

FINAL PROJECT

PRESENTATION

Predictive Pioneers

Meet Our Team



Muhammad Irfan Karim

Project Leader



Ariel Nathania

Member



Veby Pebiola Br Simanjorang

Member



Setya Dewa Nugraha

Member



Deyyana Aulia Hakim

Member



Tiya Novita Sari

Member



Anita Putri

Member



Robi Muhammad Subhan

Member

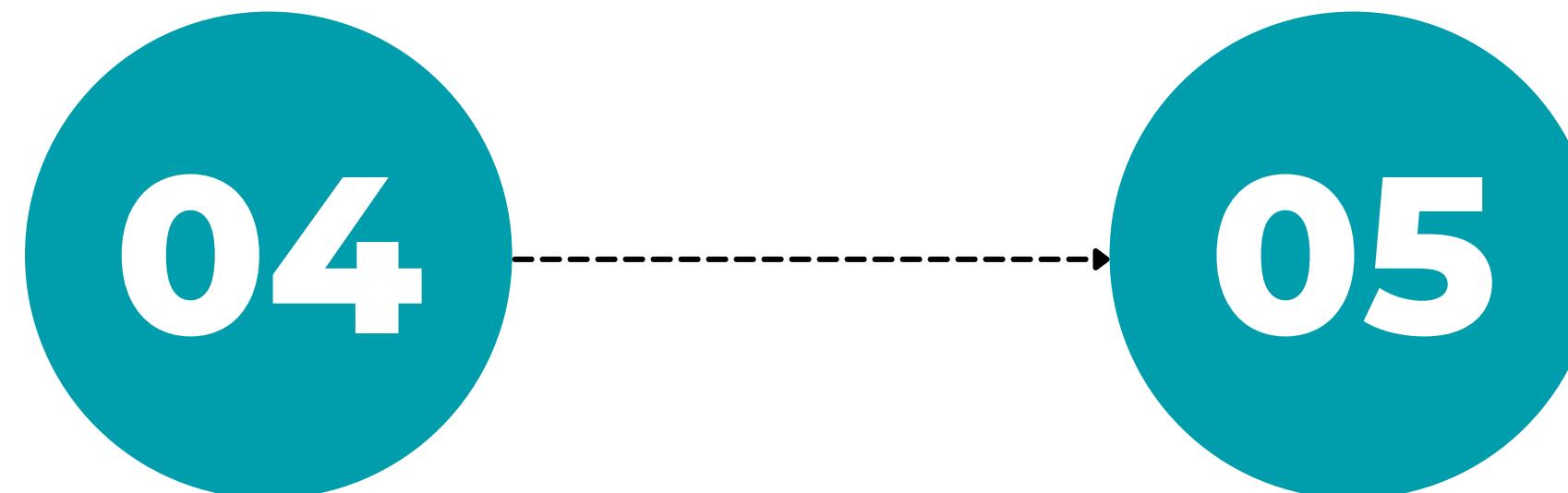
WORKFLOW



Background

EDA

Pre-Processing



Modelling

Recommendation

BACKGROUND

Profil Perusahaan

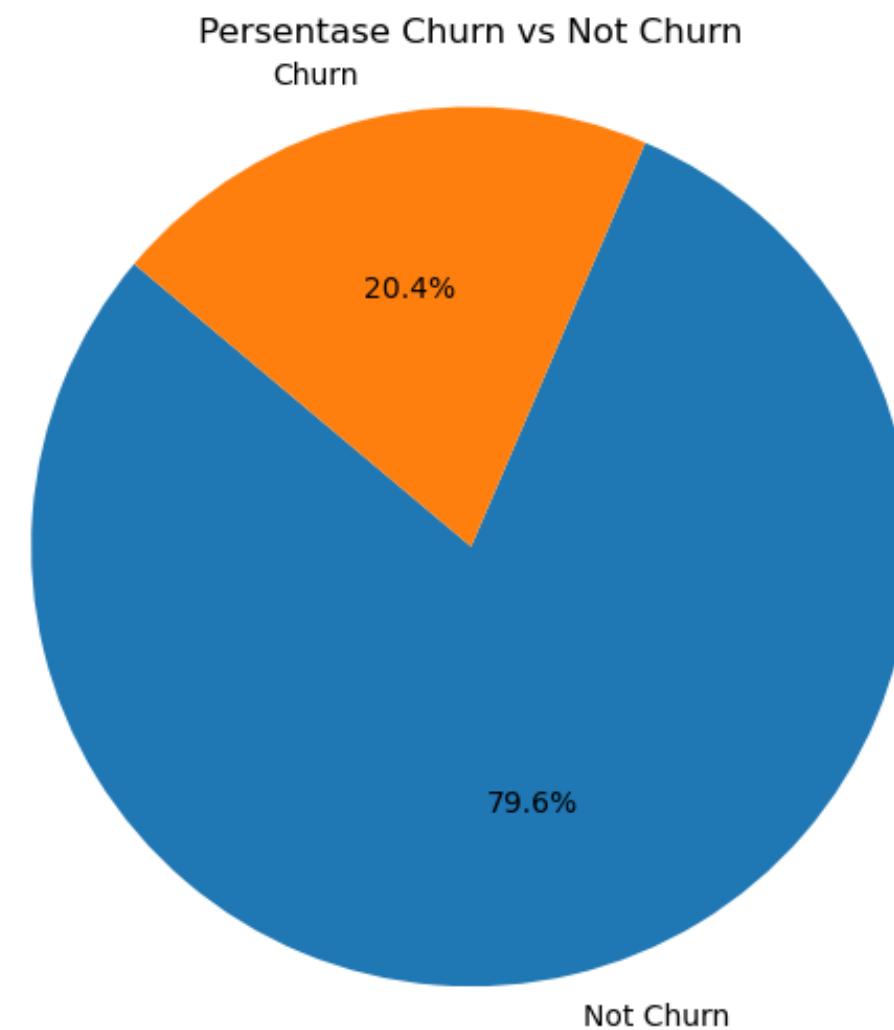
Bank Manduro

Bank Manduro, sebuah perusahaan layanan keuangan yang menawarkan berbagai layanan perbankan, sedang mengalami kerugian akibat tingkat churn yang kunjung naik hingga **20,4%** pada bulan November. Mereka berencana menerapkan teknik-teknik machine learning untuk menemukan solusi mengatasi Nasabah yang churn dari data yang mereka miliki.



Problem & Impact

Bank Manduro, menghadapi tantangan signifikan karena beberapa nasabah yang **berhenti menggunakan layanan mereka sebesar 20,4% pada bulan November**. Dampaknya akan terasa pada **pendapatan bank menurun** yang berasal dari biaya layanan, bunga pinjaman, serta berbagai produk keuangan yang dimanfaatkan oleh nasabah.



What Can We Do?



Goal

Mereduksi tingkat churn pada Bank Manduro sebesar 3,4% dalam waktu 1 bulan



Objective

- **Mencari faktor** yang mempengaruhi tingkat churn nasabah.
- **Membuat model machine learning** yang dapat memprediksi nasabah yang memiliki kecenderungan churn



Business Metrics

Churn Rate

EDA

Describe The Dataset

Data columns (total 14 columns):

#	Column	Non-Null Count	Dtype
0	RowNumber	10000	non-null int64
1	CustomerId	10000	non-null int64
2	Surname	10000	non-null object
3	CreditScore	10000	non-null int64
4	Geography	10000	non-null object
5	Gender	10000	non-null object
6	Age	10000	non-null int64
7	Tenure	10000	non-null int64
8	Balance	10000	non-null float64
9	NumOfProducts	10000	non-null int64
10	HasCrCard	10000	non-null int64
11	IsActiveMember	10000	non-null int64
12	EstimatedSalary	10000	non-null float64
13	Exited	10000	non-null object

dtypes: float64(2), int64(8), object(4)

