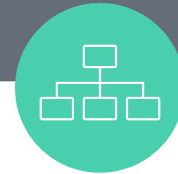
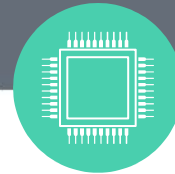




Marketing Targets Strategy on Banking Dataset




By : phpMyFeeling

Our Team Members


- Arya Octavian
- Athalla Dewanto
- Bagoes Fikri
- Intan Denovita
- Jose Christian
- Jovian Aditya
- Muhammad Hazim
- Nur Almar

Bank POV

Peminjam Uang / Surat
Berharga Negara



Aku punya
uang dingin,
Investasi yang
aman apa ya?



AHA. Aku
masukkan ke
term deposit
bank



Higher Interest rate

Interest rate
Setelah jangka waktu



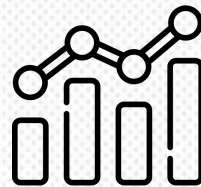
PROBLEM STATEMENT : Jumlah nasabah bank yang telah mendaftar deposito berjangka pada Perusahaan Bank ABC masih rendah. Dari 45.211 nasabah, hanya sekitar 11,7 % nasabah saja yang mendaftar deposito berjangka.



GOAL : Meningkatkan Conversion Rate User nasabah bank yang mendaftar deposito berjangka minimal sebesar 5% atau dapat dikatakan 1 dari 20 nasabah bank mendaftar deposito berjangka.



OBJECTIVE : Melakukan klasifikasi pada customer bank tersebut, sehingga nasabah yang sekiranya tertarik untuk membuka deposito berjangka dapat diprioritaskan terlebih dahulu untuk dihubungi/ditawari deposito berjangka.



BUSINESS METRICS : Conversion Rate User yang membuka deposito berjangka



Dataset Overview

Terdapat **7 Kolom Numerik**

- age
- balance
- day
- duration
- campaign
- pdays
- previous

Terdapat **17 Kolom**
Dengan **45211 baris**

Terdapat **10 Kolom Kategorikal**

- job
- marital
- education
- default
- housing
- loan
- contact
- month
- poutcome
- **y: [Target]**



Exploratory Data Analysis

Descriptive Statistics

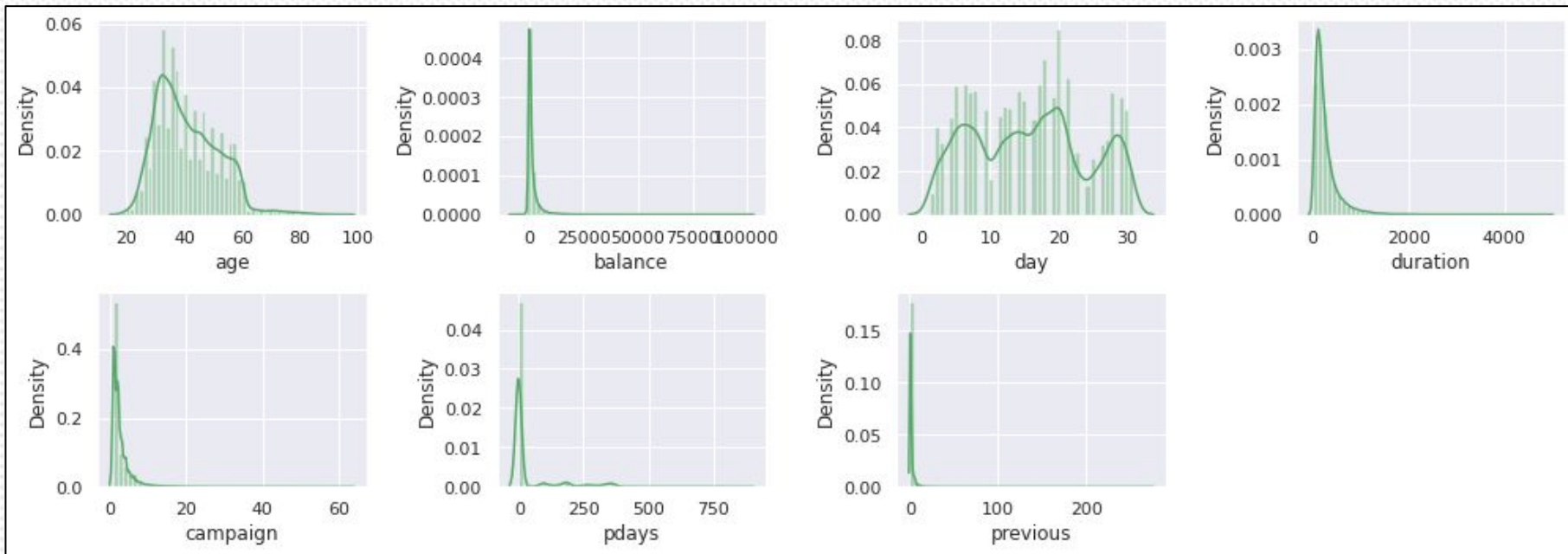
```
[ ] df[nums].describe()
```

	age	balance	day	duration	campaign	pdays	previous
count	45211.000000	45211.000000	45211.000000	45211.000000	45211.000000	45211.000000	45211.000000
mean	40.936210	1362.272058	15.806419	258.163080	2.763841	40.197828	0.580323
std	10.618762	3044.765829	8.322476	257.527812	3.098021	100.128746	2.303441
min	18.000000	-8019.000000	1.000000	0.000000	1.000000	-1.000000	0.000000
25%	33.000000	72.000000	8.000000	103.000000	1.000000	-1.000000	0.000000
50%	39.000000	448.000000	16.000000	180.000000	2.000000	-1.000000	0.000000
75%	48.000000	1428.000000	21.000000	319.000000	3.000000	-1.000000	0.000000
max	95.000000	102127.000000	31.000000	4918.000000	63.000000	871.000000	275.000000

```
[ ] df[cats].describe()
```

