Rakamin
Academy

Predict.

Term Deposit Potential Customer Prediction

Case Study from Portuguese Banking Institution

Data by Kaggle

Content Direction

1 **Background**
- Business Context
- Problem & Solutions

2 **EDA**
- Univariate Analysis
- Multivariate Analysis

3 **Pre-Processing**
- Data Cleansing
- Feature Engineering

4 **Modeling Experiment**
- Process Explained
- Modeling Result

5 **Executive Summaries & Recomendation**

Predict.

Rakamin Academy

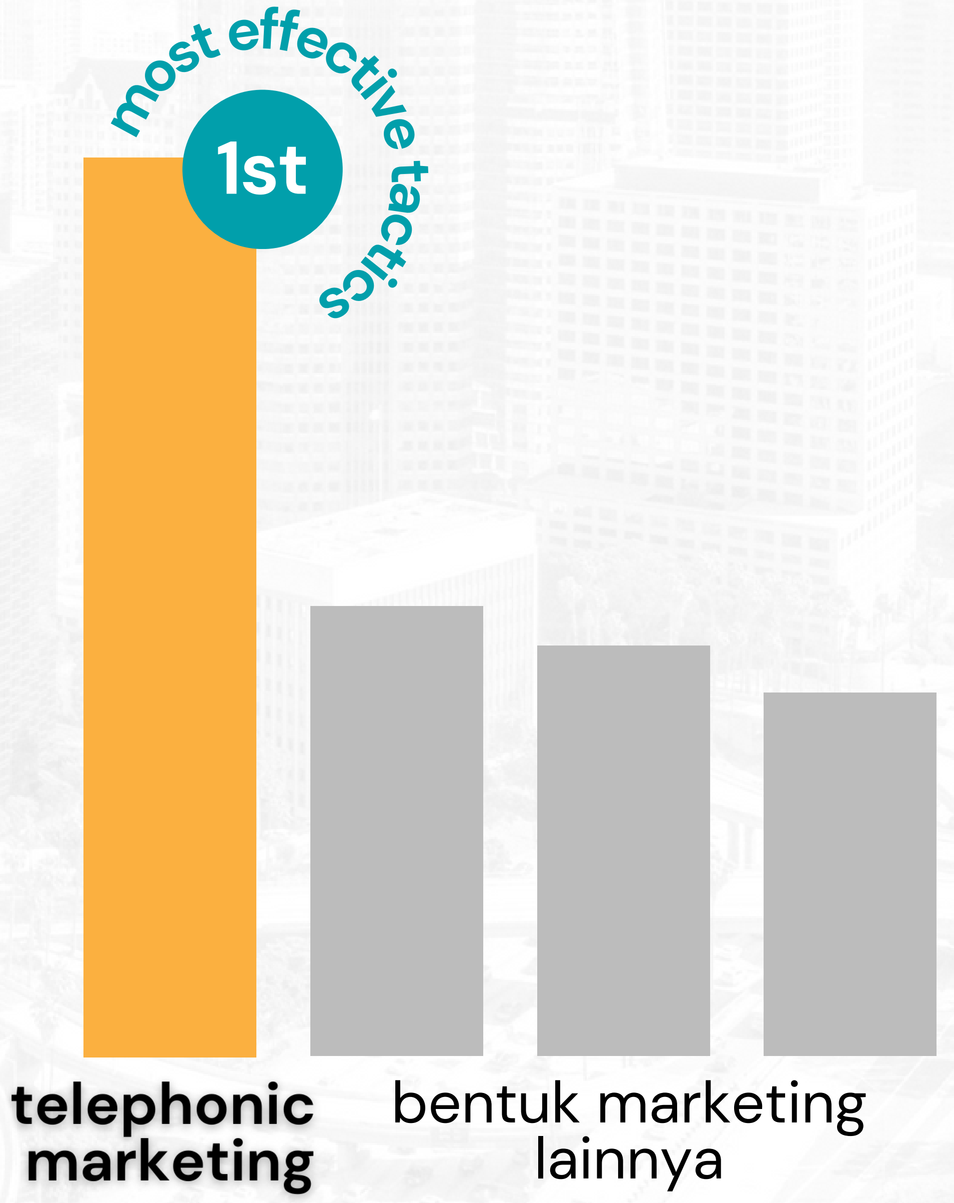
Background

Business Context

Satu Banking Institution di Portugese memiliki **major source income** dari <u>Term Deposit Program</u>

Bank memiliki banyak *marketing tactics* untuk Program yang terdiri atas,

- email marketing
- advertisements
- telephonic marketing
- digital marketing



Predict.

Rakamin Academy

Background

Problem

Marketing Capital *telephonic marketing* terlalu tinggi dengan *conversion rate* yang hanya sebesar **11.7%**

How can it be so low?

Bank belum memiliki gambaran akan Potential Customer yang akan bersedia menerima Program Marketing Term Deposit

Solutions

1. Mengetahui Potential Customer
2. Prediction Machine Learning Model yang bisa meng-klasifikasi *potential customer*

Business Metrics

1. Berhasil apabila setelah diterapkan dapat meningkatkan ***conversion rate***

Machine Learning Technical

Exploratory Data Analysis

Pre-Processing

Modeling Experiment

Exploratory Data Analysis

Descriptive Analysis

- Feature & Label
- Missing Values
- Weird Column Summaries

Univariate Analysis

- Numerical
- Categorical

Multivariate Analysis

- Heatmap Correlation
- Hued PairPlot to labels

Predict.

