

# Prediction Modeling

ID/X Partners - Data Scientist

Presented by  
Ramadhani Nurul Fazri



Halo, nama saya Ramadhani Nurul Fazri. Saya memiliki latar belakang pendidikan Teknik Geologi di Universitas Pertamina.

Melalui perkenalan ini, saya berharap dapat membangun jaringan yang lebih luas dan berkontribusi dalam proyek-proyek yang menarik di masa depan.



Kota Bekasi, jawa Barat



[rurul08@gmail.com](mailto:rurul08@gmail.com)



[www.linkedin.com/in/ramadhani-nurul-fazri](https://www.linkedin.com/in/ramadhani-nurul-fazri)

# TABLE OF CONTENT

Projek Portofolio

1

Exploratory Data Analysis

5

About Company

2

Data Preparation

6

Data Understanding

3

Data Modeling

7

Feature Engineering

4

Evaluation

8

Conclusion

9

# Project Portfolio

**Perusahaan multifinance perlu meningkatkan keakuratan penilaian risiko kredit untuk mengoptimalkan keputusan bisnis dan mengurangi kerugian. Kami mengembangkan model machine learning menggunakan data pinjaman dari Lending Club (2007-2014) untuk memprediksi risiko kredit, dengan fokus pada metrik bisnis seperti kerugian dan margin keuntungan bersih. Analisis data ini bertujuan untuk mengidentifikasi pola yang mengindikasikan pinjaman berpotensi buruk atau berisiko, tanpa asumsi yang kuat, untuk mendukung pengambilan keputusan investasi.**

Project explanation video [here!](#)

# 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 skoring, dan manajemen kinerja. Pengalaman gabungan kami telah melayani korporasi di seluruh wilayah Asia dan Australia serta di berbagai industri, khususnya layanan keuangan, telekomunikasi, manufaktur, dan ritel.**

**id/x partners menyediakan layanan konsultasi yang mengkhususkan diri dalam memanfaatkan solusi analitik data dan pengambilan keputusan (DAD) yang dikombinasikan dengan disiplin manajemen risiko dan pemasaran terintegrasi untuk membantu klien mengoptimalkan profitabilitas portofolio dan proses bisnis.**

**Layanan konsultasi yang komprehensif dan solusi teknologi yang ditawarkan oleh id/x partners menjadikannya sebagai penyedia layanan terpadu.**

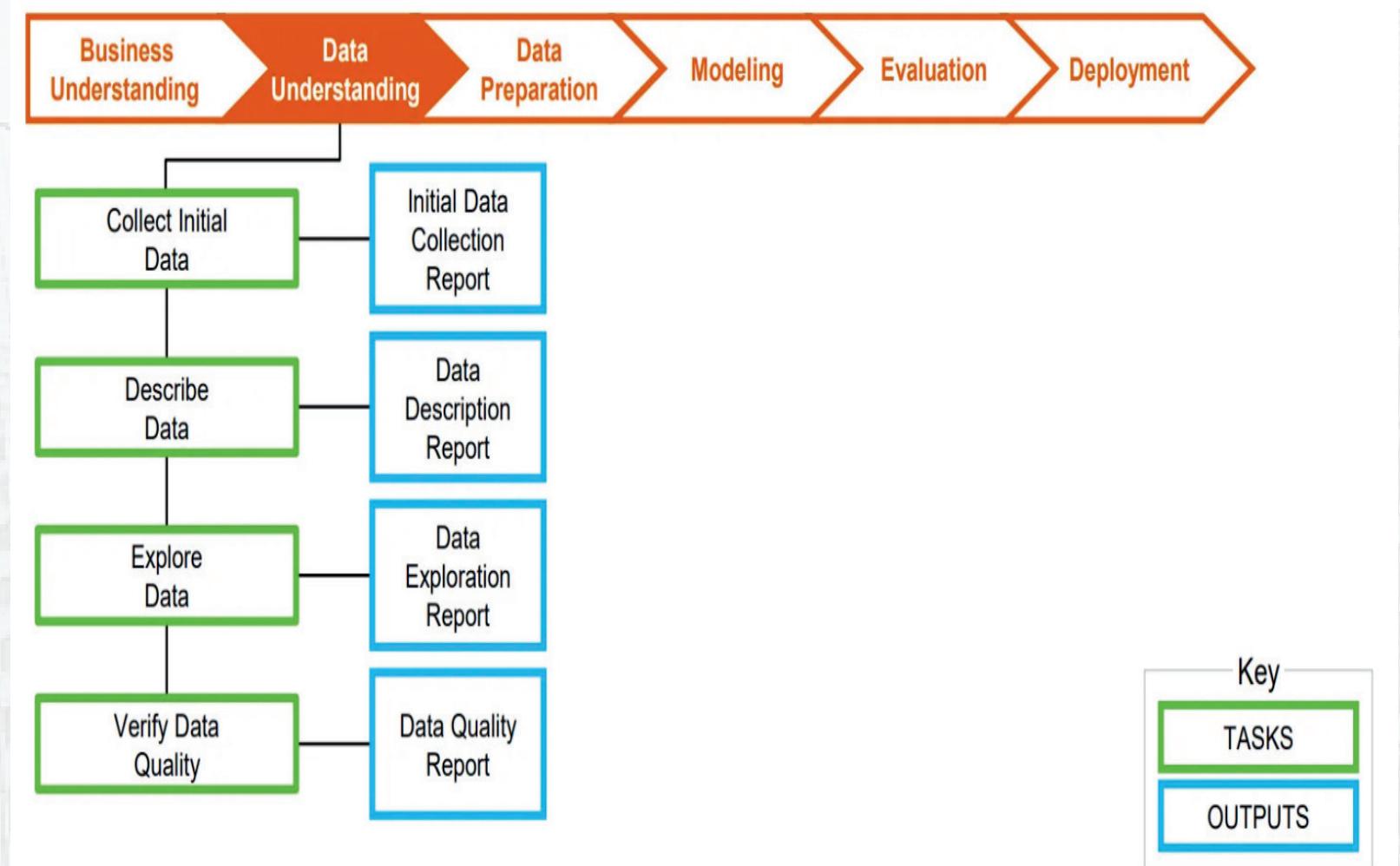


The logo consists of the text "id/x partners" in white. "id/x" is in a blue box, and "partners" is in a dark blue box. The boxes are separated by a thin white vertical line.

