

CREDIT RISK PREDICTION USING MACHINE LEARNING

PROJECT -BASED INTERNSHIP HOME CREDIT INDONESIA



Presented By
Nur Sinta Sari



PROBLEM RESEARCH

Industri pembiayaan menghadapi tantangan dalam menyeimbangkan antara risiko gagal bayar dan peluang bisnis. Kesalahan dalam keputusan kredit dapat menyebabkan kerugian finansial akibat nasabah gagal bayar atau kehilangan potensi keuntungan karena menolak nasabah yang sebenarnya layak.

Tujuan dari proyek ini adalah membangun model Machine Learning untuk memprediksi risiko gagal bayar nasabah secara lebih akurat dan konsisten, sehingga dapat mendukung pengambilan keputusan kredit yang lebih optimal.





DATA PRE-PROCESSING

Dataset terdiri dari beberapa sumber data dengan granularitas yang berbeda, di mana satu nasabah dapat memiliki banyak catatan historis. Oleh karena itu, dilakukan feature engineering dengan mengagregasi data historis seperti total hutang, rata-rata keterlambatan pembayaran, dan riwayat kredit sebelumnya ke tingkat nasabah.

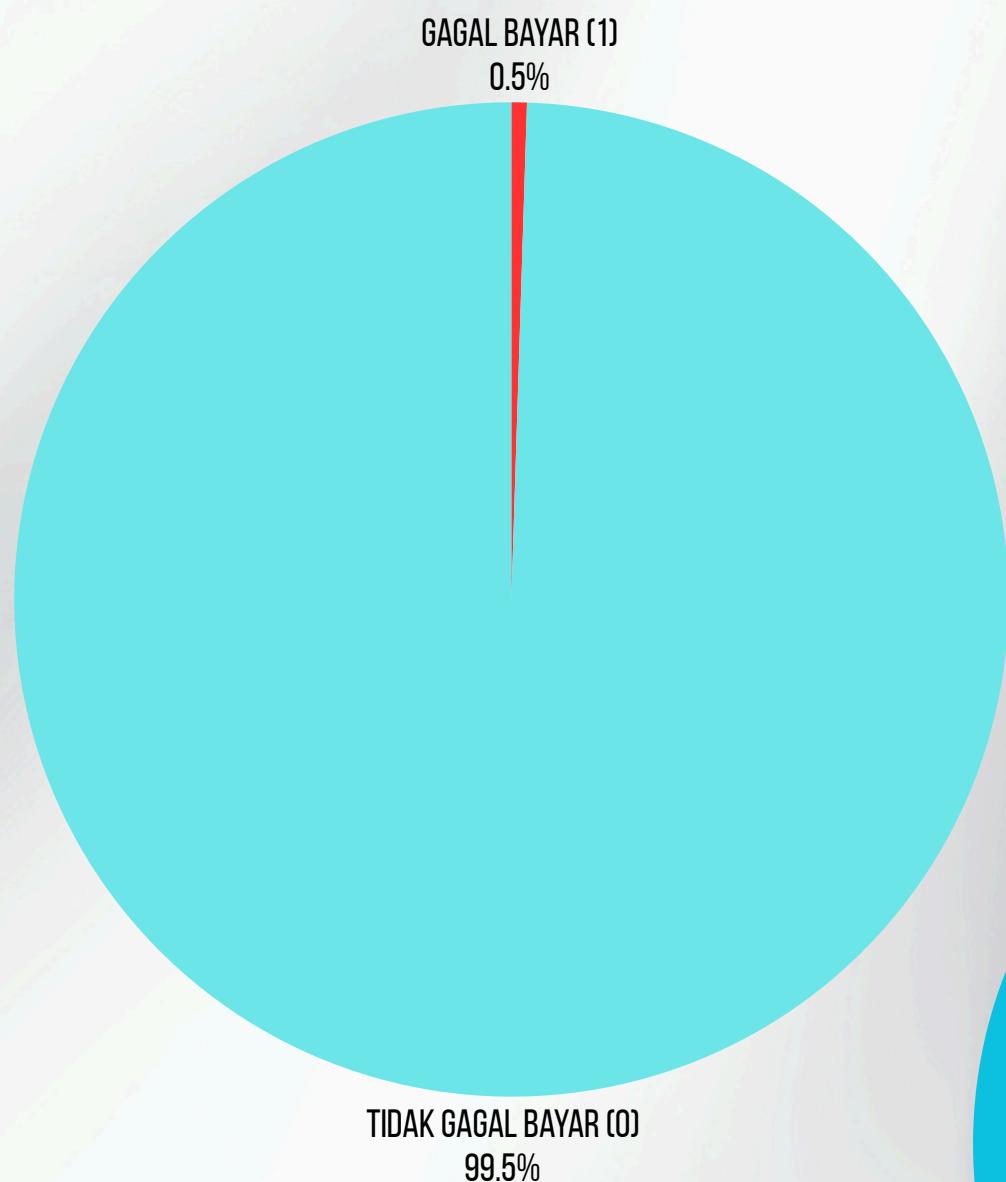
Missing value ditangani menggunakan pendekatan median untuk variabel numerik karena lebih tahan terhadap outlier. Selain itu, hanya fitur numerik yang digunakan agar kompatibel dengan model machine learning. Data kemudian dibagi menjadi data training dan validation untuk proses evaluasi model.

