

Optimizing Marketing Costs and Conversion Rates: Predicting Bank Customer Deposits from Telemarketing Campaign



Outline.

- ☐ **Problem Statement.**
- ☐ **Exploratory Data Analysis.**
- ☐ **Data Preprocessing.**
- ☐ **Machine Learning Model.**
- ☐ **Model Simulation.**
- ☐ **Recommendation.**



A circular collage of various business and technology icons in a light gray tone, including a coffee cup, line graphs, a magnifying glass, a smartphone, a rocket, a padlock, a hierarchy chart, a laptop, a piggy bank, gears, a bar chart, a pie chart, a lightbulb, a speech bubble, a robot, a mail icon, a Wi-Fi symbol, a search bar, a dollar sign, a Euro symbol, and a bar chart. The word "BUSINESS" is written in a circle in the center of the collage.

Problem Statement.

Background.



Bank Portugal Memiliki **45.211 nasabah** yang telah dihubungi Telemarketing Campaign **deposito berjangka**. Pada bisnis bank, **deposito berjangka** menjadi salah satu sumber pendapatan besarnya.

Namun.....

Background.



Namun pada **tahun 2008 hingga 2012*** negara - negara di Eropa termasuk Portugal sedang **mengalami Resesi Hebat** sehingga negara-negara tersebut menghadapi banyak tantangan ekonomi.

Menariknya.....

Background.



Menariknya, hal ini membuat Bank Metaverse berusaha mengamankan cash flow-nya dengan fokus pada strategi yang lebih terarah dan efisien dalam memilih target nasabah yang memiliki potensi terbesar.

Tetapi, ternyata.....

Current Problem.

Tetapi, Ternyata karena **keterbatasan manusia**, untuk memprediksi nasabah mana yang akan membeli produk deposito berjangka masih seringkali salah. sehingga menyebabkan

1. **Biaya Telemarketing Campaign cukup besar**, menghabiskan hingga **€3.210** untuk menawarkan deposito berjangka ke 45.211 nasabah.
2. **Return of Investment** yang didapatkan dari nasabah dinilai masih rendah. ROI telemarketing campaign baru **mencapai 16.01%**
3. **Conversion Rate** nasabah yang menyetujui deposito berjangka kurang maksimal karena **hanya 11,7%** dari total nasabah.



*Asumsi:

1. cost/customer = €0,071
portugal)
2. return/customer = €0,71

(biaya telfon salah satu provider di

(10 kali lipat dari biaya)

Goals.



Menurunkan campaign cost
yang dikeluarkan sebesar 50%



Meningkatkan Return of Investment
sebanyak 50%



Meningkatkan conversion rate
sebanyak 50%

Objective.



Membangun model Machine Learning yang dapat memprediksi nasabah yang potensial untuk menerima campaign.

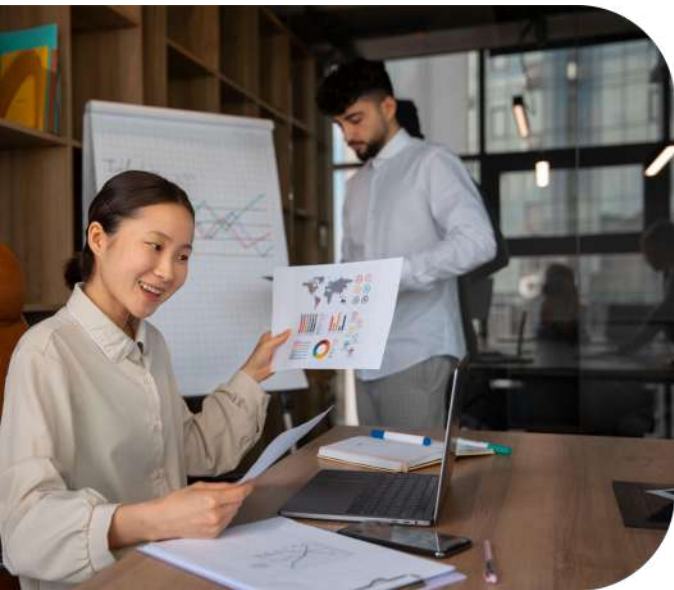


Menemukan faktor penting dan karakteristik utama dari nasabah yang menerima campaign dan menyetujui produk deposito yang ditawarkan



Memberikan rekomendasi yang dapat meningkatkan efektivitas perusahaan dalam memberikan campaign kepada nasabah terpilih

Business Metrics.



