**K-MEANS CLUSTERING METHOD FEASIBILITY OF SCHOOL BUILDING REHABILITATION IN KABUPATEN ASAHAN**

**Irma Nurhani1\*, Nuriadi Manurung2, Elly Rahayu3**

1,2,3 Sistem Informasi, Universitas Royal

*email*: irmanurhani0308@gmail.com

**Abstract:** School buildings and classrooms are an important part of educational facilities and infrastructure. The classroom is a place for the learning process and direct interaction between teaching staff and students. Students will feel comfortable with the infrastructure that supports the teaching and learning process, this will result in better quality education, especially in Asahan Regency. Therefore, to overcome this problem, decision support is needed to assist the Education Department in determining the feasibility of rehabilitation and construction of school buildings, namely by grouping existing school data using the more structured K-Means Cluster Data Mining method to manage and analyze the data optimally. The aim of this research is to show that data mining clustering increases the accuracy of grouping data on the feasibility of rehabilitation and construction of elementary school buildings. The K-Means clustering algorithm in clustering the feasibility of rehabilitating elementary school buildings, there are 31 elementary schools that are suitable for building rehabilitation and 4 that are not suitable for rehabilitating school buildings.

**Keywords:** education authorities; k-means; rehabilitation and construction of school buildings;

**Abstrak:** Bangunan/gedung sekolah dan ruang kelas merupakan bagian penting dalam sarana dan prasarana pendidikan. Ruang kelas merupakan sebuah tempat proses pembelajaran dan interaksi langsung antara tenaga pendidik dengan peserta didik. Anak didik akan merasa nyaman dengan adanya prasarana yang mendukung proses belajar mengajar, hal ini akan menghasilkan mutu Pendidikan yang lebih baik khususnya di Kabupaten Asahan. Oleh karena itu, untuk mengatasi permasalahan tersebut diperlukan suatu pendukung keputusan untuk membantu pihak Dinas Pendidikan dalam menentukan kelayakan rehabilitasi dan pembangunan gedung sekolah yaitu dengan cara mengelompokkan data sekolah yang ada dengan menggunakan Data Mining metode *K-Means Cluster* yang lebih terstruktur untuk mengelola dan menganalisis data tersebut secara optimal. Tujuan penelitian ini menunjukkan bahwa data mining *clustering* meningkatkan akurasi pengelompokan data kelayakan rehabilitasi dan pembangunan gedung sekolah dasar. Algoritma *K-Means clustering* dalam *mengcluster* kelayakan rehabilitasi pembangunan gedung sekolah dasar terdapat 31 sekolah dasar layak untuk direhabilitasi pembangan gedung dan 4 tidak layak direhabalitasi pembangunan gedung sekolah.

**Kata kunci:** dinas pendidikan; k-means; rehabilitasi dan pembangunan gedung sekolah;

**INTRODUCTION**

Educational buildings, including elementary school buildings, have a very important role in increasing the human development index. Development in the education sector, including at the elementary school level, is a process of investment in human resources which has an important role and function in supporting overall national development. Education, especially in elementary schools as the initial foundation, functions as a system that has a large influence, is interdependent, coordinated and structured, aiming to realize educational goals in order to make the life of the nation smarter, which is the main focus of every educational institution. [1]. Elementary schools are the schools with the largest number because they are a place of education for children during the first six years of formal learning. A school building that is well maintained can provide optimal support for the teaching and learning process. Adequate facilities, starting from the school environment, building conditions, to supporting equipment such as mobiles, will create a comfortable atmosphere for students, so that they can support them in achieving the maximum educational experience [2].

School buildings and classrooms are an important part of educational facilities and infrastructure. The classroom is a place for the learning process and direct interaction between teaching staff and students. Students will feel comfortable with the infrastructure that supports the teaching and learning process, this will result in better quality education, especially in Asahan Regency [3].

Asahan Regency is one of the districts/cities in North Sumatra, Indonesia. Asahan Regency has 25 (twenty five) sub-districts. The total number of elementary schools is around 452 schools. In 2024 there will be around 380 (three hundred and eighty) State Primary Schools and around 72 (seventy two) Private Primary Schools registered with the Asahan Education Service spread across the 25 sub-districts. In Asahan Regency, the regional government and the local Education Office continue to strive to improve access and quality of education. However, the condition of education in Asahan Regency is very worrying, especially in terms of infrastructure.

