

Analisis Pengelompokan Data Karyawan Terbaik Perusahaan Menggunakan Algoritma K-Means Clustering

Kharisma Silvi Amalia¹, Arum Purbaningrum², Rajnaparamitha Kusumastuti*³

^{1,2,3}STMIK AMIKOM Surakarta

^{1,2,3}Sukoharjo, Indonesia

Email: ¹kharisma.10307@mhs.amikomsolo.ac.id,

²arum.10308@mhs.amikomsolo.ac.id, ³rajna.paramitha@gmail.com

Abstract

This research implemented the K-Means Clustering algorithm on a dataset encompassing a diverse array of employee-related attributes within a company. This study aimed to identify significant patterns and relationships within this data. Moreover, the research included integrating employee sentiment data, providing in-depth insights into the moods and viewpoints of employees on various aspects of the company. The findings of this study have a positive impact in the context of human resource management and decision-making. Utilizing a value of $k=3$, with groups of best, good, and adequate employees, it was found that the best employee group consisted of individuals in the "Director" position within the IT department, working in the Manufacturing business unit. The evaluation using the silhouette coefficient yielded a score of 0.597. A value of 1 indicates well-defined clusters with a clear separation between clusters and similar members within clusters, suggesting a reasonably good clustering quality. The information from this research provides valuable insights into human resource management, assisting them in strategic decision-making related to workforce planning, human resource management, and efforts to enhance efficiency and employee retention.

Keywords: K-Means Clustering, Company Employees, Management Decisions, Data Clustering

Abstraksi

Penelitian ini mengimplementasikan algoritma K-Means Clustering pada dataset yang mencakup beragam atribut terkait karyawan dalam suatu perusahaan. Tujuan dari penelitian ini adalah untuk mengidentifikasi pola dan hubungan yang signifikan dalam data tersebut. Selain itu, penelitian ini mencakup integrasi data sentimen karyawan, yang memberikan wawasan mendalam mengenai suasana hati dan pandangan karyawan terhadap berbagai aspek perusahaan. Hasil dari penelitian ini memiliki dampak positif dalam konteks manajemen sumber daya manusia dan pengambilan keputusan. Dengan menggunakan nilai $k=3$, dengan kelompok pegawai terbaik, baik dan cukup mendapatkan hasil bahwa kelompok pegawai terbaik terdiri dari kumpulan posisi jabatan "Direktur" di departemen IT, bekerja di unit bisnis Manufacturing. Nilai evaluasi menggunakan silhouette coefficient diperoleh sebesar 0.597. Nilai yang mendekati 1 menunjukkan kluster yang baik dengan pemisahan yang jelas antara cluster dan anggota cluster yang serupa dan menunjukkan kualitas cluster yang cukup baik. Informasi yang diperoleh dari penelitian ini memberikan wawasan berharga kepada manajemen sumber daya manusia, membantu mereka dalam pengambilan keputusan strategis terkait dengan perencanaan

tenaga kerja, pengelolaan sumber daya manusia, serta upaya untuk meningkatkan efisiensi dan retensi karyawan.

Kata Kunci: *K-Means Clustering, Karyawan Perusahaan, Matriks Validasi, Keputusan Manajemen, Pengelompokan Data, Sentimen Karyawan*

1. PENDAHULUAN

Dalam era digitalisasi yang pesat, Teknologi Informasi (TI) memainkan peran sentral dalam transformasi perusahaan modern. Pengelolaan data yang kompleks mendorong perusahaan untuk mengatasi tantangan baru dalam pemahaman dan pengelolaan data besar. Guna mencapai efisiensi dalam pengelolaan data, penerapan berbagai algoritma dapat digunakan untuk membantu pengelompokan data dengan lebih mudah. Salah satu metode analisis data dalam mengidentifikasi pola dalam data adalah menggunakan algoritma K-Means Clustering.

Penelitian ini menerapkan Algoritma K-Means Clustering menggunakan data set karyawan perusahaan, dengan membagi mereka berdasarkan karakteristik dan perilaku. Pengembangan (TI) sangat penting untuk mencapai efisiensi dan inovasi dalam bisnis, sehingga analisis ini bertujuan menjelaskan hubungan antara perkembangan TI dan pengelompokan karyawan berdasarkan karakteristik pribadinya.

Data *mining* adalah proses ekstraksi pengetahuan atau pola yang berharga dari kumpulan data besar atau kompleks. Ini melibatkan penggunaan berbagai teknik analisis statistik, matematika, dan kecerdasan buatan untuk mengidentifikasi hubungan, tren, atau pola yang mungkin tidak terlihat secara langsung. Data *mining* membantu dalam pengambilan keputusan, prediksi, dan pemahaman yang lebih baik tentang data. Data *mining* adalah penemuan informasi baru dengan memeriksa pola ataupun aturan tertentu dalam data jumlah besar [1].

Penelitian serupa [2] dalam pengelompokan prestasi kerja pegawai menunjukkan hasil pengolahan data yang dapat menjadi acuan perusahaan ke depannya untuk menentukan kenaikan jabatan atau pangkat, sementara pegawai dengan kinerja kurang mendapatkan pelatihan. Hal tersebut menunjukkan penerapan model tersebut dapat membantu perusahaan dalam mengambil sebuah keputusan terhadap manajemen bisnis. Penelitian sebelumnya [3], telah mengaplikasikan metode Clustering menggunakan algoritma K-Means pada analisis pengelompokan data calon mahasiswa baru di universitas Muhammadiyah Yogyakarta berdasarkan nilai tes masuk, prestasi akademik, dan minat jurusan. Penelitian lainnya [4], membahas bagaimana data *mining* clustering dapat digunakan untuk mengelompokkan buku berdasarkan kategori, penulis, dan penerbit. Paper ini menggunakan studi kasus pada salah satu perpustakaan yang ada di kota Batam.

