**AML with Python **

**XGBoost Assignment**

We are interested in the below variables which help in deciding the quality of wine. (after opening dataset rename columns in the below sequence)

More information about the data is in this [link](https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality.names)

fixed.acidity

volatile.acidity

citric.acid

residual.sugar

chlorides

free.sulfur.dioxide

total.sulfur.dioxide

density

pH

sulphates

alcohol

quality

The response ‘**Quality’** is assumed to be a continuous variable and is predicted by the independent predictors, all of which are continuous

There are 12 predictor variables. We will treat all the variables as continuous.

1. Import and check the shape of data

2. Rename each column with the names given in this document

3. Create a heatmap

4. Use describe() to understand the distribution of data

5. take outcome attribute as target attribute

6. 7. Split the data in to train and test data in different ratios

7. Apply GradientBoostingRegressor and predict.

8. Find the Mean squared error.

9. Plot train and test error vs number of trees

10.Plot feature importance ordered by score.