



Credit Card Customer Attrition Prediction on X Company

Prasetyo Sukma Raharjo

DS32B

Overview Project

<u>Deskripsi</u>	memprediksi pelanggan kartu kredit yang berpotensi berhenti menggunakan layanan (attrition/churn)
<u>Tujuan</u>	<ul style="list-style-type: none">• Mengidentifikasi faktor-faktor utama yang menyebabkan pelanggan berhenti menggunakan kartu kredit.• Membangun model prediktif untuk mengklasifikasikan pelanggan yang berisiko attrition.• Memberikan rekomendasi bisnis untuk retensi pelanggan
<u>Metode & Analysis</u>	<ul style="list-style-type: none">• EDA (Exploratory Data Analysis), analisis statistik deskriptif, distribusi, dan korelasi, serta visualisasi.• Preprocessing Data, handling imbalanced dataset dan transformasi data (scaling/encoding).• Pemodelan Machine Learning, percobaan model dan evaluasi metrik
<u>Results</u>	<ul style="list-style-type: none">• Fitur paling berpengaruh terhadap attrition pelanggan• Model terbaik dengan akurasi dan recall tinggi• Insight bisnis yang bisa direkomendasikan untuk strategi retensi pelanggan

Table of contents

01 Datasets

This Data Consists 2 data types:

categorical (marital status, card category, income category, attrition flag) and

numericals (total transaction amount, transaction count, months on book, utilization ratio).

03 Machine Learning

Building model with high precision and recall for predicting attrition customer for next time

02 EDA

EDA is a critical step to ensures data is clean and relevant, also helps select optimal features

04 Business Insight

A recommendation and actionable strategy for customer retention

Workflow

Business Understanding

- Identify the problem,
- find the purpose, and
- define the actionable output we desire

EDA

- Find the pattern and variable relation from the dataset,
- check the distribution, outlier or missing values

Preprocessing Data

- Label Encoding
- Standardization/Normalization
- Feature Selection

MML

- Baseline Model,
- Model Interpretation,
- Evaluation model

Business Insight



01

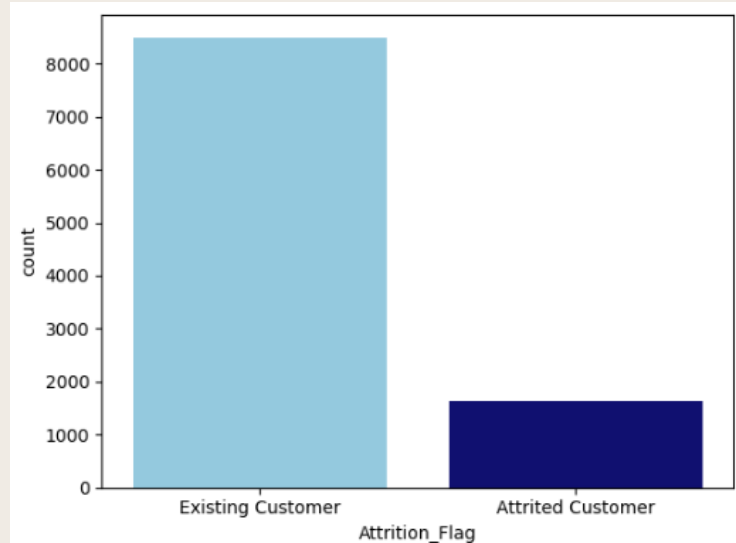
Datasets



Datasets



Datasets ini terdiri dari 10.127 pelanggan dengan fitur 23 kolom, dengan total kurang lebih 16% pelanggan yang berhenti berlangganan



