IMPLEMENTASI ALGORITMA K-MEANS UNTUK KLASTERISASI TINGKAT KEDISIPLINAN PEGAWAI PADA SISTEM KEPEGAWAIAN STMIK EL RAHMA YOGYAKARTA

Ilhan Manzis, Minarwati*
Informatika, Sekolah Tinggi Manajemen Informatika dan Komputer El Rahma Yogyakarta
Jalan SisingamangarajaNo.76, Brotokusuman, Kec. Mergangsan, Kota Yogyakarta, D.I.Yogyakarta 55671
minarwati@stmikelrahma.ac.id

ABSTRAK

Pengelolaan data presensi pada Sistem Kepegawaian STMIK El Rahma Yogyakarta saat ini masih terbatas pada pencatatan administratif sehingga belum mampu menghasilkan informasi analitis terkait kedisiplinan pegawai. Permasalahan tersebut mendorong perlunya pemanfaatan teknik *data mining* untuk mengelompokkan tingkat kedisiplinan secara objektif. Penelitian ini bertujuan untuk mengimplementasikan algoritma *K-Means Clustering* dalam mengelompokkan tingkat kedisiplinan pegawai berdasarkan data presensi. Metode yang digunakan mengacu pada tahapan *Knowledge Discovery in Databases (KDD)*, meliputi *selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*. Data yang diolah berupa persentase pemenuhan jam kerja dan kehadiran selama tiga bulan. Hasil penelitian menunjukkan bahwa data berhasil dikelompokkan menjadi tiga klaster, yaitu kedisiplinan tinggi, sedang, dan rendah. Evaluasi menggunakan *Silhouette Score* menghasilkan nilai rata-rata sebesar 0,500844 yang menunjukkan kualitas klasterisasi dalam kategori cukup baik. Hasil klasterisasi juga telah diimplementasikan ke dalam sistem sehingga mampu menyajikan informasi kedisiplinan pegawai secara otomatis dan mendukung proses evaluasi kinerja secara lebih efektif. Pengujian sistem menggunakan metode *blackbox* menunjukkan bahwa seluruh fungsi aplikasi berjalan sesuai dengan yang diharapkan

Kata kunci : *K-Means Clustering, Data Mining, Clustering, Presensi Pegawai, Evaluasi, Sistem Informasi*

1. PENDAHULUAN

Dalam institusi pendidikan tinggi, manajemen sumber daya manusia membutuhkan sistem informasi yang tidak hanya digunakan untuk keperluan administratif, tetapi juga berperan dalam menyediakan informasi strategis yang dapat menjadi dasar pertimbangan dalam proses pengambilan keputusan. Data kepegawaian yang tersimpan dalam sistem informasi memiliki potensi signifikan untuk dianalisis lebih lanjut melalui pendekatan *data mining* guna mengidentifikasi pola-pola tersembunyi yang dapat dimanfaatkan sebagai dasar dalam evaluasi kinerja pegawai [1]. Dalam konteks tersebut, data presensi pegawai menjadi salah satu indikator krusial yang dapat merepresentasikan tingkat kedisiplinan, khususnya melalui variabel kehadiran dan pemenuhan jam kerja.

Sistem Kepegawaian STMIK El Rahma Yogyakarta telah mengimplementasikan mekanisme presensi berbasis digital yang mampu merekam aktivitas kehadiran pegawai secara terintegrasi. Meskipun demikian, pemanfaatan data presensi tersebut masih terbatas pada fungsi pencatatan administratif dan belum dioptimalkan untuk menghasilkan informasi analitis yang komprehensif terkait tingkat kedisiplinan pegawai. Kondisi ini mengakibatkan belum tersedianya landasan analitis yang objektif bagi pihak pengelola dalam mengevaluasi perilaku kedisiplinan pegawai, terutama ketika data yang tersedia mencakup periode tertentu, seperti data presensi selama tiga bulan dengan variabel persentase kehadiran dan pemenuhan jam kerja.

Keterbatasan dalam pemanfaatan data presensi tersebut mengindikasikan adanya permasalahan dalam proses ekstraksi pengetahuan dari data yang tersedia. Data yang bersifat numerik dan terstruktur belum dimanfaatkan secara optimal untuk menghasilkan informasi yang mampu merepresentasikan pola kedisiplinan pegawai secara menyeluruh. Selain itu, proses evaluasi kedisiplinan yang masih bersifat deskriptif tanpa dukungan metode analitis menyebabkan hasil penilaian menjadi kurang objektif dan sulit untuk diklasifikasikan secara sistematis. Oleh karena itu, diperlukan suatu pendekatan berbasis data yang mampu mengelompokkan pegawai berdasarkan karakteristik kedisiplinan secara terukur dan konsisten.

Berbagai studi dalam bidang *data mining* menunjukkan bahwa teknik *clustering* mampu mengidentifikasi pola tersembunyi dalam data tanpa memerlukan label sebelumnya *(unsupervised learning)*. Salah satu metode yang populer dalam proses klasterisasi adalah *K-Means Clustering*, yang dikenal karena kemampuannya mengelompokkan data berdasarkan kesamaan karakteristik dengan proses komputasi yang cukup efisien. Selain itu, metode ini dinilai efektif dalam menangani data numerik serta memiliki interpretasi hasil yang sederhana sehingga mudah diimplementasikan pada berbagai bidang, termasuk analisis sumber daya manusia. Algoritma K-Means digunakan untuk mengelompokkan karyawan berdasarkan performa kerja dan mampu membantu analisis berbasis data dalam manajemen SDM [2]. Implementasi *K-Means* pada berbagai kasus, seperti

analisis kehadiran karyawan dan pengelompokan performa akademik, menunjukkan bahwa metode ini mampu menghasilkan segmentasi data yang informatif serta mendukung pengambilan keputusan berbasis data [1][3][4][5]. Selain itu, algoritma *K-Means* juga telah diterapkan dalam pengelompokan data ekonomi seperti tingkat gaji untuk mengidentifikasi perbedaan karakteristik antar kelompok data [6].

