

Wine Quality Analysis and Prediction

Objective

Use data analysis and machine learning to determine which all physiochemical variables contribute to the quality of the wine.

Data Analysis (exploration)

Data Statistics for red wine (Fig.1) -

- Density and pH are normally distributed
- fixed acidity, volatile acidity, free sulfur dioxide, total sulfur dioxide, sulphates and alcohol has got some outliers
- residual sugar and chloride has got extreme outliers

Data Statistics for white wine (Fig-2) -

- Fixed acidity, density, and pH are normally distributed
- volatile acidity, citric acid, sulphates and alcohol has got some outliers
- residual sugar, free sulfur dioxide, total sulfur dioxide, and chloride has got extreme outliers

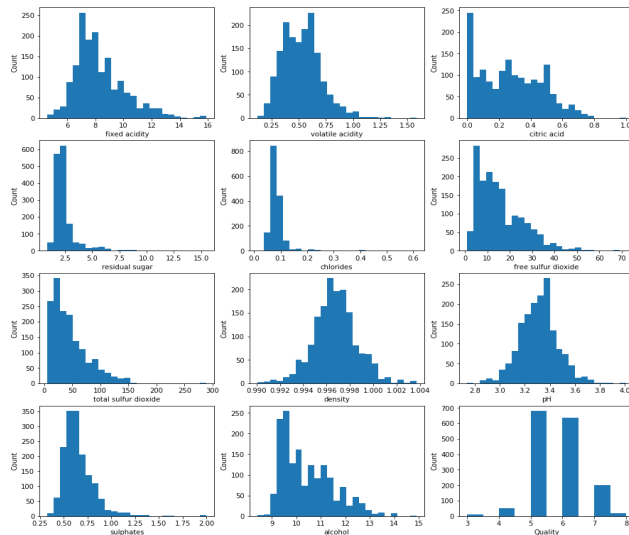


Fig-1

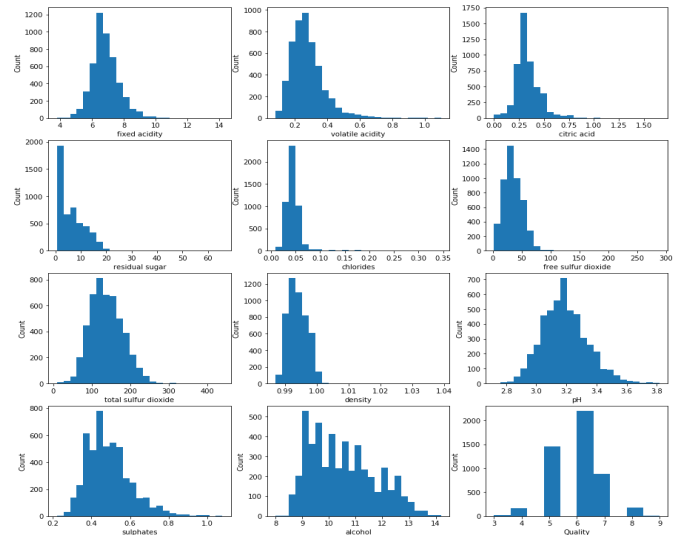
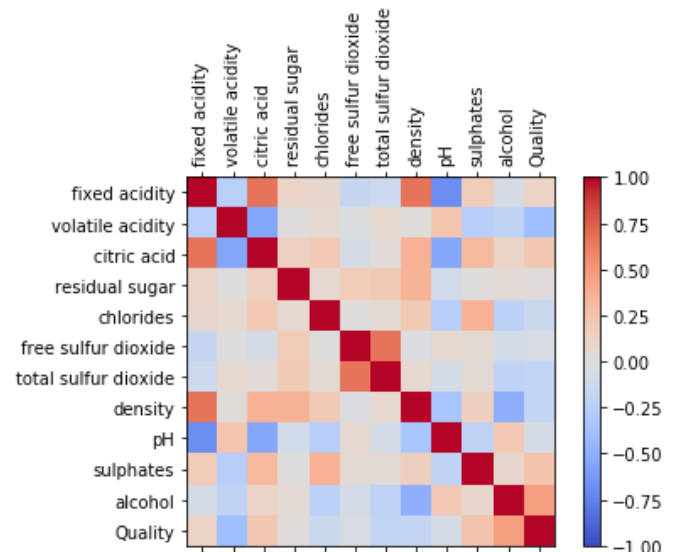


