



Perbandingan Algoritma K-Means Dengan Algoritma Fuzzy C-Means Untuk *Clustering* Tingkat Kedisiplinan Kinerja Karyawan

Nova Agustina^a, Prihandoko^a

^aProgram Studi Teknik Informatika – Sekolah Tinggi Teknologi Bandung, nova@sttbandung.ac.id

^bFakultas Ilmu Komputer dan Teknologi Informasi-Universitas Gunadarma Jakarta, pri@staff.gunadarma.ac.id

Abstract

STT Bandung is university that has great potential to become a leading university of Bandung. To achieve the purpose of college, one of the stages that must be done is the evaluation of employee performance, namely by monitoring employee discipline. To facilitate the determination of the level of employee discipline is required data mining techniques to cluster of data. In data mining there are several methods of data clusters, which is often used is the method of K-Means with Fuzzy C-Means. Based on the research conducted both methods are grouping employee performance data into 3 clusters, namely high performance level, the level of medium performance level and low performance level. The results of this study indicate that the Fuzzy C-Means method is a better method than K-Means to do data clustering on the level of employee performance in STT Bandung because the value of validation is close to 1.

Keywords: Data Mining, Comparison, K-Means, Fuzzy C-Means

Abstrak

STT Bandung merupakan perguruan tinggi swasta yang memiliki potensi besar untuk menjadi perguruan tinggi terkemuka di Kota Bandung. Untuk mencapai tujuan perguruan tinggi, salah satu tahap yang harus dilakukan adalah evaluasi kinerja karyawan, yakni dengan pemantauan kedisiplinan karyawan. Untuk memudahkan penentuan tingkat kedisiplinan karyawan maka diperlukan teknik data *mining* untuk mengklaster data. Pada data mining terdapat beberapa metode klaster data, yang sering digunakan adalah metode K-Means dengan Fuzzy C-Means. Berdasarkan penelitian yang dilakukan kedua metode tersebut mengelompokkan data kinerja karyawan menjadi 3 kluster, yakni tingkat kinerja tinggi, tingkat kinerja sedang dan tingkat kinerja rendah. Hasil dari penelitian ini menunjukkan bahwa metode Fuzzy C-Means adalah metode yang lebih baik dibandingkan K-Means untuk melakukan mengklasteran data pada tingkat kinerja karyawan di STT Bandung karena nilai validasinya bernilai mendekati 1.

Kata kunci : Data Mining, Perbandingan, K-Means, Fuzzy C-Means

© 2018 Jurnal RESTI

1. Pendahuluan

STT-Bandung (Sekolah Tinggi Teknologi Bandung) merupakan salah satu perguruan tinggi swasta di Kota Bandung yang sedang berkembang dan memiliki potensi besar untuk menjadi perguruan tinggi terkemuka di Kota Bandung. Untuk mencapai tujuan menjadi perguruan tinggi yang terkemuka, tentu STT-Bandung perlu menyeimbangkan kebutuhan dengan kemajuan teknologi.

Penilaian karyawan merupakan salah satu tahap evaluasi kerja yang dapat meningkatkan kualitas sumber daya manusia. Penilaian kedisiplinan kinerja karyawan adalah salah satu faktor untuk memperbaiki situasi kualitas sumber daya manusia. Pemantauan dan penilaian kedisiplinan karyawan harus dilakukan secara

terus menerus menjadi sebagian ciri manajemen telah berjalan dengan baik sebagai tolak ukur pertimbangan sasaran yang telah ditetapkan.

