



Churn Prediction

Bank CBA

Rakamin Academy
Studi Independen Data Science
Batch 05

READ MORE 

Meet Our Member Team

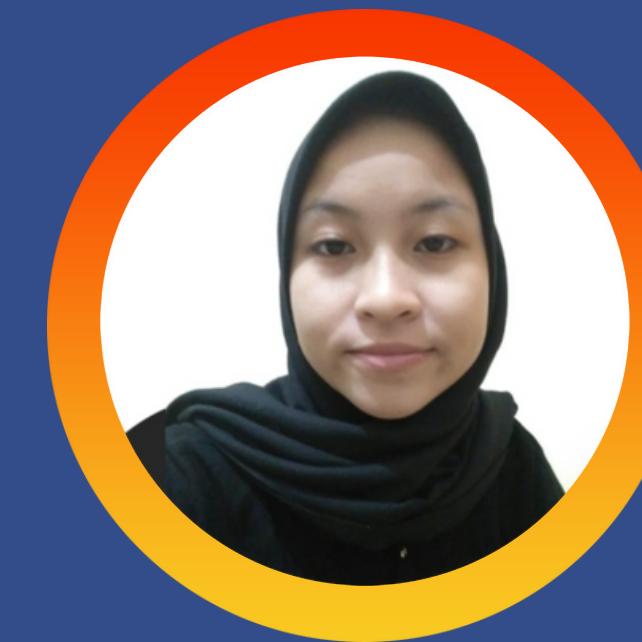
DATA GODDESSES



ALEISYA ZAHARI SALAM



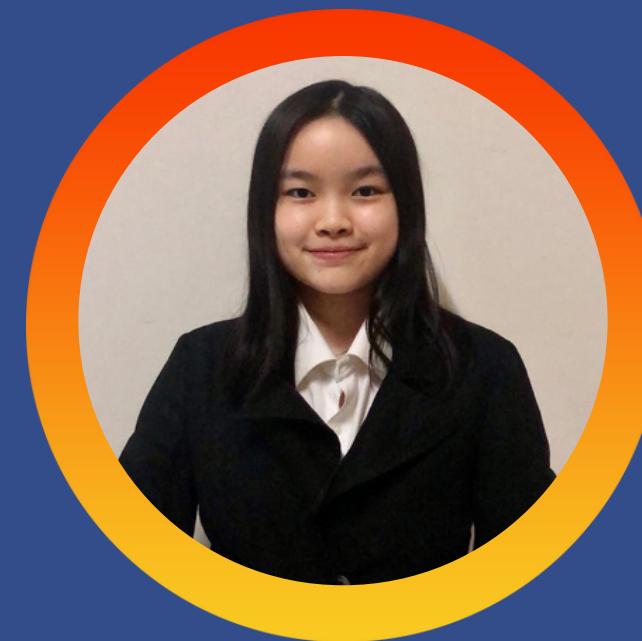
DEVI NUR AISYAH



FIORENT ARIE
HERNANTI



TISA SAFINA
ALCHALISTA



ALICIA GOFINA



DEVA KHOFIFAH
JAUHAROTUN NAQIYYAH



NAJMI LAILY FAHIRA

Outline

DATA GODDESSES



➤ BUSINESS UNDERSTANDING

➤ EXPLORATORY DATA ANALYSIS (EDA)

➤ DATA PREPROCESSING

➤ BUSINESS RECOMMENDATION

➤ BUSINESS SIMULATION

COMPANY BACKGROUND

Bank CBA adalah perusahaan yang bergerak dalam bidang keuangan, menyediakan banyak layanan keuangan kepada customer/nasabahnya. Saat ini, perusahaan ini tertarik untuk memahami faktor-faktor yang mendorong pelanggan untuk memutuskan hubungan dengan bank (churn)





