

HOME CREDIT INDONESIA

Vania Zerlina Utami

Table of **CONTENTS**

- Problem Research
- Dataset
- Data Visualization and Business Insight
- Data Pre-Processing
- Machine Learning Implementation and Evaluation
- Business Recommendation

Problem **RESEARCH**

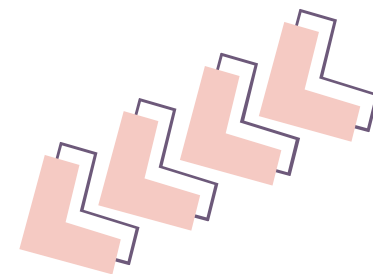
Perilaku Keuangan yang Terus Bergerak

Setiap pelanggan punya pola finansial yang unik, dan pola itu bisa berubah seiring waktu. Inilah tantangan: bagaimana membaca perubahan itu lebih cepat daripada kompetitor?



Risiko vs Peluang: Dua Sisi Mata Uang

Terlalu ketat, peluang bisnis hilang. Terlalu longgar, risiko gagal bayar meningkat. Menemukan titik seimbang antara keduanya bukan perkara mudah.



Efek Domino dari Keterlambatan Pembayaran

Satu keterlambatan bisa memicu rantai masalah: arus kas terganggu, biaya penagihan naik, hingga potensi kerugian jangka panjang.

