Ittishal Educational Research Journal
July 2024, Vol. 5, No. 02, p. 01 - 16
http://journal.ittishal.net/index.php/ierj
https://doi.org/10.51425/ierj.v5i2.84

Clustering of Mathematical Skill Improvement for Children with Special Needs Using K-Means Clsutering: A Case Study at Mutiara Al Islam Therapy Surakarta

Dwi Remawati[1], Kumaratih Sandradewi[2], Tri Irawati[3], Sri Hariyati Fitriasih[4]
[1,2,3,4]STMIK Sinar Nusantara Surakarta, Indonesia
Correspondence: dwirema@sinus.ac.id,kumaratih@sinus.ac.id,3irawati@sinus.ac.id, fitriasih@sinus.ac.id

ABSTRACT

The ability to understand mathematics in children with special needs (CSN) needs to be constantly improved. Because the capabilities of CSNs are not the same, they should be grouped based on the results of the evaluations that have been carried out. The purpose of this study is to group children with special needs who receive therapy at Mutiara Al Islam Therapy based on the progress of their mathematical skills. For grouping, use the K-Means Algorithm because this method can be effective in grouping data. Data was obtained from the Mutiara Al Islam therapy institution in Surakarta. The data is in the form of initial math skill scores and new math skill scores, the next process is calculated as a percentage increase in intentions, then clustered. The result of this study is the formation of 2 clusters, namely low improvment and high improvment, so that it can help therapists to improve CSN math skills based on the results of the cluster. Based on the results of the silhouette validity test, the accuracy level of cluster 1 is 0.575629 and cluster 2 is 0.471283.

***Keywords*:** inclusive education, mathematical skill, CSN, Clustering, K-Means

INTRODUCTION

Children with special needs (CSN) need tailored educational methods to address their different learning challenges (Chairunnisa & Rismita, 2022);(Duque et al., 2020);(Kartono et al., 2023). The choice of this method is crucial for their cognitive growth and overall development (Asya et al., 2020);(Mareza, 2019). Mathematical skills are one of the most basic academic skills (Wafa et al., 2020). This skill is essential for problem-solving, daily tasks, and logical thinking (Kalijaga, 2024);(Kefallinou et al., 2020).

Mutiara Al Islam Surakarta is an out-of-school educational institution in Surakarta or Solo that focuses on the education of children with special needs. It was established on January 1, 2003, under the name Mutiara Center. It is a non-formal education that provides services in the form of academics, communication, speech, motor skills (both fine and gross), group therapy, and social and psychological psychology. In certain groups of children with special needs, they are given additional lessons, namely basic and simple math skills that include addition, subtraction, and division. To make it easier for therapists to provide materials, it is necessary to create groups based on simple tests.

The K-Means algorithm is widely used to partition datasets into different clusters, where each cluster contains data points that are more like each other than those in other clusters (Ahmar et al., 2018). K-Means is a very useful and flexible algorithm for different types of clustering analysis, with speed and efficiency that make it a top choice in many practical applications (Nagari & Inayati, 2020). In this study, the grouping of children was based on their initial math skills, and only then was the percentage increase calculated. This percentage increase is used as data for grouping using K-means.

Previous research used the K-Means algorithm to group outstanding students based on the criteria of assignment scores, test scores, and practice scores (Dwijayanthi Nirmala et al., 2020). Other research also uses the K-Means algorithm to group students' ability levels into three categories, namely beginner, intermediate, and advanced(de la Fuente-Tomas et al., 2019).

This study focuses on the evaluation of improving the mathematical ability of children with special needs at the Mutiara Al Islam therapy institution. By applying the K-Means grouping algorithm based on their progress in math skills

using Phyton. The resulting categories are low improvement and high improvement. With the formation of these groups to improve mathematical skills, this study aims to provide actionable insights to improve educational strategies for children with special needs, especially for therapists to be able to provide appropriate treatment.

METHOD

This study uses the K-means algorithm to create a grouping. The steps of the method in this study can be seen in Figure 1(Prianto et al., 2019).

