

INTISARI

KLASIFIKASI LEVEL BANJIR MENGGUNAKAN ALGORITMA *RANDOM FOREST* DAN *SUPPORT VECTOR MACHINE*

Oleh

Larasati Syarafina Qamarani

20/455386/PA/19601

Banjir merupakan salah satu bencana alam yang umum terjadi di Indonesia. Terjadinya banjir menyebabkan kerugian besar dari berbagai aspek seperti korban makhluk hidup, ekonomi, lingkungan, hingga infrastruktur. Semakin tinggi tingkat banjir, semakin besar dampak yang diakibatkan oleh bencana ini. Rekapitulasi dampak dari setiap peristiwa banjir seperti durasi, ketinggian air, dan jumlah terdampak dianalisis untuk mengidentifikasi karakteristik banjir sesuai dengan tingkatan banjir yang terjadi. Data iklim seperti temperatur, kelembaban, curah hujan, dan kecepatan angin digunakan sebagai parameter ciri terjadinya banjir. Penelitian ini bertujuan untuk memperoleh hasil klasifikasi data setiap peristiwa banjir dengan tolak ukur hasil sistem klasifikasi menggunakan *Random Forest* dan *Support Vector Machine* serta evaluasi akurasi dengan *Confusion Matrix* yang diharapkan dapat menjadi pertimbangan untuk mengambil keputusan saat terjadi banjir dengan meminimalisir kerugian akibat banjir.

Proses penelitian menggunakan data rekapitulasi banjir oleh BPBD DKI Jakarta melalui *website* Satu Data Jakarta serta data iklim dari Stasiun Geofisika BMKG yang diambil dalam rentang waktu yang sama yaitu tahun 2013-2020. Sistem klasifikasi *Random Forest* mencapai hasil yang lebih tinggi dengan rentang evaluasi hampir sempurna, sebesar 99%. Sementara sistem klasifikasi *Support Vector Machine* mencapai hasil sedikit lebih rendah di angka 98%. Kedua sistem mengalami *overfitting* dikarenakan adanya ketidakseimbangan pada data awal sehingga model cenderung mengidentifikasi data pada kelas mayoritas, namun kondisi *overfitting* menurun seiring bertambahnya data saat pembuatan model. Data asli tidak diberi perlakuan untuk mengatasi ketidakmerataan data antar kelas untuk menjaga keaslian data dan melatih model sesuai dengan kejadian asli agar lebih aplikatif dan relevan untuk dipakai di dunia nyata.

Kata kunci—Banjir, Level, Iklim, *Machine Learning*, *Random Forest*, *Support Vector Machine*

ABSTRACT

CLASSIFICATION OF FLOOD DANGER LEVELS USING RANDOM FOREST ALGORITHM AND SUPPORT VECTOR MACHINE

by

Larasati Syarafina Qamarani

20/455386/PA/19601

Floods are one of the most common natural disasters in Indonesia, causing substantial economic, environmental, and infrastructural damage, as well as human casualties. The severity of a flood's impact correlates with its magnitude, with higher flood levels resulting in more significant consequences. This study analyzes various flood event parameters, such as duration, water level, and the number of individuals affected, to identify flood characteristics based on the severity of the event. Climate variables including temperature, humidity, rainfall, and wind speed are utilized as parameters to characterize flood occurrences. The primary aim of this research is to classify flood events using Random Forest and Support Vector Machine (SVM) algorithms and to evaluate the result of these classifications using a Confusion Matrix. The results are intended to inform decision-making processes during floods to mitigate associated losses.

Data for this study were obtained from the BPBD DKI Jakarta through the Satu Data Jakarta website, alongside climate data from the BMKG, spanning the years 2013–2020. The Random Forest classification system achieved superior performance, with an evaluation accuracy of 99%. Similarly, the SVM algorithm performed well, with an accuracy of 98%. Both models initially exhibited overfitting due to imbalanced data. However, this issue diminished as the dataset size increased during the modeling process. The study maintained the authenticity of the imbalanced data to ensure the models were trained on original events, enhancing their applicability and relevance for real-world use.

Keywords—*Flood, Level, Climate, Machine Learning, Random Forest, Support Vector Machine*