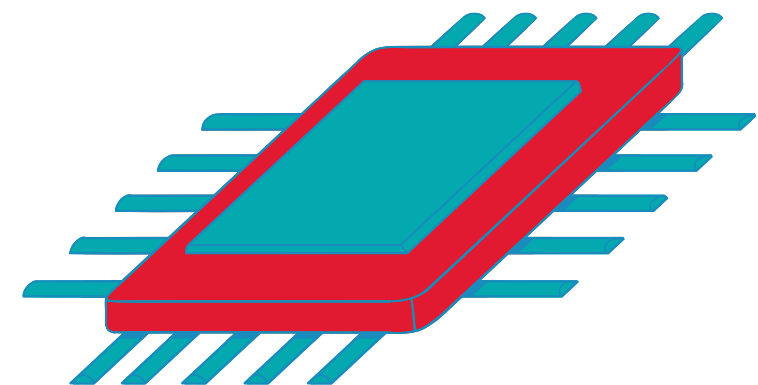


CREDIT CARD APPROVAL MACHINE LEARNING PRESENTATION

PRESENTED BY:

MUHAMMAD FAUZAN



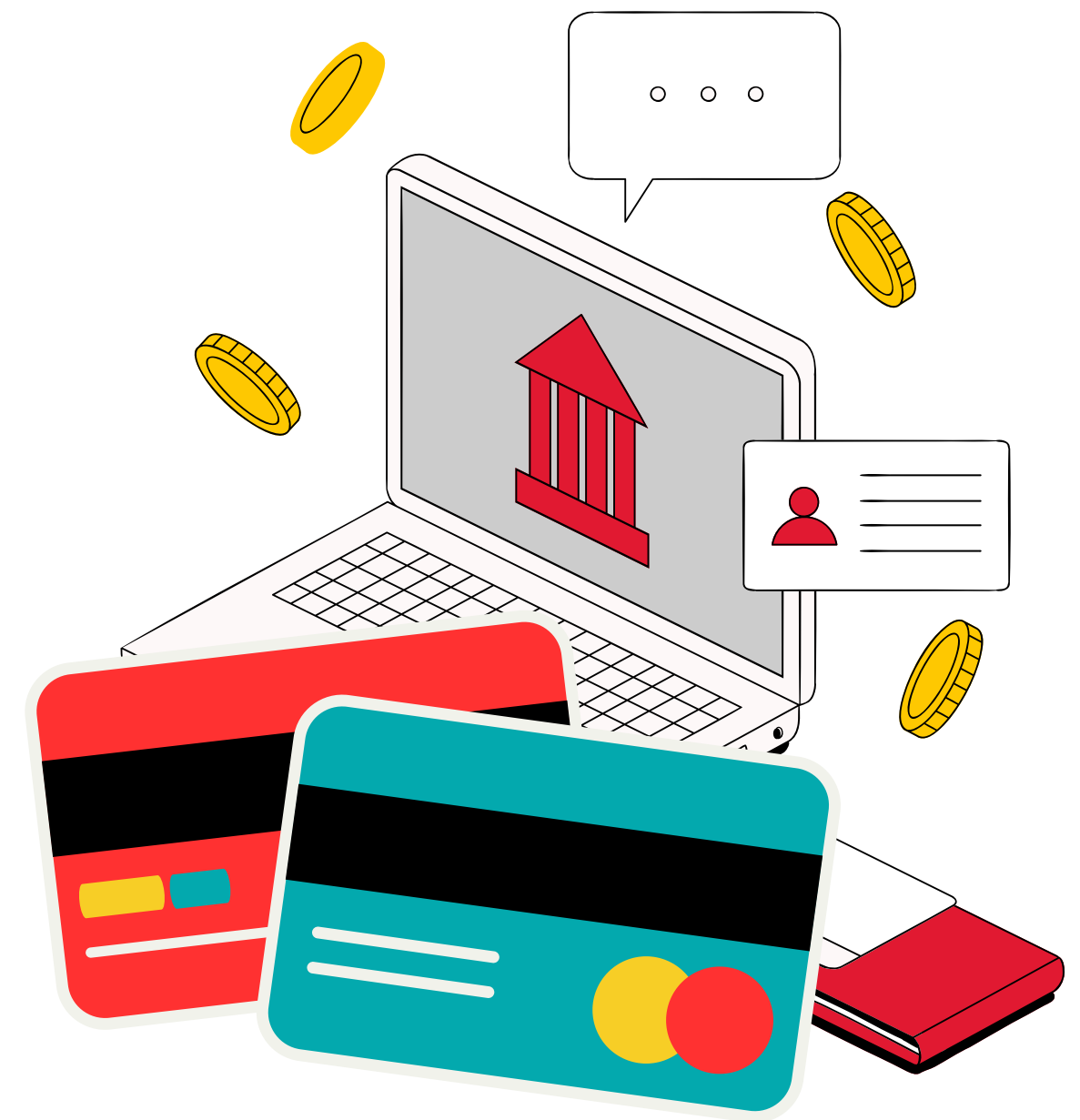
PROBLEM TO SOLVE

HOME
CREDIT

Rakamin
Academy

HOME CREDIT INDONESIA

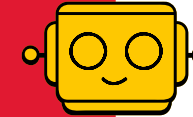
1. Mengurangi jumlah pelanggan yang layak tetapi ditolak dalam pengajuan pinjaman.
2. Memastikan pinjaman diberikan dengan syarat (principal, maturity, repayment calendar) yang optimal, sehingga meningkatkan kemungkinan pelunasan.
3. Meningkatkan akurasi model prediksi risiko kredit dengan pendekatan Machine Learning, termasuk Logistic Regression dan setidaknya satu model lain.
4. Menganalisis performa model secara menyeluruh (end-to-end) dan memberikan rekomendasi bisnis berbasis hasil analisis data.



DATASET YANG DIMILIKI

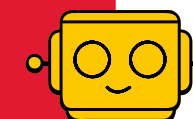
HOME
CREDIT

Rakamin
Academy



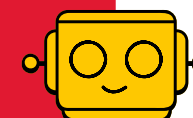
- Dataset terdiri dari `application_train.csv` dan `application_test.csv` yang menyimpan informasi utama tentang pengajuan pinjaman, termasuk detail pemohon dan status kredit dengan targetnya.

- Dataset `bureau.csv` dan `bureau_balance.csv` yang berisi riwayat kredit pelanggan dari lembaga keuangan lain, serta saldo bulanan kredit tersebut.



