

Final Project

● Project Based Internship - id/x partners

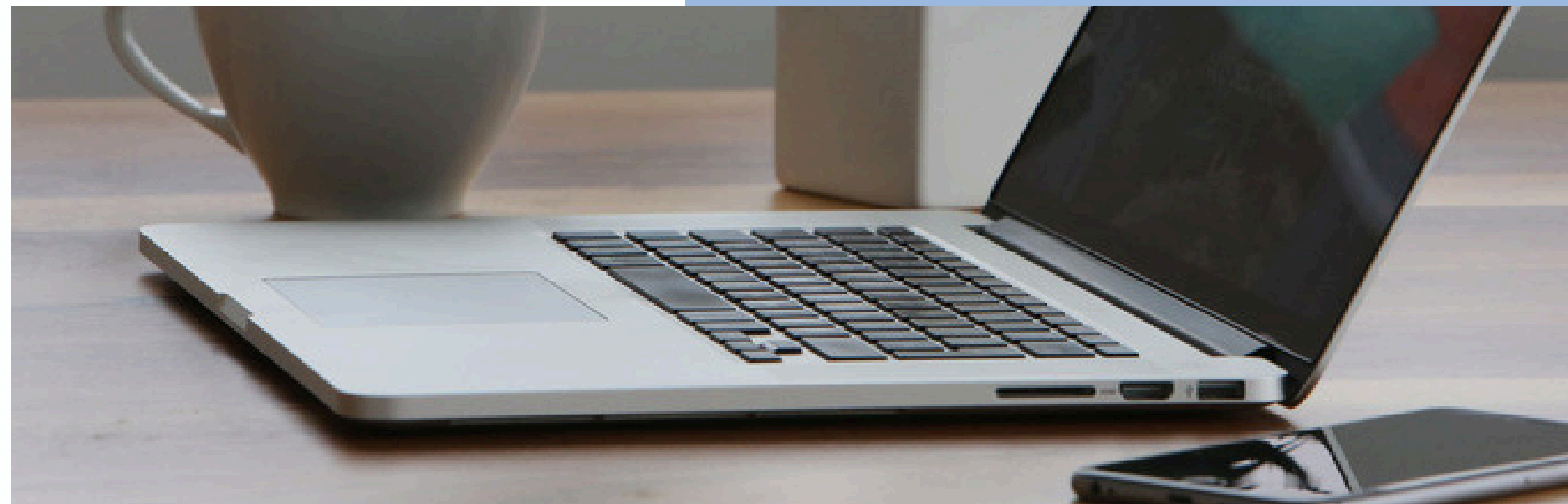
Email
thorIQ004@gmail.com

Github
rajwaAth

LinkedIn
www.linkedin.com/in/muhamad-rajwa-ath

TABLE OF CONTENT

CREDIT RISK ANALYSIS



01

Bussines Understanding

02

Data Understanding

03

Data Exploratory

04

Data Preprocessing

05

Modeling & Evaluation

06

Bussines Recommendation



BUSSINES

UNDERSTANDING

Bank menghadapi tantangan dalam menilai risiko kredit, dengan **12% pinjaman tergolong bad loan**. Tingginya rasio kredit macet ini berdampak pada stabilitas keuangan dan efisiensi operasional bank. Sistem evaluasi kredit saat ini kurang akurat dalam memprediksi risiko nasabah, sehingga diperlukan solusi berbasis machine learning untuk meningkatkan akurasi prediksi kredit.

Next



01

Problem

- **Bad loan mencapai 12%**, menunjukkan risiko kredit yang cukup tinggi.
- **Keputusan kredit kurang akurat**, menyebabkan nasabah berisiko tetap mendapat pinjaman.
- **Proses persetujuan lambat dan tidak efisien**, masih banyak bergantung pada analisis manual.

02

Objective

- **Mengembangkan model machine learning** untuk memprediksi good loan dan bad loan dengan lebih akurat.
- **Menurunkan rasio bad loan dari 12%**, meningkatkan profitabilitas bank.
- **Mempercepat dan mengotomatisasi** proses persetujuan pinjaman.

03

Goal

- **Meningkatkan akurasi prediksi** risiko kredit.
- **Mengurangi bad loan mencapai 5%** agar bank memiliki non performing loan (NPL) yang baik (Sagala & Sari, 2021).
- **Meningkatkan efisiensi operasional dan kepuasan nasabah** dengan proses persetujuan yang lebih cepat.

DATA UNDERSTANDING

DATA LENGTH → 466.285

FEATURE → 74

TARGET → LOAN_STATUS

Berdasarkan hasil deskripsi setiap kolom, **dataset ini terdiri dari 74 kolom dan 466.285 baris data**. Untuk meningkatkan kualitas data dan memastikan hanya fitur yang relevan digunakan dalam analisis lebih lanjut, beberapa langkah akan dilakukan.

Kolom dengan lebih dari 60% nilai null

dihapus karena dianggap kurang informatif, di antaranya:

desc, mths_since_last_record, annual_inc_joint, dti_joint, verification_status_joint, open_acc_6m, open_il_6m, open_il_12m, open_il_24m, mths_since_rcnt_il, total_bal_il, il_util, open_rv_12m, open_rv_24m, max_bal_bc, all_util, inq_fi, total_cu_tl, inq_last_12m, mths_since_last_major_derog, next_pymnt_d, dan mths_since_last_delinq

