

HOME CREDIT SCORECARD MODEL

oleh MY Kaharudin

[Github](#)



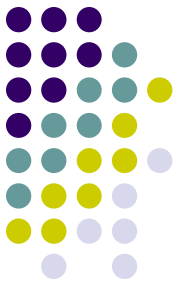
Latar Belakang

Banyak individu menghadapi kesulitan dalam mendapatkan pinjaman karena kurangnya riwayat kredit atau tidak memiliki riwayat kredit sama sekali.

Home Credit berusaha untuk meningkatkan akses keuangan bagi mereka yang tidak terlayani oleh sistem perbankan dengan memberikan pengalaman peminjaman yang positif dan aman.

Tujuannya adalah memastikan bahwa individu yang kurang terlayani ini mendapatkan pengalaman peminjaman yang baik. Home Credit menggunakan berbagai data alternatif untuk memprediksi kemampuan pelanggan dalam melunasi pinjaman.

Dengan melakukan hal ini, Home Credit memastikan bahwa individu yang mampu melunasi tidak ditolak dan pinjaman disalurkan dengan prinsip, syarat, dan jadwal pembayaran yang mendukung kesuksesan pelanggan.



DataSet

HOME CREDIT SCORE DATASET



Menangani/Tindakan



Memprediksi kemampuan pelanggan untuk melunasi pinjaman dengan menggunakan variabel TARGET.

Fokus utama evaluasi adalah pada F1-score untuk mengurangi kesalahan prediksi (positif palsu dan negatif palsu).

Prediksi didasarkan pada kemampuan pelanggan untuk melunasi (0: Mampu, 1: Tidak mampu).

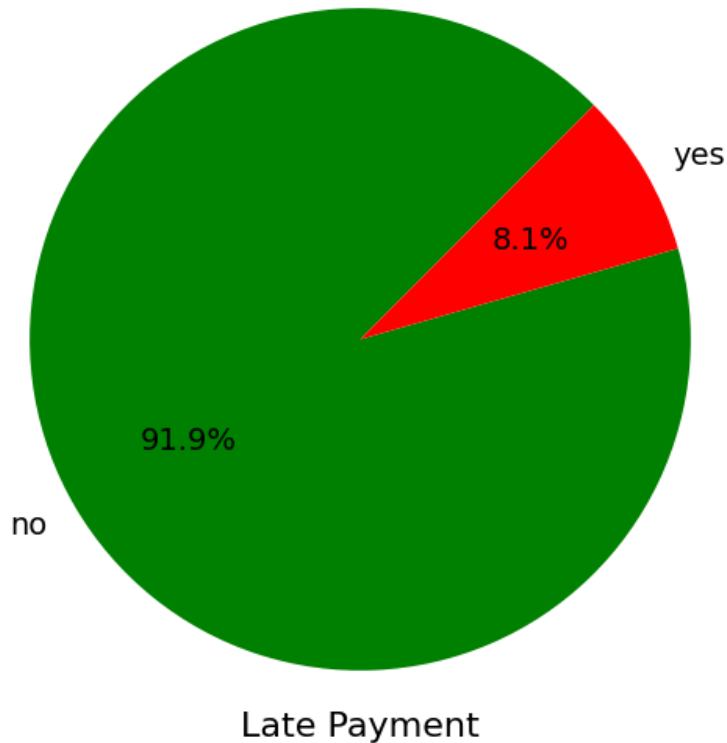
Pengujian Model

1. KNN (K-Nearest Neighbors)
2. Desicion Tree
3. Naive Bayes
 - Naive Bayes without oversampling
 - Naive Bayes oversampling

Evaluasi Tingkat Keterlambatan Pembayaran Pelanggan



Percentage of Clients by TARGET Value



Dari keseluruhan klien, sebanyak 91,9% telah memenuhi kewajiban pembayaran, sedangkan 8,1% mengalami pembayaran terlambat. Hal ini menandakan bahwa hanya sekitar 8% klien yang mengalami keterlambatan pembayaran. Ini mengindikasikan bahwa bisnis Anda memiliki tingkat keterlambatan pembayaran yang rendah. Meski demikian, perlu untuk secara rutin memantau tingkat keterlambatan pembayaran untuk memastikan tetap rendah.



