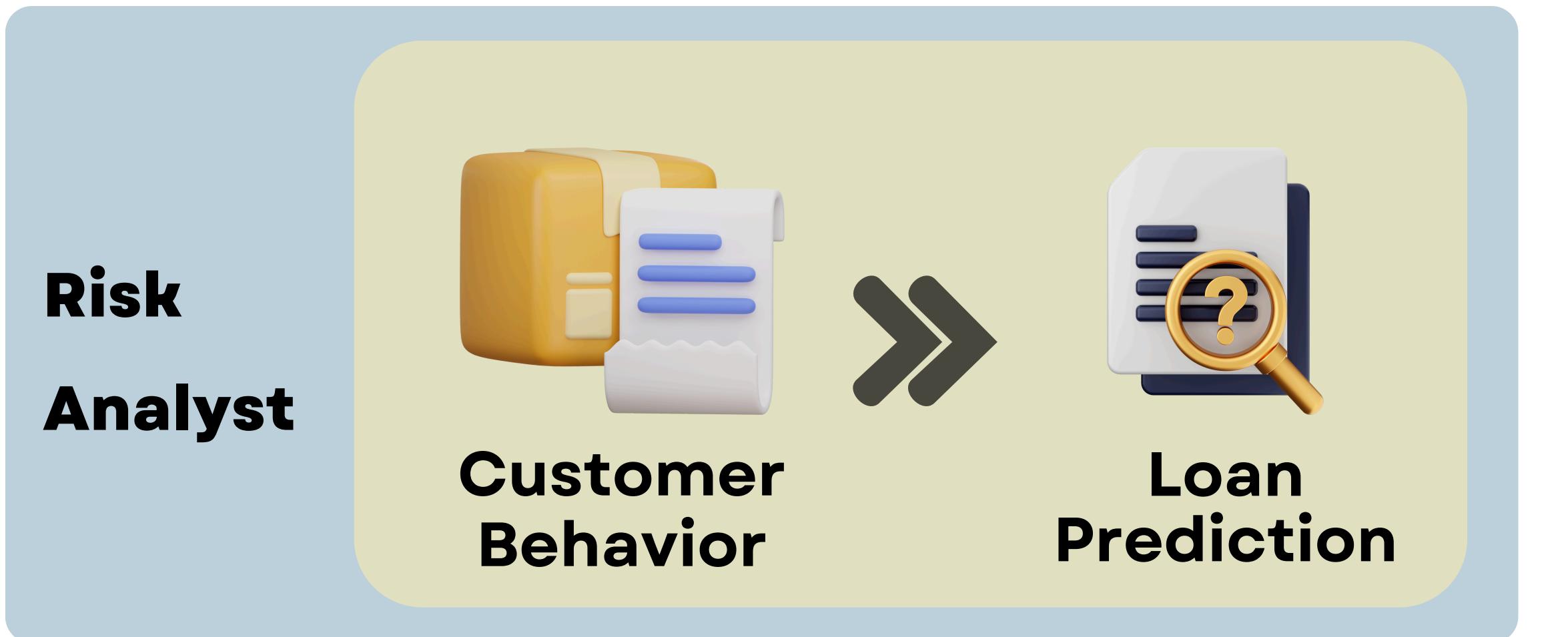


# Credit Loan Prediction Based on Customer Behavior

Project Based Internship ID/X Partners

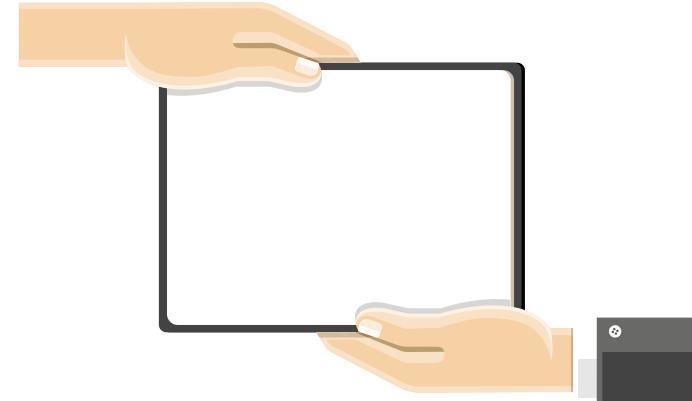
# Project Overview



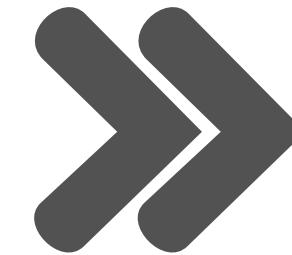
# Business Process



Pendaftaran dan  
Pemberkasan



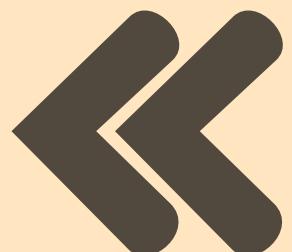
Penyerahan Berkas



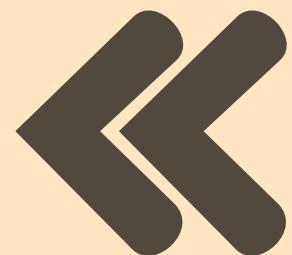
Verifikasi Data



Persetujuan  
Kredit

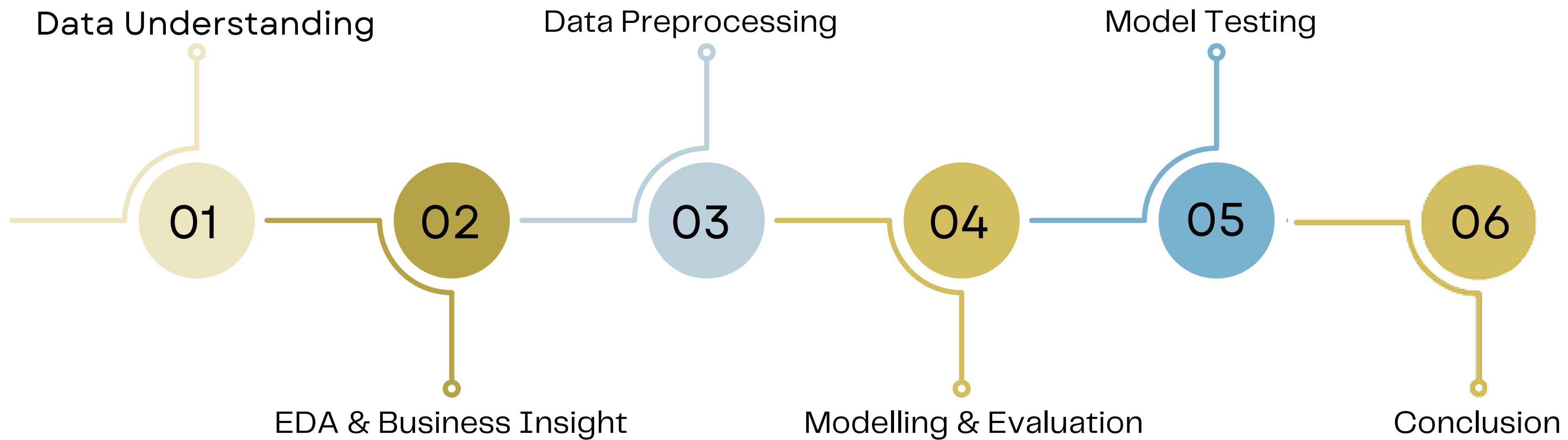


Analisa Agunan



Analisa Kelayakan  
Kredit

# Workflow

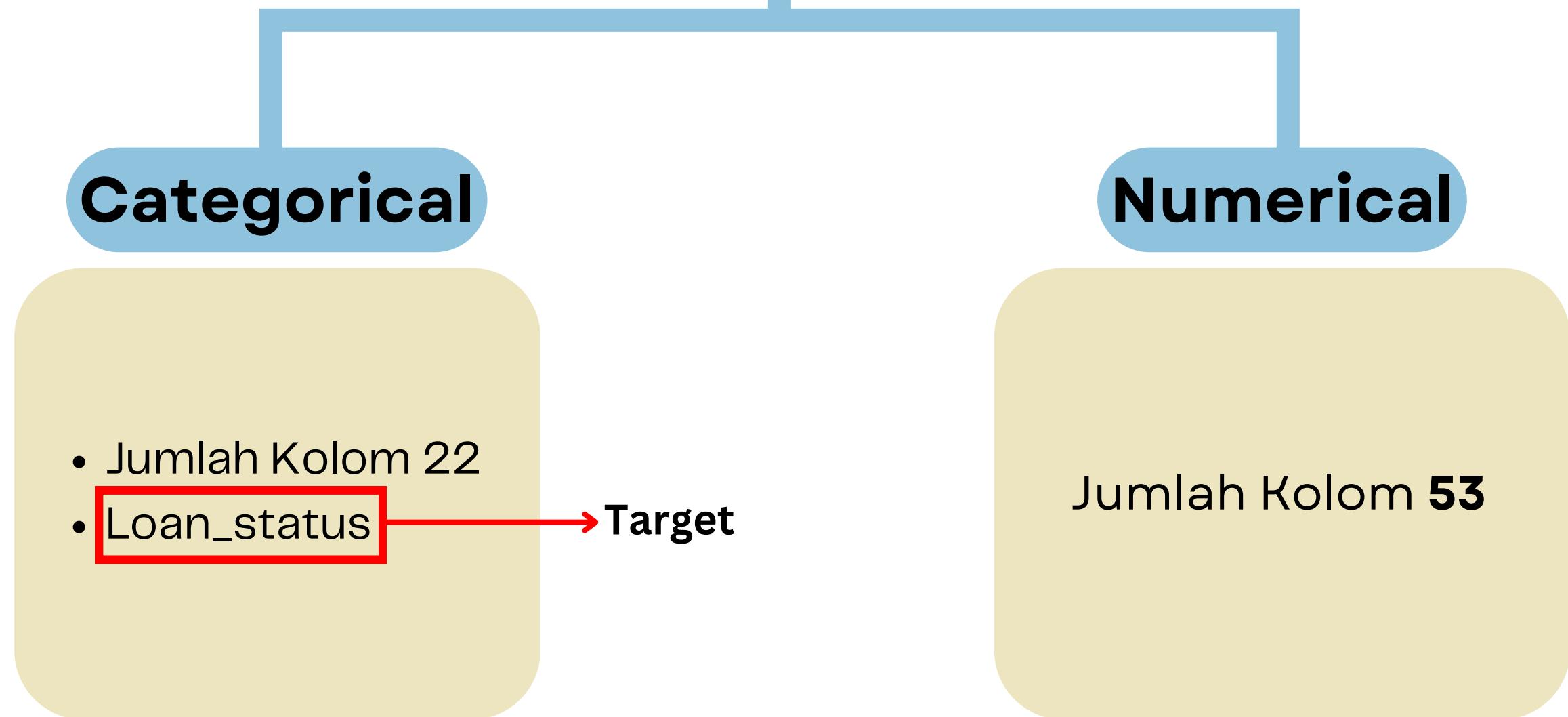
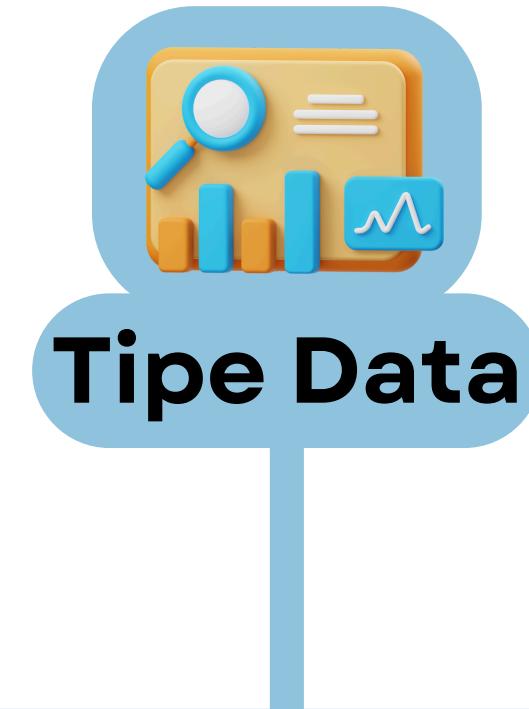


