

---

## **Determining Superior Classes Based on Academic Grades at SMK Karya Pembaharuan with the K-Means Clustering Method**

**Lydia Diffani Siregar<sup>1)\*</sup>, Arif Susilo<sup>2)</sup>, Arif Tri Widiyatmoko<sup>3)</sup>**

<sup>1)2)3)</sup>Universitas Pelita Bangsa, Bekasi, Indonesia

<sup>1)</sup>[lydiadiffanii@email.com](mailto:lydiadiffanii@email.com), <sup>2)</sup>[arif.susilo@pelitabangsa.ac.id](mailto:arif.susilo@pelitabangsa.ac.id), <sup>3)</sup>[arif.tri@pelitabangsa.ac.id](mailto:arif.tri@pelitabangsa.ac.id)

---

### **ABSTRACT**

In the educational environment, k-means clustering can help schools find the best class based on students' academic scores. By grouping students based on academic scores, schools can more easily identify groups of students who have high, medium, and low academic scores. Then the research used is All objects in one cluster have the same characteristics, but each cluster has different characteristics. Novi and Ade Mubarok wrote a journal in 2021 entitled "Application of the K-Means Algorithm to Determine Superior Classes at Junior High School Pelita Bandung" which concluded that Junior High School Pelita Bandung needed 3 clusters. After the researchers conducted experiments, they could produce 3 clusters, namely cluster 0 is the cluster with the lowest average score which will be entered into class C as many as 42 students, in cluster 1 with a medium average score will be entered into class B as many as 37 students, while in cluster 3 with the highest average score will be entered into class A as many as 40 students. The results of this study show that there are 6 students in the high category, 24 students in the moderate category, and 14 students in the low category. Evaluation of the clustering results shows quite good results, with a Davies Bouldin Index (DBI) value of 1,180, which is close to 0.

**Keywords:** K-Means, Clustering, Data mining, Academic Grades, Featured Classes.

---

### **INTRODUCTION**

Various sectors, including education, are affected by the rapid growth of information technology. Applications that assist learning and school management are among the educational innovations (Hasugian & Sagala, 2022).

The KDD (Knowledge Discovery in Database) technique includes data mining. The process of identifying patterns in data is another name for KDD. With the use of data mining, this process can be categorized, forecasted, estimated, and new information can be obtained from a sufficiently big data set that is supported by the K-Means Clustering Algorithm (Putra Primanda et al. 2021).

Based on student academic performance, to help educational institutions find outstanding classes, the k-means clustering method can be used. Schools can more quickly identify groups of children with high, medium, and poor academic scores by classifying pupils according to their academic performance (Saputra & Pakereng, 2023).

The goal of the K-means clustering technique is to separate a data set into many groups, or clusters, according to shared attributes. Because of its ease of use and capacity to handle massive volumes of data in a reasonably quick and effective manner, this approach is widely used (Sulistiyawati & Supriyanto, 2020).

Understanding this classification might encourage pupils to enhance their academic achievement in more suitable ways. The goal of this research is to give educational policymakers and schools a strong basis on which to build better learning strategies, identify kids who require extra help, and allocate educational resources as efficiently as possible (Prasetyo Aji, 2024).

The k-means Algorithm Clustering method is more scalable and efficient for processing very large objects because it is very accurate in terms of object size. This research uses the method to determine the results of cluster grouping of outstanding students based on the current problems, thus using the application of Data Mining to make it easier for the school curriculum to select students to enter superior classes (Bella Bernissa, 2020).

### **LITERATURE REVIEW**

In a previous study titled "Application of K-Means and Fuzzy C Means Algorithms in Grouping Student Achievements Based on Academic Grades," Data mining was used to aggregate student successes based on academic grades using the K-Means and Fuzzy C algorithms. While the Fuzzy C-Means algorithm yields 48 outstanding students, 29 moderately accomplished students, and 28 non-outstanding students, manual calculations and the K-Means algorithm Python program yield 52 outstanding students, 25 moderately achieved students, and 28 non-

\* Corresponding author



outstanding students (Kurniawan et al., 2023).

