

# **Banking Telesales Marketing Optimization**

**Yohanes Gabriel Valentino Manurung**

# Problem Statement

Deposito berjangka merupakan sumber pendapatan utama bagi bank. Di antara berbagai saluran yang digunakan untuk menjangkau nasabah, kampanye penjualan lewat telepon tetap menjadi salah satu yang paling efektif. Akan tetapi, biayanya tinggi. Dari referensi daring, pusat panggilan internal (4 tenaga penjualan) dapat menelan biaya sekitar 460.000 USD per tahun, sementara pusat panggilan yang dialihdayakan menelan biaya 25-65 USD per jam per tenaga penjualan untuk AS dan 8-18 USD per jam per tenaga penjualan untuk Amerika Selatan dan Asia Tenggara.

Dengan demikian, sangat penting untuk mengidentifikasi nasabah yang paling mungkin berkonversi sebelumnya sehingga mereka dapat ditargetkan secara khusus melalui panggilan. Klasifikasi ini akan membantu bank untuk mengoptimalkan kampanye penjualan lewat telepon: meningkatkan tingkat penerimaan dan/atau mengurangi biaya. Salah satu lembaga perbankan Portugis mempekerjakan Chang Corp untuk membantu proyek pengoptimalan penjualan lewat telepon mereka.

- 1.<https://www.kaggle.com/datasets/prakharrathi25/banking-dataset-marketing-targets>
- 2.<https://www.cloudtask.com/blog/how-much-does-it-cost-to-outsource-a-call-center>

# Goal

Buatlah model yang membantu mengklasifikasikan pelanggan berdasarkan kemungkinan mereka untuk menerima tawaran penjualan lewat telefon, sehingga kami dapat membantu klien untuk:

1. Meningkatkan tingkat penerimaan penjualan lewat telefon, dan/atau
2. Mengurangi biaya penjualan lewat telefon

# Objectives

- 1.Cari tahu karakteristik dan faktor lain dari nasabah yang akan tertarik dan mendaftar pada produk deposito berjangka yang ditawarkan
- 2.Bangun model pembelajaran mesin yang dapat memprediksi calon nasabah yang akan ditawarkan kampanye penjualan melalui telepon
- 3.Berikan rekomendasi kepada tim pemasaran untuk meningkatkan efektivitas kampanye bagi calon nasabah deposito berjangka

# Business Metrics

- 1.Take up rate ( Tingkat Penerimaan )
- 2.Telesales cost ( Biaya Penjualan Tele )

# Stage 1

# Feature List

- 1 - age (numeric)
  - 2 - job : type of job (categorical: "admin.", "unknown", "unemployed", "management", "housemaid", "entrepreneur", "student", "blue-collar", "self-employed", "retired", "technician", "services")
  - 3 - marital : marital status (categorical: "married", "divorced", "single"; note: "divorced" means divorced or widowed)
  - 4 - education (categorical: "unknown", "secondary", "primary", "tertiary")
  - 5 - default: has credit in default? (binary: "yes", "no")
  - 6 - balance: average yearly balance, in euros (numeric)
  - 7 - housing: has housing loan? (binary: "yes", "no")
  - 8 - loan: has personal loan? (binary: "yes", "no")
  - # related with the last contact of the current campaign:
  - 9 - contact: contact communication type (categorical: "unknown", "telephone", "cellular")
  - 10 - day: last contact day of the month (numeric)
  - 11 - month: last contact month of year (categorical: "jan", "feb", "mar", ..., "nov", "dec")
  - 12 - duration: last contact duration, in seconds (numeric)
  - # other attributes:
  - 13 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
  - 14 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)
  - 15 - previous: number of contacts performed before this campaign and for this client (numeric)
  - 16 - poutcome: outcome of the previous marketing campaign (categorical: "unknown", "other", "failure", "success")
  - Output variable (desired target):**
  - 17 - y - has the client subscribed a term deposit? (binary: "yes", "no")
- Missing Attribute Values: None

# Insights Descriptive Statistics (1 / 3)

## A. Apakah ada kolom dengan tipe data kurang sesuai, atau nama kolom dan isinya kurang sesuai?

Dari semua informasi di atas, terlihat bahwa data tersebut memiliki 17 feature/ kolom. Tipe data untuk masing-masing kolom sudah sesuai. Antara nama kolom dengan isinya juga sudah sesuai.

## B. Apakah ada kolom yang memiliki nilai kosong? Jika ada, apa saja?

Dari 17 kolom tidak ada satupun yang memiliki nilai kosong, sehingga tidak perlu ada preprocessing untuk missing values.

# Insights Descriptive Statistics (2 / 3)

## C. Apakah ada kolom yang memiliki nilai summary agak aneh? (min/mean/median/max/unique/top/freq)

### Kolom Numerik

- Age: Untuk kolom age tidak terlihat potensi skewed karena mean (40.9) dan mediannya (39.0) tidak berbeda signifikan.
- Balance: Untuk kolom balance terlihat nilai mean (1362.3) sangat jauh di atas mediannya (448.0). Nilai minimum balance memiliki nilai aneh, yaitu nilai yang minus sebesar -8019. Perlu dilakukan investigasi lanjutan terkait balance negatif.
- Day: Untuk kolom day terlihat mean (15.8) dan median (16.0) hampir sama. Kemudian nilai minimum, maksimum, kuartil 1 dan 3 terlihat normal.
- Duration: Untuk kolom duration terlihat bahwa nilai mean (258.2) lebih besar dari nilai median (180.0), ada potensi positive skew.
- Campaign: Untuk kolom campaign nampak adanya kecenderungan positive skew, di mana nilai mean (2.7) lebih besar dibanding nilai median (2.0).
- Pdays: Untuk kolom pdays terlihat data berkumpul di angka -1 (36954 dari 45211 records), yang artinya hampir semua client tidak pernah di hubungi sebelumnya.
- Previous: Untuk kolom previous terlihat kebanyakan client (36954 dari 45211 records) memiliki nilai 0, yang artinya belum pernah menerima campaign sebelumnya.

# Insights Descriptive Statistics (3 / 3)

C. Apakah ada kolom yang memiliki nilai summary agak aneh?  
(min/mean/median/max/unique/top/freq)

Kolom Kategori

- Job : terdapat 12 pekerjaan berbeda, mayoritas 'blue-collar' (21.5%).
- Marital : lebih dari 50% klien sudah menikah.
- Education : lebih dari 50% secondary.
- Housing : lebih dari 50% client memiliki pinjaman rumah.
- Loan : mayoritas client tidak memiliki personal loan.
- Contact : lebih dari 50% client menggunakan telepon seluler.
- Month : paling banyak di bulan May saat melakukan campaign.
- Poutcome : ~80% memiliki value 'unknown'
- y : 88.3% client tidak membeli deposito. Kolom yang merupakan target ini memiliki class-imbalance.

# Insights Univariate Analysis (1 / 4)

Numerical

Age:

1. slight skew positif = mean > modus/median
2. tidak ditemukan low outlier
3. ditemukan high outlier = ada yang nilainya jauh diatas normal

Day :

4. multimodal distribution
5. tidak ditemukan low outlier dan high outlier

Balance :

6. skew positif = mean > modus/median
7. ditemukan low outlier = ada yang nilainya jauh dibawah normal
8. ditemukan high outlier = ada yang nilainya jauh diatas normal

# Insights Univariate Analysis (2 / 4)

Numerical

Duration :

1. skew positif = mean > modus/median
2. tidak ditemukan low outlier = tidak ada yang nilainya jauh dibawah normal
3. ditemukan high outlier = ada yang nilainya jauh diatas normal

Campaign :

4. skew positif = mean > modus/median
5. tidak ditemukan low outlier = tidak ada yang nilainya jauh dibawah normal
6. ditemukan high outlier = ada yang nilainya jauh diatas normal

Pdays :

7. skew positif = mean > modus/median
8. tidak ditemukan low outlier = tidak ada yang nilainya jauh dibawah normal
9. ditemukan high outlier = ada yang nilainya jauh diatas normal

# Insights Univariate Analysis (3 / 4)

Numerical

Previous :

1. skew positif = mean > modus/median
2. tidak ditemukan low outlier = tidak ada yang nilainya jauh dibawah normal
3. ditemukan high outlier = ada yang nilainya jauh diatas normal
- 4.

Categorical

Job : Terdapat 12 kategori, kemungkinan terlalu banyak (perlu grouping). Top 3 nya ‘blue-collar’, ‘management’, dan ‘technician’ (masing-masing memiliki nilai di atas 7000 records). Terdapat ‘unknown’ job dengan jumlah records di bawah 500

Marital : ‘married’ menempati posisi tertinggi dengan jumlah records > 25,000, ‘single’ dan ‘divorced’ berada pada posisi kedua dan ketiga. Possibly pada tahap pre-processing dilakukan grouping menjadi married\_flag 'yes' and 'no'.

Education : ‘secondary’ menempati posisi tertinggi dengan > 20,000 records, diikuti dengan ‘tertiary’, ‘primary’, dan ‘unknown’. ‘unknown’ < 2,500 records. Pada tahap preprocessing kemungkinan dilakukan label encoding.

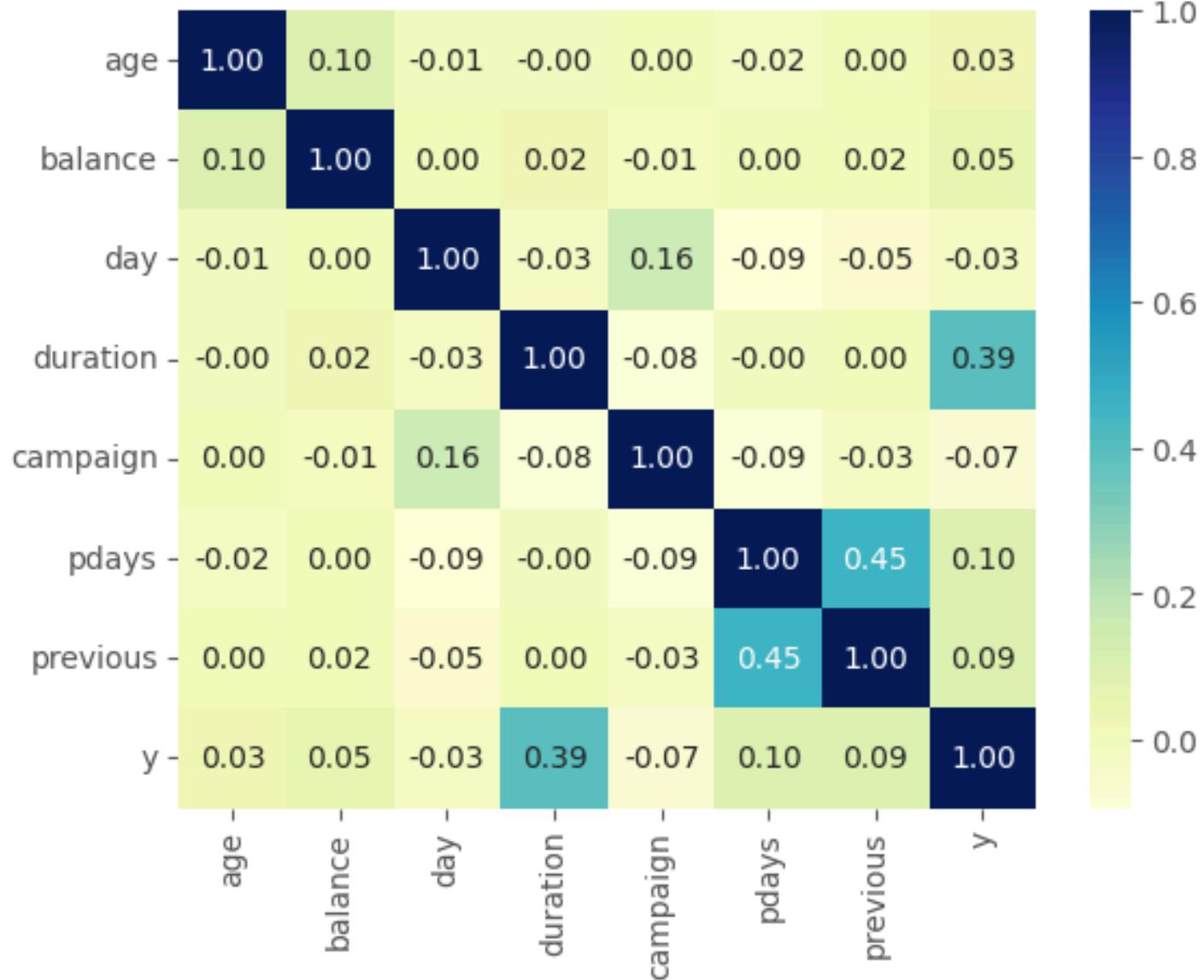
# Insights Univariate Analysis (4 / 4)

## Categorical

1. Default : hampir semua records memiliki value ‘no’
2. Housing : posisi pertama ‘yes’ dengan jumlah record sekitar 25,000, sisanya ‘no’. distribusi keduanya cukup seimbang
3. Loan : ‘no’ menempati posisi pertama (jumlah record > 35,000), sisanya ‘yes’
4. Contact : ‘cellular’ menempati posisi pertama (jumlah record > 25,000), diikuti ‘unknown’ dan ‘telephone’. Jumlah ‘unknown’ cukup banyak, sekitar 13,000 records, kemungkinan memerlukan penyesuaian pada tahap pre-processing.
5. Month : ‘may’ menempati posisi pertama (jumlah record > 12,000), diikuti ‘jul’ dan ‘aug’
6. Poutcome : ‘unknown’ menempati posisi tertinggi dengan jumlah record > 35,000. Jumlah ‘success’ < 2,500
7. y : ‘no’ berada di posisi tertinggi, dengan jumlah record sekitar 40,000. ‘yes’ sekitar 5,000 record. Kolom yang merupakan target ini memiliki class-imbalance, kemungkinan besar akan dilakukan oversampling pada tahap selanjutnya.