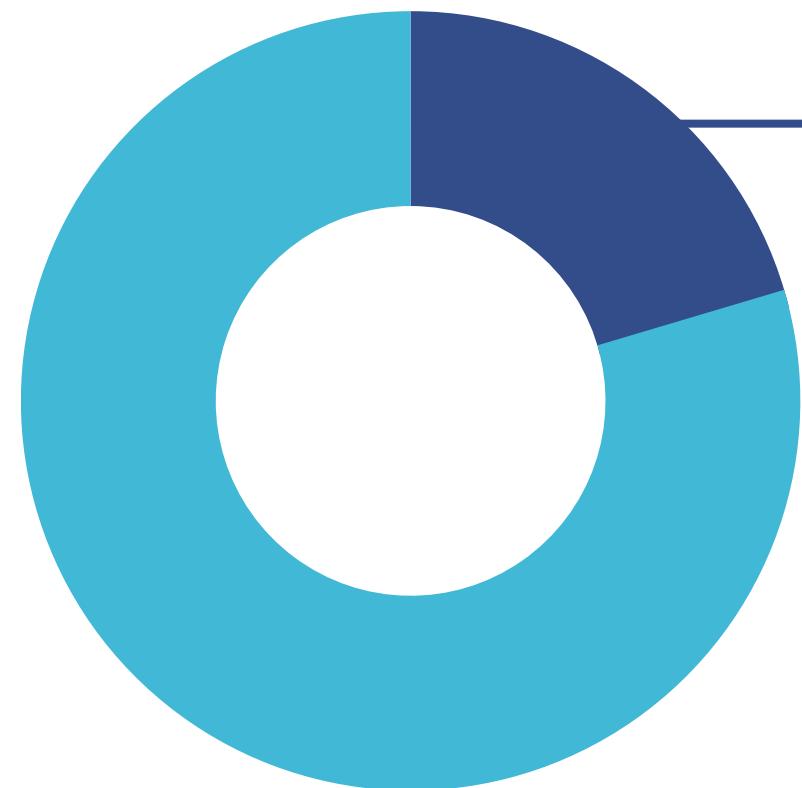
BUSINESS UNDERSTANDING

● PROBLEM STATEMENT

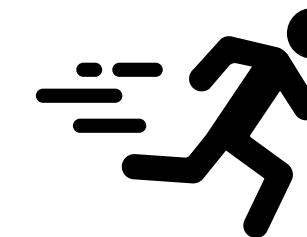
● GOALS, OBJECTIVE & BUSINESS METRICS

Problem Statement

1. Customer churn



20.4% dari
10.000 costumer
memilih untuk
menghentikan
langganan

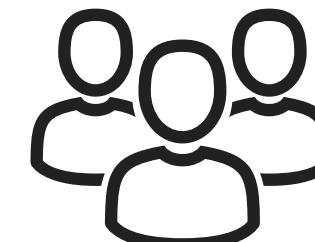
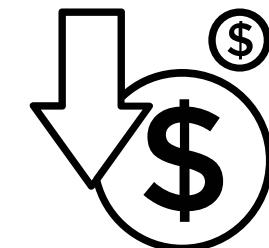


Churn rate yang dapat
diterima yaitu **5% - 7%**

Based : Lincoln Murphy SixteenVentures

Dampak?

**Revenue dan
Profit menurun**



**Keaktifan costumer
berkurang**

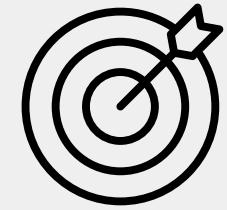


**Citra perusahaan
buruk**

Goals, Objective & Business Metrics

DATA GODDESSES

Goals



Mengurangi
tingkat churn
(< 7 %)

Objective



1. Melakukan deep analytics (EDA)
2. Membangun model machine learning klasifikasi (churn atau tidak)

Business Metris



Churn rate



EXPLORATORY DATA ANALYSIS

● DATASET INFORMATION

● PRELIMINARY INSIGHT



Dataset Information

Customer Information

RowNumber
CustomerId
Surname

Exited or Not (Target Variable)

Exited

Customer Attributes

CreditScore
Geography
Gender
Age
Tenure
Balance
NumOfProducts
HasCrCard
IsActiveMember
EstimatedSalary

