**MACHINE LEARNING – PROJECT PROPOSAL**

**Background**

Nowadays, red wine appears more frequently and its production process has been improved a lot. Due to this and many other reasons, the quality of red wine varies a lot as well. And this confused the wine enterprises about how to determine the price of their productions. In other words, how the quality varies based on the wine’s different features?

**Problem**

With 11 given values of different features of the wine, we predict the quality of it and then it’s easy for us to determine the price.

**Input**

The 11 features can be read from the dataset file using *pandas.read()*.

1 - fixed acidity

2 - volatile acidity

3 - citric acid

4 - residual sugar

5 - chlorides

6 - free sulfur dioxide

7 - total sulfur dioxide

8 - density

9 - pH

10 - sulphates

11 - alcohol

**Output**

The value standing for quality ranged [0, 10]

**Datasets**

I downloaded the training and testing dataset from the website [*https://archive.ics.uci.edu/ml/datasets/Wine+Quality*](https://archive.ics.uci.edu/ml/datasets/Wine+Quality)

**Project Design**

Do some data exploration and then construct the model based on the training data. Concrete steps have not been decided yet. I think finally we will get an equation like the following one:

t = aixi

where t is the target value standing for quality of the wine and xi represents the descriptive feature while ai is the parameter of xi

Process of my project can be seen on *https://github.com/PJYGit/ML-FinalProject*