SPEKTRUM AKADEMIKA
Volume. 1, Nomor. 2, 2026

SEGMENTASI POLA PENCATATAN PRESENSI PEGAWAI MENGGUNAKAN GAUSSIAN MIXTURE MODEL

Muhammad Handika Maulana Sifa[1], Yuniar Rachmawati[2], Muhammad Nasrudin[3], Mohammad Idhom[4]

[1,2,3,4] Program Studi Sains Data, Universitas Pembangunan Nasional Veteran Jawa Timur, Indonesia
Email: sifahandika23@gmail.com[1], yuniarrachmaaa@gmail.com[2], nasrudin.fasilkom@upnjatim.ac.id[3], idhom@upnjatim.ac.id[4]

ABSTRAK

Penelitian ini mengidentifikasi pola pencatatan presensi pegawai menggunakan Gaussian Mixture Model (GMM) pada perusahaan jasa energi yang dianonimkan. Data September 2026 mencakup 6.464 observasi layak. Hari masuk tercatat dan kelengkapan catatan pulang distandardisasi, lalu diuji pada K=2–8. K=2 dipilih dengan Silhouette 0,4799, DBI 0,9986, CH 5.211,91, dan rata-rata peringkat 3,40. K=8 memiliki BIC dan AIC terendah, tetapi pemisahan labelnya lebih lemah. Segmen 1 mencakup 32,22% observasi, dengan rata-rata 14,1 hari masuk dan kelengkapan pulang 50,7%; Segmen 2 mencakup 67,78%, dengan 22,9 hari dan kelengkapan 98,0%. Profil mendukung usulan pemeriksaan administratif dan pemeliharaan kualitas pencatatan. Hasil bersifat deskriptif satu periode dan tidak membuktikan kedisiplinan individu maupun dampak intervensi.

Kata kunci: People Analytics; presensi; GMM; segmentasi; kualitas pencatatan

ABSTRACT

This study identifies employee attendance-recording patterns using a Gaussian Mixture Model (GMM) in an anonymized energy-services company. September 2026 data included 6,464 eligible observations. Recorded check-in days and check-out completeness were standardized, and models with two to eight components were evaluated. The two-component model was selected with Silhouette 0.4799, DBI 0.9986, CH 5,211.91, and average rank 3.40. The eight-component model had the lowest BIC and AIC but weaker label separation. Segment 1 represented 32.22% of observations, averaging 14.1 recorded days and 50.7% completeness; Segment 2 represented 67.78%, averaging 22.9 days and 98.0% completeness. Profiles inform administrative verification and record-quality maintenance. Findings describe one period and do not establish employee discipline or intervention effects.

Keywords: People Analytics; attendance records; GMM; segmentation; record quality

1. Pendahuluan

Data presensi merupakan salah satu sumber informasi operasional yang paling dekat dengan kegiatan administrasi sumber daya manusia. Setiap pencatatan masuk dan pulang membentuk jejak yang dapat diringkas untuk memahami variasi ketercakupan data antarpegawai. Namun, tersedianya sistem elektronik tidak serta-merta menghasilkan informasi yang siap digunakan untuk pengambilan keputusan. Rekap yang panjang dapat memperlihatkan jumlah catatan tanpa menjelaskan kelompok mana yang membutuhkan

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

pemeriksaan lebih lanjut. Penelitian pada instansi pemerintah menunjukkan bahwa arsip kehadiran yang semula disimpan untuk administrasi dapat diolah menjadi pola melalui klasterisasi (Darma et al., 2024).

Persoalan yang dikaji dalam penelitian ini adalah heterogenitas pola pencatatan presensi pada sebuah perusahaan jasa energi di Indonesia. Sebagian pegawai mempunyai banyak hari masuk tercatat dan catatan pulang yang hampir lengkap, sedangkan sebagian lainnya mempunyai frekuensi pencatatan lebih rendah atau kelengkapan pulang yang bervariasi. Perbedaan tersebut penting bagi pengelola administrasi karena pemeriksaan dengan prosedur yang sama untuk seluruh pegawai belum tentu efisien. Segmentasi dapat menyusun prioritas pemeriksaan, selama informasi yang dihasilkan tetap dihubungkan dengan konteks pencatatan dan tidak digunakan untuk menyimpulkan perilaku yang tidak diukur.

Jumlah hari masuk tercatat berbeda dari tingkat kehadiran terhadap kewajiban kerja. Seorang pegawai dapat memiliki sedikit catatan karena cuti, perubahan penempatan, jadwal bergilir, penugasan lapangan, atau permasalahan integrasi sistem. Demikian pula, catatan pulang yang kosong belum membuktikan bahwa pegawai meninggalkan pekerjaan sebelum waktunya. Penelitian ini karena itu menggunakan istilah pola pencatatan presensi secara konsisten. Fokus analisis adalah struktur data administratif yang dapat diamati, sementara penjelasan mengenai kedisiplinan, produktivitas, dan kepatuhan jadwal memerlukan variabel serta validasi tersendiri.

Dalam konteks *People Analytics*, pengelompokan data dapat membantu mengubah ringkasan operasional menjadi dasar pertanyaan manajerial yang lebih terarah. Informasi yang diperlukan pengelola SDM tidak berhenti pada jumlah kelompok, tetapi mencakup ciri setiap kelompok, ketidakpastian penempatannya, serta pemeriksaan yang layak dilakukan. Studi klasterisasi presensi di Indonesia telah menerapkan K-Means untuk mengelompokkan pegawai berdasarkan catatan kehadiran (Mira et al., 2024). Kajian tersebut memperlihatkan relevansi data presensi sebagai objek analisis, tetapi tidak membuktikan bahwa definisi kedisiplinan dari satu organisasi dapat langsung dipindahkan ke organisasi lain.

Penelitian ini menggunakan *Gaussian Mixture Model* (GMM) untuk memperoleh pengelompokan probabilistik. Pendekatan tersebut memodelkan distribusi data sebagai campuran komponen Gaussian dan memberikan probabilitas keanggotaan setiap observasi. Estimasi parameter dapat dilakukan melalui algoritma *Expectation-Maximization* (EM) (Alyarahma et al., 2024). Sifat probabilistik sesuai untuk mengkaji data dengan karakteristik yang berdekatan atau saling bertumpang tindih. Meskipun demikian, probabilitas model tidak dapat dianggap sebagai probabilitas bahwa seorang pegawai melakukan pelanggaran administrasi.

Pemilihan jumlah komponen menjadi persoalan utama karena kriteria evaluasi tidak selalu menghasilkan rekomendasi yang sama. Indeks berbasis jarak menilai pemisahan dan kekompakan label kelompok, sedangkan kriteria informasi menilai kecocokan model probabilistik setelah mempertimbangkan kompleksitasnya. Penelitian perbandingan klasterisasi menunjukkan bahwa pembacaan beberapa indeks diperlukan untuk memahami perbedaan kualitas kelompok (Ramadhan et al., 2025). Oleh sebab itu, pemilihan model perlu dijelaskan melalui tujuan penelitian, karakteristik data, dan konsekuensi interpretasinya, bukan sekadar mengikuti satu angka minimum atau maksimum.

Analitik presensi dan batas pengukuran

Analitik presensi dapat dipahami sebagai kegiatan menyusun, mengolah, dan menafsirkan catatan kehadiran untuk mendukung pertanyaan pengelolaan tenaga kerja. Nilai praktisnya bergantung pada kesesuaian antara data yang tersedia dan keputusan yang hendak didukung. Penelitian tentang pola kehadiran menggunakan K-Means mencakup tahapan pembersihan, pembentukan atribut, normalisasi, evaluasi, serta visualisasi untuk menjelaskan kelompok pegawai (Sumantri et al., 2026). Rangkaian tersebut

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

menunjukkan bahwa klasterisasi merupakan bagian dari proses analisis, sehingga mutu sumber data dan definisi atribut tetap perlu diperhatikan.

Kajian Alfian et al. (2024) menggunakan K-Means dan Davies-Bouldin Index dalam pengelompokan pegawai, dengan memperlihatkan bahwa hasil klaster tetap dapat mengalami tumpang tindih. Temuan tersebut menjadi pengingat bahwa label kelompok perlu dibaca bersama definisi variabelnya. Dalam penelitian ini, label segmen dirumuskan sebagai kebutuhan administratif, seperti peningkatan konsistensi pencatatan pulang atau pemeliharaan kualitas pencatatan. Pilihan istilah ini mengarahkan pengguna pada tindakan pemeriksaan yang dapat dilakukan tanpa menambahkan penilaian moral atau penilaian kinerja yang belum ditopang data.

Segmentasi probabilistik dengan GMM

GMM mengasumsikan bahwa distribusi pengamatan dapat didekati oleh gabungan beberapa komponen Gaussian. Setiap komponen memiliki bobot, vektor rata-rata, serta matriks kovarians. Perbedaan parameter tersebut memungkinkan model merepresentasikan kelompok dengan sebaran yang berbeda. Penelitian di Indonesia telah menggunakan GMM untuk pemetaan karakteristik wilayah dan membandingkannya dengan K-Means berdasarkan indeks klasterisasi (Wahidah & Utari, 2023). Hasil pada domain tersebut memberi dasar penggunaan pendekatan probabilistik, tetapi tidak menjadi bukti keunggulan GMM pada data presensi penelitian ini.

