



Final Project Bank UNO Marketing Campaign Targets

created by UNO

Ketua : Daniel Rowin

Anggota :

Febby Maghfirani Aziz

Amodya Subagio

Nur Rahman Shalahudin

Fakhry Abdurrohman

Daviro Yota Nagasan Wahyudi

Putri Vina Fajriyani

Asmiyeni Islamiati



Today's Highlight



1

Background

2

3

4

5



Problem Statement

Jumlah nasabah yang membuka deposito berjangka pada UNO Bank dari total 45.211 nasabah yang ada, **hanya sekitar 5.289 nasabah atau 11,7 % nasabah saja yang membuka deposito berjangka.**

Permintaan dari manajemen UNO Bank itu terdapat peningkatan, **minimal sebesar 15%** dari total nasabah UNO bank yang membuka deposito berjangka.

Diantaranya terdapat 4.369 nasabah atau 9.66% nasabah yang dihubungi melalui telepon cellular dan 390 nasabah atau 0.86% nasabah yang dihubungi melalui telepon rumah.



Role

Sebagai tim data scientist dalam suatu perusahaan bernama UNO Bank, kami diminta oleh manajemen UNO Bank untuk memprediksi apakah nasabah akan berlangganan deposito berjangka berdasarkan data yang tersedia guna meningkatkan performa dari perusahaan tersebut.

Goal

Meningkatnya jumlah nasabah yang membuka **deposito berjangka menjadi sebesar 15%**. Dimana ini berdasarkan permintaan manajemen UNO Bank.



Objective

1. Menganalisis profil nasabah berdasarkan variabel-variabel seperti usia, pekerjaan, status perkawinan, pendidikan, dan fitur lainnya sehingga nasabah tertarik untuk membuka deposito berjangka di UNO Bank.
2. Membangun model yang dapat memprediksi dengan akurat apakah seorang nasabah akan berlangganan deposito berjangka setelah kampanye pemasaran telepon dilakukan berdasarkan klasifikasi nasabahnya.



Business Metrics

Conversion Rate : Persentase nasabah UNO Bank yang membuka deposito berjangka





Today's Highlight



1

Background

2

EDA & Insight

3

4

5



Data Overview

- Terdiri dari 45.211 baris data dengan 17 features
- Tidak memiliki missing values
- Value unknown pada contact akan di-replace dengan modus dan pada poutcome di-replace dengan never

1

Categorical Data

job, marital, education, default,
housing, loan, contact, poutcome, y

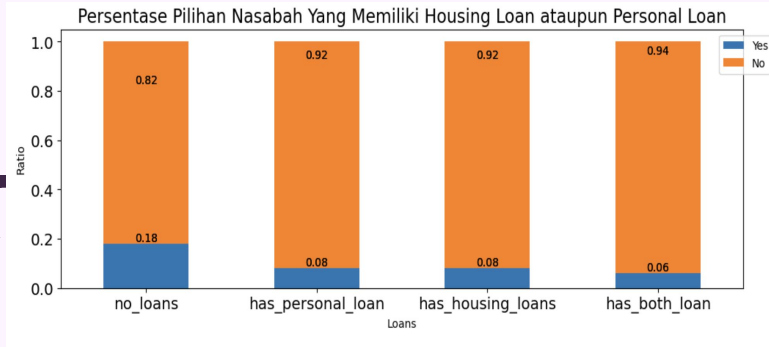
2

Numerical Data

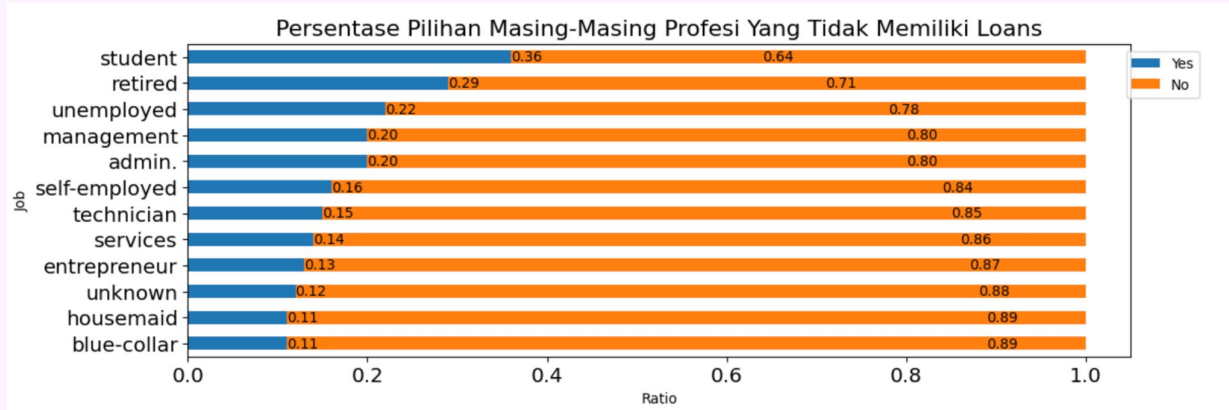
Age, balance, campaign, pdays,
previous

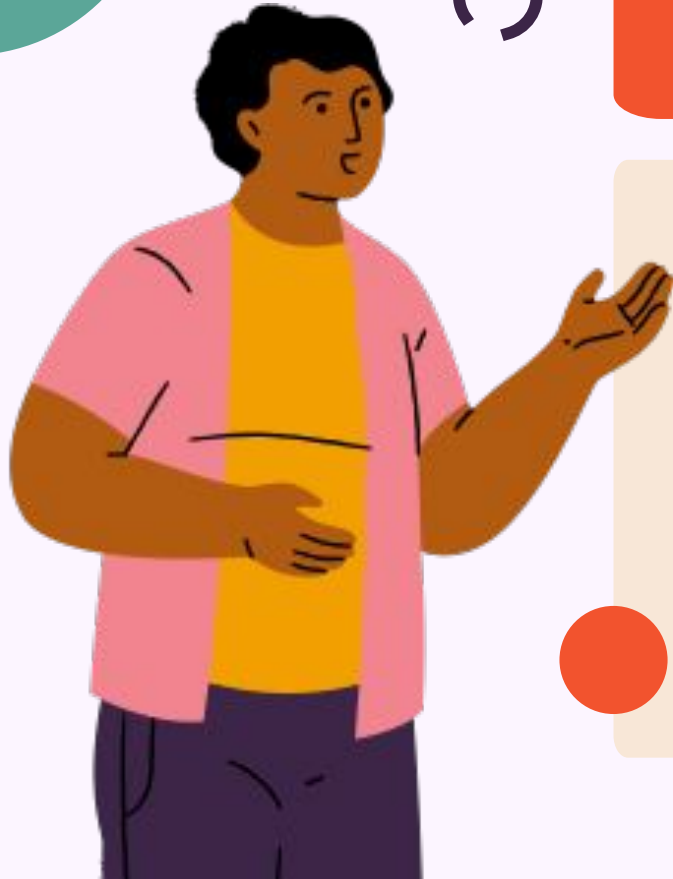


Business Insight



Nasabah yang membuka term deposit didominasi oleh nasabah yang **tidak memiliki housing loans maupun personal loans** dan jika dilihat berdasarkan pekerjaannya. Nasabah yang membuka term deposit kebanyakan berprofesi sebagai **student, retired, unemployed, admin, dan management**. Direkomendasikan kepada pihak marketing untuk mencari pendekatan yang paling baik bagi nasabah yang berprofesi sebagai blue-collar dan housemaid karena memiliki rasio terendah.

