

ANALISIS KLASTERISASI DATA SEKOLAH SMP DI KABUPATEN KOTAWARINGIN TIMUR MENGGUNAKAN METODE K-MEANS

Hana Silviana Sutanto¹⁾, Imelda Susilawaty²⁾

^{1, 2)}Universitas Darwan Ali, Jurusan Sistem Informasi Fakultas Ilmu Komputer
Jl. Batu Berlian No.10 Kalimantan Tengah, Kotawaringin Timur

han.silviana07@gmail.com¹⁾, susilawatyimelda24@gmail.com²⁾

* Korespondensi: e-mail: susilawatyimelda24@gmail.com

ABSTRAK

Dalam penelitian ini, metode K-Means digunakan untuk menilai klasterisasi data sekolah menengah pertama (SMP) di Kabupaten Kotawaringin Timur. Data tentang 112 SMP yang ada di wilayah tersebut tersedia di website Dapodik Kemendikbud. Setelah proses preprocessing selesai, data yang sesuai dan legal dipilih untuk dianalisis. Metode K-Means mengelompokkan sekolah menjadi berbagai klaster berdasarkan hal-hal yang serupa. Hasil analisis menunjukkan bahwa data SMP di Kabupaten Kotawaringin Timur terdiri dari tiga klaster penting. Klaster 0 memiliki 87 sekolah, klaster 1 memiliki 4 sekolah, dan klaster 2 memiliki 21 sekolah. Klaster 1 dan klaster 2 mendapat peringkat terbaik dalam hal kualitas pendidikan, sementara klaster 0 mendapat peringkat ketiga. Nilai $k=3$ adalah nilai yang tepat untuk jumlah klaster yang ideal, seperti yang ditunjukkan oleh penggunaan metode elbow dalam analisis. Analisis ini membantu pemangku kepentingan, termasuk pemerintah daerah, sekolah, dan masyarakat, memahami perbedaan di antara sekolah SMP dan membuat rencana yang tepat untuk meningkatkan kualitas pendidikan di daerah tersebut. Kami menggunakan perangkat lunak open source RapidMiner untuk melakukan analisis data dan klasterisasi dalam penelitian ini. Kelebihan perangkat lunak open source adalah kemampuan untuk disesuaikan dengan berbagai situasi dan kemampuan pengguna untuk bekerja sama satu sama lain. Hasilnya menunjukkan bahwa analisis klasterisasi yang dilakukan dengan metode K-Means membantu kami memahami pola dan kelompok sekolah di Kabupaten Kotawaringin Timur berdasarkan ciri-ciri tertentu. Dengan memahami dengan lebih baik perbedaan dan kualitas sekolah, pemangku kepentingan dapat membuat keputusan yang lebih baik tentang cara meningkatkan pendidikan di wilayah mereka. Penelitian ini memberikan perspektif yang bermanfaat bagi mereka yang membuat keputusan pendidikan.

Kata Kunci: Data Mining, K-Means, Rapidminer, Cluster

ABSTRACT

In this study, the K-means method was used to assess the clustering of junior secondary school (SMP) data in East Kotawaringin district. Data on 112 junior high schools in the region is available on the Dapodik website of the Ministry of Education and Culture. After the preprocessing process was completed, suitable and legal data were selected for analysis. The K-Means method groups schools into various clusters based on similarities. The results of the analysis showed that the junior secondary school data in East Kotawaringin district consisted of three important clusters. Cluster 0 has 87 schools, cluster 1 has 4 schools, and cluster 2 has 21 schools. Cluster 1 and Cluster 2 are ranked the best in terms of education quality, while Cluster 0 is ranked third. The value $k = 3$ is the right value for the ideal number of clusters, as shown by the use of the elbow method in the analysis. This analysis helps stakeholders, including local governments, schools, and communities, understand the differences among junior secondary schools and make appropriate plans to improve education quality in the area. We used the open-source software RapidMiner to conduct the data analysis and clustering in this study. The advantages of open source software are its ability to be adapted to various situations and the ability of users to cooperate with each other. The results show that the clustering analysis conducted using the K-Means method helps us understand the patterns and groups of schools in East Kotawaringin Regency based on certain characteristics. By better understanding the differences and quality of schools, stakeholders can make better decisions on how to improve education in their areas. This research provides a useful perspective for those who make education decisions.

Keywords: Data Mining, K-Means, Rapidminer, Cluster.

I. PENDAHULUAN

Hak untuk mendapatkan pendidikan yang layak adalah hak warga negara Indonesia. Fasilitas yang memadai di sekolah menjadi sangat penting untuk mencapai pendidikan yang layak. Di Indonesia, semua jenjang pendidikan telah didistribusikan ke berbagai wilayah kabupaten untuk memastikan pendidikan yang sama di seluruh wilayah. Pemerintah kabupaten mendukung upaya ini dengan mempertimbangkan perkembangan jumlah penduduk di sekitar wilayah tersebut[1].

Salah satu daerah di Indonesia yang memiliki banyak sekolah menengah pertama (SMP) adalah Kabupaten Kotawaringin Timur. Untuk mengelola pendidikan, pemerintah dan pemangku kepentingan harus memahami pola-pola yang ada dalam data sekolah menengah pertama di wilayah tersebut. Klasterisasi data adalah teknik analisis yang bertujuan untuk mengelompokkan objek data ke dalam kelompok-kelompok atau klaster-klaster yang memiliki karakteristik yang sama[2]. Salah satu cara untuk menemukan pola-pola ini adalah dengan menggunakan teknik ini. Untuk data sekolah SMP di Kabupaten Kotawaringin Timur, metode klasterisasi *K-Means* dipilih.