Algoritma serupa pada penelitian [5] dengan hasil penelitian yang menunjukkan bahwa algoritma K-means Clustering efektif dalam mengelompokkan siswa berdasarkan performa akademik mereka. Kemudian pada penelitian [6] berhasil menunjukkan bagaimana pemanfaatan data yang ada dapat secara signifikan meningkatkan efektivitas

promosi pendidikan, serta memungkinkan universitas untuk menargetkan calon mahasiswa potensial dengan cara yang lebih strategis dan personal.

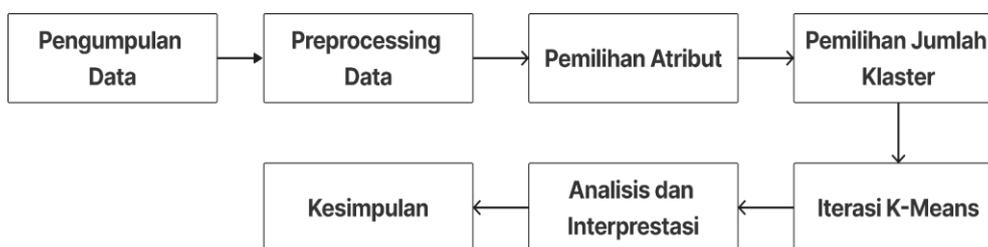
Dalam serangkaian penelitian yang berfokus pada penerapan algoritma K-Means Clustering dalam berbagai bidang, para peneliti [7],[8],[9][10]telah menunjukkan hasil yang signifikan. Penelitian [8] juga memfokuskan pada *clustering* data mahasiswa untuk promosi, sedangkan penelitian [10] menerapkannya dalam menentukan penerima bantuan bedah rumah. Di sisi lain, penelitian [7] dan [9] menggunakan Naïve Bayes dalam analisis Intrusion Detection System, dan dalam visualisasi data *mining*. Hasil-hasil ini menegaskan bahwa kedua algoritma tersebut sangat efektif dalam mengolah dan menganalisis data dalam berbagai konteks, mulai dari pendidikan, pemasaran, hingga keamanan dan bantuan sosial, menawarkan wawasan yang lebih dalam dan strategi yang lebih terarah berdasarkan analisis data.

Pada penelitian dengan topik yang sama tentang *clustering* [11] berhasil menampilkan data prediksi kelulusan mahasiswa dengan fokus pada evaluasi efektivitas K-Means *clustering* dalam meramalkan waktu kelulusan, dengan mempertimbangkan tingkat kemampuan individu mahasiswa. Pengolahan data pada penelitian ini menggunakan algoritma K-Means *Clustering* dengan perangkat lunak data *mining* WEKA. Penggunaan WEKA sendiri pada penelitian sebelumnya [12] menunjukkan bahwa antara hasil perhitungan manual dengan perhitungan menggunakan WEKA memiliki nilai yang sama. Hal tersebut menunjukkan bahwa penggunaan WEKA untuk membantu proses pengolahan data bisa dijadikan acuan untuk menarik keputusan sesuai kebutuhan bisnis

Penggunaan data *mining* dalam penelitian ini digunakan untuk pengolahan dan pengelompokan data dari data set yang kompleks, sehingga dapat membantu bisnis dalam memahami tren pasar, dan membuat keputusan yang didasarkan pada data yang lebih baik. Analisis ini memberikan gambaran tentang bagaimana pengelompokan karyawan berdasarkan karakteristik pribadi mereka dapat membantu mengidentifikasi bakat yang ada dalam perusahaan dan merancang program pengembangan yang lebih efektif. Hal ini dapat berdampak positif pada peningkatan produktivitas dan kesejahteraan karyawan dalam lingkungan kerja yang semakin kompetitif

2. METODE PENELITIAN

Penelitian ini memiliki beberapa tahap alur penelitian yang dimulai dari tahapan pengumpulan data, tahapan *preprocessing* data, pemilihan atribut, pemilihan jumlah klaster, pengolahan berdasarkan iterasi K-Means, analisis dan interpretasi, serta kesimpulan yang dapat diperhatikan pada gambar 1 berikut.



Gambar 1. Alur Penelitian

Berdasarkan gambar 1, data set yang digunakan merupakan data set *public* dari website *thespreadsheetguru.com* yang menyediakan informasi berkaitan dengan data karyawan. Alur penelitian dimulai dengan pengumpulan data karyawan perusahaan dan dilanjutkan dengan tahap *preprocessing* untuk membersihkan dan mengatasi *missing values*. Kemudian, atribut yang relevan dipilih untuk analisis. Selanjutnya, jumlah kluster ditentukan dan dilakukan iterasi *K-Means* untuk mengelompokkan data. Hasil *clustering* dianalisis untuk memahami karakteristik masing-masing kluster.

