



HOME  
CREDIT

# HOME CREDIT SCORE CARD MODEL

By Ike Devanti 

## CONTENTS

01 | Problem Research

02 | Data Preprocessing

03 | Business Insight

04 | Machine Learning  
Model

05 | Business  
Recommendation

# C Problem Research

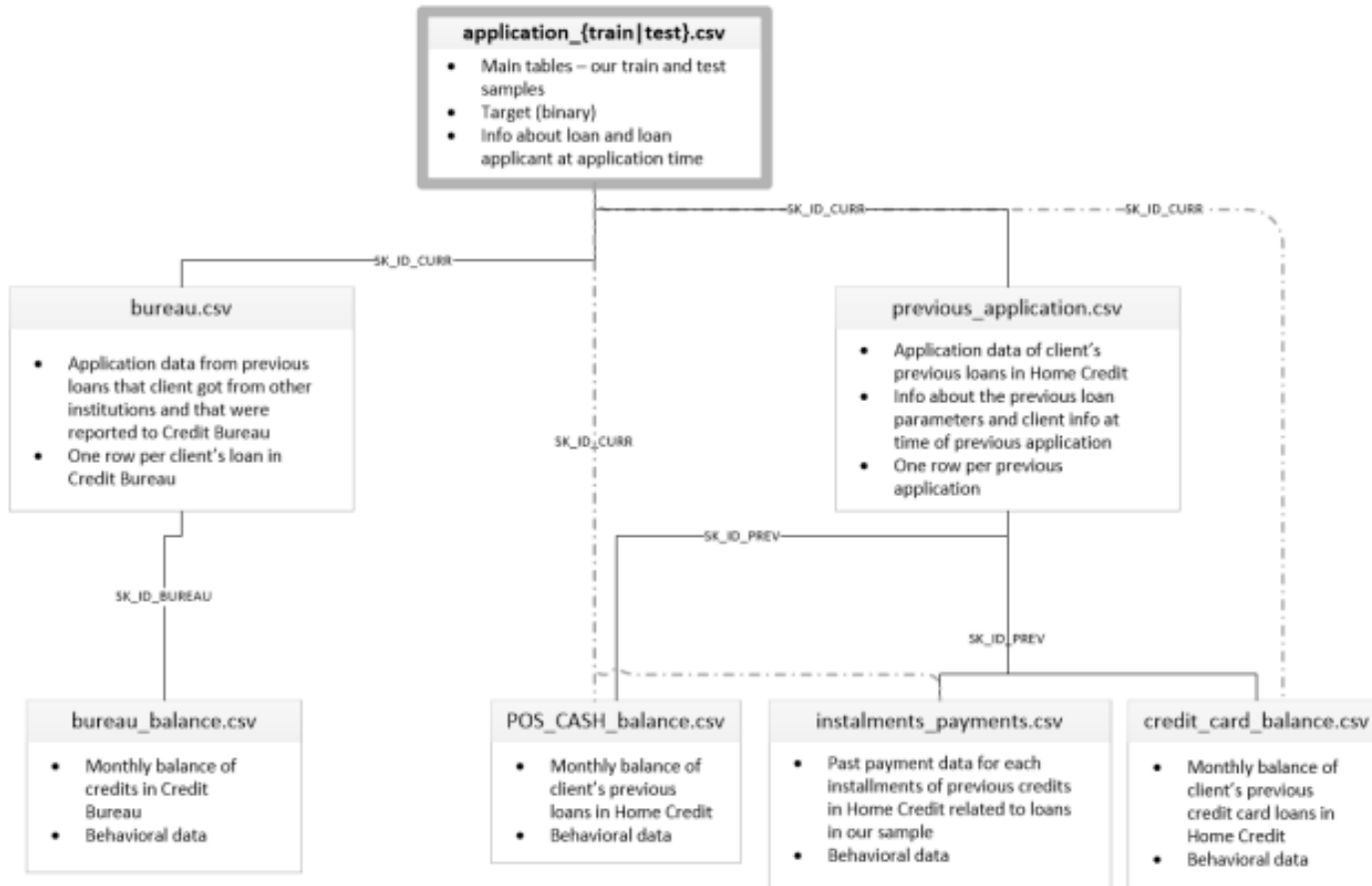
Many people struggle to get loans due to insufficient or non-existent credit histories. Home Credit strives to broaden financial inclusion for the unbanked population by providing a positive and safe borrowing experience. In order to make sure this underserved population has a positive loan experience, Home Credit makes use of a variety of alternative data to **predict their clients' repayment abilities**. Doing so will **ensure that clients capable of repayment are not rejected** and that loans are given with a principal, maturity, and repayment calendar that will empower their clients to be successful.

