Implementation of K-Means Clustering for Student Learning Outcome Analysis Using the Elbow Method at SDN 2 Tegalrejo

Muhammad Febria Hafid Syahputra [a1], Zaehol Fatah [a2]
[a]Information Systems Study Program, Faculty of Science and Technology, Ibrahimy University, East Java, Indonesia - 68374
e-mail: [1]hafid.srg@gmail.com, [2]zaeholfatah@gmail.com

Abstrak

Pendidikan merupakan upaya strategis untuk meningkatkan keterampilan individu yang dilakukan melalui proses pembelajaran yang terstruktur. Pemerataan akses pendidikan di sekolah menjadi salah satu pendekatan penting yang dapat diwujudkan melalui analisis capaian akademik setiap peserta didik. SD Negeri 2 Tegal Rejo, sebagai lembaga pendidikan formal, menghadapi variasi signifikan dalam rata-rata nilai siswa. Oleh karena itu, diperlukan metode pengelompokan untuk memahami dan memenuhi beragam kemampuan belajar siswa. Penelitian ini menggunakan Algoritma K-Means Clustering, sebuah teknik yang banyak digunakan dalam analisis data dan data mining. Data mining adalah proses sistematis untuk mengumpulkan dan mengolah data guna menghasilkan informasi yang bernilai. Secara khusus, penelitian ini menerapkan algoritma K-Means untuk mengelompokkan peserta didik berdasarkan rata-rata nilai mereka dengan tujuan mendapatkan informasi yang andal dan dapat ditindaklanjuti. Tahapan penelitian meliputi pengumpulan data, preprocessing, transformasi data, analisis clustering menggunakan algoritma K-Means, serta evaluasi menggunakan metode Elbow untuk menentukan jumlah klaster optimal. Hasil penelitian menunjukkan bahwa peserta didik dapat dikelompokkan ke dalam dua klaster. Klaster 0: Berisi 14 peserta didik dengan rata-rata nilai sebesar 53,8%. Klaster 1: Berisi 12 peserta didik dengan rata-rata nilai sebesar 46,2%. Hasil pengelompokan ini memberikan pemahaman yang komprehensif mengenai distribusi kemampuan akademik peserta didik. Informasi ini menjadi dasar dalam merancang strategi pembelajaran yang lebih efektif dan terarah, sehingga dapat mendukung peningkatan hasil belajar dan kualitas pendidikan secara keseluruhan.

Kata kunci: *Clustering, Data Mining, K-Means, Metode Elbow, Rapidminer*

Abstract

Education serves as a critical effort to enhance individual skills, primarily achieved through structured learning. Ensuring equitable access to education within schools is a key approach, which can be facilitated by analyzing students' academic achievements. SD Negeri 2 Tegal Rejo, as a formal educational institution, observes significant variations in the average performance scores among its students. To address this, a grouping method is necessary to better understand and cater to the diverse learning abilities of the students. This study employs the K-Means Clustering Algorithm, a prominent technique in the fields *of data analysis and data mining. Data mining involves the systematic collection and processing of data to uncover valuable insights. Specifically, this study utilizes the K-Means algorithm to categorize students based on their average scores, aiming to extract reliable and actionable information. The research methodology includes several stages: data collection, preprocessing, transformation, and clustering analysis using the K-Means algorithm, followed by evaluation through the Elbow method to identify the optimal number of clusters. The findings reveal two distinct clusters Cluster 0: Comprising 14 students with an average score of 53.8%. Cluster 1: Comprising 12 students with an average score of 46.2%. The clustering results offer a comprehensive understanding of the distribution of students' academic abilities. These insights serve as a basis for designing targeted and effective teaching strategies, contributing to improved educational outcomes and the overall quality of learning experiences.*

Keywords : *Clustering, Data Mining, K-Means, Elbow Method, Rapidminer.*

1. Introduction

The development of information technology has brought rapid progress, positively impacting various fields such as economics, health, arts, and education. Technology not only simplifies tasks and increases productivity but also strengthens the foundation of the education sector. At the primary education level, technology plays a crucial role in optimizing the teaching and learning process, which forms an essential foundation for shaping students' knowledge and skills from an early age. Quality education enhances human resource (HR) capabilities, ultimately supporting national progress. However, achieving quality education requires appropriate methods and approaches, one of which involves leveraging technology to analyze students' learning outcomes, enabling the full potential of each student to be realized.

