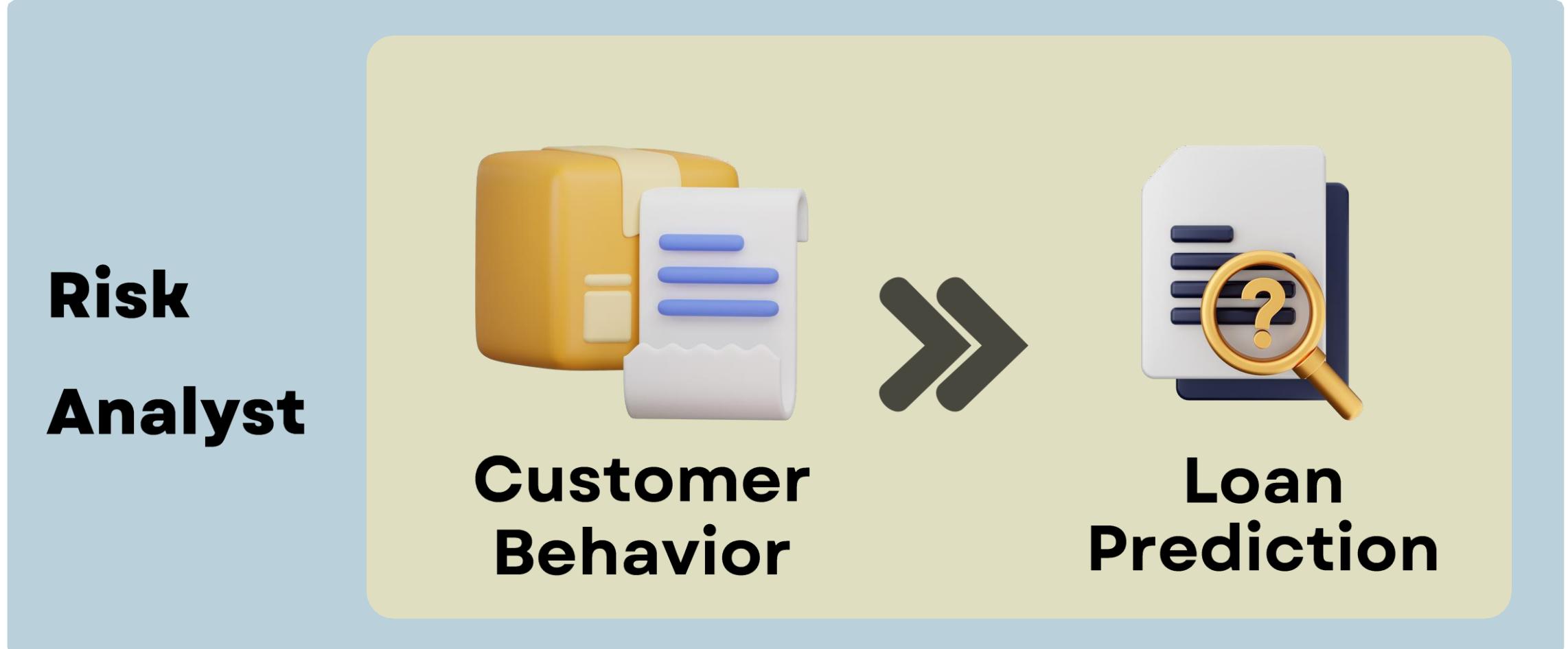


CREDIT LOAN PREDICTION ON CUSTOMER BEHAVIOR

Project Based Internship ID/X Partners

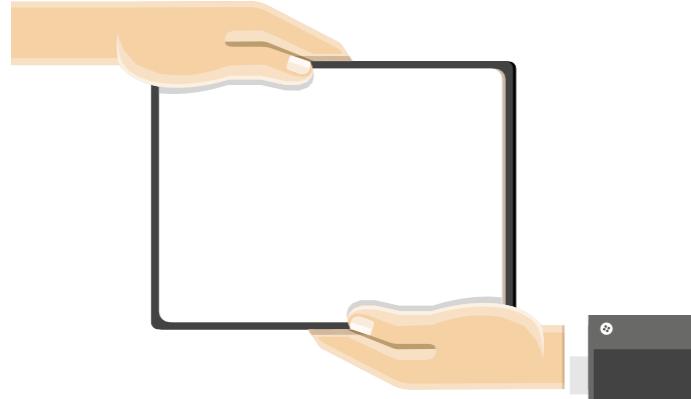
Project Overview



Business Process



Pendaftaran dan



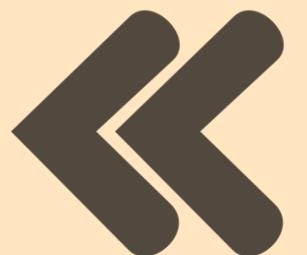
Penyerahan Berkas



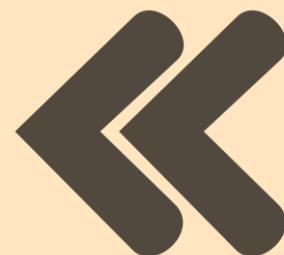
Verifikasi Data



Persetujuan
Kredit

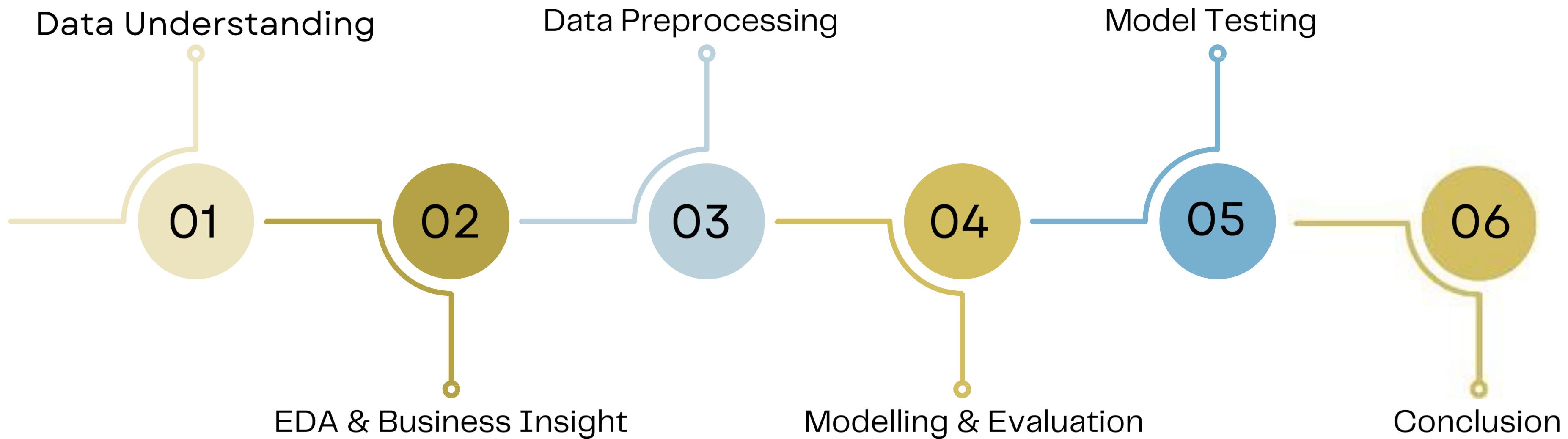


Analisa Agunan



Analisa Kelayakan
Kredit

Workflow



Data Understanding

Dataset



Kolom → 75
Baris → 252.000



Tipe Data

Categorical

Jumlah Kolom 22
dan
Loan_status (label)

Numerical

Jumlah Kolom 53

Problem



Pengajuan pinjaman semakin banyak namun Loan Approval dilakukan secara manual sehingga membutuhkan waktu yang cukup lama.

Loan Approval dalam waktu yang lama secara manual membutuhkan banyak tenaga kerja sehingga biaya dan tenaga kerja semakin tinggi



Goals

Mengurangi biaya, usaha, dan waktu proses Loan Approval melalui analisis customer behavior.



Objective

Memberikan rekomendasi loan approval berdasarkan customer behavior



Business Metrics

- Default rate
- Workforce cost

Exploratory Data Analysis & Business Insight

#0

FEATURE EXTRACTION

Melakukan klasifikasi pada setiap pinjaman dalam dataset menjadi dua kategori utama, yaitu "Good Loans" dan "Bad Loans", berdasarkan status akhir dari masing-masing pinjaman. Klasifikasi ini menjadi langkah awal dalam memahami dan mengelola risiko kredit.

Good Loans

- **Current:** Pinjaman yang masih berjalan dan pembayaran dilakukan sesuai jadwal.
- **Fully Paid:** Pinjaman yang telah dilunasi sepenuhnya oleh peminjam.
- **In Grace Period:** Pinjaman yang masih dalam masa tenggang pembayaran.
- **Does not meet the credit policy.**
Status:Fully Paid: Pinjaman yang, meskipun tidak memenuhi kebijakan kredit, tetapi telah dilunasi sepenuhnya.

