



# Final Project Data Science

Studi Independen Bersertifikat Batch 6 - CourseNet Indonesia  
(Januari - Juni 2024)

Furqon Nurbaril Yahya



# Table of contents

## 01 Overview

Latar Belakang Masalah

## 02 Exploratory Data Analysis

Memahami pola data pada  
dataset

## 03 Data Preprocessing

Penskalaan data, Handling  
missing values, Encoding,  
dan Oversampling

## 04 Machine Learning Model

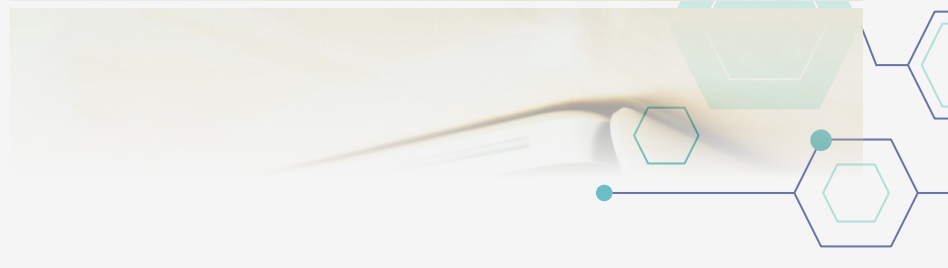
Rekomendasi Model  
Machine Learning Terbaik



# 01

# Overview

Latar belakang masalah dan tujuan proyek



# Overview

## Latar Belakang Masalah

- Tim bisnis di sebuah bank semakin resah dengan meningkatnya jumlah nasabah yang menutup layanan kartu kreditnya.
- Menjaga pelanggan tetap setia adalah kunci untuk keberlanjutan bisnis bank.

## Tujuan Proyek

- Mengembangkan model prediktif untuk mengidentifikasi nasabah yang berpotensi churn.
- Memberikan treatment/layanan khusus kepada pelanggan yang berisiko tinggi untuk churn.
- Meningkatkan retensi pelanggan dan mengurangi churn rate.

