**6103 FINAL PROJECT PROPOSAL**

We are Team 1 and our research topic is about White wine. We want to find out which chemical compositions affect the quality of White wine the most. This dataset was produced by Cortez et al. in Modeling wine preferences by data mining from physicochemical properties (https://www.sciencedirect.com/science/article/pii/S0167923609001377). It contains physicochemical and sensory preference data of several thousand Portuguese vinho verde white wine samples. The data was collected by CVRVV, the official certification organization of vinho verde.

SMART Questions:

Q1- Are fixed acidity, volatile acidity, residual sugar, density, citric acid, sulphates, chlorides, sulphur dioxide, alcohol and pH independent of each other?If not which combinations of those variables are independent of each other?

Q2- Which variables among fixed acidity, residual sugar, and total sulfur dioxide are most correlated with alcohol percentage?

Q3- Is residual sugars, or another chemical concentration (citric acid, Chlorides, Total sulfur dioxide, Sulphates, alcohol) the biggest predictor of subjective quality?

The paper tested three models for predicting wine quality based on physicochemical information (SVMs, multiple regression, Neural networks). We can test these models, as well as other models, such as random forests, and K-nearest neighbors.

Github link:

<https://github.com/chen19960615/DATS_6103_Final_Project_Team1>

Team members :

Suhas Buravalla

Steven Chen

Siva Gogineni