

INTISARI

Penelitian ini bertujuan untuk melakukan perbandingan kinerja model klasifikasi dalam mengidentifikasi perilaku repeat buyer pada platform e-commerce menggunakan algoritma XGBoost, LightGBM, dan Long Short-Term Memory (LSTM). Dataset yang digunakan merupakan data perilaku pengguna yang telah melalui tahapan preprocessing dan feature engineering guna memastikan kualitas data dan menghindari data leakage. Evaluasi model dilakukan menggunakan metrik Accuracy, Precision, Recall, F1-score, dan Area Under the Curve (AUC). Hasil penelitian menunjukkan bahwa XGBoost memperoleh performa terbaik dengan akurasi sebesar 0,79 dan AUC sebesar 0,83, diikuti oleh LightGBM dengan akurasi 0,78 dan AUC 0,80. Sementara itu, LSTM menghasilkan akurasi 0,65 dan AUC 0,65, namun memiliki recall yang relatif tinggi dalam mendeteksi kelas positif. Temuan ini menunjukkan bahwa model berbasis ensemble tree cenderung lebih unggul pada karakteristik dataset yang digunakan dibandingkan model berbasis sekuensial. Penelitian ini memberikan gambaran empiris mengenai efektivitas masing-masing algoritma dalam konteks klasifikasi perilaku konsumen e-commerce.

Kata kunci: e-commerce, LightGBM, LSTM, repeat buyer, XGBoost

ABSTRACT

This study aims to compare the performance of classification models in identifying repeat buyer behavior in e-commerce platforms using XGBoost, LightGBM, and Long Short-Term Memory (LSTM) algorithms. The dataset consists of user behavioral data that underwent preprocessing and feature engineering to ensure data quality and prevent leakage. Model evaluation was conducted using Accuracy, Precision, Recall, F1-score, and Area Under the Curve (AUC) metrics. The results indicate that XGBoost achieved the best performance with an accuracy of 0.79 and an AUC of 0.83, followed by LightGBM with an accuracy of 0.78 and an AUC of 0.80. Meanwhile, LSTM obtained an accuracy of 0.65 and an AUC of 0.65, although it demonstrated relatively high recall in detecting the positive class. These findings suggest that tree-based ensemble models perform better than sequential-based models for the characteristics of the dataset used in this study. This research provides empirical insights into the comparative effectiveness of different classification algorithms in predicting consumer behavior in e-commerce.

Keywords: e-commerce, LightGBM, LSTM, repeat buyer, XGBoost