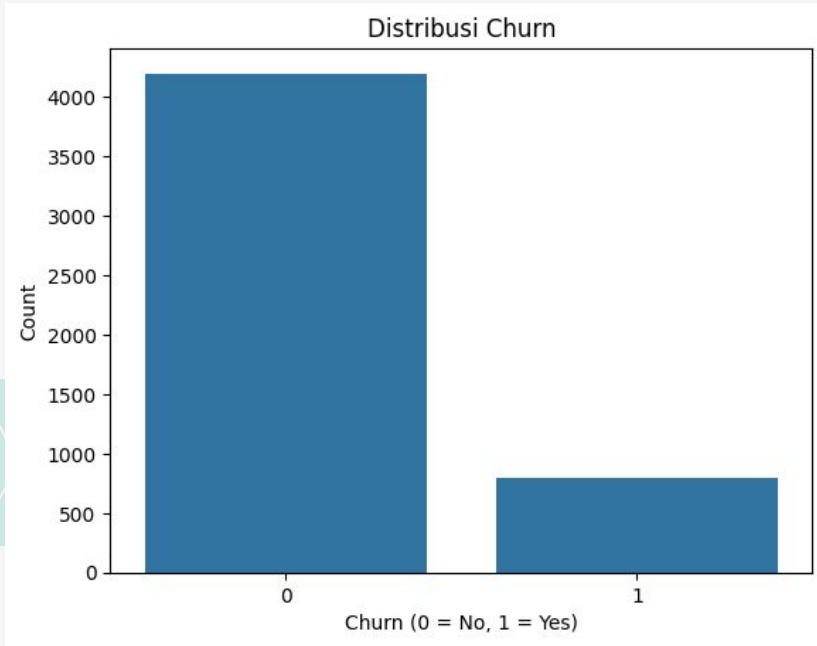
# 02

## Exploratory Data Analysis

Memahami pola data pada dataset



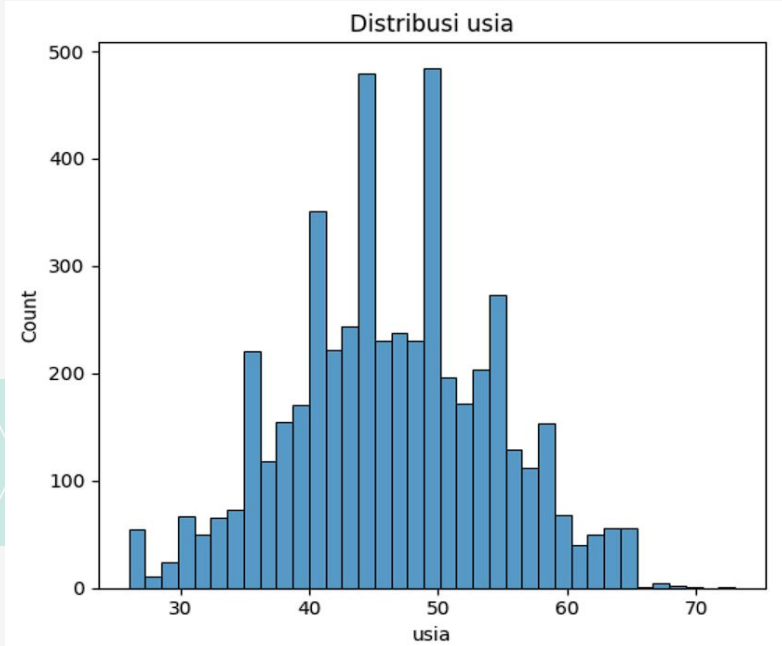
# Exploratory Data Analysis



- Label 0 (Not Churn) = 4200
- Label 1 (Churn) = 800

**Poin:** Jumlah pelanggan yang tidak churn lebih banyak dibandingkan dengan yang churn. Ini menunjukkan bahwa churn adalah kejadian yang relatif jarang dalam dataset ini.

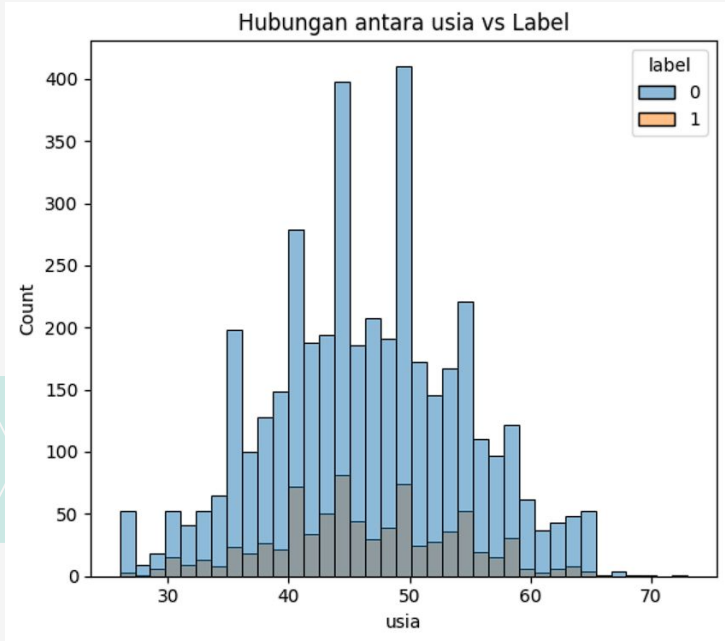
# Exploratory Data Analysis



Rata-rata usia pelanggan: 40 - 50 tahun

Poin: Mayoritas pelanggan bank berada dalam usia produktif.

