



# Inventa: Journal of Science, Technology, and Innovation

Vol 1 No 3 April 2026, Hal 265-275  
ISSN: 3123-3147 (Print) ISSN: 3123-3155 (Electronic)  
Open Access: <https://scriptaintelektual.com/inventa>

## Analisis K-Means Clustering pada Data Penempatan Karyawan di PT. Guna Cipta Prima Berbasis Web

Sella<sup>1\*</sup>, Yohanni Syahra<sup>2</sup>

<sup>1-2</sup> Universitas Muhammadiyah Sumatera Utara, Indonesia  
email: [Sella140802@gmail.com](mailto:Sella140802@gmail.com)<sup>1</sup>

### Article Info :

Received:  
25-03-2026  
Revised:  
12-04-2026  
Accepted:  
21-04-2026

### Abstract

*The advancement of information technology has made a significant contribution to decision-making processes, particularly in human resource management. One of the common challenges faced by organizations is the difficulty in determining employee placement objectively based on their competencies and performance. This study aims to implement the K-Means Clustering algorithm to classify employee data at PT Guna Cipta Prima, develop a web-based system using PHP and MySQL, and analyze the resulting clusters. The data used are real-world data consisting of job attributes such as security and cleaning service. The clustering process is conducted using three categories: high, medium, and low performance. The results indicate that the developed system is capable of effectively implementing the K-Means algorithm and producing clusters that align with the characteristics of the data. A Silhouette Coefficient value of 0.67 demonstrates a reasonably good level of clustering validity. Therefore, this system can assist the Human Resources Department (HRD) in conducting employee placement analysis in a more efficient, objective, and data-driven manner, thereby supporting more accurate decision-making.*

**Keywords:** Clustering, Data Mining, Employees, PHP, Systems.

### Abstrak

Perkembangan teknologi informasi memberikan kontribusi signifikan dalam mendukung pengambilan keputusan, khususnya pada manajemen sumber daya manusia. Salah satu permasalahan yang sering dihadapi perusahaan adalah kesulitan dalam menentukan penempatan karyawan secara objektif sesuai kemampuan dan kinerja. Penelitian ini bertujuan menerapkan algoritma K-Means Clustering untuk mengelompokkan data karyawan di PT Guna Cipta Prima, merancang sistem berbasis web menggunakan PHP dan MySQL, serta menganalisis hasil pengelompokan tersebut. Data yang digunakan merupakan data nyata dengan atribut pekerjaan seperti security dan cleaning service. Proses clustering dilakukan dengan tiga kategori, yaitu kinerja tinggi, sedang, dan rendah. Hasil penelitian menunjukkan bahwa sistem yang dibangun mampu mengimplementasikan algoritma K-Means dengan baik dan menghasilkan pengelompokan yang sesuai karakteristik data. Nilai Silhouette Coefficient sebesar 0,67 menunjukkan kualitas clustering yang cukup baik. Dengan demikian, sistem ini dapat membantu HRD dalam melakukan analisis penempatan karyawan secara lebih efisien, objektif, dan berbasis data, sehingga mendukung pengambilan keputusan yang lebih akurat.

**Kata Kunci:** Clustering, Data Mining, Karyawan, PHP, Sistem.



©2022 Authors.. This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.  
(<https://creativecommons.org/licenses/by-nc/4.0/>)

## PENDAHULUAN

Perkembangan global dalam bidang analitik data dan kecerdasan bisnis menunjukkan pergeseran paradigma dari pengambilan keputusan berbasis intuisi menuju pendekatan berbasis data yang terukur dan dapat direplikasi, khususnya dalam domain manajemen sumber daya manusia yang semakin kompleks akibat ekspansi operasional multi-lokasi dan heterogenitas kebutuhan tenaga kerja. Transformasi digital yang ditopang oleh kemajuan teknologi basis data dan komputasi web telah memungkinkan organisasi untuk mengintegrasikan data operasional secara real-time dan mengolahnya menjadi informasi strategis yang bernilai tinggi, sebagaimana ditegaskan dalam literatur mengenai sistem basis data relasional dan implementasinya dalam pengelolaan informasi organisasi (Rahimi, 2020).

Pemanfaatan teknik data mining menjadi semakin krusial sebagai instrumen analitik untuk mengekstraksi pola tersembunyi dari data skala besar yang tidak dapat diidentifikasi melalui pendekatan konvensional, sehingga membuka peluang bagi optimalisasi proses bisnis secara sistematis dan berbasis

bukti (Setiawan, 2021). Penelitian terdahulu menunjukkan bahwa algoritma K-Means Clustering memiliki kapabilitas yang signifikan dalam mengelompokkan data berdasarkan kesamaan karakteristik, yang kemudian dapat digunakan sebagai dasar pengambilan keputusan strategis dalam berbagai sektor. Studi oleh Syahra et al. (2021) mengungkapkan bahwa penerapan K-Means pada data pelanggan mampu menghasilkan segmentasi yang operasional dan relevan secara bisnis, sementara Andriyani et al. (2023) menunjukkan bahwa pendekatan serupa efektif dalam mengidentifikasi pola kemiskinan regional yang kompleks melalui pengelompokan berbasis variabel kuantitatif.

