

Perbandingan Metode K-Means Dan Metode Fuzzy C-Means (FCM) Pada Analisis Kinerja Pegawai PT. Cemara Khatulistiwa Persada Bontang

Comparison of K-Means Method and Fuzzy C-Means (FCM) Method in Employee Performance Analysis of PT. Cemara Khatulistiwa Persada Bontang

Nurul Rakhmawaty¹, Yuki Novia Nasution², dan Fidia Deny Tisna Amijaya³

Laboratorium Matematika Komputasi FMIPA Universitas Mulawarman

E-mail: nurulrakhmawaty21@gmail.com

ABSTRACT

Discipline assessment of employee performance is one of the factors to improve the situation of the quality of human resources. Monitoring and assessment of employee discipline must be carried out continuously as some of the characteristics of management that have gone well as a benchmark for considering the targets that have been set. The K-Means method and the Fuzzy C-Means (FCM) method are non-hierarchical cluster methods. Both methods attempt to partition data into one or more clusters, so that data with the same characteristics are grouped into the same cluster or groups and data with different characteristics are grouped into other groups. This study discusses the comparison of the K-Means method and the Fuzzy C-Means (FCM) method in analyzing employee performance at PT. Cemara Khatulistiwa Persada Bontang, where groups of employees with high, medium, and low levels of employee performance will be determined based on the clustering results of the two methods and determine the best method. The grouping of data for the two methods was obtained from employee attendance data in 2020. Based on the results, it was found that clustering using the K-Means method in the first cluster (low performance level) had 23 employees, the second cluster (medium performance level) had 27 employees, and cluster the third (high performance level) there are 30 employees. Then based on the results of clustering using the FCM method in the first cluster (medium performance level) there are 26 employees, the second cluster (high performance level) there are 31 employees, and the third cluster (low performance level) there are 23 employees. Based on the results of the standard deviation ratio, it was obtained that the K-Means method with a value of 2.4944 was better than the FCM method with a value of 2.7323 in clustering employee performance at PT. Cemara Khatulistiwa Persada.

Keywords: FCM, K-Means, Employee Performance Assessment, Standard Deviation Ratio

Pendahuluan

Analisis kelompok (*cluster analysis*) dalam bidang ilmu pengetahuan digunakan dengan tujuan mengelompokkan sebuah objek yaitu dapat berupa data. Terdapat beberapa metode yang digunakan untuk mengelompokkan data agar menjadi beberapa kelompok data, diantaranya adalah menggunakan cabang dari ilmu matematika, yaitu logika *fuzzy* dan *data mining*. *Data mining* ialah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam daftar data. *Data mining* adalah suatu proses yang memerlukan teknik statistik dan matematika, *machine learning* dan kecerdasan buatan untuk mengekstraksi dan mengidentifikasi informasi bermanfaat serta pengetahuan yang terkait dengan berbagai daftar data besar. *Learning* pada *data mining* terdapat dua jenis yaitu *unsupervised learning* dan *supervised learning*. Adapun metode yang merupakan *unsupervised learning* yaitu *association*, *clustering*, *knowledge discovery*, dan lain-lain. Sedangkan metode yang merupakan *supervised learning* antara lain *classification*,

regression, *summarization*, dan *predictive*. *Data clustering* merupakan metode yang termasuk kedalam *unsupervised learning*. Terdapat dua jenis *data clustering* yang dipergunakan pada proses pengelompokan data, yaitu *hierarchical* (hirarki) *data clustering* dan *non-hierarchical* (non hirarki) *data clustering* (Ramadhan, dkk., 2017).

Metode *K-Means* dan metode *Fuzzy C-Means* (FCM) merupakan metode *cluster non-hierarchical* (non hirarki). Kedua metode tersebut berupaya mempartisi data ke dalam satu ataupun lebih *cluster*, sehingga data yang mempunyai ciri yang sama dikelompokkan ke dalam *cluster* ataupun kelompok yang sama serta data yang memiliki ciri yang berbeda dikelompokkan ke dalam kelompok yang lain. Metode *K-Means* merupakan algoritma yang membagi data ke dalam *k cluster* dengan menetapkan tiap objek ke pusat *cluster* terdekatnya (nilai rata-rata dari variabel untuk seluruh objek dalam *cluster* tertentu) berdasarkan pada dimensi jarak yang digunakan. Metode FCM merupakan teknik *data clustering* dimana keberadaan pada setiap titik data dikelompokkan dalam suatu *cluster* dengan

derajat keanggotaan tertentu. Informasi ini digunakan dalam mendefinisikan fungsi-fungsi keanggotaan guna mempresentasikan nilai *fuzzy* dari masing-masing *cluster* (Yohanes, 2016).

