BANK CHURN PREDICTION

Final Project

STEADFAST AND CO.

Rakamin Academy Data Science
Batch 24

STEADFAST AND CO. TEAM

Rafif Rizki Zaidan

Ketua Kelompok

Muhammad Hibatur Akmal

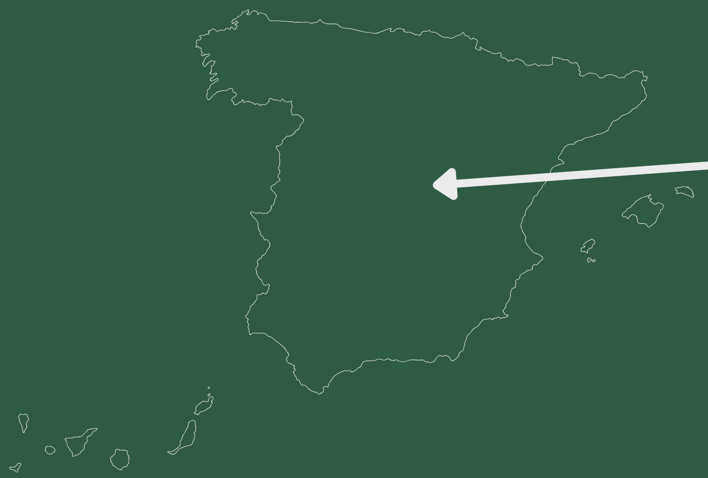
Anggota

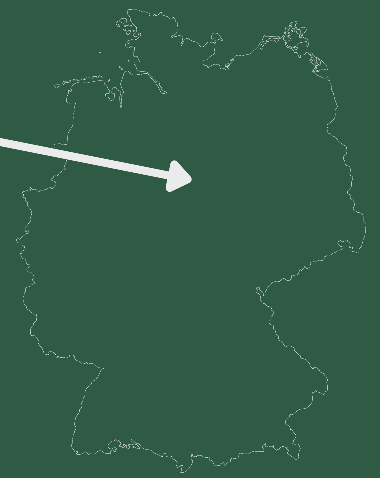
Gerald Alfa Daud Manas

Anggota

TABLE OF CONTENTS

BACKGROUND





