



Prediksi Nasabah Potensial: Solusi Data-Driven untuk Mempercepat Seleksi Awal Pengajuan Kredit

Dino Kuning

- Nadhilah Farhana
- Raditya Satria Gantara



Tim Member

Dino Kuning



Nadhilah Farhana

Data Scientist



Raditya Satria G.

Project Manager

Rangkaian Pembahasan



1

Penjelasan Home Credit

2

Tantangan Saat Ini

3

Tujuan Penelitian

4

Dataset & Sumber Data

5

Alur Analisis

6

Hasil Evaluasi Model

7

Keunggulan Proses

8

Rekomendasi & Potensi Pengembangan

Apa itu Home Credit?



Home Credit adalah perusahaan pembiayaan berbasis teknologi yang menyediakan layanan kredit konsumen, terutama untuk pembelian barang elektronik, furnitur, hingga pinjaman tunai. Home Credit beroperasi di berbagai negara, termasuk Indonesia, dengan fokus pada **nasabah unbanked dan underbanked** — yaitu mereka yang belum terjangkau layanan perbankan formal.

Apa Tantangan yang Dihadapi Home Credit Saat Ini?

» Tantangan Proses Seleksi Awal Nasabah Secara Manual

Proses manual dalam menyeleksi calon nasabah memerlukan waktu dan sumber daya yang besar. Setiap aplikasi kredit harus diperiksa secara individual, yang meningkatkan risiko kesalahan manusia dan inkonsistensi dalam penilaian. Selain itu, metode manual sulit untuk mengakomodasi volume aplikasi yang besar secara efisien.



Home Credit Indonesia
2020: 4,7 Juta Pelanggan

Dengan meningkatnya jumlah aplikasi kredit, Home Credit dihadapkan pada kebutuhan untuk menyaring aplikasi secara cepat tanpa mengorbankan akurasi. Penyaringan yang lambat dapat menurunkan kepuasan pelanggan dan meningkatkan risiko kehilangan peluang bisnis. Sebaliknya, penyaringan yang tidak akurat dapat menyebabkan pemberian kredit kepada individu yang tidak memenuhi syarat, meningkatkan risiko kredit macet.

