

Business Case Study and Report

HOME CREDIT SCORE CARD PREDICTION

Presented by:

MAKMURRY AKBAR

**HOME
CREDIT**
Kamu Bisa!



PROJECT BACKGROUND

Home Credit saat ini sedang menggunakan berbagai macam metode statistik dan Machine Learning untuk membuat prediksi skor kredit. Kami diminta untuk membuka potensi maksimal dari data yang dimiliki Home Credit. Dengan melakukannya, harapannya dapat memastikan pelanggan yang mampu melakukan pelunasan tidak ditolak ketika melakukan pengajuan pinjaman, dan pinjaman dapat diberikan dengan principal, maturity, dan repayment calendar yang akan memotivasi pelanggan untuk sukses

**HOME
CREDIT**

Objectives

Membuat prediksi kategori calon pelamar yang berpotensi tidak kesulitan membayar pinjaman.

Menggali wawasan dari data yang bisa memberi manfaat untuk keputusan bisnis dimasa mendatang

Action

- Lakukan pembersihan data, dan visualisasi untuk wawasan bisnis.
- Membangun model dengan algoritma pembelajaran mesin.
- Prediksi kemampuan pembayaran klien pada pengujian data aplikasi dengan model pembelajaran mesin terbaik.
- Memberikan rekomendasi bagi perusahaan untuk meningkatkan keberhasilan klien dalam mengajukan pinjaman.

DATASET

Dataset yang digunakan pada project ini adalah data application sebagai data utama yang terdiri dari data train dan data test

	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREI
0	100002	1	Cash loans	M	N	Y	0	202500.0	40659
1	100003	0	Cash loans	F	N	N	0	270000.0	129350
2	100004	0	Revolving loans	M	Y	Y	0	67500.0	13500
3	100006	0	Cash loans	F	N	Y	0	135000.0	31268
4	100007	0	Cash loans	M	N	Y	0	121500.0	51300

Data Untuk pelatihan memiliki 307511 baris dan 122 fitur (kolom) termasuk kolom "TARGET" (label yang ingin kita prediksi).

Jumlah baris dan kolom: (48744, 121)

