

ANALISIS STRATEGI PENERIMAAN PESERTA DIDIK BARU MENGUNAKAN K-MEANS *CLUSTERING* DENGAN OPTIMASI ELBOW

Abdul Hadi¹, Decky Ryansyah², Goenawan Brotosaputro³

Program Studi Teknologi Sistem Informasi, Fakultas Teknologi Informasi
Universitas Budi Luhur^{1,2,3}

Jalan Ciledug Raya RT.10/RW.02, Kelurahan Pertukangan Utara, Pesanggrahan, Jakarta Selatan
2111600173@student.budiluhur.ac.id¹, 2411600212@student.budiluhur.ac.id²,
goenawan.brotosaputro@budiluhur.ac.id³

Abstrak

Penerimaan Peserta Didik Baru (PPDB) merupakan kegiatan rutin tahunan di sekolah, termasuk SMP Citra Negara Depok. Setiap tahun sekolah menerima data pendaftaran dalam jumlah besar, namun belum dimanfaatkan secara optimal untuk menentukan strategi PPDB pada tahun berikutnya sehingga berdampak pada penurunan jumlah pendaftar. Penelitian ini bertujuan mengelompokkan strategi PPDB yang efektif menggunakan metode K-Means *Clustering* yang dioptimalkan dengan Elbow Method dan normalisasi Min-Max. Dataset yang digunakan berasal dari data pendaftar tahun pelajaran 2021–2022 dan 2022–2023 dengan atribut nama sekolah, jumlah rombel, jumlah pendaftar, dan selisih. Hasil Elbow Method menunjukkan jumlah cluster optimal sebanyak tiga. Proses K-Means berhenti pada iterasi ke-7 dari maksimal 10 iterasi. Hasil pengelompokan menghasilkan tiga strategi, yaitu diskon 55% pada periode Oktober–Desember, strategi penyebaran brosur, dan diskon 50% pada periode Januari–Juni. Hasil penelitian menunjukkan bahwa metode K-Means yang dioptimalkan dengan Elbow Method mampu membantu menentukan strategi PPDB yang lebih efektif bagi sekolah.

Kata Kunci : *Clustering*, Metode Min-Max, Optimasi Metode Elbow, Data Mining, K-Means *Clustering*, Strategi PPDB

Abstract

New Student Admission is an annual routine activity at schools, including SMP Citra Negara Depok. Every year, the school receives a large amount of registration data, but it has not been optimally utilized to determine the new student admission strategy for the following year, resulting in a decrease in the number of applicants. This study aims to cluster effective new student admission strategies using the K-Means clustering method optimized with the Elbow method and min-max normalization. The dataset used is derived from the registration data for the 2021–2022 and 2022–2023 academic years, with attributes including school name, number of classes, number of registrants, and difference. The Elbow Method results indicate that the optimal number of clusters is three. The K-Means process stops at the 7th iteration out of a maximum of 10 iterations. The clustering results in three strategies: a 55% discount during the October–December period, a brochure distribution strategy, and a 50% discount during the January–June period. The research findings indicate that the K-Means method optimized with the Elbow Method can help determine a more effective new student admission strategy for the school.

Keywords : *Clustering, Min-Max Method, Elbow Method Optimization, Data Mining, K-Means Clustering, New Student Enrollment Strategy*

PENDAHULUAN

Penerimaan Peserta Didik Baru (PPDB) merupakan proses seleksi dan penerimaan calon peserta didik sebagai input bagi lembaga pendidikan untuk menjamin pemerataan akses dan kualitas pendidikan. Pelaksanaan PPDB di Indonesia mengacu pada Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 1 Tahun 2021 (Tunggawijaya et al., 2024). Setiap tahun jumlah pendaftar di SMP Citra Negara Depok mengalami perubahan karena persaingan dengan sekolah swasta maupun sekolah negeri di wilayah Depok. Dalam beberapa tahun terakhir, sekolah ini juga mengalami penurunan jumlah pendaftar. Berbagai strategi promosi telah dilakukan, seperti penyebaran brosur, pemberian potongan biaya pendidikan, dan kunjungan ke sekolah dasar. Namun, strategi tersebut belum didasarkan pada analisis data sehingga efektivitasnya belum dapat diketahui secara optimal. Untuk mengatasi permasalahan tersebut, penelitian ini memanfaatkan

