

Klasterisasi Kecamatan di Kabupaten Asahan Berdasarkan Indikator Pendidikan SMA Menggunakan K-Means

Oktopanda¹, Andrian Syahputra^{1*}

¹Fakultas Ekonomi, Sains dan Teknologi, Program Studi Sistem Informasi, Universitas Muhammadiyah Asahan, Kisaran, Indonesia

Email: 1Oktopanda1@gmail.com, 2*andriansyahputra4@gmail.com

(* : coresponding author)

Abstrak—Kesenjangan antarkecamatan dalam jumlah sekolah, guru, dan peserta didik berpotensi mencerminkan ketimpangan kapasitas layanan pendidikan pada jenjang Sekolah Menengah Atas. Berangkat dari kondisi tersebut, penelitian ini dilakukan untuk mengelompokkan kecamatan-kecamatan di Kabupaten Asahan berdasarkan indikator pendidikan SMA Tahun Ajaran 2025/2026 dengan memanfaatkan algoritma K-Means. Data sekunder yang bersumber dari Badan Pusat Statistik, mencakup jumlah sekolah, guru, dan peserta didik, selanjutnya diolah menjadi tiga rasio, yaitu peserta didik terhadap guru, peserta didik terhadap sekolah, serta guru terhadap sekolah. Sebelum diproses, seluruh data distandardisasi melalui Z-score, sementara penentuan jumlah kluster dievaluasi dengan Elbow Method, Silhouette Score, dan Davies-Bouldin Index. Dari total 25 kecamatan, hanya 18 kecamatan yang dapat diikutsertakan dalam perhitungan, sedangkan tujuh kecamatan lain yang bernilai nol dikaji secara terpisah. Hasil pengujian memperlihatkan bahwa konfigurasi dua kluster merupakan yang paling optimal, dengan capaian Silhouette Score sebesar 0,491 dan Davies-Bouldin Index sebesar 0,701. Kluster pertama mencakup 16 kecamatan yang memiliki skala layanan tergolong rendah hingga menengah, sementara kluster kedua hanya beranggotakan Kota Kisaran Barat dan Kota Kisaran Timur yang menunjukkan konsentrasi sekolah, guru, dan peserta didik jauh lebih tinggi. Temuan ini diharapkan dapat menjadi informasi awal bagi upaya evaluasi pemerataan layanan pendidikan SMA di Kabupaten Asahan.

Kata Kunci: Data Mining, Indikator Pendidikan, K-Means, Klasterisasi, Sekolah Menengah Atas

Abstract—Disparities among districts in the number of schools, teachers, and students can signal unequal capacity in senior high school service provision. Building on this issue, the present study clusters the districts of Asahan Regency according to senior high school education indicators for the 2025/2026 academic year, applying the K-Means algorithm. Secondary data obtained from the Statistics of Asahan Regency, covering the number of schools, teachers, and students, were subsequently transformed into three ratios: student-to-teacher, student-to-school, and teacher-to-school. All variables were standardized through Z-scores, whereas the optimal number of clusters was determined using the Elbow Method together with the Silhouette Score and Davies-Bouldin Index. Out of 25 districts in total, only 18 could be included in the clustering process, while the remaining seven districts with zero values were examined separately. The evaluation results indicate that a two-cluster configuration performs best, yielding a Silhouette Score of 0.491 and a Davies-Bouldin Index of 0.701. The first cluster contains 16 districts characterized by a relatively low-to-moderate service scale, whereas the second cluster consists solely of West Kisaran City and East Kisaran City, both showing markedly higher concentrations of schools, teachers, and students. These findings are expected to serve as preliminary information for assessing the equitable distribution of senior high school services in Asahan Regency.

Keywords: Clustering, Data Mining, Education Indicators, K-Means, Senior High School

1. PENDAHULUAN

Sebagai jenjang yang mengantarkan peserta didik menuju pendidikan tinggi maupun dunia kerja, pendidikan menengah memegang peranan penting dalam pembentukan kompetensi mereka. Agar layanan pendidikan dapat menjangkau seluruh wilayah secara merata, dibutuhkan ketersediaan sekolah, guru, dan kapasitas layanan yang seimbang. Namun demikian, karakteristik kecamatan di Kabupaten Asahan yang beragam menyebabkan jumlah satuan pendidikan, guru, dan peserta didik SMA berbeda-beda antarwilayah. Kondisi ini tergambar jelas pada data BPS Kabupaten Asahan Tahun Ajaran 2025/2026, yang sekaligus menjadi dasar untuk melakukan analisis berbasis data (Badan Pusat Statistik Kabupaten Asahan, 2026).

