**PREDICTING QUALITY OF WINE**

**Team: Rodgers**

**Chapter 1**

**PREDICTING QUALITY OF WINE**

**1.1.Introduction to Machine Learning:**

Machine learning (ML) is a category of [algorithm](https://whatis.techtarget.com/definition/algorithm) that allows software applications to become more accurate in predicting outcomes without being explicitly programmed. The basic premise of machine learning is to build algorithms that can receive input data and use [statistical analysis](https://whatis.techtarget.com/definition/statistical-analysis) to predict an output while updating outputs as new data becomes available.

Machine learning focuses on the development of Computer Programs that can change when exposed to new data. Machine learning involves computer to get trained using a given data set, and use this training to predict the properties of a given new data.

**Python:**

Python is a popular programming language. It was created by Guido van Rossum, and released in 1991.

Python is a widely used general-purpose, high level programming language. It was initially designed by Guido van Rossum in 1991 and developed by Python Software Foundation. It was mainly developed for emphasis on code readability, and its syntax allows programmers to express concepts in fewer lines of code.

**Artificial Intelligence:**

“The science and engineering of making intelligent machines, especially intelligent computer programs”.

Artificial Intelligence is an approach to make a computer, a robot, or a product to think how smart human think. AI is a study of how human brain think, learn, decide and work, when it tries to solve problems. And finally this study outputs intelligent software systems. The aim of AI is to improve computer functions which are related to human knowledge, for example, reasoning, learning, and problem-solving.

**1.2. Objectives of Research:**

This analysis examines data from wine concentrating on relationship between a key list of sweetness, acidity, tannin levels, alcohol using them to try and predict the quality of wine.

In general, models such as this derived of much larger data sets will eventually evolve to produce better quality out comes with less invasive methods by predicting possible quality of wine.

Where we find out the Ph and alcohol quality in it to find the best quality of wine.so that the prediction of quality takes place.

The dataset contains information about red and white wine.

The dataset has 1600 data points of each wine with 12 characteristics(attributes).

Wine quality is measured on 3,4,5(low)-6,7,8(high) scale.

Management wants to understand the characteristics of these wines and how different ingredients affect the quality.



**1.3. Problem statement:**

## **Using Linear Regression to predict the quality of Wine**

World’s most famous wine Bordeaux Wine: Large differences in price and quality between years, although wine is produced in a similar way. Traditionally, experts Expert tasters predict which ones will be good. #Goal: Predicting the wine quality based on the given historical data of “Bordeaux Wines”. # Linear Regression.

**3.Data Collection:**

In wine tasting, the term **acidity** refers to the fresh, tart and sour attributes of the wine. Three primary acids are found in wine grapes — tartaric, malic and citric acids. They are evaluated in relation to how well the acidity balances out the sweetness and bitter components of the wine, such as tannins.

* **Fixed Acidity**  
    
  Titratable acidity, sometimes referred to as fixed acidity, is a measurement of the total concentration of titratable acids and free hydrogen ions present in your wine. A [litmus paper](https://en.wikipedia.org/wiki/Litmus) can be used to identify whether a given solution is acidic or basic. The most common titratable acids are tartaric, malic, citric and carbonic acid. These acids, along with many more in smaller quantities, either occur naturally in the grapes or are created through the fermentation process.
* **Volatile Acidity**  
    
  Volatile acidity is mostly caused by bacteria in the wine creating acetic acid — the acid that gives vinegar its characteristic flavour and aroma — and its by product, ethyl acetate. Volatile acidity could be an indicator of spoilage, or errors in the manufacturing processes — caused by things like damaged grapes, wine exposed to air, and so on. This causes acetic acid bacteria to enter and thrive, and give rise to unpleasant tastes and smells. Wine experts can often tell this just by smelling it!
* **Citric Acid**  
    
