**White Wine Quality Analysis**

***Background:***

Quality is a concern for both wine industry experts and customers. Wine quality evaluation using the traditional (professional) method requires time. Machine learning models are essential tools for replacing human labor in the modern world. Several characteristics can be utilized to predict the wine quality in this situation, however not all of the characteristics will be beneficial for a more precise prediction. Our thesis study is concentrated on what properties of wine are necessary in order to acquire a positive result. For this investigation, we used a White Wine quality dataset. These datasets can be used to perform regression or classification tasks. These datasets have the following 12 attributes:

Fixed acidity:

Most of the acids in wine are either fixed or nonvolatile. The total number of acids in a sample defines how acidic it is. To quantify the complete collection of them, use grouping techniques or total acidity. A category of organic acids with limited volatility, such as malic, lactic, tartaric, or citric acids, are referred to as having "fixed acidity." It belonged to the sample itself.

Volatile acidity:

The quantity of acetic acid in wine, which, at high levels, can taste unpleasant like vinegar. The group of formic acid, acetic acid, propionic acid, and butyric acid that can be extracted from a material using a distillation process is referred to as "volatile acidity."

Citric acid:

Wines include trace levels of citrus acid, which can improve flavor and freshness. Citric acid is a weak organic acid that has no color. It naturally exists in citrus fruits. It is a stage in the citric acid cycle, a biochemical pathway via which all aerobic organisms generate energy.

Residual sugar:

After fermentation has ended, it's unusual to discover wines that have less sugar per liter than 1 gram. Remaining sugar refers to the sugars in a finished wine that have not yet undergone fermentation. The sugar content is expressed in grams per liter (g/l). The residual sugar concentration (RS) of a wine affects its sweetness, and in the EU, the RS level is associated with certain labeling terms.

Chlorides:

The salt content of the wine. Because of the ions extracted from the skins during the fermentation process, white winemaking causes a greater extraction of chloride. In order to prevent finished wine from having more chloride than is permitted by law, red juice should only contain 356 mg/L of chloride ions.

Free sulfur dioxide:

The free form of SO2 can be found in equilibrium with both molecular SO2 and the bisulfite ion. Sulfites that are available for reaction are both antibacterial and antioxidant. Bound sulfites are those sulfites that have permanently and irreversibly interacted with other molecules in the wine medium. By combining the free and bound sulfites, the total sulfite concentration is calculated.

Total sulfur dioxide:

SO2amounts, both free and bound; free SO2 is largely undetectable in wine at low SO2 concentrations. Total Sulfur Dioxide (TSO2) is the sum of the sulfur dioxide (SO2) that is free in the wine and the sulfur dioxide (SO2) that is attached to other substances in the wine, such as aldehydes, pigment, or sugar.

Density:

Depending on the amount of alcohol and sugar, water has a density that is nearly identical to that of water. An instrument called a hydrometer is used to gauge the density of liquids. Winemakers use this tool to measure the density of juice, fermenting wine, and finished wine in relation to pure water. It was sealed glass tube with a weighted bulb at a end. Specific gravity refers to this proportion

PH:

Measures the acidity or basicity of a wine on a scale from 0 to 14. most wines lie around the range of 3 to 4. Wines with higher pH levels (>3.65) may experience a number of challenges during vivification and maturation. First off, bacteria have a higher propensity to taint wines with high pH values. In the past, sulfur dioxide has been used to preserve wines as they mature.

Sulphates:

A component of wine that can increase the amount of the antibacterial gas sulfur dioxide (S02). Grape sulfites, one of the tens of thousands of chemical byproducts created during fermentation, are naturally found in all wines in small proportions. The winemaker also adds sulfites to preserve and protect the wine from yeast and microbial contamination. Some wine drinkers who have sulfur allergies may have headaches and clogged sinuses after a few sips. The amount of sulfur dioxide gas (S02), which acts as an antibacterial, can be increased by adding it to wine.

Alcohol:

This is the wine's percentage alcohol concentration.

Quality: output parameter

***Research Questions:***

*Applying the White Wine dataset, which contains the 12 attributes that are described above, we are using the Naive Bayes and Decision Tree machine learning methods to predict the quality of White Wine.*

