



K-Means Clustering Untuk Mengukur Pengaruh Kompetensi Terhadap Kinerja Pegawai

Saidatus Syaqila*, M. Fakhriza

Fakultas Sains dan Teknologi, Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Deli Serdang
Jl. Lap. Golf No.120, Kp. Tengah, Kec. Pancur Batu, Kabupaten Deli Serdang, Sumatera Utara, Indonesia

Email: ^{1,*}saidatussyaqila511@gmail.com, ²fakhriza@uinsu.ac.id

Email Penulis Korespondensi: saidatussyaqila511@gmail.com

Submitted: 16/01/2025; Accepted: 31/01/2025; Published: 31/01/2025

Abstrak—Sumber daya manusia memainkan peran penting dalam meningkatkan kinerja organisasi, termasuk di lingkungan Dinas Pemuda dan Olahraga (Dispora) Provinsi Sumatera Utara. Penelitian ini bertujuan untuk mengukur pengaruh kompetensi terhadap kinerja pegawai dengan menggunakan algoritma K-Means Clustering, yang dikenal sebagai metode pengelompokan data tanpa supervisi. Dataset terdiri dari 700 data pegawai dengan 15 atribut yang mencakup kompetensi teknis, manajerial, dan sosial. Data dikumpulkan melalui survei langsung dan diolah menggunakan Python dengan proses normalisasi melalui metode StandardScaler untuk memastikan konsistensi data. Metode elbow digunakan untuk menentukan jumlah cluster optimal, menghasilkan lima kluster: kinerja terbaik, sangat baik, dan rata-rata. Hasil analisis menunjukkan bahwa Hasil clustering mengelompokkan pegawai ke dalam lima cluster, yaitu Cluster 0 dengan 145 pegawai yang memiliki kompetensi teknis tinggi, Cluster 1 dengan 160 pegawai yang menonjol dalam kompetensi sosial dan manajerial, Cluster 2 dengan 125 pegawai yang menunjukkan kompetensi rata-rata pada semua aspek, Cluster 3 dengan 135 pegawai yang memiliki kompetensi teknis sedang tetapi unggul dalam kompetensi sosial, dan Cluster 4 dengan 135 pegawai yang memiliki potensi besar untuk berkembang. Penelitian ini memberikan manfaat praktis berupa identifikasi pola kompetensi untuk pengembangan pelatihan berbasis kebutuhan kelompok, serta pengambilan keputusan strategis yang lebih objektif dalam pengelolaan sumber daya manusia. Dengan demikian, penelitian ini diharapkan dapat mendukung peningkatan kinerja pegawai melalui pendekatan berbasis data yang efektif.

Kata Kunci: Machine Learning; K-Means Clustering; Data Mining; Kompetensi Pegawai; Kinerja Pegawai

Abstract—Human resources play an important role in improving organizational performance, including in the North Sumatra Province Youth and Sports Agency (Dispora). This study aims to measure the effect of competence on employee performance using the K-Means Clustering algorithm, known as an unsupervised data clustering method. The dataset consists of 700 employee data with 15 attributes covering technical, managerial, and social competencies. Data were collected through direct surveys and processed using Python with a normalization process through the StandardScaler method to ensure data consistency. The elbow method was used to determine the optimal number of clusters, resulting in five clusters: best performance, very good, and average. The results of the analysis show that the clustering results group employees into five clusters, namely Cluster 0 with 145 employees who have high technical competence, Cluster 1 with 160 employees who excel in social and managerial competence, Cluster 2 with 125 employees who show average competence in all aspects, Cluster 3 with 135 employees who have moderate technical competence but excel in social competence, and Cluster 4 with 135 employees who have great potential for development. This research provides practical benefits in the form of identifying competency patterns for developing group-based training needs, as well as more objective strategic decision-making in human resource management. Thus, this research is expected to support improving employee performance through an effective data-based approach.

Keywords: Machine Learning; K-Means Clustering; Data Mining; Employee Competence; Employee Performance

1. PENDAHULUAN

Sumber daya manusia adalah aset yang terpenting dalam meningkatkan kinerja organisasi, baik dalam swasta, sosial maupun pemerintah. Segala proses yang diperlukan dalam pencapaian tujuan organisasi tidak terlepas dari sumber daya manusia, baik dari pengambilan keputusan, hingga pada proses evaluasi yang semuanya tidak dapat dipisahkan dari unsur sumber daya manusia [1]. Demikian pula halnya atas suatu organisasi atau instansi pemerintah yang menekankan bahwa sumber daya manusia yang dalam hal ini pegawai berperan penting dalam memberikan pelayanan kepada masyarakat.

Oleh karena itulah setiap pegawai perlu memiliki pengetahuan kompetensi dan motivasi dalam penyelesaian pekerjaan sehingga dapat menunjang peningkatan kinerja pegawai. Dalam rangka menciptakan iklim kompetensi yang tinggi di segala bidang perusahaan dituntut untuk bekerja lebih efektif dan efisien. Tingkat kompetensi yang tinggi menuntut suatu perusahaan atau instansi untuk mengoptimalkan sumber daya manusia yang dimilikinya. Hal ini disebabkan karena sumber daya manusia mempunyai peran utama dalam setiap kegiatan perusahaan atau instansi. Sebab kesuksesan sebuah perusahaan atau instansi sangat bergantung dengan sumber daya manusia yang ada dalam perusahaan atau instansi tersebut.

Kinerja sebagai hasil-hasil fungsi pekerjaan atau kegiatan seseorang atau kelompok dalam suatu organisasi yang dipengaruhi oleh berbagai faktor untuk mencapai tujuan organisasi dalam periode waktu tertentu [2]. Oleh karena itu, kinerja pegawai yang baik secara langsung akan mempengaruhi kinerja lembaga dan untuk memperbaiki kinerja pegawai tentu merupakan suatu pekerjaan yang memakan waktu dan proses yang panjang. Selain meningkatkan pengawasan dan pembinaan, juga dilakukan penilaian terhadap tingkat keberhasilan kinerja



yang telah dilakukan oleh para pegawainya. Hal tersebut bertujuan untuk mengetahui pengaruh pimpinan, dan motivasi kerja terhadap kinerja pegawai pada Dinas Pemuda Dan Olahraga Provinsi Sumatera Utara yang telah dicapai oleh melalui pelaksanaan pemerintahan dalam meningkatkan kinerja pegawai di lingkungan DISPORA Provinsi Sumatera Utara. Selain itu, penilaian terhadap kinerja juga bermanfaat sebagai tolak ukur yang dapat digunakan untuk memperbaiki kinerja pegawai yang bersangkutan. Untuk itulah penelitian ini sangat penting karena selain untuk mengetahui kompetensi motivasi dan kinerja para pegawai

Berdasarkan penelitian yang dilakukan oleh Marcelina dkk (2023), dimana hasil penelitian Studi ini menunjukkan bahwa penerapan algoritma K-Means Clustering menggunakan KNIME efektif untuk mengelompokkan UKM berdasarkan karakteristiknya. Hasil klusterisasi menghasilkan tiga kelompok: UKM mandiri, berkembang, dan binaan, yang memberikan wawasan strategis bagi DinKop-UKM dalam merancang program pengembangan. Evaluasi model melalui Regression Predictor dan Numeric Score menunjukkan akurasi prediksi, sementara visualisasi Scatter Plot mempermudah pemahaman pola kluster. Visualisasi data menggunakan Scatter Plot memungkinkan pemahaman yang lebih baik tentang pola kluster yang dihasilkan. Pengelompokan ini menjadi masukan bagi DinKop-UKM untuk menerapkan program pengembangan strategi yang lebih tepat sasaran [3].

