

# Pengelompokkan Karyawan Berdasarkan Kesalehan Menggunakan Perbandingan *Fuzzy C-Means*, *K-Means*, dan *Probabilistic Distance Clustering*

Aldi Rizal<sup>1</sup>, Dian Candra Rini Novitasari<sup>2</sup>, Moh. Hafiyusholeh<sup>3</sup>

<sup>123</sup>Program Studi Matematika, Fakultas Sains dan Teknologi, UIN Sunan Ampel Surabaya, Jl. Ahmad Yani No.117 Surabaya, Jawa Timur, Indonesia

Korespondensi; Aldi Rizal, Email: [aldirizal480@gmail.com](mailto:aldirizal480@gmail.com)

## Abstrak

Kesalehan dapat dijadikan tolak ukur untuk menentukan karyawan tersebut baik atau tidaknya. Tingkatan kesalehan karyawan tersebut dapat dikelompokkan menjadi beberapa kelompok dengan mendefinisikan tiap-tiap kelompoknya. Dengan begitu, tujuan dari penelitian ini ialah untuk mengelompokkan karyawan dengan berdasar pada tingkat kesalehannya. Pengelompokan kesalehan karyawan tersebut menggunakan metode *clustering* di antaranya ialah *K-Means*, *Fuzzy C-Means*, dan *Probabilistic Distance Clustering* dengan uji evaluasi nilai *cluster* menggunakan *silhouette coefficient* sebagai pembandingnya. Data yang digunakan sebanyak 500 data karyawan yang diambil di lingkungan bank. Hasil perbandingan menghasilkan nilai terbaik pada nilai *silhouette coefficient* dengan jumlah nilai  $k=2$  dengan rata-rata masing-masing metode sebesar 0,5134. Dengan mendefinisikan *cluster* 1 mewakili *cluster* karyawan yang tidak saleh sementara *cluster* 2 mewakili *cluster* karyawan saleh.

**Kata Kunci:** *Clustering*, Perbandingan *Clustering*, Kesalehan, Karyawan

## Abstract

Piety can be used as a benchmark to determine whether the employee is good or not. The level of employee piety can be grouped into several groups by defining each group. Therefore, the purpose of this research is to group employees based on their level of piety. The employee piety grouping uses clustering methods including K-Means, Fuzzy C-Means, and Probabilistic Distance Clustering with cluster value evaluation test using silhouette coefficient as a comparison. The data used is 500 employee data taken in a bank environment. The comparison results produce the best value on the silhouette coefficient value with the number of  $k=2$  values with an average of 0.5134 for each method. By defining cluster 1 represents a cluster of non-piety employees while cluster 2 represents a cluster of piety employees.

**Keywords:** Clustering, Clustering Comparison, Piety, Employee

---

## Pendahuluan

Perusahaan memiliki karyawan sebagai salah satu aset yang berharga. Karyawan juga merupakan sumber daya yang sangat penting bagi kelangsungan operasional perusahaan [1]. Perusahaan akan mendapatkan keuntungan jika memiliki karyawan yang kreatif, produktif, dan terampil [2]. Karyawan yang memiliki kriteria yang baik tersebut dapat dipertahankan oleh perusahaan agar dapat memberikan efek yang baik bagi suatu perusahaan [3]. Pada berbagai sektor industri, terutama di sektor perbankan, karyawan yang memiliki dedikasi tinggi dalam bekerja sangat diinginkan oleh perusahaan-perusahaan bank. Industri perbankan memiliki peran yang krusial dan berfungsi sebagai penopang dalam mendorong kemajuan ekonomi suatu negara. [4]. Untuk memenuhi peran penting tersebut sumber daya yang unggul tentu dibutuhkan oleh suatu perusahaan perbankan [5].

Sumber daya yang unggul, terutama pada sumber daya manusia dapat dilihat dari tingkat kesalehan karyawan tersebut [6]. Tolak ukur yang digunakan dapat berupa tingkat kesalehan seorang karyawan dalam peningkatan kinerja seorang karyawan, penyelesaian masalah, serta pengambilan keputusan dalam dunia kerja [7]. Dalam pandangan Junaedi (2019), seseorang dapat dinyatakan sebagai individu yang saleh dengan mempertimbangkan beberapa faktor, seperti memiliki keimanan kepada Allah, ketakutan terhadap Allah dan Rasul-Nya, berkontribusi dalam mendukung agama Islam, melakukan amal kebaikan, berjuang dalam perjuangan di jalan Allah, menjadikan akhirat sebagai tujuan utama, memiliki kesabaran dalam menghadapi cobaan dari Allah, dan taat dalam menjalankan perintah-perintah Allah, serta tidak berbuat keburukan atau kerusakan di bumi [8]. Karyawan yang saleh setidaknya mempunyai beberapa sifat dari kesembilan sifat tersebut. Dari kesembilan sifat tersebut, hal yang paling sering disebutkan adalah dengan memperhatikan ketakwaannya kepada Allah. Hal tersebut telah dijelaskan pada hadits berikut:

اَتَقْ اَنَّ اللَّهَ حِيْثُمَا كُنْتَ وَأَتَيْعُ السَّيِّئَةَ الْحَسَنَةَ تَمْحُهَا وَخَالِقُ النَّاسَ بِخُلُقٍ حَسَنٍ

artinya: “Bertakwalah kepada Allah di mana saja engkau berada. Ikutilah kejelekan dengan kebaikan niscaya ia akan menghapuskan kejelekan tersebut dan berakhhlaklah dengan manusia dengan akhlak yang baik.” [HR. Tirmidzi no. 1987 dan Ahmad 5/153]. Hadits tersebut menerangkan bahwa manusia hendaknya berbuat baik di mana pun tempatnya. Penerapan prinsip ini dapat mencakup lingkungan kerja, di mana perilaku yang baik dapat berdampak positif terhadap kinerja seorang karyawan. Di sisi lain, perilaku yang buruk akan memiliki dampak negatif terhadap kinerjanya [9].

Kesalehan dapat diamati melalui beberapa faktor yang mempengaruhinya, yang terdiri dari dua komponen, yakni *Islamic Spirituality* (IS) dan *Islamic Social Responsibility* (ISR). Skala pengukuran terdiri dari 18 jenis untuk komponen IS dan 35 jenis untuk komponen ISR dapat dilihat dari beberapa faktor yang mempengaruhi [10]. Komponen IS mempunyai sub komponen yaitu di antaranya keyakinan, ibadah, pertaubatan, dan mengingat Allah. Komponen ISR juga mempunyai sub komponen yaitu di antaranya amal saleh, sabar, integritas, pemenuhan janji, cinta keluarga, pengampunan, keadilan, pengendalian emosi, kebenaran, dan menjaga kesucian [11]. Dari beberapa sub komponen tersebut akan ditentukan karyawan tersebut akan masuk pada kategori saleh atau tidak. Untuk memperoleh informasi tersebut, diperlukan proses pengelompokan, yang biasanya disebut *clustering*. Melalui *clustering* atau pengelompokan karyawan, hasil evaluasi karyawan dapat digunakan dan menjadi pertimbangan dalam merekrut karyawan baru.

*Clustering* mempunyai beberapa macam metode yang bisa digunakan pada data yang tidak memiliki label [12]. Beberapa metode tersebut di antaranya *Fuzzy C-Means* [13], *K-Means* [14], *Probabilistic Distance Clustering* dan masih banyak algoritma lainnya [15]. Dari beberapa metode yang telah disebutkan, sistem pengelompokan akan dilakukan perbandingan ketiga metode tersebut yakni *Fuzzy C-Means* (FCM), *K-Means*, dan *Probabilistic Distance Clustering* (PDC). *K-Means* adalah metode pengelompokan yang paling umum digunakan dan memiliki algoritma yang paling sederhana [16]. Sedangkan FCM merupakan metode yang menggunakan logika *fuzzy* di dalamnya yang memanfaatkan himpunan *fuzzy set* [17]. Berbeda dengan metode PDC, metode ini menggunakan nilai probabilitas untuk menentukan *cluster* pada masing-masing data dengan membandingkan dengan nilai jarak antar *centroid* [15].

Terdapat penelitian-penelitian yang memakai ketiga metode tersebut, di antaranya adalah penelitian yang dilakukan oleh Yusuf (2021). Penelitian yang dilakukan menggunakan metode PDC dan mendapatkan tiga *cluster* terbaik dengan evaluasi cluster menggunakan *silhouette coefficient* [18]. Penelitian lain yang menggunakan metode FCM dilakukan oleh Ulinnuha (2020) dengan mengelompokkan perkebunan di Indonesia menurut provinsi pada tahun 2017 dengan 3 *cluster* optimal yang didapatkan nilai *silhouette coefficient* sebesar 0,84321191 [19]. Adapun penelitian yang menggunakan metode *K-Means*, salah satunya penelitian yang dilakukan oleh Aziz dkk (2022) di mana

hasil penelitian tersebut diperoleh evaluasi *cluster* terbaiknya dengan menggunakan *silhouette coefficient* adalah sebesar 0,73 yang mana termasuk optimal [20].

Tujuan dari penelitian ini adalah untuk mengetahui perbandingan metode *K-Means*, FCM, dan PDC di mana akan dicari nilai *silhouette coefficient* paling optimal dengan berdasarkan tingkat kesalahan. Dengan mengetahui hal tersebut, diharapkan dapat membantu perusahaan dalam meningkatkan kualitas sumber daya manusia pada suatu perusahaan.

