

INTISARI

KLASIFIKASI ERUPSI GUNUNGAPI MERAPI BERDASARKAN ENERGI AKTIVITAS SEISMIC HARIAN DENGAN *ADAPTIVE* *RANDOM FOREST*

Oleh

Andre Krismantara

23/526369/PPA/06650

Gunung Merapi memiliki karakteristik erupsi dengan pola seismisitas yang dinamis dan terus berubah seiring waktu (*concept drift*), serta memiliki distribusi data kejadian erupsi yang sangat tidak seimbang dibandingkan kondisi non-erupsi. Kondisi ini menyebabkan model klasifikasi statis konvensional mengalami kegagalan karena akurasi terus menurun akibat perubahan pola data yang tidak terprediksi, seperti yang terjadi pada perubahan karakteristik energi seismik tahun 2010, sehingga menuntut proses pelatihan ulang yang tidak efisien dari segi waktu dan sumber daya. Oleh karena itu, penelitian ini mengembangkan sistem klasifikasi erupsi berbasis energi aktivitas seismik harian menggunakan pendekatan *Online Machine Learning* yang dirancang untuk beradaptasi secara kontinu terhadap perubahan data tanpa perlu pelatihan ulang total, sekaligus menjaga stabilitas akurasi di tengah ketidakseimbangan kelas.

Sebagai solusi teknis, penelitian ini mengusulkan penggunaan algoritma *Adaptive Random Forest* (ARF) dan *Hoeffding Tree* (HT) yang diintegrasikan dengan mekanisme deteksi *drift* ADWIN serta strategi adaptif baru (disebut IV.A) yang menggabungkan *confidence-aware online SMOTE* dan *drift-aware light retraining*. Metodologi penelitian mencakup pemrosesan data energi seismik harian dari tahun 1991 hingga 2013 yang dikonversi ke format *streaming*, pelatihan model secara inkremental, serta evaluasi performa menggunakan *Cumulative F1-Score*, *Rolling F1-Score*, dan *confusion matrix*.

Hasil penelitian menunjukkan bahwa model ARF dengan strategi IV.A mencatatkan performa paling unggul dibandingkan model lainnya, dengan

cumulative F1-score rata-rata sebesar 0,89 dan tingkat stabilitas *rolling F1* terbaik (standar deviasi 0,0515). Model ini terbukti mampu menjaga konsistensi klasifikasi baik pada erupsi efusif maupun eksplosif, serta beradaptasi dengan baik saat terjadi *concept drift* ekstrem seperti pada erupsi besar tahun 2010, di mana model statis mengalami degradasi performa yang signifikan. Sebaliknya, model HT IV.A justru menunjukkan ketidakstabilan akibat *overfitting* terhadap data sintetis, sehingga pendekatan ARF IV.A disimpulkan sebagai metode yang paling efektif dan efisien untuk diimplementasikan dalam sistem peringatan dini kebencanaan berbasis data *streaming*.

Kata Kunci: Klasifikasi erupsi, *Online Machine Learning*, *Adaptive Random Forest*, *Hoeffding Tree*, *Concept Drift*, *Online SMOTE*.

ABSTRACT

CLASSIFICATION OF MERAPI VOLCANIC ERUPTIONS BASED ON DAILY SEISMIC ACTIVITY ENERGY USING ADAPTIVE RANDOM FOREST

By:

Andre Krismantara

23/526369/PPA/06650

Merapi volcano exhibits eruption characteristics with dynamic seismicity patterns that change over time (concept drift) and possesses a highly imbalanced distribution of eruption data compared to non-eruption conditions. These conditions cause conventional static classification models to fail, as their accuracy continuously degrades due to unpredictable changes in data patterns, such as the shift in seismic energy characteristics observed in 2010, thereby necessitating retraining processes that are inefficient in terms of time and resources. Consequently, this study develops an eruption classification system based on daily seismic activity energy using an Online Machine Learning approach designed to adapt continuously to data changes without the need for total retraining, while simultaneously maintaining accuracy stability amidst class imbalances.

As a technical solution, this research proposes the use of Adaptive Random Forest (ARF) and Hoeffding Tree (HT) algorithms integrated with the ADWIN drift detection mechanism and a novel adaptive strategy (termed IV.A) that combines confidence-aware online SMOTE and drift-aware light retraining. The research methodology encompasses processing daily seismic energy data from 1991 to 2013 converted into a streaming format, incremental model training, and performance evaluation using Cumulative F1-Score, Rolling F1-Score, and confusion matrices.

The results indicate that the ARF model with strategy IV.A achieved superior performance compared to other models, recording an average cumulative F1-score of 0.89 and the best rolling F1 stability (standard deviation of 0.0515).

This model proved capable of maintaining classification consistency for both effusive and explosive eruptions, as well as adapting effectively during extreme concept drift events like the major 2010 eruption, where static models experienced significant performance degradation. Conversely, the HT IV.A model demonstrated instability due to overfitting on synthetic data; thus, the ARF IV.A approach is concluded to be the most effective and efficient method for implementation in streaming data-based disaster early warning systems.

Keywords: Eruption classification, Online Machine Learning, Adaptive Random Forest, Hoeffding Tree, Concept Drift, Online SMOTE.