Namun, penelitian sebelumnya masih terbatas pada penerapan algoritma K-Means Clustering untuk pengelompokan UKM dan kinerja karyawan secara umum tanpa memberikan analisis yang lebih mendalam terkait pengaruh kompetensi terhadap kinerja pegawai di sektor tertentu, seperti instansi pemerintahan. Selain itu, penelitian-penelitian tersebut belum mengintegrasikan hasil klusterisasi dengan rekomendasi pengembangan yang dapat digunakan untuk meningkatkan kinerja pegawai secara langsung.

Selanjutnya penelitian yang dilakukan oleh Faran dan Triayudi dalam penerapan algoritma k-means dimana hasil penelitian menunjukkan bahwa penggunaan data mining untuk pembentukan kelompok kinerja karyawan merupakan pendekatan yang tepat, karena mampu menghasilkan informasi baru untuk pengambilan keputusan. Dalam proses penyelesaian dilakukan dengan menggunakan clustering dengan menggunakan algoritma K-Means, dalam hal ini proses pembentukan clustering ataupun pengelompokan kinerja pegawai berhasil dilakukan dimana terdapat 2 (dua) cluster yang didapatkan dari data. Hasil dari algoritma K-Means bahwasannya terdapat 7 (tujuh) data yang termasuk dalam Cluster 1 yaitu Karyawan 1, Karyawan 3, Karyawan 6, Karyawan 7, Karyawan 8, Karyawan 12 dan Karyawan 13 serta terdapat 8 (delapan) data yang termasuk dalam Cluster 2 yaitu Karyawan 2, Karyawan 4, Karyawan 5, Karyawan 9, Karyawan 10, Karyawan 11, Karyawan 14 dan Karyawan 15 [4].

Kemudian studi yang dilakukan oleh Mudakir (2023), berdasarkan hasil pengujian terhadap dataset penilaian karyawan untuk menentukan pengangkatan karyawan menggunakan metode K-Means dan Davies Bouldin Index untuk validasi maka dapat disimpulkan Metode K-Means dapat digunakan untuk penentuan pengangkatan karyawan, proses perhitungan dengan algoritma k-means dengan rumus Euclidean Distance. Di dapat hasil validasi DBI sebesar -2.803 hasil tersebut dapat di kategorikan cukup baik. Penentuan 2 cluster yaitu cluster 0 dengan nilai lulus terdapat 61 data karyawan dan cluster 1 dengan nilai tidak lulus terdapat 67 data karyawan, dari hasil tersebut maka peluang potensi untuk pengangkatan karyawan dengan hasil tidak lulus lebih banyak dikarenakan beberapa faktor penilaian yang sudah maksimal [5].

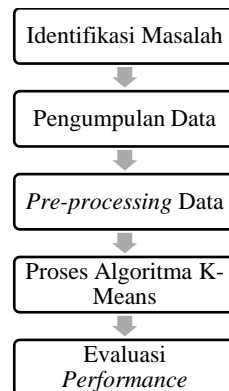
Berdasarkan latar belakang diatas pada penelitian ini digunakan algoritma K-Means Clustering mengenai mengukur pengaruh kompetensi terhadap kinerja pegawai. Penilaian kinerja karyawan menggunakan Clustering K-Means, dikarenakan K-Means Clustering merupakan metode data mining yang dapat memproses model tanpa supervisi dan proses pengelompokan dilakukan dengan sistem partisi. Metode clustering K-Means dapat mempermudah perusahaan dalam mengetahui kinerja karyawan yang memiliki kompetensi dan loyalitas yang tinggi dalam pengelompokannya terbagi menjadi 3 cluster yaitu terbaik, sangat baik dan rata-rata. K-Means clustering diharapkan dapat memberikan kontribusi terhadap perusahaan dalam pengelompokan kinerja karyawan dan perusahaan dapat mengambil keputusan dengan waktu yang singkat dan objektif. Bahwa lingkungan kerja berpengaruh signifikan terhadap kinerja pegawai. Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma K-Means Clustering guna mengukur pengaruh kompetensi terhadap kinerja pegawai pada Dinas Pemuda dan Olahraga Provinsi Sumatera Utara. Penelitian ini juga bertujuan untuk memberikan pengelompokan kinerja pegawai ke dalam tiga cluster (terbaik, sangat baik, dan rata-rata) serta memberikan rekomendasi pengembangan strategi berdasarkan hasil klusterisasi tersebut.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Kerangka penelitian merupakan tahapan-tahapan yang akan diterapkan pada saat kita melakukan penelitian. Maka dengan hal ini akan dilakukan agar tahapan penelitian yang dilakukan lebih terstruktur. Penelitian ini menggunakan metode K-Means Clustering yang merupakan metode fokus pada pengamatan yang mendalam sehingga menghasilkan suatu kejadian yang lebih komprehensif. Jenis metode penelitian deskriptif. Metode deskriptif merupakan pendekatan studi objek yang bertujuan untuk memberikan gambaran jelas mengenai suatu keadaan dengan mengumpulkan informasi dan menganalisis data. Pada penelitian yang akan dilakukan ialah menerapkan algoritma K-Means Clustering dengan rapidminer untuk melihat potensi pengaruh kinerja para

pegawai dalam bekerja . Berikut merupakan tahapan yang akan dilakukan pada proses penerapan algoritma K-Means Clustering untuk mengukur pengaruh kinerja para pegawai dalam bekerja. Alur proses penelitian ini terdiri dari beberapa tahapan utama seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Kerangka Penelitian

Gambar 1 menunjukkan alur penelitian yang terdiri dari lima tahap: Identifikasi Masalah untuk menentukan fokus penelitian, Pengumpulan Data yang relevan, Pre-processing Data untuk membersihkan dan mempersiapkan data, Proses Algoritma K-Means untuk mengelompokkan data, dan Evaluasi Performance untuk menilai kualitas hasil clustering.

Pada tahap identifikasi masalah pada penelitian dilakukan identifikasi kebutuhan yang ada dalam Clustering untuk mengukur pengaruh kompetensi terhadap kinerja pegawai seperti kurangnya objektivitas dalam proses seleksi atau kriteria evaluasi yang digunakan. Seperti proses wawancara dan studi literatur mengenai permasalahan yang terjadi didalam lingkungan kerja profesional. Sebelum itu berikut merupakan dasar teori yang digunakan pada penelitian ini.

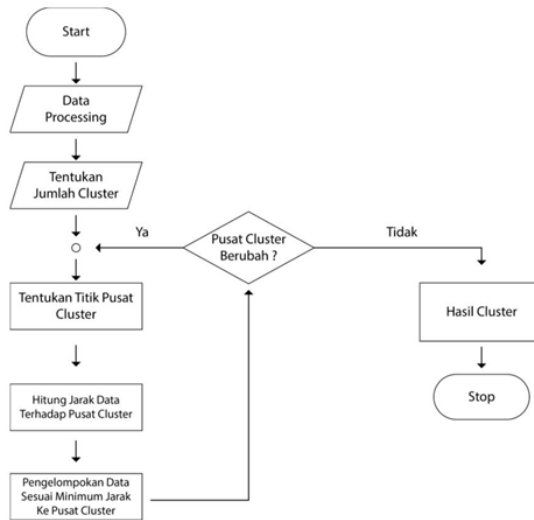
Menurut [6] Machine Learning (ML) adalah cabang dari kecerdasan buatan yang berfokus pada belajar dari data (learn from data), yaitu mengembangkan sistem yang mampu belajar secara "mandiri" tanpa perlu diprogram manusia lagi. Untuk mencapai hasil yang optimal, ML membutuhkan data yang sah sebagai bahan belajar (dalam proses pendidikan) sebelum digunakan untuk pengujian. Sedangkan menurut [7] salah satu cabang ilmu kecerdasan buatan adalah pembelajaran mesin, khususnya yang mempelajari bagaimana komputer dapat belajar dari data untuk meningkatkan kecerdasannya.