## Landasan Teori

Metode *Clustering* merupakan metode *unsupervised learning* yang biasanya digunakan untuk mengetahui pola data yang belum memiliki label. *Clustering* memiliki dua macam jenis analisis, salah satunya ialah analisis non-hierarki. Metode *clustering* non-hierarki memiliki beberapa macam metode yakni *K-Means*, FCM, dan PDC dengan penjelasan sebagai berikut [20].

### **K-Means**

*K-Means* adalah metode yang paling sering digunakan pada pengelompokan data metode *unsupervised learning* yang tidak membutuhkan label data [21]. Perhitungan *K-Means* dimulai dengan menentukan jumlah *cluster* kemudian menentukan titik awal *centroid*, kemudian dilanjut dengan perhitungan jarak tiap titik data dengan pusat *cluster* menggunakan rumus sebagai berikut:

$$D_{(i,j)} = \sqrt{\sum_{k=1}^n (X_{ik} - C_{jk})^2} \quad (1)$$

Dimana  $D$  adalah jarak cluster,  $X_{ik}$  adalah nilai data pada  $i$  dan  $j$ ,  $C_{jk}$  merupakan nilai centroid pada  $j$  dan  $k$ ,  $n$  merupakan banyaknya *cluster*. Proses selanjutnya menentukan data berdasarkan pusat *cluster* terdekat, kemudian menghitung pusat *cluster* yang baru dengan rumus:

$$C_i = \frac{\sum d_i}{n_k} \quad (2)$$

dengan  $C_i$  adalah pusat *i*-cluster ke- $i$ ,  $d$  ialah titik data ke- $i$ , dan  $n$  ialah banyaknya *cluster*  $k$

### **Fuzzy C-Means**

*Fuzzy C-Means* (FCM) ialah metode pengklasteran data yang di dalam suatu *cluster* tersebut terdapat titik-titik data yang masing-masingnya telah ditentukan oleh nilai derajat keanggotaan [22]. Pada metode ini, pusat pada awal *cluster* masih belum sepenuhnya akurat sebab pada setiap data mempunyai derajat keanggotaan untuk tiap *cluster*. Sehingga, perulangan perlu dilakukan untuk memperbaiki derajat keanggotaannya dengan pusat *cluster*. Sehingga, pusat *cluster* akan bergerak menuju posisi yang akurat dan tepat. Proses perulangan ini didasarkan pada fungsi objektif yang memiliki nilai minimal, yang menggambarkan jarak antara data dengan pusat *cluster*. [23]. Algoritma dari *fuzzy c-means* dapat dijelaskan sebagai berikut:

1. Memasukan data yang akan di-*cluster* X, ke dalam bentuk matriks berukuran  $n \times m$  ( $n$  ialah jumlah data dan  $m$  ialah atribut tiap data).  $X_{ij}$  = data ke- $i$  ( $i = 1, 2, 3, \dots, n$ ), atribut ke- $j$  ( $j = 1, 2, \dots, m$ ).
2. Menetapkan beberapa parameter yang diperlukan, di antaranya:
  - a. Jumlah *cluster* ( $c$ )
  - b. Pangkat/bobot ( $w$ )
  - c. Maksimum iterasi ( $MaxIter$ )
  - d. *Error* terkecil ( $\varepsilon$ )
  - e. Fungsi objektif ( $P_0 = 0$ )
  - f. Iterasi awal ( $t = 1$ )
3. Menentukan bilangan random  $\mu_{ik}$  ( $i = 1, 2, \dots, n$  dan  $k = 1, 2, \dots, c$ ) yang dijadikan elemen-elemen pada matriks U. Lalu, hitung jumlah tiap kolom dengan persamaan berikut:

$$Q_i = \sum_{k=1}^c \mu_{ik} \quad (3)$$

Untuk  $j = 1, 2, \dots, n$  maka menentukan matriks *random* dapat menggunakan persamaan berikut:

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \quad (4)$$

4. Mencari pusat *cluster* ke- $k$  dengan persamaan sebagai berikut:

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (5)$$

5. Menentukan fungsi objektif pada iterasi ke- $t$ ,  $P_t$  dengan persamaan:

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ([\sum_{j=1}^m (X_{ij} - V_{kj})^2] (\mu_{ik})^w) \quad (6)$$

6. Menghitung terjadinya perubahan nilai matriks partisi:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{\frac{-1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{\frac{-1}{w-1}}} \quad (7)$$

Untuk nilai  $i = 1, 2, \dots, n$  dan nilai  $k = 1, 2, \dots, c$

7. Cek kondisi hingga berhenti dengan persamaan:

$$(P_t - (P_t - 1)) < \epsilon \quad (8)$$

Atau ( $t > MaxIter$ ) maka berhenti. Jika tidak  $t = t + 1$  maka ulang langkah ke-4.

