

## Penerapan Data Mining Dalam Menentukan Tingkat Kedisiplinan Karyawan Perhotelan Menggunakan Algoritma *K-Means Clustering*

Preddy Marpaung<sup>1\*</sup>, Ibnu Febrian<sup>2</sup>, Widia Putri<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, STMIK Pelita Nusantara, Medan, Indonesia

Email: <sup>1</sup>preddymarpaung2@gmail.com, <sup>2</sup>ibnufebrian77@gmail.com, <sup>3</sup>jelekjenok@gmail.com

Email Penulis Korespondensi: <sup>1</sup>preddymaraung2@gmail.com

**Abstrak**—Perhotelan merupakan salah satu bisnis usaha di bidang pariwisata. Salah satu perhotelan yang ada di Deli Serdang adalah Hotel Halay. Dimana selama ini dalam menentukan tingkat kedisiplinan karyawannya hanya melalui observasi dan tafsiran manager, sehingga kurang efektif dan kurang professional dalam menentukan kedisiplinan karyawannya. Tujuan dalam menentukan tingkat kedisiplinan karyawan adalah untuk meningkatkan pelayanan terbaik kepada pelanggan melalui para karyawan, sehingga dibuat penilaian tingkat kedisiplinan kepada karyawannya untuk diberikan penghargaan dan pembinaan kepada karyawan yang kurang disiplin. Penentuan tingkat kedisiplinan di tentukan ke dalam 3 *cluster* yaitu Sangat Baik (C1), Baik (C2), dan Cukup (C3). Algoritma yang diterapkan adalah algoritma *k-means clustering*. Dimana hasil dari penerapan algoritma ini dari 10 karyawan dapat menghasilkan Sangat Baik (C1) sebanyak 6 data, Baik (C2) sebanyak 2 data, dan Cukup (C3) sebanyak 2 data. Sehingga dengan adanya hasil yang didapat, bisa membantu pihak perhotelan untuk mengevaluasi para karyawannya setiap episode.

**Kata Kunci:** Perhotelan, Tingkat Kedisiplinan, Karyawan, Algoritma K-Means Clustering

**Abstract**— Hospitality is one of the businesses in the tourism sector. One of the hotels in Deli Serdang is the Halay Hotel. Where so far, determining the level of discipline of employees has only been through observations and interpretations of managers, so it is less effective and less professional in determining employee discipline. The aim in determining the level of employee discipline is to improve the best service to customers through employees, so that an assessment of the level of discipline of employees is made to provide rewards and guidance to employees who lack discipline. The level of discipline is determined into 3 clusters, namely Very Good (C1), Good (C2), and Fair (C3). The algorithm applied is the k-means clustering algorithm. Where the results of applying this algorithm from 10 employees can produce Very Good (C1) with 6 data, Good (C2) with 2 data, and Fair (C3) with 2 data. So with the results obtained, it can help hoteliers to evaluate their employees every episode.

**Keywords:** Hospitality, Discipline Level, Employees, K-Means Clustering Algorithm

### 1. PENDAHULUAN

Perhotelan merupakan salah satu bisnis usaha di bidang pariwisata. Perkembangan bisnis perhotelan mengalami peningkatan yang cukup baik beberapa tahun terakhir, termasuk di daerah Deli Serdang, Sumatera Utara. Oleh karena itu, setiap perhotelan menuntut para karyawannya untuk memberikan pelayanan yang terbaik kepada pelanggan atau konsumen. Dalam memberikan pelayanan, tentu kedisiplinan karyawan menjadi dasar utama dalam memberikan pelayanan terbaiknya[1]

Salah satu perhotelan yang ada di Deli Serdang adalah Hotel Halay. Dimana selama ini dalam menentukan tingkat kedisiplinan karyawannya hanya melalui observasi dan tafsiran manager, sehingga kurang efektif dan kurang professional dalam menentukan kedisiplinan karyawannya. Tujuan dalam menentukan tingkat kedisiplinan karyawan adalah untuk meningkatkan pelayanan terbaik kepada pelanggan melalui para karyawan, sehingga dibuat penilaian tingkat kedisiplinan kepada karyawannya untuk diberikan penghargaan dan pembinaan kepada karyawan yang kurang disiplin. Oleh karena itu perlu dibuat sebuah solusi untuk memecahkan permasalahan dalam menentukan tingkat kedisiplinan karyawan, sehingga dengan solusi tersebut dapat menghasilkan tingkat kedisiplinan karyawan yang lebih objektif dan professional.

