Rakamin Academy X id/x partners

Credit Risk Prediction

ID/X Partners - Data Scientist

Presented by
Arkan Syafiq At'taqy

Surabaya, Indonesia

arkanattaqy09@gmail.com

https://www.linkedin.com/in/arkan-attaqy

Arkan Syafiq At’taqy

Mahasiswa S1 Teknologi Sains Data Universitas Airlangga

Saya merupakan mahasiswa S1 semester 4 Program Studi Teknologi Sains Data Universitas Airlangga yang memiliki ketertarikan tinggi terhadap *Data Science* dan *Data Analyst*. Saya memiliki keahlian untuk menggunakan bahasa pemrograman (Python, SQL, dan R) dan beberapa aplikasi yang erat kaitannya dengan data seperti Google Colaboratory, IBM SPSS Statistics, Minitab, Tableau, Microsoft Excel dan Microsoft Word.

Courses and Certification

Certificate of Competency - Kimia Farma Big Data Analytics Project Based Internship Program |

link certificate Maret, 2024

Certificate of Completion - Statistics for Data Science and Business Analysis by 365 Careers |

link certificate Januari, 2024

Certificate of Completion - Statistics for Data Science and Business Analysis by 365 Careers |

link certificate Januari, 2024

About Company

ID/X Partners didirikan pada tahun 2002 oleh mantan bankir dan konsultan manajemen yang memiliki pengalaman luas dalam manajemen siklus dan proses kredit, pengembangan penilaian, dan manajemen kinerja. ID/X Partners telah melayani perusahaan-perusahaan di kawasan Asia dan Australia dan di berbagai industri, khususnya jasa keuangan, telekomunikasi, manufaktur, dan ritel.

ID/X Partners menyediakan layanan konsultasi yang berspesialisasi dalam pemanfaatan solusi data analitik dan pengambilan keputusan (DAD) yang dikombinasikan dengan manajemen risiko dan disiplin pemasaran terintegrasi untuk membantu klien mengoptimalkan profitabilitas portofolio dan proses bisnis.

Layanan konsultasi komprehensif dan solusi teknologi yang ditawarkan ID/X Partners menjadikannya sebagai penyedia layanan terpadu.

Project Portfolio

Sebagai Data Scientist di ID/X Partners, disini saya mengembangkan model machine learning yang dapat memprediksi risiko kredit (credit risk) berdasarkan dataset yang disediakan, yang mencakup data pinjaman yang disetujui dan ditolak. Dalam pengembangan modelnya Anda juga perlu melakukan beberapa tahap dimulai dengan Data Understanding, Exploratory Data Analysis (EDA), Data Preparation, Data Modelling, dan Evaluation.

Link code here!

Project explanation video here!

Workflow Project

Rakamin Academy

Data Understanding

Exploratory Data Analysis

Data Preparation

Data Modelling

Evaluation

Data Understanding

Dataset

Dimensi

Kolom: 75

Baris: 466.285

Tipe

Kategorik: 22

Numerik: 53

Goals

- Meningkatkan keakuratan dalam menilai dan mengelola risiko kredit.
- Mengoptimalkan keputusan bisnis dan mengurangi potensi kerugian.

