

BAB II

LANDASAN TEORI

2.1 Kedisiplinan Pegawai

Kedisiplinan pegawai merupakan salah satu elemen kunci dalam manajemen organisasi yang memiliki peran strategis dalam menentukan efektivitas kerja serta pencapaian tujuan organisasi. [1] mendefinisikan kedisiplinan sebagai suatu kondisi di mana pegawai mematuhi norma-norma organisasi yang mencakup aturan dan standar perilaku yang ditetapkan oleh manajemen. Selanjutnya, [2] menyatakan bahwa kedisiplinan adalah alat yang digunakan oleh manajemen untuk berkomunikasi dengan pegawai guna mendorong perubahan perilaku yang sejalan dengan tujuan organisasi. [1] Kedisiplinan dapat dilihat sebagai refleksi dari sikap dan perilaku pegawai yang mematuhi kebijakan organisasi. [3] menjelaskan bahwa kedisiplinan pegawai dapat diidentifikasi melalui indikator kehadiran tepat waktu, pelaksanaan tugas sesuai standar, serta ketaatan terhadap peraturan organisasi. Dengan demikian, kedisiplinan bukan hanya terkait kepatuhan terhadap aturan tetapi juga berperan penting dalam menjaga stabilitas dan efisiensi proses kerja.

Penelitian sebelumnya menunjukkan bahwa kedisiplinan pegawai dapat dipengaruhi oleh berbagai faktor, baik internal maupun eksternal. Faktor internal meliputi motivasi kerja, kompetensi, dan gaya kepemimpinan yang diterapkan dalam organisasi [3]. Faktor eksternal seperti lingkungan kerja, budaya organisasi, serta penerapan sanksi terhadap pelanggaran disiplin juga memiliki pengaruh yang signifikan [4]; [5]. [6]) menambahkan bahwa kepemimpinan yang efektif dan

pemberian insentif yang adil dapat meningkatkan kedisiplinan pegawai dengan menciptakan lingkungan kerja yang mendukung.

Kedisiplinan yang baik sangat penting untuk mendukung produktivitas dan efektivitas organisasi. Penelitian oleh [7] menunjukkan bahwa penerapan sanksi yang adil, disertai perhatian terhadap kebutuhan pegawai, dapat meningkatkan tingkat kedisiplinan. Dalam jangka panjang, hal ini berdampak pada peningkatan produktivitas kerja serta efektivitas operasional organisasi secara keseluruhan. [8] menjelaskan bahwa keberhasilan suatu organisasi sangat bergantung pada kemampuan manajemen untuk memonitor dan memastikan tingkat kedisiplinan pegawai secara konsisten. Pengukuran kedisiplinan pegawai biasanya dilakukan dengan menganalisis beberapa indikator utama, seperti data kehadiran, keterlambatan, dan pelanggaran aturan. Hal ini memungkinkan organisasi untuk mengevaluasi tingkat kepatuhan pegawai secara objektif. Organisasi dapat mengidentifikasi area yang memerlukan perbaikan dan memberikan intervensi yang tepat guna meningkatkan kinerja pegawai [9].

Kemajuan teknologi telah memungkinkan pengelolaan kedisiplinan pegawai secara lebih efektif melalui penggunaan sistem berbasis data. [9] menunjukkan bahwa aplikasi presensi berbasis Android dapat digunakan untuk mencatat kehadiran pegawai secara real-time dan memberikan informasi yang akurat tentang kepatuhan pegawai terhadap jadwal kerja. Sistem ini tidak hanya meningkatkan efisiensi administrasi tetapi juga menyediakan data yang relevan untuk analisis lebih lanjut.

Oleh karena itu penggunaan algoritma data mining seperti K-Means *Clustering* menawarkan pendekatan berbasis teknologi untuk mengelompokkan tingkat kedisiplinan pegawai dalam kategori yang lebih objektif. [2] menyatakan bahwa sistem evaluasi kedisiplinan berbasis data memberikan transparansi yang lebih baik dan meningkatkan kepercayaan pegawai terhadap proses evaluasi. Dengan menggunakan perangkat analisis seperti RapidMiner, penelitian ini tidak hanya mampu mengevaluasi data secara cepat tetapi juga menghasilkan visualisasi yang dapat membantu manajemen dalam pengambilan keputusan.