Rakamin Academy

Exploratory Data Analysis
Descriptive Analysis

1 Feature & Labels

- 16 *feature* & 1 label

2 Missing Values

- Tidak ada **NULL Values**
- Nilai 'Other' & 'Unknown'

3 Weird Column Summaries

- **Balance** memiliki nilai minimum negatif
- **Previous** & **Pdays** memiliki nilai 0 & -1 hingga Q3
- **Poutcome** memiliki modus dengan nilai 'Unknown'

Kategorikal	Ordinal	default, month, loan, housing, education, y
	Nominal	marital, job, contact, poutcome
Numerical	Discrete	age, balance, day, duration, campaign, pdays, previous

[22]:

	balance
count	45211.000000
mean	1362.272058
std	3044.765829
min	-8019.000000
25%	72.000000
50%	448.000000
75%	1428.000000
max	102127.000000

Nilai Negatif

[30]:

	previous	pdays
count	45211.000000	45211.000000
mean	0.580323	40.197828
std	2.303441	100.128746
min	0.000000	-1.000000
25%	0.000000	-1.000000
50%	0.000000	-1.000000
75%	0.000000	-1.000000
max	275.000000	871.000000

Nilai Sama

[8]:

	poutcome
count	45211
unique	4
top	unknown
freq	36959

Nilai Modus

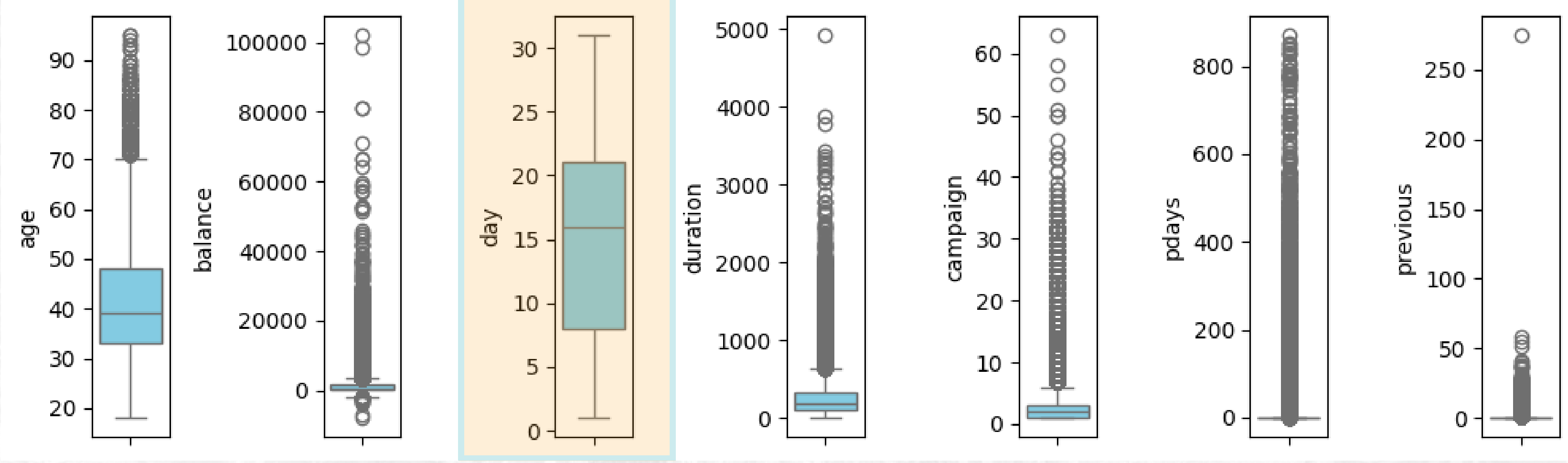
Predict.

Rakamin Academy

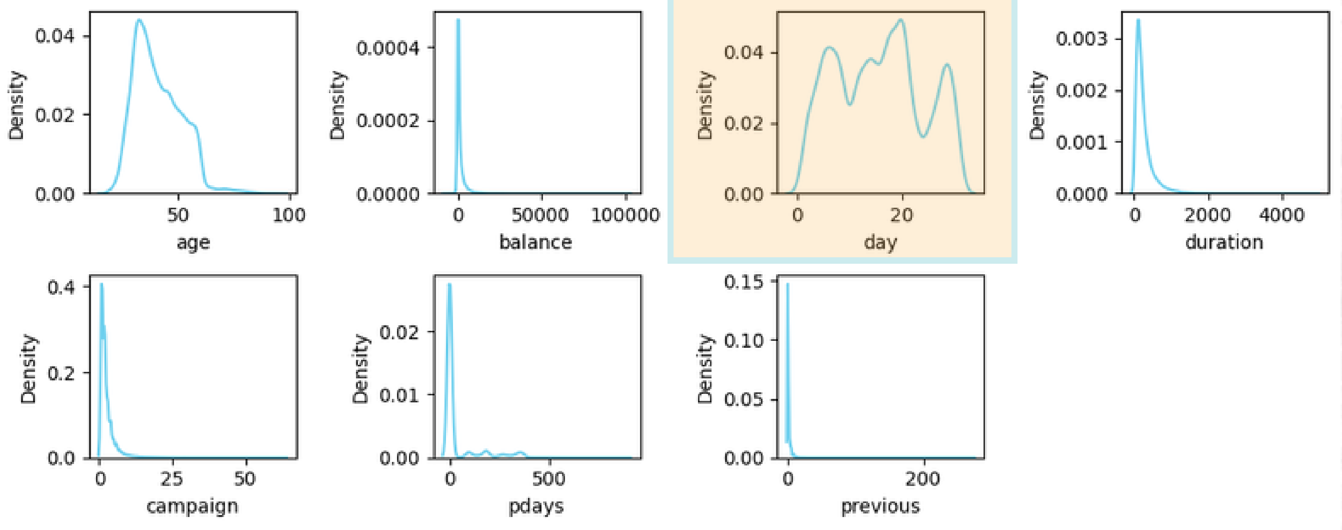
Exploratory Data Analysis
Univariate Analysis

