J-INTECH (Journal of Information and Technology)
Accredited Sinta 4 Ministry of Higher Education, Science and Technology
Republic of Indonesia SK No. 10/C/C3/DT.05.00/2025
E-ISSN: 2580-720X || P-ISSN: 2303-1425

Analyzing Students' Interest in Mathematics Through the Implementation of the K-Means Clustering Algorithm

Hamdun Sulaiman[1], Ramdhan Saepul Rohman[2], Dede Wintana[3*], Gunawan[4], Muhammad Abdul Ghani[5]

[1,5]Universitas Bina Sarana Informatika,Teknik dan Informatika,Sistem Informasi, jl. Keramat 98 Jakarta Pusat, Indonesia
[2,3,4] Universitas Bina Sarana Informatika, Fakultas Teknik dan Informatika, Informatika Kampus Kota Sukabumi, Jl. Cemerlang No.8 Sukabumi, Indonesia.

Keywords
Educational Data Mining; K-Means Clustering; Mathematics Interest; RapidMiner; Student Performance.

****Corresponding Author:***
Dede.dwe@bsi.ac.id

Abstract
This Research is motivated by the importance of understanding students' interest in mathematics, especially in State Junior High School 193 East Jakarta, considering that mathematics is often considered a difficult and frightening subject for some students. Learning interest, which is defined as the tendency of students to pay attention with a feeling of pleasure, has a significant influence on the process and results of student learning. This study aims to identify the level of student interest in mathematics using the K-Means algorithm. This method is used to group students into several clusters based on their level of interest. The results showed that students were divided into three clusters, namely the first cluster with very high interest totaling 193 students with an average Final Semester Exam score of 91.920, the second cluster with low interest totaling 18 students with an average score of 52.333, and the third cluster with high interest totaling 66 students with an average score of 87.606.

1. Introduction

Interest is an individual's tendency to enjoy and devote full attention to something.[1] Students with a high interest in learning tend to enjoy attending lessons and are able to regulate themselves during the learning process.[2] Students are more likely to enjoy each learning activity when they are actively engaged in the educational process at school.[3] However, not all students have the same level of learning interest.[4] Students with a high interest in learning feel enthusiastic and can manage their behavior to follow lessons effectively, whereas students with low learning interest tend to show a lack of enthusiasm toward learning activities provided by the teacher.[5]

Interest is a key component in school-based learning. Students with a strong interest in learning will direct their behavior toward achieving academic goals and obtaining optimal results.[6] In reality, however, not all students possess a high level of learning interest.[7] This is evident from their attitudes and behaviors that hinder the learning process and reflect a lack of interest in studying.[8]

Mathematics is an integral part of the education system, taught from elementary school through to higher education.[9]The mathematics curriculum includes the delivery of clear and concrete knowledge, the development of thinking skills, and the formation of disciplined reasoning.[10] Mathematics is a basic subject introduced at the elementary school level.[11] It is believed that through mathematics education, students' thinking abilities and logical reasoning can be developed and applied in daily life.[12]

For some students, mathematics may be perceived as a difficult and intimidating subject. Students' interest in mathematics greatly influences their learning process.[13] Therefore, it is important to understand the factors that affect students' interest in learning mathematics.[14] For many students, numbers are often seen as a major obstacle to be avoided. Quite a few students frequently express complaints about mathematics lessons.[15]

The use of the K-Means algorithm in this study aims to cluster students based on their level of interest in mathematics and explore ways to influence that interest. The K-Means algorithm is a clustering analysis method that groups data into similar clusters.[16]To identify students' interest in learning mathematics, the K-Means algorithm can be used to categorize students into various groups based on characteristics or indicators related to their level of interest. Students' opinions are categorized into three groups—good, low, and very low—in relation to their interests and aptitudes.[17]

Although there have been many studies on classification, most of them use supervised learning methods. To date, there is still limited research exploring the use of unsupervised learning. The K-Means algorithm is one of the clustering methods in data mining that falls under the category of Unsupervised Learning. This method is used to group data into one or more clusters based on a partitioning approach centered on the centroid. In this algorithm, the input data does not have class labels, so the clustering process is carried out without prior knowledge of the data's class or category.[18] The advantage of the K-Means algorithm lies in its ability to provide accurate and precise information regarding data grouping.[19] This algorithm is easy to implement and has relatively low time and space complexity.[20] Therefore, it is computationally efficient and capable of producing good and satisfactory output.[21]