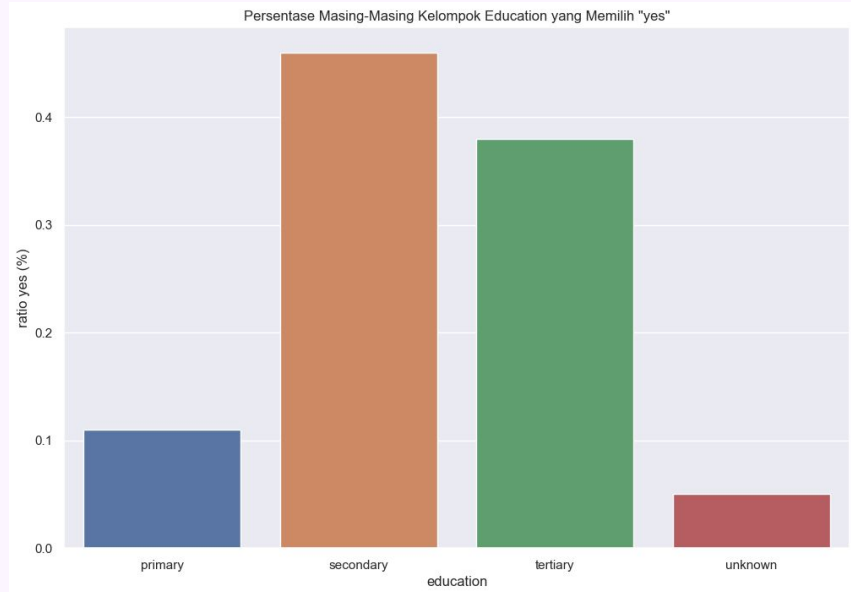
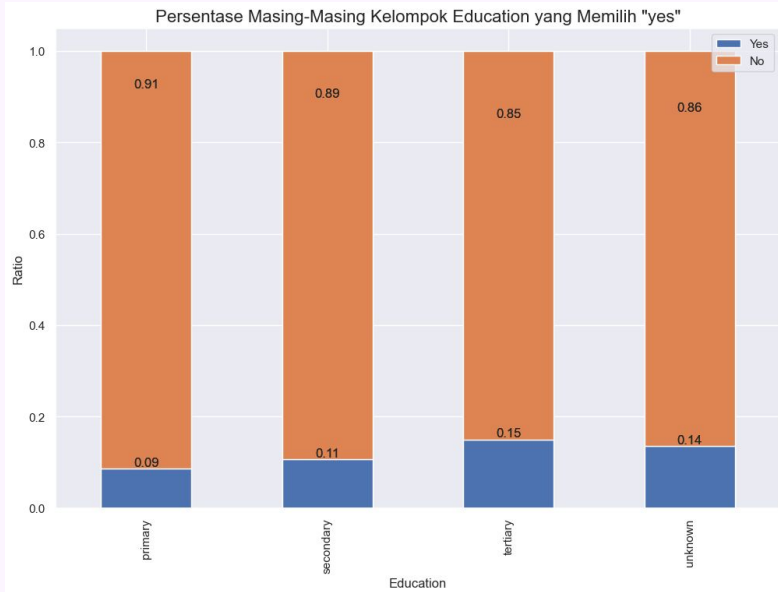




Business Insight

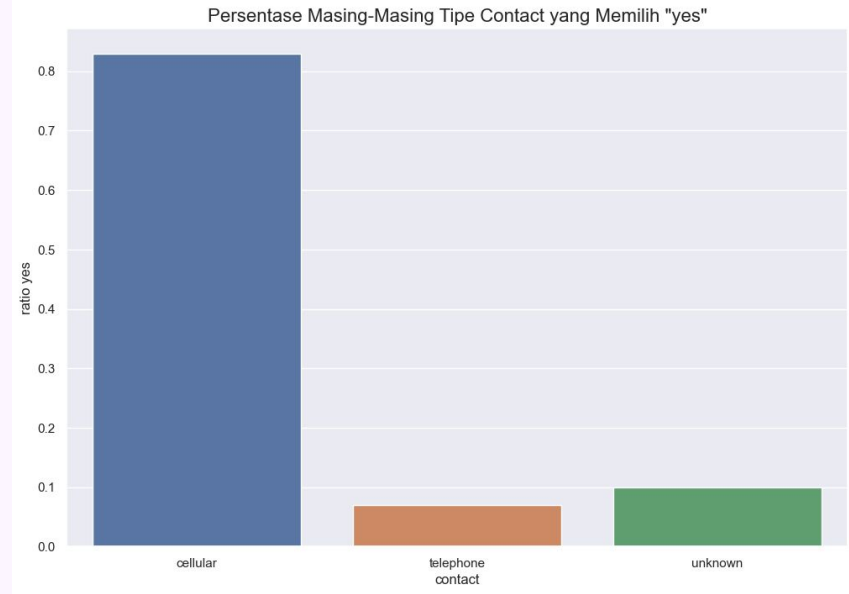
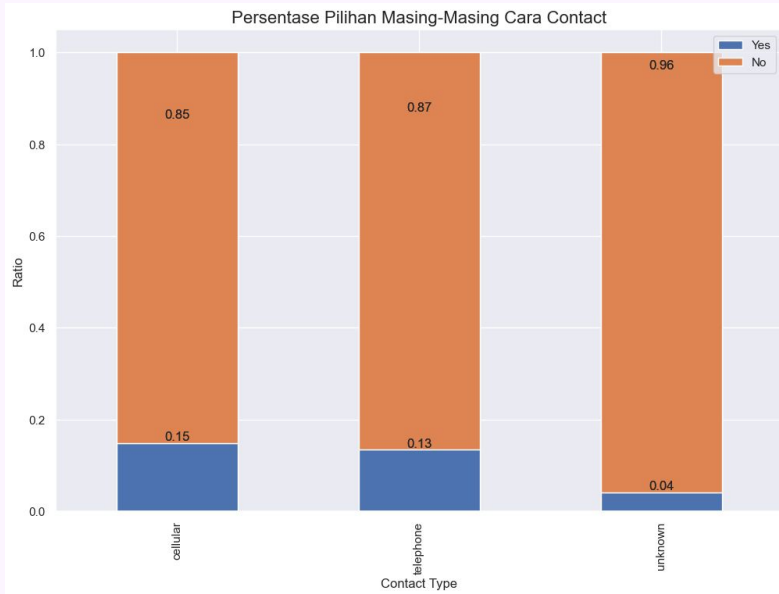
Dilakukan uji hypothesis dan ditemukan bahwa semakin banyak campaign yang diberikan maka akan semakin banyak nasabah yang akan menolak untuk membuka term deposit . Oleh karena itu direkomendasikan untuk optimalisasi pemberian campaign kepada nasabah, karena jumlah campaign dan keputusan nasabah berhubungan terbalik dan berpotensi dianggap spam.

Business Insight



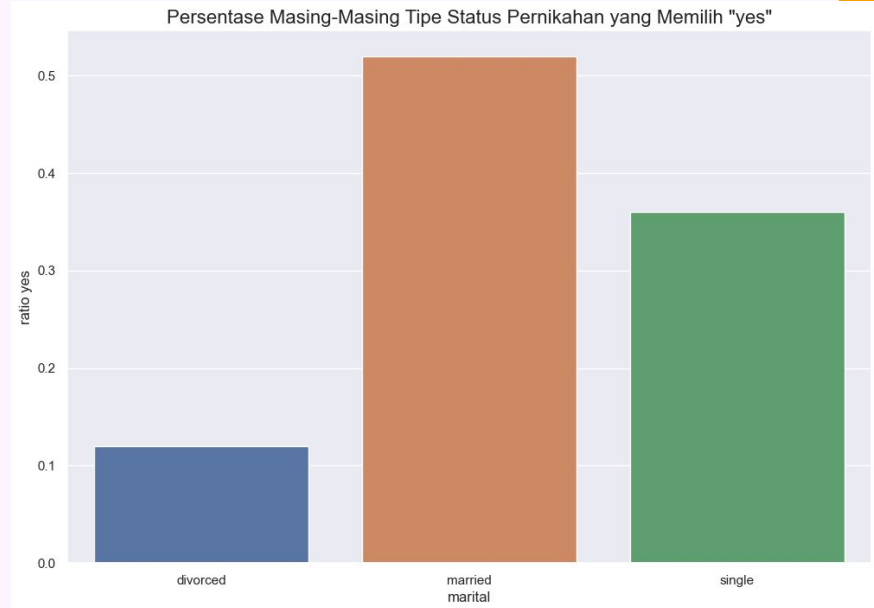
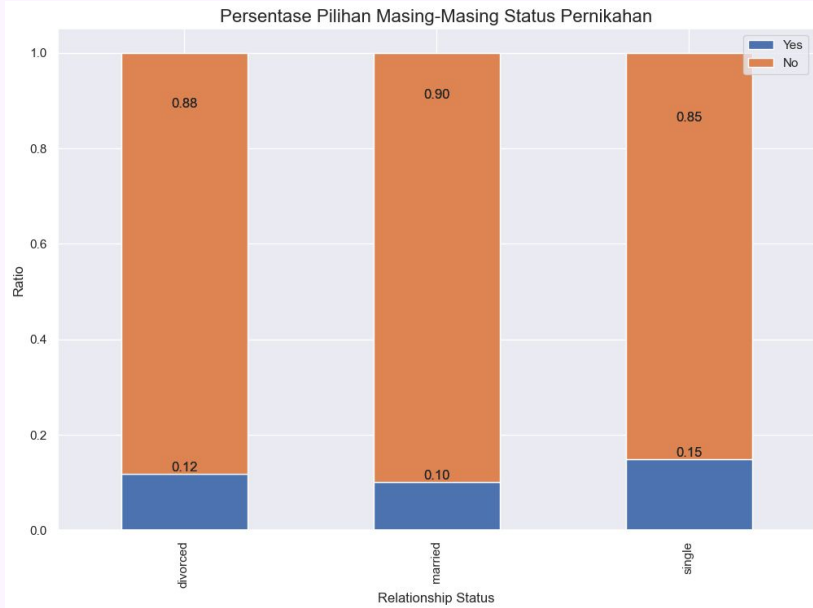
- Pada visualisasi diatas dapat dilihat bahwa kelompok education **tertiary** (lulusan S1 atau di atasnya) paling banyak memilih untuk mendaftar deposito berjangka sebanyak 15% dibanding total yang mendaftar di kelompoknya, yang diikuti oleh kelompok education **yang tidak diketahui**, lalu **secondary** (lulusan SMP dan SMA) dan terakhir **primary** (Lulusan SD ke bawah).
- Dengan membandingkan terhadap mereka yang mengambil yes saja, terlihat bahwa kelompok education **secondary** paling banyak memilih yes, diikuti dekat kelompok **tertiary**, lalu **primary** dan kelompok Education yang **tidak diketahui**.
- Dari insight tersebut, rekomendasi bisnis yang dapat diberikan adalah menargetkan campaign terhadap kelompok **tertiary untuk nasabah baru**. Bisa di daerah kampus, universitas, mengajak investasi sebagai kesempatan untuk mendapatkan passive income untuk uang yang mungkin telah mereka tabung.
- Rekomendasi ke dua, bisa ditargetkan untuk mereka yang masih di jenjang sekolah, utamanya **secondary untuk penawaran deposito berjangka**. Bisa melakukan campaign ke sekolah-sekolah, melibatkan orang tua, untuk membuka rekening dan deposito dini.

