

Abstrak

Indonesia sebagai negara dengan jumlah penduduk Muslim terbesar di dunia menghadapi tantangan dalam pengelolaan pendaftaran haji yang terus meningkat setiap tahunnya. Penelitian ini bertujuan untuk memprediksi jumlah pendaftaran haji dengan menggunakan metode algoritma *Random Forest* dan *XGBoost* yang terkenal dengan keakuratannya dalam analisis data besar dan non-linear. Dengan menggunakan data historis pendaftaran haji, model ini dirancang untuk mengidentifikasi pola dan faktor-faktor utama yang memengaruhi jumlah pendaftaran. Hasil evaluasi model menunjukkan bahwa *Random Forest* memberikan performa yang jauh lebih unggul dibandingkan *XGBoost*, dengan nilai *R-squared* sebesar 0,93 yang menunjukkan kemampuan model dalam menjelaskan variasi data, serta nilai *Mean Absolute Error* (MAE) dan *Mean Squared Error* (MSE) yang lebih rendah. Sebaliknya, *XGBoost* menunjukkan performa yang kurang optimal dengan nilai *R-squared* hanya 0,40. Temuan ini memberikan informasi penting bagi pemerintah untuk mengoptimalkan perencanaan dan pengelolaan kuota haji sehingga pelayanan kepada calon jamaah dapat menjadi lebih efisien dan tepat sasaran.

Kata kunci: Prediksi, Haji, *Random Forest*, *XGBoost*, *Data Mining*, Kabupaten Karawang
Algoritma: Prediksi Menggunakan *XGBoost*

Abstract

Indonesia as a country with the largest Muslim population in the world faces challenges in managing Hajj registration which continues to increase every year. This study aims to predict the number of Hajj registrations by using Random Forest and XGBoost algorithm methods which are well-known for their accuracy in large and non-linear data analysis. Using historical Hajj enrollment data, the model is designed to identify patterns and key factors that affect the number of enrollments. The model evaluation results showed that Random Forest performed significantly better than XGBoost, with an R-squared value of 0.93 indicating the model's ability to explain data variation, as well as lower Mean Absolute Error (MAE) and Mean Squared Error (MSE) values. In contrast, XGBoost showed less than optimal performance with an R-squared value of only 0.40. These findings provide important information for the government to optimize the planning and management of Hajj quotas so that services to prospective pilgrims can be more efficient and targeted.

Keywords: Prediction, Hajj, Random Forest, XGBoost, Data Mining, Karawang Regency
Algorithm: Prediction Using XGBoost