Tabel 2.1 Indikator Utama Dalam Pengukuran Kedisiplinan Pegawai

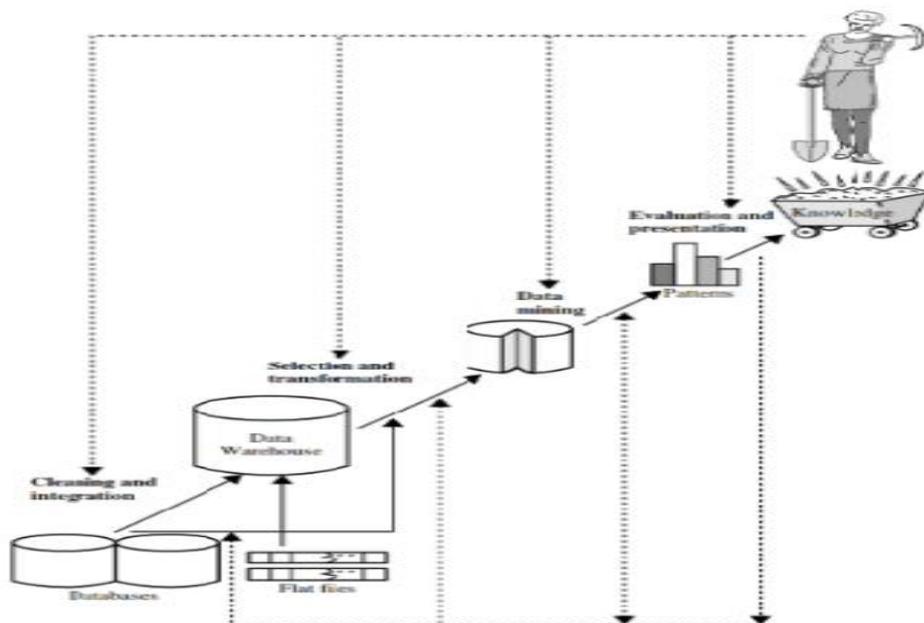
Indikator	Deskripsi	Satuan Pengukuran
Kehadiran	Jumlah hari hadir tepat waktu dalam sebulan.	Hari hadir tepat waktu per bulan
Keterlambatan	Rata-rata durasi keterlambatan per bulan.	Menit keterlambatan per bulan
Pelanggaran Aturan	Frekuensi pelanggaran yang tercatat dalam sebulan.	Jumlah pelanggaran per bulan

2.2 Data Mining

Data mining adalah suatu proses analisis data yang bertujuan untuk menggali informasi atau pola-pola tersembunyi dalam kumpulan data yang besar. Proses ini menggunakan berbagai teknik seperti statistik, matematika, kecerdasan buatan, dan machine learning untuk mengubah data mentah menjadi informasi strategis yang mendukung pengambilan keputusan. Dalam organisasi, data mining berfungsi untuk meningkatkan efisiensi operasional, memberikan wawasan strategis, serta membantu dalam merumuskan kebijakan berbasis data [10]. Penggunaan data mining memungkinkan organisasi untuk menemukan pola dan hubungan yang

mungkin tidak terlihat dengan metode analisis tradisional. Sebagai contoh, [10] menunjukkan bahwa penerapan algoritma data mining pada pengolahan data warehouse menghasilkan informasi yang relevan untuk perencanaan tenaga kerja di sektor kesehatan. Selain itu, analisis sentimen melalui data mining, seperti yang dijelaskan oleh [11], dapat memberikan wawasan tentang opini publik terhadap suatu layanan atau produk. Dengan demikian, data mining tidak hanya mendukung pengambilan keputusan yang berbasis data, tetapi juga memberikan pemahaman yang lebih baik terhadap dinamika sosial yang memengaruhi keputusan tersebut.

Salah satu keunggulan utama data mining adalah kemampuannya untuk meningkatkan efisiensi organisasi melalui otomatisasi proses analisis data., sehingga organisasi dapat menghemat waktu dan sumber daya, sekaligus meningkatkan kinerja secara keseluruhan. Adapun yang menjadi tahapan dalam proses data mining dapat dilihat pada gambar berikut:



Gambar 2.1 Tahapan Data Mining

a. Pembersihan Data (Data cleaning)

Data cleaning merupakan proses untuk mengatasi nilai yang hilang, noise, dan data yang tidak konsisten.

b. Integration Data (Data Integrasi)

