ABSTRAK

Virus Covid-19 adalah kelompok dari coronavirus yang menyerang sistem pernafasan. Namun virus ini juga dapat menyebabkan infeksi paru-paru sehingga dapat membahayakan pasien apabila tidak ditangani dengan baik. Dalam masa pandemi sejak tahun 2020, jumlah pasien terus meningkat namun jumlah ruang ICU terbatas sehingga dibutuhkan kriteria pasien agar penggunaan ruang ICU dapat optimal. Pada penelitian ini dilakukan klasifikasi data pasien Covid-19 yang membutuhkan ICU menggunakan algoritma *random forest*. Data pasien Covid-19 didapatkan dari *website kaggle* yang terdiri dari 566.602 baris dan memiliki 23 kolom atribut. Penelitian ini sendiri bertujuan untuk mengetahui hasil akurasi dan *f1-score* klasifikasi pasien Covid-19 yang membutuhkan ICU dan mengidentifikasi atribut dari pasien yang dapat digunakan untuk klasifikasi pasien Covid-19. Teknik pengujiannya menggunakan 3,5,7,dan 10 *fold cross validation,* jumlah pohon yang digunakan 10,30,dan 100 pohon. Berdasarkan pengujian yang dilakukan diketahui bahwa performa optimal didapatkan dari model yang dikenai proses *balancing*, tanpa deteksi *outlier*, 100 *tree*, dan 10 *fold* dengan *f1-score training* 85,061% dan akurasi *training* 87.950% sedangkan untuk *f1-score testing* 85,557% dan akurasi *testing* 85,319%. Selain itu didapatkan 16 kriteria atribut yang berpengaruh terhadap penentuan pasien Covid-19 untuk mendapatkan ruang ICU yaitu *intubed, pneumonia, astma, age, cardiovascular, other_disease, renal_chronic, copd, inmsupr, hypertension, tobacco, obesity, diabetes, contact_other_covid, sex, before_treatment.*

Kata Kunci : Covid-19, ICU, *Data Mining***,** *Random Forest***,** *Balancing***,** Deteksi *Outlier*

ABSTRACT

A Covid-19 virus is a group of coronaviruses that attack the respiratory system. However, this virus can also cause lung infections so it can be dangerous if not treated properly. During the pandemic period since 2020, the number of patients has continued to increase, but the number of ICU rooms is limited so the criteria required by patients are to optimize the use of the ICU. In this study, data classification of Covid-19 patients requiring ICU was carried out using the random forest algorithm. The Covid-19 patient data obtained from the Kaggle website consists of 566,602 rows and has 23 attribute columns. This study itself aims to determine the results of accuracy and f1-score classification of Covid-19 patients who require ICU and attributes of patients that can be used for classification of Covid-19 patients. The testing technique uses 3, 5, 7, and 10-fold cross-validation, the number of trees used is 10, 30, and 100 trees. Based on the tests carried out, it is known that the optimal performance is obtained from the model that is subjected to the balancing process, without outlier detection, 100 tree, and 10 fold with f1-score training of 85.061% and 87.950% accuracy training, while for testing f1-score 85.557% and testing accuracy of 85.319%. In addition, there were 16 attribute criteria that influenced the possibility of Covid-19 patients getting to the ICU, namely intubed, pneumonia, asthma, age, cardiovascular, other_disease, renal_chronic, copd, inmsupr, hypertension, tobacco, obesity, diabetes, contact_other_covid, sex, before_treatment.

Keywords: Covid-19, ICU, Data Mining, Random Forest, Balancing, Outlier Detection