

Analisis Pengaruh Distance Metric Pada Algoritma K-Medoids Dalam Pengelompokan Kinerja Guru Menggunakan Silhouette Coefficient

Rosmini^{1*}, Muhammad Fadlan², Sinawati³

^{1,2,3}Program Studi Sistem Informasi, STMIK PPKIA Tarakanita Rahmawati

*rosmini@ppkia.ac.id

Abstrak

Penelitian ini bertujuan untuk menganalisis pengaruh distance metric pada algoritma K-Medoids dalam pengelompokan kinerja guru berdasarkan indikator pedagogik, kepribadian, sosial, dan profesional. Penelitian ini dilakukan karena penelitian sebelumnya pada pengelompokan kinerja guru umumnya lebih berfokus pada penggunaan algoritma clustering tanpa mengevaluasi secara komparatif pengaruh pemilihan distance metric terhadap kualitas cluster yang dihasilkan. Data penelitian diperoleh dari hasil penilaian asesor terhadap kinerja guru di SD Islam Al-Irsyad sebanyak 16 data guru yang terlebih dahulu dinormalisasi untuk menyamakan skala antar atribut sebelum dilakukan proses clustering. Pengelompokan dilakukan menggunakan algoritma K-Medoids dengan metode jarak Euclidean, Manhattan, dan Cosine, serta variasi jumlah cluster (K) dari 2 hingga 5 dengan bantuan RapidMiner. Evaluasi hasil clustering dilakukan menggunakan metode Silhouette Coefficient yang dihitung dengan bantuan Python untuk menentukan kualitas cluster dan jumlah cluster optimal. Hasil penelitian menunjukkan bahwa nilai Silhouette Coefficient tertinggi sebesar 0,290 diperoleh pada K=4 menggunakan metode Euclidean dan Manhattan. Berdasarkan hasil tersebut, metode Euclidean dipilih sebagai metode terbaik karena lebih umum digunakan dan mampu merepresentasikan jarak antar data secara lebih efektif. Hasil clustering kemudian dianalisis dengan memberikan label pada setiap cluster berdasarkan nilai rata-rata indikator untuk mempermudah interpretasi. Dengan demikian, penelitian ini menunjukkan bahwa pemilihan distance metric berpengaruh terhadap kualitas hasil clustering pada pengelompokan kinerja guru.

Kata kunci : K-Medoids, Euclidean Distance, Manhattan Distance, Cosine Similarity, Silhouette Coefficient, Pengelompokan

Abstract

This study aims to analyze the effect of distance metrics on the K-Medoids algorithm in clustering teacher performance based on pedagogical, personality, social, and professional indicators. This study was conducted because previous research on teacher performance clustering has generally focused more on the use of clustering algorithms without comparatively evaluating the effect of distance metric selection on the quality of the resulting clusters. The research data were obtained from assessor evaluations of teacher performance at SD Islam Al-Irsyad involving 16 teacher data records, which were first normalized to standardize the scale among attributes before the clustering process was carried out. Clustering was performed using the K-Medoids algorithm with Euclidean, Manhattan, and Cosine distance methods, as well as variations in the number of clusters (K) from 2 to 5 using RapidMiner. The clustering results were evaluated using the Silhouette Coefficient method calculated with Python to determine the cluster quality and the optimal number of clusters. The results showed that the highest Silhouette Coefficient value of 0.290 was obtained at K=4 using the Euclidean and Manhattan methods. Based on these results, the Euclidean method was selected as the best method because it is more widely used and capable of representing the distance between data points more effectively. The clustering results were then analyzed by assigning labels to each cluster based on the average value of the indicators to facilitate interpretation. Thus, this study demonstrates that the selection of distance metrics affects the quality of clustering results in teacher performance grouping.

Keywords : K-Medoids, Euclidean Distance, Manhattan Distance, Cosine Similarity, Silhouette Coefficient, Clustering

1. Pendahuluan

Data mining merupakan suatu proses penggalian informasi dari sekumpulan data dengan tujuan memperoleh pengetahuan yang dapat dimanfaatkan pada berbagai bidang. Dalam praktiknya, data mining dapat digunakan untuk berbagai keperluan analisis, seperti klasifikasi, asosiasi, prediksi dan clustering. Salah satu teknik yang banyak digunakan adalah clustering yaitu proses pengelompokan data berdasarkan tingkat kemiripan karakteristik yang dimiliki tanpa menggunakan label tertentu^{[1],[2]}. Clustering adalah metode untuk mengidentifikasi struktur alami dalam data yang tidak berlabel, sehingga memungkinkan pengelompokan objek serupa berdasarkan kedekatan fitur multidimensional mereka^[3]. Dalam bidang Pendidikan, teknik clustering dimanfaatkan untuk menganalisis data penilaian kinerja guru guna mengidentifikasi pola karakteristik yang terbentuk, dengan adanya pengelompokan tersebut, pihak terkait dapat memperoleh gambaran yang lebih jelas mengenai distribusi kinerja guru, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan, evaluasi maupun perumusan strategi peningkatan kualitas Pendidikan^[4].