Dengan menggunakan analisis klasterisasi data sekolah SMP[3], pemerintah dan pemangku kepentingan pendidikan dapat memahami perbedaan di antara sekolah-sekolah SMP dan merencanakan strategi yang sesuai untuk meningkatkan kualitas pendidikan di daerah mereka.

Selain itu, untuk melakukan analisis klasterisasi, digunakan RapidMiner, salah satu program yang paling populer untuk analisis data dan penambangan data (*data mining*), yang memiliki banyak fitur dan algoritma, termasuk metode *K-Means* yang dapat digunakan untuk klasterisasi. Dengan algoritma klasterisasi yang disertakan dalam aplikasi ini, peneliti dapat melakukan proses klasterisasi data sekolah SMP dengan mudah dan efisien.

Dalam penelitian ini, peneliti akan menggunakan metode *K-Means* untuk menganalisis klasterisasi data sekolah SMP di Kabupaten Kotawaringin Timur. Kabupaten Kotawaringin Timur dapat dianggap sebagai populasi atau kelompok besar yang terdiri dari beberapa sekolah SMP.

Studi ini bertujuan untuk menentukan pola dan kelompok sekolah yang mungkin ada di Kabupaten Kotawaringin Timur berdasarkan fitur sekolah tersebut. Peneliti akan mengelompokkan sekolah-sekolah menjadi beberapa klaster berdasarkan kesamaan karakteristik menggunakan teknik *K-Means*. Dalam upaya meningkatkan kualitas pendidikan di Kabupaten Kotawaringin Timur, tujuan utamanya adalah untuk memberikan informasi yang lebih mendalam dan wawasan yang bermanfaat kepada pihak terkait, seperti pemerintah daerah, sekolah, dan masyarakat.

II. LANDASAN TEORI

A. *K-Means*

Stuart Lloyd menciptakan algoritma *K-Means* pada tahun 1982 dan telah menjadi salah satu algoritma *clustering* yang paling populer. *K-Means* membagi item yang ada ke dalam kelompok atau segmen, di mana item dalam satu kelompok memiliki kemiripan yang lebih besar satu sama lain.

Mengoptimalkan pengelompokan objek data dalam setiap klaster dengan meminimalkan varians total dari setiap klaster adalah prinsip utama algoritma *K-Means*. Akibatnya, objek-objek yang berada dalam satu klaster memiliki kemiripan yang tinggi, sedangkan objek-objek yang berada dalam klaster yang berbeda memiliki perbedaan yang signifikan[4].

Salah satu kelemahan algoritma *K-Means* adalah sensitif terhadap inisialisasi awal yang acak dan rentan terhadap konvergensi ke solusi lokal yang suboptimal[5]. Akibatnya, pengujian dilakukan dengan banyak inisialisasi awal untuk menentukan hasil terbaik.

Dalam algoritma *K-Means*, nilai *K*, yang diwakili oleh jumlah klaster yang dibutuhkan, merupakan unit analisis. Nilai *K* ditetapkan secara acak atau berdasarkan pertimbangan tertentu. Setiap kelompok data memiliki *centroid*, atau rata-rata, yang digunakan untuk menentukan posisi pusatnya. Jarak antara *centroid* terdekat setiap kelompok data dapat dihitung dengan menggunakan rumus geometri konvensional[6]. Adapun rumus *K-Means* dapat dilihat pada penerapan persamaan 1.

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

B. Elbow Method

Dalam analisis klaster, metode siku membantu menentukan berapa banyak klaster yang tepat untuk digunakan dalam suatu model. Metode ini biasanya digunakan ketika kita ingin mengklasifikasikan data ke dalam kelompok-kelompok yang berbeda berdasarkan kemiripan karakteristik mereka. Grafik hubungan kelompok dengan kesalahan yang menurun menunjukkan nilai K pada kombinasi Elbow dengan K-Means. Setiap kali nilai K meningkat, grafik akan menurun secara bertahap hingga mencapai hasil nilai K yang stabil[7].

