**Topic: Prediction of credit revolving balance**

**Introduction**

According to the business moment given the problem is to predict the revolving balance using the remaining observations in the dataset. In which the dataset contains of 36 variables and 887379 observations. Project is executed using R Language and deployed using R shiny.

**Structure of the project**

**EDA**

The dataset is imported in .CSV format and be executed in R. Firstly, to clean the data NA values is been calculated and removed the variables which has more than 50% of NA values. For the remaining variables which still have the NA values is been filled using Simple Imputer (mean, median). Variables which are in the character type is been converted into numeric and data is been standardized.

**Model building**

X and y are considering that x being the input and y being the output which is total revolving balance and remaining variables are consider as input. Data is been split into train and test; this helps the model to train the model using data first then test the remaining data.

Here Linear regression, Decision Tree, Randomforest models are been built.

**Evaluation**

We have checked the performance of the various regression/classification algorithms (based on the problem).

For regression algorithms, some of the common metrics are — MSE and R Square.

Evaluation metrics pertaining to classification are— Confusion Matrix, AUC/ROC curves etc. We have compared these scores for each algorithm to check which ones performed better than the rest.

**Result**

After different comparisons of models we came up with the Least RMSE value in linear Regression model. We got model results as Train=19391.02 and for Test=19423.51