Pada pengelompokan probabilistik, satu observasi memiliki sejumlah probabilitas keanggotaan yang jumlahnya satu. Label akhir diperoleh dari komponen dengan probabilitas terbesar. Prosedur ini memudahkan penyajian profil segmen, tetapi informasi ketidakpastian sebaiknya tetap disimpan. Dua observasi yang memperoleh label sama dapat memiliki tingkat kepastian berbeda. Rata-rata probabilitas maksimum dalam suatu segmen hanya menjelaskan kecenderungan penempatan menurut model yang dilatih; ukuran tersebut bukan akurasi klasifikasi karena penelitian tidak memiliki label kebenaran pembanding.

Algoritma EM memperbarui probabilitas keanggotaan dan parameter komponen secara bergantian. Tahap ekspektasi menghitung tanggung jawab setiap komponen terhadap observasi berdasarkan parameter sementara, sedangkan tahap maksimisasi memperbarui bobot, rata-rata, dan kovarians. Proses diulang hingga memenuhi aturan penghentian. Literatur GMM lokal menjelaskan penggunaan EM untuk mengestimasi parameter campuran (Alyarahma et al., 2024). Pada penerapan praktis, inisialisasi dan regularisasi tetap penting karena optimasi dapat menuju solusi lokal atau menghadapi komponen dengan varians sangat kecil.

Standardisasi dan karakteristik fitur

Standardisasi mengubah fitur menjadi skala yang sebanding melalui pengurangan rata-rata dan pembagian dengan simpangan baku. Jumlah hari masuk mempunyai satuan hari, sedangkan kelengkapan pulang dinyatakan dalam persen. Tanpa penyetaraan skala, perbedaan rentang numerik dapat mendominasi geometri data dan evaluasi berbasis jarak. Penggunaan standardisasi Z-score sebelum klasterisasi juga dijumpai dalam penelitian segmentasi yang membandingkan beberapa algoritma (Diana et al., 2026). Standardisasi tidak mengubah arti substantif fitur dan tidak dengan sendirinya membuat distribusi menjadi normal.

Kedua fitur penelitian mempunyai keterbatasan distribusional yang perlu dipertimbangkan. Jumlah hari merupakan bilangan cacah, sedangkan rasio pulang dibatasi antara nol dan seratus persen serta dapat menumpuk pada kedua batas tersebut. GMM digunakan sebagai pendekatan deskriptif terhadap distribusi ini, bukan sebagai pernyataan bahwa data mentah mengikuti satu distribusi Gaussian. Pemilihan kovarians

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

diagonal menyederhanakan struktur komponen. Konsekuensinya, hubungan antarfitur di dalam setiap komponen tidak dimodelkan sebagai kovarians bebas, walaupun campuran secara keseluruhan tetap dapat memperlihatkan hubungan antarvariabel.

Evaluasi struktur kelompok dan kecocokan model

Silhouette membandingkan kedekatan suatu observasi terhadap kelompoknya sendiri dengan kedekatan terhadap kelompok lain. Nilai yang lebih tinggi menunjukkan pemisahan relatif yang lebih baik menurut jarak yang digunakan. Davies-Bouldin Index (DBI) membandingkan penyebaran dalam kelompok dengan jarak antarpusat kelompok, sehingga nilai lebih rendah diutamakan. Calinski-Harabasz Index (CH) membandingkan variasi antarkelompok dan variasi di dalam kelompok. Ketiga ukuran tersebut menjelaskan aspek struktur partisi yang berbeda dan tidak boleh disamakan dengan persentase ketepatan prediksi (Ramadhan et al., 2025).

Bayesian Information Criterion (BIC) dan Akaike Information Criterion (AIC) menggabungkan kecocokan model dan penalti kompleksitas. Dalam konvensi yang digunakan pada penelitian ini, nilai lebih rendah lebih diutamakan. Penerapan BIC dan AIC untuk mengevaluasi jumlah komponen GMM juga terdapat dalam penelitian segmentasi mahasiswa (Karnaidi & Fudholi, 2026). Namun, komponen yang diperlukan untuk mendekati bentuk distribusi tidak selalu sama dengan jumlah kelompok operasional yang mudah dipisahkan. Oleh karena itu, keputusan segmentasi perlu membedakan kecocokan kepadatan probabilistik dan keterbacaan kelompok administratif.

Posisi penelitian terhadap kajian terdahulu

Tabel 1. Posisi penelitian terhadap kajian terdahulu

Kajian	Objek dan pendekatan	Kaitan dengan penelitian
Mira et al. (2024)	Presensi pegawai; K-Means	Landasan pemanfaatan arsip presensi untuk pengelompokan.
Sumantri et al. (2026)	Pola kehadiran; K-Means dan visualisasi	Pembentukan fitur dan penyajian pola; bukan bukti keunggulan GMM.
Alyarahma et al. (2024)	Indikator kemiskinan; GMM dan EM	Landasan pengelompokan probabilistik dan estimasi parameter.
Wahidah dan Utari (2023)	Indikator kemiskinan; K-Means dan GMM	Pemilihan metode perlu memperhatikan struktur data.
Karnaidi dan Fudholi (2026)	Segmentasi mahasiswa; HDBSCAN dan GMM	Penggunaan BIC/AIC dan penerjemahan profil menjadi tindakan.
Penelitian ini	Presensi September 2026; GMM K=2–8	Dua fitur pencatatan, evaluasi multikriteria, dan usulan verifikasi SDM.

Sumber: Sintesis kajian yang dirujuk, diolah peneliti (2026).

Kesenjangan yang ditangani adalah kebutuhan untuk menerjemahkan evaluasi GMM yang saling berbeda menjadi keputusan segmentasi yang dapat dijelaskan. Penelitian ini menggunakan satu periode, dua fitur pencatatan, dan rentang kandidat yang sama dalam seluruh perbandingan. Profil akhir dikaitkan dengan pemeriksaan administratif yang terukur, sementara keterbatasan mengenai jadwal, pasangan catatan harian, serta stabilitas model dinyatakan secara eksplisit. Dengan demikian, kontribusi empirisnya adalah

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

pemetaan pola pada data September, sedangkan kontribusi praktisnya adalah rancangan tindak lanjut yang tetap memerlukan verifikasi manusia.

Kajian ini dibatasi pada September 2026 dengan unit analisis satu pegawai dalam satu bulan. Pemilihan periode tersebut menjaga keseragaman rentang pengamatan dan menghindari pencampuran observasi pegawai yang sama dari beberapa bulan menjadi seolah-olah individu yang berbeda. Dua fitur digunakan, yaitu jumlah hari masuk tercatat dan persentase kelengkapan catatan pulang. Informasi identitas dipakai hanya untuk menggabungkan catatan dalam pengolahan terbatas, sedangkan naskah menampilkan statistik agregat tanpa nama pegawai, nomor induk, nama unit, atau rincian struktur organisasi.

Tujuan penelitian adalah mengevaluasi GMM dengan dua sampai delapan komponen, menetapkan model yang sesuai untuk segmentasi administratif, menjelaskan profil kelompok yang terbentuk, dan merumuskan usulan tindak lanjut SDM. Kontribusi penelitian terletak pada penjelasan terbuka mengenai perbedaan hasil evaluasi, penggunaan probabilitas keanggotaan secara hati-hati, serta hubungan antara profil segmen dan kebutuhan perbaikan pencatatan. Penelitian ini tidak menguji pengaruh sebab-akibat maupun keberhasilan intervensi. Hasilnya merupakan dasar penyusunan pemeriksaan lanjutan pada periode dan cakupan data yang dianalisis.

2. Metode

2.1 Rancangan penelitian dan sumber data

Penelitian menggunakan pendekatan kuantitatif deskriptif dengan studi kasus pada perusahaan jasa energi yang identitasnya dianonimkan. Data sekunder mencakup master pegawai bulanan dan catatan presensi September 2026. Tanggal pengambilan data adalah 1 Oktober 2026. Cakupan waktu analisis mengikuti tanggal masuk yang tercatat dalam bulan September. Dengan demikian, penelitian memotret kondisi pencatatan yang tersedia saat ekstraksi, bukan perubahan data setelah koreksi administratif pada tanggal berikutnya.

Unit analisis adalah satu identitas pegawai terkode dalam satu bulan. Master pegawai dan ringkasan presensi digabung menggunakan penggabungan luar agar identitas yang hanya muncul pada salah satu sumber tetap terdeteksi. Populasi operasional hasil penggabungan berjumlah 6.790 identitas; istilah ini tidak diasumsikan sama dengan jumlah pegawai aktif organisasi yang telah diverifikasi. Sebanyak 6.464 identitas mempunyai hari masuk tercatat dan masuk ke pemodelan. Sebanyak 326 lainnya tidak membentuk fitur yang layak menurut aturan tersebut dan dipisahkan untuk validasi data.

2.2 Penggabungan data dan pembentukan variabel

Identitas dinormalisasi secara terbatas melalui penghapusan spasi pada bagian awal dan akhir sebelum agregasi. Catatan presensi disaring menurut bulan pada tanggal masuk. Jumlah hari masuk dihitung dari banyaknya tanggal masuk yang berbeda untuk setiap identitas, sehingga beberapa catatan pada tanggal yang sama tidak otomatis menambah jumlah hari. Jumlah hari dengan pulang tercatat dihitung dari tanggal masuk berbeda yang memiliki setidaknya satu isian pulang yang dinilai terisi oleh aturan sistem. Isian kosong serta penanda seperti null, none, nan, nat, nol, dan tanda hubung dikecualikan.

