



Credit Scorecard Model Using Logistic Regression Study Case: **Home Credit**

Written by Novrizal Roynanda

- 1 Business Understanding
- 2 Business Insights
- 3 Data Processing
- 4 Data Modeling
- 5 Credit Scorecard

PART ONE

01

Business Understanding



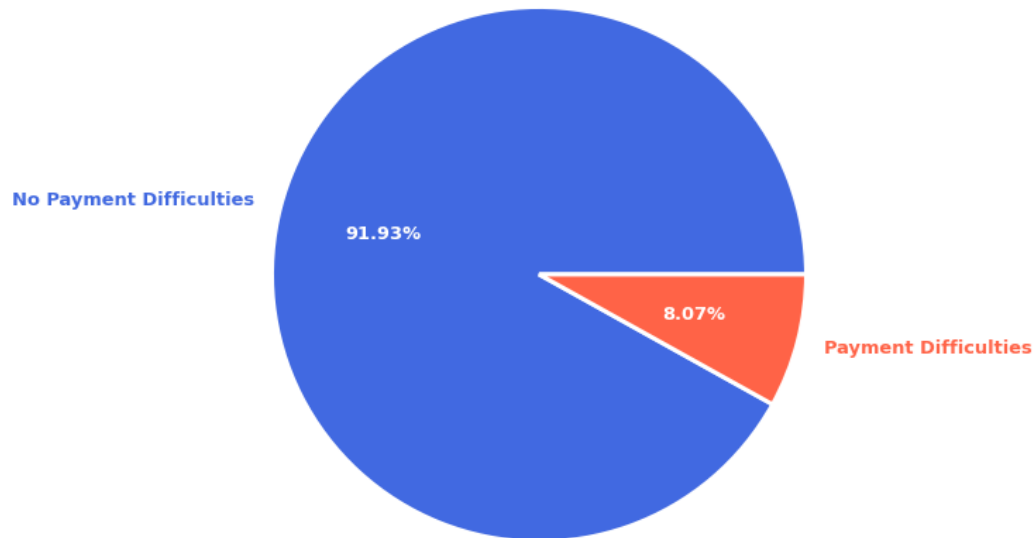
Background

Founded in 1997, Home Credit is an international consumer finance provider with operations in eight countries. Home Credit focus on responsible lending model empowers underserved customers with little or no credit history to access financing, enabling customers to borrow easily and safely, both online and offline.

Dataset Overview

- There are about **91%** (**282,686 applicants**) loans which indicates that client did not had any problems in repaying the loan in given time.
- While **9%** of the total loans (**24,825 applicants**) involved the clients having problems in repaying the loan.

Distribution of Client Repayment Abilities



Problem Statement

A major challenge for banks and other finance lending agencies is to decide for which candidates to approve loans. 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.

Business Objectives

Create a credit scoring system where the inputs are various features describing the financial and behavioral history of the loan applicants, in order to automatically predict whether the loan will be repaid or defaulted.

