

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



Implementasi Algoritma K-Means Dalam Proses Clustering Hasil Penilaian Kinerja Karyawan di PT. XYZ

Nadia Nur Saida ^a, Rakhilca Yanedika ^b, Azzahra Athifah Dzaki ^c, Fayza Zeevania Putri ^d

Jurusan Sistem Informasi, Fakultas Teknologi Informasi, Universitas Andalas, Padang, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 00 Februari 0000

Revisi Akhir: 00 Maret 0000

Diterbitkan Online: 00 April 0000

KATA KUNCI

Penilaian Kinerja Karyawan

Clustering

K-Means

Data Mining

KORESPONDENSI

E-mail: *

ABSTRACT

Kinerja karyawan merupakan hal yang penting diperhatikan oleh perusahaan. Karyawan yang memiliki kinerja tinggi cenderung bekerja lebih efisien, menghasilkan output yang lebih banyak, dan meningkatkan efektivitas operasional bagi perusahaan. Begitu juga pada PT. XYZ yang bergerak di bidang perkebunan kelapa sawit dan pengolahan minyak kelapa sawit. PT. XYZ telah melaksanakan penilaian kinerja guna mengidentifikasi karyawan-karyawan yang memiliki kinerja yang baik. Namun dalam pelaksanaannya, PT. XYZ mengalami beberapa permasalahan dalam melakukan penilaian karyawan. Penelitian ini akan membahas penerapan algoritma K-Means clustering pada data mining terhadap pengelompokan penilaian kinerja karyawan di PT. XYZ. Penelitian ini bertujuan PT. XYZ dalam pengelompokan hasil penilaian kinerja karyawan. Penilaian kinerja karyawan ini dibagi menjadi beberapa cluster yaitu baik, kurang baik, dan sangat baik. Hasil akurasi perbandingan hasil cluster manual dengan python didapati sebesar 92%.

1. PENDAHULUAN

1.1. Latar Belakang

Untuk mencapai tujuan organisasi diperlukan karyawan yang sesuai dengan persyaratan dalam organisasi, dan juga harus mampu menjalankan tugas-tugas yang telah ditentukan oleh organisasi [1]. Karyawan yang memiliki kinerja tinggi cenderung bekerja lebih efisien, menghasilkan output yang lebih banyak, dan meningkatkan efektivitas operasional bagi perusahaan. Hal ini dapat membantu perusahaan mencapai target produksi, mengurangi biaya, dan meningkatkan keuntungan. Oleh karena itu, penting bagi perusahaan untuk memantau dan mengevaluasi kinerja karyawan mereka secara berkala.

Begitu juga pada PT. XYZ yang bergerak di bidang perkebunan kelapa sawit dan pengolahan minyak kelapa

sawit. Perusahaan ini memiliki tujuan untuk menjadi industri pengolahan minyak kelapa sawit yang sehat dan inovatif serta mengutamakan mutu dan kelestarian lingkungan. Dalam upaya mencapai tujuan tersebut, PT. XYZ telah melaksanakan penilaian kinerja guna mengidentifikasi karyawan-karyawan yang memiliki kinerja yang baik. Penilaian kinerja adalah proses pengevaluasian kinerja, penyusunan rencana pengembangan, dan pengkomunikasian hasil proses tersebut kepada karyawan itu sendiri [2]. Dari hasil penilaian tersebut akan diperoleh berbagai wawasan dan informasi mengenai kinerja karyawan, di antaranya perusahaan bisa mendapat gambaran tentang sejauh mana karyawan mencapai target dan tujuan kerja yang ditetapkan, serta bagaimana karyawan berkinerja dalam memenuhi harapan perusahaan.

Namun dalam pelaksanaannya, PT. XYZ mengalami beberapa permasalahan dalam melakukan penilaian karyawan. Permasalahannya yaitu penilaian masih

dilakukan secara subjektif berdasarkan preferensi individu manajer, sehingga penilaian yang dihasilkan kurang objektif. Ketidakobjektifan ini tentu akan berdampak pada keputusan perusahaan terkait pengelolaan kinerja karyawan. Penilaian juga masih dilakukan secara semi-manual, sehingga perusahaan kesulitan dalam mengidentifikasi pola kinerja yang berulang atau tren yang berkaitan dengan faktor-faktor tertentu karena jumlah data karyawan yang cukup banyak.

1.2. Rumusan Masalah

Perumusan masalah berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah yang didapat pada penelitian ini adalah bagaimana penerapan algoritma K-Means clustering pada data mining terhadap pengelompokan penilaian kinerja karyawan di PT.XYZ menggunakan bahasa pemrograman python?

1.3. Batasan Masalah

Batasan masalah yang diangkat dari penelitian ini adalah :

- Penerapan data mining pada data hasil penilaian kinerja karyawan di PT. XYZ.

- Menggunakan salah satu algoritma data mining, yaitu K-Means Clustering.

- Pengolahan data menggunakan algoritma K-Means Clustering secara manual pada aplikasi Microsoft Excel dan bahasa pemrograman python pada Jupyter Notebook.

- Hasil pengolahan data untuk mengelompokkan hasil kinerja karyawan kedalam beberapa cluster.

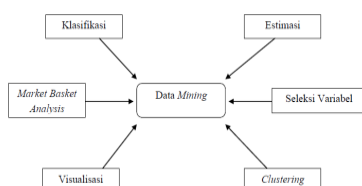
1.4. Tujuan

Tujuan dari penelitian ini adalah untuk membantu PT. XYZ dalam pengelompokan hasil penilaian kinerja karyawan guna mengevaluasi kinerja karyawan untuk perbaikan dan pengambilan keputusan terkait hasil kinerja karyawan dalam satu periode penilaian.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining adalah serangkaian proses untuk mengeksplorasi nilai tambah dari kumpulan data dalam bentuk pengetahuan yang tidak diketahui secara artifisial [6]. Karena data mining sebenarnya berasal dari bidang ilmiah seperti kecerdasan buatan, pembelajaran mesin, statistik, dan basis data[7]. Dengan kata lain, data mining adalah proses penggalian pola dari data. Data mining menjadi pusat dari beberapa kajian[8]. Diantaranya adalah estimasi, seleksi variabel, clustering, visualisasi, market basket analysis dan klasifikasi, Berikut gambaran pada gambar 2.1.