Definisi tersebut mengukur keberadaan isian pulang pada catatan yang terkait dengan tanggal masuk, bukan validasi semantik atas urutan waktu masuk dan pulang. Pengelompokan tanggal yang sama juga belum membuktikan bahwa seluruh sesi kerja pada hari tersebut lengkap. Pada pekerjaan lintas tengah malam atau beberapa sesi dalam satu hari, diperlukan aturan pasangan jadwal yang lebih rinci. Karena itu, istilah kelengkapan dalam penelitian dibatasi pada kelengkapan menurut aturan agregasi yang digunakan, sementara validitas pasangan waktu menjadi keterbatasan yang dibahas pada hasil.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Tabel 2. Definisi operasional variabel

Variabel	Definisi dan satuan	Peran
Identitas terkode	Kunci untuk penggabungan dan agregasi bulanan	Bukan fitur; tidak dipublikasikan
Hari masuk tercatat H_i	Jumlah tanggal masuk berbeda pada September; hari	Fitur model
Hari dengan pulang tercatat P_i	Tanggal masuk berbeda dengan setidaknya satu isian pulang yang memenuhi aturan; hari	Pembentuk rasio
Kelengkapan pulang R_i	$100 \times P_i/H_i$ untuk $H_i > 0$; persen	Fitur model
Anggota dengan catatan belum lengkap	Proporsi anggota segmen dengan setidaknya satu hari belum lengkap; persen	Profil pendukung
Probabilitas maksimum	Probabilitas posterior terbesar per observasi	Profil ketidakpastian model

Sumber: Definisi implementasi analisis presensi, diolah peneliti (2026).

Jika H_i menyatakan jumlah tanggal masuk berbeda dan P_i menyatakan jumlah tanggal masuk berbeda dengan isian pulang yang memenuhi aturan, rasio kelengkapan dihitung melalui Persamaan (1). Observasi dengan H_i sama dengan nol tidak diberi rasio nol untuk keperluan klasterisasi karena rasio tersebut tidak terdefinisi. Nilai rasio dibatasi pada rentang nol sampai seratus persen dalam implementasi. Pembatasan numerik ini merupakan pengamanan komputasi, sehingga tidak dapat menggantikan pemeriksaan anomali sumber data.

$$R_i = 100 \times \frac{P_i}{H_i}, \quad H_i > 0 \qquad (1)$$

2.3 Standardisasi dan estimasi GMM

Kedua fitur distandardisasi menggunakan rata-rata dan simpangan baku dari seluruh observasi layak September. Pada Persamaan (2), x_{ij} adalah nilai fitur j untuk pegawai i, μ_j adalah rata-rata fitur, dan σ_j adalah simpangan baku dengan pembagi n sebagaimana implementasi standardisasi yang digunakan. Nilai standar z_{ij} menjadi masukan pemodelan. Statistik profil kemudian dikembalikan ke satuan hari dan persen agar hasilnya dapat ditafsirkan secara operasional.

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \qquad (2)$$

Distribusi fitur standar dimodelkan sebagai campuran K komponen pada Persamaan (3). Bobot π_k bernilai nonnegatif dan seluruh bobot berjumlah satu. Vektor μ_k serta matriks Σ_k masing-masing menyatakan rata-rata dan kovarians komponen. Matriks kovarians menggunakan bentuk diagonal untuk membatasi jumlah parameter. Pilihan ini memungkinkan setiap fitur memiliki varians tersendiri dalam setiap komponen tanpa mengestimasi kovarians silang secara bebas.

$$f(z_i) = \sum_{k=1}^{K} \pi_k N(z_i \mid \mu_k, \Sigma_k) \qquad (3)$$

Probabilitas posterior keanggotaan dihitung sesuai Persamaan (4), kemudian label pegawai ditentukan oleh komponen dengan nilai posterior terbesar. Dalam tahap maksimisasi, bobot serta parameter komponen diperbarui berdasarkan pembobotan posterior. EM mengulangi kedua tahap sampai aturan penghentian

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

tercapai. Nomor segmen berfungsi sebagai penanda kelompok hasil, bukan urutan tingkat kinerja. Dalam penyajian akhir, segmen ditafsirkan berdasarkan profil fitur asli dan kebutuhan administratifnya.

$$\gamma_{ik} = \frac{\pi_k N(z_i \mid \mu_k, \Sigma_k)}{\sum_{l=1}^{K} \pi_l N(z_i \mid \mu_l, \Sigma_l)}, \quad c_i = arg\, max_k\, \gamma_{ik} \qquad (4)$$

Tabel 3. Konfigurasi pemodelan dan evaluasi

Komponen	**Pengaturan**
Periode dan unit analisis	September 2026; satu pegawai per bulan
Fitur masukan	Hari masuk tercatat dan rasio kelengkapan pulang
Standardisasi	Z-score pada observasi layak September
Kandidat dan model final	K=2, 3, 4, 5, 6, 7, 8; final K=2
Bentuk kovarians	Diagonal
Regularisasi kovarians	0,001
Jumlah inisialisasi	3
Batas iterasi	300
Pengaturan bilangan acak	42
Silhouette	Sampel 1.000; representasi awal setiap kelompok
DBI, CH, BIC dan AIC	Seluruh 6.464 observasi layak
Ambang pemeriksaan kelompok kecil	2% dari observasi layak; minimal 2 anggota

Sumber: Konfigurasi analisis yang terdokumentasi (2026).

Konfigurasi pada Tabel 3 mengikuti implementasi analisis yang terdokumentasi untuk hasil penelitian. Tiga inisialisasi dipakai untuk memilih solusi dengan kecocokan terbaik di antara percobaan dalam satu pemodelan. Pengaturan tersebut berbeda dari pengujian stabilitas melalui banyak pengulangan atau pengambilan sampel ulang. Penelitian tidak melaporkan kestabilan ARI, distribusi jumlah iterasi aktual, maupun hasil sensitivitas terhadap jenis kovarians karena keluaran pengujian tersebut tidak tersedia. Oleh sebab itu, reproduksibilitas konfigurasi tidak disamakan dengan bukti kestabilan empiris.

2.4 Evaluasi kandidat dan keputusan model final

Model dievaluasi untuk K=2 sampai K=8. Semua kandidat menggunakan fitur standar yang sama. Silhouette dihitung pada sampel sebanyak 1.000 observasi, dengan pemilihan awal hingga tiga anggota dari setiap kelompok dan pelengkapan sampel secara acak tanpa pengembalian. Bilangan acak dikendalikan dengan nilai 42. Karena pemilihan awal mengikuti label, susunan sampel dapat berbeda antar-K. DBI dan CH dihitung pada keseluruhan observasi layak, sedangkan BIC dan AIC dihitung dari model yang dilatih pada data yang sama.

Pada Persamaan (5), a(i) menyatakan rata-rata jarak observasi i terhadap anggota kelompoknya, dan b(i) merupakan rata-rata jarak terkecil terhadap kelompok lain. Rata-rata s(i) pada sampel evaluasi berukuran n_e=1.000 menghasilkan skor Silhouette yang dilaporkan. Pengambilan sampel mengurangi beban

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

perhitungan jarak, tetapi menambahkan ketidakpastian evaluasi. Selisih skor antar-K dibaca sebagai hasil prosedur yang digunakan, bukan sebagai perbedaan yang sudah diuji signifikansinya.

$$s(i) = \frac{b(i) - a(i)}{max\{a(i),b(i)\}}, \; S = \frac{1}{n_e}\sum_{i=1}^{n_e} s(i) \quad (5)$$

DBI dihitung melalui rata-rata kemiripan terbesar setiap kelompok terhadap kelompok lain, sebagaimana Persamaan (6). S_k adalah rata-rata jarak anggota kelompok k ke pusatnya, sedangkan M_{kl} adalah jarak antara pusat kelompok k dan l. CH pada Persamaan (7) membandingkan jejak matriks penyebaran antarkelompok B_k dan dalam kelompok W_k, dengan penyesuaian jumlah kelompok serta ukuran data n. Pemilihan arah evaluasi mengikuti nilai DBI yang lebih rendah dan CH yang lebih tinggi.

$$DBI = \frac{1}{K}\sum_{k=1}^{K} max_{l \neq k} \frac{S_k + S_l}{M_{kl}} \quad (6)$$

$$CH = \frac{tr(B_K)/(K-1)}{tr(W_K)/(n-K)} \quad (7)$$

BIC dan AIC dihitung menggunakan log-likelihood maksimum ℓ serta jumlah parameter bebas p, dengan rumus pada Persamaan (8). Untuk dua fitur dan kovarians diagonal, jumlah parameter adalah p=K(2d)+(K−1)=5K−1. Pada K=2 terdapat sembilan parameter bebas. Kedua kriteria menggunakan konvensi nilai lebih rendah lebih baik. Penyebutan konvensi ini diperlukan karena pelaporan BIC pada perangkat lunak lain dapat memakai tanda yang berbeda.

$$BIC = -2\ell + p\,ln(n); \quad AIC = -2\ell + 2p \quad (8)$$

Setiap kandidat diberi peringkat pada kelima indikator, kemudian rata-rata peringkat dihitung sesuai Persamaan (9). Silhouette dan CH diurutkan menurun, sedangkan DBI, BIC, serta AIC diurutkan menaik. Nilai rata-rata lebih rendah diutamakan. Metode ini memberi bobot sama kepada setiap indikator, tetapi indikator tersebut tidak independen. Tiga indikator struktur dan dua kriteria informasi juga tidak otomatis merepresentasikan keseimbangan kepentingan yang universal. Oleh karena itu, peringkat digunakan sebagai ringkasan keputusan yang dilengkapi pembacaan profil.