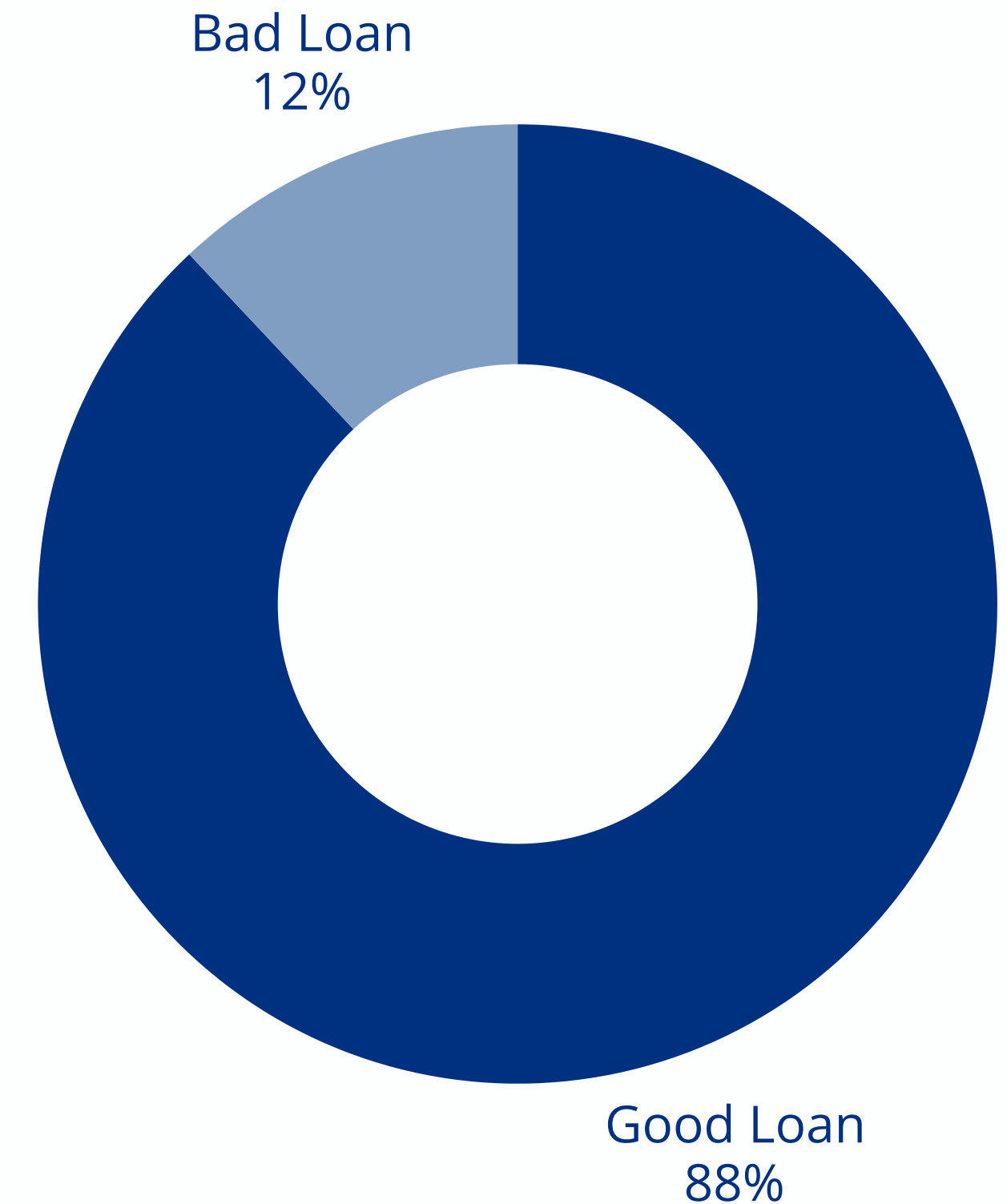
Kolom yang Kurang Relevan

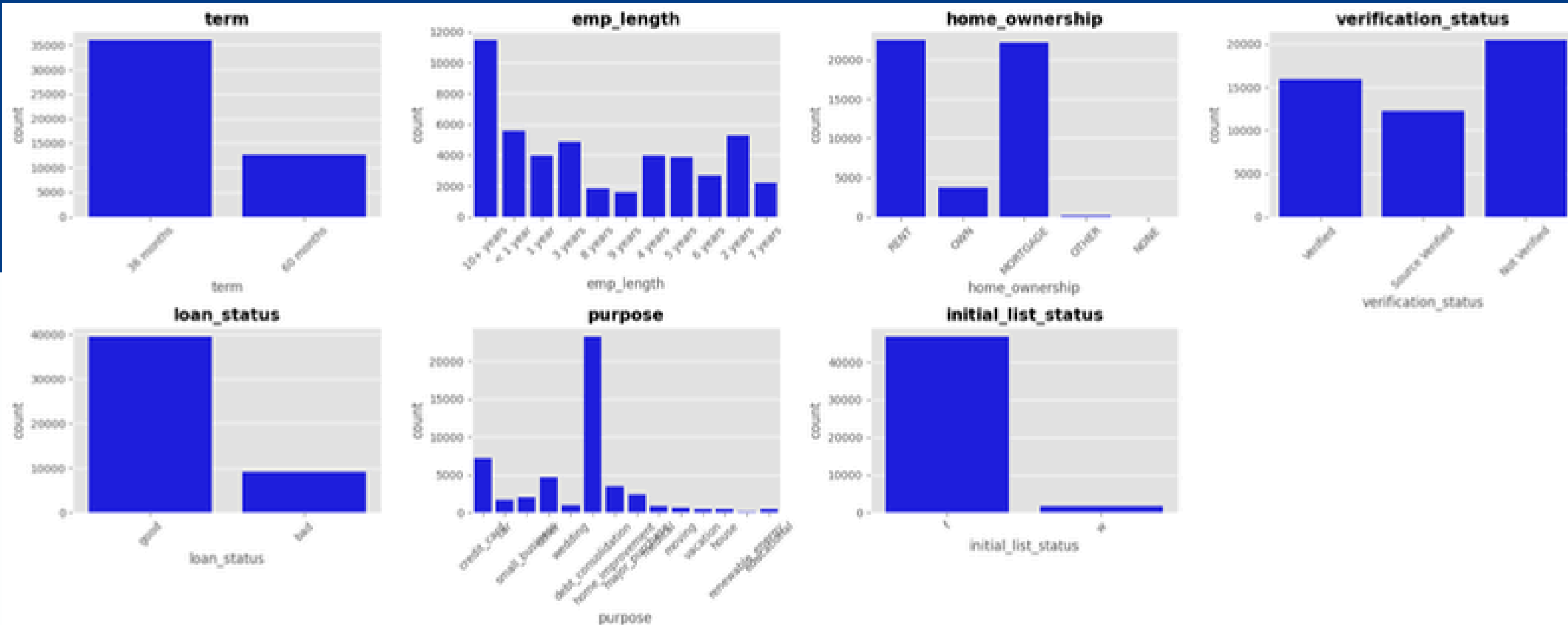
Beberapa kolom yang dianggap kurang relevan juga akan dihapus untuk menyederhanakan dataset, seperti

id, member_id, emp_title, url, desc, policy_code, application_type, zip_code, inq_last_6mths, dan delinq_2yrs

DATA EXPLORATORY

Dataset yang digunakan memiliki distribusi kelas yang tidak seimbang, yang dapat mempengaruhi performa model. Oleh karena itu, diperlukan teknik penyeimbangan seperti oversampling atau undersampling untuk meningkatkan akurasi prediksi.



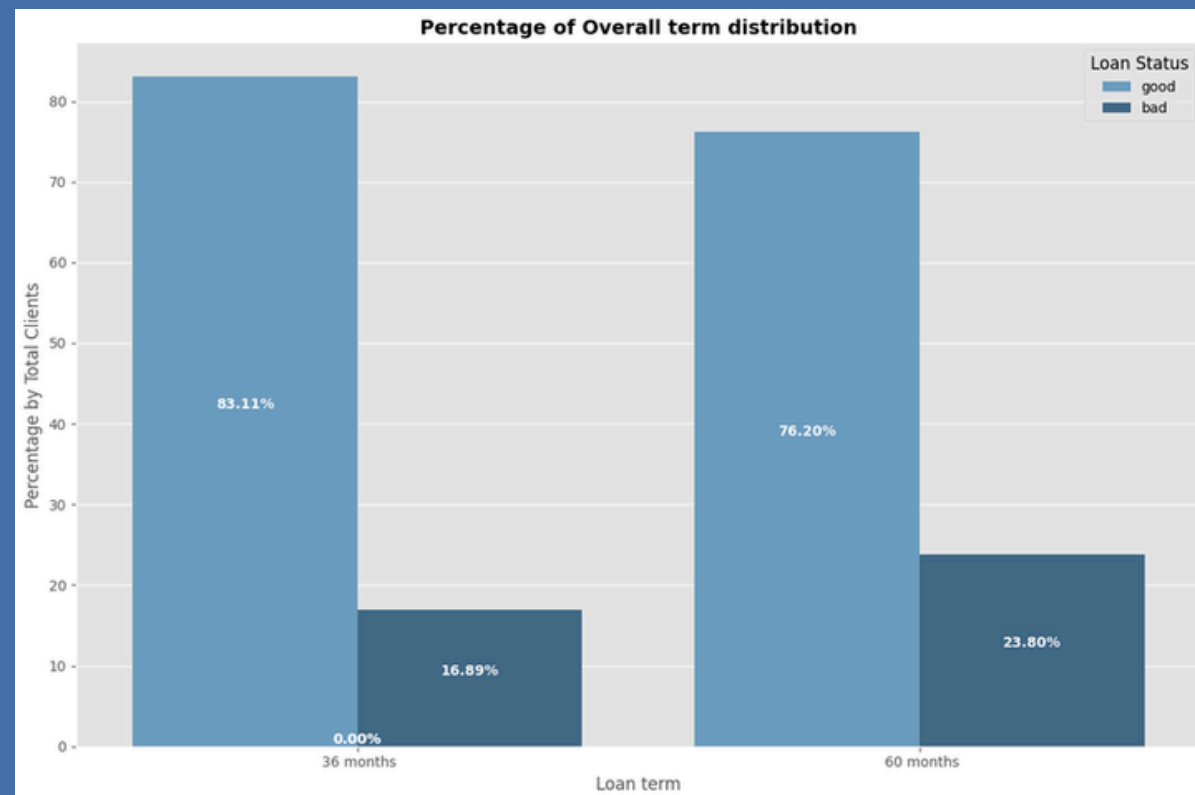


- **Term:** Mayoritas pinjaman berdurasi 36 bulan, sementara 60 bulan lebih sedikit.
- **Employment Length:** >10 tahun pengalaman mendominasi, kategori lain lebih merata.
- **Home Ownership:** Sebagian besar peminjam RENT atau MORTGAGE, sedikit yang OWN.

