

PEMETAAN DAN PENGELOMPOKAN KONDISI PENDIDIKAN KECAMATAN DI KABUPATEN KARAWANG MENGGUNAKAN *K-MEANS CLUSTERING*

Ahmad Wahdan Subhan, Bayu Priyatna, April Lia Hananto, Shofa Shofiah Hilabi

Sistem Informasi, Universitas Buana Perjuangan Karawang
Jalan Ronggo Waluyo Sirnabaya, Puseurjaya, Telukjambe Timur, Indonesia
Si22.ahmadsubhan@mhs.ubpkarawang.ac.id

ABSTRAK

Riset ini dilakukan guna memetakan wilayah kecamatan di Karawang sesuai strata pendidikan warga demi mengungkap pola serta disparitas antar daerah. Masalah utama yang dibahas adalah absennya instrumen pemetaan distribusi pendidikan yang akurat, sehingga menghambat efektivitas formulasi kebijakan publik. Analisis data mengimplementasikan algoritma *K-Means Clustering* melalui fase normalisasi *Min-Max Scaler*, identifikasi titik siku (*Elbow Method*), serta pengujian validitas menggunakan *Silhouette Coefficient*. Hasilnya menunjukkan bahwa konfigurasi tiga kelompok ($K = 3$) merupakan skenario paling optimal dengan perolehan skor *Silhouette* sebesar 0.5686, yang merepresentasikan stabilitas struktur klaster. Wilayah teridentifikasi ke dalam tipologi pendidikan dasar, berkembang, dan tinggi. Berdasarkan temuan, area dengan level pendidikan tinggi terkonsentrasi pada pusat pertumbuhan ekonomi, sementara mayoritas kecamatan lainnya masih tergolong dalam klaster pendidikan dasar. Kesenjangan ini mengindikasikan perlunya redistribusi sumber daya manusia dan infrastruktur pendidikan yang lebih inklusif. Temuan ini diharapkan mampu menjadi fondasi teknis bagi pemerintah daerah dalam merancang program intervensi yang presisi, seperti alokasi anggaran dan beasiswa yang disesuaikan dengan profil kebutuhan spesifik di tiap wilayah.

Kata kunci : *K-Means, clustering, pendidikan, Karawang, pemetaan, data mining*

1. PENDAHULUAN

Pendidikan menjadi parameter krusial dalam mengevaluasi level pengembangan sumber daya manusia pada suatu daerah. Terbukanya akses pendidikan yang merata mampu menstimulasi pertumbuhan ekonomi wilayah sekaligus mengeskalasi taraf kesejahteraan masyarakat secara komprehensif. [1].

Namun, pada kenyataannya, masih terdapat perbedaan dalam pencapaian pendidikan di antara berbagai wilayah, termasuk Kabupaten Karawang, yang memiliki campuran antara kecamatan pedesaan dan perkotaan [2].

Identifikasi daerah yang butuh perhatian khusus di sektor pendidikan masih sulit dilakukan karena datanya belum terpetakan dengan baik per wilayah [3].

Absennya skema kategorisasi yang sistematis berpotensi menyebabkan kebijakan pembangunan menjadi tidak tepat sasaran dan kurang optimal [4].

Guna memitigasi kendala tersebut, diperlukan implementasi teknik analisis data yang mampu melakukan klasifikasi wilayah berdasarkan kesamaan karakteristik pendidikan [5].

K-Means Clustering merupakan algoritma penambangan data yang efektif digunakan untuk mengorganisir objek ke dalam beberapa kelompok berdasarkan kemiripan atribut tertentu [6].

Penelitian ini bertujuan untuk melakukan kategorisasi penduduk di Kabupaten Karawang sesuai tingkat pendidikan melalui penerapan metode *K-Means*. [7].

Diharapkan hasil penelitian ini dapat menggambarkan sebaran tingkat pendidikan dan menjadi landasan bagi pelaksanaan kebijakan yang

bertujuan untuk meningkatkan tingkat pendidikan di Kabupaten Karawang [8].

2. TINJAUAN PUSTAKA

Untuk mengumpulkan informasi, data, dan kerangka teoritis bagi penelitian ini, peneliti melakukan tinjauan pustaka. Untuk mencapai hal ini, peneliti menelaah berbagai bahan tertulis yang relevan. Guna menjamin kredibilitas akademik, studi ini mengelaborasi beragam referensi primer seperti jurnal penelitian dan laporan institusional yang membedah implementasi *K-Means* dalam pemetaan kualitas pendidikan regional. Prosedur sistematis tersebut ditempuh untuk mengekstraksi parameter krusial yang berdampak pada akurasi pengelompokan, sekaligus memastikan bahwa seluruh kerangka penelitian memiliki basis teoretis yang solid dan dapat dipertanggungjawabkan. [9].

Wilayah tersebut berhasil dikelompokkan menjadi tiga klaster utama dalam sebuah studi oleh mengenai “Pengelompokan Tingkat Pendidikan di DKI Jakarta pada Tingkat Kecamatan Menggunakan Algoritma *K-Means*” [10].

