

ABSTRACT

COMPARISON OF DISTANCE CALCULATION METHODS IN K-MEANS FOR VILLAGES CLUSTERING IN BANYUMAS REGENCY BASED ON SOSIO-ECONOMIC CONDITIONS

By

Safina Nanda Kinanthi Asrie

21110028

The differences in the socioeconomic conditions of villages in Banyumas Regency highlight the need for clustering as a basis for more effective and targeted development planning. This study clusters villages in Banyumas Regency based on their socioeconomic conditions using the K-Means algorithm and the 2018 Village Potential data published by the Central Bureau of Statistics (BPS). Three distance measurement methods, namely Euclidean Distance, Manhattan Distance, and Chebyshev Distance, are compared to analyze the impact of these methods on clustering results. Based on the findings, the Euclidean distance demonstrated the best performance with a stable Silhouette Score of 0.678 and a DBI of 0.510. Manhattan achieved a highest score of 0.570 but exhibited varying results and sensitivity to initial centroids. Conversely, Chebyshev showed the lowest performance with a maximum score of 0.510. This study recommends Euclidean as the best method, achieving a silhouette score of 0.679 in experiments using centroids derived from median values.

Keywords: *Villages Clustering, K-Means, Euclidean Distance, Chebyshev Distance, Manhattan Distance, Socio-economic*