Sayangnya, data yang tersaji dalam bentuk tabel statistik belum secara langsung memperlihatkan kelompok kecamatan dengan karakteristik layanan pendidikan yang serupa. *Data mining* dengan memanfaatkan teknik *clustering* dapat dimanfaatkan untuk menemukan pola tersembunyi tanpa memerlukan label kelas sebelumnya. Di antara berbagai algoritma, K-Means

banyak dipilih karena sifatnya yang sederhana, efisien, dan mudah diinterpretasikan dalam mengelompokkan objek berdasarkan kedekatannya terhadap pusat kluster (Tampubolon et al., 2024). Penerapannya pada aspek fasilitas, pendidik, dan tenaga kependidikan juga terbukti membantu pemerintah mengidentifikasi kelompok sekolah yang membutuhkan pendampingan maupun pemerataan sumber daya (Nurahman et al., 2022).

Di bidang pendidikan, pemanfaatan K-Means telah merambah berbagai aspek, mulai dari prediksi kelulusan mata kuliah, analisis penyelesaian studi, hingga penggunaan *e-learning*, serta segmentasi minat program studi (Adistya et al., 2023; Novita et al., 2023; Vega et al., 2023; Wahyudin & Dikananda, 2023). Pada jenjang sekolah, algoritma yang sama juga diterapkan untuk mengelompokkan prestasi, menentukan jurusan, serta mengklasifikasikan capaian belajar berdasarkan karakteristik akademik maupun sosial ekonomi (Palevi & Indra, 2024; Qusyairi et al., 2024; Siregar et al., 2024; Tundo et al., 2024). Berbagai kajian tersebut memperlihatkan bahwa hasil klusterisasi mampu menjadi informasi pendukung dalam evaluasi akademik, meskipun karakteristik kluster yang dihasilkan tetap harus dimaknai sesuai konteks data yang digunakan.

Perlu dicatat, kualitas hasil K-Means sangat dipengaruhi oleh skala variabel, jumlah kluster, dan pemilihan pusat awal. Standardisasi menjadi hal yang tidak dapat diabaikan agar variabel bernilai besar tidak mendominasi perhitungan jarak. Selain itu, reduksi dan prapemrosesan variabel dapat meningkatkan keterbacaan sekaligus stabilitas hasil yang diperoleh (Rosyada & Utari, 2024). Penentuan jumlah kluster juga perlu dievaluasi menggunakan metode seperti *Elbow Method*, *Silhouette Score*, dan *Davies-Bouldin Index*, bukan semata-mata ditentukan secara subjektif (Wahyudin & Syahruf Efendi, 2024). Sebagai pengembangan lebih lanjut, kombinasi K-Means dengan analisis komponen utama juga telah diterapkan untuk menghasilkan pengelompokan kemampuan akademik yang lebih objektif (Idris & Haryono, 2025).

Sejauh ini, penelitian-penelitian terdahulu cenderung berfokus pada peserta didik, mahasiswa, atau satuan pendidikan secara individual, sementara pengelompokan kecamatan berdasarkan indikator layanan SMA di Kabupaten Asahan masih jarang dikaji. Untuk mengisi celah tersebut, penelitian ini menggunakan enam indikator, yaitu jumlah sekolah, jumlah guru, jumlah peserta didik, rasio peserta didik terhadap guru, rasio peserta didik terhadap sekolah, dan rasio guru terhadap sekolah. Secara khusus, penelitian ini diarahkan untuk menentukan jumlah kluster yang optimal, mengelompokkan kecamatan berdasarkan kesamaan karakteristiknya, serta menguraikan implikasi hasil tersebut bagi evaluasi pemerataan layanan pendidikan SMA.