Return on Investment



Revenue



Campaign Cost



Profit



Cost Reduction



Conversions



Potential Revenue Loss



% Leads Contacted

Outline.

- ☒ **Problem Statement.**
- ☐ **Exploratory Data Analysis.**
- ☐ **Data Preprocessing.**
- ☐ **Machine Learning Model.**
- ☐ **Model Simulation.**
- ☐ **Recommendation.**

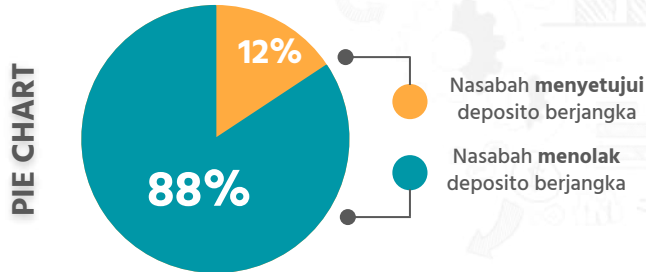


A circular collage of various business and data-related icons in a light gray, hand-drawn style. The icons include a coffee cup, a line graph, a magnifying glass, a search bar, a rocket, a padlock, a smartphone, a bar chart, a gear, a piggy bank, a speech bubble, a robot, a mail envelope, a Wi-Fi symbol, a percentage sign, a lightbulb, a document, a bar chart, a gear, a piggy bank, a speech bubble, a robot, a mail envelope, a Wi-Fi symbol, a percentage sign, a lightbulb, a document, and the word "BUSINESS" in a circle.

Exploratory Data Analysis.

Data Overview.

1. Dataset terdiri dari **45.211 baris** dan **17 kolom**
2. Terdapat **7 data numerik** dan **10 data kategorikal**
3. Tidak terdapat data **null** atau **kosong**



Target Feature.

(Feature “y”)



“Yes”

Nasabah **setuju** untuk mengambil deposito berjangka

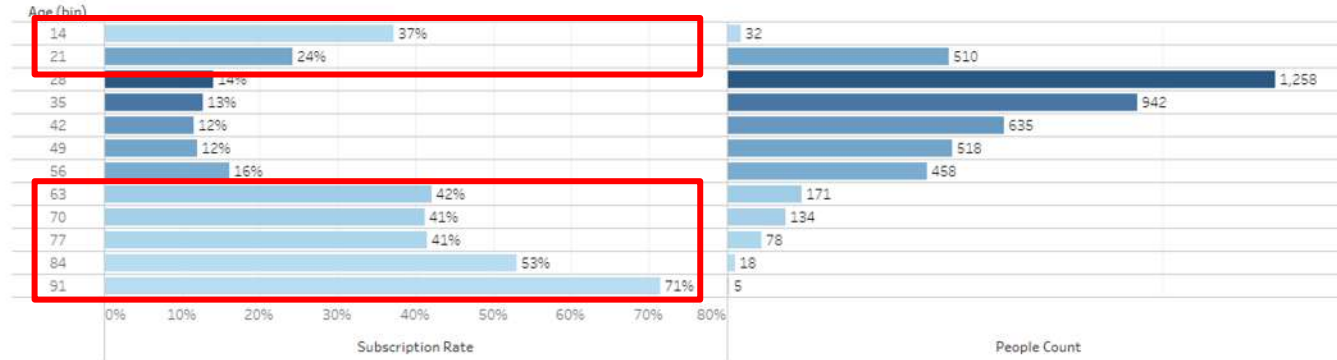


“No”

Nasabah **menolak** untuk mengambil deposito berjangka

Data Insight - Age.

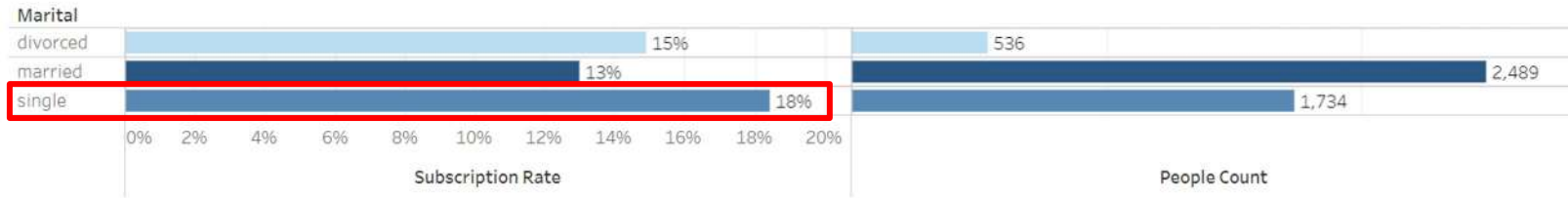
Is there a relationship between Age (bin), and **subscription rate** and **count**?