Gambar 1. Kajian Umum Data Mining

2.2. Metode Clustering

Clustering atau pengklasteran merupakan salah satu metode data mining dengan tujuan untuk mengelompokkan data dengan karakteristik yang sama di suatu ‘wilayah’ yang sama dengan karakteristik yang berbeda di ‘wilayah’ lain. Proses clustering dimana pengelompokan objek data ke dalam dua atau lebih cluster yang mempunyai tingkat kemiripan yang tinggi dan dengan data cluster berbeda tidak memiliki kemiripan dengan objek lainnya. Dan juga, dengan adanya pengelompokan data, kita dapat membangun sebuah model masalah berdasarkan pengelompokan-pengelompokan tersebut[12].

2.3. Algoritma K-Means

K-Means adalah algoritma sederhana. Dari beberapa teknik clustering yang paling sederhana dan umum dikenal adalah clustering K-Means. Algoritma *k-means* adalah algoritma pengklasteran sederhana dengan jenis clustering data non hirarki yang mempartisi dataset ke beberapa kluster *k*, algoritmanya banyak digunakan karena mudah untuk diterapkan dan menggunakan waktu yang relatif singkat. Prinsip utama dari teknik ini adalah menyusun *k* buah partisi (centroid)/rata-rata (mean) dari sekumpulan data dimulai dengan pembentukan partisi kluster kemudian secara iteratif partisi kluster ini diperbaiki hingga tidak terjadi perubahan yang signifikan pada partisi kluster[14].

3. METODE

3.1. Metode Penelitian

Metode yang digunakan pada penelitian ini adalah metode deskriptif kuantitatif. Metode deskriptif kuantitatif adalah suatu metode yang bertujuan untuk membuat gambar atau deskriptif tentang suatu keadaan secara objektif yang menggunakan angka, mulai dari pengumpulan data, penafsiran terhadap data tersebut serta penampilan dan hasilnya [13].

Metode ini tepat untuk digunakan dalam analisis clustering dengan algoritma K-Means karena metode ini menggunakan data numerik atau kuantitatif untuk menggambarkan dan menganalisis fenomena atau variabel yang sedang diteliti.

3.2. Tahap Pengumpulan Data

3.2.1. Lokasi dan Waktu Pengumpulan Data

Pengumpulan data dilakukan di PT. XYZ yang beralamat di Salareh Aia, Kecamatan Palembayan, Kabupaten Agam, Sumatera Barat. Adapun pengumpulan data ini dilakukan pada tanggal 4 Mei 2023 - 8 Mei 2023 dengan judul laporan “Implementasi Algoritma K-Means dalam Proses Clustering Hasil Penilaian Kinerja Karyawan di PT. XYZ” .

3.2.2. Kebutuhan Laporan

Dalam laporan ini kebutuhan yang diperlukan untuk menjadi bahan analisis serta pengolahan data adalah :

1. Data hasil penilaian kinerja karyawan di PT. XYZ.
2. Perusahaan PT. XYZ sebagai penyedia data.
3. Microsoft Excel sebagai alat pengolahan data manual.
4. Jupyter Notebook sebagai alat pengolahan data mining.

3.2.3. Metode Pengumpulan Data

1. Wawancara

Pada penelitian ini, dilakukan wawancara dengan seorang narasumber dari Departemen Human Resource di PT. XYZ.

2. Dokumentasi

Dokumentasi yang didapatkan berupa data rekapan hasil penilaian kinerja karyawan dalam satu periode penilaian di PT. XYZ.

3. Studi Literatur

Dokumentasi yang didapatkan berupa data rekapan hasil penilaian kinerja karyawan dalam satu periode penilaian di PT. XYZ.

3.3. Tahapan Pembuatan Laporan

1. Persiapan

Di tahap ini dilakukan penyusunan rencana pembuatan laporan dan perancangan timeline pembuatan laporan, termasuk menentukan objek, metode serta algoritma yang digunakan.

2. Studi Literatur

Di tahap ini dilakukan kajian dan studi literatur mengenai permasalahan yang diangkat, metode, serta algoritma yang digunakan.

3. Pengumpulan Data

Di tahap ini dilakukan melalui wawancara, dokumentasi, dan studi literatur.

4. Preprocessing Data

Di tahap ini dilakukan pengolahan awal termasuk data reduction untuk mengurangi data yang tidak dibutuhkan dan data cleaning untuk mengisi missing value atau yang kurang pas, mengoreksi data yang tidak konsisten, dan mendeteksi redundansi data.

5. Perhitungan Data Manual

Tahap ini dilakukan menggunakan Microsoft Excel agar di akhir dapat dibandingkan keakuratannya dengan hasil pengolahan data menggunakan Python di Jupyter Notebook.

6. Pengujian Cluster

Di tahap ini dilakukan pengujian cluster menggunakan metode Elbow Method.

7. Visualisasi Hasil Cluster

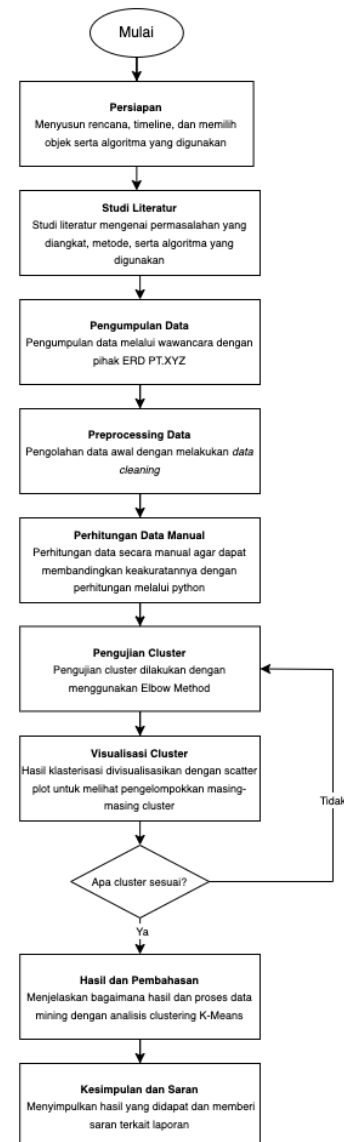
Di tahap ini hasil dari tahap clustering akan divisualisasikan dalam bentuk scatter plot untuk melihat penyebaran dan pengelompokan dari masing - masing cluster.

