



Metode *k-means clustering* untuk mengukur tingkat kedisiplinan pegawai (studi kasus di pemerintah kabupaten padang pariaman)

Rezki Kristian^{*1}, Sarjon Defit², Sumijan³

Email: ¹rezki.kr@gmail.com, ²sarjond@yahoo.co.uk, ³sumijan@sumijan.ac.id

^{1,2,3}Teknik Informatika, Ilmu Komputer, Universitas Putra Indonesia YPTK Padang

Diterima: 17 Maret 2023 | Direvisi: 17 April 2023 | Disetujui: 28 April 2023

©2023 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Knowledge Discovery In Database (KDD) merupakan sebuah proses merubah data mentah menjadi data yang bermanfaat berbentuk informasi. Data mining adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (database) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. *Clustering* adalah salah satu metode dalam data mining yang dimana objek data yang mempunyai kemiripan atau karakteristik yang sama akan dikelompokkan menjadi satu kelompok dan yang berbeda di kelompokkan pada kelompok yang lainnya. Salah satu aspek kedisiplinan yang dapat digunakan untuk mengevaluasi kinerja pegawai adalah dengan kehadiran. Metode *k-means* digunakan untuk mengelompokkan tingkat kedisiplinan pegawai kemudian mendeskripsikan nilai-nilai yang sudah didapatkan untuk menghasilkan sebuah knowledge baru mengenai pola data tingkat kedisiplinan pegawai. Data absensi di clustering menjadi 3 yaitu untuk mengukur tingkat kedisiplinan rendah, sedang, dan tinggi. proses perhitungan dari 41 sampel pegawai menghasilkan 3 kali iterasi, dan di dapatkan hasil akhir 3 cluster yaitu cluster 1 sebanyak 10 pegawai dengan kedisiplinan rendah, cluster 2 sebanyak 7 pegawai dengan kedisiplinan sedang, dan cluster 3 sebanyak 24 pegawai dengan kedisiplinan tinggi. Penelitian ini bertujuan agar pimpinan dapat mengetahui pegawai mana yang memiliki tingkat kedisiplinan tinggi, sedang dan rendah sehingga dapat memberikan apresiasi atau penghargaan dan sanksi agar dapat mempertahankan dan meningkatkan kedisiplinannya sehingga pelayanan kepada masyarakat bisa optimal dan visi misi pemerintah daerah bisa tercapai.

Kata kunci: *KDD, Data Mining, Metode K-Means Clustering, Kedisiplinan*

K-means clustering method for measuring employee discipline level (case study in padang pariaman district government)

Abstract

Knowledge Discovery In Database (KDD) is a process of converting raw data into useful data in the form of information. Data mining is a technique of digging up hidden or hidden valuable information in a very large data collection (database) so that an interesting pattern is found that was previously unknown. Clustering is a method in data mining in which data objects that have similarities or the same characteristics are grouped into one group and those that are different are grouped into another group. One aspect of discipline that can be used to evaluate employee performance is attendance. The *k-means* method is used to classify employee discipline levels and then describes the values that have been obtained to generate new knowledge regarding data patterns on employee discipline levels. The attendance data is clustered into 3, namely to measure low, medium, and high levels of discipline. After carrying out the calculation process, the 41 employee samples produced 3 iterations, and the final result was 3 clustering, namely cluster 1 of 10 employees with low discipline, cluster 2 of 7 employees with moderate discipline, and cluster 3 of 24 employees with high discipline. This is intended so that leaders can find out which employees have high, medium and low levels of discipline so that they can provide appreciation or rewards and sanctions in order to maintain and improve their discipline so that service to the community can be optimal and the vision and mission of the local government can be achieved.

Keywords: KDD, Data Mining, K-Means Clustering Method, Discipline

1. PENDAHULUAN

Knowledge Discovery In Database (KDD) merupakan sebuah proses mengubah data mentah menjadi data yang bermanfaat berbentuk informasi [1]. *Knowledge Discovery in Database* (KDD) merupakan proses analisa yang terstruktur bertujuan mendapatkan informasi yang baru dan benar, menemukan pola dari data yang kompleks, dan bermanfaat [2]. Data mining adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (database) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui [3] [4]. Data mining merupakan alternatif untuk membaca tren data yang terjadi dan mampu mengklasifikasikan, mengelompokkan, dan memperkirakan data dalam jumlah besar [5]. Data mining juga dapat digunakan untuk menganalisa sentimen masyarakat [6] [7]. *Clustering* adalah salah satu metode dalam data mining yang dimana objek data yang mempunyai kemiripan atau karakteristik yang sama akan dikelompokkan menjadi satu kelompok dan yang berbeda di kelompokkan pada kelompok yang lainnya [8]. Perbedaan *clustering* dengan klasifikasi yaitu tidak adanya variabel target dalam melakukan suatu pengelompokan pada proses clustering [9].

