3

Title:
Clustering Analysis for Classifying Student Academic Performance in Higher Education

Document Type: Article

Authors:
Ahmad Fikri Mohamed Nafuri,
Nor Samsiah Sani,
Nur Fatin Aqilah Zainudin,
Abdul Hadi Abd Rahman,
Mohd Aliff, mohdaliff@unikl.edu.my

Abstract:
There are three income categories for Malaysians: the top 20% (T20), the middle 40% (M40), and the bottom 40% (B40). The government has extended B40′s access to higher education to eliminate socioeconomic disparities and improve their lives. The number of students enrolled in bachelor's degree programmes at universities has risen annually. However, not all students who enrolled graduated. Machine learning approaches have been widely used and improved in education. However, research studies related to unsupervised learning in education are generally lacking. Therefore, this study proposes a clustering-based approach for classifying B40 students based on their performance in higher education institutions (HEIs). This study developed three unsupervised models (k-means, BIRCH, and DBSCAN) based on the data of B40 students. Several data pre-processing tasks and feature selection have been conducted on the raw dataset to ensure the quality of the training data. Each model is optimized using different tuning parameters. The observational results have shown that the optimized k-means on Model B (KMoB) achieved the highest performance among all the models. KMoB produced five clusters of B40 students based on their performance. With KMoB, this study may assist the government in reducing HEI drop-out rates, increasing graduation rates, and eventually boosting students' socioeconomic status.