02

EDA



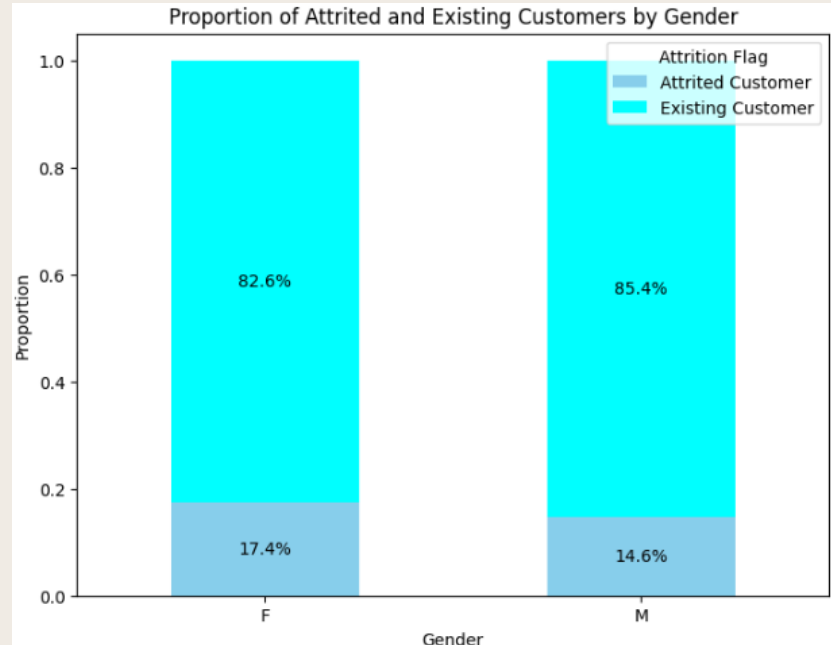
Proportion by Gender



Wanita lebih berisiko

Terdapat selisih 3% wanita lebih banyak yang berhenti berlangganan dibandingkan pria.

Hasil p-value 0,0001 menunjukkan adanya hubungan yang cukup signifikan



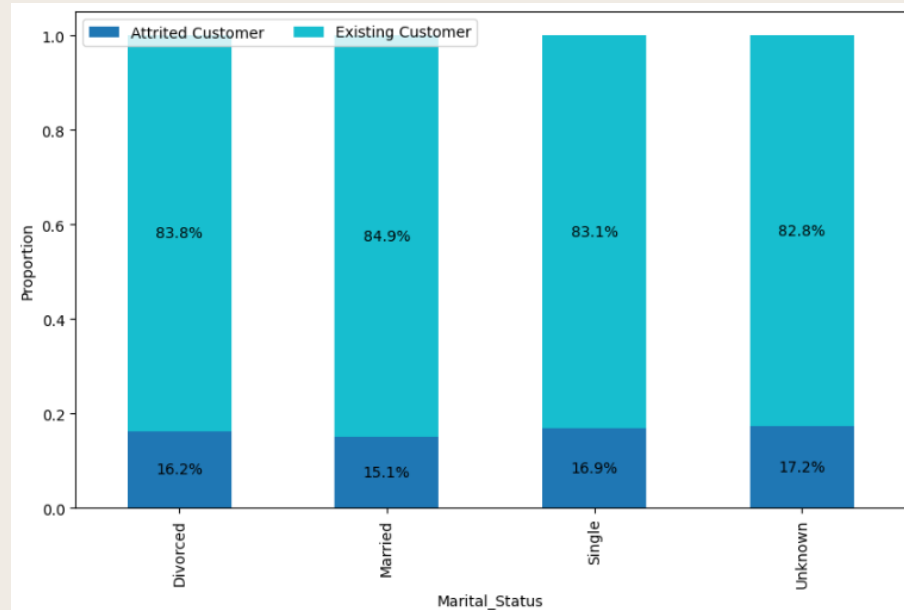
Proportion by Marital Status



Married cenderung retensi

Status married pada pelanggan membuat pelanggan lebih cenderung untuk tetap berlangganan dibandingkan dengan status single.