Untuk memudahkan pihak manajemen menentukan tingkat kedisiplinan kinerja karyawan, maka teknik *clustering* data perlu digunakan. Pengelompokan data atau *clustering* merupakan suatu metode yang digunakan dalam data mining yang cara kerjanya mencari dan mengelompokkan data yang mempunyai kemiripan karakteristik antara data satu dengan data lainnya yang diperoleh [7]. Dari beberapa teknik *clustering* data yang paling sederhana dan umum menggunakan Algoritma K-Means dan Fuzzy C-Means sebagai pembandingan dimana hasil dari kedua algoritma tersebut dibandingkan

Penelitian ini bertujuan untuk membandingkan algoritma k-means dengan fuzzy c-means untuk melihat kemampuan dari setiap algoritma dalam clustering sehingga mendapatkan hasil yang terbaik di Sekolah Tinggi Teknologi Bandung menggunakan data presensi karyawan pada tahun 2017. Manfaat penelitian ini agar menjadi referensi selanjutnya dalam menggunakan metode *clustering*.

2. Tinjauan Pustaka

2.1 Data Mining

Data mining adalah proses menemukan pola yang menarik dan pengetahuan dari sejumlah besar data. Sumber data dapat mencakup database, gudang data, web, repositori informasi lainnya, atau data yang dialirkan ke dalam sistem dinamis [8]. Data mining biasa digunakan untuk klasifikasi, clustering, prediksi, estimasi, asosiasi maupun deskripsi data.

2.2 Algoritma K-Means

Algoritma K-Means adalah bekerja dengan membagi data ke dalam k buah cluster yang telah ditentukan [9] Algoritma K-Means merupakan algoritma yang sering digunakan karena penggunaannya yang sederhana dan mudah untuk diimplementasikan. Algoritma K-means merupakan metode berbasis jarak yang membagi data dalam sejumlah cluster dan setiap tahapan tertentu terhadap suatu objek. Algoritma K-Means memiliki aturan yaitu ada jumlah *cluster* yang akan diinputkan dan hanya memiliki atribut bertipe numerik.

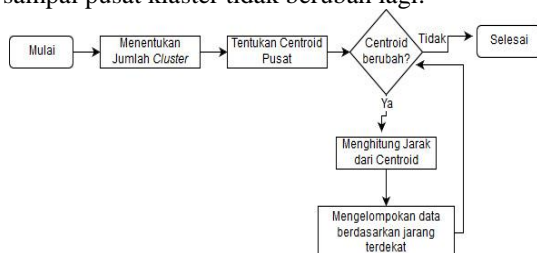
Tahapan clustering menggunakan algoritma k-means:

- Memasukan data
- Menentukan jumlah kluster
- Ambil sebarang data sebanyak jumlah kluster secara acak sebagai pusat kluster (sentroid).
- Hitung jarak antara data dengan pusat kluster, dengan menggunakan persamaan :

$$D(i, j) = \sqrt{(x_{1i} - x_{1j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (1)$$

Dimana:

- (*i, j*) = jarak data *i* ke pusat kluster *j*,
x_{ki} = data ke *i* pada atribut ke *j*,
x_{ji} = titik pusat ke *j*, pada atribut ke *i*.
 e. Hitung kembali pusat kluster dengan keanggotaan kluster yang baru
 f. Jika pusat kluster tidak berubah maka proses kluster telah selesai, jika belum maka ulangi langkah ke d sampai pusat kluster tidak berubah lagi.



Gambar 1. Flowchart K-Means

2.3 Algoritma Fuzzy C-Means

Algoritma Fuzzy C-Means merupakan algoritma yang mengelompokkan data yang kabur dan menghasilkan pengelompokan data dimana bobot keanggotaannya berasal dari data tersebut [10]. Fuzzy C-Means berhubungan dengan konsep kesamaan fungsi objek yang berdekatan dan menemukan titik pusat cluster sebagai prototype. Untuk beberapa objek data tidak memiliki batasan pada salah satu kelas saja tetapi data tersebut dapat dikelompokkan berdasarkan derajat keanggotaan yaitu antara 0 dan 1 yang menunjukkan keanggotaan parsial dari data tersebut.