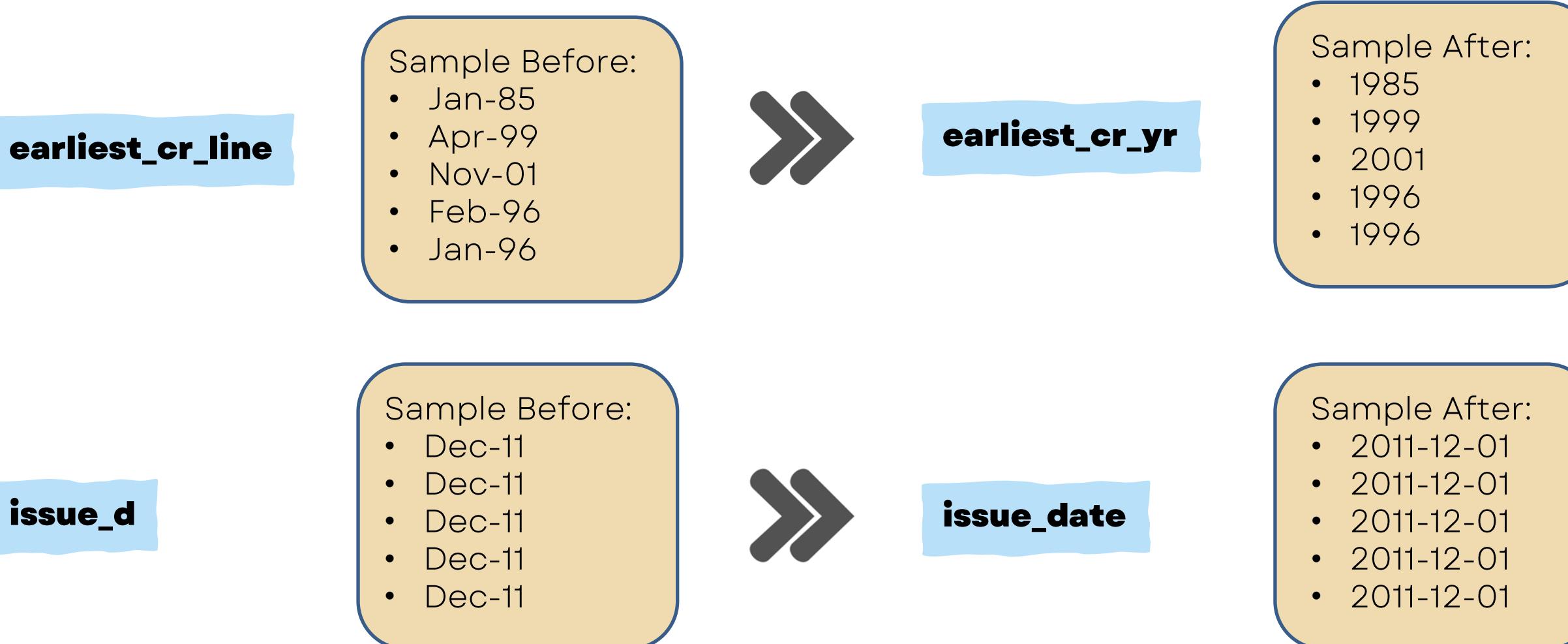
Bad Loans

- **Charged Off:** Pinjaman yang dianggap tidak dapat dipulihkan dan dihapuskan sebagai kerugian.
- **Late (16-30 days):** Pinjaman yang memiliki keterlambatan pembayaran antara 16 hingga 30 hari.
- **Late (31-120 days):** Pinjaman yang memiliki keterlambatan pembayaran antara 31 hingga 120 hari.
- **Default:** Pinjaman yang tidak dapat dipulihkan dan dianggap gagal bayar.
- **Does not meet the credit policy.**
Status:Charged Off: Pinjaman yang, meskipun tidak memenuhi kebijakan kredit, dianggap tidak dapat dipulihkan.

#0

FEATURE EXTRACTION

Membuat fitur baru dari earliest_cr_line dan issue_d untuk melihat tren jumlah peminjam baik dan buruk per tahun berdasarkan tanggal pembukaan kredit dan tanggal penerbitan pinjaman.



#1

DISTRIBUSI GOOD_BAD_LOAN BERDASARKAN VERIFIKASI_STATUS

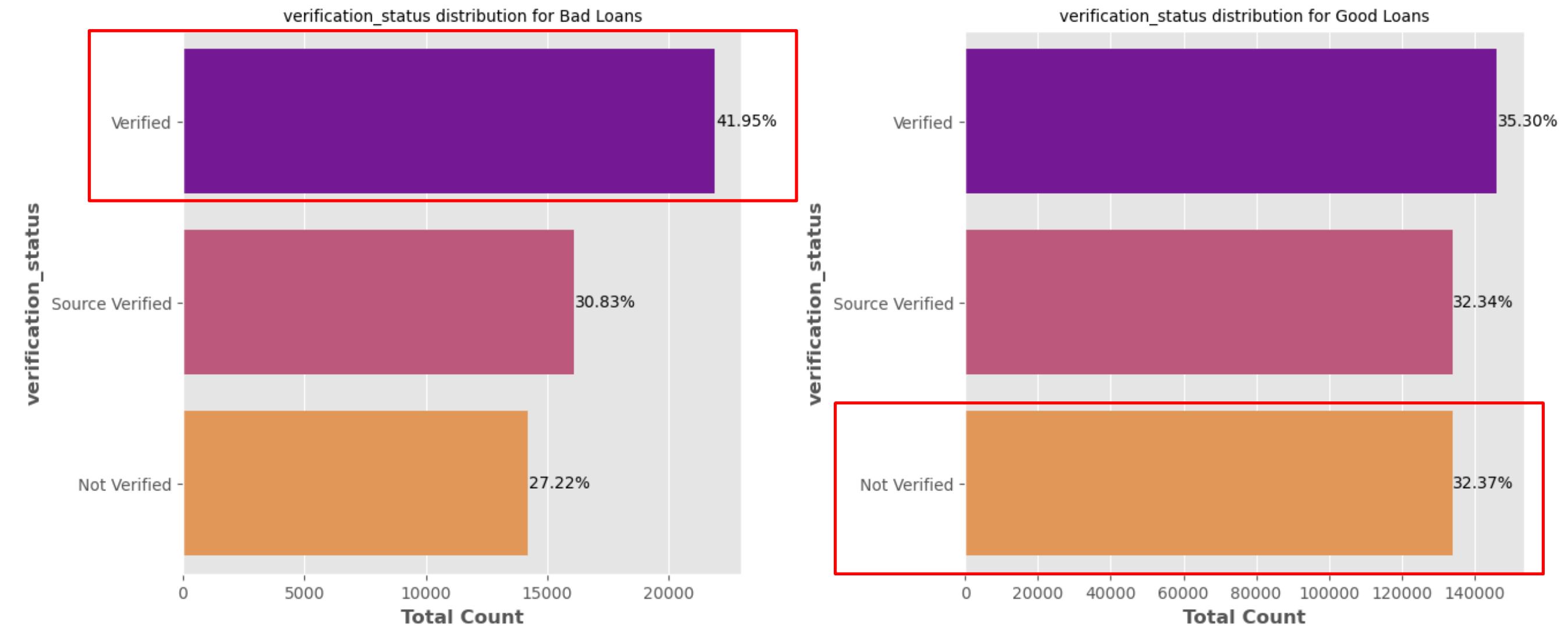
Insight:

Proporsi pinjaman yang diverifikasi (Verified & Source Verified) cenderung lebih tinggi pada pinjaman buruk dibandingkan dengan pinjaman baik (Verified: 41.91%, Source Verified: 30.83). Sebaliknya, proporsi pinjaman yang tidak diverifikasi cenderung lebih tinggi pada pinjaman baik.

Potensial Masalah



Verifikasi pinjaman
yang tidak efektif



* Memisahkan pinjaman baik atau buruk dan membuat persentase masing-masing.

#2

DISTRIBUSI GOOD_BAD_LOAN BERDASARKAN HOME_OWNERSHIP

Insight:

Individu yang menyewa rumah (Rent) cenderung lebih tinggi pada pinjaman buruk (46.57%), sedangkan proporsi individu yang memiliki hipotek (Mortgage) cenderung lebih tinggi pada pinjaman baik (51.35%).