Salah satu algoritma clustering yang dapat digunakan adalah K-Medoids. Algoritma ini merupakan pengembangan dari K-Means yang memiliki keunggulan dalam menangani data yang mengandung noise dan outlier, karena

menggunakan objek data sebagai pusat cluster (medoid). Meskipun demikian, hasil clustering menggunakan K-Medoids sangat dipengaruhi oleh metode perhitungan jarak (distance matrix) yang digunakan dalam proses pengelompokan.^[5] Beberapa metode perhitungan jarak yang umum digunakan antara lain Euclidean, Manhattan dan Cosine. Masing-masing metode memiliki karakteristik yang berbeda dalam mengukur kedekatan antar data. Perbedaan tersebut dapat mempengaruhi hasil pengelompokan yang dihasilkan, baik dari segi struktur cluster maupun tingkat akurasi pemisahan data. Sehingga diperlukan analisis untuk mengetahui metode perhitungan jarak yang paling sesuai dengan sesuai dalam menghasilkan kualitas cluster yang optimal^[6]. Untuk mengevaluasi kualitas hasil clustering, salah satu metode yang digunakan Silhouette Coefficient. Metode ini mampu mengukur tingkat kesesuaian suatu data terhadap cluster lain, sehingga dapat memberikan gambaran mengenai kualitas pemisahan dan kohesi antar cluster. Nilai Silhouette yang tinggi menunjukkan bahwa data telah dikelompokkan dengan baik.^{[6], [7]}. Meskipun penelitian terkait clustering telah banyak dilakukan menggunakan berbagai algoritma dan metode jarak, penelitian yang secara khusus membandingkan pengaruh metode jarak Euclidean, Manhattan dan Cosine pada algoritma K-Medoids dalam pengelompokan kinerja guru masih terbatas. Selain itu, belum

banyak penelitian yang mengevaluasi kualitas hasil clustering menggunakan Silhouette Coefficient pada data penilaian kinerja guru. Oleh karena itu, penelitian ini dilakukan untuk mengetahui metode jarak yang menghasilkan kualitas cluster terbaik

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis pengaruh penggunaan distance metric terhadap kualitas hasil clustering menggunakan algoritma K-Medoids pada data penilaian kinerja guru. Evaluasi dilakukan menggunakan Silhouette Coefficient untuk membandingkan performa masing-masing metode jarak. Hasil penelitian diharapkan dapat memberikan rekomendasi dalam pemilihan distance metric yang tepat untuk meningkatkan kualitas hasil clustering.

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Untuk menunjang penelitian yang dilakukan penulis menggunakan beberapa artikel terkait sebagai referensi dalam penelitian ini. Berikut ini merupakan beberapa penelitian terlebih dahulu yang relevan dengan topik penelitian yang dilakukan.

1. Penelitian pertama dilakukan oleh Nita Mirtantika (2023), melakukan segmentasi pelanggan pada XYZ Store dengan menggunakan teknik data mining clustering yaitu algoritma K-Medoids dan analisis RFM (recency,

frequency dan monetary). Selain itu pada penelitian ini menggunakan metode Silhouette Coefficient untuk penentuan jumlah cluster optimal dan evaluasi cluster menggunakan metode Davies-Boulding Index (DBI). Hasil penelitian segmentasi pelanggan diperoleh tiga jenis pelanggan yaitu loyal customer sebanyak 21 pelanggan, typical customer sebanyak 31 pelanggan dan newcomer sebanyak 61 pelanggan. Dengan adanya segmentasi ini, maka Perusahaan dapat membuat kebijakan dan strategis bisnis dengan pelanggan agar mendapatkan keuntungan yang lebih besar.^[8]

2. Penelitian kedua dilakukan oleh Solikhun (2025), dimana pada penelitian menggunakan algoritma K-Medoids yang dihitung menggunakan persamaan kuantum serta membandingkan dua metode perhitungan jarak yaitu Manhattan dan Chebyshev. Hasil pada penelitian ini menunjukkan bahwa algoritma Manhattan lebih unggul dalam konteks medis, terutama untuk deteksi kasus positif dengan recall 0,57 dan F1-score 0.695 serta nilai precision lebih rendah 0,88. Oleh karena itu Manhattan lebih cocok digunakan untuk aplikasi medis dimana false negative beresiko tinggi, seperti deteksi penyakit meskipun biaya dan MSE lebih tinggi. Sedangkan Chebyshev memiliki nilai precision sempurna dan akurasi 80% serta nilai recall lebih rendah 0.33 dimana hal ini menunjukkan banyak kasus positif yang terlewatkan.^[9]

3. Penelitian ketiga dilakukan oleh Anisa Putri Ayuni (2024), pada penelitian dilakukan analisa hasil pengelompokan kabupaten/kota berdasarkan indikator pembentuk IPM (Indeks Pembangunan Manusia) menggunakan K-Medoids dan CLARA. Metode Silhouette Coefficient digunakan untuk perbandingan cluster optimal pada algoritma K-Medoids dan CLARA. Hasil pengelompokan dengan K-Medoids diperoleh 358 kabupaten/kota yang bergabung dalam cluster nilai IPM sedang dan 156 kabupaten/kota tergabung dalam cluster nilai IPM tinggi. Sedangkan pengelompokan menggunakan CLARA diperoleh 426 kabupaten/kota yang tergabung dalam cluster dengan nilai IPM sedang dan 88 kabupaten/kota dalam cluster nilai IPM tinggi. Oleh karena itu, CLARA merupakan algoritma yang optimal digunakan dalam penelitian ini karena memiliki nilai silhouette coefficient paling tinggi yaitu 0,438 dibandingkan dengan K-Medoids.^[10]