Fig-2

Correlation between the quality and each attribute

Red Wine correlation matrix (Fig-3) -

- fixed acidity: +ve with citric acid and density, -ve with pH
- volatile acidity: -ve with citric acid and quality
- citric acid: +ve with fixed acidity, sulphates, density, and quality -ve with volatile acidity and pH
- density: +ve with fixed acidity, residual sugar, and citric acid -ve with alcohol and pH
- pH: -ve with fixed acidity and citric acid
- sulphates: +ve with chlorides, citric acid, and Quality
- alcohol: +ve with quality, -ve with density



- quality: +ve alcohol, sulphates and citric acid, -ve with volatile acidity

White wine correlation matrix (Fig-4) -

- fixed acidity: +ve with citric acid and density, -ve with pH
- volatile acidity: -ve with citric acid and quality
- citric acid: +ve with fixed acidity, -ve with volatile acidity and pH
- residual sugar: +ve with sulfur dioxide and density, -ve with alcohol and pH
- chlorides: -ve with alcohol and quality
- total sulfur dioxide: +ve with free sulfur dioxide, -ve with alcohol and quality
- density: +ve with residual sugar and sulfur dioxide, -ve with alcohol and quality
- pH: -ve with fixed acidity, citric acid, and residual sugar
- alcohol: +ve with quality, -ve with density, residual sugar, sulfur dioxide, and chlorides
- quality: +ve with alcohol, -ve with density, total sulfur dioxide, volatile acidity and chlorides

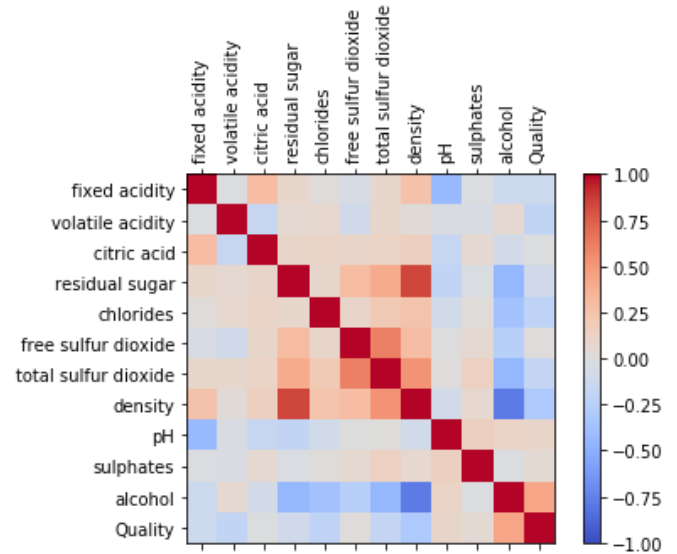


Fig-4

Final Observation-

Red Wine-

Alcohol and **sulphates** are the **main contributing factors determining the quality** of wine (Wine with higher alcohol content is better in taste). Citric acid also has a slight positive effect while volatile acidity is inversely proportional to the quality.

White Wine-

Alcohol is again the **positive** contributing factor while **density**, **total sulfur dioxide**, **volatile acidity** and **chlorides** have a significant **negative** relationship with the quality.

Prediction (Data exploitation)-

Based on the statistics of output variable we clubbed the output into **3 classes: Low [0,5), Medium [5,6] & High (6-10]**. (Difference is shown in Fig-5)

As a result, we were able to improve the performance of our Support Vector Classifier on the dataset with prediction accuracy approximately equal to 88% on the test set.

Fig-5

