

# BANK CHURN PREDICTION

Final Project

Rakamin Academy Data Science  
Batch 24

STEADFAST AND CO.



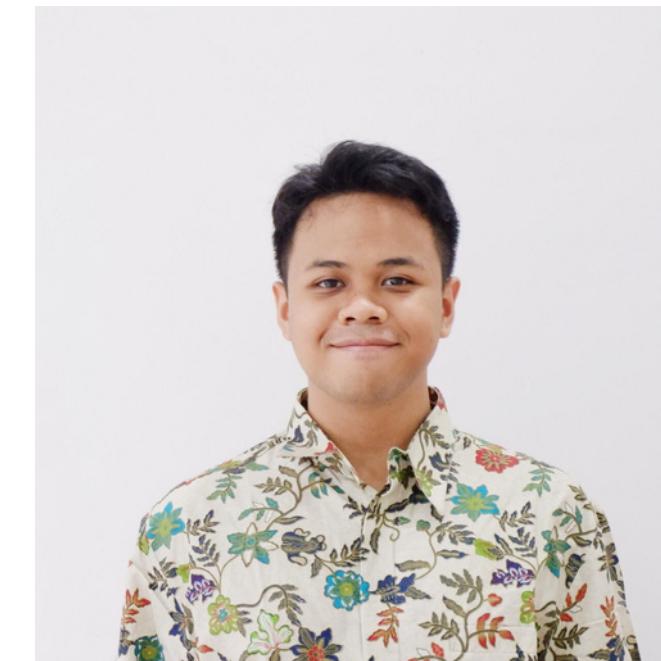
# STEADFAST AND CO. TEAM



**Rafif Rizki  
Zaidan**  
Data Science Lead



**Muhammad  
Hibatur Akmal**  
Data Science Member



**Gerald Alfa Daud  
Manas**  
Data Science Member

# TABLE OF CONTENTS



- **Bagian 1:**  
Business Understanding & Business Insights
- **Bagian 2:**  
Data Pre-Processing
- **Bagian 3:**  
Machine Learning Evaluation
- **Bagian 4:**  
Business Action & Recommendation



## Bagian 1: Business Understanding & Business Insight

# Background

Client kami adalah Bank SCB yang telah melakukan ekspansi ke Uni Eropa selama 10 tahun terakhir.

Bank SCB meminta tim Steadfast and Co. untuk memberikan insight dari database nasabah yang melakukan churn.

Insight digunakan untuk memberikan business recommendation agar pihak Bank dapat memfokuskan tindakan kepada nasabah yang melakukan churn.

