



FINAL PROJECT

PREDICTIVE APPLICATION FOR BANK X CUSTOMERS FOR TERM-DEPOSIT PRODUCTS WITH CONSIDERATION OF BUSINESS OBJECTIVES

Exploratory Data Analysis

Machine Learning

Classification

Optimization

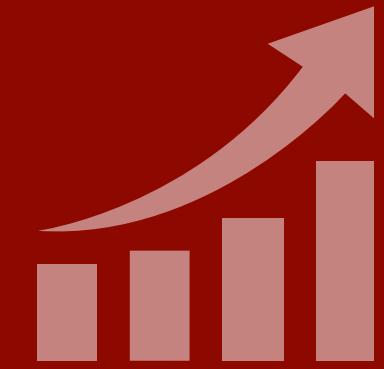
Deployment

CONTENT LIST



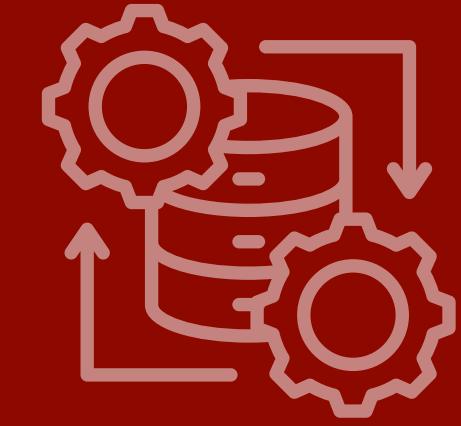
STAGE 0

Project
Introduction



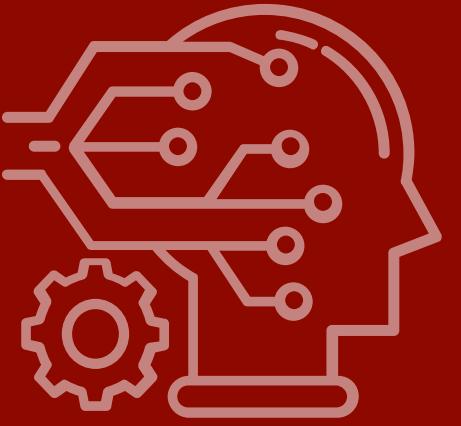
STAGE 1

EDA &
Visualization



STAGE 2

Data
Preprocessing



STAGE 3

Modeling &
Evaluation



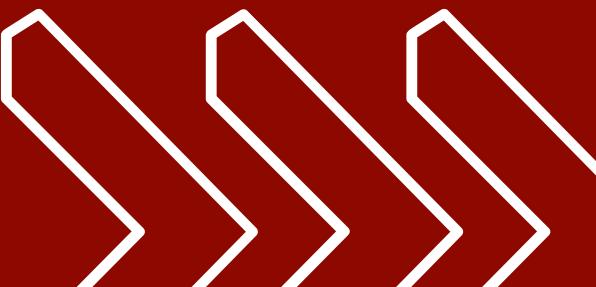
STAGE 4

Differentiation

I Problem Statement



**Bagaimana Cara Optimasi
Campaign Marketing
berdasarkan Data
Historis?**





I Solution

STAGE 0 PROJECT INTRODUCTION

>> WHY MACHINE LEARNING ?



Hou, Sipu & Cai, Zongzhen & Wu, Jiming & Du, Hongwei & Xie, Peng. (2022). **Applying Machine Learning to the Development of Prediction Models for Bank Deposit Subscription.** *International Journal of Business Analytics.* 9. 1-12. 10.4018/IJBAN.288514.

Ahmad Ilham, Laelatul Khikmah, Indra, Ulumuddin, Ida Bagus Ary, & Indra Iswara. (2019). **Long-term deposits prediction: a comparative framework of classification model for predict the success of bank telemarketing.**

Borugadda, Premkumar & Nandru, Prabhakar & Madhavaiah, C.. (2021). **Predicting the Success of Bank Telemarketing for Selling Long-term Deposits: An Application of Machine Learning Algorithms.** 7. 91-108.

Log In or Register Now

I Goal & Objectives

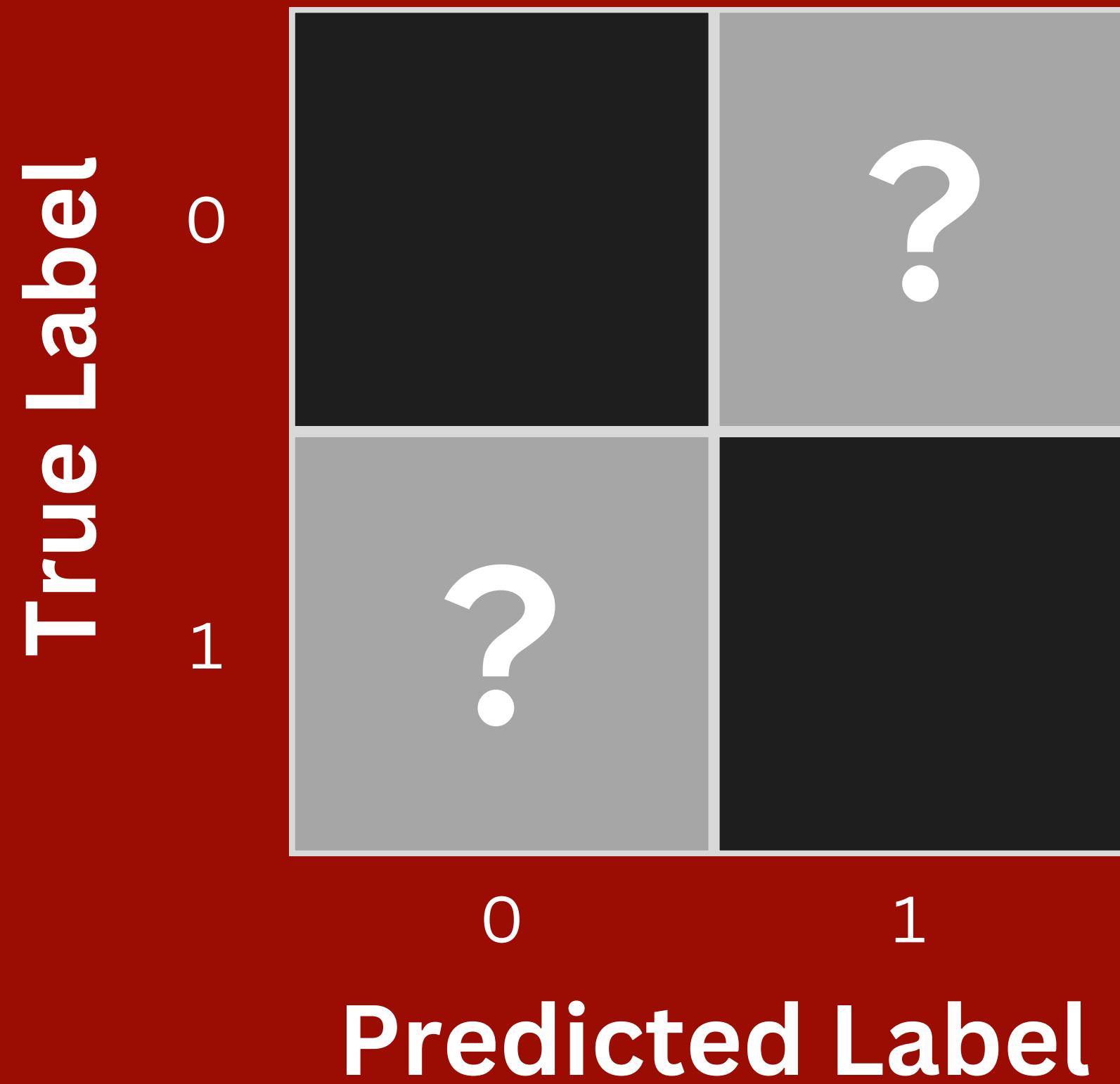
Goal

Optimasi Campaign Marketing melalui Pembuatan Model/Aplikasi Prediktif dengan Pertimbangan Bisnis (potencial profit & operational campaign)

Objectives

- 01 Melakukan **Decsriptive Analytics** untuk memperoleh informasi apa saja yang telah terjadi, hidden patterns, ataupun rekomendasi bisnis.
- 02 Melakukan **Diagnostic Analytics** untuk Mengetahui Mengapa sesuatu Terjadi.
- 03 Melakukan **Predictive Analytics** (Predictive Modelling) untuk Memperoleh Model Prediksi Terbaik.
- 04 Melakukan **Prescriptive Analytics** untuk Menyarankan Pilihan Keputusan.

| Business Metrics

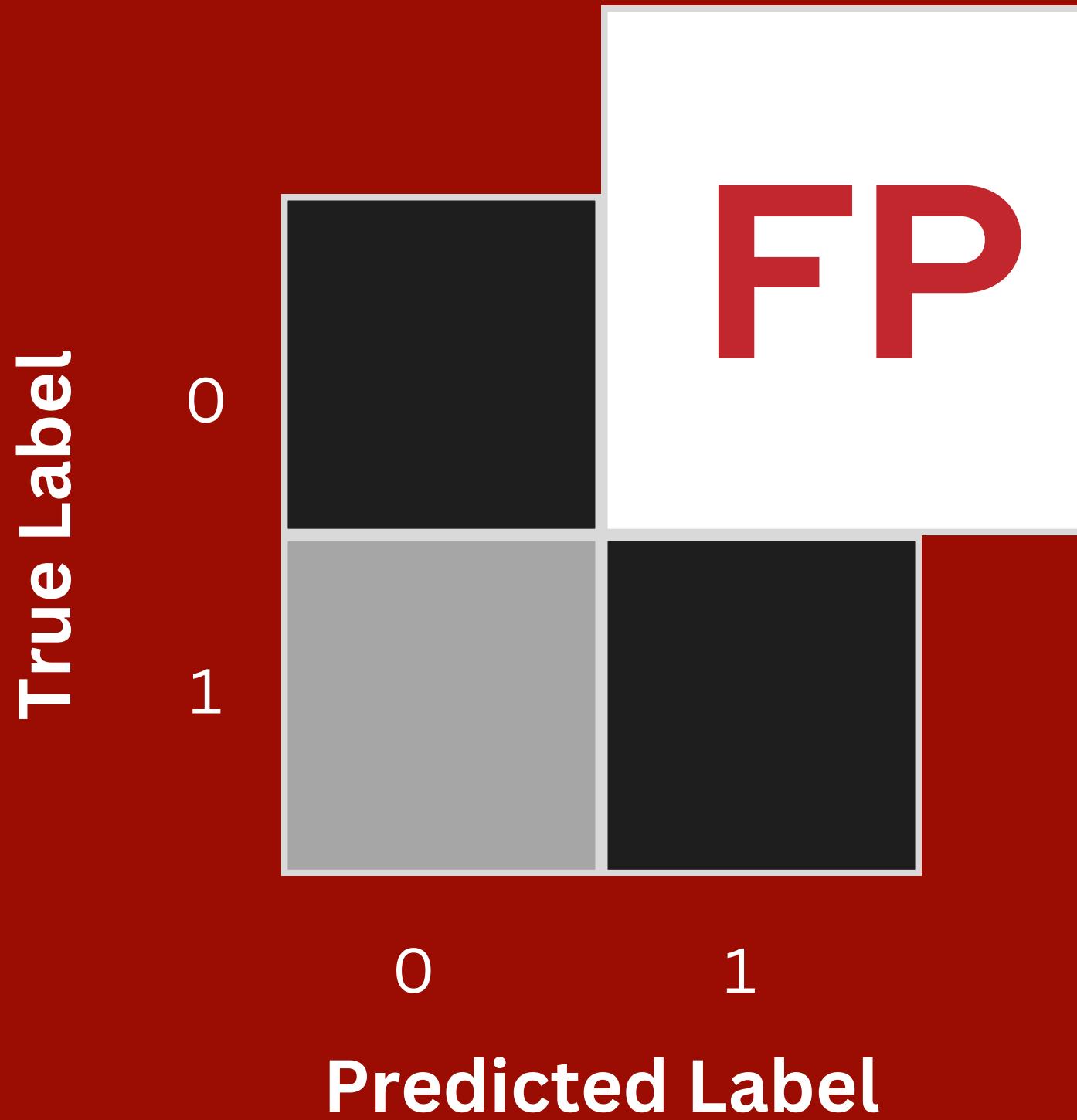


IMPACT

for business



| Business Metrics, False Positive



Treatment

Bank mengirim campaign.

Expect

Nasabah berlangganan term deposit.

Reality

Nasabah tidak berlangganan term deposit, walaupun menerima campaign.

Consequence

Terjadi *loss budget* / anggaran terbuang : ***operational campaign***.

I Business Metrics, False Negative

Treatment

Bank TIDAK mengirim campaign.

Expect

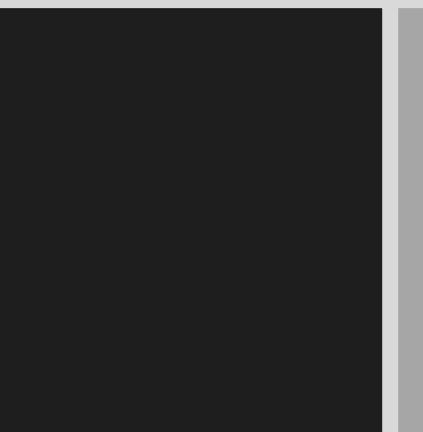
Nasabah TIDAK berlangganan term deposit,
walaupun diberi campaign.

Reality

Nasabah akan berlangganan term deposit,
jika MENERIMA campaign.