Ada beberapa teknik machine learning yaitu Supervised Learning dan Unsupervised Learning. Pembelajaran yang diawasi atau terarah (Supervised Learning) didefinisikan sebagai pembelajaran yang diawasi dengan menggunakan label pada data pembelajaran. Jenis pembelajaran yang diawasi dapat mencakup klasifikasi dan regresi [8]. Sedangkan Unsupervised learning menganalisis dan mengelompokkan dataset yang tidak dilabeli menggunakan algoritma pembelajaran mesin. Algoritma ini memungkinkan penemuan pola atau pengelompokan data yang tersembunyi tanpa intervensi manusia. Karena kemampuan untuk menemukan kesamaan dan perbedaan dalam data, sangat cocok untuk analisis data eksploratif, strategi penjualan silang, segmentasi pelanggan, dan pengenalan gambar [9]. Penelitian ini menggunakan unsupervised karena K-Means termasuk ke dalam unsupervised Learning.

Data mining adalah siklus yang memanfaatkan penalaran buatan manusia yang terukur, numerik, dan metode AI untuk memisahkan dan membedakan data berharga dan informasi terkait dari kumpulan data yang besar [10]. Clustering merupakan Proses mengorganisasikan objek-objek ke dalam kelompok-kelompok (cluster) yang anggota kelompoknya memiliki kemiripan di beberapa karakteristiknya [11]. Metode Clustering yang paling banyak dipelajari adalah metode partisi dan metode hirarki. Metode partisi bertujuan untuk menemukan pengelompokan yang terdapat dalam data dengan mengoptimalkan fungsi dan tujuan yang dapat meningkat partisi [12].

Tahap pengumpulan data dilakukan untuk memperoleh informasi yang relevan terkait kompetensi dan kinerja pegawai di Kantor Disporasu Kedepuitan Wilayah 1 Kota Medan. Data dikumpulkan melalui koordinasi langsung dengan instansi terkait (wawancara), mencakup variabel-variabel seperti pengetahuan teknis, kemampuan manajerial, dan keterampilan sosial. Proses pengumpulan data berlangsung dari Desember hingga selesai, dengan memastikan setiap data yang diterima telah diverifikasi untuk keakuratan dan konsistensi. Seluruh data yang terkumpul kemudian diolah ke dalam format csv menggunakan perangkat lunak Python untuk memudahkan analisis lebih lanjut, termasuk pre-processing data dengan standarisasi dan transformasi data yang diperlukan untuk pemodelan clustering. Proses algoritma K-Means dapat dijelaskan melalui diagram alir yang menunjukkan tahapan dari awal hingga memperoleh hasil pengelompokan data, seperti yang ditunjukkan pada Gambar 2. Gambar 2 menunjukkan langkah-langkah dalam algoritma K-Means, dimulai dari proses awal (Start), dilanjutkan dengan data processing untuk mempersiapkan data. Setelah itu, jumlah cluster ditentukan, dan titik pusat awal cluster dipilih. Algoritma menghitung jarak setiap data terhadap pusat cluster dan mengelompokkan

data berdasarkan jarak terdekat. Jika pusat cluster berubah, proses iterasi dilanjutkan hingga pusat cluster tidak mengalami perubahan. Setelah konvergen, hasil pengelompokan (Hasil Cluster) diperoleh, dan proses dihentikan (Stop). Diagram ini menggambarkan cara kerja iteratif algoritma K-Means untuk menghasilkan cluster yang optimal.



Gambar 2. Flowchart Penelitian

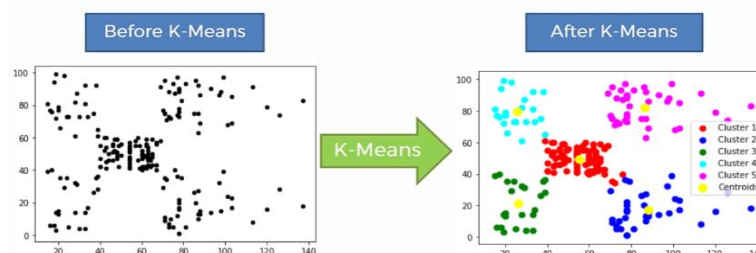
Sebelum melakukan clustering dengan K-Means, tahap preprocessing data dilakukan untuk membersihkan data dari nilai yang hilang atau outlier yang mungkin mempengaruhi hasil clustering [13]. Hal ini dapat meliputi normalisasi data jika skala variabel berbeda-beda. Disini digunakan metode StandardScaler untuk preprocessing data. Preprocessing data dalam data mining adalah proses penting yang melibatkan pembersihan, transformasi, dan integrasi data untuk mempersiapkan data tersebut agar sesuai untuk analisis [14]. Tujuan dari preprocessing data adalah untuk meningkatkan kualitas data dan membuatnya lebih cocok untuk tugas data mining tertentu [15]. Penelitian ini menggunakan metode preprocessing Standard Scaler, yang melakukan standarisasi fitur dengan menghapus rata-rata dan menskalakan unit varian. Adapun proses menghitung pre-processing pada data seperti tabel menggunakan rumus standar scaler seperti persamaan (1) berikut :

$$Z_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

Dimana Z_i digunakan untuk menghitung skor Z dari suatu data, dengan x_i sebagai nilai data individu, μ sebagai nilai rata-rata populasi atau sampel, dan σ sebagai standar deviasi populasi atau sampel.

Selanjutnya dengan melakukan penerapan algoritma K-Means clustering pada data kompetensi dan kinerja yang sudah dikumpulkan. Algoritma ini akan mengelompokkan pegawai ke dalam beberapa cluster berdasarkan kemiripan nilai kompetensi dan kinerja mereka. K-Means clustering ialah teknik analisis cluster yang bertujuan untuk membagi objek menjadi k cluster, dan kemudian mengamati bahwa setiap objek cluster memperoleh rata-rata terdekatnya [16]. Untuk memecahkan masalah pengelompokan dataset, algoritma ini mengelompokkan pengamatan ke dalam kelompok k, di mana k adalah parameter input. Berdasarkan kedekatan setiap pengamatan ke nilai rata-rata kelompok, setiap data kemudian ditetapkan pada kelompok k. K-means berusaha mempartisi data saat ini ke dalam satu atau lebih cluster. Ini adalah metode clustering non-hirarki [17].

Sebelum menggunakan algoritma K-Means, proses clustering dimulai dengan preprocessing data untuk memastikan data siap digunakan. Gambar 3 menunjukkan perubahan visual data sebelum dan sesudah diterapkan algoritma K-Means. Sebelum proses clustering, data masih dalam bentuk campuran tanpa pengelompokan yang jelas (Before K-Means). Setelah diterapkan algoritma K-Means, data terkelompok ke dalam tiga cluster yang berbeda, ditandai dengan warna yang berbeda pada visualisasi (After K-Means).



Gambar 3. Proses K-Means Clustering



Pada Gambar 3 menunjukkan proses K-Means clustering secara garis besar dimulai dengan menentukan jumlah cluster (k) yang akan dibuat, misalnya $k=5$. Selanjutnya, titik data awal untuk pusat cluster (centroid) ditentukan secara acak, dan data yang paling dekat dengan pusat cluster akan dimasukkan ke dalam cluster tersebut. Setelah itu, nilai rata-rata dari setiap data dalam cluster dihitung, yang kemudian memperbarui posisi pusat cluster (centroid). Proses ini diulang hingga semua data dalam satu cluster tidak lagi berpindah ke cluster lain, menandakan bahwa iterasi K-Means telah selesai [18].