According to the findings of the study, Application of the K-Means Algorithm to Determine Superior Classes at Junior High School Pelita Bandung, 119 data points were collected and split into three clusters. Of these, 42 students in class C were in the first cluster with the lowest average, 37 students in class B were in the second cluster with a medium average, and 40 students in class A were in the third cluster with the highest average value (Novi & Ade, 2021).

To apply the K-Means method in Rapidminer, data about product stock, especially initial stock, sold stock, and final stock, are entered into a database. Next, this data is connected to the Rapidminer tool, and the K-Means method is processed and formed according to the research entitled "Application of Data Mining K-Means Clustering Method for Sales Analysis at Banten Fashion Hijab Shop." (Nurajizah & Salbinda, 2021).

The results of applying k-means to a sample of 120 data from the report card scores of SMK Raksana 2 Medan students were published in a journal titled "Mapping Featured Class Students Using the K-Means Clustering Algorithm." By entering the centroid value and searching for the closest value, three clusters were created: cluster 1 produced 10 students who were proficient in Android programming, cluster 2 produced 62 students who were proficient in web programming, and cluster 3 produced 48 students who were proficient in desktop programming (Hutagalung et al., 2022).

The application of the k-Means Clustering method to analyze student academic grades based on knowledge and skills is one of the findings and discussions in the previous journal, "Implementation of K Means Clustering for Analysis of Student Academic Grades Based on Knowledge and Skills Values." According to this study, schools can choose the best teaching strategy by classifying pupils according to their academic standing (Oktario Dacwanda & Nataliani, 2021).

The technique for using the K-Means algorithm to identify superior clustering classes is discussed in a journal article entitled "Academic Information System and Determination of Superior Classes with K-Means Algorithm at Junior High School Negeri 3 Ubud." The system will process student grades using the K-Means technique, grouping them based on the value that is closest to the cluster center point. The superior class can be identified more quickly and effectively with this system (Kusuma & Aryati, 2019).

## METHOD

The use of the k-means clustering approach in superior class grouping is investigated by the author. The academic score information of SMK Karya Pembaharuan's grade XI pupils from the 2023–2024 academic year is the focus of this study. The aim of this study is to determine how student academic score data is used to categorize pupils into superior classrooms.

### K-Means

With the growth in the amount of data available, clustering algorithms have become increasingly prominent. K-Means is a straightforward clustering analysis technique that seeks to identify the most effective way to organize the nth entity into clusters. The following procedures are used to cluster using the k-means algorithm (Kesuma Dinata et al., 2020):

- a. Determine the number of clusters.
- b. In determining the centroid value, it is done randomly. Meanwhile, if determining the centroid value which is a stage of iteration, the following formula is used:

$$\bar{V}_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_1} x_{kj} \quad (1)$$

- $V_{ij}$  = Centroid/mean of the i-th cluster for the j-th variable  
 $N_i$  = The amount of data that is a member of the i-th cluster  
 $ik$  = Index of the cluster  
 $j$  = Indeks of the variable  
 $X_{kj}$  = The value of the kth data in the cluster for the jth variable

- c. Calculate the distance between the centroid point and the point of each object. You can use geometric distances, ie:

$$D_e = \sqrt{(x_1 - s_1)^2 + (y_1 - t_1)^2} \quad (2)$$

$D_e$  = Euclidean Distance

$i$  = number of objects

\* Corresponding author



(x, y) = object coordinates  
(s, t) = centroid coordinates

- d. Data clustering is based on the results of calculating the distance of data to the centroid. To identify cluster members, the minimum distance of objects is required. In the distance matrix, the data membership value is 0 or 1. A value of 1 for data allocated to a particular cluster and a value of 0 for data allocated to another cluster.
- e. Finds the new centroid from the clustering result.
- f. Recalculate use the new cluster center to measure the distance of each data. Repeat until the centroid value and cluster members do not move to another cluster.

**RESULT**

The data used main data used in this research comes from SMK Karya Pembaharuan school. This data consists of 44 academic grade lists of eleventh grade (XI) students during the odd semester of the 2023/2024 school year. The variables used include student names and grades for each subject that students study during one semester.

Table 1  
Dataset

No	Nama	Pabp	Pkn	B.Ind	Mtk	B.Ing	Pjok	Int	Kls
1	Aditia Maulana	70	65	75	76	72	67	74	Tkj
2	Ahmad Jahid	79	65	75	76	72	66	74	Tkj
3	Andri Maulana	70	64	72	75	72	65	75	Tkj
4	Apriliyanti	79	80	84	78	85	85	86	Tkj
...	.....	...	....	...	.....	....	....	....	.....
...	.....	...	....	...	.....	....	....	....	.....
40	Bunga Maheza Luna	78	80	81	78	83	76	79	Ak
41	Cendy Fauziya	79	70	81	79	79	83	83	Ak
42	Cindy Lustianti	79	85	88	90	83	76	86	Ak
43	Citra	79	71	83	78	79	85	78	Ak
44	Dini Anggraini	79	85	83	88	85	90	83	Ak

**Determining the Number of Clusters**

Number of groups is the number of groups to be created (Adam Rifais & Laksana, 2024). In this study, three clusters will be formed, so the value of k = 3. The resulting clusters are: cluster 0 (C0) for the high category, cluster 1 (C1) for the moderate category, and cluster 2 (C2) for the low category

**Determining the Centroid**

The initial step of k-means algorithm is randomly initiate k centroids. The number of groups should be has been determined by user (Sudrajat et al., 2022). A student's grade is randomly selected as following table shows the center point of the first cluster:

Table 2  
Determining the Random Centroid

Centroid	Pabp	Pkn	B.Ind	Mtk	B.Ing	Pjok	Int
C0	82	79	86	80	84	88	83
C1	82	80	74	78	78	85	80
C2	82	80	82	80	84	86	78

**Calculating Data Distance to Clustering Center**

The value used in the Euclidean Distance calculation is from the student dataset and the random centroid table, which

\* Corresponding author



is like the example below.

Where is the distance of the first data distance to cluster center 0 (C0):

$$d_0 = \sqrt{\frac{(70 - 82)^2 + (65 - 79)^2 + (75 - 86)^2 + (76 - 80)^2 + (72 - 84)^2 + (67 - 88)^2 + (74 - 83)^2}{}} = 33,80828 \quad (3)$$

Distance of the first data to the cluster center 1 (C1):

$$d_1 = \sqrt{\frac{(70 - 82)^2 + (65 - 80)^2 + (75 - 74)^2 + (76 - 78)^2 + (72 - 78)^2 + (67 - 85)^2 + (74 - 80)^2}{}} = 27,74887 \quad (4)$$

Distance of the first data to the cluster center 2(C2):

$$d_2 = \sqrt{\frac{(70 - 82)^2 + (65 - 80)^2 + (75 - 82)^2 + (76 - 80)^2 + (72 - 84)^2 + (67 - 86)^2 + (74 - 78)^2}{}} = 30,90307 \quad (5)$$

From the calculation of the closest distance (*min*) for each data to each cluster using the *Euclidean Distance* equation, it can be determined that C1 is the closest distance with a value of 27.74 which is referred to as cluster 1.

**Testing Process Using RapidMiner**

Classifying student grade data at SMK Karya Pembaharuan using the RapidMiner application and the research objective is the K-Means method. This is where you import data into RapidMiner, the initial step in the Rapidminer application.

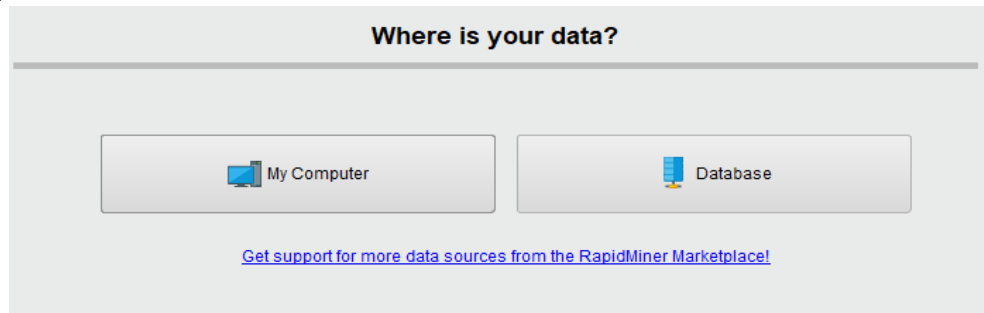


Fig. 1 Import Data

In this case the author imports data in \*xls format. If the imported data has no problems, the data import process can be continued until it is finished and successfully imported into rapidminer.

**Application of K-Means Algorithm**

The next k-means steps method for the data clustering process. First choose the operator that will be used. K-means is the operator employed in this study. First select the operator to be used. K-means is the operator employed in this study.

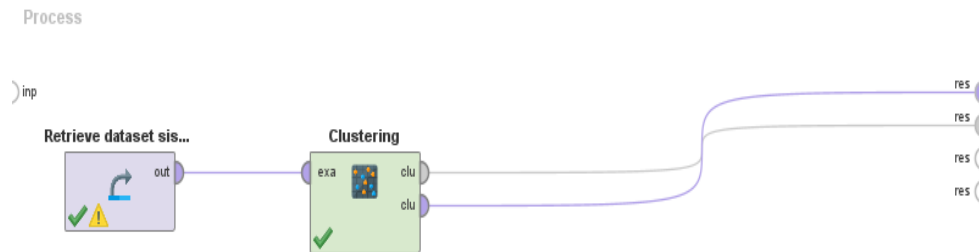


Fig. 2 Applying K-means

**Clustering Results on Rapidminer**

Here the author enters the value of 3 because in this study grouping data into 3 clusters. After that, running the process is done to get the results generated by the k-means algorithm which is has been applied to rapidminer. Here are the results of the clusrtering of imported data.

\* Corresponding author

Row No.	id ↑	cluster	PABP	PKN	B.IND	MTK	B.ING	PJOK	INT
7	7	cluster_1	70	73	75	74	71	69	75
8	8	cluster_2	79	78	81	76	84	86	84
9	9	cluster_1	70	60	71	74	68	72	74
10	10	cluster_1	70	64	76	76	70	68	72
11	11	cluster_2	82	79	86	80	84	88	83
12	12	cluster_2	79	76	78	79	82	88	80
13	13	cluster_0	74	78	76	77	82	83	66
14	14	cluster_0	74	75	74	77	78	75	66
15	15	cluster_2	82	80	74	78	78	85	80
16	16	cluster_2	78	77	76	79	80	83	82
17	17	cluster_1	70	72	76	77	76	68	75
18	18	cluster_0	70	77	78	78	80	85	66
19	19	cluster_2	78	80	77	77	80	87	80
20	20	cluster_2	79	74	73	77	80	88	82
21	21	cluster_0	70	74	77	78	79	89	75

ExampleSet (44 examples, 2 special attributes, 7 regular attributes)

Fig. 3 Clustering Result

Next, the author wants to shows the output of the cluster model above. The results are as follows.

## Cluster Model

```
Cluster 0: 15 items
Cluster 1: 8 items
Cluster 2: 21 items
Total number of items: 44
```

Fig. 4 Number Of Cluster Divisions

In the picture above, it is involved that the results of the data processing provide the results of the process into data divided into three clusters, namely clusters 0, 1, and 2. With the number of cluster 0 as much as 15 data, cluster 1 as much as 8 data and cluster 2 which is 21 data from a total of all data amounting to 44 student data.