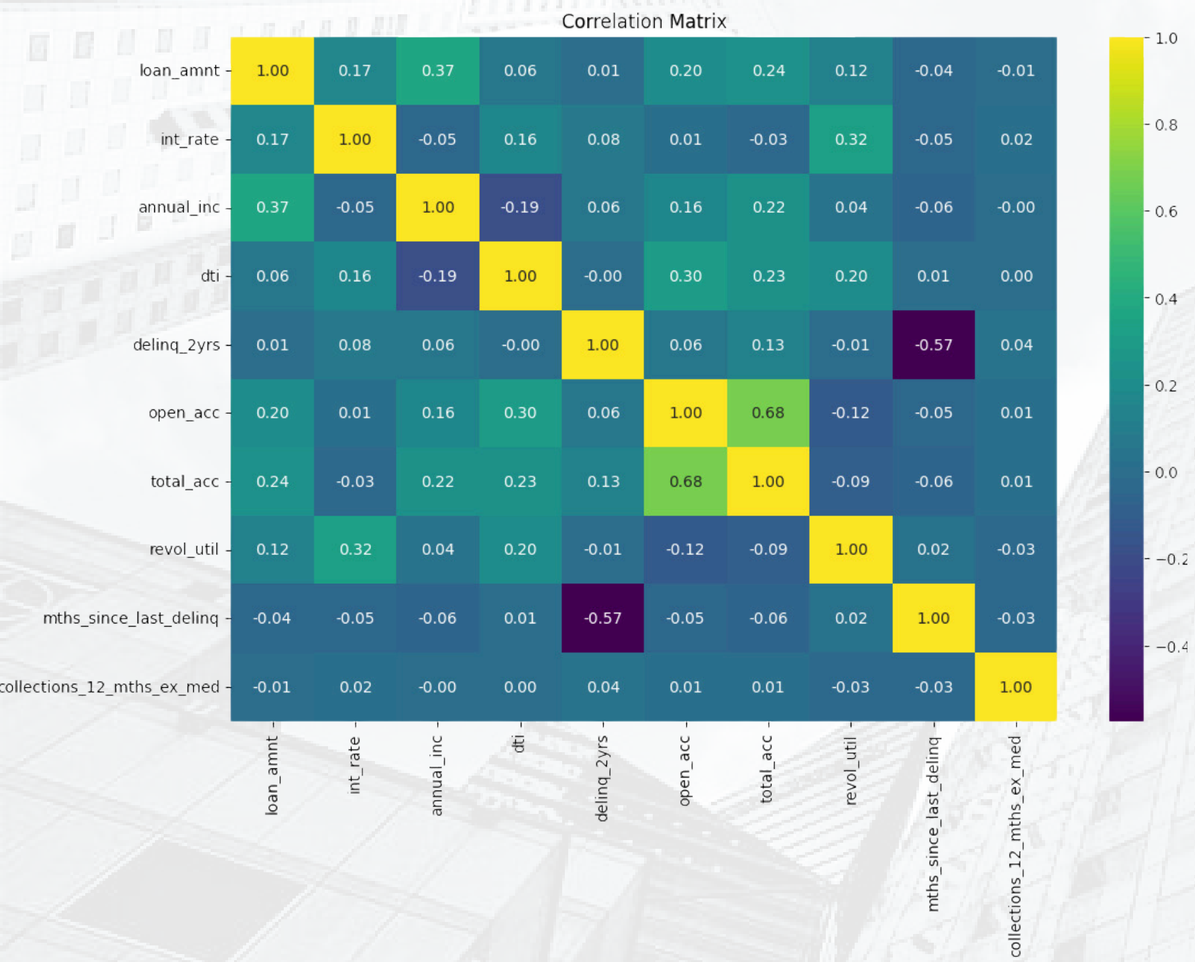
Objective

- Mengembangkan model machine learning yang dapat memprediksi risiko kredit dengan presisi tinggi.
- Mengidentifikasi faktor-faktor yang berkontribusi terhadap risiko kredit.

Business Metrics

- Accuracy
- Precision
- Recall
- F1 Score
- ROC AUC

Exploratory Data Analysis (EDA)



- **Korelasi Positif Kuat:** Variabel seperti total_acc dan open_acc menunjukkan korelasi positif yang kuat, menandakan bahwa semakin banyak akun yang dimiliki seseorang, jumlah total akun kredit juga cenderung lebih banyak.
- **Korelasi Negatif:** Terdapat korelasi negatif antara int_rate dan annual_inc, yang bisa diartikan bahwa pendapatan tahunan yang lebih tinggi cenderung berkaitan dengan suku bunga yang lebih rendah pada pinjaman.
- **Hubungan Saling Eksklusif:** Beberapa variabel seperti delinq_2yrs (jumlah keterlambatan pembayaran dalam 2 tahun) tidak menunjukkan korelasi yang signifikan dengan variabel lain, menunjukkan bahwa faktor ini beroperasi secara independen dari faktor-faktor lain dalam dataset.