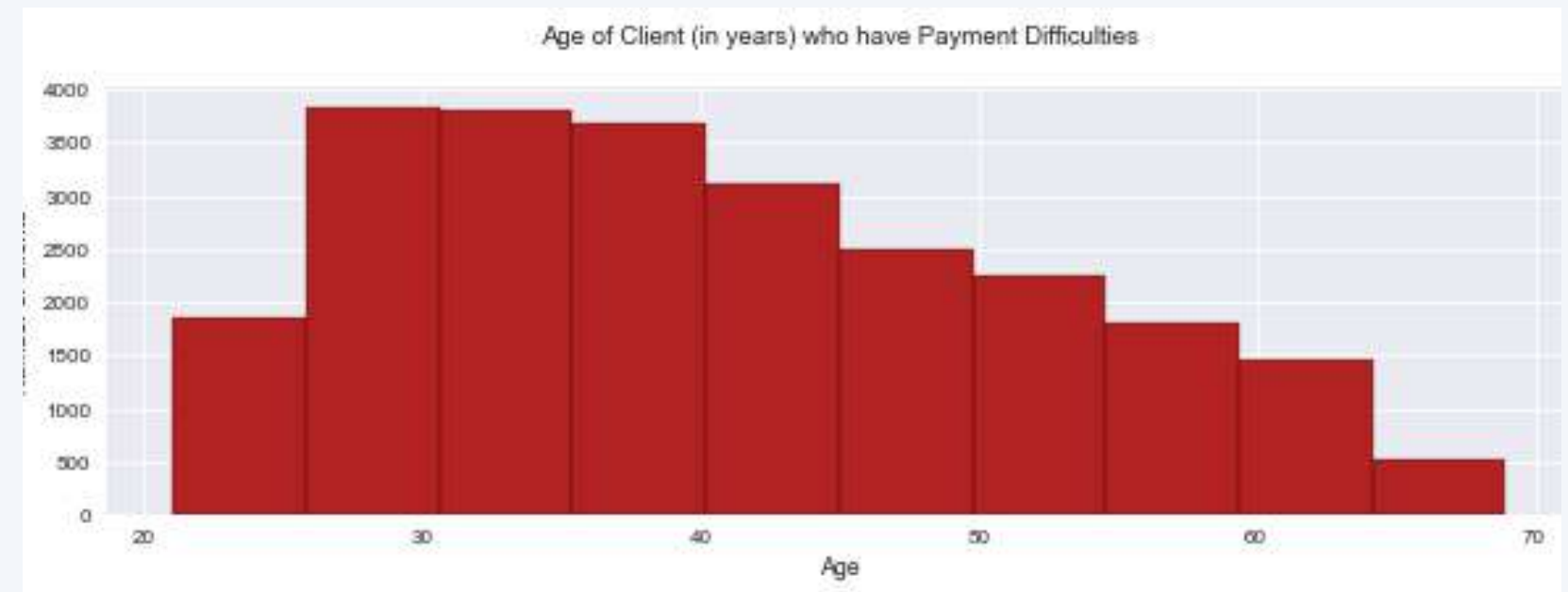
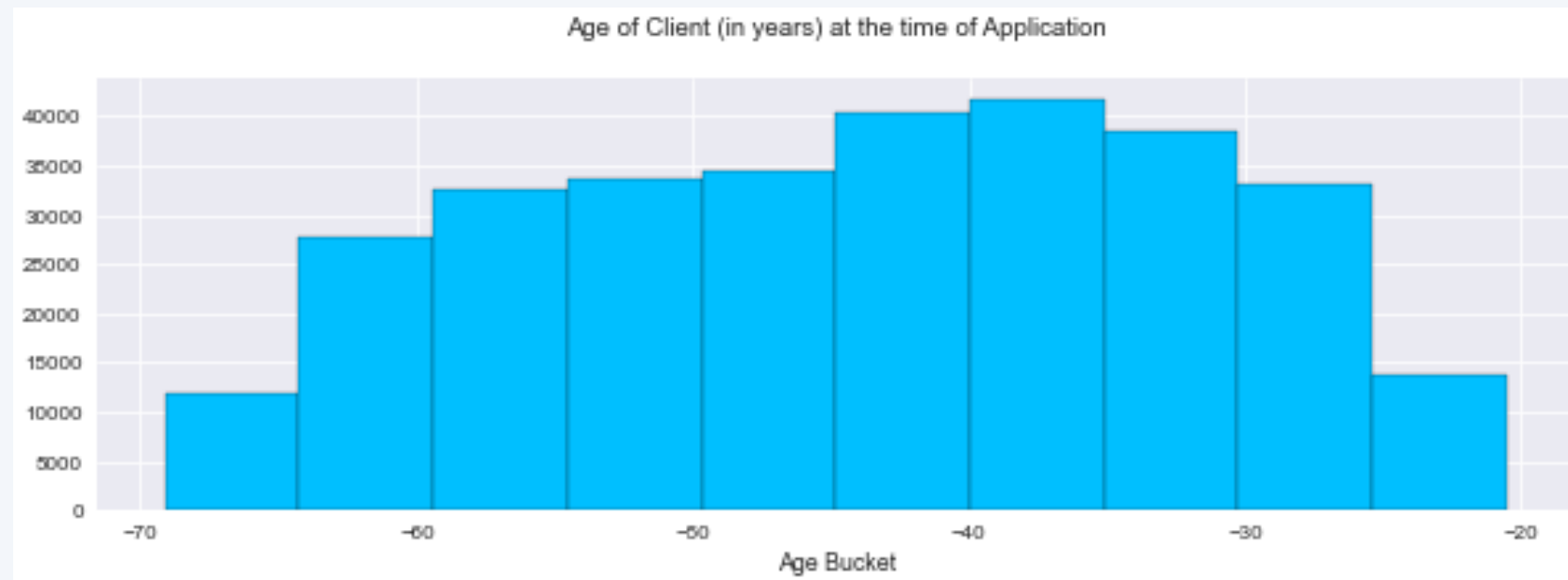
[3]:

	SK_ID_CURR	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_
0	100001	Cash loans	F	N	Y	0	135000.0	568800.0	
1	100005	Cash loans	M	N	Y	0	99000.0	222768.0	
2	100013	Cash loans	M	Y	Y	0	202500.0	663264.0	
3	100028	Cash loans	F	N	Y	2	315000.0	1575000.0	
4	100038	Cash loans	M	Y	N	1	180000.0	625500.0	

Bisa dilihat dari kolom SK_ID_CURR bahwa data test merupakan sampel dari data train yang diambil dengan cara pengacakan silang dan untuk pengujian kolom 'TARGET' sengaja dihilangkan, oleh karena itu hanya Home Credit yang tahu hasil tes untuk mencocokkannya dengan hasil prediksi kita

