

Penerapan Algoritma *K-Means Clustering* untuk Klasterisasi Pola Kehadiran Pegawai

Anom Danang Sumantri^{1*}, Satriedi Wahyu Binabar¹, Nurul Amalia¹

¹Program Studi S1 Teknik Informatika, Institut Widya Pratama Pekalongan, Indonesia

*Email: anomdees414@gmail.com

Info Artikel

Kata Kunci :

pola kehadiran pegawai, *data mining*, *k-means clustering*, visualisasi data

Keywords :

employee attendance pattern, data mining, k-means clustering, data visualization

Tanggal Artikel

Dikirim : 6 November 2025

Direvisi : 27 Desember 2025

Diterima : 30 Desember 2025

Abstrak

Analisis pola kehadiran pegawai merupakan langkah penting dalam memahami karakteristik perilaku waktu kehadiran di lingkungan kerja. Penelitian ini bertujuan untuk menganalisis pola kehadiran pegawai berdasarkan data absensi pegawai menggunakan algoritma *K-Means clustering*. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan data, klasterisasi, evaluasi dan visualisasi. Data absensi selama tiga bulan diolah melalui tahap pembersihan, diskretisasi, pembentukan, pembuatan atribut baru dan normalisasi untuk memastikan kualitas data. Proses klasterisasi dilakukan dengan jumlah 3 klaster berdasarkan hasil metode *elbow*. Evaluasi model menggunakan metode Davies–Bouldin Index (DBI) menghasilkan nilai sebesar 0,743. Hasil klasterisasi kemudian divisualisasikan menggunakan *Power BI* untuk mempermudah interpretasi pola kehadiran. Berdasarkan hasil tersebut, diperoleh 3 pola waktu kehadiran utama yaitu waktu kehadiran baik, cukup, dan kurang yang menggambarkan perbedaan karakteristik waktu kehadiran pegawai secara lebih jelas. Hasil penelitian ini memberikan pemahaman menyeluruh mengenai pola waktu kehadiran pegawai yang dapat dijadikan dasar untuk analisis lebih lanjut, seperti evaluasi kinerja dan kedisiplinan.

Abstract

Analysis of employee attendance patterns is an important step in understanding the characteristics of attendance behavior in the work environment. This study aims to analyze employee attendance patterns based on attendance data using the K-Means clustering algorithm. The research stages include data collection, data pre-processing, clustering, evaluation and visualization. Attendance data for 3 months is processed through the stages of cleaning, discretion, creation of new attributes and normalization to ensure data quality. The clustering process was carried out with a total of 3 clusters based on the results of the elbow method. The model evaluation using the Davies–Bouldin Index (DBI) method yielded a value of 0.743. The clustering results are then visualized using Power BI to make it easier to interpret attendance patterns. Based on these results, 3 main attendance time patterns were obtained, namely good, sufficient, and poor attendance time which depicts the differences in the characteristics of employee attendance time more clearly. The results of this study provide a comprehensive understanding of the pattern of employee attendance time which can be used as a basis for further analysis, such as performance and discipline evaluation.

1. PENDAHULUAN

Sumber daya manusia merupakan aset terbesar dan penting dalam sebuah organisasi. Dalam suatu instansi, pegawai merupakan aset penting yang memengaruhi tercapainya keberhasilan unit kerja [1]. Keberhasilan sebuah organisasi bukan hanya dilihat dari capaian yang telah diperoleh, tetapi juga dari tingkat kedisiplinan sumber daya manusia dalam mematuhi aturan-aturan yang telah ditetapkan. Dalam proses pembelajaran maupun pekerjaan kehadiran memiliki peran yang sangat penting [2]. Kehadiran pegawai merupakan salah satu indikator penting dalam menilai kedisiplinan dan kinerja individu dalam suatu instansi atau organisasi [3].

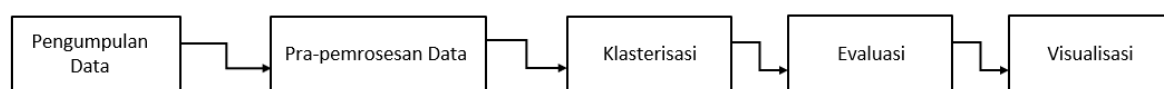
Dalam era digital seperti sekarang, sistem absensi pegawai telah beralih dari absensi manual ke sistem digital [4]. Diantaranya melalui perangkat elektronik seperti fingerprint dan pengenalan wajah, maupun menggunakan aplikasi berbasis *mobile*. Sistem digital tersebut dapat membantu meningkatkan efisiensi dalam pengumpulan dan manajemen data kehadiran pegawai. Data kehadiran pegawai seringkali hanya dijadikan arsip dan belum dimanfaatkan secara optimal untuk menghasilkan informasi yang lebih mendalam tentang perilaku kehadiran pegawai. Akibatnya, data tersebut belum dapat dieksplorasi sepenuhnya untuk menghasilkan informasi yang bernilai. Seiring berkembangnya teknologi, data tersebut bisa dioptimalkan menggunakan data mining untuk menggali informasi yang lebih luas dari data tersebut [5]. Salah satu metode data mining yaitu *K-Means*, untuk mengelompokkan data berdasarkan karakteristiknya [6]. Algoritma *K-Means* merupakan algoritma clustering yang akan mengempokkan data menjadi kluster atau kelompok-kelompok tertentu berdasarkan kemiripan karakteristik datanya [7]. Dengan menerapkan algoritma ini, instansi dapat mengetahui kelompok-kelompok pegawai berdasarkan pola kehadiran pegawai. Informasi tersebut akan sangat membantu sebagai dasar untuk analisis lanjutan, seperti evaluasi kinerja dan kedisiplinan.

