



INTISARI

Kekeringan merupakan kejadian iklim yang sering terjadi dengan dampak signifikan pada manusia, hewan, dan tumbuhan, termasuk di Kulon Progo, Daerah Istimewa Yogyakarta. Strategi mitigasi berbasis data iklim sangat penting untuk mengatasi dampaknya, dan *machine learning* telah menjadi alat yang efektif untuk memprediksi kondisi kekeringan. Penelitian ini memitigasi kekeringan menggunakan metode klasifikasi dan *time series forecasting*. Dalam pembangunan model, terdapat beberapa masalah yang menyebabkan ketidakefektifan model, salah satunya adalah data yang tidak seimbang. Dengan berdasarkan latar belakang tersebut, penelitian ini memiliki beberapa tujuan seperti pengujian penanganan *imbalanced data*, menentukan model klasifikasi terbaik untuk kekeringan, menentukan parameter iklim yang paling signifikan untuk pelatihan model klasifikasi kekeringan, menguji pengaruh model *time series forecasting* untuk parameter iklim yang berkaitan dengan kekeringan, serta menarik kesimpulan temporal mengenai kekeringan. Hal ini didasari dari mengenai pentingnya pemanfaatan teknologi yang berbasis data untuk menangkal beberapa kerugian dari terjadinya kekeringan. Metode indeks kekeringan yang dipakai pada penelitian kali ini adalah Standardized Precipitation Evapotranspiration Index (SPEI). Selain itu, dilakukan juga *imbalanced data handling* untuk data kelas kekeringan yang tidak seimbang. *Imbalanced data handling* pada penelitian kali ini menggunakan teknik SMOTE, ADASYN, *random oversampling*, dan *random undersampling*. Untuk model klasifikasi kekeringan, digunakan model Decision Tree, Random Forest, Extra Trees, CatBoost, dan LightGBM. Sementara itu, model *time series forecasting* yang digunakan adalah Prophet, SARIMA, Holt-Winters Exponential Smoothing, dan LSTM. Hasilnya, metode *imbalanced data handling* yang memberikan pengaruh akurasi paling tinggi adalah *random oversampling*. Model klasifikasi Extra Trees menunjukkan performa paling baik dalam melakukan klasifikasi kekeringan dengan akurasi 0,98. Parameter iklim yang paling signifikan dalam melakukan tugas klasifikasi kekeringan adalah *total precipitation* dan *rain precipitation*. Untuk model *time series forecasting*, didapatkan model LSTM konsisten memiliki performa yang bagus untuk variabel *rain precipitation* yang memiliki pola yang lebih baik dan variabel SPEI yang memiliki pola yang tidak beraturan. Untuk analisis temporal, didapatkan bulan Juli dan Agustus memiliki potensi kekeringan yang paling besar dibandingkan dengan bulan lainnya.

Kata kunci : klasifikasi kekeringan, prediksi kekeringan, *machine learning*, analisis temporal



ABSTRACT

Drought is a frequent climate event with significant impacts on humans, animals and plants, including in Kulon Progo, Special Region of Yogyakarta. Climate data-based mitigation strategies are essential to address its impacts, and machine learning has become an effective tool for predicting drought conditions. This research mitigates drought using classification and time series forecasting methods. In building the model, there are several problems that cause the model to be ineffective, one of which is unbalanced data. Based on this background, this research has several objectives such as testing the handling of imbalanced data, determining the best classification model for drought, determining the most significant climate parameters for training drought classification models, testing the influence of the time series forecasting model for climate parameters related to drought, as well as drawing temporal conclusions regarding drought. This is based on the importance of using data-based technology to ward off several losses from drought. The drought index method used in this research is the Standardized Precipitation Evapotranspiration Index (SPEI). Apart from that, imbalanced data handling was also carried out for unbalanced drought class data. Imbalanced data handling in this research uses the SMOTE, ADASYN, random oversampling, and random undersampling techniques. For the drought classification model, the Decision Tree, Random Forest, Extra Trees, CatBoost, and LightGBM models are used. Meanwhile, the time series forecasting models used are Prophet, SARIMA, Holt-Winters Exponential Smoothing, and LSTM. As a result, the imbalanced data handling method that has the highest impact on accuracy is random oversampling. The Extra Trees classification model shows the best performance in classifying drought with an accuracy of 0.98. The most significant climate parameters in carrying out drought classification tasks are total precipitation and rain precipitation. For the time series forecasting model, it was found that the LSTM model consistently had good performance for the rain precipitation variable which had a better pattern and the SPEI variable which had an irregular pattern. For temporal analysis, it was found that July and August have the greatest potential for drought compared to other months.

Keywords : drought classification, drought prediction, machine learning, temporal analysis