8. Hasil dan Pembahasan

Pada tahapan ini akan dijelaskan mengenai hasil dari proses data mining yang dilakukan menggunakan analisis clustering dengan Algoritma K-Means.

9. Kesimpulan dan Saran

Di tahap ini dibuat kesimpulan dari hasil yang didapatkan dan memberikan saran untuk pihak PT. XYZ terkait dengan permasalahan yang diangkat serta pembuatan laporan selanjutnya.



Gambar 2 Tahap Pembuatan Laporan

3.4. Tahap Pre-Processing Data

1. Data Cleaning

Tahap ini melibatkan identifikasi dan penanganan data yang tidak lengkap, data yang hilang, atau data yang tidak valid dari data mentah yang akan digunakan.

2. Data Integration

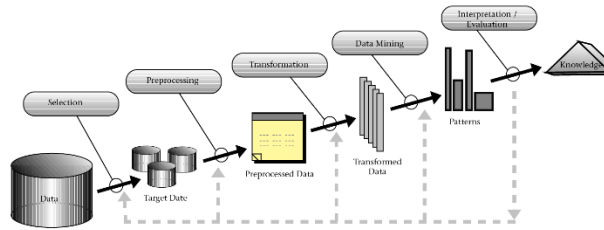
Pada tahap ini menangani perbedaan skema data, format data, dan resolusi referensi yang berbeda.

3. Transformasi

Data Pada tahap ini, dilakukan transformasi pada data untuk memperbaiki distribusi data, mengurangi noise, atau mengubah skala data.

4. Data Reduction

Tahap ini dilakukan untuk mengurangi jumlah fitur dan menghilangkan fitur yang tidak memberikan kontribusi signifikan pada analisis dengan cara mengurangi sampel data yang diambil.



Gambar 3 Tahap Pre-processing Data

3.5. Tahap Klasifikasi Data

Penelitian ini menggunakan algoritma K-Means Clustering sebagai algoritma klasifikasinya. Hal ini sesuai dengan permasalahan yang diangkat, yakni identifikasi klasifikasi hasil penilaian kinerja karyawan di PT. XYZ, di antaranya kinerja sangat baik, baik, dan kurang baik.

4. HASIL DAN PEMBAHASAN

Pada penelitian ini menggunakan data penilaian kinerja karyawan yang berasal dari PT.XYZ selama satu periode penilaian. Data yang diperoleh akan ditransformasi agar dapat diproses lebih lanjut untuk clusterisasi. Proses cluster dilakukan untuk mengetahui pola kinerja karyawan dalam satu periode penilaian sehingga dapat dilakukan perbaikan dan pengambilan keputusan terhadap hasil kinerja karyawan.

4.1. Persiapan Data

Sumber data utama yang digunakan dalam penelitian ini adalah dataset penilaian kinerja karyawan di PT. XYZ dengan jumlah keseluruhan data 121 record dari keseluruhan total karyawan yang ada pada PT.XYZ. Jumlah atribut data yang diterima ialah sebanyak 18 atribut. Dengan atribut yang dibutuhkan dalam pengclusteran ini adalah NIK, Index Performance, dan Index Potensial seperti pada tabel 1.

Tabel 1. Rekap Appraisal Penilaian Kinerja Karyawan PT.XYZ

No	NIK	Index Potensial	Index Performance
1	009AMP18068708	33	39
2	327AMP10037607	24	22
3	002AMP13077908	27	39
4	GA/GMP/0604/403	24	28
5	307AMP09087102	27	33
6	002AMP14069207	27	33
7	324AMP11028703	24	30
8	0493AMP03057303	21	25
9	313AMP01028104	18	25
10	180AMP06067908	18	14
11	329AMP08057901	24	25
12	0431AMP97036608	39	47

13	322AMP11027610	18	22
...
121	0410AMP01067604	18	42

4.2. Tahap Pre-processing Data

Data penilaian kinerja karyawan PT.XYZ yang diterima memiliki 18 atribut data, yaitu NIK, Nama, Tgl Lahir, Umur, Tgl Masuk, Masa Kerja, Jabatan, Pendidikan, Hasil Kerja (Kuantitas Kerja, Ketepatan Waktu, Kualitas Kerja), Perilaku Kerja (Pengetahuan, Komunikasi, Kerjasama Tim, Disiplin Kerja), Index (*Performance* dan *Potensial*) dan Keterangan. Dengan keterangan, Index *Performance* merupakan rerata akhir penilaian Hasil Kerja dan Index *Potensial* merupakan rerata akhir penilaian Perilaku kerja. Variable feature yang akan digunakan dalam proses klusterisasi ini adalah Index *Potensial* dan Index *Performance* yang merupakan rata-rata hasil akhir penilaian terhadap hasil kerja dan perilaku kerja. Kedua variabel ini tidak mengandung nilai yang hilang (*missing value*) ataupun *outliers*, sehingga dapat langsung dilakukan pengelompokan data.

4.3. Perhitungan Manual Algoritma K-Means Clustering

Berikut ini langkah-langkah perhitungan manual algoritma K-Means Clustering yang menggunakan 61 records sample data dengan rumus perhitungan jarak Euclidean Distance seperti berikut :

$$\text{Distance} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + (x_{ni} - x_{nj})^2} \quad \dots(1)$$

1. Menentukan Cluster Awal

Menentukan K jumlah cluster secara acak (random). Percobaan pertama ditetapkan 3 titik acak sebagai pusat awal perhitungan jarak dari seluruh kelompok cluster.

