

Propose Machine Learning

# Banking Deposit

Rendi Haryadi Dharmawan  
JCDSAH-0108

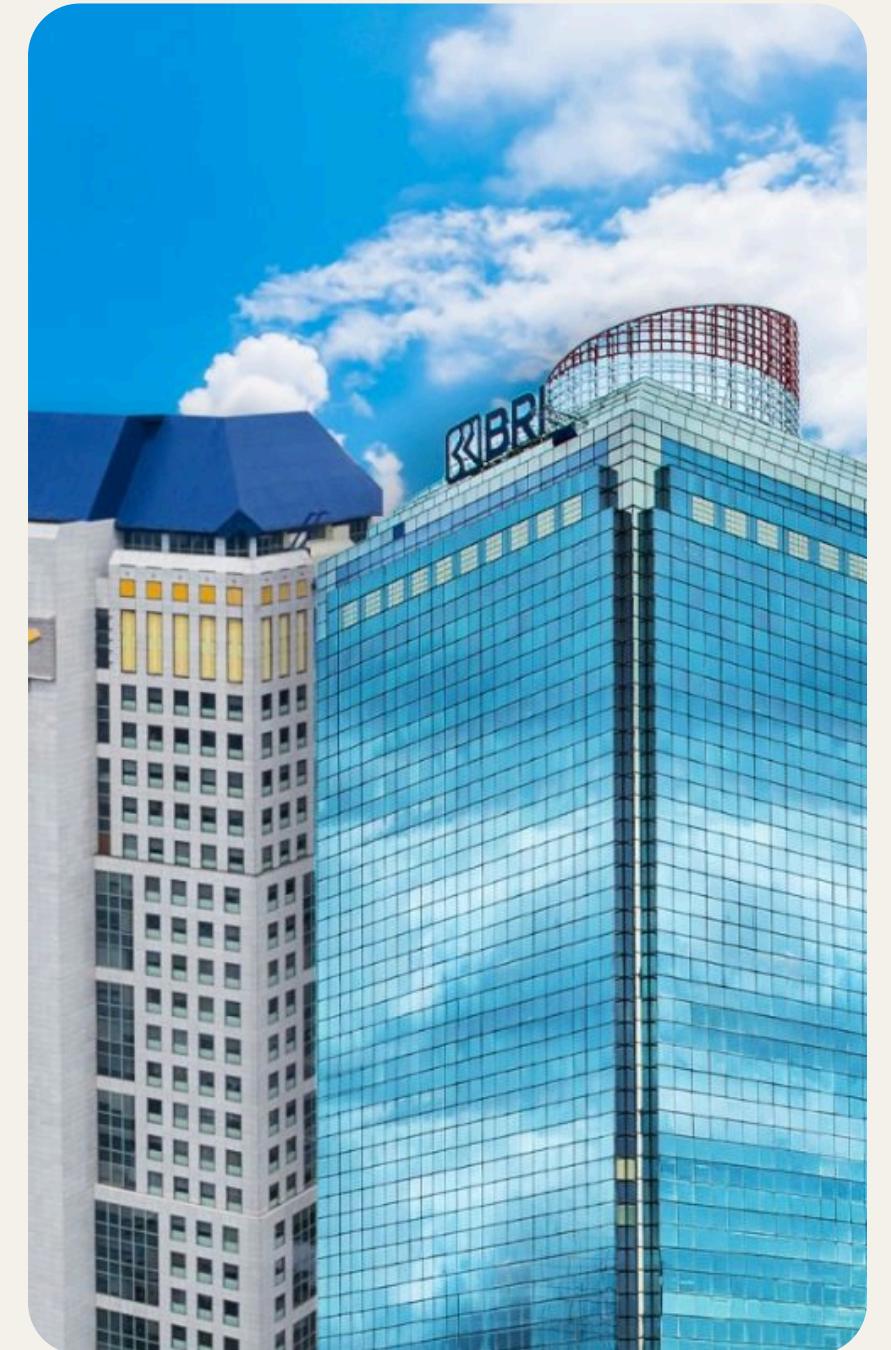
# Introduction

1. About the bank
2. Problem statement

# About The bank

The bank saat ini merupakan bank lokal yang telah memiliki 11.162 nasabah dengan fokus profit nya Nasabah melakukan deposito berjangka dari nasabah-nasabah nya.

Sepanjang berjalannya operasional bank, the bank telah menerima deposito dari 5.289 nasabah nya



# Problem statement

## **Perlu nya sistem untuk mempermudah pemberian deposito**

Dengan banyak nya konsiderasi berdasarkan portofolio nasabah, the bank memperlukan sistem yang dapat memfilter awal apakah nasabah tersebut mempunyai kans untuk melakukan deposito

### **Problem statement:**

Sistem apa yang perlu dibuat untuk menentukan  
Apakah nasabah berpotensi melakukan deposito?



# Data Understanding

1. Data cleaning
2. Summary Kondisi Eksisting  
(Exploratory Data Analysis)
3. Feature Engineering

# Data Identifier

## Nasabah ID

- **Age** – Age of customer
- **Job** – Job of customer
- **Martial** – Martial status of customer
- **Education** – Customer education level
- **Default** – Has credit in default?
- **Housing** – If costumer has housing loan
- **Loan** – Has Personal Loan
- **Balance** – Customer's individual balance
- **Contact** – Communication type
- **Deposit** – has the client subscribed a term deposit

## Campaign

- **Month** – Last contact month of year
- **Day** – Last contact day of the week
- **Duration** – Last contact duration, in seconds
- **Campaign** – Number of contacts performed during this campaign and for this client
- **Pdays** – Number of days that passed by after the client was last contacted from a previous campaign
- **Previous** – Number of contacts performed before this campaign and for this client
- **Poutcome** – outcome of the previous marketing campaign

# Data cleaning

| Kategori              | Action  |
|-----------------------|---|
| Data 'unknown'        | Nilai 'unknown' pada fitur kategorikal akan diperlakukan sebagai kategori terpisah. |
| Value -1 pada 'pdays' | -1 pada 'pdays' berarti klien tidak pernah dihubungi sebelumnya.<br>Ubah menjadi 0  |
| Duplikat data         | Tidak ada   |
| Outlier               | Gunakan IQR atau range Q1-Q3  |
| kolom 'duration'      | fitur yang diketahui hanya setelah panggilan selesai.<br>drop                       |

