

## ABSTRAK

Pada akhir tahun 2019 di Kota Wuhan, China diketahui wabah penyakit baru yang disebabkan oleh virus korona (2019-nCov) atau Covid-19. Melalui laman indonesia.go.id pada tanggal 2 Maret 2020 lalu diumumkan bahwa pertama kali kasus Covid-19 masuk ke Indonesia menjangkit 2 orang Warga Negara Indonesia asal Depok, Jawa Barat. Berawal dari kasus tersebut, setiap harinya jumlah kasus masyarakat Indonesia yang terjangkit Covid-19 ini terus bertambah. Untuk menanggulangi hal tersebut pemerintah memerlukan sebuah prediksi untuk melihat kasus dimasa mendatang. Menggunakan dataset covid-19 di Indonesia dari Kaggle dari Januari 2020-September 2021, penelitian ini melakukan pengelompokkan data dengan K-Means Clustering kemudian melakukan prediksi terhadap data dengan Metode Naive Bayes. Naive Bayes bisa memperkirakan probabilitas di masa depan berdasarkan pengalaman masa lalu serta dengan data yang relatif kecil dapat menghasilkan estimasi parameter yang baik. Parameter yang digunakan yaitu kasus, sembuh, kematian dan kepadatan penduduk. Serta acuan untuk mengukur level kasus berdasarkan referensi WHO serta data dari risiko Covid-19 di daerah tersebut. Selanjutnya dari data yang ada didapatkan nilai yang diproses untuk mengategorikan tingkat level kasus Covid-19 sesuai data yang diinputkan masuk dalam level Very Low, Low, Moderate, High, dan Very High. Didapatkan kesimpulan bahwa dilihat dari data hasil dari clustering cluster 0 Low, cluster 1 termasuk level Moderate, cluster 2 termasuk level High, cluster 3 termasuk level Very Low dan cluster 4 termasuk level Very High. Kombinasi metode K-Means Clustering untuk melakukan pengelompokkan data dengan metode Naive Bayes Classifier untuk melakukan klasifikasi data agar dapat melakukan prediksi level kasus Covid-19 di Indonesia mendapatkan hasil akurasi sebesar 90% dengan pengujian sistem menggunakan Confusion Matrix. Beberapa hal yang dapat diberikan untuk pengembangan sistem, seperti menambahkan parameter lain atau menerapkan kombinasi metode lain untuk mencari akurasi yang lebih baik lagi.

**Kata kunci:** *Covid-19, K-Means Clustering, Naive Bayes Classifier, Prediksi, Confusion Matrix*

## **ABSTRACT**

*At the end of 2019 in Wuhan City, China, a new disease outbreak caused by the corona virus (2019-nCov) or Covid-19 was discovered. Through the indonesia.go.id page on March 2, 2020, it was announced that the first Covid-19 case entered Indonesia infecting 2 Indonesian citizens from Depok, West Java. Starting from this case, every day the number of cases of Indonesian people infected with Covid-19 continues to grow. To overcome this, the government needs a prediction to see future cases. Using the covid-19 dataset in Indonesia from Kaggle from January 2020-September 2021, this study performed data clustering with K-Means Clustering and then predicted the data with the Naive Bayes Method. Naïve Bayes can estimate future probabilities based on past experience and with relatively small data can produce good parameter estimates. The parameters used are cases, recoveries, deaths and population density. As well as a reference for measuring case levels based on WHO references and data from the risk of Covid-19 in the area. Furthermore, from the existing data, a value is obtained which is processed to categorize the level of the Covid-19 case level according to the data inputted into the Very Low, Low, Moderate, High, and Very High levels. It was concluded that judging from the data results from clustering cluster 0 Low, cluster 1 including the Moderate level, cluster 2 including the High level, cluster 3 including the Very Low level and cluster 4 including the Very High level. The combination of the K-Means Clustering method for clustering data with the Naïve Bayes Classifier method for classifying data in order to predict the level of Covid-19 cases in Indonesia obtained an accuracy result of 90% with system testing using Confusion Matrix. Some things that can be given for system development, such as adding other parameters or applying other method combinations to find even better accuracy.*

**Keywords:** Covid-19, K-Means Clustering, Naïve Bayes Classifier, Prediction, Confusion Matrix