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

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

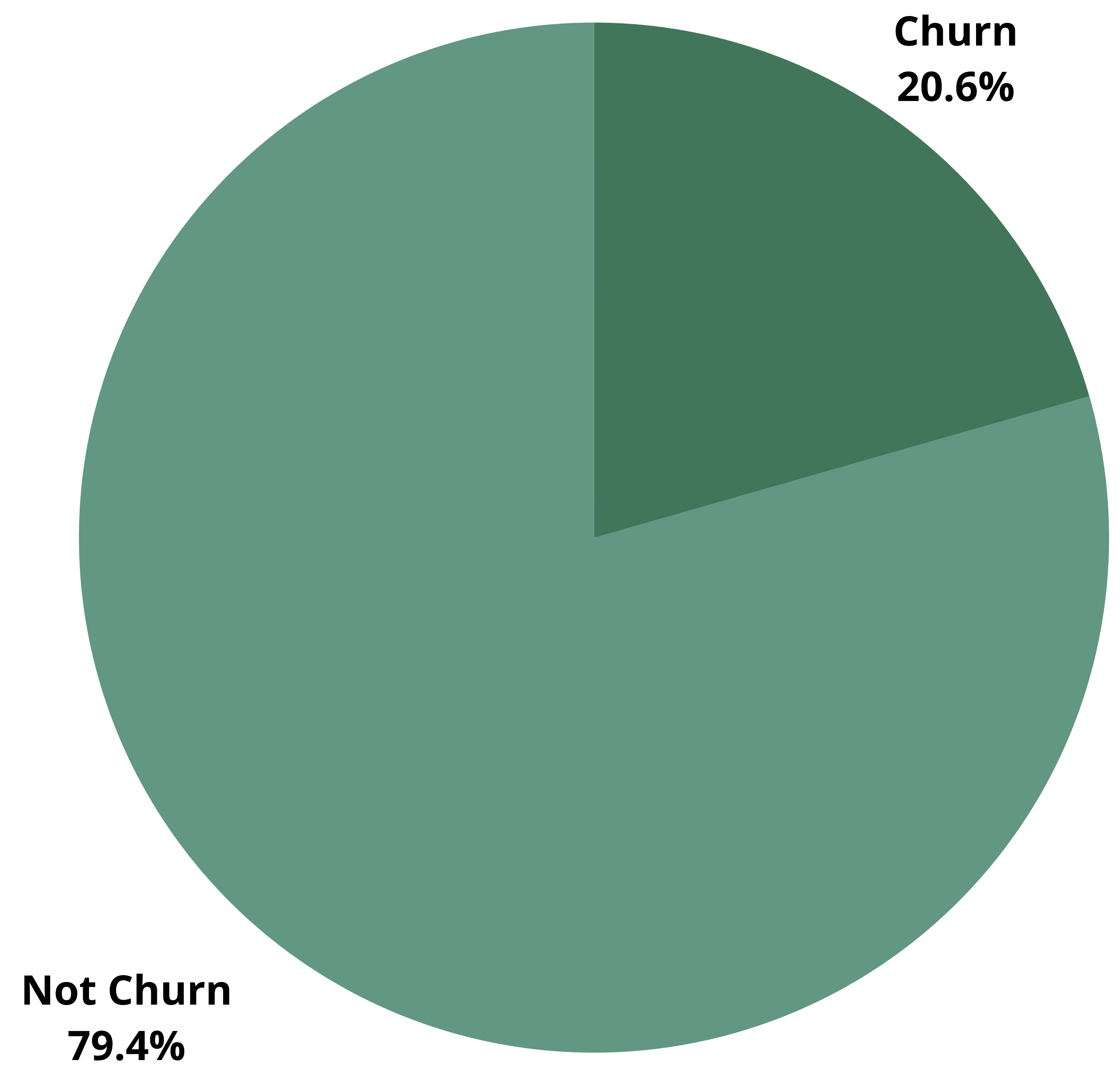
Bank SCB meminta tim Steadfast and Co. untuk memberikan *insight* dari database nasabah yang melakukan churn.

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

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

(Gandy, 2009).

