



PREDIKSI CUSTOMER CHURN

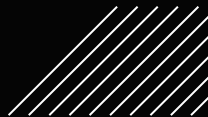

TELCO COMPANY

Pitri Handayani

Background



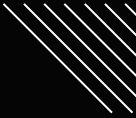
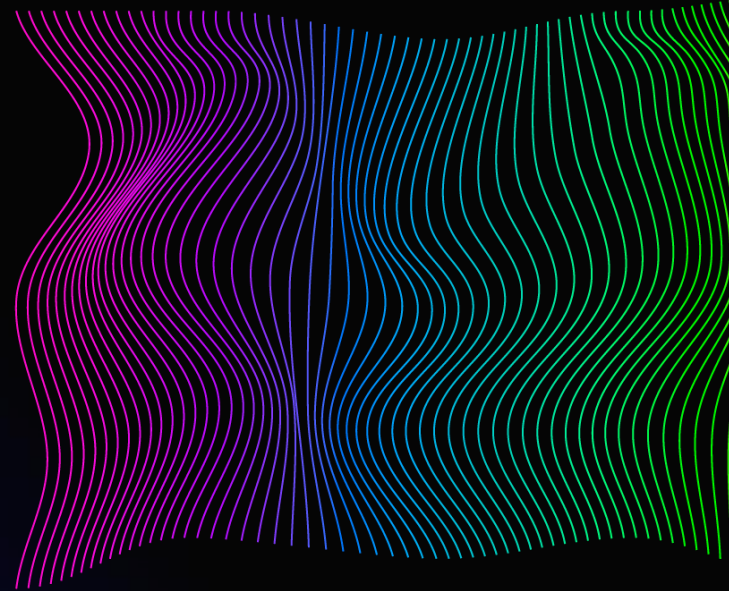
Customer churn rate adalah persentase pelanggan yang berhenti menggunakan layanan perusahaan. Perusahaan telekomunikasi memiliki kebutuhan untuk mengetahui apakah pelanggan akan berhenti berlangganan atau tidak, karena mendapatkan pelanggan baru dapat menghabiskan biaya hingga 25 kali lebih banyak daripada mempertahankan pelanggan yang sudah ada (Sumber : heavy.ai/use-case/customer-churn-analysis). Oleh karena itu diperlukan strategi yang efektif untuk mengurangi churn rate dan mempertahankan pelanggan.





What we did

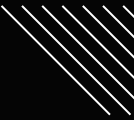
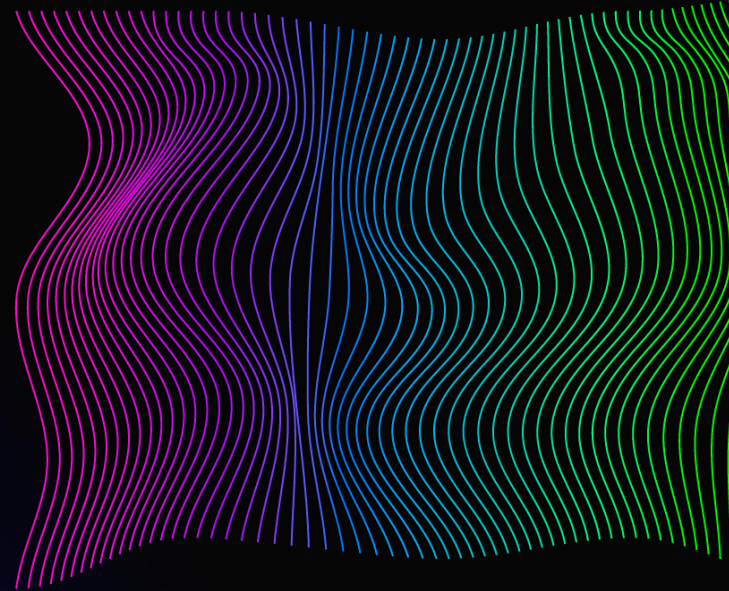
Kita akan membangun model Machine learning untuk dapat melakukan tindak preventif bagi pelanggan yang berpotensi besar untuk *churn*.





