

Implementasi *Geospatial Clustering* Menggunakan Algoritma *K-Means* Untuk Penentuan Lokasi Strategis Promosi Kampus Universitas Langlangbuana

Implementation of Geospatial Clustering Using K-Means Algorithm for Determining the Strategic Location of Campus Promotion at Langlangbuana University

Dila Siti Nurfadhilah^{*1}, Awan Setiawan², Ridwan Zulkifli³

^{1,2}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Langlangbuana

³Program Studi Teknik Informatika, Fakultas Teknik, Universitas Ma'soem

*e-mail: dilastnrf@gmail.com¹, awans2425@gmail.com², ridwan.zulkifli12@gmail.com³

Received:
10.09.2025

Revised:
22.09.2025

Accepted:
11.10.2025

Available online:
20.10.2025

Abstrak: *Data mining* merupakan teknik untuk mengekstraksi informasi bernilai dari kumpulan data berskala besar. Di lingkungan perguruan tinggi, proses seleksi mahasiswa baru menghasilkan kumpulan data yang kaya, khususnya terkait latar belakang asal sekolah calon mahasiswa. Namun, jika data tersebut tidak dianalisis dan dimanfaatkan secara maksimal, potensi informasinya akan menurun seiring berjalannya waktu. Berdasarkan data yang diperoleh dari Universitas Langlangbuana, ditemukan adanya penurunan jumlah mahasiswa baru selama tiga tahun terakhir jika dibandingkan dengan tahun 2021. Untuk merespons kondisi ini, dibutuhkan strategi berbasis data yang mampu mengidentifikasi wilayah promosi secara lebih tepat. Penelitian ini menerapkan metode *geolocation clustering* dengan algoritma *K-Means* untuk mengelompokkan data asal sekolah mahasiswa. Sistem dikembangkan menggunakan bahasa pemrograman *Python*, dengan fokus pada segmentasi wilayah berdasarkan sebaran sekolah. Hasil analisis menghasilkan tujuh kluster utama, dengan distribusi tertinggi berada di wilayah Jawa Barat, khususnya Kota Bandung dan Kabupaten Bandung. Temuan ini memberikan gambaran wilayah potensial sebagai sasaran promosi guna meningkatkan jumlah pendaftar di masa yang akan datang. Pendekatan ini memungkinkan institusi pendidikan untuk merumuskan strategi pemasaran yang lebih efektif dan efisien berdasarkan data yang tersedia.

Kata kunci: *Geospatial Clustering, K-Means, Data mining, Elbow Method*

Abstract: *Data mining* is a technique for extracting valuable information from large-scale data sets. In higher education, the new student selection process generates a rich data set, especially related to the background of prospective students' school of origin. However, if the data is not analyzed and utilized optimally, its information potential will decrease over time. Based on data obtained from Langlangbuana University, it was found that there was a decrease in the number of new students for the last three years when compared to 2021. To respond to this condition, a data-based strategy is needed that is able to identify promotional areas more precisely. This research applies the *geolocation clustering* method with the *K-Means* algorithm to cluster student school origin data. The system was developed using the *Python* programming language, with a focus on regional segmentation based on school distribution. The analysis resulted in seven main clusters, with the highest distribution being in the West Java region, particularly Bandung City and Bandung Regency. This finding provides an overview of potential areas to target for promotion to increase the number of applicants in the future. This approach allows educational institutions to formulate more effective and efficient marketing strategies based on available data.

Keywords: *Geospatial Clustering, K-Means, Data mining, Elbow Method*

1. PENDAHULUAN

Data mining merupakan teknik yang digunakan untuk mengekstraksi informasi bernilai dari kumpulan data berskala besar melalui proses penggalian pola tersembunyi dalam data. Di era transformasi digital, pemanfaatan data sebagai aset strategis menjadi kunci keberhasilan dalam menghadapi tantangan kompetitif, termasuk dalam sektor pendidikan tinggi. Institusi pendidikan yang mampu mengelola dan menganalisis data secara sistematis memiliki keunggulan dalam merancang strategi pemasaran dan pengambilan keputusan yang berbasis pada pola historis dan tren yang teridentifikasi dalam data. Sayangnya, banyak institusi pendidikan masih terbatas dalam hal pemanfaatan data secara optimal, khususnya dalam konteks promosi dan rekrutmen mahasiswa baru, sehingga strategi yang diterapkan sering kali bersifat generik dan kurang berdampak. Dalam hal ini, volume data yang besar dan terus bertambah dapat menimbulkan kesulitan dalam mencari informasi dan

berpotensi menyebabkan redundansi data, yang pada akhirnya menghambat efisiensi pengambilan keputusan [1].

