



MODEL SKOR KREDIT HOME CREDIT

FAUZAN KAMIL



BUSINESS UNDERSTANDING

BACKGROUND

Home Credit menggunakan berbagai metode statistik dan Machine Learning untuk membuat prediksi skor kredit. Perusahaan ingin mengoptimalkan potensi maksimum data yang mereka miliki. Hal ini bertujuan untuk memastikan bahwa klien yang mampu membayar dapat diterima dan pinjaman diberikan dengan prinsipal, jangka waktu, dan jadwal pembayaran yang akan memberdayakan klien mereka untuk berhasil.

TUJUAN

1. Mengidentifikasi pelanggan dengan karakteristik calon klien yang mengalami kesulitan dalam melunasi pinjaman.
2. Memprediksi kemampuan pelanggan dalam melunasi pinjaman.

TINDAKAN

1. Melakukan pembersihan data dan visualisasi untuk mendapatkan wawasan bisnis yang lebih baik.
2. Membangun model menggunakan algoritma machine learning.
3. Memprediksi kemampuan pelanggan dalam melunasi pinjaman menggunakan data aplikasi uji.
4. Memberikan rekomendasi kepada perusahaan untuk meningkatkan kesuksesan klien dalam mengajukan pinjaman.

DATA UNDERSTANDING

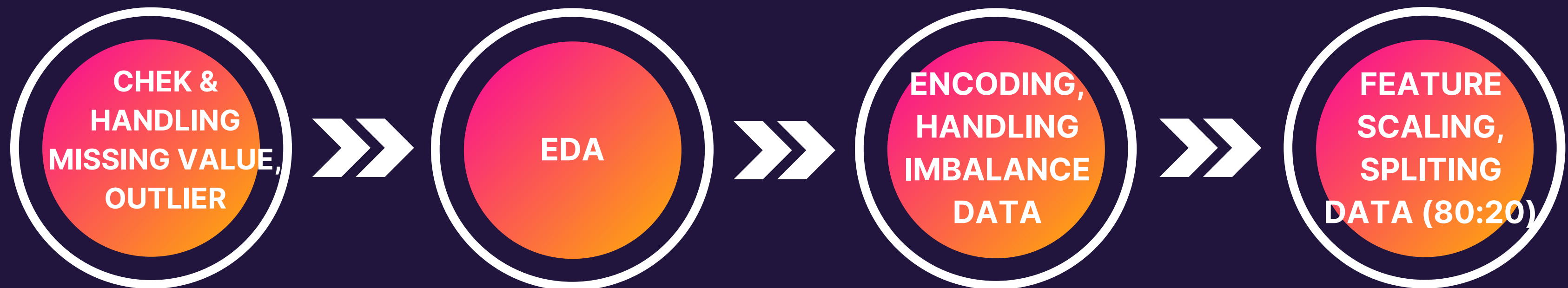
Untuk menyelesaikan masalah ini saya menggunakan Dataset `application_train|test.csv` Tabel ini dibagi menjadi dua file terpisah yaitu file Data Latih (dengan TARGET) dan file Data Uji (tanpa TARGET). File Data Latih berisi informasi tentang pinjaman beserta variabel TARGET yang menunjukkan apakah pinjaman tersebut memiliki risiko atau tidak. Sedangkan file Data