At SD Negeri 2 Tegalrejo, Subdistrict Kelumpang Hilir, Regency Kotabaru, evaluating students' learning outcomes is a primary focus in improving the quality of teaching and learning activities (KBM). As part of the effort to enhance educational quality, the school is committed to analyzing the semester exam results of Grade 1 students. This analysis is critical for identifying the competency levels of each student individually, allowing the school to adapt teaching methods that are most effective and tailored to students' needs. This approach ensures that the teaching and learning process is equitable and effective, supporting the development of students according to their potential.

One technique that can be applied to support the analysis of learning outcomes is the clustering method. Clustering is a data grouping method where each cluster contains similar data while differing from other clusters (Apriliyaningsih and Istiawan, 2016). Clustering enables data separation based on characteristic similarities, allowing students to be grouped according to their learning achievement similarities. This method also introduces a new approach to addressing dependency and feedback between criteria in decision-making processes. One commonly used algorithm in clustering techniques is K-Means. This algorithm divides data into several clusters based on certain characteristics, which in this context are the students' semester exam results. Using the K-Means algorithm, schools can group students based on learning outcomes, ensuring that those with lower achievements receive targeted attention and specific learning strategies. Similarly, high-performing students can be directed to participate in development activities that align with their comprehension levels.

The K-Means algorithm has several advantages, including its ability to process large amounts of data quickly and efficiently. However, its effectiveness highly depends on determining the optimal number of clusters. To achieve this, the Elbow method is used as an approach to determine the best number of clusters by comparing the Sum of Squared Errors (SSE) values for several cluster count trials. The Elbow method helps produce optimal and representative cluster divisions, enabling more accurate analysis and ultimately providing beneficial results for decision-making related to teaching strategies.

Clustering techniques, such as K-Means, enable educators to identify patterns in student achievement, providing opportunities to design more targeted and effective teaching strategies. At SD Negeri 2 Tegalrejo, this algorithm has been utilized to categorize Grade 1 students based on their semester exam results. This study aims to address several key questions:

1. How can students be grouped optimally using the K-Means algorithm to ensure that each cluster accurately represents their academic performance?
2. Do the resulting clusters reflect the diversity in students' academic achievement levels?
3. How can the school leverage the clustering results to improve teaching and learning quality? For instance, can these findings assist teachers in tailoring their teaching methods to meet the specific needs of each group?
4. Can the K-Means algorithm help educators identify groups of students with particular learning needs, such as those requiring remedial support or advanced challenges?
5. How can the clusters be used to develop targeted interventions, such as additional tutoring sessions, interactive teaching approaches, or enrichment programs, to enhance students' comprehension and academic growth effectively?.

2. Research Method

The approach used in this study follows the CRISP-DM method (Cross-Industry Standard Process for Data Mining). CRISP-DM consists of five stages and a framework, as illustrated in Figure 1 below :

Figure 1 CRISP-DM

3.1 Business Understanding

This stage focuses on defining the business problem and understanding the research environment. The business problem to be addressed must be clearly defined and measurable. The business understanding phase begins with identifying the long-term objectives of the business. Once these objectives are understood, the next step is to analyze the existing business problems and the factors contributing to them. After gaining insights into the business objectives and the problems encountered, the final step is to determine the methods or solutions to address these issues effectively.

3.2 Data Understanding

This stage involves collecting initial data, exploring it, and assessing its quality. The aim is to obtain a preliminary understanding of the data. The data can be sourced from databases, data warehouses, the internet, and other sources, then processed into meaningful information.