Pada proses *clustering* dalam metode *K-Means*, sebuah objek hanya akan jadi anggota satu *cluster* serta sulit untuk mencapai kekonvergenan. Oleh sebab itu, digunakan metode yang lain yaitu metode FCM. FCM merupakan sebuah metode yang merujuk pada seberapa besar kemungkinan suatu data bisa menjadi anggota ke dalam suatu *cluster*. Algoritma FCM digunakan karena pada metode FCM kemungkinan kegagalan untuk konvergen lebih kecil dibandingkan metode *K-Means* (Agusta, 2007).

Berdasarkan pada penelitian Chusyairi & Saputra (2019) dilakukan perbandingan *data clustering* antara *K-Means* dengan FCM pada pengelompokan data puskesmas dalam pemberian imunisasi untuk balita dan bayi, dimana data tersebut akan dikelompokkan puskesmas yang memberikan pelayanan imunisasi tinggi, pelayanan imunisasi sedang, dan pelayanan imunisasi yang rendah. Berdasarkan pada hasil yang telah dilakukan, didapatkan bahwa algoritma *K-Means* memiliki hasil yang lebih baik dibandingkan algoritma FCM dalam pengelompokan data puskesmas di Banyuwangi. Penelitian selanjutnya telah dilakukan Agustina & Prihandoko (2018) yaitu pada pengelompokan data kinerja karyawan dengan menggunakan variabel masuk, izin, dan terlambat menjadi tiga *cluster* yakni tingkat kinerja tinggi, tingkat kinerja sedang dan tingkat kinerja rendah. Berdasarkan hasil dari penelitian yang telah dilakukan, bahwa metode FCM merupakan metode yang lebih baik dibandingkan dengan metode *K-Means* dalam pengelompokan data kinerja karyawan di STT Bandung karena nilai validasinya bernilai mendekati 1. Selanjutnya yaitu pada penelitian Selviana & Mustakim (2016) pengukuran motivasi belajar yang dikelompokkan menjadi empat kategori yaitu *Attention*, *Relevance*, *Confidence*, *Satisfaction*. Berdasarkan hasil dari penelitian yang telah dilakukan, bahwa metode FCM lebih baik dibandingkan dengan metode *K-Means* dalam pengelompokan data pada mahasiswa Program Studi Sistem Informasi angkatan 2012 dan 2013. Berdasarkan penelitian Agustina & Prihandoko (2018), metode *clustering K-Means* dan FCM digunakan untuk pengelompokan kinerja karyawan. Penilaian kinerja karyawan ialah salah satu tahap evaluasi kerja yang bisa meningkatkan mutu sumber daya manusia. Setiap perusahaan menginginkan karyawan yang memiliki kinerja sesuai dengan standar yang telah ditetapkan perusahaan sebelumnya atau bahkan melebihinya.

Penilaian kinerja karyawan juga akan membuat karyawan termotivasi untuk berkinerja lebih baik lagi dari hari ke hari. Penilaian kedisiplinan kinerja karyawan adalah salah satu faktor untuk memperbaiki situasi kualitas sumber daya manusia. Pemantauan serta penilaian kedisiplinan karyawan wajib dilakukan secara terus menerus menjadi sebagian karakteristik manajemen sudah berjalan dengan baik sebagai tolak ukur pertimbangan target yang telah ditetapkan (Evita, dkk., 2017).

Tujuan yang ingin dicapai dalam penelitian ini adalah untuk mengetahui hasil *clustering* pegawai di PT. Cemara Khatulistiwa Persada Tahun 2020 dengan menggunakan metode *K-Means* dan metode FCM. Menentukan metode terbaik untuk *clustering* pegawai PT. Cemara Khatulistiwa Persada dengan menggunakan metode *K-Means* dan metode FCM berdasarkan nilai s_w dan s_b .

Data Mining

Data mining ialah suatu metode pengolahan data guna menemukan pola yang terkandung dari data tersebut. *Data mining* merupakan pengolahan data dengan skala besar, dengan demikian *data mining* mempunyai peranan penting dalam bidang industri, cuaca, keuangan, ilmu dan teknologi (Santosa, 2007). *Data mining* juga bisa dilakukan pada berbagai macam tipe *database* serta penyimpanan informasi, tetapi jenis pola yang hendak ditemui ditetapkan oleh bermacam fungsi *data mining* seperti fungsi deskripsi, fungsi estimasi, fungsi prediksi, fungsi klasifikasi, fungsi pengelompokkan dan fungsi asosiasi (Susanto & Suryadi, 2010).

