

Banking Dataset - Marketing Strategy

peerData



1. Muhamad Iqbal
2. A Nahda La Roiba
3. Ilham Maulana
4. Clara Natalie S
5. R. Rani Indah Salamah
6. Eka Apriyani
7. Sekar Ayu Larasati
8. Firstandy Edgar Dhafa



Table of contents

01

Background

02

Dataset

03

EDA & Insights

04

**Data
Preparation**

05

Modelling

06

Interpretation

07

Recommendation



01

Background

Informasi mengenai permasalahan yang ingin diselesaikan beserta goal dan objectives yang ingin dicapai



Problems



INEFFECTIVE CAMPAIGN

Hanya **11%** dari total nasabah yang **setuju untuk berlangganan** dari total nasabah yang dihubungi.

IMPACT



HIGH MARKETING COST!!

Perlu **efisiensi biaya marketing** karena kondisi ekonomi yang belum pulih akibat pandemi..



Roles



DATA SCIENTIST (MARKETING SUPPORT)

Responsibilities:

Membuat **model machine learning** untuk memprediksi pelanggan yang berpotensi berlangganan deposito.



Goals

1



FIND TARGETS

Mencari target pelanggan yang **potensial** berlangganan deposito.

2



COST REDUCTION

Mengurangi biaya marketing pada campaign deposito berjangka tanpa kehilangan potensi pendapatan

3

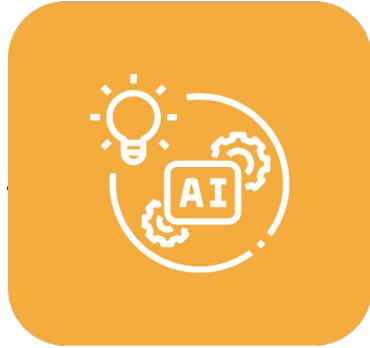


FEATURES IMPORTANCE

Mencari **fitur-fitur penting** yang mendasari keputusan berlangganan.



Objectives



MACHINE LEARNING

Menentukan algoritma **Machine Learning (ML)** dengan nilai *evaluation metric* **tertinggi** dengan model yang dihasilkan.



COST REDUCTION

Menghitung jumlah **biaya dan profit** setelah pengaplikasian rekomendasi **ML** dan membandingkannya dengan **tanpa pengaplikasian ML** dan **filtering manual**

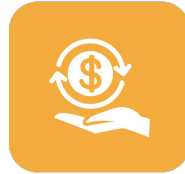


RECOMMENDATION

Menjabarkan **rekomendasi bisnis** berdasarkan fitur-fitur penting yang mendasari rekomendasi ML.



Business Metrics



Cost reduction

Mengurangi jumlah biaya yang dikeluarkan tanpa kehilangan banyak **potensi** pendapatan. Impact dihasilkan dari rekomendasi **machine learning**.





02

Dataset

BANKING MARKETING TARGET



About the dataset



Data ini terkait kampanye pemasaran langsung lembaga perbankan **Portugal** untuk memprediksi apakah klien akan **berlangganan deposito** atau tidak.



45.211 baris dan **18 kolom** diurutkan berdasarkan tanggal (dari **Mei 2008 hingga November 2010**)



Variabel **target**, ditandai sebagai **"y"**. "yes" jika klien berlangganan deposito berjangka dan "no" jika klien tidak berlangganan.



03

EDA & Insight

Descriptive, Univariate, and Multivariate Statistics Summary



Persentase Deposito Berjangka

↑ 4X

JOB

"**student**" memiliki kemungkinan **hampir 4 kali lebih tinggi** dibandingkan dengan kategori pekerjaan lainnya (dalam hal ini, "blue-collar").

↑ 1.4X

MARITAL

"**single**" memiliki kemungkinan sekitar **1.477 kali lebih tinggi** dibandingkan dengan kategori "married".

↑ 1.7X

EDUCATION

"**tertiary**" memiliki kemungkinan sekitar **1.738 kali lebih tinggi** dibandingkan dengan kategori "primary".

↑ 1.8X

DEFAULT

"**no**" (**tanpa default**) memiliki kemungkinan sekitar **1.848 kali lebih tinggi** dibandingkan dengan kategori "yes" (dengan default).

Persentase Deposito Berjangka

↑ 2X

HOUSING

"no" (tanpa kepemilikan rumah) memiliki kemungkinan sekitar **2.167 kali lebih tinggi** dibandingkan dengan kategori "yes" (dengan kepemilikan rumah).