Dataset memiliki
14 kolom dan
10.000 baris

Tidak ditemukan
nilai Null maupun
data duplikat

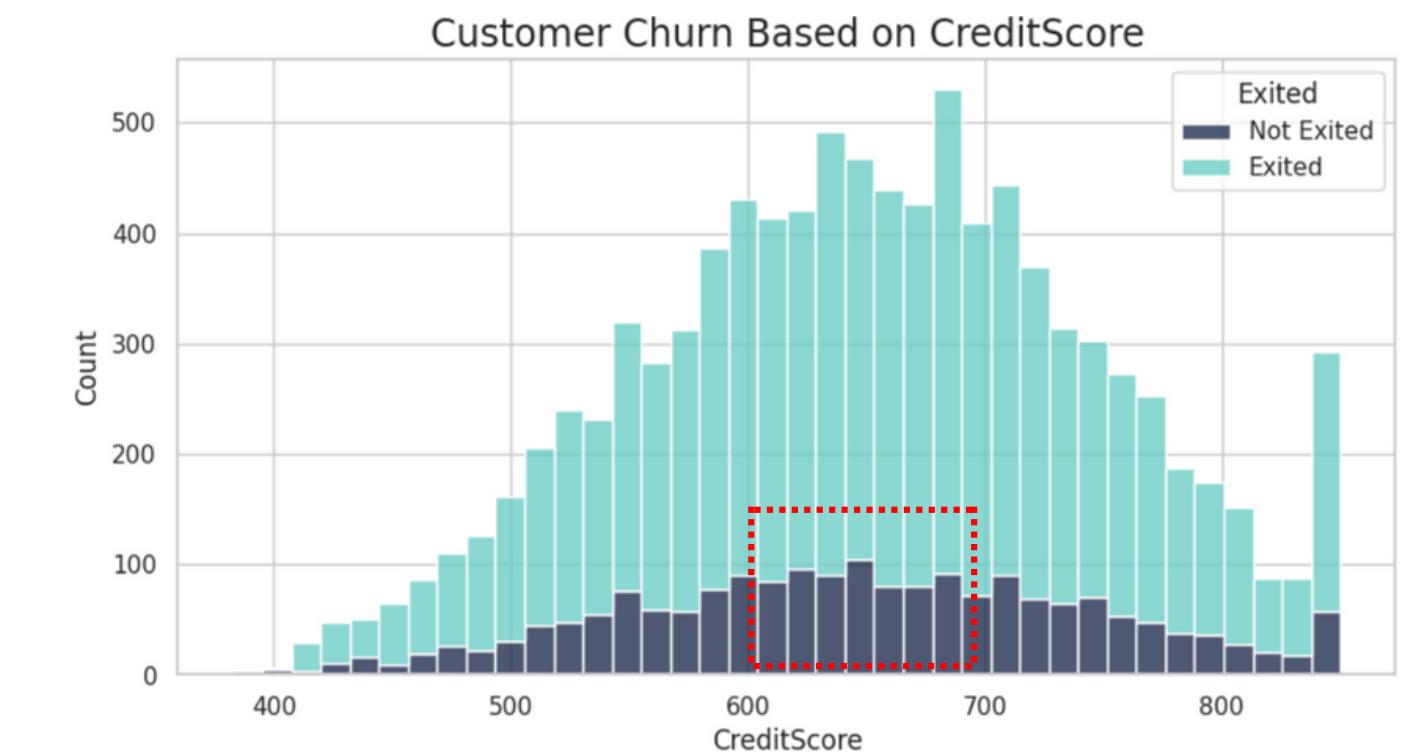
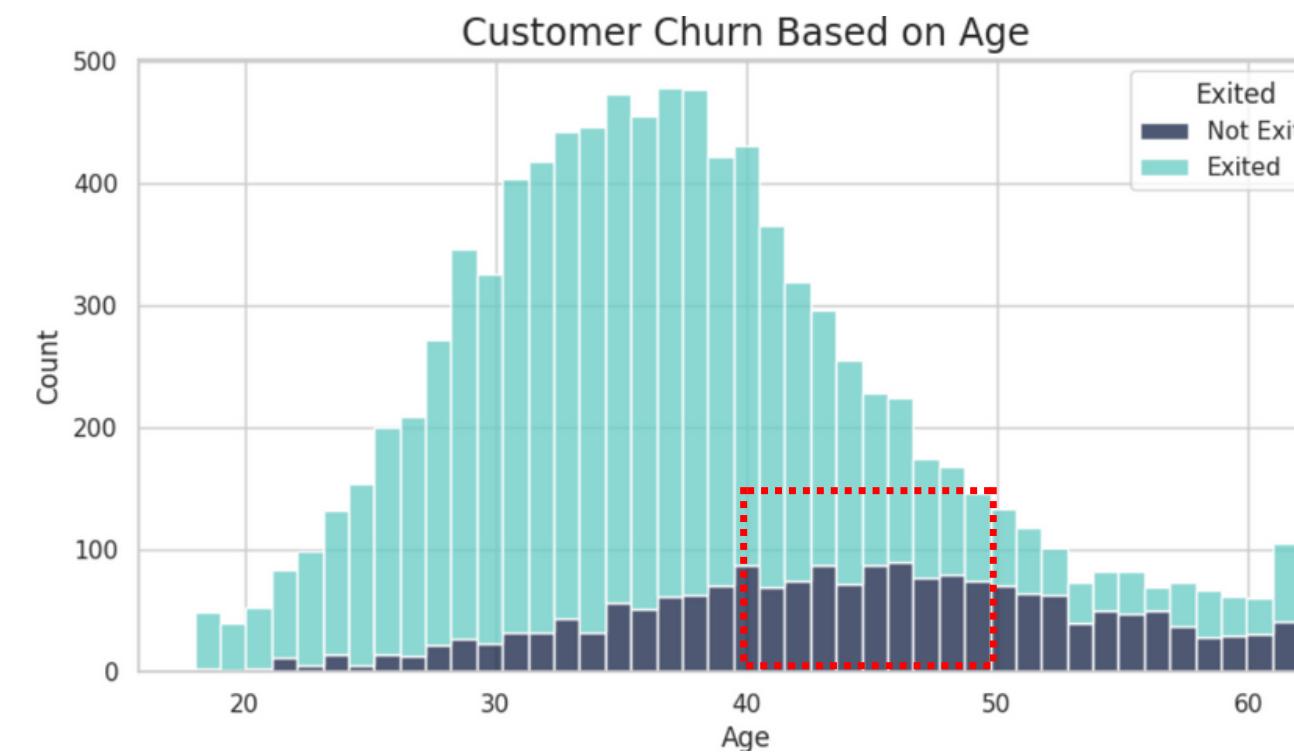
Business Insight

DATA GODDESSES

1 $40 \leq \text{Age Churn} \leq 50$



2 $600 \leq \text{Credit Score} \leq 700$



Customer churn **paling tinggi** terjadi di rentang usia **40-50 tahun**.
Namun, terdapat kecenderungan **semakin tua lebih banyak yang churn**

Terdapat kecenderungan **churn yang tinggi** pada customer dengan **creditscore rentang 600 - 700**.

