

## IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK ANALISIS POLA ABSENSI KARYAWAN PT X

### IMPLEMENTATION OF K-MEANS CLUSTERING ALGORITHM FOR ANALYZING EMPLOYEE ATTENDANCE PATTERNS AT PT X

Jessica Widodo<sup>1</sup>, Jap Tji Beng<sup>1,2\*</sup>, Desi Arisandi<sup>1</sup>, Sri Tiatri<sup>2</sup>

Fakultas Teknologi Informasi, Program Studi Sarjana Sistem Informasi, Universitas Tarumanagara<sup>1</sup>  
Laboratorium Kognisi Edukasi dan Inovasi Teknologi, Fakultas Psikologi, Universitas Tarumanagara<sup>2</sup>  
[t.jap@untar.ac.id](mailto:t.jap@untar.ac.id)\*

#### ABSTRACT

*The operational success of PT X, a manufacturing company, is largely determined by the consistent attendance of employees involved in the production process. Therefore, attendance is a crucial indicator for human resource management to maintain smooth operations. This study aims to analyze employee attendance patterns at PT X over a one-year period to support data-driven decision-making. Daily attendance data was processed using Feature Engineering techniques to generate behavioral variables, such as attendance, tardiness, average minutes of delay, early departure, and variation in arrival times. The method used was the clustering method using the K-Means algorithm, with the optimal number of clusters determined using the Elbow Method and clustering quality evaluated using the Silhouette Score. The three main patterns identified were highly disciplined employees, moderately disciplined employees, and employees with low discipline. The results revealed three optimal clusters with a Silhouette Score of 0.87, indicating good clustering. These results support the use of Business Intelligence for strategic decision-making in human resource management at PT X.*

**Keywords:** Employee Attendance, Elbow Method, Feature Engineering, K-Means Clustering, Silhouette Score.

#### ABSTRAK

Keberhasilan operasional PT X, sebuah perusahaan manufaktur, sangat ditentukan oleh konsistensi kehadiran karyawan yang terlibat dalam proses produksi. Oleh karena itu, absensi menjadi indikator penting bagi manajemen sumber daya manusia untuk menjaga kelancaran operasional. Penelitian ini bertujuan untuk menganalisis pola absensi karyawan PT X selama satu tahun guna mendukung pengambilan keputusan berbasis data. Data absensi harian diolah menggunakan teknik *Feature Engineering* untuk menghasilkan variabel perilaku absensi, seperti jumlah kehadiran, keterlambatan, rata-rata menit keterlambatan, pulang cepat, dan variasi jam masuk. Metode yang digunakan adalah metode *clustering* dengan algoritma *K-Means*, dengan jumlah *cluster* optimal ditentukan melalui *Elbow Method* dan kualitas *clustering* dievaluasi menggunakan *Silhouette Score*. Tiga pola utama yang teridentifikasi adalah karyawan yang sangat disiplin, disiplin moderat, dan disiplin rendah. Hasil menunjukkan tiga *cluster* optimal dengan *Silhouette Score* 0,87, menunjukkan pengelompokan yang baik. Hasil penelitian ini mendukung pemanfaatan *Business Intelligence* untuk pengambilan keputusan strategis dalam manajemen sumber daya manusia di PT X.

**Kata Kunci:** Absensi karyawan, Elbow Method, Feature Engineering, K-Means Clustering, Silhouette Score.

#### PENDAHULUAN

Kedisiplinan kehadiran karyawan merupakan salah satu faktor penting yang memengaruhi produktivitas di perusahaan dan membentuk kinerja yang optimal (Gregorio et al., 2023). Karyawan yang hadir tepat waktu dan konsisten dalam menjalankan jam kerja tidak hanya memastikan kelancaran operasional sehari-hari, tetapi juga menciptakan budaya kerja yang profesional dan bertanggung jawab (Wulandari & Lailiyah, 2022).

Sebaliknya, ketidakhadiran atau keterlambatan yang sering dapat mengganggu alur kerja tim, menurunkan semangat kerja, serta mempengaruhi pencapaian target perusahaan (Simarmata et al., 2025). Penelitian empiris menunjukkan bahwa disiplin kerja dan absensi berpengaruh signifikan terhadap produktivitas karyawan (Safira, 2024).

