

Perbandingan *K-Means* dan *Fuzzy C-Means* untuk Pengelompokan Data *User Knowledge Modeling*

Aditya Ramadhan¹, Zuliar Efendi², Mustakim³

^{1,2,3}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. HR. Soebrantas Km. 18 Panam Pekanbaru – Riau

e-mail: ¹aditya.ramadhan@students.uin-suska.ac.id, ²zuliar.efendi@students.uin-suska.ac.id,
³mustakim@uin-suska.ac.id

Abstrak

Pada data mining terdapat sebuah metode yang digunakan untuk mengklaster data menjadi kelompok-kelompok data, yaitu metode *K-Means* dan *Fuzzy C-Means*. Kedua metode tersebut jika dilihat dari beberapa penelitian sebelumnya mengenai clustering *K-Means* dan *FCM*, masing-masing metode mampu memberikan hasil cluster terbaik. Pengklasteran data *user knowledge modeling* menggunakan metode *K-Means* dan *Fuzzy C-Means* menghasilkan jumlah anggota klaster yang berbeda. Dapat dilihat dari jumlah klaster yang diperoleh dari kedua metode tersebut. Berdasarkan penelitian yang dilakukan, kedua metode tersebut mengelompokkan data *user knowledge modeling* menjadi 4 kluster. Perbandingan yang digunakan dalam penelitian ini adalah uji performa validitas. Untuk nilai validasi *SI* dari metode *K-Means* bernilai 0.1866, sedangkan nilai validasi *PCI* dari metode *FCM* adalah bernilai 0.2854. Hasil dari penelitian ini menunjukkan bahwa metode *FCM* adalah metode yang lebih baik daripada *K-Means* untuk melakukan clustering pada data *user knowledge modeling* dikarenakan nilai validasinya bernilai mendekati 1.

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

Abstract

In data mining, there is a method used to cluster data into groups of data, the method is *K-Means* and *Fuzzy C-Means*. That two methods when we viewed from some previous research on *K-Means* clustering and *FCM*, each method capable of providing the best cluster. Clustering data *user knowledge modeling* using *K-Means* and *Fuzzy C-Means* clustering produce a different number of members. Can be seen from the number of clusters obtained from the two methods. Both of these methods can do that for clustering of data *user modeling knowledge* into 4 clusters. Comparisons were used in this study is to test the validity of the performance. For the validation of the *SI* value of the *K-Means* method is worth 0.1866, while the value of *PCI* validation of methods *FCM* is worth 0.2854. The result of this research show that the *FCM* method is a better than *K-Means* method to perform clustering in the data *user modeling knowledge* because its validity value worth close to 1.

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

1. Pendahuluan

Pengetahuan merupakan berbagai gejala yang ditemui dan diperoleh manusia melalui pengamatan akal. Pemodelan pengguna adalah salah satu mekanisme paling ampuh untuk aplikasi adaptif berbasis web. Penelitian ini menggunakan model siswa / pengguna untuk menyesuaikan subyek sesuai dengan pengetahuan mereka. Tujuan dari sistem pemodelan pengguna adalah untuk memberikan pengetahuan yang cukup atau sesuai bagi siswa / pengguna.

Data *user knowledge modeling* merupakan data dari 258 model pengetahuan siswa / pengguna yang diidentifikasi berdasarkan tingkat waktu belajar untuk objek tujuan, tingkat jumlah pengulangan pengguna untuk objek sasaran, tingkat waktu belajar pengguna untuk objek terkait dengan objek tujuan, kinerja ujian pengguna Untuk objek terkait dengan objek tujuan, dan kinerja ujian pengguna untuk objek tujuan[1]. Dari 258 model pengetahuan siswa tersebut, penelitian sebelumnya mengelompokkan menjadi empat kelompok tingkatan, yaitu *very low*, *low*, *middle* dan *high*.

Analisis kelompok (*cluster analysis*) telah digunakan diberbagai bidang ilmu pengetahuan, dengan tujuan mengelompokkan objek. Ada beberapa metode yang dapat digunakan untuk mengelompokkan data menjadi beberapa kelompok data, diantaranya adalah

dengan menggunakan salah satu cabang dari ilmu matematika, yaitu *data mining* dan logika *fuzzy*.