Penelitian Liaza et al. (2020) memperlihatkan bahwa K-Means tidak hanya mampu mengelompokkan data secara matematis, tetapi juga menghasilkan struktur cluster yang tervalidasi secara empiris dan dapat diinterpretasikan secara kontekstual. Sintesis dari temuan-temuan tersebut mengindikasikan bahwa K-Means memiliki fleksibilitas metodologis yang tinggi dalam berbagai domain, termasuk potensi penerapannya dalam konteks pengelolaan tenaga kerja yang membutuhkan klasifikasi berbasis kebutuhan operasional. Meskipun demikian, kajian literatur juga memperlihatkan adanya keterbatasan yang cukup mendasar, terutama terkait dengan minimnya penelitian yang secara spesifik mengaplikasikan teknik clustering pada masalah penempatan karyawan dalam perusahaan jasa berbasis multi-lokasi, di mana dinamika kebutuhan tenaga kerja bersifat fluktuatif dan kontekstual. Sebagian besar penelitian terdahulu lebih berfokus pada segmentasi pelanggan atau analisis sosial-ekonomi, sehingga belum memberikan kerangka konseptual yang memadai untuk menjelaskan bagaimana clustering dapat diintegrasikan dalam sistem pengambilan keputusan sumber daya manusia secara praktis dan berkelanjutan (Putri et al., 2022).

Inkonsistensi dalam pendekatan evaluasi kualitas cluster, seperti variasi penggunaan metrik validasi tanpa justifikasi metodologis yang kuat, menunjukkan adanya celah epistemik yang perlu diisi melalui penelitian yang lebih sistematis dan terstruktur. Kondisi empiris di PT. Guna Cipta Prima merefleksikan urgensi permasalahan tersebut, di mana proses penempatan karyawan yang masih berbasis pengalaman subjektif berimplikasi pada ketidakseimbangan distribusi tenaga kerja antar lokasi, yang pada akhirnya berdampak pada efisiensi operasional dan kualitas layanan. Kebutuhan akan suatu sistem analitik yang tidak hanya mampu mengelompokkan data secara akurat, tetapi juga terintegrasi dalam platform yang mudah diakses dan digunakan oleh manajemen menjadi semakin mendesak. Pengembangan sistem berbasis web dengan memanfaatkan teknologi PHP dan MySQL menawarkan solusi yang adaptif dan scalable dalam mengelola serta memvisualisasikan data secara interaktif, sebagaimana diuraikan dalam kajian mengenai pengembangan aplikasi web dinamis (Sholichin, 2020).

Integrasi antara metode analitik dan sistem informasi ini membuka ruang bagi transformasi proses pengambilan keputusan yang lebih transparan dan berbasis data. Dalam lanskap keilmuan, penelitian ini menempati posisi sebagai upaya integratif yang menggabungkan pendekatan data mining dengan pengembangan sistem informasi berbasis web untuk menjawab permasalahan praktis dalam manajemen sumber daya manusia, khususnya terkait penempatan karyawan. Berbeda dengan studi sebelumnya yang cenderung terfragmentasi antara aspek metodologis dan implementatif, penelitian ini berupaya membangun jembatan antara analisis clustering dan aplikasi operasionalnya dalam lingkungan organisasi nyata, dengan mempertimbangkan aspek validasi model melalui metrik seperti SSE dan Silhouette Coefficient yang digunakan secara komparatif sebagaimana diadopsi dalam penelitian-penelitian terdahulu (Andriyani et al., 2023).

Pendekatan ini sekaligus memperluas cakupan penerapan K-Means ke dalam domain yang relatif kurang dieksplorasi, yaitu distribusi tenaga kerja berbasis kebutuhan lokasi. Penelitian ini bertujuan untuk menerapkan algoritma K-Means Clustering pada data penempatan karyawan guna menghasilkan pengelompokan lokasi berdasarkan kebutuhan tenaga kerja secara objektif, menentukan jumlah cluster optimal melalui pendekatan evaluatif, serta merancang dan mengimplementasikan sistem informasi berbasis web yang mampu mengintegrasikan seluruh proses analitik secara komprehensif. Kontribusi penelitian ini terletak pada pengembangan kerangka metodologis yang mengombinasikan teknik preprocessing, clustering, dan evaluasi model dalam satu alur sistematis, serta pada penyediaan solusi praktis berupa sistem yang dapat digunakan sebagai alat bantu pengambilan keputusan strategis dalam penempatan karyawan secara efisien, terukur, dan berbasis data.

## METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif terapan (applied quantitative research) yang menggunakan pendekatan komputasional berbasis data mining. Pendekatan ini dipilih karena penelitian bertujuan untuk menerapkan algoritma K-Means Clustering pada data numerik nyata guna menghasilkan pengelompokan (cluster) data penempatan karyawan di PT. Guna Cipta Prima yang bersifat objektif, terukur, dan dapat direplikasi. Tahapan penelitian dilakukan secara sistematis dan terstruktur, dimulai dari pengumpulan data riil dari perusahaan, dilanjutkan dengan preprocessing data, implementasi algoritma K-Means secara manual maupun terkomputerisasi, evaluasi hasil clustering, hingga perancangan dan pengembangan sistem berbasis web menggunakan PHP 8.1 dan MySQLi sebagai media penyajian hasil analisis kepada pengguna secara interaktif (Liaza et al., 2020).

