

ABSTRACT

Disease outbreaks have claimed many lives and had a major impact on the economic, social and lifestyle aspects of society globally. One of the major challenges in dealing with disease outbreaks is early detection, which is essential for reducing their impact. Early warning systems are a crucial component in controlling outbreaks at an early stage. This is important because outbreaks often appear as sudden and unexpected spikes in disease cases. Previous research using an unsupervised learning approach with a hybrid autoencoder deep learning model on multivariate spatio-temporal data was considered successful for detecting the Covid-19 outbreak in Italy. However, the research has not been tested to be able to detect early outbreaks in various conditions and has no reference to when the outbreak occurs and how long the optimal preparation time is needed to warn of the outbreak. Therefore, this master's research aims to create an outbreak early detection model that can be used in several outbreak case studies and that can be assessed for quantity performance.

This research builds a ground truth timeline modeling for outbreaks and alarms, so that the classification process can be carried out. Using the ground truth, the quantity performance of the baseline model (hybrid autoencoder deep learning model) can be obtained and used as the original baseline scenario. However, in order for the model to be adapted to the alarm time selection, the modification of the hybrid autoencoder deep learning model is required and used as the original modified scenario. This research also applies hyperparameter tuning using Particle Swarm Optimization (PSO) and Bayesian optimization-Gaussian processes (BO-GP). The application of hyperparameter tuning is made into two different scenarios so that the performance impact of the application can be observed. This research was tested on four datasets with different feature characteristics: Covid-19 outbreak in Italy, Covid-19 outbreak in Indonesia, Dengue outbreak in Central Java, and Tegal City Dengue outbreak. Various datasets were run on the four research scenarios previously mentioned.

In this study, all outbreak datasets were trained and tested for all scenarios using several different alarm time frames (t days before the outbreak). The results obtained in this study show that the model scenario using hyperparameter tuning was better than the baseline scenario (original). In both Particle Swarm Optimization (PSO) and Bayesian optimization-Gaussian process (BO-GP) scenarios, there is a trend of improvement in key performance metrics. The most notable improvement is seen in the fourth scenario, namely the model with BO-GP as the hyperparameter tuning method, where the highest values of accuracy and kappa are 0.90 and 0.74 on the Covid-19 case dataset in the Indonesian region with a preparation time of 7 days before the outbreak. Similarly, in other outbreak cases, the model using BO-GP obtained excellent performance and outperformed the other model scenarios in all

aspects. These results indicate that the proposed approach improves the performance measurement of early outbreak detection and opens up opportunities for further research to maintain high prediction accuracy while reducing the computational burden, thus making these models more efficient and effective in early outbreak detection.

Keywords : outbreak early detection, anomaly detection, unsupervised learning, hybrid autoencoder deep learning, Covid-19, Dengue, multivariate spatio-temporal data, hyperparameter tuning, PSO, BO-GP.

INTISARI

Wabah penyakit telah menyebabkan banyak korban jiwa dan memberikan dampak besar terhadap aspek ekonomi, sosial, dan gaya hidup masyarakat secara global. Salah satu tantangan besar dalam menghadapi wabah penyakit adalah deteksi dini, yang esensial untuk mengurangi dampaknya. Sistem peringatan dini adalah komponen krusial dalam mengendalikan wabah pada tahap awal. Hal ini penting karena wabah sering kali muncul sebagai lonjakan kasus penyakit yang mendadak dan tak terduga. Penelitian sebelumnya yang menggunakan pendekatan *unsupervised learning* dengan model *hybrid autoencoder deep learning* pada data *multivariat spatio-temporal* dianggap berhasil untuk mendeteksi wabah Covid-19 di Italia. Namun penelitian tersebut belum teruji untuk dapat mendeteksi dini wabah dalam berbagai macam kondisi dan tidak memiliki acuan kapan wabah itu terjadi dan berapa lama waktu persiapan optimal yang diperlukan dalam memberikan peringatan atas wabah tersebut. Oleh karena itu, penelitian magister ini bertujuan untuk membuat model deteksi dini wabah yang dapat digunakan dalam beberapa studi kasus wabah dan yang dapat dinilai performa kuantitasnya.

Penelitian ini membangun pemodelan *ground truth timeline* untuk wabah dan alarm, sehingga dapat dilakukan proses klasifikasi. Dengan adanya *ground truth*, performa kuantitas dari model *baseline* (model *hybrid autoencoder deep learning*) bisa didapatkan dan dijadikan *baseline* skenario orisinal. Namun agar model bisa disesuaikan dengan pemilihan waktu alarm, maka modifikasi dari model *hybrid autoencoder deep learning* diperlukan dan dijadikan sebagai skenario modifikasi orisinal. Penelitian ini juga menerapkan *hyperparameter tuning* dengan menggunakan *Particle Swarm Optimization* (PSO) dan *Bayesian Optimization - Gaussian Processes* (BO-GP). Penerapan *hyperparameter tuning* dijadikan dua skenario yang berbeda dari sebelumnya, sehingga bisa dilihat dampak performa atas penerapannya tersebut. Penelitian ini dicobakan pada empat *dataset* dengan karakteristik fitur yang berbeda, yaitu: wabah Covid-19 di Italia, wabah Covid-19 di Indonesia, wabah DBD di Jawa Tengah dan wabah DBD Kota Tegal. Berbagai *dataset* tersebut dijalankan pada empat skenario penelitian yang sudah disampaikan sebelumnya.

Dalam penelitian ini semua *dataset* wabah dilatih dan diujikan pada semua skenario dengan menggunakan beberapa kerangka waktu alarm yang berbeda-beda (t hari sebelum wabah). Hasil yang didapatkan pada penelitian tersebut menunjukkan bahwa skenario model yang menggunakan *hyperparameter tuning* lebih baik dari *baseline* skenario (orisinal). Pada skenario PSO *tuning* maupun BO-GP *tuning*, menunjukkan tren peningkatan pada metrik performa utama. Peningkatan paling menonjol terlihat pada skenario keempat yaitu model dengan BO-GP sebagai metode *hyperparameter tuning*, dimana nilai tertinggi akurasi dan kappa adalah 0,90 dan 0,74 pada *dataset* kasus Covid-19 di wilayah Indonesia dengan waktu persiapan 7 hari sebelum wabah.

Hal tersebut juga terjadi pada *dataset* kasus wabah lainnya, model yang menggunakan BO-GP memperoleh performa luar biasa dan mengungguli skenario model lain di semua aspek. Hasil ini mengindikasikan bahwa pendekatan yang diusulkan memperbaiki pengukuran kinerja deteksi dini wabah dan membuka peluang untuk penelitian lebih lanjut untuk mempertahankan akurasi prediksi yang tinggi sekaligus mengurangi beban komputasi, sehingga membuat model-model ini lebih efisien dan efektif deteksi dini wabah.

Kata kunci – deteksi dini wabah, deteksi anomali, *unsupervised learning*, *hybrid autoencoder deep learning*, Covid-19, DBD, data *multivariat spatio-temporal*, *hyperparameter tuning*, PSO, BO-GP.