DATA VISUALIZATION AND BUSINESS INSIGHT

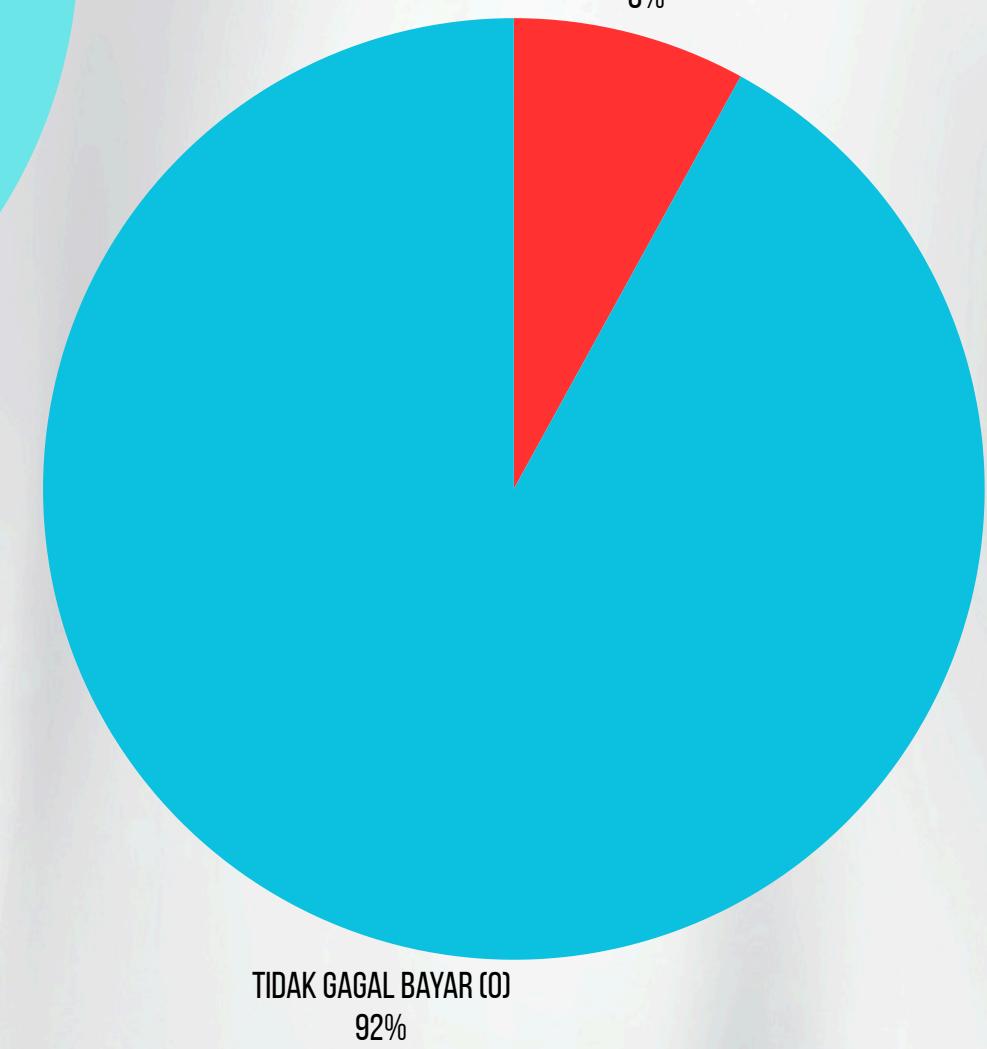
Visualisasi data menunjukkan bahwa distribusi TARGET bersifat tidak seimbang, di mana mayoritas nasabah tidak mengalami gagal bayar. Kondisi ini umum terjadi pada data kredit dan menjadi alasan penggunaan metrik evaluasi ROC AUC.