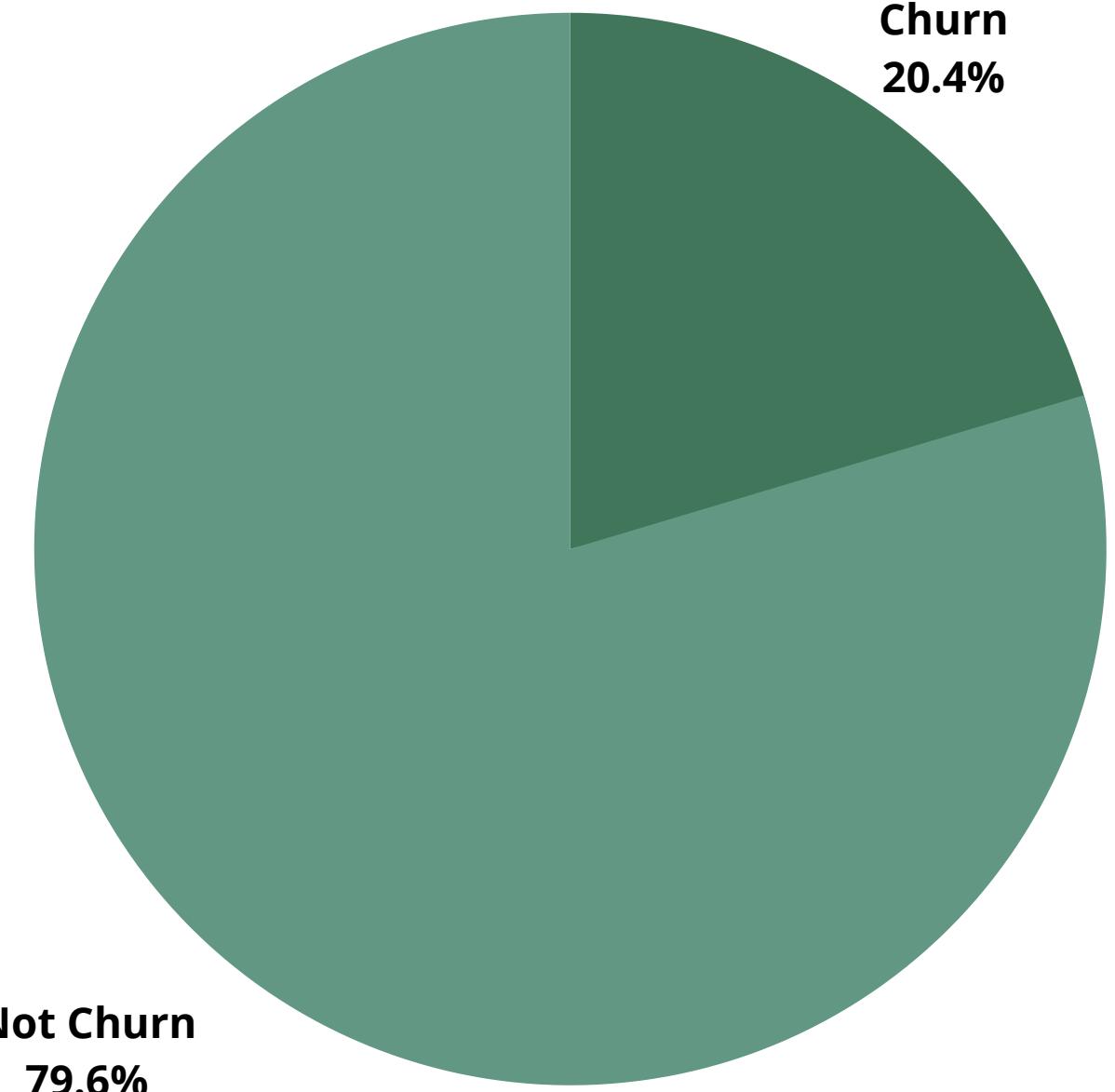


# MENGAPA?

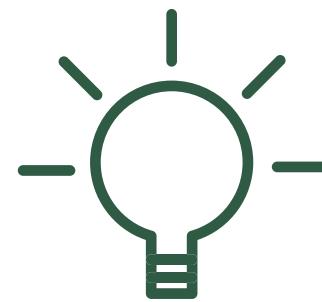
Pada tahun ini, dari 10.000 nasabah, persentase nasabah yang melakukan churn adalah 20.37% dari total nasabah yang ada di Bank SCB.

Biaya untuk mendapatkan customer baru mencapai **5-6 kali lebih mahal** dibandingkan dengan mempertahankan customer lama (Benlan *et al.*, 2014).

Menurut Gallo dalam Harvard Business Review (2014), mengurangi tingkat churn customer hingga **5%** dapat meningkatkan keuntungan perusahaan sebesar **25-95%**.



# BUSINESS UNDERSTANDING



## Problem Statement

Bagaimana cara memprediksi customer yang churn sehingga dapat mengurangi tingkat churn?

## Business Metrics

$$\text{Customer Churn Rate} = \frac{\text{Lost Customers}}{\text{Total Customers at the Start of Time Period}} \times 100$$

(Gandy, 2009).

## Goals

Membantu proses identifikasi melalui sistem prediksi sehingga tidak perlu untuk menambah biaya atau resource untuk mengidentifikasi customer yang akan churn.

## Objective

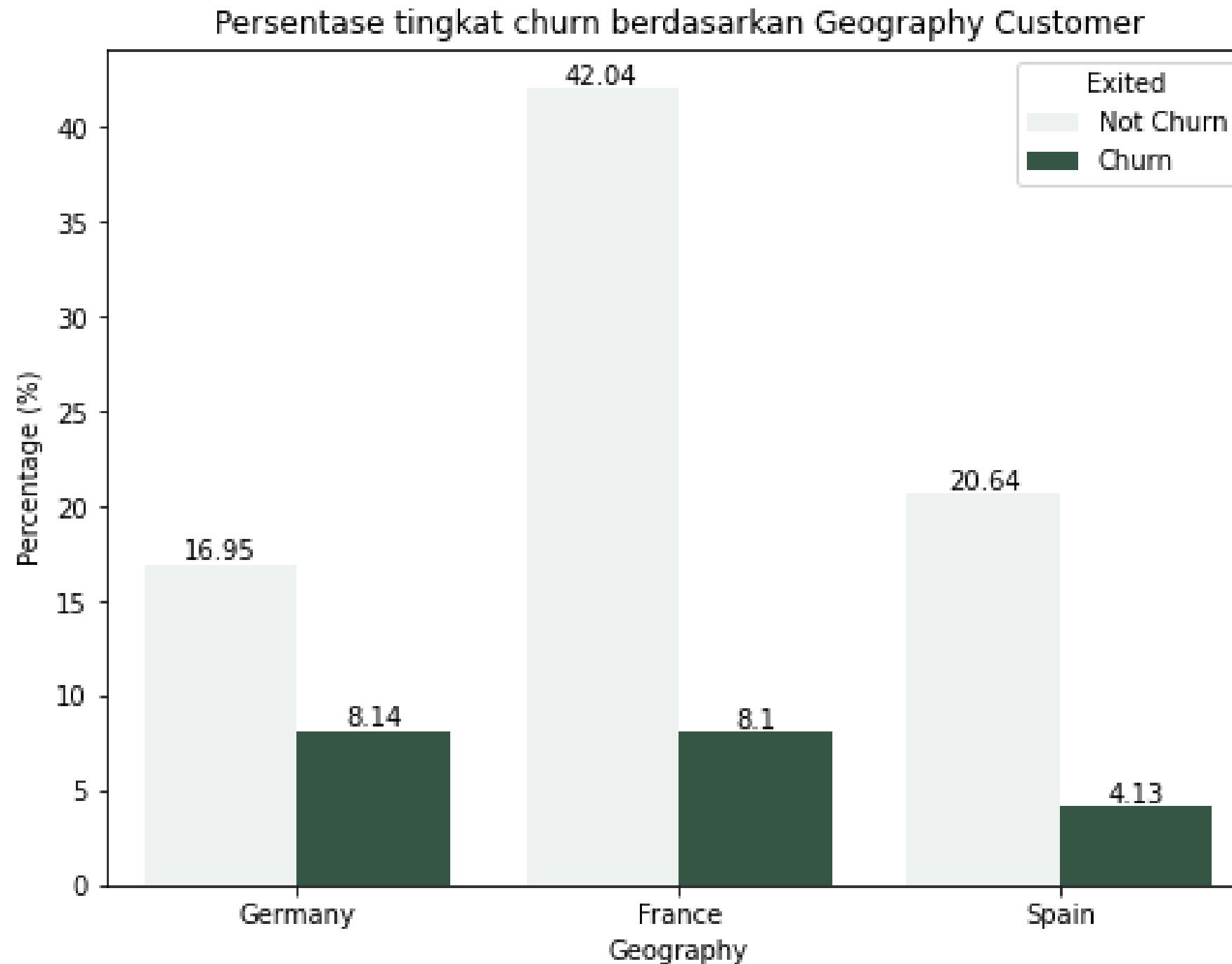
Membuat modelling untuk membantu proses identifikasi (faktor yang mempengaruhi dan customer yang berpotensi) untuk churn.

