

## Optimalisasi Kinerja Karyawan Berbasis HR Analytics dengan K-Means Clustering dan Analisis Faktor Demografi

Anandita Nabilla Ramadhani<sup>1</sup>, \*Ghita Athalina<sup>2</sup>

<sup>1,2</sup> Program studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya  
Jl. Masjid Al Gazali, Bukit Lama, Kec. Ilir Bar. I, Kota Palembang, Sumatera Selatan  
Email: <sup>1</sup>anandita0211@gmail.com, <sup>2</sup>ghitaathalina@media.is.uec.ac.jp

### ABSTRACT

*Data-driven Human Resources (HR) management is an important aspect in improving organizational productivity and efficiency in the digital era. This research aims to cluster employees based on performance using the K-Means Clustering algorithm and evaluate the influence of demographic factors on job performance. The dataset used is a public dataset from Kaggle, including employee performance information such as Key Performance Indicators (KPIs), training scores, multiple trainings, performance appraisals, awards, as well as demographic attributes such as gender, age, education level, and recruitment channels. Using the six-stage CRISP-DM framework, the data was processed using StandardScaler, and the optimal number of clusters was determined through the Elbow Method, Davies-Bouldin Index, and Silhouette Score, resulting in two main clusters. Cluster 0 includes high-performing employees with KPIs above 80%, good performance ratings, and good training scores, while Cluster 1 consists of low-performing employees, with lower KPIs, poor performance ratings, and training scores. Analysis showed demographic factors did not significantly affect employee performance. This research recommends focused training for low-performing employees and rewards for high-performing employees, so that each employee can reach their full potential and support organizational productivity.*

*Keywords: data-driven HR management; k-means clustering; employee performance; demographic factors*

### ABSTRAK

Manajemen *Human Resources (HR)* berbasis data menjadi aspek penting dalam meningkatkan produktivitas dan efisiensi organisasi di era digital. Penelitian ini bertujuan untuk mengelompokkan karyawan berdasarkan kinerja menggunakan algoritma *K-Means Clustering* serta mengevaluasi pengaruh faktor demografi terhadap performa kerja. Dataset yang digunakan merupakan dataset publik dari *Kaggle*, mencakup informasi kinerja karyawan seperti *Key Performance Indicators (KPI)*, skor pelatihan, banyak pelatihan, penilaian kinerja, penghargaan, serta atribut demografi seperti gender, usia, tingkat pendidikan, dan saluran rekrutmen. Dengan menggunakan kerangka kerja *CRISP-DM* yang terdiri dari enam tahapan, data diolah menggunakan *StandardScaler*, dan jumlah kluster optimal ditentukan melalui Metode *Elbow*, *Davies-Bouldin Index*, dan *Silhouette Score*, yang menghasilkan dua kluster utama. *Cluster 0* mencakup karyawan berkinerja tinggi dengan *KPI* di atas 80%, rating kinerja baik, dan skor pelatihan yang baik, sedangkan *Cluster 1* terdiri dari karyawan berkinerja rendah, dengan *KPI*, rating kinerja buruk, dan skor pelatihan yang lebih rendah. Analisis menunjukkan faktor demografi tidak mempengaruhi kinerja karyawan secara signifikan. Penelitian ini merekomendasikan pelatihan terfokus bagi karyawan berkinerja rendah dan pemberian penghargaan bagi karyawan berkinerja tinggi, sehingga setiap karyawan dapat mencapai potensi maksimalnya dan mendukung produktivitas organisasi.

Kata kunci : *human resources analytics; k-means clustering; analisis demografi; produktivitas karyawan*

## 1. PENDAHULUAN

Produktivitas karyawan adalah elemen penting dalam keberhasilan organisasi, terutama dalam era persaingan global yang semakin kompetitif. Organisasi yang mampu meningkatkan produktivitas tenaga kerja memiliki peluang lebih besar untuk mencapai efisiensi operasional, inovasi berkelanjutan, dan daya saing pasar (Fadilla & Triani, 2024).

Namun, produktivitas tenaga kerja Indonesia tergolong rendah, laporan *International Labour Organization* (ILO) menunjukkan Indonesia berada di peringkat ke-111 dari 189 negara, dengan produktivitas tenaga kerja hanya mencapai US\$14 per jam kerja pada tahun 2023 (International Labour Organization, 2020). Kesenjangan ini mencerminkan perlunya pendekatan strategis untuk memahami pola dan determinan produktivitas karyawan secara lebih mendalam.

Mengidentifikasi faktor-faktor yang mempengaruhi kinerja karyawan merupakan aspek penting untuk meningkatkan produktivitas karyawan. Dalam penelitian ini, kinerja karyawan dihitung berdasarkan beberapa indikator kinerja termasuk skor pelatihan, jumlah

pelatihan yang diikuti, penilaian kinerja sebelumnya, *Key Performance Indicators* (KPI) dan penghargaan yang diterima. KPI memiliki peran penting dalam mendorong kinerja karyawan dan produktivitas organisasi (Sudaryanto, 2024).

