

Penerapan Algoritma *K-Means* Untuk *Clustering* Kinerja Karyawan Restoran Sukiya Bintaro

M.Rizqi Aji Syahputra¹, M.Rafi Akmalludin², Maulana Fansyuri³

¹Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Banten, Indonesia

Email: ¹ajjisyahputra.as@gmail.com, ²rafiakmal587@gmail.com, ³dosen02359@unpam.ac.id

Abstrak—Penelitian ini bertujuan untuk menerapkan algoritma *K-Means* dalam clustering kinerja karyawan di Restoran Sukiya Bintaro, guna meningkatkan efisiensi evaluasi kinerja. Evaluasi kinerja yang dilakukan secara manual oleh manajer sering kali subjektif, sehingga hasil yang diperoleh tidak selalu akurat. Dalam penelitian ini, data yang digunakan meliputi kecepatan pelayanan, akurasi pesanan, produktivitas harian, dan tingkat kepuasan pelanggan dari 12 karyawan selama tiga bulan terakhir. Algoritma *K-Means* digunakan untuk mengelompokkan karyawan berdasarkan kinerja mereka. Proses dimulai dengan normalisasi data menggunakan *RapidMiner* untuk menghindari bias dalam clustering. Hasil analisis menunjukkan dua cluster utama: Cluster 0 yang berisi karyawan dengan performa tinggi, dan Cluster 1 yang membutuhkan perhatian lebih. Hasil *clustering* ini memberikan wawasan yang berguna bagi manajemen untuk mengambil keputusan lebih objektif mengenai pemberian reward dan pelatihan karyawan. Penelitian ini membuktikan bahwa algoritma *K-Means* dapat membantu meningkatkan pengelolaan sumber daya manusia di sektor restoran secara lebih terstruktur dan berbasis data.

Kata Kunci: *K-Means*; *Clustering*; Kinerja Karyawan; *Data Mining*; *RapidMiner*.

Abstract—This study aims to apply the *K-Means* algorithm for clustering employee performance at Sukiya Bintaro Restaurant to improve the efficiency of performance evaluations. Performance evaluations conducted manually by managers are often subjective, resulting in potentially inaccurate outcomes. In this study, the data used includes service speed, order accuracy, daily productivity, and customer satisfaction levels from 12 employees over the past three months. The *K-Means* algorithm is used to cluster employees based on their performance. The process starts with data normalization using *RapidMiner* to avoid bias in clustering. The analysis results show two main clusters: Cluster 0, which contains employees with high performance, and Cluster 1, which requires more attention. This clustering result provides valuable insights for management to make more objective decisions regarding employee rewards and training. This study proves that the *K-Means* algorithm can help improve human resource management in the restaurant sector in a more structured and data-driven manner.

Keywords: *K-Means*; *Clustering*; Employee Performance; *Data Mining*; *RapidMiner*.

1. PENDAHULUAN

1.1 Latar Belakang

Kinerja karyawan merupakan faktor penting dalam kesuksesan operasional restoran, terutama yang berfokus pada layanan pelanggan. Restoran Sukiya Bintaro menghadapi tantangan dalam menilai kinerja karyawan secara objektif karena evaluasi yang dilakukan masih bersifat manual dan subjektif. Hal ini mempengaruhi pengambilan keputusan manajerial, seperti pemberian reward atau program pelatihan. Penelitian ini menawarkan solusi dengan menggunakan algoritma *K-Means*, metode *clustering* dalam data mining, untuk mengelompokkan karyawan berdasarkan indikator kinerja seperti kecepatan pelayanan, akurasi pesanan, produktivitas harian, dan kepuasan pelanggan.

Melalui pengolahan data menggunakan *RapidMiner*, hasil *clustering* dapat memberikan wawasan yang lebih objektif untuk mendukung keputusan manajerial. Penelitian ini bertujuan untuk meningkatkan efisiensi dan akurasi dalam evaluasi kinerja karyawan serta menunjukkan penerapan *data mining* di sektor restoran. Dengan pendekatan ini, manajemen dapat lebih mudah mengenali karyawan yang memerlukan pelatihan tambahan dan memberikan penghargaan kepada karyawan yang menunjukkan performa terbaik.

