

## **Credit Score Classification, Cleaning, Exploratory data analysis and modeling.**

### **Abstract**

Most Financial firms has been collecting customers basic banking information over decades, of which a lot of credit-related information has been part of these collection. Now, these firms want to build an intelligent system to segregate the people into credit score brackets to reduce the manual efforts, and we intend to apply Random Forests technique to achieve this goal.

### **Introduction**

This study is using a banking dataset posted on Kaggle[1], and it has a train and a test dataset. It contains 100,000 customers, with 27 different customer information (columns). In general, the data set contains numerical, binary information, and categorical variables, and this information includes:

Customer ID, Month, Name, Age, SSN(social security number), Occupation, Annual\_Income, Monthly\_Inhand\_Salary, Num\_Bank\_Accounts, Num\_Credit\_Card, Interest\_Rate, Num\_of\_Loan, Type\_of\_Loan, Delay\_from\_due\_date(average number of days delayed from the payment date), Num\_of\_Delayed\_Payment(average number of payments delayed by a person), Changed\_Credit\_Limit(percentage change in credit card limit), Num\_Credit\_Inquiries, Credit\_Mix(classification of the mix of credits), Outstanding\_Debt, Credit\_Utilization\_Ratio, Credit\_History\_Age:(age of credit history of the person), Payment\_of\_Min\_Amount, Total\_EMI\_per\_month(total Equated Monthly Installments payments (in USD)), Amount\_invested\_monthly, Payment\_Behaviour, Monthly\_Balance and Credit\_Score(Poor, Standard, Good). According to [2] Financial institutions may consider your credit report and credit score to decide weather a loan should be given to an individual or not. Also, it is used to determine the amount of interest they will charge on these loans. So, we will be using this information to build a machine learning model that can classify the credit score of customers. Also, we noticed that this data was last updated June 2022. We intend to first do some data cleaning and exploratory data analysis with the hope of finding some insights.

### **Reference**

[1] <https://www.kaggle.com/datasets/parisrohan/credit-score-classification>

[2]<https://www.canada.ca/en/financial-consumer-agency/services/credit-reports-score/credit-report-score-basics.html>