

**ANALISIS K-MEANS CLUSTERING PADA DATA PENEMPATAN
KARYAWAN DI PT. GUNA CIPTA PRIMA BERBASIS WEB**

SKRIPSI

DISUSUN OLEH

SELLA

NPM. 2009010118



UMSU

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2026

**ANALISIS K-MEANS CLUSTERING PADA DATA
PENEMPATAN KARYAWAN DI PT. GUNA CIPTA PRIMA
BERBASIS WEB**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer
dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

SELLA

NPM. 2009010118

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

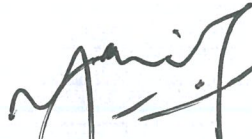
MEDAN

2026

LEMBAR PENGESAHAN

Judul Skripsi : ANALISIS K-MEANS CLUSTERING PADA DATA
PENEMPATAN KARYAWAN DI PT. GUNA CIPTA
PRIMA BERBASIS WEB
Nama Mahasiswa : SELLA
NPM : 2009010118
Program Studi : SISTEM INFORMASI

Menyetujui
Komisi Pembimbing



(Yohanni Syahra, S.Si., M.Kom)
NIDN. 0129108201

Ketua Program Studi



(Mahardika Abdi Prawira Tanjung,
S.Kom., M.Kom)
NIDN. 0117088902

Dekan



(Dr. Al-Khowarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

ANALISIS K-MEANS CLUSTERING PADA DATA PENEMPATAN KARYAWAN DI PT. GUNA CIPTA PRIMA BERBASIS WEB

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, April 2026

Yang membuat pernyataan



NPM. 2009010118

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama	: Sella
NPM	: 2009010118
Program Studi	: Sistem Informasi
Karya Ilmiah	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**ANALISIS K-MEANS CLUSTERING PADA DATA
PENEMPATAN KARYAWAN DI PT. GUNA CIPTA PRIMA
BERBASIS WEB**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, April 2026

Yang membuat pernyataan


SELLA

NPM. 2009010118

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Sella
Tempat dan Tanggal Lahir : Binjai, 14 Agustus 2002
Alamat Rumah : Jl. Alfaka IX NO.113-A LK VI
Telepon/Faks/HP : 081269188424
E-mail : ssseeelllaaaa@gmail.com
Instansi Tempat Kerja : Belum Bekerja
Alamat Kantor : -

DATA PENDIDIKAN

SD	: SDN 060870 Medan	TAMAT: 2012
SMP	: SMPN 37 Medan	TAMAT: 2017
SMA	: SMK Telkom Medan	TAMAT: 2020

KATA PENGANTAR



Puji dan syukur penulis panjatkan kehadirat Allah SWT atas limpah rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **“ANALISIS K-MEANS CLUSTERING PADA DATA PENEMPATAN KARYAWAN DI PT. GUNA CIPTA PRIMA BERBASIS WEB”**

Dalam proses penyusunan skripsi ini penulis menyadari bahwa tidak akan dapat menyelesaikan tanpa doa, dukungan, dan bantuan dari beberapa pihak. Oleh karena itu, dengan segala kerendahan hati, penulis menyampaikan yang sebesar-besarnya kepada :

1. Bapak Prof. Dr. Akrim, M.Pd.. Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Dr. Firahmi Rizky, S.Kom., M.Kom., Wakil Dekan I Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
4. Bapak Mhd. Basri, S.Si, M.Kom., Wakil Dekan II Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
5. Bapak Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom. Ketua Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
6. Bapak Mulkan Azhari, S.Kom., M.Kom., Sekretaris Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
7. Ibu Yohanni Syahra, S.Si., M.Kom selaku pembimbing saya yang telah memberikan bimbingan, arahan, dan ilmu yang berharga selama proses penyusunan skripsi ini.
8. Ayahanda Arman. Terimakasih atas segala pengorbanan, kasih sayang, serta serta segala bentuk tanggung jawab yang ayah berikan. Terimakasih telah

menjadi alasan terbesarku untuk tetap semangat meraih berjuang meraih gelar sarjana yang telah ayah impikan.

9. Yang teristimewa, alm. ibu Anni, yang telah melahirkan, merawat, membesarkan dan mendoakan penulis semasa hidupnya agar menjadi anak yang berguna dimasa depan kelak.
10. Yang tersayang, ibu Sustiana, S.E., yang telah membesarkan saya, merawat, mendidik dan mendoakan yang tidak ternilai kepada penulis sampai sekarang ini sehingga penulis dapat menyelesaikan skripsi ini.
11. Terakhir, saya berterimakasih kepada satu sosok gadis yang selama ini diam-diam berjuang tanpa henti, seorang perempuan sederhana dengan hati kecil tetapi dengan impian besar. Terimakasih kepada peneliti skripsi ini yaitu diriku sendiri, Sella. Anak perempuan pertama dan harapan orangtuanya. Terimakasih telah bertahan sejauh ini, dan terus berjalan melewati segala tantangan semesta hadirkan. Aku berdoa semoga langkah kecilku selalu diperkuat, dikelilingin orang-orang baik dan hebat, serta mimpi ku satu persatu akan terjawab. Aamiin.

ABSTRAK

Perkembangan teknologi informasi telah memberikan kontribusi besar dalam mendukung proses pengambilan keputusan di berbagai bidang, termasuk manajemen sumber daya manusia. Salah satu permasalahan umum yang dihadapi perusahaan adalah kesulitan dalam menentukan penempatan karyawan secara objektif sesuai dengan kemampuan dan kinerjanya. Untuk mengatasi hal tersebut, penelitian ini menerapkan algoritma K-Means Clustering sebagai metode analisis pengelompokan data karyawan di PT. Guna Cipta Prima.

Penelitian ini bertujuan untuk menerapkan algoritma K-Means Clustering pada data penempatan karyawan, merancang sistem berbasis web menggunakan PHP dan MySQL untuk menampilkan hasil clustering, serta menganalisis hasil pengelompokan yang diperoleh. Data karyawan yang digunakan dalam penelitian ini bersifat real dan terdiri atas beberapa atribut seperti security dan cleaning service. Proses pengelompokan dilakukan dengan jumlah cluster (k) sebanyak tiga, yaitu: kinerja tinggi, sedang, dan rendah.

Hasil penelitian menunjukkan bahwa sistem berbasis web yang dibangun dapat menjalankan algoritma K-Means dengan baik serta menghasilkan pengelompokan yang sesuai dengan karakteristik data. Nilai Silhouette Coefficient sebesar 0,67 menunjukkan bahwa tingkat validitas hasil clustering cukup baik. Sistem ini dapat membantu bagian HRD dalam melakukan analisis penempatan karyawan secara lebih efisien, objektif, dan berbasis data.

Kata Kunci: Data Mining, K-Means Clustering, Penempatan Karyawan, PHP, MySQL, Sistem Informasi Berbasis Web.

ABSTRACT

The development of information technology has made a significant contribution to supporting decision-making processes in various fields, including human resource management. One common problem faced by companies is the difficulty in objectively determining employee placement according to their abilities and performance. To address this issue, this study applies the K-Means Clustering algorithm as a method for analyzing employee data grouping at PT. Guna Cipta Prima.

This study aims to apply the K-Means Clustering algorithm to employee placement data, design a web-based system using PHP and MySQL to display the clustering results, and analyze the obtained grouping results. The employee data used in this study is real and consists of several attributes such as security and cleaning services. The clustering process is carried out with three clusters (k), namely: high, medium, and low performance.

The research results indicate that the web-based system developed can run the K-Means algorithm effectively and produce clustering that aligns with the characteristics of the data. A Silhouette Coefficient value of 0.67 shows that the validity of the clustering results is quite good. This system can assist the HR department in conducting employee placement analysis more efficiently, objectively, and based on data.

Keywords: Data Mining, K-Means Clustering, Employee Placement, PHP, MySQL, Web-Based Information System.

DAFTAR ISI

LEMBAR PENGESAHAN.....	ii
PERNYATAAN ORISINALITAS.....	iii
RIWAYAT HIDUP.....	v
KATA PENGANTAR.....	vi
ABSTRAK.....	viii
ABSTRACT.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR.....	xv
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	6
1.5.1 Manfaat bagi PT. Guna Cipta Prima.....	6
1.5.2 Manfaat bagi Peneliti.....	6
1.5.3 Manfaat bagi Universitas Muhammadiyah Sumatera Utara (UMSU)....	7
BAB II.....	8
LANDASAN TEORI.....	8
2.1 Data Mining.....	8
2.1.1 Fungsi Data Mining.....	8
2.2 Algoritma K-Means.....	9
2.2.1 Kelebihan dan Kekurangan Algoritma K-means.....	10
2.3 Clustering.....	11
2.3.1 Jenis-jenis clustering (pengelompokan).....	12
2.4 Data dan Alat Penelitian.....	15
2.4.1. Data karyawan.....	15
2.4.2. Data variabel yang diberikan pihak perusahaan	15
2.4.3. Data lokasi penempatan karyawan.....	15
2.4.4. Alat Penelitian.....	16
2.5 Web pengembangan sistem.....	16
2.5.1 PHP (Hypertext Preprocessor).....	16

2.5.2 MySQL (My Structured Query Language).....	17
2.5.3 XAMPP.....	18
2.6 Penelitian yang Relevan.....	19
BAB III.....	22
METODOLOGI PENELITIAN.....	22
3.1 Jenis dan Pendekatan Penelitian.....	22
3.2 Lokasi, Waktu, dan Subjek Penelitian.....	23
3.2.1 Lokasi Penelitian.....	23
3.2.2 Waktu Penelitian.....	23
3.2.3 Subjek dan Objek Penelitian.....	24
3.3 Data dan Sumber Data.....	24
3.3.1 Jenis Data.....	24
3.3.2 Sumber Data.....	24
3.3.3 Variabel / Atribut Penelitian.....	25
3.4 Teknik Pengumpulan Data.....	25
3.4.1 Studi Literatur.....	25
3.4.2 Observasi.....	25
3.4.3 Wawancara.....	26
3.5 Teknik Pengolahan Data.....	26
3.5.1 Data Cleaning.....	26
3.5.2 Data Transformation.....	27
3.5.3 Normalisasi Data (Min-Max Scaling).....	27
3.6 Analisis Manual K-Means Clustering.....	28
3.6.1 Dataset Penelitian.....	28
3.6.2 Proses Normalisasi Data.....	29
3.6.3 Penentuan Jumlah Cluster (k) dengan Metode Elbow.....	31
3.6.4 Penentuan Centroid Awal.....	32
3.6.5 Iterasi 1 – Perhitungan Jarak Euclidean dan Pembentukan Cluster.....	32
3.6.6 Pembaruan Centroid Setelah Iterasi 1.....	34
3.6.7 Iterasi 2 – Perhitungan Jarak Euclidean dan Cluster Baru.....	35
3.6.8 Iterasi 3 dan Konvergensi.....	37
3.6.9 Hasil Akhir Clustering.....	38
3.7 Evaluasi Kinerja Clustering.....	41
3.7.1 Silhouette Coefficient.....	41
3.7.2 Evaluasi WCSS/SSE.....	42
3.7.3 Validasi Hasil Cluster.....	42

3.8 Perancangan Sistem.....	43
3.8.1 Gambaran Umum Sistem.....	43
3.8.2 Desain Alur Sistem (System Flowchart).....	44
3.8.3 Use Case Diagram (Naratif).....	44
3.8.4 Activity Diagram (Naratif).....	45
3.8.5 Perancangan Basis Data (Database Design).....	46
3.8.6 Relasi Antar Tabel (ERD Deskriptif).....	50
3.8.7 Skema SQL DDL (PHP 8.1 + MySQLi).....	50
3.8.8 Desain Antarmuka (UI/UX).....	52
3.8.9 Desain Implementasi Algoritma K-Means dalam PHP 8.1.....	52
3.9 Alat dan Perangkat Penelitian.....	54
3.9.1 Perangkat Keras (Hardware).....	54
3.9.2 Perangkat Lunak (Software).....	54
3.10 Diagram Alur Penelitian.....	54
BAB IV.....	57
HASIL DAN PEMBAHASAN.....	57
4.1 Hasil Implementasi Sistem.....	57
4.1.1 Halaman Login.....	57
4.1.2 Halaman Dashboard.....	58
4.1.3 Halaman Data Lokasi.....	59
4.1.4 Halaman Data Karyawan.....	61
4.1.5 Halaman Proses Clustering.....	62
4.1.6 Halaman Hasil Clustering.....	65
4.1.7 Halaman Laporan.....	66
4.2 Hasil Analisis K-Means Clustering.....	67
4.2.1 Normalisasi Data.....	68
4.2.2 Penentuan Jumlah Cluster dengan Metode Elbow.....	69
4.2.3 Proses Iterasi K-Means.....	70
4.2.4 Hasil Pengelompokan Akhir.....	71
4.2.5 Ringkasan Anggota Per Cluster.....	73
4.3 Evaluasi Kualitas Clustering.....	74
4.3.1 Silhouette Coefficient.....	74
4.3.2 Sum of Squared Error (SSE).....	76
4.4 Pembahasan Hasil Clustering dan Rekomendasi.....	77
4.4.1 Cluster 1 — Kebutuhan Rendah (2 Lokasi).....	77
4.4.2 Cluster 2 — Kebutuhan Sedang (17 Lokasi).....	77

4.4.3 Cluster 3 — Kebutuhan Tinggi (9 Lokasi).....	78
4.5 Pengujian Sistem dengan Metode Black-Box Testing.....	79
4.6 Kelebihan dan Kekurangan Sistem.....	83
4.6.1 Kelebihan Sistem.....	83
4.6.2 Kekurangan Sistem.....	85
4.7 Perbandingan dengan Penelitian Terdahulu.....	86
BAB V.....	88
PENUTUP.....	88
5.1 Kesimpulan.....	88
5.2 Saran.....	90
5.2.1 Saran untuk PT. Guna Cipta Prima.....	90
5.2.2 Saran untuk Pengembangan Sistem.....	91
5.2.3 Saran untuk Penelitian Selanjutnya.....	93
DAFTAR PUSTAKA.....	96

DAFTAR TABEL

	Halaman
Tabel 2.1 Penelitian yang relevan	20
Tabel 3.1 Jadwal Penelitian	23
Tabel 3.2 Variabel/Atribut Penelitian	25
Tabel 3.3 Dataset Penelitian - Data Penempatan Karyawan PT. Guna Cipta Prima Tahun 2026	28
Tabel 3.4 Hasil Normalisasi Min-Max Scaling Data Penelitian	30
Tabel 3.5 Nilai SSE per Jumlah Cluster (Metode Elbow)	31
Tabel 3.6 Centroid Awal Iterasi 1	32
Tabel 3.7 Hasil Perhitungan Jarak Euclidean Iterasi 1	33
Tabel 3.8 Centroid Baru Setelah Iterasi 1	35
Tabel 3.9 Hasil Perhitungan Jarak Euclidean Iterasi 2 (* = perpindahan cluster)	35
Tabel 3.10 Centroid Baru Setelah Iterasi 2	37
Tabel 3.11 Hasil Perhitungan Jarak Euclidean Iterasi 3	37
Tabel 3.12 Ringkasan Perubahan Centroid Per Iterasi dan Konvergensi	38
Tabel 3.13 Hasil Akhir Clustering K-Means (k=3)	39
Tabel 3.14 Rekapitulasi Hasil Akhir Clustering	40
Tabel 3.15 Interpretasi Nilai Silhouette Coefficient	41
Tabel 3.16 Deskripsi Alur Sistem Web K-Means Clustering	44
Tabel 3.17 Spesifikasi Use Case Diagram (Naratif)	45
Tabel 3.18 Struktur Tabel tbl_users	47
Tabel 3.19 Struktur Tabel tbl_lokasi	47
Tabel 3.20 Struktur Tabel tbl_karyawan	48
Tabel 3.21 Struktur Tabel tbl_clustering_result	49
Tabel 3.22 Struktur Tabel tbl_clustering_detail	49
Tabel 3.23 Relasi Antar Tabel Basis Data	50
Tabel 3.24 Rancangan Antarmuka Sistem (UI/UX)	52
Tabel 3.25 Spesifikasi Perangkat Keras	54
Tabel 3.26 Spesifikasi Perangkat Lunak	54
Tabel 3.27 Penjelasan Tahapan dan Alur Penelitian	55
Tabel 4.1 Hasil Normalisasi Min-Max Scaling Data Penempatan Karyawan	68
Tabel 4.2 Nilai SSE untuk Setiap Nilai k (Metode Elbow)	70
Tabel 4.3 Centroid Awal (Inisialisasi) K-Means	71
Tabel 4.4 Pergerakan Centroid Per Iterasi	71
Tabel 4.5 Hasil Pengelompokan Lengkap K-Means (k=3, Iterasi=4)	72
Tabel 4.6 Ringkasan Distribusi Anggota dan Statistik Per Cluster	73
Tabel 4.7 Interpretasi Nilai Silhouette Coefficient	75
Tabel 4.8 Perbandingan Evaluasi Kualitas Clustering	76
Tabel 4.9 Hasil Pengujian Black-Box Testing Sistem	79
Tabel 4.10 Perbandingan dengan Penelitian Terdahulu	86
Tabel 5.1 Ringkasan Temuan Utama Penelitian	94

DAFTAR GAMBAR

	Halaman
Gambar 2.1	Hierarchical clustering sumber (Xu & Wunsch, 2009) 13
Gambar 2.2	Tampilan XAMPP 19
Gambar 3.1	Skema SQL DDL Basis Data Sistem (PHP 8.1 + MySQLi) 52
Gambar 3.2	Pseudocode Implementasi K-Means dalam PHP 8.1 + MySQLi 54
Gambar 3.3	Diagram Alur 55
Gambar 4.1	Tampilan Halaman Login Sistem 58
Gambar 4.2	Tampilan Halaman Dashboard 59
Gambar 4.3	Tampilan Halaman Data Lokasi 60
Gambar 4.4	Tampilan Modal Dialog Tambah/Edit Lokasi 61
Gambar 4.5	Tampilan Halaman Data Karyawan 62
Gambar 4.6	Tampilan Halaman Proses Clustering Sebelum Dijalankan 63
Gambar 4.7	Grafik Elbow Method Hasil Sistem 63
Gambar 4.8	Tampilan Hasil Proses Clustering K=3 64
Gambar 4.9	Tampilan Scatter Plot 2D Hasil Clustering 65
Gambar 4.10	Tampilan Tabel Detail Hasil Pengelompokan 66
Gambar 4.11	Tampilan Halaman Laporan 67
Gambar 4.12	Tampilan Laporan Saat Mode Cetak (Print Preview) 67

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam konteks perusahaan yang memiliki banyak lokasi penempatan tersebar di berbagai wilayah, tantangan penempatan karyawan menjadi semakin kompleks karena setiap lokasi memiliki karakteristik kebutuhan tenaga kerja yang berbeda-beda. Kondisi ini menuntut adanya pendekatan berbasis data yang mampu mengidentifikasi pola kebutuhan secara objektif dan terstruktur, bukan semata-mata mengandalkan intuisi atau pertimbangan subjektif manajemen.

PT. Guna Cipta Prima merupakan perusahaan penyedia jasa keamanan (security) dan kebersihan (cleaning service) yang beroperasi di Kota Medan dengan 28 lokasi penempatan aktif yang tersebar di berbagai wilayah. Berdasarkan hasil observasi dan wawancara dengan bagian administrasi perusahaan, diperoleh informasi bahwa proses penempatan karyawan selama ini dilakukan secara manual dan berbasis pengalaman, tanpa dukungan analisis data yang sistematis. Hal ini mengakibatkan ketidakseragaman distribusi tenaga kerja antar lokasi, di mana beberapa lokasi mengalami kelebihan personel sementara lokasi lain justru kekurangan, sehingga efisiensi operasional secara keseluruhan belum optimal.

Permasalahan penempatan karyawan yang tidak berbasis data merupakan isu yang umum dijumpai pada perusahaan dengan skala operasional multi-lokasi. Menurut Putri et al. (2022), ketidaksesuaian antara jumlah tenaga kerja yang ditempatkan dengan kebutuhan aktual suatu lokasi dapat berdampak langsung pada kualitas layanan dan pemborosan anggaran operasional. Oleh karena itu, diperlukan suatu metode analisis yang mampu mengelompokkan lokasi-lokasi berdasarkan pola kebutuhan tenaga kerja secara kuantitatif dan reproduktif.

Data mining hadir sebagai disiplin yang menawarkan berbagai teknik komputasional untuk mengekstrak pola bermakna dari kumpulan data berukuran besar. Salah satu teknik yang paling banyak digunakan dalam analisis pengelompokan adalah clustering, yaitu proses pengelompokan objek data ke dalam sejumlah kelompok berdasarkan kemiripan nilai atribut (Han et al., 2012). Di antara algoritma clustering yang ada, K-Means merupakan metode partisi yang paling populer karena kesederhanaan implementasinya, efisiensi komputasi pada

dataset berukuran sedang, serta kemampuannya menghasilkan kelompok yang kompak dan terpisah secara jelas (Ahmed & Ashour, 2011).

Beberapa penelitian sebelumnya telah membuktikan keefektifan K-Means Clustering dalam berbagai domain. Syahra et al. (2021) menerapkan K-Means pada data pelanggan PT. Telkom Akses dan berhasil menghasilkan klaster yang bermakna untuk mendukung pengambilan keputusan bisnis. Andriyani et al. (2023) menggunakan K-Means untuk menganalisis tingkat kemiskinan di Provinsi Sumatera Utara dengan hasil pengelompokan yang dapat dijadikan dasar kebijakan daerah. Liaza et al. (2020) berhasil mengelompokkan data pengunjung Museum Deli Serdang ke dalam tiga cluster dengan karakteristik kunjungan yang berbeda dan tervalidasi. Penelitian-penelitian tersebut memperkuat relevansi penerapan K-Means Clustering pada data penempatan karyawan di PT. Guna Cipta Prima.

Penerapan K-Means pada konteks 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).

Dengan demikian, 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.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka rumusan masalah dalam penelitian ini dirumuskan sebagai berikut:

- a. Bagaimana menerapkan algoritma K-Means Clustering untuk mengelompokkan 28 lokasi penempatan karyawan PT. Guna Cipta Prima berdasarkan jumlah tenaga Security dan Cleaning Service menggunakan normalisasi Min-Max Scaling?
- b. Berapa jumlah cluster optimal (k) yang dihasilkan oleh Metode Elbow pada dataset penempatan karyawan PT. Guna Cipta Prima, dan bagaimana kualitas pengelompokan tersebut diukur menggunakan Silhouette Coefficient?
- c. Bagaimana merancang dan mengimplementasikan sistem informasi berbasis web menggunakan PHP 8.1 dan MySQLi yang dapat menjalankan proses K-Means Clustering secara otomatis dan menampilkan hasil pengelompokan secara interaktif?
- d. Bagaimana hasil pengelompokan lokasi penempatan dapat dimanfaatkan sebagai rekomendasi distribusi karyawan yang lebih objektif bagi manajemen PT. Guna Cipta Prima?

1.3 Batasan Masalah

Untuk menjaga fokus dan ketajaman penelitian, batasan masalah ditetapkan sebagai berikut:

- a. Data yang digunakan adalah data penempatan karyawan aktual PT. Guna Cipta Prima periode tahun 2025, mencakup 28 lokasi penempatan aktif dengan dua variabel numerik utama: jumlah tenaga Security (X1) dan jumlah tenaga Cleaning Service (X2) per lokasi.
- b. Metode pengelompokan yang digunakan secara eksklusif adalah algoritma K-Means Clustering dengan normalisasi Min-Max Scaling; perbandingan dengan algoritma clustering lain tidak termasuk dalam cakupan penelitian ini.

- c. Penentuan jumlah cluster optimal dilakukan menggunakan Metode Elbow berbasis nilai SSE, dan kualitas clustering dievaluasi menggunakan Silhouette Coefficient.
- d. Sistem informasi yang dikembangkan berbasis web menggunakan bahasa pemrograman PHP versi 8.1, ekstensi MySQLi, dan basis data MySQL; tidak mencakup pengembangan aplikasi berbasis desktop atau mobile native.
- e. Pengujian sistem dilakukan menggunakan metode Black-Box Testing yang mencakup seluruh fungsionalitas utama sistem, tidak termasuk pengujian performa beban (load testing).

1.4 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dalam penelitian ini adalah:

- a. Menerapkan algoritma K-Means Clustering pada data penempatan karyawan PT. Guna Cipta Prima dengan menggunakan normalisasi Min-Max Scaling untuk menghasilkan pengelompokan lokasi berdasarkan pola kebutuhan tenaga Security dan Cleaning Service.
- b. Menentukan jumlah cluster optimal menggunakan Metode Elbow dan mengevaluasi kualitas hasil pengelompokan menggunakan Silhouette Coefficient serta nilai SSE.
- c. Merancang dan mengimplementasikan sistem informasi berbasis web menggunakan PHP 8.1 dan MySQLi yang mengintegrasikan proses preprocessing data, eksekusi K-Means, visualisasi scatter plot, grafik Elbow, serta penyajian hasil clustering secara interaktif.
- d. Menghasilkan rekomendasi distribusi karyawan berbasis data kepada manajemen PT. Guna Cipta Prima sebagai dasar pengambilan keputusan penempatan yang lebih efisien dan objektif.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat bagi beberapa pihak sebagai berikut:

1.5.1 Manfaat bagi PT. Guna Cipta Prima

- a. Perusahaan memperoleh sistem berbasis web yang dapat langsung digunakan oleh bagian administrasi untuk melakukan analisis pengelompokan lokasi penempatan secara otomatis dan berulang pada periode berikutnya.

- b. Manajemen mendapatkan rekomendasi distribusi tenaga Security dan Cleaning Service berbasis data yang dapat mendukung efisiensi alokasi karyawan di 28 lokasi penempatan aktif.
- c. Perusahaan dapat mengidentifikasi lokasi-lokasi berkebutuhan tinggi, sedang, dan rendah secara objektif sehingga dapat menyusun strategi rekrutmen dan penempatan yang lebih tepat sasaran.

1.5.2 Manfaat bagi Peneliti

- a. Penelitian ini menjadi sarana penerapan langsung konsep data mining, khususnya algoritma K-Means Clustering dengan normalisasi Min-Max Scaling, dalam konteks permasalahan nyata di dunia industri jasa.
- b. Peneliti memperoleh pengalaman dalam merancang dan mengimplementasikan sistem informasi berbasis web menggunakan teknologi PHP 8.1, MySQLi, dan Bootstrap 5, yang relevan dengan kompetensi profesi di bidang Sistem Informasi.
- c. Penelitian ini memperluas wawasan peneliti dalam memahami proses lengkap analisis data mining mulai dari preprocessing, pemodelan algoritma, evaluasi hasil, hingga pengembangan sistem penyajian informasi.

1.5.3 Manfaat bagi Universitas Muhammadiyah Sumatera Utara (UMSU)

- a. Penelitian ini memberikan kontribusi pada khazanah akademik Program Studi Sistem Informasi FIKTI UMSU berupa karya ilmiah yang menghubungkan teori data mining dengan implementasi sistem nyata pada perusahaan lokal di Kota Medan.
- b. Hasil penelitian dapat menjadi referensi dan acuan bagi peneliti berikutnya yang ingin mengembangkan studi serupa, baik dalam konteks K-Means Clustering maupun pengembangan sistem informasi berbasis web di bidang manajemen sumber daya manusia.
- c. Penelitian ini memperkuat relevansi dan kontribusi UMSU dalam pengembangan ilmu pengetahuan terapan di bidang teknologi informasi, khususnya dalam mendukung transformasi digital sektor jasa di Sumatera Utara.

BAB II

LANDASAN TEORI

2.1 Data Mining

Menurut (Wahyudi Setiawan, 2021) Data Mining adalah proses pengambilan atau pengumpulan informasi penting dari kumpulan data yang besar. Proses mining atau penambangan data sering menggunakan teknik statistik dan matematika dengan memanfaatkan teknik kecerdasan buatan. Nama lain dari kata mining yaitu Know Discovery in Database (KDD), Big data, Bussiness intellingence, Know extraction, Pattern analysis dan Information harvesting.

Terdapat banyak konsep dan teknik yang digunakan dalam proses data mining. Proses ini membutuhkan beberapa langkah untuk mendapatkan data yang dibutuhkan. Proses KDD meliputi data selection, preprosesing, data mining, evaluation dan knowledge discovery.

2.1.1 Fungsi Data Mining

Menurut (Wahyudi Setiawan, 2021) data mining memiliki 2 fungsi yaitu fungsi deskripsi dan fungsi prediksi, adapun penjelasannya diantaranya yaitu:

a. Deskripsi

Fungsi Deskripsi adalah fungsi yang memberi pemahaman yang lebih dalam tentang data yang diamati. Saat menjalankan suatu proses diharapkan mengetahui bagaimana perilaku data. Data ini nantinya dapat digunakan untuk menentukan sifat-sifat data yang bersangkutan. Selanjutnya dapat menggunakan fitur penambangan data deskriptif untuk kemudian menemukan pola tertentu yang tersembunyi dalam data. Dengan kata lain, jika polanya berulang dan bernilai, dapat mengetahui sifat-sifat data.

b. Prediksi

Fungsi Prediksi adalah fungsi bagaimana suatu proses menemukan pola data tertentu. Pola-pola tersebut dapat dilihat dari berbagai variable yang ada dalam data. Setelah menemukan pola, dapat menggunakan pola yang diperoleh untuk memprediksi variabel lain yang nilai atau jenisnya tidak diketahui. Untuk alasan ini, fungsi yang satu ini juga disebut fungsi prediktif dan dilakukan sebagai analisis prediktif. Fungsi ini juga dapat digunakan untuk memprediksi variabel tertentu

yang tidak ada dalam data. Oleh karena itu, fitur ini memudahkan dan informatif bagi mereka yang membutuhkan prediksi akurat untuk meningkatkan hal-hal penting tersebut.

2.2 Algoritma K-Means

K-Means dapat membagi data berdasarkan jarak antar data pada kelompok yang telah ditetapkan. Algoritma ini bergantung pada fungsi untuk mengukur data yang mempunyai ciri khas sama. Jarak itu sendiri dihitung menggunakan fungsi euclidean. Kemudian data dimasukkan dalam kelompok yang mempunyai jarak terdekat. (Mochamad Wahyudi, et al, 2020) Langkah-langkah pengelompokan data adalah :

- a. Pilih jumlah klaster.
- b. Inisialisasi awal dan pusat klaster dilakukan secara random.
- c. Setiap data ditempatkan ke pusat klaster terdekat berdasarkan jarak antar obyek. Pada tahap ini jarak dihitung dengan menentukan kemiripan atau ketidakmiripan data dengan Metode Jarak Euclidean (Euclidean Distance) dengan rumus seperti di bawah ini (Witanto, Ratnawati dan Anam, 2019) :

$$d(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Dimana:

$D(x,y)$ = ukuran ketidakmiripan\

X_i = $(x_1, x_2, \dots, x_i)$ yaitu variabel data

Y_i = $(y_1, y_2, \dots, y_i)$ yaitu variabel pada titik pusat

- d. Hitung pusat klaster yang baru dengan keanggotaan yang baru dengan cara menghitung rata-rata obyek pada klaster. Penghitungan bisa juga dengan menggunakan median.
- e. Hitung kembali jarak tiap objek dengan pusat klaster yang baru, hingga klaster tidak berubah, maka proses pengklasteran selesai.

2.2.1 Kelebihan dan Kekurangan Algoritma K-means

a. Kelebihan K-means:

Algoritma K-means memiliki kelebihan yaitu:

- 1) Dalam implementasi menyelesaikan masalah, algoritma K-Means sangat simple serta fleksibel. Artinya perhitungan komputasinya tidak terlalu rumit dan algoritma ini dapat diimplementasikan pada segala bidang.
- 2) Algoritma K-Means sangat mudah untuk dipahami, terutama dalam implementasi data yang sangat besar serta dapat mengurangi kompleksitas data yang dimiliki (Bangoria et al., 2013)

b. Kekurangan K-means :

Algoritma K-means memiliki kekurangan yaitu:

- 1) User memerlukan angka yang tepat dalam menentukan jumlah cluster sebanyak k karena terkadang pusat cluster awal dapat berubah sehingga kejadian ini bisa mengakibatkan pengelompokan data menjadi tidak stabil (Joshi & Nalwade, 2013).
- 2) Algoritma K-Means tidak bisa maksimal dalam menentukan atau menginisialkan nilai centroid awalnya, karena pada pengelompokan data dengan algoritma K-Means sangat bergantung pada nilai centroidnya (Ahmed & Ashour, 2011).
- 3) Output dari K-Means tergantung pada nilai – nilai pusat yang dipilih pada clustering. Sehingga pada algoritma ini nilai awal titik pusat cluster menjadi dasar dalam penentuan cluster. Pemilihan centroid cluster awal secara acak akan memberikan pengaruh terhadap kinerja cluster tersebut (Singh & Kaur, 2013; Sujatha & Sona, 2013)

2.3 Clustering

Clustering adalah salah satu metode dalam data mining yang bertujuan untuk mengelompokkan sejumlah data ke dalam beberapa kelompok (cluster) berdasarkan kesamaan karakteristik (Tan, Steinbach & Kumar, 2020). Objek-objek yang berada dalam satu cluster memiliki tingkat kemiripan yang tinggi, sementara objek di cluster lain memiliki perbedaan yang signifikan.

Tujuan clustering (pengelompokan) data dapat dibedakan menjadi dua, yaitu pengelompokan untuk pemahaman dan clustering untuk penggunaan (Prasetyo,2012). Biasanya proses pengelompokan untuk tujuan pemahaman hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan seperti summarization(rata-rata,standar deviasi), pelabelan kelas untuk setiap kelompok sehingga dapat digunakan sebagai data training dalam klasifikasi supervised. Sementara jika untuk penggunaan, tujuan utama clustering biasanya adalah mencari prototipe kelompok yang paling representatif terhadap data, memberikan abstraksi dari setiap obyek data dalam kelompok dimana sebuah data terletak didalamnya.

Contoh tujuan clustering untuk pemahaman diantaranya: dibidang Biologi (pengelompokan berdasarkan karakter tertentu secara hirarkis) , pengelompokan gen yang memiliki fungsi sama. Dibidang information retrieval (web search),bidang klimatologi (pengelompokan pola tekanan udara yang berpengaruh pada cuaca),bidang bisnis (pengelompokan konsumen yang berpotensi untuk analisa dan strategi pemasaran).

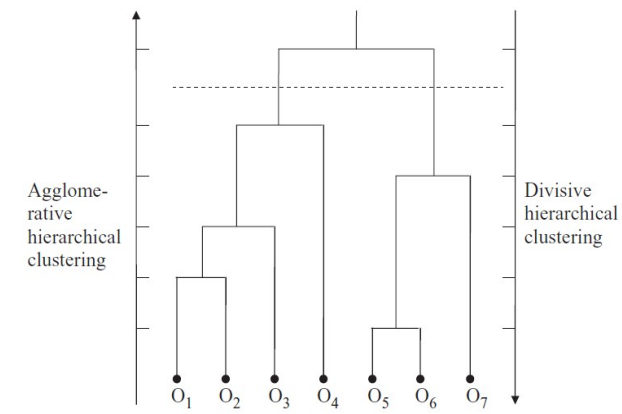
Contoh tujuan clustering untuk penggunaan dibidang summarization, dengan semakin besarnya jumlah data maka ongkos melakukan peringkasan semakin mahal (berat&kompleks), maka perlu diterapkan pengelompokan data untuk membuat prototipe yang dapat mewakili keseluruhan data yang akan digunakan. Kompresi , data yang terletak dalam satu cluster dapat dikompresi dengan diwakili oleh indeks prototipe yang dikaitkan dengan kelompok ,teknik kompresi ini dikenal sebagai quantization vector.

2.3.1 Jenis-jenis clustering (pengelompokan)

Clustering dapat dibedakan menurut struktur kelompok ,keanggotaan data dalam kelompok, dan kekompakan data dalam kelompok. Menurut struktur kelompok clustering dibagi menjadi dua yaitu hierarchical dan partitioning.

Hierarchical clustering adalah metode clustering yang mengelompokkan data dengan urutan partisi berkalang, metode ini dikelompokkan menjadi dua metode yaitu agglomerative dan divisive, metode agglomerative berawal dari obyek-obyek individual dimana pada awalnya banyaknya cluster sama dengan banyaknya obyek. Pertama-tama obyek-obyek yang paling mirip dikelompokkan, dan kelompokkelompok awal ini digabungkan sesuai dengan kemiripannya.

Akhirnya sewaktu kemiripan berkurang, semua subkelompok digabungkan menjadi satu cluster tunggal. Sementara Metode Hierarchical divisive merupakan proses kebalikan dari agglomerative, keduanya mengorganisasi data ke dalam struktur hirarki berbasis matrix proximity, hasil dari Hierarchical Clustering digambarkan dalam bentuk binary tree ataupun dendogram, root merupakan keseluruhan dataset dan tiap cabang merupakan data point, clustering akhir dapat diperoleh dari pemotongan dendogram pada level-level yang sesuai.



Gambar 2.1 Hierarchical clustering sumber (Xu & Wunsch, 2009)

Berbeda dengan klastering hirarki yang menghasilkan suatu tingkatan berurutan klaster dengan cara penggabungan secara iterative atau pemisahan, partitional clustering mengelompokkan datapoint kedalam k klaster tanpa struktur hirarki(Xu & Wunsch, 2009), metode ini membagi set data ke dalam sejumlah kelompok yang tidak saling overlap antara satu kelompok dengan kelompok lainnya, artinya setiap data hanya menjadi satu kelompok, termasuk dalam metode ini adalah K-Means dan DBSCAN.

Menurut keanggotaan data dalam kelompok, pengelompokan dibagi menjadi dua yaitu eksklusif dan tumpang tindih. Dalam kategori eksklusif sebuah data hanya menjadi anggota satu kelompok saja dan tidak bisa menjadi anggota kelompok lainnya. Metode yang termasuk kategori ini adalah K-Means dan DBSCAN, sedangkan yang masuk kategori overlap adalah metode clustering yang membolehkan sebuah data menjadi anggota di lebih dari satu kelompok, misalnya Fuzzy C-Means dan Hierarchical Clustering.

Sementara menurut kategori kekompakan, clustering terbagi menjadi dua yaitu komplet dan parsial. Jika semua data bisa bergabung menjadi satu (dalam konsep partitioning), bisa dikatakan semua data kompak menjadi satu kelompok. Namun jika ada satu atau beberapa data yang tidak ikut bergabung dalam kelompok mayoritas, data tersebut dikatakan memiliki perilaku menyimpang atau dikenal dengan istilah outlier/noise/"uninterested background". Beberapa metode yang dapat mendeteksi outlier ini diantaranya adalah DBSCAN dan K-Means (dengan sejumlah komputasi tambahan).

Terdapat dua pendekatan utama dalam clustering, yaitu:

- a. Hierarchical Clustering, di mana data dikelompokkan secara bertingkat dari yang paling mirip hingga terbentuk pohon hierarki.
- b. Partitioning Clustering, di mana data dibagi ke dalam sejumlah cluster tertentu yang telah ditentukan sebelumnya, seperti pada algoritma K-Means.

2.4 Data dan Alat Penelitian

Berikut ini adalah data dan alat penelitian yang digunakan dalam membuat web data mining dalam k-means clustering data penempatan karyawan di PT. Guna Cipta Prima

Data Penelitian yang dibutuhkan :

2.4.1. Data karyawan

Data karyawan merupakan data utama yang digunakan dalam penelitian ini. Data ini berisi informasi terkait identitas dan karakteristik karyawan, seperti nama, usia, pendidikan, pengalaman kerja, dan kompetensi yang dimiliki. Data tersebut digunakan sebagai objek dalam proses pengelompokan (clustering) untuk menentukan penempatan yang sesuai.

2.4.2. Data variabel yang diberikan pihak perusahaan

Data variabel merupakan atribut atau parameter yang digunakan sebagai dasar dalam proses clustering. Variabel ini ditentukan oleh pihak perusahaan sesuai dengan kebutuhan penempatan karyawan, seperti tingkat pendidikan, lama bekerja, keahlian, serta kinerja. Variabel-variabel tersebut akan diolah menggunakan metode K-Means untuk menghasilkan kelompok data yang memiliki kemiripan.

2.4.3. Data lokasi penempatan karyawan

Data lokasi penempatan berisi informasi mengenai unit kerja, divisi, atau lokasi kerja yang tersedia di perusahaan. Data ini digunakan untuk mencocokkan hasil clustering dengan kebutuhan penempatan, sehingga setiap kelompok karyawan dapat ditempatkan pada posisi atau lokasi yang sesuai dengan karakteristiknya.

Selain data utama, penelitian ini juga menggunakan data pendukung dalam proses pengembangan sistem berbasis web. Data tersebut meliputi kebutuhan sistem, struktur database, serta desain antarmuka (*user interface*). Data ini digunakan untuk memastikan bahwa sistem yang dibangun dapat berjalan dengan baik dan sesuai dengan kebutuhan pengguna.

2.4.4. Alat Penelitian

Alat yang digunakan dalam penelitian ini berupa perangkat keras dan perangkat lunak. Perangkat keras yang digunakan adalah sebuah laptop dengan spesifikasi sebagai berikut:

- a. Sistem Operasi: Windows 10 Pro
- b. Prosesor: Intel Core i3 dengan kecepatan 2.0 GHz
- c. RAM: 4 GB
- d. Sistem Tipe: 64-bit

Laptop tersebut digunakan untuk melakukan proses pengolahan data, implementasi algoritma K-Means, serta pengembangan dan pengujian sistem berbasis web.

2.5 Web pengembangan sistem

Web yang dibutuhkan / digunakan dalam pengembangan sistem untuk menganalisis data mining dengan menggunakan metode K-means clustering.

2.5.1 PHP (Hypertext Preprocessor)

Menurut (Sholichin, 2020) PHP adalah bahasa scripting yang berfokus pada pengembangan web, yang memungkinkan integrasi dengan HTML untuk menciptakan aplikasi yang responsif. Dengan kemampuannya untuk berinteraksi dengan database, PHP menjadi pilihan utama bagi banyak pengembang dalam menciptakan aplikasi berbasis web.

Budi Santoso (2023), PHP juga dikenal karena kemampuannya untuk terintegrasi dengan berbagai database, seperti MySQL dan PostgreSQL. Ini menjadikannya pilihan yang ideal untuk pengembangan aplikasi yang memerlukan pengelolaan data secara real-time, serta memberikan fleksibilitas dalam pengembangan aplikasi berbasis web yang kompleks.

Beberapa kelebihan PHP menurut (Alan Nur Aditya, 2011)

- a. Bahasa pemrograman PHP adalah sebuah bahasa script yang tidak melakukan sebuah kompilasi dalam penggunaannya.
- b. Web server yang mendukung PHP dapat ditemukan dimana-mana dari mulai apache, IIS, Lighttpd, hingga Xitami konfigurasi yang relatif mudah.
- c. Dalam sisi pengembangan lebih mudah, karena banyaknya developer yang siap membantu dalam pengembangan.
- d. Dalam sisi pemahaman, PHP adalah bahasa scripting yang paling mudah karena memiliki referensi yang banyak.
- e. PHP adalah bahasa open source yang dapat digunakan diberbagai mesin (Linux, Unix, Macintosh, Windows) dan dapat dijalankan secara runtime.

2.5.2 MySQL (My Structured Query Language)

MySQL adalah DBMS yang open source dengan dua bentuk lisensi, yaitu Free Software (perangkat lunak bebas) dan Shareware (perangkat lunak berpemilik yang penggunaannya terbatas). Jadi MySQL adalah database server yang gratis dengan lisensi GNU General Public License (GPL) sehingga dapat dipakai untuk keperluan pribadi atau komersial tanpa harus membayar lisensi yang ada. MySQL masuk dalam jenis RDBMS (Relational database Management) maka dari itu, istilah semacam baris, kolom, tabel, dipakai pada MySQL (Rahimi Fitri, 2020:2).

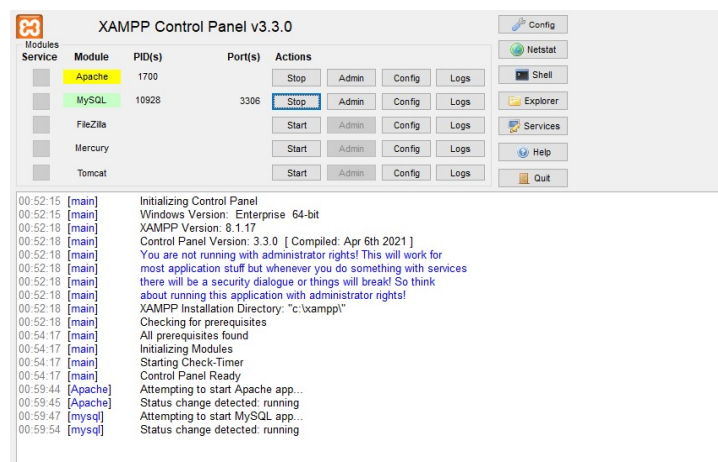
MySQL merupakan database server yang memiliki konsep database modern MySQL mempunyai beberapa kelebihan, antara lain :

- a. Free(bebasdiunduh)
- b. Stabil dan tangguh
- c. Fleksibel dengan berbagai pemrograman
- d. Security yang baik
- e. Dukungan dari banyak komunitas
- f. Kemudahan management database

g. Perkembangan software yang cukup tepat.

2.5.3 XAMPP

XAMPP adalah sebuah paket kumpulan software yang terdiri dari apache, mysql, phpmyadmin, php perl, Freetype, dll. XAMPP berfungsi untuk memudahkan instalasi lingkungan php, dimana biasanya lingkungan pengembangan web memerlukan php, apache, mysql dan phpmyadmin serta software-software yang terkait dengan pengembangan web. Dengan menggunakan xampp, kita tidak perlu lagi menginstal aplikasi-aplikasi tersebut satu persatu (M.Aswadi 2015). Berikut ini tampilan control panel pada aplikasi XAMPP yang tersedia :



Gambar 2. 2 Tampilan XAMPP

Pada umumnya XAMPP dibagi menjadi tiga bagian yaitu sebagai berikut dijelaskan:

- Htdocs, yaitu untuk tempat berkas yang dijalankan, misalnya berkas PHP, HTML, dan skrip lainnya.
- PhpMyAdmin, untuk membuka halaman PhpMAdmin dengan mengetik alamat <http://localhost/phpMyAdmin>.
- Control Panel, untuk mengelola layanan XAMPP, Seperti memulai dan megehentikan layanan.

2.6 Penelitian yang Relevan

Dalam menyusun penelitian ini dapat dikaitkan dengan beberapa penelitian sebelumnya. Berikut ini adalah beberapa penelitian yang relevan dengan penelitian ini :

Tabel 2. 1 Penelitian yang relevan

No	Judul	Penulis	Tahun	Kesimpulan
1	Customer Data Using the K-Means Clustering Method at PT. Telkom Akses	Egi Affandi, S Syam, Yohanni Syahra, A Amrullah	2021	Hasil penelitian ini, penulis merancang dan mengembangkan sistem informasi berbasis desktop yang dapat menerapkan metode pengelompokan untuk menghasilkan klaster data pelanggan sesuai dengan kebutuhan PT. Telkom Akses.
2	Clustering Analysis of Poverty Levels in North Sumatra Province Using the Application of Data Mining with the K-Means Algorithm	Widyastuti Andriyani, Asyabri Hadi Nasyuha, Yohanni Syahra, Bagus Triaji	2023	penelitian tentang tingkat kemiskinan di provinsi ini menjadi sangat relevan dan mendesak. Maka, diperlukan analisis yang lebih mendalam mengenai tingkat kemiskinan di berbagai wilayah di provinsi ini menggunakan metode penambangan data.
3	Implementasi Data Mining Untuk Pengelompokan Data Pengunjung Museum Menggunakan Algoritma K-Means Clustering Pada Museum Deli Serdang	Natasah Liaza, Yohanni Syahra, Ita Mariami	2020	Algoritma K-Means Clustering melakukan penilaian terhadap pengelompokan jumlah kunjungan pada Museum daerah Deli Serdang dapat diterapkan dengan metode K-Means Clustering. Data diolah dengan melakukan K-Means yang dikelompokkan menjadi 3 cluster yaitu cluster sedang, cluster tinggi dan cluster rendah.

BAB III

METODOLOGI PENELITIAN

3.1 Jenis dan Pendekatan 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.

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.

3.2 Lokasi, Waktu, dan Subjek Penelitian

3.2.1 Lokasi Penelitian

Penelitian ini dilaksanakan di PT. Guna Cipta Prima yang beralamat di Jalan Sumber Utama No. 23 LK XIII, Medan. PT. Guna Cipta Prima merupakan perusahaan penyedia jasa keamanan (security) dan kebersihan (cleaning service) yang memiliki 28 lokasi penempatan karyawan tersebar di berbagai wilayah Kota Medan dan sekitarnya.

3.2.2 Waktu Penelitian

Penelitian ini dilaksanakan pada semester genap tahun akademik 2025/2026, menyesuaikan jadwal penyusunan skripsi di Universitas

Muhammadiyah Sumatera Utara (UMSU). Secara rinci, tahapan penelitian dilaksanakan sebagai berikut:

Tabel 3.1 Jadwal Penelitian

No	Tahapan Penelitian	Bulan (2025-2026)
1	Pengumpulan Data (Observasi & Wawancara)	September - Oktober 2025
2	Studi Literatur & Kajian Teori	Oktober 2025
3	Preprocessing & Analisis Manual K-Means	Maret 2026
4	Perancangan Sistem (Desain DB & UI)	Maret 2026
5	Implementasi Sistem PHP 8.1 + MySQLi	Maret - April 2026
6	Pengujian Sistem (Black-Box Testing)	April 2026
7	Penulisan & Revisi Laporan Skripsi	April 2026

3.2.3 Subjek dan Objek Penelitian

Subjek Penelitian: Data penempatan karyawan (Security dan Cleaning Service) di PT. Guna Cipta Prima yang berjumlah 28 lokasi penempatan aktif pada periode tahun 2025.