Penelitian sebelumnya, menunjukkan bahwa variabel sosiodemografi, seperti usia, tingkat pendidikan, pengalaman kerja, upah, dan partisipasi angkatan kerja, secara signifikan mempengaruhi produktivitas tenaga kerja. Dengan memahami hubungan ini organisasi dapat merancang strategi yang lebih efektif untuk meningkatkan kinerja dan efisiensi karyawan (Dirgantara & Santoso, 2024).

Dalam menghadapi tantangan produktivitas tenaga kerja, diperlukan pendekatan analitik berbasis data guna memahami pola kinerja karyawan secara mendalam dan strategis. Salah satu metode yang efektif adalah algoritma *K-Means Clustering*, yang mampu mengelompokkan data berdasarkan karakteristik yang serupa. Keunggulan K-Means dikenal luas karena keunggulannya dalam mengelompokkan data berukuran besar secara efisien dan

menangani outlier dengan cepat dan akurat (Padiku & Lahinta, 2022).

Berdasarkan hasil penelitian yang dilakukan oleh Gideon Bartolomeus Kaligis dan Sri Yulianto, algoritma *K-Means* menjadi metode terbaik dalam menentukan data kinerja pegawai berdasarkan kriteria tertentu, dengan nilai *Davies Bouldin Index* (DBI) paling rendah dibandingkan *K-Medoids* dan *X-Means* (Kaligis & Yulianto, 2022).

Penelitian serupa oleh Jhiro Faran dan Agung Triayudi juga menunjukkan bahwa penerapan algoritma *K-Means* efektif dalam membentuk kelompok kinerja karyawan dan menghasilkan informasi baru yang mendukung pengambilan keputusan. (Faran & Triayudi, 2024) .

Penelitian lain, berhasil menggunakan *K-Means* untuk mengelompokkan karyawan berdasarkan tiga variabel utama dan menghasilkan pengelompokan karyawan ke dalam kategori terbaik, sangat baik, dan rata-rata (Reisandi et al., 2021).

Namun, ketiga penelitian tersebut belum menggali secara mendalam pengaruh faktor demografi seperti usia, gender, tingkat pendidikan,

dan jalur rekrutmen, yang penting untuk dianalisis agar dapat memberikan wawasan lebih komprehensif tentang faktor-faktor yang membedakan kinerja karyawan.

Adanya penelitian ini bertujuan untuk mengisi celah tersebut dengan mengintegrasikan algoritma *K-Means Clustering* dan analisis faktor demografi guna mengidentifikasi pola kinerja karyawan secara menyeluruh. Dengan menggunakan kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM), yang mengikuti enam tahapan sistematis untuk memastikan proses analisis berjalan terstruktur dan relevan dengan kebutuhan. Hasil dari penelitian ini diharapkan dapat memberikan rekomendasi berbasis data untuk peningkatan produktivitas tenaga kerja serta kontribusi signifikan dalam pengelolaan sumber daya manusia berbasis data.

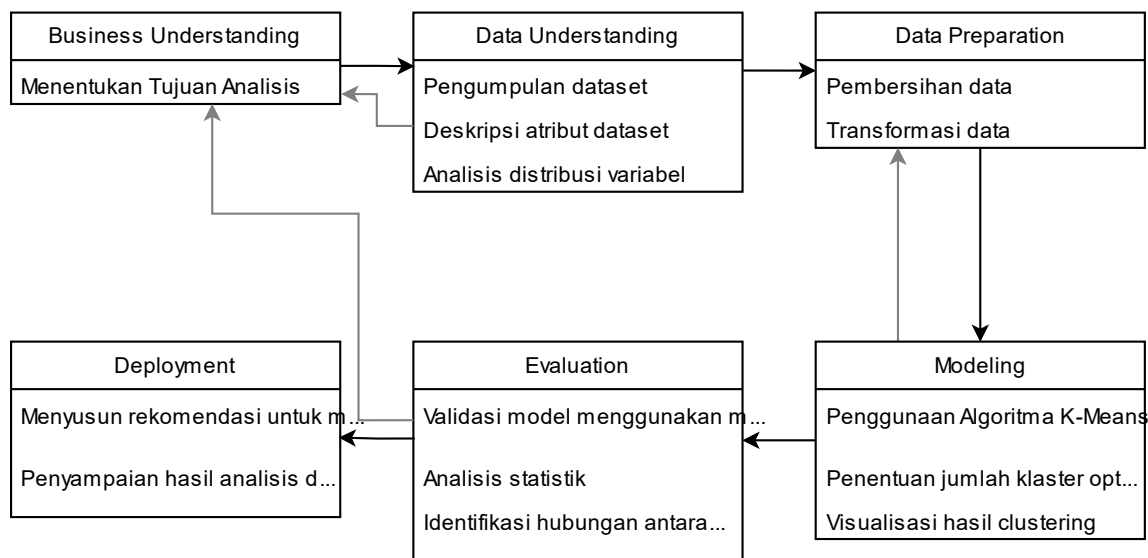
## 2. METODE

Dataset yang digunakan dalam penelitian ini diambil dari *Kaggle* berjudul "*Employees Performance for HR Analytics*," yang terdiri dari 17.417 data karyawan dengan 13 kolom variabel. Untuk menganalisis data ini,

algoritma *K-Means Clustering* diterapkan untuk mengelompokkan karyawan berdasarkan karakteristik kinerja mereka.