2. METODOLOGI PENELITIAN

2.1 Jenis, Sumber Data, dan Tahapan Penelitian

Penelitian ini dirancang dengan pendekatan kuantitatif deskriptif yang dipadukan dengan teknik *clustering*. Data sekunder yang digunakan bersumber dari tabel BPS Kabupaten Asahan, meliputi jumlah satuan pendidikan, kepala sekolah dan guru, serta peserta didik SMA menurut kecamatan pada Tahun Ajaran 2025/2026, dengan unit analisis mencakup 25 kecamatan. Secara keseluruhan, tahapan penelitian terdiri atas pengumpulan data, pembersihan data, pembentukan indikator rasio, standardisasi Z-score, evaluasi K = 2-5, klusterisasi K-Means, dan evaluasi dan interpretasi hasil.



Gambar 1. Tahapan penelitian

2.2 Variabel Penelitian

Tabel 1. Variabel penelitian

Kode	Variabel	Sumber
X1	Jumlah sekolah SMA	Data BPS
X2	Jumlah guru SMA	Data BPS

Kode	Variabel	Sumber
X3	Jumlah peserta didik SMA	Data BPS
X4	Rasio peserta didik terhadap guru	Hasil perhitungan
X5	Rasio peserta didik terhadap sekolah	Hasil perhitungan
X6	Rasio guru terhadap sekolah	Hasil perhitungan

Guna membedakan skala layanan sekaligus beban pendidikan antarkecamatan, penelitian ini menggunakan tiga indikator rasio yang perhitungannya dirumuskan sebagai berikut.

$$RPG_i = PD_i / G_i \quad (1)$$

$$RPS_i = PD_i / S_i \quad (2)$$

$$RGS_i = G_i / S_i \quad (3)$$

Dalam hal ini, RPG_i menyatakan rasio peserta didik terhadap guru, RPS_i menyatakan rasio peserta didik terhadap sekolah, dan RGS_i menyatakan rasio guru terhadap sekolah pada kecamatan ke- i , dengan PD , G , dan S berturut-turut merepresentasikan jumlah peserta didik, guru, dan sekolah.

2.3 Prapemrosesan dan Standardisasi

Tahap prapemrosesan diawali dengan pemeriksaan nama kecamatan, nilai kosong, duplikasi, dan nilai nol pada data. Kecamatan yang memiliki jumlah sekolah dan guru bernilai nol tidak diikutsertakan dalam perhitungan rasio, mengingat pembagian $0/0$ tidak terdefinisi secara matematis. Setelah itu, variabel pada kecamatan yang memenuhi syarat distandardisasi menggunakan *Z-score* agar berada pada skala yang sebanding.

$$z_{ij} = (x_{ij} - \mu_j) / \sigma_j \quad (4)$$

Pada Persamaan (4), x_{ij} menunjukkan nilai objek ke- i pada variabel ke- j , μ_j merupakan rata-rata variabel tersebut, dan σ_j menyatakan standar deviasinya.

2.4 Penentuan Klaster dan Algoritma K-Means

Untuk menentukan jumlah klaster yang tepat, nilai K diuji pada rentang 2 hingga 5. *Elbow Method* digunakan untuk mengamati pola penurunan *Within-Cluster Sum of Squares*, sedangkan *Silhouette Score* dan *Davies-Bouldin Index* dimanfaatkan untuk menilai kekompakan sekaligus keterpisahan antarklaster. Konfigurasi terbaik kemudian dipilih dengan mempertimbangkan posisi titik siku, *Silhouette Score* tertinggi, serta *Davies-Bouldin Index* terendah. Proses K-Means sendiri dijalankan menggunakan inisialisasi K-Means++ dengan 100 kali pengulangan inisialisasi serta *random state* 42 agar hasilnya dapat direplikasi.

$$d(x_i, c_k) = \sqrt{\sum_j (x_{ij} - c_{kj})^2} \quad (5)$$

Penempatan setiap kecamatan pada pusat klaster terdekat dilakukan berdasarkan jarak Euclidean sebagaimana Persamaan (5), dengan pusat klaster yang terus diperbarui melalui rata-rata anggotanya hingga keanggotaan klaster menjadi stabil.