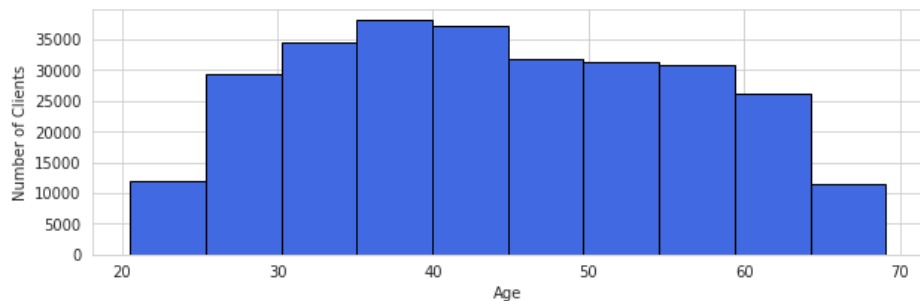


PART TWO

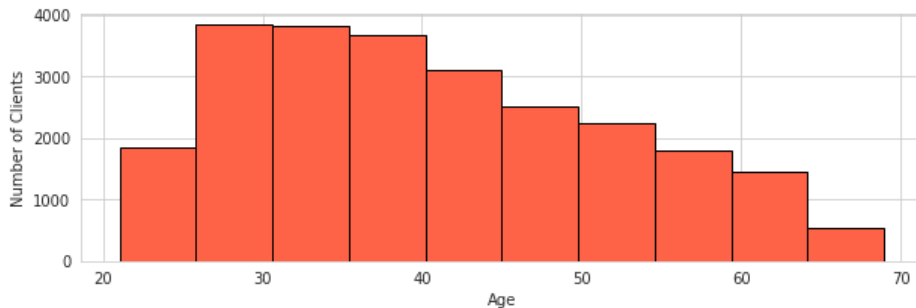
02

Business Insights

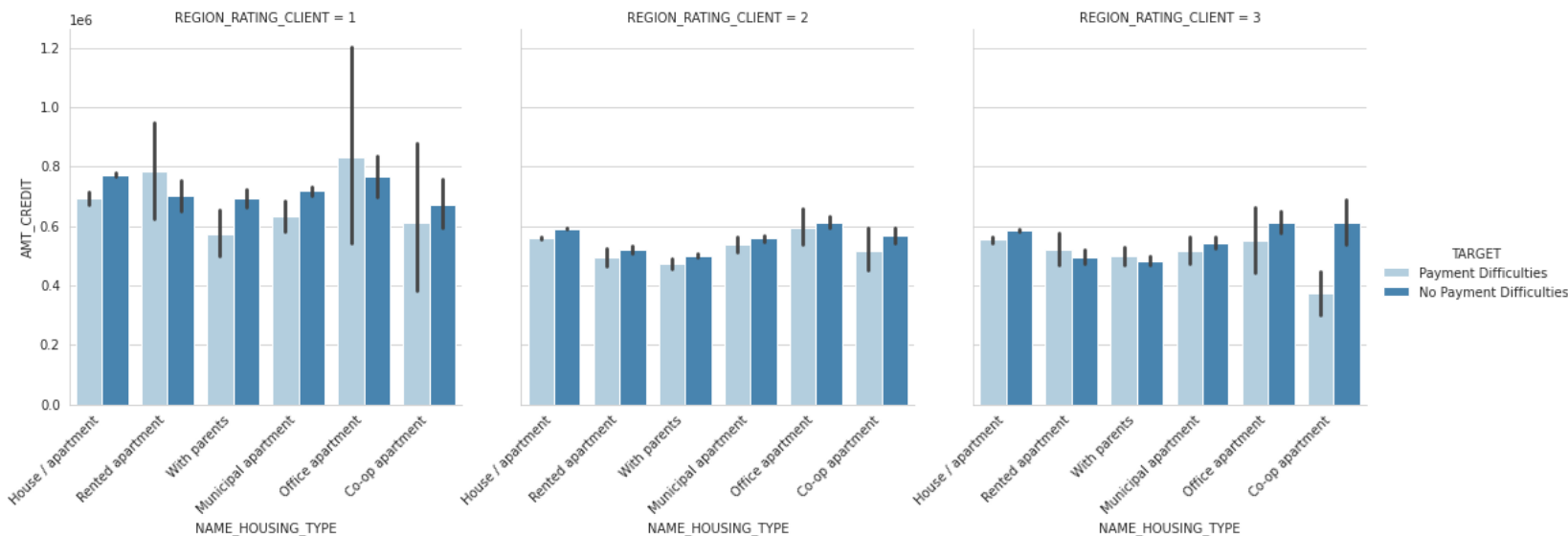
Age of Client (in years) who have No Payment Difficulties