Pendekatan ini diorganisir menggunakan kerangka kerja CRISP-DM (*Cross Industry Standard Process for Data Mining*), yang mencakup enam

tahapan utama: *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. CRISP-DM menyediakan kerangka kerja terstruktur yang memandu penelitian secara sistematis dan terarah (Dhewayani et al., 2022). Tahapan dalam penelitian ini seperti pada Gambar 1.



Gambar 1. Metode Penelitian *CRISP-DM*

### 2.1. *Business Understanding*

Tahap ini bertujuan untuk memahami tujuan, mengidentifikasi masalah yang ingin diselesaikan, dan memastikan hasil analisis selaras dengan kebutuhan (Hidayat et al., 2024). Penelitian ini bertujuan untuk:

1. Mengelompokkan karyawan menggunakan algoritma *K-Means Clustering*, berdasarkan atribut

kinerja yaitu jumlah pelatihan yang diikuti, rating kinerja tahun sebelumnya, KPI, penghargaan yang diraih, dan nilai rata-rata pelatihan.

2. Mengevaluasi pengaruh faktor-faktor demografi yaitu usia, gender, tingkat pendidikan, dan saluran rekrutmen, untuk mengidentifikasi

karakteristik yang mempengaruhi kinerja karyawan.

- Menghasilkan rekomendasi berbasis data yang dapat digunakan untuk merancang strategi pengelolaan SDM yang lebih efektif dan meningkatkan produktivitas karyawan.

## 2.2. Data Understanding

Tahap ini bertujuan untuk memahami dataset yang digunakan, mencakup eksplorasi dan pemeriksaan kualitas data. Dataset yang digunakan terdiri dari 17.417 baris data dan 13 kolom variabel, termasuk informasi terkait kinerja dan faktor-faktor demografi karyawan. Penjelasan dataset dapat dilihat pada Tabel 1.

Tabel 1. Penjelasan Dataset

No	Atribut	Tipe Data	Penjelasan
1	employee_id	Integer	ID unik yang diberikan kepada setiap karyawan untuk identifikasi.
2	department	Categorical	Departemen tempat karyawan bekerja.
3	region	Categorical	Wilayah atau lokasi tempat karyawan bekerja.
4	education	Categorical	Tingkat pendidikan terakhir yang dimiliki karyawan (S1-S3).
5	gender	Categorical	Jenis kelamin karyawan (Wanita, Pria).

6	recruitment_channel	Categorical	Cara karyawan direkrut (melalui referensi, job fair, website perusahaan).
7	number_of_trainings	Integer	Jumlah pelatihan yang telah diikuti oleh karyawan (1-9 pelatihan).
8	age	Integer	Usia karyawan dalam rentang 20-60 tahun.
9	previous_year_rating	Float	Peringkat kinerja karyawan pada tahun sebelumnya (rentang 1-5).
10	length_of_service	Integer	Lama waktu karyawan bekerja di perusahaan (1-31 tahun).
11	key_performance_indicators_more_than_80	Binary (0/1)	Indikator apakah karyawan telah mencapai lebih dari 80% dari Key Performance Indicators (KPI) yang ditetapkan (0: tidak mencapai, 1: mencapai).
12	awards_won	Binary (0/1)	Jumlah penghargaan yang telah diperoleh karyawan (0: tidak ada, 1: ada).
13	average_training_score	Float	Nilai rata-rata dari semua pelatihan yang diikuti oleh karyawan (rentang 41-99).

(Sumber:

<https://www.kaggle.com/datasets/sanjanchaudhari/employees-performance-for-hr-analytics>)

## 2.3. Data Preparation

Tahap ini mencakup pembersihan, pengolahan, dan transformasi data untuk memastikan data siap digunakan dalam analisis. Langkah-langkah yang dilakukan adalah:

1. Pembersihan data, pembersihan data dilakukan untuk memastikan kualitas dataset yang digunakan dalam penelitian memberikan hasil yang akurat dan dapat diandalkan (Widiari et al., 2020). Setelah pemeriksaan, ditemukan 771 *missing values* pada kolom pendidikan dan 1.363 *missing values* pada kolom rating kinerja tahun sebelumnya yang kemudian dihapus, sehingga menjadi 15.424 baris. Selanjutnya, dua baris duplikat juga dihapus, menghasilkan dataset bersih sebanyak 15.422 baris, dapat dilihat pada Gambar 2.

```
<class 'pandas.core.frame.DataFrame'>
Index: 15422 entries, 1 to 17416
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   employee_id           15422 non-null  int64
1   department            15422 non-null  object
2   region                15422 non-null  object
3   education             15422 non-null  object
4   gender                15422 non-null  object
5   recruitment_channel   15422 non-null  object
6   no_of_trainings       15422 non-null  int64
7   age                   15422 non-null  int64
8   previous_year_rating  15422 non-null  float64
9   length_of_service     15422 non-null  int64
10  KPIs_met_more_than_80 15422 non-null  int64
11  awards_won            15422 non-null  int64
12  avg_training_score    15422 non-null  int64
dtypes: float64(1), int64(7), object(5)
memory usage: 1.6+ MB
```