- Dataset `previous_application.csv` mencatat riwayat pengajuan pinjaman sebelumnya di Home Credit, sedangkan `POS_CASH_balance.csv`, `installments_payments.csv`, dan `credit_card_balance.csv` menyimpan informasi tentang pembayaran dan saldo pinjaman atau kredit pelanggan di Home Credit.

- Terakhir, terdapat `HomeCredit_columns_description.csv` yang berisi deskripsi setiap fitur dalam dataset.



INSIGHT FROM THE DATA

01

DEMOGRAFI DAN RIWAYAT PEKERJAAN

Faktor seperti usia (AGE) dan lama bekerja (YEAR_EMP) berperan penting dalam menilai stabilitas finansial peminjam. Umumnya, semakin tua usia dan semakin lama seseorang bekerja, semakin kecil risiko gagal bayar.

02

PENDAPATAN DAN KEMAMPUAN FINANSIAL

Variabel seperti jumlah anak , total pendapatan, dan besar pinjaman (AMT_CREDIT) membantu mengukur apakah peminjam memiliki kapasitas keuangan yang memadai untuk melunasi kredit.

03

KEPEMILIKAN ASET DAN JEJAK KREDIT

Kepemilikan mobil atau properti (FLAG_OWN_CAR, FLAG_OWN_REALTY) serta jumlah permintaan kredit sebelumnya menjadi indikator kuat untuk memahami pola keuangan peminjam dan risiko kreditnya.



DATA PRE-PROCESSING

MENGATASI MISSING VALUE

Data yang dimiliki memiliki banyak sekali missing value, untuk mengatasi hal ini. Kolom dengan lebih dari 50% missing values dihapus, sedangkan sisanya diisi dengan nilai median (untuk numerik) dan modus (untuk kategori).