memory usage: 1.1+ MB

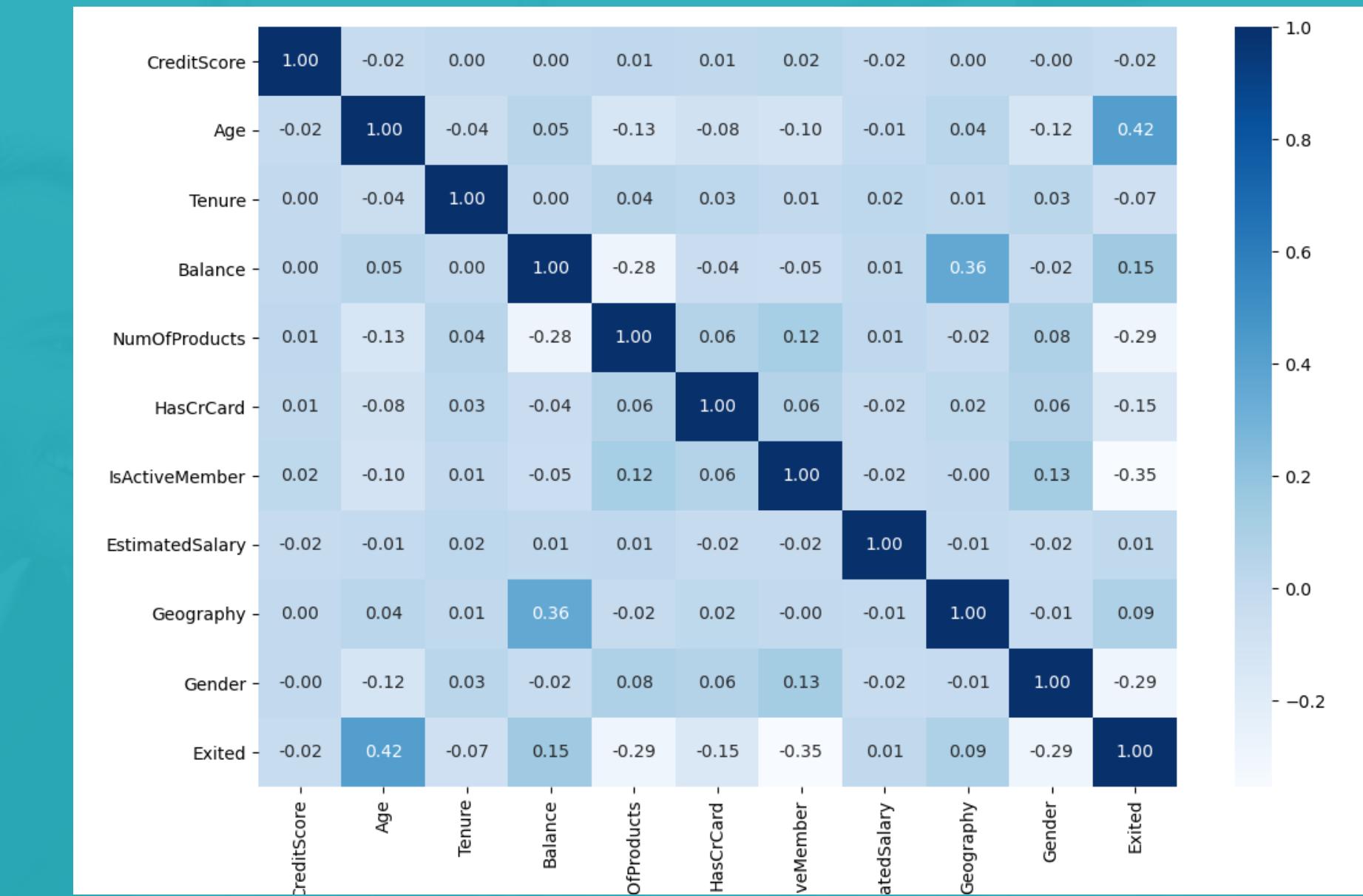


Ada 7 data numerik dan 6 data kategorik termasuk target