Berdasarkan data yang didapat, jumlah mahasiswa baru menunjukkan tren penurunan yang cukup signifikan. Tercatat 1.806 mahasiswa baru pada tahun 2021, menurun menjadi 1.608 pada tahun 2022, 1.385 pada tahun 2023, dan 1.282 pada tahun 2024. Fenomena ini menunjukkan adanya pergeseran minat pendaftar yang perlu dikaji lebih lanjut. Penurunan ini dapat disebabkan oleh beberapa faktor, seperti peningkatan persaingan antarperguruan tinggi, perubahan preferensi calon mahasiswa terhadap jurusan, serta belum optimalnya strategi promosi kampus dalam menjangkau wilayah potensial. Oleh karena itu, penelitian ini bertujuan untuk menganalisis persebaran geografis asal sekolah mahasiswa Universitas Langlangbuana menggunakan metode *K-Means clustering* agar diperoleh segmentasi wilayah strategis untuk kegiatan promosi yang lebih efektif dan berbasis data. Ruang lingkup penelitian dibatasi pada data mahasiswa baru periode 2021–2024 yang berasal dari data internal Universitas Langlangbuana. Analisis difokuskan pada pengelompokan wilayah asal sekolah berdasarkan koordinat geografis (*latitude-longitude*), tanpa membahas faktor sosial ekonomi atau minat jurusan. Penentuan lokasi promosi penerimaan mahasiswa baru menggunakan metode *data mining* dapat memberikan hasil yang lebih efektif dan terfokus, sehingga berpotensi meningkatkan jumlah pendaftar mahasiswa baru setiap tahunnya [2].

Untuk mengatasi permasalahan tersebut, dibutuhkan pendekatan berbasis data (*data-driven decision making*) yang mampu mengintegrasikan teknologi informasi dengan analisis data yang sistematis. Salah satu pendekatan yang relevan adalah penerapan teknik *data mining* dengan algoritma *clustering* seperti *K-Means*, yang mampu mengelompokkan data berdasarkan kemiripan karakteristik tertentu, seperti wilayah geografis atau asal sekolah calon mahasiswa. Metode ini telah terbukti efektif dalam berbagai penelitian untuk mendukung pengambilan keputusan dalam sektor pendidikan, termasuk dalam menentukan strategi promosi, lokasi prioritas rekrutmen, serta segmentasi calon mahasiswa potensial.

Kontribusi utama penelitian ini terletak pada penerapan analisis *geospatial* berbasis *K-Means* yang tidak hanya mengelompokkan wilayah asal sekolah mahasiswa, tetapi juga memanfaatkan data dari universitas untuk memetakan area potensial promosi secara visual dan terukur. Pendekatan ini memperkuat strategi promosi kampus melalui analisis spasial yang lebih mendalam dan berbasis data aktual. Dengan pendekatan ini, diharapkan terjadi peningkatan jumlah pendaftar di Universitas Langlangbuana pada tahun-tahun berikutnya.

2. TINJAUAN PUSTAKA

2.1. *Data mining*

Data mining adalah proses yang menggunakan satu atau lebih teknik *machine learning* untuk menganalisis dan secara otomatis mengekstraksi pengetahuan atau informasi yang bermanfaat dari sekumpulan data [3]. Selain itu, *data mining* juga bertujuan untuk menemukan pola atau hubungan tersembunyi dalam data melalui metode analisis modern yang mudah dipahami dan memberikan nilai tambah bagi pemilik data [4].

2.2. *Geospatial*

Geospatial adalah aspek spasial yang menggambarkan lokasi, posisi, dan koordinat suatu objek atau peristiwa di permukaan bumi dalam sistem koordinat tertentu. Informasi *geospatial* adalah data *geospatial* yang telah diproses sehingga dapat digunakan dalam perumusan kebijakan, pengambilan keputusan, atau pelaksanaan kegiatan yang berkaitan dengan permukaan bumi [5]. Dalam konteks big data, penanganan data *geospatial* memerlukan metode dan teori khusus untuk mengatasi tantangan volume besar, kecepatan, dan variasi data [6].

2.3. *Clustering*

Clustering digunakan untuk mengelompokkan data atau pola yang belum ditentukan labelnya melalui proses tanpa pengawasan, sehingga menghasilkan grup dengan karakteristik spesifik. Penerapannya sangat bermanfaat, misalnya pada analisis pola, proses pengambilan keputusan, pembelajaran mesin, maupun penambangan data [7]. Proses pengelompokan data atau objek ke dalam

cluster dilakukan dengan mempertimbangkan kesamaan atribut, yang dikenal sebagai *clustering*. Teknik ini merupakan salah satu pendekatan dalam *data mining*. Pengelompokan yang efektif menghasilkan objek dalam satu *cluster* yang sangat mirip, namun berbeda dengan objek pada *cluster* lainnya [8].

2.4. *K-Means Algorithm*

K-Means clustering adalah metode analisis kluster non-hierarkis yang bertujuan untuk membagi sekumpulan objek menjadi satu atau lebih kluster berdasarkan kesamaan karakteristiknya. Objek dengan karakteristik serupa akan dikelompokkan ke dalam kluster yang sama, sementara objek dengan karakteristik berbeda akan ditempatkan ke dalam kluster yang berbeda [9]. Algoritma *K-Means* telah mengalami berbagai pengembangan dan variasi untuk meningkatkan kinerjanya dalam menangani berbagai jenis dataset [10]. Metode *K-Means* membagi data menjadi beberapa kelompok berdasarkan kesamaan karakteristik, di mana data dengan karakteristik yang sama dikelompokkan bersama, sedangkan data yang berbeda ditempatkan dalam kelompok lain, sehingga memungkinkan pengelompokan dengan tingkat variasi yang kecil di dalam setiap kelompok [11].

2.5. *Metode Elbow*

Elbow Method adalah teknik yang digunakan untuk membantu menentukan jumlah *cluster* yang optimal dengan menganalisis persentase perubahan dalam jumlah *cluster* hingga membentuk pola menyerupai siku pada grafik. Metode ini memberikan pendekatan dengan mencoba berbagai nilai jumlah *cluster*, lalu menambahkannya secara bertahap untuk membentuk model yang dapat mengidentifikasi jumlah *cluster* terbaik [12].

Rumus perhitungan *Elbow Method* merepresentasikan jumlah kelompok yang digunakan dalam algoritma *K-Means*, X_i adalah jumlah data, dan C_k menunjukkan jumlah *cluster* pada *cluster* ke- K . Tujuan dari metode ini adalah meminimalkan nilai *Sum of Squared Errors (SSE)* atau *Inertia Cluster*. Prosesnya dilakukan dengan menentukan titik di mana nilai SSE mulai melandai dan membentuk sudut seperti siku. Dataset akan dievaluasi menggunakan berbagai nilai K , lalu dianalisis titik di mana grafik mulai melengkung. Hasil yang optimal ditunjukkan oleh nilai *inertia* yang paling rendah. Visualisasi ini menghasilkan *elbow curve* yang membantu dalam mengidentifikasi jumlah *cluster* terbaik berdasarkan nilai *inertia* dari setiap model [13].

2.6. *Python*

Python merupakan bahasa pemrograman berorientasi objek yang termasuk dalam kategori scripting language. Bahasa ini dapat dimanfaatkan untuk pengembangan perangkat lunak dan mendukung berbagai sistem operasi. Saat ini, *Python* menjadi salah satu bahasa yang populer di bidang data science dan analisis data. Kepopulerannya didorong oleh ketersediaan berbagai library yang menyediakan fungsi untuk analisis data, *machine learning*, data *preprocessing*, serta visualisasi data secara efektif dan efisien [14].

3. METODE

Metode penelitian ini bersifat terapan (applied research), di mana penulis mengimplementasikan teknik *data mining* dengan algoritma *K-Means Clustering* untuk mengelompokkan data mahasiswa Universitas Langlangbuana berdasarkan persebaran wilayah asal sekolah pada periode 2021–2024. Tujuan dari penerapan metode ini adalah untuk mengidentifikasi wilayah potensial promosi kampus secara berbasis data. Rangkaian tahapan penelitian secara menyeluruh disajikan pada Gambar 1.

3.1. *Data Selection*

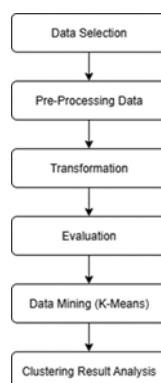
Data yang digunakan merupakan data mahasiswa baru Universitas Langlangbuana periode 2021–2024. Atribut yang dipilih mencakup NPM, nama, asal sekolah, dan tahun masuk, sesuai dengan fokus analisis. Data diperoleh dari sistem akademik universitas.