Kelebihan utama dari algoritma K-Means adalah kemudahannya dalam implementasi dan efisiensi waktu komputasi, terutama untuk dataset yang besar. Oleh karena itu, kumpulan data yang memiliki ciri-ciri yang mirip berada di cluster yang sama, sedangkan kumpulan data yang memiliki ciri-ciri yang berbeda berada di cluster yang berbeda [19]. Persamaan digunakan untuk menentukan titik pusat cluster pada data (2):

$$C_{kj} = \frac{x_{1j} + x_{2j} + \dots + x_{nj}}{n} \quad (2)$$

Dimana C_{kj} digunakan untuk menentukan pusat cluster ke- k pada variabel j dengan $j = 1, 2, \dots, p$. Dalam perhitungan ini, n merepresentasikan banyaknya data pada cluster ke- k . Dan untuk perbarui nilai titik centroid pada persamaan (3) :

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i \quad (3)$$

Keterangannya adalah rumus μ_k digunakan untuk memperbarui nilai titik centroid dari cluster ke- k . Dalam hal ini, N_k menunjukkan jumlah data pada cluster ke- k , sementara x_i adalah nilai data ke- i dalam cluster ke- k . Setelah clustering selesai, analisis dan interpretasikan hasilnya [20]. Identifikasi setiap cluster yang terbentuk dan karakteristik kompetensi serta kinerja pegawai di dalamnya. Amati apakah ada pola yang menunjukkan pengaruh kompetensi tertentu terhadap kinerja pegawai.

Penelitian ini menggunakan bahasa pemrograman python merupakan sebuah bahasa pemrograman yang dapat dijalankan pada berbagai platform, terutama yang memprioritaskan keterbacaan kode. Penggunaan Python terutama ditujukan untuk analisis data, visualisasi data, serta pengembangan dan pembuatan AI [21]. Proses implementasi bertujuan untuk menemukan perbandingan antara dua metode klastering bahasa pemrograman yang digunakan dalam penelitian ini. Bahasa pemrograman yang digunakan dalam penelitian ini adalah python, dan aplikasinya menggunakan jupyter notebook [22]. Jupyter notebook adalah editor Python berbasis web yang dapat digunakan dengan peramban. Python dipilih karena banyaknya library yang mendukung pre-processing data, modeling data, dan implementasi model [23].

3. HASIL DAN PEMBAHASAN

Proses penelitian ini diawali dengan tahap perencanaan yang mencakup penentuan topik penelitian dan identifikasi kebutuhan data. Topik yang dipilih dalam penelitian ini adalah penerapan algoritma K-Means Clustering untuk mengukur pengaruh kompetensi terhadap kinerja pegawai. Penelitian ini membutuhkan data pegawai yang relevan, termasuk informasi tentang kompetensi teknis, manajerial, dan sosial. Data tersebut diperoleh dari Kantor Disporasus Kedeputian Wilayah 1 Kota Medan, dengan proses pengumpulan data dilakukan secara bertahap selama periode Maret hingga selesai. Langkah ini bertujuan untuk memastikan bahwa data yang dikumpulkan akurat, lengkap, dan sesuai dengan kebutuhan penelitian.

Tahap selanjutnya adalah pemodelan clustering menggunakan algoritma K-Means. Pada tahap ini, data pegawai dianalisis untuk menghasilkan pengelompokan yang representatif berdasarkan kompetensi yang dimiliki. Proses pemodelan meliputi normalisasi data guna menyamakan skala variabel, pemilihan atribut kompetensi utama sebagai parameter clustering, penggunaan metode inisialisasi centroid K-Means++ untuk meningkatkan efisiensi algoritma, serta evaluasi hasil clustering menggunakan metode Elbow. Hasil akhir dari tahap ini adalah pengelompokan data pegawai ke dalam beberapa cluster yang dapat dijadikan dasar untuk menarik kesimpulan mengenai pengaruh kompetensi terhadap kinerja mereka.

3.1 Dataset Penelitian

Dataset yang digunakan dalam penelitian ini terdiri dari 700 data pegawai dengan 15 atribut/kolom yang mencerminkan berbagai aspek kompetensi. Setiap atribut dinilai menggunakan skala Likert 1–5, di mana :

1. Sangat Kurang
2. Kurang
3. Cukup
4. Baik
5. Sangat Baik

Dataset ini mencakup tiga jenis kompetensi utama, yaitu (1) Kompetensi Teknis, meliputi kemampuan menguasai pengetahuan teknis, memecahkan masalah teknis secara akurat, menggunakan teknologi terbaru, menyelesaikan tugas teknis tanpa supervisi, dan memperbarui keterampilan sesuai kebutuhan organisasi; (2) Kompetensi Manajerial, yang mencakup pengelolaan waktu, perencanaan dan prioritas kerja, pengambilan

keputusan sesuai prosedur, kepemimpinan, serta pengelolaan konflik; dan (3) Kompetensi Sosial, meliputi kemampuan berkomunikasi dengan efektif, bekerja dalam tim, menunjukkan empati, beradaptasi dengan berbagai kepribadian, dan menerima kritik secara konstruktif. Data ini menjadi dasar analisis untuk memahami bagaimana ketiga jenis kompetensi tersebut memengaruhi kinerja pegawai dalam organisasi.

Tabel 1. Data Mentah

No.	Nama Pegawai	Menguasai pengetahuan teknis terkait pekerjaan	Mampu memecahkan masalah teknis dengan cepat dan akurat	Menggunakan teknologi terbaru dalam pekerjaan	Efektif dalam menyelesaikan tugas teknis tanpa supervisi	Terus memperbarui keterampilan teknis sesuai kebutuhan organisasi	...	Mampu menerima kritik dengan baik dan melakukan perbaikan
1	Musa Manurung	5	5	5	4	5	...	5
2	Yowan Alfitra	4	5	5	4	4	4	4
...
700	Andika	1	4	3	2	1	1	2

Setelah tabel 1 dataset asli dibuat, langkah awal dalam pengolahan data adalah menghapus kolom "Nama Pegawai" karena tidak relevan untuk analisis numerik. Fokus diarahkan pada atribut kompetensi yang telah dinilai dalam skala 1–5. Untuk mempermudah proses analisis dan visualisasi, nama kolom yang panjang diganti dengan label yang lebih singkat. Sebagai contoh, kolom “Menguasai pengetahuan teknis terkait pekerjaan” diubah menjadi “C1”, dan seterusnya hingga “Mampu menerima kritik dengan baik dan melakukan perbaikan” menjadi “C15”. Transformasi ini bertujuan untuk meningkatkan efisiensi dalam pemrosesan data tanpa mengurangi informasi yang terkandung dalam setiap variabel. Berikut adalah hasil data transformasi pada Tabel 2.

Tabel 2. Data Transformasi

No.	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
1	5	5	5	4	5	5	5	5	5	5	4	5	4	3	5
2	4	5	5	4	4	4	4	4	5	5	5	5	5	5	5
...
700	1	4	2	5	5	4	1	1	5	2	5	2	2	4	3

Tabel 2 menampilkan data awal yang diperoleh dari responden melalui Google Form. Data ini merepresentasikan penilaian kompetensi pegawai terhadap 15 atribut kompetensi (C1 hingga C15) dalam skala 1 hingga 5. Setiap baris dalam tabel merepresentasikan individu (responden) yang memberikan penilaian terhadap atribut-atribut tersebut. Data ini belum melalui proses standarisasi sehingga masih dalam bentuk nilai asli yang dikumpulkan dari responden.

3.2 Pre-processing

Tahap pre-processing dilakukan untuk memastikan data yang dikumpulkan melalui Google Form terkait Pengaruh Kompetensi Terhadap Kinerja Pegawai siap digunakan dalam analisis clustering. Data ini diisi langsung oleh pegawai yang bersangkutan dan mencakup penilaian terhadap 15 atribut kompetensi dalam skala 1–5. Karena setiap atribut memiliki skala yang sama, proses standarisasi data dilakukan menggunakan metode StandardScaler. Standarisasi ini bertujuan untuk mengubah nilai setiap atribut agar memiliki rata-rata 0 dan standar deviasi 1, sehingga variabel dengan skala besar tidak mendominasi hasil clustering. Proses ini memastikan bahwa algoritma K-Means Clustering bekerja secara optimal dengan memberikan bobot yang sama pada setiap atribut kompetensi seperti yang ditunjukkan pada Tabel 3 dibawah.

Tabel 3. Hasil Pre-Processing Data

No.	C1	C2	C3	C4	C5	C6	C7	...	C15
1	1.487391	1.382653	1.369403	0.679288	1.375229	1.421418	1.379622	...	1.449910
2	0.7771769	1.3826532	1.3694028	0.6792879	0.6730698	0.7009308	0.6747056	...	1.4499095
...
700	1.3534647 35	0.6635913 92	0.6759356 93	1.3900526 09	1.3752291 44	0.7009308 05	1.4400433 24	...	0.033389

Tabel 3 menunjukkan hasil transformasi data setelah melalui proses standarisasi menggunakan metode StandardScaler. Nilai pada tabel ini diubah sedemikian rupa sehingga memiliki rata-rata 0 dan standar deviasi 1. Proses ini bertujuan untuk memastikan bahwa semua atribut kompetensi memiliki bobot yang seimbang dalam analisis clustering, sehingga tidak ada atribut yang mendominasi hasil karena memiliki skala yang lebih besar.

Adapun proses menghitung pre-processing pada data seperti tabel menggunakan rumus standar scaler seperti persamaan berikut :

$$Z_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

Dimana awalnya adalah menghitung rata-rata tiap column dengan contoh column “C1” atau “Menguasai pengetahuan teknis terkait pekerjaan” yaitu :

$$\mu = \frac{5 + 4 + 4 + 3 + 1 + 1 + 3 + 5 + 4 + 4 + 2 + 4 + 2 + 2 + 1 + \dots + 1}{700}$$

$$\mu = 2.905714286$$

Selanjutnya adalah menghitung variance (S^2) :

$$S^2 = \frac{(5 - 2.905714286) + (4 - 2.905714286) + (4 - 2.905714286) + \dots + (1 - 2.905714286)}{700}$$

$$S^2 = 1.409033366$$

Langkah terakhir adalah menghitung Standard Scaler

$$x_i = 5$$

$$Z_i = \frac{5 - 2.905714286}{1.409033366} = 1.487391$$

$$x_i = 4$$

$$Z_i = \frac{4 - 2.905714286}{1.409033366} = 0.777176902$$

$$x_i = 4$$

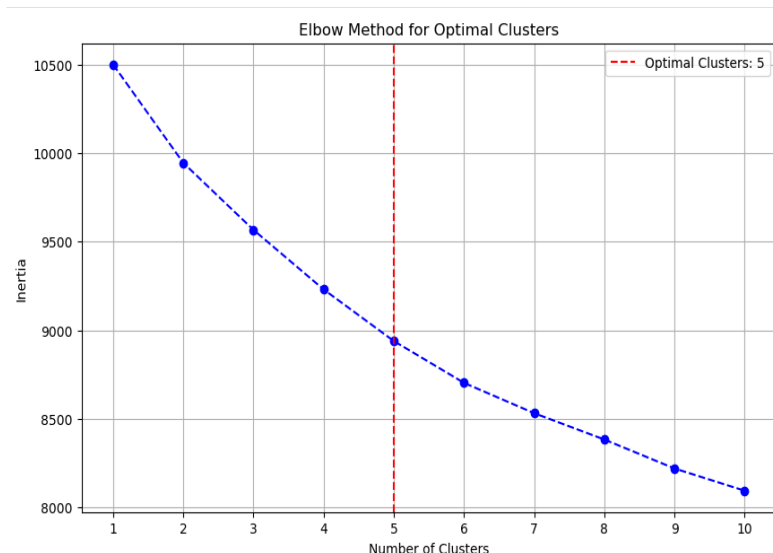
$$Z_i = \frac{5 - 2.905714286}{1.409033366} = 0.777176902$$

Proses perhitungan dilakukan secara terus-menerus sampai “C12” dan hasilnya ada pada tabel 3 diatas.

3.3 Menggunakan Metode Elbow untuk menentukan K terbaik

Dalam proses analisis clustering, langkah penting adalah menentukan jumlah cluster optimal (k) yang akan digunakan. Untuk tujuan ini, digunakan Elbow Method, yaitu metode evaluasi yang mengukur total inertia atau jumlah kuadrat jarak setiap titik data ke centroid cluster-nya. Proses ini melibatkan perhitungan nilai inertia untuk berbagai jumlah cluster, dari k=1 hingga k=10.

Hasilnya divisualisasikan dalam bentuk grafik, di mana sumbu horizontal menunjukkan jumlah cluster (k) dan sumbu vertikal menunjukkan nilai inertia. Titik siku atau "elbow" pada grafik menjadi indikator jumlah cluster optimal, karena pada titik tersebut terjadi penurunan signifikan dalam inertia sebelum akhirnya menjadi lebih lambat. Dalam penelitian ini, jumlah cluster optimal ditentukan dengan bantuan pustaka KneeLocator untuk memastikan akurasi dalam mengidentifikasi titik elbow. Hasil visualisasi dengan bahasa python didapat seperti pada gambar 4 dibawah.



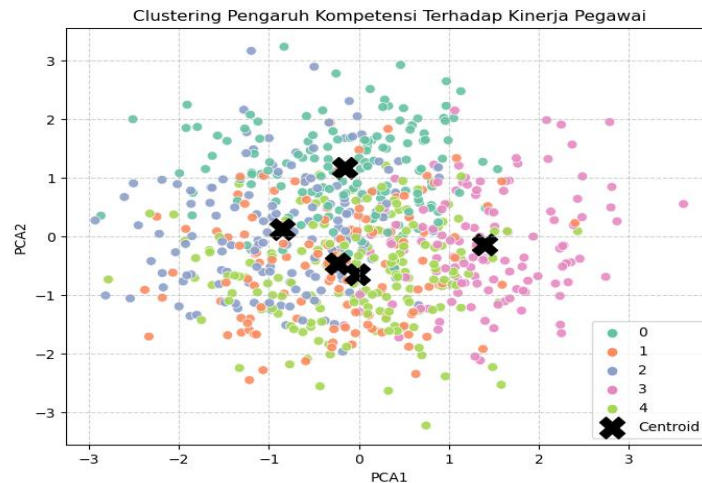
Gambar 4. Metode Elbow Untuk Menentukan K Optimal

Gambar 4 menunjukkan hasil metode Elbow untuk menentukan jumlah kluster optimal dalam algoritma K-Means. Grafik memplot nilai inertia (jumlah jarak kuadrat antara setiap titik data dengan pusat kluster) terhadap jumlah kluster. Dari grafik, terlihat penurunan inertia yang signifikan saat jumlah kluster bertambah, namun mulai melambat setelah kluster ke-5. Titik "elbow" pada grafik, yang ditandai dengan garis vertikal merah putus-putus, menunjukkan bahwa jumlah kluster optimal adalah 5. Ini adalah titik di mana penambahan kluster tidak lagi menghasilkan pengurangan inertia yang signifikan, menandakan keseimbangan antara kompleksitas model dan kualitas kluster.

3.4 Penerapan K-Means Clustering

Pada penelitian ini, algoritma K-Means Clustering diterapkan dengan menggunakan inisialisasi K-Means++ untuk memastikan pemilihan centroid awal yang lebih baik, sehingga meningkatkan konvergensi algoritma. Jumlah cluster yang digunakan berdasarkan hasil dari Elbow Method, yang menunjukkan nilai optimal. Setelah model K-Means diterapkan, data dikelompokkan ke dalam cluster sesuai dengan pola kompetensi pegawai. Hasil clustering ini kemudian ditambahkan ke dalam dataset asli sebagai kolom baru bernama "Cluster". Penambahan ini mempermudah analisis dan visualisasi lebih lanjut untuk memahami bagaimana kompetensi pegawai terdistribusi dalam setiap kelompok.