Data mining adalah adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam daftar data. *Data mining* merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai daftar data besar. *Data Clustering* merupakan salah satu metode *data mining* yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis *data clustering* yang sering dipergunakan dalam proses pengelompokan data yaitu *hierarchical* (hirarki) *data clustering* dan *non-hierarchical* (non hirarki) *data clustering* [2].

Proses clustering adalah proses pengelompokan data ke dalam cluster berdasarkan parameter tertentu sehingga obyek-obyek dalam sebuah cluster memiliki tingkat kemiripan yang tinggi satu sama lain dan sangat tidak mirip dengan obyek yang lain pada cluster yang berbeda [3]. Sampai saat ini, para ilmuwan masih terus melakukan berbagai usaha untuk melakukan perbandingan model *cluster* sehingga dapat menghasilkan cluster yang cocok digunakan pada sebuah penelitian.

Dalam *data mining* terdapat sebuah metode yang digunakan untuk mengkluster data menjadi kelompok-kelompok data, yaitu metode *K-Means*. Algoritma *K-Means* adalah algoritma klustering yang paling sederhana dibanding algoritma klustering yang lain [4]. Algoritma ini mempunyai kelebihan mudah diterapkan dan dijalankan, relatif cepat, mudah untuk diadaptasi, dan paling banyak dipraktekkan dalam tugas *data mining*.

Selain *data mining*, terdapat cabang ilmu matematika yang mempunyai metode untuk mengkluster data yaitu logika *fuzzy*. Dalam teori *fuzzy*, keanggotaan sebuah data tidak diberi nilai secara tegas dengan nilai 1 (menjadi anggota) dan 0 (tidak menjadi anggota), melainkan dengan suatu nilai derajat keanggotaan yang jangkauan nilainya 0 sampai 1 [5]. Dalam pengklasteran data, metode *Fuzzy C-Means* adalah salah satu metode yang digunakan dalam logika *fuzzy*.

Beberapa penelitian sebelumnya menerapkan teknik *K-Means Clustering* dan *Fuzzy C Means Clustering* sebagai penelitian dalam hal Pengelompokan data, diantaranya : Mustakim (2012) dalam penelitiannya yang berjudul "Pemetaan Digital dan Pengelompokan Lahan Hijau di Wilayah Provinsi Riau Berdasarkan *Knowledge Discovery in Databases* (KDD) dengan Teknik *K-Means Mining*" menjelaskan bahwa kelebihan *K-Means* adalah beban komputasi relatif lebih ringan sehingga klusterisasi bisa dilakukan dengan cepat walaupun relatif tergantung pada banyak jumlah data dan jumlah cluster yang ingin dicapai [6]. Ratna Ekawati (2013) menjelaskan dalam penelitiannya yang berjudul "Klasifikasi Usaha Kecil Dan Menengah (UKM) Sektor Industri Dengan Metode *Fuzzy C-Means Clustering* Wilayah Kota Cilegon" bahwa FCM memberikan hasil pengelompokan yang halus atau tidak banyak menggeser pusat *cluster* [7].

K-Means dan *Fuzzy C-Means* jika dilihat dari beberapa riset sebelumnya yang dilakukan Mustakim dan Ekawati mampu memberikan hasil *cluster* terbaik. Kedua metode tersebut sama-sama memiliki hasil yang signifikan serta mempunyai beberapa perbedaan dalam hal pola *cluster*. Pengujian performasi algoritma adalah salah satu cara untuk menemukan metode mana yang lebih baik yang digunakan dalam *clustering*.

Di dalam penelitian ini akan ditunjukkan perbandingan pengklasteran data *user knowledge modeling* menggunakan metode *K-Means* dan *Fuzzy C-Means*. Penelitian ini bertujuan untuk membandingkan dua algoritma *clustering* yaitu *K-Means* dan *Fuzzy C-Means* untuk melihat kemampuan dari setiap algoritma dalam *clustering* sehingga mendapatkan hasil algoritma terbaik. Manfaat dari penelitian ini adalah agar menjadi referensi bagi penelitian selanjutnya dalam pemelihan algoritma *clustering*.

2. Metode Penelitian

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]. Dalam *data mining* terdapat sebuah metode yang digunakan untuk mengkluster data, yaitu *K-Means*.