3.2. *Pre-Processing Data*

Tahap ini bertujuan untuk membersihkan dan menyiapkan data sebelum dianalisis. Proses dilakukan menggunakan Microsoft Excel dan *Python* dengan tahapan:

- Menghapus duplikasi dan data kosong.

- Menyeragamkan nama sekolah (contoh: “SMAN 1 Bandung” dan “SMA Negeri 1 Bandung”).
- Menyimpan data bersih dalam format .csv.



Gambar 1. Metode Penelitian

3.3. Transformation

Data asal sekolah mahasiswa dikonversi ke koordinat geografis (latitude dan longitude) menggunakan layanan *Google Maps*. Hasil konversi disimpan dalam format .csv untuk selanjutnya divisualisasikan dalam peta digital menggunakan library *Folium*.

3.4. Data Mining

Analisis pengelompokan dilakukan dengan algoritma *K-Means Clustering*, yang bertujuan mengelompokkan data ke dalam sejumlah cluster berdasarkan kesamaan lokasi geografis. Proses algoritma *K-Means* secara umum Adalah [15]:

1. Menentukan jumlah cluster (K) awal.
2. Menentukan titik pusat awal (*centroid*) secara acak.
3. Menghitung jarak antara setiap data terhadap *centroid* menggunakan *Euclidean Distance*:

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^n (x_{ik} - c_{jk})^2}$$

4. Menetapkan data ke cluster dengan jarak terdekat.
5. Menghitung ulang posisi *centroid* baru:

$$\mu_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i$$

6. Mengulangi proses hingga nilai *centroid* stabil (*convergent*).

Tujuan optimisasi *K-Means* adalah meminimalkan *Sum of Squared Errors (SSE)*:

$$J = \sum_{j=1}^k \sum_{i=1}^{n_j} ||x_i - \mu_j||^2$$

Dalam penelitian ini, algoritma dijalankan menggunakan **Python 3.11** dengan library *scikit-learn*. Jumlah iterasi maksimum ditetapkan sebanyak **300**, dengan metode inisialisasi **K-Means++** untuk menghindari kesalahan pemilihan *centroid* awal.

3.5. Evaluation

Untuk menentukan jumlah cluster optimal, digunakan *Elbow Method* dengan menghitung nilai SSE pada berbagai nilai K (1–10). Nilai K yang menghasilkan titik siku (*elbow point*) dianggap sebagai jumlah cluster terbaik. Selain itu, hasil cluster divalidasi menggunakan *Silhouette Score*, dengan rumus:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

di mana:

- $a(i)$ = rata-rata jarak antar titik dalam satu cluster,
- $b(i)$ = rata-rata jarak titik ke cluster terdekat.

Interpretasi nilai Silhouette:

- 0.71–1.00 : Cluster sangat baik
- 0.51–0.70 : Cukup baik
- 0.26–0.50 : Sedang
- < 0.25 : Kurang baik

3.6. *Clustering Result Analysis*

Hasil *clustering* divisualisasikan dalam bentuk peta digital interaktif untuk menampilkan persebaran wilayah asal sekolah tiap cluster. Setiap cluster ditandai dengan warna berbeda sesuai dengan hasil pengelompokan. Selain itu, data hasil *clustering* diekspor ke format .csv yang berisi informasi tambahan seperti Cluster ID dan koordinat masing-masing sekolah.

Analisis dilakukan dengan mengamati pola dominasi wilayah setiap cluster untuk mengidentifikasi area promosi potensial, seperti wilayah dengan kontribusi pendaftar tertinggi atau yang mengalami tren peningkatan dari tahun ke tahun.

4. HASIL DAN PEMBAHASAN

4.1. *Data Selection*

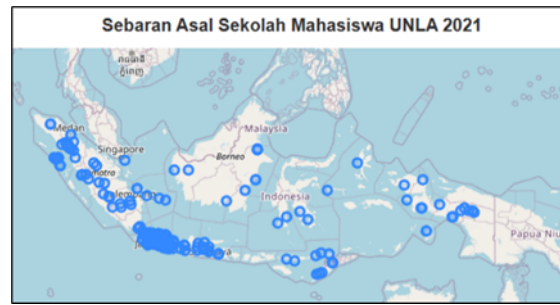
Penelitian ini menggunakan data mahasiswa Universitas Langlangbuana dari tahun akademik 2021 hingga 2024 sebagai objek analisis. Jumlah mahasiswa pada setiap tahun akademik adalah: 1.803 mahasiswa pada 2021, 1.608 pada 2022, 1.385 pada 2023, dan 1.282 pada 2024. Data ini digunakan untuk melihat tren penurunan jumlah mahasiswa sekaligus menganalisis sebaran geografis asal sekolah mahasiswa.