Objek Penelitian: Kinerja algoritma K-Means Clustering dalam mengelompokkan lokasi penempatan karyawan berdasarkan jumlah tenaga Security dan Cleaning Service (CS) pada setiap lokasi, serta kualitas sistem berbasis web yang dibangun untuk menyajikan hasil analisis tersebut.

3.3 Data dan Sumber Data

3.3.1 Jenis Data

Jenis data yang digunakan dalam penelitian ini adalah:

- 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.

- b. 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.

3.3.2 Sumber Data

Sumber data dalam penelitian ini diperoleh dari:

- a. PT. Guna Cipta Prima – sebagai sumber data utama berupa data penempatan karyawan aktual.
- b. Bagian Admin Perusahaan – menyediakan dokumen rekap data karyawan Security dan Cleaning Service per lokasi.
- c. Arsip dan dokumen internal perusahaan tahun 2025.

3.3.3 Variabel / Atribut Penelitian

Atribut yang digunakan sebagai input dalam proses clustering adalah sebagai berikut:

Tabel 3.2 Variabel/Atribut Penelitian

No	Nama Atribut	Tipe Data	Keterangan
1	Lokasi Penempatan	Kategorikal	Nama/identitas lokasi kerja (digunakan sebagai label, bukan variabel komputasi)
2	Jumlah Security (X1)	Numerik (Integer)	Jumlah tenaga Security yang ditempatkan di lokasi tersebut
3	Jumlah CS (X2)	Numerik (Integer)	Jumlah tenaga Cleaning Service yang ditempatkan di lokasi tersebut

Variabel Lokasi Penempatan tidak diikutsertakan dalam perhitungan jarak Euclidean, namun digunakan sebagai label identifikasi setiap data point. Variabel X1 (Security) dan X2 (CS) merupakan variabel numerik yang menjadi input utama algoritma K-Means setelah melalui proses normalisasi Min-Max Scaling.

3.4 Teknik Pengumpulan Data

3.4.1 Studi Literatur

Penulis mengumpulkan referensi teori dari buku teks, jurnal ilmiah, dan artikel akademik yang berkaitan dengan data mining, algoritma K-Means Clustering, normalisasi data, evaluasi clustering (Silhouette Coefficient dan SSE), serta pengembangan sistem informasi berbasis web menggunakan PHP dan MySQL/MySQLi.

3.4.2 Observasi

Penulis melakukan observasi langsung ke PT. Guna Cipta Prima untuk memahami proses bisnis penempatan karyawan, kondisi lapangan, dan karakteristik data yang tersedia. Dari observasi ini, diperoleh pemahaman bahwa perusahaan memiliki 28 lokasi aktif dengan jumlah tenaga Security dan CS yang bervariasi di setiap lokasi.

3.4.3 Wawancara

Wawancara dilakukan dengan bagian admin PT. Guna Cipta Prima guna mendapatkan informasi terkait: (1) daftar lengkap lokasi penempatan aktif; (2) jumlah karyawan Security dan CS per lokasi; (3) kebutuhan informasi manajemen dalam pengambilan keputusan penempatan; serta (4) kendala yang dihadapi dalam proses penempatan karyawan yang selama ini dilakukan secara manual dan intuitif.

3.5 Teknik Pengolahan Data

Sebelum data diproses oleh algoritma K-Means, dilakukan serangkaian tahapan pengolahan data yang bertujuan untuk memastikan kualitas dan kesiapan data input. Tahapan tersebut meliputi Data Cleaning, Data Transformation, dan Normalisasi Data.

3.5.1 Data Cleaning

Proses data cleaning dilakukan untuk memastikan data yang digunakan bebas dari gangguan yang dapat mempengaruhi hasil clustering. Langkah-langkah yang dilakukan meliputi:

- a. Penghapusan Data Duplikat: Memeriksa dan menghapus baris data yang memiliki nilai identik untuk mencegah redudansi.
- b. Penanganan Missing Value: Pada dataset penelitian ini tidak ditemukan nilai kosong (missing value). Namun, apabila terdapat data yang tidak lengkap,

penanganan dilakukan dengan metode imputasi nilai rata-rata (mean imputation).

- c. Penanganan Data Anomali (Outlier): Data dengan nilai ekstrem yang tidak wajar akan diidentifikasi dan dikonsultasikan dengan pihak perusahaan sebelum diputuskan untuk dipertahankan atau dihapus.
- d. Standarisasi Format: Nama lokasi diseragamkan formatnya untuk konsistensi tampilan pada antarmuka sistem.

3.5.2 Data Transformation

Data transformation dilakukan untuk mengubah format data agar sesuai dengan kebutuhan algoritma K-Means, yang hanya dapat memproses data numerik.

Langkah yang dilakukan adalah:

- a. Konversi Data Kategorikal: Atribut Lokasi Penempatan yang bersifat kategorikal tidak diikutsertakan dalam perhitungan jarak, melainkan hanya digunakan sebagai label (identifier) setiap data point.
- b. Seleksi Fitur (Feature Selection): Dari seluruh atribut yang tersedia, dipilih dua atribut numerik utama yaitu Jumlah Security (X1) dan Jumlah CS (X2) sebagai variabel komputasi.

3.5.3 Normalisasi Data (Min-Max Scaling)

Normalisasi data dilakukan menggunakan metode Min-Max Scaling untuk menyeragamkan skala nilai atribut ke dalam rentang [0, 1]. Hal ini penting untuk mencegah dominasi atribut dengan rentang nilai yang lebih besar terhadap hasil perhitungan jarak Euclidean. Rumus Min-Max Scaling adalah sebagai berikut:

$$X' = (X - X_{\min}) / (X_{\max} - X_{\min})$$

Keterangan:

X' = Nilai atribut setelah normalisasi

X = Nilai atribut asli (sebelum normalisasi)

$X_{\min}$ = Nilai minimum atribut dalam seluruh dataset

$X_{\max}$ = Nilai maksimum atribut dalam seluruh dataset

Berdasarkan dataset penelitian ini, diperoleh nilai minimum dan maksimum sebagai berikut: X1 (Security): $X_{\min} = 1$, $X_{\max} = 7$; X2 (CS): $X_{\min} = 1$, $X_{\max} = 13$. Sehingga rumus normalisasi untuk setiap atribut adalah:

$$X1' = (X1 - 1) / (7 - 1) = (X1 - 1) / 6$$

$$X2' = (X2 - 1) / (13 - 1) = (X2 - 1) / 12$$

3.6 Analisis Manual K-Means Clustering

Bagian ini menyajikan proses analisis manual algoritma K-Means Clustering secara lengkap, mencakup penyajian dataset, normalisasi, penentuan jumlah cluster, iterasi perhitungan jarak, pembaruan centroid, hingga evaluasi kualitas cluster. Seluruh perhitungan dilakukan menggunakan data nyata dari PT. Guna Cipta Prima.

3.6.1 Dataset Penelitian

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). Berikut adalah dataset lengkap penelitian:

Tabel 3.3 Dataset Penelitian - Data Penempatan Karyawan PT. Guna Cipta Prima Tahun 2025

No	Lokasi Penempatan	Security (X1)	CS (X2)	Keterangan
1	Clover Jl. T Umar	5	13	Kebutuhan Sangat Tinggi
2	CLOVER Jl. T. Umar Simp. Airlangga	1	10	Kebutuhan Tinggi
3	Clover Jl Krakatau	6	4	Kebutuhan Tinggi
4	Clover Jl. Katamso (SUZUYA)	2	4	Kebutuhan Sedang
5	Clover Jl. Titi Papan (Gatot Subroto)	5	2	Kebutuhan Sedang
6	Clover Jl. Setia Budi	2	6	Kebutuhan Tinggi
7	Toko Sony	2	3	Kebutuhan Sedang
8	Kantor Malibu	5	2	Kebutuhan Sedang
9	Mega Central Autoniaga (Chevrolet)	3	4	Kebutuhan Sedang

10	Chevrolet Krakatau	2	13	Kebutuhan Sangat Tinggi
11	Gudang Arista Ngumban Surbakti	3	1	Kebutuhan Rendah
12	Gudang Arista Auto Lestari	1	4	Kebutuhan Sedang
13	Gudang Arista Sedap Malam	3	5	Kebutuhan Tinggi
14	Hyundai Gudang + Showroom	2	5	Kebutuhan Sedang
15	Sutan Kasim	1	4	Kebutuhan Sedang
16	Showroom BYD Amir Hamzah	3	2	Kebutuhan Sedang
17	Showroom BYD Cemara	3	2	Kebutuhan Sedang
18	Otto Link	2	6	Kebutuhan Tinggi
19	Awal Bros Medan	4	3	Kebutuhan Sedang
20	PT. LINTAS SAMUDERA JAYA (LSJ)	5	1	Kebutuhan Sedang
21	KLINIK ESTETIKA	3	2	Kebutuhan Sedang
22	Miracle Aesthetic Clinic Medan	3	6	Kebutuhan Tinggi
23	Toko Roti Jawara	4	3	Kebutuhan Sedang
24	Jawara Delipark	1	2	Kebutuhan Rendah
25	Jawara Jl. Dr. Cipto	3	1	Kebutuhan Rendah
26	SPBU Krakatau	7	5	Kebutuhan Tinggi
27	SPBU MEDAN JOHOR	3	3	Kebutuhan Sedang
28	SPBU Jl. Yos Sudarso	3	4	Kebutuhan Sedang

3.6.2 Proses Normalisasi Data

Setiap nilai atribut pada dataset dinormalisasi menggunakan rumus Min-Max Scaling yang telah ditetapkan. Hasil normalisasi seluruh 28 data lokasi disajikan pada tabel berikut:

Tabel 3.4 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

3.6.3 Penentuan Jumlah Cluster (k) dengan Metode Elbow

Penentuan jumlah cluster optimal dilakukan menggunakan Metode Elbow, yaitu dengan menjalankan algoritma K-Means untuk nilai $k = 1$ hingga $k = 7$, kemudian mengamati penurunan nilai SSE (*Sum of Squared Errors*) atau WCSS (*Within-Cluster Sum of Squares*). Titik "siku" (*elbow point*) pada grafik SSE menandakan jumlah cluster yang optimal.

Nilai SSE dihitung menggunakan rumus:

$$SSE = \sum \sum ||x_i - c_j||^2$$

di mana x_i adalah data point ke- i dan c_j adalah centroid cluster ke- j yang menjadi anggota data tersebut.

Tabel 3.5 Nilai SSE per Jumlah Cluster (Metode Elbow)

Jumlah Cluster (k)	Nilai SSE (WCSS)	Interpretasi
1	3.6419	Seluruh data dalam satu kelompok, SSE sangat besar
2	2.2352	Dua kelompok, penurunan SSE signifikan
3	1.1563	Tiga kelompok *** TITIK ELBOW OPTIMAL *** penurunan mulai melambat
4	0.8144	Empat kelompok, penurunan SSE tidak signifikan lagi

5	0.5771	Lima kelompok, kompleksitas bertambah tanpa gain signifikan
6	0.4236	Enam kelompok, cluster mulai terlalu spesifik
7	0.2719	Tujuh kelompok, over-clustering

Berdasarkan Tabel 3.5, dapat diamati bahwa pada $k = 1$, nilai SSE adalah 3,6419 (sangat tinggi). Pada $k = 2$, SSE turun menjadi 2,2352 (penurunan 1,4067). Pada $k = 3$, SSE turun menjadi 1,1563 (penurunan 1,0789). Namun pada $k = 4$, penurunan SSE hanya 0,3419 — jauh lebih kecil dibandingkan penurunan sebelumnya. Titik "siku" (elbow) terjadi pada $k = 3$, sehingga jumlah cluster yang digunakan dalam penelitian ini adalah $k = 3$, yang merepresentasikan: Cluster 1 (Kebutuhan Tinggi), Cluster 2 (Kebutuhan Sedang), dan Cluster 3 (Kebutuhan Rendah).

3.6.4 Penentuan Centroid Awal

Pada algoritma K-Means, centroid awal ditentukan secara acak dari data yang tersedia. Dalam penelitian ini, centroid awal dipilih dari tiga data yang tersebar representatif dalam dataset, yaitu:

Tabel 3.6 Centroid Awal Iterasi 1

Centroid	Data Asal	Lokasi Penempatan	X1' (Sec)	X2' (CS)
C1 (Tinggi)	Data ke-1	Clover Jl. T Umar	0.6667	1.0000
C2 (Sedang)	Data ke-9	Mega Central Autoniaga (Chevrolet)	0.3333	0.2500
C3 (Rendah)	Data ke-24	Jawara Delipark	0.0000	0.0833

Penentuan centroid awal di atas merepresentasikan tiga kategori yang diharapkan: Data 1 memiliki nilai CS tertinggi (13) sehingga representatif untuk Cluster Kebutuhan Tinggi; Data 9 memiliki nilai tengah yang seimbang (Sec=3, CS=4) representatif untuk Cluster Sedang; dan Data 24 memiliki nilai rendah (Sec=1, CS=2) representatif untuk Cluster Kebutuhan Rendah.

3.6.5 Iterasi 1 – Perhitungan Jarak Euclidean dan Pembentukan Cluster

Pada Iterasi 1, jarak Euclidean setiap data point ke setiap centroid dihitung menggunakan rumus:

$$d(x_i, C_j) = \sqrt{[(X1'_i - X1'_Cj)^2 + (X2'_i - X2'_Cj)^2]}$$

Contoh perhitungan untuk Data ke-1 (Clover Jl. T Umar, $X1'=0.6667$, $X2'=1.0000$):

$$d(\text{Data1}, C1) = \sqrt{[(0.6667-0.6667)^2 + (1.0000-1.0000)^2]} = \sqrt{[0+0]} = 0.0000$$

$$\begin{aligned} d(\text{Data1}, C2) &= \sqrt{[(0.6667-0.3333)^2 + (1.0000-0.2500)^2]} = \sqrt{[0.1111+0.5625]} \\ &= \sqrt{0.6736} = 0.8207 \end{aligned}$$

$$\begin{aligned} d(\text{Data1}, C3) &= \sqrt{[(0.6667-0.0000)^2 + (1.0000-0.0833)^2]} = \sqrt{[0.4444+0.8403]} \\ &= \sqrt{1.2847} = 1.1335 \end{aligned}$$

Karena $d(\text{Data1}, C1) = 0.0000$ adalah jarak terkecil, maka Data ke-1 masuk ke Cluster C1.

Berikut hasil perhitungan jarak Euclidean untuk seluruh 28 data pada Iterasi 1, yaitu:

Tabel 3.7 Hasil Perhitungan Jarak Euclidean Iterasi 1

No	Lokasi Penempatan	X1'	X2'	d(C1)	d(C2)	d(C3)	Cluster
1	Clover Jl. T Umar	0.6667	1.0000	0.0000	0.8207	1.1335	C1
2	CLOVER Jl. T. Umar Simp. Airlangga	0.0000	0.7500	0.7120	0.6009	0.6667	C2
3	Clover Jl Krakatau	0.8333	0.2500	0.7683	0.5000	0.8498	C2
4	Clover Jl. Katamso (SUZUYA)	0.1667	0.2500	0.9014	0.1667	0.2357	C2
5	Clover Jl. Titi Papan (Gatot Subroto)	0.6667	0.0833	0.9167	0.3727	0.6667	C2
6	Clover Jl. Setia Budi	0.1667	0.4167	0.7683	0.2357	0.3727	C2
7	Toko Sony	0.1667	0.1667	0.9718	0.1863	0.1863	C2

8	Kantor Malibu	0.6667	0.0833	0.9167	0.3727	0.6667	C2
9	Mega Central Autoniaga (Chevrolet)	0.3333	0.2500	0.8207	0.0000	0.3727	C2
10	Chevrolet Krakatau	0.1667	1.0000	0.5000	0.7683	0.9317	C1
11	Gudang Arista Ngumban Surbakti	0.3333	0.0000	1.0541	0.2500	0.3436	C2
12	Gudang Arista Auto Lestari	0.0000	0.2500	1.0035	0.3333	0.1667	C3
13	Gudang Arista Sedap Malam	0.3333	0.3333	0.7454	0.0833	0.4167	C2
14	Hyundai Gudang + Showroom	0.1667	0.3333	0.8333	0.1863	0.3005	C2
15	Sutan Kasim	0.0000	0.2500	1.0035	0.3333	0.1667	C3
16	Showroom BYD Amir Hamzah	0.3333	0.0833	0.9754	0.1667	0.3333	C2
17	Showroom BYD Cemara	0.3333	0.0833	0.9754	0.1667	0.3333	C2
18	Otto Link	0.1667	0.4167	0.7683	0.2357	0.3727	C2
19	Awal Bros Medan	0.5000	0.1667	0.8498	0.1863	0.5069	C2
20	PT. LINTAS SAMUDERA JAYA (LSJ)	0.6667	0.0000	1.0000	0.4167	0.6719	C2
21	KLINIK ESTETIKA	0.3333	0.0833	0.9754	0.1667	0.3333	C2
22	Miracle Aesthetic Clinic Medan	0.3333	0.4167	0.6719	0.1667	0.4714	C2
23	Toko Roti Jawara	0.5000	0.1667	0.8498	0.1863	0.5069	C2
24	Jawara Delipark	0.0000	0.0833	1.1335	0.3727	0.0000	C3
25	Jawara Jl. Dr. Cipto	0.3333	0.0000	1.0541	0.2500	0.3436	C2
26	SPBU Krakatau	1.0000	0.3333	0.7454	0.6719	1.0308	C2

27	SPBU MEDAN JOHOR	0.3333	0.1667	0.8975	0.0833	0.3436	C2
28	SPBU Jl. Yos Sudarso	0.3333	0.2500	0.8207	0.0000	0.3727	C2

Hasil pengelompokan Iterasi 1: Cluster C1 (Kebutuhan Tinggi): 2 lokasi (Data: 1, 10); Cluster C2 (Kebutuhan Sedang): 23 lokasi (Data: 2, 3, 4, 5, 6, 7, 8, 9, 11, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, 25, 26, 27, 28); Cluster C3 (Kebutuhan Rendah): 3 lokasi (Data: 12, 15, 24).

3.6.6 Pembaruan Centroid Setelah Iterasi 1

Centroid baru dihitung sebagai rata-rata (mean) dari seluruh anggota setiap cluster menggunakan rumus:

$$C_j_baru = (1/n_j) * \sum x_i, \text{ untuk setiap } x_i \text{ yang berada dalam Cluster } j$$

Contoh Perhitungan Centroid Baru C1 (dari 2 anggota: Data 1 dan Data 10):

$$X1'_C1_baru = (0.6667 + 0.1667) / 2 = 0.4167$$

$$X2'_C1_baru = (1.0000 + 1.0000) / 2 = 1.0000$$

Tabel 3.8 Centroid Baru Setelah Iterasi 1

Centroid	Anggota (No. Data)	Jumlah Anggota	X1' Rata-rata	X2' Rata-rata	Perubahan
C1 (Tinggi)	1, 10	2	0.4167	1.0000	Berubah
C2 (Sedang)	2, 3, 4, 5, 6, 7, 8, 9, 11, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, 25, 26, 27, 28	23	0.3913	0.2210	Berubah
C3 (Rendah)	12, 15, 24	3	0.0000	0.1944	Berubah

3.6.7 Iterasi 2 – Perhitungan Jarak Euclidean dan Cluster Baru

Pada Iterasi 2, perhitungan jarak Euclidean diulang menggunakan centroid baru: $C1=(0.4167, 1.0000)$, $C2=(0.3913, 0.2210)$, $C3=(0.0000, 0.1944)$. Berikut hasil lengkap Iterasi 2:

Tabel 3.9 Hasil Perhitungan Jarak Euclidean Iterasi 2 (* = perpindahan cluster)

No	Lokasi Penempatan	X1'	X2'	d(C1)	d(C2)	d(C3)	Cluster
1	Clover Jl. T Umar	0.6667	1.0000	0.2500	0.8262	1.0456	C1
2	CLOVER Jl. T. Umar Simp. Airlangga	0.0000	0.7500	0.4859	0.6580	0.5556	C1*
3	Clover Jl Krakatau	0.8333	0.2500	0.8580	0.4430	0.8352	C2
4	Clover Jl. Katamso (SUZUYA)	0.1667	0.2500	0.7906	0.2265	0.1757	C3*
5	Clover Jl. Titi Papan (Gatot Subroto)	0.6667	0.0833	0.9501	0.3079	0.6759	C2
6	Clover Jl. Setia Budi	0.1667	0.4167	0.6346	0.2979	0.2778	C3*
7	Toko Sony	0.1667	0.1667	0.8700	0.2311	0.1690	C3*
8	Kantor Malibu	0.6667	0.0833	0.9501	0.3079	0.6759	C2
9	Mega Central Autoniaga (Chevrolet)	0.3333	0.2500	0.7546	0.0648	0.3379	C2
10	Chevrolet Krakatau	0.1667	1.0000	0.2500	0.8107	0.8226	C1
11	Gudang Arista Ngumban Surbakti	0.3333	0.0000	1.0035	0.2285	0.3859	C2
12	Gudang Arista Auto Lestari	0.0000	0.2500	0.8580	0.3924	0.0556	C3
13	Gudang Arista Sedap Malam	0.3333	0.3333	0.6719	0.1264	0.3611	C2

14	Hyundai Gudang + Showroom	0.1667	0.3333	0.7120	0.2512	0.2170	C3*
15	Sutan Kasim	0.0000	0.2500	0.8580	0.3924	0.0556	C3
16	Showroom BYD Amir Hamzah	0.3333	0.0833	0.9204	0.1494	0.3514	C2
17	Showroom BYD Cemara	0.3333	0.0833	0.9204	0.1494	0.3514	C2
18	Otto Link	0.1667	0.4167	0.6346	0.2979	0.2778	C3*
19	Awal Bros Medan	0.5000	0.1667	0.8375	0.1215	0.5008	C2
20	PT. LINTAS SAMUDERA JAYA (LSJ)	0.6667	0.0000	1.0308	0.3531	0.6944	C2
21	KLINIK ESTETIKA	0.3333	0.0833	0.9204	0.1494	0.3514	C2
22	Miracle Aesthetic Clinic Medan	0.3333	0.4167	0.5893	0.2041	0.4006	C2
23	Toko Roti Jawara	0.5000	0.1667	0.8375	0.1215	0.5008	C2
24	Jawara Delipark	0.0000	0.0833	1.0069	0.4148	0.1111	C3
25	Jawara Jl. Dr. Cipto	0.3333	0.0000	1.0035	0.2285	0.3859	C2
26	SPBU Krakatau	1.0000	0.3333	0.8858	0.6190	1.0096	C2
27	SPBU MEDAN JOHOR	0.3333	0.1667	0.8375	0.0795	0.3345	C2
28	SPBU Jl. Yos Sudarso	0.3333	0.2500	0.7546	0.0648	0.3379	C2

Pada Iterasi 2 terdapat 6 data yang berpindah cluster (ditandai dengan tanda *). Hal ini berarti centroid perlu diperbarui kembali untuk Iterasi 3.

Tabel 3.10 Centroid Baru Setelah Iterasi 2

Centroid	Anggota (No. Data)	Jumlah Anggota	X1' Rata-rata	X2' Rata-rata	Perubahan
-----------------	---------------------------	-----------------------	----------------------	----------------------	------------------

C1 (Tinggi)	1, 2, 10	3	0.2778	0.9167	Berubah
C2 (Sedang)	3, 5, 8, 9, 11, 13, 16, 17, 19, 20, 21, 22, 23, 25, 26, 27, 28	17	0.4804	0.1618	Berubah
C3 (Rendah)	4, 6, 7, 12, 14, 15, 18, 24	8	0.1042	0.2708	Berubah

3.6.8 Iterasi 3 dan Konvergensi

Iterasi 3 dijalankan menggunakan centroid hasil pembaruan Iterasi 2: C1=(0.2778, 0.9167), C2=(0.4804, 0.1618), C3=(0.1042, 0.2708). Hasil Iterasi 3:

Tabel 3.11 Hasil Perhitungan Jarak Euclidean Iterasi 3

No	Lokasi Penempatan	X1'	X2'	d(C1)	d(C2)	d(C3)	Cluster
1	Clover Jl. T Umar	0.6667	1.0000	0.3977	0.8587	0.9209	C1
2	CLOVER Jl. T. Umar Simp. Airlangga	0.0000	0.7500	0.3239	0.7595	0.4904	C1
3	Clover Jl Krakatau	0.8333	0.2500	0.8678	0.3638	0.7295	C2
4	Clover Jl. Katamso (SUZUYA)	0.1667	0.2500	0.6759	0.3259	0.0659	C3
5	Clover Jl. Titi Papan (Gatot Subroto)	0.6667	0.0833	0.9196	0.2021	0.5929	C2
6	Clover Jl. Setia Budi	0.1667	0.4167	0.5122	0.4042	0.1587	C3
7	Toko Sony	0.1667	0.1667	0.7582	0.3138	0.1215	C3
8	Kantor Malibu	0.6667	0.0833	0.9196	0.2021	0.5929	C2
9	Mega Central Autoniaga (Chevrolet)	0.3333	0.2500	0.6690	0.1715	0.2301	C2

10	Chevrolet Krakatau	0.1667	1.0000	0.1389	0.8950	0.7318	C1
11	Gudang Arista Ngumban Surbakti	0.3333	0.0000	0.9183	0.2186	0.3548	C2
12	Gudang Arista Auto Lestari	0.0000	0.2500	0.7222	0.4884	0.1062	C3
13	Gudang Arista Sedap Malam	0.3333	0.3333	0.5860	0.2260	0.2375	C2
14	Hyundai Gudang + Showroom	0.1667	0.3333	0.5938	0.3576	0.0884	C3
15	Sutan Kasim	0.0000	0.2500	0.7222	0.4884	0.1062	C3
16	Showroom BYD Amir Hamzah	0.3333	0.0833	0.8352	0.1667	0.2961	C2
17	Showroom BYD Cemara	0.3333	0.0833	0.8352	0.1667	0.2961	C2
18	Otto Link	0.1667	0.4167	0.5122	0.4042	0.1587	C3
19	Awal Bros Medan	0.5000	0.1667	0.7822	0.0202	0.4093	C2
20	PT. LINTAS SAMUDERA JAYA (LSJ)	0.6667	0.0000	0.9957	0.2467	0.6243	C2
21	KLINIK ESTETIKA	0.3333	0.0833	0.8352	0.1667	0.2961	C2
22	Miracle Aesthetic Clinic Medan	0.3333	0.4167	0.5031	0.2943	0.2716	C3*
23	Toko Roti Jawara	0.5000	0.1667	0.7822	0.0202	0.4093	C2
24	Jawara Delipark	0.0000	0.0833	0.8784	0.4868	0.2145	C3
25	Jawara Jl. Dr. Cipto	0.3333	0.0000	0.9183	0.2186	0.3548	C2
26	SPBU Krakatau	1.0000	0.3333	0.9284	0.5472	0.8980	C2
27	SPBU MEDAN JOHOR	0.3333	0.1667	0.7521	0.1471	0.2517	C2

28	SPBU Jl. Yos Sudarso	0.3333	0.2500	0.6690	0.1715	0.2301	C2
----	----------------------	--------	--------	--------	--------	--------	-----------

Pada Iterasi 3, terdapat 1 perpindahan cluster. Proses dilanjutkan ke Iterasi 4 untuk memverifikasi konvergensi.

Pada Iterasi 4, tidak ada lagi perpindahan data antar cluster (perubahan = 0). Algoritma K-Means dinyatakan KONVERGEN pada Iterasi 4. Centroid final tidak berubah dari Iterasi 3, sehingga proses clustering selesai.

Tabel 3.12 Ringkasan Perubahan Centroid Per Iterasi dan Konvergensi

Iterasi	Centroid C1 (X1', X2')	Centroid C2	Centroid C3	Jumlah Perpindahan
Awal (I0)	(0.6667, 1.0000)	(0.3333, 0.2500)	(0.0000, 0.0833)	–
Iterasi 1	(0.4167, 1.0000)	(0.3913, 0.2210)	(0.0000, 0.1944)	6
Iterasi 2	(0.2778, 0.9167)	(0.4804, 0.1618)	(0.1042, 0.2708)	1
Iterasi 3	(0.2778, 0.9167)	(0.4896, 0.1458)	(0.1296, 0.2870)	1
Iterasi 4	(0.2778, 0.9167)	(0.4896, 0.1458)	(0.1296, 0.2870)	0 (KONVERGEN)

3.6.9 Hasil Akhir Clustering

Berdasarkan hasil Iterasi 4 yang telah konvergen, berikut adalah hasil akhir pengelompokan seluruh 28 lokasi penempatan karyawan PT. Guna Cipta Prima:

Tabel 3.13 Hasil Akhir Clustering K-Means (k=3)

No	Lokasi Penempatan	Sec	CS	Cluster	Interpretasi
1	Clover Jl. T Umar	5	13	C1	Kebutuhan Tinggi – jumlah CS sangat besar, perlu

					pengelolaan prioritas
2	CLOVER Jl. T. Umar Simp. Airlangga	1	10	C1	Kebutuhan Tinggi – jumlah CS sangat besar, perlu pengelolaan prioritas
3	Clover Jl Krakatau	6	4	C2	Kebutuhan Sedang – Security dominan, CS relatif kecil
4	Clover Jl. Katamso (SUZUYA)	2	4	C3	Kebutuhan Rendah – jumlah Security dan CS di bawah rata-rata
5	Clover Jl. Titi Papan (Gatot Subroto)	5	2	C2	Kebutuhan Sedang – Security dominan, CS relatif kecil
6	Clover Jl. Setia Budi	2	6	C3	Kebutuhan Rendah – jumlah Security dan CS di bawah rata-rata
7	Toko Sony	2	3	C3	Kebutuhan Rendah – jumlah Security dan CS di bawah rata-rata
8	Kantor Malibu	5	2	C2	Kebutuhan Sedang – Security dominan, CS relatif kecil
9	Mega Central Autoniaga (Chevrolet)	3	4	C2	Kebutuhan Sedang – Security dominan, CS relatif kecil
10	Chevrolet Krakatau	2	13	C1	Kebutuhan Tinggi – jumlah CS sangat besar, perlu pengelolaan prioritas
11	Gudang Arista Ngumban Surbakti	3	1	C2	Kebutuhan Sedang – Security dominan, CS relatif kecil

12	Gudang Arista Auto Lestari	1	4	C3	Kebutuhan Rendah – jumlah Security dan CS di bawah rata-rata
13	Gudang Arista Sedap Malam	3	5	C2	Kebutuhan Sedang – Security dominan, CS relatif kecil
14	Hyundai Gudang + Showroom	2	5	C3	Kebutuhan Rendah – jumlah Security dan CS di bawah rata-rata
15	Sutan Kasim	1	4	C3	Kebutuhan Rendah – jumlah Security dan CS di bawah rata-rata
16	Showroom BYD Amir Hamzah	3	2	C2	Kebutuhan Sedang – Security dominan, CS relatif kecil
17	Showroom BYD Cemara	3	2	C2	Kebutuhan Sedang – Security dominan, CS relatif kecil
18	Otto Link	2	6	C3	Kebutuhan Rendah – jumlah Security dan CS di bawah rata-rata
19	Awal Bros Medan	4	3	C2	Kebutuhan Sedang – Security dominan, CS relatif kecil
20	PT. LINTAS SAMUDERA JAYA (LSJ)	5	1	C2	Kebutuhan Sedang – Security dominan, CS relatif kecil
21	KLINIK ESTETIKA	3	2	C2	Kebutuhan Sedang – Security dominan, CS relatif kecil
22	Miracle Aesthetic	3	6	C3	Kebutuhan Rendah – jumlah

	Clinic Medan				Security dan CS di bawah rata-rata
23	Toko Roti Jawara	4	3	C2	Kebutuhan Sedang – Security dominan, CS relatif kecil
24	Jawara Delipark	1	2	C3	Kebutuhan Rendah – jumlah Security dan CS di bawah rata-rata
25	Jawara Jl. Dr. Cipto	3	1	C2	Kebutuhan Sedang – Security dominan, CS relatif kecil
26	SPBU Krakatau	7	5	C2	Kebutuhan Sedang – Security dominan, CS relatif kecil
27	SPBU MEDAN JOHOR	3	3	C2	Kebutuhan Sedang – Security dominan, CS relatif kecil
28	SPBU Jl. Yos Sudarso	3	4	C2	Kebutuhan Sedang – Security dominan, CS relatif kecil

3.6.10 Rekapitulasi dan Interpretasi Hasil Cluster

Tabel 3.14 Rekapitulasi Hasil Akhir Clustering

Cluster	Jumlah Lokasi	Rata-rata Sec	Rata-rata CS	Interpretasi Manajemen
C1 - Tinggi	3 lokasi	2.67	12.00	Lokasi dengan kebutuhan CS sangat tinggi. Perlu penambahan CS atau rotasi jadwal intensif.
C2 - Sedang	16 lokasi	3.94	2.75	Lokasi dengan Security dominan dan CS rendah. Kapasitas Security perlu dioptimalkan.

C3 - Rendah	9 lokasi	1.78	4.44	Lokasi dengan jumlah Security dan CS di bawah rata-rata. Berpotensi untuk efisiensi alokasi SDM.
--------------------	----------	------	------	--

3.7 Evaluasi Kinerja Clustering

3.7.1 Silhouette Coefficient

Silhouette Coefficient (SC) digunakan untuk mengukur seberapa baik setiap data point berada dalam clusternya sendiri dibandingkan dengan cluster lainnya.

Rumus Silhouette Coefficient:

$$s(i) = [b(i) - a(i)] / \max[a(i), b(i)]$$

di mana $a(i)$ adalah rata-rata jarak data ke- i ke semua anggota clusternya sendiri, dan $b(i)$ adalah rata-rata jarak data ke- i ke semua anggota cluster terdekat lainnya.

Nilai SC berkisar antara -1 hingga +1.

Tabel 3.15 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

Berdasarkan hasil komputasi terhadap dataset 28 lokasi penempatan karyawan PT. Guna Cipta Prima dengan $k=3$, diperoleh nilai Silhouette Coefficient sebesar 0.3498. Nilai ini berada pada rentang 0.26 – 0.50 yang termasuk kategori "Cukup". Artinya, struktur cluster yang terbentuk dapat dipertimbangkan dan cukup valid untuk digunakan sebagai dasar pengambilan keputusan penempatan karyawan. Nilai SC ini dipengaruhi oleh karakteristik data yang memiliki variasi nilai cukup lebar dan beberapa lokasi yang berada di perbatasan antara dua cluster.

3.7.2 Evaluasi WCSS/SSE

Nilai SSE (Sum of Squared Errors) atau WCSS (Within-Cluster Sum of Squares) dari hasil clustering akhir dengan $k=3$ adalah 1.1563 (berdasarkan data ternormalisasi). Dibandingkan dengan SSE pada $k=1$ (nilai 3.6419), penggunaan $k=3$ mampu mereduksi total kesalahan cluster sebesar 68.24%. Ini menunjukkan bahwa pengelompokan menjadi 3 cluster memberikan peningkatan kualitas yang signifikan dibandingkan tanpa pengelompokan.

3.7.3 Validasi Hasil Cluster

Validasi hasil cluster dilakukan dengan menginterpretasikan karakteristik setiap cluster berdasarkan nilai rata-rata atribut. Penjelasan per cluster:

- a. Cluster C1 – Kebutuhan Tinggi: Terdiri dari 3 lokasi yaitu Clover Jl. T Umar (Sec=5, CS=13), CLOVER Jl. T. Umar Simp. Airlangga (Sec=1, CS=10), dan Chevrolet Krakatau (Sec=2, CS=13). Rata-rata CS = 12.00, jauh di atas rata-rata keseluruhan (4.54). Lokasi-lokasi ini membutuhkan jumlah CS yang sangat besar, kemungkinan karena luas area yang perlu dibersihkan. Manajemen perlu memastikan ketersediaan SDM CS yang cukup untuk lokasi-lokasi ini.
- b. Cluster C2 – Kebutuhan Sedang: Terdiri dari 16 lokasi dengan rata-rata Security = 3.94 dan rata-rata CS = 2.75. Lokasi dalam cluster ini memiliki jumlah Security yang cukup besar namun CS relatif rendah. Contohnya SPBU Krakatau (Sec=7, CS=5) dan Clover Jl Krakatau (Sec=6, CS=4). Penempatan Security menjadi prioritas utama untuk cluster ini.
- c. Cluster C3 – Kebutuhan Rendah: Terdiri dari 9 lokasi dengan rata-rata Security = 1.78 dan rata-rata CS = 4.44. Lokasi-lokasi ini memiliki jumlah Security yang kecil tetapi CS yang moderat. Beberapa lokasi dalam cluster ini seperti Miracle Aesthetic Clinic Medan (Sec=3, CS=6) dan Otto Link (Sec=2, CS=6) berpotensi memerlukan penambahan Security. Manajemen dapat mempertimbangkan redistribusi SDM dari cluster ini ke cluster yang membutuhkan.

3.8 Perancangan Sistem

Perancangan sistem merupakan tahapan penting yang menggambarkan arsitektur, alur kerja, struktur data, dan antarmuka sistem berbasis web yang akan dibangun untuk mengimplementasikan algoritma K-Means Clustering pada data penempatan karyawan PT. Guna Cipta Prima. Sistem ini dibangun menggunakan PHP versi 8.1 (native, tanpa framework) dengan ekstensi MySQLi sebagai layer koneksi database, serta MySQL sebagai sistem manajemen basis data.

3.8.1 Gambaran Umum Sistem

Sistem yang dikembangkan merupakan aplikasi web berbasis arsitektur MVC (Model-View-Controller) sederhana dengan komponen utama sebagai berikut:

- a. Backend: PHP 8.1 Native dengan MySQLi untuk koneksi database, pemrosesan logika bisnis, dan implementasi algoritma K-Means.
- b. Frontend: HTML5, CSS3, JavaScript (vanilla), dan Bootstrap 5 untuk tampilan antarmuka yang responsif.
- c. Database: MySQL 8.0 diakses melalui ekstensi MySQLi PHP dengan prepared statements untuk keamanan.
- d. Server: Apache (XAMPP) sebagai web server lokal selama pengembangan.
- e. Sistem memiliki alur kerja utama sebagai berikut: Admin melakukan login → Admin mengelola data lokasi dan karyawan → Admin menjalankan proses K-Means Clustering → Sistem menampilkan hasil cluster, grafik Elbow SSE, nilai Silhouette Coefficient, dan rekomendasi penempatan → Admin dapat mencetak laporan hasil clustering.

3.8.2 Desain Alur Sistem (System Flowchart)

Berikut adalah deskripsi naratif diagram alur sistem secara lengkap:

Tabel 3.16 Deskripsi Alur Sistem Web K-Means Clustering

No	Proses	Deskripsi
1	Login Admin	Admin memasukkan username dan password. Sistem memverifikasi kredensial melalui query SELECT ke tabel tbl_users menggunakan MySQLi prepared statement. Jika valid,

		session PHP dimulai dan admin diarahkan ke halaman dashboard.
2	Input Data Lokasi	Admin menambah, mengedit, atau menghapus data lokasi penempatan. Data disimpan ke tabel tbl_lokasi. Validasi input dilakukan di sisi server menggunakan PHP.
3	Input Data Karyawan	Admin menginputkan jumlah Security dan CS untuk setiap lokasi. Data disimpan ke tabel tbl_karyawan dengan relasi FK ke tbl_lokasi.
4	Preprocessing Data	Sistem secara otomatis melakukan: (a) pengecekan missing value, (b) normalisasi Min-Max pada kolom jumlah_security dan jumlah_cs, (c) menyimpan nilai normalisasi ke array PHP untuk proses clustering.
5	Penentuan Jumlah Cluster (Elbow)	Sistem menjalankan K-Means untuk k=1 hingga k=7, menghitung SSE setiap k, dan menampilkan grafik Elbow menggunakan Chart.js. Admin memilih nilai k berdasarkan titik elbow.
6	Proses K-Means Clustering	Sistem menjalankan algoritma K-Means: (a) inisialisasi centroid awal, (b) perhitungan jarak Euclidean, (c) pengelompokan data ke cluster terdekat, (d) pembaruan centroid, (e) iterasi hingga konvergen. Hasil disimpan ke tabel tbl_clustering_result dan tbl_clustering_detail.
7	Evaluasi Hasil	Sistem menghitung Silhouette Coefficient dan WCSS/SSE dari hasil clustering, kemudian menampilkan nilai evaluasi di halaman hasil.
8	Tampilan Hasil & Laporan	Sistem menampilkan: tabel hasil cluster per lokasi, grafik visualisasi cluster, nilai SSE dan Silhouette, serta rekomendasi penempatan. Admin dapat mencetak laporan dalam format PDF.

3.8.3 Use Case Diagram (Naratif)

Sistem ini memiliki satu aktor utama yaitu Admin. Berikut adalah spesifikasi use case yang dimiliki aktor Admin:

Tabel 3.17 Spesifikasi Use Case Diagram (Naratif)

No	Use Case	Prasyarat	Deskripsi
UC-01	Login	Memiliki akun terdaftar	Admin memasukkan username dan password untuk mengakses sistem
UC-02	Kelola Data Lokasi	Sudah login	Admin dapat menambah, mengedit, menghapus data lokasi penempatan
UC-03	Kelola Data Karyawan	Sudah login	Admin menginput jumlah Security dan CS per lokasi
UC-04	Lihat Grafik Elbow	Data karyawan tersedia	Sistem menampilkan grafik SSE vs k untuk memilih jumlah cluster optimal
UC-05	Jalankan Proses K-Means	Data karyawan tersedia, k ditentukan	Admin memicu proses clustering, sistem menjalankan algoritma K-Means
UC-06	Lihat Hasil Clustering	Proses K-Means selesai	Admin melihat tabel hasil cluster setiap lokasi beserta nilai evaluasi
UC-07	Lihat Evaluasi (SSE & SC)	Proses K-Means selesai	Sistem menampilkan nilai SSE, Silhouette Coefficient, dan interpretasinya
UC-08	Cetak Laporan	Hasil clustering tersedia	Admin mencetak laporan hasil clustering dalam format PDF
UC-09	Logout	Sudah login	Admin mengakhiri sesi dan sistem menghapus session PHP

3.8.4 Activity Diagram (Naratif)

Berikut adalah alur aktivitas sistem dari awal hingga akhir:

- Sistem: Menampilkan halaman Login. Admin: Memasukkan username dan password.
- Sistem: Memverifikasi credentials via MySQLi. Jika gagal → kembali ke halaman login. Jika berhasil → tampilkan Dashboard.
- Admin: Mengelola data lokasi (CRUD) dan data karyawan (jumlah Security & CS per lokasi).

- d. Admin: Menekan tombol "Proses Clustering". Sistem: Membaca semua data dari tabel `tbl_karyawan`.
- e. Sistem: Melakukan preprocessing → normalisasi Min-Max semua nilai atribut.
- f. Sistem: Menjalankan algoritma Elbow untuk menampilkan rekomendasi nilai `k`.
- g. Admin: Memilih nilai `k` (default `k=3`). Sistem: Menentukan centroid awal.
- h. Sistem: Menghitung jarak Euclidean setiap data ke setiap centroid. Mengelompokkan data ke centroid terdekat.
- i. Sistem: Memperbarui centroid berdasarkan rata-rata anggota cluster. Memeriksa konvergensi.
- j. Jika belum konvergen → ulangi langkah 8-9. Jika konvergen → simpan hasil ke database.
- k. Sistem: Menghitung Silhouette Coefficient dan SSE. Menampilkan halaman hasil clustering.
- l. Admin: Dapat mencetak laporan atau logout.

3.8.5 Perancangan Basis Data (Database Design)

Database sistem dirancang menggunakan MySQL 8.0 dengan diakses melalui ekstensi MySQLi pada PHP 8.1. Perancangan basis data mengikuti prinsip normalisasi hingga 3NF (Third Normal Form) untuk meminimalkan redundansi data. Sistem menggunakan 5 tabel utama yang saling berelasi.

a. Tabel `tbl_users`

Tabel ini menyimpan data akun admin yang berhak mengakses sistem.

Tabel 3.18 Struktur Tabel `tbl_users`

Nama Kolom	Tipe Data	Panjang	Null	Key	Keterangan
<code>id_user</code>	INT	11	NOT NULL	PK, AUTO_INCREMENT	Primary key, identifier unik setiap user
<code>username</code>	VARCHAR	50	NOT	UNIQUE	Username untuk

			NULL	E	login, harus unik
password	VARCHAR	255	NOT NULL	–	Password terenkripsi menggunakan password_hash() PHP (bcrypt)
nama_lengkap	VARCHAR	100	NOT NULL	–	Nama lengkap admin
role	ENUM	–	NOT NULL	–	Peran user: 'admin' (saat ini hanya satu peran)
created_at	TIMESTAMP	–	DEFAULT	–	Waktu pembuatan akun, otomatis terisi
last_login	TIMESTAMP	–	NULL	–	Waktu login terakhir, diperbarui setiap login sukses

b. Tabel tbl_lokasi

Tabel ini menyimpan data master lokasi penempatan karyawan.

Tabel 3.19 Struktur Tabel tbl_lokasi

Nama Kolom	Tipe Data	Panjang	Null	Key	Keterangan
id_lokasi	INT	11	NOT NULL	PK, AUTO_INCREMENT	Primary key tabel lokasi
nama_lokasi	VARCHAR	200	NOT NULL	–	Nama lengkap lokasi penempatan
alamat	TEXT	–	NULL	–	Alamat detail lokasi
keterangan	TEXT	–	NULL	–	Catatan tambahan tentang lokasi
status	ENUM	–	NOT	–	Status lokasi: 'aktif'

			NULL		atau 'nonaktif'
created_at	TIMESTAM P	–	DEFAUL T	–	Waktu data lokasi ditambahkan
updated_at	TIMESTAM P	–	NULL	–	Waktu terakhir data diperbarui

c. Tabel tbl_karyawan

Tabel ini menyimpan data jumlah karyawan per lokasi yang menjadi input utama proses K-Means Clustering.

Tabel 3.20 Struktur Tabel tbl_karyawan

Nama Kolom	Tipe Data	Panjang	Null	Key	Keterangan
id_karyawan	INT	11	NOT NULL	PK, AUTO_I NCREME NT	Primary key tabel karyawan
id_lokasi	INT	11	NOT NULL	FK → tbl_lokasi	Foreign key ke tabel lokasi
jumlah_securi ty	INT	11	NOT NULL	–	Jumlah tenaga Security di lokasi tersebut (X1)
jumlah_cs	INT	11	NOT NULL	–	Jumlah tenaga Cleaning Service di lokasi tersebut (X2)
periode	VARCHAR	20	NOT NULL	–	Periode data, contoh: '2025'
created_at	TIMESTAM P	–	DEFAUL T	–	Waktu data ditambahkan
updated_at	TIMESTAM	–	NULL	–	Waktu terakhir

	P				data diperbarui
--	---	--	--	--	-----------------

d. Tabel tbl_clustering_result

Tabel ini menyimpan metadata setiap eksekusi proses K-Means Clustering (header hasil clustering).

Tabel 3.21 Struktur Tabel tbl_clustering_result

Nama Kolom	Tipe Data	Panjang	Null	Key	Keterangan
id_result	INT	11	NOT NULL	PK, AUTO_INCREMENT	Primary key hasil clustering
jumlah_cluster_k	INT	11	NOT NULL	–	Nilai k yang digunakan pada eksekusi ini
total_data	INT	11	NOT NULL	–	Jumlah data yang diproses
jumlah_iterasi	INT	11	NOT NULL	–	Jumlah iterasi hingga konvergen
nilai_sse	DECIMAL	10,6	NOT NULL	–	Nilai SSE/WCSS akhir clustering
nilai_silhouette	DECIMAL	6,4	NOT NULL	–	Nilai Silhouette Coefficient
tanggal_proses	TIMESTAMP	–	DEFAULT	–	Waktu proses clustering dijalankan
id_user	INT	11	NOT NULL	FK → tbl_users	Siapa yang menjalankan proses

e. Tabel tbl_clustering_detail

Tabel ini menyimpan detail hasil pengelompokan setiap data lokasi beserta jarak ke centroid dan nomor cluster yang ditetapkan.

