



Use of the K-Means Method in Student Mapping

Penggunaan Metode K-Means dalam Pemetaan Siswa

Yohan Rahman S. Ramli
Universitas Indraprasta PGRI
yohanrahman@gmail.com

Abstract

High school (SMA) has several majors and entry requirements for these majors. One of them is senior high school. One of the majors in this high school is the Mathematics and Natural Sciences department, usually there are class groups that are tailored to the knowledge and abilities of prospective students. but there are requirements for prospective new students to be accepted into the Mathematics and Natural Sciences class based on their abilities and knowledge, one of the requirements is the National Examination Score at the Junior High School level, the Mathematics National Examination scores and the Science National Examination scores. The K-Means method is one of the methods used to determine the grouping based on predetermined data.

Keywords: *k-means, cluster ing, student, data mining*

Abstrak

Setiap sekolah menengah ke Atas (SMA) memiliki beberapa jurusan dan syarat masuk ke jurusan tersebut. Salah satunya adalah Jenjang SMA . Salah satu jurusan yang ada di SMA ini adalah jurusan MIPA, biasanya terdapat kelompok kelas yang disesuaikan dengan pengetahuan dan kemampuan calon siswa. namun terdapat syarat untuk calon siswa baru agar diterima masuk ke kelas MIPA berdasarkan kemampuan dan pengetahuannya, salah satu syaratnya adalah Nilai Ujian Nasional di jenjang SMP, nilai Ujian Nasional Matematika dan Nilai Ujian Nasional IPA. Metode K-Means adalah salah satu metode yang digunakan untuk mengetahui pengelompokan berdasarkan data yang telah ditentukan. Untuk itu metode K-Means dapat digunakan untuk mengetahui kelompok kelas MIPA berdasarkan nilai Ujian Nasional calon siswa SMA. Namun metode K-Means memiliki nilai akurasi yang rendah yaitu sekitar 34% saja.

Kata kunci: *k-means, pengelompokan, siswa, pengumpulan data*

Page | 349

PENDAHULUAN

Pada jenjang SMA terdapat beberapa jurusan yang akan diambil oleh calon siswa baru sesuai dengan minat dan kemampuannya. Salah satu jurusan yang ada di Jenjang SMA adalah jurusan Matematika dan Ilmu Pengetahuan Alam (MIPA) untuk menerima calon siswa baru Jenjang SMA mempunyai syarat-syarat tertentu, salah satu syaratnya adalah berdasarkan nilai ujian yang terdiri dari nilai Ujian Nasional Matematika, Nilai Ujian Nasional IPA dan Nilai rata-rata ujian nasional calon siswa tersebut. Nilai ujian nasional calon siswa akan menentukan pengelompokan kelas MIPA berdasarkan syarat yang telah ditentukan. Salah satu bentuk analisis yang digunakan untuk penelitian ini dengan cara mengelompokkan data atau biasa disebut dengan *Cluster ing*.

Algoritma K-means dirasa dapat digunakan dalam penelitian ini karena dapat menentukan jumlah *cluster* yang optimal sehingga memiliki akurasi yang lebih baik. selain itu, dalam implementasinya, algoritma ini memiliki beberapa kelemahan salah satunya adalah tingkat akurasi yang rendah.

Permasalahan tingkat akurasi yang rendah dapat disebabkan karena pemilihan *cluster* yang kurang tepat sehingga perlu adanya implementasi sebuah pendekatan baru sebagai solusi untuk menentukan jumlah *cluster* yang dinamis.



METODE

Menurut Kusriani (2009) data *Mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Data *mining* adalah proses yang menggunakan Teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk menginteraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

Menurut Tan (2006) Data *mining* merupakan salah satu disiplin ilmu yang digunakan untuk menentukan suatu informasi tertentu dalam sekumpulan data sebagai pendukung pengambilan keputusan. Data *mining* juga sebagai salah satu proses untuk memperoleh informasi yang memiliki nilai guna dari sekumpulan data.