Data Overview

Data yang digunakan merupakan data dari sebuah perusahaan telekomunikasi yang diperoleh dari Kaggle. Dataset tersebut berisikan data untuk 7043 pelanggan yang meliputi demografis pelanggan, informasi pembayaran akun, serta produk layanan yang didaftarkan oleh tiap pelanggan

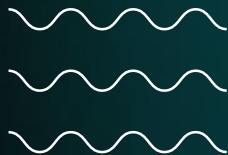


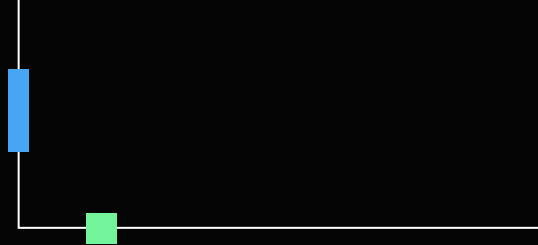
ALUR



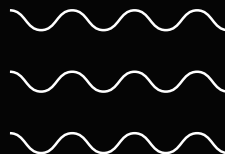
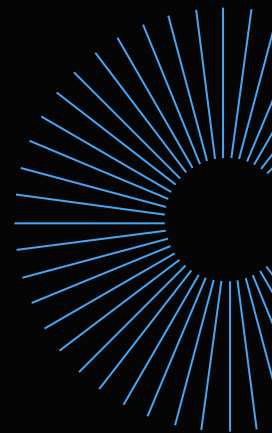


Data Understanding





Dataset yang ada
● meliputi 7043 pengamatan
dengan 20 fitur dan 1 label (Churn)

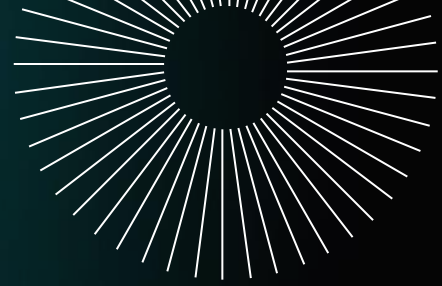


No	Nama Kolom	Keterangan	Type Data
1	CustomerID	ID unik setiap pelanggan	Categorical
2	Gender	Gender pelanggan pria atau wanita	Categorical
3	SeniorCitizen	Apakah pelanggan merupakan senior citizen di atas 65 tahun (0: No, 1: Yes)	numeric, int
4	Partner	Apakah pelanggan memiliki partner atau tidak (Yes, No)	categorical
5	Dependents	Apakah pelanggan memiliki tanggungan atau tidak (Yes, No)	categorical
6	Tenure	Jumlah bulan dalam menggunakan produk perusahaan	numeric, int
7	PhoneService	Apakah pelanggan memiliki layanan telepon atau tidak (Yes, No)	categorical
8	MultipleLines	Apakah pelanggan menggunakan layanan multiple lines atau tidak (Yes, No, No phone service)	categorical
9	InternetService	Internet service provider yang digunakan pelanggan (DSL, Fiber optic, No)	categorical
10	OnlineSecurity	Apakah pelanggan memiliki layanan online security atau tidak (Yes, No, No internet service)	categorical
..
..
..
21	Churn	Apakah pelanggan churn/beralih atau tidak (Yes or No)	categorical



Exploratory Data Analysis

Detail Data



Dimensi data Telco Customer Churn adalah 7043 baris dan 21 kolom

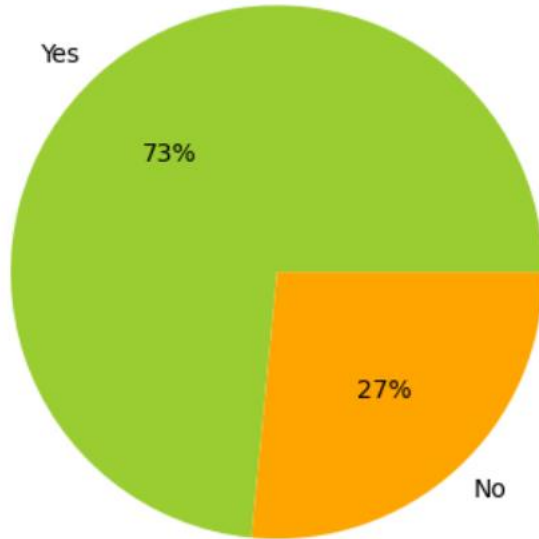


Target data yaitu field 'Churn', dimana jumlah yes-churn adalah 1869 atau 73% dan No-churn adalah 5174 atau 27%.