# Summary Kondisi Eksisting (EDA)

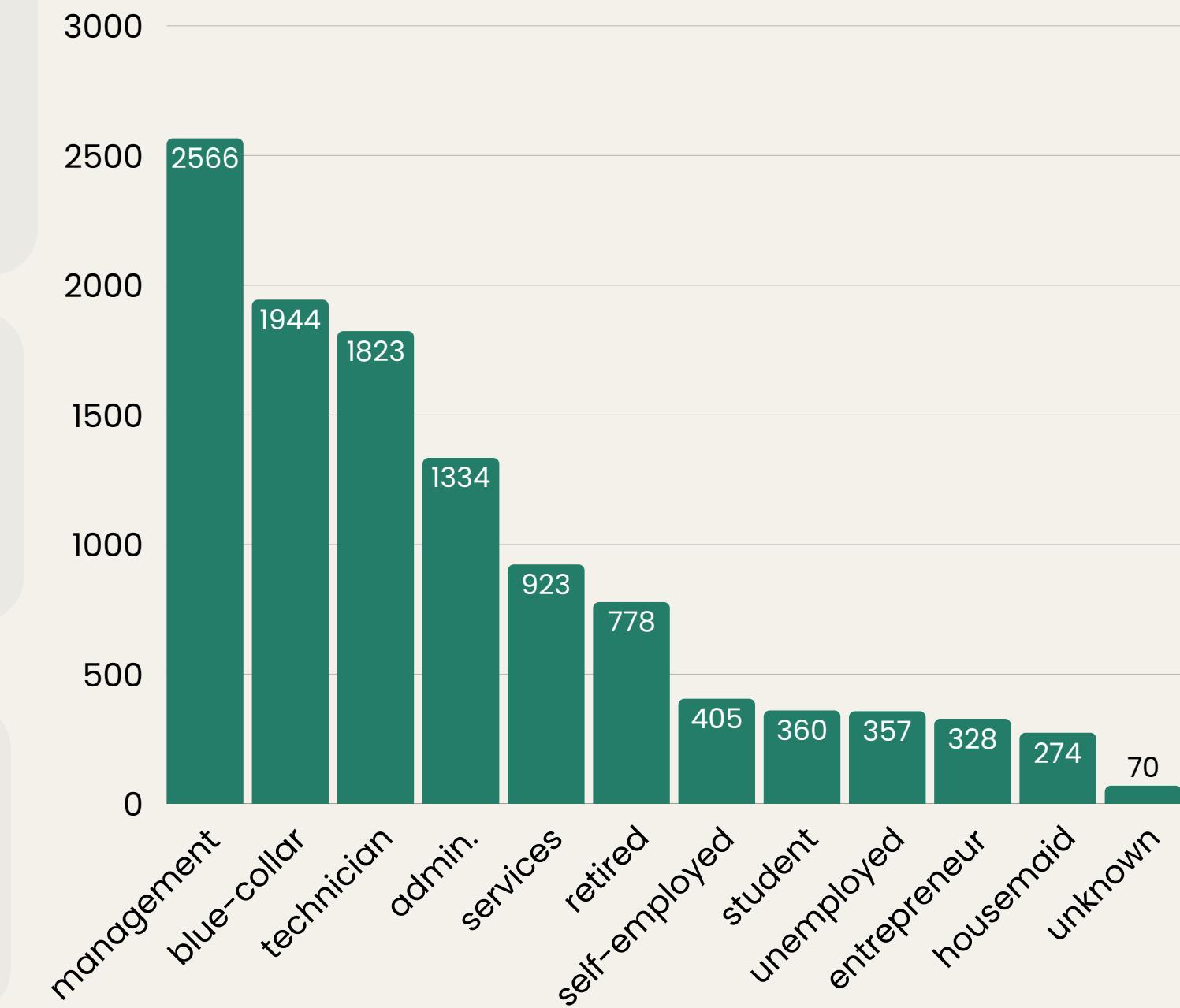
## TOTAL DATA

**11.162**  
**Nasbah**

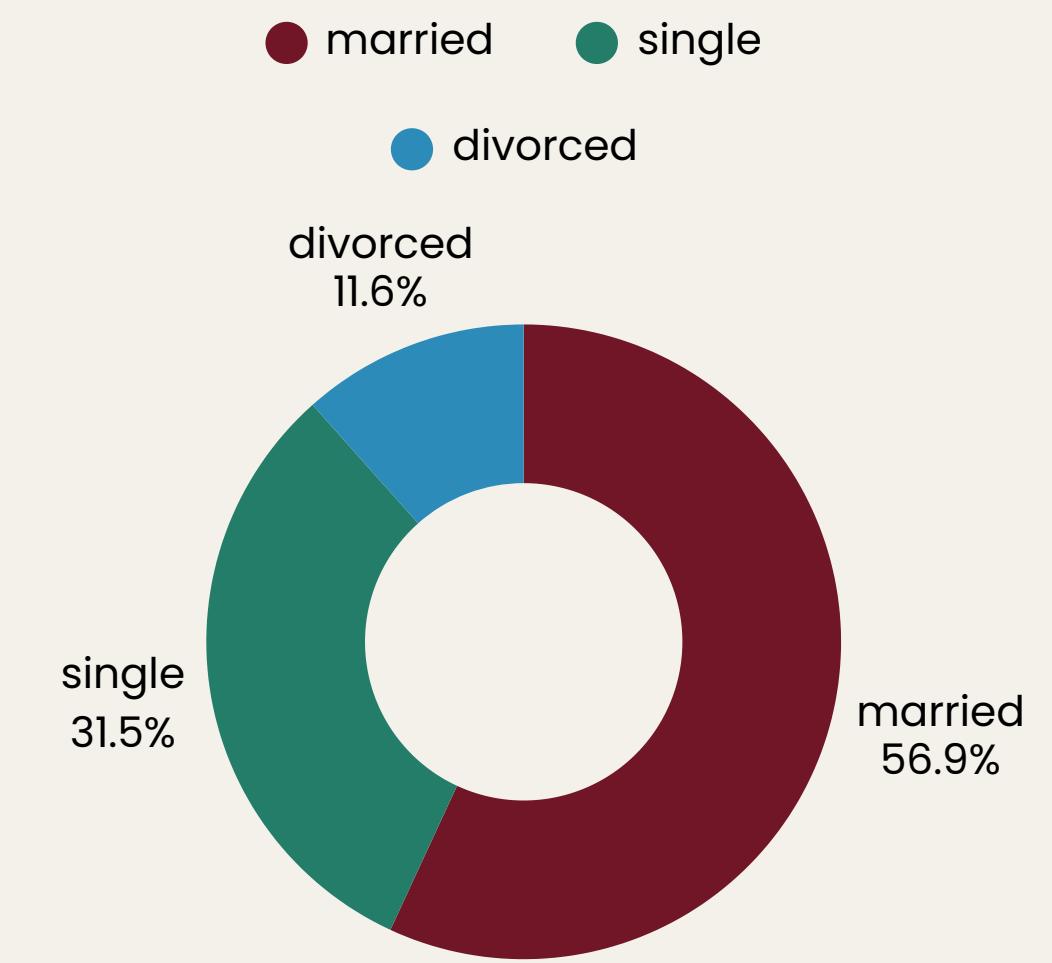
**41**  
**RATA-RATA UMUR  
NASABAH**

**5.906** | **5.256**  
**PUNYA  
PINJAMAN** | **TIDAK  
PUNYA**

## JOB DISTRIBUTION



## MARTIAL STATUS



# Feature engineering

Original data