  Citric acid is generally found in very small quantities in wine grapes. It acts as a preservative and is [added to wines](http://wineserver.ucdavis.edu/industry/enology/methods_and_techniques/reagents/citric_acid.html) to increase acidity, complement a specific flavor or prevent ferric hazes. It can be added to finished wines to increase acidity and give a “fresh” flavor. Excess addition, however, can ruin the taste.
* **Residual Sugars**  
    
  **Residual Sugar**, or RS for short, refers to any natural grape **sugars** that are left over after fermentation ceases (whether on purpose or not). The juice of **wine** grapes starts out intensely sweet, and fermentation uses up that **sugar** as the yeasts feast upon it.  
    
  During winemaking, yeast typically converts all the sugar into alcohol making a **dry** wine. However, sometimes not all the sugar is fermented by the yeast, leaving some [sweetness leftover](http://winefolly.com/update/sugar-in-wine-misunderstanding/).  
    
  To make a wine that tastes good, the key is to have a perfect balance between the sweetness and the sourness in the drink.
* **Chloride**  
    
  The amount of chlorides present in a wine is usually an indicator of its “[saltiness](http://www.aromadictionary.com/articles/salt_article.html).” This is [usually influenced](https://www.researchgate.net/publication/276444447_Chloride_concentration_in_red_wines_Influence_of_terroir_and_grape_type) by the territory where the wine grapes grew, cultivation methods, and also the grape type. Too much saltiness is considered undesirable. The right proportion can make the wine [more savoury](http://wineoscope.com/2015/10/02/when-a-wine-is-salty-and-why-it-shouldnt-be/).
* **Total sulphur dioxide**

amount of free and bound forms of S02; in low concentrations, SO2 is mostly undetectable in wine, but at free SO2 concentrations over 50 ppm, SO2 becomes evident in the nose and taste of wine

* **Free sulphur dioxide**

the free form of SO2 exists in equilibrium between molecular SO2 (as a dissolved gas) and bisulfide ion; it prevents microbial growth and the oxidation of wine.

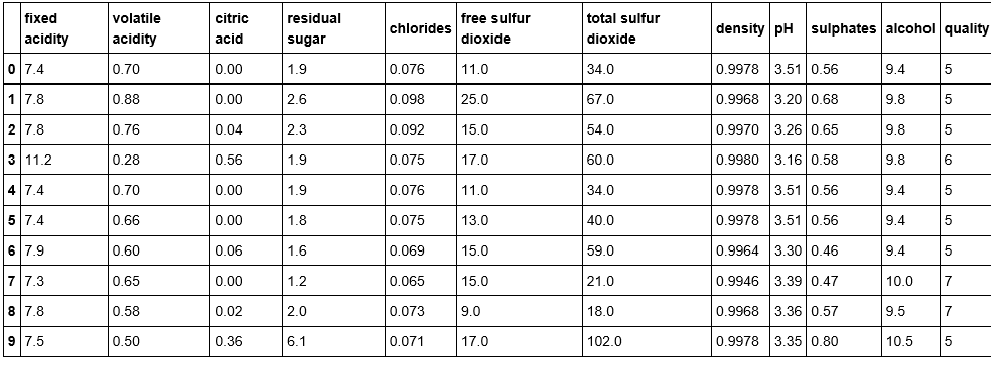
* **Density**  
    
  Also known as specific gravity, it can be used to measure the alcohol concentration in wines. During fermentation, the sugar in the juice is converted into ethanol with carbon dioxide as a waste gas. Monitoring the density during the process allows for optimal control of this conversion step for highest quality wines. Sweeter wines generally have higher densities.
* **pH**  
    
  pH stands for power of hydrogen, which is a measurement of the hydrogen ion concentration in the solution. Generally, solutions with a pH value less than 7 are considered acidic, with some of the strongest acids being close to 0. Solutions above 7 are considered alkaline or basic. The pH value of water is 7, as it is neither an acid nor a base.
* **Sulphates**  
    