# Exploratory Data Analysis

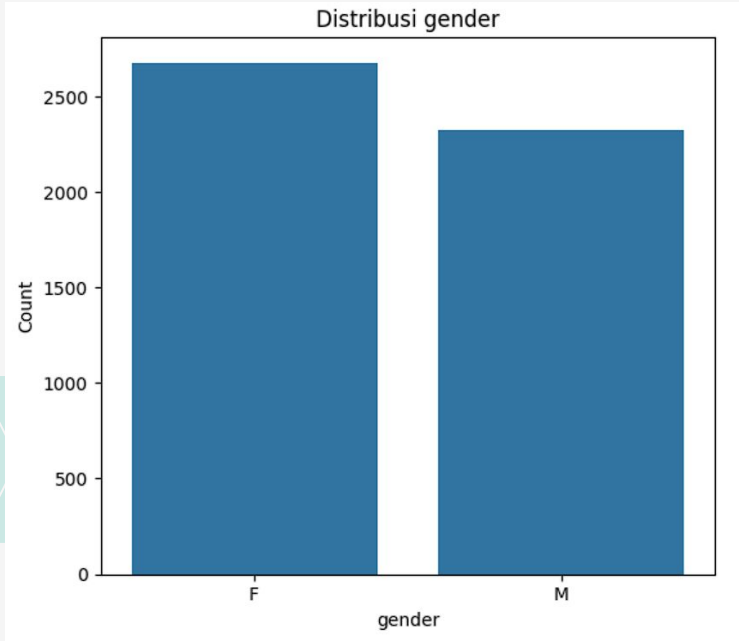


Rata-rata usia pelanggan: 40 - 50 tahun

Poin: Tidak ada perbedaan churn yang signifikan berdasarkan usia pelanggan.

