

ABSTRAK

Covid-19 adalah penyakit menular yang disebabkan oleh jenis virus corona yang baru ditemukan. Pasien positif Covid-19 semakin hari semakin meningkat. Banyak pasien yang mempunyai riwayat penyakit serius diantaranya diabetes, asma, hipertensi dan lain-lain. Tidak sedikit diantara pasien yang positif akhirnya meninggal dunia, namun beberapa diantaranya sembuh meski masih harus melewati beberapa penanganan khusus dari dokter dan perawat. Pada penelitian ini digunakan dataset *patient pre-condition dataset* (data COVID-19 dari pemerintah Mexico) (Mukherjee, 2020) yang didapat dari *website* Kaggle. Dilakukan klasifikasi terhadap pasien yang hidup dan meninggal dengan menggunakan algoritma *Random Forest* untuk mengetahui seberapa akurat algoritma tersebut bekerja pada data *real*. Pada penelitian ini dilakukan *balancing data* terhadap data tidak seimbang menggunakan SMOTE . Percobaan dilakukan dengan pengujian menggunakan *3-fold cross validation*, *5-fold cross validation*, dan *7-fold cross validation* terhadap data yang sudah dikenai proses *balancing* maupun data yang belum dikenai proses *balancing*. Penelitian ini mendapatkan hasil akurasi tertinggi sebesar 82.401% pada data yang belum dikenai proses *balancing*, dengan *7-fold cross validation*.

Kata kunci : Covid-19, klasifikasi, *Random Forest*, SMOTE

ABSTRACT

Covid-19 is an infectious disease caused by a newly discovered type of coronavirus. The number of positive COVID-19 patients is increasing day by day. Many patients have a history of serious illnesses including diabetes, asthma, hypertension and others. Not a few of the positive patients died, but some of them recovered eventhough still had to go through some special treatment from doctors and nurses. In this research, the patient pre-condition dataset (COVID-19 data from the Mexican government) (Mukherjee, 2020) was used which was obtained from the Kaggle website. The researches classified living and dead patients using the random forest algorithm to find out how accurately the algorithm worked on real data . In this study, data balancing was carried out against imbalance data using SMOTE. The experiment was carried out by testing using 3-fold cross validation, 5-fold cross validation, and 7-fold cross validation towards balance and imbalance data as well. This study obtained the highest accuracy result of 82.401% on imbalance data, using 7-fold cross validation.

Keywords : Covid-19, classification, Random Forest, SMOTE