3.3 Data Preparation

This stage involves transforming raw data into a more interpretable format. It includes several processes: data integration, which combines data from multiple sources; data cleaning, which removes or replaces null values; and data selection, where relevant data is chosen based on the research focus. Additionally, data adjustment is crucial to ensure the data aligns with the research requirements.

3.4 Modeling

Modeling is the stage where models are selected and applied using data mining algorithms. This stage aims to optimize the research outcomes. In this study, two modeling approaches were implemented.

a. Elbow Method

The Elbow method is a technique used to determine the optimal number of clusters (c) by calculating the Sum of Square Error (SSE) for each cluster. The larger the difference in SSE values between adjacent clusters, forming a sharp angle, the better the chosen number of clusters. The value of c in the Elbow and K-Means combination is represented in a plot showing the relationship between the number of clusters and the reduced error. Increasing the value of c causes the graph to decline gradually until the value of k stabilizes. SSE is used to determine the optimal number of c. This method tests the SSE values for different cluster counts and searches for the largest SSE difference, forming a right angle on the elbow plot, to find the best number of clusters. The formula for SSE is shown in Equation 1.

$$SSE = \sum_{K=1}^{K} \sum_{x_i e S_k} ||x_i - c_k||_2^2$$

b. K-Means Clustering

K-Means is a non-hierarchical data clustering technique that involves dividing data into one or more groups or clusters. Data is grouped into a cluster based on the similarity of attributes. This similarity can be measured by calculating the distance of each data point to the cluster center (centroid). The K-Means Clustering method is a widely used approach for clustering large datasets due to its quick and efficient process. The resulting clusters provide valuable knowledge or information for decision-makers in policy formulation. The basic concept of clustering is to group several objects into clusters, where a good cluster is one that has a high degree of similarity among the objects within it and a high level of dissimilarity with objects from other clusters. The following outlines the calculation steps for the K-Means algorithm.

1. Determine the number of clusters (c) or groups.
2. Randomly determine the initial cluster center (centroid)
3. Measure the distance of each data from the cluster center (centroid), namely using the Euclidean Distance formula in equation 2.
   $$D_{(ij)} = \sqrt{\left\{\left(x_{(1p)} - x_{(pq)}\right)^2 + \left(x_{(2p)} - x_{(2q)}\right)^2 + \ldots + \left(x_{(rp)} - x_{(rq)}\right)^2\right\}}$$
   Explanation:
   D_{i,j} = the distance between the p-th data point and the center of cluster q
   x_{r,p} = the p-th data point in the r-th attribute
   x_{r,q} = the center point of cluster q in the r-th attribute
4. Group each data into clusters based on the minimum distance.
5. Perform the iteration process by determining the new cluster center (centroid) using the formula in equation 3.
   $$v = \frac{\sum_{p=1}^{n} x_i}{n}\ ; p = 1,2,3, \ldots, n$$
   Explanation:
   v = centroid of the cluster
   x(p) (p) = the p-th object
   n = the number of objects/total number of objects
6. Repeat step 3 until there are no cluster changes in each data from the previous iteration process.

3.5 Evaluation

The evaluation stage aims to measure the performance and accuracy of the model obtained in accordance with the target set in the first stage. In this study, the performance and accuracy of the developed model are measured using the Davies-Bouldin Index**.**

a. Davies Bouldin Indeks

The Davies-Bouldin Index is a clustering evaluation technique based on the similarity within clusters and the dissimilarity between clusters in the distribution of clusters. The Davies-Bouldin Index aims to assess the performance of a model in determining the optimal number of clusters. The smaller the DBI value, the more optimal the number of clusters. The formula for DBI can be seen in the following Equation 4**.**

$$DBI = \frac{1}{k} \sum_{i=1}^{k} max_{i \neq j}(R_{ij})$$

3. Result and Discussion

3.1 Business Understanding

Business understanding is the initial stage in the CRISP-DM method. During the business understanding phase, the objectives of the research are determined, and strategies are prepared to achieve the set goals (Defiyanti, Jajuli, and Rohmawati, 2017). The object of this study is

