

ABSTRAK

Tanah longsor merupakan salah satu bencana alam yang paling merusak karena dapat mengakibatkan perubahan drastis dalam morfologi lingkungan dan kerusakan pada struktur alami dan buatan di bumi. Di wilayah Provinsi Daerah Istimewa Yogyakarta (DIY), bencana tanah longsor telah menghasilkan kerugian ekonomi dan korban jiwa yang besar dalam kurun waktu tahun 2015-2019. Dengan melihat dampak tersebut, maka diperlukan usaha untuk meminimalkan dampak dengan mendeteksi secara dini tanda-tanda dan kemungkinan terjadinya longsor. *Random Forest* (RF) adalah sebuah algoritma *machine learning* yang dapat digunakan untuk memprediksi kejadian tanah longsor. Pada penelitian ini, algoritma RF diterapkan ke dalam dua bentuk model, yaitu model dengan parameter *default* dan model dengan parameter *tuning*. Algoritma *Decision Tree* digunakan sebagai metode pembandingan dalam penilaian kinerja sistem.

Penelitian ini menggunakan data kejadian bencana yang terdiri dari 3.848 *record* data dan 21 atribut. Data tersebut diolah dan dianalisis untuk mengetahui korelasi antara faktor pengondisian tanah longsor dengan kejadian bencana. *Dataset* kejadian tanah longsor yang terbentuk dibagi menjadi 70% data pelatihan untuk melatih model pengklasifikasi *machine learning*, serta 30% data pengujian dan data validasi untuk pengujian model tersebut. Parameter prediksi yang digunakan antara lain, curah hujan harian (CHH), curah hujan kumulatif tiga hari (CHK_3H), curah hujan kumulatif satu bulan (CHK_1B), jenis tanah, kemiringan lereng, ketinggian wilayah, dan penggunaan lahan.

Hasil pengujian kinerja menunjukkan bahwa algoritma RF dapat diterapkan untuk memprediksi tanah longsor yang dipicu oleh curah hujan di wilayah Provinsi Daerah Istimewa Yogyakarta. Hal tersebut terlihat pada model *Random Forest* dengan parameter *tuning* yang memperoleh hasil kinerja terbaik dengan akurasi sebesar 87,65%, presisi sebesar 89,66%, dan *recall* sebesar 60,47% dalam pengujian sistem menggunakan data uji. Model tersebut juga menghasilkan akurasi dan presisi tertinggi dengan nilai masing-masing 84,62% dan 45,95% pada pengujian sistem menggunakan data validasi. Sementara itu, nilai *recall* tertinggi sebesar 46,34% diperoleh oleh model *Random Forest* dengan parameter *default*.

Kata Kunci: *machine learning, prediksi, tanah longsor, random forest (RF), decision tree (DT).*

ABSTRACT

Landslides are one of the most destructive natural disasters because they can cause drastic changes in environmental morphology and damage to natural and artificial structures on earth. In the Special Region of Yogyakarta (DIY), landslides have resulted in large economic losses and casualties in the 2015-2019 period. By looking at these impacts, efforts are needed to minimize the impact by detecting early signs and possibility of landslides. Random Forest (RF) is a machine learning algorithm that can be used to predict landslides. In this study, the RF algorithm is applied to two models, which are a model with default parameters and a model with tuning parameters. Decision Tree Algorithm is used as a comparison method in assessing system performance.

This study uses disaster event data consisting of 3,848 data records and 21 attributes. The data is processed and analyzed to determine the correlation between the conditioning factor of landslides and the occurrence of disasters. The landslide event dataset that is formed is divided into 70% training data for training the machine learning classifier models, and 30% testing data and validation data for testing the model. Prediction parameters used are daily rainfall (CHH), three-day cumulative rainfall (CHK_3H), one-month cumulative rainfall (CHK_1B), soil type, slope, area elevation, and land use.

The results of the performance test show that the RF algorithm can be applied to predict landslides triggered by rainfall in the Special Region of Yogyakarta Province. This can be seen in the Random Forest model with tuning parameters that get the best performance results with 87.65% accuracy, 89.66% precision, and 60.47% recall in system testing using test data. The model also produces the highest accuracy and precision with values of 84.62% and 45.95% respectively in system testing using validation data. Meanwhile, the highest recall value of 46.34% is obtained by the Random Forest model with default parameters.

Keyword: *machine learning, prediction, landslide, random forest (RF), decision tree (DT).*