Home Credit Scorecard Model

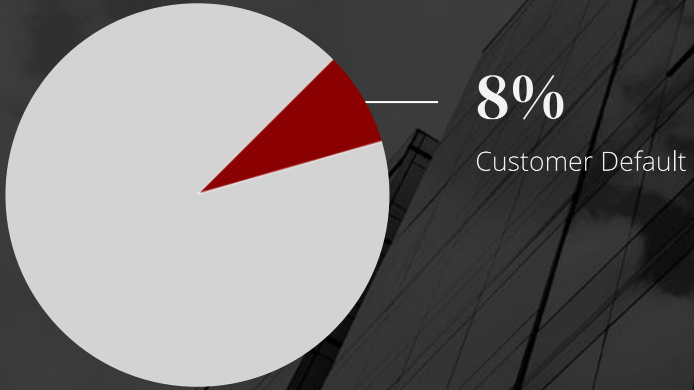
Anik Maulia Tri Handayani
Project Based Virtual Intern: Data Scientist

Our company

PT Home Credit Indonesia is a technology-based financing company that has been in operation since 2013, with an extensive network of partner stores in over 200 cities across Indonesia. Home Credit continues to evolve into a trusted financial partner for millions of customers. Currently, we are designing statistical methods and machine learning techniques to generate credit score predictions. By doing so, we ensure that customers capable of making payments are not rejected when applying for loans. Loans can be granted with a principal, maturity, and repayment calendar designed to motivate customers toward success.

Problem

The Distribution of Customers Loan Status



8% of the total 307,511 customers are experiencing loan defaults, posing a potential risk of financial loss.

Goals

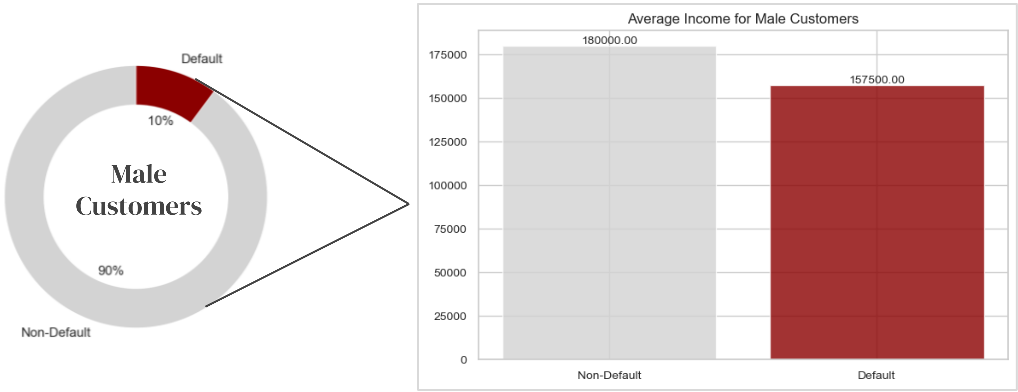
Decrease the default rate among customers to minimize the potential risk of financial loss.

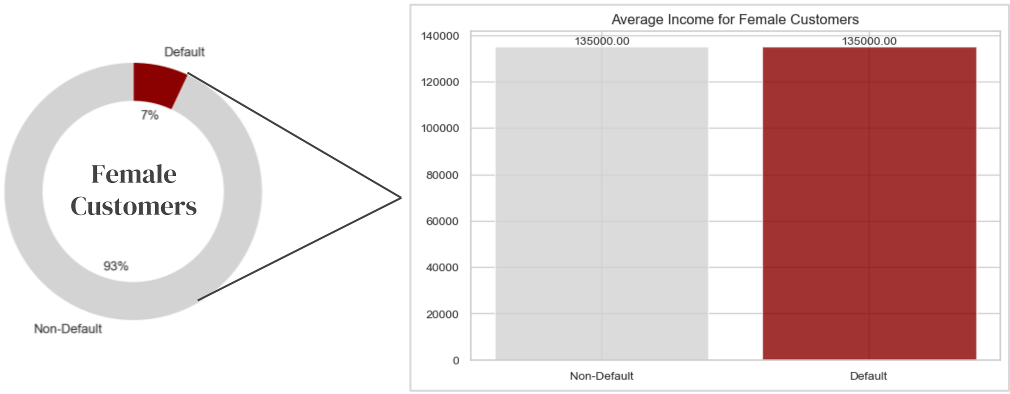
Objective

Develop a classification model to generate credit score predictions for binary values (0 indicating clients without payment difficulties and 1 indicating clients with payment difficulties), and selectively approve loans based on predictions of 0.

Business Metrics

Default rate.





