COMPILER
Vol. 13, No. 1, May 2024, pp. 27~38
ISSN: 2252-2403 (Print), ISSN: 2549-2403 (Online)
Accredited Fourth Grade by Kemenristekdikti, Decree No:200/M/KPT/2020
DOI: 10.28989/compiler.v13i1.1962

Recommendation System for Clustering to Allocate Classes for New Students Using The K-Means Method

Yuri Ariyanto[1], Wilda Imama Sabilla[2], Zidan Shabira As Sidiq[3,*]
[1,2,3]Departement of Information Technology, State Polytechnic of Malang, Indonesia

Article Info

Article history:

Received November 23, 2023
Accepted February 19, 2024
Published May 31, 2024

Keywords:

Recommendation System
Clustering System
K-Means
New Student Class Division

ABSTRACT

SMAN 1 Durenan has a plan to organize the allocation of classes for new students using a system to achieve practical and efficient student grouping. The reason for implementing this class allocation system is SMAN 1 Durenan aims to create a new system to process student data for class allocation according to specific needs. This research involves the development of a Recommendation System for Clustering to Allocate Classes for New Students using the K-Means method. The system processes data of newly enrolled students at SMAN 1 Durenan based on specific attributes. The results of this student data processing serve as considerations and references for SMAN 1 Durenan to perform class allocation as needed. The analysis in this research utilizes the K-Means method to obtain data clusters that maximize the similarity of characteristics within each group and maximize the differences between the collections created. The developed recommendation system website provides information about the student data clustering results from the K-Means process at SMAN 1 Durenan.

Corresponding Author:

Zidan Shabira As Sidiq
Departement of Information Technology
State Polytechnic of Malang
Jalan Soekarno Hatta No. 9, Kota Malang, Jawa Timur, Indonesia
Email: *zidanshabira@gmail.com

1. INTRODUCTION

The rapid growth of technology and information in the era of globalization has had a significant impact on all sectors of life. The existence of information technology is a very important requirement for various institutions, including government, industry, and educational institutions [1]. Education plays a very crucial role in the context of national development. The importance of education is not only limited to the individual level but has an impact that covers broader dimensions, such as the social, economic and cultural development of a nation [2]. The use of information and communication technology in the educational environment is used as a tool to access all needs and input data to support the communication process to achieve effective and communicative goals [3].

Data mining is an iterative and interactive process that identifies new patterns or models in massive databases. The goal is to provide significant support in future decision-making by looking for desirable trends or patterns. The process involves using specialized software for valuable and insightful data analysis. The results of data mining can be further analyzed using decision support tools, creating opportunities for deep learning and a comprehensive understanding of the patterns identified, enabling users to make informed decisions that are robust and insightful [4]. To make the process more relevant, clustering is one of the techniques in data mining that can be applied.

The clustering process involves grouping data based on the principle of class similarity and reducing the similarity between classes. A cluster itself can be explained as a group or set of data objects that are like each other within one cluster while being significantly different from objects belonging to other clusters. In this process, objects will be grouped into one or more clusters, creating a high similarity between one object and another object in a particular cluster [5]. In this research, a Clustering Recommendation System for New Student Class Division at SMAN 1 Durenan using the K-Means Method will be made in the hope of improving and creating a teaching and learning process to be more effective and more interactive.

Research conducted by Rozzi Kesuma which examines Clustering on Motorcycle Data using the K-Means method which focuses on improving the application of K-Means through the use of more data and development through integration with feature selection methods to improve the performance of clustering results [6]. In addition, Gustientiedina examined the Application of the K-Means Algorithm for Clustering Drug Data at Pekanbaru Regional Hospital which resulted in clustering of drug data, namely drug groups with low usage, drug groups with moderate usage, and groups with high usage. Another study by Haviluddin on the application of the K-Means Method in Clustering Final Project Recommendations is an important reference for understanding the value relationship between compulsory courses (MKW) and students' final assignments (TA) [7]. In his experiment, Haviluddin managed to identify 3 groups (clusters) of MKW, namely A, B, and C. His research highlights that the K-Means algorithm can be an effective alternative for analyzing the relationship between MKW and TA grades, providing a basis for decisions or policies at the student, lecturer, and study program levels. As a future research direction, it is suggested to explore the comparison and optimization of K-Means algorithm to achieve more diverse accuracy [8].

In this study, the first focus of the researcher is the object of research that is being carried out is the average data of report cards and psychological scores. Second, data clustering is done to divide equally from students who have high, medium, and low scores. Some of the advantages of the research being carried out, the first is that the clustering results are still very much related to the criteria attributes that are carried out by the cluster, the second is that the number of iterations is not limited, and the determination of the initial Centroid is based on an even division of the number of clusters to produce an optimal cluster. So, the purpose of this research is to apply the K-Means method for clustering the grouping of students at SMAN 1 Durenan based on Full Name, Junior High School Origin, Average Value of Junior High School Report Card, Average Value of School Examination, IQ Score, Learning Style Score, Academic Score.

2. RESEARCH METHOD

2.1. Data Collection Technique

In the data acquisition phase of this research, the approach used is a literature review. This method involves exploring literature obtained through document searches, references, books, information from the internet, and other sources. The data used are student admission data from the year 2022, obtained through interviews with SMAN 1 Durenan. The author will use the K-Means algorithm for clustering the data. The clustered data includes the average scores of junior high school reports, the average scores of the junior high school final exam, IQ scores, learning style scores, and academic ability scores. The excerpted data table that has been collected and will undergo the calculation process, can be shown in Table 1.

Table 1 Student Data Cut Table

Full name	From middle school	Middle School Report Card Grades	Middle School Exam Scores	IQ value	Assess Learning Styles	Academic Value
Adinda Dwi Rahmawati	SMPN 2 BANDUNG	91.34	90.47	114	46.15	50.5
Ahmad Alfarizzi	SMPN 3 BANDUNG	90.71	91.07	97	38.46	28.5
Alan Nelsen Febriano	SMPN 1 DURENAN	85.78	88.59	105	46.15	44.5
Amela Bunga Priantika	SMPN 2 BANDUNG	88.72	83.19	112	53.85	66.5
Anggita Novi Anisa	SMPN 1 DURENAN	87.30	86.91	108	38.46	45.5
Anggun Dwi Yunitasari	SMPN 1 BANDUNG	85.34	84.68	118	69.23	79
Ani Dwi Febriana Putri	SMPN 2 BANDUNG	86.70	85.62	112	53.85	65.5
Anita Rahmanda Sari	SMPN 1 DURENAN	89.73	86.67	120	69.23	86
Arfiati Gita Wahyuningsih	SMPN 1 DURENAN	87.20	86.55	115	46.15	67.5
Bunga Violina Eka Yulianti	MTSN 4 TULUNGAGUNG	80.43	77.27	103	38.46	57.5
Dewi Nofita Azalia	SMPN 1 DURENAN	86.49	87.03	117	46.15	71
Diah Ayu Mustika Ninhrum	SMPN 1 CAMPURDARAT	84.17	83.17	121	69.23	88
Dova Febrano Nedilata	SMPN 1 BANDUNG	84.72	83.71	105	38.46	34
Eni Wijayanti	SMPN 1 DURENAN	85.34	85.09	115	46.15	50.5
Ervina Khoirun Nisak	SMPN 1 DURENAN	90.84	91.12	117	53.85	86.5
Fahmi Nasrudin	SMPN 1 WATULIMO	86.19	86.66	101	38.46	47
Fatma Zanita	SMPN 2 WATULIMO	83.45	82.65	112	46.15	51.5

Fika Khiyarotul Ummah	SMPN 2 DURENAN	86.40	86.04	120	69.23	87.5
Flora Marthalena Florenza	SMPN 1 BANDUNG	86.49	86.62	108	46.15	60.5
Gaung Prada Kaffa Angkasa	SMPN 2 DURENAN	87.56	88.93	117	53.85	82.5
Gistria Citra Maharani	SMPN 2 DURENAN	84.31	84.27	115	43.15	80
Gita Khanza Az Zahra	SMPN 1 DURENAN	91.59	90.57	109	38.46	54
Jihan Nurviana	SMPN 1 DURENAN	87.50	86.46	121	69.23	88.5
Jimmy Ramadinnata	SMPN 1 BANDUNG	87.50	82.31	110	46.15	59
Khosh Jawahir	SMP ISLAM DURENAN	88.75	87.22	117	53.85	74.5
Linda Yuliana	MTS JAMM'IYATUL KHAIR	81.98	78.66	113	46.15	62.5
Marcella Elvia Hardiyanti	SMPN 1 POGALAN	87.30	86.43	119	53.85	83.5
Mohammad Hudan Daldiri	SMPN 1 DURENAN	87.54	86.99	97	46.15	37.5
Muhammad Fauziyah Fernanda Putra	MTSN 4 TULUNGAGUNG	86.33	85.54	120	53.85	87
Nabila Alfa Rahma	SMPN 1 DURENAN	84.56	83.69	116	46.15	78.5
Naylatul Lestari	MTSN 4 TULUNGAGUNG	89.82	89.57	115	53.85	66.5
Nela Fatihatussholehah	SMP 2 DURENAN	91.87	91.85	118	69.23	87.5
Putri Zulfiana Anggraini	SMPN 2 DURENAN	87.55	87.46	107	53.85	53.5
Rani Nor Triyani	SMPN 1 DURENAN	86.78	86.02	105	46.15	56
Revanda Marita Erlana	SMPN 1 POGALAN	91.49	90.94	112	53.85	77.5
Revani Evalia Romadhoni	SMPN 1 BANDUNG	83.41	82.96	109	46.15	56.5
Risqi Tri Wahyuni	SMPN 1 POGALAN	90.20	89.78	110	46.15	54
Salfina Salsabila Azalia	SMPN 1 DURENAN	88.41	89.56	121	53.85	88.5
Salsa Putri Berlian	SMPN 1 KAMPAK	84.79	85.02	111	46.15	69.5
Saskia Zannuba Zakkiyah	SMPN 1 POGALAN	89.48	88.16	114	53.85	71.5
Septi Rahmawati	MTSN 4 TRENGGALEK	91.21	90.62	120	53.85	88
Septy Regita Pramesti	MTSN 4 TULUNGAGUNG	85.42	85.81	107	43.15	49
Septyan Arneda Lina	SMPN 1 DURENAN	86.73	85.05	115	46.15	84.5
Syafrina Okti Ramdhani	SMPN 3 WATULIMO	87.91	88.46	112	53.85	79
Syahrila Fratista	SMP 1 DURENAN	85.34	87.56	109	46.15	57
Tria Wahyu Saputri	SMP 1 DURENAN	83.49	82.75	105	53.85	53
Trisa Arinda	SMPN 2 WATULIMO	83.98	84.77	117	53.85	86
Widyana Alfiati Alqurni	SMP ISLAM DURENAN	89.76	89.38	119	53.85	88.5
Wiyana Choirunisa	MTSN 1 TRENGGALEK	84.51	83.95	110	38.46	65.5
Zetta Renata Supono	SMPN 2 DURENAN	90.57	91.67	114	38.46	75.5

2.2. Data Processing Techniques

The data to be clustered is obtained from SMAN 1 Durenan, including student names, junior high school origins, average scores of the junior high school exam, and the results of the Psychological Test data consisting of IQ scores, learning style scores, and academic ability scores. The data needs to be processed to make it more easily usable in the clustering process with the K-Means method. In the image below is the display of the processed data in the database used to store and present data on students confirmed by the system. In this image, there are nine attributes, each serving different functions, which can be shown in Figure 1.

	#	Nama	Jenis	Penyortiran	Atribut	Tak Ternilai	Bawaan	Komentar	Ekstra	Tindakan		
☐	1	**id**	int(20)		UNSIGNED	Tidak	*Tidak ada*			Ubah	Hapus	Lainnya
☐	2	**nama**	varchar(255)	utf8mb4_general_ci		Ya	*NULL*			Ubah	Hapus	Lainnya
☐	3	**jenis_kelamin**	varchar(50)	utf8mb4_general_ci		Ya	*NULL*			Ubah	Hapus	Lainnya
☐	4	**asal_sekolah**	varchar(50)	utf8mb4_general_ci		Ya	*NULL*			Ubah	Hapus	Lainnya
☐	5	**kode**	bigint(20)		UNSIGNED	Tidak	*Tidak ada*			Ubah	Hapus	Lainnya
☐	6	**nilai_sekolah**	float			Tidak	*Tidak ada*			Ubah	Hapus	Lainnya
☐	7	**iq**	int(11)			Tidak	*Tidak ada*			Ubah	Hapus	Lainnya
☐	8	**gaya_belajar**	float			Tidak	*Tidak ada*			Ubah	Hapus	Lainnya
☐	9	**akademik**	float			Tidak	*Tidak ada*			Ubah	Hapus	Lainnya

Figure 1 Data Processing Database

2.3. K-Means Method

K-Means is one of the non-hierarchical data clustering methods that partition existing data into one or more clusters/groups. This method involves partitioning data into clusters so that data with similar characteristics are grouped into the same cluster, while data with different characteristics are grouped into other clusters [9]. The K-Means algorithm is an iterative clustering procedure that divides a predefined set of data into K clusters. Its advantages include the continuity of implementation, high execution speed, ease of adaptation, and popularity in practical usage. The primary goal of this clustering algorithm is to group the entire dataset into relatively homogeneous subgroups or clusters [10].

The primary steps in clustering data using the K-Means algorithm involve the formation of initial Centroid points (cj). Typically, these initial Centroid points are randomly generated, and their quantity is adjusted according to the predetermined number of clusters. Once K Centroid points are established, the next step is to calculate the distance between each data point (Xi) and the Centroids from j to k, symbolized as d(Xi, cj). There are various distance metrics that can serve as indicators of similarity between a data instance, and one of them is the Euclidean distance metric. The data is then assigned to clusters in this method using the principle that data with similar characteristics are grouped into the same cluster [11]. The following are the steps in the K-Means algorithm using the Euclidean Distance equation:

1. Determine the number of clusters (k) in the dataset.
2. Determine the centre point (Centroid) randomly at the initial stage.
3. Calculate each data's closest distance to the Centroid. By using the formula below:

   De = $\boxed{\sqrt{(x_i - y_i)^2 + \cdots + (x_i - y_i)^2}}$

   (1)

   Information:
   De = *Euclidean Distance*
   (x) = object coordinates
   (y) = Centroid coordinates
   i = many objects

4. The next step is to calculate the cluster points again with the current cluster members. The cluster point is the average value of all data in a cluster. Can be calculated using the following formula:

   $\boxed{C_{kj} = \frac{\sum_{i}^{P} = 1\, x_{ij}}{P}}$

   (2)

   Information:
   Xij = into cluster
   P = many objects in cluster

5. The final step is to recalculate each object with the Cluster points (new Centroids). If the cluster calculation shows no further changes, then the clustering calculation is considered complete. However, if there are still changes in the cluster calculation, the calculation is repeated, following steps c until the cluster members no longer move. If there is no difference between the new and old cluster centres, the iteration is stopped.[2]

2.4. System Trial

The system testing is conducted after completing all the design and implementation stages, with the aim of ensuring that all built system functions can operate as expected. The system testing process in this research is carried out with the following steps:

2.4.1. Black Box

Black Box Testing is a software testing method in which testing is conducted without considering the internal details of the system or application. This testing is performed to determine whether a program has met predefined functional standards or not [12]. Black Box testing is a software testing method that focuses on testing the functions of an application related to the software's operational structure. In this research, black box testing will be conducted through a series of test cases. Each test case will involve simulating various functions within the system, excluding the forecasting function. This testing aims to evaluate the extent to which the output of the system's processing aligns with the predefined expectations.

2.4.2. Determining the Optimal Number of Clusters Using the Elbow Method

This research utilizes the Elbow method to determine the optimal number of clusters. The Elbow method considers the comparison of the Sum Square of Error (SSE) values for each cluster value, focusing on the point where adding clusters does not result in a significant decrease in SSE. In other words, as the number of clusters (k) increases, the SSE value decreases. The Elbow method guides determining the best number of clusters by examining the point where adding clusters does not significantly contribute to the reduction in SSE. This approach offers relevant information in determining the optimal number of clusters for data analysis [13]. The Elbow method is a useful approach for determining the optimal number of clusters in the formation of a cluster. This method involves observing the changes in the reduction of cluster error (inertia) with an increasing number of clusters. The point where the reduction in cluster error starts to level off or forms an elbow-like bend, resembling the "elbow" in a human arm, is considered the optimal number of clusters. The Elbow method assists researchers or analysts in determining the most representative and efficient number of clusters for a specific data structure [14]. Having the lowest possible SSE (Sum of Squared Error) value indicates good clustering. The lower the Sum of the Squared Error value, the more optimal it is [15]. The SSE formula is as follows:

$$\boxed{SSE = \sum_{k=1}^{k} \sum_{x1 \in Sk} \| x_1 - C_k \|_2^2} \quad (3)$$

Information:

K = total clusters
Xi = into the data
Ck = beginning of the cluster
n = total data

Elbow Method algorithm for determining the K value in K-Means:

1. Start
2. Initialization of K value
3. Increase the K value
4. Calculate the sum of square errors for each K value
5. Look at the results of the sum of square error of the K value which has dropped drastically
6. Determine the value of K in the form of an angle
7. Finished.

3. RESULTS AND ANALYSIS

Based on the research method employed, the results and discussion in this study begin with the calculation process using the Elbow method to determine the optimal number of clusters. Referring to the calculation results conducted in Excel, the outcomes can be identified. For instance, in trial 1 with a dataset consisting of 50 student data, the SSE (Sum of Square Error) values can be shown in Table 2.

Table 2 SSE Results

Cluster	SSE	Difference
2	975.1968	975.1968
3	567.2453	407.9515
4	422.1423	145.103
5	406.5561	15.58625
6	406.5561	0

3.1. Clustering Test Results

Here are the results of determining the optimal cluster using the Elbow method by conducting tests with 9 clusters. The testing outcomes can be shown in Figure 2.

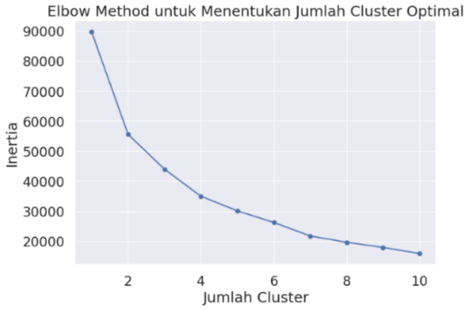

Figure 2 Optimal Cluster Results

From the conducted data trials, it is observed that the largest difference in SSE values in the testing is found in cluster 6. Therefore, the optimal cluster optimization is 6. To determine the largest difference, you can refer to the table number 2 above.

3.2. Blackbox Testing

Black Box testing focuses on the functional specifications of the software. Its purpose is to discover functional errors in the program. Here are the results of the testing using Black Box Testing, can be shown in Table 3.

Table 3 Blackbox Testing

No	Input Data	Hope	System Results	Conclusion
1	Menu Data Murid	Admin can import and CRUD student data	Admin can import data and CRUD can add, change, and delete data in the database	Valid
2	Class Data Menu	Admin can perform CRUD on class data	CRUD can be carried out by admins and can add, change, and delete data in the database	Valid
3	Clustering Menu	Admin can perform calculations to determine clustering of student data	The system displays the results of the data calculation process	Valid

Based on the discussed testing, it can be concluded that black box testing of the application focuses only on the inputs and outputs that inform the conformity of the created application with the specified specifications. The test results indicate that the system can perform CRUD, import data, carrying out the clustering process, and displaying the results of the clustering process.

3.3. Method Analysis

Based on the conducted research, the process of calculating data using the K-Means method involves the following steps, can be shown in Table 4.

1. Determining the number of clusters to be created. In the experiment with 50 data points, a total of 6 clusters will be chosen.
2. Determining the centroid or central point of each cluster randomly. The data is selected randomly, as shown in the table below:

Table 4 Centroid Data

No	Average Report Card	Average School Test Scores	IQ Test Scores	Assess Learning Styles	Academic Ability	Full name
3	85.78	88.59	105	46.15	44.5	Alan Nelsen Febriano
17	83.45	82.65	112	46.15	51.5	Fatma Zanita
30	84.56	83.69	116	46.15	78.5	Nabila Alfa Rahma
38	88.41	89.56	121	53.85	88.5	Salfina Salsabila Azalia
46	83.49	82.75	105	53.85	53	Tria Wahyu Saputri

50	90.57	91.67	114	38.46	75.5	Zetta Renata Supono

From the table above there are 6 centroid data randomly selected by the researcher, the optimal number of centroids is 6 data based on the test results using the Elbow Method which can be seen in Figure 2.

3. Calculate the distance of each data point to the centroid or cluster center using the Euclidean distance equation. Here are the results of the calculations using the Euclidean Distance Equation, can be shown in Table 5.

Table 5 Euclidean Distance Calculation Results

Full name	Distance 1	Distance 2	Distance 3	Distance 4	Distance 5	Distance 6
Adinda Dwi Rahmawati	12.30642109	11.33157094	29.66372869	39.51838813	16.36340124	26.19482773
Ahmad Alfarizzi	20.23836456	30.60621015	54.88554454	66.48625572	31.97581743	49.98379337
Alan Nelsen Febriano	0	11.7776271	36.0901427	47.53049337	13.07247872	33.66848081
Amela Bunga Priantika	25.10166528	17.67355369	15.38978882	24.61042462	16.08715326	19.92950075
Anggita Novi Anisa	8.617940589	12.00442418	35.07152834	47.57212104	18.27358202	31.13439416
Anggun Dwi Yunitasari	43.67399799	36.50519689	23.20613927	19.21024987	32.99532997	32.42307666
Ani Dwi Febriana Putri	23.64227781	16.57327367	15.89322183	25.06885917	14.95964572	19.8098839
Anita Rahmanda Sari	49.9922524	42.92476208	25.30900433	15.93389155	40.06055916	33.44769798
Arfiati Gita Wahyuningsih	25.20273795	17.15437262	11.71107168	23.38405867	19.94352276	12.71642245
Bunga Violina Eka Yulianti	19.72057301	14.63560385	26.97052836	41.67225216	17.33482333	27.48053129
Dewi Nofita Azalia	29.14082531	20.82488896	8.492967679	19.78957554	23.55012527	11.24998222
Diah Ayu Mustika Ninhrum	52.08555366	44.12170894	25.46289261	17.19314107	41.45097345	35.57039359
Dova Febrano Nedilata	13.94001793	20.42357951	46.48023343	59.25319148	24.50074693	43.59855617
Eni Wijayanti	12.18374327	4.418789427	28.06364909	39.60676457	13.20750166	27.49155143
Ervina Khoirun Nisak	44.71347112	37.88987992	14.79639483	5.323391776	37.2871211	19.16318084
Fahmi Nasrudin	9.234668375	14.96575424	35.8868207	48.70770473	17.64738508	32.02396759
Fatma Zanita	11.7776271	0	27.33703898	39.76983405	10.5143521	27.77032409
Fika Khiyarotul Ummah	51.12311904	43.73649506	25.27042738	15.96793349	40.87892611	34.29108776
Flora Marthalena Florenza	16.4129522	11.04547419	20.0077435	32.00968603	12.186751	19.03378575
Gaung Prada Kaffa Angkasa	40.62725686	33.19066887	10.61826728	7.288305702	32.69567708	17.64709041
Gistria Citra Maharani	37.28368678	28.87237434	3.556529207	16.35646967	30.76479807	11.71297144
Gita Khanza Az Zahra	14.25000351	14.26085902	28.37501366	39.77731263	19.50934392	22.12465593
Jihan Nurviana	52.27027549	44.87321027	25.96156582	15.71567689	42.22130505	34.66066791
Jimmy Ramadinnata	16.66273687	8.761740695	20.65909001	33.22545109	11.68604724	21.08128554
Khosh Jawahir	33.37645577	25.73450796	10.3105286	14.75097285	25.57143915	16.43067254
Linda Yuliana	22.38403226	11.83558195	17.23244904	30.97490759	15.24953114	21.72980212
Marcella Elvia Hardiyanti	42.22873429	34.07933245	10.40697843	6.326847556	33.97511589	19.07885217
Mohammad Hudan Daldiri	10.89300693	21.36735126	45.40672197	56.95306664	19.94793473	42.69917329
Muhammad Fauziyah	45.827339	37.419066	12.41351683	4.872042693	37.3744525	21.4629122
Nabila Alfa Rahma	36.0901427	27.33703898	0	15.28297746	28.85426312	13.11245972
Naylatul Lestari	25.70159528	19.53850813	16.32580779	22.84706108	19.20498112	17.99512712
Nela Fatihatussholehah	50.97452403	44.94688866	27.16078239	16.24069272	41.81900047	33.2943734
Putri Zulfiana Anggraini	12.19425274	11.32413794	28.0792272	37.76439593	6.551160203	28.22581443
Rani Nor Triyani	11.82602638	9.575792395	25.25088711	37.23880369	9.476655528	23.80488815
Revanda Marita Erlana	35.14849357	29.47296558	13.29952631	14.60790197	27.9343176	15.69176217
Revani Evalia	14.04684306	5.839323591	23.126941	35.99791661	9.358979645	23.92207767

Romadhoni						
Risqi Tri Wahyuni	11.67058268	10.32711964	26.5546173	37.06470693	13.40652826	23.26145094
Salfina Salsabila Azalia	47.53049337	39.76983405	15.28297746	0	39.83506621	21.53995822
Salsa Putri Berlian	25.9754692	18.23218308	10.38372765	23.53733205	19.34897672	13.48209924
Saskia Zannuba Zakkiyah	29.71809045	23.02196777	12.5090887	18.46902542	22.09973303	16.3205484
Septi Rahmawati	47.01218778	39.74013714	16.05575909	3.195872338	39.64259452	20.75124575
Septy Regita Pramesti	6.411552074	7.356391779	31.07223841	43.51715294	12.14818917	28.88093143
Septyan Arneda Lina	41.39364806	33.38440354	6.599886363	11.59536545	34.16617626	14.13174087
Syafrina Okti Ramdhani	36.09839055	29.48198942	10.46496058	13.14191767	27.87723982	16.4463309
Syahrila Fratista	13.17211069	8.181087947	22.95289306	34.77017256	10.8558095	21.69403374
Tria Wahyu Saputri	13.07247872	10.5143521	28.85426312	39.83506621	0	30.88357007
Trisa Arinda	44.08369767	35.76751739	10.86475034	8.051024779	35.1755668	21.14568987
Widyana Alfiati Alqurni	46.98674813	39.52719317	15.08993373	2.419690063	39.23671495	20.89861957
Wiyana Choirunisa	23.41534967	16.18486021	16.25442094	30.55379191	20.50810815	14.57127311
Zetta Renata Supono	33.66848081	27.77032409	13.11245972	21.53995822	30.88357007	0

The results of the table above are obtained from the calculation process as follows:
Based on the existing data, an example of data number one can be taken for the cluster calculation process (C), where:
Number of Clusters = 6
C1= (85.78, 88.59, 105, 46.15, 44.5)
C2= (83.45, 82.65, 116, 46.15, 51.5)
C3= (84.56, 83.69, 116, 46.15, 78.5)
C4= (88.41, 89.56, 121, 53.85, 88.5)
C5= (83.49, 82.75, 105, 53.85, 53)
C6= (90.57, 91.67, 114, 38.46, 75.5)
Then:
Cluster search for the first data:

$$\sqrt{(91.34-85.78)^2+(90.47-88.59)^2+(114-105)^2+(46.15-46.15)^2+(50.5-44.5)^2}$$
= 12.30642109

$$\sqrt{(91.34-83.45)^2+(90.47-82.65)^2+(114-116,46)^2+(46.15-46.15)^2+(50.5-51.5)^2}$$
= 11.33157094

$$\sqrt{(91.34-84.56)^2+(90.47-83.69)^2+(114-116)^2+(46.15-46.15)^2+(50.5-78.5)^2}$$
= 29.66372869

$$\sqrt{(91.34-88.41)^2+(90.47-89.56)^2+(114-121)^2+(46.15-53.85)^2+(50.5-88.5)^2}$$
= 39.51838813

$$\sqrt{(91.34-83.49)^2+(90.47-82.75)^2+(114-105)^2+(46.15-53.85)^2+(50.5-53)^2}$$
= 16.36340124

$$\sqrt{(91.34-90.57)^2+(90.47-91.67)^2+(114-114)^2+(46.15-38.46)^2+(50.5-75.5)^2}$$
= 26.19482773

The above method is done for each data until the last data, resulting in the data distance as shown in Table 5.

4. Grouping each data based on the nearest distance between the data and its centroid or cluster center. The table below displays the results of the minimum distance calculations along with the assigned cluster groups, can be shown in Table 6.

Table 6 Minimum Value and Cluster

No	Min Value	Cluster	No	Min Value	Cluster	No	Min Value	Cluster	No	Min Value	Cluster
1	11.33157094	2	14	10.3105286	3	27	11.83558195	2	40	10.38372765	3
2	20.23836456	1	15	4.418789427	2	28	6.326847556	4	41	12.5090887	3
3	0	1	16	5.323391776	4	29	10.89300693	1	42	3.195872338	4
4	15.38978882	3	17	9.234668375	1	30	4.872042693	4	43	6.411552074	1
5	8.617940589	1	18	0	2	31	0	3	44	6.599886363	3
6	19.21024987	4	19	15.96793349	4	32	16.32580779	3	45	10.46496058	3
7	14.95964572	5	20	11.04547419	2	33	16.24069272	4	46	8.181087947	2
8	15.93389155	4	21	7.288305702	4	34	6.551160203	5	47	0	5
9	11.71107168	3	22	3.556529207	3	35	9.476655528	5	48	8.051024779	4
10	14.63560385	2	23	14.25000351	1	36	13.29952631	3	49	2.419690063	4
11	8.492967679	3	24	15.71567689	4	37	5.839323591	2	50	10.38372765	3
12	17.19314107	4	25	8.761740695	2	38	10.32711964	2			
13	13.94001793	1	26	11.33157094	2	39	0	4			

5. Determining the new centroid point. The calculation is done by calculating the average value of each cluster member. The results of the value of the new centroid point can be seen in the table 7 below:

Table 7 New Centroid

No	Average Report Card	Average School Test Scores	IQ Test Scores	Assess Learning Styles	Academic Ability	Full name
3	87.40625	87.53875	103.625	40.96875	42.5	Alan Nelsen Febriano
17	85.548	84.337	110.3	45.381	55.95	Fatma Zanita
30	87.52083	86.59583	114.25	49.75	73.875	Nabila Alfa Rahma
38	87.88571	87.51571	119.1429	60.44143	86.21429	Salfina Salsabila Azalia
46	86.13	85.4625	107.25	51.925	57	Tria Wahyu Saputri
50	87.54	87.81	112	38.46	70.5	Zetta Renata Supono

The results in table 7 above are obtained by calculating the average value of each cluster member, such as the following example of data in cluster 1:

Table 8 Data Cluster 1

Name	Average Report Card	Average School Test Scores	IQ Test Scores	Assess Learning Styles	Academic Ability
Ahmad Alfarizzi	90.71	91.07	97	38.46	28.5
Alan Nelsen Febriano	85.78	88.59	105	46.15	44.5
Anggita Novi Anisa	87.30	86.91	108	38.46	45.5
Dova Febrano Nedilata	84.72	83.71	105	38.46	34
Fatma Zanita	83.45	82.65	112	46.15	51.5
Khosh Jawahir	88.75	87.22	117	53.85	74.5
Muhammad Fauziyah Fernanda Putra	86.33	85.54	120	53.85	87
Septyan Arneda Lina	86.73	85.05	115	46.15	84.5
Cluster Mean Value	**87.40625**	**87.53875**	**103.625**	**40.96875**	**42.5**

The result of the average value of cluster 1 members is used as the new centroid value.

6. After obtaining the values of the latest centroid points, compare these values with the previously determined centroid points. If there is a change in the centroid values, the process returns to the step of calculating the distance of each data point to the centroid. However, if there is no change in the centroid values, the process stops or concludes. The table below shows the process stopping after iterations up to the 4th iteration and produces the following numbers, can be shown in Table 8.

Table 9 Clustering Results

No	Full Name	Cluster	No	Full Name	Cluster	No	Full Name	Cluster	No	Full Name	Cluster
1	Adinda Dwi Rahmawati	2	14	Eni Wijayanti	2	27	Marcella Elvia Hardiyanti	4	40	Saskia Zannuba Zakkiyah	3
2	Ahmad Alfarizzi	1	15	Ervina Khoirun Nisak	4	28	Mohammad Hudan Daldiri	1	41	Septi Rahmawati	4
3	Alan Nelsen Febriano	1	16	Fahmi Nasrudin	1	29	Muhammad Fauziyah Fernanda Putra	4	42	Septy Regita Pramesti	2
4	Amela Bunga Priantika	3	17	Fatma Zanita	2	30	Nabila Alfa Rahma	3	43	Septyan Arneda Lina	3
5	Anggita Novi Anisa	1	18	Fika Khiyarotul Ummah	4	31	Naylatul Lestari	3	44	Syafrina Okti Ramdhani	3
6	Anggun Dwi Yunitasari	4	19	Flora Marthalen a Florenza	2	32	Nela Fatihatussholehah	4	45	Syahrila Fratista	2
7	Ani Dwi Febriana Putri	3	20	Gaung Prada Kaffa Angkasa	4	33	Putri Zulfiana Anggraini	5	46	Tria Wahyu Saputri	5
8	Anita Rahmanda Sari	4	21	Gistria Citra Maharani	3	34	Rani Nor Triyani	2	47	Trisa Arinda	4
9	Arfiati Gita Wahyunings ih	3	22	Gita Khanza Az Zahra	2	35	Revanda Marita Erlana	3	48	Widyana Alfiati Alqurni	4
10	Bunga Violina Eka Yulianti	2	23	Jihan Nurviana	4	36	Revani Evalia Romadhoni	2	49	Wiyana Choirunisa	2
11	Dewi Nofita Azalia	3	24	Jimmy Ramadinn ata	2	37	Risqi Tri Wahyuni	2	50	Zetta Renata Supono	3
12	Diah Ayu Mustika Ninhrum	4	25	Khosh Jawahir	3	38	Salfina Salsabila Azalia	4			

13	Dova Febrano Nedilata	1	26	Linda Yuliana	2	39	Salsa Putri Berlian	3

From the results of calculations using the K-Means method in the table above, namely using 50 student data and 6 optimal clusters, it produces the following student grouping data:

Table 10 Number of Each Cluster

Cluster	Amount
1	6
2	14
3	14
4	14
5	2
6	0

Based on the research conducted, the use of the K-means method for clustering problems is considered less suitable because this method has a weakness in determining the initial center of the cluster or centroid which results in the results of the clusters formed depending on the determination of the cluster center value, this makes it very difficult to obtain clear and unique initial centroid results. According to [16] the K-Means method has a weakness when applied to clustering problems where it must be iterated continuously until results are obtained that do not change significantly and need to use many parameters to obtain more optimal results.

According to some previous studies, clustering problems can be overcome by using the K-Nearest Neighbors method. This method is widely implemented for clustering problems because it can classify new objects based on the attributes used where the results of testing new samples will be classified based on the K-Nearest Neighbors category [17]. The K-Nearest Neighbors algorithm has the advantage that the way this algorithm works is simpler than other clustering methods and easy to understand [18].

4. CONCLUSION

Based on the results of the research and discussion carried out, it is concluded that the system can apply the K-means method for grouping or clustering student data using attributes such as name, school origin, average junior high school report card score, average junior high school exam score, iq score, learning style score and academic score. In making the system that has been done, it turns out that the system can only perform calculations and display the results of the clustering process as can be seen in table 10 which contains the final results of the clustering process, namely cluster 1 contains as much as 6 data, cluster 2 contains 14 data, cluster 3 contains 14 data, cluster 4 contains 14 data, cluster 5 contains 2 data, and cluster 6 contains 0 data, so the system still cannot divide or classify into each class because the data for each cluster is still uneven. According to several previous studies, clustering problems can be overcome by using the K-Nearest Neighbors method.