↑ 1.9X

LOAN

"no" (tanpa pinjaman) memiliki kemungkinan sekitar **1.893 kali lebih tinggi** dibandingkan dengan kategori "yes" (dengan pinjaman).

↑ 3.6X

CONTACT

"cellular" memiliki kemungkinan **3.661x** lebih tinggi dibandingkan dengan kategori "unknown".

↑ 7X

POUTCOME

"success" memiliki kemungkinan **7.069x** lebih tinggi dibandingkan dengan kategori "unknown".

4 fitur ini bisa menjadi dasar dilakukannya filtering manual untuk melihat cost-reduction yang dihasilkan



04

Data Preparation

Data Cleansing | Feature Engineering





Missing Value
(0 Missing Values)



Duplicated Data
(0 Duplicated Data)



Outliers

Keep ✓
Drop ✗



Feature Transformation
Mengubah nilai negatif
pada "**Balance**" menjadi
nilai absolut



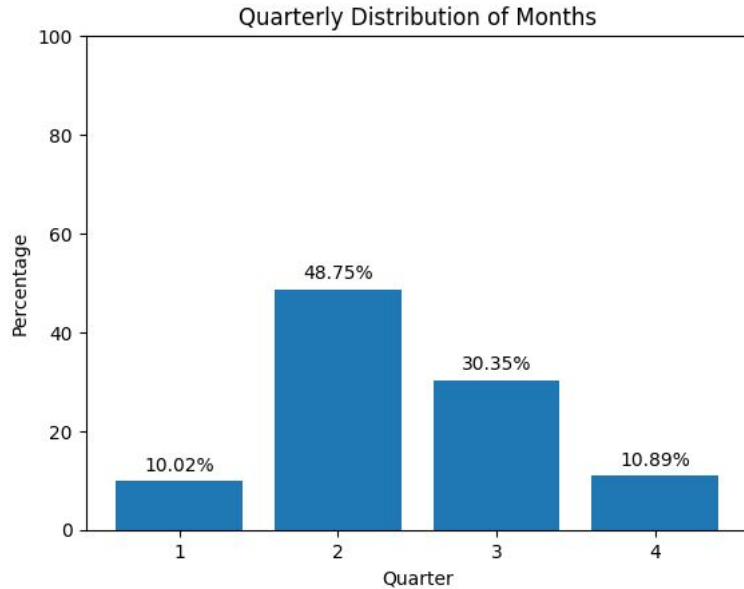
Feature Encoding

Label Encoding ✓
OHE ✓



Handling Class Imbalance
SMOTE

Feature Extraction



Konversi kolom “Month” ke kuartal

Sebagian besar customer
dihubungi pada Kuartal 2
(April, Mei, Juni)

Feature Selection (1)

menghilangkan fitur-fitur yang tidak relevan atau tidak digunakan dalam memprediksi target variabel

01

'pdays'

Korelasi tinggi dengan 'previous'

02

'age'

Kurang berkontribusi terhadap model

03

'balance'

Kurang berkontribusi terhadap model

04

'previous'

Bias dengan fitur **'campaign'**

05

'day'

Tidak memberikan info penting

06

'education'

Berisi nilai **"unknown"** yang tidak terdefinisi

07

'contact'

Berisi nilai **"unknown"** yang tidak terdefinisi



Feature Selection (2)

RFE (Recursive Feature Elimination)

Melakukan seleksi fitur lanjutan menggunakan RFE yang didasari pada kontribusi tiap fitur terhadap model, sehingga tersisa **18 fitur** untuk dipakai pada modeling :

1. Housing
2. loan
3. duration
4. campaign
5. job_admin.
6. job_blue-collar
7. job_management
8. job_technician
9. marital_divorced

10. marital_married
11. marital_single
12. poutcome_failure
13. poutcome_other
14. poutcome_unknown
15. month_quartal_1
16. month_quartal_2
17. month_quartal_3
18. month_quartal_4



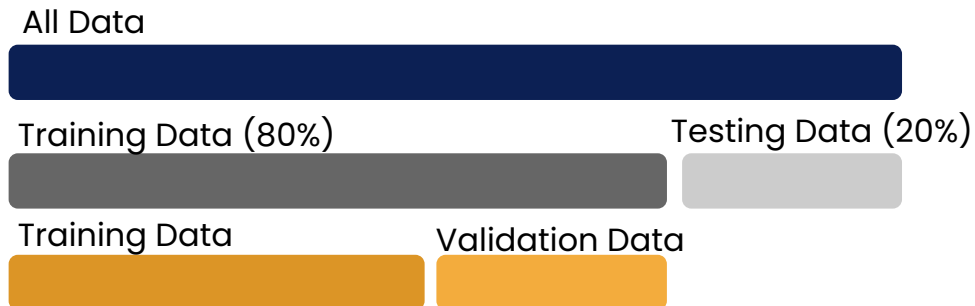
05

Modeling

Tiga jenis model yang akan kami gunakan adalah Logistic Regression, Random Forest, dan XGBoost.