# PENJELASAN DATASET

Dataset memiliki 14 kolom dan 10.000 baris. Target dari dataset ini adalah Exited.

Column Name	Explanation
RowNumber	Corresponds to the record (row) number and has no effect on the output.
CustomerId	Contains random values and has no effect on customer leaving the bank.
Surname	The surname of a customer has no impact on their decision to leave the bank.
CreditScore	Can influence customer churn, since a customer with a higher credit score is less likely to leave the bank.
Geography	A customer's location can affect their decision to leave the bank.
Gender	It's interesting to explore whether gender plays a role in a customer leaving the bank.
Age	This is certainly relevant, since older customers are less likely to leave their bank than younger ones.
Tenure	Refers to the number of years that the customer has been a client of the bank.
Balance	Also, a very good indicator of customer churn, as people with a higher balance in their accounts are less likely to leave the bank compared to those with lower balances.
NumOfProducts	Refers to the number of products that a customer has purchased through the bank.
HasCrCard	Denotes whether a customer has a credit card. This column is also relevant, since people with a credit card are less likely to leave the bank.
IsActiveMember	Active customers are less likely to leave the bank.
EstimatedSalary	As with balance, people with lower salaries are more likely to leave the bank compared to those with higher salaries.
Exited	Whether or not the customer left the bank.

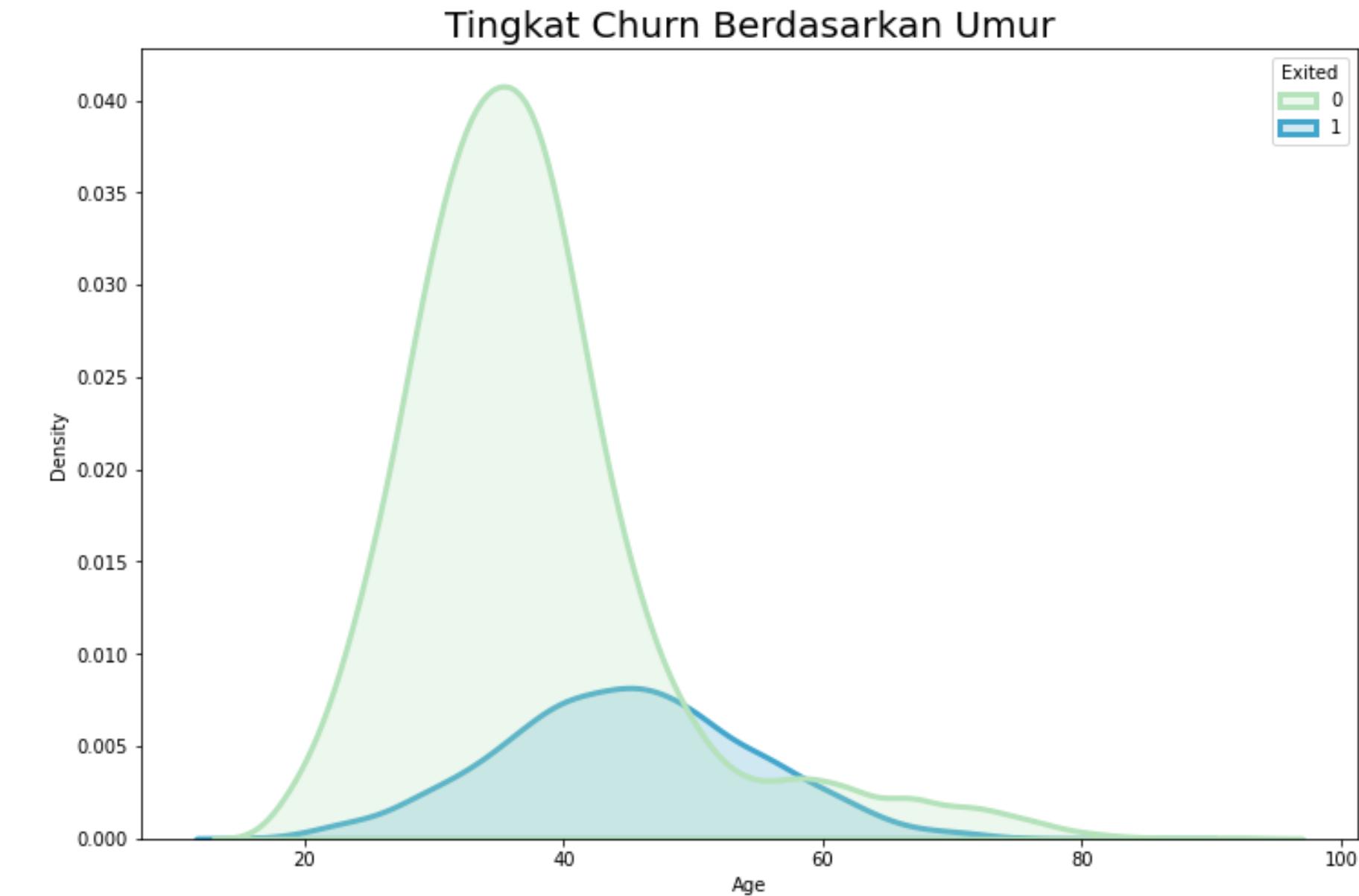
# EXPLORATORY DATA ANALYSIS



Nasabah yang churn adalah nasabah yang berada pada wilayah **Geography Germany**.

