Application of K-Means Algorithm in Clustering Model for Learning Management System Usage Evaluation

Muhammad Sholeh*a, Suraya*b, Dina Andayati*c

aDepartment of Informatic, Institut Sains & Teknologi AKPRIND, Yogyakarta, Indonesia
bDepartment of Computer Systems Engineering, Institut Sains & Teknologi AKPRIND, Yogyakarta, Indonesia
cDepartment of Digital Business, Institut Sains & Teknologi AKPRIND, Yogyakarta, Indonesia

Abstract
The use of a learning management system (LMS) is one of the media that can be used to disseminate lecturer materials to students. Materials that can be uploaded on the LMS can be in the form of lecture materials in the form of files, videos, or questions. The effectiveness of LMS can be evaluated by looking at activities in using LMS. The effectiveness of using LMS can be seen from the log. Log results from LMS can be evaluated in various ways and one way is to use data mining clustering models. The clustering model can be used to create student groupings and the clustering results can be labeled in the form of categories, such as very good, good, and bad categories. This labeling depends on the clustering results that will be processed in the modeling. The research method uses CRISP DM which consists of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The beginning of the research process is carried out by taking log data in the Moodle LMS. The clustering model in this research will use the K-Means algorithm and the evaluation of clustering results will be evaluated for performance using the Davies-Bouldin method. Implementation of data mining processing using Rapid Miner application. The datasheet used is a datasheet taken from the LMS log of the Computer Programming course in the Mechanical Engineering study program - AKPRIND Institute of Science & Technology Yogyakarta odd semester of the 2021/2022 and 2022/2023 academic years. The results of the study resulted in the best clustering based on the Davies Bouldin method of 2. The clustering results, cluster 0 consists of 28 data named the category of frequent access to LMS and cluster 1 consists of 54 with the category of not frequent access to LMS.

Keywords: Clustering, LMS, Computer Programming, Labeling

DOI: https://doi.org/10.35145/jabt.v4i3.130

1.0 INTRODUCTION

The use of a Learning Management System (LMS) is one of the media used in the teaching and learning process. LMS can be used as a medium of interaction between lecturers and students in supporting the learning process. LMS is a web-based application that can be used in the management, documentation, monitoring, reporting, administration, and distribution of educational content (Rusli et al., 2020). Activities in the LMS are strongly influenced by the role of lecturers, especially creating materials, giving assignments, giving feedback, and other activities. Although influenced by the lecturer's activity, the amount of material in the form of video content uploaded by lecturers cannot be a measure of the success of using the LMS. Another evaluation is the extent to which students are active in using the LMS. Student activeness can be monitored from the log on the LMS. This LMS log can be evaluated to conclude the effectiveness of a course using LMS (Cenka et al., 2022),(Okike & Mogorosi, 2020) (Maloney et al., 2022), [4]. (Maloney et al., 2022),(Agusriandi et al., 2022).

LMS is a learning management system that allows teachers and learners to access learning materials and activities online. LMS has many benefits in the teaching and learning process, including facilitating teaching and learning, optimizing time and place, facilitating adaptive and flexible teaching, increasing learner engagement and participation, improving the efficiency and effectiveness of teaching and learning, and providing accurate data and information related to learner performance.
By using LMS, teachers, and learners can utilize technology to improve effectiveness and efficiency in the teaching-learning process. Learners can study materials and complete assignments anywhere and anytime as long as they are connected to the internet, and choose the right time and place to participate in online discussions or forums. Teachers can tailor learning materials to learners' needs and abilities, create a variety of tasks and activities, and provide online feedback to facilitate more effective learning.

The use of LMS can increase learners' engagement and participation in the learning process, by providing interesting and interactive online discussion forums and activities. Teachers can save time and effort in terms of administration and learning organization, and monitor learners' performance online to improve efficiency and effectiveness in the teaching-learning process. In addition, LMS can provide accurate data and information related to learners' performance, which can help teachers and learners in evaluating and improving the quality of learning. Overall, LMS is a very useful learning management system for improving effectiveness and efficiency in the teaching and learning process.