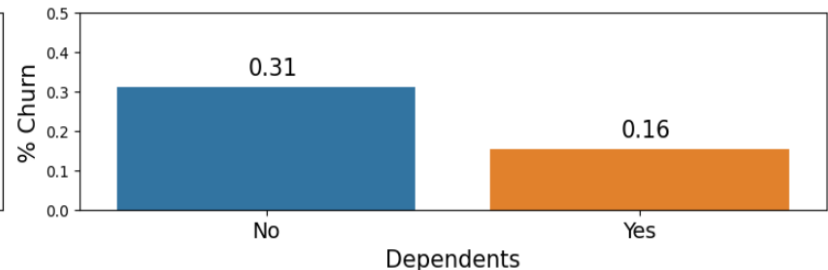
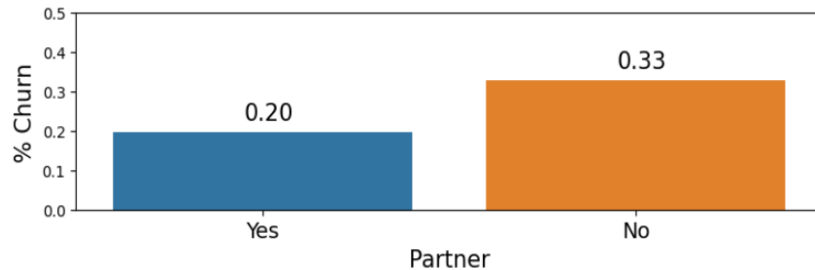
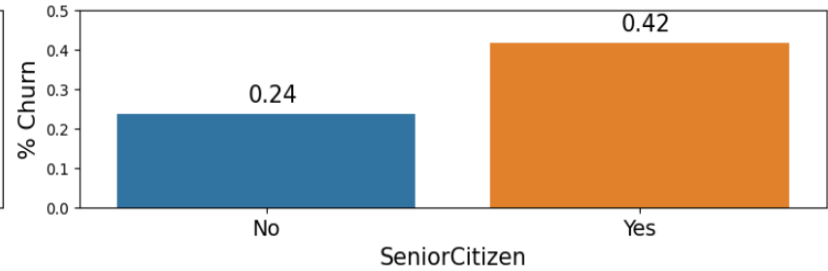
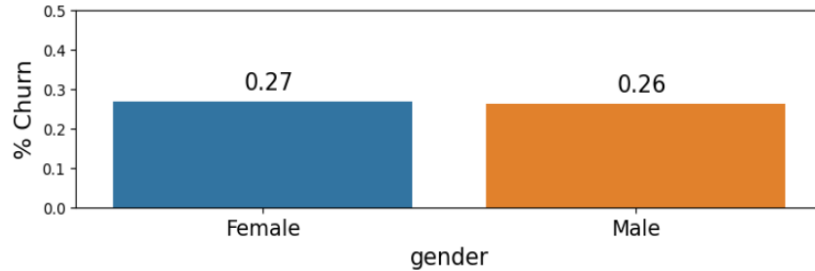


Data Preprocessing

- Melakukan pengecekan missing value
- Melakukan perubahan data pada Total Charge dari kategorikal ke numerical dan juga mengeluarkan data yang missing dari dataset
- Label encoding Churn supaya bisa menggunakan sns.barplot