Jumlah cluster = 3

Jumlah data = 61 record

Jumlah atribut = 2

Tabel 2. Nilai Pusat Cluster awal

No	Titik Pusat	Index Performance	Index Potensial
1	Centroid 0	25.5	32
2	Centroid 1	28.5	41.5
3	Centroid 2	20	39

Hitung jarak tiap data ke masing-masing pusat cluster menggunakan persamaan (1) yaitu persamaan Euclidean Distance.

Jarak setiap data ke tiap titik pusat pertama :

$$S1|66\ 68 = ((33 - 25.5)^2 + (39 - 32)^2) = 105.25$$

$$S2|66\ 68 = ((24 - 25.5)^2 + (22 - 32)^2) = 102.25$$

$$S3|66\ 68 = ((27 - 25.5)^2 + (39 - 32)^2) = 51.25$$

$$S4|66\ 68 = ((24 - 25.5)^2 + (28 - 32)^2) = 18.25$$

$$S5|66\ 68 = ((27 - 25.5)^2 + (33 - 32)^2) = 13.25$$

Dst.

Jarak setiap data ke tiap titik pusat kedua :

$$\begin{aligned}
S1|13 &= ((33 - 30)^2 + (39 - 39)^2) = 9 \\
S2|13 &= ((24 - 30)^2 + (22 - 39)^2) = 325 \\
S3|13 &= ((27 - 30)^2 + (39 - 39)^2) = 9 \\
S4|13 &= ((24 - 30)^2 + (28 - 39)^2) = 157 \\
S5|13 &= ((27 - 30)^2 + (33 - 39)^2) = 45 \\
\text{Dst.}
\end{aligned}$$

Jarak setiap data ke tiap titik pusat ketiga :

$$\begin{aligned}
S1|2733 &= ((33 - 28.5)^2 + (39 - 41.5)^2) = 26.5 \\
S2|2733 &= ((24 - 28.5)^2 + (22 - 41.5)^2) = 400.5 \\
S3|2733 &= ((27 - 28.5)^2 + (39 - 41.5)^2) = 8.5 \\
S4|2733 &= ((24 - 28.5)^2 + (28 - 41.5)^2) = 202.5 \\
S5|2733 &= ((27 - 28.5)^2 + (33 - 41.5)^2) = 74.5 \\
\text{Dst.}
\end{aligned}$$

Perhitungan diatas menghasilkan tabel distance seperti tabel 3.

Tabel 3. Perhitungan Euclidean Distance Awal

No	NIK	C1	C2	C3
1	009AMP18068708	105.25	26.5	9
2	327AMP10037607	102.25	400.5	325
3	002AMP13077908	51.25	8.5	9
4	GA/GMP/0604/403	18.25	202.5	157
5	307AMP09087102	3.25	74.5	45
6	002AMP14069207	3.25	74.5	45
7	324AMP11028703	6.25	152.5	117
8	0493AMP03057303	69.25	328.5	277
9	313AMP01028104	105.25	382.5	340
10	180AMP06067908	380.25	866.5	769
11	329AMP08057901	51.25	292.5	232
12	0431AMP97036608	407.25	140.5	145
13	322AMP11027610	156.25	490.5	433
...
61	325AMP05078303	51.25	8.5	9

Anggota terkecil dipilih diantara 3 cluster. Jika terkecil pada bagian C1 maka termasuk sebagai anggota C1 yaitu sebanyak 41 data, jika terkecil pada bagian C2 maka termasuk sebagai anggota C2 yaitu sebanyak 9 data, dan jika terkecil pada bagian C3 maka termasuk sebagai anggota C3 yaitu sebanyak 11 data.

2. Menentukan Cluster Iterasi ke-1

Menentukan titik centroid baru dengan pengambilan sampel secara acak. dengan mempertimbangkan rentang nilai index performance dan index potensial

Tabel 4. Nilai Pusat Cluster Iterasi ke-1

No	Titik Pusat	Index Performance	Index Potensial
1	Centroid 0	27	33
2	Centroid 1	25.5	29
3	Centroid 2	34.5	43

Hitung jarak tiap data ke masing-masing pusat cluster baru menggunakan persamaan Euclidean Distance.

Jarak setiap data ke tiap titik pusat pertama :

$$\begin{aligned}
S1|1722 &= ((33 - 34.5)^2 + (39 - 45)^2) = 18.25 \\
S2|1722 &= ((24 - 34.5)^2 + (22 - 45)^2) = 551.25
\end{aligned}$$

$$\begin{aligned}
S3|1722 &= ((27 - 34.5)^2 + (39 - 45)^2) = 72.25 \\
S4|1722 &= ((24 - 34.5)^2 + (28 - 45)^2) = 335.25 \\
S5|1722 &= ((27 - 34.5)^2 + (33 - 45)^2) = 156.25 \\
\text{Dst.}
\end{aligned}$$

Jarak setiap data ke tiap titik pusat kedua :

$$\begin{aligned}
S1|28 &= ((33 - 25.5)^2 + (39 - 29)^2) = 156.25 \\
S2|28 &= ((24 - 25.5)^2 + (22 - 29)^2) = 51.25 \\
S3|28 &= ((27 - 25.5)^2 + (39 - 29)^2) = 102.25 \\
S4|28 &= ((24 - 25.5)^2 + (28 - 29)^2) = 3.25 \\
S5|28 &= ((27 - 25.5)^2 + (33 - 29)^2) = 18.25 \\
\text{Dst.}
\end{aligned}$$

Jarak setiap data ke tiap titik pusat ketiga :

$$\begin{aligned}
S1|116 &= ((33 - 27)^2 + (39 - 33)^2) = 72 \\
S2|116 &= ((24 - 27)^2 + (22 - 33)^2) = 130 \\
S3|116 &= ((27 - 27)^2 + (39 - 33)^2) = 36 \\
S4|116 &= ((24 - 27)^2 + (28 - 33)^2) = 34 \\
S5|116 &= ((27 - 27)^2 + (33 - 33)^2) = 0 \\
\text{Dst.}
\end{aligned}$$

Perhitungan diatas menghasilkan tabel distance seperti tabel 5.