Business Insight



Bisa dilihat melalui visualisasi diatas bahwa tipe contact **cellular** (HP) merupakan tipe contact yang paling banyak menghasilkan customer untuk mendaftar deposito berjangka. Hal ini bisa terjadi karena di era sekarang ini orang-orang lebih banyak melakukan komunikasi melalui telepon celullar dibandingkan telepon biasa (telepon rumah), dan hampir setiap orang pasti mempunyai telepon cellular. Dari sini pihak bank **direkomendasi** mulai merubah strategi campaign untuk memprioritaskan melakukan contact melalui telepon celullar. Selain itu pihak bank sebelum melakukan panggilan secara langsung ke nomor telepon celullar, bisa juga bisa melakukan campaign melalui email atau melalui whatsapp sebagai tahapan awal dalam menawarkan deposito berjangka. Dengan demikian peluang customer yang mendaftar deposito berjangka akan meningkat.

Business Insight



- Pada visualisasi diatas dapat dilihat bahwa **yang masih single** paling banyak memilih untuk mendaftar deposito berjangka sebanyak 15% dibanding total yang mendaftar di antara nasabah yang masih single, lalu diikuti oleh yang sudah **bercerai**, dan terakhir mereka yang sudah **menikah** dalam 10%.
- Dengan membandingkan terhadap mereka yang mengambil yes saja, terlihat bahwa **yang sudah menikah** paling banyak memilih yes, diikuti mereka yang masih **single**, dan terakhir mereka yang sudah **cerai**.
- Rekomendasi untuk meningkatkan mereka yang mendaftar deposito berjangka pihak bank saat melakukan campaign kepada masing-masing individu dapat melakukan promosi dengan strategi pendekatan yang berbeda. Misalkan kepada kelompok **"single"** untuk nasabah baru bisa melakukan promosi seperti "dengan bunga deposito dalam setahun adalah x% maka kira-kira dalam y tahun dana menikah akan dapat terkumpul". Sedangkan untuk kelompok **"married"** untuk penawaran deposito berjangka bisa melakukan promosi seperti "dengan bunga deposito dalam setahun adalah x% maka dana pendidikan untuk anak/dana pensiun saat masa tua akan terjamin"



Today's Highlight



1

Background

2

EDA & Insight

3

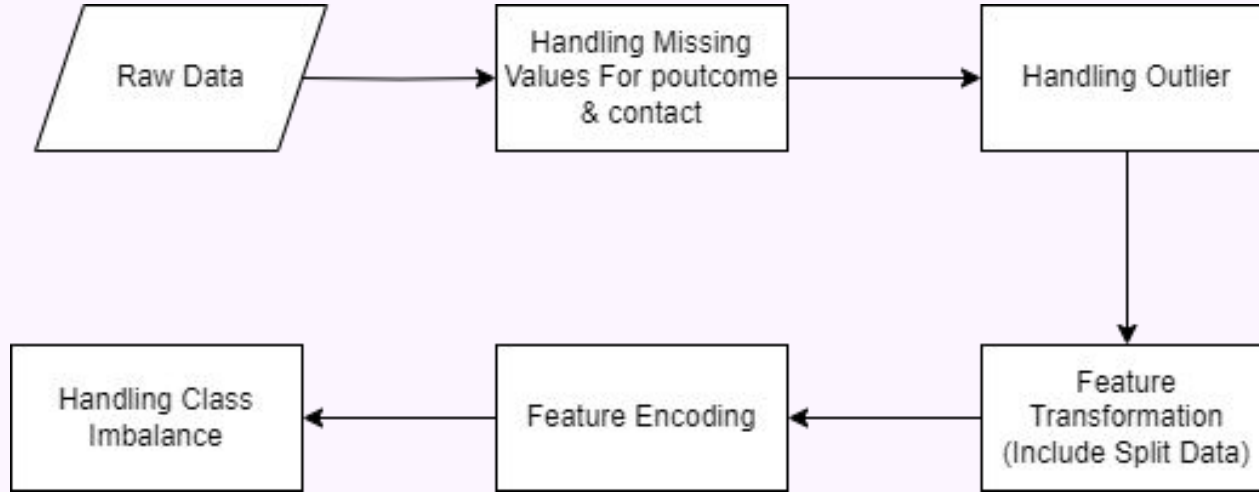
Data Pre-Processing

4

5



Diagram Pre-Processing



Feature Transformation

StandardScaler()

Oversampling

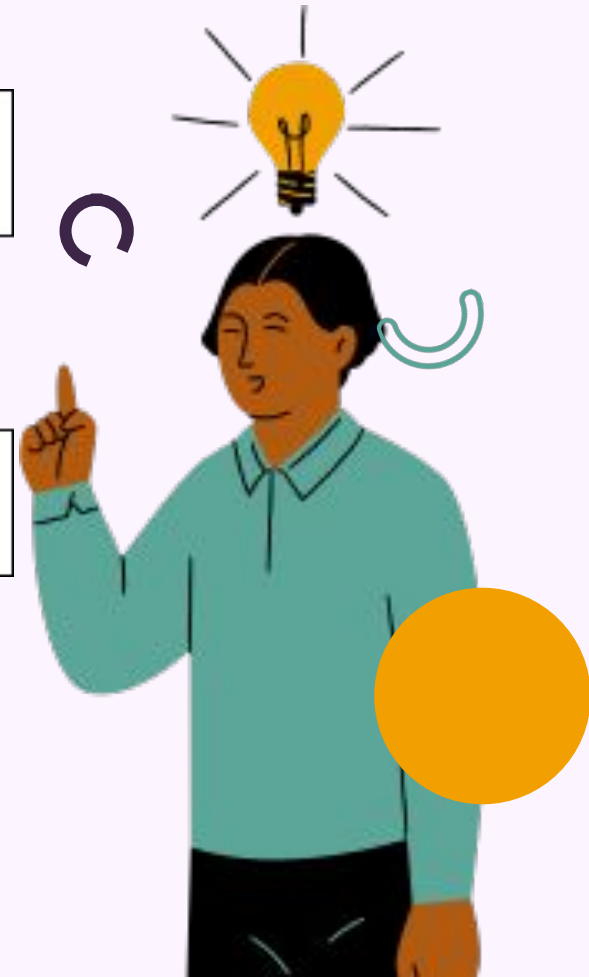
SMOTE

Split Train-Test Data

random state = 42
Split = 0.2

Number of Train Data:
35828

Number of Test Data: 8957



Feature Engineering

17 Feature Raw Data

Age
Job
Marital
Education
Default
Balance
Housing
Loan
Contact
Day
Month
Duration
Campaign
Pdays
Previous
Poutcome
y

35 Engineered Feature

age
default
balance
housing
loan
day
duration
campaign
pdays
previous
pendidikan_primary
pendidikan_secondary
pendidikan_tertiary
pendidikan_unknown
kerja_admin
kerja_blue-collar
kerja_entrepreneur
kerja_housemaid
kerja_management
kerja_retired
kerja_self-employed
kerja_services
kerja_student
kerja_technician
kerja_unemployed
kerja_unknown
status_divorced
status_married
status_single
contact_cellular
contact_telephone
poutcome_failure
poutcome_never
poutcome_other
poutcome_success