- **Verification Status:** Banyak peminjam tidak diverifikasi, sebagian terverifikasi.
- **Loan Status:** Ketidakseimbangan data, mayoritas good, sedikit bad.
- **Purpose:** Didominasi debt_consolidation, tujuan lain jauh lebih sedikit.
- **Initial List Status:** Mayoritas f, hanya sedikit w.

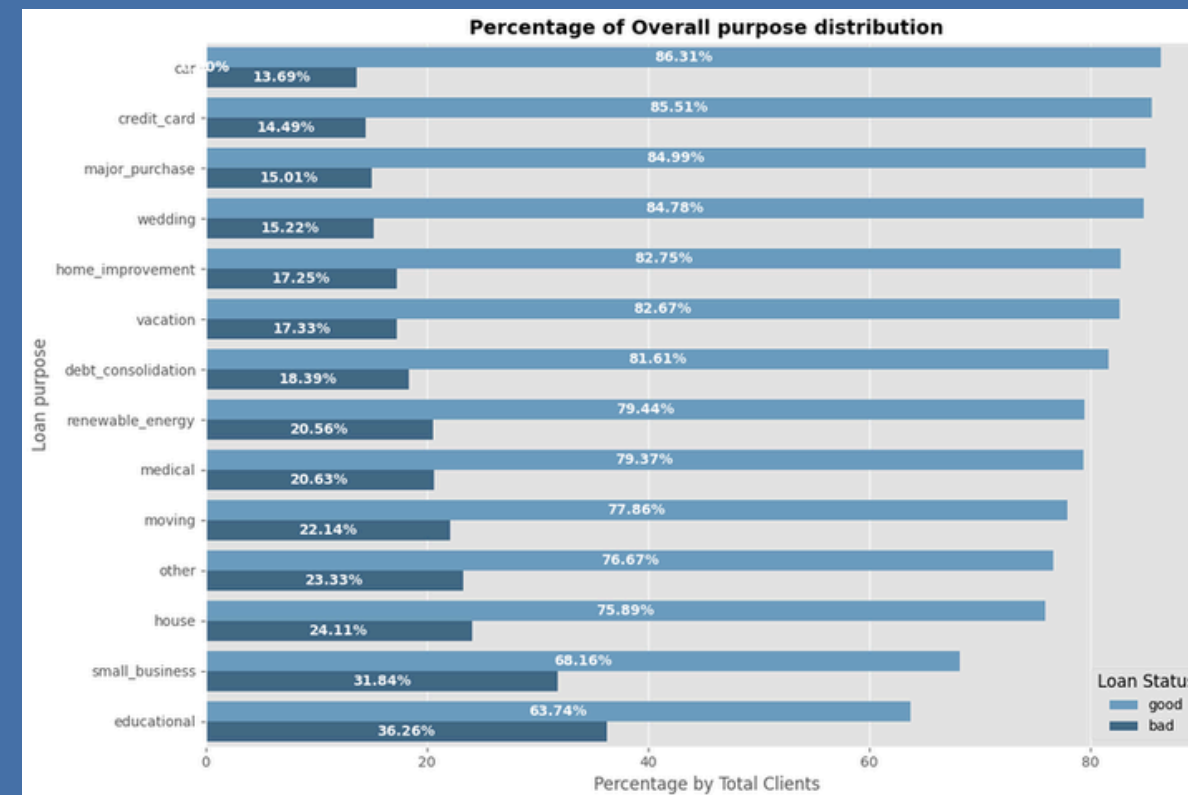
MULTIVARIATE

ANALYSIS



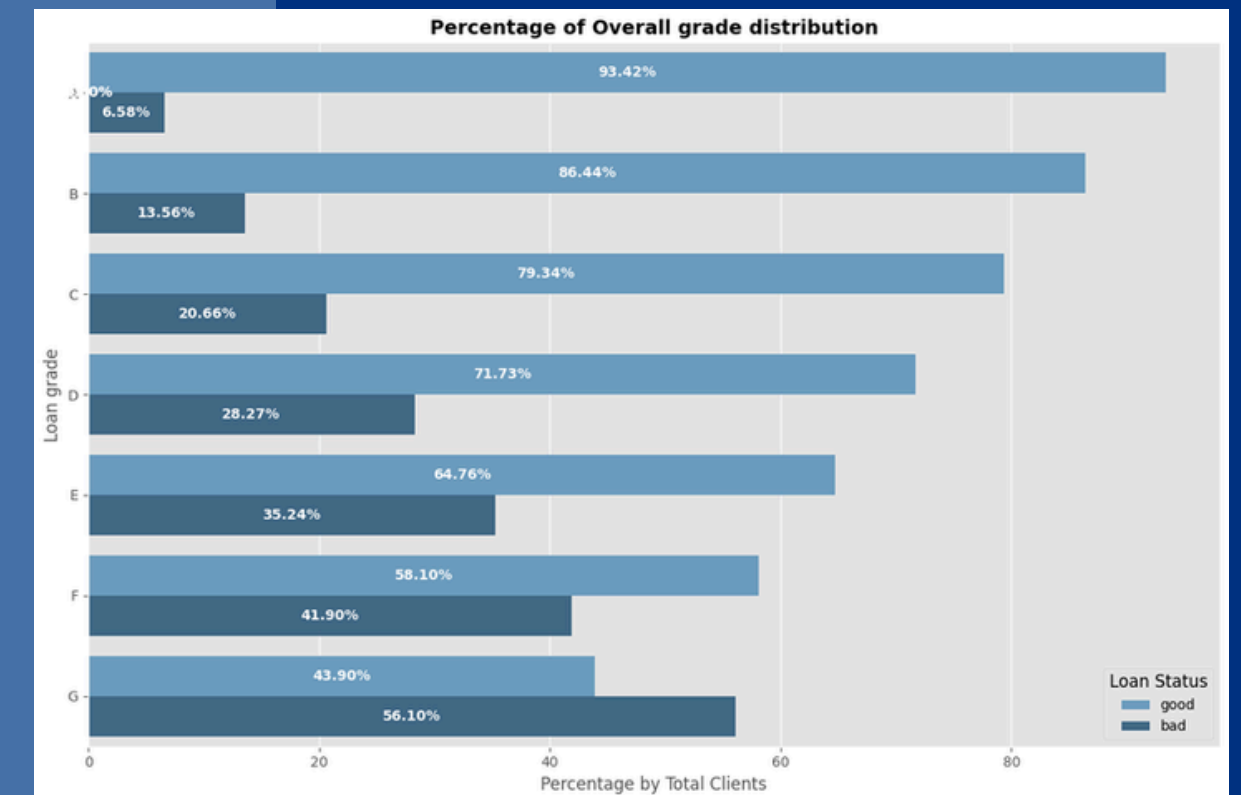
Trem

Semakin lama tenor pinjaman, semakin tinggi proporsi peminjam dengan status "bad", menunjukkan risiko gagal bayar lebih besar pada pinjaman jangka panjang.