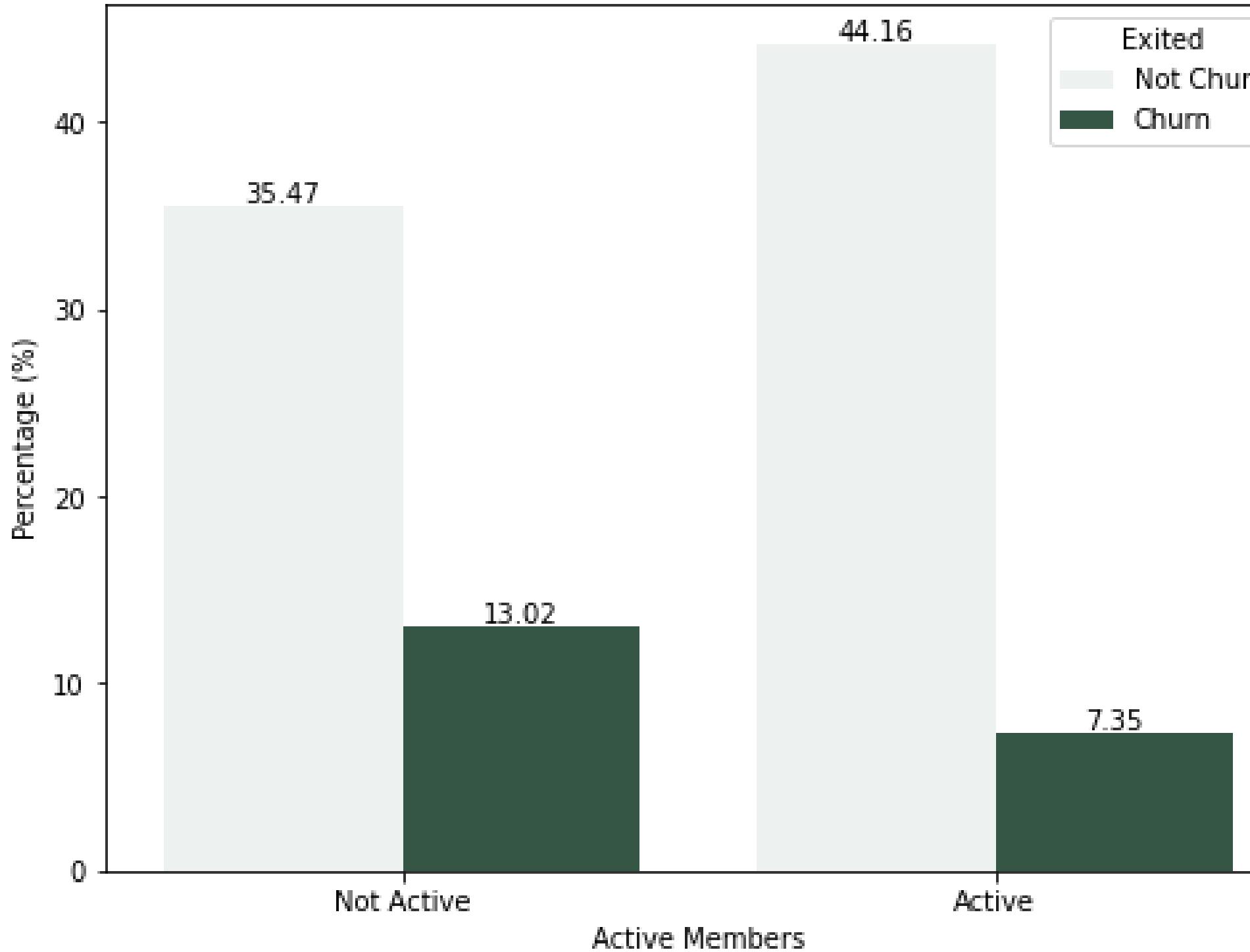
# EXPLORATORY DATA ANALYSIS

Customer dengan **kisaran umur di atas 20 tahun hingga di bawah 40** cenderung tidak Churn. Akan tetapi, **dari umur 40 mulai terlihat tingkat churn yang semakin besar** hingga pada kisaran umur **60 tahun.**