Seiring dengan berkembangnya pendekatan *Business Intelligence*, yaitu *data-driven decision making*, data kini

tidak hanya dipandang sebagai hasil sampingan dari proses operasional, tetapi juga sebagai aset strategis yang dapat dimanfaatkan untuk menghasilkan wawasan yang mendalam. Implementasi sistem informasi yang terintegrasi dapat memperbaiki proses analisis sekaligus meningkatkan efisiensi dan membawa pengaruh yang signifikan (Yehezkiel et al., 2025; Wijaya et al., 2024). Melalui teknik *machine learning* dan kecerdasan buatan, data dapat diolah untuk menemukan pola tersembunyi, membuat prediksi, serta mendukung proses pengambilan keputusan secara lebih cepat, akurat, dan berbasis bukti (Alpaydin, 2020). Pendekatan ini memungkinkan organisasi untuk meningkatkan efisiensi operasional, mengoptimalkan strategi bisnis, dan meminimalkan risiko (Ardiansyah et al., 2024). Hal ini juga berlaku bagi data absensi yang sebelumnya hanya digunakan untuk keperluan administrasi rutin. Kini, data tersebut dapat dimanfaatkan untuk analisis perilaku karyawan melalui teknik *machine learning* yang mampu mengidentifikasi pola absensi dan mendeteksi indikasi masalah kedisiplinan secara lebih sistematis.

Dalam studi *machine learning*, dikenal tiga pendekatan utama, yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning* (Alpaydin, 2020). Setiap pendekatan memiliki fokus yang berbeda di mana *supervised learning* berfokus pada pembuatan prediksi dari data berlabel, *unsupervised learning* untuk mengidentifikasi pola tersembunyi pada data yang tidak berlabel, dan *reinforcement learning* untuk melatih agen membuat keputusan optimal melalui sistem *reward* dan *penalty*. Untuk tujuan analisis pola pada data historis dalam pengambilan keputusan bisnis, perhatian utama seringkali diberikan pada *supervised learning* dan *unsupervised learning* (Alloghani et al., 2020). Penerapan algoritma *classification* yang merupakan bagian dari *supervised learning*, telah

menunjukkan efektivitasnya dalam pengambilan keputusan berbasis data.

Contohnya, pada studi Barito et al. (2022), algoritma C4.5 digunakan untuk mengklasifikasikan mahasiswa penerima bantuan sosial COVID-19, menghasilkan model yang sangat akurat dalam memisahkan kategori penerima bantuan. Sementara itu, studi Kenny et al. (2024) menggunakan algoritma Naïve Bayes untuk evaluasi penilaian kinerja karyawan, berhasil membangun sistem prediksi yang efektif untuk mendukung keputusan manajemen sumber daya manusia. Di sisi lain, algoritma *clustering*, yang merupakan bagian dari *unsupervised learning*, terbukti mampu menemukan pola tersembunyi pada data yang kompleks. Sebagai contoh, studi Lusiana et al. (2020) menerapkan *Hierarchical Clustering* untuk mengelompokkan objek wisata menggunakan foto *geotagged*, mengidentifikasi kluster-kluster dengan karakteristik yang unik dan memberikan wawasan baru bagi pengembangan pariwisata daerah Bantul dan Sleman. Hierarchical Clustering merupakan teknik pengelompokan yang bertujuan menyusun data ke dalam struktur hirarki (Pratama et al., 2020). Selain itu, studi Andrian et al. (2024) menggunakan metode K-Means dan Fuzzy C-Means untuk *clustering* data meteorologi di Indonesia Timur, berhasil mengelompokkan wilayah berdasarkan pola cuaca yang serupa yang bermanfaat untuk analisis iklim. Efektivitas metode *classification* dan *clustering* pada berbagai studi memberikan justifikasi kuat untuk penerapannya dalam analisis pola perilaku tenaga kerja. Dalam konteks analisis pola absensi, pendekatan *clustering* menjadi krusial karena kemampuannya mengidentifikasi dan mengelompokkan pola perilaku karyawan yang tersembunyi, sehingga manajemen sumber daya manusia dapat memperoleh wawasan komprehensif, mendukung pengambilan keputusan proaktif yang berbasis data, serta meningkatkan produktivitas dan kesejahteraan.

PT X, sebuah perusahaan manufaktur, sangat bergantung pada konsistensi kehadiran karyawan dalam operasionalnya. Perusahaan ini memiliki 24 orang karyawan pada divisi produksi. Meskipun data absensi harian telah tercatat secara rutin melalui sistem informasi, pemanfaatan data tersebut masih terbatas pada fungsi administrasi, seperti perhitungan gaji dan rekapitulasi bulanan. Akibatnya, perusahaan belum memiliki mekanisme analisis yang mendalam untuk memahami pola absensi karyawan secara keseluruhan yang terakumulasi. Padahal, pola absensi yang kurang baik, seperti misalnya sering terlambat atau sering tidak hadir, dapat berdampak signifikan pada produktivitas, pembagian beban kerja, hingga penurunan kualitas layanan. Kondisi inilah yang membuka peluang untuk memanfaatkan teknik analisis data guna menemukan pola absensi tersembunyi, yang dapat mendukung pengambilan keputusan manajemen sumber daya manusia secara proaktif.

