



## Home Credit Score Card Model

By: Tyovendi Arisandy



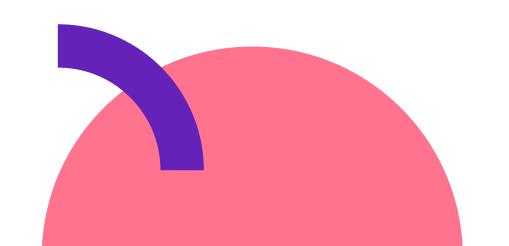
### Table of Contents

- Business Problem
- Data Preprocessing
- Exploratory Data Analysis (EDA)
- Machine Learning Modeling
- Recomendation



#### **Business Context**

Banyak orang membutuhkan kredit dari lembaga pinjaman karena tidak memiliki cukup dana untuk memenuhi kebutuhan atau untuk mengembangkan bisnis. Namun, terkadang sulit untuk mendapatkan kredit dari lembaga pinjaman karena tidak memiliki jaminan atau karena tingkat bunga yang tinggi. Oleh karena itu, Home Credit hadir untuk memberikan layanan kredit yang mudah, aman, dan tentu memberikan manfaat bagi klien.

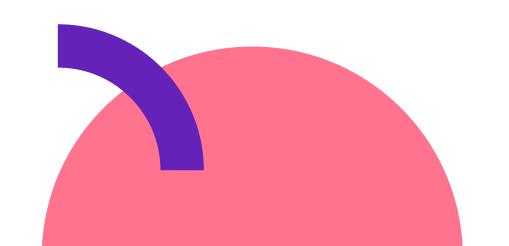


#### **Problems**

Terkadang ada orang yang kesulitan untuk membayar kembali kredit yang telah diambil. Hal ini dapat terjadi karena beberapa faktor, seperti terjadinya masalah keuangan yang tak terduga, mengalami kesulitan ekonomi, atau bahkan karena adanya perubahan dalam situasi keuangan pribadi. Ketidakmampuan untuk membayar kembali kredit dapat merugikan pihak kreditur dalam jangka panjang, terutama jika terdapat banyak orang yang mengalami masalah tersebut.

#### Objective/Goals

Dengan bantuan Machine Learning (ML) diharapkan pihak pemberi layanan kredit dapat dengan mudah menentukan klien yang sanggup membayar pinjaman dan klien yang tidak sanggup membayar (defaulted). memodelkan algoritma ML tertentu dan memilih dengan performa yang baik akan menjadi prioritas perusahaan untuk menanggulangi kerugian jangka panjang akibat kredit bermasalah

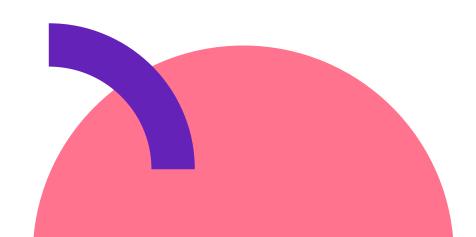


#### **Analytical Approach**

- menganalisis data untuk dapat menemukan pola dari fitur-fitur yang ada, yang membedakan klien yang cenderung akan kesulitan membayar kredit dan yang tidak.
- Selanjutnya, kita akan membangun suatu model klasifikasi yang akan membantu perusahaan untuk dapat menyediakan 'tool' prediksi klien yang kesulitan membayar kredit dan tidak