Beberapa penelitian sebelumnya telah menerapkan algoritma *K-Means* untuk mengolah data kehadiran pegawai. Penelitian Muharmi dan Nadriati tahun 2022 menerapkan algoritma *K-Means Clustering* untuk menganalisis data kehadiran karyawan dengan tujuan untuk mengelompokkan pegawai berdasarkan pola kehadiran [8]. Penelitian Darma, Yusman, dan Hendrawan tahun 2024 yang melakukan klasterisasi data kehadiran pegawai di Dinas Pekerjaan Umum dan Penataan Ruang Kabupaten Langkat [9]. Penelitian Virgo, Defit, dan Yuhandri tahun 2020 menggunakan algoritma *K-Means* untuk mengelompokkan tingkat kehadiran dosen non-PNS di IAIN Batusangkar [10]. Penelitian Adih, Pangestu, Akbar, Purnamasari dan Saprudin tahun 2025 melakukan analisis *K-Means Clustering* pada sistem presensi mobile di Puskesmas Kosambi [11].

Dari beberapa penelitian terdahulu terbukti bahwa *K-Means* efektif untuk melakukan klasterisasi kehadiran pegawai. Pada penelitian ini akan memanfaatkan algoritma *K-Means* untuk menganalisis data absensi pegawai di Dinas Kesehatan Kota Pekalongan. Penelitian ini akan melakukan pengelompokkan dengan mempertimbangkan kluster optimal yang digunakan melalui metode *elbow* dan evaluasi akurasi hasil kluster menggunakan *Davies-Bouldin Index* (DBI) sebagai evaluasi model. *Davies-Bouldin index* (DBI) merupakan salah satu metode evaluasi kluster, dimana kualitas hasil kluster akan dievaluasi berdasarkan kedekatan dan kemiripan data dalam suatu kluster serta jarak antar kluster yang berbeda [12]. Proses klasterisasi yang dilakukan bukan hanya menghasilkan kelompok berdasarkan pola kehadiran pegawai, tetapi juga diintegrasikan ke dalam visualisasi data. Dengan demikian, bukan hanya jumlah kluster dan anggotanya saja yg diperoleh, tetapi juga gambaran yang lebih detail mengenai karakteristik masing-masing kluster. Sehingga dapat mempermudah analisis dan interpretasi pola kehadiran pegawai. Penelitian ini bertujuan untuk menghasilkan klasterisasi kehadiran pegawai yang memberikan gambaran kondisi kehadiran secara objektif dan mempermudah pemantauan serta identifikasi pola kehadiran pegawai, sehingga dapat digunakan sebagai dasar dalam analisis lanjutan, seperti evaluasi kedisiplinan dan kinerja pegawai.

2. METODE PENELITIAN

Untuk memberikan gambaran yang lebih jelas mengenai tahapan penelitian, alur metodologi penelitian disajikan pada Gambar 1.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Data

Proses pengumpulan data dilakukan dengan teknik dokumentasi, yaitu menyalin dan mengolah data absensi yang tercatat secara elektronik dalam sistem absensi instansi tersebut. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari sistem absensi pegawai Dinas Kesehatan Kota Pekalongan. Data absensi yang dikumpulkan mencakup informasi identitas pegawai, tanggal kehadiran, waktu kehadiran, serta kode kehadiran yang merepresentasikan kondisi kehadiran pegawai, termasuk kehadiran tepat waktu dan keterlambatan. Dari keseluruhan data tersebut, penelitian ini akan memfokuskan analisis pada variabel waktu kehadiran dan keterlambatan pegawai karena dinilai paling relevan untuk mengidentifikasi kecenderungan perilaku waktu kehadiran pegawai. Adapun rentang waktu data yang digunakan mencakup periode 3 bulan, yakni bulan april hingga juni tahun 2025. Pemilihan periode tersebut dimaksudkan untuk memberikan gambaran representatif mengenai pola kehadiran pegawai dalam satu triwulan.

2.2 Pra-pemrosesan data

Pra-pemrosesan data merupakan tahap pembersihan dan penyesuaian data untuk meningkatkan kualitas data yang digunakan [13]. Proses pembersihan data dilakukan dengan menghilangkan data kosong agar data yang digunakan bersih dan konsisten. Selanjutnya data akan di transformasi dengan mencakup 3 langkah utama yaitu *discretization*, *attribute construction* dan normalisasi data. Proses *discretization* dilakukan untuk mengubah data waktu kehadiran pegawai yang bersifat kontinu menjadi data kategorikal. Tahapan ini bertujuan untuk menyederhanakan representasi data sehingga pola distribusi waktu kehadiran pegawai dapat diidentifikasi dengan lebih jelas. Pada proses ini, waktu kehadiran dibagi ke dalam tiga keterangan waktu yang akan dikodekan menjadi W1, W2 dan W3 yang merepresentasikan tingkat ketepatan waktu kedatangan pegawai. Sementara itu, informasi keterlambatan pegawai diperoleh dari kode keterlambatan yang telah tersedia pada data absensi, yaitu M2, M3, dan M4. Tahap selanjutnya adalah *attribute construction*, yaitu pembentukan atribut baru yang merepresentasikan frekuensi kemunculan masing-masing kode hasil *discretization* waktu kehadiran dan kode keterlambatan pegawai. Melalui proses *discretization* dan *attribute construction* tersebut, terbentuk enam fitur baru yang berasal dari variabel waktu kehadiran dan keterlambatan pegawai yaitu W1, W2, W3, M2, M3 dan M4, yang selanjutnya digunakan sebagai input dalam proses klasterisasi. Selanjutnya, dilakukan normalisasi untuk mengubah data ke dalam skala tertentu. Proses normalisasi data tersebut bertujuan agar setiap fitur memiliki kontribusi yang seimbang terhadap proses pembentukan klaster.