Mengingat kebutuhan untuk mengidentifikasi pola absensi tersembunyi tanpa label kategori yang telah ditentukan sebelumnya, pendekatan *unsupervised learning*, khususnya metode *clustering*, menjadi pilihan metodologi yang paling tepat. Algoritma K-Means dipilih karena dikenal efisien, mudah diinterpretasikan, dan merupakan metode paling umum digunakan untuk analisis pola data dalam konteks bisnis. K-Means bekerja dengan mengelompokkan data ke dalam  $k$  cluster, di mana setiap cluster berisi data dengan karakteristik serupa. Tujuan utama K-Means adalah meminimalkan jarak antara titik data dengan pusat cluster (*centroid*), sehingga pola tersembunyi dalam data dapat terungkap (Syahkur et al., 2024).

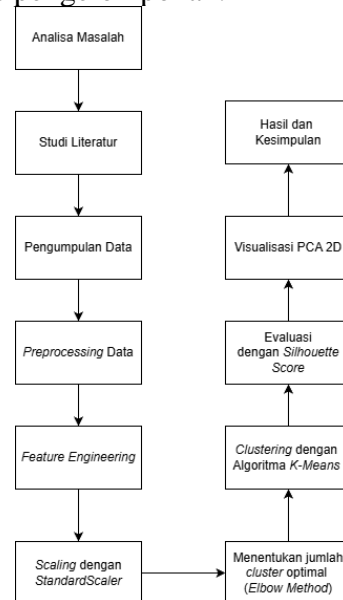
Pemanfaatan K-Means sangat luas, khususnya dalam manajemen sumber daya manusia. Hal ini ditunjukkan oleh studi Muharmi & Nadriati (2022), yang menganalisis disiplin karyawan berdasarkan data absensi digital dan berhasil mengidentifikasi tiga kelompok

disiplin (sangat baik, baik, dan kurang baik) yang efektif untuk perencanaan intervensi SDM. Selain itu, algoritma ini juga diterapkan dalam studi mengenai analisis pola transaksi, serupa dengan studi Noval et al. (2025) yang mengimplementasikan K-Means untuk mengelompokkan data penjualan guna mengungkap pola beli yang efektif di Toko Monisa.

Oleh karena itu, tujuan utama dari penelitian ini adalah menerapkan algoritma K-Means *Clustering* untuk menganalisis dan menemukan pola absensi yang berbeda-beda di kalangan 24 karyawan divisi produksi PT X, guna memberikan wawasan yang dibutuhkan manajemen sumber daya manusia untuk pengambilan keputusan berbasis data.

## METODE

Penelitian ini menggunakan metode *clustering* dengan algoritma K-Means untuk mengelompokkan karyawan berdasarkan pola absensi, *Elbow Method* untuk menentukan jumlah cluster optimal, dan *Silhouette Score* untuk mengevaluasi kualitas pengelompokan.



**Gambar 1. Alur Penelitian**

Sumber: dokumentasi pribadi

Pada Gambar 1 terlihat alur penelitian ini yang dimulai dengan analisis masalah, studi literatur, pengumpulan data, *preprocessing data*, *feature engineering*,

*scaling* dengan *StandardScaler*, menentukan jumlah *cluster* optimal dengan *Elbow Method*, *clustering* dengan algoritma *K-Means*, evaluasi dengan *Silhouette Score*, visualisasi PCA 2D (*Principal Component Analysis 2 Dimension*), hingga mencapai hasil dan kesimpulan.

### Pengumpulan Data

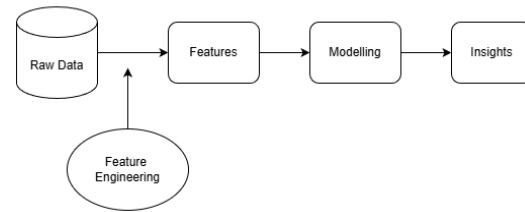
Data yang digunakan untuk penelitian ini diperoleh dari sistem *database* yang terdapat di perusahaan PT X. Penelitian ini memfokuskan populasi pada divisi produksi. Teknik pengambilan sampel yang digunakan adalah sensus atau *total population sampling*, di mana seluruh populasi terfokus yang berjumlah 24 orang karyawan dijadikan sebagai sampel. Instrumen utama penelitian adalah data absensi harian karyawan yang bersumber langsung dari sistem *database*. Dataset mencakup data absensi selama periode satu tahun, yaitu dari bulan Juli 2024 hingga Juli 2025. Data absensi yang diperoleh mencakup *id* karyawan, tanggal absensi, waktu *clock in*, dan waktu *clock out*. Dalam periode satu tahun terdapat 4.720 baris data absensi.

