

zenius

Kampus  
Merdeka  
INDONESIA JAYA

# Final Project Presentation

Nomor Kelompok: 9

Nama Mentor: Rauzan Sumara

Nama:

- Siti Bunga Rohiyatun Nufus
- Zheryl Zabrina Hibatullah

Machine Learning Class

Program Studi Independen Bersertifikat  
Zenius Bersama Kampus Merdeka



# Petunjuk

- Waktu presentasi adalah 5 menit (tentatif, tergantung dari banyaknya kelompok yang mendaftarkan diri)
- Waktu tanya jawab adalah 5 menit
- Silakan menambahkan gambar/visualisasi pada slide presentasi
- Upayakan agar tetap dalam format poin-poin (ingat, ini presentasi, bukan esai)
- Jangan masukkan *code* ke dalam slide presentasi (tidak usah memasukan screenshot jupyter notebook)

- 1. Latar Belakang**
- 2. Explorasi Data dan Visualisasi**
- 3. Modelling**
- 4. Kesimpulan**

# Latar Belakang

# Latar Belakang Project

Sumber Data: <<https://www.kaggle.com/datasets/barun2104/telecom-churn?datasetId=567482>>

Problem: **classification**

Tujuan:

- Melakukan klasifikasi guna memperkirakan dan mengelompokkan kelas pada data Customer churn
- Untuk menganalisis faktor apa saja yang dapat mempengaruhi churn

# Explorasi Data dan Visualisasi

# Business Understanding



Pesatnya perkembangan industri telekomunikasi



Perusahaan berbasis pelanggan berusaha memenuhi kebutuhan bertahan dalam lingkungan kompetitif

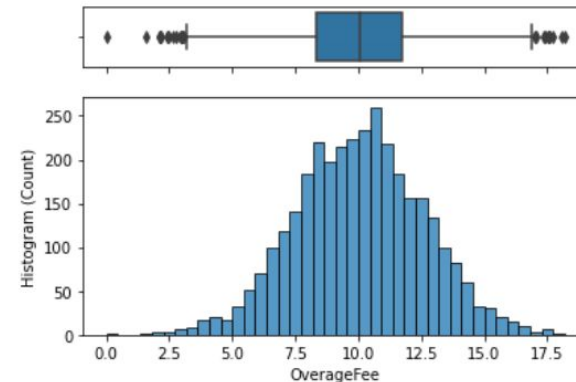
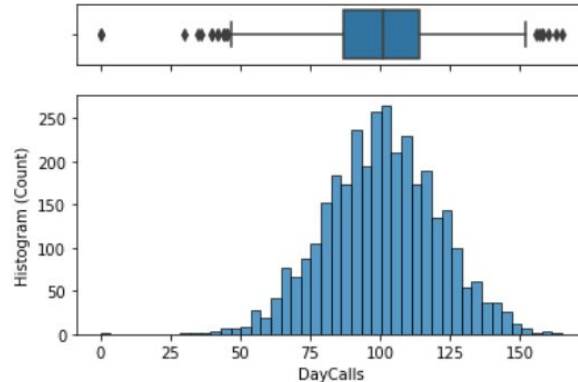
Perusahaan memahami perilaku konsumen dan memprediksi asosiasi pelanggan



Biaya mendapatkan pelanggan baru jauh lebih mahal

# Data Cleansing

- Data Churn memiliki 3333 baris dan 11 kolom tanpa adanya data missing maupun data duplikat

