**American International University-Bangladesh**



**Course:** INTRODUCTION TO DATA SCIENCE

**Assignment Title:** Final-term assignment

**Submitted by:**

Name: Mir Faris  
Id: 20-42241-1  
Section: B

**Submitted to:**

TOHEDUL ISLAM

Assistant Professor, Computer Science  
American International University-Bangladesh

**Submission Date:** August 16, 2023

**Dataset Description**

**Name**: Red Wine Quality

**Source**: https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009

**Attributes**: There are total 12 attributes in the dataset.

1. **Ph:** PH is an important parameter in evaluating the acid–base balance of water. WHO has recommended maximum permissible limit of pH from 6.5 to 8.5. The current investigation ranges were 6.52–6.83 which are in the range of WHO standards.
2. **Fixed Acidity:** This characteristic indicates how much non-volatile acids are present in the wine. These acids support the wine's overall flavor and harmony. Higher fixed acidity wines may taste harsh or sour.
3. **Citric Acid:** Citrus fruits naturally contain citric acid, a mild organic acid. It is occasionally added to wines to increase their acidity and impart a zesty, fresh flavor.
4. **Volatile Acidity:** The term "volatile acidity" describes the wine's presence of volatile acids, particularly acetic acid. A vinegar-like flavor and disagreeable scent might arise from higher amounts of volatile acidity.
5. **Residual Sugar:** It indicates the amount of sugar left over after fermentation is complete. Wines with higher residual sugar are sweeter, while those with lower levels are drier.
6. **Chlorides:** Chlorides in wine can contribute to its saltiness or salinity. Elevated chloride levels might impact the taste negatively, making the wine taste salty.
7. **Free Sulfur Dioxide:** Free sulfur dioxide acts as a preservative and anti-oxidant in wine. It helps prevent the growth of unwanted microorganisms and oxidation.
8. **Total Sulfur Dioxide:** Total sulfur dioxide is the sum of both free and bound forms of sulfur dioxide in the wine. It's an important parameter in winemaking to maintain wine quality and stability.
9. **Density:** Density is a measure of the mass of the wine per unit volume. It can provide insights into the wine's overall composition and potentially its alcohol content.
10. **Sulphates:** Sulfites (sometimes known as sulphates) are substances that are frequently used as preservatives in wine. They may help support the stability of the wine's flavor and aroma.
11. **Alcohol:** Alcohol content in wine affects its body, mouthfeel, and perceived warmth. Higher alcohol content can make the wine feel fuller and richer.
12. **Quality:** On a scale of 3 to 9, the quality attribute most likely signifies a subjective evaluation of the wine's general quality. Higher values typically represent higher quality in terms of taste and other attributes

**Instances**: There are total 1599 instances in the dataset.

**Classes**: The class attribute of the dataset is “Quality”. The dataset is classified with 6 classes;

* 3, 4, 5, 6, 7, 8

**Performed Tasks**

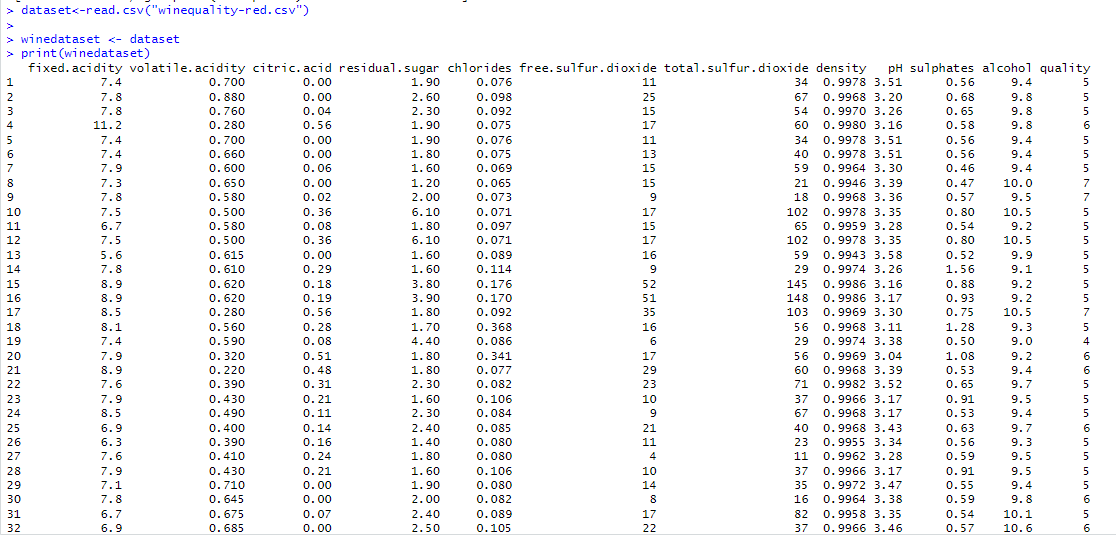
1. Import, clone and view dataset

Fig 1: Dataset

1. Checking for values:





Missing value free dataset



There are 6 classes

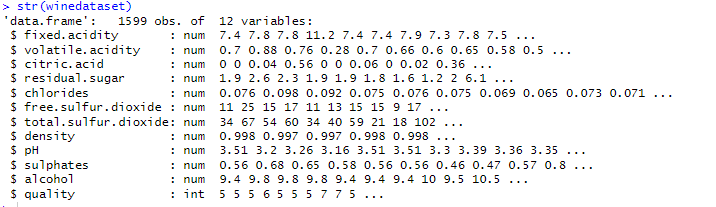


Fig 2: Checking the types if needed to convert numeric.