# 3. Data Understanding

Data Understanding adalah tahap kedua dalam proses CRISP-DM (Cross-Industry Standard Process for Data Mining) yang fokus pada pengumpulan dan penilaian kualitas data. Tahap ini melibatkan empat tugas utama:

1. **Mengumpulkan Data Awal:** Mengidentifikasi data yang tersedia, metode pengambilan, dan masalah yang mungkin dihadapi.
2. **Mendeskripsikan Data:** Memeriksa properti data yang diperoleh, termasuk format, kuantitas, dan isi dari setiap tabel atau dataset.
3. **Menjelajahi Data:** Menggunakan pertanyaan ilmu data untuk mendapatkan wawasan awal melalui kueri, visualisasi, dan laporan ringkasan.
4. **Memverifikasi Kualitas Data:** Memastikan data cukup bersih dan relevan untuk analisis yang akan dilakukan



## 03



466285 Observasi  
75 Variabel

- Data yang digunakan memiliki nilai hilang
- Tidak memiliki data duplikat

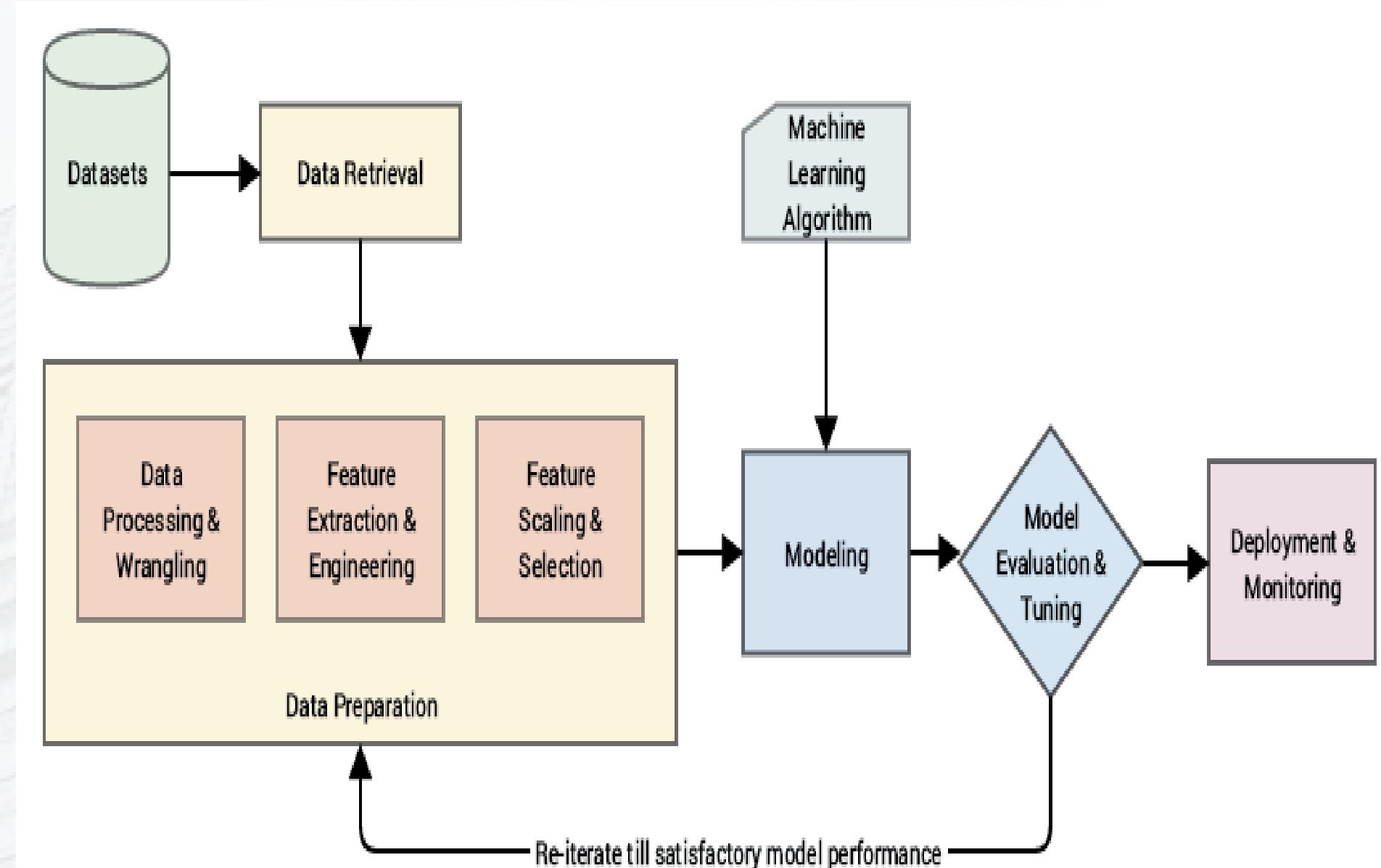
Variabel dependen adalah **loan\_status** dimana dibagi dalam 5 kategori sebagai berikut:

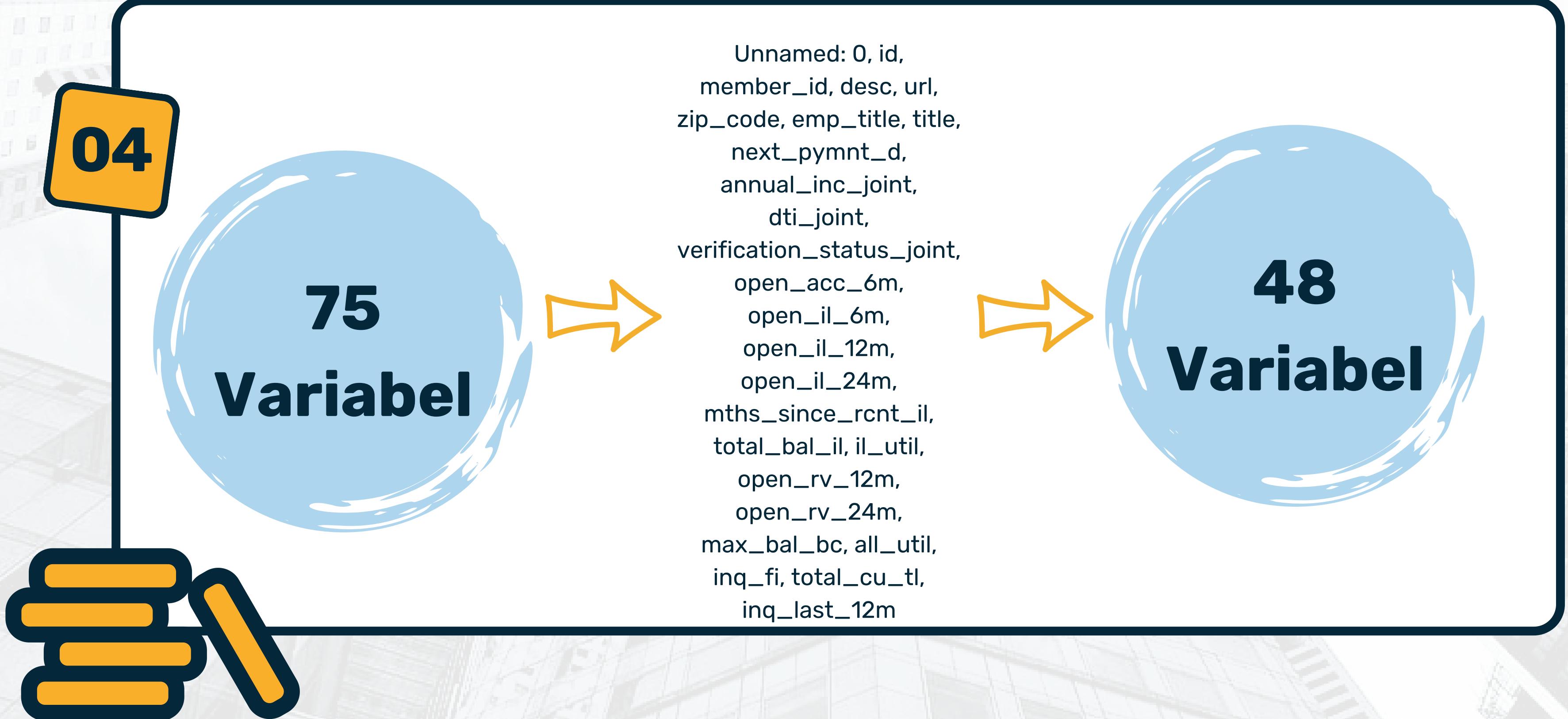
1. Status "Fully Paid" dan "Does not meet the credit policy. Status:Fully Paid" masuk dalam kategori '**Fully Paid**'
2. Status "Charged Off" dan "Does not meet the credit policy. Status:Charged Off" masuk dalam kategori '**Charged Off**'
3. Status "Default" dan "Late (31-120 days)" masuk dalam kategori '**High Risk**'
4. Status "In Grace Period","Late (16-30 days)", dan "Current" masuk dalam kategori '**Moderate Risk**'
5. Status lainnya masuk dalam kategori '**Status**'

# 4. Feature Engineering

Feature Engineering dalam ilmu data adalah proses kreatif yang melibatkan pemilihan, transformasi, dan pembuatan fitur baru dari data mentah untuk meningkatkan kinerja model pembelajaran mesin. Ini termasuk:

- 1. Seleksi Fitur:** Memilih fitur yang paling relevan dengan masalah yang sedang dihadapi.
- 2. Transformasi Fitur:** Mengubah skala atau distribusi fitur untuk meningkatkan interpretasi oleh model.
- 3. Penciptaan Fitur:** Menggabungkan atau memodifikasi fitur untuk menghasilkan informasi yang lebih berguna.
- 4. Ekstraksi Fitur:** Mengidentifikasi dan mengekstrak informasi penting dari data mentah.