FEATURE ENGINEERING

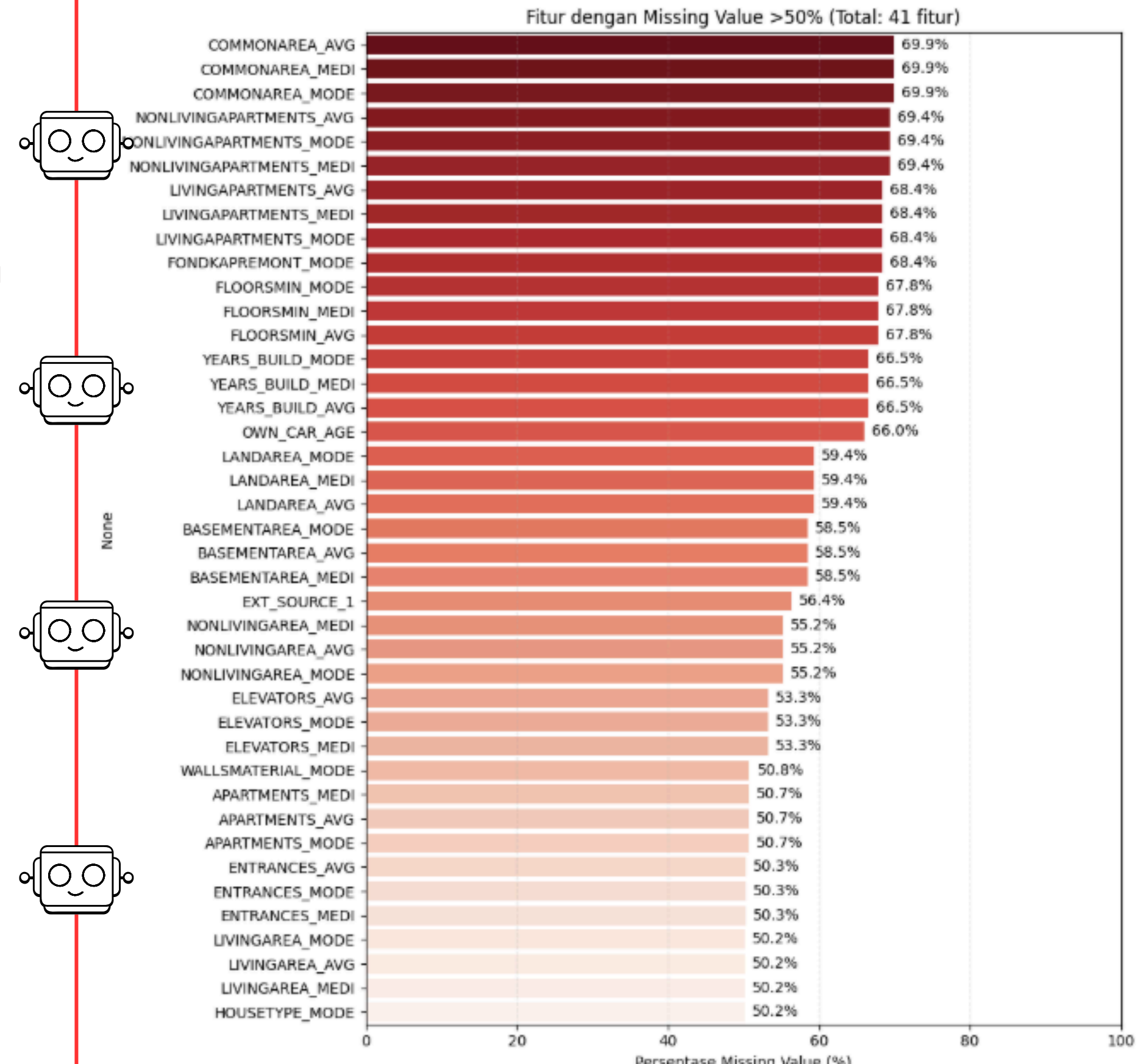
Beberapa variabel seperti umur, lama bekerja, dan total permintaan kredit dihitung ulang untuk merepresentasikan informasi dengan lebih baik.

HANDLING OUTLIERS & FEATURE SELECTION

Distribusi data divisualisasikan untuk mendeteksi outlier yang dapat memengaruhi hasil prediksi. Seleksi fitur dilakukan menggunakan Weight of Evidence (WoE) untuk meningkatkan relevansi variabel.