2. Research Method

To complete this study, a quantitative secondary research method was employed. Teachers and students at SMPN 193 East Jakarta were interviewed to gather input regarding student interest. The data collected was then processed to extract patterns. In this process, Knowledge Discovery in Databases (KDD) was applied, and the K-Means algorithm was selected as an effective method for clustering large datasets in order to determine the level of students' interest in learning mathematics. The stages of the Knowledge Discovery in Databases (KDD) process are as follows data selection, preprocessing, data transformation, data mining process, and interpretation. The data selected for the data mining process was stored in a separate file from the operational database. The dataset used was the mathematics scores of 7th-grade students at SMPN 193 East Jakarta. At preprocessing stage, duplicate and incorrect data were removed as part of the data preparation procedures. In data trasformation stage, incomplete or undefined data entries were transformed based on selected attributes. The data was then grouped into three interest clusters: *very high interest*, *relatively high interest*, and *relatively low interest* to assess students' learning levels in mathematics. In data mining process stage involved handling the student performance data provided by the school, specifically the scores of 7th-grade students at SMPN 193 East Jakarta. The K-Means algorithm, which combines similar data points into clusters, was applied to process this data. In interpretation stage, students' mathematics score data was tested and analyzed using RapidMiner software.

The research stages in this study, titled *"Determining Students' Interest in Learning Mathematics Using the K-Means Algorithm"*, can be illustrated as follows:

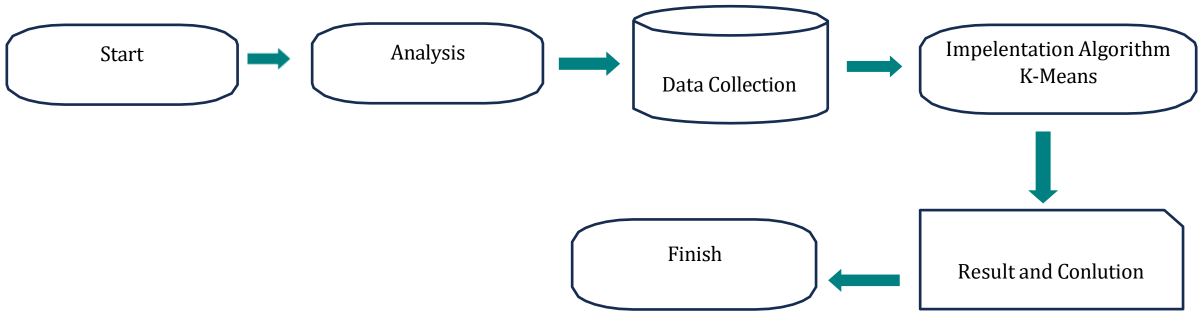

Figure 1. Stap Of the Art

In need analysis phase, the actual problems are identified along with the essential elements needed for the clustering process in determining students' interest in learning mathematics using the *K-Means algorithm.* During the data collection process, the researcher reviewed several references and journals. The academic performance data were obtained from the school for research purposes. In data analysis stage, the provided data was analyzed using the K-Means algorithm, using school-provided data. In this research implementation, the K-Means algorithm was used to classify students' learning interest based on the information provided. During the implementation phase, to identify patterns and group the data based on students' interest in learning. Data testing was conducted to determine students' learning interests, which were divided into three categories: very interested, interested, and less interested. This grouping was carried out using the Cluster Centroids method, which functions to determine the center of each group based on the data distribution pattern.

The knowledge obtained from the clustering results provides an overview of the level of student interest in studying mathematics at school, so that it can be used as a basis for designing more effective learning strategies. To ensure that the number of clusters used is optimal, validation was carried out using the Elbow Method. This method works by analyzing changes in the Sum of Squared Errors (SSE) value, where the elbow point—the point where the curve experiences a sharp change—marks the best number of clusters. In this study, the elbow point indicates that three clusters are the most appropriate choice, because after this number, the decrease in SSE is no longer significant. With this approach, the clustering results obtained have a higher level of accuracy in representing the pattern of students' learning interests in mathematics.

3. Result and Discussions

This research uses the K-Means algorithm to analyze the interest in learning mathematics among 7th- grade students at SMPN 193 East Jakarta. K-Means is an algorithm that uses distance to group data into several clusters. This algorithm is more effective for numerical attributes. This research applies K-Means clustering, with the Elbow Method employed to optimize the clustering process. The elbow method is used to determine the optimal number of clusters in the data grouping process. The details of the steps to be investigated are explained in the form of a diagram.

In data collection section discusses the data used in the research on determining students' interest in learning mathematics. The data consist of 222 entries of 7th-grade student grades provided by SMPN 193 East Jakarta.