Tabel 5. Perhitungan Euclidean Distance Awal

No	NIK	C1	C2	C3
1	009AMP18068708	72	156.5	18.25
2	327AMP10037607	130	51.5	551.25
3	002AMP13077908	36	102.5	72.25
4	GA/GMP/0604/403	34	3.5	335.25
5	307AMP09087102	0	18.5	156.25
6	002AMP14069207	0	18.5	156.25
7	324AMP11028703	18	3.5	279.25
8	0493AMP03057303	100	36.5	506.25
9	313AMP01028104	145	72.5	596.25
10	180AMP06067908	442	281.5	1113.25
11	329AMP08057901	73	18.5	434.25
12	0431AMP97036608	340	506.5	36.25
13	322AMP11027610	202	105.5	713.25
...
61	325AMP05078303	36	102.5	72.25

Terdapat perubahan cluster data pada perhitungan iterasi ke-2. Anggota terkecil dipilih diantara 3 cluster. Jika terkecil pada bagian C1 maka termasuk sebagai anggota C1 yaitu sebanyak 20 data, jika terkecil pada bagian C2 maka termasuk sebagai anggota C2 yaitu sebanyak 27 data, dan jika terkecil pada bagian C3 maka termasuk sebagai anggota C3 yaitu sebanyak 14 data.

3. Menentukan Cluster Iterasi ke-2

Karena hasil iterasi ke-1 dan cluster awal tidak sama, sehingga perlu dilakukan kembali perhitungan iterasi ke-2 dan seterusnya hingga mendapatkan hasil yang sama.

Menentukan titik centroid baru dengan pengambilan sampel berdasarkan rata-rata rentang nilai index performance dan index potensial yang terbagi tiga, yaitu rendah, sedang, dan tinggi

Tabel 6. Nilai Pusat Cluster Iterasi ke-2

No	Titik Pusat	Index Performance	Index Potensial
1	Centroid 0	25.5	33
2	Centroid 1	21	25
3	Centroid 2	34.5	41.5

Hitung jarak tiap data ke masing-masing pusat cluster baru menggunakan persamaan Euclidean Distance.

Jarak setiap data ke tiap titik pusat pertama :

$$\begin{aligned} S1|1\ 96 &= ((33 - 25.5)^2 + (39 - 33)^2) = 92.25 \\ S2|1\ 96 &= ((24 - 25.5)^2 + (22 - 33)^2) = 123.25 \\ S3|1\ 96 &= ((27 - 25.5)^2 + (39 - 33)^2) = 38.25 \\ S4|1\ 96 &= ((24 - 25.5)^2 + (28 - 33)^2) = 27.25 \\ S5|1\ 96 &= ((27 - 25.5)^2 + (33 - 33)^2) = 2.25 \end{aligned}$$

Dst.

Jarak setiap data ke tiap titik pusat kedua :

$$\begin{aligned} S1|4\ 13 &= ((33 - 21)^2 + (39 - 25)^2) = 340 \\ S2|4\ 13 &= ((24 - 21)^2 + (22 - 25)^2) = 18 \\ S3|4\ 13 &= ((27 - 21)^2 + (39 - 25)^2) = 232 \\ S4|4\ 13 &= ((24 - 21)^2 + (28 - 25)^2) = 18 \\ S5|4\ 13 &= ((27 - 21)^2 + (33 - 25)^2) = 100 \end{aligned}$$

Dst.

Jarak setiap data ke tiap titik pusat ketiga :

$$\begin{aligned} S1|1\ 33 &= ((33 - 34.5)^2 + (39 - 41.5)^2) = 74.5 \\ S2|1\ 33 &= ((24 - 34.5)^2 + (22 - 41.5)^2) = 416.5 \\ S3|1\ 33 &= ((27 - 34.5)^2 + (39 - 41.5)^2) = 266.5 \\ S4|1\ 33 &= ((24 - 34.5)^2 + (28 - 41.5)^2) = 416.5 \\ S5|1\ 33 &= ((27 - 34.5)^2 + (33 - 41.5)^2) = 266.5 \end{aligned}$$

Dst.

Perhitungan diatas menghasilkan tabel distance seperti tabel 7.

Tabel 7. Perhitungan Euclidean Distance Awal

No	NIK	C1	C2	C3
1	009AMP18068708	92.25	340	74.5
2	327AMP10037607	123.25	18	416.5
3	002AMP13077908	38.25	232	266.5
4	GA/GMP/0604/403	27.25	18	416.5
5	307AMP09087102	2.25	100	266.5
6	002AMP14069207	2.25	100	416.5
7	324AMP11028703	11.25	34	602.5
8	0493AMP03057303	84.25	0	824.5
9	313AMP01028104	120.25	9	824.5
10	180AMP06067908	417.25	130	416.5
11	329AMP08057901	66.25	9	602.5
12	0431AMP97036608	378.25	808	824.5
13	322AMP11027610	177.25	18	824.5
...
61	325AMP05078303	38.25	232	266.5

Terdapat perubahan cluster data pada perhitungan iterasi ke-2. Anggota terkecil dipilih diantara 3 cluster. Jika terkecil pada bagian C1 maka termasuk sebagai anggota C1 yaitu sebanyak 29

data, jika terkecil pada bagian C2 maka termasuk sebagai anggota C2 yaitu sebanyak 22 data, dan jika terkecil pada bagian C3 maka termasuk sebagai anggota C3 yaitu sebanyak 10 data.

4. Menentukan Cluster Iterasi ke-3

Karena hasil iterasi ke-2 dan iterasi ke-1 tidak sama, sehingga perlu dilakukan kembali perhitungan iterasi ke-3 dan seterusnya hingga mendapatkan hasil yang sama.