BANK CBA - DATA GODDESSES

Business Insight

DATA GODDESSES

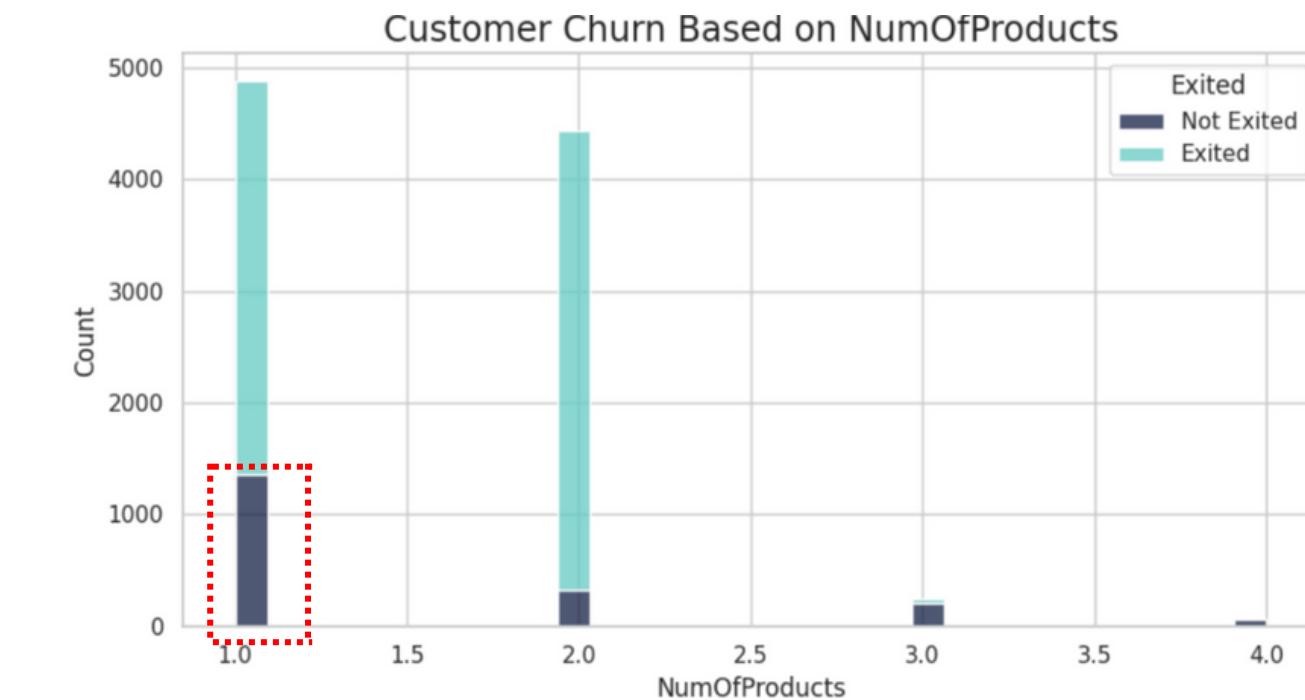
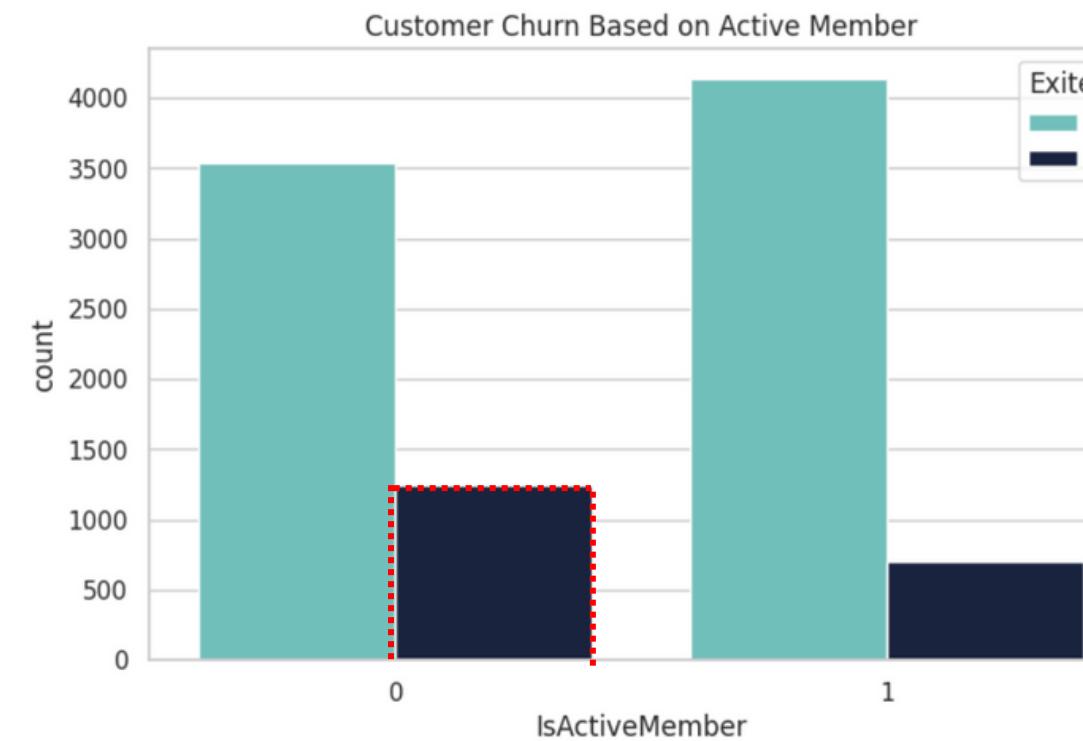
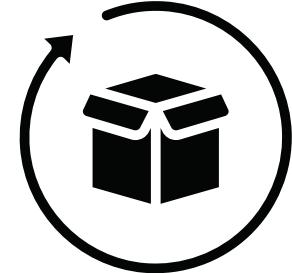
3

Churn when not active



4

Num of Products = 1



Churn **lebih tinggi** pada customer yang tidak aktif

Churn **lebih tinggi** pada customer yang hanya membeli **satu produk**.