4. Penelitian Keempat dilakukan oleh Dinda Ramadhaniaty dimana pada penelitian ini menggunakan K-Medoids untuk mengelompokan status karyawan kontrak yang layak diperpanjang atau tidak layak secara objektif sehingga mempermudah perusahaan dalam menganalisis data agar hasil lebih tepat dan dapat menghindari kecurangan dalam prosesnya. Pada penelitian ini menggunakan tiga metode perhitungan jarak yaitu Euclidean Distance, Manhattan Distance

dan Chebyshev Distance dan untuk uji akurasi menggunakan Silhouette Coefficient. Hasil penelitian diperoleh nilai Silhouette Coefficient untuk chebyshev distance menghasilkan nilai yang mendekati 1 dengan nilai 0,3220214. Sedangkan, hasil nilai cluster SSE diperoleh bahwa nilai error terkecil dimiliki oleh 2 cluster serta hasil pengelompokan yang diperoleh yaitu 20 data ke dalam kategori Tidak Layak dan 22 data ke dalam kategori Layak.^[11]

Berdasarkan penelitian terdahulu, pemilihan metode jarak dan algoritma clustering berpengaruh terhadap kualitas hasil pengelompokan data. Selain itu, metode evaluasi Silhouette Coefficient banyak digunakan untuk menentukan jumlah cluster optimal dan mengukur kualitas cluster yang dihasilkan. Oleh karena itu, penelitian ini berfokus pada analisis pengaruh distance matrix pada algoritma K-Medoids dalam pengelompokan kinerja guru menggunakan evaluasi Silhouette Coefficient..

2.2. Landasan Teori

1. Data Mining

Data mining merupakan sekumpulan teknik untuk mengesktrasikan pengetahuan yang bermanfaat dengan cara menemukan pola-pola tertentu dari kumpulan data, sehingga dapat menghasilkan informasi bernilai tambah yang sebelumnya tidak diketahui.^{[12],[13]}

2. Clustering

Clustering adalah mengelompokkan sekumpulan data menjadi beberapa kelompok (cluster) dari data yang belum diketahui kelas atau kelompok sebelumnya, serta menentukan keanggotaan setiap data ke dalam cluster tertentu. Cluster sendiri merupakan suatu metode atau kelas dalam data, yang dapat digunakan untuk mengidentifikasi struktur taksonomi atau hubungan antar data yang ada.^{[12], [13]}

3. K-Medoids

K-Medoids atau Partitioning Around Medoids (PAM) merupakan algoritma clustering yang menentukan pusat cluster dimana setiap cluster menggunakan sebuah objek. Objek cluster merupakan Medoid yang paling central dan memiliki jarak paling dekat ke titik lain. Penentuan awal nilai K pada K-Medoids dilakukan secara acak atau random. Selain itu K-Medoids sangat efisien untuk Kumpulan data kecil. Kemudian data pada cluster yang sudah ada akan diperbarui dengan memperbarui medoid yang mewakili tiap cluster. Iterasi dilakukan sampai medoid tidak ada perubahan. Berikut tahapan dari K-Medoids.^[11]

Menentukan jumlah cluster (k) serta melakukan inisialisasi pusat cluster (medoid) sebanyak k

Mengalokasikan setiap data atau objek kedalam cluster terdekat menggunakan persamaan ukur jarak. Menghitung jarak antara setiap objek ke dalam cluster dengan kandidat medoid yang baru
Menghitung nilai total simpangan (S), yaitu selisih antara total jarak baru dan total jarak sebelumnya,

Apabila nilai $S < 0$, maka kandidat medoid akan menggantikan medoid sebelumnya sehingga diperoleh hasil cluster yang stabil beserta anggota masing-masing cluster.

4. Euclidean Distance

Euclidean Distance merupakan metode perhitungan jarak yang paling umum digunakan, yaitu untuk mengukur jarak antara dua titik dalam ruang Euclidean yang mempresentasikan hubungan antara jarak dan sudut antar titik.^{[11], [14]}

5. Manhattan Distance

Manhattan Distance merupakan metode perhitungan jarak yang digunakan untuk mengukur nilai absolut antara koordinat dari sepasang objek.^{[11], [14]}

6. Cosine Similarity

Cosine Similarity adalah metode yang digunakan untuk mengukur tingkat kemiripan antara dua objek data dengan menghitung nilai cosinus dari sudut yang terbentuk antara dua vektor, sehingga mencerminkan tingkat kesamaan arah antar data.^{[14], [15]}

7. Silhouette Coefficient

Silhouette Coefficient merupakan metode yang digunakan untuk menentukan jumlah cluster yang optimal berdasarkan nilai rata-rata silhouette. Metode ini mengevaluasi kualitas hasil pengelompokan dengan membandingkan rata-rata jarak suatu data terhadap data lain dalam cluster yang sama dengan jarak terhadap data pada cluster yang berbeda. Semakin tinggi nilai

Silhouette Coefficient untuk suatu jumlah K, maka semakin baik kualitas cluster yang terbentuk dan semakin optimal jumlah cluster tersebut.[6], [16], [17]

