



Estimasi Bobot Parameter M pada Fuzzy C-Means menggunakan Analisis Robust dengan Simulasi Data Spasial

Clinton Heskia Tambuwun¹, Yohanes A.R. Langi¹, Altien J. Rindengan^{1*}

¹Jurusan Matematika-Fakultas Matematika dan Ilmu Pengetahuan Alam-Universitas Sam Ratulangi Manado, Indonesia

*Corresponding Author : altien@unsrat.ac.id

ABSTRAK

Bobot parameter m (fuzzifier) dapat mempengaruhi performa dari fuzzy c-means (FCM). Pedoman-pedoman baru untuk memilih parameter m telah di sarankan dengan nilai dari m disarankan pada $m \in [1.5, 2.5]$ serta $m \in [1.5, 4]$. Sehingga akan diestimasi nilai m lainnya dengan data berbeda dan diaplikasikan pada data spasial. Hasil penelitian ini nilai parameter m pada data numerik $m \in [1.1, 2.894]$ dan nilai parameter m pada data spasial $m \in [1.1, 2.923]$.

INFO ARTIKEL

Diterima : 18 Januari 2020
Diterima setelah revisi : 24 Januari 2020
Tersedia online : 25 Januari 2020

Kata Kunci

Fuzzy C-Means
Parameter M
Analisis Robust
Data Spasial

ABSTRACT

The weighting exponent m (fuzzifier) can influence the performance of fuzzy c-means (FCM). New guidelines for choosing the parameter m have been suggested with the value of m suggested in $m \in [1.5, 2.5]$ and $m \in [1.5, 4]$. Therefore, the other m values will be estimated with different data and applied to spatial data. The results of this study, the value of the m parameter in numerical dataset is $m \in [1.1, 2.894]$ and the value of m parameter in spatial data is $m \in [1.1, 2.923]$.

ARTICLE INFO

Accepted : 18 January 2020
Accepted after revision: 24 January 2020
Available online: 25 January 2020

Keyword

Fuzzy C-Means
M Parameter
Robust Analysis
Spatial Data

1. PENDAHULUAN

Clustering data adalah suatu teknik yang bertujuan untuk pengelompokan sejumlah data atau objek kedalam *cluster* (kelompok) sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin, dan membuat jarak antar *cluster* sejauh mungkin. *fuzzy clustering* adalah proses menentukan derajat keanggotaan, dan kemudian menggunakannya dengan memasukannya ke dalam elemen data satu kelompok *cluster* atau lebih [1]. Ada banyak metode untuk *fuzzy clustering*, tetapi metode yang paling terkenal untuk *fuzzy clustering* adalah algoritma *fuzzy c-means* (FCM) karena pada metode FCM kemungkinan kegagalan untuk konvergen lebih kecil dibanding metode lainnya dan juga sudah banyak ekstensi yang telah diusulkan untuk memvalidasi partisi dari objek data yang dihasilkan oleh *fuzzy c-means*. *Fuzzy C-Means* (FCM) adalah suatu teknik pengelompokan data yang keberadaan tiap-tiap data

dalam suatu kelompok ditentukan oleh nilai atau derajat keanggotaan tertentu [2]. Telah diselidiki faktor penting lain yang mempengaruhi efektivitas *fuzzy c-means* yaitu eksponen pembobot m , dan disarankan $m \in [1.5, 2.5]$ [2]. Telah dianalisis pemilihan parameter untuk *fuzzy c-means* yang menjelaskan eksponen pembobot m (fuzzifier) dapat mempengaruhi kinerja *fuzzy c-means* (FCM) dan menyarankan parameter $m \in [1.5, 4]$ [3]. m adalah parameter fuzzy yang rata-rata derajat keaburan dari setiap data derajat keanggotaan lebih besar dari 1,0 [4]. Berdasarkan analisis robus FCM diatas, pedoman-pedoman baru untuk memilih parameter m telah di sarankan. Sehingga dalam penelitian ini akan diestimasi parameter m lainnya dengan menambahkan data dan juga akan di terapkan ke dalam data spasial. Penelitian ini memberikan alternatif parameter m pada FCM karena biasanya dalam program menggunakan $m = 2$ sebagai parameter standar.