3. ANALISA DAN PEMBAHASAN

3.1 Deskripsi Data dan Penanganan Nilai Nol

Berdasarkan data BPS, tercatat sebanyak 43 SMA, 849 guru, dan 15.940 peserta didik yang tersebar di 25 kecamatan. Kota Kisaran Barat tercatat memiliki jumlah sekolah terbanyak, sementara Kota Kisaran Timur unggul dalam jumlah guru maupun peserta didik. Rincian data selengkapnya disajikan pada Tabel 2.

Tabel 2. Data pendidikan SMA menurut kecamatan

No.	Kecamatan	Sekolah	Guru	Peserta Didik
1	Bandar Pasir Mandoge	2	30	702
2	Bandar Pulau	1	25	377
3	Aek Songsongan	4	50	720
4	Rahuning	0	0	0
5	Pulau Rakyat	2	64	1.522
6	Aek Kuasan	1	31	655
7	Aek Ledong	0	0	0
8	Sei Kepayang	0	0	0
9	Sei Kepayang Barat	1	16	235
10	Sei Kepayang Timur	0	0	0
11	Tanjung Balai	1	24	522
12	Simpang Empat	1	45	810
13	Teluk Dalam	0	0	0
14	Air Batu	3	44	912
15	Sei Dadap	1	3	93
16	Buntu Pane	1	31	620
17	Tinggi Raja	2	12	192
18	Setia Janji	1	8	67
19	Meranti	5	74	1.391
20	Pulo Bandring	1	5	218
21	Rawang Panca Arga	0	0	0
22	Air Joman	2	47	918
23	Silau Laut	0	0	0
24	Kota Kisaran Barat	9	164	2.560
25	Kota Kisaran Timur	5	176	3.426
	Jumlah	43	849	15.940

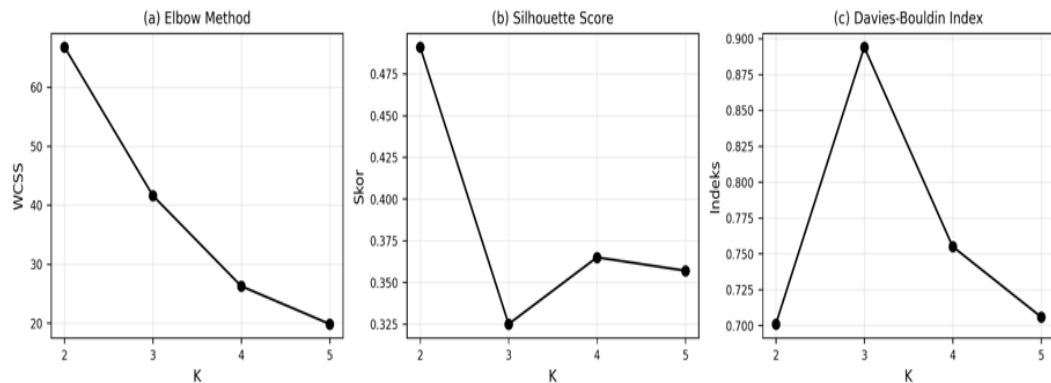
Terdapat tujuh kecamatan yang mencatatkan nilai nol pada seluruh indikator, yaitu Rahuning, Aek Ledong, Sei Kepayang, Sei Kepayang Timur, Teluk Dalam, Rawang Panca Arga, dan Silau Laut. Nilai nol tersebut hanya menandakan tidak adanya SMA dalam cakupan tabel yang digunakan, bukan berarti tidak terdapat sama sekali bentuk pendidikan menengah di wilayah tersebut. Oleh sebab itu, ketujuh kecamatan ini dianalisis sebagai kelompok tersendiri dan tidak diikutsertakan dalam proses K-Means berbasis rasio.

3.2 Hasil Indikator Rasio dan Evaluasi Jumlah Klaster

Dari 18 kecamatan yang berhasil diproses, rasio peserta didik terhadap guru tertinggi ditemukan di Pulo Bandring dengan nilai 43,60, disusul Sei Dadap sebesar 31,00. Sementara itu, rasio peserta didik terhadap sekolah tertinggi tercatat di Simpang Empat sebesar 810,00, kemudian Pulau Rakyat sebesar 761,00, dan Kota Kisaran Timur sebesar 685,20. Temuan ini menegaskan bahwa jumlah absolut saja tidak cukup untuk menggambarkan beban layanan; indikator rasio perlu dibaca bersamaan agar beban tersebut tidak tertutupi oleh besar-kecilnya ukuran kecamatan.