Menentukan titik centroid baru dengan pengambilan sampel berdasarkan rata-rata rentang nilai index performance dan index potensial yang terbagi tiga, yaitu rendah, sedang, dan tinggi

Tabel 8. Nilai Pusat Cluster Iterasi ke-2

No	Titik Pusat	Index Performance	Index Potensial
1	Centroid 0	25.5	33
2	Centroid 1	21	26
3	Centroid 2	33	41.5

Hitung jarak tiap data ke masing-masing pusat cluster baru menggunakan persamaan Euclidean Distance.

Jarak setiap data ke tiap titik pusat pertama :

$$\begin{aligned} S1|29\ 96 &= ((33 - 25.5)^2 + (39 - 33)^2) = 92.25 \\ S2|29\ 96 &= ((24 - 25.5)^2 + (22 - 33)^2) = 123.25 \\ S3|29\ 96 &= ((27 - 25.5)^2 + (39 - 33)^2) = 38.25 \\ S4|29\ 96 &= ((24 - 25.5)^2 + (28 - 33)^2) = 27.25 \\ S5|29\ 96 &= ((27 - 25.5)^2 + (33 - 33)^2) = 2.25 \end{aligned}$$

Dst.

Jarak setiap data ke tiap titik pusat kedua :

$$\begin{aligned} S1|29\ 96 &= ((33 - 25.5)^2 + (39 - 33)^2) = 92.25 \\ S2|29\ 96 &= ((24 - 25.5)^2 + (22 - 33)^2) = 123.25 \\ S3|29\ 96 &= ((27 - 25.5)^2 + (39 - 33)^2) = 38.25 \\ S4|29\ 96 &= ((24 - 25.5)^2 + (28 - 33)^2) = 27.25 \\ S5|29\ 96 &= ((27 - 25.5)^2 + (33 - 33)^2) = 2.25 \end{aligned}$$

Dst.

Jarak setiap data ke tiap titik pusat ketiga :

Dst.

$$\begin{aligned} S1|29\ 96 &= ((33 - 25.5)^2 + (39 - 33)^2) = 92.25 \\ S2|29\ 96 &= ((24 - 25.5)^2 + (22 - 33)^2) = 123.25 \\ S3|29\ 96 &= ((27 - 25.5)^2 + (39 - 33)^2) = 38.25 \\ S4|29\ 96 &= ((24 - 25.5)^2 + (28 - 33)^2) = 27.25 \\ S5|29\ 96 &= ((27 - 25.5)^2 + (33 - 33)^2) = 2.25 \end{aligned}$$

Dst.

Perhitungan diatas menghasilkan tabel distance seperti tabel 9.

Tabel 9. Perhitungan Euclidean Distance Awal

No	NIK	C1	C2	C3
1	009AMP18068708	92.25	313	72.25
2	327AMP10037607	123.25	25	387.25
3	002AMP13077908	38.25	205	246.25
4	GA/GMP/0604/403	27.25	13	387.25
5	307AMP09087102	2.25	85	246.25
6	002AMP14069207	2.25	85	246.25
7	324AMP11028703	11.25	25	387.25
8	0493AMP03057303	84.25	1	564.25

9	313AMP01028104	120.25	10	777.25
10	180AMP06067908	417.25	153	777.25
11	329AMP08057901	66.25	10	387.25
12	0431AMP97036608	378.25	765	42.25
13	322AMP11027610	177.25	25	777.25
...
61	325AMP05078303	38.25	205	246.25

Terdapat perubahan cluster data pada perhitungan iterasi ke-2. Anggota terkecil dipilih diantara 3 cluster. Jika terkecil pada bagian C1 maka termasuk sebagai anggota C1 yaitu sebanyak 29 data, jika terkecil pada bagian C2 maka termasuk sebagai anggota C2 yaitu sebanyak 22 data, dan jika terkecil pada bagian C3 maka termasuk sebagai anggota C3 yaitu sebanyak 10 data.

5. Hasil Akhir Pengelompokan

Karena hasil iterasi ke-3 dan iterasi ke-2 clusternya sudah sama, maka tidak perlu lagi melanjutkan iterasi-4. Perhitungan perulangan penentuan centroid berhenti pada iterasi ke-3.

Hasil clusterisasi penilaian kinerja karyawan dengan algoritma k-means menggunakan perhitungan manual dengan 61 records sample data yang digunakan, didapati hasil clusternya yaitu :

- Cluster 0, sebagai cluster pertama dengan kelompok hasil penilaian kinerja karyawan 'baik' sebanyak 29 data.
- Cluster 1, sebagai cluster kedua dengan kelompok hasil penilaian kinerja karyawan 'kurang baik' sebanyak 22 data.
- Cluster 2, sebagai cluster ketiga dengan kelompok hasil penilaian kinerja karyawan 'sangat baik' sebanyak 10 data.

4.4. Penerapan Algoritma K-Means Clustering Pada Python

Python merupakan bahasa pemrograman interpretatif yang multiguna. Bahasa pemrograman ini banyak digunakan dalam pengolahan data.

Setelah dilakukannya perhitungan secara manual dilanjutkan pada tahap pengujian algoritma ini dengan bahasa pemrograman python pada jupyter notebook. Pengujian dilakukan untuk meningkatkan mutu pembelajaran menggunakan algoritma k-means clustering.

Algoritma k-means membutuhkan parameter untuk jumlah kelas (cluster) yang dibutuhkan. Penentuan jumlah kelas (cluster) dapat menggunakan metode elbow. Metode Elbow adalah metode empiris untuk menentukan jumlah cluster yang optimal dalam dataset. Dengan menggunakan metode ini, dapat menentukan jumlah kandidat k dan kemudian menerapkan pada pengelompokan k-means menggunakan nilai k yang ditetapkan.