ENCODING, FINAL DATA

One-hot encoding digunakan untuk mengubah variabel kategori, menghasilkan total 90 fitur dari 23 fitur awal.



MODELING

MODEL YANG DICoba

- Logistic Regression Tanpa balancing
- Logistic Regression Oversampling
- Logistic Regression Undersampling
- XGBoost (Dengan undersampling)

EVALUASI MODEL

Tanpa Balancing & Oversampling (LogReg):

- Akurasi tinggi (92%), tetapi recall kelas 1 sangat rendah (0.00) → model tidak mampu mengenali peminjam yang gagal bayar.
- ROC AUC: ~0.7335

Undersampling (LogReg & XGBoost):

- Akurasi turun (68%) tetapi recall kelas 1 meningkat drastis (67%), artinya model lebih baik dalam mendeteksi peminjam berisiko.
- ROC AUC tetap stabil (~0.7335 untuk LogReg, 0.7329 untuk XGBoost)

KESIMPULAN

- Model tanpa balancing & oversampling gagal mengenali peminjam gagal bayar.
- Undersampling meningkatkan recall kelas 1 tetapi mengorbankan akurasi.
- Terdapat 2 kandidat model yang paling baik, yaitu Logistic Regression dengan Undersampling, dan juga XGBoost.

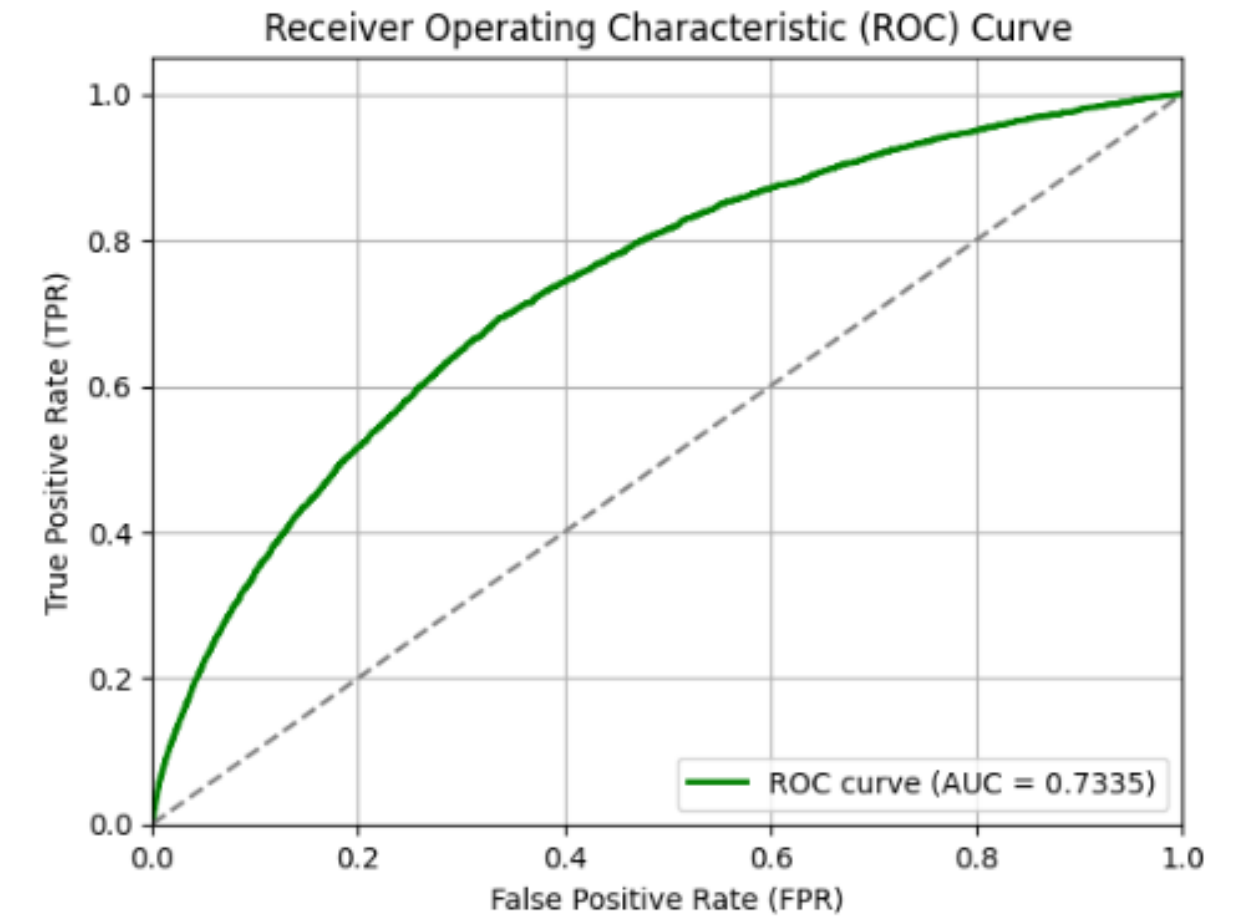
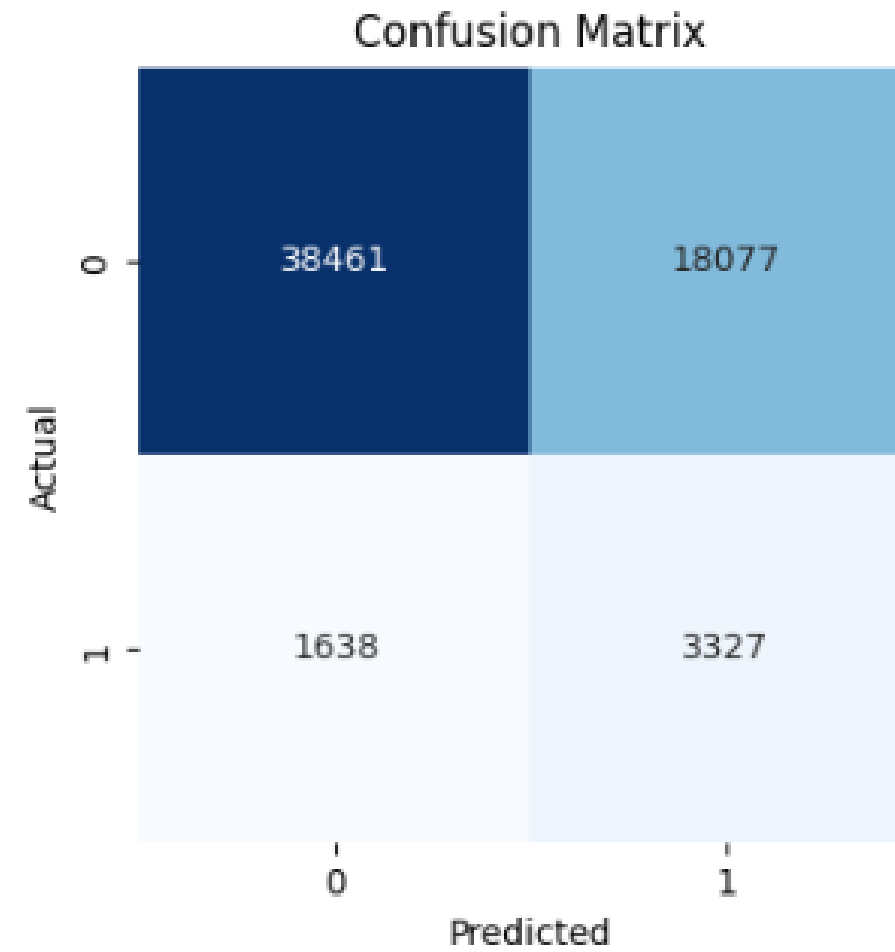
**HOME
CREDIT**

