observation = pd.DataFrame({

'fixed acidity': [7.0],

'volatile acidity': [0.4],

'citric acid': [0.25],

'residual sugar': [2.0],

'chlorides': [0.045],

'free sulfur dioxide': [35.0],

'total sulfur dioxide': [120.0],

'density': [0.99],

'pH': [3.2],

'sulphates': [0.6],

'alcohol': [11.0]

})

# Use the trained model to predict the quality of the new observation

predicted\_quality = clf.predict(new\_observation)

print(f'Predicted Wine Quality: {predicted\_quality[0]}')

Output:

