Abstract

Forecasting wave height is essential for managing coastal activities, as it mitigates risks and losses in maritime transportation and informs coastal infrastructure design. Numerical wave modeling is frequently employed in forecasting wave heights. However, this methodology requires significant computational time and high grid resolution to obtain more accurate forecasts. Meanwhile, statistical approaches employing trend analysis of historical data for wave height prediction exhibit limitations in addressing extreme scenarios and characterising intricate nonlinear interactions. This work proposes using a machine learning model called CatBoost to predict wave height time series. We choose high-resolution wave data from nested wave simulations utilising the SWAN model at Pacitan Beach, East Java, Indonesia. We forecasted for intervals of 1, 5, 7, and 14 days, utilising diverse durations of wave data. Furthermore, the Catboost model's efficacy is evaluated compared to other boosting models, including Adaboost and XGboost. We conducted hyperparameter optimisation with GridSearch to get optimal performance with Catboost. The results demonstrated that the CatBoost model attained the maximum accuracy in all forecasting situations, with an RMSE of 0.0121 and an R2 value of 0.9986 over a 14-day prediction interval.

Keywords—*Wave Height Forecasting, CatBoost, AdaBoost, XGBoost, Hyperparameter Tuning*