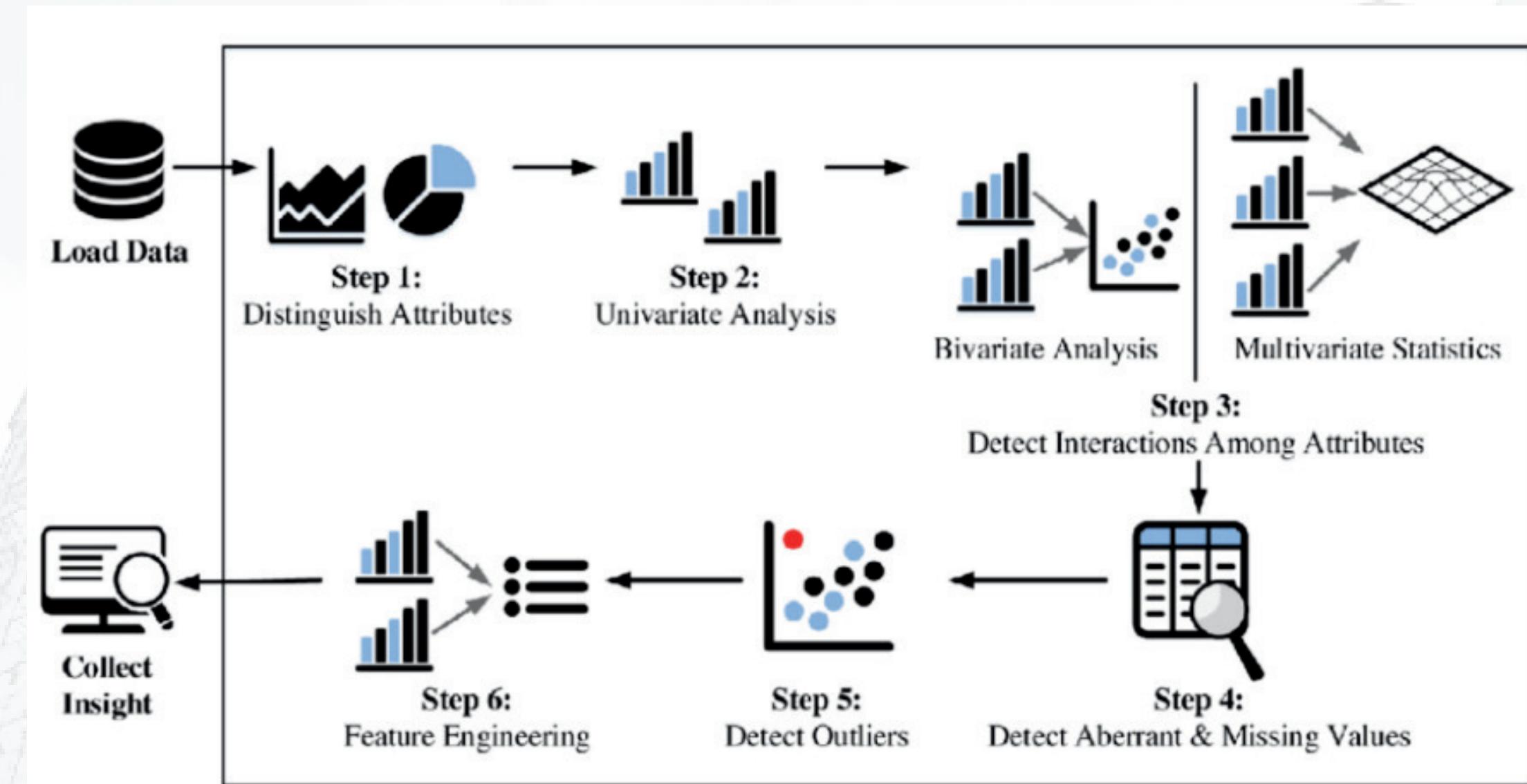




# 5. Exploratory Data Analysis

Exploratory Data Analysis (EDA) dalam ilmu data adalah proses analisis awal data untuk memahami karakteristik utama, menemukan pola, anomali, dan hubungan antar variabel. Proses EDA biasanya meliputi:

1. **Gambaran Umum Dataset:** Memahami jumlah observasi, jenis fitur, dan data yang hilang.
2. **Statistik Deskriptif:** Meringkas data numerik melalui ukuran tendensi sentral dan dispersi.
3. **Visualisasi Data:** Menggunakan grafik dan diagram untuk menggambarkan distribusi dan hubungan data.
4. **Evaluasi Kualitas Data:** Memeriksa kebersihan dan konsistensi data.