Menurut AK Jain (1999) *Cluster* ing diartikan sebagai klasifikasi pola tanpa pengawasan ke dalam sebuah *cluster* . Menurut Kusriani (2009) jarak antar *cluster* didefinisikan dengan menggunakan beberapa metode-metode untuk menentukan *cluster* mana yang saling berdekatan. *Single linkage* adalah jarak terkecil antar satu elemen *cluster* dalam satu *cluster* dengan elemen suatu *cluster* dengan elemen lain di *cluster* di *cluster* yang berbeda.

Complete Linkage adalah jarak terbesar antara satu elemen dalam suatu *cluster* dengan elemen lain di *cluster* yang berbeda. Average adalah jarak rata-rata antara satu elemen dalam suatu *cluster* dengan elemen lain di *cluster* yang berbeda. Centroid adalah jarak antar centroid dari tiap *cluster* dengan centroid *cluster* lainnya. Medoid adalah jarak antar medoid dari tiap *cluster* dengan medoid *cluster* lainnya.

Menurut Syafnidawati (2020) K-means *cluster* ing merupakan salah satu metode *cluster* analisis non hirarki yang berusaha untuk memartisi objek yang ada kedalam satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam *cluster* yang lain.

Langkah dalam menyelesaikan penelitian dengan metode k-means :

1. Menentukan jumlah *Cluster* yang akan digunakan.
2. Alokasi data ke dalam *Cluster* secara acak.
3. Hitung jarak antar centroid ke data dari masing-masing *cluster* dengan rumus euclidean.

$$d(x_i, c_j) = \sqrt{\sum_{i=1}^r (x_i - c_j)^2}$$

Keterangan :

x = data

c = centroid

i = data ke -

j = centroid ke -

Alokasi data masing-masing ke centroid/rata-rata terdekat dengan rumus ditentukan centroid baru untuk menghitung iterasi selanjutnya hingga nilai initial *cluster* centre sama atau *cluster* berpindah. Dengan rumus

$$c = \frac{\sum_{i=1}^r (x_{i1} + x_{i1} + \dots x_{in})}{j}$$

Keterangan :

C = centroid

X = data ke

DOI UNTUK ARTIKEL INI

<https://doi.org/10.37010/int.v3i2.1072>

Scan barcode untuk
mengunjungi OJS
kami





Intelektium adalah jurnal yang diterbitkan oleh Neoelectura, diterbitkan dua kali dalam satu tahun. Intelektium adalah media publikasi ilmiah dalam bentuk makalah konseptual dan penelitian lapangan yang terkait dengan bidang pendidikan. Diharapkan Intelektium dapat menjadi media bagi akademisi dan peneliti untuk menerbitkan karya ilmiah mereka dan menjadi sumber referensi untuk pengembangan ilmu pengetahuan.

J = jumlah data pada *cluster*

Kembali ke tahap ke-3, lakukan perulangan hingga anggota *cluster* tidak berpindah ke *cluster* yang lain dan nilai centroid memiliki hasil yang tetap.

HASIL DAN PEMBAHASAN

Hasil

Tujuan penelitian ini adalah untuk mengetahui penempatan calon siswa baru di kelas MIPA berdasarkan nilai ujian nasional saat berada SMP. Penelitian ini menggunakan metode K-Means dengan jarak perhitungan Single Linkage untuk menentukan jarak centroid terdekat dari data. Dalam penelitian ini memiliki beberapa kriteria diantaranya adalah Nilai Ujian Nasional Matematika, Nilai Ujian Nasional IPA dan Rata-rata dari seluruh Nilai Ujian Nasional calon siswa MIPA Jenjang SMA. Sebagai data uji, jumlah data yang digunakan dalam penelitian ini adalah 10 data calon siswa Jenjang SMA. Data ini dapat dilihat pada tabel 1.