# Insights Multivariate Analysis (1 / 2)

## Correlation Heatmap - Numerical Features



**A. Bagaimana korelasi antara masing-masing feature dan label. Kira-kira feature mana saja yang paling relevan dan harus dipertahankan?**

Korelasi antar feature numerik dengan target

- y sebagai variabel target / variabel dependent / label, y adalah variabel yang ingin diprediksi dari variabel bebas lainnya.
- Tidak ada korelasi yang kuat antara target y dengan feature numerikal yang ada, mengindikasikan penggunaan model non-linear lebih tepat untuk dataset ini. Nilai korelasi tertinggi target-feature ada pada y dengan duration, yaitu 0.39 (tidak cukup kuat).

Hubungan antar feature kategorikal dengan target

- Klien yang paling tertarik dengan produk deposito adalah klien dengan status menikah, pendidikan lanjutan 'secondary' dan 'tertiary', tidak memiliki pinjaman KPR, tidak memiliki hutang di bank (personal loan), dikontak menggunakan cellular/HP, tidak pernah gagal membayar hutang (default = 'no'), memiliki rata-rata balance positif

Dari penjabaran plot multivariat numerik dan kategorikal di atas, semua feature numerik dan kategorikal akan dipertahankan untuk modeling iterasi pertama.

## Insights Multivariate Analysis (2 / 2)

**B. Bagaimana korelasi antar-feature, apakah ada pola yang menarik? Apa yang perlu dilakukan terhadap feature itu?**

- Terdapat korelasi positif dari feature numerik yang cukup kuat (0.45), yaitu antara pdays dengan previous. Dari hasil univariate analysis sebelumnya, diketahui pada sample data mayoritas client belum mendapat campaign, sehingga ber-value -1 pada pdays dan 0 pada previous. Kemungkinan besar hal ini menyebabkan tingginya nilai korelasi antara kedua feature tersebut
- Dari heatmap korelasi feature numerik tidak ada feature pasangan dengan nilai korelasi  $>0.7$ , kemungkinan besar tidak ada data yang redundant.

# Business Insights (1 / 3)

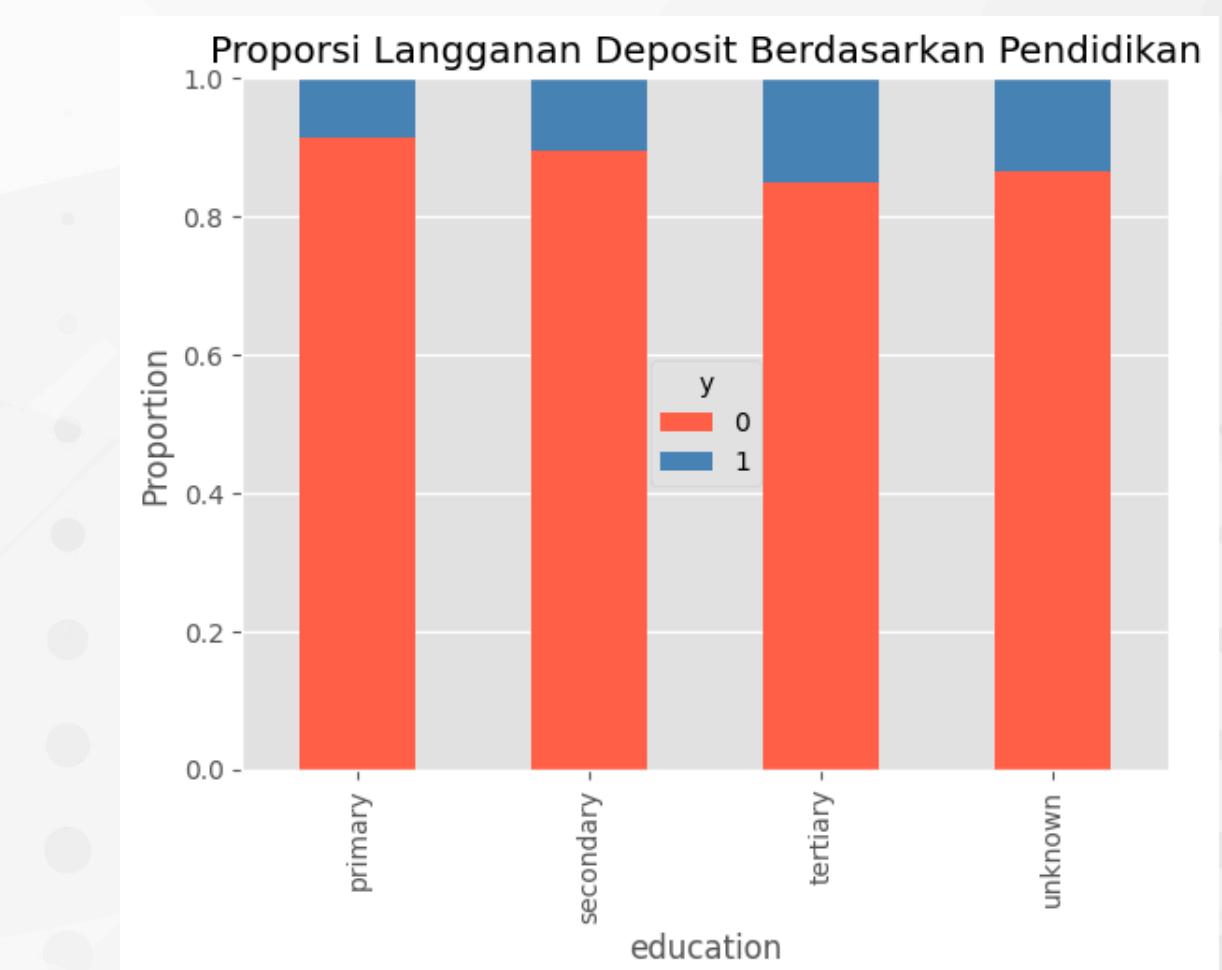
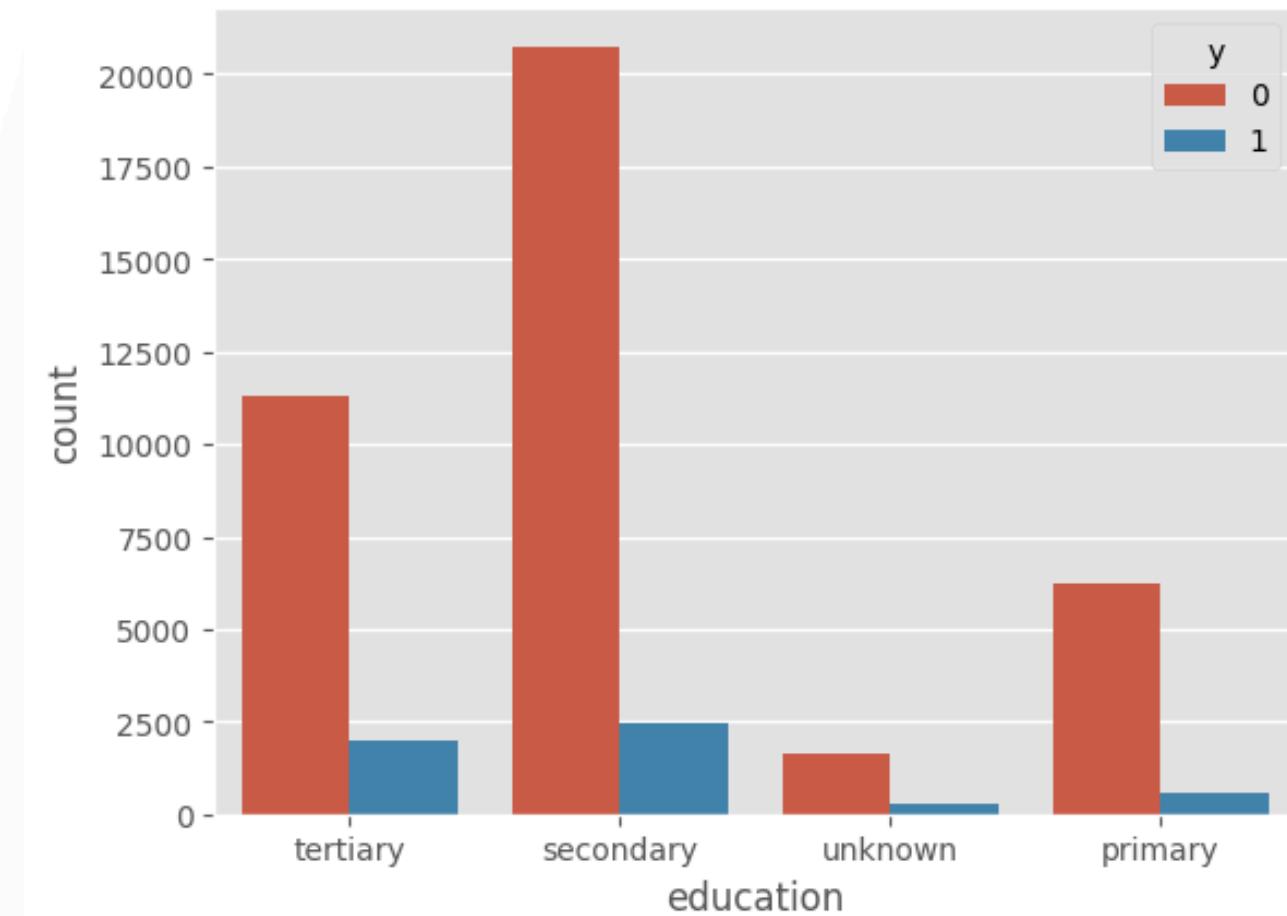
Dari hasil EDA yang dilakukan, terdapat beberapa quick insights yang bisa dipertimbangkan dalam upaya meningkatkan performa campaign term-deposit oleh telesales:

1. Secara umum, client yang tertarik dengan produk deposito adalah klien dengan karakteristik

- usia produktif: 20-60 b. menikah
- pendidikan lanjutan 'secondary' dan 'tertiary'
- tidak memiliki cicilan rumah
- tidak memiliki hutang di bank

Tim Marketing dapat memusatkan effort terhadap orang-orang dengan karakteristik seperti di atas.

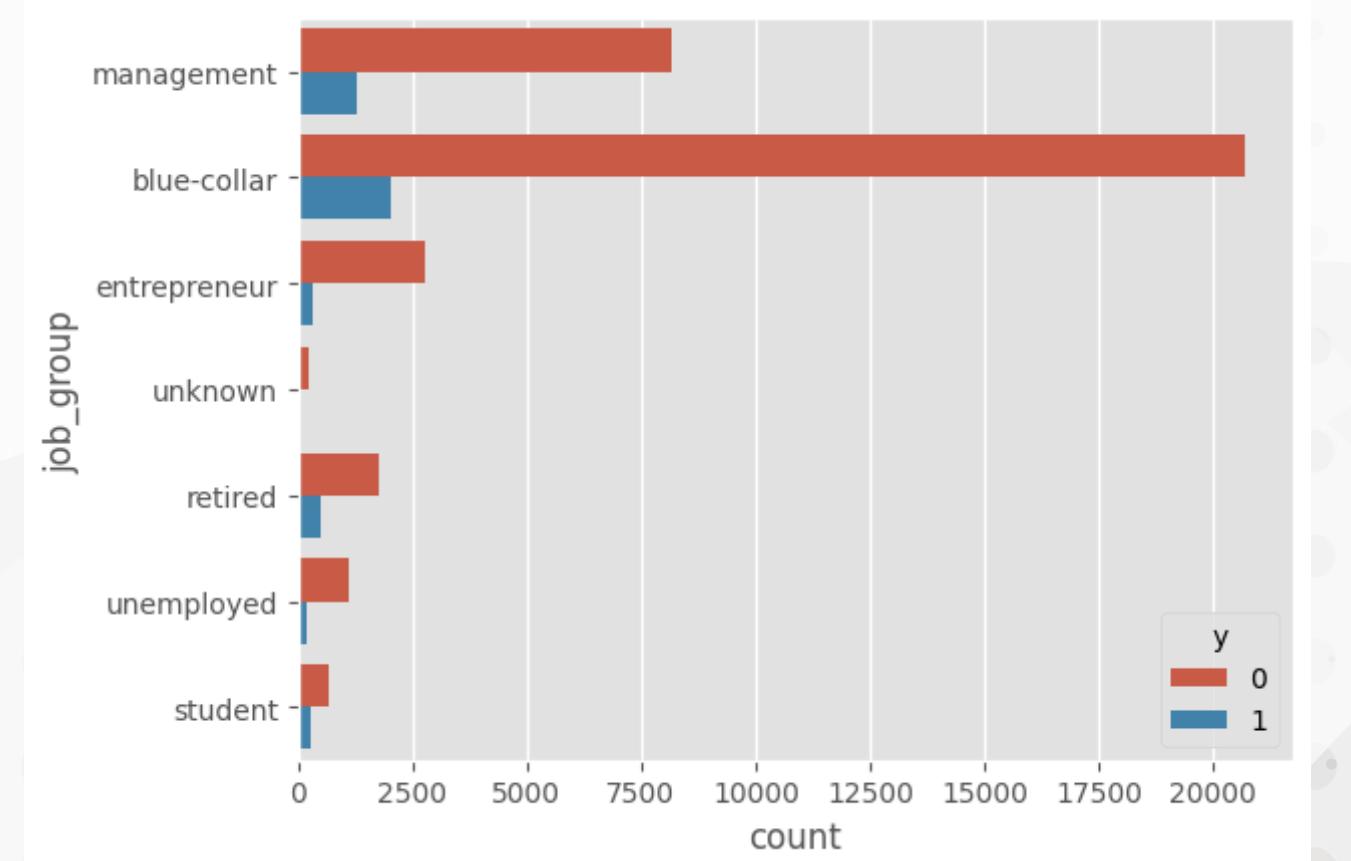
2. Pendidikan. Dari grafik di samping dapat dilihat bahwa mayoritas client yang sign-up untuk term-deposit memiliki pendidikan lanjutan (secondary dan tertiary). Jika dilihat stacked bar 100% (proporsi yang mengambil term-deposit untuk tiap jenjang pendidikan), % yes meningkat seiring dengan bertambah tingginya pendidikan. Tim marketing mungkin bisa memberikan approach yang berbeda kepada client dengan pendidikan lebih rendah, yaitu dengan menjelaskan manfaat term-deposit menggunakan bahasa yang mudah dipahami.