Rakamin
Academy

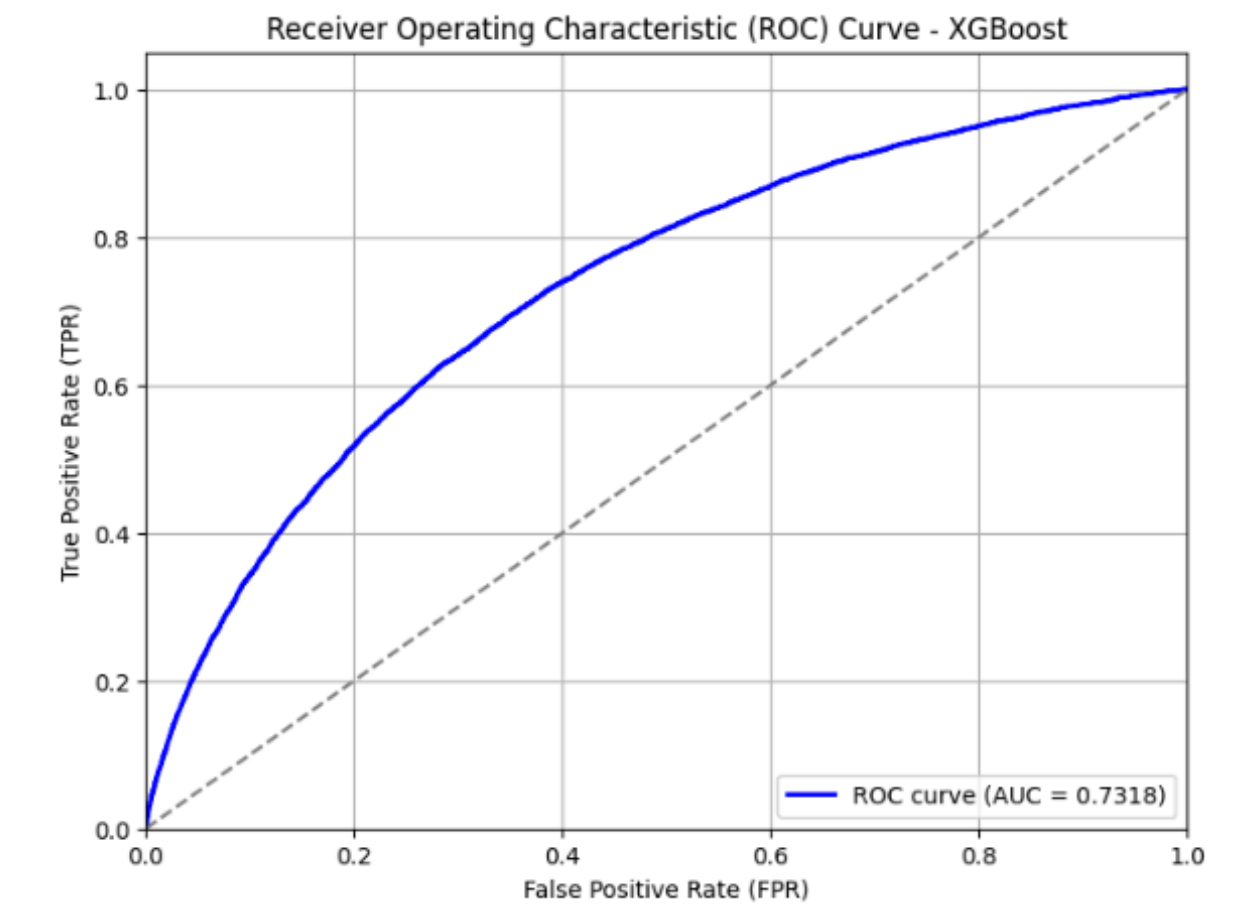
