



id/x partners

Credit Risk Prediction Model

Final Project – VIX ID/X Partners



Rika Rostika Afipah – Data Scientist Intern



Business Problem

- Perusahaan lending menghadapi risiko kredit akibat gagal bayar
- Keputusan approval pinjaman yang kurang akurat dapat menyebabkan kerugian
- Diperlukan sistem prediksi risiko kredit berbasis data





Project Objective

- Membangun model machine learning untuk memprediksi credit risk
- Mengklasifikasikan peminjam menjadi:
 1. Good Loan
 2. Bad Loan
- Membantu proses pengambilan keputusan pinjaman secara objektif





Dataset Overview

- Sumber: Lending Club
- Periode data: 2007–2014
- Total data: Loan approved & rejected
- Target variable: loan_status

Target mapping:



Fully Paid → Good Loan



Charged Off → Bad Loan





Data Preparation

- Filtering target (Fully Paid & Charged Off)
- Handling missing values
- Feature encoding & scaling
- Train-test split (80% : 20%)

Fitur utama:

- Loan amount
- Interest rate
- Annual income
- Debt-to-income ratio
- Employment length



Modeling Approach

- Logistic Regression (baseline model)
- Random Forest Classifier (final model)
- Model dipilih karena:
 - Performa baik
 - Mampu menangkap pola non-linear





Model Evaluation

1. Significant Improvement (Threshold 0.2)

- Recall (Bad Loan): 56% (Naik drastis dari 15%).
- ROC-AUC Score: 0.66 (Kemampuan membedakan kelas cukup baik).

2. Business Impact Analysis

- Trade-off: False Positive meningkat (lebih banyak nasabah aman yang perlu dicek manual).
- Justification: Lebih baik "berhati-hati" (mitigasi risiko di awal) daripada "kebobolan" (kehilangan uang pokok pinjaman).

Confusion Matrix (Threshold 0.2)

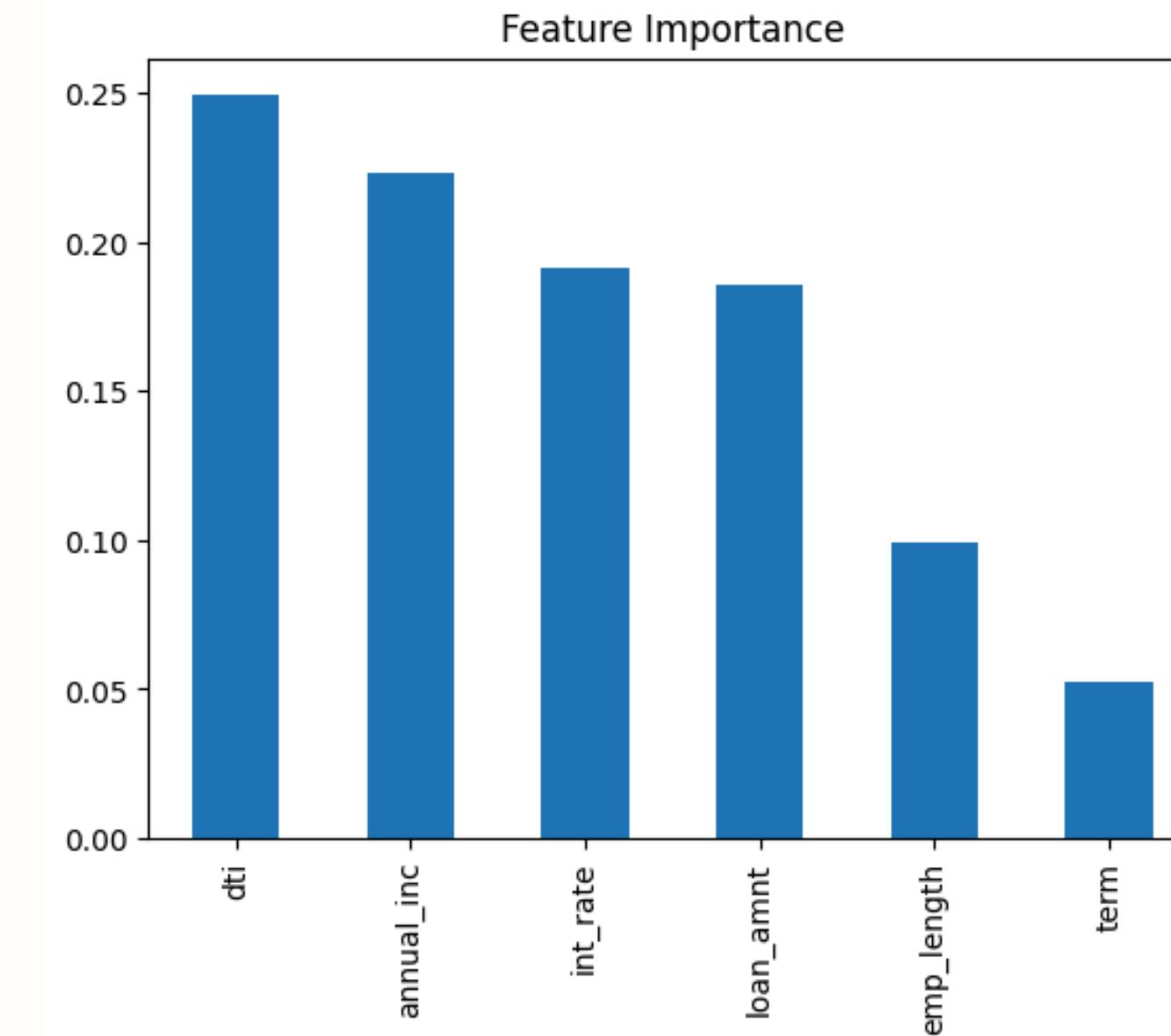
		Predicted Label	
		0	1
True Label	0	18484	8911
	1	2666	3325
0	18484	8911	1



Feature Importance

Faktor paling berpengaruh terhadap risiko kredit:

- Debt-to-Income (DTI)
- Annual income
- Interest rate
- Loan amount





Business Insight & Recommendation

Insight:

- Interest Rate yang tinggi dan DTI (Debt-to-Income) yang besar adalah indikator terkuat seseorang akan gagal bayar.
- Nasabah dengan kepemilikan rumah "Rent" (Sewa) cenderung memiliki risiko lebih tinggi.

Recommendation (Strategic):

1. Terapkan Threshold Agresif (0.2): Jangan gunakan threshold standar (0.5). Gunakan batas 0.2 (20%). Artinya, jika model memprediksi probabilitas gagal bayar seseorang $> 20\%$, sistem harus menandainya sebagai High Risk.
2. Sistem "Two-Layer Approval":
 - Low Risk ($\text{Prob} < 0.2$): Auto-Approval (Proses cepat).
 - High Risk ($\text{Prob} > 0.2$): Jangan langsung ditolak, tapi alihkan ke Manual Review oleh tim analis kredit untuk memverifikasi dokumen lebih lanjut. Ini menanggulangi jumlah False Positive yang tinggi.





Conclusion



- Reliable Detection: Model Random Forest dengan Weighted Class dan Threshold 0.2 terbukti efektif meningkatkan keamanan portofolio pinjaman dengan Recall 56%.
- Business Impact: Solusi ini menyeimbangkan antara keamanan dana perusahaan dan peluang bisnis, dengan memprioritaskan pencegahan kerugian finansial akibat gagal bayar (Charged Off).
- Ready for Deployment: Model telah disimpan (.pkl) dan siap diintegrasikan ke dalam sistem backend perusahaan untuk scoring otomatis.