Algoritma *K-Means Clustering* sendiri digunakan dalam pengelompokan karyawan berdasarkan kesamaan atribut-atribut tersebut, dengan tujuan mengidentifikasi pola dan struktur dalam data karyawan serta memahami karakteristik kelompok-kelompok karyawan yang terbentuk. Gambar 2 merupakan sampel data set yang digunakan dalam penelitian.

EEID	Full Name	Job Title	Department	Business Unit	Gender	Ethnicity	Age	Hire Date	Annual Salary	Bonus %	Country	City	Exit Date
E02387	Emily Davis	Sr. Manger	IT	Research & Development	Female	Black	55	08/04/2016	\$141.604	15%	United States	Seattle	16/10/2021
E04105	Theodore Dinh	Technical Architect	IT	Manufacturing	Male	Asian	59	29/11/1997	\$99.975	0%	China	Chongqing	
E02572	Luna Sanders	Director	Finance	Speciality Products	Female	Caucasian	50	26/10/2006	\$163.099	20%	United States	Chicago	
E02832	Penelope Jordan	Computer Systems Manager	IT	Manufacturing	Female	Caucasian	26	27/09/2019	\$84.913	7%	United States	Chicago	
E01639	Austin Vo	Sr. Analyst	Finance	Manufacturing	Male	Asian	55	20/11/1995	\$95.409	0%	United States	Phoenix	
E00644	Joshua Gupta	Account Representative	Sales	Corporate	Male	Asian	57	24/01/2017	\$50.994	0%	China	Chongqing	
E01550	Ruby Barnes	Manager	IT	Corporate	Female	Caucasian	27	01/07/2020	\$119.746	10%	United States	Phoenix	
E04332	Luke Martin	Analyst	Finance	Manufacturing	Male	Black	25	16/05/2020	\$41.336	0%	United States	Miami	20/05/2021
E04533	Easton Bailey	Manager	Accounting	Manufacturing	Male	Caucasian	29	25/01/2019	\$113.527	6%	United States	Austin	
E03838	Madeline Walker	Sr. Analyst	Finance	Speciality Products	Female	Caucasian	34	13/06/2018	\$77.203	0%	United States	Chicago	
E00591	Savannah Ali	Sr. Manger	Human Resources	Manufacturing	Female	Asian	36	11/02/2009	\$157.333	15%	United States	Miami	
E03344	Camila Rogers	Controls Engineer	Engineering	Speciality Products	Female	Caucasian	27	21/10/2021	\$109.851	0%	United States	Seattle	
E00530	Eli Jones	Manager	Human Resources	Manufacturing	Male	Caucasian	59	14/03/1999	\$105.086	9%	United States	Austin	
E04239	Everleigh Ng	Sr. Manger	Finance	Research & Development	Female	Asian	51	10/06/2021	\$146.742	10%	China	Shanghai	
E03496	Robert Yang	Sr. Analyst	Accounting	Speciality Products	Male	Asian	31	04/11/2017	\$97.078	0%	United States	Austin	09/03/2020
E00549	Isabella Xi	Vice President	Marketing	Research & Development	Female	Asian	41	13/03/2013	\$249.270	30%	United States	Seattle	
E00163	Bella Powell	Director	Finance	Research & Development	Female	Black	65	04/03/2002	\$175.837	20%	United States	Phoenix	
E00884	Camila Silva	Sr. Manger	Marketing	Speciality Products	Female	Latino	64	01/12/2003	\$154.828	13%	United States	Seattle	
E04116	David Barnes	Director	IT	Corporate	Male	Caucasian	64	03/11/2013	\$186.503	24%	United States	Columbus	
E04625	Adam Dang	Director	Sales	Research & Development	Male	Asian	45	09/07/2002	\$166.331	18%	China	Chongqing	
E03680	Elias Alvarado	Sr. Manger	IT	Manufacturing	Male	Latino	56	09/01/2012	\$146.140	10%	Brazil	Manaus	
E04732	Eva Rivera	Director	Sales	Manufacturing	Female	Latino	36	02/04/2021	\$151.703	21%	United States	Miami	
E03484	Logan Rivera	Director	IT	Research & Development	Male	Latino	59	24/05/2002	\$172.787	28%	Brazil	Rio de Janer	
E00671	Leonardo Dixon	Analyst	Sales	Speciality Products	Male	Caucasian	37	05/09/2019	\$49.998	0%	United States	Seattle	

Gambar 2. Sampel Data set

Berdasarkan gambar 2, data set yang digunakan adalah *The SpreadsheetGuru* yang berjumlah 1000 data mencakup atribut *Employee ID* (EEID), Nama lengkap (*Full Name*), Posisi pekerjaan (*Job Title*), Departemen (*Department*), Unit *business* (*Business Unit*), Jenis kelamin (*Gender*), Latar belakang etnis (*Ethnicity*), Usia (*Age*), Tanggal mulai bekerja (*Hire Date*), Pendapatan tahunan (*Annual Salary*), Percentage bonus (Bonus %), Negara (*Country*), Kota (*City*), dan Tanggal keluar (*Exit Date*). Sedangkan pada penelitian ini hanya menggunakan atribut Departemen, *Unit Business*, Pendapatan tahunan (*Annual Salary*), dan Percentage bonus (Bonus %).