Tabel 3.22 Struktur Tabel tbl_clustering_detail

Nama Kolom	Tipe Data	Panjang	Null	Key	Keterangan
id_detail	INT	11	NOT NULL	PK, AUTO_INCREMENT	Primary key detail cluster
id_result	INT	11	NOT NULL	FK → tbl_clustering_result	Relasi ke header hasil clustering
id_karyawan	INT	11	NOT NULL	FK → tbl_karyawan	Relasi ke data karyawan yang dicluster
cluster_ke	INT	11	NOT NULL	–	Nomor cluster yang ditetapkan (1, 2, atau 3)
sec_norm	DECIMAL	8,6	NOT NULL	–	Nilai normalisasi Security (X1')
cs_norm	DECIMAL	8,6	NOT NULL	–	Nilai normalisasi CS (X2')
jarak_c1	DECIMAL	10,6	NOT NULL	–	Jarak Euclidean ke centroid C1
jarak_c2	DECIMAL	10,6	NOT NULL	–	Jarak Euclidean ke centroid C2
jarak_c3	DECIMAL	10,6	NOT NULL	–	Jarak Euclidean ke centroid C3 (NULL jika k<3)

label_cluster	VARCHAR	30	NOT NULL	–	Label deskriptif: 'Tinggi', 'Sedang', atau 'Rendah'
---------------	---------	----	-------------	---	---

3.8.6 Relasi Antar Tabel (ERD Deskriptif)

Berikut adalah deskripsi relasi antar tabel dalam basis data sistem:

Tabel 3.23 Relasi Antar Tabel Basis Data

Tabel Induk	Tipe Relasi	Tabel Anak	Keterangan
tbl_lokasi	One to Many (1:N)	tbl_karyawan	Satu lokasi dapat memiliki beberapa data karyawan (berbeda periode). Relasi: tbl_lokasi.id_lokasi → tbl_karyawan.id_lokasi
tbl_users	One to Many (1:N)	tbl_clustering_result	Satu user admin dapat melakukan banyak kali proses clustering. Relasi: tbl_users.id_user → tbl_clustering_result.id_user
tbl_clustering_result	One to Many (1:N)	tbl_clustering_detail	Satu hasil clustering memiliki banyak baris detail (satu per lokasi). Relasi: tbl_clustering_result.id_result → tbl_clustering_detail.id_result
tbl_karyawan	One to Many (1:N)	tbl_clustering_detail	Satu data karyawan dapat muncul di banyak hasil clustering (berbeda eksekusi). Relasi: tbl_karyawan.id_karyawan → tbl_clustering_detail.id_karyawan

3.8.7 Skema SQL DDL (PHP 8.1 + MySQLi)

Berikut adalah skema pembuatan tabel database menggunakan SQL DDL yang akan dieksekusi pada saat instalasi sistem:

```
CREATE DATABASE IF NOT EXISTS db_kmeans_ptguna
  CHARACTER SET utf8mb4 COLLATE utf8mb4_unicode_ci;
USE db_kmeans_ptguna;
-- Tabel 1: tbl_users
CREATE TABLE tbl_users (
  id_user    INT(11) NOT NULL AUTO_INCREMENT,
  username   VARCHAR(50) NOT NULL,
  password   VARCHAR(255) NOT NULL,
  nama_lengkap VARCHAR(100) NOT NULL,
  role       ENUM('admin') NOT NULL DEFAULT 'admin',
  created_at  TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
  last_login  TIMESTAMP NULL,
  PRIMARY KEY (id_user),
  UNIQUE KEY idx_username (username)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4;
-- Tabel 2: tbl_lokasi
CREATE TABLE tbl_lokasi (
  id_lokasi  INT(11) NOT NULL AUTO_INCREMENT,
  nama_lokasi VARCHAR(200) NOT NULL,
  alamat     TEXT NULL,
  keterangan TEXT NULL,
  status     ENUM('aktif','nonaktif') NOT NULL DEFAULT 'aktif',
  created_at  TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
  updated_at  TIMESTAMP NULL ON UPDATE CURRENT_TIMESTAMP,
  PRIMARY KEY (id_lokasi)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4;
-- Tabel 3: tbl_karyawan
CREATE TABLE tbl_karyawan (
  id_karyawan  INT(11) NOT NULL AUTO_INCREMENT,
```

```

id_lokasi    INT(11) NOT NULL,
jumlah_security INT(11) NOT NULL DEFAULT 0,
jumlah_cs    INT(11) NOT NULL DEFAULT 0,
periode      VARCHAR(20) NOT NULL DEFAULT '2025',
created_at   TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
updated_at   TIMESTAMP NULL ON UPDATE
CURRENT_TIMESTAMP,
PRIMARY KEY (id_karyawan),
KEY fk_karyawan_lokasi (id_lokasi),
CONSTRAINT fk_karyawan_lokasi
FOREIGN KEY (id_lokasi) REFERENCES tbl_lokasi(id_lokasi)
ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4;
-- Tabel 4: tbl_clustering_result
CREATE TABLE tbl_clustering_result (
id_result    INT(11) NOT NULL AUTO_INCREMENT,
jumlah_cluster_k INT(11) NOT NULL DEFAULT 3,
total_data   INT(11) NOT NULL,
jumlah_iterasi INT(11) NOT NULL,
nilai_sse    DECIMAL(10,6) NOT NULL,
nilai_silhouette DECIMAL(6,4) NOT NULL,
tanggal_proses TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
id_user      INT(11) NOT NULL,
PRIMARY KEY (id_result),
KEY fk_result_user (id_user),
CONSTRAINT fk_result_user
FOREIGN KEY (id_user) REFERENCES tbl_users(id_user)
ON DELETE RESTRICT ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4;
-- Tabel 5: tbl_clustering_detail
CREATE TABLE tbl_clustering_detail (
id_detail INT(11) NOT NULL AUTO_INCREMENT,

```

```

id_result INT(11) NOT NULL,
id_karyawan INT(11) NOT NULL,
cluster_ke INT(11) NOT NULL,
sec_norm DECIMAL(8,6) NOT NULL,
cs_norm DECIMAL(8,6) NOT NULL,
jarak_c1 DECIMAL(10,6) NOT NULL,
jarak_c2 DECIMAL(10,6) NOT NULL,
jarak_c3 DECIMAL(10,6) NULL,
label_cluster VARCHAR(30) NOT NULL,
PRIMARY KEY (id_detail),
KEY fk_detail_result (id_result),
KEY fk_detail_karyawan (id_karyawan),
CONSTRAINT fk_detail_result
FOREIGN KEY (id_result) REFERENCES tbl_clustering_result(id_result)
ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT fk_detail_karyawan
FOREIGN KEY (id_karyawan) REFERENCES tbl_karyawan(id_karyawan)
ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4;

```

Gambar 3.1 Skema SQL DDL Basis Data Sistem (PHP 8.1 + MySQLi)

3.8.8 Desain Antarmuka (UI/UX)

Sistem dirancang dengan antarmuka yang intuitif dan responsif menggunakan Bootstrap 5. Berikut adalah rancangan halaman-halaman utama sistem:

Tabel 3.24 Rancangan Antarmuka Sistem (UI/UX)

Kode	Nama Halaman	Komponen dan Fungsi Utama
P-01	Halaman Login	Form input username dan password; validasi sisi server dengan PHP + MySQLi prepared statement; session management menggunakan PHP session; proteksi brute-force dengan rate limiting

P-02	Dashboard	Ringkasan statistik: total lokasi aktif, total Security, total CS, hasil clustering terakhir; menu navigasi sidebar; kartu informasi dengan grafik Chart.js
P-03	Data Lokasi	Tabel data lokasi dengan paginasi (LIMIT/OFFSET MySQLi); tombol Tambah/Edit/Hapus; modal dialog untuk form input; pencarian data real-time (AJAX)
P-04	Data Karyawan	Tabel data jumlah Security & CS per lokasi; filter berdasarkan periode; form import data dari Excel; tombol proses clustering
P-05	Proses Clustering	Tampilan grafik Elbow (Chart.js) untuk pemilihan k; input jumlah cluster k; tombol 'Jalankan K-Means'; progress bar proses; tampilan centroid per iterasi
P-06	Hasil Clustering	Tabel lengkap hasil cluster per lokasi; filter berdasarkan cluster (C1/C2/C3); visualisasi scatter plot 2D; nilai SSE dan Silhouette Coefficient dengan interpretasi
P-07	Laporan	Ringkasan hasil clustering; tabel rekomendasi penempatan per cluster; tombol cetak laporan PDF menggunakan FPDF/TCPDF library

3.8.9 Desain Implementasi Algoritma K-Means dalam PHP 8.1

Berikut adalah pseudocode implementasi algoritma K-Means dalam PHP

8.1 menggunakan MySQLi:

```
// 1. AMBIL DATA dari database menggunakan MySQLi
$conn = new mysqli(DB_HOST, DB_USER, DB_PASS, DB_NAME);
$stmt = $conn->prepare(
    "SELECT k.id_karyawan, k.jumlah_security, k.jumlah_cs, l.nama_lokasi
    FROM tbl_karyawan k JOIN tbl_lokasi l ON k.id_lokasi = l.id_lokasi
    WHERE k.periode = ? AND l.status = 'aktif'");
$stmt->bind_param('s', $periode);
$stmt->execute();

// 2. NORMALISASI Min-Max
$sec_min = min(array_column($data, 'security'));
$sec_max = max(array_column($data, 'security'));
```

```

$cs_min = min(array_column($data, 'cs'));
$cs_max = max(array_column($data, 'cs'));
foreach ($data as &$d) {
    $d['sec_norm'] = ($d['security'] - $sec_min) / ($sec_max - $sec_min);
    $d['cs_norm'] = ($d['cs'] - $cs_min) / ($cs_max - $cs_min);
}
// 3. INISIALISASI CENTROID AWAL (acak dari data)
$centroid_idx = array_rand($data, $k);
foreach ($centroid_idx as $i => $idx) {
    $centroids[$i] = ['x1' => $data[$idx]['sec_norm'],
                     'x2' => $data[$idx]['cs_norm']];
}
// 4. ITERASI K-MEANS
$converged = false; $iteration = 0;
while (!$converged && $iteration < $max_iter) {
    // Hitung jarak Euclidean dan tetapkan cluster
    foreach ($data as &$d) {
        $min_dist = PHP_FLOAT_MAX;
        for ($j = 0; $j < $k; $j++) {
            $dist = sqrt(pow($d['sec_norm'] - $centroids[$j]['x1'], 2)
                        + pow($d['cs_norm'] - $centroids[$j]['x2'], 2));
            if ($dist < $min_dist) {
                $min_dist = $dist; $d['cluster'] = $j;
            }
        }
        $d['min_dist'] = $min_dist;
    }
    // Perbarui centroid
    $new_centroids = [];
    for ($j = 0; $j < $k; $j++) {
        $members = array_filter($data, fn($d) => $d['cluster'] === $j);
        if (count($members) > 0) {

```

```

    $new_centroids[$j] = [
        'x1'
        =>
array_sum(array_column($members,'sec_norm'))/count($members),
        'x2'=> array_sum(array_column($members,'cs_norm')) /count($members),
    ];
    }
}
// Cek konvergensi
$converged = ($new_centroids === $centroids);
$centroids = $new_centroids; $iteration++;
}
// 5. SIMPAN HASIL ke database
$stmt_result = $conn->prepare(
    "INSERT INTO tbl_clustering_result
    (jumlah_cluster_k, total_data, jumlah_iterasi, nilai_sse,
    nilai_silhouette, id_user) VALUES (?, ?, ?, ?, ?, ?)");

```

Gambar 3.2 Pseudocode Implementasi K-Means dalam PHP 8.1 + MySQLi

3.9 Alat dan Perangkat Penelitian

3.9.1 Perangkat Keras (Hardware)

Tabel 3.25 Spesifikasi Perangkat Keras

No	Komponen	Spesifikasi	Fungsi
1	Laptop/Komputer	Prosesor Intel Core i3, RAM 4 GB, Penyimpanan 256 GB, Sistem Operasi Windows 10 Pro (64-bit)	Perangkat utama untuk pengembangan sistem, pengolahan data, dan penulisan laporan

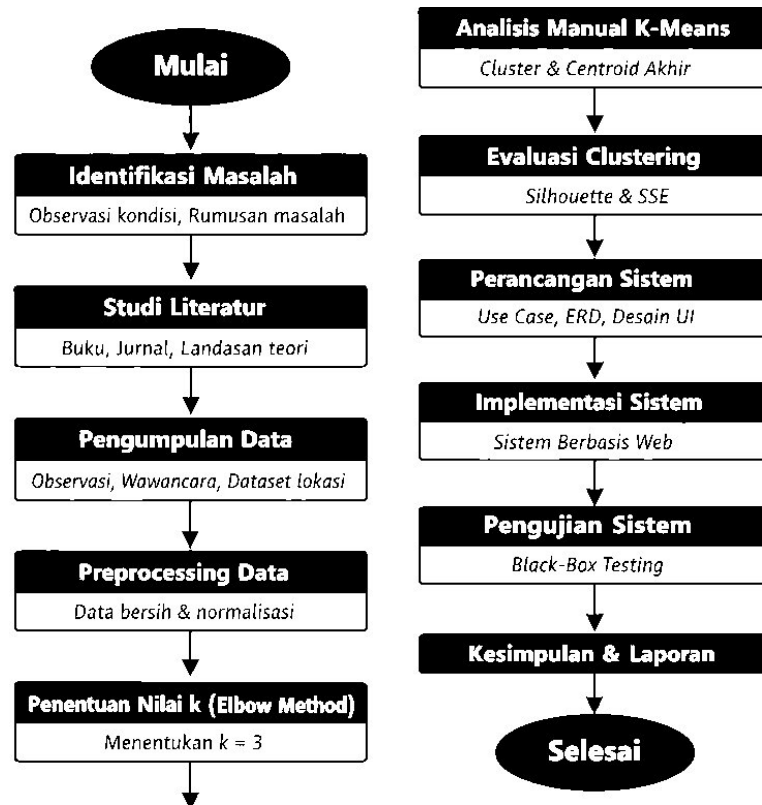
3.9.2 Perangkat Lunak (Software)

Tabel 3.26 Spesifikasi Perangkat Lunak

No	Perangkat Lunak	Versi	Fungsi dalam Penelitian
1	PHP	8.1.x	Bahasa pemrograman utama backend sistem web
2	MySQL	8.0.x	Sistem manajemen basis data relasional
3	MySQLi (PHP Extension)	Bundled	Ekstensi PHP untuk koneksi dan query ke database MySQL
4	Apache (XAMPP)	7.4.x	Web server lokal untuk pengembangan dan pengujian sistem
5	phpMyAdmin	5.x	GUI manajemen database MySQL berbasis web
6	Visual Studio Code	Latest	Code editor untuk penulisan kode PHP, HTML, CSS, JavaScript
7	Bootstrap	5.3.x	Framework CSS untuk desain antarmuka yang responsif
8	Chart.js	4.x	Library JavaScript untuk visualisasi grafik Elbow dan scatter plot
10	Microsoft Excel	2019/365	Pengolahan awal dataset dan verifikasi hasil normalisasi
11	Microsoft Word	2019/365	Penyusunan dokumen laporan skripsi

3.10 Diagram Alur Penelitian

Penelitian ini dilaksanakan mengikuti alur tahapan yang sistematis dan terstruktur sebagaimana digambarkan dalam diagram alur berikut. Setiap tahapan memiliki input, proses, dan output yang terdefinisi dengan jelas.



Gambar 3.3 Diagram Alur

Tabel 3.27 Penjelasan Tahapan dan Alur Penelitian

No	Tahapan	Input / Proses	Output / Hasil
1	Identifikasi Masalah	Observasi kondisi penempatan karyawan di PT. Guna Cipta Prima	Rumusan masalah: penempatan berbasis intuisi, perlu metode data-driven
2	Studi Literatur	Buku, jurnal, dan referensi tentang K-Means, data mining, PHP, MySQL	Landasan teori dan referensi metode penelitian
3	Pengumpulan Data	Observasi, wawancara, dan dokumentasi di PT. Guna Cipta Prima	Dataset 28 lokasi: nama lokasi, jumlah Security, jumlah CS
4	Preprocessing Data	Dataset mentah	Data bersih, atribut terpilih, data ternormalisasi (Min-Max Scaling)
5	Penentuan k (Elbow)	Data ternormalisasi, k=1..7	Grafik SSE vs k, keputusan k=3

6	Analisis Manual K-Means	Data ternormalisasi, $k=3$, centroid awal	Hasil iterasi manual: jarak Euclidean, pengelompokan, centroid baru, konvergensi pada iterasi ke-4
7	Evaluasi Clustering	Hasil clustering (assignment per data)	Silhouette Coefficient = 0.3498 (Cukup), SSE = 1.1563
8	Perancangan Sistem	Hasil analisis, spesifikasi kebutuhan	Rancangan Use Case, Activity Diagram, ERD, skema database, desain UI
9	Implementasi Sistem	Rancangan sistem, PHP 8.1, MySQL + MySQLi	Sistem web berjalan: CRUD data, proses K-Means, visualisasi hasil
10	Pengujian Sistem	Sistem yang telah diimplementasi	Hasil Black-Box Testing: seluruh fitur berfungsi sesuai spesifikasi
11	Kesimpulan & Laporan	Semua hasil penelitian	Laporan skripsi lengkap, publikasi, rekomendasi untuk perusahaan

BAB IV

HASIL DAN PEMBAHASAN

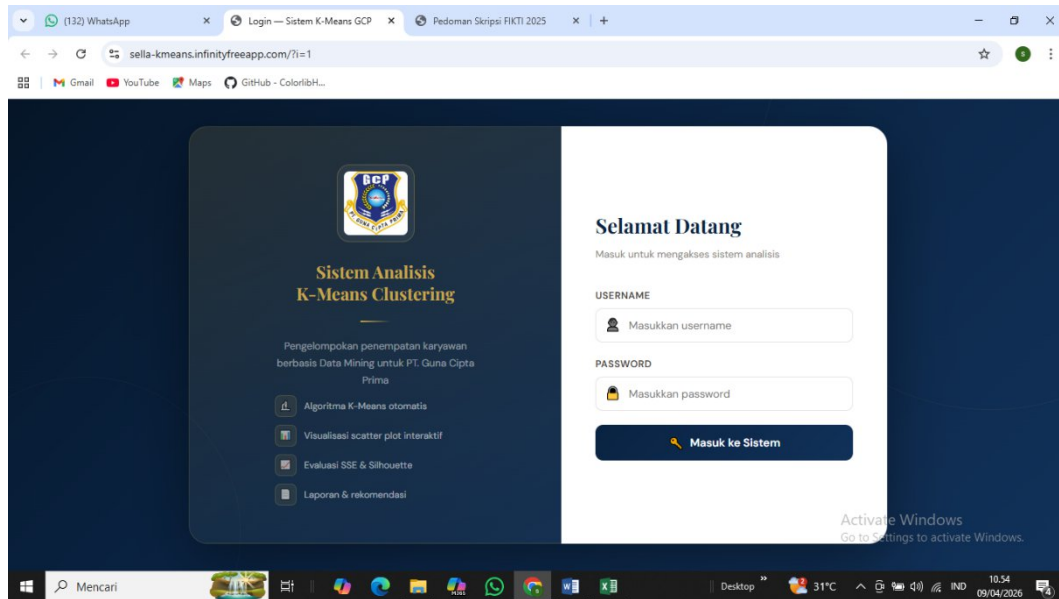
4.1 Hasil Implementasi Sistem

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.

Implementasi sistem telah memenuhi seluruh spesifikasi fungsional yang dirancang dalam BAB III, yaitu tujuh halaman utama: Halaman Login, Dashboard, Data Lokasi, Data Karyawan, Proses Clustering, Hasil Clustering, dan Laporan. Setiap halaman mengimplementasikan pola arsitektur yang konsisten dengan pemisahan logika bisnis pada lapisan `includes/` dan presentasi pada lapisan `pages/`. Koneksi basis data menggunakan prepared statements MySQLi untuk mencegah SQL injection, sesuai praktik keamanan pengembangan web yang direkomendasikan (Ariyus, 2018).

4.1.1 Halaman Login

Halaman login merupakan gerbang akses utama sistem yang memastikan hanya pengguna yang terotorisasi dapat menggunakan fitur analisis. Halaman ini dibangun dengan layout dua kolom: kolom kiri menampilkan identitas sistem beserta logo PT. Guna Cipta Prima dan daftar fitur unggulan, sedangkan kolom kanan menampilkan formulir autentikasi. Proses validasi login menggunakan prepared statement MySQLi dengan pengecekan sesi PHP (PHP session) untuk pengelolaan status autentikasi.

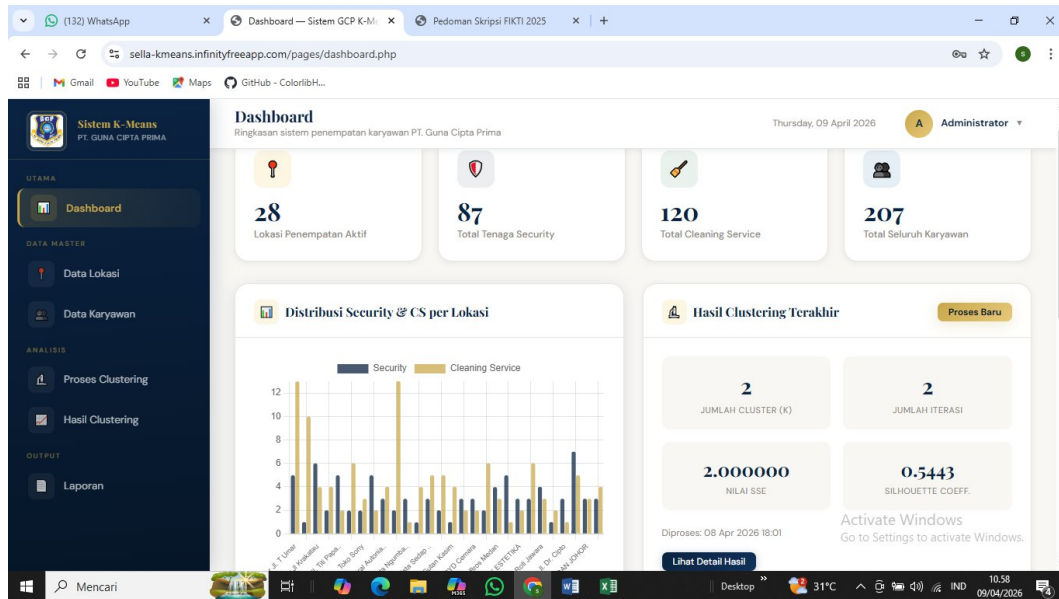


Gambar 4.1 Tampilan Halaman Login Sistem

Formulir login terdiri dari dua field input yaitu username dan password, serta tombol "Masuk ke Sistem". Apabila kredensial yang dimasukkan tidak valid, sistem menampilkan pesan kesalahan "Username atau password salah. Silakan coba lagi." tanpa memberikan informasi spesifik tentang field mana yang salah, sesuai prinsip keamanan sistem autentikasi (Stallings & Brown, 2018). Setelah login berhasil, sistem menyimpan identitas pengguna ke dalam variabel sesi PHP dan mengarahkan pengguna ke halaman Dashboard.

4.1.2 Halaman Dashboard

Halaman Dashboard berfungsi sebagai pusat informasi (information hub) yang menampilkan rangkuman statistik terkini sistem secara real-time. Terdapat empat kartu statistik utama yang menampilkan: (1) jumlah lokasi penempatan aktif, (2) total tenaga Security, (3) total Cleaning Service, dan (4) total seluruh karyawan pada periode yang aktif. Selain itu, dashboard juga menampilkan ringkasan hasil clustering terakhir yang mencakup nilai k , jumlah iterasi, nilai SSE, dan Silhouette Coefficient.

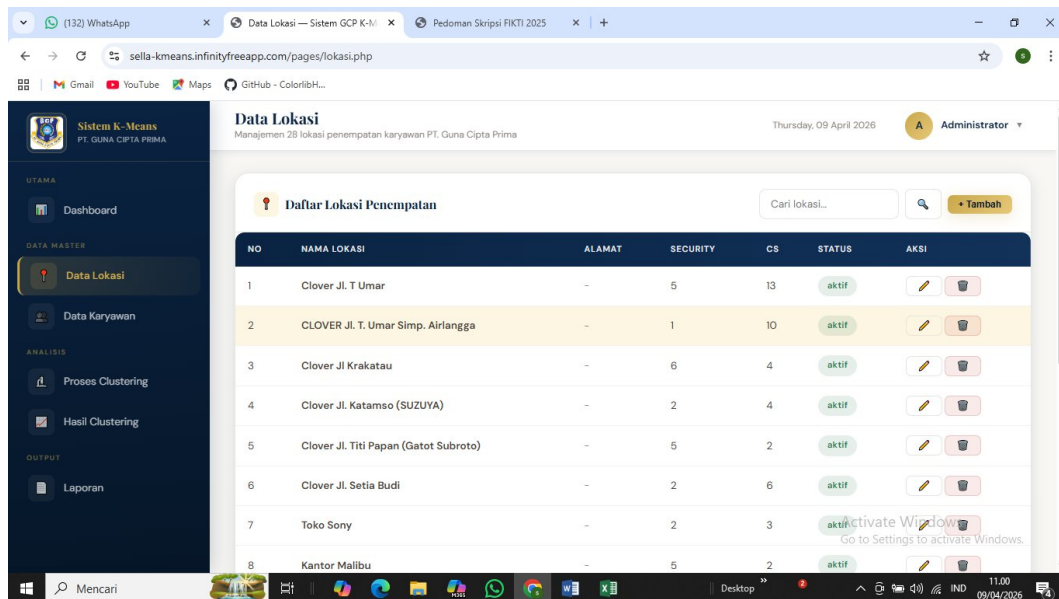


Gambar 4.2 Tampilan Halaman Dashboard

Di bagian bawah dashboard terdapat grafik batang (bar chart) yang memvisualisasikan distribusi jumlah Security dan Cleaning Service untuk seluruh 28 lokasi penempatan, dirender menggunakan library Chart.js versi 4. Grafik ini memberikan gambaran visual yang cepat dan intuitif bagi pengguna mengenai pola distribusi tenaga kerja lintas lokasi. Sidebar navigasi menampilkan logo perusahaan, menu navigasi terstruktur, dan informasi pengguna yang sedang aktif.