Consequence

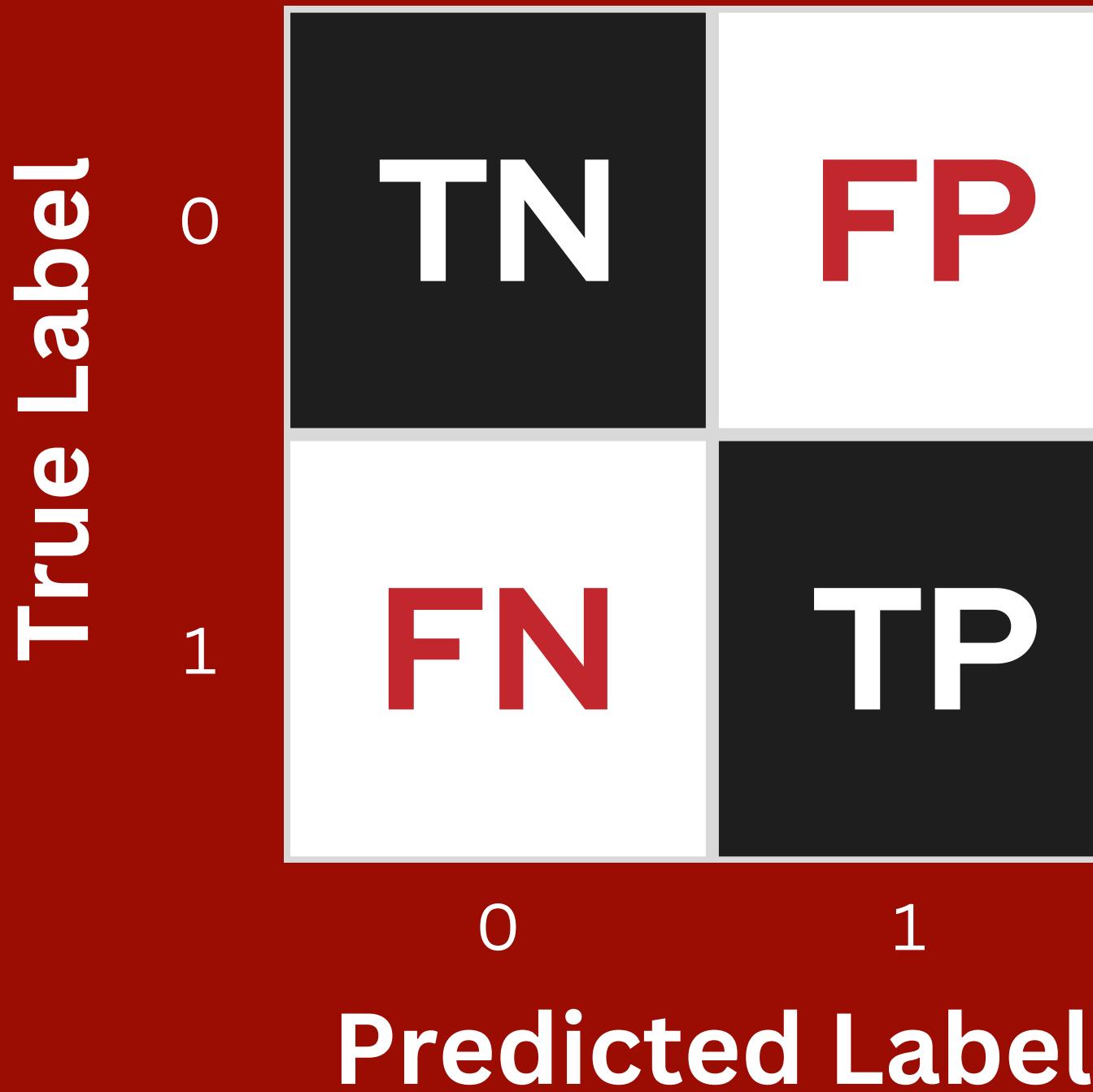
Gagal peroleh nasabah berlanggan term deposit:
Adanya potensi loss ***potencial profit***.

		Predicted Label	
		0	1
True Label	0		
	1		

FN



I Business Metrics, Scenario Option



Minimasi Operational Campaign

Minimasi kondisi FP --> **Precision**

Minimasi marketing campaign --> **operational_campaign**

Maksimasi Potencial Profit

Minimasi kondisi FN --> **Recall**

Maksimasi TP --> # gain **potencial_profit**

Optimasi Operational Campaign & Potencial Profit

Minimasi kondisi FP & FN --> **F1, ROC-AUC**

Minimasi marketing campaign --> **operational_campaign**

Maksimasi TP --> # gain **potencial_profit**

I Roles



Muhammad Iqbal A.
Data Manager Lead



Tsabitah Karimah
Data Scientist



Liliani Priestiwanty
Data Scientist



Cholifatur Rohman
Data Scientist



Khoshid Mahendra
Data Analyst



Hafifah Amalia
Data Analyst



Roffe Rizqullah D.
Data Analyst



Arieska Restu
Data Analyst

CONTENT LIST



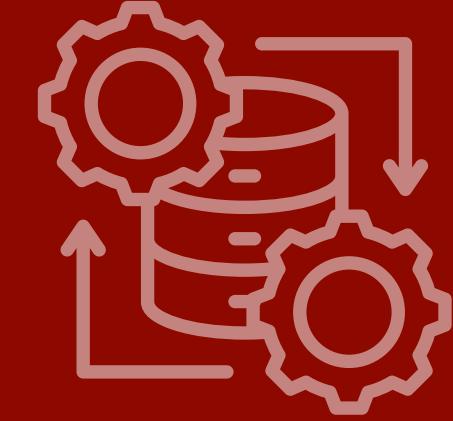
STAGE 0

Project
Introduction



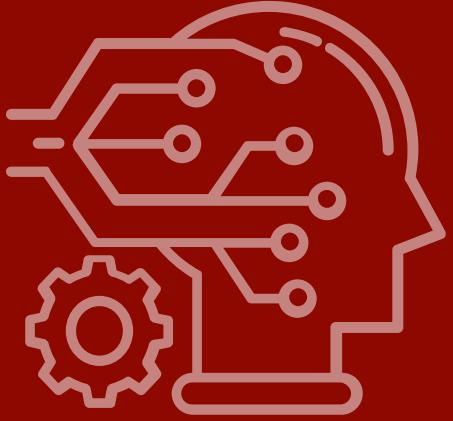
STAGE 1

EDA &
Visualization



STAGE 2

Data
Preprocessing



STAGE 3

Modeling &
Evaluation

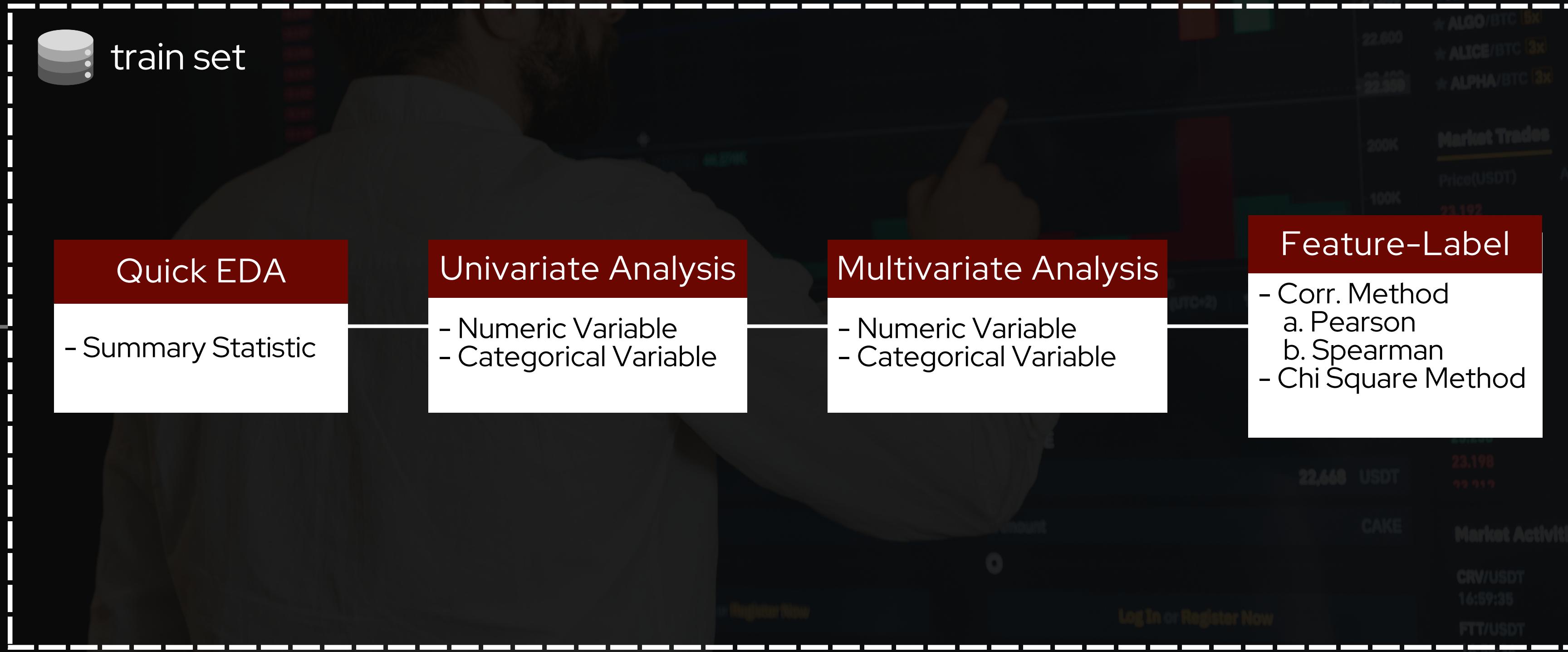


STAGE 4

Differentiation

| Flowchart

STAGE 1 EDA & VISUALIZATION



| Quick EDA

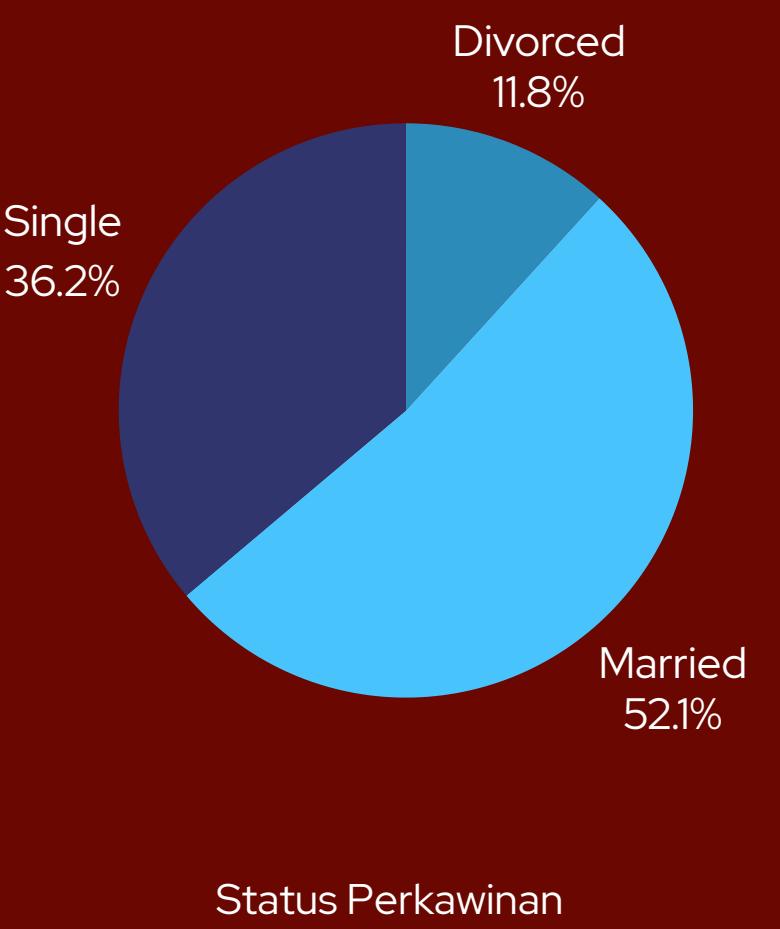
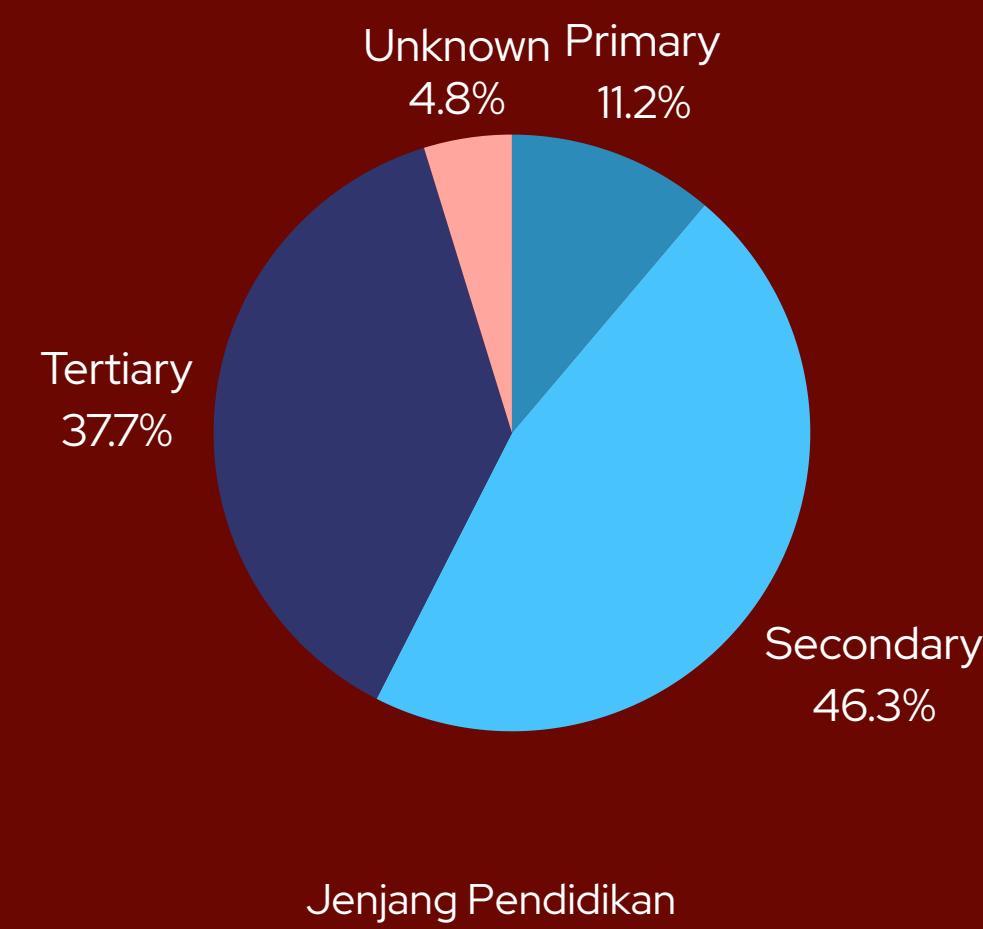
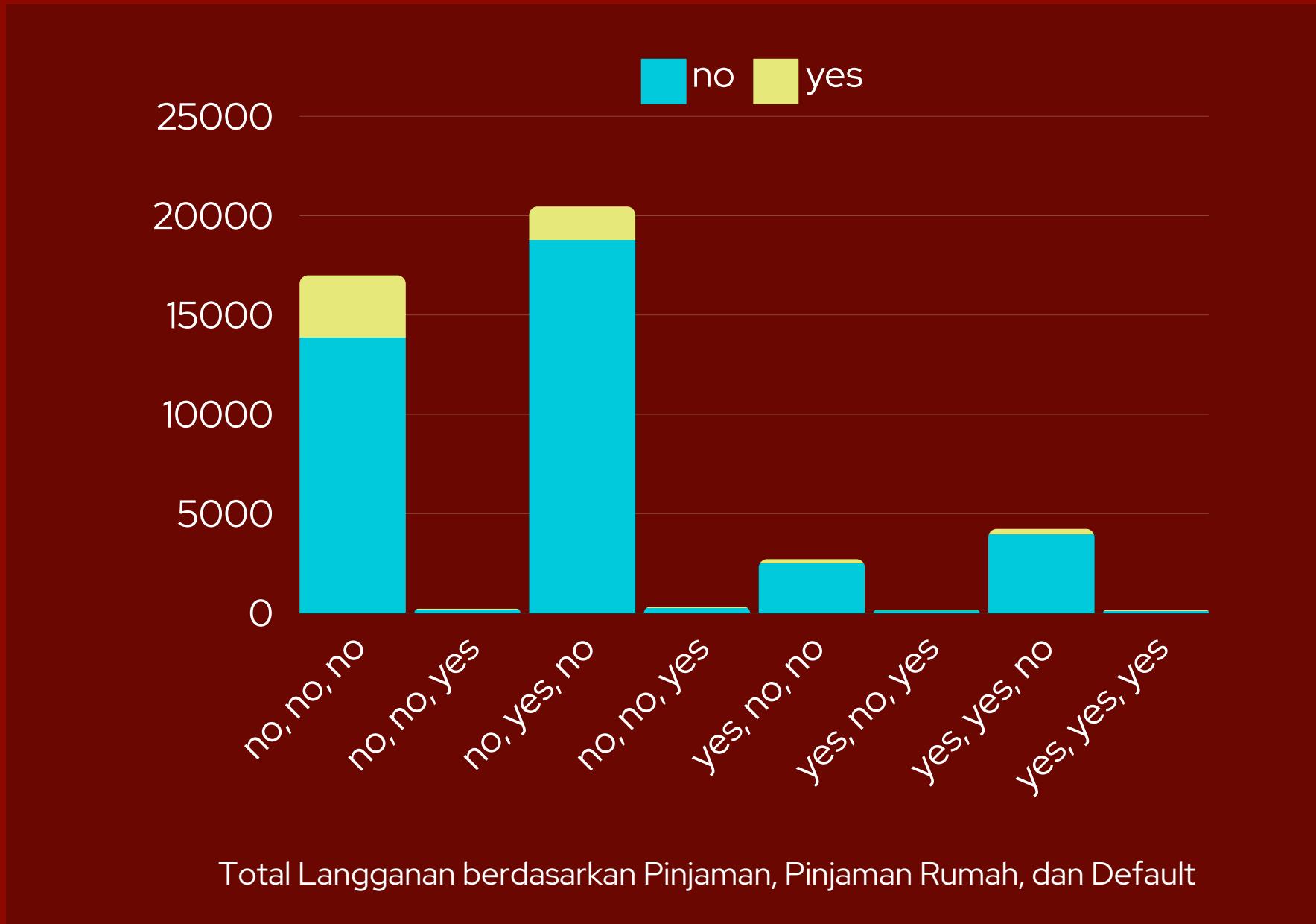
Dataset