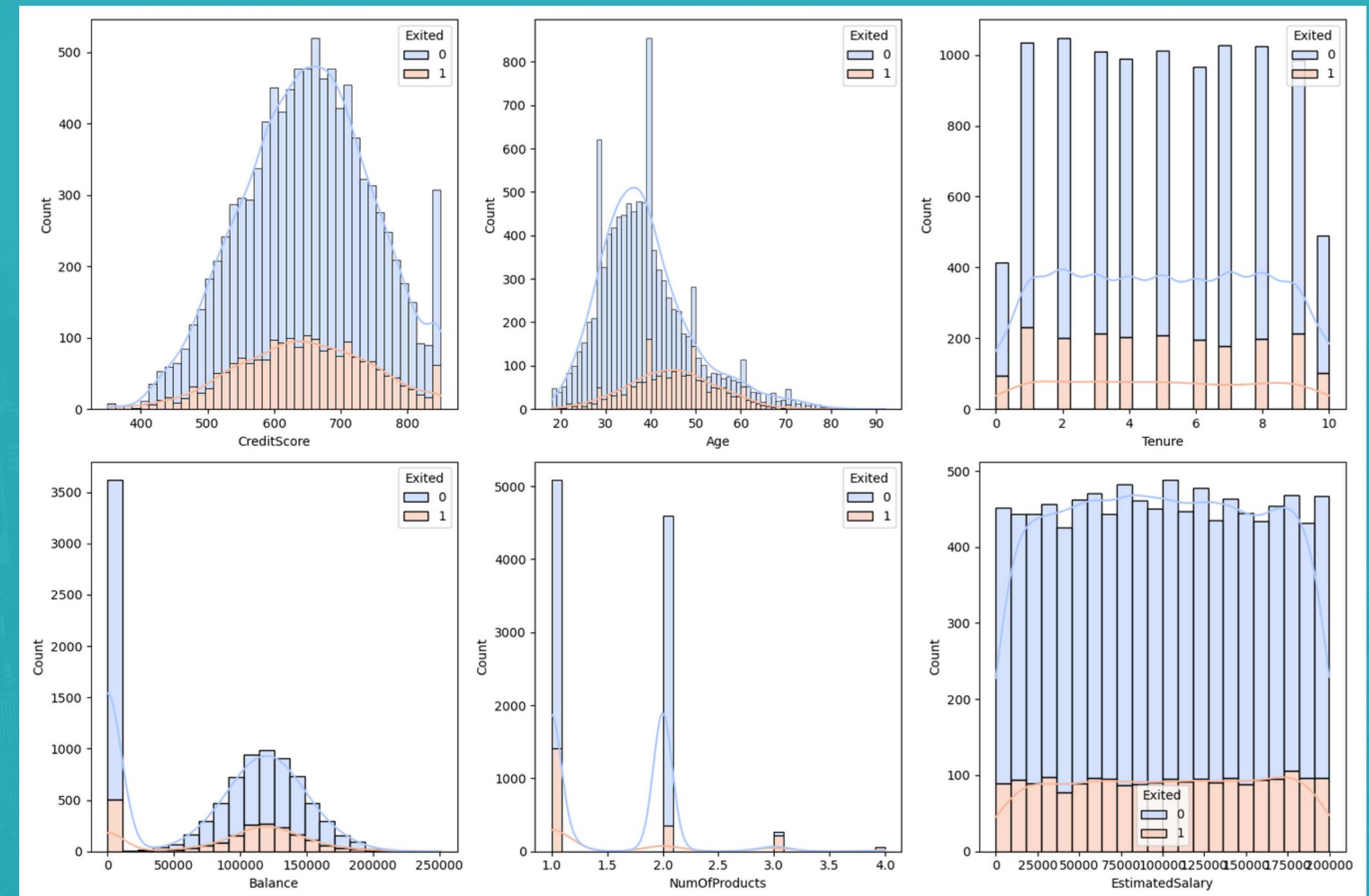
Data terdiri dari 13 kolom dan 10.000 baris tanpa nilai kosong dan duplikat.