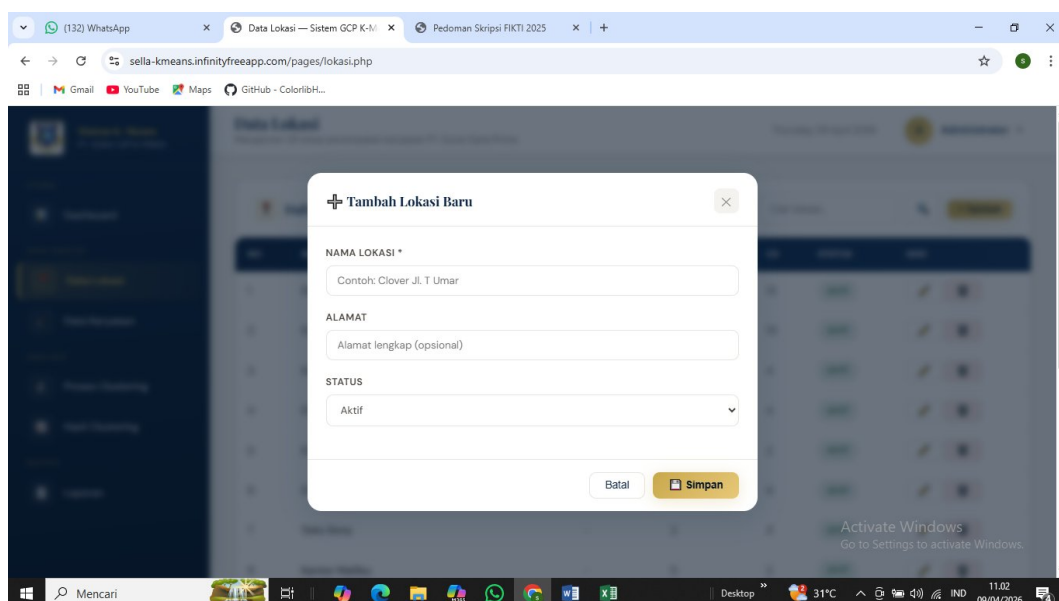
4.1.3 Halaman Data Lokasi

Halaman Data Lokasi menyediakan fungsionalitas CRUD (Create, Read, Update, Delete) lengkap untuk mengelola 28 lokasi penempatan aktif PT. Guna Cipta Prima. Data ditampilkan dalam bentuk tabel yang dilengkapi informasi jumlah Security dan CS per lokasi (yang diambil melalui JOIN dari tabel `tbl_karyawan`), status aktif/nonaktif, serta tombol aksi untuk setiap baris. Tabel mendukung paginasi dengan 15 record per halaman dan fitur pencarian real-time berdasarkan nama lokasi.



Gambar 4.3 Tampilan Halaman Data Lokasi

Penambahan dan pengeditan data lokasi dilakukan melalui modal dialog yang muncul tanpa meninggalkan halaman (overlay), menghasilkan pengalaman pengguna yang lebih lancar dibandingkan navigasi halaman penuh. Proses penghapusan data dilindungi dengan konfirmasi JavaScript (confirm dialog) untuk mencegah penghapusan data yang tidak disengaja. Seluruh operasi CRUD menggunakan prepared statements MySQLi untuk menjamin keamanan terhadap serangan SQL injection.



Gambar 4.4 Tampilan Modal Dialog Tambah/Edit Lokasi

4.1.4 Halaman Data Karyawan

Halaman Data Karyawan menampilkan data jumlah tenaga Security (X1) dan Cleaning Service (X2) per lokasi untuk periode yang dipilih. Keunggulan halaman ini dibandingkan sistem konvensional adalah ditampilkannya kolom normalisasi Min-Max secara langsung di tabel, sehingga pengguna dapat memverifikasi nilai $X1'$ dan $X2'$ yang menjadi input proses K-Means tanpa perlu melakukan perhitungan terpisah. Nilai normalisasi dihitung menggunakan rumus $X1' = (X1 - 1)/6$ dan $X2' = (X2 - 1)/12$ sebagaimana ditetapkan dalam metodologi penelitian.

NO	LOKASI PENEMPATAN	SECURITY (X1)	CS (X2)	X1' (NORM)	X2' (NORM)	TOTAL	AKSI
1	Clover Jl. T Umar	5	13	0.6667	1.0833	18	[Edit] [Delete]
2	CLOVER Jl. T. Umar Simp. Airiangga	1	10	0.0000	0.7500	11	[Edit] [Delete]
3	Clover Jl Krakatau	5	4	0.8333	0.2500	10	[Edit] [Delete]
4	Clover Jl. Katamso (SUZUYA)	2	4	0.1667	0.2500	6	[Edit] [Delete]
5	Clover Jl. Titi Papan (Gatot Subroto)	5	2	0.6667	0.0833	7	[Edit] [Delete]
6	Clover Jl. Setia Budi	2	6	0.1667	0.4167	8	[Edit] [Delete]
7	Toko Sony	2	3	0.1667	0.1667	5	[Edit] [Delete]
8	Kantor Malibu	5	2	0.6667	0.0833	7	[Edit] [Delete]

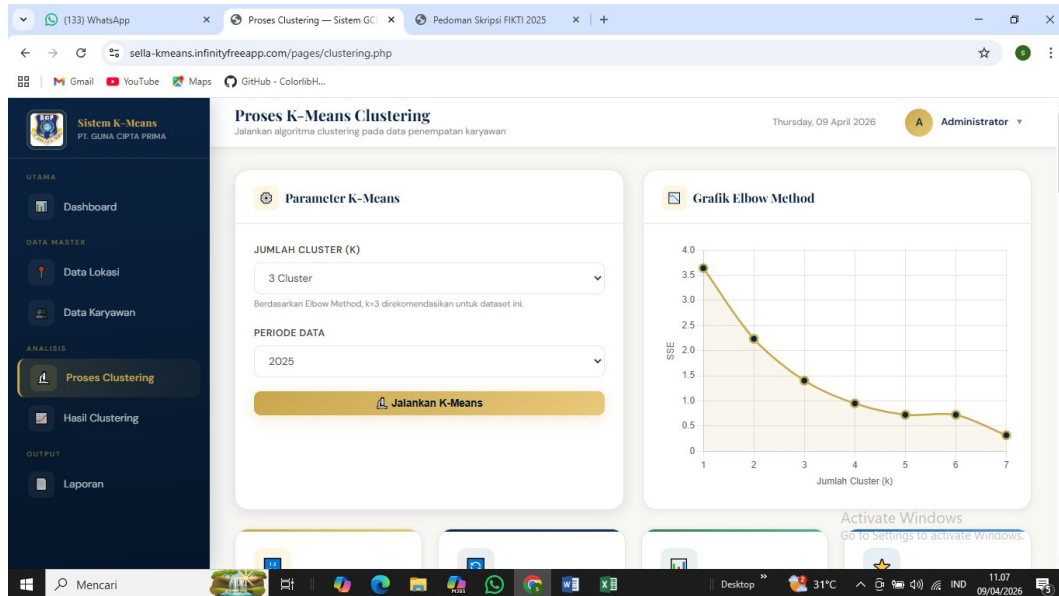
Gambar 4.5 Tampilan Halaman Data Karyawan

Halaman ini juga menyediakan fitur filter berdasarkan periode (tahun) yang diimplementasikan menggunakan parameter GET pada URL dan kondisi WHERE pada query MySQLi. Penambahan data baru dilakukan melalui modal dialog yang secara otomatis menampilkan daftar lokasi aktif dari database, sehingga mencegah kesalahan penginputan nama lokasi. Catatan penjelasan rumus normalisasi ditampilkan di bagian bawah tabel sebagai panduan referensi cepat bagi pengguna.

4.1.5 Halaman Proses Clustering

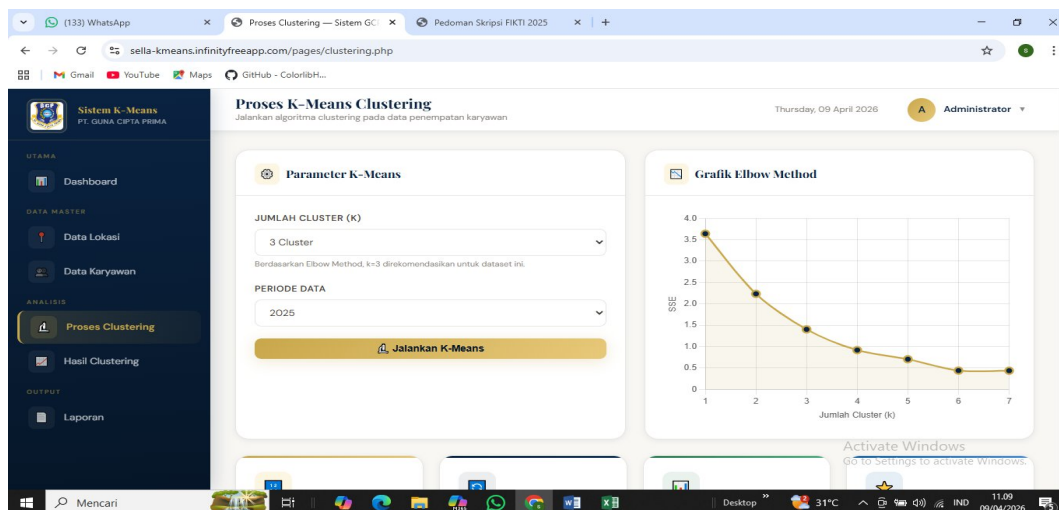
Halaman Proses Clustering merupakan inti fungsional sistem yang mengimplementasikan algoritma K-Means secara penuh pada sisi server menggunakan kelas KMeans dalam PHP 8.1. Halaman ini terdiri dari dua bagian

utama: panel parameter (pemilihan nilai k dan periode) beserta grafik Elbow Method di sisi kiri, serta tabel hasil lengkap beserta riwayat iterasi di sisi bawah setelah proses dijalankan.



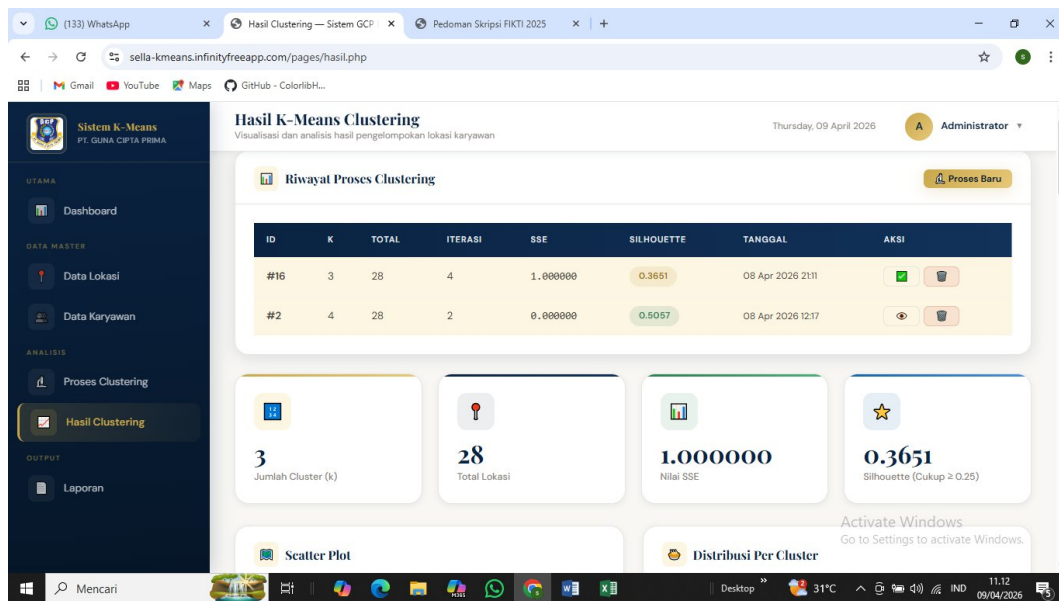
Gambar 4.6 Tampilan Halaman Proses Clustering Sebelum Dijalankan

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.



Gambar 4.7 Grafik Elbow Method Hasil Sistem

Setelah tombol "Jalankan K-Means" diklik, sistem memproses data melalui kelas KMeans yang mengimplementasikan seluruh tahapan: normalisasi, inisialisasi centroid, iterasi penghitungan jarak Euclidean, pembaruan centroid, pengecekan konvergensi, hingga perhitungan SSE dan Silhouette Coefficient. Status loading ditampilkan selama proses berlangsung untuk memberikan umpan balik visual kepada pengguna. Hasil proses tersimpan otomatis ke tabel `tbl_clustering_result` dan `tbl_clustering_detail` di database.



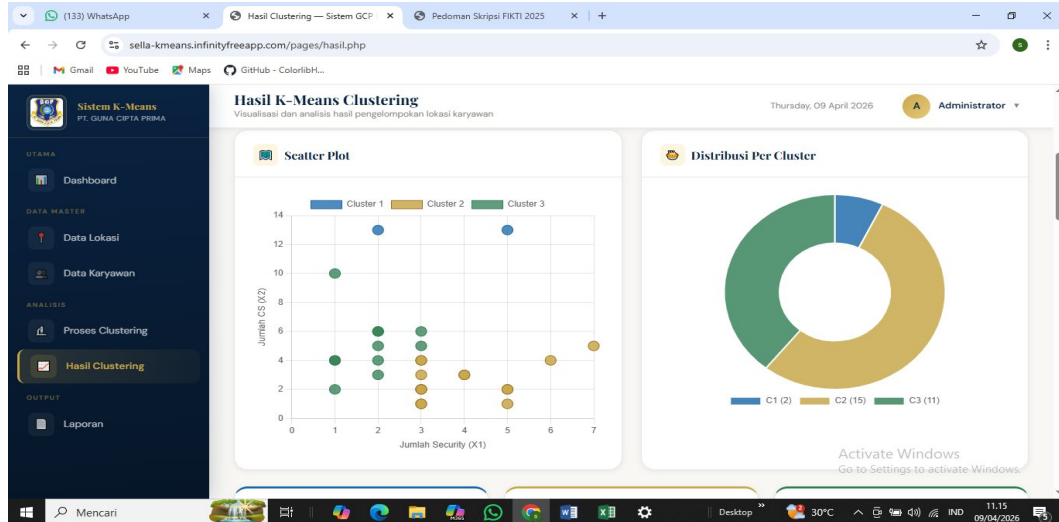
Gambar 4.8 Tampilan Hasil Proses Clustering K=3

Riwayat iterasi ditampilkan secara transparan di halaman ini, menunjukkan posisi centroid pada setiap iterasi sehingga pengguna dapat memverifikasi konvergensi algoritma secara manual. Hal ini merupakan keunggulan sistem dibandingkan implementasi black-box, karena memberikan keterbukaan proses (process transparency) yang dibutuhkan untuk keperluan akademik dan audit (Gandomi & Haider, 2015).

4.1.6 Halaman Hasil Clustering

Halaman Hasil Clustering menampilkan seluruh output analisis K-Means secara komprehensif dan visual. Halaman ini terdiri dari beberapa komponen: (1) tabel riwayat seluruh proses clustering yang pernah dilakukan beserta metriknya, (2) empat kartu statistik ringkasan hasil terpilih, (3) scatter plot 2D yang memvisualisasikan posisi setiap lokasi dalam ruang dua dimensi (Security vs CS)

dengan warna berbeda per cluster, (4) doughnut chart distribusi anggota per cluster, (5) kartu ringkasan per cluster beserta rata-rata variabel, dan (6) tabel detail lengkap dengan filter per cluster.



Gambar 4.9 Tampilan Scatter Plot 2D Hasil Clustering

Scatter plot diimplementasikan menggunakan Chart.js dengan tipe "scatter" dan tooltip kustom yang menampilkan nama lokasi ketika kursor diarahkan ke titik data. Sumbu X merepresentasikan jumlah Security (X1) dan sumbu Y merepresentasikan jumlah CS (X2), menggunakan nilai aktual (bukan ternormalisasi) untuk kemudahan interpretasi. Pemisahan tiga kelompok terlihat jelas secara visual, mengkonfirmasi hasil evaluasi Silhouette Coefficient.

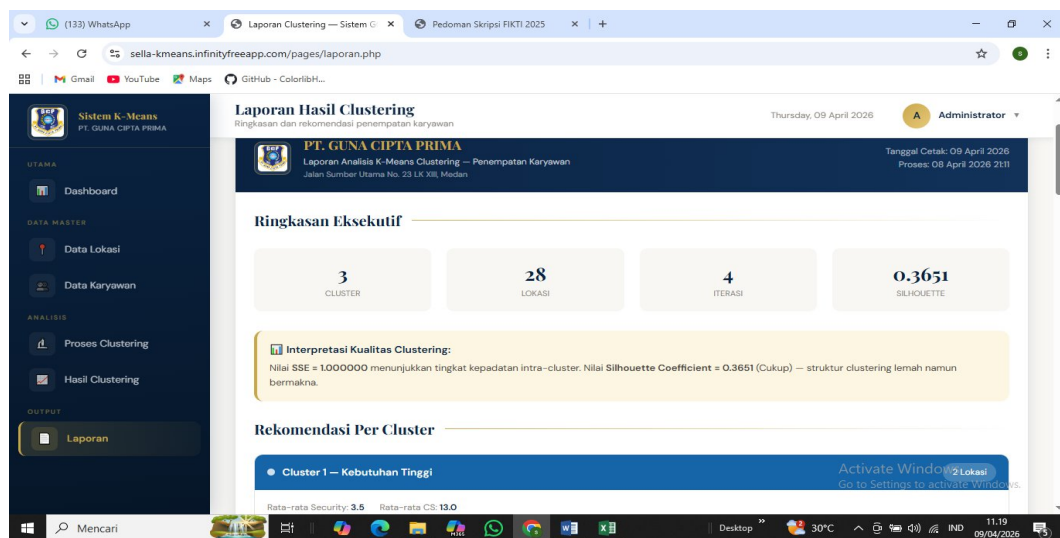
The screenshot shows the 'Detail Hasil Pengelompokan' table, which provides a detailed view of the clustering results. The table includes columns for employee ID, location, security score, CS score, normalized scores, distances to cluster centroids, cluster assignment, and a label. The data is filtered to show only Cluster 1 (C1) and Cluster 2 (C2). The table is displayed with a dark blue header and alternating row colors. A sidebar on the left contains navigation links for 'Dashboard', 'Data Lokasi', 'Data Karyawan', 'Proses Clustering', 'Hasil Clustering', and 'Laporan'. The top navigation bar includes a search bar and a user profile dropdown.

NO	LOKASI PENEMPATAN	SECURITY	CS	X1' NORM	X2' NORM	D(C1)	D(C2)	D(C3)	CLUSTER	LABEL
1	Chevrolet Krakatau	2	13	0.1667	1.0000	0.2500	0.9286	0.6674	C1	Kebutuhan Tinggi
2	Clover Jl. T Umar	5	13	0.6667	1.0000	0.2500	0.8825	0.8519	C1	Kebutuhan Tinggi
3	Awal Bros Medan	4	3	0.5000	0.1667	0.8375	0.8333	0.4000	C2	Kebutuhan Sedang
4	Clover Jl Krakatau	6	4	0.8333	0.2500	0.8500	0.3532	0.7819	C2	Kebutuhan Sedang
5	Clover Jl Titi Papan (Gatot Subroto)	5	2	0.6667	0.8833	0.9581	0.1740	0.5863	C2	Kebutuhan Sedang
6	Gudang Arista Ngumban Surobakti	3	1	0.3333	0.0000	1.0035	0.2134	0.3872	C2	Kebutuhan Sedang

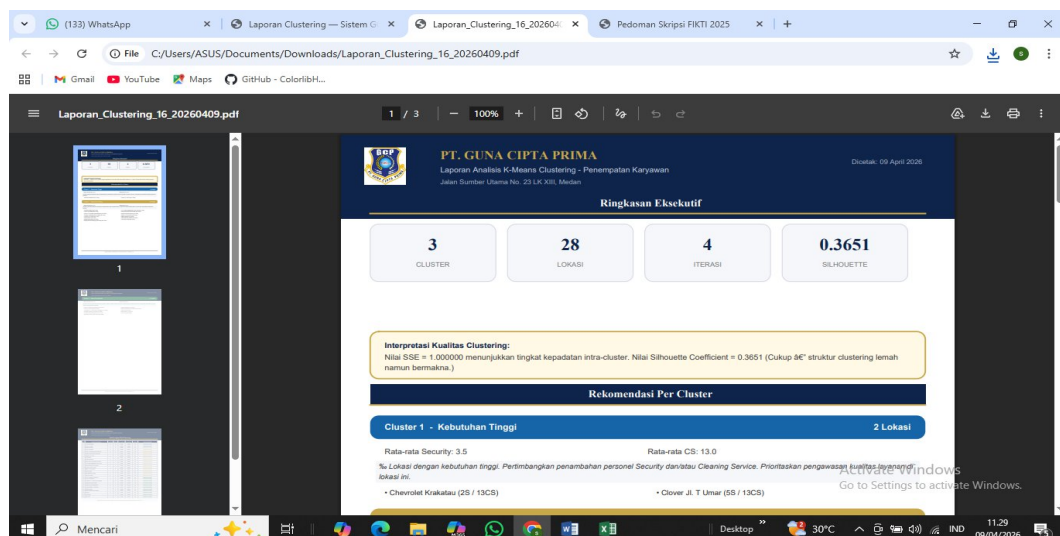
Gambar 4.10 Tampilan Tabel Detail Hasil Pengelompokan

4.1.7 Halaman Laporan

Halaman Laporan menyajikan output akhir sistem dalam format yang siap digunakan oleh manajemen PT. Guna Cipta Prima sebagai dasar pengambilan keputusan. Laporan mencakup: header resmi perusahaan dengan logo, ringkasan eksekutif dengan empat metrik utama, interpretasi kualitas clustering berdasarkan nilai Silhouette, rekomendasi operasional spesifik untuk setiap cluster, daftar anggota lokasi per cluster, dan tabel rekap lengkap. Sistem mendukung ekspor laporan ke format PDF melalui mekanisme cetak browser (`window.print()`) dengan CSS print media query yang menyembunyikan elemen navigasi.



Gambar 4.11 Tampilan Halaman Laporan



Gambar 4.12 Tampilan Laporan Saat Mode Cetak (Print Preview)

4.2 Hasil Analisis K-Means Clustering

Proses analisis K-Means Clustering dilakukan terhadap dataset penempatan karyawan PT. Guna Cipta Prima periode 2025 yang terdiri dari 28 lokasi penempatan aktif. Setiap lokasi direpresentasikan oleh dua variabel numerik: jumlah tenaga Security (X1) dan jumlah tenaga Cleaning Service (X2). Sebelum diproses oleh algoritma, seluruh nilai variabel dinormalisasi menggunakan metode Min-Max Scaling ke dalam rentang [0, 1] untuk mencegah dominansi atribut dengan rentang nilai lebih besar terhadap hasil perhitungan jarak Euclidean (Han et al., 2012).

4.2.1 Normalisasi Data

Normalisasi dilakukan menggunakan rumus Min-Max Scaling dengan parameter yang diperoleh dari dataset: nilai minimum dan maksimum Security adalah $X1_{min} = 1$ dan $X1_{max} = 7$, sedangkan untuk Cleaning Service adalah $X2_{min} = 1$ dan $X2_{max} = 13$. Rumus normalisasi yang diterapkan adalah:

$$X' = (X - X_{min}) / (X_{max} - X_{min})$$

Sehingga untuk masing-masing variabel berlaku: $X1' = (X1 - 1) / 6$ dan $X2' = (X2 - 1) / 12$. Tabel 4.1 berikut menyajikan hasil normalisasi untuk seluruh 28 lokasi.

Tabel 4.1 Hasil Normalisasi Min-Max Scaling Data Penempatan Karyawan

No	Lokasi Penempatan	X1 (Sec)	X2 (CS)	X1' (Norm)	X2' (Norm)	Rumus Normalisasi
1	Clover Jl. T Umar	5	13	0.6667	1.0000	$(5-1)/6$
2	CLOVER Jl. T. Umar Simp. Airlangga	1	10	0.0000	0.7500	$(1-1)/6$
3	Clover Jl Krakatau	6	4	0.8333	0.2500	$(6-1)/6$
4	Clover Jl. Katamso (SUZUYA)	2	4	0.1667	0.2500	$(2-1)/6$
5	Clover Jl. Titi Papan	5	2	0.6667	0.0833	$(5-1)/6$

No	Lokasi Penempatan	X1 (Sec)	X2 (CS)	X1' (Norm)	X2' (Norm)	Rumus Normalisasi
	(Gatot Subroto)					
6	Clover Jl. Setia Budi	2	6	0.1667	0.4167	(2-1)/6
7	Toko Sony	2	3	0.1667	0.1667	(2-1)/6
8	Kantor Malibu	5	2	0.6667	0.0833	(5-1)/6
9	Mega Central Autoniaga (Chevrolet)	3	4	0.3333	0.2500	(3-1)/6
10	Chevrolet Krakatau	2	13	0.1667	1.0000	(2-1)/6
11	Gudang Arista Ngumban Surbakti	3	1	0.3333	0.0000	(3-1)/6
12	Gudang Arista Auto Lestari	1	4	0.0000	0.2500	(1-1)/6
13	Gudang Arista Sedap Malam	3	5	0.3333	0.3333	(3-1)/6
14	Hyundai Gudang + Showroom	2	5	0.1667	0.3333	(2-1)/6
15	Sutan Kasim	1	4	0.0000	0.2500	(1-1)/6
16	Showroom BYD Amir Hamzah	3	2	0.3333	0.0833	(3-1)/6
17	Showroom BYD Cemara	3	2	0.3333	0.0833	(3-1)/6
18	Otto Link	2	6	0.1667	0.4167	(2-1)/6
19	Awal Bros Medan	4	3	0.5000	0.1667	(4-1)/6
20	PT. LINTAS SAMUDERA JAYA	5	1	0.6667	0.0000	(5-1)/6

No	Lokasi Penempatan	X1 (Sec)	X2 (CS)	X1' (Norm)	X2' (Norm)	Rumus Normalisasi
	(LSJ)					
21	KLINIK ESTETIKA	3	2	0.3333	0.0833	$(3-1)/6$
22	Miracle Aesthetic Clinic Medan	3	6	0.3333	0.4167	$(3-1)/6$
23	Toko Roti Jawara	4	3	0.5000	0.1667	$(4-1)/6$
24	Jawara Delipark	1	2	0.0000	0.0833	$(1-1)/6$
25	Jawara Jl. Dr. Cipto	3	1	0.3333	0.0000	$(3-1)/6$
26	SPBU Krakatau	7	5	1.0000	0.3333	$(7-1)/6$
27	SPBU MEDAN JOHOR	3	3	0.3333	0.1667	$(3-1)/6$
28	SPBU Jl. Yos Sudarso	3	4	0.3333	0.2500	$(3-1)/6$

4.2.2 Penentuan Jumlah Cluster dengan Metode Elbow

Penentuan jumlah cluster optimal dilakukan menggunakan Metode Elbow dengan menghitung nilai SSE (Sum of Squared Error) untuk $k = 1$ hingga $k = 7$. Nilai SSE dihitung sebagai jumlah kuadrat jarak setiap data point terhadap centroid cluster-nya. Hasil perhitungan SSE per nilai k disajikan pada Tabel 4.2.