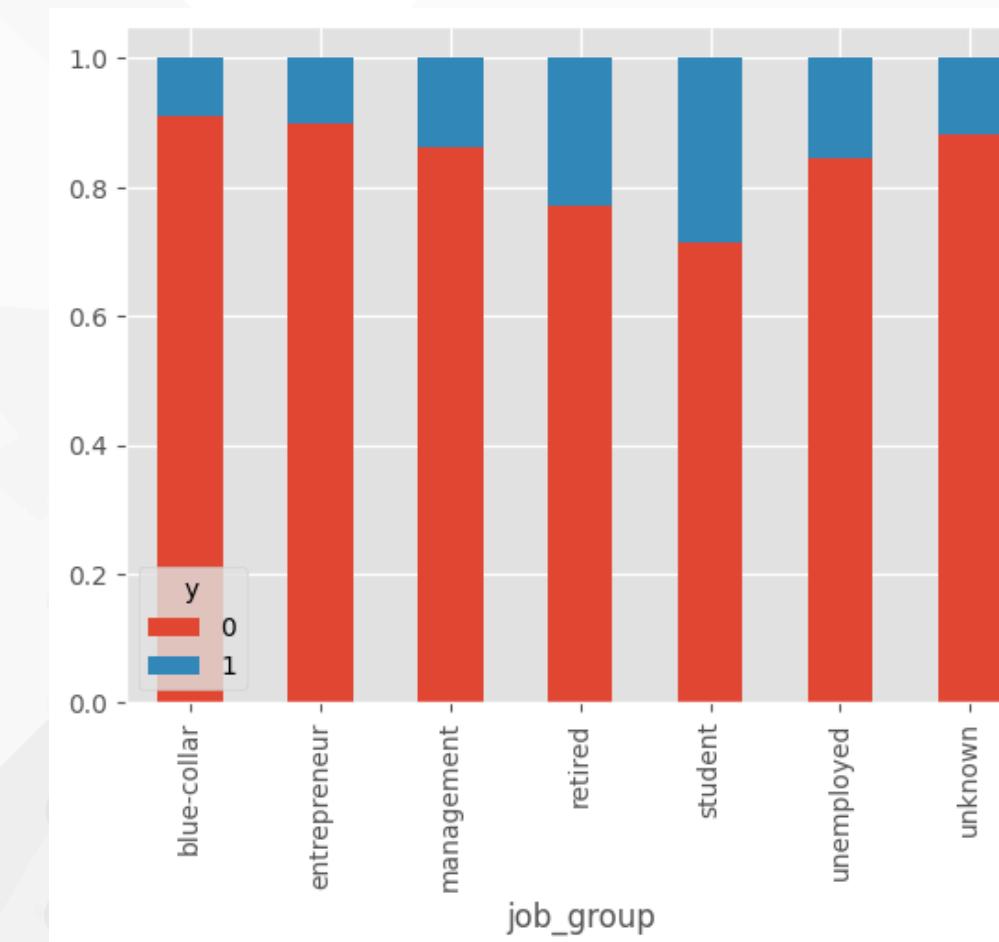
# Business Insights (2 / 3)

## 3. Jenis pekerjaan.

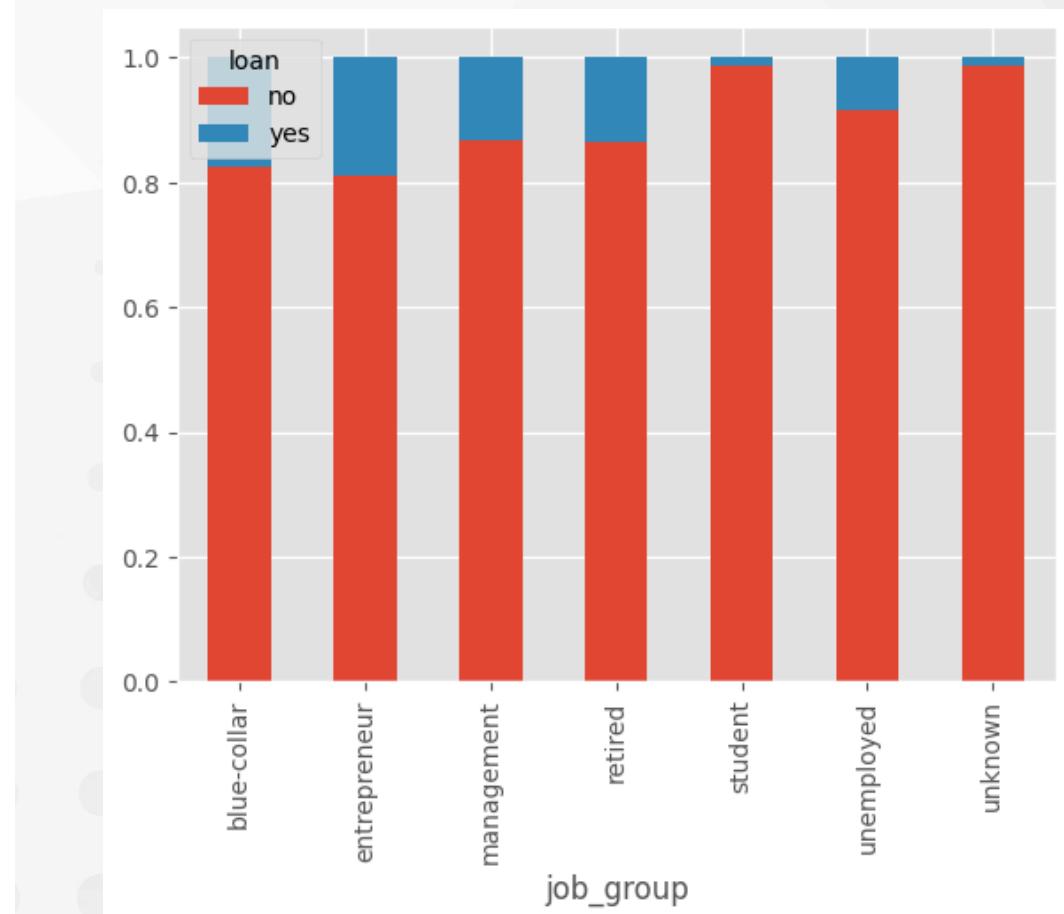
- a. Setelah dilakukan pengelompokan jenis pekerjaan, dari plot di bawah dapat dilihat client yang mengambil term-deposit mayoritas memiliki pekerjaan: blue-collar, management. Namun demikian, jika dilihat % yes dari masing-masing pekerjaan, % yes blue-collar relatif rendah. Jika sample menggambarkan populasi, maka kemungkinan banyak client bank yang memiliki pekerjaan blue-collar. Tim marketing perlu re-assess apakah product term-deposit dan approach yang dilakukan sudah tepat untuk segment ini.
- b. Terlihat % yes untuk student dan retired lebih tinggi dibanding job group lain. Pengamatan ini dapat dijadikan basis agar tim marketing dapat merancang product dan marketing approach yang lebih menarik lagi untuk student dan retired. Diharapkan dengan adanya product yang lebih tailored untuk 2 segment ini, akan ada akuisisi baru dan kedua segment bisa bertambah besar
- c. % yes untuk entrepreneur pun relatif rendah. Segment ini kemungkinan besar memerlukan modal usaha. Nampak dari visualisasi 3C bahwa hampir 20% entrepreneur memiliki personal loan (relatif lebih tinggi dibanding segment lain). Sebaiknya tim marketing mempertimbangkan apakah lebih menguntungkan dan menarik bila segment ini difokuskan untuk ditawarkan loan