1 Numerical Feature Plot

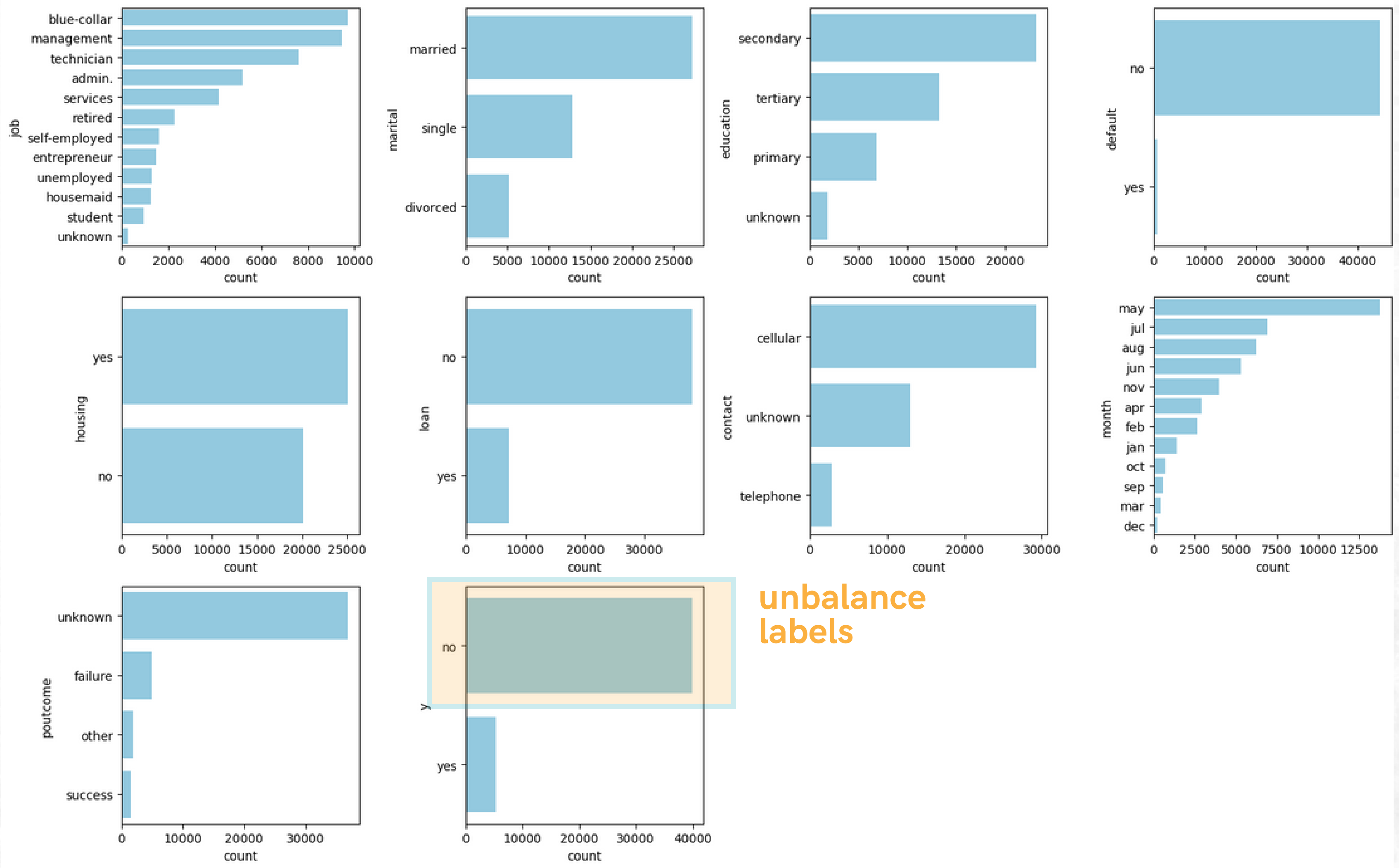
kecuali day semua ada outlier



multimodal ditribution



2 Categorical Feature Plot



unbalance labels

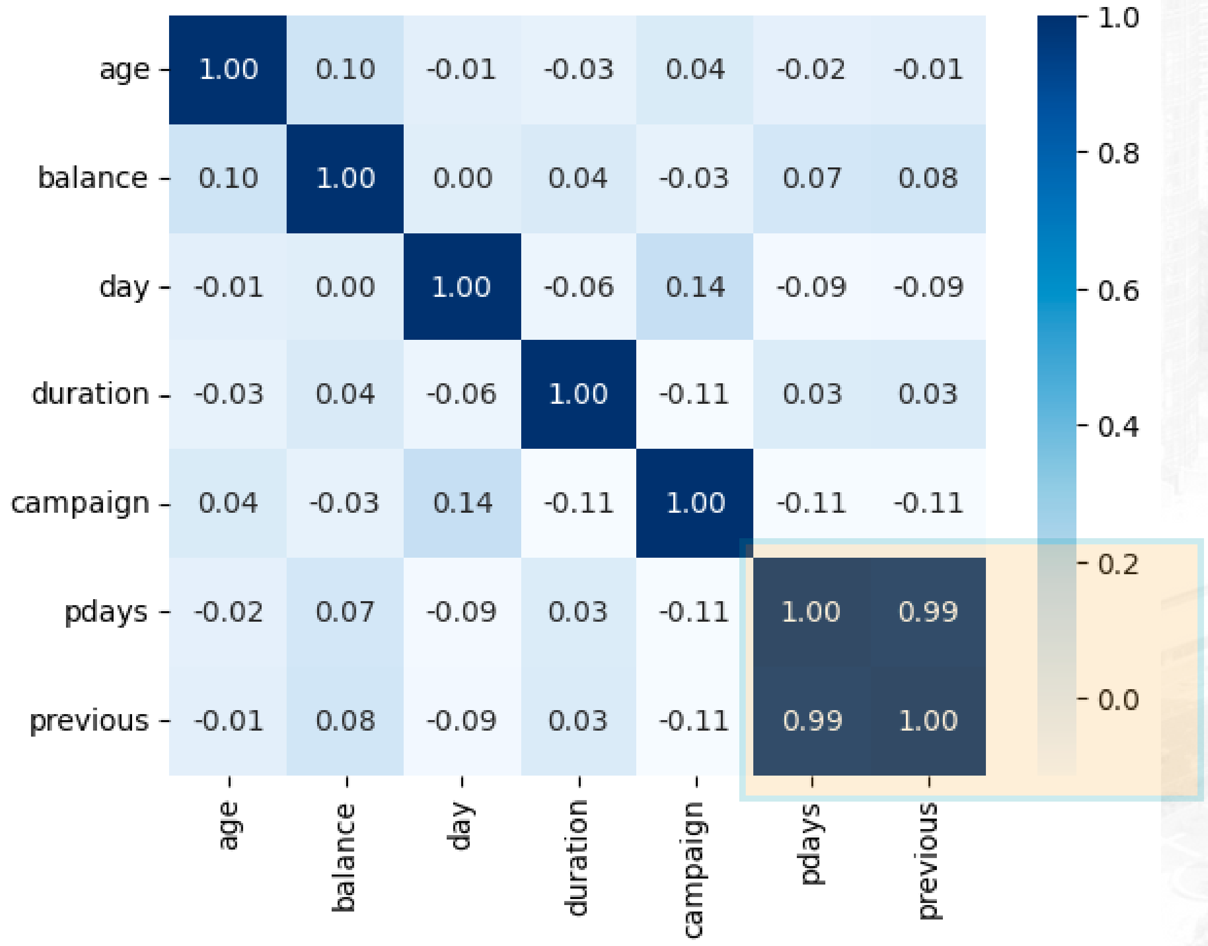
Predict.