8. Kinerja Guru

Kinerja Guru adalah hasil kerja yang dicapai guru dalam melaksanakan tugas profesionalnya di sekolah, terutama dalam kegiatan pembelajaran. Kinerja ini mencerminkan kemampuan guru dalam menjalankan perannya secara efektif dan sesuai dengan standar pendidikan yang berlaku. Penilaian kinerja guru dalam penelitian ini menggunakan empat indikator utama, yaitu kompetensi pedagogik, kepribadian, sosial dan profesional. Kompetensi kepribadian mencerminkan sikap dan perilaku guru sebagai teladan. Kompetensi sosial berkaitan dengan kemampuan berinteraksi dengan siswa, rekan kerja dan masyarakat. Sementara itu, kompetensi profesional berhubungan dengan penguasaan materi ajar dan pengembangan diri sebagai pendidik. [18]

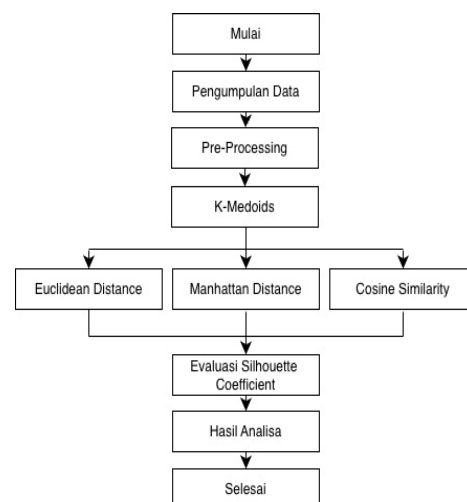
9. RapidMiner

RapidMiner adalah perangkat lunak berbasis visual untuk data mining yang memudahkan proses pengolahan dan analisis data tanpa pemrograman yang kompleks. RapidMiner menyediakan berbagai operator seperti preprocessing, clustering dan analisis data lainnya. Dalam penelitian ini, RapidMiner digunakan untuk penerapan algoritma K-Medoids dengan berbagai distance metric.[13]

3. Metode Penelitian

3.1. Tahapan Penelitian

Metode penelitian ini menjelaskan langkah-langkah yang digunakan dalam proses pengolahan data untuk mencapai tujuan penelitian. Pada penelitian ini, metode yang digunakan adalah pendekatan kuantitatif dengan teknik data mining clustering menggunakan algoritma K-Medoids. Penelitian ini bertujuan untuk menganalisis pengaruh penggunaan distance metric yang berbeda terhadap hasil pengelompokan data kinerja guru. Seluruh proses penelitian dilakukan melalui tahapan yang sistematis mulai dari pengumpulan data hingga evaluasi hasil clustering menggunakan Silhouette Coefficient. Berikut adalah tahapan penelitian untuk pengelompokan keaktifan anggota dengan menggunakan algoritma K-Medoids :



Gambar 1. Tahapan Penelitian

1. Pengumpulan Data

Data dalam penelitian ini dilakukan dengan metode dokumentasi, yaitu menggunakan data

sekunder berupa hasil penilaian asesor pada periode tertentu. Data yang diperoleh terdiri dari 14 kriteria penilaian dan 4 indikator utama, yaitu pedagogik, kepribadian, sosial dan professional yang digunakan sebagai variabel dalam proses analisis clustering

2. Pre-Processing

Tahap ini preprocessing data meliputi data reduction, data cleaning dan data transformation. Data transformation merupakan proses normalisasi menggunakan metode Min-Max. Berikut rumus yang digunakan yaitu pada persamaan 1-[14], [19]

3. Penentuan Parameter

Pada tahap ini dilakukan penentuan parameter yang digunakan dalam proses clustering menggunakan algoritma K-Medoids. Parameter yang digunakan meliputi jenis distance metric yaitu Euclidean Distance, Manhattan Distance dan Cosine Similarity serta jumlah cluster (K) yang ditentukan pada rentang $K = 2$ hingga $K = 5$. Setiap kombinasi distance metric dan nilai K digunakan untuk mengetahui pengaruhnya terhadap kualitas hasil clustering yang diukur menggunakan Silhouette Coefficient.

4. Penerapan Algoritma K-Medoids

Pada tahap ini dilakukan proses clustering data kinerja guru menggunakan algoritma K-Medoids dengan bantuan RapidMiner. Proses clustering dilakukan menggunakan beberapa jenis distance metric yaitu Euclidean Distance, Manhattan

Distance dan Cosine Similarity serta variasi jumlah cluster (K) dari 2 hingga 5. Setiap kombinasi parameter dijalankan untuk menghasilkan kelompok data berdasarkan tingkat kemiripan karakteristik kinerja guru.

5. Evaluasi Hasil Clustering

Pada tahap ini dilakukan evaluasi terhadap hasil clustering menggunakan metode Silhouette Coefficient dengan bantuan Python. Evaluasi dilakukan untuk mengukur kualitas cluster yang dihasilkan dari setiap kombinasi distance metric dan jumlah cluster (K). Nilai Silhouette Coefficient digunakan untuk mengetahui tingkat kedekatan data dalam satu cluster dan perbedaan antar cluster.

6. Analisis Hasil

Tahap analisis hasil dilakukan dengan membandingkan nilai Silhouette Coefficient dari seluruh kombinasi distance metric dan jumlah cluster yang digunakan. Hasil analisis digunakan untuk menentukan distance metric dan nilai K yang menghasilkan kualitas clustering terbaik pada data kinerja guru.