Metode penelitian yang digunakan mencakup tiga pendekatan utama, yaitu: (1) pendekatan empiris melalui observasi dan wawancara langsung di PT. Guna Cipta Prima untuk memperoleh data primer; (2) pendekatan analitis melalui perhitungan manual dan komputasi algoritma K-Means menggunakan data aktual; serta (3) pendekatan rekayasa perangkat lunak (software engineering) dalam merancang dan membangun sistem informasi berbasis web. Jenis data yang digunakan dalam penelitian ini adalah data primer dan skunder, dimana Data Primer: Data yang diperoleh secara langsung dari PT. Guna Cipta Prima melalui kegiatan observasi dan wawancara dengan bagian admin perusahaan. Data primer meliputi informasi lokasi penempatan, jumlah tenaga Security, dan jumlah tenaga Cleaning Service (CS) pada setiap lokasi aktif. Data Sekunder: Data pendukung yang diperoleh melalui studi literatur, jurnal ilmiah, dan referensi akademik terkait algoritma K-Means Clustering, data mining, serta pengembangan sistem informasi berbasis web (Santoso, 2023).

## HASIL DAN PEMBAHASAN

### Perancangan Variabel, Preprocessing Data, dan Integrasi Sistem Berbasis Web dalam Analisis K-Means Clustering

Penelitian ini menggunakan dua variabel numerik utama yaitu jumlah tenaga Security (X1) dan jumlah tenaga Cleaning Service/CS (X2) per lokasi penempatan. Pemilihan kedua variabel ini didasarkan pada pertimbangan bahwa keduanya merupakan indikator langsung beban kerja dan kebutuhan tenaga di setiap lokasi. Sebelum diproses, data dinormalisasi menggunakan metode Min-Max Scaling untuk menghilangkan efek dominansi atribut dengan rentang nilai yang lebih besar. Penentuan jumlah cluster optimal dilakukan menggunakan Metode Elbow berdasarkan nilai SSE (Sum of Squared Error), yang menghasilkan keputusan  $k=3$  sebagai jumlah cluster terbaik. Hasil clustering kemudian dievaluasi menggunakan Silhouette Coefficient untuk mengukur kualitas pemisahan antar cluster (Rousseeuw, 1987).

Untuk menjadikan hasil analisis dapat diakses dan dimanfaatkan secara praktis oleh pihak manajemen PT. Guna Cipta Prima, penelitian ini juga mengembangkan sistem informasi berbasis web menggunakan PHP 8.1 dan MySQLi. Sistem tersebut mengintegrasikan seluruh tahapan analisis—mulai dari manajemen data, proses clustering, visualisasi hasil, hingga pemberian rekomendasi—dalam satu antarmuka yang interaktif dan responsif. Pengembangan sistem berbasis web dipilih karena kemudahan akses lintas perangkat dan tidak memerlukan instalasi perangkat lunak khusus di sisi pengguna (Sholichin, 2020; Santoso, 2023).

Penelitian ini bermaksud menjawab kebutuhan nyata PT. Guna Cipta Prima dalam melakukan analisis pola penempatan karyawan secara ilmiah dan berbasis data, sekaligus menghasilkan sistem yang dapat digunakan secara berkelanjutan oleh bagian administrasi perusahaan untuk mendukung pengambilan keputusan yang lebih efisien, objektif, dan terukur.

Sistem informasi berbasis web yang dikembangkan dalam penelitian ini berhasil diimplementasikan menggunakan PHP 8.1, ekstensi MySQLi, dan basis data MySQL yang dijalankan pada web server Apache melalui XAMPP. Sistem dirancang untuk menjalankan seluruh tahapan analisis K-Means Clustering secara terintegrasi, mulai dari manajemen data, preprocessing, eksekusi algoritma, visualisasi hasil, hingga penyajian laporan dan rekomendasi. Antarmuka sistem dibangun menggunakan kerangka kerja CSS Bootstrap 5 yang dipadukan dengan desain kustom bertema Navy-Gold, sehingga menghasilkan tampilan yang profesional, responsif, dan mudah digunakan.

Untuk mengevaluasi kinerja dan kualitas hasil clustering, penelitian ini menggunakan beberapa metrik evaluasi, yaitu *Silhouette Coefficient*, *Sum of Squared Errors* (SSE), dan validasi cluster. Metode evaluasi ini digunakan untuk menilai seberapa baik algoritma clustering dalam mengelompokkan data

serta sejauh mana cluster yang dihasilkan mampu merepresentasikan tingkat prestasi siswa secara bermakna. *Silhouette Coefficient* digunakan untuk mengukur seberapa baik setiap titik data berada dalam cluster-nya dibandingkan dengan cluster lain.

Metrik ini mempertimbangkan dua aspek utama, yaitu kedekatan data dalam satu cluster (*intra-cluster cohesion*) dan jarak antar cluster (*inter-cluster separation*). Nilai silhouette yang lebih tinggi menunjukkan bahwa suatu data cocok berada dalam cluster-nya sendiri dan memiliki perbedaan yang jelas dengan cluster tetangga. Nilai silhouette keseluruhan diperoleh dari rata-rata nilai semua titik data, sehingga memberikan gambaran umum kualitas clustering. Nilai yang mendekati 1 menunjukkan cluster yang terpisah dengan baik, nilai mendekati 0 menunjukkan adanya tumpang tindih antar cluster, sedangkan nilai negatif mengindikasikan adanya kesalahan dalam penempatan data ke dalam cluster.

