

Clustering Value of Sman 1 Sumenep Student's Reports Using K-Means Method

Mariska Regina
Faculty of Computer Science
Information System
Surabaya, Indonesia
19082010069@student.upnjatim.ac.id

Fanni Silvana
Faculty of Computer Science
Information System
Surabaya, Indonesia
19082010091@student.upnjatim.ac.id

Dhian Satria Yudha Kartika
Faculty of Computer Science
Information System
Surabaya, Indonesia
dhian.satria@upnjatim.ac.id

Abstract— Education is the basis for the formation of human character. There are two types of education in schools, namely formal education and extracurricular education. Extracurricular activities are additional activities that are usually carried out outside school hours. The existence of this extracurricular activity helps students to develop their talents and potential. This research was conducted at SMA Negeri 1 Sumenep which facilitates 23 extracurricular activities for its students. Based on these activities, there are 72 students as a dataset. This study determines the grouping of extracurricular selection using the K-Means Clustering method. This method is expected to provide an understanding of the grouping of students according to their interests and potential based on the extracurricular options they choose. The results obtained are 3 clusters, namely the category of bad, good and very good with a value of 76.12 each. 85.25 and 79.43. Based on the three clusters formed with centroid points C0 (3.8, 77.9), C1 (1.4, 87.4), and C2 (1.8, 82.4).

Keywords : clustering, extracurricular, value

I. INTRODUCTION

Extracurricular is one of the activities carried out by students outside the school curriculum learning hours. The existence of this activity is expected to be able to develop the interests and potential possessed by students outside the school curriculum. These extracurricular activities themselves can be in the form of activities in the arts, sports or other activities that do have a positive purpose to support students' interests and potential. It is hoped that this extracurricular program will also make students accustomed and skilled in organizing, managing, solving problems and adding to existing insights according to the extracurriculars they are engaged in.

At SMA Negeri 1 Sumenep, the implementation of extracurricular activities is menso one of the important aspects that are considered to support the students. There are various types of extracurriculars, namely Nasa, Artline, Journalism, Robotics, Patas, Tiresfinger, Kates, Pik-r, PMR, Ragzsi, Hits,

Gumsa, Futsal, Badminton, Paskibra, Volly, Basketball, Tahfidz, Tata Culinary, Liris, Remus, MPK, and Boy Scouts. Each type of extracurricular has its own membership in which activity management is managed by each extracurricular. As for the delivery of extracurricular information through online media blogs, while member registration still uses the manual method by filling out a registration form. However, because the school is still not maximal in integrating data management for each student to group students' interests by looking at the extracurriculars chosen by students.

To solve this problem and to see the grouping of students, researchers conducted research using the K-Means Clustering method, because this method is a valid method to apply when solving various complex problems in computer science and statistics [1].

Clustering itself is a technique of data mining which later aimsto form data groups based on similarity characteristics with one data and another [2]. One of the clusteromg that can be used when grouping data with reference to similarity characteristics is K-Means Clustering. Several studies using the K-Means algorithm are often used because of its easy application and quite a short process in the process, so this research is expected to help Sumenep 1 Public High School in the process of seeing how it influences the number of specializations taken (extracurricular) with an average value report card

II. METHODOLOGY

This study uses a clustering technique with the K-Means method to find out the types of clusters that exist in the interests of SMA Negeri 1 Sumenep students. The stages in his own research include :

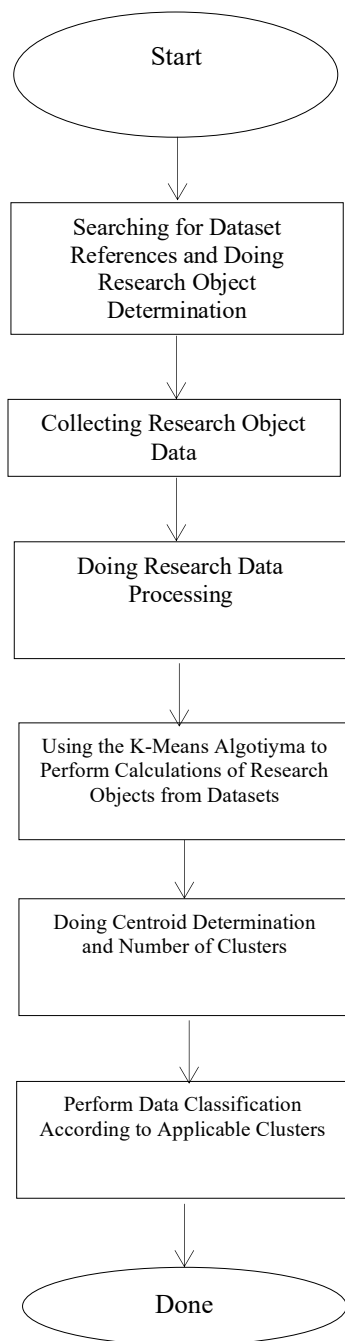


Fig 1. Methodology

The purpose of the above methodological stages is :

- 1) Collecting Research Object Data is a stage that is carried out to collect data in the form of extracurricular specialization data taken and data on the average value of class 10 and 11 class report cards for SMA Negeri 1 Sumenep.
- 2) Performing data calculations that have been obtained can be seen later from the number of Specializations Followed and the Average Value of Report Cards

3) Determine how many clusters (k) from the number of datasers previously obtained and then determine the centroids which are done randomly.

4) Perform calculations of the K-Means algorithm on specialization data followed by using the calculation of the Euclidean Distance on the available dataser with predetermined centroids. This time the calculation uses the following formula.

III. RESULT AND DISCUSSION

3.1 Display Data on the Number of the Number of Specializations Followed and the Average Value of Report Cards (Semester 1)

Jumlah Klaster = 2 Nilai rata-rata silhouette = 0.5570662821567479
 Jumlah Klaster = 3 Nilai rata-rata silhouette = 0.5020725174687367
 Jumlah Klaster = 4 Nilai rata-rata silhouette = 0.5144376023861307
 Jumlah Klaster = 5 Nilai rata-rata silhouette = 0.48791912851262076

Fig 2. Cluster Average

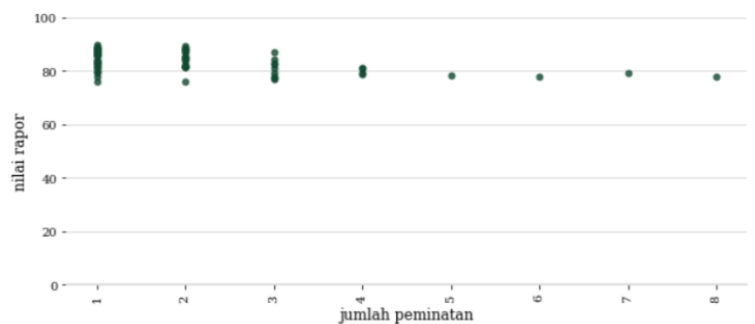


Fig 3. Number of interest in report cards

The stages before determining clustering are required to know how the data collection is in graphic form so that it is able to provide a description of how the data must be processed.

3.2 Perform Determination of the Optimal Number of Cluster

When determining the optimal of cluster, , this research uses Google Collabs with the application of the Elbow Method so that you can see how the data statistics are, which will later be processed how many clusters can be applied.

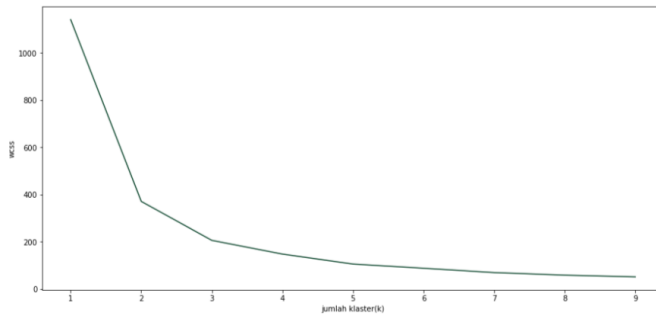


Fig 4. Elbow Method

3.3 Perform Silhouette Determinations

The largest number of clusters and silhouettes are two clusters, but in this study 3 clusters were used because there was a lot of data in order to meet the suitability of the objectives, namely, very good, good enough, not good.

3.4 Perform Determination of the Optimal Number of Clusters

a) Data Entry

	Nama Siswa	Jumlah Peminatan yang Diikuti	Nilai Rata-rata Raport (Semester 1)
0	Afiqatun Nabila Rahman Pramuka	1	82.86
1	Riskiayatul Nur Oktarani	2	81.44
2	SHINTA DEWI NURUL IZZA	5	78.31
3	Putri Isyam Meylani	3	76.93
4	N. Fauzan Rofiqi	2	81.75
...	...	...	...
65	Neva Zevanya Nana Naulita	3	83.00
66	Malita wardiah	1	81.13
67	Muhammad Rizky ramadhani	6	77.94
68	hirda balqis listyawati	1	87.44
69	Nayla Syam Arifah	1	84.25

Fig 5. Data Entry

b) Determine the number of clusters and centroid values

After inputting the data, it determines the number of clusters or groups of initial centroid values, which results from the formulation automatically, the data results will appear as follows.

TABLE I. CENTROID VALUE

Centroids	0	1	2
Number of Interests Followed	3.8	1.4	1.8
Average	77.9	87.4	82.4

c) Perform Group Determination and Recalculate Centroid Values

The system for calculating the distance between the cluster center point and the point on the object will then be grouped on the basis of comparing and selected through the closet distance between the data and the cluster center, this distance shows that the data is in a group with the nearest cluster center by comparing the cluster results and taken from the smallest.

d) Grouping Results

When you have got the results by K-Means Clustering, you will get the results from grouping of each cluster, namely

TABLE II. INTERESTS STUDENT

No.	Student's Name	Number Of Interests Followed	Report Card Average (Semester 1)	Clusters
1.	Shinta Dewi Nurul Izza	5	78,31	0
...	...	...	...	...
15.	Muhammad Rizky Ramadhani	6	77,94	0
16.	Priyanka Diaz Muhammad	2	85.25	1
...	...	...	...	...
43.	Hirda Balqis Listyawati	1	87,44	1
44.	Afiqatun Nabila Rahman	1	82,86	2
...	...	...	...	...
70.	Nayla Syam Arifah	1	84.25	2

TABLE III. CLUSTER AVERAGE SCORE

Clusters	Category	Minimum Average Score
Clusters 0	Not good	76,12
Clusters 1	Very good	85.25
Clusters 2	Pretty good	79,43

IV. CONCLUSION

So conclusions can be drawn based on the results obtained using the K-Means algorithm on 72 types of data, the number of specializations attended and the average value of the report cards taken from grades 10 and 11 students of SMA Negeri 1 Sumenep resulting in three clusters that are not well-interpreted. very good and good enough. Based on the three clusters formed with centroid points C0 (3.8, 77.9), C1 (1.4, 87.4), and C2 (1.8, 82.4).

Based on the results, cluster 0 or which is not good has 13 students, cluster 1 in the very good category has 28 students and cluster 2 in the fairly good category has 29 students.

The advice that we can give for the future is for the involvement of the BK in extracurricular selection for each student, so that students remain focused on academic grades but can still develop their interests to the fullest.

- [1] D. Anggarwati, O. Nurdiawan, I. Ali and D. A. Kurnia, "Penerapan Algoritma K-Means Dalam Prediksi Penjualan Karoseri," JURNAL DATA SCIENCE & INFORMATIKA (JDSI), vol. Vol. 1 No. 2, pp. 58-62, 2021.
- [2] J. Hutagalung and F. Sonata, "Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah Asuransi," JURNAL MEDIA INFORMATIKA BUDIDARMA, Vol. Volume 5, Nomor 3, pp. 1187-1194, 2021.
- [3] D.O. Dacwanda and Y. Nataliani, "Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan," JURNAL TEKNOLOGI INFORMASI, Vol. 18 No.2, pp. 125-138, 2021.
- [4] A. S. Kusuma and K. S. Aryati, "Sistem Informasi Akademik Serta Penentuan Kelas Unggulan Dengan Metode Clusttering Dengan Algoritama K-Means Di Smp Negeri 3 Ubud," J. Sist. Inf. dan Komput. Terap. Indones., vol. 1, no. 3, pp. 143–152, 2019.
- [5] M. A. W. K. Murti, "Penerapan Metode K-Means Clustering Untuk Mengelompokan Potensi Produksi Buah – Buah Di Provinsi Daerah Istimewa Yogyakarta," Universitas Sanata Dharma, 2017
- [6] P.U.F. Aulia and S.Saepudin, "PENERAPAN DATA MINING K-MEANS CLUSTERING UNTUK MENGELOMPOKKAN BERBAGAI JENIS MEERK LAPTOP", SISTEMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika), Universitas Nusa Putra, 7 Agustus 2021.
- [7] Vania, E., Nuraini, S., & Kartika, D. S. Y. (2022, September). PENGGUNAAN ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN REKOMENDASI FILM INDONESIA. In Prosiding Seminar Nasional Teknologi dan Sistem Informasi (Vol. 2, No. 1, pp. 207-214)..
- [8] Al Rasyid, H., Soebari, B. F. K., & Kartika, D. S. Y. (2022, September). IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN PENJUALAN PRODUK PADA ONLINE SHOP TOKO GIZI. In Prosiding Seminar Nasional Teknologi dan Sistem Informasi (Vol. 2, No. 1, pp. 242-248).
- [9] Musyarofah, U. L., Alima, S. N., & Kartika, D. S. Y. (2022, September). KLASIFIKASI TOP 50 SPOTIFY TAHUN 2010-2019 MENGGUNAKAN METODE K-MEANS CLUSTERING. In Prosiding Seminar Nasional Teknologi dan Sistem Informasi (Vol. 2, No. 1, pp. 215-220).