4. Hasil dan Pembahasan

4.1. Hasil Penelitian

1. Data Penilaian Kinerja Guru

Pada bagian ini disajikan data awal yang digunakan dalam penelitian sebelum dilakukan proses normalisasi. Data tersebut merupakan hasil pengolahan dari 14 kriteria penilaian yang

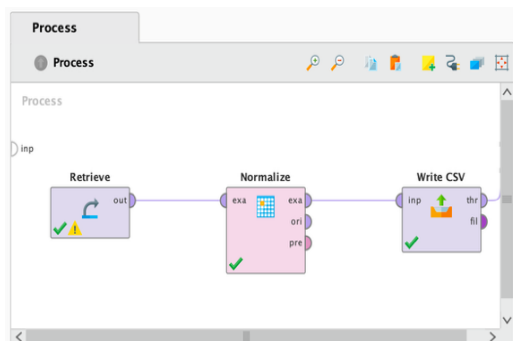
telah diagregasi menjadi 4 indikator utama, yaitu pedagogik, kepribadian, sosial dan profesional. Penyajian data awal ini bertujuan untuk memberikan gambaran nilai asli dari masing-masing indikator sebelum dilakukan proses normalisasi, sehingga perbedaan skala antar atribut dapat terlihat jelas.

Tabel 1. Contoh Data Penilaian Kinerja Guru

No	Alternatif	Ped	Kep	Sos	Pro
1	G001	14,29	16	7	11
2	G002	16	14,67	7	10,5
3	G003	16	15	6,5	11
4	G004	16,57	14,67	6,5	10,5
5	G005	16,14	14,33	7	10
6	G006	16,14	13,67	6,5	12
7	G007	16	14	7	10,5
8	G008	14,71	15	7	11
9	G009	15,29	15	7	11
10	G010	15,57	14	6,5	12,5
11	G011	16	14,67	8	11
12	G012	16,14	15,67	8	11
13	G013	16	14,67	7,5	11
14	G014	14,86	15,33	7	11
15	G015	15,43	15,67	7	9
16	G016	14,57	14,33	7,5	11

2. Proses Normalisasi dengan RapidMiner

Pada bagian ini ditampilkan data hasil normalisasi yang dilakukan dengan RapidMiner.



Gambar 2. Pemodelan Min-Max Normalisasi

Normalisasi bertujuan untuk menyamakan skala nilai pada setiap indikator, sehingga setiap atribut

memiliki kontribusi yang seimbang dalam proses analisis.

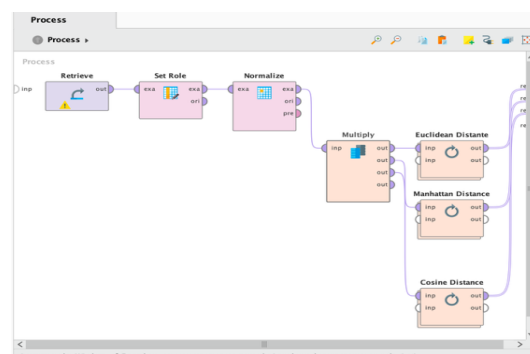
Row No.	Pedagogik	Kepribadian	Sosial	Profesional	Alternatif
1	-1.948	1.846	-0.131	0.164	G001
2	0.581	-0.187	-0.131	-0.491	G002
3	0.581	0.317	-1.175	0.164	G003
4	1.424	-0.187	-1.175	-0.491	G004
5	0.788	-0.707	-0.131	-1.146	G005
6	0.788	-1.716	-1.175	1.473	G006
7	0.581	-1.211	-0.131	-0.491	G007
8	-1.326	0.317	-0.131	0.164	G008
9	-0.469	0.317	-0.131	0.164	G009
10	-0.055	-1.211	-1.175	2.128	G010
11	0.581	-0.187	1.958	0.164	G011
12	0.788	1.341	1.958	0.164	G012
13	0.581	-0.187	0.914	0.164	G013
14	-1.105	0.822	-0.131	0.164	G014
15	-0.262	1.341	-0.131	-2.455	G015
16	-1.534	-0.707	0.914	0.164	G016

Gambar 3. Hasil Normalisasi

Data pada gambar 2 ini merupakan data hasil yang diperoleh dari tahap pre-processing data menggunakan 4 indikator utama dengan 16 data guru yang dinormalisasi menggunakan RapidMiner dengan metode Min-Max Normalization. Data ini kemudian digunakan sebagai dasar dalam proses pengelompokan menggunakan algoritma K-Medoids.