Proses menambah data yang sudah ada dengan data atau informasi lain yang relevan

c. Seleksi Data (Data Selection)

Pemilihan data yang relevan dan dapat dilakukan analisis dari data operasional yang ada. Data hasil seleksi disimpan dalam database yang terpisah.

d. Transformasi Data (Data Transformation)

Proses perubahan atau penggabungan data kedalam format yang sesuai untuk diproses dalam data mining.

e. Data Mining

Data mining adalah proses menentukan pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Metode atau algoritma dalam data mining dapat bervariasi. Data yang digunakan untuk bisa menjadi sebuah model yang baik idealnya mencukupi sebagai data riset. Semakin banyak data dan semakin sedikit kesalahan (error) maka semakin bagus model yang dijadikan patokan.

f. Evaluasi Pola (Pattern Evaluation)

Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta pola atau informasi yang ada

sebelumnya. Proses interpretation berguna untuk menerjemahkan pola-pola yang didapat dari data mining menjadi bentuk yang mudah dimengerti.

g. Presentasi pengetahuan (Knowledge presentation)

Presentasi pengetahuan menampilkan pola informasi yang dihasilkan dari proses data mining, visualisasi ini membantu mengkomunikasikan hasil data mining dalam bentuk yang mudah dimengerti.

2.3 Clustering

Clustering adalah salah satu metode data mining yang berfungsi untuk mengelompokkan data berdasarkan kesamaan tertentu tanpa memerlukan label awal. Metode ini termasuk dalam kategori *unsupervised learning* yang digunakan untuk menemukan struktur dalam data yang tidak terlabel. Dengan *clustering*, data dapat dibagi ke dalam kelompok-kelompok yang memiliki karakteristik serupa, sehingga lebih mudah untuk dianalisis. [12] menyebutkan bahwa *clustering* memudahkan interpretasi pola data dengan mengidentifikasi karakteristik yang relevan dalam setiap cluster, menjadikannya alat yang fleksibel untuk eksplorasi data.

Clustering memberikan manfaat signifikan dalam menganalisis perilaku dan kinerja pegawai. [13] menunjukkan bahwa pengelompokan pegawai berdasarkan respons terhadap pengaturan kerja fleksibel membantu organisasi merancang kebijakan yang lebih efektif. [14] juga menyatakan bahwa *clustering* bermanfaat untuk mengelompokkan pegawai berdasarkan tingkat keterlibatan dan perilaku inovatif mereka, sehingga mendukung pengembangan sumber daya manusia.

Selain itu, clustering memungkinkan organisasi untuk mendeteksi anomali dalam data. [15] menjelaskan bahwa dengan mengidentifikasi pola dalam cluster normal, organisasi dapat menemukan data yang tidak sesuai dengan pola tersebut. Hal ini membantu dalam pengambilan keputusan yang lebih baik, terutama dalam konteks pengelolaan risiko atau pengawasan kinerja.

2.4 Algoritma K-Means

Algoritma K-Means merupakan salah satu teknik clustering yang paling populer dalam analisis data. Prinsip dasar dari algoritma ini adalah mengelompokkan data ke dalam sejumlah k kluster berdasarkan kedekatannya dengan pusat kluster (*centroid*). "K-Means bekerja dengan cara menginisialisasi sejumlah centroid, kemudian menempatkan data ke dalam kluster berdasarkan jarak terdekat, dan terus memperbarui posisi centroid hingga mencapai titik konvergensi". Proses kerja algoritma K-Means melibatkan beberapa tahapan utama sebagai berikut:

1. Inisialisasi Centroid

Algoritma dimulai dengan memilih K centroid secara acak dari dataset. Pemilihan centroid awal sangat memengaruhi hasil akhir klusterisasi. [16] mencatat bahwa metode inisialisasi seperti *k-means* dapat meningkatkan efisiensi dan akurasi algoritma.

2. Pengelompokan Data

Setiap data dikelompokkan ke dalam kluster berdasarkan jarak terdekatnya dengan centroid. Jarak ini dihitung menggunakan *Euclidean Distance*, dengan rumus:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

di mana x dan y adalah koordinat data dan centroid.

