VOL. 03. NO. 5 SEPTEMBER 2026

JOURNAL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE (JCSAI)

ANALISIS ALGORITMA K-MEANS MENGGUNAKAN METODE ELBOW DAN SILHOUETTE UNTUK KLASTERISASI GAJI KARYAWAN

Fiqri Miftakhul Huda[1], Nana Suarna[2], Agus Bahtiar[3], Ade Rizki Rinaldi[4]

Program Studi Teknik Informatika[12]
Program Studi Sistem Informasi[3]
Program Studi Rekayasa Perangkat Lunak[4]

STMIK IKMI Cirebon
https://ikmi.ac.id/page/18/?lang=de
fiqrimiftakhulhuda.kbm@gmail.com

(*) Corresponding Author : fiqrimiftakhulhuda.kbm@gmail.com
Published : 30 Januari 2026

Abstract—*This research aims to classify employee salary levels into low (Tier 1), medium (Tier 2), and high (Tier 3) categories by grouping employees based on key characteristics such as Education level, age, joining year, work experience, and Payment tier using the K-Means Clustering algorithm. The study also focuses on optimizing the number of clusters using the Elbow Method and Silhouette Score to obtain the most representative grouping. The dataset used in this research consists of 4,653 employee records with nine main attributes. The research stages include data preprocessing, categorical variable conversion using LabelEncoder, feature normalization using MinMaxScaler, determination of the optimal number of clusters, and visualization of clustering results. The Elbow Method indicates that the optimal cluster number is k = 4. The Silhouette Score also supports k = 4 as the most interpretable and compact grouping (Silhouette Score = 0.2063). Therefore, four clusters were selected as the optimal number. The clustering results reveal four distinct groups of employees based on patterns of age, work experience, education level, and Payment tier. The clusters formed include: Cluster 0 (Junior/Low Salary, approx. ₹3–6 LPA) consisting of younger employees with low work experience; Cluster 1 (Highest Salary/Developing Career, approx. ₹15–30 LPA) consisting of young employees with the highest PaymentTier; Cluster 2 (Junior/Medium-High Salary, approx. ₹7–14 LPA) consisting of the youngest employees with medium-high salary; and Cluster 3 (Highly Educated/High Potential) consisting of employees with the highest education levels and significant career growth potential.Tier Payment Tier Cluster Tier Education Payment Tier Overall, this research successfully applies and evaluates the K-Means algorithm to group employees effectively and provides valuable insights that can support companies in salary distribution analysis, career development planning, and data-driven decision-making. This study also opens opportunities for further development by adding more features and comparing other clustering algorithms.*

Keywords*: K-Means, Elbow Method, Silhouette Score, Clustering, Data Mining, Employee, Salary Classification.*

Abstrak—Penelitian ini bertujuan untuk menerapkan tingkat gaji karyawan ke dalam kategori rendah (*Tier* 1), menengah (*Tier* 2), dan tinggi (*Tier* 3) dengan mengelompokkan karyawan berdasarkan karakteristik utama seperti tingkat pendidikan, usia, tahun bergabung, pengalaman kerja, dan tingkat pembayaran menggunakan algoritma *K-Means Clustering*. Selain itu, penelitian ini juga mengoptimalkan jumlah klaster dengan menggunakan Metode *Elbow* dan *Silhouette* untuk memperoleh pembagian kelompok yang paling *representatif*. Data yang digunakan dalam penelitian ini berasal dari dataset *Employee.csv* yang memuat 4.653 data karyawan dengan sembilan atribut utama. Tahap penelitian meliputi pra-pemrosesan data, konversi variabel kategorikal menggunakan *LabelEncoder*, normalisasi fitur dengan *MinMaxScaler*, penentuan jumlah klaster optimal, hingga visualisasi hasil klasterisasi. Nilai untuk k = 4 memberikan keseimbangan terbaik antara nilai SSE dan *Silhouette* serta menghasilkan klaster yang paling mudah diinterpretasikan. Oleh karena itu, empat klaster dipilih sebagai jumlah klaster optimal.Hasil klasterisasi menunjukkan adanya empat kelompok karyawan berdasarkan pola usia, pengalaman kerja, tingkat pendidikan, dan tingkat pembayaran yang berbeda. Klaster yang terbentuk meliputi: Klaster 0 (Berpendidikan Tinggi/Gaji Rendah, *Payment Tier* rendah, kisaran ₹3–6 *LPA*) yaitu karyawan usia muda dengan pengalaman kerja rendah; Klaster 1 (Gaji Tertinggi/Karyawan Muda, *Payment Tier* menengah, kisaran ₹7–14 *LPA*) yaitu karyawan dengan pengalaman dan karir yang berkembang; Klaster 2 (Junior/Gaji