Formula

Data Add-on

**month**

**Mapping from Month to number, e.g ='jan':1,**

**'month\_num'**

**previous**

**pday**

**Previous / pdy (if pdy >0)**

**previous\_per\_pdays**

**Age**

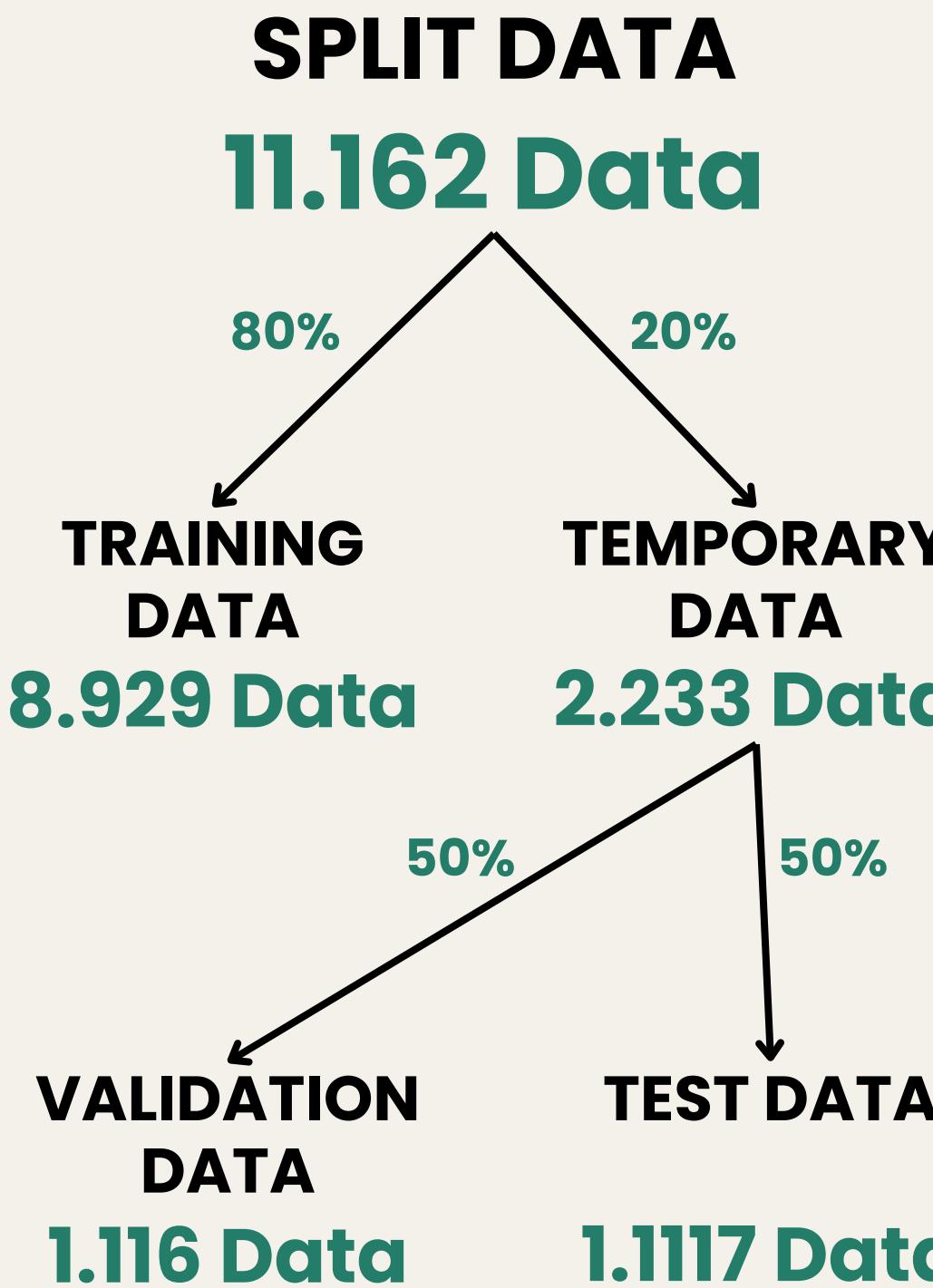
**bins Age ('<25', '25-35', '35-45', '45-55', '55-65', '65+')**

