

INTISARI

Perbandingan Metode Klasifikasi *Rotation Forest* dan *Random Forest* pada Data Tanah Longsor

Oleh

Reta Salma Adhani

19/442606/PA/19355

Pada studi *machine learning*, analisis klasifikasi bertujuan untuk meminimalkan kesalahan klasifikasi dan juga memaksimalkan hasil akurasi prediksi. Klasifikasi adalah pembelajaran penting untuk pengenalan pola (karakteristik pada data). Studi ini menganalisis dan membandingkan model kerentanan tanah longsor menggunakan algoritma *Random Forest* (RF), dan *Rotation Forest* (RoF). Penelitian ini dilakukan pada set data tertentu, di Italia tengah yang terdiri dari total 7.360 unit lereng yang mencakup area 4095 km^2 . Unit-unit lereng diklasifikasikan menjadi unit-unit dengan atau tanpa longsor. Selain itu, setiap unit lereng juga telah diberi 26 atribut yang digunakan sebagai variabel independen.

Random Forest merupakan algoritma *machine learning* dengan membentuk lebih dari satu pohon keputusan sehingga seperti hutan (*forest*) dan klasifikasi dengan metode ini menggunakan suara keputusan terbanyak (*majority vote*). Sedangkan *Rotation Forest* merupakan pengembangan dari *Random Forest* dengan menggunakan PCA untuk membentuk matriks rotasi yang bertujuan mentransformasikan himpunan data awal yang akan disusun ulang menjadi data untuk membangun pohon keputusan yang saling bebas satu sama lain. Akan dibandingkan performa dari model klasifikasi metode *random Forest* dan *Rotation Forest* dengan *Confusion Matrix* dan *Receiver Operating Characteristic* (ROC). Hasil penelitian menunjukkan bahwa *Rotation Forest* memiliki performa yang lebih baik dibandingkan dengan *Random Forest*.

Kata Kunci: *Machine Learning*, Klasifikasi, *Rotation Forest*, *Random Forest*

ABSTRACT

Comparison of Random Forest and Rotation Forest Classification Methods on Landslide Data

By

Reta Salma Adhani

19/442606/PA/19355

In machine learning studies, classification analysis aims to minimize classification error and also maximize prediction accuracy results. Classification is an important learning for pattern recognition (characteristics in data). This study analyzes and compares landslide susceptibility models using Random Forest (RF), and Rotation Forest (RoF) algorithms. The study was conducted on a specific dataset, in central Italy consisting of a total of 7.360 slope units covering an area of 4.095 km². The slope units were classified into units with or without landslides. In addition, each slope unit has also been assigned 26 attributes that are used as independent variables.

Random Forest is a machine learning algorithm by forming more than one decision tree so that it is like a forest and classification with this method uses the majority vote. While Rotation Forest is the development of Random Forest by using PCA to form a rotation matrix that aims to transform the initial data set which will be rearranged into data to build decision trees that are independent of each other. The performance of the classification model of the Random Forest and Rotation Forest methods will be compared with the Confusion Matrix and Receiver Operating Characteristic (ROC). The results showed that Rotation Forest has better performance than Random Forest.

Keywords: *Machine Learning, Classification, Rotation Forest, Random Forest*