3A



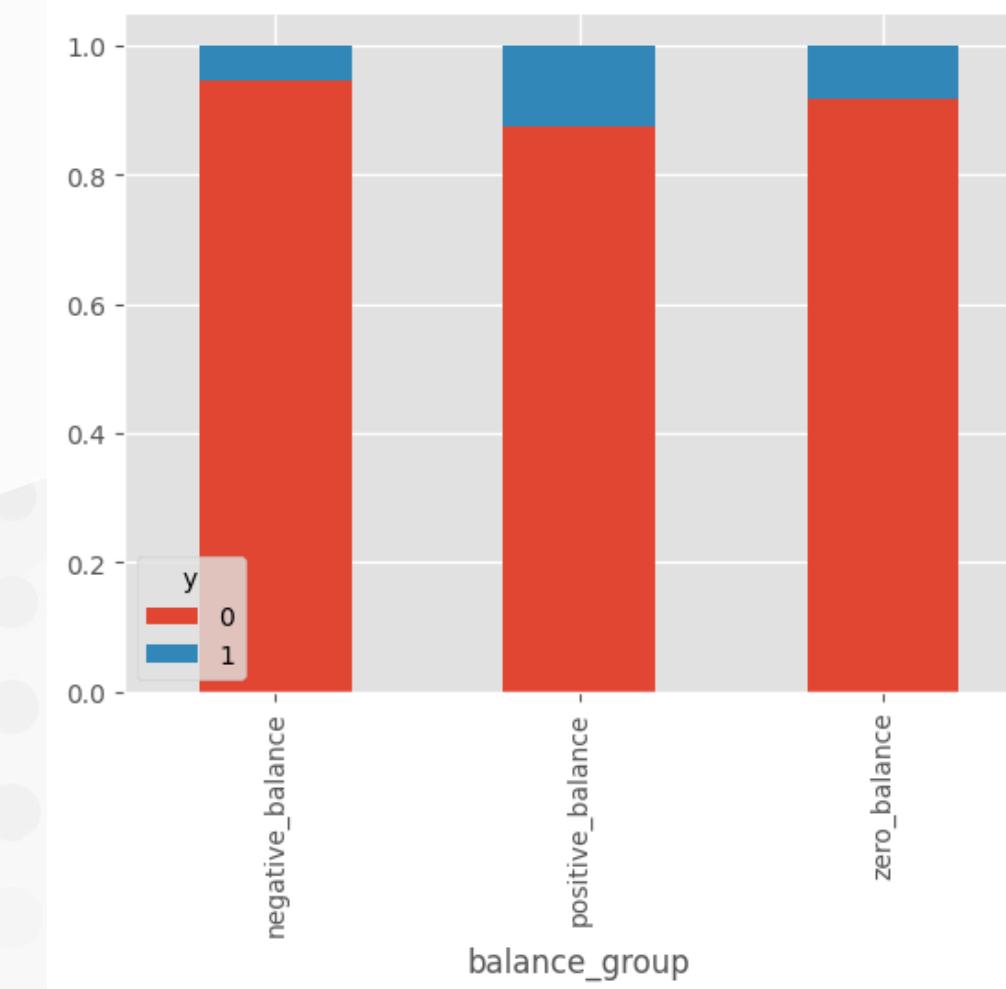
3B



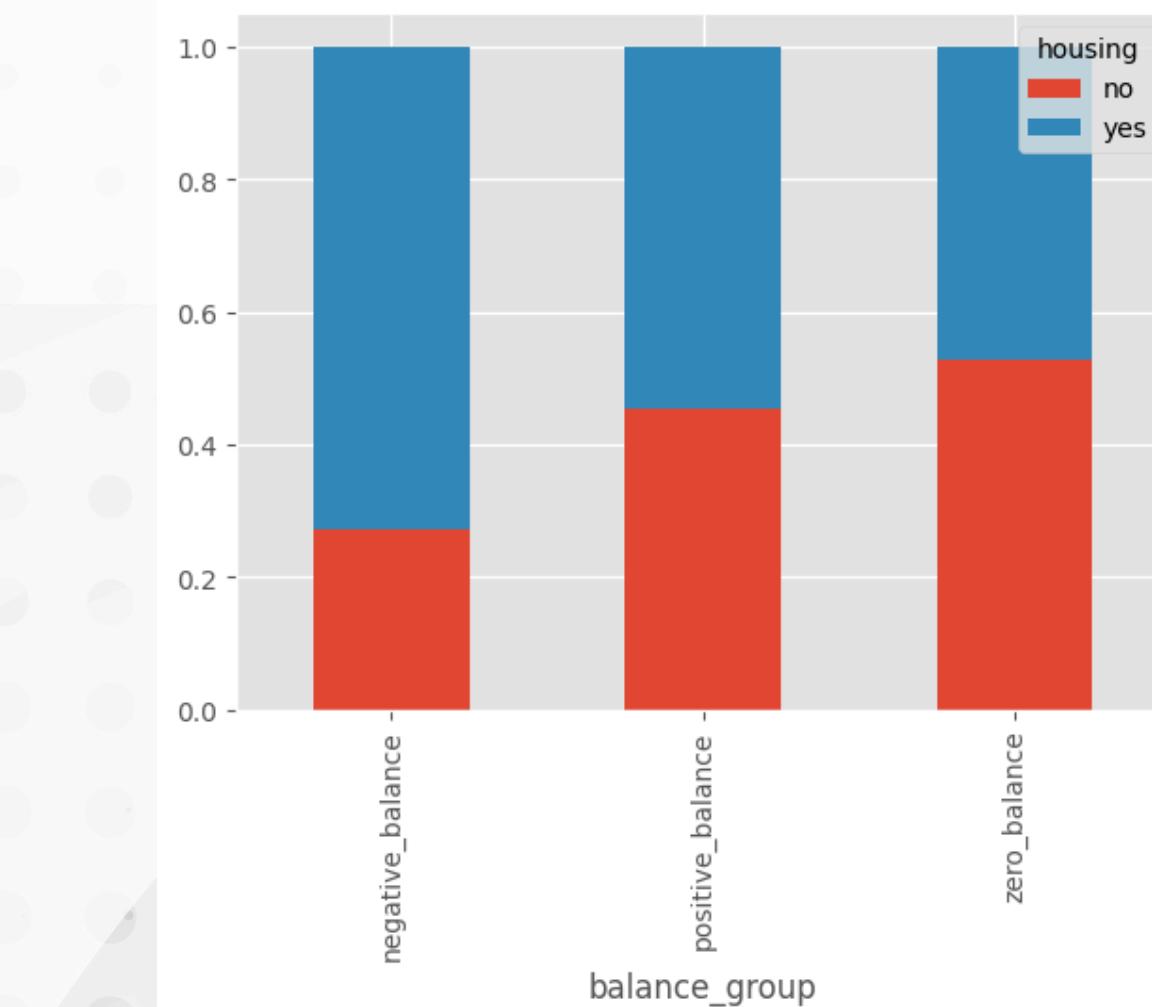
3C

# Business Insights (3 / 3)

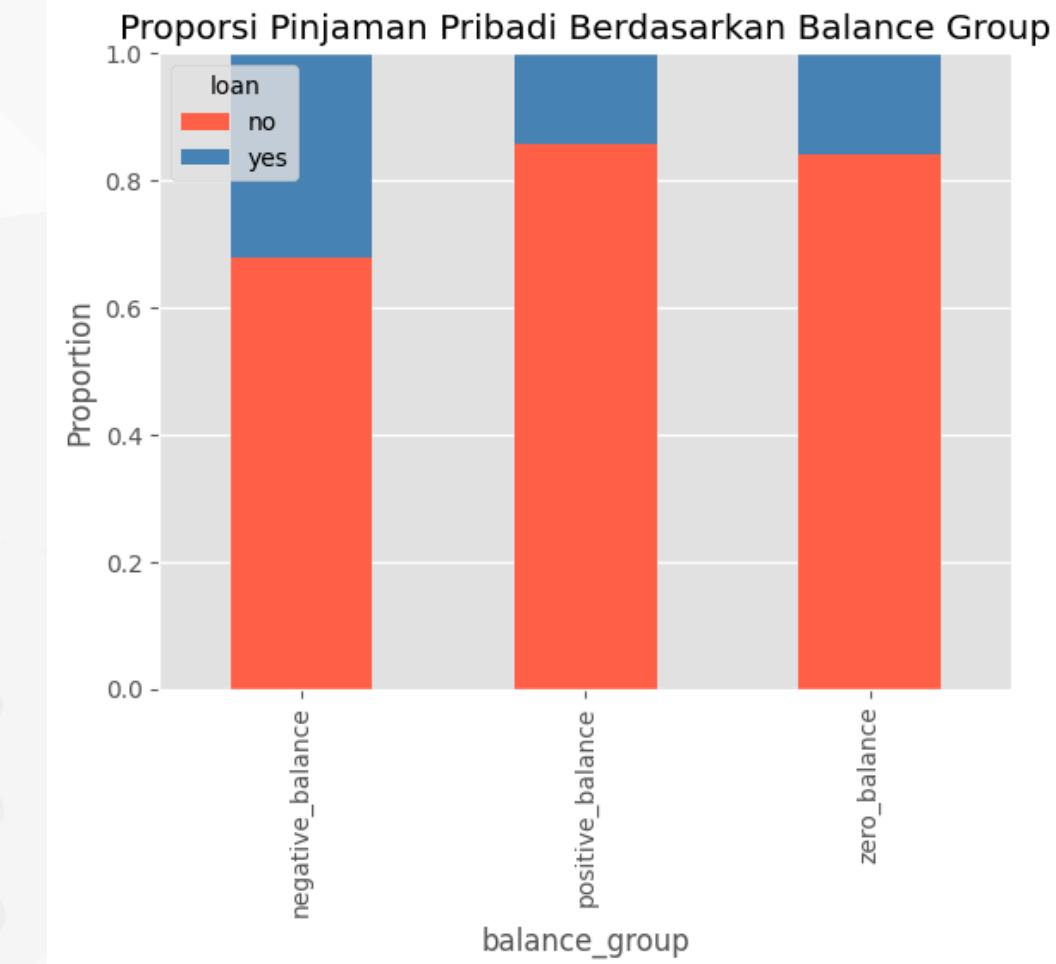
4. Average yearly balance. Untuk performa campaign yang lebih baik, sebaiknya orang-orang dengan average yearly balance negative diexclude dari whitelist telemarketing campaign. Selain % yes take up term-depositnya paling rendah dibanding 2 segment lain (visualisasi 4A), sekitar 70% dari orang2 ini memiliki housing loan dan 30% memiliki personal loan (vis 4B dan 4C). Adanya loan mengindikasikan rendah atau tidak adanya 'cold money' yang bisa digunakan untuk investasi (seperti pada term-deposit). Sementara itu, client dengan average yearly balance = 0 mengindikasikan bahwa akun mereka lebih bersifat transaksional. Client hanya akan menambah saldo ketika perlu melakukan pembayaran atau transfer. Client seperti ini mungkin memiliki 'cold money' di insititusi lain. Jika kita bisa memperoleh data biro (seperti SLIK) yang mengcompile data asset dan liability client di insititusi lain, kita dapat membuat program untuk menarik minat mereka agar berinvestasi pada bank kita.



4A



4B



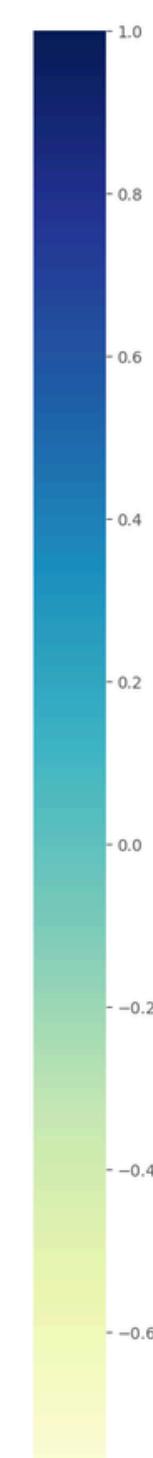
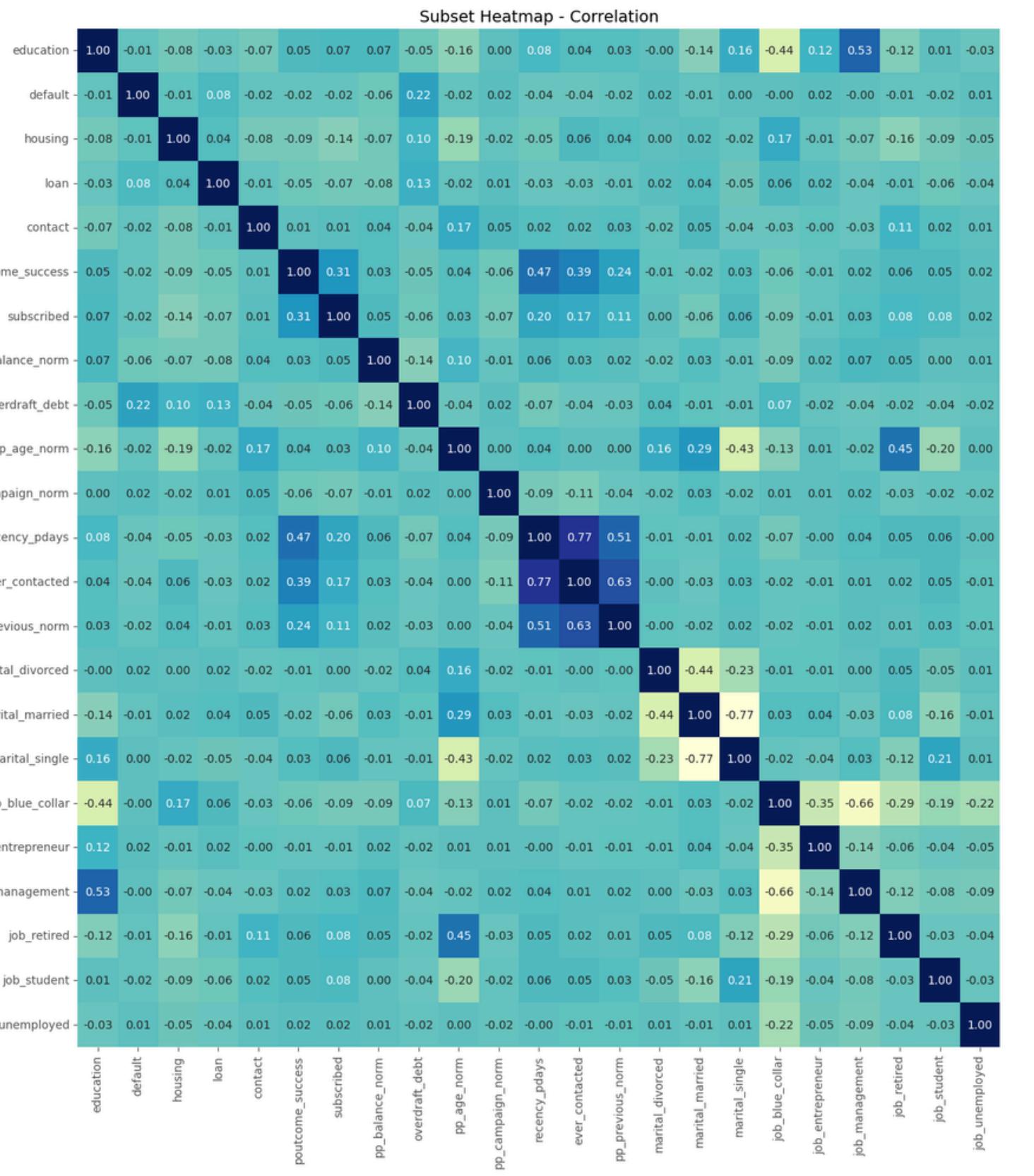
4C