Menengah-Tinggi, *Payment Tier* tinggi, kisaran ₹15–30 *LPA*) yaitu karyawan senior dengan pengalaman panjang; dan Klaster 3 (Berpendidikan Tinggi/Potensial) yaitu karyawan dengan tingkat pendidikan tertinggi dan potensi karir besar. Secara keseluruhan, penelitian ini berhasil menerapkan dan mengevaluasi algoritma K-*Means* secara efektif untuk mengelompokkan karyawan, serta memberikan wawasan penting yang dapat digunakan oleh perusahaan dalam analisis distribusi gaji, perencanaan pengembangan karier, dan pengambilan keputusan berbasis data. Penelitian ini juga membuka peluang pengembangan lebih lanjut dengan menambah fitur dan membandingkan algoritma klasterisasi lainnya.

Kata Kunci: K-*Means*, *Elbow*, *Silhouette*, Klasterisasi, Data Mining, Karyawan, Salary Classification.

INTRODUCTION

Analisis data gaji berbasis *data mining* dan *clustering* telah menjadi salah satu pendekatan penting dalam mendukung pengambilan keputusan berbasis data di bidang sumber daya manusia (SDM). Sejumlah penelitian menunjukkan bahwa teknik validasi klaster seperti *Silhouette*, *Within-Cluster Sum of Squares* (WCSS) atau *Elbow*, serta *Gap Statistic* berperan penting dalam menentukan jumlah klaster optimal (k) pada algoritma K-*Means*. Selain itu, berbagai pengembangan algoritmik seperti *genetic algorithm* dan *metaheuristic initialization* terbukti meningkatkan stabilitas dan interpretabilitas hasil klasterisasi pada dataset sosio-ekonomi yang heterogen. Penerapan metodologi yang terstandardisasi dalam *quality assurance* (QA) dan *reproducibility pipeline* menjadikan segmentasi gaji lebih dapat dipercaya untuk keperluan pemerataan upah, perencanaan kompensasi, dan deteksi anomali dalam sistem HR (Damos, Zhu, Li, Khalifa, Hassan, & Elhabob, 2024; Dzimińska et al., 2021; Khan et al., 2024a; Ogrizović et al., 2024; Shutaywi & Kachouie, 2021);

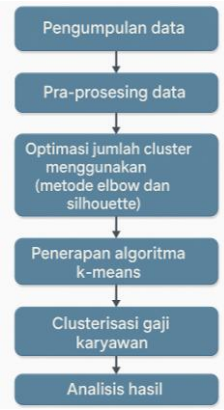
Analisis data gaji karyawan merupakan salah satu aspek penting dalam manajemen sumber daya manusia, terutama untuk mendukung kebijakan kompensasi yang adil dan berbasis data. Dengan meningkatnya *volume* dan kompleksitas data kepegawaian, metode *data mining* seperti algoritma K-*Means* menjadi alat yang efektif untuk mengelompokkan karyawan berdasarkan kesamaan atribut, termasuk gaji. Namun, sejumlah tantangan masih dihadapi dalam penerapan K-*Means* untuk klasterisasi gaji, antara lain sensitivitas terhadap inisialisasi pusat klaster, penentuan jumlah klaster (k) yang optimal, serta risiko bias dan ketidakadilan dalam hasil pengelompokan [2][3], [4].

Metode *Elbow* dan *Silhouette* dikembangkan untuk membantu menentukan jumlah klaster yang paling representatif dengan menilai keseimbangan antara kompaksi dan pemisahan klaster [4], [5]. Penelitian terkini menunjukkan bahwa kombinasi kedua metode ini mampu meningkatkan akurasi hasil *clustering*, mengurangi kesalahan klasterisasi, serta meningkatkan stabilitas algoritma K-*Means* pada dataset berskala besar [3], [6]. Selain itu, pendekatan inisialisasi adaptif dan teknik optimisasi berbasis metaheuristik telah terbukti mengatasi permasalahan local minima serta memperbaiki reprodusibilitas hasil [2], [7] Dengan demikian, optimalisasi algoritma K-*Means* menggunakan metode *Elbow* dan *Silhouette* diharapkan dapat menghasilkan pengelompokan gaji karyawan yang lebih akurat, adil, dan bermakna untuk mendukung pengambilan keputusan strategis organisasi.