2. Pengertian Clustering

Clustering adalah proses pengelompokan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas atau cluster [5]. Terdapat dua tahapan yang harus dilakukan dalam analisis cluster, yaitu [6]:

- Memutuskan apakah jumlah cluster ditentukan atau tidak, dan
- Menentukan algoritma yang akan digunakan dalam clustering. Untuk memutuskan berapa jumlah cluster yang akan dibentuk.

2.1 Pengertian Fuzzy Clustering

Menurut Bezdek, *Fuzzy Clustering* merupakan salah satu metode analisis cluster dengan mempertimbangkan tingkat keanggotaan yang mencakup himpunan fuzzy sebagai dasar pembobot bagi pengelompokan [7].

a. Ukuran Fuzzy

Ukuran fuzzy menunjukkan derajat keaburan dari himpunan fuzzy. Secara umum ukuran keaburan dapat dituliskan sebagai berikut:

$$f: P(X) \rightarrow R$$

Dengan P(X) adalah himpunan semua subset dari X dan f(A) adalah suatu fungsi yang memetakan subset A ke karakteristik derajat keaburannya. Dalam mengukur nilai keaburan, fungsi f harus mengikuti hal-hal sebagai berikut:

- $f(A) = 0$ jika dan hanya jika A adalah himpunan crisp.
- Jika $A < B$, maka $f(A) < f(B)$, dimana $A < B$ lebih kabur dibanding A dengan kata lain A lebih tajam dibandingkan B. Relasi ketajaman $A < B$ didefinisikan dengan:
 $[x] \leq [y]$, jika $\mu_B[x] \leq 0,5$;
 dan $\mu_A[x] \geq \mu_B[x]$, jika $\mu_B[x] \geq 0,5$
- $f(A)$ akan mencapai maksimum jika dan hanya jika A benar-benar kabur secara maksimum. Tergantung pada interpretasi derajat keaburan, nilai fuzzy maksimal biasanya terjadi pada saat $\mu_A[x] = 0,5$ untuk setiap x.

b. Indeks Keaburan

Indeks keaburan adalah jarak antara suatu himpunan fuzzy A dengan himpunan crisp C yang terdekat. Himpunan crisp C terdekat dari himpunan fuzzy A dinotasikan sebagai:

$$[x] = 0, \text{ jika } \mu_A[x] \leq 0,5; \text{ dan}$$

$$[x] = 1, \text{ jika } \mu_A[x] \geq 0,5$$

2.2 Pengertian Fuzzy C-Means

Fuzzy C-Means (FCM) adalah salah satu teknik peng-cluster-an data yang mana keberadaan tiap-tiap titik data dalam suatu cluster ditentukan oleh derajat keanggotaannya. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981 [6].

2.2.1 Algoritma Clustering Fuzzy C-Means dan Pemilihan Parameter

Algoritma fuzzy c-means (FCM) adalah metode pengelompokan fuzzy yang terkenal dengan fungsi objektif. Rumus fuzzy c-means:

$$J_{FCM}(\mu, a) = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m d(x_j, a_i) = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m \|x_j - a_i\|^2 \quad (1)$$