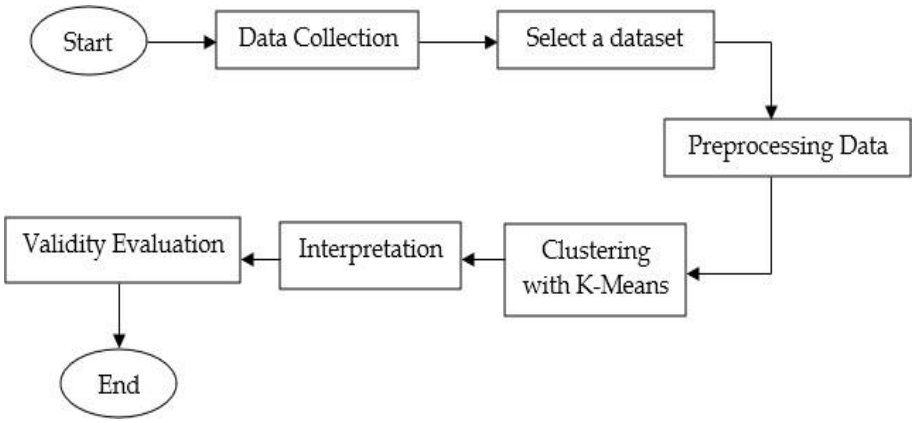


Figure 1 Research Method

1. Data Collection

   Data Collection At this stage, the researcher collected the data needed for the study, namely data from the Mutiara Al Islam Therapy Institute. To collect data using several data collection methods, namely:

   a) Observation

      The observation referred to in this study is to conduct research directly with the aim of better understanding and knowing what steps must be taken in solving the problems found. The steps at the time of observation are as follows:

      1. Preparation Before Observation:

         - Determination of observation objectives: clearly understand what you want to achieve with observation. For example, observe the behavior of children with special needs during assessment.

- Permits: ensuring that all necessary permits have been obtained, both from the authorities at the research site and from the subject to be observed.
- Equipment: prepare necessary tools such as laptops, cameras, or recording devices to record findings during observations.

2. Implementation of Observations:
   - Initial Approach: introduce yourself to the relevant parties and explain the purpose and procedure of the observation.
   - Direct Observation: directly observe the situation and activities that occur at the research site. For example, look at how children with special needs are assessed when they are assessed in basic mathematics.
   - Data Logging: record all relevant findings in detail. This can include field notes, photographs, or video footage.
   - Interaction with Subjects: conduct interviews or discussions with related subjects or parties if necessary. For example, ask teachers or caregivers about their observations of the assessments.
3. Analysis of Findings:
   - Data Compilation: this organizes all the data that has been collected during the observation.
   - Identify Patterns and Trends: analyze data to identify patterns or trends that emerge from observations.
   - Evaluation of Findings: evaluating findings involves referring to the original purpose of the observation to see if there are any conclusions that can be drawn or hypotheses that can be formed.

Problems that may be observed during observation:

1. Access Limitations:
   - It is difficult to get full access to the research venue or subject.
   - Solution: Arrange appointments and communication with the authorities in advance to ensure adequate access.
2. Difficulties in Recording:

- The challenge is accurately recording all important details as the observation progresses.
- Solution: use various note-taking tools, such as audio or video recordings.

b) Interview

It is data collection by coming to the research site to conduct a direct question and answer with the Mutiara Al Islam Therapy Institute to obtain the needed information. We met directly with Mrs. Kurnia Fitrotin, S.Pd., M. Psi. as the head of the Mutiara Al Islam Institute.

c) Literature Research

To support the research to be carried out, the author conducts library research, namely by looking for data from books and similar research journals related to the method used in this study, namely the K-Means Clustering method.

3. Dataset

In this study, a dataset of children with special needs (CSN) was used at the Mutiara Al Islam therapy institution in Surakarta, Central Java.