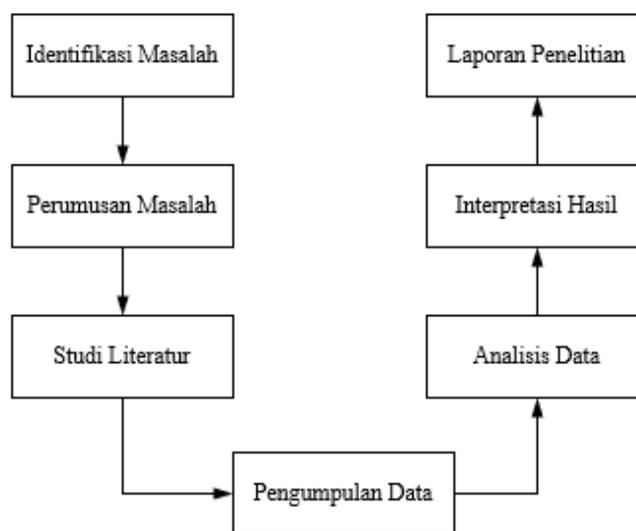
C. Rapidminer

RapidMiner adalah platform perangkat lunak open source, terbuka untuk siapa saja [8]. Karena itu, kode sumbernya dapat diakses dan diubah oleh pengguna sesuai kebutuhan. *RapidMiner* adalah alat untuk data mining dan analisis prediktif yang memiliki banyak fitur dan algoritma untuk melakukan berbagai tugas, seperti pemrosesan data, eksplorasi data, visualisasi, pemodelan prediktif, dan evaluasi model. Dengan menggunakan platform ini, pengguna dapat menggali wawasan dari data, menemukan pola atau tren, membuat prediksi masa depan, dan menggunakan analisis data untuk membuat keputusan[9]. Kelebihan *RapidMiner* sebagai perangkat lunak open source adalah kemampuannya untuk disesuaikan dengan berbagai situasi[10]. Perangkat lunak dapat disesuaikan dengan kebutuhan pengguna, dan mereka juga dapat membuat ekstensi atau komponen tambahan untuk meningkatkan fungsionalitasnya. Selain itu, sifat terbuka *RapidMiner* memungkinkan komunitas pengguna bekerja sama dan berbagi informasi[11].

III. METODOLOGI PENELITIAN

Proses yang diperlukan untuk merencanakan, merancang, melaksanakan, dan menganalisis penelitian termasuk dalam kerangka kerja yang dikenal sebagai metodologi penelitian, yang mencakup kumpulan langkah-langkah yang diperlukan untuk melakukan penelitian secara sistematis. Metodologi penelitian membantu peneliti mendapatkan data yang relevan dan valid dan membuat kesimpulan yang dapat diandalkan.

Mengidentifikasi masalah, membuat perumusan masalah, melakukan penelitian literatur, mengumpulkan data, menganalisis data, menginterpretasikan temuan, dan menyusun laporan penelitian adalah kerangka penelitian.



Gambar 1. Kerangka Penelitian

Untuk penjelasan deksripsi dari kerangka penelitian, ialah:

1. Peneliti harus menemukan dan merumuskan masalah penelitian yang relevan, jelas, dan terdefinisi dengan baik.
2. Tujuan penelitian harus didefinisikan dengan jelas. Tujuan ini menentukan tujuan penelitian dan berfungsi sebagai landasan untuk pembuatan metodologi yang tepat.
3. Peneliti melakukan penelitian literatur untuk mengevaluasi penelitian sebelumnya yang berkaitan dengan subjek tersebut. Ini membantu menemukan celah pengetahuan yang ada, membangun fondasi teoritis, dan mencegah penelitian yang tidak perlu diulangi.
4. Peneliti mengumpulkan data sesuai dengan desain penelitian. Ini bisa mencakup penggunaan observasi, wawancara, kuesioner, atau analisis dokumen. Data dikumpulkan dengan melakukan pencarian di situs web dapodik.

5. Untuk menemukan pola, tren, dan hubungan antara variabel, metode dan teknik analisis yang tepat digunakan untuk menganalisis data yang dikumpulkan. Ini termasuk penyusunan, pengkodean, dan pengolahan data.
6. Untuk menjawab pertanyaan penelitian, hasil analisis data dievaluasi dan ditafsirkan. Kesimpulan dibuat berdasarkan hasil penelitian dan dibandingkan dengan kerangka teoritis yang telah dikembangkan.
7. Peneliti harus menulis laporan penelitian dengan jelas, sistematis, dan mudah dipahami. Laporan harus mencakup pendahuluan, metodologi yang digunakan, hasil, pembahasan, saran, dan kesimpulan.

IV. HASIL DAN PEMBAHASAN

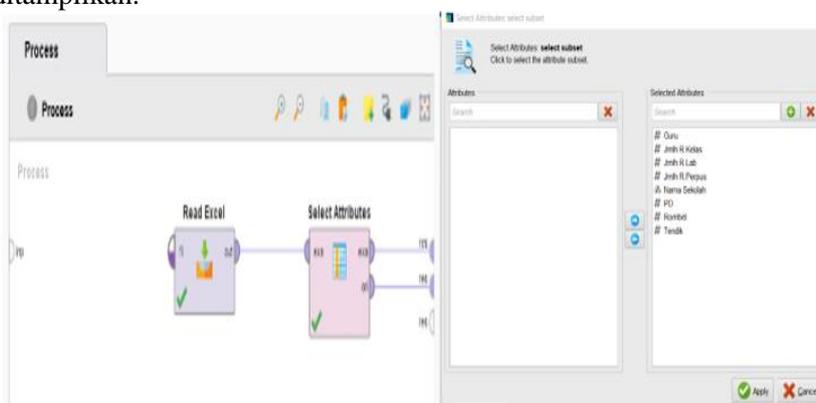
Dalam metodologi penelitian, penelitian dilakukan dalam beberapa tahapan. Pertama, masalah diidentifikasi dengan melihat masalah di dunia pendidikan di kabupaten Kotawaringin Timur. Pengamatan menunjukkan masalah ketidakseimbangan fasilitas. Sekolah dan kebutuhan lain yang disediakan oleh pemerintah. Setelah kasus diteliti secara teoritis dan ilmiah, literatur dipelajari untuk menambah informasi dalam pengamatan. Setelah kasus diketahui secara menyeluruh, dilakukan perumusan masalah untuk memfokuskan penelitian pada masalah yang diteliti, seperti kesetaraan fasilitas pendidikan dan kebutuhan lainnya. Hasil studi literatur menunjukkan bahwa analisis data mining dengan metode *clustering* diperlukan untuk mengelompokkan satuan pendidikan berdasarkan klasternya.

