

## ABSTRAK

Di awal tahun 2020 ini, dunia dikagetkan dengan kejadian infeksi berat dengan penyebab yang belum diketahui, yang berawal dari laporan dari Cina kepada World Health Organization (WHO) yaitu COVID-19. COVID-19 pertama dilaporkan di Indonesia pada tanggal 2 Maret 2020 sejumlah dua kasus. Data 31 Maret 2020 menunjukkan kasus yang terkonfirmasi berjumlah 1528 kasus dan 136 kasus kematian. Tingkat mortalitas COVID-19 di Indonesia sebesar 8,9%, angka ini merupakan yang tertinggi di Asia Tenggara dan sampai saat ini pandemi COVID-19 masih belum terselesaikan. Oleh karena itu, diperlukanlah sebuah model untuk mempermudah pemerintah dan masyarakat dalam mencari informasi tingkat penyebaran COVID-19 di setiap wilayahnya. Model tersebut berupa sebuah pengelompokan tingkat penyebaran COVID-19 di Indonesia. Sehingga dengan pengelompokan tingkat penyebaran ini bisa diketahui daerah yang banyak terdapat kasus ataupun sebaliknya.

Pada penelitian ini akan mengimplementasikan algoritma *K-Medoids* atau *Partitioning Around Medoids (PAM)* untuk mengelompokkan tingkat penyebaran COVID-19 di Indonesia. Metode tersebut merupakan metode *partisional clustering* dimana bertujuan untuk menemukan satu set *k-cluster* di antara data yang paling mencirikan objek dalam kumpulan suatu data.

Pada penelitian ini, data yang digunakan berjumlah 7797 *record* dan 2835 *record* untuk prediksi. Data tersebut belum memiliki label serta memiliki 37 atribut. Dalam penelitian ini menggunakan berbagai variasi pemodelan kemudian akan dibandingkan hasilnya, serta akan dilakukan prediksi menggunakan hasil pemodelan *medoids* dari hasil *clustering* sebelumnya. Variasi tersebut antara lain, jumlah *cluster* 3 dan 4, jumlah atribut dengan PCA, menggunakan 3 jenis *distance measure* serta 2 jenis normalisasi di dalam *preprocessing*-nya. Hasil dari penelitian menyatakan jika penggunaan variasi jumlah *cluster*, penggunaan PCA, penggunaan variasi *distance measure* dan variasi normalisasi cukup berpengaruh.

**Kata kunci:** Pengelompokan, *K-Medoids* atau *Partitioning Around Medoids (PAM)*, COVID-19

## ABSTRACT

*At the beginning of 2020, the world was surprised by the incidence of severe infections with unknown causes, which started from a report from China to the World Health Organization (WHO) namely COVID-19. COVID-19 was first reported in Indonesia on March 2, 2020 in two cases. Data from March 31, 2020 shows 1528 confirmed cases and 136 deaths. The mortality rate of COVID-19 in Indonesia is 8.9%, this figure is the highest in Southeast Asia and to date the COVID-19 pandemic remains unresolved. Therefore, a model is needed to make it easier for the government and the public to find information on the extent of the spread of COVID-19 in each region. The model is a grouping of the spread of COVID-19 in Indonesia. So that by grouping the level of this spread can be known areas that can be many cases or not.*

*In this study will implement the algorithm K-Medoids or Partitioning Around Medoids (PAM) to group the rate of spread of COVID-19 in Indonesia. This method is a partitional clustering method that aims to find a set of k-clusters among the data that most characterize objects in a data set.*

*In this study, the data used amounted to 7797 records and 2835 records for predictions. The data doesn't have a label yet and has 37 attributes. In this study use various variations of modeling will then be compared the results, as well as will be done predictions using the results of modeling medoids from the previous clustering results. The variations include, the number of clusters 3 and 4, the number of attributes, using 3 types of distance measure and 2 types of normalization in its preprocessing. The results of the study stated that the use of variations in the number of clusters, the use of PCA, the use of distance measure variations and variations in normalization had an effect.*

**Keywords:** *Grouping, K-Medoids or Partitioning Around Medoids (PAM), COVID-19*