Attribute	cluster_0	cluster_1	cluster_2
PABP	74.933	71.125	79.429
PKN	75.867	67.375	78.619
B.IND	77.600	74.875	80.667
MTK	77.933	75	79.381
B.ING	77.800	71.625	81.190
PJOK	81.667	67.750	83.524
INT	72.733	74.125	81.190

Fig. 5 Centoid Table View

At this point, the center point value of each cluster is displayed. By measuring this value, it will be used as a reference for calculations for each dataset with each cluster center point.

\* Corresponding author



## PerformanceVector

```
PerformanceVector:  
Avg. within centroid distance: 65.345  
Avg. within centroid distance_cluster_0: 58.658  
Avg. within centroid distance_cluster_1: 52.734  
Avg. within centroid distance_cluster_2: 74.925  
Davies Bouldin: 1.180
```

Fig. 6 Vector Performance Results

In this last stage is the display of the Davies Bouldin Index values and how small they are value, better cluster collected from clustering clustering method. The decision of calculations This value of 1,180 numbers is displayed with the K-Means algorithm has quite good results because it is close to 0.

The visualization results of the clustering results in the form of a diagram are as follows:

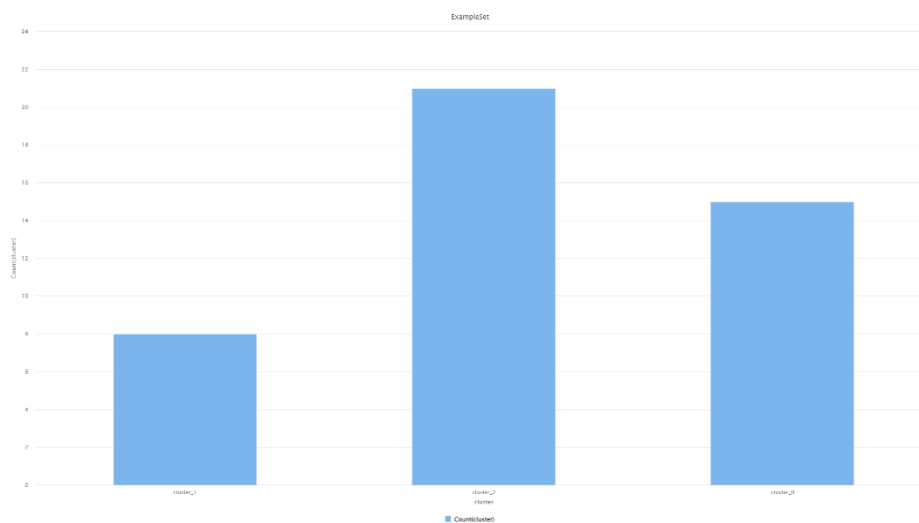


Fig. 7 Visualization Results

### DISCUSSIONS

A dataset of SMK Karya Pembaharuan grade XI students from the 2023–2024 academic year is used in this investigation. After then, the score dataset is divided into three categories low, medium, and high. Based on student academic grades, this study determines the top class using the K-mens clustering algorithm. according to aforementioned test results, it can be inferred that, of the three clusters, C0 (students in the low category), C1 (students in the moderate category), and C2 (students in the high category) comprise eight, eight, and twenty-eight students, respectively, based on clustering and testing with a fast breaking machine application on student grade data.

### CONCLUSION

It is easy to group students based on superior classes automatically using K-Means clustering technique, according to research that has been done using this method to group students based on academic grades. The results of three clusters of forty-four students demonstrate that SMK Karya Pembaharuan has a suitable method with k-means clustering, as evidenced by the Davies Bouldin Index (DBI) test results, which are close to zero with a result of 1,180.