p-value bernilai 0,108 menunjukkan hubungan yang tidak signifikan



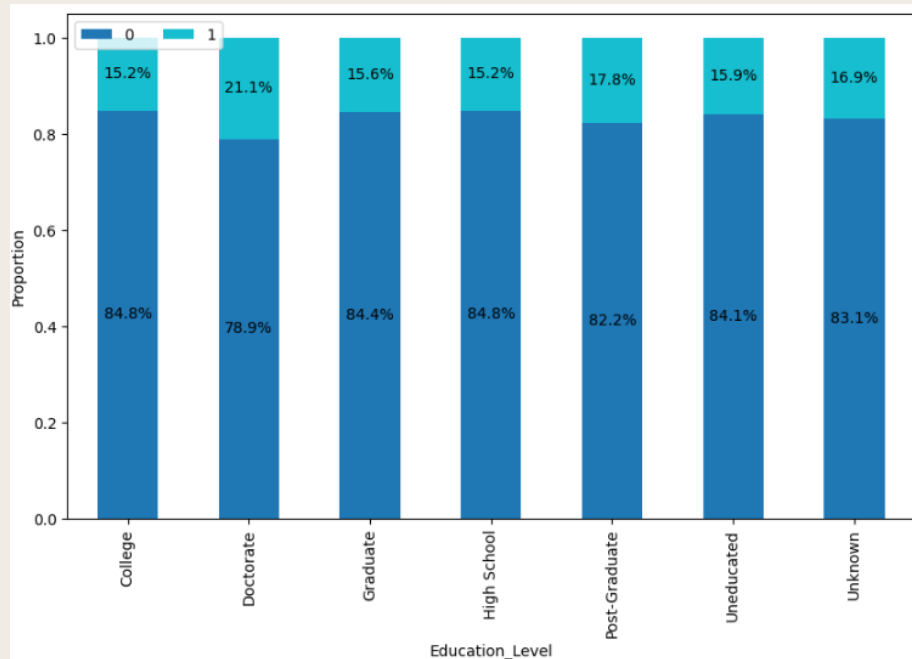
Proportion by Education Level



Doctorate level cenderung berhenti

Doctorate level memiliki kecenderungan attrition yang lebih tinggi dibanding level Pendidikan lainnya (21%)

p-value bernilai 0,052 menunjukkan hubungan yang tidak signifikan



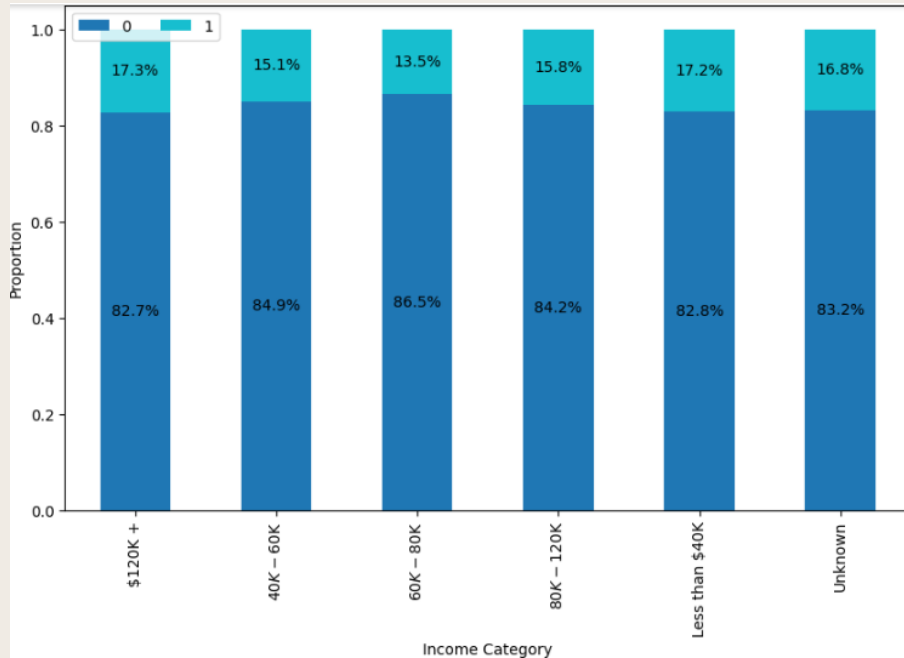
Proportion by Income Category



Penghasilan 120K memiliki resiko

Kategori penghasilan tinggi 120K memiliki resiko kehilangan pelanggan yang cukup besar (17,3%) dibanding kategori lainnya.

p-value bernilai 0,025 menunjukkan hubungan yang signifikan sehingga kategori ini perlu diperhatikan



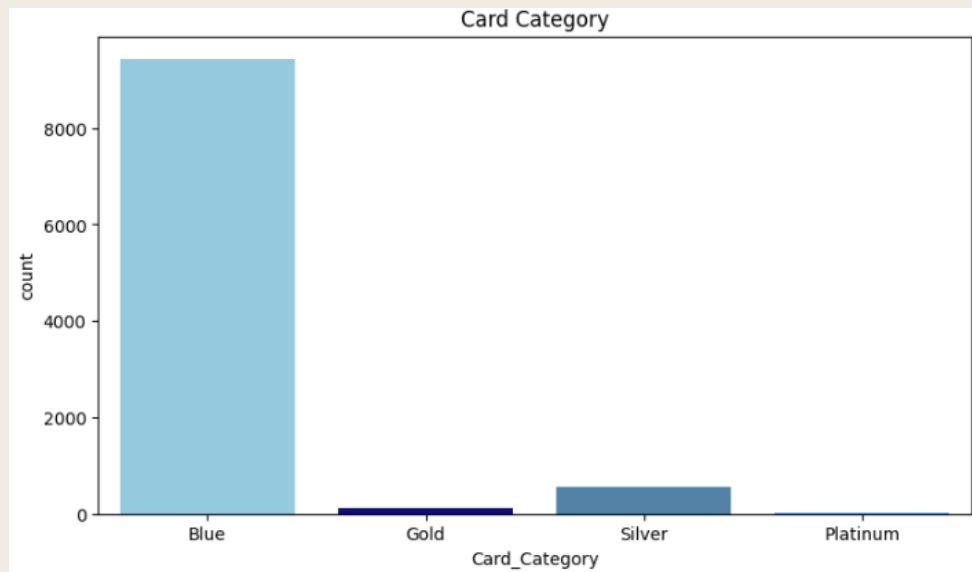
Proportion by Card Category



Blue Card mendominasi

Kategori Blue card mendominasi secara keseluruhan pada data pelanggan (>90%)