**age\_group**

# Machine Learning

1. Modelling data
2. Pipeline Machine Learning
3. Validasi Model
4. Evaluasi Model

# Modelling data



## MODEL-MODEL YANG AKAN DICOBAA

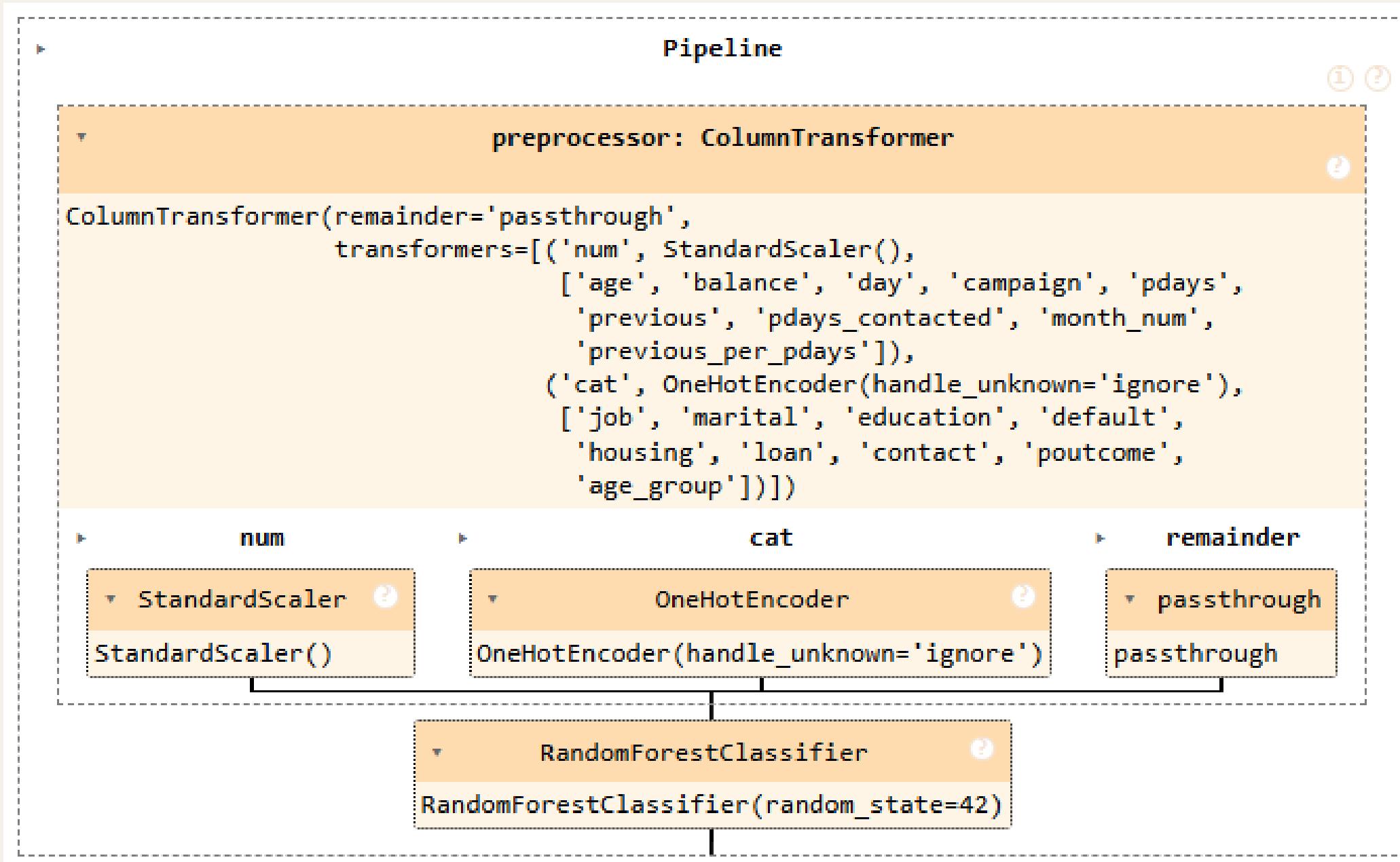
Menggunakan cross validation untuk mencari recall tertinggi

- Logistic Regression
  - Recall = 60% Accuracy = 69%
- Decision Tree
  - Recall = 61% Accuracy = 63%
- Random Forest
  - **Recall = 63% Accuracy = 71%**
- Gradient Boosting
  - Recall = 62% Accuracy = 73%
- Support Vector Machine
  - Recall = 58% Accuracy = 70%
- Naive Bayes
  - Recall = 49% Accuracy = 65%

Model yang  
diambil  
Random forest

# Pipeline ML

## 2 Komponen utama pipeline ini:



### 1. Preprocessing

- Transformasi kolom menjadi numerik (contoh age, balance dsb)
- Transformasi kolom menjadi category (contoh job, marital status, dsb)
- Atribut selain diatas akan tetap di passthrough

### 2. Model

- Model klasifikasi yang digunakan adalah Random Forest

# Validasi Model

## 1. Optimalisasi Recall Hyperparameter Tuning

- Recall score before: 63,2%
- Recall score after: 63,3%

## 2. Classification Report

- Klasifikasi Decision “**No**”