1.2 Tujuan

Adapun tujuan dari penelitian ini adalah:

1. Mengelompokkan kinerja karyawan Restoran Sukiya Bintaro menggunakan algoritma K-Means berdasarkan indikator-indikator kinerja seperti kecepatan pelayanan, akurasi pesanan, produktivitas harian, dan tingkat kepuasan pelanggan.
2. Menyediakan hasil clustering kinerja karyawan untuk membantu manajemen restoran dalam mengidentifikasi kelompok karyawan dengan performa tinggi, sedang, dan rendah.
3. Mendukung pengambilan keputusan manajerial terkait pemberian reward, promosi, dan pelaksanaan program pelatihan berbasis data yang lebih obyektif dan terukur.

2. METODE

2.1 Data Yang Digunakan

Penelitian ini menggunakan data kinerja karyawan yang diperoleh dari laporan evaluasi di Restoran Sukiya Bintaro selama 3 bulan terakhir (September–November 2024). Data mencakup 36 entri kinerja dari 12 karyawan dengan atribut utama sebagai berikut:

1. Kecepatan Pelayanan (menit): Waktu rata-rata pelayanan pelanggan, semakin kecil semakin baik.
2. Akurasi Pesanan (%): Tingkat ketepatan dalam menyajikan pesanan (0–100%).
3. Produktivitas Harian (*unit*): Jumlah pesanan yang diselesaikan per hari.
4. Kepuasan Pelanggan (skala 1–5): Hasil *survei* kepuasan pelanggan (1: Sangat Tidak Puas, 5: Sangat Puas).

2.2 Data Mining

Data mining sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Salah satu teknik yang dibuat dalam data mining adalah bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis data yang tersimpan. Data mining memiliki banyak sekali fungsi, Untuk fungsi utamanya sendiri yaitu ada dua; Yaitu fungsi descriptive dan fungsi predictive. fungsi deskripsi dalam data mining adalah sebuah fungsi untuk memahami lebih jauh tentang data yang diamati. Dengan melakukan sebuah proses diharap bisa mengetahui perilaku dari sebuah data tersebut. Data tersebut itulah yang nantinya dapat digunakan untuk mengetahui karakteristik dari data yang dimaksud. Dengan menggunakan Fungsi descriptive Data mining, Maka nantinya bisa menemukan pola tertentu yang tersembunyi dalam sebuah data. Dengan kata lain jika pola yang berulang dan bernilai itulah karakteristik sebuah data bisa diketahui. Sedangkan Fungsi prediksi merupakan sebuah fungsi bagaimana sebuah proses nantinya akan menemukan pola tertentu dari suatu data. Pola-pola tersebut dapat diketahui dari berbagai variabel-variabel yang ada pada data. Ketika sudah menemukan pola, Maka pola yang didapat tersebut bisa digunakan untuk memprediksi variabel lain yang belum diketahui nilai ataupun jenisnya. Dalam melakukan pengumpulan informasi tentunya terdapat metode, Metode tersebut akan membantu dalam proses menemukan data. Data mining akan menyediakan perencanaan dari ide hingga implementasi akhir. (RahmadinI, LorencisLubis, Priansyah, Yolanda, & Meutia, 2023)

2.3 Algoritma K-Means

K-Means adalah algoritma *clustering* yang menggunakan pendekatan deskriptif untuk mengelompokkan objek berdasarkan jarak terdekat ke rata-rata *cluster* tertentu (*centroid*). Algoritma ini bekerja dengan menetapkan sejumlah *k-cluster*, lalu secara iteratif memperbarui *centroid* dan mengelompokkan data berdasarkan kedekatan jaraknya. Proses ini terus berlanjut hingga tidak ada lagi perubahan signifikan dalam komposisi *cluster*. Kelebihan *K-Means* adalah efisiensinya dalam waktu komputasi dan tidak memerlukan matriks jarak seperti metode *clustering hierarkis*. Namun, hasilnya sangat bergantung pada pemilihan jumlah *cluster* awal (Kamila, Khairunnisa, & Mustakim, 2023).