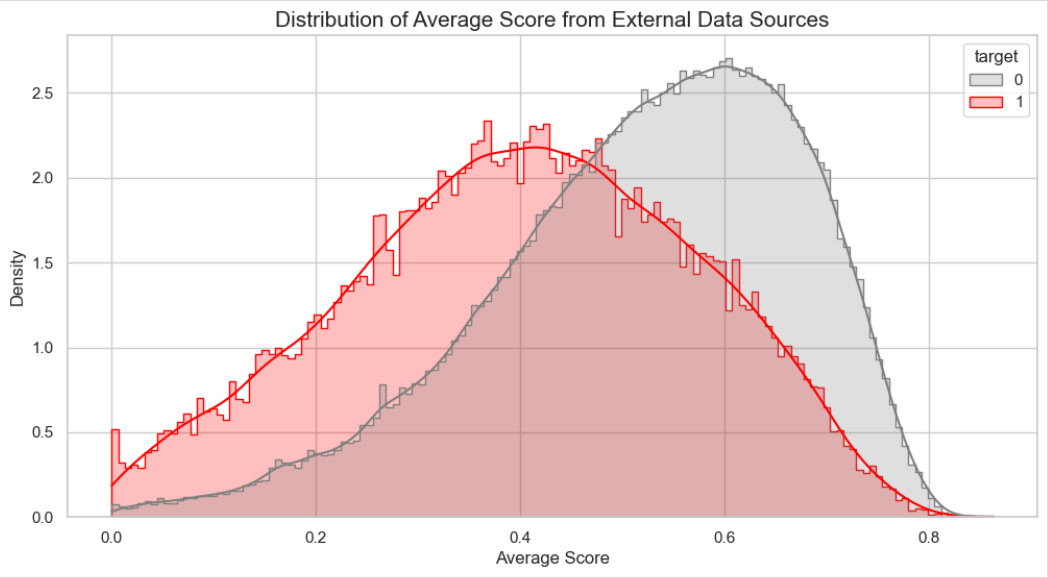
01 Business Insight

Default Rate by Gender and Income

- Male Customers: Higher default rates.
  Those facing payment challenges have a significantly lower average income, highlighting the pronounced impact of income on defaults.
- Female Customers: Lower default rates.
  The data suggests that income may not have a significant impact for females, as both default and non-default customers exhibit similar average incomes.
- Recommendation: Consider income as a critical factor, especially for **male** customers.
  - ➢ Action: High income may warrant loan approval, while low income may justify rejection to reduce default risk and potential losses.

02 Business Insight

Default by Average Score from External Data Sources



- Customers facing payment difficulties tend to have a lower average score from external data sources compared to customers without payment difficulties.

- Recommendation: Consider the external score as a critical factor.
  - ➢ Action: A high score may justify loan approval, while a low score may warrant rejection to mitigate default risk and potential losses.

Data Pre-Processing

01

Data Splitting

Data Train: 80% (246,008 rows)
Data Validation: 20% (61,503 rows)

02

Handle Missing Values

Drop column (missing values >35%)
SimpleImputer (missing values <35%)

