

## Group 12 Project Proposal: Wine Quality

Noah Eisenberg, Michael Gentleman, Gavin Fancher, Matteo Magcalas, Ethan Ferger

**Our Data:** Our dataset is focused on predicting the quality of wine from its chemical makeup. Our dataset contains two subsets of data: Red wine and white wine. The dataset includes 11 features, and a label which denotes the wine quality on a scale from 1-10.

<https://www.kaggle.com/datasets/xuzihe2010/wine-quality-red?select=winequality-white.csv>

### **Our questions:**

- What features are most indicative of quality (and what aren't)?
- Are those indicative features different for white and red wine respectively?
  - Differences in wine type, predicting wine type, etc.
    - The only thing with predicting wine type from the variables, is there are regulations for certain chemical makeup for red and white wine. We may want to exclude these regulated features from that specific classification case.

### **Our variables:**

- Fixed acidity (g(tartaricacid)/L)
- Volatile acidity (g(aceticacid)/L)
- Citric acid (g/L)
- Residual sugar (g/L)
- Chlorides (g(sodiumchloride)/L)
- Free sulfur dioxide (mg/L)
- Total Sulfur Dioxide (mg/L)
- Density (g/mL)
- pH
- Sulphates (g(potassiumsulphate)/L)
- Alcohol (% vol)
- Quality (score 0-10)

### **Our methods:**

- Classification (Red vs White)
  - kNN
  - Decision Trees (Random Forests)
  - Logistic Regression
  - SVM
- Regression (Quality 1-10)
  - Linear Regression
  - Logistic Regression (data splitting)
  - Gradient boosting

## Reading in Data:

```
1 import pandas as pd
2
3 red = pd.read_csv('winequality-red.csv', delimiter=';')
4 white = pd.read_csv('winequality-white.csv', delimiter=';')
5
6 red
```

✓ 0.0s

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56	9.4	5
1	7.8	0.880	0.00	2.6	0.098	25.0	67.0	0.99680	3.20	0.68	9.8	5
2	7.8	0.760	0.04	2.3	0.092	15.0	54.0	0.99700	3.26	0.65	9.8	5
3	11.2	0.280	0.56	1.9	0.075	17.0	60.0	0.99800	3.16	0.58	9.8	6
4	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56	9.4	5
...	...	...	...	...	...	...	...	...	...	...	...	...
1594	6.2	0.600	0.08	2.0	0.090	32.0	44.0	0.99490	3.45	0.58	10.5	5
1595	5.9	0.550	0.10	2.2	0.062	39.0	51.0	0.99512	3.52	0.76	11.2	6
1596	6.3	0.510	0.13	2.3	0.076	29.0	40.0	0.99574	3.42	0.75	11.0	6
1597	5.9	0.645	0.12	2.0	0.075	32.0	44.0	0.99547	3.57	0.71	10.2	5
1598	6.0	0.310	0.47	3.6	0.067	18.0	42.0	0.99549	3.39	0.66	11.0	6

1599 rows x 12 columns