SD Negeri 2 Tegal Rejo in Kelumpang Hilir Subdistrict, Kotabaru Regency. The main issue at SD Negeri 2 Tegal Rejo is the need for an evaluation of the first-grade students' learning outcomes to identify their competencies and standardize the quality of the Teaching and Learning Activities (KBM). The purpose of this research is to assess the academic abilities of the students to help improve the quality of teaching. Therefore, this study will cluster the first-grade students' learning outcomes based on their semester exam scores using the K-Means method.

3.2 Data Understanding

Based on the understanding of the existing problems, the data required for this research are the semester exam scores of first-grade students at SD Negeri 2 Tegal Rejo. This data is collected from several core subjects during the semester to explore patterns in academic achievement. This stage aims to understand and analyze the collected data so that it can be used in the clustering process. The data gathered includes the average scores of first-grade students in several core subjects from the first-semester exam.

3.3 Data Preparation

This stage involves the data preparation process, which includes data integration, data cleaning, and data selection. In the data integration stage, the subject scores of each student are combined into one table for analysis. The data used includes scores from the subjects of PABP, PKn, Indonesian, Mathematics, IPAS, PJOK, SB, PAQ, English, BTA, and ICT for first-grade students at SD Negeri 2 Tegal Rejo.

Next, data cleaning is performed to ensure there are no missing or invalid values. In this case, the data provided is already clean and complete, with no missing values or input errors, allowing it to be processed further.

In the data selection stage, from the available subjects, relevant attributes are selected for the clustering analysis. The attributes chosen for this analysis are PABP, PKn, Indonesian, Mathematics, IPAS, and English, which are considered to represent the students' academic competencies in various areas.

The table below shows some sample data of average scores from students at SD Negeri 2 Tegal Rejo:

Table 1 Sample Data of Average Student Scores

No	PABP	PKn	B.Ind	Math	IPAS	PJOK	SB	PAQ	English	BTA	IT
1	82	73	75	77	78	78	82	78	75	75	75
2	89	77	76	76	83	82	82	79	82	79	82
3	90	90	85	81	84	89	84	95	91	87	87
4	90	85	91	84	89	91	85	97	95	86	87
5	85	89	87	87	77	85	85	87	87	84	84

6	84	92	87	88	83	88	87	98	88	84	83
7	86	76	76	79	79	82	79	86	80	82	77
8	82	96	91	91	84	91	82	98	96	92	90
9	75	73	73	71	74	82	80	70	70	77	76
10	77	76	76	74	81	76	78	76	76	78	77
11	85	93	90	88	86	90	82	92	92	90	92
12	97	93	91	91	96	91	85	96	95	92	86
13	85	85	87	84	84	81	81	86	80	83	81
14	92	74	76	77	76	77	77	77	78	77	78
15	77	76	76	76	74	76	78	78	77	78	77
16	84	76	74	79	78	78	82	79	79	78	77
17	85	79	77	79	82	79	82	79	82	82	81
18	89	76	76	80	83	83	80	80	80	80	80
19	89	97	93	91	96	90	90	96	96	96	88
20	82	93	92	90	96	96	91	92	92	92	86
21	86	84	74	84	83	88	85	91	86	84	80
22	78	74	77	78	75	74	81	81	75	74	74
23	81	77	81	77	77	77	78	77	76	78	81
24	84	78	77	81	78	75	81	82	81	81	80
25	88	92	84	82	81	82	93	93	92	93	92
26	88	92	90	89	83	83	92	91	91	92	90

The data preparation process using the RapidMiner application, as illustrated in the following image.