03

Handle Invalid Values

13 columns, include invalid values and data type

04

Feature Extraction

2 features

05

Feature Scaling

Yeo-Johnson Transformation

06

Feature Encoding

Label Encoding,
One Hot Encoding

07

Handle Imbalance Class

Resampling

08

Feature Selection

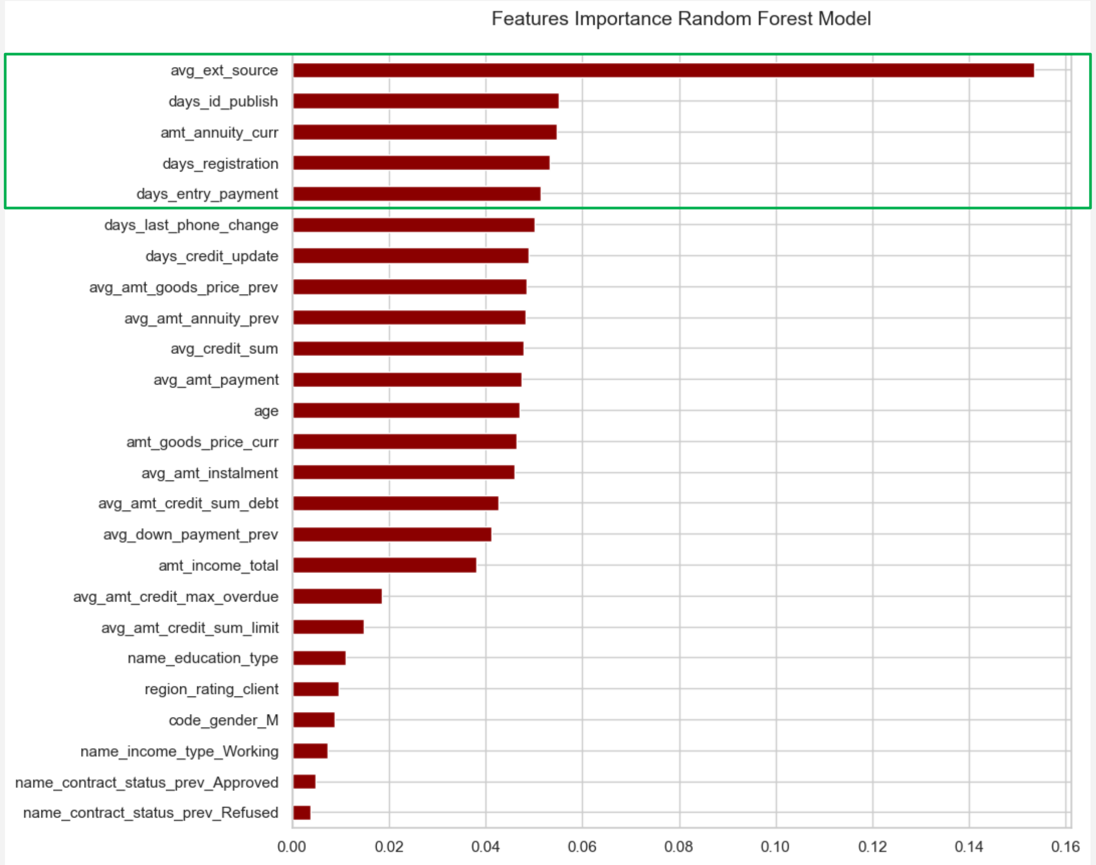
Anova, Variance, MI, Select KBest, ExtraTreesClassifier, Multicollinearity

Current recommendations focus on male clients and those with lower external scores. To address wider payment challenges, a machine learning model has been developed for accurate predictions and effective recommendation tailoring. Now, let's review each model's performance.

Modeling

	Models	Recall (Train)	Recall (Val)	Precision (Train)	Precision (Val)	Accuracy (Train)	Accuracy (Val)	ROC AUC (Train)	ROC AUC (Val)
0	Adaboost Classifier	0.680000	0.690000	0.680000	0.900000	0.680000	0.690000	0.680000	0.680000
1	Decision Tree	1.000000	0.860000	1.000000	0.860000	1.000000	0.660000	1.000000	0.540000
2	Gradient Boosting Classifier	0.690000	0.690000	0.690000	0.900000	0.690000	0.690000	0.690000	0.670000
3	Logistic Regression	0.480000	0.660000	0.590000	0.870000	0.580000	0.660000	0.580000	0.570000
4	Random Forest	1.000000	0.920000	1.000000	0.880000	1.000000	0.920000	1.000000	0.510000
5	XGBoost Classifier	0.790000	0.740000	0.760000	0.890000	0.770000	0.740000	0.770000	0.670000

The model with the best and consistently good performance is Random Forest. Therefore, predictions for clients with payment difficulties and without payment difficulties on the test data will be conducted using the Random Forest model.



Business Recommendations

1. Client Prioritization:
   Target clients with robust external scores (> 0.4).
2. Document Turnover Rule:
   Enforce a streamlined 9-year document turnover policy.
3. Annuity Adjustment Strategy:
   Fine-tune annuity amounts for clients in payment challenges.
4. Registration Stability:
   Ensure a maximum 12-year registration stability.
5. Credit Payment Cap:
   Implement a 1.8-year cap for previous credit payments.

Business Simulation

TP
19,986
6.5%

FP
4839
1.57%

FN
172
0.06%

TN
282,514
91.87%

Before Modeling

- All 307,511 loan applications accepted.
- No rejections for clients without payment difficulties.
- 24,825 clients face payment difficulties (8% default rate).
- Potential loss: 12,748,407,075 IDR (Assuming 513,531 IDR loss per client).
- Adjusted potential loss (50% treated): 6,373,946,772 IDR.

After Modeling

- Only 287,353 predicted non-difficulty clients' (TN + FP) loans accepted.
- 172 non-difficulty clients' loans rejected.
- 4,839 clients face difficulties (1.7% default rate).
- Potential loss: 2,484,976,509 IDR (Assuming 513,531 IDR loss per client).
- Adjusted potential loss (50% treated): 1,242,231,489 IDR.

Model significantly reduces default rate and potential losses.

Thanks

For more detail, click here

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution