

# VISUALIZATION & MACHINE LEARNING

Kelompok DS-3-J  
FGA x Binar Academy – Data Science 1

The background is a dark blue gradient. It is decorated with various geometric elements: small squares in white, pink, and teal, some of which are solid and others are outlines. Thin white vertical lines of varying lengths are scattered across the page. A small teal horizontal line is located near the top center.

# CHALLENGE 1

## COVID-19 DASHBOARD

## Problem

Pandemi Covid-19 telah melanda dunia sejak tahun 2020. Hingga sekarang, tren dari kasus covid di Indonesia sempat mengalami proses naik dan turun.

## Objective

Membuat dashboard yang dibuat ditujukan agar masyarakat umum dapat mengetahui perkembangan kasus covid yang ada.

## Dataset

Data penderita COVID-19 (termasuk total kasus, kasus aktif, kasus kesembuhan, dan kasus kematian) di Indonesia dari awal pandemi hingga 16 September 2022





# INDONESIA COVID-19 DASHBOARD

Location

Province

Pilih rentang tanggal



Total Population  
242,7 M



Confirmed  
6,4 jt



Recovered  
6,2 jt



Cases Active  
34,9 rb



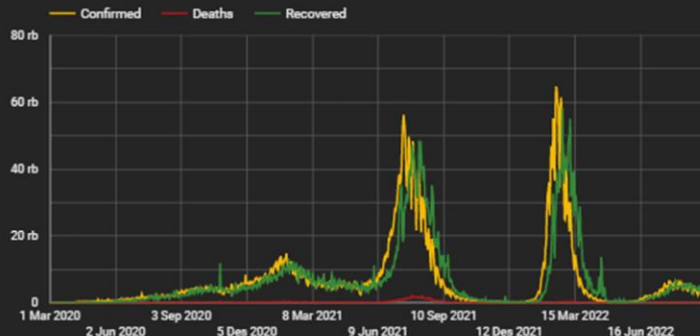
Deaths  
157,8 rb

Recovered Ratio  
93,63%

## Total Confirmed Cases by Location

Location	Total Confirmed	Recovered
DKI Jakarta	1 jt	98,13%
Jawa Barat	1 jt	97,49%
Jawa Tengah	636 rb	94,52%
Jawa Timur	602 rb	94,54%
Banten	334 rb	98,35%
Daerah Istimew...	224 rb	97,06%
Kalimantan Timur	209 rb	97,13%
Bali	167 rb	96,88%
Sumatera Utara	159 rb	97,51%
Riau	153 rb	96,94%
Sulawesi Selatan	144 rb	98,17%
Sumatera Barat	105 rb	97,54%
Nusa Tenggara ...	94 rb	98,27%
Kalimantan Sela...	87 rb	96,87%
Sumatera Selatan	82 rb	95,51%
Lampung	75 rb	94,16%
Kepulauan Riau	71 rb	97,15%
Kepulauan Bang...	66 rb	97,45%
Kalimantan Barat	66 rb	98,12%
Sulawesi Tengah	61 rb	97,02%
Kalimantan Ten...	58 rb	96,94%
Sulawesi Utara	53 rb	96,63%
Papua	50 rb	98,37%
Kalimantan Utara	45 rb	98,05%

## Daily Cases Trend



## Top 10 Total Cases of Deaths by Location

Location	Deaths
Jawa Tengah	33 rb
Jawa Timur	32 rb
Jawa Barat	16 rb
DKI Jakarta	15 rb
Daerah Istimewa Yogy...	6 rb
Kalimantan Timur	6 rb
Bali	5 rb
Riau	4 rb
Lampung	4 rb
Sumatera Selatan	3 rb

## Total Cases Distribution in Indonesia



Total Case 13.951 1.412.474

## Recovered Rate and Fatality Rate by Province

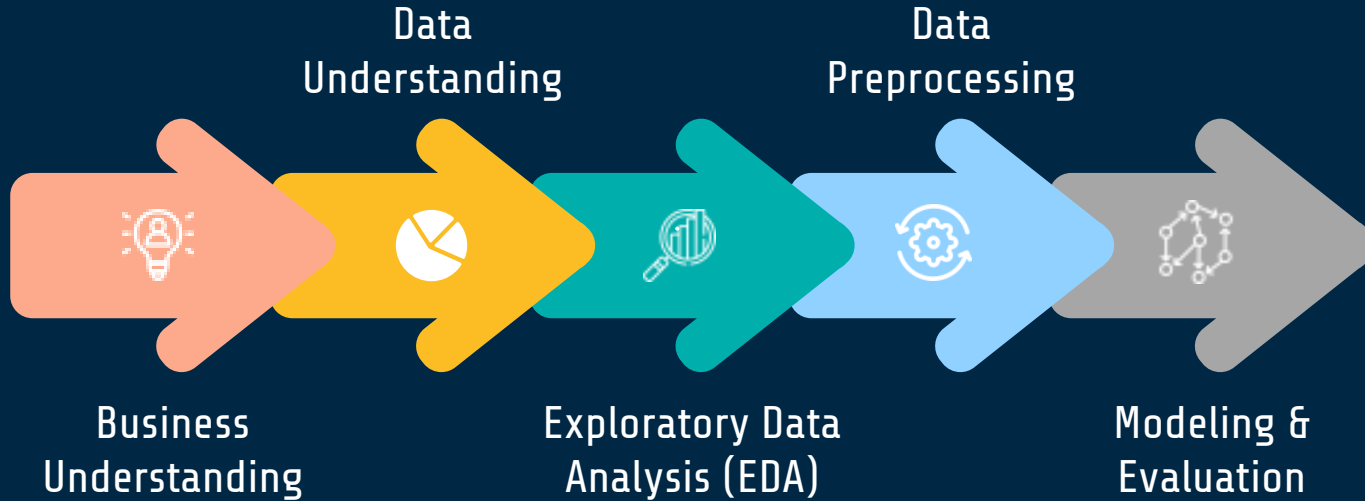


The background is a dark blue gradient. It features several vertical white lines of varying lengths. Scattered throughout are small squares in teal, pink, and orange. Some squares are solid, while others are outlined. The overall aesthetic is modern and tech-oriented.

# CHALLENGE 2

Predict Telco Customer  
Churn

# Diagram Proses



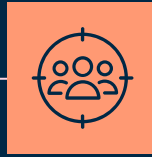
# Business Understanding

01

# Problem Statement



Perkembangan industri telekomunikasi memperketat persaingan antar provider



Perusahaan harus mampu mempertahankan pelanggannya



Perusahaan harus dapat mengetahui pelanggan yang akan churn dan faktor yang mempengaruhinya

Pelanggan memiliki hak dalam memilih provider yang sesuai dan dapat beralih dari provider sebelumnya yang diartikan sebagai **Customer Churn**. Peralihan ini dapat menyebabkan berkurangnya pendapatan bagi perusahaan telekomunikasi sehingga penting untuk ditangani.



## Goal

Melakukan prediksi agar bisa memetakan strategi bisnis untuk mempertahankan pelanggan.

## Objective

- Membuat model machine learning yang dapat memprediksi pelanggan yang berpotensi churn
- Mengetahui pola/parameter pelanggan churn

## Matric

Churn Rate



# Data Understanding

02

Fitur Name	Description
state	US State
account_length	Total bulan customer menjadi user telco provider
area_code	Kode Area
International_plan	Customer memiliki plan international
voice mail plan	Customer memiliki plan voice mail
number_vmail_messages	Total pesan voice mail
total_day_minutes	Total minutes pada day calls
total_day_calls	Total day calls
total_day_charge	Total charge dari day calls
total_eve_minutes	Total menit pada evening call
total_eve_calls	Total evening call
total_eve_charge	Total charge pada evening call
total_night_minutes	Total menit pada night call
total_night_calls	Total night call
total_night_charge	Total charge pada night call
total_intl_minutes	Total menit pada international call
total_intl_calls	Total international call
total_intl_charge	Total charge pada international call
number_customer_service_calls	Total call kepada customer service

## Dataset Info



20 kolom & 4250 baris



5 fitur kategorikal & 15 fitur Numerikal

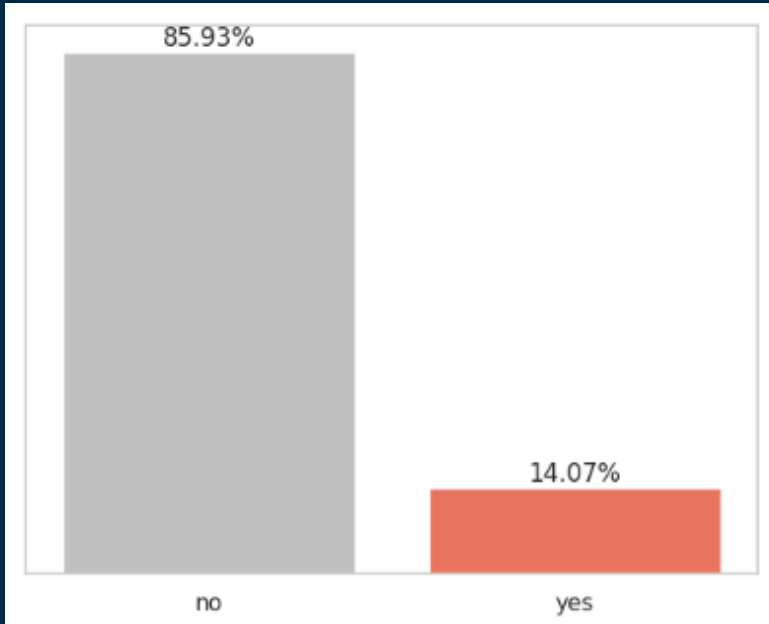


Dataset tidak memiliki null/missing value  
Dataset tidak memiliki baris yang duplikat