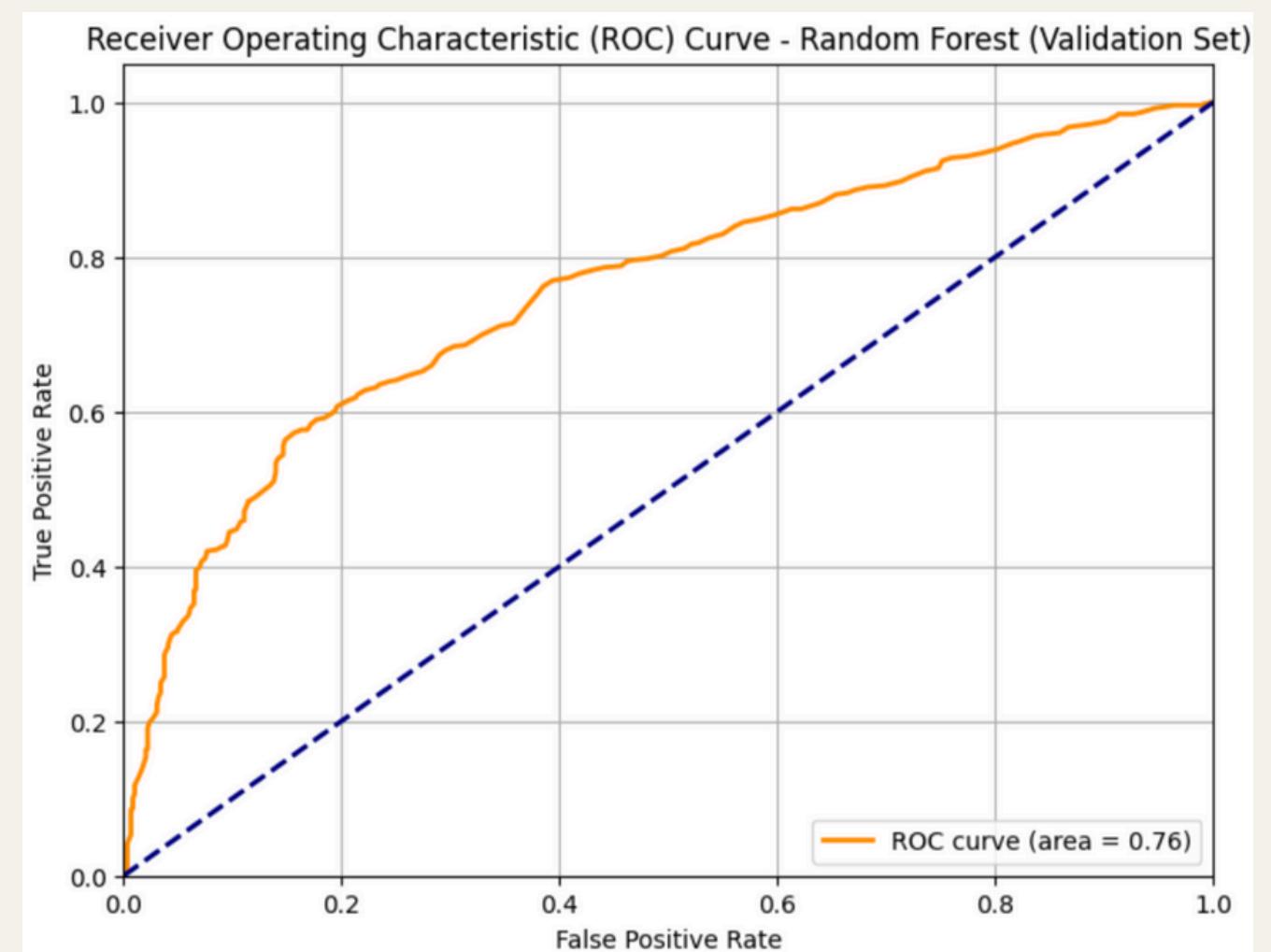
Model lebih baik dalam mendeteksi kelas “no”. Recall yang tinggi (0.79), alaupun precision-nya sedikit lebih rendah (70%)

- Klasifikasi Decision “**Yes**”

recall lebih rendah (**61%**) yang berarti ada potensi nasabah yang bisa terlewat. Sedangkan Precision tinggi 73% menunjukan model bisa memprediksi “**yes**”

