

ABSTRACT

Airfare prices are influenced by various complex and dynamic factors, requiring a prediction method capable of accurately capturing nonlinear patterns. This study aims to build an airfare price prediction model using the Extreme Gradient Boosting (XGBoost) algorithm and improve its performance through Hyperparameter Tuning using Bayesian Optimization. The dataset used is historical airline ticket price data that has undergone preprocessing, including data cleaning, data transformation using One-Hot Encoding, normalization of numerical features, and division of data into training and test data with a ratio of 85:15. The XGBoost model was built as a baseline model, then hyperparameter Optimization was performed using the BayesSearchCV method to obtain the best parameter configuration, with performance evaluation using the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 Score metrics. The results show that the application of Bayesian Optimization can significantly improve model performance compared to the default configuration, where the XGBoost model after Tuning produces an MSE value of 3738706.7500000, an RMSE of 1933.5735698, an MAE of 912.3685303, and an R^2 value of 0.9926839, indicating a very high level of prediction accuracy. In addition, feature importance analysis shows that flight class and number of transits are dominant factors in determining airfare prices. Therefore, it can be concluded that the combination of XGBoost and Bayesian Optimization is an effective and reliable approach for predicting airfare prices and has the potential to be further developed into an applicable prediction system.

Keywords: XGBoost, Bayesian Optimization, Machine Learning, Regression

ABSTRAK

Harga tiket pesawat dipengaruhi oleh berbagai faktor yang kompleks dan bersifat dinamis, sehingga diperlukan metode prediksi yang mampu menangkap pola *nonlinier* secara akurat. Penelitian ini bertujuan untuk membangun model prediksi harga tiket pesawat menggunakan algoritma *Extreme Gradient Boosting (XGBoost)* serta meningkatkan kinerjanya melalui *Hyperparameter Tuning* menggunakan *Bayesian Optimization*. Dataset yang digunakan merupakan data historis harga tiket pesawat yang telah melalui tahapan *preprocessing*, meliputi pembersihan data, transformasi data menggunakan *One-Hot Encoding*, normalisasi fitur numerik, serta pembagian data menjadi data latih dan data uji dengan proporsi 85:15. Model *XGBoost* dibangun sebagai model *baseline*, kemudian dilakukan optimasi *hyperparameter* menggunakan metode *BayesSearchCV* untuk memperoleh konfigurasi parameter terbaik, dengan evaluasi performa menggunakan metrik *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, *Mean Absolute Error (MAE)*, dan *R² Score*. Hasil penelitian menunjukkan bahwa penerapan *Bayesian Optimization* mampu meningkatkan kinerja model secara signifikan dibandingkan dengan konfigurasi default, di mana model *XGBoost* setelah *Tuning* menghasilkan nilai MSE sebesar 3738706.7500000 RMSE sebesar 1933.5735698 MAE sebesar 912.3685303, dan nilai *R²* sebesar 0.9926839, yang mengindikasikan tingkat akurasi prediksi yang sangat tinggi. Selain itu, analisis *feature importance* menunjukkan bahwa kelas penerbangan dan jumlah transit merupakan faktor dominan dalam menentukan harga tiket pesawat, sehingga dapat disimpulkan bahwa kombinasi *XGBoost* dan *Bayesian Optimization* merupakan pendekatan yang efektif dan andal untuk prediksi harga tiket pesawat serta berpotensi dikembangkan lebih lanjut ke dalam sistem prediksi yang bersifat aplikatif.

Kata Kunci: *XGBoost, Bayesian Optimization, Machine Learning, Regression.*