

Artikel Jurnal Computing

PREDIKSI CHURN PELANGGAN MENGGUNAKAN METODE XGBOOST PADA DATA TELCO CUSTOMER CHURN

Penulis
Ardi Kamal Karima (301230023)

ABSTRAK: Churn pelanggan merupakan permasalahan penting dalam industri jasa, khususnya pada sektor telekomunikasi, karena berdampak langsung terhadap pendapatan perusahaan. Penelitian ini bertujuan untuk membangun model prediksi churn pelanggan menggunakan pendekatan machine learning berbasis supervised learning. Dataset yang digunakan adalah Telco Customer Churn yang diperoleh dari Kaggle dan memiliki karakteristik data tidak seimbang (imbalanced). Metode utama yang digunakan adalah Extreme Gradient Boosting (XGBoost), yang dikenal efektif dalam menangani data imbalanced dan hubungan nonlinier antar fitur. Sebagai pembanding, digunakan Logistic Regression sebagai model baseline. Proses penelitian meliputi tahap preprocessing data, eksplorasi data, pemodelan, evaluasi model menggunakan metrik F1-Score dan ROC-AUC, serta deployment model ke dalam aplikasi web berbasis Streamlit. Hasil penelitian menunjukkan bahwa model XGBoost memberikan performa yang lebih baik dibandingkan Logistic Regression dalam memprediksi churn pelanggan. Aplikasi web yang dikembangkan memungkinkan pengguna melakukan prediksi churn secara interaktif dan dapat dijadikan sebagai alat bantu pengambilan keputusan bisnis.

Kata kunci: Churn Pelanggan, Machine Learning, XGBoost, Data Imbalanced, Streamlit

ABSTRACT: Customer churn is a critical issue in the service industry, particularly in the telecommunications sector, as it directly affects company revenue. This study aims to develop a customer churn prediction model using a supervised machine learning approach. The dataset used is the Telco Customer Churn dataset obtained from Kaggle, which exhibits an imbalanced class distribution. The primary method employed is Extreme Gradient Boosting (XGBoost), known for its effectiveness in handling imbalanced data and nonlinear feature relationships. Logistic Regression is used as a baseline model for comparison. The research stages include data preprocessing, exploratory data analysis, model training, evaluation using F1-Score and ROC-AUC metrics, and deployment through a Streamlit-based web application. The results show that the XGBoost model outperforms Logistic Regression in predicting customer churn. The developed web application allows users to perform interactive churn predictions and serves as a decision-support tool for business analysis.

Keywords: Customer Churn, Machine Learning, XGBoost, Imbalanced Data, Streamlit

PENDAHULUAN

Persaingan yang ketat dalam industri telekomunikasi menyebabkan perusahaan harus mampu mempertahankan pelanggan yang sudah ada. Salah satu tantangan utama yang dihadapi adalah churn pelanggan, yaitu kondisi ketika pelanggan berhenti menggunakan layanan yang diberikan. Biaya untuk memperoleh pelanggan baru umumnya lebih tinggi dibandingkan mempertahankan pelanggan lama, sehingga prediksi churn menjadi aspek penting dalam strategi bisnis.

Machine learning menjadi salah satu pendekatan yang efektif untuk memprediksi churn pelanggan berdasarkan data historis. Namun, permasalahan churn sering kali memiliki karakteristik data tidak seimbang, di mana jumlah pelanggan yang tidak churn jauh lebih banyak dibandingkan pelanggan churn. Oleh

karena itu, diperlukan algoritma yang mampu menangani kondisi tersebut dengan baik.

Penelitian ini menggunakan metode XGBoost untuk memprediksi churn pelanggan karena kemampuannya dalam menangani data imbalanced dan kompleksitas fitur. Selain itu, hasil prediksi diimplementasikan dalam bentuk aplikasi web interaktif agar mudah digunakan oleh pengguna non-teknis.

METODE

Metode penelitian yang digunakan terdiri dari beberapa tahap. Tahap pertama adalah pengumpulan dataset Telco Customer Churn dari Kaggle yang berisi informasi pelanggan seperti durasi berlangganan, biaya bulanan, jenis kontrak, serta status churn. Tahap kedua adalah preprocessing data, yang meliputi penghapusan atribut yang tidak relevan, penanganan missing value, encoding data kategorikal, serta normalisasi fitur numerik.

Tahap selanjutnya adalah pemodelan menggunakan dua algoritma machine learning, yaitu Logistic Regression sebagai baseline dan XGBoost sebagai model utama. Evaluasi model dilakukan menggunakan metrik F1-Score dan ROC-AUC untuk memastikan performa model pada data imbalanced. Tahap terakhir adalah deployment model ke dalam aplikasi web berbasis Streamlit yang memungkinkan pengguna melakukan prediksi churn secara interaktif.