Berdasarkan temuan studi tersebut, lokasi lokasi di Klaster 0 memiliki indikator pendidikan yang rendah sehingga memerlukan dukungan infrastruktur dari pemerintah. Demikian pula, sebuah studi oleh mengenai tingkat pendaftaran sekolah di Jawa Tengah juga berhasil mengidentifikasi pola distribusi kabupaten/kota ke dalam tiga kelompok, sehingga memberikan gambaran objektif kepada pemerintah provinsi untuk menurunkan angka putus sekolah melalui strategi yang lebih terfokus [11].

Meskipun penelitian lain telah mengkaji pengelompokan data pendidikan di berbagai lokasi, penelitian ini berbeda karena berfokus pada analisis di tingkat kecamatan di Kabupaten Karawang. Penelitian ini memanfaatkan keunggulan metode *K-Means* untuk mengembangkan model pengelompokan wilayah yang disesuaikan dengan karakteristik unik Karawang yang belum sepenuhnya dieksplorasi dalam penelitian sebelumnya. Hasil klasifikasi ini diproyeksikan sebagai instrumen pengambilan keputusan yang objektif bagi otoritas daerah di Kabupaten Karawang dalam menyusun regulasi berbasis data. Dengan memanfaatkan temuan tersebut, distribusi anggaran serta fasilitas pendidikan dapat dilakukan secara lebih proporsional guna menjamin keadilan sosial dan mempercepat transformasi mutu sekolah di seluruh wilayah. [12].

2.1. Pendidikan

Pembangunan kualitas modal manusia sebagai pilar masa depan bangsa sangat bergantung pada instrumen pendidikan. Performa sistem pendidikan formal, yang mencakup seluruh hierarki mulai dari level dasar hingga jenjang pendidikan tinggi, menjadi penentu utama dalam mencetak generasi yang kompetitif., sering kali dijadikan tolok ukur kualitas tersebut [13].

Pemanfaatan data pendidikan telah menjadi hal yang sangat penting di era digital untuk memastikan bahwa infrastruktur dan sumber daya dialokasikan kepada masyarakat secara efisien dan adil [14].

Tabel 1. Jenjang Pendidikan di Indonesia

Jenjang Pendidikan	Status Pendidikan
Pendidikan Dasar	SD, SMP (dan sederajat).
Pendidikan Menengah	SMA, SMK (dan sederajat).
Pendidikan Tinggi	Universitas

2.2. K-Means

Implementasi *K-Means* menitikberatkan pada teknik pembagian data melalui skema partisi guna menghasilkan kelompok-kelompok yang kohesif. Esensi dari algoritma ini terletak pada kemampuannya mengukur tingkat kemiripan antar fitur, yang kemudian digunakan sebagai landasan untuk mengelompokkan data secara mandiri dan sistematis.[15].

Minimisasi deviasi di dalam kelompok serta maksimisasi jarak pemisahan antar kluster merupakan target utama dari skema perhitungan jarak *Euclidean* ini. Dengan mengevaluasi kedekatan setiap titik data terhadap *centroids* secara berulang, sistem mampu mengorganisir data ke dalam struktur pengelompokan yang paling stabil dan memiliki kohesi internal yang kuat. [16].

Secara teknis *K-Means* banyak digunakan karena kompleksitas komputasinya yang rendah dan kemudahan implementasinya, terutama saat menghasilkan pengelompokan yang andal pada

kumpulan data dengan distribusi data yang padat dan terpisah dengan baik[17].

$$SSE = \sum_{j=1}^k \sum_{i=1}^n ||x_i^j - c_j||^2 \quad (1)$$

Keterangan :

SSE : Akumulasi kuadrat deviasi antara setiap observasi data terhadap titik pusat (*centroid*) pada masing-masing kelompok.

k : Jumlah total kluster yang terbentuk.

n : Jumlah data yang terdapat di dalam kluster ke *j*.

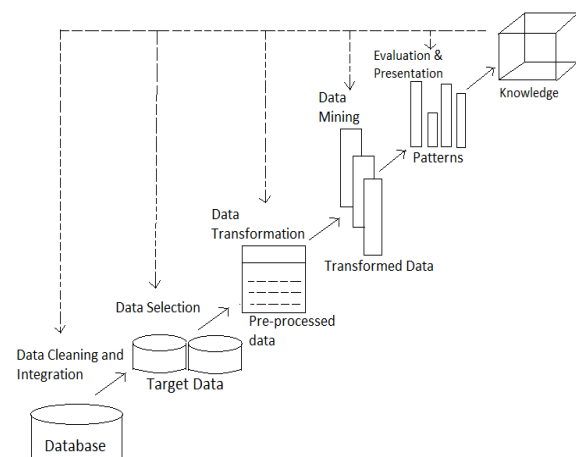
x_i^j : Entitas data ke *i* yang terasimilasi dalam kluster ke *j*

c_j : Titik tengah koordinat (*centroid*) yang merepresentasikan kelompok ke *j*.

