Final Project Proposal

Eric Lu, Andy Luo, Ted Chen, Dali Su

11/12/2020

Project Title

Wine Master

Group Member

Andy Luo
Dali Su
Ted Chen
Eric Lu

Dataset Description

The dataset contains 6497 observations with 13 variables

```
Fixed_acidity: fixed acidity (Numeric)
Volatile_acidity: volatile acidity (Numeric)
Citric_acid: citric acid (Numeric)
Redisual_sugar: residual sugar Numeric)
Chlorides:  chlorides (Numeric)
Free_sulfur_dioxide: free sulfur dioxide(Numeric)
Total_sulfur_dioxide: total sulfur dioxide (Numeric)
Density: density (Numeric)
PH: pH (Numeric)
Sulphates: sulphates (Numeric)
Alcohol: alcohol (Numeric)
Quality: quality level of wines ranging from 1 to 10 (Categorical)
Type: Red = Red Wine or White = White Wine (Categorical)
```

Blackground

We found two datasets: the first one represents the white wine and the second represents the red wine in northern Portugal. Both datasets contain the information of thousands of wine samples, including their alcohol level and density. We combine these two datasets and add a new variable indicating their types of red or white wine.

Brief Statement

We are trying to find the best model that predicts or determines the quality of wines by estimating from the explanatory variables; some explanatory variables may be alcohol, PH level, and density.

Data Snippet

```
redwine = read.csv("winequality-red.csv")
whitewine = read.csv("winequality-white.csv")
redwine$type = "red"
whitewine$type = "white"
total <- rbind(redwine, whitewine)
head(total)
```

```
##   fixed.acidity volatile.acidity citric.acid residual.sugar chlorides
## 1           7.4             0.70        0.00            1.9     0.076
## 2           7.8             0.88        0.00            2.6     0.098
## 3           7.8             0.76        0.04            2.3     0.092
## 4          11.2             0.28        0.56            1.9     0.075
## 5           7.4             0.70        0.00            1.9     0.076
## 6           7.4             0.66        0.00            1.8     0.075
##   free.sulfur.dioxide total.sulfur.dioxide density   pH sulphates alcohol
## 1                  11                   34  0.9978 3.51      0.56     9.4
## 2                  25                   67  0.9968 3.20      0.68     9.8
## 3                  15                   54  0.9970 3.26      0.65     9.8
## 4                  17                   60  0.9980 3.16      0.58     9.8
## 5                  11                   34  0.9978 3.51      0.56     9.4
## 6                  13                   40  0.9978 3.51      0.56     9.4
##   quality type
## 1       5  red
## 2       5  red
## 3       5  red
## 4       6  red
## 5       5  red
## 6       5  red
```

Reference

https://archive.ics.uci.edu/ml/datasets/wine+quality
P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553, 2009.