Today's Highlight



1

Background

2

EDA & Insight

3

Data Pre-Processing

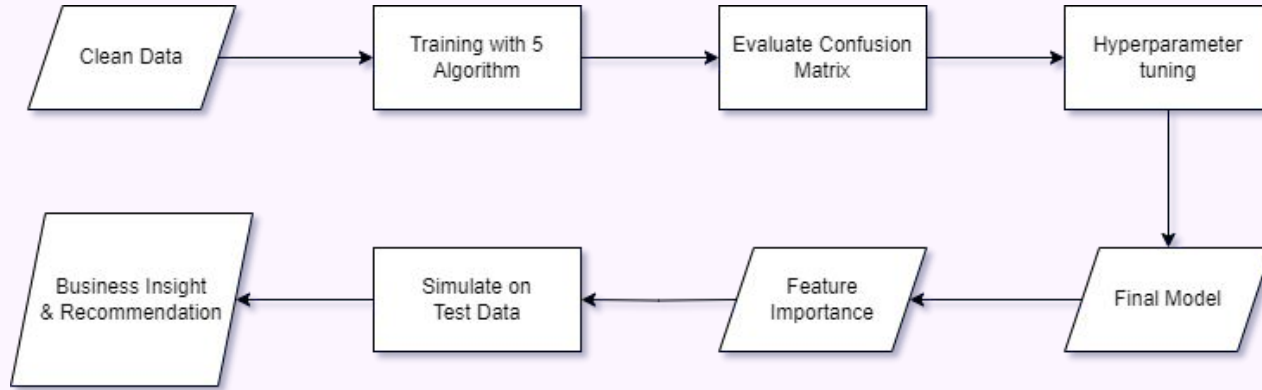
4

Modeling

5



Diagram Modelling



Algorithm

1. Logistic Regression
2. Decision Tree
3. Random Forest
4. AdaBoost
5. XGBoost



Evaluasi Confusion Matrix

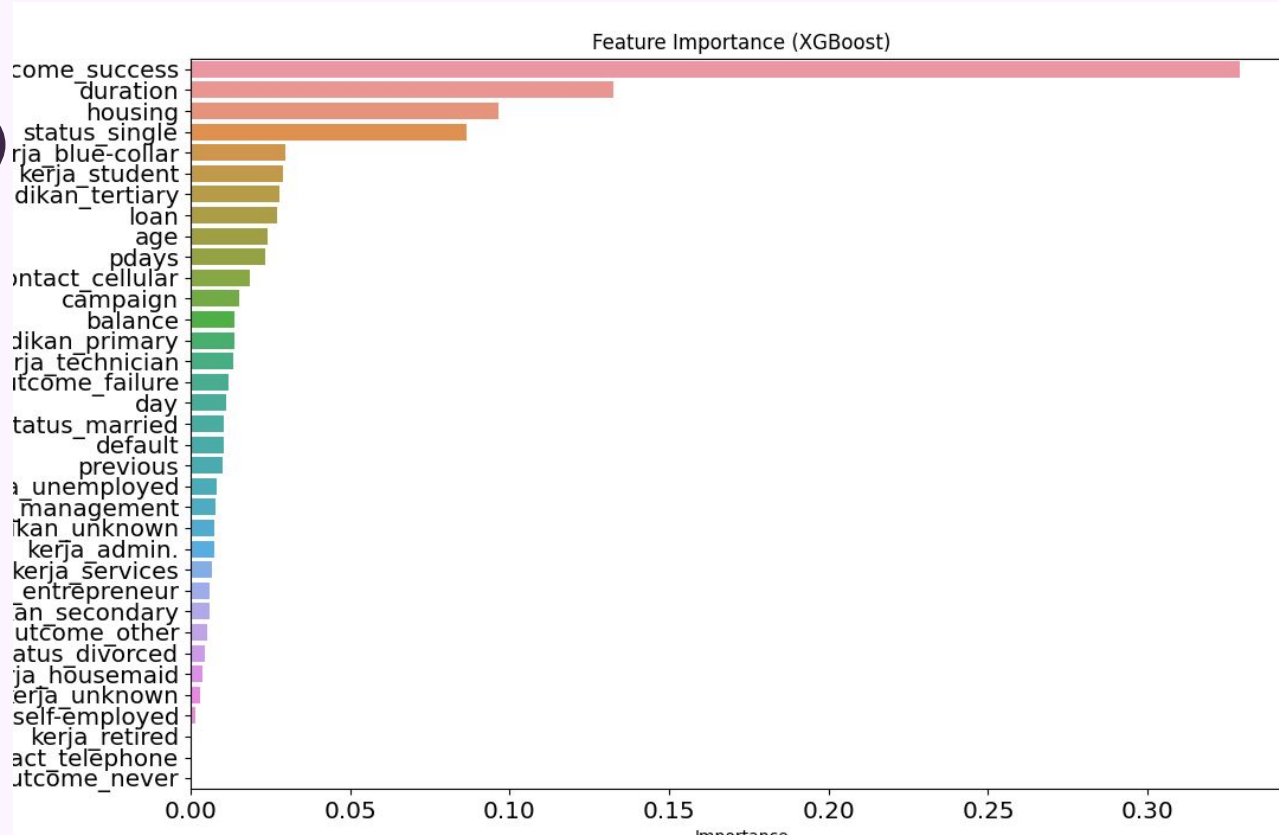
Train-Test Split						
Algorithm	AUC-Score Train	AUC-Score Test	Recall	Precision	F1	Accuracy
Logistic Regression	0.89	0.88	0.65	0.78	0.69	0.90
Decision Tree	1.00	0.67	0.67	0.66	0.66	0.86
Random Forest	1.00	0.88	0.65	0.76	0.68	0.90
AdaBoost	0.90	0.89	0.66	0.77	0.69	0.90
XGBoost	0.97	0.89	0.69	0.75	0.71	0.90

Evaluasi Confusion Matrix

Hyperparameter Tuning

Algorithm	AUC-Score Train	AUC-Score Test	Recall	Precision	F1	Accuracy
Logistic Regression	0.88	0.89	0.65	0.78	0.68	0.90
Decision Tree	0.77	0.78	0.65	0.77	0.69	0.90
Random Forest	0.89	0.92	0.58	0.82	0.61	0.90
AdaBoost	0.90	0.90	0.66	0.77	0.69	0.90
XGBoost	0.90	0.92	0.68	0.77	0.72	0.90

Feature Importance XGBoost



Feature-feature yang memiliki importance kurang dari 0,030 akan dihapus, guna mengurangi feature 'noise' yang ada

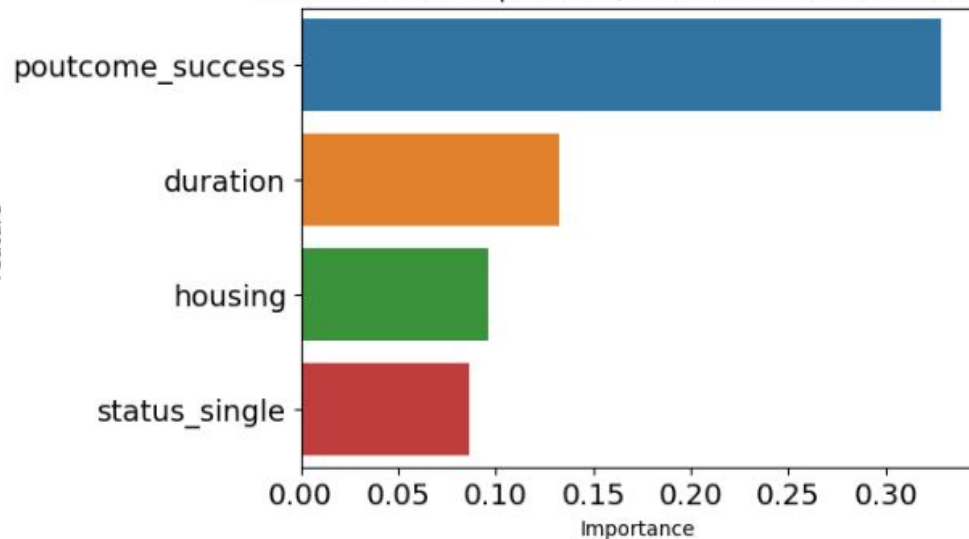
Setelah itu dilakukan modelling menggunakan XGBoost dengan data test menggunakan feature-feature yang sudah dikurangi.

Final Model XGBoost

Final Model XGBoost					
AUC-Score Train	AUC-Score Test	Recall	Precision	F1	Accuracy
0.88	0.86	0.66	0.78	0.69	0.90

Selected Feature Importance

Selected Feature Importance (XGBoost after Feature Selection)



Feature-feature yang memiliki pengaruh besar pada model yaitu,

Poutcome_success

Duration

Housing

status_single



Feature Importance

Review definisi feature:

poutcome_success: diambil dari feature encoding untuk feature "poutcome" yang berarti hasil/outcome dari campaign previous(p) / sebelumnya, yang mana hasilnya berupa success (bukan failure, unknown, other)

Duration: feature berkaitan tentang durasi campaign yang dilakukan pada waktu sebelumnya, dalam detik

Housing: feature yang merepresentasikan apakah pengguna memiliki pinjaman uang untuk rumah atau tidak (y/n)

Status_single : feature yang mempresentasikan pengguna yang statusnya belum menikah (y/n)

Business Insight

1. Poutcome_success:

Insight: Keberhasilan kampanye pemasaran sebelumnya memiliki dampak yang signifikan. Bisnis dapat fokus pada strategi yang **telah terbukti berhasil** dalam kampanye sebelumnya, memperkuat taktik yang menghasilkan tingkat keberhasilan yang tinggi.