Berdasarkan kondisi tersebut, terdapat kesenjangan penelitian yang terletak pada belum diterapkannya metode *clustering* untuk mengelompokkan tingkat kedisiplinan pegawai secara objektif dalam sistem kepegawaian terintegrasi. Sebagian besar penelitian sebelumnya lebih berfokus pada analisis kehadiran atau performa dalam konteks yang berbeda, sehingga belum secara spesifik membahas pengelompokan kedisiplinan pegawai berbasis data presensi sebagai indikator utama, khususnya dengan memanfaatkan variabel persentase kehadiran dan pemenuhan jam kerja dalam sistem kepegawaian terintegrasi. Oleh karena itu, diperlukan suatu pendekatan yang mampu mengisi kesenjangan tersebut dengan memanfaatkan data presensi sebagai sumber utama dalam proses klasterisasi.

Penelitian ini bertujuan untuk mengimplementasikan algoritma *K-Means Clustering* dalam mengelompokkan tingkat kedisiplinan pegawai berdasarkan data presensi selama tiga bulan terakhir. Variabel yang digunakan meliputi persentase pemenuhan jam kerja dan persentase kehadiran. Jumlah klaster pada penelitian ini ditentukan sebanyak tiga kategori, yaitu tingkat kedisiplinan tinggi, sedang, dan rendah, yang didasarkan pada konsep pengelompokan tingkat kedisiplinan pegawai sehingga memudahkan proses interpretasi hasil. Pemilihan algoritma *K-Means* didasarkan pada kemampuannya dalam mengolah data numerik secara efisien serta menghasilkan pengelompokan yang mudah diinterpretasikan [4][5]. Hasil klasterisasi yang diperoleh diharapkan mampu menyediakan informasi yang objektif, terstruktur, dan dapat dimanfaatkan sebagai dasar dalam proses evaluasi serta pengambilan keputusan pada sistem kepegawaian.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Sejumlah penelitian terkait penerapan *data mining* dan algoritma *K-Means* telah digunakan secara luas di berbagai bidang. Salah satu penelitian menunjukkan bahwa *data mining* dapat dimanfaatkan untuk menganalisis pola kehadiran karyawan dengan menggunakan algoritma *K-Means* melalui pengelompokan data absensi. Hasil pengelompokan tersebut menghasilkan kategori karyawan disiplin dan kurang disiplin berdasarkan indikator keterlambatan, kehadiran, dan ketidakhadiran, sehingga dapat digunakan sebagai dasar dalam evaluasi dan pengambilan keputusan [7]. Selain itu, penelitian lain menunjukkan bahwa algoritma *K-Means* dapat diterapkan dalam analisis pola kehadiran karyawan berdasarkan ketepatan waktu dan tingkat keterlambatan, sehingga menghasilkan pengelompokan karyawan ke dalam kategori disiplin dan kurang disiplin yang dapat digunakan untuk mendukung pengambilan keputusan dalam pengelolaan kepegawaian [1].

Penelitian lain menunjukkan bahwa algoritma *K-Means* mampu mengelompokkan data ke dalam beberapa klaster berdasarkan karakteristik tertentu, sehingga hasilnya dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan [8]. Selain itu, teknik *clustering* juga dapat digunakan untuk mengidentifikasi pola tertentu dalam data berdasarkan tingkat kemiripan antar objek yang dianalisis [9].

2.2. Data Mining

Data mining dapat dipahami sebagai suatu teknik analitis yang digunakan untuk menemukan pola, hubungan, atau pengetahuan tersembunyi dari kumpulan data berukuran besar. Kegiatan ini tidak berdiri sendiri, melainkan merupakan bagian dari rangkaian proses *Knowledge Discovery in Databases (KDD)*. Dalam tahapan *KDD*, terdapat beberapa langkah yang harus dilakukan, yaitu pemilihan data yang relevan, pembersihan dan persiapan data, transformasi data ke dalam bentuk yang sesuai, penerapan metode *data mining*, serta evaluasi hasil untuk memastikan informasi yang diperoleh dapat dimanfaatkan sebagai dasar pengambilan keputusan [10]. Dalam konteks pengelolaan sumber daya manusia, *data mining* dapat dimanfaatkan untuk mengidentifikasi pola perilaku pegawai, termasuk dalam menganalisis tingkat kedisiplinan berdasarkan data presensi melalui pengelompokan data kehadiran dan keterlambatan [7].

2.3. K-Means Clustering

Algoritma *K-Means* merupakan metode *clustering* yang digunakan untuk membagi data ke dalam beberapa kelompok dengan mempertimbangkan kedekatan nilai terhadap titik pusat klaster *(centroid)*. Pengelompokan dilakukan berdasarkan tingkat kemiripan karakteristik antar data sehingga setiap objek berada pada klaster dengan jarak terdekat terhadap *centroid* [11]. Dalam penerapannya, algoritma ini diawali dengan penentuan sejumlah titik pusat *(centroid)* sebagai representasi awal setiap klaster. Selanjutnya, jarak masing-masing data terhadap *centroid* dihitung untuk menentukan kedekatan terdekat, sehingga data ditempatkan pada kelompok yang sesuai. Proses perhitungan dan pembaruan *centroid* dilakukan secara berulang sampai tidak terjadi perubahan signifikan pada hasil pengelompokan dan diperoleh struktur klaster yang stabil. Algoritma *K-Means* banyak digunakan karena memiliki metode yang sederhana dan mudah diimplementasikan dalam berbagai kasus pengolahan data [8].