See The Correlation



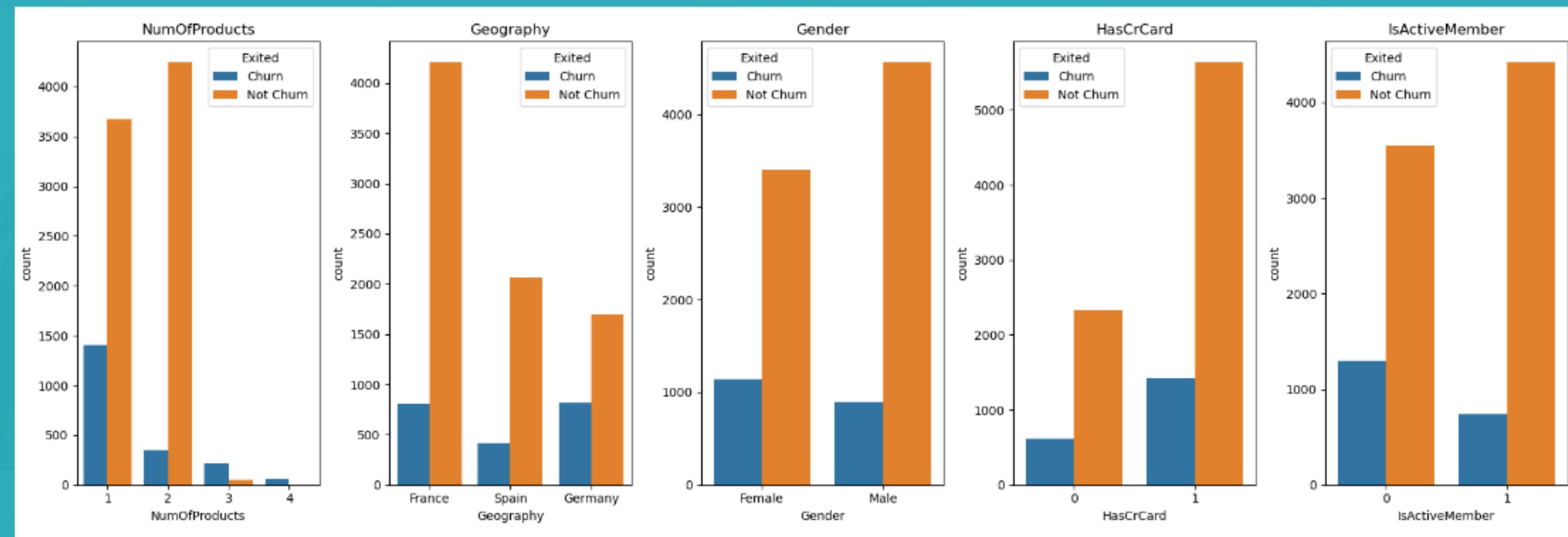
- Fitur **Age** dan **Balance** memiliki korelasi positif terhadap label target
- Fitur **NumOfProducts**, **HasCrcard**, **IsActiveMember**, dan **Gender** memiliki korelasi negatif yang kuat terhadap label target

GG



- Dalam kategori CreditScore, nasabah dengan rentang **score 580-699** memiliki kecenderungan **CHURN** yang **besar**.
- Dalam kategori Age, nasabah yang cenderung **CHURN** rata-rata berusia **40-49** tahun
- Dalam kategori Balance, nasabah yang cenderung **CHURN** memiliki saldo sebanyak **0 (tidak memiliki saldo)**