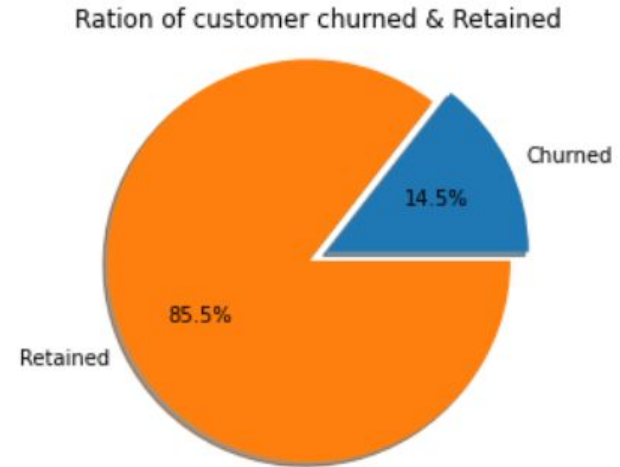


**Terdapat outlier juga pada kolom 'DayCalls' dan 'Overage Fee', sehingga outlier tersebut dihapus**



# Exploratory Data Analysis

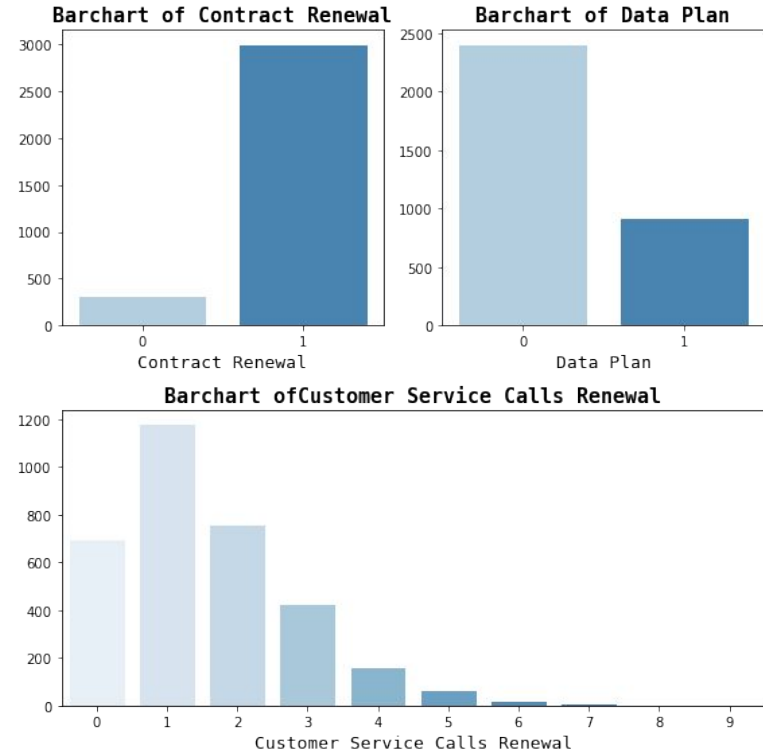
Perbandingan customer yang membatalkan layanan telekomunikasi lebih kecil dibandingkan dengan yang masih tetap memakai servis telekomunikasi. Sehingga perusahaan masih bisa mempertahankan customer yang sudah ada.



# Exploratory Data Analysis

Berdasarkan bar chart disamping, diketahui bahwa :

- Customer yang memperbarui kontrak dengan layanan telekomunikasi lebih banyak sekitar 3000 orang dibanding yang tidak memperbarui kontrak.
- Customer lebih banyak tidak memakai paket data yang berjumlah sekitar 2400
- Customer paling banyak melakukan panggilan customer service dalam 1 kali panggilan