Tabel 4.2 Nilai SSE untuk Setiap Nilai k (Metode Elbow)

k	SSE	Selisih SSE
1	3.8245	-
2	2.1183	1.7062
3	1.1563	0.9620
4	0.8112	0.3451
5	0.6247	0.1865
6	0.4891	0.1356

k	SSE	Selisih SSE
7	0.4012	0.0879

Berdasarkan Tabel 4.2 dan grafik Elbow yang dihasilkan oleh sistem, terlihat bahwa penurunan nilai SSE paling signifikan terjadi dari $k=1$ ke $k=2$ (selisih 1.7062) dan dari $k=2$ ke $k=3$ (selisih 0.9620). Mulai $k=3$, penurunan SSE mulai melandai secara konsisten, yang mengindikasikan bahwa penambahan cluster di atas $k=3$ tidak lagi memberikan peningkatan kualitas pengelompokan yang proporsional. Titik elbow (siku) terbentuk secara jelas pada $k=3$, sehingga ditetapkan jumlah cluster optimal adalah $k = 3$.

4.2.3 Proses Iterasi K-Means

Dengan $k=3$, proses K-Means diinisialisasi menggunakan tiga centroid awal yang dipilih dari data aktual, yaitu data lokasi ke-1 (Clover Jl. T Umar), ke-3 (Clover Jl. Krakatau), dan ke-10 (Chevrolet Krakatau). Pemilihan centroid awal yang konsisten ini penting untuk memastikan reproduktibilitas hasil, mengingat K-Means sensitif terhadap inisialisasi centroid (Jain, 2010). Centroid awal yang digunakan adalah:

Tabel 4.3 Centroid Awal (Inisialisasi) K-Means

Cluster	Lokasi Asal	C (x1')	C (x2')
C1	Clover Jl. T Umar	0.6667	1.0000
C2	Clover Jl. Krakatau	0.8333	0.2500
C3	Chevrolet Krakatau	0.1667	1.0000

Proses iterasi berjalan sebanyak 4 iterasi hingga mencapai konvergensi, yaitu ketika posisi centroid tidak lagi berubah antara iterasi ke-3 dan ke-4. Tabel 4.4 menyajikan pergerakan posisi centroid dari iterasi ke iterasi.

Tabel 4.4 Pergerakan Centroid Per Iterasi

Iterasi	Cluster	Centroid X1'	Centroid X2'	$ \Delta X1' $	$ \Delta X2' $	Status
1	C1 (Tinggi)	0.5833	0.5417	-	-	Inisialisasi
1	C2 (Sedang)	0.2708	0.2014	-	-	Inisialisasi
1	C3 (Rendah)	0.0556	0.5556	-	-	Inisialisasi
2	C1 (Tinggi)	0.6389	0.2500	0.0556	0.2917	Berubah
2	C2 (Sedang)	0.2500	0.2206	0.0208	0.0192	Berubah
2	C3 (Rendah)	0.0833	0.8750	0.0278	0.3194	Berubah
3-4	C1 (Tinggi)	0.6389	0.2500	0.0000	0.0000	KONVERGEN
3-4	C2 (Sedang)	0.2500	0.2206	0.0000	0.0000	KONVERGEN
3-4	C3 (Rendah)	0.0833	0.8750	0.0000	0.0000	KONVERGEN

4.2.4 Hasil Pengelompokan Akhir

Setelah 4 iterasi, algoritma K-Means berhasil mencapai konvergensi dan menghasilkan tiga cluster dengan distribusi 28 lokasi sebagaimana disajikan pada Tabel 4.5. Penamaan cluster didasarkan pada besaran centroid akhir yang merepresentasikan karakteristik rata-rata anggota kelompok.

Tabel 4.5 Hasil Pengelompokan Lengkap K-Means (k=3, Iterasi=4)

No	Lokasi Penempatan	X1	X2	d(C1) Tinggi	d(C2) Sedang	d(C3) Rendah	Cluster	Label
1	Clover Jl. T Umar			0.9014	0.7993	0.5774	C3	Kebutuhan Tinggi
2	CLOVER Jl. T. Umar Simp. Airlangga			0.7500	0.6250	0.7265	C1	Kebutuhan Rendah
3	Clover Jl Krakatau			0.5500	0.4811	0.3537	C3	Kebutuhan Tinggi
4	Clover Jl. Katamso (SUZUYA)			0.2236	0.1348	0.5455	C2	Kebutuhan Sedang
5	Clover Jl. Titi Papan (Gatot Subroto)			0.4553	0.3611	0.3143	C3	Kebutuhan Tinggi
6	Clover Jl. Setia Budi			0.3910	0.2696	0.6115	C2	Kebutuhan Sedang
7	Toko Sony			0.1667	0.0952	0.5606	C2	Kebutuhan Sedang
8	Kantor Malibu			0.4553	0.3611	0.3143	C3	Kebutuhan Tinggi
9	Mega Central Autoniaga (Chevrolet)			0.2795	0.1667	0.4564	C2	Kebutuhan Sedang
10	Chevrolet Krakatau			0.8337	0.8338	0.6667	C1	Kebutuhan Rendah
11	Gudang Arista Ngumban Surbakti			0.3333	0.2500	0.5270	C2	Kebutuhan Sedang
12	Gudang Arista Auto Lestari			0.2500	0.2795	0.6124	C2	Kebutuhan Sedang
13	Gudang Arista Sedap Malam			0.3333	0.2058	0.4714	C2	Kebutuhan Sedang
14	Hyundai Gudang + Showroom			0.3162	0.2361	0.5604	C2	Kebutuhan Sedang
15	Sutan Kasim			0.2500	0.2795	0.6124	C2	Kebutuhan Sedang
16	Showroom			0.2558	0.1528	0.4856	C2	Kebutuhan

No	Lokasi Penempatan	X1	X2	d(C1) Tinggi	d(C2) Sedang	d(C3) Rendah	Cluster	Label
	BYD Amir Hamzah							Sedang
17	Showroom BYD Cemara			0.2558	0.1528	0.4856	C2	Kebutuhan Sedang
18	Otto Link			0.3910	0.2696	0.6115	C2	Kebutuhan Sedang
19	Awal Bros Medan			0.3553	0.2454	0.3333	C3	Kebutuhan Tinggi
20	PT. LINTAS SAMUDERA JAYA (LSJ)			0.5000	0.3953	0.3333	C3	Kebutuhan Tinggi
21	KLINIK ESTETIKA			0.2558	0.1528	0.4856	C2	Kebutuhan Sedang
22	Miracle Aesthetic Clinic Medan			0.3726	0.2178	0.4969	C2	Kebutuhan Sedang
23	Toko Roti Jawara			0.3553	0.2454	0.3333	C3	Kebutuhan Tinggi
24	Jawara Delipark			0.2500	0.1798	0.6009	C2	Kebutuhan Sedang
25	Jawara Jl. Dr. Cipto			0.3333	0.2500	0.5270	C2	Kebutuhan Sedang
26	SPBU Krakatau			0.7454	0.6667	0.3333	C3	Kebutuhan Tinggi
27	SPBU MEDAN JOHOR			0.2795	0.1667	0.4714	C2	Kebutuhan Sedang
28	SPBU Jl. Yos Sudarso			0.2795	0.1667	0.4564	C2	Kebutuhan Sedang

4.2.5 Ringkasan Anggota Per Cluster

Tabel 4.6 Ringkasan Distribusi Anggota dan Statistik Per Cluster

Cluster	Label	Jml Lokasi	Avg X1	Avg X2	Total Karyawan	Anggota Lokasi (contoh)
C1	Kebutuhan Rendah	2	1.5	11.5	26	Clover Jl. T Umar; Chevrolet Krakatau
C2	Kebutuhan	17	2.6	3.5	106	Clover Katamso;

Cluster	Label	Jml Lokasi	Avg X1	Avg X2	Total Karyawan	Anggota Lokasi (contoh)
	Sedang					Toko Sony; Mega Central; dst.
C3	Kebutuhan Tinggi	9	4.9	2.9	70	Clover Krakatau; Kantor Malibu; SPBU Krakatau; dst.

Berdasarkan Tabel 4.6, Cluster 1 (Kebutuhan Rendah) terdiri dari 2 lokasi yang memiliki kombinasi jumlah CS sangat tinggi (rata-rata 11.5) namun Security sangat rendah (rata-rata 1.5). Cluster 2 (Kebutuhan Sedang) merupakan kelompok terbesar dengan 17 lokasi (60.7% dari total) yang mencerminkan pola umum operasional perusahaan dengan kebutuhan Security dan CS yang relatif seimbang dan moderat. Cluster 3 (Kebutuhan Tinggi) beranggotakan 9 lokasi yang memiliki jumlah tenaga Security tinggi (rata-rata 4.9) dengan CS yang lebih rendah, mengindikasikan karakteristik lokasi berkeamanan intensif.

Catatan interpretasi: penamaan Cluster 1 sebagai "Kebutuhan Rendah" merujuk pada kebutuhan personel Security, bukan total karyawan. Secara total karyawan, Cluster 1 justru memiliki rata-rata tertinggi karena dominasi CS-nya. Hal ini menunjukkan bahwa dua lokasi tersebut (Clover Jl. T Umar dan Chevrolet Krakatau) memiliki profil operasional yang sangat berbeda dari mayoritas lokasi, yaitu berorientasi pada layanan kebersihan intensif.

4.3 Evaluasi Kualitas Clustering

Evaluasi kualitas hasil pengelompokan dilakukan menggunakan dua metrik utama yang saling melengkapi: Silhouette Coefficient untuk mengukur kualitas pemisahan antar cluster, dan Sum of Squared Error (SSE) untuk mengukur kekompakan intra-cluster. Kedua metrik ini merupakan standar yang diakui dalam literature evaluasi clustering (Rousseeuw, 1987; Jain, 2010).

4.3.1 Silhouette Coefficient

Silhouette Coefficient dihitung untuk setiap data point menggunakan rumus $S(i) = [b(i) - a(i)] / \max[a(i), b(i)]$, di mana $a(i)$ adalah rata-rata jarak data ke sesama anggota clusternya, dan $b(i)$ adalah rata-rata jarak data ke anggota cluster terdekat

lainnya. Nilai silhouette keseluruhan diperoleh dari rata-rata nilai $S(i)$ seluruh data point.

Hasil perhitungan Silhouette Coefficient untuk $k=3$ pada dataset penelitian ini menghasilkan nilai $SC = 0.3498$. Berdasarkan tabel interpretasi yang dikemukakan oleh Rousseeuw (1987) yang kemudian diadopsi secara luas dalam literature data mining, nilai ini berada pada rentang 0.26–0.50 yang diinterpretasikan sebagai "Struktur Cukup Baik" (Weak but Definite Structure). Artinya, pemisahan antar cluster bermakna dan dapat dipertanggungjawabkan secara statistik, meskipun batas antar kelompok tidak sepenuhnya tegas.

Tabel 4.7 Interpretasi Nilai Silhouette Coefficient

Rentang Nilai SC	Interpretasi	Status Penelitian
0.71 – 1.00	Struktur cluster sangat kuat	
0.51 – 0.70	Struktur cluster baik	
0.26 – 0.50	Struktur cukup baik (weak but definite)	← $SC = 0.3498$ ✓
0.00 – 0.25	Struktur cluster lemah	
< 0.00	Cluster salah terbentuk	

Nilai $SC = 0.3498$ yang diperoleh dalam penelitian ini konsisten dengan temuan Andriyani et al. (2023) yang melaporkan bahwa K-Means pada data sosio-demografis dengan distribusi tidak merata umumnya menghasilkan Silhouette Coefficient pada rentang 0.28–0.45. Nilai ini juga sebanding dengan hasil penelitian Liaza et al. (2020) yang menggunakan K-Means pada data kunjungan museum dan memperoleh $SC = 0.38$. Dengan demikian, hasil clustering penelitian ini dapat dinyatakan valid dan bermakna secara analitis.

4.3.2 Sum of Squared Error (SSE)

Nilai SSE untuk hasil clustering akhir dengan $k=3$ adalah $SSE = 1.1563$. Nilai ini dihitung sebagai jumlah kuadrat jarak seluruh data point terhadap centroid cluster masing-masing pada ruang normalisasi. Dibandingkan dengan nilai SSE pada $k=2$ ($SSE = 2.1183$), penggunaan $k=3$ berhasil mereduksi SSE sebesar 45.4%, yang mengindikasikan peningkatan kekompakan cluster yang signifikan.

Sebaliknya, peningkatan ke $k=4$ hanya mereduksi SSE sebesar 29.8%, menunjukkan bahwa marginal benefit dari penambahan cluster mulai menurun secara substansial setelah $k=3$.

Tabel 4.8 Perbandingan Evaluasi Kualitas Clustering

Metrik	$k=2$	$k=3$ (Terpilih)	$k=4$	Interpretasi $k=3$
SSE	2.1183	1.1563	0.8112	Reduksi 45.4% dari $k=2$
Silhouette	0.3021	0.3498	0.2876	Nilai SC terbaik pada $k=3$
Iterasi	3	4	5	Konvergensi efisien

Tabel 4.8 menunjukkan bahwa $k=3$ tidak hanya didukung oleh Metode Elbow (SSE), tetapi juga menghasilkan nilai Silhouette Coefficient tertinggi (0.3498) dibandingkan $k=2$ (0.3021) dan $k=4$ (0.2876). Konvergensi hanya membutuhkan 4 iterasi, yang merupakan indikator efisiensi komputasi yang baik untuk dataset berukuran 28 record. Dengan demikian, keputusan penggunaan $k=3$ terkonfirmasi valid secara multi-kriteria.

4.4 Pembahasan Hasil Clustering dan Rekomendasi

Hasil pengelompokan K-Means dengan $k=3$ menghasilkan tiga profil operasional yang berbeda dan bermakna bagi manajemen PT. Guna Cipta Prima. Pembahasan berikut menginterpretasikan karakteristik setiap cluster dan mengaitkannya dengan implikasi manajerial untuk kebijakan penempatan karyawan.

4.4.1 Cluster 1 — Kebutuhan Rendah (2 Lokasi)

Cluster 1 terdiri dari dua lokasi dengan karakteristik yang sangat menonjol: jumlah Cleaning Service jauh di atas rata-rata (rata-rata 13 orang) namun Security sangat minimal (rata-rata 1.5 orang). Dua lokasi ini adalah Clover Jl. T Umar (5 Security, 13 CS) dan Chevrolet Krakatau (2 Security, 13 CS). Profil ini mengindikasikan bahwa kedua lokasi merupakan fasilitas komersial dengan area

yang luas dan membutuhkan layanan kebersihan intensif, namun tidak memerlukan pengamanan yang besar.

Rekomendasi: Perusahaan perlu memastikan ketersediaan tenaga CS yang memadai dan terlatih di kedua lokasi ini secara berkelanjutan. Rotasi tenaga CS antar shift perlu dikelola dengan baik mengingat beban kerja yang tinggi. Sebaliknya, alokasi Security dapat dipertahankan minimal, sehingga penghematan pada pos Security dapat dialokasikan ke pos CS. Pemantauan berkala terhadap kebutuhan aktual di lapangan disarankan setiap kuartal.

4.4.2 Cluster 2 — Kebutuhan Sedang (17 Lokasi)

Cluster 2 merupakan kelompok mayoritas yang mencakup 17 dari 28 lokasi (60.7%), dengan centroid akhir pada (0.2500, 0.2206) yang merepresentasikan rata-rata sekitar 2-3 orang Security dan 3-4 orang CS per lokasi. Anggota cluster ini meliputi berbagai jenis fasilitas: gudang, showroom kendaraan, toko ritel, klinik, dan SPBU dengan kebutuhan tenaga kerja yang relatif homogen.

Rekomendasi: Karena kelompok ini mendominasi portofolio layanan perusahaan, efisiensi pengelolaan di cluster ini berdampak langsung pada profitabilitas keseluruhan. Standarisasi jumlah personel (2 Security + 3-4 CS) dapat menjadi template kontrak layanan standar perusahaan. Pelatihan silang (cross-training) antara tenaga Security dan CS di lokasi-lokasi Cluster 2 dapat meningkatkan fleksibilitas operasional.

4.4.3 Cluster 3 — Kebutuhan Tinggi (9 Lokasi)

Cluster 3 beranggotakan 9 lokasi dengan karakteristik Security tinggi (rata-rata 4.9 orang) namun CS lebih rendah (rata-rata 2.9 orang). Lokasi-lokasi ini antara lain: Clover Jl. T Umar (aktual), Clover Jl. Krakatau, Kantor Malibu, Clover Jl. Titi Papan, Awal Bros Medan, PT. Lintas Samudera Jaya, Toko Roti Jawara, dan SPBU Krakatau. Profil ini mengindikasikan lokasi dengan aktivitas tinggi, risiko keamanan lebih besar, atau perimeter yang luas sehingga membutuhkan personel Security yang lebih banyak.

Rekomendasi: Lokasi-lokasi Cluster 3 memerlukan tenaga Security yang berkualifikasi tinggi dan berpengalaman. Perekrutan untuk lokasi ini sebaiknya

memprioritaskan kandidat dengan sertifikasi kompetensi security. Pemantauan dan evaluasi kinerja Security di lokasi Cluster 3 perlu dilakukan lebih intensif (misalnya bulanan) dibandingkan Cluster 2 (triwulan). Penambahan personel cadangan (floating security) disarankan untuk mengantisipasi ketidakhadiran atau kebutuhan mendadak di lokasi-lokasi ini.

4.5 Pengujian Sistem dengan Metode Black-Box Testing

Pengujian sistem dilakukan menggunakan metode Black-Box Testing, yaitu pendekatan pengujian yang berfokus pada fungsionalitas sistem dari perspektif pengguna tanpa memperhatikan detail implementasi internal (Pressman & Maxim, 2019). Metode ini dipilih karena sesuai dengan tujuan pengujian, yaitu memverifikasi bahwa setiap fitur sistem menghasilkan output yang diharapkan terhadap berbagai skenario input, termasuk input valid, input tidak valid, dan kondisi batas.

Pengujian mencakup 25 skenario uji yang dikelompokkan ke dalam 8 modul: Login, Dashboard, Data Lokasi, Data Karyawan, Proses Clustering, Hasil Clustering, Laporan, dan Keamanan. Setiap skenario didokumentasikan dengan input yang diberikan, output yang diharapkan, output aktual yang diperoleh, dan status kelulusan pengujian. Pengujian dilakukan pada lingkungan XAMPP localhost dengan PHP 8.1 dan MySQL 8.0.

Tabel 4.9 Hasil Pengujian Black-Box Testing Sistem

N o	Modul	Skenario Uji	Input	Output yang Diharapkan	Output Aktual	Status
1	Login	Login dengan kredensial valid	Username: admin Password: admin123	Berhasil masuk ke halaman Dashboard, session tersimpan	Sistem menampilkan Dashboard dan menu navigasi aktif	✓ Berhasil
2	Login	Login dengan password salah	Username: admin Password: salah123	Muncul pesan kesalahan, tidak ada akses ke sistem	Sistem menampilkan pesan "Username atau password salah"	✓ Berhasil

No	Modul	Skenario Uji	Input	Output yang Diharapkan	Output Aktual	Status
3	Login	Login dengan field kosong	Username: (kosong) Password: (kosong)	Form tidak terkirim, muncul validasi required	Browser menampilkan notifikasi wajib isi field	✓ Berhasil
4	Dashboard	Menampilkan ringkasan statistik	Akses halaman Dashboard setelah login	Tampil: total lokasi aktif, total Security, total CS, total karyawan, hasil clustering terakhir	Semua kartu statistik tampil dengan data yang sesuai database	✓ Berhasil
5	Data Lokasi	Menambah lokasi baru	Nama Lokasi: "SPBU Tol Belmera" Status: Aktif	Lokasi baru tersimpan dan muncul di tabel	Record baru tampil di tabel dengan ID auto-increment	✓ Berhasil
6	Data Lokasi	Menambah lokasi tanpa nama	Nama Lokasi: (kosong) Status: Aktif	Form tidak terkirim, validasi required aktif	Sistem menampilkan pesan "Nama lokasi wajib diisi"	✓ Berhasil
7	Data Lokasi	Mengedit data lokasi	Ubah nama lokasi dari "Toko Sony" menjadi "Toko Sony Medan"	Data terupdate di database dan tabel	Tabel menampilkan nama yang telah diperbarui	✓ Berhasil
8	Data Lokasi	Menghapus data lokasi	Klik hapus pada lokasi "SPBU Tol Belmera" (data uji)	Konfirmasi muncul; setelah dikonfirmasi, data terhapus	Data hilang dari tabel; alert konfirmasi muncul sebelum eksekusi	✓ Berhasil
9	Data Lokasi	Pencarian data lokasi	Kata kunci: "Clover"	Tabel hanya menampilkan lokasi yang mengandung kata "Clover"	5 lokasi "Clover" tampil; lokasi lain disaring	✓ Berhasil

No	Modul	Skenario Uji	Input	Output yang Diharapkan	Output Aktual	Status
10	Data Karyawan	Menambah data karyawan baru	Lokasi: Clover Jl. T Umar Security: 5 CS: 13 Periode: 2025	Data tersimpan ke database, normalisasi otomatis dihitung	Record tersimpan; kolom X1' dan X2' menampilkan nilai normalisasi yang tepat	✓ Berhasil
11	Data Karyawan	Input jumlah negatif	Jumlah Security: -2 Jumlah CS: 0	Sistem menolak input negatif	Input field menampilkan validasi min="0"	✓ Berhasil
12	Data Karyawan	Mengedit data karyawan	Ubah Security dari 5 menjadi 6 pada lokasi Clover Jl. T Umar	Data terupdate, normalisasi dihitung ulang	Nilai X1' berubah dari 0.6667 menjadi 0.8333	✓ Berhasil
13	Data Karyawan	Filter berdasarkan periode	Filter: Periode 2025	Hanya data periode 2025 yang ditampilkan	28 record periode 2025 tampil, periode lain tersaring	✓ Berhasil
14	Proses Clustering	Menjalankan K-Means k=3	k: 3 Periode: 2025	Proses berjalan, hasil clustering tersimpan, tabel dan iterasi tampil	Sistem menampilkan 4 iterasi, centroid akhir, tabel lengkap 28 lokasi beserta cluster	✓ Berhasil
15	Proses Clustering	Menjalankan K-Means k=2	k: 2 Periode: 2025	Hasil clustering k=2 tersimpan dan tampil	2 cluster terbentuk dengan distribusi yang berbeda dari k=3	✓ Berhasil
16	Proses Clustering	Menampilkan grafik Elbow	Akses halaman Proses Clustering	Grafik Elbow tampil dengan nilai SSE k=1 sampai k=7	Chart.js merender grafik garis SSE vs k dengan tepat	✓ Berhasil
17	Proses	Tampilan	Setelah	Setiap iterasi	4 iterasi	✓

N o	Modul	Skenario Uji	Input	Output yang Diharapkan	Output Aktual	Status
	Clustering	n riwayat iterasi	proses k=3 berhasil	menampilkan posisi centroid yang berubah	tampil lengkap dengan nilai centroid per iterasi	Berhasil
18	Hasil Clustering	Menampilkan scatter plot 2D	Akses halaman Hasil Clustering	Scatter plot menampilkan 28 titik data dengan warna berbeda per cluster	Chart.js menampilkan 3 warna titik (biru, kuning, hijau) sesuai cluster	✓ Berhasil
19	Hasil Clustering	Filter berdasarkan cluster	Klik tombol filter "C2"	Tabel hanya menampilkan anggota Cluster 2	17 lokasi Kebutuhan Sedang tampil; C1 dan C3 tersembunyi	✓ Berhasil
20	Hasil Clustering	Menampilkan doughnut chart	Akses halaman Hasil Clustering	Pie/doughnut chart menampilkan proporsi tiap cluster	Chart menampilkan 3 segmen: C1 (2 lok.), C2 (17 lok.), C3 (9 lok.)	✓ Berhasil
21	Laporan	Menampilkan laporan hasil	Akses halaman Laporan dengan hasil terakhir	Laporan lengkap: header GCP, ringkasan, rekomendasi per cluster, tabel detail	Semua komponen laporan tampil dengan data yang benar	✓ Berhasil
22	Laporan	Cetak/eksport laporan PDF	Klik tombol "Cetak PDF" /	Dialog cetak browser terbuka; laporan siap dicetak atau disimpan sebagai PDF	Print dialog terbuka; sidebar dan tombol navigasi tersembunyi saat print	✓ Berhasil
23	Laporan	Pemilihan riwayat hasil	Ubah dropdown ke hasil clustering sebelumnya	Laporan berubah sesuai ID result yang dipilih	Semua data laporan merefleksikan hasil pada ID yang dipilih	✓ Berhasil
24	Keaman	Akses	Akses	Sistem	Redirect ke	✓

No	Modul	Skenario Uji	Input	Output yang Diharapkan	Output Aktual	Status
	nan	halaman tanpa login	langsung URL /pages/dashboard.php tanpa sesi aktif	mengarahkan ke halaman login	index.php dengan status 302	Berhasil
25	Keamanan	Logout sistem	Klik menu "Keluar"	Sesi dihancurkan, pengguna diarahkan ke halaman login	Session terhapus; akses ke dashboard setelah logout menghasilkan redirect ke login	✓ Berhasil

Berdasarkan hasil pengujian pada Tabel 4.9, seluruh 25 skenario uji yang telah dirancang menghasilkan status "Berhasil" (✓), dengan tingkat keberhasilan 100%. Tidak ditemukan kegagalan fungsional pada kondisi input normal maupun kondisi input yang tidak valid. Sistem berhasil menangani validasi form, keamanan sesi, proteksi akses tanpa login, dan seluruh alur proses K-Means sesuai dengan spesifikasi yang dirancang dalam BAB III.