pendekatan data mining, yaitu proses menemukan pengetahuan dari basis data melalui analisis data untuk menghasilkan informasi yang bermanfaat, seperti kategorisasi, asosiasi, pengelompokan, estimasi, dan prediksi Agneresa dkk (Agneresa et al., 2022). Salah satu teknik dalam data mining adalah *clustering*, yaitu proses pengelompokan data berdasarkan tingkat kemiripan karakteristik sehingga objek dalam satu kelompok memiliki kesamaan yang lebih tinggi dibandingkan dengan objek pada kelompok lainnya (Alhapizi et al., 2020). Salah satu metode *clustering* yang banyak digunakan adalah K-Means *Clustering*, yaitu metode pengelompokan non-hierarki yang membagi data ke dalam beberapa kelompok berdasarkan kesamaan karakteristik objek (Trisya Amanda et al., 2025). Beberapa penelitian sebelumnya telah menggunakan metode K-Means untuk mengelompokkan data. Namun, penentuan jumlah cluster pada penelitian tersebut sering dilakukan secara langsung tanpa metode evaluasi yang jelas, sehingga hasil pengelompokan belum tentu optimal. Oleh karena itu, penelitian ini menggunakan Elbow Method untuk menentukan jumlah cluster yang paling optimal berdasarkan nilai *Sum of Square Error* (SSE) (Orisa, 2022). Selain itu, penelitian ini juga menerapkan normalisasi Min-Max pada tahap pra-pemrosesan data karena algoritma berbasis jarak seperti K-Means sensitif terhadap perbedaan skala data. Normalisasi ini bertujuan menyamakan rentang nilai antar atribut sehingga perhitungan jarak menjadi lebih akurat. Seluruh proses penelitian dilakukan menggunakan metodologi CRISP-DM (Cross-Industry Standard Process for Data Mining) yang terdiri dari enam tahapan, yaitu pemahaman bisnis, pemahaman data, pengolahan data, pemodelan, evaluasi, dan penyebaran (Ordila et al., 2020) Melalui pendekatan tersebut, penelitian ini bertujuan membantu tim PPDB SMP Citra Negara Depok dalam menentukan strategi promosi yang lebih efektif dan berbasis data..

PENELITIAN RELEVAN

Beberapa penelitian sebelumnya telah memanfaatkan teknik data mining, khususnya metode *clustering*, untuk mendukung strategi promosi atau penerimaan mahasiswa baru. Penelitian yang dilakukan oleh Andrea (Tri et al., 2019), menggunakan Hierarchical *Clustering* metode Agglomerative untuk mengelompokkan data mahasiswa berdasarkan asal daerah, asal sekolah, dan penghasilan orang tua. Hasil penelitian menunjukkan bahwa strategi promosi yang paling efektif adalah melalui penyebaran brosur serta rekomendasi dari teman atau kerabat. Keunggulan penelitian ini adalah penggunaan metode hierarchial untuk melihat hubungan antar data, namun penelitian ini belum membandingkan hasilnya dengan metode *clustering* lain.

Penelitian kedua berjudul “Penerapan Data Mining Menggunakan Algoritma K-Means *Clustering* Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang” yang dilakukan oleh (Herlina et al., 2024). yang menerapkan algoritma K-Means *Clustering* untuk menentukan strategi promosi mahasiswa baru di Universitas Bina Darma Palembang. Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah empat cluster, yang kemudian dikelompokkan berdasarkan kemampuan ekonomi calon mahasiswa. Penelitian ini mampu memberikan segmentasi yang jelas, namun belum menggunakan teknik normalisasi data maupun metode evaluasi cluster secara khusus.

Penelitian ke tiga Teknik *clustering* dengan Algoritma K-means untuk menangani strategi promosi di Politeknik TEDC Bandung. Penelitian ini dilakukan oleh (Farahdinna et al., 2019) uga menggunakan K-Means *Clustering* untuk menentukan strategi promosi di Politeknik TEDC Bandung. Hasil penelitian menunjukkan bahwa empat cluster dapat mewakili karakteristik calon mahasiswa. Penelitian ini memberikan gambaran segmentasi calon mahasiswa, namun analisis yang dilakukan masih terbatas pada pengelompokan tanpa optimasi jumlah cluster.

Penelitian ke 4 *Clustering* tingkat promosi kampus dengan menggunakan menggunakan algoritma k-means, penelitian ini dilakukan oleh Ramadhani et al (2021), menggunakan algoritma K-Means untuk mengelompokkan tingkat efektivitas promosi kampus. Hasil penelitian menunjukkan bahwa promosi lebih efektif dilakukan pada sekolah yang memiliki jumlah pendaftar tinggi. Penelitian ini memberikan insight lokasi promosi, tetapi belum menggunakan metode evaluasi untuk menentukan jumlah cluster yang optimal.

Penelitian ke 5 Menentukan strategi promosi menggunakan algoritma *clustering* k-means (Syahputra et al., 2020). Penerapan K-Means *Clustering* untuk menentukan strategi promosi di SMA dan SMK Harapan Bangsa. Hasil penelitian menunjukkan adanya dua strategi promosi utama

yang dapat diterapkan oleh tim promosi. Namun, penelitian ini masih terbatas pada pengelompokan strategi tanpa analisis optimasi data.

Penelitian ke 6 dilakukan oleh Farozi (2021) yaitu menggunakan *K-Means Clustering* untuk menganalisis data pendaftar mahasiswa baru di STIE Serelo Lahat berdasarkan asal daerah, jenis sekolah, gelombang pendaftaran, dan program studi. Hasil penelitian menghasilkan tiga cluster yang dapat digunakan sebagai dasar dalam menentukan strategi promosi yang lebih terarah. Penelitian ini memberikan pemetaan karakteristik calon mahasiswa, tetapi belum menggunakan teknik normalisasi data

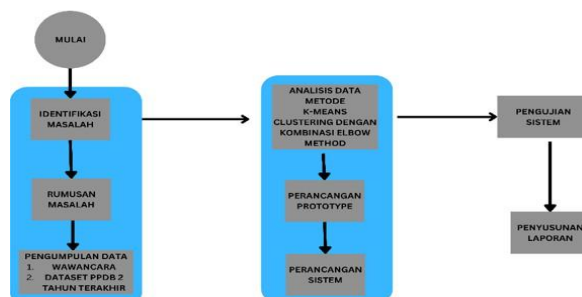
Penelitian ke 7 dilakukan oleh Silaban et al (2024), yaitu memanfaatkan *K-Means Clustering* untuk mengelompokkan calon mahasiswa STMIK Kaputama Binjai berdasarkan daerah asal, asal sekolah, dan jurusan yang diminati. Hasil *clustering* menghasilkan tiga cluster yang dapat digunakan untuk menentukan lokasi promosi yang lebih strategis. Penelitian ini mampu memberikan gambaran potensi wilayah promosi, namun belum mengoptimalkan proses *clustering* dengan metode evaluasi cluster.

Berdasarkan penelitian-penelitian tersebut dapat disimpulkan bahwa sebagian besar penelitian hanya menggunakan algoritma *K-Means Clustering* tanpa menerapkan normalisasi data maupun metode evaluasi untuk menentukan jumlah cluster yang optimal, serta sebagian besar objek penelitian masih berfokus pada tingkat perguruan tinggi. Oleh karena itu, penelitian ini mengisi celah tersebut dengan menerapkan *K-Means Clustering* yang dioptimalkan menggunakan *Elbow Method* serta normalisasi *Min-Max* pada studi kasus PPDB di tingkat Sekolah Menengah Pertama (SMP). Berdasarkan penelitian terdahulu, hipotesis dalam penelitian ini adalah diduga penerapan *K-Means Clustering* yang dioptimalkan dengan *Elbow Method* dan normalisasi *Min-Max* dapat membantu mengidentifikasi pola data pendaftar sehingga menghasilkan strategi penerimaan peserta didik baru yang lebih efektif.

METODE PENELITIAN

Metode *K-Means Clustering* digunakan dalam penelitian ini untuk melakukan proses pengelompokan (*clustering*) data sehingga dapat diketahui pola sebaran asal pendaftar peserta didik baru. Data penelitian diperoleh dari panitia PPDB SMP Citra Negara berdasarkan data pendaftar selama dua tahun terakhir. Dataset yang digunakan berjumlah 720 data dari tahun ajaran 2020/2021 hingga 2023/2024 dengan atribut meliputi nama sekolah, jumlah rombel akhir, jumlah pendaftar tahun 2021/2022, jumlah pendaftar tahun 2022/2023, serta selisih jumlah pendaftar. Sampel penelitian menggunakan teknik total sampling, yaitu seluruh populasi data pendaftar yang tersedia. Sebelum proses *clustering* dilakukan, data terlebih dahulu melalui tahap preprocessing dengan melakukan normalisasi menggunakan metode *Min-Max Normalization* untuk menyamakan rentang nilai antar atribut sehingga perhitungan jarak pada algoritma *K-Means* menjadi lebih akurat (Pamungkas et al., 2023). Selanjutnya, *Elbow Method* digunakan untuk menentukan jumlah cluster yang optimal berdasarkan nilai *Sum of Square Error* (SSE). Setelah jumlah cluster optimal diperoleh, proses *clustering* dilakukan menggunakan algoritma *K-Means* dengan menghitung jarak data terhadap centroid menggunakan *Euclidean Distance*. Proses iterasi dilakukan hingga centroid stabil atau tidak mengalami perubahan. Hasil *clustering* kemudian dianalisis untuk mengidentifikasi pola asal sekolah pendaftar serta kelompok sekolah yang memiliki potensi tinggi dalam menyumbang jumlah pendaftar. Hasil tersebut diharapkan dapat menjadi dasar dalam menentukan strategi penerimaan peserta didik baru yang lebih efektif di SMP Citra Negara.