Goals

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

Objective

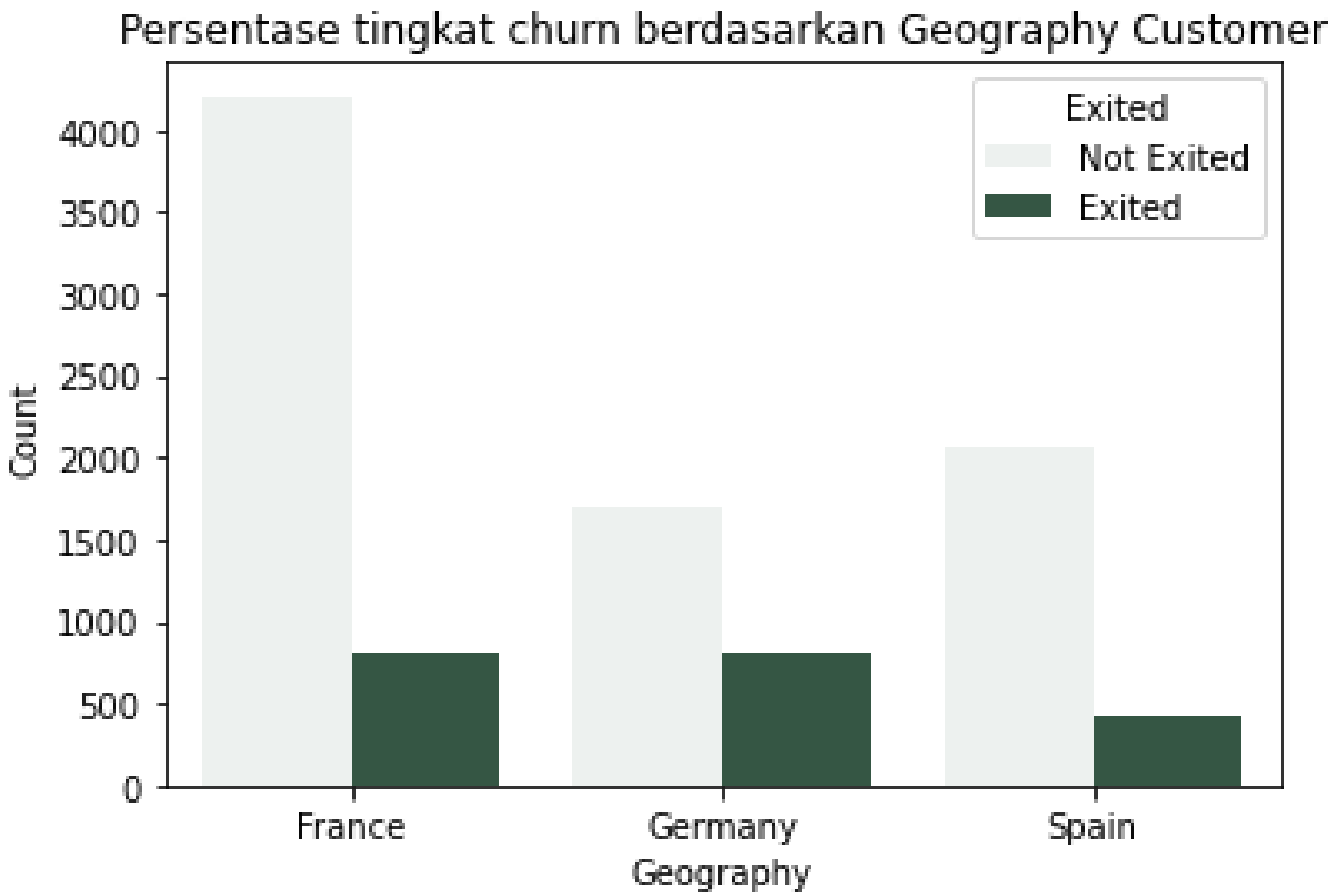
Membuat modelling untuk membantu proses identifikasi (faktor yang mempengaruhi dan customer yang berpotensi) untuk churn.