2. Duration:

Insight: Durasi kontak memiliki dampak besar. Dengan mengoptimalkan durasi panggilan menjadi **minimal 4 menit**, bank dapat meningkatkan efektivitas kampanye pemasaran teleponnya, memastikan bahwa setiap interaksi memberikan nilai maksimal bagi pelanggan dan meningkatkan peluang konversi.

3. Housing:

Insight: Informasi tentang kepemilikan rumah mempengaruhi keputusan pelanggan. Ini bisa dikaitkan dengan kondisi keuangan pelanggan atau prioritas finansial mereka. Bisnis dapat menyusun strategi pemasaran khusus untuk kelompok **pelanggan yang tidak memiliki kredit rumah**, menawarkan produk atau layanan yang lebih sesuai dengan kebutuhan mereka.

4. status_single:

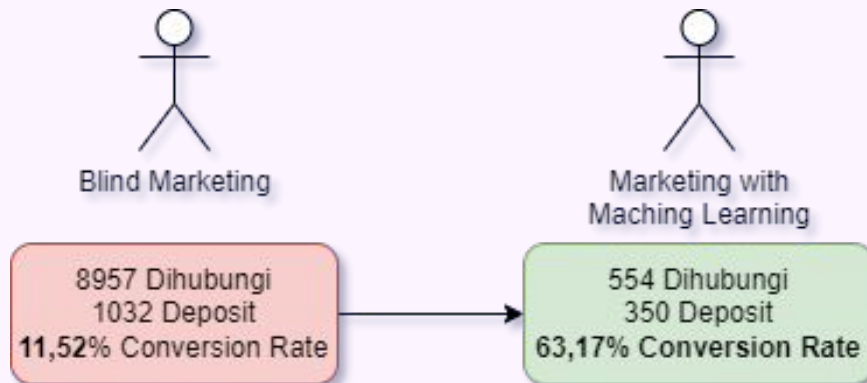
Insight: **Pelanggan dengan status sudah menikah atau divorce sangat berpotensi** untuk menerima campaign, pasangan yang menikah/divorce lebih fokus pada kestabilan finansial jangka panjang bagi keluarga mereka, termasuk tabungan untuk pendidikan anak, pensiun, atau pembelian properti, membuat produk seperti deposito berjangka menjadi pilihan yang menarik.

Simulasi

Dari Simulasi Pada Model, didapatkan hasil sebagai berikut.

Simulated on Data Test	Predicted Yes	Predicted No
Actual Yes	350	682
Actual No	204	7721

Kami merekomendasikan untuk **menghubungi yang terprediksi Yes saja**, dengan conversion rate sebesar 63,17%. **Goal tercapai dengan minimal meningkatkan sampai 15%.**





Today's Highlight



1

Background

2

EDA & Insight

3

Data Pre-Processing

4

Modeling

5

Rekomendasi Bisnis



Rekomendasi Bisnis

1. Optimalkan Kampanye Pemasaran:

Identifikasi faktor-faktor **keberhasilan dari kampanye pemasaran sebelumnya** yang menghasilkan outcome_success tinggi. Tingkatkan strategi pemasaran dengan mempertimbangkan pendekatan yang sama untuk meningkatkan kesuksesan kampanye mendatang.

2. Peningkatan Durasi Kontak:

Implementasikan **minimal durasi panggilan 4 menit** sebagai standar untuk semua interaksi dengan pelanggan dalam kampanye pemasaran telepon. Ini berdasarkan analisis data yang menunjukkan durasi interaksi yang lebih lama berkorelasi dengan peningkatan peluang konversi pelanggan.

3. Penargetan Kelompok Tanpa Kredit Rumah:

Fokus pada kelompok klien yang **tidak memiliki kredit rumah**. Pertimbangkan kondisi keuangan pelanggan dalam menyesuaikan penawaran. Pelanggan dengan pinjaman perumahan besar mungkin lebih berhati-hati dalam membuat komitmen finansial baru. Rancang kampanye khusus atau tawarkan insentif yang sesuai untuk menarik perhatian kelompok ini.

4. Pemasaran Berfokus Keluarga:

Gunakan materi pemasaran yang menyoroti bagaimana deposito berjangka dapat membantu dalam perencanaan masa depan keluarga, seperti pendidikan anak atau dana darurat keluarga. Tawarkan paket atau promosi khusus untuk **pasangan yang menikah**, mungkin dengan suku bunga yang lebih baik atau manfaat tambahan ketika membuka deposito bersama.



Thank You!

Descriptive Statistics

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 45211 entries, 0 to 45210
Data columns (total 17 columns):
#   Column      Non-Null Count  Dtype
---  -
0   age         45211 non-null  int64
1   job         45211 non-null  object
2   marital     45211 non-null  object
3   education   45211 non-null  object
4   default     45211 non-null  object
5   balance     45211 non-null  int64
6   housing     45211 non-null  object
7   loan        45211 non-null  object
8   contact     45211 non-null  object
9   day         45211 non-null  int64
10  month       45211 non-null  object
11  duration    45211 non-null  int64
12  campaign    45211 non-null  int64
13  pdays      45211 non-null  int64
14  previous    45211 non-null  int64
15  poutcome    45211 non-null  object
16  y           45211 non-null  object
dtypes: int64(7), object(10)
memory usage: 5.9+ MB
```

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

	age	balance	campaign	pdays	previous
count	45211.000000	45211.000000	45211.000000	45211.000000	45211.000000
mean	40.936210	1362.272058	2.763841	40.197828	0.580323
std	10.618762	3044.765829	3.098021	100.128746	2.303441
min	18.000000	-8019.000000	1.000000	-1.000000	0.000000
25%	33.000000	72.000000	1.000000	-1.000000	0.000000
50%	39.000000	448.000000	2.000000	-1.000000	0.000000
75%	48.000000	1428.000000	3.000000	-1.000000	0.000000
max	95.000000	102127.000000	63.000000	871.000000	275.000000

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

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

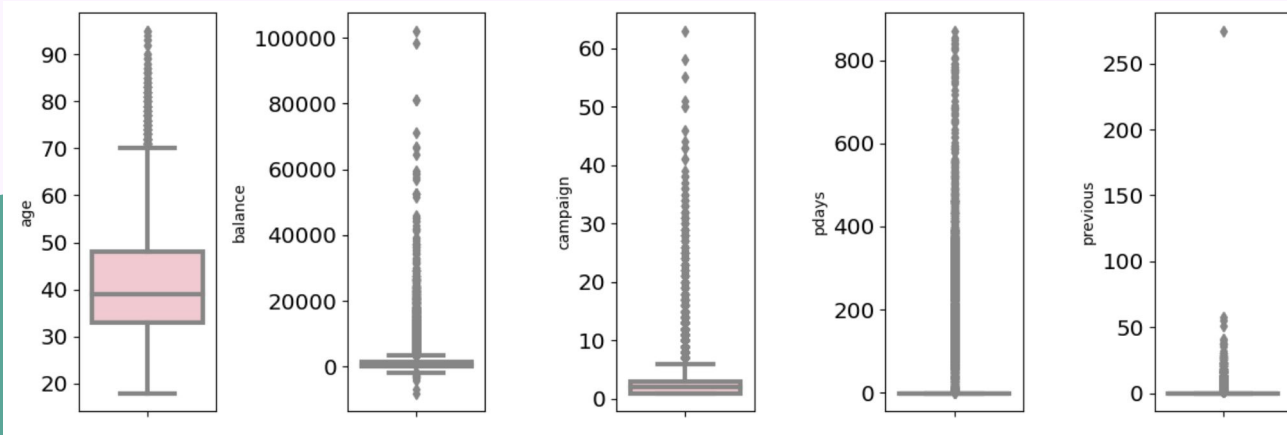
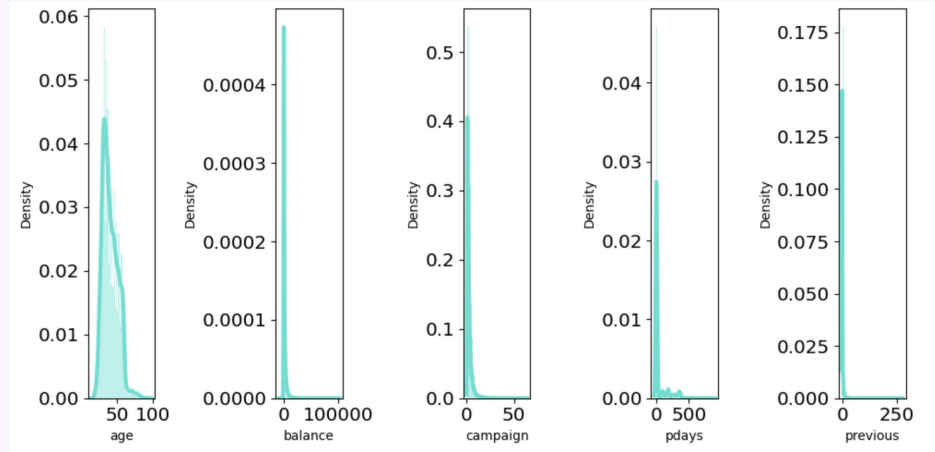
```
df.isna().sum()
```

