

zenius

Kampus
Merdeka
INDONESIA JAYA

Final Project Presentation

Nomor Kelompok: 5

Nama Mentor: Erwin Fernanda

Nama:

- Sekar Puspita Sari
- Arijuddin Muhammad Rahmatul Qalbi

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:

- Untuk mengetahui faktor yang mempengaruhi Customer Churn
- Untuk memprediksi Customer tersebut akan churn atau tidak

Explorasi Data dan Visualisasi

Business Understanding

Topik : Customer Churn Classification on Telco Sampled Data

Customer Churn adalah situasi dimana Customer berpindah atau berhenti berlangganan per periode bulan terakhir

Dalam beberapa tahun terakhir, Customer Churn terjadi akibat:

- 1. Pembukaan pasar kepada kompetitor baru**
- 2. Layanan dan teknologi baru**

Customer Churn dapat menyebabkan kerugian besar terhadap perusahaan telekomunikasi sehingga harus diantisipasi

Sumber: Hussein Mouselhy (2019)

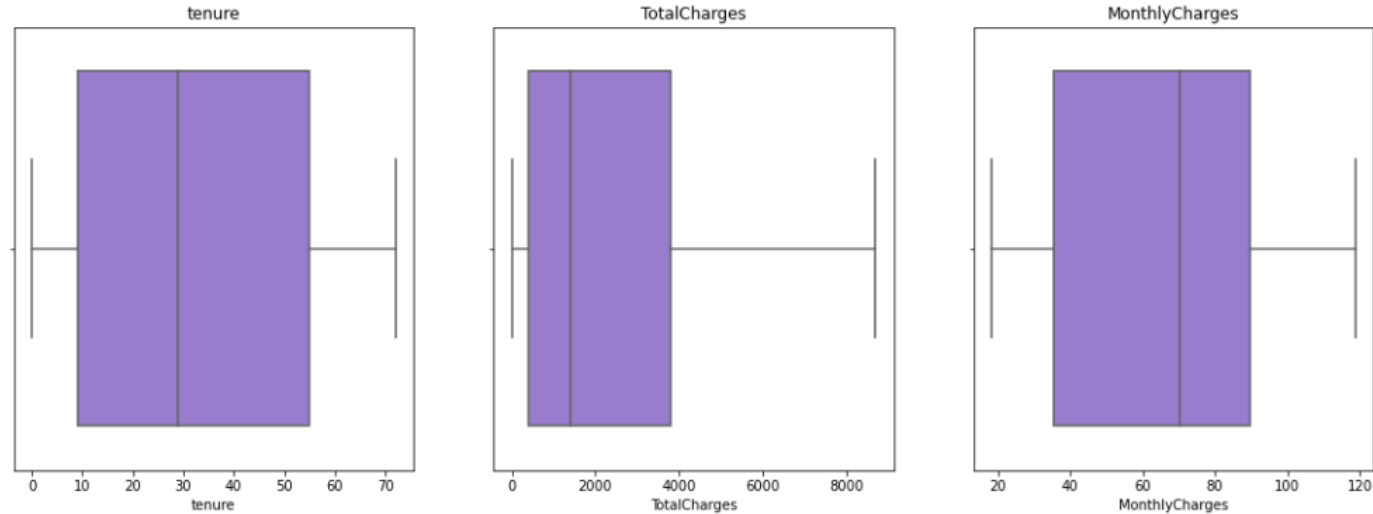
Data Cleansing

- Dimensi data pada dataset yang digunakan yaitu 7.423 baris dan 21 kolom
- Setelah mengecek jumlah data kosong pada dataset, tidak ditemukan adanya data kosong.
- Ditemukan kolom yang seharusnya numerik tetapi tipe datanya object (Total Charges) dan juga sebaliknya (Senior Citizen)
- Setelah dicari, ternyata ditemukan adanya NaN value pada kolom "TotalCharges"
- Agar menjadi tipe data numerik, maka NaN Values diganti dengan angka 0
- Kolom-kolom yang dianggap tidak berpengaruh terhadap Customer Churn dihapuskan.

#	Column	Non-Null Count		Dtype
0	customerID	7043	non-null	object
1	gender	7043	non-null	object
2	SeniorCitizen	7043	non-null	int64
3	Partner	7043	non-null	object
4	Dependents	7043	non-null	object
5	tenure	7043	non-null	int64
6	PhoneService	7043	non-null	object
7	MultipleLines	7043	non-null	object
8	InternetService	7043	non-null	object
9	OnlineSecurity	7043	non-null	object
10	OnlineBackup	7043	non-null	object
11	DeviceProtection	7043	non-null	object
12	TechSupport	7043	non-null	object
13	StreamingTV	7043	non-null	object
14	StreamingMovies	7043	non-null	object
15	Contract	7043	non-null	object
16	PaperlessBilling	7043	non-null	object
17	PaymentMethod	7043	non-null	object
18	MonthlyCharges	7043	non-null	float64
19	TotalCharges	7043	non-null	object
20	Churn	7043	non-null	object