A. Dataset

Dataset ini berasal dari website dapodik kemendikbud kabupaten Kotawaringin Timur. Pada situs web dapodik, data awal hasil pilihan memiliki 9 fitur, yaitu sebagai berikut: : Nama Sekolah, Status, PD, Rombel, Guru, Tendik, Jmlh R.Kelas, Jmlh R.Lab, Jmlh R.Perpus. Dataset memiliki rekam data sebanyak 112 record.

B. Selections

Setelah data dikumpulkan, seleksi dilakukan untuk memenuhi kebutuhan penelitian. Sebagian besar atribut yang diolah oleh Website Dapodik Kemendikbud kabupaten Kotawaringin Timur tidak digunakan dalam proses pengolahan data. sehingga proses pemilihan atribut diperlukan. Operator Select Attributes digunakan untuk menyelesaikan proses pemilihan melalui aplikasi rapid miner. Gambar 2 menunjukkan bagaimana pilihan di aplikasi rapid miner ditampilkan.



Gambar 2. menunjukkan cara menampilkan opsi aplikasi quick miner.

Gambar2 menunjukkan bagaimana atribut yang akan digunakan untuk pengolahan data dipilih. Dalam pengolahan data, sembilan atribut dipilih, yaitu Nama Sekolah, Status, PD, Rombel, Guru, Tendik, Jmlh R.Kelas, Jmlh R.Lab, Jmlh R.Perpus. Atribut Nama Satuan Pendidikan diubah menjadi tipe identitas sehingga tidak diperhitungkan saat memproses data karena atribut ini adalah atribut objek penelitian untuk mengetahui hasil *clustering*.

C. Preprocessing

Preprocessing mengubah data yang tidak konsisten atau rusak menjadi data yang dapat diolah, serta mengurangi data yang sulit dipahami. Ini dapat dilakukan dengan menghilangkan duplikat, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan. *Preprocessing* data ditunjukkan pada Gambar 3.



Gambar 3. Preprocessing

Pada Gambar 3, preprocessing data dilakukan dengan menggunakan operator pengganti nilai yang tidak ada dan contoh filter operator. Proses ini menghapus record yang tidak memiliki tujuan yang ditemukan atau data yang sulit dipahami dan memperbaiki data yang tidak konsisten. Tahapan preprocessing menyelesaikan 114 dataset semula, dan data yang valid saat ini adalah 112 rekaman.

D. Data Mining

Dataset yang telah melewati tahapan preprocessing dan transformasi sekarang siap untuk tahapan berikutnya, mining process. Saat ini, ada 112 record yang valid dengan 9 atribut, dan Dataset yang valid atau siap untuk proses mining ditunjukkan dalam Tabel 1.

Tabel 1. Dataset valid

Nama Sekolah	Stat us	P D	Rom bel	Gu ru	Ten dik	Jmlh R.Kelas	Jmlh R.Lab	Jmlh R.Perpus
SMP NEGERI 1 ANTANG KALANG	Neg eri	16 2	6	11	2	7	2	1
SMP NEGERI 2 ANTANG KALANG	Neg eri	10 6	5	7	1	6	1	1
SMP NEGERI 3 ANTANG KALANG	Neg eri	34	3	8	1	3	0	1
SMP NEGERI 3 MENTAYA HULU	Neg eri	16 0	6	7	2	5	1	1
SMP NEGERI SATU ATAP 1 ANTANG KALANG	Neg eri	32	3	7	1	3	0	1
SMP NEGERI SATU ATAP 2 ANTANG KALANG	Neg eri	11 1	4	5	1	2	1	0
SMP NEGERI SATU ATAP 3 ANTANG KALANG	Neg eri	61	3	3	1	3	0	0
SMP SWASTA IT AL-ISHLAH WARINGIN AGUNG	Swa sta	35	3	3	1	3	0	0
SMP NEGERI 11 SAMPIT	Neg eri	30 1	8	17	2	6	2	1
SMP NEGERI 3 SAMPIT	Neg eri	87 6	26	45	12	24	4	1
SMP NEGERI 9 SAMPIT	Neg eri	55 7	16	30	6	16	2	2
SMP NEGERI SATU ATAP 1 BAAMANG	Neg eri	73	3	11	1	3	0	0
SMP NEGERI SATU ATAP 2 BAAMANG	Neg eri	76	3	9	3	3	1	1
SMP PGRI 4 SAMPIT	Swa sta	14 1	5	9	1	8	1	1
SMP PLUS AL BUSTHOMI NAHDLATUL WATHAN	Swa sta	17 2	6	12	1	6	0	1

