Visualization & Machine Learning

Dayana Gita Putra

Machine Learning

Telco Customer Churn Prediction

Outline

Memahami Business Problem → Data Cleansing, EDA & Featurisation → Membuat Machine Learning Model → Kesimpulan

PROBLEM

Customer berhenti berlangganan (Churn) layanan telekomunikasi sehingga dapat menyebabkan kerugian bagi perusahaan.

GOAL

Membangun sebuah
model machine learning untuk membantu
dalam mencegah customer churn

Exploratory Data Analysis

Data Overview

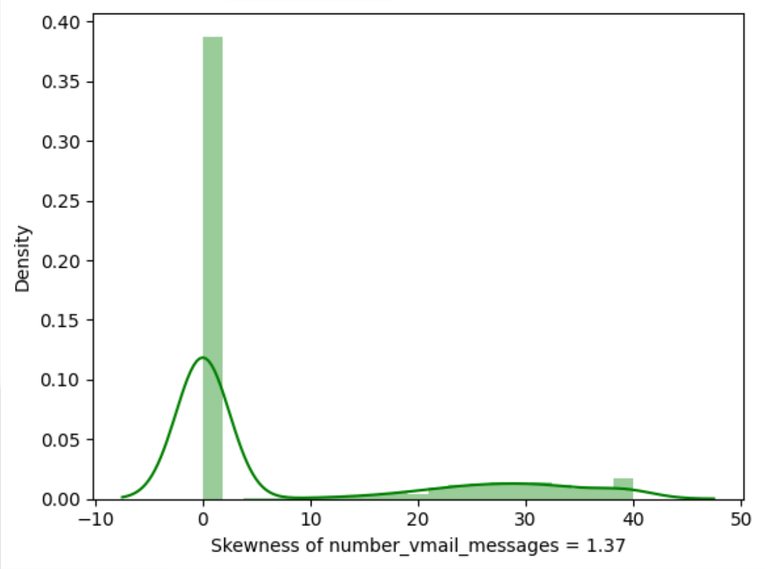
1. Dataset terdiri dari **4.250 baris** dan **20 kolom**
2. Terdapat **15 data numerik** dan **5 data kategorik**