No	Keterangan	Source Code
1	Library Python	import pandas as pd import numpy as np import matplotlib.pyplot as plt from sklearn.cluster import KMeans from sklearn.preprocessing import MinMaxScaler
2	Grafik Elbow	wcss_list = [] for i in range (1,10):

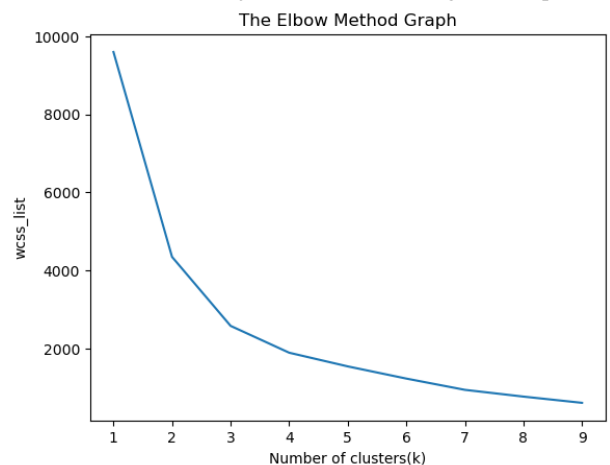
7

<https://doi.org/10.25077/TEKNOSI>

		kmeans = KMeans(n_clusters = i, init='k-means++', random_state = 42) kmeans.fit(dataClus) wcss_list.append(kmeans.inertia_)
3	Variable Feature	X = df[['INDEX Performance', 'INDEX Potensial']]
4	Model Algoritma K-Means	kmeans = KMeans(n_clusters=3, init='k-means++', random_state = 42) predict = kmeans.fit_predict(X)
5	Label Cluster	labels = kmeans.labels_
6	Pusat Cluster	kmeans.cluster_centers_
7	Jumlah data tiap Cluster	dataOlah['Clusters'].value_counts()
8	Visualisasi barchart Jumlah data tiap Cluster	ax = dataOlah['Clusters'].value_counts().plot(kind='bar', figsize=(14,8), title="Total Karyawan tiap Clusters", color='palevioletred')
9	Visualisasi cluster dengan scatter plot	plt.scatter(X.iloc[:, 0], X.iloc[:, 1], c=kmeans.labels_, cmap='rainbow') plt.scatter(kmeans.cluster_centers_[0], kmeans.cluster_centers_[1], s=300, c='orange')
10	Perbandingan dengan hasil manual	ars = adjusted_rand_score(ground_truth_labels, labels)

4.5. Evaluasi Hasil

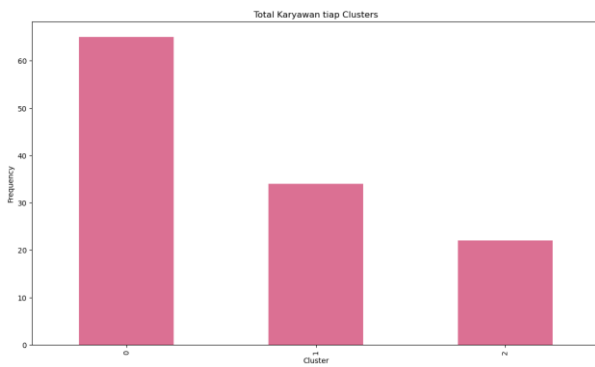
Grafik Elbow — Penentuan jumlah k yang digunakan dengan mengidentifikasi titik siku atau "elbow" pada grafik, yaitu titik di mana penurunan inersia menjadi lebih lambat. Jika penurunan inersia sebelum titik siku signifikan, tetapi penurunan inersia setelah titik siku hanya kecil, maka titik siku tersebut dapat dianggap sebagai jumlah optimal k. Terlihat pada grafik metode elbow Gambar 2. menunjukkan nilai k = 3 sebagai nilai optimal.



Gambar 4 Grafik Metode Elbow

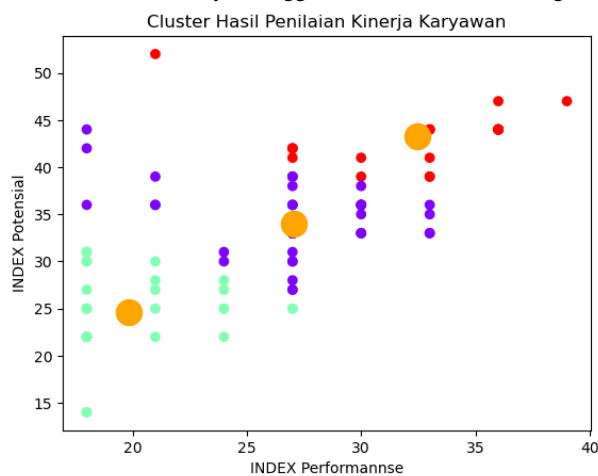
Pusat Cluster— Titik pusat centroid cluster pertama dengan nilai 27.046 dan 33.984, kemudian centroid cluster kedua dengan nilai 19.852 dan 24.676, dan centroid cluster ketiga dengan nilai 32.454 dan 43.272.

Visualisasi Barchart — Diagram perbandingan jumlah data pada masing-masing jumlah cluster.



Gambar 5 Visualisasi Diagram Perbandingan Jumlah data Cluster

Visualisasi Scatterplot — Visualisasi titik data setiap cluster menggunakan scatter plot untuk melihat persebaran data di setiap titik centroid. Grafik ini menunjukkan hasil prediksi yang telah dimodelkan sebelumnya menggunakan k-means clustering.



Gambar 6 Persebaran Data

Akurasi Perbandingan Hasil — Berdasarkan hasil penelitian pada data hasil penilaian kinerja karyawan PT.XYZ menggunakan dataset penilaian dalam satu periode penilaian, dilakukan perhitungan secara manual dan python untuk menentukan kelompok masing-masing data. Hasil akurasi perbandingan hasil cluster manual dengan python didapati sebesar 0.9280109924183162.

Cluster hasil penilaian kinerja karyawan PT.XYZ yang terdiri dari 121 records :

- Cluster 0, sebagai cluster pertama dengan kelompok hasil penilaian kinerja karyawan ‘baik’ sebanyak 65 data.
- Cluster 1, sebagai cluster kedua dengan kelompok hasil penilaian kinerja karyawan ‘kurang baik’ sebanyak 34 data.
- Cluster 2, sebagai cluster ketiga dengan kelompok hasil penilaian kinerja karyawan ‘sangat baik’ sebanyak 22 data.