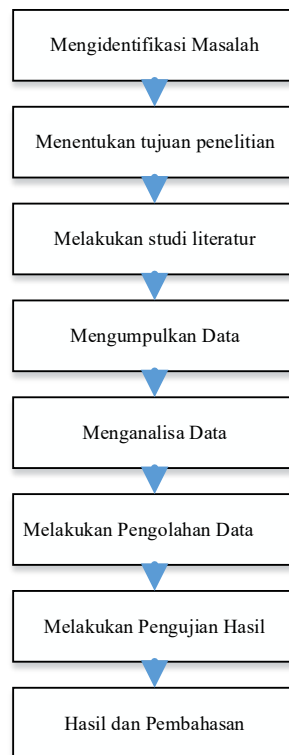
K-Means Clustering adalah teknik pengelompokan data non-hirarki yang memisahkan data ke dalam cluster, mengelompokkan data dengan fitur yang sama, bersama-sama dan mengelompokkan data dengan karakteristik yang berbeda ke dalam kelompok yang berbeda [10] [11]. *K-Means clustering* merupakan algoritma *unsupervised learning* yang membagi data berdasarkan cluster centroid, diikuti dengan menghitung jarak antara data dan centroid yang sesuai dan menetapkan data ke centroid tertentu berdasarkan nilai jarak minimum, serta memperbarui centroid dengan menemukan nilai jarak rata-rata data yang ditetapkan dari centroid [12].

Penelitian terdahulu yang berkaitan dengan metode *k-means clustering* telah banyak dilakukan, seperti dalam penilaian kedisiplinan siswa [13]. Menganalisa tingkat kedisiplinan pegawai pada kantor desa [14]. Melakukan penentuan mata kuliah peminatan mahasiswa [15]. Mengelompokkan perilaku mahasiswa berdasarkan ketidakhadiran, keterlambatan, nilai akademik [16]. Penelitian lainnya yang menggunakan *algoritma k-means clustering* dalam menentukan kepadatan permintaan dari berbagai jenis pekerjaan [17], melakukan zonasi suhu air waduk [18]. Menganalisis kasus Covid 19 di Indonesia [19] dan menganalisa populasi ayam petelur di Indonesia [20].

Pemerintah Padang Pariaman berkomitmen untuk meningkatkan kedisiplinan pegawai dengan melakukan pemotongan tunjangan penghasilan pegawai yang dihitung berdasarkan kinerjanya. Indikator dari kinerja pegawai adalah 40% kehadiran dan 60% tugas dan fungsi yang dituangkan kedalam laporan kinerja harian (Perbub No. 36 Tahun 2021 Tentang TPP ASN). Indikator kehadiran dilihat dari data absensi pegawai yang direkam tiap hari melalui perangkat fingerprint. Saat ini di Pemerintah Kabupaten Padang Pariaman belum melakukan klasterisasi absensi pegawai, sehingga tidak diketahui berapa jumlah pegawai yang tingkat kedisiplinannya tinggi, sedang, dan rendah. Penelitian ini bertujuan untuk mengelompokkan tingkat kedisiplinan pegawai kemudian mendeskripsikan nilai-nilai yang sudah didapatkan untuk menghasilkan sebuah knowledge baru mengenai pola data tingkat kedisiplinan pegawai dan memberikan informasi kepada pimpinan tentang tingkat kedisiplinan pegawai juga sebagai dasar dalam mengambil kebijakan untuk meningkatkan kedisiplinan pegawai.

2. METODE PENELITIAN

Metode penelitian dalam penelitian ini meliputi 8 tahapan yang digunakan untuk memecahkan permasalahan yaitu mengidentifikasi masalah, menentukan tujuan penelitian, melakukan studi literatur, mengumpulkan data, menganalisa data, melakukan pengolahan data, melakukan pengujian hasil, hasil dan pembahasan. Analisis menggambarkan masalah yang ada dan akan diselesaikan dalam penelitian. Desain menggambarkan bagaimana menyelesaikan permasalahan dan disajikan dalam bentuk diagram dengan penjelasan lengkap.



Gambar 1. Kerangka Kerja Penelitian

Berdasarkan gambar rincian kerangka kerja diatas, peneliti menguraikan tahapan kerangka kerja penelitian sebagai berikut :

2.1. Mengidentifikasi Masalah

Tahapan identifikasi masalah merupakan tahap awal dalam melakukan penelitian. Tahapan ini penting karena peneliti melakukan perumusan masalah dari masalah yang ditemukan pada objek penelitian.

2.2. Menentukan Tujuan Penelitian

Menentukan tujuan penelitian diperlukan agar peneliti tidak menyimpang dari tujuan yang ingin dicapai untuk mendapatkan hasil optimal. Tahap ini memperjelas ruang lingkup dan batasan masalah.

2.3. Melakukan Studi Literatur

Pada tahap ini penulis mencari referensi dan landasan teori yang didapatkan dari beberapa sumber seperti buku, jurnal ilmiah dan bahan referensi lainnya yang berhubungan dengan judul dan metode yang digunakan. Proses pencarian bahan tersebut digunakan untuk melengkapi penelitian ini sehingga penulis memiliki bahan referensi yang baik dan relevan.