ACKNOWLEDGEMENTS

We thank the Department of Information Technology State Polytechnic of Malang, my supervisor, the Research Institute at SMAN 1 Durenan, and everyone else who made this study possible.

REFERENCES

[1] Z. Nabila *et al.*, "Sistem Informasi Kearsipan Menggunakan Framework Laravel (Studi Kasus: Prodi Sistem Informasi Universitas Peradaban)," *INFORMAL Informatics J.*, vol. 2, no. 1, pp. 53–57, 2021, doi: 10.19184/isj.v5i1.17071.

[2] M. S. Fauzi and S. Samsudin, "Smart School Berbasis Web Interaktif di SD Swasta Amaliyah Sunggal dengan Algoritma K-Means Cluster," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 3, pp. 332–341, 2022, doi: 10.32736/sisfokom.v11i3.1479.

[3] M. A. Saputra and Soedjarwo, "Implementasi sistem informasi manajemen berbasis aplikasi mobile pada jenjang sma," *J. Inspirasi Manaj. Pendidik.*, vol. Vol. 09, no. No. 02, pp. 361–376, 2021.

[4] E. D. Sikumbang, "Penerapan Data Mining Dengan Algoritma Apriori," *J. Tek. Komput. AMIK BSI*, vol. 9986, no. September, pp. 1–4, 2018.

[5] F. Indriyani and E. Irfiani, "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means," *JUITA J. Inform.*, vol. 7, no. 2, p. 109, 2019, doi: 10.30595/juita.v7i2.5529.

[6] R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," *INFORMAL Informatics J.*, vol. 5, no. 1, p. 10, 2020, doi: 10.19184/isj.v5i1.17071.

[7] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-

Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
[8] H. Haviluddin, S. J. Patandianan, G. M. Putra, N. Puspitasari, and H. S. Pakpahan, "Implementasi Metode K-Means Untuk Pengelompokkan Rekomendasi Tugas Akhir," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 16, no. 1, p. 13, 2021, doi: 10.30872/jim.v16i1.5182.
[9] S. Sukamto, I. D. Id, and T. R. Angraini, "Penentuan Daerah Rawan Titik Api di Provinsi Riau Menggunakan Clustering Algoritma K-Means," *JUITA J. Inform.*, vol. 6, no. 2, p. 137, 2018, doi: 10.30595/juita.v6i2.3172.
[10] M. Syahril, S. Kusnasari, A. Muhazir, and A. Syahputri, "Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD Implementasi Data Mining Untuk Rekomendasi Jurusan Menggunakan Algoritma K-Means Clustering Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD," *Teknol. Sist. Inf. dan Sist. Komput. TGD*, vol. 6, pp. 235–245, 2023.
[11] R. F. Saputra, Y. Agus Pranoto, and R. Primaswara P., "Implementasi Metode K-Means Clustering Pada Tes Psikologi Untuk Menentukan Kelompok Belajar Siswa Berbasis Mobile," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 1, pp. 328–333, 2021, doi: 10.36040/jati.v5i1.3290.
[12] T. S. Jaya, "Pengujian Aplikasi dengan Metode Blackbox Testing Boundary Value Analysis (Studi Kasus: Kantor Digital Politeknik Negeri Lampung)," *J. Inform. J. Pengemb. IT*, vol. 3, no. 1, pp. 45–48, 2018, doi: 10.30591/jpit.v3i1.647.
[13] A. W. Fuadah, F. N. Arifin, and O. Juwita, "Optimasi K-Klasterisasi Ketahanan Pangan Kabupaten Jember Menggunakan Metode Elbow," *INFORMAL Informatics J.*, vol. 6, no. 3, p. 136, 2021, doi: 10.19184/isj.v6i3.28363.
[14] E. Muningsih and S. Kiswati, "Sistem Aplikasi Berbasis Optimasi Metode Elbow Untuk Penentuan Clustering Pelanggan," *Joutica*, vol. 3, no. 1, p. 117, 2018, doi: 10.30736/jti.v3i1.196.
[15] I. Wahyudi, M. B. Sulthan, and L. Suhartini, "Analisa Penentuan Cluster Terbaik Pada Metode K-Means Menggunakan Elbow Terhadap Sentra Industri Produksi Di Pamekasan," *J. Apl. Teknol. Inf. dan Manaj.*, vol. 2, no. 2, pp. 72–81, 2021, doi: 10.31102/jatim.v2i2.1274.
[16] C. Kamila, M. Adiyatma, G. R. Namang, R. Ramadhan, F. Syah, and D. Redaksi, "Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik," *J. Intech*, vol. 2, no. 1, pp. 1–6, 2021.
[17] M. D. Alkhussayid and F. Ferdiansyah, "Implementasi Algoritma K-Nearest Neighbors Pada Penentuan Jurusan Siswa," *J. Sist. Komput. dan Inform.*, vol. 4, no. 1, p. 25, 2022, doi: 10.30865/json.v4i1.4759.
[18] D. Chasanah, A. Siregar, and ..., "Klasifikasi Kelayakan Siswa dalam Menentukan Kelas Unggulan Menggunakan Algoritma K-Nearest Neighbor," *... Student J. ...*, vol. III, pp. 51–58, 2022, [Online]. Available: https://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/421%0Ahttps://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/download/421/335.