## OBJECTIVE :

1. Identify characteristics of potential clients who will have difficulty repaying loans and who will not.
2. Predict client's repayment abilities.

## ACTION :

1. Perform data cleaning, and
2. visualization for business insights.
3. Build a models with machine learning algorithms.
4. Provide recommendations for company to increase their clients succeed in applying for loans.



- The data used are application train and application test. There are our main table, broken into two files for train (with TARGET) and test (without TARGET).
- The Main dataset consist of 122 columns and 307511 rows. With 1 column target feature ["TARGET"]

# Q Data Preprocessing

## EDA

Discover patterns, and the structure of the dataset

- Bivariate Visualization
- Multivariate Visualization

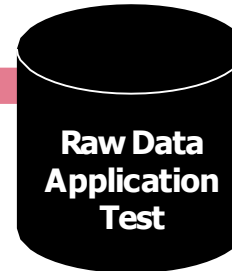
## Data Cleaning

Discover patterns, and the structure of the dataset

- Detecting Duplication
- Handling Missing Values
- Detecting Outliers

## Model Building

- Label Encoding
- Feature Selection
- Handling Imbalanced Data
- Model Building
- Model Evaluation



## Prediction

Predict client repayment abilities with best machine learning model

## Data Cleaning

- Detecting Duplication
  - ✓ No Duplicate rows
- Handling Missing Values
- Label encoding

### • Data Application Train

- ❖ 122 Numbers of Columns
- ❖ 307.511 Numbers of Rows

### • Data Application Test

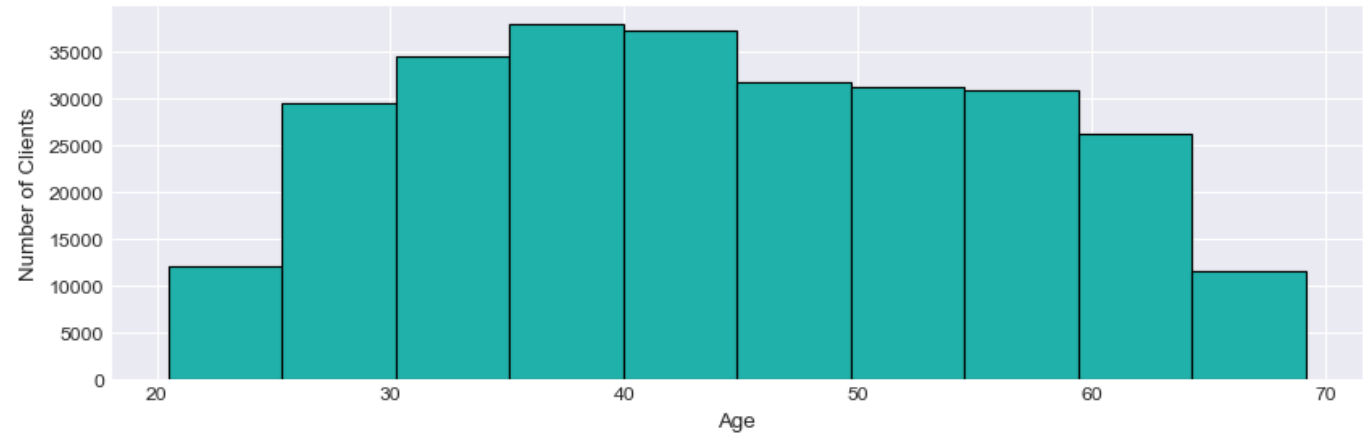
- ❖ 121 Numbers of Columns
- ❖ 48.744 Numbers of Rows