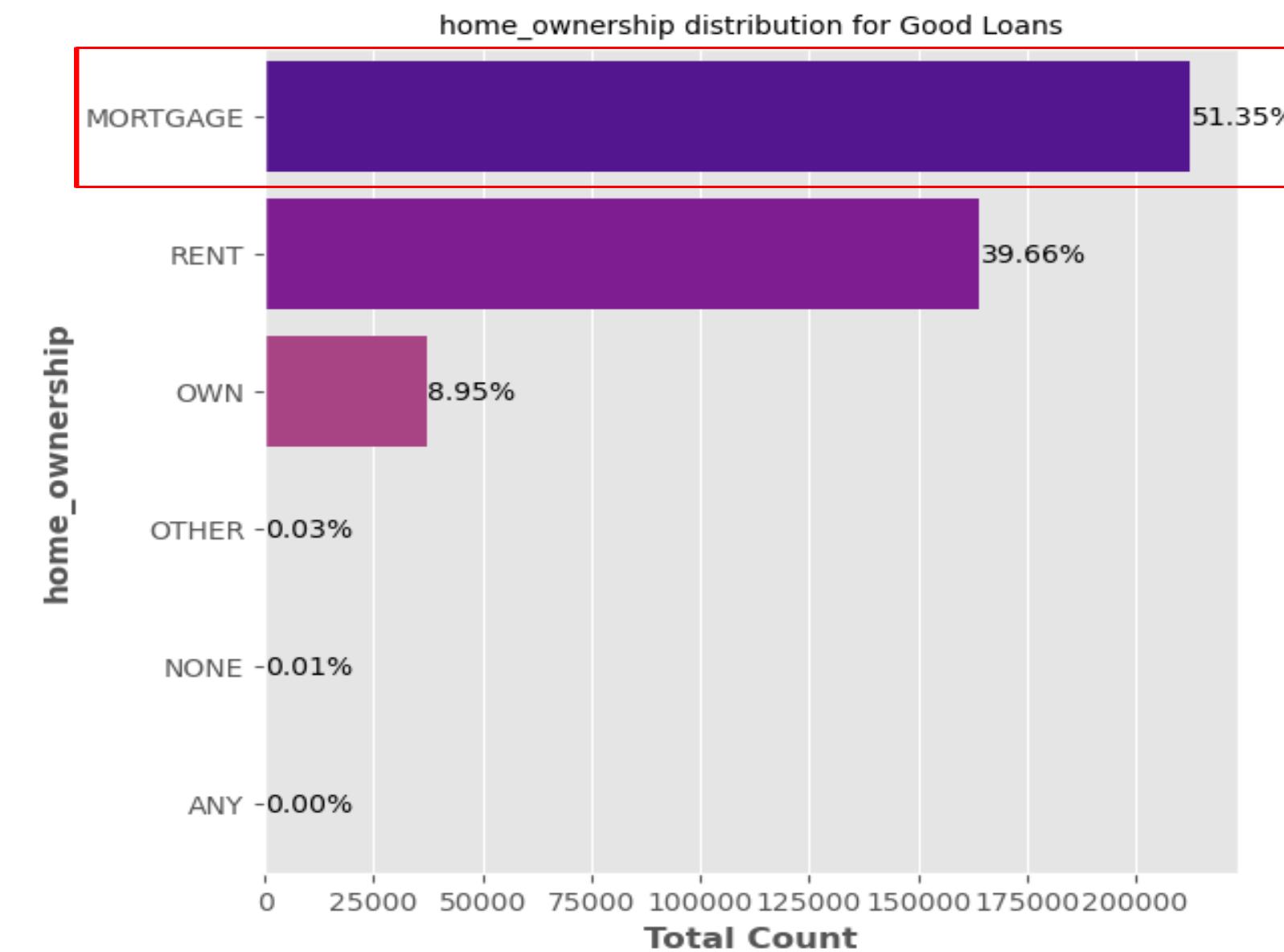
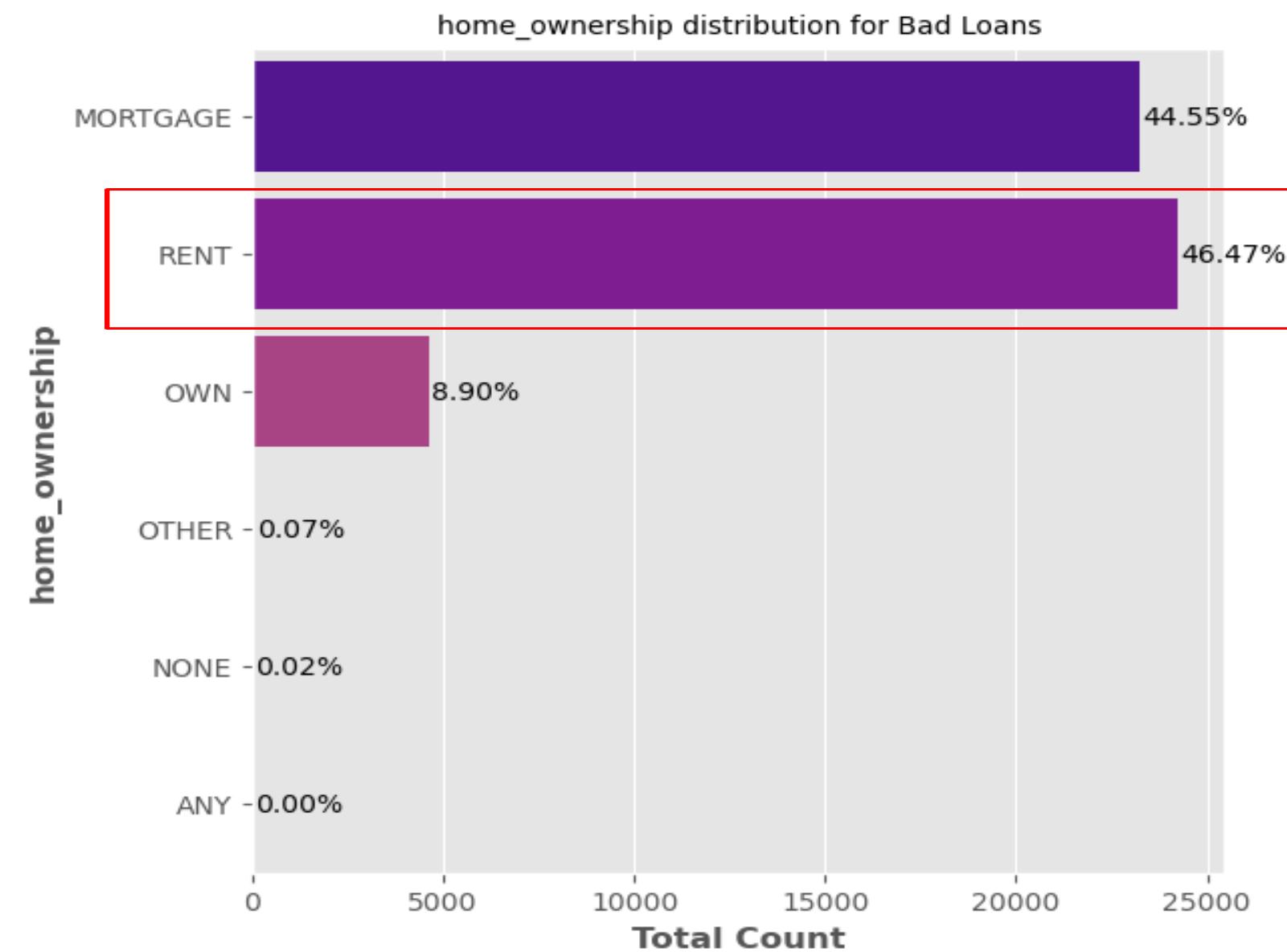
Potensial Masalah



Rendahnya Stabilitas keuangan



Tanggungan Finansial Tinggi



#3

DISTRIBUSI GOOD_BAD_LOAN BERDASARKAN TERM

Insight:

Proporsi pinjaman dengan durasi 36 bulan cenderung lebih tinggi pada pinjaman baik (74.50%), sedangkan proporsi pinjaman dengan durasi 60 bulan cenderung lebih tinggi pada pinjaman buruk (35.66%).