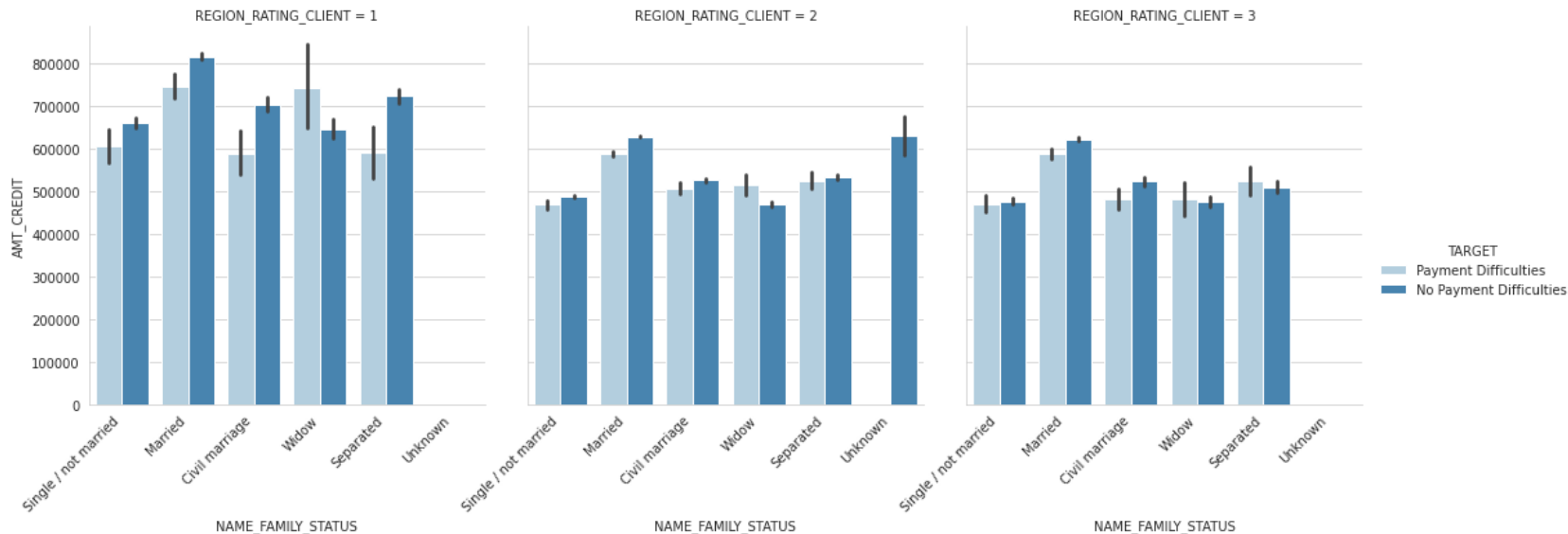
Age of Client (in years) who have Payment Difficulties



Clients who have no payment difficulties are client in the range of **35-45 years**. While clients who have payment difficulties are client in the range of **25-35 years**.



Clients who live in rented apartment and office apartment and their region have a rating of 1, have a problem repaying the loans compared to client in region with rating of 2 or 3.



Clients who have a family status as widow, whether they live in a region with a rating of 1, 2, or 3, have problems repaying loans for moderate to higher amounts credit of the loan. For clients who have a family status as separated and live in region with a rating of 3, also have problems repaying loans for a moderate amount credit of the loan.

PART THREE

03

Data Processing



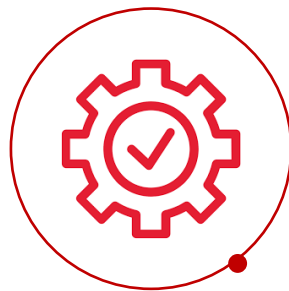
Data Cleansing

- Check Data Duplicates
- Check Missing Values
- Drop Feature
- Simple Imputer Median.



Feature Selection

- Data Splitting (80:20)
- Categorical Data
(Chi-Square)
- Numerical Data
(ANOVA)



Feature Engineering

- Information Value (IV)
- Weight of Evidence Binning
- Weight of Evidence Transform



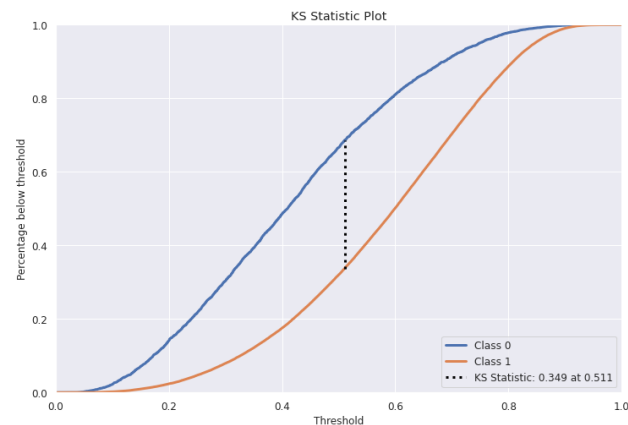
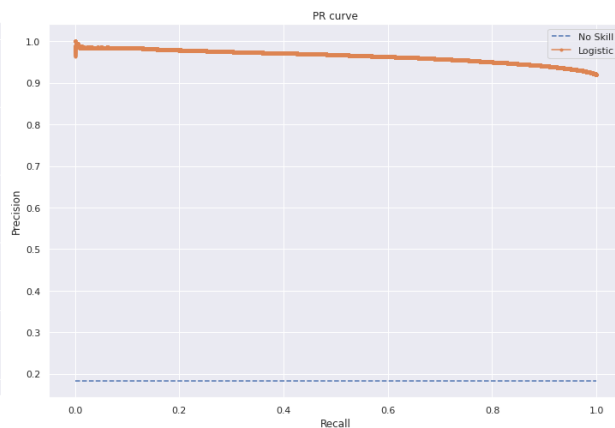
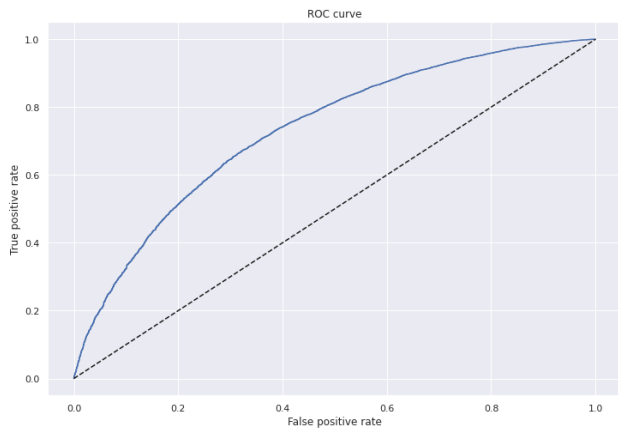
Modeling and Evaluation

