

Jumlah Cluster Optimal dalam Pengelompokan Siswa SMK dengan Metode Elbow K-Means Clustering

Ninik Tri Hartanti^{1*}, Erni Seniwati², Rina Pramitasari³, Irma Rofni Wulandari³

^{1,2,3,4}Universitas Amikom Yogyakarta

e-mail: [1ninik.t@amikom.ac.id](mailto:ninik.t@amikom.ac.id), [2erni@amikom.ac.id](mailto:erni@amikom.ac.id), [3rina.pramitasari@amikom.ac.id](mailto:rina.pramitasari@amikom.ac.id), [4irma@amikom.ac.id](mailto:irma@amikom.ac.id)
e-mail: ninik.t@amikom.ac.id

Abstrak - Jenjang pendidikan di Indonesia setelah sekolah lanjutan pertama adalah Sekolah Menengah Tingkat Atas (SMA) atau yang sederajat yaitu sekolah kejuruan SMK. Perkembangan teknologi telah memberikan berbagai kemudahan, termasuk dalam bidang pendidikan. Salah satu bentuk inovasi teknologi dalam pendidikan adalah penerapan data mining, khususnya educational data mining. Penerapan educational data mining dengan metode K-Means untuk mengelompokkan data siswa SMK di kabupaten Magelang, merupakan inti dari penelitian ini. Adapun data yang diterapkan merujuk pada data publik <https://dapo.kemendikbud.go.id> Berdasarkan data tersebut, Metode Elbow diterapkan untuk mengidentifikasi jumlah cluster yang paling sesuai. Jumlah cluster ini kemudian menjadi dasar penghitungan menggunakan algoritma K-Means. Hasil analisis menunjukkan tiga kelompok: Cluster 1, yang meliputi siswa dengan minat rendah untuk melanjutkan ke SMK, terdapat di Kecamatan Muntilan dan Mertoyudan; Cluster 2, terdiri dari siswa dengan kecenderungan minat cukup dalam memilih lanjut studi ke SMK, tersebar di 10 kecamatan, yaitu Salaman, Tegalrejo, Secang, Salam, Borobudur, Tempuran, Windusari, Kaliangkrik, Kajoran dan Ngablak, serta Cluster 3 mencakup siswa dengan kecenderungan minat tertinggi untuk lanjut studi ke SMK, terdiri dari 9 kecamatan, yaitu Mungkid, Grabak, Bandongan, Sawangan, Pakis, Candimulyo, Dukun, Srumbung, dan Ngluwar. Penelitian ini sangat penting dalam mendukung kinerja pemerintah daerah Kabupaten Magelang dalam sistem pemantauan distribusi jumlah siswa Sekolah Menengah Kejuruan di setiap kecamatan. Dengan demikian, akan lebih mudah untuk mendeteksi kecamatan mana yang memiliki jumlah siswa yang kurang atau bahkan terlalu banyak. Aspek ini penting untuk diperhatikan karena berkaitan langsung dengan ketersediaan sarana prasarana dan tenaga pengajar di sekolah-sekolah menengah tersebut.

Kata kunci: Educational Data Mining; Cluster; Centroid; K-Means; Elbow

Abstract - Educational level in Indonesia after junior high school is Senior High School (SMA) or the equivalent, namely vocational school. Technological developments have provided various conveniences, including in the field of education. One form of technological innovation in education is application data mining, specifically educational data mining. Implementation of Educational data mining using the K-Means method to group vocational school student data in Magelang district is the core of this research. The data applied refers to public data <https://dapo.kemendikbud.go.id> Based on this data, the Elbow Method is applied to identify quantities cluster the most suitable one. Amount cluster This then becomes the basis for calculations using the K-Means algorithm. The analysis results show three groups: Cluster 1, which includes students with low interest in continuing to vocational school, is found in Muntilan and Mertoyudan Districts; Cluster 2, consisting of students with sufficient interest in choosing to continue their studies at vocational school, spread across 8 sub-districts, namely Secang, Tegalrejo, Salam, Bandongan, Mungkid, Salaman, Grabag, and Borobudur, as well as Cluster 3 includes students with the highest interest in continuing their studies at vocational school, consisting of 11 sub-districts, Srumbung, Dukun, Ngluwar, Kaliangkrik, Tempuran, Pakis, Candimulyo, Windusari, Kajoran, Ngablak, and Sawangan. This research is very important in supporting the performance of the local government of Magelang Regency in the monitoring system for the distribution of the number of Vocational High School students in each sub-district. Thus, it will be easier to detect which sub-districts have too few or too many students. This aspect is important to note because it is directly related to the availability of facilities and infrastructure and teaching staff in these high schools.