Table 1. The Datasets

NO.	STUDENT NAME	CLASS	ASSIGNMENT	MID-SUM	FINAL EXAM	MULTIPLICATION
1	ADHITYA	7A	100	74	96	86
2	ADINDA KANNAYA SAPHIRA	7A	80	74	98	93
3	AIRA ALFATIHANI BANGSAWAN PUTRI	7A	93	74	90	80
4	AISYA DARMA	7A	87	74	75	86
5	ALINIA ZAHFA AZZALIA	7A	100	90	80	80
...	...	...	...	...	...	...
...	...	...	...	...	...	...
222	ZAHIN MAULINA	7F	98	90	90	100

The Elbow Method is applied in K-Means clustering to determine the optimal number of clusters to be formed. The elbow method helps in determining the right number of clusters for a given dataset. To obtain the average distance value within the centroid, the Performance Vector (Performance) operator needs to be configured with appropriate parameters for cluster analysis in RapidMiner. In this context, Performance will be used to evaluate the performance of the clustering model that has been created by considering the distance between data points and their corresponding cluster centers. By using seven clusters, Performance will provide metrics that describe how well the clusters group the data in terms of the average distance to each cluster center.

Table 2. Determining the Average Within Centroid Distance

Jumlah *Cluster*	Avg. within centroid distance
2	572.433
3	**451.571**
4	378.484
5	317.477
6	277.094
7	251.982

The results from Performance will be visualized in the form of a graph, where the most prominent curve represents the elbow method. The elbow method is used to determine the most appropriate number of clusters by identifying the point on the graph where the gradient drop decreases significantly—in other words, where the curve forms an "elbow." The number of clusters corresponding to this elbow point is considered the optimal number of clusters for analyzing the data using the elbow method.

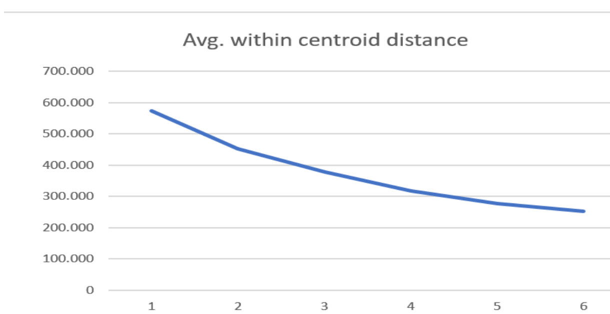

Figure 2. Visualization of Centroid Distance Graph

From the Performance graph visualization, it can be seen that the curve reaches its most prominent point or "elbow" at the value of 1,200,000. Using the elbow method, we determined that the optimal number of clusters for analyzing students' interest in learning mathematics is three clusters. This indicates that the student data can be grouped into three categories based on their interest in learning mathematics.

In this implementation and testing, we used RapidMiner Studio version 9.10, where the data is processed using the K-Means algorithm for 3 clusters. At data processing stage, the clustering process is carried out using the K-Means algorithm to group data based on similarity patterns between attributes, with attributes used for assignments, mid-term assessments, final semester exams, and multiplication. The number of clusters is set at three, according to the characteristics and structure of the data being analyzed. Because the dataset used only consists of numeric attributes, the distance measurement method used is the appropriate distance measure. The selection of this method aims to ensure that the calculation of the distance between data runs optimally, so that the clustering process produces accurate and representative group divisions.

After the data is grouped into each cluster, an evaluation of the quality of the clustering results is carried out using Cluster Distance Performance, which measures how far the data points are spread in each group. This evaluation is important to determine the effectiveness of cluster formation and to see the extent to which the data is distributed homogeneously in each group.

To improve the quality of the clustering that has been formed, optimization is carried out by applying the Maximize parameter. This optimization aims to ensure that the results of data separation do not produce negative values, as well as to increase the average separation between clusters so that the segmentation quality is better.

The results of the clustering process are then visualized in the form of an X-Y curve, where the elbow method is used to determine the most optimal number of clusters. The elbow point is marked by a sharp change in the curve, indicating that the number of clusters used—three—is the best choice based on the distribution and pattern of the data formed.

The result displayed in the Performance output shows that the Average with centroid distance is 451.571, as the number of clusters is set to 3. The output also appears in the ExampleSet under the Model section, including the cluster assignments for each data entry. You can then choose to view the results in terms of *Data*, *Statistics*, or *Visualizations* via the ExampleSet tab, or check the *Description* in the Cluster Model tab.