Tahapan clustering menggunakan algoritma fuzzy c-means:

- Masukkan data yang akan dikluster, berupa matriks berukuran $n \times m$
- Tentukan:
 - Jumlah *cluster* (*c*)
 - Pangkat (*w*)
 - Maksimum Iterasi (Max Iter)
 - Error terkecil (ϵ)
 - Fungsi objektif awal = $P_0 = 0$
 - Iterasi awal = $t = 1$
- Bangkitkan bilangan acak (μ_{ik}), dengan $i = 1, 2, \dots, n$; $k = 1, 2, \dots, n$; dan c sebagai elemen-elemen matriks partisi awal *U*.

Hitung jumlah setiap kolom :

$$Q_i = \sum_{k=1}^c \mu_{ik} \quad (2)$$

Dengan $j = 1, 2, \dots, n$

Hitung:

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \quad (3)$$

- Hitung pusat kluster ke- *k* : V_{kj}

$$V_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^w \times x_{ij}}{\sum_{i=1}^n (\mu_{ik})^w} \quad (4)$$

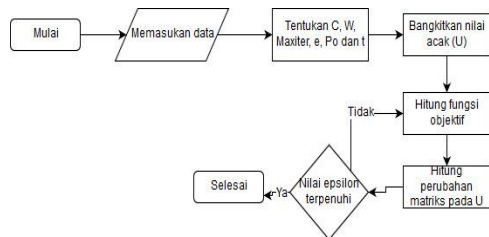
- Hitung fungsi objektif pada iterasi ke- *t*

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \quad (5)$$

- Hitung perubahan matriks partisi

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

- Cek kondisi berhenti: a. Jika: $(|P_t - P_{t-1}| \text{ MaxIter})$ maka berhenti, b. Jika tidak: $t = t + 1$, ulangi langkah ke-4 Output yang dihasilkan dari Fuzzy C-Means (FCM) merupakan deretan pusat kluster dan beberapa derajat keanggotaan untuk tiap-tiap titik data.



Gambar 2. Flowchart Fuzzy C-Means

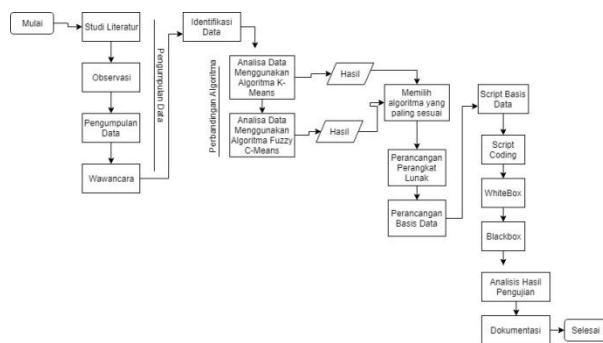
3. Metodologi Penelitian

3.1 Jenis, Sifat dan Pendekatan Penelitian

Metode analisis data yang dipakai adalah Penelitian Tindakan. Tahap awal adalah melakukan pengumpulan data melalui observasi, wawancara dan studi literatur. Selanjutnya adalah membandingkan Algoritma K-Means dengan Fuzzy C-Means, kemudian membuat rancangan sistem dan pengujian diakhiri dengan implementasi.

3.2 Metode Analisis Data

Metode analisis data yang dipakai adalah Metode Perbandingan Ekspensial. Tingkat kedisiplinan diukur melalui perbandingan hasil nilai penilaian tingkat kinerja menggunakan Algoritma K-Means dengan Algoritma Fuzzy C-Means. Pengukuran dipilih berdasarkan hasil dari teknik implementasi algoritma yang paling sesuai atau mendekati. Data yang berkaitan dengan presensi karyawan di STT Bandung terlebih dahulu diambil kemudian dilakukan *clustering*. Hasil *cluster* merupakan pengelompokan tingkat kinerja karyawan, terdiri dari Tingkat Kinerja Tinggi, Tingkat Kinerja Sedang dan Tingkat Kinerja Rendah.



Gambar 3. Alur Penelitian

4. Hasil dan Pembahasan

Pada penelitian ini akan menjelaskan mengenai perbandingan clustering data kinerja karyawan di STT Bandung menggunakan metode K-Means dan Fuzzy C-Means. Untuk menentukan pusat centroid setiap cluster memiliki cara yang berbeda pada metode Algoritma K-Means dengan Fuzzy C-Means. Data yang dianalisa adalah data jumlah masuk, jumlah ijin dan jumlah terlambat karyawan di STT-Bandung pada tahun 2017. Jumlah cluster yang akan dibentuk terdiri dari 3 klaster,

dimana klaster 1 dikategorikan tingkat kedisiplinan kinerja tinggi, klaster 2 dikategorikan tingkat kedisiplinan tingkat sedang, dan klaster 3 dikategorikan tingkat kedisiplinan kinerja rendah. Tabel data presensi karyawan adalah sebagai berikut:

Tabel 1.Data Presensi Karyawan

No	NIK	Masuk	Ijin	Terlambat
1	06.77.003	259	54	59
2	08.82.004	257	56	12
3	09.81.007	281	32	55
4	17.94.049	112	201	18
5	91.63.001	235	78	107
....
34	97.73.002	265	28	75

Pada tabel 1 dapat dilihat jumlah masuk, ijin dan terlambat karyawan di Sekolah Tinggi Teknologi Bandung selama satu tahun pada tahun 2017.

4.1 Metode Algoritma K-Means

Tahap pertama untuk *clustering* pada algoritma K-Means adalah menentukan pusat centroid setiap klaster. Pusat centroid ditentukan secara *random* pada iterasi 1 berdasarkan data yang sudah ada. Pada tahap ini ditentukan pusat centroid secara random sebagai berikut :

Tabel 2.Pusat Centroid Klaster

	Masuk	Ijin	Terlambat
C1	296	17	52
C2	240	71	12
C3	205	108	65

Selanjutnya adalah menghitung jarak terhadap pusat klaster menggunakan rumus :

$$D(i, j) = \sqrt{(x_{1i} - x_{1j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (7)$$

Contoh menghitung jarak terhadap pusat untuk nik 06.77.003:

Klaster 1 (C1)

$$D(06.77.x, C1) = \sqrt{(259 - 246)^2 + (54 - 17)^2 + (69 - 52)^2} = 55.018$$

Klaster 2 (C2)

$$D(06.77.x, C1) = \sqrt{(259 - 240)^2 + (54 - 71)^2 + (69 - 12)^2} = 40.012$$

Klaster 3(C3)

$$D(06.77.x, C1) = \sqrt{(259 - 205)^2 + (54 - 108)^2 + (69 - 65)^2} = 256.166$$

Sehingga didapatkan hasil jarak berikut ini untuk Iterasi 1:

Tabel 3. Jarak Terhadap Centroid Pusat

No	NIK	C1	C2	C3
1	06.77.003	55.018	40.012	286.166
2	08.82.004	68.132	31.112	282.449
3	09.81.007	21.424	57	315.941
4	17.94.049	262.427	186.134	78.460
5	91.63.001	102.308	71.895	260.190
....
34	97.73.002	49.507	49.889	295.328

Selanjutnya adalah menentukan keanggotaan klaster dan jarak minimum.

Tabel 4. Keanggotaan Klaster

No	NIK	MJ	Ka	KMJ
1	06.77.003	40.012	C2	286.166
2	08.82.004	31.112	C2	282.449
3	09.81.007	21.424	C1	315.941
4	17.94.049	78.460	C3	78.460
5	91.63.001	71.893	C2	260.190
....
34	97.73.002	49.507	C1	295.328

Dimana :

Ka = Keanggotaan

MJ = Minimal Jarak

KMJ = Kuadrat Min Jarak

Selanjutnya adalah menentukan pusat klaster baru:

Tabel 4. Jarak Pusat Klaster

No	NIK	C1	C2	C3
		M, I, T	M, I, T	M, I, T
1	06.77.003		259,54,59	
2	08.82.004		257,56,12	
3	09.81.007	281,32,55		
4	17.94.049			112,201,18
5	91.63.001		235,78,2017	
....
34	97.73.002	265,48,75		

Dimana:

M,I,T = Masuk, Ijin, Terlambat

Pusat klaster baru ditentukan dengan menjumlahkan seluruh jumlah masuk C1, jumlah Ijin C1, Terlambat C1 kemudian dibagi dengan jumlah keanggotaan. Begitupun dengan Klaster 2 dan Klaster 3. yang ada pada pusat klaster baru, contoh:

$$M1 = (281 + .. + 265)/12 = 278.833$$

$$I1 = (32 + .. + 48)/12 = 34.166$$

$$T2 = (55 + .. + 75)/12 = 65.75$$

Didapatkan hasil sebagai berikut:

Tabel 5. Pusat Centroid Klaster

	Masuk	Ijin	Terlambat
C1	278.833	34.166	65.75
C2	238.894	74	59.631
C3	93.666	219.333	27.333

Kemudian lanjutkan ke Iterasi selanjutnya menggunakan Pusat Centroid Klaster Baru hingga keanggotaan baru dengan keanggotaan sebelumnya tidak berubah.

Berdasarkan data yang telah diperoleh dan diolah menggunakan algoritma K-means didapatkan hasil clustering pada tabel 6.

Tabel 6. Tabel Hasil Clustering Menggunakan Algoritma K-Means

No	NIK	C1	C2	C3	Hasil
1	06.77.003	20,378	47,022	23,750	C1
2	08.82.004	53,766	67,443	23,149	C1
3	09.81.007	13,762	78,218	26,637	C1
4	17.94.049	230,80	167,45	27,556	C3
5	91.63.001	69,868	45,463	215,16	C2
....
34	97.73.002	17,239	49,889	246,94	C1

Klaster 1, 2 dan 3 dibandingkan kemudian ambil klaster dengan nilai terkecil untuk menentukan keanggotaan klaster.

4.2 Metode Algoritma Fuzzy C-Means

Tahap pertama untuk *clustering* pada algoritma Fuzzy C-Means adalah menentukan jumlah cluster yang dicari, maksimum iterasi, nilai pembobot dan nilai kesalahan terkecil sesuai dengan kebutuhan, dimana pada studi ini peneliti menentukan nilai-nilai tersebut dapat dilihat pada tabel 7.

Tabel 7. Penentuan nilai awal Fuzzy C-Means

No	Kebutuhan	Jumlah
1	Jumlah Cluster yang Dicari	3
2	Maksimum Iterasi	100
3	Nilai Pembobot (Pangkat)	2
4	Nilai Kesalahan Terkecil	0.000001

Selanjutnya adalah menentukan matriks keanggotaan secara random, keanggotaan yang dilakukan oleh peneliti pada studi ini adalah sebagai berikut:

Tabel 8. Matriks Keanggotaan

NIK	C1	C2	C3	Jumlah
06.77.003	0.3	0.3	0.4	1
08.82.004	0.3	0.5	0.2	1
09.81.007	0.8	0.1	0.1	1
17.94.049	0.5	0.2	0.3	1
91.63.001	0.3	0.3	0.4	1
.....
97.73.002	0.3	0.5	0.2	1

Dimana :

$$C1 + C2 + C3 = 1.$$

Setelah menentukan matriks secara random, selanjutnya adalah menentukan μ^2 berdasarkan matriks random yang telah ditentukan, dapat dilihat pada tabel 9.

Tabel 9. μ^2

NIK	C1	C2	C3
06.77.003	0.09	0.09	0.016
08.82.004	0.09	0.025	0.04
09.81.007	0.064	0.01	0.01
17.94.049	0.025	0.04	0.09
91.63.001	0.09	0.09	0.016
.....
97.73.002	0.09	0.025	0.04
TOTAL	6.41	3.57	4.48

Setelah mendapatkan hasil dari μ^2 , selanjutnya adalah mengalikan *cluster* pada masing-masing μ^2 dengan data presensi karyawan (masuk, ijin, terlambat) sehingga didapatkan hasil yang dapat dilihat pada tabel 10, 11 dan 12.