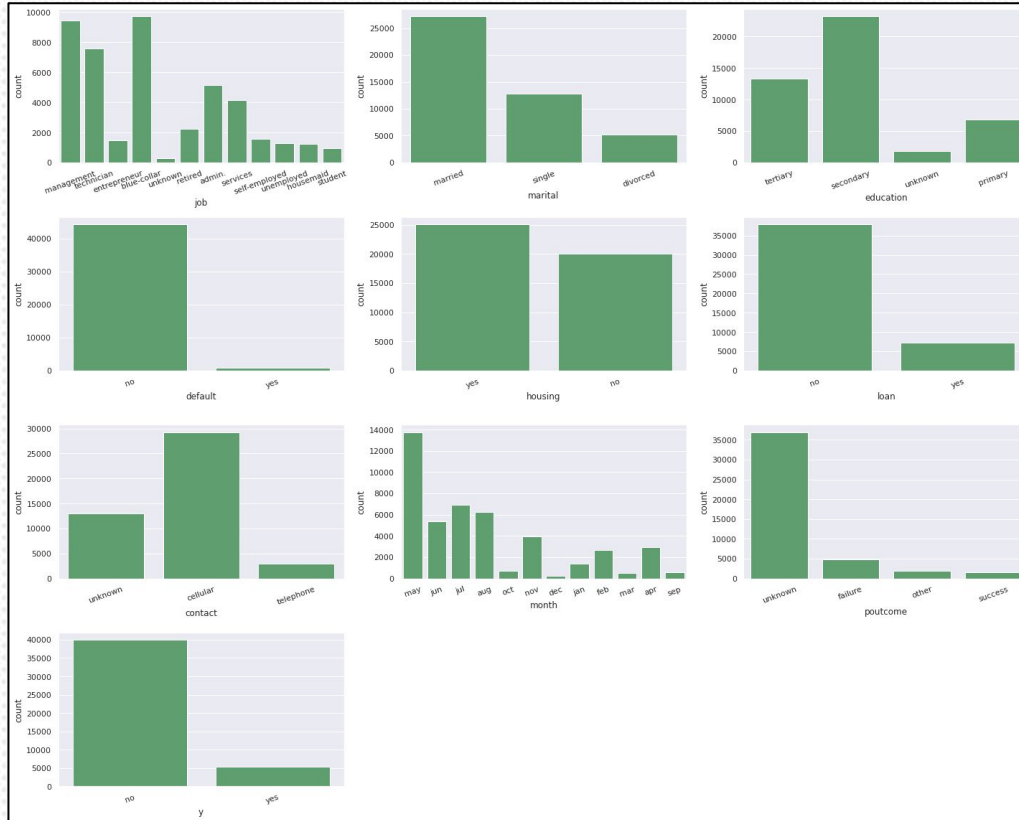
	job	marital	education	default	housing	loan	contact	month	poutcome	y
count	45211	45211	45211	45211	45211	45211	45211	45211	45211	45211
unique	12	3	4	2	2	2	3	12	4	2
top	blue-collar	married	secondary	no	yes	no	cellular	may	unknown	no
freq	9732	27214	23202	44396	25130	37967	29285	13766	36959	39922

Distribusi Data Numerik



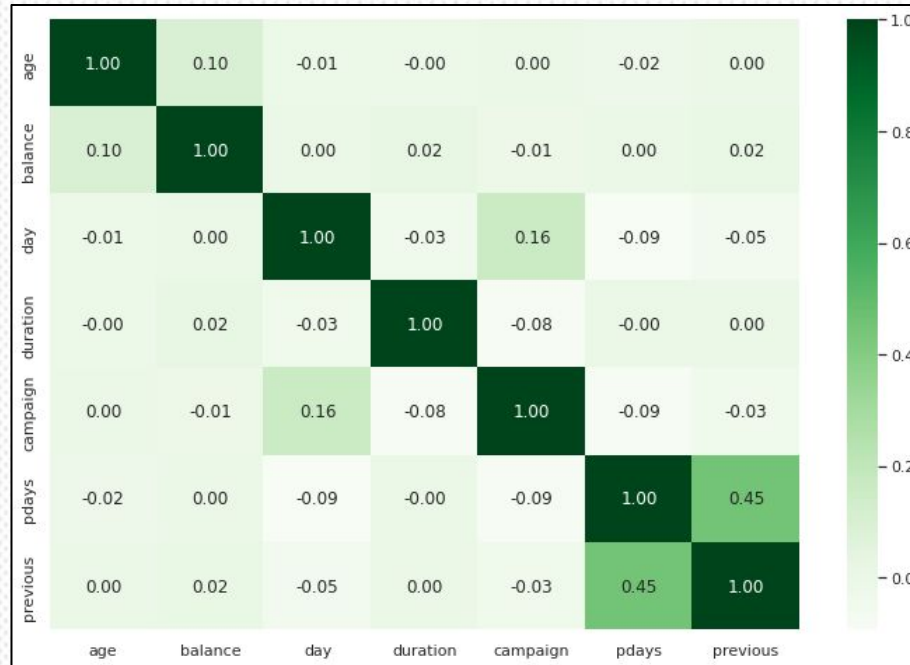
- Kolom age memiliki bentuk distribusi mendekati normal
 - Kolom day memiliki distribusi multimodal
 - Selebihnya mengalami positif skewness

