**Building a Credit Risk Analyser**

This project tries to analyze the data and various features present in the data.

And then based on the analysis, a classification model is build. The code and

the report are merged together into a single document, so as to ensure a clear

understanding to whomsoever reading this document.

1. **Objective** : This project aims at building a classification model to predict whether a person is going to be a defaulter or not based on various parameters like age, education and income etc. This helps the companies to perform credit risk analysis i.e possibility of the borrower's repayment failure and the loss caused to the financer when the borrower does not for any reason repay the contractual loan obligations.
2. **Data and Variables**: age, gender, education, occupation, organization\_type, seniority, annual\_income, disposable\_income, house\_type, vehicle\_type, marital\_status, number\_of\_cards\_holding, Defaulter. Here, Defaulter is the response variable. The data is structured.
3. **Methods and algorithm used**: Here we use Decision Tree algorithm to predict whether a future person applying for a credit card is likely to be a defaulter or not based on the given data.
4. **Partitioning of the data**: We partition the given dataset into two sets: train dataset (for training our model) and test dataset (for prediction and checking the accuracy of our model).
5. **Libraries used**: pandas, numpy, matplotlib, seaborn, sklearn,.
6. **Development Process**: We perform univariate analysis of the variables Age of the persons, Gender, Education, Occupation, Organization type, Seniority, Annual income, Disposable income, House type, Vehicle type, Marital status, Number of cards. We then perform few bivariate analyses to check the correlation between some variables. For the categorical variables, we create dummy variables to convert them into numerical type for use in our model. We then perform cross-validation. We then calculate the accuracy of our model.

Train accuracy is: 85.92962798758616

Test accuracy is: 85.84688302284246

CV accuracy is: 85.63099492194847