

ABSTRACT

The use of digital data now saves a lot of information, but still raw and not in the form of knowledge that can be directly seen, therefore we need a data processing to get useful particular understanding from raw data. Data Mining has an important role to process and find useful information from data. This process is also called Knowledge Discovery. Educational Data Mining is a Knowledge Discovery process in the world of education using data mining techniques. One method used in data mining is clustering. By using clustering analysis techniques, data can be grouped into groups without the need for prior, so that the data is grouped based on similarity of patterns and data characteristics.

This study compares the K-means, Hierarchical and Louvain clustering methods to see the most appropriate clustering technique in analyzing log activity data in Moodle for eLisa Learning Management System (LMS). Group of student activity from quiz and task score are create from data log. Data transformation process are create from Learning Management System log become student score based on quiz and task from student activity. Clustering process result are used for Machine Learning (ML) process model to create groups from student score Learning Management System log.

The results of clustering are measured using the Silhouette Coefficient, and then compare the values and distribution between clusters. K-means clustering produces the highest Silhouette Coefficient value with 0,541. Hierarchical clustering produces the biggest number of cluster compare to others, and also this algorithm can detect outlier data as new cluster. Louvain clustering perform very well to find cluster groups in new dataset and the highest SC is 0,246, also the algorithm does not required the number of clusters to be specified before. Louvain clustering can used for ML model because this algorithm divides more evenly compare to K-means and Hierarchical clustering algorithm.

Keywords: Educational Data Mining, Silhouette Coefficient, Log Data, Louvain, Clustering

INTISARI

Penggunaan data digital menyimpan banyak informasi, tetapi masih mentah dan tidak dalam bentuk pengetahuan yang dapat dilihat secara langsung, sehingga dibutuhkan pemrosesan data untuk mendapatkan pemahaman khusus yang berguna dari data mentah. *Data mining* memiliki peran penting untuk memproses dan menemukan informasi dari data, proses ini juga disebut *Knowledge Discovery*. *Educational Data Mining* adalah proses *data mining* pada bidang pendidikan. Salah satu metode yang digunakan dalam *data mining* adalah *clustering*. Dengan *clustering*, data dikelompokkan tanpa perlu pengetahuan sebelumnya, sehingga data tersebut dikelompokkan berdasarkan kesamaan pola dan karakteristik dari data.

Penelitian ini membandingkan metode *K-means*, *Hierarchical* dan *Louvain* untuk melihat teknik yang tepat dalam menganalisis data aktivitas log Sistem Manajemen Pembelajaran Moodle (LMS) pada *eLearning* eLisa. Dari data log tersebut dibuat pengelompokan berdasarkan nilai yang diambil dari aktivitas mahasiswa dalam mengerjakan kuis dan tugas. Data log *Learning Management System* ditransformasikan menjadi data nilai mahasiswa berdasarkan aktivitas quiz dan tugas yang sudah dikerjakan. Dari pengelompokan data tersebut hasilnya digunakan untuk proses model *Machine Learning* (ML) pada kelompok nilai mahasiswa di dalam *log Learning Management System* (LMS).

Hasil pengelompokan dievaluasi menggunakan *Silhouette Coefficient*, kemudian dibandingkan nilai dan distribusi klaster. *K-means clustering* menghasilkan nilai *Silhouette Coefficient* (SC) tertinggi yaitu 0,541, menghasilkan jumlah klaster yang paling sedikit yaitu dua. *Hierarchical clustering* menghasilkan jumlah klaster terbanyak dan dapat mendeteksi data *outlier* sebagai klaster baru. *Louvain clustering* sangat baik untuk menemukan kelompok klaster dalam *dataset* baru dengan SC tertinggi 0,246 dan algoritme ini tidak memerlukan jumlah klaster yang ditentukan sebelumnya. *Louvain clustering* dapat digunakan untuk model ML karena membagi kelompok nilai secara merata dibandingkan dengan *K-means* dan *Hierarchical clustering*.

Kata Kunci: *Educational Data Mining, Silhouette Coefficient, Log Data, Louvain, Clustering*