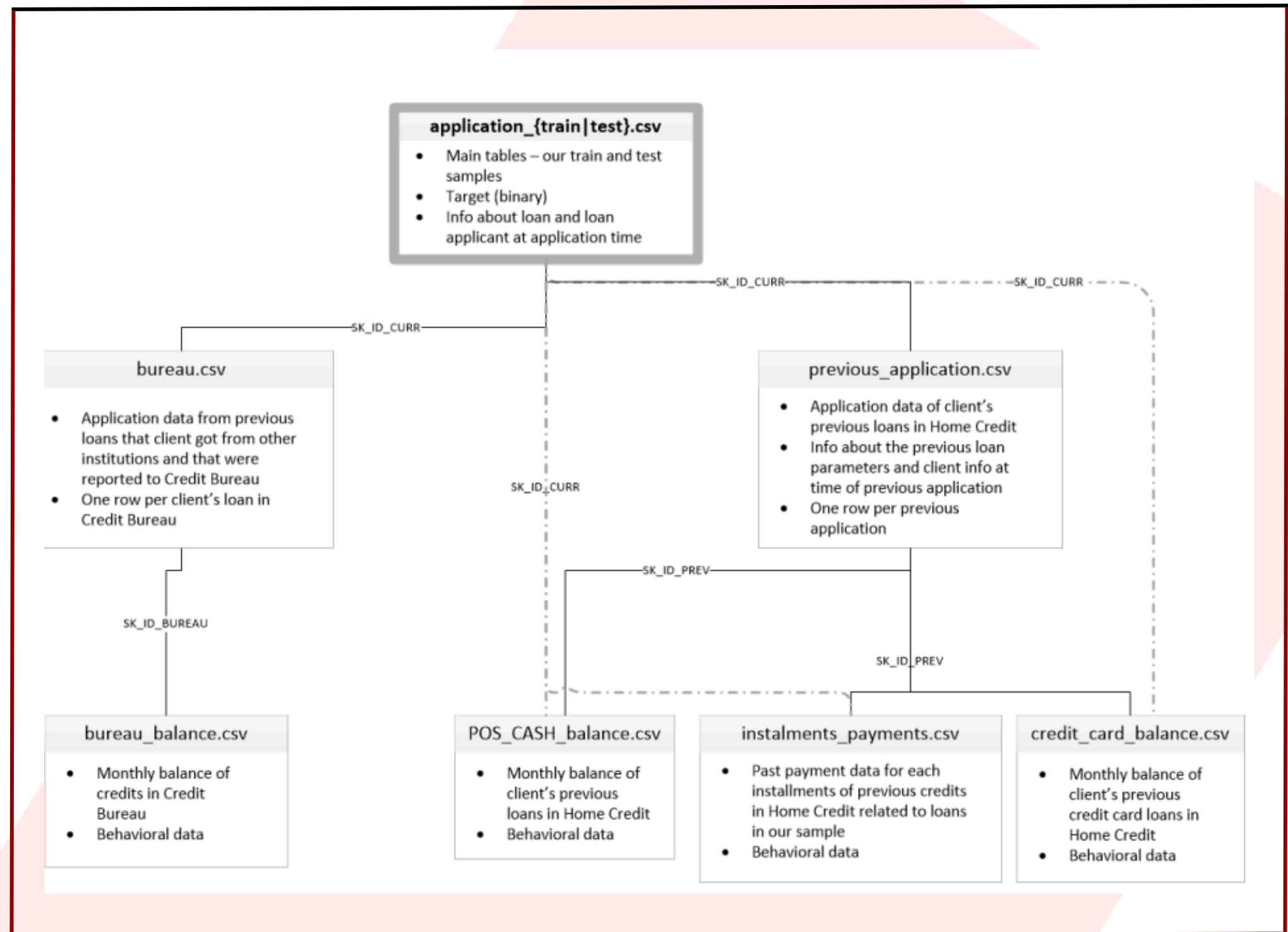


Butuh Model yang Lebih Gesit & Presisi

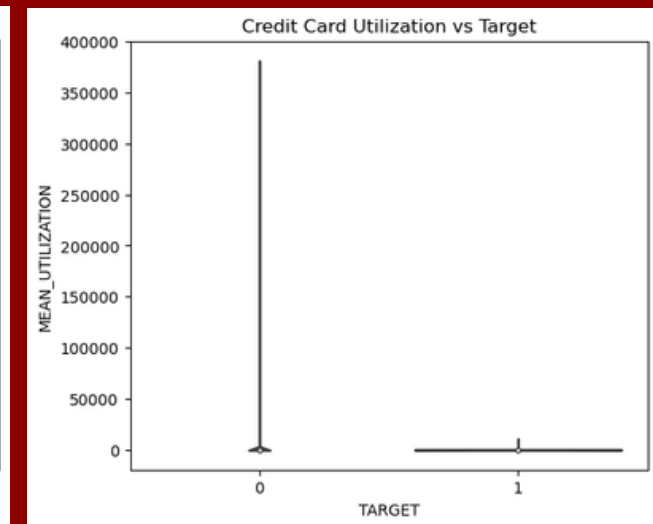
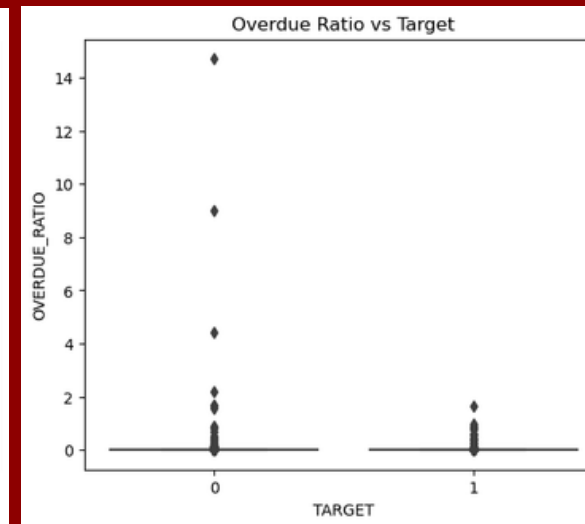
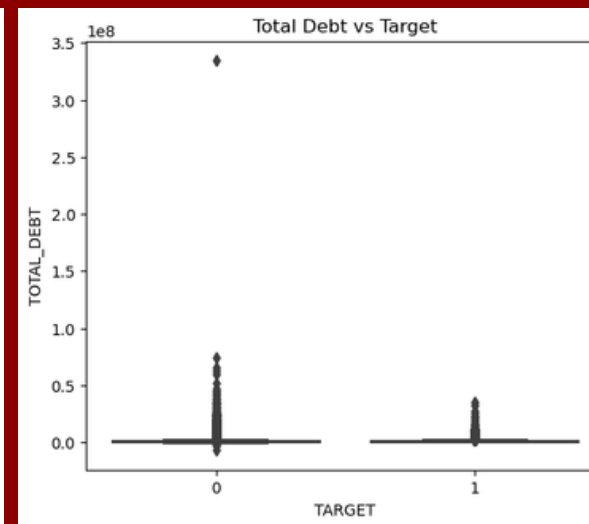
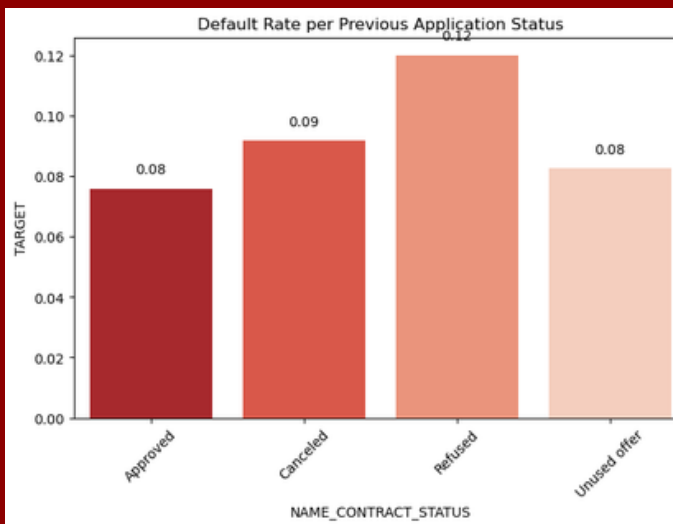