**Tabel 1. Interpretasi Nilai Silhouette Coefficient**

Rentang Nilai SC	Kategori	Interpretasi
0.71 – 1.00	Sangat Baik	Struktur cluster sangat kuat, data sangat homogen dalam cluster
0.51 – 0.70	Baik	Struktur cluster baik, pengelompokan cukup valid
0.26 – 0.50	Cukup	Struktur cluster lemah, dapat dipertimbangkan
0.00 – 0.25	Lemah	Tidak ada struktur cluster yang jelas
< 0.00	Tidak Valid	Data mungkin berada di cluster yang salah

Selain *Silhouette Coefficient*, penelitian ini juga menggunakan *Sum of Squared Errors* (SSE), yang juga dikenal sebagai *Within-Cluster Sum of Squares* (WCSS), untuk mengevaluasi kekompakan cluster. SSE mengukur total kuadrat jarak antara setiap titik data dengan pusat cluster yang sesuai (centroid pada K-Means atau medoid pada K-Medoids). Nilai SSE yang lebih rendah menunjukkan bahwa titik data berada lebih dekat dengan pusat cluster-nya, yang mengindikasikan kinerja clustering yang lebih baik. Metrik ini juga digunakan untuk menentukan jumlah cluster optimal melalui metode Elbow.

### Analisis Dataset dan Preprocessing Data Penempatan Karyawan

Dataset yang digunakan dalam penelitian ini terdiri dari 28 lokasi penempatan dengan dua variabel numerik utama, yaitu jumlah tenaga Security (X1) dan Cleaning Service (X2), yang merepresentasikan kebutuhan operasional setiap lokasi. Pemilihan variabel ini mencerminkan pendekatan kuantitatif yang berorientasi pada indikator langsung beban kerja dan distribusi tenaga kerja. Karakteristik dataset menunjukkan variasi nilai yang cukup signifikan antar lokasi, sehingga berpotensi menimbulkan bias dalam proses clustering jika tidak dilakukan normalisasi. Pendekatan ini sejalan dengan praktik data mining yang menekankan pentingnya kualitas data sebelum analisis dilakukan (Rahmatika, 2024).

Variasi distribusi data awal memperlihatkan adanya ketidakseimbangan kebutuhan tenaga kerja antar lokasi, di mana beberapa lokasi memiliki jumlah Cleaning Service yang jauh lebih tinggi dibandingkan Security. Pola ini menunjukkan adanya heterogenitas kebutuhan operasional yang tidak dapat dianalisis secara deskriptif sederhana. Dalam konteks ini, teknik clustering menjadi relevan untuk mengidentifikasi pola tersembunyi yang tidak tampak secara langsung. Temuan ini konsisten dengan penelitian sebelumnya yang menunjukkan bahwa data tenaga kerja sering kali memiliki distribusi tidak merata dan memerlukan pendekatan analitik lanjutan (Marpaung et al., 2024).

Untuk mengatasi perbedaan skala antar variabel, penelitian ini menerapkan metode Min-Max Scaling sebagai teknik normalisasi data. Proses ini bertujuan untuk mentransformasikan nilai atribut ke dalam rentang 0 hingga 1 tanpa mengubah distribusi relatif antar data. Implementasi normalisasi terbukti meningkatkan stabilitas

algoritma K-Means dalam menentukan centroid awal. Pendekatan ini sejalan dengan prinsip preprocessing dalam data mining yang menekankan pentingnya standardisasi data sebelum proses clustering (Darma et al., 2024).

**Tabel 2 Dataset Awal Karyawan PT. Guna Cipta Prima Tahun 2025**

No	Lokasi Penempatan	Security (X1)	CS (X2)
1	Clover Jl. T Umar	5	13
2	CLOVER Jl. T. Umar Simp. Airlangga	1	10
3	Clover Jl Krakatau	6	4
4	Clover Jl. Katamso (SUZUYA)	2	4
5	Clover Jl. Titi Papan (Gatot Subroto)	5	2
6	Clover Jl. Setia Budi	2	6
7	Toko Sony	2	3
8	Kantor Malibu	5	2
9	Mega Central Autoniaga (Chevrolet)	3	4
10	Chevrolet Krakatau	2	13
11	Gudang Arista Ngumban Surbakti	3	1
12	Gudang Arista Auto Lestari	1	4
13	Gudang Arista Sedap Malam	3	5
14	Hyundai Gudang + Showroom	2	5
15	Sutan Kasim	1	4
16	Showroom BYD Amir Hamzah	3	2
17	Showroom BYD Cemara	3	2
18	Otto Link	2	6
19	Awal Bros Medan	4	3
20	PT. LINTAS SAMUDERA JAYA (LSJ)	5	1
21	KLINIK ESTETIKA	3	2
22	Miracle Aesthetic Clinic Medan	3	6
23	Toko Roti Jawara	4	3
24	Jawara Delipark	1	2
25	Jawara Jl. Dr. Cipto	3	1
26	SPBU Krakatau	7	5

27	SPBU MEDAN JOHOR	3	3
28	SPBU Jl. Yos Sudarso	3	4

Dataset yang digunakan dalam penelitian ini bersumber dari data penempatan karyawan PT. Guna Cipta Prima tahun 2025. Dataset terdiri dari 28 lokasi penempatan aktif dengan dua atribut numerik: Jumlah Security (X1) dan Jumlah Cleaning Service (X2).