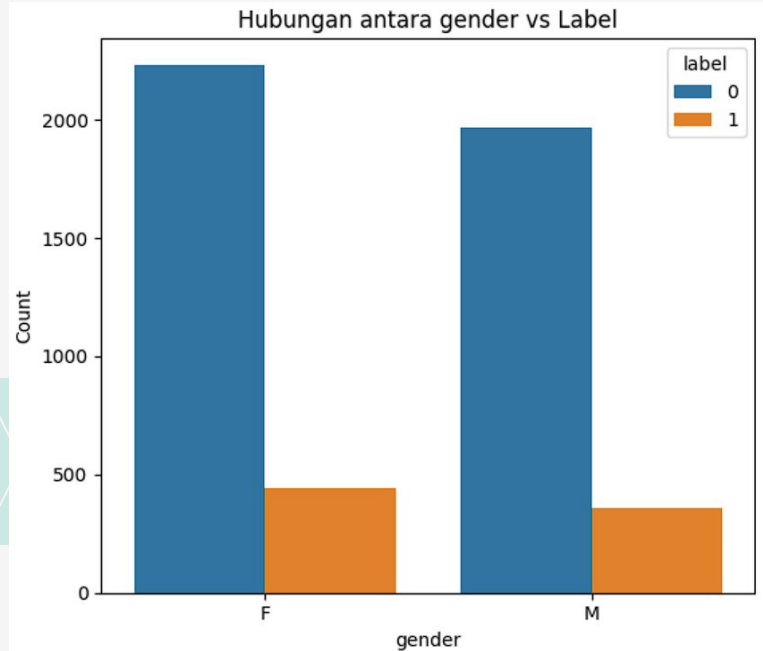


# Exploratory Data Analysis



Poin: Proporsi gender pelanggan relatif seimbang antara laki-laki dan perempuan

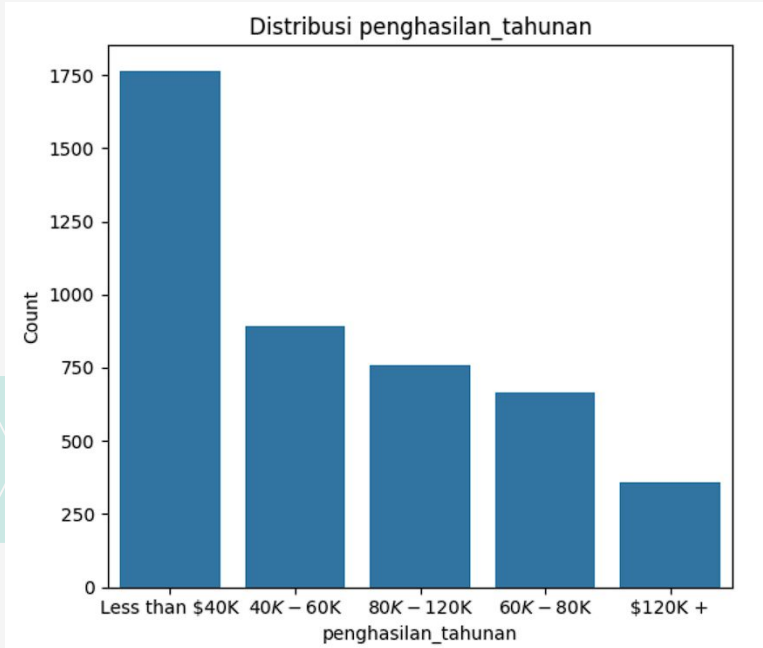
# Exploratory Data Analysis



Proporsi gender pelanggan relatif seimbang antara laki-laki dan perempuan

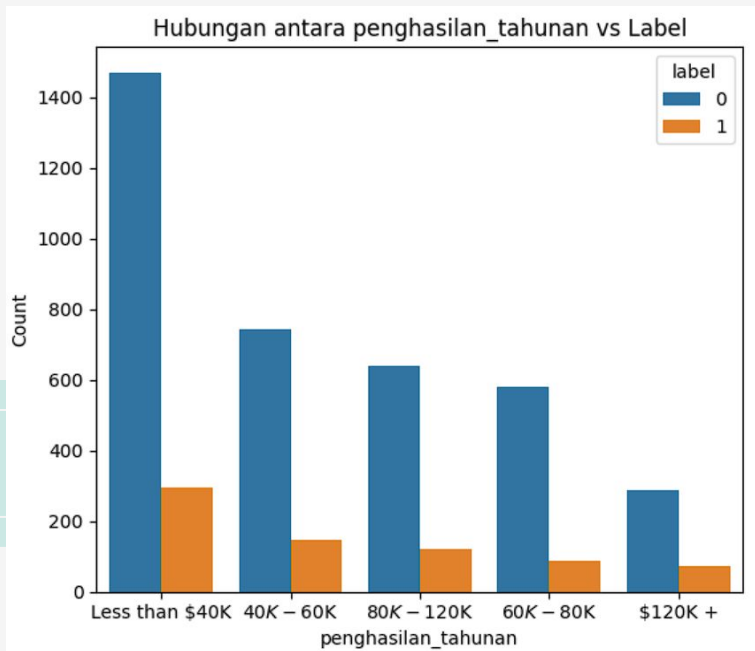
Poin: Tidak ada perbedaan churn yang signifikan berdasarkan gender pelanggan.

# Exploratory Data Analysis



Poin: Mayoritas pelanggan memiliki penghasilan tahunan < 40 ribu dolar per tahun.

# Exploratory Data Analysis

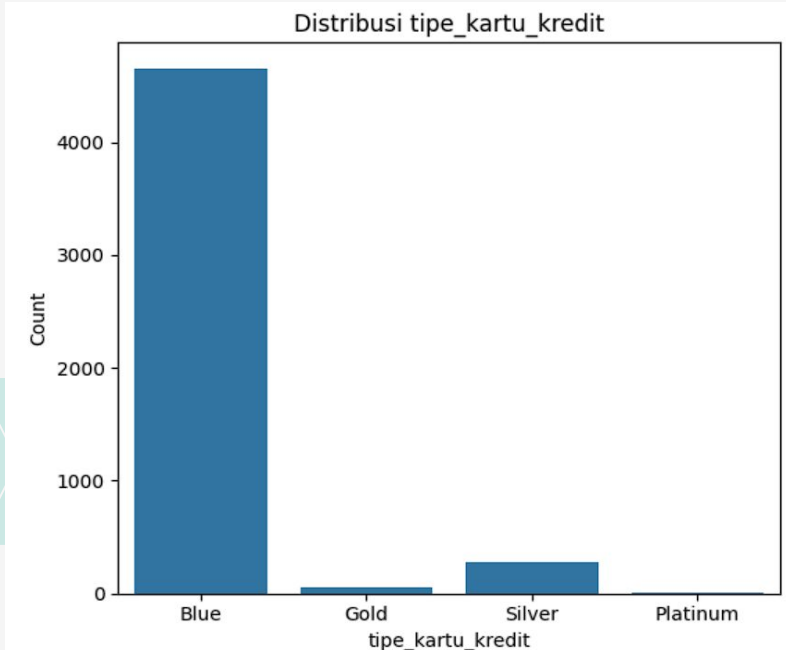


## Poin:

1. Segmen pelanggan berpenghasilan <\$40K memiliki jumlah churn (label 1) tertinggi di antara semua kelompok penghasilan.
2. Segmen pelanggan berpenghasilan >\$120K memiliki jumlah churn terendah.
3. Semakin tinggi penghasilan, semakin rendah kecenderungan untuk churn, dan sebaliknya.

