

zenius

Kampus
Merdeka
INDONESIA JAYA

Final Project Presentation

Nomor Kelompok: 07

Nama Mentor: Diajeng Ciptaning Ayu

Nama:

- Nurfadilah Araswati
- Zahra Firdianti Zeny

Machine Learning Class

Program Studi Independen Bersertifikat
Zenius Bersama Kampus Merdeka



- 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 mamperkirakan dan mengelompokkan kelas pada data Customer churn
- Untuk menganalisis faktor apa saja yang dapat mempengaruhi churn

Explorasi Data dan Visualisasi

Business Understanding



Perusahaan berbasis pelanggan berusaha memenuhi kebutuhan bertahan dalam lingkungan kompetitif



Pesatnya perkembangan industri telekomunikasi

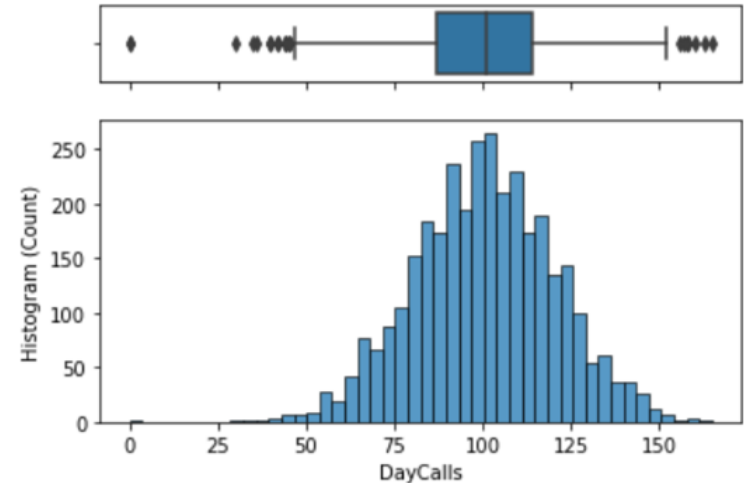
Perusahaan memahami perilaku konsumen dan memprediksi asosiasi pelanggan

Biaya mendapatkan pelanggan baru jauh lebih mahal

Data Cleansing

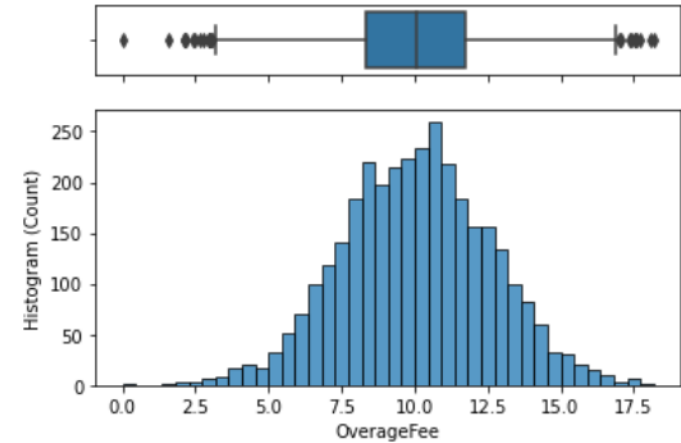
- Data Churn memiliki 3333 baris dan 11 kolom tanpa adanya data missing maupun data duplikat

Data Churn dibutuhkan Proses Data Cleansing, dikarenakan pada kolom DayCalls terdapat outlier dimana terdeteksi 2 customer yang memiliki rata-rata telpon harian = 0, namun sudah berlangganan selama beberapa minggu dan membayar sewa bulanan. Oleh karena itu, kita dapat menghilangkan outlier tersebut