MODELING STRATEGIES



Strateginya adalah, pertama, membagi data menjadi dua bagian: Training Data(80%) dan Testing Data (20%). Dari Training Data, kita akan membaginya lagi menjadi dua bagian: Training Data dan Validation Data.

Data Validasi akan digunakan untuk mengevaluasi kemampuan prediktif model.



Model	Precision Score	Precision Cross-Val (Train)	Precision Cross-Val (Test)
logistic	0.627692	0.958771	0.958853
RandomForest	0.471471	0.996779	0.925333
XGBoost	0.558524	0.964807	0.946204
Tuned_LogisticRegressor	0.623704	0.956841	0.956622

Precision Recall Threshold

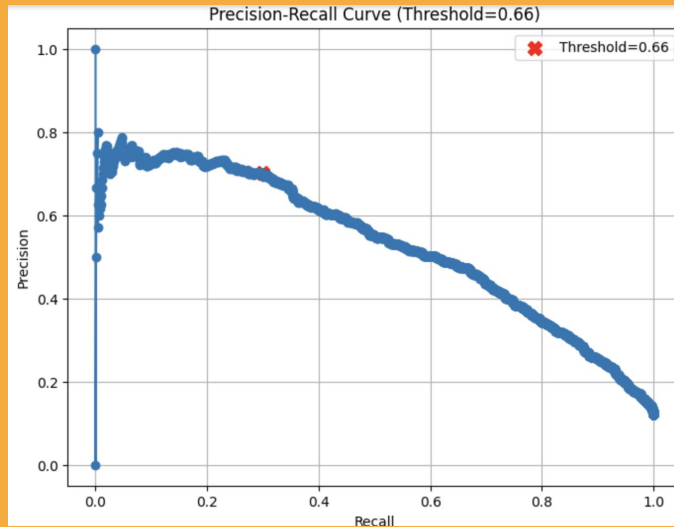
0.701299 0.298343 0.662721

0.700651 0.297422 0.664309

0.702174 0.297422 0.664853

0.701525 0.296501 0.666774

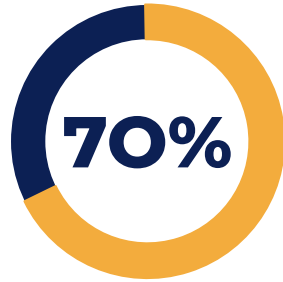
0.700873 0.295580 0.667835



Dengan mencari nilai **precision 0.70**, nilai **threshold minimal** yang tersedia adalah sebesar **0.662721**. Alhasil, nilai **precision naik menjadi 0.7012** dengan recall 0.2983 atau 0.3.

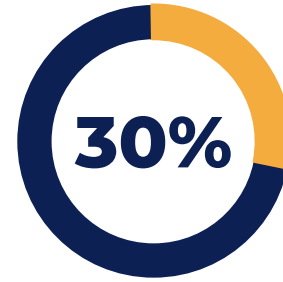


Logistic Regression



Precision

- **Meminimalisir** False Positive (**Memprediksi pelanggan akan tertarik berlangganan padahal tidak.**)
- Dari pelanggan yang diprediksi akan berlangganan, sekitar 70% di antaranya memang benar-benar akan berlangganan



Recall

- **Meminimalisir** False Negative (**Memprediksi pelanggan tidak berlangganan padahal tertarik.**)
- Dari pelanggan yang diprediksi tidak berlangganan, hanya sekitar 30% yang benar-benar tidak berlangganan.



Precision is our main metric

Dengan menggunakan **precision**, diharapkan false positive akan berkurang, sehingga perusahaan tidak perlu mengeluarkan **biaya lebih besar** (misal biaya telepon, campaign, karyawan) pada nasabah yang **tidak tertarik** pada deposito berjangka.

Kami **sedikit mengabaikan false negative** karena kami berasumsi bahwa nasabah yang tertarik akan langsung menghubungi kami tanpa kami harus melakukan campaign.