p-value bernilai 0,52 menunjukkan hubungan yang tidak signifikan

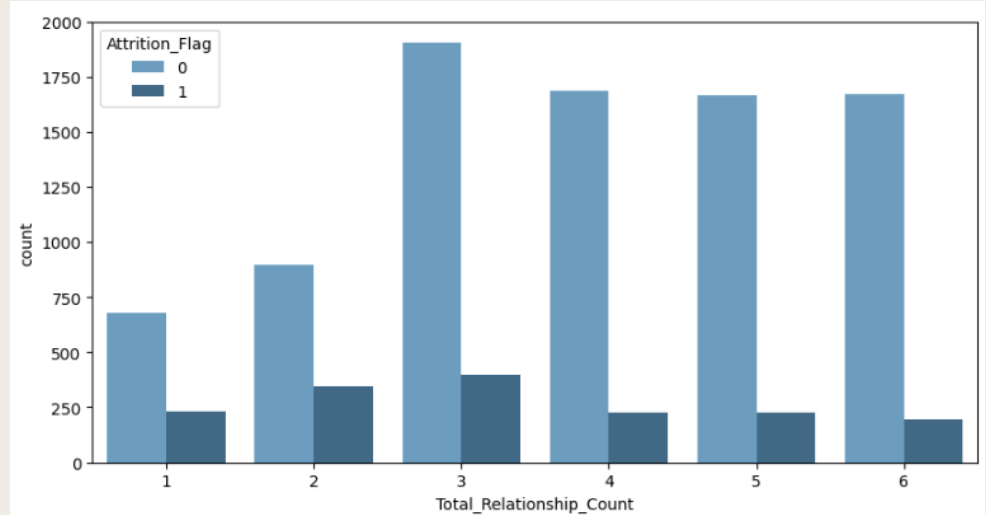


Total Relationships Distribution

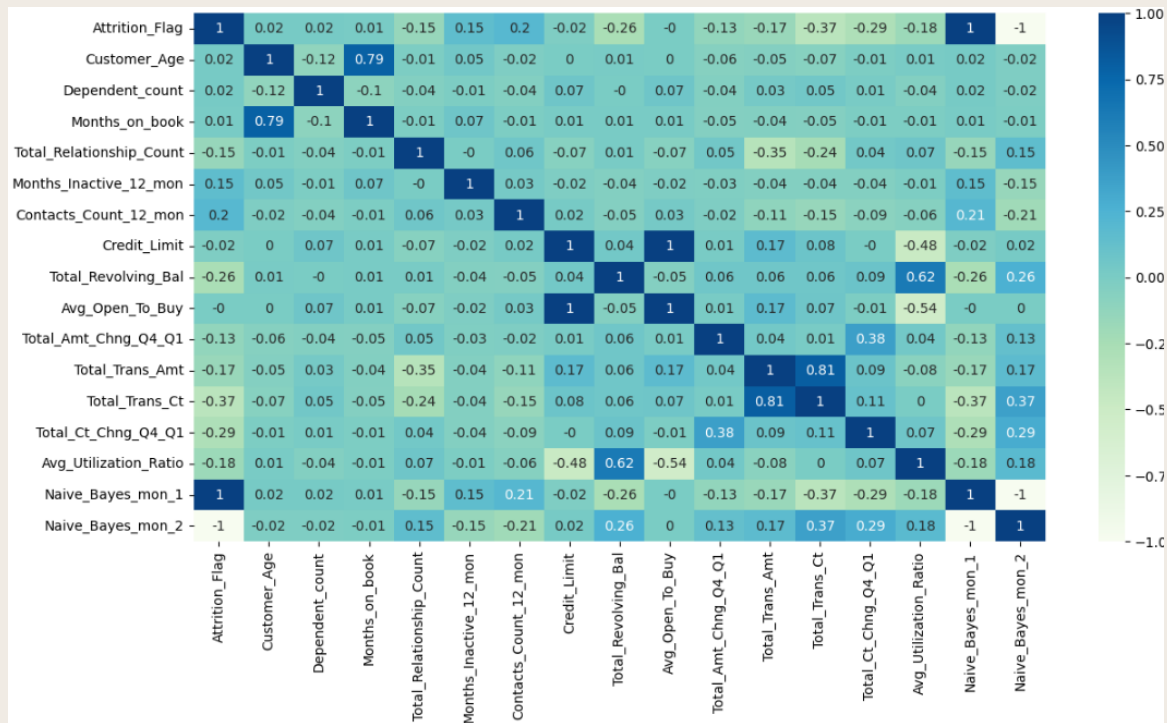