### Probabilistic Distance Clustering (PDC)

*Probabilistic Distance Clustering* merupakan salah satu metode pengelompokan non-hirarkis di mana data diklasterisasi berdasarkan nilai probabilitas terbesar yang mempengaruhi setiap pusat *cluster* [18]. Algoritma PDC ini pertama kali diusulkan oleh Ben-Israel dan Iyigun (2008) yang secara umum algoritmanya dapat dijabarkan sebagai berikut [24]:

1. Inisialisasi nilai  $D$  sebagai data,  $C_1, C_2, \dots, C_k$  sebagai nilai pusat cluster, dan nilai  $\epsilon > 0$
2. Buat nilai random pusat *cluster* awal pada perhitungan iterasi pertama.
3. Menghitung jarak  $d_1(x), d_2(x)$  untuk setiap  $x \in D$  dengan menggunakan *euclidean distance*.

$$d(x_i, c_k) = (\sum_{i=1}^N (x_i - c_k)^2)^{1/2} \quad (9)$$

Untuk  $N$  adalah banyak data,  $c_k$  ialah pusati *cluster* ke  $k$ , dan  $d(x, c)$  ialah nilai jarak *Euclidean Distance* masing-masing data ke pusat *cluster*.

4. Menghitung pusat cluster baru akan digunakan rumus sebagai berikut.

- a) Hitung nilai probabilitas pada tiap data ke *cluster*  $C_1, C_2, \dots, C_k$

$$p_k(x_i) = \frac{\prod_{j \neq k} d_j(x_i)}{\sum_{t=1}^K \prod_{j \neq t} d_j(x_i)}, k = 1, \dots, K \quad (10)$$

Untuk nilai  $k$  ialah cluster ke- $k$ , dan  $p_k(x)$  ialah probabilitas data *cluster*  $k$ , dan  $d_j(x)$  ialah nilai jarak tiap data ke- $j$

- b) Hitung perubahan nilai pusat *cluster* yang baru pada tiap *cluster*

$$c_k = \sum_{i=1}^N \left( \frac{u_k(x_i)}{\sum_{j=1, \dots, N} u_k(x_j)} \right) x_i \quad (11)$$

Untuk  $N$  ialah banyak data,  $c_k$  merupakan pusat *cluster* ke- $k$ ,  $u_k(x_i)$  ialah nilai  $u$  data ke- $i$  cluster  $k$ . dan  $u_k(x_j)$  adalah nilai  $u$  data ke- $j$  cluster  $k$

Untuk nilai  $u$  diperoleh dengan cara sebagai berikut:

$$u_k(x_i) = \frac{p_k(x_i)^2}{d_k(x_i, c_k)} \quad (12)$$

Untuk  $p_k(x_i)$  ialah nilai probabilitas data ke- $i$  cluster  $k$  dan  $d_k(x_i, c_k)$  ialah nilai *Euclidean Distance* data ke pusat *cluster*  $k$  pada data ke- $i$

5. Hitung toleransi error,

$$\|C_1^+ - C_1\| + \|C_2^+ - C_2\| < \epsilon \quad (13)$$

Jika nilai toleransi error memenuhi maka perhitungan iterasi berhenti dan kembali lagi pada Langkah

1. Dengan  $C_1, C_2$  adalah pusat *cluster* yang lama dan  $C_1^+, C_2^+$  merupakan pusat *cluster* yang baru. Nilai probabilitas akan dihitung untuk setiap anggota  $D$  dalam masing-masing *cluster*. Kemudian,

probabilitas tersebut akan digunakan dalam proses pelabelan nilai anggota D setelah perhitungan pusat *cluster*.

6. Dalam setiap iterasinya, perhitungan pusat-pusat *cluster* dan fungsi jarak akan menghasilkan nilai berbeda. Hal yang sama juga berlaku untuk fungsi *Joint Distance Function* (JDF), yang akan berkurang secara bertahap setiap iterasi. JDF digunakan untuk mengukur jarak total antara setiap data dengan masing-masing *cluster*, dan fungsi ini bermanfaat dalam memantau setiap iterasi yang dilakukan. Semakin kecil nilai JDF yang diperoleh maka semakin baik hasil *clustering* yang diperoleh. Berikut untuk menentukan nilai JDF.

$$F(c_1, c_2, \dots, c_k) = \sum_{i=1}^N \frac{\prod_{j \neq k} d_j(x_i, c_k)}{\sum_{t=1}^K \prod_{j \neq t} d_j(x_i, c_t)} \quad (14)$$

Untuk nilai  $d(x, c)$  adalah nilai *Euclidean Distance* data ke- $i$  ke pusat cluster  $c$ ,  $x_i$  ialah data ke- $i$ ,  $c_k$  adalah pusat *cluster* dari *cluster*  $k$ ,  $p_k(x_i)$  ialah probabilitas data ke- $i$  cluster  $k$ ,  $u_k(x_i)$  ialah nilai  $u$  data ke- $i$  *cluster*  $k$ .