2.3 Klasterisasi

Pada tahap ini akan dilakukan proses pengelompokan data kehadiran pegawai dengan menggunakan *K-Means* sebagai metode klasterisasi. Algoritma ini akan untuk membagi data ke dalam sejumlah klaster berdasarkan kemiripan karakteristik antar data [14]. Proses pengelompokan data pada algoritma *K-Means* dilakukan dengan menghitung jarak antara setiap data dan pusat klaster (centroid) yang dirumuskan sebagai berikut:

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

Keterangan :

- $d(x_i, c_j)$ = Jarak antara titik data x dan centroid C.
- $x_{(ik)}$ = nilai atribut ke-k dari data ke-i
- $x_{(jk)}$ = nilai atribut ke-k dari centroid ke-j
- n = jumlah atribut

Sehingga data dengan pola atau karakteristik yang serupa akan tergabung dalam klaster yang sama. Proses klasterisasi akan dilakukan menggunakan aplikasi Altair AI Studio. Sebelum dilakukan klasterisasi, terlebih dahulu ditentukan jumlah klaster optimal menggunakan metode *elbow*. Hal ini bertujuan untuk mengidentifikasi titik optimal jumlah klaster berdasarkan titik *elbow* yang terbentuk. Dengan demikian, jumlah klaster yang dipilih merupakan hasil analisis yang paling representatif terhadap struktur data yang digunakan.

2.4 Evaluasi

Hasil klaster yang telah terbentuk kemudian dilakukan proses evaluasi untuk menilai kualitas klasterisasi yang dihasilkan. Proses evaluasi akan dilakukan menggunakan *Davies-Bouldin Index* (DBI) yang akan mengukur tingkat kohesi (kedekatan data dalam satu klaster) serta separasi (jarak atau perbedaan antar klaster). Nilai DBI yang dihasilkan akan menjadi indikator kualitas klaster, di mana semakin kecil nilai DBI menunjukkan bahwa klaster yang terbentuk memiliki struktur yang lebih baik [15].

2.5 Visualisasi

Tahap visualisasi dilakukan untuk memahami pola dan karakteristik dari hasil klasterisasi yang diperoleh [16]. Visualisasi berfungsi sebagai sarana eksplorasi dan penyajian hasil sehingga lebih mudah dipahami, dianalisis dan diinterpretasikan. Melalui visualisasi, hubungan antar klaster serta perbedaan karakteristik di antara kelompok data dapat ditampilkan secara jelas dan informatif. Pada penelitian ini, proses visualisasi dilakukan dengan menggunakan Microsoft Power BI untuk menampilkan hasil klasterisasi dalam bentuk diagram yang interaktif. Penggunaan Power BI memungkinkan peneliti menyoroti pola-pola tertentu pada data, seperti distribusi waktu kehadiran pegawai pada setiap klaster.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Berdasarkan proses pengumpulan data, diperoleh data absensi pegawai Dinas Kesehatan Kota Pekalongan untuk periode April hingga Juni 2025. Dataset tersebut mencakup informasi nama pegawai, tanggal kehadiran, waktu kehadiran, keterlambatan serta kode kehadiran harian pegawai.

Total data yang berhasil dikumpulkan dan digunakan dalam penelitian ini berjumlah 66 data kehadiran pegawai. Dari keseluruhan atribut yang tersedia variabel waktu kehadiran dan keterlambatan pegawai menjadi fitur utama dalam klasterisasi. Variabel tersebut dipilih karena mampu merepresentasikan pola dan kecenderungan perilaku waktu kehadiran pegawai. Tabel 1 menyajikan contoh data kehadiran pegawai selama satu bulan sebagai representasi dari keseluruhan data tiga bulan yang digunakan dalam penelitian.

Tabel 1. Contoh data kehadiran pegawai (Periode juni 2025)

Pegawai	Tanggal										
	1	2	3	4	5	26	27	28	29	30
1	HL	06:57	07:03	07:20	07:02	07:07	HL			DL
2	HL	07:11	07:13	DD	07:29	M2	HL			07:21
3	HL	CB	CB	CB	CB	07:08	HL			07:12
4	HL	07:13	07:16	07:15	07:16	07:14	HL			CT
5	HL	07:14	07:24	07:27	M2	07:27	HL			07:24
.....
62	HL	07:13	07:17	07:10	07:16		07:13	HL			07:17
63	HL	07:22	07:24	07:22	07:25	CT	HL			07:25
64	HL	07:01	07:03	07:02	06:53	06:50	HL			06:49
65	HL	07:07	07:10	06:59	CT	07:08	HL			07:12
66	HL	07:12	07:25	07:28	07:19	07:23	HL			07:16

Dalam data tersebut terdapat beberapa data kosong pada tanggal-tanggal tertentu, hal ini menandakan perlunya tahapan pra-pemrosesan data sebelum dilakukan analisis lebih lanjut. Data waktu kehadiran yang diperoleh masih dalam format waktu (HH:MM) dan keterlambatan masih bersifat non-numerik, sehingga belum dapat digunakan secara langsung dalam proses klasterisasi. Oleh karena itu, diperlukan tahap preprocessing dan transformasi data untuk mengonversi data ke dalam bentuk numerik.