2.3. Data Mining

Ekstraksi pola, model, maupun pengetahuan bernilai guna yang tersimpan dalam himpunan data bervolume masif melalui mekanisme otomatis secara umum didefinisikan sebagai *data mining*. Implementasi teknologi ini menitikberatkan pada proses transformasi entitas data mentah menjadi instrumen informasi strategis yang esensial, guna mendukung ketepatan serta objektivitas dalam setiap tahapan pengambilan keputusan. proses ini menggunakan berbagai pendekatan, termasuk klasifikasi, asosiasi, dan pengelompokan [18].

Untuk memastikan bahwa data yang dihasilkan selama fase fase mulai dari prapemrosesan hingga evaluasi dapat diandalkan dan relevan dengan tujuan penelitian, data mining sering kali mengikuti kerangka kerja *Knowledge Discovery in Databases (KDD)* [19].



Gambar 1. Knowledge Discovery in Databases

2.4. Normalisasi Data

Penyetaraan kontribusi antar variabel melalui normalisasi merupakan tahapan prapemrosesan yang sangat krusial, khususnya pada implementasi algoritma yang mengandalkan kalkulasi jarak seperti *K-Means*. Prosedur ini dilakukan untuk memastikan bahwa seluruh atribut berada pada skala numerik yang

seragam, sehingga proses analisis dapat berjalan secara objektif tanpa terpengaruh oleh dominasi variabel yang memiliki rentang nilai lebih besar. [19].

Implementasi teknik *Min-Max-Normalization* dalam studi ini difokuskan pada transformasi nilai setiap atribut ke dalam *interval* [0, 1]. Langkah strategis tersebut diambil guna mengeliminasi potensi bias akibat dominasi variabel berskala besar seperti agregat lulusan pendidikan dasar terhadap variabel dengan skala lebih kecil, layaknya lulusan pascasarjana. Prosedur ini krusial untuk menjaga integritas serta presisi hasil pengelompokan agar tidak terdistorsi oleh perbedaan magnitudo data antar atribut. [20].

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2)$$

Keterangan :

x' : Representasi nilai hasil transformasi skala (*normalized value*).

x : Input data awal sebelum melalui tahapan standarisasi.

$\min(x)$: Besaran terendah dalam himpunan observasi pada variabel yang bersangkutan.

$\max(x)$: Nilai tertinggi yang ditemukan dalam distribusi data pada atribut tersebut.

3. METODE PENELITIAN

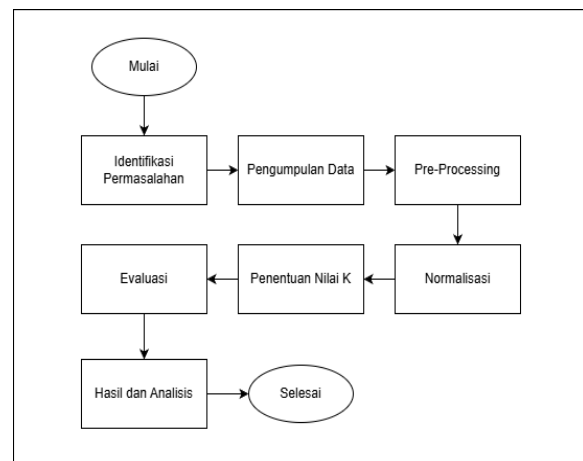
Kajian ini menitikberatkan pada evaluasi pola persebaran jenjang pendidikan di wilayah Kabupaten Karawang. Pemilihan lokus tersebut didasarkan pada signifikansi Karawang sebagai salah satu episentrum industri utama di provinsi Jawa Barat, yang memiliki karakteristik sosioekonomi beragam. Objek penelitian mencakup seluruh kecamatan di Kabupaten Karawang dengan variabel utama berupa jumlah penduduk pada berbagai jenjang pendidikan, mulai dari tidak sekolah hingga lulusan pascasarjana (S3) pada periode tahun 2020-2021.

Proses pengolahan data dilakukan menggunakan alat bantu Google Colab dengan lini masa penelitian yang dimulai dari pengumpulan data pada Oktober 2025 hingga penyusunan laporan akhir pada Januari 2026. Melalui pemetaan komprehensif terkait pola distribusi kualitas modal manusia di level kecamatan, studi ini diproyeksikan dapat menjadi instrumen evaluasi bagi otoritas wilayah dalam mengoptimalkan kebijakan publik. Hal tersebut krusial agar setiap intervensi yang dilakukan oleh pemerintah daerah memiliki akurasi yang lebih tinggi serta berdampak langsung pada sasaran yang dituju.

3.1. Prosedur Penelitian

Kerangka kerja penelitian ini dirancang secara sistematis guna menjamin aspek reproduksibilitas bagi peneliti selanjutnya yang bermaksud melakukan studi dengan parameter data sejenis. Urutan tahapan riset dipaparkan melalui bagan alir yang mengilustrasikan rangkaian proses secara holistik, mulai dari akuisisi

data hingga tahap kategorisasi dan interpretasi temuan. Integritas setiap tingkatan sangat krusial demi menjamin bahwa algoritma *K-Means* mampu menghasilkan kelompok data yang stabil, valid, serta memiliki makna ilmiah yang mendalam. Secara komprehensif, metodologi yang diterapkan meliputi akuisisi serta prapengolahan data, transformasi skala (*normalisasi*), estimasi jumlah kelompok optimal, implementasi *K-Means*, uji validitas kluster, hingga sintesis temuan akhir.