# EXPLORATORY DATA ANALYSIS

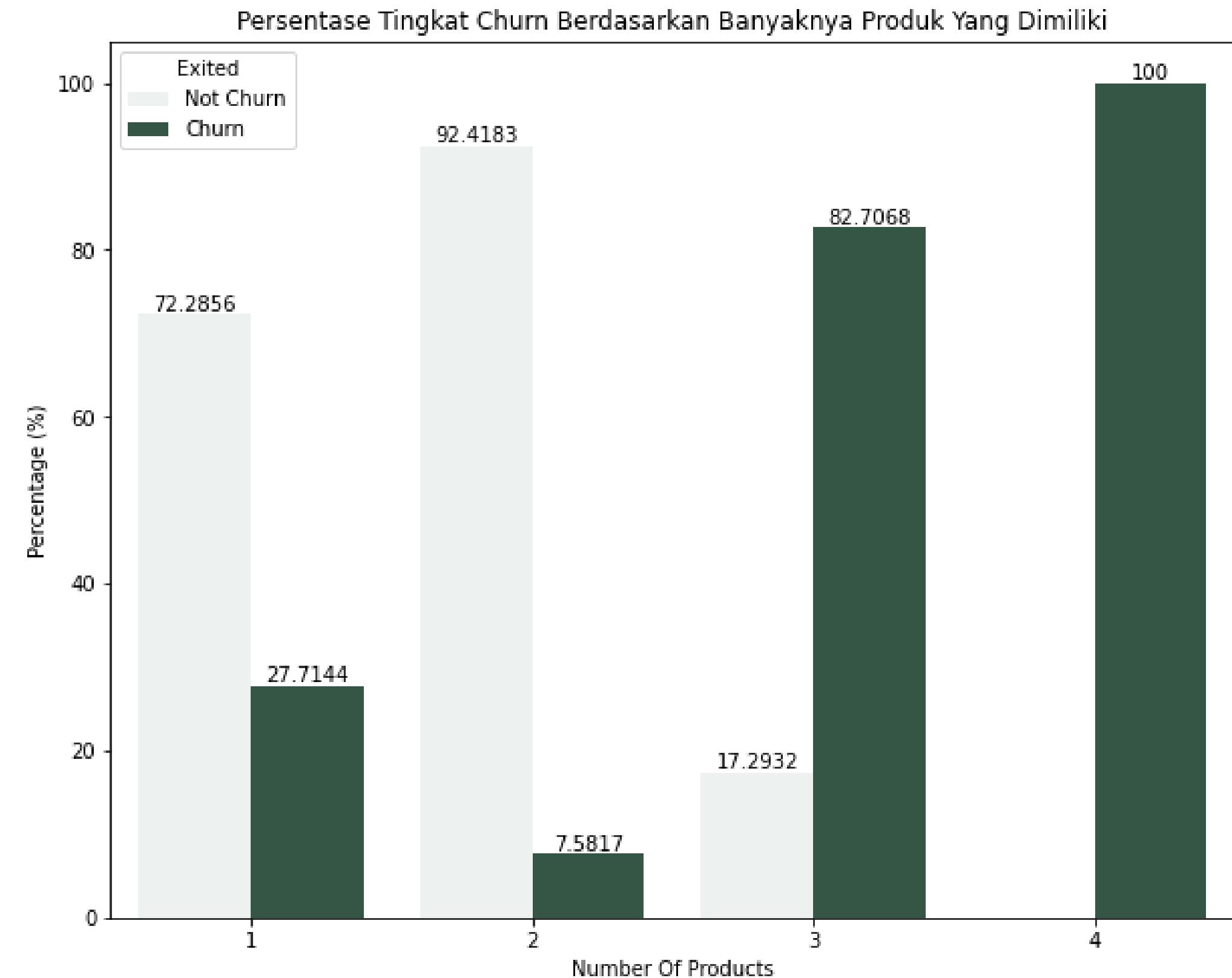
Percentase Customer Churn berdasarkan Active Member



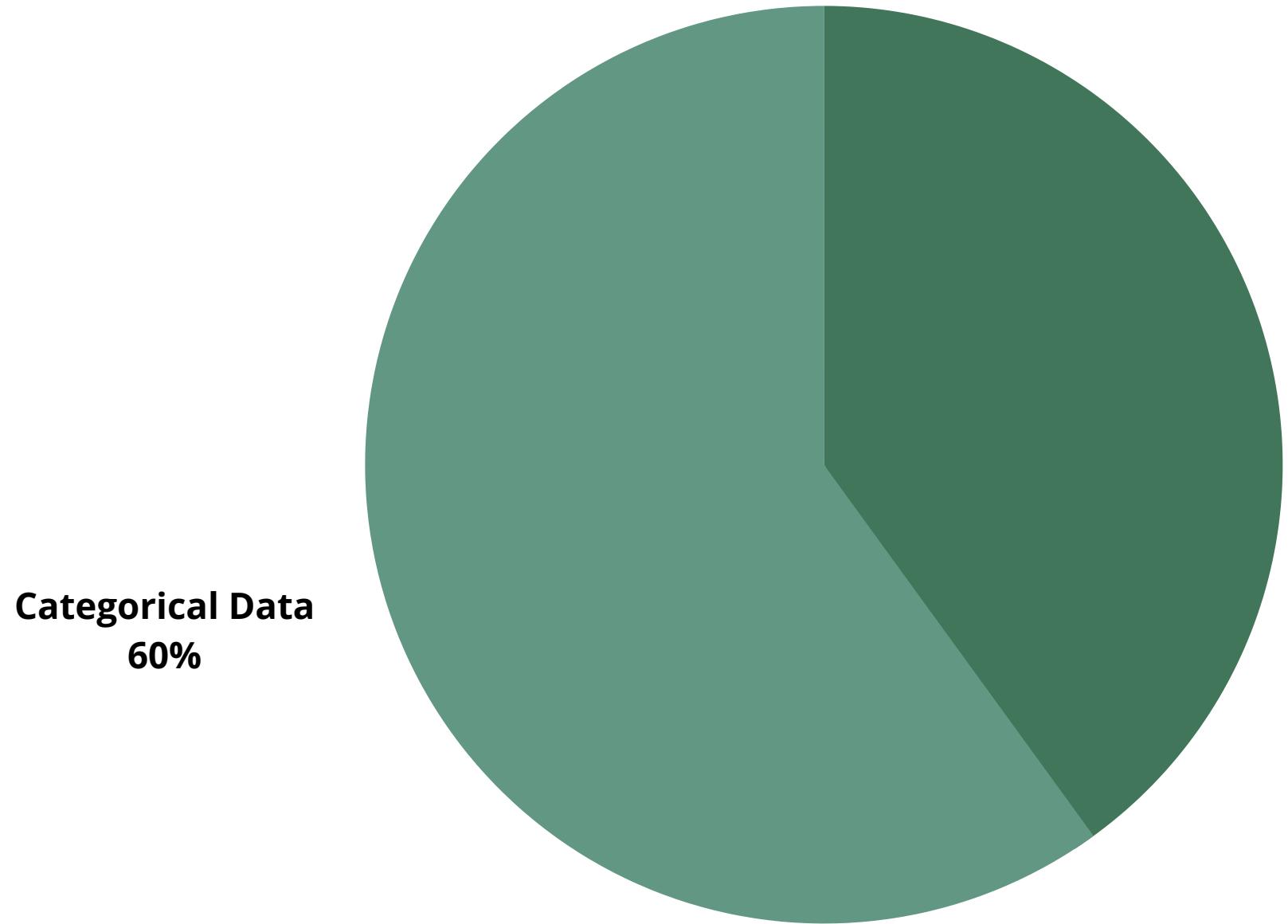
Sebanyak **13.02%** Customer yang melakukan churn adalah adalah **customer yang tidak aktif**.