5. KESIMPULAN

Berdasarkan hasil penelitian pada PT. XYZ, maka dapat diambil beberapa kesimpulan sebagai berikut :

- Metode Clustering K-Means dapat membantu mengelompokkan kinerja karyawan berdasarkan faktor-faktor yang mempengaruhi pada PT. XYZ.
- Pengelompokkan kinerja karyawan pada penelitian ini dibagi atas: karyawan kinerja sangat baik, baik, dan kurang baik.
- Pengelompokkan kinerja karyawan baik, kurang baik, sangat baik diperoleh setelah melalui proses perhitungan dengan menerapkan algoritma K-Means Clustering selesai.
- Hasil dari pengujian yang telah dilakukan dengan 61 data sampel, maka terbentuk data kelompok karyawan dengan kinerja baik terdiri dari 29 (dua puluh sembilan) karyawan, data kelompok karyawan dengan kinerja kurang baik terdiri dari 22 (dua puluh dua) karyawan, dan data kelompok sangat baik 10 (sepuluh) karyawan.
- Hasil dari pengelompokkan kinerja ini akan membantu PT. XYZ dalam pengambilan keputusan untuk pengembangan karir karyawan sehingga target dan tujuan perusahaan dapat tercapai.

DAFTAR PUSTAKA

- [1] T. Penulis et al., KINERJA KARYAWAN. 2021. [Online]. Available: www.penerbitwidina.com
- [2] A. Ismail et al., “INDONESIA KLASIFIKASI JAJANAN TRADISIONAL JAWA TENGAH DENGAN METODE TRANSFER LEARNING DAN MOBILNETV2 IMPLEMENTASI METODE K-MEANS CLUSTERING PADA PENILAIAN 66 KINERJA KARYAWAN PT KOPETRI CITRA ABADI INTEGRASI SISTEM PRESENSI BIOMETRIK DENGAN SISTEM INFORMASI SUMBER DAYA MANUSIA UNIVERSITAS MUHAMMADIYAH SEMARANG ANALISIS PENGARUH PENERAPAN CRM PADA APLIKASI POSAJA TERHADAP LOYALITAS PENGGUNA.” [Online]. Available: <http://e-journal.janabadra.ac.id/>
- [3] S. Regina, E. Sutinah, and N. Agustina, “Clustering Kualitas Kinerja Karyawan Pada Perusahaan Bahan Kimia Menggunakan Algoritma K-Means,” JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 5, no. 2, p. 573, Apr. 2021, doi: 10.30865/mib.v5i2.2909.
- [4] A. Dan, N. Yona, S. Munti, G. W. Nurcahyo, and J. Santony, “Analisis Dan Penerapan Data Mining Untuk Menentukan Gaji Karyawan Tetap Dan Karyawan Kontrak Menggunakan Algoritma K-Means Clustering (Studi Kasus Di Pt Indomex Dwijaya Lestari).”
- [5] A. Nur Khomarudin, “Teknik Data Mining : Algoritma K-Means Clustering,” 2003. [Online]. Available: <https://agusnkhom.wordpress.com>
- [6] V. S. Moertini, “DATA MINING SEBAGAI SOLUSI BISNIS,” 2002.

[7] Inc. Jhon Willey & Sons, *Discovering Knowledge In Data: An Introduction To Data Mining*. Larose, D. T. , 2005.

[8] B. S. Budi Santosa, *Data mining : Teknik pemanfaatan data untuk keperluan bisnis / Budi Santosa*. Yogyakarta: Garah Ilmu, 2007.

[9] "Machine Learning, Data Mining, and Knowledge Discovery: An Introduction." [Online]. Available: http://www.kdnuggets.com/data_mining_course/

[10] Retno Tri Wulandari, *Data Mining Teori dan Aplikasi Rapidminer*. 2017.

[11] J. P. H. J. L. I. kosasih Kotter, *Dampak budaya perusahaan terhadap kinerja / John P. Kotter, James I. Heskett; alih bahasa, Benyamin Molan; penyunting bahasa, Kosasih Iskandarsyah*. Jakarta : Prenhallindo, 1997.

[12] M. Simanjuntak, E. Manik, T. Supratman, and S. Kaputama, "PENERAPAN DATA MINING PENGELOMPOKAN KEJAHATAN ELEKTRONIK SESUAI UU ITE DENGAN MENGGUNAKAN METODE CLUSTERING," *Jurnal Mahajana Informasi*, vol. 3, no. 2, 2018.

[13] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, pp. 17–24, Apr. 2019, doi: 10.25077/teknosi.v5i1.2019.17-24

[14] N. Y. S. Munti, G. W. Nurcahyo dan J. Santony, "Analisis Dan Penerapan Data Mining Untuk Menentukan Gaji Karyawan Tetap Dan Karyawan Kontrak Menggunakan Algoritma K-Means Clustering (Studi Kasus Di Pt Indomex Dwijaya Lestari)," *Jurnal Inovasi Teknik Informatika*, vol. 1 no. 1, 2018



Azzahra Athifah Dzaki
Mahasiswa Jurusan Sistem Informasi
Universitas Andalas angkatan tahun 2021.
Berasal dari Kota Payakumbuh, Sumatera
Barat, Indonesia.



Fayza Zeevania Putri
Mahasiswa Jurusan Sistem Informasi angkatan
2021 Universitas Andalas, Padang, Sumatera
Barat, Indonesia.

BIODATA PENULIS



Nadia Nur Saida
Mahasiswa Jurusan Sistem Informasi
Universitas Andalas angkatan tahun 2021.
Berasal dari Kota Padang, Sumatera Barat,
Indonesia.



Rakhilca Yanedika
Mahasiswa Jurusan Sistem Informasi
Universitas Andalas angkatan tahun 2021.
Berasal dari Kota Padang, Sumatera Barat,
Indonesia.