

ABSTRAK

Curah Hujan memberikan informasi tingkat turunnya hujan ke dalam beberapa kategori. Informasi tersebut disimpan di dalam dataset yang dikumpulkan oleh BMKG Jatiwangi, Majalengka. Fitur di dalam dataset antara lain yaitu suhu udara maksimal dan suhu udara minimal, lama penyinaran matahari, kelembapan udara, dan kecepatan angin. Variabel kelas kategori curah hujan termasuk multi kelas yang berjumlah 5 label yaitu Tidak Hujan/Berawan, Hujan Ringan, Hujan Sedang, Hujan Lebat, dan Hujan Sangat Lebat. Metode yang digunakan untuk mengklasifikasi curah hujan adalah *Naïve Bayes* dan *Random Forest*. Model *Random Forest* memiliki akurasi, presisi, *recall* dan nilai F1 tertinggi yang didapatkan dari *5-fold* dengan parameter *tree* 80 yaitu sebesar 65,795%, 61,360%, 65,795%, dan 63,124%. Sedangkan model *Gaussian Naïve Bayes* memiliki akurasi, presisi, *recall*, dan nilai F1 terbesar yang berasal dari *3-Fold* dengan parameter *alpha* 0.10 sebesar 67,666%, 63,510%, 67,666%, dan 64,685%. Dengan demikian, model klasifikasi *Naïve Bayes* lebih baik daripada model klasifikasi *Random Forest*.

Kata Kunci: klasifikasi, curah hujan, *naïve bayes*, *random forest*

ABSTRACT

Rainfall provided information on the level of rainfall in several categories. This information was stored in a dataset collected by BMKG Jatiwangi, Majalengka. Features in the dataset included maximum air temperature, minimum air temperature, duration of sunlight, air humidity, and wind speed. The rainfall category as a class variable included multiple classes with five labels: *Tidak Hujan/Berawan*, *Hujan Ringan*, *Hujan Sedang*, *Hujan Lebat*, and *Hujan Sangat Lebat*. The methods used to classify rainfall were Naïve Bayes and Random Forest. The Random Forest model had the highest accuracy, precision, recall, and F1 score obtained from 5-fold cross-validation with a tree parameter of 80, which were 65.795%, 61.360%, 65.795%, and 63.124%, respectively. Meanwhile, the Gaussian Naïve Bayes model had the greatest accuracy, precision, recall, and F1 score, coming from 3-Fold cross-validation with an alpha parameter of 0.10, which were 67.666%, 63.510%, 67.666%, and 64.685%, respectively. Thus, the Naïve Bayes classification model was better than the Random Forest classification model.

Keyword: classification, rainfall, *naïve bayes*, *random forest*