Eksponen pembobot $m > 1$ adalah indeks fuzzy (fuzzifier), $\mu = \{\mu_1, \dots, \mu_c\}$ dengan $\mu_{ij} = \mu_i(x_j)$ is a fuzzy c-partition, dan $a = \{a_1, \dots, a_c\}$ adalah bagian dari pusat cluster c. Kondisi yang diperlukan untuk minimizer (μ, a) dari J_{FCM} adalah persamaan pembaruan berikut:

$$\mu_{ij} = \frac{\|x_j - a_i\|^{-2/(m-1)}}{\sum_{k=1}^c \|x_j - a_k\|^{-2/(m-1)}} \quad (2)$$

dan

$$a_i = \frac{\sum_{j=1}^n \mu_{ij}^m x_j}{\sum_{j=1}^n \mu_{ij}^m} \quad (3)$$

Di mana $j = 1, \dots, n, i = 1, \dots, c$. Perhatikan bahwa $d(x_j, a_i) = \|x_j - a_i\|^2$ adalah yang paling banyak digunakan. Namun, tipe lain dari matriks $d(x_j, a_i)$ dapat digunakan untuk meningkatkan penggunaan dan efektifitas fuzzy c-means [2]. Ingat bahwa pendekatan yang disatukan secara luas yang terlalu menggeneralisasi FCM telah di pelajari [8]. Di sisi lain, faktor penting yang mempengaruhi efektifitas FCM adalah parameter pembobot m , yang sebelumnya telah diselidiki secara menyeluruh oleh Pal & Bezdek dan Yu beserta kawan-kawan. Eksponen pembobot m disebut indeks fuzzy (fuzzifier); itu dapat mempengaruhi kinerja pengelompokan FCM. Yu dkk telah mengusulkan batas atas yang dapat mencegah rata-rata sampel sehingga menjadi pengoptimal yang unik (unique optimizer) fungsi tujuan fuzzy c-means. Aturannya adalah bahwa $\forall i, a_i = \bar{x}$ adalah stabil untuk FCM jika $\lambda_{max}(C_x) < 0.5$ dan $m \geq (1 - 2\lambda_{max}(C_x))^{-1}$, Dimana $C_x = \frac{\sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x})^T}{n \|x_j - \bar{x}\|^2}$ dan $\lambda_{max}(C_x)$ adalah nilai eigen maksimum dari matriks C_x . Sehingga, untuk fuzzy c-means, kita harus menetapkan $m < (1 - 2\lambda_{max}(C_x))^{-1}$ jika $\lambda_{max}(C_x) < 0.5$.

Jika $\lambda_{max}(C_x) \geq 0.5$, kita dapat menganggap m sebagai nilai positif; lebih lanjut, Pal & Bezdek menyaranakan $m \in [1.5, 2.5]$.

Tabel 1 menunjukkan batas atas m untuk beberapa dataset yang di ambil dari UCI Repository of Machine Learning Databases. Untuk dataset $\lambda_{max}(C_x) < 0.5$, rata-rata sampel akan menjadi unique optimizer of FCM ketika fuzzifier $m \geq (1 - 2\lambda_{max}(C_x))^{-1}$ (batas atas). Pada dataset Iris, waktu nilai eigen maksimum dari matriks C_x lebih besar dari 0,5, maka batas atas m untuk dataset Iris adalah tak berhingga positif. Ini adalah hasil dari Yang dan Wu [8]

Table 1. Batas atas m untuk dataset numerik

Dataset	Jumlah sample	Jumlah features	Jumlah Cluster	Batas atas m
Vowel	990	10	11	1.7787
Waveform	5000	21	3	2.8935
Glass	214	9	6	3.1726
Iris	150	4	3	∞

Table 2. Mean square error dan jumlah 100 iterasi dari normal dataset yang dihasilkan secara acak

	MSE		Jumlah Iterasi			
	Kasus terbaik	Rata-rata	Kasus terbaik	Kasus terbaik	Rata-rata	Kasus terbaik
$m = 1,5$	0.0579	0.145	0.332	10	14.93	23
	1	54	01			
$m = 2$	0.1615	0.459	0.910	13	20.20	35
	3	12	60			
$m = 2,5$	3.8775	5.956	6.5217	50	50.00	50
	4	04	4			
$m = 3$	6.6752	6.748	6.796	36	42.24	50
	9	13	83			
$m = 3,5$	6.7391	6.754	6.7851	27	30.80	36
	6	95	3			

Algoritma fuzzy c-means:

- Input data yang akan di-cluster X, berupa matriks berukuran $n \times p$ (n =jumlah sampel data, p =atribut

setiap data). X_{kj} = data sampel ke- k ($k = 1, 2, \dots, n$), atribut ke- j ($j = 1, 2, 3, \dots, m$).

