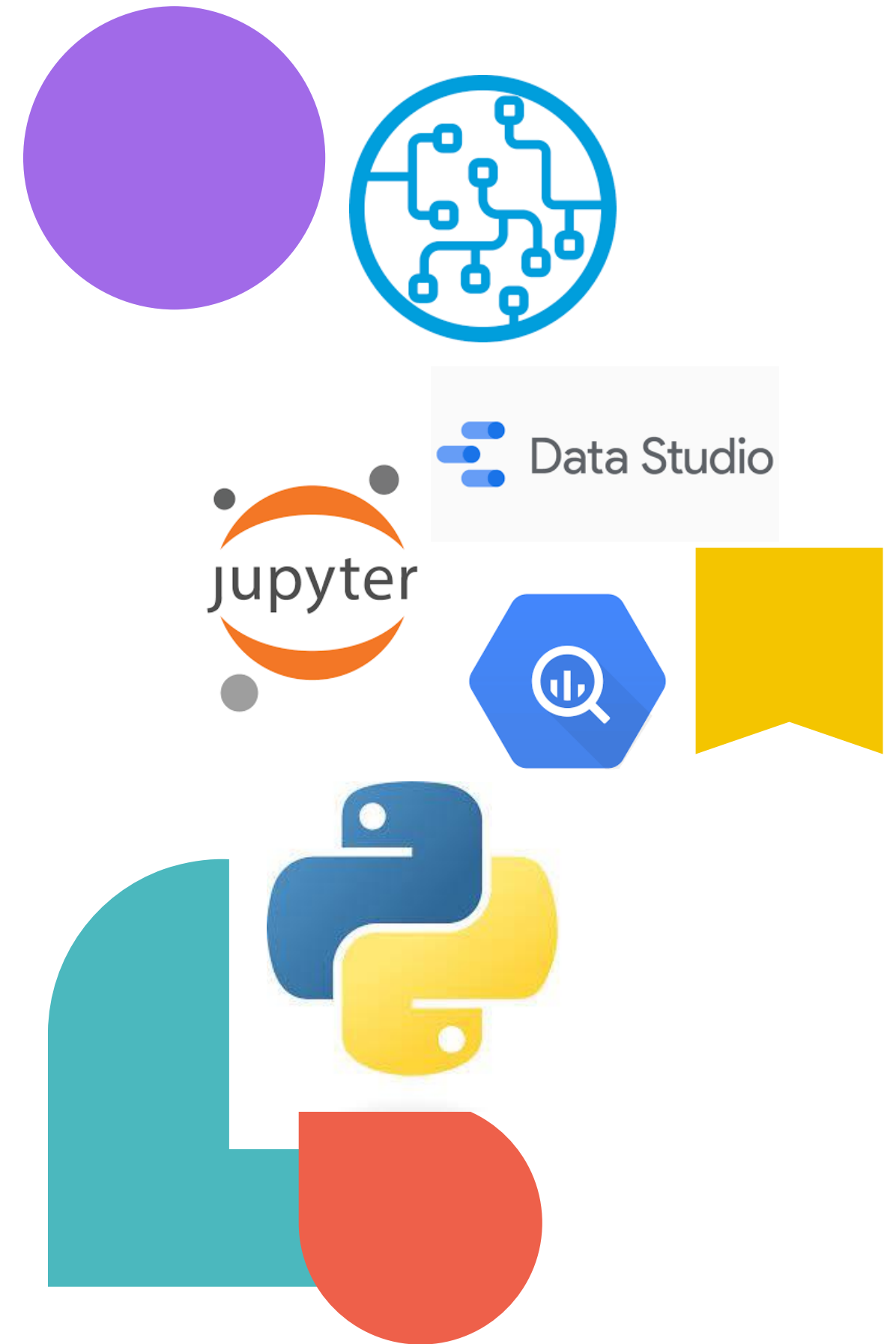
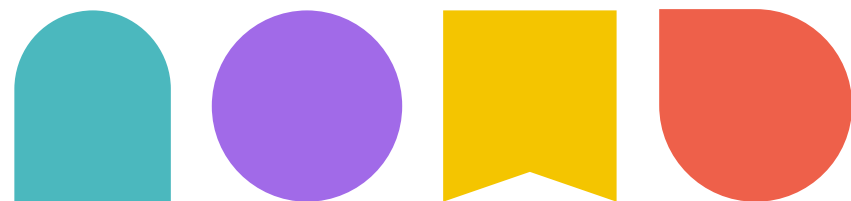
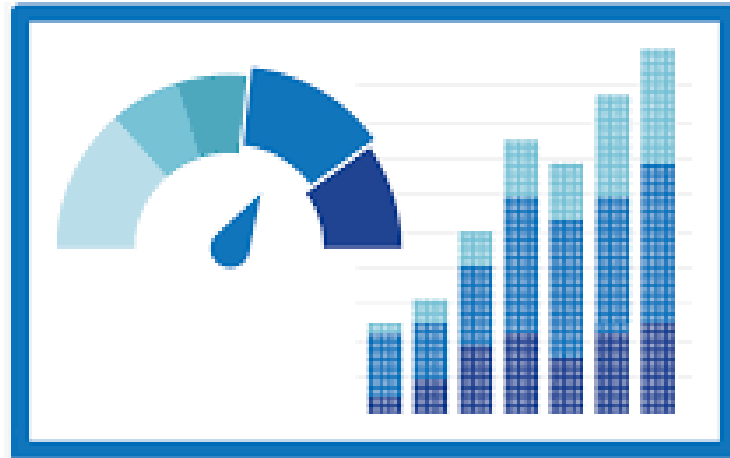