# **Data Understanding**

# Dataset



Kolom → 75  
Baris → 466.285



# Problem



Proses penilaian kredit manual memakan waktu, tidak konsisten, dan rawan kesalahan, yang dapat menyebabkan pemberian kredit yang tidak optimal dan meningkatkan risiko kredit macet.

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



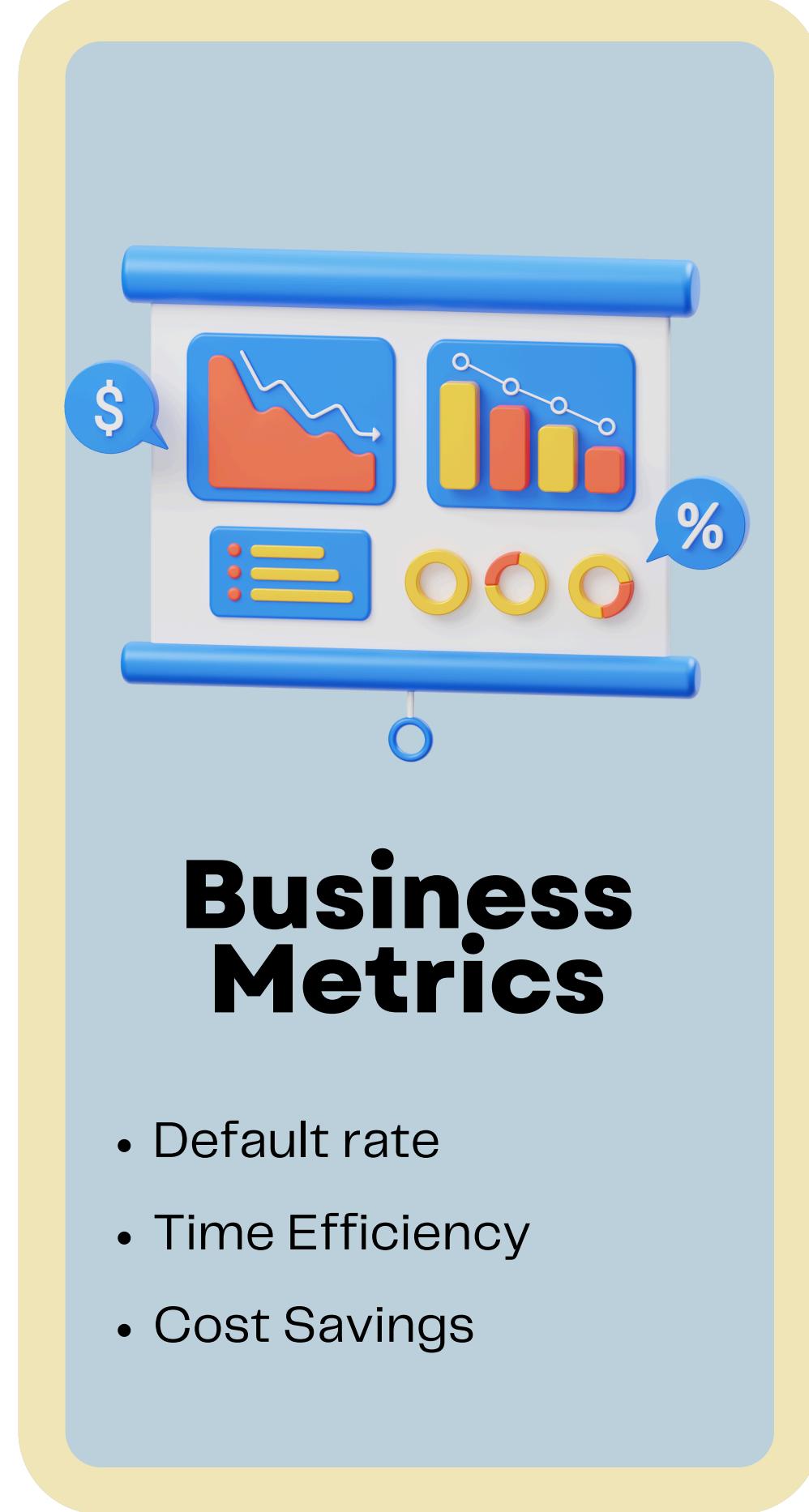
## Goals

Pengurangan biaya operasional, dan waktu proses Credit Approval melalui analisis customer behavior



## Objective

Memberikan rekomendasi credit loan approval berdasarkan customer behavior



## Business Metrics

- Default rate
- Time Efficiency
- Cost Savings

# **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
- Fully Paid
- In Grace Period
- Does not meet the credit policy. Status: Fully Paid

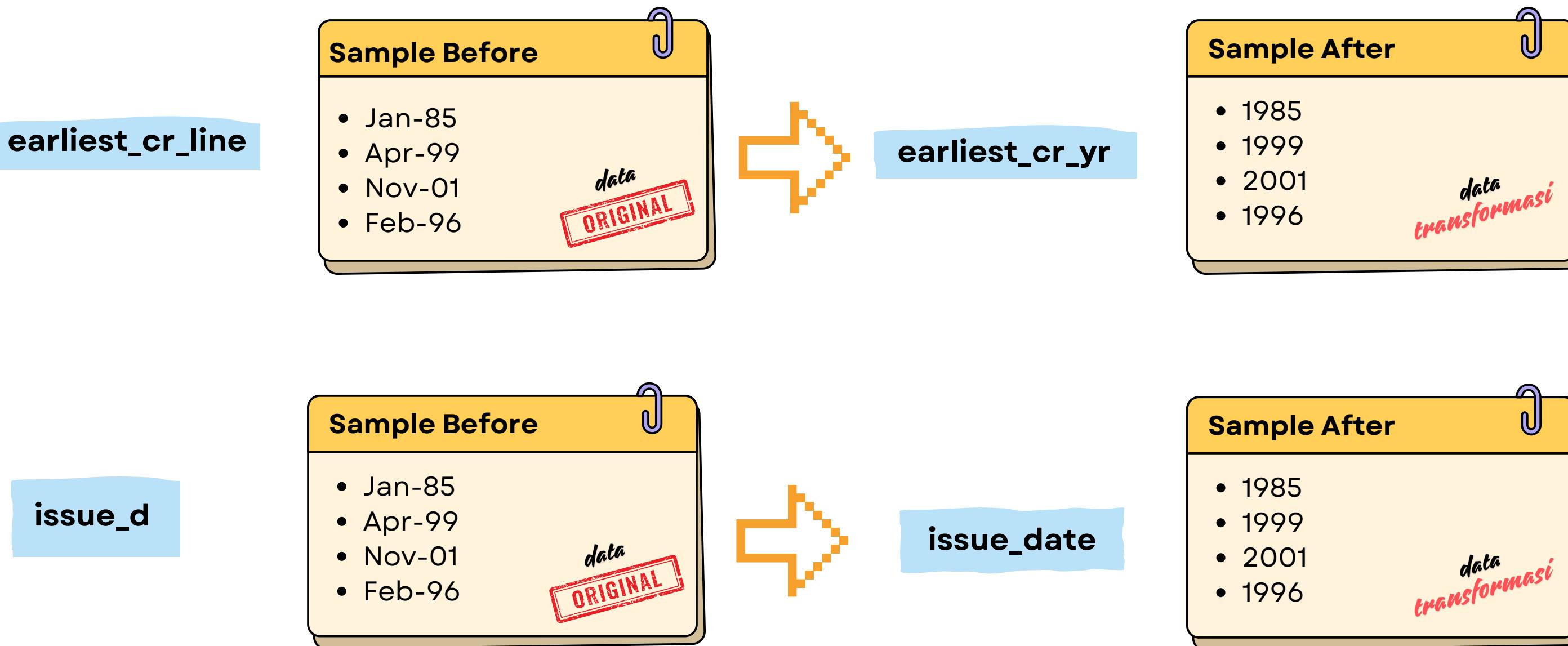
### Bad Loans

- Charged Off
- Late (16-30 days)
- Late (31-120 days)
- Default
- Does not meet the credit policy. Status: Charged Off