# Exploratory Data Analysis

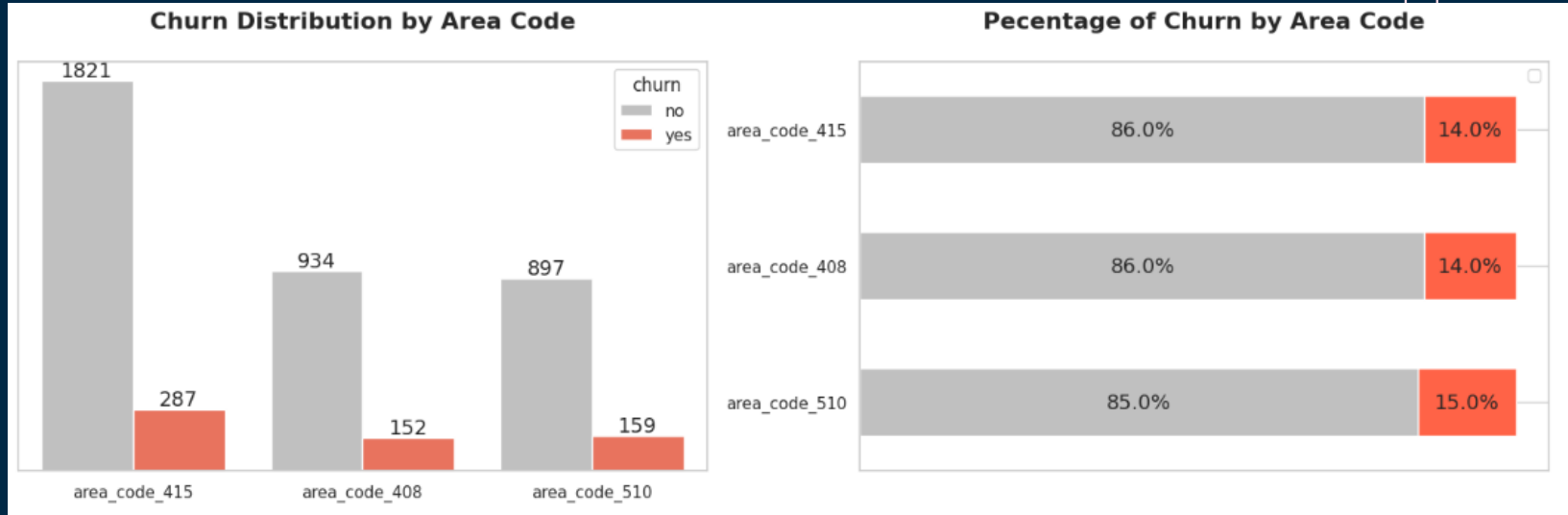
03

# Persentase Pelanggan Churn



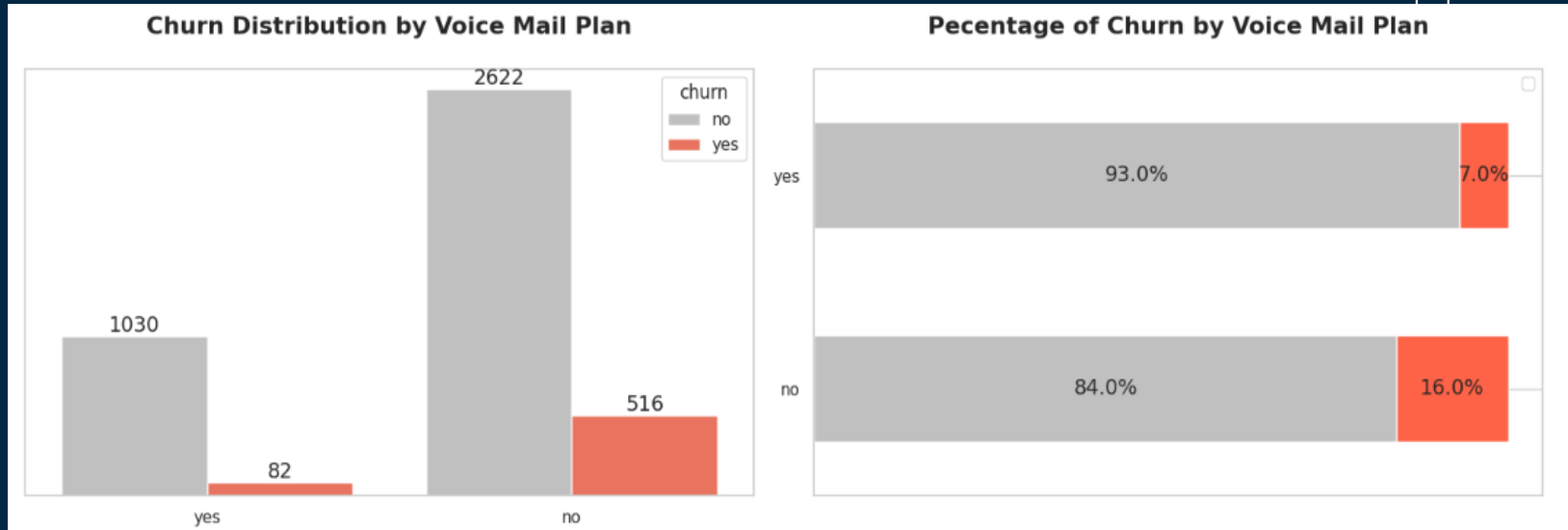
Pada dataset 14.07% adalah customer churn. Menurut sumber rata-rata tingkat churn untuk perusahaan telekom adalah 22%\*. Berdasarkan hal tersebut tingkat churn rate di perusahaan ini masih tergolong wajar. Namun tidak salah untuk tetap meminimalisir tingkat churn sehingga perusahaan dapat meningkatkan profit lebih tinggi.