# Stage 2

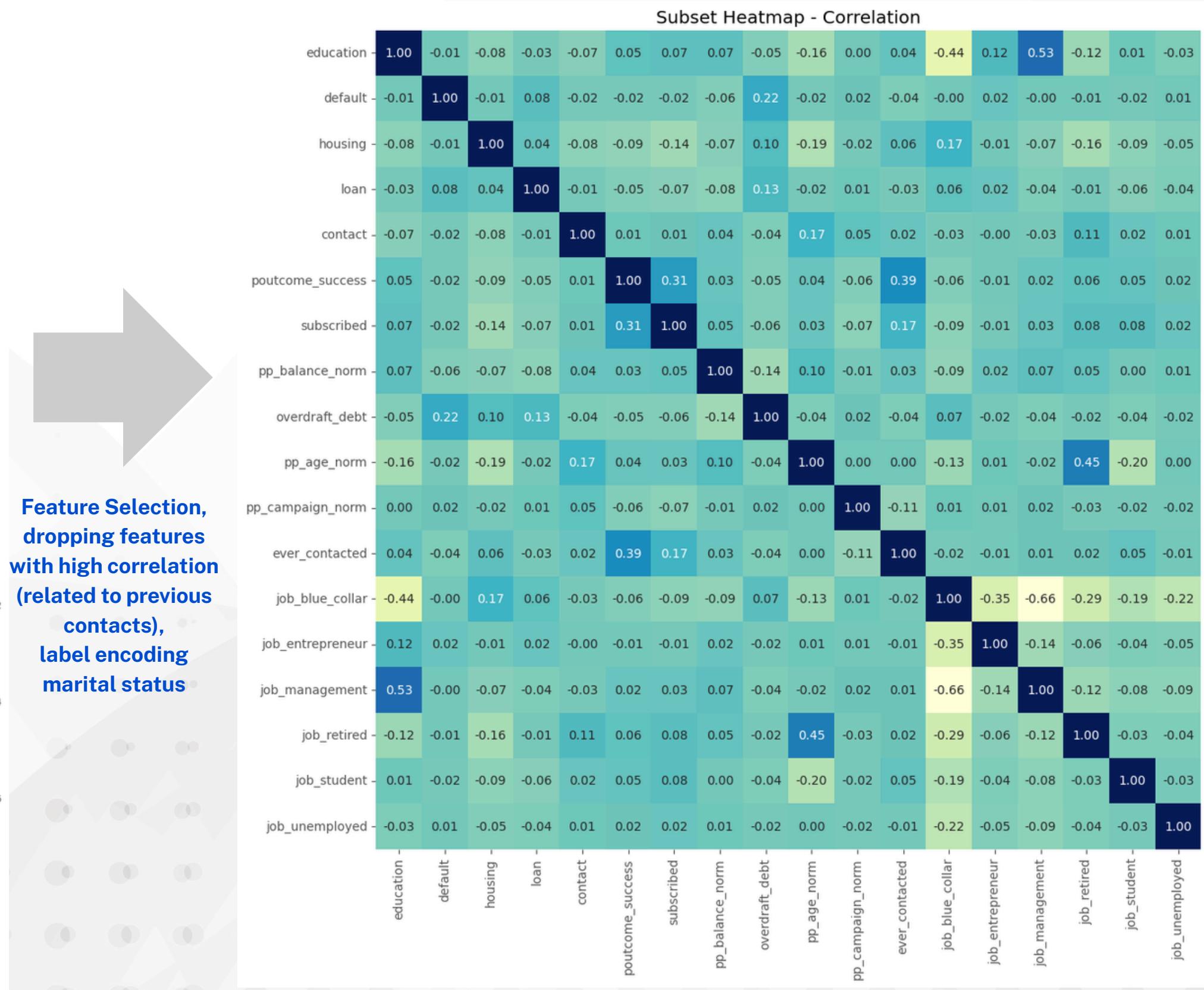
# Summary Preprocessing (1A - 2B) - Version 1

Feature	Type	Preprocessing
age	numerical	Handle Outliers, Feature transformation
job	categorical	Handle missing value (using mode from education), Feature extraction (create job group to reduce unique values), Feature encoding - OHE Job Group (non ordinal category)
marital	categorical	Feature encoding (Label encoding)
education	categorical	Handle missing value, Feature encoding
default	categorical	Feature encoding (Label encoding)
balance	numerical	Handle Outliers, Feature transformation,
housing	categorical	Feature encoding (Label encoding)
loan	categorical	Feature encoding (Label encoding)
contact	categorical	Handle missing value (using mode ),Feature encoding (Label encoding)
day	numerical	Handle Outliers, Feature transformation
month	categorical	Feature encoding (Label encoding), Feature extraction (ever_contacted)
duration	numerical	Handle Outliers, Feature transformation, Feature extraction (partner)
campaign	numerical	Handle Outliers, Feature transformation
pdays	numerical	Feature selection ( Drop Feature)
previous	numerical	Feature selection (Drop Feature)
poutcome	categorical	Feature encoding (Label encoding)
y	categorical	Feature encoding, handle class-imbalance

# Summary Preprocessing (1A - 2B) - Version 1



Feature Selection,  
dropping features  
with high correlation  
(related to previous  
contacts),  
label encoding  
marital status



# Summary Preprocessing (1A - 2B) - Version 1

## df\_train\_fin

```
→ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 45211 entries, 0 to 45210
Data columns (total 19 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   education        45211 non-null   int64  
 1   default          45211 non-null   int64  
 2   housing          45211 non-null   int64  
 3   loan              45211 non-null   int64  
 4   contact           45211 non-null   int64  
 5   poutcome_success 45211 non-null   int64  
 6   subscribed        45211 non-null   int64  
 7   pp_balance_norm  45211 non-null   float64 
 8   overdraft_debt   45211 non-null   int64  
 9   pp_age_norm      45211 non-null   float64 
 10  pp_campaign_norm 45211 non-null   float64 
 11  ever_contacted   45211 non-null   int64  
 12  is_married        45211 non-null   int64  
 13  job_blue_collar  45211 non-null   bool   
 14  job_entrepreneur 45211 non-null   bool   
 15  job_management    45211 non-null   bool   
 16  job_retired       45211 non-null   bool   
 17  job_student        45211 non-null   bool   
 18  job_unemployed    45211 non-null   bool  
dtypes: bool(6), float64(3), int64(10)
memory usage: 4.7 MB
```

## handling class-imbalance

```
→ Original
False    39922
True     5289
Name: count, dtype: int64

UNDERSAMPLING
False    10578
True     5289
Name: count, dtype: int64

OVERSAMPLING
False    39922
True     19961
Name: count, dtype: int64

SMOTE
False    39922
True     19961
Name: count, dtype: int64
```

## 2C. Feature Engineering - Tambahan Fitur

- **Jumlah hari sejak tanggal transaksi terakhir** —> sebagai proxy untuk ‘recency’, orang yang baru melakukan transaksi kemungkinan lebih responsif dibandingkan orang yang sudah lama tidak mengakses akunnya. Pada dataset ada pdays, yang menunjukan recency contact customer, namun demikian sebagian besar customer belum pernah dikontak. Oleh karena itu, mungkin feature ini lebih berguna untuk mewakili recency
- **Jumlah tanggungan (manusia) yang dimiliki** —> bisa menjadi indikator apakah orang tersebut kemungkinan memiliki uang untuk diinvestasikan
- **Rata-rata frekuensi transaksi bulanan** —> untuk menggambarkan seberapa sering (frequency) client melakukan interaksi dengan akun. Jika interaksinya sering, bisa jadi client tersebut nyaman dengan bank kita, dan akan lebih mudah menawarkan produk deposito
- **Riwayat term-deposit client pada bank kita** —> jika pernah melakukan deposito, dilihat recency, frequency dan monetary nya. Semakin baik skor-skor tersebut, makin mungkin client berminat untuk menambah/ menaruh uang ulang pada product term deposito
- **Jumlah asset yang dimiliki di insititusi finansial lain (misalnya menggunakan data biro SLIK)** —> bisa mengetahui kekayaan (potensi memiliki uang untuk diinvestasikan), apakah akun pada bank kita hanya untuk transaksi (misal di bank lain saldonya tinggi, tapi di kita kecil), apakah punya deposito di bank lain (familiaritas dengan produk deposito, lebih mudah untuk ditawari)
- **Jumlah liabilitas yang dimiliki di insititusi finansial lain (misalnya menggunakan data biro SLIK)** —> orang-orang yang memiliki banyak hutang di tempat lain kemungkinan tidak tertarik membeli produk deposito karena harus melunasi hutangnya

# Stage 3

# Model Evaluation and Hyperparameter Tuning

## Decision Tree

Algoritma ini digunakan karena robust terhadap data yang memiliki outlier

## Hasil Model Evaluasi

```
Model Evaluation (Decision Tree)
Accuracy (Test Set): 0.78
Precision (Test Set): 0.21
Recall (Test Set): 0.31
F1-Score (Test Set): 0.25
roc_auc (test-proba): 0.57
roc_auc (train-proba): 1.00
```

## Setelah Hyperparameter tuning

```
Accuracy (Test Set): 0.76
Precision (Test Set): 0.23
Recall (Test Set): 0.44
F1-Score (Test Set): 0.30
roc_auc (test-proba): 0.66
roc_auc (train-proba): 0.80
```

- Metrics model yang diprioritaskan: Recall, karena tujuan dari modeling ini adalah memaksimalkan prediksi klien yang benar-benar tertarik (positif) terhadap penawaran produk bank.
- Setelah dilakukan hyperparameter tuning, nilai Recall justru meningkat sebesar 13%, dari sebelumnya 0.31 menjadi 0.44.

# Model Evaluation and Hyperparameter Tuning

## Random Forest

Algoritma ini digunakan karena robust, prediksi model lebih akurat dapat mengurangi variance / mencegah overfitting

## Hasil Model Evaluasi

```
Model Evaluation (Random Forest)
Accuracy (Test Set): 0.82
Precision (Test Set): 0.27
Recall (Test Set): 0.33
F1-Score (Test Set): 0.30
roc_auc (test-proba): 0.66
roc_auc (train-proba): 1.00
```

## Setelah Hyperparameter tuning

```
Accuracy (Test Set): 0.89
Precision (Test Set): 0.67
Recall (Test Set): 0.15
F1-Score (Test Set): 0.24
roc_auc (test-proba): 0.74
roc_auc (train-proba): 0.81
```

- Metrics yang difokuskan tetap Recall, karena target modeling adalah menemukan sebanyak mungkin klien potensial.
- Sayangnya, recall justru menurun drastis setelah tuning, yang berarti banyak klien potensial tidak berhasil dikenali oleh model.
- Overfitting memang berkurang, tetapi trade-off-nya adalah model menjadi terlalu konservatif, hanya menandai klien-klien yang sangat pasti (makanya precision naik).

# Model Evaluation and Hyperparameter Tuning

## Logistic Regression

Algoritma ini digunakan karena pada dataset yang digunakan tidak ada multikolinieritas dan algoritma ini memiliki kemampuan komputasi yang cepat.

## Hasil Model Evaluasi

```
Model Evaluation (Logistic Regression)
Accuracy (Test Set): 0.70
Precision (Test Set): 0.22
Recall (Test Set): 0.61
F1-Score (Test Set): 0.32
roc_auc (test-proba): 0.72
roc_auc (train-proba): 0.72
```

## Setelah Hyperparameter tuning

```
Accuracy (Test Set): 0.89
Precision (Test Set): 0.68
Recall (Test Set): 0.14
F1-Score (Test Set): 0.24
roc_auc (test-proba): 0.70
roc_auc (train-proba): 0.71
```

- Metrik utama tetap Recall, karena goal utama adalah mendeteksi sebanyak mungkin calon pelanggan potensial.
- Setelah tuning, recall justru menurun drastis, walau precision meningkat.
- Trade-off precision vs recall ini menunjukkan bahwa tuning berhasil mengurangi false positive, tapi dengan konsekuensi melewatkannya terlalu banyak klien potensial.

# Model Evaluation and Hyperparameter Tuning

## K-Nearest Neighbor

Algoritma ini digunakan karena memiliki cara kerja yang sederhana selain itu algoritma ini cocok digunakan untuk data yang bersifat non-linear.

## Hasil Model Evaluasi

```
Model Evaluation (KNN)
Accuracy (Test Set): 0.72
Precision (Test Set): 0.21
Recall (Test Set): 0.50
F1-Score (Test Set): 0.30
roc_auc (test-proba): 0.65
roc_auc (train-proba): 0.94
```

## Setelah Hyperparameter tuning

```
Accuracy (Test Set): 0.85
Precision (Test Set): 0.29
Recall (Test Set): 0.22
F1-Score (Test Set): 0.25
roc_auc (test-proba): 0.63
roc_auc (train-proba): 1.00
```

- Metrik utama tetap Recall, dan seharusnya nilai ini meningkat setelah tuning – tetapi yang terjadi justru sebaliknya.
- Model overfit setelah tuning (train AUC = 1.00), artinya model terlalu mengikuti pola di data training.

# Model Evaluation and Hyperparameter Tuning

## Naive Bayes

Algoritma ini digunakan karena memiliki waktu komputasi yang cepat.

### Hasil Model Evaluasi

```
Model Evaluation (Naive Bayes)
Accuracy (Test Set): 0.81
Precision (Test Set): 0.28
Recall (Test Set): 0.38
F1-Score (Test Set): 0.32
roc_auc (test-proba): 0.70
roc_auc (train-proba): 0.70
```

- Metrics model : Recall. Karena tujuan dari modeling ini kita ingin mendapatkan rasio prediksi data yang positif dari keseluruhan data yang benar positif

# Model Evaluation and Hyperparameter Tuning

## XGBoost

Algoritma ini digunakan karena robust, prediksi model lebih akurat dapat mengurangi bias/ mencegah underfitting

## Hasil Model Evaluasi

```
Model Evaluation (XGBoost)
Accuracy (Test Set): 0.86
Precision (Test Set): 0.38
Recall (Test Set): 0.33
F1-Score (Test Set): 0.35
roc_auc (test-proba): 0.71
roc_auc (train-proba): 0.81
```

## Setelah Hyperparameter tuning

```
Accuracy (Test Set): 0.84
Precision (Test Set): 0.31
Recall (Test Set): 0.28
F1-Score (Test Set): 0.29
roc_auc (test-proba): 0.66
roc_auc (train-proba): 0.99
```

- Recall justru menurun sebesar 5%, yang menunjukkan bahwa kemampuan model dalam mendekripsi kelas positif menjadi lebih buruk setelah tuning.
- Terjadi overfitting, terlihat dari selisih yang sangat besar antara roc\_auc training dan testing (0.99 vs 0.66). Model terlalu fit terhadap data training, tetapi gagal melakukan generalisasi dengan baik ke data test.

