



Credit Risk Prediction

Dimas Jabbar Rosul



Agenda

01 Problem Research & Business Understanding

02 Business Insights

03 Data Pre-processing

04 Machine Learning Model

05 Business Recommendation



Problem Research

- Data yang diperoleh berisi informasi applicants pinjaman dan apakah mereka dilabeli risiko yang baik atau risiko yang tidak baik. Ketika seorang applicant mengajukan pinjaman, terdapat dua tipe risiko, yaitu:
 1. Good risk, artinya dibayar dengan full, current, dan In grace period. Applicants dengan label ini cenderung di approved di kemudian hari.
 2. Bad Risk, artinya terlambat membayar pinjaman, default, dan charged off. Applicants dengan label ini, pinjamannya cenderung tidak di approve di kemudian hari.





Business Understanding



Problem Statement

Meminjamkan pinjaman pada applicants dengan risiko yang buruk dapat menjadi sumber kerugian finansial terbesar. Credit loss adalah jumlah uang yang hilang dari peminjam ketika applicant menolak untuk membayar.

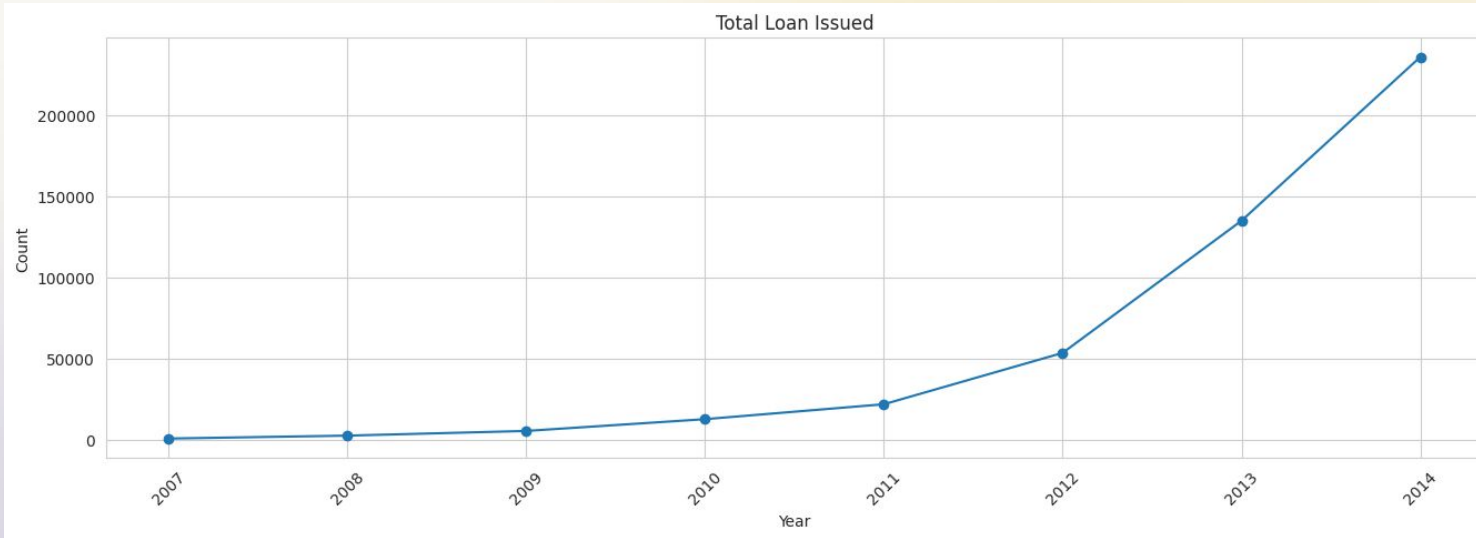


Objective

- Mengidentifikasi pola yang mengindikasikan apakah seseorang tidak membayar kembali pinjamannya atau di label sebagai bad risk.
- Mengimplementasikan machine learning untuk membangun model prediksi untuk memprediksi risiko pinjaman.



Business Insight



- Total pinjaman yang bermasalah bertambah tiap tahun, paling tinggi di tahun 2014, dengan 235.628 pinjaman yang bermasalah.



Business Insight



Loan status	Total Loss	% of total loss	Total Applicants	Average loss
Charged Off	\$574,356,330	83,49%	43,236	\$13,284
Late	\$102,293,296	14,87%	8,118	\$12,600
Default	\$11,299,446	1,64%	832	\$13,581

- Pinjaman dengan status “Charged off” adalah sumber loss terbesar (83%) untuk perusahaan dengan total loss 574 juta dari 43,236 applicant.
- Pinjaman dengan status “Late” berkontribusi 14% terhadap kerugian perusahaan dengan total loss adalah 102 juta dari 8118 applicant.
- Loan dengan status “Default” hanya berkontribusi 1.64% pada kerugian perusahaan dengan total kerugian 11 juta dari 832 applicant. Namun status pinjaman ini mempunyai rata-rata loss terbesar yaitu \$13,581 per applicant.



