

PREDICTING CHURN FOR BANK CUSTOMERS

7A-OCTAGRAM TEAM



OCTAGRAM'S TEAM :



KETUA

Cicilia Rumanat Siahaan



ANGGOTA

Yulia Rohmawati



ANGGOTA

Arief Sulistyo Wibowo



ANGGOTA

Wilson Firmando Manurung



ANGGOTA

Yansen Cristopel Sinaga



ANGGOTA

Marifatul Hasanah



ANGGOTA

Alfaroqih Swandi



ANGGOTA

Rahmat Diza Ramadhan

TABLE OF CONTENTS



01 Background

02 Exploratory Data Analysis & Insight

03 Pre-processing Data

04 Modelling (Classification & Clustering)

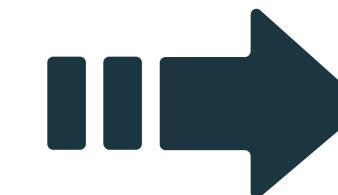
05 Recommendation & Simulation

BACKGROUND

PROBLEM STATEMENT & GOAL



Jumlah nasabah bank yang churn meningkat



Pendapatan perusahaan bank menurun

GOAL!



Mempertahankan customer atau nasabah bank agar tidak jadi churn (churn rate menurun) sehingga **pendapatan perusahaan perbankan dapat meningkat kembali.**

OBJECTIVES & BUSINESS METRIC



OBJECTIVES

- Mengidentifikasi **faktor-faktor** yang memengaruhi nasabah bank *churn*
- Melakukan **segmentasi** nasabah berdasarkan atribut yang tersedia
- Mengembangkan **model Machine Learning** untuk memprediksi nasabah berpotensi *churn*.

CHURN RATE

Churn rate dalam data ini merujuk pada **tingkat nasabah bank yang berhenti menggunakan layanan**. Churn rate ini bertujuan untuk memberikan wawasan tentang seberapa baik bank **mempertahankan nasabah dan mencegah mereka beralih** ke pesaing. Semakin rendah Churn Rate, semakin baik kinerja bisnis dalam mempertahankan pelanggan.

PROJECT ASSIGNMENT



Membuat model machine learning untuk prediksi nasabah bank yang berpotensi churn

Membuat model machine learning untuk mengetahui behaviour profil nasabah bank

Benefit:

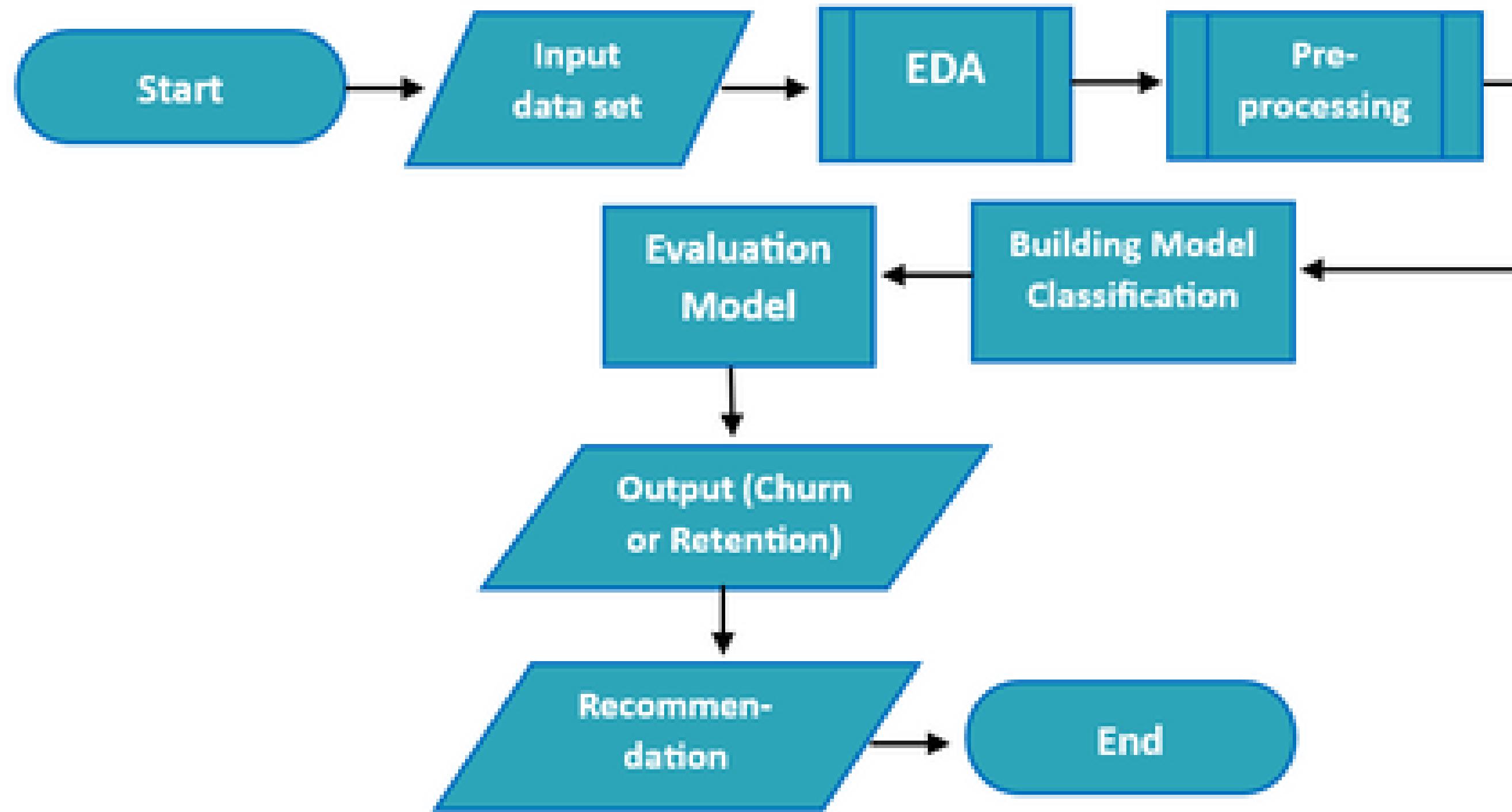
Lebih Efektif & Efisien

Rekomendasi Bisnis yang Lebih Tepat

Pendapatan Meningkat

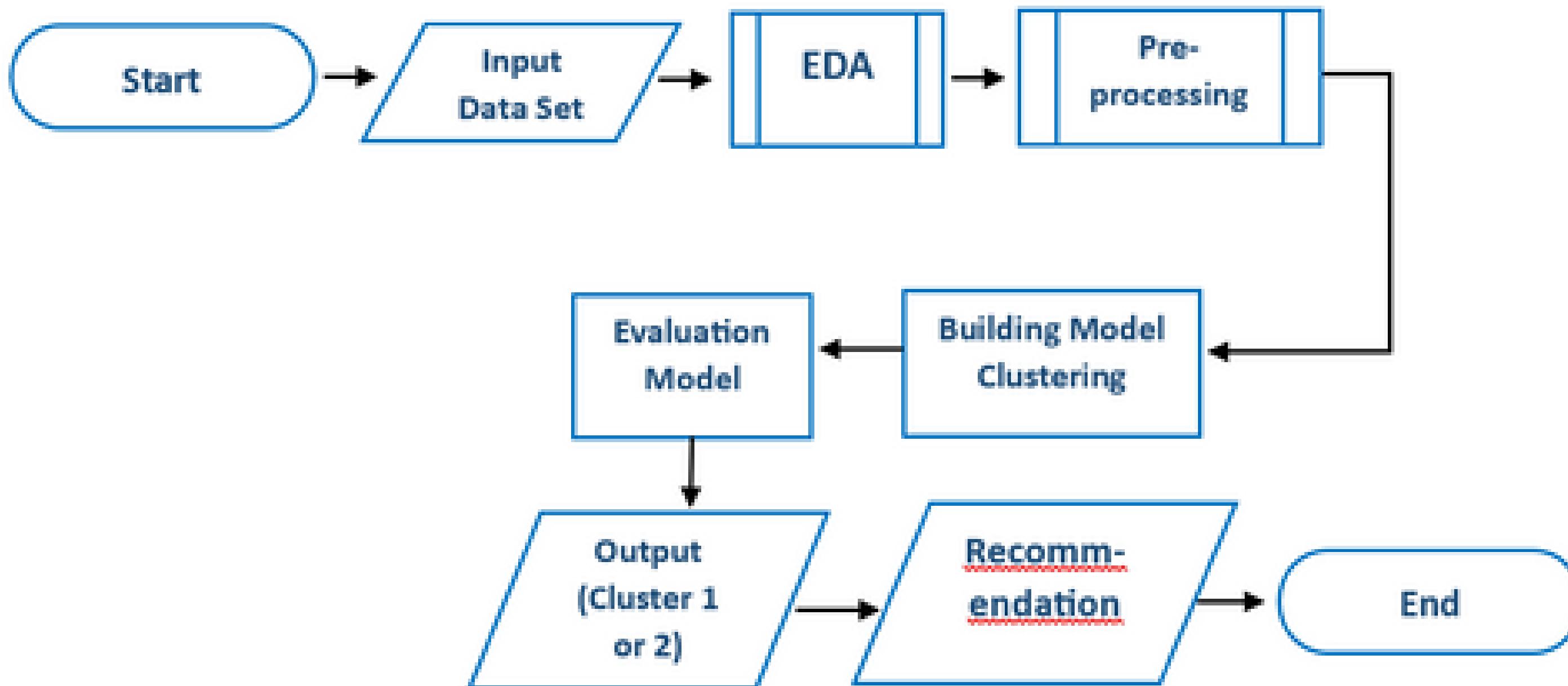
FLOWCHART

CLASSIFICATION



FLOWCHART

CLUSTERING



EDA & INSIGHT

- EXPLORATORY DATA ANALYSIS
- INSIGHT



DATASET

Sumber: Kaggle (Predicting Churn For Bank Customers)

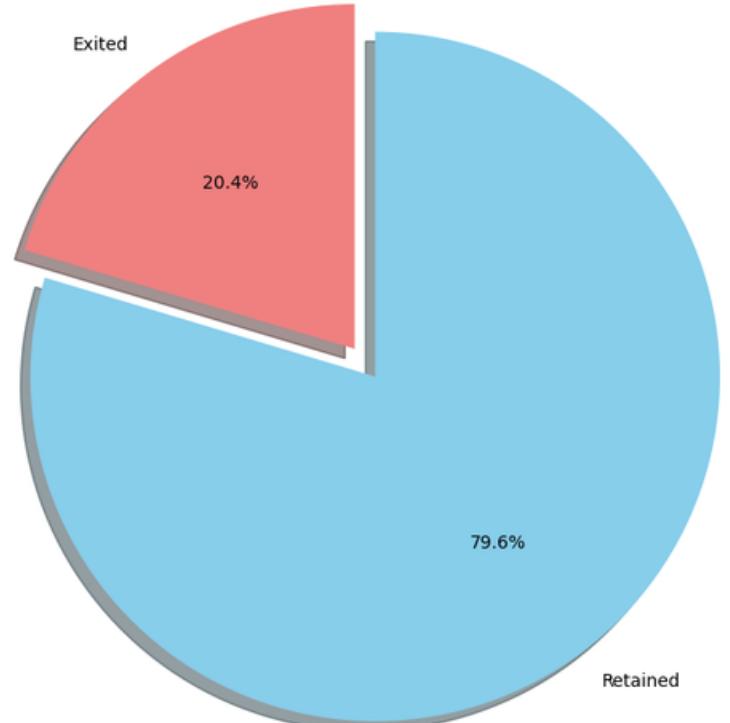
Dataset memiliki 10.000 baris data yang terdiri dari profil nasabah bank yang berhubungan dengan potensi churn-nya dengan atribut kolom sebagai berikut:

Customer ID	Merepresentasikan identifikasi unik nasabah
Surname	Merepresentasikan last name nasabah
CreditScore	Merepresentasikan angka yang mengestimasi kemampuan nasabah dalam melunasi utangnya tepat waktu
Geography	Merepresentasikan wilayah nasabah
Gender	Merepresentasikan jenis kelamin nasabah (male or female)
Age	Merepresentasikan usia nasabah
Tenure	Merepresentasikan jangka waktu pinjaman nasabah
Balance	Merepresentasikan saldo pada akun nasabah
NumOfProducts	Merepresentasikan jumlah produk yang digunakan atau dimiliki oleh nasabah
HasCrCard	Merepresentasikan kepemilikan nasabah akan kartu kredit
IsActiveMember	Merepresentasikan apakah member nasabah aktif atau tidak
EstimatedSalary	Merepresentasikan estimasi gaji nasabah
Exited	Merepresentasikan nasabah churn atau tidak oleh nasabah