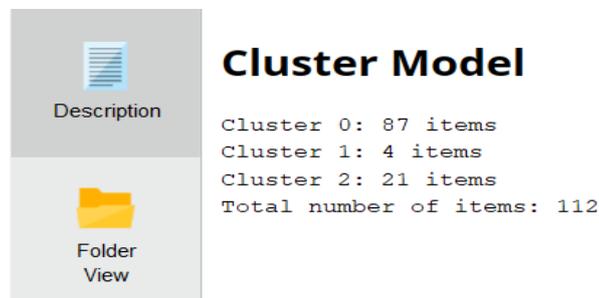
Nama Sekolah	Stat us	P D	Rom bel	Gu ru	Ten dik	Jmlh R.Kelas	Jmlh R.Lab	Jmlh R.Perpus
SMPS IT DARUL MA RIFAH	Swa sta	25 0	6	8	1	6	1	1
SMP NEGERI 7 MENTAYA HULU	Neg eri	15 4	6	6	1	6	1	1
SMP NEGERI 8 MENTAYA HULU	Neg eri	27	3	6	1	3	1	1
SMP SEKAR SARI SARPATIM	Swa sta	11 4	5	6	2	5	2	1
SMPS EKA TJIPTA KUAYAN	Swa sta	21 2	7	15	2	8	2	1
SMP NEGERI 1 CEMPAGA	Neg eri	42 1	13	22	7	13	2	1
SMP NEGERI 4 CEMPAGA	Neg eri	18 6	7	19	4	8	1	1
SMP NEGERI 6 CEMPAGA	Neg eri	10 8	4	10	2	6	2	1
SMP NEGERI 7 CEMPAGA	Neg eri	25 4	9	20	4	11	1	1
SMP NEGERI 2 CEMPAGA	Neg eri	32 5	11	19	5	12	2	1
SMP NEGERI 3 CEMPAGA	Neg eri	33 8	9	18	3	8	1	1
SMP NEGERI 5 CEMPAGA	Neg eri	16 5	6	11	3	6	3	1
SMP NEGERI SATU ATAP 1 CEMPAGA HULU	Neg eri	69	3	10	2	3	1	0
SMP NEGERI SATU ATAP 2 CEMPAGA HULU	Neg eri	40	3	4	1	2	0	0
SMP BUMITAMA CEMPAGA HULU	Swa sta	20 1	6	8	2	6	2	1
SMP BUMITAMA METRO CEMPAGA	Swa sta	16 5	6	10	2	6	2	1
SMP NEGERI 1 KOTA BESI	Neg eri	41 0	14	24	5	14	2	1
SMP NEGERI 2 KOTA BESI	Neg eri	14 5	6	8	2	3	5	1
SMP NEGERI 4 KOTA BESI	Neg eri	65	3	7	3	6	1	1
SMP NEGERI 8 KOTA BESI	Neg eri	37	3	8	1	3	0	1
SMP NEGERI SATU ATAP 1 KOTA BESI	Neg eri	50	3	3	2	3	0	1
SMPS 08 BEST AGRO	Swa sta	11 8	6	7	1	3	0	1
SMP NEGERI 1 SAMPIT	Neg eri	92 5	26	47	9	24	5	1
SMP NEGERI 10 SAMPIT	Neg eri	18 9	7	11	2	7	1	1
SMP NEGERI 2 SAMPIT	Neg eri	81 0	23	44	9	23	3	1
SMP NEGERI 4 SAMPIT	Neg eri	77 0	23	38	9	23	3	1
SMP NEGERI 7 SAMPIT	Neg	25	9	15	6	9	1	1

Nama Sekolah	Stat us	P D	Rom bel	Gu ru	Ten dik	Jmlh R.Kelas	Jmlh R.Lab	Jmlh R.Perpus
	eri	8						
SMP NEGERI 8 SAMPIT	Neg eri	44 3	12	23	3	11	1	1
SMP DARUL AMIN	Swa sta	54 1	18	7	1	14	1	1
SMP IDRISIA	Swa sta	13 0	4	9	2	3	2	1
SMP ISLAM TERPADU ARAFAH	Swa sta	27 5	9	11	1	11	3	1
SMP KATHOLIK AL BERTUS SAMPIT	Swa sta	33 4	12	17	2	12	3	1
SMP KRISTEN SAMPIT	Swa sta	38	3	9	1	3	1	1
SMP MUHAMMADIYAH SAMPIT	Swa sta	23 4	8	15	1	8	1	1
SMP PGRI 1 SAMPIT	Swa sta	75	3	5	1	3	1	1
SMP SWASTA CITA BUNDA JHS SAMPIT	Swa sta	90	3	15	1	3	1	1
SMPS ASSAADAAT	Swa sta	66	2	10	1	4	0	0
SMP NEGERI 1 MENTAYA HILIR SELATAN	Neg eri	23 6	9	17	7	16	2	1
SMP NEGERI 2 MENTAYA HILIR SELATAN	Neg eri	18 0	6	11	5	6	2	1
SMP NEGERI 3 MENTAYA HILIR SELATAN	Neg eri	55	3	7	3	3	1	1
SMP Islam Terpadu Al Madaniyah Samuda	Swa sta	13 9	6	16	3	6	0	1
SMP NEGERI 1 MENTAYA HILIR UTARA	Neg eri	34 2	12	21	6	13	2	2
SMP NEGERI 2 MENTAYA HILIR UTARA	Neg eri	42	3	6	1	6	1	1
SMP NEGERI 3 MENTAYA HILIR UTARA	Neg eri	96	3	7	3	6	1	1
SMP SWASTA BINA BANGSA 01	Swa sta	42 1	14	18	2	14	1	1
SMP TUNAS PERTIWI	Swa sta	16 1	6	9	5	6	1	1
SMP NEGERI 1 MENTAYA HULU	Neg eri	27 1	10	21	3	12	1	1
SMP NEGERI 4 MENTAYA HULU	Neg eri	12 9	6	10	4	6	1	1
SMP NEGERI 5 MENTAYA HULU	Neg eri	82	3	4	1	6	2	1
SMP NEGERI 6 MENTAYA HULU	Neg eri	12 3	6	11	1	3	1	2
SMP NEGERI SATU ATAP 2 MENTAYA HULU	Neg eri	91	3	10	2	3	0	0
SMP KRIDATAMA LANCAR I	Swa sta	85	3	3	2	3	0	1
SMP PGRI 2 TEGUH SEMPURNA	Swa sta	66	3	2	1	3	0	0
SMP PGRI TEGUH SAMPURNA	Swa sta	13 3	6	8	2	9	2	1