### **Silhouette Coefficient**

Suatu *cluster* perlu pengoptimalan dengan tujuan agar pengelompokan tersebut dapat dikatakan baik dan optimal. Nilai *silhouette coefficient* pada suatu *cluster* data diperoleh dengan mencari nilai suatu jarak antara objek dengan menggunakan rumus jarak yaitu *euclidean distance*. *Silhouette coefficient* ini diperoleh dari nilai *silhouette index* pada tiap-tiap *cluster*. Terdapat tiga tahapan yang dilakukan untuk memperoleh nilai *silhouette coefficient* [21].

1. Menghitung  $a(x_i)$  yang merupakan perhitungan jarak rata-rata antara objek yang termasuk dalam *cluster* yang sama.

$$a(x_i) = \frac{1}{|C_a|-1} \sum_{j \in C_a, i \neq j} d(x_i, x_j) \quad (15)$$

Untuk  $x_i$  ialah datai ke- $i$ ,  $C_a$  ialah jumlahi dari anggotai *cluster*  $a$ , dan  $d(x_i, x_j)$  ialah fungsi jarak tiap datai ke- $i$  pada *cluster*  $j$

2. Menghitung  $b(i)$  yang merupakan perhitungan jarak rata-rata antara objek yang berasal dari *cluster* yang berbeda.

$$b(x_i) = \min_{C_b \neq C_a} \frac{1}{|C_b|} \sum_{j \in C_b} d(x_i, x_j) \quad (16)$$

Untuk  $x_i$  ialah datai ke- $i$ ,  $C_b$  ialah jumlahi dari anggotai *cluster*  $b$ , dan  $d(x_i, x_j)$  ialah fungsi jarak tiap datai ke- $i$  pada *cluster*  $j$

3. Menghitung *sillhouette coefficient* dari nilai  $a(x_i)$  dan  $b(x_i)$  yang telah diketahui pada tiap-tiap data.

$$S_{x_i} = \frac{(b_{x_i} - a_{x_i})}{\max \{a_{x_i}, b_{x_i}\}} \quad (17)$$

### **Metode Penelitian**

Penelitian ini dilakukan dalam bentuk penelitian deskriptif kuantitatif. Data yang digunakan dalam penelitian ini berasal dari dataset yang telah dikumpulkan oleh Raji Maham dan Omar Khalid Bhatti pada tahun 2020 [25], menggunakan metode pengumpulan data berupa survei atau kuesioner (dalam bentuk kategori). Data yang digunakan dalam penelitian ini terdiri dari kuesioner dengan pertanyaan-pertanyaan mengenai *Islamic Piety* atau kesalehan, yang terbagi menjadi dua kategori, yaitul *Islamic Spirituality* (IS) dan *Islamic Social Responsibility* (ISR). Data tersebut awalnya berbentuk kategorik, namun telah diubah menjadi data numerik yang belum memiliki kelas. Data tersebut kemudian dibagi menjadi dua jenis, yaitu data kategorik dan data numerik. Berikut ini merupakan contoh sampel data yang terdiri dari data kategorik dan data numerik.

**Tabel 1.** Sampel Data

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	5	1	4	2	2	3	2	39	34	16	15	48	18	18	26	31	14
2	1	2	3	2	1	1	2	30	27	9	19	53	18	23	20	35	16
1	3	1	3	2	3	3	1	32	27	24	18	49	15	16	25	24	11
1	2	2	2	2	1	3	1	49	42	22	21	78	21	28	29	42	21
1	2	2	2	1	2	2	1	38	29	20	15	59	13	19	29	32	16
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
2	3	1	3	1	2	3	1	19	9	6	7	48	3	7	11	18	9

Pada variabel A-H merupakan data kategorik dan variabel I-R merupakan data numerik dengan penjabaran di antaranya adalah. A: Gender (1 adalah laki-laki dan 2 adalah perempuan). B: Usia (1 menyatakan 20-24, 2 untuk 25-29, 3 untuk 30-34, 4 untuk 35-39, dan 5 untuk  $\geq 40$ ). C: Status (1 menyatakan Menikah dan 2 belum menikah). D: Pendidikan (1 adalah sarjana (S1), 2 adalah Magister (S2), dan 3 adalah MS/MPhil(S3)). E: Designation (1 untuk Nonmangerial, 2 untuk Line Manager, dan 3 untuk Middle Manager). F: Lama Bekerja Saat Ini (1 untuk  $\leq 1$  Tahun, 2 untuk 1-5 Tahun, dan 3 untuk 6-10 Tahun). G: Pengalaman Kerja (1 menyatakan  $\leq$  Tahun, 2 menyatakan 2-5 tahun dan 3  $\geq 6$  Tahun). H: *No of Employees* (1 untuk 1-10, 2 untuk 11-20, 3 untuk  $>20$ ). I: Ingat Kepada Allah (rentang 1-49). J: Shadaqah (rentang 1-42). K: Emosi (rentang 1-28). L: Memaafkan (rentang 1-21). M: Integritas (rentang 1-84). N: Cinta Keluarga (rentang 1-21). O: Kesabaran (rentang 1-28). P: Tobat (rentang 1-35). Q: Rituals (rentang 1-42). R: Kejujuran (Range 1-21)

Terdapat beberapa proses yang dilakukan untuk mengelompokkan karyawan berdasarkan kesalehan pada penelitian ini, proses yang dimaksudkan adalah ketiga metode yang digunakan dalam penelitian yaitu *K-Means*, FCM, PDC. Dimana pada hasil *cluster* pada ketiga metode tersebut akan dievaluasi nilai akhir *cluster*-nya dengan melihat nilai *Silhouette Coefficient* sebagai pembandingnya.