4. Preprocessing

Data preprocessing is the process of transforming raw data into a form that is easier to understand. This process is important because raw data often does not have a regular format. In addition, data mining cannot process raw data. There are several stages that need to be done when doing data preprocessing. Here are some of the stages:

a) Data cleaning

The first step that needs to be done when preprocessing data is data cleaning. This means that the raw data that has been obtained needs to be re-selected. Then, delete or eliminate incomplete, irrelevant, and inaccurate data.

b) Data integration

Because data preprocessing will combine several pieces of data in a dataset, it must check the data coming from these various sources so that it has the same format. This process is one of the important steps.

c) Data transformation

The next process that must be carried out is data transformation. All collected data must be equalized to facilitate the data analysis process. Data transformation is the process of transforming data into a form suitable for analysis, in this case data that shows the ability of mathematical skills. Here, the high improvement (C1) and low improvement (C2) groupings use mathematical values.

d) Reduce data.

The last step that needs to be taken is to reduce the amount of data (data reduction). The point is to reduce the sample of data taken, but with a note, it will not change the results of the data analysis. At this stage, to get good data, the data selection stage is carried out. Data selection is carried out by changing several pieces of data, which aims to facilitate understanding by referring to data stability, inconsistent, or stacked data. When it is finished, data that has been processed or mature data that is ready to move on to the next stage is obtained.

5. Clustering

In the K-Means Clustering stage, it is grouping existing data into two groups, namely low and high. The definition of clustering or data grouping is a stage to classify datasets whose class attributes have not been described; conceptually, clustering is to maximize and minimize the similarities between classes. A cluster can also be interpreted as a group. So clustering analysis basically produces several clusters. Before conducting the analysis, it is necessary to understand that a set of certain data already has similarities among its members. Therefore, each member with the same characteristics is grouped into one or more groups. The K-Means Clustering algorithm is a data mining analysis method where the modeling process is unattended and how to group data based on partitions. The data is grouped by the K-means method into several groups, and each group has similar or the same characteristics as the others, but the other groups have different characteristics. With the aim of minimizing the difference between each data point in one cluster and maximizing the difference between other clusters. The steps taken to cluster in the K-Means method are:

Specifies the number of clusters (k) in the data set used.

a) Determining the value of k as a centorid is usually done randomly
b) Calculating the distance between the data and the centroid using the Euclidean formula (equation 1).

$$\text{Distance: } (x\ \text{i}, \mu\ \text{j}) = \sqrt{\sum (x\ \text{i}, \mu\ \text{j})^2} \quad \text{(1)}$$

Where:

d = document point

xi= Criteria data

μj= centroid on the j-cluster

c) Group the data based on the value closest to the centroid, then update the new centroid value with the location from the center of the cluster using the Euclidean formula (equation 2).

i. $\mu j\ (t+1) = {}^{1}/Nsj \sum j \in sj\ xj$ ..... (2)

Where:

μJ (t+1) = new centroid in iteration to (t+1)

Nsj = the amount of data in the cluster

d) Perform steps 2 to 4 until each cluster of each member has not changed.

6. Interpretation

   The analysis is the last stage that discusses the results of the analysis of clustering disability data from severe, medium, and light clusters with values.

7. Validity Evaluation

   Validity Evaluation in clustering or classification algorithms refers to the process of assessing how well the model performs. This involves evaluating different aspects of the model to ensure it is accurate, reliable, and generalizable.

FINDINGS/RESULTS AND DISCUSSION

The dataset of children with special needs was obtained from the Mutiara Al Islam Therapy Institute for the 2022-2023 school year using simple mathematics training values that include addition, subtraction, and division. The

value of mathematics skills consists of the initial score before being given the question and the new value after being given a simple math problem. The data obtained is already in the form of a spreadsheet file; it is clean, and there is no need to preprocess the data. The data will be calculated as a percentage of the increase in mathematical skills. The percentage yield will be mined for lateralization. The next stage is clustering mining using the K-Means algorithm to find out if each student enters the cluster with a low improvement or a high improvement.