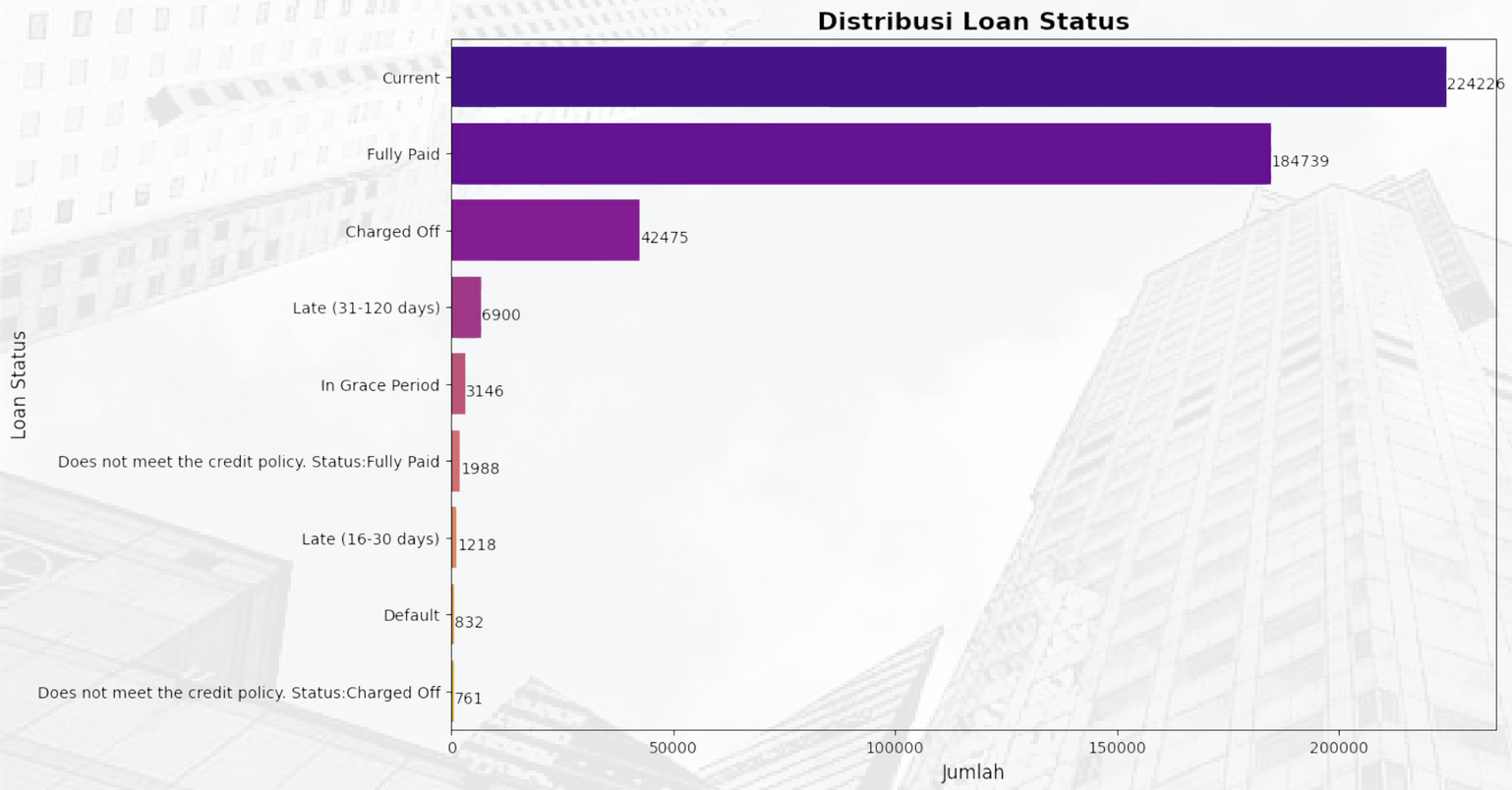
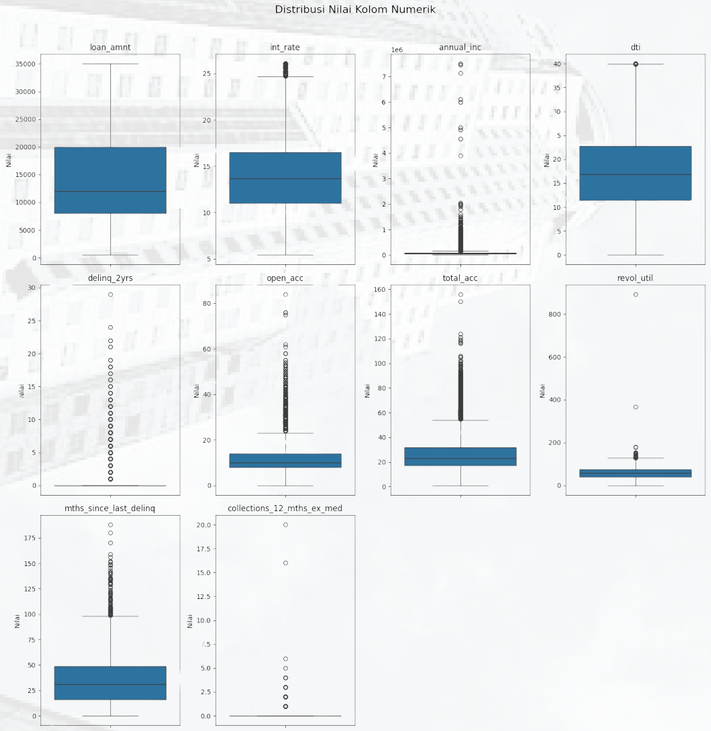
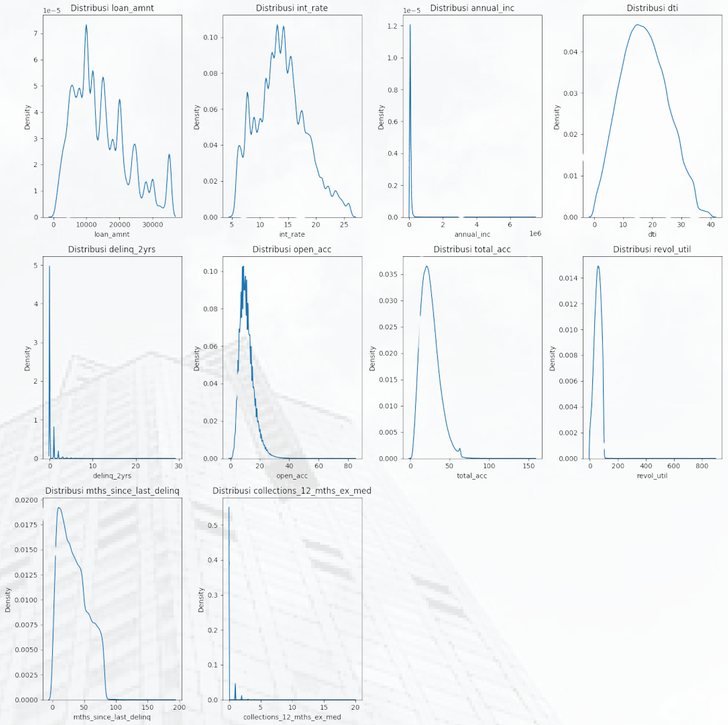