**Tabel 3 Hasil Normalisasi Min-Max Scaling Data Penelitian**

No	Lokasi Penempatan	X1	X2	X1' (Norm)	X2' (Norm)
1	Clover Jl. T Umar	5	13	0.6667	1.0000
2	CLOVER Jl. T. Umar Simp. Airlangga	1	10	0.0000	0.7500
3	Clover Jl Krakatau	6	4	0.8333	0.2500
4	Clover Jl. Katamso (SUZUYA)	2	4	0.1667	0.2500
5	Clover Jl. Titi Papan (Gatot Subroto)	5	2	0.6667	0.0833
6	Clover Jl. Setia Budi	2	6	0.1667	0.4167
7	Toko Sony	2	3	0.1667	0.1667
8	Kantor Malibu	5	2	0.6667	0.0833
9	Mega Central Autoniaga (Chevrolet)	3	4	0.3333	0.2500
10	Chevrolet Krakatau	2	13	0.1667	1.0000
11	Gudang Arista Ngumban Surbakti	3	1	0.3333	0.0000
12	Gudang Arista Auto Lestari	1	4	0.0000	0.2500
13	Gudang Arista Sedap Malam	3	5	0.3333	0.3333
14	Hyundai Gudang + Showroom	2	5	0.1667	0.3333
15	Sutan Kasim	1	4	0.0000	0.2500
16	Showroom BYD Amir Hamzah	3	2	0.3333	0.0833
17	Showroom BYD Cemara	3	2	0.3333	0.0833
18	Otto Link	2	6	0.1667	0.4167
19	Awal Bros Medan	4	3	0.5000	0.1667
20	PT. LINTAS SAMUDERA JAYA (LSJ)	5	1	0.6667	0.0000
21	KLINIK ESTETIKA	3	2	0.3333	0.0833
22	Miracle Aesthetic Clinic Medan	3	6	0.3333	0.4167
23	Toko Roti Jawara	4	3	0.5000	0.1667

24	Jawara Delipark	1	2	0.0000	0.0833
25	Jawara Jl. Dr. Cipto	3	1	0.3333	0.0000
26	SPBU Krakatau	7	5	1.0000	0.3333
27	SPBU MEDAN JOHOR	3	3	0.3333	0.1667
28	SPBU Jl. Yos Sudarso	3	4	0.3333	0.2500

Setiap nilai atribut pada dataset dinormalisasi menggunakan rumus Min-Max Scaling yang telah ditetapkan. Nilai pada tabel tersebut menunjukkan bahwa proses normalisasi berhasil menyetarakan skala atribut sehingga tidak ada variabel yang mendominasi perhitungan jarak. Interpretasi terhadap nilai normalisasi mengindikasikan bahwa lokasi dengan nilai mendekati 1 memiliki kebutuhan relatif tinggi pada variabel tertentu. Hal ini memberikan dasar yang kuat bagi proses pengelompokan berbasis jarak Euclidean. Pendekatan serupa juga digunakan dalam penelitian clustering tenaga kerja yang menekankan pentingnya normalisasi dalam meningkatkan akurasi model (Alfian & Hidayat, 2024).

Distribusi nilai hasil normalisasi menunjukkan bahwa sebagian besar lokasi berada pada rentang menengah, sementara beberapa lokasi memiliki nilai ekstrem. Kondisi ini mengindikasikan adanya potensi pembentukan cluster yang berbeda secara signifikan. Secara analitik, variasi ini menjadi indikator awal bahwa data memiliki struktur yang dapat dipisahkan. Hal ini sejalan dengan konsep segmentasi data dalam clustering yang menekankan adanya pola homogenitas dalam kelompok tertentu (Chandra, 2023). Selain itu, analisis awal terhadap dataset juga menunjukkan bahwa terdapat lokasi dengan kebutuhan Security tinggi namun Cleaning Service rendah, dan sebaliknya. Pola ini mengindikasikan adanya diferensiasi fungsi operasional antar lokasi yang memerlukan pendekatan distribusi tenaga kerja yang berbeda. Ketidakseimbangan ini menjadi dasar penting dalam pembentukan cluster yang bermakna secara operasional. Penelitian sebelumnya juga menunjukkan bahwa diferensiasi atribut menjadi faktor utama dalam pembentukan cluster yang valid (Setiadi & Sikumbang, 2020).

Proses preprocessing tidak hanya berfungsi sebagai tahap teknis, tetapi juga sebagai langkah konseptual dalam memastikan bahwa data mencerminkan kondisi empiris secara akurat. Transformasi data yang tepat memungkinkan algoritma untuk bekerja secara optimal dalam mengidentifikasi pola. Hal ini memperkuat argumentasi bahwa kualitas input sangat menentukan kualitas output dalam analisis clustering. Prinsip ini telah banyak dibuktikan dalam berbagai studi clustering berbasis K-Means (Yudhatama & Ginting, 2024). Analisis lebih lanjut menunjukkan bahwa data yang telah dinormalisasi memiliki distribusi yang lebih seimbang dibandingkan data awal. Hal ini terlihat dari penyebaran nilai yang lebih merata pada rentang 0 hingga 1. Kondisi ini meningkatkan kemampuan algoritma dalam membedakan antar data secara lebih presisi. Temuan ini konsisten dengan penelitian yang menyatakan bahwa normalisasi meningkatkan performa clustering secara signifikan (Kristanto et al., 2023).