Solusi yang tepat diterapkan adalah dengan menerapkan ilmu komputer yakni data mining. Data mining adalah proses pengumpulan sebuah informasi penting pada suatu data yang berukuran besar. Untuk pengumpulan data tersebut dapat dilakukan melalui proses perhitungan statistika, matematika, maupun penggunaan teknologi AI (Artificial Intelligence). Istilah lain dari data mining sendiri dapat berarti penambangan data yang berbentuk sebuah tool untuk melakukan analisa dengan teknik penyaringan informasi secara lebih akurat[2][3]. Dalam data mining banyak algoritma yang dihasilkan, namun menyelesaikan masalah yang ada pada penelitian ini hanya ada sebuah algoritma.

Algoritma yang digunakan dalam menyelesaikan permasalahan adalah algoritma *k-means clustering*. Melalui algoritma ini akan di tentukan tingkat kedisiplinan karyawan kedalam 3 cluster, yaitu Sangat Baik (C1), Baik (C2), Cukup (C3). Alasan penenliti menentukan algoritma k-means clustering merupakan salah satu algoritma pengelompokan yang paling populer dan sederhana, dimana pertama kali diterbitkan pada tahun 1955 dan sampai sekarang K-means masih banyak digunakan. K-Means Clustering juga merupakan suatu metode untuk menganalisis data, yang sering digunakan dalam metode data mining. pengelompokan dapat dimanfaatkan untuk mengidentifikasi

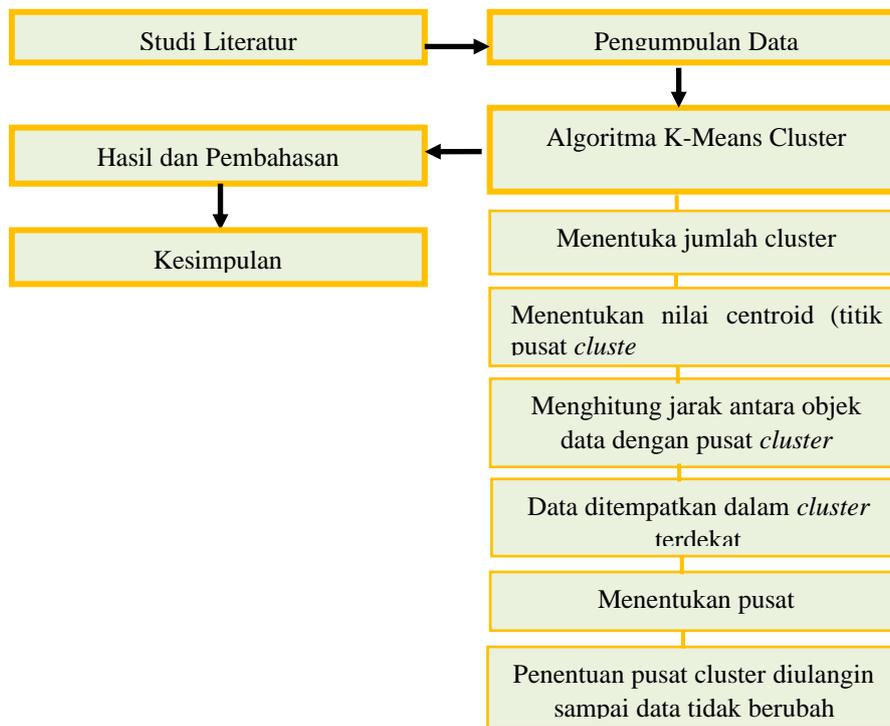
kasus berdasarkan kelompok atribut, atau dapat dikatakan mengelompokkan data yang memiliki atribut yang sama. Metode clustering yang paling populer yaitu metode K-Means Clustering[4][5]