# Maulana Kavaldo – Visualization & Machine Learning

149368779100-121



# Outline



## Challenge 1 - Visualization

Dashboard COVID-19 in Indonesian and function

## Challenge 2 - Machine Learning

Churn Prediction

- Data Preparation
- Data Preprocessing
- Classification Models

# Visualization – Dashboard



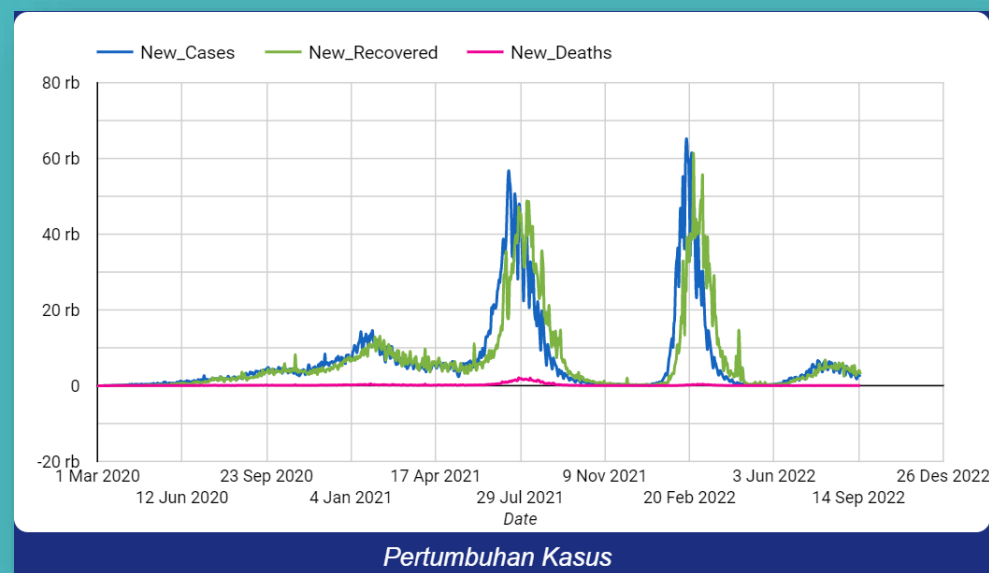
# Visualization – Dashboard



Total dari kasus, total meninggal, total sembuh, tingkat sembuh dan tingkat kematian yang terjadi di seluruh provinsi di Indonesia. Nilai dapat berubah jika ingin melihat provinsi tertentu dan melakukan filter waktu dari tanggal tertentu.



Perkembangan COVID-19 yang terjadi di Indonesia di tahun 2022. Dimana data terakhir menampilkan nilai total kasus, kematian dan sembuh terlihat pada gambar.



Perkembangan COVID-19 mulai dari tahun 2020 s/d 2022. Terjadi lonjakan kasus terbesar pada bulan Juni-Juli 2020 dan bulan Februari pada tahun 2022.

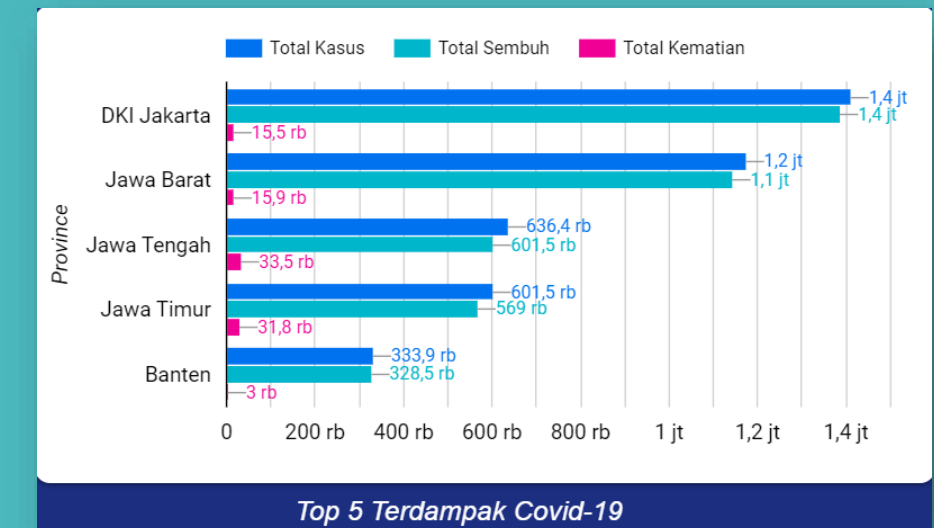


Provinsi di Indonesia dengan warna pudar hingga lebih pekat, semakin pekat warna maka nilai/total kasus semakin tinggi. Untuk grafik yang ditampilkan yaitu, Total Meninggal dan Total Sembuh.