- Dalam kategori NumOfProducts, nasabah dengan 1 produk memiliki kecenderungan **CHURN** yang **besar**.
- Dalam kategori Geography, nasabah yang cenderung **CHURN** bertempat tinggal di **Germany**
- Dalam kategori Gender, nasabah yang cenderung **CHURN** ialah **Perempuan**
- Dalam kategori HasCrCard, nasabah yang cenderung **CHURN** ialah nasabah yang mempunyai **Credit Card**
- Dalam kategori IsActiveMember, nasabah yang cenderung **CHURN** ialah member yang **tidak aktif**

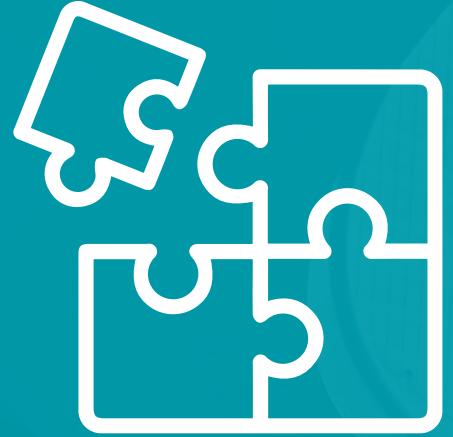
RECAP



Attribute nasabah yang cenderung banyak melakukan **Churn** adalah sebagai berikut:

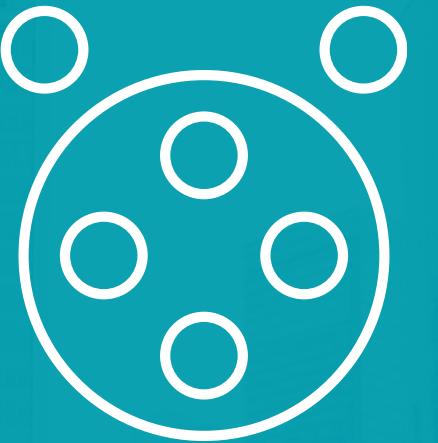
- **Balance** : 0
- **Age** : 40-49
- **Member** : Not Active Member
- **Credit Card** : has credit card
- **Credit score** : 580-699
- **Gender** : Female
- **Geography** : Germany

Pre-Processing



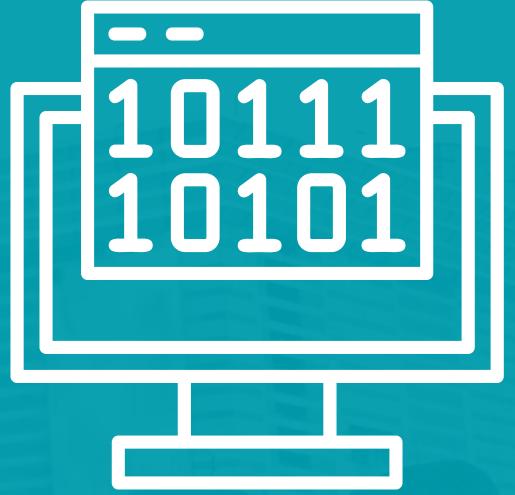
MISSING VALUE AND DUPLICATE DATA

Tidak ada nilai yang hilang dan duplikat pada data.



OUTLIERS

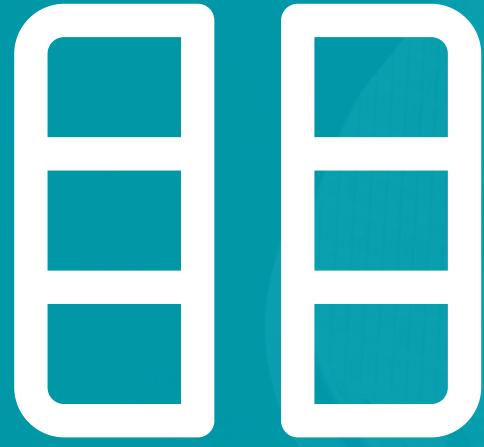
Terdapat outlier pada 3 fitur, yaitu **Age**, **CreditScore** dan **NumOfProducts** tetapi kami putuskan untuk tidak menghapus outliernya dikarenakan merepresentasikan real life case.