Even though there have been various efforts to develop educational infrastructure, the challenges faced related to the inadequate condition of school buildings are one of the main challenges in efforts to improve the quality of education. Many school buildings are unfit for use, are damaged, or lack basic facilities, such as inadequate classrooms and other supporting infrastructure. This problem not only hampers the teaching and learning process but also has the potential to endanger the safety of students and teaching staff.

Therefore, serious attention from various parties, including the government and related institutions, is needed to ensure effective handling steps are taken. Providing adequate school facilities is an important foundation in creating an educational environment that is safe, comfortable and supports maximum achievement of learning goals.

Initial mapping carried out by the Education Department shows that there are a number of schools that require rehabilitation, as well as several schools that require new classrooms to be built in response to the growth in student numbers. The government, through various programs, has allocated a budget for the rehabilitation and construction of school buildings. However, the main challenge in implementation is how to prioritize school buildings that must be repaired or rebuilt efficiently. This decision grouping process is often hampered by limited structured data, causing budget allocation to be less than optimal.

The conventional manual approach in determining the feasibility of rehabilitation and construction of school buildings is often subjective and takes a long time. In this case, evaluating the feasibility of rehabilitation and construction of school buildings needs to be carried out using a more systematic and measurable approach. Apart from considering physical conditions such as damage to buildings and facilities, other aspects such as room capacity, age of the building, light, medium and heavy damage must also be taken into account [4].

With the complexity of the data that needs to be analyzed, conventional manual approaches become less effective and efficient. The lack of a comprehensive system for determining the appropriateness grouping for school building rehabilitation often results in inequality, where schools that actually need rehabilitation more often do not get it. On the other hand, there are cases where schools with unclear land status still receive rehabilitation. This often triggers conflicts during the physical construction process with parties who claim ownership of the school land [5].

Therefore, to overcome this problem, decision support is needed to assist the Education Department in determining the feasibility of rehabilitation and construction of school buildings, namely by grouping existing school data using the more structured K-Means Cluster Data Mining method to manage and analyze the data optimally.

Data Mining is the process of collecting and processing data to understand interesting information or patterns in data, using certain algorithms, techniques and techniques, with the aim of finding important information in data [6]. Data Mining can also be explained as processing to obtain information using mathematical, statistical, machine learning and artificial intelligence techniques to identify useful and binding knowledge from various large data warehouses and databases [7]. The process of grouping records that have useful meaning is called Clustering. Clustering is grouping a number of data or objects into clusters (groups) so that each cluster will contain data that is as similar as possible and different from objects in other clusters [8]. One of the Clustering algorithms found in data mining is used to obtain groups that have large amounts of data using an efficient and fast separation technique, namely the K-Means Algorithm. The selection of this technique is very easy in the process of application or use and adapts according to the case [9].

Research using the same technique under the title Application of the K-Means Technique on Underprivileged Data in Blitar District, The number of samples used was 22 sub-districts in Blitar Regency and produced 2 clusters, namely cluster 0 and cluster 1 [10]. The next research is through the title of student data clustering using the K-Means algorithm to support promotional strategies. The test results were carried out using 171 datasets via centroids. Promotional strategies for prospective new students will later follow clusters that are formed based on the most popular study programs at each school [11]. Next, the research is entitled Grouping Book Borrowing Activities in the STMIK Widya Pratama Library Using the K-Means Technique. In this research, the K-Means algorithm can be used to classify book borrowing activity through the GPA variables and the amount borrowed. The number of clusters is divided into 2 groups [12]. Next, the research is entitled Application of the Clustering Algorithm for Grouping Poverty Levels in Banten Province. In this research, 3 regional groups were selected based on poverty level. The algorithm used in this research has a process and the cluster division vector looks the same, however based on the DBI results obtained k-medoid is 0.582, and k-means is 0.602.[13]. Clustering is the process of grouping data into several clusters or groups, where the data in one cluster has high similarity, while the similarity between clusters is minimal [14].

K-means is a non-hierarchical data clustering method that groups data in the form of one or more clusters/groups. Data that has the same characteristics is grouped and data that has different characteristics is grouped with other clusters/groups so that data in one cluster/group has a small level of variation. The aim is to divide the data into one or more clusters. The advantage of K-Means lies in its ability to group large numbers of objects quickly, thereby speeding up the grouping process. The purpose of data clustering is to minimize the objective function set in the clustering process [15].

This research focuses on the application of the K-Means Clustering algorithm to analyze the feasibility of rehabilitation and construction of elementary school buildings in Asahan Regency. It is hoped that the results of this research will not only provide recommendations based on data and facts, but will also serve as a guide for policy makers in planning and developing educational infrastructure in the region. Apart from that, it is hoped that this effort can support improving the quality of education in Asahan Regency and create a safer and more comfortable learning environment for students. Based on the background of the problem, the author wants to design and create a computerized system that can cluster the feasibility of rehabilitation and construction of elementary school buildings.

**METHOD**

Data mining and machine learning techniques can be used to make predictions based on past data. Data mining is the process of finding useful patterns in large data sets. From other sources, data mining is the study of collecting, cleaning, processing, analyzing, and obtaining useful insights from data [16].

Data grouping the feasibility of rehabilitation and construction of school buildings by taking samples from the Asahan District Education Office, the following researchers created as an illustration of how researchers process the existing data which will later become a summary per school to then be entered into the calculation formula for the K-Means Clustering Analysis Algorithm. Can be seen in the table 1.

Table 1. Grouping Criteria Scores at the Education Service Kab. Asahan

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Criteria (C) | Data Source | Weight Value Terms | |
| 1 | Study Room Damage (C1) | Dinas Pendidikan Kabupaten Asahan | 1 | Low |
| 2 | Currently |
| 3 | Heavy |
| 2 | Damage to Student Support Facilities (C2) | Dinas Pendidikan Kabupaten Asahan | 1 | Low |
| 2 | Currently |
| 3 | Heavy |
| 3 | Damage to Toilets and Places of Worship (C3) | Dinas Pendidikan Kabupaten Asahan | 1 | Low |
| 2 | Currently |
| 3 | Heavy |

The calculation algorithm using the K-Means method is an algorithm for clustering and attribute-based objects into k partitions, where k < n. In general, K-means clustering is a non-hierarchical data clustering technique that groups data in the form of one or more groups [17].

The aim is to minimize the objective function set in the grouping process, which generally seeks to minimize variation within a group and maximize variation between groups. The steps in the K-Means Clustering Algorithm are by deciding the number of clusters and deciding the centroid value [18].

In deciding the centroid value for the start of the solution, the initial centroid value is done randomly. Meanwhile, if the centroid value is the stage of completion, then the following formula is used [19]:

(1)

Calculates the distance between the centroid points through the points of each object.

Grouping objects to determine cluster members is by calculating the minimum distance of objects. Return to stage 2, repeat until the resulting centroid value remains constant and cluster members do not move to another cluster.

**RESULTS AND DISCUSSION**

In working on the implementation of the K-Means algorithm in this research, the author used the clustering function which is part of data mining which is useful for dividing a set of data into several clusters/groups based on similarities or similarities based on community data in table 1, to be normalized based on the criteria in table 2. This normalization can be seen in table 3.

Table 2. Alternative Value Data Per Criteria

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Alternative Name | C`1 | C2` | C3 |
| 1 | UPTD SDN 017127 Sei Nangka | Heavy | Currently | Heavy |
| 2 | UPTD SDN 010043 Perk. Air Batu III/IX | Heavy | Currently | Heavy |
| 3 | UPTD SDN 010076 Pondok Bungur | Heavy | Currently | Currently |
| 4 | UPTD SDN 010110 Desa Ambalutu | Currently | Heavy | Heavy |
| 5 | UPTD SDN 013819 Desa Sei Lama | Heavy | Heavy | Heavy |
| 6 | UPTD SDN 013830 Desa P. Mahondang | Currently | Heavy | Heavy |
| 7 | UPTD SDN 014649 Desa P. Rakyat Tua | Heavy | Heavy | Currently |
| 8 | UPTD SDN 014667 Danau Sijabut Gardu | Heavy | Heavy | Heavy |
| 9 | UPTD SDN 015933 Lobu Rappa | Heavy | Currently | Currently |
| 10 | UPTD SDN 014671 Sentang | Heavy | Currently | Heavy |
| - | - | - | - | - |
| 35 | [UPTD SDN 013842 Sei Silau Tua](https://dapo.dikdasmen.go.id/sekolah/535AD53045AC469DEADE) | Low | Low | Currently |

Table 3. Criterion Data Normalization

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Alternative Name | C1 | C2 | C3 |
| 1 | UPTD SDN 017127 Sei Nangka | 3 | 2 | 3 |
| 2 | UPTD SDN 010043 Perk. Air Batu III/IX | 3 | 2 | 3 |
| 3 | UPTD SDN 010076 Pondok Bungur | 3 | 2 | 2 |
| 4 | UPTD SDN 010110 Desa Ambalutu | 2 | 3 | 3 |
| 5 | UPTD SDN 013819 Desa Sei Lama | 3 | 3 | 3 |
| 6 | UPTD SDN 013830 Desa P. Mahondang | 2 | 3 | 3 |
| 7 | UPTD SDN 014649 Desa P. Rakyat Tua | 3 | 3 | 2 |
| 8 | UPTD SDN 014667 Danau Sijabut Gardu | 3 | 3 | 3 |
| 9 | UPTD SDN 015933 Lobu Rappa | 3 | 2 | 2 |
| 10 | UPTD SDN 014671 Sentang | 3 | 2 | 3 |
| - | - | - | - | - |
| 35 | [UPTD SDN 013842 Sei Silau Tua](https://dapo.dikdasmen.go.id/sekolah/535AD53045AC469DEADE) | 1 | 1 | 2 |

To determine nominal data, it must be initialized first in the form of numbers by sorting the data based on the frequency of occurrence and initializing the data starting from the highest data with a value of 1, then the next data 2, 3 and so on. To decide on the initial center of the cluster, deciding on the centroid value for the start of the iteration is done randomly. In this case the author made an example for 2 elementary schools, namely, PTD SDN 013819 Sei Lama Village and UPTD SDN 015893 Air Putih. Can be seen in the tablel 5.

Table 4. Iteration Centroid Point 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Kode | Name | C1 | C2 | C3 |
| A05 | SDN 013819 Desa Sei Lama | Heavy | Heavy | Heavy |
| A32 | SDN 015893 Air Putih | Low | Low | Low |

After deciding on the initial center of the cluster, deciding on the centroid value for the start of the iteration is done randomly. In this case the author made examples for 2 elementary schools, namely, SDN 013819 Sei Lama Village (Data 5), ​​SDN 015893 Air Putih (Data 32). Calculate the distance between the centroid points through the points of each object using the Euclidian Distance formula, namely;

One example is the calculation of Code A01 (SDN 017127 Sei Nangka) to later Code A35 (SDN 013842 Sei Silau Tua):

Because C1 is looking for a new cluster calculation it becomes:

A1=

A2 =

A3 =

A4 =

A5 =

and so on until A35

Because C2 is looking for a new cluster calculation becomes:

A1 =

A2 =

A3 =

A4 =

A5 =

and so on until A35

Then calculate the distance from data 1 to data 35 to the center of the cluster. So the result of the distance calculation from the first and second iterations is that there are similarities in the clustering results between the two, especially in the names of the elementary schools produced. This shows that there is no need for further iteration or repetition. So that in Cluster 1, there are 31 elementary schools that are suitable for rehabilitating school buildings, currently available for Cluster 2, which is not suitable for rehabilitating school buildings, are 4 elementary schools, which can be seen in table 5.

Table 5. Grouping Cluster Members

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Alternative Name | *Cluster 1* | *Cluster 2* | Information |
| 1 | SDN 017127 Sei Nangka | C1 |  | Worthy |
| 2 | SDN 010043 Perk. Air Batu III/IX | C1 |  | Worthy |
| 3 | SDN 010076 Pondok Bungur | C1 |  | Worthy |
| 4 | SDN 010110 Desa Ambalutu | C1 |  | Worthy |
| 5 | SDN 013819 Desa Sei Lama | C1 |  | Worthy |
| 6 | SDN 013830 Desa P. Mahondang | C1 |  | Worthy |
| 7 | SDN 014649 Desa P. Rakyat Tua | C1 |  | Worthy |
| 8 | SDN 014667 Danau Sijabut Gardu | C1 |  | Worthy |
| 9 | SDN 015933 Lobu Rappa | C1 |  | Worthy |
| 10 | SDN 014671 Sentang | C1 |  | Worthy |
| - | - | - | - | - |
| 35 | [UPTD SDN 013842 Sei Silau Tua](https://dapo.dikdasmen.go.id/sekolah/535AD53045AC469DEADE) |  | C2 | Not feasible |

The following is a graph of the K-Means method in clustering the feasibility of rehabilitating elementary school buildings, there are 31 elementary schools that are suitable for building rehabilitation and 4 are not suitable for rehabilitating school buildings. Can be seen in pictures 1.

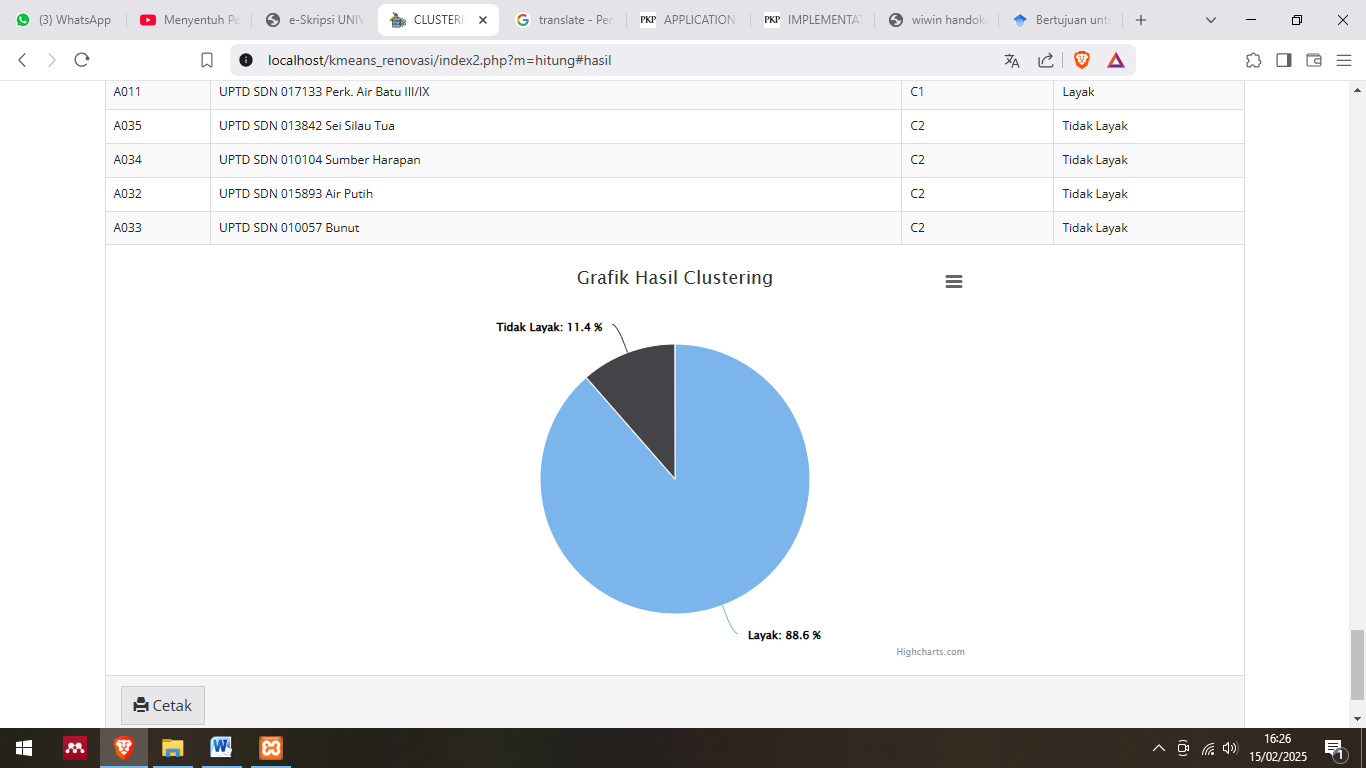


Figure 1. Calculation Results for Rehabilitation and Construction of Elementary School Buildings

**CONCLUSION**

The clustering data mining system accelerates the grouping of data on the feasibility of rehabilitation and construction of elementary school buildings to rehabilitate elementary school buildings in the Asahan District Education Office. The system shows that data mining clustering increases the accuracy of grouping data on the feasibility of rehabilitation and construction of elementary school buildings. The K-Means clustering algorithm in clustering the feasibility of rehabilitating elementary school buildings, there are 31 elementary schools that are suitable for building rehabilitation and 4 are not suitable for rehabilitating school buildings.

**BIBLIOGRAPHY**

[1] S. R. J. Rajagukguk, S. Tumanggor, J. G. Malau, and H. Turnip, “Pentingnya Pemerhatian Sarana dan Prasarana bagi Pendidikan di Sekolah yang Terpencil,” *PEDIAQU J. Pendidik. Sos. dan Hum.*, vol. 2, no. 1, pp. 204–215, 2023.

[2] A. F. Hutauruk *et al.*, “Media Pembelajaran dan TIK,” 2022.

[3] A. Mosepe, “Analisa Tingkat Kerusakan BangunanGedung Sekolah Dasar Di KecamatanPomona Tenggara Kabupaten Poso,” pp. 1–5, 2022.