Tahap-tahap penelitian yang dilakukan oleh peneliti dapat dilihat pada gambar dibawah ini :



Gambar 1. Langkah - langkah penelitian

HASIL DAN PEMBAHASAN

Persiapan Data

Data yang digunakan dalam penelitian ini merupakan data pendaftar peserta didik baru (PPDB) SMP Citra Negara yang diperoleh dari panitia PPDB. Data tersebut berasal dari dua tahun ajaran terakhir, yaitu tahun ajaran 2021/2022 dan 2022/2023. Dataset yang digunakan berjumlah 720 record dengan lima atribut utama, yaitu nama sekolah, jumlah rombel akhir, jumlah pendaftar tahun 2021/2022, jumlah pendaftar tahun 2022/2023, serta selisih jumlah pendaftar. Data tersebut berasal dari beberapa sekolah dasar yang berada di tiga kecamatan di Kota Depok dan satu kecamatan di Jakarta Selatan. Dataset ini digunakan sebagai dasar dalam proses analisis *clustering* untuk mengetahui pola sebaran asal sekolah pendaftar serta mengidentifikasi kelompok sekolah yang memiliki potensi tinggi dalam menyumbang jumlah pendaftar ke SMP Citra Negara.

Contoh dataset yang digunakan dalam penelitian ini dapat dilihat pada Tabel 2 berikut.:

Tabel 1. Dataset

Nama_sekolah	Jumlah_rom bel_akhir	2021_ 2022	2022_ 2023	selisih
SDN Tanah Baru 3	3	13	21	8
SDN Tanah Baru 2	4	21	18	-3
SDN Tanah Baru 1	2	9	14	5
SDN Mampang 3	3	1	9	8
SDN Pancoran Mas 3	4	2	8	6
MIT Nurul Iman 1	1	0	7	7
SDN Beji Timur 1	6	0	5	5
SDN Mampang 1	4	4	5	1
SDN Depok Jaya 1	4	9	4	-5
SDN Pancoran Mas 3	4	2	8	6

Metode Normalisasi Min-Max

Metode normalisasi dilakukan supaya nilai menjadi terskala pada jaungkauan kecil dan spesifik seperti 0,0 ke 0,1. Pada umum nya dapat dilakukan dengan beberapa cara, untuk penelitian ini menggunakan metode normalisasi min- max. Metode ini menskala data dari nilai d menjadi d' dalam jangkauan data minimum (p) hingga maksimum (p) sebagaimana persamaan 1

$$d' = \frac{d - \min(p)}{\max(p) - \min(p)} \quad (1)$$

Dimana:

d' : data ternormalisasi

d : data awal

Min (p) : data minimum pada suatu atribut Max (p) : data maksimum pada suatu atribut

Dengan hasil normalisasi yang dilakukan oleh sistem dataset data data numerik tersebut di konversikan menjadi nilai dengan skala 0-1 sesuai kriteria nya masing-masing fungsi ny adalah mempermudah untuk penggunaan optimasi pada metode elbow, hasil dari normalisasi ini berbentuk sebuah data – data yang sesuai dengan tahun pelajaran calon pendaftar sesuai atribut yang dipilih (Palinggik Allorerung et al., 2024).