BANK CBA - DATA GODDESSES

Business Insight

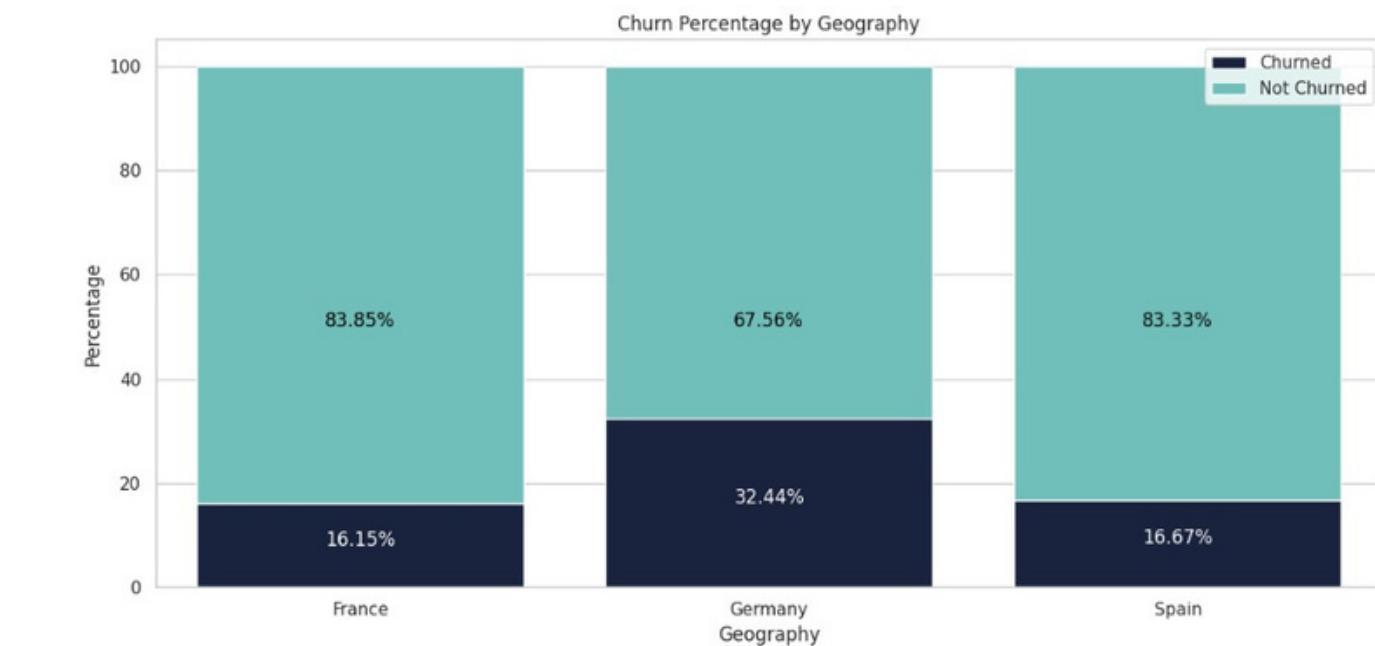
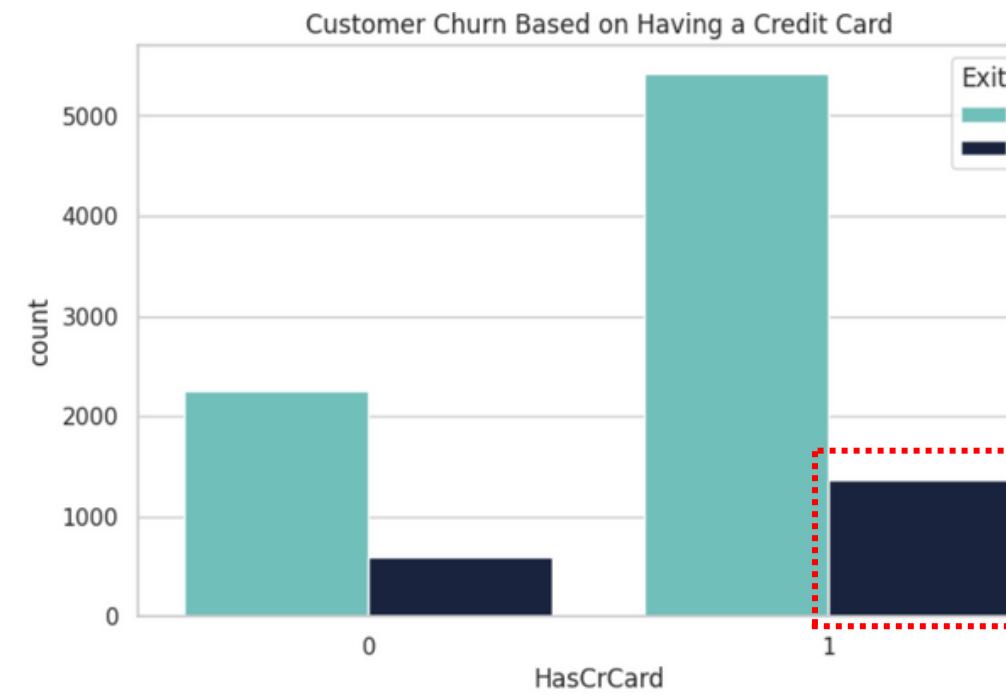
5

Credit Card cenderung churn



6

Geography



Pemilik kartu kredit lebih cenderung untuk **Churn** dibandingkan dengan yang tidak memiliki kartu kredit.

Customer **Germany** memiliki tingkat **churn** lebih tinggi dibanding Spain dan France



