Title: Credit Card default Prediction

# **Bhargava Gaggainpali DSC680 - Applied Data Science** [bgaggainpali/bgaggainpali\_DSC680 (github.com)](https://github.com/bgaggainpali/bgaggainpali_DSC680)

Portfolio: [Bhargava Gaggainpali - Portfolio | Bhargava-Gaggainpali (bgaggainpali.github.io)](https://bgaggainpali.github.io/Bhargava-Gaggainpali/)

https://bgaggainpali.github.io/Bhargava-Gaggainpali/

# Which Domain?

In financial industry, banks are playing important role in challenging times like now, with COVID pandemic across the globe. People are losing jobs and financial institutions are facing more delinquency rate on credit card loans. The increase in delinquency rate will result in significant financial loss to commercial banks. It is very critical for lending institutions like banks to have a prediction model to be able to predict customers for credit card default.

# Which Data?

I have identified UCI\_Credit\_Card.csv as source for my work, below is the Kaggle link. There are 30,000 observations in the dataset, each row in the dataset represents a credit card client. Given is the list of variables in the dataset.

Source File: <https://www.kaggle.com/ainslie/credit-card-default-prediction-analysis>

**Variable Description**

ID Credit Card ID - Sequence Number

LIMIT\_BAL Credit Limit

SEX 1 = male, 2 = female

EDUCATION 1 = graduate school, 2 = university, 3 = high school

MARRIAGE 1 = married, 2 = single, 3 = others

AGE Customer Age

PAY\_0 Repayment status September 2005

PAY\_2 Repayment status August 2005

PAY\_3 Repayment status July 2005

PAY\_4 Repayment status June 2005

PAY\_5 Repayment status May 2005

PAY\_6 Repayment status April 2005

BILL\_AMT1 Bill Amount September 2005

BILL\_AMT2 Bill Amount August 2005

BILL\_AMT3 Bill Amount July 2005

BILL\_AMT4 Bill Amount June 2005

BILL\_AMT5 Bill Amount May 2005

BILL\_AMT6 Bill Amount April 2005

PAY\_AMT1 Payment Amount September 2005

PAY\_AMT2 Payment Amount August 2005

PAY\_AMT3 Payment Amount July 2005

PAY\_AMT4 Payment Amount June 2005

PAY\_AMT5 Payment Amount May 2005

PAY\_AMT6 Payment Amount April 2005

default.payment.next.month 1 = default, 0 = On time payment

# Research Questions? Benefits? Why analyze these data?

With the current pandemic Covid situation lots of people are falling as default for credit card monthly payments and it would impact financial budgeting for banks as it has limited assets which can handle the defaults until certain extent only, so finding out the customers who are possibly to fall in as credit card default will help banks to mitigate and balance the risk of the defaults. In this paper I will try to predict to find, who are possibly to fall in as credit card default and identify the customers to reduce this risk factor of bank efforts and assets.

# What Method?

I see this problem as a classification issue, where we should try to understand and able to predict the customers, who have high Credit Card default chances.

Planning to use supervised machine learning algorithm to work on the classification problem to be trained with algorithms like:

1. Logistic Regression

2. Decision Tree

3. Random Forest

Start with loading data into a data frame and then understand the data, then perform Exploratory Data Analysis (EDA) on the data set. EDA involves doing Univariate and Bivariate Analysis, identify missing values and outliers and fill the gaps with appropriate values. In the next step, building model with starting from logistic regression and observe the accuracy of the model. When the accuracy of the of the model is not high, then planning to use Decision Tree and Random Forest to achieve higher accuracy.

# Potential Issues?

Due to the low volume of the data that is available, I would assume model accuracy would be low. I am planning to use multiple algorithms, at least one would give expected accuracy in the range of 80-85. Also, I might encounter some data type errors while working with variables.

# Concluding Remarks

By taking a sample data set, with set of features, I would like to explore the features to find the pattern which majorly influence the ‘Credit Card default Prediction’ process and the order of importance. For this to achieve, will try to clean, validate, perform exploratory data analysis and also use the machine learning algorithm to build the predicting model. Based on the data and model efficiency, I would assume the accuracy of the model will be between 80%-85%.

**Reference:**

LATOYA IRBY - February 10, 2020 - What You Can Do About Credit Card Default

[What You Can Do About Credit Card Default (thebalance.com)](https://www.thebalance.com/what-is-credit-card-default-960209)

Jenny Wang - Jun 24, 2020 - Will You Be Able to Make Your Credit Card Payment?

[Will You Be Able to Make Your Credit Card Payment? | by Jenny Wang | Towards Data Science](https://towardsdatascience.com/catch-me-if-you-can-predicting-credit-card-default-using-classification-analysis-28b2955f7f7d)

Marcos Dominguez - Feb 26,2021- Predicting Credit Card Defaults with Machine Learning

[Predicting Credit Card Defaults with Machine Learning | by Marcos Dominguez | The Startup | Feb, 2021 | Medium](https://medium.com/swlh/predicting-credit-card-defaults-with-machine-learning-fcc8da2fdafb)

Yashna Sayjadah, Ibrahim Abaker Targio Hashem, Faiz Alotaibi, Khairl Azhar Kasmiran - October 2018 - Credit Card Default Prediction using Machine Learning Techniques

[(PDF) Credit Card Default Prediction using Machine Learning Techniques (researchgate.net)](https://www.researchgate.net/publication/334765725_Credit_Card_Default_Prediction_using_Machine_Learning_Techniques)

Bank Rate – 2021 - Credit card default: How it happens, what to do about it

[Credit Card Default: What to Do About It | Bankrate.com](https://www.bankrate.com/finance/credit-cards/credit-card-default/)

Equifax – 2021 - What Happens If I Default on a Loan or Credit Card Debt?

[Process & Potential Effects of Defaulting on a Loan | Equifax](https://www.equifax.com/personal/education/covid-19/default-loan-credit-card-debt/)