

ABSTRAK

Menentukan jumlah stok yang tepat diberbagai wilayah menjadi salah satu tantangan yang dihadapi oleh Industri penerbitan buku, dimana ketergantungan pada perkiraan manual sering kali menyebabkan ketidakseimbangan pasokan dan permintaan. Penelitian ini bertujuan untuk membangun sistem prediksi penjualan buku dengan menggunakan algoritma *Extreme Gradient Boosting* (XGBoost) berbasis website. Data yang digunakan adalah data historis penjualan periode 2020 -2023 sebagai data latih dan tahun 2024 sebagai data validasi. Tahapan *preprocessing* meliputi imputasi bulan dan tahun yang tidak dicatat, imputasi eksemplar yang kosong dengan *Zero Imputation* serta *feature engineering* yaitu pembuatan fitur *Lag* dan *Rolling Mean* untuk menangkap pola musiman penjualan.

Hasil evaluasi menunjukkan bahwa model XGBoost mampu memprediksi penjualan dengan nilai *Mean Absolute Error* (MAE) sebesar 59,7 dan koefisien determinasi (R^2) sebesar 49%. Hasil pengujian menunjukkan bahwa tingkat akurasi prediksi belum mencapai nilai mendekati 100% pada R^2 . Kondisi ini dipengaruhi oleh pola penjualan yang fluktuatif, keberadaan nilai ekstrem (outlier) yang tidak dihilangkan karena dipertahankan untuk menangkap pola musiman, serta adanya periode tertentu tanpa aktivitas penjualan. Meskipun model cenderung kurang akurat untuk memprediksi penjualan ekstrem, tetapi fitur *Lag* dan *Rolling Mean* terbukti efektif dalam menangkap pola musiman dan tren perubahan pada algoritma *machine learning*. Selain pengujian teknis, sistem ini juga telah melalui tahap pengujian pengguna (*Beta Testing*) kepada 14 responden dari pihak Kanisius, dengan hasil nilai rata-rata sebesar 79,2%, yang mengindikasikan bahwa sistem berada pada kategori "Layak" untuk diimplementasikan sebagai alat bantu pengambilan Keputusan menentukan stok buku ditahun berikutnya.

Kata kunci : Prediksi Penjualan Buku, XGBoost, *Machine Learning*, *Feature Engineering*, *Beta Testing*

ABSTRACT

Determining the right amount of stock in various regions is one of the challenges faced by the book publishing industry, where reliance on manual estimates often leads to imbalances in supply and demand. This study aims to develop a book sales prediction system using a website-based Extreme Gradient Boosting (XGBoost) algorithm. The data used is historical sales data from 2020 to 2023 as training data and 2024 as validation data. The preprocessing stages include imputation of unrecorded months and years, imputation of empty copies with Zero Imputation, and feature engineering, namely the creation of Lag and Rolling Mean features to capture seasonal sales patterns.

The evaluation results show that the XGBoost model is capable of predicting sales with a Mean Absolute Error (MAE) value of 59.7 and a coefficient of determination (R^2) of 49%. The test results show that the prediction accuracy level has not reached a value close to 100% on R^2 . This condition is influenced by fluctuating sales patterns, the existence of extreme values (outliers) that are not removed because they are retained to capture seasonal patterns, and the existence of certain periods without sales activity. Although the model tends to be less accurate in predicting extreme sales, the Lag and Rolling Mean features have proven effective in capturing seasonal patterns and trends in machine learning algorithms. In addition to technical testing, this system has also undergone user testing (Beta Testing) with 14 respondents from Kanisius, with an average score of 79.2%, indicating that the system is "Suitable" for implementation as a decision-making tool for determining book stock for the following year.

Keyword: Predict book sales, XGBoost, Machine Learning, Feature Engineering, Beta Testing

