## The first stage of this project was data pre processing. This included:

## Filling up the missing values with a global constant value .

## using the attribute selection method called "correlation attribute eval" for ranking and selecting the important attributes , which have large contribution.

## replacing the string values of the instances with simple numeric values like 0,1,2,3....

## These all tasks were achieved using weka tool.

## Ranking of attributes:

## 

## We only considered those attributes whose correlation value is greater than or equal to 0.06 .

## snapshot of initial dataset:

## 

## snapshot of dataset after preprocessing:

## 

## There are approximately 45200 instances in our dataset. We will first shuffle the dataset randomly . After which, the 31640 (70%) instances will be used for training purpose and the rest will be used for testing.

## Visualization :

## Given below are the graphs of all the attributes :

## 

## attribute "default" (least ranked) :

## 

**Output after analysis (using python) :**

We will the decision tree , accuracy , confusion matrix and duration of each algorithm as an output.