  Sulphates are salts of sulfuric acid. They aren’t involved in wine production, but some beer makers use calcium sulphate — also known as brewers’ gypsum — to correct mineral deficiencies in water during the brewing process. It also adds a bit of a “sharp” taste.
* **Alcohol**  
    
  Ah yes, alcohol — the key to a successful party! Alcoholic beverages existed from at least the Neolithic period (10,000 BC). Drinking it in small amounts gives you warm fuzzy feelings inside, and makes you more sociable. Of-course, higher doses can also make you pass out.
* **Quality**

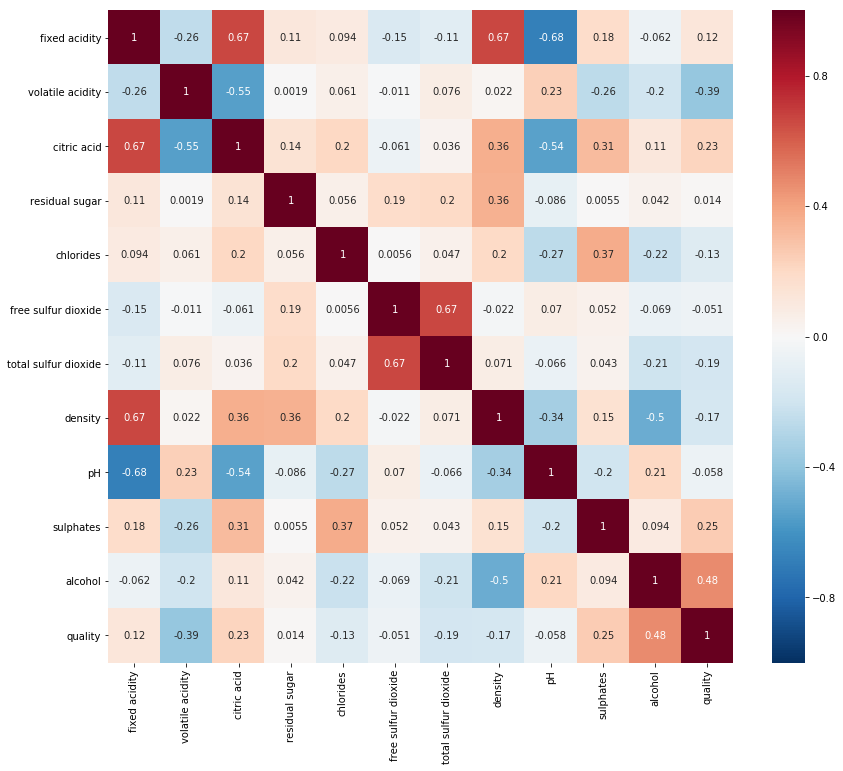
output variable (based on sensory data, score between 3 and 8).

**4.Methodology:**

**4.1 Exploratory Data Analysis:**

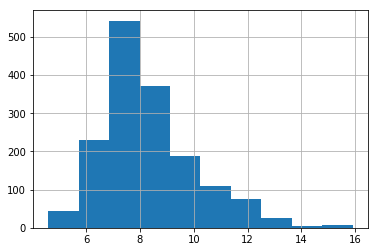
**4.1.1 Figures and Tables:**

****

****

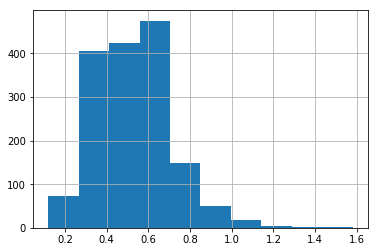
data1['fixed acidity'].hist(bins=10)

<matplotlib.axes.\_subplots.AxesSubplot at 0x18ffd6f5c50>

****

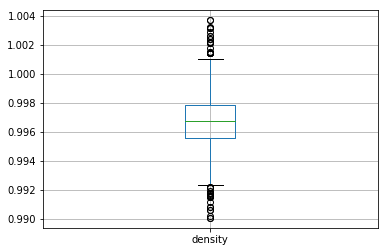
data1['volatile acidity'].hist(bins=10)