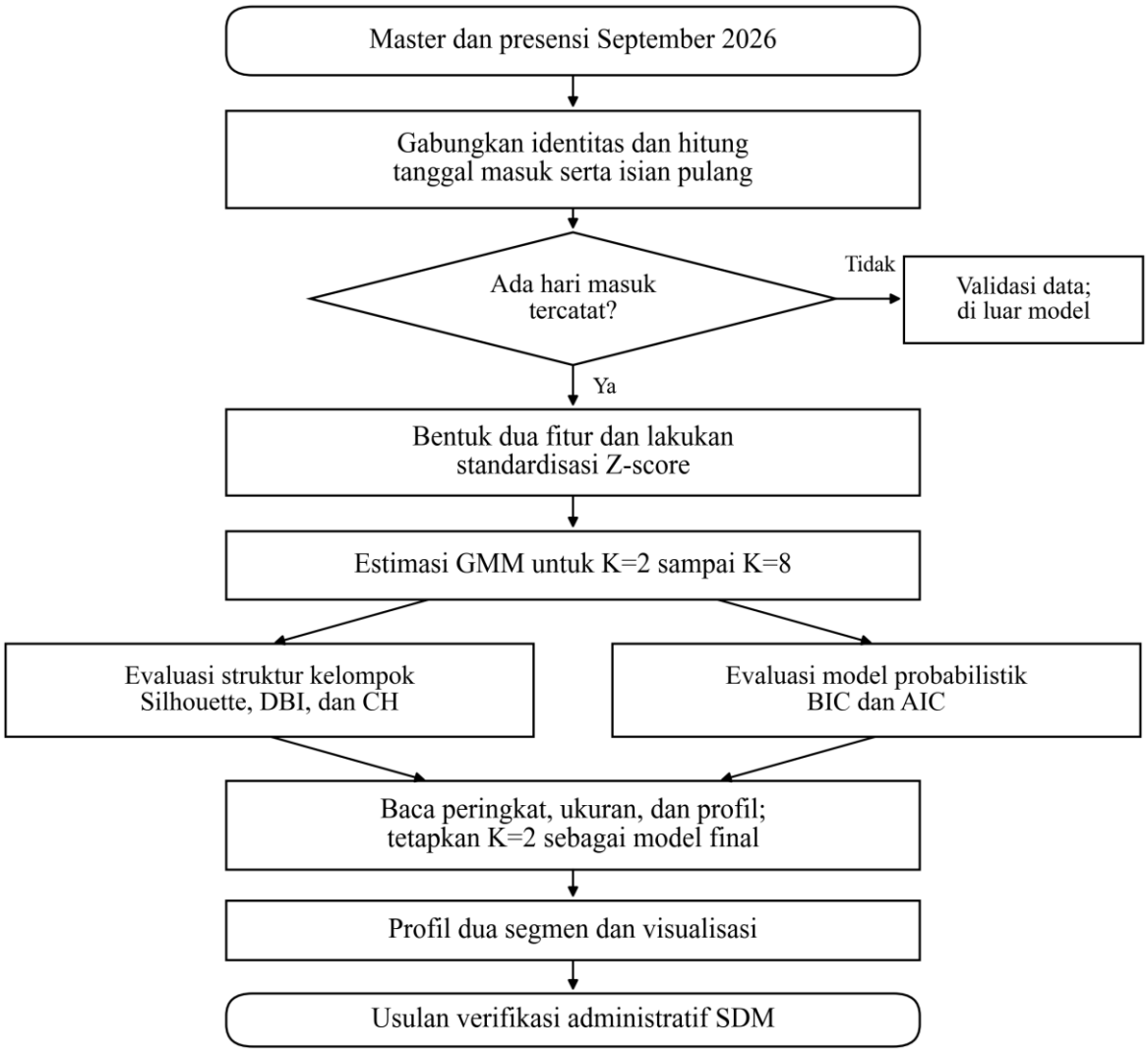
$$\bar{R}(K) = \frac{1}{5}\sum_{m=1}^{5} r_m(K) \quad (9)$$

Ukuran kelompok terkecil diperiksa agar penambahan komponen tidak hanya menghasilkan kelompok yang sangat kecil. Ambang pemeriksaan ukuran dalam sistem adalah dua persen dari observasi layak, dengan batas minimal dua anggota. Seluruh kandidat September menghasilkan ukuran minimum di atas ambang ini. Status valid pada tabel berarti partisi dapat dievaluasi menurut aturan sistem; status tersebut tidak berarti kandidat unggul pada seluruh indikator. K=2 ditetapkan sebagai model final dengan mempertimbangkan peringkat, pemisahan, ukuran, dan kejelasan profil administratif.

2.5 Alur penelitian dan penyajian hasil

Alur pada Gambar 1 menghubungkan penggabungan data, pemeriksaan kelayakan, pembentukan fitur, pemodelan, evaluasi, dan penyusunan rekomendasi. Percabangan kelayakan memisahkan observasi yang memerlukan validasi dari observasi pemodelan. Dua jalur evaluasi ditampilkan untuk membedakan indeks struktur kelompok dan kriteria informasi. Hasil keduanya dipertemukan pada keputusan K final sebelum profil dijelaskan dalam satuan asli.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



Gambar 1. Alur penelitian segmentasi pola pencatatan presensi

Sumber: Rancangan alur analisis penelitian (2026).

Grafik sebar dua indikator asli digunakan untuk membaca posisi kelompok dalam satuan yang dapat dikenali pengguna. Proyeksi *Principal Component Analysis* (PCA) dua dimensi digunakan sebagai tampilan tambahan atas dua fitur yang sama, bukan sebagai masukan baru untuk melatih GMM. Pergeseran kecil titik pada visualisasi hanya bertujuan mengurangi tumpang tindih tampilan. Angka evaluasi dan profil berasal dari nilai fitur sebelum pergeseran visual tersebut.

Pelaporan publik membatasi informasi pada hasil agregat, indikator evaluasi, profil dua segmen, dan usulan tindak lanjut. Nama perusahaan, nama unit, identitas pegawai, rincian roster, serta konfigurasi akses sistem tidak ditampilkan. Data individual dan pemetaan identitas tidak disertakan dalam artikel. Rekomendasi disusun sebagai usulan analitis; belum terdapat bukti bahwa rekomendasi tersebut telah divalidasi atau dilaksanakan oleh pengelola SDM.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

3. Hasil dan Pembahasan

3.1 Cakupan observasi dan kelayakan data

Penggabungan sumber bulanan menghasilkan 6.790 identitas dalam populasi operasional. Sebanyak 6.464 identitas, atau 95,20%, memenuhi aturan adanya hari masuk tercatat. Sebanyak 326 identitas, atau 4,80%, berada di luar pemodelan. Persentase tersebut dihitung terhadap populasi hasil penggabungan, bukan terhadap daftar pegawai aktif yang telah diaudit secara independen. Ringkasan ini menjaga agar pembaca tidak menganggap seluruh angka populasi mempunyai definisi administratif yang sama.

Tabel 4. Cakupan data September 2026

Cakupan	Jumlah	Persentase
Identitas hasil penggabungan sumber	6.790	100,00%
Observasi layak segmentasi	6.464	95,20%
Observasi di luar pemodelan	326	4,80%

Sumber: Hasil September 2026; persentase dihitung terhadap 6.790 identitas.

Observasi yang tidak masuk pemodelan tetap mempunyai arti operasional. Ketiadaan hari masuk dapat muncul karena status pada master belum diperbarui, ketidaktersediaan data pada sistem presensi, atau kondisi lain yang perlu ditelusuri. Penelitian tidak membagikan 326 identitas tersebut menurut penyebab karena rincian audit penyebab belum tersedia. Selisih jumlah sumber juga tidak digunakan untuk menyimpulkan absensi tanpa izin. Pemisahan ini menjaga agar ketidaktersediaan fitur tidak dipaksakan menjadi kelompok perilaku tertentu.

3.2 Perbandingan kandidat K=2 sampai K=8

Tabel 5 menyajikan seluruh hasil evaluasi September. K=2 memperoleh Silhouette tertinggi, yaitu 0,4799, DBI terendah, yaitu 0,9986, dan CH tertinggi, yaitu 5.211,91. Nilai tersebut menunjukkan bahwa pembagian dua kelompok mempunyai dukungan paling konsisten dari ketiga indeks struktur dalam rentang kandidat yang diuji. Nilai Silhouette 0,4799 tidak ditafsirkan sebagai akurasi 47,99% dan tidak digunakan untuk menyatakan bahwa pemisahan seluruh anggota sudah sempurna.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