Dashboard Customer Risk

Number of Customer

307.507K

Number of Credit

184.21bn

Number of Income

51.91bn

Type Risk

☐ Normal Cust

☐ Risk Cust

Gender

☐ F

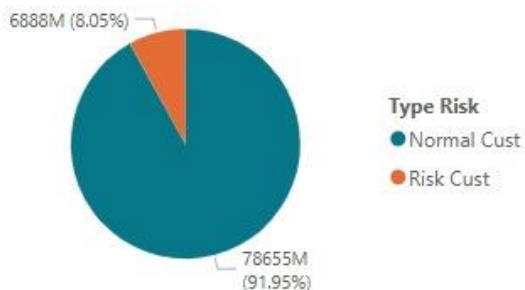
☐ M

Own Car

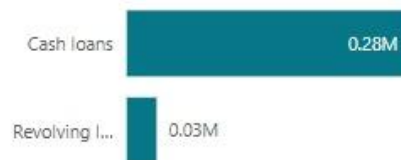
☐ N

☐ Y

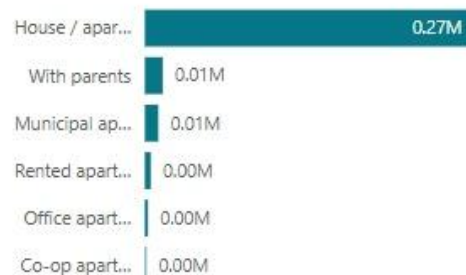
Customer by Risk Type



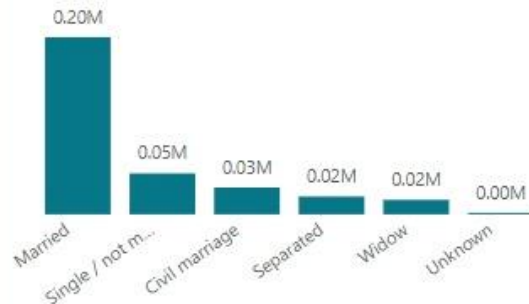
Customer by Payment Type



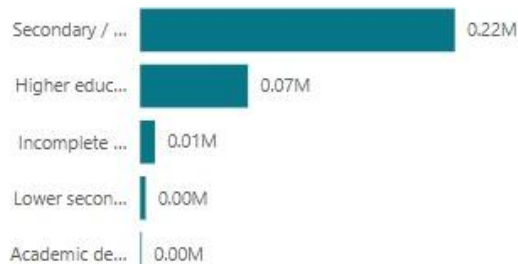
Customer by Housing Type



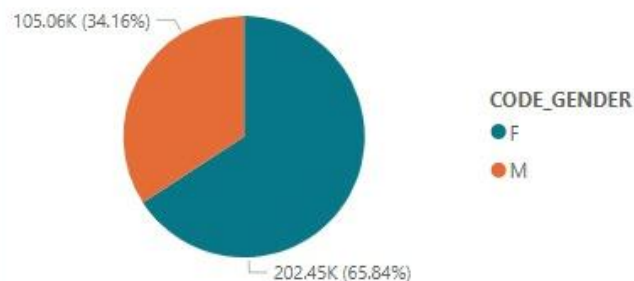
Customer by Status



Customer by Education



Customer by Gender

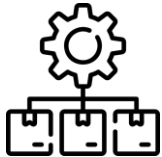


Tujuan Penelitian



Mengembangkan model prediksi nasabah potensial berbasis data

membangun model prediktif berbasis data historis dan perilaku nasabah menggunakan teknik *machine learning* sehingga Home Credit dapat mengurangi risiko gagal bayar dan meningkatkan efisiensi dalam proses pemberian pinjaman.



Mengautomasi proses seleksi awal yang efisien dan objektif

Otomatisasi ini memungkinkan penyaringan awal terhadap calon peminjam berdasarkan skor prediksi pembayaran, sehingga hanya nasabah dengan potensi baik yang melanjutkan ke tahap verifikasi lebih lanjut. Hasilnya adalah proses yang lebih cepat, efisien, dan dapat diandalkan untuk skala besar.



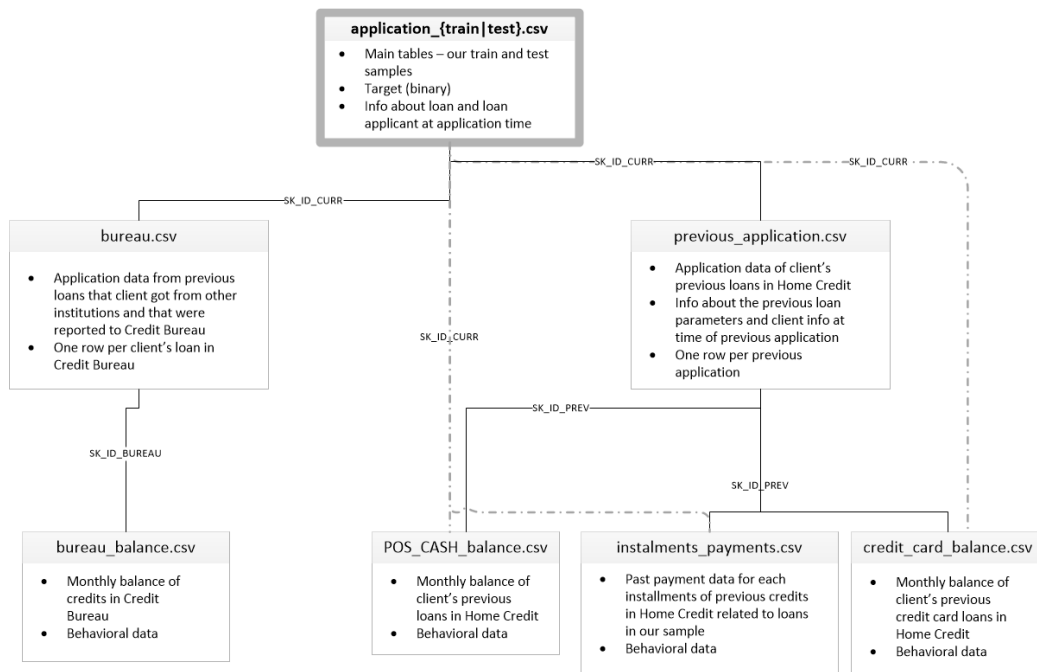
Menyediakan rekomendasi strategis berbasis hasil prediksi