Tabel 2. Dataset Hasil Normalisasi Metode Min-Max

Nama_sekolah	Jumlah_rombel_akhir	2021-2022	2022-2023	Selisih
SDN Tanah Baru 3	0.4	0.61904761904762	1	1
SDN Tanah Baru 2	0.6	1	0.85714285714286	0.38888888888889
SDN Tanah Baru 1	0.2	0.42857142857143	0.66666666666667	0.83333333333333
SDN Mampang 3	0.4	0.047619047619048	0.42857142857143	1
SDNPancoran Mas 3	0.6	0.095238095238095	0.38095238095238	0.88888888888889
MIT Nurul Iman 1	0	0	0.33333333333333	0.94444444444444
SDN Beji Timur 1	1	0	0.23809523809524	0.83333333333333
SDN Mampang 1	0.6	0.19047619047619	0.23809523809524	0.61111111111111
SDN Depok Jaya 1	0.6	0.42857142857143	0.19047619047619	0.27777777777778

Tabel hasil normalisasi menggunakan metode min-max ini menghasilkan keluaran data dari skala 0-1, dimana hasil data ini akan diteruskan ke tahap selanjutnya untuk membuat pengelompokan data di optimasi elbow.

Optimasi Elbow

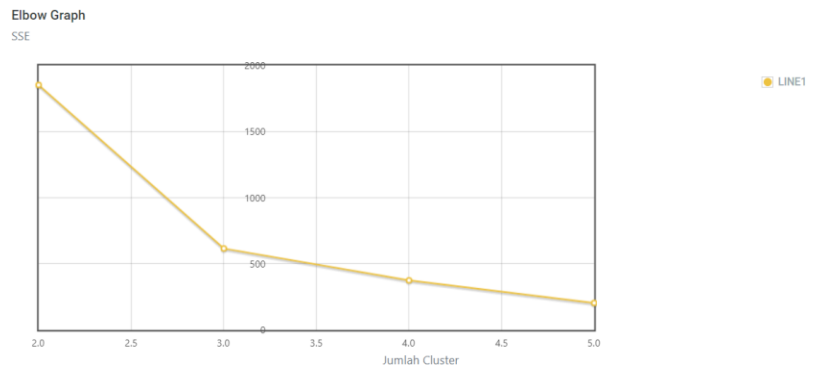
Hasil data yang telah dinormalisasi kemudian dilanjutkan dengan proses optimasi menggunakan **Elbow Method**. Metode ini digunakan untuk menentukan jumlah cluster yang paling optimal dalam proses *clustering*. Penentuan jumlah cluster dilakukan dengan menghitung nilai **Sum of Square Error (SSE)** pada beberapa nilai cluster (*k*). Perhitungan dilakukan secara berulang untuk melihat penurunan nilai SSE pada setiap penambahan jumlah cluster (Maori & Evanita, 2023) Untuk mengetahui kualitas hasil pengelompokan, dilakukan evaluasi cluster menggunakan pendekatan *Sum of Square Error (SSE)* melalui Elbow Method. Nilai SSE menunjukkan tingkat kedekatan data terhadap centroid pada setiap cluster. Semakin kecil nilai SSE, maka kualitas pengelompokan yang dihasilkan semakin baik. Berdasarkan hasil evaluasi tersebut diperoleh jumlah cluster optimal sebanyak tiga cluster yang kemudian digunakan dalam proses K-Means *Clustering*.

Tabel 3. Hasil Algoritma Elbow

Cluster (k)	Nilai (SSE)	Selisih Penurunan SSE	Cluster (k)
2	1852.5913043478	1852.5913043478	2
3	1236.5354895105	616.05581483734	3
4	861.45663650075	375.07885300974	4
5	657.44110275689	204.01553374386	5

SSE Graphic Elbow Method

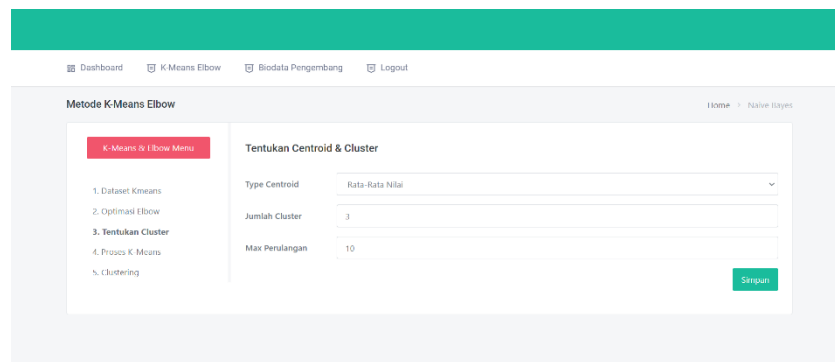
SSE (*Sum of Squared Errors*) atau disebut juga sebagai within-cluster sum of squares (WCSS) merupakan hasil visualisasi dari perhitungan metode Elbow. Grafik yang dihasilkan menunjukkan hubungan antara jumlah cluster dengan nilai SSE. Grafik tersebut digunakan untuk melihat titik penurunan yang paling signifikan (titik siku/elbow) yang menunjukkan jumlah cluster paling optimal. Berdasarkan hasil grafik Elbow yang diperoleh dari dataset yang telah dinormalisasi menggunakan metode Min-Max, terlihat bahwa penurunan nilai SSE yang paling signifikan terjadi pada *k* = 3. Oleh karena itu, jumlah cluster yang digunakan dalam penelitian ini adalah tiga cluster (Sa'diah et al., 2023)