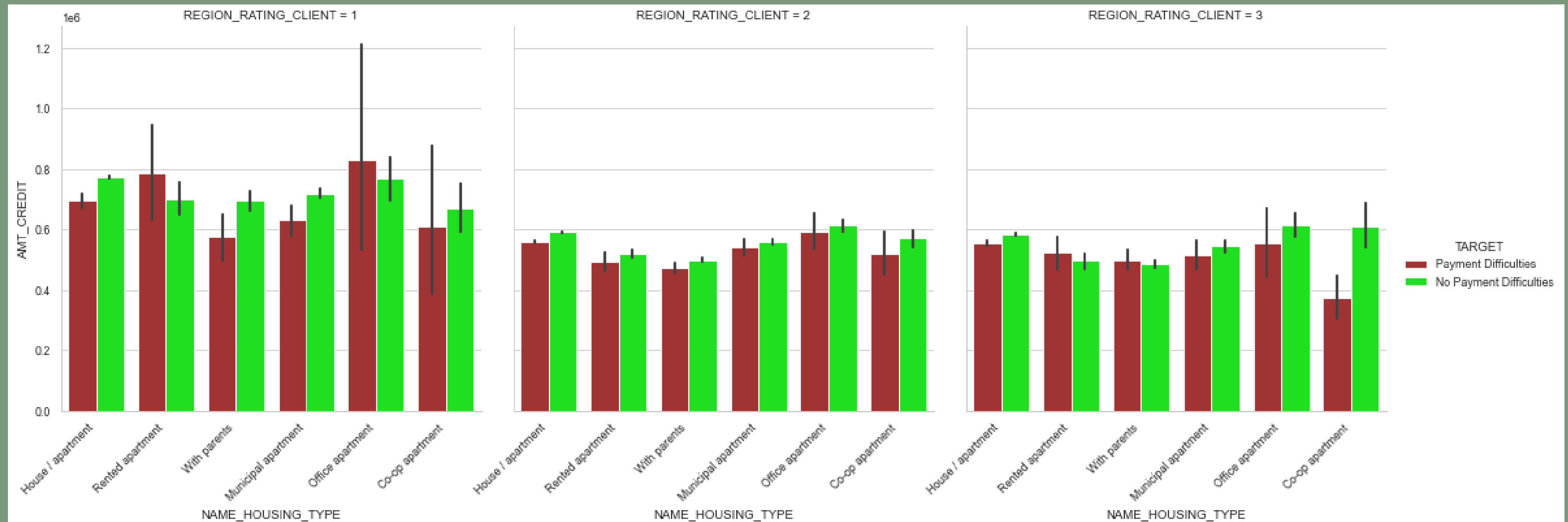
**HOME
CREDIT**

Exploratory Data Analysis (EDA)



Jumlah nasabah yang mengajukan pinjaman berdasarkan tanggal aplikasi paling banyak berada pada kisaran 30-40 tahun, diikuti nasabah pada kisaran 40-45 tahun. Sementara itu, jumlah pelamar untuk klien berusia <25 tahun atau usia >65 tahun sangat sedikit. Klien yang mengalami kesulitan pembayaran adalah klien dengan rentang usia 35-45 tahun. Sedangkan klien yang berusia 60-70 tidak terlalu banyak yang kesulitan

Exploratory Data Analysis (EDA)



