

ABSTRACT

The Group Investigation (GI) model is a widely used cooperative learning strategy in higher education. However, challenges such as large class sizes and diverse student profiles make manual group formation difficult. While previous studies have employed clustering algorithms like K-Means, the use of K-Medoids—which is more robust to noise—remains underexplored, particularly in the context of GI. This study proposes a data-driven approach using the K-Medoids algorithm to form student groups for GI-based learning. Adopting the KDD framework, the process includes data selection, preprocessing, transformation, clustering, and evaluation. The group formation process consists of three main stages: (1) students are grouped based on their learning interests, (2) each interest-based subset is clustered using K-Medoids, with the optimal distance metric and number of clusters (k) determined by the highest Silhouette Score, and (3) heterogeneous groups are formed. In the second clustering stage, the use of Manhattan distance combined with PCA produced the highest Silhouette Score, with score 0.555 for Pengukuran Kinerja ($k=3$) dan 0.618 for Pengembangan Data Warehouse ($k=4$). The evaluation results show that the proposed method produced groups with higher heterogeneity compared to the baseline method, Cluster-Based Assembly, namely 0.607 vs. 0.555 for the Pengukuran Kinerja subset, and 0.622 vs. 0.543 for the Pengembangan Data Warehouse subset. Hence, the proposed method successfully formed internally diverse yet interest-cohesive groups, supporting the objectives of GI-based learning.

Keywords: Cooperative Learning, Group Investigation, K-Medoids, Student Group Formation