3.2 Pra-pemrosesan data

Berdasarkan hasil dari pengumpulan data, data absensi pegawai yang diperoleh masih mengandung beberapa data kosong pada tanggal tertentu serta memiliki atribut yang belum sepenuhnya siap untuk dianalisis menggunakan metode klusterisasi. Oleh karena itu, dilakukan tahapan pra-pemrosesan data untuk memastikan kualitas dan konsistensi data yang digunakan dalam penelitian ini. Tahap pra-pemrosesan data dilakukan dengan menghapus data yang memiliki nilai kosong dan penyesuaian kolom data, sehingga hanya data yang lengkap dan valid yang digunakan dalam proses analisis selanjutnya. Contoh hasil pra-pemrosesan data dapat dilihat pada tabel 2.

Tabel 2. Hasil pembersihan dan penyesuaian data

Pegawai	1	2	3	4	5	23	24	25	26	27	30
1	HL	06:57	07:03	07:20	07:02	07:13	07:08	06:50	07:07	HL	DL
2	HL	07:11	07:13	DD	07:29	M2	07:23	DD	M2	HL	07:21
3	HL	CB	CB	CB	CB	CB	CB	CB	07:08	HL	07:12
4	HL	07:13	07:16	07:15	07:16	07:17	07:10	07:16	07:14	HL	CT
5	HL	07:14	07:24	07:27	M2	07:25	07:24	07:29	07:27	HL	07:24
.....
62	HL	07:13	07:17	07:10	07:16	07:14	07:13	07:26	07:13	HL	07:17
63	HL	07:22	07:24	07:22	07:25	07:22	07:22	07:24	CT	HL	07:25
64	HL	07:01	07:03	07:02	06:53	CD	06:48	07:11	06:50	HL	06:49
65	HL	07:07	07:10	06:59	CT	07:15	CT	07:20	07:08	HL	07:12
66	HL	07:12	07:25	07:28	07:19	07:22	07:23	L2	07:23	HL	07:16

Pada tahap transformasi data, waktu kehadiran dan keterlambatan yang semula bersifat non-numerik dikonversi ke dalam bentuk numerik agar dapat digunakan dalam proses analisis klusterisasi. Atribut waktu kehadiran yang awalnya direpresentasikan dalam format waktu (HH:MM) ditransformasikan melalui proses *discretization* menjadi beberapa keterangan waktu. Dalam penelitian ini, waktu kehadiran dibagi ke dalam 3 interval waktu yang merepresentasikan tingkat ketepatan waktu kedatangan pegawai. Setiap interval tersebut selanjutnya dikodekan ke dalam bentuk numerik, sebagaimana disajikan pada Tabel 3.

Tabel 3. Discretization

Kode	Keterangan
W1	<07:00
W2	07:00 - 07:15
W3	07:15 – 07:30

Selain variabel hasil diskretisasi, data absensi juga memiliki variabel bawaan dari sistem berupa kode keterlambatan otomatis yang dihasilkan secara langsung oleh perangkat absensi. Kode ini tidak melalui proses diskretisasi manual, melainkan ditetapkan oleh sistem berdasarkan rentang waktu keterlambatan yang telah dikonfigurasi sebelumnya. Atribut ini memiliki kode M2, M3, dan M4 sebagaimana disajikan pada Tabel 4.

Tabel 4. Variabel keterlambatan

Kode	Keterangan
M2	Terlambat 1 - 15 menit
M3	Terlambat 16 – 30 menit
M4	Terlambat >30 menit

Setelah dilakukan proses diskretisasi terhadap atribut jam kehadiran (W) dan identifikasi kode keterlambatan dari sistem (M), proses selanjutnya adalah *attribute construction* yang bertujuan untuk membentuk atribut baru melalui penjumlahan frekuensi dari setiap kode keterangan. Dalam hal ini kode W1, W2, W3, M2, M3 dan M4 yang terbentuk sebelumnya akan dijadikan atribut baru. Pada tahap ini, setiap kode dihitung berdasarkan frekuensi kemunculannya dalam data kehadiran masing-masing pegawai selama periode 3 bulan, Hasil pembentukan atribut tersebut disajikan pada Tabel 5.

Tabel 5. Hasil *discretization* dan *attribute construction*

Pegawai	W1	W2	W3	M2	M3	M4
1	6	30	8	0	0	0
2	6	23	12	5	0	0
3	0	17	0	0	0	0
4	3	29	14	0	0	0
5	1	2	42	1	0	0
.....
62	17	24	7	0	0	0
63	0	1	46	0	0	0
64	19	31	0	0	0	0
65	1	36	9	0	0	0
66	0	7	27	9	0	0

Tabel 5 merupakan tabel hasil transformasi data yang telah dilakukan. Proses tersebut diakhiri dengan melakukan rekapitulasi pada kode keterangan tiap data pegawai. Proses ini menghasilkan data numerik pada tiap-tiap atribut. Dengan data numerik ini, setiap pegawai dapat dikelompokkan berdasarkan pola waktu yang dimiliki, serta memungkinkan identifikasi pola kehadiran yang jelas.

Setelah melakukan transformasi, kemudian data tersebut akan dilakukan normalisasi dengan menggunakan teknik normalisasi *Min-Max Normalization* atau *range transformation*. Proses ini akan mentransformasi data menjadi nilai baru dengan rentang 0 – 1 seperti pada Tabel 6.