Tabel 10. $\mu^2 \times 1$

NIK	C1	C2	C3
06.77.003	23.31	4.86	6.21
08.82.004	23.13	5.04	1.08
09.81.007	179.84	20.48	35.2
17.94.049	65	13.25	25.75
91.63.001	21.15	7.02	9.63
.....
97.73.002	23.85	4.32	6.75
TOTAL	1546.65	458.4	366.13

Tabel 11. Miu² X2

NIK	C1	C2	C3
06.77.003	23.31	4.86	6.21
08.82.004	64.25	14	3
09.81.007	2.81	0.32	0.55
17.94.049	10.4	2.12	4.12
91.63.001	21.15	7.02	9.63
.....
97.73.002	66.25	12	18.75
TOTAL	827,15	290,24	222,91

Tabel 12. Miu² X3

NIK	C1	C2	C3
06.77.003	41.44	8.64	11.04
08.82.004	10.28	2.24	0.48
09.81.007	2.81	0.32	0.55
17.94.049	23.4	4.77	9.27
91.63.001	37.6	12.48	17.12
.....
97.73.002	10.6	1.92	3
TOTAL	1109,78	292,44	257,68

Pusat kluster baru ditentukan dengan menjumlahkan seluruh jumlah masuk C1, jumlah Ijin C1, Terlambat C1 pada Miu² kemudian dibagi dengan jumlah keanggotaan Miu² X1, X2 dan X3. Contoh perhitungan dan hasil dapat dilihat berikut:

$$V_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^w \times x_{ij}}{\sum_{i=1}^n (\mu_{ik})^w} \quad (8)$$

$$M1 = 1546.65/6.41 = 241.28$$

$$I1 := 458.4/3.57 = 71.51$$

$$T1 = 366.13/4.48 = 57.11$$

Tabel 13. Pusat Cluster

	Masuk	Ijin	Terlambat
C1	241.28	71.51	57.11
C2	231.69	81.29	62.43
C3	247.71	65.27	57.51

Langkah selanjutnya adalah menghitung jumlah setiap kolom, contoh:

$$Q_i = \sum_{k=1}^c$$

$$M1 = ((259 - 241.28)^2 + (54 - 241.28)^2 + (59 - 241.28)^2) = 761.63$$

$$I1 = ((259 - 71.51)^2 + (54 - 71.51)^2 + (59 - 71.51)^2) = 1533.89$$

$$T1 = ((259 - 57.51)^2 + (54 - 57.51)^2 + (59 - 57.51)^2) = 386.27$$

Selanjutnya adalah menghitung nilai L, dimana L= X_V dikalikan dengan Miu². Hasilnya dapat dilihat pada tabel 14.

Tabel 14. Nilai L

NIK	C1	C2	C3	Total
06.77.003	68.54	138.05	61.80	268.40
08.82.004	227.09	956.15	89.76	1273.00
09.81.007	2011.45	49.16	22.21	2082.84
17.94.049	699.50	129.88	213.31	1042.70
91.63.001	231.27	180.66	443.54	855.48
.....
97.73.002	129.14	593.96	36.11	759.222

Fungsi Objective =

$$P_t = \sum_{t=1}^n \sum_{k=1}^c (|\sum_{j=1}^m (X_{ij} - V_{kj})^2| (\mu_{ik})^w)$$

$$= 268.40 + 1273 + 2081.84 + 1042.70 + \dots + 855.48$$

$$= 85089.07$$

Selanjutnya cek kondisi berhenti, dengan melakukan :

- Jika: $(|P_t - P_{t-1}| \text{ MaxIter})$ maka berhenti
- Jika tidak: $t = t + 1$. Hasilnya dapat dilihat pada tabel 15.