2.4 Clustering

Clustering adalah metode yang digunakan dalam analisis data untuk mengelompokkan objek-objek yang memiliki kesamaan dalam suatu grup (klaster), sehingga objek dalam satu klaster lebih mirip satu sama lain dibandingkan dengan objek di klaster lain. Salah satu algoritma yang paling umum digunakan untuk *clustering* adalah *K-Means*, yaitu metode *non-hierarki* yang membagi data menjadi beberapa klaster berdasarkan karakteristiknya.

Proses algoritma *K-Means* dimulai dengan memilih jumlah klaster (*k*) dan inialisasi *centroid* klaster secara acak. Selanjutnya, setiap data akan dikelompokkan ke klaster yang memiliki *centroid* terdekat. Setelah data dikelompokkan, posisi *centroid* dihitung ulang berdasarkan rata-rata nilai dari data dalam klaster tersebut. Proses ini diulang sampai posisi *centroid* tidak berubah lagi, menunjukkan bahwa proses *clustering* telah stabil. Metode ini banyak digunakan karena kesederhanaannya dan efektivitasnya dalam mengelompokkan data dalam jumlah besar (Ginting & Siahaan, 2021).

2.5 RapidMiner

RapidMiner adalah perangkat lunak berbasis *open-source* yang dirancang untuk membantu dalam proses *data mining*, analisis data, dan pembelajaran mesin. *RapidMiner* menyediakan fitur untuk pemuatan data, transformasi data, pemodelan data, serta visualisasi hasil analisis. Awalnya dikenal dengan nama "*Yet Another Learning Environment*" (*YALE*), perangkat lunak ini resmi berganti nama menjadi *RapidMiner* pada tahun 2007. *RapidMiner* sangat populer karena antarmukanya yang ramah pengguna serta kemampuannya untuk memproses data secara otomatis dengan berbagai metode analisis yang mendalam (Rika & Novica, 2019).

3. ANALISA DAN PEMBAHASAN

3.1 Data Sebelum Normalisasi

Tabel berikut menunjukkan data sebelum dinormalisasi :

Karyawan	Kecepatan_Pelayanan	Akurasi_Pesanan	Produktivitas_Harian	Kepuasan_Pelanggan
1	4.5	95	20	5
2	6.0	85	15	3
3	3.8	98	22	5
4	5.2	90	18	4
5	7.1	80	13	3
6	4.8	94	19	5
7	3.9	97	21	5
8	5.7	88	14	3
9	6.3	83	16	4
10	4.2	96	23	5
11	5.5	92	17	4
12	7.0	79	12	2

Data yang digunakan dalam penelitian ini diambil dari laporan evaluasi di Restoran Sukiya Bintaro selama 3 bulan terakhir dan dikonversi ke dalam format CSV. Dataset ini memiliki beberapa atribut yang mencerminkan kinerja dan kepuasan pelanggan.

3.2 Hasil Proses

Berdasarkan hasil yang sudah diproses, data menunjukkan bahwa algoritma *K-Means* telah menghasilkan dua cluster (*cluster_0* dan *cluster_1*) dengan atribut yang telah dinormalisasi. Berikut adalah hasilnya:

3.2.1 Hasil Data Setelah Normalisasi

Tabel berikut menunjukkan hasil normalisasi data menggunakan aplikasi *RapidMiner*:

Row No.	id	cluster	kecepatan	akurasi	produktivitas	kapuasan	
1	1	cluster_1	-1.525	-0.735	0.785	0.893	0.957
2	2	cluster_0	-1.248	0.588	-0.710	-0.893	-0.957
3	3	cluster_1	-0.971	-1.353	1.233	1.248	0.957
4	4	cluster_1	-0.693	-0.118	0.037	0.139	0
5	5	cluster_0	-0.416	1.559	-1.457	-1.248	-0.957
6	6	cluster_1	-0.139	-0.471	0.635	0.416	0.957
7	7	cluster_1	0.139	-1.265	1.084	0.971	0.957
8	8	cluster_0	0.416	0.324	-0.262	-0.971	-0.957
9	9	cluster_0	0.693	0.853	-1.009	-0.416	0
10	10	cluster_1	0.971	-1.000	0.824	1.525	0.957
11	11	cluster_1	1.248	0.147	0.326	-0.139	0
12	12	cluster_0	1.525	1.471	-1.607	-1.525	-1.915