Gambar 2. Alur Proses Penelitian

3.2. Identifikasi Permasalahan

Tahap awal penelitian dimulai dengan identifikasi permasalahan yang menjadi dasar dilakukannya penelitian. Pada tahap ini, permasalahan difokuskan pada belum adanya pemetaan kecamatan berdasarkan tingkat pendidikan penduduk yang dapat menggambarkan kesamaan dan perbedaan karakteristik pendidikan secara objektif. Identifikasi permasalahan dilakukan melalui kajian awal terhadap data pendidikan dan studi literatur terkait penerapan metode *clustering* dalam analisis data pendidikan. Hasil dari tahap ini digunakan untuk merumuskan fokus penelitian, menentukan tujuan penelitian, serta menetapkan batasan penelitian agar analisis yang dilakukan tetap terarah dan relevan.

3.3. Pengumpulan Data

Tahap pengumpulan data dilakukan setelah permasalahan penelitian teridentifikasi dengan jelas. Guna menjamin validitas informasi, akuisisi data dalam studi ini dilakukan melalui repositori publik resmi milik otoritas pemerintah yang telah tervalidasi.. Dataset yang dikumpulkan mencakup informasi jumlah penduduk berdasarkan jenjang pendidikan pada tingkat kecamatan. Prosedur akuisisi data dilaksanakan dengan menitikberatkan pada aspek komprehensivitas cakupan wilayah serta standarisasi struktur informasi. Hal tersebut dilakukan guna menjamin bahwa seluruh unit observasi memiliki atribut yang utuh dan seragam, sehingga mempermudah fase integrasi data serta meminimalisir risiko anomali pada tahapan pengolahan selanjutnya.

Data yang telah dikumpulkan selanjutnya disimpan dalam format terstruktur untuk memudahkan proses analisis.

3.4. Pre Processing

Guna meningkatkan reliabilitas hasil analisis, dataset yang telah dihimpun wajib melalui fase prapengolahan sebelum dieksekusi ke tahapan komputasi utama. Prosedur ini diorientasikan untuk memitigasi anomali statistik yang dipicu oleh data yang bersifat parsial, inkonsisten, atau tidak relevan. Rangkaian aktivitasnya mencakup eliminasi nilai invalid, manajemen data kosong (*missing values*), serta sinkronisasi format data secara menyeluruh. Di samping itu, dilakukan reduksi terhadap variabel non-numerik yang tidak memiliki relevansi terhadap kalkulasi metrik jarak, seperti pengkodean administratif wilayah. *Output* dari fase ini adalah dataset yang telah tervalidasi dan siap untuk diintegrasikan ke dalam model normalisasi serta klusterisasi.

3.5. Normalisasi Data

Tahap normalisasi dilakukan untuk mengatasi perbedaan skala antar atribut dalam dataset. Jumlah penduduk pada jenjang pendidikan tertentu dapat memiliki rentang nilai yang jauh berbeda, Guna memitigasi distorsi pada kalkulasi jarak *K-Means*, riset ini mengadopsi teknik *Min-Max Normalization* untuk menyelaraskan seluruh atribut ke dalam interval $[0, 1]$. Prosedur transformasi ini esensial untuk menjamin proporsionalitas bobot setiap variabel saat penghitungan jarak *Euclidean* dilakukan. Dengan meniadakan dominasi skala antar-atribut, segmentasi yang dihasilkan menjadi lebih objektif serta mampu merepresentasikan kondisi data yang sebenarnya.

3.6. Penentuan Nilai K

Estimasi kuantitas kelompok (*K*) merupakan fase krusial dalam implementasi *K-Means* mengingat signifikansinya terhadap konfigurasi klaster yang dihasilkan. Studi ini melakukan analisis mendalam untuk mengidentifikasi jumlah kelompok paling optimal melalui pendekatan *Elbow Method*. Prosedur ini melibatkan kalkulasi variasi nilai *Sum of Squared Error (SSE)* pada berbagai tingkatan *K*. Penentuan titik *K* terbaik didasarkan pada area di mana grafik penurunan *SSE* mulai menunjukkan pola konvergensi atau melandai, yang merepresentasikan titik keseimbangan antara efisiensi jumlah kelompok dengan minimalisasi deviasi data.

3.7. Evaluasi

Tahapan validasi hasil klastering diorientasikan untuk menjamin tercapainya tingkat homogenitas internal yang optimal serta separasi antar-kelompok yang signifikan. Implementasi *Silhouette Coefficient* dalam riset ini berfungsi sebagai instrumen ukur untuk mengevaluasi derajat kohesi suatu observasi terhadap klasternya dibandingkan dengan klaster terdekat

lainnya. Perolehan skor *Silhouette* yang tinggi merepresentasikan struktur pengelompokan yang solid, di mana setiap objek telah teratribusikan pada klaster yang paling relevan secara statistik dengan batas pemisahan yang tegas.