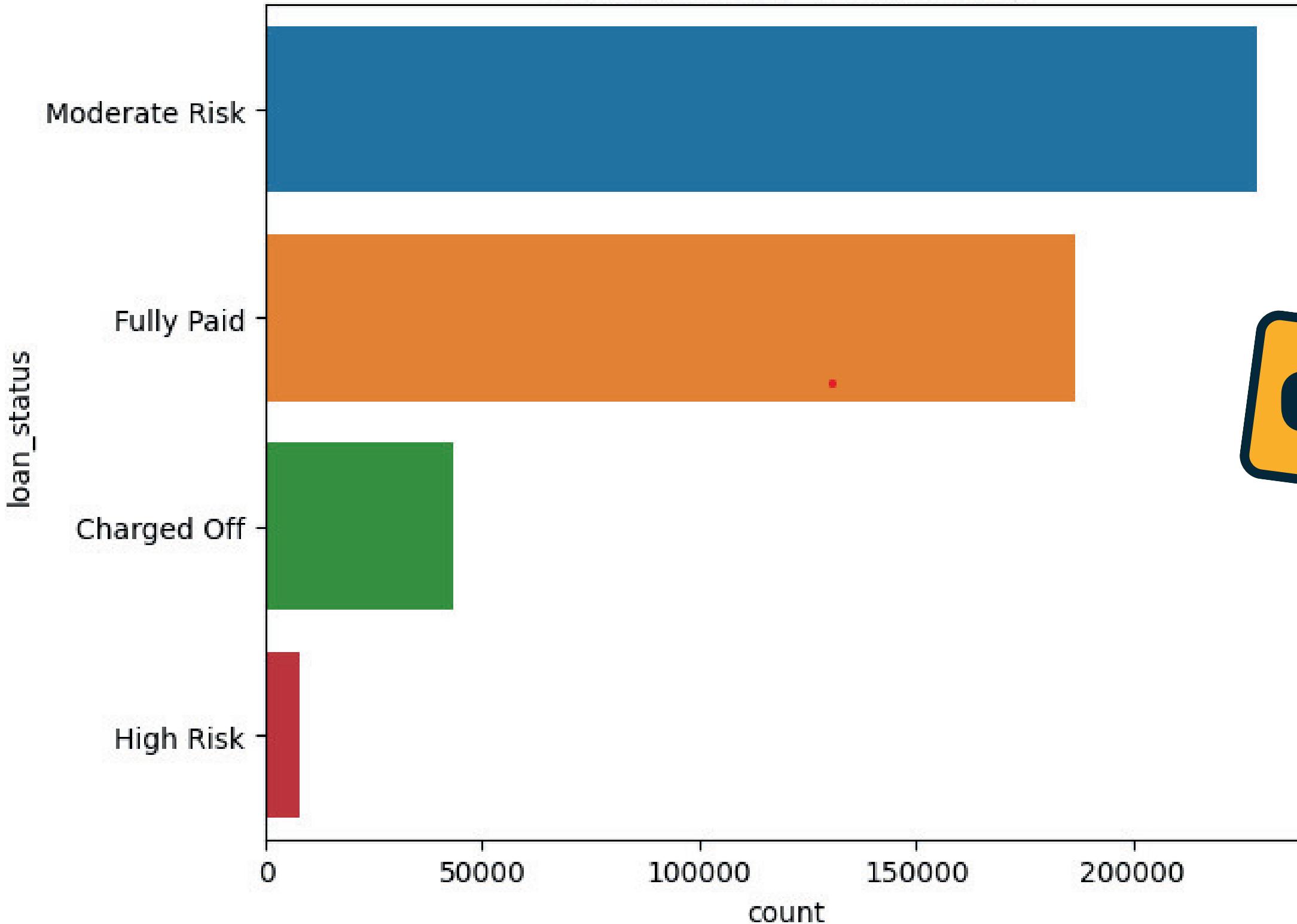
04

## Handle Missing Value

Nilai kosong pada tipe kolom numerik  
di isi dengan nilai median

Nilai kosong pada tipe kolom kategori  
di isi dengan nilai modus

Countplot for loan\_status



05

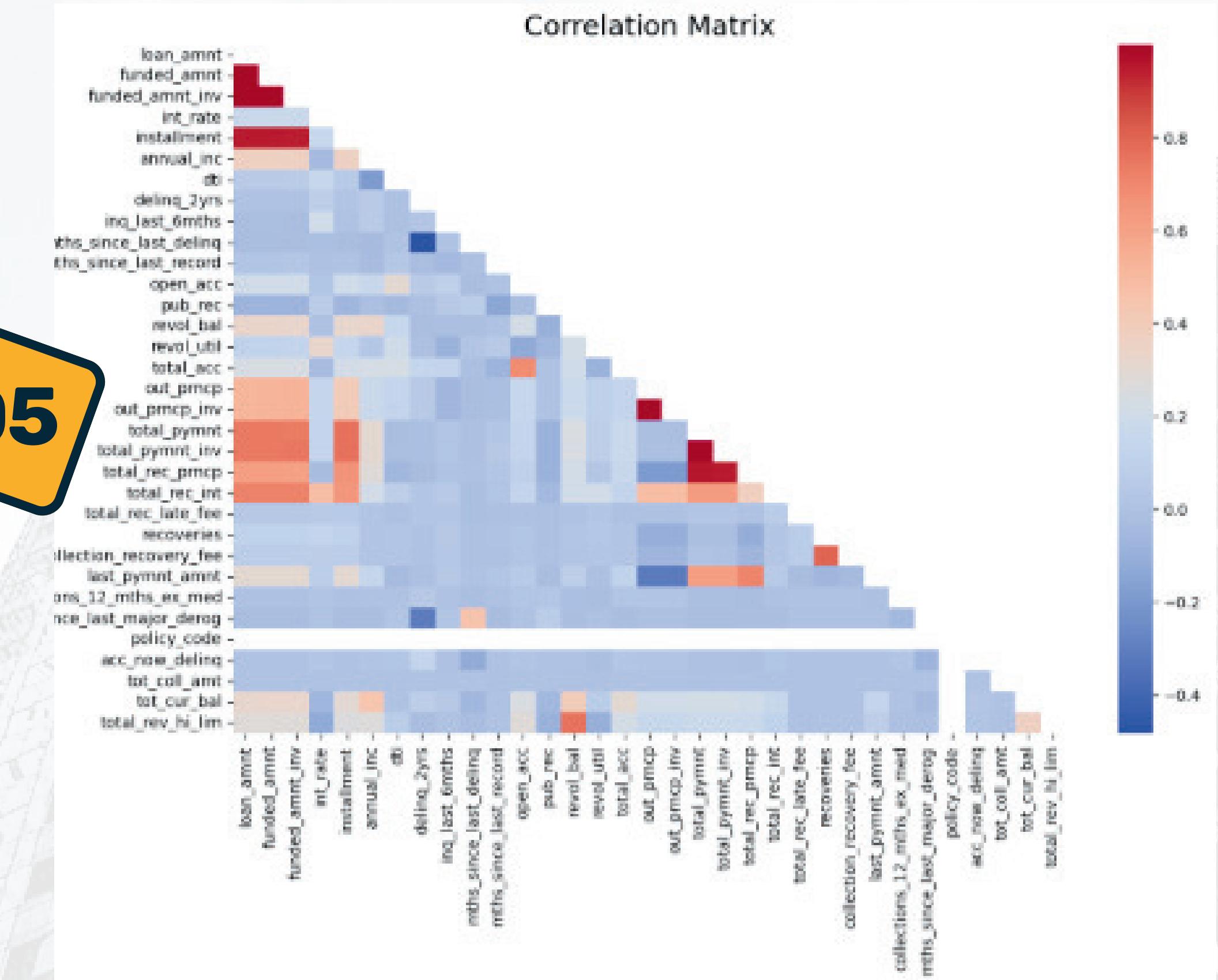
**loan\_status** dengan kategori  
**Moderate Risk** dan **Fully Paid**  
yang paling banyak