PENJELASAN DATASET

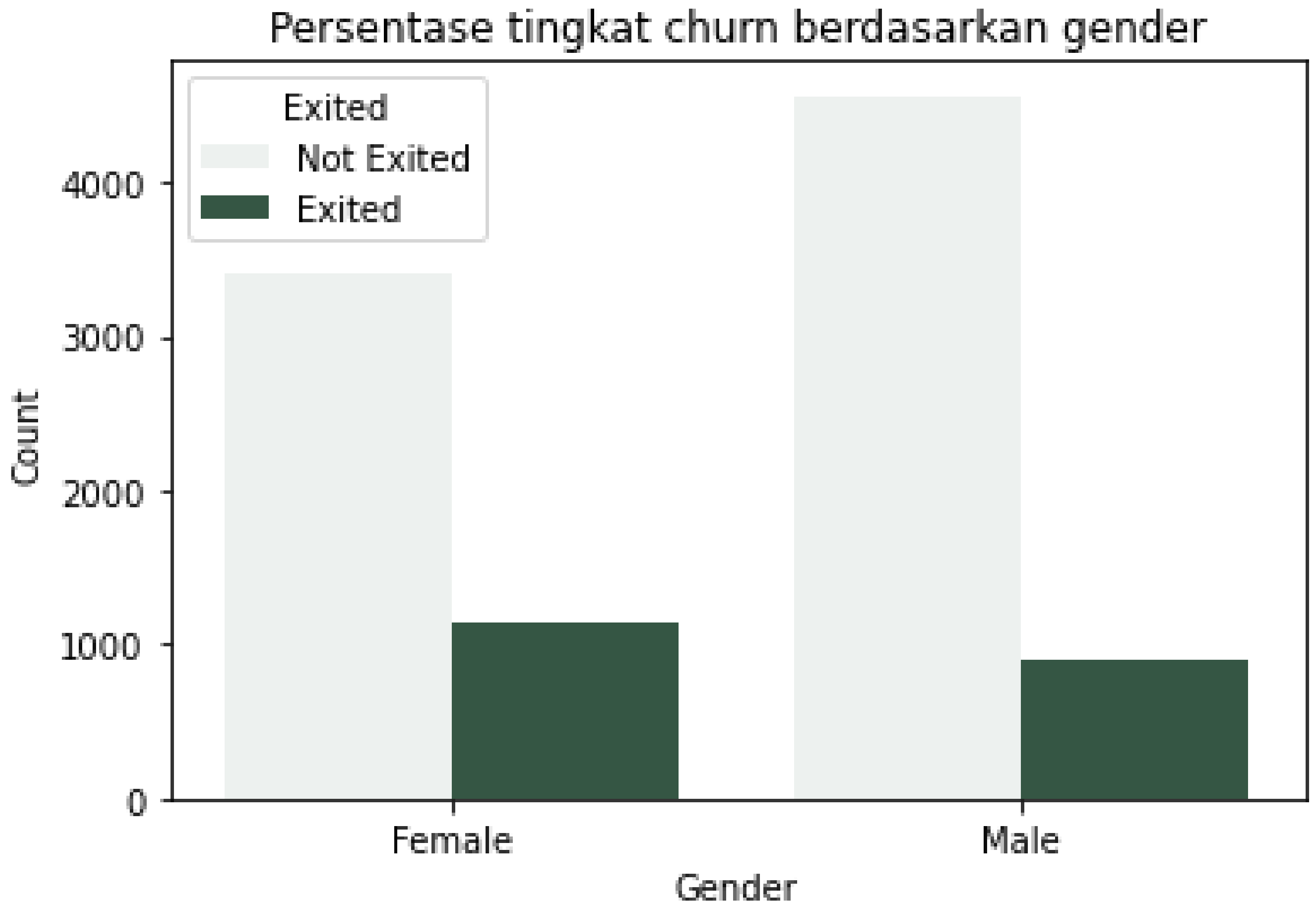
Dataset memilik 14 kolom dan 10.000 baris. Target dari dataset ini adalah Exited.

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.