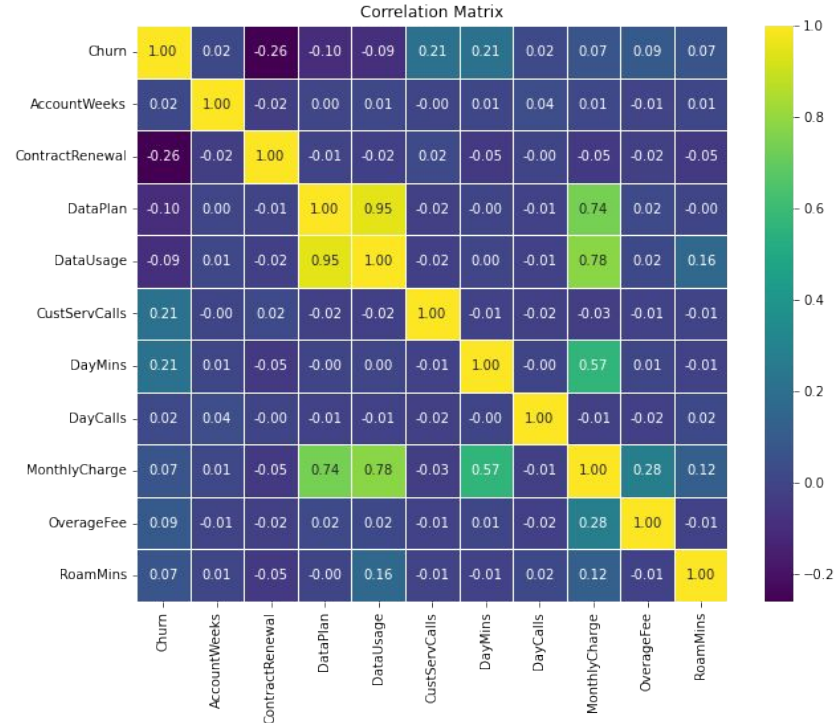
# Exploratory Data Analysis

Korelasi paling kuat antara variabel :

- Data Usage dengan Data Plan
- Data plan dengan Monthly Charge
- Data usage dengan Monthly Charge

Korelasi cukup kuat antara variabel

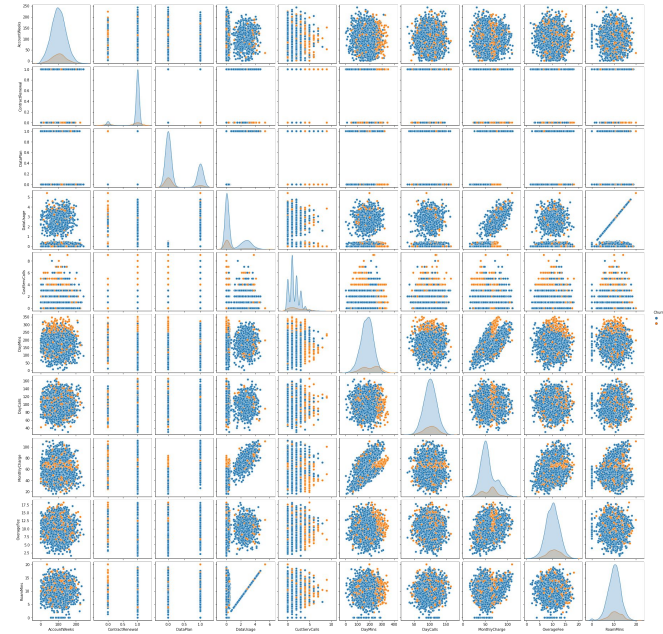
- DayMins dengan Monthly Charge



# Exploratory Data Analysis

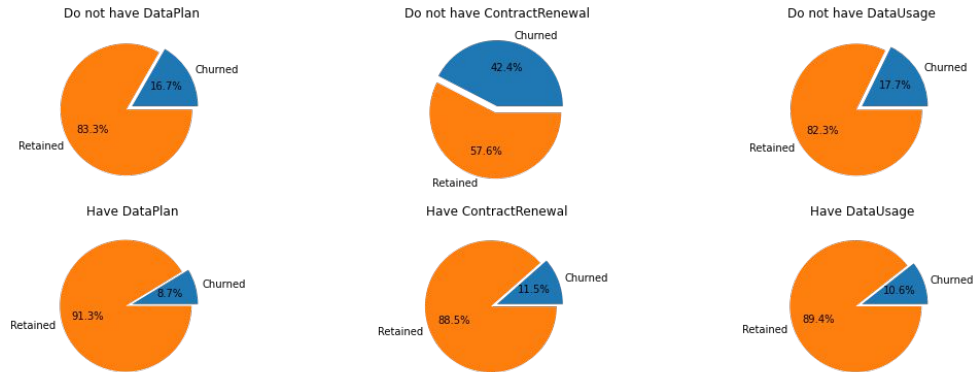
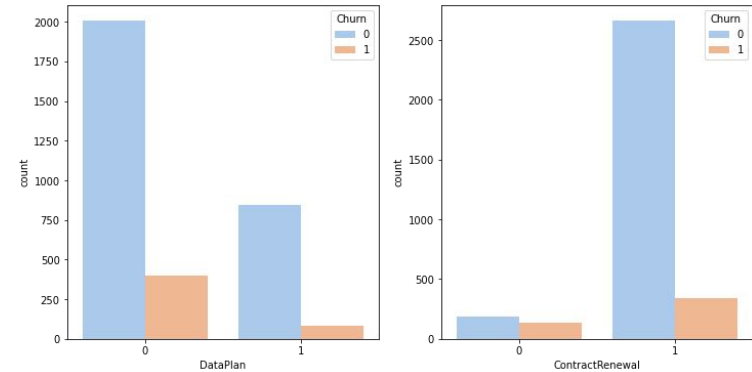
Berdasarkan hasil pairplot antar variabel juga didapat :