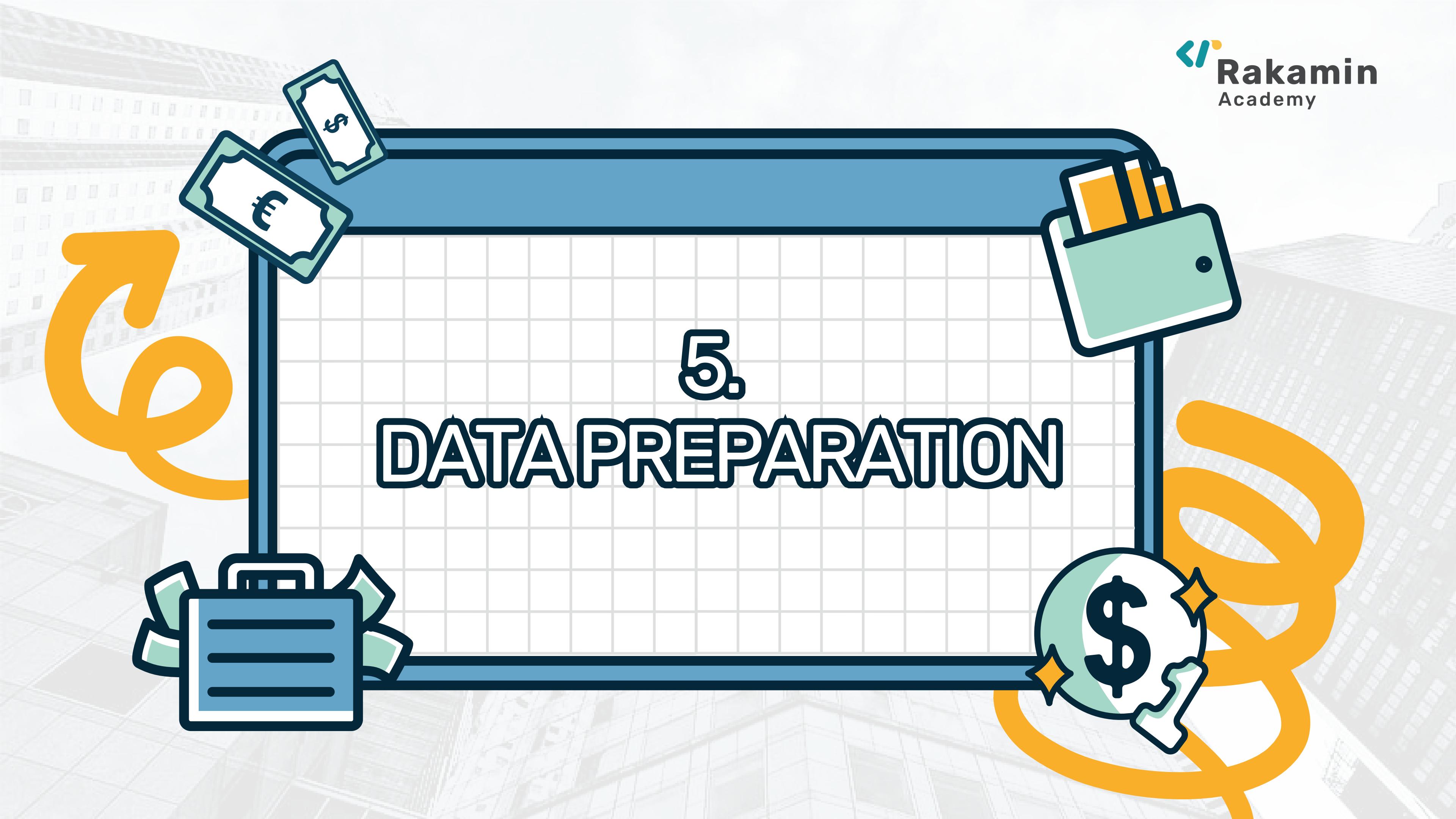
**Insight:**

- Peminjam '**Moderate Risk**' menjadi target untuk program manajemen risiko
- Kategorisasi peminjam '**High Risk**' perlu ditinjau lebih lanjut untuk memahami faktor penyebab dan meminimalkan potensi kerugian

- Fitur **loan\_amnt** dan **funded\_amnt** memiliki korelasi yang sangat tinggi
  - Terdapat korelasi negatif yang signifikan antara **loan\_amnt** dan **int\_rate**

