An Analysis of the Distribution of Educational Resources in Sri Lanka Using Data Mining Techniques

A. E. Madurapperumage^{1(*)}, J. B. Ekanayake¹

¹Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka

^(*)<u>Email</u>:anuradha@appsc.sab.ac.lk

The Sri Lankan government invests 2.8% of GDP for the education sector. Therefore, it is vital to analyze the dispersion of educational facilities of the country on the district basis, as the administration of the education sector follows the same basis. The objective of this research is to cluster the districts based on the existing educational resources. The data for the study was gathered from the annual report issued by the Census and Statistical Department established in the Ministry of Education. The dataset consists of 21 features, which describe the properties of schools, teachers and students. The gathered data were clustered using the K-means clustering algorithm by assigning random initial seeds. The resulted clusters are validated using the silhouette coefficient. Clustering and data visualization were conducted using the orange data mining toolkit. The reordered silhouette values were decreasing, respectively, with 5, 4, 3, and 6 clusters. The clustering resulted in five clusters of hierarchical clustering algorithm with the method of average linkage being very similar to the clusters constructed from the K- means algorithm. This indicates that educational resources in Sri Lanka can be accurately mapped into five clusters, which express the diversity of dispersion of educational facilities among districts. As this disparity existing in distributing the educational facilities, it is worthwhile to select students for inter-zonal competitions based on the uniformity of educational facilities. This study emphasizes the importance of the consideration of the distribution of educational resources in Sri Lanka while revealing the existing diversity where the facilities are idling and overused among districts.

Keywords: Educational data mining, hierarchical clustering, K-means clustering, orange data mining tool, silhouette coefficient

75

06 No⁴