Implementation of the K-Means Algorithm on Learning Outcomes and Self-Regulated Learning

Anisatul Farida¹* and Nugroho Arif Sudibyo²

¹,²Teknik Informatika, Universitas Duta Bangsa, Indonesia
Jl. Bhayangkara No 55, Tipes, Surakarta, Indonesia

*Corresponding Author: anisatul_farida@udb.ac.id

ABSTRACT

This study aims to apply data mining to Student Self-Regulated Learning and learning outcomes in online learning in higher education using the k-means algorithm. The cluster analysis method in this research is the K-means method. The research was conducted by taking data from students who took the Statistical course Informatics Engineering, the faculty of Computer Science at Duta Bangsa University Surakarta. The number of student data was 22 data. The results show that the number of clusters according to the k-means algorithm is 3.

Keywords: K-Means Algorithm, learning outcomes, self-regulated learning
INTRODUCTION

One of the important parts in online learning today is Self-Regulated Learning (SRL). Many researchers focus on research on self-regulated learning during the current Covid-19 pandemic. This is because the level of SRL greatly determines self-motivation and self-regulation skills in learning. The students are no exception. With self-regulated learning, it is hoped that it can improve student learning outcomes in online learning in accordance with learning objectives (Kusuma, 2020). This is the reason why a lesson should be designed to help students, including students, to realize their own needs. Learning must also encourage students' enthusiasm for learning to achieve learning objectives. In this case, it is the result of learning. Lecturers must also understand and understand self-regulated learning so that lecturers understand the needs of students so that lecturers in their duties do not only teach to carry out their obligations but more so so that students are able to achieve maximum potential in the learning process.

There are four assumptions regarding self-regulated learning according to Woltres, in (Santosa, 2020). First, the foundation for dynamic and constructive thinking. Students who actively construct understanding, goals, and methods from information contained in their learning environment and from their own minds. Second, self-regulated learning makes students able to control themselves. This means that the characteristics of the student learning environment can be supervised and controlled by the students themselves. Students can arrange certain parts based on consistent cognition, motivation, and conduits. Third, the view of self-regulated learning as a goal, criterion, or provision. That is why self-regulated learning can be used as a tool to assess whether a learning process needs to be continued or not. If one of the criteria or standards changes, adjustments are needed according to the conditions of the learning environment. Fourth, the assumption of self-regulated learning as a mediator between personal characteristics and context and learning outcomes.

Self-regulated learning has three important aspects that will determine the level of self-regulated learning. According to Zimmerman in (Chotimah & Nurmufida, 2020) these three aspects are the first aspect of metacognition. In this aspect, the individual begins to plan, set goals, and evaluate tasks. The second aspect of motivation, namely when individuals have high confidence and enthusiasm in doing a task. The three behavioral aspects are more about individual efforts to choose, structure, and create an environment that optimizes learning. Based on the results of interviews with Informatics Engineering students, Faculty of Computer Science, Universitas Duta Bangsa who took the Statistics
course on self-regulated learning. The results showed that some students had not been able to implement self-regulated learning properly. Some students who have a high level of self-regulated learning are able to make priorities in doing assignments, while for students who are active in organizations, as much as possible they can catch up in lectures by being active in class, asking friends and lecturers.

Other research shows that self-regulated learning is influenced by students' emotions and motivations which in turn affects student learning outcomes. This is in accordance with research (Sihombing, 2020) that in online learning as it is today, self-regulated learning greatly affects student learning motivation online. Therefore, in online learning, it is necessary to have the right learning approach in order to increase self-regulated learning in students to improve learning outcomes. The right learning approach will greatly affect self-regulated learning. Research (Meri, 2020) also emphasizes that self-regulated learning really supports students' mathematical abilities during a pandemic.

Regarding the role of self-regulated learning in online learning, researchers gave different responses regarding learning outcomes (Santosa, 2020). According to (Azizah, 2021) the role of the teacher is to moderate the contribution of self-regulated learning to student satisfaction in distance learning. This is in accordance with the conditions on the ground that in online learning, including Informatics Engineering FIKOM UDB students who take statistics courses, almost 75% also state that the role of lecturers greatly influences learning outcomes. Students stated that independent learning is very necessary and is determined by how the lecturer teaches in online learning.

To bridge self-regulated learning in supporting student learning outcomes in the pandemic era, clustering is needed to determine the grouping of self-regulated learning. Clustering is a method used in data mining which works by finding and grouping data that have similar characteristics between one data and other data that has been obtained. The hallmark of this data mining technique is that it has an unsupervised nature, which means that this technique is applied without the need for training data and without a teacher and does not require a target output (Ma’rifatin, 2020).

One of the clustering methods that have efficient and fast properties that can be used is the k-means method (Nurcahyo et. al., 2020). This method aims to create a cluster of objects based on attributes into k partitions. The way this method works is that at first the cluster to be formed is determined, the first element in each cluster can be selected to be used as the center point (centroid) (Suprihatiningsih & Sudibyo, 2020). Furthermore, the
Implementation of the K-Means Algorithm on Learning Outcomes and Self-Regulated Learning
(Farida, Anisatul. & Sudibyo, N.A.)

steps will be repeated until there are no objects that can be moved anymore (Pribadi et. al., 2020).

Based on the problems described above, this research will apply the k-means method to generate information regarding the grouping of self-regulated learning on the learning outcomes of Informatics Engineering students at Universitas Duta Bangsa Surakarta who are taking Statistics courses in the 2020/2021 academic year. Then these results can be used as material or basis for lecturers and stakeholders to determine the right learning approach and direction of leadership policies.