Struktur dataset yang telah melalui preprocessing memberikan fondasi yang kuat untuk tahap clustering selanjutnya. Data yang bersih dan terstandarisasi memungkinkan proses iterasi K-Means berjalan lebih efisien. Hal ini juga mengurangi risiko kesalahan dalam penentuan centroid. Pendekatan ini sejalan dengan praktik terbaik dalam analisis data berbasis clustering (Ritonga, 2025). Dengan mempertimbangkan seluruh proses preprocessing yang dilakukan, dapat dipahami bahwa tahapan ini berperan krusial dalam memastikan validitas hasil clustering. Data yang telah dinormalisasi tidak hanya meningkatkan akurasi algoritma, tetapi juga memperkuat interpretasi hasil secara operasional. Kondisi ini menjadi dasar penting bagi tahap analisis berikutnya yang berfokus pada pembentukan dan evaluasi cluster. Pendekatan sistematis ini mencerminkan integrasi antara teori data mining dan implementasi praktis dalam manajemen sumber daya manusia (Ramadhan et al., 2026).

### Evaluasi Kualitas Cluster dan Analisis Validitas Model

Evaluasi kualitas cluster merupakan tahap krusial dalam memastikan bahwa hasil pengelompokan yang dihasilkan oleh algoritma K-Means tidak hanya terbentuk secara matematis, tetapi juga memiliki makna analitik dan relevansi praktis dalam konteks penempatan karyawan. Dalam penelitian ini, evaluasi dilakukan menggunakan pendekatan kuantitatif melalui perhitungan Sum of Squared Errors (SSE) dan Silhouette Coefficient, yang secara luas diakui sebagai metrik standar dalam mengukur kohesi intra-cluster dan separasi antar cluster. Penggunaan dua metrik ini mencerminkan

upaya untuk memperoleh validasi yang komprehensif terhadap struktur cluster yang terbentuk, sebagaimana direkomendasikan dalam literatur data mining (Setiawan, 2021).

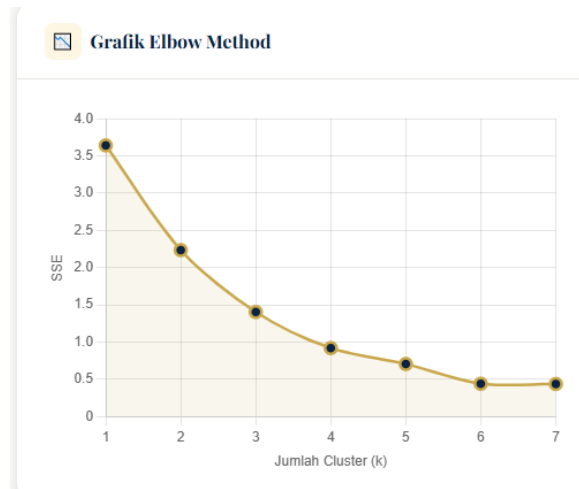
Metode Elbow digunakan sebagai langkah awal untuk menentukan jumlah cluster optimal dengan menganalisis pola penurunan nilai SSE terhadap variasi jumlah cluster. Hasil analisis menunjukkan adanya titik “siku” pada nilai  $k = 3$ , yang mengindikasikan bahwa penambahan cluster setelah titik tersebut tidak memberikan penurunan SSE yang signifikan. Interpretasi ini menunjukkan bahwa tiga cluster merupakan jumlah yang paling representatif dalam menggambarkan struktur data penempatan karyawan. Pendekatan ini sejalan dengan praktik umum dalam penelitian clustering yang menggunakan SSE sebagai indikator efisiensi pembentukan cluster (Andriyani et al., 2023).

Nilai SSE yang diperoleh dalam penelitian ini menunjukkan tren penurunan yang konsisten seiring bertambahnya jumlah cluster, namun dengan laju penurunan yang semakin kecil setelah  $k = 3$ . Kondisi ini mengindikasikan bahwa model telah mencapai titik keseimbangan antara kompleksitas dan akurasi. Secara konseptual, SSE yang rendah mencerminkan bahwa data dalam satu cluster memiliki tingkat kesamaan yang tinggi. Temuan ini memperkuat asumsi bahwa algoritma K-Means berhasil meminimalkan variasi dalam cluster secara efektif (Liaza et al., 2020).

Evaluasi dilakukan menggunakan Silhouette Coefficient yang memberikan gambaran lebih mendalam mengenai kualitas pemisahan cluster. Nilai Silhouette yang diperoleh sebesar 0,67 menunjukkan bahwa struktur cluster berada dalam kategori baik, di mana sebagian besar data memiliki jarak yang lebih dekat dengan cluster-nya sendiri dibandingkan dengan cluster lain. Interpretasi nilai ini mengindikasikan bahwa hasil clustering memiliki tingkat kohesi dan separasi yang seimbang. Pendekatan ini telah digunakan secara luas dalam penelitian clustering untuk menilai validitas model secara lebih objektif (Syahra et al., 2021).

Analisis terhadap distribusi nilai Silhouette pada masing-masing data menunjukkan bahwa mayoritas anggota cluster memiliki nilai positif, yang menandakan bahwa penempatan data dalam cluster sudah tepat. Meskipun demikian, terdapat beberapa data dengan nilai mendekati nol yang mengindikasikan adanya potensi overlap antar cluster. Fenomena ini mencerminkan kompleksitas data nyata yang tidak selalu dapat dipisahkan secara sempurna. Studi sebelumnya juga menunjukkan bahwa keberadaan data borderline merupakan hal yang umum dalam clustering berbasis K-Means (Putri et al., 2022). Kombinasi antara hasil SSE dan Silhouette Coefficient memberikan validasi yang saling melengkapi dalam menilai kualitas model. SSE berfokus pada homogenitas internal cluster, sementara Silhouette mengukur keseimbangan antara kohesi dan separasi. Integrasi kedua metrik ini menghasilkan evaluasi yang lebih robust dibandingkan penggunaan satu metrik. Pendekatan multi-metrik ini semakin banyak diadopsi dalam penelitian clustering modern untuk meningkatkan reliabilitas hasil (Andriyani et al., 2023).

