



Home Credit : Credit Score

Mohammad Sulaeman

About this Presentation

Home Credit menggunakan berbagai metode statistik dan Machine Learning untuk membuat prediksi skor kredit. Membuka potensi maksimal dari data yang dimiliki Home Credit Indonesia dapat dilakukan dengan membuat model yang memastikan nasabah yang mampu membayar tidak ditolak saat mengajukan pinjaman. Selain itu, Home Credit dapat memberikan pinjaman dengan pokok, jatuh tempo, dan kalender pelunasan yang akan memotivasi nasabah untuk berhasil.

Dengan menggunakan metode CRISPM-DM, kami akan membangun model Regresi Logistik, Advanced Model XGBoost, dan Advanced Model LightGBM dan membandingkan model mana yang paling baik dalam memprediksi skor kredit.

Business Understanding

Problem Statement :

Skor kredit adalah angka dari 300 hingga 850 yang menggambarkan kelayakan kredit konsumen. Kelayakan kredit adalah cara pemberi pinjaman menentukan apakah seseorang akan gagal membayar kewajiban utangnya atau seberapa layak orang tersebut menerima kredit baru. Kelayakan kredit adalah apa yang dilihat oleh pemberi pinjaman sebelum menyetujui kredit baru untuk calon peminjam. Skor kredit memainkan peran kunci dalam keputusan pemberi pinjaman untuk menawarkan kredit. Semakin tinggi skornya, semakin baik seorang peminjam terlihat di mata pemberi pinjaman. Skor kredit didasarkan pada riwayat kredit: jumlah rekening yang terbuka, jumlah total utang, riwayat pembayaran, dan faktor lainnya. Pemberi pinjaman menggunakan skor kredit untuk mengevaluasi probabilitas bahwa seseorang akan membayar kembali pinjaman secara tepat waktu.

Business Understanding

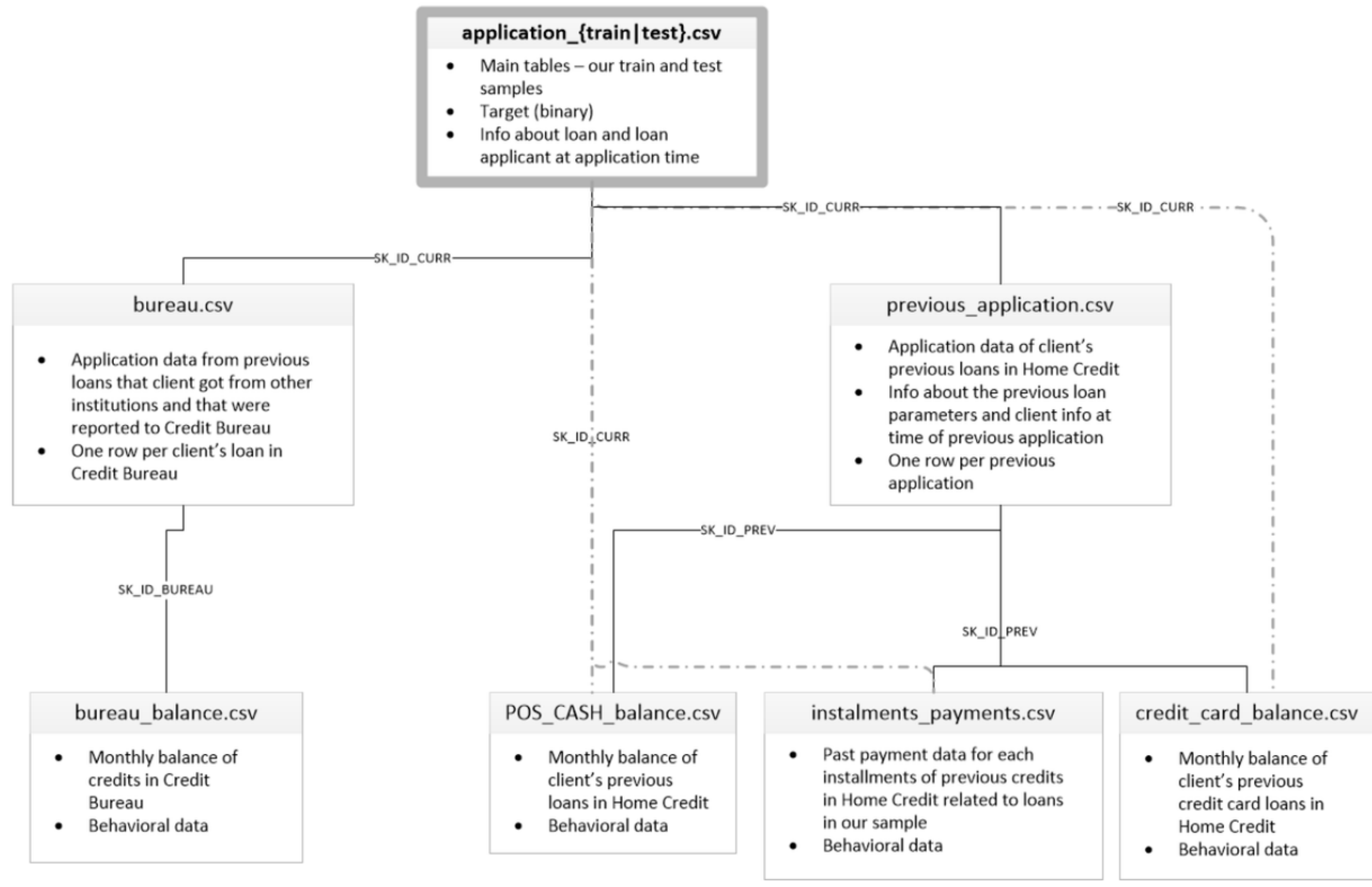
Business Metric and Objectives :