Selain itu, visualisasi hasil prediksi XGBoost memperlihatkan proporsi nasabah yang diklasifikasikan sebagai berisiko dan tidak berisiko. Insight ini penting bagi bisnis untuk memahami seberapa besar potensi risiko yang dihadapi dan menentukan strategi mitigasi risiko kredit.

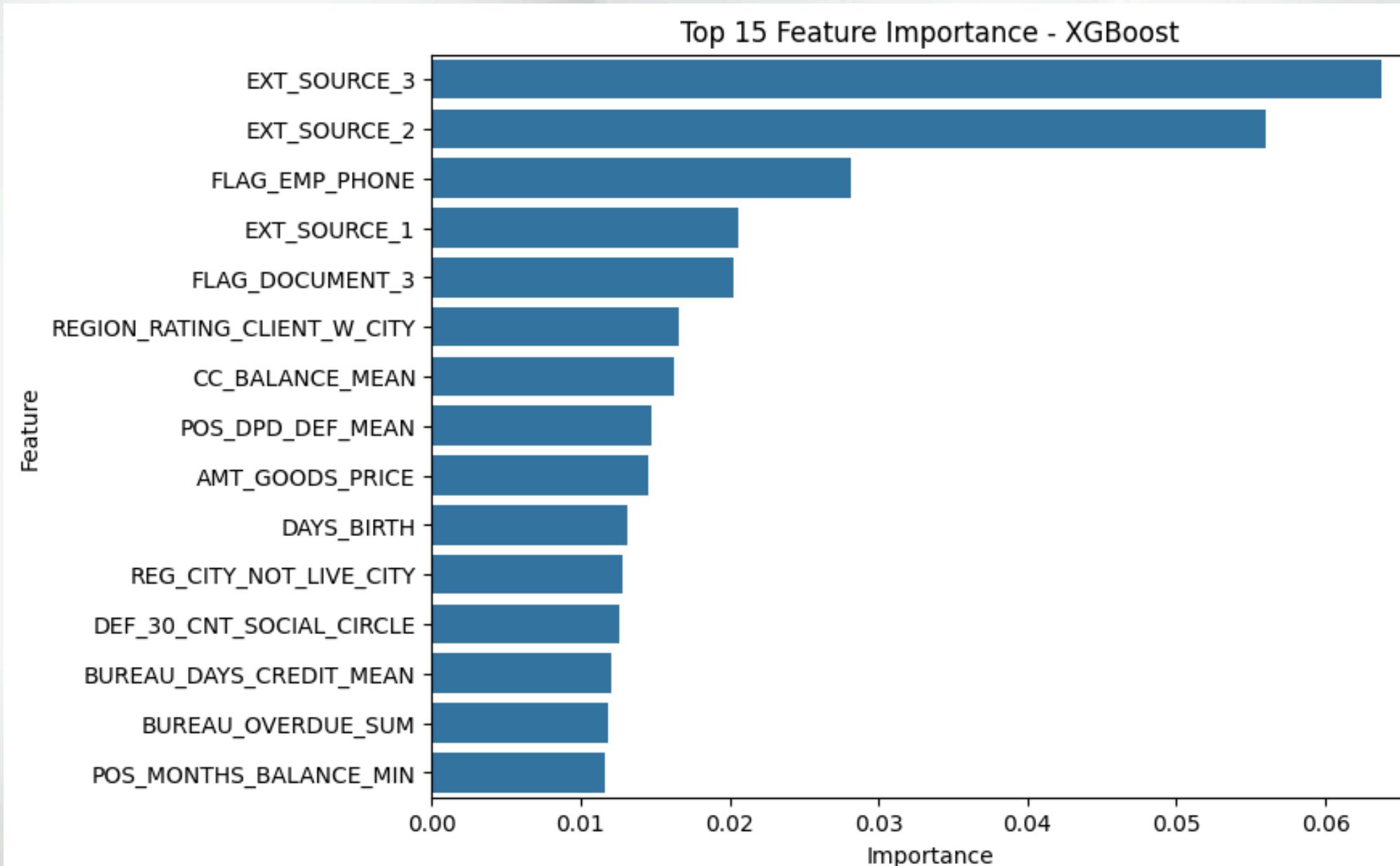
DISTRIBUSI PREDIKSI NASABAH - XGBOOST



DISTRIBUSI TARGET AKTUAL



MACHINE LEARNING INSIGHT



Model XGBoost menunjukkan kemampuan yang baik dalam mengidentifikasi pola kompleks dari perilaku kredit nasabah. Analisis feature importance menunjukkan bahwa faktor yang paling berpengaruh terhadap risiko gagal bayar adalah riwayat hutang, keterlambatan pembayaran cicilan, serta histori kredit sebelumnya.

Hal ini mengindikasikan bahwa perilaku pembayaran historis merupakan indikator yang lebih kuat dibandingkan hanya informasi demografis. Insight ini sejalan dengan praktik industri yang menempatkan histori kredit sebagai faktor utama dalam penilaian risiko.

MACHINE LEARNING IMPLEMENTATION AND EVALUATION

Beberapa model machine learning diuji, yaitu Logistic Regression sebagai baseline, Random Forest, dan XGBoost sebagai model utama. Evaluasi dilakukan menggunakan ROC AUC karena metrik ini cocok untuk data yang tidak seimbang.