Tabel 15. Nilai LT

NIK	C1	C2	C3	Total
06.77.003	0.0013	0.0006	0.025	0.0045
08.82.004	0.0003	0.0002	0.0004	0.0110
09.81.007	0.0003	0.0002	0.0004	0.0009
17.94.049	0.0003	0.0003	0.0004	0.0010
91.63.001	0.0003	0.0004	0.0003	0.0012
.....
97.73.002	0.0006	0.0004	0.0001	0.0022

Selanjutnya adalah menentukan keanggotaan baru dimana setiap kluster pada data LT dibagi dengan Total LT, sehingga didapatkan hasil berikut :

Tabel 16. Keanggotaan Baru

NIK	C1	C2	C3
06.77.003	0.288	0.143	0.568
08.82.004	0.359	0.236	0.403
09.81.007	0.327	0.209	0.463
17.94.049	0.328	0.283	0.388
91.63.001	0.311	0.399	0.289
.....
97.73.002	0.313	0.189	0.497

Selanjutnya menghitung selisih Objective pada iterasi pertama, dengan menggunakan nilai mutlak :

$$\begin{aligned} \text{Selisih Objective} &= \text{ABS}(\text{Fungsi Objective} - 0) \\ &= \text{ABS}(85089.07 - 0) \\ &= 85089.07 \end{aligned}$$

Proses Iterasi ke 1 pada Fuzzy C-Means telah selesai, selanjutnya melanjutkan proses Iterasi Ke 2 dengan menggunakan keanggotaan baru pada tabel 16 (sebelumnya keanggotaan ditentukan secara random) dan dilakukan seterusnya hingga selisih fungsi objective kurang dari kesalahan terkecil. Perhitungan fungsi objective pada Iterasi Kedua, menggunakan rumus :

$$\text{Selisih Objective} = \text{ABS}(\text{Fungsi Objective Iterasi ke 2} - \text{selisih objective Iterasi ke 1}).$$

Untuk proses clustering menggunakan Fuzzy C-Means didapatkan pada iterasi ke 23, berikut adalah penjelasannya :

$$\begin{aligned} \text{Selisih Fungsi Objective} &= \text{ABS}(\text{Fungsi Objective ke 23} - \text{Selisih Fungsi Objective ke 22}) \\ &= \text{ABS}(30059.512716802 - 30059.512717362) \\ &= 0.00000056 \end{aligned}$$

Selisih fungsi objective pada Iterasi ke 23 < Kesalahan terkecil yang diharapkan.

Tabel 17. Tabel Hasil Clustering Fuzzy C-Means

No	NIK	K1	K2	K3	Hasil
1	06.77.003	0.155	0.005	0.879	C3
2	08.82.004	0.464	0.032	0.503	C3
3	09.81.007	0.047	0.004	0.984	C3
4	17.94.049	0.248	0.019	0.731	C3
5	91.63.001	0.564	0.038	0.3997	C1
....
34	97.73.002	0.044	0.002	0.953	C3

4.3 Validitas

4.3.1 Validasi Algoritma K-Means

Untuk mendapatkan validitas dari algoritma K-means, maka digunakan validitas Silhouette Index. Hasil didapatkan dengan nilai Indeks rata-rata yang didapat adalah 0,528.