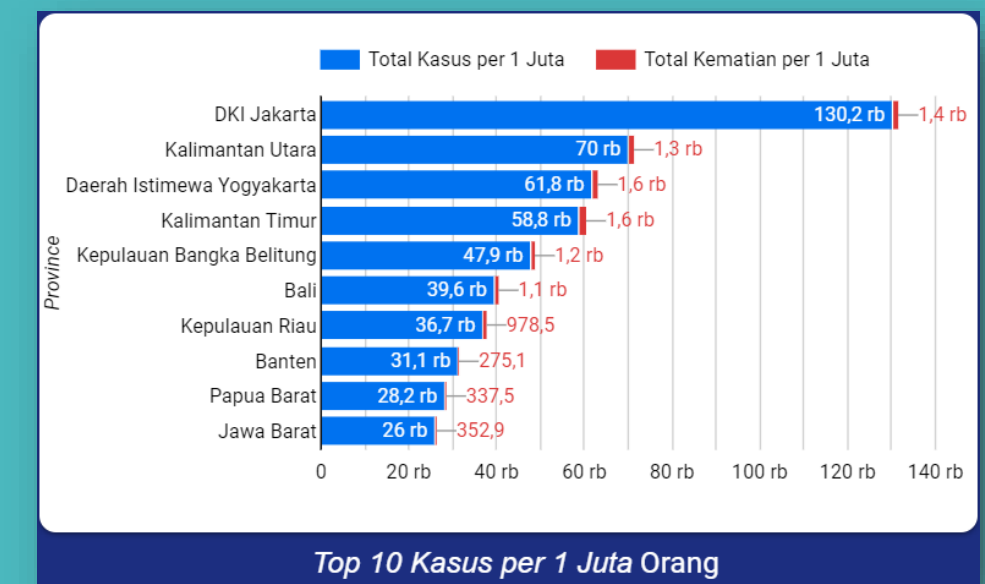
Provinsi	Total Kasus	Total Meninggal	Total Sembuh
1. Aceh	44.038	2.223	41.713
2. Bali	166.831	4.731	161.629
3. Banten	333.875	2.950	328.482
4. Bengkulu	29.173	522	28.617
5. DKI Jakarta	1.412.511	15.513	1.386.134
6. Daerah Istimewa Yogyakarta	224.307	5.928	217.712
7. Gorontalo	13.951	487	13.433
8. Jambi	38.643	889	37.681

**Tabel Data Covid-19 di Indonesia**

Untuk dapat melihat secara rinci data akibat covid-19 dapat melihat tabel data covid. Table data dapat diurutkan sesuai kebutuhan dengan men-hit pada kolom tertentu baik dapat terurut secara menurun atau menaik.



Beberapa provinsi teratas terdampak COVID-19 berdasarkan total kasus. DKI Jakarta menduduki posisi pertama dengan total kasus sebanyak 1.4 juta sekian, total sembuh 1.4 juta dan total kematian 15.5 ribu



Diurutkan berdasar total kasus per 1 juta orang dan DKI Jakarta lagi-lagi menempati posisi pertama dengan total kematian sebanyak 1.4 ribu jiwa..