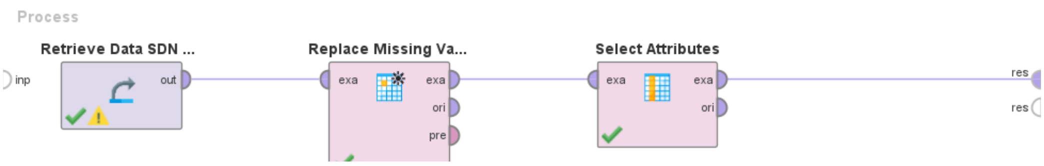

Figure 2 Data Preparation Process Using the RapidMiner Application

3.4 Modelling

After the data is ready for use, the next step is modeling. In this study, the optimal number of clusters (c) was determined using the elbow method. With the elbow method, the number of clusters is considered optimal when the graph comparing the Sum of Squared Errors (SSE) and the number of clusters forms an 'elbow,' indicating a significant change. Table 2 below shows the SSE values for cluster counts ranging from c=2 to c=8.

Table 2 Comparison of SSE Values with the Number of Clusters

No	*Number of Clusters* (c)	*Sum of Square Error (SSE)*
1	2	143.340
2	3	110.484
3	4	90.502
4	5	75.230
5	6	68.195
6	7	54.585
7	8	49.360

Figure 3 below illustrates the graph comparing SSE values with the number of clusters (c).

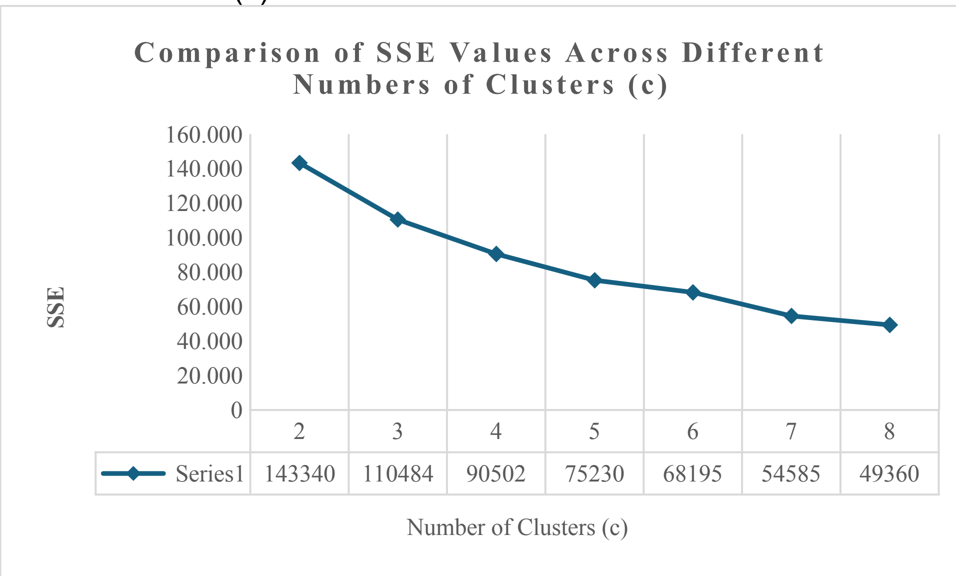

Figure 3 Graph Comparing SSE Values with the Number of Clusters (c)

Based on Figure 3, the line forming an elbow or the sharpest angle is observed at a cluster count of c=2. This indicates that c=2 is the optimal number of clusters, with an SSE value of 143.340. Therefore, the data will be grouped into 2 clusters. After determining the value of c, the next step is to apply the K-Means algorithm to the average subject scores at SD Negeri 2 Tegal Rejo with c=2. The results of clustering these average subject scores using the RapidMiner application are shown in Figure 4 below.

