

CREDIT CARD APPROVAL USE MACHINE LEARNING

PRESENTED BY
KEVIN CAESAR SIMORANGKIR



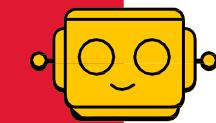
Problem Statement

HOME CREDIT INDONESIA



1. Mengurangi jumlah pelanggan yang layak tetapi ditolak dalam pengajuan pinjaman.
2. Memastikan pinjam dan berikan 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



- 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.



Pembersihan dan Pra Pemrosesan data

1 handling Missing Values

fitur dengan missing value, untuk mengatasi hal ini. fitur dengan lebih dari 50% missing values dihapus, sedangkan sisanya diisi dengan nilai median/mean melihatnya dari skewnesss (numerik) dan modus (kategori).

2 Feature Engineering

fitur seperti umur, lama bekerja, dantotal permintaan kredit dihitung ulang untuk merepresentasikan informasi dengan lebih baik.

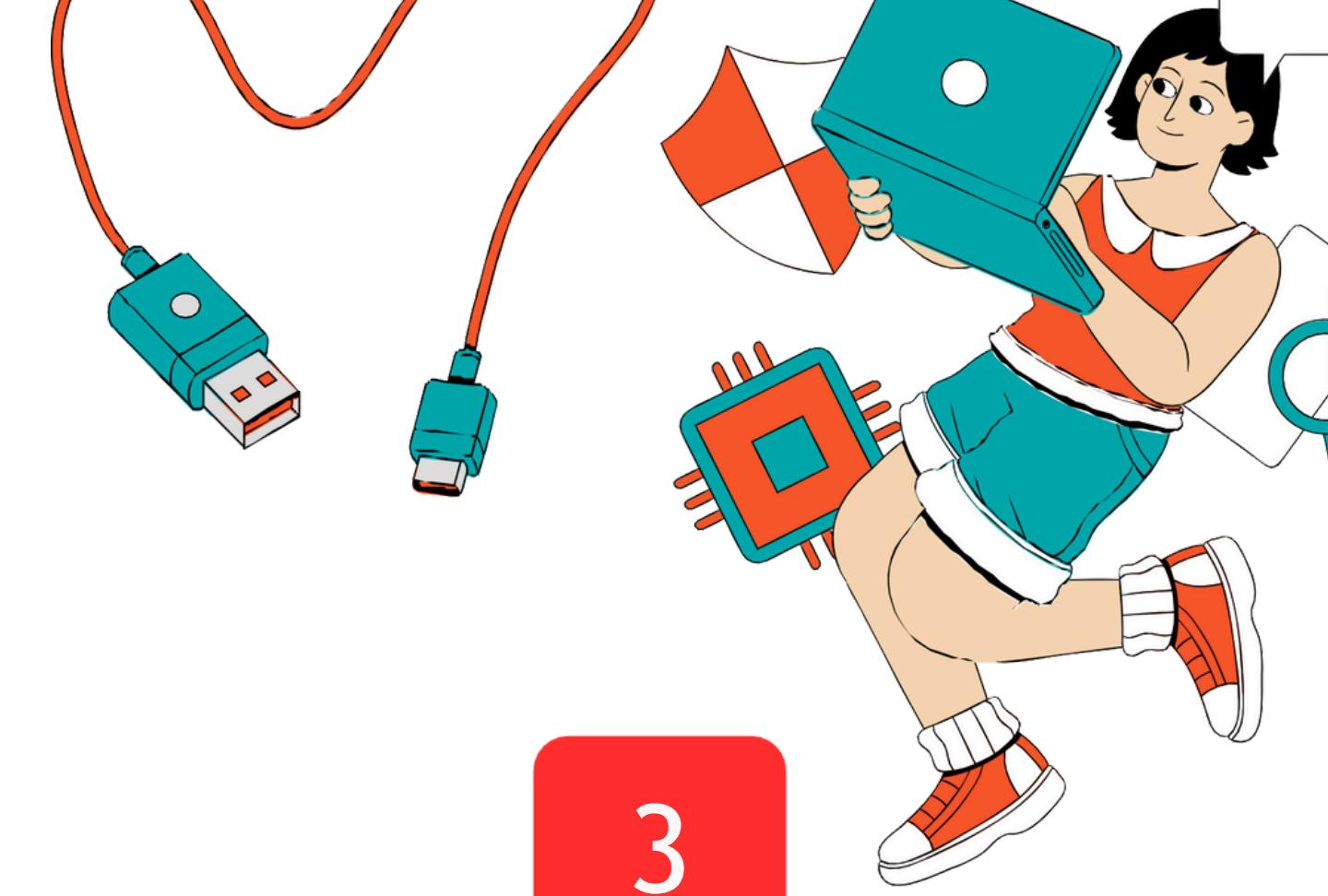
3 Scaling fitur numerik

Variabel dengan skala besar memiliki pengaruh lebih besar dibanding skala kecil, meskipun kedua variabel sama pentingnya. Oleh karena itu, menyamakan skala variabel agar tidak ada fitur mendominasi analisis

4 encoding fitur kategorik

karena machine learning tidak dapat mengolah data kategorik maka dari itu kita perlu mengubahnya menjadi numerik

Knowledge from data



1

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 bekerja maka semakin kecil risiko gagal bayar

2

PENDAPATAN DAN KEMAMPUAN FINANSIAL