# Machine Learning



## Churn Prediction



# Churn Prediction

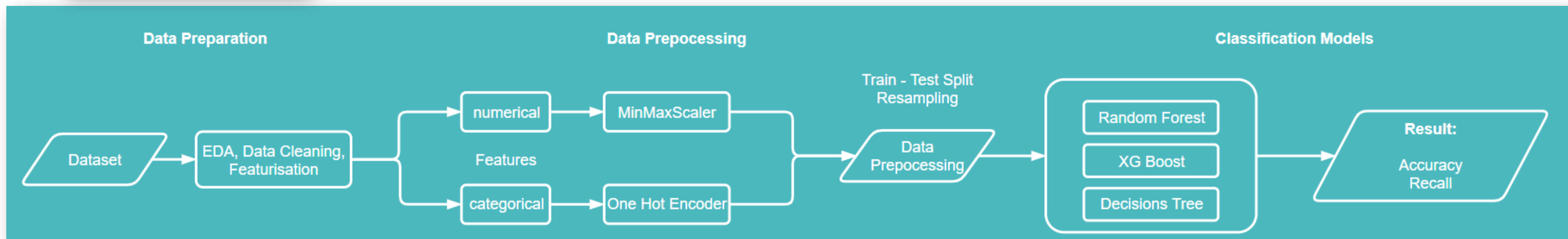
Telecom



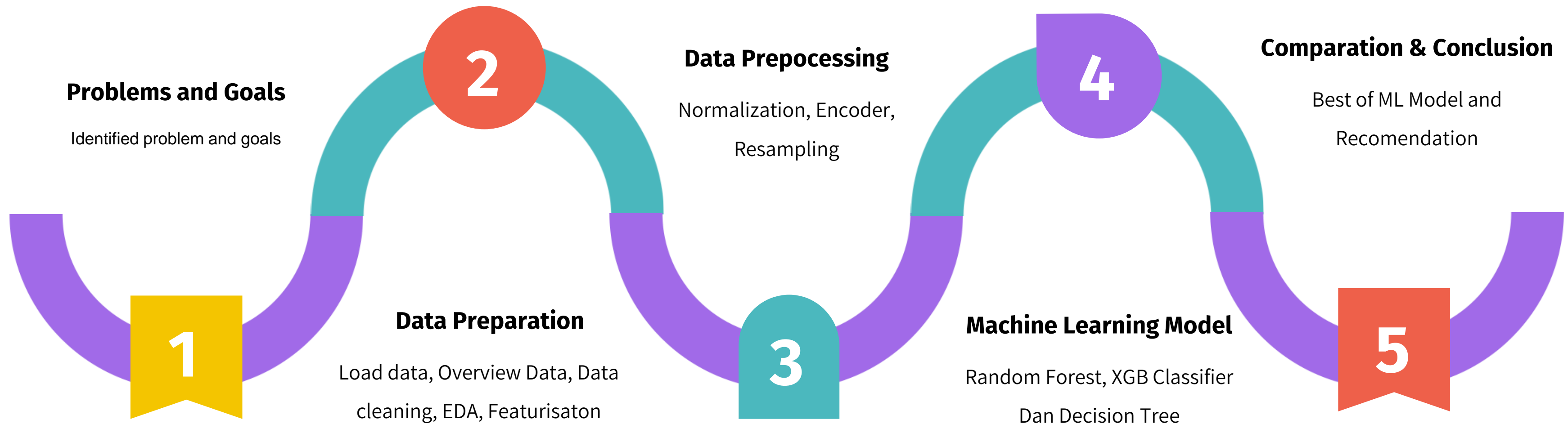
- Random Forest
- XGB Classifier
- Decision Tree

AUC Score

95%



Machine Learning Workflow



## Problem

Peningkatan jumlah jaringan telekomunikasi di pasar menyebabkan peningkatan jumlah pelanggan yang beralih ke jaringan lain.



## Goals

Memprediksi pelanggan dengan potensi churn atau tidak churn secara akurat dan memberikan rekomendasi.



