

ABSTRACT

USING K-MEANS AND K-MEANS++ METHODS GROUP DATA AS COVID-19 ON THE ISLAND OF JAVA

Oleh
Nursatio Nugroho
18102208

Corona virus (covid-19) is an infectious disease between animals and humans. At the end of December 2019, the virus was identified in Wuhan Province, China. The spread data presented by covid19.go.id is only in the form of aggregated data for each province but there is no information on the distribution of district/city covid cases. District and city covid data information is very much needed to find out the cluster of covid cases. This study aims to cluster data on the spread of COVID-19 in every district on the island of Java so as to produce zone clusters that must be implemented by PPKM based on positive cases, the first dose of vaccine, and the second dose. vaccine. This study will use the K-Means and K-Means++ algorithms to determine the level of spread of COVID-19 on the island of Java. Based on the number of positive cases, the first vaccine, and the second vaccine, the cases were categorized. After grouping and getting clusters in each group, each cluster will be evaluated for its quality using the silhouette coefficient to choose the best. The results of the study are expected to reveal the extent of the spread of the Covid-19 virus in every district/city on the island of Java, as well as the cluster with the highest Silhouette Coefficient score. For the results of cluster testing using the Silhouette Coefficient, the K-Means method K=3 produces 0.825, K=4 produces 0.873, K=5 produces 0.862, and K=6 produces 0.841, for the K-Means++ method, k=3 produces 0.822, K =4 produces 0.865, K = 5 produces 0.882, and K=6 produces 0.858. The results showed that West Jakarta City, South Jakarta City, North Jakarta City, and Central Jakarta City are areas that are very prone to COVID-19 cases. Based on the test results using the Silhouette Coefficient, the K-Means method is better for cluster formation with a lower k value, while K-Means++ is superior for higher cluster formation.

Keyword: K-Means, K-Means++, Clustering, Covid-19, Silhouette Coefficient