Clustering

Clustering ialah pengelompokan *item* data ke dalam beberapa kecil grup sedemikian sehingga tiap-tiap grup memiliki suatu persamaan yang esensial (Garcia-Molina, dkk., 2002). *Clustering* mengacu pada pengelompokan observasi, catatan, ataupun permasalahan ke dalam kelas yang seragam. Suatu *cluster* merupakan kumpulan dari catatan yang mirip satu sama lain serta berbeda dengan catatan pada *cluster* lain. *Clustering* berbeda dengan *classification* yang tidak terdapat variabel sasaran untuk *clustering*. Algoritma *clustering* mencari ke segmen data seluruh kelompok menjadi sub kelompok yang relatif homogen ataupun kelompok, dimana kesamaan catatan dalam *cluster* dioptimalkan, serta kesamaan catatan luar *cluster* ini diminimalkan (Larose, 2005).

Algoritma Min-Max

Algoritma Min-Max ialah algoritma yang digunakan dalam normalisasi data dengan cara

membandingkan serta memperkecil rentang pada data. Algoritma Min-Max ialah metode sederhana dimana metode tersebut bisa secara khusus cocok dengan data dalam batasan yang ditetapkan sebelumnya. Algoritma ini dirumuskan sebagai berikut (Suyanto, 2018):

$$\hat{x}_{ik} = \frac{x_{ik} - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Keterangan :

\hat{x}_{ik} = nilai baru hasil normalisasi untuk data ke- i atribut/dimensi ke- k

x_{ik} = data ke- i atribut/dimensi ke- k , $k = 1, \dots, dim$

x_{min} = data minimum atau data yang paling kecil

x_{max} = data maksimum atau data yang paling besar.

Metode K-Means

Tahapan *clustering* dengan menggunakan algoritma *K-Means* adalah sebagai berikut:

1. Menentukan banyak *cluster* (m), ambang batas fungsi obyektif (T), fungsi obyektif awal (F^0), dan pusat *cluster* (c) secara acak yang dinamakan c_1, c_2, \dots, c_n
2. Menghitung jarak tiap data dengan masing-masing pusat *cluster* menggunakan rumus Jarak Euclidean:

$$D(x_i, c_j) = \sqrt{\sum_{k=1}^{dim} (x_{ik} - c_{jk})^2} \quad (2)$$

Keterangan:

$D(x_i, c_j)$ = Jarak Euclid data ke- i dengan pusat *cluster* ke- j

x_{ik} = data ke- i atribut/dimensi ke- k , $k = 1, \dots, dim$

c_{jk} = pusat *cluster* ke- j atribut/dimensi ke- k , $k = 1, \dots, dim$

Kemudian berdasarkan hasil jarak tiap data dengan masing-masing pusat *cluster*, setiap data dialokasikan ke *cluster* yang mempunyai pusat *cluster* dengan jarak terdekat dari data tersebut.

3. Memperbarui tiap pusat *cluster* dengan menggunakan rumus rata-rata sebagai berikut:

$$c_{jk}^t = \sum_{i=1, x_i \in c_j}^n \frac{x_{ik}}{n_j} \quad (3)$$

Keterangan:

c_{jk}^t = pusat *cluster* baru ke- j atribut/dimensi ke- k pada iterasi ke- t , $k = 1, \dots, dim$

$i, x_i \in c_j$ = indeks i yang termasuk ke dalam *cluster* ke- j

n = banyak data

x_{ik} = data ke- i atribut/dimensi ke- k , $k = 1, \dots, dim$

n_j = banyak data pada pusat *cluster* ke- j

4. Menghitung fungsi obyektif yang digunakan untuk metode *K-Means* ditentukan berdasarkan jarak masing-masing data dengan tiap *cluster* baru sesuai dengan keanggotaan data dalam kelompok, fungsi obyektif yang digunakan adalah sebagai berikut :

$$F = \sum_{i=1}^n \sum_{j=1}^m a_{ij} D^2(x_i, c_j) \quad (4)$$

n adalah banyak data dan m adalah banyak *cluster*. a_{ij} adalah nilai keanggotaan data ke- i dengan pusat *cluster* ke- j . a mempunyai nilai 0 atau 1. Apabila data merupakan anggota suatu kelompok maka nilai $a_{ij} = 1$ dan jika sebaliknya $a_{ij} = 0$. Selanjutnya hasil perhitungan tersebut dijumlahkan keseluruhan dan dihitung selisih fungsi obyektif = $|F \text{ baru} - F \text{ lama}|$. Nilai F lama ialah fungsi obyektif awal (F^0) yang telah ditentukan sebelumnya dan F baru ialah hasil perhitungan yang telah didapatkan pada langkah ke empat.

5. Membandingkan selisih fungsi obyektif dengan nilai T yaitu apabila selisih fungsi obyektif $< T$ maka hentikan perulangan, namun jika selisih fungsi obyektif $> T$ maka ulangi langkah 2 sampai dengan 5 (Prasetyo, 2012).

Metode Fuzzy C-Means

Tahapan *clustering* dengan menggunakan metode *Fuzzy C-Means* yaitu sebagai berikut:

1. Memasukkan data berupa matriks \mathbf{X} berukuran $n \times dim$, dengan n adalah banyaknya jumlah sampel data dan dim merupakan banyak atribut setiap data.