3.8. Hasil dan Analisis

Sebagai tahapan final, penelitian ini melakukan visualisasi serta interpretasi mendalam terhadap pola-pola yang terbentuk dari proses klusterisasi. Karakteristik tingkat pendidikan pada setiap klaster kecamatan dievaluasi secara sistematis untuk mengungkap diferensiasi maupun kemiripan kondisi antarlokus wilayah. Seluruh hasil analisis tersebut digunakan sebagai basis argumentasi dalam penarikan kesimpulan akhir. Untuk menjamin kemudahan pemanfaatan hasil riset, penyajian data dilakukan secara hibrida melalui grafik visual dan deskripsi kualitatif yang mendukung kejelasan narasi penelitian.

4. HASIL DAN PEMBAHASAN

Paparan mengenai luaran dari seluruh rangkaian komputasi data sektor pendidikan di wilayah Kabupaten Karawang disajikan secara komprehensif dalam bab ini., mulai dari pembersihan data mentah hingga terbentuknya model pengelompokan Mengadopsi algoritma *K-Means*. Orientasi sentral dalam bab ini menitikberatkan pada mentransformasi data statistik yang bersifat teknis menjadi informasi strategis yang dapat memberikan gambaran nyata mengenai sebaran kualitas pendidikan di tingkat kecamatan. Melalui analisis ini, setiap temuan dijabarkan secara sistematis guna memastikan akurasi hasil yang diperoleh serta mempermudah interpretasi bagi pemetaan wilayah secara objektif.

Setiap tahapan yang disajikan merupakan hasil pemrosesan menggunakan perangkat lunak Google Colab, yang mencakup tahap normalisasi, penentuan nilai *K* optimal, hingga evaluasi kualitas klaster. Penjelasan dalam bab ini dibagi menjadi beberapa sub bahasan utama yang mengikuti alur metodologi penelitian, yang bertujuan untuk menunjukkan korelasi antara teori data mining dengan kondisi riil di lapangan. pemerintah daerah dalam merumuskan kebijakan pemerataan pendidikan di Karawang.

4.1. Objek Penelitian

Dataset yang diolah dalam kajian ini mencakup profil pendidikan masyarakat pada level kecamatan di wilayah Kabupaten Karawang. Klasifikasi data ini mengacu pada strata edukasi masyarakat yang mencakup spektrum luas, mulai dari penduduk tanpa latar belakang pendidikan formal hingga mereka yang telah menyelesaikan jenjang pendidikan tinggi, baik pada level diploma maupun sarjana dan pascasarjana (D3 hingga S3).

Data yang diperoleh masih bersifat deskriptif dan tersebar, sehingga memerlukan tahapan persiapan sebelum memasuki fase pengolahan yang lebih komprehensif. Pada tahap ini dilakukan pembersihan

data, termasuk penyesuaian nilai kosong yang menunjukkan tidak adanya penduduk pada jenjang pendidikan tertentu dengan nilai nol. Selanjutnya, data ditransformasikan ke dalam bentuk *pivot table* dengan struktur baris berupa kecamatan dan kolom berupa jenjang pendidikan. Implementasi transformasi ini diorientasikan untuk mengonstruksi skema input yang seragam, sehingga data memiliki validitas yang cukup untuk diproses pada tahapan segmentasi.

Kecamatan	BANYUSARI	BATUJAYA	CIAMPEL	CIBUAYA	CIKAMPEK
BELUM TAMAT SD	10035	8631	4260	17805	26230
D1/D2	55	98	44	35	250
D3	257	263	173	138	2875
S1	742	2148	457	414	2923
S2	17	48	8	5	97
S3	0	4	1	2	2
TAMAT SD	46181	60822	29638	42546	43894
TAMAT SLTA	15028	22572	7401	8356	70555
TAMAT SLTP	20215	23956	14723	14823	39769
TIDAK/BELUM SEKOLAH	18319	32101	25508	20416	43208

Gambar 3. Dataset

4.2. Normalisasi Data

Guna meminimalisir bias pada algoritma *K-Means* yang sangat sensitif terhadap magnitudo nilai, dataset dalam studi ini melalui fase standarisasi menggunakan teknik *Min-Max Normalization*. Transformasi ini menyelaraskan rentang nilai antar-indikator pendidikan ke dalam interval $[0, 1]$ tanpa mengintervensi pola distribusi orisinal data antar wilayah. Penyeragaman skala tersebut merupakan langkah krusial agar setiap variabel memberikan kontribusi bobot yang adil dalam kalkulasi metrik jarak selama proses segmentasi kecamatan dilaksanakan.