# EXPLORATORY DATA ANALYSIS

Customer dengan jumlah produk lebih dari 2 cenderung untuk melakukan churn.



# DATA PRE-PROCESSING



**13 Kolom Dan 10000 Baris**

## Dropped Features

- RowNumber
- CustomerId
- Surname

## Categorical

- Geography
- Gender
- Tenure
- NumOfProducts
- HasCrCard
- IsActiveMember

## Numerical

- CreditScore
- Age
- Balance
- EstimatedSalary

## Target

- Exited

# DATA PRE-PROCESSING: IN DETAIL

Missing Values	—————	<b>Tidak Ada</b>
Duplicate Data	—————	<b>Tidak Ada</b>
Handling Outliers	—————	<b>10000 → 9859</b>
Log Transformation	—————	<b>1 Kolom</b>
Normalisasi	—————	<b>Tidak Ada</b>
Standarisasi	—————	<b>4 Kolom</b>
Feature Encoding	—————	<b>Dari 10 Kolom → 13 Kolom</b>
Feature Selection	—————	<b>10 Kolom</b>
Feature Extraction	—————	<b>Tidak Ada</b>

# EVALUATION METRICS



AUC Score

**Kami menggunakan Evaluation Metrics AUC Score.**

AUC Score dipilih karena Metric tersebut memperlihatkan seberapa besar kemungkinan Model dapat membedakan nasabah churn dan tidak churn.

Selain itu, Metrics tersebut juga cocok untuk digunakan karena jumlah variabel target pada model dataset yang dilakukan processing imbalanced.

# MODEL EVALUATION

Model	Accuracy	Precision	Recall	F1	Test Score	Train Score
XGBoost Classifier	0.87	0.79	0.42	0.55	0.86	0.89
Random Forest Classifier	0.86	0.78	0.36	0.49	0.86	0.93
Adaboost Classifier	0.85	0.70	0.38	0.49	0.84	0.86
Decision Tree	0.85	0.74	0.38	0.50	0.81	0.86
K-nearest Neighbor	0.81	0.54	0.35	0.42	0.79	0.83
Logistic Regression	0.82	0.57	0.27	0.37	0.78	0.89

Berdasarkan semua model yang telah di-tune, Model yang memiliki hasil metric AUC terbaik adalah **XGBoost Classifier**.

# Top Rank Feature Importance

Dari model Machine Learning yang sudah di tuning, kami memilih empat (4) Feature Importance teratas untuk kami berikan Business Recommendations.

**1st**



**Number of Product**

**2nd**



**Age**

**3rd**



**Active Members**

**4th**



**Geography**

# BUSINESS RECOMMENDATIONS

1st



Number of Product

**Menambahkan value dan/atau skema bonus pada pembelian lebih dari dua produk** (paket bundling, mengurangi biaya pajak dan bunga, memberikan voucher/reward skala kecil, menambahkan layanan produk).

2nd



Age

**Memberikan penawaran finansial kepada nasabah pada rentang umur 40-60 tahun** (program kredit rumah, kredit mobil) dan **menawarkan Financial Wellness Program** untuk masing-masing nasabah.

# BUSINESS RECOMMENDATIONS

3rd

**Memberikan notifikasi non-provokatif** untuk member yang tidak aktif dan memberikan **beberapa promosi/one-time big promotion** (cashback/discount/bundling deal) berdasarkan pemakaian/pembelian produk sebelumnya.