Kolom	Jumlah Data	Tipe Data
age	45211	int64
job	45211	object
marital	45211	object
education	45211	object
default	45211	object
balance	45211	int46
housing	45211	object
loan	45211	object
contact	45211	object
day	45211	int46
month	45211	object
duration	45211	int46
campaign	45211	int46
pdays	45211	int46
previous	45211	int46
poutcome	45211	object
y	45211	object

Fitur ←

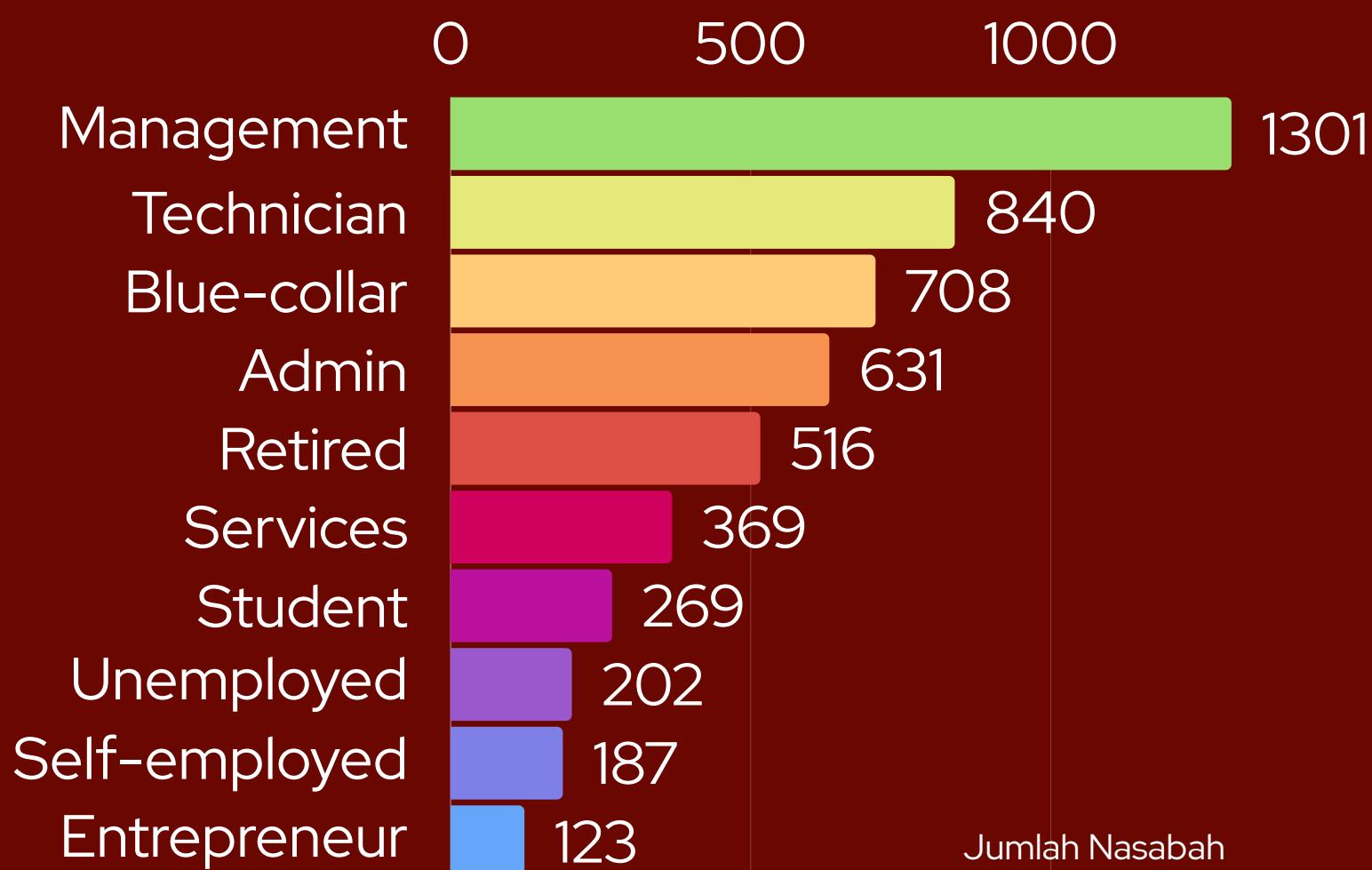
← **Target**

I Persona Nasabah Berlangganan

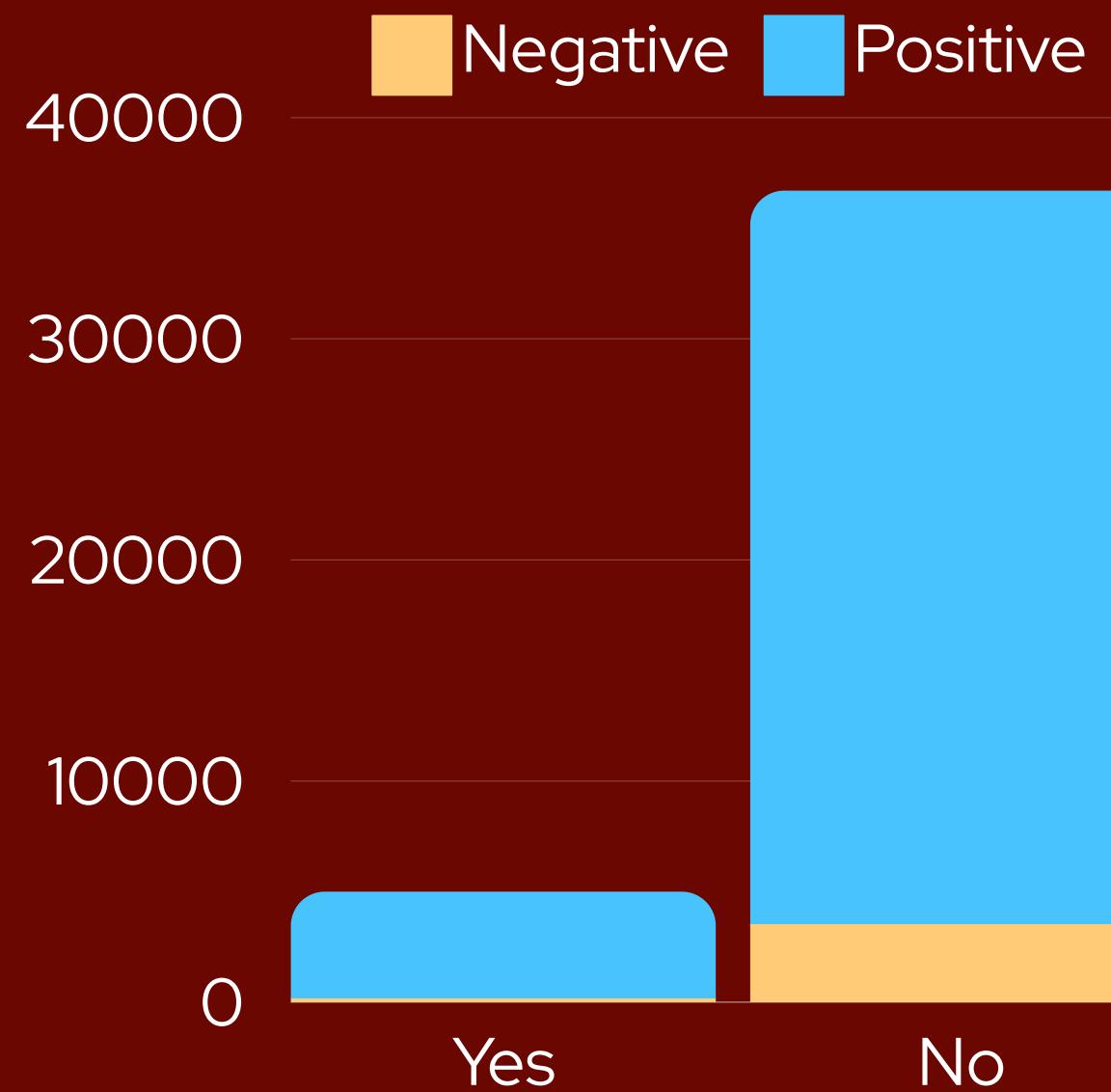


I Persona Nasabah Berlangganan

Jumlah Nasabah untuk Setiap Pekerjaan

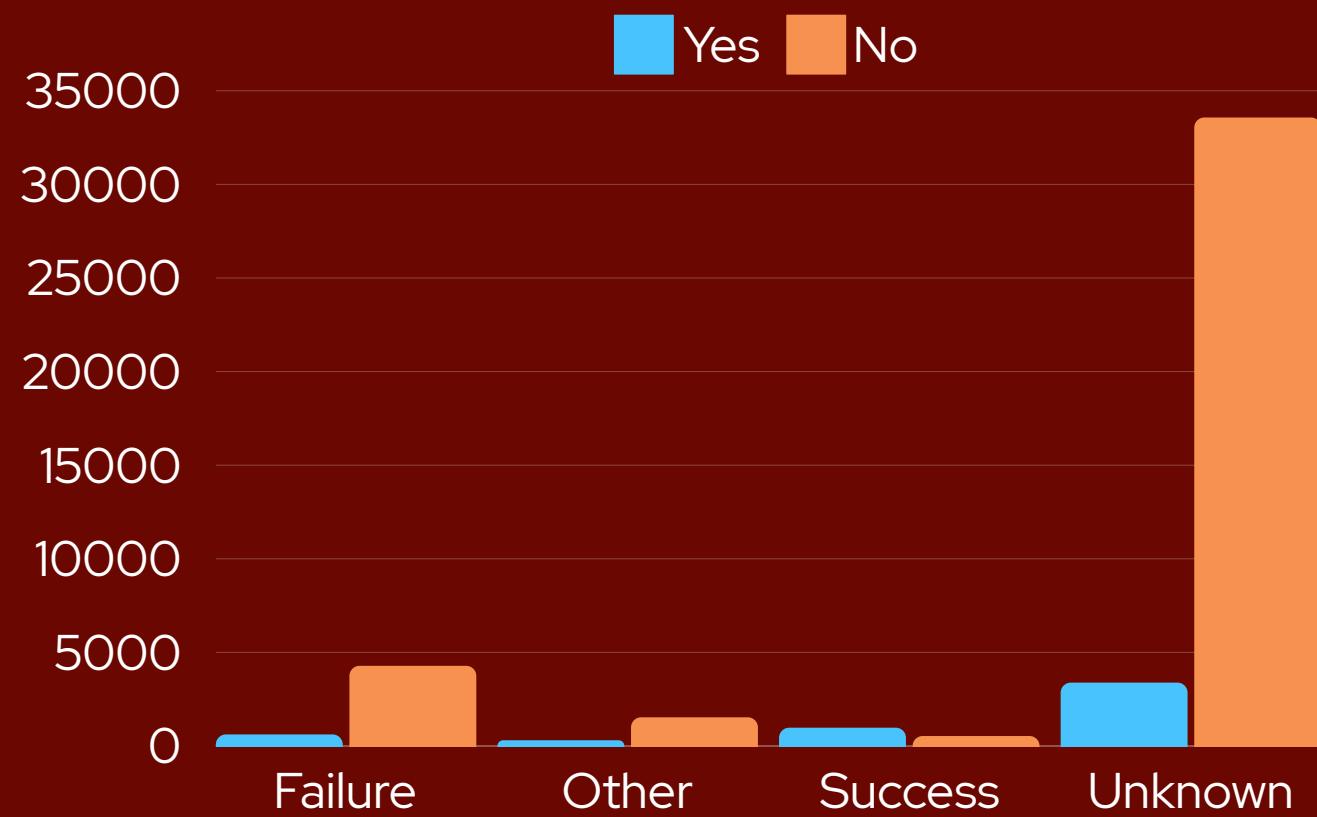


Distribusi Balance Terhadap Target Y



I Persona Nasabah Berlangganan

Status Subscription berdasarkan
Previous Campaign Outcome



I Uji Chi Square

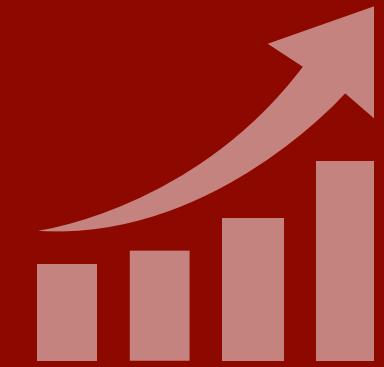
FITUR	Nilai Chi-Square	Nilai P-Value	Tidak Berpengaruh Signifikan	Berpengaruh Signifikan
Job & Education	29699.78	0.0		
Loan & Housing	280.27	1.85		
Job & Loan	1376.01	3.01		
Job & P-Outcome	1180.42	4.89		
Job & Contact	3005.24	0.0		