Tabel 6. Hasil normalisasi

Pegawai	W1	W2	W3	M2	M3	M4
1	0.182	0.833	0.160	0	0	0
2	0.182	0.639	0.240	0.156	0	0
3	0	0.472	0	0	0	0
4	0.091	0.806	0.280	0	0	0
5	0.030	0.056	0.840	0.031	0	0
.....
62	0.515	0.667	0.140	0	0	0
63	0	0.028	0.920	0	0	0
64	0.576	0.861	0	0	0	0
65	0.030	1	0.180	0	0	0
66	0	0.194	0.540	0.281	0	0

Setelah dilakukan normalisasi, semua data memiliki skala yang sama sehingga tidak ada lagi data yang terlalu besar maupun terlalu kecil dibandingkan variabel lainnya. Dengan demikian, hasil analisis pada tahap berikutnya menjadi lebih adil karena setiap variabel memberi pengaruh yang seimbang dalam pembentukan kluster.

3.3 Klusterisasi

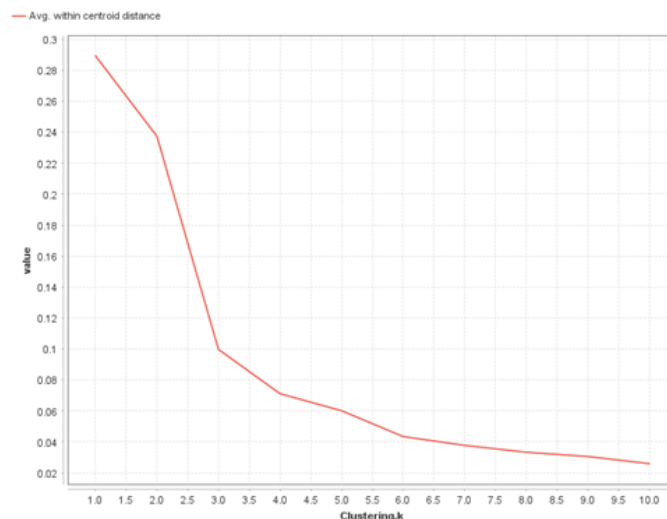
Sebelum melakukan proses klusterisasi, akan dilakukan penentuan nilai *k*, langkah ini dilakukan untuk mendapatkan jumlah kluster optimal. Pada penelitian ini penentuan nilai *k* dilakukan dengan menggunakan metode *elbow*, dimana metode ini akan memberikan rekomendasi jumlah kluster optimal dengan melihat penurunan nilai *Avg. Within Centroid Distance* pada tiap nilai *k*. Tabel merupakan hasil perhitungan *Avg. Within Centroid Distance* yang dilakukan pada *k* = 1 hingga *k* = 10. Adapun nilai yang diperoleh seperti pada Tabel 7.

Tabel 7. Nilai *avg. within centroid distance*

Clustering	<i>Avg. Within Centroid Distance</i>
1	0.289

Clustering	Avg. Within Centroid Distance
2	0.237
3	0.100
4	0.071
5	0.060
6	0.044
7	0.039
8	0.033
9	0.030
10	0.026

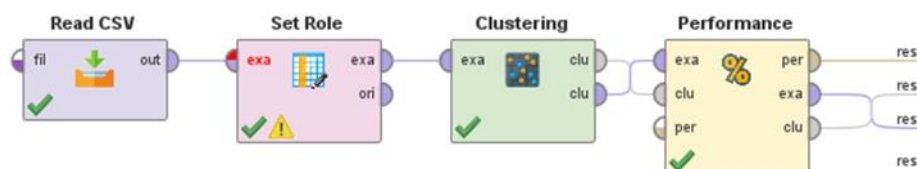
Dari hasil perhitungan pada Tabel 6, berikutnya di visualisasikan dalam bentuk grafik pada gambar. Visualisasi ini dilakukan untuk mempermudah identifikasi titik elbow pada penurunan nilai *Avg. Within Centroid Distance* seiring penambahan nilai k. Grafik hasil visualisasi tersebut dapat dilihat pada Gambar 2.



Gambar 2. Grafik Elbow

Pada Gambar 2 menggambarkan penurunan nilai ekstrim terjadi pada titik titik k=2 dan titik k = 3, kemudian nilai tersebut mulai turun secara perlahan setelah titik k = 3 hingga k = 10. Jadi dapat dimaknai bahwa titik elbow berada di k = 3, karena pada titik tersebut terjadi transisi penurunan nilai yang signifikan menuju penurunan yang melandai. Dengan demikian proses klusterisasi akan diterapkan jumlah kluster optimal yaitu 3 kluster.

Setelah melalui tahapan proses transformasi data penentuan nilai k optimal, kini data tersebut siap digunakan untuk klusterisasi. Proses klusterisasi akan dilakukan menggunakan K-Means clustering. Dengan menggunakan aplikasi Altair AI Studio, proses ini akan melakukan pengelompokkan data menjadi 3 kluster seperti yang telah ditentukan pada proses sebelumnya. Tahapan proses tersebut ditunjukkan pada Gambar 3.



Gambar 3. Proses Klusterisasi K-Means

Gambar 3 merupakan alur proses klusterisasi pada penelitian ini yang terdiri dari 4 tahap. Tahap pertama adalah mengambil dan membaca data hasil dari transformasi menggunakan operator Read CSV. Dimana atribut Pegawai, W1, W2, W3, M2, M3 dan M4 akan dimuat. Setelah data diambil selanjutnya menentukan peran suatu atribut menggunakan operator *set role*. Pada konteks ini atribut Pegawai akan dijadikan id. Proses ini dilakukan agar algoritma mengetahui atribut mana yang akan dipakai untuk perhitungan sehingga hasil perhitungannya menjadi optimal. Kemudian melakukan klusterisasi dengan operator

Clustering. Pada proses ini data akan di kelompokkan kedalam 3 klaster berdasarkan pola atau karakteristik tertentu. Pada tahap terakhir menggunakan operator performance yang akan mengidentifikasi tingkat akurasi model yang telah dibuat.