[4] E. P. A. Yahya and F. A. B. Pradipta, “Asesmen dan analisa gedung eksisting (Studi kasus bangunan Johar Shopping Center Semarang).” Universitas Islam Sultan Agung, 2022.

[5] J. Komputer, “Fakultas ilmu komputer,” *J. Sist. Inf. Telemat.*, vol. 7, no. 1, pp. 1–17, 2013.

[6] P. K. Heatubun and M. Fansyuri, “Analisa Data Mining Untuk Prediksi Penjualan Produk Menggunakan Algoritma FP-Growth Berbasis Web Studi Kasus Online Shop Muslim Galeri,” *BINER J. Ilmu Komputer, Tek. dan Multimed.*, vol. 1, no. 6, pp. 1376–1382, 2024.

[7] D. S. Angreni and M. Susanti, “Implementasi Data Mining Untuk Rekomendasi Kenaikan Pangkat Pegawai Negeri Sipil Menggunakan Algoritma Naïve Bayes Pada Biro Administrasi Pimpinan Sekretariat Daerah Provinsi Sulawesi Tengah,” *Innov. J. Soc. Sci. Res.*, vol. 4, no. 1, pp. 9661–9674, 2024.

[8] A. A. Zulyani, A. S. Y. Irawan, and A. Jamaludin, “Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Menentukan Tingkat Vaksinasi Pada Kecamatan Tambun Selatan,” *Innov. J. Soc. Sci. Res.*, vol. 3, no. 3, pp. 7037–7050, 2023.

[9] J. Informasi, S. Oktarian, and S. Defit, “Klasterisasi Penentuan Minat Siswa dalam Pemilihan Sekolah Menggunakan Metode Algoritma K-Means Clustering,” vol. 2, 2020, doi: 10.37034/jidt.v2i3.65.

[10] D. N. Alfiansyah, V. R. S. Nastiti, and N. Hayatin, “Penerapan Metode K-Means pada Data Penduduk Miskin Per Kecamatan Kabupaten Blitar,” *J. Repos.*, vol. 4, no. 1, pp. 49–58, 2022.

[11] W. Lestari, “Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus: STMIK Bina Bangsa Kendari),” *J. Sist. Inf. dan Sist. Komput.*, vol. 4, no. 2, pp. 35–48, 2019.

[12] N. I. Royanti and B. Ismanto, “Pengelompokkan Keaktifan Peminjaman Buku Di Perpustakaan Stmik Widya Pratama Dengan Metode K-Means,” *J. Chem. Inf. Model.*, vol. 15, no. 1, pp. 81–84, 2020.

[13] T. A. Munandar, “Penerapan Algoritma Clustering Untuk Pengelompokan Tingkat Kemiskinan Provinsi Banten,” *JSiI (Jurnal Sist. Informasi)*, vol. 9, no. 2, pp. 109–114, 2022.

[14] M. Triandini, S. Defit, and G. W. Nurcahyo, “Data Mining dalam Mengukur Tingkat Keaktifan Siswa dalam Mengikuti Proses Belajar pada SMP IT Andalas Cendekia,” *J. Inf. dan Teknol.*, vol. 3, pp. 167–173, 2021, doi: 10.37034/jidt.v3i3.120.

[15] S. N. Arofah and F. Marisa, “Penerapan Data Mining untuk Mengetahui Minat Siswa pada Pelajaran Matematika menggunakan Metode K-Means Clustering,” *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 3, no. 2, pp. 693–702, 2018, doi: 10.31328/jointecs.v3i2.787.

[16] W. Handoko and A. Nasution, “Implementation of Data Mining By Using K-Means To Classify Marriage Age,” *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 9, no. 2, pp. 161–166, 2023, doi: 10.33330/jurteksi.v9i2.2045.

[17] R. Supardi and I. Kanedi, “Implementasi Metode Algoritma K-Means Clustering pada Toko Eidelweis,” *J. Teknol. Inf.*, vol. 4, no. 2, pp. 270–277, 2020, doi: 10.36294/jurti.v4i2.1444.

[18] R. Harum, F. Fathurrohman, and A. R. Rinaldi, “Pemodelan Pola Penjualan Sepatu Impor Melalui Pendekatan Algoritma K-Means,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 4223–4230, 2024.

[19] “Implementasi metode k-means clustering untuk pengelompokan kelas berdasarkan pemahaman siswa pada bimbingan belajar matematika sachio banyuwangi skripsi,” 2022.