# Model Evaluation and Hyperparameter Tuning

## Adaboost

Algoritma ini digunakan karena robust, prediksi model lebih akurat dapat mengurangi bias/ mencegah underfitting

## Hasil Model Evaluasi

```
Model Evaluation (Adaboost)
Accuracy (Test Set): 0.78
Precision (Test Set): 0.28
Recall (Test Set): 0.52
F1-Score (Test Set): 0.36
roc_auc (test-proba): 0.72
roc_auc (train-proba): 0.72
```

## Setelah Hyperparameter tuning

```
Accuracy (Test Set): 0.89
Precision (Test Set): 0.68
Recall (Test Set): 0.13
F1-Score (Test Set): 0.22
roc_auc (test-proba): 0.73
roc_auc (train-proba): 0.73
```

- Metrics model : Recall. Karena tujuan dari modeling ini kita ingin mendapatkan rasio prediksi data yang positif dari keseluruhan data yang benar positif
- Setelah dilakukan hyperparameter tuning, nilai recall menurun sebesar 39%, meskipun akurasi dan precision meningkat tetapi tidak terjadi overfitting

# **Stage 3**

## **PCA**

# Model Evaluation and Hyperparameter Tuning

## Decision Tree

Algoritma ini digunakan karena robust terhadap data yang memiliki outlier

## Hasil Model Evaluasi

```
Model Evaluation (Decision Tree with PCA)
Accuracy (Test Set): 0.75
Precision (Test Set): 0.19
Recall (Test Set): 0.35
F1-Score (Test Set): 0.24
roc_auc (test-proba): 0.57
roc_auc (train-proba): 1.00
```

## Setelah Hyperparameter tuning

```
--- Hasil Setelah Tuning dengan PCA ---
Accuracy (Test Set): 0.73
Precision (Test Set): 0.21
Recall (Test Set): 0.47
F1-Score (Test Set): 0.29
roc_auc (test-proba): 0.65
roc_auc (train-proba): 0.88
```

- **Metrics model: Recall.** Karena tujuan utama dari modeling ini adalah untuk memaksimalkan rasio deteksi positif yang benar dari keseluruhan data yang seharusnya positif (**false negative**).
- **Performa awal:** Recall masih rendah (35%), menunjukkan bahwa model awal cukup lemah dalam mendekripsi data positif meskipun training AUC tinggi (overfitting). Namun setelah hyperparameter tuning, Recall meningkat sebesar 12% (dari 35% → 47%), yang merupakan peningkatan positif.

# Model Evaluation and Hyperparameter Tuning

## Random Forest

Algoritma ini digunakan karena robust, prediksi model lebih akurat dapat mengurangi variance / mencegah overfitting

## Hasil Model Evaluasi

```
Model Evaluation (Random Forest with PCA)
Accuracy (Test Set): 0.77
Precision (Test Set): 0.22
Recall (Test Set): 0.35
F1-Score (Test Set): 0.27
roc_auc (test-proba): 0.64
roc_auc (train-proba): 1.00
```

## Setelah Hyperparameter tuning

```
--- Hasil Setelah Tuning dengan PCA ---
Accuracy (Test Set): 0.74
Precision (Test Set): 0.23
Recall (Test Set): 0.50
F1-Score (Test Set): 0.31
roc_auc (test-proba): 0.68
roc_auc (train-proba): 0.86
```

- Setelah tuning, recall meningkat sebesar 15% (dari 35% → 50%), dan overfitting berkurang (train AUC turun dari 1.00 ke 0.86).
- Meski accuracy sedikit menurun, model menjadi lebih generalizable dan tidak overfit.

# Model Evaluation and Hyperparameter Tuning

## Logistic Regression

Algoritma ini digunakan karena pada dataset yang digunakan tidak ada multikolinieritas dan algoritma ini memiliki kemampuan komputasi yang cepat.

## Hasil Model Evaluasi

```
Model Evaluation (Logistic Regression with PCA)
Accuracy (Test Set): 0.65
Precision (Test Set): 0.20
Recall (Test Set): 0.63
F1-Score (Test Set): 0.30
roc_auc (test-proba): 0.69
roc_auc (train-proba): 0.70
```

## Setelah Hyperparameter tuning

```
--- Hasil Setelah Tuning dengan PCA ---
Accuracy (Test Set): 0.65
Precision (Test Set): 0.19
Recall (Test Set): 0.63
F1-Score (Test Set): 0.30
roc_auc (test-proba): 0.70
roc_auc (train-proba): 0.70
```

- Setelah dilakukan hyperparameter tuning, nilai recall tetap stabil di angka 63 persen dan performa lainnya tidak mengalami perubahan signifikan.
- Tidak terjadi overfitting karena nilai AUC pada data pelatihan hampir sama dengan nilai AUC pada data pengujian.

# Model Evaluation and Hyperparameter Tuning

## K-Nearest Neighbor

Algoritma ini digunakan karena memiliki cara kerja yang sederhana selain itu algoritma ini cocok digunakan untuk data yang bersifat non-linear.

## Hasil Model Evaluasi

```
Model Evaluation (KNN with PCA)
Accuracy (Test Set): 0.76
Precision (Test Set): 0.23
Recall (Test Set): 0.42
F1-Score (Test Set): 0.30
roc_auc (test-proba): 0.65
roc_auc (train-proba): 0.91
```

## Setelah Hyperparameter tuning

```
--- Hasil Setelah Tuning dengan PCA ---
Accuracy (Test Set): 0.78
Precision (Test Set): 0.21
Recall (Test Set): 0.34
F1-Score (Test Set): 0.26
roc_auc (test-proba): 0.62
roc_auc (train-proba): 1.00
```

- Setelah tuning, recall justru menurun dari 42% menjadi 34%.
- Terjadi overfitting karena AUC train jauh lebih tinggi daripada AUC test.

# Model Evaluation and Hyperparameter Tuning

## Naive Bayes

Algoritma ini digunakan karena memiliki waktu komputasi yang cepat.

## Hasil Model Evaluasi

```
Model Evaluation (Naive Bayes)
Accuracy (Test Set): 0.71
Precision (Test Set): 0.20
Recall (Test Set): 0.47
F1-Score (Test Set): 0.28
roc_auc (test-proba): 0.66
roc_auc (train-proba): 0.67
```

- Metrics model : Recall. Karena tujuan dari modeling ini kita ingin mendapatkan rasio prediksi data yang positif dari keseluruhan data yang benar positif
- Tidak dilakukan tuning, namun model sudah cukup stabil.
- Tidak terjadi overfitting karena AUC train hampir sama dengan AUC test.

# Model Evaluation and Hyperparameter Tuning

## XGBoost

Algoritma ini digunakan karena robust, prediksi model lebih akurat dapat mengurangi bias/ mencegah underfitting

### Hasil Model Evaluasi

```
Model Evaluation (XGBoost)
Accuracy (Test Set): 0.74
Precision (Test Set): 0.23
Recall (Test Set): 0.52
F1-Score (Test Set): 0.32
roc_auc (test-proba): 0.69
roc_auc (train-proba): 0.82
```

### Setelah Hyperparameter tuning

```
Accuracy (Test Set): 0.74
Precision (Test Set): 0.22
Recall (Test Set): 0.48
F1-Score (Test Set): 0.30
roc_auc (test-proba): 0.68
roc_auc (train-proba): 0.87
```

- Setelah tuning, recall menurun sedikit sebesar 4%.
- Tidak terjadi overfitting karena AUC train dan test masih tergolong seimbang.

# Model Evaluation and Hyperparameter Tuning

## Adaboost

Algoritma ini digunakan karena robust, prediksi model lebih akurat dapat mengurangi bias/ mencegah underfitting

## Hasil Model Evaluasi

```
Model Evaluation (Adaboost)
Accuracy (Test Set): 0.69
Precision (Test Set): 0.21
Recall (Test Set): 0.58
F1-Score (Test Set): 0.31
roc_auc (test-proba): 0.70
roc_auc (train-proba): 0.71
```

- Metrics model : Recall. Karena tujuan dari modeling ini kita ingin mendapatkan rasio prediksi data yang positif dari keseluruhan data yang benar positif
- Setelah tuning, recall menurun sedikit sebesar 4%.
- Tidak terjadi overfitting karena AUC train dan test masih tergolong seimbang.

## Setelah Hyperparameter tuning

```
Accuracy (Test Set): 0.71
Precision (Test Set): 0.22
Recall (Test Set): 0.55
F1-Score (Test Set): 0.31
roc_auc (test-proba): 0.69
roc_auc (train-proba): 0.71
```

# Perbandingan Hasil Akhir Model Evaluation

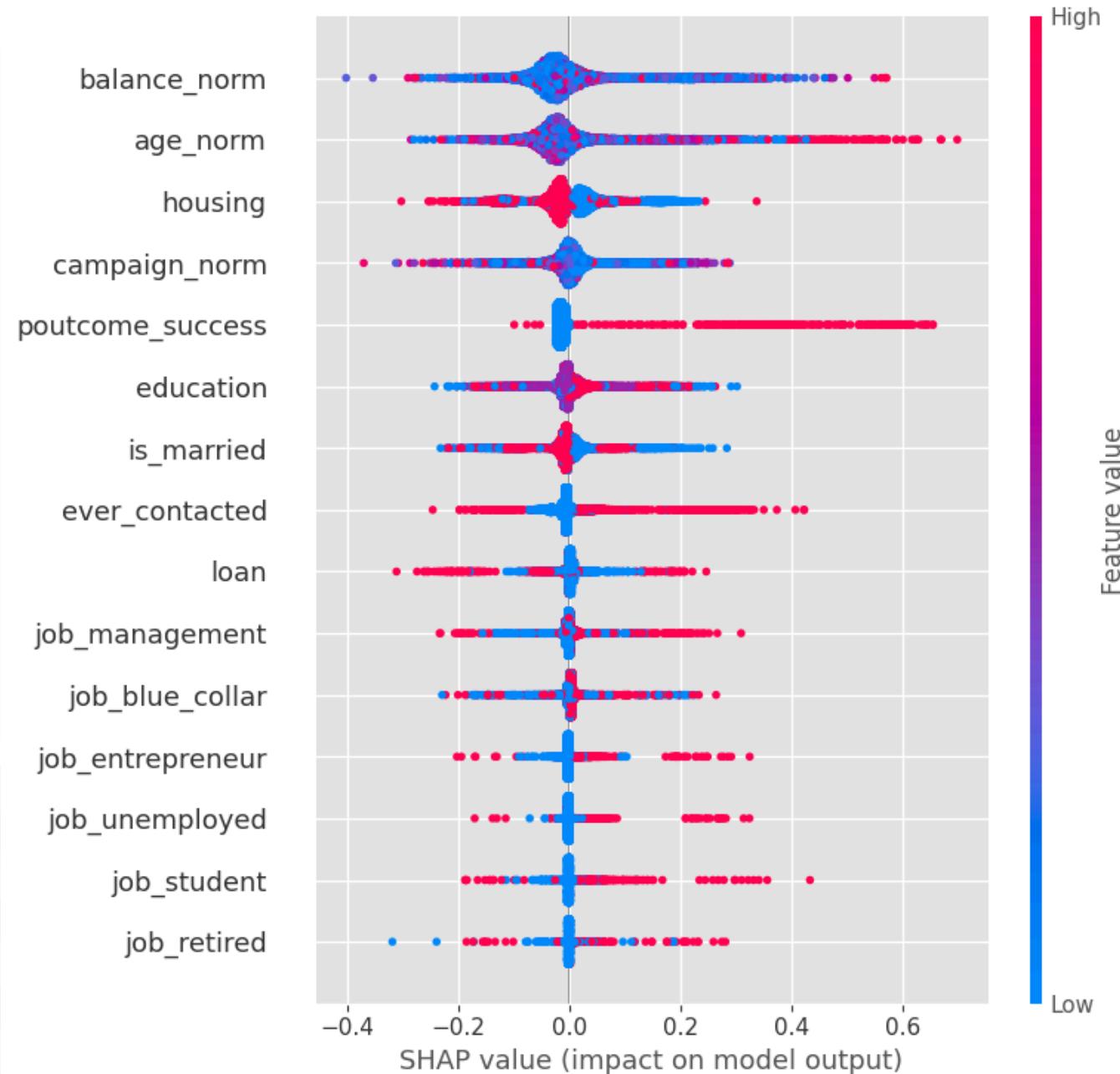
## TANPA PCA

Algoritma	Hyperparameter Tuning	Recall (Tanpa PCA)
XGBoost	Yes	32%
K-Nearest Neighbor	Yes	22%
Decision Tree	Yes	44%
Random Forest	Yes	15%
Adaboost	Yes	13%
Naive Bayes	-	-
Logistic Regression	Yes	14%