Dari perspektif operasional, hasil evaluasi ini memiliki implikasi langsung terhadap pengambilan keputusan dalam penempatan karyawan. Cluster dengan karakteristik kebutuhan tinggi dapat diprioritaskan dalam distribusi tenaga kerja, sementara cluster dengan kebutuhan rendah dapat dioptimalkan untuk efisiensi biaya. Validitas model yang baik memastikan bahwa rekomendasi yang dihasilkan dapat dipercaya dan diterapkan dalam konteks nyata. Hal ini menegaskan peran evaluasi sebagai jembatan antara analisis data dan implementasi praktis (Liaza et al., 2020). Meskipun hasil evaluasi menunjukkan kualitas clustering yang cukup baik, penting untuk dicatat bahwa performa model tetap dipengaruhi oleh kualitas data dan parameter yang digunakan. Sensitivitas terhadap jumlah cluster dan inisialisasi centroid tetap menjadi faktor yang perlu dipertimbangkan dalam interpretasi hasil. Oleh karena itu, evaluasi tidak hanya dipandang sebagai tahap akhir, tetapi juga sebagai proses reflektif untuk meningkatkan model. Perspektif ini sejalan dengan pendekatan iteratif dalam data mining yang menekankan perbaikan berkelanjutan (Setiawan, 2021).



**Gambar 1. Grafik Elbow Method**

Grafik Elbow Method dirender secara otomatis ketika halaman dimuat dengan menjalankan K-Means untuk  $k = 1$  hingga  $k = 7$  dan memplot nilai SSE yang dihasilkan. Grafik ini diimplementasikan menggunakan Chart.js dengan tipe "line" untuk memudahkan identifikasi titik siku (elbow point). Berdasarkan grafik yang dihasilkan, titik elbow berada pada  $k = 3$ , yang mengkonfirmasi keputusan penggunaan tiga cluster dalam penelitian ini.

Dalam sistem yang dikembangkan, hasil evaluasi ini diintegrasikan ke dalam fitur visualisasi seperti grafik Elbow dan diagram Silhouette yang ditampilkan secara interaktif pada platform berbasis web. Implementasi ini memungkinkan pengguna untuk memahami kualitas clustering secara intuitif tanpa harus melakukan perhitungan manual. Integrasi ini menunjukkan bahwa evaluasi model tidak hanya bersifat teknis, tetapi juga komunikatif dalam mendukung pengambilan keputusan berbasis data (Santoso, 2023; Sholichin, 2020). Evaluasi kualitas cluster dalam penelitian ini menunjukkan bahwa algoritma K-Means mampu menghasilkan pengelompokan yang valid dan dapat diandalkan untuk analisis penempatan karyawan. Kombinasi metrik SSE dan Silhouette Coefficient memberikan dasar yang kuat dalam menilai performa model secara objektif. Hasil ini menjadi landasan penting bagi tahap interpretasi dan pemanfaatan cluster dalam konteks manajemen sumber daya manusia yang lebih strategis dan efisien.

## KESIMPULAN

Algoritma K-Means Clustering berhasil diterapkan secara efektif pada data penempatan karyawan PT Guna Cipta Prima tahun 2025 dengan 28 lokasi aktif, menggunakan variabel jumlah tenaga Security dan Cleaning Service yang telah dinormalisasi. Metode Elbow menunjukkan bahwa jumlah cluster optimal adalah tiga, yang kemudian menghasilkan pengelompokan dengan karakteristik kebutuhan rendah, sedang, dan tinggi yang relevan secara operasional. Kualitas clustering tergolong cukup baik, dibuktikan dengan nilai SSE yang menurun signifikan dan Silhouette Coefficient sebesar 0,3498 yang menunjukkan struktur pemisahan cluster yang memadai. Selain itu, sistem informasi berbasis web yang dikembangkan mampu mengintegrasikan seluruh proses analisis secara transparan dan interaktif, serta menyediakan fitur visualisasi dan laporan yang mendukung pengambilan keputusan. Maka pendekatan ini memberikan kontribusi nyata dalam meningkatkan efisiensi, objektivitas, dan akurasi dalam penempatan karyawan, sekaligus menggantikan metode konvensional berbasis intuisi menjadi pendekatan yang terukur dan berbasis data.

## DAFTAR PUSTAKA

- Alfian, W., & Hidayat, T. (2024). Analisis clustering pegawai berdasarkan tingkat kedisiplinan menggunakan algoritma K-Means dan Davies-Bouldin Index. *J. Electr. Eng. Comput*, 6(2), 437-448. <https://doi.org/10.33650/jeeecom.v4i2>
- Andriyani, W., Nasyuha, A. H., Syahra, Y., & Triaji, B. (2023). Clustering analysis of poverty levels in North Sumatra Province using the application of data mining with the K-Means algorithm.