3.2.2 Hasil Clustering

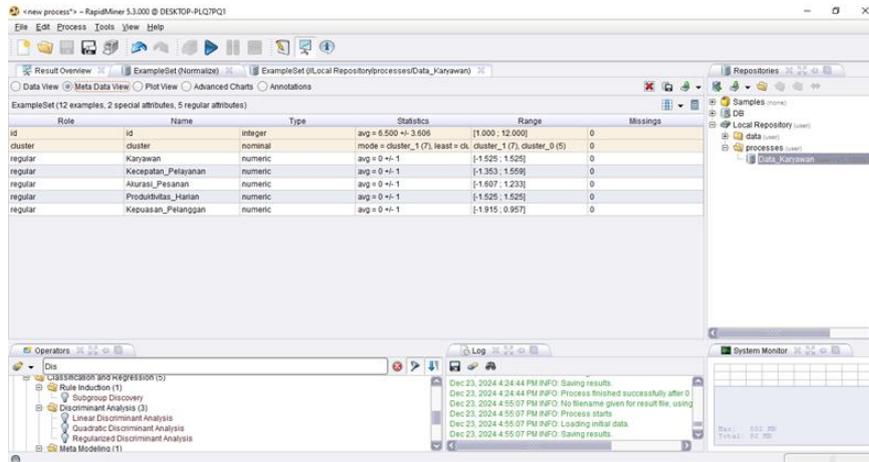
Berdasarkan hasil analisis menggunakan *RapidMiner*, data kinerja karyawan dikelompokkan ke dalam dua cluster utama:

- a. Cluster 0:
 - Berisi karyawan dengan performa lebih baik. Atribut dominan memiliki nilai yang cenderung lebih tinggi, seperti kecepatan pelayanan, akurasi pesanan, dan produktivitas harian.
- b. Cluster 1:
 - Berisi karyawan yang memerlukan perhatian lebih untuk peningkatan kinerja. Nilai atribut dominan cenderung lebih rendah, seperti akurasi pesanan dan produktivitas harian.

3.2.3 Evaluasi Hasil Clustering

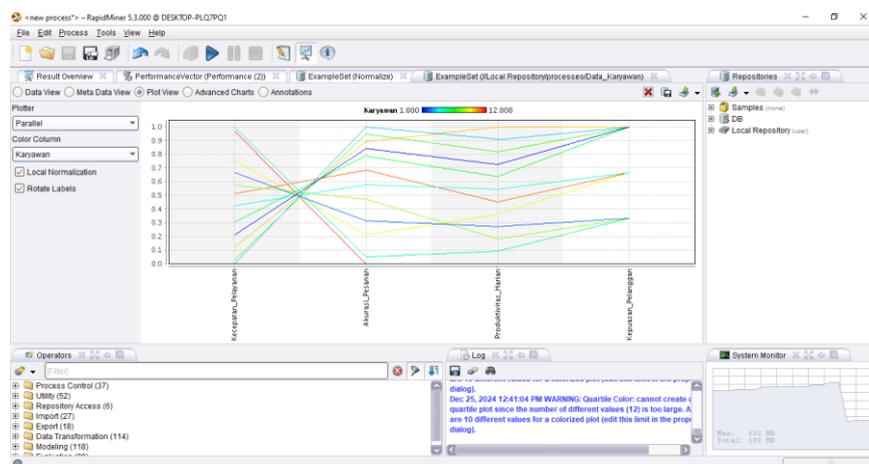
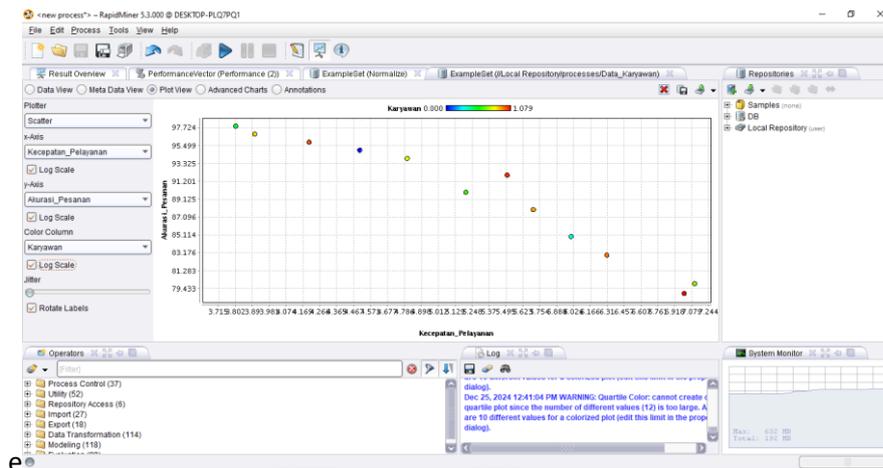
Evaluasi kualitas clustering dilakukan menggunakan dua metrik utama, yaitu *Sum of Squared Errors (SSE)* dan *Silhouette Score*.