Menentukan jumlah klaster optimal merupakan langkah krusial dalam analisis data gaji karena kesalahan dalam menentukan k dapat menyebabkan pemisahan kelompok yang keliru, rendahnya kohesi internal, serta pemisahan antar-klaster yang tidak jelas. Pendekatan seperti metode *Elbow*, *Silhouette*, dan Gap Statistic menawarkan cara pelengkap untuk memperkirakan k yang optimal. Penelitian terkini menunjukkan bahwa kombinasi antara *indeks validitas* internal (*Silhouette*, CH, DB) dengan peningkatan algoritmik seperti *adaptive initialization* dan *gap-based standardization* menghasilkan pemilihan klaster yang lebih stabil dan dapat diinterpretasikan dengan baik untuk aplikasi praktis seperti klasterisasi gaji karyawan [1], [2], [8], [9], [10].

MATERIALS AND METHODS

Penelitian ini juga melakukan optimasi jumlah *cluster* dengan metode *Elbow* dan *Silhouette* untuk mendapatkan jumlah kelompok yang paling representatif. Pendekatan ini diharapkan dapat menghasilkan klasifikasi gaji karyawan yang lebih akurat dan membantu organisasi dalam pengambilan keputusan terkait kompensasi dan kebijakan SDM.Berikut urutan Desain penelitian seperti pada gambar 1



Gambar 1. Alur Penelitian

Tahap pertama dalam proses ini adalah pengumpulan data, yaitu mengumpulkan berbagai informasi yang relevan dengan penelitian. Data yang digunakan biasanya berasal dari sumber internal perusahaan, seperti data karyawan yang berisi nama, jabatan, masa kerja, pendidikan terakhir, serta gaji yang diterima. Proses pengumpulan data bertujuan untuk memperoleh dataset yang akurat, lengkap, dan representatif agar hasil analisis yang diperoleh dapat mencerminkan kondisi sebenarnya di lapangan. Kualitas data pada tahap ini sangat menentukan hasil akhir proses analisis karena data yang tidak lengkap atau tidak valid dapat mengganggu proses pengolahan dan interpretasi selanjutnya.

Pra-prosesing Data Setelah data terkumpul, dilakukan tahap pra-prosesing data untuk memastikan data siap digunakan dalam proses analisis. Tahap ini mencakup beberapa langkah penting seperti pembersihan data (data cleaning) untuk menghapus duplikasi dan menangani nilai kosong (*missing values*), serta normalisasi data agar seluruh variabel berada dalam skala yang sebanding. Selain itu, dilakukan pula seleksi fitur (*feature selection*) dengan memilih atribut yang relevan terhadap penelitian, misalnya variabel gaji dan masa kerja. Langkah-langkah ini penting karena algoritma K-*Means* sangat sensitif terhadap skala data dan kehadiran nilai-nilai ekstrem. Dengan demikian, pra-prosesing membantu meningkatkan kualitas dan konsistensi data sebelum dilakukan pengelompokan.

Optimasi Jumlah *Cluster* (Metode *Elbow* dan *Silhouette*) Tahap berikutnya adalah menentukan jumlah *cluster* yang optimal sebelum algoritma K-*Means* dijalankan. Dalam penelitian ini digunakan dua metode, yaitu metode *Elbow* dan metode *Silhouette*. Metode *Elbow* dilakukan dengan menghitung nilai *Sum of Squared Error* (SSE) untuk berbagai kemungkinan jumlah *cluster*, kemudian memilih titik di mana penurunan nilai SSE mulai melambat (titik siku). Sedangkan metode *Silhouette* digunakan untuk mengukur seberapa baik setiap data cocok dengan *cluster*-nya dibandingkan dengan *cluster* lain. Nilai *Silhouette* yang tinggi menunjukkan pembentukan *cluster* yang lebih baik. Melalui kedua metode ini, jumlah *cluster* yang paling optimal dapat ditentukan, sehingga hasil pengelompokan menjadi lebih akurat dan bermakna.