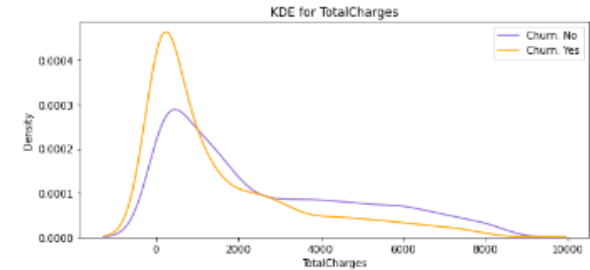
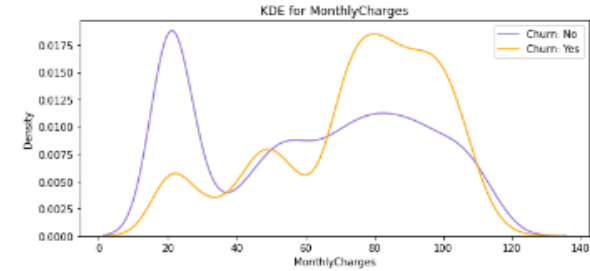
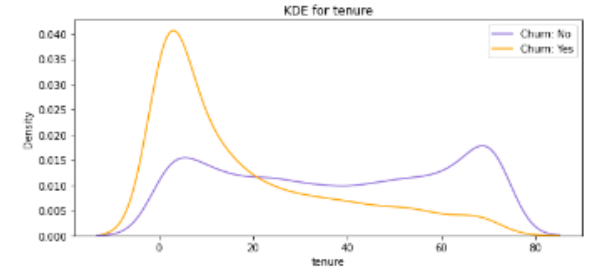
Data Cleansing

- Checking Anomaly Data



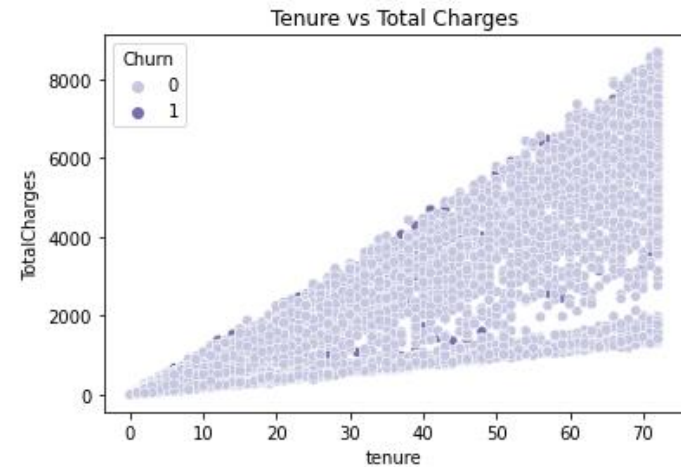
Exploratory Data Analysis

KDE Plot Tenure, MonthlyCharges, TotalCharges terhadap Churn



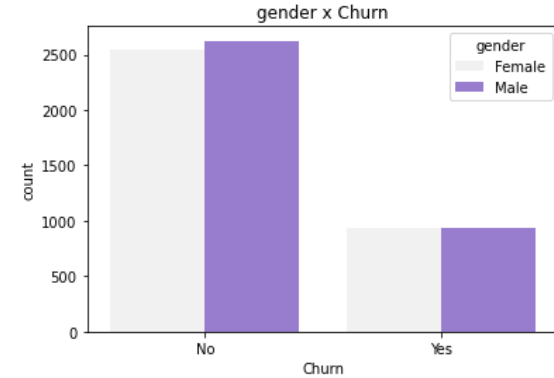
Exploratory Data Analysis

Hubungan Tenure dengan TotalCharges



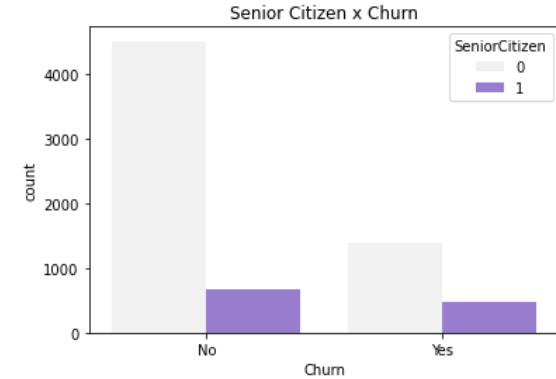
Exploratory Data Analysis

Hubungan Gender dan Churn



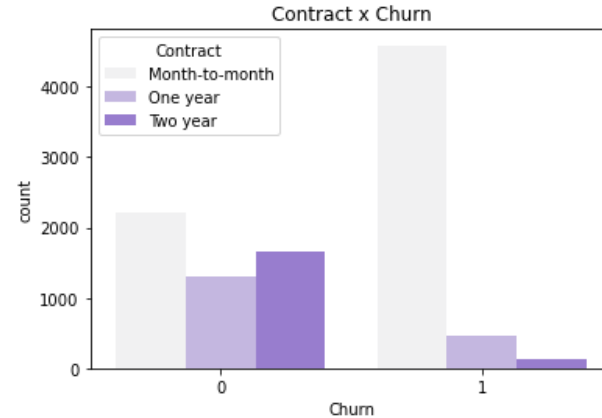
Exploratory Data Analysis

Hubungan Senior Citizen dan Churn



Exploratory Data Analysis

Hubungan Contract dan Churn



Exploratory Data Analysis

Hubungan Payment Method dan Churn

PaymentMethod	Bank transfer (automatic)	Credit card (automatic)	Electronic check	Mailed check	All
Churn					
No	1286	1290	1294	1304	5174
Yes	258	232	1071	308	1869
All	1544	1522	2365	1612	7043

