

1. Pendahuluan

Latar Belakang

COVID-19 merupakan penyakit pandemik yang menyerang seluruh dunia mulai Desember 2019 dan pertama kali muncul di Wuhan, China. Pandemi ini menyebar begitu cepat, bahkan WHO menyatakan jumlah kasus per 11 November 2020 mencapai sekitar 52 juta orang di lebih dari 200 negara [2][9]. Kasus COVID-19 juga terjadi di Indonesia dan pada akhir tahun 2021 kasus COVID-19 di Indonesia mencapai 4 juta orang dengan angka kematian lebih dari 140 ribu orang. Jawa Barat merupakan salah satu provinsi di Indonesia dengan kasus COVID-19 tertinggi, khususnya Bandung. Tercatat lebih dari 100.000 kasus terkonfirmasi positif hingga Januari 2023. Oleh karena itu, pencegahan terhadap faktor-faktor yang mempercepat penyebaran virus ini harus segera dihentikan, salah satunya dengan membuat klasifikasi penyebaran dan menganalisis faktor apa saja yang mempengaruhinya [10].

Banyak teknik yang digunakan untuk mengklasifikasikan penyebaran penyakit, salah satunya dengan menggunakan machine learning atau deep learning. Metode jaringan saraf adalah metode yang populer digunakan dalam pembelajaran mendalam. ANN (Artificial Neural Network) adalah sistem komputasi yang diadaptasi dari sel-sel saraf di otak manusia. ANN telah digunakan dalam berbagai pemecahan masalah seperti identifikasi, klasifikasi, dan pengenalan data. Sedangkan CNN (Convolutional Neural Network) merupakan deep neural network pada *computer vision* yang dapat digunakan dalam pengolahan citra dan klasifikasi data serta dapat digunakan untuk mengklasifikasikan data time series [1][3][25]. Pada beberapa makalah terkait prediksi klasifikasi penyakit menggunakan metode ANN dan CNN memiliki kinerja yang baik, antara lain penelitian yang memprediksi penyebaran DBD di Indonesia menggunakan ANN (algoritma back propagation dengan 9 node) dan regresi berganda dengan nilai akurasi sebesar 87,16% [5]. Kemudian pada paper lain, CNN dibandingkan dengan metode lain yaitu vanilla CNN, vanilla NN, XGBOOST, random forest, dan LR dalam klasifikasi data khususnya time series dalam multivariate time series atau disingkat MVCNN. Hasilnya, CNN memiliki nilai akurasi yang lebih baik sebesar 0,02% dibandingkan dengan random forest. Melihat akurasi kinerja kedua metode tersebut, ANN dan CNN dapat digunakan untuk memprediksi penyebaran COVID-19.

Dalam tugas akhir ini bertujuan untuk menentukan kinerja mana yang lebih baik antara ANN dan CNN dalam model prediksi klasifikasi penyebaran COVID-19 yang dikombinasikan dengan teknik perluasan fitur pada data *time series*. Teknik perluasan fitur (*feature expansion*) adalah metode yang digunakan untuk memilih fitur terbaik dan menghilangkan atribut yang tidak relevan. Kemudian masing-masing model dilatih dan hasil akurasinya dianalisis untuk menentukan model terbaik. Metrik evaluasi yang digunakan adalah menghitung nilai *Akurasi*, *precision*, *recall*, *f1-score*, dan RMSE. Berikut beberapa kontribusi penelitian kami: Gunakan metode perluasan fitur untuk membangun model data terlatih, dan bandingkan kinerja metode ANN dan CNN pada model prediksi penyebaran COVID-19.

Topik dan Batasannya

Topik yang dibahas dalam tugas akhir ini adalah menentukan model terbaik untuk melakukan prediksi klasifikasi terhadap kasus COVID-19. Model yang dibandingkan adalah model ANN dan CNN dimana keduanya menggunakan teknik perluasan fitur (*feature expansion*) yang dibagi menjadi beberapa model, yaitu model 2, 3, 4, dan 5 bulan sebelum tahun target. Batasan masalah pada tugas akhir ini adalah data yang kami gunakan adalah data penduduk dan konfirmasi kasus COVID-19 di Kota Bandung. Data ini kami dapatkan dari Badan Pusat Statistik kota Bandung dan Dinas Kesehatan kota Bandung. Rentang waktu dari data ini dimulai dari November 2021 sampai dengan april 2022.

Tujuan

Tujuan dari penelitian pada tugas akhir ini adalah untuk membandingkan kinerja model ANN dan CNN yang menggunakan teknik perluasan fitur (*feature expansion*) pada kasus COVID-19 di kota Bandung. Nantinya penelitian ini diharapkan dapat memberi manfaat bagi pemerintah kota Bandung dan tenaga kesehatan agar dapat mengetahui faktor-faktor utama yang memicu naiknya kasus COVID-19 di kota Bandung.

Organisasi Tulisan

Organisasi tulisan pada laporan tugas akhir yang kami susun dimulai dengan pendahuluan, kemudian studi pustaka mengenai COVID-19, metode ANN, metode CNN, dan perluasan fitur (*feature expansion*). Dilanjutkan dengan sistem yang dibangun dan hasil pengujian berupa metrik evaluasi, yaitu *Akurasi*, *precision*, *recall*, *f1-score*, dan RMSE. Terakhir ditutup dengan kesimpulan.