## horizontal line



Credit Card Fraud Detection

02.07.2021

**─**

Olawin, Ridwan

Wu, William

Ta, Quan

Lee, Hajun

Drexel University

# Introduction

Credit cards are widely used in our time. Nowadays, people look at your credit to lend you money. Banks, companies, and lenders look at your credit information in order to assess the amount of money that they can lend. However, even if one’s credit looks fine at the beginning, one can become a credit card defaulter in a minute, making the lenders lose their profit. Our application seeks to find out what interesting rules lie in people who borrow money from others and become credit card defaulters. This application seeks to allow lenders to have an almost accurate evaluation of the person who is borrowing money and assume whether lending to such individuals would be good or not. This application would be extremely useful especially if you are a prospective lender.

# Data Description

One important note is that the current data set has 122 variables. The variables that we have below are part of an abridged version of the data set that was found on Kaggle(<https://www.kaggle.com/mishra5001/credit-card?select=application_data.csv>).

Descriptive Statistics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Observations** | **Mean** | **Std. Dev** | **Min** | **Max** |
| SK\_ID\_PREV | 1,670,214.00 | 1,923,089.00 | 532,598.00 | 1,000,001.00 | 2,845,382.00 |
| SK\_ID\_CURR | 1,670,214.00 | 278,357.20 | 102,814.80 | 100,001.00 | 456,255.00 |
| NAME\_CONTRACT\_TYPE | 0.00 | - | - | - | - |
| AMT\_ANNUITY | 1,297,979.00 | 15,955.12 | 12,782.14 | 0.00 | 418,058.20 |
| AMT\_APPLICATION | 1,670,214.00 | 175,233.90 | 292,779.80 | 0.00 | 6,905,160.00 |
| AMT\_CREDIT | 1,670,213.00 | 196,114.00 | 318,574.60 | 0.00 | 6,905,160.00 |
| AMT\_DOWN\_PAYMENT | 774,370.00 | 6,697.40 | 20,921.50 | -0.90 | 3,060,045.00 |
| AMT\_GOODS\_PRICE | 1,284,699.00 | 227,847.30 | 315,396.60 | 0.00 | 6,905,160.00 |
| WEEKDAY\_APPR\_PROCESS\_START | 0.00 | - | - | - | - |
| HOUR\_APPR\_PROCESS\_START | 1,670,214.00 | 12.48 | 3.33 | 0.00 | ` |
| FLAG\_LAST\_APPL\_PER\_CONTRACT | 0.00 | - | - | - | - |
| NFLAG\_LAST\_APPL\_IN\_DAY | 1,670,214.00 | 1.00 | 0.06 | 0.00 | 1.00 |
| RATE\_DOWN\_PAYMENT | 774,370.00 | 0.08 | 0.11 | 0.00 | 1.00 |
| RATE\_INTEREST\_PRIMARY | 5,951.00 | 0.19 | 0.09 | 0.03 | 1.00 |
| RATE\_INTEREST\_PRIVILEGED | 5,951.00 | 0.77 | 0.10 | 0.37 | 1.00 |
| NAME\_CASH\_LOAN\_PURPOSE | 0.00 | - | - | - | - |
| NAME\_CONTRACT\_STATUS | 0.00 | - | - | - | - |
| DAYS\_DECISION | 1,670,214.00 | -880.68 | 779.10 | -2,922.00 | -1.00 |
| NAME\_PAYMENT\_TYPE | 0.00 | - | - | - | - |
| CODE\_REJECT\_REASON | 0.00 | - | - | - | - |
| NAME\_TYPE\_SUITE | 0.00 | - | - | - | - |
| NAME\_CLIENT\_TYPE | 0.00 | - | - | - | - |
| NAME\_GOODS\_CATEGORY | 0.00 | - | - | - | - |
| NAME\_PORTFOLIO | 0.00 | - | - | - | - |
| NAME\_PRODUCT\_TYPE | 0.00 | - | - | - | - |

Variable Characteristics

|  |  |
| --- | --- |
| **Variable** | **Description** |
| SK\_ID\_PREV | ID of previous credit in Home credit related to loan in our sample. |
| SK\_ID\_CURR | ID of loan in our sample |
| NAME\_CONTRACT\_TYPE | Contract product type (Cash loan, consumer loan [POS] ,...) of previous application |
| AMT\_ANNUITY | Annuity of previous application |
| AMT\_APPLICATION | For how much credit did client ask on the previous application |
| AMT\_CREDIT | Final credit amount on the previous application. |
| AMT\_DOWN\_PAYMENT | Down payment on the previous application |
| AMT\_GOODS\_PRICE | Goods price that client asked for on the previous application |
| WEEKDAY\_APPR\_PROCESS\_START | On which day of the week did the client apply for previous application |
| HOUR\_APPR\_PROCESS\_START | The day and hour the client applied for the previous application |
| FLAG\_LAST\_APPL\_PER\_CONTRACT | Flag if it was the last application for the previous contract. |
| NFLAG\_LAST\_APPL\_IN\_DAY | Flag if the application was the last application per day of the client. |
| NFLAG\_MICRO\_CASH | Flag Microfinance loan |
| RATE\_DOWN\_PAYMENT | Down payment rate normalized on previous credit |
| RATE\_INTEREST\_PRIMARY | Interest rate normalized on previous credit |
| RATE\_INTEREST\_PRIVILEGED | Interest rate normalized on previous credit |
| NAME\_CASH\_LOAN\_PURPOSE | Purpose of the cash loan |
| NAME\_CONTRACT\_STATUS | Contract status (approved, cancelled, ...) of previous application |
| DAYS\_DECISION | When was the decision about previous application made |
| NAME\_PAYMENT\_TYPE | Payment method that client chose to pay for the previous application |
| CODE\_REJECT\_REASON | Why was the previous application rejected |
| NAME\_TYPE\_SUITE | Who accompanied client when applying for the previous application |
| NAME\_CLIENT\_TYPE | Was the client an old or new client when applying for the previous application |
| NAME\_GOODS\_CATEGORY | What kind of goods did the client apply for in the previous application |
| NAME\_PORTFOLIO | Was the previous application for CASH, POS, CAR, … |
| NAME\_PRODUCT\_TYPE | Was the previous application x-sell o walk-in |

# Preliminary Plan

## Problem formulation

Credit card companies on a daily basis receive many different applications for a loan. It is therefore very important for these companies to be able to identify the risk associated with sanctioning out a loan to a customer. These would enable the company to avoid risk and save a lot of money.

## Possible techniques/model

Given the nature of the data set, the technique intended to be used in solving this problem is classification based learning. The reason is because the given data set contains frequent patterns that have been established with a lender. Hence, it would be possible to establish the correlation between classifying them whether they are low, medium or high risk.

# Timeline

|  |  |
| --- | --- |
| Time | Goal |
| By the end of week 5 | * Preprocess and prepare the dataset * Determine the model we are going to use for classification on training set |
| By the end of week 7 | * Construct the model to use on the training and test sets * Calculate the accuracy of our model |
| By the end of week 9 | * Improve the accuracy of our model to meet our goal |
| By the end of week 10 | * Finish our report |