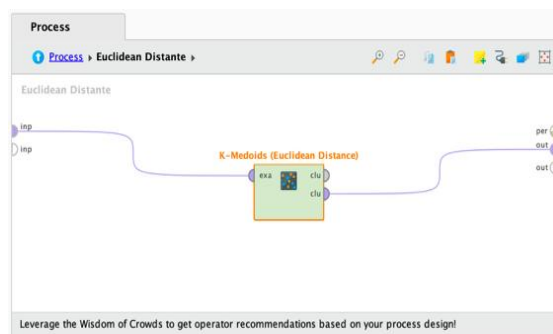
3. Proses Penerapan K-Medoids

Pada pada tahap ini dilakukan proses clustering di RapidMiner dengan 3 perhitungan jarak yaitu Euclidean, Manhattan dan Cosine. Hasil proses ditampilkan pada gambar berikut :



Gambar 4. Proses perhitungan jarak menggunakan tiga metode

Pada gambar diatas proses pengolahan data dirancang dalam satu alur kerja agar seluruh perhitungan dapat dilakukan secara otomatis dalam satu kali proses. Data terlebih dahulu dinormalisasikan, kemudian hasilnya digunakan sebagai input untuk 3 Loop Paramter yang mempresentasikan metode perhitungan jarak, yaitu Euclidean, Manhattan dan Cosine. Di dalam masing-masing Loop Parameter terdapat algoritma K-Medoids seperti pada gambar berikut



Gambar 5. Contoh Clustering K-Medoids dalam Loop Paramter Euclidean Distance

Operator K-Medoids digunakan untuk melakukan proses clustering dengan variasi jumlah cluster (k) dari 2 hingga 5. Dengan pendekatan ini, seluruh kombinasi metode jarak dan nilai k dapat diuji secara bersamaan, sehingga menghasilkan perbandingan hasil clustering yang lebih efisien dan konsisten. Berikut contoh hasil clustering K-Medoids menggunakan perhitungan jarak Euclidean dengan nilai K= 2.

Row No.	Abstraksi	cluster	Pengaruh	Kapabilitas	Sosial	Profesional
1	G001	cluster_2	-1.366	1.866	-0.331	0.154
2	G002	cluster_2	0.181	-0.187	-0.231	-0.491
3	G003	cluster_0	0.181	0.217	-1.175	0.164
4	G004	cluster_0	1.424	-0.187	-1.175	-0.491
5	G005	cluster_2	0.788	-0.787	-0.331	-1.146
6	G006	cluster_0	0.788	-1.716	-1.175	1.471
7	G007	cluster_2	0.181	-1.175	-0.331	-0.491
8	G008	cluster_2	-1.126	0.217	-0.331	0.164
9	G009	cluster_2	-0.469	0.217	-0.331	0.164
10	G010	cluster_0	-0.051	-1.175	-1.175	1.118
11	G011	cluster_2	0.181	-0.187	1.108	0.164
12	G012	cluster_2	0.788	1.361	1.108	0.164
13	G013	cluster_2	0.181	-0.187	0.164	0.164
14	G014	cluster_2	-1.105	0.802	-0.331	0.164
15	G015	cluster_2	-0.262	1.361	-0.331	-1.115
16	G016	cluster_2	-1.124	-0.787	0.164	0.164

Gambar 6. Hasil Perhitungan Jarak Euclidean dengan K=2

Selanjutnya, seluruh hasil pengujian dari setiap metode jarak dan variasi nilai K digunakan untuk proses evaluasi menggunakan metode Silhouette Coefficient yang dihitung dengan bantuan Python guna menentukan kualitas cluster serta jumlah cluster yang optimal.

4. Proses Perhitungan Silhouette Coefficient menggunakan Python

Pada tahap ini, hasil clustering dievaluasi menggunakan metode Silhouette Coefficient dengan bantuan Python. Hal ini dilakukan karena RapidMiner tidak menyediakan evaluasi clustering untuk metode Silhouette Coefficient.

Berdasarkan hasil perhitungan menggunakan Python, nilai Silhouette Coefficient digunakan untuk evaluasi kualitas cluster yang terbentuk. Semakin tinggi nilai silhouette, maka semakin baik kualitas pengelompokan. Dari hasil tersebut dapat menunjukkan jumlah cluster dan metode jarak yang paling optimal. Berikut hasil perhitungan silhouette coefficient.

Tabel 3. Hasil Perhitungan Silhouette Coefficient

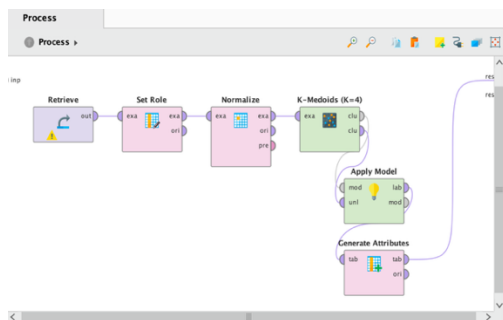
Metode	K2	K3	K4	K5
Euclidean	0,217	0,208	0,290	0,231

Metode	K2	K3	K4	K5
Manhattan	0,197	0,208	0,290	0,267
Cosine	0,082	0,248	0,171	0,202

Berdasarkan hasil yang diperoleh pada tabel 3, nilai Silhouette Coefficient tertinggi diperoleh pada nilai $K = 4$ untuk metode Euclidean dan Manhattan dengan nilai sebesar 0,290. Hal ini menunjukkan bahwa cluster tersebut merupakan yang paling optimal. Sementara itu, metode Cosine menghasilkan nilai terbaik pada $K=3$, namun masih lebih rendah dibandingkan metode lainnya.

5. Proses Clustering dengan nilai $K = 4$

Pada tahap ini, setelah diperoleh nilai K optimal, dilakukan Kembali proses clustering dengan menggunakan nilai tersebut. Hasil cluster kemudian diberi label berdasarkan nilai rata-rata masing-masing indikator. Pelabelan ini bertujuan untuk mempermudah interpretasi hasil clustering. Untuk mempermudah pemahaman, contoh perhitungan ditampilkan pada hasil terbaik dimana $K=4$ dengan metode Euclidean berdasarkan nilai evaluasi yang paling optimal.