Fitur memiliki anak, Pendapatan, dan besar pinjaman membantu mengukur kemampuan peminjam

3

KEPEMILIKAN ASET DAN JEJAK KREDIT

kepemilikan properti (flag_own_car, flag_own_realty) menjadi indikator kuat untuk memahami seseorang gagal bayar

MODELING

MODEL YANG DICOBAB

- 1. Logistic Regression
 Tanpa balancing
- 2. Logistic Regression
 Oversampling
- 3. Logistic
- 3. Catboost dengan
 smote

EVALUASI MODEL

Akurasi Terbaik ialah algoritma Catboost

Akurasi testing sebesar 92% dan akurasi

- training sebesar 95% yang menandakan tidak terjadinya overfitting kemudian model juga tidak bias ke kelas mayoritas karena sudah dilakukan balanced antar kelas menggunakan smote

Algoritma Logistik Regression

Pada algoritma ini, dilihat bahwa akurasi testing sebelum melakukan balanced data sangat baik

- tetapi hasilnya bias ke mayoritas sehingga tidak disarankan menggunakan model ini

Algoritma ini dengan data smote, dilihat bahwa akurasi testing setelah balanced data

- akurasi turun signifikan ini menandakan bahwa pelatihan sebelum smote bias ke kelas mayoritas

kesimpulan

- 1.algoritma catboost dengan menggunakan teknik smote menghasilkan model terbaik untuk memodelkan credit card approval
- 2.algoritma logistik tanpa melakukan balanced pada kelas minoritas menghasilkan bias ke kelas mayoritas
- 3.algoritma logistik dengan kelas balanced hasilnya turun signifikan perlu dilakukan kombinasi parameter dengan optuna/randomized