Pada algoritma *K-Means*, proses pembentukan klaster dilakukan dengan menghitung tingkat

kedekatan setiap data terhadap titik pusat *(centroid)*. Perhitungan jarak tersebut umumnya menggunakan metode *Euclidean Distance* yang dapat dinyatakan dalam bentuk persamaan berikut:

$$d = \sqrt{\sum_{i=1}^{n} (x_i - c_i)^2} \quad (1)$$

Di mana x_i merupakan nilai data dan $c_i$merupakan nilai *centroid*. Setiap objek data ditempatkan pada klaster yang memiliki nilai jarak paling kecil terhadap *centroid*. Setelah proses pengelompokan tersebut, posisi *centroid* dihitung kembali berdasarkan nilai rata-rata seluruh anggota dalam masing-masing klaster. Tahapan ini dilakukan berulang hingga tidak terjadi lagi perubahan posisi *centroid* atau sistem mencapai kondisi stabil.

2.4. Knowledge Discovery in Databases (KDD)

Knowledge Discovery in Databases (KDD) adalah suatu kerangka kerja yang digunakan untuk memperoleh pengetahuan dari kumpulan data melalui tahapan yang terstruktur dan berurutan. Proses ini diawali dengan pemilihan data yang relevan, dilanjutkan dengan tahap pembersihan dan persiapan data, kemudian transformasi data ke dalam bentuk yang sesuai untuk dianalisis. Setelah itu dilakukan penerapan teknik *data mining*, dan pada tahap akhir hasil yang diperoleh dievaluasi serta diinterpretasikan agar menghasilkan informasi yang bermakna [10]. Setiap tahapan dalam *KDD* memiliki peran penting dalam meningkatkan kualitas data, khususnya melalui proses pembersihan dan transformasi sehingga data menjadi lebih konsisten dan siap untuk dianalisis. Penerapan metode *KDD* dalam penelitian *data mining* membantu menghasilkan informasi yang berguna dari data yang telah diproses melalui tahapan-tahapan tersebut [12].

2.5. Silhouette Score

Silhouette Score adalah salah satu metrik evaluasi internal pada proses klasterisasi yang digunakan untuk menilai seberapa baik suatu data tergabung dalam klasternya sendiri dibandingkan dengan klaster lainnya. Pengukuran ini mempertimbangkan tingkat kekompakan dalam satu klaster *(cohesion)* serta jarak pemisah antar klaster *(separation)* [9]. *Silhouette Score* memiliki nilai dalam interval −1 sampai 1. Semakin mendekati angka 1, semakin baik kualitas pemisahan dan kekompakan klaster yang terbentuk. Perhitungan *Silhouette Score* dinyatakan dengan rumus:

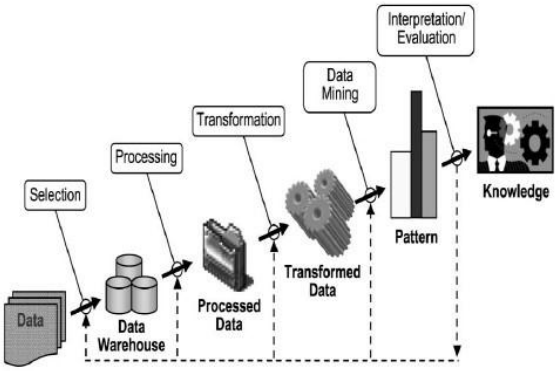
$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

di mana a(i) merupakan rata-rata jarak antara data ke-i dengan anggota dalam klaster yang sama *(intra-cluster distance)*, sedangkan b(i) merupakan rata-rata jarak antara data ke-i dengan klaster terdekat *(inter-cluster distance)* [13]. Penggunaan metode ini digunakan untuk mengevaluasi kualitas hasil klasterisasi berdasarkan kedekatan dan pemisahan antar klaster.

3. METODE PENELITIAN

3.1. Tahap Penelitian

Proses pengolahan data menggunakan *Knowledge Discovery in Databases* (KDD) yang terdiri dari tahapan *selection*, *preprocessing*, *transformation*, *data mining*, serta *evaluation* atau *interpretation*. Tahapan tersebut ditunjukkan pada Gambar 1, dimulai dari *selection* untuk memilih data yang relevan, dilanjutkan *preprocessing* (processing) untuk pembersihan dan validasi data, kemudian *transformation* melalui normalisasi, selanjutnya *data mining* menggunakan algoritma K-Means untuk menghasilkan klaster, dan diakhiri dengan *evaluation* atau *interpretation* untuk menilai serta memaknai hasil klasterisasi.



Gambar 1. Tahapan Metode KDD[14]

3.2. Sumber Data

Data yang digunakan berupa data presensi pegawai yang diperoleh dari Sistem Kepegawaian STMIK El Rahma Yogyakarta selama tiga bulan terakhir. Data tersebut telah melalui proses agregasi dan anonimisasi sehingga tidak memuat identitas pribadi pegawai, serta disusun dalam bentuk numerik berupa persentase kehadiran dan pemenuhan jam kerja yang digunakan sebagai dasar dalam proses analisis.