4.2. *Pre-processing Data*

Proses pra-pemrosesan dilakukan untuk memastikan data bersih dan konsisten. Atribut yang digunakan hanya *Nomor Identitas Mahasiswa* dan *asal sekolah*. Setelah pembersihan data, diperoleh data valid sebanyak 1.491 (2021), 1.350 (2022), 1.092 (2023), dan 933 (2024) entri. Langkah ini penting untuk menghindari bias pada hasil *clustering* dan memastikan bahwa hasil pengelompokan benar-benar mencerminkan kondisi sebenarnya.

4.3. *Transformation*

Data asal sekolah dikonversi menjadi koordinat geografis (latitude dan longitude) menggunakan *Google Maps*. Koordinat ini divisualisasikan menggunakan library *Folium*, menghasilkan peta sebaran asal sekolah mahasiswa tiap tahun. Dari visualisasi awal, tampak bahwa mayoritas asal sekolah mahasiswa tersebar di wilayah Jawa Barat, terutama Kota Bandung dan Kabupaten Bandung. Berikut contoh hasil implementasi data asal sekolah mahasiswa UNLA 2021 kedalam map disajikan pada Gambar 2.



Gambar 2. Hasil Implementasi Data Asal Sekolah Mahasiswa UNLA 2021 Kedalam Map

4.4. Evaluation

Langkah selanjutnya adalah menentukan jumlah cluster optimal menggunakan *Elbow Method*. Sistem membaca data koordinat geografis dari variabel “X”, menentukan *centroid* secara acak, lalu menghitung nilai *Sum of Squared Errors (SSE)* untuk setiap jumlah cluster yang diuji dalam rentang $K = 1$ hingga 101. Tujuan tahap ini adalah menemukan jumlah cluster dengan nilai SSE yang relatif kecil dan efisien. Berikut hasil implementasi *Elbow Method* menunjukkan jumlah cluster optimal ditunjukkan pada Tabel 1.

Tabel 1. Hasil Evaluasi Penentuan Jumlah Cluster Menggunakan *Elbow Method*

Tahun	Rentang K	Jumlah Cluster Optimal (K)	Nilai SSE	<i>Silhouette Score</i>	Keterangan
2021	1–101	7	0.662	0.56	Nilai SSE menurun tajam hingga $K=7$ kemudian mulai stabil, menunjukkan titik siku (<i>elbow point</i>) yang jelas.
2022	1–101	7	0.640	0.54	Penurunan SSE signifikan hingga $K=7$, dengan nilai <i>Silhouette</i> yang cukup baik antar-cluster.
2023	1–101	7	0.617	0.53	Pola grafik menunjukkan kestabilan setelah $K=7$, mengindikasikan jumlah cluster yang efisien.
2024	1–101	6	0.583	0.55	Kurva SSE mulai landai setelah $K=6$, sehingga jumlah cluster optimal berada pada $K=6$.

Dari hasil tersebut, terlihat bahwa jumlah cluster optimal berada pada $K=7$ untuk tahun 2021–2023 dan $K=6$ untuk tahun 2024. Nilai *Silhouette Score* yang berkisar antara 0.53 hingga 0.56 menunjukkan bahwa hasil pengelompokan memiliki kualitas yang cukup baik, dengan jarak antar-cluster yang jelas dan kemiripan tinggi antar data dalam cluster yang sama.

Rentang pengujian yang lebih luas ($K=1-101$) memberikan gambaran yang lebih komprehensif terhadap stabilitas model *clustering*. Peningkatan jumlah cluster di atas $K=7$ tidak lagi memberikan penurunan *SSE* yang signifikan, sehingga dapat disimpulkan bahwa jumlah cluster tersebut merupakan konfigurasi paling efisien dan representatif terhadap data yang digunakan.

4.5. Data Mining

Setelah jumlah cluster optimal diperoleh melalui *Elbow Method*, tahap selanjutnya adalah penerapan algoritma *K-Means Clustering* pada data hasil *pre-processing* dan *transformation*. Proses ini dilakukan menggunakan bahasa pemrograman *Python 3.11* dengan library *scikit-learn*, *numpy*, dan *pandas*. Analisis dilakukan secara terpisah untuk setiap tahun akademik, yaitu 2021–2024. Parameter yang digunakan dalam proses *clustering* adalah sebagai berikut:

- Metode inisialisasi *centroid*: *K-Means++* untuk memastikan pemilihan titik pusat awal yang efisien,
- Jumlah iterasi maksimum: 300 kali,
- Metode pengukuran jarak: *Euclidean Distance*,
- Nilai random state: 42 untuk menjaga konsistensi hasil,

- Kriteria berhenti: perubahan nilai *centroid* lebih kecil dari 0.0001 (convergence threshold).

