



MEMPREDIKSI RISIKO GAGAL BAYAR KREDIT RUMAH

PENDEKATAN ILMU DATA UNTUK MENDUKUNG KEPUTUSAN PEMBERIAN KREDIT



LATAR BELAKANG

Masalah

Industri perbankan perlu memprediksi risiko kredit secara akurat. Home Credit menghadapi tantangan dalam menilai peminjam tanpa riwayat kredit, sehingga menggunakan kumpulan data yang komprehensif untuk membangun model prediktif yang lebih presisi daripada metode konvensional.

Proyek ini bertujuan untuk mengurangi utang macet dan memperluas akses kredit. Dengan menggunakan metode pembelajaran mesin seperti jaringan saraf tiruan dan metode ansambel, pola data yang kompleks diubah menjadi kecerdasan untuk mengoptimalkan keputusan kredit.

PERNYATAAN MASALAH

Bagaimana kita dapat secara akurat memprediksi kemungkinan gagal bayar pinjaman menggunakan informasi pemohon dan data kredit historis?



TUJUAN

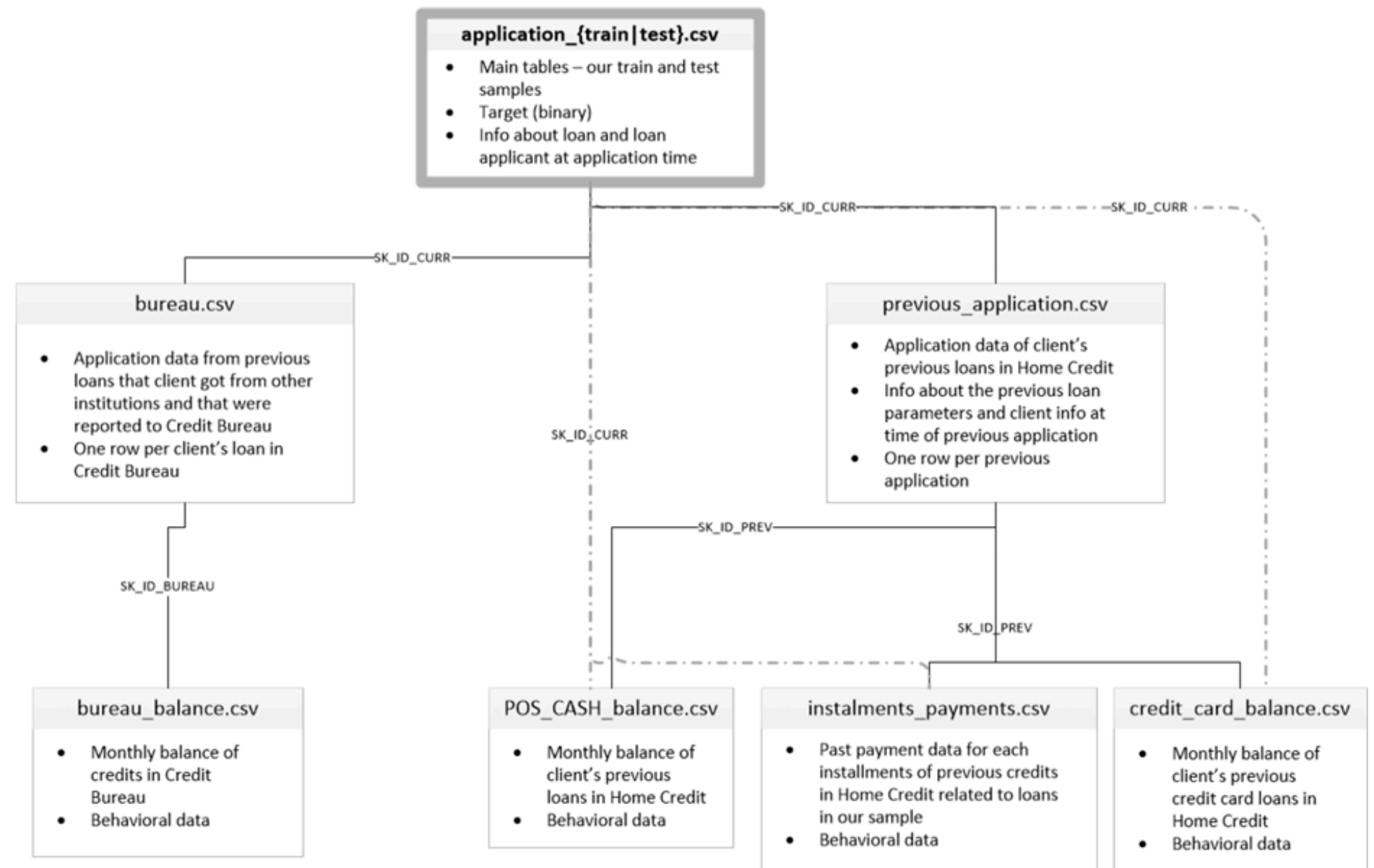
- 01 Identifikasi peminjam potensial yang berisiko gagal bayar sejak dini.
- 02 Tingkatkan akurasi prediksi untuk meminimalkan risiko.
- 03 Memberikan rekomendasi praktis untuk mendukung pengambilan keputusan bisnis.



