

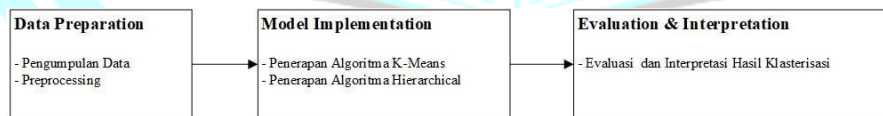
BAB III METODE PENELITIAN

3.1 Objek Penelitian

Penelitian ini berfokus pada data kehadiran karyawan dari sebuah perusahaan selama periode waktu satu tahun. Data yang digunakan mencakup informasi yang relevan dengan karakteristik kehadiran, seperti jumlah hari hadir, jumlah sakit, jumlah cuti mendadak, jumlah cuti terencana, dan variabel lain yang mendukung analisis kinerja karyawan. Penelitian ini membandingkan performa dua algoritma klasterisasi, yaitu *K-Means* dan *Hierarchical Clustering*. Pendekatan ini dipilih karena bertujuan untuk menganalisis dan mengukur efektivitas algoritma dalam mengelompokkan data kehadiran karyawan berdasarkan karakteristik kinerja mereka.

3.2 Prosedur Penelitian

Penelitian ini terdiri dari beberapa tahapan utama yang disusun secara sistematis, seperti yang digambarkan pada diagram flowchart di bawah. Setiap tahap dijelaskan dengan rinci pada gambar 3.1 untuk memberikan gambaran proses secara keseluruhan.



Gambar 3. 1. Tahapan Penelitian

Untuk memberikan pemahaman yang lebih terstruktur, setiap tahap penelitian dijelaskan secara rinci sebagai berikut:

3.2.1 Data Preparation

Pada tahap ini, dilakukan persiapan data yang meliputi proses pengumpulan data kehadiran karyawan dan preprocessing data untuk memastikan data yang digunakan dalam penelitian bersih, relevan, dan siap untuk dianalisis lebih lanjut.

3.2.2 Pengumpulan Data

Data yang diterapkan dalam studi ini berasal dari perusahaan XYZ, yang mencatat data kehadiran karyawan selama periode Januari hingga Desember 2024.

Dataset ini terdiri dari 441 data karyawan dan mencakup berbagai atribut seperti pada tabel 3.1 yang menggambarkan karakteristik kehadiran mereka.

Tabel 3. 1. Atribut Dataset

No	Atribut	Keterangan
1	Total Sakit	Jumlah hari karyawan tidak masuk kerja karena sakit selama tahun 2024.
2	Total Cuti Mendadak	Jumlah hari cuti yang diambil secara mendadak tanpa perencanaan selama tahun 2024.
3	Total Cuti Terencana	Jumlah hari cuti yang telah direncanakan dan disetujui sebelumnya selama tahun 2024.
4	Total Tidak Masuk	Jumlah keseluruhan ketidakhadiran karyawan yaitu total dari sakit, cuti mendadak, dan cuti terencana selama tahun 2024.
5	Total Kehadiran	Jumlah hari kerja yang dihadiri oleh karyawan, dihitung dari total hari kerja (240 hari) dikurangi total tidak masuk.
6	Rasio Kehadiran	Proporsi kehadiran karyawan dalam setahun, dihitung sebagai total kehadiran dibagi 240 hari kerja.
7	Cuti Keseluruhan	Jumlah total cuti yang diambil oleh karyawan, termasuk cuti mendadak dan cuti terencana selama tahun 2024.
8	Persentase Cuti	Persentase cuti yang diambil oleh karyawan dalam setahun, dihitung sebagai $(\text{Cuti Keseluruhan} \div 240) \times 100\%$.

3.2.3 Preprocessing

Proses *preprocessing* dilakukan dengan memeriksa apakah terdapat missing value dalam dataset. Selanjutnya, dipilih fitur yang paling relevan untuk analisis klusterisasi, yaitu Rasio Kehadiran dan Persentase Cuti. Pemilihan atribut ini bertujuan untuk menyederhanakan analisis serta lebih fokus dalam mengidentifikasi karakteristik kehadiran karyawan.

3.2.4 Model Implementation

Tahap ini merupakan penerapan algoritma klusterisasi terhadap data yang telah dipersiapkan. Algoritma yang digunakan dalam penelitian ini adalah *K-Means* dan *Hierarchical*, di mana masing-masing algoritma diterapkan untuk membentuk kluster berdasarkan karakteristik kehadiran karyawan.

3.2.5 Evaluation & Interpretation

Setelah penerapan algoritma, dilakukan evaluasi terhadap hasil klasterisasi menggunakan metrik evaluasi seperti *Silhouette Score*, *Calinski-Harabasz Index*, dan *Dunn Index*. Tahap ini bertujuan untuk menginterpretasikan kualitas klaster yang terbentuk dan menentukan algoritma mana yang memberikan hasil terbaik dalam mengelompokkan data kehadiran karyawan.

3.3 Alat dan Bahan

Penelitian ini menggunakan alat dan perangkat sebagai berikut:

1. *Google Colab* untuk analisis data dan implementasi algoritma.
2. Python dengan Menggunakan library seperti *pandas*, *numpy*, *sklearn*, *matplotlib*, dan *seaborn*.
3. Microsoft Excel untuk pemeriksaan awal dan pembersihan data sederhana.