### Preprocessing Data

*Preprocessing* merupakan tahap persiapan data mentah agar siap untuk dianalisis. Pada penelitian ini, langkah-langkah *preprocessing* dimulai dengan *parsing* tanggal dan waktu menjadi tipe data *datetime*, sehingga perhitungan durasi kerja dan keterlambatan dapat dilakukan secara akurat. Selanjutnya, ditentukan jam kerja normal untuk menghitung keterlambatan dan pulang cepat karyawan. Nilai negatif pada keterlambatan atau pulang cepat diabaikan agar hanya tercatat nilai yang relevan. Selain itu, dilakukan pemeriksaan untuk memastikan tidak terdapat data yang hilang yang dapat memengaruhi hasil analisis.

### Feature Engineering dan Scaling dengan *StandardScaler*

*Feature engineering* adalah proses transformasi data mentah menjadi fitur-fitur yang lebih informatif dan relevan untuk digunakan dalam pemodelan (Dong & Liu, 2018).



**Gambar 2. Alur Feature Engineering**

Sumber: dokumentasi pribadi

*Feature engineering* berperan sebagai jembatan antara data mentah dan proses pemodelan, karena kualitas dan relevansi fitur yang dihasilkan akan sangat memengaruhi kinerja algoritma dan ketepatan wawasan yang diperoleh dari analisis. Pada **Gambar 2** terlihat *feature engineering* terletak di antara *raw data* dan *feature*, menyiapkan data mentah menjadi bentuk yang siap digunakan untuk pemodelan sehingga *insight* yang dihasilkan lebih akurat dan bermakna. Karena setiap fitur memiliki skala yang berbeda, maka dilakukan standarisasi menggunakan *StandardScaler*. *Scaling* memastikan setiap fitur memiliki rata-rata 0 dan standar deviasi 1, sehingga tidak ada fitur yang mendominasi proses *clustering*.

### Elbow Method

*Elbow Method* digunakan untuk menentukan jumlah *cluster* optimal (*k*) dalam algoritma *K-Means* (Maori & Evanita, 2023). Prinsipnya adalah mencoba beberapa nilai *k* dan mengevaluasi kualitas pengelompokan menggunakan WCSS (*Within Cluster Sum of Squares*), yaitu jumlah kuadrat jarak titik data terhadap *centroid* masing-masing *cluster* (Syahkur et al., 2024). Nilai WCSS yang kecil menunjukkan *cluster* yang rapat dan homogen. Dengan memplot WCSS terhadap jumlah *cluster*, titik *elbow* yang menunjukkan perlambatan penurunan WCSS dapat dijadikan jumlah *cluster* optimal (Maori & Evanita, 2023). Metode ini membantu menyeimbangkan kerapatan

*cluster* dan kompleksitas model (Syahkur et al., 2024).

### K-Means Clustering

*K-Means* adalah algoritma *clustering* yang membagi data menjadi *k cluster* berdasarkan kedekatan antar titik (Ikotun et al., 2023). Algoritma ini bekerja dengan menempatkan *k centroid* secara acak, kemudian menugaskan setiap titik data ke *centroid* terdekat, dan memperbarui posisi *centroid* berdasarkan rata-rata titik di tiap *cluster*. Proses ini diulang sampai *centroid* stabil atau perubahan sangat kecil. *K-Means* sangat efektif untuk menemukan pola kelompok dalam dataset besar dan numerik (Neco et al., 2025).

### Silhouette Score

*Silhouette Score* adalah metrik untuk mengevaluasi kualitas clustering dengan mengukur seberapa mirip tiap titik data dengan *cluster* sendiri dibandingkan dengan *cluster* lain. Skor berkisar dari -1 hingga 1, di mana nilai mendekati 1 menunjukkan titik data sangat sesuai dengan *cluster*, nilai mendekati 0 menunjukkan titik berada di batas antar *cluster*, dan nilai negatif menunjukkan kemungkinan salah *cluster* (Utami et al., 2023).

### PCA 2D

PCA 2D adalah teknik reduksi dimensi yang merangkum informasi dari variabel-variabel asli menjadi beberapa komponen utama yang saling ortogonal, sambil mempertahankan sebagian besar variasi data (Rosyada & Utari, 2024). Dalam konteks *clustering*, PCA 2D mengekstraksi dua komponen utama sehingga data dapat divisualisasikan dalam scatter plot dua dimensi, memudahkan pengamatan distribusi cluster, pola, dan potensi *outlier* (Garg & Dwivedi, 2024).