Untuk memahami distribusi cluster dalam ruang data yang lebih sederhana, digunakan metode Principal Component Analysis (PCA) untuk mereduksi dimensi data ke dalam dua komponen utama. Visualisasi PCA memberikan gambaran bagaimana data terdistribusi dalam dua dimensi dengan mempertahankan sebanyak mungkin informasi dari dataset asli. Dalam scatter plot hasil PCA, setiap titik mewakili seorang pegawai, dan warna menunjukkan cluster tempat pegawai tersebut tergabung. Centroid setiap cluster juga divisualisasikan dalam ruang PCA untuk menyoroti posisi pusat dari masing-masing kelompok. Visualisasi ini membantu mengidentifikasi pola distribusi, seperti kedekatan atau keterpisahan antar-cluster, yang memberikan wawasan lebih lanjut mengenai kesesuaian clustering yang telah dilakukan. Hasil visualisasi dapat dilihat pada gambar 5 dibawah ini.



Gambar 5. Visualisasi Clustering

Pada gambar 5, setiap cluster diberi nama atau label deskriptif berdasarkan analisis rata-rata kompetensi, seperti "Cluster A: Karyawan Berprestasi Tinggi" atau "Cluster B: Karyawan Berpotensi Berkembang". Penamaan ini bertujuan untuk memberikan konteks yang lebih mudah dipahami oleh pembaca atau pengambil keputusan. Selain itu, jarak Euclidean dihitung untuk setiap pegawai terhadap centroid cluster-nya. Jarak ini menunjukkan seberapa dekat atau representatif seorang pegawai terhadap cluster tempat ia berada. Pegawai dengan jarak yang lebih kecil dianggap lebih konsisten dengan karakteristik cluster, sementara pegawai dengan jarak yang lebih besar dapat menjadi perhatian untuk analisis lebih lanjut. Penamaan setiap cluster yaitu :

- 1) C0 : Cluster A: Karyawan Berprestasi Tinggi
- 2) C1 : Cluster B: Karyawan Berprestasi Rendah
- 3) C2 : Cluster C: Karyawan yang Berkembang
- 4) C3 : Cluster D: Karyawan Berprestasi Tidak Konsisten
- 5) C4 : Cluster E: Karyawan Berpotensi Tinggi

Pada visualisasi hasil clustering menggunakan PCA, seperti terlihat pada Gambar 4.2, data terbagi menjadi 5 cluster berdasarkan kompetensi pegawai. Cluster 0, yang ditandai dengan warna hijau muda, mewakili kelompok dengan tingkat kompetensi tertentu, sedangkan cluster 1, berwarna oranye, menunjukkan kelompok dengan nilai dan berat kompetensi yang lebih tinggi. Cluster 2, yang berwarna biru, cluster 3, yang berwarna merah jambu, dan cluster 4, yang berwarna hijau, masing-masing mewakili kelompok dengan karakteristik kompetensi yang beragam. Titik centroid dari setiap cluster divisualisasikan dengan simbol silang hitam untuk menandai pusat setiap kelompok. Berikut adalah proses perhitungan manual K-Means, dimana centroid data awal ditentukan pada Tabel 4 berikut.

Tabel 4. Centroid Data Awal

Cluster	C1	C2	C3	...	C15
[6] SRI SUSANTI (C0)	-1.35346	0.663591	-1.35772	...	1.44991



Cluster	C1	C2	C3	...	C15
[25] KIKI SYAHPUTRA (C1)	-0.64325	1.382653	1.369403	...	-1.38313
[78] MANIUR ROSPITA HUTAPEA (C2)	1.487391	-0.05547	0.687623	...	-0.674871
[147] EFENDI SIRAIT (C3)	0.777177	-0.77453	-1.35772	...	0.741649
[266] DANIEL ASMON (C4)	0.066963	1.382653	0.687623	...	1.44991

Kemudian menghitung jarak setiap data menggunakan Euclidean Distance:

$$\begin{aligned}
 Data_{(0,6)} &= \sqrt{(1.48739078 - (-1.353464735))^2 + (1.382653272 - 0.663591392)^2 + (1.369402859 - (-1.35771521))^2 + \dots + (1.449909581 - 1.449909581)^2} \\
 &= 5.3013 \\
 Data_{(1,25)} &= \sqrt{(0.777176902 - (-0.643250856))^2 + (1.382653272 - 1.382653272)^2 + (1.369402859 - 1.369402859)^2 + \dots + (1.449909581 - (-1.383130772))^2} \\
 &= 3.1692 \\
 \dots \\
 Data_{(2,78)} &= \sqrt{(0.777176902 - 1.48739078)^2 + (-1.493594248 - (-0.055470488))^2 + (1.369402859 - 0.687623342)^2 + \dots + (-0.674870684 - (-0.674870684))^2} \\
 &= 2.1898
 \end{aligned}$$

Tabel 5 menyajikan hasil perhitungan jarak setiap pegawai terhadap pusat cluster (centroid) pada iterasi pertama. Data ini digunakan untuk menentukan cluster terdekat berdasarkan jarak minimum. Kolom 'Jarak ke Centroid' menunjukkan jarak setiap data pegawai ke masing-masing pusat cluster (C0, C1, C2, dst.), sementara kolom 'Jarak Terdekat' mengidentifikasi cluster yang memiliki jarak paling kecil, yang menjadi hasil pengelompokan awal pada iterasi pertama.

Tabel 5. Hasil Perhitungan Jarak Pusat Cluster Iterasi Pertama

No	Nama Pegawai	Jarak ke Centroid					Jarak Terdekat
		C0	C1	C2	C3	C4	
1	MUSA MANURUNG	6.096378737	5.626781933	6.74908203	6.465318015	4.899987701	C4
2	YOWAN ALFITRA	5.431270206	5.669169851	6.081000907	6.498709848	4.181989194	C4
...
700	ANDIKA	6.118048241	5.400112962	5.910756337	5.389682325	5.111746801	C4

Data pada Tabel 5 menunjukkan bahwa setiap pegawai dikelompokkan ke cluster terdekat berdasarkan jarak minimum. Pada iterasi berikutnya, nilai centroid akan diperbarui berdasarkan rata-rata posisi data yang tergabung dalam setiap cluster, hingga algoritma mencapai konvergensi. Langkah selanjutnya adalah menentukan centroid baru untuk iterasi selanjutnya dengan menghitung rata-rata (mean) pada masing-masing cluster pada Tabel 6 berikut.

Tabel 6. Centroid Data Baru

Cluster	C1	C2	C3	...	C15
[6] SRI SUSANTI (C0)	-0.5234558	0.117797435	-0.4623662	...	0.229654248
[25] KIKI SYAHPUTRA (C1)	-0.36363909	0.43145330	0.59636152	...	-0.60237162
[78] MANIUR ROSPITA HUTAPEA (C2)	0.40714950	-0.04338541	0.4412660	...	-0.4606071

Cluster	C1	C2	C3	...	C15
[147] EFENDI SIRAIT (C3)	0.12324412	-0.3023937	-0.41608765	...	0.20711357
[266] DANIEL ASMON (C4)	0.08036329	0.19552281	0.192368409	...	0.541198524

Ulangi langkah 3-4 hingga posisi data sudah tidak mengalami perubahan. Sehingga hasil akhir posisi cluster terdapat pada Tabel 7.