2.0 LITERATURE REVIEW

The use of learning by using e-learning media is starting to use lot and continues to increase (Cenka et al., 2022). The use of LMS is not only to support the learning process in higher education but is also used in primary education. The use of LMS during the Covid 19 Pandemic became one of the solutions to support the learning process carried out online (Devia Kartika & Hezy Kurnia, 2022). (Dhika et al., 2021). Research that explores the use of e-learning was conducted by (Ajiatmojo, 2021), (Wiragnawanan & Ngurah, 2022), (Setiawati, 2021), (Fitriani, 2020). Devia’s research concluded that the use of e-learning can increase the effectiveness, understanding and knowledge, and skills of teachers in using online learning systems. Ajiatmojo’s research [8], explores the advantages of e-learning including lower costs, more flexible place and time, independence in learning for students, increasing competitiveness for teachers and students, and the quality of material with varied sources and material that can be updated every time. Wiragnawanan's research (Wiragnawan & Ngurah, 2022), concluded that the use of e-learning can help teachers in monitoring activities carried out in detail for learners who use LMS. The effectiveness of e-learning usage requires an evaluation process to see the benefits of usage. Evaluation can be done by looking at the weaknesses and obstacles of e-learning usage. Research that evaluates the use of e-learning was conducted (Rauf & Amin, 2021),(Rizal et al., 2022),(Lestariningih et al., 2020),(Oktaria & Hadiwinarto, 2020).

One way to evaluate the use of LMS can be seen from the activeness of students in viewing material uploaded by lecturers or working on questions by uploading answers on the LMS. The LMS evaluation process in terms of student activity can use the data mining process. One of the data mining models used is clustering. The clustering model is a method of grouping data. Clustering is a process for grouping data into several clusters or groups so that data in one cluster has the maximum level of similarity and data between clusters has minimum similarity (Cielen et al., 2016). (Job, 2018).

Clustering research with LMS log data objects was conducted (Nurdiani et al., 2019), the results of the study resulted in 3 types of clusters with categories of students with a lot of activity and getting high scores, students who do moderate activities and get high scores, and categories of students with a small amount of activity and with low scores. Other research was conducted (Alawi & Shaharanee, Izwan Nizal Mohd Jamil, 2023),(Ademi & Loshkovska, 2020). Ademi’s research (Ademi & Loshkovska, 2020), presented a cluster analysis of Moodle data in terms of students’ preferences for various assessment methods.

Research in the field of data mining that uses data on LMS in addition to clustering methods can also use classification models. Classification models are used to make predictions. Researchers who use classification models with LMS log data include (Tamada et al., 2022), (Ramawasmi & SuSanak, 2022), (Dutt & Ismai, 2019), (Evale, 2017),(Ljubobratović, 2019),(Yağcı, 2022). Tamada (Tamada et al., 2022), using a classification model with the Random Forest algorithm, Ramawasmi [22], using the Random Forest algorithm, Naive Bayes, Logistic Regression, and k-Nearest Neighbours, Dutt (Dutt & Ismai, 2019), using the k-Nearest Neighbour (KNN) algorithm, Support Vector Machine (SVM), and Random Forest (RF).

The use of LMS at the AKPRIND Institute of Science & Technology Yogyakarta as one of the media to support the learning process has been used for a long time. The content in the LMS can be in the form of learning materials or test/exam questions. The LMS at IST AKPRIND was very useful during online lectures during the Covid 19 Pandemic. After the Covid-19 Pandemic, the use of LMS remains the media used in the learning process at IST AKPRIND.

Based on the background and literature review, one of the efforts to evaluate the LMS process is by looking at student logs in using the LMS. Logs on the LMS can be used to record activities carried out by students on the online learning platform. Activities that can be done include activity monitoring. Logs can be used to track user activities on the LMS, including login and logout, access to learning materials, interaction with instructors and other students, and completion of assignments. One way for the evaluation process is to cluster students who use the LMS.
3.0 METHODOLOGY

The data mining model used for the LMS evaluation process is the clustering method. Clustering is one of the exploratory data analysis techniques used to identify subgroups in data with similar data points into the same subgroup and other subgroups that have different points. The implementation of the clustering model uses the K-means algorithm and the performance evaluation process uses Davies Bouldin. The flowchart of the research process is presented in Figure 1.

![Flowchart of the research process](image)

The research methodology on data mining uses CRISP-DM (Cross-Industry Standard Process for Data Mining). This methodology consists of six main stages and is designed to assist project teams in understanding, identifying, and solving business problems by collecting and analyzing data. The CRISP-DM stages are Understanding the business, data understanding, data preparation, modeling, evaluation, and deployment.

4.0 RESULTS AND DISCUSSION

Business Understanding

The purpose of creating an LMS log data clustering model for LMS users, especially computer programming courses, is to map students in using the LMS. The results of this clustering are expected to be an evaluation of the effectiveness of LMS in supporting the lecture process.

Data Understanding

Data understanding is a stage to understanding the processed data. The data processed is the LMS user activity log. The LMS activity log is presented in Figure 2.

![LMS log data](image)
Data preparation
Data preparation is done by checking the datasheet of incorrect data or other processes to determine the data needed in the modeling process. Steps taken in the data preparation process include:

1. Determine attributes that are used in processing but not used in clustering modeling calculations. In making clustering models, all attributes must be of numeric type and all attributes will be calculated in determining the distance.