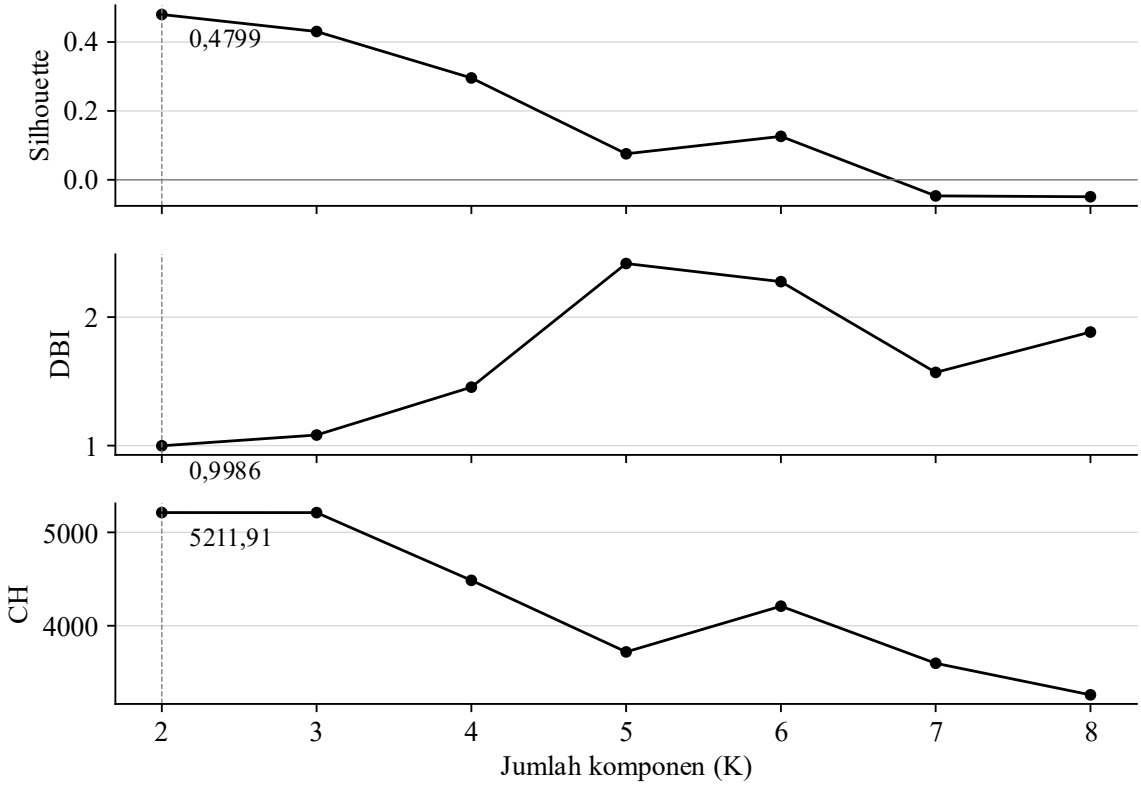
Tabel 5. Evaluasi seluruh kandidat GMM September 2026

K	Silhouette ↑	DBI ↓	CH ↑	BIC ↓	AIC ↓	Segmen terkecil
2	**0,4799**	**0,9986**	**5.211,91**	**20.611,99**	**20.551,02**	**2.083 (32,2%)**
3	0,4305	1,0824	5.211,45	17.012,77	16.917,93	494 (7,6%)
4	0,2957	1,4549	4.486,20	15.495,53	15.366,82	494 (7,6%)
5	0,0755	2,4170	3.720,06	11.667,43	11.504,85	494 (7,6%)
6	0,1261	2,2761	4.209,70	11.340,46	11.144,01	360 (5,6%)
7	−0,0467	1,5697	3.598,22	11.235,02	11.004,70	395 (6,1%)
8	−0,0492	1,8838	3.260,53	10.502,18	10.237,99	323 (5,0%)

Sumber: Hasil evaluasi September 2026. ↑ lebih tinggi; ↓ lebih rendah. Seluruh partisi valid menurut aturan sistem; K=2 model final.

K=3 merupakan kandidat terdekat dengan Silhouette 0,4305, DBI 1,0824, dan CH 5.211,45. Selisih CH terhadap K=2 hanya 0,46, sehingga keunggulan K=2 menurut CH sangat kecil secara numerik. Penetapan dua kelompok tidak bertumpu pada selisih tersebut saja. Penurunan Silhouette sebesar 0,0494 dan peningkatan DBI sebesar 0,0838 pada K=3 turut dipertimbangkan, bersama penyederhanaan interpretasi menjadi dua kebutuhan administratif utama.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



Gambar 2. Perbandingan indeks struktur kelompok pada K=2 sampai K=8

Sumber: Digambar ulang dari hasil evaluasi September 2026.

Pada K=4, Silhouette turun menjadi 0,2957 dan DBI naik menjadi 1,4549. K=5 menghasilkan Silhouette 0,0755 serta DBI tertinggi, yaitu 2,4170. K=6 menunjukkan perbaikan relatif terhadap K=5 pada beberapa indeks, tetapi masih berada di bawah K=2 menurut ketiga ukuran struktur. Dengan demikian, pola perubahan tidak sepenuhnya monoton. Penambahan komponen dapat mengubah pembagian lokal dan membentuk partisi baru yang kualitas jaraknya tidak selalu lebih baik daripada partisi sebelumnya.

K=7 dan K=8 mempunyai Silhouette negatif, masing-masing −0,0467 dan −0,0492. Secara rata-rata pada sampel evaluasi, sebagian hubungan jarak lebih sesuai dengan kedekatan terhadap kelompok lain daripada label yang diberikan. Hal ini menunjukkan lemahnya pemisahan menurut ukuran tersebut, bukan berarti semua anggota pasti ditempatkan keliru. Model tetap menghasilkan partisi yang dapat dievaluasi. Perbedaan antara validitas teknis partisi dan kualitas substantif kelompok menjadi penting ketika pengguna dashboard memilih kandidat secara manual.

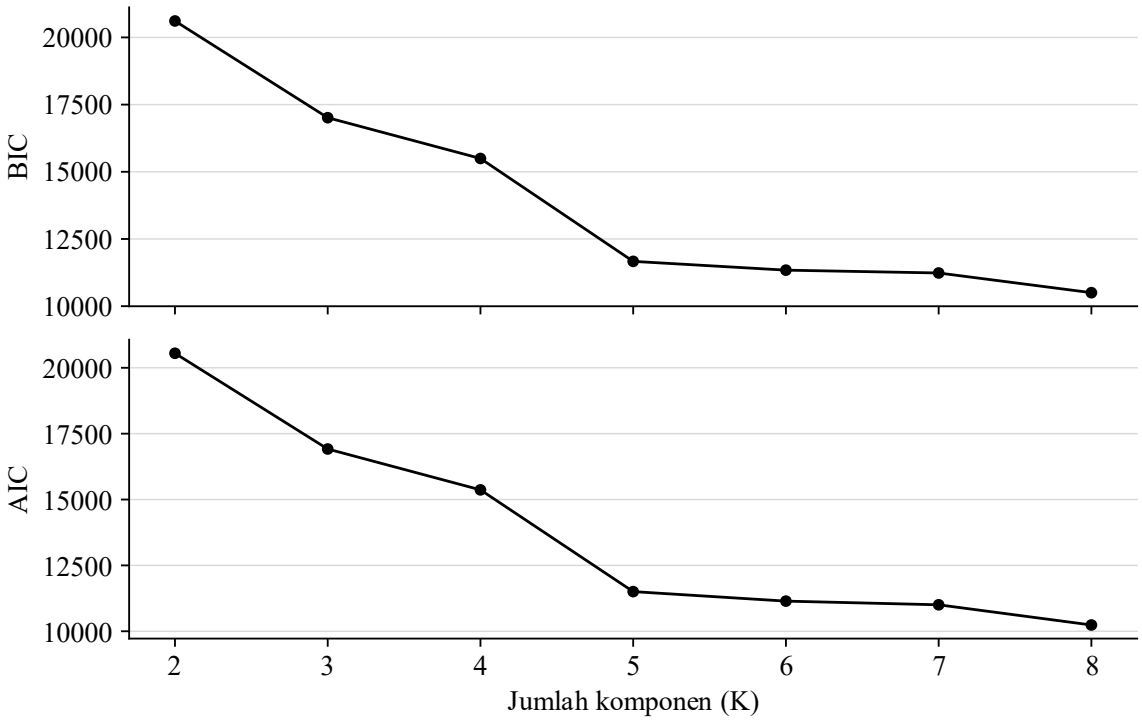
Ukuran segmen terkecil menurun dari 2.083 anggota pada K=2 menjadi 323 anggota pada K=8. Pada K=3, K=4, dan K=5, ukuran terkecil sama-sama 494 anggota. Kesamaan ukuran ini tidak membuktikan bahwa anggota kelompoknya identik karena tidak dilakukan perbandingan keanggotaan lintas-K. Angka ukuran minimum hanya dipakai untuk memahami ketidakseimbangan skala kelompok. Seluruh ukuran tersebut berada di atas ambang pemeriksaan dua persen, sehingga keputusan model tidak berasal dari pengguguran kandidat karena kelompok terlalu kecil.

3.3 Perbedaan BIC dan AIC dengan indeks struktur

BIC menurun dari 20.611,99 pada K=2 menjadi 10.502,18 pada K=8, sedangkan AIC menurun dari 20.551,02 menjadi 10.237,99. Dalam konvensi yang digunakan, K=8 lebih baik menurut kedua kriteria

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

informasi. Hasil tersebut menunjukkan bahwa penambahan komponen memberi keuntungan kecocokan yang masih melampaui penalti kompleksitas dalam rentang pengujian. Penelitian mengakui keuntungan ini dan tidak menyebut K=2 sebagai model yang optimal menurut BIC atau AIC.



Gambar 3. Perbandingan BIC dan AIC pada kandidat GMM

Sumber: Digambar ulang dari hasil evaluasi September 2026.

Secara metodologis, satu kelompok administratif dapat terdiri atas beberapa bagian distribusi yang memerlukan lebih dari satu komponen Gaussian untuk didekati. Penumpukan rasio pada nol dan seratus persen, sifat diskret jumlah hari, dan kovarians diagonal dapat menjadi alasan mengapa representasi kepadatan membutuhkan lebih banyak komponen. Penjelasan tersebut merupakan kemungkinan yang konsisten dengan sifat fitur, bukan hasil pengujian penyebab pada penelitian ini. Tidak dilakukan uji pembanding kovarians penuh atau model distribusi terbatas untuk memastikan sumber perbedaan.