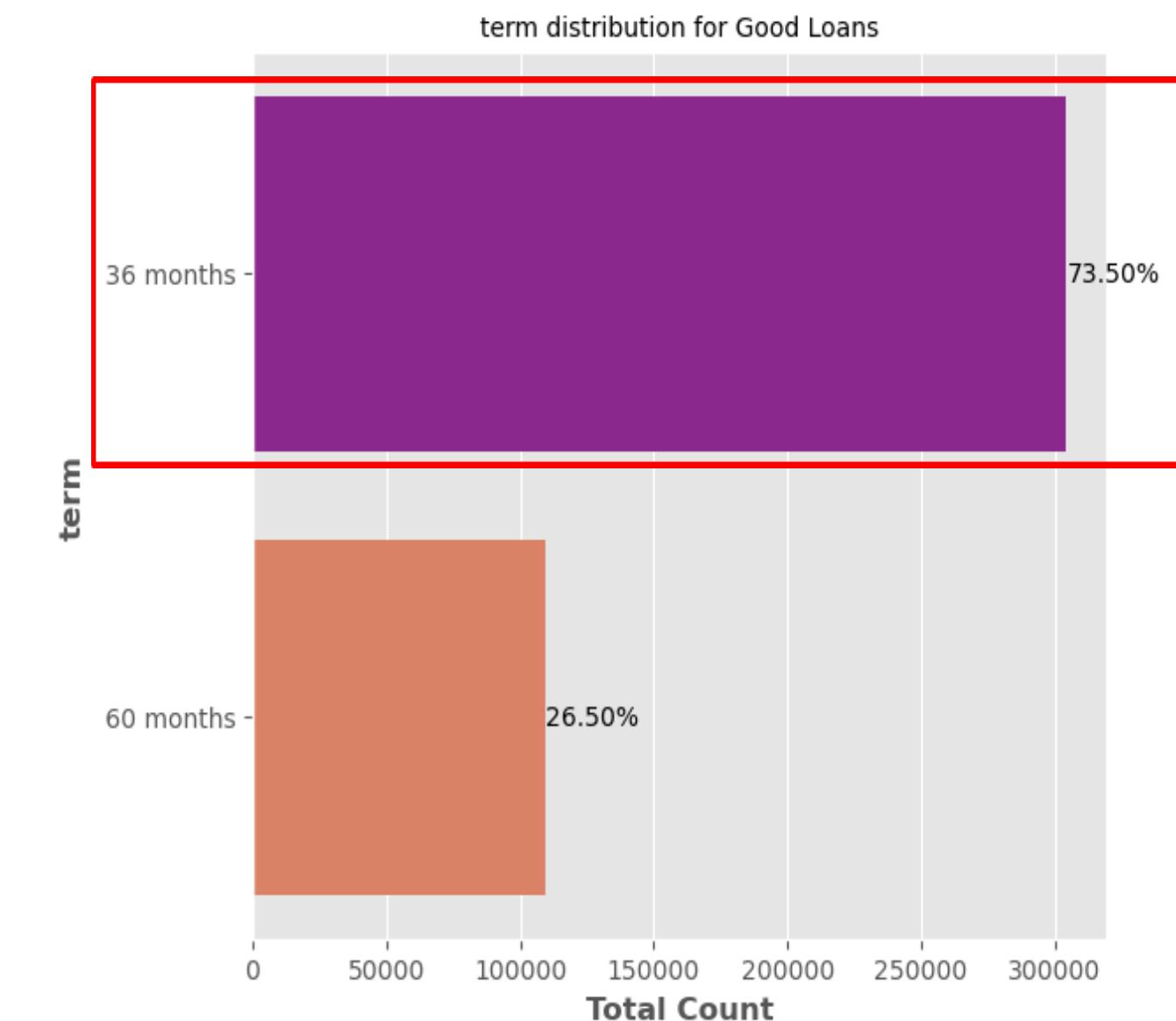
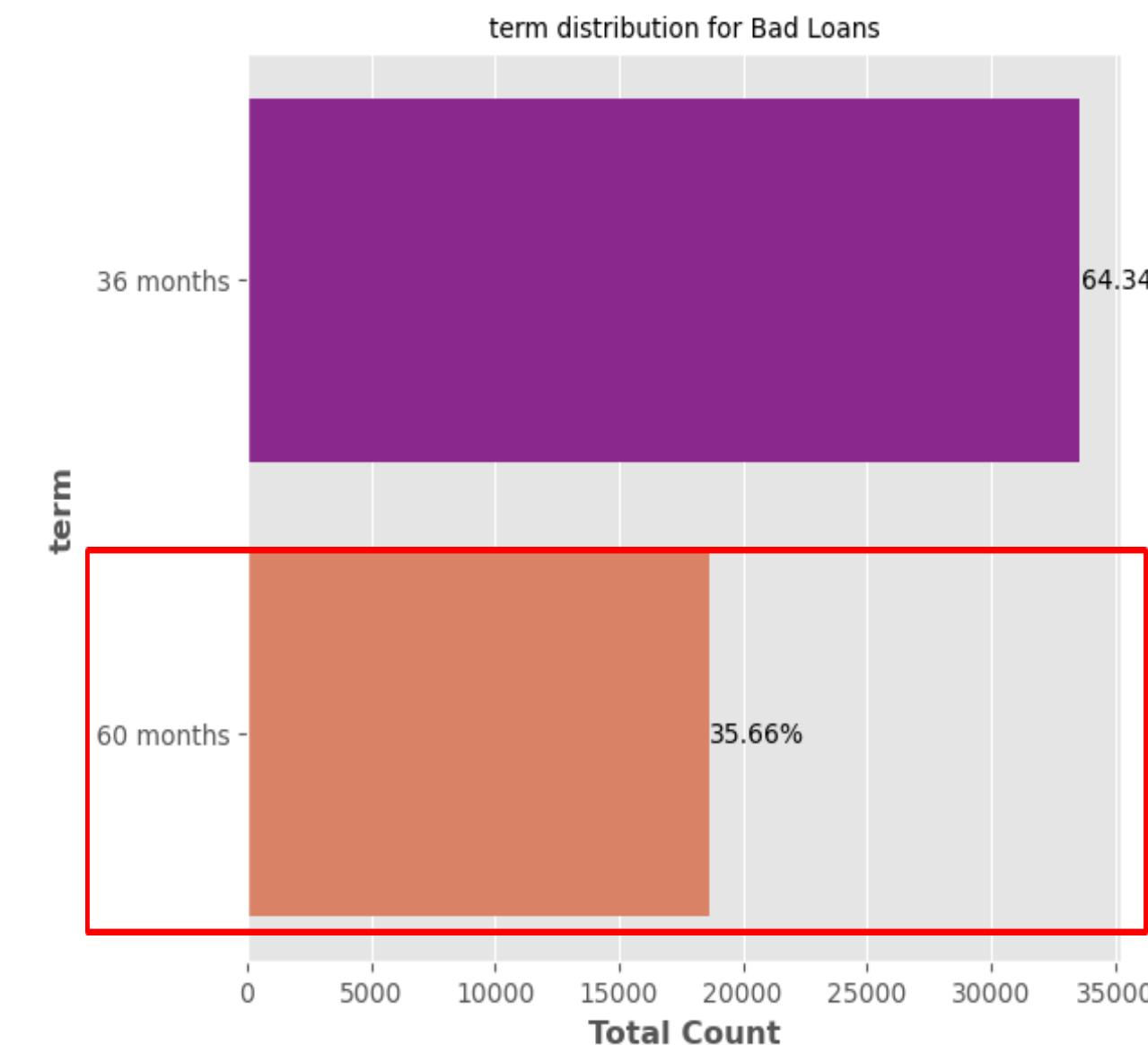
Potensial Masalah



Ketidakpastian ekonomi



Resiko jangka panjang



#4

DISTRIBUSI GOOD_BAD_LOAN BERDASARKAN ISSUE_DATE_YEAR

Insight:

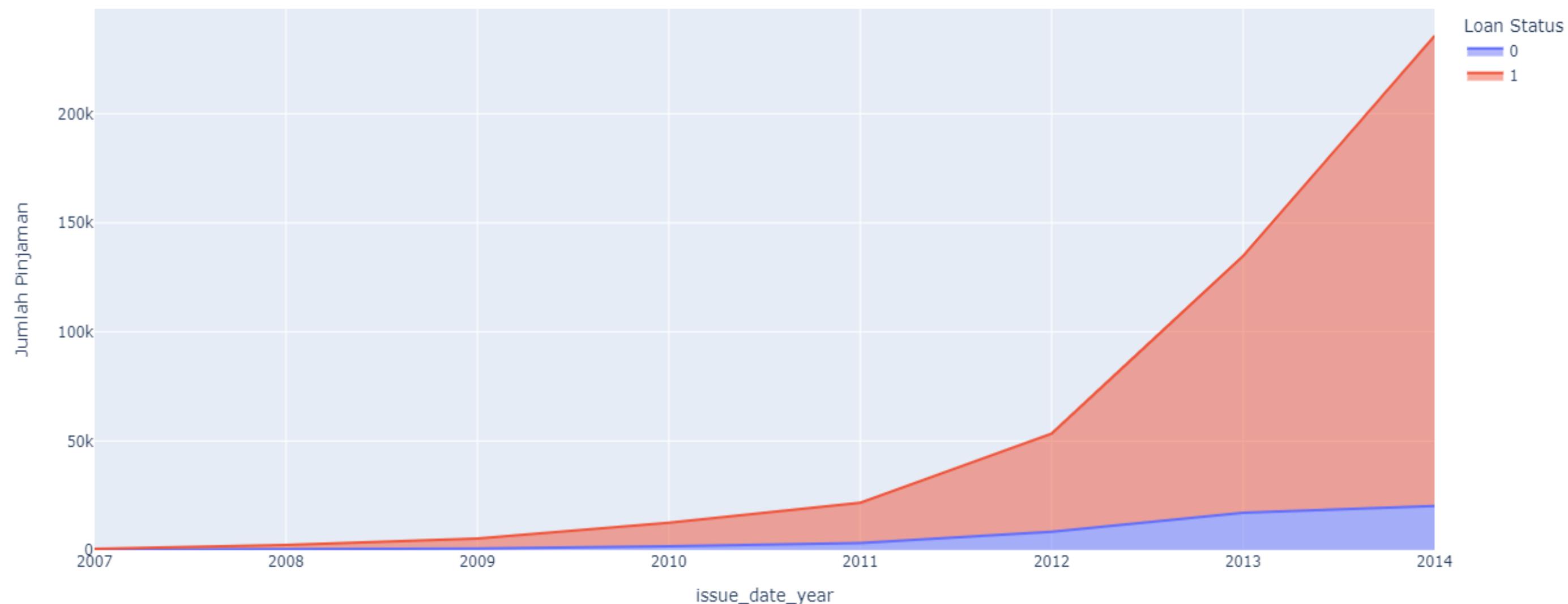
Berdasarkan Berdasarkan Tanggal Pembukaan Kredit Pertama, sejak tahun 2011 jumlah pinjaman mengalami kenaikan lebih dari dua kali lipat setiap tahunnya dan angka bad loan-pun masih terkendali.

Potensial Masalah



Peningkatan risiko

Tren Jumlah Pinjaman Baik dan Buruk per Tahun Berdasarkan Tanggal Pembukaan Kredit Pertama



Data Preprocessing

Data PreProcessing

Data Cleaning

Menghapus kolom dengan
lebih dari 70% data null

Menghapus kolom yang
memiliki korelasi kuat

Menghapus kolom dengan IV
(information value) < 0,02

Data PreProcessing

Data Transformation

- Melakukan normalisasi kolom yang memiliki distribusi skew > 0.5 atau < -0.5
- Melakukan Standarisasi kolom yang memiliki distribusi normal antara 0.5 sampai dengan -0.5
- Melakukan One-Hot Encoding untuk kolom-kolom kategorikal

