ORIE 5741 Project: Wine Analysis

Yu Jin, yj465
Heru Wang, hw743
Jinyuan Yu, jy478

Question

From the perspective of the wine industry, how can we predict human wine taste preferences based on the physicochemical properties of wine? This is an important question for wine certification and quality assessment in this industry.

With the COVID-19 pandemic, the last year has been quite a dozy for the wine market. The industry has challenges to face and needs new adaption in strategies to be successful in meeting the new demand. Advanced technologies are expected to be utilized for both wine making and selling processes. In this project, our goal is to apply machine learning techniques to predict wine quality based on its physicochemical properties.

We believe this project will provide additional values in the following fields:

- Improving the wine production process to enhance wine quality
- Supporting oenologist wine tasting evaluation
- Stratifying wines for premium brands and pricing
- Making better marketing decisions to attract potential customers

Dataset

This dataset is related to red and white variants of the Portugal "Vinho Verde" wine, which is a unique product in Portugal and particularly appreciated due to its freshness. The data was collected from May 2004 to Feb 2007 using samples tested by the official certification entity (CVRVV), with a total of 4898 white and 1599 red samples.

Due to privacy and logistic issues, only physicochemical variables were considered as inputs, including acidity, sugar, chlorides, sulfur dioxide, density, pH, sulphates, alcohol and so on. For outputs, it uses median sensory score to measure the wine quality. Each sample was evaluated by a minimum of three sensory assessors using blind tastes, which graded the wine in a scale that ranges from 0 (bad) to 10 (excellent).

The dataset can be downloaded from the UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/