Actual	0	TN	FP
	1	FN	TP
		0	1
		Predicted	

$$\text{PRECISION : } \frac{TP}{TP+FP}$$





06

Interpretation



Conversion Rate

12%

Before ML

Conversion rate = (Jumlah Nasabah yang Berlangganan Secara **Aktual** / Jumlah Nasabah yang **Dihubungi**) * 100
Conversion rate = $(5289 / 45211) * 100 = \sim 12\%$

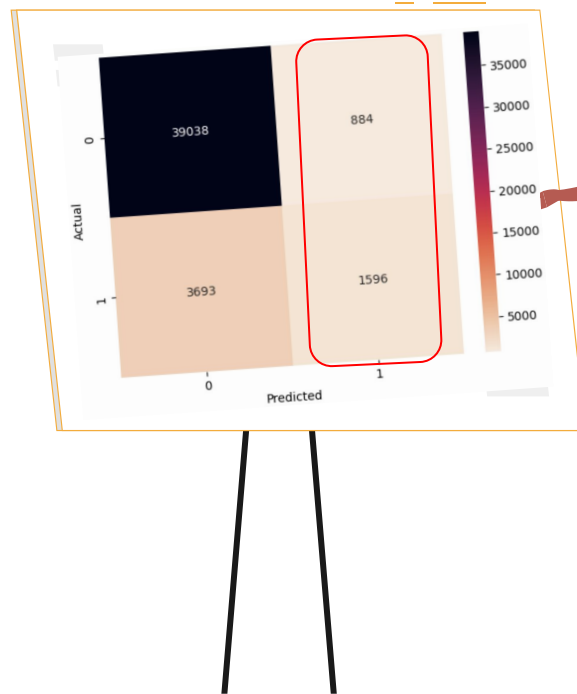
Hanya **12%** yang benar-benar berlangganan deposito berjangka.

64%

After ML

Conversion rate = (Jumlah Nasabah yang **aktual** Berlangganan / Jumlah Nasabah yang **diprediksi** berlangganan) * 100
Conversion rate = $(1596 / \{884 + 1596\}) * 100 = 0.643 \sim 64\%$

Berdasarkan prediksi model, sebanyak **64%** dari nasabah yang diprediksi benar-benar berlangganan deposito berjangka.



Cost Efficiency

- **Dataset awal**, dari 45211 nasabah, total campaign dilakukan 124.956 kali.
- **filtering manual** menggunakan (4 fitur) loan, housing, contact dan poutcome, didapatkan sebanyak 998 nasabah.
- **Menggunakan ML**, didapatkan 2.480 nasabah.

Jika asumsi campaign dilakukan **2 kali** per nasabah dan menggunakan “lead cost” sebesar **25** pounds, maka:

	Total Campaign	Price	Total Cost
Before ML	124956	£25	£ 3,123,900
Filter Manual	(998 x 2)	£25	£ 49,900
After ML	(2480 x 2)	£25	£ 124,000
Cost Reduction After ML - before ML			£ 2,999,900 (96%)
Cost Reduction After ML - Filter manual			£ 74,100 (60%) *filter manual lebih mereduksi cost



Profit Comparison

- **Dataset awal**, dari **45.211** nasabah, total nasabah deposit **5.289** nasabah.
- **filtering manual**, dari **998** nasabah yang terfilter, total deposit **704** nasabah.
- **Menggunakan ML**, dari rekomendasi **2.480** nasabah, total prediksi deposit **1.596** nasabah.

Jika asumsi nominal deposit sebesar **250 euro/216 pounds per nasabah**, maka:

	Total Deposit	Price	Total Revenue	Total Profit
Before ML	5.289	£216	£ 1,142,424	-1,981,476
Filter Manual	704	£216	£ 152,064	102,164
After ML	1.596	£216	£ 344,736	220,736
Profit After ML - before ML				2,202,212 (997%)
Profit After ML - Filter manual				118,572 (53%)