Berhenti ketika memiliki 3 kartu

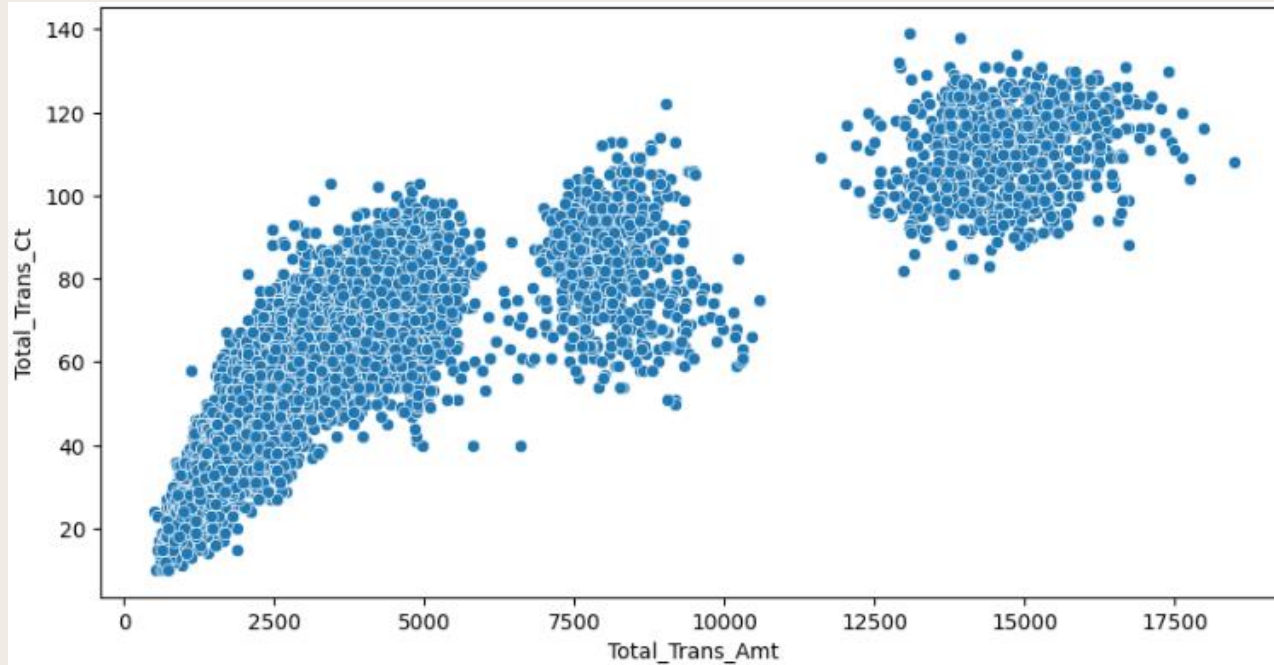
Pelanggan memiliki kecenderungan berhenti berlangganan ketika memiliki 3 kartu