Dalam penelitian ini, jumlah kluster yang digunakan adalah 3 untuk memastikan konsistensi hasil *clustering*. 3 Klaster ini terdiri kelompok Terbaik, Baik, dan Cukup. Algoritma K-Means Clustering sendiri digunakan dalam pengelompokan karyawan berdasarkan kesamaan atribut-atribut tersebut, dengan tujuan mengidentifikasi pola dan struktur dalam data karyawan serta memahami karakteristik kelompok-kelompok karyawan yang terbentuk.

2.1. Data Mining

Data *mining* adalah ekstraksi informasi berharga dan pola tersembunyi dari data besar. Tujuannya adalah mengidentifikasi hubungan, tren, dan pola untuk mendukung pengambilan keputusan. Metode penyelesaian melibatkan *clustering*, klasifikasi, dan *Association Rule Mining*. Data Mining dapat digunakan oleh perusahaan besar untuk menggali data dan mendapatkan informasi yang dapat menunjang serta meningkatkan proses bisnis perusahaan tersebut [13].

Data *mining* adalah suatu proses ekstraksi atau penggalian data dan informasi yang bersumber dari *database* yang besar. Proses ini bertujuan untuk mengidentifikasi pola atau pengetahuan yang sebelumnya tidak diketahui namun dapat dipahami dan bermanfaat. Informasi yang ditemukan melalui data *mining* dapat digunakan untuk membuat keputusan bisnis yang sangat penting. Dengan menganalisis data set yang luas, data *mining* membantu mengungkap wawasan baru dan hubungan yang dapat mendukung pengambilan keputusan strategis dalam berbagai bidang [14].

2.2. K-Means Clustering

K-Means Clustering adalah algoritma pengelompokan yang digunakan untuk membagi kumpulan data menjadi kelompok-kelompok yang serupa berdasarkan karakteristik tertentu. Pokok dari proses *clustering* adalah dengan menghitung jarak kedekatan antar tiap data untuk bisa dikategorikan sebagai kelompok atau kluster yang sama. *K-Means* merupakan salah satu metode *clustering non* hierarki yang bertujuan untuk memartisi data ke dalam kelompok sehingga data dengan karakteristik serupa dikelompokkan bersama [15], [16].

Dalam tahapannya, jumlah kluster K dipilih terlebih dahulu, dan algoritma *K-Means* kemudian mengikuti langkah-langkah berikut.

1. Langkah pertama adalah memilih jumlah kluster K yang diinginkan.
2. Kemudian, inisialisasi k pusat kluster dilakukan, biasanya dengan cara *random*, di mana pusat-pusat kluster diberi nilai awal secara acak.
3. Seluruh data atau objek dialokasikan ke kluster terdekat berdasarkan jarak di antara objek-objek tersebut.
4. Proses selanjutnya melibatkan perhitungan ulang pusat kluster, yang dapat diukur dengan menggunakan rata-rata atau median dari semua data dalam kluster tertentu.
5. Setiap objek kemudian ditugaskan kembali ke pusat kluster yang baru.

6. Jika pusat kluster tidak mengalami perubahan, proses *clustering* dianggap selesai. Namun, jika pusat kluster masih berubah, kembali ke langkah nomor 3 dan ulangi proses hingga pusat kluster stabil.

Kemudian, pada untuk menghitung jarak antar tiap kluster pada penelitian ini menggunakan rumus perhitungan jarak *Euclidean*. Jarak *Euclidean* digunakan dalam berbagai aplikasi seperti komputasi grafis, optimasi, analisis data, dan terutama dalam algoritma pembelajaran mesin dan data *mining* seperti *K-Means Clustering*, di mana ia digunakan untuk mengukur kesamaan atau kedekatan antara data poin. Berikut adalah rumus perhitungan jarak menggunakan *Euclidean* pada persamaan (1) [17].

$$d_{ij} = \sqrt{\sum_{j=1}^m (u_j - v_j)^2} \quad (1)$$

Dimana u dan v adalah dua objek yang akan dihitung jaraknya, sedangkan u_j merupakan komponen ke i dari u secara berurutan.

Tahapan akhir dari proses *clustering* adalah melakukan perhitungan evaluasi. Satu cara yang dapat digunakan untuk mengetahui kekuatan kluster dan melihat kualitasnya adalah menggunakan evaluasi *silhouette coefficient* [18]. *Silhouette Coefficient* adalah metrik yang digunakan untuk mengukur seberapa baik sebuah objek telah diklasifikasikan saat menggunakan algoritma *clustering*. Koefisien ini memberikan wawasan tentang pemisahan dan kohesi dari kluster yang terbentuk. Nilai dari metrik ini berkisar antara -1 hingga 1, di mana nilai yang mendekati 1 menunjukkan bahwa objek tersebut jauh dari kluster lain dan dekat dengan pusat klusternya, nilai mendekati 0 menunjukkan bahwa objek tersebut berada di dekat batas keputusan antara dua kluster, dan nilai negatif menunjukkan bahwa objek tersebut mungkin telah ditempatkan di kluster yang salah. Rumus *Silhouette Coefficient* dapat diperhatikan pada persamaan (2) [19]

$$s(i) = \frac{b(i) - a(i)}{(\max \{a(i) - b(i)\})} \quad (2)$$

Dimana :

A = banyaknya data kluster A

Nilai $s(i)$ berada antara -1 dan 1, dimana setiap nilai diinterpretasi sebagai berikut:

$s(i) \approx 1 \Rightarrow$ data ke- i digolongkan dengan baik (dalam A)

$s(i) \approx 0 \Rightarrow$ data ke- i berada di tengah antara dua kluster (A dan B)

$s(i) \approx -1 \Rightarrow$ data ke- i digolongkan dengan lemah (dekat ke kluster B daripada A)

2.3. Waikato Environment for Knowledge Analysis (WEKA)

Waikato Environment for Knowledge Analysis (WEKA), adalah perangkat lunak sumber terbuka yang dikembangkan di Universitas Waikato, Selandia Baru, untuk

memfasilitasi tugas analisis data dan eksperimen di bidang data *mining* dan *machine learning*. Dengan antarmuka pengguna grafis yang intuitif, WEKA menyediakan berbagai algoritma pembelajaran mesin dan alat visualisasi yang membantu pengguna menjalankan eksperimen dengan mudah dan memahami struktur data.