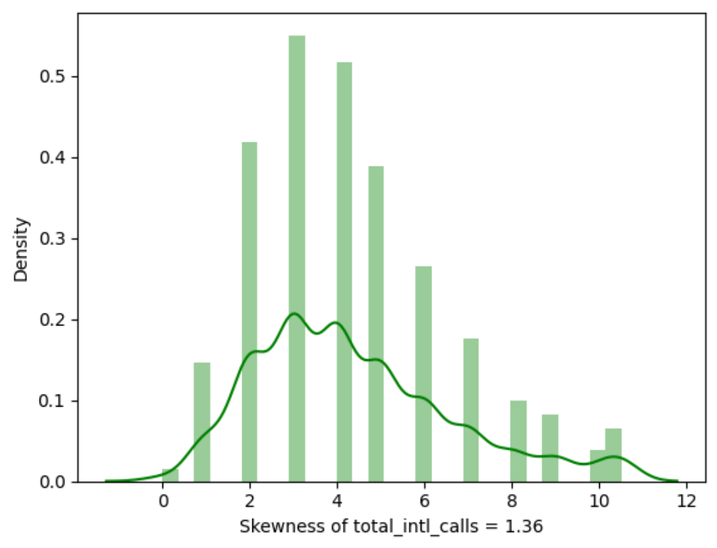
14 %

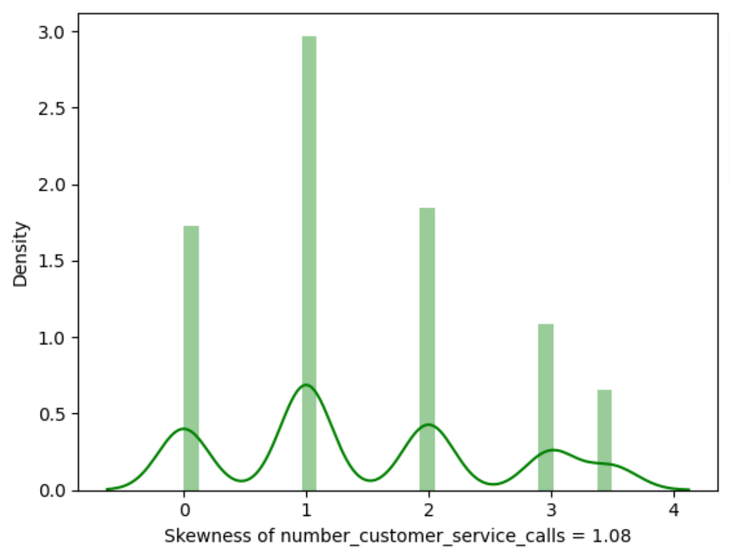
Yes Customer Churn

No Customer tidak Churn

Data Distribution

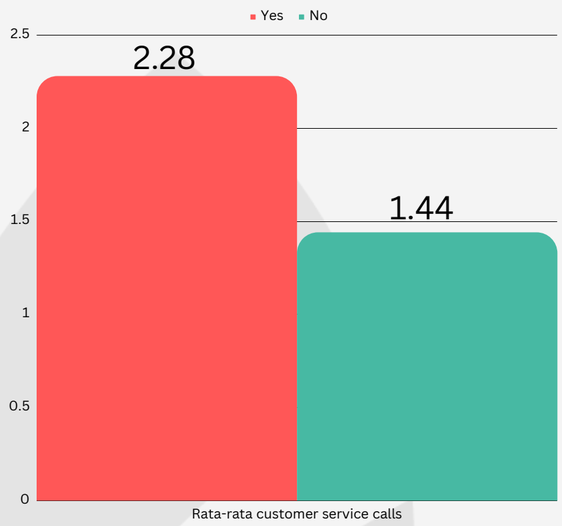




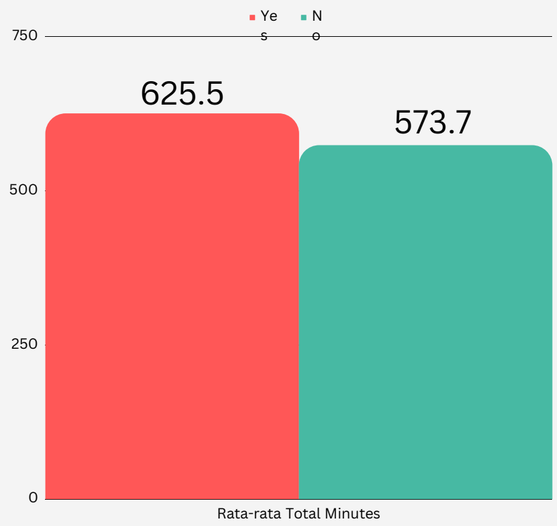


Seluruh feature memiliki distribusi yang normal kecuali feature :
number_vmail_messages, total_intl_calls, dan **number_customer_service_calls.**

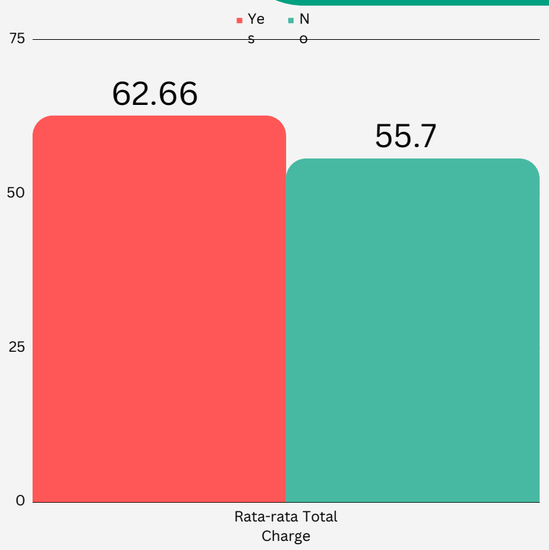
Bivariate Analysis



Customer yang melakukan panggilan lebih dari **2x panggilan pada customer service** memiliki kecenderungan untuk Churn