2.4. Mengumpulkan Data

Data yang digunakan dalam penelitian ini adalah data absensi pegawai dari 2 Organisasi Perangkat Daerah Pemerintah Kabupaten Padang Pariaman yaitu Badan Kepegawaian dan Pengembangan Sumber Daya Manusia (BKPSDM) dan Dinas Kependudukan dan Pencatatan Sipil (Disdukcapil) yang diperoleh dari admin aplikasi E-TPP yang ada di Badan Kepegawaian dan Sumber Daya Manusia Kabupaten Padang Pariaman.

2.5. Menganalisa Data

Pada tahap ini dilakukan analisis data yang dikumpulkan. Analisa data dilakukan untuk menseleksi data yang akan digunakan sehingga data yang diolah nantinya bebas dari noise dan data yang tidak konsisten atau data yang tidak relevan. Proses pada analisa data yaitu *data cleaning*, *data integration* dan *data selection*.

2.6. Melakukan Pengolahan Data

Pengolahan data ini dilakukan untuk mendapatkan hasil dari pengelompokan data yang ada, dimana data tersebut nantinya akan menjadi 3 cluster, yang meliputi tingkat kedisiplinan pegawai yang tinggi, sedang dan rendah. Adapun tahapan pengolahan data menggunakan algoritma *K-Means Clustering* antara lain :

- Masukan data yang akan di *cluster*
- Tentukan jumlah *cluster*
- Tentukan nilai centroid awal secara acak sebagai pusat *cluster*, menentukan nilai centroid iterasi berikutnya menggunakan rumus :

$$CI = (R1 + R2 + R3 + \dots + Rn)/(\Sigma R) \quad (1)$$

Keterangan :

CI = centroid baru

R1 = Nilai data ke-1

Rn = Nilai data ke-n

ΣR = jumlah data.

- d. Menghitung jarak data dengan pusat *centroid* dengan persamaan *Eucliden Distance*

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2} \quad (2)$$

Keterangan :

De = *Euclidean Distance*

i = banyaknya objek

(x,y) = koordinat objek

(s,t) = koordinat centroid.

- e. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang baru
f. Jika pusat *cluster* tidak ada perubahan maka proses *cluster* telah selesai. Jika belum maka ulangi langkah ke 4 hingga nilai centroid yang dihasilkan tetap dan anggota *cluster* tidak berpindah ke *cluster* lain

2.7. Melakukan Pengujian Hasil

Pada tahap ini peneliti menemukan dan menganalisis hasil dari perhitungan *k-means clustering* baik itu hasil dari perhitungan secara manual maupun dengan menggunakan aplikasi data mining yaitu dengan *RapidMiner*. Tujuan tahap ini adalah untuk mengetahui kesesuaian antara hasil output dari perhitungan manual dengan perhitungan menggunakan aplikasi.

2.8. Hasil dan Pembahasan

Hasil dari penelitian di uraikan dengan menganalisa hasil pengujian yang diperoleh dari data yang telah diolah menggunakan algoritma *k-means clustering* berdasarkan permasalahan pada tingkat kedisiplinan pegawai. Kemudian hasil yang didapat bisa dijadikan masukan bagi pimpinan dalam mengambil keputusan terhadap output tingkat kedisiplinan pegawai. Sehingga peneliti dapat menentukan kesimpulan dari hasil penelitian. Tahapan ini juga menjelaskan tentang kesesuaian antara rumusan masalah dengan tujuan penelitian.

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah data absensi pegawai pada Badan Kepegawaian dan Pengembangan Sumber Daya Manusia (BKPSDM) dan Dinas Kependudukan dan Pencatatan Sipil (Disdukcapil) pada bulan Juni, Juli dan Agustus Tahun 2022. Melalui metode *K-Means Clustering*, data kehadiran pegawai akan diolah dan diclusterkan untuk mendapatkan cluster pegawai yang tingkat kedisiplinan nya rendah, sedang dan tinggi. Hasil *cluster* dapat dijadikan acuan bagi pimpinan untuk melihat tingkat kedisiplinan pegawai dan sebagai bahan bagi pimpinan dalam membuat kebijakan untuk meningkatkan kedisiplinan pegawai. Atribut yang digunakan terdiri dari 5 atribut yaitu hadir, terlambat, pulang sebelum waktunya, tanpa keterangan dan cuti.