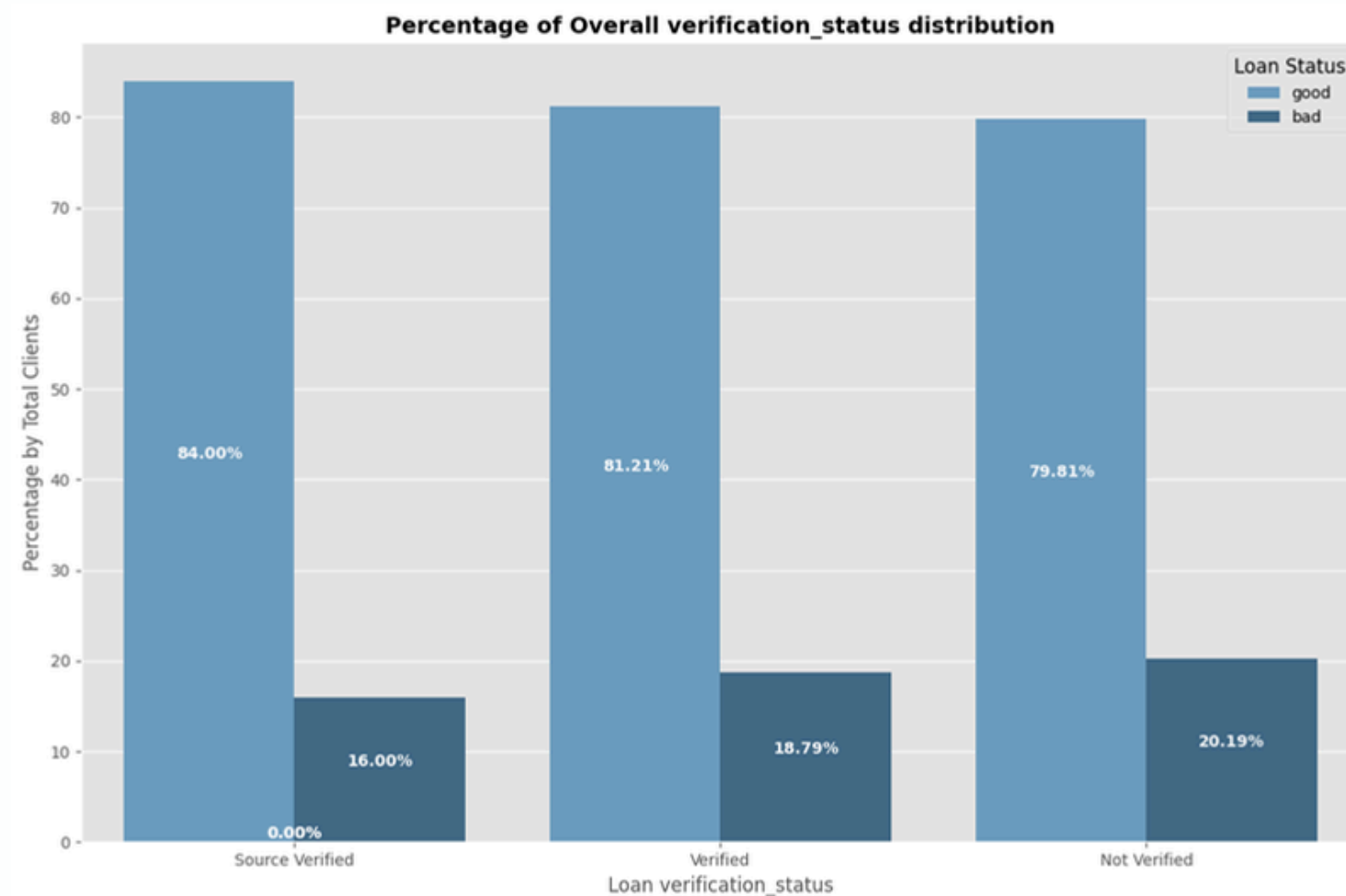
Purpose

Pinjaman pendidikan dan usaha kecil berisiko gagal bayar tertinggi, sementara kendaraan dan kartu kredit lebih rendah.



Grade

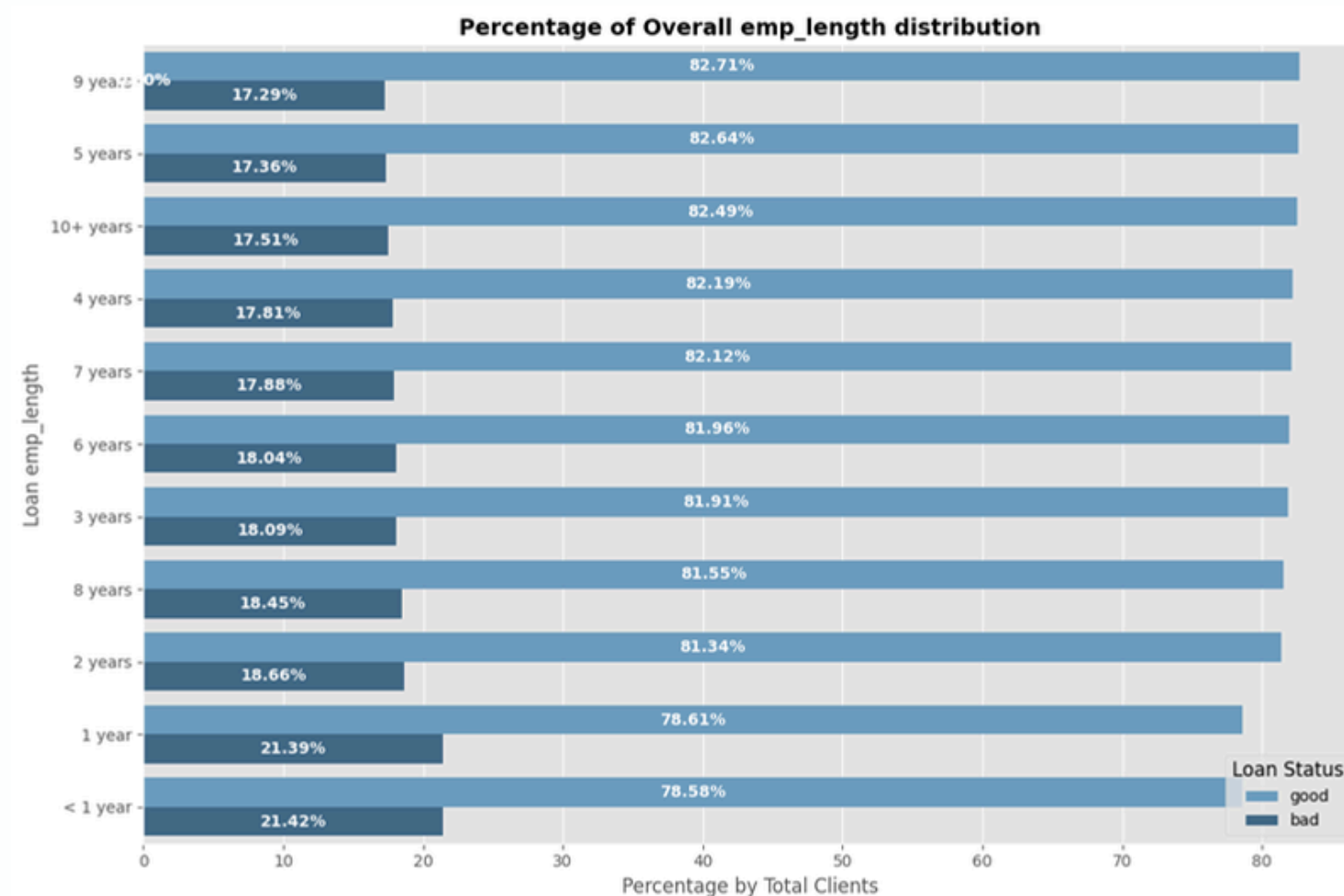
Grade rendah (F dan G) berisiko tinggi gagal bayar (>40%), sementara grade A dan B lebih aman dengan risiko minimal.