Penerapan Algoritma K-*Means* Setelah jumlah *cluster* optimal diperoleh, tahap selanjutnya adalah menerapkan algoritma K-*Means*. Proses ini dimulai dengan menentukan titik pusat awal (centroid) secara acak, kemudian menghitung jarak setiap data terhadap masing-masing centroid menggunakan metode jarak *Euclidean*. Setiap data akan dikelompokkan ke dalam *cluster* dengan jarak terdekat. Setelah semua data terklasifikasi, centroid setiap *cluster* akan diperbarui berdasarkan rata-rata posisi data yang ada di dalamnya. Proses ini diulang hingga posisi *centroid* tidak lagi berubah (*konvergen*), yang menandakan bahwa pembentukan *cluster* telah stabil. Hasil dari tahap ini berupa pembagian data ke dalam beberapa kelompok berdasarkan kemiripan karakteristiknya.

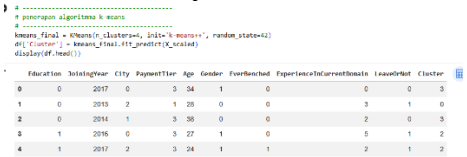
***Cluster*isasi Gaji Karyawan** Tahap ini merupakan hasil dari penerapan algoritma K-*Means*, di mana data karyawan telah terbagi ke dalam beberapa *cluster* berdasarkan kesamaan atribut tertentu, seperti masa kerja dan gaji. Misalnya, hasil pengelompokan dapat membentuk tiga *cluster*, yaitu *cluster* karyawan baru dengan gaji rendah, *cluster* karyawan menengah dengan gaji sedang, dan *cluster* karyawan senior dengan gaji tinggi. Melalui hasil *cluster*isasi ini, perusahaan dapat memahami pola distribusi gaji karyawan dengan lebih baik. Tahap ini juga dapat membantu pihak manajemen atau HRD dalam mengambil keputusan strategis, seperti menentukan kebijakan kenaikan gaji, promosi jabatan, atau perencanaan pelatihan berdasarkan kelompok karyawan.

Analisis Hasil Tahap terakhir adalah analisis hasil, yaitu proses interpretasi terhadap hasil *cluster* yang telah terbentuk. Pada tahap ini dilakukan evaluasi untuk mengetahui karakteristik masing-masing *cluster* dan melihat sejauh mana hasil pengelompokan sesuai dengan tujuan penelitian. Analisis hasil dapat dilakukan dengan menampilkan grafik sebaran data tiap *cluster* serta mengidentifikasi perbedaan utama di antara kelompok yang terbentuk. Dari hasil analisis ini, dapat diperoleh berbagai insight yang berguna bagi perusahaan, seperti pemetaan tingkat kesejahteraan karyawan, efektivitas sistem penggajian, serta potensi ketimpangan antar kelompok. Dengan demikian, tahap analisis hasil

menjadi landasan penting dalam penyusunan rekomendasi kebijakan berbasis data.

RESULTS AND DISCUSSION

Dalam proses pengelompokan menggunakan algoritma K-Means, digunakan metode pengukuran jarak untuk menentukan kedekatan antara data dengan centroid. Salah satu metode yang digunakan adalah *Euclidean Distance*, yaitu jarak lurus antara dua titik dalam ruang multidimensi. Jarak ini dihitung dengan cara mengukur selisih antar atribut data, kemudian dikuadratkan, dijumlahkan, dan diakarkan. Semakin kecil nilai jarak yang dihasilkan, maka semakin dekat hubungan antara data dengan centroid *cluster* tersebut. Sebaliknya, jika jarak yang dihasilkan besar, maka data tersebut dianggap kurang sesuai dengan *cluster* tersebut. Oleh karena itu, setiap data akan dikelompokkan ke dalam *cluster* yang memiliki nilai jarak terkecil terhadap centroid, sehingga menghasilkan pengelompokan yang optimal berdasarkan kemiripan karakteristik data.



```
# penerapan algoritma k means
kmeans_final = KMeans(n_clusters=4, init='k-means++', random_state=42)
df['Cluster'] = kmeans_final.fit_predict(X_scaled)
display(df.head())
```