## PCA

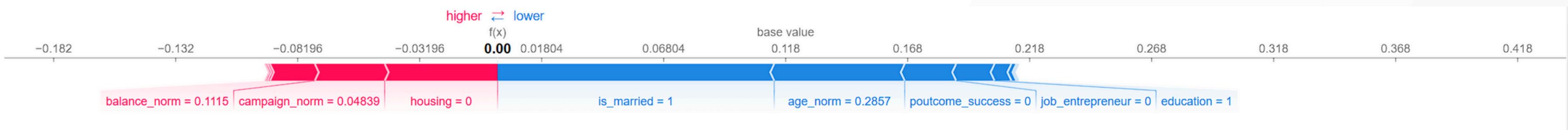
Algoritma	Hyperparameter Tuning	Recall (Dengan PCA)
XGBoost	Yes	48%
K-Nearest Neighbor	Yes	34%
Decision Tree	Yes	47%
Random Forest	Yes	50%
Adaboost	Yes	55%
Naive Bayes	No	47%
Logistic Regression	Yes	63%

# Interpretasi Model - Decision Tree (SHAP) (1 / 2)



1. campaign\_norm semakin kecil, semakin mungkin subscribe
2. age\_norm berpengaruh, namun titik merah berkumpul di dua sudut
3. Balance\_norm memiliki pengaruh terhadap model, namun titik merah tersebar cukup merata. Bisa dilihat di ujung kiri semuanya biru, indikasi balance rendah maka tidak subscribe
4. Housing, titik merah terpusat di sisi kiri. Jika memiliki housing loan, tidak subscribe
5. Poutcome\_success, titik merah terpusat di kanan. Jika pernah take up campaign program sebelumnya, kemungkinan akan take up lagi
6. Is\_married, titik merah berpusat di kiri. Relasi negatif dengan target. Orang yang single lebih mungkin subscribe
7. Education, titik merah berpusat di kanan. Orang dengan tingkat edukasi lebih tinggi, lebih mungkin subscribe
8. Fitur-fitur lain tidak menunjukkan sebaran yang konklusif

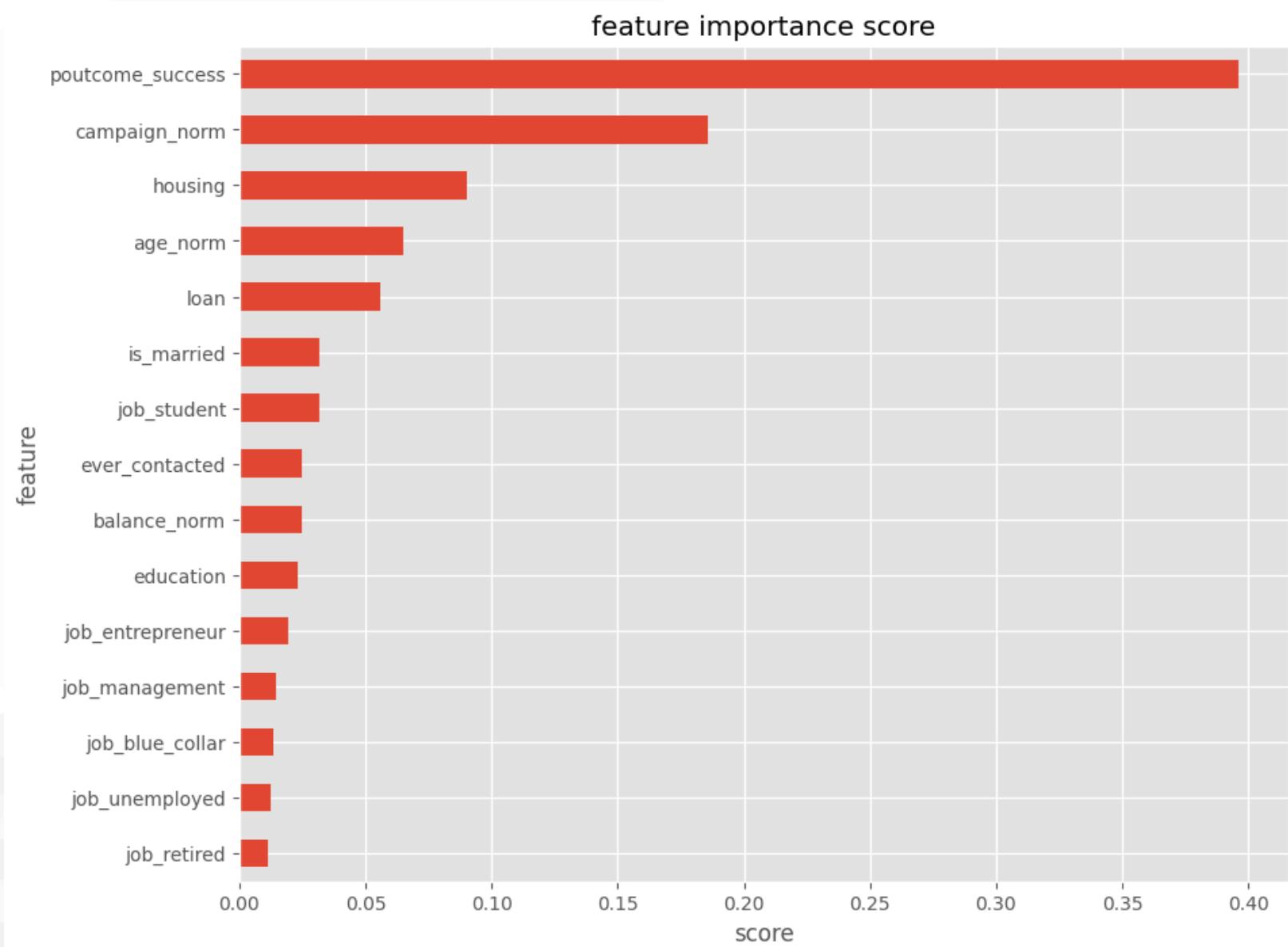
# Interpretasi Model - Decision Tree (SHAP) (2 / 2)



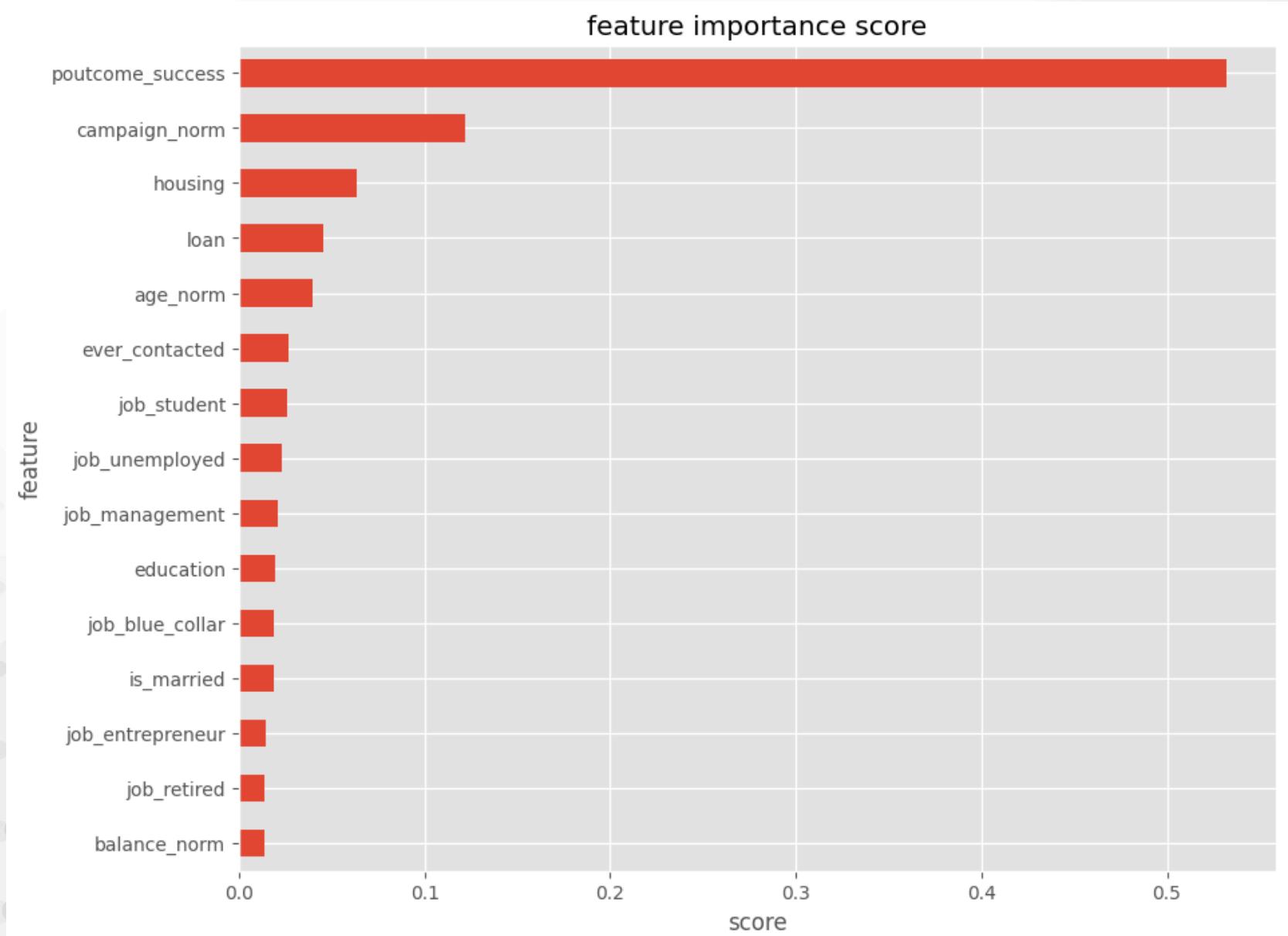
1. `is_married = 1` menjadi faktor negatif terkuat. Status menikah individu ini secara signifikan menurunkan kemungkinan ia untuk berlangganan.
2. `age_norm = 0.2857` juga memberikan pengaruh negatif yang kuat, menunjukkan profil usianya membuat ia cenderung tidak berlangganan.
3. `housing = 0` secara menarik menjadi pendorong negatif. Untuk orang ini, tidak memiliki pinjaman rumah justru membuatnya lebih kecil kemungkinannya untuk berlangganan.
4. `poutcome_success = 0` merupakan faktor pendorong positif. Artinya, meskipun kampanye sebelumnya tidak berhasil, model melihat ini sebagai sinyal positif untuk individu spesifik ini, yang mungkin berlawanan dengan tren umum.
5. Faktor negatif (`is_married`, `age`) jauh lebih dominan daripada faktor positif (`poutcome_success`), sehingga hasil prediksi akhir sangat rendah (0.0104).

# Interpretasi Model - XGBoost (Feature Importance) (1 / 2)

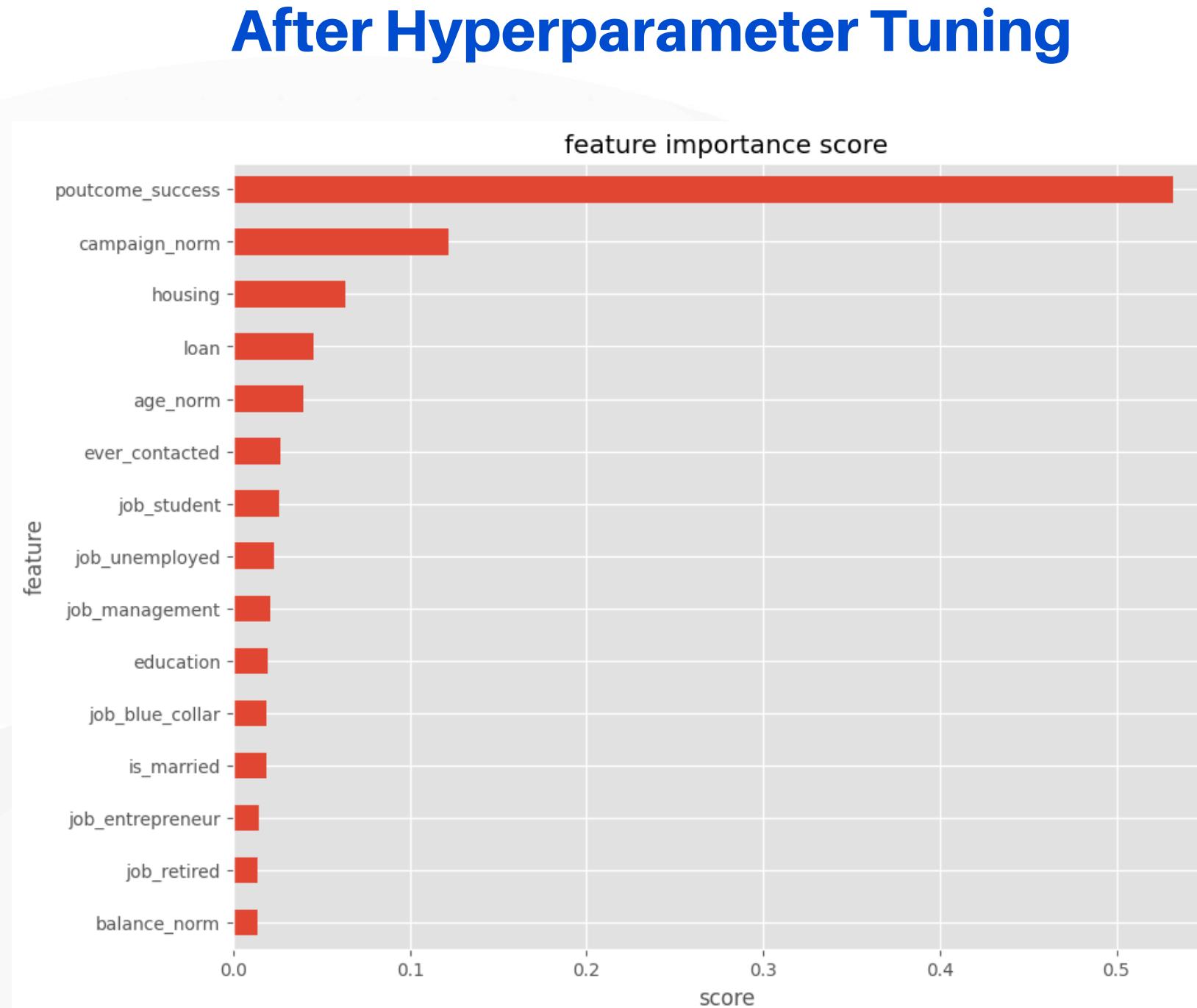
## Before Hyperparameter Tuning



## After Hyperparameter Tuning



# Interpretasi Model - XGBoost (Feature Importance) (2 / 2)



Setelah Hyperparameter Tuning, XGBoost menghasilkan recall sebesar 0.28, sedikit menurun dari model awal (0.33), namun model menunjukkan indikasi overfitting (AUC train jauh lebih tinggi dari test).