stem_model_eval_tts_test

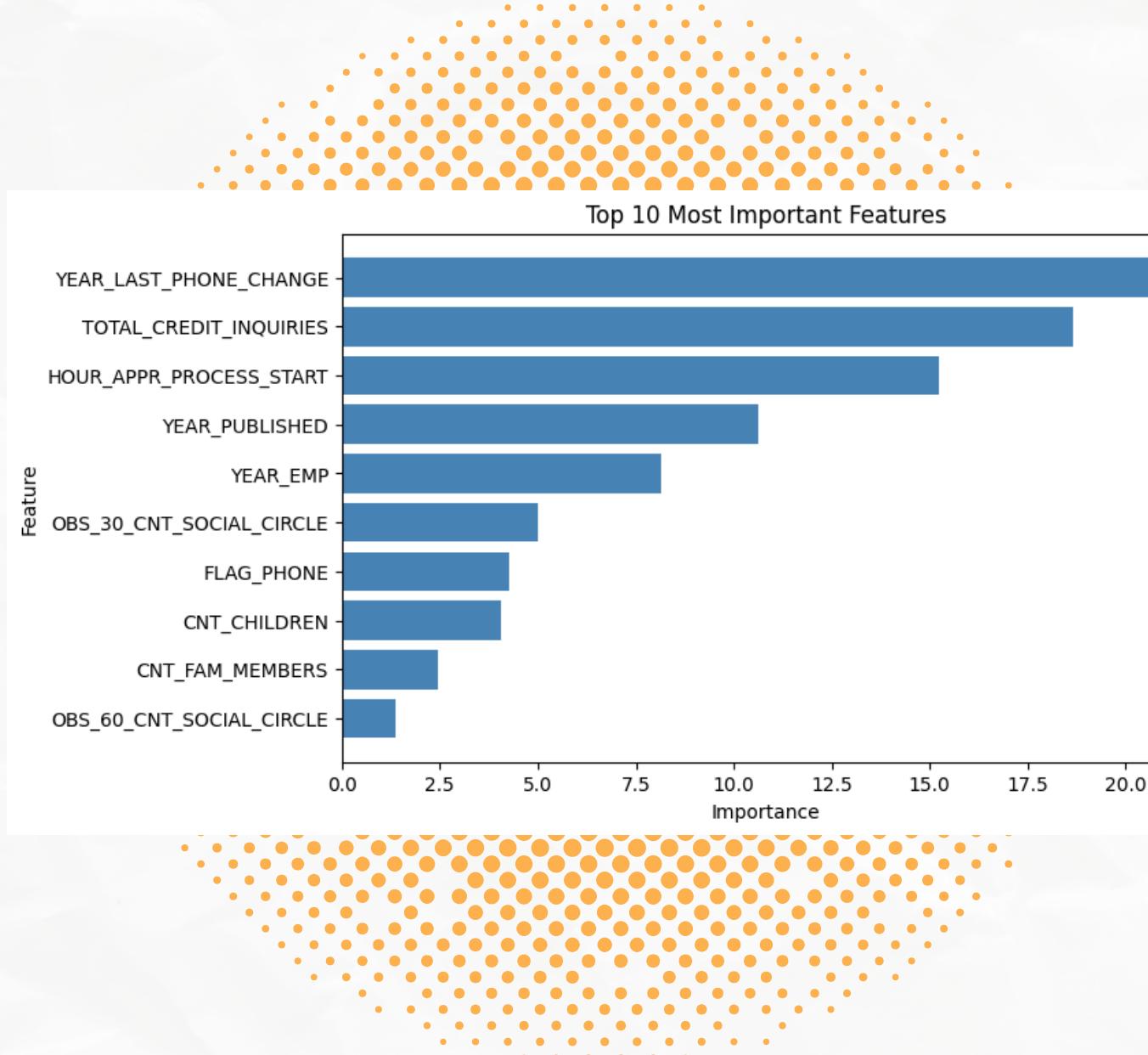
	Accuracy
Catboost Test 20%	92.0
Logistik Test 20%	91.9
Logistik Test smote 20%	70.9

stem_model_eval_tts_train

	Accuracy
Catboost Train 80%	95.9
Logistik Train 80%	91.9
Logistik Train smote 80%	73.3



