

Telco Customer Churn Classification

Nizam Avif Anhari, Al Fitra Nur Ramadhani

{nizamanhari,alfitranurr}@webmail.umm.ac.id

Program Studi Informatika, Universitas Muhammadiyah Malang

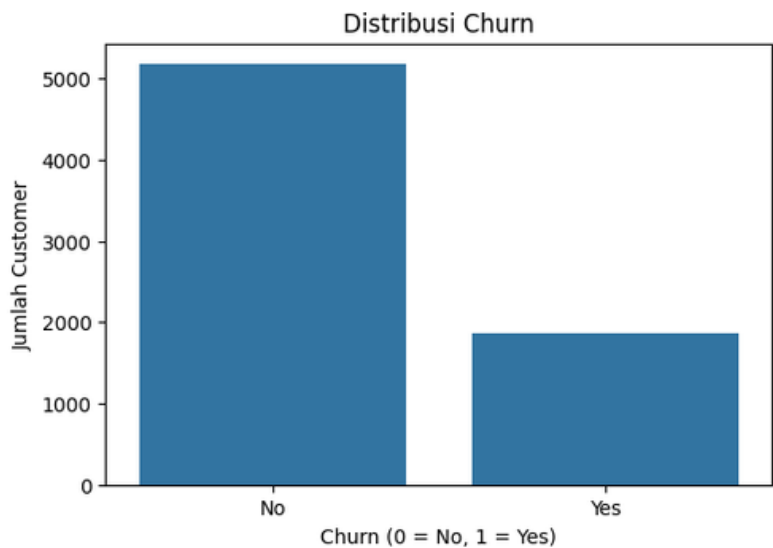
Overview

- Churn pelanggan di telekomunikasi menggerus pendapatan berulang dan menaikkan biaya akuisisi. Penelitian ini memanfaatkan decision tree pada dataset untuk mengungkap faktor risiko churn, seperti durasi kontrak, biaya bulanan, dan ketersediaan layanan lain.
- Projek ini penting karena dengan mengenali faktor utama churn, perusahaan bisa menawarkan kontrak jangka panjang fleksibel dan layanan tambahan untuk mempertahankan pelanggan.

Dataset

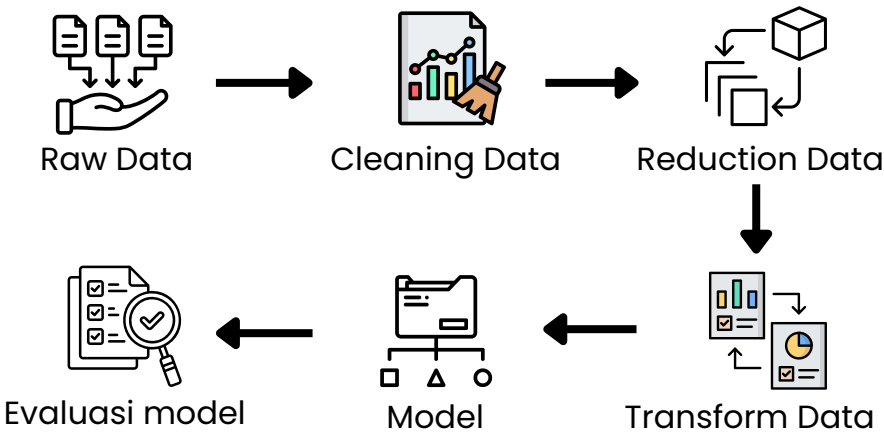
Telco Customer Churn

- Dataset memiliki 7043 instance dengan 21 colom diantaranya yaitu, CustomerID, Gender, Monthly Charges, Total Charges.
- Target class terbagi menjadi 2 yaitu, churn dan not churn.