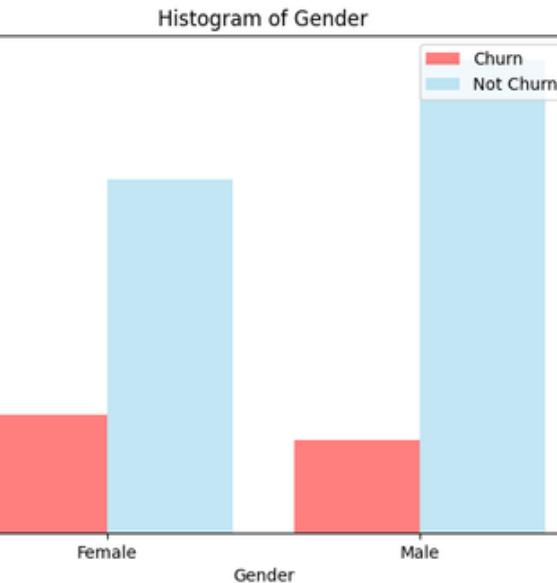
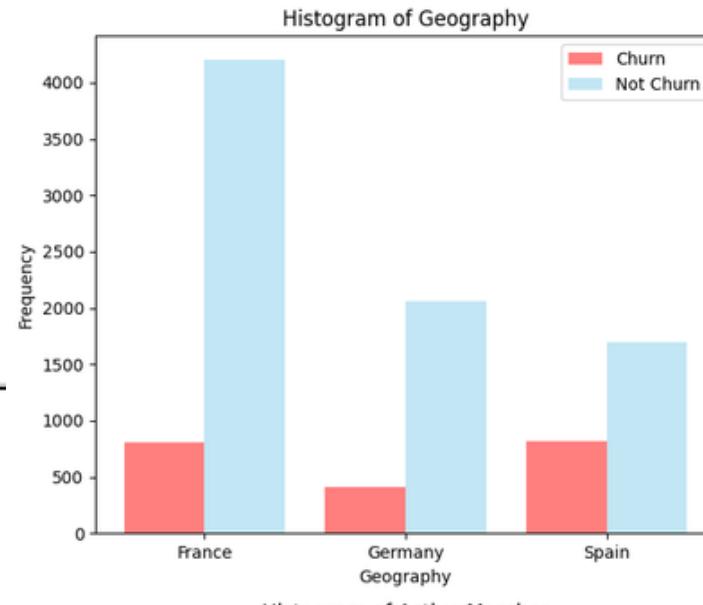
EDA & INSIGHT

Proporsi Customer Churn dan Tidak Churn

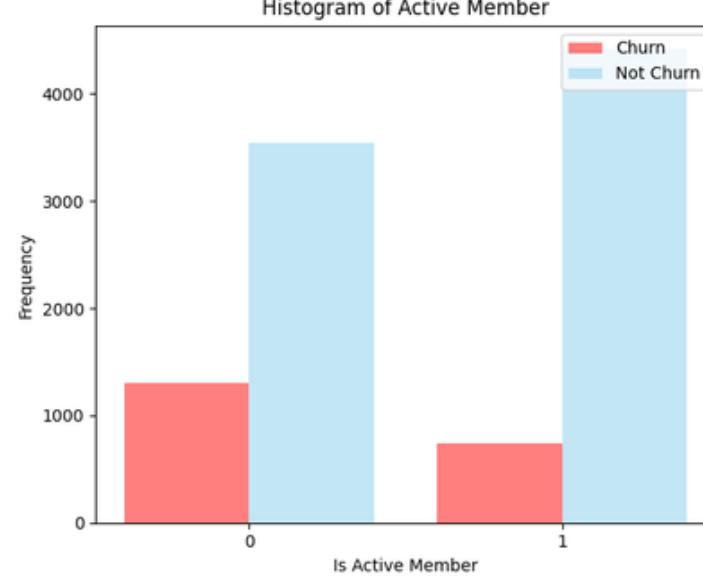


Berdasarkan *pie chart* di atas, proporsi nasabah atau customer yang **tidak churn lebih besar** dibandingkan nasabah yang churn.

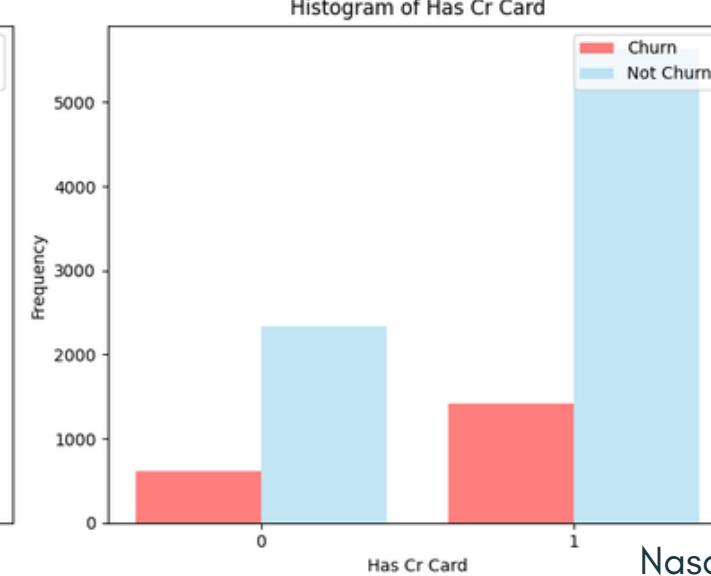
Nasabah dari **France** dan **Spain** memiliki **tingkat churn lebih tinggi** dibanding **Germany**.
Nasabah dari France angka "tidak churn"-nya paling tinggi.



Nasabah **perempuan** cenderung memiliki **tingkat churn lebih tinggi** dibanding nasabah laki-laki



Nasabah yang tidak aktif memiliki **tingkat churn lebih tinggi** dibanding nasabah yang aktif.