Distribusi Data Kategorikal



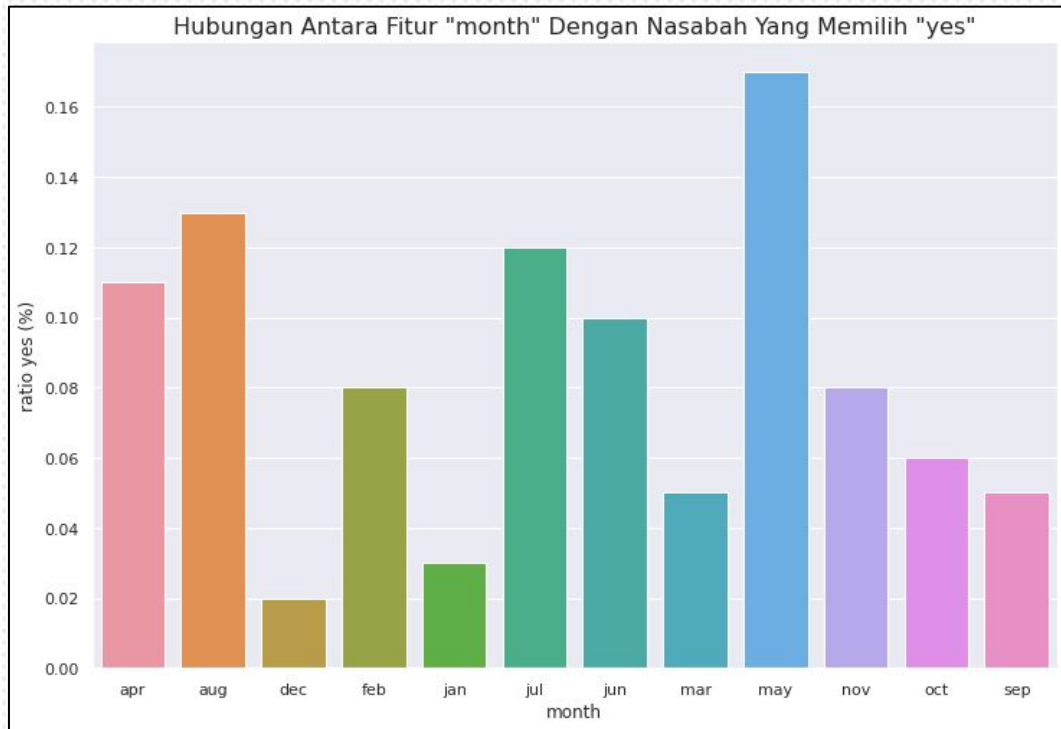
- Feature job didominasi 5 pekerjaan yaitu "blue-collar", "management", "technician", "admin", dan "services".
- Kolom target yaitu kolom y memiliki 2 frekuensi antara value "yes" dan "no" tidak seimbang, dimana value "yes" hanya berkisar 11,7% dari keseluruhan value.
- Untuk kolom education,contact yang memiliki nilai unknown, pada saat preprocessing bisa digantikan dengan nilai mode nya
- Saat proses preprocessing kolom-kolom kategorikal ini akan dilakukan encoding.

Matriks Correlation Antara Feature Numerikal



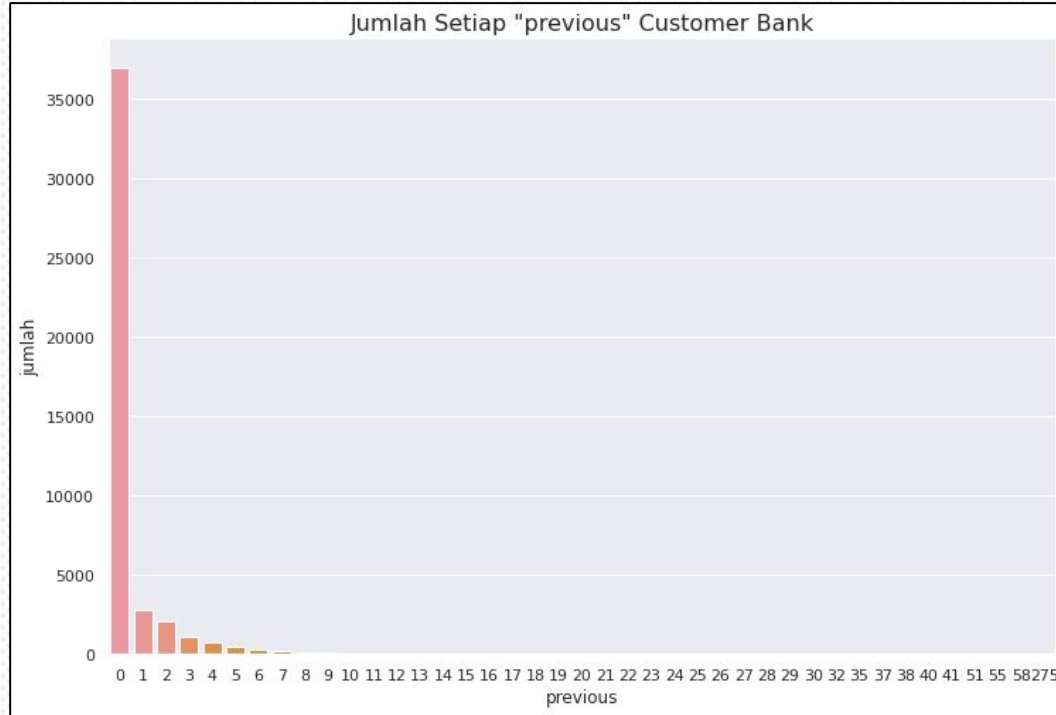
Tidak ada features numerical yang berkorelasi kuat dan dapat dilihat juga bahwa pada features numerical tidak ada yang redundant

Insight For Preprocessing (1/2)



Meskipun dapat dilihat bahwa bulan May memiliki persentase customer yang membuka deposito paling tinggi dibandingkan dengan bulan lainnya namun perbedaannya tidak signifikan. Bahkan tidak ada yang memiliki persentase $> 0.2\%$