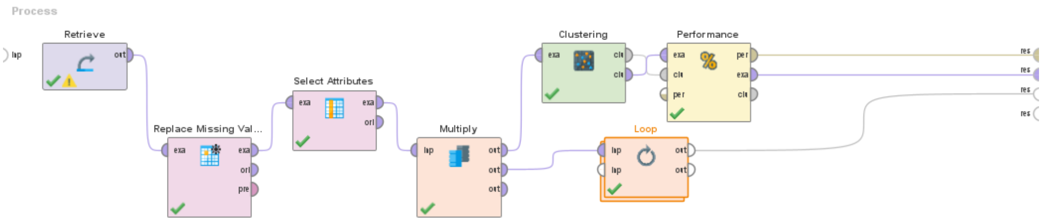

Figure 4 Clustering Process Using the RapidMiner Application

Figure 5 below shows the Cluster Model results generated using the RapidMiner application.

Index	Nominal value	Absolute count	Fraction
1	cluster_0	14	0.538
2	cluster_1	12	0.462

Figure 5 Cluster Model Results Using the RapidMiner Application

Based on Figure 5, it can be observed that Cluster 0 consists of 14 students, while Cluster 1 consists of 12 students. Cluster 0 has the highest number of members, followed by Cluster 1 with the fewest members. To determine the category of each cluster, a visualization using a bar chart is performed, as shown in Figure 6.

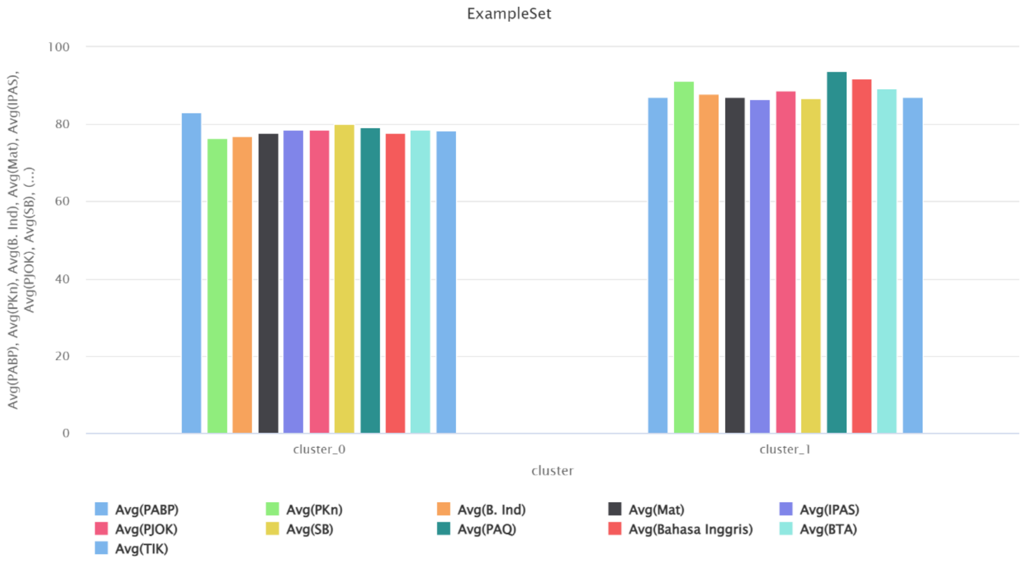

Figure 6 Graph Comparing the Average Scores of Subjects in Each Cluster

Figure 6 shows that Cluster 0 is dominated by students with relatively low average scores, consisting of 28 members. Meanwhile, Cluster 2 is dominated by students with moderate average scores, with 42 members. Lastly, Cluster 1 represents the group with high average scores, consisting of only 1 student, as detailed in Table 3.