# Area Code



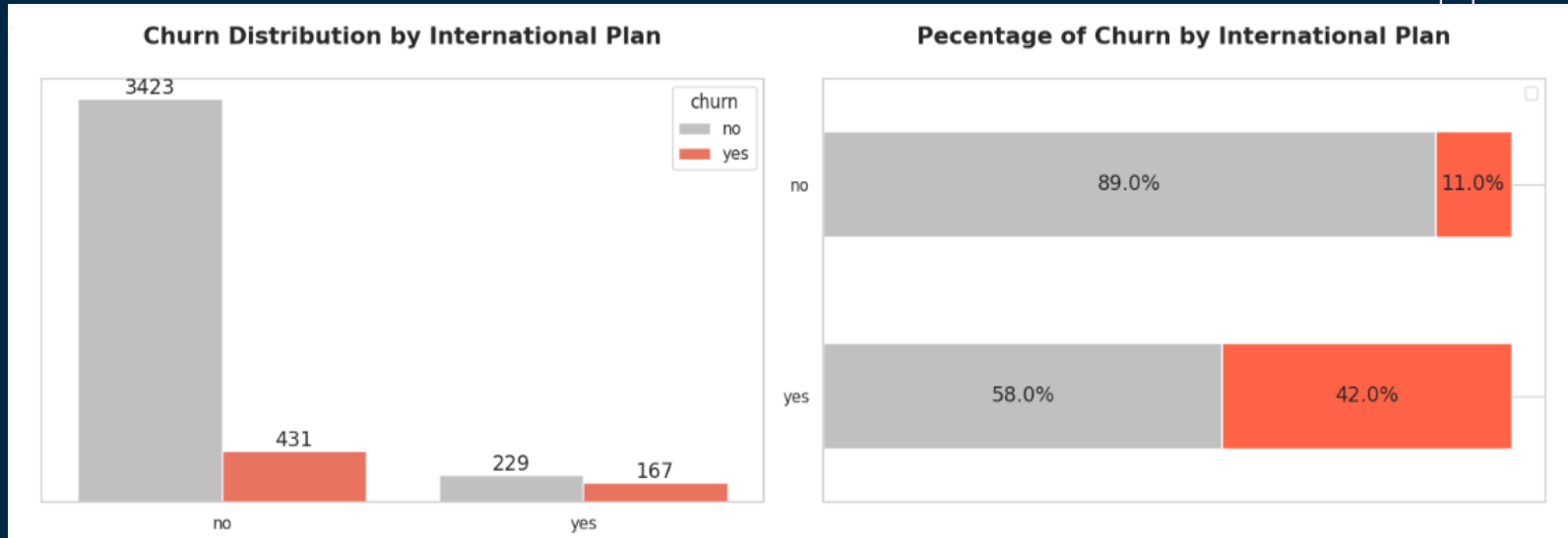
- Area 415 memiliki jumlah (count) customer paling tinggi.
- Secara rasio, pada setiap area memiliki persentase customer yang churn tidak jauh berbeda yaitu 14 - 15 %

# Voice Mail Plan



- Mayoritas pelanggan tidak memiliki voice mail plan, dan pelanggan yang tidak memiliki voice mail plan lebih berpotensi untuk churn
- Kemungkinan pelanggan churn karena jika pelanggan tidak memiliki voice mail plan, mereka tidak akan dapat menerima pesan suara dari orang-orang yang mencoba menghubungi mereka ketika mereka tidak dapat menjawab telepon. Ini dapat menyebabkan ketidaknyamanan bagi pelanggan.