Nama Sekolah	Stat us	P D	Rom bel	Gu ru	Ten dik	Jmlh R.Kelas	Jmlh R.Lab	Jmlh R.Perpus
SMP SWASTA 07 BEST AGRO	Swa sta	71	3	7	1	3	1	1
SMP NEGERI 1 PARENGGEAN	Neg eri	35 9	12	23	1	12	1	1
SMP NEGERI 2 PARENGGEAN	Neg eri	21 3	8	15	3	8	1	1
SMP NEGERI 3 PARENGGEAN	Neg eri	15 1	6	11	3	10	2	1
SMP NEGERI 4 PARENGGEAN	Neg eri	11 1	3	4	1	6	1	1
SMP NEGERI 6 PARENGGEAN	Neg eri	57	3	8	1	3	1	1
SMP NEGERI 7 PARENGGEAN	Neg eri	16 4	6	11	3	6	1	1
SMP NEGERI 8 PARENGGEAN	Neg eri	85	4	9	1	6	1	1
SMPN SATAP 1 PARENGGEAN	Neg eri	84	3	5	2	4	0	0
SMP WIJAYA KUSUMA	Swa sta	16 2	6	12	3	6	2	1
SMP NEGERI 1 PULAU HANAUT	Neg eri	62	3	9	3	4	1	1
SMP NEGERI 2 PULAU HANAUT	Neg eri	69	3	7	2	4	2	1
SMP NEGERI 3 PULAU HANAUT	Neg eri	87	5	8	4	6	1	1
SMP NEGERI SATU ATAP 1 PULAU HANAUT	Neg eri	37	3	2	1	3	0	0
SMP NEGERI SATU ATAP 2 PULAU HANAUT	Neg eri	45	3	3	1	3	0	0
SMP NEGERI SATU ATAP 3 PULAU HANAUT	Neg eri	89	3	7	2	3	0	0
SMP NEGERI 5 SAMPIT	Neg eri	19 8	6	10	2	6	1	2
SMP NEGERI 6 SAMPIT	Neg eri	18 3	6	13	2	10	2	1
SMP NEGERI SATU ATAP 1 SERANAU	Neg eri	10 2	5	6	1	3	0	0
SMP NEGERI SATU ATAP 2 SERANAU	Neg eri	12	3	6	1	3	1	1
SMP MERANTI MUSTIKA	Swa sta	45	3	4	1	3	0	0
SMP NEGERI 2 MENTAYA HULU	Neg eri	19 5	7	15	2	12	3	1
SMP NEGERI 4 ANTANG KALANG	Neg eri	18 0	6	10	1	6	1	1
SMP NEGERI SATU ATAP 1 MENTAYA HULU	Neg eri	68	3	5	1	3	0	1
SMP NEGERI SATU ATAP 4 ANTANG KALANG	Neg eri	35	3	7	1	3	0	0
SMP PGRI BUKIT INDAH	Swa sta	13 2	3	6	1	3	0	0
SMPS 1 BUMITAMA ANTANG	Swa	17	6	11	3	6	2	1

Nama Sekolah	Stat us	P D	Rom bel	Gu ru	Ten dik	Jmlh R.Kelas	Jmlh R.Lab	Jmlh R.Perpus
KALANG	sta	5						
SMP NEGERI 1 TELAWANG	Neg eri	46	2	6	1	2	0	0
SMP NEGERI 3 KOTA BESI	Neg eri	21 5	7	16	3	9	1	1
SMP NEGERI 5 KOTA BESI	Neg eri	31 1	10	16	2	9	1	1
SMP NEGERI 6 KOTA BESI	Neg eri	25	3	3	1	3	0	1
SMP NEGERI 7 KOTA BESI	Neg eri	77	3	7	3	6	1	1
SMP HAMPARAN 2	Swa sta	11 2	5	9	1	6	2	1
SMP PGRI TANAH PUTIH	Swa sta	20	3	1	1	3	0	0
SMPS ANWAR KARIM IV	Swa sta	12 0	4	7	1	4	1	1
SMPS BINA BANGSA 02	Swa sta	26 4	8	11	2	9	2	1
SMP NEGERI 1 TELUK SAMPIT	Neg eri	96	3	7	3	7	1	1
SMP NEGERI 2 TELUK SAMPIT	Neg eri	54	3	7	3	6	2	1
SMP NEGERI SATAP 2 TELUK SAMPIT	Neg eri	10 5	4	8	1	3	0	0
SMP NEGERI SATU ATAP 1 TELUK SAMPIT	Neg eri	64	3	7	1	3	0	0
SMP NEGERI 5 PARENGGEAN	Neg eri	28	3	8	1	6	1	1
SMP NEGERI SATU ATAP 2 PARENGGEAN	Neg eri	25	3	3	1	3	0	0
SMPS TUNAS AGRO SUBUR KENCANA	Swa sta	18 5	6	11	3	8	2	1

Semua data telah dirubah ke angka numerik, seperti yang ditunjukkan dalam Tabel 2 dataset valid. Dengan demikian, alat pengurangan cepat dapat digunakan untuk menerapkan algoritma K-Means. Setelah mengetahui bahwa nilai $k=3$, dataset yang diproses dibagi menjadi tiga *cluster*. Pembagian dari tiga *cluster* ini menghasilkan *cluster* model, seperti yang ditunjukkan pada Gambar 4.