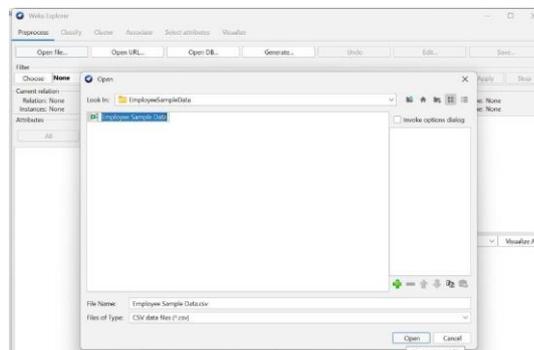
WEKA menonjol dalam proses *preprocessing* data yang efektif, evaluasi model yang komprehensif, dan ekstensibilitas melalui paket ekstensi, memberikan pengguna kemampuan untuk memperluas fungsionalitas sesuai kebutuhan mereka

3. HASIL DAN PEMBAHASAN

3.1. Implementasi Algoritma K-Means Clustering dengan Weka

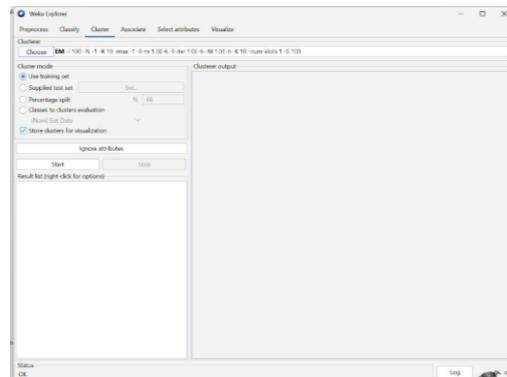
Dalam analisis *clustering* menggunakan Algoritma K-Means pada data set karyawan perusahaan, setiap kluster menggambarkan kelompok karyawan dengan karakteristik yang berbeda-beda, ditentukan oleh kombinasi atribut tertentu. Proses implementasi pada lingkungan WEKA melibatkan beberapa tahap. Pertama, data set karyawan perusahaan diimpor ke dalam platform. Selanjutnya, algoritma *K-Means* dipilih dan dikonfigurasi dengan menentukan jumlah kluster *K* sesuai dengan sifat data set. Inisialisasi pusat kluster dapat dilakukan otomatis atau secara manual oleh pengguna. Algoritma kemudian dieksekusi, dan setelah proses *clustering* selesai, hasilnya dianalisis dengan memeriksa atribut dan karakteristik masing-masing kluster.

1. Langkah pertama adalah melakukan *input* data set yang digunakan pada gambar 3 berikut



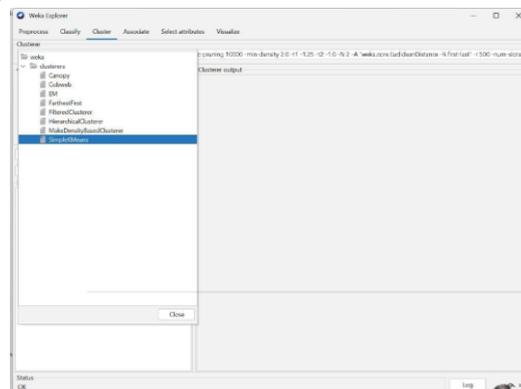
Gambar 3. *Input* Data set pada WEKA

2. Langkah berikutnya pilih menu pengolahan kluster pada bagian atas program seperti pada gambar 4.



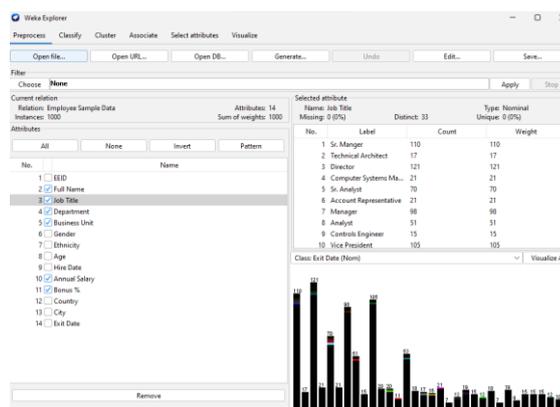
Gambar 4. Pengolahan metode kluster pada WEKA

- Langkah selanjutnya, menentukan metode yang digunakan yaitu *SimpleKmeans* seperti pada gambar 5



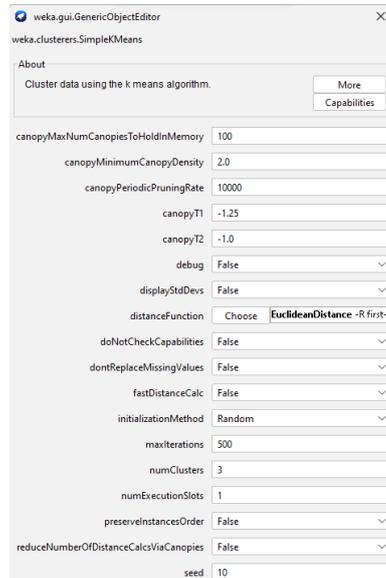
Gambar 5. Pemilihan Algoritma K-Means pada WEKA

- Berikutnya menentukan atribut data set yang digunakan untuk pengolahan data pada gambar 6.