Algoritma *K-Means* merupakan algoritma yang membutuhkan parameter *input* sebanyak k dan membagi sekumpulan n objek kedalam k *cluster* sehingga tingkat kemiripan antar anggota dalam satu *cluster* tinggi sedangkan tingkat kemiripan dengan anggota pada *cluster* lain sangat rendah. Kemiripan anggota terhadap *cluster* diukur dengan kedekatan objek terhadap nilai *mean*

pada *cluster* atau disebut sebagai *centroid cluster* [9]. Metode *K-Means* merupakan metode *clustering* yang paling sederhana dan umum. Hal ini dikarenakan *K-Means* mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien [10]. Tetapi metode ini mempunyai kelemahan dalam menganalisis persebaran data serta bergantung pada inialisasi *centroid*. *K-Means* hanya melihat jarak data ke masing-masing *centroid* pada setiap *cluster* [11].

Berikut adalah algoritma dari metode *K-Means*:

- a. Masukkan data yang akan dikluster.
- b. Tentukan jumlah kluster.
- c. Ambil sebarang data sebanyak jumlah kluster secara acak sebagai pusat kluster (*sentroid*).
- d. Hitung jarak antara data dengan pusat kluster, dengan menggunakan persamaan :

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

Dimana:

- $D(i, j)$ = jarak data i ke pusat kluster j ,
- x_{ki} = data ke i pada atribut ke j ,
- x_{ji} = titik pusat ke j , pada atribut k

- 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.

2.2. Logika Fuzzy

Logika *fuzzy* pertama kali diperkenalkan oleh Prof. Lotfi A. Zadeh pada tahun 1965. Dalam banyak hal, logika *fuzzy* digunakan sebagai suatu cara untuk memetakan permasalahan dari *input* menuju ke *output* yang diharapkan. Dalam logika *fuzzy* terdapat *fuzzy clustering* yang merupakan salah satu metode untuk menentukan kluster optimal dalam suatu ruang vektor yang didasarkan pada bentuk normal *Euclidian* untuk jarak antar vektor. Tujuan dari *Fuzzy C-Means* adalah untuk mendapatkan pusat *cluster* yang nantinya akan digunakan untuk mengetahui data yang masuk ke dalam sebuah *cluster* [12].

Dalam logika *fuzzy* terdapat metode yang sering digunakan untuk mengkluster data, yaitu metode *Fuzzy C-Means*. *Fuzzy C-Means* adalah suatu metode pengklasteran data yang ditentukan oleh derajat keanggotaan. Berikut adalah algoritma *Fuzzy C-Means*:

- a. Masukkan data yang akan dikluster, berupa matriks berukuran $n \times m$
- b. Tentukan :
 - i. Jumlah kluster = c
 - ii. Pangkat = w
 - iii. Maksimum Iterasi = $Max\ Iter$
 - iv. Error Terkecil = ϵ
 - v. Fungsi objektif awal = $P_0 = 0$
 - vi. Iterasi awal = $t = 1$

- c. 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:

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

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

Hitung :

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

- d. Hitung pusat kluster ke- k : V_{kj}

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

dengan $k = 1, 2, \dots, c$ dan $j = 1, 2, \dots, m$

- e. Hitung fungsi objektif pada iterasi ke- t ,

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

f. Hitung perubahan matriks partisi:

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

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

g. Cek kondisi berhenti:

a. Jika: $(|P_t - P_{t-1}| < \epsilon)$ atau $(t > MaxIter)$ 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.

2.3. Silhouette Index (SI)

Analisa metode *silhouette* dapat digunakan untuk validasi algoritma *K-Means*. Hal ini dilakukan dengan melihat besar nilai s dari hasil perhitungan dengan menggunakan bantuan software MatLab. Hasil perhitungan nilai *silhouette coefficient* dapat bervariasi antara -1 hingga 1. Jika $si = 1$ berarti objek i sudah berada dalam cluster yang tepat. Jika nilai $si = 0$ maka objek i berada di antara dua cluster sehingga objek tersebut tidak jelas harus dimasukkan ke dalam cluster A atau cluster B. Akan tetapi, jika $si = -1$ artinya struktur *cluster* yang dihasilkan *overlapping*, sehingga objek i lebih tepat dimasukkan ke dalam *cluster* yang lain.