Gambar 3 memperlihatkan penurunan kriteria informasi yang masih berlanjut hingga batas K terbesar yang diuji. Oleh karena itu, K=8 disebut kandidat dengan nilai terbaik di antara K=2 sampai K=8, bukan minimum global atas seluruh jumlah komponen yang mungkin. Perluasan rentang K dapat saja menghasilkan kriteria lebih rendah, tetapi belum tentu menghasilkan kelompok operasional yang lebih mudah dijelaskan. Batas pencarian model merupakan bagian dari desain penelitian dan harus dibedakan dari penemuan jumlah kelompok yang berlaku universal.

Penggunaan BIC dan AIC dalam penelitian ini sesuai dengan fungsi keduanya sebagai kriteria model probabilistik, sebagaimana juga digunakan pada penerapan GMM dalam segmentasi pendidikan tinggi (Karnaidi & Fudholi, 2026). Namun, tujuan penelitian presensi adalah merumuskan kelompok yang dapat diterjemahkan menjadi pemeriksaan administratif. Karena itu, kecocokan distribusi dibaca bersama pemisahan label dan profil. Keputusan akhir merupakan pertimbangan terhadap beberapa tujuan yang berbeda, bukan penolakan terhadap salah satu indikator karena menghasilkan rekomendasi yang tidak diharapkan.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

3.4 Konsensus peringkat dan alasan pemilihan K=2

Peringkat seluruh indikator ditampilkan pada Tabel 6. K=2 memperoleh peringkat satu untuk Silhouette, DBI, serta CH, dan peringkat tujuh untuk BIC serta AIC. Rata-rata peringkatnya adalah (1+1+1+7+7)/5=3,40. K=3 menghasilkan rata-rata 3,60, disusul K=4 sebesar 3,80. Selisih kecil antara dua kandidat teratas menunjukkan perlunya menjelaskan tujuan penggunaan, alih-alih menyajikan peringkat sebagai keputusan statistik yang tidak dapat diperdebatkan.

Tabel 6. Peringkat setiap indikator dan rata-rata peringkat

K	Silhouette	DBI	CH	BIC	AIC	Rata-rata
2	**1**	**1**	**1**	**7**	**7**	**3,40**
3	2	2	2	6	6	3,60
4	3	3	3	5	5	3,80
5	5	7	5	4	4	5,00
6	4	6	4	3	3	4,00
7	6	4	6	2	2	4,00
8	7	5	7	1	1	4,20

Sumber: Perhitungan peringkat dari Tabel 5. Peringkat 1 merupakan yang terbaik.

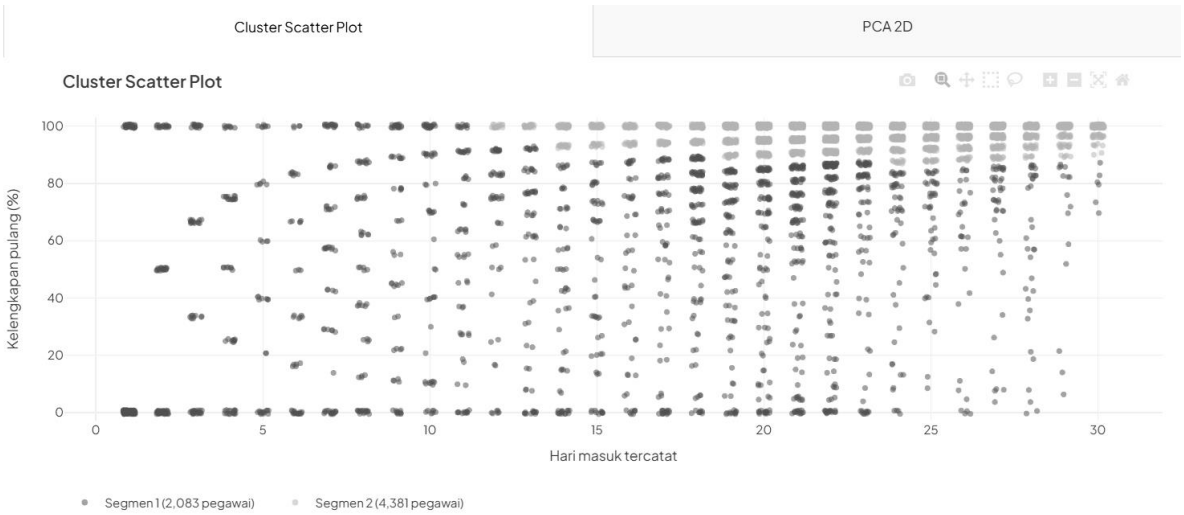
Konsensus peringkat membuat skala setiap indikator sebanding, tetapi menghilangkan informasi besar kecilnya selisih. Perbedaan CH yang sangat kecil tetap menghasilkan beda satu tingkat. BIC dan AIC juga berbagi dasar log-likelihood dan penalti, sehingga keduanya membawa informasi yang saling berkaitan. Indeks struktur pun tidak sepenuhnya independen. Dengan demikian, rata-rata peringkat diperlakukan sebagai aturan keputusan operasional, bukan statistik inferensial atau bukti bahwa satu model secara signifikan lebih baik.

K=2 dipertahankan karena memperoleh dukungan dari seluruh indeks struktur, membentuk dua kelompok dengan ukuran yang memadai, dan menghasilkan karakteristik administratif yang dapat dijelaskan. K=3 tetap merupakan alternatif yang masuk akal untuk dikaji jika organisasi membutuhkan rincian lebih jauh. Akan tetapi, penelitian ini tidak menyajikan profil K=3 sehingga tidak menyimpulkan manfaat tambahan dari pemecahan tersebut. Model final ditetapkan dalam ruang lingkup data, fitur, konfigurasi, dan tujuan penelitian September 2026.

3.5 Pola pada visualisasi fitur asli dan PCA

Grafik sebar pada Gambar 4 memperlihatkan dua pola yang berbeda dalam ruang hari masuk tercatat dan kelengkapan pulang. Segmen 2 banyak terlihat pada kelengkapan mendekati seratus persen dengan jumlah hari tercatat relatif tinggi. Segmen 1 tersebar lebih luas pada sumbu kelengkapan dan juga mencakup observasi dengan jumlah hari yang rendah. Sebagian anggota Segmen 1 tetap mempunyai rasio pulang tinggi, sehingga label segmen tidak dapat diganti dengan aturan satu ambang kelengkapan saja.

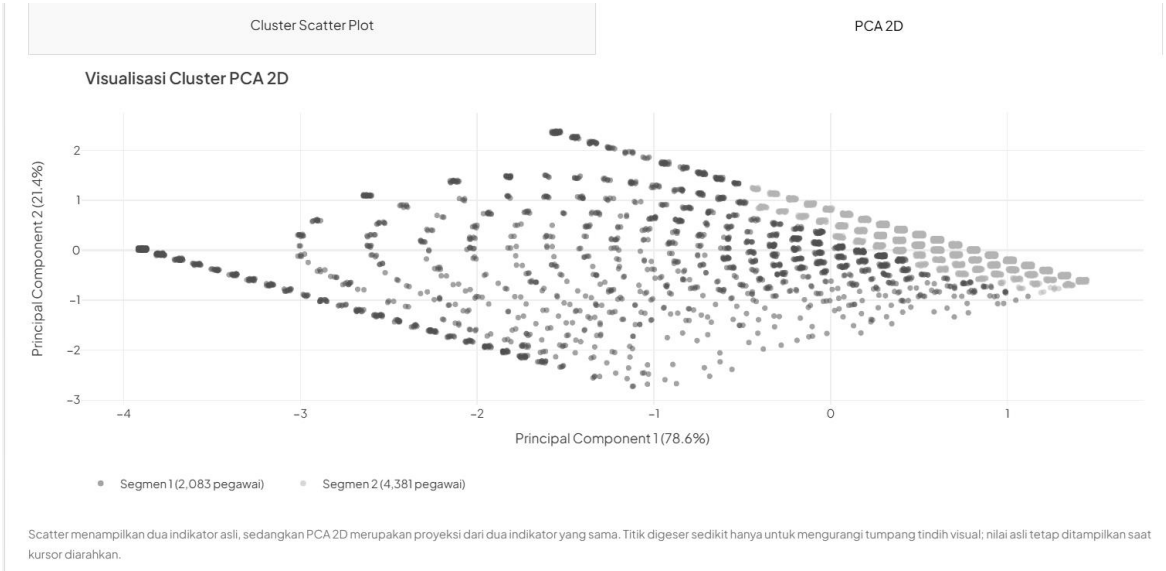
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



Gambar 4. Sebaran anggota pada dua indikator pencatatan asli

Sumber: Visualisasi asli hasil GMM September 2026. Versi cetak menggunakan skala abu-abu; titik gelap mewakili Segmen 1 dan titik terang Segmen 2.

Titik-titik pada grafik dapat berimpit karena jumlah hari merupakan bilangan bulat dan rasio dibentuk dari pasangan bilangan cacah. Pergeseran kecil pada tampilan membantu keterbacaan, tetapi kepadatan visual tetap tidak identik dengan jumlah pegawai yang tepat pada setiap koordinat. Oleh sebab itu, ukuran kelompok diambil dari tabel profil, bukan dihitung dari banyaknya titik yang tampak. Visualisasi dipakai untuk menjelaskan pola dan tumpang tindih, sedangkan simpulan kuantitatif bersumber dari hasil agregasi serta evaluasi.



