ABSTRACT

Health issues remain a major challenge in Indonesia, particularly because infectious diseases continue to be an unresolved problem. Low public awareness of the importance of health exacerbates this condition. If left uncontrolled, infectious diseases can increase the risk of extraordinary events (KLB). Therefore, infectious diseases must be addressed immediately to prevent a broader impact. One approach to understanding the spread of infectious diseases is by clustering provinces in Indonesia based on their infectious disease characteristics. This study aims to classify provinces in Indonesia using the Cluster Ensemble method, incorporating Single Linkage, Complete Linkage, Average Linkage, Centroid, Ward, K-Means, and Fuzzy C-Means. The variables used in clustering include Tuberculosis, HIV, AIDS, Leprosy, Measles, Dengue Fever (DHF), and Malaria. The core principle of Cluster Ensemble is to combine the results of various clustering methods to achieve more stable and accurate groupings. In this study, the merging process is conducted by constructing a Connected Triple-based Similarity (CTS) matrix, which considers both the relationships between data points and between clusters within the ensemble. The results indicate that provinces in Indonesia are divided into two groups. The first group consists of 33 provinces with relatively lower levels of infectious diseases, while the second group comprises 5 provinces with higher levels of infectious diseases. By applying the cluster ensemble approach, the clustering results become more stable and are not influenced by the limitations of any single clustering method.

Keywords: Infectious Disease, Cluster Analysis, Cluster Ensemble