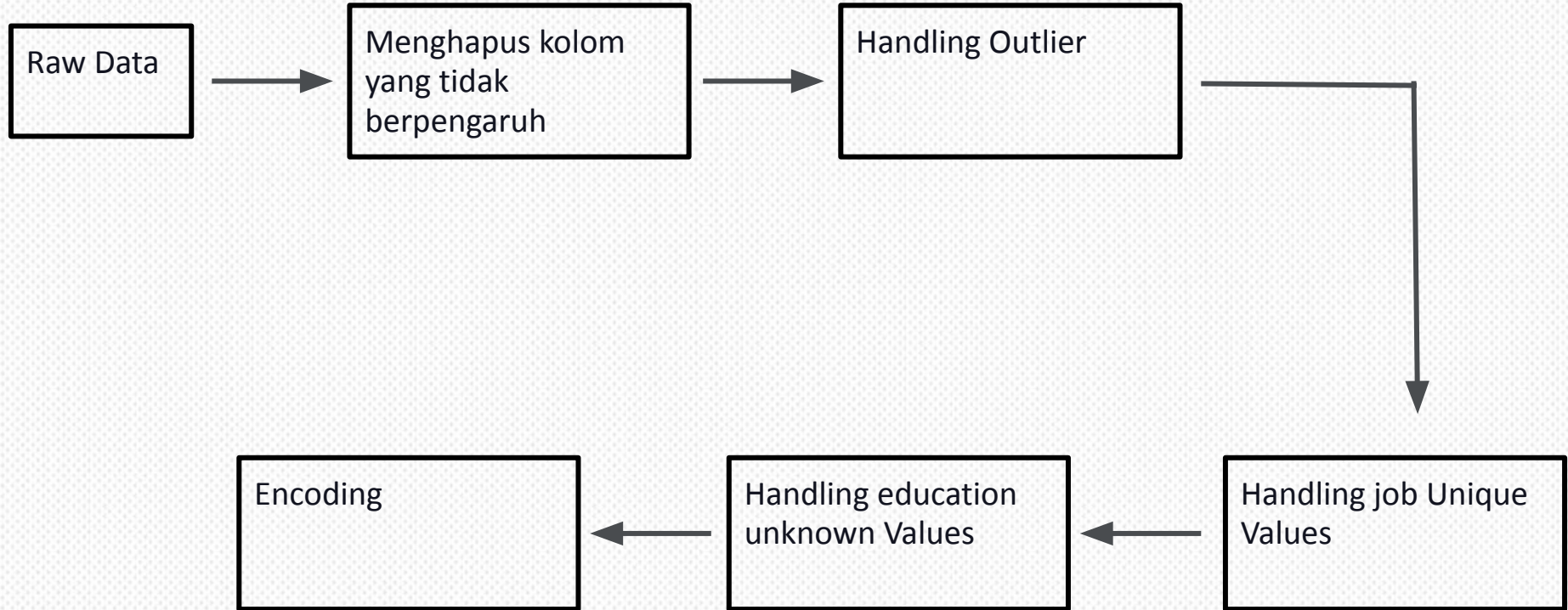
Insight For Preprocessing (2/2)



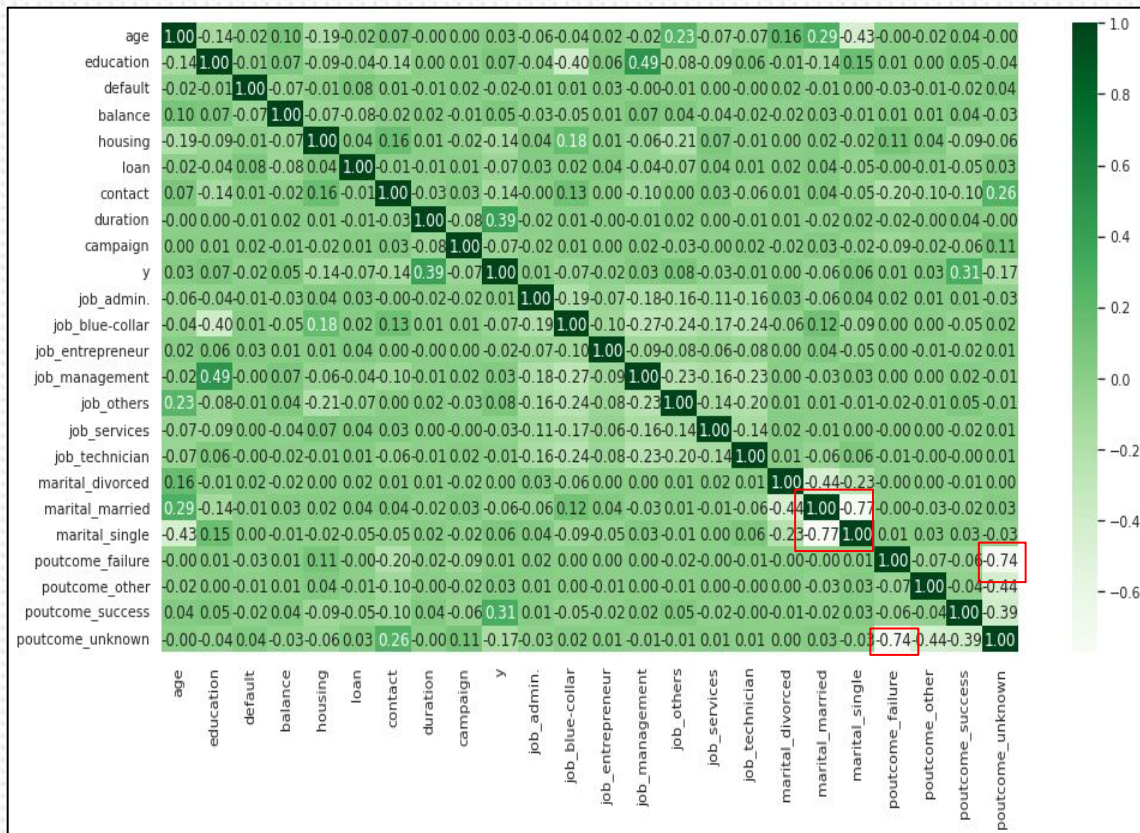
Dapat dilihat bahwa sebanyak lebih dari **35000 nasabah (tepatnya 36954)** memiliki **previous 0**, dimana hal tersebut berarti nasabah belum pernah dihubungi pada campaign sebelumnya. Artinya sekitar **81% nasabah belum pernah dihubungi** pada campaign sebelumnya.