4.3.2 Validasi Algoritma Fuzzy C-Means

Untuk mendapatkan validitas dari algoritma Fuzzy C-Means, maka digunakan validitas menggunakan Partition Coefficient Index (PCI) dengan rumus sebagai berikut:

a. Validasi PCI (U_i^2)

Tabel 18. Nilai validasi metode Fuzzy C-Means

No	NIK	K1	K2	K3
1	06.77.003	0,0134	2,6190	0,7727
2	08.82.004	0,2153	0,0010	0,2537
3	09.81.007	0,0022	1,7231	0,8993
4	17.94.049	0,0004	0,9530	0,0001
5	91.63.001	0,3187	0,0014	0,1577
....
34	97.73.002	0,0019	6,8709	0,9082
Jumlah Semua Kluster				25,784

b. Nilai derajat keanggotaan

Tabel 19. Nilai derajat keanggotaan Algoritma Fuzzy C-Means

No	NIK	K1	K2	K3
1	06.77.003	0,1159	0.0051	0.8790
2	08.82.004	0.4640	0.0322	0.5037
3	09.81.007	0.0475	0.0041	0.9483
4	17.94.049	0.0218	0.9669	0.0111
5	91.63.001	0.5646	0.0381	0.3971
....
34	97.73.002	0.0443	0.0026	0.9530

$$PCI = (1/34)(25,784) = 0,758$$

5. Kesimpulan

5.1 Simpulan

Berdasarkan hasil penelitian yang telah diperoleh, dapat disimpulkan bahwa hasil cluster dari data

presensi karyawan menggunakan metode K-Means dan Fuzzy C-Means berbeda. Hal ini dapat dilihat dari jumlah cluster yang diperoleh dari kedua metode tersebut. Dilihat dari hasil validasi, Fuzzy C-Means dominan menghasilkan metode yang lebih baik, dengan nilai validasinya adalah 0,758 dikarenakan nilai validasinya lebih mendekati nilai 1, dibandingkan dengan metode K-Means dengan nilai validasinya adalah 0,528.

5.2 Saran

Penelitian ini masih dapat dikembangkan dengan membandingkan beberapa algoritma clustering lainnya, sehingga semakin banyak membandingkan algoritma lainnya, maka semakin terlihat algoritma mana yang memang paling cocok untuk digunakan di studi kasus *clustering* kinerja karyawan.

Daftar Rujukan

- [1] H.T. Kahraman et al. *The development of intuitive knowledge classifier and the modeling of domain dependent data. Knowledge-Based Systems* 37 (2013) 283–295
- [2] Agusta, Yudhi. K-Means – Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika*. 2007; Vol. 3, 47–60.
- [3] Simbolon, Cary Lineker. Clustering Lulusan Mahasiswa Matematika Fmipa Untan Pontianak Menggunakan Algoritma Fuzzy C-Means. *Buletin Ilmiah Mat. Stat. Dan Terapannya (Bimaster)*. 2013; Volume 02, No.1, 21–26.Referensi yang berasal dari **situs web**
- [4] Muzakir, Ari. Analisa Dan Pemanfaatan Algoritma K-Means Clustering pada Data Nilai Siswa Sebagai Penentuan Penerima Beasiswa. *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)* 2014; Pp 195–200.
- [5] Selviana, Nur Indah dan Mustakim. Analisis Perbandingan K-Means dan Fuzzy C-Means Untuk Pemetaan Motivasi Balajar Mahasiswa. *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI)* 8. 2016. *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI)* 9 ISSN (Printed) : 2579-7271 Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau ISSN (Online) : 2579-5406 Pekanbaru, 18-19 Mei 2017 226
- [6] Mustakim. Pemetaan Digital dan Pengelompokan Lahan Hijau di Wilayah Provinsi Riau Berdasarkan Knowledge Discovery in Databases (KDD) dengan Teknik K-Means Mining. *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI)*. 2012.
- [7] J. O. Ong, “Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University,” *Ilmiah Teknik Industri*, vol. 12, 2013.
- [8] Jiawei, Han, *data mining Concepts and Techniques*. USA: Elsevier Inc. All rights reserved. 2012.
- [9] Han, J dan Kamber, M., “Data Mining Concept and Technique”, Morgan Kaufmann, 2001.
- [10] Ghosh, S., Dubey, S.K., 2013. *Comparative Analysis of K-Means and Fuzzy C-Means Algorithms*. *International Journal of Advanced Computer Science and Applications*, Vol. 4, No.4