

ABSTRACT

Customer churn poses a significant challenge for businesses, particularly in subscription-based industries such as telecommunications, where retaining customers is critical to maintaining profitability. Predicting and mitigating churn effectively requires advanced predictive modeling approaches. This research proposal focuses on investigating the impact of various feature selection techniques on customer churn prediction using Deep Neural Networks (DNNs). The study aims to compare the performance of four feature selection methods—Filter (ANOVA F-test), Wrapper (Random Forest), Dimensionality Reduction (PCA), and Embedded (XGBoost)—to understand their influence on model accuracy, computational efficiency, and interpretability.

The research will utilize the Telco Customer Churn Dataset, which contains comprehensive customer data encompassing demographic, service-related, and account-specific attributes. The proposed study will systematically apply each feature selection method individually to train DNN models and evaluate their performance using metrics such as accuracy, precision, recall, F1-score, and Matthews Correlation Coefficient (MCC). A baseline model using all features will also be established for comparative analysis.

The expected outcomes of this research include identifying the impact of feature selection method on DNN customer churn prediction model. The findings aim to provide actionable insights into the role of feature selection in optimizing churn prediction models, contributing to the development of scalable and accurate tools for customer retention strategies. This study seeks to advance the understanding of feature selection in high-dimensional datasets, offering valuable contributions to machine learning applications in customer analytics.

Keywords: Customer Churn Prediction, Deep Neural Networks, Feature Selection, ANOVA F-test, Random Forest, PCA, XGBoost

ABSTRAK

Customer churn merupakan tantangan besar bagi perusahaan, terutama di industri berbasis langganan seperti telekomunikasi, di mana mempertahankan pelanggan sangat penting untuk menjaga profitabilitas. Memprediksi dan mengurangi churn secara efektif memerlukan pendekatan pemodelan prediktif yang canggih. Proposal penelitian ini berfokus pada penyelidikan dampak berbagai teknik seleksi fitur terhadap prediksi customer churn dengan menggunakan Deep Neural Networks (DNN). Studi ini bertujuan untuk membandingkan kinerja empat metode seleksi fitur—Filter (ANOVA F-test), Wrapper (Random Forest), Dimensionality Reduction (PCA), dan Embedded (XGBoost)—untuk memahami pengaruhnya terhadap akurasi model, efisiensi komputasi, dan keterbacaan (*interpretability*).

Penelitian ini akan menggunakan Telco Customer Churn Dataset, yang mencakup data pelanggan secara komprehensif, termasuk atribut demografis, layanan, dan informasi akun. Studi ini akan menerapkan setiap metode seleksi fitur secara sistematis untuk melatih model DNN dan mengevaluasi kinerjanya menggunakan metrik seperti akurasi, presisi, recall, F1-score, dan Matthews Correlation Coefficient (MCC). Sebagai pembandingan, model dasar (*baseline*) yang menggunakan semua fitur juga akan dibangun.

Hasil yang diharapkan dari penelitian ini mencakup identifikasi dampak metode seleksi fitur terhadap model prediksi *churn* berbasis DNN. Temuan ini bertujuan untuk memberikan wawasan yang dapat diimplementasikan terkait peran seleksi fitur dalam mengoptimalkan model prediksi churn, serta berkontribusi terhadap pengembangan alat yang *scalable* dan akurat untuk strategi retensi pelanggan. Studi ini juga bertujuan untuk memperluas pemahaman mengenai seleksi fitur dalam *dataset* berdimensi tinggi, memberikan kontribusi yang bernilai dalam penerapan pembelajaran mesin di bidang analitik pelanggan.