Data Preprocessing



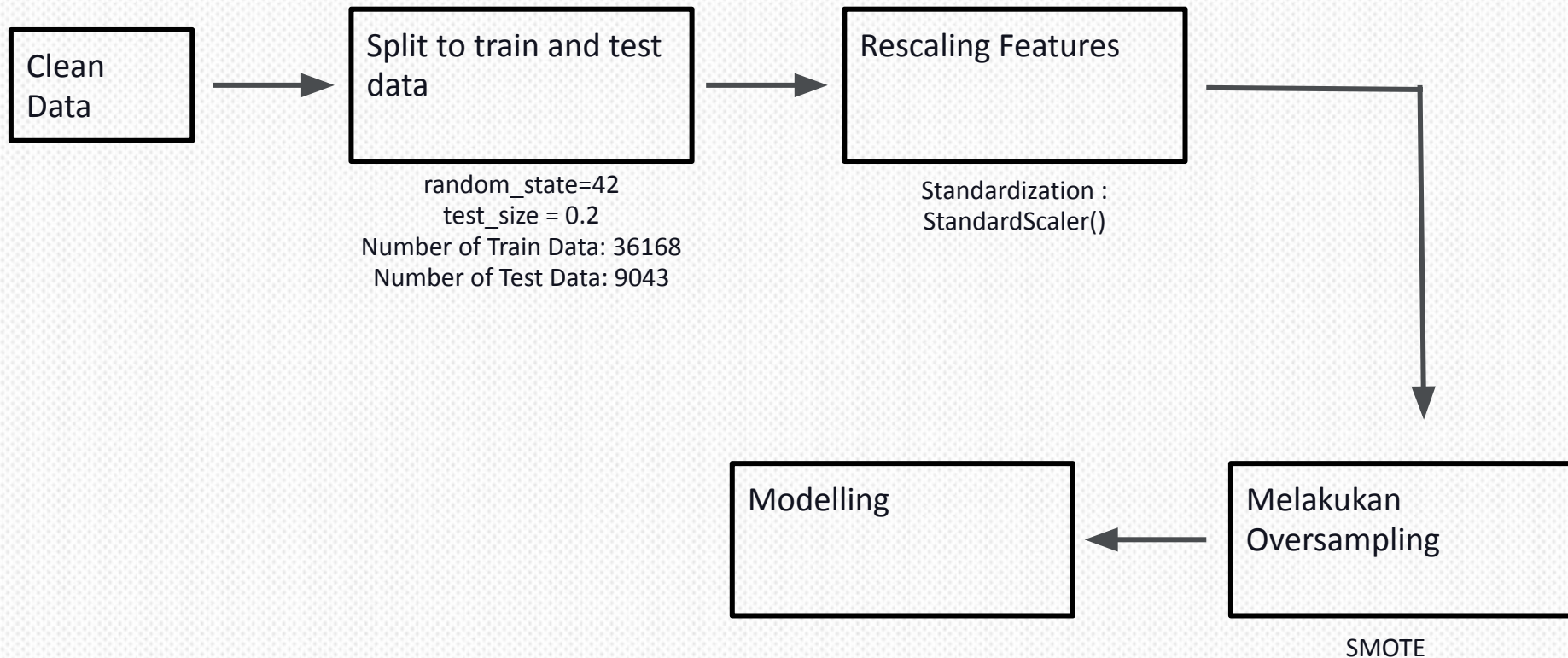
Hasil Heatmap



- Tidak ada feature yang berkorelasi kuat dengan Target.
- Terdapat fitur yang redundan yaitu : **marital_married** dan **marital_single** serta **poutcome_failure** dan **poutcome_unknown**.



Modeling



AUC Score Gap With Train Test Split & Cross Validation

Train Test Split		
Algorithm	AUC-Proba (Train)	AUC-Proba (Test)
Logistic Regression	91.05%	87.45%
Decision Tree	100%	71.05%
Random Forest	100%	88.63%
AdaBoost	91.82%	87.93%
XGBoost	93.23%	89.51%
Cross Validation		
Algorithm	AUC-Proba (Train)	AUC-Proba (Test)
Logistic Regression	91.20%	85.20%
Decision Tree	95.60%	76.40%
Random Forest	97.10%	79.10%
AdaBoost	95.80%	80.80%
XGBoost	95.30%	82.00%

Other Metrics Evaluation Score Results With Train Test Split (Test Dataset)

Test Results				
Algorithm	Recall	Precision	F1	Accuracy
Logistic Regression	77.54%	37.95%	50.96%	82.00%
Decision Tree	57.29%	34.12%	42.76%	81.50%
Random Forest	66.73%	44.36%	53.29%	85.89%
AdaBoost	77.09%	37.63%	50.57%	81.82%
XGBoost	82.22%	37.55%	51.55%	81.36%

Evaluasi Sementara

- Gap Score AUC train & test tiap algoritma melalui metode train test split maupun cross validation masih besar → indikasi Overfitting → Hyperparameter Tuning
- Selain AUC metrics yang ingin dimaksimalkan adalah Recall (False Negatif rendah = opportunity loss minimum) → Hyperparameter Tuning



Model Comparison

(AFTER Hyperparameter Tuning)

Model Evaluation After Hyperparameter Tuning

Algorithm	AUC-Proba (Train)	AUC-Proba (Test)	Recall	Precision	F1	Accuracy
Logistic Regression	87.68%	87.47%	77.64%	37.93%	50.96%	81.98%
Decision Tree	83.48%	82.89%	79.56%	27.56%	40.94%	72.31%
Random Forest	89.51%	88.43%	80.93%	35.26%	49.12%	79.77%
AdaBoost	88.90%	88.30%	82.58%	35.93%	50.07%	80.13%
XGBoost	90.25%	89.65%	82.68%	37.10%	51.22%	81.00%

