Rakamin Academy X HOME CREDIT

Machine Learning Project

Scorecard Model

Data Science - Project Based Internship

Presented by
Oktavian Dwi Putra

Oktavian Dwi Putra

About Me

Result oriented professional with background in digital marketing, especially SEO and a strong desire to transition into the field of Data Science. Possessing a solid foundation in statistics, machine learning, and data analysis. Eager to apply my analytical mindset, problem solving abilities, and passion for data driven insights to drive meaningful outcomes as a Data Scientist.

Rakamin Academy

Experiences

SEO Specialist cmlabs
Apr 2022 Mar 2023

Improve website visibility and performance in search engines for several clients, including:

- Do keyword research for new content weekly.
- Perform onpage optimization for existing content to improve their performance.
- Monitor the ranking of targeted keywords from existing content.

Problem Statements:

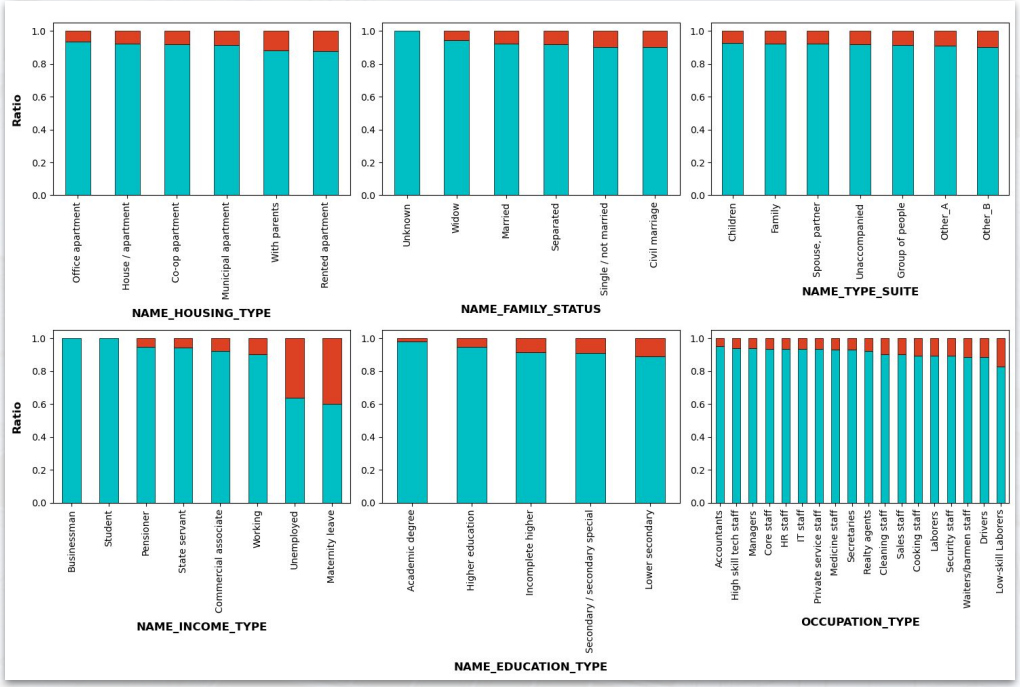
Home Credit is currently using various statistical methods and Machine Learning to make credit score predictions to ensure clients who are able to make payments are not rejected when applying for the loan.

Goals & Objectives:

Minimize the number of clients who are approved but actually defaulters and create predictive model to determine potential client and default client.

1. Exploratory Data Analysis

1.1. Categorical Data

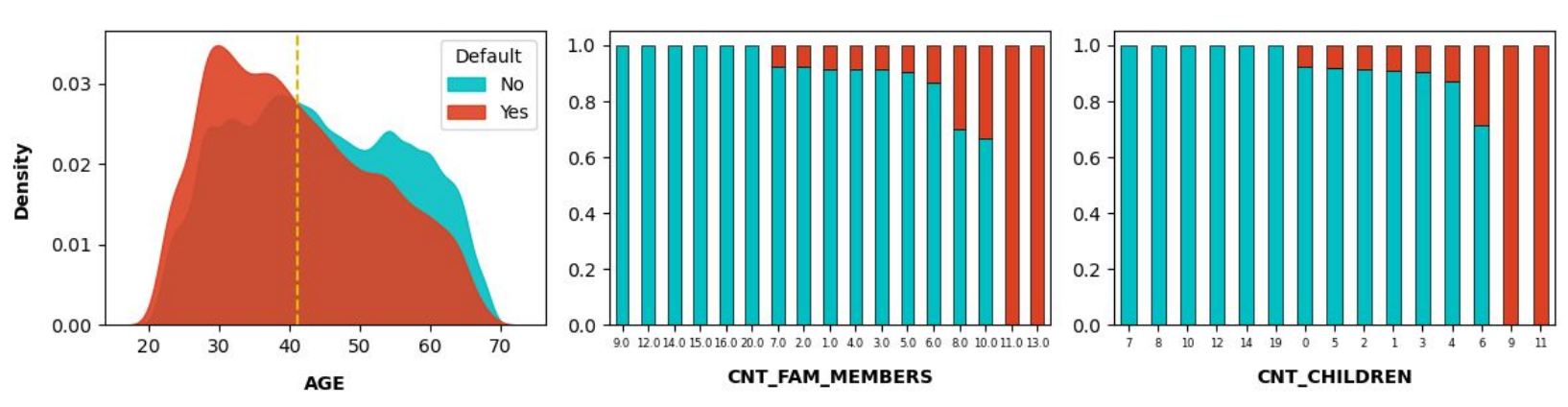


Observation:

1. Clients who own a house less likely to become default clients.
2. The highest ratio of default clients comes from NAME_FAMILY_STATUS Civil marriage and the lowest comes from Widow.
3. Clients who are accompanied by family or partner when applying for the loan are less likely to become default clients.
4. The highest ratio of default clients comes from Maternity leave and Unemployed clients and the lowest comes from Businessman.
5. The higher the education, the less likely to become the default clients.
6. The highest ratio of default clients comes from clients working as Low-skill Laborers and the lowest comes from Accountants.

1. Exploratory Data Analysis

1.2. Numerical Data



Observation:

1. The older the clients, the less likely to become default clients.
2. The more family members and children, the more likely, to become the default clients.

2. Data Preprocessing

There are some steps that we used before the data ready for modeling process, such as:

- Feature selection using Predictive Power Score.
- Handle missing value.
- Handle duplicate data.
- Feature encoding.
- Split data.
- Handle outliers.
- Handle class imbalance.
- Standardization.

3. Modeling

3.1. Model Training & Validation

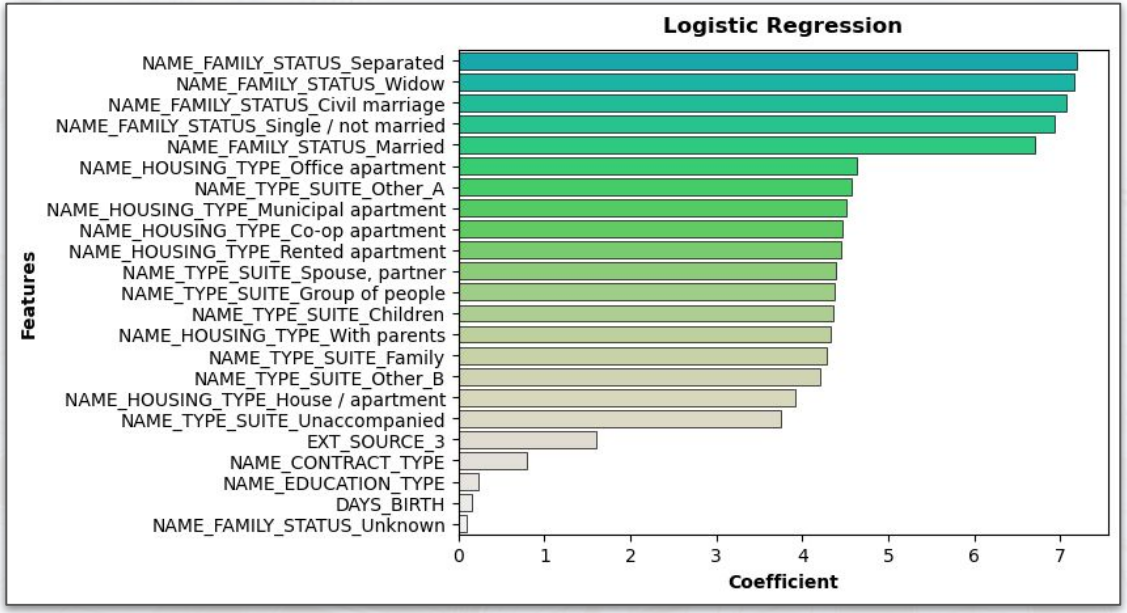
No.	Model	Acc (Train)	Acc (Test)	Δ Acc	Time Elapsed
1	Logistic Regression	0.84939	0.87734	0.02795	169.779793
2	Extra Trees	0.95340	0.87382	-0.07958	417.120225
3	Random Forest	0.96629	0.86977	-0.09652	865.018647
4	Gradient Boosting	0.86552	0.85087	-0.01465	1245.015871
5	Decision Tree	0.91870	0.82055	-0.09815	15.366524
6	AdaBoost	0.84746	0.80384	-0.04362	1248.123892

Observation:

Based on the performance results, we will choose Logistic Regression as the model because it has the highest accuracy on the test data.

3. Modeling

3.2. Feature Importances



Observation:

Based on feature importances (coefficients) from Logistic Regression model. The top 3 features to predict the default clients are NAME_FAMILY_STATUS, NAME_HOUSING_TYPE and NAME_TYPE_SUITE.

4. Business Recommendation

- Focus on the top 3 features to predict the default clients that are NAME_FAMILY_STATUS, NAME_HOUSING_TYPE and NAME_TYPE_SUITE.
- Focus on clients aged 40 years and over, widow or separated, who live in apartments, are accompanied by another person when applying for the loan and have high education or an academic degree because most of them do not have problems when repaying the loan.
- Be careful with clients who are young or under 40 years old, single or married, still live with their parents, are unaccompanied when applying for the loan, and have low education because most of them have problems when repaying the loan.

5. Documentation