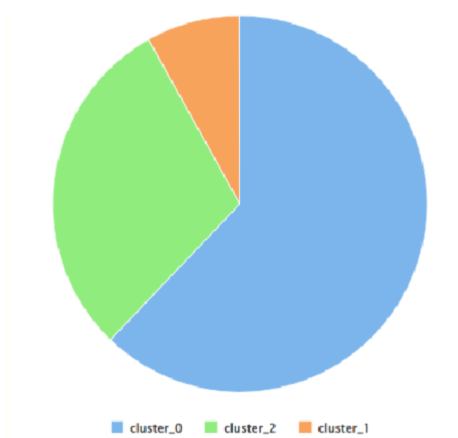

Figure 3. Visualization of Cluster

Based on RapidMiner's analysis of student data from SMPN 193 East Jakarta regarding their interest in mathematics, three clusters were identified. Cluster 0 consists of 193 students with a very high interest in mathematics, with an average UAS (final exam) centroid score of 91.920. This cluster includes students with the highest level of interest in mathematics compared to the other clusters. Cluster 1 consists of 18 students with relatively low interest in mathematics, with an average UAS centroid score of 52.333. Students in this cluster show less enthusiasm for mathematics compared to those in other clusters. Cluster 2 consists of 66 students with relatively high interest in mathematics, with an average UAS centroid score of 87.606. Although not as high as Cluster 0, students in this group still demonstrate a considerable interest in the subject.

4. Conclusions and Future Works

Based on these results, it can be concluded that there is variation in students' interest in mathematics at SMPN 193 East Jakarta. The students can be grouped into clusters that reflect their level of interest in mathematics. The students' interest in mathematics at SMPN 193 East Jakarta can be categorized into three groups: very high interest in Cluster 0 with 193 students, low interest in Cluster 1 with 18 students, and moderate interest in Cluster 2 with 66 students. Future studies are recommended to enhance cluster analysis by applying alternative algorithms such as DBSCAN or Hierarchical Clustering to obtain a more comprehensive comparison of clustering results. Additionally, it is important to include categorical attributes, such as gender and students' social backgrounds, to gain a deeper understanding of the factors influencing students' interest in learning mathematics..

5. References

[1] M. Mahdalena, "PENGARUH MINAT BELAJAR, DUKUNGAN ORANG TUA DAN LINGKUNGAN BELAJAR TERHADAP PERILAKU BELAJAR SISWA DAN HASIL BELAJAR SISWA PADA MATA PELAJARAN IPA (Studi Faktor Yang Mempengaruhi Prilaku Belajar dan Hasil Belajar Siswa Kelas 4,5 dan 6 Pada SDN Binuang 4 da," *KINDAI*, vol. 18, no. 2, pp. 332–351, Oct. 2022, doi: 10.35972/kindai.v18i2.803.

[2] M. Azhar and H. Wahyudi, "Motivasi Belajar: Kunci Pengembangan Karakter dan Keterampilan Siswa," *Uluwwul Himmah Educational Research Journal*, vol. 1, no. 1, 2024.

[3] A. Lufianti, A. Saefudin, and M. K. Umam, "Supervisi Penddikan Di Pondok Pesantren Khozinatul Hikmah Bawu Jepara," *Journal of Learning and Teaching*, vol. 1, no. 1, pp. 5–15, 2024.

[4] I. S. Hidayati, P. O. Putri, and Y. A. Sarumaha, "PENINGKATKAN MINAT BELAJAR MATEMATIKA SISWA KELAS V SD NEGERI PREMBULAN DENGAN MODEL PEMBELAJARAN KOOPERATIF TIPE GROUP INVESTIGATION (GI)," *Intersections : Jurnal Pendidikan Matematika dan Matematika* , vol. 6, no. 2, pp. 30–37, 2021.

[5] J. A. Starosta and B. Izydorczyk, "Understanding the Phenomenon of Binge-Watching—A Systematic Review," *Int J Environ Res Public Health*, vol. 17, no. 12, p. 4469, Jun. 2020, doi: 10.3390/ijerph17124469.

[6] Kumala Wurdaningrum, Sigit Wibowo, Surayanah, and Faisadini Intan Bahari, "PERAN TEMAN SEBAYA TERHADAP MOTIVASI DAN HASIL BELAJAR MATEMATIKA SISWA KELAS 2 SDN KARANGTENGAH KOTA BLITAR," *Jurnal Citra Pendidikan Anak*, vol. 4, no. 1, pp. 15–26, Feb. 2025, doi: 10.38048/jcpa.v4i1.4906.

[7] N. A. Agustin, Kanom, and R. N. Darmawan, "ANALISIS POTENSI WISATA KULINER IKAN BAKAR SEBAGAI DAYA TARIK WISATA DI PANTAI BLIMBINGSARI BANYUWANGI," *Jurnal Inovasi Penelitian*, vol. 1, no. 6, 2020.