# International Plan

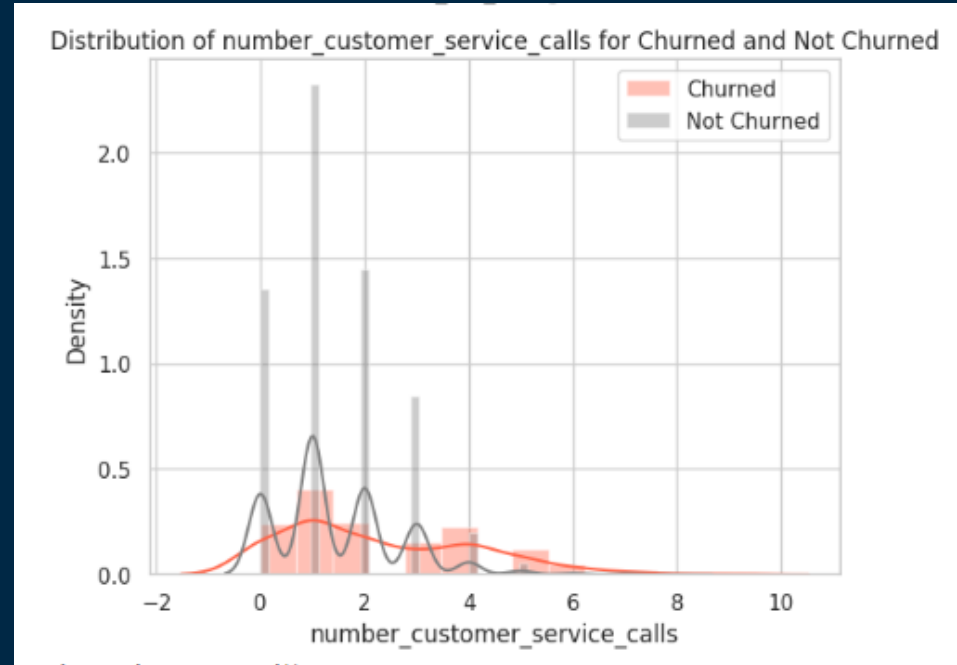


- Angka pelanggan yang memiliki International Plan cukup rendah namun memiliki tingkat churn yang tinggi
- Churn pada pelanggan yang memiliki Internatiol plan ini kemungkinan dapat terjadi karena biaya roaming yang tinggi atau masalah kualitas jaringan provider



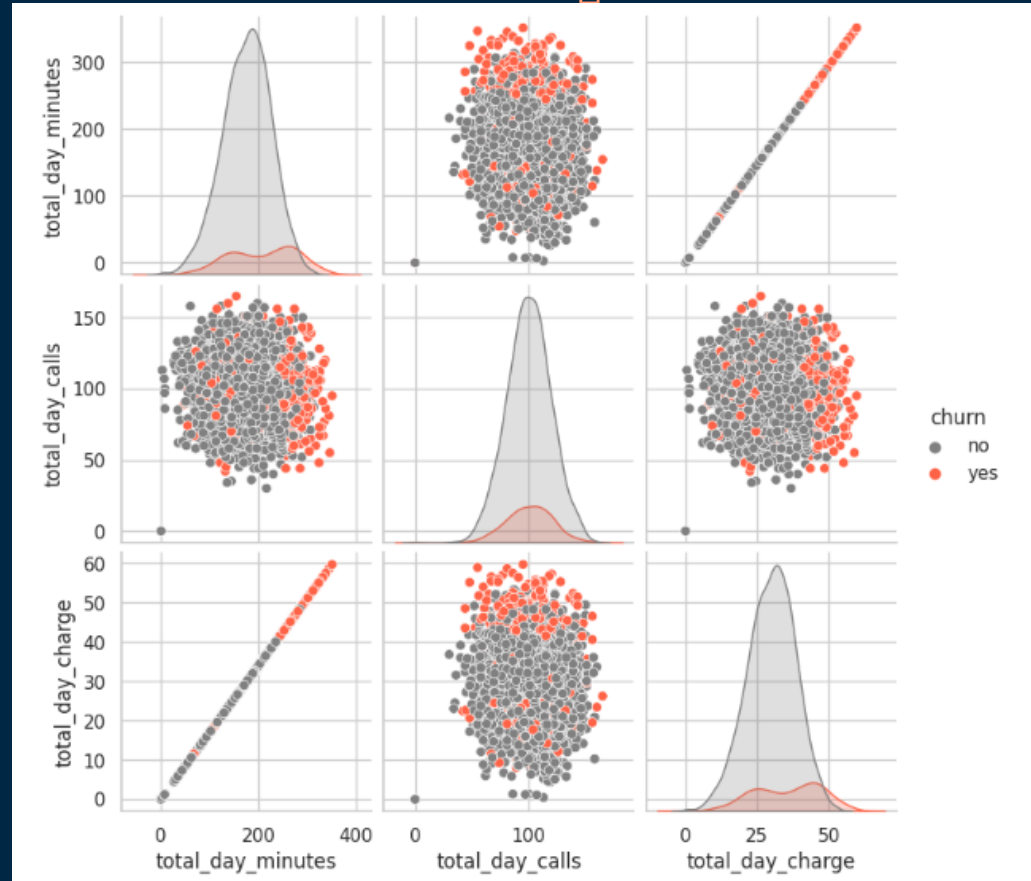
# Number Customer Service Calls

- Pelanggan yang menghubungi customer service lebih dari 3 kali sangat berpotensi untuk churn
- Hal tersebut dapat diduga karena ketidakpuasan pelanggan terhadap layanan atau pemecahan masalah/keluhan yang diberikan



# Total charge, calls, minute

Total harga yang disebabkan pada panggilan di pagi hari memiliki pengaruh yang cukup besar terhadap tingkat churn pelanggan. Terlihat bahwa mayoritas pelanggan churn yang memiliki durasi menit panggilan lebih lama, mendapatkan harga panggilan yang lebih besar. Hal tersebut kemungkinan dapat disebabkan karena pelanggan tidak puas terhadap harga untuk telepon dengan durasi yang lama (mungkin terlalu mahal untuk pagi hari).