## Hasil dan Pembahasan

Proses perbandingan pengelompokan atau *clustering* akan dimulai dengan menginisialisasi jumlah *cluster* terlebih dahulu dengan nilai *cluster* atau  $k=2$  dan toleransi error ( $\epsilon = 0.001$ ). Setelah nilai  $k$  dan toleransi error ditentukan, langkah selanjunya akan ditentukan pada masing-masing metode dan dengan bantuan perhitungan *software* pengolah komputasi maka akan didapatkan hasil sebagai berikut.

Pada *K-Means* awal pusat cluster akan *di-random* yang kemudian dicari masing-masing jarak datanya dengan pusat cluster dan iterasi akan berhenti jika tidak ada perubahan pada pusat cluster. Berikut hasil pusat *cluster* pada iterasi ke-10 (terakhir) perhitungan *K-Means*.

**Tabel 2.** Pusat *Cluster K-Means*

	A	B	C	...	Q	R
Cluster 1	1,5979	1,4911	1,4911	...	31,1886	13,5708
Cluster 2	1,3379	2,7900	1,4840	...	22,7854	18,1850

Berbeda dengan *K-means*, dalam perhitungan FCM ada beberapa parameter yang perlu ditetapkan terlebih dahulu. Awal perhitungan pada FCM akan dihitung dengan matriks random U dengan ukuran 500x18. Pada pusat *cluster* juga menggunakan matriks *random* awal perhitungan yang akan memengaruhi pusat *cluster*. Berikut didapatkan pusat *cluster* pada iterasi ke-17 (terakhir).

**Tabel 3.** Pusat *Cluster* FCM

	<b>A</b>	<b>B</b>	<b>C</b>	...	<b>Q</b>	<b>R</b>
<i>Cluster 1</i>	1,2565	2,6014	1,4756	...	31,3195	13,7789
<i>Cluster 2</i>	1,3378	2,7964	1,4905	...	23,3215	18,0826

Pada PDC, setiap perhitungan iterasi terdapat nilai probabilitas pada tiap data untuk mengetahui data yang dihitung masuk ke *cluster* mana dengan melihat nilai probabilitas terbesarnya. Pusat cluster juga bergantung pada nilai probabilitas tersebut, berikut pusat cluster pada iterasi ke-31 (terakhir).

**Tabel 4.** Pusat *Cluster* PDC

	<b>A</b>	<b>B</b>	<b>C</b>	...	<b>Q</b>	<b>R</b>
<i>Cluster 1</i>	1,3212	2,7847	1,4756	...	25,1631	14,8367
<i>Cluster 2</i>	1,2619	2,6482	1,4905	...	29,9791	17,5454

Perhitungan akhir iterasi pada masing-masing metode diperoleh nilai akhir cluster sebagai berikut.

**Tabel 5.** Hasil Akhir *Cluster*

<b>Data ke-<i>i</i></b>	<b>Cluster</b>		
	<b>K-means</b>	<b>FCM</b>	<b>PDC</b>
1	1	2	1
2	2	1	1
3	2	1	1
4	1	2	2
:	:	:	:
500	2	1	1

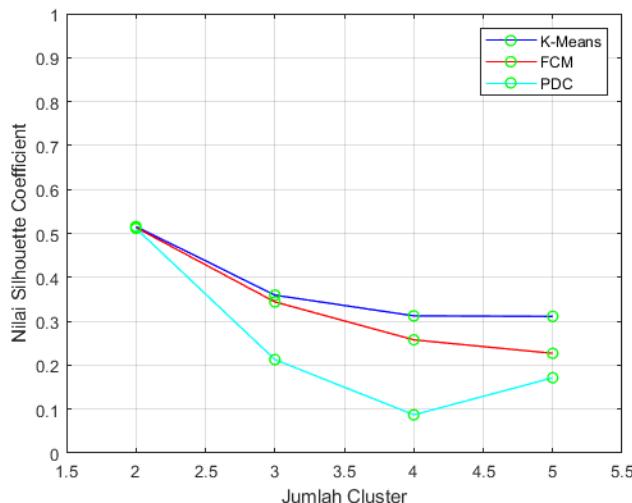
Berdasarkan pada Tabel 5 diperoleh bahwa pada masing-masing metode didapatkan data dengan kelompok yang berbeda. Pada *K-Means* sebanyak 281 data masuk pada *cluster* 1 dan 219 data masuk pada *cluster* 2. Sedangkan pada metode FCM *cluster* 1 beranggotakan 224 data dan *cluster* 2 sebanyak 276 data. Sementara pada PDC sebanyak 237 data masuk *cluster* 1 dan 263 masuk *cluster* 2. *Cluster* 1 dilabeli sebagai kelompok karyawan tidak saleh, sementara *cluster* 2 dilabeli sebagai kelompok karyawan saleh. Perincian data yang masuk pada masing-masing *cluster* dapat dilihat pada Tabel 6 berikut.