Berdasarkan masalah diatas dan refresnsi jurnal penelitian terdahulu dalam penyelesaian masalah, maka melalui penelitian ini perlu ditentukan tingkat kedisiplinan karyawan hotel ke dalam 3 cluster yakni, sangat baik (c1), baik (c2), dan cukup(c3). Sehingga penelian ini diangkat dengan judul “Penerapan Data Mining Dalam Menentukan Tingkat Kedisiplinan Karyawan Perhotelan Menggunakan Algoritma K-Means Clustering”. Dengan demikian, akan dapat membantu pihak perhotelan dalam menentukan kedisiplinan karyawannya setiap efisode.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian dimulai dengan melakukan literature kepada mitra , kemudian dilanjutkan literature dari jurnal terkait masalah utama penentuan tingkat kedisiplinan karyawan hotel, dan juga algoritma yang sesuai untuk digunakan dalam penyelesaian masalah. Algoritma K-Means Clustering digunakan dalam menentukan tingkat kedisiplinan pada karyawan hotel halay kedalam kelompok (cluster) dengan kategori sangat baik(C1),baik(C2), dan cukup (C3). Tahapan yang akan dilakukan dalam penelitian ini secara keseluruhan dapat dilihat sebagai berikut:



Gambar 1: Tahapan Penelitian

### 2.2 Pengumpulan Data

Adapun pengumpulan data yang digunakan dalam penelitian ini adalah hasil dari hasil wawancara dan data dari jurnal sebelumnya yang sudah pernah dilakukan dalam menentukan waiters terbaik[6]:

Adapun dataset yang digunakan pada penelitian ini sebagai berikut:

Tabel 1. Rekapulasi Data dan Nilai karyawan

No	Nama Karyawan	Absensi	Penampil an Diri	Pelayanan	Sikap & Loyalitas	Tugas & Tanggung Jawab	Kerjasama Tim
1	Bangbang	70	70	80	90	90	75
2	Jenerius S	60	80	75	80	80	80
3	Hansen S	60	65	70	70	75	80

4	Erwin M	60	75	60	80	60	60
5	Lastrodo	80	80	70	80	80	80
6	Horas S	80	70	70	70	80	80
7	Freddy H	80	70	75	70	80	70
8	Wanda S	60	60	70	65	70	70
9	Ronaldo	80	80	80	80	80	90
10	Walfin	80	80	80	90	80	90

**2.3. Algoritma K-Means Clustering**

Untuk menentukan tingkat kedisiplinan karyawan perhotelan dalam hal ini hotel halay inn ke dalam 3 cluster yaitu sangat baik (C1), Baik(C2), dan Kurang(C3),maka di terapkanlah metode k-means sebagai model pengelompokan data dengan langka langka sebagai berikut[2]:

a). Menentukan jumlah cluster

Dari 10 data karawan akan ditentukan ke dalam 3 cluster, yakni Sangat Baik (C1), Baik (C2), dan Cukup(C3)

b). Menentukan nilai centroid (titik pusat cluster) secara acak.

c). Menghitung jarak antara data atau objek dengan pusat cluster menggunakan teori Euclidian Distance yang dirumuskansebagai berikut:

$$D(i,j) = \sqrt{(X1i-X1j)^2 + (X2i-X2j)^2 + \dots + (Xki-Xkj)^2} \tag{2.1}$$

Dimana

$D(i,j)$  : Jarak data ke  $i$  kepusat cluster  $j$

$Xki$  : Data ke  $i$  pada atribut data ke  $k$

$Xkj$  : Titik pusat ke  $j$  pada atribut data ke  $k$

d). Data ditempatkan dalam *cluster* yang terdekat.

e). Pusat *cluster* baru akan ditentukan bila semua data telah ditetapkan dalam *cluster* terdekat.

Setelah data ditempatkan ke dalam cluster terdekat, maka selanjutnya akan dilakukan perhitungan data yang ditempatkan ke dalam masing masing cluster dengan rumus berikut:  $D = 1/n$  (2.2)

f). Proses penentuan pusat cluster dan penempatan data dalam cluster diulangi sampai nilai centroid tidak berubah lagi.

**3. HASIL DAN PEMBAHASAN**

Data yang digunakan pada penelitian ini adalah data pada tabel 1 yang sudah direkap. Dan selanjutnya akan diproses dengan algoritma *k-means clustering* yaitu menentukan 3 *cluster* untuk menentukan tingkat kedisiplinan karyawan halay inn, yaitu Sangat Baik (C1), Baik (C2), dan Cukup (C3).

**3.1 Menentukan Nilai Centroid (Titik Pusat Cluster) Secara Acak**

Pada iterasi pertama akan ditentukan titik pusat cluster secara ajak sebanyak tiga *cluster* yaitu Sangat Baik (C1), Baik (C2), dan Cukup (C3). Pusat *Cluster* “Centroid” dapat dilihat pada tabel berikut;

Tabel 2. Titik Pusat *Cluster* “Centroid”

Data Ke	Pusat Cluster	Abse nsi	Penampil an Diri	Pelayan an	Sikap & Loyalitas	Tugas & Tanggung Jawab	Kerjasa ma Tim
9	C1	80	80	80	80	80	90
1	C2	70	70	80	90	90	75
4	C3	60	75	60	80	60	60

**3.2 Perhitungan Data Ke Pusat Cluster**

Akan dilakunan perhitungan antara data pertama ke pusat *cluster* “centroid” sesuai pada rumus 2.1 diatas sebagai iterasi pertama. Berikut ini merupakan contoh perhitungan antara data ke-1 pada table 2.1 terhadap pusat *cluster* c1,c2, dan c3 yang ada pada table 2.

$$C1 = \sqrt{(80 - 70)^2 + (80 - 70)^2 + (80 - 80)^2 + (80 - 90)^2 + (80 - 90)^2 + (90 - 75)^2} = 535$$

$$C2 = \sqrt{(70 - 70)^2 + (70 - 70)^2 + (80 - 80)^2 + (90 - 90)^2 + (90 - 90)^2 + (75 - 75)^2} = 0$$

$$C3 = \sqrt{(60 - 70)^2 + (75 - 70)^2 + (60 - 80)^2 + (80 - 90)^2 + (60 - 90)^2 + (65 - 75)^2} = 1660$$

Setelah itu akan dilakukan perhitungan dari data ke-2 ke semua pusat *cluster*, sampai data ke-10.

### 3.3 Menempatkan Data ke Dalam Cluster Terdekat

Setelah semua perhitungan dilakukan dari data ke-1 sampai data ke-10 ke semua pusat *cluster* menggunakan rumus 2.1, maka hasilnya dapat dilihat pada tabel berikut:

Tabel 3. Hasil Perhitungan Data Pada Pusat Cluster Pada Iterasi Pertama

No	Nama Karyawan	Abse nsi	Pen amp ilan Diri	Pela yan an	Sikap & Loyalitas	Tugas & Tanggu ng Jawab	Kerja sama Tim	C1	C2	C3	Jarak Terdek at	Cluster Data
1	Bangbang	70	70	80	90	90	75	535	0	1660	0	Cluster2
2	Jenerius S	60	80	75	80	80	80	145	360	1050	145	Cluster1
3	Hansen S	60	65	70	70	75	80	570	785	925	570	Cluster1
4	Erwin M	60	75	60	80	60	60	1745	1660	0	0	Cluster3
5	Lastrodo	80	80	70	80	80	80	200	435	945	200	Cluster1
6	Horas S	80	70	70	70	80	80	400	635	1045	400	Cluster1
7	Freddy H	80	70	75	70	80	70	625	560	870	560	Cluster2
8	Wanda S	60	60	70	65	70	70	1245	1260	750	750	Cluster3
9	Ronaldo	80	80	80	80	80	90	0	535	1745	0	Cluster1
10	Walfin	80	80	80	90	80	90	100	435	1845	100	Cluster1