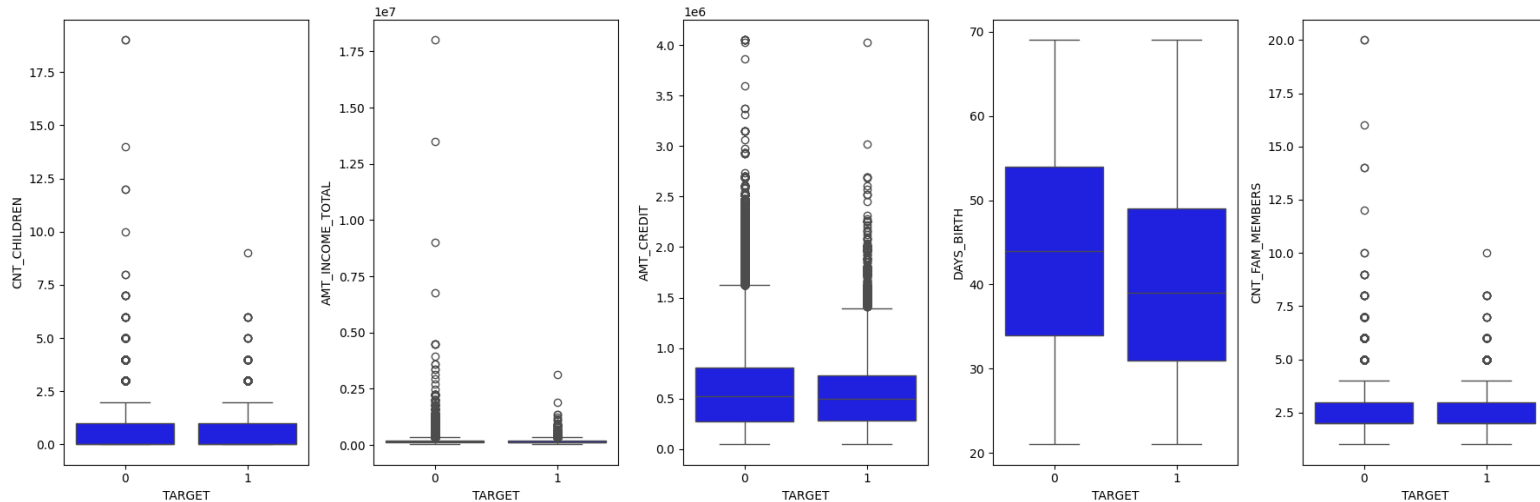
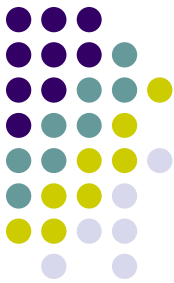
Keterlambatan Pembayaran Pelanggan

Sebagian besar pelanggan yang membayar keterlambatan melakukannya dalam kurun waktu 30 hari setelah jatuh tempo.

Hanya sekitar 1% dari pelanggan yang membayar terlambat lebih dari 90 hari.

	count	unique	top	freq
NAME_CONTRACT_TYPE	307505	2	Cash loans	278232
CODE_GENDER	307505	2	F	202447
FLAG_OWN_CAR	307505	2	N	202920
FLAG_OWN_REALTY	307505	2	Y	213306
WEEKDAY_APPR_PROCESS_START	307505	7	TUESDAY	53900

Pola Jumlah pada Rentang Usia Pelanggan



Grafik ini memberikan gambaran terkait pola jumlah anak pada berbagai rentang usia pelanggan. Terdapat beberapa observasi khusus yang dapat diambil dari data ini:

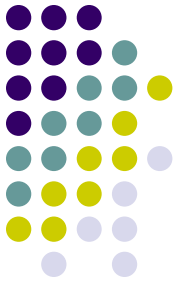
1. Kelompok usia 18-24 tahun memiliki jumlah anak tertinggi. Hal ini mungkin disebabkan oleh faktor-faktor seperti:

- Masih muda dan belum memiliki karier yang mapan.
- Mungkin masih kuliah atau belum bekerja penuh waktu.
- Mungkin belum memiliki tanggung jawab keuangan yang besar.