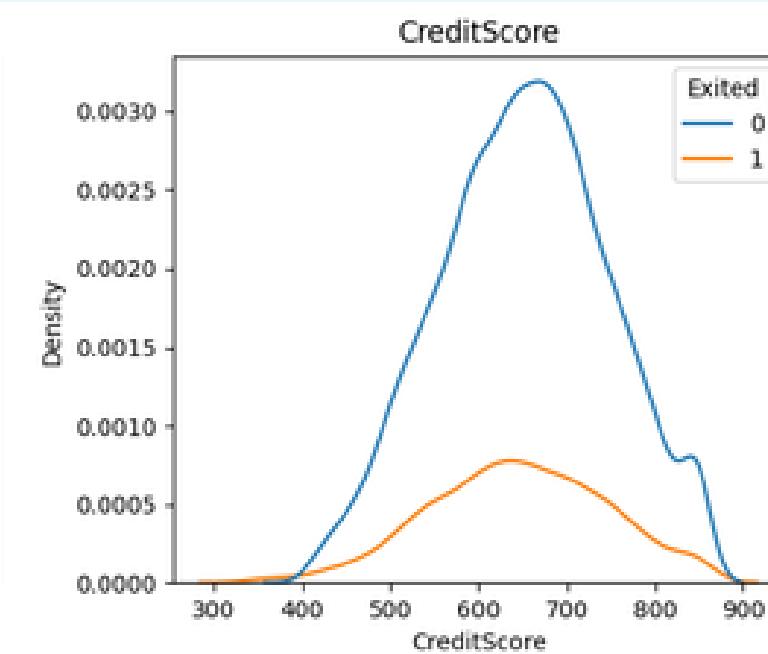
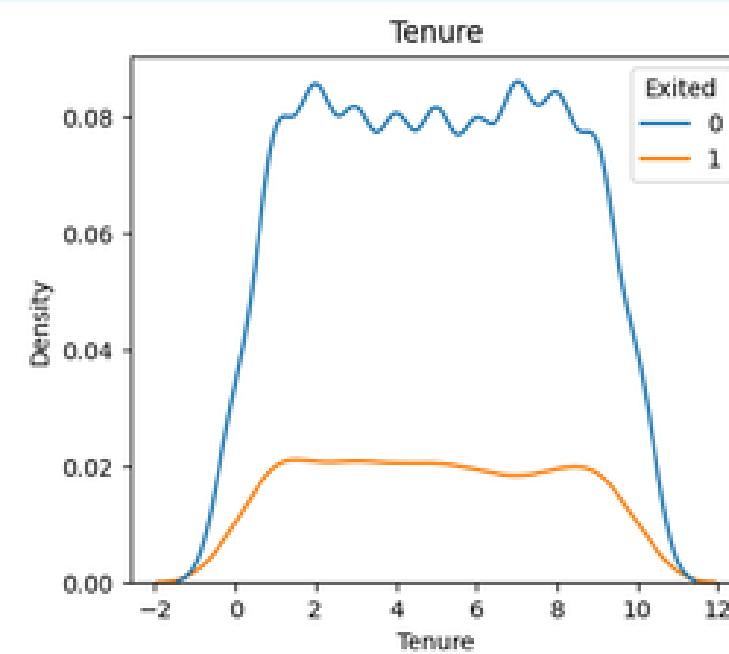
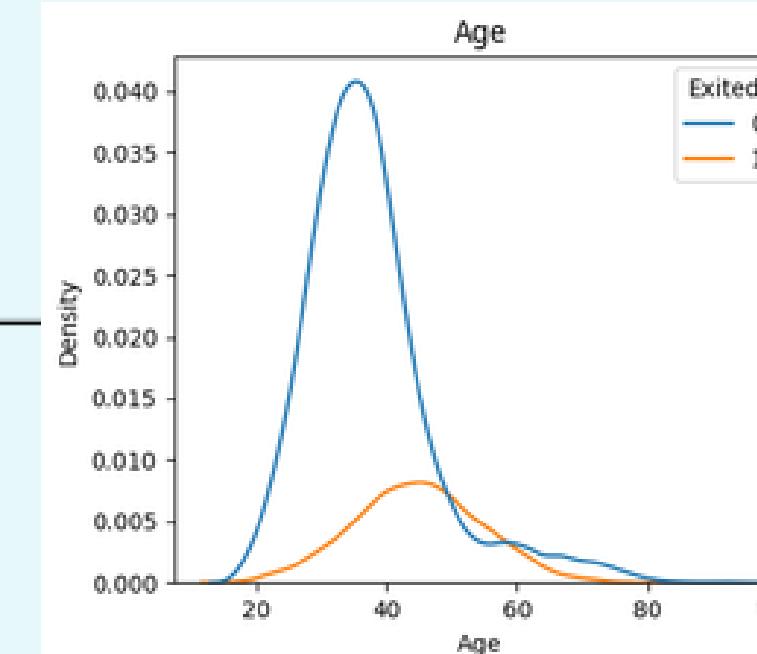
Nasabah yang memiliki credit card **tingkat churn-nya lebih tinggi** dibanding nasabah yang tidak memiliki credit card. Akan tetapi, nasabah yang memiliki credit card juga merupakan nasabah **yang loyal** kepada bank.

Keterangan:

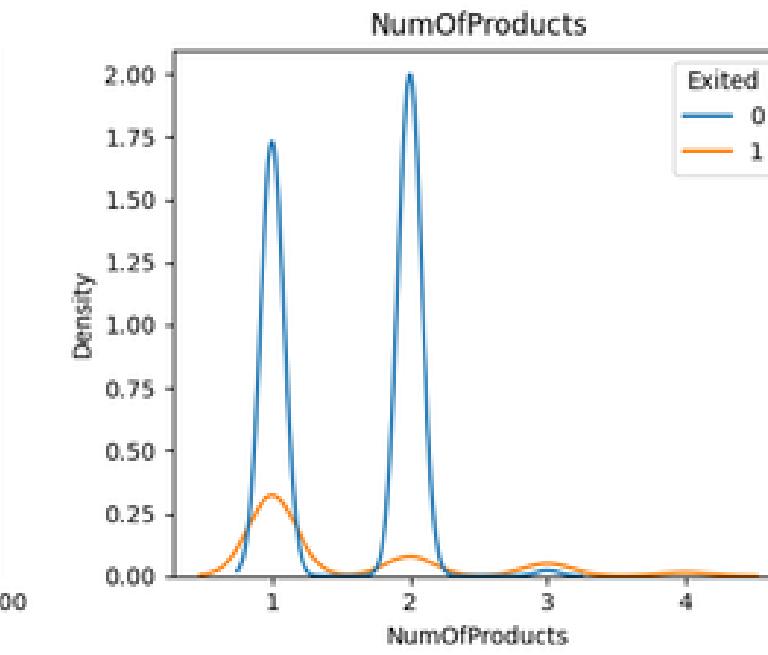
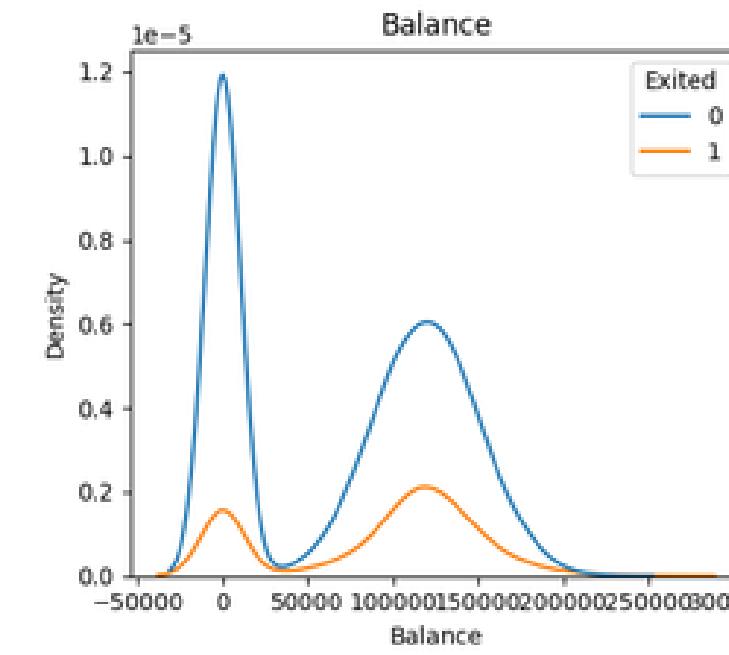
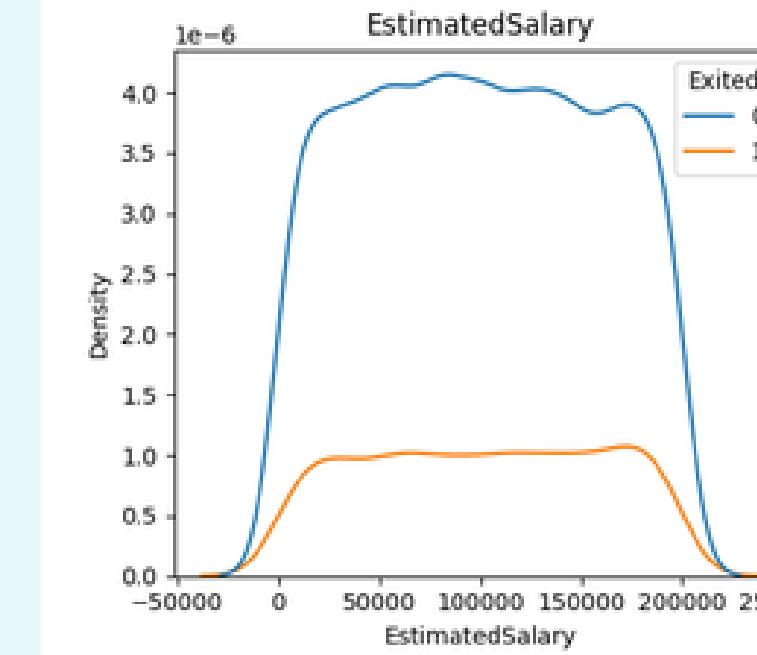
Angka 0 menunjukkan: (1) nasabah tidak aktif, (2) nasabah tidak memiliki kartu kredit
Angka 1 menunjukkan: (1) nasabah aktif, (2) nasabah memiliki kartu kredit