RINGKASAN DATASET

Home Credit Dataset

- **Sumber data:** Home Credit (beberapa tabel: application, bureau, credit card balance, previous application, Pos Cash Balance, and installments payments).
- **Data Latih:** 307511 data.
- **Data Test :** 48744 data.
- **Total ukuran dataset:** 2559.36 MB



METODOLOGI

Pendekatan Analitis



01 Data Preprocessing

- **Missing value imputation** (median for numeric, 'MISSING' for categorical)
- **Feature engineering** (Age, Income ratios, Credit utilization)
- **Multi-table aggregation** for enriched features

02 Feature Engineering

- **Derived variables:** AGE, YEARS_EMPLOYED, INCOME_PER_FAMILY
- **Financial ratios:** CREDIT_TO_INCOME, ANNUITY_TO_INCOME
- **Bureau aggregates:** Historical credit behavior metrics

03 Model Strategy

- **Traditional ML** (Random Forest, LightGBM)
- **Deep Learning** (Simple NN, Deep NN)
- **Hyperparameter tuning** for optimization

04 Hyperparameter Tuning

- GridSearchCV & RandomizedSearchCV

05 Model Evaluation

- AUC, Accuracy, Precision, Recall, F1

HASIL ANALISIS

01

Gambaran Risiko Dasar

Secara keseluruhan, kami menemukan bahwa rata-rata nasabah Home Credit memiliki tingkat **gagal bayar** (tidak mampu melunasi pinjaman) sekitar **8.07%**. Ini adalah angka dasar yang perlu kita coba kurangi melalui sistem prediksi yang lebih baik.

02

Segmen Nasabah Paling Berisiko

Analisis data menunjukkan bahwa risiko tidak merata. Dua kelompok nasabah yang paling berisiko tinggi adalah:

- Nasabah yang mencantumkan tipe pendapatan sebagai '**Maternity leave**' (**Cuti Hamil**), yang memiliki tingkat **gagal bayar sangat tinggi**, yaitu **40.00%**.
- Kelompok **usia muda (0 hingga 25 tahun)** memiliki tingkat gagal bayar tertinggi di antara semua kelompok usia, mencapai **12.29%**.

HASIL ANALISIS

03 Faktor Kunci Penentu Risiko (EXT_SOURCE)

Kami menemukan bahwa **faktor** terpenting yang menentukan apakah seorang peminjam akan gagal bayar atau tidak adalah **Skor Kredit Eksternal (diwakili oleh fitur EXT_SOURCE_1, 2, dan 3)**. **Semakin baik skor** eksternal ini, **semakin kecil** kemungkinan peminjam tersebut mengalami **gagal bayar**.

04 Kinerja Model Terbaik

Dari berbagai model prediksi yang dibangun (termasuk Random Forest dan Neural Networks), **model LightGBM** menunjukkan **kinerja prediksi risiko** yang **paling unggul**. Model ini mencapai skor akurasi prediksi risiko (**AUC**) **0.7628**, menjadikannya alat yang paling andal untuk mengidentifikasi peminjam berisiko.

REKOMENDASI

1. Prioritaskan Data Eksternal

Karena **Skor Kredit Eksternal (EXT_SOURCE)** terbukti menjadi **prediktor terpenting**, perusahaan harus memastikan bahwa **data ini selalu terintegrasi, valid, dan up-to-date** dalam proses **persetujuan kredit**.

2. Strategi Pengelolaan Segmen Berisiko

Perlu ada kebijakan yang lebih ketat atau penyesuaian penawaran untuk segmen yang sangat berisiko.

- Untuk **pemohon usia muda (0-25 tahun)**, pertimbangkan untuk **menawarkan limit pinjaman yang lebih kecil** atau mengenakan **suku bunga yang sedikit lebih tinggi** untuk **mengimbangi risiko** yang ada.
- Perlu **peninjauan ulang** terhadap **pemohon** dengan status pendapatan '**Maternity leave**', mengingat **tingkat gagal bayar** mereka yang **sangat tinggi**.

3. Optimalisasi Keputusan Otomatis

Saat **mengimplementasikan model terbaik (LightGBM)**, kita perlu **hati-hati** dalam **menentukan batas ambang prediksi risiko**. Meskipun LightGBM akurat secara keseluruhan (**AUC 0.7628**), saat ini model cenderung lebih fokus untuk mengidentifikasi sebanyak mungkin **kasus gagal bayar (Recall tinggi)** daripada benar-benar yakin bahwa prediksi gagal bayar itu benar (**Precision rendah, 0.1893**). Kita harus **mengatur ambang batas** ini untuk **menyeimbangkan** antara **meminimalkan kerugian** (dengan menolak pemohon yang benar-benar berisiko) dan tidak **kehilangan terlalu banyak nasabah potensial yang baik**.

KESIMPULAN

01.

Capaian Utama

Proyek ini berhasil mengidentifikasi dan membangun model prediksi risiko pinjaman terbaik. Model LightGBM unggul dibandingkan semua metode lain, termasuk model Random Forest dan jaringan saraf tiruan yang lebih kompleks (Deep NN)

02.

Insight Kritis

Analisis ini mengkonfirmasi bahwa data eksternal dan usia adalah pendorong utama risiko gagal bayar

03.

Perbandingan Model

Menariknya, kami menemukan bahwa model Machine Learning tradisional yang canggih seperti LightGBM masih mengungguli (lebih akurat) daripada model Deep Learning yang rumit untuk jenis data nasabah ini, meskipun Deep Learning mampu menangkap pola non-linear yang kompleks



TERIMA KASIH

● ATAS PERHATIAN YANG BAIK