1. Nasabah dengan umur **diatas 63 tahun** cenderung memiliki ***subscription rate* lebih besar** daripada umur di bawahnya.
2. Selain itu, *subscription rate* untuk umur dibawah 28 tahun juga lebih tinggi dibandingkan dengan umur 28-62 tahun.

Data Insight - Marital.

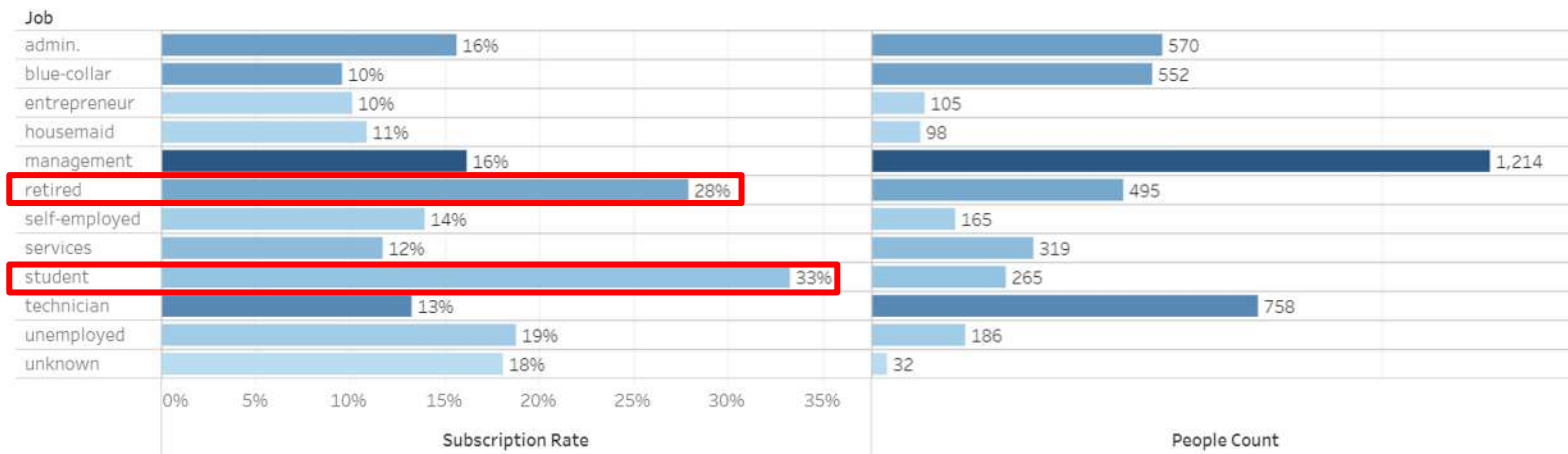
Is there a relationship between **Marital status**, and **subscription rate** and **count**?



1. Dapat dilihat bahwa **status marital** yang memiliki **subscription rate terbesar** adalah **status single**. Jika dilihat dari frekuensi maka status menikah memiliki nilai terbanyak.

Data Insight - Job.

Is there a relationship between **Job**, and **subscription rate** and **count**?

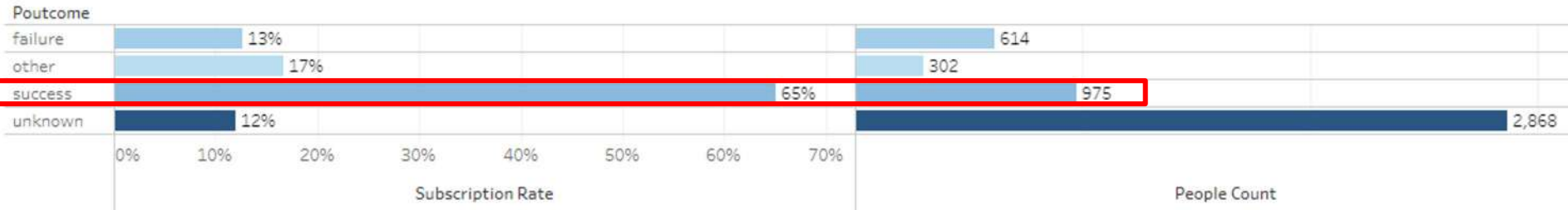


students & retired lebih tertarik untuk deposito dimana :

1. ***students*** yang mencoba membangun kebiasaan menabung,
2. ***retired*** yang mencari cara untuk mengelola tabungan mereka dengan bijak dimasa pensiun.