Gambar 6. Penentuan Atribut pada WEKA

- Langkah selanjutnya *input* jumlah kluster yang sudah ditentukan dengan *double* klik pada fitur algoritma *simpleKMeans*. Pada penelitian ini jumlah kluster yang digunakan terbagi menjadi 3 kluster seperti yang tertera pada gambar 7



Gambar 7. Mengatur Jumlah Cluster pada WEKA

- Langkah selanjutnya dalam proses ini adalah melihat hasil pengolahan kluster data yang telah diproses, seperti yang terlihat pada gambar 8 di bawah ini. Pada tahap ini, data berhasil terbagi menjadi 3 kluster setelah melalui proses pengelompokan. Proses perhitungan berlangsung hingga semua data terhitung, memastikan pembentukan kelompok dalam kluster dengan jarak minimal. Penghentian pengulangan terjadi ketika angka pusat kluster tetap konsisten pada iterasi ke-3.

Hasil ini mencerminkan konvergensi algoritma K-Means, menunjukkan bahwa pengelompokan data telah mencapai titik optimal dengan pusat kluster yang stabil. Selama proses ini, kestabilan pusat kluster yang tercapai menandakan bahwa titik tengah masing-masing kluster tidak berubah secara signifikan setelah iterasi ke-3. Konvergensi ini menjadi kunci dalam memastikan bahwa kelompok yang terbentuk sudah mewakili pola yang ada dalam data dengan sebaik mungkin. Hal ini penting untuk mendapatkan pemahaman yang akurat terhadap karakteristik setiap kluster, yang dapat berguna dalam berbagai bidang, seperti analisis pasar, pengenalan pola dalam pengolahan citra, serta dalam identifikasi hubungan antar variabel dalam penelitian ilmiah. Untuk visualisasi dari hasil pengklusteran dapat dilihat pada gambar 9.

```

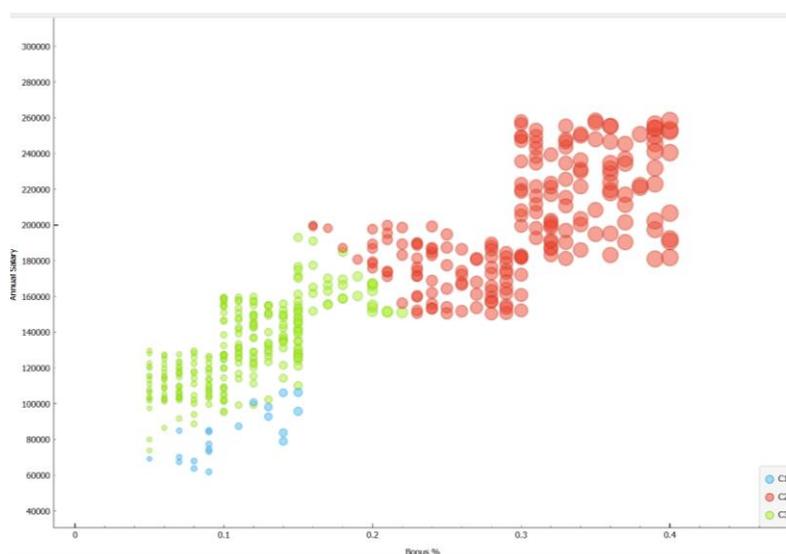
Final cluster centroids:
Attribute                               Cluster#
                                         0           1           2
                                         (586.0)    (325.0)    (89.0)
-----
Full Name                               Rylee Yu    Rylee Yu    Ezra Simmons Savannah Ali
Job Title                               Director    Director    Sr. Analyst  Business Partner
Department                              IT          Sales       IT           Human Resources
Business Unit                           Manufacturing Research & Development Manufacturing Manufacturing
Annual Salary                           $146,140   $157,057   $70,165     $157,333
Bonus %                                 0%         0%         0%         0%

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances
0      586 ( 59%)
1      325 ( 33%)
2       89 (  9%)
    
```

Gambar 8. Hasil Pengolahan Data kluster pada WEKA



Gambar 9. Visualisasi Hasil Clustering

3.2. Data yang Telah Terkluster

Hasil dari *output Clustered Instances* menunjukkan bahwa data telah berhasil dikelompokkan menjadi tiga kluster utama seperti pada Tabel 1 di bawah ini.

