

ABSTRAK

Banjir merupakan salah satu bencana alam yang sering terjadi di Kabupaten Bandung dan memberikan dampak signifikan terhadap aspek sosial, ekonomi, serta lingkungan. Berdasarkan data dari BNPB periode 2019 – 2023, tercatat 446 kejadian banjir dengan berbagai data kerugian. Penelitian sebelumnya menerapkan metode prediksi banjir seperti *K-Nearest Neighbor*, *Naive Bayes*, *Support Vector Machine*, namun masih memiliki keterbatasan dari segi akurasi, efisiensi komputasi dan parameter data. Sebagai solusi, penelitian ini mengimplementasikan algoritma *Random Forest*, yang dikenal memiliki performa lebih baik. Tujuan dari penelitian ini untuk melakukan klasifikasi dan prediksi tingkat kerawanan banjir di Kabupaten Bandung secara akurat, serta menghasilkan sistem yang dapat mendukung perencanaan mitigasi bencana berbasis data. Variabel yang digunakan dalam model meliputi data curah hujan, jenis tanah, penggunaan lahan, dan data historis banjir dari tahun 2019 hingga 2023. Sebagai Metodologinya penelitian ini dilakukan dengan tahapan : pengumpulan data dari beberapa instansi yang terkait yaitu BMKG, BNPB, dan Bappeda. *Preprocessing* data dan Eksplorasi data, pemodelan menggunakan algoritma *Random Forest* dengan bantuan *library Pycaret* sebagai perbandingan algoritma yang tepat berdasarkan akurasi tertinggi dan *scikit-learn* sebagai penentuan algoritma terbaik, serta evaluasi model menggunakan metrik klasifikasi seperti akurasi, *precision*, *recall*, *F1-score*, AUC, Kappa dan MCC. Hasil penelitian menunjukkan bahwa model *Random Forest* yang telah di optimasi memiliki akurasi sebesar 96%, AUC 99%, dan rata-rata *Precision*, *recall*, dan *F1-score* di atas 90%. Model ini terbukti mampu membedakan wilayah dengan tingkat risiko banjir tinggi dan rendah secara sangat baik. Model prediksi tersebut kemudian diintegrasikan ke dalam sebuah aplikasi berbasis web bernama CaaIR (Cek Area Caah dan Informasi Rawan), yang dibangun menggunakan *framework React* untuk antarmuka pengguna dan *Flask* untuk *backend*. Aplikasi ini memungkinkan pengguna memasukkan data parameter wilayah dan memperoleh prediksi tingkat kerawanan banjir secara instan.

Kata Kunci : Banjir, *Machine Learning*, *Random Forest*, Prediksi Bencana, *PyCaret*, Kabupaten Bandung.

ABSTRACT

Floods are one of the natural disasters that frequently occur in Bandung Regency, causing significant impacts on social, economic, and environmental aspects. Based on data from BNPB for the period 2019–2023, a total of 446 flood incidents were recorded, along with various associated damages. Previous studies have applied flood prediction methods such as K-Nearest Neighbor, Naive Bayes, and Support Vector Machine; however, these approaches still have limitations in terms of accuracy, computational efficiency, and data parameters. As a solution, this study implements the Random Forest algorithm, which is known for its superior performance. The purpose of this research is to classify and predict the level of flood vulnerability in Bandung Regency accurately, and to develop a system that supports data-driven disaster mitigation planning. The model uses variables including rainfall, soil type, land use, and historical flood data from 2019 to 2023. The methodology includes several stages: data collection from relevant institutions (BMKG, BNPB, and Bappeda), data preprocessing and exploration, modeling using the Random Forest algorithm with the help of PyCaret to compare multiple algorithms based on highest accuracy, and Scikit-Learn to finalize the best model. The model is then evaluated using classification metrics such as accuracy, precision, recall, F1-score, AUC, Kappa, and MCC. The results show that the optimized Random Forest model achieved an accuracy of 96%, an AUC score of 99%, and average precision, recall, and F1-score above 90%. This model is proven to effectively distinguish between areas with high and low flood risk. The predictive model is then integrated into a web-based application called CaaIR (Cek Area Caah dan Informasi Rawan), built using React for the user interface and Flask for the backend. This application enables users to input regional parameters and instantly receive predictions of flood risk levels.

Keywords: *Flood, Machine Learning, Random Forest, Disaster Prediction, PyCaret, Bandung Regency*