01

Verification Status

Klien dengan status verifikasi "Not Verified" memiliki risiko gagal bayar tertinggi (20,19%), sedangkan "Source Verified" paling rendah (16%), menunjukkan bahwa verifikasi meningkatkan kualitas kredit.

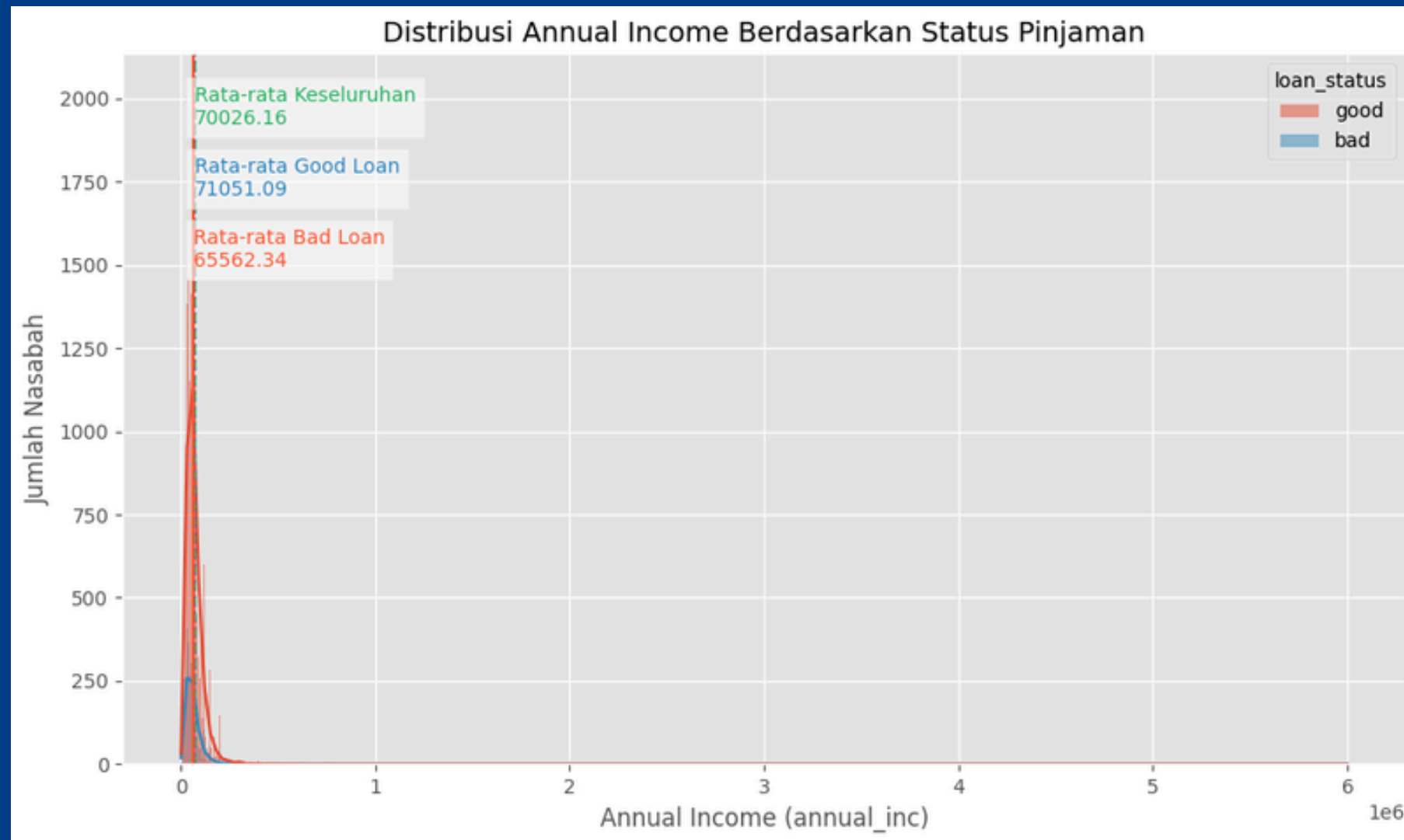


02

EMP Length

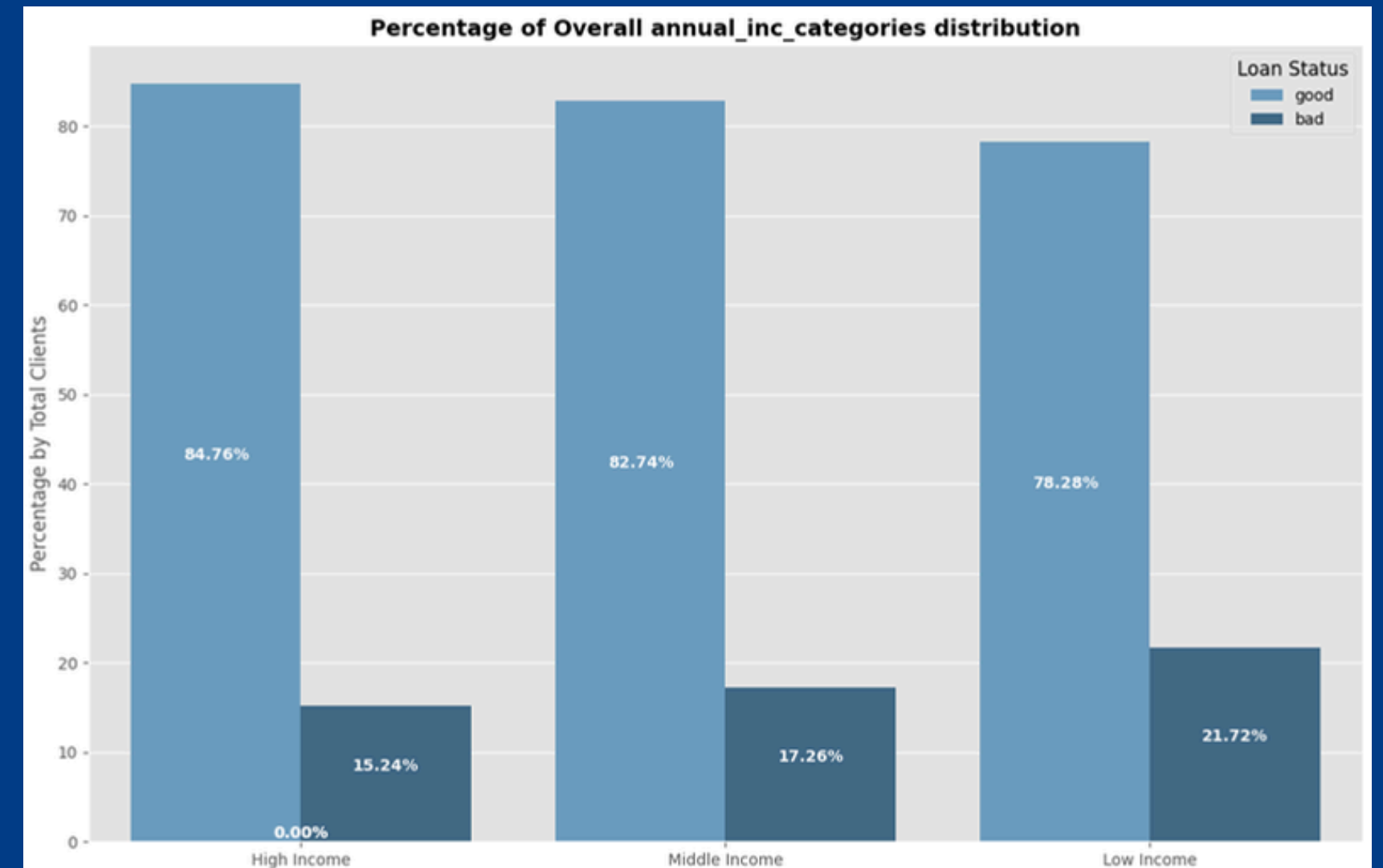
Semakin lama masa kerja, semakin rendah risiko gagal bayar. Klien dengan masa kerja <1 tahun memiliki tingkat gagal bayar tertinggi (21,42%), menunjukkan bahwa stabilitas pekerjaan berkontribusi pada kesehatan kredit.