Gambar 4. Model *Cluster* Pengelompokan Sekolah di Kotawaringin Timur

Keanggotaan masing-masing kelompok ditunjukkan pada Gambar 4. Kelompok 0 memiliki 87 item, yang menunjukkan bahwa ada 87 sekolah yang tergabung dalam satu kelompok; Kelompok 1 memiliki 4 item, yang menunjukkan bahwa ada 4 sekolah; dan Kelompok 2 memiliki 21 item, yang menunjukkan bahwa ada 21 sekolah.

Setelah mengetahui nilai *cluster* berikutnya, titik *centroid* dipilih. Pada setiap atribut *cluster*, titik *centroid* dipilih secara acak. Pemilihan titik *centroid* dapat dilihat pada Tabel 2.

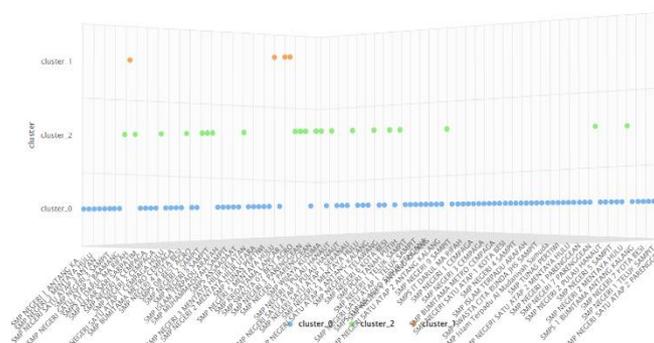
Tabel 2. Titik *centroid*

Attribute	<i>cluster</i> <u>0</u>	<i>cluster</i> <u>1</u>	<i>cluster</i> <u>r_2</u>
PD	103.13	845.25	340.23
Rombel	8	0	8
Guru	4.264	24.500	10.905
Tendik	8.092	43.500	17.762
Jmlh R.Kelas	1.828	9.750	3.333
Jmlh R.Lab	4.828	23.500	11.238
Jmlh R.Perpus	0.954	3.750	1.571
	0.770	1	1.095

Keanggotaan pada setiap *cluster* dihasilkan dari nilai *cluster* dan titik *centroid* yang dipilih. *Cluster-2* memiliki 21 sekolah yang terdiri dari SMP Negeri 11 Sampit, SMP Negeri 9 Sampit, SMPS It Darul Ma Rifah, SMP Negeri 1 Cempaga, SMP Negeri 7 Cempaga, SMP Negeri 2 Cempaga, SMP Negeri 3 Cempaga, SMP Negeri 1 Kota Besi, SMP Negeri 7 Sampit, SMP Negeri 8 Sampit, SMP Darul Amin, SMP Islam Terpadu Arafah, SMP Katholik Al Bertus Sampit, SMP Muhammadiyah Sampit, SMP Negeri 1 Mentaya Hilir Selatan, SMP Negeri 1 Mentaya Hilir Utara, SMP Swasta Bina Bangsa 01, SMP Negeri 1 Mentaya Hulu, SMP Negeri 1 Parenggean, SMP Negeri 5 Kota Besi, SMPS Bina Bangsa 02. *Cluster-1* berjumlah 4 sekolah yang terdiri dari SMP Negeri 3 Sampit, SMP Negeri 1 Sampit, SMP Negeri 2 Sampit, SMP Negeri 4 Sampit. Sedangkan pada *Cluster-0* berjumlah 87 sekolah, dari total 112 sekolah di Kabupaten Kotawaringin timur, data sekolah *Cluster-2* dan *Cluster-1* tidak ada tercantum

E. Analisis Cluster

Gambar 5 menunjukkan bahwa merah berada di peringkat pertama, hijau di peringkat kedua, dan biru di peringkat ketiga. Oleh karena itu, berdasarkan hasil visualisasi dari Gambar 5 dan Gambar 4, dapat dijelaskan bahwa warna orange berada di posisi Cluster-1, yang menunjukkan bahwa Cluster 1 memiliki empat sekolah di peringkat pertama; Gambar 4 menunjukkan hasil visualisasi dari Cluster-2, yang menunjukkan bahwa Cluster 1 memiliki empat sekolah di peringkat kedua.