EDA & INSIGHT

Nasabah yang **churn** kebanyakan ada pada usia **40-50** tahun.



Nasabah yang **churn** kebanyakan memiliki credit score antara **600 - 700**.



Semakin banyak num of products, semakin **rendah** tingkat churn.

Nasabah **paling banyak churn** adalah nasabah yang memiliki balance **100.000 - 150.000**.

PRE-PROCESSING



Classification:

PRE-PROCESSING

Handling Missing Values, Invalid Values

- Dataset "Churn Bank Customers" **tidak terdapat missing values.**
- Dataset "Churn Bank Customers" memiliki **data type & values yang valid.**

Handling Duplicated & Redundant

- Dataset "Churn Bank Customers" **tidak terdapat duplicated data.**
- Dataset "Churn Bank Customers" memiliki **redundant data.**

Handling Outliers

- Dataset "Churn Bank Customers" **memiliki outlier** pada kolom "**Age**", "**Credit Score**" dan "**Num of Products**".
- **Outlier tetap digunakan** tanpa dihandle agar informasi tidak ada yang hilang.

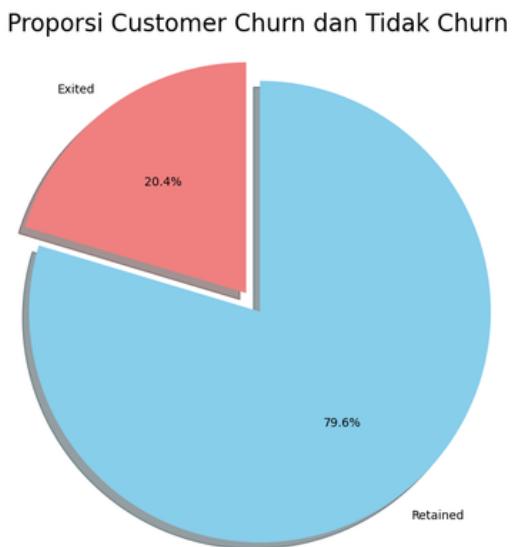
Remove Features & Feature Encoding

- Features yang tidak berhubungan pada Dataset "Churn Bank Customers" **dihapus** yaitu "**Row of Number**", "**Surname**", "**Customer Id**".
- **Label Encoding** pada "**Gender**" dengan **0: Female** dan **1: Male**.
- **OHE** pada "**Geography**"



PRE-PROCESSING : CLASSIFICATION

HANDLING CLASS IMBALANCE



Feature target “**Exited**” pada Dataset “Churn Bank Customers” mengalami **class imbalance** dengan proporsi antara customer churn dan retained adalah **20.4%** dan **70.9%**.

- Sebelum class imbalance di-handle, dataset “Churn Bank Customer” di-**split** menjadi data train dan data test dengan proporsi **80:20**.
- Handling Class Imbalance digunakan metode **SMOTE (Synthetic Minority Over-sampling Technique)**.

FEATURE TRANFORMATION

Metode Feature Transformation adalah **Robust Scalling**.

Robust Scalling dipilih sebagai metode feature transformation pada modelling classification karena data yang digunakan mempertahankan outlier. Oleh karena itu, perlu metode scalling yang robust terhadap outlier.

Clustering:

PRE-PROCESSING

Handling Missing Values, Invalid Values

- Dataset “Churn Bank Customers” **tidak terdapat missing values.**
- Dataset “Churn Bank Customers” memiliki **data type & values yang valid.**

Handling Duplicated & Redundant

- Dataset “Churn Bank Customers” **tidak terdapat duplicated data.**
- Dataset “Churn Bank Customers” memiliki **redundant data.**

Handling Outliers

- Dataset “Churn Bank Customers” **memiliki outlier** pada kolom **“Age”, “Credit Score”** dan **“Num of Products”**

 1. **Outlier tetap digunakan** tanpa dihandle.
 2. **Outlier di-handle dengan metode IQR.**
 3. Data Outlier di-drop atau **dihapus.**