Active Members

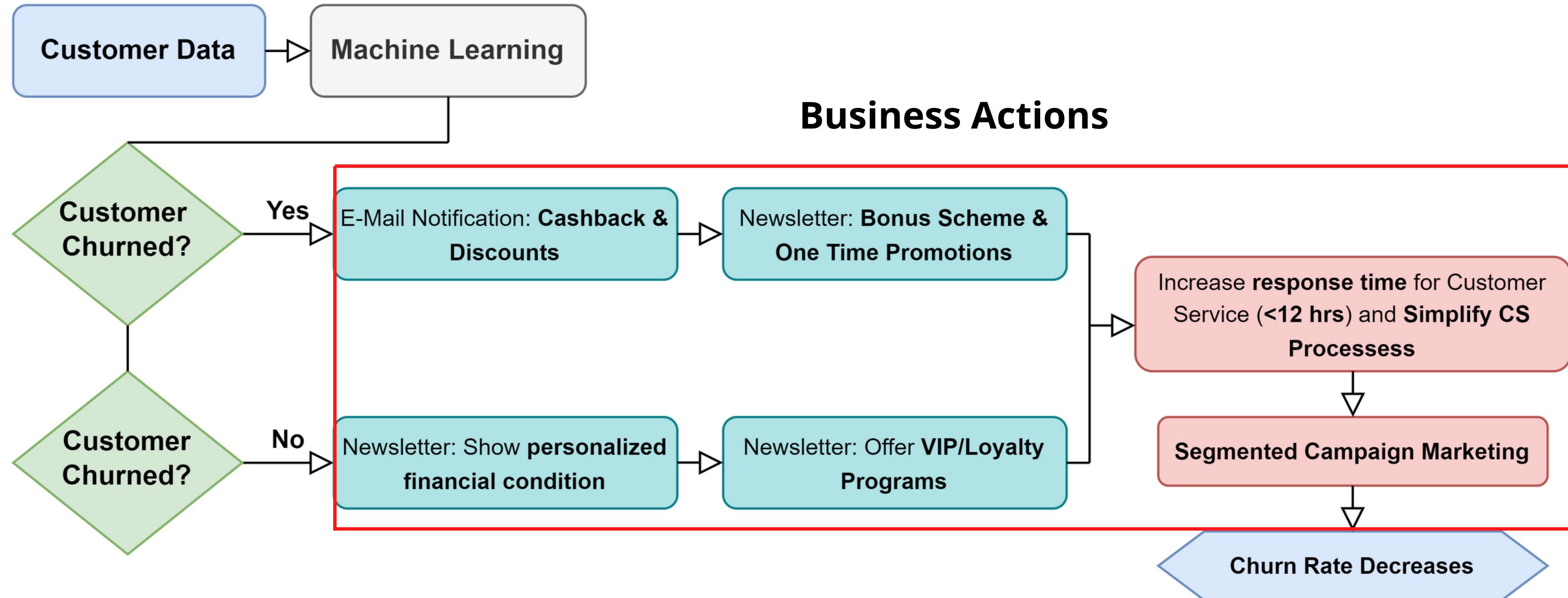
4th

**Menambahkan value produk** pada cabang di **wilayah Germany** (paket bundling, mengurangi biaya pajak dan bunga, memberikan voucher/reward skala kecil, menambahkan layanan produk), **mempercepat dan mempermudah skema Customer Service** agar lebih nyaman untuk di akses.



Geography

# BUSINESS ACTION FLOW





# IMPACT SIMULATION

## Hasil Deteksi Machine Learning

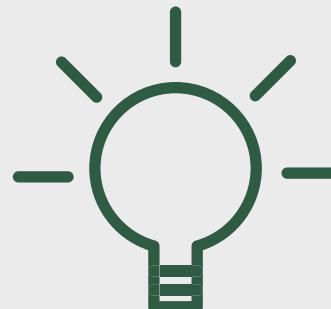
Skor AUC Data Testing: 86%

**Jumlah Churn terdeteksi = Jumlah Churn x Skor AUC**

$$\begin{aligned} &= 2037 \times 86/100 \\ &= \mathbf{1751 \text{ dari } 2037} \end{aligned}$$

Sisa nasabah yang tidak terdeteksi dengan benar: 286 Orang

Dari skor AUC Data Testing, yaitu 86%, kami asumsikan bahwa Machine Learning dapat mendeteksi tepat sebanyak 86% dari seluruh customer yang churn, yaitu **1751 orang**.



Menurut website **Forbes**, Probabilitas untuk menjual ke customer yang sudah ada (Existing Customer) adalah **60-70%**.

## Perhitungan Keberhasilan Business Action

Menggunakan asumsi probabilitas sebelumnya, kami memperhitungkan jumlah customer yang berhasil dilakukan penanganan dari Business Recommendation.

**Keberhasilan BA = Jumlah Churn - (Jumlah Churn x Probabilitas Selling)**

$$\begin{aligned} &= 1751 - (1751 \times 60\%) \\ &= \mathbf{1050} \end{aligned}$$

Dari total 2037 orang yang Churn, **sebanyak 1050 orang** dapat berubah menjadi tidak Churn. Jumlah akhir customer yang masih churn adalah sebanyak kurang lebih 986 orang.