# Q Business Insight

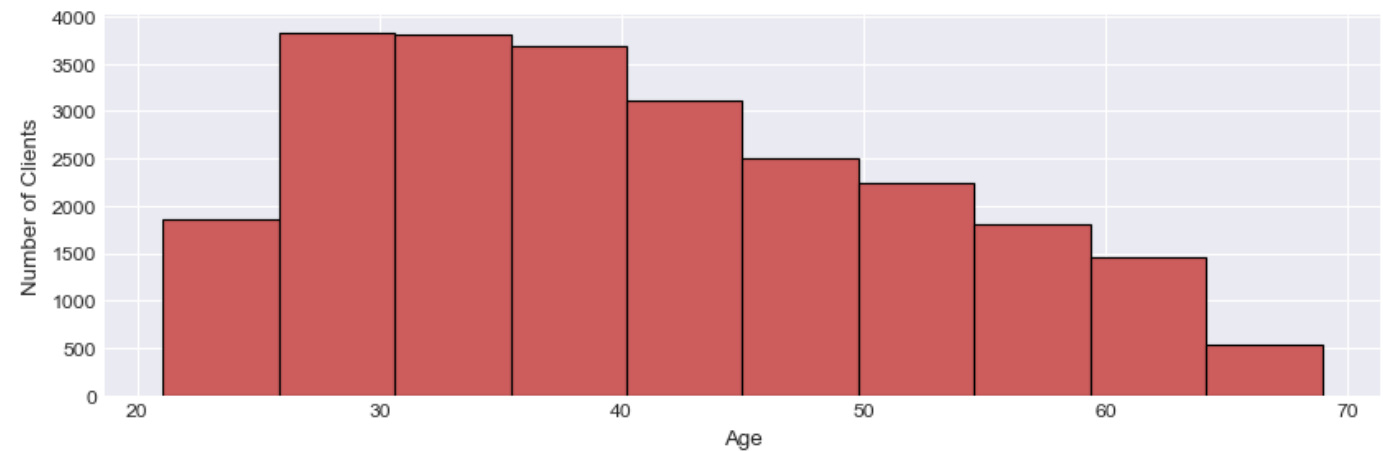
- Most number of clients who apply for loans are in the range of 35-40 years.
- Meanwhile, the number of applicants for clients aged <25 or age >65 is very low.

- Clients who have no payment difficulties are clients in the range of 35-45 years. You can target these clients as your priority.
- While clients who have payment difficulties are client the range of 25-35 years.

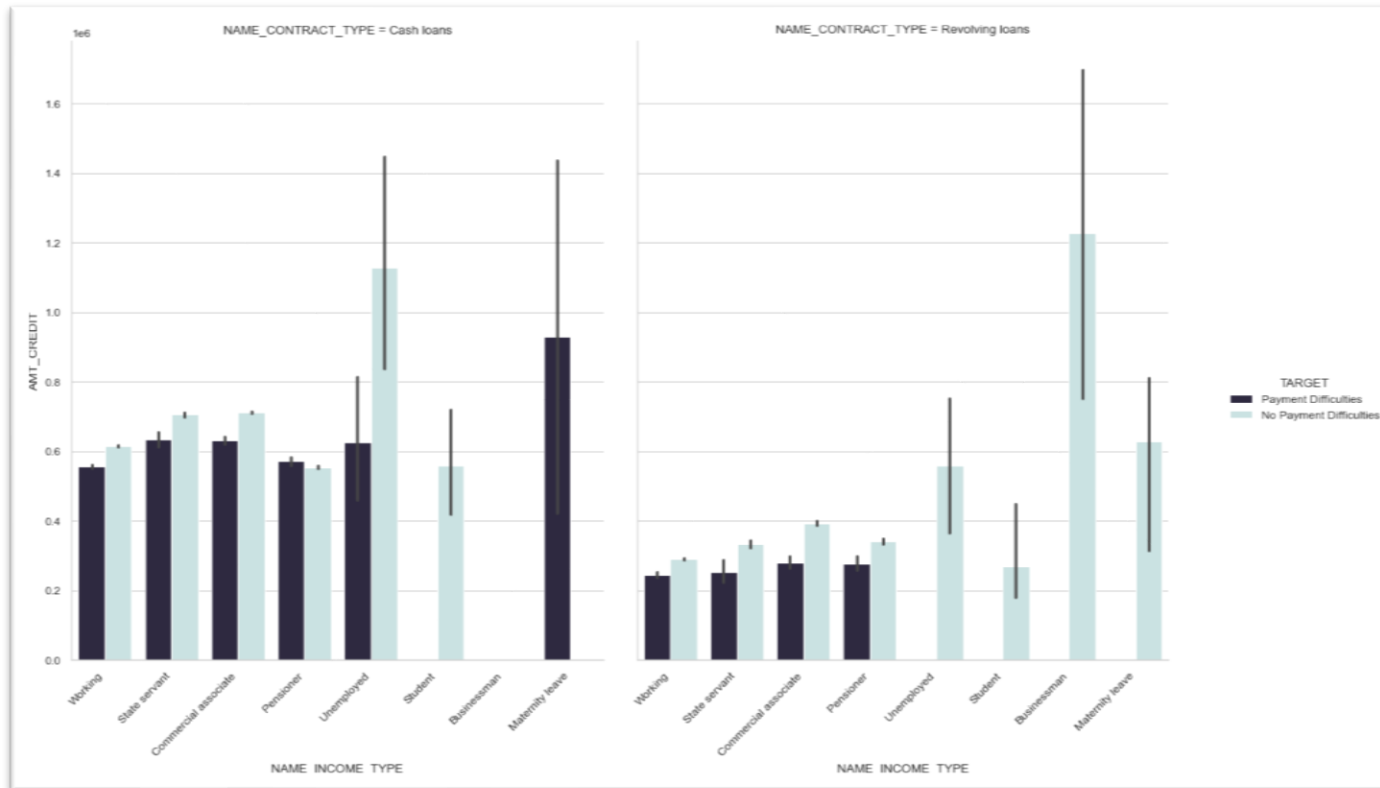
Age of Client (in years) who have No Payment Difficulties