Faktor-faktor yang dipertimbangkan dalam penilaian kredit termasuk riwayat pembayaran, jenis pinjaman, lama riwayat kredit, dan total utang seseorang. Faktor-faktor tersebut dapat digunakan dalam membangun model untuk memprediksi kemampuan setiap pemohon dalam melunasi pinjaman (skor kredit). Setelah model prediksi skor kredit dibuat, model tersebut dievaluasi dengan metrik; Akurasi, Presisi, Recall, F1-score, dan AUC.

Model Definition :

Model yang dibuat adalah memprediksi skor kredit dengan menggunakan dataset yang disediakan oleh Home Credit Indonesia yang terdiri dari 8 data primer, yaitu `application_train`, `application_test`, `biro`, `bureau_balance`, `previous_appplication`, `POS_CASH_balance`, `cicilan_bayar`, dan `credit_card_balance`.

Data Understanding






Data Preparation

Data Gather

Menyatukan data dari beberapa csv menggunakan key column ['SK_ID_CURR', 'SK_ID_PREV'] dengan tujuan mempermudah dalam melakukan preparasi data dan juga memodeling data. Selain itu data gathering juga bertujuan untuk mengefisienkan semua variable yang terdapat pada semua data csv.

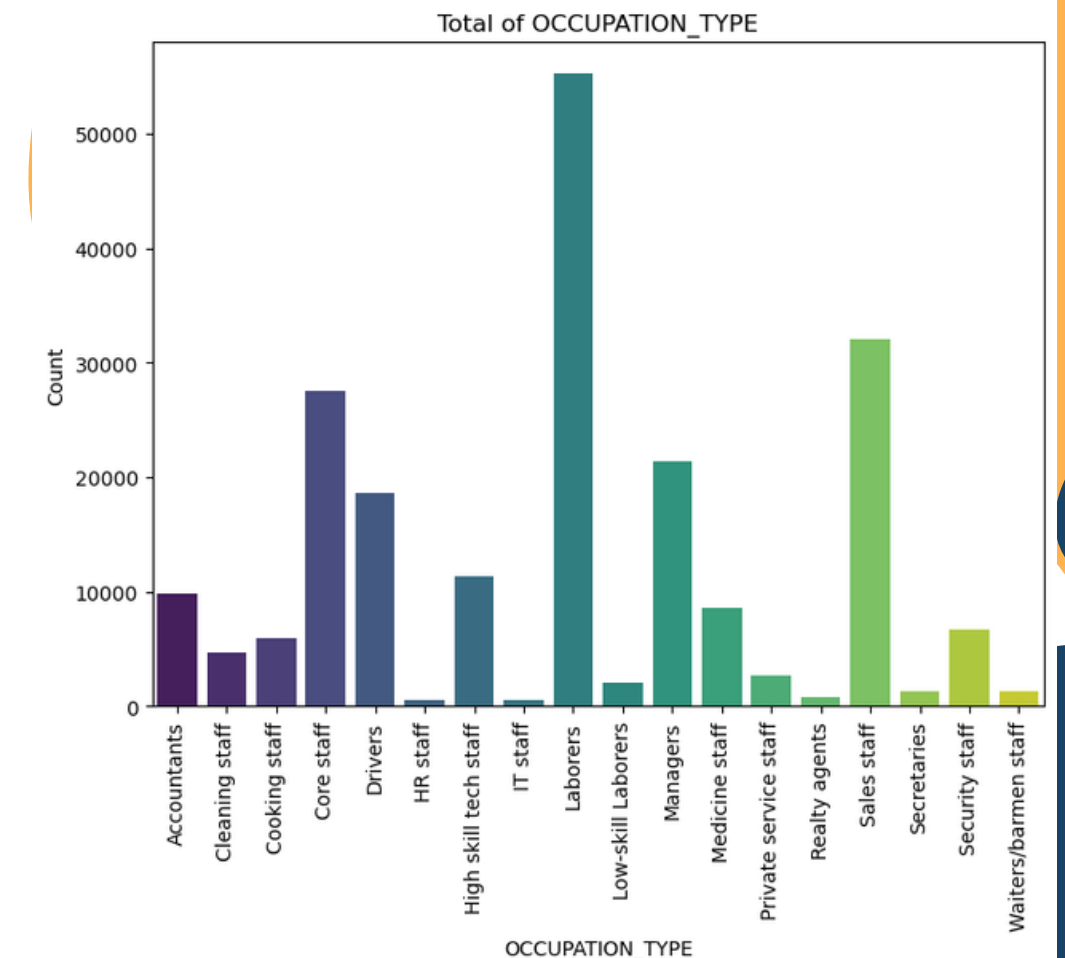
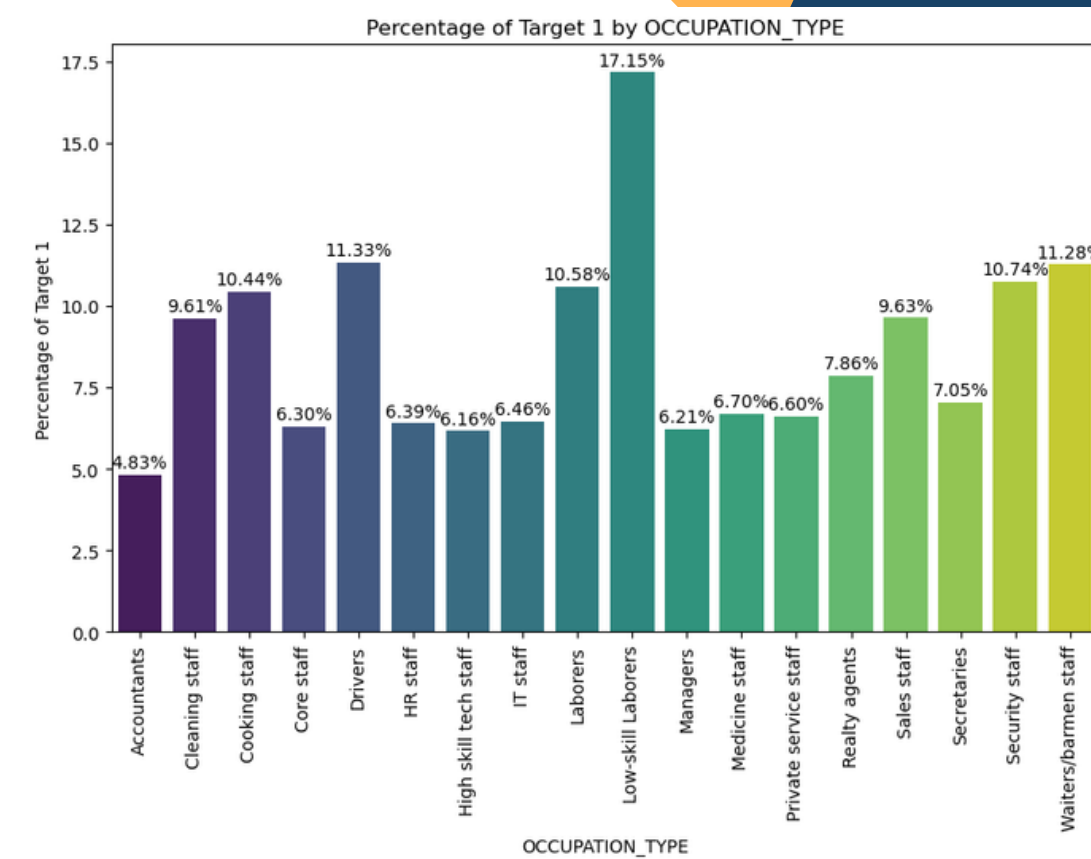
Cleaning Data