DATA PREPROCESSING

● DATA CLEANSING

● FEATURE ENGINEERING

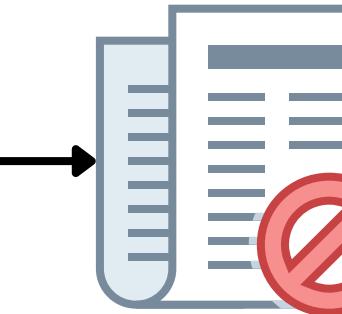
Data Preprocessing

DATA GODDESSES



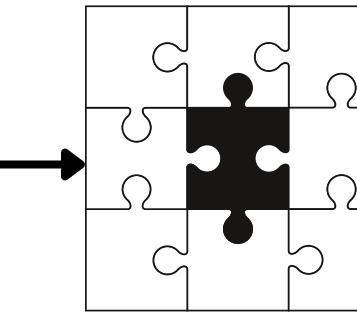
Handling Duplicate

0 Duplicate



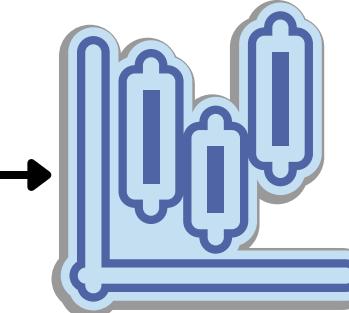
Invalid Data

Semua tipe kolom sesuai



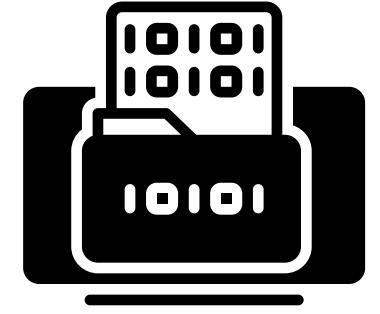
Missing Value

0 Missing Value



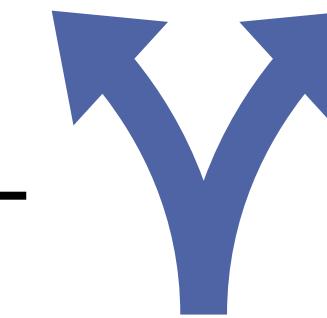
Handling Outlier

IQR



Feature Encoding

One Hot Encoding



Data Split

Train (80%) & Test (20%)



Feature Selection

Chi - Square
EDA



Feature Extraction

Add new feature for
exploration purpose



Handling Imbalance Data

SMOTE



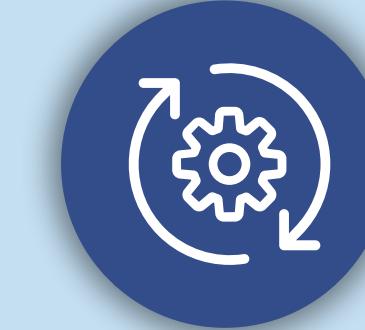
Data Transformation

Standard Scaler

BANK CBA - DATA GODDESSES

Data Preprocessing

DATA GODDESSES



10 kolom dijadikan fitur
untuk modeling

AGE
ISACTIVEMEMBER
BALANCE
NUMOFPRODUCTS
CREDITSCORE
IS_MALE
IS_FEMALE
GEO_FRANCE
GEO_GERMANY
GEO_SPAIN

BANK CBA - DATA GODDESSES



MODELING AND EVALUATION

- MACHINE LEARNING TECHNIQUES
- MODEL EVALUTION PARAMETERS
- MODEL COMPARISON
- APAKAH MODEL SUDAH BEST-FIT?
- PEMILIHAN MODEL
- FEATURE IMPORTANCE

Modeling and Evaluation: Machine Learning Techniques

DATA GODDESSES



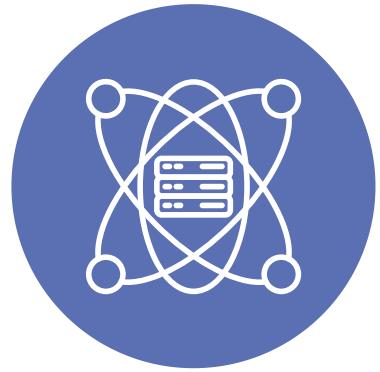
Linear SVC



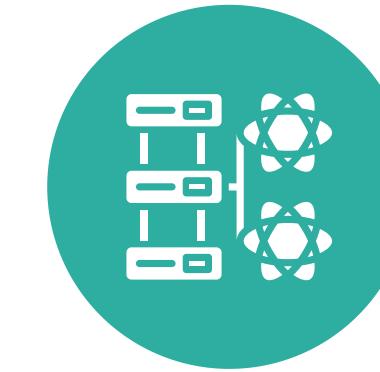
Decision Tree



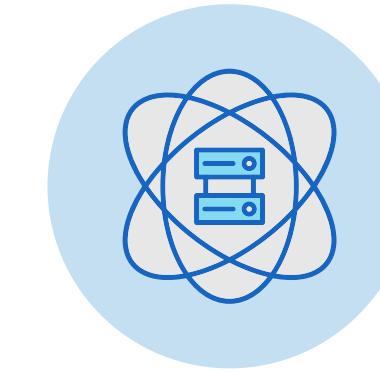
Logistic Regression



Random Forest



KNN
(K-Nearest Neighbors)