CONTENT LIST



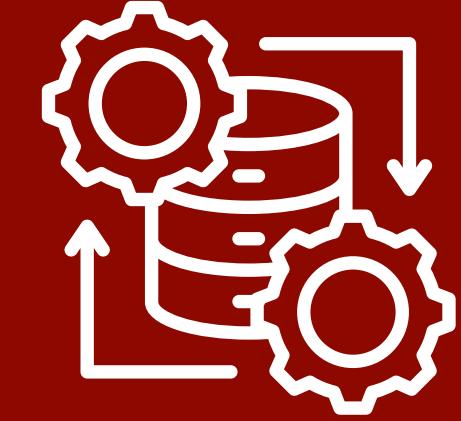
STAGE 0

Project
Introduction



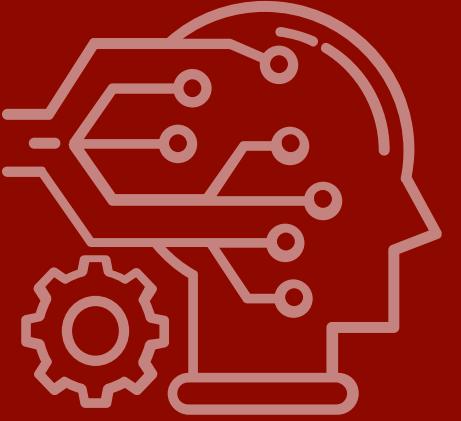
STAGE 1

EDA &
Visualization



STAGE 2

Data
Preprocessing



STAGE 3

Modeling &
Evaluation

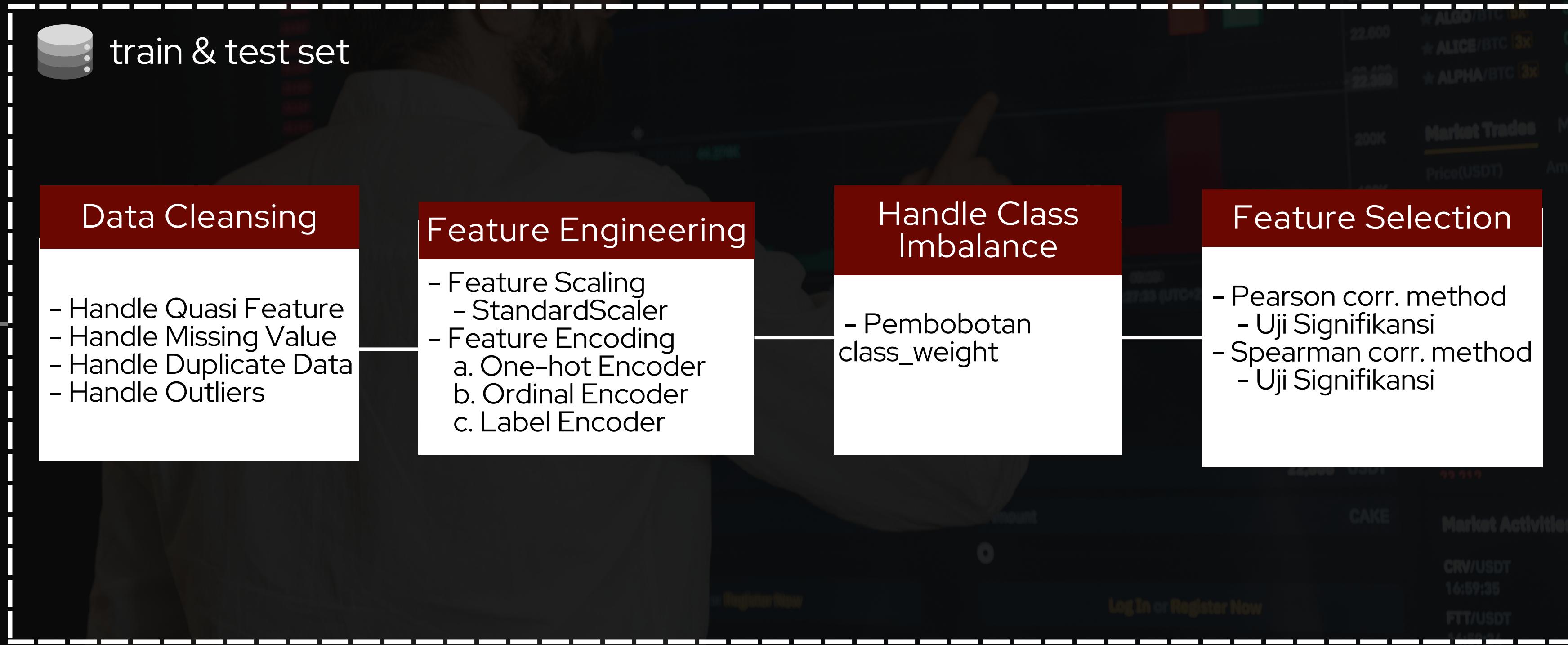


STAGE 4

Differentiation
Stage

| Flowchart

STAGE 2 DATA PREPROCESSING



I Data Pre-Processing

Tidak ada missing value untuk semua variabel

Missing Value Handling

Penggunaan StandardScaler diterapkan untuk train data dan test data

Feature Scaling

Tidak terdapat baris yang mengalami duplicate data

Duplicate Handling

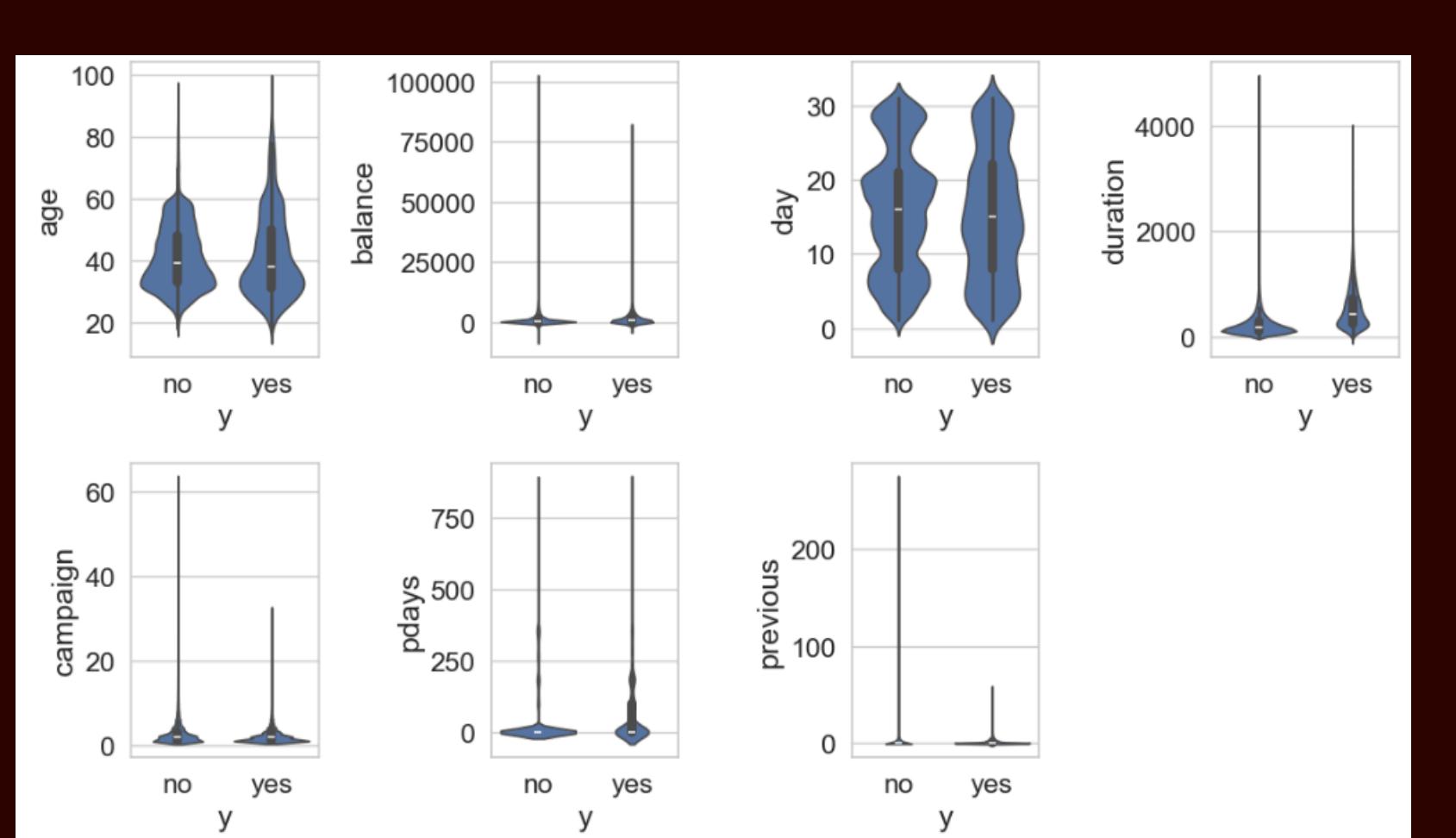


PDSI

PERJUANGAN

| Outliers Handling

STAGE 2 DATA PREPROCESSING



EXPLANATION

- Fitur 'balance', 'duration', 'campaign', 'pdays', 'previous' terlihat jelas memiliki outlier.
- Tidak dilakukan penghapusan outlier untuk menghindari perubahan distribusi data dan mencegah informasi yang hilang

I Feature Encoding

One-hot Encoding

Job, Marital,
Contact, Month,
Poutcome

Ordinal Encoding

Education, Default,
Housing, Loan

Label Encoding

Target Y

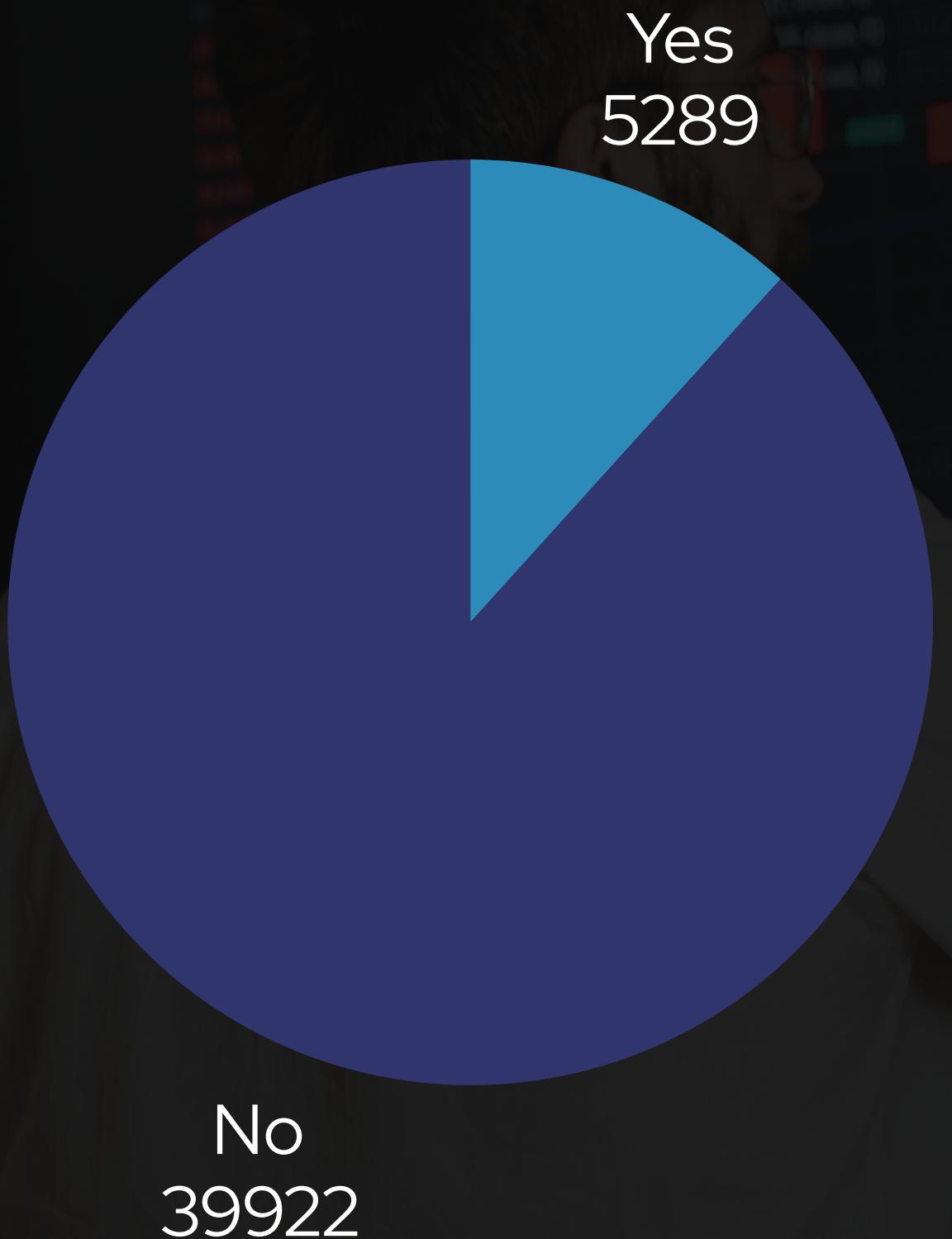
“

- Terdapat dua data yaitu train data & test data
- Pada tahap encoding, dilakukan perubahan pada kedua train data & test data