<matplotlib.axes.\_subplots.AxesSubplot at 0x18fffadef28>



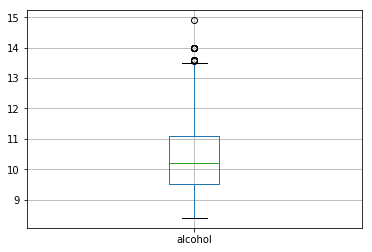
data1.boxplot(column='density')

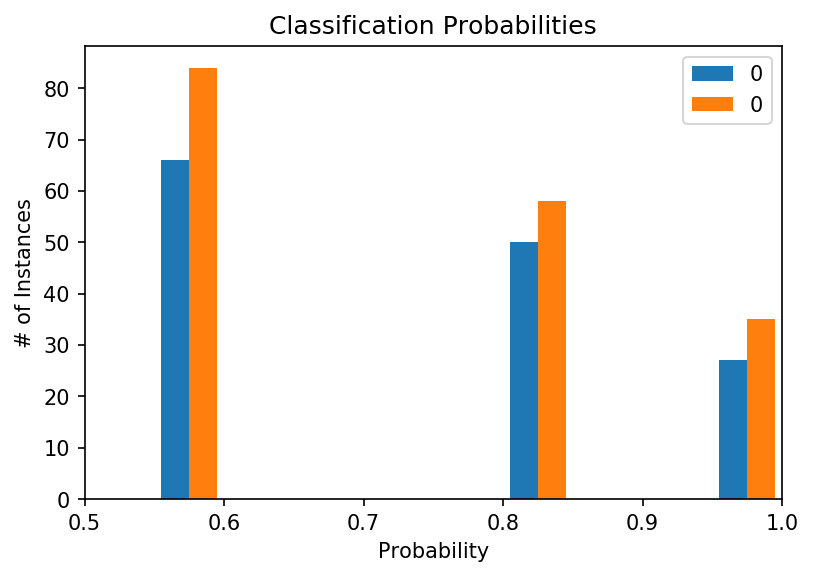
<matplotlib.axes.\_subplots.AxesSubplot at 0x18fffb75eb8>



data1.boxplot(column='alcohol')

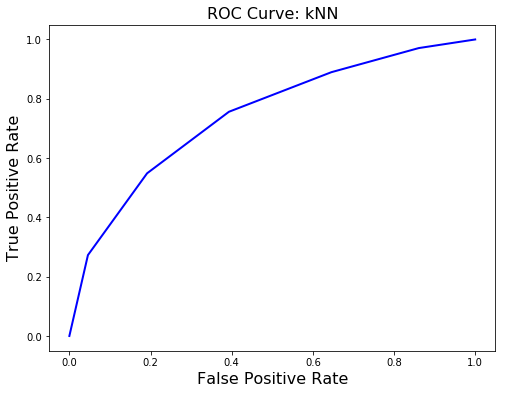
<matplotlib.axes.\_subplots.AxesSubplot at 0x18fffbd0ef0>





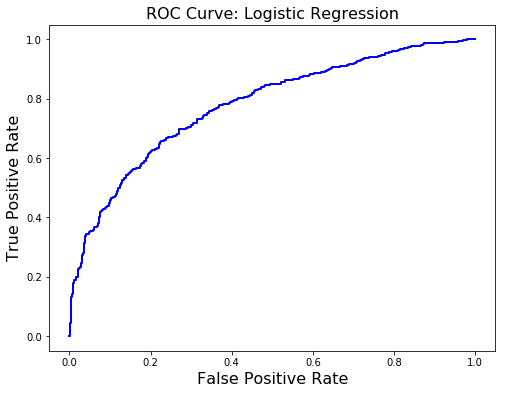
ROC\_curve('kNN', y\_train, clf\_scores)