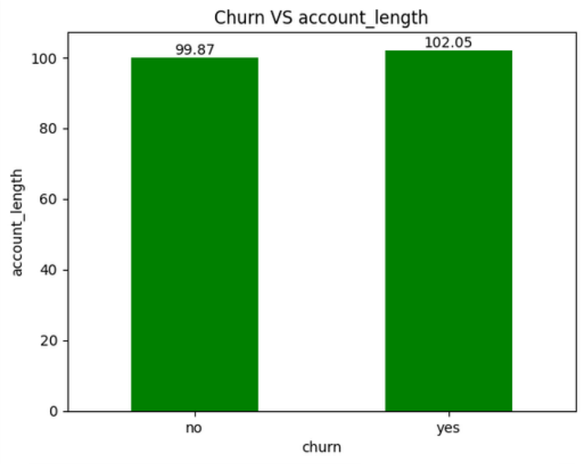


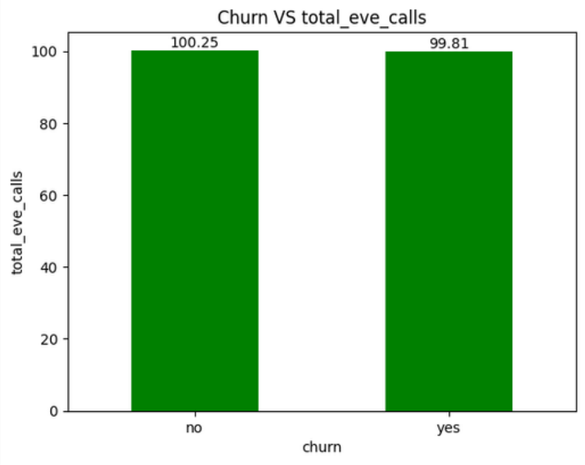
Customer yang memiliki durasi panggilan**rata-rata 625.5 minute (±10 jam)** memiliki kecenderungan untuk Churn

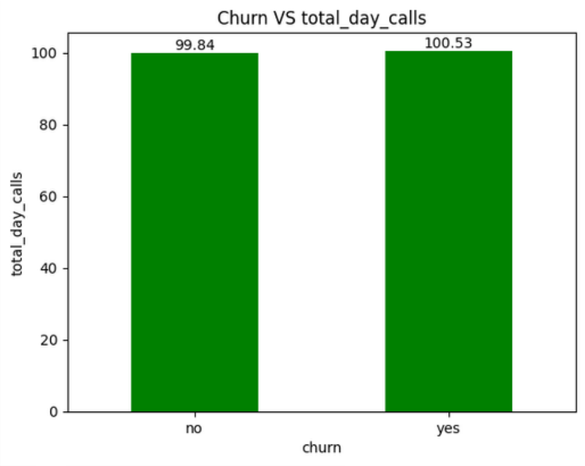


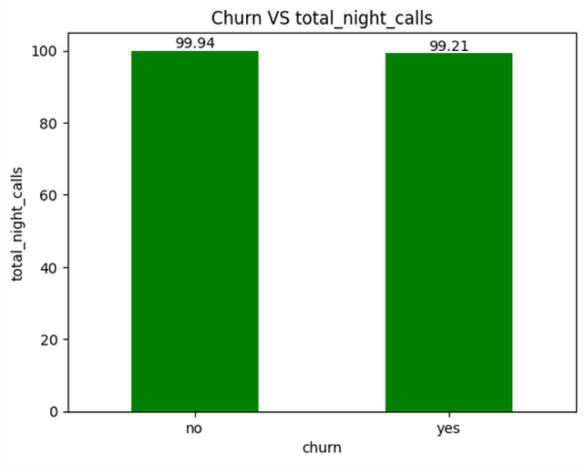
Customer yang memiliki tagihan (charge) **rata-rata 62.66 dolar** memiliki kecenderungan untuk Churn