- Handling missing values : menghapus column yang memiliki persentase data missing diatas 70%
 - Handling Variables with Small Variances : menghapus column yang memiliki sedikit informasi yang dibutuhkan
 - Fill the blank row : mengisi row yang kosong dengan nilai median dari tiap tiap column.
- 

Analyst Data

Dari data disamping bisa dilihat bahwa seseorang dengan posisi Accountants, HR Staff, dan IT Staff memiliki tingkat untuk credit risk yang rendah dan berpotensi menjadi customer yang baik. Namun tidak banyak dari mereka yang melakukan peminjaman. Oleh karena itu diperlukan adanya campaign untuk beberapa posisi tersebut agar mereka lebih tertarik untuk melakukan credit/ pinjaman.

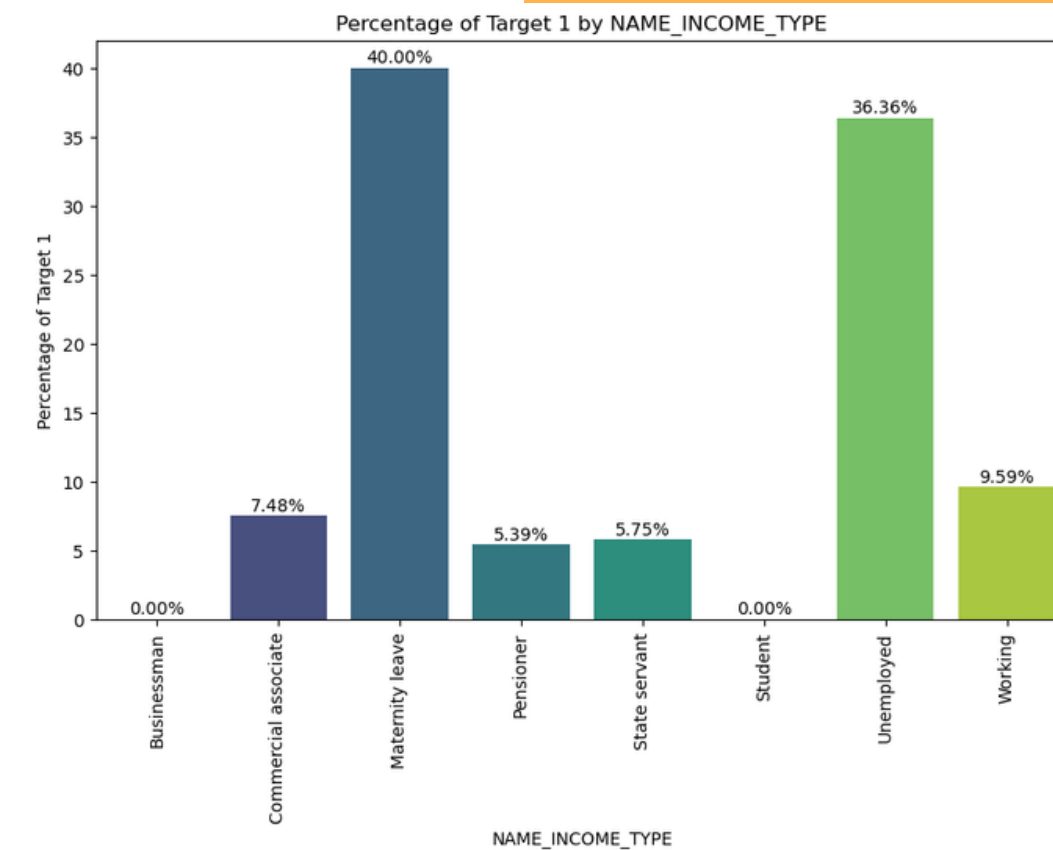
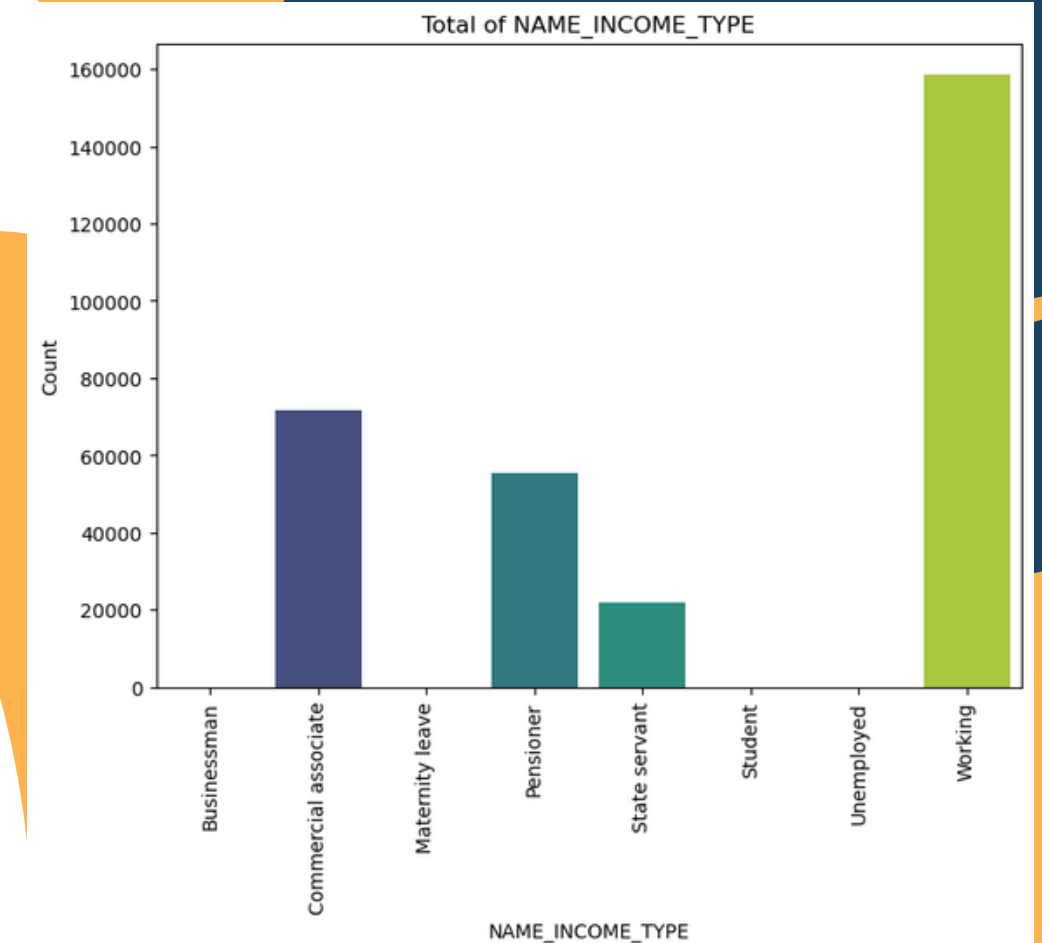
Selain itu bisa dilihat bahwa posisi Low Skill Laborers memiliki angka yang tinggi dalam kesulitan melakukan pembayaran. Oleh karena itu company harus lebih berhati hati dalam melakukan penerimaan credit posisi tersebut.