3.3. Selection

Proses *selection* dilakukan dengan memilih data yang relevan sesuai dengan tujuan analisis, yaitu data presensi pegawai yang telah diolah menjadi dua variabel utama berupa persentase pemenuhan jam kerja dan persentase kehadiran, sehingga data yang digunakan lebih terfokus untuk proses klasterisasi.

3.4. Preprocessing (Processing)

Proses *preprocessing* dilakukan untuk memastikan kualitas data melalui pembersihan dan validasi, dengan melakukan pembatasan nilai persentase pemenuhan jam kerja hingga maksimum 120% dan persentase kehadiran pada rentang 0 hingga

100%, sehingga data terhindar dari nilai yang tidak wajar.

3.5. Transformation

Data diubah ke dalam bentuk yang sesuai melalui normalisasi menggunakan metode *Min-Max* untuk menyamakan skala antar variabel, sehingga tidak terjadi dominasi salah satu variabel dalam proses klasterisasi.

3.6. Data Mining

Algoritma *K-Means Clustering* diterapkan untuk mengelompokkan data berdasarkan tingkat kemiripan, dengan jumlah klaster sebanyak tiga yang merepresentasikan tingkat kedisiplinan tinggi, sedang, dan rendah melalui perhitungan jarak dan pembaruan centroid secara iteratif hingga mencapai kondisi stabil.

3.7. Evaluation

Pengukuran kualitas hasil klasterisasi dilakukan menggunakan metode Silhouette Score untuk menilai tingkat kedekatan data dalam satu klaster serta pemisahan antar klaster.

3.8. Interpretation

Hasil klasterisasi diinterpretasikan dengan mengelompokkan data ke dalam kategori tingkat kedisiplinan tinggi, sedang, dan rendah, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan.

4. HASIL DAN PEMBAHASAN

4.1. Sumber Data

Data yang digunakan berupa presensi pegawai dari Sistem Kepegawaian STMIK El Rahma Yogyakarta selama Desember 2025 hingga Februari 2026. Data mencakup atribut kehadiran, ketidakhadiran, dan pemenuhan jam kerja yang diolah menjadi persentase. Untuk menjaga kerahasiaan, data telah dianonimkan menggunakan kode Pegawai A–Z. Data presensi ditampilkan pada Tabel 1.

Tabel 1. Data presensi pegawai

Nama	Hadir	Tidak	Memenuhi
Pegawai A	34	22	7
Pegawai B	53	3	53
Pegawai C	53	3	44
Pegawai D	52	4	49
Pegawai E	35	21	21
Pegawai F	48	8	38
Pegawai G	55	1	42
Pegawai H	34	22	28
Pegawai I	44	12	6
Pegawai J	43	13	1
Pegawai K	49	7	0
Pegawai L	56	0	56
Pegawai M	26	30	22
Pegawai N	54	2	31
Pegawai O	32	24	14
Pegawai P	54	2	7
Pegawai Q	49	7	49
Pegawai R	36	20	21
Pegawai S	52	4	28
Pegawai T	26	30	6
Pegawai U	28	28	28
Pegawai V	50	6	50
Pegawai W	16	40	0
Pegawai X	55	1	55
Pegawai Y	56	0	50
Pegawai Z	2	16	2

4.2. Selection

Selection dilakukan dengan memilih atribut yang digunakan dalam proses analisis, yaitu persentase pemenuhan jam kerja sebagai x_1 dan persentase kehadiran sebagai x_2, sehingga data yang digunakan lebih terfokus untuk proses klasterisasi.

4.3. Preprocessing (Processing)

Data kemudian diolah menjadi bentuk persentase untuk masing-masing variabel, yaitu $x_1$sebagai persentase pemenuhan jam kerja dan $x_2$sebagai persentase kehadiran. Hasil pengolahan data ditampilkan pada Tabel 2.

Tabel 2. Data hasil preprocessing

Nama	X1	X2
Pegawai A	20,59	60,71
Pegawai B	100	94,64
Pegawai C	83,02	94,64
Pegawai D	94,23	92,86
Pegawai E	60	62,5
Pegawai F	79,17	85,71
Pegawai G	76,36	98,21
Pegawai H	82,35	60,71
Pegawai I	13,64	78,57
Pegawai J	2,33	76,79
Pegawai K	0	87,5
Pegawai L	100	100
Pegawai M	84,62	46,43
Pegawai N	57,41	96,43
Pegawai O	43,75	57,14
Pegawai P	12,96	96,43
Pegawai Q	100	87,5
Pegawai R	58,33	64,29
Pegawai S	53,85	92,86
Pegawai T	23,08	46,43
Pegawai U	100	50
Pegawai V	100	89,29
Pegawai W	0	28,57
Pegawai X	100	98,21
Pegawai Y	89,29	100
Pegawai Z	100	11,11

4.4. Transformation

Data selanjutnya dinormalisasi menggunakan metode *Min-Max* untuk menyamakan skala antar variabel. Normalisasi dilakukan dengan menggunakan rumus sebagai berikut:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

Sebagai contoh, dilakukan normalisasi pada data Pegawai A untuk variabel $x_1$dan x_2. Misalkan:

- $x_1 = 20{,}59$, $x_{1min} = 0$, $x_{1max} = 100$
- $x_2 = 60{,}71$, $x_{2min} = 11{,}11$, $x_{2max} = 100$

Maka diperoleh:

$$x_1' = \frac{20{,}59 - 0}{100 - 0} = 0{,}2059$$

$$x_2' = \frac{60{,}71 - 11{,}11}{100 - 11{,}11} = 0{,}557993025$$

Hasil normalisasi ditampilkan pada Tabel 3.