Machine Learning Model

MODELLING

Process of Choosing the Algorithm

SPLITTING



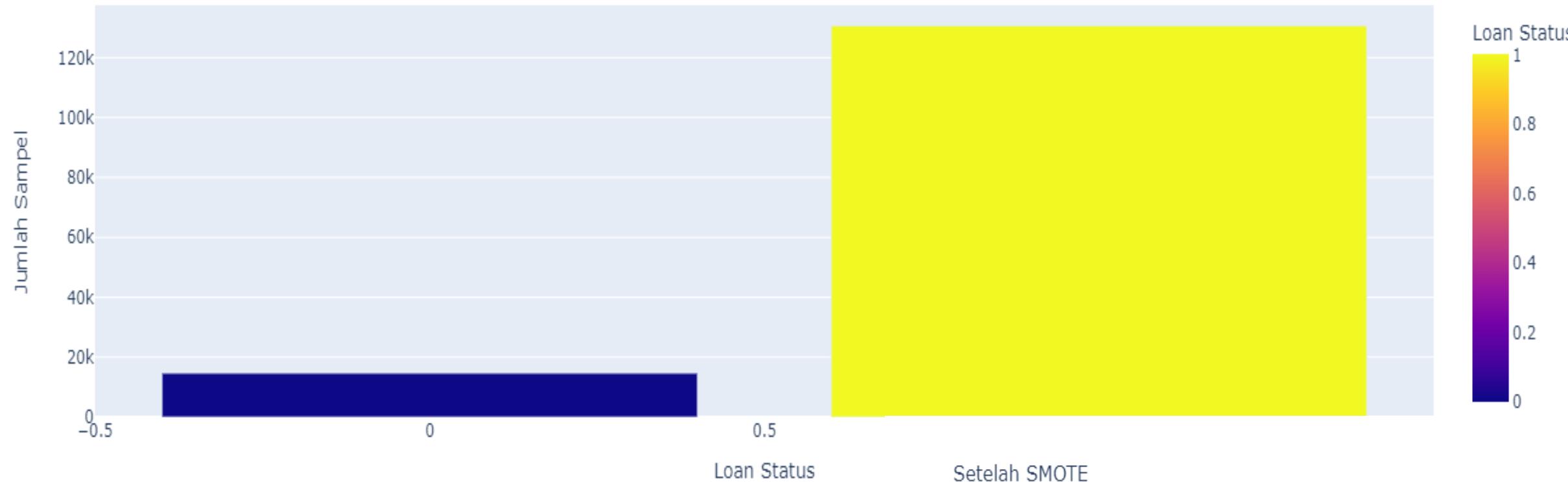
ALGORITHM SCORE

No	Algoritma	Precision	Recall	F1-Score
1	XGBoost	0.97	1.00	0.98
2	Random Forest	0.95	0.99	0.97

