

PREDIKSI RISIKO GAGAL BAYAR PINJAMAN

Analisis Data & Implementasi Machine Learning
untuk Home Credit

Syamsul Maarip



https://github.com/syamsulmaarip05/projek_Rakamin.git



LATAR BELAKANG

1. **Masalah (Problem):** Tingkat gagal bayar (kredit macet) menyebabkan kerugian finansial yang signifikan bagi perusahaan.
2. **Tantangan:** Proses persetujuan pinjaman perlu membedakan antara klien berisiko tinggi dan klien berisiko rendah secara akurat dan efisien.
3. **Tujuan (Goal):**
 - Membangun model machine learning untuk memprediksi probabilitas klien akan gagal bayar (TARGET=1).
 - Memberikan rekomendasi bisnis untuk mengoptimalkan proses persetujuan pinjaman.

DATA SET

Sumber Data: 8 tabel data relasional (CSV) yang mencakup:

- application_train.csv: Data aplikasi pinjaman utama (307.511 baris).
- bureau.csv & bureau_balance.csv: Riwayat kredit klien di lembaga lain.
- previous_application.csv: Riwayat pinjaman klien sebelumnya di Home Credit.
- Data Perilaku (POS_CASH, credit_card, installments_payments).

INSIGHT 1

Insight: Perilaku pembayaran angsuran di masa lalu adalah prediktor terkuat. Klien yang telat bayar (DAYS_PAST_DUE) atau kurang bayar (PAYMENT_DIFF) pada pinjaman sebelumnya memiliki risiko gagal bayar yang jauh lebih tinggi.

Rekomendasi Aksi (Action):

- Untuk Persetujuan Pinjaman: Jadikan fitur DAYS_PAST_DUE dan PAYMENT_DIFF sebagai fitur utama dalam model scoring.
- Untuk Manajemen Klien: Gunakan data ini untuk sistem "Peringatan Dini" (Early Warning) pada klien aktif untuk mencegah keterlambatan.

INSIGHT 2

Insight: Skor dari sumber eksternal (EXT_SOURCE_1, EXT_SOURCE_2, EXT_SOURCE_3) memiliki korelasi yang sangat kuat dengan kemampuan bayar. Klien dengan skor eksternal yang rendah cenderung gagal bayar.

Rekomendasi Aksi (Action):

- **Investasi Data:** Perusahaan harus terus berinvestasi untuk mendapatkan data skor eksternal yang akurat dan up-to-date.
- **Segmentasi:** Buat jalur "Jalur Hijau" (Fast-Track) untuk persetujuan otomatis bagi klien dengan skor eksternal di atas threshold tertentu.

PROSES PEMBUATAN MMODELA

1. **Feature Engineering:** Menggabungkan 8 tabel data menjadi 1 dataset utama melalui agregasi (mean, max, sum) per SK_ID_CURR.
2. **Feature Selection (Tahap 1):** Menghapus kolom-kolom yang memiliki missing values lebih dari 60% untuk mengurangi noise.
3. **Data Cleaning (Tahap 2):** Melakukan One-Hot Encoding pada sisa fitur kategorikal (mengubah teks menjadi angka).
4. **Penanganan Imbalance:** Menerapkan SMOTE (Synthetic Minority Over-sampling) pada data training untuk menyeimbangkan rasio 92:8.

PROSES PEMBUATAN MODEL

1. **Eksperimen:** Beberapa model dievaluasi (termasuk Regresi Logistik).
2. **Model Final:** LightGBM (LGBMClassifier).
3. **Alasan Pemilihan:** LightGBM adalah algoritma Gradient Boosting yang dikenal sangat kuat, cepat, dan efektif untuk data tabelar yang besar dan imbalance seperti ini.
4. **Tuning:** Model disempurnakan menggunakan GridSearchCV pada data SMOTE untuk menemukan hyperparameter terbaik (misal: `n_estimators`, `learning_rate`) yang memaksimalkan ROC AUC.

PERFORMA MODEL

--- Hasil Model Terbaik (F1-Score Optimal) ---

Classification Report (Optimal):

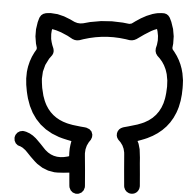
	precision	recall	f1-score	support
Lunas (0)	0.95	0.88	0.91	56538
Gagal Bayar (1)	0.23	0.42	0.30	4965
accuracy			0.84	61503
macro avg	0.59	0.65	0.60	61503
weighted avg	0.89	0.84	0.86	61503

REKOMENDASI BISNIS

1. **Interpretasi Hasil:** Model ini memiliki Recall 42% (mampu menangkap 42% kasus gagal bayar) namun Presisi 23% (hanya 1 dari 4 yang ditandai benar-benar gagal bayar).
2. **Masalah:** Presisi 23% terlalu rendah untuk Penolakan Otomatis (akan banyak menolak pelanggan baik).
3. **Rekomendasi Utama:** Implementasikan sebagai Sistem Pendukung Keputusan.
4. **Jalur Hijau (Aman):** Prediksi "Lunas (0)" -> Persetujuan dipercepat.
5. **Jalur Merah (Berisiko):** Prediksi "Gagal Bayar (1)" -> Otomatis dialihkan ke Tinjauan Manual (Manual Review) oleh tim credit analyst.

KESIMPULAN

Model LightGBM yang diusulkan mampu mengurangi risiko kredit dengan cara mengidentifikasi 42% potensi gagal bayar untuk ditinjau lebih lanjut. Dampak: Meningkatkan efisiensi tim credit analyst dengan memfokuskan waktu mereka pada aplikasi yang paling berisiko.



https://github.com/syamsulmaarip05/projek_Rakamin.git

TERIMA KASIH