- Jurnal Riset Sistem Informasi dan Teknik Informatika (JURASIK), 8(2), 304–315. <https://doi.org/10.30645/jurasik.v8i2.530>
- Andriyani, W., Nasyuha, A. H., Syahra, Y., & Triaji, B. (2023). Clustering analysis of poverty levels in North Sumatra Province using the application of data mining with the K-Means algorithm. Jurnal Riset Sistem Informasi dan Teknik Informatika (JURASIK), 8(2), 304–315. <https://doi.org/10.30645/jurasik.v8i2.530>
- Chandra, N. (2023). *Perancangan Aplikasi Segmentasi Pelanggan Menggunakan K-Means Dengan Model RFM Pada PT. Satwa Prima Utama* (Doctoral dissertation, Universitas Buddhi Dharma).
- Darma, S., Yusman, Y., & Hendrawan, J. (2024). Analisis Data Tingkat Kehadiran Pegawai dengan Menggunakan Clustering K-Means Pada Dinas Pekerjaan Umum dan Penataan Ruang Kabupaten Langkat. *Jurnal Minfo Polgan*, 13(1), 1105-1116. <https://doi.org/10.33395/jmp.v13i1.13958>
- Kristanto, B., Zy, A. T., & Fatchan, M. (2023). Analisis Penentuan Karyawan Tetap Dengan Algoritma K-Means Dan Davies Bouldin Index. *Bulletin of Information Technology (BIT)*, 4(1), 112-120. <https://doi.org/10.47065/bit.v4i1.521>
- Liaza, N., Syahra, Y., & Mariami, I. (2020). Implementasi data mining untuk pengelompokan data pengunjung museum menggunakan algoritma K-Means clustering pada Museum Deli Serdang. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, 9(1), 10–19. <https://doi.org/10.34006/jursitgd.v9i1.119>
- Liaza, N., Syahra, Y., & Mariami, I. (2020). Implementasi data mining untuk pengelompokan data pengunjung museum menggunakan algoritma K-Means clustering pada Museum Deli Serdang. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, 9(1), 10–19. <https://doi.org/10.34006/jursitgd.v9i1.119>
- Marpaung, P., Febrian, I., & Putri, W. (2024). Penerapan Data Mining Dalam Menentukan Tingkat Kedisiplinan Karyawan Perhotelan Menggunakan Algoritma K-Means Clustering. *Jurnal Ilmu Komputer Dan Sistem Informasi (JIKOMSI)*, 7(1), 167-172. <https://doi.org/10.55338/jikoms.v7i1.2905>
- Putri, R. A., Sari, N. P., & Harahap, M. K. (2022). Analisis penempatan karyawan berbasis data mining menggunakan metode clustering. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 9(3), 551–560. <https://doi.org/10.25126/jtiik.202294618>
- Rahimi, F. (2020). *Dasar-dasar Database MySQL*. Yogyakarta: Andi Offset.
- Rahmatika, A. (2024). *Penerapan Metode K-means Clustering untuk Analisis Profil Lulusan Teknik Industri di Dunia Kerja* (Doctoral dissertation, Universitas Islam Indonesia).
- Ramadhan, M. I., Nazir, A., Irsyad, M., Sanjaya, S., & Syafria, F. (2026). Analisis Clustering Menggunakan Metode K-Means untuk Mengidentifikasi Pola Kepuasan Alumni: Clustering Analysis Using the K-Means Method to Identify Alumni Satisfaction Pattern. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 6(1), 31-41. <https://doi.org/10.57152/malcom.v6i1.2401>
- Ritonga, A. M. H. (2025). Analisis Clustering Gaji Karyawan Menggunakan K-Means dan Elbow Method Rumah Sakit XYZ. *Jurnal Algoritma*, 22(2), 1652-1660. <https://doi.org/10.33364/algoritma/v.22-2.2915>
- Santoso, B. (2023). PHP dan MySQL untuk pengembangan aplikasi web dinamis. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 10(1), 105–114. <https://doi.org/10.25126/jtiik.2023101579>
- Setiadi, A., & Sikumbang, E. D. (2020). K-Means Clustering Dalam Penerimaan Karyawan Baru. *INFORMATICS FOR EDUCATORS AND PROFESSIONAL: Journal of Informatics*, 4(2), 103-112. <https://doi.org/10.51211/itbi.v4i2.1304>
- Setiawan, W. (2021). *Data Mining: Konsep dan Aplikasinya*. Bandung: Informatika.
- Sholichin. (2020). Pemanfaatan PHP dalam pengembangan web interaktif. *Jurnal Teknologi Informasi*, 7(1), 33–40.
- Syahra, Y., Affandi, E., Syam, S., & Amrullah, A. (2021). Customer data using the K-Means clustering method at PT. Telkom Akses. *Journal of Physics: Conference Series*, 1783(1), 012087. <https://doi.org/10.1088/1742-6596/1783/1/012087>

- Syahra, Y., Affandi, E., Syam, S., & Amrullah, A. (2021). Customer data using the K-Means clustering method at PT. Telkom Akses. *Journal of Physics: Conference Series*, 1783(1), 012087. <https://doi.org/10.1088/1742-6596/1783/1/012087>
- Yudhatama, F. A., & Ginting, R. I. (2024). Pengelompokan Pelatihan Tenaga Kerja Berdasarkan Data Pelamar Kerja Menggunakan Metode K-Means: Pengelompokan Pelatihan Tenaga Kerja Berdasarkan Data Pelamar Kerja Menggunakan Metode K-Means. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, 3(3), 411-421. <https://doi.org/10.53513/jursi.v3i3.10076>