Hasil pengujian menunjukkan bahwa *Silhouette Score* berada pada kisaran 0.53–0.56, yang menandakan kualitas *clustering* tergolong cukup baik (*fair clustering quality*). Artinya, data dalam setiap cluster memiliki kemiripan tinggi dan terpisah dengan jelas dari cluster lainnya.

Hasil pengelompokan kemudian divisualisasikan menggunakan library *Folium* dalam bentuk peta digital interaktif. Setiap titik mewakili asal sekolah mahasiswa, sedangkan perbedaan warna menandakan perbedaan cluster. Berikut contoh hasil implementasi konversi data mahasiswa UNLA 2021, yang telah dikelompokkan ke dalam peta, ditampilkan pada Gambar 3.



Gambar 3. Hasil Implementasi Konversi Data kedalam Map

Dari hasil visualisasi terlihat bahwa sebagian besar titik dengan warna serupa terkonsentrasi di wilayah Kota Bandung dan Kabupaten Bandung, menandakan dominasi asal sekolah dari wilayah tersebut. Sebaran yang lebih jarang ditemukan di wilayah luar Jawa Barat, menunjukkan area dengan potensi promosi yang masih dapat dikembangkan.

4.6. Clustering Result Analysis

Tahap ini berfokus pada analisis interpretatif hasil *clustering* untuk memahami pola sebaran mahasiswa berdasarkan asal sekolah dan wilayah geografisnya. Setelah proses *clustering* selesai, setiap data diberi label cluster dan disimpan dalam format .csv agar mudah diolah dan divisualisasikan kembali untuk analisis spasial.

Hasil pengelompokan dari tahun akademik 2021–2024 menunjukkan pola persebaran yang relatif konsisten. Sebaran data terbagi menjadi 6 hingga 7 cluster, di mana setiap cluster memiliki karakteristik wilayah yang berbeda. Rangkuman hasil analisis per cluster disajikan pada Tabel 2.

Tabel 2. Hasil Ringkasan Cluster Persebaran Asal Sekolah Mahasiswa

Tahun	Jumlah Cluster	Wilayah Dominan	Jumlah Sekolah Terbanyak	Karakteristik Cluster
2021	7	Kota Bandung & Kab. Bandung	SMKN 6 Bandung (113 siswa)	Konsentrasi tinggi di Bandung Raya, indikasi wilayah utama asal mahasiswa.
2022	7	Bandung Raya, Garut, Tasikmalaya	SMKN 3 Bandung (75 siswa)	Sebaran meluas ke Priangan Timur dengan peningkatan variasi asal sekolah.
2023	7	Bandung, Subang, Purwakarta	SMKN 4 Bandung (73 siswa)	Terjadi penyebaran lebih merata ke wilayah sekitar Bandung.
2024	6	Bandung & Kab. Bandung Barat	SMAN 1 Soreang (67 siswa)	Cluster lebih sedikit namun lebih padat, menandakan fokus asal pendaftar.

4.7. Hasil Analisis Data Mahasiswa Keseluruhan

Setelah proses *clustering* selesai dilakukan untuk setiap tahun akademik, seluruh hasil pengelompokan dari tahun 2021 hingga 2024 digabungkan dan dianalisis secara komprehensif. Tujuan tahap ini adalah untuk mengidentifikasi pola umum persebaran asal sekolah mahasiswa Universitas Langlangbuana serta menentukan wilayah strategis yang berpotensi menjadi fokus kegiatan promosi.

Data hasil pengelompokan disimpan dalam format .csv yang berisi informasi asal sekolah, koordinat geografis (latitude dan longitude), serta label cluster hasil *K-Means*. Penggabungan data dari keempat tahun tersebut menghasilkan total 5.119 entri mahasiswa yang telah terkelompok ke dalam 6–7 cluster utama.

Hasil analisis menunjukkan bahwa sebaran mahasiswa Universitas Langlangbuana didominasi oleh asal sekolah yang berlokasi di wilayah Bandung Raya, mencakup Kota Bandung, Kabupaten

Bandung, dan Kabupaten Bandung Barat. Wilayah ini memiliki kepadatan data tertinggi pada semua tahun akademik, yang berarti menjadi wilayah inti (*core cluster*) bagi penerimaan mahasiswa baru.

