Capstone Project 1 Proposal: Wine Quality
Audrey Rager

What is the problem you want to solve?

In addition to the principle of supply and demand, the price of wine depends
opinion of the wine's quality by experts and quality assessments by physicochemical tests. Expert opinions can vary widely.

Correlating expert human quality assessment to the chemical properties of wine could be of benefit to the wine industry. Doing so would make the certification and quality assessment and assurance process more controlled and predictable.

Who is your client and why do they care about this problem?

The client is a winery owner looking for an efficient way to assess wine quality in order to set pricing and market the wine.

The client may be able to use models developed during this study to predict wine quality to price the vintage and market to appropriate distributors (Cortez et al, 2009; McMilan, R., 2020). This would make the process of pricing and marketing wine more economical (McMilan, R., 2020).

What data are you using? How will you acquire the data?

For my Capstone Project 1, I will be analyzing a Wine Quality data set from UC Irvine Machine Learning Repository. The data are related to white wine variants of Portuguese "Vinho Verde" wine (Cortez et al, 2009). The data set includes 11 physicochemical variables (fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol) and one output variable of quality (score between 0 and 1).

Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.

1. Download and, if necessary, clean the data.
2. Exploratory Data Analysis (EDA): Display and examine descriptive statistics, histograms, box-plots, scatter plots, and correlation matrices.
3. Perform Linear Regression, Support Vector Machines (SVM), KNN Classifier, Decision Tree Classifier, Random Forest Regression, and Random Forest Classifier.
4. Hyper-parameter tuning for each model.
5. Evaluate each model and make recommendations to the client.

What are your deliverables?

- GitHub including
    - Code
    - Final Report and all intermediary assignments
    - Slide Deck
    - ReadMe.html file with description and links to all components of the Capstone

REFERENCES

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

Wine Quality Dataset from UC Irvine Machine Learning Repository
https://archive.ics.uci.edu/ml/datasets/Wine+Quality

McMilan, R., 2020, State of the U.S. Wine Industry, Silicon Valley Bank Wine Division, https://www.svb.com/globalassets/library/uploadedfiles/reports/svb-2020-state-of-the-wine-industry-report-final.pdf, 71 p.

PSU, 2020, Analysis of Wine Quality Data, Applied Data Mining and Statistical Learning (STAT 508), Department of Statistics, Eberly College of Science, Pennsylvania State University, open.ed@psu, https://online.stat.psu.edu/stat508/lesson/analysis-wine-quality-data, accessed February 12, 2020.