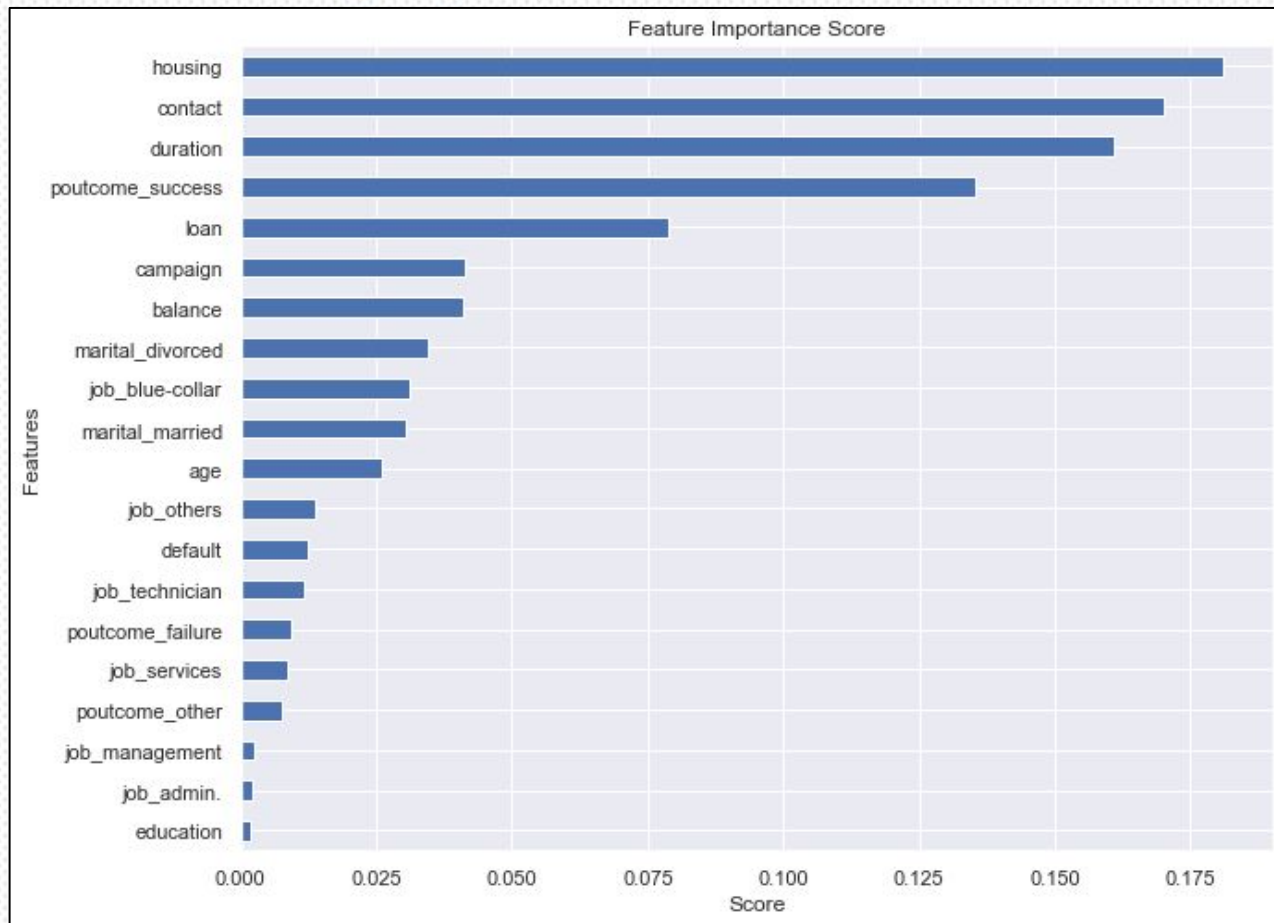
Model XGBoost dipilih karena :

- Memiliki hasil score evaluasi AUC-Proba Test tertinggi yaitu 89.65%
- Gap score evaluasi AUC-Proba Train dan AUC-Proba Test cukup rendah yaitu sekitar 0.6%.
- Memiliki hasil score evaluasi Recall yang terbaik yaitu 82.68%



Feature Importance

Buang feature
yang importance
score < 0.025

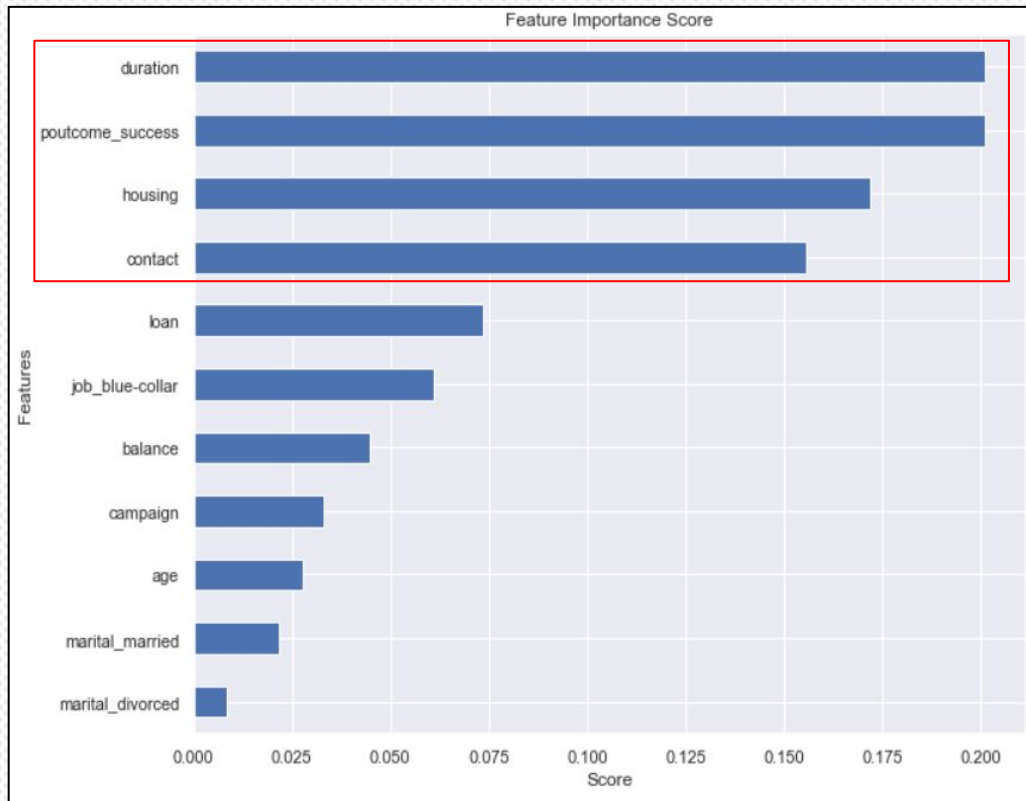


Perbandingan Sebelum dan Sesudah Membuang Fitur

XGBoost						
Condition	AUC-Proba (Train)	AUC-Proba (Test)	Recall	Precision	F1	Accuracy
Before Drop Features	90.25%	89.65%	82.68%	37.10%	51.22%	81.00%
After Drop Features	90.33%	89.66%	83.59%	36.39%	50.71%	80.39%