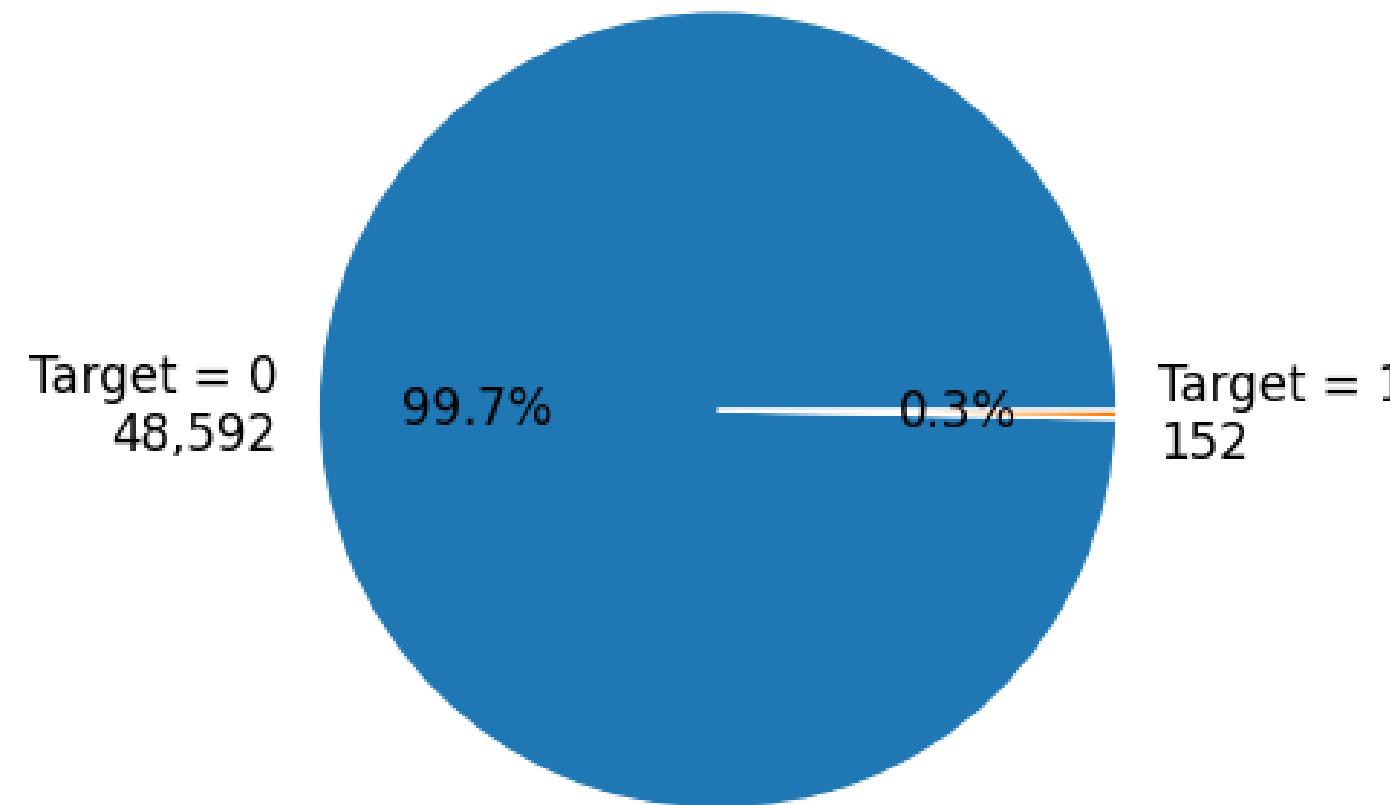
3 Fitur yang berpengaruh dalam pemodelan



- 1 Year last Phone Change, ini menunjukkan bahwa semakin sering berganti hape maka sangat berpengaruh dalam approval
- 2 Total_Credit_inquiries, fitur yang didapat dai feauture engineering amt_req_berau dari hour,day,week, month, qrt, year
- 3 Hour_appr_process_start, mengetahui perkiraan waktu pengajuan pinjaman Mendeteksi pola perilaku pengguna (apakah pengajuan malam hari punya risiko gagal bayar lebih tinggi)



Distribusi Nilai TARGET



implementasi model dalam prediksi

dengan menggunakan algoritma catboost diperoleh bahwa terdapat 48.592 pelamar yang diindikasi tidak gagal bayar sehingga dapat dilakukan approval, sedangkan sebanyak 152 peserta tidak layak mendapatkan approval sesuai request yang diajukan karena diprediksi gagal bayar

Target 0 → memiliki tidak memiliki kendala pembayaran sehingga dapat diapproval

Target 1 → pelanggan yang diprediksi sulit bayar selama periode meminjaman

Recommendation and Business Insight



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. Lama penggunaan telepon Sangat Berpengaruh

Temuan:

Fitur terkait flag_phone & year_last_phone_change memiliki pengaruh besar dalam model prediksi.

Action:

Jika lama penggunaan telepon cepat tetapi faktor lain mendukung, dapat diberikan opsi pinjaman lebih rendah dari yang diajukan.

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 lamanya penggunaan telepon → dapat menggunakan data ini yang sering diabaikan, data penggunaan telepon yang penting dalam pemodelan untuk meningkatkan akurasi prediksi selain faktor lain.

Thank You



[https://github.com/kevinsimorangkir2001/
ProyekAkhir_DataScientist_HCI_Rakamin](https://github.com/kevinsimorangkir2001/ProyekAkhir_DataScientist_HCI_Rakamin)



[https://www.linkedin.com/in/kevin-
caesarr/](https://www.linkedin.com/in/kevin-caesarr/)