Tabel 3. Evaluasi Jumlah Kluster

K	WCSS	Silhouette Score	Davies-Bouldin Index
2	66.762	0.491	0.701
3	41.583	0.325	0.894
4	26.255	0.365	0.755
5	19.841	0.357	0.706


Gambar 2. Evaluasi Jumlah Kluster Menggunakan Elbow, Silhouette, Dan Davies-Bouldin

K = 2 memberikan *Silhouette Score* tertinggi, yakni 0,491, sekaligus *Davies-Bouldin Index* terendah sebesar 0,701. Meskipun nilai WCSS terus menurun seiring bertambahnya K, kedua ukuran evaluasi internal tersebut secara konsisten menunjukkan bahwa dua kluster menghasilkan struktur pengelompokan yang paling stabil. Atas dasar itu, K = 2 dipilih sebagai konfigurasi akhir dalam proses pengelompokan.

3.3 Hasil Klasterisasi dan Karakteristik Kluster

Tabel 4. Keanggotaan Hasil Klasterisasi

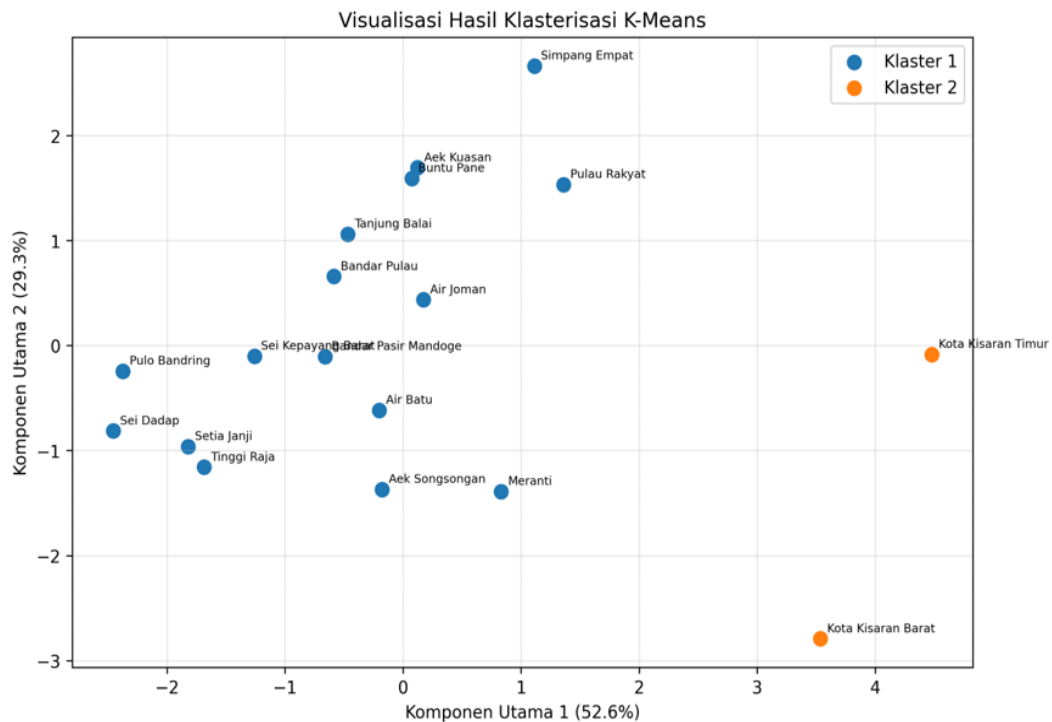
Kluster	Jumlah	Kecamatan
Kluster 1	16	Bandar Pasir Mandoge, Bandar Pulau, Aek Songsongan, Pulau Rakyat, Aek Kuasan, Sei Kepayang Barat, Tanjung Balai, Simpang Empat, Air Batu, Sei Dadap, Buntu Pane, Tinggi Raja, Setia Janji, Meranti, Pulo Bandring, Air Joman
Kluster 2	2	Kota Kisaran Barat, Kota Kisaran Timur