The following are the steps to complete the classification of math skills for children with special needs using the K-means algorithm:

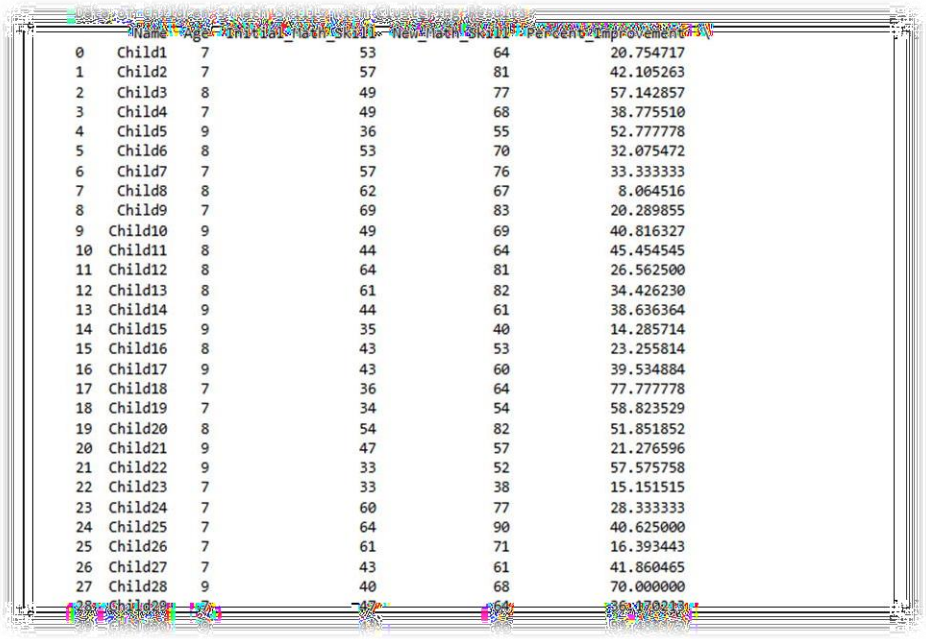
1. Determine the dataset to be in the cluster.

Table 1 Sample Data of Math Skilss Scores

No.	Name	Age	Initial Math Skill	New Math Skill
1.	Child1	7	53	64
2.	Child2	7	57	81
3.	Child3	8	49	77
4.	Child4	7	49	68
5.	Child5	9	36	55
6.	Child6	8	53	70
7.	Child7	7	57	76
8.	Child8	8	62	67
9.	Child9	7	69	83
10.	Child10	9	49	69
11.	Child11	8	44	64
12.	Child12	8	64	81
13.	Child13	8	61	82
14.	Child14	9	44	61
15.	Child15	9	35	40
16.	Child16	8	43	53
17.	Child17	9	43	60
18.	Child18	7	36	64
19.	Child19	8	34	54
20.	Child20	8	54	82
21.	Child21	9	47	57
22.	Child22	9	33	52
23.	Child23	7	33	38
24.	Child24	7	60	77
25.	Child25	7	64	90
26.	Child26	7	61	71
27.	Child27	7	43	61
28.	Child28	9	40	68
29.	Child29	7	47	64
30.	Child30	9	43	66

2. Calculating the percent improvement

In this process, calculate the percentage increase obtained by each student. This percentage value will be clustered in the next step. The results of the calculation of the percentage improvement are as shown in figure 2.