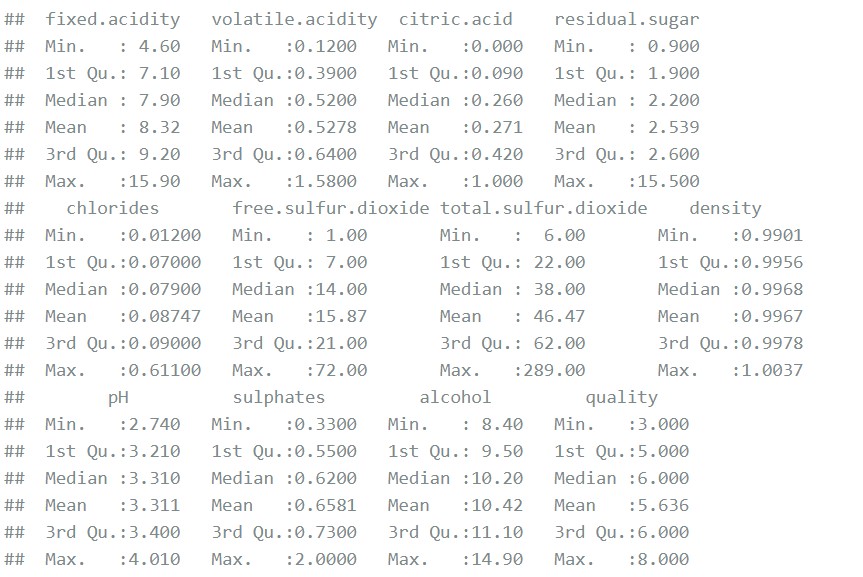
***Potential Solution:***

*We consider the quality score between 7 and 10 to be "Excellent" by using the Quality attribute in the White Wine dataset, and we may estimate the White Wine quality using this information.*

***Methods and Results:***

***Exploratory Data Analysis :***

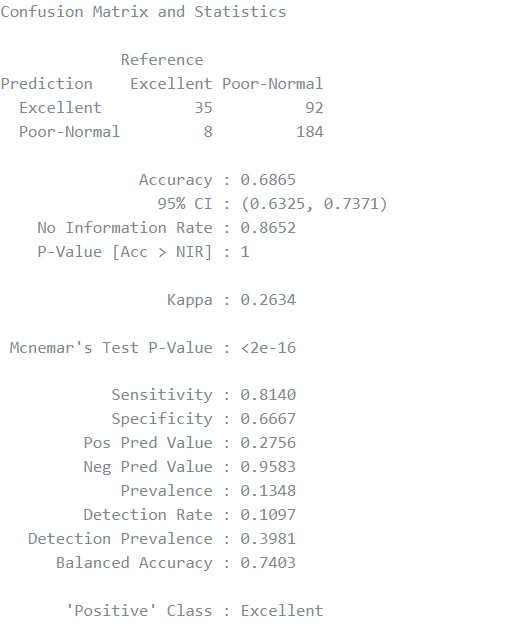
By using White Wine data set the below results are obtained



***Naïve Bayes:***

*It is a classification method that relies on the independence of predictors in Bayes' Theorem. To put it simply, a Naive Bayes classifier believes that the presence of one feature in a class has nothing to do with the presence of any other features.*

*Based on Naïve Bayes technique we can able to generate confusion matrix and statistics as follows* ***Result:***

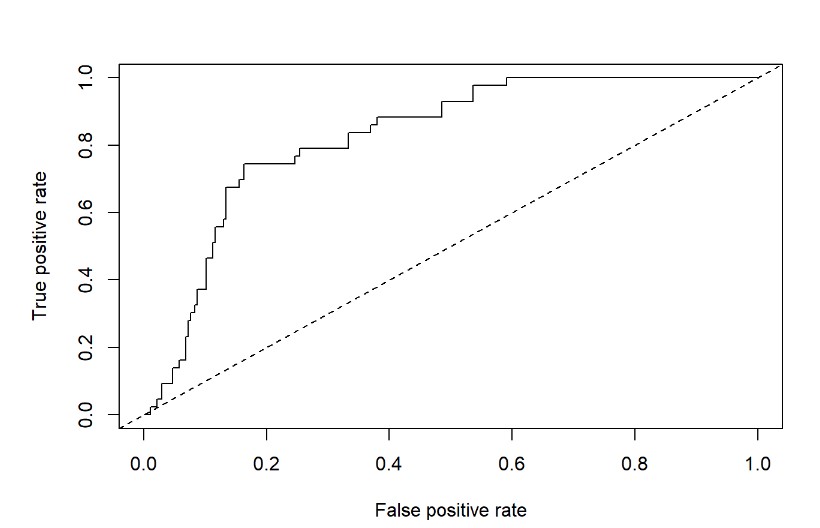


**Confusion Matrix :**

|  |  |  |
| --- | --- | --- |
| Prediction | Excellent | Poor-Normal |
| Excellent | 35 | 92 |
| Poor-  Normal | 8 | 184 |

**Receiver Operating Curve(ROC):**

By using this receiver operating curve, we can able to generate a correlation between True positivity rate and false positivity rate.



Area Under ROC Curve(AUC):

AUC displays a big area under the ROC curve; if the parameter's value is close to 1, the model is good.