Tabel 5. Centroid Kluster Dalam Skala Asli

Indikator	Kluster 1	Kluster 2
Jumlah sekolah	1.81	7.00
Jumlah guru	31.81	170.00
Jumlah peserta didik	622.12	2993.00
Rasio peserta didik/guru	20.64	17.54
Rasio peserta didik/sekolah	376.64	484.82
Rasio guru/sekolah	19.15	26.71

Kluster 1 beranggotakan 16 kecamatan dengan rata-rata 1,81 sekolah, 31,81 guru, dan 622,12 peserta didik. Secara umum, kelompok ini mencerminkan skala layanan yang tergolong rendah hingga menengah, meskipun variasi internal di dalamnya tetap cukup tinggi. Sebagai contoh, Pulo Bandring dan Sei Dadap menunjukkan rasio peserta didik terhadap guru yang tinggi, sedangkan Simpang Empat dan Pulau Rakyat justru menanggung beban peserta didik per sekolah yang besar.

Berbeda dengan kluster pertama, Kluster 2 hanya beranggotakan Kota Kisaran Barat dan Kota Kisaran Timur, dengan rata-rata 7 sekolah, 170 guru, dan 2.993 peserta didik. Kedua kecamatan ini bahkan menampung 32,56% sekolah, 40,05% guru, dan 37,55% peserta didik SMA se-Kabupaten Asahan, yang mengindikasikan tingginya konsentrasi layanan pendidikan SMA di kawasan Kota Kisaran. Meskipun rasio peserta didik terhadap guru pada Kluster 2 lebih rendah dibandingkan Kluster 1, beban peserta didik per sekolah justru lebih tinggi sehingga kapasitas sekolah pada kluster ini tetap perlu mendapat perhatian.



Gambar 3. Visualisasi hasil klasterisasi menggunakan Principal Component Analysis

Melalui visualisasi dua dimensi, terlihat jelas bahwa Kota Kisaran Barat dan Kota Kisaran Timur berada terpisah dari sebagian besar kecamatan lainnya. Pemisahan ini sejalan dengan nilai centroid yang jauh lebih tinggi pada jumlah sekolah, guru, dan peserta didik di kedua kecamatan tersebut. Meskipun demikian, visualisasi ini semata-mata digunakan untuk keperluan penyajian, sementara proses K-Means tetap mengacu pada keenam variabel yang telah distandardisasi.

3.4 Pembahasan

Secara keseluruhan, hasil klasterisasi berhasil menyederhanakan variasi indikator pendidikan menjadi dua pola utama, yakni kecamatan dengan skala layanan rendah-menengah dan kecamatan dengan konsentrasi layanan tinggi. Temuan ini sejalan dengan penelitian klasterisasi sekolah sebelumnya, yang menunjukkan bahwa algoritma K-Means mampu membantu mengidentifikasi kelompok dengan kebutuhan sumber daya yang berbeda-beda (Nurahman et al., 2022). Penggunaan indikator rasio dalam penelitian ini turut memperkuat interpretasi hasil, mengingat kecamatan dengan jumlah sekolah yang sama sekalipun dapat memikul beban layanan yang sangat berbeda.

Meskipun demikian, keanggotaan kluster tidak selayaknya dijadikan satu-satunya dasar pengambilan kebijakan. Pulo Bandring, Sei Dadap, Sim pang Empat, dan Pulau Rakyat, misalnya, tetap memerlukan pemeriksaan indikator secara individual sekalipun berada pada kluster yang sama. Demikian pula, tujuh kecamatan bernilai nol perlu ditelaah lebih lanjut dengan mempertimbangkan data SMA dari pengelola lain, SMK, Madrasah Aliyah, jumlah penduduk usia sekolah, jarak ke sekolah, serta akses ke kecamatan tetangga. Dengan kata lain, hasil K-Means dalam penelitian ini lebih tepat diposisikan sebagai informasi awal untuk menentukan prioritas analisis, bukan sebagai kesimpulan tunggal mengenai mutu pendidikan.