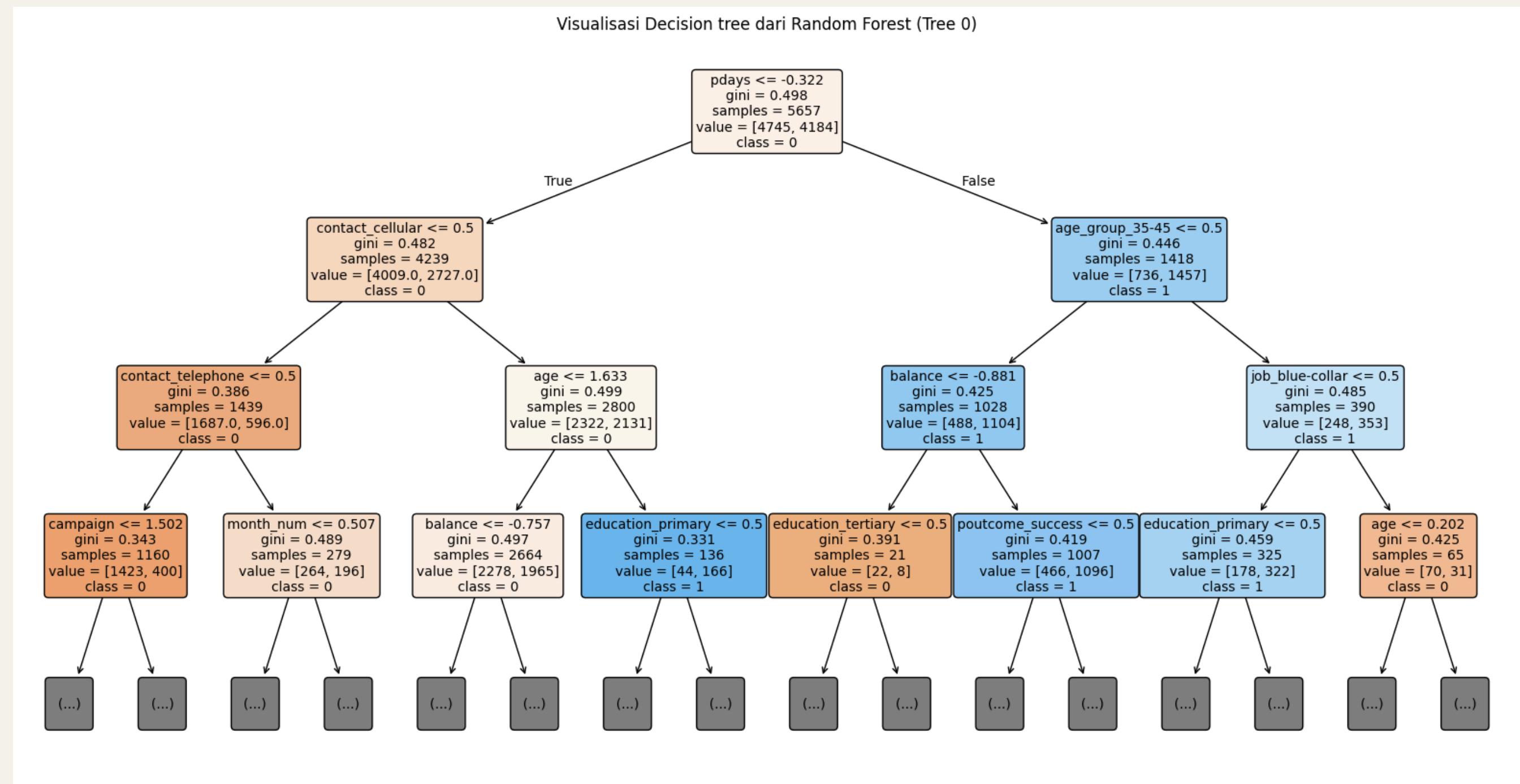
## 3. ROC Curve

**76% peluang model** dapat membedakan antara kelas positif dan negatif dengan benar



# Validasi Model

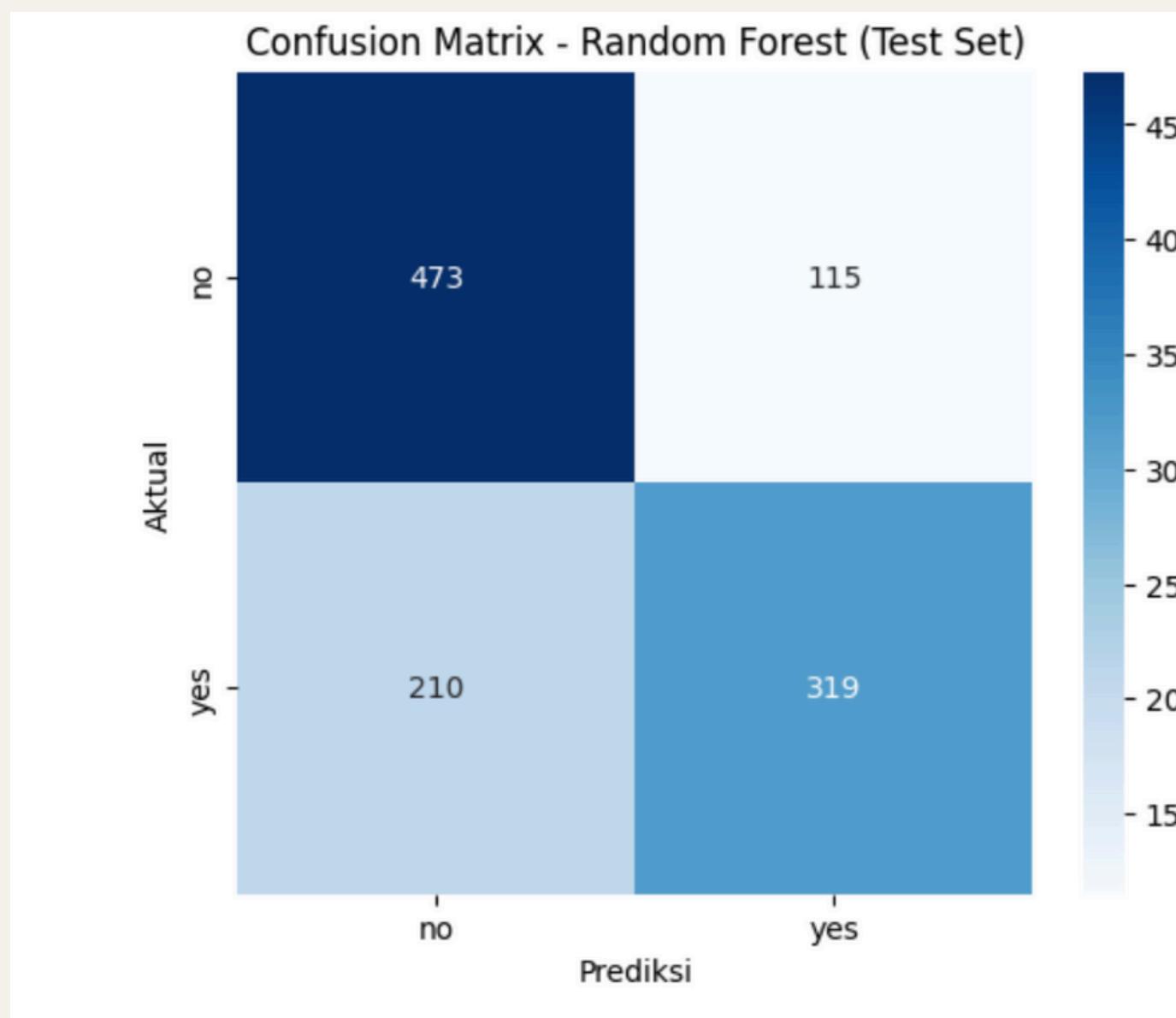