#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

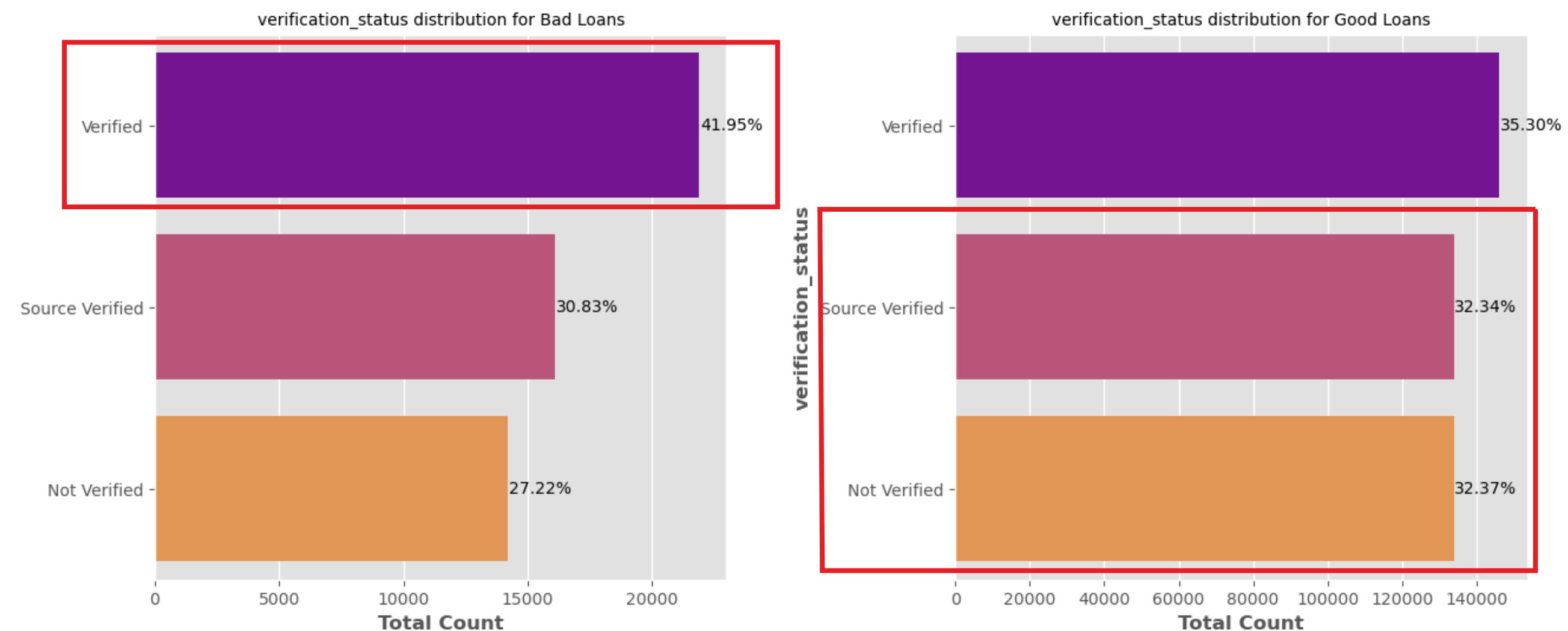
### Insight:

Sebagian besar (41,95%) bad loans adalah pinjaman yang terverifikasi, yang lebih tinggi dibandingkan dengan 35,30% dalam good loans. Ini bisa menunjukkan bahwa proses verifikasi saja bukanlah indikator kuat dari kualitas pinjaman.

### Potensial Masalah



Verifikasi  
pinjaman yang  
tidak efektif



\* Memisahkan Good Loans atau Bad Loans dan membuat persentase masing-masing

# #2

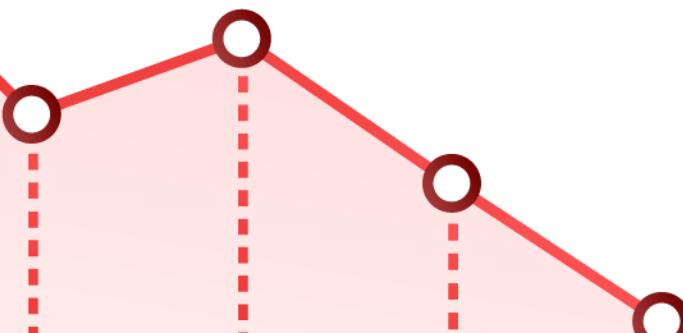
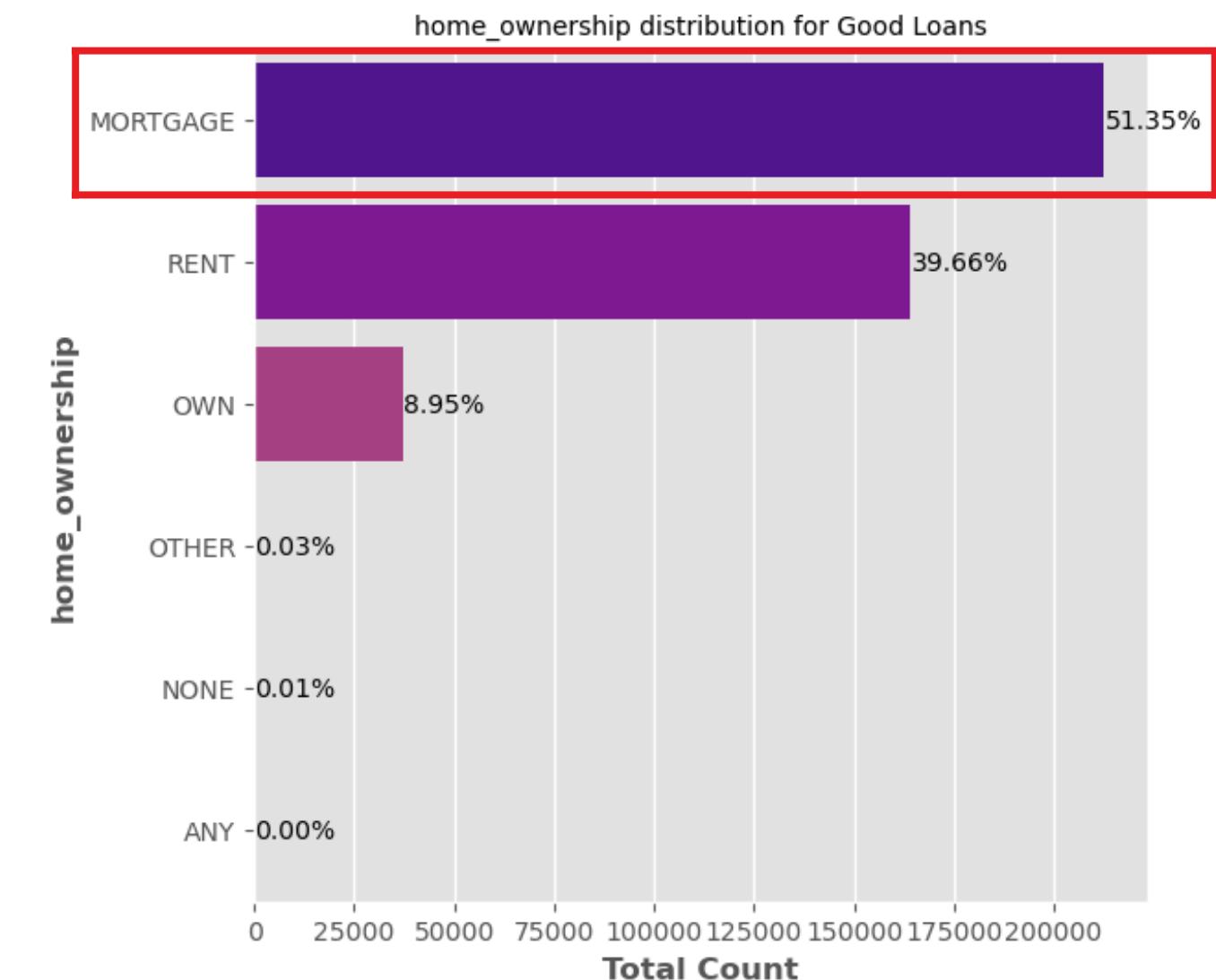
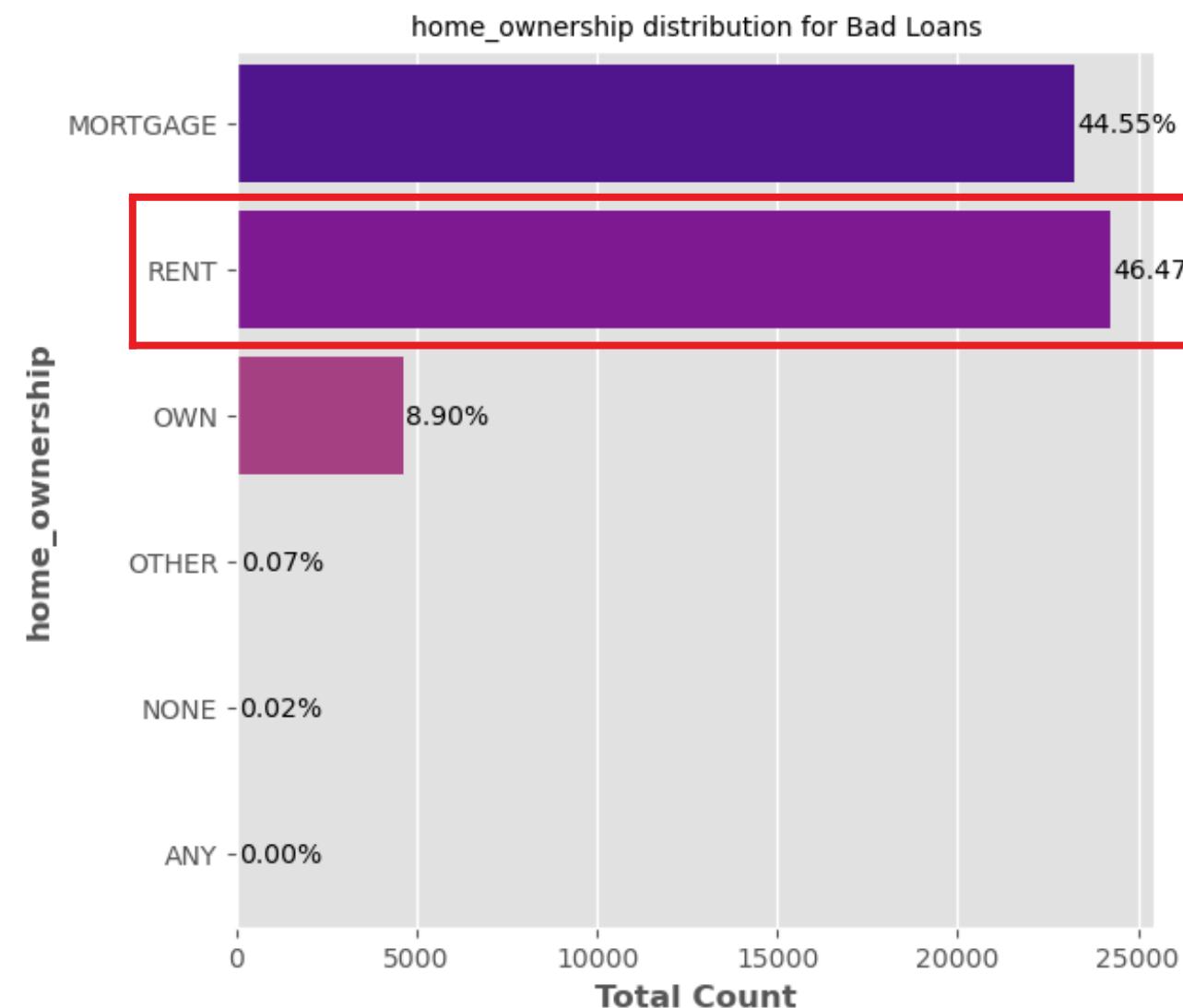
## DISTRIBUSI GOOD\_BAD\_LOAN BERDASARKAN HOME\_OWNERSHIP

### Insight

Peminjam yang menyewa rumah memiliki risiko gagal bayar yang lebih tinggi dibandingkan dengan peminjam yang memiliki hipotek atau kepemilikan rumah.

