Home Credit Scorecard Classification Model

Deni Yuniawan

Final Task of Project-Based Internship Home Credit Indonesia x Rakamin Academy Batch December 2024

Github : https://github.com/deniywn/rakamin-hci_pbi

Problem Research

Dataset sourced from PBI HCI x Rakamin Academy.

Home Credit Indonesia aims to develop a model to predict the credit scores of their customers.

The dataset includes 122 columns, with one TARGET column.

Objective & Goal

Objective

Speed up credit score calculations using a machine learning model to avoid rejecting potential customers.
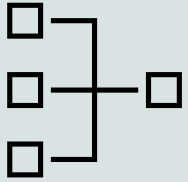
Goals

- Develop a machine learning model for credit score classification.

- Provide business recommendations.

## Business Metrics

Metric: credit score ratio

Data Pre-Processing

- Identify and remove features with a significant amount of missing values.

- Conduct univariate analysis and clean the features.

- Convert the 'days' column to 'age/year'.

- Transform binary categorical features into binary numerical ones.

- Check the correlation among numerical features.
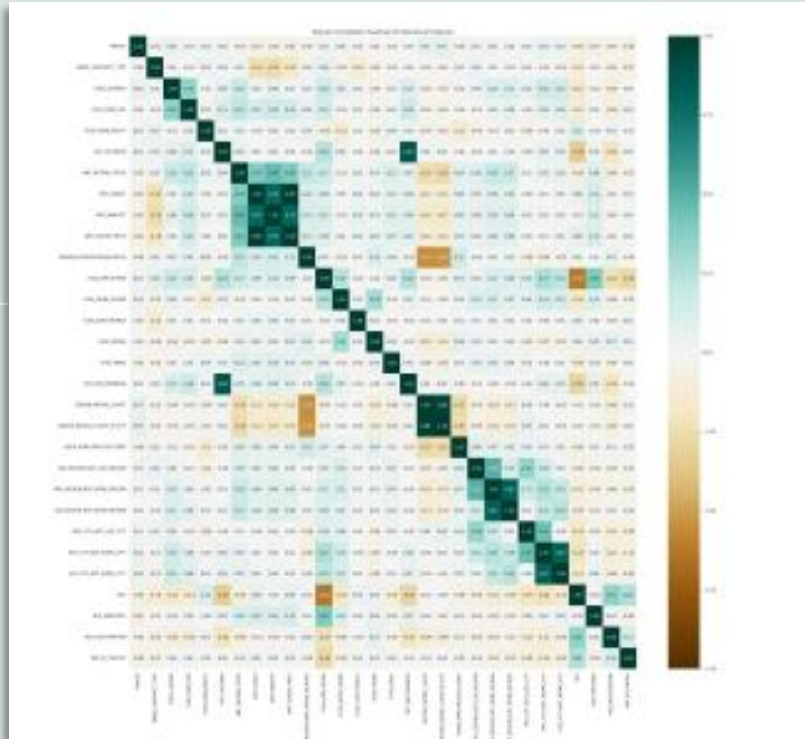
- Remove features that have a zero correlation with the TARGET variable.

- Eliminate features with a correlation greater than 0.6 with another feature to prevent multicollinearity.
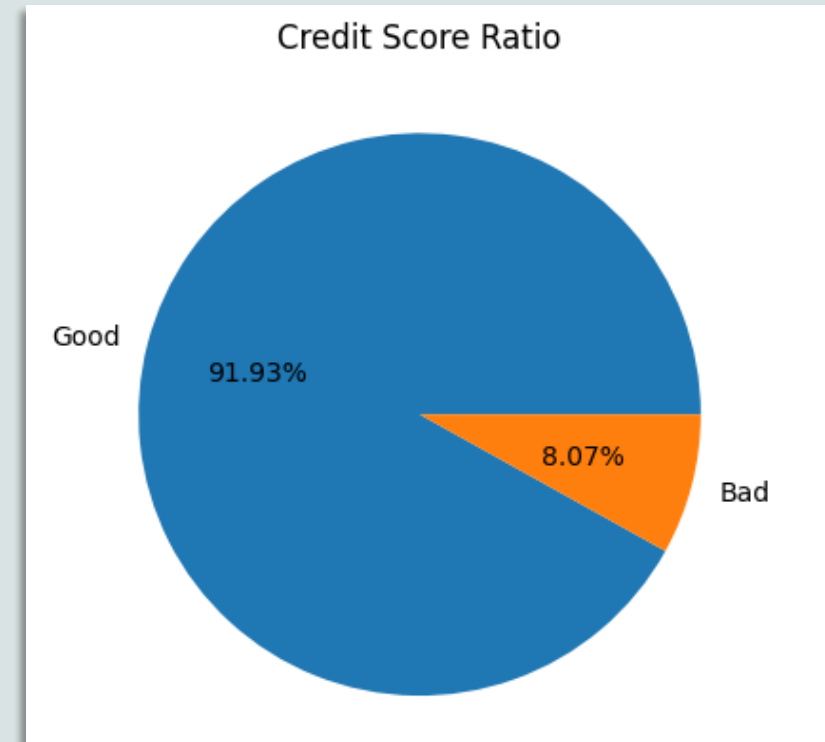
- Apply StandardScaler to non-binary numerical features.