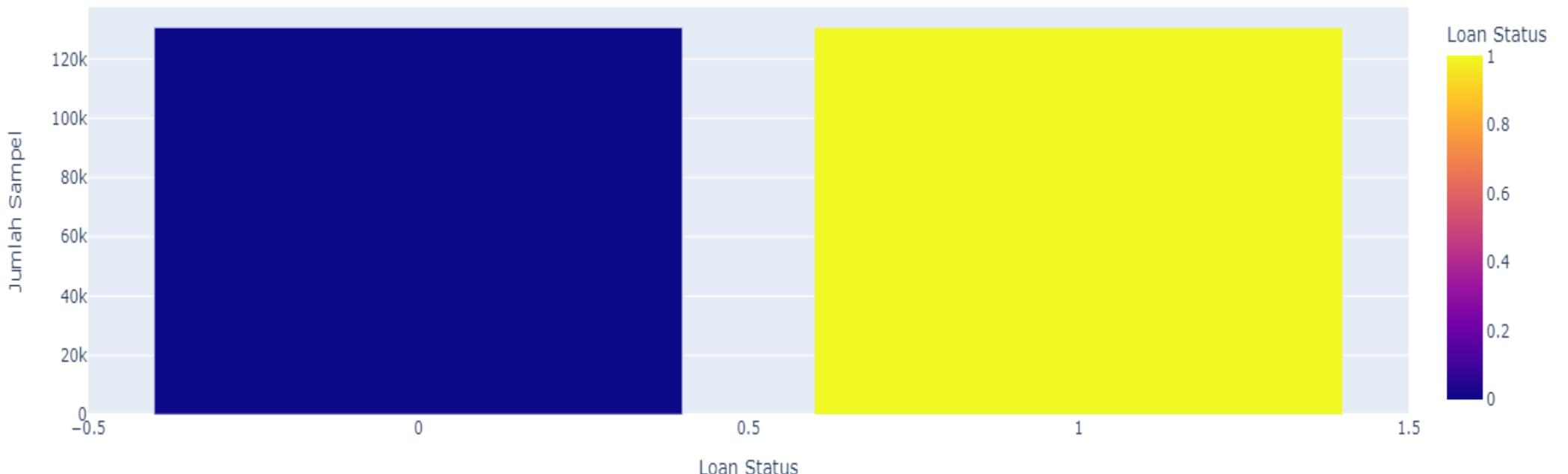
MODELLING

Handle Class Imbalance Sampling SMOTE pada Data Train

Sebelum SMOTE

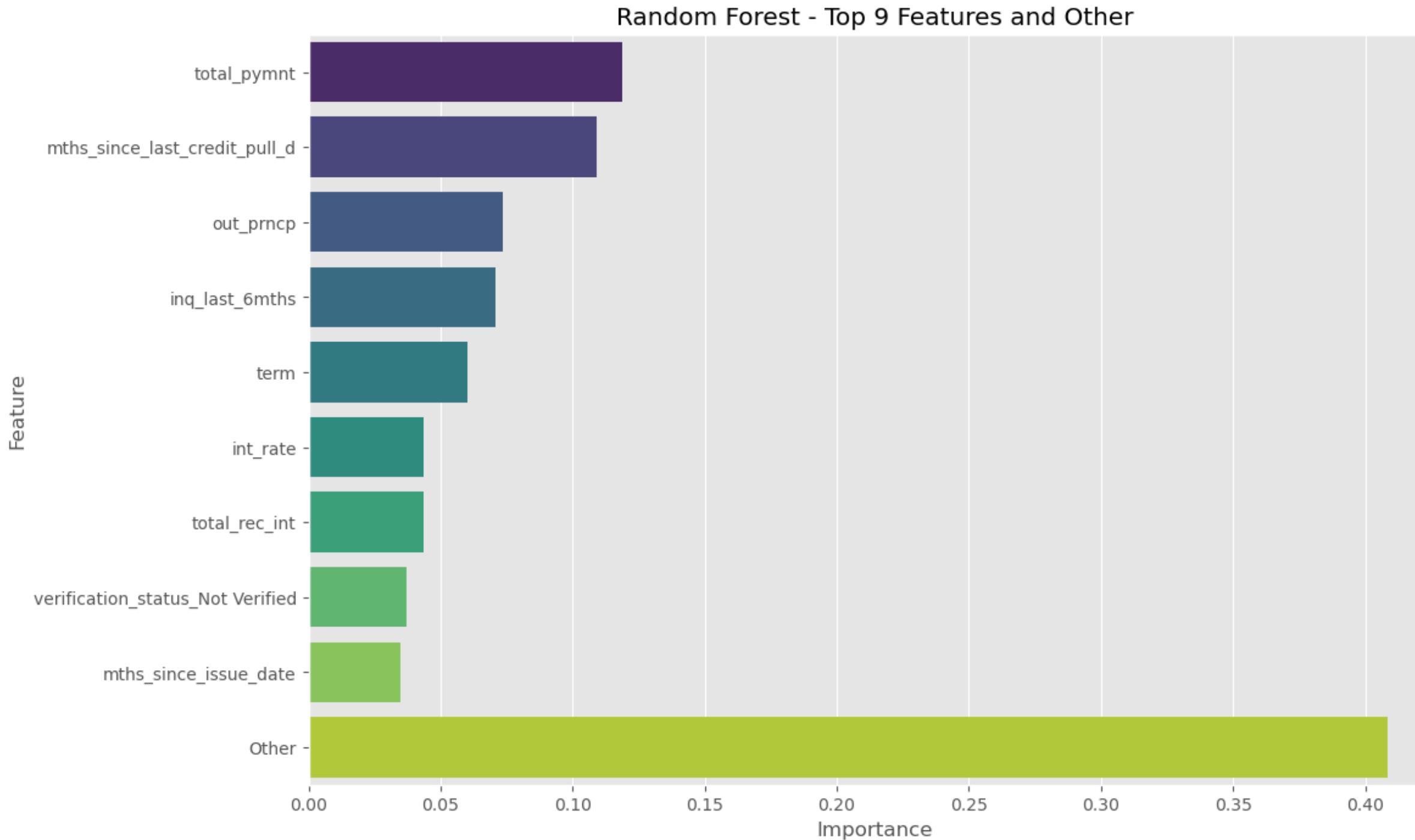


Setelah SMOTE



MODELLING

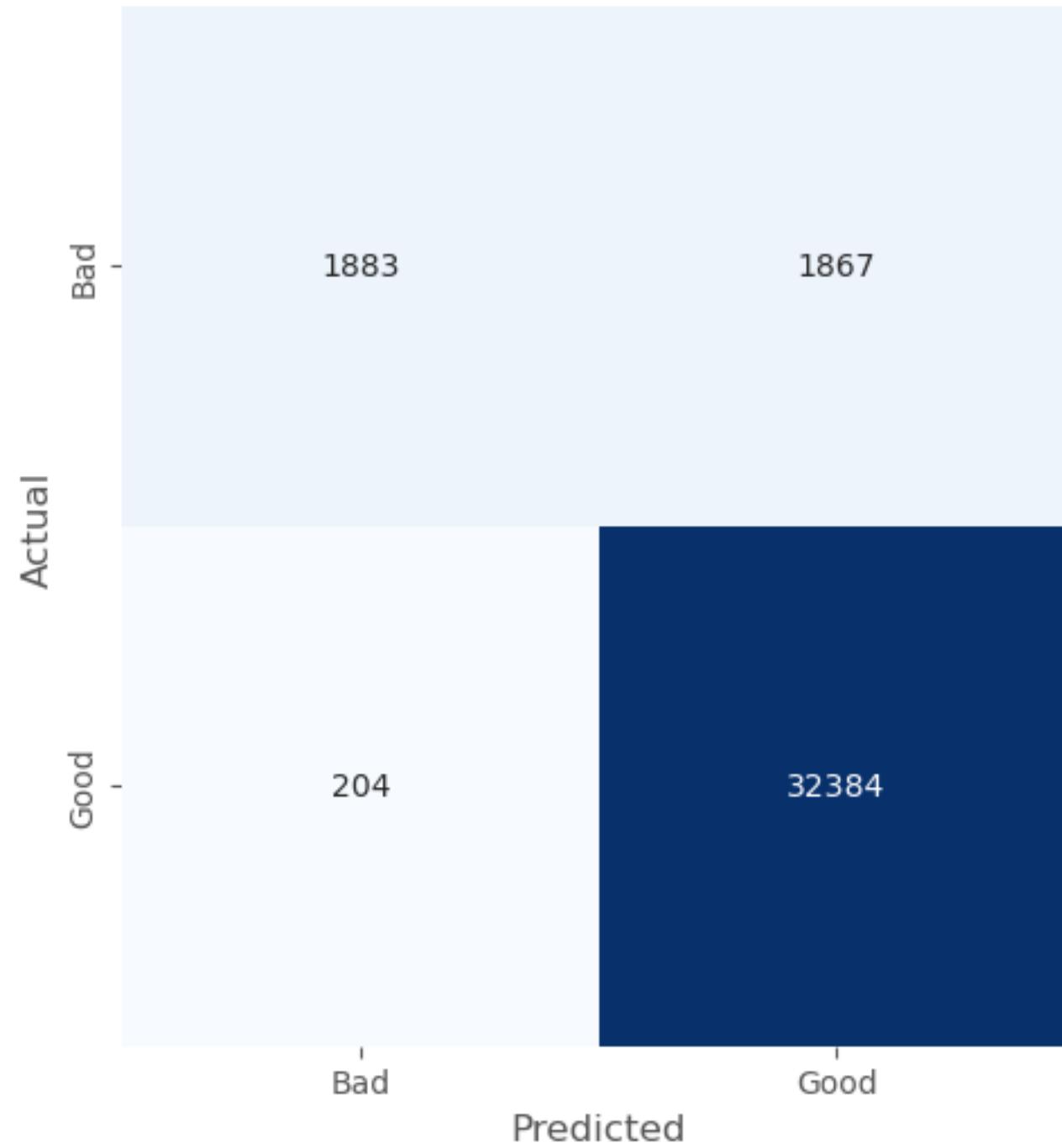
Feature Importance



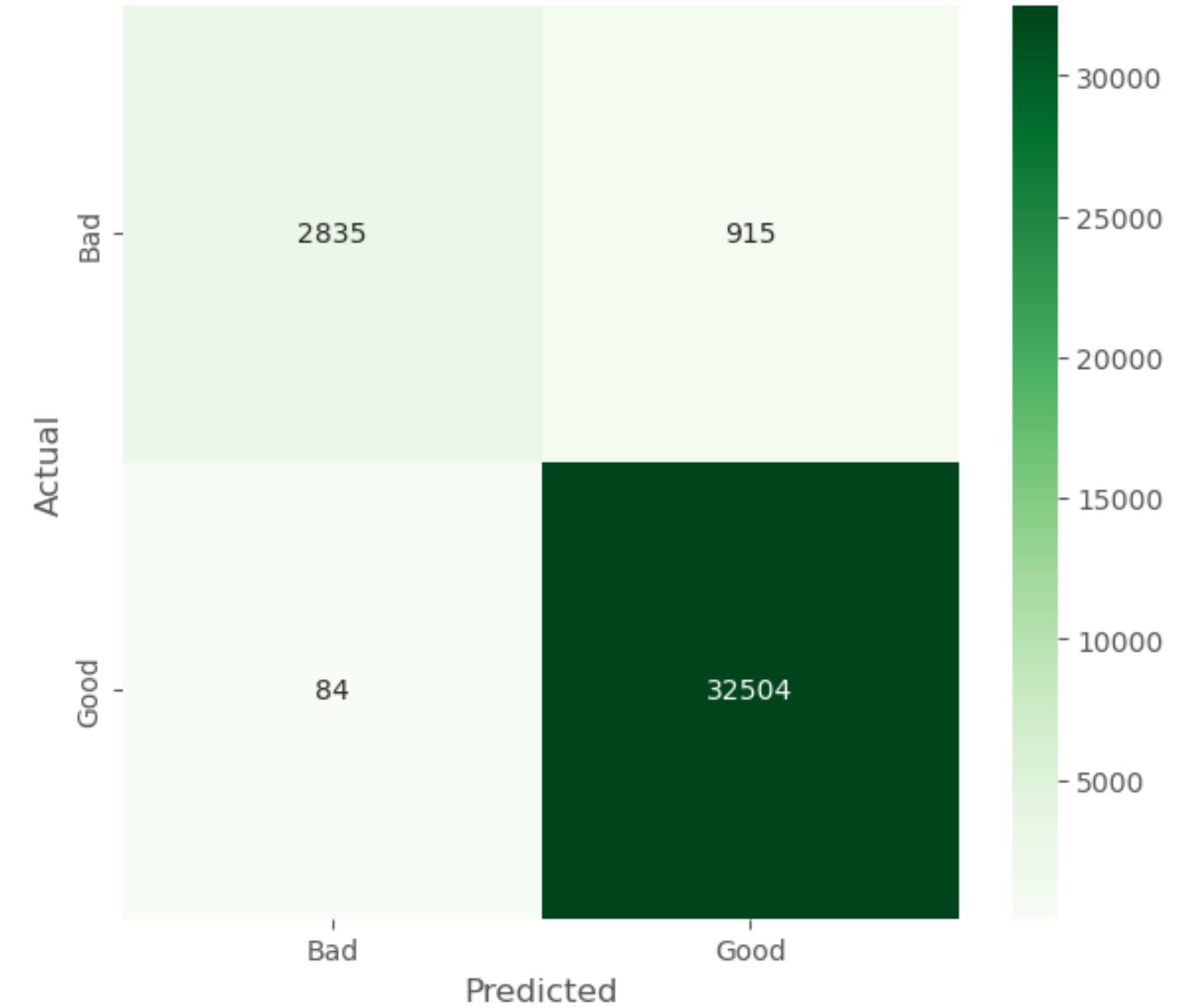
Fitur paling berpengaruh dalam Random Forest adalah **total_pymnt, mths_since_last_credit_pull_d** dan **out_prncp**.