Data Cleansing

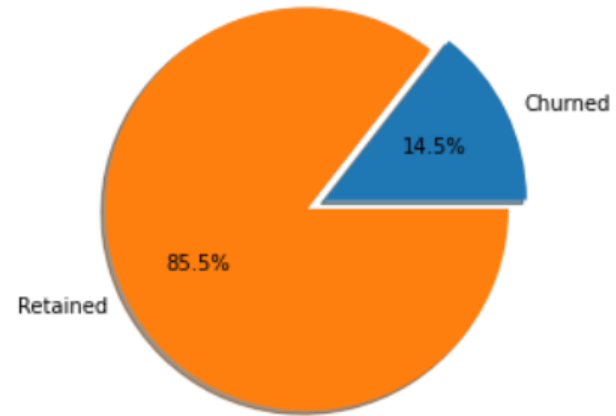
Terdapat outlier juga pada kolom 'Overage Fee' dimana terdapat 1 customer yang memiliki biaya langganan terbesar yang dikeluarkan selama berlangganan = 0, padahal customer sudah berlangganan selama beberapa minggu dan membayar sewa bulanan. Oleh karena itu, kita menghilangkannya outlier tersebut



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.

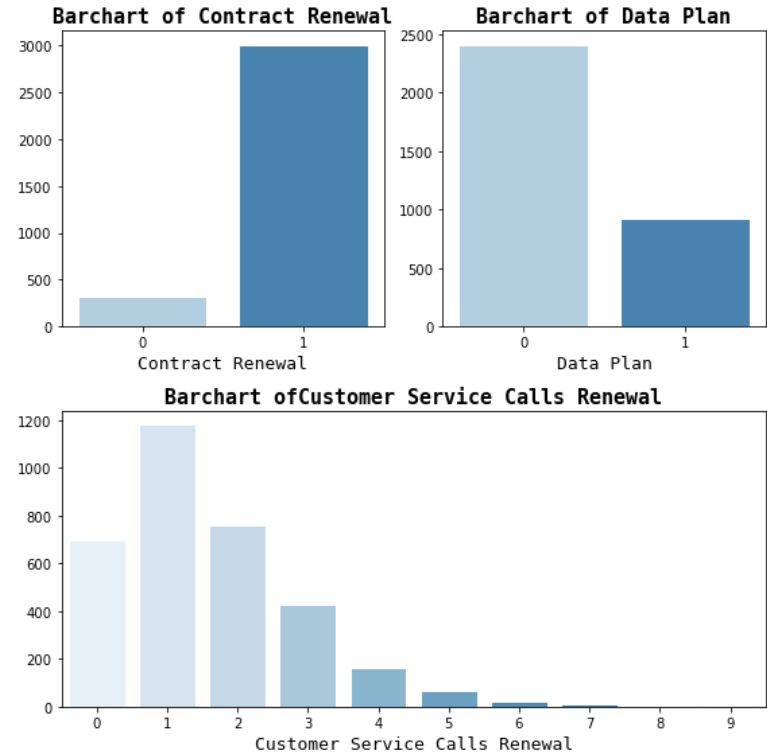
Ration of customer churned & Retained



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



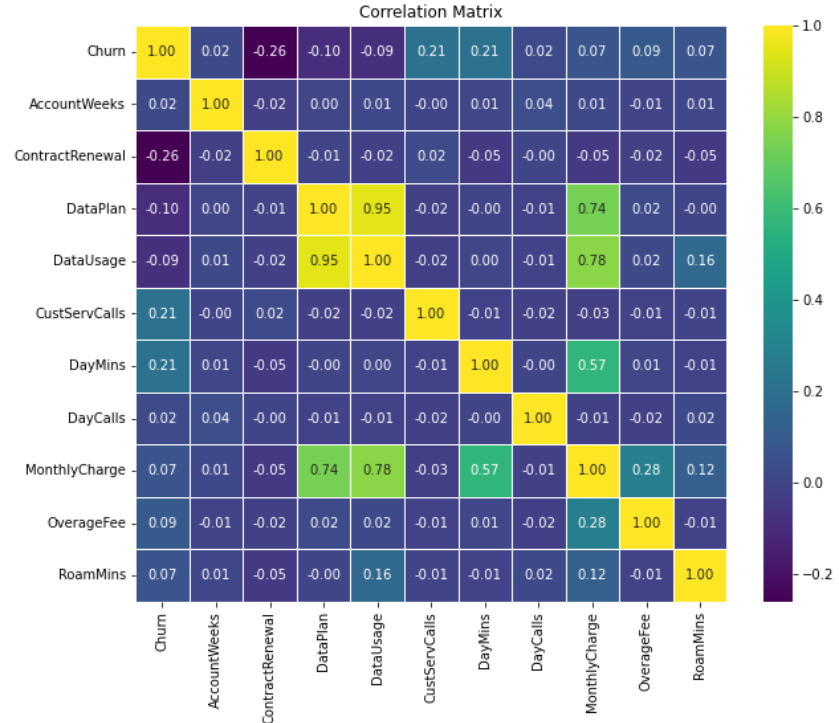
Exploratory Data Analysis

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

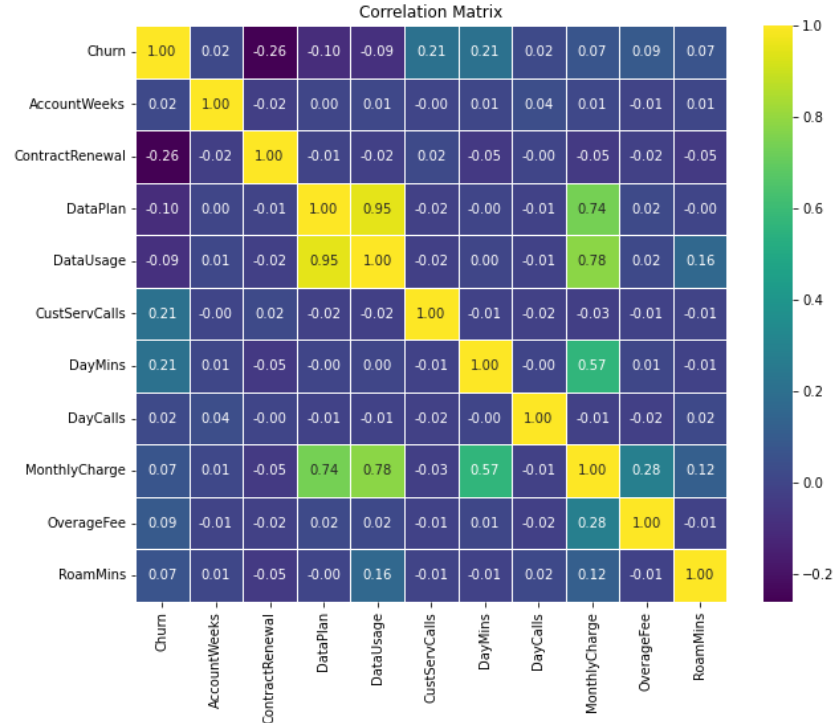
- DayMins dengan Monthly Charge



Exploratory Data Analysis

Korelasi paling kuat antara variabel :

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



Modelling



Classification dengan Random Forest dan Decision Tree

hal-hal yang disarankan untuk dipresentasikan:

- Untuk menjalankan model dan prediksi menggunakan metode random forest dan decision tree, membagi train set dan test set sebanyak 80 : 20, 80:15, dan 90:10. Sehingga terdapat 6 model yang dihasilkan. Dimana train set digunakan untuk melakukan fiiting model dan test set digunakan untuk prediksi.
- Jenis model awal yang dicoba
- Jenis model lain yang turut dicoba, serta tindakan-tindakan apa saja yang dilakukan untuk mencoba menambah akurasi model (hyperparameter tuning, dst)
- Model final
- Kolom-kolom apa saja yang menjadi prediktor dan target variable untuk model final

Model	Metode	AUC	Accuracy
20:80	Random Forest	0.8188899097895089	0.948948948948949
	Decision Tree	0.8080730036752424	0.9024024024024024
15:85	Random Forest	0.7945814220183486	0.944
	Decision Tree	0.7882024082568807	0.898
10:90	Random Forest	0.8451782121299223	0.9519519519519519
	Decision Tree	0.8254527162977867	0.9039039039039038

Conclusion

- Metode yang cocok atau rekomendasi untuk data ini adalah metode random forest, karena memiliki nilai accuracy dan AUC yang lebih tinggi dibandingkan decision tree.

**Terima
kasih!**
Ada pertanyaan?

zenius



**Kampus
Merdeka**
INDONESIA JAYA