Data Visualization Variabel Kategorik

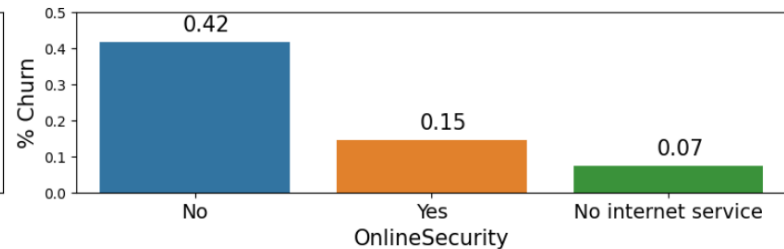
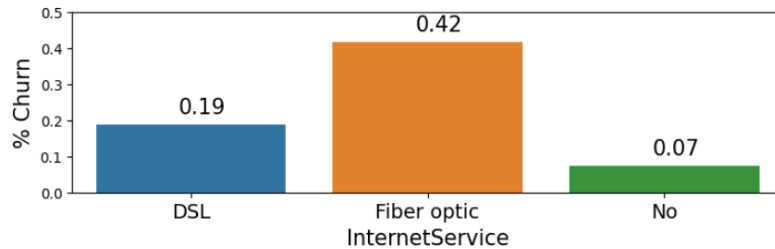
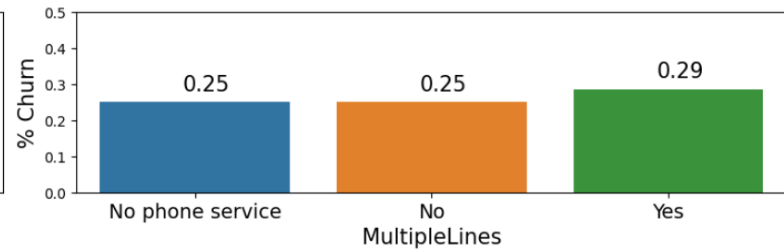
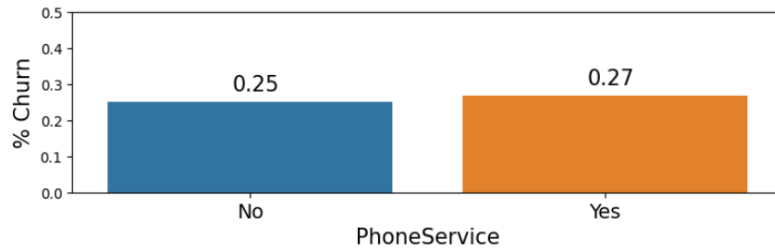


Gender : Baik untuk laki-laki maupun perempuan hampir tidak ada perbedaan persentase churn

SeniorCitizen: Memiliki persentase Churn 42% mengindikasikan tingginya churn

Partner: Pelanggan yang tidak memiliki partner cenderung memiliki persentase churn yang cukup tinggi dibandingkan pelanggan yang memiliki partner

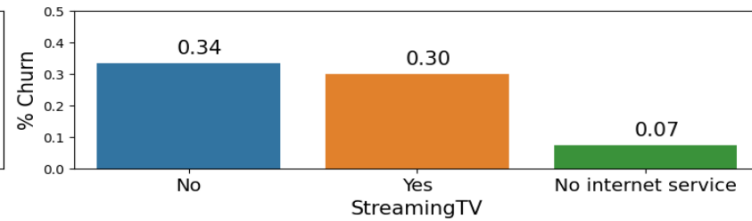
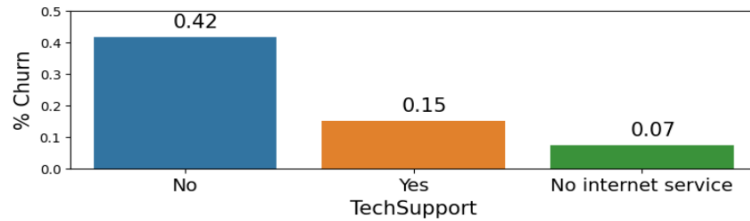
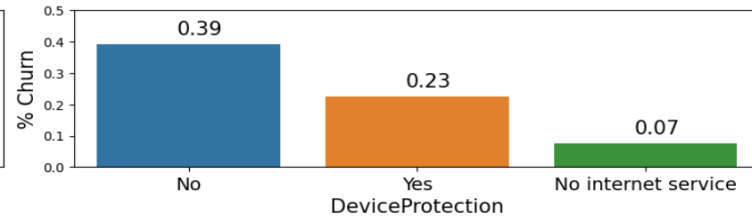
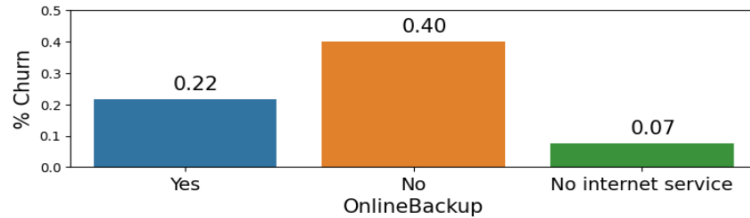
Dependents: Pelanggan yang tidak memiliki tanggungan cenderung memiliki persentase churn yang cukup tinggi dibandingkan pelanggan yang memiliki tanggungan