**Tabel 6.** Data Masing-Masing *Cluster*

Metode	Banyaknya Data	
	<i>Cluster 1</i>	<i>Cluster 2</i>
<i>K-Means</i>	281	219
FCM	224	276
PDC	237	263

Setelah didapatkan hasil pada masing-masing metode maka dilakukan evaluasi *cluster* dengan menggunakan *Silhouette Coefficient* (SC), dimana dilakukan uji coba dengan nilai  $k=\{2,3,4,5\}$ . Didapatkan nilai SC pada masing-masing metode dengan  $k=2$  sebesar 0,5137 pada K-Means, dan

0,5132 pada FCM, serta 0,5114 pada PDC. Sehingga didapatkan nilai SC terbaiknya pada  $k=2$  dengan rata-rata masing-masing metode sebesar 0,5134. Pada uji coba nilai  $k$  menunjukkan bahwa didapatkan pada uji coba jumlah *cluster* diperoleh nilai di bawah 0,5 yang menandakan jika  $k > 2$  maka *cluster* tidak dapat dikatakan optimal. Hasil uji coba nilai  $k$  dapat dirangkum dengan sajian Gambar 2 berikut.



**Gambar 1.** Perbandingan Nilai SC

## Kesimpulan

Berdasarkan hasil pengelompokan dengan menggunakan *K-Means*, FCM, dan PDC didapatkan nilai evaluasi terbaiknya dengan jumlah cluster  $k = 2$  pada masing-masing metode dengan iterasi paling sedikit ditunjukkan pada metode *K-means* sebanyak 10 iterasi. Pada FCM perhitungan iterasi sebanyak 17 dan PDC sebanyak 31 iterasi. *Cluster 1* dikategorikan sebagai kelompok karyawan tidak saleh sementara *cluster 2* dikategorikan sebagai kelompok karyawan saleh. Dengan masing-masing perincian pada *K-means* sebanyak 281 data termasuk *cluster 1* dan 219 data termasuk *cluster 2*. Pada FCM sebanyak 224 data termasuk *cluster 1* dan 276 termasuk *cluster 2*. Sementara pada PDC sebanyak 237 data termasuk *cluster 1* dan 263 data termasuk *cluster 2*.

## Referensi

- [1] R. F. Oley, D. P. E. Saerang, and M. Pandowo, “Analyzing The Effect Of Work Discipline And Employee Morale On Employee Performance PT. Bank Sulutgo,” *J. EMBA*, vol. 7, no. 1, pp. 351–360, 2019.
- [2] B. Al Kurdi, M. Alshurideh, and T. Al afaishat, “Employee retention and organizational performance: Evidence from banking industry,” *Manag. Sci. Lett.*, vol. 10, no. 16, pp. 3981–3990, 2020, doi: 10.5267/j.msl.2020.7.011.
- [3] M. S. Rahman, M. N. Haq, and M. M. Karim, “An Analysis on Job Satisfaction of Employees in Govt. Banks: A Study on Janata Bank Limited,” *IOSR J. Bus. Manag.*, vol. 19, no. 06, pp. 18–25, 2018, doi: 10.9790/487x-1906031825.
- [4] R. Rahmayati, “Competition Strategy in the Islamic Banking Industry: An Empirical Review,” *Int. J. Business, Econ. Soc. Dev.*, vol. 2, no. 2, pp. 65–71, 2021, doi: 10.46336/ijbesd.v2i2.133.
- [5] W. S. Sari, M. I. Fasa, and Suharto, “Implementasi Prinsip Pengembangan Sumber Daya Manusia (Sdm) Dalam Penerapannya Dibeberapa Bidang Berbasis Ekonomi Dan Syariah,” *KarismaPro*, vol. 13, no. 1, pp. 32–44, 2022, doi: 10.53675/karismapro.v13i1.254.
- [6] O. K. Bhatti, U. S. Aslam, A. Hassan, and M. Sulaiman, “Employee motivation an Islamic perspective,” *Humanomics*, vol. 32, no. 1, pp. 33–47, 2016, doi: 10.1108/H-10-2015-0066.
- [7] A. Sani and V. M. Ekowati, “Spirituality at work and organizational commitment as moderating variables in relationship between Islamic spirituality and OCB IP and influence toward employee performance,” 2019, doi: 10.1108/JIMA-08-2018-0140.
- [8] D. Junaedi, “Karakteristik Orang Shaleh Dalam Surat Muhammad,” *J. Andi Djemma J. Pendidik.*, vol. 3, no. 1, pp. 32–45, 2019.