AUC Score (kNN): 0.74



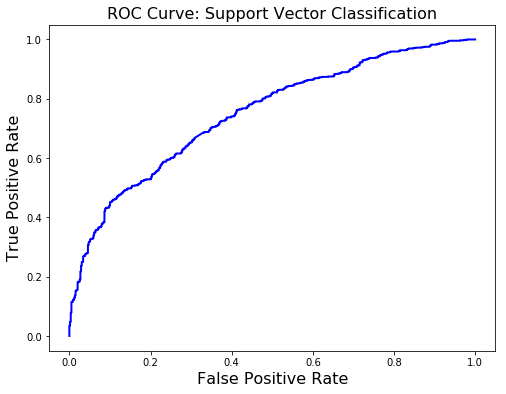
ROC\_curve('Logistic Regression', y\_train, lr\_scores)

AUC Score (Logistic Regression): 0.78



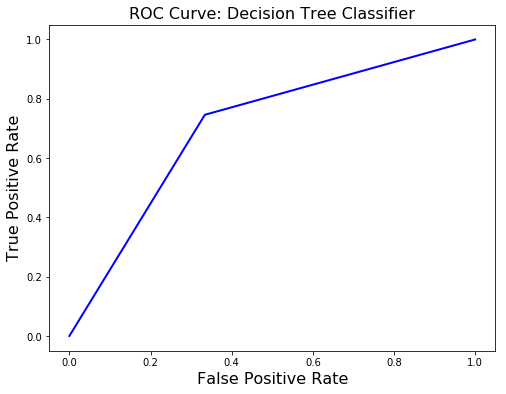
ROC\_curve('Support Vector Classification',y\_train,svc\_scores)

AUC Score (Support Vector Classification): 0.75



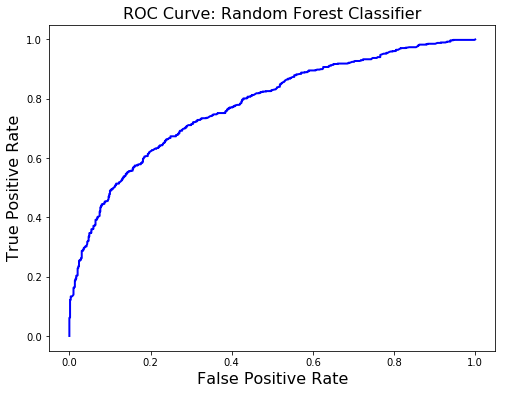
ROC\_curve('Decision Tree Classifier',y\_train,dtc\_clf\_scores)

AUC Score (Decision Tree Classifier): 0.71



ROC\_curve('Random Forest Classifier',y\_train,rmf\_clf\_scores)

AUC Score (Random Forest Classifier): 0.78



**4.2. Data Modelling:**

**TASK 1: Convert data in excel file or CSV format**

At first convert data from text to column in excel sheet before importing the data. Because, unstructured data were found in the dataset. Then, save the data in csv format.

**TASK 2: Import data**

Import data to data in csv format. No clean-up is required as the data has already been cleaned previously.

wine quality <- read.csv(file. Choose(), header=T, sep=",", check. names=TRUE)

**TASK 3: Check characteristics and data relationship**

* **Check data characteristics with following codes -**

head( wine quality)

str(wine quality)

summary(wine quality)

class(wine quality)

n row(wine quality)

n col(wine quality)