Modelling



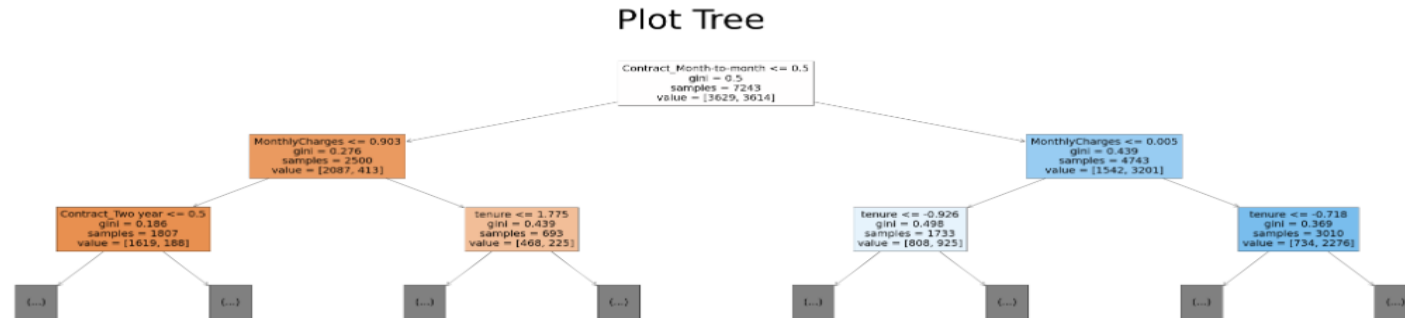
Prediction Model Customer Churn

Train Test split

- Train test split dilakukan dengan 70% data train dan 30% data test yang sebelumnya telah di-Oversampling

Model Awal

- Model awal yang digunakan adalah Decision Tree



Prediction Model Customer Churn

Model Lain

- Model yang turut digunakan yaitu Random Forest, Logistic Regression dan Random Forest dengan Feature Selection dengan melakukan hyperparameter tuning pada ketiga model tersebut.

Metrik Evaluasi

- Metrik yang digunakan yaitu Accuracy, Precision, Recall dan AUC

	Dtree	LR	RF	RF(Feat. Select)
Accuracy	0.8718196457326892	0.7529790660225443	0.7539452495974235	0.7198067632850241
Precision	0.8820911633195556	0.7542761205802153	0.7606060786976072	0.7436592923662283
Recall	0.8714245705750561	0.7528005974607916	0.753553647000249	0.7190440627333832
AUC	0.8566529748568583	0.7569952701020662	0.7645164301717701	0.7251338063231266

Prediction Model Customer Churn

Model Akhir

- Model Akhir yang digunakan yaitu Decision Tree Model

Feature yang digunakan

- Kolom-kolom yang digunakan pada model Machine Learning yaitu:

['SeniorCitizen', 'Partner', 'Dependents', 'tenure', 'PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport', 'StreamingTV', 'StreamingMovies', 'PaperlessBilling', 'MonthlyCharges', 'TotalCharges', 'Churn', 'Contract_Month-to-month', 'Contract_One year', 'Contract_Two year', 'PaymentMethod_Bank transfer (automatic)', 'PaymentMethod_Credit card (automatic)', 'PaymentMethod_Electronic check', 'PaymentMethod_Mailed check']

Conclusion

Conclusion & Recommendation

Conclusion

- Gender tidak berpengaruh pada Customer Churn
- Customer yang berlangganan lebih lama cenderung lebih setia
- Customer yang berlangganan Month to month cenderung lebih gampang untuk churn
- Semakin lama customer berlangganan maka total pembayaran semakin tinggi
- Customer yang Monthly Charges-nya kecil cenderung lebih setia
- Berdasarkan beberapa evaluasi yang dilakukan, model yang paling optimal adalah Decision Tree
- Customer yang bukan Senior Citizen cenderung lebih gampang untuk churn
- Customer yang memilih electronic check payment method memiliki rasio churn yang tinggi

Recommendation

- Untuk mengurangi Churn Rate, perusahaan seharusnya berusaha mempengaruhi customer agar mengambil kontrak yang lebih panjang.
- Perusahaan berusaha menggiring customer untuk mengambil Payment Method selain Electronic check

**Terima
kasih!**
Ada pertanyaan?

zenius



**Kampus
Merdeka**
INDONESIA JAYA