Adapun bentuk matriksnya sebagai berikut:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \dots & x_{1dim} \\ x_{21} & x_{22} \dots & x_{2dim} \\ \vdots & \vdots \ddots & \vdots \\ x_{n1} & x_{n2} \dots & x_{ndim} \end{bmatrix} \quad (5)$$

x_{ik} = data sampel ke- i ($i = 1, 2, \dots, n$), atribut ke- k ($k = 1, 2, \dots, dim$).

2. Menentukan *initial value* berikut:

Jumlah *cluster* (m)

Bobot dalam FCM (w)

Maksimum iterasi (Maxiter)

Error terkecil yang diharapkan (ϵ)

Fungsi obyektif awal ($F^0 = 0$)

Iterasi awal ($t = 1$)

3. Membangkitkan bilangan *random* $(\mu_{ji})_{m \times n}$, $j = 1, 2, \dots, m$ dan $i = 1, 2, \dots, n$;

berukuran $m \times n$ sebagai elemen-elemen matriks partisi awal \mathbf{U} . Dimana μ_{ji} merupakan bilangan *random* dari 0 sampai dengan 1 yang menyatakan

derajat keanggotaan dari setiap data. Berikut merupakan bentuk matriks partisi awal \mathbf{U} :

$$\mathbf{U} = \begin{bmatrix} \mu_{11}(x_1) & \mu_{12}(x_2) & \dots & \mu_{1n}(x_n) \\ \mu_{21}(x_1) & \mu_{22}(x_2) & \dots & \mu_{2n}(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{m1}(x_1) & \mu_{m2}(x_2) & \dots & \mu_{mn}(x_n) \end{bmatrix} \quad (6)$$

Selanjutnya, pada matriks partisi awal \mathbf{U} dinormalisasi menggunakan rumus:

$$Q_i = \sum_{j=1}^m \mu_{ji} \text{ dengan } i = 1, 2, \dots, n \quad (7)$$

Kemudian menghitung elemen matriks yang ternormalisasi agar jumlah setiap kolom matriks sama dengan atau mendekati satu dengan rumus:

$$\mu_{ji}^{norm} = \frac{\mu_{ji}}{Q_i} \quad (8)$$

Berikut merupakan bentuk matriks partisi awal yang telah ternormalisasi :

$$\mathbf{U}^0 = \begin{bmatrix} \mu_{11}^{norm}(x_1) & \mu_{12}^{norm}(x_2) & \dots & \mu_{1n}^{norm}(x_n) \\ \mu_{21}^{norm}(x_1) & \mu_{22}^{norm}(x_2) & \dots & \mu_{2n}^{norm}(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{m1}^{norm}(x_1) & \mu_{m2}^{norm}(x_2) & \dots & \mu_{mn}^{norm}(x_n) \end{bmatrix} \quad (9)$$

Jumlah elemen-elemen pada setiap kolom matriks partisi awal \mathbf{U}^0 harus mendekati atau sama dengan satu.

4. Menghitung pusat *cluster* ke- j ; V_{jk} dengan $j = 1, 2, \dots, m$ dan $k = 1, 2, \dots, dim$ dalam bentuk matriks berukuran $m \times dim$ menggunakan persamaan berikut:

Persamaan untuk nilai awal pusat *cluster*:

$$V_{jk} = \frac{\sum_{i=1}^n ((\mu_{ji}^{norm})^w \cdot x_{ik})}{\sum_{i=1}^n (\mu_{ji}^{norm})^w} \quad (10)$$

Kemudian hasil perhitungan V_{jk} ditulis dalam bentuk matriks berukuran $m \times dim$ sebagai berikut:

$$\mathbf{V}^t = \begin{bmatrix} V_{11} & V_{12} & \dots & V_{1dim} \\ V_{21} & V_{22} & \dots & V_{2dim} \\ \vdots & \vdots & \ddots & \vdots \\ V_{m1} & V_{m2} & \dots & V_{mdim} \end{bmatrix} \quad (11)$$

5. Menghitung fungsi obyektif pada iterasi ke- t , F^t menggunakan rumus sebagai berikut :

$$F^t = \sum_{i=1}^n \sum_{j=1}^m \left(\left[\sum_{k=1}^{dim} (x_{ik} - V_{jk}^t)^2 \right] (\mu_{ji}^{norm})^w \right); \text{ untuk } t = 1 \quad (12)$$

6. Memperbaiki derajat keanggotaan setiap data dalam setiap *cluster* dengan menghitung perubahan matriks partisi:

$$\mu_{ji}^t = \frac{\left[\sum_{k=1}^{dim} (x_{ik} - V_{jk}^t)^2 \right]^{-\frac{1}{w-1}}}{\sum_{j=1}^m \left[\sum_{k=1}^{dim} (x_{ik} - V_{jk}^t)^2 \right]^{-\frac{1}{w-1}}} \quad (13)$$