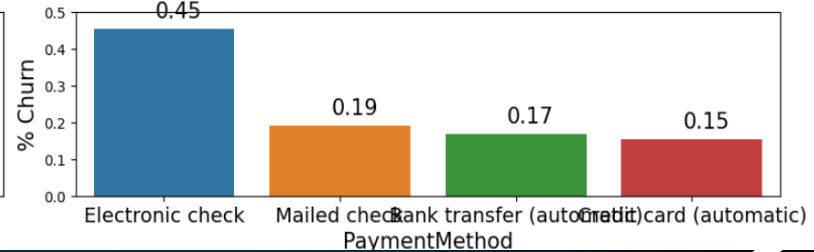
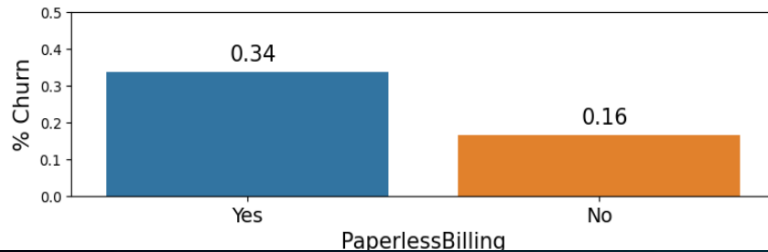
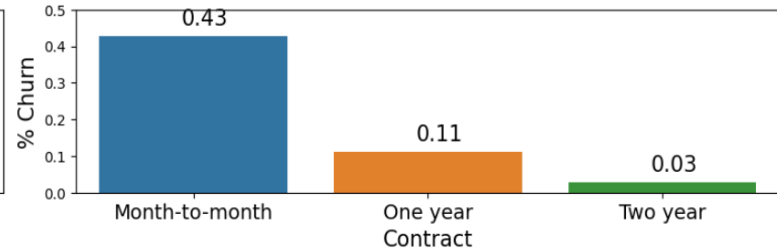
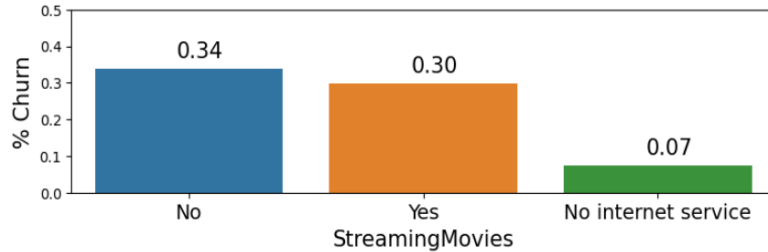
PhoneService: memiliki perbedaan yang cukup tipis dalam persentase churn antara pelanggan dengan atau tanpa Phone Service.

InternetService: pelanggan yang menggunakan DSL dalam layanan internet memiliki persentase yang kecil untuk churn

Online Security: Pelanggan yang menggunakan layanan tersebut cenderung tidak memiliki churn yang tinggi



- Online Backup, Device Protection, Tech Support:** Pelanggan yang menggunakan layanan-layanan tersebut cenderung tidak memiliki churn yang tinggi
- Streaming TV :** Baik pengguna maupun tidak, perbedaan pelanggan yang churn hanya sedikit.