Pengelompokkan data menjadi 3 klaster menghasilkan cluster model yang memiliki perbedaan karakteristik atau pola antar klaster sebagaimana ditunjukkan pada Gambar 4.

Cluster Model

```
Cluster 0: 17 items
Cluster 1: 3 items
Cluster 2: 46 items
Total number of items: 66
```

Gambar 4. Hasil Klasterisasi

Gambar 4 menunjukkan keanggotaan dari masing-masing klaster yang terbentuk. Dimana pada klaster 0 terdapat 17 data pegawai. Kemudian pada klaster 1 memiliki anggota yang paling sedikit yaitu terdapat 3 data pegawai. Kluster 2 menjadi yang paling dominan dalam klasterisasi ini dengan 46 data pegawai.

3.4 Evaluasi

Setelah dilakukan klasterisasi menggunakan algoritma K-means, tahap berikutnya adalah melakukan evaluasi hasil klasterisasi untuk melihat performa model yang telah dihasilkan. Proses evaluasi akan menggunakan *Davies-Bouldin Index* (DBI). Hasil evaluasi tersebut dapat dilihat pada Gambar 5.

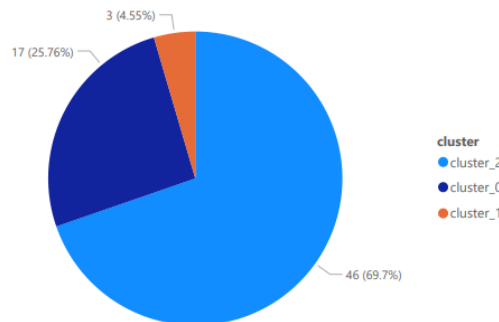


Gambar 5. Nilai DBI

Bedasarkan gambar 5 diperoleh nilai DBI sebesar 0.743. Nilai ini menunjukkan bahwa hasil klasterisasi yang terbentuk tergolong cukup baik, karena nilai DBI berada di bawah angka 1 yang umumnya digunakan sebagai batas kualitas klaster yang dapat diterima.

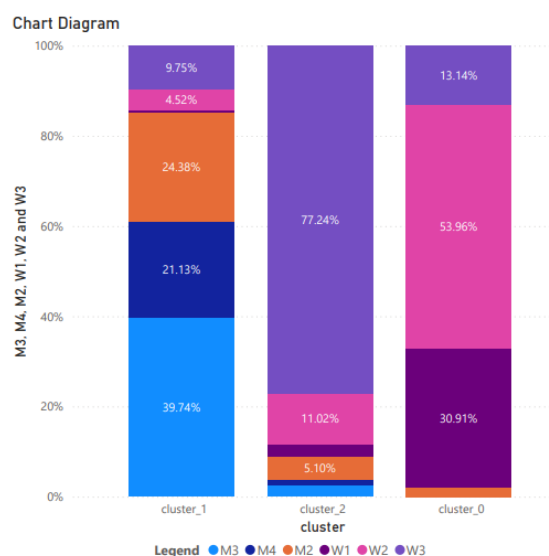
3.5 Visualisasi

Data yang telah dikelompokkan dengan algoritma K-Means pada proses sebelumnya, kemudian divisualisasikan untuk mempermudah dalam interpretasi. Visualisasi hasil klasterisasi tersebut dapat dilihat pada Gambar 6, Gambar 7 dan Gambar 8.



Gambar 6. Distribusi Data pada Kluster

Pada Gambar 6 menggambarkan hasil visualisasi persebaran atribut dan data pada setiap kluster. Pada diagram pie menunjukkan distribusi data dalam kluster berdasarkan jumlah data pegawai pada tiap kluster. Terlihat bahwa kluster 2 mendominasi dengan proporsi sebesar 69,7% atau 46 data pegawai. Diikuti dengan kluster 0 dengan proporsi sebesar 25,76% dari total data atau 17 data pegawai. Sementara itu kluster 1 menjadi bagian terkecil yaitu 4,55% dari total data atau hanya terdapat 3 data pegawai yang termasuk kedalam kluster tersebut. Hal ini menggambarkan bahwa sebagian besar data dikelompokkan kedalam kluster 2, sedangkan kluster 1 memiliki distribusi data yang relatif kecil sehingga dapat dianggap sebagai kelompok minoritas dalam pembagian kluster.

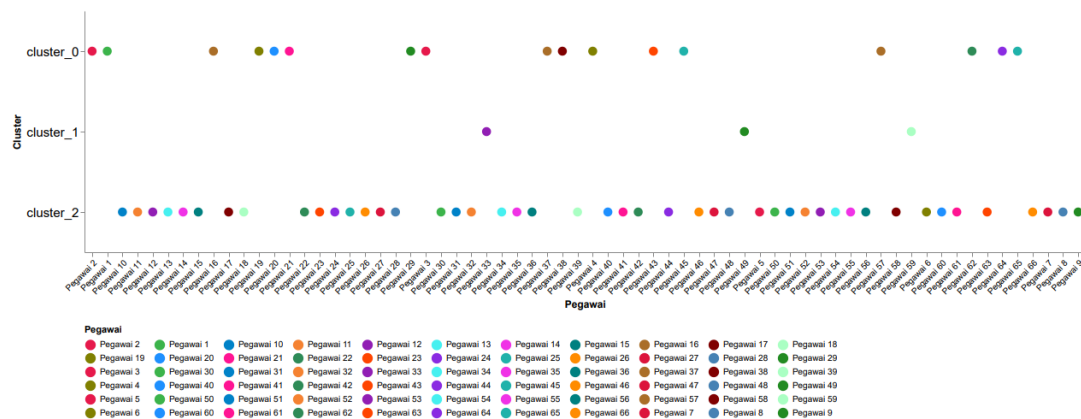


Gambar 7. Distribusi Atribut pada Kluster

Gambar 7 merupakan hasil visualisasi distribusi atribut pada masing-masing kluster yang dibuat dengan diagram chart. Dengan visualisasi akan memudahkan pemahaman karakteristik pada masing-masing kluster. Adapun interpretasi pada masing-masing kluster adalah seperti berikut.