	Education	JoiningYear	City	PaymentTier	Age	Gender	EverBenched	ExperienceInCurrentDomain	LeaveOrNot	Cluster
0	0	2017	0	3	34	1	0	0	0	3
1	0	2013	2	1	28	0	0	3	1	0
2	0	2014	1	3	38	0	0	2	0	3
3	1	2016	0	3	27	1	0	5	1	2
4	1	2017	2	3	24	1	1	2	1	2

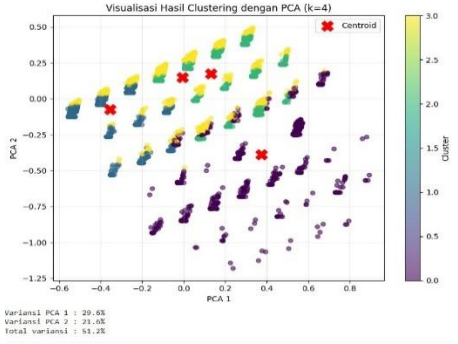
Gambar 2 penerapan algoritma k means

Pada gambar tersebut ditampilkan proses penerapan algoritma K-Means dengan jumlah *cluster* sebanyak 4 (n-*clusters* = 4) menggunakan metode inisialisasi *k-means++* dan *random state* sebesar 42 untuk memastikan hasil yang konsisten. Dataset yang digunakan terdiri dari beberapa variabel, yaitu *Education, JoiningYear, City, PaymentTier, Age, Gender, EverBenched, ExperienceInCurrentDomain,* dan *LeaveOrNot*. Setelah proses *cluster*ing dilakukan, sistem menambahkan satu kolom baru yaitu *Cluster*, yang berisi label hasil pengelompokan data ke dalam *cluster* tertentu, yaitu *cluster* 0, 1, 2, dan 3.

Setiap baris data pada tabel tersebut merepresentasikan satu individu atau entitas, yang kemudian dikelompokkan ke dalam *cluster* berdasarkan kemiripan karakteristik dari variabel-variabel yang digunakan. Sebagai contoh, pada data baris pertama masuk ke dalam *cluster* 3, sedangkan baris kedua masuk ke *cluster* 0, dan seterusnya. Perbedaan label *cluster* ini menunjukkan bahwa masing-masing data memiliki pola atau karakteristik yang berbeda sehingga dikelompokkan ke dalam *cluster* yang berbeda pula.

Hasil ini menunjukkan bahwa algoritma K-Means berhasil melakukan segmentasi data menjadi 4 kelompok utama. Setiap *cluster* nantinya dapat dianalisis lebih lanjut untuk mengetahui karakteristik masing-masing kelompok, misalnya berdasarkan rata-rata usia, tingkat pengalaman kerja, atau kemungkinan karyawan untuk keluar (*LeaveOrNot*). Dengan adanya pengelompokan ini, diharapkan dapat memberikan insight yang lebih jelas terkait pola data yang ada dalam dataset.

Pada tahap ini dilakukan visualisasi hasil *cluster*isasi untuk mempermudah dalam memahami pola pengelompokan data yang telah dihasilkan oleh algoritma K-Means. Visualisasi ini bertujuan untuk menampilkan distribusi data ke dalam masing-masing *cluster* sehingga perbedaan karakteristik antar kelompok dapat terlihat dengan lebih jelas. Umumnya, visualisasi dilakukan dalam bentuk grafik scatter plot dengan menggunakan dua atau lebih variabel yang telah dinormalisasi, di mana setiap titik merepresentasikan satu data dan warna yang berbeda menunjukkan *cluster* yang berbeda. Selain itu, centroid dari masing-masing *cluster* juga ditampilkan sebagai titik pusat yang menjadi representasi rata-rata dari data dalam *cluster* tersebut. Dari hasil visualisasi, dapat diamati bahwa data dalam satu *cluster* cenderung berkumpul dan terpisah dari *cluster* lainnya, yang menunjukkan bahwa proses *cluster*ing telah berjalan dengan baik. Dengan adanya visualisasi ini, interpretasi terhadap hasil pengelompokan menjadi lebih mudah, terutama dalam mengidentifikasi pola distribusi dan perbedaan karakteristik antar *cluster* yang terbentuk.



