Machine Learning Project Plan – Red Wine Quality Prediction

Kurtis Shane Wilhoite

1 Topic Info

The purpose of this machine learning project would be to create a program with which is able to predict the quality of Red Wine using previously established tables comparing the quality to the fixed acidity, volatile acidity, citric acidity, residual sugar, chloride, free sulfur dioxide, total sulfur dioxide, density, pH, sulphate, and alcohol levels.

2 Data Info

The dataset's information can be seen at: https://archive.ics.uci.edu/ml/datasets/wine+quality

To give an overview, the dataset being used are specifically that of Red wine, though both White and Red wine are available. Having a personal preference for Red wine, I chose to do Red wine. Red and White wine have different composition and thus accuracy would be negatively affected if both datasets were combined.

The data set is using the Red variant of specifically Portuguese "Vinho Verde" wine, as a comprehensive quality and makeup list of all wines would be immense.

Input variables (based on physicochemical):
1 - fixed acidity
2 - volatile acidity
3 - citric acid
4 - residual sugar
5 - chlorides
6 - free sulfur dioxide
7 - total sulfur dioxide
8 - density
9 - pH
10 - sulphates
11 - alcohol
Output variable (based on sensory data):
12 - quality (score between 0 and 10)

3 GOAL

The goal of this project is to make a program capable of having a user input the values of a given, or imaginary/ideal, red wine's levels of the various aspects listed above. As an output, the quality of the wine should be given a numerical rating as well as a general idea of if the wine would be high or lower in quality.

It would also be good to show the data in an informative way, such as comparing the various aspects and how they correlate to the quality overall (I.e., is it better to have a high alcohol level? What about sugar content?)