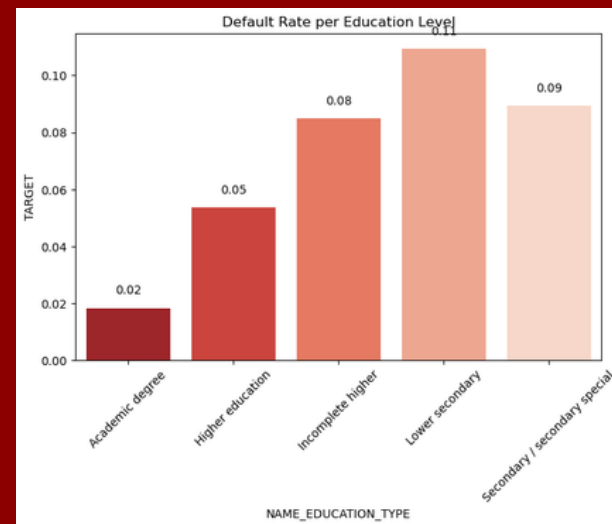
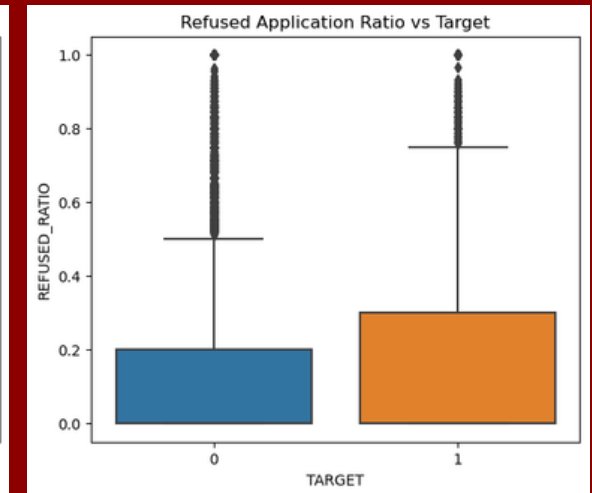
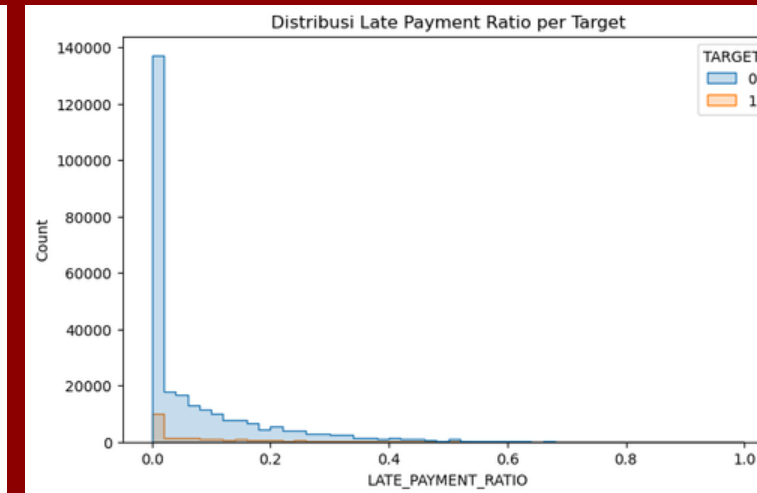
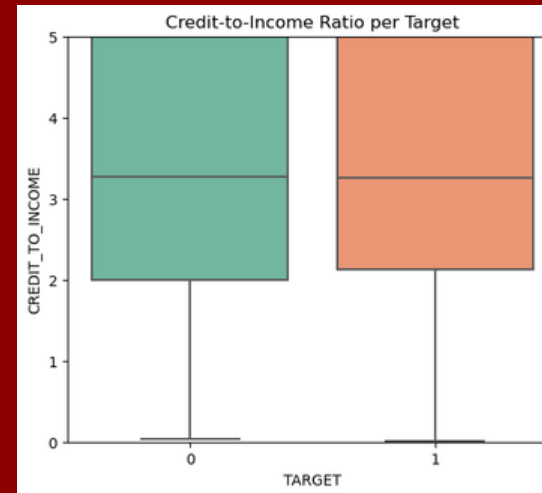
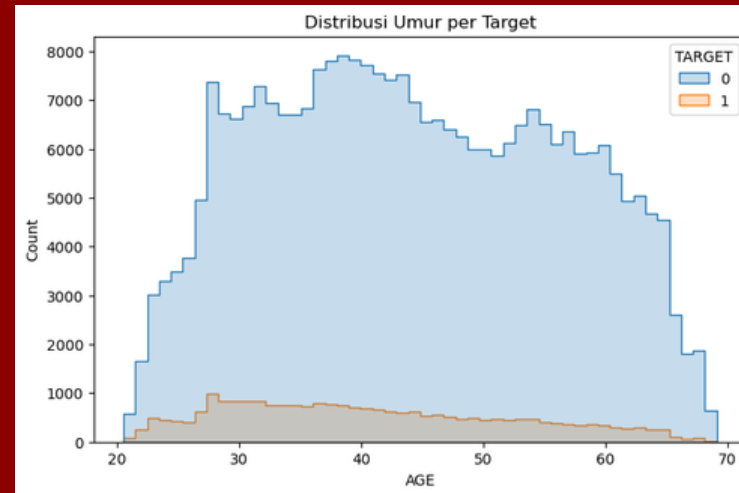