## Model Random Forest



# Evaluasi Model

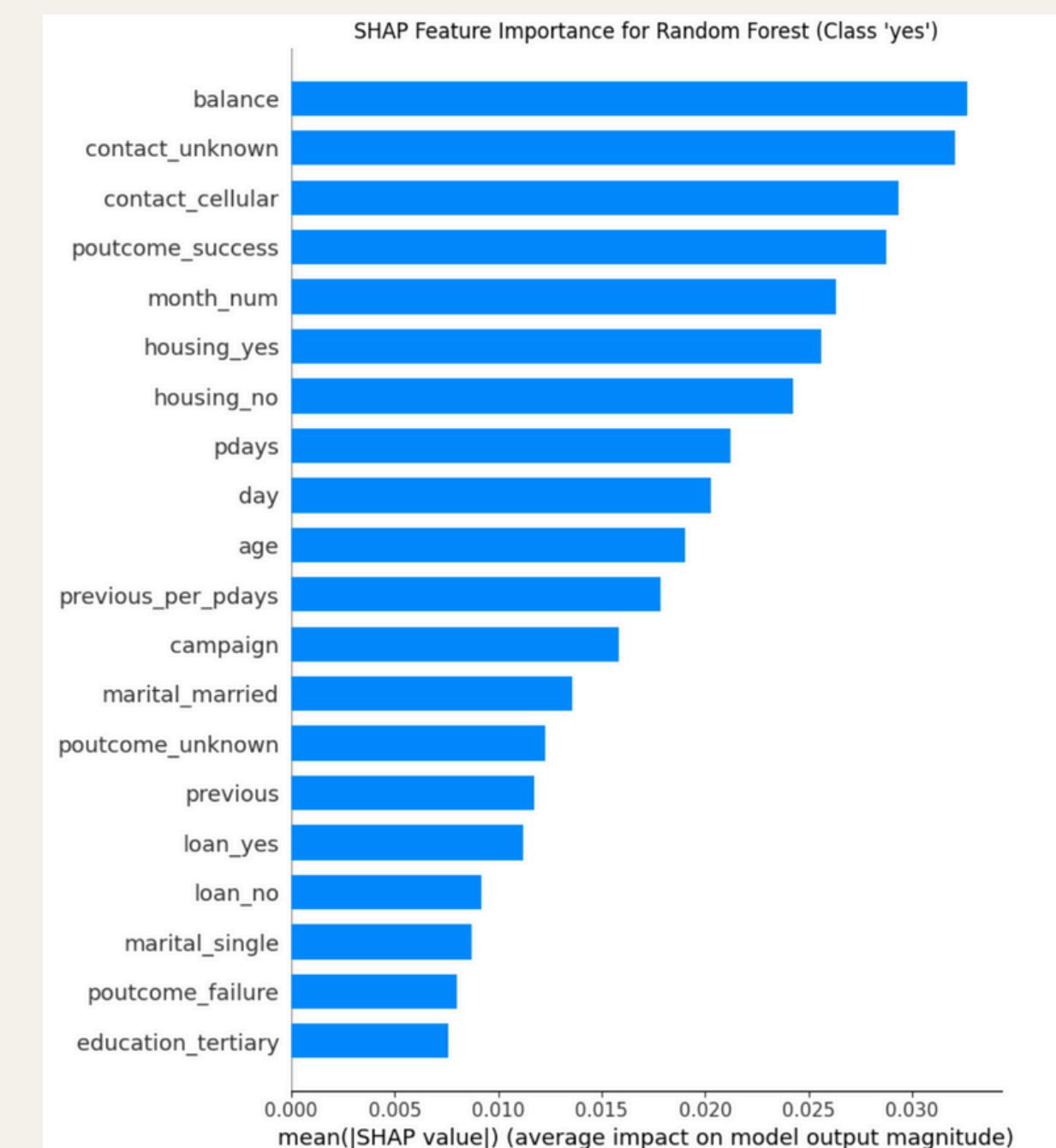
## 1. Confusion Matrix

Model bisa benar memprediksi 473 data "NO" dan 319 data "YES"