”

| Class Imbalance Handling

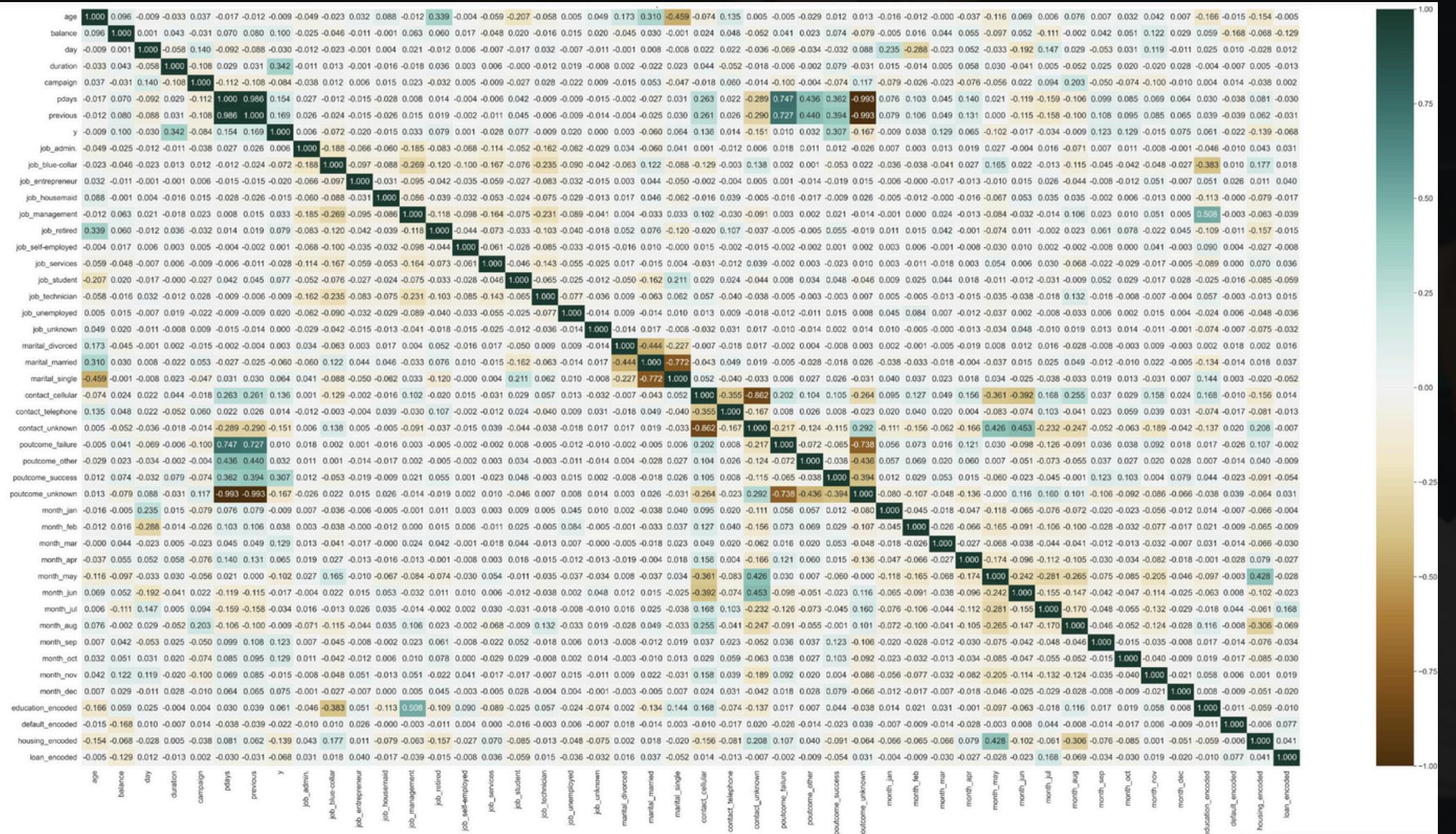
STAGE 2 DATA PREPROCESSING



EXPLANATION

Untuk meng-handle kasus ini, kami akan 'mengatur' ulang bobot atau lebih dikenal `class_weight`.

| Feature Selection, Spearman Method [NL]

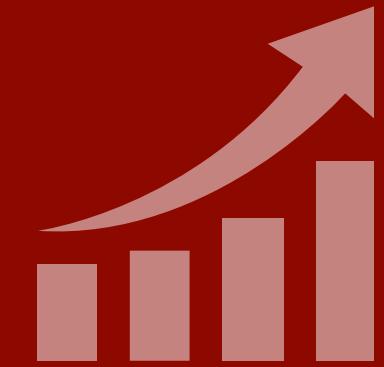


CONTENT LIST



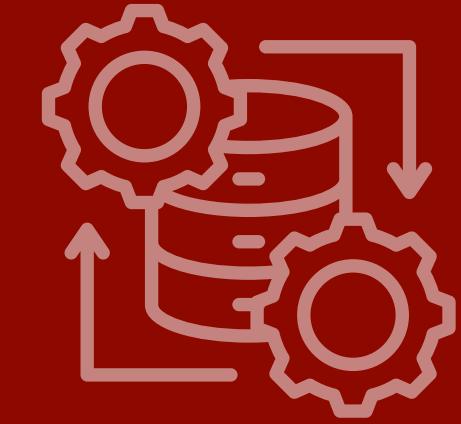
STAGE 0

Project
Introduction



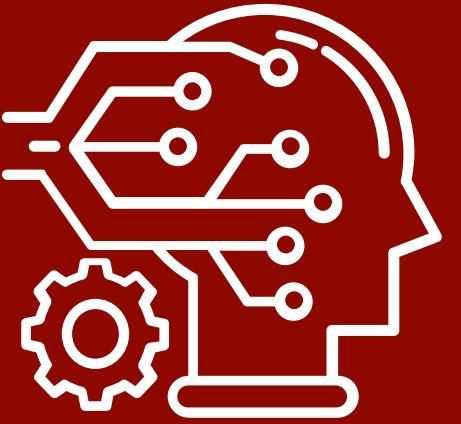
STAGE 1

EDA &
Visualization



STAGE 2

Data
Preprocessing



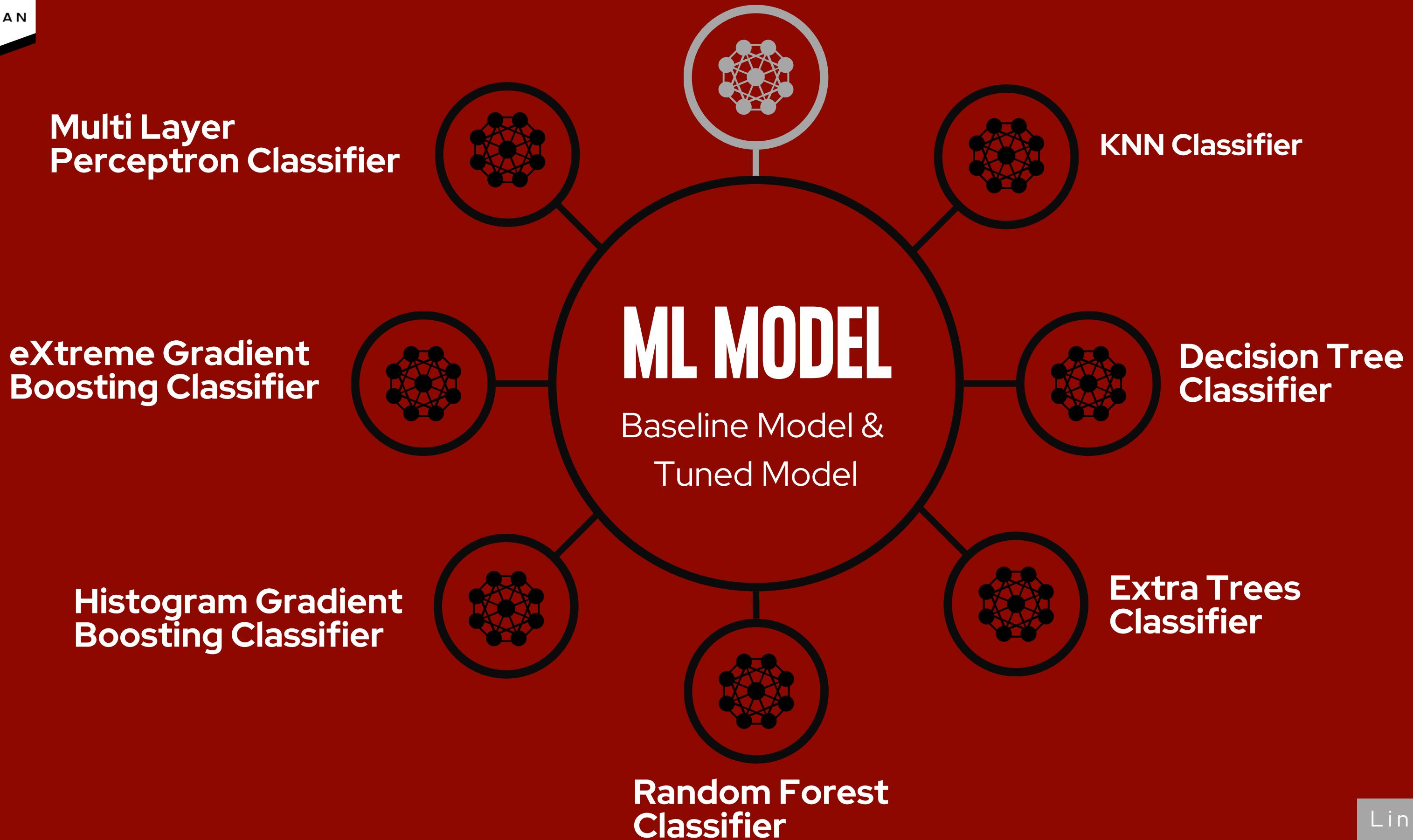
STAGE 3

Modeling &
Evaluation



STAGE 4

Differentiation



I Modeling Process

Modeling berbagai Algoritma

Secara general, model yang digunakan terbagi menjadi 2 yaitu Linear [L] dan Non-linear [NL], secara default.



Tuning Hyperparameter

Melakukan tuning hyperparameter untuk setiap jenis algoritma model



Komparasi semua model

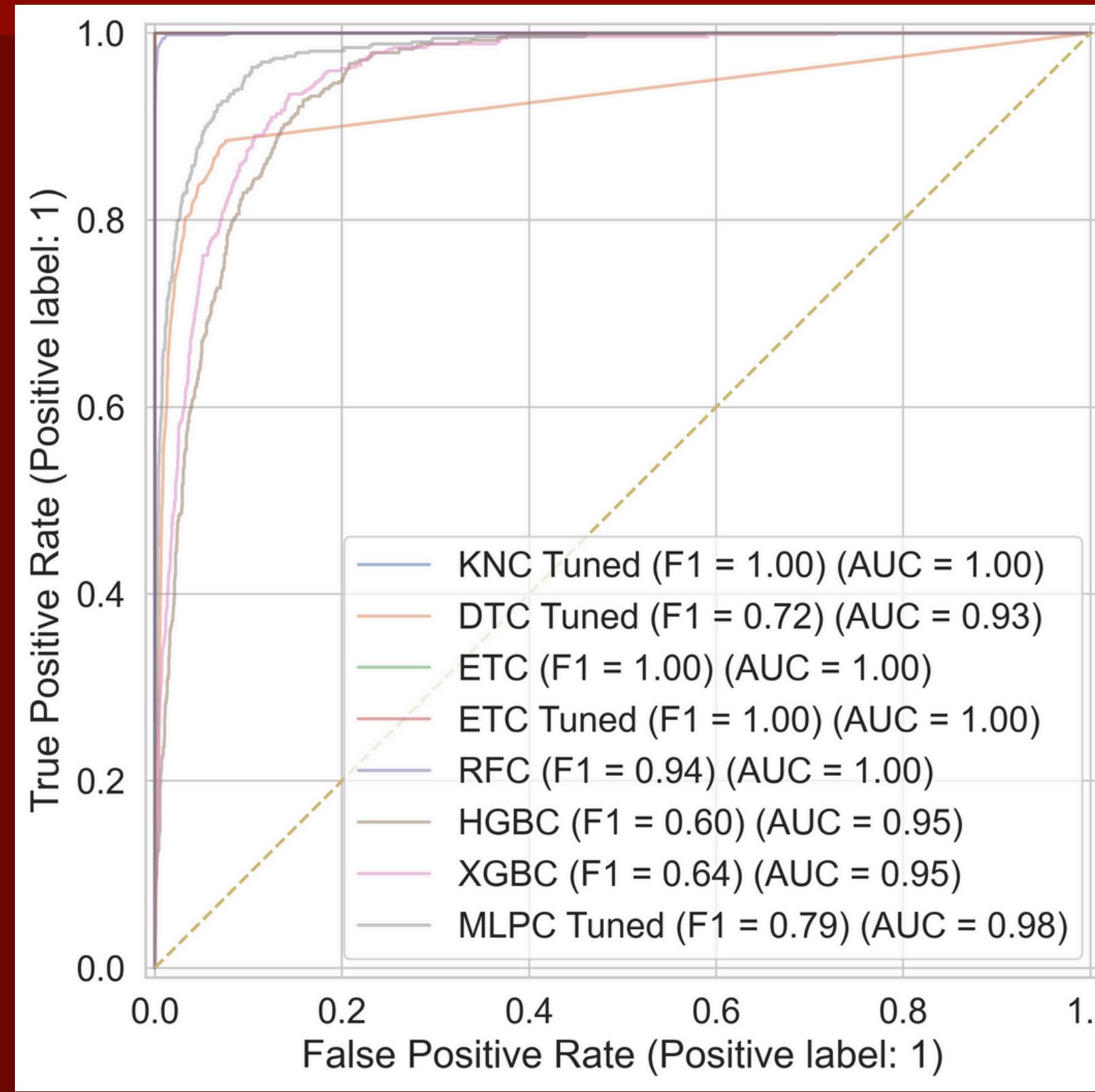
Membandingkan dan memilih model terbaik, dari setiap algoritma model dengan hyperparameter terbaik.