Maka diputuskan akan membuang feature yang memiliki feature importance score < 0.025

Final Model Feature Importance



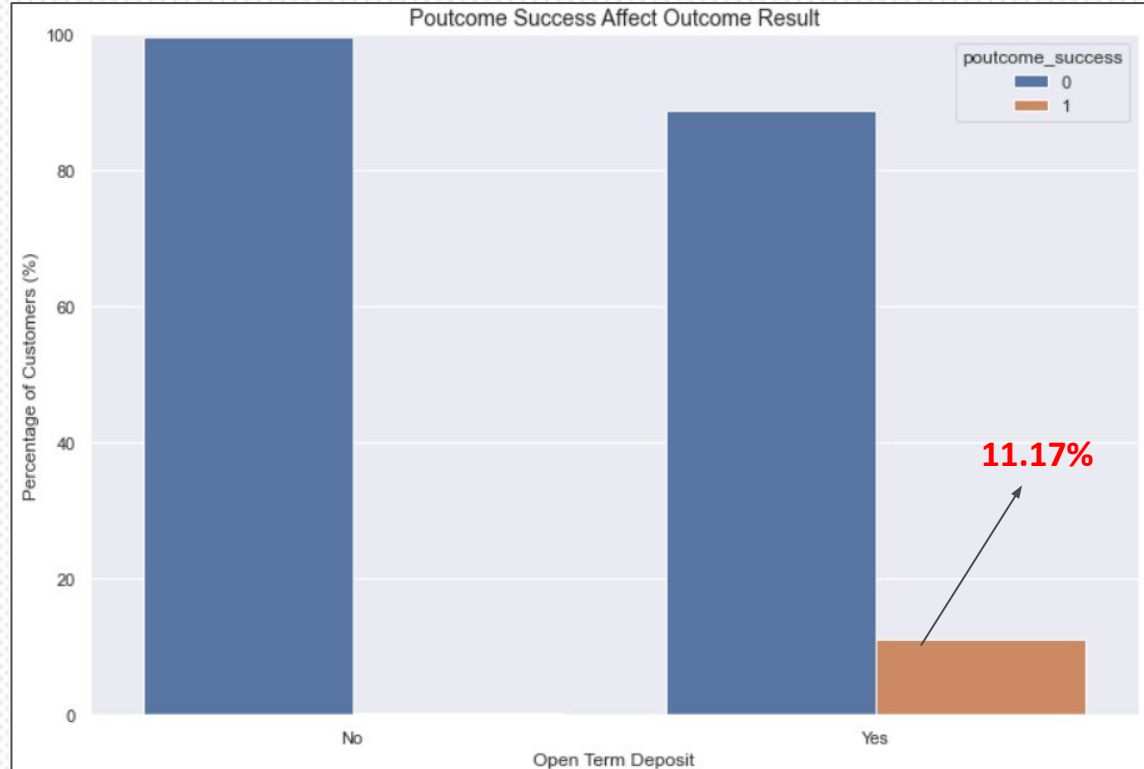
Top 4 Features:

- duration
- poutcome_success
- housing
- contact



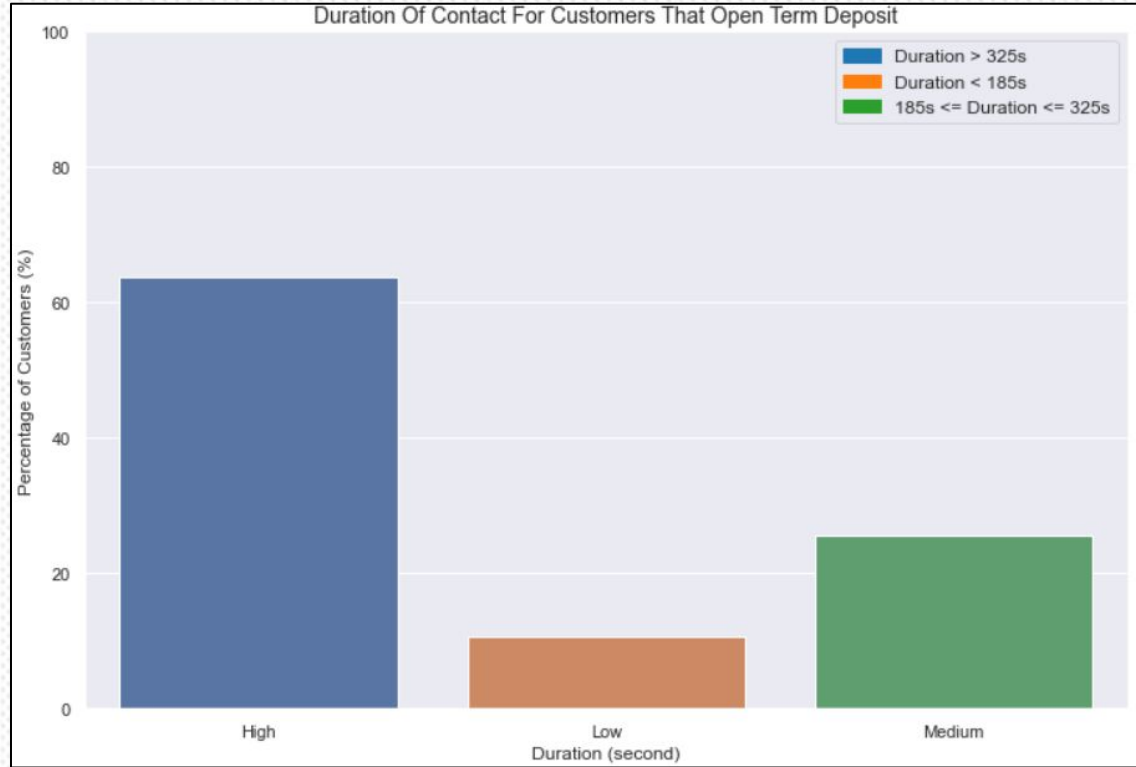
Business Insights

Business Insights (1/3)



