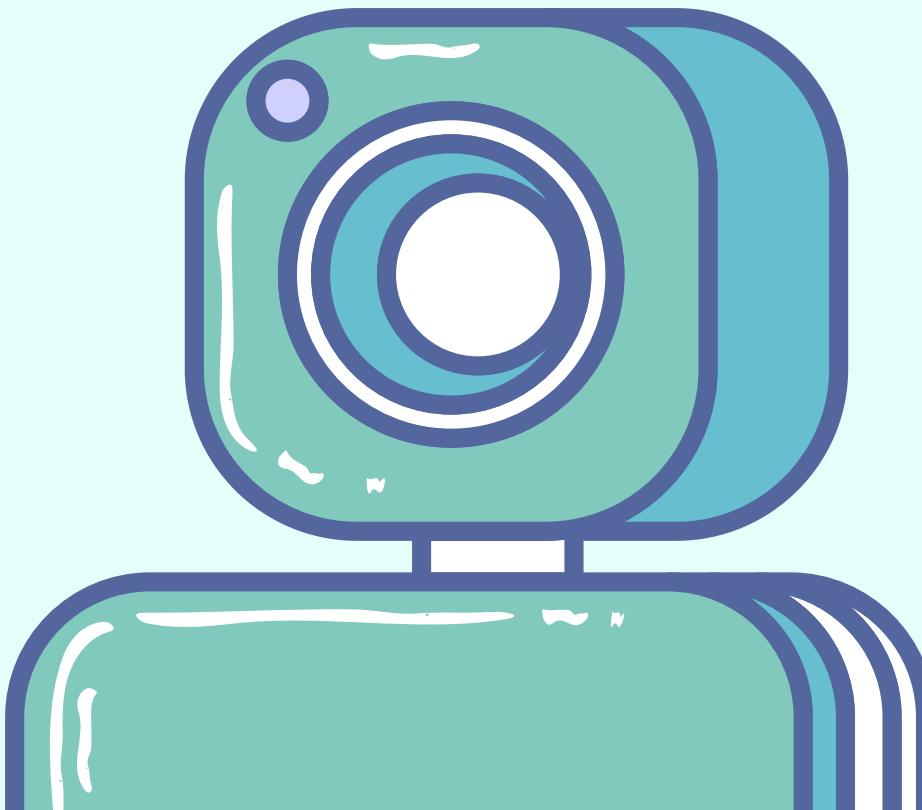


HOME CREDIT SCORECARD MODEL

**HOME CREDIT INDONESIA
DATA SCIENTIST**

Juan Anemao Sokhi Zidomi



Problem Research

Problem

Home Credit Indonesia saat ini menghadapi tantangan dalam mendorong nasabah untuk melakukan pelunasan kredit secara tepat waktu. Hal ini terlihat dari meningkatnya jumlah pelanggan yang mengalami keterlambatan hingga gagal melunasi kewajiban pinjaman mereka. Kondisi ini menunjukkan bahwa strategi pengelolaan dan pengawasan pelunasan kredit belum sepenuhnya efektif dalam memastikan nasabah mampu dan bersedia menyelesaikan kewajibannya sesuai jadwal.

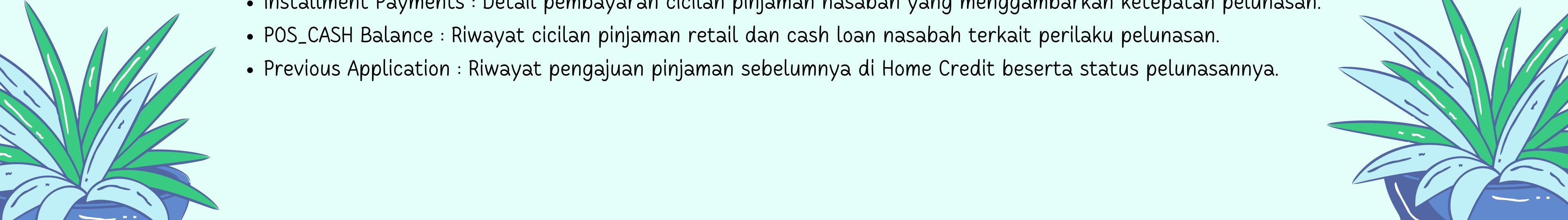
Tujuan

Membangun model yang dapat memprediksi pelanggan yang berpotensi melakukan pelunasan kredit tidak tepat waktu, sehingga perusahaan dapat lebih mudah menentukan strategi penagihan dan pengelolaan kredit secara efektif.