Tabel 1. Kluster Instances

Cluster	Instance	Persentase
0	586	59%
1	324	32%
2	89	9%

Kelompok Terbaik, yang ditandai oleh jabatan "*Director*" di departemen IT, bekerja di unit bisnis *Manufacturing*, dan menerima gaji tinggi tanpa bonus, mencerminkan tim pimpinan senior yang berfokus pada pengembangan teknologi dan strategi bisnis. Kelompok Baik, yang terdiri dari individu dengan jabatan "*Director*" di departemen *Sales*,

bekerja di unit bisnis *Research & Development*, dan memiliki gaji yang lebih rendah, mencakup tim penjualan yang berfokus pada inovasi dan pengembangan produk. Sementara itu, Kelompok Cukup, dengan jabatan "*Sr. Analyst*" di departemen IT, bekerja di unit bisnis *Manufacturing*, dan menerima gaji tinggi tanpa bonus, terdiri dari individu dengan keahlian dan tanggung jawab tingkat tinggi meskipun tidak memiliki jabatan direktur.

Output Clustered Instances mengindikasikan bahwa data telah berhasil dikelompokkan ke dalam tiga kluster utama. Kluster 0, yang merupakan kluster terbesar, memuat 586 *instance* atau sekitar 59% dari total data set. Hasil ini menunjukkan bahwa terdapat sejumlah besar data *point* yang memiliki kesamaan dalam karakteristik atau atribut tertentu, sehingga dikelompokkan bersama dalam kluster 0. Kluster ini mungkin mencerminkan pola atau tren khusus yang ada di dalamnya. Selanjutnya pada kluster 1, meskipun lebih kecil dari kluster 0, tetap memiliki signifikansi dengan 324 *instance* atau sekitar 32% dari total data set. Hal ini mengisyaratkan adanya kelompok data yang berbagi karakteristik tertentu, namun mungkin berbeda dari kelompok yang ada di kluster 0. Analisis lebih lanjut pada kluster ini dapat membuka wawasan tambahan terkait dengan pola khusus atau hubungan antar-*instance*.

Sementara itu, kluster 2 adalah yang terkecil dengan 89 *instance* atau sekitar 9% dari total data set. Meskipun ukurannya lebih kecil, kluster ini memiliki keunikan karakteristik yang membedakannya dari kluster lainnya. Kluster ini mungkin mencakup data *point* yang memiliki sifat atau atribut khusus yang membedakannya secara signifikan. Dengan menggali lebih dalam ke dalam *output "Final Cluster Centroids,"* tiga kelompok kluster dapat diterjemahkan sebagai representasi dari struktur pekerjaan dan karakteristik karyawan dalam data set. Pertama-tama, Kelompok Terbaik (kluster 0) menonjol dengan kehadiran individu yang memiliki jabatan "*Director*" di departemen IT, bekerja di unit bisnis *Manufacturing*, dan menerima gaji tahunan sekitar \$157,057 tanpa adanya bonus. kluster ini mungkin mencerminkan kelompok karyawan senior atau pemimpin yang terlibat dalam strategi bisnis dan pengembangan teknologi.

Di sisi lain, Kelompok Baik (kluster 1) menunjukkan adanya individu dengan jabatan "*Director*" di departemen *Sales*, bekerja di unit bisnis *Research & Development*, dengan gaji tahunan sekitar \$70,165 dan tanpa bonus. Meskipun gaji tahunan lebih rendah dibandingkan dengan Kelompok Terbaik, kluster ini mungkin mencakup tim penjualan yang berfokus pada inovasi dan pengembangan produk. Sementara itu, Kelompok Cukup (kluster 2) mewakili individu dengan jabatan "*Sr. Analyst*" di departemen IT, bekerja di unit bisnis *Manufacturing*, dan menerima gaji tahunan sekitar \$157,333 tanpa adanya bonus. Meskipun jabatan ini mungkin tidak sejajar dengan jabatan direktur, gaji yang tinggi dapat mencerminkan tingkat keahlian dan tanggung jawab yang signifikan.

Secara keseluruhan, hasil *output* memberikan pandangan rinci tentang struktur internal organisasi, menyoroti tingkat tanggung jawab, jabatan, dan unit bisnis yang berbeda. Pengkategorian ke dalam kelompok Terbaik, Baik, dan Cukup memberikan

pemahaman lebih lanjut tentang distribusi karyawan dan potensi kontribusi mereka terhadap tujuan perusahaan.

4. KESIMPULAN

Dari analisis menggunakan algoritma K-Means Clustering pada data set karyawan perusahaan, dapat disimpulkan bahwa pengelompokan data tersebut memberikan gambaran yang kaya tentang struktur internal organisasi dan karakteristik karyawan. Hasil penelitian ini menunjukkan bahwa penelitian yang dilakukan berhasil melakukan clustering data menjadi 3 kelompok karyawan menjadi kelompok terbaik, baik dan cukup. Kluster 1 berisikan kelompok jabatan "Director" di departemen IT, bekerja di unit bisnis Manufacturing. Kluster 2 berisikan kelompok jabatan "Director" di departemen Sales, bekerja di unit bisnis Research & Development. Sedangkan kelompok 3 pada posisi jabatan "Sr. Analyst" di departemen IT, bekerja di unit bisnis Manufacturing.

Nilai evaluasi yang diperoleh menggunakan Silhouette Coefficient sebesar 0.597. Hal ini menunjukkan nilai evaluasi yang hampir mendekati nilai +1, sehingga penelitian ini cukup baik dalam melakukan clustering. Penggunaan algoritma K-Means Clustering pada data karyawan menjadi alat yang efektif dalam memahami pola-pola ini dan mendukung upaya perusahaan untuk merancang strategi berdasarkan pemahaman yang lebih mendalam terhadap keberagaman dan kontribusi karyawan.