- DataUsage memiliki hubungan linier dengan RoamMins untuk pelanggan yang tidak melakukan churn.
- MonthlyCharge memiliki hubungan positif dengan DataUsage, DayMins, dan OverageFee



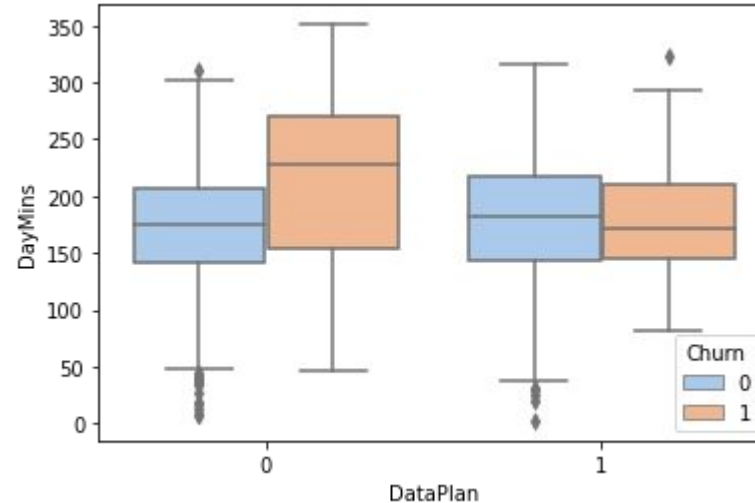
# Exploratory Data Analysis

- Pelanggan yang tidak mempunyai data plan lebih banyak melakukan churn dibanding yang mempunyai data plan
- Pelanggan yang tidak melakukan contract renewal lebih banyak melakukan churn dibanding yang tidak melakukan contract renewal
- Pelanggan yang tidak menggunakan data sama sekali memiliki tingkat churn 7% lebih besar. Ini bisa membantu kita dalam prediksi



# Exploratory Data Analysis

- Rata-rata, orang yang tidak membeli paket data dan DayMinsnya sangat tinggi sehingga sangat mungkin pelanggan untuk melakukan churn.



# Modelling



# Classification dengan Random Forest dan Decision Tree

- Untuk menjalankan model dan prediksi menggunakan metode random forest dan decision tree, membagi train set dan test set sebanyak 80 : 20, 85:15, dan 90:10. Sehingga terdapat 6 model yang dihasilkan. Dimana train set digunakan untuk melakukan fitting model dan test set digunakan untuk prediksi.
- Base line model
  - y train:** Churn
  - x train :** AccountWeeks, ContractRenewal, DataPlan, DataUsage, CustServCalls, DayMins, DayCalls, MonthlyCharge, OverageFee, RoamMins
- Model awal yang di coba yaitu model dengan pembagian train set dan test set sebanyak 80:20
- Jenis model lain yang dicoba yaitu model dengan pembagian train set dan test set sebanyak 85:15 dan 90:10



# Classification dengan Random Forest dan Decision Tree

Berikut merupakan hasil yang didapatkan:

Model	Metode	AUC	F1 Score	Accuracy
80:10	Random Forest	0.8188899097895089	0.97	0.948948948948949
	Decision Tree	0.8218760441029067	0.95	0.9174174174174174
85:15	Random Forest	0.7945814220183486	0.97	0.944
	Decision Tree	0.7847620412844037	0.94	0.892
90:10	Random Forest	0.8451782121299223	0.97	0.9519519519519519
	Decision Tree	0.8117275079045703	0.94	0.8948948948948949

# Classification dengan Random Forest dan Decision Tree

Berdasarkan tabel tersebut dapat diketahui bahwa:

- Pada model 80:20 dengan metode random forest dan decision tree menghasilkan nilai akurasi sebesar 95% dan 92%.
- Pada model 85:15 dengan metode random forest dan decision tree menghasilkan nilai akurasi sebesar 94% dan 89%.
- Pada model 90:10 dengan metode random forest dan decision tree menghasilkan nilai akurasi sebesar 95,2% dan 89,5%.

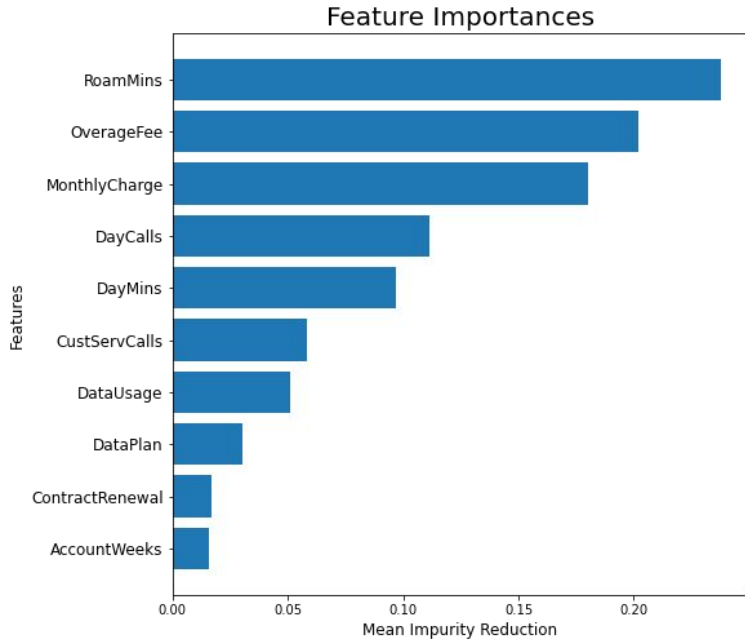
# Conclusion

# Model Terbaik dan Faktor yang Mempengaruhi Churn

Model terbaik:

Berdasarkan evaluasi model yang telah didapatkan diketahui bahwa nilai accuracy tertinggi yaitu pada model dengan perbandingan train set dan test set sebesar 90:10. Dengan menggunakan metode Random Forest yaitu didapati nilai accuray sebesar 95% yang berarti ketepatan model sebesar 95%. Dan nilai AUC (*Area Under Curve*) sebesar 0.85. Serta nilai f1 Score dengan base 0 yaitu sebesar 0.97.

# Model Terbaik dan Faktor yang Mempengaruhi Churn



Faktor yang mempengaruhi:

Berdasarkan gambar disamping diketahui bahwa faktor yang paling berpengaruh terhadap nilai Churn yaitu RoamMins, sehingga dapat disimpulkan bahwa untuk mengurangi nilai Churn kita harus mengurangi RoamMins pada data Customer Churn.

# Saran untuk stakeholder

- Perusahaan atau industri telekomunikasi yang berbasis pelanggan untuk memenuhi kebutuhannya agar bertahan dalam lingkungan yang kompetitif, mereka dapat lebih memperhatikan tiga variabel yang paling mempengaruhi Churn yaitu variabel RoamMins, OverageFee, dan Monthly Charge.
- Perusahaan juga bisa melakukan inovasi dengan adanya pembelian paket data untuk Call Service jika melakukan panggilan yang lama supaya monthly charge pelanggan nantinya tidak akan semakin besar dan minat pelanggan untuk retained semakin besar.

# Terima kasih!

Ada pertanyaan?

zenius



Kampus  
Merdeka  
INDONESIA JAYA