Bivariate Analysis



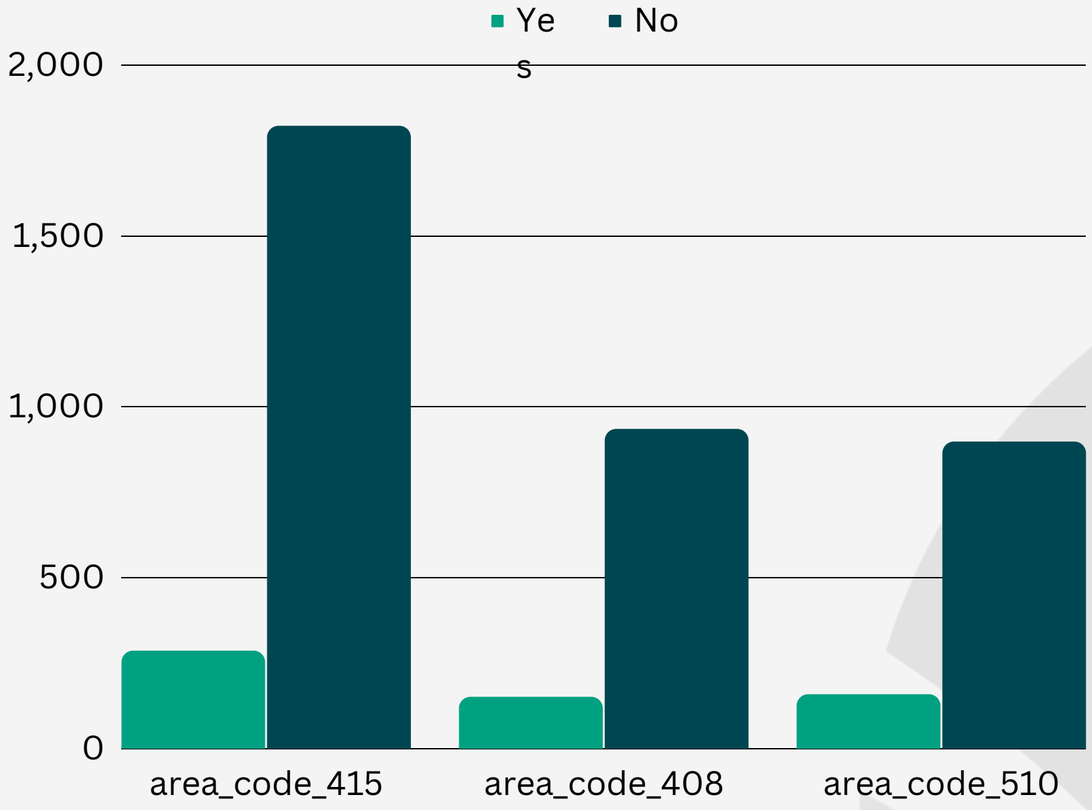






Setelah dilakuakan uji statistik terhadap feature numeric terdapat beberapa feature prediktor yang tidak menyebabkan perbedaan yang signifikan terhadap feature target, diantaranya adalah **account_length, total_day_calls, total_eve_calls, total_night_calls**.

Bivariate Analysis



Pada feature categoric setelah dilakuakan uji statistik didapati bahwa feature **area_code** tidak mempengaruhi feature target