PENDAPATAN TAHUNAN



Distribusi Pendapatan Tahunan

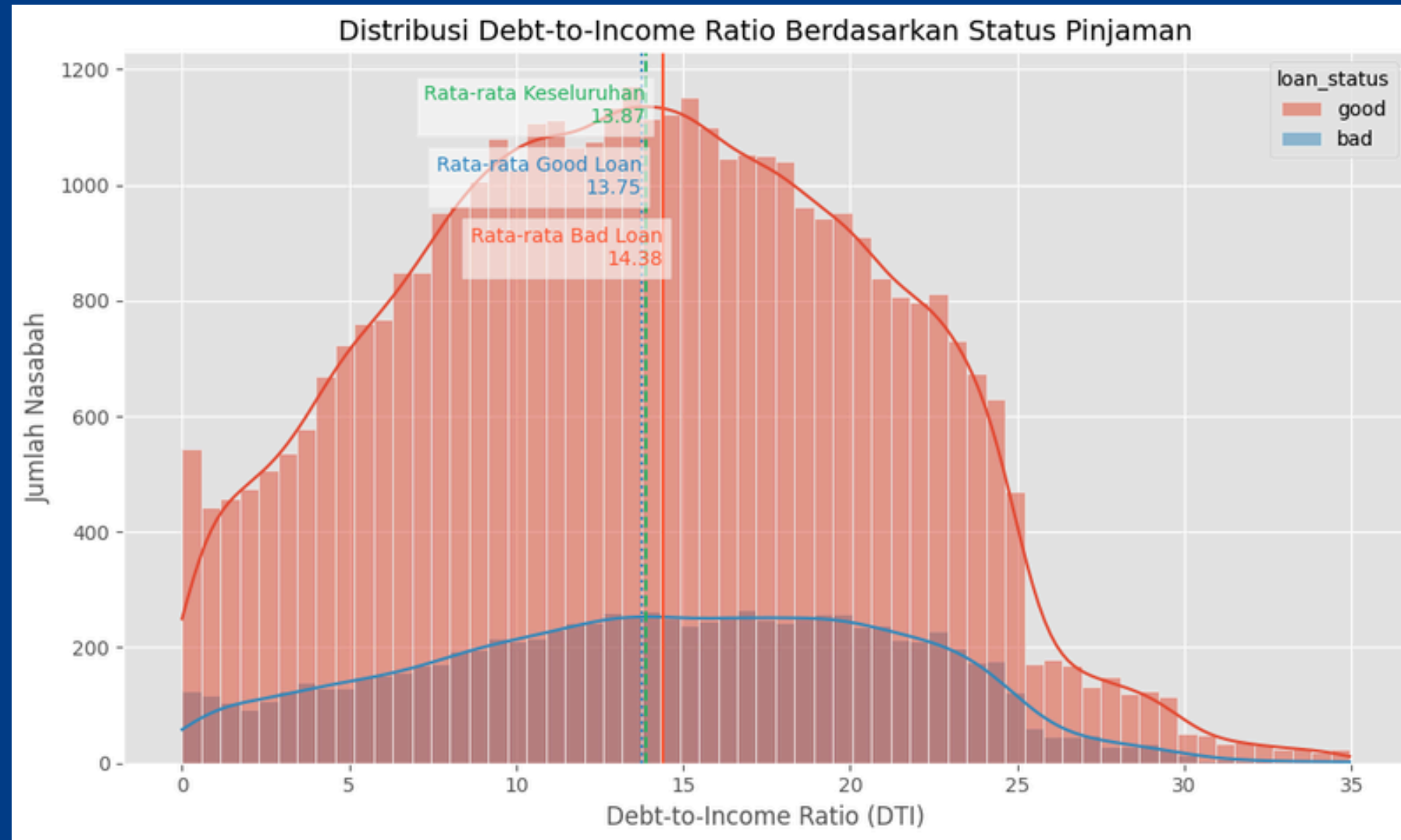
Klien dengan kredit baik memiliki pendapatan rata-rata lebih tinggi (\$71.051) dibandingkan kredit buruk (\$65.562), menunjukkan bahwa penghasilan lebih besar berkontribusi pada kelayakan kredit.



Persentase Segmentasi

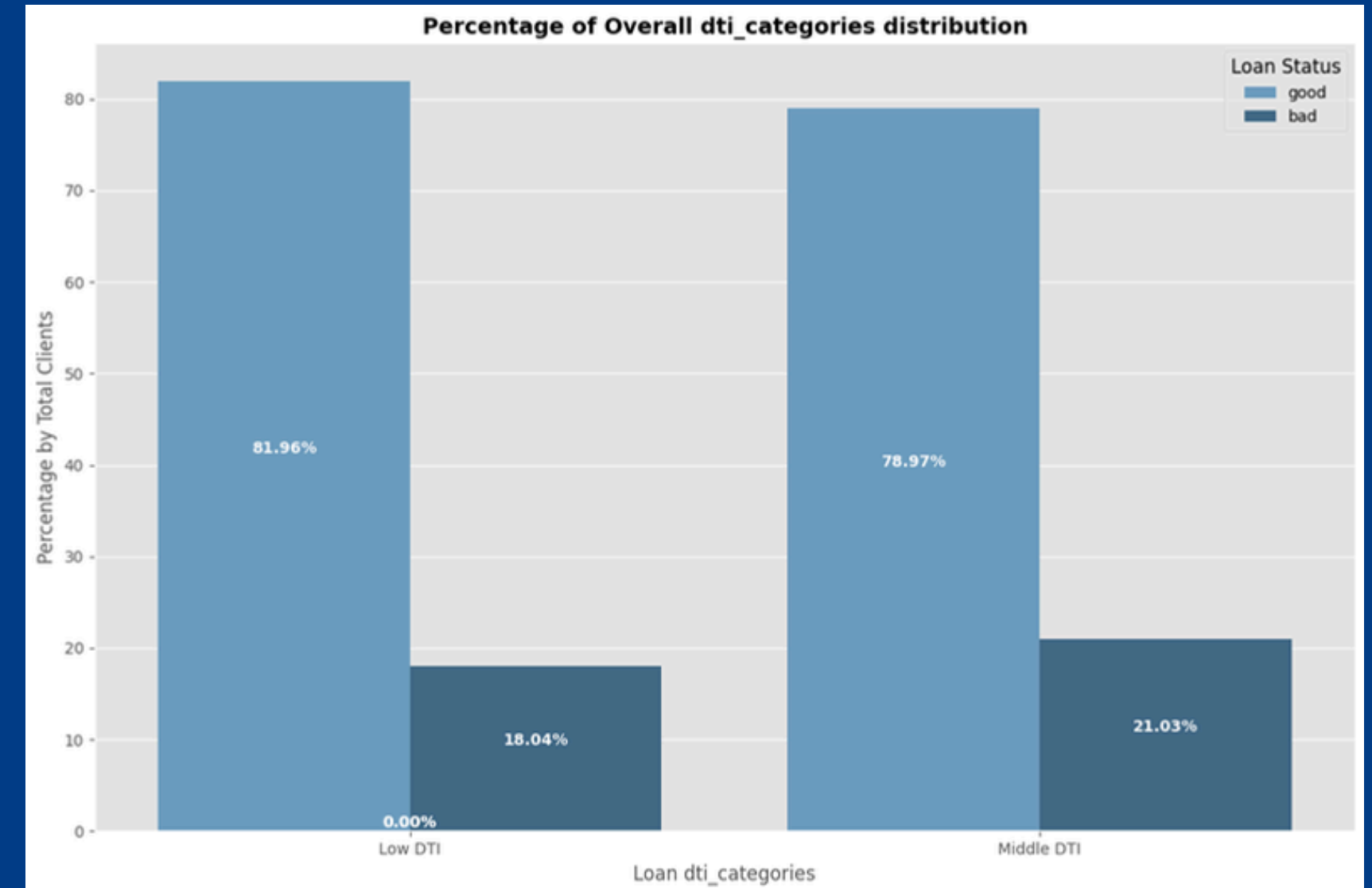
Semakin tinggi pendapatan, semakin rendah risiko gagal bayar. Klien berpenghasilan tinggi memiliki tingkat kredit macet terendah (15,24%), sedangkan berpenghasilan rendah tertinggi (21,72%).