Pada tahap perhitungan iterasi pertama ini dapat dilihat bahwa tingak kedisiplinan karyawan pada tabel diatas sebagai berikut: Sangat Baik (C1) sebanyak 6 data yaitu ke 2,3,5,6,9 dan 10), Baik (C2) sebanyak 2 data, yakni data ke1 dan 7, sedangkan Cukup (C3) sebanyak 2 data yakni data ke 4 dan 8.

### 3.4 Menentukan Pusat Cluster Baru

Langka selanjutnya adalah menentukan pusat cluster baru dengan menjumlahkan semuan data setia cluster sesuai rumus 2.2. Pada C1(Sangat Baik) Memiliki 6 data anggota dengan nilai masing masing, maka perhitungannya sebagai berikut:

$$C1 = \frac{60 + 60 + 80 + 80 + 80 + 80}{6} = 73 \text{ Untuk Nilai Absensi}$$

$$C1 = \frac{80 + 65 + 80 + 70 + 80 + 80}{6} = 76 \text{ Untuk Nilai Penampilan diri}$$

Dan seterusnya dilakukan seperti diatas untuk mendapatkan nilai setiap kriteria pada C1,C2 dan C3.

Setelah dilakukan perhitungan kesuma data setiap cluster sesuai rumus 2.2 maka dapat dihasilkan pusat *cluster* “centroid “ baru sebagai iterasi kedua sebagai berikut:

Tabel 4. Pusat Cluster “Centroid” Baru

Pusat cluster Iterasi 2	Absensi	Penampilan Diri	Pelayanan	Sikap & Loyalitas	Tugas & Tanggu ng Jawab	Kerjasama Tim
C1	73	76	74	78	79	83
C2	75	70	78	80	85	73
C3	60	68	65	73	65	65

**3.5 Proses Penentuan Pusat Cluster Sampai Data Tidak Berubah**

Pada bagian ini akan dilakukan perhitungan sesuai rumus 2.1 antara pusat cluster baru iterasi kedua yang ada di tabel 3.3 ke data karyawan pada tabel 2.1. Dimana hasil perhitungannya jika didapatkan berbeda dengan hasil perhitungan pada iterasi pertama yaitu tabel 3, maka akan ditentukan pusat *cluster* baru sesuai rumus 2.2. Namun jika hasil perhitungannya tidak berubah sesuai data pada tabel 3, maka perhitungan menentukan tingkat kedisiplinan karyawan halay inn akan di hentikan. Oleh karena ini akan dilakukan perhitungan sama seperti iterasi satu. Maka hasil perhitungan pusat *cluster* iterasi2 ke data karyawan pada tabel 2.1 sebagai berikut:

Tabel 5. Hasil Perhitungan Data Ke Pusat *Cluster* Baru Iterasi kedua

No	Nama Karyawan	Ab sen si	Penam pilan Diri	Pelay anan	Sikap & Loyalitas	Tugas & Tanggun g Jawab	Kerja sama Tim	c1	c2	c3	Jarak Terde kat	Cluster Data
1	Bangbang	70	70	80	90	90	75	394	143	1273	143	Cluster2
2	Jenerius S	60	80	75	80	80	80	46	203	763	46	Cluster1
3	Hansen S	60	65	70	70	75	80	246	353	363	246	Cluster1
4	Erwin M	60	75	60	80	60	60	1129	1128	188	188	Cluster3
5	Lastrodo	80	80	70	80	80	80	56	243	708	56	Cluster1
6	Horas S	80	70	70	70	80	80	139	243	508	139	Cluster1
7	Freddy H	80	70	75	70	80	70	289	143	383	143	Cluster2
8	Wanda S	60	60	70	65	70	70	721	628	188	188	Cluster3
9	Ronaldo	80	80	80	80	80	90	106	443	1308	106	Cluster1
10	Walfin	80	80	80	90	80	90	239	543	1558	239	Cluster1