3. Pembaruan Centroid

Setelah data dikelompokkan, centroid baru dihitung sebagai rata-rata dari data dalam setiap kluster menggunakan rumus:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q$$

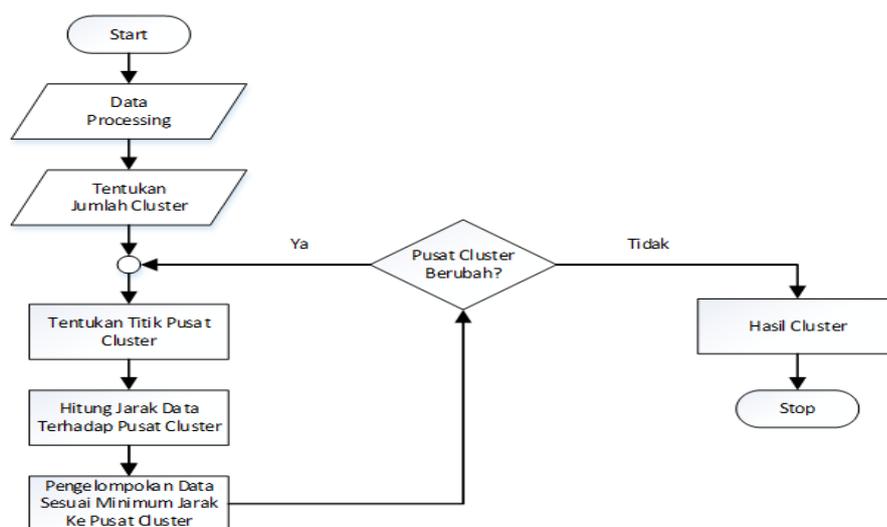
Dimana : μ_k = titik centroid dari cluste ke-K

N_k = banyaknya data padacluster ke-K

x_q = data ke-q pada cluster ke - K

4. Iterasi

Langkah pengelompokan dan pembaruan centroid diulang hingga tidak ada perubahan signifikan pada posisi centroid atau jumlah iterasi



Gambar 2. 2 Diagram Tahapan Algoritma K-Means

Algoritma K-Means memiliki beberapa keunggulan utama, antara lain:

1. Kesederhanaan Implementasi

Algoritma ini mudah dipahami dan diterapkan pada berbagai jenis data.

2. Efisiensi Komputasi

Kompleksitas waktu algoritma adalah $O(n \cdot k \cdot i)$ di mana n adalah jumlah data, k adalah jumlah kluster, dan i adalah jumlah iterasi [17].

3. Kemampuan Mengelompokkan Data Besar

Algoritma ini dapat memproses dataset besar dengan cepat, terutama untuk data yang memiliki batas antar kluster yang jelas [18].

Algoritma ini juga memiliki beberapa keterbatasan yaitu:

1. Sensitivitas terhadap Inisialisasi Centroid

Pemilihan centroid awal secara acak dapat menghasilkan hasil yang suboptimal.

2. Kebutuhan Penentuan Jumlah Kluster (KKK)

Jumlah kluster harus ditentukan sebelumnya, yang sering kali sulit tanpa evaluasi tambahan [19].

3. Kesulitan Mengatasi Data Non-Linier

Algoritma ini kurang efektif untuk data dengan distribusi yang kompleks atau tidak linier [20].

Algoritma K-Means telah digunakan secara luas dalam analisis perilaku dan kinerja pegawai. Algoritma ini dapat membantu mengelompokkan pegawai berdasarkan indikator seperti kehadiran, keterlambatan, dan pelanggaran aturan. Penelitian oleh [13] menunjukkan bahwa pengelompokan pegawai berdasarkan

respons terhadap kebijakan organisasi dapat membantu manajemen merancang kebijakan yang lebih efektif. [14] mencatat bahwa K-Means dapat digunakan untuk mengidentifikasi kelompok pegawai berdasarkan tingkat keterlibatan dan perilaku inovatif. Dengan mengelompokkan pegawai ke dalam kategori yang relevan, organisasi dapat merancang program pengembangan yang lebih sesuai dengan kebutuhan masing-masing kelompok. Keunggulan utama K-Means dalam menganalisis data kedisiplinan adalah kemampuannya untuk menyederhanakan data kompleks menjadi kluster yang lebih mudah dianalisis. [15] menggunakan algoritma ini untuk mendeteksi pola anomali dalam data kedisiplinan. Hasil klusterisasi dapat membantu organisasi mengambil tindakan korektif lebih awal untuk meningkatkan kinerja pegawai.