**ML mereduksi cost tanpa
menghilangkan potensi pendapatan**

Machine Learning Performance

Cost Reduction

96%



Compared by non ML

Profit

997%

Compared by non ML

53%

Compared by manual
filtering

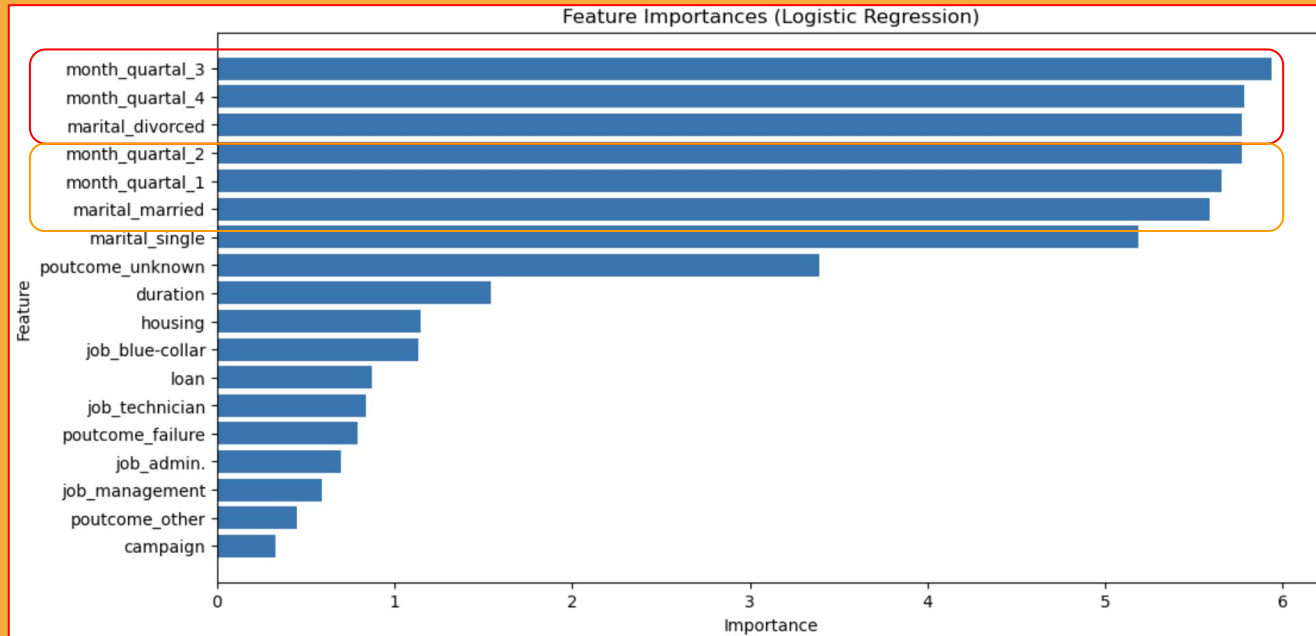


07

Recommendation



Feature Importance



Insight



Month quartal 3 & 4 (semester 2)

- Semester 2 waktu yang **krusial untuk keputusan deposit**, hal ini bisa jadi karena banyak nasabah menysikan banyak uang **“senggang”** di waktu ini misal untuk **liburan**.
- Pada waktu ini bank juga biasanya melakukan **penyesuaian** suku bunga dan perencanaan anggaran tahun depan.



Month quartal 1 & 2 (semester 1)

Bulan **maret** memiliki tingkat konversi langganan deposit **tertinggi** dan bulan **april** juga **cukup banyak**, kedua bulan ini bisa menjadi perhatian sepanjang semester 1.



Marital Divorced

- Kategori ini merujuk kepada nasabah yang **cukup dewasa** dan **telah melewati** masa pernikahan.
- Tipe ini memiliki **keleluasaan** dana karena banyak yang hidup **sendiri**, sehingga bisa lebih baik spend untuk **invest perencanaan masa depan**.



Marital Married

Tipe ini biasanya memiliki **perencanaan** keuangan untuk masa depan berdua, sehingga beberapa dari mereka kemungkinan besar akan **tertarik** terhadap deposito sbg media **investasi yang aman**



Recommendation



Month

- Bank perlu melakukan **campaign masing-masing 1 kali** tiap nasabah pada **kuartal 3 dan 4**
- Dalam campaign bisa menyesuaikan **promo liburan** (return lebih tinggi) dengan **pencairan** menjelang **akhir tahun** di kuartal 4.
- Perlu mencoba melakukan opsi campaign 1 kali pada bulan **maret** karena memiliki konversi tertinggi dalam langganan deposito.



Marital Divorced & Married

- Selain promo liburan, bank juga perlu memberikan **program khusus** untuk **tabungan pendidikan anak dan hari tua**, karena mungkin banyak nasabah dari kategori ini yang merupakan **janda memiliki anak** dan mungkin **janda tanpa tanggungan** yang biasanya butuh tabungan pensiun untuk dirinya sendiri.
- **Promo liburan** yang ditawarkan bisa **dikhususkan** juga pada pasangan yang **sudah menikah**



Git hub project

- <https://github.com/roquewindrunner/FinalProject-Rakamin>

Reference

- <https://www.kaggle.com/datasets/prakharrathi25/banking-dataset-marketing-targets>
- <https://immigrantinvest.com/blog/best-portugal-banks-en/>
- https://media.techtarget.com/rms/computerweekly/DowntimePDF/Buyers_guide/Telemarketing_Buyers_Guide_new.pdf

THANK YOU!