Model prediksi tidak hanya berfungsi untuk seleksi, tetapi juga memberikan wawasan strategis bagi manajemen. Misalnya, karakteristik demografis yang mendukung pembayaran tepat waktu. Dari sini, Home Credit dapat merancang strategi pemasaran, penetapan limit kredit, dan kebijakan pinjaman yang lebih tepat sasaran dan berbasis data.

Dataset & Sumber Data



- **application_train.csv / application_test.csv**
Data utama berisi informasi aplikasi pinjaman, termasuk demografi, pekerjaan, dan kondisi keuangan nasabah.
- **bureau.csv & bureau_balance.csv**
Riwayat kredit nasabah dari lembaga keuangan lain.
- **previous_application.csv**
Data aplikasi pinjaman sebelumnya di Home Credit.
- **POS_CASH_balance.csv, credit_card_balance.csv, installments_payments.csv**
Aktivitas pembayaran dari berbagai produk keuangan seperti kartu kredit dan cicilan.



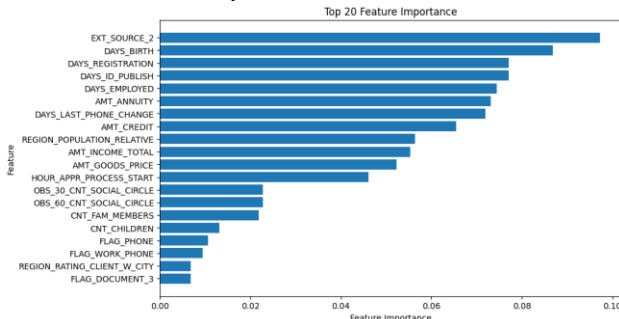
Alur Analisis

1 Data Cleaning & Preprocessing

- Menghapus kolom dengan missing values >10%
- Feature Engineering:
 - Ubah satuan data negatif menjadi positif (DAYS_BIRTH)
 - Menambahkan kolom DAYS_EMPLOYED_ANOM
- Encoding data kategorik

4 Feature Selection

Memilih Top 20 berdasarkan hasil feature importance



2 Train-Test Split

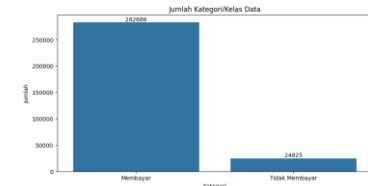
80% data train,
20% data test

5 Scaling

Data train: fit_transform,
Data test: transform

3 Oversampling

Hanya diterapkan untuk data train



6 Training Model

XGBoost, Random Forest,
Voting Ensemble

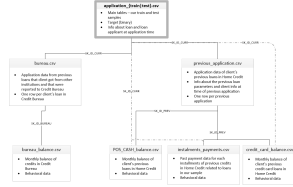
7 Testing Model & Evaluasi

Accuracy, ROC-AUC, Precision,
Recall

Hasil Evaluasi Model

Skema 1

Merge semua dataset
Ukuran data: 307511×150



| Model | Akurasi | ROC-AUC | Recall (0) | Precision (0) |
|--------------------|---------|---------|------------|---------------|
| XGBoost | 92% | 52% | 99% | 92% |
| Random Forest (RF) | 92% | 50% | 100% | 92% |

Skema 2

Data application train
Ukuran data: 307511×20

| Model | Akurasi | ROC-AUC | Recall (0) | Precision (0) |
|--------------------|---------|---------|------------|---------------|
| XGBoost | 93% | 50% | 100% | 93% |
| Random Forest (RF) | 93% | 50% | 100% | 93% |

