Tyler Perry, Mehmet Ozmen, Fabio Nasseh

Dr. Zhe He

LIS 4761 Data Mining

September 19, 2021

**Team-Based Project Proposal**

**PROBLEM STATEMENT:**

We will attempt to predict the quality of different Vinho Verde red wine varieties based on some or all of the variables provided to us in the dataset *(fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulfates, alcohol).*

**DATASETS TO BE USED:** *Red Wine Quality:* [*www.kaggle.com/uciml/red-wine-quality*](https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009)*.*

This dataset contains physicochemical and sensory data on red wine variants of the Portuguese “Vinho Verde” wine.

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.

**DATA MINING SOFTWARE:**

We will use Weka to create the linear regression model and Orange3 for data cleanup and visualization. Jupyter Notebook will be useful to build/evaluate the prediction models. The reason we chose Jupyter Notebook is that it provides us with the ability to collectively write the codes and test them in real time. We can collaborate on the codes and debug our codes in real time just as we work on written assignments on Google Doc.

**TEAM MEMBER RESPONSIBILITIES:**

All members are expected to put in an appropriate amount of work and effort in contributing to the project. Since this is a small size group with only three members, each person will be responsible for one third of the work required for each week. This will be determined at the start of each week based on the workload for that week.

**TEAM MEETING SCHEDULE:**

We will utilize class meeting time and a group chat for most communications. We will schedule zoom meetings through our group chat as needed.