# Data Preprocessing

04

Tidak dilakukan  
handling outlier,  
karena outlier masih  
dapat ditoleransi  
atau bukan  
kesalahan input  
data/nilai ekstrim

## DATA DUPLIKAT

Tidak terdapat  
data duplikat

Dilakukan Label  
Encoding pada fitur  
'international\_plan', 'voic  
ce\_mail\_plan', 'churn'  
(Yes - 1, No - 0), dan  
area\_code

## FEATURE ENCODING

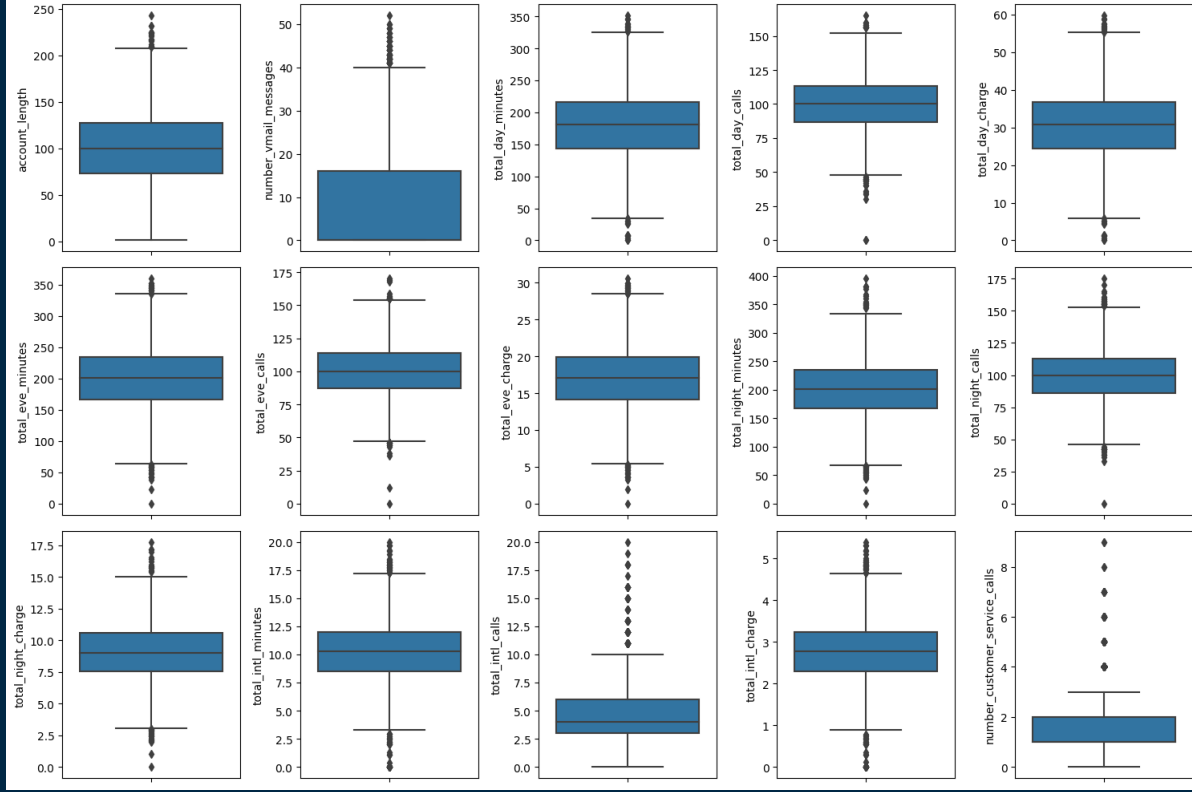
## FEATURE SELECTION

Memilih fitur yang akan  
digunakan pada model

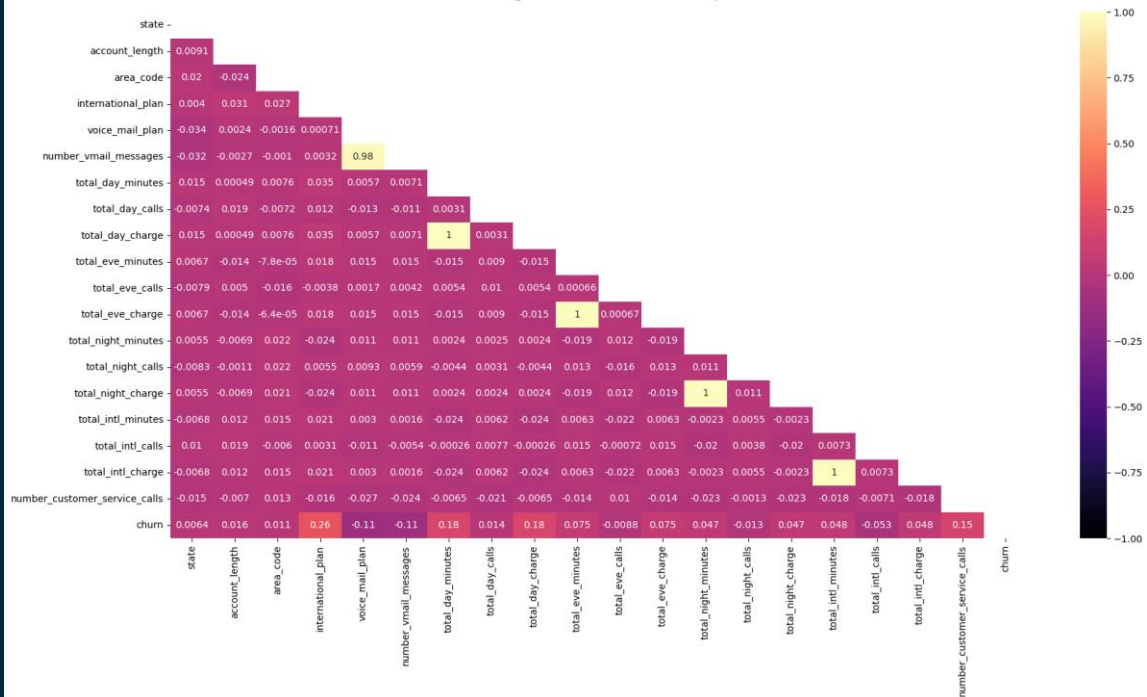
Feature  
Standardization,  
split train test,  
model experiment

## MODELING

## OUTLIER



Triangle Correlation Heatmap



```
# Print the list of redundant feats
for feature in to_drop:
    corr_value = corr_matrix[feature][corr_matrix[feature] >= 0.8]
    print(f"{feature}: {list(corr_value.index)} (Correlation Values: {list(corr_value)})")
```

```
number_vmail_messages: ['voice_mail_plan', 'number_vmail_messages'] (Correlation Values: [0.9843505975685432, 1.0])