Tabel 3. Data hasil normalisasi

Nama	X1	X2
Pegawai A	0,2059	0,557993025
Pegawai B	1	0,939700754
Pegawai C	0,8302	0,939700754
Pegawai D	0,9423	0,919676004
Pegawai E	0,6	0,578130273
Pegawai F	0,7917	0,83923951
Pegawai G	0,7636	0,979862752
Pegawai H	0,8235	0,557993025
Pegawai I	0,1364	0,758915514
Pegawai J	0,0233	0,738890764
Pegawai K	0	0,859376758
Pegawai L	1	1
Pegawai M	0,8462	0,397345033
Pegawai N	0,5741	0,959838002
Pegawai O	0,4375	0,517831027
Pegawai P	0,1296	0,959838002
Pegawai Q	1	0,859376758
Pegawai R	0,5833	0,598267522
Pegawai S	0,5385	0,919676004
Pegawai T	0,2308	0,397345033
Pegawai U	1	0,437507031
Pegawai V	1	0,879514006
Pegawai W	0	0,196422545
Pegawai X	1	0,979862752
Pegawai Y	0,8929	1
Pegawai Z	1	0

4.5. Data Mining

Data mining dilakukan menggunakan algoritma *K-Means Clustering* untuk mengelompokkan data berdasarkan tingkat kedisiplinan. Jumlah klaster ditentukan sebanyak tiga ($k = 3$), yaitu C1 sebagai kedisiplinan tinggi, C2 sebagai kedisiplinan sedang, dan C3 sebagai kedisiplinan rendah. Penentuan *centroid* awal menggunakan metode *Min–Mean–Max Initialization*, di mana nilai *centroid* ditentukan berdasarkan nilai minimum, rata-rata, dan maksimum dari masing-masing variabel. Perhitungan *centroid* dilakukan dengan rumus sebagai berikut:

$$C_{min} = \min(x), C_{mean} = \frac{\sum x}{n}, C_{max} = \max(x) \quad (4)$$

Berdasarkan data hasil normalisasi pada Tabel 3, diperoleh:

- $x_1^{min} = 0$, $x_1^{max} = 1$
- $x_2^{min} = 0$, $x_2^{max} = 1$

Selanjutnya dihitung nilai rata-rata dari masing-masing variabel, sehingga diperoleh:

- $x_1^{mean} \approx 0{,}628838462$
- $x_2^{mean} \approx 0{,}722011648$

Berdasarkan hasil tersebut, diperoleh *centroid* awal yang ditampilkan pada Tabel 4.

Tabel 4. Centroid awal

Cluster	X1	X2
C1	1	1
C2	0,628838	0,722012
C3	0	0

Selanjutnya dilakukan perhitungan jarak antara setiap data dengan *centroid* menggunakan metode *Euclidean Distance* untuk mengetahui kedekatan data terhadap masing-masing klaster. Adapun rumus yang digunakan adalah sebagai berikut:

$$d = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2} \quad (5)$$

Sebagai contoh, dilakukan perhitungan jarak untuk Pegawai A dengan nilai $(0{,}2059; 0{,}558)$ terhadap *centroid* awal, yaitu C1 = (1 ; 1), C2 = (0,6288 ; 0,7220), dan C3 = (0 ; 0).

$$d_{C1} = \sqrt{(0{,}2059 - 1)^2 + (0{,}558 - 1)^2} = 0{,}908826$$

$$d_{C2} = \sqrt{(0{,}2059 - 0{,}6288)^2 + (0{,}558 - 0{,}7220)^2} = 0{,}453628759$$

$$d_{C3} = \sqrt{(0{,}2059 - 0)^2 + (0{,}558 - 0)^2} = 0{,}59477$$

Berdasarkan hasil tersebut, Pegawai A masuk ke klaster C2 karena memiliki jarak terkecil. Hasil perhitungan jarak seluruh data terhadap *centroid* awal ditampilkan pada Tabel 5.

Tabel 5. Hasil perhitungan jarak

Nama	C1	C2	C3	Cluster
Pegawai A	0,908826	0,453629	0,59477	C2
Pegawai B	0,060299	0,43029	1,372238	C1
Pegawai C	0,180189	0,296538	1,253902	C1
Pegawai D	0,0989	0,37058	1,316713	C1
Pegawai E	0,581355	0,146743	0,833207	C2
Pegawai F	0,263121	0,200665	1,153738	C2
Pegawai G	0,237256	0,290943	1,242262	C1
Pegawai H	0,475944	0,254549	0,99474	C2
Pegawai I	0,89662	0,493819	0,771076	C2
Pegawai J	1,011	0,605774	0,739258	C2
Pegawai K	1,009839	0,643667	0,859377	C2
Pegawai L	0	0,463722	1,414214	C1
Pegawai M	0,621971	0,39071	0,934846	C2
Pegawai N	0,427789	0,244044	1,118427	C2
Pegawai O	0,740873	0,279822	0,677905	C2
Pegawai P	0,871326	0,552992	0,968548	C2
Pegawai Q	0,140623	0,395765	1,318533	C1
Pegawai R	0,578816	0,131857	0,835561	C2
Pegawai S	0,468438	0,21733	1,065733	C2

Nama	C1	C2	C3	Cluster
Pegawai T	0,97717	0,513657	0,459512	C3
Pegawai U	0,562493	0,467658	1,091518	C2
Pegawai V	0,120486	0,403197	1,331745	C1
Pegawai W	1,282863	0,819562	0,196423	C3
Pegawai X	0,020137	0,451938	1,400047	C1
Pegawai Y	0,1071	0,383414	1,340623	C1
Pegawai Z	1	0,811826	1	C2

Selanjutnya dilakukan pembaruan *centroid* menggunakan rata-rata setiap variabel pada masing-masing klaster dengan rumus:

$$C_k = \frac{\sum x}{n} \quad (6)$$

Hasil perhitungan *centroid* baru ditampilkan pada Tabel 6.

