

GitHub Link: <https://github.com/Nat-Han1999/CS3244-CreditCard.git>

# Project Title: To approve or not to approve: Credit Cards?

## Project Description

The amount of credit card applications a bank has to handle every day is enormous. Not all applications are approved, however, with hypothesised reasons like high-loan balances, low-income levels, or poor credit track records. There are also those applications that were approved, but some borrowers have gone into credit default or displayed undesirable credit behaviour, and to the banks, that's a loss on the books. Default risk is a very real problem for banks and financial institutions.

We think that it is important to understand what attributes are important and indicative of a borrower's undesirable credit behaviour. We believe the lessons learnt may help to better inform decisions made by banks in approving credit card applications, controlling credit risk. On the consumer side, we may also get a better understanding of our application outcome since banks are unlikely to reveal rejection reasons directly.

This project aims classify an applicant as a potential good or bad borrower based on their personal information.

## Table of Contents

1. [Dataset](#)
2. [Models](#)
3. [Techniques](#)
4. [Contributors](#)

## Dataset

<https://www.kaggle.com/datasets/rikdifos/credit-card-approval-prediction>

The dataset is supposedly real-world bank data hosted on Kaggle, sourced from a machine learning course. It includes 2 separate CSV (Comma-separated Values) files. The first file, credit\_record.csv is the credit repayment records of customers approved for a credit card, recorded by an ID. The other file, application\_record.csv, is personal

information about the customers, and consists of attributes like income level, education level, employment information among many others. We will have to define what is deemed as "good"/"bad" customers as labels are not provided.

## Models

The following models will be applied in this project:

- Supervised
  - Logistic Regression
  - Decision Tree
  - Random Forest
  - XGBoost
  - AdaBoost
- Unsupervised
  - K-Means Clustering

## Techniques

- Feature Extraction
  - Feature Engineering
- Splitting of Data into Test/Train
  - Stratified Split
- Handling Class Imbalance
  - Upsampling and Downsampling
  - SMOTENC
- Hyperparameter Tuning
  - RandomizedSearchCV
  - GridSearchCV

## Contributors

- Bokai: <https://github.com/heimianbb>
- Jia Yee: <https://github.com/jyo2022>
- Junjing: <https://github.com/Radix-huang>
- Nathaniel: <https://github.com/Nat-Han1999>
- Sandy: <https://github.com/sandyseahjs>
- YiChi: <https://github.com/CDDnoICE>