Data Pre-processing

Handling Outlier

Menghapus nilai outlier

Scaling

MinMaxScaller

Featurisation

Label encode
Feature Engineering

Split Data

Train : 3187
Test : 1063

Handling Imbalance

Over Sampling
SMOTE

Modeling

Features

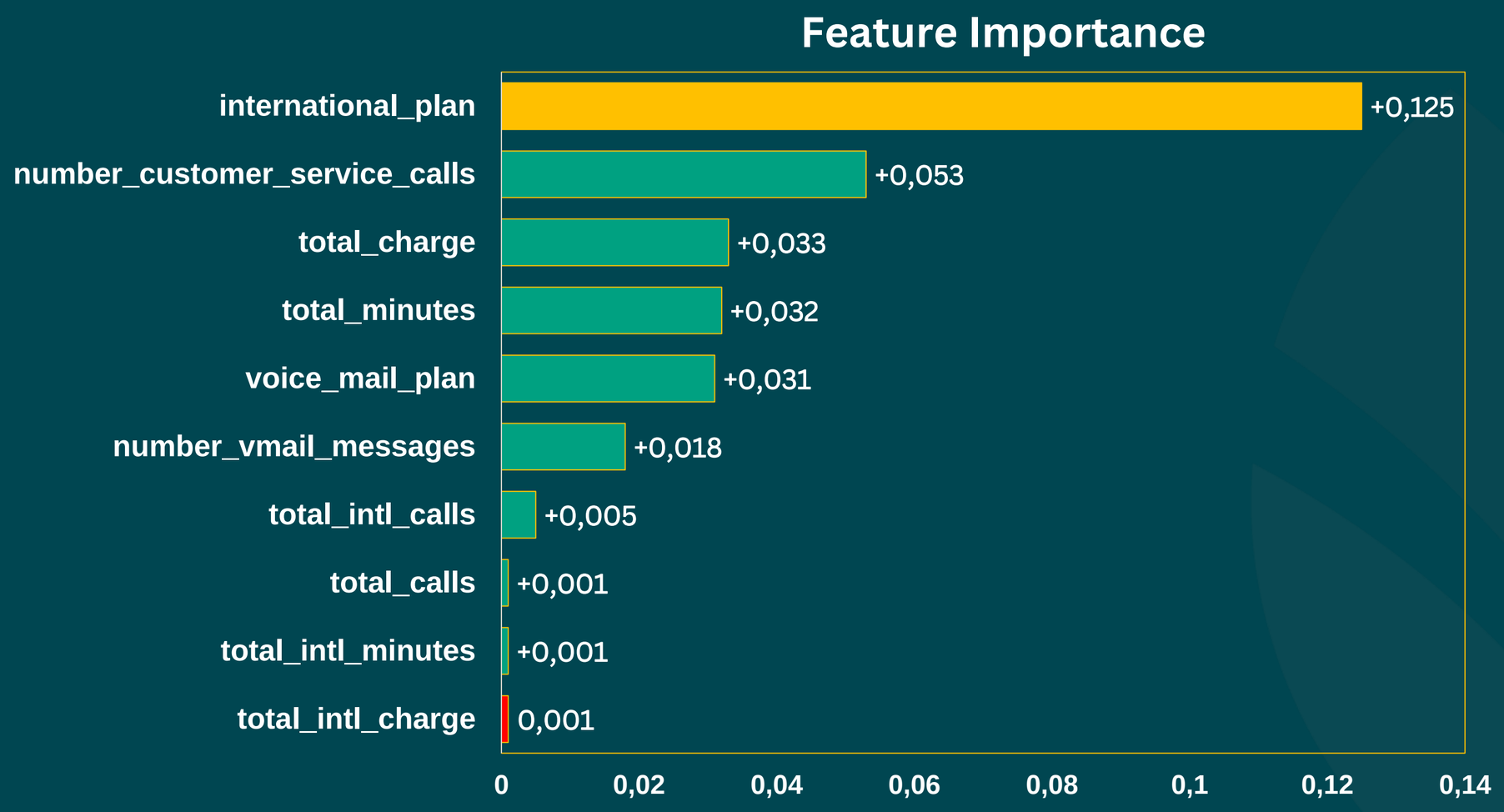
Dataset Features	Model Features
Id	
State	
Total_intl_charge	Total_intl_charge
Total_intl_calls	Total_intl_calls
Total_intl_minute	Total_intl_minute
Total_night_charge	Total_charge
Total_eve_charge	
Total_day_charge	
Total_night_calls	Total_calls
Total_eve_calls	
Total_day_calls	
Total_night_minutes	Total_minutes
Total_eve_minutes	
Total_day_minutes	
number_vmail_messages	Number_vmail_messages
Voice_mail_plan	Voice_mail_plan
International_plan	International_plan
Area_code	
Account_length	
Number_customer_service_calls	Number_customer_service_calls

Classification Models

Model	Train AUC	Test AUC
DecisionTreeClassifier	1.00	0.81
RandomForestClassifier	1.00	0.86
KNeighborsClassifier	0.93	0.76
GaussianNB	0.78	0.77
LogisticRegression	0.75	0.74

Algoritma terbaik adalah **GaussianNB** dimana nilai AUC training dan Test memiliki perbedaan yang kecil sehingga dapat dikatan **model fit**

Feature Importance



Kesimpulan

Model machine learning Naive-Bayes dengan nilai **AUC 78 % pada** Train dan **77 % Test**, merupakan pilihan terbaik untuk memprediksi customer churn.

Terima Kasih.

Design by Canva