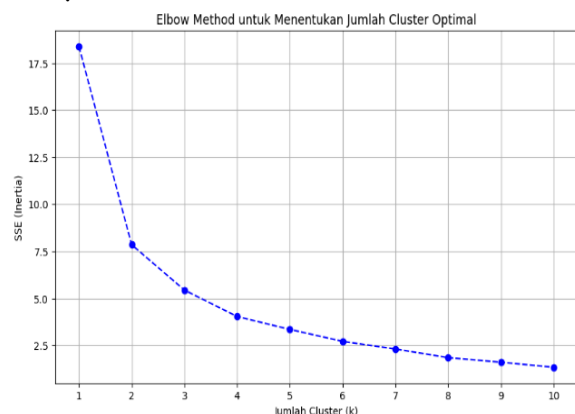
Kecamatan	BANYUSARI	BATUJAYA	CIAMPEL	CIBUAYA	CIKAMPEK
BELUM TAMAT SD	0.259	0.208	0.050	0.539	0.842
D1/D2	0.032	0.100	0.014	0.000	0.340
D3	0.019	0.020	0.008	0.003	0.372
S1	0.028	0.139	0.006	0.002	0.200
S2	0.022	0.079	0.006	0.000	0.169
S3	0.000	0.129	0.032	0.065	0.065
TAMAT SD	0.435	0.687	0.152	0.373	0.396
TAMAT SLTA	0.074	0.147	0.001	0.010	0.607
TAMAT SLTP	0.326	0.393	0.228	0.230	0.674
TIDAK/BELUM SEKOLAH	0.153	0.351	0.257	0.183	0.511

Gambar 4. Normalisasi Data

4.3. Penentuan Jumlah K

Estimasi kuantitas kelompok optimal dalam studi ini diimplementasikan melalui pendekatan *Elbow Method* dengan mengevaluasi fluktuasi nilai *Sum of Squared Error (SSE)* pada berbagai skenario jumlah kluster. Prosedur ini ditujukan untuk mengidentifikasi "titik siku" di mana reduksi nilai *SSE* tidak lagi

menunjukkan penurunan drastis, yang mengindikasikan bahwa penambahan kluster tidak lagi memberikan peningkatan kualitas segmentasi yang substansial. Berdasarkan observasi grafik, teramati adanya penurunan *SSE* yang signifikan hingga koordinat $K = 3$, sebelum akhirnya pola grafik mulai melandai. Meskipun terdapat indikasi pembentukan empat kelompok, hasil validasi menunjukkan bahwa konfigurasi $K = 4$ menghasilkan kluster tunggal (*single-member cluster*) yang hanya mencakup satu kecamatan. Oleh sebab itu, riset ini menetapkan $K = 3$ sebagai jumlah kluster terpilih guna menjamin proporsionalitas distribusi anggota serta memudahkan proses interpretasi dalam konteks disparitas pendidikan di Kabupaten Karawang.



Gambar 5. Penentuan Jumlah K

4.4. Evaluasi Model (*Silhouette Coefficient*)

Setelah menentukan jumlah kluster melalui *Elbow Method*, tahap selanjutnya adalah menguji validitas dan kekuatan struktur kluster menggunakan metode *Silhouette Coefficient*. Berdasarkan hasil pengujian terhadap nilai $K = 3$, Perolehan skor rata-rata 0,5686 menunjukkan struktur kluster yang solid dan konsisten karena telah melampaui ambang batas 0,5. Angka tersebut menjadi bukti matematis bahwa tiap kecamatan telah teralokasi pada kelompok yang tepat dengan kohesi internal tinggi dan separasi antar-kluster yang tegas. Validasi ini mengonfirmasi stabilitas kategori pendidikan (Tinggi, Berkembang, dan Dasar) sebagai landasan kebijakan daerah yang andal.

```

Silhouette Score buat k=2: 0.5953
Silhouette Score buat k=3: 0.5686
Silhouette Score buat k=4: 0.2920
Silhouette Score buat k=5: 0.2854
Silhouette Score buat k=6: 0.2750
Silhouette Score buat k=7: 0.2304
Silhouette Score buat k=8: 0.2447
Silhouette Score buat k=9: 0.2553
Silhouette Score buat k=10: 0.2590

```

. Gambar 6. *Silhouette Coefficient Score*

4.5. Hasil Klasterisasi

Proses klasterisasi data tingkat pendidikan penduduk di Kabupaten Karawang dilakukan dengan menerapkan metode *K-Means* dengan Konfigurasi tiga kelompok dipilih sebagai jumlah klaster optimal dalam penelitian ini guna menggambarkan sebaran data secara representatif. Penetapan tiga klaster dilakukan untuk memperoleh pengelompokan yang jelas, stabil, dan mudah diinterpretasikan dalam analisis kondisi pendidikan antar kecamatan. Setiap kecamatan dikelompokkan berdasarkan kemiripan distribusi penduduk menurut spektrum pendidikan, Mencakup rentang dari masyarakat tanpa latar belakang pendidikan formal sampai dengan kualifikasi pendidikan tinggi..

Hasil klasterisasi menunjukkan perbedaan karakteristik pendidikan yang nyata pada masing masing klaster. Klaster pertama mencerminkan wilayah dengan tingkat pendidikan relatif rendah, klaster kedua menggambarkan kondisi pendidikan sedang, sedangkan klaster ketiga menunjukkan dominasi pendidikan menengah dan tinggi. Hasil ini menjadi dasar analisis lebih lanjut mengenai ketimpangan pendidikan antar kecamatan serta perumusan rekomendasi kebijakan pendidikan daerah secara berkelanjutan dan terarah nasional.