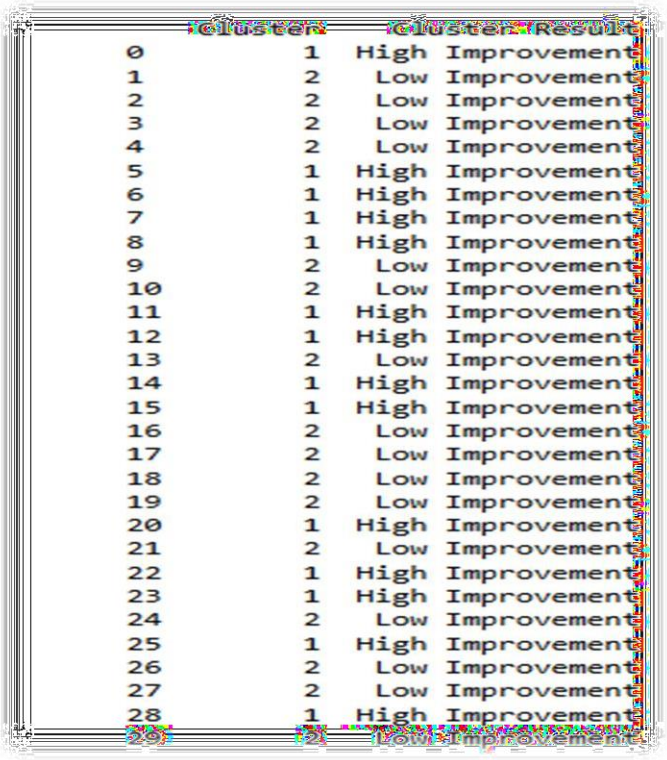


	Name	Age	Initial_Math_Skill	New_Math_Skill	Percent_Improvement
0	Child1	7	53	64	20.754717
1	Child2	7	57	81	42.105263
2	Child3	8	49	77	57.142857
3	Child4	7	49	68	38.775510
4	Child5	9	36	55	52.777778
5	Child6	8	53	70	32.075472
6	Child7	7	57	76	33.333333
7	Child8	8	62	67	8.064516
8	Child9	7	69	83	20.289855
9	Child10	9	49	69	40.816327
10	Child11	8	44	64	45.454545
11	Child12	8	64	81	26.562500
12	Child13	8	61	82	34.426230
13	Child14	9	44	61	38.636364
14	Child15	9	35	40	14.285714
15	Child16	8	43	53	23.255814
16	Child17	9	43	60	39.534884
17	Child18	7	36	64	77.777778
18	Child19	7	34	54	58.823529
19	Child20	8	54	82	51.851852
20	Child21	9	47	57	21.276596
21	Child22	9	33	52	57.575758
22	Child23	7	33	38	15.151515
23	Child24	7	60	77	28.333333
24	Child25	7	64	90	40.625000
25	Child26	7	61	71	16.393443
26	Child27	7	43	61	41.860465
27	Child28	9	40	68	70.000000

Figure 2 percent improvment

3. **Determine the number of clusters.**

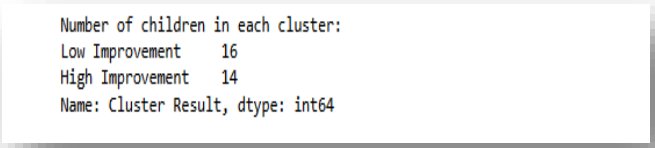
In this process, the value of k, the number of clusters of children with special needs, is determined as many as 2 clusters (k-2). The clusters formed are the high-improvement cluster and the low-improvment cluster. Figure 3 shows the results of clustering using Python.



	Cluster	Cluster Result
0	1	High Improvement
1	2	Low Improvement
2	2	Low Improvement
3	2	Low Improvement
4	2	Low Improvement
5	1	High Improvement
6	1	High Improvement
7	1	High Improvement
8	1	High Improvement
9	2	Low Improvement
10	2	Low Improvement
11	1	High Improvement
12	1	High Improvement
13	2	Low Improvement
14	1	High Improvement
15	1	High Improvement
16	2	Low Improvement
17	2	Low Improvement
18	2	Low Improvement
19	2	Low Improvement
20	1	High Improvement
21	2	Low Improvement
22	1	High Improvement
23	1	High Improvement
24	2	Low Improvement
25	1	High Improvement
26	2	Low Improvement
27	2	Low Improvement
28	1	High Improvement
29	2	Low Improvement

Figure 3 Cluster results

Figure 4 shows the clustering results, which show that 14 students entered Cluster 1 or High improvement, and 16 entered Cluster 2 or Low improvement cluster.



```
Number of children in each cluster:
Low Improvement     16
High Improvement    14
Name: Cluster Result, dtype: int64
```

Figure 4 Number of each cluster

The visualization of the clustering results is shown in Figure 5, showing that the number of clusters of low improvement is more than that of high improvement.