1. Normalizing min-max excluding targeted attribute:

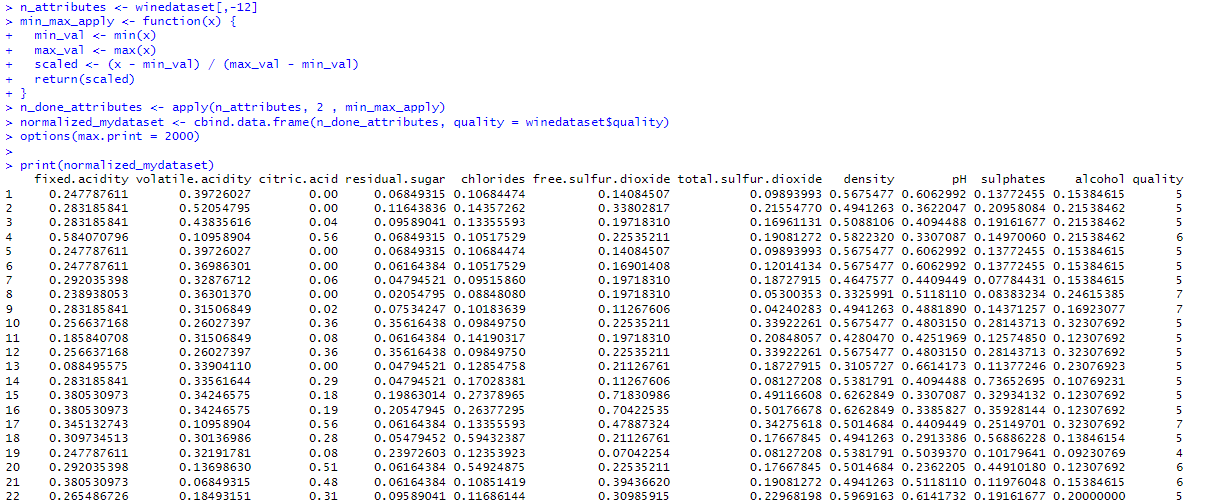
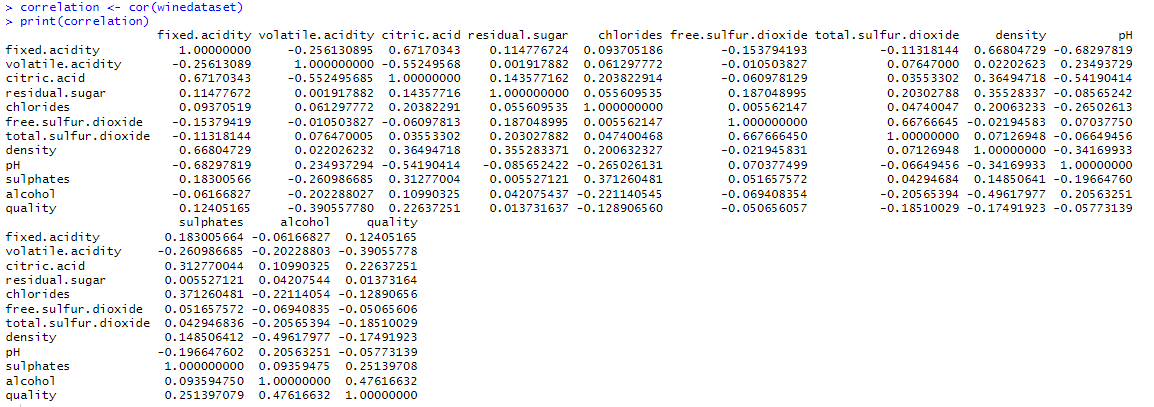
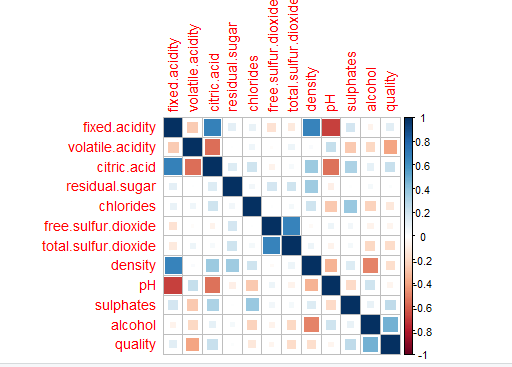