4. IMPLEMENTASI

4.1 Lingkungan dan Parameter Implementasi

Seluruh proses pengolahan data dilakukan dengan memanfaatkan Python pada platform Google Colab. Pustaka Pandas berperan dalam pengelolaan data, Scikit-learn digunakan untuk standardisasi, penerapan K-Means, serta perhitungan *Silhouette Score*, *Davies-Bouldin Index*, dan *Principal Component Analysis*, sedangkan Matplotlib dimanfaatkan untuk keperluan visualisasi.

Tabel 6. Lingkungan dan Parameter Implementasi

Komponen	Konfigurasi
Platform	Google Colab
Bahasa	Python
Data	Microsoft Excel
Inisialisasi	K-Means++
n_init	100
random_state	42
K yang diuji	2, 3, 4, dan 5
Ukuran jarak	Euclidean Distance

4.2 Pengujian Proses

Tabel 7. Hasil Pengujian Proses Pengolahan

No.	Skenario	Hasil	Status
1	Membaca data	25 kecamatan terbaca	Berhasil
2	Mendeteksi nilai nol	7 kecamatan teridentifikasi	Berhasil
3	Menghitung indikator rasio	Rasio terbentuk pada 18 kecamatan	Berhasil
4	Standardisasi	Enam variabel mempunyai skala sebanding	Berhasil
5	Evaluasi K = 2–5	WCSS, Silhouette, dan DBI diperoleh	Berhasil
6	Klasterisasi K = 2	18 kecamatan memperoleh label	Berhasil
7	Inverse standardization	Centroid kembali ke skala asli	Berhasil
8	Ekspor hasil	Tabel dan grafik tersimpan	Berhasil

Secara umum, seluruh tahapan yang dirancang berhasil menghasilkan keluaran sesuai dengan yang diharapkan. Pada pengujian ulang dengan parameter yang sama, komposisi klaster yang dihasilkan tetap konsisten, yaitu 16 kecamatan pada Klaster 1 dan dua kecamatan pada Klaster 2. Implementasi ini pada akhirnya menghasilkan data rasio, tabel evaluasi K, label klaster, centroid, serta visualisasi yang dapat direplikasi kembali.

5. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma K-Means untuk mengelompokkan kecamatan di Kabupaten Asahan berdasarkan enam indikator pendidikan SMA Tahun Ajaran 2025/2026. Dari total 25 kecamatan, hanya 18 kecamatan yang dapat diproses, sementara tujuh kecamatan bernilai nol dianalisis secara terpisah karena indikator rasionya tidak dapat dihitung. Evaluasi terhadap nilai K = 2 hingga K = 5 memperlihatkan bahwa konfigurasi dua klaster merupakan pilihan terbaik, dengan *Silhouette Score* sebesar 0,491 dan *Davies-Bouldin Index* sebesar 0,701. Klaster 1 beranggotakan 16 kecamatan dengan skala layanan tergolong rendah hingga menengah, sedangkan Klaster 2 hanya terdiri atas Kota Kisaran Barat dan Kota Kisaran Timur yang menunjukkan konsentrasi sekolah, guru, dan peserta didik lebih tinggi. Hasil klasterisasi ini dapat dimanfaatkan sebagai informasi awal dalam mengevaluasi pemerataan layanan pendidikan, meskipun

rekomendasi kebijakan tetap perlu mempertimbangkan indikator masing-masing kecamatan secara individual. Untuk penelitian selanjutnya, disarankan menambahkan variabel lain seperti jumlah penduduk usia sekolah, ruang kelas, sarana, angka partisipasi, jarak sekolah, SMK, dan Madrasah Aliyah, serta membandingkan performa K-Means dengan metode lain seperti K-Medoids, *hierarchical clustering*, atau *Fuzzy C-Means*.