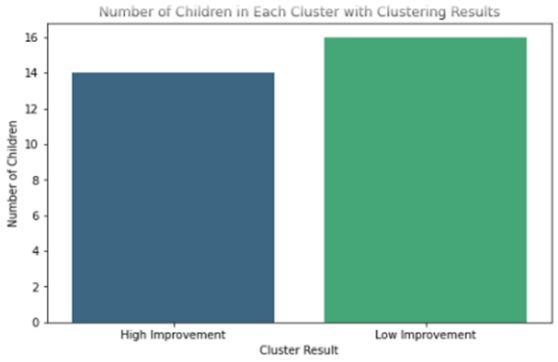


Figure 5 Number of students per cluster

The following is a bar chart that illustrates how each student is grouped based on the percentage of improvement in their math skills. Students with high improvement are grouped in the High Improvement cluster (blue), while students with low improvement are grouped in the Low Improvement cluster (green). In Figure 6, with this visualization, it is easy to see the difference in the level of improvement between the students in the two clusters.

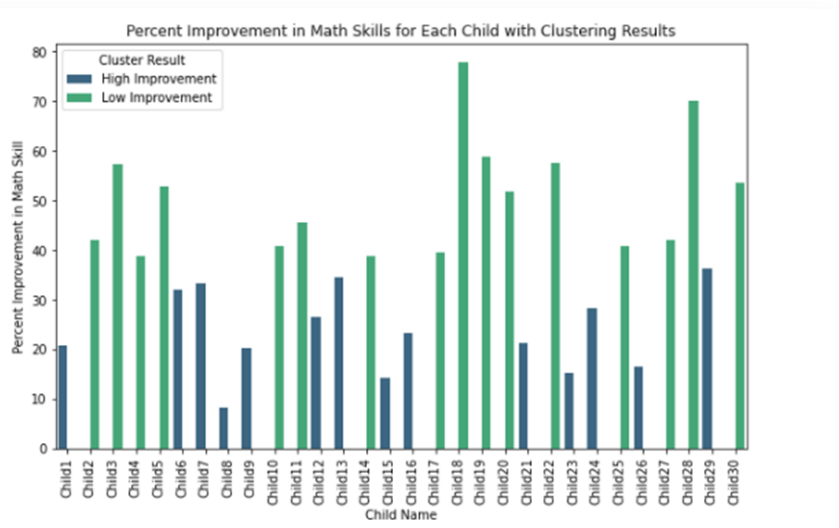


Figure 6 Percent Improvement

4. Visualization with scatter plot

This visualization helps understand how data is distributed and how clusters are formed based on the percentage improvement in math skills. Cluster 1 was identified as a group with a high increase (blue dots), while Cluster 2 was identified as a group with a low rise (red dots). The centroid shows the center of each cluster and helps visualize how the data is distributed within each cluster. The centroid (black X) helps show the midpoint of each cluster, providing an understanding of the average location of each cluster member.

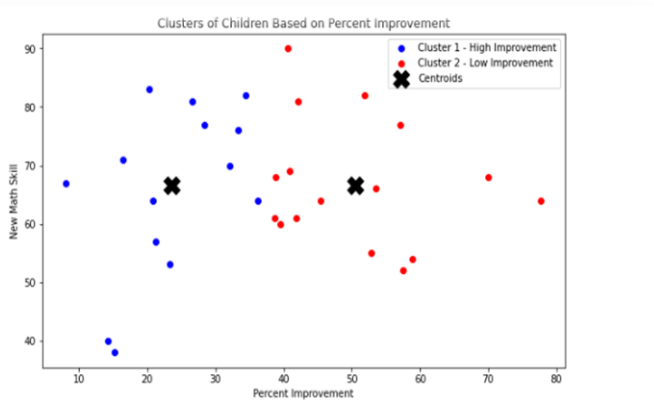


Figure 6 Scatter Plot of cluster results

6. Euclidean Distances

This visualization shows the Euclidean distance from each child to their centroid cluster. The bar chart in Figure 7 helps understand how close or far each child is from their centroid cluster based on the percentage improvement in their math skills. Smaller distances indicated that the percentage increase in those children was more representative of their clusters. In comparison, more considerable distances suggested that the percentage of improvement in those children was more outlier in their cluster.