Gambar 2. Hasil Pembersihan Data

2. Transformasi data, transformasi data dalam penelitian ini berfokus pada normalisasi variabel kinerja karyawan menggunakan *StandardScaler*. Proses ini mengubah nilai fitur menjadi distribusi dengan rata-rata 0 dan

deviasi standar 1. Tujuannya memastikan kontribusi yang proporsional dalam algoritma *K-Means Clustering*. Normalisasi ini menyelaraskan skala fitur, seperti skor pelatihan (41–99) dan penghargaan (0/1), sehingga bobotnya seimbang dan hubungan antar fitur lebih mudah dipahami (Tresi Aprilia, 2024)

```
no_of_trainings  previous_year_rating  KPIs_met_more_than_80  awards_won \
0      -0.420522          -0.281183          -0.752543          -0.156568
1      -0.420522          -1.863812          -0.752543          -0.156568
2       2.977585          -1.072497          -0.752543          -0.156568
3      -0.420522          0.518132           0.753543          -0.156568
4      -0.420522          -0.281183          -0.752543          -0.156568

avg_training_score
0      -0.929770
1      -1.238513
2       0.122831
3      -0.177912
4       0.348389
```

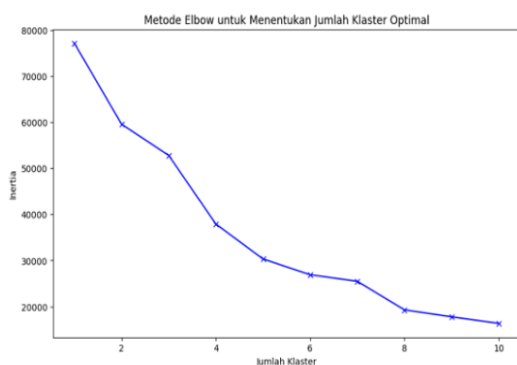
Gambar 3. Hasil Transformasi Data

## 2.4. Modeling

Tahap Modeling berfokus pada penerapan algoritma *K-Means Clustering* untuk mengelompokkan karyawan berdasarkan kinerja dengan variabel yang sudah dinormalisasi.

*K-Means* adalah algoritma yang bekerja dengan meminimalkan jarak antar titik data dalam satu kluster dan memaksimalkan jarak antar kluster. Dalam algoritma ini, jumlah kluster direpresentasikan sebagai nilai  $K$ , dengan tahap pertama yang dilakukan adalah penentuan jumlah kluster optimal. Untuk itu, digunakan dua metode evaluasi yaitu *Elbow Method* dan *Davies-Bouldin Index (DBI)*.

*Elbow Method* berfungsi untuk menentukan jumlah kluster optimal dengan menganalisis perubahan *Within-Cluster Sum of Squares (WCSS)*, seiring bertambahnya jumlah kluster. Penurunan signifikan dalam nilai WCSS menunjukkan bahwa tambahan kluster tidak lagi memberikan peningkatan yang berarti dalam variasi total yang dijelaskan oleh model (Vredizon et al., 2024). Hasil *Elbow Method* pada penelitian ini dapat dilihat pada Gambar 4 dan nilai WCSS dapat dilihat pada tabel 2.



Gambar 4. Hasil *Elbow Method*

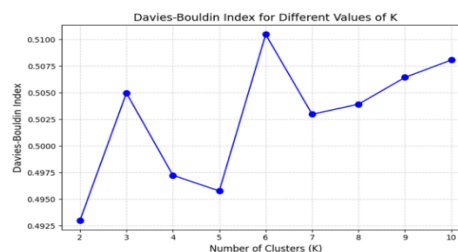
Tabel 2. Nilai K dan Avg. WCSS

Jumlah Kluster (K)	Avg. WCSS (Inertia)
1	77110.000000
2	59557.975450
3	52792.566601
4	37904.278731
5	30325.247177
6	26902.409423
7	25481.368981
8	19286.354512
9	17778.042776
10	16364.208245

Berdasarkan hasil ini, jumlah kluster yang optimal adalah  $K=4$ . Kluster 4 adalah titik di mana penurunan WCSS

mulai melambat, dan dapat dianggap sebagai titik "elbow" dalam grafik *Elbow Method*.

Untuk menguji kebenaran kluster optimal dari *Elbow Method* digunakan *Davies-Bouldin Index*. *Davies-Bouldin Index (DBI)* digunakan untuk menilai kualitas kluster dengan mengukur sejauh mana kluster-kluster tersebut terpisah dan seberapa rapat data dalam kluster tersebut. Nilai DBI yang lebih rendah menunjukkan kualitas kluster yang lebih baik (Umagapi et al., 2023). Hasil perhitungan DBI untuk berbagai jumlah kluster dapat dilihat pada Gambar 5 dan Tabel 3.