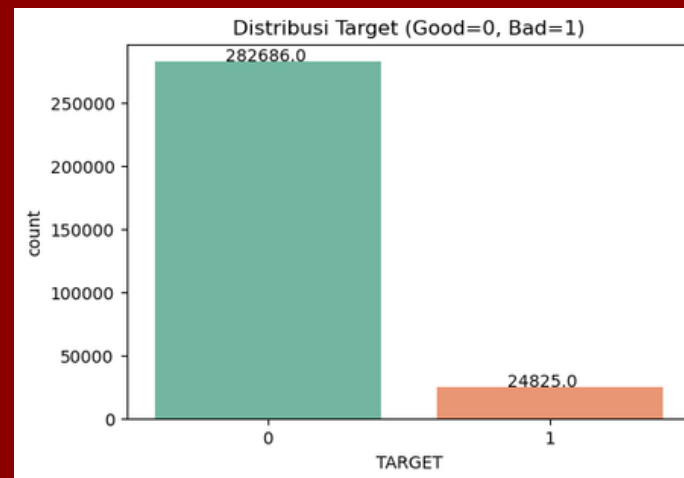
Solusinya? Model prediksi yang tak hanya akurat, tapi juga gesit mengikuti perubahan tren. Dengan begitu, risiko bisa ditekan, peluang bisa dimaksimalkan, dan hubungan dengan pelanggan tetap terjaga.