**Insight:**  
Perusahaan dapat  
mengembangkan model  
penetapan harga yang lebih  
dinamis





# 5. **DATA PREPARATION**

05

## Proses Encoding untuk loan\_status

Charged Off : 0

Fully Paid : 1

High Risk : 2

Moderate Risk : 3

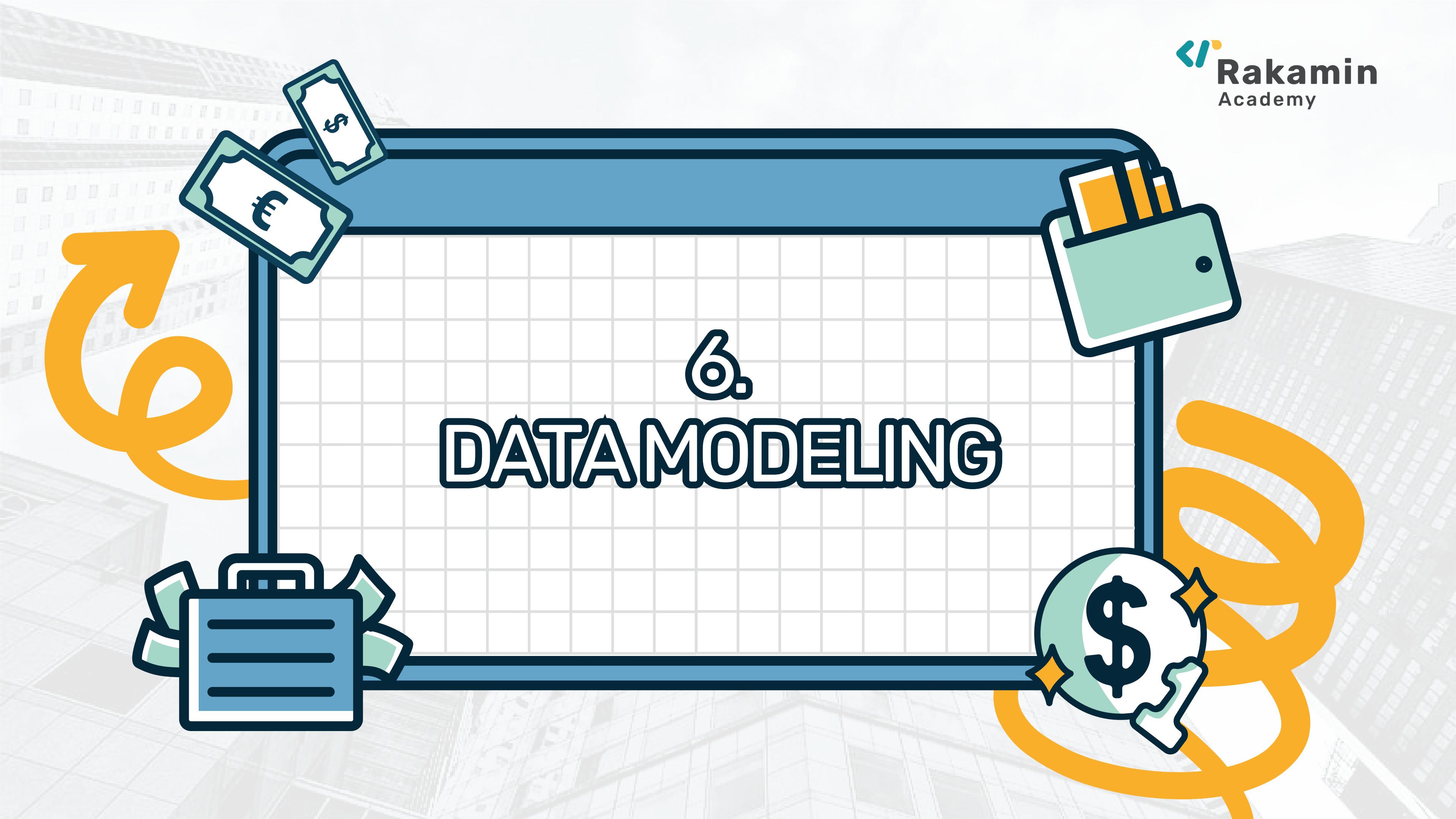




05

80%  
Data Latih

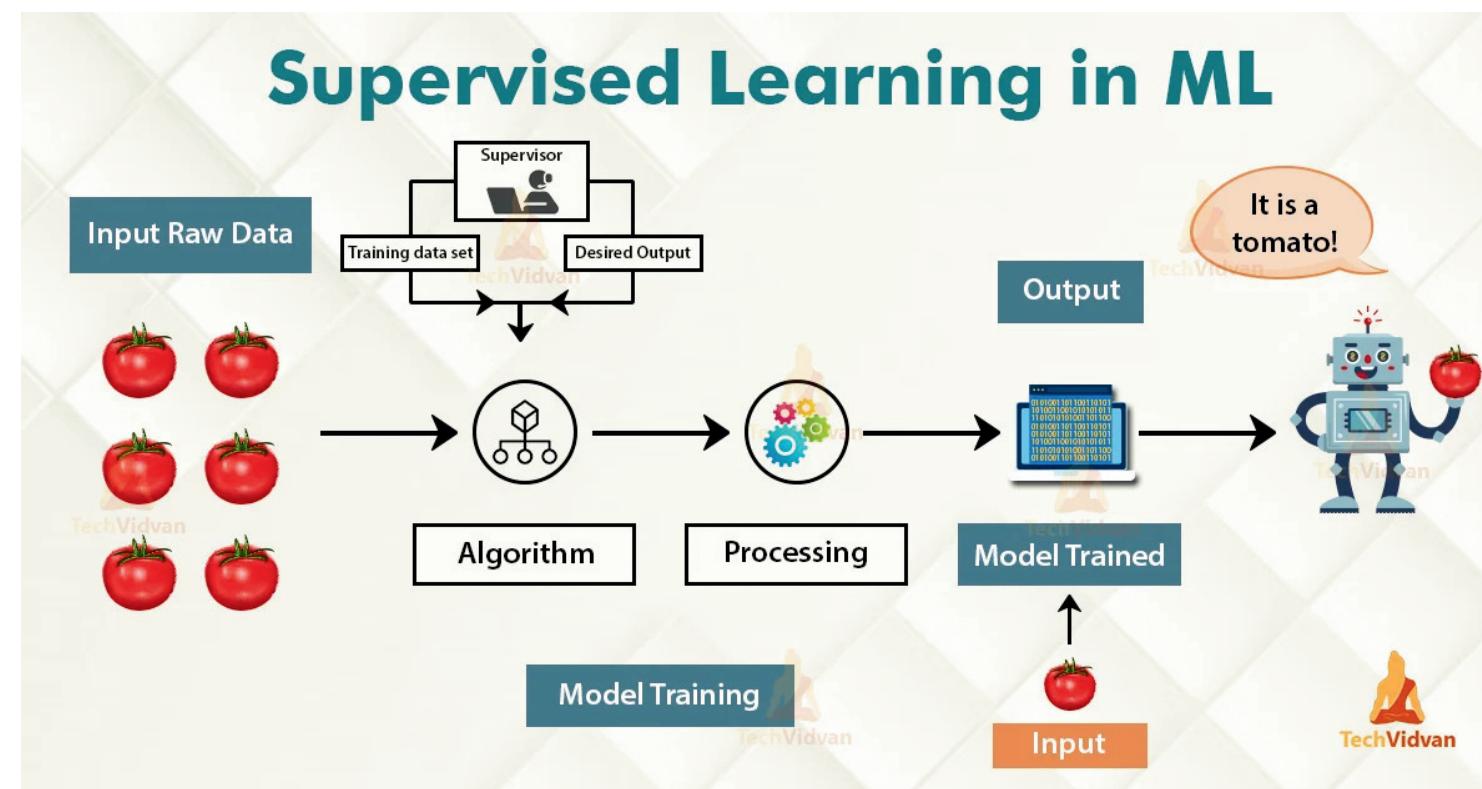
80%  
Data Uji



# 6. **DATA MODELING**

06

Model yang digunakan untuk memprediksi adalah:



## Algoritma Machine Learning:

1. Logistic Regression
2. Decision Tree
3. Random Forest
4. XGBoosts

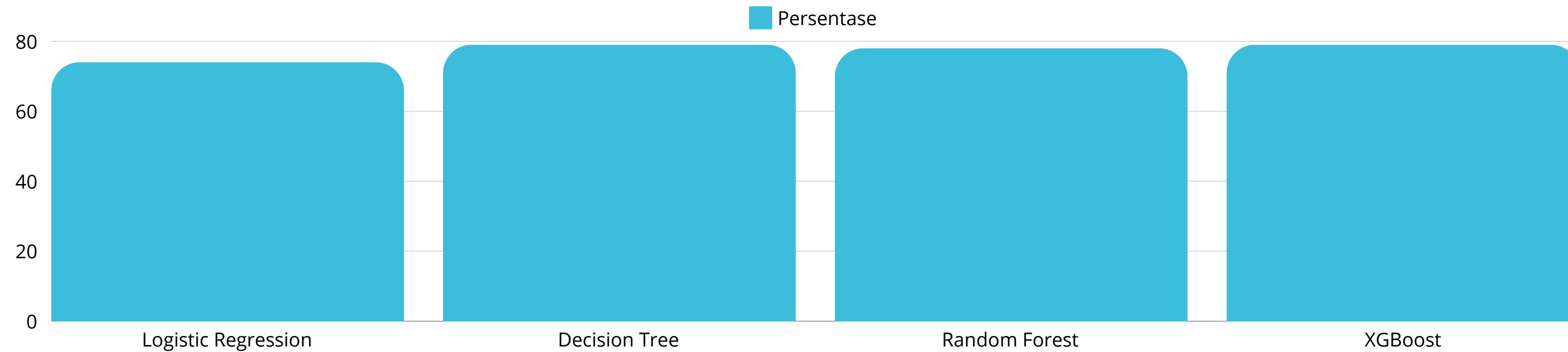
**06**

<b>Algoritma Machine Learning</b>	<b>Akurasi</b>	<b>Precision</b>	<b>Recall</b>	<b>F1 Score</b>	<b>roc-auc</b>
Logistic Regression	0.9806	0.7401	0.7457	0.7428	0.9456
Decision Tree	0.9710	0.7918	0.7975	0.7944	0.8935
Random Forest	0.9845	0.9762	0.7806	0.8027	0.9606
XGBoost	0.9842	0.9508	0.7904	0.8155	0.9701

07

## False Negative

Jumlah kasus dimana model memprediksi bahwa status pinjaman adalah 'buruk' tetapi ternyata benar 'baik'





## 8. **EVALUATION**

07

<b>Algoritma Machine Learning</b>	<b>Mean</b>	<b>Std</b>
Logistic Regression	0,9728	0.0012
Decision Tree	0.9468	0.0072
Random Forest	0.9748	0.0017
XGBoost	0.9671	0.0120



# 9. Conclusion

- Dari 100 peminjam, sekitar 76 orang akan lolos untuk mendapatkan pinjaman, dengan 92.1% adalah peminjam baik dan 7.9% adalah peminjam berisiko.
- Model dengan recall sebesar 79% cukup baik dalam mendekripsi sebagian besar peminjam berisiko, tetapi masih menyisakan risiko kecil dari peminjam berisiko yang tidak terdeteksi.

## Rekomendasi Bisnis:

Fokus pada minimizing **false negatives** untuk memaksimalkan akurasi deteksi peminjam berisiko tinggi

# Thank You



**Rakamin**  
Academy



**id/x** partners