Table 3 Results of Student Score Clustering

No	PABP	PKn	B. Ind	Mat	IPAS	PJOK	SB	PAQ	English	BTA	TIK	Cluster
1	82	73	75	77	78	78	82	78	75	75	75	cluster_1
2	89	77	76	76	83	82	82	79	82	79	82	cluster_1
3	90	90	85	81	84	89	84	95	91	87	87	cluster_0
4	90	85	91	84	89	91	85	97	95	86	87	cluster_0
5	85	89	87	87	77	85	85	87	87	84	84	cluster_0
6	84	92	87	88	83	88	87	98	88	84	83	cluster_0
7	86	76	76	79	79	82	79	86	80	82	77	cluster_1
8	82	96	91	91	84	91	82	98	96	92	90	cluster_0
9	75	73	73	71	74	82	80	70	70	77	76	cluster_1
10	77	76	76	74	81	76	78	76	76	78	77	cluster_1
11	85	93	90	88	86	90	82	92	92	90	92	cluster_0
12	97	93	91	91	96	91	85	96	95	92	86	cluster_0
13	85	85	87	84	84	81	81	86	80	83	81	cluster_1
14	92	74	76	77	76	77	77	77	78	77	78	cluster_1
15	77	76	76	76	74	76	78	78	77	78	77	cluster_1
16	84	76	74	79	78	78	82	79	79	78	77	cluster_1
17	85	79	77	79	82	79	82	79	82	82	81	cluster_1
18	89	76	76	80	83	83	80	80	80	80	80	cluster_1
19	89	97	93	91	96	90	90	96	96	96	88	cluster_0
20	82	93	92	90	96	96	91	92	92	92	86	cluster_0
21	86	84	74	84	83	88	85	91	86	84	80	cluster_0
22	78	74	77	78	75	74	81	81	75	74	74	cluster_1
23	81	77	81	77	77	77	78	77	76	78	81	cluster_1
24	84	78	77	81	78	75	81	82	81	81	80	cluster_1
25	88	92	84	82	81	82	93	93	92	93	92	cluster_0
26	88	92	90	89	83	83	92	91	91	92	90	cluster_0

Based on the clustering results of students' average scores at SD Negeri 2 Tegal Rejo, it is observed that there is a disparity in the number of members among Cluster 1, Cluster 0, and Cluster 2. Cluster 1, comprising students with moderate average scores, has a larger number of members compared to Cluster 0 and Cluster 2. Cluster 0 consists of students with high average scores, while Cluster 2 includes students with very high average scores.

These findings highlight differences in academic achievement among the student groups. SD Negeri 2 Tegal Rejo may consider developing more targeted teaching strategies to address this gap, focusing more attention on students in Cluster 1 to improve their learning outcomes and narrow the achievement gap with students in Clusters 0 and 2. With a more inclusive approach, it is hoped that students' learning outcomes at SD Negeri 2 Tegal Rejo can become more balanced across all subjects.

3.5 Evaluation

The evaluation phase involves measuring the clustering performance of students' scores for three clusters (c=3) using the Davies-Bouldin Index (DBI). The DBI value generated for this model using the K-Means algorithm is 0.114 or 2202, as shown in Figure 7.

Davies Bouldin

Davies Bouldin: 0.647

Figure 7 DBI Value with Number of Clusters (c) = 2

To determine whether c=3 is the most optimal number of clusters, an experiment was conducted by calculating the DBI values for cluster counts ranging from c=2 to c=7, as shown in Table 4.

Table 4 Comparison of Number of Clusters (c) with DBI

No	Number of Clusters (c)	*Davies Bouldin Index (DBI)*
1	2	0.647
2	3	1.142
3	4	1.204
4	5	1.133
5	6	1.097
6	7	0.970
7	8	1.184

Based on Table 4, the smallest DBI value is observed when the number of clusters is 3. A smaller DBI value, closer to 0, indicates that the data points within a cluster are more similar. Conversely, a larger DBI value, closer to 2, suggests that the data points within a cluster are less similar. Therefore, it is confirmed that c=3 is the most optimal number of clusters.