Age of Client (in years) who have Payment Difficulties



# Q Business Insight



All student clients have no difficulty repaying the loans whether with cash loan or revolving loan for a low to medium credit amount of the loan.

- ❖ All student clients have no difficulty repaying the loans whether with cash loan or revolving loan for a low to medium credit amount of the loan. For the income type of maternity leave with cash loans, all the clients have problems repaying the loans for a medium credit amount of the loan. While all clients with maternity leaves and revolving loans have no difficulty repaying the loans.
- ❖ For unemployed clients with cash loans, more than 50% of clients have problems repaying loans with medium credit amounts of the loan. While all unemployed clients with revolving loans have no difficulty repaying the loan.

# Machine Learning Model

## Model Comparison

Algorithm	Training Accuracy Score	Testing Accuracy Score	Error Margin
Logistic Regression	67.16%	67.29%	0.13%
Gaussian Naive Bayes	60.24%	60.39%	0.15%
Decision Tree	100%	83.9%	11.74%
Random Forest	100%	99.65%	0.35%
K-Nearest Neighbor	91.56%	88.07%	3.79%
Neural Network	70.01%	69.48%	0.58%

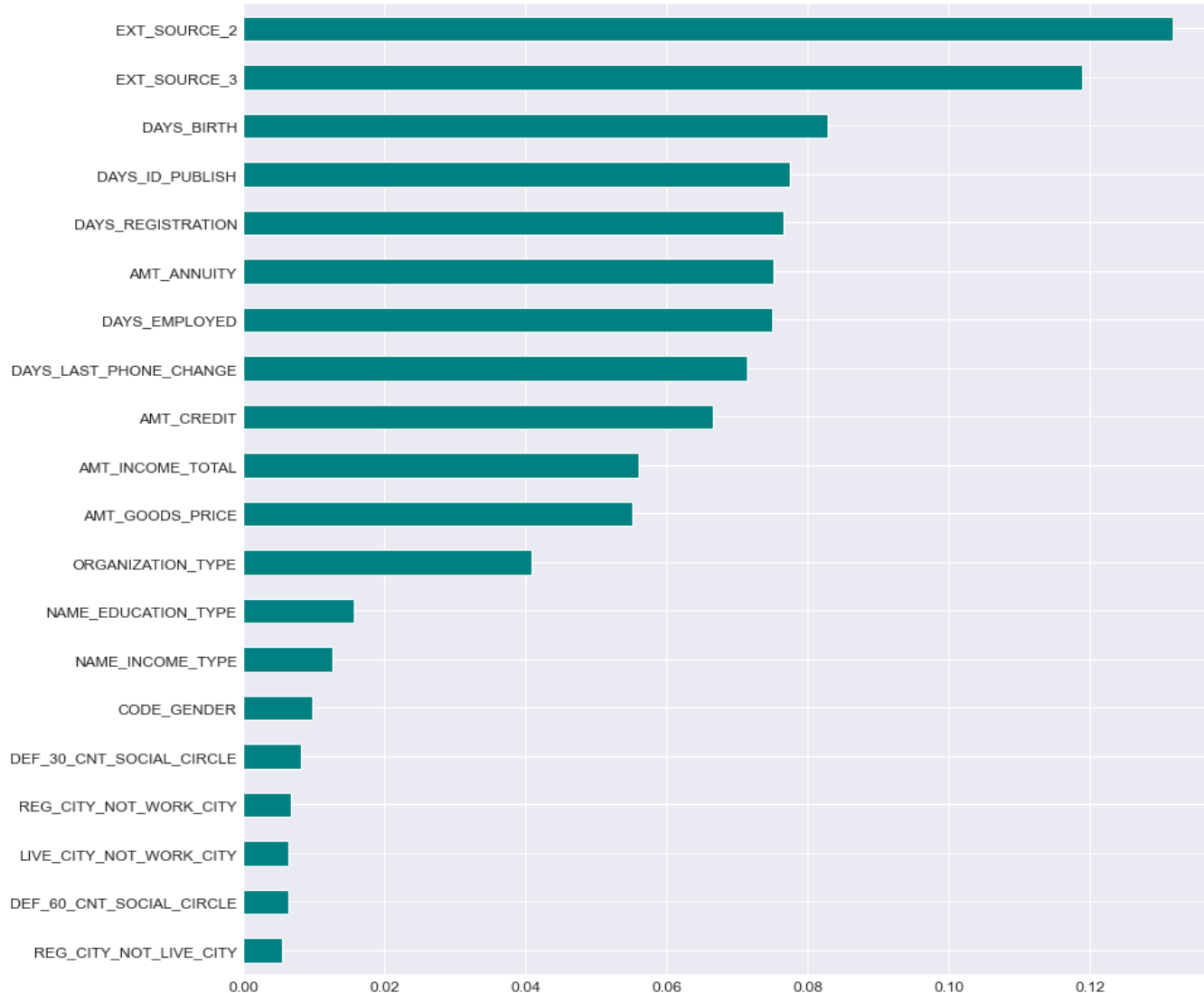
The prediction accuracy of the train and test data in Random Forest model has a value that is not much different, it can be said that the model is very good, which is there is no underfitting or overfitting. So the Random Forest model was chosen as the best model to predict client's repayment abilities.



# Machine Learning Model

Best Model

Features Importance Plot



1

## Algorithm

Random Forest Classifier

2

## Performance

- Random Forest model gives 100% correct result
- There is 0.35% error margin

3

## The 5 most important features

- Score from external data source 2
- Score from external data source 3
- Client's age in days