Gradient Boosting
Classifier



BEST FIT MODELS



Modeling and Evaluation: Model Evaluation Parameters

DATA GODDESSES

1

Recall

Primary Metrics
Evaluation

Mereduksi False Negative

Customer yang diprediksi tidak churn ✗,
namun sebenarnya churn ✓

Increase
revenue & profit



2

Precision
Secondary Metrics
Evaluation

Mereduksi False Positive

Customer yang diprediksi churn ✓,
namun sebenarnya tidak churn ✗

Reduce Cost
pencegahan customer churn



3

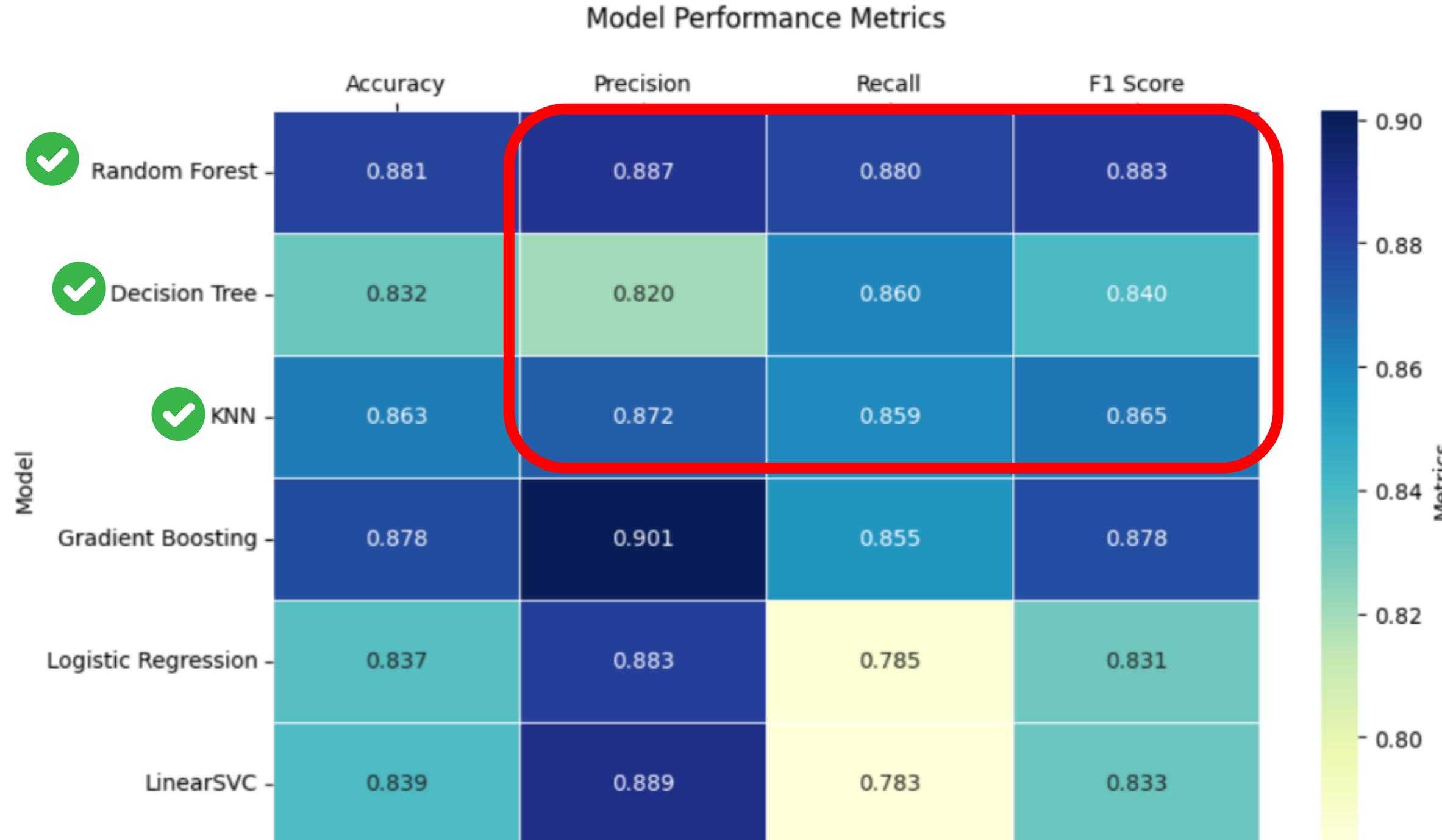
F1 Score
Combined Metrics
Evaluation

Dibutuhkan untuk permasalahan ini, karena diperlukan
sebuah **Trade-off** antara Recall dan Precision.
F1 Score berfungsi sebagai harmonisasi rata-rata dari
Recall dan Precision.



Modeling and Evaluation: Model Comparison

DATA GODDESSES



Nilai Recall, Precision, dan F1 Score mendapatkan evaluasi yang baik pada
Random Forest, Decision Tree, dan KNN

Modeling and Evaluation: Pemilihan Model



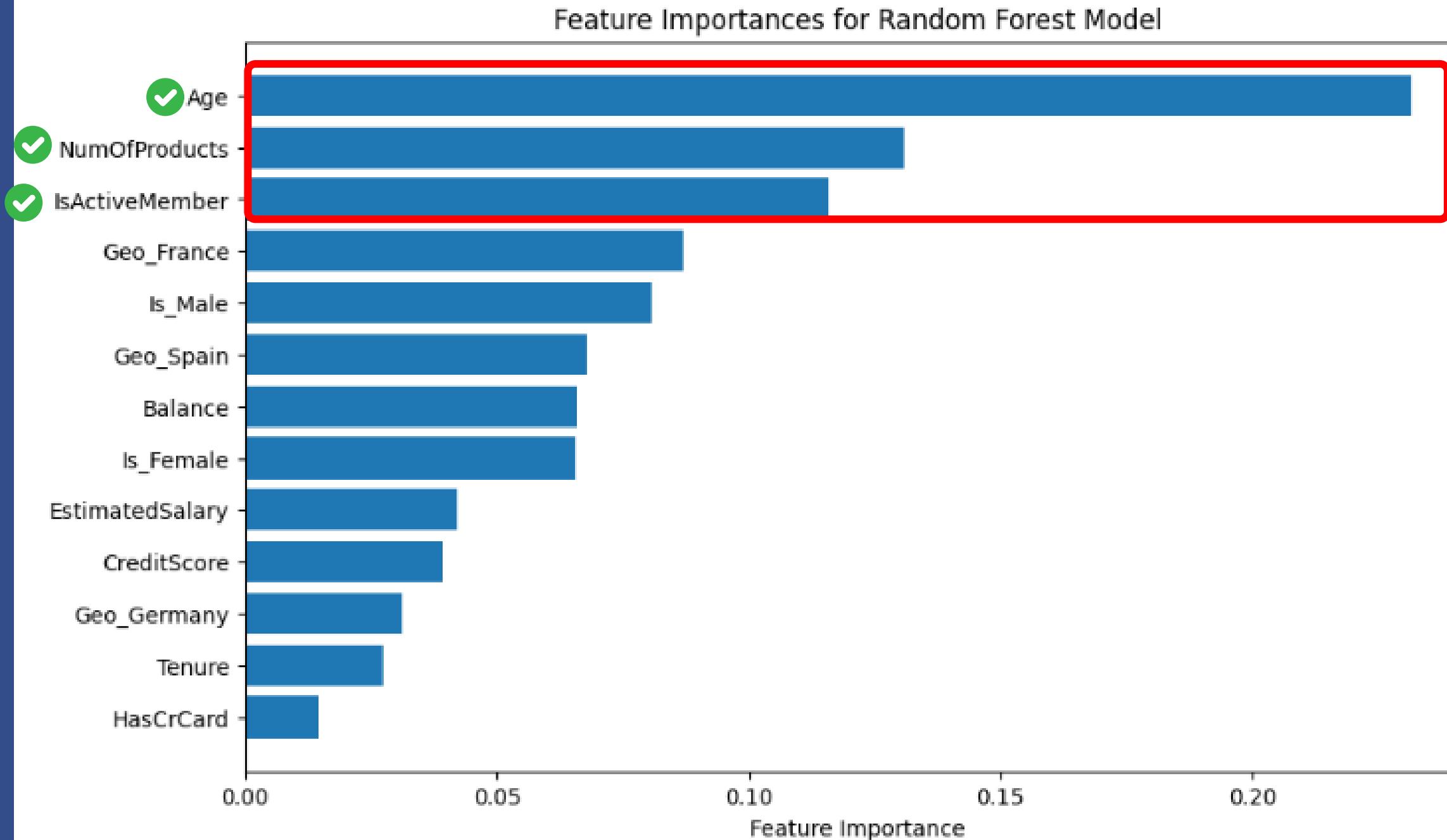
1. TUJUAN MODEL METRIKS
memiliki **tingkat recall** pada data uji **tertinggi** di antara model-model yang diuji (sekitar **88%**).

