

## ABSTRAK

*Active cases* COVID-19 dapat menjadi salah satu penyebab penularan COVID-19, karena *active cases* merupakan sekelompok orang yang dinyatakan positif COVID-19 dan masih dalam masa perawatan. Model regresi dapat diterapkan untuk memprediksi angka *active cases* COVID-19 yang baru (*new active cases*), berdasarkan *feature* yang berkorelasi dengan *label*. Tujuan dari penelitian ini adalah untuk mengetahui model regresi mana yang lebih baik dalam memprediksi angka *new active cases*. Pembangunan model regresi pada penelitian ini menggunakan algoritma *Multiple Linear Regression* dan *Neural Network*, dengan menambahkan proses optimasi *hyperparameter* pada setiap algoritma. Pemilihan *feature* dilakukan berdasarkan pada seberapa besar nilai korelasi antar *feature* dan *label*. Pengujian dilakukan dengan melihat hasil akurasi yang diperoleh dari model *Multiple Linear Regression* dan *Neural Network* yang sudah di-*hyperparameter tuning* dengan data yang displit secara *linear* dan *shuffled*. Berdasarkan hasil pengujian yang dilakukan, diperoleh bahwa penggunaan metode *shuffled sampling* pada *split* data memberikan penurunan nilai *error* sebesar >50% dengan model yang memiliki akurasi yang lebih baik, yaitu model yang dibangun menggunakan algoritma *Neural Network* yang menghasilkan nilai  $RMSE = 925.452$ ,  $Absolute Error = 669.729$  dan  $R^2 = 0.866$ . Nilai ini menunjukkan bahwa *error* yang dihasilkan oleh model ini masih cukup besar, namun nilai determinasi korelasinya dianggap dapat menjelaskan hubungan antara variabel bebas dan *label* yang kuat, serta model dapat mengikuti pola *trend* dari data aktualnya.

Kata kunci : *COVID-19*, Model Regresi, *New active cases*, *Neural Network*, *Multiple Linear Regresi*, *Hyperparameter Tuning*

## ABSTRACT

*Active cases of COVID-19 can be one of the causes of COVID-19 transmission because active cases are a group of people who have tested positive for COVID-19 and are still being treated. Regression models can be applied to predict the number of new active cases of COVID-19 (new active cases), based on features that are correlated with labels. This study aims to find out which regression model is better at predicting the number of new active cases. The development of the regression model in this study uses Multiple Linear Regression and Neural Network algorithms, by adding a parameter optimization process to each algorithm. Feature selection is based on how big the correlation value between features and labels is. The test is carried out by looking at the accuracy results obtained from the Multiple Linear Regression and Neural Network models that have been hyperparameters tuned with linearly split and shuffled data. Based on the results of the tests carried out, it was found that the use of the shuffled sampling method on split data gave a decrease in the error value of >50% with a model that had better accuracy, namely a model built using the Neural Network algorithm which resulted in an RMSE value of 925,452, Absolute Error = 669,729 and  $R^2 = 0.866$ . This value indicates that the error generated by this model is still quite large, but the correlation determination value is considered to be able to explain the relationship between independent variables and strong labels, and the model can follow the trend pattern of the actual data.*

**Kata kunci :** *COVID-19, Regression Model, New active cases, Neural Network, Multiple Linear Regresi, Hyperparameter Tuning*