**Insight:** Pelanggan berpenghasilan rendah mungkin lebih sensitif terhadap biaya layanan atau produk, sehingga lebih rentan untuk churn.

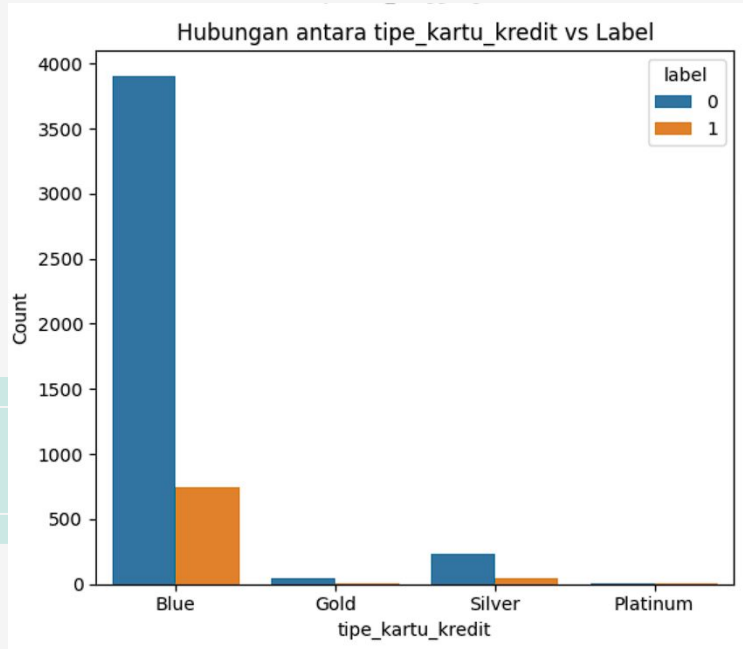
# Exploratory Data Analysis



Poin: Mayoritas pelanggan menggunakan kartu kredit tipe **Blue**



# Exploratory Data Analysis

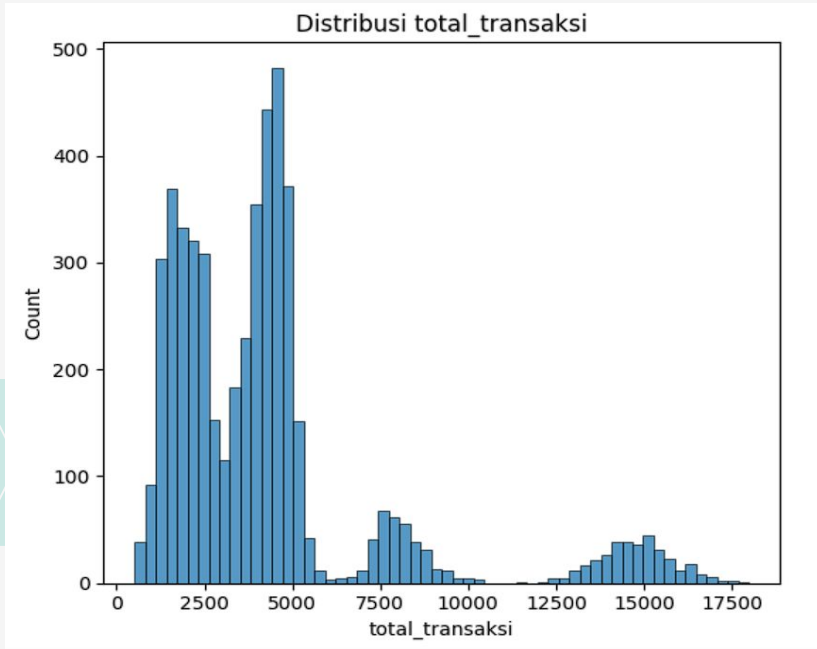


## Poin:

1. Dari grafik, terlihat perbedaan yang signifikan antara jumlah churn untuk kartu Blue dibandingkan jenis kartu lainnya (Gold, Silver, Platinum).
2. Untuk kartu Gold, Silver, dan Platinum, tingkat churn relatif rendah dan hampir sama.
3. Ini mengindikasikan adanya faktor atau karakteristik khusus terkait kartu Blue yang memengaruhi kecenderungan pelanggan untuk tidak melanjutkan layanan (churn).

