

Clustering Regions in Central Java Based on the Human Development Index Using the K-Means Algorithm

Dian Rahmawati Iman Sari

Data Science Study Program, Faculty of Science and Technology

University of Technology Yogyakarta

Jl. Ringroad Utara Jombor Sleman Yogyakarta

E-mail: dianrahmaa17@gmail.com

Abstract

The Human Development Index (HDI) is a key indicator used to assess the success of development efforts across different regions. However, disparities in HDI achievement between areas present a significant challenge that requires deeper analysis. This study aims to cluster regencies and cities in Central Java Province based on similarities in HDI characteristics using the K-Means clustering algorithm. The dataset utilized includes 2024 HDI data and supporting variables such as Life Expectancy (LE), Mean Years of Schooling (MYS), Expected Years of Schooling (EYS), and Per Capita Expenditure. The research process follows standard Data Mining methodology: business understanding, data understanding, data preparation, K-Means modelling, model evaluation, and deployment. To determine the optimal number of clusters, two evaluation methods were employed: the Elbow Method and Silhouette Coefficient. The Elbow Method suggested three optimal clusters, while the Silhouette Method indicated two clusters with a Silhouette Score of 0.55 (interpreted as good). Upon evaluating actual regional data, the three-cluster model was deemed more representative, offering segmentation that better reflects HDI conditions in Central Java. The Elbow-based clustering resulted in: Cluster 0 (medium HDI) comprising 15 regions, Cluster 1 (high HDI) with 4 regions, and Cluster 2 (low HDI) consisting of 16 regions.

Keywords: Human Development Index, Clustering, K-Means, Elbow Method, Silhouette Coefficient.