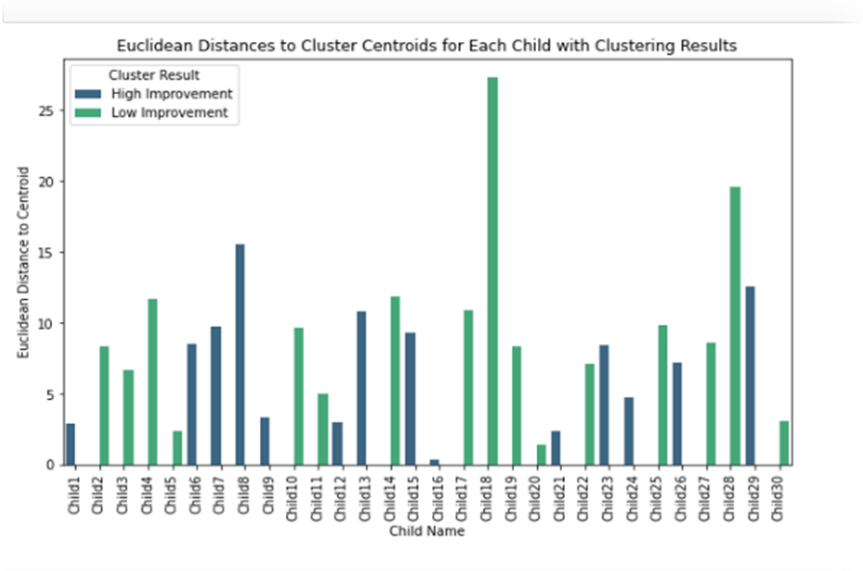


Figure 7 Euclidean Distances

7. Validity of Clustering

Clustering validity is used to measure the clustering results produced. In this study, the validity of clustering used Silhouette. Based on Figure 8, the results obtained are pretty good, showing that the clusters formed are quite good at separating data. Cluster 1: This cluster has a higher average silhouette score value, indicating that the members of this cluster are more similar to each other and more different from other clusters. Cluster 2: This cluster has a lower average silhouette score than Cluster 1 but is still in a positive range, indicating that the cluster members are also well grouped, although not as good as Cluster 1.

```
Silhouette Score: 0.5198100028772259
Average Silhouette Score per Cluster:
Cluster
1    0.575269
2    0.471283
Name: Silhouette_Value, dtype: float64
```

Figure 8 Silhouette Score

Figure 9 shows how well each child fits into their cluster. Students in high improvement clusters are generally grouped very well, with some having high silhouette scores. Although it also had a positive silhouette score, some students in this low improvement cluster had lower scores, indicating that they were not as strong in clustering as the High improvement cluster.

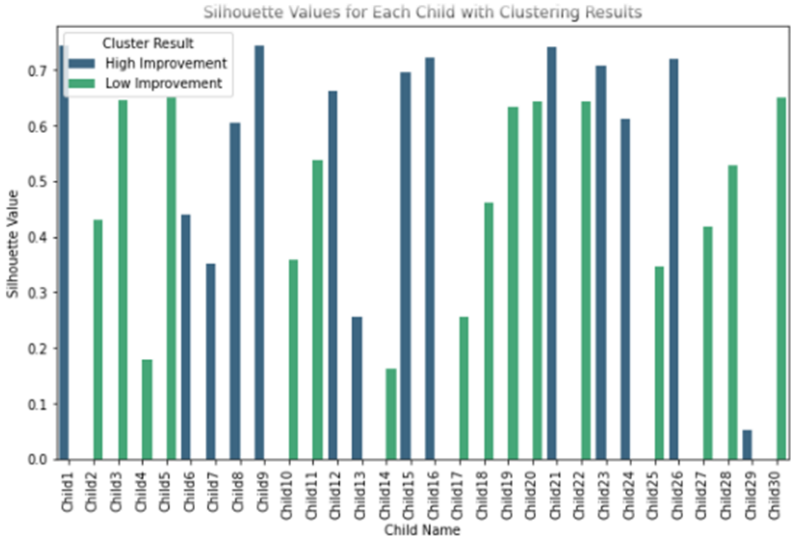


Figure 9 Silhouette Values

CONCLUSIONS

Applying the K-Means algorithm to the percentage of improvement data of children with special needs mathematics skills has divided it into two pretty good clusters, namely High Improvement and Low Improvement. While the clustering results are pretty good, with silhouette values supporting cluster quality, there are indications that the Low Improvement cluster may need further improvement or analysis to address outliers or more distributed distributions. Overall, these results suggest that the K-Means method can effectively group educational data of

children with special needs based on skill enhancement criteria, providing valuable insights for more targeted educational interventions.

