

Abstract

Weather is an essential aspect of life because it can affect human activities. Therefore, it is important for weather prediction to have high accuracy. One of the methods used to predict rainfall is data mining. In this study, a classification model was developed using the C5.0 algorithm to forecast rainfall in Bandung Regency. Then, the SMOTE algorithm was used to overcome imbalanced datasets. Weather data for the model development were obtained from the Meteorological, Climatological, and Geophysical Agency (BMKG) of Bandung for the years 2005 until 2017. Subsequently, the model was validated using a k-fold cross-validation. The results of the C5.0 test produced the highest accuracy of 92% for the imbalance dataset, while the accuracy of the addition of data using the SMOTE technique was 99%.