Tabel 1. Data Calon Siswa Mipa Jenjang SMA

No	Calon Siswa Ke	Matematika	IPA	Rata-Rata Nilai
1	Calon Siswa Ke-1	37.50	35.00	46.43
2	Calon Siswa Ke-2	36.31	40.94	45.78
3	Calon Siswa Ke-3	68.43	73.10	75.49
4	Calon Siswa Ke-4	38.98	42.38	48.00
5	Calon Siswa Ke-5	75.00	68.13	77.45
6	Calon Siswa Ke-6	36.02	39.03	44.44
7	Calon Siswa Ke-7	46.50	56.50	52.08
8	Calon Siswa Ke-8	70.80	67.71	74.97
9	Calon Siswa Ke-9	74.72	68.80	78.77
10	Calon Siswa Ke-10	93.79	88.01	91.14

Page | 351

Selanjutnya, data calon siswa MIPA Jenjang SMA dikelompokkan menjadi 3 kelompok. Dengan menggunakan metode K-Means, berikut adalah langkah-langkah penyelesaiannya :

Menentukan jumlah *Cluster*, jumlah *cluster* merupakan jumlah kelompok yang akan dihasilkan. Dalam pengelompokkan calon siswa MIPA terdapat 3 *cluster* yang dibutuhkan.

Menentukan centroid awal data calon siswa MIPA Jenjang SMA pada tabel 1 ditentukan centroid secara random. Centroid 1 terdapat pada data calon siswa ke-2, centroid 2 terdapat pada data calon siswa ke-5, centroid 3 terdapat pada data calon siswa ke-9. Penentuan centroid awal dapat dilihat pada tabel 2.

Tabel 2. Penentuan Centroid

No	Calon Siswa Ke	Matematika	IPA	Rata-Rata Nilai	Centroid
1	Calon Siswa Ke-1	37.50	35.00	46.43	centroid 1
2	Calon Siswa Ke-2	36.31	40.94	45.78	
3	Calon Siswa Ke-3	68.43	73.10	75.49	
4	Calon Siswa Ke-4	38.98	42.38	48.00	centroid 2
5	Calon Siswa Ke-5	75.00	68.13	77.45	
6	Calon Siswa Ke-6	36.02	39.03	44.44	
7	Calon Siswa Ke-7	46.50	56.50	52.08	centroid 3
8	Calon Siswa Ke-8	70.80	67.71	74.97	
9	Calon Siswa Ke-9	74.72	68.80	78.77	
10	Calon Siswa Ke-10	93.79	88.01	91.14	



DOI PUBLIKASI

<https://doi.org/10.37010/int.v3i2>

SEPTEMBER

Vol. 3 No. 2

2022

Menghitung jarak setiap data ke centroid yang telah ditentukan. Dengan menggunakan persamaan Eclidean Distance Space. Adapun contoh perhitungan jarak setiap data ke centroid yang telah ditentukan adalah sebagai berikut :

a. Jarak calon siswa ke-1 ke centroid 1 :

$$\begin{aligned} (x_1, c_2) &= \sqrt{\sum_{i=1}^r (x_{1i} - c_{1i})^2} \\ &= \sqrt{(37,50 - 36,31)^2 + (40,94 - 35,00)^2 + (46,43 - 45,78)^2} \\ &= 6,09 \end{aligned}$$

b. Jarak calon siswa ke-1 ke centroid 2 :

$$\begin{aligned} (x_1, c_3) &= \sqrt{\sum_{i=1}^r (x_{1i} - c_{2i})^2} \\ &= \sqrt{(37,50 - 75,00)^2 + (40,94 - 68,13)^2 + (46,43 - 77,45)^2} \\ &= 58,87 \end{aligned}$$

c. Jarak calon siswa ke-1 ke centroid 3 :

$$\begin{aligned} (x_1, c_3) &= \sqrt{\sum_{i=1}^r (x_{1i} - c_{3i})^2} \\ &= \sqrt{(37,50 - 74,72)^2 + (40,94 - 68,80)^2 + (46,43 - 78,77)^2} \\ &= 59,78 \\ &= \sqrt{((37,50-74,72)^2 + (40,94-68,80)^2 + (46,43-78,77)^2)} \\ &= 59,78 \end{aligned}$$

Adapun hasil dari perhitungan jarak antar data setiap ke centroid untuk menentukan *cluster* , dapat dilihat pada tabel 3.