Gambar 5. Hasil *Davies-Bouldin Index (DBI)*

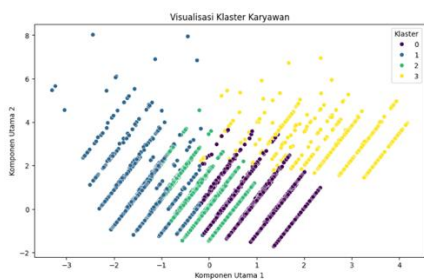
Tabel 3. Nilai K dan DBI

Jumlah Kluster (K)	DBI
2	0.4930
3	0.5050
4	0.4972
5	0.4958
6	0.5105
7	0.5030
8	0.5039
9	0.5064
10	0.5081

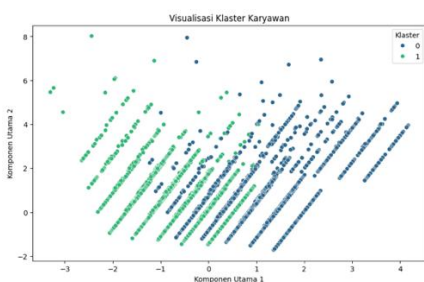
Berdasarkan hasil DBI, jumlah kluster  $K = 2$  menghasilkan nilai DBI

terkecil, yang menunjukkan bahwa pembagian kluster pada  $K = 2$  adalah yang paling optimal di antara pilihan yang diuji.

Untuk mengatasi perbedaan hasil antara *Elbow Method* ( $K = 4$ ) dan *Davies Bouldin Index* atau DBI ( $K = 2$ ) dalam menentukan jumlah kluster optimal, dilakukan visualisasi hasil clustering. Visualisasi ini akan membantu untuk memilih kluster yang paling representatif dan optimal berdasarkan kedua metode tersebut. Hasil visualisasi dapat dilihat pada Gambar 6 dan Gambar 7.



Gambar 6. Hasil Visualisasi *Elbow Method* ( $K=4$ )



Gambar 7. Hasil Visualisasi DBI ( $K = 2$ )

Berdasarkan hasil visualisasi didapati bahwa kluster yang paling representatif dan optimal adalah  $K = 2$  karena menghasilkan pemisahan yang

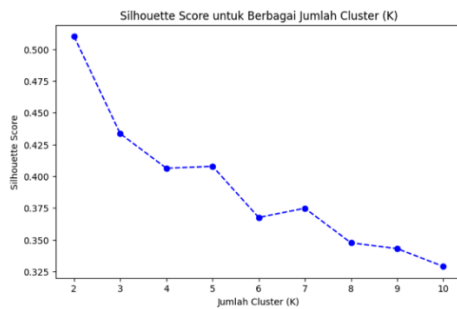
jelas dengan dua kluster yang terpisah secara signifikan jika dibandingkan  $K = 4$ . Sehingga jumlah kluster yang optimal dalam mengelompokkan karyawan berdasarkan kinerja adalah 2 kluster.

## 2.5. Evaluation

Tahap evaluasi bertujuan untuk memastikan bahwa hasil clustering yang dihasilkan oleh algoritma *K-Means Clustering* sesuai dengan tujuan analisis dan memiliki kualitas yang optimal. Penelitian menggunakan *Silhouette Score* untuk mengukur seberapa baik elemen dalam kluster dikelompokkan. Pengujian *Silhouette Score* juga digunakan untuk menentukan jumlah kluster yang paling optimal dalam proses *clustering* suatu dataset (Hendrastuty, 2024).

*Silhouette Score* bekerja dengan membandingkan kedekatan setiap data terhadap klasternya sendiri dan kluster lain. Skor ini memiliki rentang nilai dari -1 hingga 1, di mana nilai mendekati 1 menunjukkan pengelompokan yang sangat baik, nilai mendekati 0 menunjukkan bahwa data berada di perbatasan antar kluster, dan nilai negatif menunjukkan kemungkinan bahwa data salah dikelompokkan. Hasil evaluasi

menggunakan *Silhouette Score* ditampilkan pada Gambar 8.



Gambar 8. Evaluasi *Silhouette Score*

Grafik *Silhouette Score* menunjukkan bahwa  $K = 2$  memiliki skor tertinggi sekitar 0.5 yang menandakan kualitas kluster terbaik dibanding jumlah kluster lainnya. Penurunan skor pada  $K > 2$  menunjukkan menurunnya kualitas pemisahan antar kluster. Dengan demikian, berdasarkan evaluasi ini kluster optimal adalah  $K = 2$ .

Setelah jumlah kluster optimal ditentukan, selanjutnya dilakukan analisis statistik untuk memahami distribusi fitur kinerja di setiap kluster. Analisis ini dilakukan dengan memvisualisasikan perbedaan data dalam fitur kinerja antar kluster. Selain itu, hubungan antara kluster dengan faktor-faktor demografi juga diidentifikasi untuk memberikan wawasan lebih mendalam mengenai karakteristik masing-masing kluster.

## 2.6. Deployment

*Deployment* merupakan tahap akhir yang bertujuan untuk mengintegrasikan hasil analisis *clustering* ke dalam rekomendasi strategis yang dapat langsung diimplementasikan oleh organisasi. Langkah ini memastikan bahwa *insight* yang diperoleh dari model digunakan secara efektif untuk mendukung pengambilan keputusan strategis berbasis data, sehingga memberikan nilai tambah yang nyata bagi organisasi.