5. SARAN

Penelitian menggunakan algoritma K-Means Clustering pada data set karyawan perusahaan menghasilkan tiga kelompok kluster yang mewakili tingkatan karyawan berdasarkan jabatan dan gaji. Saran untuk penelitian lebih lanjut melibatkan validasi model, integrasi faktor tambahan seperti pendidikan dan pengalaman kerja, eksplorasi metode pengelompokan alternatif, analisis dinamika temporal, pemahaman penyebab di balik pengelompokan, pengaruh terhadap kinerja bisnis, dan implementasi strategi pengembangan karyawan. Penelitian mendalam dalam aspek-aspek ini dapat memberikan pemahaman yang lebih kaya tentang struktur organisasi dan membantu perusahaan mengoptimalkan pengelolaan sumber daya manusia untuk mencapai tujuan bisnis yang lebih baik.

DAFTAR PUSTAKA

- [1] Tyas, T. M. M., & Purnamasari, A. I. (2023). Penerapan Algoritma K-means dalam Mengelompokkan Demam Berdarah Dengue Berdasarkan Kabupaten. *Blend Sains Jurnal Teknik*, 1(4), 277-283.
- [2] Natalis, A., & Nataliani, Y. (2022). Pemanfaatan k-Means Clustering dan Analytic Hierarchy Process terhadap Penilaian Prestasi Kerja Pegawai. *Jurnal Teknik Informatika dan Sistem Informasi*, 8(1), 88-99.

- [3] Asroni, A., Fitri, H., & Prasetyo, E. (2018). Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). *Semesta Teknika*, 21(1), 60-64.
- [4] Nasir, J. (2020). Penerapan Data Mining Clustering Dalam Mengelompokkan Buku Dengan Metode K-Means. *Jurnal Simetris*, 11(2).
- [5] Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan. *Jurnal Tekno Kompak*, 15(2), 25-36.
- [6] Alhapizi, M. R., Nasir, M., & Effendy, I. (2020). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang. *Journal of Software Engineering Ampera*, 1(1), 1-14.
- [7] Suryadewiansyah, M. K., & Tju, T. E. E. (2022). Naïve Bayes dan Confusion Matrix untuk Efisiensi Analisa Intrusion Detection System Alert. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 8(2), 81-88.
- [8] Lestari, W. (2019). Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus: STMIK Bina Bangsa Kendari). *Jurnal Sistem Informasi dan Sistem Komputer*, 4(2), 35-48.
- [9] Irmayani, W. (2021). Visualisasi Data Pada Data Mining Menggunakan Metode Klasifikasi Naïve Bayes. *Jurnal Khatulistiwa Informatika*, 9(1).
- [10] Kusnadi, Y., & Putri, M. S. (2021). Clustering Menggunakan Metode K-Means Untuk Menentukan Prioritas Penerima Bantuan Bedah Rumah (Studi Kasus: Desa Ciomas Bogor). *Jurnal Teknologi Informatika dan Komputer*, 7(1), 17-24.
- [11] Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 5(1), 62.
- [12] Pratama, Y., Pramudya, Y., Albert, E., Ferdinand, R., Juniansyah, V., & Rasywir, E. (2022). Klasterisasi Data Pertanian di Tingkat Provinsi Jambi Tahun 2021 Menggunakan Algoritma K-Means. *Bulletin of Computer Science Research*, 3(1), 57-63.
- [13] Asyuti, S., & Setyawan, A. A. (2023). DATA MINING DALAM PENGGUNAAN PRESENSI KARYAWAN DENGAN CLUSTER MEANS. *Jurnal Ilmiah Sains Teknologi Dan Informasi*, 1(1), 1-10.
- [14] Hartono, B., Eniyati, S., & Hadiono, K. (2023). Perbandingan Metode Perhitungan Jarak pada Nilai Centroid dan Pengelompokan Data Menggunakan K-Means Clustering. *Jurnal Sistem Komputer dan Informatika (JSON)*, 4(3), 503-509.
- [15] S. Butsianto and N. T. Mayangwulan, "Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 3, no. 3, 2020, doi: 10.32672/jnkti.v3i3.2428].
- [16] Juliawati, F., Buaton, R., & Saragih, R. (2023). Pengelompokan Data Mining Penerimaan Bantuan Pangan Non Tunai (BPNT) Menggunakan Metode Clustering (Studi Kasus: Kantor Desa Payabakung Hamparan Perak). *Explorer*, 3(2), 69-76.
- [17] Zuhdianto, R., & Mukti, F. S. (2023). A CLUSTERING OPTIMIZATION FOR ENERGY EFFICIENCY IN WIRELESS SENSOR NETWORK USING K-MEANS ALGORITHM. *Jurnal Teknik Informatika (Jutif)*, 4(1), 225-234.

- [18] Reinaldi, Y., Ulinnuha, N., & Hafiyusholeh, M. (2021). Comparison of Single Linkage, Complete Linkage, and Average Linkage Methods on Community Welfare Analysis in Cities and Regencies in East Java. *Jurnal Matematika, Statistika dan Komputasi*, 18(1), 130-140.
- [19] Kusumastuti, R., Bayunanda, E., Rifa'i, A.M., Asgar, M.R., Ilmawati, F.I., & Kusrini, K. (2022). Clustering Titik Panas Menggunakan Algoritma Agglomerative Hierarchical Clustering (AHC). *CogITo Smart Journal*.