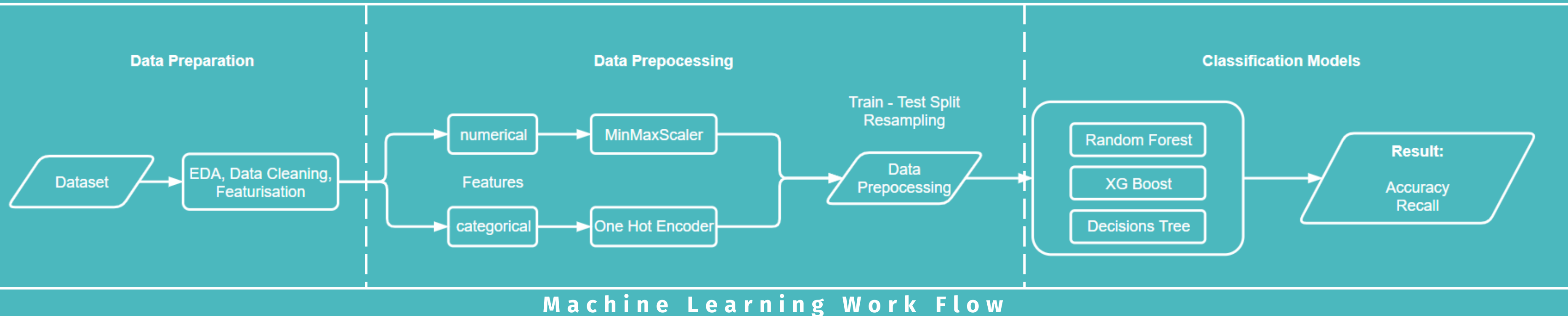


## Dataset Churn

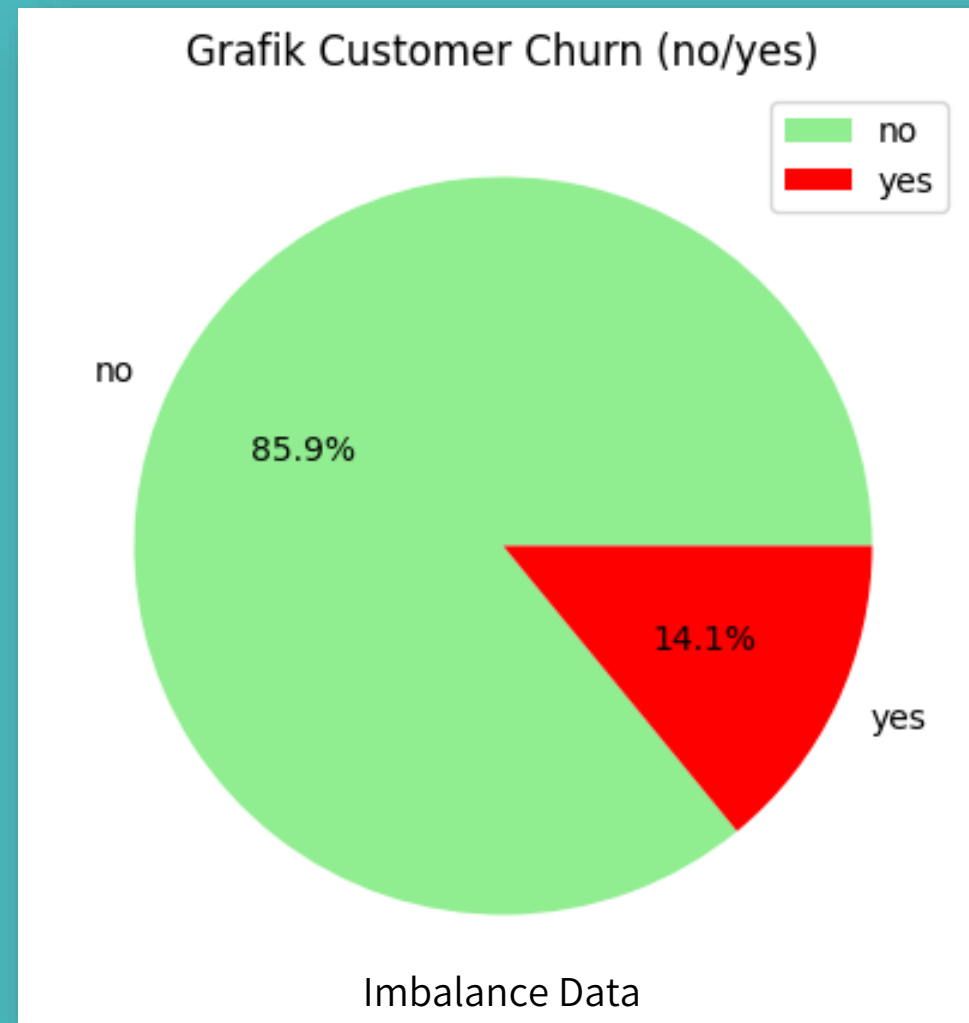
Berisi 4.250 sampel. Terdapat 19 fitur dan 1 variabel boolean "churn" yang menunjukkan kelas sampel. 19 fitur input dan 1 variabel target.

