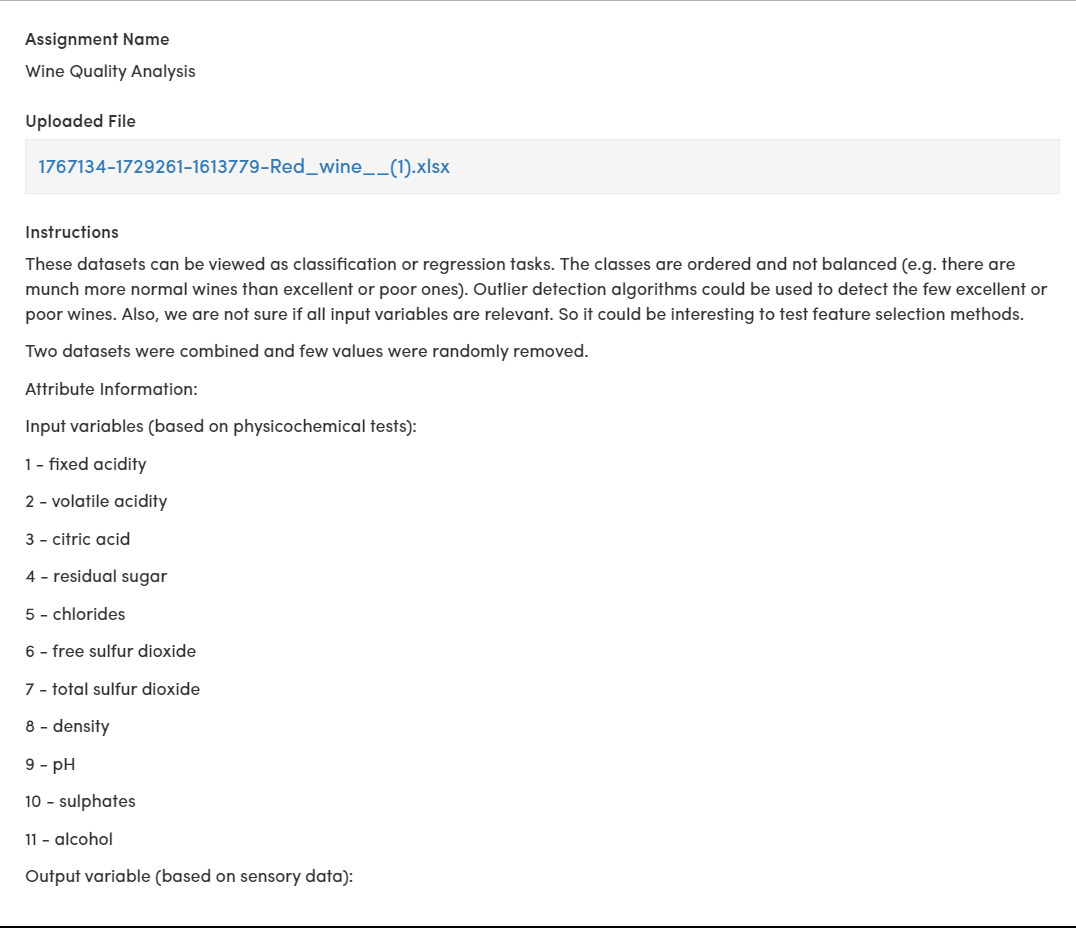
MAJOR PROJECT

WINE QUALITY ANALYSIS



CODE:

# Import necessary libraries

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

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

from sklearn.preprocessing import StandardScaler

from sklearn.ensemble import RandomForestClassifier

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

# Load the dataset

data = pd.read('Wine.csv')

print(data.head()) # To view the first few rows of the dataset

print(data.info()) # To get information about data types and missing values

print(data.describe()) # To get statistical summary of the dataset

print(data.isnull().sum())

X = data.iloc[:, :-1] # All columns except the last one (quality)

y = data['quality'] # The quality column

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

sc = StandardScaler()

X\_train = sc.fit\_transform(X\_train)

X\_test = sc.transform(X\_test)

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

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

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

print("Confusion Matrix:\n", confusion\_matrix(y\_test, y\_pred))

print("Classification Report:\n", classification\_report(y\_test, y\_pred))

print("Accuracy Score:", accuracy\_score(y\_test, y\_pred))

feature\_importances = pd.Series(classifier.feature\_importances\_, index=X.columns)

feature\_importances.nlargest(10).plot(kind='barh')

plt.title("Top 10 Important Features")

plt.show()