

ABSTRAK

Wabah virus corona versi SARS-nCov2 telah terjadi di akhir tahun 2019 di Wuhan, Tiongkok. Karena kemampuan penularan yang supercepat, virus ini sudah bertransmisi lintas negara. Hampir semua negara terkena wabah virus ini, hingga PBB melalui WHO menjadikan wabah virus ini sebagai pandemi global. Penelitian ini menggunakan teknik klasifikasi *data mining* dengan algoritme C4.5 dengan alat bantu *Orange data mining*. Juga dilakukan pengolahan dengan bahasa pemrograman Python menggunakan *library* standar yang sudah disediakan seperti matplotlib, seaborn, numpy, pandas, dan decisionTree *Classification*. Klasifikasi dilakukan untuk mengelompokkan data kasus terkonfirmasi, meninggal, dan sembuh dari covid 19 menjadi tinggi dan rendah. Hasil klasifikasi didapatkan nilai AUC 0.871, yang menunjukkan klasifikasi kategori baik, nilai *accuracy* sebesar 93%, nilai *precision* sebesar 95%, dan nilai *recall* 96%. Pada pemodelan menggunakan Python, nilai akurasi yang didapatkan adalah 94%, nilai presisi adalah 98%, dan nilai *recall* yang didapatkan adalah 95%. Penelitian juga berhasil memprediksi tren kasus di Indonesia dengan menggunakan model fungsi *logistic*, model standar yang digunakan untuk memprediksi laju pertumbuhan populasi. Hasil prediksi yang didapatkan adalah rata-rata laju pertumbuhan virus adalah 0.0255, puncak pandemi terjadi pada hari ke-196 sampai 200 atau tanggal 14 sampai 18 September 2020, akhir pandemi diprediksi pada hari ke-717 atau tanggal 17 Februari 2022 mendatang.

Kata Kunci: *data mining*, python, klasifikasi, prediksi, pandemi, covid19



ABSTRACT

The outbreak of corona virus disease (covid-19) in the end of 2019 has occurred in Wuhan, Tiongkok. This virus has an ability to spread quickly. Thus, WHO makes this outbreak as a global pandemic. This research conducted based on the data of covid-19 that reported globally accessed from kaggle.com. The C4.5 algorithm is used to classify cases, confirmed, deaths, and recovered as high, and low. This algorithm has successfully classify the cases. We use Orange data mining tools and obtain the AUC 0.871, it indicates the classification is good, the accuracy is 93%, precision is 95%, and recall 96%. Then the classification conducted with Python using standard provided libraries such as matplotlib, seaborn, numpy, pandas, and decisionTree classification. The accuracy with Python is 94%, precision is 98%, and recall is 95%. We also predict the cases trend in Indonesia and predict the end of pandemic using logistic model function, a standard model for population increasing measurement. In this prediction we know that the infection speed rate is 0.0255, pandemic day peak is September 14-18, 2020, it means the case is in a high condition but will reduce time by time, and the end of pandemic is 17 February, 17 2022.

Keyword: data mining, python, classification, prediction, pandemic, covid19