Correlation Matrix

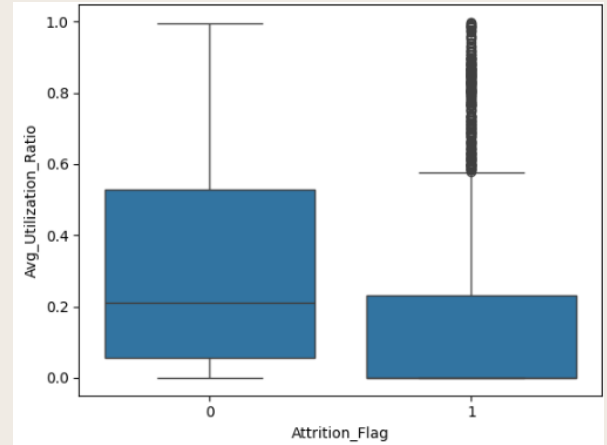
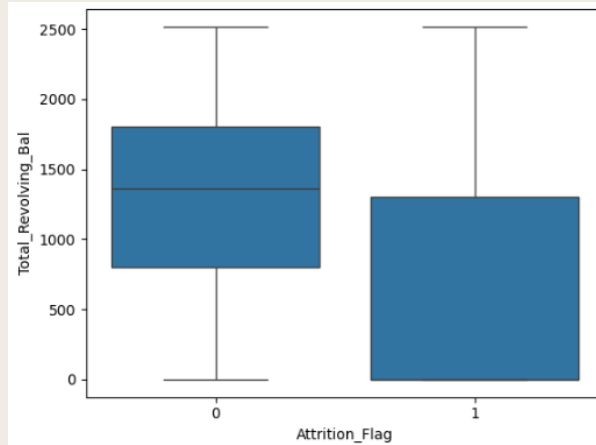
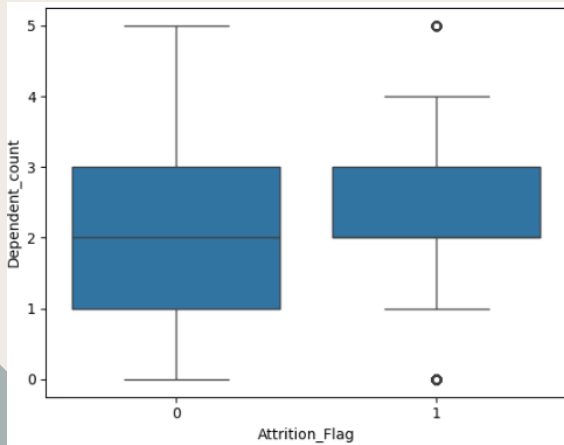


Transaction Scatter Plot



Box Plot

Dependent Count, Revolving Balance, dan Utilization Ratio memiliki distribusi yang cukup berbeda antara attrition customer dengan yang tidak, sehingga bisa dijadikan predictor dalam machine learning



Candidate Variables for Feature



Numerical

- Dependent_count
- Total_Relationship_Count
- Months_Inactive_12_mon
- Contacts_Count_12_mon
- Total_Trans_Ct
- Avg_Utilization_Ratio
- Revolving_Balance



Categorical

- Gender
- Income_category



03

Preprocessing Model



Preprocessing Data



- **Label Encoding**, konversi variable kategorikal menjadi nilai numerik
- **Standardization Scaling**
- **Imbalance Data Handling**, menyeimbangkan dataset agar tidak bias ke kelas mayoritas



04

Machine Learning Models



Machine Learning

01 Naïve-Bayes

klasifikasi probabilistik yang didasarkan pada Teorema Bayes dengan asumsi bahwa semua fitur bersifat independen

03 KNN

Prediksi kelas ditentukan oleh **mayoritas tetangga terdekat**

02 Logistic Regression

klasifikasi berbasis **regresi linear** yang memprediksi probabilitas suatu kelas

04 XGBoost

Menggabungkan banyak tree decision secara bertahap untuk mengurangi error (boosting)

Underfitting/Overfitting?

01 Naïve-Bayes

Training Accuracy: 87%; Test Accuracy: 87%

Training F1-score: 51%; Test F1-score: 50%

03 KNN

Training Accuracy: 90%; Test Accuracy: 88%

Training F1-score: 65%; Test F1-score: 61%

02 Logistic Regression

Training Accuracy: 88%; Test Accuracy: 88%

Training F1-score: 55%; Test F1-score: 52%

04 XGBoost

Training Accuracy: 96%; Test Accuracy: 90%

Training F1-score: 88%; Test F1-score: 68%

Handle Overfitting Model



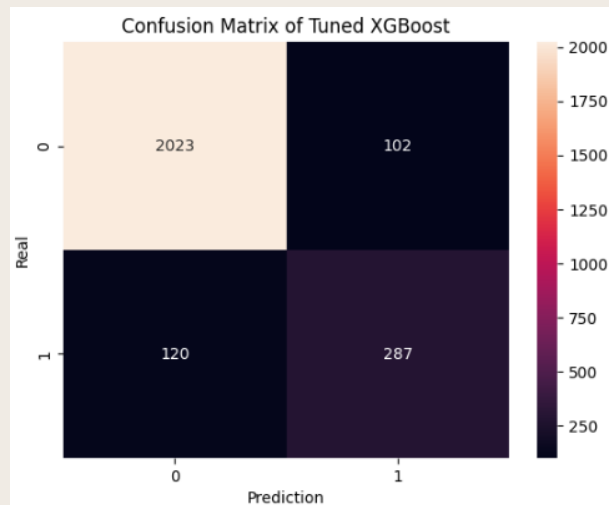
Hyperparameter Tuning

- Cross Validation, membagi dataset menjadi beberapa bagian lalu melatih secara berulang
- Grid search, menemukan kombinasi hyperparameter terbaik dengan mencoba semua kombinasi parameter