Komparasi Baseline vs Tuned

Membandingkan antara hasil baseline & tuned model serta memilih mana yang terbaik.

I ROC Curves



MODEL COMPARISON

- Diperoleh ada 3 model dengan nilai AUC = 1.00 yaitu KNC CV, ETC, ETC CV, RFC, dan RF CV.

Best Model	F1 Train	F1 Test	Gap F1	ROC-AUC Train	ROC-AUC Test	Gap ROC-AUC
KNC, Tuned Model	1.000	1.000	0.000	1.000	1.000	0.000
ETC, Baseline Model	1.000	1.000	0.000	1.000	1.000	0.000
ETC, Tuned Model	1.000	1.000	0.000	1.000	1.000	0.000
RFC, Baseline Model	1.000	0.940	0.060	1.000	0.943	0.057
DTC, Tuned Model	0.814	0.725	0.090	0.969	0.904	0.065
HGBC, Baseline Model	0.618	0.600	0.018	0.896	0.881	0.015
XGBC, Baseline Model	0.708	0.636	0.072	0.933	0.890	0.043
MLPC, Tuned Model	0.798	0.786	0.012	0.860	0.852	0.000

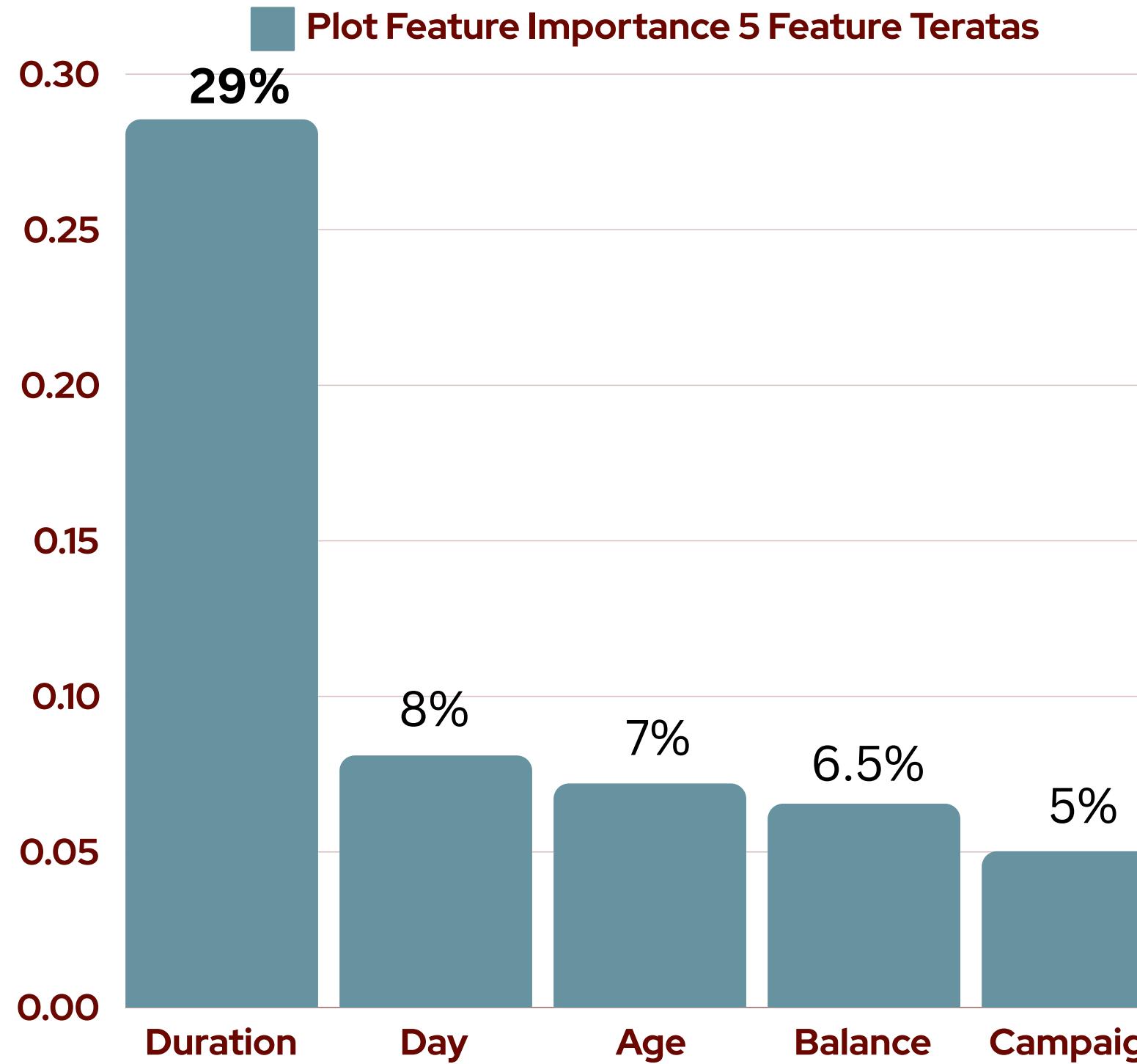
MODEL COMPARISON



EXPLANATION

- Diperoleh 3 model yang mampu memperoleh f1 score sempurna (train & test) yaitu KNC Tuned, ETC dan ETC Tuned.
- Model yang terpilih adalah ETC.

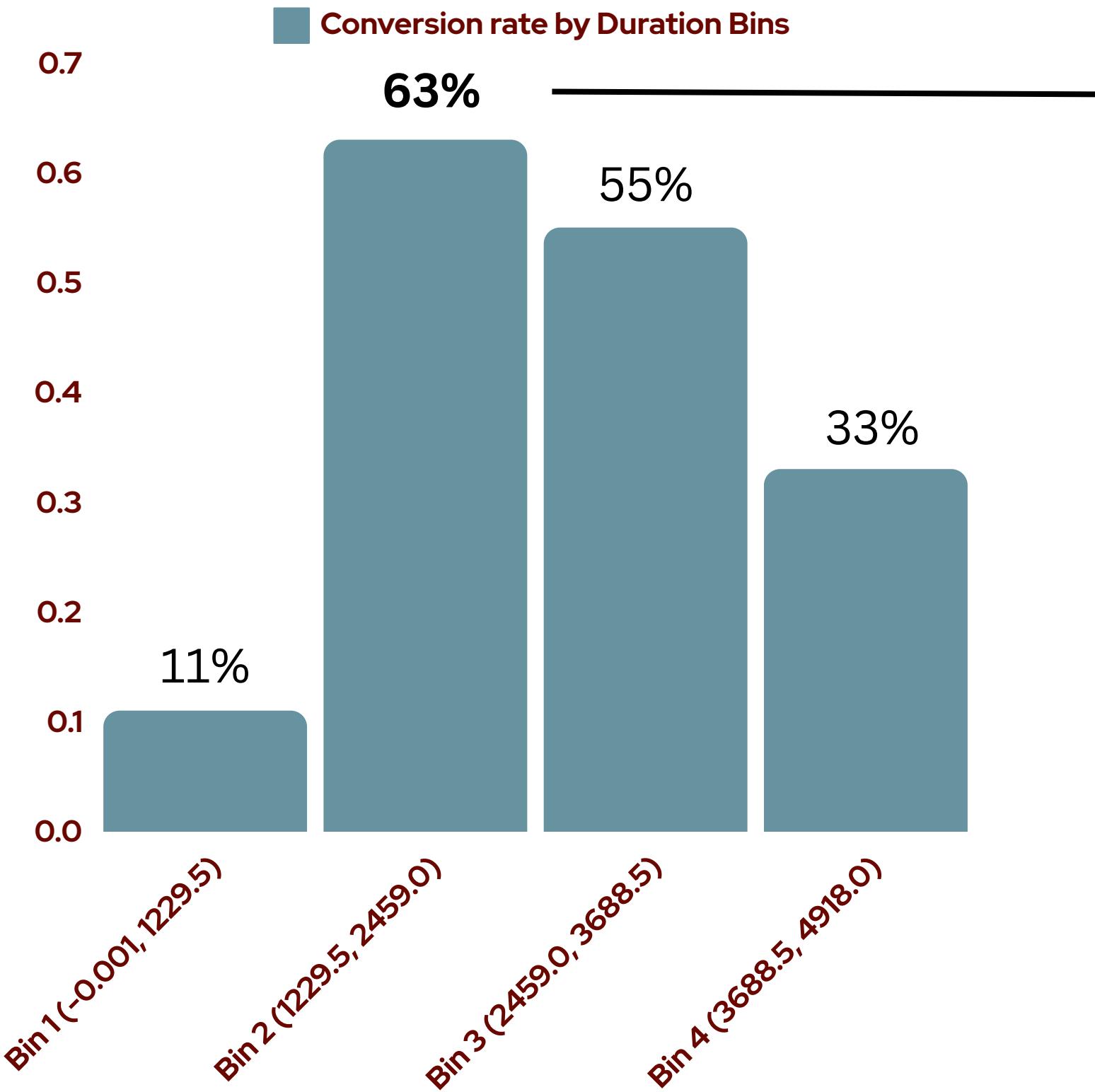
I Feature Importance



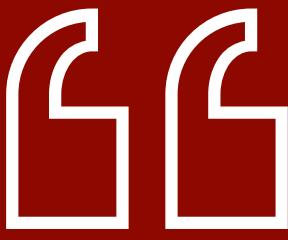
⚠️ Things to Note

- Fitur 'duration' adalah fitur paling berpengaruh untuk menentukan apakah calon nasabah akan menggunakan produk term-deposit atau tidak
- Gap mean decrease antara fitur paling berpengaruh ('duration') dengan ke-2 ('day') cukup jauh.

I Feature Importance, Duration Feature



- Conversion rate tertinggi pada Bin 2: interval (1229.5, 2459.0) detik ATAU (20.5, 40.98] menit
- Selain itu, duration pada Bin 2 memiliki perbedaan signifikan dengan interval pada Bin 1 ($p < 0.05$).
- Sedangkan, interval lain tidak memiliki perbedaan signifikan terhadap interval dengan conversion rate tertinggi

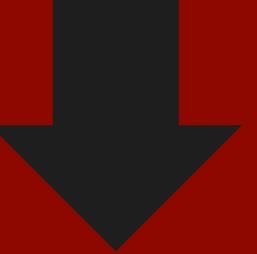


BUSINESS RECOMMENDATION

1

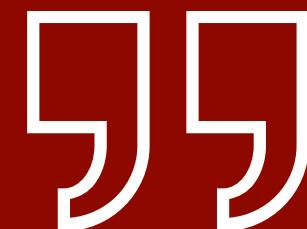
Fokus pada Durasi Kontak

Durasi kontak (duration) adalah fitur paling penting. Ini menunjukkan bahwa terdapat interval duration yang efektif dalam convert berlangganan term deposit. Tim pemasaran harus memastikan bahwa percakapan dengan nasabah tidak terburu-buru dan memberikan waktu yang cukup untuk menjelaskan manfaat dari term deposit.

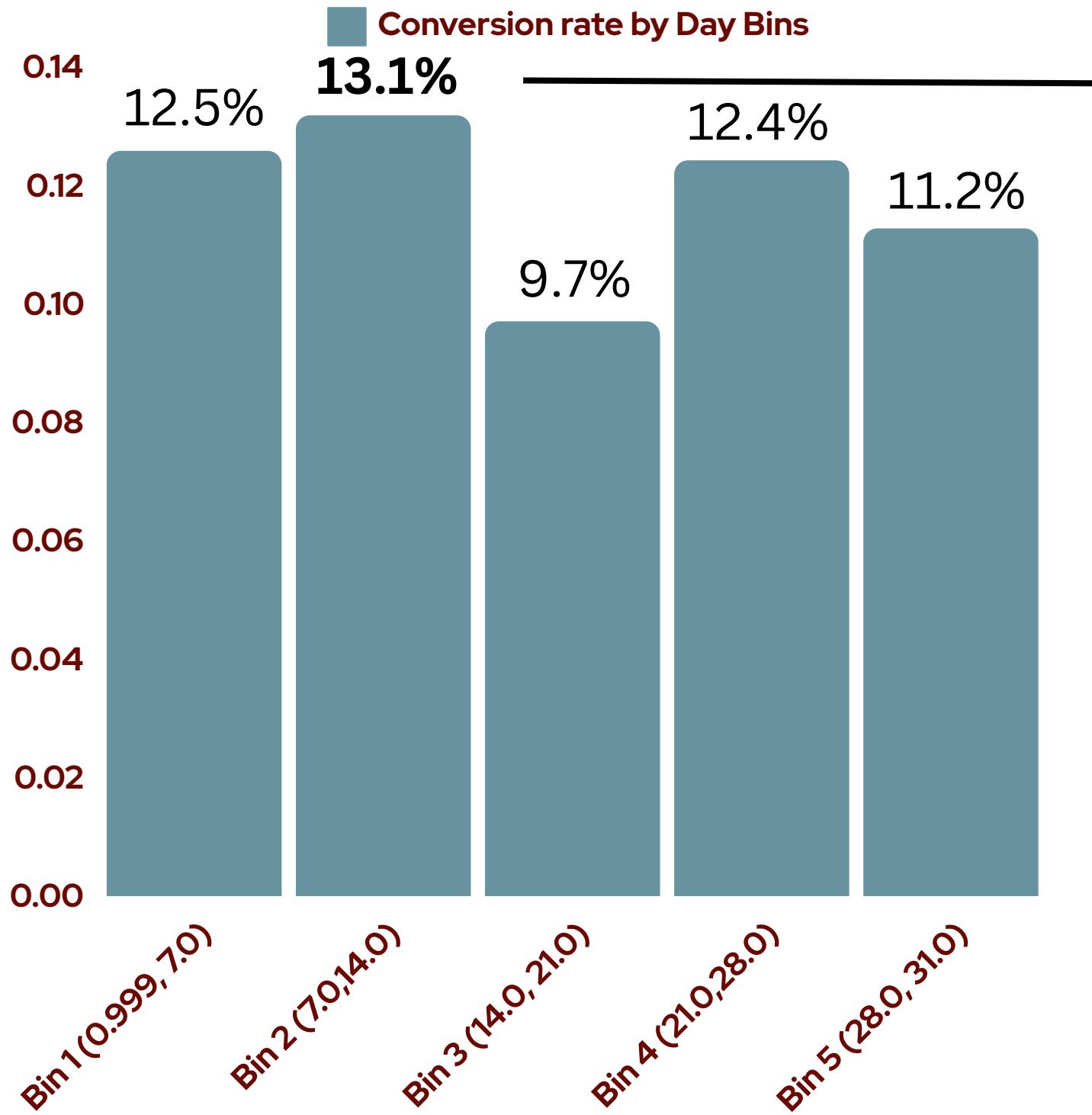


Action Items

- **Pelatihan Tim:** Berikan pelatihan kepada tim telemarketing untuk memastikan bahwa mereka mampu melakukan percakapan yang lebih panjang dan informatif dengan nasabah.
- **Skrip Panggilan:** Kembangkan skrip panggilan yang lebih rinci untuk membantu agen menjaga percakapan tetap berjalan dengan baik dan informatif.