Data Pre-processing



Dataset :

Data yang digunakan berasal dari data id/x. Dengan 75 fitur dan 466.285 row data.

Data Cleansing

- * Cek duplikasi data
- * Handling missing values
- * Detecting Outliers

Model Building

- * Label Encoding
Mengganti data non-numeric → numeric
- * Feature selection
- * Handle imbalance label
Lakukan SMOTE sehingga labelnya seimbang.
- * Lakukan Normalisasi dengan MinMaxScaler.

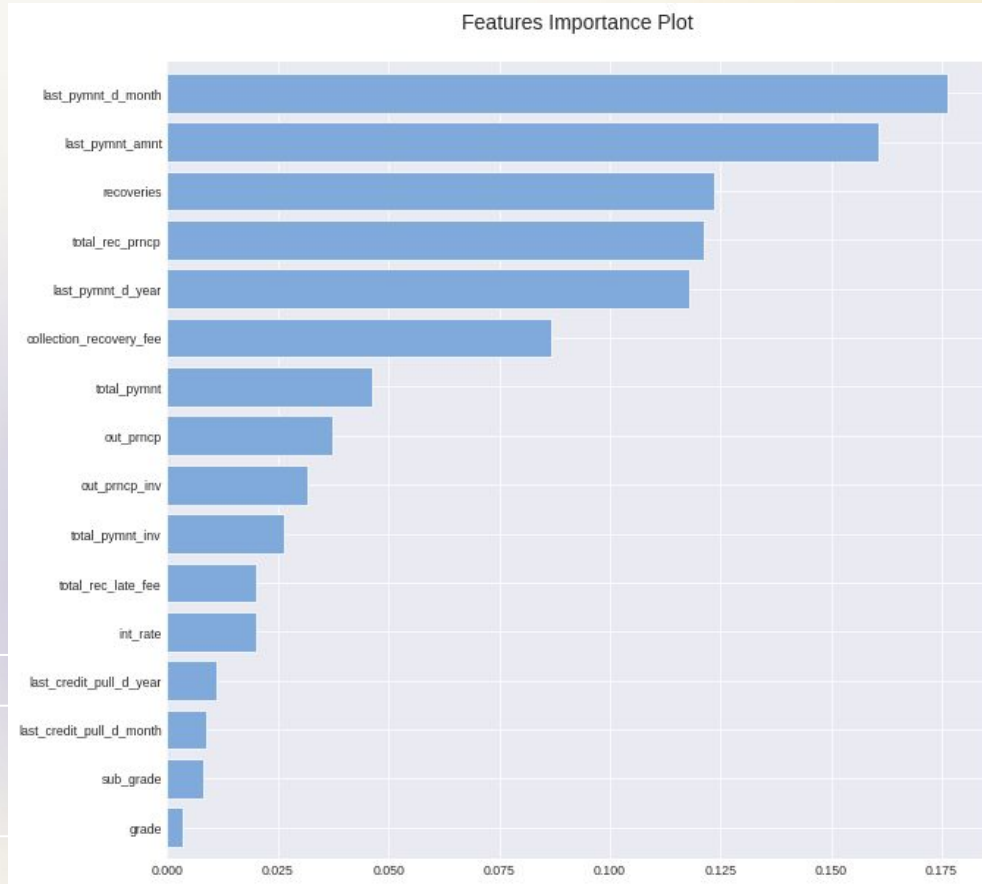




Machine Learning Model

Algoritma	Training Accuracy score	Testing Accuracy score	Error Margin
Decision Tree	99.98%	98.32%	1.66%
Random Forest	99.98%	99.11%	0.87%
Logistic Regression	88.38%	88.41%	0.03%
KNN	97.61%	96.7%	0.91%
Gaussian Naive Bayes	75.08%	74.98%	0.1%
Neural Network	98.99%	98.72%	0.27%
XG Boost	99.39%	99.08%	0.31%
Gradient Boosting	97.56%	97.52%	0.04%

Features Important



- Algoritma terbaik, yang digunakan untuk memprediksi adalah Random Forest.
- Top 5 fitur yang berpengaruh pada prediksi model adalah: “Last payment d month”, “Last payment amount”, “recoveries”, “total rec pmcp”, “last payment d year”



Business Recommendations

- Last Payment month, last payment amount, recoveries value, total principal received, dan last payment year adalah fitur yang sangat penting untuk mengidentifikasi apakah applicant mempunyai kemungkinan untuk tidak membayar kembali pinjamannya.

Solution:

- Perusahaan harus lebih memonitor indikator di atas untuk mengurangi kerugian/loss.
- Di kemudian hari, jika ada applicant dengan indikator diatas yang berpotensi risiko buruk, maka perusahaan dapat mengambil tindakan dengan me-reject/menolak pinjaman, mengurangi jumlah pinjaman, atau meminjamkan dengan rate bunga yang tinggi untuk menghindari total loss/kerugian bagi perusahaan.





Thanks!

Github:

<https://github.com/Dimasuxx/ID-X-Credit-Risk-Prediction>

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon** and infographics & images by **Freepik**

Please keep this slide for attribution