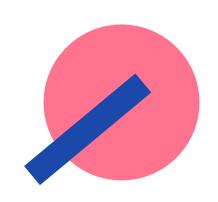
#### **Evaluation Metric**

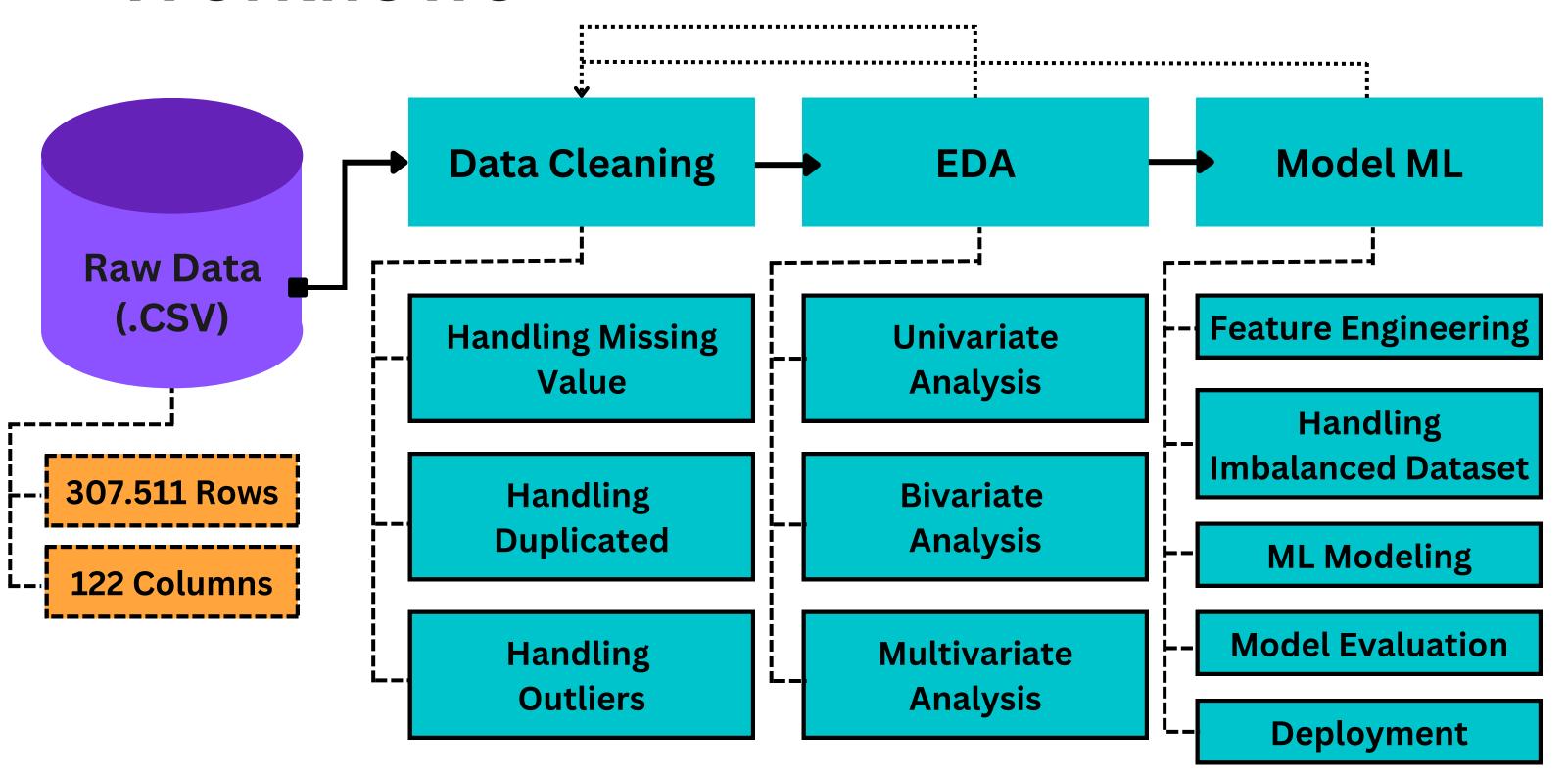
- Mengurangi <u>Type Error I</u> (fakta bahwa klien kesulitan membayar credit, tetapi ML memprediksi tidak kesulitan membayar)
- Meningkatkan porbabilitas <u>True Positif</u>
   dan <u>True Negatif</u>
- Menggunakan ROC-AUC untuk menilai performa ML