### HASIL DAN PEMBAHASAN

Dalam penelitian ini, *feature engineering* dilakukan pada dua level, yaitu harian dan per karyawan.

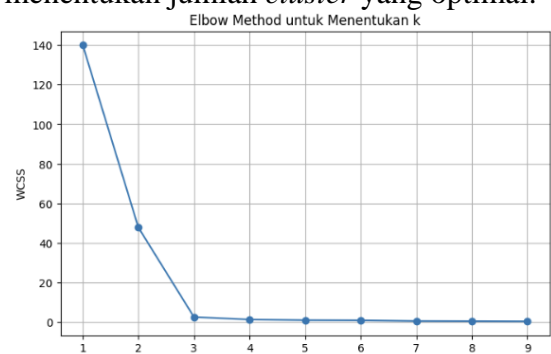
**Tabel 1. Variabel Feature Engineering**

Level	Fitur	Deskripsi
-------	-------	-----------

Harian	terlambat_hadir	Menit keterlambatan masuk
Harian	pulang_cepat_minit	Menit pulang lebih awal
Harian	durasi_jam	Lama jam kerja hari tersebut
Harian	jam_masuk_decimal	Jam masuk dalam format desimal
Harian	jam_pulang_decimal	Jam pulang dalam format desimal
Per Karyawan	total_hadir	Total hadir selama setahun
Per Karyawan	total_terlambat	Total hari terlambat
Per Karyawan	rata2_terlambat	Rata-rata keterlambatan per hari
Per Karyawan	total_pulang cepat	Total hari pulang lebih awal
Per Karyawan	rata2_durasi	Rata-rata lama jam kerja per hari
Per Karyawan	std_jam_masuk	Variabilitas jam masuk (standar deviasi)
Per Karyawan	std_jam_pulang	Variabilitas jam pulang (standar deviasi)
Per Karyawan	total_tidak_hadir	Total hari tidak hadir

Sumber : dokumentasi pribadi

Pada level harian, fitur dibuat dari data absensi tiap hari, sedangkan pada level per karyawan, data harian diolah menjadi ringkasan statistik tahunan. Rincian fitur dapat dilihat pada Tabel 1. Kemudian dilakukan scaling dengan *StandardScaler* untuk memastikan dataset siap diolah lebih lanjut dengan *Elbow Method* untuk menentukan jumlah *cluster* yang optimal.

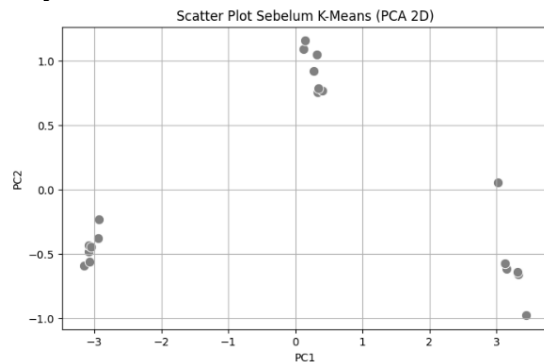


**Gambar 3. Elbow Plot**

Sumber : dokumentasi pribadi

Berdasarkan hasil visualisasi pada Gambar 3, kurva WCSS menunjukkan penurunan yang sangat signifikan hingga  $k = 3$ . Penurunan drastis ini menunjukkan bahwa pembentukan tiga kelompok *cluster* pola absensi efektif dalam memisahkan varian perilaku absensi karyawan. *Elbow point* yang paling jelas terlihat berada pada  $k = 3$ . Setelah titik ini, penambahan jumlah

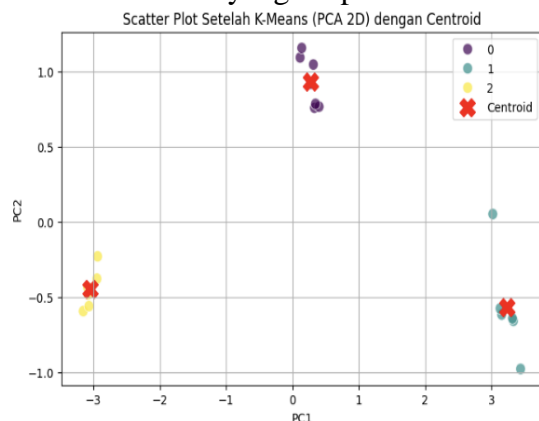
*cluster* hanya menghasilkan penurunan nilai WCSS yang sangat minimal atau relatif datar. Hal ini mengindikasikan bahwa pembagian data menjadi lebih dari tiga *cluster* tidak memberikan peningkatan yang substansial terhadap kekompakan *cluster* dan hanya akan mempersulit interpretasi pola yang dihasilkan. Oleh karena itu, jumlah *cluster* optimal yang dipilih untuk menganalisis pola absensi karyawan PT X adalah  $k = 3$ .