In this process, the username attribute is one of the attributes that is not used in the process of determining the distance but to display the clustering model results the username attribute is used to see the clustering results.

The process for determining the username attribute so that it is not used in the process of calculating the distance in the clustering process can be determined by defining the username attribute as an attribute with the target role as ID. The process of determining the set role with the target role ID for the username attribute is presented in Figure 3.

2. Determining the attributes used
In making a model, attribute selection is one of the most important steps. Not all existing attributes are used in making the model. The results of taking LMS log data are around 227 attributes. Attributes that are not used include attributes of departments, institutions, and attributes that contain the date of access by students in viewing material content. Figure 4, the process of selecting the attributes used.

3. Converting nominal type to Numeric
The main requirement in the K-Means algorithm process is that all attribute types must be numeric. This is because the K-Means algorithm will calculate the distance of each point to a centroid. Figure 5, the process of converting nominal type to numeric type. In this modeling, all attributes will be converted to numeric.

The process of using the Nominal to Numerical operator converts nominal data into 0 and 1. Figure 6, datasheet before the nominal to numerical operator process.
The nominal to numerical operator process changes the Completed information to 0, which means the student viewed the available content and changes the No Completed information to 1, which means the student did not view the learning content. Figure 7, the result of using the nominal to numerical operator.

4. Checking empty data and outlier data
Empty data is very influential in the data mining process. The processed datasheet must be checked for empty data.

Modeling
The clustering model using the K-means algorithm is tested by simulating clustering from 2 to 6. The process of creating a clustering model is shown in Figure 8.
Model Evaluation
Model evaluation is done by testing the performance of each cluster formed. The process to determine a good cluster is done using the Davies-Bouldin method. An example of evaluation results with Davies Bouldin for $K = 5$ is presented in Figure 9.

![Performance Vector](image)

**Figure 9. Davies Bouldin Evaluation Results for K=5**

The results of Figure 9, the Davies Bouldin value of 0.549, this Davies Bouldin value will be compared with the Davies Bouldin result value for values $K = 1$ to $K = 6$. A good Davis Bouldin value is the smallest (Septiani et al., 2022), (Wijaya et al., 2021). The results of the Davies-Bouldin value comparison are in Table 1.

<table>
<thead>
<tr>
<th>$K$</th>
<th>Davies Bouldin</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.37</td>
</tr>
<tr>
<td>3</td>
<td>0.746</td>
</tr>
<tr>
<td>4</td>
<td>0.690</td>
</tr>
<tr>
<td>5</td>
<td>0.549</td>
</tr>
<tr>
<td>6</td>
<td>0.632</td>
</tr>
</tbody>
</table>

Based on Table 1, the smallest Davies Bouldin value is at the value of $k = 2$, so the clustering chosen is cluster as much as 2. The results of students entering cluster 0 are presented in Figure 10.

![Figure 10. Clustering results in cluster 0](image)

The results of clustering are as in Figure 12, username 191031042, 191031050, and so on enter cluster 0, and username 201034046, 191031010, and so on enter cluster 1. An example of student data entering cluster 1 is presented in Figure 11.
Cluster Labelling
After the clustering is formed, the next process is to create a label with the desired category. The process of creating labels can see the visualization results of the clusters formed. Figure 12, is a visualization between the user name and the course description content material.

The visualization results in Figure 12, cluster 0 is a cluster that contains students who often use the content on the LMS and Cluster 1 contains students who are currently accessing LMS content. So that the cluster results are labeled as cluster 0 with the category of frequently accessing LMS and cluster 1 with the category of not frequently accessing LMS.

5.0 CONCLUSION

Conclusion
The use of LMS as a learning tool is now widely used in education. The use of LMS must be evaluated to see the effectiveness of its use. One way to evaluate is by looking at user activity logs in utilizing the LMS. The more users utilize the LMS, the more they are expected to understand the material.

Recommendation
The recommendation of many clusters is 2. The selection of clusters as many as 2, is the result of evaluating the performance of the clustering results using Davies Bouldin. The process of measuring performance evaluation is tried by creating clusters ranging from 2-6 clusters.

The results of the 2 cluster recommendations are then grouping student data that is included in each cluster and giving category names for each cluster, Naming the category of the cluster as cluster 0 category with the name of the category often accessing the LMS. There are 28 students included in cluster 0 and cluster 1 with the name of the category that does not often access the LMS as many as 58.
These results show that the use of LMS still needs to be improved. Efforts to increase student activity in using the LMS can be done by providing motivation and support, creating interesting tasks and activities, providing quick responses, creating interesting discussions, scheduling regular use of the LMS, and making materials that are easily accessible.

References


