ABSTRAK

REVA NURNADIA SYAPA. Perbandingan Algoritma Random Forest dan XGBoost dalam Prediksi Lama Penyinaran Matahari untuk Mendukung Analisis Energi Surya di Kota Kupang. Dibimbing oleh Dr. Meilia Nur Indah Susanti, S.T., M.Kom.

Kota Kupang, Nusa Tenggara Timur, memiliki potensi energi surya yang tinggi sehingga informasi mengenai lama penyinaran matahari (*sunshine duration*) diperlukan sebagai salah satu parameter pendukung dalam analisis potensi energi surya dan perencanaan Pembangkit Listrik Tenaga Surya (PLTS). Namun, penelitian yang membandingkan secara langsung performa algoritma Random Forest dan XGBoost dalam memprediksi lama penyinaran matahari sebagai variabel target masih terbatas. Penelitian ini bertujuan menganalisis dan membandingkan performa algoritma Random Forest Regressor dan XGBoost Regressor dalam memprediksi lama penyinaran matahari di Kota Kupang. Penelitian ini menerapkan metodologi CRISP-DM menggunakan data harian Stasiun Meteorologi El Tari Kupang periode 2016–2020 sebanyak 1.793 observasi dengan enam parameter meteorologi sebagai variabel prediktor. Data dibagi menggunakan *time-based split* dengan rasio terbaik 90:10, sedangkan optimasi hyperparameter dilakukan menggunakan *Grid Search Cross-Validation*. Hasil penelitian menunjukkan bahwa XGBoost Regressor menghasilkan nilai MAE sebesar 1,3210 jam, RMSE sebesar 1,6911 jam, dan R^2 sebesar 0,5225, sedangkan Random Forest Regressor memperoleh nilai MAE sebesar 1,3259 jam, RMSE sebesar 1,7118 jam, dan R^2 sebesar 0,5107. Hasil uji Wilcoxon Signed-Rank Test menunjukkan bahwa perbedaan performa kedua model tidak signifikan secara statistik ($p = 0,809$). Meskipun demikian, XGBoost Regressor direkomendasikan sebagai model yang lebih sesuai untuk kasus prediksi pada penelitian ini karena secara konsisten menghasilkan nilai MAE dan RMSE yang lebih rendah serta nilai R^2 yang lebih tinggi dibandingkan Random Forest Regressor.

Kata kunci: energi surya, lama penyinaran matahari, regresi, *Random Forest*, *XGBoost*

ABSTRACT

REVA NURNADIA SYAPA. *A Comparison of Random Forest and XGBoost Algorithms for Predicting Sunshine Duration to Support Solar Energy Analysis in Kupang City*.
Supervised by Dr. Meilia Nur Indah Susanti, S.T., M.Kom.

Kupang City, East Nusa Tenggara, has high solar energy potential; therefore, information on sunshine duration is needed as one of the supporting parameters for solar energy potential analysis and the planning of Solar Power Plants (PLTS). However, studies directly comparing the performance of the Random Forest and XGBoost algorithms in predicting sunshine duration as the target variable remain limited. This study aims to analyze and compare the performance of the Random Forest Regressor and XGBoost Regressor algorithms in predicting sunshine duration in Kupang City. The study adopted the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology using 1,793 daily observations collected from the El Tari Meteorological Station, Kupang, covering the period from 2016 to 2020, with six meteorological parameters used as predictor variables. The dataset was divided using a time-based split with the best ratio of 90:10, while hyperparameter optimization was performed using Grid Search Cross-Validation. The results showed that the XGBoost Regressor achieved a Mean Absolute Error (MAE) of 1.3210 hours, a Root Mean Squared Error (RMSE) of 1.6911 hours, and an R^2 score of 0.5225, whereas the Random Forest Regressor obtained an MAE of 1.3259 hours, an RMSE of 1.7118 hours, and an R^2 score of 0.5107. The Wilcoxon Signed-Rank Test indicated that the performance difference between the two models was not statistically significant ($p = 0.809$). Nevertheless, the XGBoost Regressor is recommended as the more suitable model for the prediction task in this study because it consistently achieved lower MAE and RMSE values as well as a higher R^2 score than the Random Forest Regressor.

Keywords: Random Forest, regression, solar energy, sunshine duration, XGBoost