- [9] A. A. Larasati, V. R. Zainal, and L. C. Nawangsari, "Konsep Penilaian Kinerja Karyawan Berbasis Akhlak," *J. Penelit. dan Pengabd. Masy.*, vol. 2, no. 6, 2022, doi: 10.36418/comserva.v2i6.352.
- [10] R. Maham, O. K. Bhatti, and A. O. Öztürk, "Impact of Islamic spirituality and Islamic social responsibility on employee happiness with perceived organizational justice as a mediator," *Cogent Bus. Manag.*, vol. 7, no. 1, 2020, doi: 10.1080/23311975.2020.1788875.
- [11] G. R. Zandi, N. M. Kamil, M. Sulaiman, and I. C. Ishak, "Islamic Spirituality and Continuance Organizational Commitment : The Case of Malaysian Organizations," *J. Islam. Manag. Stud.*, vol. 1, no. 1, pp. 54–71, 2018.
- [12] R. Garcia-Dias, S. Vieira, W. H. Lopez Pinaya, and A. Mechelli, "Clustering analysis," *Mach. Learn. Methods Appl. to Brain Disord.*, pp. 227–247, 2020, doi: 10.1016/B978-0-12-815739-8.00013-4.
- [13] J. Tamaela, E. Sediyono, and A. Setiawan, "Cluster Analysis Menggunakan Algoritma Fuzzy C-means dan K-means Untuk Klasterisasi dan Pemetaan Lahan Pertanian di Minahasa Tenggara," *J. Buana Inform.*, vol. 8, no. 3, pp. 151–160, 2017, doi: 10.24002/jbi.v8i3.1317.
- [14] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [15] C. Tortora and F. Palumbo, "Clustering Mixed-Type Data Using a Probabilistic Distance Algorithm," *SSRN Electron. J.*, vol. 2, no. 50, 2022, doi: 10.2139/ssrn.4102539.
- [16] T. Alfina and B. Santosa, "Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan Gabungan Keduanya dalam Membentuk Cluster Data (Studi Kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS)," *Anal. Perbandingan Metode Hierarchical Clust. K-means dan Gabungan Keduanya dalam Clust. Data*, vol. 1, no. 1, pp. 1–5, 2012.
- [17] R. Ghaniy and F. Indriyaningsih, "Penerapan Metode Fuzzy C-Means dalam Pemilihan Program Studi Mahasiswa Baru di Perguruan Tinggi," *Teknois J. Ilm. Teknol. Inf. dan Sains*, vol. 10, no. 2, pp. 19–30, 2020, doi: 10.36350/jbs.v10i2.84.
- [18] A. Yusuf *et al.*, "Identify Elementary Student Distribution Based on Kompetisi Sains Madrasah Data Using Probabilistic Distance Clustering," *Smart Innov. Syst. Technol.*, vol. 182, pp. 281–289, 2021, doi: 10.1007/978-981-15-5224-3\_27.
- [19] N. Ulinnuha, "Provincial Clustering in Indonesia Based on Plantation Production Using Fuzzy C-Means," *J. Ilm. Teknol. dan Inf.*, vol. 9, no. 1, pp. 8–12, 2020.
- [20] S. M. Aziz and N. A. K. Rifai, "Pengelompokan Ekspor Kopi Menurut Negara Tujuan Menggunakan Metode K-Means Clustering dengan Silhouette Coefficient," *Bandung Conf. Ser. Stat.*, vol. 2, no. 2, pp. 416–424, 2022.
- [21] J.-O. Palacio-Niño and F. Berzal, "Evaluation Metrics for Unsupervised Learning Algorithms," *Comput. Sci. Artif. Intell. Univ. Granada*, 2019.
- [22] N. Ulinnuha, "Clustering Kinerja Akademis Mahasiswa Menggunakan Fuzzy C-means," *Syst. Inf. Syst. Informatics J.*, vol. 2, no. 2, pp. 29–34, 2016, doi: 10.29080/systemic.v2i2.112.
- [23] S. Kusumadewi and H. Purnomo, "Aplikasi Logika Fuzzy untuk Pendukung Keputusan," vol. 2009, no. semnasIF, pp. 1–7, 2010.
- [24] A. Ben-israel and C. Iyigun, "Probabilistic D-Clustering," vol. 26, pp. 5–26, 2008, doi: 10.1007/s00357-008-.
- [25] R. Maham and O. K. Bhatti, "Dataset on the scale validation of Islamic piety," *Data Br.*, vol. 33, p. 106360, 2020, doi: 10.1016/j.dib.2020.106360.