2.4. Partition Coefficient Index (PCI)

Bedzek (1981) mengusulkan validasi *fuzzy clustering* dengan menghitung koefisien partisi atau PC sebagai evaluasi nilai keanggotaan data pada setiap *cluster*. Nilai PCI hanya mengavaluasi nilai derajat keanggotaan, tanpa memandang vektor (data) yang biasanya mengandung informasi geometrik. Nilai dalam rentang $[0, 1]$, nilai yang semakin besar mendekati 1 mempunyai arti bahwa kualitas *cluster* yang didapat semakin baik [4].

3. Metode

3.1. Data yang digunakan

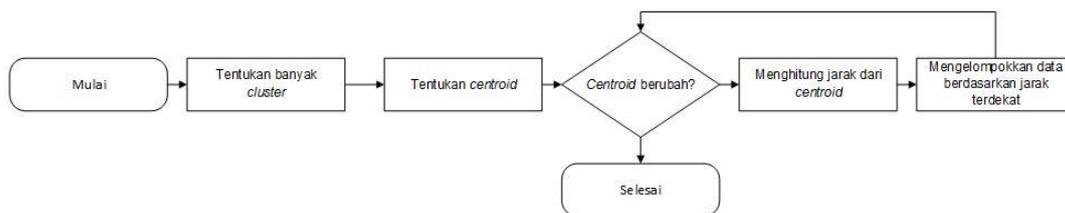
Data yang digunakan pada penelitian ini adalah data *user knowledge modeling*, yang di dapatkan pada *UCI Machine Learning Repository*. Data *user knowledge modeling* memiliki 258 *record* data dan peneliti-peneliti sebelumnya mengelompokkan menjadi empat kelompok tingkatan siswa / pengguna. Variabel yang digunakan adalah *The degree of study time for goal object materials* (STG), *The degree of repetition number of user for goal object materials* (SCG), *The degree of study time of user for related objects with goal object* (STR), *The exam performance of user for related objects with goal object* (LPR), dan *The exam performance of user for goal objects* (PEG). Dengan data sebagai berikut:

Tabel 1. Data *user knowledge modelling*

No	STG	SCG	STR	LPR	PEG
1	0	0	0	0	0
2	0.0800	0.0800	0.1000	0.2400	0.9000
3	0.0600	0.0600	0.0500	0.2500	0.3300
4	0.1000	0.1000	0.1500	0.6500	0.300
5	0.0800	0.0800	0.0800	0.9800	0.2400
....
No	STG	SCG	STR	LPR	PEG
258	0.6600	0.9000	0.7600	0.8700	0.7400

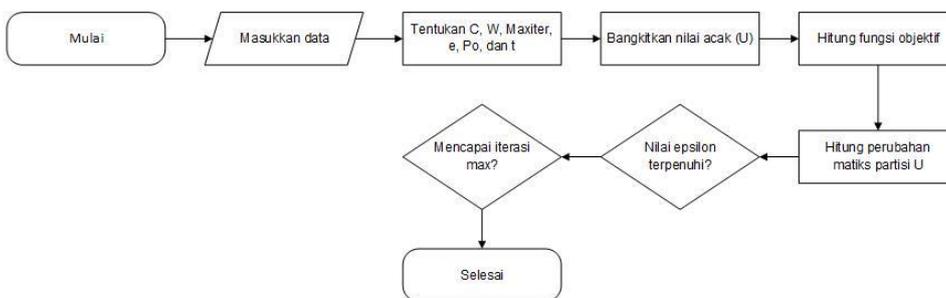
3.2. Perbandingan K-Means dan FCM

Pada penelitian ini, pengklasteran data *user knowledge modeling* menggunakan dua metode kluster, yaitu metode *K-Means* dan *Fuzzy C-Means*. Untuk lebih memahami perbedaan kedua algoritma tersebut, dapat dilihat dari *flowchart* algoritma *K-Means* berikut.



Gambar 1. Flowchart K-Means

Dalam pengklasteran data *user knowledge modeling* menggunakan metode *K-Means*, hal yang pertama dilakukan adalah memasukkan data *user knowledge modeling* terlebih dahulu. Setelah itu, tentukan jumlah kluster yang diharapkan. Lalu tentukan pula titik pusat kluster yang secara acak diambil dari data. Selanjutnya hitung jarak data ke pusat kluster. Setelah itu, kelompokkan data berdasarkan hasil minimum perhitungan jarak data kepusat kluster. Lalu ulangi lagi langkah awal untuk mengecek apakah titik pusat kluster yang telah dihasilkan sudah tepat dengan mengambil sebarang data dari data baru hasil dari perhitungan jarak data ke pusat kluster. Jika titik pusat kluster berubah maka kita ulangi lagi langkah-langkah sebelumnya sehingga titik pusat kluster tidak berubah [13].