Gambar 5. Proyeksi PCA dua dimensi dari fitur yang sama

Sumber: Visualisasi asli hasil GMM September 2026. Komponen pertama 78,6% dan kedua 21,4%; titik gelap Segmen 1 dan titik terang Segmen 2.

PCA pada Gambar 5 menunjukkan kontribusi variasi sebesar 78,6% pada komponen pertama dan 21,4% pada komponen kedua. Karena hanya terdapat dua fitur dan keduanya dipertahankan, proyeksi ini memuat seluruh variasi dalam ruang standar dua dimensi. Tampilan PCA tidak menambahkan bukti dari variabel

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

baru dan tidak merupakan pengujian validasi yang independen. Perubahan arah sumbu membantu melihat hubungan geometris yang sama, sedangkan arah positif atau negatif komponen tidak memiliki makna normatif mengenai baik buruknya pegawai.

3.6 Profil dua segmen pada skala asli

Profil K=2 ditampilkan pada Tabel 7. Segmen 1 terdiri atas 2.083 anggota atau 32,22% observasi layak, sedangkan Segmen 2 terdiri atas 4.381 anggota atau 67,78%. Kedua jumlah tepat membentuk keseluruhan 6.464 observasi pemodelan. Persentase dua desimal dihitung ulang dari jumlah anggota; tampilan dashboard sebelumnya membulatkannya menjadi 32,2% dan 67,8%. Persentase profil tidak menggunakan penyebut populasi hasil penggabungan karena cakupan kelompok hanya observasi yang tersegmentasi.

Tabel 7. Profil dua segmen GMM pada skala asli

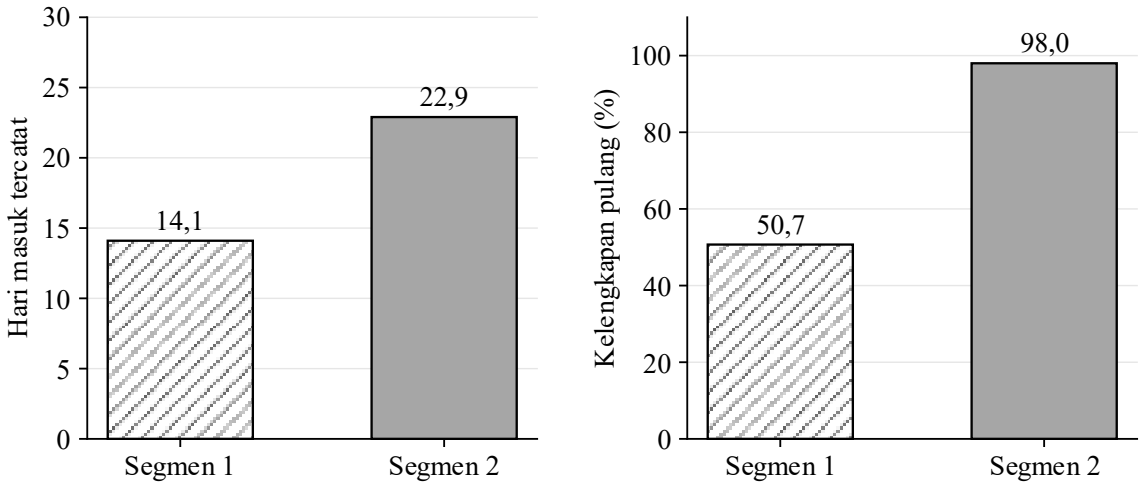
Indikator	**Segmen 1**	**Segmen 2**
Fokus administratif	Peningkatan konsistensi pencatatan pulang	Pemeliharaan kualitas pencatatan
Jumlah anggota	2.083	4.381
Proporsi observasi layak	32,22%	67,78%
Rata-rata hari masuk tercatat	14,1 hari	22,9 hari
Rata-rata kelengkapan pulang	50,7%	98,0%
Selisih dari kelengkapan penuh	49,3 poin persentase	2,0 poin persentase
Anggota dengan ≥1 hari belum lengkap	92,5%	33,9%
Rata-rata probabilitas maksimum	97,2%	97,0%

Sumber: Profil GMM September 2026. Proporsi anggota dihitung terhadap 6.464 observasi; rata-rata fitur mengikuti pembulatan keluaran sumber.

Segmen 1 mempunyai rata-rata 14,1 hari masuk tercatat dan kelengkapan pulang 50,7%. Selisih terhadap kelengkapan penuh adalah 49,3 poin persentase. Sebanyak 92,5% anggota memiliki setidaknya satu hari dengan catatan pulang belum lengkap menurut indikator sistem. Kombinasi ini mendukung penamaan peningkatan konsistensi pencatatan pulang. Istilah tersebut menunjukkan arah pemeriksaan data, bukan kesimpulan bahwa seluruh anggota mempunyai penyebab masalah yang sama atau tidak memenuhi kewajiban kerja.

Segmen 2 mempunyai rata-rata 22,9 hari masuk tercatat dan kelengkapan pulang 98,0%. Meski rata-ratanya mendekati penuh, 33,9% anggota masih mempunyai setidaknya satu hari dengan catatan pulang belum lengkap. Kedua angka tersebut tidak bertentangan: indikator pertama mengukur rata-rata tingkat kelengkapan, sedangkan indikator kedua mengukur jangkauan anggota yang pernah mengalami kekosongan. Kelompok dengan rata-rata tinggi tetap dapat membutuhkan audit berkala untuk mengidentifikasi masalah kecil yang tersebar pada banyak anggota.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



Gambar 6. Perbandingan rata-rata fitur pada dua segmen

Sumber: Digambar ulang dari profil September 2026; dua panel memakai satuan berbeda.

Selisih rata-rata antarkelompok adalah 8,8 hari masuk tercatat dan 47,3 poin persentase kelengkapan pulang. Angka ini merupakan perbedaan deskriptif, bukan ukuran pengaruh keanggotaan terhadap perilaku. Rata-rata fitur adalah hasil pembentukan kelompok sehingga perbedaannya tidak dapat diperlakukan sebagai hubungan kausal. Penelitian juga tidak melaporkan simpangan baku atau kuartil tiap segmen karena ringkasan tersebut tidak tersedia. Oleh sebab itu, rata-rata harus dibaca bersama grafik sebar yang memperlihatkan heterogenitas anggota.

Rata-rata probabilitas maksimum mencapai 97,2% pada Segmen 1 dan 97,0% pada Segmen 2. Nilai tinggi menunjukkan kecenderungan model memberi satu label yang dominan pada banyak anggota, tetapi tidak menghapus tumpang tindih geometris maupun keterbatasan fitur. Probabilitas tersebut bergantung pada spesifikasi distribusi serta parameter terlatih. Tanpa label pembanding dan pemeriksaan kalibrasi, tidak tepat menyebutnya sebagai akurasi 97% atau tingkat kepastian administratif bahwa rekomendasi untuk setiap pegawai sudah benar.

3.7 Implikasi administratif dan usulan tindak lanjut SDM

Perbedaan profil menunjukkan dua fokus pemeriksaan. Pada Segmen 1, pengelola dapat menelusuri tanggal pulang kosong, lalu mencocokkannya dengan roster, cuti, izin, penugasan, dan gangguan sistem. Prioritas ditentukan pula oleh keberulangan kejadian serta hasil verifikasi sumber, sehingga masalah pencatatan dapat dibedakan dari kondisi kerja yang sah.

Pada Segmen 2, fokusnya adalah pemeliharaan kualitas dan deteksi penyimpangan baru. Audit sampel perlu memeriksa kewajaran pasangan waktu serta kesesuaiannya dengan jadwal. Praktik administrasi yang berhasil dapat didokumentasikan. Audit tetap diperlukan karena rata-rata kelengkapan tinggi tidak menjamin mutu catatan seluruh anggota.

Tabel 8. Usulan tindak lanjut berdasarkan profil segmen

Segmen	Pemeriksaan awal	Tindak lanjut setelah verifikasi
1	Telusuri tanggal pulang kosong; cocokkan roster, cuti, izin, penugasan, dan gangguan sistem.	Perbaiki sumber atau sinkronisasi data; evaluasi prosedur pencatatan; pantau kejadian berulang.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Segmen	Pemeriksaan awal	Tindak lanjut setelah verifikasi
2	Audit sampel pasangan masuk–pulang; periksa pengecualian meskipun rata-rata tinggi.	Pertahankan prosedur yang sesuai; dokumentasikan praktik; pantau penurunan mutu.

Sumber: Usulan peneliti berdasarkan profil September 2026; belum merupakan kebijakan SDM yang tervalidasi.

Tabel 8 menyusun usulan tindakan yang mengikuti informasi teramati. Seluruh usulan memerlukan peninjauan pengelola SDM dan pemilik proses presensi. Penelitian tidak menetapkan sanksi, penilaian performa, atau keputusan kepegawaian otomatis. Pemisahan tersebut mempunyai konsekuensi praktis: dashboard dapat menjadi daftar awal pemeriksaan, tetapi keputusan akhir harus merujuk pada bukti sumber serta konteks kerja yang sah. Model dapat memperkecil ruang pencarian awal tanpa menggantikan kewenangan dan pertimbangan administratif.