Gambar 2. Metode Elbow

Berdasarkan grafik Elbow yang diperoleh, penurunan nilai SSE yang paling signifikan terjadi pada $k = 3$ sehingga jumlah cluster yang digunakan dalam penelitian ini adalah tiga cluster

Tentukan Centeroid



Gambar 3. Penentuan Centroid

Setelah mengetahui jumlah cluster yang optimal adalah $k = 3$ maka selanjutnya adalah dengan cara menentukan centeroid untuk melanjutkan pada algoritma k-means dengan rumus rata-rata nilai untuk membuat dan meningkatkan kualitas klaster yang dihasilkan oleh algoritma k-means (Arief Soeleman & Ilmu Komputer, 2021) Dengan menggunakan Rata-rata nilai dalam penentuan centeroid agar mendapatkan klaster yang terbaik dengan meminimalkan jarak antara titik data dalam klaster dan centeroid nya. Maka jumlah cluster yang dipilih adalah 3 dengan melakukan perhitungan perulangan sebanyak 10 kali. Dengan perhitungan rata-rata nilai peneliti dapat memastikan bahwa centeroid yang dipilih cenderung lebih baik dalam pesbaran nya. Penentuan centeroid juga berfungsi untuk mempermudah dalam hal proses klasterisasi dengan K-means, jadi setiap elmen di protipe ini berfungsi secara structural.

Proses K-means

Dengan menggunakan metode perhitungan jarak yaitu Euclidean distance, untuk mengukur seberapa dekat atau seberapa cocok suatu titik data dengan centeroid terbaru, dilakukan perulangan sampai dengan hasil klaster dan perulangan bisa menemui persamaan antara nilai centeroid dan perulangan menggunakan metode Euclidean distance disini dari 3 centeroid dan melakukan maksimal 10 kali perulangan oleh sistem terdapat hasil 7 kali perulangan. Rumus perhitungan Euclidean distance adalah sebagai berikut (Pribadi et al., 2022):

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^n (x_j - c_j)^2} \quad (2)$$

- d = jarak
- j = banyaknya data
- c = centeroid
- x = data

Tabel 4. Tabel Centroid

Centroid 1	3.2	9.2	14	4.8
Centroid 2	2.3636363636364	6.8181818181818	1.8181818181818	-5
Centroid 3	2.7211538461538	0.82692307692308	0.82692307692308	-0.0096153846153846

Dari hasil 3 kluster yang di input kepada system menghasilkan 3 centroid dengan nilai rata-rata seperti tabel diatas, perhitungan jarak dengan Euclidian distance berhasil melakukan sebanyak 7 kali perulangan, Perulangan ini dilakukan untuk membuat hasil akhir menjadi pengelompokan untuk bisa membuat peneliti mengetahui hasil akhir dari proses k-means *clustering* yang dioptimalkan oleh metode elbow

Perulangan Ke - 7

Perulangan 7 - Penentuan Centroid

Centroid 1	3.2	9.2	14	4.8
Centroid 2	2.3636363636364	6.8181818181818	1.8181818181818	-5
Centroid 3	2.7211538461538	0.82692307692308	0.81730769230769	-0.0096153846153846

Gambar 4. Perulangan Centeroid Perulangan 7

Hasil *Clustering K Means*

Dari hasil proses k-means dengan perhitungan jarak Euclidean distance mendapatkan hasil pengelompokan kluster yang optimal yaitu dengan 3 kluster diantaranya :

Kluster 1 dengan 5 data

Kluster 3 dengan 104 data

Kluster 2 dengan 11 data

Dengan total dari keseluruhan data ialah 120 data.