Gambar 8. Proses Clustering dengan nilai $K=4$

Berikut hasil setelah konfigurasi proses clustering dengan $K = 4$ dan metode Euclidean ditampilkan pada RapidMiner

Row No.	Alternatif	cluster	Pedagogik	Kepribadian	Sosial	Profesional	Keterangan
1	G001	cluster_3	-1.948	1.846	-0.131	0.164	Pedagogik rendah
2	G002	cluster_1	0.581	-0.187	-0.131	-0.491	Seimbang
3	G003	cluster_1	0.581	0.317	-1.175	0.164	Seimbang
4	G004	cluster_1	1.424	-0.187	-1.175	-0.491	Seimbang
5	G005	cluster_1	0.788	-0.707	-0.131	-1.146	Seimbang
6	G006	cluster_0	0.788	-1.716	-1.175	1.473	Profesional tinggi
7	G007	cluster_1	0.581	-1.211	-0.131	-0.491	Seimbang
8	G008	cluster_3	-1.326	0.317	-0.131	0.164	Pedagogik rendah
9	G009	cluster_1	-0.469	0.317	-0.131	0.164	Seimbang
10	G010	cluster_0	-0.055	-1.211	-1.175	2.128	Profesional tinggi
11	G011	cluster_1	0.581	-0.187	1.958	0.164	Seimbang
12	G012	cluster_1	0.788	1.341	1.958	0.164	Seimbang
13	G013	cluster_1	0.581	-0.187	0.914	0.164	Seimbang
14	G014	cluster_3	-1.105	0.822	-0.131	0.164	Pedagogik rendah
15	G015	cluster_2	-0.262	1.341	-0.131	-2.455	Kepribadian dominan
16	G016	cluster_3	-1.534	-0.707	0.914	0.164	Pedagogik rendah

Gambar 9. Hasil Clustering dengan nilai $K=4$ dengan perhitungan jarak Euclidean

Nilai rentang setiap cluster diperoleh dari hasil perhitungan rata-rata atribut pada masing cluster menggunakan RapidMiner setelah proses K-Medoids clustering menggunakan operator Aggregate. Berikut karakteristik cluster pada perhitungan jarak Euclidean:

1) Cluster 0: "Profesional Tinggi"

- Nilai rata-rata pada atribut profesional lebih tinggi dibandingkan cluster lain
- Rentang nilai dominan : 0,5 – 1,8 (tinggi pada skala normalisasi)
- Atribut lain seperti kepribadian dan sosial cenderung lebih rendah

2) Cluster 1: "Seimbang"

- Nilai rata-rata seluruh atribut berada pada kisaran mendekati nol
- Rentang nilai dominan : -0,2 – 1,6

- Tidak ada atribut yang sangat dominan atau sangat rendah

3) Cluster 2: "Kepribadian Dominan"

- Nilai rata-rata atribut kepribadian paling tinggi dibandingkan atribut lain
- Rentang nilai dominan : 0,8 – 1,3 (kepribadian tinggi)
- Atribut profesional cenderung rendah (hingga negatif)

4) Cluster 3: "Pedagogik Rendah"

- Nilai rata-rata pedagogic paling rendah dibandingkan cluster lain
- Rentang nilai dominan : -1,4 – 0,5 (rendah)
- Atribut lain berada pada kategori sedang

5. Kesimpulan

Berdasarkan hasil penelitian, distance metric atau perhitungan jarak berpengaruh terhadap hasil clustering menggunakan K-Medoids. Metode Euclidean dan Mahattan menghasilkan kualitas cluster terbaik dengan nilai Silhouette Coefficient tertinggi sebesar 0,290 pada K= 4. Meskipun metode Euclidean dan Manhattan menghasilkan nilai yang sama, metode Euclidean dipilih karena lebih umum digunakan dan lebih representatif dalam mengukur jarak antar data. Oleh karena itu, kombinasi tersebut dipilih sebagai konfigurasi optimal dalam pengelompokan kinerja guru.

6. Daftar Pustaka

- [1] E. Rahmah, E. Haerani, A. Nazir, and S. Ramadhani, 'Penerapan Algoritma K-Medoids Clustering Untuk Menentukan Strategi Promosi Pada Data Mahasiswa (Studi Kasus: Stikes Perintis Padang)', Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI), vol. 5, no. 3, pp. 556–564, May 2022, doi: 10.32672/jnkti.v5i3.4355.
- [2] A. A. D. Sulistyawati and M. Sadikin, 'Penerapan Algoritma K-Medoids Untuk Menentukan Segmentasi Pelanggan', SISTEMASI, vol. 10, no. 3, p. 516, May 2021, doi: 10.32520/stmsi.v10i3.1332.
- [3] Oktavian, D. Hartama, and Selli, 'Optimization Of K-Means And K-Medoids Clustering Using Dbi Silhouette Elbow On Student Data', JURTEKSI (Jurnal Teknologi dan Sistem Informasi), vol. 11, no. 2, pp. 289–296, May 2025, doi: 10.33330/jurteks.v11i2.3531.
- [4] M. F. Tri and Y. Nataliani, 'Analisis Pengaruh Penilaian Asesor terhadap Kinerja Guru Mata Pelajaran dengan k-Means Clustering', Indonesian Journal of Computing and Modeling, vol. 4, no. 1, pp. 14–22, May 2021, doi: 10.24246/icm.v4i1.5063.
- [5] S. Bahri and D. M. Midyanti, 'Penerapan Metode K-Medoids untuk Pengelompokan Mahasiswa Berpotensi Drop Out', Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 10, no. 1, pp. 165–172, May 2023, doi: 10.25126/jtiik.20231016643.
- [6] R. Hidayati, A. Zubair, A. H. Pratama, and L. Indana, 'Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering', Techno Com, vol. 20, no. 2, pp. 186–197, May 2021, doi: 10.33633/tc.v20i2.4556.
- [7] R. Rahmawati, W. Prihartono, and Fathurrohman, 'Optimasi Stok Dengan Clustering Data Transaksi Penjualan Menggunakan Algoritma K-Means Di Konter Agung Cell', Jurnal Informatika dan