## 3. HASIL DAN PEMBAHASAN

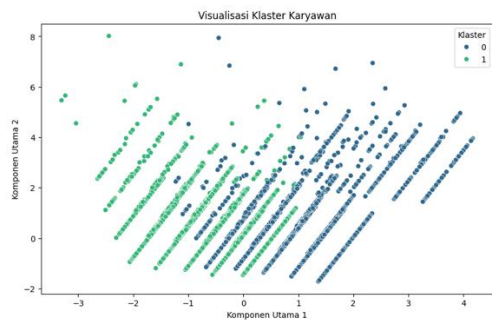
### 3.1. Hasil *K-Means Clustering*

Setelah dipastikan bahwa dua kluster merupakan jumlah optimal berdasarkan *Silhouette Score*. Model *K-Means Clustering* dilatih menggunakan dataset yang telah dinormalisasi ( $X_{scaled}$ ). Prosesnya dimulai dengan inisialisasi pusat kluster (*centroids*) secara acak, diikuti dengan pengelompokan data ke kluster terdekat berdasarkan jarak *Euclidean* (Ari Sulistiyawati & Eko Supriyanto, 2021). *Centroid* kemudian diperbarui secara iteratif berdasarkan rata-rata posisi data dalam masing-masing kluster hingga mencapai konvergensi.

Pelatihan model *K-Means Clustering* dapat dilihat pada Gambar 9, sementara visualisasi hasil klusterisasi menggunakan PCA (*Principal Component Analysis*) ditampilkan pada Gambar 10 dan distribusi jumlah anggota pada masing-masing kluster ditunjukkan pada Tabel 4.

```
# Melatih K-means Clustering dengan Jumlah Kluster Optimal (k=2)
kmeans = KMeans(n_clusters=2, random_state=42)
data['cluster'] = kmeans.fit_predict(X_scaled)
```

Gambar 9. Pelatihan Model *K-Means Clustering*

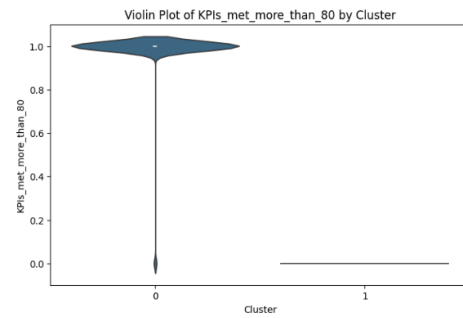


Gambar 10. Visualisasi *Cluster*

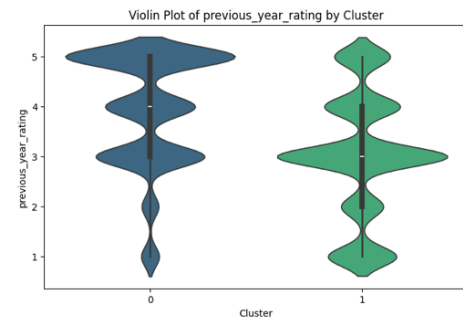
Tabel 4. Distribusi *Cluster*

Kluster	Jumlah Karyawan
0	5668
1	9754

Model *K-Means Clustering* berhasil mengelompokkan data menjadi dua kluster, yaitu Kluster 0 dengan 5.668 karyawan dan Kluster 1 dengan 9.754 karyawan. Untuk memahami karakteristik masing-masing kluster berdasarkan atribut kinerja, dilakukan analisis statistik menggunakan *violin plot* yang ditunjukkan pada Gambar 11 sampai Gambar 15.



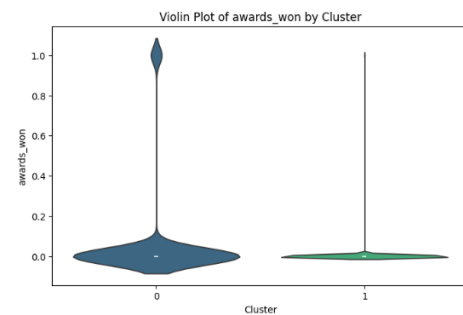
Gambar 11. Penyebaran Kluster berdasarkan Banyak Pelatihan



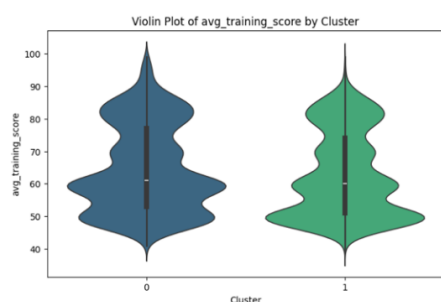
Gambar 12. Penyebaran Kluster berdasarkan Rating Kinerja Tahun Sebelumnya



Gambar 13. Penyebaran Kluster berdasarkan Banyak Pelatihan



Gambar 14. Penyebaran Kluster berdasarkan Penghargaan



Gambar 15. Penyebaran Kluster berdasarkan Rata-rata Skor Pelatihan

Berdasarkan hasil visualisasi persebaran kluster dengan *violin plot*. Didapatilah hasil analisis karakteristik kluster yang ditampilkan pada Tabel 5.