```
age      0
job      0
marital  0
education 0
default  0
balance  0
housing  0
loan     0
contact  0
day      0
month    0
duration 0
campaign 0
pdays   0
previous 0
poutcome 0
y        0
dtype: int64
```

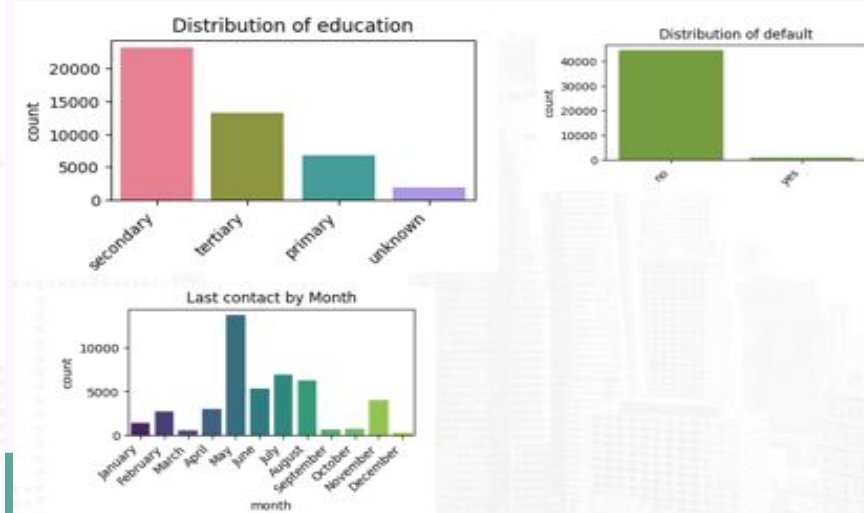
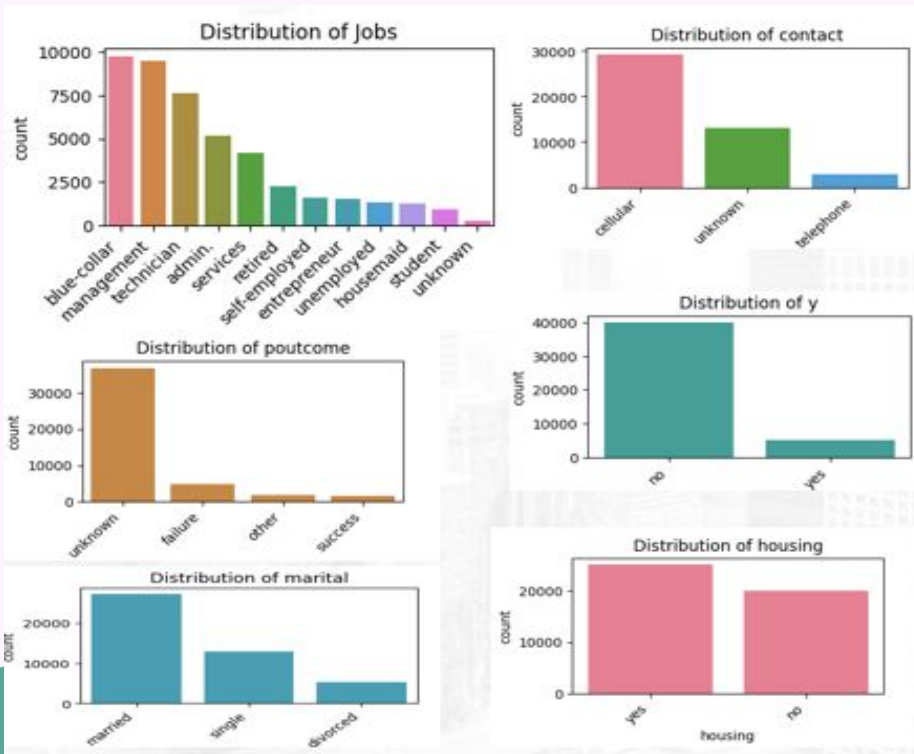
```
df['duration'].describe()
```

```
count    45211.000000
mean      258.163080
std       257.527812
min        0.000000
25%       103.000000
50%       180.000000
75%       319.000000
max       4918.000000
Name: duration, dtype: float64
```

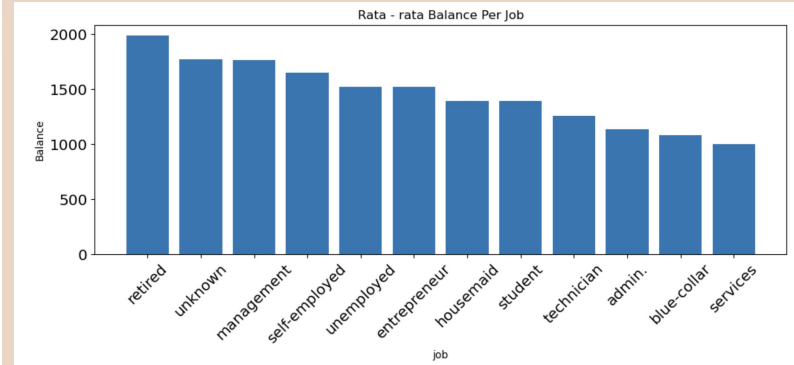
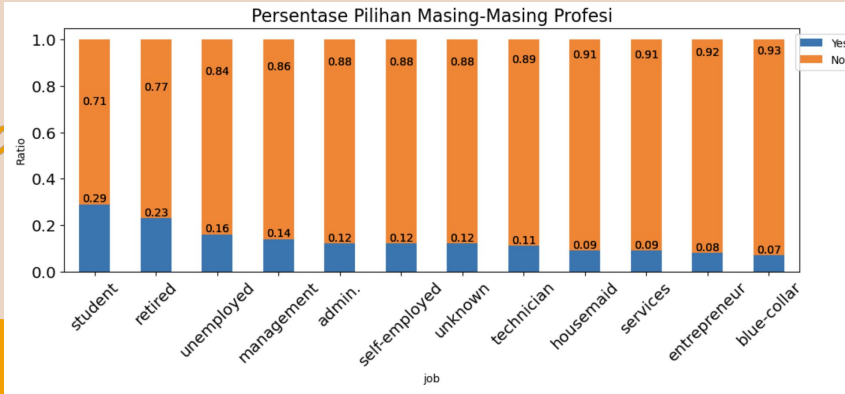
Distribution of Numerical Data



Distribution of Categorical Data



Business Insight



- Nasabah yang membuka term deposit didominasi oleh nasabah yang berprofesi sebagai **student, retired, unemployed, dan management**.
- Nasabah yang memiliki saldo rata - rata tahunan tertinggi dimiliki oleh nasabah dengan profesi **retired, unknown, management, self-employed, unemployed, dan entrepreneur** dengan saldo rata - rata tahunan di atas 1500.

Business Insight

y campaign		
0	no	2.846350
1	yes	2.141047

```
import scipy.stats as st
# Hypothesis Testing using mann whitney
stat, p_value = st.mannwhitneyu(yes['campaign'], no['campaign'], alternative='two-sided')
p_value
```

1.9484904873905108e-71

```
alpha = 0.05
print('P-Value :', p_value)

if p_value >= alpha:
    print('Tidak cukup bukti jumlah campaign mampu membedakan user untuk membuka akun atau tidak')
else:
    print('cukup bukti jumlah campaign mampu membedakan user untuk membuka akun atau tidak')
```

P-Value : 1.9484904873905108e-71

cukup bukti jumlah campaign mampu membedakan user untuk membuka akun atau tidak

Hasil uji hipotesis menunjukkan bahwa $p\text{-value} < \alpha$, maka kita akan mengambil keputusan bahwa jumlah campaign berpengaruh terhadap nasabah untuk membuka akun atau tidak secara signifikan.

Namun berdasarkan berdasarkan rata-rata ternyata semakin banyak campaign yang diberikan ternyata user akan semakin menolak membuka akun.

Jadi kesimpulannya, jumlah campaign berhubungan terbalik dengan user membuka akun.

Handle missing values

```
df.isna().sum()
```

```
age      0
job      0
marital  0
education 0
default  0
balance  0
housing  0
loan     0
contact  0
day      0
month    0
duration 0
campaign 0
pdays   0
previous 0
poutcome 0
y        0
dtype: int64
```