### Potensial Masalah

Rendahnya stabilitas keuangan



#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%).

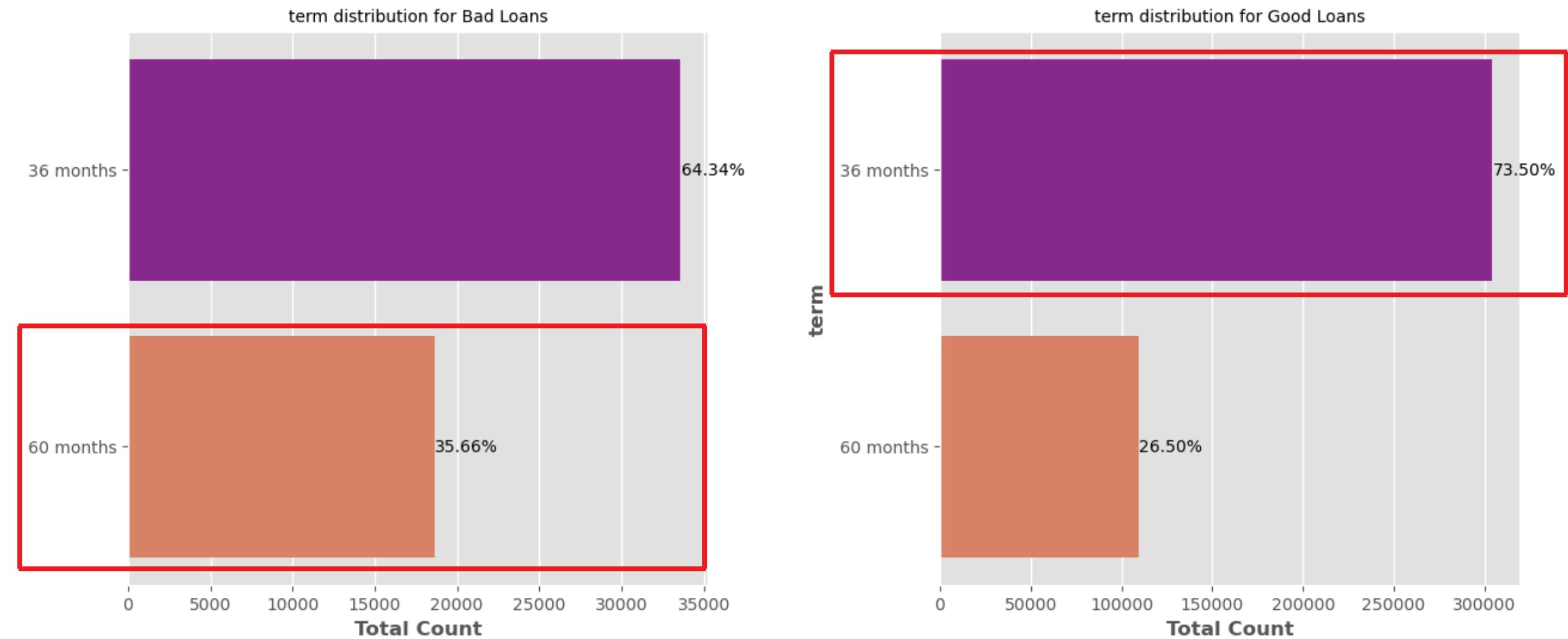
### Potensial Masalah



Suku Bunga  
dan  
Profitabilitas



Resiko jangka  
panjang

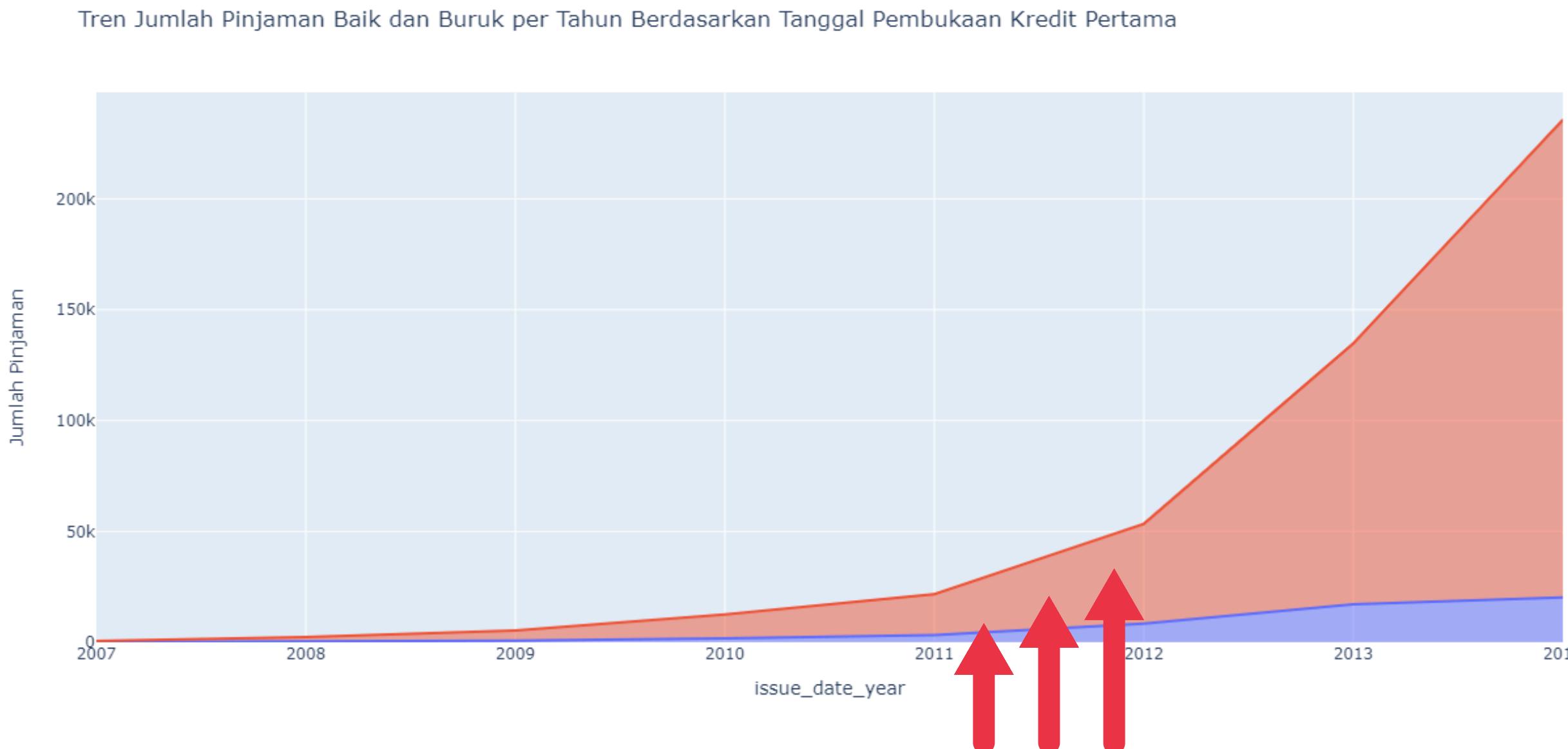


#4

## DISTRIBUSI GOOD\_BAD\_LOAN BERDASARKAN ISSUE\_DATE\_YEAR

### Insight

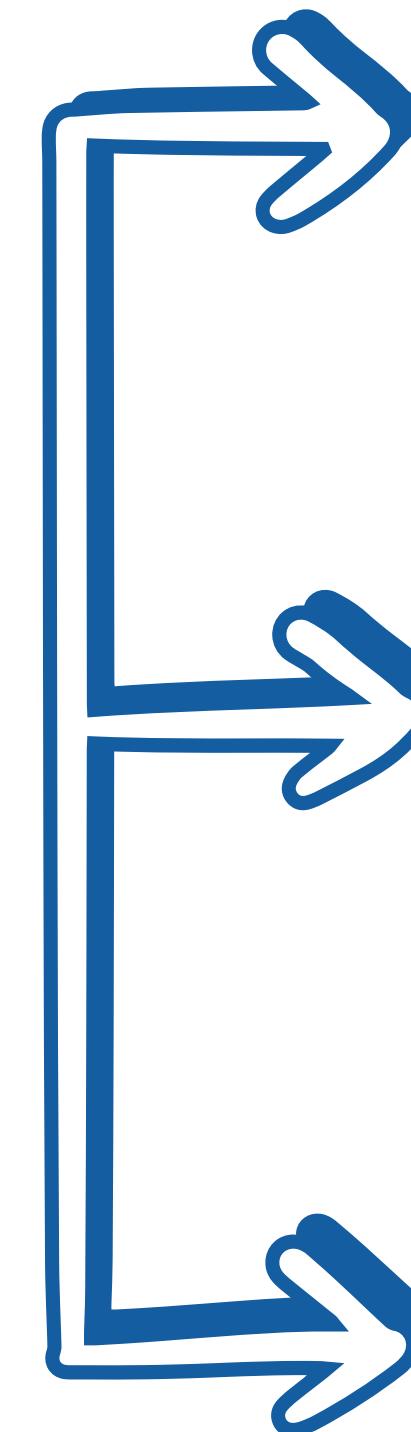
Berdasarkan Berdasarkan Tanggal Pembukaan Kredit Pertama, sejak tahun 2011 jumlah pinjaman mengalami kenaikan lebih dari dua kali lipat setiap tahunnya dan angka bad loan-pun masih terkendali.