Tabel 1. Data Absensi Pegawai Bulan Juni, Juli dan Agustus 2022

NO	Nama	OPD	H	TL	PSW	TK	C
1	Maizar	BKPSDM	36	23	5	0	0
2	Eliza	BKPSDM	57	2	2	1	2
3	Revid Hidayat	BKPSDM	49	10	0	0	5
4	Dewi Anggraini	BKPSDM	46	18	0	0	0
5	Fitri Agustin Azhar	BKPSDM	53	9	0	0	2
6	Suarni	BKPSDM	48	13	0	0	3
7	Metra Nelida	BKPSDM	39	18	1	0	6
8	Masfufah	BKPSDM	31	27	1	0	5
9	Mukhtar Luthfi	BKPSDM	50	8	1	0	5
10	Riyan Fikri	BKPSDM	36	23	2	0	3
11	Alfakhri Syukri	BKPSDM	50	4	6	0	4
12	Syanti Fitri Angraini	BKPSDM	38	16	5	0	5
13	Hardi Fardian	BKPSDM	36	22	0	1	5
14	Sal Amini Eka Putra	BKPSDM	62	1	0	0	1
15	Armen Marta	BKPSDM	53	8	3	0	0
16	Lucy Andria	BKPSDM	28	26	5	0	5
17	Nia Lailatul Azmi	BKPSDM	21	34	2	0	7
18	Nila Purnama	BKPSDM	60	3	0	0	1
19	Satria Ms	BKPSDM	46	11	0	1	6
20	Reni Oktavia	BKPSDM	26	33	3	0	2
21	Febria Hasni Folia	BKPSDM	24	37	0	0	3
22	Agustina Murni Nur	BKPSDM	57	7	0	0	0
23	Dian Novita	BKPSDM	49	11	0	0	4
24	Jon Kenedi	BKPSDM	26	37	1	0	0
25	Ratna Sari Astuti	BKPSDM	31	25	1	0	7
26	Zeta Hidayati	Disdukcapil	62	0	2	0	0
27	Anda Marzuki	Disdukcapil	55	4	0	0	5

NO	Nama	OPD	H	TL	PSW	TK	C
28	Andriyani	Disdukcapi	52	9	3	0	0
29	Wirman Wahyudi	Disdukcapi	63	1	0	0	0
30	Mariati	Disdukcapi	41	23	0	0	0
31	Yusneli Roza	Disdukcapi	55	8	0	1	0
32	Rina Angriani	Disdukcapi	29	28	0	0	7
33	Rahmi Marisa	Disdukcapi	50	14	0	0	0
34	Ali Muzakar	Disdukcapi	58	3	0	0	3
35	Yeni Zulvia	Disdukcapi	53	7	3	0	1
36	Ikhwan Pardamean S	Disdukcapi	58	5	0	1	0
37	Mirawati	Disdukcapi	20	36	4	0	4
38	Andri	Disdukcapi	57	6	0	1	0
39	Zulhari	Disdukcapi	27	29	5	3	0
40	Renol Pajri	Disdukcapi	61	1	2	0	0
41	Riky Yuliardi	Disdukcapi	57	5	2	0	0

Keterangan :

OPD = Organisasi Perangkat Daerah
H = Hadir
TL = Terlambat
PSW = Pulang Sebelum Waktunya
TK = Tanpa Keterangan
C = Cuti

3.1 Menentukan jumlah cluster

Pada penelitian ini penulis melakukan klusterisasi data menjadi 3 kluster, yaitu C1 (Pegawai dengan tingkat kedisiplinan rendah), C2 (Pegawai dengan tingkat kedisiplinan sedang), dan C3 (Pegawai dengan tingkat kedisiplinan tinggi).

3.2. Menentukan nilai centroid awal

Untuk menentukan nilai centroid awal dilakukan secara acak dari sampel yang digunakan dalam penelitian. Centroid awal ini merupakan titik pusat cluster pertama. Centroid untuk C1 di ambil dari sampel data ke 8, C2 diambil dari sampel data ke 13 dan C3 di ambil dari sampel data ke 19. Bentuk centroid awal yang digunakan dapat dilihat pada tabel berikut :

Tabel 2. Nilai Centroid Awal

Cluster	Data ke-	H	TL	PSW	TK	C
C1	8	31	27	1	0	5
C2	13	36	22	0	1	5
C3	19	46	11	0	1	6

3.3 Menghitung jarak data dengan centroid iterasi 1

Menghitung jarak data ke 1 terhadap pusat cluster iterasi 1 menggunakan rumus *Euclidean Distance* (2) seperti berikut :

$$\begin{aligned}
 \text{a. } C1 &= \sqrt{(36 - 31)^2 + (23 - 27)^2 + (5 - 1)^2 + (0 - 0)^2 + (0 - 5)^2} \\
 &= \sqrt{(5)^2 + (-4)^2 + (4)^2 + (0)^2 + (-5)^2} \\
 &= \sqrt{82} \\
 &= 9,005 \\
 \text{b. } C2 &= \sqrt{(36 - 36)^2 + (23 - 22)^2 + (5 - 0)^2 + (0 - 1)^2 + (0 - 5)^2} \\
 &= \sqrt{(0)^2 + (1)^2 + (5)^2 + (-1)^2 + (-5)^2} \\
 &= \sqrt{52} \\
 &= 7,211 \\
 \text{c. } C3 &= \sqrt{(36 - 46)^2 + (23 - 11)^2 + (5 - 0)^2 + (0 - 1)^2 + (0 - 6)^2} \\
 &= \sqrt{(-10)^2 + (12)^2 + (5)^2 + (-1)^2 + (-6)^2} \\
 &= \sqrt{306} \\
 &= 17,493
 \end{aligned}$$

Selanjutnya melakukan pencarian jarak data hingga data ke 41 dan mengelompokkan keanggotaan cluster dengan memilih nilai minimum dari jarak data terhadap pusat cluster. Hasil pencarian dan pengelompokan data dapat dilihat pada tabel berikut :

