

AI Algorithms: Final Project

Statement of Work

Name: Sravan Pingali

Student ID: 100809884

Rational and Problem Statement

'The older a wine is, the better its quality.' This is a statement that is commonly used for wine. While it is true there are many other factors that are responsible for the quality analysis of wines.

The wine market all over the world is around 355 billion US dollars. Global wine market has seen steady growth over the last decade and the global wine market is expected to worth around 429 billion US dollars in three years. That is a projected growth of 21%. This market isn't going decline in the coming years, if not decades and crafting quality wine is vital in maintaining this steady growth. Much like artwork, quality wines can be extremely rewarding. Wine making is a time-consuming process. Many commercially successful wine companies are very meticulous in how they make wine. This includes using the accurate measurements of ingredients for fermentation. Until a few decades ago, making quality wines would have taken excruciating amount of time, as people had to heavily depend on trial and error approaches to create new flavors. Wine companies had to make batches with different ingredient ratios to figure out the best possible combination to make good wine. This has been the traditional way in which new wine flavors were found. But now, with the advent of data analytics and AI, this process is made simpler. Using data analytics to predict the quality of wine, can be used as a preliminary way of discarding ingredient ratios that may now work. This will help in reducing the number of trail sessions required to find new flavors. Hence, predicting the quality of red wine can help save resources of a wine making company. Essentially, the computer algorithm will select which combination of the features constitute for a wine to be of good quality.

This problem can be solved using classification. The goal of this project is to predict wine quality based on certain characteristics of the wine.

Data Sources and Data Requirements

The dataset that will be used here is obtained from Kaggle. This data was initially obtained by UCI Machine Learning repository. The complete list of features is provided below.

1.	Fixed acidity
2.	Volatile acidity
3.	Citric Acid
4.	Residual Sugar
5.	Chlorides
6.	Free Sulphur dioxide
7.	Total Sulphur dioxide
8.	Density
9.	pH
10.	Sulphates
11.	Alcohol
12.	Quality

Of these, the first eleven are the features that will be used to predict the quality of wine. The data set will be divided into test and training set to check the accuracy of the method applied.

Though there may be other factors influencing the quality of wine, it is assumption here that these 11 features will be enough to get an accurate prediction for the quality of a wine. It is also assumed that this dataset is a representative of most wines.

Some of the limitations are:

- i) Incomplete data
- ii) Inconsistent data (outliers)
- iii) Duplicate data

Test Process to guarantee quality of work

Since, this is a classification problem. Accuracy will be used as the primary factor in determining the quality of the work. Accuracy will be used to before and after feature selection to check if there is an improvement or a decrement in the performance of the prediction. Training set will be used to train the classification models. After training, test sets will be used to check the success of a given model. The prediction accuracy over these test sets will determine the quality of work. Higher accuracy means good quality of work and lower accuracy means poor quality of work. Cross validation will be used to check which classification models have performed the best.