

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam manajemen sumber daya manusia, analisis data kehadiran karyawan merupakan elemen penting yang dapat meningkatkan produktivitas dan efisiensi organisasi. Perusahaan yang mampu melakukan analisis kehadiran secara efektif akan dapat mengidentifikasi pola-pola kehadiran yang berkaitan dengan kinerja, sehingga mendukung pengambilan keputusan berbasis data (Wibowo & Tjahjono, 2023). Dengan kemajuan teknologi, pencatatan kehadiran kini telah beralih ke sistem digital, seperti fingerprint, RFID, dan aplikasi berbasis cloud, yang memungkinkan pengumpulan data dalam volume besar (Pratiwi & Ahmadi, 2025). Data ini tidak hanya berguna untuk administrasi absensi, namun juga dapat dianalisis lebih lanjut untuk mendeteksi pola kehadiran dan tren kerja karyawan.

Salah satu metode yang banyak diterapkan dalam analisis data kehadiran adalah teknik klasterisasi, yang memungkinkan perusahaan untuk mengelompokkan karyawan berdasarkan karakteristik kehadiran mereka (Arda et al., 2024). Melalui teknik ini, manajemen dapat mengenali kelompok karyawan dengan kebiasaan kehadiran yang serupa, seperti mereka yang selalu tepat waktu, sering terlambat, atau yang memiliki pola kerja fleksibel (Friano Bunga et al., 2024). Hasil dari klasterisasi ini dapat digunakan untuk merancang jadwal kerja yang lebih efisien, serta menetapkan kebijakan berbasis data yang bertujuan meningkatkan kinerja karyawan.

Namun, tantangan utama yang dihadapi dalam analisis data kehadiran adalah pemilihan algoritma klasterisasi yang tepat. Algoritma *K-Means* dikenal karena kemampuannya dalam mengelola data besar, tetapi memiliki kelemahan dalam menentukan total klaster yang terbaik dan sensitif atas pemilihan titik awal pusat klaster (Salsabila et al., 2024). Di sisi lain, algoritma *Hierarchical* tidak memerlukan definisi jumlah klaster di awal dan lebih intuitif dalam visualisasi hubungan antar data, tetapi memiliki keterbatasan dalam kapasitas komputasi saat diterapkan pada dataset yang besar (Februariyanti et al., 2021). Oleh karena itu,

studi yang membandingkan kedua algoritma ini dalam pengelolaan kehadiran karyawan sangat diperlukan untuk menentukan pendekatan yang lebih efektif.

Beberapa penelitian sebelumnya telah membandingkan metode klusterisasi dalam konteks yang berbeda. Misalnya, penelitian oleh Nahya Nur dan kolega (2023) tentang pengelompokan daerah berisiko stunting di Indonesia menggunakan *K-Means* dan *Hierarchical Clustering* membuktikan *K-Means* lebih optimal, dengan nilai *Silhouette Coefficient* sebesar 0.48 dan *Calinski-Harabasz Index* sebesar 10.49, dibandingkan dengan *Hierarchical Clustering* yang mencatat nilai 0.47 dan 9.54 (Indra et al., 2023). Penelitian ini menggambarkan efektivitas *K-Means* dalam mengelompokkan wilayah berdasarkan parameter tertentu. Namun, penelitian tersebut tidak secara spesifik membahas pengelompokan data kehadiran karyawan, sehingga masih terdapat celah untuk penelitian lebih lanjut dalam penerapan kedua algoritma ini untuk analisis kehadiran. Selain itu, studi yang membandingkan algoritma klusterisasi dalam konteks data kehadiran masih terbatas, khususnya dalam penilaian performa berdasarkan *Silhouette Score*, *Calinski-Harabasz Index*, dan *Dunn Index*.

Dengan demikian, studi ini bertujuan untuk membandingkan algoritma *K-Means* dan *Hierarchical* dalam klusterisasi data kehadiran karyawan, sehingga dapat menentukan metode mana yang paling efektif. Diharapkan temuan analisis ini dapat memberikan keterlibatan dalam proses analitik sumber daya manusia dan memberikan wawasan praktis bagi perusahaan dalam memilih algoritma klusterisasi yang sesuai untuk pengelolaan kehadiran karyawan.

1.2 Rumusan Masalah

1. Bagaimana performa algoritma *K-Means* dan *Hierarchical* dalam melakukan klusterisasi data kehadiran karyawan di perusahaan XYZ?
2. Algoritma manakah yang memberikan hasil klusterisasi terbaik berdasarkan evaluasi metrik, seperti nilai *Silhouette Score*, *Calinski-Harabasz Index*, dan *Dunn Index*?

1.3 Tujuan Penelitian

1. Menganalisis performa algoritma *K-Means* dan *Hierarchical* dalam klusterisasi data kehadiran karyawan.

2. Membandingkan hasil klasterisasi kedua algoritma untuk menentukan algoritma yang paling efektif dalam mengidentifikasi karakteristik kinerja karyawan.

1.4 Manfaat

1. Memberikan kontribusi pada pengembangan ilmu pengetahuan di bidang data mining, khususnya terkait analisis klasterisasi dengan algoritma *K-Means* dan *Hierarchical*.
2. Membantu perusahaan XYZ dalam memahami pola kehadiran karyawan secara lebih efektif.