Gambar 3 visualisasi hasil *cluster* dengan pca

PCA (*Principal Component Analysis*) digunakan untuk mereduksi data dari beberapa variabel menjadi dua komponen utama, yaitu PCA 1 dan PCA 2, sehingga data dapat divisualisasikan dalam bentuk dua dimensi. Dalam penelitian ini, PCA menggabungkan variabel seperti *Age, PaymentTier, ExperienceInCurrentDomain, JoiningYear*, dan variabel lainnya menjadi komponen baru yang tetap merepresentasikan informasi utama dari data.

Komponen PCA 1 merupakan kombinasi dari variabel-variabel yang memiliki pengaruh paling besar dalam membedakan data karyawan, seperti usia, pengalaman kerja, atau tingkat gaji (*PaymentTier*). Sedangkan PCA 2 merupakan kombinasi variabel lain yang juga berkontribusi, namun dengan tingkat pengaruh yang lebih kecil dibandingkan PCA 1. Berdasarkan hasil perhitungan, PCA 1 menjelaskan 29,6% variansi data, PCA 2 menjelaskan 21,6% variansi, sehingga total variansi yang dapat dijelaskan oleh kedua komponen adalah 51,2%. Artinya, lebih dari separuh informasi dalam data berhasil direpresentasikan dalam visualisasi dua dimensi ini.

Berdasarkan hasil visualisasi PCA, setiap titik merepresentasikan satu data karyawan, sedangkan warna menunjukkan hasil pengelompokan *cluster* dari algoritma K-Means. Jika titik-titik dengan warna berbeda terlihat cukup terpisah, maka hal ini menunjukkan bahwa hasil *cluster*ing sudah cukup baik. Namun, jika masih terdapat tumpang tindih antar *cluster*, maka menunjukkan bahwa beberapa data memiliki karakteristik yang mirip sehingga pemisahan *cluster* belum sepenuhnya optimal. Berdasarkan hasil visualisasi PCA pada penelitian ini, terlihat adanya beberapa tumpang tindih antar klaster, terutama antara Klaster 1 dan Klaster 2 yang memiliki karakteristik serupa. Hal ini konsisten dengan nilai Silhouette Score sebesar 0,2063 yang menunjukkan pemisahan klaster yang moderat, dan dapat dijelaskan oleh fakta bahwa total variansi yang dijelaskan oleh dua komponen PCA hanya sebesar 51,2% (PCA 1 = 29,6%, PCA 2 = 21,6%), sehingga sebagian informasi data tidak tertangkap dalam visualisasi dua dimensi ini.

CONCLUSION

Berdasarkan hasil penelitian yang telah dilakukan mengenai penerapan algoritma K-*Means* untuk pengelompokan gaji karyawan dengan optimasi jumlah klaster menggunakan metode *Elbow* dan *Silhouette*, dapat disimpulkan beberapa hal sebagai berikut:

Prosedur penerapan algoritma K-*Means* pada penelitian ini dilakukan melalui tahapan praproses data yang meliputi label encoding dan normalisasi MinMaxScaler, dilanjutkan dengan penentuan jumlah klaster optimal, penerapan algoritma, serta visualisasi dan analisis hasil klasterisasi. Algoritma K-*Means* berhasil mengelompokkan 4.653 data karyawan ke dalam empat klaster utama berdasarkan atribut *Education , Age, Experience In Current Domain, Joining Year,* dan *Payment Tier*. Proses normalisasi dan label encoding pada tahap praproses data berperan penting dalam memastikan bahwa setiap atribut memiliki skala yang seimbang dan dapat diolah secara numerik oleh algoritma.

Berdasarkan evaluasi kombinasi Metode *Elbow* dan *Silhouette*, jumlah klaster optimal yang diperoleh adalah k = 4. Metode *Elbow* menunjukkan titik siku pada grafik *Within Cluster Sum of Squares (WCSS)* menunjukkan penurunan yang signifikan hingga titik siku (*elbow*). Hasil ini kemudian dikonfirmasi melalui metode *Silhouette*, yang menunjukkan bahwa *Silhouette Score* untuk k = 4 memiliki nilai tertinggi sebesar 0,2063 dan representatif, menandakan klasterisasi yang baik dan tidak tumpang tindih secara signifikan.

