

## ABSTRAK

Cuaca merupakan salah satu aspek penting yang berpengaruh terhadap aktivitas manusia. Adanya perubahan cuaca yang dipengaruhi oleh berbagai faktor seperti suhu, kelembapan udara, angin, waktu, dan lokasi, menjadikan pentingnya untuk mengetahui kemungkinan cuaca yang akan terjadi guna menghindari dan mempersiapkan solusi dari dampak yang ditimbulkan. Kemungkinan cuaca yang akan terjadi dapat ditentukan dengan lebih akurat menggunakan metode klasifikasi cuaca yang baik. Pada penelitian ini metode klasifikasi yang digunakan adalah regresi logistik dan *random forest* dengan menerapkan teknik *oversampling* SMOTE (*synthetic minority oversampling technique*) pada masing-masing metode. Peneliti membandingkan kedua metode tersebut menggunakan data cuaca di wilayah Jawa Tengah yang kemudian dibagi dalam tiga proporsi data latih yang berbeda, yaitu 60%, 70% dan 80%, dengan melihat kebaikan modelnya menggunakan nilai *area under curve* (AUC). Rata-rata AUC dari regresi logistik, regresi logistik dengan SMOTE, *random forest*, dan *random forest* dengan SMOTE adalah 0,6923, 0,5794, 0,7419, dan 0,7192. Berdasarkan hasil analisis kedua metode tersebut, dapat disimpulkan bahwa metode SMOTE tidak cocok digunakan untuk klasifikasi cuaca di wilayah Jawa Tengah karena menjadikan nilai AUC lebih kecil dari sebelumnya. Namun demikian, baik sebelum maupun sesudah dilakukan *oversampling* dengan SMOTE, *random forest* merupakan metode yang lebih baik dibandingkan dengan regresi logistik. Nilai AUC metode *random forest* menghasilkan nilai yang cukup tinggi sehingga metode *random forest* layak digunakan untuk mengklasifikasi cuaca wilayah Jawa Tengah.

**Kata kunci:** klasifikasi cuaca, regresi logistik, *random forest*, SMOTE, AUC.



## **ABSTRACT**

*Weather is one important aspect that influences human activity. There are changes in the weather that are influenced by various factors such as temperature, humidity, wind, time, and location, making it important to know the possibility of the weather that will occur in order to avoid and prepare solutions for the impacts. The possibility of weather that will occur can be determined more accurately using good weather classification methods. In this research, the classification method used was logistic regression and random forest by applying the SMOTE (synthetic minority oversampling technique) oversampling techniques to each method. Researchers compared the two methods using weather data in the Central Java region which were then divided into three different proportions of training data, namely 60%, 70% and 80%, by looking at the goodness of the model using the area under curve (AUC) value. The average AUC of logistic regression, logistic regression with SMOTE, random forest, and random forest with SMOTE are 0.6923, 0.5694, 0.7419, and 0.7192. Based on the results of the analysis of the two methods, it can be concluded that the SMOTE method is not suitable for weather classification in the Central Java region because it makes the AUC value smaller than before. However, both before and after oversampling with SMOTE, random forest is a better method than logistic regression. The AUC value of the random forest method produces a high enough value so that the random forest method is feasible to use to classify the weather in the Central Java region..*

**Keywords:** *weather classification, logistic regression, random forest, SMOTE, AUC.*