Dataset awal yang diperoleh dari panitia PPDB berjumlah 720 data pendaftar. Data tersebut kemudian direkap berdasarkan asal sekolah, sehingga diperoleh 120 data sekolah yang digunakan sebagai dataset pada proses *clustering*. Berdasarkan proses K-Means *Clustering* dengan nilai k = 3 diperoleh hasil pengelompokan yaitu Cluster 1 sebanyak 5 sekolah, Cluster 2 sebanyak 11 sekolah, dan Cluster 3 sebanyak 104 sekolah.

Jumlah Cluster

Cluster	Jumlah
1	5
3	104
2	11

[Export](#)

Gambar 5. Hasil Pengelompokan Kluster

Analisis Hasil *Clustering*

Hasil *clustering* menunjukkan bahwa Cluster 3 memiliki jumlah anggota paling banyak yaitu 104 sekolah. Hal ini menunjukkan bahwa sebagian besar sekolah memiliki jumlah pendaftar pada kategori sedang. Sementara itu, Cluster 1 dengan jumlah anggota sebanyak 5 sekolah menunjukkan kelompok sekolah dengan jumlah pendaftar yang relatif tinggi.

Implikasi terhadap Strategi PPDB

Berdasarkan hasil pengelompokan tersebut, pihak SMP Citra Negara dapat mengidentifikasi sekolah-sekolah yang memiliki potensi tinggi dalam menyumbang jumlah pendaftar. Sekolah yang berada pada cluster dengan jumlah pendaftar tinggi dapat dijadikan target utama dalam kegiatan sosialisasi PPDB. Sementara itu, sekolah pada cluster dengan jumlah pendaftar rendah dapat menjadi fokus dalam peningkatan promosi untuk menarik minat calon peserta didik baru..

SIMPULAN

Berdasarkan hasil penelitian, penerapan metode K-Means *Clustering* dengan normalisasi Min-Max dan optimasi Elbow Method mampu mengelompokkan data pendaftar PPDB SMP Citra Negara Depok secara efektif. Hasil optimasi menunjukkan jumlah klaster optimal sebanyak tiga klaster ($k = 3$) dengan distribusi 5 data pada Klaster 1, 11 data pada Klaster 2, dan 104 data pada Klaster 3. Hasil pengelompokan ini menunjukkan bahwa sebagian besar sekolah asal pendaftar berada pada Klaster 3 yang memiliki potensi pendaftar paling tinggi. Temuan ini memberikan gambaran pola asal sekolah pendaftar sehingga dapat digunakan sebagai dasar dalam menentukan strategi promosi dan penerimaan peserta didik baru yang lebih terarah dan berbasis data.

Penelitian ini masih memiliki beberapa keterbatasan, di antaranya jumlah data yang digunakan terbatas pada data pendaftar dari dua tahun terakhir serta wilayah asal sekolah yang hanya mencakup beberapa kecamatan di Kota Depok dan Jakarta Selatan. Selain itu, variabel yang digunakan dalam proses *clustering* masih terbatas pada atribut jumlah rombel dan jumlah pendaftar pada tahun tertentu. Oleh karena itu, penelitian selanjutnya disarankan menggunakan data dengan rentang waktu yang lebih panjang serta menambahkan variabel lain yang relevan agar hasil pengelompokan dapat memberikan analisis yang lebih komprehensif.