Gambar 2. Flowchart FCM

Pada algoritma pengklasteran data menggunakan metode *Fuzzy C-Means*, sebagai langkah awal yang perlu dilakukan adalah memasukkan data yang akan diklaster dalam bentuk matriks $n \times m$. Lalu tentukan beberapa indikator yang diperlukan pada metode *Fuzzy C-Means*. Setelah itu bangkitkan bilangan *random*. Lalu, hitung pusat kluster data. Dari perhitungan pusat kluster, hitung fungsi objektif pada iterasi. Setelah itu, hitung perubahan matriks partisi. Lalu, cek kondisi berhenti dengan dilihat dari apakah nilai epsilon yang merupakan salah satu indikator telah terpenuhi atau tidak. Jika sudah terpenuhi maka iterasi selesai, jika iterasi telah maksimal maka kondisi berhenti [13].

Perbandingan dari metode *K-Means* dan *Fuzzy C-Means* tidak berhenti pada algoritma perhitungannya, akan tetapi perbandingannya terlihat ketika dihitung nilai validasi nya.

4. Hasil dan Pembahasan

Pada penelitian ini akan menjelaskan mengenai perbandingan pengklasteran data *user knowledge modeling* menggunakan metode *K-Means* dan *Fuzzy C-Means*. Dengan menggunakan *tools Matlab*, didapatkan hasil dari clustering dengan menggunakan metode *K-Means* dan *FCM*. Berikut adalah hasil kluster dari metode *K-Means* dan *FCM*.

4.1. K-Means

Dengan menggunakan *tools Matlab* dengan jumlah kluster yaitu 4, didapat hasil sebagai berikut:

Tabel 2. Tabel hasil *clustering* menggunakan metode *K-Means*

No	Klaster 1	Klaster 2	Klaster 3	Klaster 4	Hasil Klaster
1	0.6872	1.0951	1.0063	1.1838	1
2	0.7822	0.7920	0.5235	1.0162	3
3	0.4437	0.8345	0.6643	0.9008	1
4	0.3719	0.8375	0.7025	0.6502	1
5	0.6525	1.1022	0.9657	0.7574	1
...
258	1.1054	0.8629	0.9424	0.6903	4

Setelah dilakukan *clustering* pada data *user knowledge modeling* dengan menggunakan metode *K-Means* dan *tools* Matlab, maka didapatkan hasil klaster 1 berjumlah 75 anggota, klaster 2 berjumlah 67 anggota, klaster 3 berjumlah 55 anggota, klaster 4 berjumlah 61 anggota.

4.2. FCM

Dengan menggunakan *tools* Matlab, dan melakukan *input* pangkat (*w*) adalah 2, jumlah klaster (*k*) berjumlah 4 dan Maxiter berjumlah 50, didapatkan hasil klaster untuk data *user knowledge modeling* sebagai berikut:

Tabel 3. Tabel hasil *clustering* menggunakan metode *FCM*

No	Klaster 1	Klaster 2	Klaster 3	Klaster 4	Hasil Klaster
1	0.2367	0.2200	0.2616	0.2814	4
2	0.3340	0.2694	0.2042	0.1922	1
3	0.2464	0.2117	0.2599	0.2818	4
4	0.1709	0.1570	0.2938	0.3781	4
5	0.1798	0.1729	0.2917	0.3553	4
...
258	0.2471	0.2694	0.2540	0.2293	2

Setelah dilakukan *clustering* pada data *user knowledge modeling* dengan menggunakan metode *FCM* dan *tools* Matlab, maka didapatkan hasil klaster 1 berjumlah 66 anggota, klaster 2 berjumlah 66 anggota, klaster 3 berjumlah 41 anggota, klaster 4 berjumlah 85 anggota.

4.3. Validitas

4.3.1. Validasi *K-Means* dengan *Silhouette Index* (SI)

Untuk mendapatkan nilai validitas dari algoritma metode *K-Means*, maka digunakan validitas *Silhouette Index*. Dengan menggunakan *tools* Matlab didapatkan nilai validitas untuk algoritma *K-Means* yaitu 0.1866.