Tabel 7. Hasil Pengelompokan Clustering K-Means

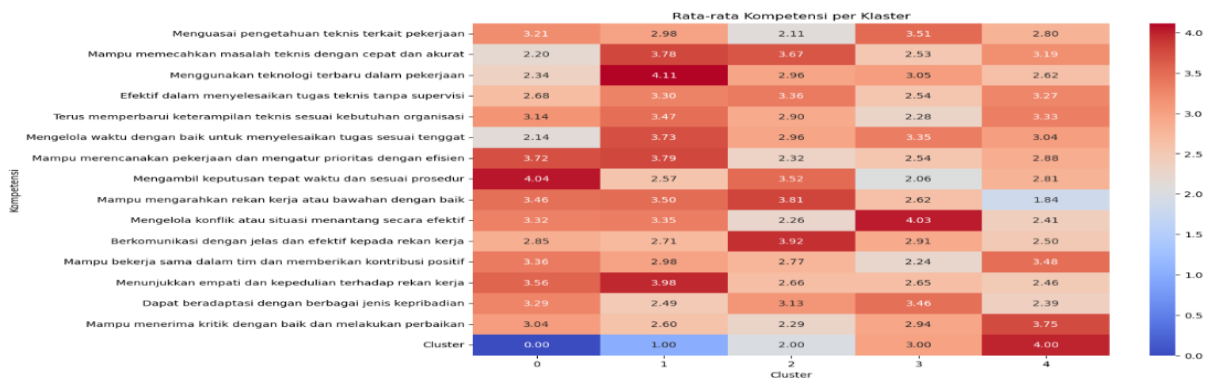
No.	C1	C2	C3	...	C15	Cluster	Nama Cluster
1	1.487391	1.382653	1.369403	...	1.449910	1	Cluster B: Karyawan Berprestasi Rendah
2	0.777176902	1.382653272	1.369402859	...	1.449909581	1	Cluster B: Karyawan Berprestasi Rendah
...		
700	-1.353464735	0.663591392	-0.675935693		0.033389	2	Cluster C: Karyawan yang Berkembang

Setelah hasil clustering diperoleh, rata-rata nilai dari setiap variabel kompetensi dihitung untuk masing-masing cluster. Analisis ini memberikan wawasan mengenai karakteristik utama setiap kelompok, seperti cluster dengan kompetensi teknis yang lebih dominan atau cluster dengan kelemahan tertentu dalam kompetensi sosial. Ringkasan rata-rata setiap variabel kompetensi per cluster ditampilkan dalam Tabel 8 berikut, yang menggambarkan perbedaan signifikan di antara cluster dan memberikan dasar untuk interpretasi lebih lanjut.

Tabel 8. Ringkasan Hasil Cluster

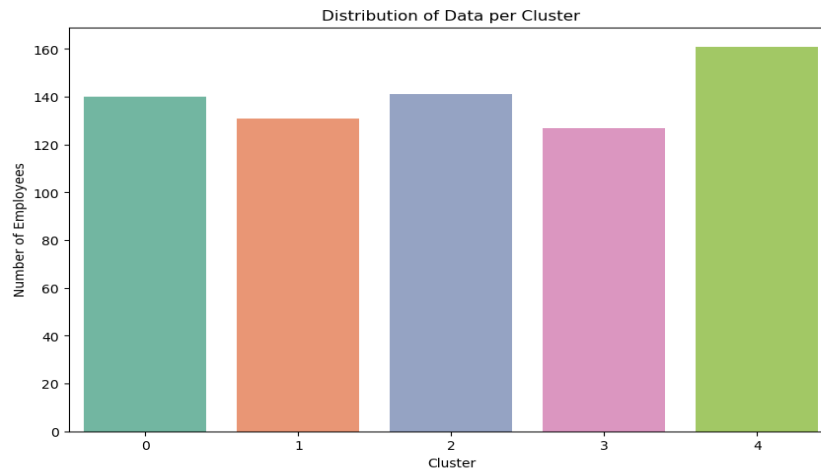
Cluster	C1	C2	C3	...	C15
0	3.214285714	2.2	2.342857143	...	3.042857143
1	2.977099237	3.778625954	4.114503817	...	2.603053435
2	2.106382979	3.666666667	2.964539007	...	2.290780142
3	3.511811024	2.527559055	3.047244094	...	2.94488189
4	2.801242236	3.186335404	2.621118012	...	3.745341615

Untuk memberikan pemahaman visual yang lebih jelas, rata-rata kompetensi per cluster divisualisasikan dalam bentuk heatmap. Dalam heatmap ini, setiap baris mewakili variabel kompetensi, dan setiap kolom mewakili cluster. Warna pada heatmap menunjukkan nilai rata-rata, dengan gradasi yang lebih gelap menunjukkan nilai yang lebih tinggi. Visualisasi ini membantu mengidentifikasi pola yang mencolok, seperti cluster dengan skor tinggi dalam kompetensi teknis atau cluster yang memiliki keseimbangan antara kompetensi teknis, manajerial, dan sosial. Heatmap memberikan gambaran intuitif mengenai distribusi kompetensi di antara cluster. Gambar 6 menunjukkan heatmap rata-rata skor kompetensi untuk setiap cluster yang dihasilkan dari analisis clustering. Heatmap ini memberikan gambaran intuitif tentang distribusi kompetensi antar cluster dan memudahkan analisis pola pada masing-masing kelompok.



Gambar 6. Rata-rata Kompetensi Per Klaster

Gambar 6 adalah heatmap yang menampilkan rata-rata skor kompetensi untuk masing-masing cluster hasil analisis clustering. Sebagai contoh, kompetensi "Menggunakan teknologi terbaru dalam pekerjaan" memiliki skor rata-rata tertinggi di Cluster 1 (4.11), menunjukkan bahwa anggota cluster ini unggul dalam kompetensi tersebut dibandingkan cluster lain. Sebaliknya, kompetensi "Menguasai pengetahuan teknis terkait pekerjaan" memiliki skor rata-rata tertinggi di Cluster 3 (3.51). Kompetensi "Mengelola konflik atau situasi menantang secara efektif" memiliki skor tertinggi di Cluster 4 (4.03), sedangkan Cluster 0 dan Cluster 1 menunjukkan nilai rendah (masing-masing 3.32 dan 3.26). Skor rendah juga terlihat pada kompetensi "Mampu menerima kritik dengan baik dan melakukan perbaikan," di mana Cluster 0 hanya mencapai skor rata-rata 2.00. Heatmap ini membantu dalam memahami pola distribusi kompetensi di setiap cluster. Dengan mengidentifikasi kompetensi unggulan atau rendah pada cluster tertentu, hasil ini dapat digunakan untuk mendukung pengambilan keputusan terkait pengembangan pegawai. Adapun distribusi data dalam masing-masing cluster yaitu :



Gambar 7. Distribusi Data

Dari hasil visualisasi pada Gambar 7 menunjukkan distribusi jumlah karyawan dalam setiap cluster hasil analisis clustering. Cluster 4 memiliki jumlah karyawan terbanyak, yaitu sekitar 160 orang, diikuti oleh Cluster 2 dan Cluster 0, masing-masing dengan jumlah sekitar 145 orang. Cluster 1 memiliki jumlah karyawan yang sedikit lebih rendah, yaitu sekitar 135 orang, sedangkan Cluster 3 memiliki jumlah karyawan paling sedikit, yaitu sekitar 125 orang. Grafik ini membantu untuk memahami bagaimana data karyawan terdistribusi ke dalam masing-masing cluster, yang dapat memberikan gambaran ukuran relatif setiap kelompok dalam analisis ini.

4. KESIMPULAN

Penelitian ini berhasil mengelompokkan 700 data pegawai dengan 15 atribut kompetensi ke dalam lima cluster optimal menggunakan algoritma K-Means Clustering, yang ditentukan melalui metode Elbow. Proses pengolahan data meliputi transformasi dan normalisasi dengan StandardScaler untuk memastikan keseragaman skala. Hasil clustering menunjukkan pembagian kompetensi pegawai berdasarkan aspek teknis, manajerial, dan sosial dengan karakteristik unik di setiap cluster. Penelitian ini memberikan gambaran mengenai potensi pengembangan pegawai di setiap kelompok dan dapat menjadi dasar untuk strategi pengelolaan sumber daya manusia. Penelitian selanjutnya disarankan untuk memperluas cakupan dataset dan menambahkan atribut baru guna meningkatkan generalisasi dan nilai analisis. Proses clustering menghasilkan lima kelompok dengan karakteristik yaitu dimana Cluster 0 terdiri dari 145 pegawai yang memiliki kompetensi teknis tinggi, kemudian cluster 1 dengan jumlah 160 pegawai merupakan kelompok yang menonjol dalam kompetensi sosial dan manajerial, lalu cluster 2, yang mencakup 125 pegawai, menunjukkan pegawai dengan kompetensi rata-rata pada semua aspek, dan cluster 3 dengan 135 pegawai memiliki kompetensi teknis sedang namun unggul di kompetensi social, serta cluster 4, yang juga berjumlah 135 pegawai, terdiri atas pegawai yang memiliki potensi besar untuk berkembang. Untuk penelitian selanjutnya, disarankan menggunakan dataset yang lebih beragam dan mencakup berbagai sektor atau instansi untuk meningkatkan generalisasi hasil. Selain itu, penting untuk mengintegrasikan atribut tambahan, seperti tingkat pendidikan atau pengalaman kerja, guna memperkaya analisis kompetensi. Implementasi dashboard interaktif juga disarankan agar hasil clustering dapat divisualisasikan dengan lebih jelas dan mendukung pengambilan keputusan secara real-time.