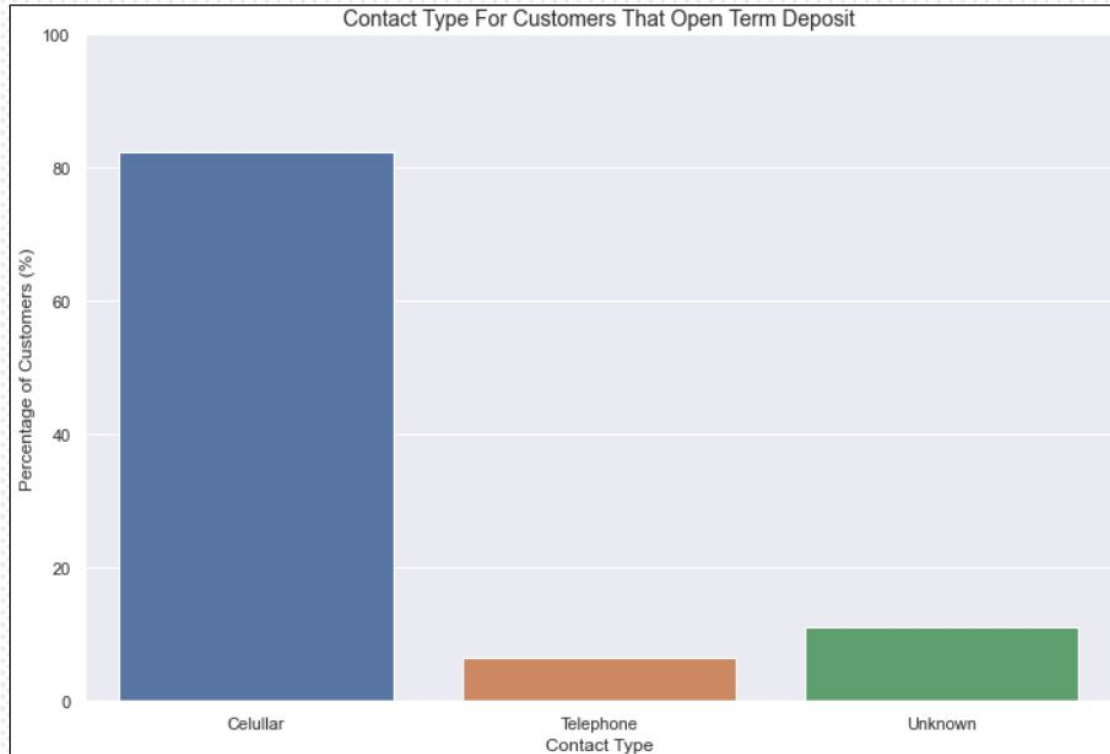
- Customer yang memutuskan membuka deposit sekitar 11% customer tersebut juga mengikuti campaign sebelumnya yang ditawarkan oleh pihak bank.
- Hasil campaign periode sebelumnya cukup mempengaruhi keputusan customer untuk mengikuti campaign kali ini.

Business Insights (2/3)



- Customer yang memutuskan membuka deposito berjangka, banyak yang memiliki `duration` dengan kategori High.
- Team marketing dari bank memiliki peran yang penting dalam membuat campaign ini berjalan dengan sukses.

Business Insights (3/3)



- Customer yang memilih untuk membuka deposito berjangka paling banyak dihubungi melalui telepon selular.
- Era sekarang ini orang-orang lebih banyak melakukan komunikasi melalui telepon selular & hampir semua orang mempunyai telepon cellular.

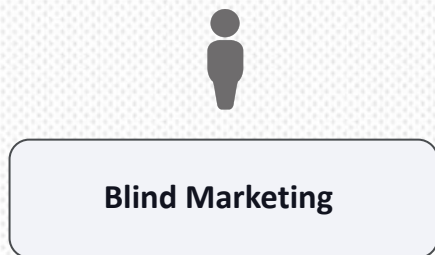


Simulation

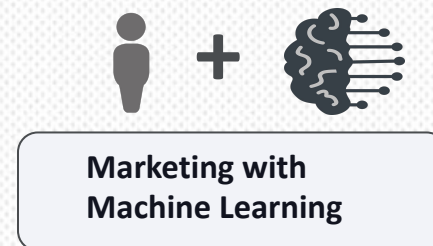
Blind Marketing vs XGBoost Modelling

Asumsi Simulasi :

- Campaign starts from 1 November 2022 - 16 December 2022 (34 effective days)
- Call duration/customers = 5 minutes
- Office hours 9:00am - 17:00pm (effective hours = 7 hours = 420mins)
- 1 Marketing Officer can call 84 customers/day



341 customers
Conversion Rate = 3.77%



912 customers
Conversion Rate = 10.08%

Increase 6.31% conversion rate



Business Recommendation

Business Recommendation

- Menerapkan model XGBoost untuk memaksimalkan kampanye deposito berjangka.
- Bank mulai sekarang harus mulai memaksimalkan setiap campaign yang ada kepada nasabah.
- Bank harus mulai meningkatkan kualitas pelayanan yang dimiliki oleh tim pemasarannya.
- Bank dapat mulai mengubah strategi kampanye dengan memprioritaskan melakukan kontak melalui telepon seluler.