**307511 BARIS DAN
122 KOLOM**

**TIPE DATA FLOAT SEBANYAK
65 DAN INTEGER 41 DAN
OBJECT16**

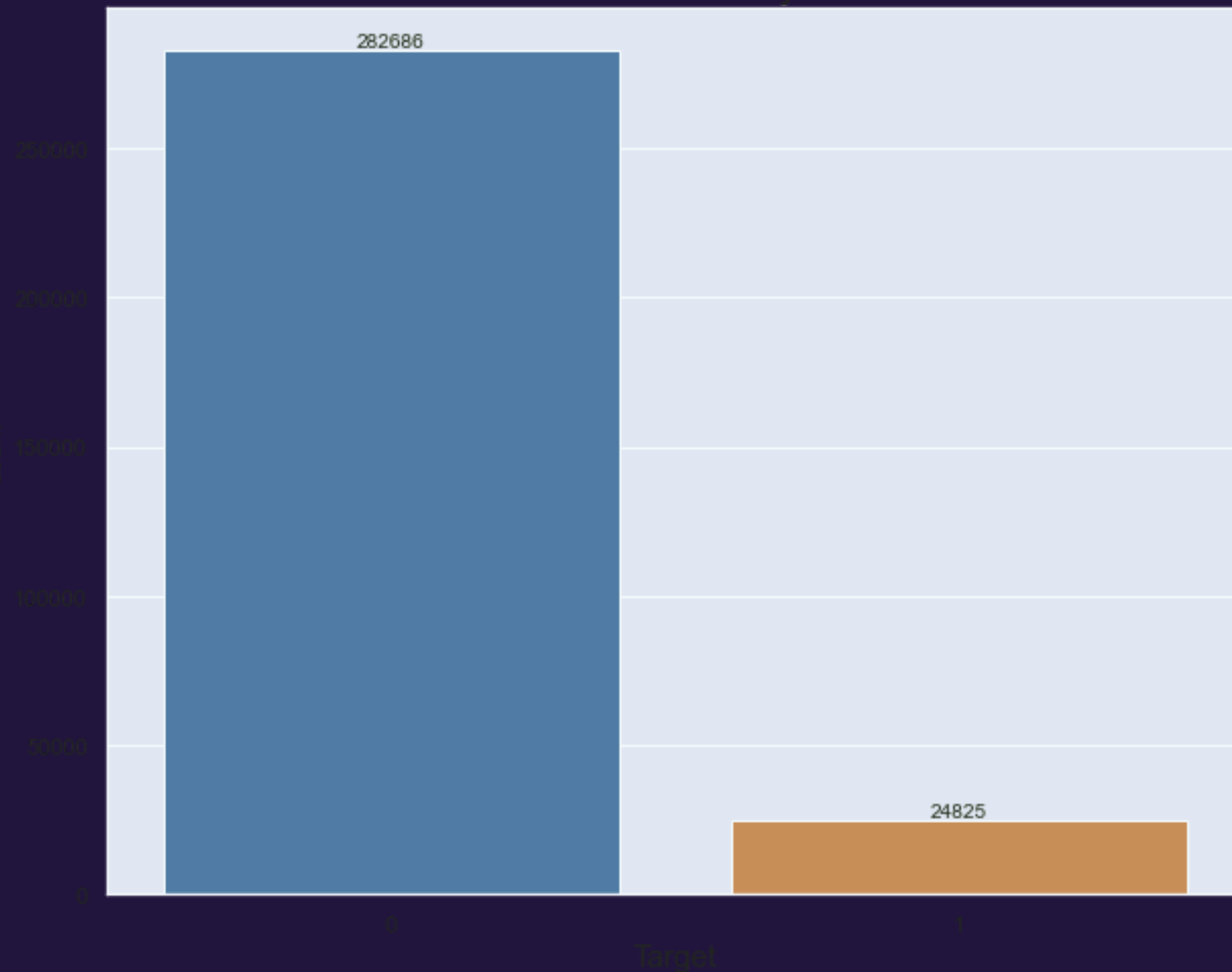
DATA PREPROCESSING



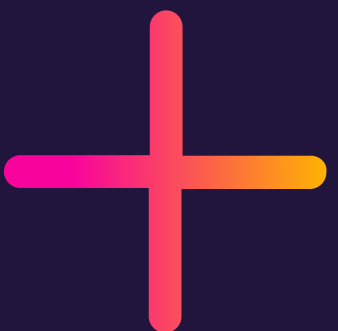


EXPLORATORY DATA ANALYSIS

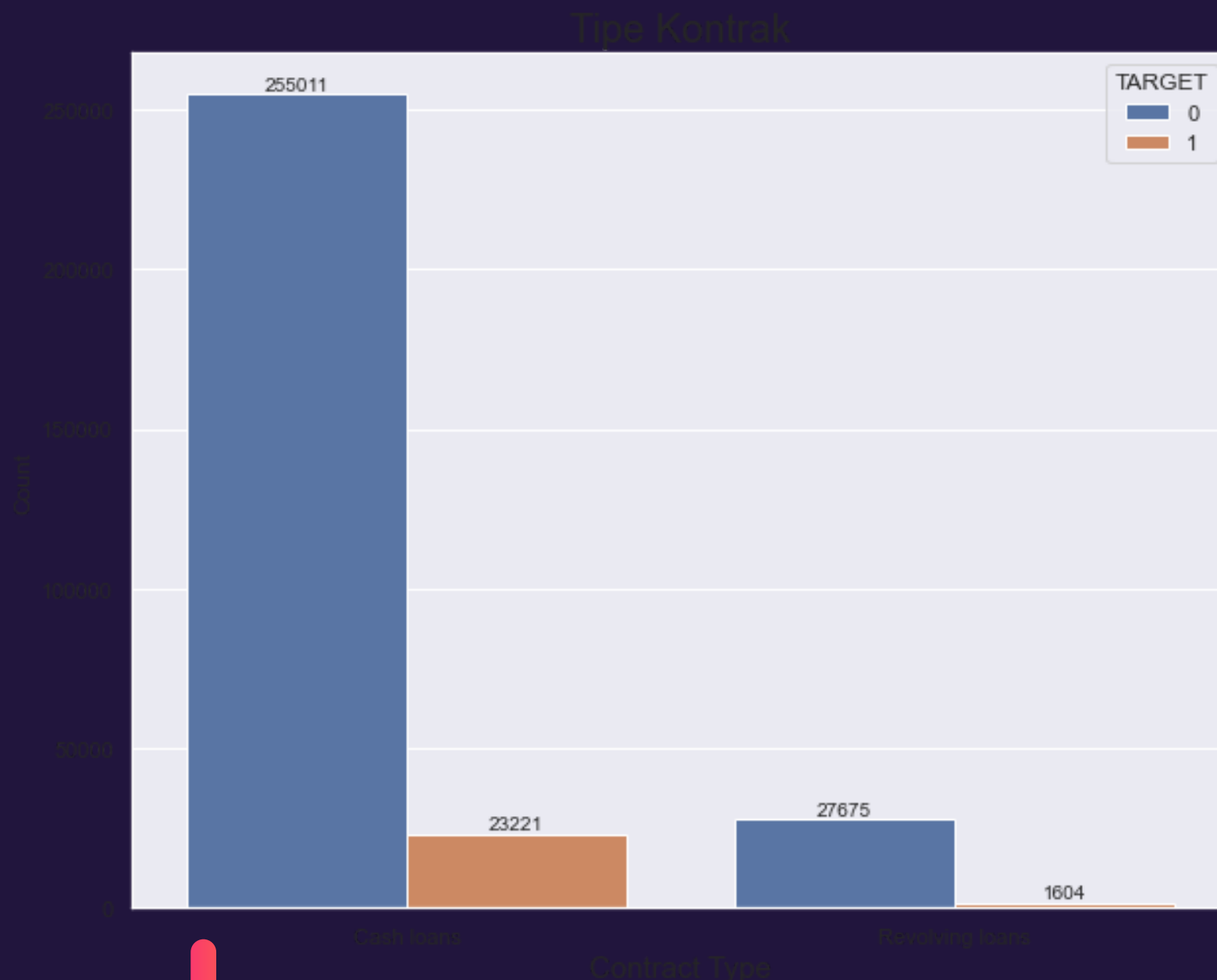
Kesulitan Pembayaran



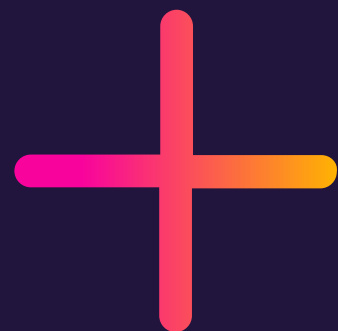
- Terdapat 282.686 klien yang termasuk dalam kategori 0 pada variabel Target. Kategori 0 menunjukkan bahwa klien tersebut tidak mengalami kesulitan dalam pembayaran pinjaman. Mereka melakukan pembayaran tepat waktu pada setiap angsuran pinjaman yang pertama.
- Terdapat 24.825 klien yang termasuk dalam kategori 1 pada variabel Target. Kategori 1 menunjukkan bahwa klien tersebut mengalami kesulitan dalam pembayaran pinjaman. Mereka memiliki keterlambatan pembayaran lebih dari X hari pada setidaknya satu dari Y angsuran pertama pinjaman dalam contoh data kita.