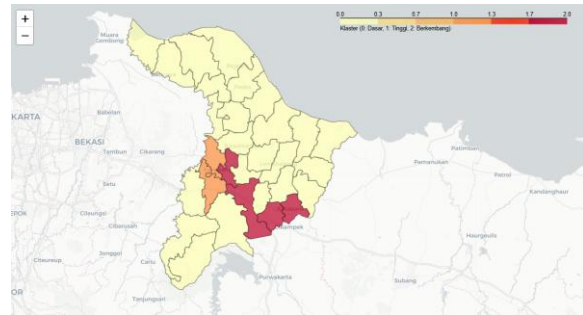
Cluster	0	1	2
BELUM TAMAT SD	11859.29	9428.5	25924.75
D1/D2	80.17	599.5	355.75
D3	350.50	6506.5	4402.00
S1	1311.38	11217.0	6619.25
S2	31.92	520.5	184.75
S3	2.79	28.5	7.00
TAMAT SD	42827.96	45133.5	54909.00
TAMAT SLTA	17781.54	88509.0	77691.75
TAMAT SLTP	17629.08	21695.5	44369.75
TIDAK/BELUM SEKOLAH	24713.12	47995.5	59433.75

Gambar 7. Hasil Klasterisasi

4.6. Visualisasi Geografis

Visualisasi geografis hasil pengelompokan pada peta Kabupaten Karawang di Gambar 7 memperlihatkan pola sebaran spasial yang sangat kontras antara wilayah pusat pertumbuhan dengan wilayah pinggiran. Klaster 1 (Tinggi Merah) dan Klaster 2 (Berkembang Oranye) tampak terkonsentrasi di area pusat administrasi dan zona industri utama seperti Kecamatan Karawang Barat, Karawang Timur, Kotabaru, dan Cikampek. Sebaliknya, Klaster 0 (Dasar Kuning) mendominasi hampir seluruh wilayah bagian utara dan pesisir, yang mengindikasikan bahwa akses serta capaian pendidikan tinggi masih berpusat pada area dengan infrastruktur ekonomi yang lebih maju. Peta ini memberikan bukti visual yang kuat bagi pemerintah daerah mengenai adanya kesenjangan kualitas sumber

daya manusia secara kewilayahan di Kabupaten Karawang.



Gambar 8. Visualisasi Geograis

5. KESIMPULAN DAN SARAN

Merujuk pada temuan studi ini, implementasi algoritma *K-Means* berhasil memetakan 30 kecamatan di Kabupaten Karawang ke dalam tiga klaster tingkat pendidikan secara objektif. Penentuan jumlah klaster optimal melalui *Elbow Method* menghasilkan nilai $K = 3$ dengan tingkat kesalahan (SSE) yang minimal, yang menunjukkan struktur kelompok yang stabil. Hasil pemetaan mengidentifikasi adanya kesenjangan kualitas SDM, di mana klaster pendidikan tinggi (Klaster 1) berpusat di area industri, sedangkan wilayah pinggiran masih didominasi klaster pendidikan dasar (Klaster 0). Sebagai saran, Pemerintah Kabupaten Karawang perlu memprioritaskan bantuan pendidikan di wilayah Klaster 0, sementara penelitian selanjutnya bisa mencoba membandingkan hasil ini Melalui komparasi dengan pendekatan alternatif seperti *Fuzzy C-Means* atau menambah variabel seperti jumlah guru agar hasil pemetaannya makin tajam.

DAFTAR PUSTAKA

- [1] E. Simatupang And I. Yuhertiana, "Merdeka Belajar Kampus Merdeka Terhadap Perubahan Paradigma Pembelajaran Pada Pendidikan Tinggi: Sebuah Tinjauan Literatur," 2021.
- [2] N. L. Made Ariasih And N. N. Yuliarini, "Pengaruh Tingkat Pendidikan, Tingkat Kesehatan Dan Pengangguran Terbuka Terhadap Tingkat Kemiskinan Di Provinsi Bali," *Cerdika: Jurnal Ilmiah Indonesia*, Vol. 1, No. 7, Pp. 802–839, Jul. 2021, Doi: 10.36418/Cerdika.V1i7.131.
- [3] A. Ika Deswanti, M. Asbari, D. Novitasari, And A. Purwanto, "Pengaruh Tingkat Pendidikan Dan Pengalaman Kerja Terhadap Kinerja Karyawan: Narrative Literature Review," *Journal Of Information Systems And Management*, Vol. 02, No. 03, 2023, [Online]. Available: <https://jisma.org>
- [4] A. Vega, I. Vanezha, J. Agustia, M. Rizqi, And R. Uzdah, "Kesetaraan Akses Pendidikan: Analisis Pengimplementasian Nilai Pancasila Dalam Pemerataan Akses Pendidikan Di Indonesia," *Jli*, 2024.