# **Data Preprocessing**

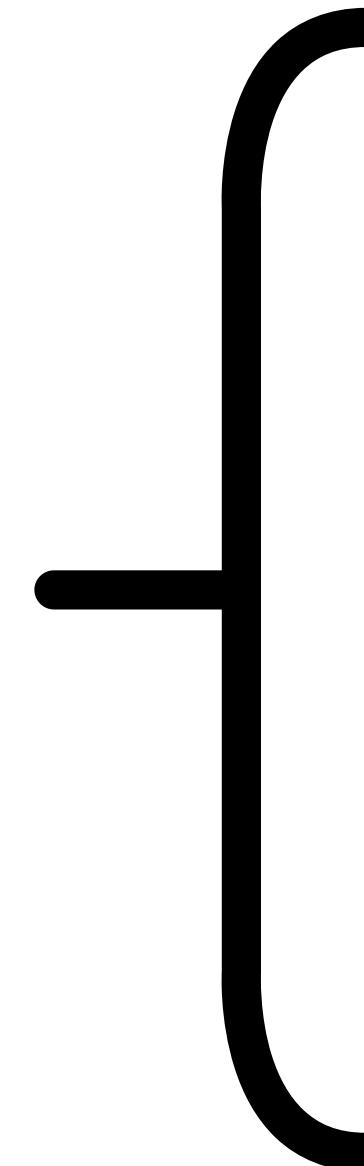
# Data PreProcessing

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  
TRANSFORMASI



- Melakukan normalisasi kolom yang memiliki distribusi skew > 0.5 atau > -0.5
- Melakukan Standarisasi kolom yang memiliki distribusi normal antara 0.5 - (-0.5)