Data Insight - P-outcome

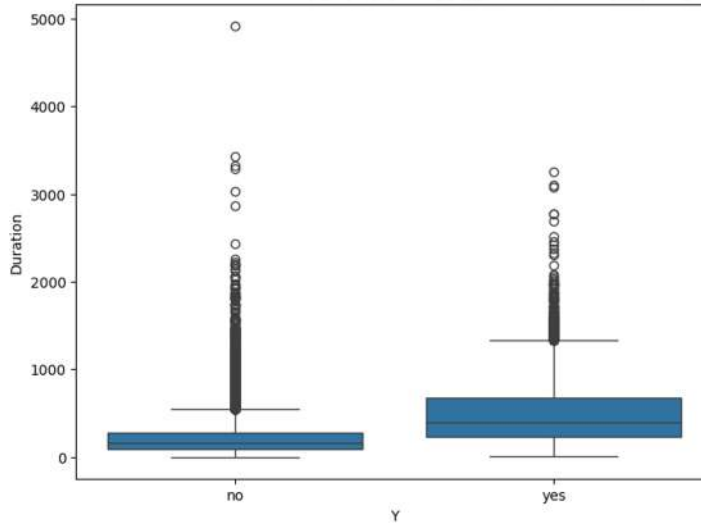
Is there a relationship between Previous campaign **outcome**, and **subscription rate** and **count**?



P-outcome *success* lebih tertarik untuk deposito dimana :

1. ***trust*** yang telah dibangun sebelumnya berdasarkan penggunaan produk,
2. ***keinginan untuk menambah profit*** yang ada pada nasabah yang menggunakan produk yang lain

Data Insight - Duration.



1.

Durasi menelpon antara *subscriber* dan yang tidak itu **cenderung lebih tinggi** yang *subscriber*,

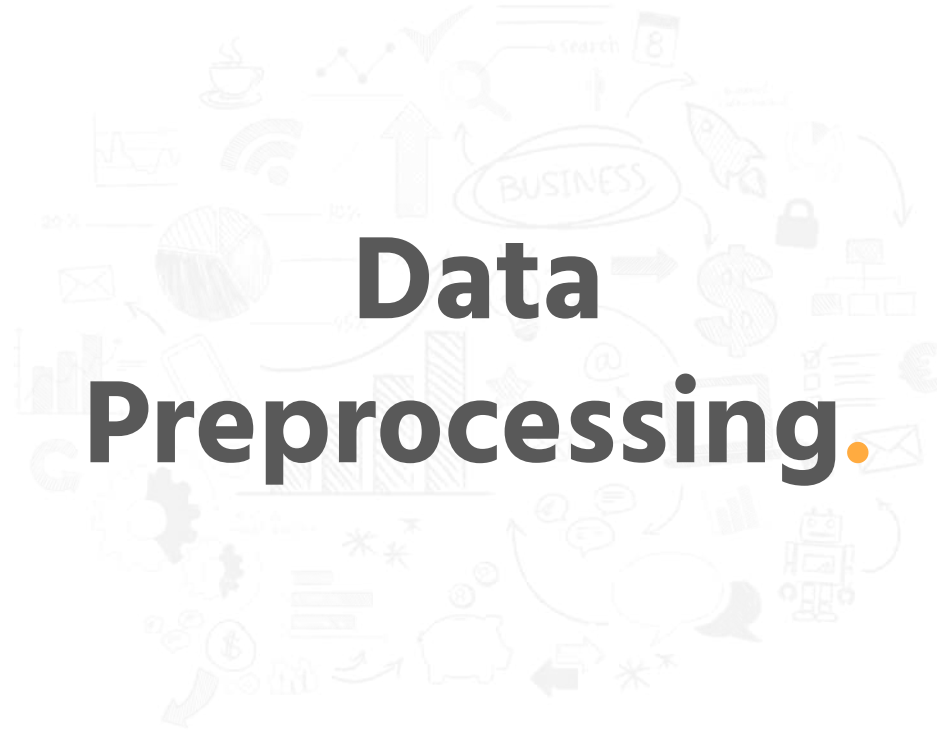
Dimana dilihat durasinya

- sedikit distribusinya pada durasi < 50 detik
- padat pada angka sekitar 1700 detik.

Outline.