Diagram disamping memberikan gambaran tentang kinerja pinjaman yang digunakan untuk analisis keuangan. Informasi ini juga bisa menjadi indikator kesehatan ekonomi, di mana banyak pinjaman yang berstatus current atau lunas menunjukkan kondisi yang lebih baik dibandingkan dengan banyaknya pinjaman yang bermasalah.





Plot diatas membantu dalam memahami kesehatan keuangan dan perilaku peminjam pada umumnya, seperti jumlah pinjaman umum, suku bunga, dan penggunaan kredit. Distribusi dan outlier menunjukkan tingkat risiko dan stabilitas keuangan yang berbeda-beda di kalangan peminjam

Data Preparation

Handling Column

Menghapus kolom yang memiliki missing values lebih dari 70%

Mengimputasi nilai median terhadap kolom numerik

Mengimputasi nilai modus terhadap kolom kategorik

Data Transformation

- Menambah variabel dengan Feature Enginnering
- Menambah variabel dengan One-Hot Encoding
- Mengurangi variabel dengan Feature Importane
- Train-Test Split sebesar 80% Train & 20% Test terhadap dataset
- Mengaplikasikan SMOTE terhadap data imbalance

Data Modelling

ALGORITHM SCORE

No	Algoritma	Precision	Recall	F1 Score
1	Extra Trees	0.96	0.91	0.94
2	Random Forest	0.98	0.88	0.93
3	LGBM	1.00	0.86	0.92
4	Logistic Regression	0.75	0.66	0.70

Evaluation

SUGGESTION

Berdasarkan hasil analisa, terdapat beberapa saran yang bisa saya berikan, diantaranya:

- **Pemahaman Bisnis yang Mendalam:** Memahami bisnis klien secara mendalam sangat penting. Ini termasuk memahami tujuan bisnis, proses kerja, dan tantangan yang dihadapi oleh klien. Dengan pemahaman yang baik, solusi yang disediakan akan lebih relevan dan efektif.

- **Pendekatan yang Terintegrasi:** Mengintegrasikan data analitik, pengambilan keputusan, manajemen risiko, dan pemasaran untuk memberikan solusi yang holistik kepada klien. Pendekatan ini akan memastikan bahwa semua aspek bisnis klien diperhitungkan dalam pengambilan keputusan.

Thank You

Rakamin Academy X id/x partners