Sementara itu, daerah seperti Garut, Tasikmalaya, dan Ciamis membentuk cluster menengah yang menunjukkan potensi pertumbuhan, meskipun dengan jumlah mahasiswa yang lebih sedikit dibandingkan wilayah inti. Adapun wilayah seperti Subang, Purwakarta, dan Cianjur menunjukkan persebaran dengan kepadatan rendah, namun tetap penting untuk dikembangkan melalui strategi promosi berbasis digital atau kerja sama dengan sekolah mitra.

Untuk mengetahui kontribusi setiap sekolah terhadap jumlah mahasiswa baru Universitas Langlangbuana, dilakukan analisis frekuensi berdasarkan asal sekolah yang dikelompokkan dari hasil *clustering*. Hasil analisis menunjukkan lima sekolah dengan jumlah mahasiswa terbanyak selama periode 2021–2024 sebagaimana ditunjukkan pada Tabel 3 berikut.

Tabel 3. Sekolah Asal dengan Jumlah Pendaftar Terbanyak (2021–2024)

Peringkat	Nama Sekolah Asal	Jumlah Mahasiswa	Lokasi	Keterangan
1	SMKN 6 Bandung	113	Kota Bandung	Penyumbang mahasiswa terbesar, masuk wilayah inti (Cluster 1).
2	SMKN 3 Bandung	75	Kota Bandung	Wilayah inti dengan kontribusi stabil setiap tahun.
3	SMKN 4 Bandung	73	Kota Bandung	Konsisten menyumbang mahasiswa baru sejak 2021.
4	SMAN 1 Soreang	67	Kab. Bandung	Wilayah pinggiran Bandung dengan potensi tinggi.
5	SMAN 1 Banjaran	63	Kab. Bandung	Termasuk wilayah penyangga dengan tren pertumbuhan positif

Dari hasil rekapitulasi tersebut dapat dilihat bahwa mayoritas mahasiswa baru berasal dari sekolah menengah kejuruan (SMK), terutama yang berlokasi di Kota Bandung. Hal ini menunjukkan bahwa SMK menjadi segmen potensial utama dalam rekrutmen mahasiswa Universitas Langlangbuana.

Selain itu, pola persebaran geografis yang ditampilkan melalui peta *clustering* memperlihatkan konsistensi antara tahun satu dengan tahun lainnya. Wilayah dengan tingkat kepadatan tinggi tetap berada di area Bandung Raya, sementara sebaran di luar provinsi menunjukkan peningkatan kecil dari tahun ke tahun.

Temuan ini memberikan dasar kuat bagi pihak universitas untuk menyusun strategi promosi berbasis wilayah (*geo-targeted promotion*). Misalnya, dengan memprioritaskan kegiatan sosialisasi dan kunjungan sekolah di wilayah yang berada dalam cluster dominan, serta memperluas promosi digital ke wilayah dengan potensi pertumbuhan (seperti Garut dan Tasikmalaya).

Dengan demikian, hasil analisis keseluruhan menunjukkan bahwa penerapan algoritma *K-Means Clustering* tidak hanya mampu mengelompokkan data secara objektif, tetapi juga menghasilkan peta spasial yang dapat dijadikan acuan praktis dalam meningkatkan efektivitas promosi penerimaan mahasiswa baru Universitas Langlangbuana.

5. KESIMPULAN

Berdasarkan analisis yang telah dilakukan, dapat disimpulkan bahwa penelitian ini diharapkan mampu menjadi alat bantu strategis dalam pengambilan keputusan terkait penentuan lokasi promosi bagi Universitas Langlangbuana, dengan tujuan utama meningkatkan jumlah calon mahasiswa baru di tahun-tahun mendatang. Dari hasil analisis, teridentifikasi bahwa Universitas Langlangbuana perlu memperkuat upaya promosi di beberapa wilayah kunci. Daerah prioritas utama adalah Jawa Barat, khususnya Kota Bandung, Kabupaten Bandung, serta Jabodetabek dan sekitarnya, mengingat wilayah-wilayah ini merupakan kontributor mahasiswa terbanyak selama periode 2021–2024. Selain itu, Pulau Sumatra, mencakup Medan, Padang, dan sekitarnya, juga merupakan area penting yang perlu menjadi fokus promosi selanjutnya.