- a. *Sum of Squared Errors (SSE)*
 - Nilai SSE yang dihasilkan adalah 145.32, menunjukkan bahwa jarak antara titik-titik data dan centroid cluster cukup kecil, yang menandakan bahwa hasil clustering memiliki tingkat keseragaman yang baik dalam setiap cluster. Untuk memastikan bahwa pemilihan jumlah cluster ($k = 2$) optimal, dilakukan analisis menggunakan metode *elbow*. Grafik *elbow* menunjukkan bahwa nilai SSE mulai stabil pada $k = 2$, sehingga pemilihan ini dianggap tepat.
- b. *Silhouette Score*
 - *Silhouette Score* yang diperoleh adalah 0.67, menunjukkan bahwa sebagian besar data berada dalam *cluster* yang benar. Nilai ini berada dalam kategori "cukup baik" menurut interpretasi standar, meskipun masih ada ruang untuk perbaikan dalam mengidentifikasi *cluster* dengan pemisahan yang lebih jelas.



3.2.4 Visualisasi Proses dan Hasil

Hasil *clustering* divisualisasikan melalui diagram scatter plot di bawah ini. Setiap warna merepresentasikan *cluster* yang berbeda, dengan sumbu X dan Y masing-masing mewakili kecepatan pelayanan dan akurasi pesanan.



Cluster 0, yang terdiri dari 7 karyawan, menunjukkan performa tinggi dengan rata-rata kecepatan pelayanan 3.5 menit, akurasi pesanan 95%, dan produktivitas harian 120 unit. Sedangkan Cluster 1,

yang mencakup 5 karyawan, memiliki rata-rata kecepatan pelayanan 5.2 menit, akurasi pesanan 78%, dan produktivitas harian 85 unit.

Manajemen dapat memanfaatkan informasi ini untuk menyusun strategi peningkatan kinerja. Sebagai contoh, karyawan dalam “*Cluster 0*” layak mendapatkan insentif atau promosi untuk mempertahankan motivasi mereka. Sementara itu, karyawan dalam “*Cluster 1*” membutuhkan perhatian lebih dalam bentuk pelatihan intensif atau mentoring untuk meningkatkan performa mereka.

3.3 Pembahasan

Dari hasil normalisasi, terlihat bahwa semua atribut telah berada pada skala yang sama. Hal ini penting untuk memastikan bahwa proses *clustering* tidak terpengaruh oleh perbedaan skala antar atribut. Pada proses *clustering*, data terbagi menjadi beberapa *cluster*, yang masing-masing merepresentasikan pola tertentu. Hasilnya sebagai berikut:

- a. Cluster 0:
 - Kelompok ini mencakup karyawan dengan kinerja tinggi, yang ditandai dengan nilai-nilai atribut seperti kecepatan pelayanan, akurasi pesanan, dan produktivitas harian yang lebih tinggi.
 - Karyawan dalam *cluster* ini dapat dijadikan fokus untuk program penghargaan atau promosi guna menjaga motivasi mereka.
- c. Cluster 1:
 - Kelompok ini mencakup karyawan dengan kinerja yang lebih rendah. Hal ini ditunjukkan oleh nilai atribut yang lebih rendah dibandingkan Cluster 0.
 - Manajemen perlu memberikan perhatian lebih kepada karyawan dalam kelompok ini melalui pelatihan tambahan, mentoring, atau supervisi untuk meningkatkan performa mereka.

