

INTISARI

OPTIMASI MODEL KLASIFIKASI GEMPA VULKANIK GUNUNG MERAPI DENGAN DATA SINTETIK MENGGUNAKAN CGAN DAN CVAE UNTUK MENGATASI *IMBALANCE DATASET*

Oleh

Ilham Sidik

23/531065/PPA/06758

Ketidakseimbangan data merupakan permasalahan umum dalam klasifikasi gempa vulkanik, terutama pada kelas minoritas seperti *Volcano-Tectonic A* (VTA) dan *Volcano-Tectonic B* (VTB) yang berperan sebagai prekursor aktivitas erupsi. Ketidakseimbangan ini menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas dan berpotensi mengabaikan sinyal penting. Penelitian ini bertujuan untuk mengevaluasi efektivitas pemanfaatan data sintetik dalam mengatasi permasalahan tersebut dengan membandingkan tiga pendekatan, yaitu augmentasi berbasis nilai fisis, *Conditional Generative Adversarial Network* (CGAN), dan *Conditional Variational Autoencoder* (CVAE). Penelitian ini menggunakan data seismik Gunung Merapi periode Oktober 2019 dengan ekstraksi tiga fitur utama, yaitu frekuensi dominan, amplitudo maksimum, dan durasi sinyal. Model klasifikasi XGBoost digunakan untuk mengevaluasi performa pada setiap skenario pelatihan. Hasil eksperimen menunjukkan bahwa pendekatan CGAN memberikan peningkatan kinerja terbaik, dengan peningkatan *recall* kelas minoritas hingga lebih dari 85% dan peningkatan nilai akurasi menjadi 80% dibandingkan *baseline* yang hanya 76%, serta peningkatan nilai *F1-score* secara konsisten. Selain itu, uji ketahanan terhadap *noise* menunjukkan bahwa model berbasis CGAN mampu mempertahankan akurasi hingga sekitar 64% pada *noise level* 0.5, lebih stabil dibandingkan augmentasi fisis dan CVAE. Evaluasi kualitas data sintetik menggunakan *Mean Squared Error* (MSE) menunjukkan nilai di bawah 0.01 untuk seluruh kelas gempa, yang mengindikasikan kesesuaian distribusi fitur antara data asli dan data sintetik. Berdasarkan hasil tersebut, penelitian ini menyimpulkan bahwa pemanfaatan data sintetik berbasis CGAN efektif dalam meningkatkan kinerja dan *robustness* klasifikasi gempa vulkanik pada dataset tidak seimbang.

Kata Kunci: Klasifikasi gempa vulkanik, data sintetik, ketidakseimbangan data, augmentasi data, GAN, VAE, *machine learning*, Gunung Merapi.

ABSTRACT

OPTIMIZATION OF VOLCANIC EARTHQUAKE CLASSIFICATION MODEL OF MOUNT MERAPI WITH SYNTHETIC DATA USING CGAN AND CVAE TO OVERCOME DATASET IMBALANCE

By

Ilham Sidik

23/531065/PPA/06758

Data imbalance is a common problem in volcanic earthquake classification, especially in minority classes such as Volcano-Tectonic A (VTA) and Volcano-Tectonic B (VTB), which act as precursors to eruptive activity. This imbalance causes the classification model to tend to be biased towards the majority class and potentially ignore important signals. This study aims to disseminate the effectiveness of utilizing synthetic data in addressing this problem by comparing three approaches: physical value-based augmentation, Conditional Generative Adversarial Network (CGAN), and Conditional Variational Autoencoder (CVAE). This study uses seismic data from Mount Merapi for the period of October 2019 with the extraction of three main features: dominant frequency, maximum amplitude, and signal duration. The XGBoost classification model is used to examine the performance in each training scenario. Experimental results show that the CGAN approach provides the best performance improvement, with an increase in minority class recall of more than 85% and an increase in accuracy to 80% compared to the baseline of only 76%, as well as a consistent increase in F1-score. Furthermore, noise robustness tests show that the CGAN-based model is able to maintain accuracy up to approximately 64% at a noise level of 0.5, more stable than physical augmentation and CVAE. Synthetic data quality evaluation using Mean Squared Error (MSE) shows values below 0.01 for all earthquake classes, indicating consistent feature distribution between the original and synthetic data. Based on these results, this study concludes that the use of CGAN-based synthetic data is effective in improving the performance and robustness of volcanic earthquake classification on imbalanced datasets.

Keywords: Volcanic earthquake classification, synthetic data, data imbalance, data augmentation, GAN, VAE, machine learning, Mount Merapi.