Tabel 6. Centroid baru

Cluster	X1	X2
C1	0,936556	0,944188
C2	0,512667	0,645389
C3	0,1154	0,296884

Proses perhitungan jarak dan pembaruan *centroid* dilakukan secara berulang hingga tidak terjadi perubahan klaster. Pada penelitian ini, proses konvergensi tercapai pada iterasi ke-5, di mana tidak terdapat perubahan anggota klaster pada iterasi terakhir. Berdasarkan hasil akhir proses klasterisasi, diperoleh pembagian data ke dalam tiga klaster. Hasil pengelompokan akhir ditampilkan pada Tabel 7.

Tabel 7. Hasil akhir klasterisasi

Nama	Cluster
Pegawai A	C3
Pegawai B	C1
Pegawai C	C1
Pegawai D	C1
Pegawai E	C2
Pegawai F	C1
Pegawai G	C1
Pegawai H	C2
Pegawai I	C3
Pegawai J	C3
Pegawai K	C3
Pegawai L	C1
Pegawai M	C2
Pegawai N	C1
Pegawai O	C2
Pegawai P	C3
Pegawai Q	C1
Pegawai R	C2
Pegawai S	C1
Pegawai T	C3
Pegawai U	C2
Pegawai V	C1
Pegawai W	C3
Pegawai X	C1
Pegawai Y	C1
Pegawai Z	C2

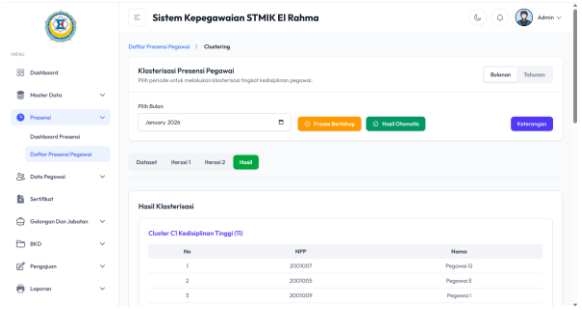
4.6. Evaluation

Evaluasi hasil klasterisasi dilakukan menggunakan metode *Silhouette Score* untuk mengukur kualitas pengelompokan data berdasarkan tingkat kedekatan dalam klaster dan pemisahan antar klaster. Perhitungan dilakukan menggunakan *Microsoft Excel* dengan menghitung nilai a(i), b(i), dan s(i) pada setiap data. Hasil perhitungan *Silhouette Score* ditampilkan pada Tabel 8.

Tabel 8. Hasil perhitungan Silhouette Score

Nama	Cluster	a(i)	b(i)	s(i)
Pegawai A	C3	0,30372	0,580098	0,476434
Pegawai B	C1	0,177647	0,60103	0,70443
Pegawai C	C1	0,166179	0,547963	0,696734
Pegawai D	C1	0,15807	0,561358	0,718415
Pegawai E	C2	0,309393	0,465863	0,335871
Pegawai F	C1	0,208479	0,45233	0,539099
Pegawai G	C1	0,199978	0,576485	0,653109
Pegawai H	C2	0,302872	0,41019	0,26163
Pegawai I	C3	0,275065	0,705976	0,610377
Pegawai J	C3	0,280401	0,80054	0,649735
Pegawai K	C3	0,333421	0,866033	0,615002
Pegawai L	C1	0,193122	0,651636	0,703635
Pegawai M	C2	0,301616	0,560339	0,461725
Pegawai N	C1	0,326649	0,575535	0,432444
Pegawai O	C2	0,414464	0,42373	0,021868
Pegawai P	C3	0,394182	0,733707	0,462753
Pegawai Q	C1	0,198841	0,536609	0,62945
Pegawai R	C2	0,323806	0,459523	0,295342
Pegawai S	C1	0,357852	0,549858	0,349193
Pegawai T	C3	0,388315	0,560226	0,306859
Pegawai U	C2	0,374835	0,537067	0,302071
Pegawai V	C1	0,188949	0,552388	0,65794
Pegawai W	C3	0,546858	0,825598	0,337622
Pegawai X	C1	0,184316	0,634552	0,709534
Pegawai Y	C1	0,169434	0,616315	0,725086
Pegawai Z	C2	0,607581	0,957693	0,365579
rata-rata Silhouette Score				**0,500844**

Berdasarkan hasil perhitungan, diperoleh nilai rata-rata *Silhouette Score* sebesar 0,500844, yang menunjukkan bahwa kualitas klasterisasi berada pada kategori cukup baik dan mampu memisahkan data dengan cukup jelas. Berdasarkan hasil tersebut, proses pengelompokan data kemudian diterapkan ke dalam Sistem Kepegawaian STMIK El Rahma Yogyakarta. Penerapan ini bertujuan untuk mempermudah proses klasterisasi serta penyajian hasil pengelompokan tingkat kedisiplinan pegawai secara otomatis melalui sistem. Tampilan hasil klasterisasi pada sistem dapat dilihat pada Gambar 2.



