**DDS DS 6306 Project 2: Wine Quality**

**A computer on a table with a vineyard and wine glass

Description automatically generated**

**Introduction**

Wine quality is a multifaceted attribute influenced by a myriad of factors, including chemical properties, production techniques, and environmental conditions. The goal of this data science project is twofold:

1. **Objective A**: Build a predictive model that minimizes the Mean Absolute Error (MAE) on the test set.
2. **Objective B**: Analyze the data to uncover key factors that contribute to high-quality wine, providing actionable insights for vintners to enhance their production processes.

**Objective A: Building a Predictive Model**

The primary objective is to develop a robust predictive model that accurately forecasts wine quality based on various features. We aim to achieve the lowest possible Mean Absolute Error (MAE) on the test set, ensuring high precision in our predictions.

**Mean Absolute Error (MAE)** is a commonly used metric for evaluating regression models. It measures the average magnitude of errors in a set of predictions, without considering their direction. The formula for MAE is:

A math equations and numbers

Description automatically generated with medium confidence

Minimizing MAE ensures that our model predictions are as close as possible to the actual values, providing a reliable tool for assessing wine quality.

The student with the lowest MAE on the test set will win an extra 5 points on the project and Bragging Rights! ☺

**Objective B: Mining Data for Insights**

Beyond prediction, this project aims to delve into the dataset to identify the key determinants of high-quality wine. By analyzing the relationships between various chemical properties and wine quality, we can provide valuable insights to vintners, potentially leading to improvements in wine production.

Key factors to investigate may include:

* **Acidity Levels**: Both fixed and volatile acidity could influence the taste and preservation of wine.
* **Sugar Content**: Residual sugar levels may affect the sweetness and fermentation process.
* **pH Level**: The pH of wine can impact its color, stability, and overall balance.
* **Alcohol Content**: Higher alcohol levels might correlate with perceived wine quality.

By examining these and other variables, we aim to identify patterns and correlations that distinguish high-quality wines from lower-quality counterparts. These insights can then inform production techniques, helping vintners optimize their processes to produce superior wines.

**Feature Details:**

Input variables (based on physicochemical tests):

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)

The features above include objective tests (e.g. PH values) and the output is based on sensory data (median of at least 3 evaluations made by wine experts). Each expert graded the wine quality between 0 (very bad) and 10 (very excellent).

**Dates:**

Final Project Deadline: Sunday, August 11th at 11:59pm CST

**Deliverables:**

7 min recorded presentation with the link posted on this Google Doc (assume the audience is a technical audience of vintners):

<https://docs.google.com/document/d/16McabKpyhpx0U-bUsEmJqWny-SQ0Bc3kgqG_sC5j3_E/edit?usp=sharing>

Predicted quality for wines test set. This should be in CSV format and should have two columns: ID and quality.

RShiny Dashboard that at least shows the predicted quality of a wine given certain specs.

Rmarkdown that describes your analysis.

GitHub Repo with all files that pertain to the deliverables above.