EXPLORATORY DATA ANALYSIS



- Terdapat 255.011 pinjaman yang memiliki tipe kontrak "0" (cash) dan kategori target "0". Ini berarti mayoritas pinjaman dengan tipe kontrak tunai memiliki klien yang tidak mengalami kesulitan dalam pembayaran pinjaman.
- Terdapat 27.675 pinjaman yang memiliki tipe kontrak "1" (revolving) dan kategori target "0". Ini menunjukkan bahwa mayoritas pinjaman dengan tipe kontrak revolving juga memiliki klien yang tidak mengalami kesulitan dalam pembayaran pinjaman.
- Terdapat 23.221 pinjaman yang memiliki tipe kontrak "0" (cash) dan kategori target "1". Ini menunjukkan bahwa sejumlah pinjaman dengan tipe kontrak tunai mengalami kesulitan dalam pembayaran pinjaman.
- Terdapat 1.604 pinjaman yang memiliki tipe kontrak "1" (revolving) dan kategori target "1". Ini menunjukkan bahwa sejumlah pinjaman dengan tipe kontrak revolving juga mengalami kesulitan dalam pembayaran pinjaman.



MODELING

Model	Mean Absolute Error	Accuracy	ROC AUC
Logistic Regression	0.169410	0.830590	0.830597
Random Forest	0.047588	0.952412	0.952462
XGBoost	0.045598	0.954402	0.954448
Decision Tree	0.099978	0.900022	0.900010
Gradient Boosting	0.055998	0.944002	0.944054

MODEL TERBAIK

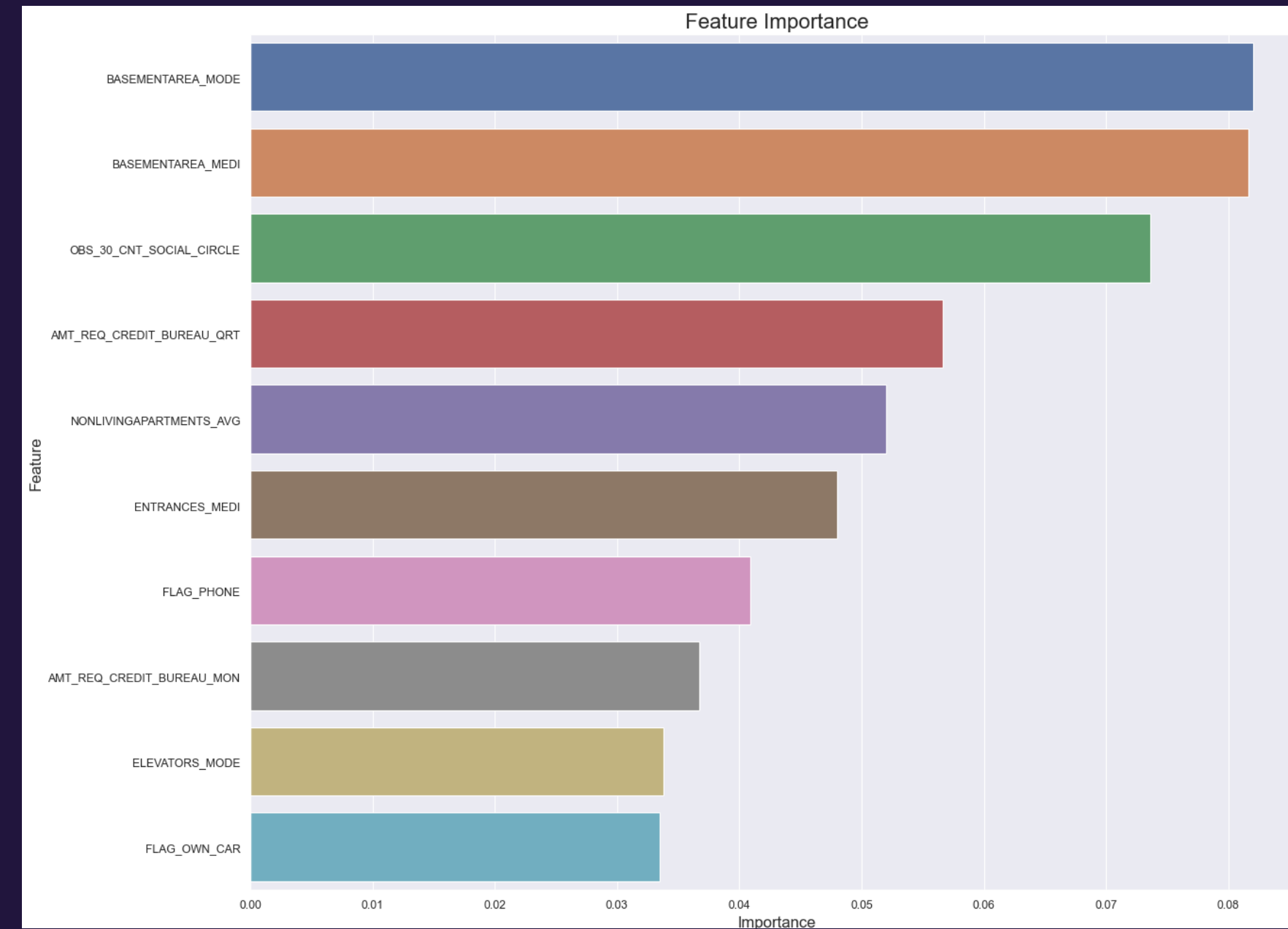
Berdasarkan tabel disamping, model dengan performa terbaik dapat ditentukan dengan membandingkan nilai-nilai metrik Mean Absolute Error (MAE), Accuracy, dan ROC AUC. Semakin rendah nilai MAE dan semakin tinggi nilai Accuracy serta ROC AUC, semakin baik performa model tersebut. Model **XGBoost** memiliki nilai **MAE yang paling rendah** dan nilai **Accuracy** serta **ROC AUC** yang paling tinggi dibandingkan dengan model-model lainnya. Oleh karena itu, model XGBoost dapat dipilih sebagai model **terbaik**.