I Feature Importance, Day Feature



- Interval feature day dengan conversion rate tertinggi (13%) berada pada hari ke-(7.0, 14.0] atau week ke-2
- Berdasarkan z test proportion, interval hari (0.999, 7.0] dan (21.0, 28.0] memiliki nilai conversion rate yang efektif dengan interval (7.0, 14.0], walaupun secara conversion rate tetap berada dibawah.

BUSINESS RECOMMENDATIONS

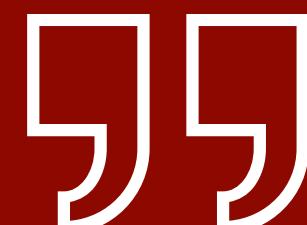
2

Optimalisasi Hari Kontak

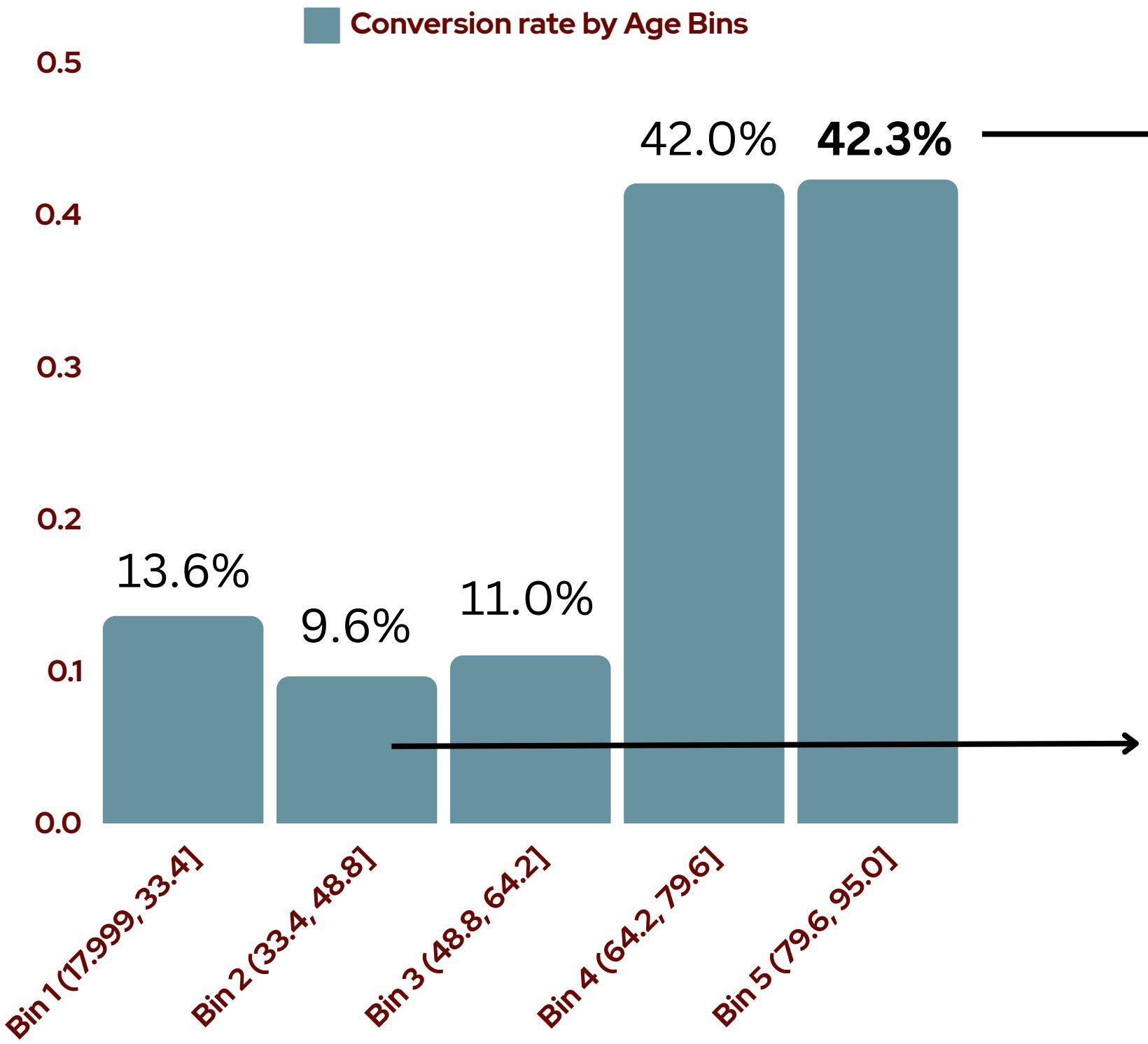
Hari kontak (day) adalah fitur kedua yang paling penting. Analisis lebih lanjut diperlukan untuk menentukan hari-hari tertentu yang lebih efektif untuk menghubungi nasabah. Kampanye pemasaran bisa diatur untuk memaksimalkan kontak pada hari-hari tersebut.

Action Items

- **Analisis Data:** Lakukan analisis tambahan untuk menemukan hari-hari tertentu dalam sebulan yang paling efektif untuk menghubungi nasabah.
- **Penjadwalan:** Sesuaikan jadwal kampanye pemasaran untuk memaksimalkan kontak pada hari-hari tersebut.

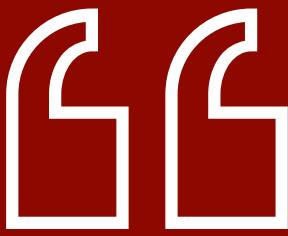


I Feature Importance, Age Feature



- Conversion rate tertinggi fitur age adalah interval (79.6, 95.0], dengan conversion rate 42%
- Terlihat perbedaan signifikan antara interval lain terhadap interval dengan conversion rate tertinggi.
- Distribusi umur nasabah bank terbanyak

BUSINESS RECOMMENDATIONS



3

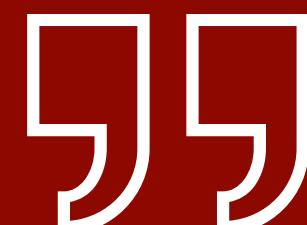
Segmentasi Berdasarkan Usia

Usia (age) adalah fitur penting lainnya. Kampanye pemasaran bisa disesuaikan berdasarkan kelompok usia yang berbeda. Misalnya, pesan pemasaran yang berbeda mungkin lebih efektif untuk nasabah yang lebih muda dibandingkan dengan nasabah yang lebih tua.



Action Items

- **Personalisasi Pesan:** Gunakan pesan yang lebih dinamis dan digital untuk nasabah yang lebih muda, dan pendekatan yang lebih konservatif dan detail untuk nasabah yang lebih tua.
- **Targeting Spesifik:** Gunakan segmentasi usia untuk mengarahkan kampanye pemasaran ke demografi yang berbeda dengan penawaran yang sesuai.



CONTENT LIST



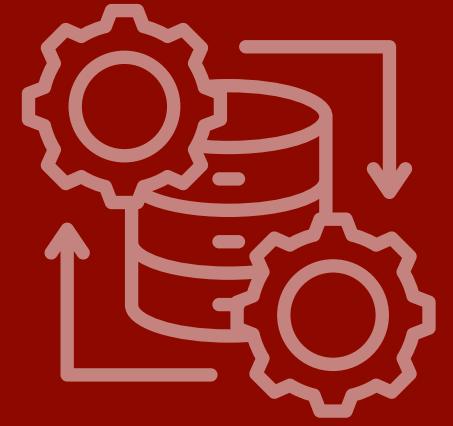
STAGE 0

Project
Introduction



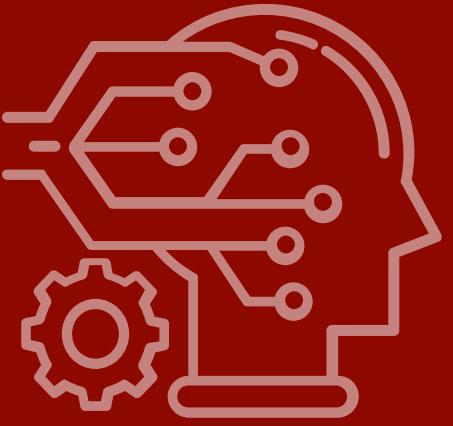
STAGE 1

EDA &
Visualization



STAGE 2

Data
Preprocessing



STAGE 3

Modeling &
Evaluation



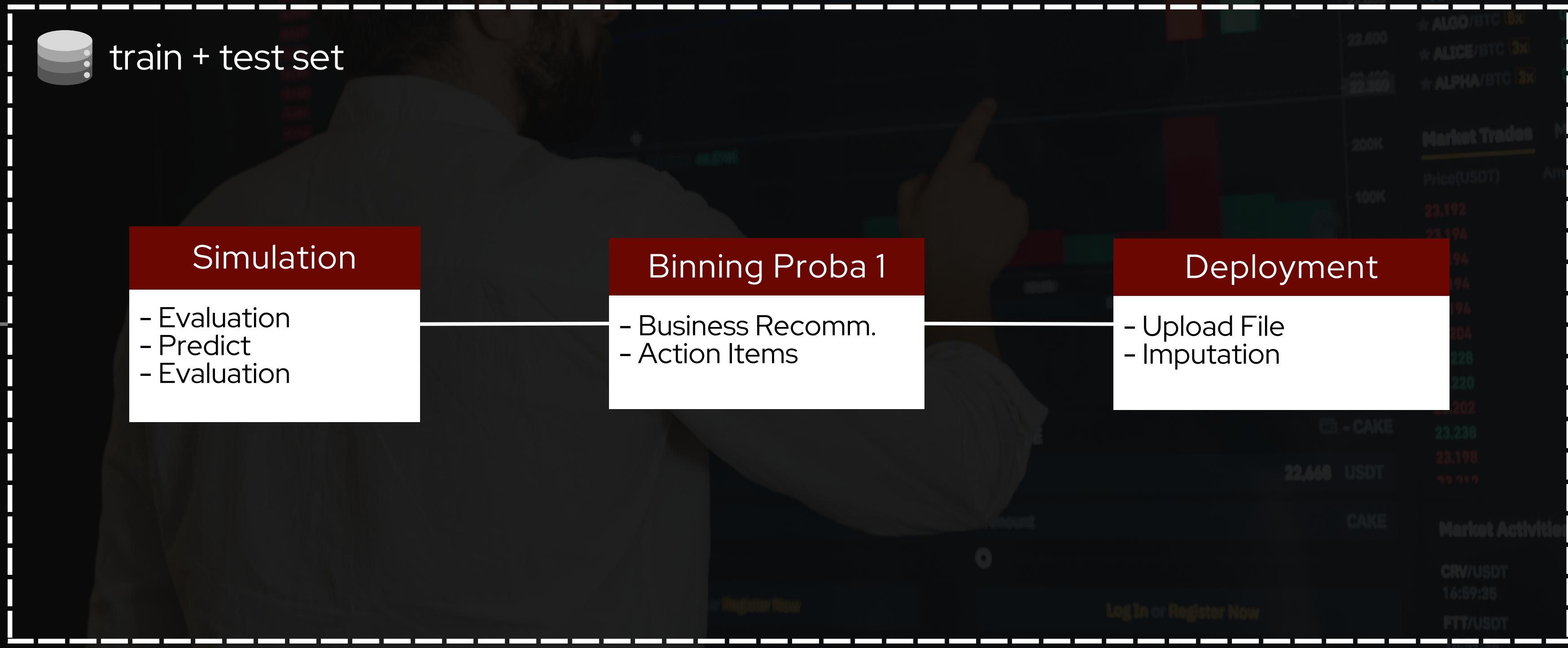
STAGE 4

Differentiation



| Flowchart

STAGE 4 DIFFERENTIATION

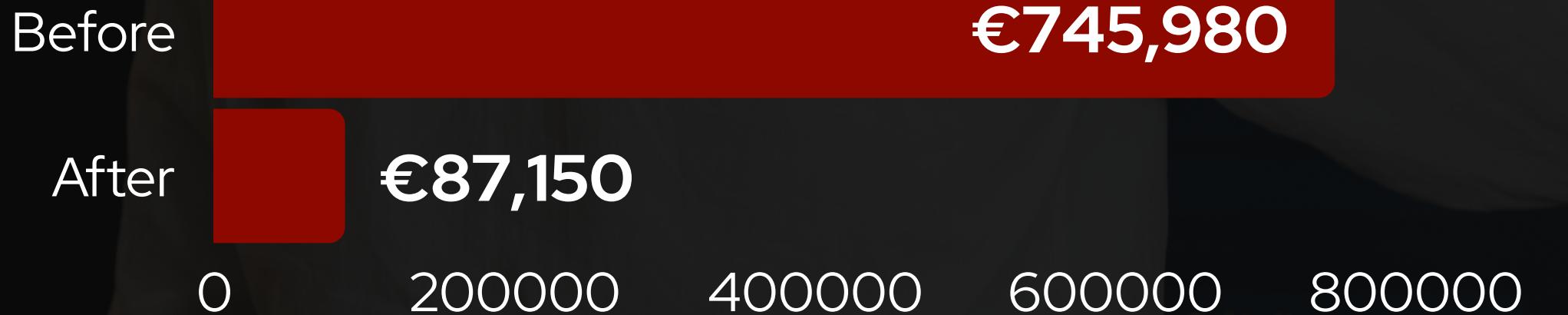


| Simulation

1

Optimasi Operational Campaign

Data yang dipakai adalah data yang tersedia di internet untuk setiap orang = \$15/person
Dataset yang digunakan adalah penggabungan data train dan data test dengan jumlah
49.732 rows



Efisiensi Operational
Campaign Hingga

88%

Log In or Register Now

| Simulation

2

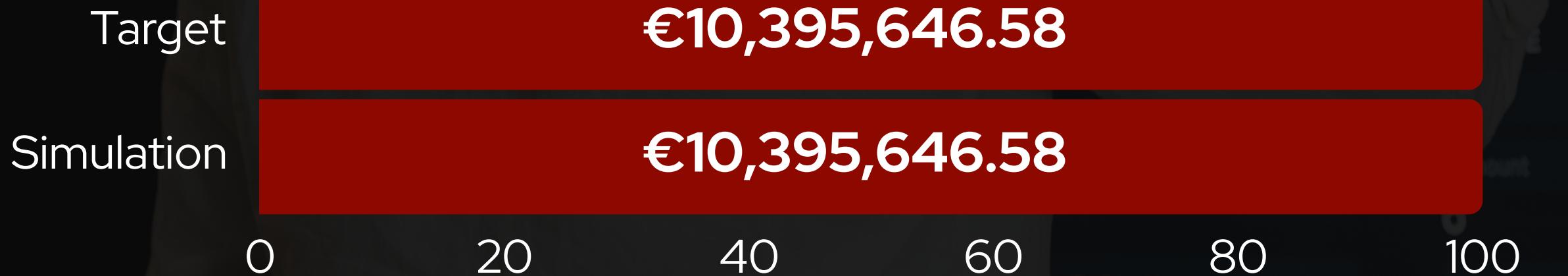
Optimasi Potencial Profit

Rata-rata balance nasabah yang sudah berlangganan (€1804.26) digunakan sebagai asumsi potencial profit yang dapat diperoleh Bank secara historis & realistik.