METHOD

The subjects of this research are students of Universitas Duta Bangsa in the even semester of 2020/2021 who takes statistics courses. In this study, only one class was taken to see student learning outcomes and SRL. Student SRL data are taken from questionnaires that have been tested for validity and reliability. The questionnaire was taken using a Google Form that has been processed in such a way that there are high, medium and low levels.

After obtaining student learning outcomes and SRL data, the next step is to apply the K-Means algorithm. Here are the steps of the algorithm.

1. Start by forming k clusters (groups), where k is the number of clusters (groups) to be formed.
2. Get the initial centroid of the minimum, medium, and maximum values of the dataset.
3. Determine the distance between each data point and its center of mass.
4. Sort the data into clusters depending on the minimum centroid distance.
5. If the distance between the new Centroid and the old Centroid does not change, then the clustering process using the Density k-means algorithm is complete; however, if the distance between the new Centroid and the old Centroid continues to change, repeat the distance calculation process to the unchanged value.
6. Clustering results from the previous iteration

After applying the K-Means algorithm to learning outcomes and SRL, how many clusters are formed.

RESULTS AND DISCUSSION

The results of the questionnaire showed that there were two students with low SRL, eighteen students with moderate SRL and two students with high SRL. The SRL level will
be combined with student learning outcomes. Figure 1 is the Orange software workflow in this study.

![Orange software workflow](image)

**Figure 1. Workflow software Orange**

The determination of the number of clusters in this study was assisted by the Orange software. Orange software uses Silhouette Scores in determining the number of clusters. After the data is entered and processed the results are as shown in Figure 2.

![Silhouette Scores](image)

**Figure 2. Silhouette Scores in determining the number of clusters**

The next step is to know the characteristics in the cluster. Figure 3 is the result of the cluster that has been done with the Orange software.
The first cluster is a score of 70s with a Low and Medium SRL, while the second cluster is a value of 80s with a Medium SRL. Furthermore, the third cluster is a value above 85 with a High SRL. This shows that the K-Means algorithm can be used (Aggarwal & Sharma, 2019) and (Fernandes et al., 2019). This is also in line with research that has been conducted by (Shovon et al., 2012) and (Oyelade et al., 2010). Furthermore, there is an assumption that there is a relationship between student learning outcomes and SRL. To prove it, a Spearman correlation test will be conducted between learning outcomes and SRL. Table 1 is the result of the correlation analysis.

Table 1. The results of the Spearman correlation analysis between learning outcomes and SRL

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Silhouette</th>
<th>Hasil Belajar</th>
<th>SRL</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>0.59987</td>
<td>77.0</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C2</td>
<td>0.663161</td>
<td>80.5</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C3</td>
<td>0.648826</td>
<td>70.0</td>
<td>SRL Rendah</td>
</tr>
<tr>
<td>C4</td>
<td>0.694053</td>
<td>81.5</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C5</td>
<td>0.709707</td>
<td>83.3</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C6</td>
<td>0.696434</td>
<td>85.0</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C7</td>
<td>0.75</td>
<td>86.8</td>
<td>SRL Tinggi</td>
</tr>
<tr>
<td>C8</td>
<td>0.694053</td>
<td>85.0</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C9</td>
<td>0.696434</td>
<td>85.0</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C10</td>
<td>0.75</td>
<td>86.8</td>
<td>SRL Tinggi</td>
</tr>
<tr>
<td>C11</td>
<td>0.708887</td>
<td>83.0</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C12</td>
<td>0.70465</td>
<td>82.3</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C13</td>
<td>0.606725</td>
<td>75.3</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C14</td>
<td>0.663161</td>
<td>80.5</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C15</td>
<td>0.70465</td>
<td>82.3</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C16</td>
<td>0.696434</td>
<td>85.0</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C17</td>
<td>0.628229</td>
<td>86.0</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C18</td>
<td>0.709707</td>
<td>83.3</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C19</td>
<td>0.706592</td>
<td>84.0</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C20</td>
<td>0.75</td>
<td>86.8</td>
<td>SRL Tinggi</td>
</tr>
<tr>
<td>C21</td>
<td>0.696434</td>
<td>85.0</td>
<td>SRL Sedang</td>
</tr>
<tr>
<td>C22</td>
<td>0.638673</td>
<td>74.6</td>
<td>SRL Rendah</td>
</tr>
</tbody>
</table>

Figure 3. The results of the cluster that has been done with the Orange software.
The results show that the number of clusters according to the k-means algorithm obtained is 3. Furthermore, there is an assumption that there is a relationship between student learning outcomes and SRL. This is in line with research (Pratama, 2017) which states that there is a relationship between learning outcomes and SRL. It is hoped that with this research, both teachers and lecturers will further improve SRL in learning.

CONCLUSION

Based on the results of the analysis of the K-Means Algorithm, Self-Regulated Learning (SRL) on learning outcomes there are 3 clusters. The first cluster with medium and low SRL, second cluster with medium SRL, third cluster with high SRL. The results of this study can be used as a reference for further research with the suggestion that clustering with the K-Means algorithm can be used as a reference and consideration for lecturers and university leaders in relation to learning strategies for lecturers and policy directions for PT. The selection of learning strategies chosen by lecturers should also be able to improve students' SRL.

REFERENCES


