**Final Project: Wine Quality Prediction**

**Course: Data Science Essentials**

Siddanth Sudarshan email: ssuda3@uis.edu

## Abstract

Quality of wine is essential in the process of brewing and selling. However, this process is time-consuming and costly to have every bottle of wine's quality assessed by humans. It is meaningful to develop a way that can predict the wine quality based on some of its chemical features. The features used in this problem are fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol. All of them are real-valued. We have 1599 red wine samples and 4898 white wine samples. The value need to be predicted is wine quality, which is evaluated from 0(worst) to 10(best). Although the values of 'quality' are integers, it is wiser to treat the learning process as a regression problem.

## Problem Statement

The problem here is to use the chemical properties of wine to predict its quality and also find the features that determine the quality of white and red wines respectively.

## Goal

1. My goal is to write a machine learning model to predict the quality of both white and red wines.
2. Secondly, to determine the features that are important in assessing the quality of wine.

## Machine Learning

For both wines, white and red, I have used Logistic Regression to train my model. Although quality is a continuous variable, I have grouped the wines based on the quality. Anything above below 5 is considered as bad and is represented by 0 and wines with quality greater than 5 are good wines and are represented by 1. I have then split the data into test and train data. To predict the accuracy of my model I have used cross validation using 5 folds.

Accuracy of models

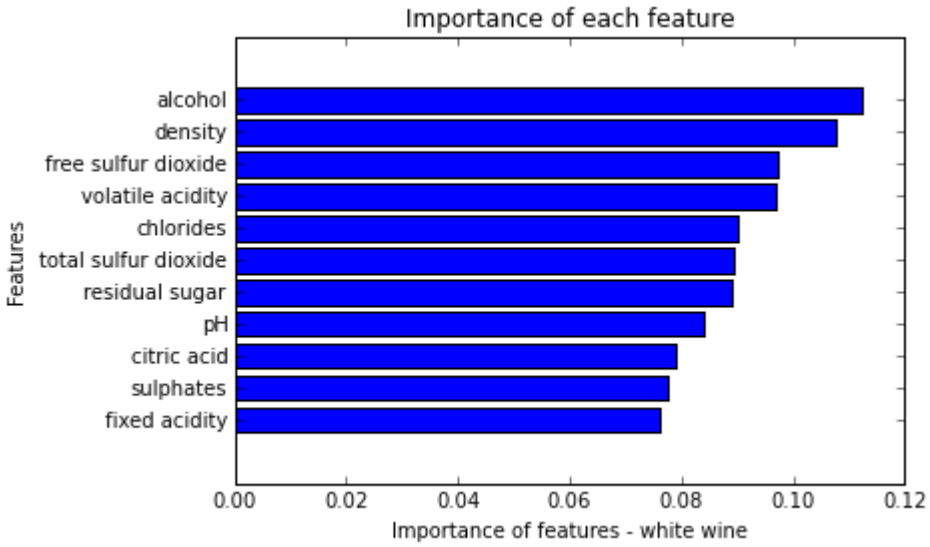
White wine – 73%

Red Wine – 72%

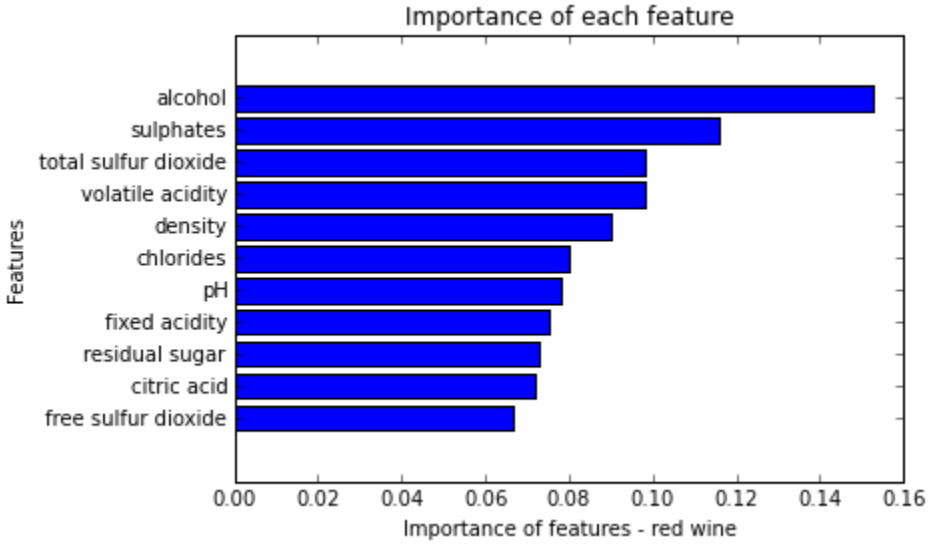
## Important features

I have used Random Forest Classifier to understand the factors that help determine the quality of wine.

For white wine –




For Red wine –



## Conclusion

Assessing the quality of wine is very important. It helps in sales and helps generate more business to the winery but human assessment is always the best. Quality of wine is subjective. What one person likes may not be liked by another. Nonetheless, I have tried to my best to predict the quality of wine using machine learning.

## Source of dataset

https://archive.ics.uci.edu/ml/datasets/Wine+Quality

Paulo Cortez, University of Minho, Guimarães, Portugal, http://www3.dsi.uminho.pt/pcortez
A. Cerdeira, F. Almeida, T. Matos and J. Reis, Viticulture Commission of the Vinho Verde Region(CVRVV), Porto,Portugal.