DATASET

- Application (Train/Test) – Data utama tiap pinjaman.
- Bureau & Bureau Balance – Riwayat kredit sebelumnya & saldo bulanan.
- POS_CASH & Credit Card Balance – Saldo pinjaman dan kartu kredit bulanan.
- Previous Application – Riwayat aplikasi kredit klien.
- Installments Payments – Riwayat cicilan dibayar / telat.
- Columns Description – Keterangan semua kolom.



DATA VISUALIZATION



BUSINESS INSIGHT

Profil Nasabah Berisiko Tinggi (High Risk)

Berdasarkan berbagai metrik yang dianalisis, nasabah dengan ciri-ciri berikut menunjukkan kecenderungan lebih tinggi untuk mengalami gagal bayar:

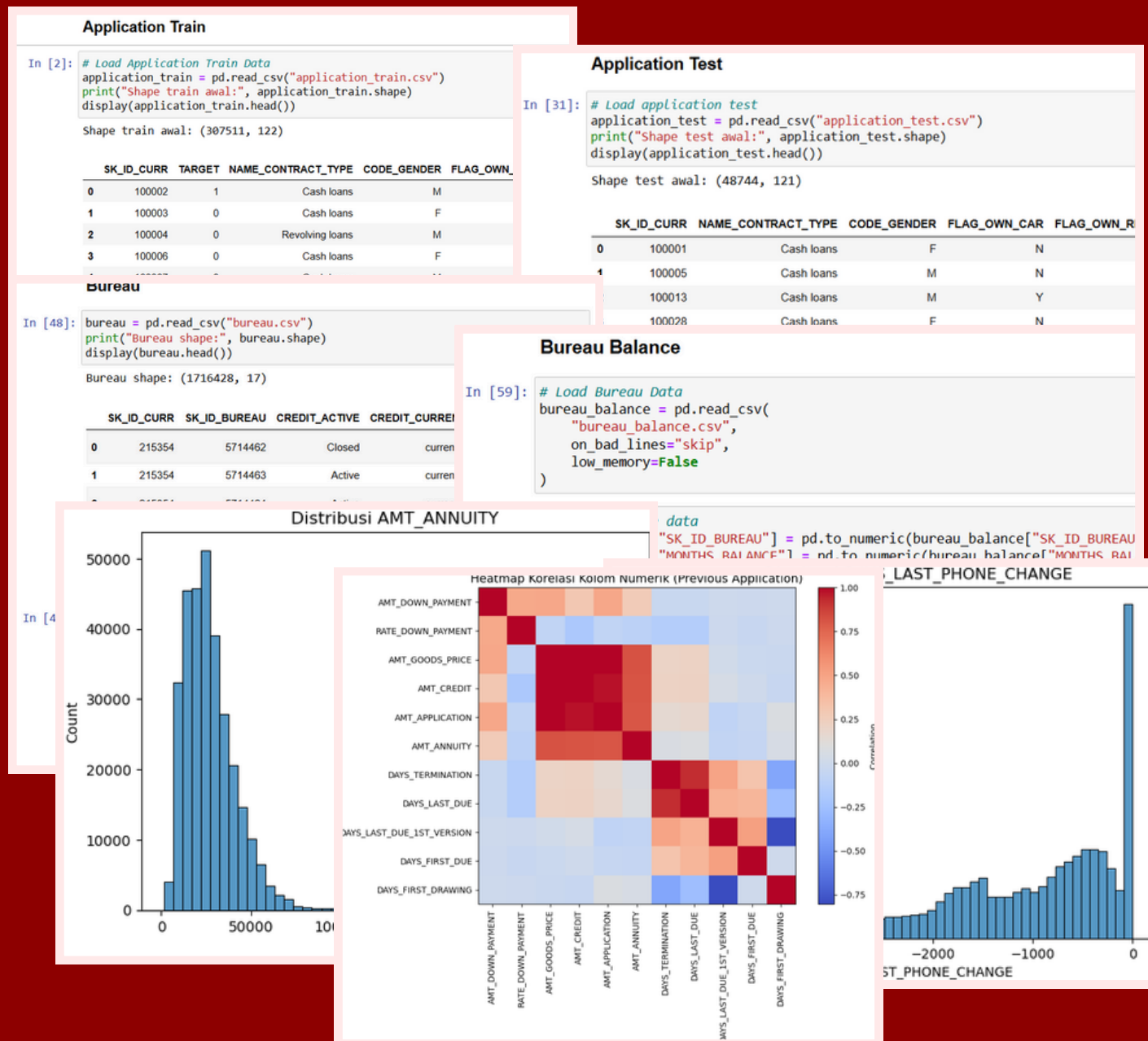
- Usia muda (khususnya < 40 tahun)
- Rasio kredit terhadap pendapatan (CTI) tinggi
- Tingkat pendidikan rendah (terutama "lower secondary" dan "secondary special")
- Riwayat pengajuan kredit yang ditolak sebelumnya
- Riwayat keterlambatan pembayaran
- Utilisasi kartu kredit tinggi
- Overdue ratio tinggi (meskipun jarang terjadi, tapi sangat relevan pada kasus default)