- **Recall 100%**: semua nasabah yang benar-benar mampu bayar (kelas 0) berhasil terdeteksi oleh model
- **Precision 92–93%** : mayoritas dari prediksi “mampu bayar” memang benar-benar mampu bayar, sehingga sangat efisien untuk skrining awal
- **Akurasi juga tinggi (92–93%)**, menunjukkan bahwa model dapat diandalkan untuk seleksi otomatis nasabah potensial
- **Nilai ROC-AUC masih rendah**, artinya kemampuan model untuk membedakan antara nasabah bayar dan tidak bayar secara umum masih kurang tajam
- **Skema 2 lebih efisien** karena hanya menggunakan fitur dari application_train, tetapi tetap memberi hasil recall dan precision maksimal

Keunggulan Proses

- ✓ Efisiensi waktu & biaya proses seleksi awal
- ✓ Meningkatkan rasio approval berkualitas
- ✓ Mengurangi risiko penyaluran kredit ke nasabah tidak layak
- ✓ Mendukung keputusan bisnis yang lebih presisi

Data Historis Bisa Membantu Mengenali Pola Nasabah Potensial

Pemanfaatan data historis memungkinkan perusahaan untuk mengidentifikasi pola dan karakteristik nasabah potensial. Dengan menganalisis data sebelumnya, perusahaan dapat mengembangkan model prediktif yang membantu dalam **proses seleksi nasabah secara lebih efisien dan akurat**. Hal ini sejalan dengan upaya pengelolaan risiko kredit yang bertujuan untuk **memprofilkan nasabah potensial secara lebih baik**.



*"Dengan strategi yang kuat yang dapat beradaptasi dengan segala perubahan gaya hidup dan pola belanja masyarakat, kami dapat **meningkatkan pertumbuhan pelanggan** kami ... Home Credit Indonesia senantiasa memperkuat kepercayaan pelanggan **tidak hanya melalui layanan digital, namun disertai peningkatan kualitas layanan** ... di seluruh wilayah di Indonesia," jelas **Animesh – Direktur Utama Home Credit Indonesia**.*

Rekomendasi Bisnis & Potensi Pengembangan

✓ Implementasi Model sebagai Screening Tool Awal

✓ Integrasi ke Sistem Front-End / Aplikasi Sales

Model dapat diintegrasikan langsung ke dalam **sistem aplikasi pengajuan kredit**, baik oleh tim sales maupun sistem online yang dapat memberikan **prediksi real-time** apakah nasabah termasuk kategori “layak lanjut”.

✓ Penyesuaian Proses Bisnis Berdasarkan Prediksi

- **Fast-track approval**: untuk nasabah yang diprediksi mampu bayar (kelas 0).
- **Review manual lebih lanjut**: untuk nasabah yang tidak yakin diprediksi layak.

✓ Monitoring & Evaluasi Berkala

Meskipun model saat ini menunjukkan performa sangat baik, tetap diperlukan **monitoring performa secara periodik**:

- Pantau apakah recall dan precision tetap tinggi di data baru.
- Lakukan retraining jika terjadi perubahan tren atau profil nasabah.

Penggunaan Model untuk Segmentasi Produk

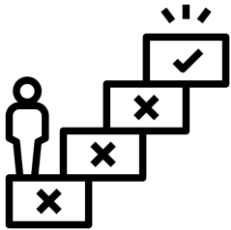
- Model prediksi tidak hanya berguna untuk menyaring siapa yang layak, tapi juga bisa diperluas untuk **segmentasi penawaran produk**.
- Nasabah yang diprediksi **sangat layak** bisa ditawarkan:
 - Limit kredit lebih besar
 - Tenor lebih panjang
 - Produk premium (misal: cicilan barang mahal atau bundling produk)
- Segmentasi berbasis prediksi akan mendorong **efisiensi penawaran** dan meningkatkan **conversion rate serta loyalitas nasabah**.



Kesulitan dan Kendala



- Total dataset yang digunakan ukurannya tergolong besar, yaitu 2,47 GB
- Variabel cukup banyak (221), sehingga butuh effort yang besar untuk mengulik insight dari datanya



- Sulit membagi waktu antar anggota tim dan menemukan waktu yang tepat untuk berdiskusi dengan tim
- Kehilangan 2 anggota tim sejak stage awal
- Keterbatasan memori laptop untuk running data yang besar, sehingga membutuhkan waktu yang lebih lama untuk menganalisis keseluruhan data

Terima Kasih