

ABSTRACT

CLUSTER ANALYSIS OF WARD AND K-MEANS METHODS TO GROUP REGENCY-CITY IN WEST JAVA PROVINCE BASED ON PARTICIPATION OF WOMEN IN DEVELOPMENT

By

Rahma Nurkholiz

Cluster analysis is one of the multivariate statistical methods used to group objects based on their similar characteristics without prior labeling. In this study, cluster analysis is used to group regencies-cities in West Java Province based on women's participation in development in 2021-2023 using the Ward and K-Means cluster methods. To overcome the problem of multicollinearity between variables, data transformation is carried out using Principal Component Analysis (PCA) in order that the resulting new variables are independent to each other. The principal component scores were used for clustering the regencies-cities data using the two cluster methods, and the results were then evaluated using the Davies-Bouldin Index (DBI) and Calinski-Harabasz Index (CHI) to determine the optimal number of clusters. The results show that both the Ward and K-Means methods produce the same number of optimal clusters, which are four clusters for data 2021, three clusters for data 2022 and 2023, and four clusters for the average of data 2021-2023.

Keyword: Cluster Analysis, Ward, K-Means, PCA, Davies-Bouldin Index, Calinski-Harabasz Index, Women's Participation.