DAFTAR PUSTAKA

- Agneresa, Lia Hananto, A., Shofiah Hilabi, S., Hananto, A., & Tukino. (2022). *Strategi Promosi Penerapan Data Mining Mahasiswa Baru Dengan Metode K-Means Clustering*.
- Alhapizi, R., Nasir, M., & Effendy, I. (2020). Penerapan Data Mining Menggunakan Algoritma K-Means *Clustering* Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang. In *Journal of Software Engineering Ampera* (Vol. 1, Number 1). <https://journal-computing.org/index.php/journal-sea/index>
- Arief Soeleman, M., & Ilmu Komputer, F. (2021). Penentuan Centroid Awal Pada Algoritma K-Means Dengan Dynamic Artificial Chromosomes Genetic Algorithm Untuk Tuberculosis Dataset Pre-Centroid Determination in K-Means Algorithm using Dynamic Artificial Chromosomes Genetic Algorithm for Tuberculosis Dataset. In *Februari* (Vol. 20, Number 1).
- Farahdinna, F., Nurdiansyah, I., Suryani, A., & Wibowo, A. (2019). Perbandingan Algoritma K-Means Dan K-Medoids Dalam Klasterisasi Produk Asuransi Perusahaan Nasional. *Jurnal Ilmiah FIFO*, 11(2), 208. <https://doi.org/10.22441/fifo.2019.v11i2.010>
- Farozi, M. (2021). *Metode K-Means Clustering Dalam Merancang Strategi Promosi Penerimaan Mahasiswa Baru Pada STIE Serelo Lahat*.
- Herlina, H., Siti Mundari, Elsa Amanda Putri, & Amanda Dian Rahmawati. (2024). Penerapan K-Means *Clustering* Untuk Strategi Promosi Untuk Meningkatkan Jumlah Perolehan Mahasiswa Baru. *Journal Of Industrial And Manufacture Engineering*, 8(1), 98–103. <https://doi.org/10.31289/jime.v8i1.10928>.
- Maori, N. A., & Evanita. (2023). Metode Elbow Dalam Optimasi Jumlah Cluster Pada K-Means *Clustering*. *Jurnal SIMETRIS*, 14.
- Ordila, R., Wahyuni, R., Irawan, Y., & Yulia Sari, M. (2020). Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma *Clustering* (Studi Kasus : Poli Klinik Pt.Inecda). *Jurnal Ilmu Komputer*, 9(2), 148–153. <https://doi.org/10.33060/jik/2020/vol9.iss2.181>.
- Palinggik Allorerung, P., Erna, A., Bagussahrir, M., & Alam, S. (2024). Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit. In *Jurnal Informatika Sunan Kalijaga* (Vol. 9, Number 3).
- Pamungkas, T. B., Maesaroh, S., Ardiansyah, P., Komputer, F. I., Buana, U. M., Sains, F., & Teknologi, D. (2023). Implementasi Data Mining Pada Stok Penggunaan Barang Di Gmf Aeroasia Menggunakan Algoritma K-Means *Clustering*. In *Jurnal Ilmiah Sains dan Teknologi P* (Vol. 7, Number 2).
- Orisa, M. (2022). *Optimasi Cluster pada Algoritma K-Means*. 13, 2022.

- Sa'diah, H., Enri, U., & Nur Padilah, T. (2023). *Penerapan Algoritme K-Means Dalam Segmentasi Daerah Rawan Kekerasan Anak Di Jawa Barat*.
- Pribadi, W. W., Yunus, A., & Sartika Wiguna, A. (2022). Perbandingan Metode K-Means Euclidean Distance Dan Manhattan Distance Pada Penentuan Zonasi Covid-19 Di Kabupaten Malang. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 6, Number 2).
- Ramadhani, A., Fauziah, R., & Royal, S. (2021). *Clustering Tingkat Promosi Kampus Dengan Menggunakan Algoritma K-Means*. In *Journal of Science and Social Research* (Number 1). <http://jurnal.goretanpena.com/index.php/JSSR>.
- Silaban, R., Fauzi, A., Arliana, L., & Kadim, N. (2024). Penerapan K-Means *Clustering* untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru (Studi Kasus: STMIK Kaputama Kota Binjai). *Jurnal Riset Sistem Informasi Dan Teknik Informatika*, 2(5). <https://doi.org/10.61132/mercurius.v2i4.322>.
- Syahputra, S., Ramadani, S., Manaor, A., & Pardede, H. (2020). Menentukan Strategi Promosi Menggunakan Algoritma *Clustering* K-Means. *JOISIE Journal Of Information System And Informatics Engineering*, 4(1), 7–14.
- Tri, A., Dani, R., Wahyuningsih, S., & Rizki, N. A. (2019). Penerapan Hierarchical *Clustering* Metode Agglomerative pada Data Runtun Waktu. *Jambura Journal of Mathematics*, 1. <http://ejurnal.ung.ac.id/index.php/jjom>.
- Trisyia Amanda, S., Berlianindita, K., Jahid Alfarizi, A., Arifin, W. A., Kelautan, L., & Daerah Serang, K. (2025). *Analisis Komparatif Svm Dan K-Means Dalam Data Mining Untuk Promosi Perguruan Tinggi*.
- Tunggawijaya, W., Liestyasari, S. I., & Budiati, A. C. (2024). McDonaldisasi Sistem Zonasi: Studi Fenomenologi Implementasi PPDB SMP di Kabupaten Karanganyar. *Ideas: Jurnal Pendidikan, Sosial, Dan Budaya*, 10(1), 23. <https://doi.org/10.32884/ideas.v10i1.1604>.