4. Conclusion

The results of the clustering analysis using the K-Means algorithm show that the dataset is divided into two clusters, namely Cluster 0 with the largest number of members of 14 students and Cluster 1 which is smaller with 12 students. The imbalance in the number of members in each cluster can reflect differences in the level of academic achievement among students. Therefore, it is recommended to provide additional classes for students in groups with lower achievements and apply more interactive teaching methods to improve their understanding. Further research is recommended to add more attributes and explore other clustering algorithms and techniques in order to gain deeper insights into student achievement patterns, which can ultimately support the development of more optimal teaching strategies.

5. Acknowledgment

We extend our heartfelt gratitude to everyone who provided guidance and valuable input during the preparation of this research methodology. It is

our hope that the explanation of the methods used offers readers a comprehensive understanding of the research process undertaken.

References

[1] Apriliyaningsih, I., Dewi, I., & Istiawan, D. (2016). Penerapan seleksi atribut berdasarkan koefisien variansi dan korelasi untuk inisialisasi pusat awal klaster pada algoritma K-Means dalam pemetaan e-government tahun 2016. University Research Colloquium, 1(11), 245–250. https://journal.unimma.ac.id/index.php/urecol/article/view/1074/753.

[2] Dayera, M. B. P., & Ohello, F. (2024). G-Tech: Jurnal Teknologi Terapan, 8(1), 186–195. https://ejournal.uniramalang.ac.id/index.php/g-tech/article/view/1823/1229.

[3] Defiyanti, S., Jajuli, M., & Rohmawati, N. (2017). Optimalisasi K-MEDOID dalam pengklasteran mahasiswa pelamar beasiswa dengan CUBIC CLUSTERING CRITERION. Jurnal Nasional Teknologi dan Sistem Informasi, 3(1), 211–218. https://doi.org/10.25077/teknosi.v3i1.2017.211-218.

[4] Jollyta, D., Siddik, M., Mawengkang, H., & Efendi, S. (n.d.). Teknik evaluasi cluster solusi menggunakan Python dan Rapidminer. Deepublish.

[5] Jollyta, D., Efendi, S., Zarlis, M., & Mawengkang, H. (2019). Optimasi cluster pada data stunting: Teknik evaluasi cluster sum of square error dan Davies Bouldin Index. Prosiding Seminar Nasional Riset Information Science (SENARIS), 1, 918–926.

[6] Farissa, R. A., Mayasari, R., & Umaidah, Y. (2021). Perbandingan algoritma K-Means dan K-Medoids untuk pengelompokkan data obat dengan silhouette coefficient di Puskesmas Karangsambung. Journal of Applied Informatics and Computing, 5(2), 109–116. https://doi.org/10.30871/jaic.v5i1.3237.

[7] Iskandar, M. S., & Fatah, Z. (2024). Implementasi metode algoritma K-Means clustering untuk menentukan penerima program Indonesia Pintar (PIP). Gudang Jurnal Multidisiplin Ilmu, 2(11), 1–8.

[8] Nabila, Z., Isnain, A. R., & Abidin, Z. (2021). Analisis data mining untuk clustering kasus Covid-19 di Provinsi Lampung dengan algoritma K-Means. Jurnal Teknologi dan Sistem Informasi (JTSI), 2(2), 100. http://jim.teknokrat.ac.id/index.php/JTSI.

[9] Khormarudin, N. A., Panjaitan, O. C. K. S., Septianingsih, A., Faisal, M., Utami, W. S., Ningsi, L. N., Tambunan, H. S., Zai, C., & Teknik Komputer. (2022). Teknik data mining: Algoritma K-Means clustering. Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika, 1(2), 116–123. https://djournals.com/klik%0Ahttps://ilmukomputer.org/category/datamining/.

[10] Oktaviany, N., Suarna, N., & Prihartono, W. (2024). Implementasi algoritma K-Means clustering dalam mengelompokkan kepadatan penduduk di Provinsi DKI Jakarta. JATI (Jurnal Mahasiswa Teknik Informatika), 8(1), 19–126. https://doi.org/10.36040/jati.v8i1.8241.