Informasi Dataset



Dataset ini terdiri dari 8 data bertipe csv. Adapun informasinya sebagai berikut:

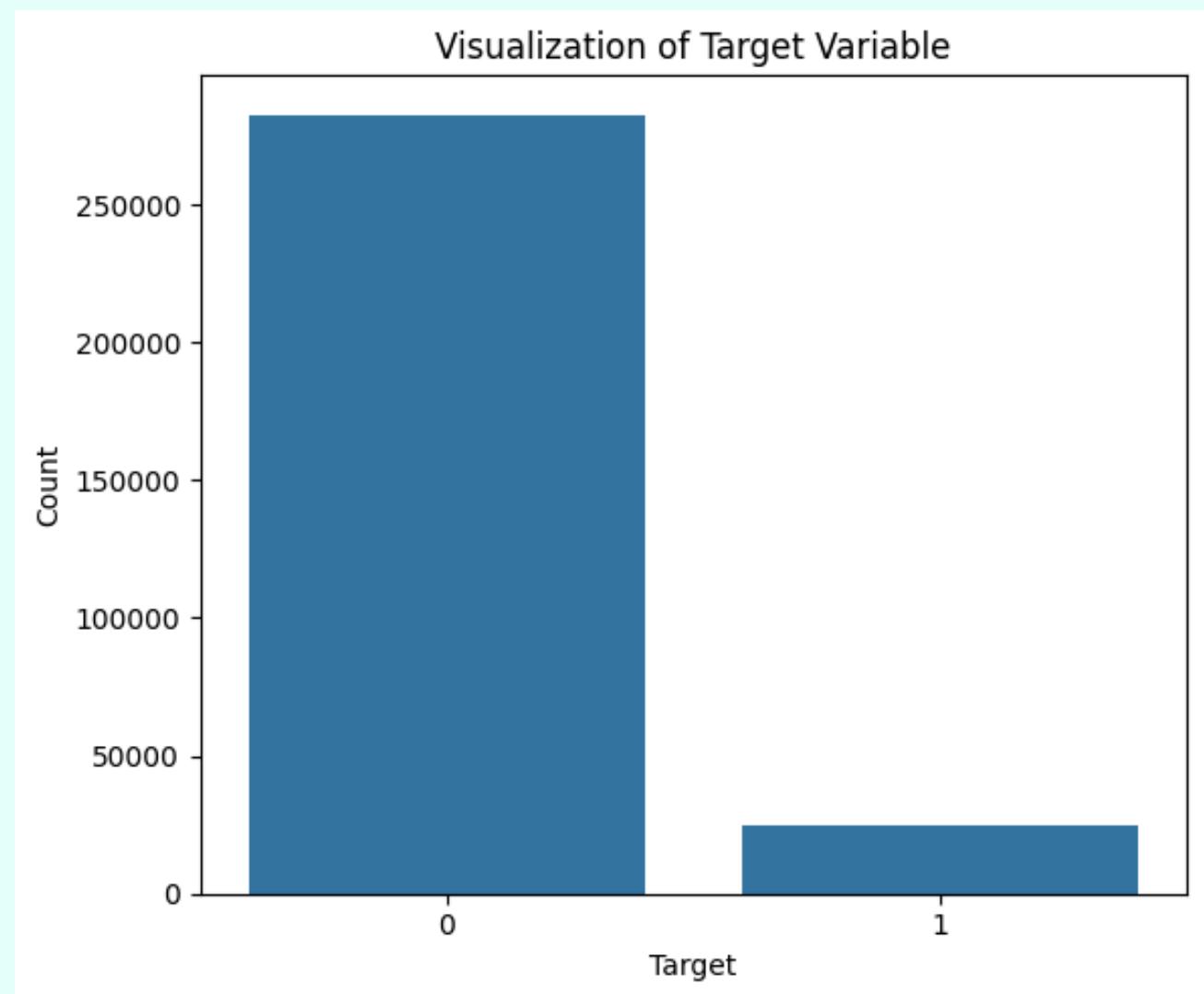
- Application train dan Application test: Data utama aplikasi pinjaman nasabah (profil, pekerjaan, pendapatan dan lain-lain). Data train memiliki target, sedangkan data test tidak memiliki target
- Bureau : Riwayat kredit nasabah
- Bureau Balance : Status bulanan di kredit bureau
- Credit Card Balance : Riwayat penggunaan dan pembayaran kartu kredit nasabah
- Installment payments : Detail pembayaran cicilan
- POS_CASH Balance : Riwayat cicilan pinjaman retail & cash loan nasabah.
- Previous application : Riwayat pengajuan pinjaman sebelumnya di Home CreditApplication
- Bureau : Riwayat kredit nasabah yang mencerminkan performa pelunasan kredit di lembaga lain.
- Bureau Balance : Status bulanan kredit nasabah pada biro kredit yang menunjukkan perkembangan pelunasan.
- Credit Card Balance : Riwayat penggunaan dan pembayaran kartu kredit nasabah sebagai indikator kedisiplinan pelunasan.
- Installment Payments : Detail pembayaran cicilan pinjaman nasabah yang menggambarkan ketepatan pelunasan.
- POS_CASH Balance : Riwayat cicilan pinjaman retail dan cash loan nasabah terkait perilaku pelunasan.
- Previous Application : Riwayat pengajuan pinjaman sebelumnya di Home Credit beserta status pelunasannya.