### Features:

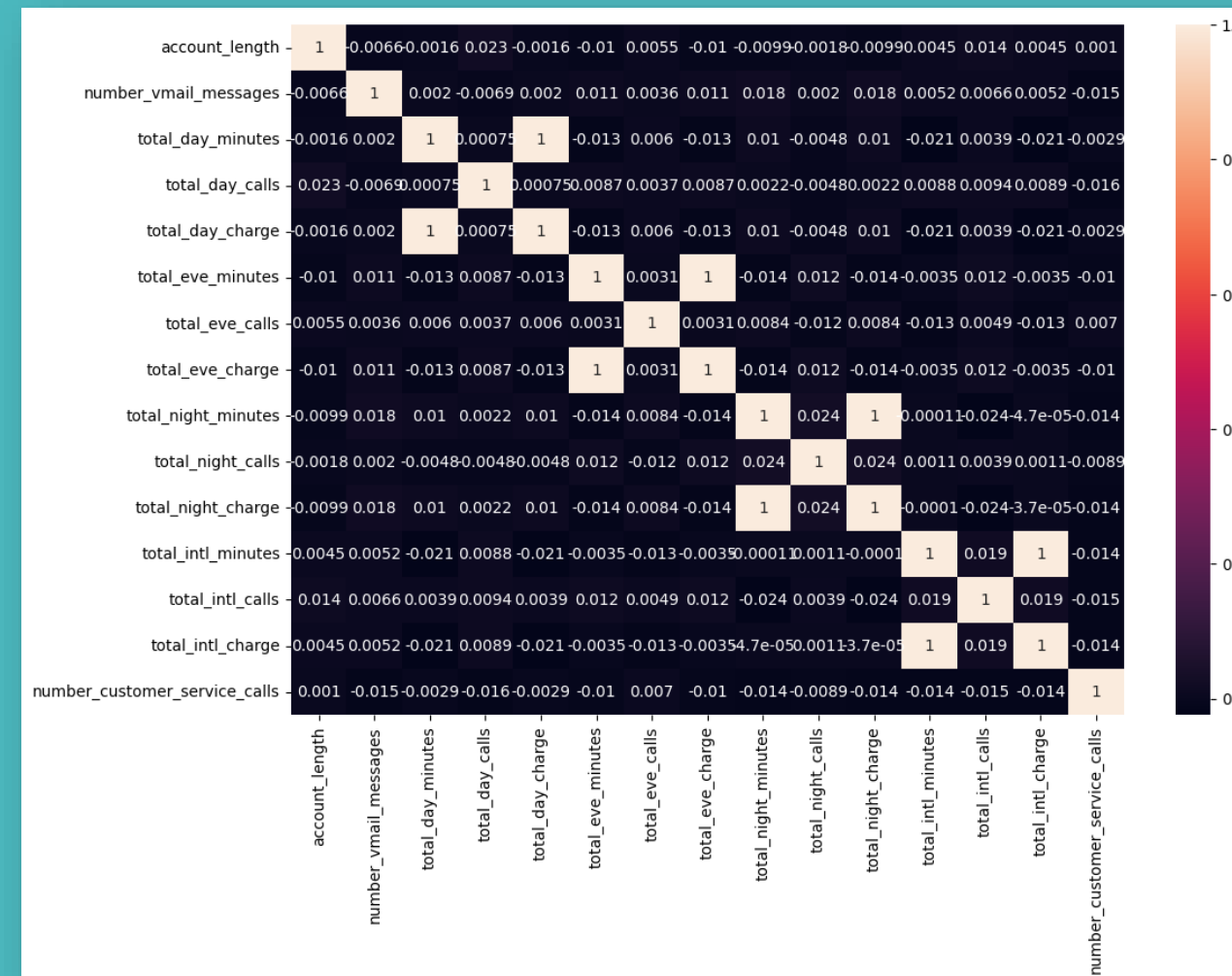
- |                          |                         |                                   |
|--------------------------|-------------------------|-----------------------------------|
| 1) state                 | 8) total_day_calls      | 15) total_night_charge            |
| 2) account_length        | 9) total_day_charge     | 16) total_intl_minutes            |
| 3) area_code             | 10) total_eve_minutes   | 17) total_intl_calls              |
| 4) international_plan    | 11) total_eve_calls     | 18) total_intl_charge             |
| 5) voice_mail_plan       | 12) total_eve_charge    | 19) number_customer_service_calls |
| 6) number_vmail_messages | 13) total_night_minutes | 20) churn                         |
| 7) total_day_minutes     | 14) total_night_calls   |                                   |