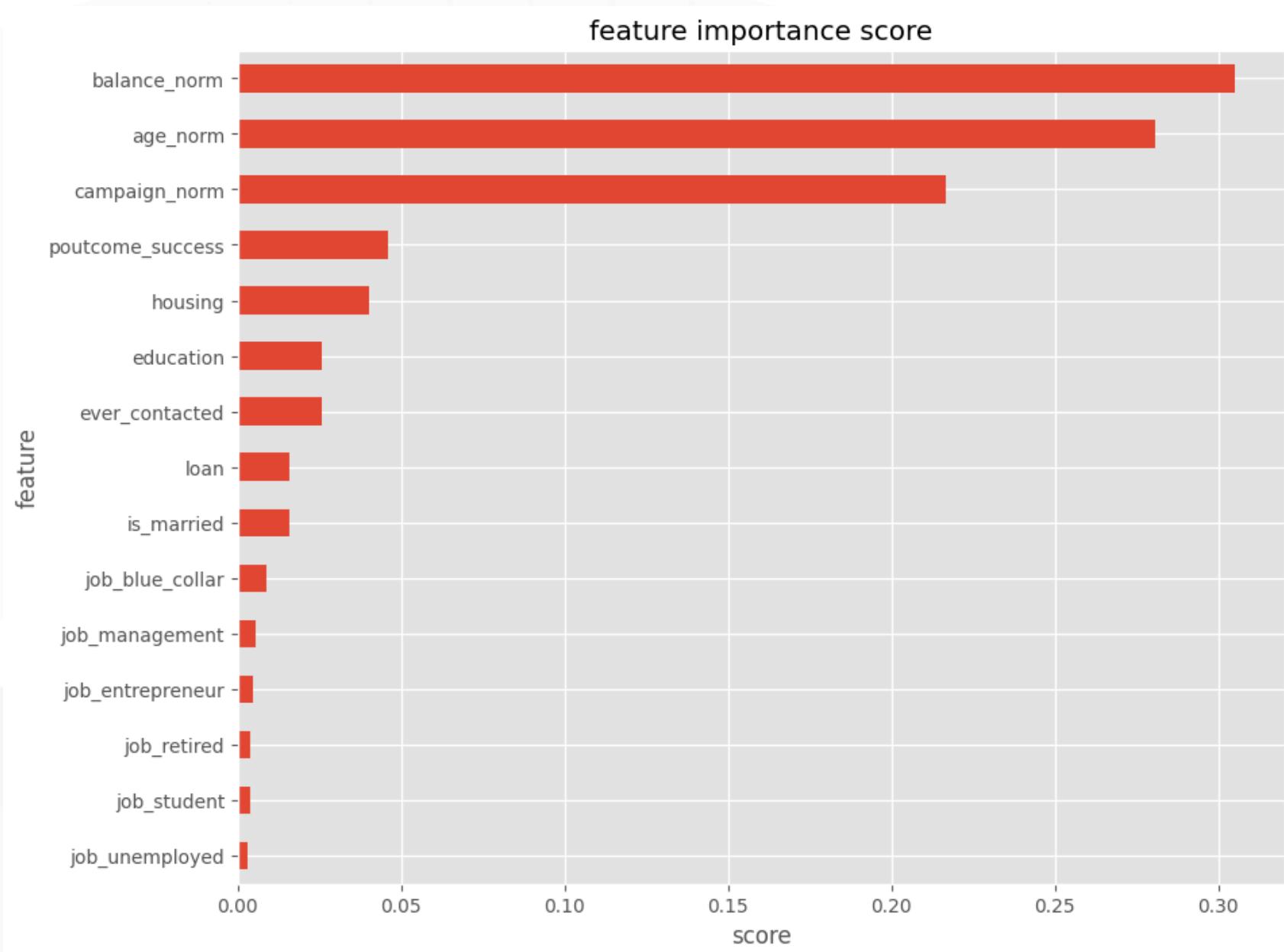
Jika kita lihat chart Feature Importance XGBoost:

- poutcome\_success menjadi fitur yang sangat dominan, dengan skor di atas 0.5.
- Fitur-fitur lain seperti campaign\_norm, housing, dan loan hanya memberi kontribusi kecil.
- Fitur seperti education, age\_norm, dan sebagian besar kategori job hampir diabaikan oleh model.

Dari hasil pengamatan ini, model sangat bergantung pada satu fitur – poutcome\_success, yang value 'yes'-nya hanya mencakup sekitar 3% dari data asli. Ketergantungan ini membuat model rapuh dan kurang general, meski nilai recall terlihat cukup baik di atas kertas.

# Interpretasi Model - Random Forest (Feature Importance)

## Before Hyperparameter Tuning



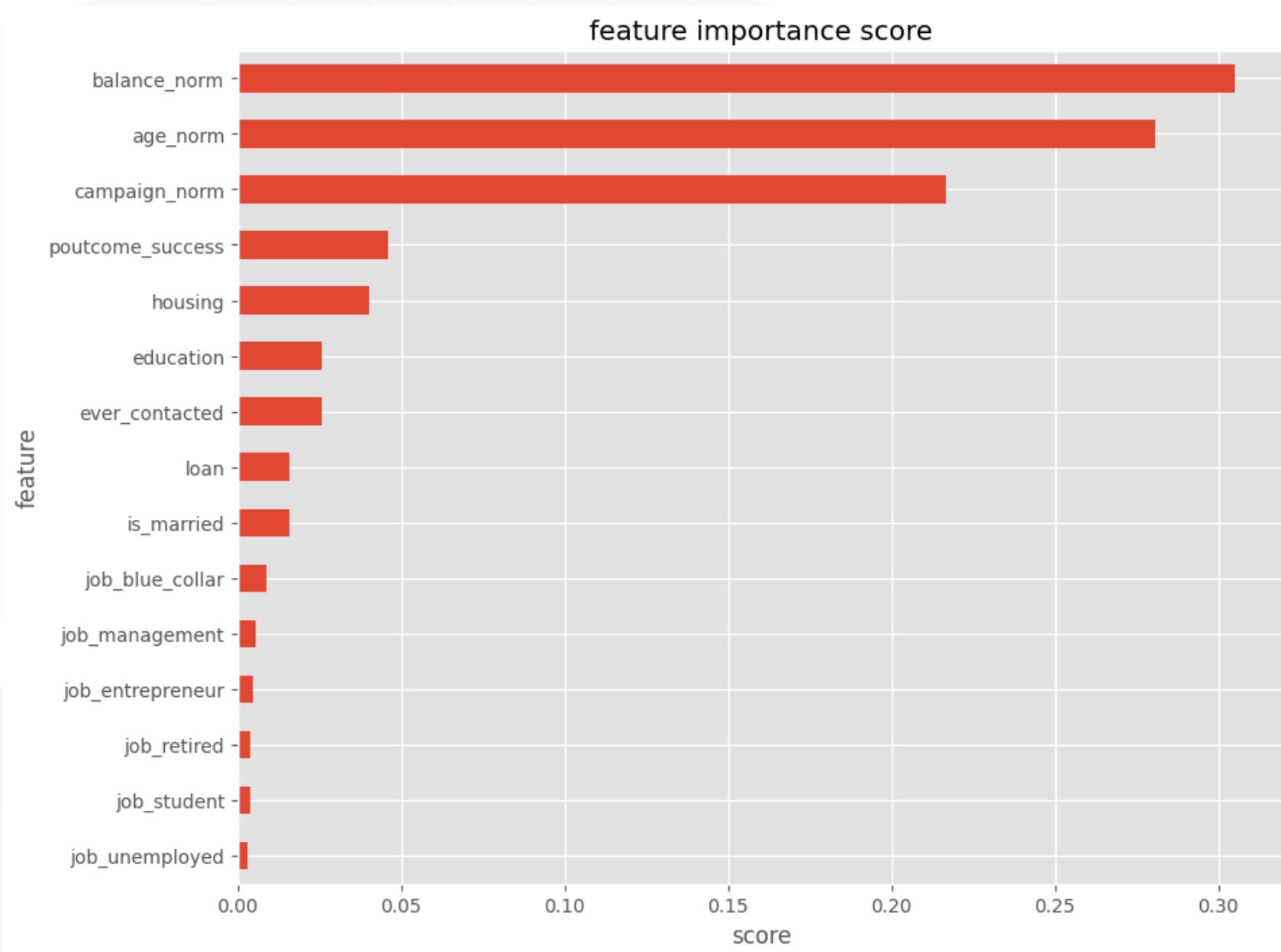
Model Random Forest yang dipilih adalah yang belum mengalami Hyperparameter Tuning. Karena setelah dilakukan tuning, overfitting berkurang, namun nilai recall juga turun.

Jika kita lihat chart Feature Importance Random Forest before Hyperparameter Tuning, top 5 fitur dengan pengaruh paling besar: balance\_norm, age\_norm, campaign\_norm, poutcome\_success, housing.

Karena sama-sama tree classifier, interpretasi interaksi masing-masing fitur dengan target dapat mengikuti interpretasi pada decision tree. Top 5 fitur yang berperan penting terhadap target bersifat masuk akal.

# Interpretasi Model - Random Forest (Feature Importance)

## Before Hyperparameter Tuning



Model Random Forest yang dipilih adalah yang belum mengalami Hyperparameter Tuning. Karena setelah dilakukan tuning, overfitting berkurang, namun nilai recall juga turun.

Jika kita lihat chart Feature Importance Random Forest before Hyperparameter Tuning, top 5 fitur dengan pengaruh paling besar: balance\_norm, age\_norm, campaign\_norm, poutcome\_success, housing.

Karena sama-sama tree classifier, interpretasi interaksi masing-masing fitur dengan target dapat mengikuti interpretasi pada decision tree. Top 5 fitur yang berperan penting terhadap target bersifat masuk akal.

# Kesimpulan

Dalam proyek ini, berbagai model klasifikasi digunakan untuk memprediksi keberhasilan kampanye pemasaran dengan fokus pada recall, yaitu kemampuan model dalam mengenali nasabah yang tertarik. Hasil menunjukkan bahwa XGBoost memberikan nilai recall tertinggi (0.74) setelah tuning, tetapi model ini terlalu bergantung pada satu fitur utama, yaitu poutcome\_success, yang sebenarnya hanya dimiliki oleh sebagian kecil data. Hal ini bisa berisiko karena model menjadi kurang fleksibel saat menghadapi data baru. Sementara itu, model Random Forest tanpa tuning menunjukkan performa yang cukup baik dengan recall 0.71, dan pembagian kontribusi fitur yang lebih merata serta hasil yang stabil tanpa overfitting.

Dari semua model yang dicoba, penggunaan PCA (reduksi dimensi) justru menurunkan performa model secara umum. Model dengan PCA cenderung memiliki recall yang lebih rendah dan menunjukkan gejala overfitting. Oleh karena itu, model Random Forest tanpa PCA dan tanpa hyperparameter tuning dinilai sebagai pilihan paling seimbang. Meskipun tidak memiliki recall tertinggi, model ini cukup akurat, tidak overfitting, dan mudah dipahami. Ini membuatnya cocok untuk digunakan dalam kasus nyata yang membutuhkan prediksi yang andal dan stabil.

# Thank You

**Yohanes Gabriel Valentino Manurung**  
**+62 812-9912-2397**  
**yohanesgabriels18@gmail.com**