Tabel 3. Hasil Perhitungan jarak data calon siswa pada masing-masing centroid.

Data Ke-	Centroid 1	Centroid 2	Centroid 3
1	6.09	58.87	59.78
2	0.00	56.91	57.79
3	54.30	8.47	8.30
4	3.76	53.18	54.06
5	56.91	0.00	1.51
6	2.35	58.79	59.69
7	19.64	39.89	40.74
8	52.52	4.90	5.57
9	57.79	1.51	0.00
10	87.05	30.59	29.76

Setelah penentuan nilai centroid terdekat, langkah selanjutnya adalah menentukan *cluster* pada data calon siswa MIPA Jenjang SMA dengan jarak antar data ke centroid yang paling dekat, sehingga diperoleh hasil seperti pada tabel 4.

Tabel 4. Hasil Penentuan Cluster .

No.	Calon siswa	C1	C2	C3
1	Calon Siswa Ke-1	*		
2	Calon Siswa Ke-2	*		
3	Calon Siswa Ke-3			*
4	Calon Siswa Ke-4	*		
5	Calon Siswa Ke-5		*	
6	Calon Siswa Ke-6	*		
7	Calon Siswa Ke-7	*		
8	Calon Siswa Ke-8		*	
9	Calon Siswa Ke-9			*
10	Calon Siswa Ke-10			*

Setelah *cluster* ditentukan, langkah selanjutnya adalah menentukan centroid baru di masing-masing *cluster* menggunakan rumus yang ada pada langkah ke-5.





Intelektium adalah jurnal yang diterbitkan oleh Neolectura, diterbitkan dua kali dalam satu tahun. Intelektium adalah media publikasi ilmiah dalam bentuk makalah konseptual dan penelitian lapangan yang terkait dengan bidang pendidikan. Diharapkan Intelektium dapat menjadi media bagi akademisi dan peneliti untuk menerbitkan karya ilmiah mereka dan menjadi sumber referensi untuk pengembangan ilmu pengetahuan.

Centroid 1 :

matematika= $(37,50+36,31+38,98+36,02+46,50)/5=67,75$

IPA= $(35,00+40,94+42,38+39,03+56,50)/5=42,77$

Rata-rata Nilai UN= $(46,43+45,78+48,00+44,44+52,08)/5=47,35$

Hasil penentuan nilai centroid baru pada 10 data calon siswa MIPA Jenjang SMA dapat dilihat pada tabel 5.

Tabel 5. Nilai Centroid baru

Centroid	Matematika	IPA	Rata-rata Nilai UN
Centroid 1	39.06	42.77	47.35
Centroid 2	72.90	67.92	76.21
Centroid 3	78.98	81.80	81.80

Setelah penentuan centroid baru, langkah selanjutnya adalah membandingkan nilai centroid sebelumnya dengan menggunakan nilai Initial *Cluster* Centre. Apabila nilai sama dan penempatan custer tidak berubah maka iterasi tidak dilanjutkan. Namun jika nilai tidak sama atau terdapat *cluster* yang berubah maka iterasi akan dilanjutkan hingga nilai initial *cluster* centre sama atau *cluster* tidak berubah.

Hasil *cluster* ing iterasi 2 dihitung menggunakan Microsoft Excel dan hasil *cluster* ing dapat dilihat pada tabel 6.

Tabel 6. Cluster iterasi 2

No.	Calon siswa ke-	Jarak data ke pusat Cluster	Cluster
1	Calon Siswa Ke-1	7.98	1
2	Calon Siswa Ke-2	3.66	1
3	Calon Siswa Ke-3	6.88	2
4	Calon Siswa Ke-4	0.76	1
5	Calon Siswa Ke-5	2.45	2
6	Calon Siswa Ke-6	5.63	1
7	Calon Siswa Ke-7	16.32	1
8	Calon Siswa Ke-8	2.45	2
9	Calon Siswa Ke-9	3.26	2
10	Calon Siswa Ke-10	18.58	3

Pada tabel 6 terdapat perubahan *cluster* dari iterasi sebelumnya, antara lain data calon siswa ke-3 berpindah *cluster* dari *cluster* 3 ke *cluster* 2, dan data calon siswa ke-9 berpindah *cluster* dari *cluster* 2 ke *cluster* 3. Sehingga iterasi harus dilanjutkan karena terdapat perubahan *cluster* atau nilai initial *cluster* centre berubah Dengan menggunakan nilai centroid baru. Hasil perhitungan *cluster* pada iterasi 3 dengan menggunakan Microsoft excel disajikan pada tabel 7.