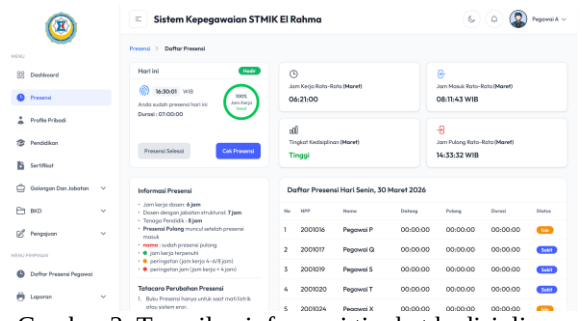
Gambar 2. Tampilan hasil klasterisasi pada sistem

Hasil yang ditampilkan pada sistem menunjukkan pembagian pegawai ke dalam tiga klaster, yaitu C1 (kedisiplinan tinggi), C2 (kedisiplinan sedang), dan C3 (kedisiplinan rendah), yang sesuai dengan hasil perhitungan menggunakan algoritma *K-Means*.

4.7. Interpretation

Berdasarkan hasil klasterisasi, data pegawai terbagi menjadi tiga kelompok, yaitu C1 (kedisiplinan tinggi), C2 (kedisiplinan sedang), dan C3 (kedisiplinan rendah). Klaster C1 didominasi oleh pegawai dengan tingkat kehadiran dan pemenuhan jam kerja yang tinggi, sedangkan klaster C2 menunjukkan kondisi yang cukup stabil namun belum optimal. Sementara itu, klaster C3 terdiri dari pegawai dengan tingkat kehadiran dan pemenuhan jam kerja yang relatif rendah.

Hasil ini menunjukkan bahwa metode *K-Means* mampu mengelompokkan tingkat kedisiplinan pegawai secara jelas dan terstruktur, sehingga dapat digunakan sebagai dasar dalam evaluasi serta pengambilan keputusan dalam pengelolaan sumber daya manusia. Selain itu, hasil klasterisasi juga ditampilkan pada sistem dalam bentuk informasi tingkat kedisiplinan pegawai. Tampilan tersebut dapat dilihat pada Gambar 3, di mana sistem menampilkan kategori kedisiplinan pegawai berdasarkan hasil pengelompokan yang telah dilakukan.



Gambar 3. Tampilan informasi tingkat kedisiplinan pegawai pada sistem

Informasi tersebut memudahkan pegawai maupun pihak manajemen dalam mengetahui tingkat kedisiplinan secara langsung, sehingga dapat mendukung proses monitoring dan evaluasi secara lebih efektif.

4.8. Pengujian Sistem

Pengujian sistem dilakukan menggunakan *blackbox testing* untuk memastikan fungsi aplikasi berjalan sesuai dengan yang diharapkan. Pengujian difokuskan pada proses klasterisasi dan tampilan hasil pada sistem, seperti ditunjukkan pada Tabel 9.

Tabel 9. Hasil pengujian sistem

Skenario Pengujian	Hasil yang Diharapkan	Hasil	Status
Memilih periode bulan	Dataset presensi sesuai periode dimuat	Sesuai	Valid
Memilih periode tahun	Dataset presensi sesuai periode dimuat	Sesuai	Valid
Tombol "Proses Bertahap"	Proses *K-Means* ditampilkan hingga selesai	Sesuai	Valid
Tombol "Hasil Otomatis"	Proses *K-Means* hingga *konvergen* dan hasil klaster (C1, C2, C3) ditampilkan	Sesuai	Valid
Dashboard pegawai	Informasi kedisiplinan tampil sesuai data	Sesuai	Valid

Hasil pengujian menunjukkan seluruh fungsi sistem berjalan sesuai harapan, sehingga aplikasi dapat digunakan untuk mendukung proses klasterisasi dan evaluasi kedisiplinan pegawai.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil yang diperoleh, algoritma *K-Means Clustering* berhasil diimplementasikan untuk mengelompokkan tingkat kedisiplinan pegawai pada Sistem Kepegawaian STMIK El Rahma Yogyakarta. Proses klasterisasi menghasilkan tiga kelompok, yaitu kedisiplinan tinggi (C1), sedang (C2), dan rendah (C3), berdasarkan variabel persentase kehadiran dan pemenuhan jam kerja. Hasil evaluasi menggunakan metode *Silhouette Score* menunjukkan nilai sebesar 0,500844 yang mengindikasikan bahwa kualitas klasterisasi berada pada kategori cukup baik. Implementasi pada sistem yang dikembangkan mampu menyajikan hasil klasterisasi secara otomatis dan konsisten dengan perhitungan manual, sehingga dapat mendukung proses evaluasi kedisiplinan pegawai.

Penelitian selanjutnya disarankan untuk menggunakan jumlah data yang lebih besar dan periode waktu yang lebih panjang agar hasil klasterisasi lebih akurat dan representatif. Selain itu, dapat dikembangkan dengan menambahkan variabel lain yang berkaitan dengan kinerja pegawai agar hasil pengelompokan lebih komprehensif. Penggunaan metode *clustering* lain atau kombinasi metode juga dapat dipertimbangkan untuk membandingkan hasil dan meningkatkan kualitas pengelompokan data. Dari sisi sistem, pengembangan fitur visualisasi seperti grafik atau dashboard interaktif dapat ditambahkan untuk mempermudah analisis dan pengambilan keputusan oleh pihak manajemen.