Rakamin Academy

Exploratory Data Analysis

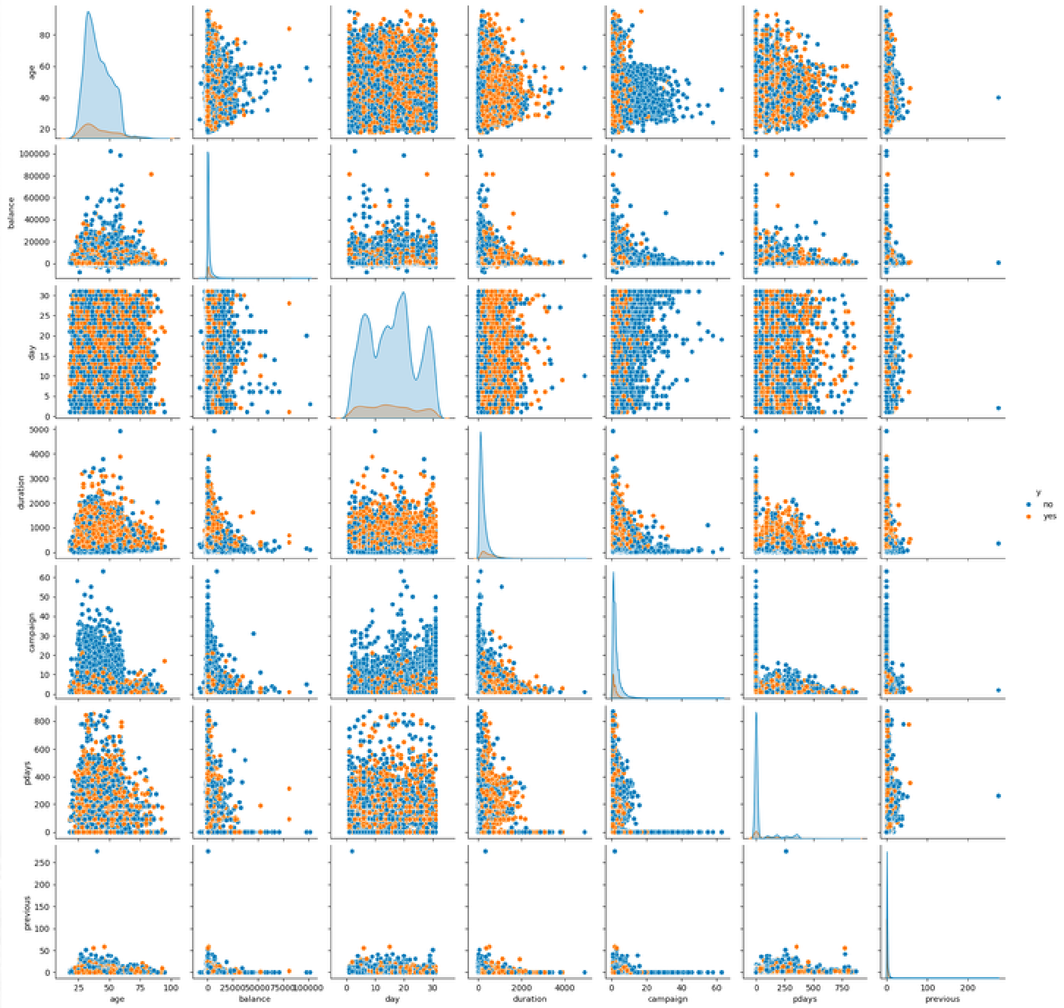
Multivariate Analysis

1 Correlation Heatmap Numerical Feature



redundan feature

2 Pairplot to identify cluster in label values



no cluster found in right top or bottom left axes

Pre-Processing

Data Cleansing

- Missing Value Handling
- Duplicated Data
- Outlier Handling
- Feature Extraction

Feature Engineering

- Feature Encoding
- Feature Transformation
- Feature Selection
- Class Imbalance Handling

Pre-Process & Modeling Iteration

Predict.

Rakamin Academy

1st Iteration

Pre-Processing		Modeling Experiment	
Train Data	Test Data	Train Data 90.92%	Test Data 9.08%
	(-) Outlier Handling	Default Parameter	
	(-) Balance Handling	GridSearchCV( )	
		Random Parameter	

2nd Iteration

Pre-Processing		Modeling Experiment
Train Data	Test Data	Concat Data
		Split (80 : 20)
Concat Data		Default Parameter
Duplicated Data Handling		GridSearchCV( )
		Random Parameter

Predict.