Keywords: Educational Data Mining; Cluster; Centroid; K-Means; Elbow

PENDAHULUAN

Seiring dengan perkembangan teknologi yang sangat cepat, seluruh negara di dunia akan merasakan dampak dari perkembangan tersebut. Khususnya di Indonesia, perkembangan teknologi sangat dibutuhkan dalam segala bidang kehidupan. Dampak

ini bisa terlihat dan dapat dirasakan dalam kemudahannya sebuah informasi sampai kepada masyarakat. Melalui beberapa media sosial semua jenis informasi akan dengan mudah sampai ke masyarakat, terlepas bahwa informasi tersebut berupa ujaran kebajikan, kebenaran ataupun informasi

sebaliknya yang mengarah keburukan, kebencian atau *hoax*. Seluruh informasi yang diterima oleh masyarakat tersebut, diharapkan akan membawa dampak positif dalam kehidupannya sehingga masyarakat dapat hidup aman, nyaman, rukun dan bahagia. Salah satu bentuk informasi yang diharapkan bisa memberikan pencerahan dalam kehidupan masyarakat adalah bidang pendidikan, seperti status sekolah, ketersediaannya tenaga pendidik atau guru dan informasi terkait kualitas dari sekolah. Pemanfaatan algoritma data mining dalam bidang pendidikan salah satunya adalah algoritma *clustering*. Bidang ilmu baru yang berhubungan dengan penerapan data mining untuk mengeksplorasi data pendidikan yang besar, misalnya data sekolah, tempat kursus, maupun perguruan tinggi untuk memahami pentingnya perhatian atas masalah pendidikan (Palacios et al., 2021).

Educational data mining telah menjadi alat yang efektif dalam tujuan eksplorasi data berskala besar yang tidak diketahui maupun belum terdeteksi dalam data pendidikan seperti data prestasi siswa (Yağcı, 2022).

Clustering adalah teknik analisis data yang bertujuan membantu menyelesaikan masalah pengelompokan. K-Means, yang merupakan metode clustering non-hierarki, beroperasi pada atribut numerik dengan mengukur jarak antar titik data untuk membentuk cluster. Metode ini menjadikan data dalam satu cluster atau lebih, untuk data yang berkarakteristik sama akan dijadikan satu kelompok dengan cluster yang sama, sedangkan data berkarakteristik lain akan disatukan ke cluster yang berbeda pula (Maulana et al., n.d.).

K-Means clustering memiliki keunggulan dalam mengelompokkan data dalam jumlah besar secara cepat dan efisien. Namun, metode ini memiliki kelemahan, yaitu bergantung pada pemilihan awal titik pusat cluster. Oleh karena itu, hasil clustering dengan K-Means akan lebih optimal jika titik pusat cluster ditentukan dengan tepat (Hartanti, 2022).

Dalam pengelompokan data, evaluasi clustering menjadi bagian penting dari proses tersebut. Metode penilaian yang sering ditemukan adalah metode Elbow yang efektif untuk mengetahui seberapa baik kualitas dari hasil *clustering*. Penilaian ini bertujuan untuk menentukan jumlah cluster yang optimal melalui perhitungan, seperti indeks Davies-Bouldin, Elbow, dan Silhouette (Sholeh & Aeni, n.d.).

Evaluasi clustering bertujuan untuk menilai kualitas hasil pengelompokan data. Metode Elbow digunakan untuk membantu menentukan jumlah cluster yang optimal dengan menganalisis persentase perubahan pada titik di mana grafik membentuk sudut siku. Proses ini melibatkan pemilihan nilai cluster berikutnya sebagai model untuk menentukan cluster terbaik, diikuti oleh perhitungan persentase perubahan yang menjadi dasar perbandingan jumlah cluster (Nainggolan et al., 2019). Merujuk informasi

pendidikan [dapodikdasmen](https://dapodikdasmen.kemdikbud.go.id) Magelang, <https://dapodikdasmen.kemdikbud.go.id>, di wilayah Kabupaten Magelang terdapat 21 kecamatan dan terdapat 46 SMK, yaitu 3 SMK Negeri dan 43 SMK Swasta.

Artikel tentang penerapan K-Means dan Elbow telah dilakukan untuk mengelompokkan obat berdasarkan review data. Keluaran dari penelitian tersebut berupa usulan model terbaik yang digunakan untuk mengklaster ulasan pengguna terhadap obat-obatan dengan menerapkan K-Means, Silhouette dan Elbow. Usulan model dapat merekomendasikan metode clustering K-Means yang terbaik untuk memproses data ulasan dan rating yang luas (Juanita & Cahyono, 2024).

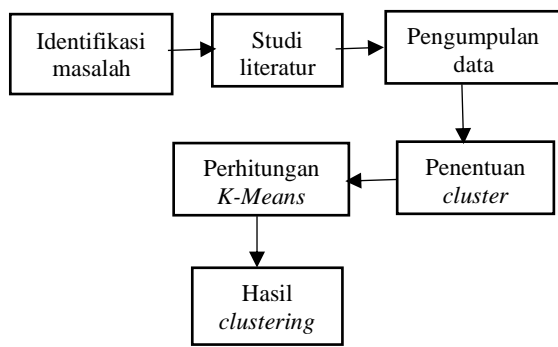
Penelitian dalam bidang kesehatan yang menerapkan metode clustering lainnya adalah penentuan pada pemrosesan citra untuk mengidentifikasi penyakit kanker (Sammouda & El-Zaart, 2021). Hasil penelitian ini berhasil mengatasi kendala untuk menentukan jumlah cluster yang paling tepat dalam segmentasi citra kanker prostat, serta efektif dalam mengelompokkan piksel pada citra kanker prostat.

Selanjutnya, dalam penerapan metode clustering, centroid ditentukan berdasarkan nilai rata-rata dan median dari data. Hasilnya menunjukkan bahwa penentuan pusat cluster awal mengurangi jumlah iterasi yang dibutuhkan sebesar 22,58% dibandingkan dengan penentuan cluster secara acak. Selain itu, penentuan jumlah cluster optimal menggunakan metode siku mengurangi iterasi yang diperlukan sebesar 25% dibandingkan dengan penggunaan jumlah cluster lainnya (Umargono et al., 2020).

Penelitian ini diharapkan dapat mendukung kinerja pemerintah daerah Kabupaten Magelang dalam sistem pemantauan distribusi jumlah siswa Sekolah Menengah Kejuruan di setiap kecamatan. Dengan demikian, akan lebih mudah untuk mendeteksi kecamatan mana yang memiliki jumlah siswa yang kurang atau bahkan terlalu banyak. Aspek ini penting untuk diperhatikan karena berkaitan langsung dengan ketersediaan sarana prasarana dan tenaga pengajar di sekolah-sekolah menengah tersebut.

METODE PENELITIAN

Tahapan dalam metode penelitian tentang penerapan metode elbow dalam K-Means clustering terlampir pada Gambar 1.



Gambar 1. Langkah Penelitian

1. Identifikasi Masalah

Banyaknya siswa lulusan SMP yang berkeinginan melanjutkan ke sekolah kejuruan atau SMK di kabupaten Magelang relatif tinggi. Oleh karena itulah perlu penataan dan pengaturan oleh pihak terkait yang disesuaikan dengan kondisi wilayah masing-masing, sehingga pemerintah dapat memantau kondisi sekolah, ketersediaan sarana prasarana di setiap sekolah SMK yang diimbangi dengan jumlah siswa di sekolah tersebut.

2. Studi Literatur

Pada tahapan ini, dilakukan pencarian artikel dari jurnal nasional atau internasional yang sesuai dengan tema penelitian kemudian mempelajari teknik dan metode yang diterapkan. Adapun topik dalam penelitian ini adalah penerapan metode optimasi clustering pada K-Means sebagai salah satu teknik pengelompokan data dalam data mining

3. Pengumpulan Data

Di tahap ini, pengumpulan data dilakukan dengan mengunduh data publik yang bisa diakses dari laman <https://dapo.kemdikbud.go>. Data tersebut mencakup beberapa tabel yaitu tabel siswa lulusan SMP, SMA, dan SMK tahun 2023-2024 semester ganjil dan genap di Kabupaten Magelang. Selain itu, data mengenai jumlah sekolah di Kabupaten Magelang juga diperlukan.

4. Penentuan Cluster

Pada proses penentuan *clustering*, dilakukan penilaian clustering menggunakan metode *Elbow*. Hal ini bertujuan untuk menetapkan kuantitas dari *cluster* optimal yang akan digunakan dalam perhitungan K-Means. Proses ini melibatkan perbandingan nilai untuk beberapa nilai k yang telah diuji dengan SSE (*Sum of Square Error*), sehingga akan diperoleh grafik berbentuk siku di titik tertentu (Jollyta et al., 2019). Nilai SSE ditentukan berdasarkan pada persamaan (1)

$$SSE = \sum_k^k = 1 \sum x \in S_k \|x_1 - C_k\|_2^2 \dots\dots\dots (1)$$

SSE *Sum of Square Error* merupakan nilai error dari jarak *centroid* yang telah ditetapkan untuk setiap k (Jollyta et al., 2019).

5. Perhitungan K-Means

Berdasarkan hasil perhitungan clustering pada tahap sebelumnya, diperoleh jumlah cluster optimal, yang selanjutnya dilakukan perhitungan dengan menggunakan metode K-Means. Perhitungan jumlah cluster di langkah sebelumnya, akan menghasilkan jumlah cluster optimal, kemudian nilai tersebut akan dijadikan sebagai nilai k dalam perhitungan K-Means. Langkah-langkah perhitungan dengan metode K-Means adalah sebagai berikut : (Firman Ashari et al., 2023)

1. Menempatkan objek pada bagian tertentu, berdasarkan perbedaan lokasi objek diantara keduanya. Persamaan yang digunakan untuk menentukan jarak terdekat adalah

$$d_{(x,y)} = \|x - y\|^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots\dots\dots (2)$$

Keterangan:

$d_{(x,y)}$ = jarak antara objek pertama (x) dengan objek kedua (y).

x = simbol untuk objek awal

y = simbol untuk objek berikutnya

2. Menetapkan *centroid* atau titik pusat untuk anggota *cluster* yang baru.
3. Melakukan langkah yang sama dengan point 1 sesuai proses dengan nilai *centroid* yang baru.
4. Jika anggota *cluster* tidak ada perubahan, maka proses berhenti.
5. Melakukan langkah yang sama dengan point 3 sampai dengan *centroid* atau pusat *cluster* tidak mengalami perubahan lagi.

6. Hasil Clustering

Melalui hasil perhitungan langkah-langkah sebelumnya akan dihasilkan kelompok atau cluster dikarenakan data kelompok dalam *cluster* tidak ada perubahan lagi.

HASIL DAN PEMBAHASAN

Berdasarkan uraian langkah dan proses dalam penelitian di atas, pada tahap ini adalah pembahasan terkait hasil dari proses penelitian tersebut.

1. Data yang digunakan

Adapun data diambil dari data publik yang bisa diakses dari <https://dapo.kemdikbud.go.id> 2023-2024 Genap dan Ganjil kabupaten Magelang.

2. Penentuan Clustering

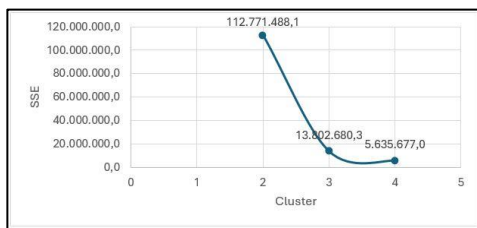
Agar jumlah *cluster* lebih optimal, banyaknya *clustering* ditentukan berdasarkan hasil perhitungan dengan metode *Elbow* melalui persamaan (1). Adapun hasil perhitungan dengan

metode *Elbow* ada di Tabel 1.

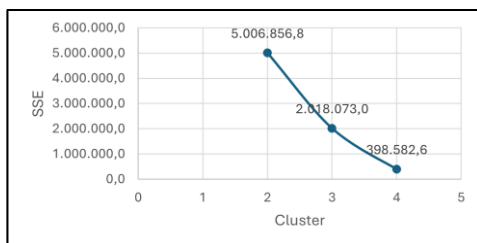
Tabel 1. Hasil Perhitungan metode Elbow

Uji Coba	Jml Data	Jumlah Cluster		
		2	3	4
1	21	112.771.488	13.802.680	5.635.677
2	15	5.006.857	2.018.073	398.583
3	17	12.086.316	4.670.726	767.806

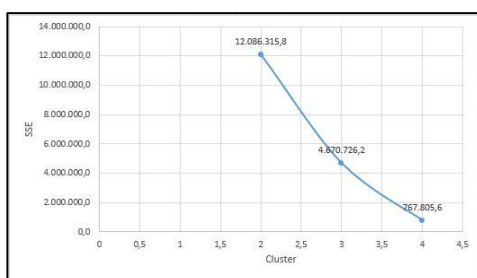
Merujuk Tabel 1, melakukan proses uji coba sebanyak 3 kali, banyaknya data uji yang digunakan adalah 21, 15 dan 17 data, sehingga hasil perhitungannya memberikan hasil yang berbeda-beda untuk setiap pengujianya. Sedangkan banyaknya *cluster* yang diujikan meliputi 2, 3 dan 4 *cluster*. Grafik perhitungan Elbow disajikan pada Gambar 2, Gambar 3 dan Gambar 4



Gambar 2 Grafik Hasil perhitungan *Elbow* Ujicoba1



Gambar 3 Grafik Hasil perhitungan *Elbow* Ujicoba2



Gambar 4 Grafik Hasil perhitungan *Elbow* Ujicoba3

Merujuk hasil grafik untuk pengujian 1, 2 dan 3 terlihat bahwa posisi *Elbow* yang berupa garis lengkung menyerupai lengan berada di titik angka 3, oleh karena itu banyaknya *cluster* optimal yang akan digunakan adalah 3.

3. Perhitungan K-Means

Langkah berikutnya, perhitungan dengan menerapkan algoritma K-Means. Adapun banyaknya *cluster* yang ditentukan adalah 3 atau $k=3$, sesuai dengan hasil langkah sebelumnya. Berdasarkan pada persamaan (2), berikut proses perhitungan *K-Means*.

Diawali dari penentuan data yang akan digunakan, kemudian ditentukan *centroid* awal, sehingga hasil disajikan di Tabel 2 dan Tabel 3.

Tabel 2. Data Perhitungan K-Means

Wilayah	Lulus SMP	Masuk SMA	Masuk SMK
Muntilan	3783	2637	3530
Mertoyudan	3378	3172	1740
Salaman	2580	1277	2248
Mungkid	2308	1115	1370
Grabag	1839	1946	545
Bandongan	2714	1131	1496
Tegalrejo	2402	707	1347
Secang	1962	340	1259
Salam	2162	86	3059
Borobudur	1587	129	963
Sawangan	1957	51	229
Tempuran	1768	61	453
Pakis	1500	261	0
Candimulyo	1458	898	96
Windusari	890	0	1230
Kaliangkrik	1016	134	106
Kajoran	1274	0	358
Ngablak	698	0	866
Dukun	1014	852	262
Srumbung	1730	67	0
Ngluwar	868	904	169

Semula ditentukan jumlah kelompok (k) adalah 3, kemudian diambil acak nilai *centroid* untuk 3 kelompok adalah seperti pada Tabel 3, yaitu Mertoyudan, Tempuran, dan Srumbung. Selanjutnya akan dihasilkan jarak untuk ketiga *centroid* tersebut di Tabel 4.

Tabel 3. Penentuan *Centroid* Awal untuk *Cluster* Awal

Centroid Awal	Lulus SMP	Masuk SMK	Masuk SMA
Mertoyudan	3378	3172	1740
Tempuran	1768	61	453
Srumbung	1730	67	0

Tabel 4. Hasil Perhitungan Jarak

C1	C2	C3
1910	4204	4771
0	3531	3730
2118	1518	2719
2348	1221	1887
2302	1834	1473

2160	1508	2285
2680	999	1944
3203	430	1644
3570	577	3270
3615	0	1322
3747	378	1385
3732	193	1265
3877	158	917
3400	780	595
4063	709	1394
4181	571	787
4049	339	1009
4244	898	1154
3627	923	181
3922	156	1213
3730	1057	0

Setelah diperoleh data jarak pada Tabel 4, langkah berikutnya adalah menghitung jarak yang terdekat diantara ketiganya. Adapun hasil yang diperoleh disajikan pada Tabel 5.

Tabel 5. Hasil Jarak Terdekat

C1	C2	C3
1		
1		
	2	
	2	
		3
	2	
	2	
	2	
	2	
	2	
	2	
	2	
	2	
	2	
		3
	2	
	2	
	2	
		3
	2	
		3

Merujuk dari data hasil perhitungan jarak terdekat diantara ketiga *centroid* (Mertoyudan, Tempuran, dan Srumbung), hasil yang diperoleh terdapat 2 wilayah untuk *cluster1* (C1), 15 wilayah untuk *cluster2* (C2), dan 4 wilayah untuk *cluster3* (C3). Perhitungan *K-Means* selanjutnya digunakan untuk memperoleh data kelompok yang jumlah anggotanya tidak mengalami perubahan alias tetap.

Perhitungan *K-Means* iterasi ke-dua menggunakan titik pusat (*centroid*) berdasarkan hasil

iterasi ke-satu, yaitu data C1, C2 dan C3 berikut.

Tabel 6. Data Centroid ke-dua

Centroid	Lulus SMP	Masuk SMK	Masuk SMA
C1	3578	2905	2635
C2	1770	357	999
C3	1295	1150	268

Tabel 7. Hasil Jarak Terdekat Centroid ke-dua

C1	C2	C3
1		
1		
	2	
	2	
		3
	2	
	2	
	2	
	2	
	2	
	2	
	2	
		3
		3
	2	
		3
	2	
	2	
		3
	2	
		3

Merujuk perhitungan pada data di Tabel 6, maka dihasilkan Tabel 7 sebagai hasil jarak terdekat untuk *centroid* ke-dua. Apabila proses perhitungan di atas menghasilkan jarak terdekat yang sama, maka proses akan berhenti di iterasi tersebut.

Terlihat Tabel 8 bahwa hasil perhitungan di iterasi empat tidak ada perubahan dalam jumlah anggota kelompoknya, oleh karena itu proses perhitungan *K-Means* dihentikan.

Tabel 8. Hasil Iterasi

Iterasi ke-	Banyaknya data		
	C1	C2	C3
1	2	15	4
2	2	13	6
3	2	10	9
4	2	10	9

4. Hasil Pengelompokan

Dari semua langkah perhitungan di atas, dihasilkan bahwa jumlah optimal *cluster* adalah 3, untuk *cluster1* diperoleh data 2 kecamatan, untuk *cluster2* didapatkan hasil terdapat 10 kecamatan, dan untuk *cluster3* diperoleh hasil terdapat 9 kecamatan. Hasil tersebut tersaji pada Tabel 9.

Tabel 9. Data Wilayah Kecamatan Per Cluster

Wilayah/Kecamatan		
Cluster1	Cluster2	Cluster3
Muntilan	Salaman	Mungkid
Mertoyudan	Tegalrejo	Grabag
	Secang	Bandongan
	Salam	Sawangan
	Borobudur	Pakis
	Tempuran	Candimulyo
	Windusari	Dukun
	Kaliangkrik	Srumbung
	Kajoran	Ngluwar
	Ngablak	
	Salaman	

KESIMPULAN

Berdasarkan semua tahapan dalam penelitian tentang jumlah *cluster* optimal untuk pengelompokan jumlah siswa SMK di Kabupaten Magelang diperoleh hasil antara lain:

Hasil perhitungan dengan metode *Elbow* didapatkan bahwa banyaknya *cluster* optimal adalah 3. Keluaran dari proses perhitungan dengan menerapkan algoritma K-Means adalah terdapat 3 *cluster*, diperoleh 2 wilayah yaitu Mertoyudan, Muntilan di *cluster1*, di *cluster2* adalah 10 wilayah yaitu, Salaman, Tegalrejo, Secang, Salam, Borobudur, Tempuran, Windusari, Kaliangkrik, Kajoran dan Ngablak sedangkan di *cluster3* adalah 9 wilayah yaitu Mungkid, Grabak, Bandongan, Sawangan, Pakis, Candimulyo, Dukun, Srumbung, dan Ngluwar.