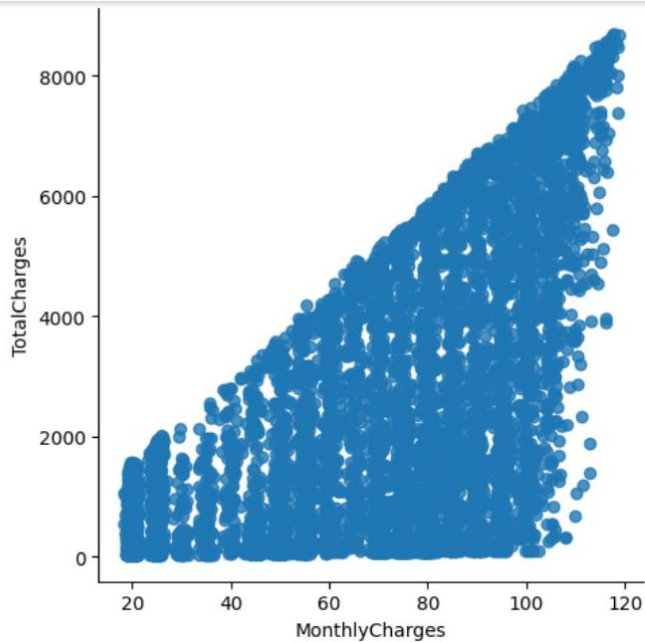
Streaming Movies: Baik pengguna maupun tidak, perbedaan pelanggan yang churn hanya sedikit.

Contract: Pelanggan yang membayar bulanan memiliki churn yang paling tinggi

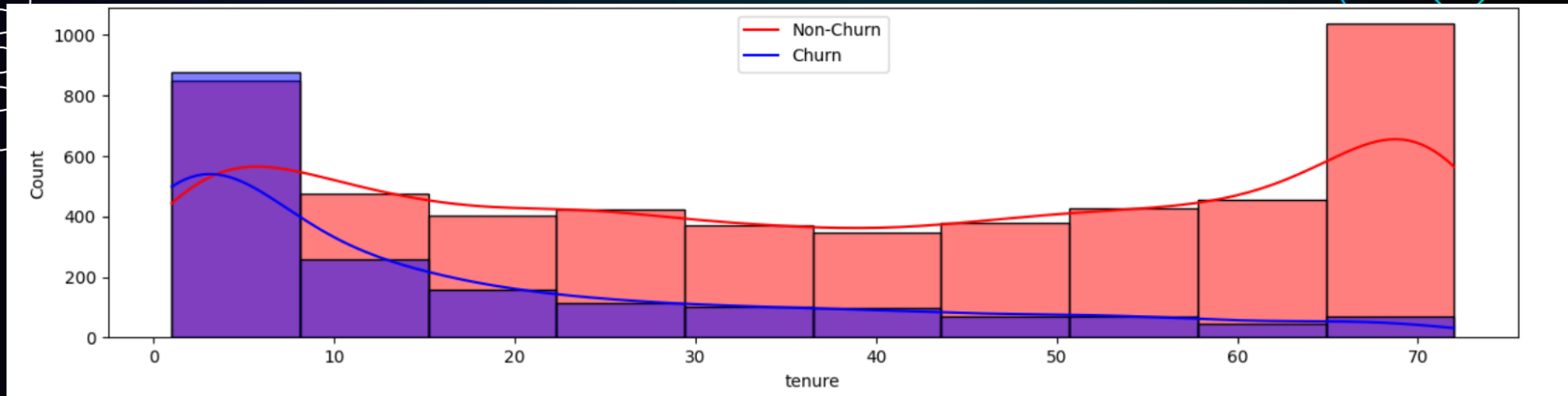
Paperless Billing: Pelanggan yang menggunakan paperless billing lebih banyak yang churn daripada yang tidak.

Payment Methode : pelanggan yang enggunakan electronic check (bagian dari paperles billing) sebagai metode pembayaran memiliki churn yang tinggi dibanding metode pembayaran lainnya.