PERBANDINGAN KINERJA MODEL

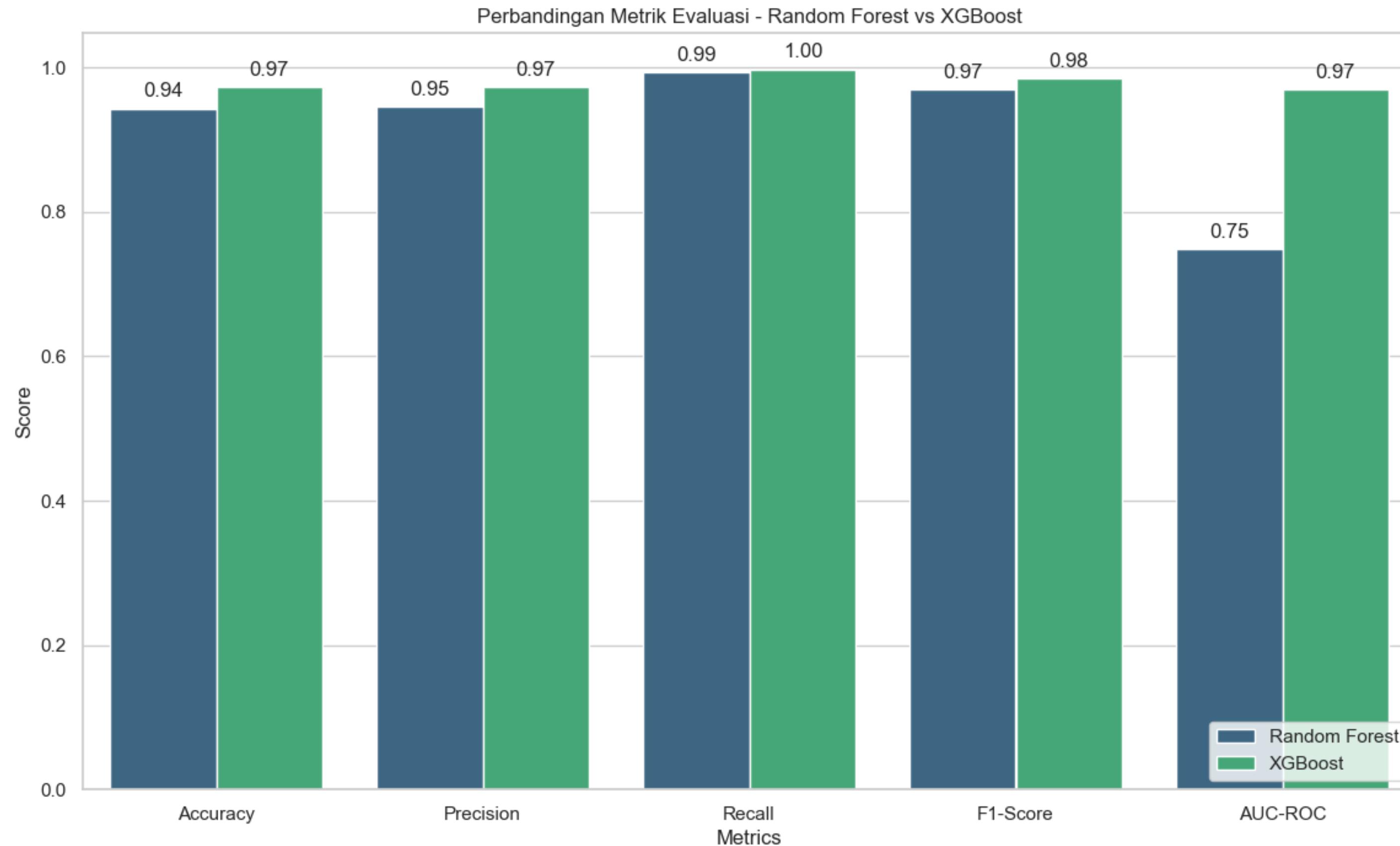
Confusion Matrix - Random Forest



Confusion Matrix - XGBoost



PERBANDINGAN KINERJA MODEL



PERBANDINGAN GOOD_BAD_LOAN DATA FINAL VS DATA TEST

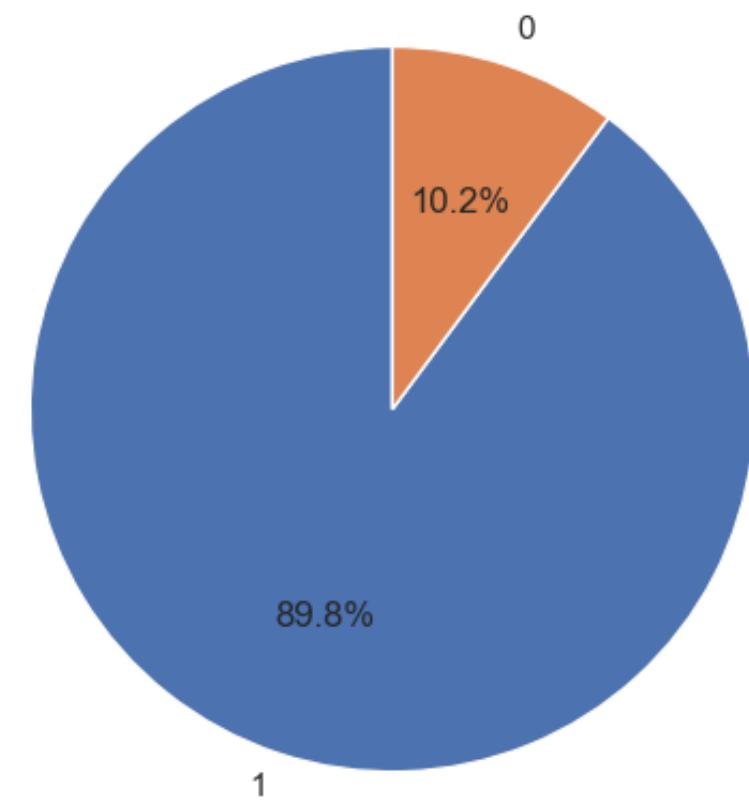
```
Data Final Good Bad Loan Distribution:  
1    163226  
0    18460  
Name: good_bad_loan, dtype: int64
```

Komposisi good_bad_loan pada Data Final

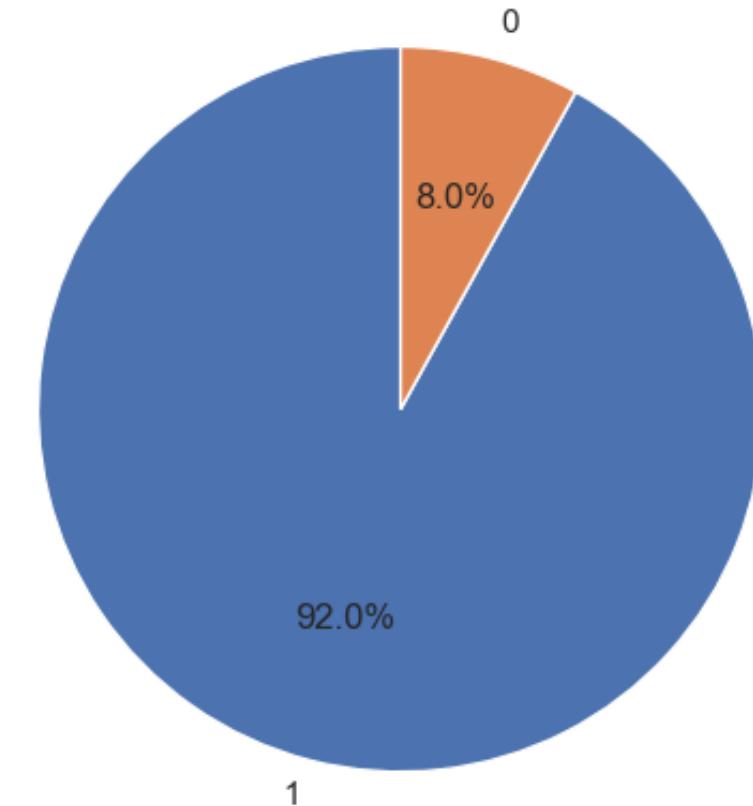
```
Test Data Good Bad Loan Distribution:  
1    33419  
0    2919  
Name: good_bad_loan, dtype: int64
```

Komposisi good_bad_loan pada Data test

Final Data Good Bad Loan Distribution



Test Data Good Bad Loan Distribution



Conclusion

RECCOMENDATION - SUGGESTION

Dengan Machine Learning dari data behavior, dapat direkomendasikan kepada perusahaan untuk mempertimbangkan data sebagai berikut :

term

36 months

home_ownership

rent

Verifikation_status

Verified &
Source Verified

Bussines Metrics

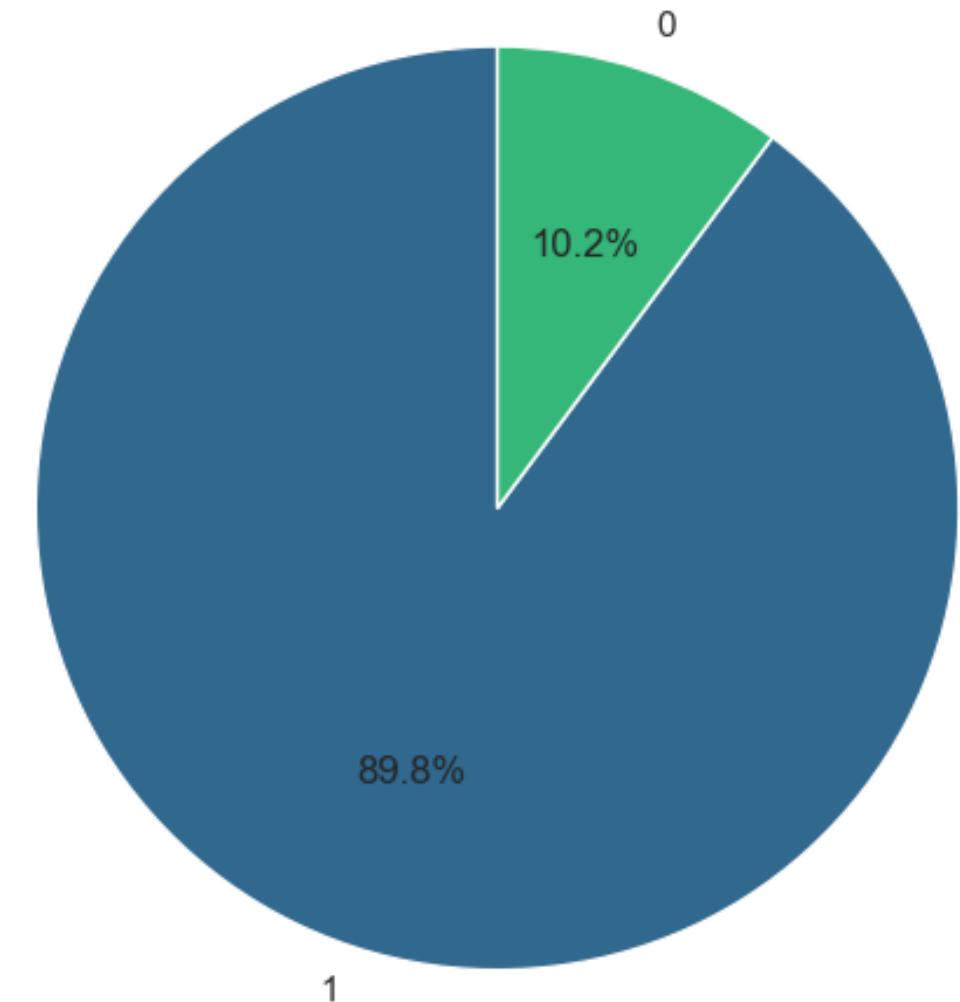
Default rate

Default rate:

Terdapat 89.8% dari total
466.285 nasabah

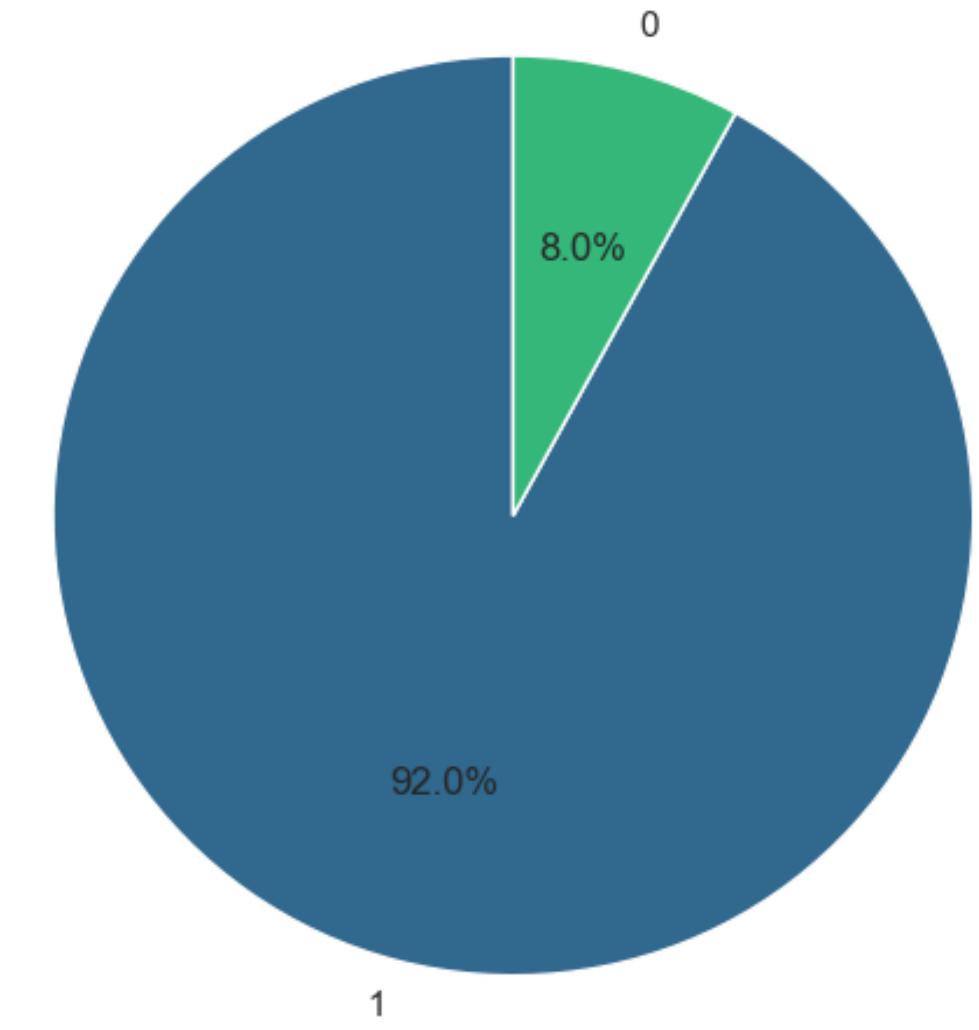
Sementara pada data test
(Test Data), persentase
nasabah yang dapat
disetujui adalah 92% dari
36.338 calon nasabah

Final Data Good Bad Loan Distribution



Test Data

Test Data Good Bad Loan Distribution



Final Data

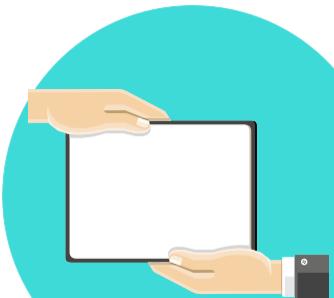
WORKFORCE TIME

MANUAL

Penyerahan
berkas



Pendaftaran dan
pemberkasan



Verifikasi data

Analisa
kelayakan

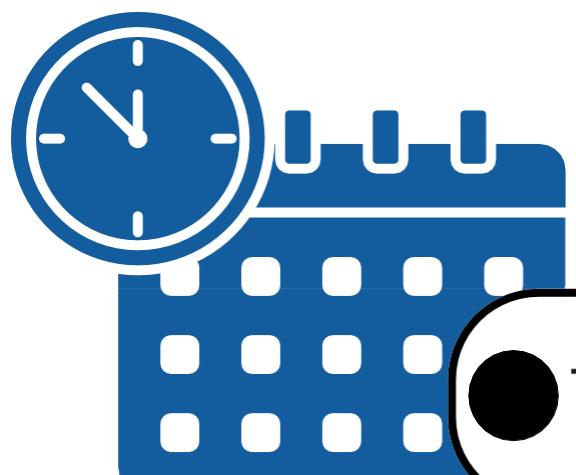


Analisa agunan

Persetujuan
kredit



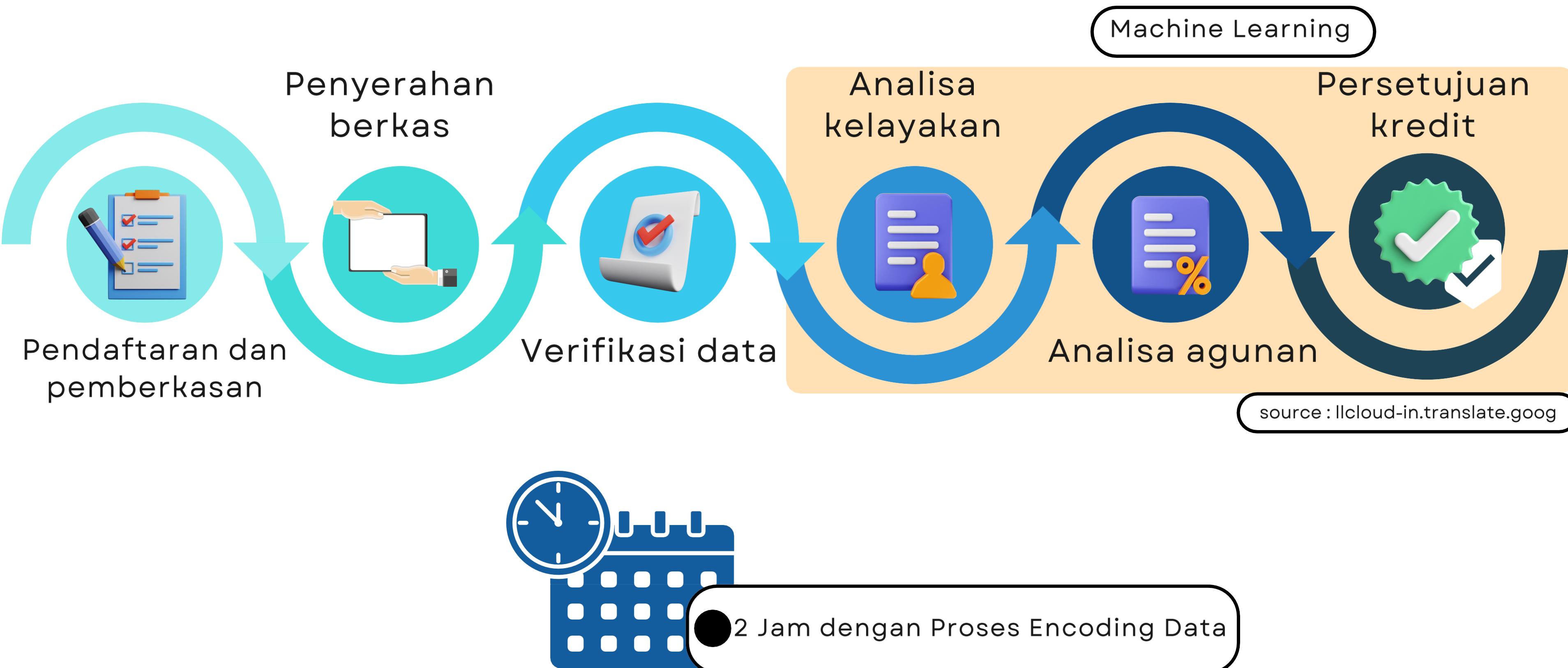
source : llcloud-in.translate.goog



1 Minggu atau beberapa bulan

WORKFORCE TIME

MANUAL & MACHINE LEARNING



WORKFORCE COST

cost

Rp4 juta - Rp8 juta

Penyerahan
berkas

Rp3,5 juta - Rp8 juta

Analisa
kelayakan

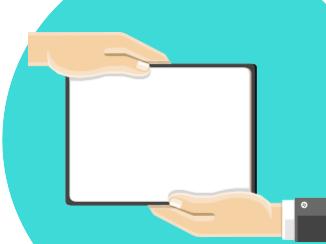
Rp4.5 juta - Rp8 juta

Persetujuan
kredit



Pendaftaran dan
pemberkasan

Rp3,5 juta - Rp8 juta



Verifikasi data

Rp4 juta - Rp8 juta



Analisa argumen

Rp 8 juta - Rp 14juta

ML Process = 16 JUTA
Data Scientist= 9 - 10 JUTA

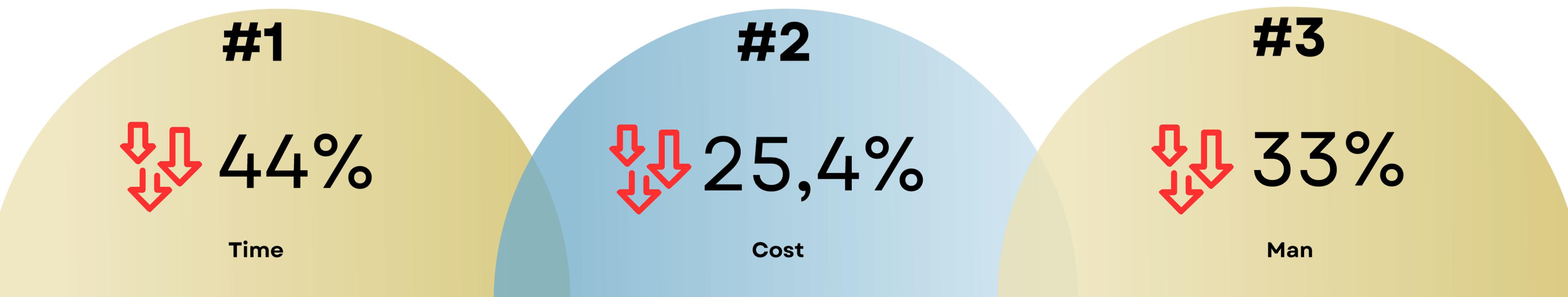
RP 20,5 -21,5
JUTA

RP 27,5JUTA →

<https://www.detik.com/>

The Efficiency

- **TIME** : 6 hari kerja menjadi 3 hari dan 3 jam kerja
- **COST** : 27,5 Juta menjadi 20,5-21,5 Juta
- **MAN** : 1 orang per proses (6 orang) menjadi 3 orang dan 1 Data Scientist (4 orang)



Thank You