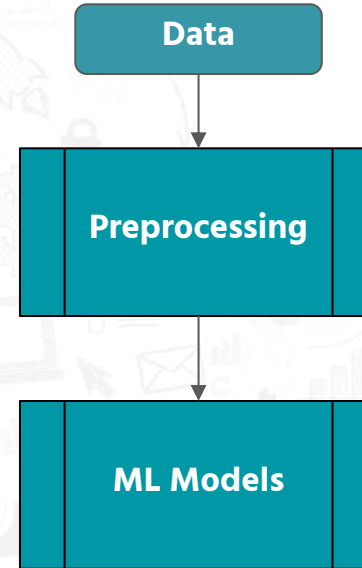
- ☒ **Problem Statement.**
- ☒ **Exploratory Data Analysis.**
- ☐ **Data Preprocessing.**
- ☐ **Machine Learning Model.**
- ☐ **Model Simulation.**
- ☐ **Recommendation.**





Data Preprocessing.

1. Handle Negative Values
2. Handle Unknown Values
3. Perform Feature Extraction
4. Handle Outliers
5. Apply Label and One Hot Encoding
6. Handle Class Imbalance



Handle Unknown Values.

1. **Contact** unknown values pada contact dihapus.
2. **Education** unknown pada education diubah menjadi other.
3. **Poutcome** unknown pada poutcome diubah menjadi other.
4. **Job** unknown pada job diubah menjadi other.
5. **Poutcome** unknown dan other values pada poutcome dihapus.

Perform Feature Extraction.

1. **day** mengubah day menjadi week
2. **Pdays dan previous** membuat fitur baru contacted berdasarkan fitur pdays dan previous.
Kemudian menghapus fitur pdays dan previous

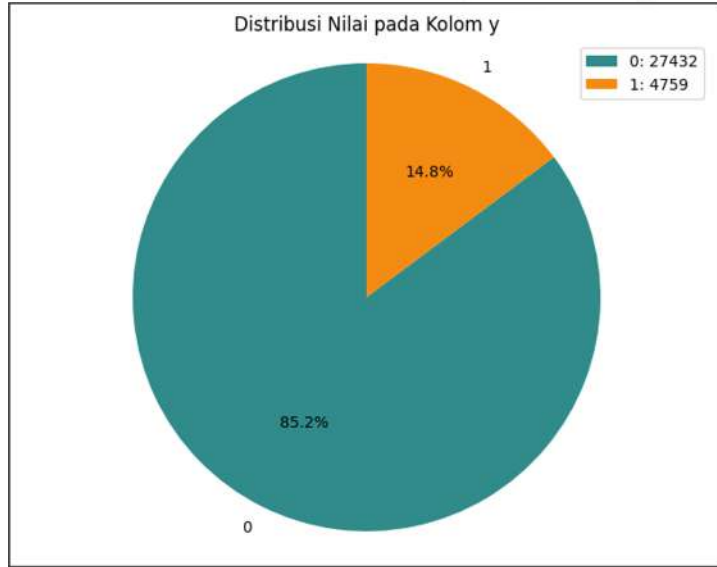
Handle Outliers.

1. **Robust Scaler** dipilih karena kemampuannya yang baik dalam menangani outlier dan memperbaiki distribusi data, sehingga memungkinkan model machine learning untuk belajar dengan lebih baik dari data yang diproses menggunakan skala yang konsisten dan lebih tahan terhadap variasi dalam data.

Perform Encoding.

1. **Label Encoding** : Fitur yang memakai label encoding adalah default, housing, loan, dan contacted.
2. **One Hot Encoding** : Fitur yang memakai label encoding adalah job, marital, education, contact, day, month, poutcome.

Handle Class Imbalance.



Metode **SMOTEENN**

SMOTEENN membantu dalam menciptakan dataset yang seimbang dan bersih, yang dapat meningkatkan performa model, terutama dalam situasi dengan data tidak seimbang seperti prediksi deposito pelanggan

Outline.

- ☒ **Problem Statement.**
- ☒ **Exploratory Data Analysis.**
- ☒ **Data Preprocessing.**
- ☐ **Machine Learning Model.**
- ☐ **Model Simulation.**
- ☐ **Recommendation.**

Machine Learning Model.

Model Evaluation

		Aktual	
		Tidak Setuju	Setuju
Prediksi	Tidak Setuju	True Negative	False Negative
	Setuju	False Positive	True Positive

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Recall

Dengan meminimalkan jumlah pelanggan yang menolak membuka deposito, kita dapat fokus untuk meningkatkan revenue (orang yang mau membuka deposito). Oleh karena itu, fokus pada optimasi Recall akan membantu kita mencapai tujuan utama yaitu meningkatkan revenue keseluruhan dengan cara menjaga sebanyak mungkin pelanggan yang ingin membuka rekening deposito.