1. Pada *Cluster0* didominasi oleh W2 (53,96%), disusul oleh M3 (30,91%) dan W3 (13,14%). Proporsi ini menunjukkan bahwa sebagian besar pegawai dalam kluster ini datang pada pukul 07.00–07.15, namun masih ada sebagian yang terlambat antara 16–30 menit. Dengan demikian, kluster 0 dapat dikategorikan sebagai kelompok pegawai dengan waktu kehadiran cukup baik.
2. Pada *Cluster1* didominasi oleh kategori M3 (39,74%), yaitu pegawai yang terlambat antara 16–30 menit, diikuti oleh M2 (21,13%) yang terlambat antara 1–15 menit, serta M4 (24,38%) yang terlambat lebih dari 30 menit. Hanya sebagian kecil yang datang lebih awal (W1 4,52% dan W3 9,75%). Hal ini menunjukkan bahwa kluster 1 berisi pegawai dengan kecenderungan sering datang terlambat. Dengan sebagian besar pegawai datang terlambat antara 16-30 menit. Sehingga pada cluster 1 bisa dikategorikan sebagai kelompok pegawai dengan waktu kehadiran yg kurang baik.

3. Pada *Cluster2* secara dominan terdiri dari W3 (77,24%), yaitu pegawai yang datang antara pukul 07.16–07.30, diikuti oleh W2 (11,02%) dan W1 (5,10%) dengan proporsi M2, M3, M4 sangat kecil. Dengan demikian, klaster 2 menggambarkan kelompok pegawai yang datang tepat waktu, dengan sebagian besar pegawai datang mendekati jam masuk yaitu pukul 07:16-07:30, menunjukkan disiplin waktu yang tinggi dibandingkan dua klaster lainnya. Sehingga pada cluster 1 bisa dikategorikan sebagai kelompok pegawai dengan waktu kehadiran yg baik.



Gambar 8. Distribusi pegawai dalam klaster

Gambar 8 menunjukkan hasil visualisasi data pegawai yang telah dikelompokkan kedalam 3 klaster. Pada sumbu horizontal merepresentasikan data pegawai dari Pegawai 1 hingga Pegawai 66, sedangkan sumbu vertikal merepresentasikan hasil klaster yang terbentuk. Setiap titik pada grafik merepresentasikan satu pegawai yang tergolong ke dalam klaster tertentu berdasarkan kemiripan pola kehadirannya. Berikut adalah interpretasi visualisasi tersebut.

1. Pada cluster 0 atau kelompok pegawai dengan waktu kehadiran cukup baik, terdapat 17 data pegawai yaitu pegawai 1, 2, 3, 4, 16, 19, 20, 21, 29, 37, 38, 43, 45, 57, 62, 64, dan 65.
2. Pada cluster 1 atau kelompok pegawai dengan waktu kehadiran yang kurang baik, terdapat 3 pegawai yaitu pegawai 33, 49, dan 59.
3. Pada cluster 2 atau kelompok pegawai dengan waktu kehadiran yg baik terdapat 46 pegawai yaitu Adapun anggota pada klaster ini yaitu pegawai 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 34, 35, 36, 39, 40, 41, 42, 44, 46, 47, 48, 50, 51, 52, 53, 54, 55, 56, 58, 60, 61, 63, dan 66.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa proses pengelompokan data kehadiran pegawai menggunakan algoritma *K-Means clustering* berhasil dilakukan dengan baik. Hasil klasterisasi menunjukkan bahwa data terbagi kedalam 3 kelompok, yaitu waktu kehadiran baik, cukup baik dan kurang baik. Nilai *Davies-Bouldin Index* sebesar 0,743 menunjukkan bahwa hasil klasterisasi yang terbentuk tergolong cukup baik dan dapat merepresentasikan pola kehadiran pegawai. Analisis dan visualisasi menggunakan *Power BI* membantu dalam memperjelas perbedaan karakteristik tiap klaster dan memudahkan interpretasi hasil secara visual dan interaktif. Hasil penelitian ini memberikan gambaran umum mengenai kecenderungan waktu kehadiran pegawai yang dapat dijadikan dasar untuk analisis lanjutan, seperti evaluasi kinerja dan kedisiplinan. Penelitian selanjutnya dapat mengembangkan analisis dengan menggunakan algoritma klasterisasi lain, seperti *Hierarchical Clustering*, *Fuzzy C-Means*, atau *DBSCAN*, guna membandingkan tingkat akurasi dan kualitas hasil klasterisasi terhadap pola kehadiran pegawai.

DAFTAR PUSTAKA

- [1] K. L. P. Situmorang and J. Manurung, "Sistem Pendukung Keputusan Penilaian Kinerja Pegawai RSUD Dr. Hadrianus Sinaga Dengan Menggunakan Metode Multi Factor Evaluation Process," *Jurnal Teknik Informatika UNIKA Santo Thomas*, vol. 06, pp. 354–366, 2021, doi: 10.54367/jtiust.v6i2.1557.