Tabel 3. Hasil Cluster Iterasi 1

NO	C1	C2	C3	Keanggotaan Cluster		
				C1	C2	C3
1	9,055	7,211	17,493		C2	
2	36,222	29,223	14,900			C3
3	24,779	17,720	3,464			C3

NO	C1	C2	C3	Keanggotaan Cluster		
				C1	C2	C3
4	18,221	11,916	9,274			C3
5	28,601	21,633	8,367			C3
6	22,136	15,166	4,243			C3
7	12,083	5,292	10		C2	
8	0,000	7,211	22	C1		
9	26,870	19,849	5,292			C3
10	6,782	3,162	16,062		C2	
11	30,265	23,622	10,296			C3
12	13,638	8,124	10,770		C2	
13	7,211	0,000	14,900		C2	
14	40,669	33,675	19,545			C3
15	29,563	22,804	10,198			C3
16	5,099	10,296	24,000	C1		
17	12,410	19,442	34,059	C1		
18	37,868	30,887	16,912			C3
19	22,000	14,900	0,000			C3
20	8,602	15,492	30,166	C1		
21	12,410	19,339	34,205	C1		
22	33,196	26,306	13,191			C3
23	24,125	17,088	3,742			C3
24	12,247	18,762	33,377	C1		
25	2,828	6,325	20,591	C1		
26	41,425	34,496	20,445			C3
27	33,257	26,192	11,489			C3
28	28,178	21,448	9,274			C3
29	41,545	34,583	20,640			C3
30	11,916	7,211	14,353		C2	
31	31,048	24,125	11,225			C3
32	3,162	9,487	24,083	C1		
33	23,580	16,912	7,874			C3
34	36,194	29,155	14,765			C3
35	30,067	23,238	10,000			C3
36	35,214	28,249	14,697			C3
37	14,560	21,679	36,359	C1		
38	33,823	26,870	13,491			C3
39	8,367	13,565	27,386	C1		
40	40,025	33,106	19,131			C3
41	34,438	27,568	14,071			C3

Dari hasil pencarian yang dilakukan pada iterasi 1 didapatkan pengelompokan data anggota cluster 1 (C1) terdiri dari 10 pegawai dengan nomor: 8,16,17,20,21,24,25,32,37 dan 39. Anggota cluster 2 (C2) terdiri dari 6 pegawai dengan nomor: 1,7,10,12,13,30 dan anggota cluster 3 (C3) terdiri dari 25 pegawai dengan nomor: 2,3,4,5,6,9,11,14,15,18,19,22,26,27,28,29,31,33,34,35,36,38, 40,41.

3.4 Menghitung nilai centroid iterasi 3

Dari hasil perhitungan pada Iterasi 3 didapatkan anggota cluster yang tidak berpindah antara hasil cluster iterasi 2 dengan hasil cluster iterasi 3, maka proses perhitungan selesai dan berhenti pada Iterasi ke 3. Nilai centroid iterasi ke 3 dihitung berdasarkan nilai rata – rata masing – masing cluster dari hasil proses iterasi 2, perhitungan untuk menentukan nilai centroid baru untuk iterasi 3 menggunakan rumus persamaan (1) :

- $$C1 = (31+28+21+26+24+26+31+29+20+27)/10 = 26,300$$

$$= (27+26+34+33+37+37+25+28+36+29)/10 = 31,200$$

$$= (1+5+2+3+0+1+1+0+4+5)/10 = 2,200$$

$$= (0+0+0+0+0+0+0+0+0+3)/10 = 0,300$$

$$= (5+5+7+2+3+0+7+7+4+0)/10 = 4,000$$
- $$C2 = (36+46+39+36+38+36+41)/7 = 38,857$$

$$= (23+18+18+23+16+22+23)/7 = 20,429$$

$$= (5+0+1+2+5+0+0)/7 = 1,857$$

$$= (0+0+0+0+0+1+0)/7 = 0,143$$

$$= (0+0+6+3+5+5+0)/7 = 2,714$$
- $$C3 = (57+49+53+48+50+50+62+53+60+46+57+49+62+55+52+53+55+50+58+53+58+57+62+57)/24 = 54,792$$

$$= (2+10+9+13+8+4+1+8+3+11+7+11+0+4+9+1+8+14+3+7+5+6+1+1)/24 = 6,250$$

$$= (2+0+0+0+1+6+0+3+0+0+0+2+0+3+0+0+0+3+0+0+2+2)/24 = 1$$

$$= (1+0+0+0+0+0+0+0+1+0+0+0+0+0+1+0+0+0+1+1+0+0)/24 = 0,208$$

$$= (2+5+2+3+5+4+1+0+1+6+0+4+0+5+0+0+0+0+3+1+0+0+0+0)/24 = 1,750$$

Hasil perhitungan nilai centroid untuk iterasi ke 3 dapat dilihat pada tabel berikut :

Tabel 4. Nilai Centroid Iterasi 3

Cluster	H	TL	PSW	TK	C
C1	26,300	31,200	2,200	0,300	4,000
C2	38,857	20,429	1,857	0,143	2,714
C3	54,792	6,250	1,000	0,208	1,750