12%

88%



Nasabah **menyetujui** membuka rekening deposito

Nasabah **menolak** membuka rekening deposito

Hasil Data Split

Model	Recall Test	Recall Train
Logistic Regression	0.85	0.82
RandomForest Classifier	0.82	0.80
XGBoost Classifier	0.71	0.78

Algoritma terbaik adalah **Logistic Regression** jika dibandingkan model lainnya. Dikarenakan nilai score recall sudah diatas threshold 80%, maka kita akan lihat nilai precision untuk pertimbangan biaya yang akan dikeluarkan pada saat eksekusi campaign marketing.

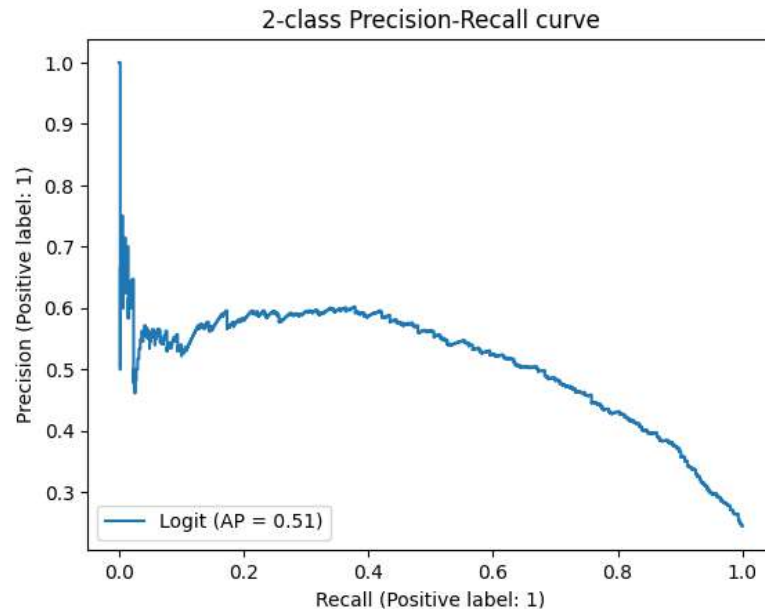
Aktual	Tidak setuju	1070	374
	Setuju	71	396
		Tidak Setuju	Setuju