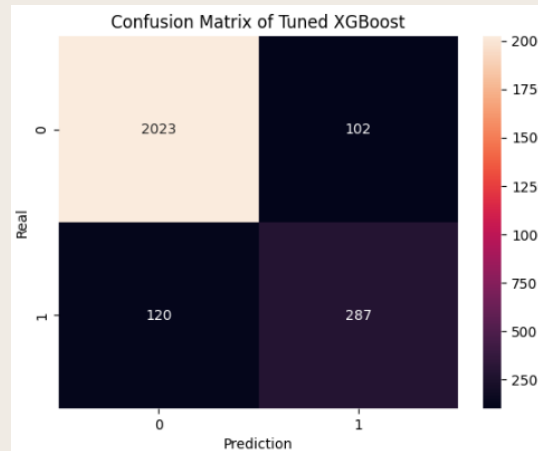
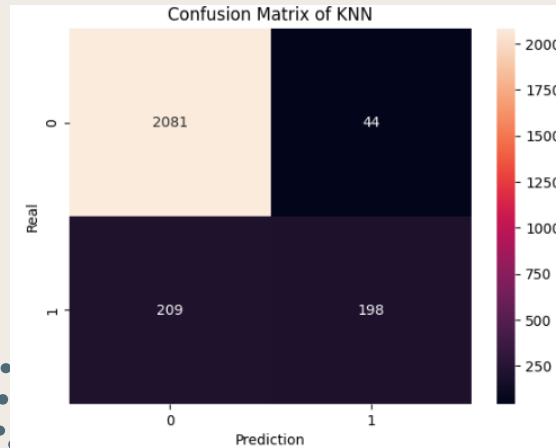
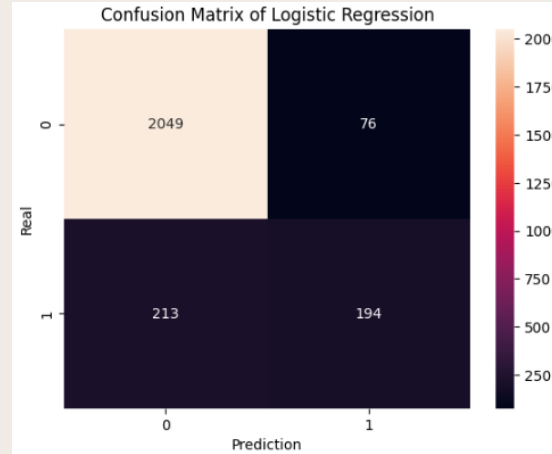
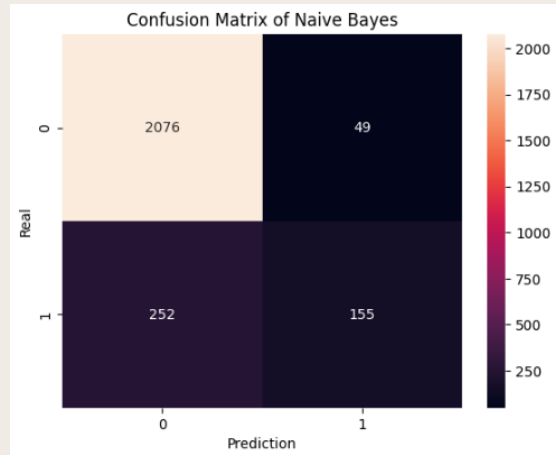


Evaluasi Metrics

- F1-score meningkat pesat, Training f1-score 80%; Test f1-score 72%
- Accuracy lebih realistis, Training Accuracy 93%; Test Accuracy 91%



Confusion Matrix



- Jumlah **False Negatives (124)** yang secara drastis menurun dibandingkan semua model sebelumnya. XGBoost dengan SMOTE paling efektif dalam mengurangi kesalahan tipe II (tidak mendeteksi pelanggan yang akan churn).
- Penurunan FN yang besar ini berdampak langsung pada peningkatan Recall (**287 True Positives berhasil dideteksi**).