Rakamin Academy

Pre–Processing
Data Cleansing

1 Missing Values Handling

Replace with Mode Value

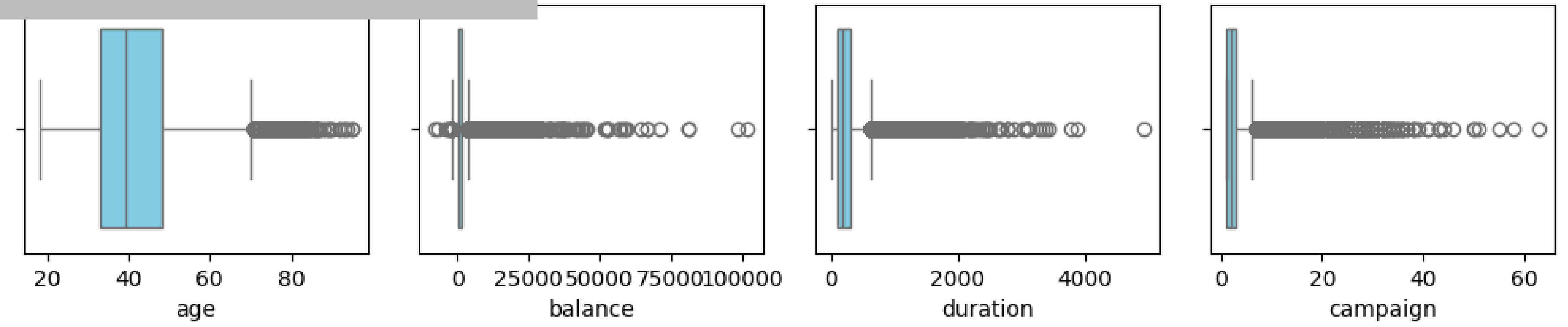
Drop the Feature

2 Duplicated Data Handling

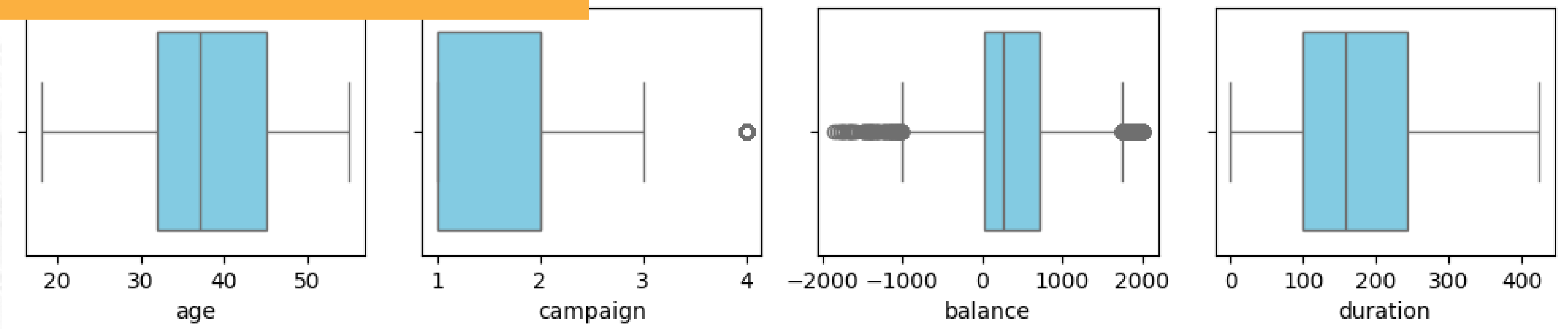
Perform hanya di second iteration, tidak di first iteration

3 Outlier Handling

Before Outlier handling



After Outlier handling



Predict.

Rakamin Academy

Pre-Processing
Data Cleansing

4 Feature Extraction

variable	before feature-extraction	after feature-extraction
`previous`	definition : Number of contacts performed before this campaign and for this client **(numerical)**	definition : Has the client contacted on the previous campaign? **(categorical)** (0 = no, 1 = yes)

Predict.

Rakamin Academy

Pre-Processing
Feature Engineering

1 Feature Encoding

7 categorical feature
21 categorical feature

Categorical Feature	Ordinal	label encoding
	Nominal	One Hot Encoding

2 Feature Transformation

StandardScaler()

Categorical	month, education
Numerical	age, balance, day, duration, campaign

3 Feature Selection

Before Pre-Processing	16 feature
After Pre-Processing	27 feature

4 Class Imbalance

Perform **Oversampling** using **SMOTE**

Before SMOTE	Out[80]: 0 22638 1 1459
After SMOTE	Out[83]: 0 22638 1 22638

Predict.