**Gambar 4. Scatter Plot Sebelum K-Means**

Sumber : dokumentasi pribadi

Pada Gambar 4, titik-titik data sudah tampak terbagi menjadi beberapa kelompok yang relatif terpisah sebelum dilakukan *K-Means Clustering*, namun tidak berlabel atau diberi warna yang membedakan kelompok. Ini menunjukkan bahwa data memang memiliki pola alami yang mendekati *cluster*, meskipun tanpa informasi *cluster* yang eksplisit.



**Gambar 5. Scatter Plot Setelah K-Means**

Sumber : dokumentasi pribadi

Setelah penentuan jumlah *cluster* optimal telah dilakukan, maka *clustering* dengan algoritma *K-Means* dapat dilakukan. Pada Gambar 5, titik-titik data sudah diberi warna sesuai dengan hasil

pengelompokan *cluster* yang terbentuk. Terdapat tiga *cluster* berbeda yang diidentifikasi, masing-masing dengan warna berbeda. Titik besar berwarna merah adalah *centroid* atau pusat *cluster* yang merepresentasikan rata-rata posisi anggota *cluster* tersebut dalam ruang fitur yang sudah direduksi dimensinya dengan PCA. *Centroid* ini membantu menggambarkan karakteristik pusat dari masing-masing *cluster*.

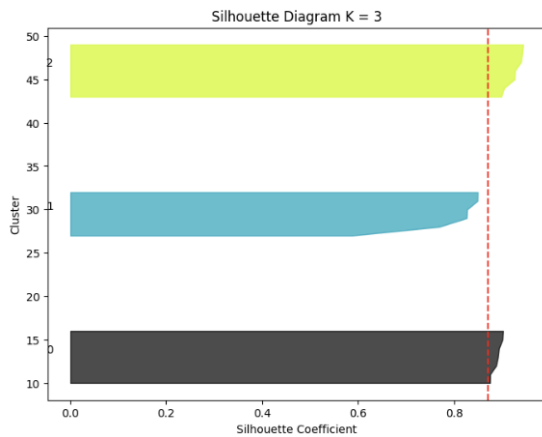
**Tabel 2. Variabel Cluster**

Cluster	Telat	Durasi Kerja	Hadir	Interpretasi
0	Rendah	Stabil	Sangat Tinggi	Sangat Disiplin
1	Sedang	Cukup stabil	Tinggi	Disiplin Moderat
2	Tinggi	Tidak stabil	Sedang-rendah	Disiplin Rendah

Sumber : dokumentasi pribadi

Berdasarkan *clustering* dengan algoritma *K-Means*, terdapat tiga *cluster* yaitu karyawan sangat disiplin, disiplin moderat, dan disiplin rendah. Penjelasan variabel *cluster* dapat dilihat pada Tabel 2. Secara sederhana, PCA memproyeksikan data multi-dimensi ke dalam dua dimensi agar bisa divisualisasikan, sedangkan *K-Means* melakukan pengelompokan berdasarkan kesamaan fitur. Meskipun data sudah tampak berkelompok di plot PCA, tanpa *K-Means* tidak ada cara sistematis untuk menentukan kelompok mana yang mana dan untuk menganalisis pola masing-masing *cluster* secara kuantitatif. *K-Means* memberikan hasil pengelompokan yang jelas dan bisa digunakan untuk analisis lebih lanjut, seperti membandingkan karakteristik tiap *cluster* dan membuat keputusan berbasis data.





**Gambar 6. Silhouette Plot**

Sumber : dokumentasi pribadi

Selain visualisasi PCA, kualitas hasil pengelompokan juga dievaluasi secara kuantitatif menggunakan *Silhouette Score*, yang pada penelitian ini memperoleh nilai 0,87. Nilai tersebut mengindikasikan bahwa *cluster* yang terbentuk memiliki kualitas sangat baik, karena menunjukkan adanya pemisahan yang jelas antar *cluster* sekaligus konsistensi yang tinggi di dalam masing-masing *cluster*. Hal ini diperlihatkan juga pada Gambar 6, di mana sebagian besar data memiliki nilai koefisien positif dengan bentuk batang yang stabil, menandakan bahwa mayoritas karyawan sesuai dengan *cluster* yang menaungi mereka. Berdasarkan hasil tersebut, algoritma *K-Means* terbukti mampu mengelompokkan karyawan PT X secara efektif berdasarkan pola absensi, sekaligus memberikan kontribusi baru berupa validasi model 3 *cluster* yang *actionable* untuk konteks perusahaan manufaktur yang datanya sebelumnya hanya digunakan untuk administrasi. Temuan ini memperkuat keyakinan bahwa hasil *clustering* dapat dimanfaatkan secara strategis dalam proses pengambilan keputusan, seperti pemantauan kedisiplinan, penyusunan kebijakan kehadiran, dan deteksi dini potensi masalah perilaku kerja.