REFERENCES

- [1] I. Buil, E. Martínez, and J. Matute, "Transformational Leadership and Employee Performance: The Role of Identification, Engagement, and Proactive Personality," *International Journal of Hospitality Management*, vol. 77, pp. 64–75, 2019, doi: 10.1016/j.ijhm.2018.06.014.



- [2] K. Sendawula, S. N. Kimuli, J. Bananuka, and G. N. Muganga, “Training, Employee Engagement and Employee Performance: Evidence from Uganda’s Health Sector,” *Cogent Business and Management*, vol. 5, no. 1, pp. 1–12, 2018, doi: 10.1080/23311975.2018.1470891.
- [3] D. Marcelina, A. Kurnia, and T. Terttiaavini, “Analisis Klaster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering,” *MALCOM Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 293–301, 2023, doi: 10.57152/malcom.v3i2.952.
- [4] J. Faran and A. Triayudi, “Penerapan Algoritma K-Means Data Mining untuk Clustering Kinerja Karyawan Koperasi,” *KLIK Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 4, pp. 2096–2108, 2024, doi: 10.30865/klik.v4i4.1728.
- [5] Mudakir, A. T. Zy, and A. S. Sunge, “Penerapan Data Mining untuk Klasifikasi Pengangkatan Karyawan Menggunakan Algoritma K-Means,” *Jurnal Informatika Teknologi dan Sains*, vol. 5, no. 3, pp. 489–497, 2023, doi: 10.51401/jinteks.v5i3.3369.
- [6] I. Cholissodin and A. A. Soebroto, *AI, Machine Learning & Deep Learning (Teori & Implementasi)*, 2021.
- [7] T. Wahyono, *Fundamental of Python for Machine Learning: Dasar-Dasar Pemrograman Python untuk Machine Learning dan Kecerdasan Buatan*, Revisi ed., 2021.
- [8] A. Septhiani and H. Hendry, “Analisis Perbandingan Algoritma Supervised Learning untuk Prediksi Kasus Covid-19 di Jakarta,” *Jurnal Sains Komputer dan Informatika*, vol. 7, no. 2, pp. 583–594, 2023, [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jsakti/article/view/668/643>.
- [9] A. Supriyadi, A. Triayudi, and I. D. Sholihati, “Perbandingan Algoritma K-Means dengan K-Medoids pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas,” *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 6, no. 2, pp. 229–240, 2021, doi: 10.29100/jupi.v6i2.2008.
- [10] I. S. Tinendung and I. Zufria, “Pengelompokan Status Stunting pada Anak Menggunakan Metode K-Means Clustering,” *Media Informatika Budidarma*, vol. 7, pp. 2014–2023, 2023, doi: 10.30865/mib.v7i4.6908.
- [11] W. Rahmansyah, I. Zufria, and M. Fakhriza, “Sistem Pendukung Keputusan Pemilihan SSD Laptop Menggunakan Kombinasi Metode AHP dan SAW,” *KLIK Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 2, pp. 1192–1199, 2023, doi: 10.30865/klik.v4i2.1257.
- [12] P. Pangestu, S. Maarip, Y. N. Addinsyah, and V. Purwayoga, “Clustering and Trend Analysis of Priority Commodities in the Archipelago Capital Region (IKN) Using a Data Mining Approach,” *International Journal of Applied Sciences and Smart Technology*, vol. 6, no. 1, pp. 169–182, 2024, doi: 10.24071/ijasst.v6i1.7798.
- [13] M. S. Hasibuan, A. H. Lubis, and M. N. Sari, “Perbandingan Algoritma Clustering DBSCAN dan K-Means dalam Pengelompokan Siswa Terbaik,” *Jurnal Ilmu Komputer dan Informatika*, vol. 5, pp. 301–309, 2024, doi : 10.37373/infotech.v5i2.1457
- [14] J. A. Noyari, A. Aprillia, R. G. Munthe, and A. Sutarmanto, “Optimasi Kinerja Sistem Informasi Manajemen Kampus Menggunakan Teknik Data Mining,” *Jurnal Informatika dan Sistem Informasi*, vol. 3, no. 1, pp. 52–63, 2024, doi : 10.33050/mentari.v3i1.613
- [15] M. Y. Nurzaman and B. N. Sari, “Implementasi K-Means Clustering dalam Pengelompokan Banyaknya Jumlah Petani Berdasarkan Kecamatan di Provinsi Jawa Barat,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, no. 3, pp. 131–144, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>.
- [16] S. Mutiah, Y. Hasnataeni, A. Fitrianto, and L. M. R. D. Jumansyah, “Perbandingan Metode Klastering K-Means dan DBSCAN dalam Identifikasi Kelompok Rumah Tangga Berdasarkan Fasilitas Sosial Ekonomi di Jawa Barat,” *Teorema: Teori dan Riset Matematika*, vol. 9, no. 2, pp. 247–260, 2024, [Online]. Available: <http://dx.doi.org/10.25157/teorema.v9i2.16290>.
- [17] S. Xia et al., “Ball K-Means: Fast Adaptive Clustering with No Bounds,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 87–99, 2022, doi: 10.1109/TPAMI.2020.3008694.
- [18] H. A. Ulvi and M. Ikhsan, “Comparison of K-Means and K-Medoids Clustering Algorithms for Export and Import Grouping of Goods in Indonesia,” *Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 3, pp. 1641–1655, 2024, [Online]. Available: <https://doi.org/10.33395/sinkron.v8i3.13815>.
- [19] D. Arbain, Sriyanto, and J. Triloka, “Perbandingan Kinerja Algoritma K-Medoids dan K-Means untuk Klasifikasi Penyakit Kanker Serviks,” *Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat*, no. 93, pp. 118–131, 2023, doi : 10.30865/mib.v5i1.2547.
- [20] M. Sulaiman, R. Yudistira, R. Narasati, and R. Herdiana, “Penerapan Data Mining dengan Metode Clustering untuk Menentukan Strategi Peningkatan Penjualan Berdasarkan Data Transaksi,” *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 6, no. 1, 2024, doi : 10.36499/jinrpl.v6i1.10329
- [21] M. Ikhsan, A. Armansyah, and A. A. Tamba, “Implementasi Jaringan Syaraf Tiruan Backpropagation pada Klasifikasi Grade Teh Hitam,” *Jurnal Sistem Komputer dan Informatika*, vol. 4, no. 2, p. 387, 2022, doi: 10.30865/json.v4i2.5312.
- [22] R. R. Asyrofi and R. Asyrofi, “Implementasi Aplikasi Jupyter Notebook sebagai Analisis Kriteria Plagiasi dengan Teknik Semantik,” *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 2, pp. 627–637, 2023, doi: 10.29100/jupi.v8i2.3699.
- [23] M. G. S. Finka, “Implementasi K-Nearest Neighbor (KNN) untuk Klasifikasi Citra Serat Kayu,” *Jurnal Ilmu Komputer dan Aplikasi*, 2023, doi : 10.56854/jt.v2i3.331.