Tantangan dalam penerapan K-Means pada data kedisiplinan adalah sensitivitas terhadap *outlier* dan kebutuhan untuk menentukan jumlah kluster (KKK) secara tepat. Oleh karena itu, pemilihan parameter yang optimal sangat penting untuk memastikan akurasi hasil klusterisasi.

Tabel 2.2 Kelebihan Dan Kekurangan Algoritma K-Means Dalam Penerapannya Untuk Analisis Kedisiplinan Pegawai

Aspek	Kelebihan	Kekurangan
Efisiensi	Cepat untuk dataset besar	Sensitif terhadap posisi awal centroid
Fleksibilitas	Mudah diterapkan pada berbagai jenis data	Membutuhkan jumlah kluster yang ditentukan
Interpretasi	Memudahkan analisis pola dalam data	Kurang efektif untuk data non-linier

2.5 RapidMiner

RapidMiner adalah perangkat lunak berbasis data mining yang dirancang untuk mendukung proses analisis data secara menyeluruh, mulai dari pengumpulan data hingga visualisasi hasil. Perangkat ini menawarkan fleksibilitas dalam menangani berbagai algoritma analisis, seperti algoritma Apriori untuk analisis pola, algoritma K-Means untuk clustering, dan algoritma klasifikasi seperti C4.5 dan Naïve Bayes [21]; [22]. RapidMiner juga dikenal karena kemampuannya untuk mengintegrasikan metode statistik dengan kecerdasan buatan, yang memungkinkan pengguna untuk mendapatkan informasi berkualitas tinggi dari data yang diolah [23].

RapidMiner memiliki antarmuka visualnya yang intuitif, yang memungkinkan pengguna untuk merancang alur analisis data tanpa memerlukan keahlian pemrograman yang mendalam. Fitur ini menjadikannya alat yang populer di berbagai bidang, seperti bisnis, pendidikan, dan kesehatan [24]

RapidMiner mendukung penerapan algoritma clustering, seperti K-Means, yang memungkinkan pengelompokan data berdasarkan kesamaan karakteristik. Tahapan utama dalam proses *clustering* menggunakan RapidMiner mencakup:

1. Pengumpulan Data

Data dapat diimpor dari berbagai format, seperti file CSV, database SQL, atau API web, sehingga memudahkan pengguna dalam mengintegrasikan data dari berbagai sumber [25].

2. Pra-pemrosesan Data

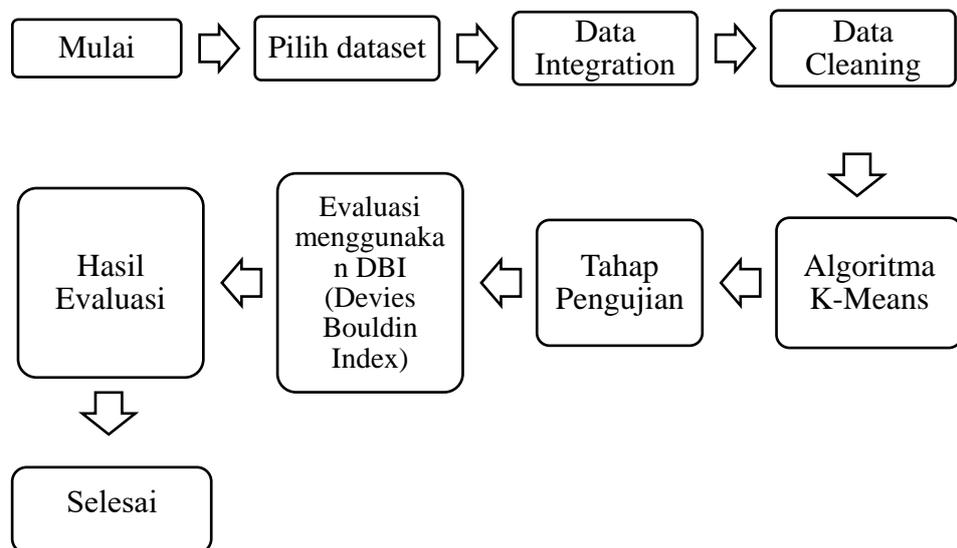
RapidMiner menyediakan alat untuk membersihkan data, mengatasi nilai yang hilang, dan melakukan normalisasi, yang penting untuk memastikan data siap digunakan dalam analisis lebih lanjut.

3. Pemilihan Algoritma

Pengguna dapat memilih algoritma clustering, seperti K-Means, menentukan jumlah kluster, dan mengatur parameter lain yang diperlukan untuk analisis.