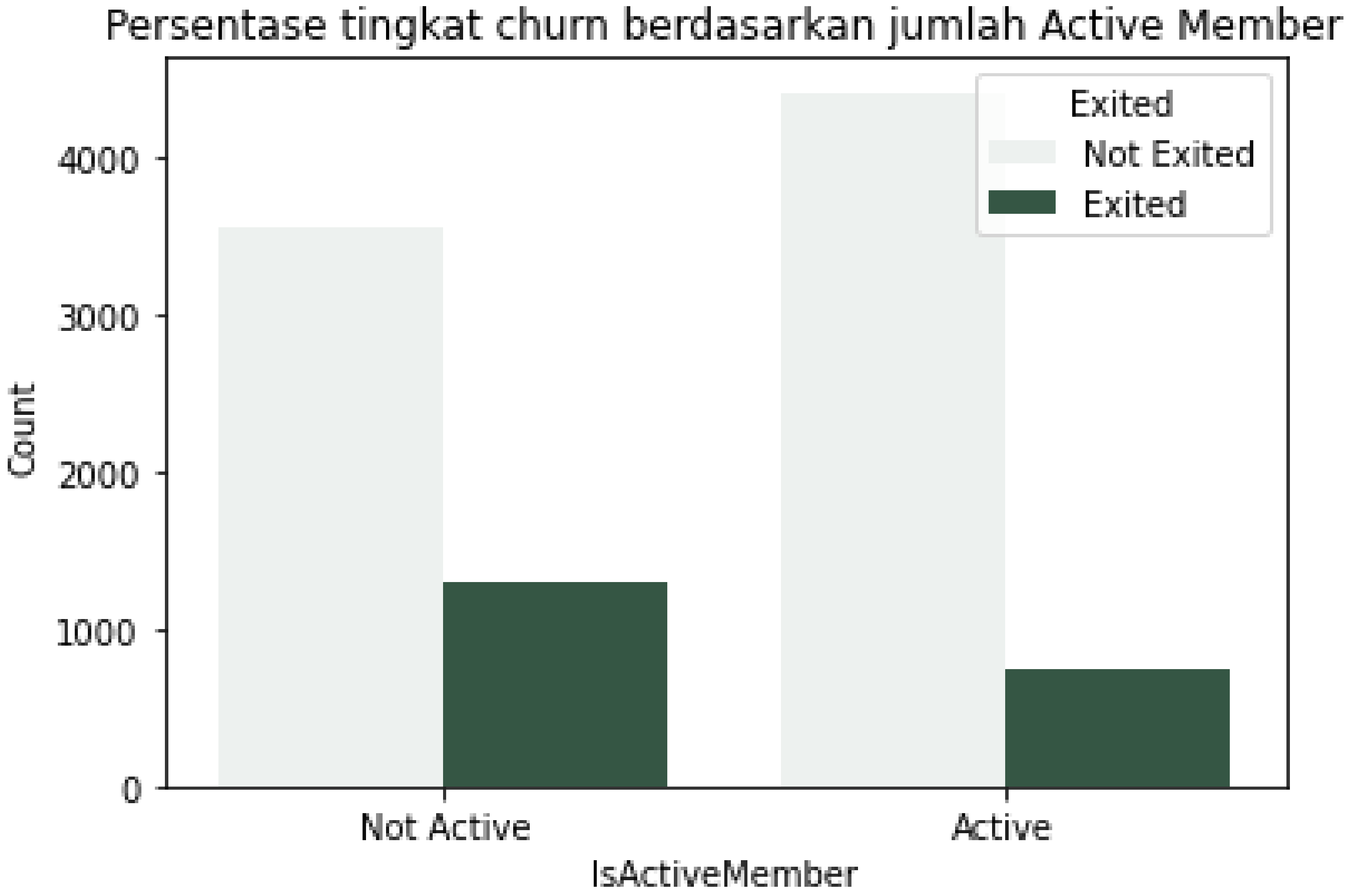
EXPLORATORY DATA ANALYSIS



Nasabah yang berada di wilayah Jerman akan lebih cenderung melakukan churn.