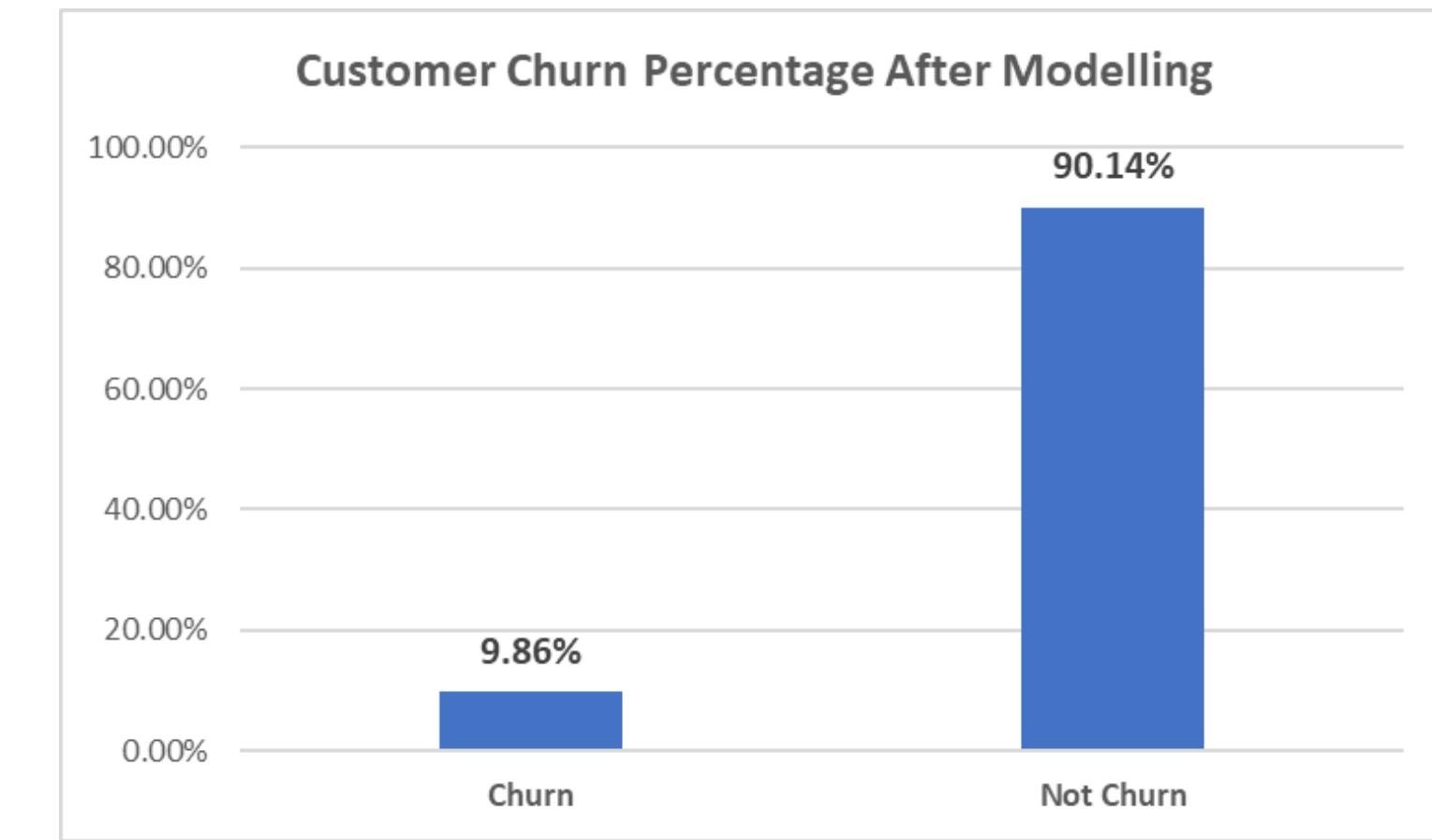
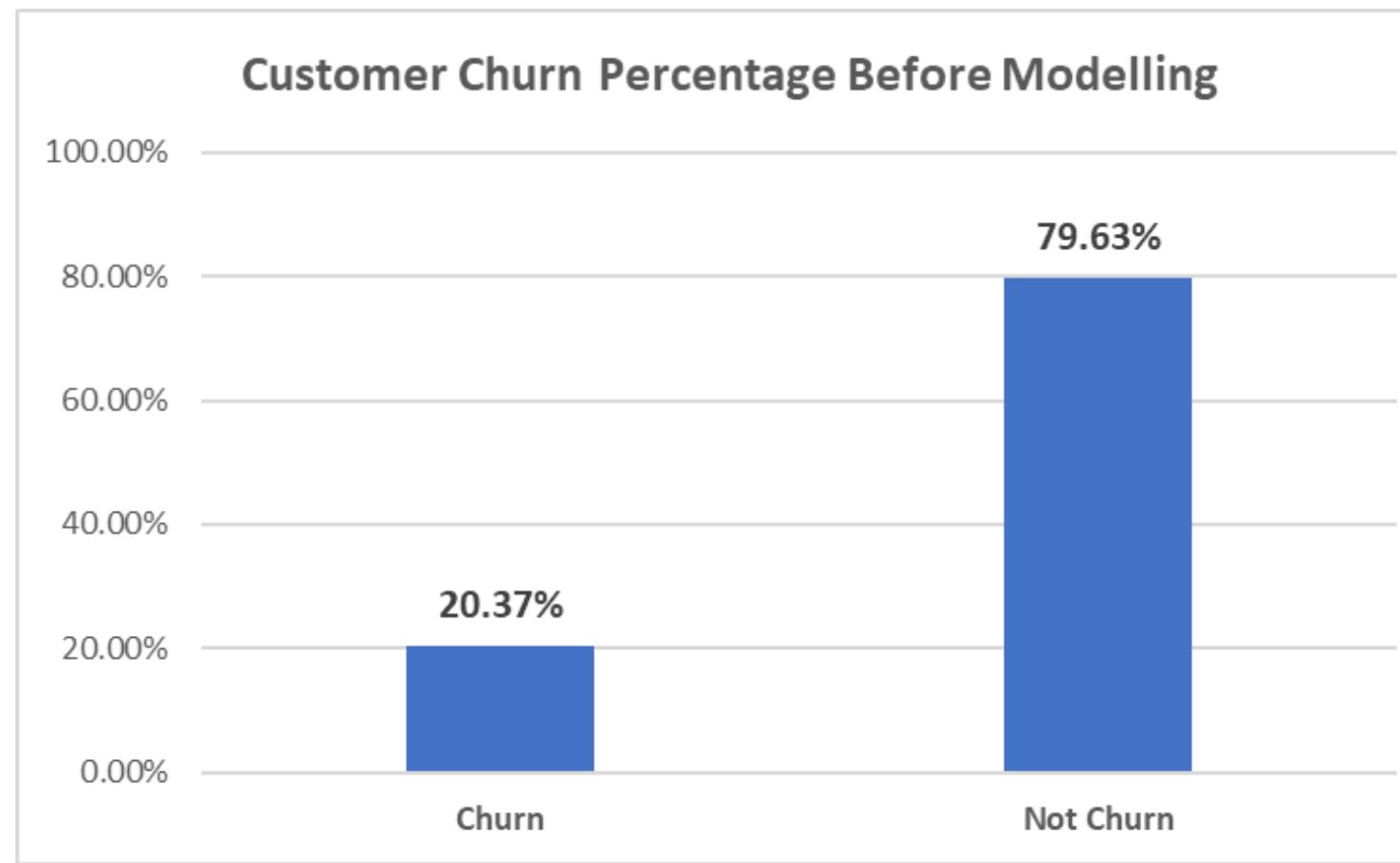
## Perhitungan Churn Rate Akhir

**Customer Churn = (Lost Customer/Total Customer) x 100%**

$$\begin{aligned} &= (986/10.000) * 100\% \\ &= \mathbf{9.86\%} \end{aligned}$$

Jadi, hasil Churn Rate setelah modelling dan penanganan dengan Business Action adalah **9.86%**.

# FINAL RESULT



Dari hasil modelling machine learning, rekomendasi bisnis dan perhitungan asumsi kesuksesan implementasi aksi bisnis, kami dapat menurunkan persentase churn dari **20.37% menjadi 9.86%**. Kami dapat **menurunkan persentase churn hingga 10.51%**.

Dengan menurunkan churn, Bank SCB dapat menambahkan revenue tambahan melalui nasabah yang dipertahankan kembali dengan aksi bisnis dari *business recommendation*.



---

Thank you!

---