Kontribusi penelitian ini terletak pada penerapan analisis *geospasial* berbasis *clustering* yang diintegrasikan dengan data dari Universitas Langlangbuana untuk mendukung strategi promosi kampus. Berbeda dari penelitian sebelumnya yang hanya menampilkan sebaran wilayah, penelitian ini menyertakan evaluasi kuantitatif menggunakan *Elbow Method* dan *Silhouette Score* untuk mengukur kualitas cluster yang dihasilkan. Pendekatan ini dapat diterapkan oleh institusi pendidikan lain sebagai model analitik berbasis data untuk perencanaan promosi dan rekrutmen mahasiswa.

DAFTAR PUSTAKA

- [1] Utami, L. L., Nurfadhilah, D. S., & Zulkifli, R. (2025). RANCANG BANGUN SISTEM INFORMASI MANAJEMEN RELASI MITRA PERUSAHAAN PADA PT SARANA KREASINDO TEKNOLOGI. *Journal of Information System Management (JOISM)*, 6(2), 171–178.
- [2] Budiman, R. (2019). Penerapan *Data mining* Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode *K–Means Clustering*). *ProTekInfo: Jurnal Pengembangan Riset dan Observasi Teknik Informatika*, 6, 2019.
- [3] Eska, J. (2018). Penerapan *data mining* untuk prediksi penjualan wallpaper menggunakan algoritma C4.5. *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, 2.
- [4] Metisen, B. M., & Sari, H. L. (2015). Analisis *clustering* menggunakan metode *K–Means* dalam pengelompokkan penjualan produk pada Swalayan Fadhila. *Jurnal media infotama*, 11(2).
- [5] Utari, W. (2021). ANALISIS IMPLEMENTASI KOMPETENSI PENGUJI KENDARAAN DI UNIT PELAKSANA UJI BERKALA KENDARAAN BERMOTOR DINAS PERHUBUNGAN KABUPATEN PROBOLINGGO. *MAP (Jurnal Manajemen dan Administrasi Publik)*, 4(1), 120–130.
- [6] Shirkhorshidi, A. S., Aghabozorgi, S., Wah, T. Y., & Herawan, T. (2014, June). Big data *clustering*: a review. In *International conference on computational science and its applications* (pp. 707–720). Cham: Springer International Publishing.
- [7] Yolandari, N. A., Butarbutar, L. E., Rajagukguk, G. C. H., Zulfi, M. F., & Ramadhani, F. (2025). ANALISIS PERBANDINGAN *K–MEANS* DAN DBSCAN DALAM PENGELOMPOKAN DATA TRAVEL REVIEW RATINGS MENGGUNAKAN EVALUASI SILHOUETTE INDEX DAN DAVIES–BOULDIN INDEX. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(3).
- [8] I. Pii, N. Suarna, and N. Rahaningsih, “Penerapan *Data mining* Pada Penjualan Produk Pakaian Dameyra Fashion Menggunakan Metode *K–Means Clustering*,” *JATI (Jurnal Mhs. Tek. Inform., vol. 7, no. 1, pp. 423–430, 2023.*
- [9] Khormarudin, A. N. (2016). Teknik *data mining*: Algoritma *K–Means clustering*. *J. Ilmu Komput*, 1–12.
- [10] Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). *K–Means clustering* algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210.
- [11] Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma *K–Means Clustering* dalam Penentuan Siswa Kelas Unggulan. *Jurnal Tekno Kompak*, 15(2), 25.
- [12] Metisen, B. M., & Sari, H. L. (2015). Analisis *clustering* menggunakan metode *K Means* dalam pengelompokkan penjualan produk pada Swalayan Fadhila. *Jurnal media infotama*, 11(2).
- [13] Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018, April). Integration *K–Means clustering* method and *Elbow Method* for identification of the best customer profile *cluster*. In IOP Conference Series: Materials Science and Engineering (Vol. 336, No. 1, p. 012017). IOP Publishing.
- [14] Muningsih, E., & Yogyakarta, A. B. S. I. (2017). Optimasi jumlah *cluster K–Means* dengan metode *elbow* untuk pemetaan pelanggan. In Pros. Semin. Nas. ELINVO (pp. 105–114).
- [15] Kusuma, D. T., & Agani, N. (2015). Prototipe Komparasi Model *Clustering* Menggunakan Metode *K–Means* Dan FCM untuk Menentukan Strategi Promosi: Study Kasus Sekolah Tinggi Teknik–PLN Jakarta. *Jurnal TICom*, 3(3), 93460.