Melakukan One-Hot Encoding untuk kolom-kolom kategorikal

# **Machine Learning Model**

# MODELLING

Process of Choosing the Algorithm

## SPLITTING

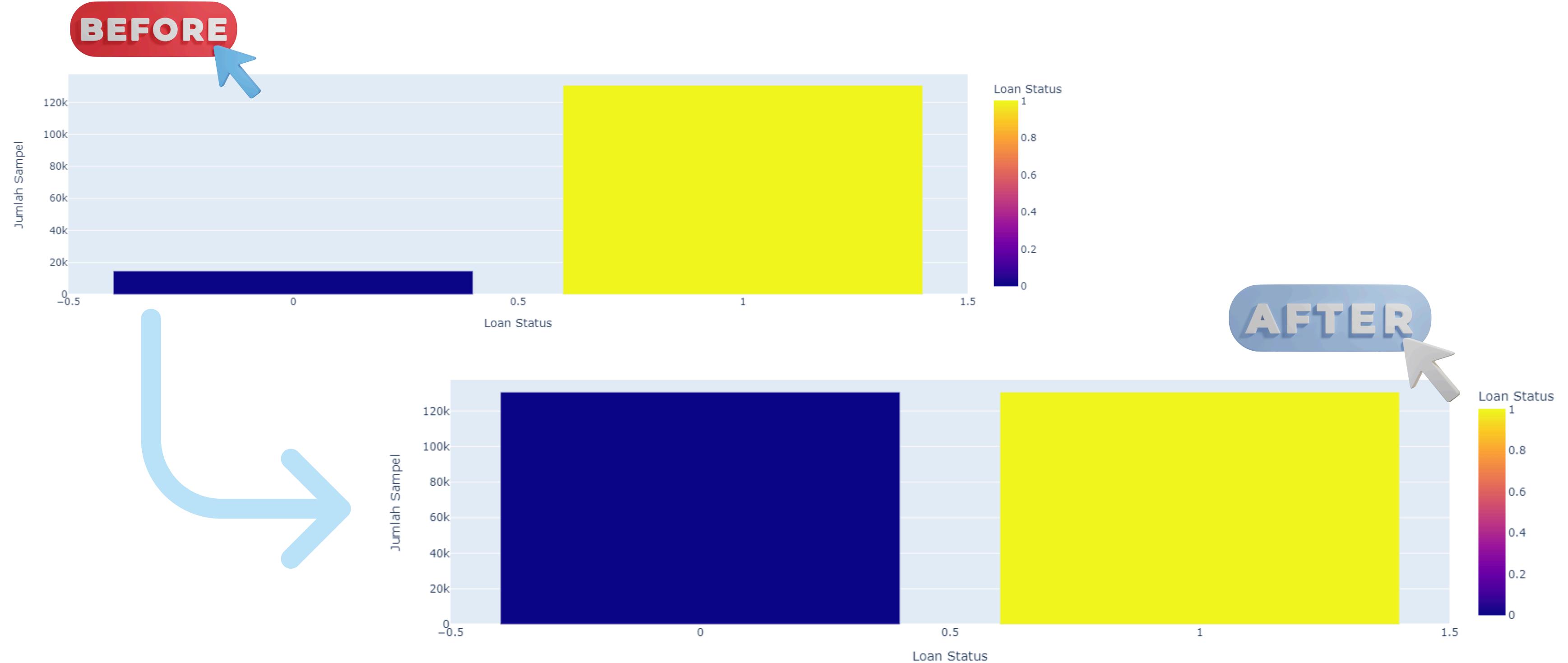


## ALGORITHM SCORE

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

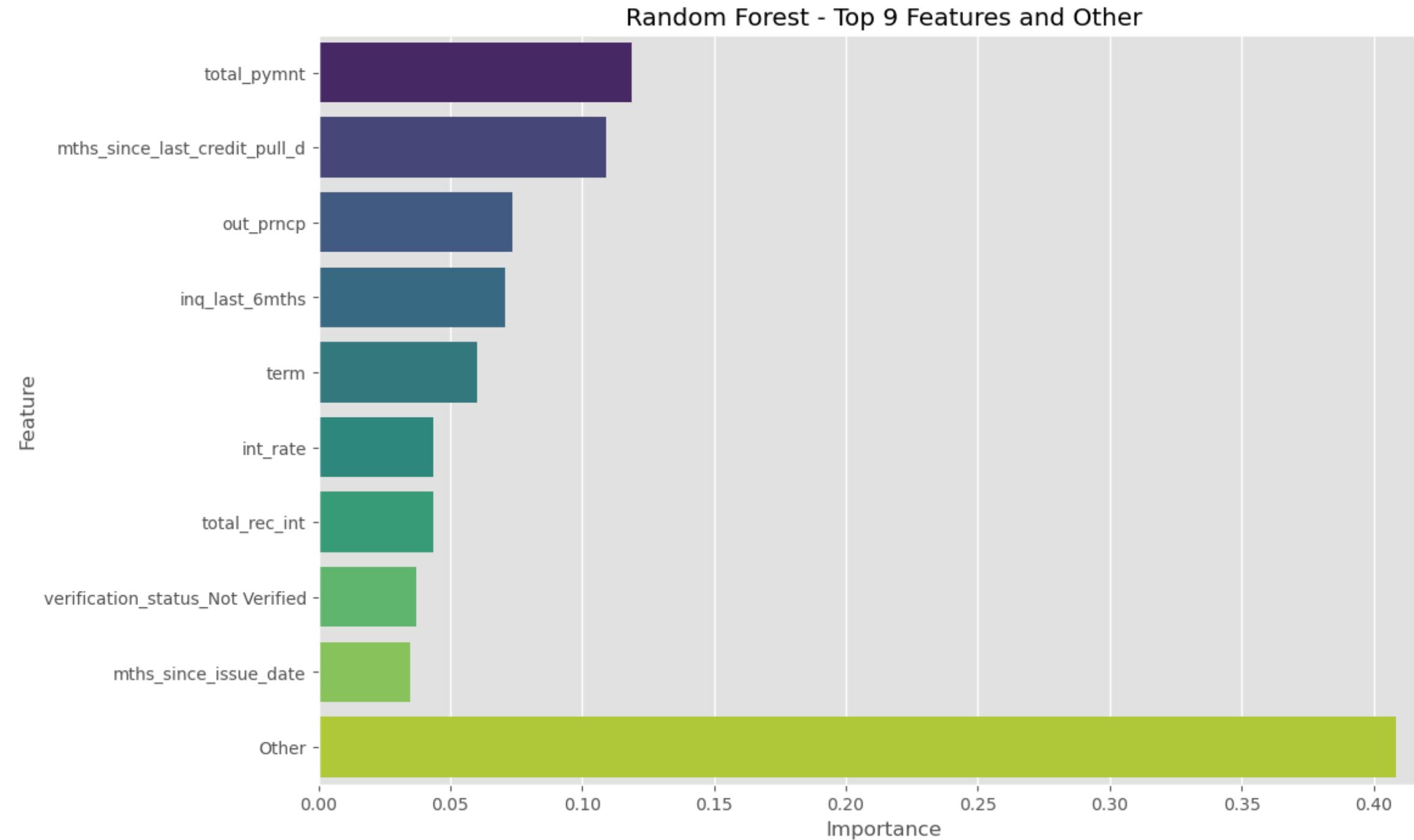
# MODELLING

Handle Class Imbalance Sampling SMOTE pada Data Train



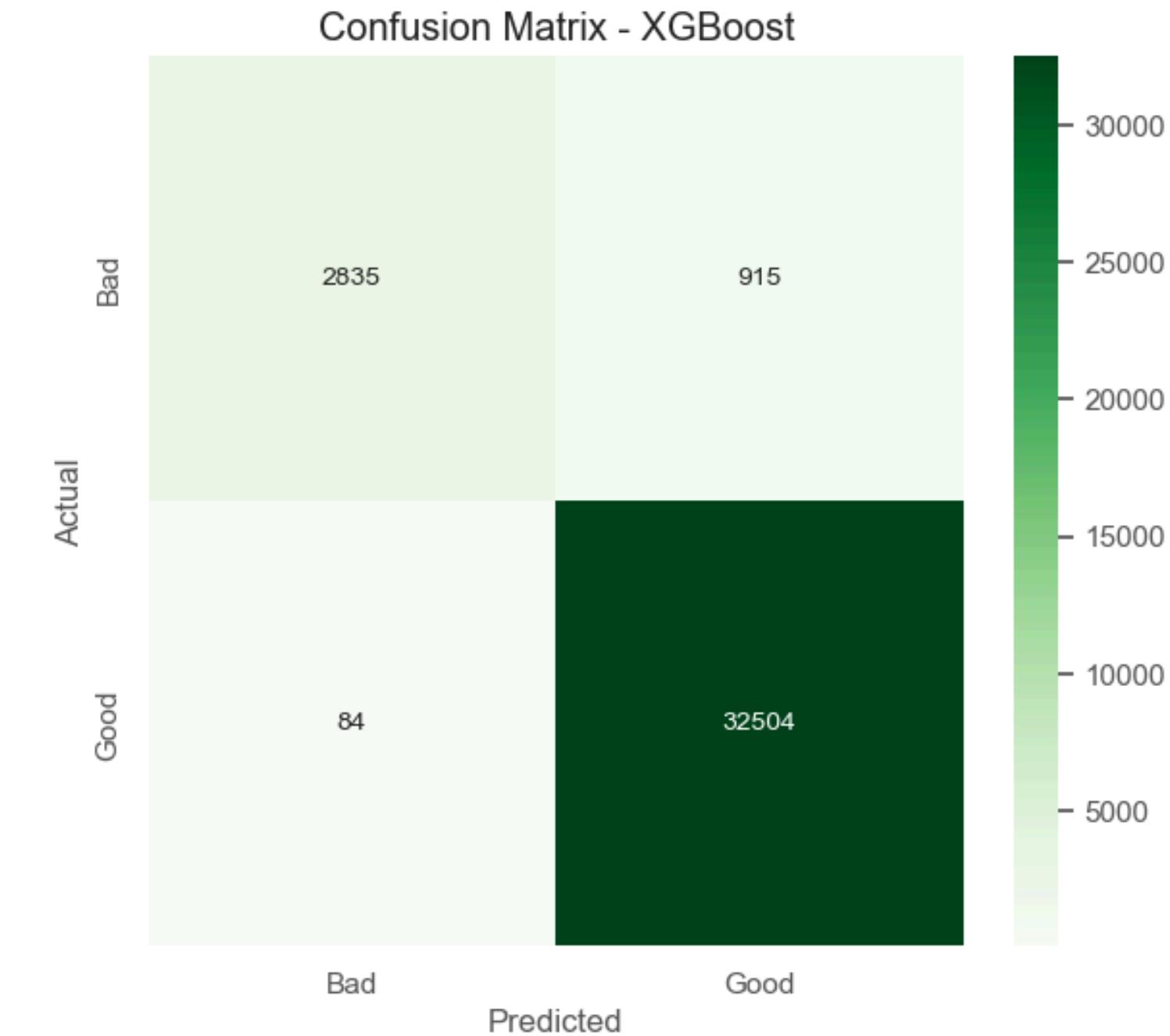
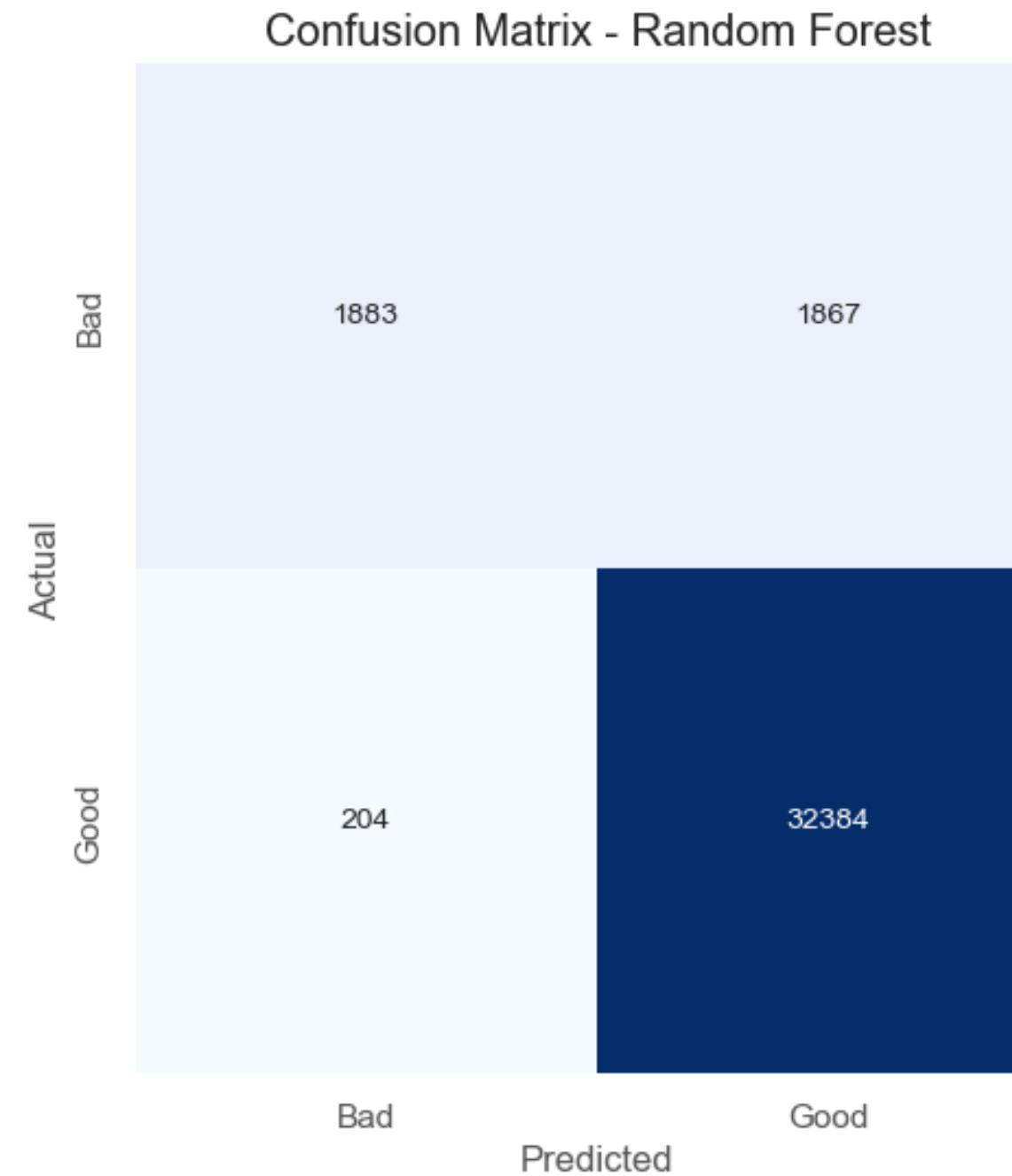
# MODELLING

## Feature Importance

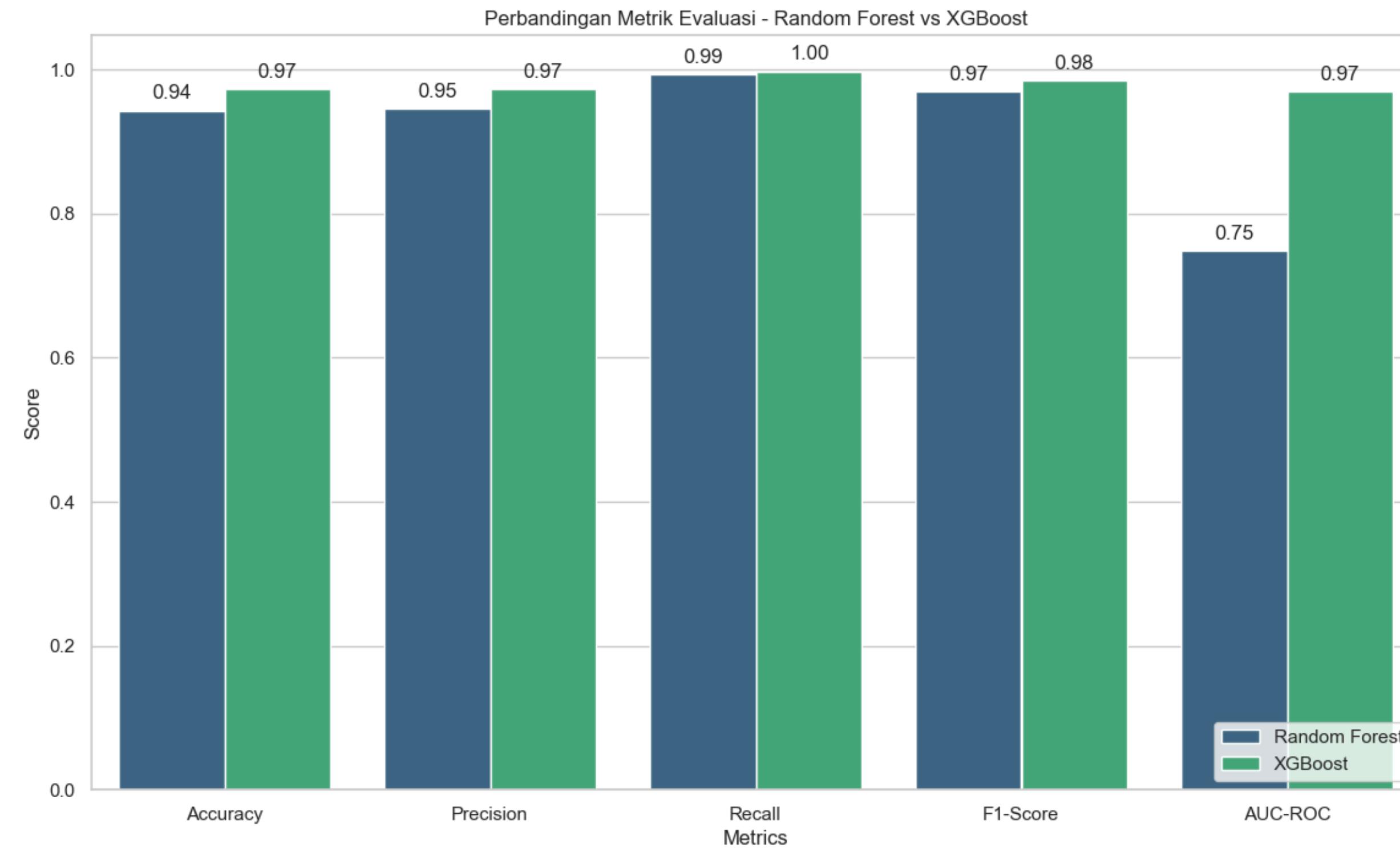


Fitur paling berpengaruh dalam Random Forest adalah **total\_pymnt**,  
**mths\_since\_last\_credit\_pull\_d** dan **out\_prncp**.