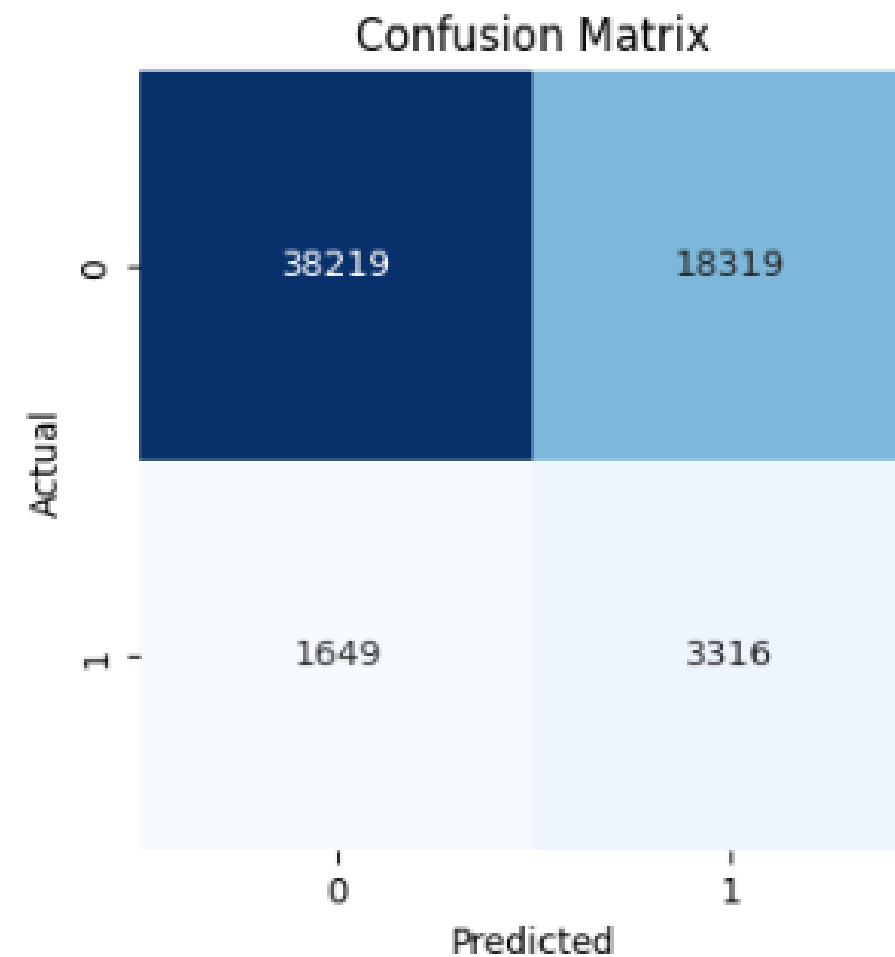