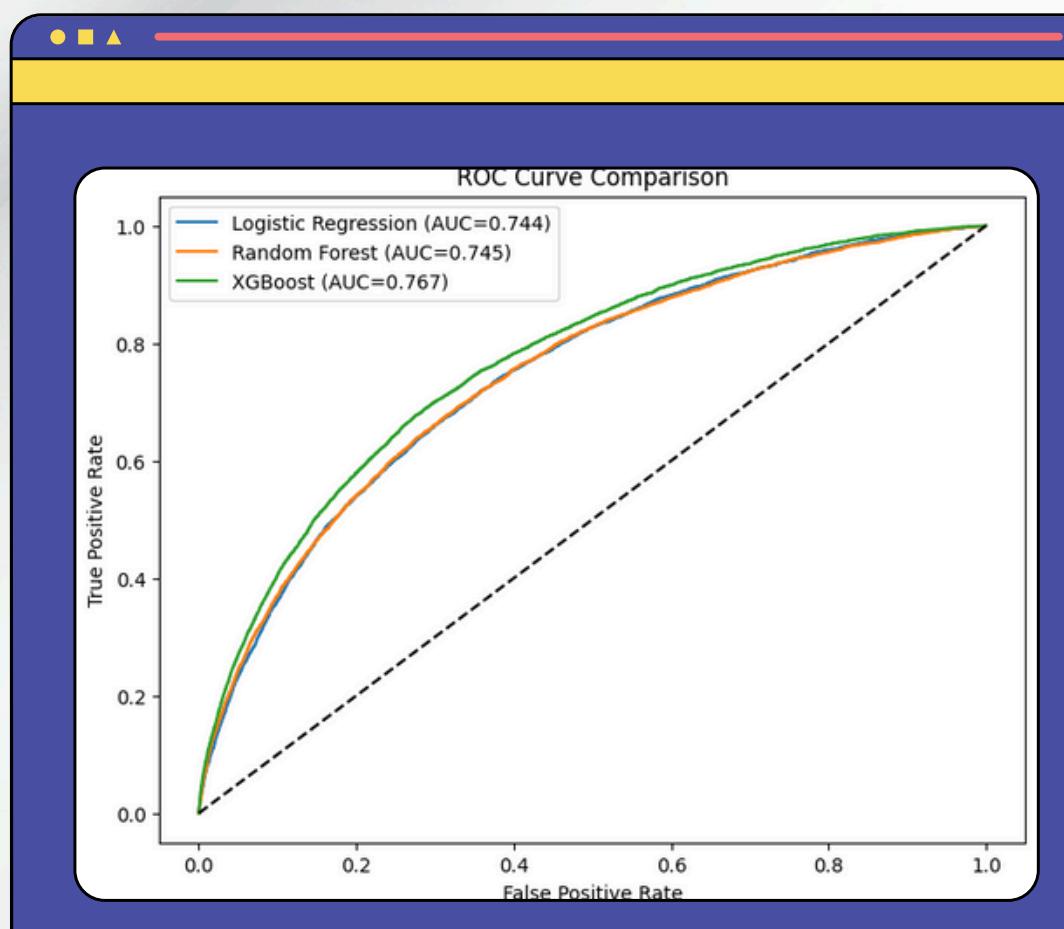
Hasil evaluasi menunjukkan bahwa XGBoost memberikan performa terbaik dalam membedakan nasabah berisiko dan tidak berisiko. Confusion matrix juga menunjukkan bahwa model mampu menangkap sebagian besar nasabah gagal bayar, yang sangat penting dalam konteks manajemen risiko kredit.



TABEL PERBANDINGAN ROC AUC

Model	ROC AUC
0 Logistic Regression	0.744302
1 Random Forest	0.745079
2 XGBoost	0.767264