DEBT-TO-INCOME RATIO



Distribusi Debt-to-Income Ratio

Semakin tinggi rasio Debt-to-Income (DTI), semakin besar risiko gagal bayar. Pinjaman bermasalah memiliki rata-rata DTI lebih tinggi (14.38) dibandingkan pinjaman lancar (13.75).



Persentase Segmentasi

Grafik menunjukkan bahwa nasabah dengan Debt-to-Income (DTI) rendah memiliki lebih sedikit pinjaman bermasalah (18.04%) dibandingkan nasabah dengan DTI menengah (21.03%). Ini menunjukkan bahwa semakin tinggi DTI, semakin besar risiko gagal bayar.

DATA PREPROCESSING

Impute Data



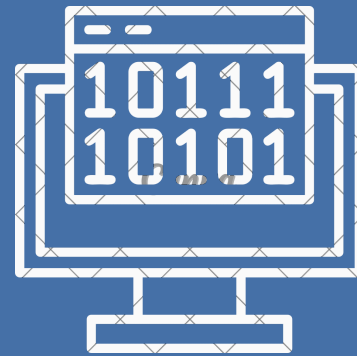
Mengisi null value dengan nilai median untuk feature numerikal dan mode untuk feature kategorikal

Feature Engineering



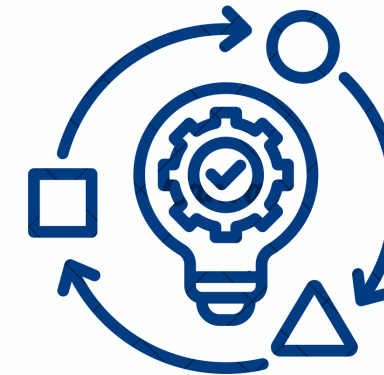
Menciptakan Feature baru dari feature yang ada sebelumnya untuk menemukan pola baru pada data

Encoding Values



Ubah data kategorikal menjadi numerik: gunakan Label Encoding untuk data ordinal dan One-Hot Encoding untuk data tanpa urutan.

Transformation Values



Mengubah skala nilai pada setiap feature agar memiliki ukuran yg serempak

Splitting Data & Data Balancing



Menyeimbangkan data antara data minorirtas dan mayoritas agar model dapat dilatih secara optimal

Kesimpulan Evaluasi Model

Dari beberapa model yang diuji, **Logistic Regression dengan SMOTE** menunjukkan performa terbaik dengan **F1-score 96%**, mengungguli **Decision Tree** yang mengalami **overfitting** dan **AdaBoost** yang memiliki **skor lebih rendah**. Oleh karena itu, **Logistic Regression dengan SMOTE** di implementasikan sebagai **model terbaik** untuk analisis kelayakan pinjaman.

MODELING

ABOUT OUR VISION AND MISSION

WITHOUT SMOTE

Without SMOTE											
	model	acc_train	acc_test	prec_train	prec_test	rec_train	rec_test	f1_train	f1_test	roc_train	roc_test
0	Logistic Regression	0.96	0.96	0.96	0.96	1.0	1.00	0.98	0.98	0.87	0.86
1	Decision Tree	1.00	0.96	1.00	0.98	1.0	0.98	1.00	0.98	1.00	0.91
2	Ada Boost	0.94	0.94	0.94	0.94	1.0	1.00	0.97	0.97	0.77	0.77

WITH SMOTE

With SMOTE											
	model	acc_train	acc_test	prec_train	prec_test	rec_train	rec_test	f1_train	f1_test	roc_train	roc_test
0	Logistic Regression	0.9	0.93	0.86	0.98	0.95	0.95	0.9	0.96	0.9	0.89
1	Decision Tree	1.0	0.94	1.00	0.97	1.00	0.96	1.0	0.97	1.0	0.89
2	Ada Boost	0.9	0.90	0.89	0.97	0.92	0.92	0.9	0.94	0.9	0.85

HYPER PARAMETER TUNING

Setelah mendapatkan algoritma terbaik yaitu **Logistic Regression**, selanjutnya dilakukan Hyperparameter Tuning untuk melihat parameter mana sajakah yang dapat memberikan output yang lebih maksimal

Parameter yang diuji:

- **C:** [0.001, 0.01, 0.1, 1, 10, 100]

Mengontrol kekuatan regularisasi (semakin kecil nilai C, semakin kuat regulasi).

- **Penalty:** ['l1', 'l2', 'elasticnet', None]

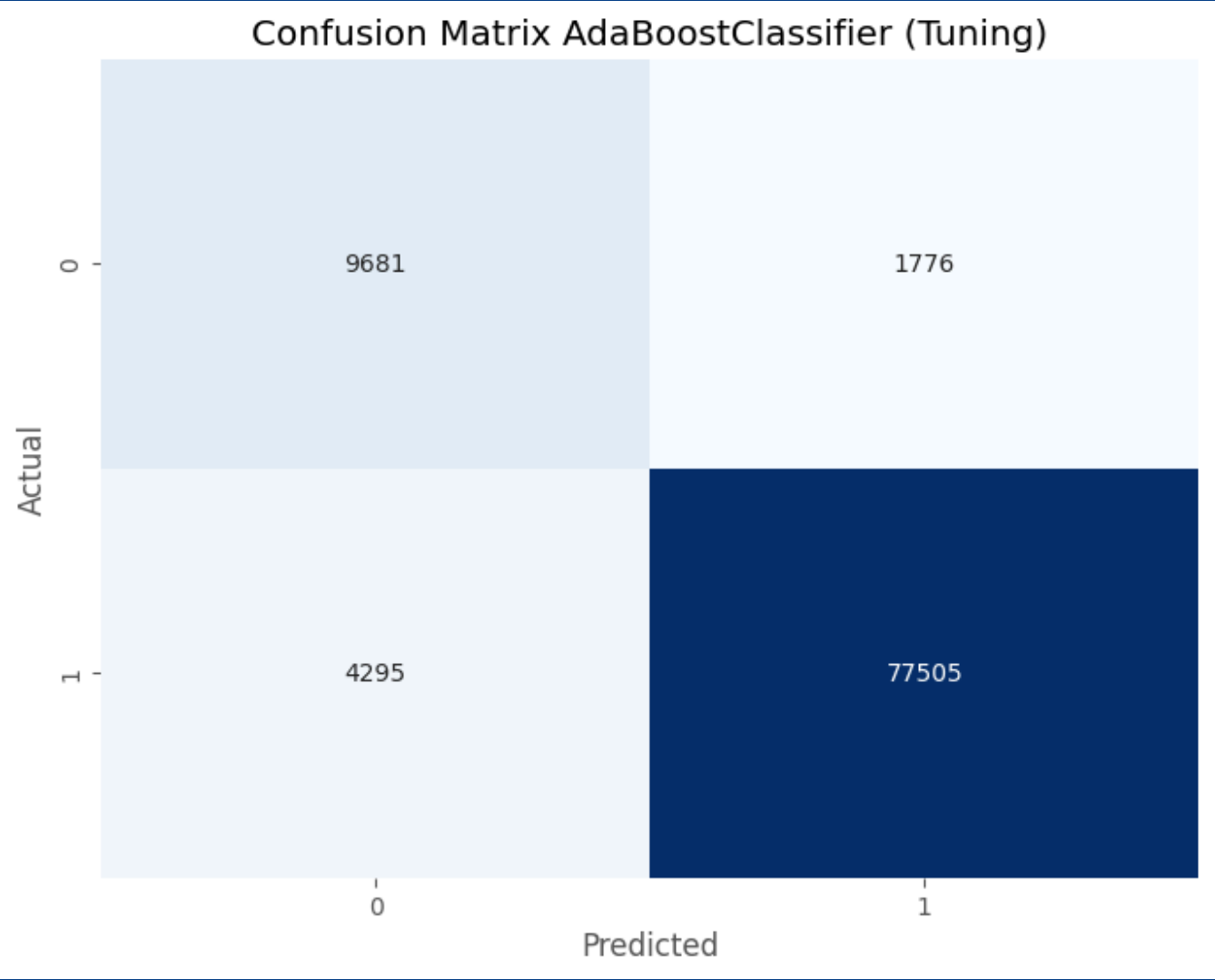
Jenis regularisasi yang digunakan untuk mencegah overfitting.