2. Menentukan:
 - a) Jumlah cluster = c ;
 - b) Pangkat pembobot = m ;
 - c) Maksimum iterasi = $MaxIter$;
 - d) Error terkecil yang diharapkan = ξ ;
 - e) Fungsi Objektif awal = $PO = 0$;
 - f) Iterasi awal = $t = 1, 2$.
3. Bangkitkan bilangan random ($\mu_{ik}, i = 1, 2, \dots, c; k = 1, 2, \dots, n$), sebagai elemen-elemen matriks partisi awal U

$$U_0 = \begin{bmatrix} \mu_{11}(x_1) & \mu_{12}(x_2) & \dots & \mu_{1c}(x_c) \\ \vdots & \vdots & & \vdots \\ \mu_{11}(x_1) & \mu_{12}(x_2) & \dots & \mu_{nc}(x_c) \end{bmatrix} \quad (4)$$

Matriks partisi pada fuzzy clustering harus memenuhi kondisi sebagai berikut:

$$\begin{aligned} \mu_{ik} &= [0, 1]; (1 \leq i \leq c; 1 \leq k \leq n) \\ \sum_{i=1}^n \mu_{ik} &= 1; 1 \leq i \leq c \\ 0 < \sum_{i=1}^c \mu_{ik} &< c; 1 \leq k \leq n \end{aligned} \quad (5)$$

Hitung jumlah setiap kolom (atribut):

$$Q_j = \sum_{i=1}^c (\mu_{ik}) \quad (6)$$

dengan $j = 1, 2, 3, \dots, m$ kemudian hitung:

$$\mu_{ik} = \frac{\mu_{ik}}{Q_j} \quad (7)$$

4. Hitunglah pusat cluster ke- k : V_{ij} , dimana $i = 1, 2, 3, \dots, c$ dan $j = 1, 2, 3, \dots, m$

$$V_{ij} = \frac{\sum_{k=1}^n ((\mu_{ik})^m * X_{kj})}{\sum_{k=1}^n (\mu_{ik})^m} \quad (8)$$

$$V = \begin{bmatrix} v_{11} & \dots & v_{1m} \\ \vdots & \ddots & \vdots \\ v_{c1} & \dots & v_{cm} \end{bmatrix}$$

5. Hitung fungsi objektif pada iterasi ke- t , P_t dengan menggunakan persamaan sebagai berikut:

$$P_t = \sum_{k=1}^n \sum_{i=1}^c \left(\left[\sum_{j=1}^m (X_{kj} - V_{ij})^2 \right] (\mu_{ik})^m \right) \quad (9)$$

6. Hitung perubahan matriks partisi:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{kj} - V_{ij})^2]^{-\frac{1}{p-1}}}{\sum_{i=1}^c [\sum_{j=1}^m (X_{kj} - V_{ij})^2]^{-\frac{1}{p-1}}} \quad (10)$$

7. Cek kondisi berhenti:

- a) Jika ($|P_t - P_{t-1}| < \xi$) atau ($t < \text{iterasi maksimal}$) maka berhenti;
- b) Jika tidak: maka $t = t + 1$ kemudian ulang langkah ke-4.