Remove Features & Feature Encoding

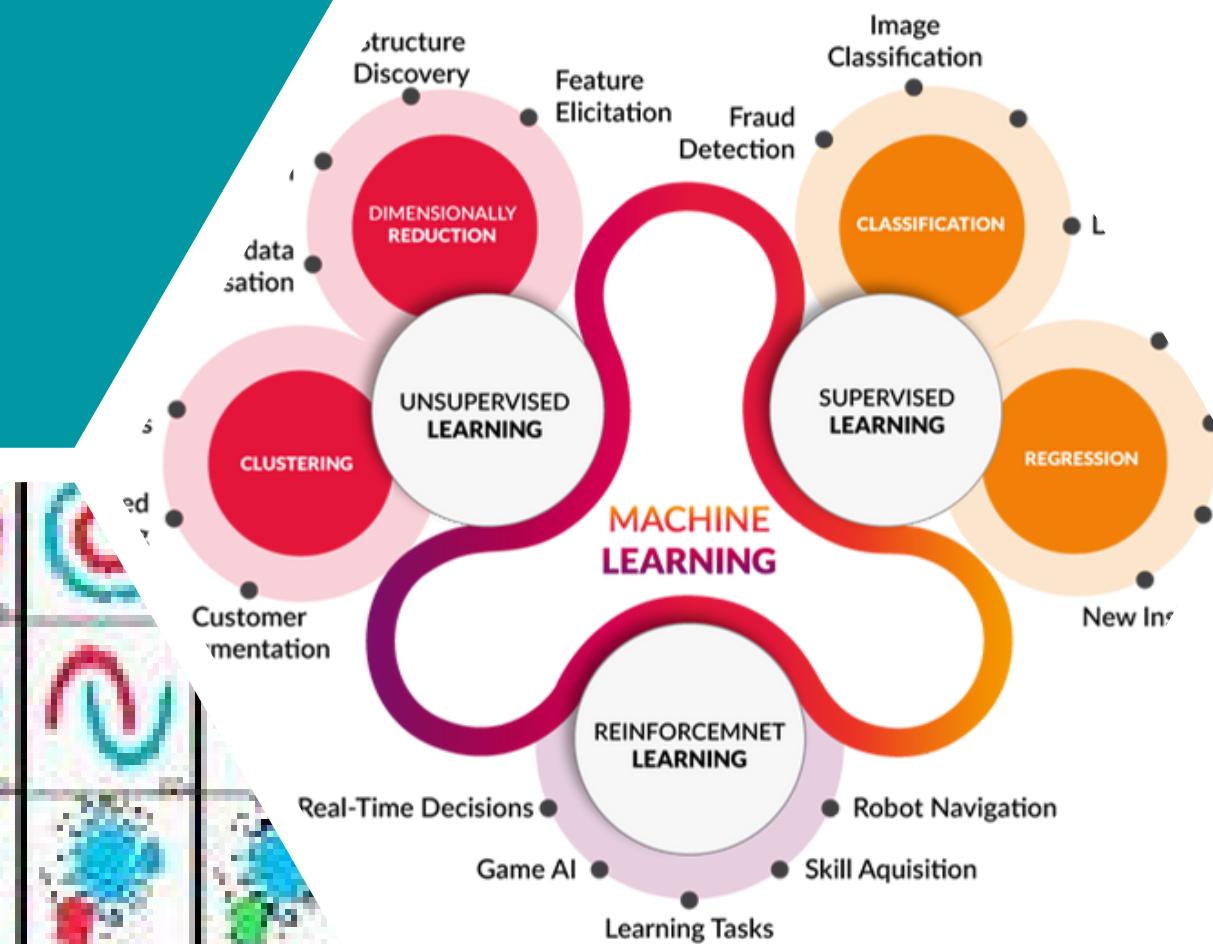
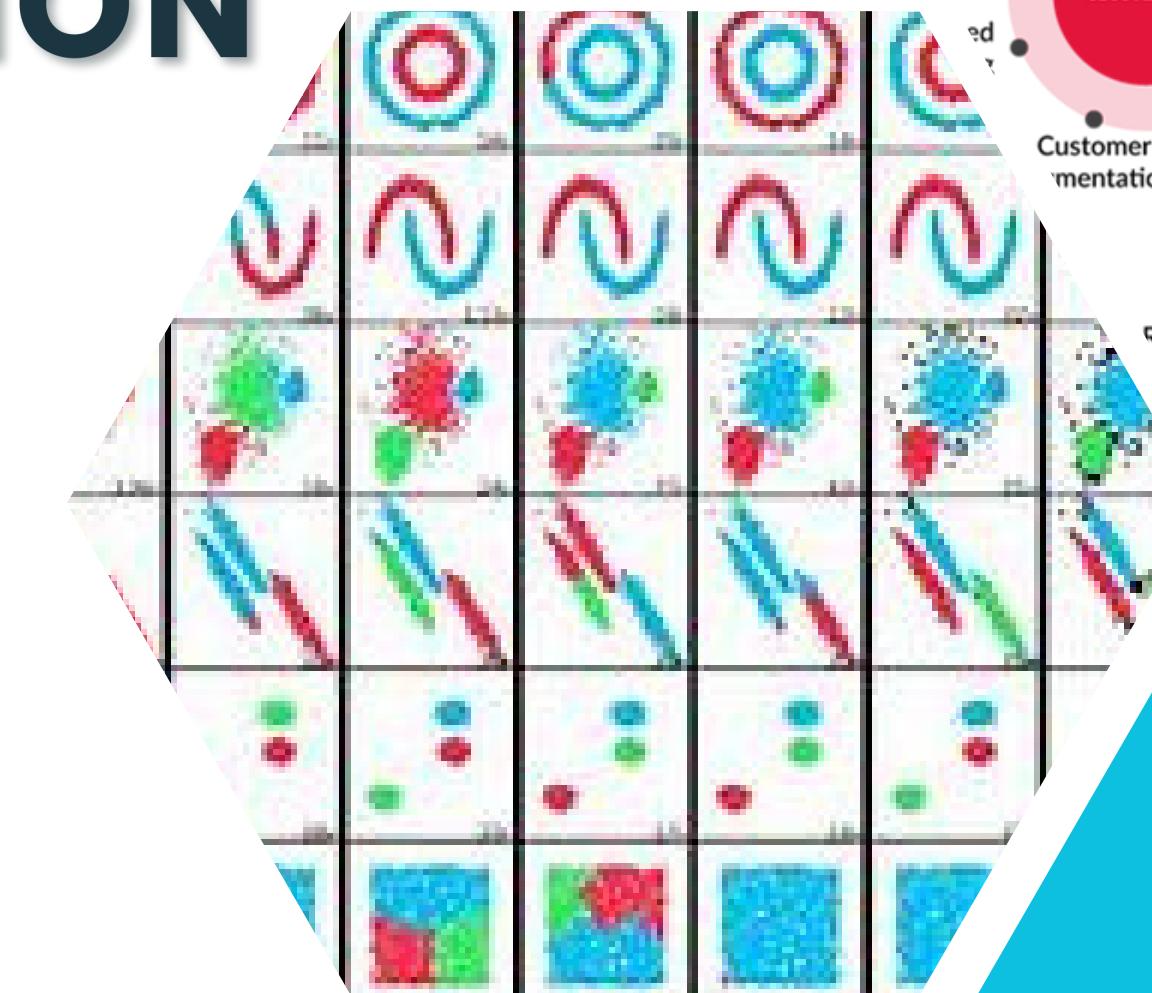
- Features yang tidak berhubungan pada Dataset “Churn Bank Customers” **dihapus** yaitu **“Row of Number”, “Surname”, “Customer Id”.**
- **Label Encoding** pada **“Gender”** dengan **0: Female dan 1: Male.**
- **OHE** pada **“Geography”**

Feature Transformation

- Dataset yang masih mempertahankan outlier dilakukan feature transformation dengan metode **Robust Scalling**.
- Dataset yang outliernya di-handle dengan IQR atau yang dihapus dilakukan feature transformation dengan metode **Standarization**.

MODELLING

- CLASSIFICATION
- CLUSTERING



Modelling:

CLASSIFICATION

Eval Metrics \ Classifier	Random Forest	Decision Tree	SVM	Naive Bayes
Accuracy	0.91	0.83	0.87	0.81
Precision	0.93	0.83	0.88	0.80
Recall	0.89	0.82	0.84	0.80
F1-Score	0.91	0.83	0.86	0.80
AUC	0.97	0.91	0.94	0.88

- Model terbaik : **RANDOM FOREST**
- **RANDOM FOREST** memiliki nilai F1-Score tertinggi diantara model lainnya. F1-Score adalah rata-rata harmonis antara presisi (precision) dan recall (sensitivitas).
- Pemilihan **F1-Score** sebagai evaluation metric karena dataset mengalami ketidakseimbangan kelas sehingga perlu untuk mengidentifikasi nasabah yang akan churn dengan akurasi yang tinggi dan meminimalkan jumlah nasabah yang salah diklasifikasikan sebagai churn.