### 2. Data Preprocessing

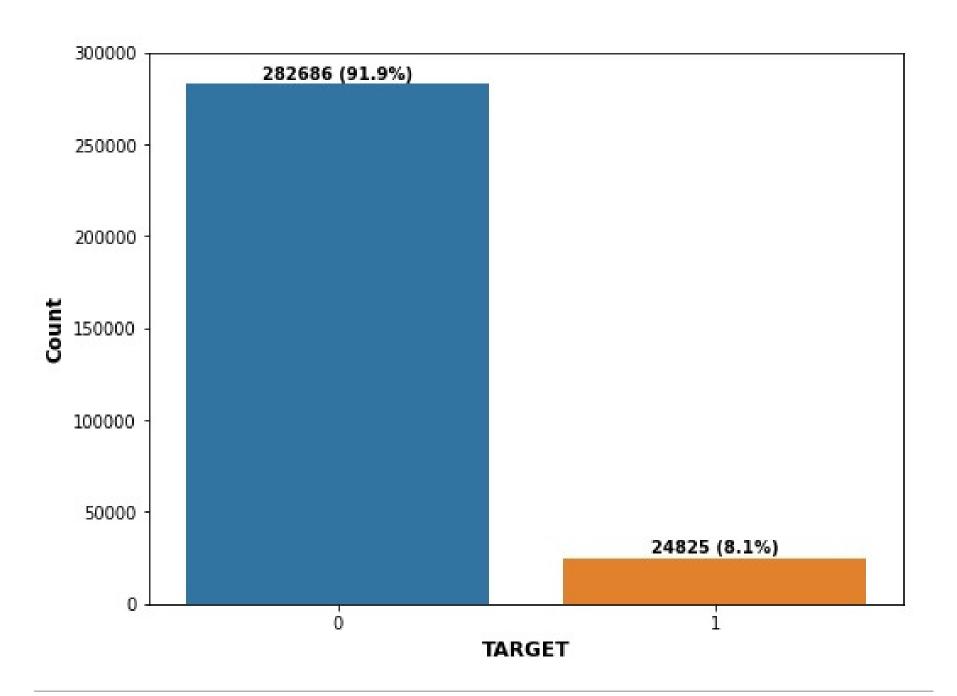
## Step by Step Data Science Workflows





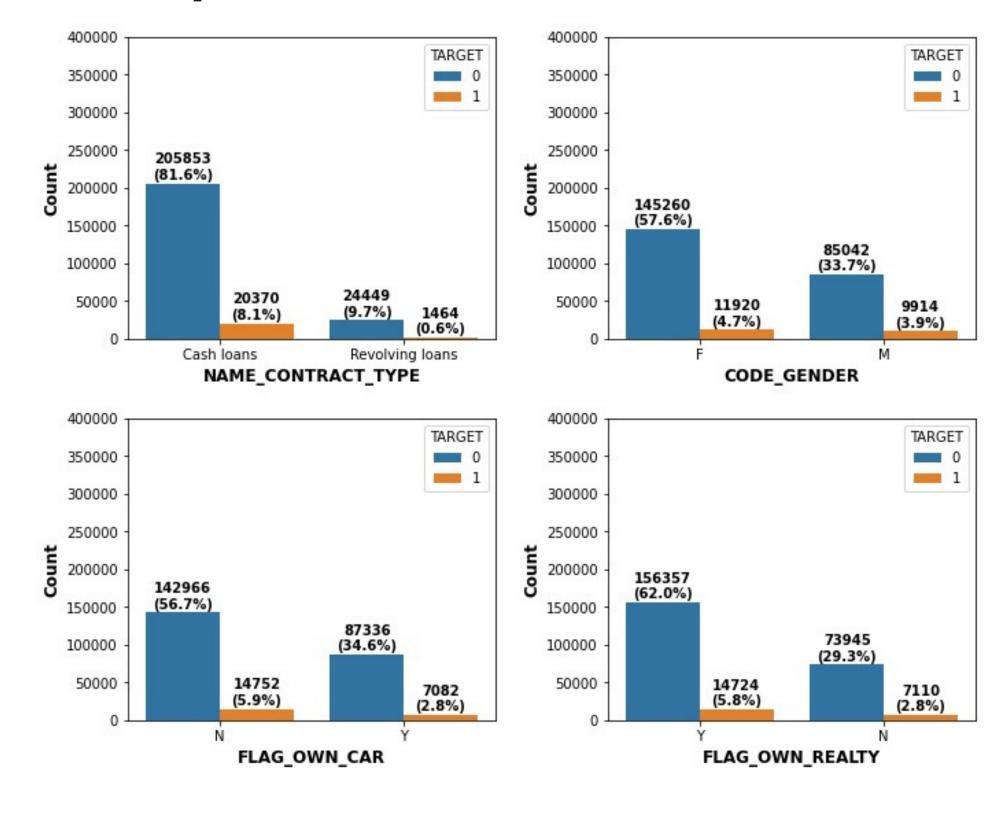
# 3. Exploratory Data Analysis (EDA)