[8] S. Safaruddin, M. Mutmainnah, N. Tahir, N. Iftika, and J. Juhaeni, "Pelatihan Sistem Evaluasi Pembelajaran PAI Guna Meningkatkan Minat dan Motivasi Belajar Siswa Sekolah Menengah Atas," *Jurnal Inovasi Penelitian dan Pengabdian Masyarakat*, vol. 3, no. 2, pp. 71–79, Dec. 2023, doi: 10.53621/jippmas.v3i2.268.

[9] P. R. Rani *et al.*, "Pengaruh Metode PJBL Terhadap Hasil Belajar Matematika di Sekolah Dasar," *Journal for Lesson and Learning Studies*, vol. 4, no. 2, pp. 264–270, Sep. 2021, doi: 10.23887/jlls.v4i2.34570.

[10] H. Saputra, "Perkembangan Berpikir Matematis Pada Anak Usia Sekolah Dasar," *JEMARI: Jurnal Edukasi*

Madrasah Ibtidaiyah, vol. 6, no. 2, 2024.

[11] N. Setiani, W. Wakinah, S. Nurazizah, and E. Andriani, “Kesulitan Belajar Matematika di Sekolah Dasar,” *Mutiara: Multidiciplinary Scientifict Journal*, vol. 2, no. 8, pp. 634–638, Aug. 2024, doi: 10.57185/mutiara.v2i8.225.

[12] A. T. Tajuddin and R. Harimukti, “Pengembangan modul berbasis pendidikan matematika realistik untuk meningkatkan kemampuan berpikir kritis siswa SMA,” in *Artikel Pemakalah Paralel*, 2018, pp. 564–567.

[13] Reza Lestari, Habibi, and Syaiful Bastari, “Persepsi Siswa Terhadap Mata Pelajaran Matematika (Studi Kasus Siswa Kelas VI SD Negeri 03 Gumay Ulu),” *JURNAL ILMU PENDIDIKAN*, vol. 3, no. 1, pp. 21–28, Jan. 2024, doi: 10.58222/jurip.v3i1.777.

[14] M. R. Sinta, N. Gusrianti, and I. Artikel, “" ",” *Edukasi Tematik : Jurnal Pendidikan Sekolah Dasar*, vol. 4, no. 20, pp. 1–10, 2024.

[15] W. Handayani, “Permasalahan Guru dan Peserta Didik dalam Pembelajaran Matematika di Sekolah Dasar Negeri 101800 Ujung Batu Jae,” 2024.

[16] N. Rohman and A. Wibowo, “Perbandingan Metode K-Medoids dan Metode K-Means Dalam Analisis Segmentasi Pelanggan Mall,” *SINTECH (Science and Information Technology) Journal*, vol. 7, no. 1, pp. 49–58, Apr. 2024, doi: 10.31598/sintechjournal.v7i1.1507.

[17] B. Alif Prayudha, R. Kurniawan, Y. Wijaya, and U. Hayati, “ALGORITMA K-MEANS UNTUK MENINGKATKAN MODEL KLASTERISASI DATA SISWA SMK SAMUDRA NUSANTARA KABUPATEN CIREBON BERDASARKAN NILAI AKADEMIK,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 1, pp. 1314–1321, Jan. 2025, doi: 10.36040/jati.v9i1.12689.

[18] H. Effendi, A. Syahrial, S. Prayoga, and W. D. Hidayat, “Penerapan Metode K-Means Clustering Untuk Pengelompokan Lahan Sawit Produktif Pada PT Kasih Agro Mandiri,” *Teknomatika*, vol. 11, no. 02, pp. 117–126, Oct. 2021, [Online]. Available: https://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/543

[19] V. P. Ramadhan and G. M. Namung, “Klasterisasi Komentar Cyberbullying Masyarakat di Instagram berdasarkan K-Means Clustering,” *J-INTECH*, vol. 11, no. 1, pp. 32–39, Jul. 2023, doi: 10.32664/j-intech.v11i1.846.

[20] V. R. Sari, E. Buulolo, and Fadlina, “Implementasi Algoritma K-Means dengan Normalisasi Sigmoidal Untuk Klastering Data Ternak Sapi,” *JIKTEKS: Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 2, no. 01, pp. 1–10, 2023, [Online]. Available: https://jurnal.faatuatua.com/index.php/JIKTEKS/article/view/52

[21] C. S. D. B. Sembiring, L. Hanum, and S. P. Tamba, “PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK MENENTUKAN JUDUL SKRIPSI DAN JURNAL PENELITIAN (STUDI KASUS FTIK UNPRI),” *Jurnal Sistem Informasi dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 80–85, Feb. 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2393.