```
[ ] df['poutcome'].replace({'unknown': 'never'}, inplace=True)
```

Tidak terdapat missing values pada dataset, namun untuk fitur poutcome yang memiliki value unknown dilakukan replacement menjadi never yang menunjukkan bahwa customer tidak pernah dihubungi karena memiliki pdays = -1.

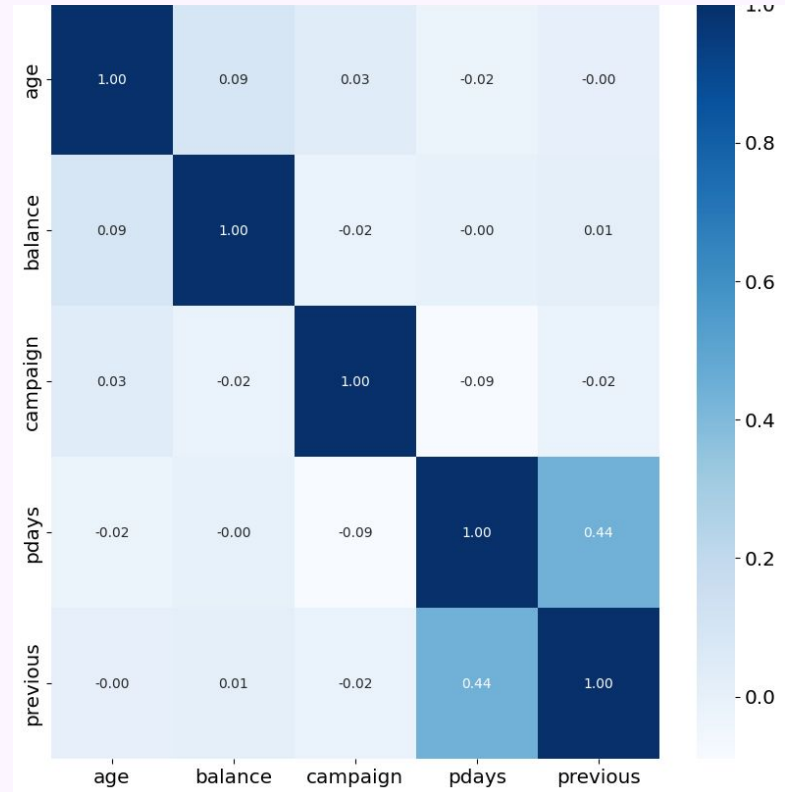
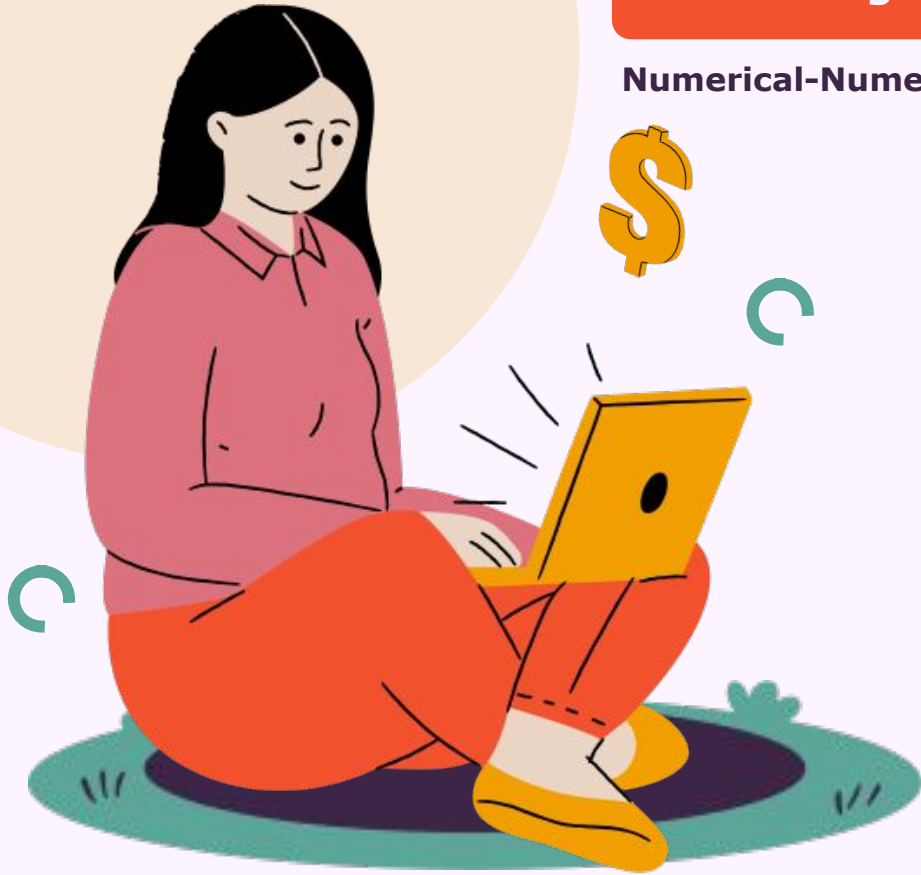
Pada **contact**, terdapat unknown yang diartikan tidak diketahui cara kontak melalui apa, diikuti dengan cara seluler dan telepon (rumah). Di sini, value unknown akan dijadikan seluler disebabkan terdapat feature duration, yang berarti durasi campaign yang dilakukan, dan lebih mudah untuk mengasumsi suatu nomor telepon itu seluler dibandingkan telepon rumah (sebab terdapat distinct dari 3 atau 4 angka pertama [seperti 031, 021, dst]).



Diagram Pre-Processing



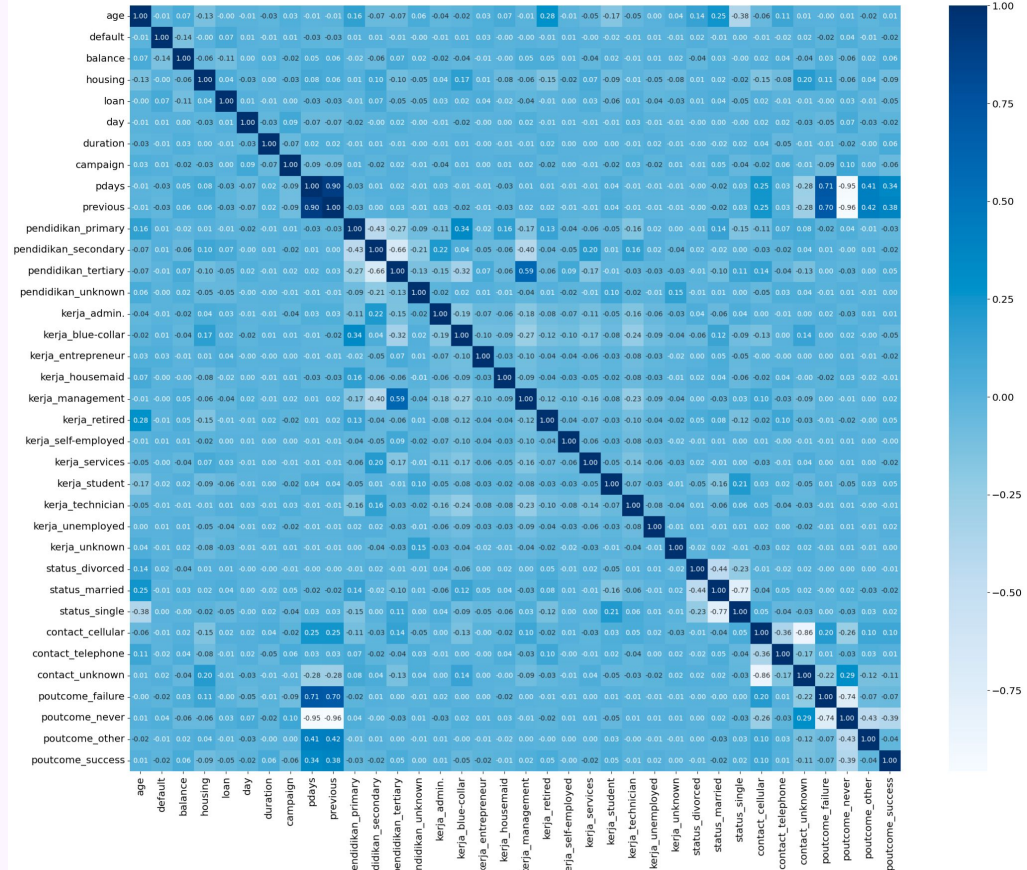
Numerical-Numerical Correlation Heatmap using Pearson Method



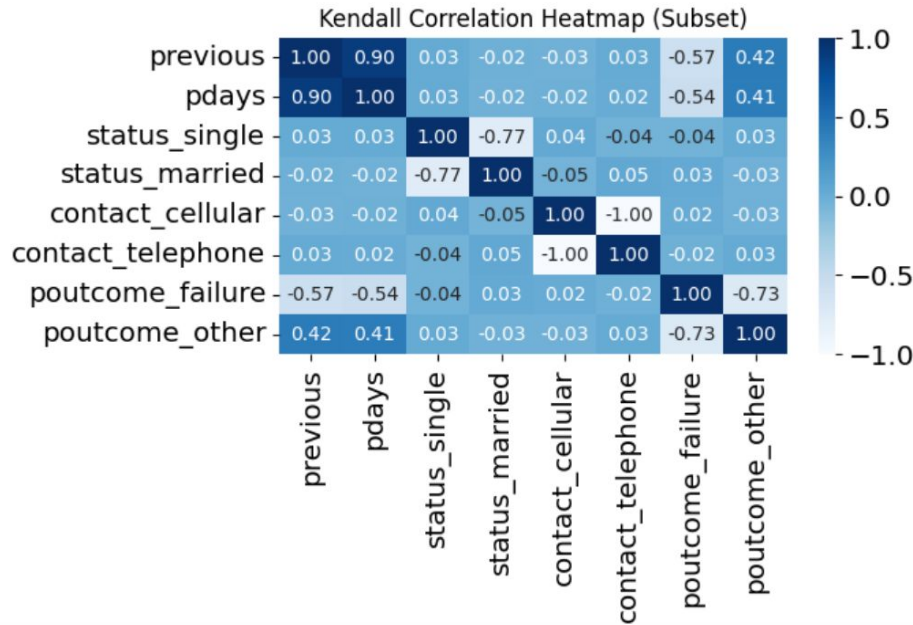
Feature Engineering



Categorical-Numerical and Categorical-Categorical Correlation Heatmap using Kendall Method



Feature Engineering



Delete Redundant Data

Feature **'previous'** dengan **'pdays'**,
'status_single' dengan **'status_married'**,
'contact_cellular' dengan
'contact_unknown', **'poutcome_failure'**
dengan **'poutcome_never'**, memiliki korelasi
di atas 0.7 yang menyebabkan mereka
redundant untuk dijadikan feature bersama,
sehingga akan digunakan salah satu saja.

Feature Extraction

Dengan adanya 32 Feature, dirasa telah cukup feature yang dibutuhkan, dengan feature ini memiliki relevansi tersendiri terhadap 'target' yang ingin dicapai. Sehingga kami **tidak akan mengeluarkan feature baru.**

Modelling

Algorithm Supervised Machine Learning

Untuk Modelling, kami menggunakan beberapa Algorithm yang telah dipelajari, yaitu,

1. Logistic Regression

- Interpretability: Logistic Regression adalah model yang relatif mudah diinterpretasikan, yang membuatnya berguna untuk memahami hubungan antara fitur dan probabilitas kejadian target.
- Efisiensi: Cocok untuk skenario ketika asumsi bahwa hubungan antara fitur dan target bersifat linier.
- Ketahanan terhadap Overfitting: Logistic Regression cenderung kurang rentan terhadap overfitting pada data yang relatif sederhana.

2. Decision Tree

- Kemampuan Memprediksi Nonlinearitas: Decision Tree dapat menangani hubungan non-linier antara fitur dan target.
- Keputusan yang Mudah Diinterpretasikan: Decision Tree menghasilkan struktur pohon yang mudah diinterpretasikan, memungkinkan pemahaman yang baik tentang faktor-faktor yang mempengaruhi keputusan.

Modelling

Algorithm Supervised Machine Learning

3. Random Forest

- Kombinasi Kelebihan Decision Tree: Random Forest menggabungkan keunggulan Decision Tree dan mengatasi kelemahan-kelemahan seperti overfitting dengan menggunakan ensemble learning.
- Mampu Menangani Banyak Fitur: Cocok untuk data dengan banyak fitur karena dapat menangani sejumlah besar variabel prediktif.

4. AdaBoost

- Peningkatan Kinerja: AdaBoost dapat meningkatkan kinerja model dengan fokus pada kasus yang sulit diidentifikasi oleh model sebelumnya, sehingga efektif untuk meningkatkan recall.
- Kemampuan Menyesuaikan: Mampu menyesuaikan diri dengan model lemah yang lebih baik untuk menangani ketidakseimbangan kelas.

5. XGBoost

- Performa yang Tinggi: XGBoost merupakan algoritma yang sangat efisien dan sering kali memberikan kinerja yang sangat baik, terutama dalam hal akurasi dan waktu komputasi.
- Ketahanan terhadap Overfitting: Memiliki teknik penanganan overfitting dan kemampuan untuk menangani kompleksitas model.

Evaluasi Confusion Matrix

Dari tabel hasil evaluasi model menggunakan beberapa metrik kinerja, terutama AUC-Score Train, AUC-Score Test, Recall, Precision, F1, dan Accuracy, tampaknya XGBoost memberikan kinerja yang baik secara keseluruhan. Berikut adalah beberapa alasan kami memilih XGBoost berdasarkan hasil tersebut:

AUC-Score (Area Under the Curve):

XGBoost memiliki nilai AUC-Score yang tinggi baik pada data latih maupun data uji. AUC adalah metrik yang baik untuk mengukur kemampuan model memisahkan kelas positif dan negatif. Semakin tinggi AUC, semakin baik model dalam membedakan antara kelas.

Recall:

Recall pada XGBoost juga relatif tinggi. Recall mengukur sejauh mana model dapat mengidentifikasi keseluruhan kasus positif. Pada konteks kampanye pemasaran, recall yang tinggi berarti model mampu mendeteksi sebanyak mungkin pelanggan yang benar-benar berlangganan, yang merupakan hal yang penting.

Precision:

Precision XGBoost terbilang baik, menunjukkan bahwa dari prediksi positif yang dilakukan model, sebagian besar adalah benar. Ini penting untuk meminimalkan jumlah kontak yang tidak perlu kepada pelanggan yang sebenarnya tidak berlangganan.

Evaluasi Confusion Matrix

F1-Score:

F1-Score yang baik pada XGBoost menunjukkan keseimbangan yang baik antara recall dan precision. F1-Score menggabungkan kedua metrik tersebut menjadi satu skor, memberikan gambaran komprehensif tentang kinerja model.

Accuracy:

Tingkat akurasi XGBoost juga cukup tinggi, menunjukkan seberapa baik model dalam membuat prediksi yang benar secara keseluruhan.

Konsistensi Kinerja (Train dan Test):

Model XGBoost menunjukkan konsistensi kinerja baik pada data latih maupun data uji, yang menunjukkan kemampuan umumnya dalam menggeneralisasi dari data pelatihan ke data baru.

Dengan kombinasi nilai yang tinggi pada berbagai metrik, XGBoost dapat dianggap sebagai pilihan yang baik untuk model pada kasus ini

Evaluasi Confusion Matrix

Meskipun hasil evaluasi model setelah penyetelan hyperparameter menunjukkan kinerja yang tinggi untuk beberapa model, ada beberapa perbedaan yang dapat memengaruhi pemilihan model. Berikut adalah beberapa alasan mengapa kami memilih XGBoost berdasarkan hasil tersebut:

AUC-Score (Area Under the Curve):

XGBoost memiliki nilai AUC-Score yang konsisten baik pada data latih maupun data uji, menunjukkan kemampuan model untuk membedakan antara kelas positif dan negatif. AUC-Score yang tinggi adalah indikator keunggulan model dalam hal ini.

Recall:

Recall pada XGBoost lebih tinggi dibandingkan dengan model lainnya. Recall yang tinggi merupakan faktor penting dalam konteks kampanye pemasaran, karena kita ingin mendeteksi sebanyak mungkin pelanggan yang benar-benar berlangganan.

F1-Score:

XGBoost memiliki F1-Score yang lebih tinggi dibandingkan dengan model lainnya. F1-Score menyatukan recall dan precision, dan nilai yang tinggi menunjukkan keseimbangan yang baik antara kedua metrik tersebut.

Evaluasi Confusion Matrix

Konsistensi Kinerja (Train dan Test):

Model XGBoost menunjukkan konsistensi kinerja baik pada data latih maupun data uji, menandakan kemampuan umumnya untuk menggeneralisasi dari data pelatihan ke data baru. Konsistensi ini dapat membantu mengurangi risiko overfitting.

Presisi:

Meskipun presisi XGBoost sama dengan Logistic Regression dan AdaBoost, kinerja secara keseluruhan yang lebih baik pada metrik lainnya menjadikan XGBoost sebagai pilihan yang menarik.

AUC-Score Test yang Tinggi:

XGBoost memiliki nilai AUC-Score pada data uji yang cukup tinggi, menunjukkan bahwa model mampu memberikan prediksi yang baik bahkan pada data yang belum pernah dilihat sebelumnya.

Dengan mempertimbangkan aspek-aspek tersebut, XGBoost tetap menjadi pilihan yang baik dan konsisten untuk kasus ini. Meskipun performa model Random Forest cukup baik dalam beberapa metrik, XGBoost masih unggul dalam hal F1-Score dan Recall, yang relevan dalam skenario kampanye deposito untuk meningkatkan identifikasi pelanggan berpotensi.