Derajat keanggotaan yang telah diperbaiki ini merupakan elemen matriks \mathbf{U}^t berukuran $m \times n$. Berikut merupakan bentuk matriksnya:

$$\mathbf{U}^t = \begin{bmatrix} \mu_{11}^t(x_1) & \mu_{12}^t(x_2) & \dots & \mu_{1n}^t(x_n) \\ \mu_{21}^t(x_1) & \mu_{22}^t(x_2) & \dots & \mu_{2n}^t(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{m1}^t(x_1) & \mu_{m2}^t(x_2) & \dots & \mu_{mn}^t(x_n) \end{bmatrix} \quad (14)$$

7. Mengecek kondisi iterasi berhenti

Jika $(|F^t - F^{t-1}| < \epsilon)$ ataupun $(t > MaxIter)$ maka iterasi dihentikan; Jika tidak, maka iterasi dinaikkan $t = t + 1$ dengan mengulangi langkah ke-4 untuk mengawali iterasi baru (Handayanto & Herlawati, 2020).

Ukuran Pemilihan Metode Terbaik

Suatu metode yang digunakan dalam membentuk sebuah *cluster* dapat dikatakan memiliki kinerja baik jika nilai rasio simpangan baku minimum s_w terhadap s_b . Adapun rumus (s_w) sebagai berikut (Santoso, 2007):

$$s_w = \frac{1}{m} \sum_{j=1}^m s_j \quad (15)$$

Keterangan:

s_w = nilai simpangan baku dalam *cluster*

s_j = simpangan baku *cluster* ke- j

m = banyak *cluster*

Jika diberikan *cluster* ke- j , dimana $j = 1, \dots, m$ dan setiap *cluster* mempunyai anggota x_i dimana $i = 1, \dots, n$ dan variabel k dimana $k = 1, 2, \dots, dim$ serta n adalah jumlah anggota dari setiap variabel, dan $\bar{x}_{j,k}$ adalah rata-rata dari *cluster* ke- j variabel ke- k sehingga untuk mencari nilai simpangan baku *cluster* ke- j (s_j) dengan mencari nilai simpangan baku *cluster* ke- j variabel ke- k ($s_{j,k}$) terlebih dahulu digunakan rumus sebagai berikut:

$$s_{j,k} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{i,k} - \bar{x}_{j,k})^2} \quad (16)$$

Berdasarkan nilai $s_{j,k}$ maka didapatkan nilai simpangan baku dalam *cluster* ke- j (s_j) yaitu dengan menggunakan rumus sebagai berikut:

$$s_j = \frac{1}{dim} \sum_{k=1}^{dim} s_{j,k} \quad (17)$$

Dalam menentukan nilai rata-rata *cluster* ke- j (\bar{x}_j) dirumuskan sebagai berikut:

$$\bar{x}_j = \frac{1}{dim} \sum_{k=1}^{dim} \bar{x}_{j,k} \quad (18)$$

Apabila terdapat rata-rata variabel dalam setiap cluster ke- $j(\bar{x}_j)$ maka komponen dari setiap cluster akan berbeda dan simpangan baku antar cluster (s_b) dirumuskan sebagai berikut:

$$s_b = \sqrt{\frac{1}{m-1} \sum_{j=1}^m (\bar{x}_j - \bar{x})^2} \quad (19)$$

Keterangan:

s_b = nilai simpangan baku antar cluster

\bar{x}_j = Rataan cluster ke- j

\bar{x} = Rataan keseluruhan cluster

m = banyak cluster

Cluster dikatakan baik apabila menghasilkan nilai rasio simpangan baku minimum s_w terhadap s_b dengan rumus sebagai berikut:

$$s = \frac{s_w}{s_b} \times 100\% \quad (20)$$

PT. Cemara Khatulistiwa Persada

PT. Cemara Khatulistiwa Persada berkedudukan di Jl. Sutan Syahrir No. 3 Bontang. Perusahaan

ini bergerak dalam bidang perdagangan besar alat tulis dan gambar, perdagangan besar barang percetakan dan penerbitan dalam berbagai bentuk, perdagangan besar komputer dan perlengkapan komputer, perdagangan besar berbagai macam material bangunan, aktivitas penyediaan tenaga kerja waktu tertentu, aktivitas kebersihan umum bangunan, dan aktivitas perawatan dan pemeliharaan taman. PT. Cemara Khatulistiwa Persada perlu menyeimbangkan kebutuhan perusahaan dengan kemajuan teknologi untuk meningkatkan produktifitas perusahaan.

