ABSTRACT

Rainfall with varying intensity can have significant impacts on various sectors of life, particularly in urban areas such as Surabaya, which exhibits relatively fluctuating rainfall patterns. This variability makes rainfall prediction essential for supporting various activities and decision-making processes related to weather conditions. In recent years, machine learning approaches have been widely utilized to improve the accuracy of meteorological predictions. One of the commonly used algorithms is Extreme Gradient Boosting (XGBoost), an ensemble boosting-based algorithm capable of achieving high predictive performance and effectively handling nonlinear relationships among meteorological variables. However, XGBoost models tend to be difficult to interpret, thus requiring model interpretability methods such as SHapley Additive exPlanations (SHAP) to explain the contribution of each variable to the prediction results. This study aims to develop a rainfall classification model in Surabaya using XGBoost with SHAP-based interpretation. The data used consist of daily meteorological data from the Perak I Maritime Meteorological Station of BMKG from January 2018 to September 2025, including variables such as temperature, humidity, sunshine duration, as well as wind direction and speed. The target variable is categorized into four classes: no rain, light rain, moderate rain, and heavy to very heavy rain. The data were split with an 80:20 ratio, and hyperparameter tuning was performed using GridSearchCV and stratified k-fold cross-validation (k = 5). The Synthetic Minority Oversampling Technique (SMOTE) was applied to the training data due to class imbalance. The results show that the XGBoost model is able to classify rainfall categories with an accuracy of 80.04%, while SHAP analysis indicates that minimum temperature is the most influential variable in predicting moderate to heavy rainfall in Surabaya. This approach not only produces an accurate model but also provides interpretability regarding the meteorological factors influencing rainfall.

Keywords: Rainfall, XGBoost, SHAP, Classification, BMKG.