# Machine Learning - Exploratory Data Analysis (EDA)

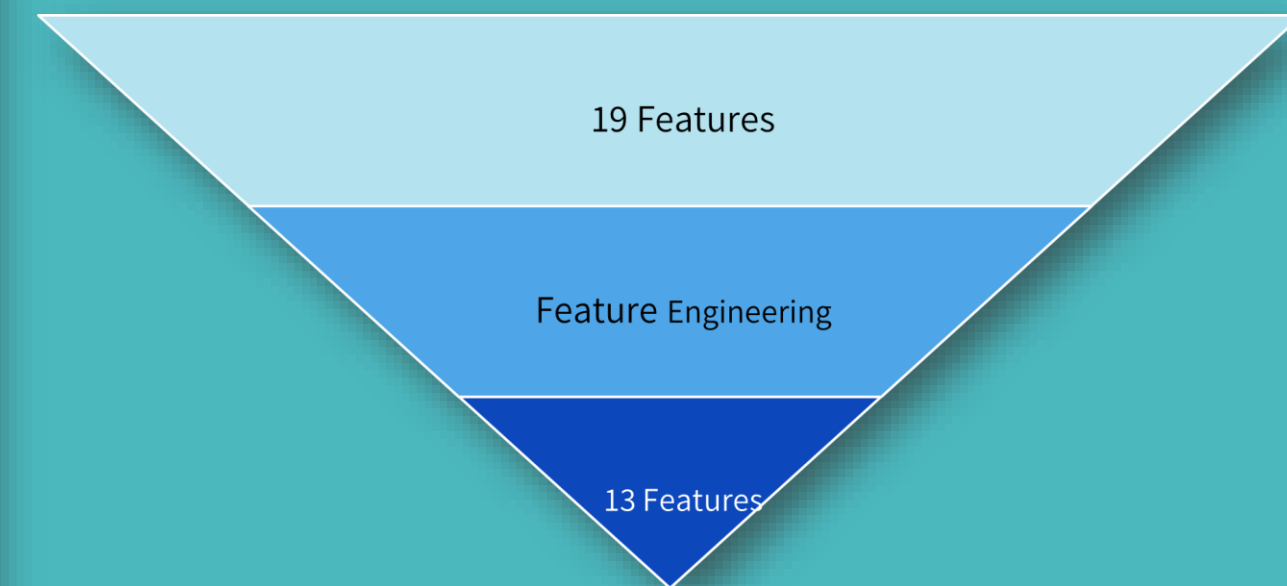


Ratio  
1 : 6

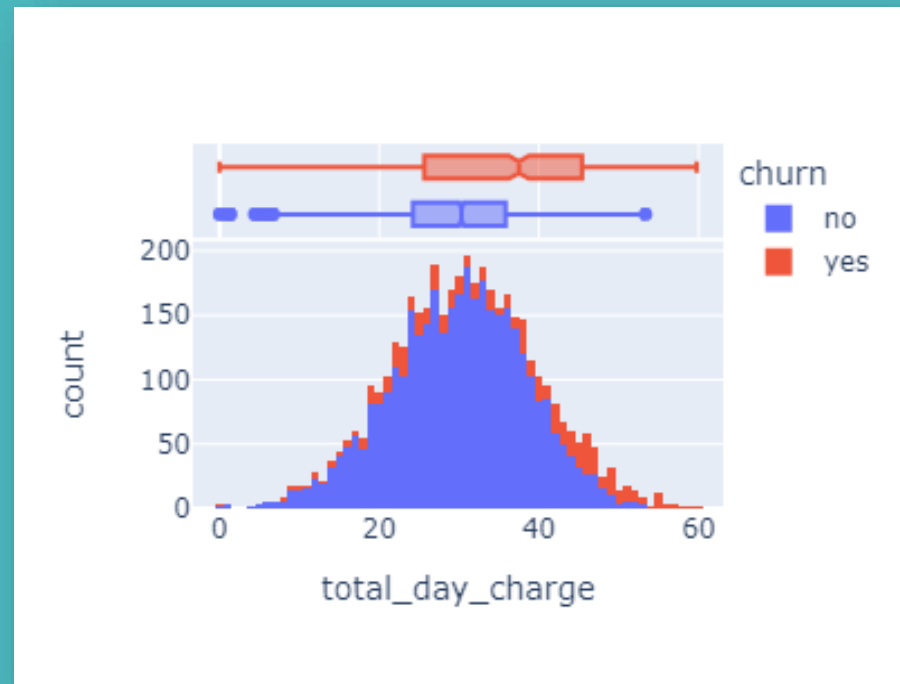


Beberapa fitur memiliki korelasi yang sama,  
sehingga dapat dihilangkan salah satu fitur