DAFTAR PUSTAKA

[1] M. Saputra *et al.*, "Implementasi K-Means Dalam Aplikasi Absensi Wajah Berbasis Laravel Dan Flutter Di PT Selaras Lawang Sewu," *Jurnal Riset Informatika dan Inovasi*, vol. 3, pp. 83–89, Feb. 2025, [Online]. Available: https://www.jurnalmahasiswa.com/index.php/jrin/article/view/2464

[2] Anandita Nabilla Ramadhani and Ghita Athalina, "Optimalisasi Kinerja Karyawan Berbasis HR Analytics dengan K-Means Clustering dan Analisis Faktor Demografi," *Jurnal Saintekom : Sains, Teknologi, Komputer dan Manajemen*, vol. 15, pp. 1–14, Mar. 2025, Accessed: Mar. 27, 2026. [Online]. Available: https://doi.org/10.33020/saintekom.v15i1.779

[3] Muhammad Raikhan Al Fatah, Aang Alim Murtopo, and Erni Unggul Sedya Utami, "Implementasi Algoritma K-Means untuk Pengelompokan Risiko Drop out Mahasiswa," *Journal of Artificial Intelligence and Digital Business (RIGGS)*, vol. 4, pp. 2719–2726, 2025, Accessed: Feb. 25, 2026. [Online]. Available: https://doi.org/10.31004/riggs.v4i3.2374

[4] Fathoni, Mutia Sahira, Adella Salsabila, Shofi Salsabila, Aulia Najibah Putri, and Ali Ibrahim, "Implementasi Algoritma K-Means Dalam Analisis Distribusi Pangkalan LPG 3kg Di Kota Palembang," *Jurnal Buffer Informatika*, vol. 11, Oct. 2025, Accessed: Feb. 25, 2026. [Online]. Available: https://doi.org/10.25134/buffer.v11i2.385

[5] Haviluddin, Suryani Junita Patandianan, Gubtha Mahendra Putra, Novianti Puspitasari, and Herman Santoso Pakpahan, "Implementasi Metode K-Means untuk Pengelompokkan Rekomendasi Tugas Akhir," *Jurnal Ilmiah Ilmu Komputer*, vol. 16, pp. 13–18, Feb. 2021, Accessed: Feb. 25, 2026. [Online]. Available: http://dx.doi.org/10.30872/jim.v16i1.5182

[6] Endah Widiarahman, Bambang Irawan, and Agus Bahtiar, "Implementasi Algoritma K-Means untuk Mengelompokan Tingkat Gaji di Indonesia Berdasarkan Provinsi," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, pp. 2824–2829, Jun. 2024, Accessed: Feb. 25, 2026. [Online]. Available: https://doi.org/10.36040/jati.v8i3.9565

[7] Yulya Muharmi and Sri Nadriati, "Analysis Of Employee Discipline Based On Digital Attendance With The K-Means Algorithm Method," *JURNAL TEKNOLOGI DAN OPEN SOURCE* , vol. 5, pp. 115–135, Dec. 2022, Accessed: Mar. 27, 2026. [Online]. Available: https://doi.org/10.36378/jtos.v5i2.2628

[8] Milla Rochmawati, Ganes Wisnu Cahya Bagaskara, Ismail Adhiya Adha, Yuyun Umaidah, and Apriade Voutama, "Implementasi Algoritma K-Means dalam Klasterisasi Penjualan pada Sebuah Perusahaan menggunakan Metodologi KDD," *SISTEMASI: Jurnal Sistem Informasi*, vol. 13, pp. 54–62, Jan. 2024, Accessed: Mar. 27, 2026. [Online]. Available: https://doi.org/10.32520/stmsi.v13i1.3074

[9] Septiantoro Sogema Fourteen, Al Hakim Putra Bintang Anugrah, Muhamad Niko Aninda Putra, Jihan Aulia Shabrina, Mulyo Adji Winoto, and ZurnanAlfian, "Perbandingan Hasil Klasterisasi K-Means Berdasarkan Silhouette Score dan Inertia Pada 5 Dataset Berbeda," *Jurnal Pendidikan Tambusai*, vol. 9, pp. 23356–23368, Jul. 2025, Accessed: Mar. 27, 2026. [Online]. Available: https://jptam.org/index.php/jptam/article/view/30330

[10] BINUS UNIVERSITY School of Information Systems, "Proses Data Mining KDD," BINUS UNIVERSITY School of Information Systems. Accessed: Mar. 27, 2026. [Online]. Available: https://sis.binus.ac.id/2021/09/30/proses-data-mining-kdd

[11] Aslam Fatkhudin, Ahmad Khambali, Fenilinas A. Artanto, Mundriyah, and Nabil A. Putra Zade, "Implementasi Algoritma Clustering K-Means Dalam Pengelompokan Mahasiswa," *Jurnal Minfo Polgan*, vol. 12, pp. 777–783, Jun. 2023, Accessed: Mar. 27, 2026. [Online]. Available: https://doi.org/10.33395/jmp.v12i1.12494

[12] Fauzan Alghifari and Didi Juardi, "Penerapan Data Mining pada Penjualan Makanan dan Minuman Menggunakan Metode Algoritma Naïve Bayes," *Jurnal Ilmiah Informatika*, vol. 9, pp. 75–81, Sep. 2021, Accessed: Mar. 27, 2026. [Online]. Available: https://doi.org/10.33884/jif.v9i02.3755

[13] Heti Mulyani, Ricak Agus Setiawan, and Halimil Fathi, "Optimization of K Value in Clustering Using Silhouette Score (Case Study: Mall Customers Data)," *Journal of Information Technology and its Utilization*, vol. 6, pp. 45–50, Dec. 2023, Accessed: Mar. 27, 2026. [Online]. Available: https://doi.org/10.56873/jitu.6.2.5243

[14] Binus University School of Accounting, "Memahami apa itu data mining," BINUS University. Accessed: Feb. 24, 2026. [Online]. Available: https://accounting.binus.ac.id/2019/10/03/memahami-apa-itu-data-mining