## SIMPULAN

Penelitian ini berhasil menerapkan algoritma *K-Means Clustering* untuk mengelompokkan pola absensi karyawan berdasarkan data absensi selama satu tahun

di divisi produksi PT X. Tujuan utamanya adalah menyediakan dasar yang lebih objektif bagi perusahaan dalam menilai tingkat kedisiplinan, merencanakan kebutuhan SDM, dan mengidentifikasi potensi permasalahan kehadiran. Pengelompokan dilakukan menggunakan sejumlah indikator yang diolah melalui proses *feature engineering*, seperti tingkat keterlambatan, frekuensi pulang cepat, variasi jam masuk dan pulang, durasi kerja harian, dan jumlah ketidakhadiran. Melalui *Elbow Method* ditentukan bahwa jumlah *cluster* optimal adalah tiga, dan hasil ini diperkuat oleh nilai *Silhouette Score* sebesar 0,87, yang menunjukkan bahwa *cluster* yang terbentuk memiliki pemisahan yang jelas dan konsistensi internal yang baik.

Ketiga *cluster* tersebut menggambarkan variasi pola absensi yang kemudian diinterpretasikan menjadi tingkat kedisiplinan yang berbeda. *Cluster 0* menunjukkan kelompok karyawan dengan disiplin tinggi, ditandai dengan keterlambatan yang hampir tidak ada dan kehadiran yang tinggi. *Cluster 1* berisi karyawan dengan tingkat kedisiplinan moderat, di mana keterlambatan maupun pulang cepat terjadi, tetapi masih dalam batas yang dapat ditoleransi. Sementara itu, *Cluster 2* mencerminkan kelompok dengan tingkat kedisiplinan rendah, ditandai dengan frekuensi terlambat yang lebih sering, durasi kerja yang tidak stabil, serta ketidakhadiran yang lebih tinggi dibanding *cluster* lain. Ketiga *cluster* ini memberikan gambaran yang cukup jelas mengenai variasi perilaku kehadiran di lingkungan kerja.

Dari keseluruhan 24 karyawan yang dianalisis, proses *clustering* mampu menunjukkan pola absensi yang konsisten dengan data historis. Temuan ini tidak hanya menunjukkan bahwa *K-Means* efektif digunakan dalam analisis perilaku kerja, tetapi juga memberikan dasar pertimbangan yang kuat bagi manajemen PT X dalam melakukan evaluasi kedisiplinan, menyusun kebijakan

kehadiran yang lebih tepat, dan menyiapkan langkah pembinaan yang sesuai kebutuhan tiap kelompok. Dengan demikian, penggunaan *K-Means* dalam penelitian ini terbukti mampu memberikan manfaat praktis dalam mendukung pengelolaan sumber daya manusia berbasis data.