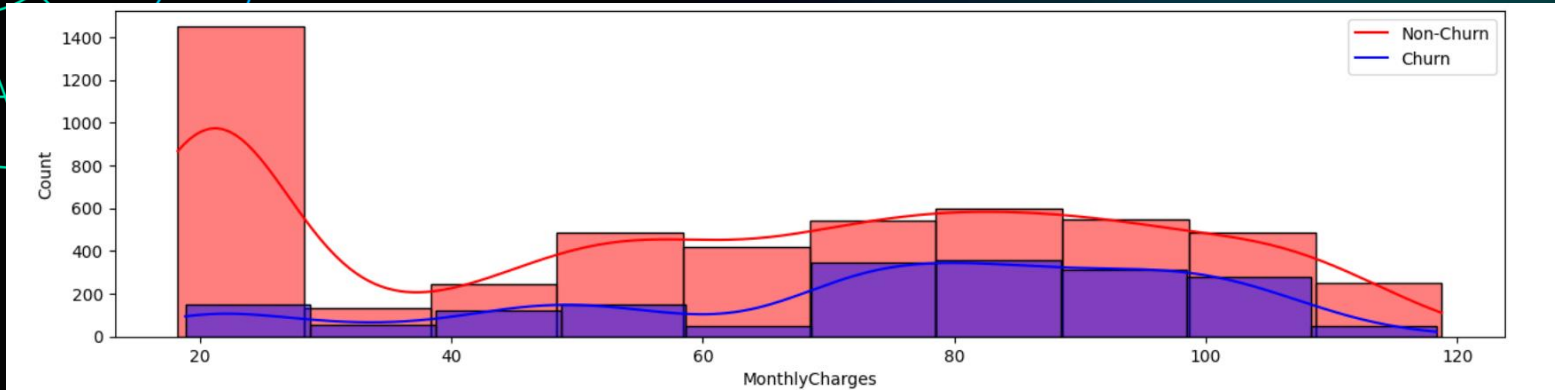
Data Visualization Variabel Numerik



TotalCharges meningkat seiring dengan peningkatan MonthlyCharges

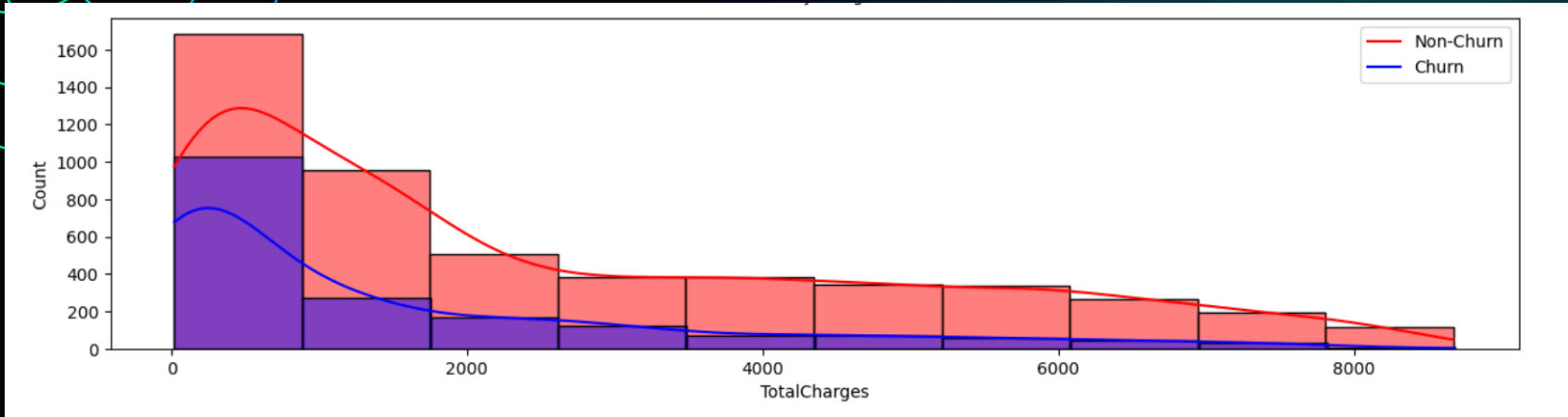


Di bulan awal pada fitur **tenure** memiliki churn paling banyak (semakin cepat berlangganan customer, semakin besar kecenderungan churn)



Untuk **MonthlyCharges** ada kecenderungan semakin kecil nilai biaya bulanan yang dikenakan, semakin kecil juga kecenderungan untuk