- Logistic Regression
- Class Weight = Balanced
- AUROC
- PR Curve
- KS Statistics

PART FOUR

04

Data Modeling



Model Result

| Models | MEAN AUROC | GINI |
|---------------------|------------|--------|
| Random Forest | 0.6578 | 0.3155 |
| Decision Tree | 0.5351 | 0.0702 |
| Logistic Regression | 0.7310 | 0.4620 |

Based on AUROC score **0.7310**, PR AUC score **0.9642**, and, KS score **0.349**, the Logistic Regression model is considered as a good performance model.

PART FIVE

05

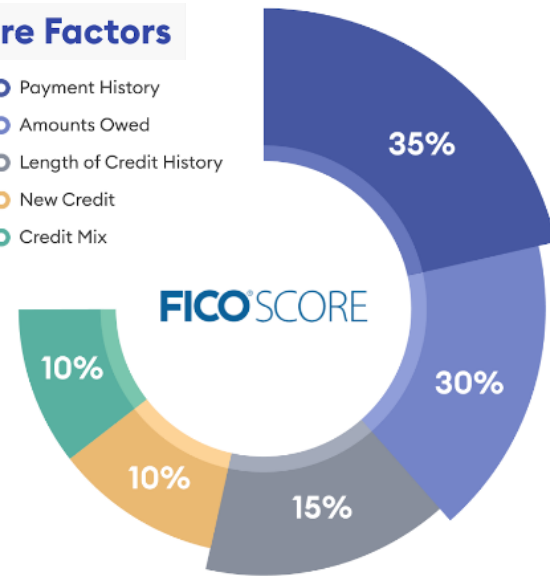
Credit Scorecard

Using FICO Score

A FICO score is a credit score created by the Fair Isaac Corporation (FICO). Lenders use borrowers' FICO scores along with other details on borrowers' credit reports to assess credit risk and determine whether to extend credit. Most credit scores have a 300-850 score range. The higher the score, the lower the risk to lenders.

FICO Score Factors

- Payment History
- Amounts Owed
- Length of Credit History
- New Credit
- Credit Mix



Threshold = 0.5

| Accept Score | N Approved | N Rejected | Approval Rate | Rejection Rate |
|--------------|------------|------------|---------------|----------------|
| 560.0 | 39840 | 21663 | 0.647773 | 0.352227 |

Best Threshold

| Accept Score | N Approved | N Rejected | Approval Rate | Rejection Rate |
|--------------|------------|------------|---------------|----------------|
| 512.0 | 55663 | 5840 | 0.905045 | 0.094955 |

We will choose the best threshold with accept score **512** to avoid higher rejection rate with threshold **0.5** but will monitoring model's performance in production.