Tabel 5. Karakteristik Kluster

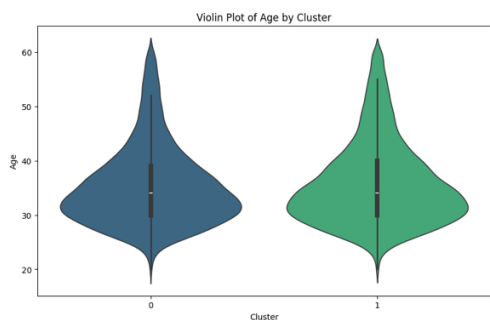
Kluster	Karakteristik	Segmen Kinerja
0	Kluster ini berisi karyawan berkinerja tinggi, yang ditandai dengan pencapaian KPI lebih dari 80%, rata-rata skor pelatihan yang tinggi dibanding kluster 0 (rata-rata skor pelatihan sebesar 88), serta rating kinerja tahun sebelumnya yang baik (Rating rata-rata Cluster 0 berada pada level 4-5). Kluster ini juga berisi karyawan yang menerima penghargaan.	Kinerja Tinggi
1	Kluster ini berisi karyawan berkinerja rendah, yang ditandai dengan pencapaian KPI kurang dari 80%, rata-rata skor pelatihan yang rendah dibanding kluster 1 (skor rata-rata pelatihan 68), dan rating kinerja tahun sebelumnya yang kurang memuaskan (didominasi oleh rating 3). Meskipun demikian, kluster ini	Kinerja Rendah

memiliki jumlah pelatihan yang lebih banyak dibandingkan dengan kluster 0.

### 3.2. Analisis *Clustering* dengan Faktor Demografi

Berdasarkan hasil analisis kluster, teridentifikasi dua segmen karyawan dengan kinerja berbeda, yaitu karyawan berkinerja tinggi di Kluster 0 dan karyawan berkinerja rendah di Kluster 1. Selanjutnya, dilakukan evaluasi terhadap faktor demografi yaitu gender, usia, tingkat pendidikan, dan saluran rekrutmen untuk memahami pengaruhnya terhadap kinerja karyawan di masing-masing kluster.

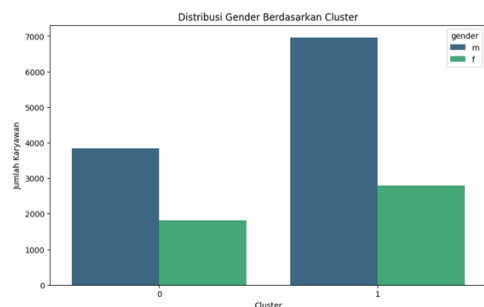
Evaluasi faktor demografi berdasarkan kluster dilakukan dengan *count plot* dan *violin plot*. *Violin plot* digunakan untuk melihat penyebaran kluster berdasarkan umur, karena umur memiliki rentang nilai yang lebih besar. Sementara itu, *count plot* digunakan untuk faktor gender, tingkat pendidikan, dan saluran rekrutmen, karena faktor-faktor ini memiliki nilai yang terbatas dan kategorikal. Hasil visualisasi ditampilkan pada Gambar 16 sampai Gambar 19.



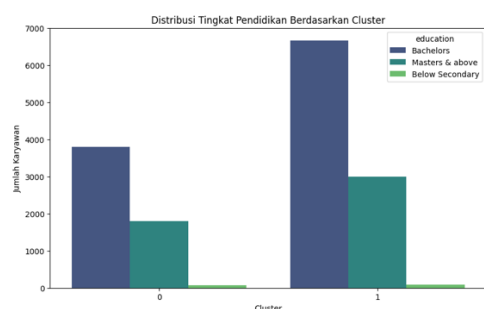
Gambar 16. Hasil *Violin Plot* berdasarkan Umur



Gambar 17. Hasil *Count Plot* berdasarkan Saluran Rekrutmen



Gambar 18. Hasil *Count Plot* berdasarkan Gender



Gambar 19. Hasil *Countplot* berdasarkan Tingkat Pendidikan

Visualisasi faktor demografi berdasarkan kluster menunjukkan bahwa

faktor demografi tidak berpengaruh signifikan, karena hasilnya serupa meskipun dengan skala yang berbeda karena jumlah data dari kluster berbeda.

#### 4. KESIMPULAN

Penelitian ini berhasil mengaplikasikan algoritma *K-Means Clustering* untuk mengelompokkan karyawan berdasarkan kinerja, dengan hasil bahwa jumlah kluster optimal adalah 2 kluster. Kluster pertama terdiri dari karyawan berkinerja tinggi dengan pencapaian *Key Performance Indicators* (KPI) di atas 80%, skor pelatihan tinggi, karyawan dengan prestasi, dan penghargaan yang diterima. Kluster kedua berisi karyawan dengan kinerja rendah, ditandai dengan KPI kurang dari 80%, skor pelatihan rendah, dan rating kinerja yang buruk. Meskipun demikian, jumlah pelatihan yang diikuti tidak menunjukkan perbedaan signifikan antara kluster. Faktor demografi seperti usia, gender, tingkat pendidikan, dan saluran rekrutmen tidak berpengaruh signifikan terhadap pengelompokan karyawan.