Analyst Data

Dari data disamping bisa dilihat bahwa seorang unemployed dan maternity leave memiliki tingkat untuk credit risk yang sangat tinggi. Namun sangat sedikit dari mereka yang melakukan credit/peminjaman. Dari data ini, company harus lebih mempertimbangkan untuk memberikan credit kepada 2 posisi ini.

Disisi lain, seseorang yang memiliki pekerjaan/working sangat sering melakukan peminjaman dan memiliki credit tisk yang rendah, campaign untuk posisi ini harus dibanyakkan agar lebih banyak yang tertarik melakukan credit.



Modeling Data

Baseline Model Logistic Regression

	SK_ID_CURR	TARGET
307511	100001	0.485442
307512	100005	0.531552
307513	100013	0.466850
307514	100028	0.475320
307515	100038	0.551509

AUC Score 0.68 from kaggle

Advanced Model LightGBM

	SK_ID_CURR	TARGET
307511	100001	0.619357
307512	100005	0.768226
307513	100013	0.072469
307514	100028	0.173089
307515	100038	0.882501

AUC Score 0.70516 from Kaggle

Advanced Model XGBoost

	SK_ID_CURR	TARGET
307511	100001	0.437404
307512	100005	0.631641
307513	100013	0.177599
307514	100028	0.262168
307515	100038	0.615681

AUC Score 0.72373 from kaggle

Dari 3 method yang digunakan untuk melakukan credit scoring, didapatkan bahwa Advance Model XGBoost memiliki tingkat akurasi yang lebih bagus dibandingkan 2 method lainnya dengan score AUC 0.72372



- Terima Kasih -

Link Github