Hasil dan Pembahasan

Analisis Statistika Deskriptif

Analisis statistika deskriptif pada data presensi pegawai berdasarkan 5 variabel pada 80 pegawai yang ada di PT. Cemara Khatulistiwa Persada pada tahun 2020 menggunakan nilai minimum, nilai maksimum, rata-rata, dan standar deviasi. Adapun hasil analisis statistika deskriptif dapat dilihat pada Tabel 1.

Tabel 1 Statistika Deskriptif

Variabel	Masuk	Izin	Terlambat	Sakit	Lembur
Minimum	1561	24	8	32	384
Maksimum	1985	240	55	232	696
Rata-Rata	1.759,8375	128,9	32,6625	142,6	522,2625
Standar Deviasi	87,53292	63,35106	13,14471	55,53915	97,88572

Tabel 2 Data Normalisasi

No	Nomor Badge	Masuk	Izin	Terlambat	Sakit	Lembur
1	300741	0,61792453	0,37037037	0,19148936	0,44	0,83974359
2	300726	0,4740566	0,85185185	0,63829787	0,12	0,6474359
3	300729	0,375	0,40740741	0,5106383	0,84	0,94871795
⋮	⋮	⋮	⋮	⋮	⋮	⋮
80	301593	0,12264151	0,92592593	0,57446809	0,8	0,10897436

Tabel 3 Pusat Cluster

Variabel	Pusat Cluster		
	1	2	3
Masuk	0	0,52122642	1
Izin	1	0,07407407	0,03703704
Terlambat	1	0,21276596	0,14893617
Sakit	0,88	0,96	0
Lembur	0,79487179	0,08333333	1

Tabel 4 Jarak Data Dengan Masing- Masing Pusat Cluster

Data ke-I	Jarak ke Centroid (Pusat)			Termasuk ke dalam Cluster
	1	2	3	
1	1,275761	0,96961	0,69151	3
2	0,988354	1,34612	1,14835	1
3	0,86979	0,99223	1,16913	1
⋮	⋮	⋮	⋮	⋮
80	0,823687	1,02059	1,78182	1

Tabel 5 Pusat Cluster Baru

Variabel	Pusat Cluster		
	1	2	3
Masuk	0,2893082	0,4981656	0,6924251
Izin	0,7805213	0,3395062	0,3267974
Terlambat	0,785658	0,3829787	0,4105131
Sakit	0,554074	0,682222	0,277647
Lembur	0,5172127	0,2374466	0,7611237

Normalisasi Data

Adapun hasil normalisasi data menggunakan Persamaan (1) dapat dilihat pada Tabel 2.

Metode Clustering K-Means

Berikut merupakan langkah-langkah menggunakan metode *K-Means*:

Penentuan Nilai-Nilai Parameter Metode K-Means

Banyaknya cluster ($m = 3$),
Ambang batas fungsi obyektif ($T = 0.1$),
Fungsi obyektif awal ($F^0 = 0$),
Pusat Cluster dapat dilihat pada Tabel 3.

Perhitungan Jarak Tiap Data Dengan Masing-Masing Pusat Cluster

Adapun hasil jarak tiap data dengan masing-masing pusat cluster menggunakan Persamaan (2) dapat dilihat pada Tabel 4. Suatu data termasuk kedalam cluster ke- i jika salah satu hasil jarak data dengan pusat cluster lebih minimum.

Pembaharuan Pusat Cluster

Adapun hasil pusat cluster baru menggunakan Persamaan (3) dapat dilihat pada Tabel 5.

Perhitungan Fungsi Obyektif

Adapun hasil fungsi obyektif menggunakan Persamaan (4) pada iterasi pertama didapatkan selisih fungsi obyektif = $|F \text{ baru} - F \text{ lama}| = 16,8496 - 0 = 16,8496$.

Perbandingan Selisih Fungsi Dengan Nilai

T

Pada iterasi pertama didapatkan selisih fungsi obyektif $T: 16,8496 > 0,1$, maka ulangi langkah 2 sampai dengan 5. Pada penelitian ini diperoleh hasil akhir clustering dari data presensi pegawai dengan menggunakan *Software Python* dapat dilihat pada Tabel 6.