Churn



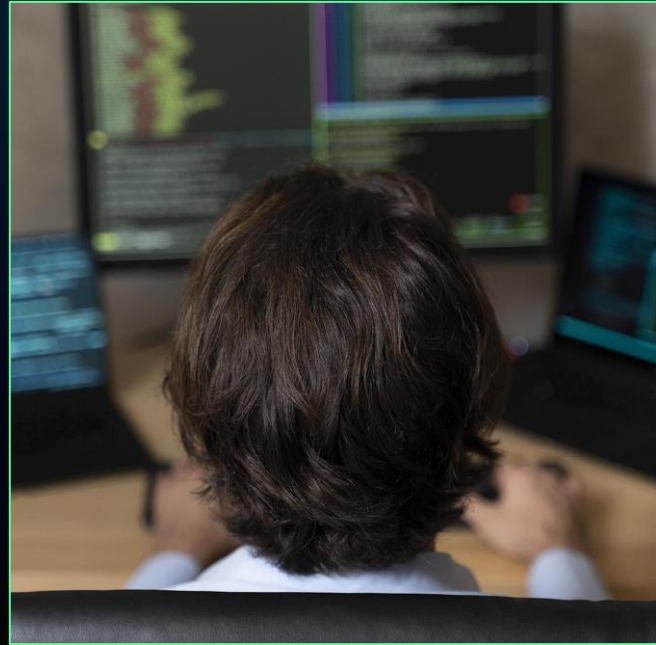
Churn lebih tinggi terdapat di **TotalCharges** yang rendah



Business Insight

- Perusahaan perlu memperhatikan senior citizen karena memiliki kecenderungan churn yang tinggi.
- Perusahaan bisa menggunakan layanan TechSupport & OnlineSecurity sebagai sarana promosi kepada pelanggan yang berpotensi untuk churn (karena pelanggan yang menggunakan kedua layanan itu memiliki churn yang rendah).
- Perusahaan lebih memperhatikan harga fiber optic dan kualitas internet service fiber optic.
- di bulan-bulan awal (12 bulan pertama) perusahaan sebisa mungkin memberikan pelayanan yang baik agar membuat pelanggan merasa nyaman dan tidak memutuskan untuk churn.

Model Building & Evaluasi





Splitting Data

Data Train sebesar 70%

Data Test sebesar 30%



Melakukan pemodelan

Decision Tree Classifier

Random Forest Classifier

Logistic Regression

K-Neighbors Classifier

Support Vector Machine



Decision Tree

	precision	recall	f1-score	support
0	0.94	0.90	0.92	799
1	0.92	0.95	0.94	967
accuracy			0.93	1766
macro avg	0.93	0.93	0.93	1766
weighted avg	0.93	0.93	0.93	1766

Random Forest

	precision	recall	f1-score	support
0	0.95	0.94	0.94	773
1	0.95	0.96	0.96	993
accuracy			0.95	1766
macro avg	0.95	0.95	0.95	1766
weighted avg	0.95	0.95	0.95	1766

Logistic Regression

	precision	recall	f1-score	support
0	0.72	0.99	0.83	772
1	0.99	0.70	0.82	994
accuracy			0.83	1766
macro avg	0.85	0.84	0.83	1766
weighted avg	0.87	0.83	0.83	1766

K-Neighbors

	precision	recall	f1-score	support
0	0.95	0.93	0.94	791
1	0.95	0.96	0.95	975
accuracy			0.95	1766
macro avg	0.95	0.95	0.95	1766
weighted avg	0.95	0.95	0.95	1766

Support Vector Machine

	precision	recall	f1-score	support
0	0.74	0.81	0.77	762
1	0.85	0.78	0.81	1004
accuracy			0.80	1766
macro avg	0.79	0.80	0.79	1766
weighted avg	0.80	0.80	0.80	1766

Kesimpulan

Dari analisa diatas dapat disimpulkan untuk memprediksi churn dari pelanggan telco dengan menggunakan dataset ini model terbaiknya menggunakan algoritma KNN. Hal ini dikarenakan performa dari model KNN pada fase training dan testing memiliki akurasi yang sama-sama baik yaitu 97 % (fase training) dan 95% (fase testing), sedangkan algoritma yang lain seperti decision tree cenderung over-Fitting performanya, algoritma random forest cenderung under-Fitting , dan algoritma lainnya memiliki akurasi yang lebih rendah dari KNN.

THANKS

Do you have any questions?

pitrihandayani90@gmail.com

Visit My Github:

https://github.com/Phandayani20/Customer-Churn/blob/main/Latihan_Customer_Churn_Telco.ipynb

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon** and infographics & images by **Freepik**

Please keep this slide for attribution