# PERBANDINGAN KINERJA MODEL



# PERBANDINGAN KINERJA MODEL



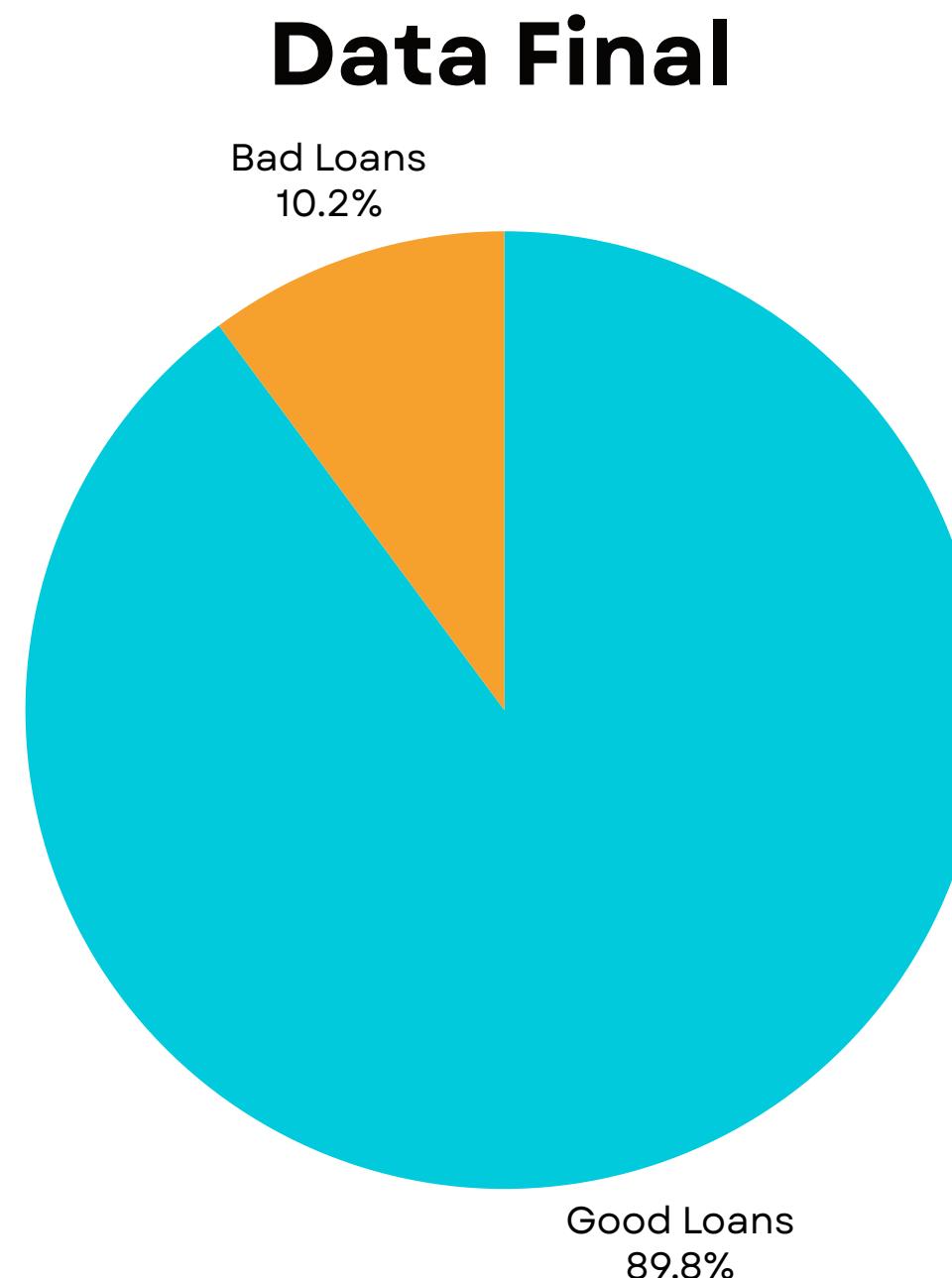
# PERBANDINGAN GOOD\_BAD\_LOAN DATA FINAL VS DATA TEST

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

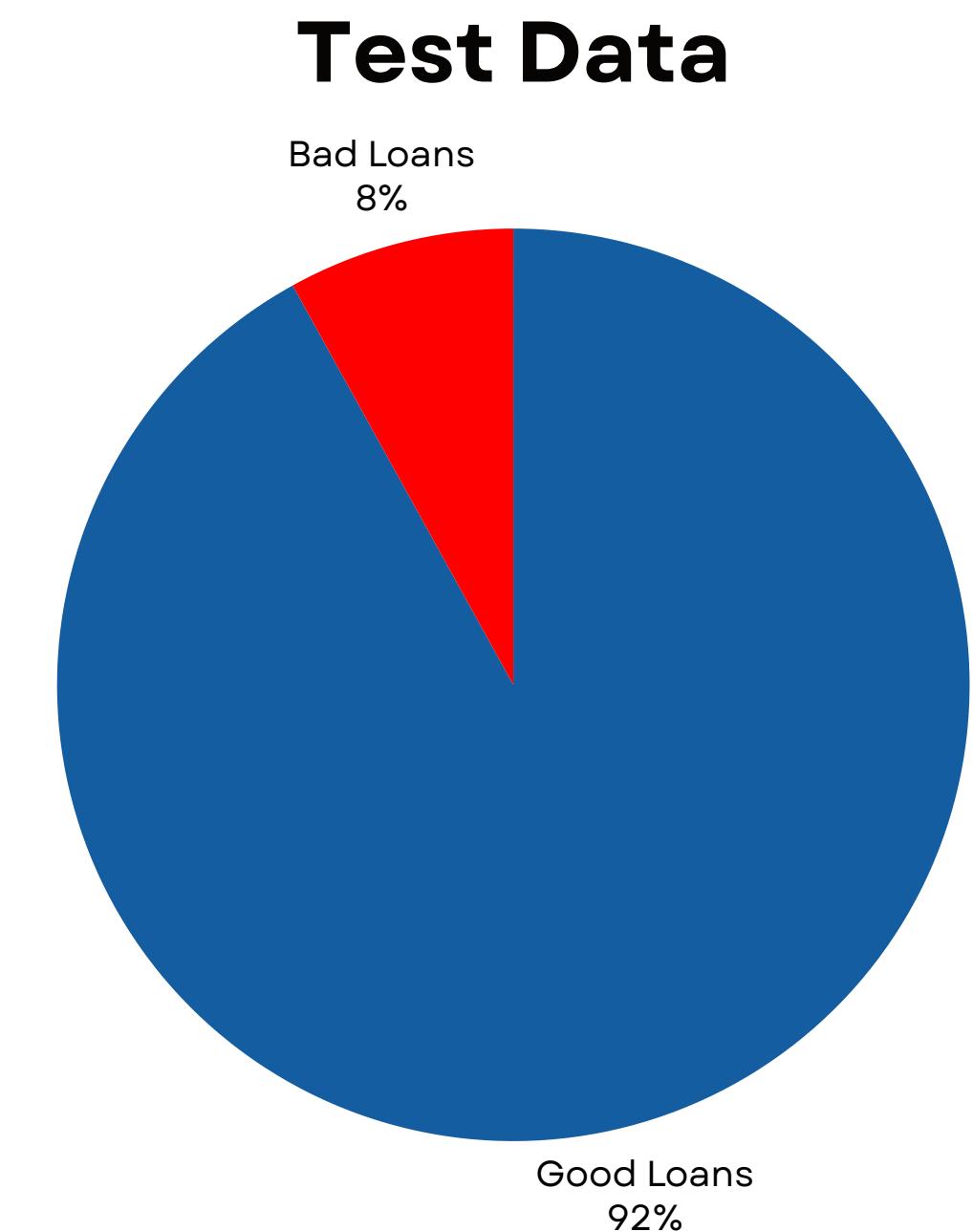
Komposisi good\_bad\_loan pada Data Final

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

Komposisi good\_bad\_loan pada Data test



Total data final: 181.686 row



Total Test Data: 36.338 row

# **Conclusion**

# **RECCOMENDATION - SUGGESTION**

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

**term**

**home\_ownership**

**Verifikation\_status**

**36 months**

**rent**

**Verified &  
Source Verified**

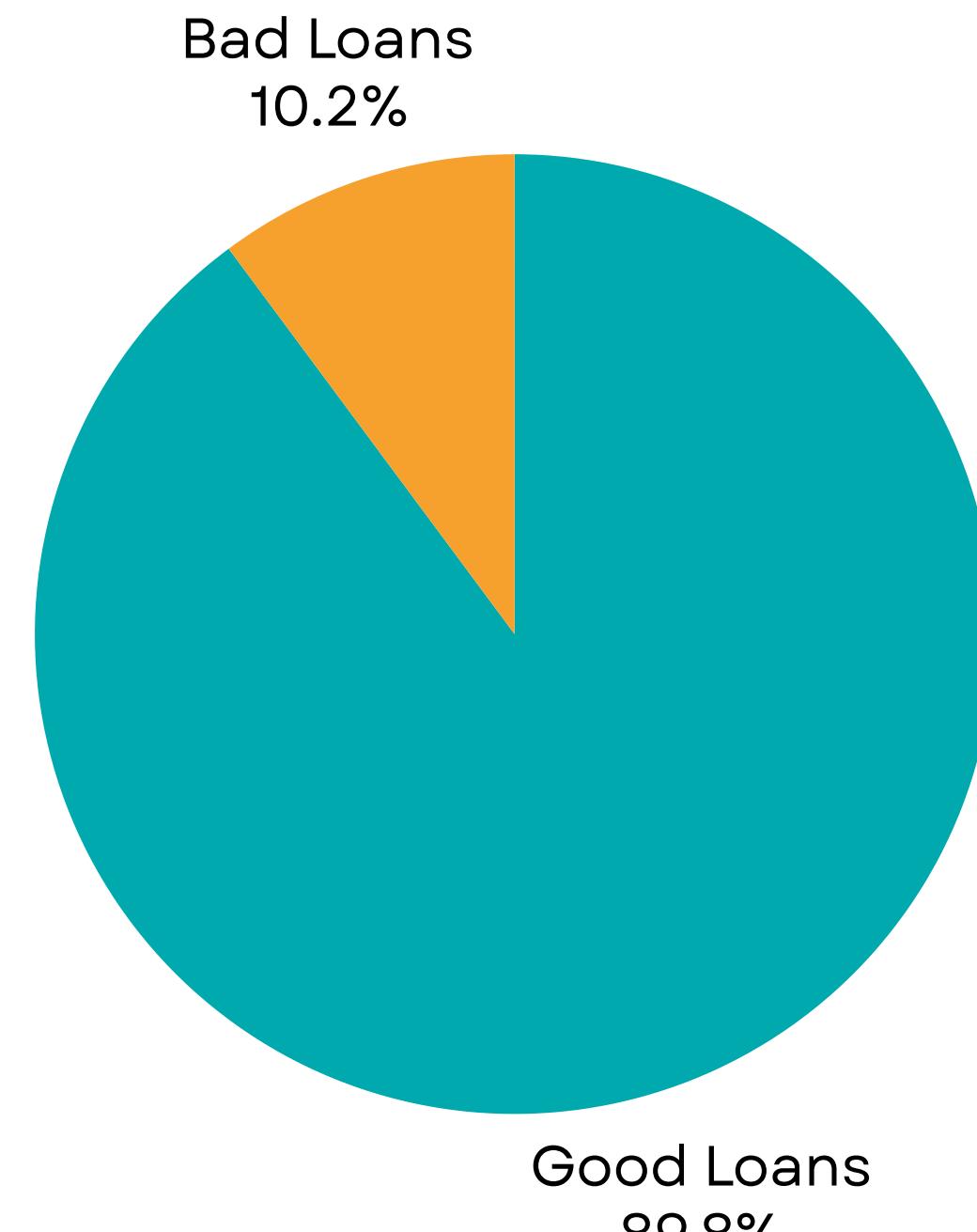
# Bussines Metrics

## Default rate

Default rate :