REFERENCES

- Adistya, A. P., Luthfyani, N., Tara, P., Adriyan, R., Rifaldi, & Rosyani, P. (2023). Klasterisasi menggunakan algoritma K-Means clustering untuk memprediksi kelulusan mata kuliah mahasiswa. *OKTAL: Jurnal Ilmu Komputer dan Sains*, 2(8), 2301–2306. <https://journal.mediapublikasi.id/index.php/oktal/article/view/1610>
- Badan Pusat Statistik Kabupaten Asahan. (2026). Jumlah satuan pendidikan, kepala sekolah dan guru, serta peserta didik Sekolah Menengah Atas menurut kecamatan di Kabupaten Asahan, 2025/2026 [Tabel statistik]. <https://asahankab.bps.go.id/>
- Idris, M. F., & Haryono, W. (2025). Penerapan teknologi machine learning untuk otomatisasi pengelompokan siswa berdasarkan kemampuan akademik di SMK Negeri 8 Kabupaten Tangerang dengan metode Principal Component Analysis dan K-Means clustering. *BINER: Jurnal Ilmu Komputer, Teknik dan Multimedia*, 3(2), 203–215.
- Novita, R., Khomarudin, A. N., Aulia, R., Jamaluddin, Yuditihwa, A., & Ayuri, A. (2023). Penerapan algoritma K-Means dan analisisnya untuk menentukan kebijakan strategis penyelesaian studi mahasiswa. *Jurnal SAINTIKOM*, 22(2), 401–413. <https://doi.org/10.53513/jis.v22i2.8461>
- Nurahman, N., Purwanto, A., & Mulyanto, S. (2022). Klasterisasi sekolah menggunakan algoritma K-Means berdasarkan fasilitas, pendidik, dan tenaga pendidik. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 21(2), 337–350. <https://doi.org/10.30812/matrik.v21i2.1411>
- Palevi, M. R., & Indra, Z. (2024). Implementasi algoritma K-Means clustering dengan pendekatan active learning pada siswa SMA untuk menentukan jurusan ke perguruan tinggi. *Jurnal SAINTIKOM*, 23(1), 26–36. <https://doi.org/10.53513/jis.v23i1.9553>
- Qusyairi, M., Hidayatullah, Z., & Sandi, A. (2024). Penerapan K-Means clustering dalam pengelompokan prestasi siswa dengan optimasi metode elbow. *Infotek: Jurnal Informatika dan Teknologi*, 7(2), 500–510. <https://doi.org/10.29408/jit.v7i2.26375>
- Rosyada, I. A., & Utari, D. T. (2024). Penerapan Principal Component Analysis untuk reduksi variabel pada algoritma K-Means clustering. *Jambura Journal of Probability and Statistics*, 5(1), 6–13. <https://ejurnal.ung.ac.id/index.php/jps/article/view/18733>
- Siregar, H. L., Hidayanthi, R., & Sakti, A. L. D. (2024). Implementation of K-Means clustering on student learning achievements based on social economic and social related. *Research and Development in Education*, 4(2), 1447–1459. <https://doi.org/10.22219/raden.v4i2.36742>
- Tampubolon, R. J., Ariesandes, T., Hafizd, M. I., & Hidayatullah, N. (2024). Literature review: Implementasi algoritma K-Means clustering dalam berbagai sektor. *BINER: Jurnal Ilmu Komputer, Teknik dan Multimedia*, 2(5), 803–809. <https://journal.mediapublikasi.id/index.php/Biner/article/view/4837>
- Tundo, Raihanah, S., Wahyudi, T., & Sugiyono. (2024). K-Means clustering of social studies performance at junior high school. *International Journal on Informatics for Development*, 13(2), 460–472. <https://doi.org/10.14421/ijid.2024.4632>
- Vega, M. A., Savitri, V. K., & Terttiaavini. (2023). Penerapan clustering menggunakan metode K-Means untuk penggunaan e-learning di dunia. *OKTAL: Jurnal Ilmu Komputer dan Sains*, 2(5), 1478–1482. <https://journal.mediapublikasi.id/index.php/oktal/article/view/2915>
- Wahyudin, E., & Dikananda, F. (2023). Segmentasi minat mahasiswa terhadap program studi menggunakan algoritma K-Means clustering. *BULLET: Jurnal Multidisiplin Ilmu*, 2(1), 271–276. <https://journal.mediapublikasi.id/index.php/bullet/article/view/5223>
- Wahyudin, E., & Syahruf Efendi, M. (2024). Optimasi klasterisasi data peserta didik SDN 2 Sarajaya menggunakan algoritma K-Means dan grid search. *BULLET: Jurnal Multidisiplin Ilmu*, 3(1), 156–161. <https://journal.mediapublikasi.id/index.php/bullet/article/view/5183>