Tabel 7. Hasil iterasi 3

No.	Calon siswa ke-	Jarak data ke pusat Cluster	Cluster
1	Calon Siswa Ke-1	7.98	1
2	Calon Siswa Ke-2	3.66	1
3	Calon Siswa Ke-3	5.41	2
4	Calon Siswa Ke-4	0.76	1
5	Calon Siswa Ke-5	3.15	2
6	Calon Siswa Ke-6	5.63	1
7	Calon Siswa Ke-7	16.32	1
8	Calon Siswa Ke-8	2.82	2
9	Calon Siswa Ke-9	3.31	2
10	Calon Siswa Ke-10	0.00	3



Pada tabel 7 tidak terdapat perubahan *cluster* dari iterasi sebelumnya atau nilai initial *cluster* centre sama, maka iterasi dihentikan. Hasil akhir pembagian kelas MIPA untuk calon siswa Jenjang SMA dapat dilihat pada tabel 8.

Tabel 8. Hasil akhir pembagian kelas MIPA calon siswa Jenjang SMA

Calon siswa	Kelas MIPA
Calon Siswa Ke-1	1
Calon Siswa Ke-2	1
Calon Siswa Ke-4	1
Calon Siswa Ke-6	1
Calon Siswa Ke-7	1
Calon Siswa Ke-3	2
Calon Siswa Ke-5	2
Calon Siswa Ke-8	2
Calon Siswa Ke-9	2
Calon Siswa Ke-10	3

PENUTUP

Metode K-means dapat digunakan untuk menentukan pengelompokan kelas calon siswa MIPA berdasarkan nilai Ujian Nasional Matematika, nilai Ujian Nasional IPA dan rata-rata Nilai Ujian Nasional siswa di jenjang SMP. Namun metode K-Means hanya memiliki nilai akurasi kebenaran sebesar 34% karena nilai dapat berubah sesuai dengan jumlah data latih. Penelitian selanjutnya seharusnya juga merangkum bagaimana penilaian formatif dilakukan untuk mendapatkan rekam nilai yang komprehensif (Robbani, Megayanti, & Prasmoro, 2020). Saat ini sekolah dengan kurikulum merdeka pun sebaiknya tidak hanya memberikan pilihan jurusan yang terbatas, namun juga mempertimbangkan bagaimana kebutuhan dunia kerja di *Era Society* 5.0 ini dihadapi para lulusan. (Robbani, Megayanti, & Rosadi, 2020) menjelaskan pentingnya pengenalan kompetensi sebagai variable pemetaan jurusan

DAFTAR PUSTAKA

- Jain, Anil K., M. Narasimha Murty, dan Patrick J. Flynn. (1999). Pengelompokan data: ulasan. *Survei komputasi ACM (CSUR)* 31.3 (1999): 264-323.
- Kusrini. (2007) konsep dan aplikasi sistem pendukung keputusan. Yogyakarta : Andi
- Robbani, H., Megayanti, W., & Prasmoro, A. V. (2020). Formative Assessment Strategies Using Elearning. *NUCLEUS*, 1(1), 45–49. <https://doi.org/10.37010/nuc.v1i1.85>
- Robbani, H., Megayanti, W., & Rosadi, N. (2020). Strategy For Implementing Certification Development Of Vocational School Graduates According To Inpres No. 9/2016. *Judicious*, 1(1), 48–54. <https://doi.org/10.37010/jdc.v1i1.102>
- Syafnidawati. (2020) K-Means *Cluster* ing. <https://raharja.ac.id/2020/04/19/k-means-cluster-ing/>. diakses tanggal 05 November 2021.