Berdasarkan hasil ini, organisasi dapat merancang strategi SDM yang lebih terarah, seperti memberikan penghargaan atau promosi kepada

karyawan berkinerja tinggi, serta pelatihan atau arahan lebih lanjut bagi karyawan dengan kinerja rendah. Dataset yang digunakan dalam penelitian menunjukkan kualitas kinerja yang rendah, dengan sedikit pelatihan, rating kinerja rata-rata yang berada di level 3, dan banyak karyawan yang tidak memenuhi KPI. Oleh karena itu, disarankan untuk penelitian selanjutnya menggunakan dataset yang lebih representatif dan berkualitas untuk mendapatkan wawasan yang lebih akurat.

#### DAFTAR PUSTAKA

- Alga Vredizon, P., Firmansyah, H., Shafira Salsabila, N., & Eko Nugroho, W. (2024). Penerapan Algoritma K-Means Untuk Mengelompokkan Makanan Berdasarkan Nilai Nutrisi. *Journal of Technology and Informatics (JoTI)*, 5(2), 108–115. <https://doi.org/10.37802/joti.v5i2.577>
- Ari Sulistiyawati, & Eko Supriyanto. (2021). *Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan*. 15(2). <https://doi.org/https://doi.org/10.33365/jtk.v15i2.1162>
- Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi*. <https://doi.org/10.34010/jati.v12i1>
- Dini Fadilla, & Mike Triani. (2024). *Analisis Determinan Produktivitas Tenaga Kerja Di Indonesia*. <https://doi.org/10.24036/ecosains.13122857.00>
- Faran, J., & Triayudi, A. (2024). KLIK: Kajian Ilmiah Informatika dan Komputer Penerapan Algoritma K-Means Data Mining untuk Clustering Kinerja Karyawan Koperasi. *Media Online*, 4(4). <https://doi.org/10.30865/klik.v4i4.1728>
- Gideon Bartolomeus Kaligis, & Sri Yulianto. (2022). *Analisa Perbandingan Algoritma K-Means, K-Medoids, dan X-Means untuk Pengelompokan Kinerja Pegawai (Studi Kasus: Sekretariat DPRD Provinsi Sulawesi Utara)*. <https://doi.org/10.24246/itexplore.v1i3.2022.pp179-193>
- Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM)*, 3(1), 46–56. <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Hidayat, R. S., Muttaqin, M. R., & Irmayanti, D. (2024). PENGELOMPOKAN DAERAH RAWAN BENCANA DI JAWA TENGAH MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 5). <https://doi.org/https://doi.org/10.36040/jati.v8i5.10880>
- Indhitya Padiku, & Agus Lahinta. (2022). Penerapan Clustering K-Means Untuk Mendukung Pengelolaan Koleksi Pada Perpustakaan Fakultas Teknik

- Universitas Negeri Gorontalo. *Jurnal Teknik*, 20(1), 54–62. <https://doi.org/https://doi.org/10.37031/jt.v20i1.206>
- Indra Reisandi, Daryana, Fatimah Sri Mulyati, & Muchammad Fauzi. (2021). *IMPLEMENTASI CLUSTERING-MEANSTERHADAP PENILAIAN KINERJA KARYAWAN PT.XYZ*. <https://doi.org/10.59188/jurnalsostech.v1i8.162>
- International Labour Organization. (2020). *Program Pekerjaan Layak Nasional (DWCP) untuk Indonesia 2020-2025*. [https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@asia/@ro-bangkok/@ilo-jakarta/documents/publication/wcms\\_757814.pdf](https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@asia/@ro-bangkok/@ilo-jakarta/documents/publication/wcms_757814.pdf)
- Ni Putu Ayu Widiari, I I Made Agus Dwi Suarjaya, & Dwi Putra Githa. (2020). *Teknik Data Cleaning Menggunakan Snowflake untuk Studi Kasus Objek Pariwisata di Bali*. <https://doi.org/10.24843/JIM.2020.v08.i02.p07>
- Sudaryanto. (2024). Peran KPI Dalam Mendorong Kinerja Karyawan Dan Produktivitas Organisasi. *Bisnis Dan Digital (JIMaKeBiDi)*, 1(2). <https://doi.org/https://doi.org/10.61132/jimakebidi.v1i2.67>
- Teddy Dirgantara, & Rokhedi Priyo Santoso. (2024). Sosiodemografis dan rata-rata produktivitas tenaga kerja di Indonesia. *Jurnal Kebijakan Ekonomi Dan Keuangan*, 97–108. <https://doi.org/10.20885/jkek.vol3.iss1.art13>
- Tresi Aprilia. (2024). *Klasifikasi Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor dan Metode Naive Bayes*. 4(2), 156–163. <https://doi.org/10.54259/satesi.v4i2.3167>
- Umagapi, I. T., Umaternate, B., Komputer, S., Pasca Sarjana Universitas Handayani, P., Kepegawaian Daerah Kabupaten Pulau Morotai, B., & Riset dan Inovasi, B. (2023). *Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa*. <https://doi.org/https://doi.org/10.36040/jati.v8i2.9005>