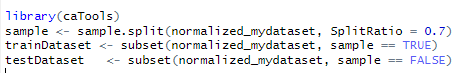
Fig:3 Min-Max Normalization



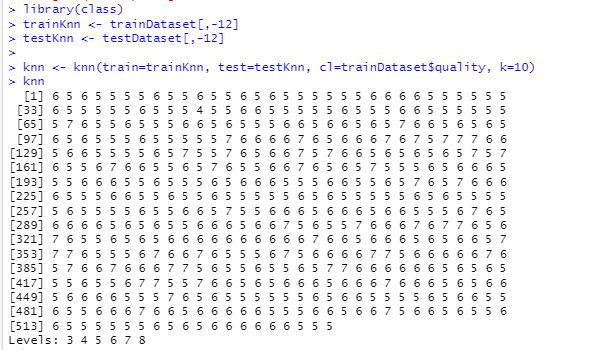
Fig 4: Correlation



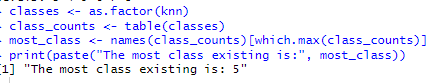
1. Splitting the dataset into two and apply KNN:



Train and Test Dataset



10-fold Cross Validation



Class 5 maximum appeared



Calculating accuracy

The accuracy value of 56.57895% means that the k-nearest neighbors (KNN) model

correctly predicted the quality of approximately 56.58% of the instances in test dataset.

Fig 5: Performing KNN steps

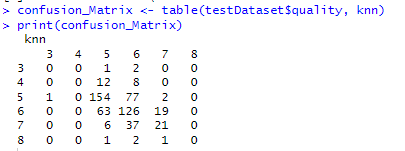


Fig 6: Generate confusion matrix

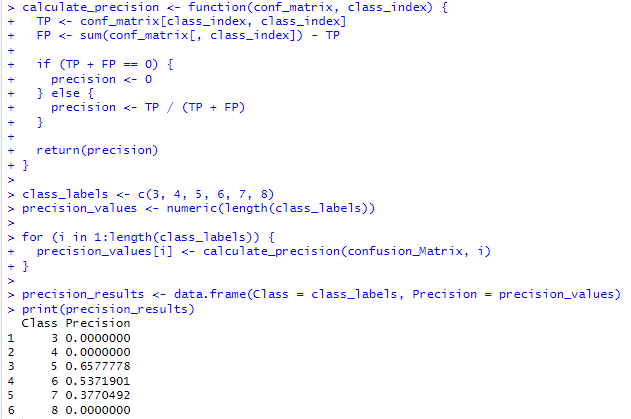


Fig 7: Calculating Precision

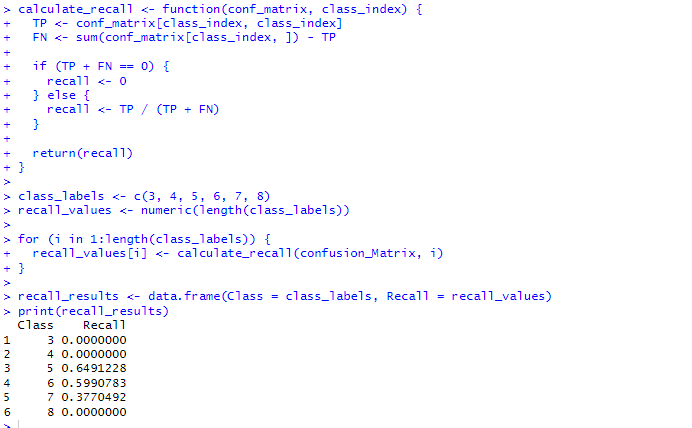


Fig 8: Calculating Recall

A confusion matrix is a table that is used to define the performance of a classification algorithm. Here the class quality determines if wine is perceived for wines quality rankings.

3: Low quality, 4: Below average quality, 5: Average quality, 6: Above average quality, 7: High quality, 8: Very high quality.

**True Positives (TP):** These values correspond to instances that were correctly predicted as their true quality class. For example, in class 5, 148 instances were correctly predicted as class 5, in class 6, 130 instances were correctly predicted as class 6, and in class 7, 23 instances were correctly predicted as class 7.

**True Negatives (TN):** Since there are no instances predicted as classes 3, 4, and 8 in the matrix, there are no true negatives values present.

**False Positives (FP):** These values represent instances that were incorrectly predicted as a certain class. For example, for class 4, 2 instances of class 3 were wrongly predicted as class 4, and for class 5, there were instances of class 3, 4, and 6 wrongly predicted as class 5. Similarly, false positives occurred for classes 6 and 7 as well.

**False Negatives (FN):** These values represent instances that belong to a certain class but were wrongly predicted as something else. For example, for class 3, 2 instances of class 4 were wrongly predicted as class 3. In class 4, there were instances of class 5, 6, and 7 wrongly predicted as class 4. Likewise, false negatives occurred for classes 5, 6, 7, and 8.