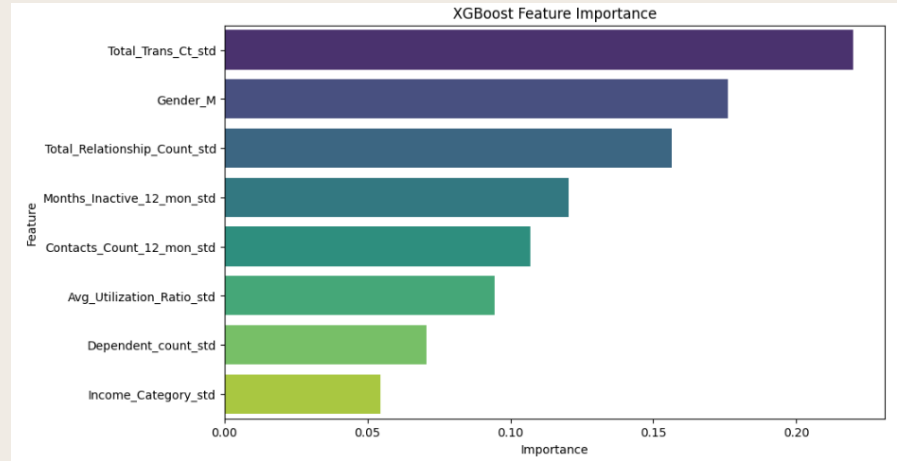
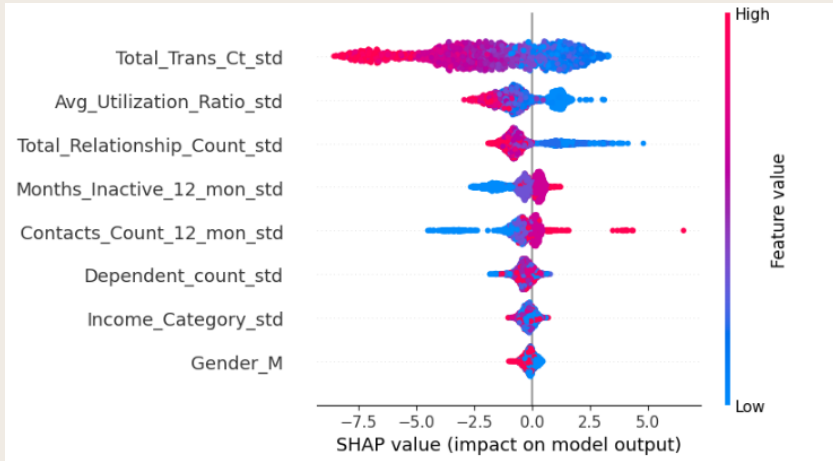
Metrics Evaluation Model

Recall yang rendah mengindikasikan bahwa model ini hanya mampu menangkap kurang dari setengah dari total pelanggan yang sebenarnya berhenti. F1-score yang relatif rendah mencerminkan ketidakseimbangan antara precision dan recall

Model XGBoost dengan menggunakan SMOTE memberikan performa terbaik untuk prediksi attrition pelanggan ini.

Model	Accuracy	Precision	Recall	F1-score	AUC
Naive Bayes	0.8772	0.7069	0.4029	0.5133	0.88
Logistic Regression	0.8859	0.7185	0.4767	0.5731	0.90
KNN (k=19)	0.8997	0.8097	0.4914	0.6116	0.92
XGBoost (SMOTE)	0.9143	0.7474	0.7052	0.7257	0.95

Feature Importance



- jumlah transaksi tinggi memiliki nilai SHAP negatif (mendorong prediksi tidak churn), yang berarti pelanggan dengan banyak transaksi cenderung tidak berhenti.
- Pelanggan yang memiliki sedikit hubungan dengan bank cenderung lebih mudah berhenti berlangganan.
- Gender Male mendorong prediksi ke arah tidak attrition dibandingkan dengan Gender Female.
- Pelanggan yang sering menghubungi bank mungkin memiliki masalah atau keluhan yang menyebabkan mereka mempertimbangkan untuk berhenti.

Business Insight



Boosting Transactions

- Berikan cashback 5% untuk transaksi minimal 3x/bulan.
- Berikan tips transaksi via email atau media social secara interaktif
- Berikan diskon marketplace yang relevan



Strengthening Customer Relationships

- Tawarkan produk simpanan dengan suku bunga tinggi jika mereka hanya memiliki kartu kredit.
- Upgrade otomatis ke kartu Silver/Gold untuk pelanggan Blue dengan total transaksi > Rp10 juta/bulan



Improving Customer Service

- pelatihan ulang CS untuk resolusi cepat (misalnya: batas waktu 24 jam)
- Review CRM Data dan lakukan evaluasi



Thanks!

Does anyone have any questions?

Prasetyo.sukmaraharjo@gmail.com

+6287889078546

[LinkedIn.com/in/prasetyosukmaraharjo](https://www.linkedin.com/in/prasetyosukmaraharjo)

