

ABSTRAK

PERBANDINGAN PERFORMA MODEL PREDIKSI *CUSTOMER CHURN* BERBASIS *MACHINE LEARNING* PADA *FASHION E-COMMERCE*

Oleh

DWI LILYAWATI

Dalam era perkembangan bisnis *e-commerce*, terutama di industri *fashion*, persaingan yang semakin ketat antara beragam platform *e-commerce* menjadi salah satu tantangan utama yang tidak dapat diabaikan. Adanya berbagai macam produk dan layanan pada *e-commerce* memungkinkan pelanggan secara bebas dapat meninggalkan perusahaan untuk mencari produk atau layanan yang sesuai dengan kebutuhan. Fenomena ini dikenal dengan istilah *churn*. Berdasarkan permasalahan tersebut, penelitian ini berfokus untuk mengembangkan model yang dapat memprediksi pelanggan yang akan *churn* pada *fashion e-commerce* sehingga perusahaan dapat mengambil tindakan proaktif untuk menjaga pelanggan yang ada agar tidak *churn*. Metode yang digunakan dalam penelitian ini adalah *Cross Industry Standard Process for Data Mining* (CRISP-DM). Model prediksi ini dibangun dengan pendekatan klasifikasi menggunakan algoritma *machine learning* yaitu *Logistic Regression*, *Random Forest* dan *XGBoost*. Kategori *churn* dibagi menjadi dua yaitu *churn* dan *not churn*. Hasil penelitian menunjukkan bahwa *XGBoost* memiliki performa yang paling tinggi di antara ketiga model, dengan akurasi sebesar 97%, *precision* sebesar 97%, *recall* sebesar 98%, *f1-score* sebesar 98% dan nilai AUC mencapai 0.995. Hasil tersebut menjadikan *XGBoost* sebagai model terbaik dalam memprediksi pelanggan yang *churn*. Selain itu, penggunaan seleksi fitur dalam membangun ulang model melalui *feature importance* pada *XGBoost* berhasil mempercepat waktu komputasi dari 2.471 detik menjadi 0.584 detik dan menghasilkan performa kinerja model yang sama seperti menggunakan seluruh fitur. Hal tersebut membuktikan bahwa penggunaan seleksi fitur melalui *feature importance* dapat mengoptimalkan kinerja model sehingga lebih efisien dari segi waktu komputasi dan tetap efektif. Penelitian ini juga menghasilkan visualisasi dan rekomendasi untuk mengurangi *customer churn*.

Kata kunci : *customer churn*, prediksi, *logistic regression*, *random forest*, *xgboost*

ABSTRACT

PERFORMANCE COMPARISON OF CUSTOMER CHURN PREDICTION MODELS BASED ON MACHINE LEARNING IN FASHION E-COMMERCE

By

DWI LILYAWATI

In the era of e-commerce business development, especially in the fashion industry, the intensifying competition among various e-commerce platforms has emerged as a significant challenge that cannot be overlooked. The wide array of products and services in e-commerce enables customers to freely depart from a company in pursuit of products or services that align with their needs. This phenomenon is commonly referred to as churn. In light of these issues, this research is focused on developing a model for predicting customer churn in fashion e-commerce, enabling companies to take proactive measures to retain existing customers and prevent churn. The research employs the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology. The predictive model is constructed using a classification approach, employing machine learning algorithms such as Logistic Regression, Random Forest, and XGBoost. Churn is categorized into two groups: churn and not churn. The research findings indicate that XGBoost outperforms the other models, achieving an accuracy rate of 97%, precision of 97%, recall of 98%, an F1-score of 98%, and an AUC value of 0.995. These results establish XGBoost as the best model for predicting customer churn. Furthermore, the utilization of feature selection in model reconstruction through feature importance in XGBoost significantly reduces computation time from 2.471 seconds to 0.584 seconds, while maintaining identical model performance. This demonstrates that the use of feature selection through feature importance optimizes model efficiency in terms of computation time, all while remaining effective. The research also yields visualizations and recommendations aimed at reducing customer churn.

Keywords : Customer Churn, Prediction, Logistic Regression, Random Forest, XGBoost