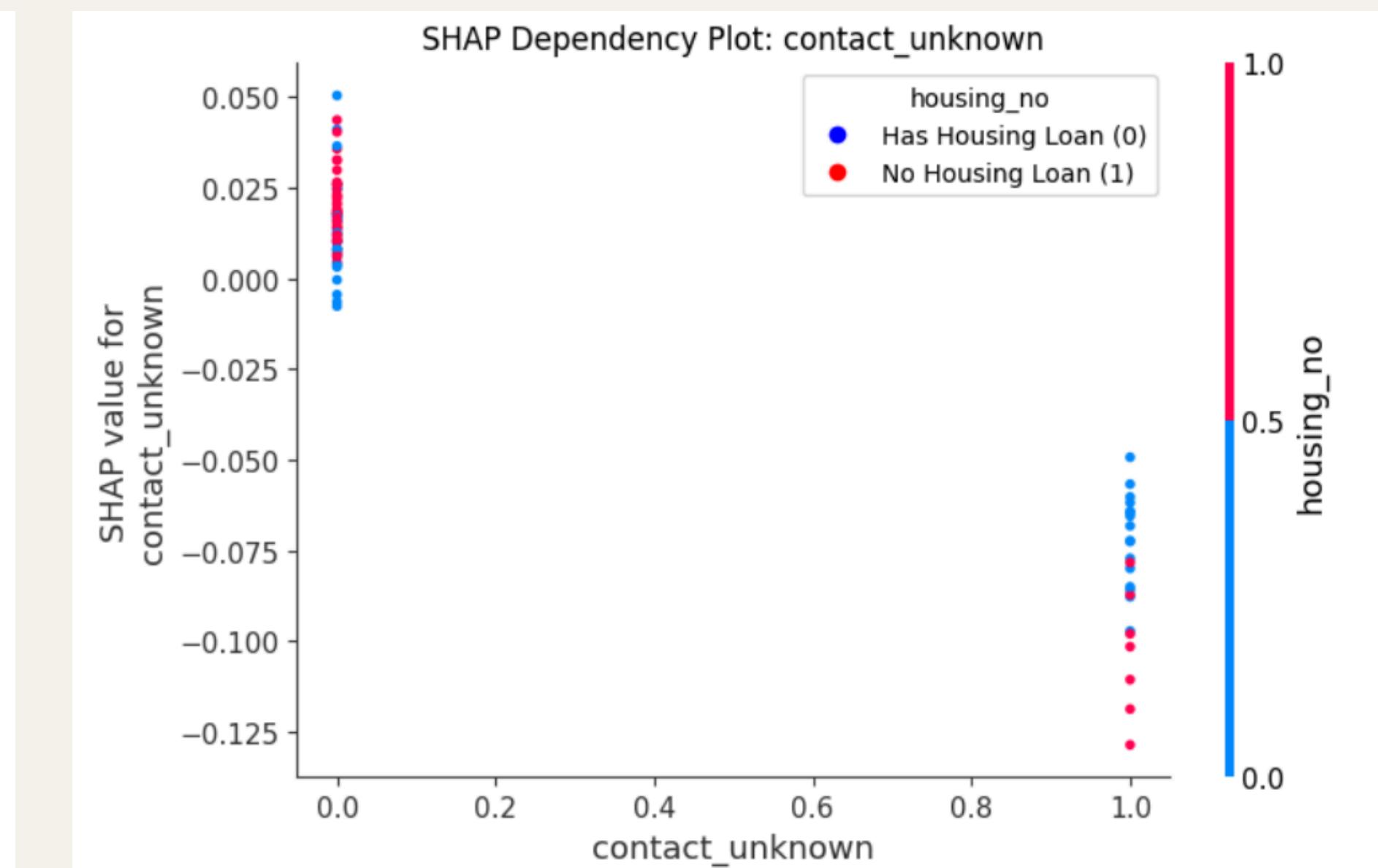
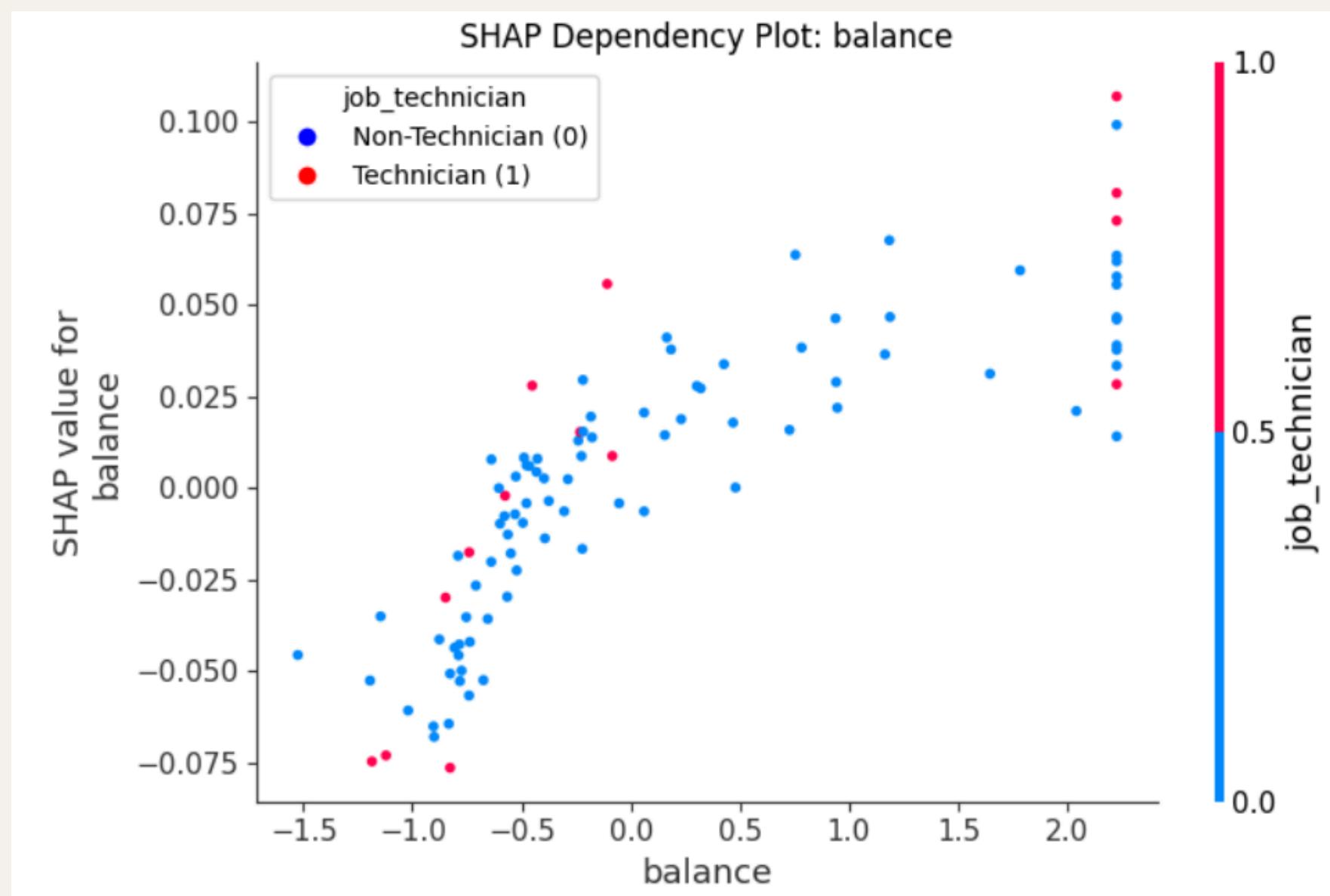
## 2. SHAP

Fitur apa yang mempengaruhi hasil prediksi



# Evaluasi Model

**2 fitur terbesar dari SHAP yaitu Balance dan contact unknown**



Semakin tinggi balance seseorang, probabilitas untuk melakukan deposito makin tinggi

data yang lengkap mempengaruhi probabilitas untuk melakukan deposito

# Evaluasi Model

## 3. LIME

Menjelaskan prediksi model machine learning untuk satu observasi (1 baris data)

Model ini dapat memprediksi 60% "NO" atau tidak melakukan deposito dengan berbagai faktor yang mendukung



# Implikasi Bisnis

## Implikasi Bisnis Strategis:

- Fokus Pemasaran: Bank harus memprioritaskan upaya pemasaran pada klien yang:
- Telah dihubungi melalui telepon seluler.
- Memiliki riwayat keberhasilan dalam kampanye sebelumnya.
- Lebih tua dan memiliki saldo tinggi.

## The Bank perlu berhati-hati dengan nasabah yang:

- Memiliki riwayat kegagalan dalam kampanye sebelumnya.
- Memiliki riwayat yang tidak diketahui (mungkin perlu riset lebih lanjut atau pendekatan berbeda).
- Terlalu sering dihubungi dalam satu kampanye.
- Memiliki pinjaman perumahan.





# Thank you