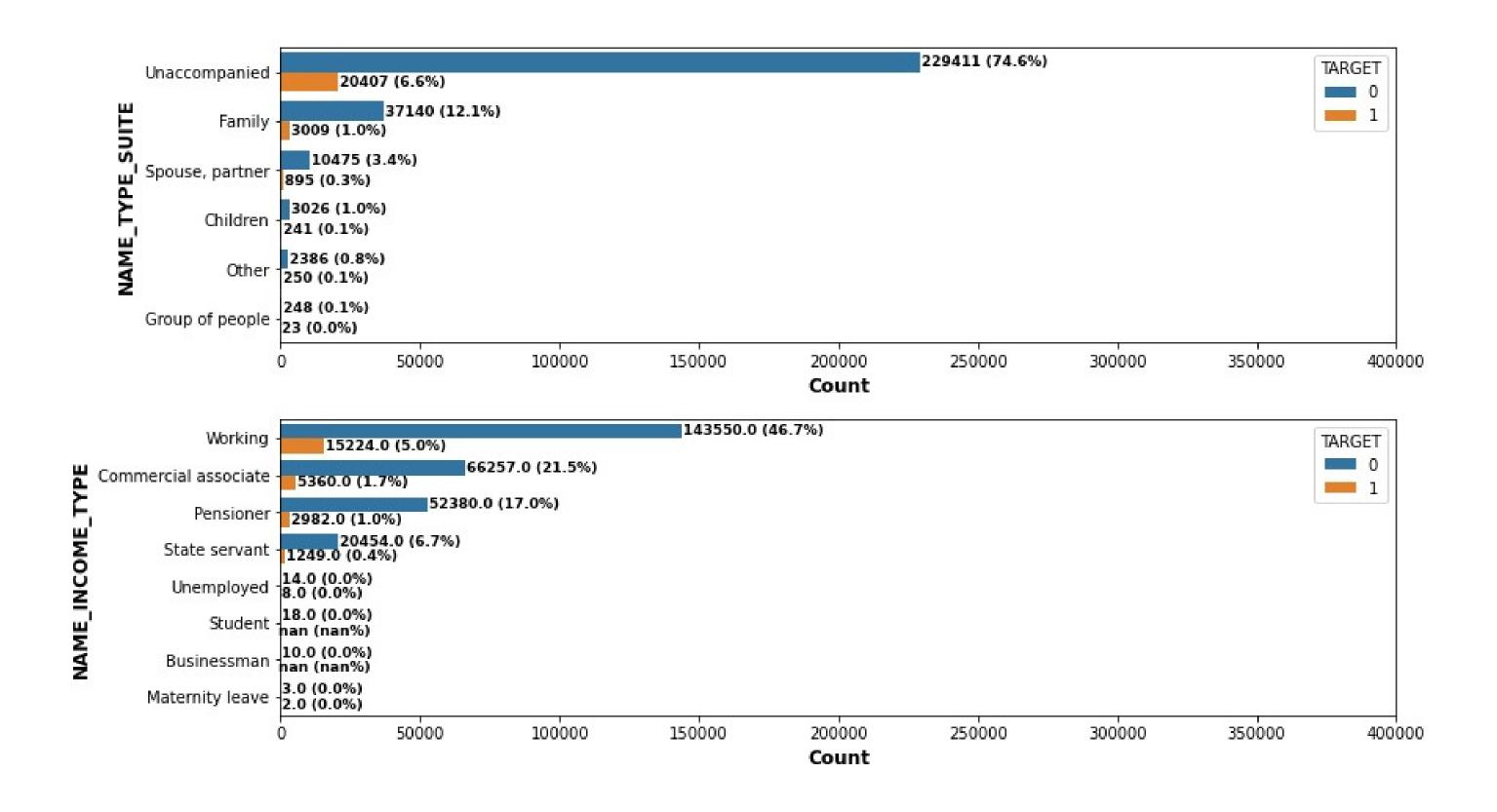
#### **Target Analysis**

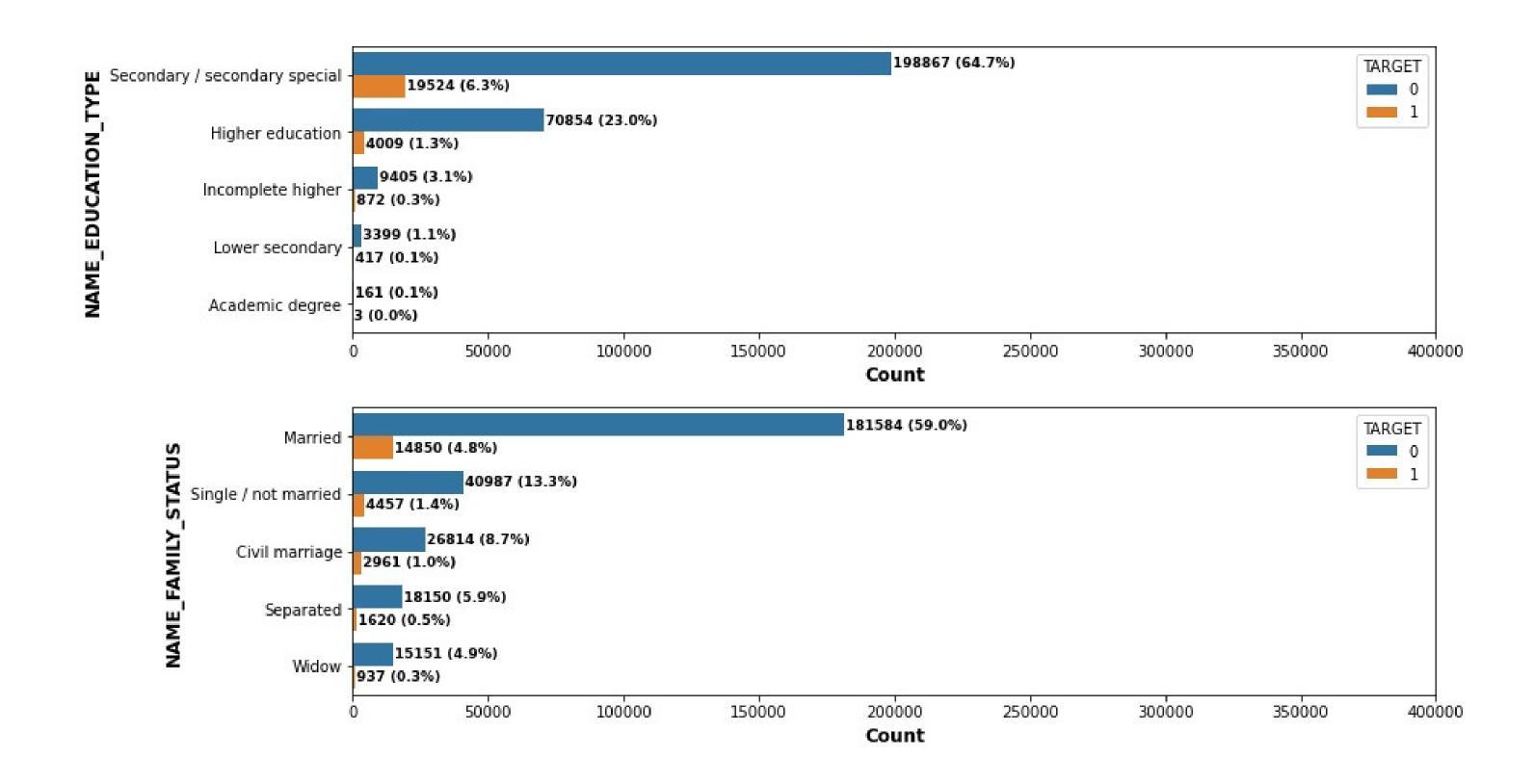


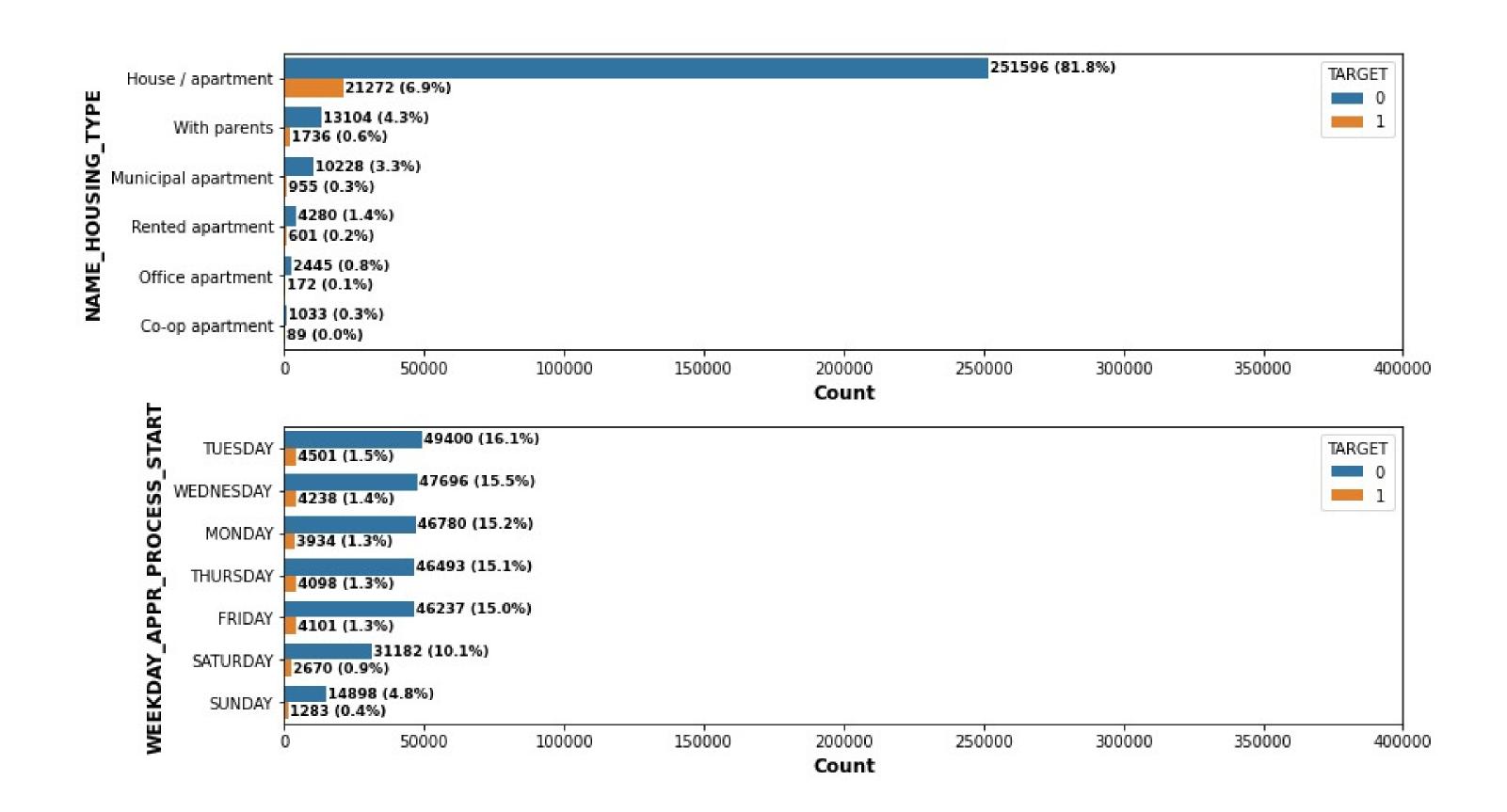
- 0 = Klien tidak mengalami kesulitan membayar pinjaman, 1 = Klien mengalami kesulitan membayar pinjaman.
- Terdapat 24825 klien atau sebesar 8.1% yang mengalami kesulitan pembayaran pinjaman (memiliki keterlambatan pembayaran tempo/hari jatuh berdasarkan telah yang ditentukan setidaknya pinjaman satu dari pinjaman yang diajukan).
- Sedangkan sisanya terdapat 282686 klien atau sebesar 91.9% yang tidak mengalami kesulitan pembayaran pinjaman.
- Dapat diartikan bahwa lebih banyak jumlah klien yang tidak kesulitan membayar pinjaman/lancar dalam membayar kredit. sedangkan klien yang kesulitan membayar kredit jumlahnya lebih sedikit
- Perbandingannya sekitar ±90:10 (data imbalance)