4. Evaluasi dan Visualisasi

Hasil clustering dapat dievaluasi menggunakan metrik seperti indeks Davies-Bouldin, dan hasilnya divisualisasikan dalam grafik yang informatif untuk memudahkan interpretasi [26].



Gambar 2. 3 Alur Proses Clustering dengan RapidMiner

Keunggulan :

1. Antarmuka Visual yang Intuitif

RapidMiner mempermudah pengguna dari berbagai tingkat keahlian untuk menjalankan analisis data tanpa memerlukan keahlian pemrograman yang mendalam [23].

2. Dukungan Beragam Algoritma

RapidMiner mendukung berbagai algoritma analisis data, termasuk clustering, klasifikasi, dan analisis pola [27].

3. Kemampuan Visualisasi yang Kuat

Dengan alat visualisasi yang kuat, pengguna dapat memahami data dan hasil analisis dengan lebih cepat dan intuitif [24].

4. Aksesibilitas dan Biaya Efisien

RapidMiner tersedia dalam versi open-source, sehingga cocok untuk pengguna dengan anggaran terbatas [25].

Keterbatasan :

1. Kinerja pada Dataset Besar

Pada dataset yang sangat besar, RapidMiner dapat mengalami keterbatasan kinerja dibandingkan perangkat lunak yang dirancang khusus untuk skala besar [27].

2. Ketergantungan pada Antarmuka Visual

Pengguna tingkat lanjut mungkin merasa terbatas dalam fleksibilitas analisis dibandingkan dengan perangkat lunak berbasis skrip [28]

Tabel 2.3 Keunggulan dan keterbatasan RapidMiner berdasarkan fitur antarmuka, dukungan algoritma, dan kemampuan visualisasi

Aspek	Keunggulan	Keterbatasan
Antarmuka	Intuitif, cocok untuk pemula	Terbatas bagi pengguna tingkat lanjut
Dukungan Algoritma	Beragam algoritma tersedia	Tidak dirancang khusus untuk analisis kompleks
Visualisasi	Memiliki alat visualisasi yang kuat	Kurang fleksibel untuk kustomisasi tingkat lanjut

2.6 Davies-Bouldin Index (DBI)

Evaluasi model adalah tahap penting dalam penelitian untuk menilai kualitas hasil clustering yang dilakukan dengan algoritma K-Means. Proses evaluasi ini bertujuan untuk mengukur seberapa baik objek data terklasifikasi dalam kluster yang tepat, serta memastikan bahwa hasil clustering optimal. Terdapat beberapa metode evaluasi yang umum digunakan untuk menilai kualitas hasil clustering menggunakan K-Means

Davies-Bouldin Index (DBI) adalah salah satu metode yang digunakan untuk mengevaluasi kualitas hasil clustering dengan mengukur rasio antara jarak antar kluster dan jarak dalam kluster. Semakin rendah nilai DBI, semakin baik kualitas clustering yang dihasilkan. DBI mengidentifikasi kluster yang saling tumpang tindih, di mana nilai DBI yang lebih tinggi menunjukkan bahwa kluster tersebut kurang terpisah dengan jelas. Metode ini memberikan gambaran objektif tentang kualitas cluster berdasarkan jarak antar kluster dan internalitas kluster. Semakin kecil nilai DBI, semakin baik pemisahan antara kluster dan semakin kuat struktur dalam kluster tersebut ([29]; [30]). Penelitian oleh [29] menunjukkan bahwa

penggunaan DBI dalam evaluasi clustering K-Means membantu dalam menentukan pemisahan yang jelas antar klaster, di mana nilai DBI yang lebih rendah mengindikasikan clustering yang lebih baik dan lebih terstruktur.

2.7 Kelebihan Penelitian

Penelitian ini memiliki keunggulan signifikan dibandingkan penelitian terdahulu karena menggunakan RapidMiner sebagai alat bantu analisis yang lebih efisien dan terintegrasi. Dengan memperhitungkan faktor kehadiran, keterlambatan, dan pelanggaran aturan, penelitian ini menawarkan hasil clustering yang lebih komprehensif. Penelitian ini menawarkan pendekatan yang lebih holistik dalam pengukuran kedisiplinan pegawai, dengan mempertimbangkan berbagai variabel kedisiplinan secara bersamaan.