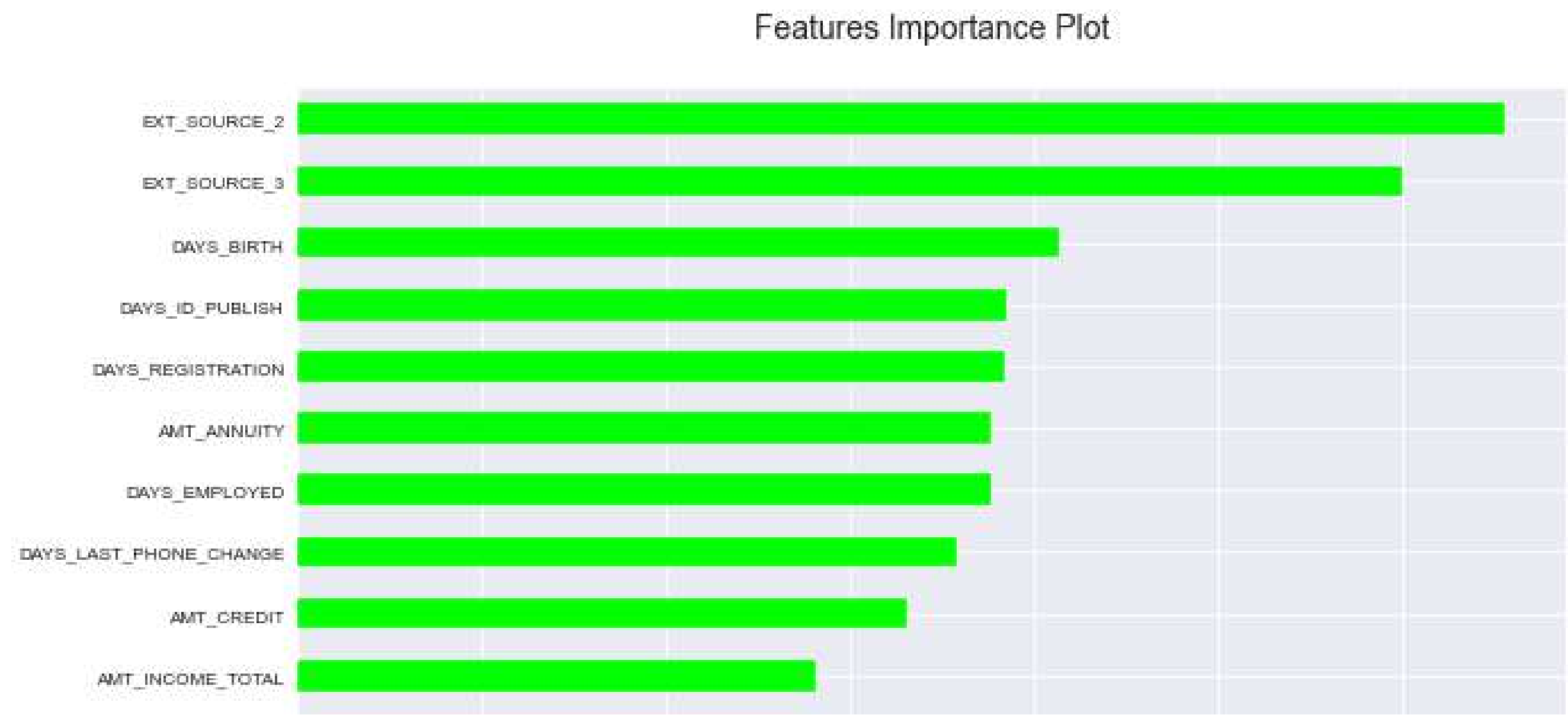
Klien yang tinggal di wilayah Rating 3 Sebenarnya lebih banyak yang tidak kesulitan dalam pelunasan dibanding dengan wilayah rating 2 dan 3 namun juga sebanding dengan jumlah klien yang tidak mampu melunasi. Hal ini mengindikasikan bahwa wilayah rating 1 tinggi akan kebutuhan kredit dan yang perlu dikaji lebih lanjut adalah penyebab klien yang sulit membayar di wilayah rating 1 ini.

MACHINE LEARNING MODEL

Models	Training Accuracy Score	Testing Accuracy Score	ROC Score
Random Forest	1.000000	0.996500	0.996500
Decision Tree	1.000000	0.882600	0.882600
Neural Network	0.700100	0.694800	0.694800
Logistic Regression	0.671600	0.672900	0.672900
Gaussian Naive Bayes	0.602400	0.603900	0.604000

setelah dilakukan metrik evaluasi dengan Akurasi score , Confused matric dan ROC-AUC maka model Random Forest dipilih sebagai model terbaik untuk memprediksi kemampuan pembayaran klien.

FEATURES IMPORTANCE



Plot di atas menginformasikan 10 fitur penting teratas yang dapat menjadi acuan minimal dalam pemeriksaan kelayakan calon klien sebelum dianalisa lebih dalam.

BUSINESS Recommendations

Lakukan Promosi dengan menyertakan edukasi keuangan untuk orang-orang yang berusia 20-25 tahun karena calon nasabah di usia ini memiliki potensi yang baik jika mereka mendapat edukasi keuangan dan jenis pinjaman yang tepat.

Lakukan survey mengenai fenomena wilayah rating 1 lebih banyak jumlah pelamarnya dibanding wilayah rating lain. Jika kita mengetahui penyebab hampir 50 % klien mengalami kesulitan pembayaran di wilayah itu, kita bisa meningkatkan calon nasabah yang potensial.

**HOME
CREDIT**

Terima Kasih

GitHub

<https://github.com/Makmurry/Makmurry/blob/main/HCI-TASK5.ipynb>