Tabel 6 Hasil Cluster K-Means Clustering

Cluster 1	Cluster 2	Cluster 3
300726	300739	300741
300729	3007742	300728
⋮	⋮	⋮
301579	301593	301588

Metode Clustering Fuzzy C-Means

Adapun langkah-langkah menggunakan metode *Fuzzy C-Means* sebagai berikut:

Penginputan Data

Adapun bentuk data dalam matriks dapat dilihat sebagai berikut:

X

$$= \begin{bmatrix} 0,61792453 & 0,37037037 & 0,19148936 & 0,44 & 0,83974359 \\ 0,4740566 & 0,85185185 & 0,63829787 & 0,12 & 0,6474359 \\ 0,375 & 0,40740741 & 0,5106383 & 0,84 & 0,94871795 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0,12264151 & 0,92592593 & 0,57446809 & 0,8 & 0,10897436 \end{bmatrix}$$

Penentuan Initial Value

Jumlah cluster ($m = 3$)
Bobot dalam FCM ($w = 2$)
Maksimum iterasi (Maxiter) = 1000
Error terkecil yang diharapkan ($\epsilon = 10^{-5}$)
Fungsi obyektif awal ($F^0 = 0$)
Iterasi awal ($t = 1$)

Pembangkitan Bilangan Random

Adapun bentuk matriks partisi awal **U** dapat dilihat sebagai berikut:

$$U_{3,80} = \begin{bmatrix} 0,43 & 0,36 & 0,51 & 0,24 & 0,09 & 0,26 & 0,82 & \dots & 0,36 \\ 0,28 & 0,22 & 0,43 & 0,64 & 0,78 & 0,32 & 0,25 & \dots & 0,52 \\ 0,47 & 0,58 & 0,26 & 0,39 & 0,23 & 0,67 & 0,06 & \dots & 0,41 \end{bmatrix}$$

Matriks partisi awal *random* dinormalisasi menggunakan Persamaan (7) dan (8), sehingga diperoleh hasil matriks partisi awal berukuran 3×80 yang telah ternormalisasi sebagai berikut:

$$U^0 = \begin{bmatrix} 0,36 & 0,31 & 0,42 & 0,18 & 0,08 & 0,20 & 0,72 & \dots & 0,27 \\ 0,23 & 0,18 & 0,35 & 0,50 & 0,70 & 0,25 & 0,22 & \dots & 0,40 \\ 0,39 & 0,5 & 0,21 & 0,30 & 0,20 & 0,53 & 0,05 & \dots & 0,31 \end{bmatrix}$$

Perhitungan Pusat Cluster

Adapun hasil matriks nilai awal pusat *cluster* berukuran 3×5 menggunakan Persamaan (10) sebagai berikut:

$$V^1 = \begin{bmatrix} 0,465048292 & 0,488232328 & \dots & 0,450017896 \\ 0,527524012 & 0,432206786 & \dots & 0,41253724 \\ 0,439146197 & 0,495497674 & \dots & 0,459588976 \end{bmatrix}$$

Perhitungan Fungsi Obyektif

Adapun hasil fungsi obyektif menggunakan Persamaan (12) pada iterasi pertama didapatkan 15,3739367.

Perbaikan Derajat Keanggotaan

Adapun hasil matriks derajat keanggotaan yang telah diperbaiki berukuran 3×80 menggunakan Persamaan (13) sebagai berikut:

$$U^1 = \begin{bmatrix} 2,926362759 & 3,180308669 & \dots & 3,16034234 \\ 3,445614737 & 2,890356635 & \dots & 2,555656011 \\ 2,716986026 & 2,944753085 & \dots & 3,421264195 \end{bmatrix}$$

Pengecekan Kondisi Iterasi Berhenti

Pada kondisi ini $t = 1 < 1000$ dan $|F^t - F^{t-1}| = |F^1 - F^0| = |15,3739367 - 0| = 15,3739367 > 10^{-5}$ yang artinya tidak memenuhi kedua kondisi pemberhentian, sehingga iterasi harus dilanjutkan dan mengulangi langkah ke-4. Pada penelitian ini diperoleh hasil akhir *clustering* dari data presensi pegawai dengan menggunakan *Software Python* dapat dilihat pada Tabel 7.

Tabel 7 Hasil Cluster Fuzzy C-Means Clustering

Cluster 1	Cluster 2	Cluster 3
300726	300741	300739
300732	300729	3007742
⋮	⋮	⋮
301579	301588	301593

Pemilihan Metode Terbaik Dengan Simpangan Baku

Pemilihan metode terbaik dari metode *K-Means* dan metode FCM berdasarkan nilai rasio minimum (s_w) terhadap (s_b) untuk metode yang baik dapat dilihat dari nilai rasio yang minimum, adapun perhitungannya dengan menggunakan Persamaan (15) sampai dengan Persamaan (20) didapatkan hasil yang dapat dilihat pada Tabel 8.

Tabel 8 nilai rasio simpangan baku (s)

Metode	Nilai (s)
<i>K-Means</i>	2,4944
<i>Fuzzy C-Means</i>	2,7323

Metode Terbaik

Berdasarkan nilai rasio simpangan baku (s) metode *K-Means* dan metode *Fuzzy C-Means* pada Tabel 8. Diperoleh bahwa metode terbaik pada *clustering* pegawai menggunakan data presensi pegawai adalah metode *K-Means*. Hal ini dapat dilihat dari nilai rasio simpangan baku s_w terhadap s_b bahwa pada metode *K-Means* hasil yang didapatkan paling minimum yaitu dengan nilai 2,4944.