- [2] R. Fitria, I. Sahputra, and R. Maulana, "Optimization of Employee Attendance System Using Face Recognition and Geotagging Based on Mobile Android," *Journal of Artificial Intelligence and Software Engineering*, vol. 5, no. 2, pp. 510–527, 2025, doi: 10.30811/jaise.v5i2.6892.
- [3] L. Stie and D. Dahanai Buntok, "PENGARUH DISIPLIN KERJA TERHADAP KINERJA PEGAWAI DALAM PENENTUAN TITIK BATAS DESA WILAYAH KABUPATEN BARITO SELATAN The Effect of Work Discipline on Employee Performance in Determining the Village Boundary Points in the South Barito Regency", doi: 10.33084/restorica.v5i2.
- [4] P. S. Nabillah and M. A. Rizqi, "EFEKTIVITAS SISTEM DISIPLIN KERJA PADA PT RAVANA JAYA," 2024, [Online]. Available: <https://e-jurnal.nobel.ac.id/index.php/akmen>
- [5] L. Firlian, R. Fadila, M. K. Ridho, E. Sunandi, and U. Aflah, "Pendekatan Data Mining dalam Optimalisasi Margin Penjualan Adidas: Studi Klasterisasi dengan K-Means dan Fuzzy C-Means," *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi*, vol. 13, no. 2, pp. 229–237, Jul. 2025, doi: 10.37905/euler.v13i2.32417.
- [6] U. Arfan and N. Paraga, "Perbandingan Algoritma K-Means, Naïve Bayes dan Decision Tree Dalam Memprediksi Penjualan Bahan Bakar Minyak," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1379–1389, Jul. 2024, doi: 10.57152/malcom.v4i4.1566.
- [7] W. Alfian, Kusriani, and T. Hidayat, "Analisis Clustering Pegawai Berdasarkan Tingkat Kedisiplinan Menggunakan Algoritma K-Means dan Davies-Bouldin Index," *Journal of Electrical Engineering and Computer (JEECOM)*, vol. 6, no. 2, pp. 437–448, 2024, doi: 10.33650/jeecon.v6i2.9556.
- [8] Y. Muharmi and S. Nadriati, "Analysis Of Employee Discipline Based On Digital Attendance With The K-Means Algorithm Method," *Jurnal Teknologi Dan Open Source*, vol. 5, no. 2, pp. 115–135, 2022, doi: 10.36378/jtos.v5i2.2628.
- [9] S. Darma, Y. Yusman, and J. Hendrawan, "Analisis Data Tingkat Kehadiran Pegawai dengan Menggunakan Clustering K-Means Pada Dinas Pekerjaan Umum dan Penataan Ruang Kabupaten Langkat," *Jurnal Minfo Polgan*, vol. 13, no. 1, pp. 1105–1116, 2024, doi: 10.33395/jmp.v13i1.13958.
- [10] I. Virgo, S. Defit, and Y. Yuhandri, "Klasterisasi Tingkat Kehadiran Dosen Menggunakan Algoritma K-Means Clustering," *Jurnal Sistim Informasi dan Teknologi*, vol. 2, pp. 23–28, 2020, doi: 10.37034/jsisfotek.v2i1.17.
- [11] Adih, W. A. D. Pangestu, M. F. Akbar, Purnamasari, and Saprudin, "Analisis K-Means Clustering pada Sistem Presensi Mobile dengan Fitur GPS Radius dan Foto Selfie untuk Pegawai Non-PNS di Puskesmas Kosambi," *Modem: Jurnal Informatika dan Sains Teknologi*, vol. 3, no. 1, pp. 15–30, Dec. 2024, doi: 10.62951/modem.v3i1.324.
- [12] N. Eliza, I. Husein, and S. Dur, "Determining Zoning of Areas Affected by Flood Disasters in Medan City Using Silhouette Coefficient and Davies Bouldin Index Analysis," *Jurnal Pijar Mipa*, vol. 19, no. 3, pp. 558–563, May 2024, doi: 10.29303/jpm.v19i3.6707.
- [13] A. Yufita, R. Kurniawan, Y. W. Arie, and T. Suprpti, "Optimalisasi Klasterisasi Tenaga Kesehatan Menggunakan K-Means dan Davies Bouldin Index," *IJAI (Indonesian Journal of Applied Informatics)*, vol. 9, pp. 452–461, Feb. 2025, doi: <https://dx.doi.org/10.20961/ijai.v9i2.96645>.
- [14] A. Fitriani, E. Arfi, and A. Huda, "Penerapan Algoritma K-Means Clustering dalam Memetakan Produktivitas Lokasi Perkebunan Nanas PT Great Giant Pineapple," *Journal of Mathematics, Computations and Statistics*, vol. 7, no. 2, pp. 215–231, Sep. 2024, doi: 10.35580/jmathcos.v7i2.4200.
- [15] F. K. Wardani and R. Kurniawan, "PENGELOMPOKAN RESPONSIVITAS HASIL KUESIONER TERHADAP PENGGUNAAN PERPUSTAKAAN DIGITAL MENGGUNAKAN ALGORITMA K-MEANS," 2024, doi: 10.33795/jip.v10i2.4850.
- [16] D. Marcelina, A. Kurnia, and T. Terttiaavini, "Analisis Klaster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 293–301, Nov. 2023, doi: 10.57152/malcom.v3i2.952.