

zenius

Kampus
Merdeka
INDONESIA JAYA

Final Project Presentation

Nomor Kelompok: 4

Nama Mentor: Ramadhan Hidayat

Nama:

- Abdul Rahman
- Hartika

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:

- Memprediksi faktor yang mempengaruhi churn

Explorasi Data dan Visualisasi

Business Understanding

Churn didefinisikan dalam istilah bisnis sebagai 'ketika klien membatalkan langganan layanan yang telah mereka gunakan'. Pelanggan yang berbeda menunjukkan perilaku dan preferensi yang berbeda, sehingga mereka membatalkan langganan karena berbagai alasan. Misalnya ada pesaing baru di pasar yang menawarkan harga lebih baik atau mungkin layanan yang mereka dapatkan belum sesuai standar, dan lain sebagainya.

Oleh karena itu, sangat penting untuk berkomunikasi secara proaktif dengan masing-masing dari mereka untuk mempertahankan dalam daftar pelanggan. Untuk itu, perlu tindakan pemasaran apa yang harus diambil untuk setiap pelanggan untuk memaksimalkan peluang bahwa pelanggan akan tetap tinggal.

Maka dari itu, Prediksi Churn penting untuk mendeteksi pelanggan yang kemungkinan besar akan meninggalkan layanan atau membatalkan langganan layanan. Karena mendapatkan klien baru seringkali lebih mahal daripada mempertahankan klien yang sudah ada.

Data Cleansing

Data tersebut perlu dibersihkan dan terdapat:

- 33 baris dan 11 kolom
- Dtype float64 ada 5 dan int64 ada 6
- Tidak terdapat Missing value
- Tidak terdapat duplicated data
- Terdapat outliers pada kolom DataUsage, DayMins, OverageFee, dan MonthlyCharge.
Solusi pada problem tersebut: Menggunakan visualisasi untuk mengecek Outliers dengan Z score atau IQR dan remove rows untuk menghandle outliers.

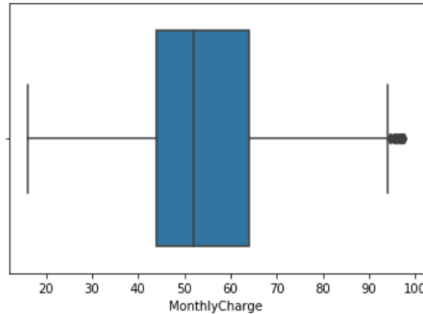
Exploratory Data Analysis

- Pada Data tersebut setiap baris mewakili pelanggan dan setiap kolom saling berkaitan dengan pelanggan seperti yang terdapat pada variable kolom.
- Pelanggan yang cenderung churn lebih sedikit dibanding pelanggan yang tidak churn.
- Korelasi Pearson DataUsage dan MonthlyCharges memiliki korelasi tinggi
- Pada cleaning data tersebut hanya terdapat outliers sehingga membutuhkan visualisasi

Exploratory Data Analysis

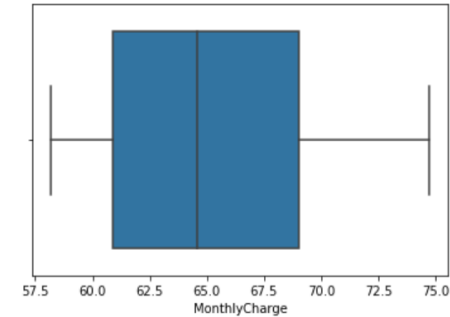
- Before Cleansing

Out[25]: <AxesSubplot:xlabel='MonthlyCharge'>



- After Cleansing

Out[28]: <AxesSubplot:xlabel='MonthlyCharge'>



Modelling



Customer Churn

Hal-hal yang disarankan untuk dipresentasikan:

- Metode train test split / cross validation:
 - a. Linear Regression
- Metrik untuk melakukan evaluasi:
 - a. Accuracy
 - b. Precision
 - c. Recall (True Positive Rate)
 - d. False Positive Rate
- Jenis model awal yang dicoba:
 - a. Logistic Regression

Customer Churn

- Jenis model lain yang turut dicoba:
 - a. Decision Tree
 - b. Random Forest
 - c. Hyperparameter Tuning

Untuk menambah akurasi model (hyperparameter tuning) : Sklearn model dengan GridSearchCV.

- Model final: Classification Report
- Kolom-kolom predictor: Churn
- Target variable untuk model final: MonthlyCharge

Conclusion

Conclusion

Pada proyek ini, Pelanggan yang tidak Churn lebih banyak dari pelanggan yang Churn. Namun ini penting untuk diperhatikan oleh perusahaan karena Churn juga sangat berpengaruh pada Perusahaan itu sendiri. Untuk itu, Perusahaan dapat memberikan gift berupa paket data atau lainnya. Karena salah satu faktor yang mempengaruhi Churn yaitu Paket data yang digunakan oleh pelanggan. Disamping itu, Perusahaan perlu memaksimalkan peluang yang ada agar pelanggan tetap tinggal dan meminimalkan kelemahan yang ada.

**Terima
kasih!**
Ada pertanyaan?

zenius



**Kampus
Merdeka**
INDONESIA JAYA

