

ABSTRACT

The Village Potential Survey (Podes) conducted by BPS is used to support regional development, but often faces the problem of non-response, which causes the data to be biased. To overcome this, villages are clustered using K-Means, although random centroid initialization may be suboptimal. Pillar algorithm is introduced to improve centroid selection. This study evaluates the performance of K-Means + Pillar algorithm combination using Podes data from BPS Banyumas, by evaluating SSE, Variance Within Cluster, Silhouette Score, and Variance Between Clusters. Lower SSE and Variance Within Cluster indicate more homogeneous clusters, while higher Silhouette Score and Variance Between Cluster indicate clearer separation. Results show K-Means excels in cluster homogeneity, achieving a lower SSE of 4,328,194.60 and a smaller Variance Within Cluster of 208.15. In contrast, K-Means + Pillar Algorithm formed more separated clusters, evidenced by a higher Variance Between Clusters of 217.54 and a larger Silhouette Score of 0.02. In conclusion, K-Means is optimal for centralized and homogeneous clusters, while K-Means + Pillar Algorithm is superior in distinguishing cluster boundaries.

Keywords: *K-Means, Pillar Algorithm, Silhouette Score, Sum of Square Error, Variance Between Cluster, Variance Within Cluster*