Modelling:

CLASSIFICATION

Feature	Feature Importance
Age	0.2252352
Num of Products	0.18396855
Is Active Member	0.09250019
Balance	0.09296625
Estimated Salary	0.084213
Credit Score	0.07821962
Tenure	0.0627265
France	0.05348106
Gender	0.04822749
Spain	0.04100564
Germany	0.02391431
Has Cr Card	0.01354218

- **Feature Importance** secara berurutan dari yang tertinggi ke terendah beserta coefisien-nya adalah seperti pada gambar di samping.
- 6 Feature yang memiliki importance tertinggi adalah **Age, Num of products, Is Active Member, Balance, Estimated Salary, Credit Score dan Tenure.**

Modelling:

CLUSTERING (WITH OUTLIER HANDLE)

Clustering	K-Medoids	DBSCAN	Complete Linkage	GMM
Eval Metrics				
Silhouette Score	0.0888	-0.2742	0.0897	0.1390
Davies-Bouldin	2.2371	1.7395	2.5953	2.0678
Calinski-Harabasz	979.6441	94.4534	921.0601	1575.6142

Model Clustering terbaik dengan menggunakan data outlier adalah model GMM (Gaussian Mixture Model) dengan Silhouette Score terbesar.

Modelling:

CLUSTERING (OUTLIER HANDLE WITH IQR)

Eval Metrics	Clustering	K-Means	K-Prototype
Silhouette Score		0.1556	0,1557
Davies-Bouldin		1.9757	1.9740
Calinski-Harabasz		1458.3229	1454.9214

Model Clustering terbaik dengan data outlier yang dihandle dengan IQR adalah model K-Prototype dengan Silhouette Score terbesar.

Modelling:

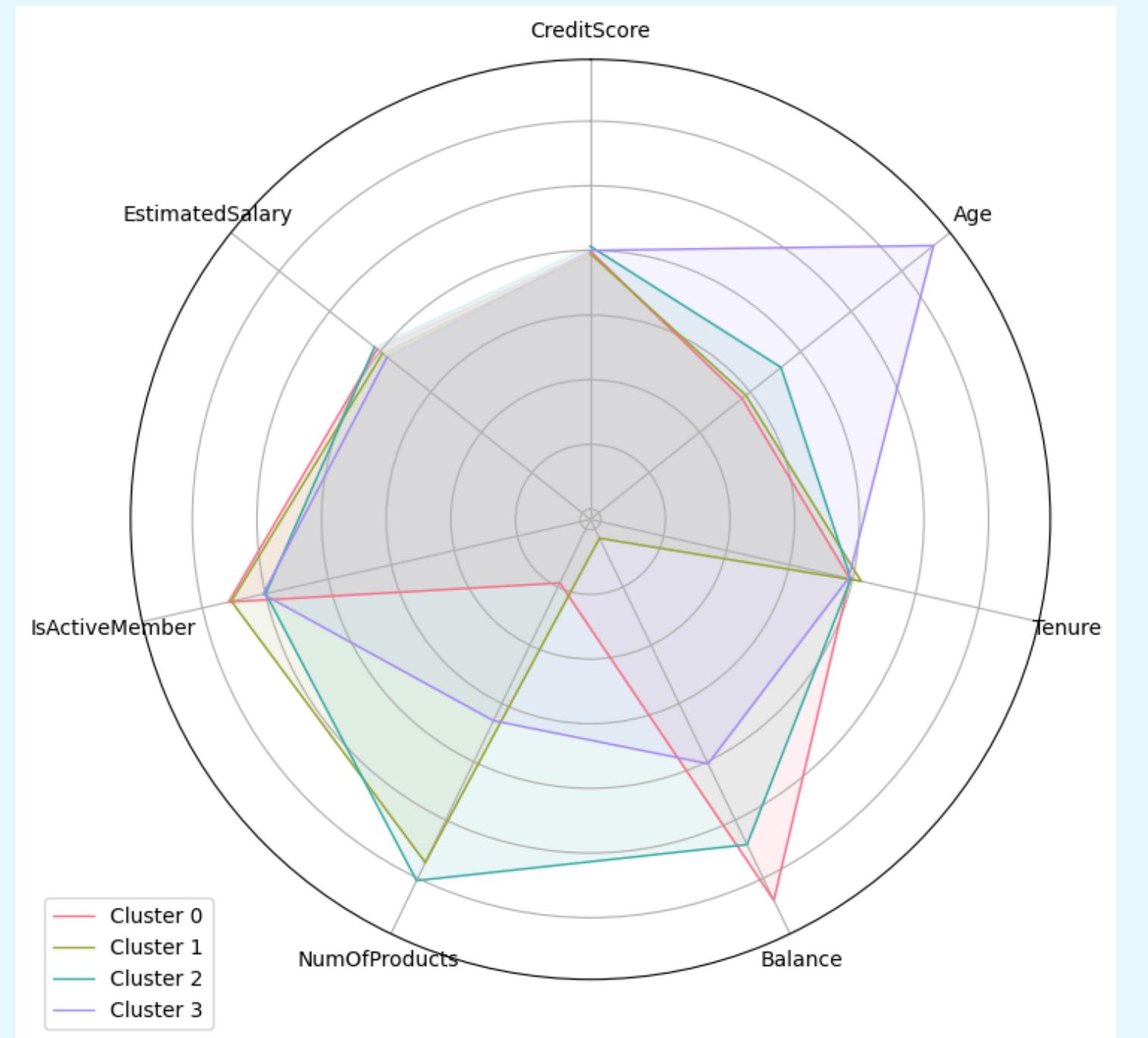
CLUSTERING (DROP OUTLIER)

Eval Metrics	Clustering	K-Means	K-Prototype
Silhouette Score		0.1378	0,1366
Davies-Bouldin		2.2109	2.2190
Calinski-Harabasz		1513.0909	1503.2883

Model Clustering terbaik dengan data outlier yang dihandle dengan IQR adalah model K-Means dengan Silhouette Score terbesar.