#### Features Analysis







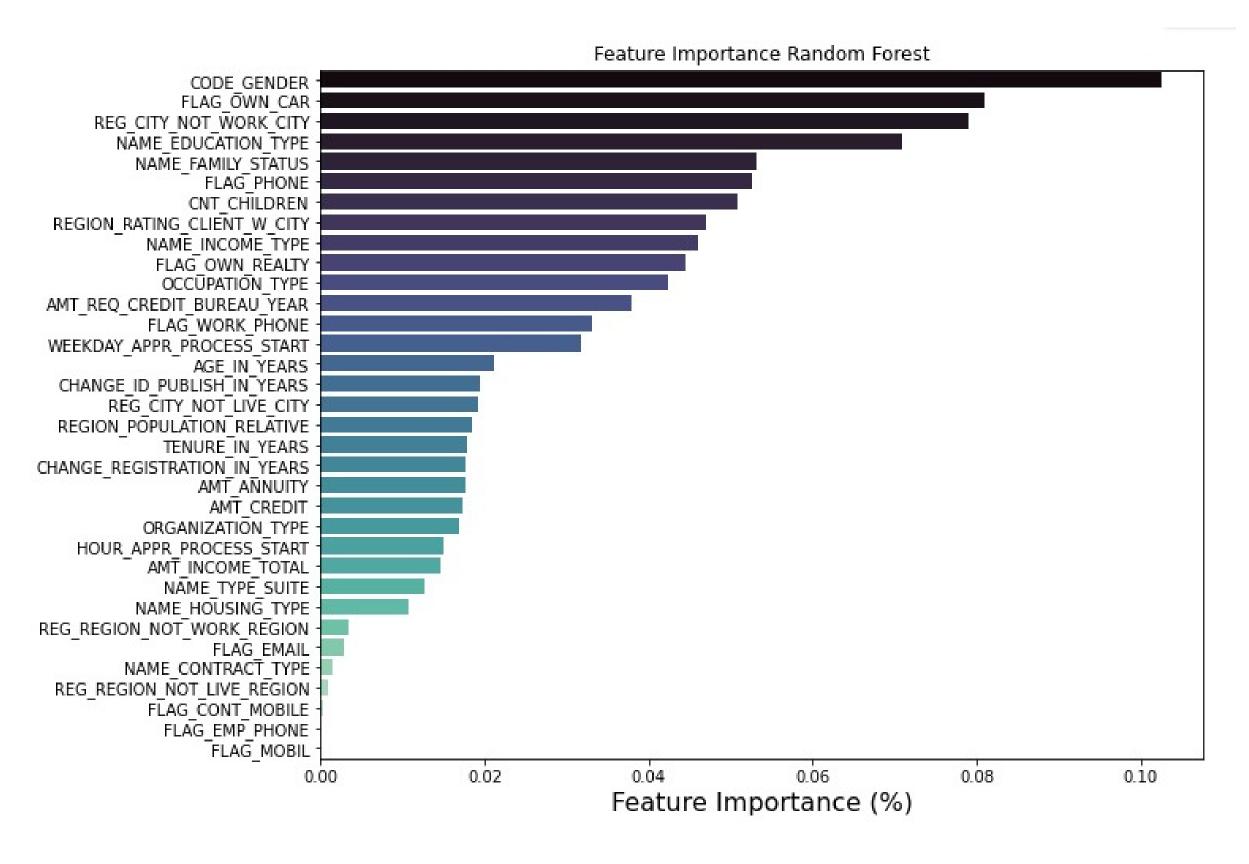


# 4. Machine Learning Modeling

#### Model ML Comparison Metric Evaluation

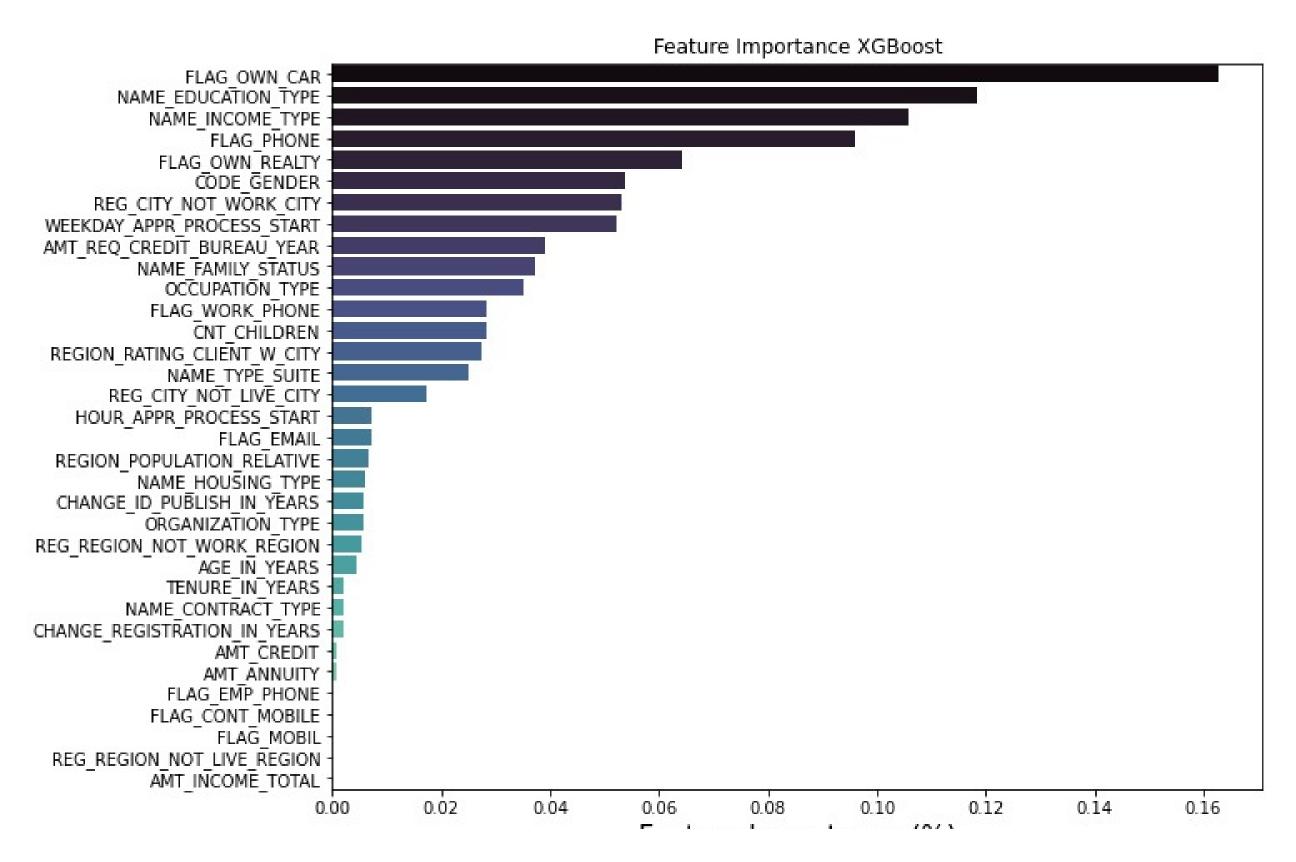
Models	ROC-AUC	Accuracy	Precision	Recall	f1-Score
Random Forest	0.950000	0.950000	1.000000	0.910000	0.950000
XGBoost	0.950000	0.950000	1.000000	0.900000	0.950000
XGBoost Tuned	0.950000	0.950000	1.000000	0.900000	0.950000
Random Forest Tuned	0.940000	0.940000	1.000000	0.890000	0.940000
Decision Tree	0.900000	0.910000	0.900000	0.920000	0.910000
K-Nearest Neighbors (KNN)	0.800000	0.800000	0.770000	0.860000	0.810000
Logistic Regression (LogReg)	0.530000	0.530000	0.550000	0.400000	0.460000