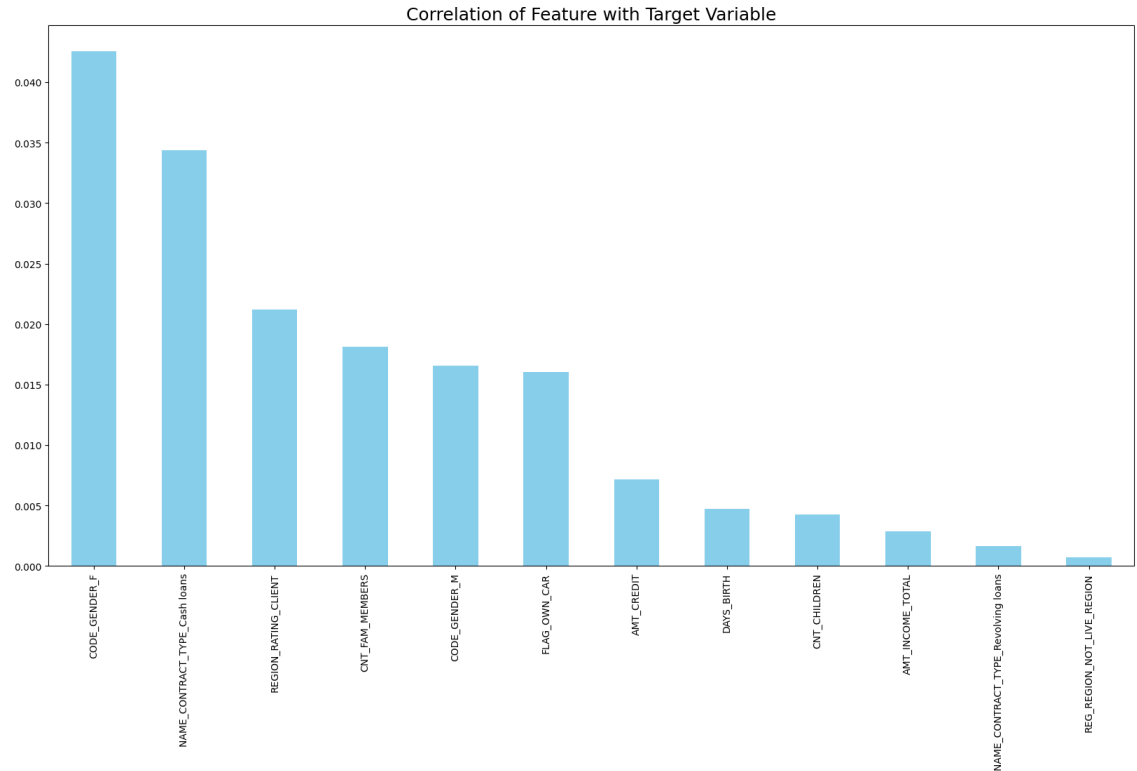
2. Pelanggan berusia 35-44 tahun memiliki jumlah anak paling sedikit. Hal ini bisa disebabkan oleh faktor-faktor seperti:

- Sudah memiliki karier yang mapan.
- Mungkin sudah memiliki tanggung jawab keuangan yang besar, seperti rumah, mobil, atau biaya pendidikan anak.
- Mungkin memiliki pandangan yang berbeda tentang keluarga.