- Encode selected categorical features.

- Use oversampling to handle imbalanced data.
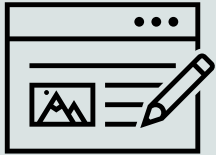
Data Visualization & Business Insight
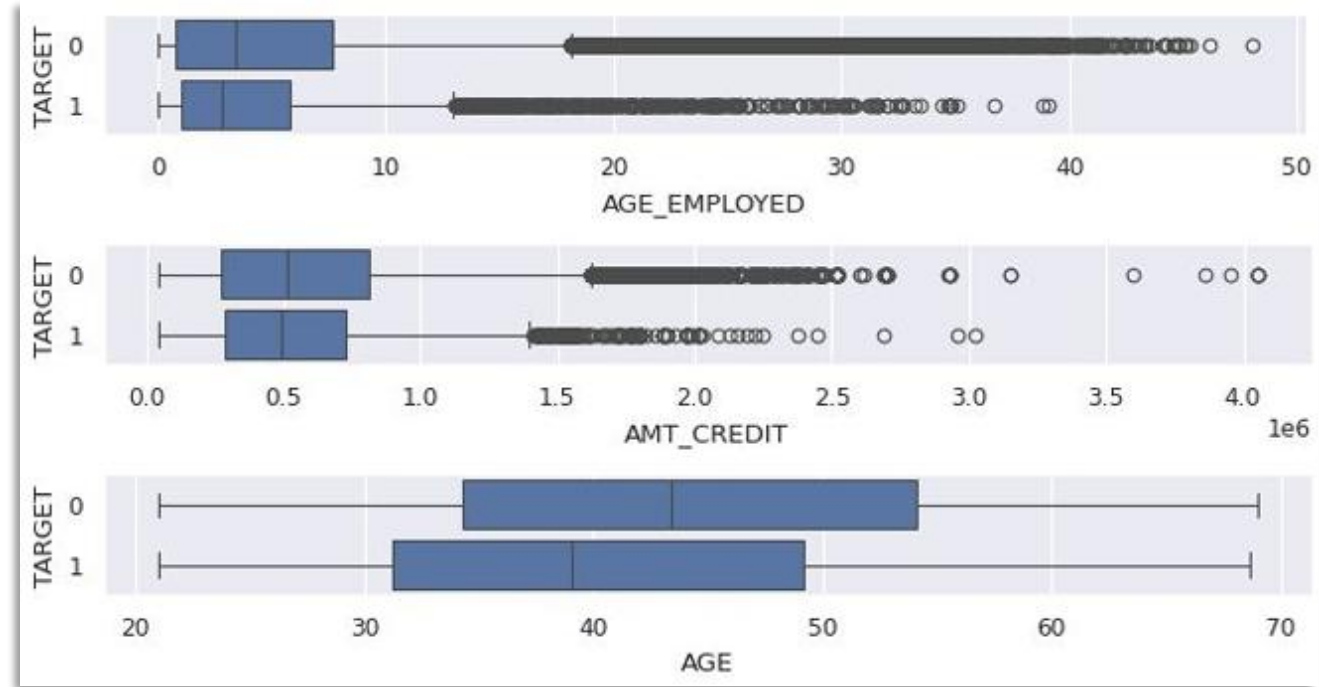
- Target value: 0 indicates good credit scores, 1 indicates bad credit scores.

- Customers with higher credit amounts tend to have better scores.