FEATURE ENCODING

One Hot Encoding
Gender & Geography

Pre-Processing



Train Test Split

Train (80 %) : Test (20%)



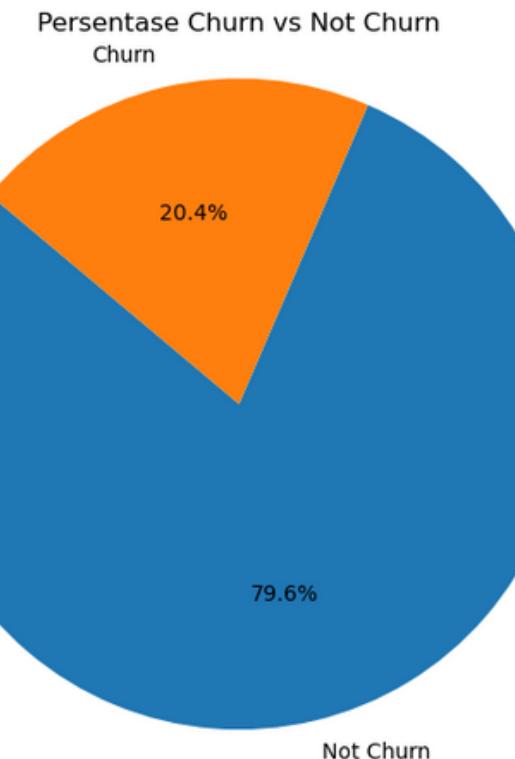
FEATURE SCALING

Normalization with MinMaxScaler

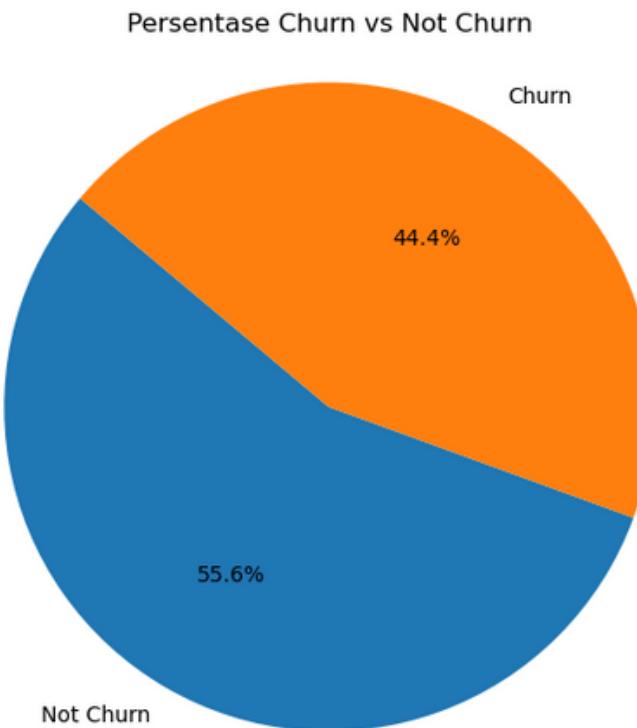


HANDLING IMBALANCE CLASS

Oversampling with SMOTE



Before & After Imbalancing



Modeling

Model yang dipakai untuk melakukan klasifikasi yaitu RandomForest

WHY RANDOMFOREST?



Baik dalam menangani
kasus ketidakseimbangan
class



Performa yang tinggi



Pencegahan Overfitting



Modeling

LANGKAH-LANGKAH:

1

Menentukan metrics yang tepat:

Pada model yang kami buat memakai metrics **F1-SCORE**

2

Mencari hyperparameter yang optimal untuk model
kemudian melakukan **Tuning Hyperparameter**

Random Forest	
F1-Score (Train)	0.857946
F1-Score (Test)	0.799675
F1-Score (Crossval Train)	0.853433
F1-Score (Crossval Test)	0.795851