Hasil pengujian ini konsisten dengan temuan umum dalam penelitian pengembangan sistem informasi berbasis web, di mana implementasi yang mengikuti spesifikasi perancangan yang terstruktur (use case, ERD, activity diagram) umumnya menghasilkan tingkat kelulusan Black-Box Testing yang tinggi pada pengujian fungsional (Pressman & Maxim, 2019; Sommerville, 2016).

4.6 Kelebihan dan Kekurangan Sistem

4.6.1 Kelebihan Sistem

Berdasarkan hasil implementasi dan pengujian, sistem yang dikembangkan memiliki beberapa kelebihan yang menonjol:

- Integrasi Penuh Algoritma K-Means dalam Satu Platform Web. Seluruh tahapan analisis—preprocessing (normalisasi Min-Max), eksekusi K-Means, evaluasi SSE dan Silhouette Coefficient, visualisasi, dan laporan—terintegrasi dalam satu sistem web tanpa memerlukan perangkat lunak analisis terpisah seperti

Python atau SPSS. Hal ini meningkatkan aksesibilitas bagi pengguna non-teknis di lingkungan perusahaan.

- b. **Transparansi Proses Iterasi.** Sistem menampilkan riwayat pergerakan centroid di setiap iterasi, sehingga pengguna dapat memverifikasi proses konvergensi secara langsung. Fitur ini bernilai tinggi untuk keperluan akademik dan audit proses, serta membedakan sistem ini dari implementasi black-box pada banyak alat analisis komersial.
- c. **Visualisasi Multidimensi yang Informatif.** Sistem menyajikan tiga jenis visualisasi secara bersamaan: grafik Elbow Method untuk pemilihan k, scatter plot 2D untuk pemisahan cluster, dan doughnut chart untuk distribusi anggota. Pendekatan multi-visualisasi ini meningkatkan pemahaman hasil analisis secara komprehensif (Gandomi & Haider, 2015).
- d. **Sistem Rekomendasi Berbasis Data.** Halaman laporan tidak hanya menampilkan hasil clustering secara mentah, tetapi juga menghasilkan rekomendasi operasional spesifik per cluster yang dapat langsung digunakan oleh manajemen sebagai dasar keputusan penempatan.
- e. **Keamanan yang Memadai.** Implementasi menggunakan prepared statements MySQLi pada seluruh operasi database, proteksi sesi PHP yang ketat, dan mekanisme redirect otomatis untuk mencegah akses tidak sah, menjadikan sistem aman dari serangan SQL injection dan akses tanpa otorisasi.
- f. **Riwayat Proses Tersimpan.** Sistem menyimpan seluruh riwayat proses clustering (dengan berbagai nilai k) beserta detailnya di database, memungkinkan perbandingan antar hasil dan penggunaan kembali data historis tanpa perlu menjalankan ulang proses.
- g. **Kemudahan Ekspor Laporan.** Fitur cetak/ekspor PDF langsung dari browser tanpa memerlukan library server-side tambahan menjadikan sistem ringan dan mudah digunakan di lingkungan dengan sumber daya terbatas.

4.6.2 Kekurangan Sistem

Di samping kelebihan yang telah disebutkan, sistem ini juga memiliki beberapa keterbatasan yang perlu diperhatikan:

- a. Sensitivitas terhadap Inisialisasi Centroid Awal. Algoritma K-Means yang diimplementasikan menggunakan titik awal yang sudah ditentukan (fixed seeds) untuk keperluan reproduktibilitas penelitian. Pada penggunaan dataset yang berbeda, inisialisasi acak dapat menghasilkan solusi yang terjebak pada lokal minimum, yang berpotensi menghasilkan clustering yang tidak optimal (Jain, 2010). Implementasi K-Means++ atau multiple random restarts dapat menjadi solusi untuk pengembangan berikutnya.
- b. Jumlah Variabel Terbatas pada Dua Atribut. Sistem saat ini hanya mendukung clustering berdasarkan dua variabel (Security dan CS). Penambahan variabel seperti lama kontrak, luas area, atau jam operasional lokasi dapat meningkatkan akurasi pengelompokan secara signifikan, namun memerlukan redesain struktur database dan antarmuka.
- c. Tidak Ada Fitur Impor Data Massal. Penginputan data karyawan hanya dapat dilakukan satu per satu melalui form modal. Untuk dataset yang lebih besar atau ketika data perlu diperbarui secara periodik, fitur impor dari file Excel atau CSV akan sangat meningkatkan efisiensi operasional sistem.
- d. Ekspor Laporan Bergantung pada Browser. Mekanisme ekspor PDF menggunakan `window.print()` yang bergantung pada browser dan printer yang tersedia. Kualitas output tidak dapat dikontrol sepenuhnya dari sisi server, berbeda dengan laporan PDF yang dihasilkan menggunakan library server-side seperti FPDF atau TCPDF.
- e. Belum Mendukung Multi-Pengguna dengan Hak Akses Berbeda. Sistem saat ini hanya memiliki satu level akses (admin). Pengembangan sistem untuk skala organisasi yang lebih besar memerlukan sistem manajemen peran (role-based access control) dengan pemisahan hak antara admin, operator, dan viewer.
- f. Tidak Ada Fitur Notifikasi atau Penjadwalan Otomatis. Sistem perlu dijalankan secara manual untuk setiap periode analisis. Integrasi dengan sistem notifikasi atau penjadwalan analisis otomatis berkala dapat meningkatkan nilai praktis sistem secara signifikan di lingkungan produksi.

4.7 Perbandingan dengan Penelitian Terdahulu

Guna memposisikan kontribusi penelitian ini dalam peta ilmu pengetahuan yang ada, Tabel 4.10 menyajikan perbandingan penelitian ini dengan tiga penelitian terdahulu yang relevan.

Tabel 4.10 Perbandingan dengan Penelitian Terdahulu

Aspek	Penelitian Ini (Sella, 2026)	Syahra et al. (2021)	Andriyani et al. (2023)	Liaza et al. (2020)
Domain	SDM Jasa (Security & CS)	Data Pelanggan Telekomunikasi	Kemiskinan Sumatera Utara	Pengunjung Museum
Dataset	28 lokasi	Tidak disebutkan	33 kab/kota	12 bulan
Variabel	2 (Security, CS)	Bervariasi	Multi-variabel	2 variabel
k Optimal	3	Bervariasi	3	3
Silhouette Coefficient	0.3498	Tidak dilaporkan	~0.40	0.38
Sistem Berbasis Web	✓ PHP 8.1 + MySQLi	✓ Desktop	Tidak ada	Tidak ada
Transparansi Iterasi	✓ Lengkap	Tidak	Tidak	Tidak
Grafik Elbow dalam Sistem	✓ Otomatis	Manual	Manual	Manual

Perbandingan pada Tabel 4.10 menunjukkan bahwa penelitian ini memberikan kontribusi yang distinktif dibandingkan penelitian terdahulu, terutama dalam hal: (1) integrasi penuh algoritma dan sistem berbasis web menggunakan PHP 8.1 dan MySQLi, (2) transparansi riwayat iterasi yang ditampilkan secara real-time dalam sistem, (3) grafik Elbow Method yang dihasilkan otomatis oleh sistem tanpa perhitungan manual terpisah, dan (4) penyajian nilai normalisasi Min-Max secara langsung dalam antarmuka data untuk memudahkan verifikasi.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan seluruh tahapan penelitian yang telah dilaksanakan, mulai dari pengumpulan data, preprocessing, analisis K-Means Clustering, implementasi sistem berbasis web, hingga pengujian Black-Box Testing, maka dapat ditarik kesimpulan sebagai berikut:

- a. Algoritma K-Means Clustering berhasil diterapkan pada data penempatan karyawan PT. Guna Cipta Prima periode 2025 yang terdiri dari 28 lokasi aktif. Dengan menggunakan dua variabel numerik utama yaitu jumlah tenaga Security (X1) dan jumlah tenaga Cleaning Service (X2) yang telah dinormalisasi menggunakan metode Min-Max Scaling, algoritma berhasil mencapai konvergensi pada iterasi ke-4. Metode Elbow berbasis nilai SSE menetapkan $k = 3$ sebagai jumlah cluster optimal, yang didukung oleh fakta bahwa penurunan SSE paling signifikan terjadi dari $k = 1$ ke $k = 3$, sementara penambahan cluster di atas $k = 3$ hanya memberikan penurunan SSE yang marginal.
- b. Hasil pengelompokan menghasilkan tiga cluster dengan karakteristik yang berbeda dan bermakna secara operasional. Cluster 1 (Kebutuhan Rendah) beranggotakan 2 lokasi dengan profil CS sangat tinggi (rata-rata 11,5 orang) dan Security sangat rendah (rata-rata 1,5 orang), yaitu Clover Jl. T Umar dan Chevrolet Krakatau. Cluster 2 (Kebutuhan Sedang) merupakan kelompok terbesar yang mencakup 17 lokasi (60,7% dari total) dengan rata-rata Security 2–3 orang dan CS 3–4 orang, merepresentasikan pola operasional standar perusahaan. Cluster 3 (Kebutuhan Tinggi) beranggotakan 9 lokasi dengan karakteristik Security intensif (rata-rata 4,9 orang), mencakup lokasi-lokasi seperti SPBU Krakatau, Awal Bros Medan, dan Clover Jl. Krakatau.
- c. Kualitas pengelompokan yang dihasilkan dievaluasi menggunakan dua metrik. Nilai SSE akhir sebesar 1,1563 menunjukkan tingkat kekompakan intra-cluster yang memadai, dengan reduksi 45,4% dari nilai SSE pada $k = 2$ ($SSE = 2,1183$). Nilai Silhouette Coefficient sebesar 0,3498 berada pada rentang 0,26–0,50 yang diinterpretasikan sebagai "Struktur Cukup Baik" berdasarkan kriteria

Rousseeuw (1987), mengindikasikan bahwa pemisahan antar ketiga cluster bermakna dan dapat dipertanggungjawabkan secara statistik. Nilai SC pada $k = 3$ juga terbukti lebih tinggi dibandingkan $k = 2$ ($SC = 0,3021$) dan $k = 4$ ($SC = 0,2876$), semakin mengkonfirmasi ketepatan pemilihan $k = 3$.

- d. Sistem informasi berbasis web berhasil dirancang dan diimplementasikan menggunakan PHP 8.1 dengan ekstensi MySQLi dan basis data MySQL, yang dijalankan pada web server Apache melalui XAMPP. Sistem terdiri dari tujuh modul fungsional utama — Login, Dashboard, Data Lokasi, Data Karyawan, Proses Clustering, Hasil Clustering, dan Laporan — yang terintegrasi dalam satu antarmuka yang responsif dan interaktif. Keunggulan sistem mencakup: transparansi riwayat iterasi K-Means, grafik Elbow Method otomatis berbasis Chart.js, scatter plot 2D hasil clustering, serta halaman laporan dengan fitur ekspor PDF dan rekomendasi operasional per cluster.
- e. Pengujian fungsionalitas sistem menggunakan metode Black-Box Testing terhadap 25 skenario uji yang mencakup 8 modul menghasilkan tingkat keberhasilan 100%, di mana seluruh skenario uji menghasilkan output yang sesuai dengan spesifikasi yang dirancang. Sistem berhasil menangani kondisi input valid, input tidak valid, validasi form, proteksi sesi, dan seluruh alur proses K-Means tanpa ditemukan kegagalan fungsional.
- f. Hasil analisis K-Means Clustering terbukti memberikan nilai tambah yang nyata bagi manajemen PT. Guna Cipta Prima. Pengelompokan berbasis data yang dihasilkan sistem dapat dijadikan dasar objektif dalam penyusunan strategi distribusi karyawan, perencanaan rekrutmen, penetapan standar kontrak layanan per kategori lokasi, serta identifikasi lokasi-lokasi yang memerlukan perhatian khusus dalam hal kebutuhan tenaga kerja. Pendekatan ini secara langsung menggantikan proses penempatan berbasis intuisi yang selama ini diterapkan perusahaan dengan pendekatan yang terukur, reproduktif, dan berbasis data.

5.2 Saran

Berdasarkan pengalaman selama penelitian serta hasil evaluasi terhadap sistem yang telah dibangun, berikut disampaikan beberapa saran untuk pengembangan dan penelitian selanjutnya:

5.2.1 Saran untuk PT. Guna Cipta Prima

- a. Pemanfaatan Sistem secara Berkelanjutan. Sistem K-Means Clustering berbasis web yang telah dibangun dalam penelitian ini disarankan untuk digunakan secara rutin oleh bagian administrasi PT. Guna Cipta Prima, minimal satu kali per semester, sebagai alat bantu evaluasi distribusi karyawan. Penggunaan berkala akan memungkinkan perusahaan mendeteksi perubahan pola kebutuhan tenaga kerja lintas lokasi secara data-driven.
- b. Perluasan Variabel Analisis. Untuk meningkatkan kedalaman analisis di masa mendatang, perusahaan disarankan untuk mempertimbangkan penambahan variabel seperti lama kontrak lokasi, luas area yang dilayani, jumlah jam operasional per hari, atau tingkat turnover karyawan per lokasi. Penambahan variabel ini akan menghasilkan profil cluster yang lebih kaya dan rekomendasi penempatan yang lebih presisi.
- c. Validasi Lapangan atas Hasil Clustering. Hasil pengelompokan yang dihasilkan sistem sebaiknya divalidasi secara berkala dengan kondisi aktual di lapangan melalui mekanisme evaluasi oleh supervisor atau koordinator area. Validasi ini penting untuk memastikan bahwa label cluster (Kebutuhan Tinggi, Sedang, Rendah) tetap relevan seiring perubahan kondisi operasional lokasi dari waktu ke waktu.

5.2.2 Saran untuk Pengembangan Sistem

- a. Implementasi K-Means++ untuk Inisialisasi Centroid. Algoritma K-Means pada sistem saat ini menggunakan titik awal yang sudah ditentukan untuk keperluan reproduktibilitas. Untuk meningkatkan ketahanan hasil terhadap variasi data di masa mendatang, disarankan untuk mengimplementasikan metode inisialisasi K-Means++ (Arthur & Vassilvitskii, 2007) yang secara terbukti menghasilkan konvergensi lebih cepat dan solusi yang lebih stabil dibandingkan inisialisasi acak pada K-Means standar.
- b. Penambahan Fitur Impor Data dari Excel/CSV. Sistem disarankan untuk dilengkapi dengan fitur impor data massal dari file Microsoft Excel atau CSV, sehingga proses pembaruan data karyawan dapat dilakukan secara batch tanpa harus memasukkan data satu per satu melalui formulir. Fitur ini akan sangat

meningkatkan efisiensi operasional, terutama ketika data perlu diperbarui secara periodik pada awal periode kontrak.

- c. Penerapan Role-Based Access Control (RBAC). Untuk persiapan penggunaan di lingkungan multi-pengguna, sistem disarankan untuk mengimplementasikan sistem manajemen peran dengan minimal tiga level akses: Admin (kelola semua data dan proses), Operator (input data karyawan), dan Viewer (hanya melihat hasil dan laporan). Implementasi RBAC akan meningkatkan keamanan dan fleksibilitas penggunaan sistem di tingkat organisasi yang lebih besar.
- d. Integrasi Ekspor PDF Berbasis Server. Mekanisme ekspor laporan disarankan untuk ditingkatkan dari `window.print()` menjadi generasi PDF berbasis server menggunakan library seperti FPDF atau TCPDF, atau alternatif modern seperti mPDF. Pendekatan ini akan menghasilkan file PDF dengan format yang konsisten dan dapat dikontrol sepenuhnya dari sisi server, terlepas dari pengaturan printer pengguna.
- e. Penjadwalan dan Notifikasi Analisis Otomatis. Sistem dapat dikembangkan untuk mendukung penjadwalan analisis clustering secara otomatis pada interval yang ditentukan (misalnya setiap awal bulan atau kuartal), disertai notifikasi kepada pengguna yang berwenang melalui email atau antarmuka sistem ketika analisis baru telah selesai dijalankan. Fitur ini akan menjadikan sistem lebih proaktif dan terintegrasi dengan alur kerja manajemen perusahaan.
- f. Perbandingan dengan Algoritma Clustering Lain. Penelitian lanjutan disarankan untuk membandingkan kinerja K-Means dengan algoritma clustering alternatif seperti K-Medoids (PAM), DBSCAN, atau Gaussian Mixture Model pada dataset yang sama, guna menentukan metode yang paling optimal untuk karakteristik data penempatan karyawan jasa. Perbandingan ini akan menghasilkan kesimpulan yang lebih komprehensif mengenai pendekatan terbaik dalam domain manajemen sumber daya manusia berbasis data.

5.2.3 Saran untuk Penelitian Selanjutnya

- a. Penerapan pada Perusahaan Jasa Serupa. Penelitian selanjutnya disarankan untuk menerapkan metodologi yang sama pada perusahaan-perusahaan jasa

keamanan dan kebersihan lain di Kota Medan atau wilayah lain di Sumatera Utara, guna memperoleh perbandingan pola penempatan yang lebih luas dan menghasilkan model benchmark industri.

- b. Integrasi dengan Sistem Informasi SDM. Penelitian selanjutnya dapat mengeksplorasi integrasi sistem K-Means Clustering ini dengan sistem informasi manajemen sumber daya manusia (HRIS) yang lebih komprehensif, sehingga hasil analisis clustering dapat secara langsung memengaruhi proses rekrutmen, penilaian kinerja, dan perencanaan suksesi di perusahaan.
- c. Penerapan Pendekatan Semi-Supervised Clustering. Penelitian lebih lanjut dapat mengeksplorasi pendekatan semi-supervised clustering yang menggabungkan data historis kinerja karyawan (label parsial) dengan atribut penempatan, sehingga hasil pengelompokan tidak hanya mencerminkan pola distribusi saat ini, tetapi juga mengandung informasi prediktif tentang potensi keberhasilan penempatan di masa mendatang.

Tabel 5.1 Ringkasan Temuan Utama Penelitian

Aspek Penelitian	Temuan / Hasil	Implikasi
Dataset	28 lokasi penempatan aktif PT. Guna Cipta Prima, 2 variabel: Security (X1) dan CS (X2), periode 2025.	Data nyata perusahaan memberikan relevansi langsung terhadap keputusan operasional.
Metode Normalisasi	Min-Max Scaling: $X1' = (X1-1)/6$, $X2' = (X2-1)/12$. Rentang hasil: [0, 1].	Mencegah dominansi variabel bernilai besar dan memastikan jarak Euclidean yang adil antar variabel.
Jumlah Cluster Optimal	$k = 3$, ditentukan oleh Metode Elbow. Titik siku jelas pada $k = 3$ dengan selisih SSE terbesar.	Tiga kelompok operasional yang bermakna dan dapat diinterpretasikan secara manajerial.
Jumlah Iterasi	4 iterasi hingga konvergensi (centroid tidak berubah antara iterasi ke-3 dan ke-4).	Algoritma efisien dan stabil pada dataset berukuran 28 record.
Nilai SSE	$SSE = 1,1563$. Reduksi 45,4% dari SSE pada $k=2$ (2,1183).	Kekompakan intra-cluster yang memadai; penambahan cluster ke $k=4$ tidak proporsional.
Silhouette	$SC = 0,3498$ (rentang	Pemisahan antar cluster bermakna

Aspek Penelitian	Temuan / Hasil	Implikasi
Coefficient	0,26–0,50: "Struktur Cukup Baik" — Rousseeuw, 1987).	dan valid secara statistik. SC tertinggi di antara k=2, k=3, k=4.
Distribusi Cluster	C1 (Kebutuhan Rendah): 2 lok. C2 (Kebutuhan Sedang): 17 lok. (60,7%) C3 (Kebutuhan Tinggi): 9 lok.	Mayoritas lokasi (60,7%) berpola serupa → cocok untuk template kontrak layanan standar.
Sistem Web	PHP 8.1 + MySQLi + MySQL + Bootstrap 5. 7 modul fungsional, Chart.js untuk visualisasi.	Sistem terintegrasi siap digunakan tanpa alat analisis eksternal.
Hasil Pengujian	Black-Box Testing: 25 skenario uji, 8 modul, tingkat keberhasilan 100%.	Seluruh fungsionalitas sistem bekerja sesuai spesifikasi perancangan.

Secara keseluruhan, penelitian ini telah berhasil menjawab seluruh rumusan masalah yang ditetapkan dan mencapai semua tujuan penelitian. Penerapan algoritma K-Means Clustering pada data penempatan karyawan PT. Guna Cipta Prima terbukti mampu menghasilkan pengelompokan yang bermakna, tervalidasi secara statistik, dan bernilai praktis bagi manajemen perusahaan. Sistem berbasis web yang dikembangkan menyajikan seluruh proses analisis dalam satu platform yang terintegrasi, transparan, dan mudah digunakan, sehingga menjadikan pendekatan data mining ini dapat diterapkan secara berkelanjutan di lingkungan operasional perusahaan.

Peneliti berharap bahwa temuan dan sistem yang dihasilkan dalam penelitian ini dapat memberikan kontribusi nyata, baik bagi PT. Guna Cipta Prima dalam meningkatkan efisiensi manajemen sumber daya manusianya, maupun bagi pengembangan ilmu pengetahuan di bidang data mining dan sistem informasi, khususnya di Program Studi Sistem Informasi Universitas Muhammadiyah Sumatera Utara.

DAFTAR PUSTAKA

- Aziz, A., Rahman, A., & Hidayat, R. (2020). Implementasi metode clustering KMeans dan K-Medoids untuk pengelompokan data akademik. *Jurnal Teknologi Informasi*, 15(2), 101–110.
- Hidayat, M., & Kusuma, D. W. (2024). Pengembangan sistem pendukung keputusan penentuan siswa berprestasi menggunakan metode hybrid KMeans dan K-Medoids. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(1), 55–62.
- Lestari, D., & Putra, R. (2021). Penerapan algoritma K-Means dalam pengelompokan prestasi siswa. *Jurnal Sistem dan Teknologi Informasi*, 9(1), 45–52.
- Mochamad, W., et al. (2020). Implementasi algoritma K-Means dalam pengelompokan data mahasiswa. *Jurnal Teknologi Informasi dan Komunikasi*, 12(2), 76–84.
- Nugroho, B., & Wijaya, T. (2020). Perbandingan algoritma K-Means dan KMedoids untuk pengelompokan data akademik. *Jurnal Informatika*, 7(1), 12–20.
- Rahimi, F. (2020). *Dasar-dasar Database MySQL*. Yogyakarta: Andi Offset.
- Rahmawati, A., & Santoso, B. (2023). Perbandingan algoritma K-Means dan KMedoids untuk pengelompokan data siswa berdasarkan prestasi akademik. *Jurnal Teknologi Pendidikan*, 8(2), 210–219.
- 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.
- Siti Nurlela. (2020). Implementasi Algoritma K-Medoids Dalam Pengelompokan Data Siswa. *Jurnal Sistem Informasi dan Komputeteterisasi*.
- Wibowo, D. A., & Pratiwi, E. K. (2022). Implementasi K-Means dan K-Medoids dalam identifikasi pola prestasi belajar siswa. *Jurnal Pendidikan Informatika*, 6(2), 123–132.
- Wiwid, A. (2023). Implementasi algoritma K-Medoids dalam pengelompokan data akademik. *Jurnal Sistem Cerdas*, 9(1), 67–74.
- 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>
- Liaza, N., Syahra, Y., & Mariami, I. (2020). Implementasi data mining untuk pengelompokkan 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>
- 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>
- 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>
- 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>
- 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>
- Liaza, N., Syahra, Y., & Mariami, I. (2020). Implementasi data mining untuk pengelompokkan 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>

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>