MODELING (FEATURE IMPORTANCE)

Berdasarkan analisis fitur importance, terdapat 10 fitur yang memiliki kontribusi penting dalam memprediksi target variabel. Berikut adalah narasi mengenai 10 fitur importance tersebut:

1. BASEMENTAREA_MODE: Luas area basement dengan importance 0.082021.
2. BASEMENTAREA_MEDI: Luas area basement dengan metrik median, importance 0.081614.
3. OBS_30_CNT_SOCIAL_CIRCLE: Jumlah interaksi sosial dalam 30 hari terakhir dengan importance 0.073601.
4. AMT_REQ_CREDIT_BUREAU_QRT: Jumlah permintaan kredit dari kreditur pada kuartal tertentu dengan importance 0.056666.
5. NONLIVINGAPARTMENTS_AVG: Rata-rata luas apartemen non-tinggal dengan importance 0.051993.
6. ENTRANCES_MEDI: Tingkat keberhasilan pintu masuk dengan importance 0.048012.
7. FLAG_PHONE: Kepemilikan telepon dengan importance 0.040936.
8. AMT_REQ_CREDIT_BUREAU_MON: Jumlah permintaan kredit dari kreditur dalam sebulan dengan importance 0.036697.
9. ELEVATORS_MODE: Tingkat keberhasilan elevator dengan importance 0.033803.
10. FLAG_OWN_CAR: Kepemilikan mobil dengan importance 0.033468.



RECOMMENDATION

1

Fokus pada klien dengan kesulitan pembayaran pinjaman (kategori 1) dan berikan dukungan finansial.

2

Evaluasi faktor yang berkontribusi pada kesulitan pembayaran pinjaman pada tipe kontrak tunai dan revolving.

3

Perbaiki persyaratan dan proses pemberian pinjaman untuk mengurangi risiko pembayaran yang tidak lancar pada tipe kontrak tunai.

4

Pahami karakteristik klien dengan kesulitan pembayaran pada tipe kontrak revolving.

5

Lakukan penelitian dan analisis lebih lanjut untuk mendapatkan wawasan dalam mengelola risiko pembayaran pinjaman.



THANK YOU

Link Github:

https://github.com/Fauzan-Kamil/vix_Home_Credit_Rakamin