3.5 Menghitung jarak data dengan centroid iterasi 3

Menghitung kembali jarak data 1 ke centroid / pusat *cluster* iterasi 3. Proses perhitungan jarak data menggunakan rumus *Euclidean Distance* (2). Proses perhitungan sebagai berikut :

Menghitung jarak data ke 1 terhadap pusat *cluster* iterasi 3

- $$C1 = \sqrt{(36 - 26,300)^2 + (23 - 31,200)^2 + (5 - 2,200)^2 + (0 - 0,300)^2 + (0 - 4,000)^2}$$

$$= \sqrt{(9,700)^2 + (-8,200)^2 + (2,800)^2 + (-0,300)^2 + (-0,400)^2}$$

$$= \sqrt{169,42}$$

$$= 13,611$$
- $$C2 = \sqrt{(36 - 38,857)^2 + (23 - 20,429)^2 + (5 - 1,857)^2 + (0 - 0,143)^2 + (0 - 2,714)^2}$$

$$= \sqrt{(-2,857)^2 + (2,571)^2 + (3,143)^2 + (-0,143)^2 + (-2,714)^2}$$

$$= \sqrt{32,037}$$

$$= 5,660$$
- $$C3 = \sqrt{(36 - 54,792)^2 + (23 - 6,250)^2 + (5 - 1)^2 + (0 - 0,208)^2 + (0 - 1,750)^2}$$

$$= \sqrt{(-18,792)^2 + (16,75)^2 + (4)^2 + (-0,208)^2 + (-1,750)^2}$$

$$= \sqrt{652,808}$$

$$= 25,550$$

Kemudian melakukan pencarian jarak data hingga data ke 41 dan mengelompokkan keanggotaan cluster dengan memilih nilai minimum dari jarak data terhadap pusat cluster. Hasil pencarian dan pengelompokan data pada iterasi 3 dapat dilihat pada tabel berikut:

Tabel 5. Hasil Cluster Iterasi 3

NO	C1	C2	C3	Keanggotaan Cluster		
				C1	C2	C3
1	13,611	5,660	25,550		C2	
2	42,422	25,885	4,963			C3
3	31,155	14,843	7,695			C3
4	24,151	8,231	14,814		C2	
5	34,852	18,292	3,447			C3
6	28,426	11,930	9,711			C3
7	18,468	4,180	20,138		C2	
8	6,501	10,531	31,737	C1		
9	33,203	16,870	6,052			C3
10	12,746	3,860	25,225		C2	
11	36,278	20,320	7,624			C3
12	19,413	5,956	20,091		C2	
13	13,604	4,477	24,766		C2	
14	46,909	30,323	9,007			C3
15	35,607	19,057	3,658			C3
16	6,234	12,808	33,682	C1		
17	6,713	22,836	44,052	C1		
18	44,101	27,517	6,269			C3
19	28,381	12,446	10,934			C3
20	2,839	18,033	39,353	C1		
21	6,698	22,336	43,546	C1		
22	39,358	22,811	3,090			C3
23	30,467	14,032	7,887			C3
24	7,160	21,167	42,162	C1		
25	8,430	10,087	30,744	C1		
26	47,582	30,989	9,753			C3
27	39,616	23,220	4,088			C3
28	34,206	17,665	4,739			C3
29	47,748	31,164	9,952			C3
30	17,443	4,695	21,792		C2	
31	37,192	20,655	2,792			C3

NO	C1	C2	C3	Keanggotaan Cluster		
				C1	C2	C3
32	5,609	13,279	34,160	C1		
33	29,639	13,279	9,334			C3
34	42,498	25,957	4,844			C3
35	36,170	19,612	2,895			C3
36	41,384	24,820	4,068			C3
37	8,128	24,583	45,931	C1		
38	39,986	23,429	3,103			C3
39	6,038	15,475	36,288	C1		
40	46,176	29,584	8,379			C3
41	40,559	23,971	3,247			C3

Hasil yang diperoleh dari iterasi 3 terdapat pengelompokan anggota cluster 1 (C1) adalah 10 pegawai dengan nomor: 8,16,17, 20,21,24,25,32,37 dan 39. Anggota cluster 2 (C2) adalah 7 pegawai dengan nomor: 1,4,7,10,12,13,30 serta anggota cluster 3 (C3) adalah 24 pegawai dengan nomor: 2,3,5,6,9,11,14,15,18,19,22,23,26,27,28,29,31,33,34,35,36,38,40 dan 41.

3.6 Hasil metode *k-means clustering*

Dari perhitungan manual yang dilakukan dalam pengolahan data menggunakan metode *k-means clustering* didapatkan hasil anggota tiap cluster. Cluster 1 merupakan cluster yang anggotanya adalah pegawai dengan tingkat kedisiplinan rendah. Hasil dari cluster 1 dapat dilihat pada table berikut :

Tabel 6. Cluster 1 (Tingkat Kedisiplinan Rendah)