Rakamin Academy

Pre-Processing
Output

1st Iteration

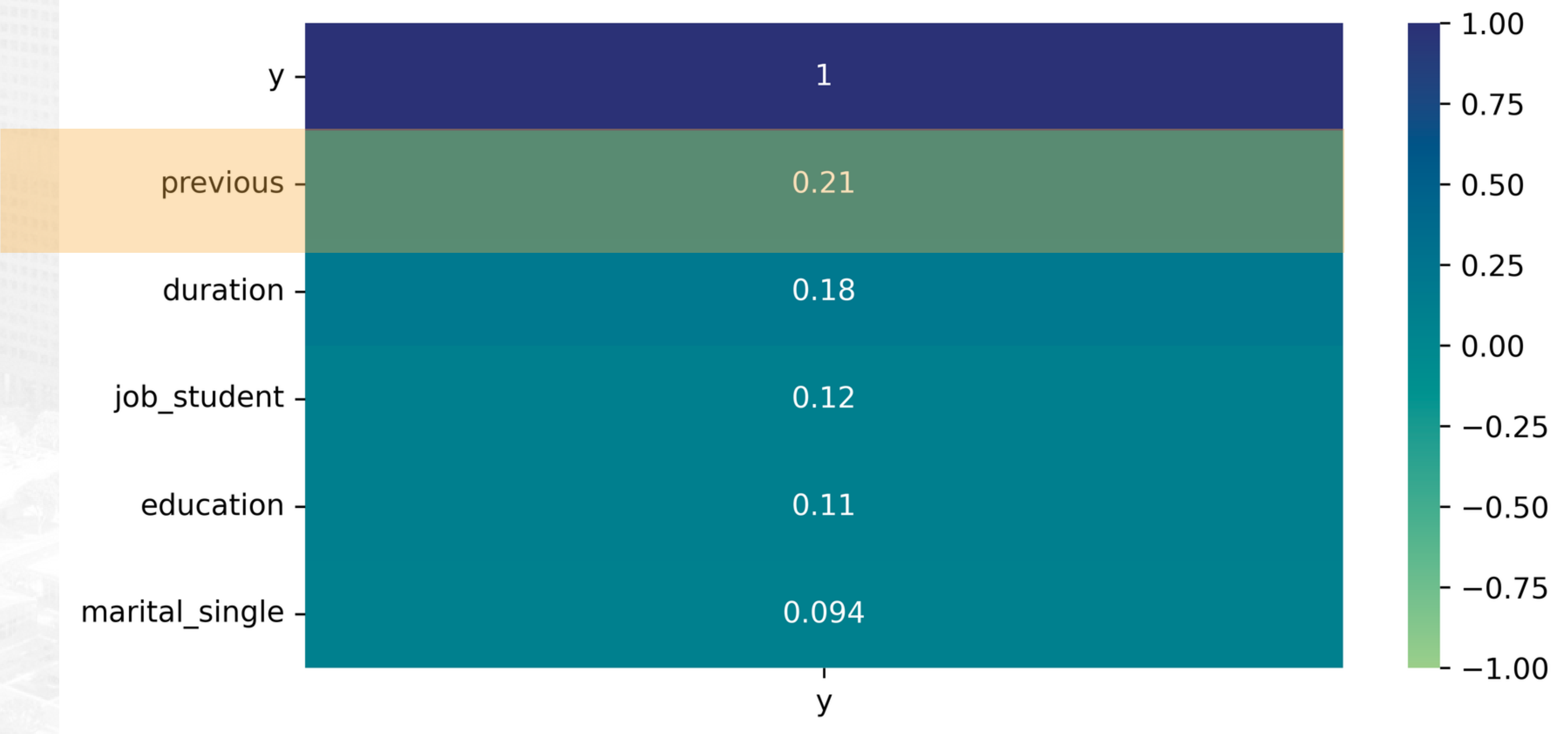
Train
90.92%

Test
9.08%

2nd Iteration

Concated Train + Test

Feature Correlated to Potential Customer



Predict.

Rakamin Academy

Modeling Experiment
Evaluation Metrics & Method

False Positive (+)

Model mudah menerapkan kategorisasi nasabah ke kelompok 'yes'

Marketing Capital **tidak** ***convert*** **dan sia-sia, padahal seharusnya bisa ditekan.**

False Negative (-)

Model mudah menerapkan kategorisasi nasabah ke kelompok 'no'

Potential revenue loss

1 Accuracy

Fokus terhadap **ketepatan prediksi terhadap** ***true positive*** **&** ***true negative*** *customer*.

2 Precision

Fokus terhadap pengurangan jumlah kasus ***false positive.***

Model hati-hati untuk mengklasifikasi *customer* ke kelompok **'yes'**

Predict.

Rakamin Academy

Modeling Experiment

1st Iteration

Method Used Explain

default parameter

Hyperparameter Tuning
GridSearchCV()

1 Logistric Regression

solver	C	max_iter	penalty
lbfgs	1	210	l2

solver	C	max_iter	penalty
liblinear	0.0001, 0.05, 1, 10, 50, 100	210, 1000, 10000	l1, l2

2 k-NN

n_neighbors	metrics (p)	algorithm
5	euclidean (2)	auto

n_neighbors	metrics (p)	algorithm
10, 50, 100, 250, 450	euclidean (2), manhattan (1)	auto

Predict.

Rakamin Academy

Modeling Experiment

1st Iteration

Method Used Explain

default parameter

Hyperparameter Tuning GridSearchCV()

3 Decision Tree

max_depth	min_samples_split	criterion	splitter
None	2	gini	best

max_depth	min_samples_split	criterion	splitter
None, 5, 10, 20	2, 5, 10	gini	best

4 Random Forest

n_estimator	max_depth	min_samples_split	criterion
100	None	2	gini

n_estimator	max_depth	min_samples_split	criterion
50, 100, 200	None, 5, 10	2	gini

Predict.

1st Iteration

Modeling Experiment

The Best Result

Rakamin Academy

Logistic Regression

solver	C	max_iter	penalty
lbfgs	1	210	l2

Model Overfitting

accuracy_train	0.86
accuracy_test	0.83

precision_train	0.86
precision_test	0.31

Predict.