HASIL DAN PEMBAHASAN

Hasil pengujian menunjukkan bahwa model XGBoost memiliki performa yang lebih baik dibandingkan Logistic Regression dalam memprediksi churn pelanggan. Nilai F1-Score dan ROC-AUC pada model XGBoost lebih tinggi, yang menandakan kemampuan model dalam mengklasifikasikan pelanggan churn dan non-churn secara lebih seimbang.

Visualisasi feature importance menunjukkan bahwa faktor seperti durasi berlangganan (tenure), biaya bulanan (monthly charges), dan jenis kontrak memiliki pengaruh besar terhadap keputusan churn. Aplikasi web yang dikembangkan memungkinkan pengguna memasukkan data pelanggan dan memperoleh prediksi churn beserta probabilitasnya, sehingga memudahkan analisis risiko pelanggan secara praktis.

OUTPUT PROGRAM

Identitas Pembuat

Nama : Ardi Kamal Karima
NIM : 301230023
Kelas : 5C
Semester : 5
Universitas : Universitas Bale Bandung (UNIBBA)

Tugas Besar Mata Kuliah Machine Learning

Contoh Input

- Tenure: 12 bulan
- Monthly Charges: 75
- Total Charges: 900
- Contract: Month-to-month

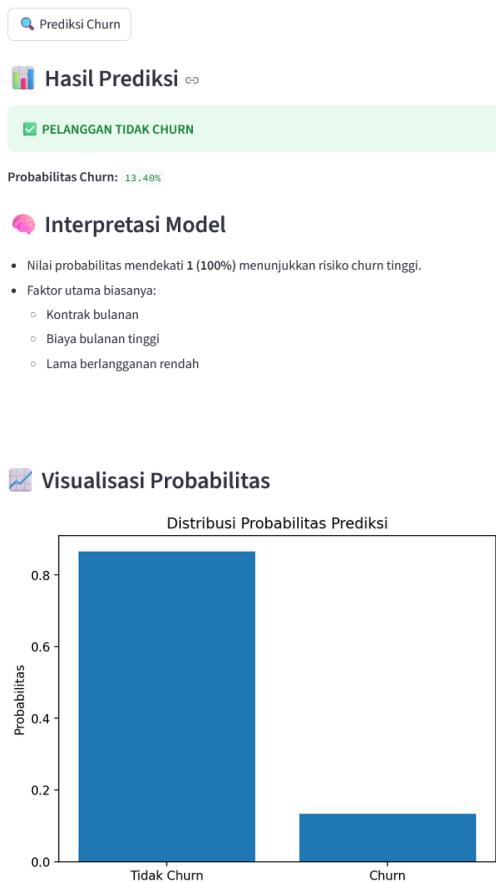
Biasanya berpotensi churn

Aplikasi Prediksi Churn Pelanggan

Aplikasi ini digunakan untuk memprediksi kemungkinan pelanggan berhenti berlangganan (churn) menggunakan Machine Learning (XGBoost) dengan data pelanggan perusahaan telekomunikasi.

Masukkan Data Pelanggan

Jenis Kelamin	Layanan Telepon	Lama Berlangganan (bulan)
Male	Yes	12
Pelanggan Lansia	Multi Line	Biaya Bulanan
0	Yes	0.00
Memiliki Pasangan	Layanan Internet	Total Biaya
Yes	DSL	0.00
Memiliki Tanggungan	Jenis Kontrak	Metode Pembayaran
Yes	Month-to-month	Electronic check



Untuk penelitian selanjutnya, disarankan untuk mencoba metode lain seperti ensemble learning atau deep learning serta menerapkan teknik penanganan imbalanced data yang lebih lanjut, seperti SMOTE, guna meningkatkan performa model.

LINK PROGRAM PROJECT

GITHUB:

https://github.com/Ardikamal/churn_project.git

DATASET KAGGLE:

<www.kaggle.com/datasets/palashfendarkar/wafnusec-telcocustomerchurn/data>

LINK PROGRAM DEPLOY STREAMLIT:

PUSTAKA ACUAN

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD*.

Kaggle. (2018). Telco Customer Churn Dataset.

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer.

SIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa metode XGBoost efektif digunakan untuk memprediksi churn pelanggan pada data yang bersifat imbalanced. Model yang dibangun mampu memberikan hasil prediksi yang lebih akurat dibandingkan Logistic Regression. Aplikasi web berbasis Streamlit yang dikembangkan mempermudah pengguna dalam melakukan analisis churn secara interaktif.