## DAFTAR PUSTAKA

- Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2020). A systematic review on supervised and unsupervised machine learning algorithms for data science. *Supervised and Unsupervised Learning for Data Science*, 3–21.
- Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
- Andrian, G., Arisandi, D., & Handhayani, T. (2024). Clustering Data Meteorologi Wilayah Indonesia Timur Dengan Metode K-Means Dan Fuzzy C-Means. *INTI Nusa Mandiri*, 18(2), 100–106.
- Ardiansyah, R., Pangestu, B. A., Hidayat, R., Kusumasari, I. R., Pembangunan, U., Veteran, N., & Timur, J. (2024). Revolusi Pengambilan Keputusan: Dari Intuisi ke Sistem Informasi. *Jurnal Ilmiah Multidisiplin*, 2(2), 7–11. <https://doi.org/10.62017/merdeka>
- Barito, E. E., Beng, J. T., & Arisandi, D. (2022). Penerapan Algoritma C4. 5 Untuk Klasifikasi Mahasiswa Penerima Bantuan Sosial Covid-19. *Jurnal Ilmu Komputer Dan Sistem Informasi*, 10(1).
- Dong, G., & Liu, H. (2018). *Feature engineering for machine learning and data analytics*. CRC press.
- Garg, N., & Dwivedi, P. (2024). A Novel Approach for Exploring Data-Driven Nutritional Insights Using Clustering and Dimensionality Reduction Techniques. *SN Computer Science*, 5(8), 1019.
- Gregorio, K., Beng, J. T., & Tiatri, S. (2023). Apakah psychological capital dan work engagement membentuk job performance karyawan di era Society 5.0? *Jurnal Muara Ilmu Sosial, Humaniora, dan Seni*, 7(3), 540–549.
- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210.
- Kenny, K., Arisandi, D., & Sutrisno, T. (2024). Evaluasi penilaian kinerja karyawan dengan metode naïve bayes. *Computatio: Journal of Computer Science and Information Systems*, 8(1), 110–118.
- Lusiana, F., Beng, J. T., & Wasino. (2020). Grouping of tourism objects using geotagged photo with hierarchical clustering method in Bantul and Sleman. *IOP Conference Series: Materials Science and Engineering*, 852(1). <https://doi.org/10.1088/1757-899X/852/1/012166>
- Maori, N. A., & Evanita, E. (2023). Metode elbow dalam optimasi jumlah cluster pada k-means clustering. *Jurnal Simetris*, 14(2), 1–11.
- Muharmi, Y., & Nadriati, S. (2022). Analysis Of Employee Discipline Based On Digital Attendance With The K-Means Algorithm Method. *Jurnal Teknologi Dan Open Source*, 5(2), 115–135.
- Neco, A., Saputra, F. A., Abdullah, N. F., Ramadhani, R., Hermansyah, T. T., & Sitio, S. L. M. (2025). PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK ANALISIS POLA DATA EKONOMI HISTORIS. *JRIS: Jurnal Rekayasa Informasi Swadharma*, 5(2), 95–107.
- Noval, M., Windarsyah, W., & Marleny, F. D. (2025). Implementasi Algoritma K-Means Untuk Analisis Pola Penjualan Pada Toko Monisa. *Jurnal Media Informatika*, 6(3), 1996–2002.
- Pratama, A., Arisandi, D., & Pradana, N. J. (2021). Implementasi metode



- agglomerative hierarchical clustering pada website pemilihan tempat futsal: Studi kasus Kota Depok. *Jurnal Ilmu Komputer dan Sistem Informasi*, 9(1), 1–6.
- Rosyada, I. A., & Utari, D. T. (2024). Penerapan Principal Component Analysis untuk Reduksi Variabel pada Algoritma K-Means Clustering. *Jambura Journal of Probability and Statistics*, 5(1), 6–13.
- Safira, F. A. (2024). PENGARUH DISIPLIN KERJA, ABSENSI, KEMAMPUAN KERJA DAN INSENTIF TERHADAP PRODUKTIVITAS KARYAWAN PT. BUMI PANDAAN PLASTIK. *Soetomo Management Review*, 2(4), 435–442.
- Simarmata, S., Sinurat, E. J., & Saragih, H. (2025). Pengaruh Employee Participation, Budaya Organisasi, dan Kedisiplinan Terhadap Produktivitas Kerja Pegawai. *Jurnal Ilmiah Manajemen, Bisnis Dan Kewirausahaan*, 5(2), 501–514. <https://doi.org/10.55606/jurimbik.v5i2.1133>
- Syahkur, M. R., Hartama, D., & Solikhun, S. (2024). Evaluasi Jumlah Cluster pada Algoritma K-Means++ Menggunakan Silhouette dan Elbow dengan Validasi Nilai DBI dalam Mengelompokkan Gizi Balita. *JST (Jurnal Sains Dan Teknologi)*, 13(3).
- Utami, I. T., Suryaningrum, F., & Ispriyanti, D. (2023). K-means cluster count optimization with silhouette index validation and davies bouldin index (CASE study: Coverage of pregnant women, childbirth, and postpartum health services in indonesia in 2020). *BAREKENG J. Ilmu Mat. Dan Terap*, 17(2), 707–716.
- Wijaya, A. C., Marcydiaz, A. H., Fitri, F. N., Arisandi, D., & Beng, J. T. (2024). Perancangan sistem dashboard penjualan berbasis web untuk Toko Online Caro Cara. *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, 10(2), 2114–2125.
- Wulandari, W., & Lailiyah, N. (2022). Produktivitas Kerja dan Hubungannya dengan Motivasi, Upah dan Kedisiplinan Kerja Karyawan. *JPEKA: Jurnal Pendidikan Ekonomi, Manajemen Dan Keuangan*, 6(2), 149–160.
- Yehezkiel, R., Beng, J. T., Arisandi, D., & Salsabila, T. M. (2025). Perbandingan biaya operasional sebagai upaya peningkatan efisiensi melalui dashboard Power BI. *INTECOMS: Journal of Information Technology and Computer Science*, 8(6). <https://doi.org/10.31539/gep60y68>