- Teknik Elektro Terapan, vol. 13, no. 2, May 2025, doi: 10.23960/jitet.v13i2.6302.
- [8] N. Mirantika, T. S. Syamfithriani, and R. Trisudarmo, 'Implementasi Algoritma K-Medoids Clustering Untuk Menentukan Segmentasi Pelanggan', vol. 17, pp. 2614–5405, doi: 10.25134/nuansa.
- [9] Solikhun, M. R. Siregar, L. Pujiastuti, M. Wahyudi, and D. Kurniawan, 'Comparison of Manhattan and Chebyshev Distance Metrics in Quantum-Based K-Medoids Clustering', SISTEMASI, vol. 14, no. 4, p. 1562, May 2025, doi: 10.32520/stmsi.v14i4.5193.
- [10] A. P. Ayuni, D. Kusnandar, and S. Martha, 'Implementasi Algoritma K-Medoids Dan Clustering Large Applications (Clara) Dengan Optimasi Silhouette Coefficient (Studi Kasus: Pengelompokan Indeks Pembangunan Manusia Berdasarkan Kabupaten/Kota di Indonesia)', 2024.
- [11] D. Ramadhaniaty, N. Puspitasari, and A. Septiarini, 'Clustering Status Keberlanjutan Karyawan Kontrak Menggunakan Algoritma K-Medoids', TELKA - Telekomunikasi Elektronika Komputasi dan Kontrol, vol. 10, no. 2, pp. 97–108, May 2024, doi: 10.15575/telka.v10n2.97-108.
- [12] A. M. Nur, Yahya, Suhartini, and R. Filikhaer, 'Penerapan Algoritma K-Means Dalam Mengelompokkan Ruang Pasien BPJS Untuk Menentukan Pola Perawatan Kesehatan Yang Efektif', Infotek Jurnal Informatika dan Teknologi, vol. 9, no. 1, pp. 174–185, May 2026, doi: 10.29408/jit.v9i1.33079.
- [13] H. Sulaiman, Y. Yuliani, K. Panggalih, M. I. Alifudin, and K. Widiyanto, 'Pengelompokan Keaktifan Anggota Perpustakaan Menggunakan Algoritma K-Means', Infotek Jurnal Informatika dan Teknologi, vol. 8, no. 1, pp. 56–65, May 2025, doi: 10.29408/jit.v8i1.27978.
- [14] M. S. Pangestu and M. A. Fitriani, 'Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K-Means', Sainteks, vol. 19, no. 2, p. 141, May 2022, doi: 10.30595/sainteks.v19i2.14495.
- [15] E. L. Amalia, A. J. Jumadi, I. A. Mashudi, and D. W. Wibowo, 'Analisis Metode Cosine Similarity Pada Aplikasi Ujian Online Otomatis (Studi Kasus JTI POLINEMA)', Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 8, no. 2, pp. 343–348, May 2021, doi: 10.25126/jtiik.2021824356.
- [16] N. Sutantri, V. Yunitasari, G. R. Sa'adi, S. Ananda, and Z. Alfian, 'Analisis Konsistensi Metode Elbow dan Silhouette Score dalam Klasterisasi pada Dataset Multisektor', Journal of Innovative and Creativity (Joecy), vol. 5, no. 2, pp. 10731–10743, May 2025, doi: 10.31004/joecy.v5i2.1690.
- [17] B. Wira, A. E. Budianto, and A. S. Wiguna, 'Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang', Rainstek Jurnal Terapan Sains dan Teknologi, vol. 1, no. 3, pp. 53–68, May 2019, doi: 10.21067/jtst.v1i3.3046.
- [18] I. Ulfailah, A. Darmiyanti, and N. Munafiah, 'Pengembangan Kompetensi Profesionalisme Guru Dalam Peningkatan Mutu Pendidikan Dan Pembelajaran', JEA (Jurnal Edukasi AUD), vol. 8, no. 2, p. 169, May 2023, doi: 10.18592/jea.v8i2.7735.
- [19] N. Widiawati, B. N. Sari, and T. N. Padilah, 'Clustering Data Penduduk Miskin Dampak Covid-19 Menggunakan Algoritma K-Medoids', Journal of Applied Informatics and Computing, vol. 6, no. 1, pp. 55–63, May 2022, doi: 10.30871/jaic.v6i1.3266