- 80% Train / 20% Test

Methodology



- Preprocessing data dengan 4 tahapan
 - Data cleaning, menangani nilai NaN, dan lain lain
 - Data reduction, menghapus fitur "Customer ID"
 - Transformation, menggunakan label encoder, binning, minmax scaling.
 - Feature Selection dengan chi2 dan SFS forward
- Evaluasi Model dengan XAI (LIME)
- Algoritma klasifikasi yang digunakan, yaitu:
 - Decision Tree
 - Naive Bayes

Results & Discussion

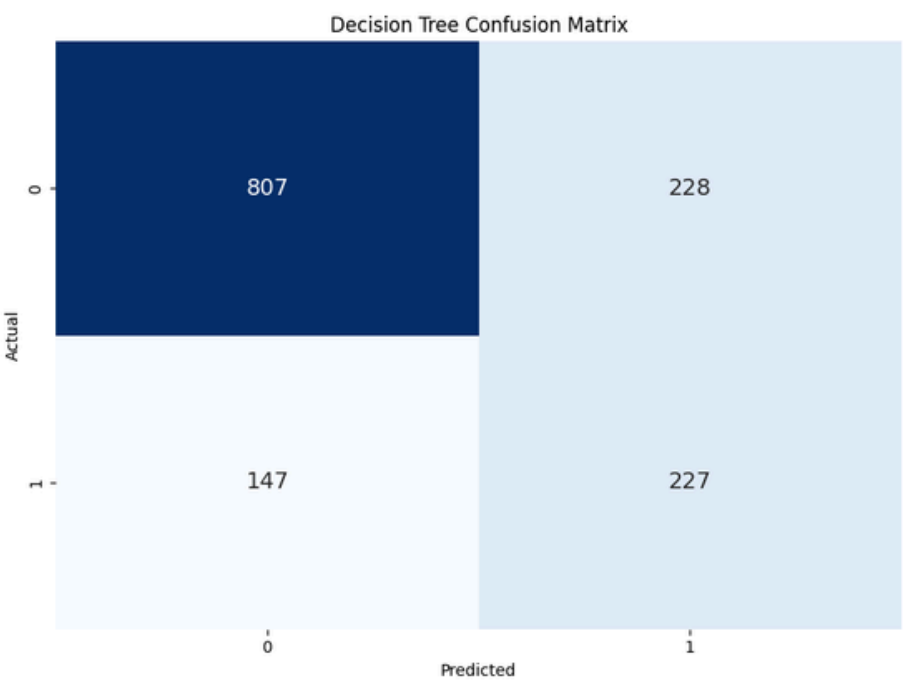
Results

Decision Tree Classification Report

Models	Accuracy
Decission Tree Non Seleksi Fitur	0.73
Naive Bayes Non Seleksi Fitur	0.72
Decission Tree + Seleksi Fitur Chi2	0.72
Naive Bayes + Seleksi Fitur Chi2	0.71
Decision Tree + Seleksi Fitur SFS	0.72
Naive Bayes + Seleksi Fitur SFS	0.72

- Model Terbaik : **Decission Tree Non Seleksi Fitur**
- Model Terburuk : **Naive Bayes + Seleksi Fitur Chi2**

Visualisasi Confusion Matrix Decission Tree Non Seleksi Fitur



Discussion

- Performa Model:** Model Decision Tree Non-Seleksi Fitur mencapai akurasi tertinggi (0.73), namun perbedaan akurasi antar model kecil, menunjukkan performa yang kompetitif.
- Analisis Kesalahan:** Confusion Matrix menunjukkan kesalahan prediksi lebih tinggi pada kelas minoritas (churn = yes), dengan 228 data salah diklasifikasikan.
- Ketidakseimbangan Data:** Dataset tidak seimbang (80% no churn, 20% churn), menyebabkan model cenderung bias ke kelas mayoritas.
- Keterbatasan Konteks:** Fitur yang digunakan terbatas pada data pelanggan, kurang memperhitungkan faktor eksternal seperti kompetitor atau kepuasan layanan.

Future

- Uji model ensemble untuk meningkatkan akurasi dan mengurangi bias pada kelas minoritas.
- Pengembangan model dengan fitur tambahan, seperti data kepuasan pelanggan atau tren pasar.
- Pengujian model secara berkala untuk memprediksi churn dalam konteks yang lebih dinamis.

Reference

Kaggle :
<https://www.kaggle.com/datasets/blastchar/telco-customer-churn>