- Days ID publish
- Days registration

# Business Recommendation

- A client with an income type of student can be said to be a client who is capable of repaying the loans whether with a cash loan or revolving loan (100% of applications approved). But there only 0.005% of applications come from the student.
- A client who works as an accountant can be said to be a client who is capable of repaying the loans (95% of applications approved). But, there is only 3.19% of applications come from an accountant. So do, the client who work as high skill tech staff and manager, they are capable of repaying the loans, but there are only a few applications



**Create a campaign** so that **more** student, accountant, high skill tech staff, manager **interested in applying for a loan**

- Clients with maternity leaves and cash loans can be said to be a client who is incapable of repaying the loan (100% of applications rejected). On the contrary, all clients with maternity leave but taking revolving loans to have their applications approved.
- For unemployed clients, more than 50% of them have a problem repaying their loans if they take cash loan contracts. Meanwhile, all unemployed client who takes revolving loans is capable of repaying the loan.



**Need further analysis**, you can **survey** to find out if there is a problem if a client with maternity leaves or unemployed takes a cash loans contract. So, in the future, if there are clients with that type of income, you **can recommend the right contract type** so that their applications will be approved

A collage of several images showing diverse groups of people in a meeting or collaborative work environment, overlaid with a semi-transparent red filter. The collage is positioned at the top left of the slide.

# Thank You!



Project Repositories

<https://github.com/devadenti/hci-final-project>

