GEORGE WASHINGTON UNIVERSITY

Introduction to Data Science (DATS 6101)

Team 4

GitHub repo: <https://github.com/daqianshijie2000/We-R-4-3->

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

**Final Project Proposal**

Title: Credit Card Approval Prediction Based on the Applicants Lifestyle

RESEARCH TOPIC:

Credit scores are a common method to tackle risk in the financial sector. It uses personal information and any data submitted by credit card users/applicants to predict the probability of future defaults and credit card borrowings.

It allows the bank to decide whether to issue a credit card to the applicant or simply decline.

Logistic regression modelling is a common method for credit scoring as it is suitable for binary classification tasks.

SMART QUESTIONS:

Overall

* What factors or variables affect or increase the likelihood of a person being issued a credit card?

EDA

* Do variables like income category and education level affect a person being issued a credit card?
* Does marriage status increase the likelihood of being issued a credit card?

METHODS:

The target risk variable in the dataset is chosen as users who are overdue for more than 60 days. These users are marked as 1, while the rest are marked as 0. If a user is a target risk, then they are not issued a credit card.

EDA will consist of plotting graphs using R libraries to find any relation between possibility of being a target risk with other variables.

Since most of the variables in the data set are categorical variables, our focus for modeling is to use logistic regression. This also helps as the target variable is binary. Other classification algorithms such as Classification trees and SVM will also be used to train the model.

DESCRIPTION OF AVAILABLE DATA:

The source of the dataset is Kaggle Sample Dataset where it was extracted in a CSV

format. The data consists of 2 CSV files: application\_record.csv and credit\_record.csv

There are two tables which are connected by the column ‘ID’.

Application\_record.csv contains an applicant’s personal information and other socio-economic details which can be helpful for building a predictive model.

Credit\_record.csv contains a user’s recorded behavior of the credit card, capturing any defaults and the balance due.

------------------------------------------------------------------------------------------------------------------------------------------