Informasi Dataset

Statistik:

- Dataset terdiri dari 307511 baris dan 122 kolom yang mana Kolom TARGET sebagai kolom target
- Pada kolom target, terdapat imbalance data dimana target 0 (Lancar) lebih besar daripada target 1

Link Penggerjaan: [Github](#)



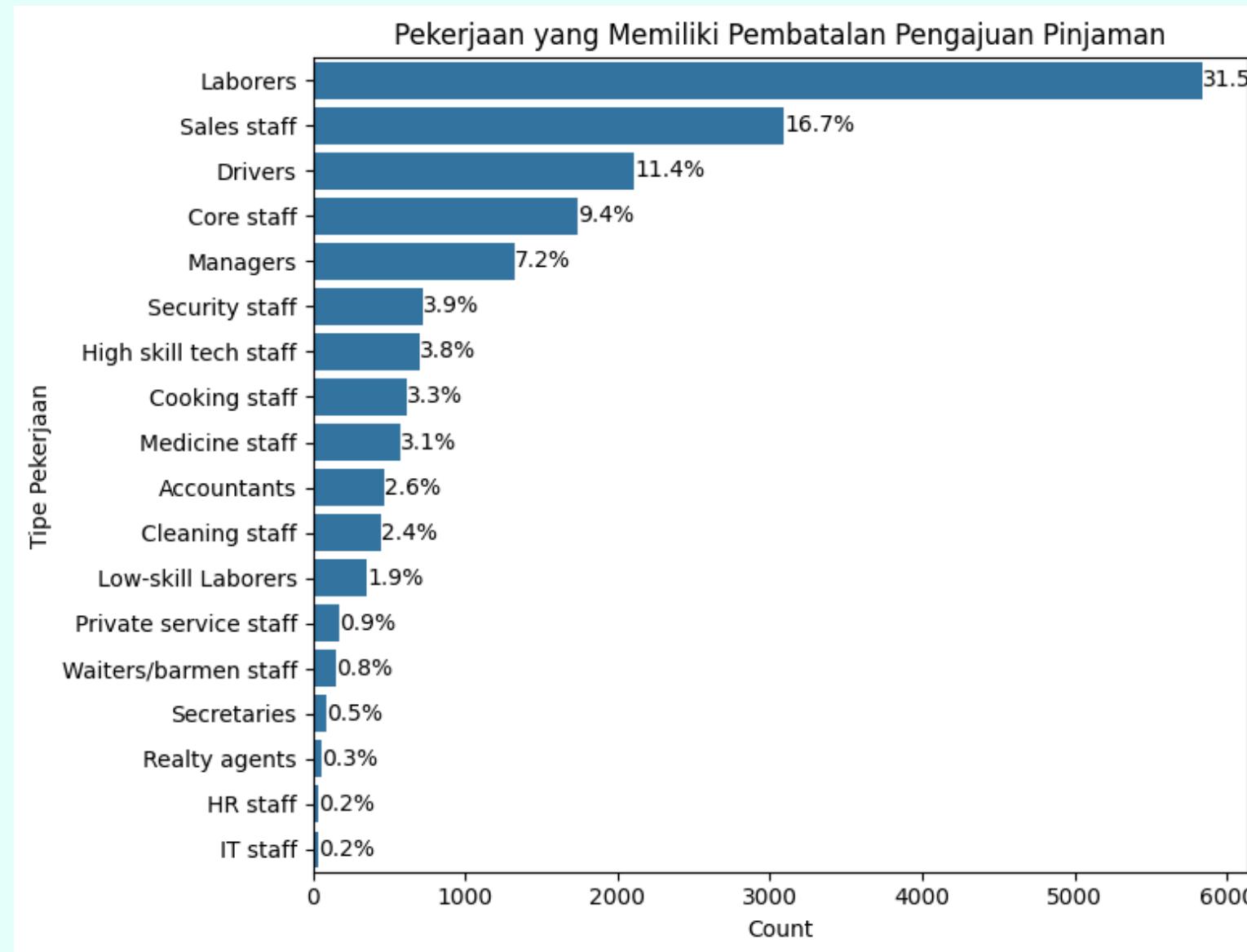
Preprocessing Data

Adapun tahapan preprocessing yang dilakukan yaitu:

1. Missing Data Handling
2. Encoding menggunakan encode manual dan One Hot Encoding
3. Feature Engineering
4. Merge Data



Data Visualization



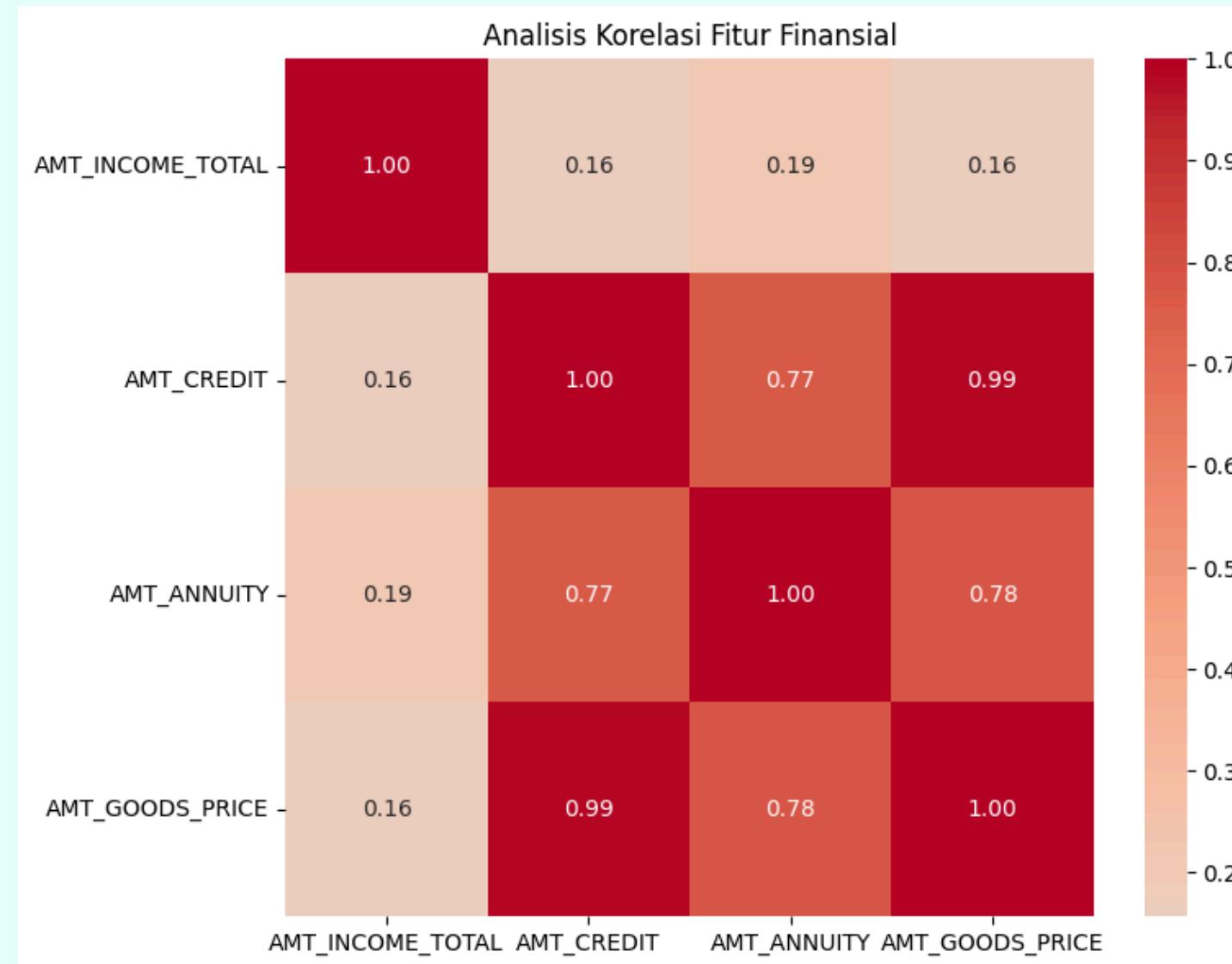
Temuan Data:

- Pada jenis pekerjaan Labores memiliki resiko tinggi terhadap gagal bayar
- Pada jenis pekerjaan Private-service staff, Waiters/barman staff, Secretaries, Realty Agents, HR Staff, IT Staff memiliki resiko rendah terhadap gagal bayar

Action:

- Melakukan limit pada kredit yang jenis pekerjaan yang berisiko gagal bayar

Data Visualization



Temuan Data:

- Korelasi yang paling besar adalah AMT_CREDIT dengan AMT GOODS_PRICE yaitu 0.99
- Korelasi yang paling kecil adalah AMT_GOODS_PRICE dengan AMT_INCOME_TOTAL

Action:

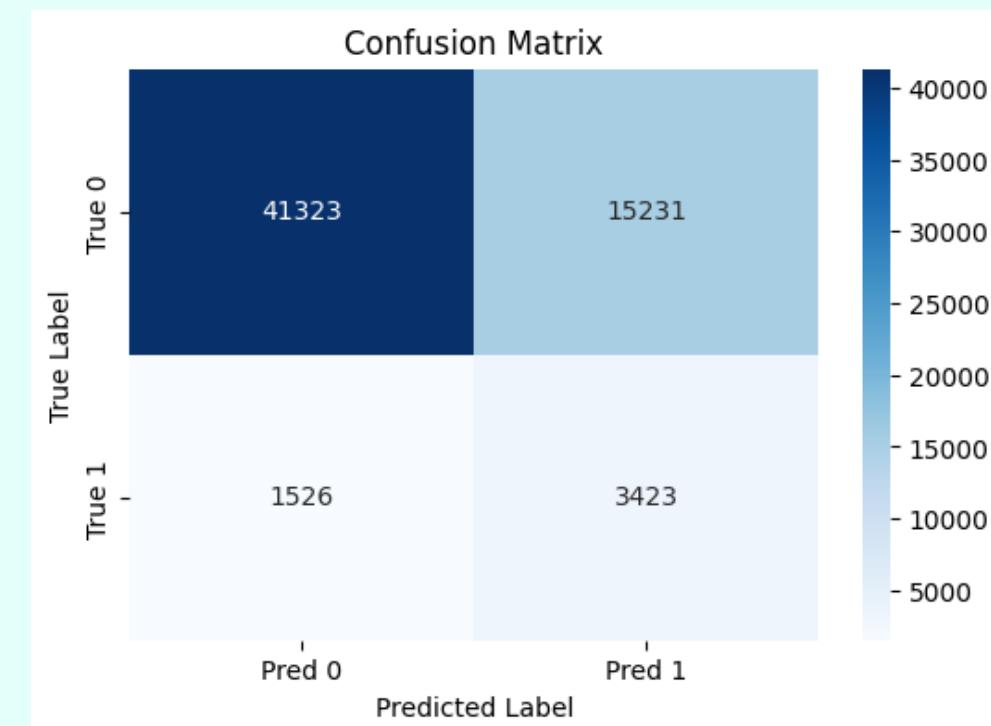
- Terapkan aturan DTI (Debt-to-Income) pada saat approval

EVALUASI DAN HASIL MODEL

	Accuracy	Precisious	Recall	F1 Score	AUC
XGBoost	0.719981	0.179439	0.694080	0.285157	0.708164
Cat Boost Classifier	0.727542	0.183500	0.691655	0.290048	0.711169
Light GBM	0.745086	0.190075	0.664781	0.295624	0.708447

Perbandingan Algoritma:

- Cat Boost Algoritma memiliki performa yang terbaik dibandingkan algoritma lain sehingga algoritma tersebut dipilih



Kesimpulan pada Confusion Matrix:

1. Model dapat mendeteksi nasabah yang berisiko dengan 69%, tetapi false positive yang tinggi sehingga model kurang efektif

Rekomendasi Bisnis

Adapun rekomendasi bisnis:

- Pada saat penentuan approval gunakan aturan DTI (Debt-To-Income)
- Melakukan edukasi pada nasabah yang melakukan peminjaman supaya melunasi pinjaman
- Mengimplementasikan model tersebut pada penentuan keputusan peminjaman.



Kesimpulan Dan Langkah Selanjutnya

Kesimpulan:

- Pada project tersebut, model yang dapat memprediksi nasabah yang dapat membayar tagihan credit yaitu Cat Boost Classofier yang memiliki performa dibandingkan model lain
- Pada jenis pekerjaan Labores memiliki resiko paling tinggi dibandingkan jenis pekerjaan lain dalam pelunasan tagihan credit.

Next Step:

- Melakukan Hyperparameter pada algoritma yang digunakan.
- Melakukan preprocessing dengan benar seperti feature engineering dll