Pembagian kategori pada setiap *cluster* ditentukan berdasarkan kecenderungan siswa untuk memilih melanjutkan studi ke Sekolah Menengah Kejuruan (SMK). *Cluster1* terdapat 2 wilayah yang terdefiniskan dengan rendahnya antusiasme siswa dalam pemilihan sekolah menengah kejuruan sebagai pilihannya, *cluster3* terdiri dari 9 wilayah terdefiniskan minat siswa yang cukup besar untuk melanjutkan ke SMK, sedangkan *cluster2* merupakan tingginya antusiasme siswa dalam pemilihan sekolah menengah kejuruan di 10 wilayah.

Berdasarkan hasil perhitungan *cluster* tersebut, sangat penting dalam mendukung kinerja pemerintah daerah Kabupaten Magelang dalam sistem pemantauan distribusi jumlah siswa Sekolah Menengah Kejuruan di setiap kecamatan. Dengan demikian, akan lebih mudah untuk mendeteksi

kecamatan mana yang memiliki jumlah siswa yang kurang atau bahkan terlalu banyak. Aspek ini penting untuk diperhatikan karena berkaitan langsung dengan ketersediaan sarana prasarana dan tenaga pengajar di sekolah-sekolah menengah tersebut.

REFERENSI

- Firman Ashari, I., Dwi Nugroho, E., Baraku, R., Yanda, I. N., & Liwardana, R. (2023). Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 7, Issue 1). <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Hartanti, N. T. (2022). Mengukur Tingkat Pemahaman Mahasiswa pada Mata Kuliah Pemrograman dengan Algoritma K-Means Clustering Measuring Students' Level of Understanding in Programming Courses with the K-Means Clustering Algorithm. *Jurnal Sisfotenika*, 12(1), 62–73.
- Jollyta, D., Efendi, S., Zarlis, M., & Mawengkang, H. (2019). *Prosiding Seminar Nasional Riset Information Science (SENARIS) Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index*.
- Juanita, S., & Cahyono, R. D. (2024). K-Means Clustering With Comparison of Elbow and Silhouette Methods for Medicines Clustering Based on User Reviews. *Jurnal Teknik Informatika (JUTIF)*, 5(1), 283–289. <https://doi.org/10.52436/1.jutif.2024.5.1.1349>
- Maulana, R., Azmi, F., & Daru Kusuma, P. S. (n.d.). *PENGELOMPOKAN DATA SISWA DI INDONESIA MENGGUNAKAN K-MEANS CLUSTERING STUDENT DATA GROUPING IN INDONESIA USING K-MEANS CLUSTERING*.
- Nainggolan, R., Perangin-Angin, R., Simarmata, E., & Tarigan, A. F. (2019). Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method. *Journal of Physics: Conference Series*, 1361(1). <https://doi.org/10.1088/1742-6596/1361/1/012015>
- Palacios, C. A., Reyes-Suárez, J. A., Bearzotti, L. A., Leiva, V., & Marchant, C. (2021). Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile. *Entropy*, 23(4), 1–23. <https://doi.org/10.3390/e23040485>
- Sammouda, R., & El-Zaart, A. (2021). An Optimized Approach for Prostate Image Segmentation Using K-Means Clustering Algorithm with

- Elbow Method. *Computational Intelligence and Neuroscience*, 2021.
<https://doi.org/10.1155/2021/4553832>
- Sholeh, M., & Aeni, K. (n.d.). *STRING (Satuan Tulisan Riset dan Inovasi Teknologi) PERBANDINGAN EVALUASI METODE DAVIES BOULDIN, ELBOW DAN SILHOUETTE PADA MODEL CLUSTERING DENGAN MENGGUNAKAN ALGORITMA K MEANS*.
<https://archive.ics.uci.edu/ml/datasets/Trave>
- Umargono, E., Suseno, J. E., & Gunawan, V. (2020). *K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula*.
- Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1).
<https://doi.org/10.1186/s40561-022-00192-z>