3.8 Komunikasi hasil dan pemantauan periode berikutnya

Pemantauan berikutnya perlu mempertahankan definisi fitur dan mencatat perubahan sistem sumber. Jika metode atau aturan agregasi berubah, perubahan skor tidak dapat langsung dianggap sebagai perubahan perilaku pegawai. Demikian pula, pergantian anggota populasi dapat mengubah proporsi segmen tanpa adanya perubahan pada individu yang sama. Evaluasi lintas bulan perlu memisahkan perubahan komposisi populasi, perubahan mutu data, dan perpindahan label pada pegawai yang benar-benar dapat dicocokkan antarperiode.

Tabel 9. Rancangan indikator pemantauan lanjutan

Indikator	Dasar pembacaan	Kebutuhan validasi
Kelengkapan per pegawai	Perubahan rasio dengan definisi yang sama	Konsistensi aturan agregasi dan sumber
Jangkauan catatan belum lengkap	Proporsi anggota dengan setidaknya satu kejadian	Pemisahan kejadian berulang dan insidental
Observasi di luar pemodelan	Jumlah dan alasan tidak mempunyai fitur layak	Rekonsiliasi master dan data presensi
Penyelesaian koreksi	Kasus yang telah diperiksa dan diperbaiki	Bukti sebelum dan setelah perbaikan
Konsistensi segmentasi	Profil dan perpindahan anggota antarperiode	Populasi yang dapat dicocokkan dan pengujian stabilitas

Sumber: Rancangan evaluasi lanjutan oleh peneliti (2026).

Indikator pada Tabel 9 merupakan rancangan evaluasi lanjutan, bukan hasil sesudah intervensi. Pelaksanaannya dapat membantu menguji apakah segmentasi memberi nilai tambah bagi administrasi presensi. Target numerik tidak ditetapkan tanpa kesepakatan operasional karena kelayakannya bergantung pada roster, jenis pekerjaan, dan mutu integrasi sistem. Ukuran keberhasilan yang lebih kuat adalah perubahan kualitas catatan yang telah diverifikasi, bukan sekadar bertambahnya proporsi anggota dalam segmen dengan rata-rata tinggi.

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

3.9 Keterbatasan dan agenda pengembangan

Penelitian mencakup satu periode sehingga belum menjelaskan kestabilan struktur sepanjang waktu. Perubahan jadwal, komposisi tenaga kerja, dan sistem pencatatan dapat mengubah kelompok. Pengembangan berikutnya perlu memakai definisi fitur yang sama pada bulan lain, membandingkan profil, dan menelusuri penyebab perpindahan anggota.

Keterbatasan kedua berkaitan dengan cakupan fitur. Jumlah hari dan rasio pulang tidak mengukur durasi kerja, keterlambatan, produktivitas, atau kepatuhan terhadap roster. Rata-rata rasio memberi bobot sama kepada setiap pegawai; nilainya berbeda dari rasio total hari pulang terhadap total hari masuk segmen. Informasi jadwal dan penugasan diperlukan untuk mengukur pemenuhan kewajiban kerja.

Keterbatasan ketiga adalah pendekatan Gaussian diagonal terhadap fitur diskret dan berbatas. Penumpukan rasio ekstrem dapat mendorong penambahan komponen. Pengujian kovarians alternatif, transformasi, atau model campuran lain dapat memperjelas kecocokan distribusi. Penelitian ini belum membandingkan GMM dengan K-Means maupun metode hierarkis, sehingga tidak menyimpulkan keunggulan GMM atas metode tersebut.

Keterbatasan keempat adalah evaluasi internal tanpa pengujian stabilitas. Silhouette memakai satu prosedur sampel dan satu pengaturan acak, tanpa interval variasi dari pengulangan. Tiga inisialisasi dan rata-rata posterior bukan validasi eksternal. Selisih peringkat 0,20 antara K=2 dan K=3 masih memerlukan analisis sensitivitas serta pengambilan sampel ulang.

Keterbatasan kelima adalah belum adanya verifikasi lapangan dan pengukuran dampak rekomendasi. Label tidak menjelaskan penyebab catatan yang kosong. Penilaian manfaat segmentasi memerlukan peninjauan SDM, dokumentasi tindakan, dan pengamatan setelah tindak lanjut. Artikel ini membatasi manfaatnya pada penyusunan prioritas pemeriksaan berdasarkan profil data.

4. Kesimpulan

Penelitian ini menghasilkan segmentasi pola pencatatan presensi September 2026 menggunakan GMM pada dua fitur, yaitu jumlah hari masuk tercatat dan kelengkapan catatan pulang. Dari 6.790 identitas dalam populasi hasil penggabungan sumber, 6.464 memenuhi aturan kelayakan pemodelan. Evaluasi tujuh kandidat menunjukkan K=2 mempunyai Silhouette 0,4799, DBI 0,9986, CH 5.211,91, dan rata-rata peringkat 3,40. Model tersebut dipilih sebagai segmentasi akhir untuk tujuan pemeriksaan administratif.

Pemilihan K=2 tidak berarti bahwa seluruh indikator memberikan dukungan yang sama. K=8 menghasilkan BIC dan AIC terendah dalam rentang pengujian, sedangkan K=2 lebih baik menurut ketiga indeks struktur. Perbedaan ini menunjukkan perlunya membedakan kecocokan model probabilistik dari keterpisahan serta kegunaan kelompok. Keputusan penelitian mempertimbangkan pemisahan, ukuran kelompok, dan kejelasan profil, dengan tetap mengakui bahwa K=3 merupakan kandidat terdekat dan hasil pemilihan belum diuji kestabilannya lintas periode.

Segmen 1 mencakup 2.083 anggota dengan rata-rata 14,1 hari masuk dan kelengkapan pulang 50,7%, sehingga diarahkan pada pemeriksaan dan peningkatan konsistensi pencatatan. Segmen 2 mencakup 4.381 anggota dengan rata-rata 22,9 hari masuk dan kelengkapan 98,0%, sehingga diarahkan pada pemeliharaan mutu serta audit berkala. Perbedaan tersebut merupakan karakteristik administratif agregat. Hasil segmentasi tidak membuktikan kedisiplinan, produktivitas, atau pelanggaran individu dan tidak menjadi dasar keputusan kepegawaian otomatis.

Pengembangan penelitian disarankan melalui validasi pasangan catatan terhadap jadwal, penelusuran data tidak lengkap, evaluasi periode lain, dan analisis sensitivitas model. Tindak lanjut perlu ditinjau SDM serta

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

diukur hasilnya. GMM dengan demikian berfungsi sebagai alat pemetaan variasi pencatatan dan penyusunan pemeriksaan administratif.

Daftar Pustaka

Alfian, W., Kusrini, & Hidayat, T. (2024). Analisis clustering pegawai berdasarkan tingkat kedisiplinan menggunakan algoritma K-Means dan Davies-Bouldin Index. *Journal of Electrical Engineering and Computer (JEECOM), 6*(2), 437–448. https://doi.org/10.33650/jeecom.v6i2.9556

Alyarahma, N. N., Kholijah, G., & Sormin, C. (2024). Pengelompokan provinsi di Indonesia menggunakan Gaussian Mixture Model berdasarkan indikator kemiskinan. *Journal of Mathematics: Theory and Applications, 6*(2), 158–167. https://doi.org/10.31605/jomta.v6i2.4032

Darma, S., Yusman, Y., & Hendrawan, J. (2024). Analisis data tingkat kehadiran pegawai dengan menggunakan clustering K-Means pada Dinas Pekerjaan Umum dan Penataan Ruang Kabupaten Langkat. *Jurnal Minfo Polgan, 13*(1), 1105–1116. https://doi.org/10.33395/jmp.v13i1.13958

Diana, P., Yusup, R., & Pebrianti, D. (2026). Analisis segmentasi konsumen e-commerce menggunakan model RFM melalui metode K-Means, DBSCAN, dan Agglomerative Clustering. *JEPIN (Jurnal Edukasi dan Penelitian Informatika), 12*(2), 111–124. https://jurnal.untan.ac.id/index.php/jepin/article/view/111025

Karnaidi, A., & Fudholi, D. H. (2026). Optimasi strategi pemasaran perguruan tinggi berbasis segmentasi data menggunakan HDBSCAN dan GMM. *SAINSTEK, 14*(1), 168–179. https://ejournal.sttp-yds.ac.id/index.php/js/article/view/426

Mira, M., Nurcahyo, A. C., Gudiato, C., & Kusnanto, K. (2024). Implementasi data mining metode K-Means menggunakan framework Python dalam mengelompokkan pegawai berdasarkan data presensi. *Jurnal Media Informatika Budidarma, 8*(3), 1334–1343. https://doi.org/10.30865/mib.v8i3.7716

Ramadhan, B. F., Abdillah, B. M., Hidayatullah, M., Faizal, M. N., Winata, R. A., & Alfian, Z. (2025). Analisis performa clustering: K-Means dan similarity matrix dalam evaluasi Silhouette, DBI, CHI, dan Dunn Index. *Jurnal Pendidikan Tambusai, 9*(2), 22961–22967. https://jptam.org/index.php/jptam/article/view/30265

Sumantri, A. D., Binabar, S. W., & Amalia, N. (2026). Penerapan algoritma K-Means clustering untuk klasterisasi pola kehadiran pegawai. *IJAI (Indonesian Journal of Applied Informatics), 10*(1), 167–177. https://doi.org/10.20961/ijai.v10i1.110323

Wahidah, Z., & Utari, D. T. (2023). Comparison of K-Means and Gaussian mixture model in profiling areas by poverty indicators. *BAREKENG: Journal of Mathematics and Its Applications, 17*(2), 717–726. https://doi.org/10.30598/barekengvol17iss2pp0717-0726

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)