Potencial profit didefinisikan dengan pengurangan rata-rata balance nasabah yang sudah berlangganan dan operasional campaign untuk setiap orang.



SIMULATION



Potencial profit dapat mencapai nilai **MAXIMUM!**

I Binning Proba 1 & Deployment, Upload File

Model Prediksi - Bank Marketing Target

Aplikasi untuk menentukan apakah nasabah berpotensi berlangganan "Term Deposit" atau tidak. ↗

Modeling by PDSI-Perjuangan | Deployment by Muhammad Iqbal A.

Pilih Tipe Prediksi

Prediksi dari file .csv

Pilih file .csv

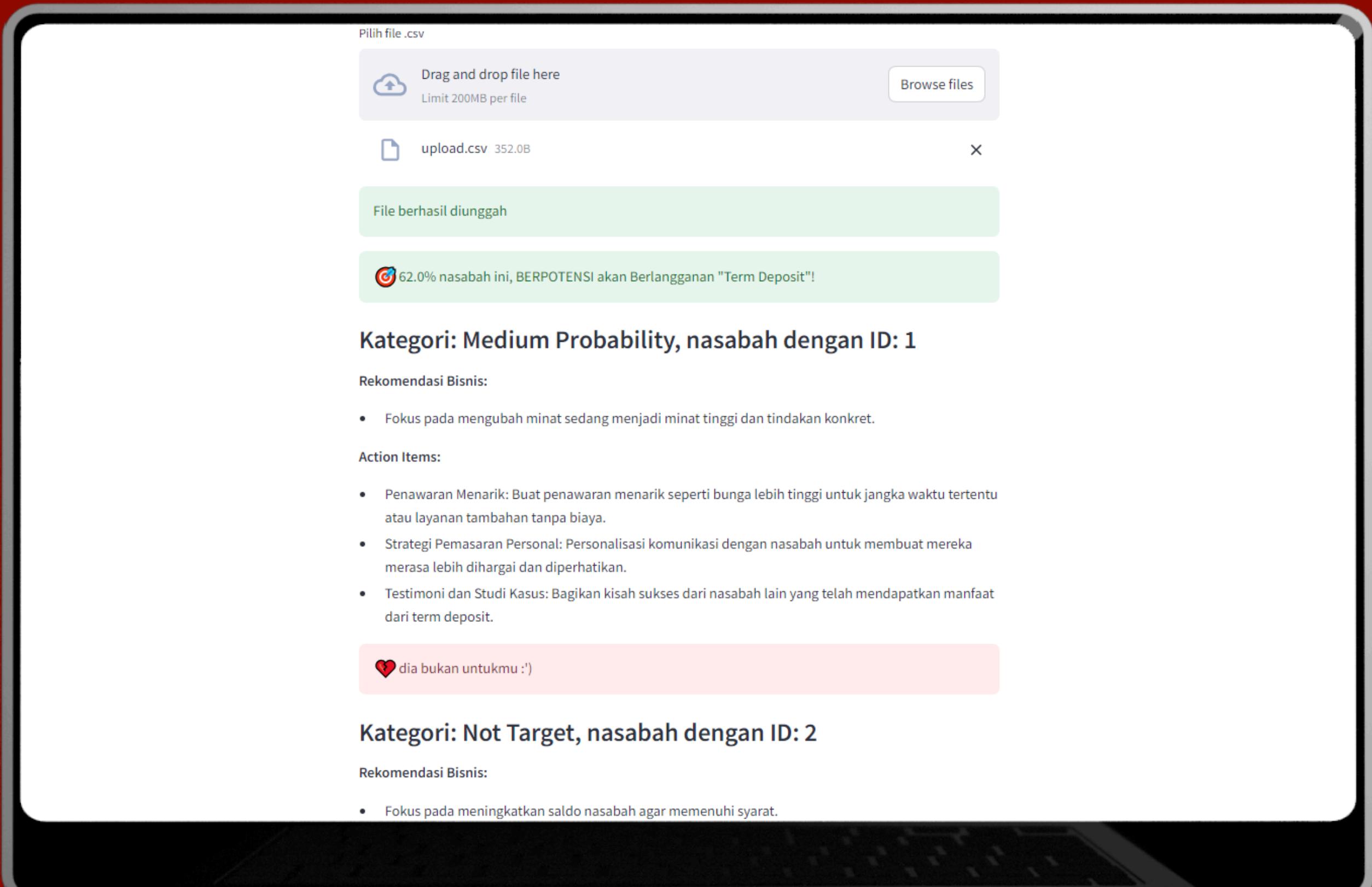


Drag and drop file here
Limit 200MB per file

Browse files

File yang diunggah kosong, silakan pilih file yang valid.

I Binning Proba 1 & Deployment, Upload File



Pilih file .csv

Drag and drop file here
Limit 200MB per file

Browse files

upload.csv 352.0B ×

File berhasil diunggah

🎯 62.0% nasabah ini, BERPOTENSI akan Berlangganan "Term Deposit"!

Kategori: Medium Probability, nasabah dengan ID: 1

Rekomendasi Bisnis:

- Fokus pada mengubah minat sedang menjadi minat tinggi dan tindakan konkret.

Action Items:

- Penawaran Menarik: Buat penawaran menarik seperti bunga lebih tinggi untuk jangka waktu tertentu atau layanan tambahan tanpa biaya.
- Strategi Pemasaran Personal: Personalisasi komunikasi dengan nasabah untuk membuat mereka merasa lebih dihargai dan diperhatikan.
- Testimoni dan Studi Kasus: Bagikan kisah sukses dari nasabah lain yang telah mendapatkan manfaat dari term deposit.

💔 dia bukan untukmu :')

Kategori: Not Target, nasabah dengan ID: 2

Rekomendasi Bisnis:

- Fokus pada meningkatkan saldo nasabah agar memenuhi syarat.

I Binning Proba 1 & Deployment, Imputation

Model Prediksi - Bank Marketing Target

Aplikasi untuk menentukan apakah nasabah berpotensi berlangganan "Term Deposit" atau tidak.

Modeling by PDSI-Perjuangan | Deployment by Muhammad Iqbal A.

Pilih Tipe Prediksi

Imputasi

Client Data

Umur (tahun) [age]

55

Campaign Data

Jenis Komunikasi [contact]

telephone

Pekerjaan [job]

management

Hari ketika last Contact [day]

1

Pernikahan [marital]

married

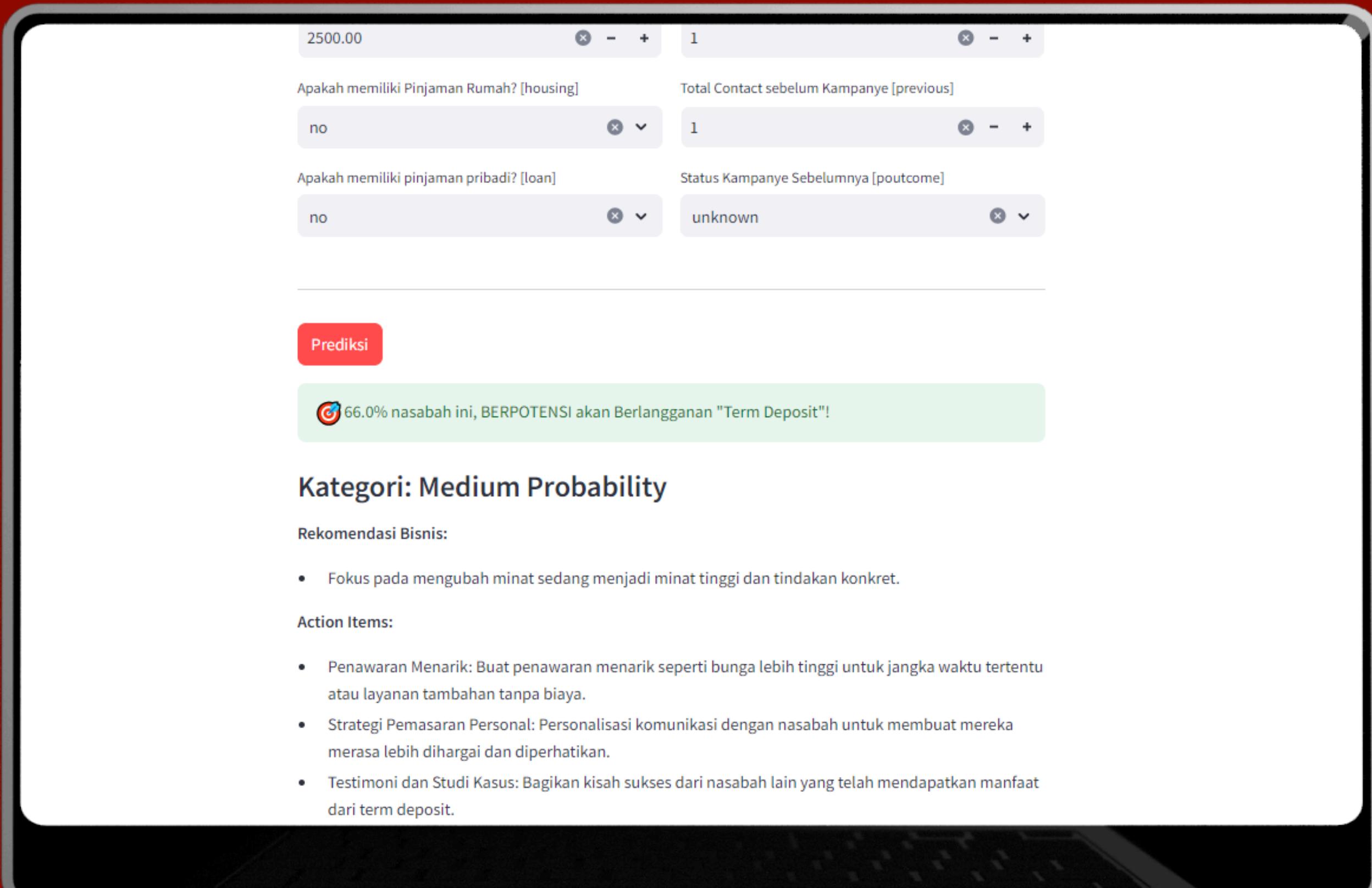
Bulan ketika last Contact [month]

mar

Pendidikan [education]

Durasi last Contact (detik) [duration]

I Binning Proba 1 & Deployment, Imputation



The screenshot shows a user interface for a predictive model. At the top, there are four input fields with numerical values and dropdown menus:

- 2500.00 (with a multiplier of 1)
- Apakah memiliki Pinjaman Rumah? [housing] (no)
- Total Contact sebelum Kampanye [previous] (1)
- Apakah memiliki pinjaman pribadi? [loan] (no)
- Status Kampanye Sebelumnya [poutcome] (unknown)

Below these inputs is a red button labeled "Prediksi". Underneath the button is a green callout box containing the text: "🎯 66.0% nasabah ini, BERPOTENSI akan Berlangganan "Term Deposit"!"

Following the callout box, the text "Kategori: Medium Probability" is displayed in bold.

Under "Rekomendasi Bisnis:", there is a bulleted list:

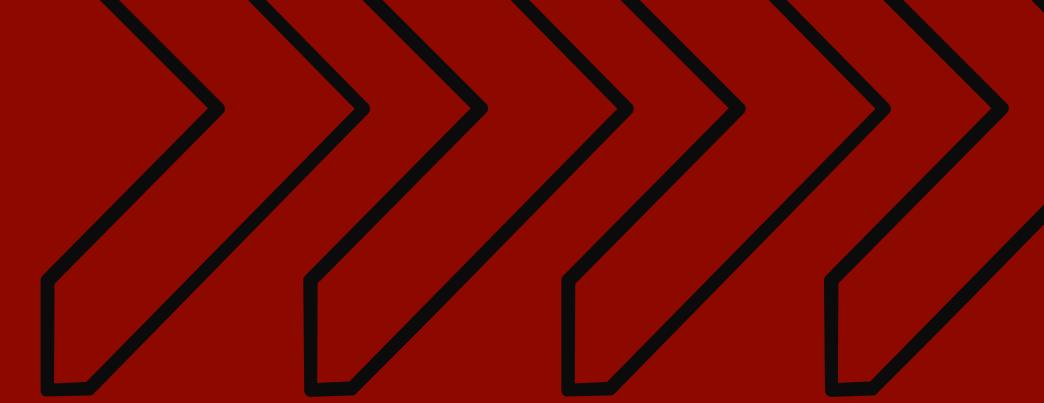
- Fokus pada mengubah minat sedang menjadi minat tinggi dan tindakan konkret.

Under "Action Items:", there is another bulleted list:

- Penawaran Menarik: Buat penawaran menarik seperti bunga lebih tinggi untuk jangka waktu tertentu atau layanan tambahan tanpa biaya.
- Strategi Pemasaran Personal: Personalisasi komunikasi dengan nasabah untuk membuat mereka merasa lebih dihargai dan diperhatikan.
- Testimoni dan Studi Kasus: Bagikan kisah sukses dari nasabah lain yang telah mendapatkan manfaat dari term deposit.

THANK YOU

HAVE A NICE DAY!



PDSI
PERJUANGAN