#### Feature Importance of Random Forest



- Fitur yang dianggap penting untuk memprediksi "TARGET" oleh model RF adalah :
- jenis kelamin
- kempemilikan mobil pribadi
- registrasi kota dan kota tempat kerja
- jenjang pendidikan
- status keluarga
- kepemilikan nomor handphone pribadi
- jumlah anak

#### Feature Importance of XGBoost



- Fitur yang dianggap penting untuk memprediksi "TARGET" oleh model XGB adalah :
- kempemilikan mobil pribadi
- jenjang pendidikan
- Jenis pendapatan
- kepemilikan nomor handphone pribadi
- kepemilikan property
- jenis kelamin

### 5. Recommendation

#### • For Business Improvement:

- Penting bagi tim analis dan tim keuangan untuk melakukan efisiensi credit agar defaulted tidak akan sering terjadi yang dapat merugikan pihak kreditur.
- o menggunakan bantuan machine learning dengan roc\_auc sebesar 95% dan kesalahan/error sebesar 5% maka berikut ini perhitungan efisiensinya:
- o misal asumsikan:
  - total dana kreditur: 100.000.000
  - credit terbayar: 100.000.000 x 95% = 95.000.000 (hanya angsuran, belum +bunga (jika ada))
  - credit tak terbayar: 100.000.000 x 5% = 5.000.000 (kerugian akibat ML gagal menentukan debitur gagal bayar)
- Rentang usia lebih muda 18-30 tahun cenderung kesulitan dalam melakukan pembayaran kredit. Sedangkan usia 35-45 adalah klien yang tidak memiliki kesulitan pembayaran kredit, sehingga dapat prioritas untuk diberi pinjaman.
- Jenjang pendidikan klien dan jenis pendapatannya perlu disoroti lebih lanjut, karena banyak klien yang memiliki kecenderungan kesulitan membayar berasal dari tingkat pendidikan secondary, dan masih mendapatkan income dari bekerja.
- Strategi pinjaman dapat diarahkan pada student, IT Staff, dan Businessman, karena terlihat 3 klien ini sedikit sekali yang mengalami kesulitan pembayaran pinjaman bahkan tidak ada.

- For Machine Learning Improvement:
  - Hasil machine learning yang memiliki performa baik yaitu XGBoost Classifier dan Random Forest Classifier. Baik berdasarkan score ROC-AUC maupun dari accuracy.
  - Setelah dilakukan Hyperparameter Tuning, tidak terlalu terjadi perubahan yang signifikan pada performa ML
  - Dapat menggunakan Pipeline untuk memudahkan dan mempercepat proses feature engineering dan deployment kedepannya.

# rerima kasih atas waktu Anga!