total_day_charge: ['total_day_minutes', 'total_day_charge'] (Correlation Values: [1.0, 1.0])
total_eve_charge: ['total_eve_minutes', 'total_eve_charge'] (Correlation Values: [0.999999786466728, 1.0])
total_night_charge: ['total_night_minutes', 'total_night_charge'] (Correlation Values: [0.9999988534695725, 1.0])
total_intl_charge: ['total_intl_minutes', 'total_intl_charge'] (Correlation Values: [1.0, 1.0])
```

## Feature Selection

Drop beberapa fitur seperti  
'number\_vmail\_messages',  
'total\_day\_minutes',  
'total\_eve\_minutes',  
'total\_night\_minutes',  
'total\_intl\_minutes',  
'state'

# Modeling & Evaluation

04

# Model Evaluation

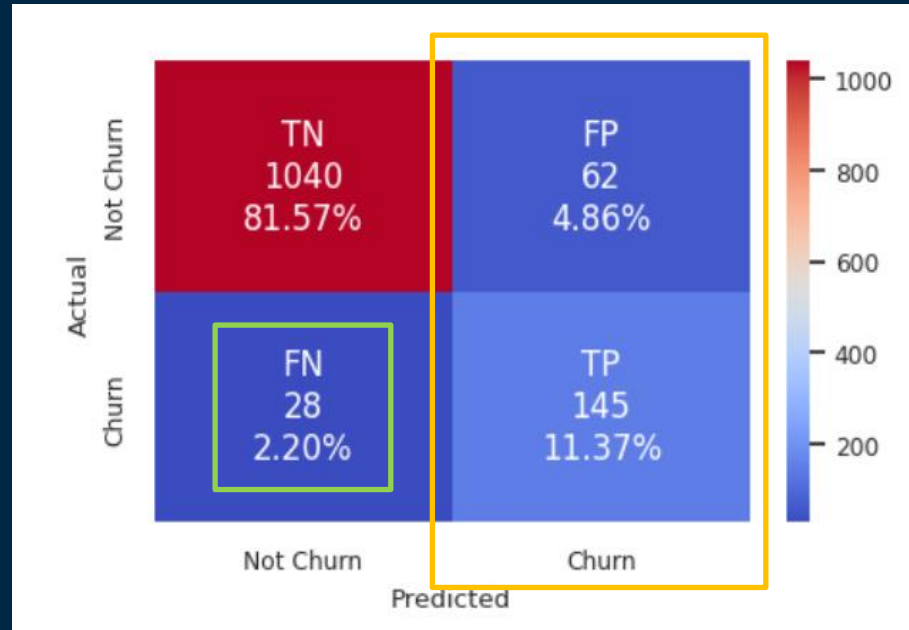
Model	Accuracy	Recall	AUC Test	AUC Train
Random Forest	93%	89%	93%	94%
XGBoost	85%	86%	92%	95%
Logistic Regression	77%	78%	84%	82%

- Dari beberapa model, Random Forest menjadi model terbaik
- Menggunakan algoritma Supervised learning
- Memiliki nilai AUC tinggi untuk memaksimalkan prediksi