Prediksi

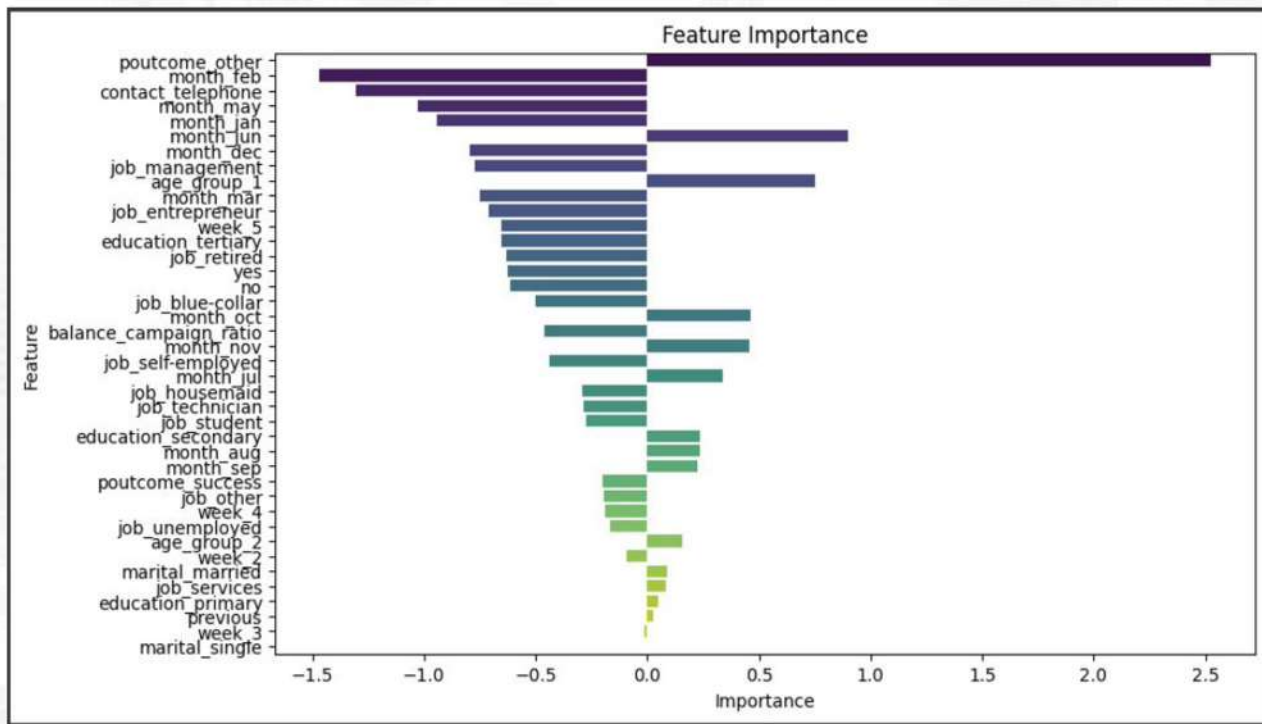
Hasil Data Split

Model	Recall	Precision
Logistic Regression (threshold precision recall tuning)	0.80	0.23

Setelah kita menetapkan memakai base model logistic regression. kita lakukan tuning threshold untuk menyesuaikan dengan goals perusahaan.



Feature Importance




Fitur **poutcome_other**, **month_jun**, dan **job_management** menjadi fitur penting yang menentukan akan konversinya *leads*.

Fitur **month_feb**, **contact_telephone**, dan **month_may** menjadi fitur penting yang menentukan tidak akan konversinya *leads*.

Outline.

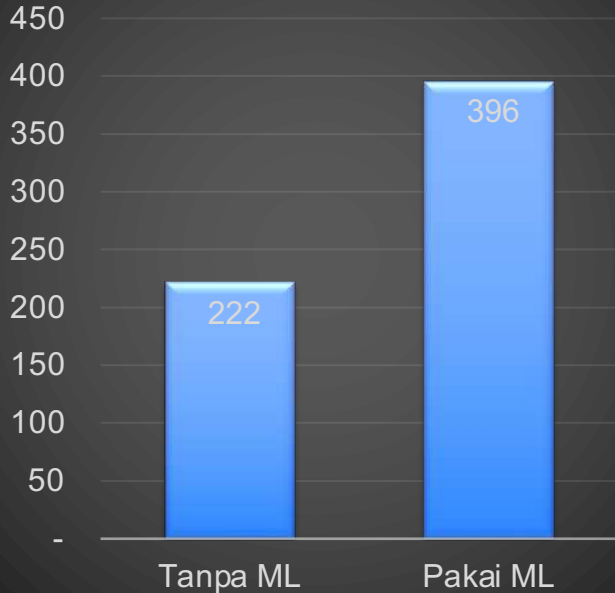
- ☒ **Problem Statement.**
- ☒ **Exploratory Data Analysis.**
- ☒ **Data Preprocessing.**
- ☒ **Machine Learning Model.**
- ☐ **Model Simulation.**
- ☐ **Recommendation.**





Model Simulation

Conversion



Conversion

Conversion bisa meningkat sebanyak 174 Customer karena bisa memprediksi customer lebih baik.

Dengan memakai model ML...

Conversion Tanpa ML = Total Target Nasabah * 11,6%

Conversion Pakai ML = TP

*Note: 11,6% dari conversion rate eksisting pada data
Total Target Nasabah adalah sum dari confusion matrix

Cost (€)