NO	Nama	OPD	H	TL	PSW	TK	Cuti	Cluster
8	Masfufah	BKPSDM	31	27	1	0	5	C1
16	Lucy Andria	BKPSDM	28	26	5	0	5	C1
17	Nia Lailatul Azmi	BKPSDM	21	34	2	0	7	C1
20	Reni Oktavia	BKPSDM	26	33	3	0	2	C1
21	Febria Hasni Folia	BKPSDM	24	37	0	0	3	C1
24	Jon Kenedi	BKPSDM	26	37	1	0	0	C1
25	Ratna Sari Astuti	BKPSDM	31	25	1	0	7	C1
32	Rina Angriani	Disdukcapil	29	28	0	0	7	C1
37	Mirawati	Disdukcapil	20	36	4	0	4	C1
39	Zulhari	Disdukcapil	27	29	5	3	0	C1
Jumlah								10

Selanjutnya cluster 2 merupakan cluster yang anggotanya pegawai dengan tingkat kedisiplinan sedang. anggota pada cluster 2 dapat dilihat pada tabel berikut:

Tabel 7 Cluster 2 (Tingkat Kedisiplinan Sedang)

NO	Nama	OPD	H	TL	PSW	TK	Cuti	Cluster
1	Maizar	BKPSDM	36	23	5	0	0	C2
4	Dewi Anggraini	BKPSDM	46	18	0	0	0	C2
7	Metra Nelida	BKPSDM	39	18	1	0	6	C2
10	Riyan Fikri	BKPSDM	36	23	2	0	3	C2
12	Syanti Fitri Angraini	BKPSDM	38	16	5	0	5	C2
13	Hardi Fardian	BKPSDM	36	22	0	1	5	C2
30	Mariati	Disdukcapil	41	23	0	0	0	C2
Jumlah								7

Cluster 3 merupakan cluster yang anggotanya pegawai dengan tingkat kedisiplinan tinggi. Data hasil pada cluster 3 adalah terlihat pada table berikut:

Tabel 8. Cluster 3 (Tingkat Kedisiplinan Tinggi)

NO	Nama	OPD	H	TL	PSW	TK	Cuti	Cluster
2	Eliza	BKPSDM	57	2	2	1	2	C3
3	Revid Hidayat	BKPSDM	49	10	0	0	5	C3
5	Fitri Agustin Azhar	BKPSDM	53	9	0	0	2	C3
6	Suarni	BKPSDM	48	13	0	0	3	C3
9	Mukhtar Luthfi	BKPSDM	50	8	1	0	5	C3
11	Alfakhri Syukri	BKPSDM	50	4	6	0	4	C3
14	Sal Amini Eka Putra	BKPSDM	62	1	0	0	1	C3
15	Armen Marta	BKPSDM	53	8	3	0	0	C3
18	Nila Purnama	BKPSDM	60	3	0	0	1	C3
19	Satria Ms	BKPSDM	46	11	0	1	6	C3
22	Agustina Murni Nur	BKPSDM	57	7	0	0	0	C3
23	Dian Novita	BKPSDM	49	11	0	0	4	C3
26	Zeta Hidayati	Disdukcapil	62	0	2	0	0	C3
27	Anda Marzuni	Disdukcapil	55	4	0	0	5	C3
28	Andriyani	Disdukcapil	52	9	3	0	0	C3
29	Wirman Wahyudi	Disdukcapil	63	1	0	0	0	C3
31	Yusneli Roza	Disdukcapil	55	8	0	1	0	C3

33	Rahmi Marisa	Disdukcapil	50	14	0	0	0	C3
34	Ali Muzakar	Disdukcapil	58	3	0	0	3	C3
35	Yeni Zulvia	Disdukcapil	53	7	3	0	1	C3
36	Ikhwan Pardamean S	Disdukcapil	58	5	0	1	0	C3
38	Andri	Disdukcapil	57	6	0	1	0	C3
40	Renol Pajri	Disdukcapil	61	1	2	0	0	C3
41	Riky Yuliardi	Disdukcapil	57	5	2	0	0	C3
Jumlah								24

4. KESIMPULAN

Dari hasil penelitian yang telah diuraikan disimpulkan bahwa metode *k-means clustering* untuk mengukur tingkat kedisiplinan pegawai, menghasilkan sebuah informasi mengenai data pengelompokan tingkat kedisiplinan rendah, sedang dan tinggi. Dari 41 data absensi terdapat anggota C1 (tingkat disiplin rendah) 10 pegawai, C2 (tingkat disiplin sedang) 7 pegawai, dan C3 (tingkat disiplin tinggi) 24 pegawai. Penelitian ini sebagai dasar bagi pimpinan untuk membuat kebijakan dalam hal meningkatkan kedisiplinan pegawai.

