**PROJECT PROPOSAL**

**Project Members** : Alekya Kumar, Monalisa Mishra, Phaneendra Ramachandraiah, Trupti Jadhav

**Project Topic :** Predicting Wine Quality for two types of wine : Red and White

**Data Set :** Wine Quality Data set. Two data sets are included in the following link.

<https://archive.ics.uci.edu/ml/datasets/Wine+Quality>

**Problem Description:** The aim of the project is to predict the quality ratings for both red and white wine and to determine which predictors have significant relationship with respect to the response variable. The data set utilised for the analysis has been obtained from UCI repository. The data set contains around 6000 data points and a brief description of the attributes are given below.

**Input variables (based on physicochemical tests):** Fixed acidity, Volatile acidity , Citric acid, Residual sugar , chlorides , Free sulfur dioxide , Total sulfur dioxide , Density, pH, Sulphates, Alcohol   
**Output variable (based on sensory data):**   
12 - quality (score between 0 and 10)

**Analysis Methods:**

1. Data Cleaning and Pre-processing the data – outlier identification and noise removal
2. Using the cleaned data set to perform Exploratory Data Analysis and understand the relationship among the variables.
3. The following approaches are employed to predict the response variable :
4. Multiple Regression
5. K-Nearest neighbor
6. Random Forests
7. Since there are two different data sets, we shall perform all three methods in both the data sets. In addition to this, we are also planning to do feature selection(Automatic and Manual).