Rekomendasi Aksi

Untuk menangani nasabah yang berisiko tinggi seperti yang masih muda, memiliki rasio kredit terhadap penghasilan yang tinggi, tingkat pendidikan yang rendah, pernah ditolak saat mengajukan pinjaman, memiliki riwayat keterlambatan pembayaran, serta penggunaan kartu kredit dan tunggakan yang cukup besar, perusahaan sebaiknya memperketat proses persetujuan pinjaman. Langkah yang dapat dilakukan adalah membatasi jumlah pinjaman dan jangka waktu cicilan, serta memasukkan faktor-faktor risiko tersebut ke dalam sistem penilaian kredit. Setelah pinjaman disetujui, perlu dilakukan pemantauan secara aktif seperti pengingat sebelum jatuh tempo, mewajibkan pembayaran otomatis, dan menawarkan produk pinjaman yang lebih sederhana dan mudah dipahami. Selain itu, edukasi keuangan juga penting agar nasabah dapat mengelola pinjamannya dengan lebih bijak dan membantu mengurangi risiko gagal bayar.

DATA PRE-PROCESSING

- **Pemeriksaan awal dataset**
 - Melihat daftar kolom (`columns.tolist()`) dan bentuk data (`shape`) serta menghitung persentase missing values.
- **Penanganan missing values**
 - Jika $>50\%$ maka kolom di-drop, jika $\leq 50\%$ dilakukan imputasi berdasarkan distribusi data (mean, median, modus), skewness, tipe data, dan makna kolom.
- **Penggabungan dataset**
 - Mengecek pasangan `SK_ID_CURR` antar tabel, melakukan agregasi per `SK_ID_CURR`, lalu menggabungkan hasilnya dengan left join ke data utama (data train & test).
- **Pemeriksaan ulang missing values**
 - Mengecek kembali setelah join, lalu drop kolom dengan missing $>50\%$ dan lakukan imputasi untuk kolom dengan missing $\leq 50\%$.
- **Seleksi fitur**
 - Memilih top 100 fitur yang paling relevan terhadap target.



MODEL & EVALUATION

Random Forest menunjukkan performa yang sangat unggul dengan Accuracy 95%, Precision 63,5%, Recall 90,3%, F1 Score 74,5%, dan ROC-AUC 0,983. Model ini mampu mendeteksi risiko gagal bayar dengan sangat akurat, meminimalkan false alarm, dan memberikan prediksi yang dapat diandalkan untuk pengambilan keputusan kredit. Dengan performa tinggi ini, Random Forest memungkinkan perusahaan untuk mengelola portofolio kredit secara lebih aman, mengurangi risiko kerugian, dan meningkatkan efisiensi proses evaluasi kredit dibandingkan model lainnya.

Metric	Logistic Regression	LightGBM	Random Forest
Accuracy	0.703	0.7299	0.9502
Precision	0.1704	0.1959	0.6347
Recall	0.6925	0.7559	0.9027
F1 Score	0.2735	0.3112	0.7453
ROC-AUC	0.7651	0.8236	0.983

Contact

 www.linkedin.com/in/vaniazerliinaa

 vaniazerlinautamii@gmail.com

 <https://github.com/vaniazerlina/HCI-Credit-Risk-Project>



THANK YOU