

ABSTRAK

HANISA. Prediksi Curah Hujan Menggunakan Teknik *Stacking* dengan Algoritma *Random Forest* dan *XGBoost* (Studi Kasus: Kec. Serpong). Dibimbing oleh Martaleli Bettiza dan Feri Irawan.

Kecamatan Serpong, Kota Tangerang Selatan merupakan wilayah rentan banjir akibat curah hujan ekstrem. Salah satu kejadian signifikan terjadi pada 11 November 2024, ketika curah hujan mencapai 150 mm/hari dan menyebabkan banjir setinggi 50-150 cm yang menggenangi permukiman, jalan utama, dan fasilitas umum. Kondisi ini mendorong pengembangan model prediksi curah hujan 30 menit ke depan sebagai dasar sistem peringatan dini. Penelitian ini menggunakan teknik *stacking ensemble* dengan *Random Forest* dan *XGBoost* sebagai *base learner* serta *XGBoost* sebagai *meta learner*. Data yang digunakan berasal dari Stasiun AQMS Kota Tangerang Selatan tahun 2024, sebanyak 16.766 data dengan interval 30 menit. Pemodelan mencakup *hyperparameter tuning* menggunakan *Grid Search CV* dengan skema *time series split* ($K=5$), pembentukan *meta feature* berbasis *out of fold prediction*, dan pelatihan *meta learner*. Hasil pengujian menunjukkan model *stacking* mencapai RMSE 0,0517, MAE 0,0085 dan R^2 0,8257. Sebagai perbandingan model tunggal, *XGBoost* menghasilkan performa terbaik dengan RMSE 0,0317, MAE 0,0048 dan R^2 0,9343, sedangkan *Random Forest* menghasilkan RMSE 0,0424, MAE 0,0122 dan R^2 sebesar 0,8828. Hasil ini menunjukkan bahwa teknik *stacking* pada penelitian ini belum mampu mengungguli performa model tunggal. Hal ini dipengaruhi oleh karakteristik data, kombinasi *base learner* dan *meta learner*, serta kualitas *meta feature* yang dihasilkan.

Kata kunci: Prediksi curah hujan, *Stacking Ensemble*, *Random Forest*, *XGBoost*, *Machine Learning*

ABSTRACT

HANISA. *Rainfall Prediction Using Stacking Technique with Random Forest and XGBoost Algorithms (Case Study: Serpong District)*. Supervised by Martaleli Bettiza and Feri Irawan.

Serpong District, South Tangerang City, is a region prone to flooding due to extreme rainfall. One significant incident occurred on November 11, 2024, when heavy rainfall caused flooding that inundated residential areas, main roads, and public facilities. This situation motivated the development of a model to predict rainfall 30 minutes ahead as a basis for an early warning system. This study employed a stacking ensemble approach using Random Forest and XGBoost as the base learners and XGBoost as the meta learner. The dataset was obtained from the AQMS Station of South Tangerang City in 2024, consisting of 16,766 observations recorded at 30-minute intervals. The modeling process included hyperparameter tuning using Grid Search Cross-Validation with a Time Series Split scheme ($K = 5$), followed by the generation of meta-features based on out-of-fold predictions. The results showed that the stacking model achieved an RMSE of 0.0517, an MAE of 0.0085, and an R^2 of 0.8257. In comparison, the single XGBoost model achieved the best performance, with an RMSE of 0.0317, an MAE of 0.0048, and an R^2 of 0.9343, while the Random Forest model achieved an RMSE of 0.0424, an MAE of 0.0122, and an R^2 of 0.8828. These results indicate that the stacking technique employed in this study did not outperform the best single model. This outcome is likely influenced by the characteristics of the dataset, the combination of the base learners and the meta learner, and the quality of the generated meta-features.

Keywords: Rainfall Forecasting, Stacking Ensemble, Random Forest, XGBoost, Machine Learning