\* Corresponding author





**REFERENCES**

- Adam Rifais, & Laksana, T. G. (2024). Penerapan Algoritma K-Means Clustering Untuk Mengetahui Kemampuan Akademik Siswa Berbasis WEB. *Jurnal Ilmiah Multidisiplin*, 1(2), 157–183. <https://doi.org/10.62282/juilmu.v1i2.157-183>
- Bella Bernissa, A. (2020). IMPLEMENTASI ALGORITMA K-MEANS UNTUK MENENTUKAN KELAS UNGGULAN PADA SMPN 1 BOJONG. In *eProsiding Sistem Informasi (POTENSI)* (Vol. 1, Issue 1). <http://eprosiding.ars.ac.id/index.php/psi>
- Hasugian, P. S., & Raphita Sagala, J. (2022). KLIK: Kajian Ilmiah Informatika dan Komputer Penerapan Data Mining Untuk Pengelompokan Siswa Berdasarkan Nilai Akademik dengan Algoritma K-Means. *Media Online*, 3(3), 262–268. <https://djournals.com/klik>
- Hutagalung, J., Hendro Syahputra, Y., Pertiwi Tanjung, Z., Triguna Dharma, S., & Pintu Air, J. I. (2022). Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering. *Hal AH Nasution*, 9(1). <http://jurnal.mdp.ac.id>
- Kesuma Dinata, R., Hasdyna, N., & Azizah, N. (2020). Analisis K-Means Clustering pada Data Sepeda Motor. In *Informatics Journal* (Vol. 5, Issue 1).
- Kurniawan, S., Siregar, A. M., & Novita, H. Y. (2023). Penerapan Algoritma K-Means dan Fuzzy C-Means Dalam Mengelompokan Prestasi Siswa Berdasarkan Nilai Akademik. *IV*(1).
- Kusuma, A. S., & Aryati, K. S. (2019). Sistem Informasi Akademik Serta Penentuan Kelas Unggulan Dengan Metode Clustering Dengan Algoritma K-Means Di Smp Negeri 3 Ubud. *Jurnal Sistem Informasi Dan Komputer Terapan Indonesia (JSIKTI)*, 1(3), 143–152. <https://doi.org/10.33173/jsikti.29>
- Novi, & Ade, M. (2021). Penerapan Algoritma K-Means untuk Menentukan Kelas Unggulan di SMP Pelita Bandung. *Infomatek*, 32(2), 97–106. <https://doi.org/10.23969/infomatek.v23i2.4351>
- Nurajizah, S., & Salbinda, A. (2021). Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten. *Jurnal Teknik Komputer AMIK BSI*, 7(2). <https://doi.org/10.31294/jtk.v4i2>
- Oktario Dacwanda, D., & Nataliani, Y. (2021). Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan. *AITI: Jurnal Teknologi Informasi*, 18(Agustus), 125–138.
- Prasetyo Aji, W. (2024). Analisa Pengelompokan Data Nilai Rapot Siswa Menggunakan Pendekatan Metode K-Means Di SMK Ponpes Manba'ul Ullum Cirebon. *Kopertip : Jurnal Ilmiah Manajemen Informatika Dan Komputer*, 8(1), 14–18. <https://doi.org/10.32485/kopertip.v8i1.352>
- Putra Primanda, R., Alwi, A., & Mustikasari, D. (2021). *url : http://studentjournal.umpo.ac.id/index.php/komputek DATA MINING SELEKSI SISWA BERPRESTASI UNTUK MENENTUKAN KELAS UNGGULAN MENGGUNAKAN METODE K-MEANS CLUSTERING (Studi Kasus di MTS Darul Fikri )*. <http://studentjournal.umpo.ac.id/index.php/komputek>
- Saputra, D. A., & Pakereng, M. A. I. (2023). Analisis Data Nilai Siswa Kelas 8 Berbasis Nilai Pengetahuan Untuk Menentukan Siswa Berprestasi dengan K-Means Clustering (Kasus SMP Negeri 4 Salatiga). In *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 7, Issue 2).
- Sudrajat, W., Cholid, I., & Petrus, J. (2022). Penerapan Algoritma K-Means Clustering untuk Pengelompokan UMKM Menggunakan Rapidminer.
- Sulistiyawati, A., & Supriyanto, E. (2020). Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan. *Jurnal TEKNO KOMPAK*, 15(2), 1–12.

\* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)