

Yi Yang, Guangyu Yang, Sohaib Hasan

Dr. Rumi Chunara

GPH 2338 Section 002

March, 20th, 2023

Project Proposal

Problem: The goal is to explore the secret equation for making a wonderful beetle of white wine and assessing the quality of wines.

Data: The data comes from Kaggle.com named Wine quality data, the temporal coverage start date is 12/31/2022, and collected by Ghassen Khaled. The dataset contains 13 variables which listed below:

fixed_acidity: scale of acidity of the wine, numeric variable, range from 3.8 to 15.9.

volatile_acidity: Scale of volatile acidity, numeric variable, range from 0.08 to 1.58.

citric_acid: Scale of citric acidity, numeric variable, range from 0.08 to 1.66.

residual_sugar: Scale of citric acidity, numeric variable, range from 0.6 to 65.8.

chlorides: Scale of chlorides, numeric variable, range from 0.009 to 0.611.

free_sulfur_dioxide: Scale of free sulfur dioxide, numeric variable, range from 1 to 289.

total_sulfur_dioxide: Scale of total sulfur dioxide, numeric variable, range from 6 to 440.

density: Scale of density, numeric variable, range from 0.98711 to 1.03898.

pH: Scale of PH, numeric variable, range from 2.72 to 4.01.

sulphates: Scale of sulphates, numeric variable, range from 0.22 to 2.00.

alcohol: Scale of alcohol, numeric variable, range from 8.0 to 14.9.

quality: Scale of quality, numeric variable, range from 3 to 9.

color: represent the color of wine, categorical variable, contains “Red” and “White”.

Approach: The model has not been decided, so we will later compare different models to find which model is the most appropriate. Our plan is to use the LOOCV approach to compute the CV error for each model (including decision tree, k nearest neighbor, linear regression model etc.) and decide which model is the best.

Evaluation: We are going to calculate the test error and BIC, and select the model from the model which has the lowest BIC, test error, and the highest adjusted R-square.

Timeline:

Date	Yi Yang	Guangyu Yang	Sohaib Hasan
3/20/2023	Finish Proposal of project	Finish Proposal of project	Finish Proposal of project
3/21/2023	Start Building Model		
4/1/2023	Finish Model Building	Start Validation	
4/10/2023		Finish Validation	Start Model Selection
4/20/2023			Finish Model Selection
4/21/2023	Start Project Report	Start Project Report	Start Project Report
5/1/2023	Finish Project Report	Finish Project Report	Finish Project Report
5/8/2023	Presentation Recording	Presentation Recording	Presentation Recording