2. WAKTU KOMPUTASI
lebih **cepat** dan **efisien**, sehingga dapat **menghemat waktu dan sumber daya**.

3. NON-ASUMSIONAL
tidak bergantung pada asumsi bahwa data harus **berdistribusi normal**.

Modeling and Evaluation:

Feature Importance



3 Faktor utama yang mempengaruhi customer churn yaitu **AGE**, **NUM OF PRODUCTS**, dan **ACTIVE MEMBER**. 3 Faktor ini akan digunakan dalam business recommendation.

Modeling and Evaluation: Feature Importance

EVALUASI FEATURE IMPORTANCE RANDOM FOREST

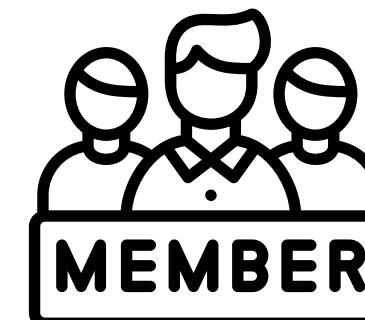
Usia customer adalah faktor paling penting yang memengaruhi churn.



Customer yang menggunakan lebih banyak produk mungkin lebih cenderung untuk tetap menjadi customer.



Customer yang tidak aktif akan cenderung lebih churn dibanding yang aktif.





BUSINESS **RECOMMENDATION**



RECOMMENDATIONS

BANK CBA - DATA GODDESSES





Pertahankan Fokus pada Segment Usia

Ditujukan untuk **seluruh customer dari berbagai usia**.

Sesuaikan **strategi retensi berdasarkan segmentasi usia atau life stages financial**. **early life** (15-30), **mid life** (31-50), **pra pensiun** (51-60), **pensiun** (60+).

Menawarkan produk atau program yang sesuai kebutuhan usia, seperti menawarkan program pensiun untuk customer yang lebih tua.

Business Recommendations



Optimalkan Penawaran Produk

Diutamakan untuk **customer yang hanya menggunakan satu jenis produk**.

Menawarkan paket produk yang menarik dengan keuntungan tambahan untuk customer yang memiliki lebih dari satu produk. Misal, memberikan diskon/cashback pada customer yang memiliki 2 produk sekaligus.



Meningkatkan Komunikasi dengan Customer

Diutamakan untuk **customer yang kurang aktif/tidak aktif**.

Berkomunikasi melalui berbagai saluran, seperti email atau pesan langsung ke nomor telefon untuk **menjelaskan nilai tambah** jika mereka aktif dalam menggunakan layanan bank.



BUSINESS **SIMULATION**



BUSINESS SIMULATION

Penting untuk melakukan simulasi bisnis guna mengevaluasi sejauh mana kesesuaian dengan metrik bisnis.

Business Simulation

DATA GODDESSES

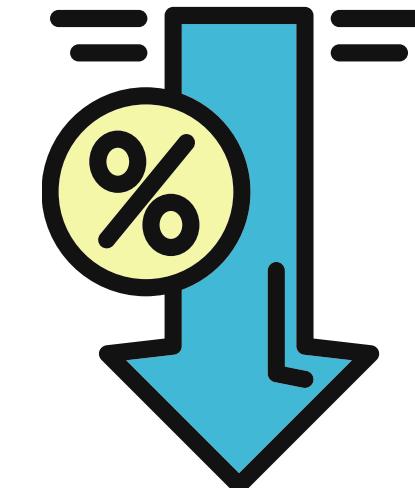
Churn Rate

20.4%

Before model

Menurun Hingga :

15.56%



4.84%

After model

Churn rate : FN / Total customer

Sebanyak 484 customer yang tidak terdeteksi akan churn (FN) dari 10.000 customer.

Dalam kasus ini, Bank CBA berpotensi menurunkan churn dari 20.4% hingga 4.84%

BANK CBA - DATA GODDESSES





Thank You

For Your Attention