Rakamin Academy

HOME CREDIT

SCORECARD MODEL

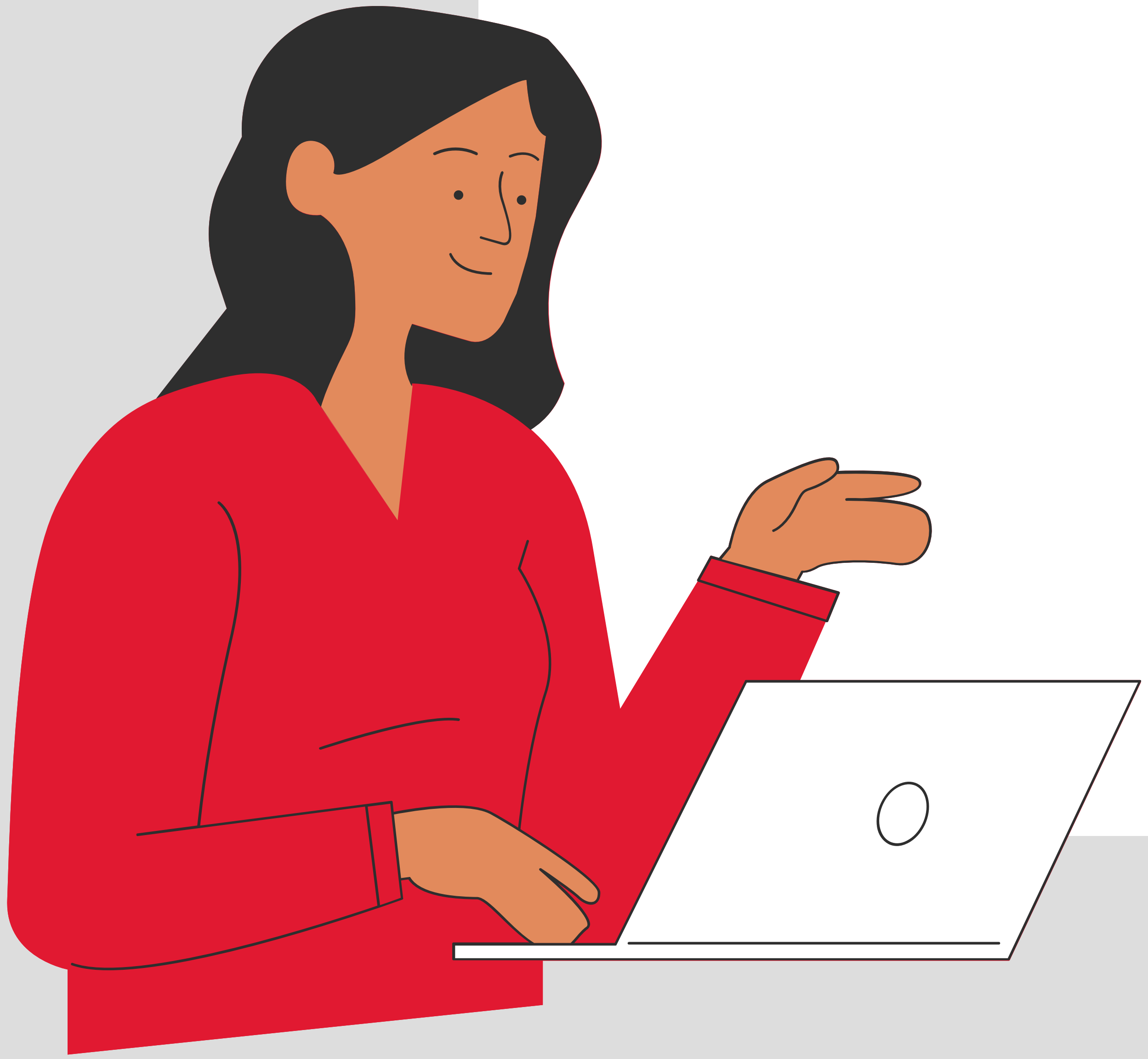
BY M. RIFKI OSKAR

TODAY'S AGENDA

1. Background
2. Exploratory Data Analysis
3. Data Preprocessing
4. Modelling

LET'S BEGIN!

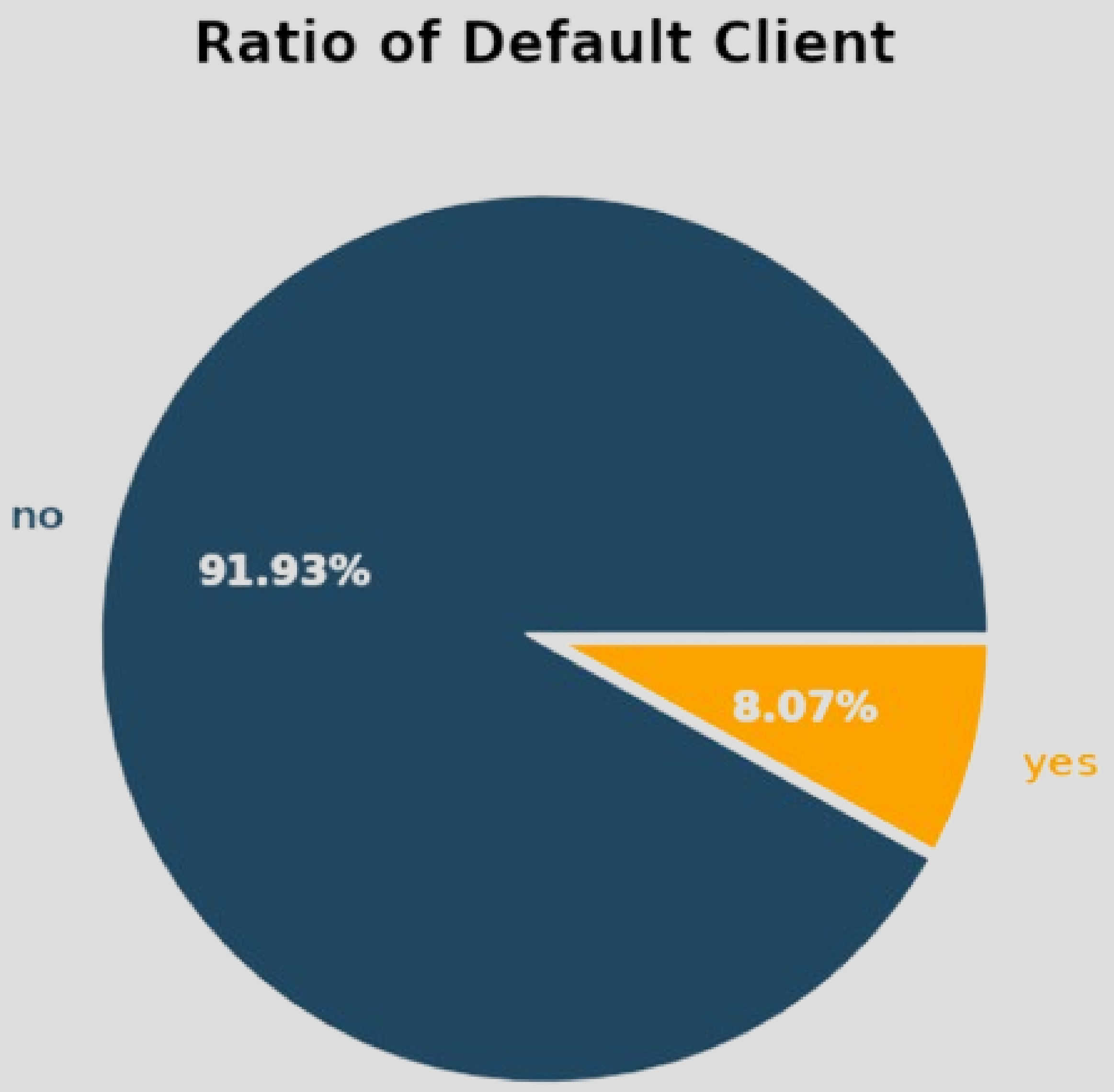
Are You Ready?



BACKGROUND

PROBLEM STATEMENT

We get 91.93% potential client and 8.07% default client



GOAL

Reducing the number of customers who are approved but actually defaulters

OBJECTIVE

Create machine learning to determine potential client and default client

BUSINESS METRIC

Default Loans (Low or High)

EXPLORATORY DATA ANALYSIS

Loan data	122 Number of Features	307510 Number of Rows

Numerical Features: 106

Categorical Features: 16

1.1.1 Numerical Features

```
[5]: ## Descriptive Statistics Numerical
num_features = df.select_dtypes(include=['int64', 'float'])
print('Total Numerical Features = {}'.format(num_features.shape[1]))
```

```
Total Numerical Features = 106
```

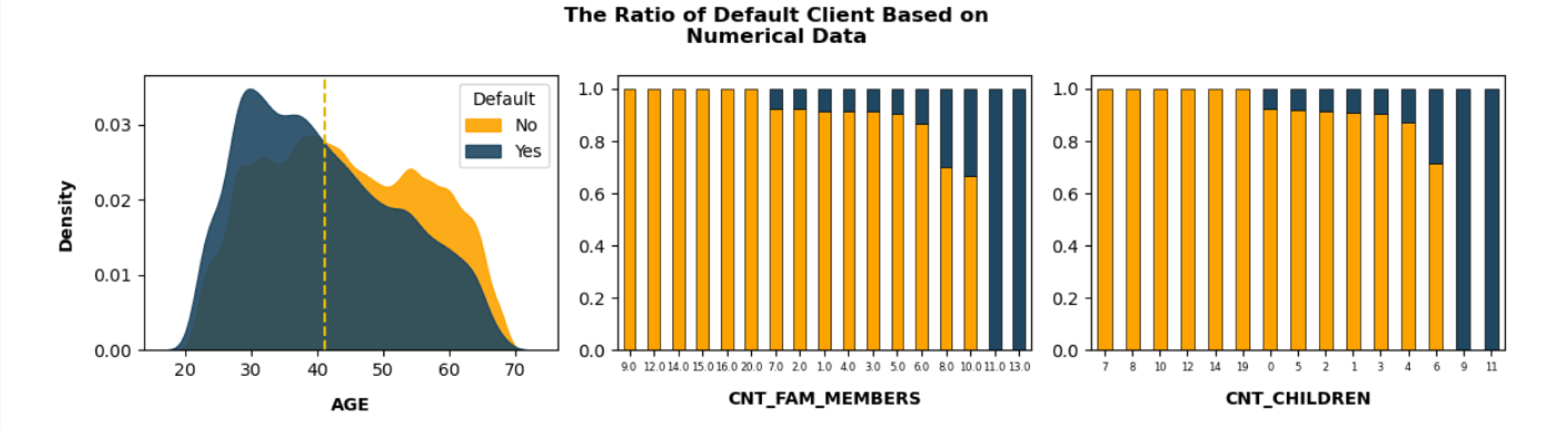
1.1.2 Categorical Features

```
[9]: ## Descriptive Statistics Categorical

cat_features = df.select_dtypes(include=['object'])
print('Total categorical features = {}'.format(cat_features.shape[1]))
```

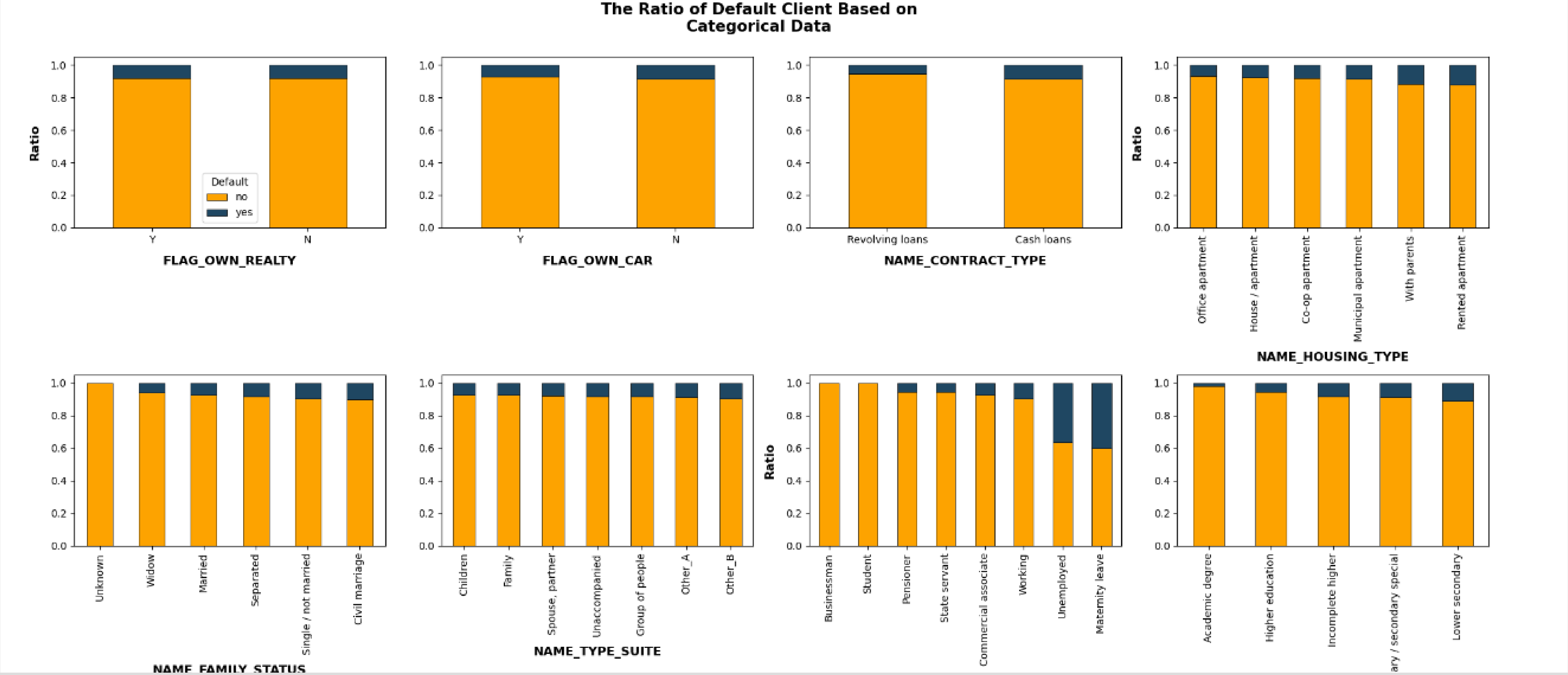
```
Total categorical features = 16
```

EXPLORATORY DATA ANALYSIS



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

EXPLORATORY DATA ANALYSIS



- The ratio of default clients who own a house or car is not much different from clients who do not own a house or car.
- Cash loans have a slightly higher default ratio compared to Revolving loans.
- Clients who own a house less likely to become default clients.
- The highest ratio of default clients comes from NAME_FAMILY_STATUS Civil marriage and the lowest comes from Widow.
- Clients who are accompanied by family or partner when applying for the loan are less likely to become default clients.
- The highest ratio of default clients comes from Maternity leave and Unemployed clients and the lowest comes from Businessman.
- The higher the education, the less likely to become the default clients.

PREPROCESSING

Handling Missing Value

- Drop null values
- Handling error values

Handling duplicated data

- tidak terdapat data yang duplicate

Feature selection

- Menggunakan metode predictive power score

Feature transformation

- melakukan standarization

Feature encoding

- mengubah feature **NAME_CONTRACT_TYPE** menjadi numeric, 0 dan 1 secara manual
- mengubah feature categorical menjadi numeric menggunakan one hot encoding

Handle class imbalance

- Menggunakan oversampling

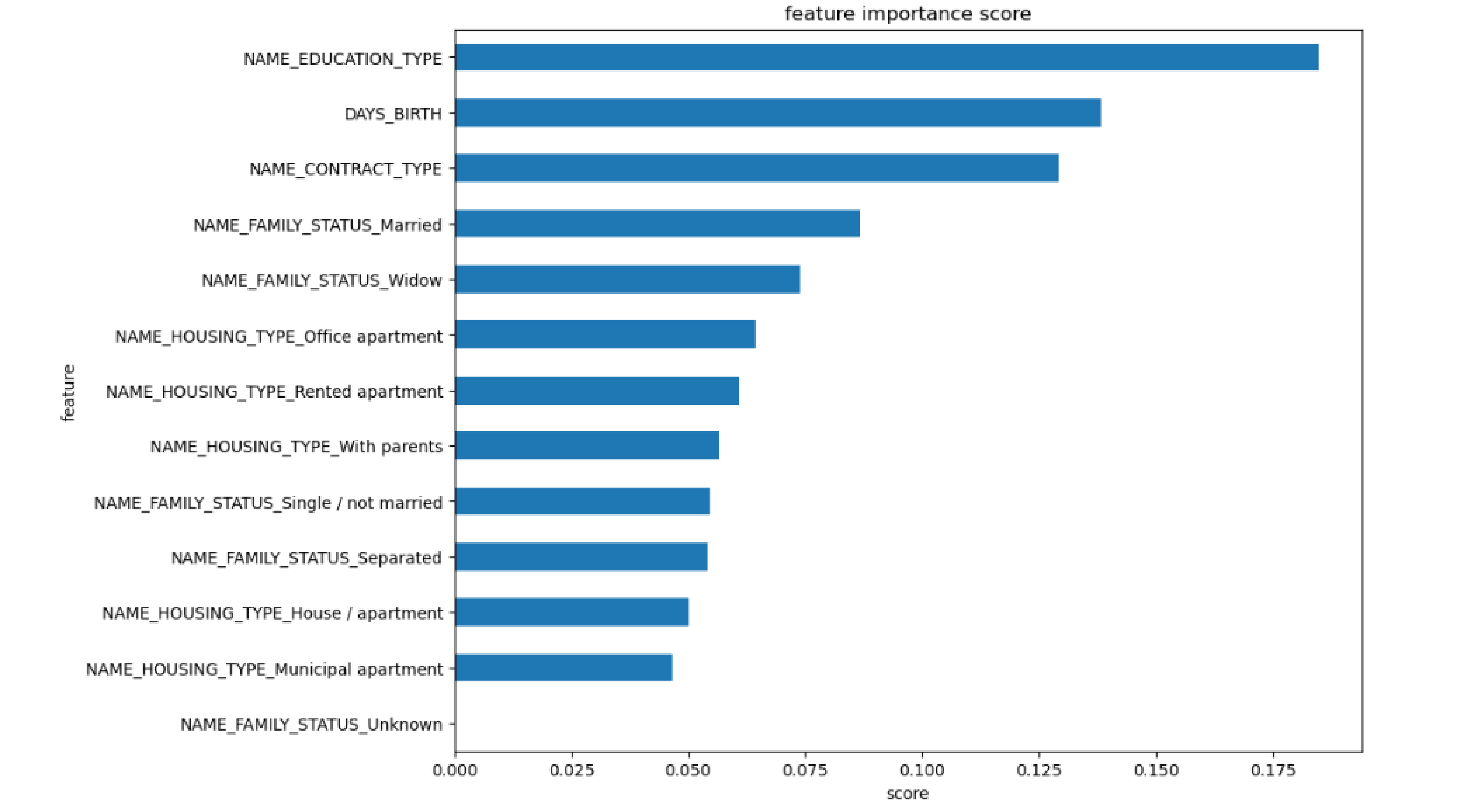
MODELLING

Algorithm	Evaluation Model				
	Accuracy	**Precision**	**Recall**	**AUC**	AUC **(Crossval)**
GradientBoost	**0.63**	**0.12**	**0.54**	**0.63**	**0.62**
AdaBoost	0.57	0.11	0.62	0.63	0.62
XGBoost	0.60	0.11	0.56	0.61	0.66
DecisionTree	0.77	0.09	0.20	0.52	0.97

- Metrics evaluasi yang digunakan adalah AUC**.**
- **GradientBoost** memiliki algoritma yang **paling baik** dengan gap antara AUC train dan test yang sangat kecil dengan hasil cross validation yang paling tinggi dari yang lainnya.

FEATURE IMPORTANCE

TOP Feature Importance From GradientBoost



BUSINESS RECOMMENDATION

Feature Importance

Focusing on feature importance could be a valuable business opportunity. By identifying the most important features and building a focused and effective model, this could help to reduce credit risk and make more informed lending decisions.

Outstanding Principal

Aligning products and interest rates with age groups, we're deploying targeted marketing and analyzing contracts for efficiency. Additionally, tailored financial education programs are introduced for diverse education levels, enhancing offerings and empowering customers.

Interest Rate

It can be beneficial to offer flexible interest rates that can be adjusted based on the borrower's age and education. This allows for a more personalized approach to lending and can help lenders manage their credit risk more effectively.

“

Believe. You're halfway there. Keep hustling to the finish line.

THANK YOU!

Have a
great day
ahead.