- [5] A. Yudistira And R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode *K-Means* Clustering," *Journal Of Artificial Intelligence And Technology Information (Jaiti)*, Vol. 1, No. 1, Pp. 20–28, Mar. 2023, Doi: 10.58602/Jaiti.V1i1.22.
- [6] D. Marcelina, A. Kurnia, And T. Terttiaavini, "Analisis Klaster Kinerja Usaha Kecil Dan Menengah Menggunakan Algoritma *K-Means* Clustering," *Malcom: Indonesian Journal Of Machine Learning And Computer Science*, Vol. 3, No. 2, Pp. 293–301, Nov. 2023, Doi: 10.57152/Malcom.V3i2.952.
- [7] M. R. Putri, G. Satya Nugraha, And R. Dwiyanaputra, "Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Pendidikan Menggunakan Metode *K-Means* Clustering Grouping Provinces In Indonesia Based On Education Indicators Using The *K-Means* Clustering Method," 2023. [Online]. Available: [Http://Jcosine.If.Unram.Ac.Id/](http://Jcosine.If.Unram.Ac.Id/)
- [8] R. E. Pawening, "Algoritma *K-Means* Untuk Mengukur Kepuasan Mahasiswa Menggunakan E Learning," *Journal Of Technology And Informatics (Joti)*, Vol. 3, No. 1, Pp. 27–33, Oct. 2021, Doi: 10.37802/Joti.V3i1.201.
- [9] A. Lia Hananto *Et Al.*, "Analysis Of Drug Data Mining With Clustering Technique Using *K-Means* Algorithm," In *Journal Of Physics: Conference Series*, Iop Publishing Ltd, Jul. 2021. Doi: 10.1088/1742 6596/1908/1/012024.
- [10] Wenny, "Bulletin Of Information System Research (Bios) Normalisasi Data Kependudukan Dengan Model Min Max Dan Algoritma *K-Means* Untuk Pengelompokkan Tingkat Ekonomi Masyarakat," 2024, [Online]. Available: [Https://Journal.Grahamitra.Id/Index.Php/Bios](https://Journal.Grahamitra.Id/Index.Php/Bios)
- [11] M. F. Zahrotun, " *K-Means* Centroid Optimization With Genetic Algorithm For Clustering Micro, Small, Medium Enterprises In Yogyakarta," 2025.
- [12] M. Ahmed, R. Seraj, And S. M. S. Islam, "The *K-Means* Algorithm: A Comprehensive Survey And Performance Evaluation," Aug. 01, 2022, *Mdpi Ag*. Doi: 10.3390/Electronics9081295.
- [13] A. Kurniawan *Et Al.*, *Dasar Dasar Ilmu Pendidikan*. 2022. [Online]. Available: [Www.Globaleksekutifteknologi.Co.Id](http://www.Globaleksekutifteknologi.Co.Id)
- [14] R. Kurniawan And M. Mukarrobini, "Klasterisasi Tingkat Pendidikan Di Dki Jakarta Pada Tingkat Kecamatan Menggunakan Algoritma *K-Means*," 2021. [Online]. Available: [Https://Data.Jakarta.Go.Id](https://Data.Jakarta.Go.Id).
- [15] M. Djaka Permana, A. Lia Hananto, E. Novalia, B. Huda, And T. Paryono, "Klasterisasi Data Jamaah Umrah Pada Tanurmutmainah Tour Menggunakan Algoritma *K-Means*," *Jurnal Komtekinfo*, Pp. 15–20, Feb. 2023, Doi: 10.35134/Komtekinfo.V10i1.332.
- [16] A. L. Hananto, F. Nurapriani, And B. Huda, "Implementasi Algoritma *K-Means* Untuk Klasterisasi Data Stunting Di Kabupaten Bekasi," *Remik: Riset Dan E Jurnal Manajemen Informatika Komputer*, Vol. 9, No. 2, 2025, Doi: 10.33395/Remik.V9i2.14748.
- [17] I. Melani, B. Priyatna, F. Nurapriani, And S. Shofia Hilabi, "Implementasi Metode *K-Means* Clustering Pada Penilaian Kinerja Karyawan Pt Kopetri Citra Abadi," 2023. [Online]. Available: [Http://E Journal.Janabadra.Ac.Id/](http://E Journal.Janabadra.Ac.Id/)
- [18] S. Sfofiah Hilabi *Et Al.*, "Pemanfaatan Data Analitik Dalam Big Data: Studi Kasus Implementasi Di Pemerintahan," Vol. 12, No. 1, Pp. 378–390, 2025, [Online]. Available: [Http://Jurnal.Mdp.Ac.Id](http://Jurnal.Mdp.Ac.Id)
- [19] N. Nuraeni, "Klasifikasi Data Mining Untuk Prediksi Potensi Nasabah Dalam Membuat Deposito Berjangka Data Mining Classification For Predicting Customer Potential In Making Term Deposits," *Jurnal Ilmiah Intech : Information Technology Journal Of Umus*, Vol. 3, No. 01, Pp. 65–75, 2021.
- [20] R. A. Farissa, R. Mayasari, And Y. Umaidah, "Perbandingan Algoritma *K-Means* Dan *K Medoids* Untuk Pengelompokkan Data Obat Dengan Silhouette Coefficient," 2021. [Online]. Available: [Http://Jurnal.Polibatam.Ac.Id/Index.Php/Jaic](http://Jurnal.Polibatam.Ac.Id/Index.Php/Jaic)