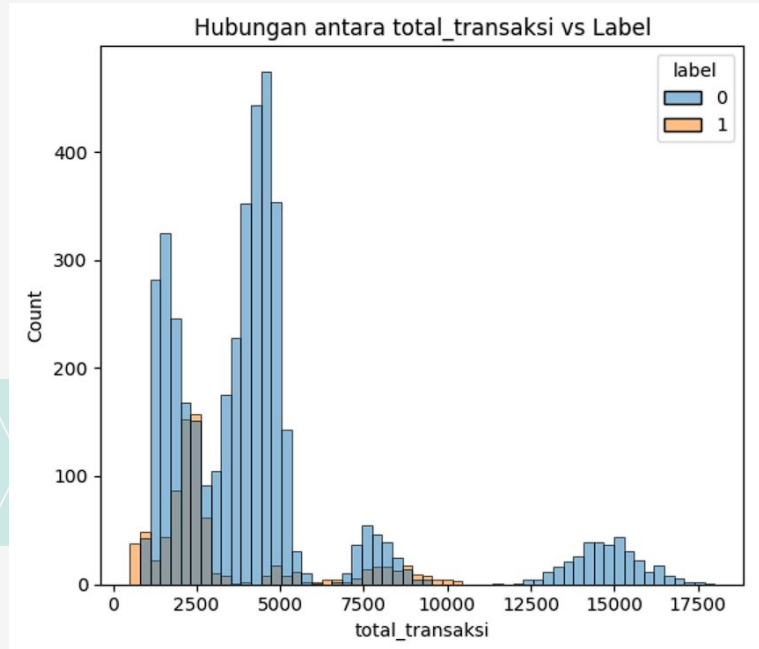
**Insight:** Pelanggan dengan kartu Blue mungkin memiliki pengalaman atau persepsi yang kurang memuaskan terhadap layanan atau dukungan yang diberikan oleh Bank.

# Exploratory Data Analysis



Poin: Total transaksi dengan jumlah sekitar 500-5000 banyak dilakukan oleh mayoritas pelanggan