Code:

auc <- performance(prediction.obj = wine\_roc, measure = "auc") auc@y.values Result:



Interpretation:

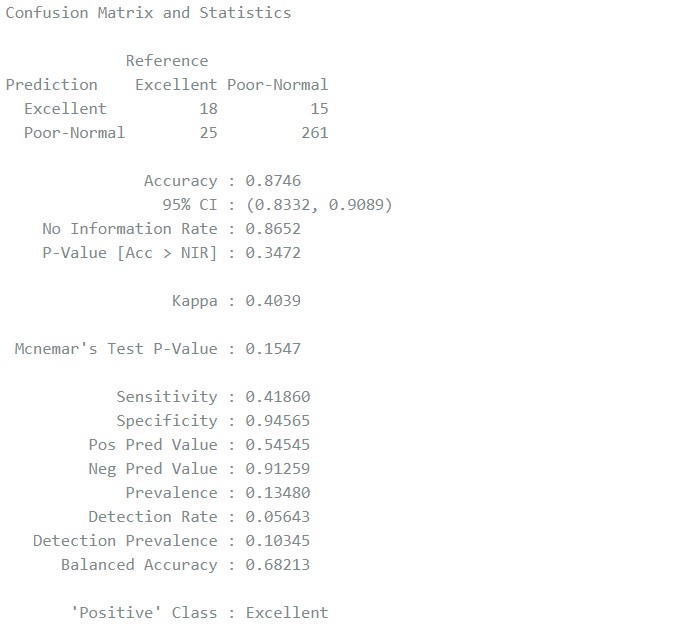
|  |  |
| --- | --- |
| Measure | value |
| Accuracy | 68.65% |
| Sensitivity | 81.4% |
| Specificity | 66.57% |
| Precision | 27.56% |

**Decision Tree:**

In supervised machine learning, which trains models using labelled input and output datasets, decision trees are a method employed. The method is mostly employed to address classification issues, which include categorizing or classifying an object using a model.

*Based on Decision tree technique we can able to generate confusion matrix and statistics as follows*

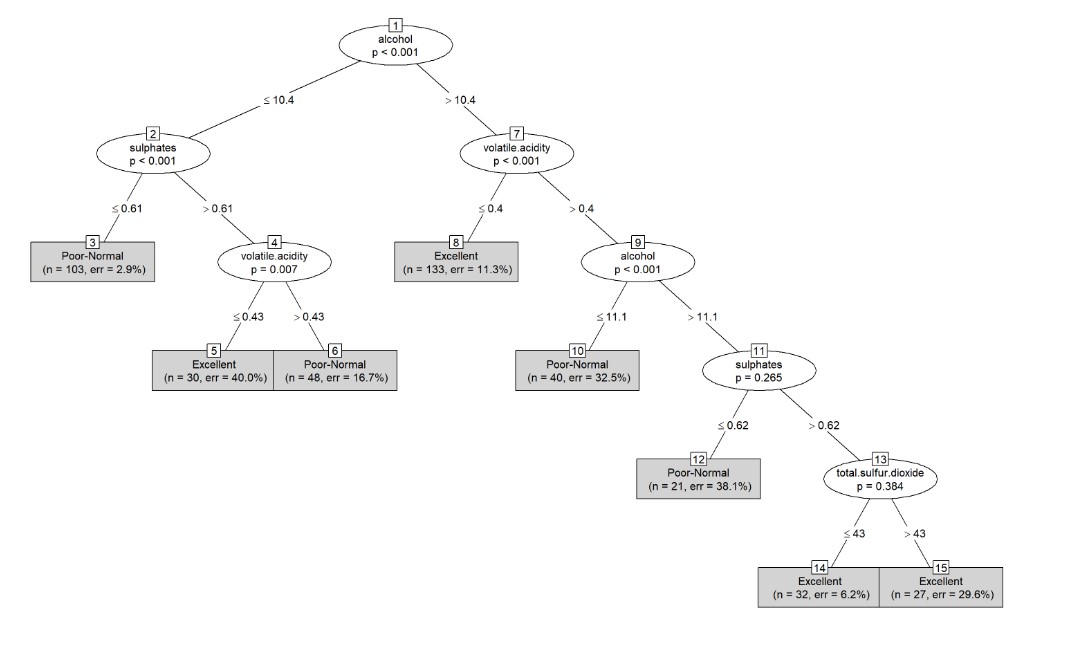
**Result:**



**Confusion matrix:**

|  |  |  |
| --- | --- | --- |
| Prediction | Excellent | Poor-Normal |
| Excellent | 18 | 15 |
| Poor-  Normal | 25 | 261 |

**Graph:**



**Interpretation:**

|  |  |
| --- | --- |
| Measure | value |
| Accuracy | 73.98% |
| Sensitivity | 79.07% |
| Specificity | 73.19% |
| Precision | 31.48% |

**Model Compare:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Sensitivity | Specificity | Precision |
| Naïve Bayes | 0.6865 | 0.8140 | 0.6667 | 0.2756 |
| Decision Tree | 0.7398 | 0.7907 | 0.7319 | 0.3148 |

**Discussion:**

Based on results mentioned above we can state that by using Decision tree technique, we can able to get good results in Accuracy, Specificity, Precision except Sensitivity when compared to that of Naïve Bayes technique. In order to improve White Wine quality analysis we may use Random forest technique to get best results.

**Limitations:**

Only physiochemical and sensory characteristics are available due to logistical and privacy concerns, as the dataset's author noted. For instance, there are no data on grape varieties, wine brands, etc.

**References:**

1. <https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/>
2. <https://www.kaggle.com/>
3. P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties.