

**HOME  
CREDIT**

 **Rakamin**  
Academy

# HOME CREDIT SCORECARD

Galih Setyo Priatbodo



# Table of Contents

01

Problem  
Research

03

Data  
Preprocessing

05

Business  
Recommendation

02

Data Visualizations &  
Business Insight

04

Machine Learning Models  
& Model Evaluation

# Problem Research

## Background

Perusahaan meminta untuk mengeksplor data perusahaan dengan berbagai macam metode statistik dan machine learning untuk membuat prediksi skor kredit.

## Objective

Membuat model prediktif untuk menentukan pelanggan potensial

## Goal

Membangun model untuk memprediksi kemampuan pelanggan untuk melakukan pelunasan pinjaman (credit score)

## Action

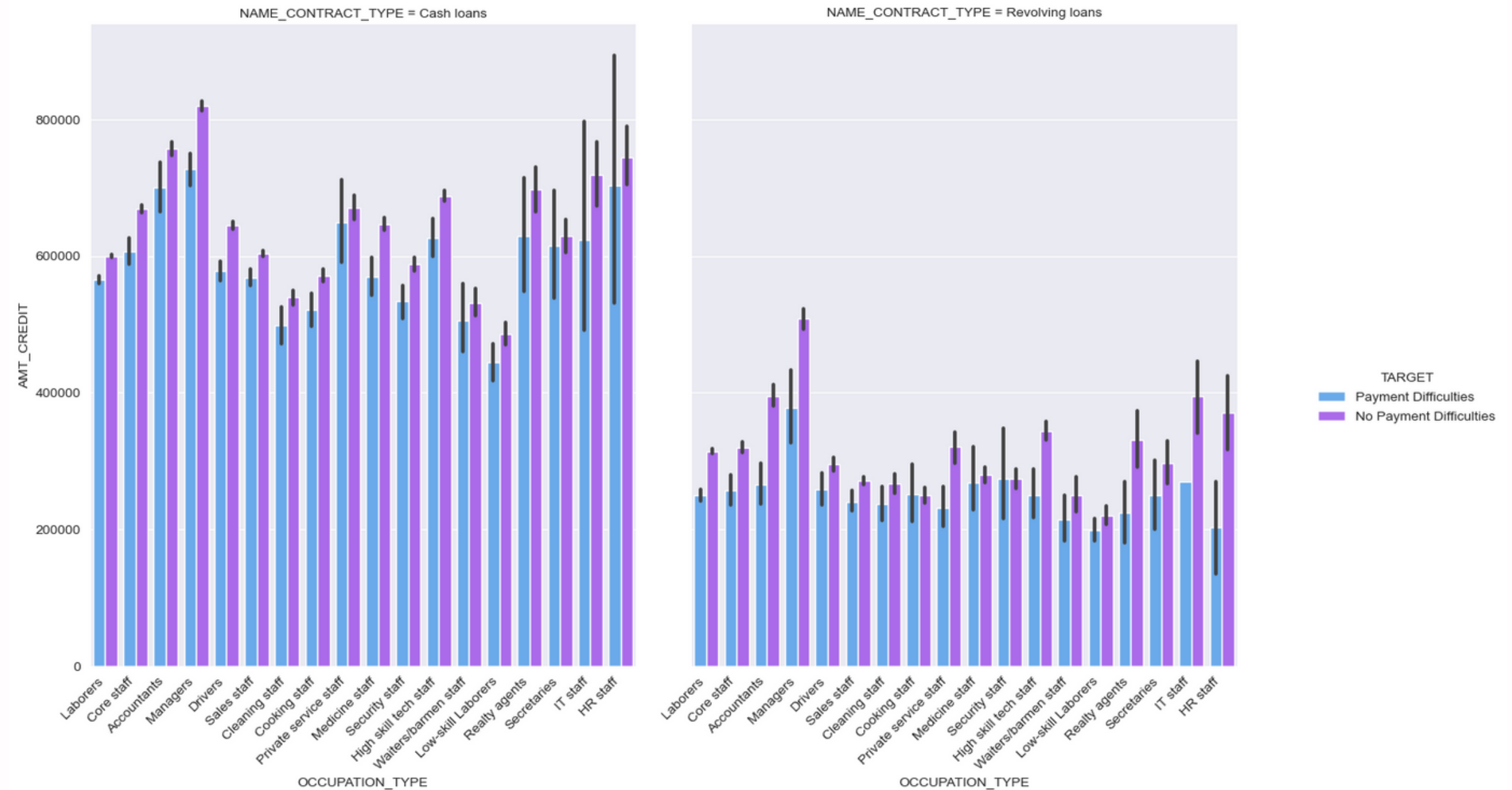
- Melakukan cleaning data
- Membuat visualisasi data
- Membuat model machine learning
- Memberikan rekomendasi

# Flow Analysis

<b>EXPLORATORY DATA ANALYSIS</b> <ul style="list-style-type: none"><li>• Cek Data Duplikat</li><li>• Cek Missing Value</li><li>• Univariate Analysis</li></ul>	<b>DATA PRE-PROCESSING</b> <ul style="list-style-type: none"><li>• Data Cleaning</li><li>• Feature Engineering</li><li>• Handling Imbalanced Data</li><li>• Data Splitting</li></ul>
<b>MODEL BUILDING</b> <ul style="list-style-type: none"><li>• Logistic Regression</li><li>• Gaussian Naive Bayes</li><li>• Decision Tree</li><li>• Random Forest</li><li>• Neural Network</li></ul>	<b>MODEL EVALUATION</b> <ul style="list-style-type: none"><li>• Classification Report</li><li>• Confusion Matrix</li><li>• ROC Score</li></ul>

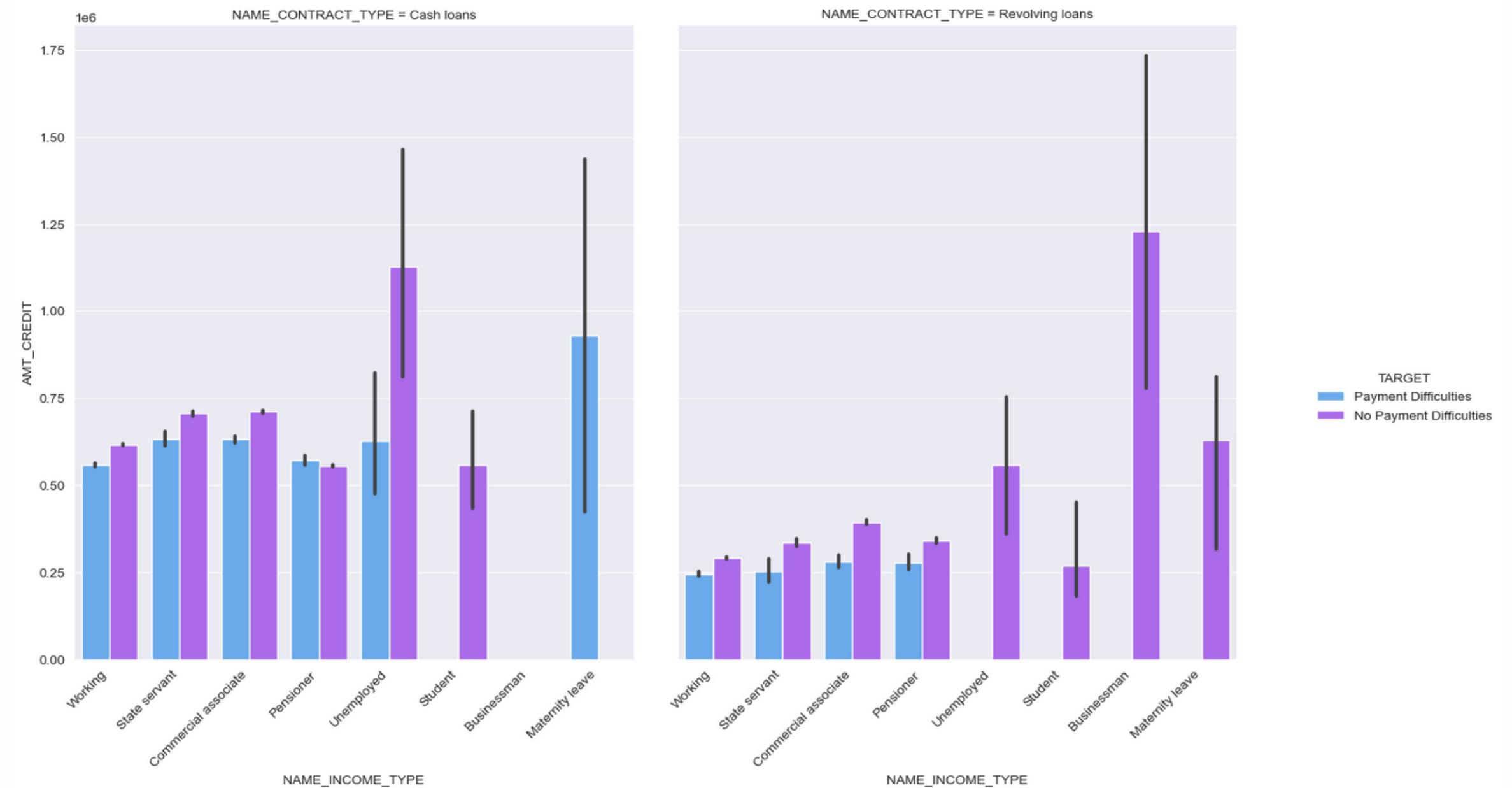
# Data Visualization

Pelanggan dengan jenis pekerjaan akuntan, manajer, HR staff, dan IT Staff memiliki tingkat kesulitan membayar yang **relatif rendah** apabila menggunakan pinjaman bergulir.



# Data Visualization

Pelanggan dengan kategori "**maternity leave**" memiliki kesulitan dalam membayar pinjaman tunai. Namun, ia tidak memiliki kesulitan dalam membayar pinjaman bergulir.

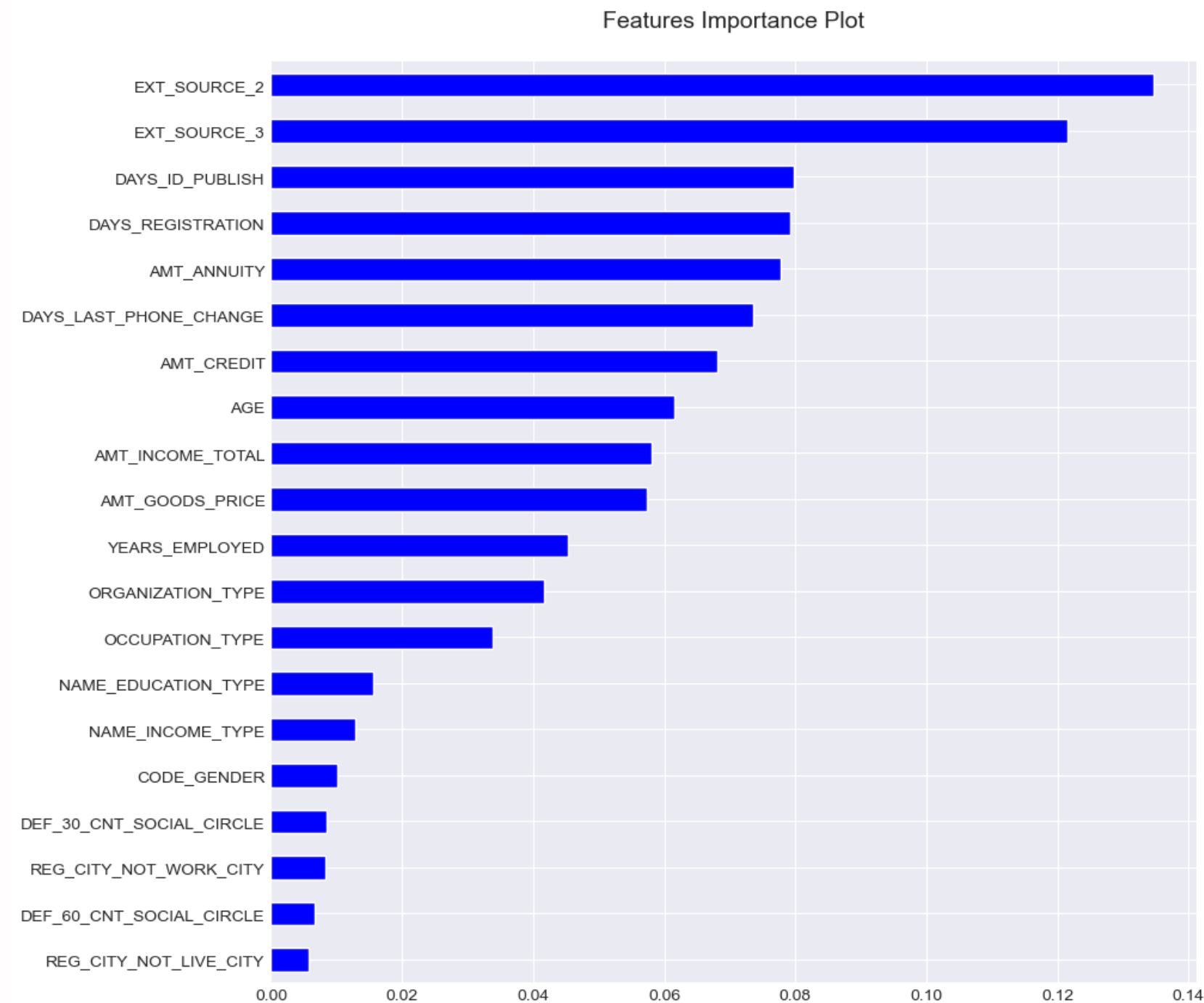


# Machine Learning Models

Algorithm	Training Accuracy Score	Testing Accuracy Score	ROC Score
Logistic Regression	67.28%	67.43%	67.43%
Gaussian Naive Bayes	60.31%	60.49%	60.51%
Decision Tree	100%	91.48%	91.48%
Random Forest	100%	99.64%	99.64%
Neural Network	69.89%	69.26%	69.6%



# Feature Importance



Top 3 important features

- Score from external data source 2
- Score from external data source 3
- Days ID Publish

## Model Selection

Setelah membandingkan 5 machine learning model, model machine learning terbaik dengan tingkat akurasi tertinggi adalah Random Forest.



# Business Recommendation

Memprioritaskan pelanggan dengan rentang umur **30 – 45 tahun** yang sudah memiliki pekerjaan, karena mereka cenderung memiliki tingkat kredit yang tinggi dan situasi keuangan yang lebih stabil. Dengan memanfaatkan insentif atau promosi untuk meningkatkan kredit.

Sangat penting untuk menjaga hubungan baik dengan **businessman** yang memiliki **revolving loans** dan riwayat pembayaran yang positif. Hal itu karena mereka dapat menjadi pelanggan yang berharga yang mungkin dapat tertarik dengan produk lain.

Penting untuk memperhatikan pelanggan yang mengambil jenis pinjaman tunai dengan tingkat kesulitan bayar yang tinggi seperti **maternity leave dan pensioner** untuk kemudian mengevaluasi situasi keuangan mereka dan menawarkan kembali produk yang sesuai.

**HOME  
CREDIT**

 **Rakamin**  
Academy

# THANK YOU

[https://github.com/galihhsetyo/home\\_credit\\_scorecard](https://github.com/galihhsetyo/home_credit_scorecard)