**Setting Loan
Approval Cut-offs**

Base (Intercept) = 560

Min Score = 300

Max Score = 850

Threshold = 0.5

Best Threshold = 0.29682

Accept Threshold = 512

Application Test = 61503 Applicants

Model = Logistic Regression

Scorecard Result

HOME
CREDIT

| Feature | Specific Feature | Score |
|------------------------|-------------------------------------|-------|
| CODE_GENDER | Male | -10 |
| | Female or XNA | 10 |
| NAME_EDUCATION_TYPE | Academic degree | 60 |
| | Higher education | 3 |
| | Incomplete higher | -9 |
| | Lower secondary | -31 |
| | Secondary / secondary special | -21 |
| NAME_FAMILY_STATUS | Single or Unknown | -2 |
| | Civil marriage | -4 |
| | Married | 7 |
| | Separated | -3 |
| | Widow | 3 |
| NAME_INCOME_TYPE | Businessman or Commercial Associate | 8 |
| | Pensioner or Maternity leave | 7 |
| | Student or Unemployed | -40 |
| | State servant | 22 |
| | Working | 3 |
| REG_CITY_NOT_LIVE_CITY | 0 | 7 |
| | 1 | -6 |
| FLAG_DOCUMENT_3 | 0 | 9 |
| | 1 | -8 |

| | | |
|-----------------------------|-----------------|-----|
| REGION_RATING_CLIENT_W_CITY | 0 | 0 |
| | 1 | 18 |
| | 2 | 9 |
| REGION_POPULATION_RELATIVE | <0.0147 | 2 |
| | 0.0147-0.0292 | 1 |
| | 0.0292-0.0436 | -2 |
| | 0.0436-0.0581 | 1 |
| | >0.0581 | -1 |
| YEAR_LAST_PHONE_CHANGE | <2 | -7 |
| | 2-4 | 3 |
| | 4-6 | 2 |
| | 6-8 | 5 |
| | 8-10 | 5 |
| | >10 | 0 |
| YEAR_BIRTH | <30 | -2 |
| | 30-40 | -10 |
| | 40-50 | -2 |
| | 50-60 | 6 |
| | >60 | 8 |
| AMT_CREDIT | <846000 | -4 |
| | 846000-1647000 | 0 |
| | 1647000-2448000 | 9 |
| | 2448000-3249000 | 3 |
| | >3249000 | -7 |

| | | |
|-------------------|--------------|-----|
| YEAR_ID_PUBLISH | <4 | -11 |
| | 4-8 | -6 |
| | 8-12 | -4 |
| | 12-16 | 4 |
| | >16 | 17 |
| YEAR_REGISTRATION | <17 | -3 |
| | 17-34 | 0 |
| | 34-51 | -3 |
| | >51 | 6 |
| EXT_SOURCE_2 | <0.0855 | -53 |
| | 0.0855-0.171 | -35 |
| | 0.171-0.256 | -24 |
| | 0.256-0.342 | -13 |
| | 0.342-0.427 | -5 |
| | 0.427-0.513 | 3 |
| | 0.513-0.598 | 10 |
| | 0.598-0.684 | 21 |
| | 0.684-0.769 | 38 |
| | >0.769 | 59 |
| EXT_SOURCE_3 | <0.0901 | -68 |
| | 0.0901-0.18 | -49 |
| | 0.18-0.269 | -33 |
| | 0.269-0.359 | -14 |
| | 0.359-0.448 | -1 |
| | 0.448-0.538 | 5 |
| | 0.538-0.627 | 26 |
| | 0.627-0.717 | 36 |
| | 0.717-0.806 | 48 |
| | >0.806 | 51 |

| Name: Novrizal Roynanda | | | Score Calculation |
|-----------------------------|-----------------|--|-------------------|
| Base Score | | | 560 |
| CODE_GENDER | Male | | -10 |
| NAME_EDUCATION_TYPE | Academic Degree | | 60 |
| NAME_FAMILY_STATUS | Single | | -2 |
| NAME_INCOME_TYPE | Working | | 3 |
| REG_CITY_NOT_LIVE_CITY | 0 | | 7 |
| FLAG_DOCUMENT_3 | 0 | | 9 |
| REGION_RATING_CLIENT_W_CITY | 2 | | 9 |
| REGION_POPULATION_RELATIVE | 0.0600 | | -1 |
| YEAR_LAST_PHONE_CHANGE | 7 | | 5 |
| YEAR_BIRTH | 26 | | -2 |
| AMT_CREDIT | 800000 | | -4 |
| YEAR_ID_PUBLISH | 5 | | -6 |
| YEAR_REGISTRATION | 5 | | -3 |
| EXT_SOURCE_2 | 0.420 | | -5 |
| EXT_SOURCE_3 | 0.420 | | -1 |
| Total | | | 619 |

Simulation

- If Total Score > Accept Threshold: Approve
- If Total Score < Accept Threshold: Reject
- Novrizal's Score (619) is higher than Accept Threshold (512), then his loan would be Approved



Visit My GitHub!

You can see the entire project documentation here from my github

[@novrizalrnd](#)