4.3.2. Validasi *Fuzzy C-Means* dengan *Partition Coefficient Index* (PCI)

Untuk menghitung nilai PCI, maka digunakan persamaan:

$$PCI = \frac{1}{N} (\sum_{i=1}^N \sum_{j=1}^K U_{ij}^2) \dots\dots\dots(7)$$

Nilai PCI menggunakan metode *FCM* adalah ai berikut:

a) Nilai Derajat keanggotaan

Tabel 4. Nilai derajat keanggotaan metode *FCM*

No	U1	U2	U3	U4
1	0.2367	0.2200	0.2616	0.2814
2	0.3340	0.2694	0.2042	0.1922
3	0.2464	0.2117	0.2599	0.2818

4	0.1709	0.1570	0.2938	0.3781
5	0.1798	0.1729	0.2917	0.3553
...
258	0.2471	0.2694	0.2540	0.2293

b) Validasi PCI(Ui²)

Tabel 5. Nilai validasi metode FCM

No	U1	U2	U3	U4
1	0.0560	0.0484	0.0684	0.0792
2	0.1115	0.0725	0.0417	0.0369
3	0.0607	0.0448	0.0675	0.0794
4	0.0292	0.0246	0.0863	0.1430
5	0.0323	0.0299	0.0851	0.1263
...
258	0.0610	0.0725	0.0645	0.0525

$$PCI = \frac{1}{258} (73.6476) = 0.2854$$

5. Kesimpulan

Dari pembahasan yang telah disampaikan, dapat disimpulkan bahwa hasil pengklasteran data *user knowledge modeling* menggunakan metode *K-Means* dan *Fuzzy C-Means* berbeda. Hal ini dapat dilihat dari jumlah kluster yang diperoleh dari kedua metode tersebut. Jika dilihat dari nilai validitas, nilai SI pada metode *K-Means* bernilai 0.1866, sedangkan nilai PCI pada metode *FCM* bernilai 0.2854. Dengan ini, metode *FCM* adalah metode yang lebih baik untuk melakukan *clustering* pada data *user knowledge modeling* dikarenakan nilai validitasnya lebih mendekati nilai 1.

Ucapan Terima Kasih

Terima kasih yang sebesar-besarnya kepada Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau atas fasilitas dan dukungan dari pimpinan. Demikian juga ucapan terima kasih kepada Tim Puzzle Research Data Technology (PREDATECH) Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau atas segala masukan, dukungan moral dan materi, koreksi dan bantuan dalam kegiatan penelitian ini sehingga penelitian ini dapat diselesaikan dengan baik.

Daftar Pustaka

- [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.
- [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 Belajar Mahasiswa*. Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 8. 2016.

- [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] Ekawati, Ratna & Yulis, Nurul. *Klasifikasi Usaha Kecil dan Menengah (UKM) Sektor Industri Dengan Metode Fuzzy C-Means Clustering Wilayah Kota Cilegon*. Seminar Nasional IENACO. 2013.
- [8] Jiawei, Han, dkk. *Data mining Concepts and Techniques*. USA: Elsevier Inc. All rights reserved. 2012.
- [9] Muzakir, Ari. *Analisa dan Pemanfaatan Algoritma K-Means Clustering Pada Data Nilai Siswa Sebagai Penentuan Penerima*. Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014.
- [10] Alfina, Tahta., Santosa, Budi., dan Barakbah, Ali Ridho. *Analisa Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data (Studi kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS)*. JURNAL TEKNIK ITS. 2012; Vol. 1, Pp 521-525.
- [11] Yusuf, Amad & Tjandrasa, Handayani. *Prediksi Nilai Dengan Metode Spectral Clustering dan Clusterwise Regression*. Jurnal Simantec. 2012; Vol. 4, No. 1.
- [12] Rizal, Annas Syaiful. *Metode K-Means Cluster Dan Fuzzy C-Means Cluster*. Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS 2015.
- [13] Febrianti, Fitria, dkk. *Perbandingan Pengklusteran Data Iris Menggunakan Metode K-Means Dan Fuzzy C-Means*. Jurnal Matematika "MANTIK". 2016; Vol. 02 No. 01.