Modelling:

CLUSTERING -> MODEL TERBAIK



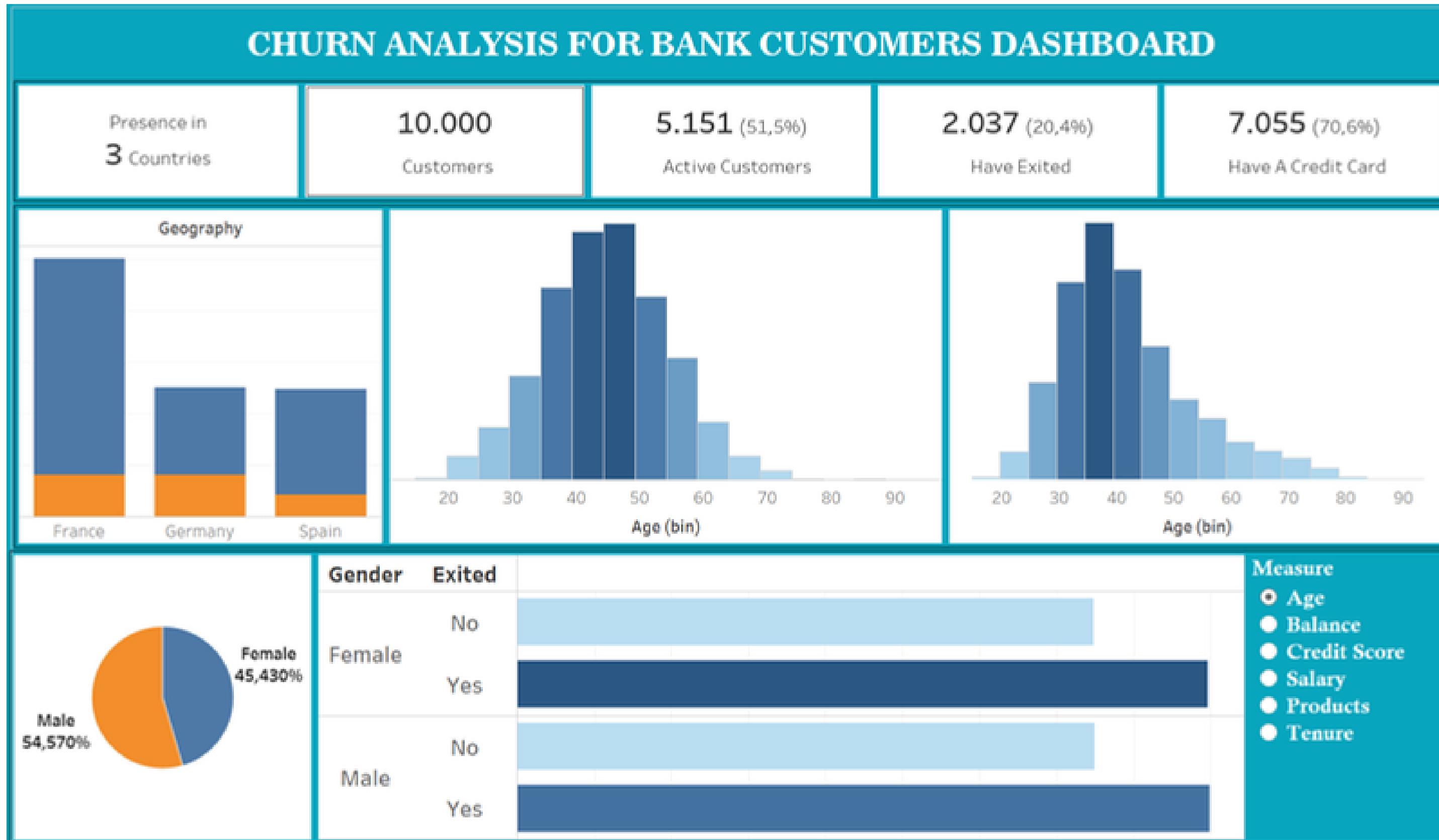
- Model terbaik : **K-Prototype** dengan data outlier di-handle dengan IQR.
- **K-Prototype** memiliki nilai **Silhouette Score** tertinggi di antara model lainnya. Silhouette Score adalah metrik evaluasi yang digunakan untuk mengukur sejauh mana objek dalam cluster serupa satu sama lain dan berbeda dari objek dalam cluster lain.

INTERPRETASI CLUSTER

- Pada **Cluster 0** profil nasabah bank adalah nasabah yang aktif, dengan balance paling tinggi, tenure lebih lama, dan memiliki credit score serta estimated salary yang tinggi tinggi. Cluster ini memiliki rentang usia yang cenderung beragam. Sementara, untuk kepemilikan produk lebih rendah.
- Pada **Cluster 1** profil nasabah bank adalah nasabah aktif dengan kepemilikan jumlah produk yang tinggi, tenure sedang, estimated salary credit score sedang. Selain itu, nasabah pada cluster ini memiliki balance rendah dan usia nasabah cenderung beragam.
- Pada **Cluster 2** profil nasabah adalah nasabah pasif yang kepemilikan produknya tinggi dibandingkan dengan cluster lain. Nasabah pada cluster ini juga memiliki balance lebih tinggi dibanding cluster 1 dan 3, dengan tenure sedang, estimated salary dan credit score sedang. Untuk usia, nasabah pada cluster ini cenderung beragam.
- Pada **Cluster 3** nasabah cenderung pasif. Nasabah pada cluster ini memiliki usia lebih tua. Kepemilikan produk dan balance sedang. Sementara itu, lama tenure, estimated salary, dan credit score juga sedang.

Output Tambahan:

DASHBOARD (TABLEAU)



RECOMENDATION

BUSINESS RECOMENDATION SIMULATION

Business Recommendation



- Meningkatkan **komunikasi** antara bank dengan nasabah sehingga dapat meningkatkan keaktifan dan loyalitas nasabah.
- Meningkatkan **retensi pelanggan** dengan menawarkan insentif atau promosi pada nasabah yang memiliki banyak produk.
- Melakukan **survey kepuasan** atas produk yang dimiliki nasabah yang mencakup evaluasi produk, sehingga bank dapat mengembangkan produk sesuai dengan preferensi nasabah.
- Memberikan **limit kredit yang panjang** bagi nasabah yang membayar tepat waktu.

Business

Recommendation



Layanan Cepatt..



- Memberi **reward** kepada nasabah yang setia agar memperkuat loyalitas dan memotivasi nasabah untuk tetap setia
- Menghadirkan **produk dan layanan sesuai umur** sehingga meningkatkan efektivitas pemasaran dan memastikan bahwa bank secara efisien memenuhi harapan nasabah.
- Melakukan **promosi beragam produk bank** serta manfaatnya sehingga dapat meningkatkan jumlah penggunaan produk bank oleh nasabah.
- Memberikan **layanan yang lebih mudah dan cepat**, seperti layanan perbankan digital yang lebih canggih dan efisien.

Business Recommendation



Setelah mencoba melakukan rekomendasi:

- **Evaluasi Kebijakan Penanganan Churn**
Berdasarkan hasil model pihak bank bisa mengevaluasi kebijakan penindakan bagi nasabah yang cenderung akan churn.
- **Menyempurnakan Strategi Pemasaran**
Berdasarkan hasil evaluasi kebijakan, Bank dapat menyempurnakan strategi pemasaran terkait nasabah yang memiliki kecenderungan akan churn.

SIMULATION

Simulation : CHURN RATE

Before Modelling

20.4%



TURUN

After Modelling

5.3%

15.1%

Dengan modelling, Perusahaan Bank berpotensi menurunkan churn rate sebesar 15,1%. Oleh karena itu, revenue perusahaan tidak akan turun.



TERIMA KASIH