Rentang antara train dan test tidak jauh yang
artinya model **BEST FIT**

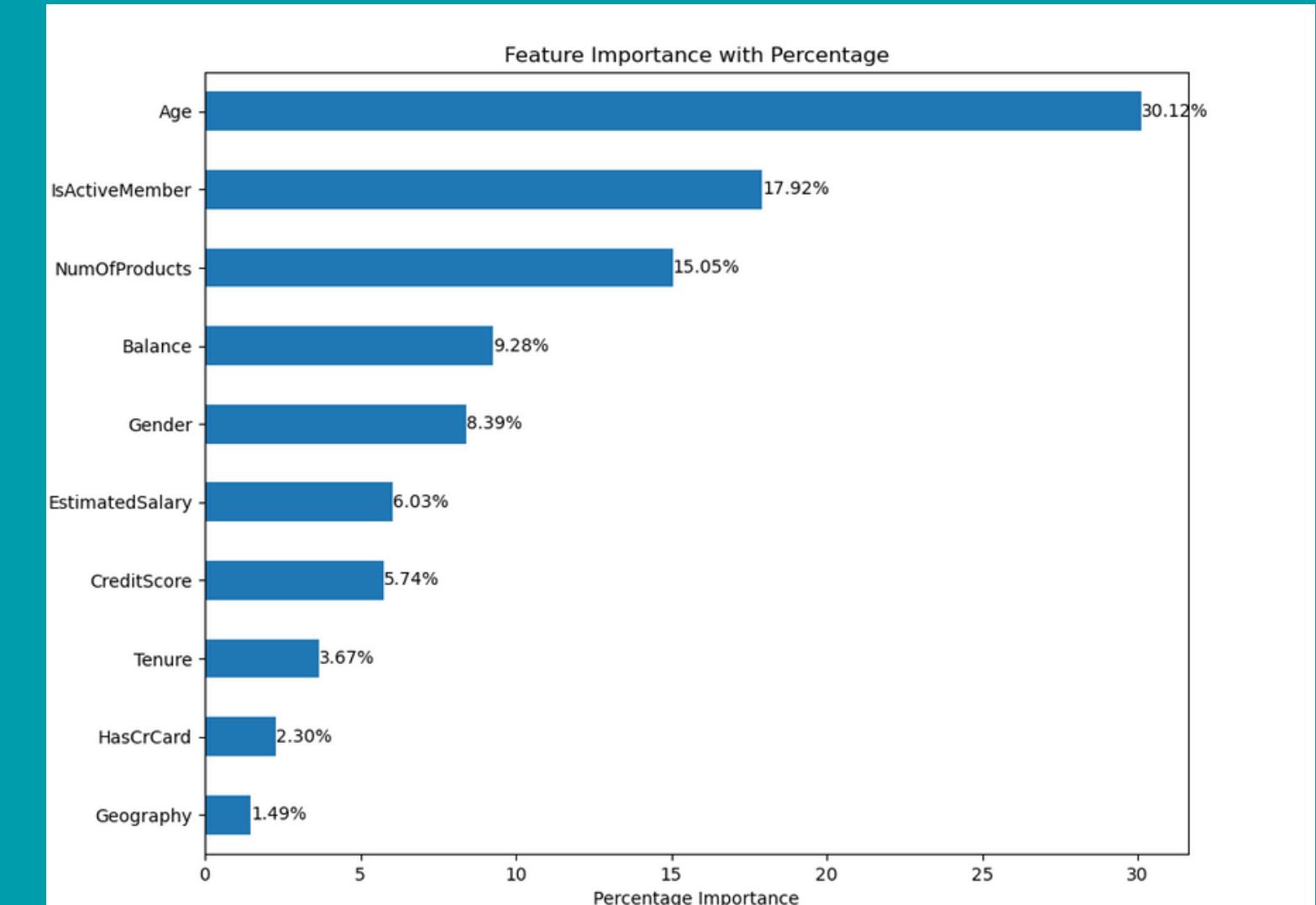
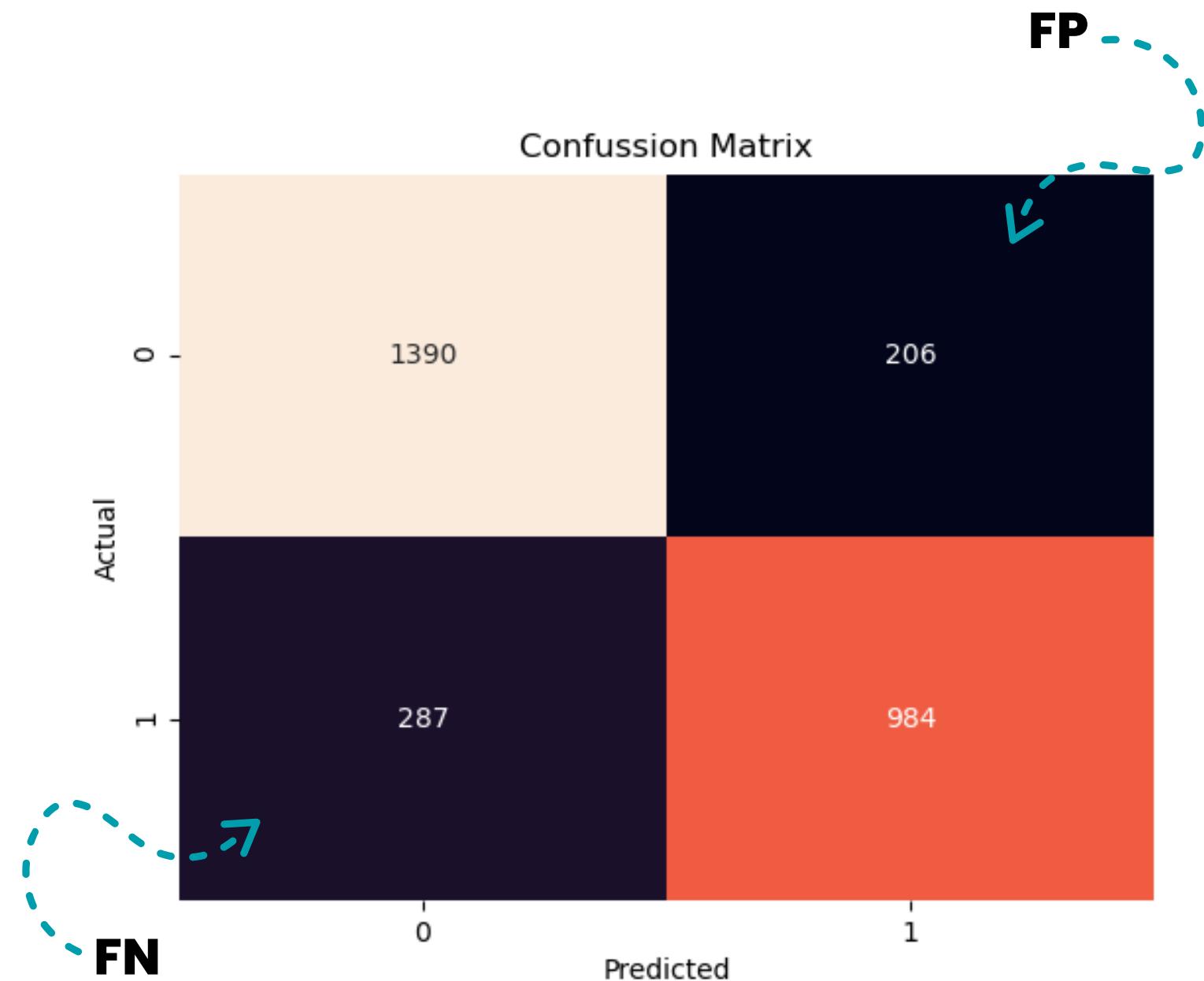
Why F1-SCORE?