# Exploratory Data Analysis



**Insight:** Pelanggan dengan total transaksi tinggi mungkin lebih cocok dengan layanan atau produk yang ditawarkan.

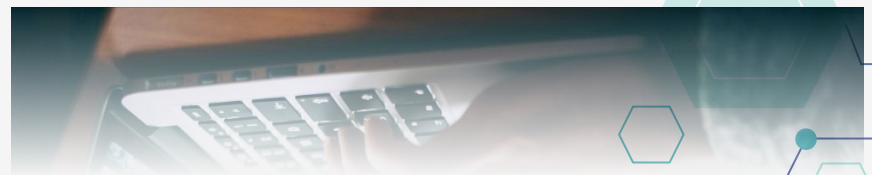




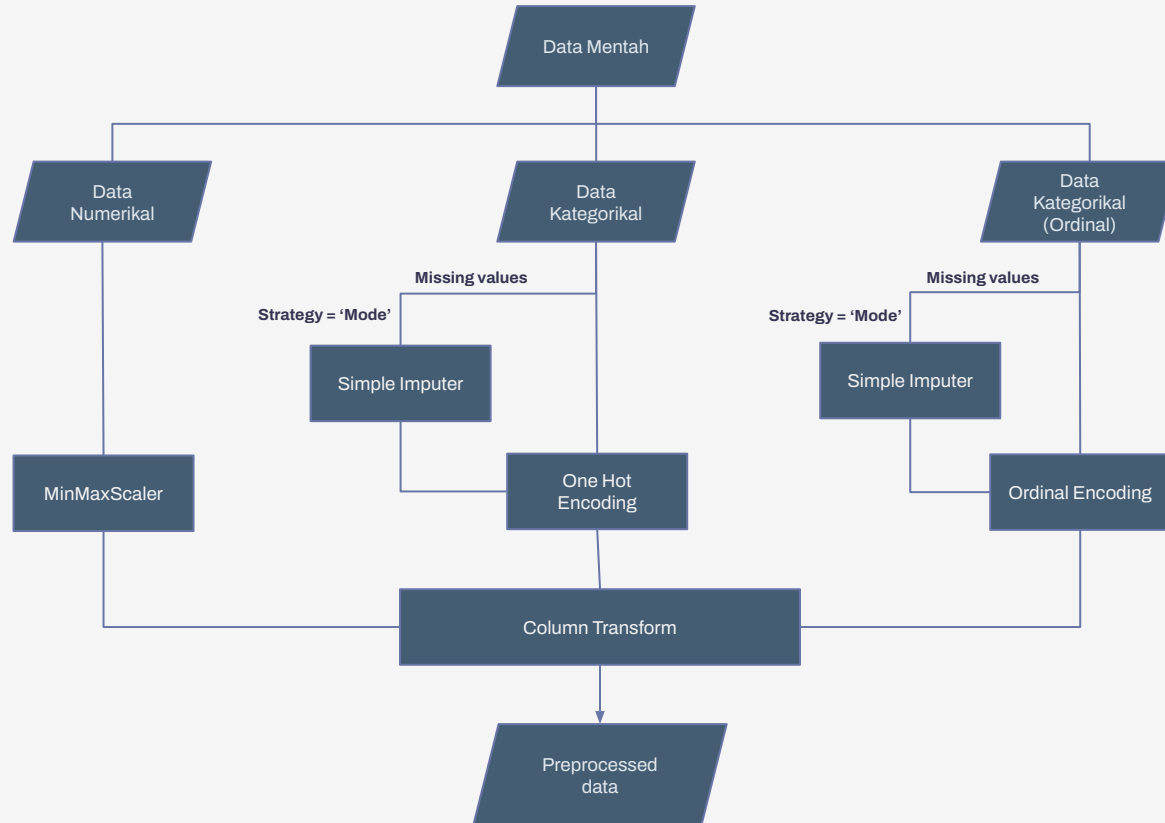
# 03

## Data Preprocessing

Penskalaan data, Handling missing values, Encoding, dan Handling Imbalance Label

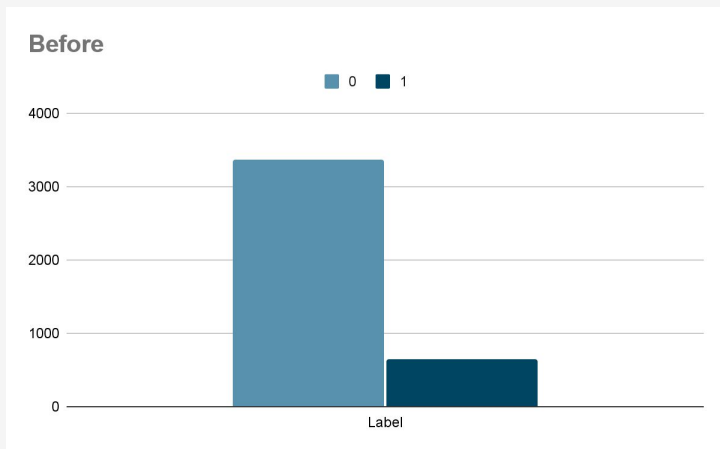


# Data Preprocessing

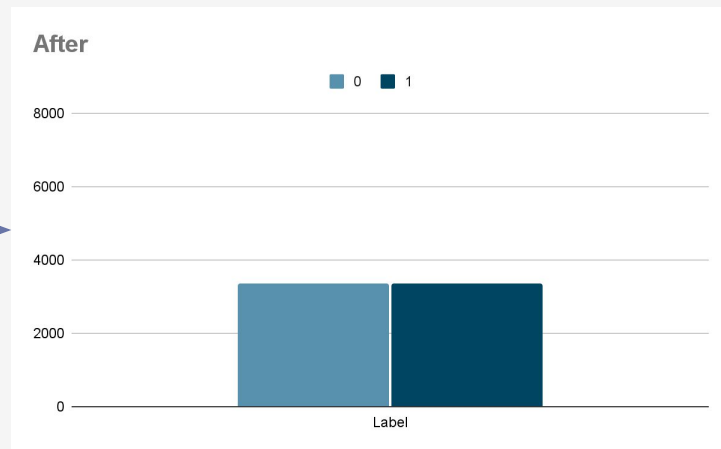


# Data Preprocessing

## Handling Imbalance Label



SMOTENC

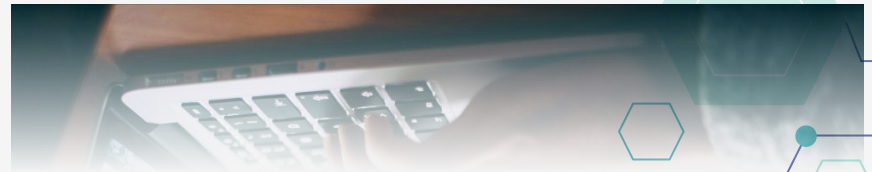




# 04

## Machine Learning Model

Rekomendasi Model Machine Learning Terbaik



# Machine Learning Model

Model	Hyperparameter	Precision		Recall		F1-score		Accuracy	
		Training	Testing	Training	Testing	Training	Testing	Training	Testing
SVM	C: 100, kernel: poly	0: 0.98 1: 0.75	0: 0.95 1: 0.64	0: 0.94 1: 0.91	0: 0.92. 1: 0.76	0: 0.96 1: 0.83	0: 0.93 1: 0.69	0.94	0.89
Decision Tree	max_depth: 10, min_samples_leaf: 1, min_samples_split: 2	0: 1.00 1: 0.90	0: 0.96 1: 0.75	0: 0.98 1: 0.99	0: 0.95 1: 0.81	0: 0.99 1: 0.94	0: 0.96 1: 0.78	0.98	0.93
Random Forest	max_depth: 10, min_samples_leaf: 1, min_samples_split: 2, n_estimators': 100	0: 1.00 1: 0.93	0: 0.98 1: 0.84	0: 0.98 1: 0.99	0: 0.97 1: 0.88	0: 0.99 1: 0.96	0: 0.97 1: 0.86	0.99	0.95