Cost

Sekitar
60%

**Biaya
Marketing
bisa
direduksi**

Cost Tanpa ML = Total Target Nasabah * biaya

Conversion Pakai ML = TP * biaya

*Note: biaya = biaya telfon per menit di Portugal sebesar
€0,071 per customer

Goals

Conversion Rate

Sekitar
5x

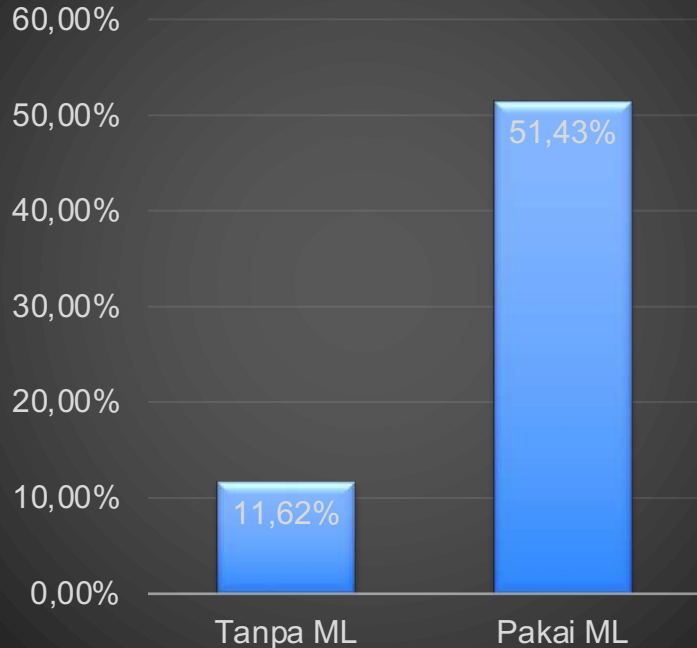
**Conversion
Rate
meningkat**

Conversion Rate Tanpa ML = Total Target Nasabah * 11,6% / Total Target Nasabah

Conversion Rate Pakai ML = $TP / (TP + FP)$

*Note: 11,6% dari conversion rate eksisting pada data Total Target Nasabah adalah sum dari confusion matrix

Conversion Rate



Financial Metrics

Profit (€)



Profit

Sekitar
8 X
lipat
Profit
meningkat

Profit Tanpa ML = revenue - cost

Conversion Pakai ML = revenue – (cost + potential lost revenue)

*Note: revenue per customer = 10 x biaya per customer

revenue = conversion x revenue per customer

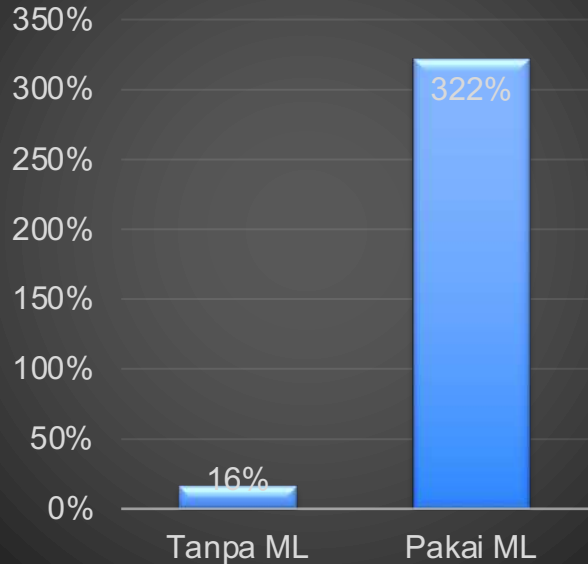
Potential lost revenue = TN * revenue per customer

Goals

Return on Investment (ROI)

Sekitar
8 X ROI meningkat
lipat

ROI (€)



ROI Tanpa ML = Profit / cost

ROI Pakai ML = profit / cost

Outline.

- ✓ **Problem Statement.**
- ✓ **Exploratory Data Analysis.**
- ✓ **Data Preprocessing.**
- ✓ **Machine Learning Model.**
- ✓ **Model Simulation.**
- ☐ **Recommendation.**



A circular collage of various business and technology icons in a light gray, hand-drawn style. The icons include a coffee cup, a line graph, a magnifying glass, a search bar, a rocket, a padlock, a laptop, a robot, a piggy bank, a bar chart, a pie chart, a lightbulb, a gear, a Wi-Fi symbol, a smartphone, a speech bubble, a dollar sign, a bar chart, a pie chart, a line graph, a magnifying glass, a search bar, a rocket, a padlock, a laptop, a robot, a piggy bank, a bar chart, a pie chart, a lightbulb, a gear, a Wi-Fi symbol, a smartphone, a speech bubble, a dollar sign, and the word "BUSINESS" in a circle.

Recommendation.

Business Recommendation.



Mengimplementasikan telemarketing yang ditargetkan menggunakan model machine learning.



Memberikan return yang lebih tinggi pada bulan Juni-November



Kegiatan telemarketing lebih fokuskan ke calon pelanggan yang memiliki status tidak menolak (other & success)

memberikan produk yang mendukung kebutuhan customer di rentang usia 28 - 63 (return yang tinggi, deposito pensiun, deposito tabungan investasi, dsb)

Outline.

- ✓ **Problem Statement.**
- ✓ **Exploratory Data Analysis.**
- ✓ **Data Preprocessing.**
- ✓ **Machine Learning Model.**
- ✓ **Model Simulation.**
- ✓ **Recommendation.**



Terima kasih.

Rakamin Final Project

