Predicting the Quality of Wine

Breonna Moore

# Domain

Wine is an alcoholic drink made from fermented grapes mainly of the Vitis vinifera species. Yeast consumes the sugar in the grapes and converts it into ethanol, carbon dioxide, and heat. Different types of grapes and strains of yeast produce different varieties of wine. Red wine is usually created using black grapes and fermented with the grape skins and seeds. However, white wines are usually made with white grapes and fermented without the grape skins.

The study that this is based on contains data that has both physiochemical data as well as sensory data on two types of Portugal wine, Red and White. Both physicochemical and sensory tests are collected on wine during wine certification and quality assessments, which are put into place to assure that illegal tampering with wine has not occurred and some standard is met by the wines being sold. This study examined wine to support the wine industry growth by identifying trends and patterns that could be used to optimize the chances of success in the wine industry.

# Motivation

I will investigate how the physiochemical data of wine relates to its sensory data. The results of this can be used by winemakers to potentially improve their decision processes and marketing strategy and thus their revenue. This can also be used by consumers of wine to know if the quality of the wine agrees with the price.

# MVP

An MVP will predict that quality of wine based on the physiochemical features found in the table below. Although wine quality is numerical and ranges from 0 – 10, which I will divide into three categories: Poor, Normal, and Excellent.

Foreseen Difficulties

This is an unbalanced dataset with more normal quality wines than excellent or poor. It is imperative that I train on wines from all categories to make the best prediction.

# Data

|  |  |  |
| --- | --- | --- |
| **Variable** | **Data Type** | **Description** |
| Quality (Dependent) | Categorical | Based on sensory data and scored 0 – 10, which will be converted to Poor, Normal, or Excellent |
| Fixed Acidity (Independent) | Numerical | Most acids involved with wine or fixed or nonvolatile, do not evaporate readily |
| Volatile Acidity (Independent) | Numerical | The amount of acetic acid in wine, which at too high of levels can lead to an unpleasant, vinegar taste |
| Citric Acid (Independent) | Numerical | Found in small quantities, citric acid can add 'freshness' and flavor to wines |
| Residual sugar (Independent) | Numerical | The amount of sugar remaining after fermentation stops, it's rare to find wines with less than 1 gram/liter and wines with greater than 45 grams/liter are considered sweet |
| Chlorides (Independent) | Numerical | The amount of salt in the wine |
| Free Sulfur Dioxide (Independent) | Numerical | The free form of SO2 exists in equilibrium between molecular SO2 (as a dissolved gas) and bisulfite ion; it prevents microbial growth and the oxidation of wine |
| Total Sulfur Dioxide (Independent) | numerical | The 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 |
| Density (Independent) | Numerical | The density of water is close to that of water depending on the percent alcohol and sugar content |
| pH (Independent) | Numerical | Measures how acidic or basic a wine is on a scale from 0 (very acidic) to 14 (very basic); most wines are between 3-4 on the pH scale |
| Sulfates (Independent) | Numerical | A wine additive which can contribute to sulfur dioxide gas (S02) levels, which acts as an antimicrobial and antioxidant |
| Alcohol (Independent) | Numerical | The percent alcohol content of the wine |

# Data Sources

1. Data Sets: <https://archive.ics.uci.edu/ml/datasets/Wine+Quality>
2. Sources: <http://winemakersacademy.com/understanding-wine-acidity/>
3. Wine Feature Descriptions: <https://github.com/jadianes/data-science-your-way/blob/master/apps/wine-quality-data-analysis/README.md>