Kami memilih metrics evaluation **F1-SCORE** karena
metrics ini yang paling sejalan dengan business
metrics kami yaitu menurunkan jumlah **churn
nasabah**.

Penggunaan **F1-SCORE** kami harapkan dapat
menghasilkan model dengan **FN dan FP** sekecil
mungkin sehingga model ini dapat bekerja dengan
maksimal untuk menjadi solusi bagi business
metrics kami.



Model Results



- Karena menggunakan **F1-Score** maka **FP** dan **FN** menjadi prioritas untuk dikurangi

- Fitur **Age** menjadi fitur terpenting dalam model **RandomForest**.

CHURN RATE

BEFORE MODEL

20,4%



-3,4%

AFTER MODEL

17,0%

Insight & Recommendation

Insight from model

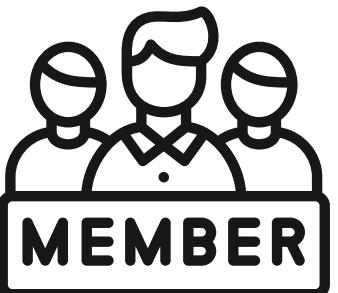


Usia berperan signifikan dalam model **RandomForest**, sehingga dapat menjadi faktor penting dalam memprediksi churn.

Dari modelling diperoleh bahwa Nasabah dalam jangkauan umur **40-49 memiliki potensi churn lebih tinggi** dibandingkan dengan umur nasabah lainnya



Recommendation



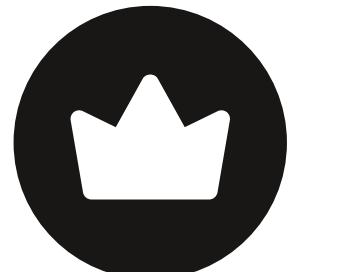
Strategi Retensi

Fokus pada strategi retensi untuk **nasabah berusia 40-49 tahun**, karena mereka memiliki kecenderungan untuk churn. Bank dapat **menawarkan insentif khusus, layanan tambahan, atau program loyalitas** untuk mempertahankan mereka.



Pengoptimalan Strategi Pemasaran

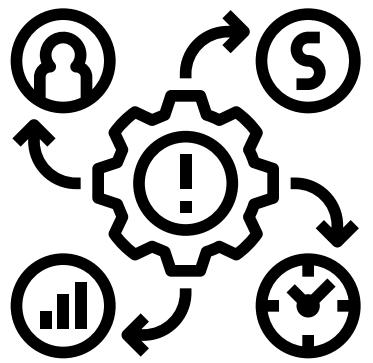
Bank perlu mengoptimalkan strategi pemasaran dan layanan yang lebih spesifik di **Germany** seperti melakukan **kampanye pemasaran** dan **memberikan penawaran menarik** untuk daerah tersebut



Peningkatan Keterlibatan Nasabah

Dorong keaktifan nasabah dengan memberikan **dorongan atau insentif** bagi nasabah yang **kurang aktif dan hanya memiliki 1 produk**, seperti **diskon untuk transaksi rutin** atau **program loyalitas**.

Impacts for Bank Manduro



1. Mampu membantu **menurunkan tingkat churn rate** pada bulan November sebesar 3,4%.
2. Mampu memudahkan Bank Manduro untuk mencari **faktor terbesar penyebab tingginya churn rate** pada Bank Manduro
3. Membantu **mengurangi kehilangan Nasabah, kepercayaan dan hilangnya pendapatan** pada Bank Manduro





TERIMA KASIH