- **Solver:** ['liblinear', 'lbfgs', 'saga']

Algoritma optimasi yang digunakan untuk menyesuaikan bobot model.

- **Max Iter:** [100, 200, 500, 1000]

Jumlah iterasi maksimum untuk memastikan model.

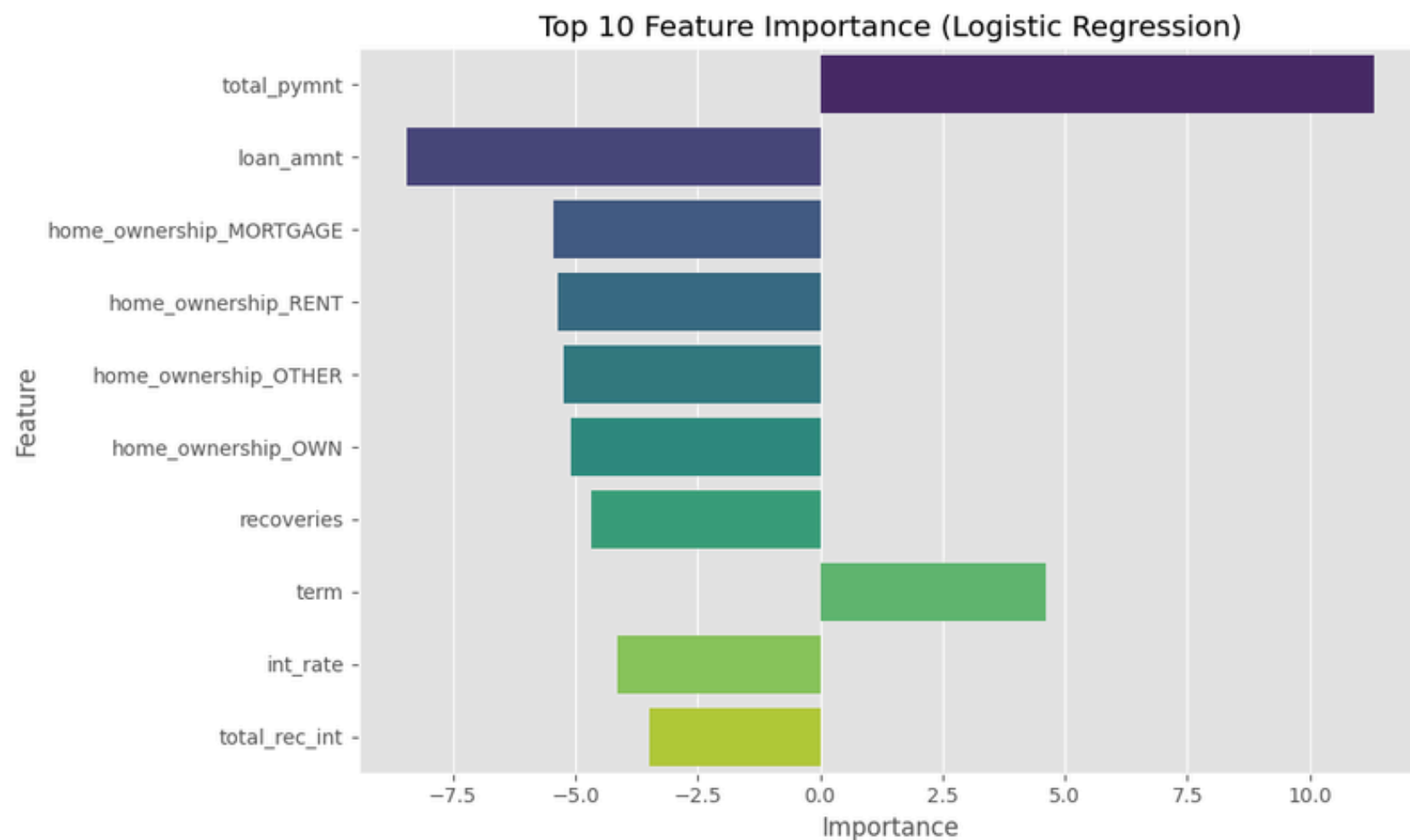


Best parameters:
{'solver':
'liblinear',
'penalty': 'l2',
'max_iter': 100,
'C': 10}

Acc Train	Acc Test	Prec Train	Prec Test	Rec Train	Rec Test	F1 Train	F1 Test	ROC Train	ROC Test
0.91	0.93	0.88	0.98	0.95	0.95	0.91	0.96	0.91	0.90

FEATURE

IMPORTANCE



Rekomendasi Strategi

1. **Total Pembayaran (↑)** – Terapkan sistem reward bagi nasabah dengan riwayat pembayaran baik.
2. **Loan Amount (↓)** – Tetapkan batas pinjaman konservatif dengan kenaikan bertahap.
3. **Home Ownership** – Lebih memperhatikan peminjam yang memiliki tunggangan pembayaran KPR/Penyewaan rumah
4. **Recoveries (↓)** – Perkuat tim penagihan proaktif sebelum jatuh tempo.
5. **Term (↑)** – Fokus pada tenor optimal dengan tingkat penyelesaian tinggi.
6. **Interest Rate (↓)** – Gunakan model pricing dinamis berbasis riwayat kredit.
7. **Total Received Interest (↓)** – Berikan insentif pelunasan cepat dengan diskon bunga.

IMPACT TO BUSSINESS



Model yang dibangun memiliki **F1-score 96%** dan **Recall 95%**, menunjukkan **kemampuan yang kuat dalam membedakan Good Loan dan Bad Loan**. Dengan keseimbangan yang baik antara precision dan recall, model ini dapat **mengidentifikasi risiko peminjam dengan lebih akurat dan efisien**.

88%

Tanpa menggunakan Machine Learning bank mendapatkan 88% peminjaman dengan status Good

95%

Dengan menggunakan Machine Learning bank dapat mengidentifikasi peminjaman yang berpotensi berstatus Good hingga 95%, naik 7%

THANK YOU

● Project Based Internship – id/x partners

Email
thoriq004@gmail.com

Github
rajwaAth

LinkedIn
www.linkedin.com/in/muhamad-rajwa-ath