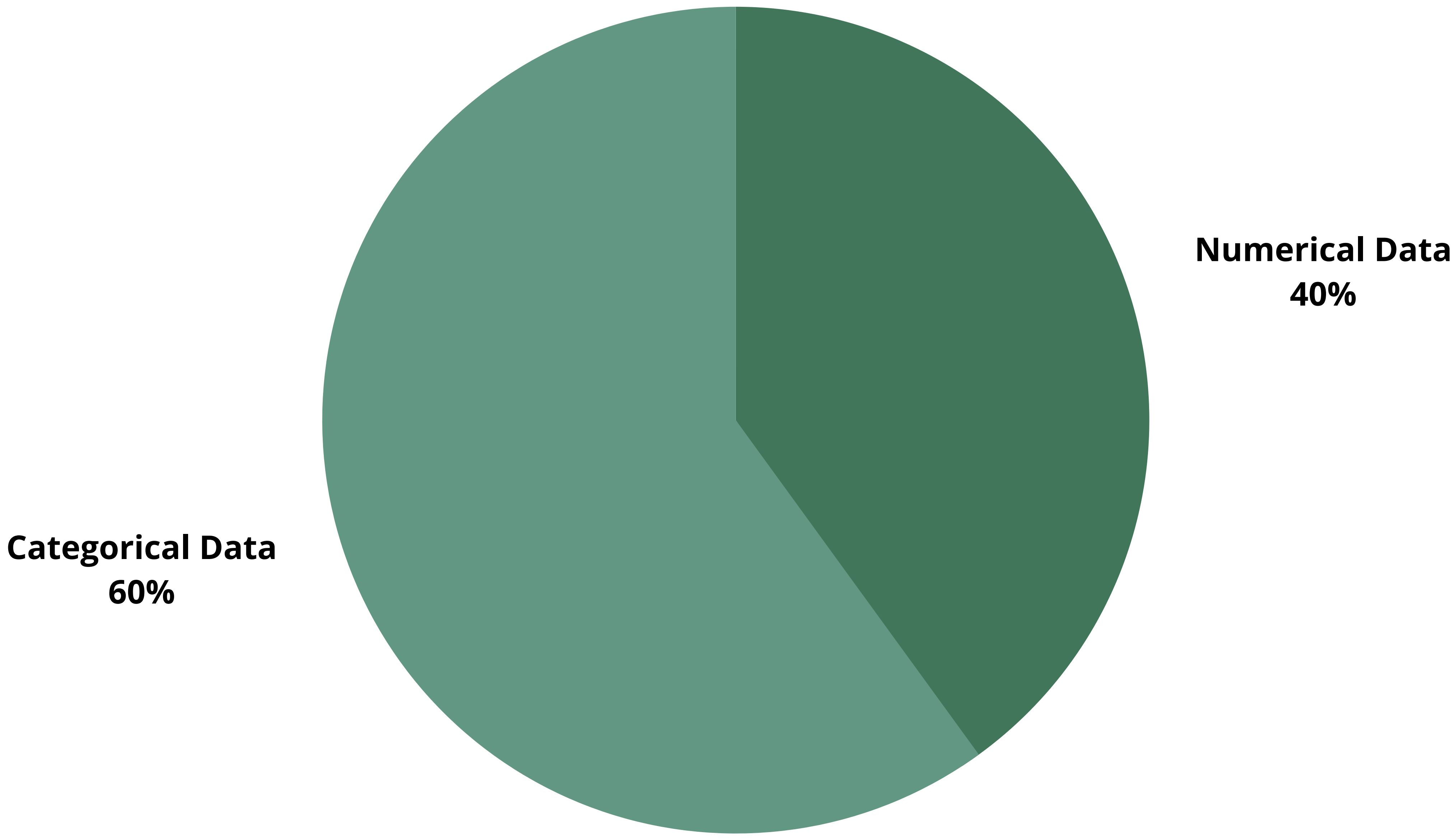


Nasabah wanita akan lebih cenderung melakukan churn.



Nasabah yang tidak aktif akan lebih 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

Target

- Exited

DATA PRE-PROCESSING: IN DETAIL

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

EVALUATION METRICS

AUC

AUC Score

P

Precision Score

Kami menggunakan Evaluation Metrics AUC Score dan Precision Score.

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

Precision Score juga dipilih karena Metric tersebut dapat meminimalisir False Positive, sehingga dapat meminimalisir jumlah nasabah yang terdeteksi churn yang dalam real casenya tidak churn.

MODEL EVALUATION

Model	Accuracy	Precision	Recall	F1	Train Score	Test Score
Logistic Regression	0.82	0.58	0.27	0.37	0.78	0.79
K-nearest Neighbor	0.82	0.56	0.36	0.44	0.79	0.83
Decision Tree	0.85	0.84	0.29	0.43	0.80	0.84
Random Forest Classifier	**0.85**	**0.82**	**0.27**	**0.40**	**0.84**	**0.87**
Adaboost Classifier	0.84	0.69	0.35	0.47	0.84	0.86

Berdasarkan semua model yang telah di-*tune,* Model yang memiliki hasil terbaik adalah **Random Forest Classifier.**