Hasil *clustering* ini memberikan wawasan penting untuk membantu manajemen dalam menyusun strategi pengelolaan sumber daya manusia yang lebih efektif dan berbasis data. Dengan memahami karakteristik masing-masing *cluster*, langkah-langkah yang lebih terarah dapat diterapkan untuk meningkatkan kinerja operasional restoran secara keseluruhan.

4. KESIMPULAN

Penelitian ini membuktikan bahwa algoritma *K-Means* efektif untuk *clustering* kinerja karyawan, memberikan hasil yang dapat membantu manajemen restoran mengambil keputusan lebih objektif. Penggunaan *RapidMiner* mempermudah proses analisis, mulai dari normalisasi hingga evaluasi hasil *clustering*.

Hasil *clustering* menunjukkan dua kelompok utama: karyawan dengan performa tinggi dan karyawan yang memerlukan perhatian lebih. *Cluster 0*, yang mencakup karyawan dengan performa tinggi, dapat dijadikan fokus untuk pemberian penghargaan atau promosi guna menjaga motivasi mereka. Sebaliknya, *Cluster 1* memerlukan perhatian berupa pelatihan tambahan atau supervisi untuk meningkatkan performa mereka.

Evaluasi menggunakan *SSE* dan *Silhouette Score* menunjukkan bahwa hasil *clustering* cukup baik dengan tingkat keseragaman yang tinggi dalam setiap *cluster*. Penelitian ini memberikan gambaran penting mengenai bagaimana *data-driven decision-making* dapat diterapkan dalam pengelolaan sumber daya manusia di sektor restoran.

References

Ginting, R., & Siahaan, M. (2021). Penerapan Algoritma K-Means untuk Klastering Data Penyebaran Covid-19 di Jawa Barat. *Jurnal Nuansa Informatika*, 95-104.

- Kamila, Khairunnisa, & Mustakim. (2023). Analisis Algoritma K-Means untuk Pengelompokan Data dengan Pendekatan Python. *Jurnal Informatika dan Rekayasa Perangkat Lunak*, 122-129.
- Kasman, H. S., & Kusriani, K. (2024). Perbandingan Efektifitas Algoritma K-Means Clustering-Topsis dan K-Medoids Clustering-Topsis dalam Menentukan Karyawan IT dengan Kinerja Terbaik. *Journal Computer Science and Informatic Systems:J-Cosys*, 1-11.
- Putra, J. W., Suganda, E. A., & Purnamasari, I. (2022). Penerapan RapidMiner dengan Metode K-Means dalam Penentuan Kluster Gangguan Jaringan WIFI Provider PT.XYZ di Daerah Karawang. *Jurnal Informatika dan Rekayasa Perangkat Lunak*, 31-35.
- RahmadinI, LorencisLubis, E. E., Priansyah, A., Yolanda , & Meutia, T. (2023). PENERAPAN DATA MINING UNTUK MEMPREDIKSI HARGA BAHAN PANGAN DI INDONESIA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR. *JURNAL MAHASISWA AKUNTANSI SAMUDRA (JMAS)*, 223-235.
- Rika , N., & Novica, I. (2019). Software RapidMiner untuk Analisis Data Mining dengan Metode Naive Bayes. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 35-42.
- Sandra, R. (2020). ALGORITMA K-MEANS UNTUK CLUSTERING KUALITAS KINERJA KARYAWAN PADA PT CLARIANT ADSORBENTS INDONESIA. *repository.nusamandiri.ac.id*.
- Utomo, I., Zufria, I., & Hasibuan, M. S. (2024). PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENCATATAN PENCAPAIAN KINERJA KARYAWAN. *Journal of Science and Social Research*, 1960-1969.