Gambar 5. Peringkat *cluster*

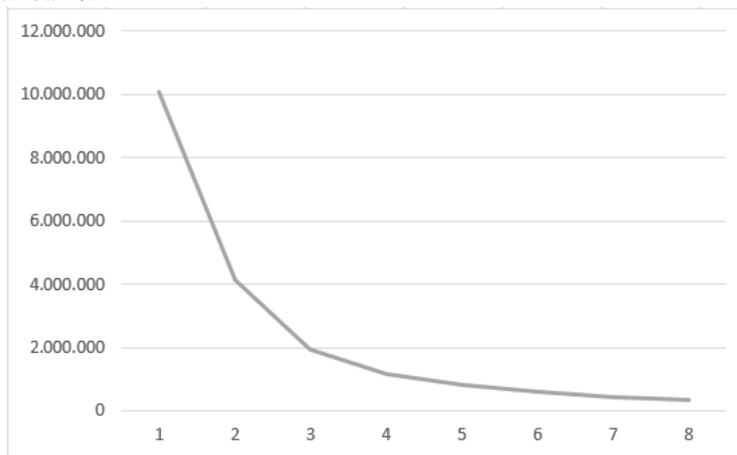
Penelitian ini menunjukkan kinerja algoritma dari hasil analisis dengan menunjukkan nilai rata-rata di sekitar jarak *centroid*, yang menghasilkan nilai 4143.863 dengan nilai K=3. Tabel 3 menunjukkan nilai perbandingan untuk kinerja pengolahan data masing-masing algoritma.

Tabel 3. Perbandingan *Performance*

Nilai K	Avg. within centroid distance
2	10080.811
3	4143.863
4	1935.004
5	1154.457
6	814.113
7	621.272
8	449.957
9	334.531

Pengujian kinerja algoritma untuk pengolahan data dilakukan sebanyak delapan kali. Pengujian pertama melihat nilai k=2, yang menghasilkan nilai 10080.811 pada Avg. within centroid distance. Kinerja algoritma diteliti dalam pengujian berikutnya dengan nilai k=3 dan seterusnya hingga nilai k=9.

Nilai Avg. within centroid distance menggunakan elbow method. Dari hasil nilai Avg. within centroid distance yang ditampilkan pada Gambar 6.



Gambar 6. Elbow method

Nilai rata-rata dalam jarak centroid berada pada nilai k=3, dan nilai rata-rata dalam jarak centroid = 4143.863. Oleh karena itu, dapat disimpulkan bahwa, dibandingkan dengan kelompok lain, data yang dibagi menjadi 3 kelompok memiliki nilai keakuratan yang lebih tinggi.

V. KESIMPULAN

Data awal untuk penelitian ini berjumlah 114 daftar, dan setelah preprocessing, 112 daftar dapat dianalisis. Hasil penelitian menunjukkan bahwa pengujian Algoritma K-Means untuk mengolah data telah diperbarui, menunjukkan nilai Avg. di sekitar jarak centroid. Ini menunjukkan bahwa penelitian sebelumnya tidak menguji kemampuan Algoritma K-Means untuk mengolah data, tetapi menyarankan untuk melakukan analisis untuk menghitung jumlah cluster.

Hasil pembagian cluster menunjukkan bahwa 87 sekolah berada dalam cluster 0, 4 sekolah berada dalam cluster 1, dan 21 sekolah berada dalam cluster 2. Selanjutnya, visualisasi algoritma K-Means menunjukkan pemeringkatan sekolah di cluster satu, kedua, dan ketiga secara berurutan. Oleh karena itu, diketahui bahwa 87 sekolah termasuk dalam kumpulan tertinggi.

DAFTAR PUSTAKA

- [1] N. Nurahman, A. Purwanto, and S. Mulyanto, “Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 337–350, 2022, doi: 10.30812/matrik.v21i2.1411.
- [2] D. D. C. Nugraha, Z. Naimah, M. Fahmi, and N. Setiani, “Klasterisasi Judul Buku dengan Menggunakan Metode K-Means,” *Semin. Nas. Apl. Teknol. Inf. Yogyakarta*, vol. 21, no. 1, pp. 1907–5022, 2014.
- [3] A. Sulistiyawati and E. Supriyanto, “Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan,” *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.
- [4] A. Aditya, I. Jovian, and B. N. Sari, “Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019,” *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 51, 2020, doi: 10.30865/mib.v4i1.1784.
- [5] R. C. PRIHANDARI, “Data Mining: Konsep Dan Aplikasi Menggunakan Rapidminer (Series: Supervised Learning Dan Unsupervised Learning),” 2022, [Online]. Available: [http://repository.uin-suska.ac.id/63073/1/REGITA CAHYANI PRIHANDARI.pdf](http://repository.uin-suska.ac.id/63073/1/REGITA%20CAHYANI%20PRIHANDARI.pdf)
- [6] Rahmawati, “Menentukan Tingkat Kesejahteraan Provinsi Kalimantan Tengah Dengan Penerapan Algoritma K-Means Clustering Menggunakan RapidMiner,” *J. Penerapan Sist. Inf.*, vol. 4, no. 1, pp. 105–115, 2023.
- [7] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, “Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 336, no. 1, 2018, doi: 10.1088/1757-899X/336/1/012017.
- [8] M. Algoritma, A. Di, P. T. Selatan, and I. Mandiri, “Implementasi Data Mining Pada Penjualan,” vol. 03, 2022.
- [9] B. Harto and I. G. I. Sudipa, *Data Science for Business (Pengantar dan Penerapan Berbagai Sektor)*, no. May. 2023. [Online]. Available: www.sonpedia.com
- [10] S. Lambertus, “Optimisasi Kinerja Firewall Menggunakan Teknik Data Mining,” *Appl. Netw.*, pp. III1–III19, 2018.
- [11] B. Lutkevich, *Social Media Analytics: Konsep dan Penerapannya dengan Rapid miner/Orange*, vol. 58, no. 10. 2010. [Online]. Available: <https://www.unodc.org/southasia/en/topics/frontpage/2009/trafficking-in-persons-and-smuggling-of-migrants.html>