Dari hasil perhitungan yang di dapat pada tabel 3.4 sama persis hasilnya dengan tabel 3.2 yakni sebagai berikut:Sangan Baik (C1) sebanyak 6 data yaitu ke 2,3,5,6,9 dan 10), Baik (C2) sebanyak 2 data, yakni data ke1 dan 7, sedangkan Cukup (C3) sebanyak 2 data yakni data ke 4 dan 8. Berarti perhitungan dihentikan pada pusat *cluster* iterasi ke dua.

Maka, hasil penentuan tingkat kedisiplinan karyawan perhotelan dalam ha ini halay inn dengan algoritmma *K-means clustering* adalah sebagai berikut:

Tabel 6. Hasil Penentuan Tingkat Kedisiplinan Karyawan

Cluster 1 (Sangat Baik)	Cluster 2 (Baik)	Cluster 3 (Cukup)
Tingkat kedisiplinan ka-ryawan dengan cluster1 adalah sebagai berikut	Tingkat kedisiplinan ka-ryawan dengan cluster2 adalah sebagai berikut	Tingkat kedisiplinan ka-ryawan dengan cluster3 adalah sebagai berikut
1. Jenerius	1. Bangbang	1. Erwin
2. Hansen	2. Freddy H	2. Wanda
3. Lastro		
4. Horas		
5. Ronaldo		
6. Walfin		

**4. KESIMPULAN**

Hasil Penelitian ini dapat disimpulkan, bahwa dengan menerapkan algoritma *k-means clustering* dapat membantu pihak hotel dalam menyelesaikan permasalahan yang dialami yaitu dengan cara menentukan tingkat kedisiplinan karyaannyan kedalam 3 *cluster*, yakni Sangat baik(C1) dengan hasil sebanyak 6 data, Baik (C2) dengan hasil sebanyak 2 data, dan Kurang(C3) dengan hasil sebanyak 2 data.

Dengan adanya hasil yang didapat ini, pihak hotel bisa terbantu dalam melakukan evaluasi terhadap karyawannya setiap episode.

## UCAPAN TERIMAKASIH

Tim peneliti mengucapkan banyak terimakasih kepada pihak pihak yang mendukung kesuksesan penelitian ini dari tahap awal sampai akhir. Kiranya Tuhan Memberkati.

## REFERENCES

- [1] V. N. Amadhy, *Perancangan Sistem Pengukuran Kinerja Menggunakan Framework Balanced Scorecard Pada Hotel Merdeka Madiun*. repository.its.ac.id, 2020.
- [2] Muhammad Robith Adani, "Mengenal Konsep Data Mining Beserta Contoh Implementasinya," *Sekawanmedia.Co.Id*, 2021. .
- [3] ... Preddy, P. Marpaung, I. Pebrian, and W. Putri, "Penerapan Data Mining Untuk Pengelompokan Kepadatan Penduduk Kabupaten Deli Serdang Menggunakan Algoritma K-Means," *J. Ilmu Komput. dan Sist. Inf.*, vol. 6, no. 2, pp. 64–70, 2023.
- [4] Q. I. Mawarni and E. S. Budi, "Implementasi Algoritma K-Means Clustering Dalam Penilaian Kedisiplinan Siswa," *J. Sist. Komput. dan Inform.*, vol. 3, no. 4, p. 522, 2022.
- [5] P. Marpaung and R. F. Siahaan, "Penerapan Algoritma K-Means Clustering Untuk Pemetaan Kepadatan Penduduk Berdasarkan Jumlah Penduduk Kota Medan," *J-SAKTI (Jurnal Sains Komput. dan Inform.)*, 2021.
- [6] P. Marpaung and H. Pandiangan, "Utilization of the MOORA Method for Recommended Selection of Best Waiters in Hospitality," vol. 4, no. 36, pp. 566–573, 2020.