**5. Findings and Suggestions:**

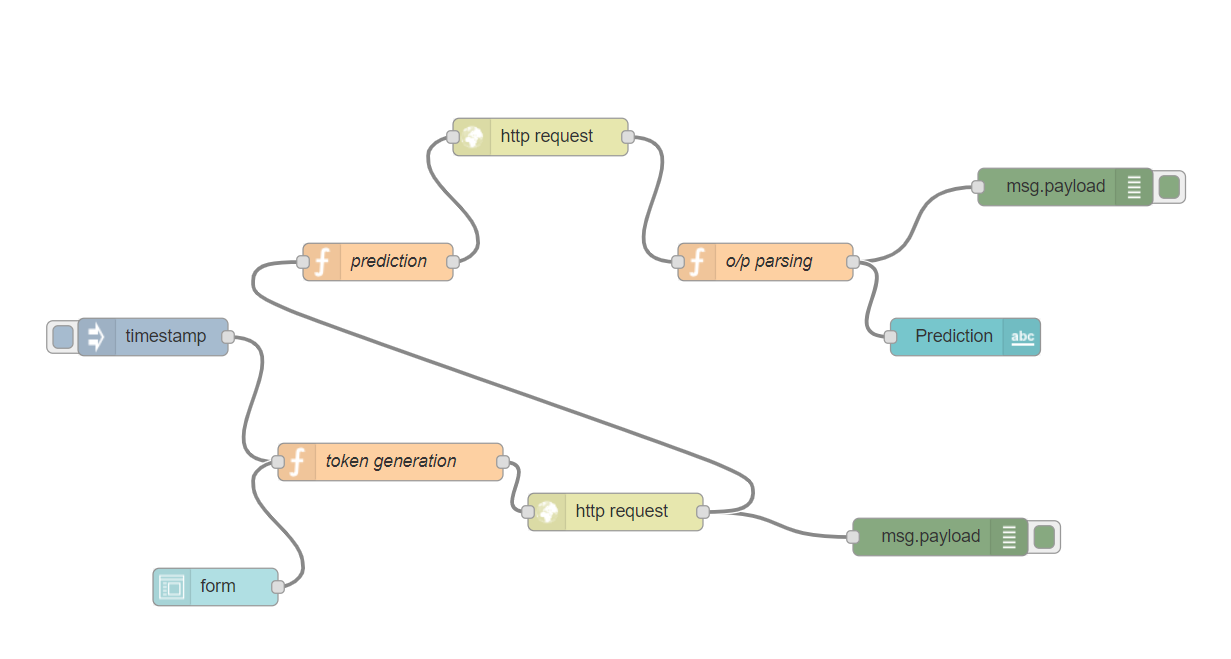
**Findings:** Datatype is data frame. Number of attributes is 12.

**Find out missing values from dataset**

supply(wine quality, function(x) sum( [is.n](http://is.na/)a(x))) sum(!complete. cases(wine quality))

**Findings:** There is no missing values exist in the dataset as set has already been cleaned

**7. flow chart**



**6. Conclusion:**

This paper was motivated by the need for research that could improve the understanding of how the quality of the wine is influenced by its different physicochemical present in it. Out of the thirteen attributes, the statistically significant attribute that influence the quality of the wine is an essential finding. By employing linear regression analysis, we come up with a model that highlights the significant attributes in both sets. The result of this regression analysis will be helpful in production and in quality prediction by studying the impact of those significant attributes in predicting the quality. There is space for further analysis to reveal the more interesting pattern and to employ rigor analytical tool to augment a more sophisticated model.

### **Managerial Relevance**

An important insight is that how much each of the chemical components contributes to the quality of the wine and how can we grade the quality level of a newly produced wine. This helps the producer to identify the distinguishing factors affecting the quality level and thereby fix a reasonable price for wine. It’s easier to filter and choose the raw material (grapes, fruits or vegetable) with a prior laboratory test to check required chemicals needed for certain quality of wine which is demanded in the market. Moreover, In different culture and nation, there is demand for the specific type of wine due to variation in climatic condition or taste preference of peoples residing in this areas. In such scenarios, the producer can collect the data that reveals the taste preferences of different culture and consequently prepare a suitable wine for such area to improve his marketing, sales and of course revenue.