Kesimpulan

Berdasarkan hasil analisis data dan pembahasan yang telah dilakukan, diperoleh kesimpulan sebagai berikut:

1. Dari hasil analisis *clustering* menggunakan metode *K-Means* bahwa didapatkan hasil pada *cluster* pertama (tingkat kinerja rendah) terdapat 23 pegawai, *cluster* kedua (tingkat kinerja sedang) terdapat 27 pegawai, dan *cluster* ketiga (tingkat kinerja tinggi) terdapat 30 pegawai. Kemudian berdasarkan hasil *clustering* menggunakan metode *Fuzzy C-Means* didapatkan hasil pada *cluster* pertama (tingkat kinerja sedang) terdapat 26 pegawai, *cluster* kedua (tingkat kinerja tinggi) terdapat 31 pegawai, dan *cluster* ketiga (tingkat kinerja rendah) terdapat 23 pegawai. Pada hasil *clustering* pegawai di PT. Cemara Khatulistiwa Persada, sehingga PT. Cemara Khatulistiwa Persada belum bisa mencapai tujuan untuk meningkatkan produktifitas perusahaan.
2. Berdasarkan hasil evaluasi dengan menggunakan nilai s_w dan s_b , metode terbaik dalam *clustering* pegawai PT. Cemara Khatulistiwa Persada menggunakan data presensi pegawai yaitu metode *K-Means*. Dalam metode *K-Means* nilai rasio simpangan baku (s) didapatkan paling minimum dengan nilai 2,4944.

Daftar Pustaka

Agusta, Y. (2007). K-Means – Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem Dan Informatika*, 3(Februari), 47–60.

Agustina, N., & Prihandoko. (2018). Perbandingan Algoritma K-Means dengan Fuzzy C-Means Untuk *Clustering* Tingkat Kedisiplinan Kinerja Karyawan. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 2(3), 621–626. <https://doi.org/10.29207/resti.v2i3.492>

- Bezdek, J. C., & Dunn, J. C. (1975). *Optimal Fuzzy Partitions: A Heuristic for Estimating the Parameters in a Mixture of Normal Distributions*. IEEE Transactions on Computers, hal 835-838.
- Evita, S. N., Muizu, W. O. Z., & Atmojo, R. T. W. (2017). Penilaian Kinerja Karyawan Dengan Menggunakan Metode Behaviorally Anchor Rating Scale dan Management By Objectives (Studi kasus pada PT Qwords Company International). *Pekbis Jurnal*, 9(1), 18–32.
- Garcia-Molina, H., Ullman, J. D., & Widom, J. (2002). *DATABASE SYSTEMS The Complete Book*. New Jersey, Prentice Hall.
- Han, J., Kamber, M., & Pei, J. (2011). Data mining Concepts and Techniques. In *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. Morgan Kaufmann. <https://doi.org/10.1109/ICMIRA.2013.45>
- Handayanto, R. T., & Herlawati. (2020). *Data Mining dan Machine Learning Menggunakan Matlab dan Python*. Informatika Bandung.
- Hidayat, T., & Istiadah, N. (2011). *Panduan Lengkap SPSS 19 untuk Mengolah Data Statistik Penelitian*. Jakarta: Media Kita.
- Kusumadewi, S., & Purnomo, H. (2004). *Aplikasi Logika Fuzzy Untuk Pendukung Keputusan Edisi Pertama*. Yogyakarta: Graha Ilmu.
- Kusumadewi, S., & Purnomo, H. (2010). *Aplikasi Logika Fuzzy Untuk Pendukung Keputusan Edisi 2*. Yogyakarta: Graha Ilmu.
- Larose, D. T. (2005). *Discovering knowledge in data an Introduction to Data Mining*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118874059>
- Prasetyo, E. (2012). *Data Mining : Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi.
- Ramadhan, A., Efendi, Z., & Mustakim. (2017). Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling. *Jurnal Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI) 9*, 219–226.
- Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Santoso, S. (2007). *Menguasai Statistik di Era Informasi Dengan SPSS 15*. Jakarta: PT Elex Media Komputindo Gramedia.
- Santoso, S. (2010). *Statistik Multivariat*. Jakarta: Elex Media Komputindo.
- Susanto, S., & Suryadi, D. (2010). *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta: C.V. Andi Offset.
- Suyanto. (2018). *Machine Learning Tingkat Dasar dan Lanjut*. Bandung: Informatika Bandung.
- Suyanto. (2019). *Data Mining Untuk Klasifikasi dan Klasterisasi Data Edisi Revisi*. Informatika Bandung.
- Yohanes. (2016). Analisis Perbandingan Algoritma Fuzzy C-Means dan K-Means. *Jurnal Annual Research Seminar 2016*, 2(1), 151–155.