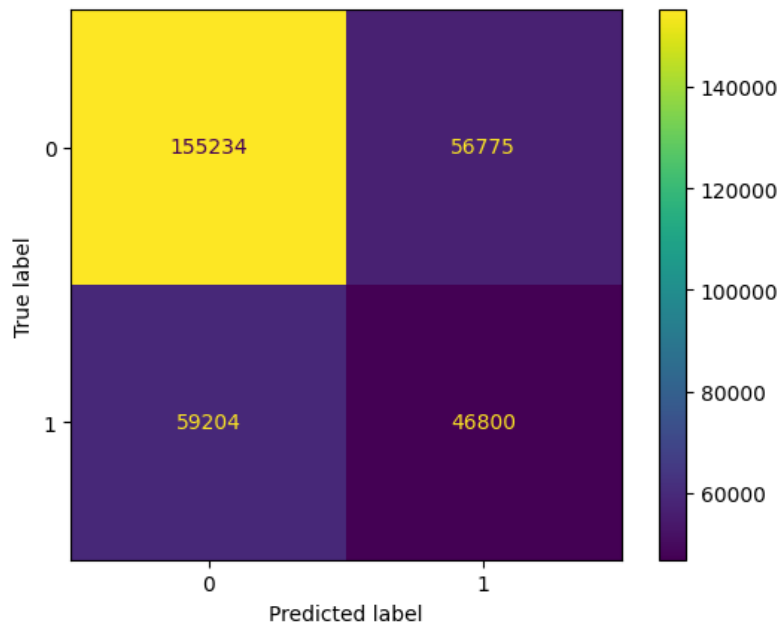
Kemampuan Pembayaran Peminjam



Menunjukkan bahwa semakin tinggi nilai fitur tersebut, semakin kecil kemungkinan peminjam untuk gagal membayar pinjaman.



Analisis Model Klasifikasi untuk Prediksi Keterlambatan Pembayaran



Meskipun Naive Bayes memiliki akurasi yang lebih rendah, performa model secara keseluruhan tetap baik. Berdasarkan analisis tersebut, ada beberapa wawasan yang dapat digunakan oleh perusahaan:

- Fitur menjadi indikator utama untuk memprediksi probabilitas keterlambatan pembayaran. Perusahaan dapat memberikan prioritas kepada pelanggan dengan nilai rendah untuk mendapatkan pinjaman.
- Jumlah data pada kelas yang tidak akan mengalami keterlambatan pembayaran jauh lebih besar. Perusahaan perlu menggunakan teknik oversampling atau undersampling untuk menyeimbangkan jumlah data di kedua kelas.

KESIMPULAN



Dari analisis model klasifikasi yang dilakukan, beberapa wawasan bisnis dapat diambil:

1. Prioritaskan Fitur Utama: Fitur-fitur seperti Days birth, External score, Income total, Age, dan Employment duration memiliki korelasi yang kuat dengan kemungkinan keterlambatan pembayaran. Ini memungkinkan perusahaan untuk fokus pada aspek-aspek ini saat menilai risiko kredit pelanggan.
2. Penggunaan Model: Meskipun Naive Bayes memiliki akurasi lebih rendah dari KNN dan Decision Tree, model ini masih dapat memberikan prediksi yang baik. Hal ini menunjukkan bahwa meskipun tidak memiliki akurasi tertinggi, Naive Bayes masih layak digunakan untuk memprediksi kemungkinan keterlambatan pembayaran.
3. Penyeimbangan Data : Terdapat perbedaan jumlah data yang signifikan antara kelas yang tidak akan mengalami keterlambatan pembayaran dengan kelas yang akan mengalami keterlambatan. Oleh karena itu, perusahaan perlu menggunakan teknik oversampling atau undersampling untuk menyeimbangkan jumlah data di kedua kelas agar model dapat mempelajari pola dengan lebih baik.
4. Keakuratan Model : Meskipun model memiliki performa yang baik, tetap perlu diingat bahwa keakuratan model bukanlah satu-satunya faktor penentu. Pemantauan berkala terhadap performa model dan penyesuaian terhadap perubahan data yang baru akan membantu meningkatkan ketepatan prediksi.

Dengan mempertimbangkan wawasan-wawasan ini, perusahaan dapat mengambil langkah-langkah yang lebih tepat dalam menentukan kebijakan kredit, menyesuaikan strategi pemasaran, dan memaksimalkan penggunaan model prediksi untuk meminimalkan risiko keterlambatan pembayaran.