Hasil klasterisasi berhasil mengklasifikasikan karyawan ke dalam empat kategori: (1) *Cluster* 0 (Junior/Gaji Rendah, kisaran ₹3–6 *LPA* atau setara Rp54 juta–Rp108 juta/tahun) yaitu karyawan usia muda dengan pengalaman kerja rendah dan *PaymentTier* rendah; (2) *Cluster* 1 (Gaji Tertinggi/Karyawan Muda, kisaran ₹15–30 *LPA* atau setara Rp270 juta–Rp540 juta/tahun) yaitu karyawan muda dengan PaymentTier tertinggi dan *PaymentTier* menengah; (3) *Cluster* 2 (Junior/Gaji Menengah-Tinggi, kisaran ₹7–14 *LPA* atau setara Rp126 juta–Rp252 juta/tahun) yaitu karyawan termuda dengan pengalaman menengah dan *PaymentTier* tinggi; dan (4) *Cluster* 3 (Berpendidikan Tinggi/Potensial) yaitu karyawan dengan tingkat pendidikan tertinggi dan potensi karir besar. Implikasi dari klasifikasi ini bagi kebijakan kompensasi perusahaan adalah bahwa perusahaan dapat merancang strategi pengembangan karier berbasis data, menyesuaikan struktur gaji secara proporsional berdasarkan profil karyawan, serta mengidentifikasi kelompok karyawan yang potensial untuk mendapatkan kenaikan kompensasi.

Cluster 0 terdiri dari karyawan dengan tingkat pembayaran (*PaymentTier*) terendah (kisaran ₹3–6 *LPA*), usia lebih muda, dan pengalaman kerja yang relatif sedikit.

Cluster 1 menggambarkan kelompok menengah dengan *PaymentTier* sekitar 2,94 (kisaran ₹7–14 *LPA*), yang umumnya memiliki pengalaman sedang dan masa kerja lebih lama.

Cluster 2 merupakan kelompok dengan *PaymentTier* tertinggi (kisaran ₹15–30 *LPA*), berisi karyawan dengan usia dan pengalaman kerja lebih besar dibanding klaster lainnya.

Cluster 3 merupakan kelompok karyawan dengan tingkat pendidikan tertinggi dan potensi karir paling besar, meskipun pengalaman kerja belum setinggi Klaster 2.

REFERENCE

[1] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," *Entropy*, vol. 23, no. 6, p. 759, 2021, doi: 10.3390/e23060759.

[2] J. Yang, Y.-K. Wang, X. Yao, and C.-T. Lin, "Adaptive initialization method for K-means algorithm," *Front. Artif. Intell.*, vol. 4, p. 740817, 2021, doi: 10.3389/frai.2021.740817.

[3] M. Gul and M. A. Rehman, "Big data: An optimized approach for cluster initialization," *J. Big Data*, vol. 10, p. 120, 2023, doi: 10.1186/s40537-023-00798-1.

[4] C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, and J. Liu, "A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm," *EURASIP J. Wirel. Commun. Netw.*, p. 31, 2021, doi: 10.1186/s13638-021-01910-w.

[5] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics (Basel).*, vol. 9, no. 8, p. 1295, 2020, doi: 10.3390/electronics9081295.

[6] P. Bombina, D. Tally, Z. B. Abrams, and K. R. Coombes, "SillyPutty: Improved clustering by optimizing the silhouette width," *PLoS One*, vol. 19, no. 6, p. e0300358, 2024, doi: 10.1371/journal.pone.0300358.

[7] A. M. Ikotun, M. S. Almutari, and A. E. Ezugwu, "K-means-based nature-inspired metaheuristic algorithms for automatic data clustering problems: Recent advances and future directions," *Applied Sciences*, vol. 11, no. 23, p. 11246, 2021, doi: 10.3390/app112311246.

[8] B. Anhar and A. Taupiq, "Korelasi gaji, disiplin kerja dan budaya organisasi terhadap kinerja karyawan," vol. 13, no. 1, pp. 81–89, 2021, doi: 10.30872/jurnalmanajemen.vVViI.XXXX.

[9] A. A. Wani, "Comprehensive analysis of clustering algorithms: Exploring limitations and innovative solutions," *PeerJ Comput. Sci.*, vol. 10, p. e2286, 2024, doi: 10.7717/peerj-cs.2286.

[10] I. K. Khan *et al.*, "Determining the optimal number of clusters by Enhanced Gap Statistic in K-mean algorithm," *Egyptian Informatics Journal*, vol. 27, p. 100504, 2024, doi: 10.1016/j.eij.2024.100504.