Rakamin Academy

Modeling Experiment

2nd Iteration

The Best Result

Random Forest

n_estimator	max_ depth	min_samples _split	criterion
200	None	2	gini

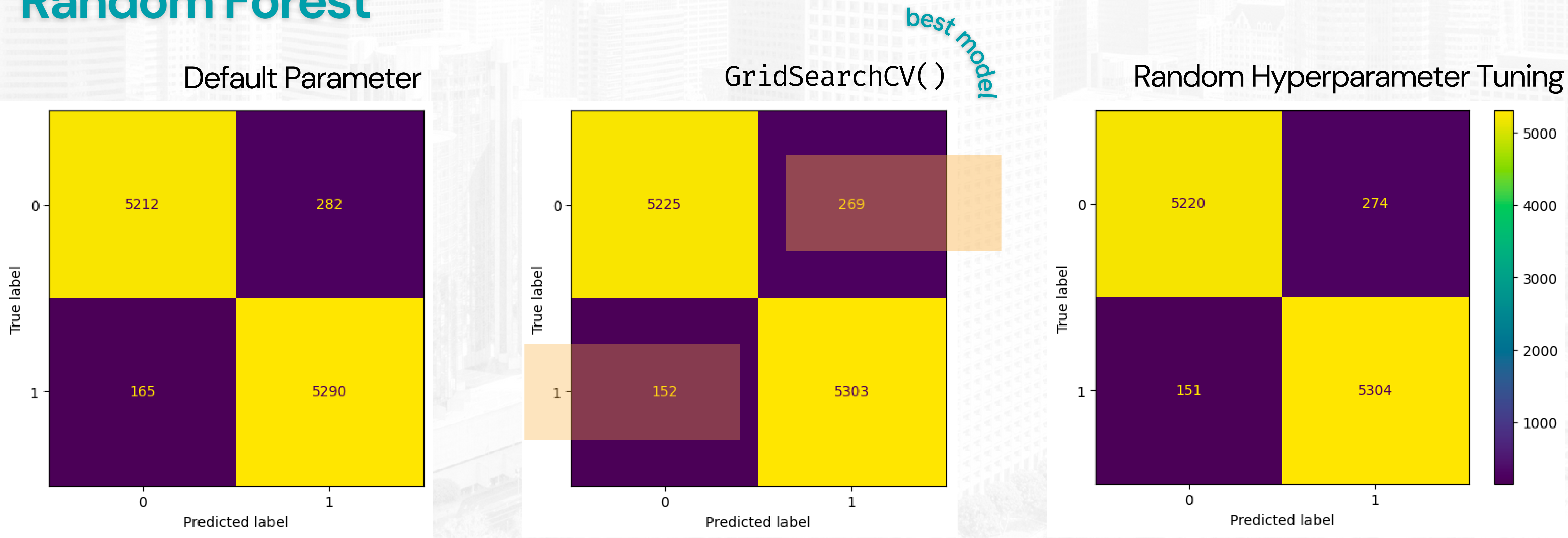
accuracy_train	1.00
accuracy_test	0.96

precision_train	1.00
precision_test	**0.95**

recall_train	1.00
recall_test	**0.97**

Modeling Experiment
Confusion Matrix Best Model

Random Forest

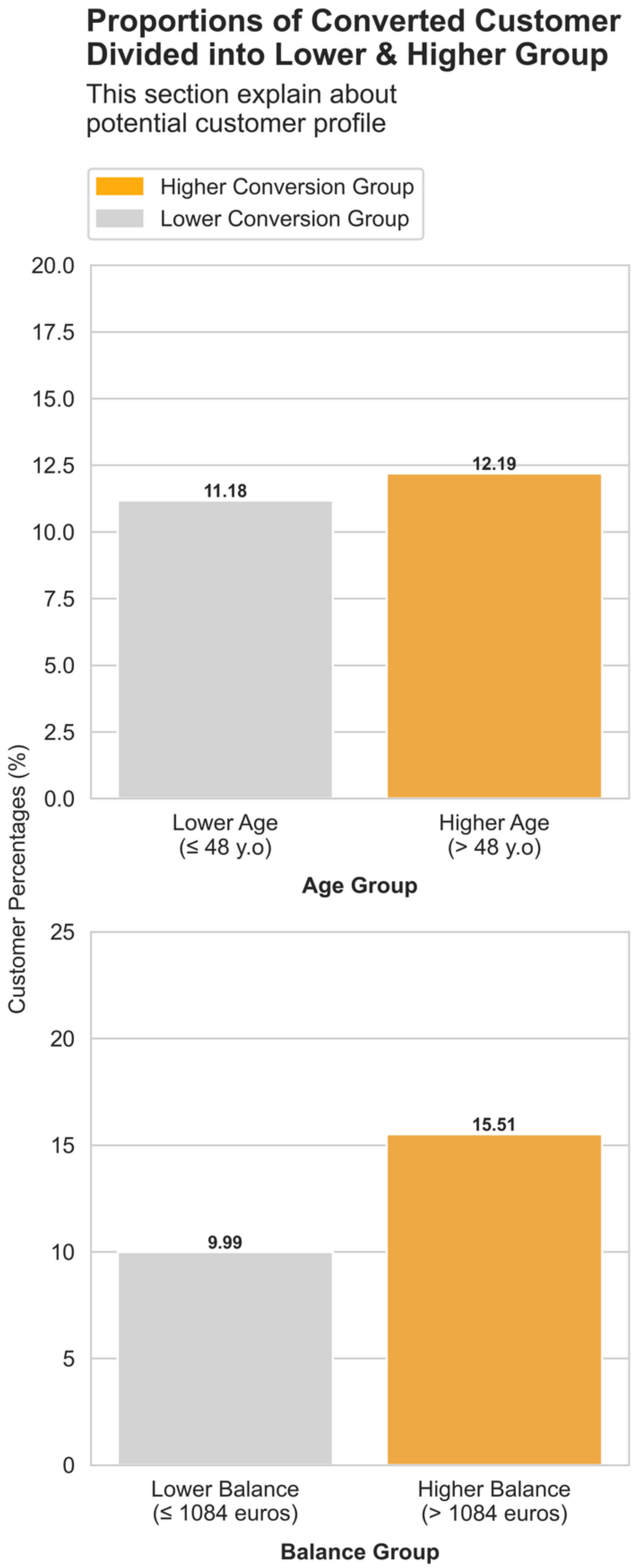


Business

Business Insights

Excecutive Summary

Recomendation



Business Insights
Potential Customer Profile

Age Group

- Converted Customer Percentages lebih tinggi condong dari kelompok usia 48 keatas.