# Confusion Matrix

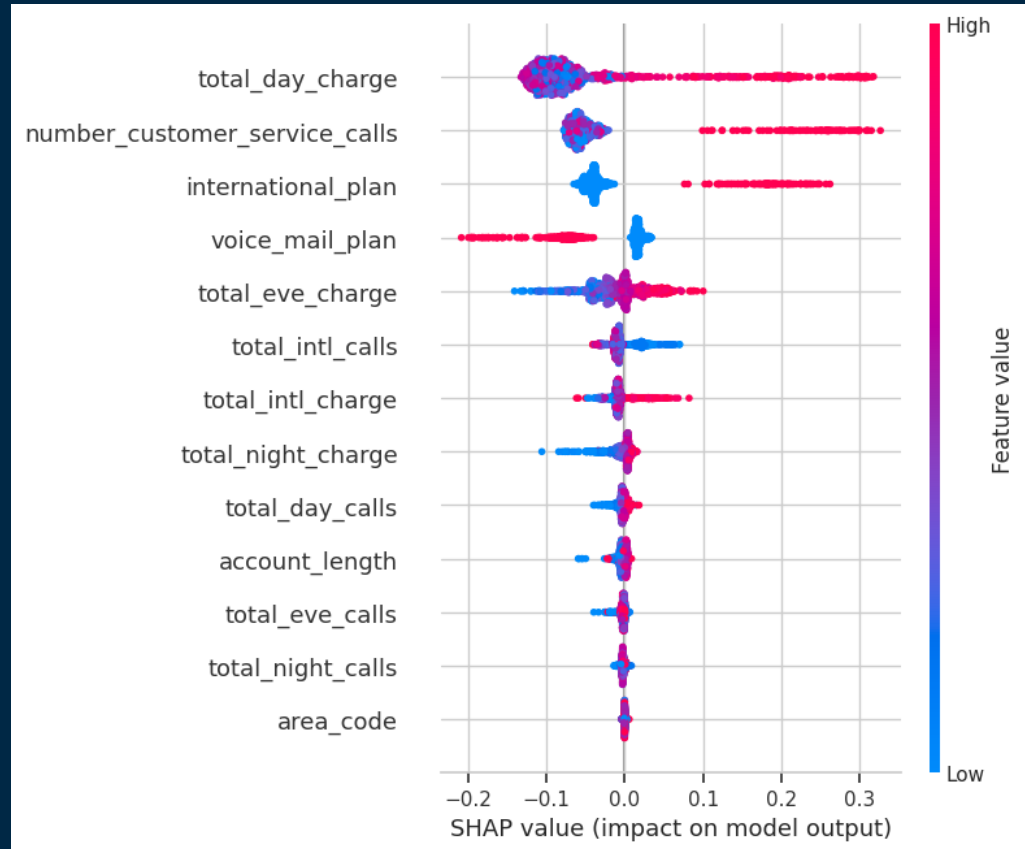
Recall memfokuskan seberapa banyak pelanggan yang sebenarnya churn yang dapat diidentifikasi. Model dapat mengidentifikasi sebagian besar pelanggan yang berpotensi berhenti menggunakan layanan (TP+FP). Hal ini dilakukan agar perusahaan dapat mengambil tindakan pencegahan dan mempertahankan pelanggan. Dari hasil confusion matriks, kesalahan dalam prediksi customer yang tidak churn (FN) juga memiliki nilai paling rendah.



# Feature Importance

Terlihat bahwa total tagihan siang hari, jumlah panggilan pelanggan ke customer service, dan international plan merupakan 3 fitur teratas yang mempengaruhi prediksi bahwa customer akan churn.

Dari hasil plot dapat dilihat juga bahwa fitur tersebut memiliki korelasi yang positif. Artinya semakin besar nilai dari ketiga fitur tersebut maka pelanggan semakin berpotensi untuk churn



# Business Recomendation

04



Perusahaan dapat mengevaluasi dan memperbaiki kualitas layanan. Melakukan survei atau wawancara dengan pelanggan yang telah berhenti berlangganan dapat membantu memahami kelemahan dalam produk atau layanan sehingga dapat diperbaiki.



Perusahaan dapat meningkatkan layanan customer service dan melakukan penyederhanaan proses penyelesaian keluhan pelanggan.



Perusahaan dapat menawarkan paket-paket menarik yang menggabungkan layanan telepon dengan produk lain (misal data internet) atau membuat paket telepon berdasarkan durasi waktu dengan harga yang lebih terjangkau. Hal ini dapat membantu meningkatkan nilai yang dirasakan oleh pelanggan dan mendorong mereka untuk tetap menggunakan layanan.



Perusahaan dapat memberikan paket roaming untuk panggilan internasional dengan biaya yang terjangkau dan meningkatkan kualitas layanan diluar negeri.



# THANKS

VIEW  
CODE

CREDITS: This presentation template was created by [Slidesgo](#),  
including icons by [Flaticon](#), and infographics & images by [Freepik](#)