DAFTAR PUSTAKA

- [1] ud Hermansyah, R. Aditia Hamdan, F. Sidik, and A. Wibowo, "Klasterisasi Data Travel Umroh di Marketplace Umroh.com Menggunakan Metode K-Means".
- [2] D. A. Fakhri, S. Defit, and Sumijan, "Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-Means Clustering," *Jurnal Informasi dan Teknologi*, pp. 160–166, Sep. 2021, doi: 10.37034/jidt.v3i3.137.
- [3] H. Parasian Doloksaribu and Y. T. Samuel, "KOMPARASI ALGORITMA DATA MINING UNTUK ANALISIS SENTIMEN APLIKASI PEDULILINDUNG1," vol. 16, no. 1, 2022, doi: 10.47111/JTL.
- [4] N. Pradita, "Penerapan Data Mining sebagai Cara untuk Memprediksi Prestasi Siswa Berdasarkan Status Ekonomi dan Kedisiplinan Menggunakan Metode Regresi Linier Berganda." [Online]. Available: <https://jurnal.universitaspurabangsa.ac.id/index.php/jimmmba/index>
- [5] S. Ika Murpratiwi, D. Ayu Indah Cahya Dewi, and A. Aranta, "Accuracy Analysis of Predictive Value in Transaction Data of Service Company Using Combination of K-Means Clustering and Time Series Methods." [Online]. Available: <http://jcosine.if.unram.ac.id/>
- [6] E. B. Susanto, Paminto Agung Christianto, Mohammad Reza Maulana, and Satriedi Wahyu Binabar, "Analisis Kinerja Algoritma Naïve Bayes Pada Dataset Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 234–241, Dec. 2022, doi: 10.37859/coscitech.v3i3.4343.
- [7] Aditya Quantano Surbakti, Regiolina Hayami, and Januar Al Amien, "Analisa Tanggapan Terhadap Psbb Di Indonesia Dengan Algoritma Decision Tree Pada Twitter," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 2, no. 2, pp. 91–97, Dec. 2021, doi: 10.37859/coscitech.v2i2.2851.
- [8] A. A. Aldino, D. Darwis, A. T. Prastowo, and C. Sujana, "Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jan. 2021. doi: 10.1088/1742-6596/1751/1/012038.
- [9] S. Sindi *et al.*, "ANALISIS ALGORITMA K-MEDOIDS CLUSTERING DALAM PENGELOMPOKAN PENYEBARAN COVID-19 DI INDONESIA," *Jurnal Teknologi Informasi*, vol. 4, no. 1, 2020.
- [10] T. Amalina, D. Bima, A. Pramana, and B. N. Sari, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 15, pp. 574–583, 2022, doi: 10.5281/zenodo.7052276.
- [11] R. T. Vlandari, W. Laksito, Y. Saptomo, and D. W. Aditama, "Application of K-Means Clustering in Mapping of Central Java Crime Area," 2020.
- [12] M. Nator and W. Obaid, "Detection and localization of early-stage multiple brain tumors using a hybrid technique of patch-based processing, k-means clustering and object counting," *Int J Biomed Imaging*, vol. 2020, 2020, doi: 10.1155/2020/9035096.
- [13] Q. I. Mawarni and E. S. Budi, "Implementasi Algoritma K-Means Clustering Dalam Penilaian Kedisiplinan Siswa," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 3, no. 4, p. 522, Jun. 2022, doi: 10.30865/json.v3i4.4242.
- [14] D. Murni, B. Efendi, N. Rahmadani, S. Informasi, and S. Tinggi Manajemen Informatika dan Komputer Royal Kisaran, "IMPLEMENTATION OF EMPLOYEE DISCIPLINE CLUSTERING AT GOTTING SIDODADI VILLAGE OFFICE BANDAR PASIR MANDOGUE USING K-MEANS ALGORITHM," *Jurnal Teknik Informatika (JUTIF)*, vol. 3, no. 2, pp. 295–304, 2022, doi: 10.20884/1.jutif.2022.3.2.236.
- [15] L. N. Rani, S. Defit, and L. J. Muhammad, "Determination of Student Subjects in Higher Education Using Hybrid Data Mining Method with the K-Means Algorithm and FP Growth," *International Journal of Artificial Intelligence Research*, vol. 5, no. 1, Dec. 2021, doi: 10.29099/ijair.v5i1.223.
- [16] J. G. Saren, H. Talirongan, F. Jean B. Talirongan, and C. L. Malicay, "Mining Student Behavioral Concern Through Referrals Using K-Means Clustering," *International Journal of Research Publications*, vol. 71, no. 1, Feb. 2021, doi: 10.47119/ijrp10071120211768.

- [17] F. M. Javed Mehedi Shamrat, Z. Tasnim, I. Mahmud, N. Jahan, and N. Islam Nobel, "Application Of K-Means Clustering Algorithm To Determine The Density Of Demand Of Different Kinds Of Jobs," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, vol. 9, p. 2, 2020, [Online]. Available: www.ijstr.org
- [18] W. Liu, P. Zou, D. Jiang, X. Quan, and H. Dai, "Zoning of reservoir water temperature field based on K-means clustering algorithm," *J Hydrol Reg Stud*, vol. 44, Dec. 2022, doi: 10.1016/j.ejrh.2022.101239.
- [19] T. Hardiani, "Analisis Clustering Kasus Covid 19 di Indonesia Menggunakan Algoritma K-Means," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 11, no. 2, pp. 156–165, Aug. 2022, doi: 10.23887/janapati.v11i2.45376.
- [20] E. Ramadanti and M. Muslih, "PENERAPAN DATA MINING ALGORITMA K-MEANS CLUSTERING PADA POPULASI AYAM PETELUR DI INDONESIA," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 7, no. 1, pp. 1–7, Jan. 2022, doi: 10.36341/rabit.v7i1.2155.