GRAFIK ROC CURVE



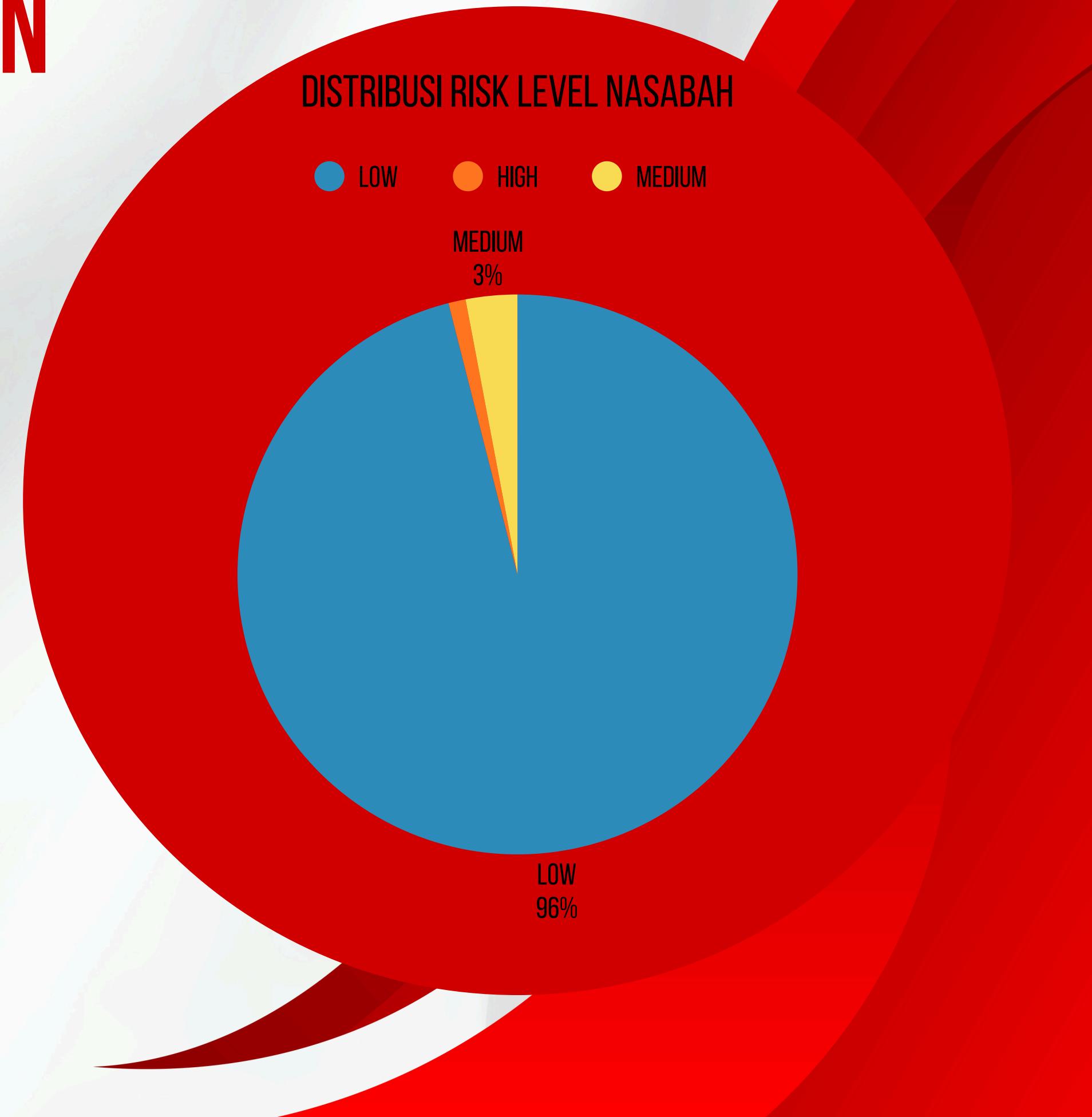
CONFUSION MATRIX XGBOOST

		Predicted	
		0	1
Actual	0	56445	109
	1	4781	168

BUSINESS RECOMMENDATION

Berdasarkan hasil model, nasabah dapat dikelompokkan ke dalam tiga tingkat risiko, yaitu Low Risk, Medium Risk, dan High Risk. Nasabah dengan risiko rendah dapat langsung disetujui secara otomatis, sementara nasabah dengan risiko menengah dapat melalui proses review manual. Nasabah dengan risiko tinggi sebaiknya ditolak atau diberikan syarat tambahan seperti tenor lebih pendek atau bunga lebih tinggi.

Implementasi model ini berpotensi meningkatkan kualitas keputusan kredit, mengurangi risiko gagal bayar, serta meningkatkan efisiensi proses persetujuan pinjaman melalui otomasi berbasis data.



THANK YOU

Thank you to Home Credit Indonesia and Rakamin Academy for the opportunity and learning experience. This program provided valuable insights into the application of machine learning for credit risk analysis and data-driven business decision-making.



<https://github.com/NurSinta/Rakamin-X-Home-Credit/tree/main>



www.linkedin.com/in/nur-sinta-sari-5a8567306