- Older customers and those with longer employment tend to have better credit scores, possibly due to increased maturity, responsibility, and financial stability.
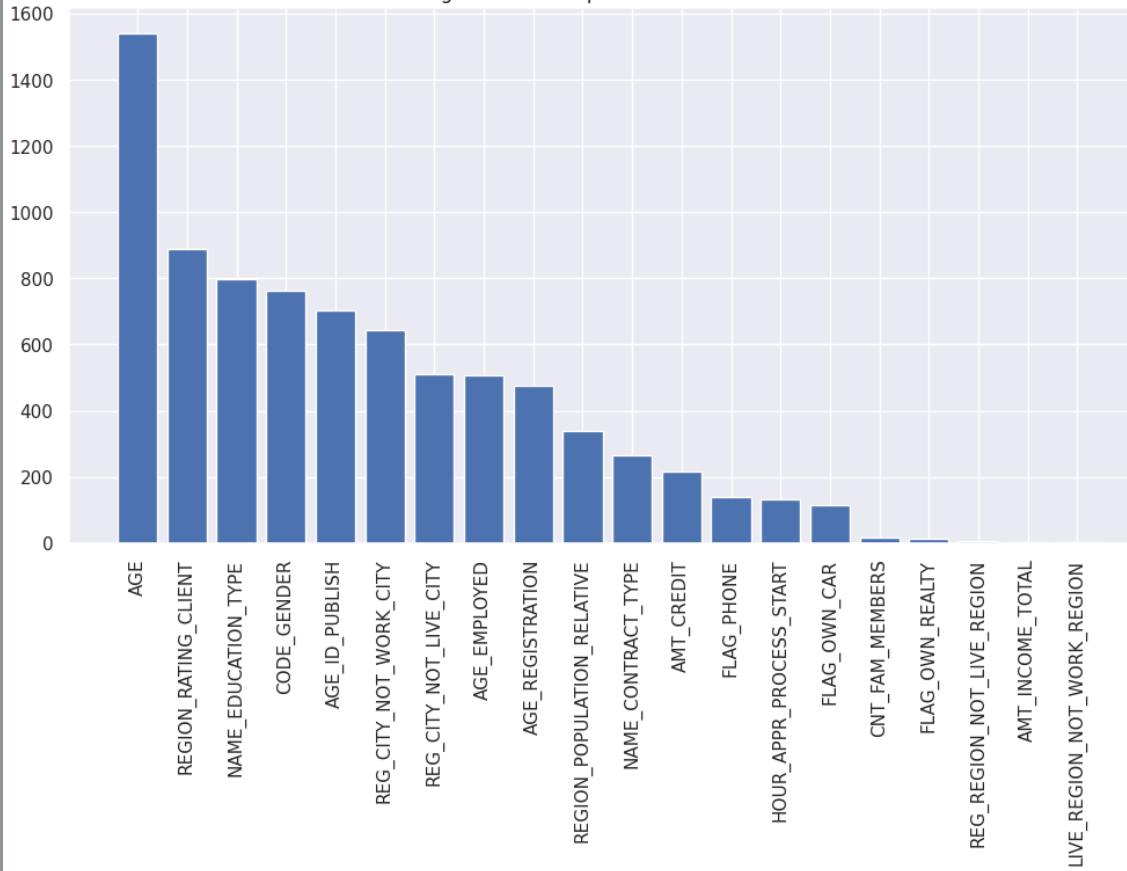
Insight

- Age is the most significant feature according to the ANOVA method.

- Older customers are generally more likely to have better credit scores. It is possible because most older people are more mature, responsible, and financially stable.

- Other important features include region rating and education type.

- Customers with education levels higher than secondary rarely have bad credit scores.
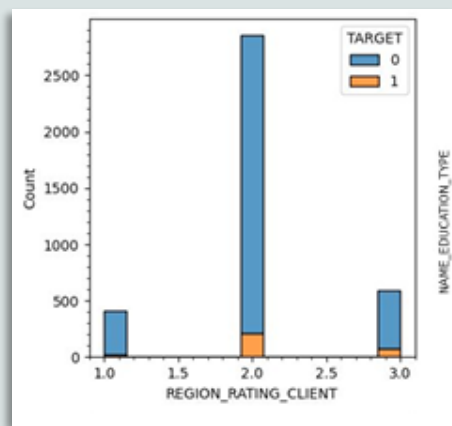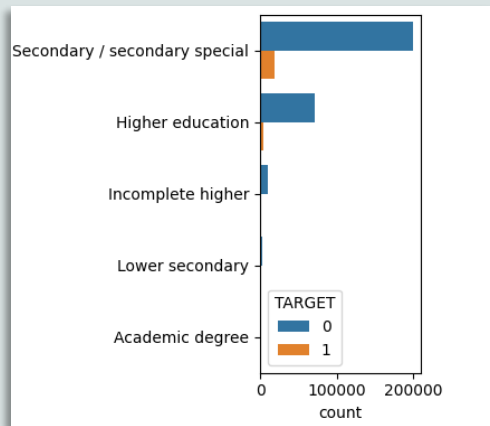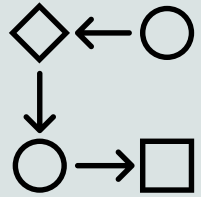
Machine Learning Implementation & Evaluation

Metode:

- Logistic Regression

- Decision Tree

- Hyperparameter Tuning

The results

Model	Before Hyperparameter Tuning		After Hyperparameter Tuning	
	Accuracy (train)	Accuracy (test)	Accuracy (train)	Accuracy (test)
Logistic Regression	65.735%	60.047%	65.735%	60.047%
Decision Tree	100%	82.464%	91.181%	87.877%

For this case, decision tree with hyperparameter tuning is the best prediction model.

Business Recommendation

- A Decision Tree model provides better results for this data.

- Older individuals, experienced professionals, and those with higher education levels are more likely to have good credit scores.

- Home Credit can target these customer segments to offer their products.