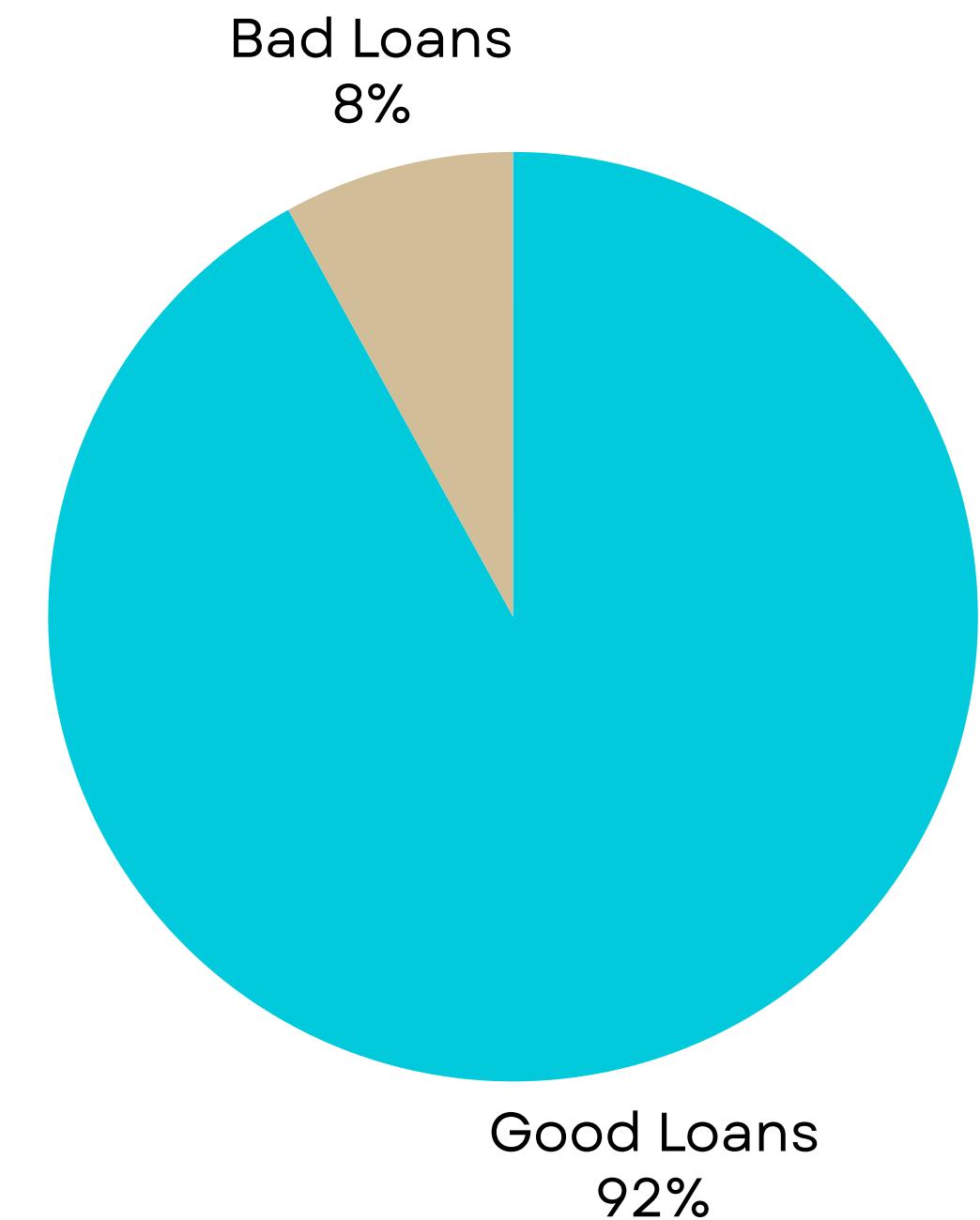
Terdapat **89.8%** dari total  
**466.285 nasabah**

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



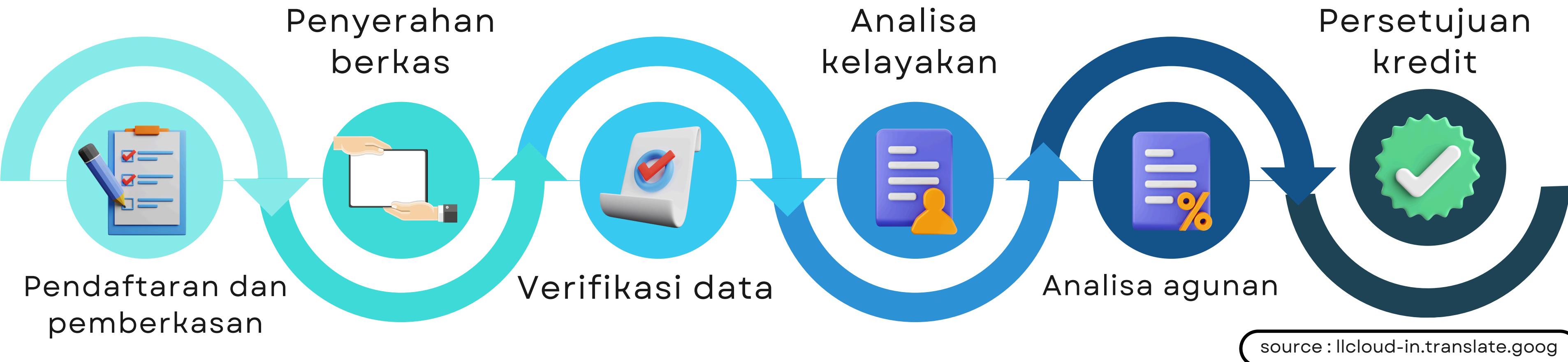
Final Data

Test Data



# WORKFORCE TIME

## MANUAL

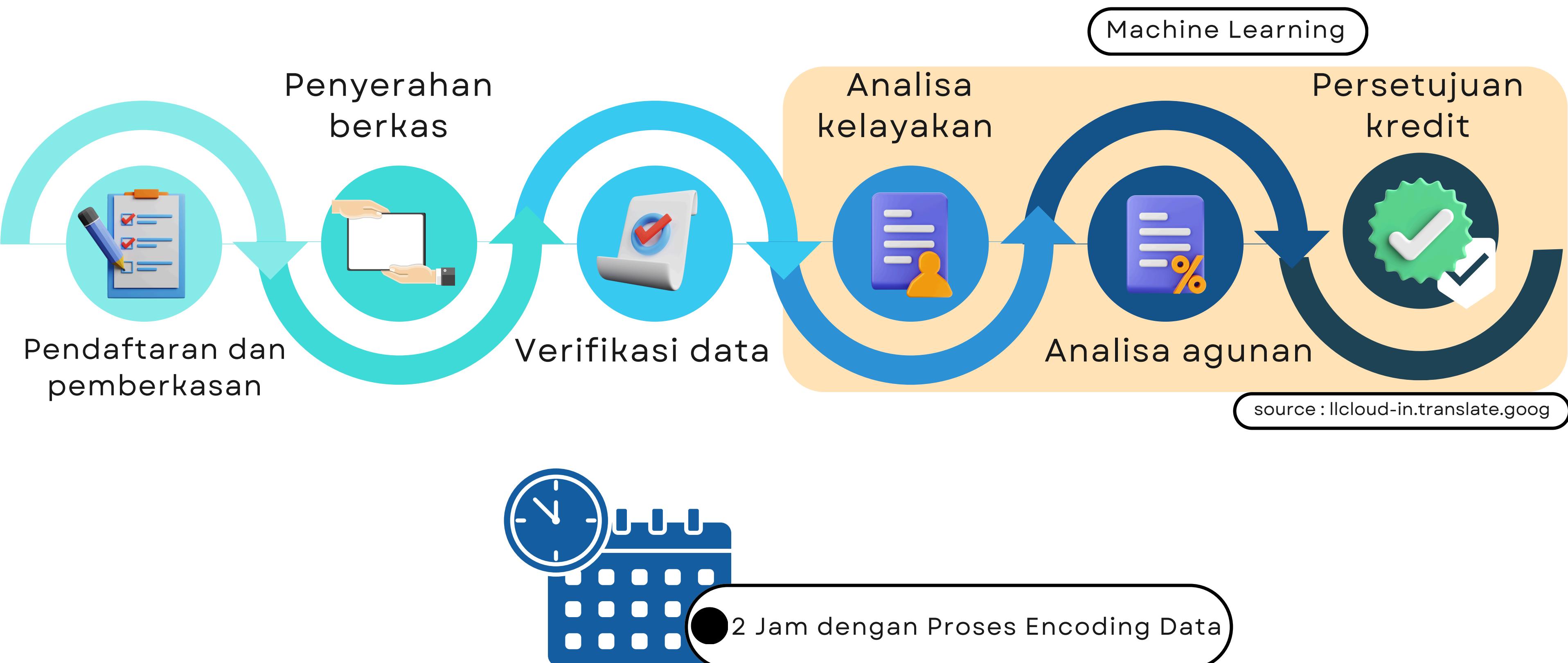


source : llcloud-in.translate.goog



# TIME EFFICIENCY

## MANUAL & MACHINE LEARNING



# COST SAVINGS

cost

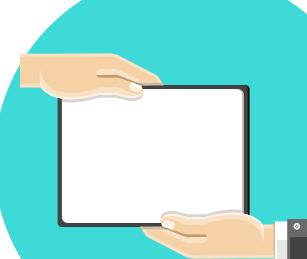
Rp4 juta - Rp8 juta

Penyerahan  
berkas



Pendaftaran dan  
pemberkasan

Rp3,5 juta - Rp8 juta



Verifikasi data

Rp4 juta - Rp8 juta

Rp3,5 juta - Rp8 juta

Analisa  
kelayakan



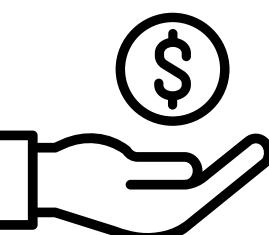
Analisa argumen

Rp 8 juta - Rp 14juta

Rp4.5 juta - Rp8 juta

Persetujuan  
kredit



 RP 27,5JUTA →

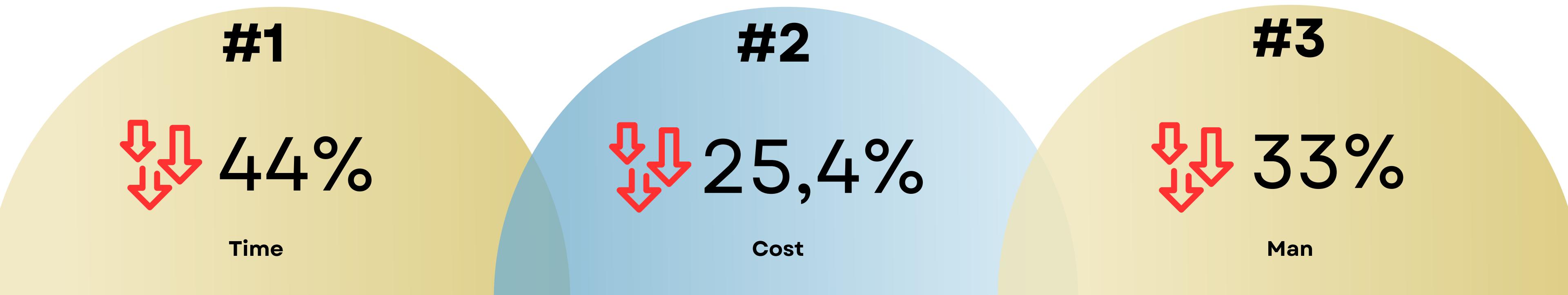
ML Process = 16 JUTA  
Data Scientist= 9 - 10 JUTA

→ RP 20,5 -21,5  
JUTA

<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**