ACKNOWLEDGMENTS

We want to thank our partners, Lembaga Terapi Mutiara Al Islam Surakarta, who supported the implementation.

REFERENCES

Ahmar, A. S., Napitupulu, D., Rahim, R., Hidayat, R., Sonatha, Y., & Azmi, M. (2018). Using K-Means Clustering to Cluster Provinces in Indonesia. *Journal of Physics: Conference Series*, *1028*(1). https://doi.org/10.1088/1742-6596/1028/1/012006

Asya, U. A., Huriati, Amal, A. A., & Arbianingsih. (2020). Occupational Therapy in Children with Special Needs: Systematic Review. *Journal Of Nursing Practice*, *3*(2), 268–278. https://doi.org/10.30994/jnp.v3i2.95

Chairunnisa, C., & Rismita, R. (2022). Educational challenges for children with special needs in inclusive primary schools. *Jurnal Ilmiah Sekolah Dasar*, *6*(1), 48–56. https://doi.org/10.23887/jisd.v6i1.39722

de la Fuente-Tomas, L., Arranz, B., Safont, G., Sierra, P., Sanchez-Autet, M., Garcia-Blanco, A., & Garcia-Portilla, M. P. (2019). Classification of patients with bipolar disorder using k-means clustering. *PLoS ONE*, *14*(1), 1–15. https://doi.org/10.1371/journal.pone.0210314

Duque, E., Gairal, R., Molina, S., & Roca, E. (2020). How the Psychology of Education Contributes to Research With a Social Impact on the Education of Students With Special Needs: The Case of Successful Educational Actions. *Frontiers in Psychology*, *11*(March). https://doi.org/10.3389/fpsyg.2020.00439

Dwijayanthi Nirmala, I., Prima, ;, & Atika, D. (2020). Implementation of K-Means Algorithm As a Clustering Method for Selecting Achievement Students Based on Academic Grade. *Jurnal PILAR Nusa Mandiri*, *16*(02), 199–204. https://doi.org/10.33480/pilar.v16i2.1575

Kalijaga, U. I. N. S. (2024). *Evaluation of Inclusive Education in University : Gaps in Program Implementation*. *5*(2), 202–211.

Kartono, D. T., Suryadinata, T. A., Zuber, A., Mundayat, A. A., Akbar, R. D., Andriani, L., Bawono, B. S., & Hidayah, U. (2023). Development of Rumah Inklusi support group for children with special needs. *Ittishal Educational Research Journal*,

4(2), 30–42. https://doi.org/10.51425/ierj.v4i2.45

Kefallinou, A., Symeonidou, S., & Meijer, C. J. W. (2020). Understanding the value of inclusive education and its implementation: A review of the literature. *Prospects*, *49*(3–4), 135–152. https://doi.org/10.1007/s11125-020-09500-2

Mareza, L. (2019). Implementation of Counseling Guidance for Children with Special Needs (Viewed from Psychological Aspects, Socio-Culture and Science and Technology Development). *Dinamika Jurnal Ilmiah Pendidikan Dasar*, *11*(2), 111. https://doi.org/10.30595/dinamika.v11i2.5730

Prianto, C., Nuraini, R., & Wali, A. T. (2019). Implementation of K-Means Methods In Clustering Students Ability Levels in English Language. *The IJICS (International Journal of Informatics and Computer Science)*, *3*(2), 49. https://doi.org/10.30865/ijics.v3i2.1382

Wafa, M. N., Moqim Hussaini, S. A., & Pazhman, J. (2020). Evaluation of students' mathematical ability in Afghanistan's schools using cognitive diagnosis models. *Eurasia Journal of Mathematics, Science and Technology Education*, *16*(6). https://doi.org/10.29333/EJMSTE/7834