Melalui Confusion Matrix dan ROC Curve pada gambar, Logistic Regression dengan undersampling dipilih karena memiliki TN dan TP lebih tinggi serta AUC sedikit lebih besar dibanding XGBoost, menunjukkan performa yang lebih baik.

L
O
G
I
S
T
I
C



X
G
B
O
O
S
T



IMPLEMENTATION

```
[ ] predictions = best_model_logreg_undersampled.predict(data_test_pred)

[ ] data_test_result = data_test.copy()

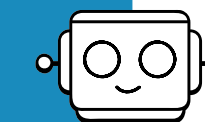
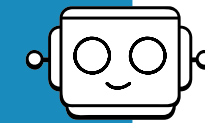
    data_test_result['TARGET'] = predictions

[ ] result.to_csv('result.csv', index=False)
```

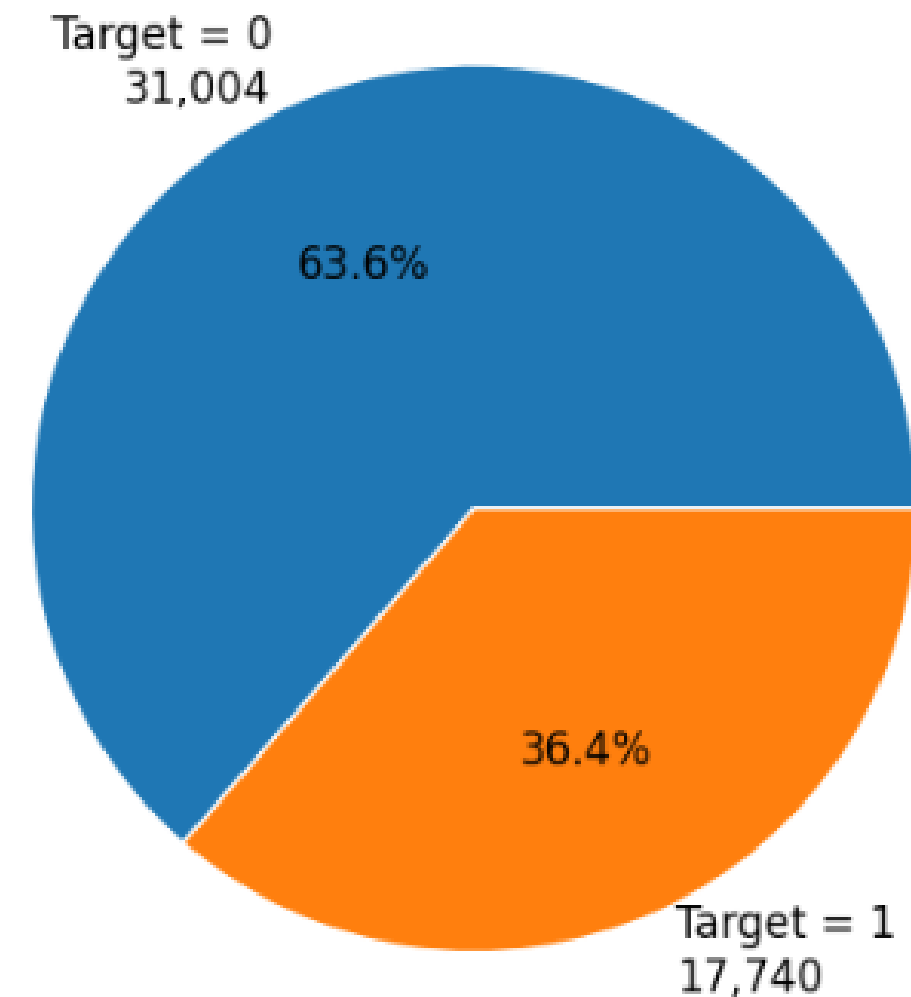
- output

	SK_ID_CURR	TARGET
0	100001	1
1	100005	1
2	100013	0
3	100028	0
4	100038	1

- Memprediksi data baru menggunakan model Logistic Regression
- Sesuai dengan output yang diinginkan, file hasil prediksi terhadap data baru dengan disimpan dengan fitur atau kolom SK_ID_CURR dan TARGET



DISTRIBUSI TARGET HASIL PREDIKSI



- TARGET = 0 berarti pelanggan tidak mengalami kesulitan pembayaran → Kemungkinan besar kreditnya berjalan lancar (approved & lancar).
- TARGET = 1 berarti pelanggan mengalami kesulitan pembayaran → Pelanggan ini berisiko gagal bayar.

Business Insight

1. Kepemilikan Aset Berpengaruh dalam Keputusan Kredit

- Temuan:
 - Fitur kepemilikan aset (contoh: kendaraan) muncul sebagai salah satu faktor penting dalam model.
 - Ini menunjukkan bahwa memiliki aset mungkin dikaitkan dengan kemampuan finansial yang lebih baik.
- Action:
 - Tawarkan produk pinjaman berbasis aset, seperti pinjaman dengan jaminan kendaraan, untuk nasabah dengan risiko menengah.
 - Lakukan analisis lebih lanjut apakah faktor kepemilikan aset berkontribusi signifikan terhadap kredit macet atau hanya sebagai indikasi stabilitas finansial.

2. Skor Eksternal Sangat Berpengaruh

- Temuan:
 - Fitur terkait EXT_SOURCE_3 & EXT_SOURCE_2 memiliki pengaruh besar dalam model prediksi.
- Action:
 - Jika skor eksternal rendah tetapi faktor lain mendukung, dapat diberikan opsi pinjaman dengan syarat lebih fleksibel.

Recomendation

- Penyesuaian kebijakan persetujuan kredit → Pelanggan dengan risiko sedang bisa diberikan syarat pinjaman khusus, misalnya tenor lebih pendek atau suku bunga yang disesuaikan.
- Otomasi evaluasi risiko kredit → Model Machine Learning dapat digunakan sebagai sistem awal untuk menyaring aplikasi kredit sebelum dianalisis lebih lanjut oleh analis kredit.
- Penggunaan data eksternal → Integrasi data dari sumber eksternal (misalnya riwayat transaksi digital) untuk meningkatkan akurasi prediksi.

Thank You !

A huge thank you to Home Credit Indonesia and Rakamin Academy for the valuable learning experience and opportunity to work on this project. Your support and guidance have been instrumental in this journey!

Resources & Links:



<https://github.com/Muzann11/Home-Credit-Indonesia>



<https://youtu.be/al2YzyWLZKo>



<https://www.linkedin.com/in/mfauzandsml/>