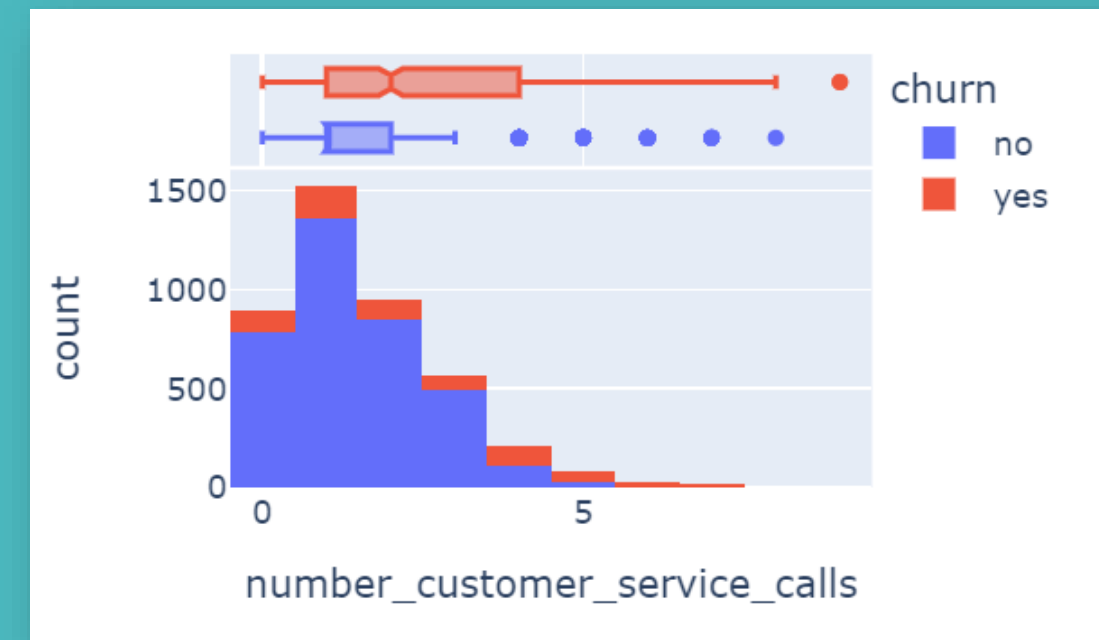
## Features Reduction



# Machine Learning - Exploratory Data Analysis (EDA)



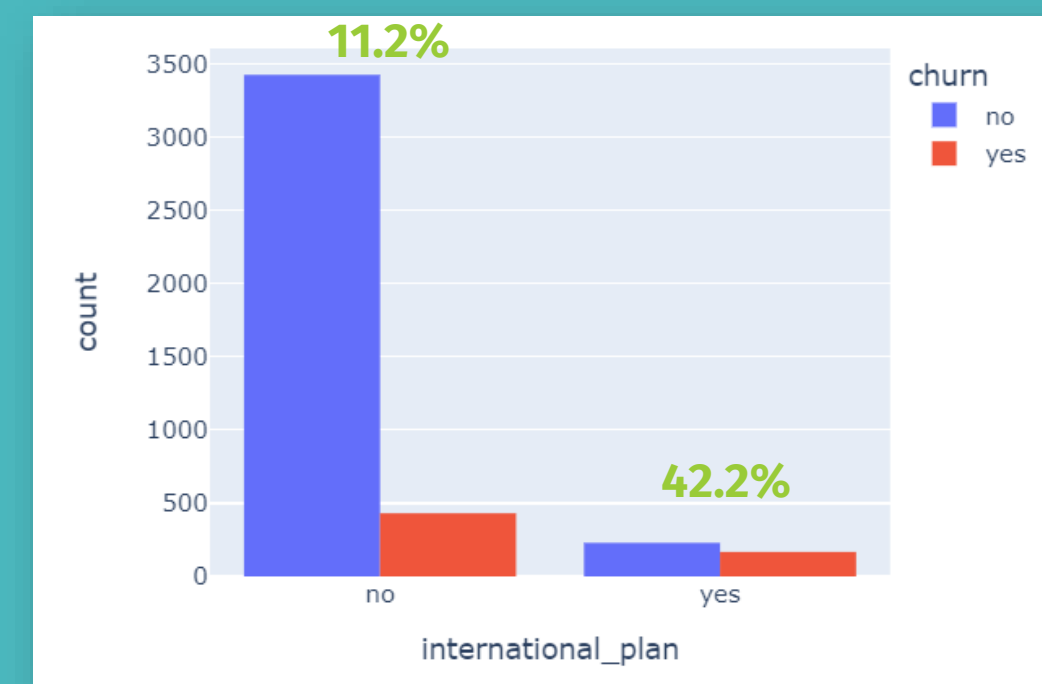
Biaya harian meningkat, churn ikut meningkat.



Churn meningkat seiring dengan meningkatnya customer call.



Customer yang sering melakukan voice mail lebih berpotensi churn



Churn rate pelanggan dengan international plane lebih tinggi

## Numerical Features

### MinMax Scaller

#### Features:

- ['account\_length',
- 'number\_vmail\_messages',
- 'total\_day\_calls',
- 'total\_eve\_calls',
- 'total\_night\_calls',
- 'total\_intl\_calls',
- 'number\_customer\_service\_calls',
- 'total\_day\_minutes',
- 'total\_eve\_minutes',
- 'total\_night\_minutes',
- 'total\_intl\_minutes']

## Categorical Features

### One Hot Encoder

#### Features:

- ['international\_plan',
- 'voice\_mail\_plan']





## Train - Test Split

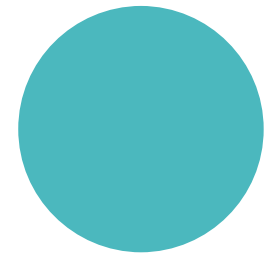
- Original Data  
1 : 8
- Resampling  
1 : 1

**Train**  
**70%**

**Test**  
**30%**



## Models & Metric



### Random Forest

Can overcome noise and missing values and can handle large amounts of data



### XGB Classifier

Parallel processing that can speed up computing, has a high flexibility of setting objectives, built in cross validation, has a regularization feature, and overcomes splits when negative loss.



### Decision Tree

Model simple and easy to understand, can be used to search for all possible decisions, and no need to do a lot of data cleaning or augmentation.

## Metric

Area Under the Curve  
(AUC)

## Comparison Models

Model	Data	Train		Test	
		Accuracy (%)	Recall (%)	Accuracy (%)	Recall (%)
Random Forest	Original Data	100	100	86	72
	Random Under Sampling	95	100	87	86
XGB Classifier	Original Data	100	100	89	78
	Random Under Sampling	96	100	87	85
Decision Tree	Original Data	100	100	82	69
	Random Under Sampling	92	100	81	82

Random Forest

## Model:

Karena data imbalance lebih disarankan model machine learning Random Forest dengan metode resampling (Random Under Sampling).

## Recommendation:

Kategori	Strategi
Potensial to Churn	Memberikan notifikasi secara berkala setiap program promosi yang masih tersedia.
	Menawarkan program anggota untuk membuat mereka loyal.
	Memberikan diskon atau promo khusus pada layanan yang ditargetkan kepada pelanggan yang secara potensial untuk churn.
Loyal	Menerapkan sistem member tier seperti silver, gold dan platinum, setiap grade memiliki manfaat yang berbeda.
	Memberikan notifikasi secara berkala setiap program promosi yang masih tersedia agar mereka tetap setia.



# Thank you!

•  
•