Recomendation

Fokuskan *marketing* pada *customer* dengan kelompok usia lebih tinggi

Balance Group

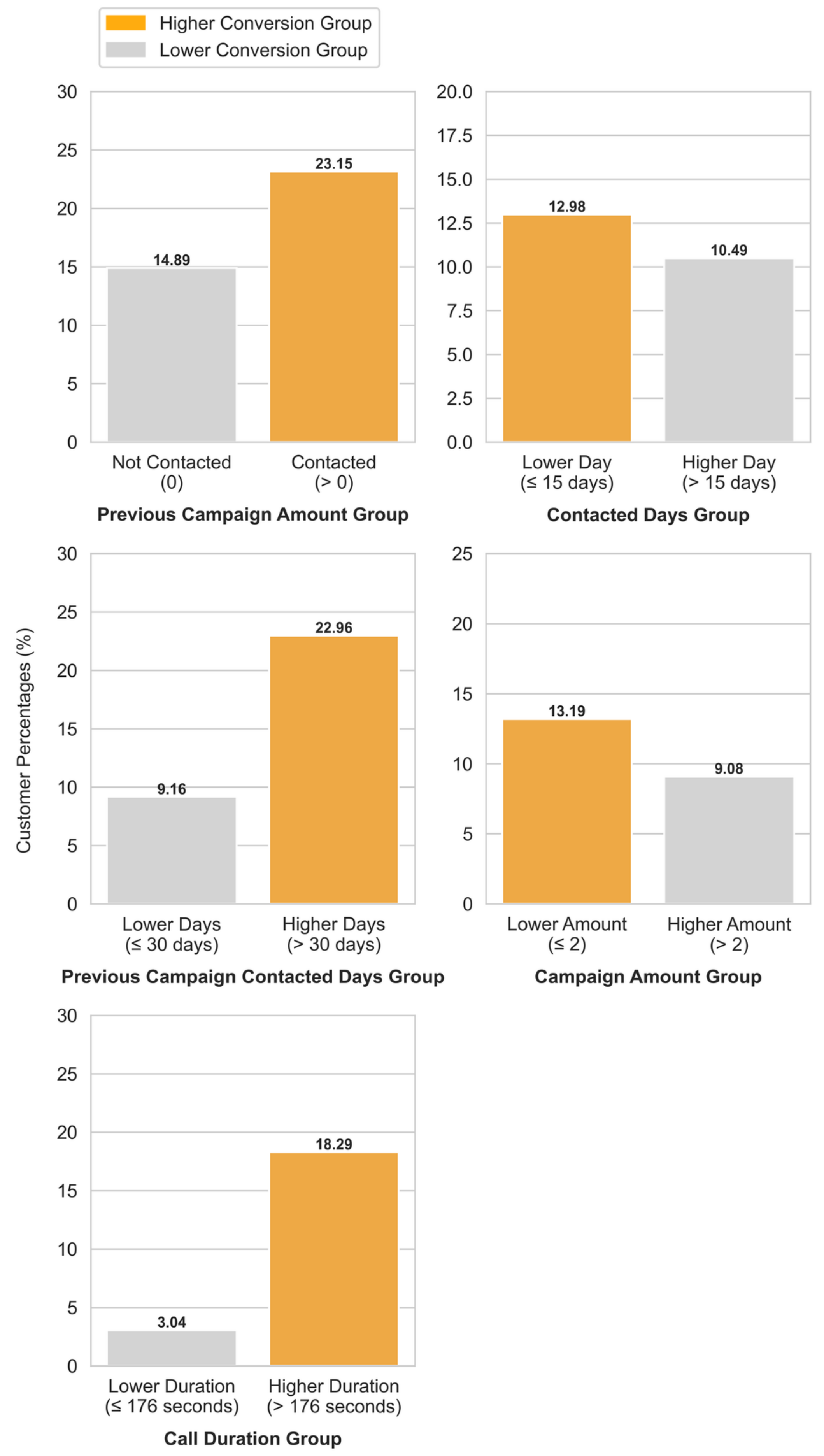
- Converted Customer lebih condong dari *higher balance value*.

Recomendation

Buat program *term deposit* dengan jangka yang lebih pendek & deposit yang lebih terjangkau.

Proportions of Converted Customer Divided into Lower & Higher Group

This section explain about customer treatment strategy



Business Insights
Marketing Strategy

Previous Campaign Amount Group

Recomendation

Fokuskan *marketing* pada *customer* yang pernah dikontak sebelumnya.

Previous Campaign Contacted Days Group

Recomendation

Buat jarak minimal antar periode campaign menjadi 30 hari.

Campaign Amount Group

Recomendation

Lakukan kontak maksimal 2 kali selama satu periode masa *campaign*.

Contacted Days Group

Recomendation

Kontak *customer* maksimal dengan jarak 15 hari sekali. Jangan terlalu sering dihubungi.

Predict.

Rakamin Academy

Executive Summary

Implementation Vision

Before Prediction Model Implementation

Customer targeted randomly & unsegmented

Customer Base : 45.211
Conversion Rate : 11.7%

Marketing Capital*
estimated a mean of € 50 for 1 customer

€ 2.260.550

Marketing Capital Loss*
nominal of unconverted customer base

€ 1.996.065,65

Predict.

Rakamin Academy

Executive Summary

Implementation Vision

Possibility After Prediction Model Implementation

Using Predictive Model & Segmented

Customer Base : 45.211

Conversion Rate : 95%*

based on model precision

Marketing Capital*

estimated a mean of € 50 for 1 customer

€ 2.260.550

Trade-Off

Marketing Capital Loss*

nominal of unconverted customer base

€ 113.027,5

Potential Customer Loss*

supposedly labeled converted but not

1.38%

Predict.

Rakamin Academy

Business Recomendation

Machine Learning Model Implementation Process

1. Deciding Campaign Scope
2. Gather Customer Data Needed
3. Model Implementation
4. Segment the potential customer from result gained
5. Execute Marketing Campaign
6. Marketing Evaluation
7. Remodel the Prediction Machine if under perform

Predict.

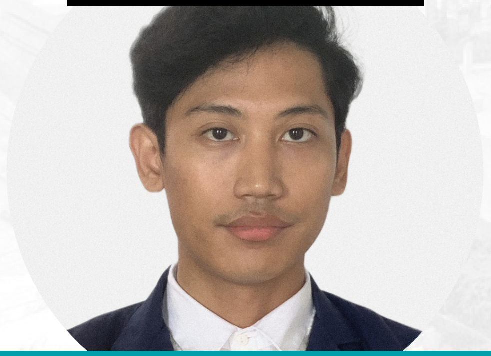
Rakamin Academy

Team Collaborator

Faalih

Project Mentor

Bagus



Project Leader/QA

Roby

Fatin

Louis

Technical ML Division

Tasya

Ramzy

Rifa

Sholdan

Business Division

Rakamin Academy

Predict.

Thank You

bagusatya08/**Term-Deposit-Potential-...**

Python based ML Project to Predict Potential Customer for Term Deposit

4 Contributors | 0 Issues | 0 Stars | 0 Forks

bagusatya08/Term-Deposit-Potential-Customer-Prediction: Python based ML Project to Predict Potential Customer for Term Deposit

Python based ML Project to Predict Potential Customer for Term Deposit - GitHub - bagusatya08/Term-Deposit-Potential-Customer-Prediction: Python based ML Project to Predict Potential Customer for T...

GitHub