Hasil ini menunjukkan bahwa Random Forest adalah pilihan optimal untuk memprediksi data kami dengan tingkat ketepatan yang tinggi, baik pada data training maupun testing.

# Slide Source:

- [https://slidesgo.com/theme/business-management-and-analytics-thesis-defense?utm\\_source=whatsapp&utm\\_medium=social-organic&utm\\_campaign=template-whatsapp](https://slidesgo.com/theme/business-management-and-analytics-thesis-defense?utm_source=whatsapp&utm_medium=social-organic&utm_campaign=template-whatsapp)

# Image Source:

- [https://unsplash.com/photos/person-using-laptop-computer-holding-card-Q59HmzK38eQ?utm\\_content=creditShareLink&utm\\_medium=referral&utm\\_source=unsplash](https://unsplash.com/photos/person-using-laptop-computer-holding-card-Q59HmzK38eQ?utm_content=creditShareLink&utm_medium=referral&utm_source=unsplash)
- [https://unsplash.com/photos/turned-on-black-and-grey-laptop-computer-mcSDtbWXUZU?utm\\_content=creditShareLink&utm\\_medium=referral&utm\\_source=unsplash](https://unsplash.com/photos/turned-on-black-and-grey-laptop-computer-mcSDtbWXUZU?utm_content=creditShareLink&utm_medium=referral&utm_source=unsplash)
- [https://unsplash.com/photos/person-using-macbook-pro-npxXWgQ33ZQ?utm\\_content=creditShareLink&utm\\_medium=referral&utm\\_source=unsplash](https://unsplash.com/photos/person-using-macbook-pro-npxXWgQ33ZQ?utm_content=creditShareLink&utm_medium=referral&utm_source=unsplash)

**Thank You**