2.2.2 Analisis Robust Dari Fuzzy C-Means

Pengaruh penyembuhan dapat membantu kita menilai pengaruh relative dari pengamatan individu terhadap nilai estimasi. Fungsi pengaruh M-estimator sebanding dengan fungsi φ nya [3]. Jika pengaruh fungsi estimasi tidak terikat, outlier dapat menyebabkan masalah ketika fungsi φ digunakan untuk menunjukkan tingkat pengaruh. Biarkan titik data x_j hilang dan pusat cluster a_i akan menjadi:

$$\rho_{FCM}(x_j - a_i) = \mu_{ij}^m \|x_j - a_i\|^2 \quad (11)$$

dan,

$$\varphi_{FCM}(x_j - a_i) = \frac{d}{da_i} \rho(x_j - a_i) = -2\mu_{ij}^m (x_j - a_i) \quad (12)$$

Dengan menyelesaikan persamaan $\sum_{j=1}^n \varphi_{FCM}(x_j - a_i) = 0$, kita memperoleh hasilnya di persamaan 3. Oleh karena itu, estimasi pusat cluster FCM adalah M-estimator dengan fungsi kehilangan (11) dan fungsi φ (12).

Perhatikan bahwa persamaan pusat cluster terbaru (3) dapat dilihat sebagai rata-rata tertimbang untuk dataset $\{x_1, \dots, x_n\}$. Bobot dari titik data x_j adalah

$$\frac{\mu_{ij}^m}{\sum_{j=1}^n \mu_{ij}^m} \quad (13)$$

Yang sebanding dengan μ_{ij}^m . Untuk noise atau outlier x_j , beratnya sebanding dengan $(1/c)^m$. Dalam kasus besar dari m , berat akan cenderung ke-0 dan karenanya, pengaruh kebisingan dan outlier pada pusat-pusat cluster akan menjadi sangat kecil. Ini bias menjelaskan mengapa fcm kuat terhadap noise dan outlier ketika m menjadi besar.

Ketika m cenderung ke tak hingga, bobot dari fungsi φ dari titik hingga x_j akan cenderung ke 0. Namun, dalam aplikasi nyata, kami tidak pernah mengamati titik data tak terhingga. Kami kemudian memiliki $\lim_{n \rightarrow \infty} \varphi(x_j - a_i) = 0$ untuk real titik data. Itu adalah nilai m yang sangat besar yang akan membuat FCM sangat kuat. Namun, ini bukan panduan yang baik untuk memilih m di FCM. Meskipun FCM menjadi sangat kuat dalam kasus besar, nilai keanggotaan untuk setiap titik data akan sangat dekat ke $1/c$, dan sampel rata-rata dapat menjadi pengotimal yang unik [3].

2.3 Data Spatial

Sebagian besar data yang akan ditangani dalam SIG merupakan data spasial yaitu sebuah data yang berorientasi geografis, memiliki sistem koordinat tertentu sebagai dasar referensinya dan mempunyai dua bagian penting yang membuatnya berbeda dari data lain, yaitu informasi lokasi (spasial) dan informasi deskriptif (atribut) yang dijelaskan berikut ini [9]:

2.3.1 Informasi Lokasi

Informasi lokasi (spasial), berkaitan dengan suatu koordinat baik koordinat geografi (lintang dan bujur) dan koordinat XYZ, termasuk di antaranya informasi datum dan proyeksi. Informasi lokasi atau geometri milik suatu objek spasial dapat dimasukkan ke dalam beberapa bentuk seperti berikut :

- a. Titik
- b. Garis (satu dimensi – line atau polyline)
- c. Polygon (dua dimensi – area)
- d. Permukaan (3D)

2.3.2 Informasi Atribut

Informasi deskriptif (atribut) atau informasi non spasial, suatu lokasi memiliki beberapa keterangan yang berkaitan dengannya, Format tabel kata-kata, kode alfanumerik, angka-angka.

- a. Format laporan Teks, deskripsi.
- b. Format pengukuran Angka-angka, hasil.
- c. Format grafik anotasi Kata-kata, angka-angka, simbol.

2.3.3 Format Data Spasial

Secara sederhana format dalam bahasa komputer berarti bentuk dan kode penyimpanan data yang berbeda antara file satu dengan lainnya. Dalam SIG, data spasial dapat direpresentasikan dalam dua format, yaitu :

- a. Data Vektor
- b. Data raster

2.3.4 Sumber Data Spasial

Salah satu syarat SIG adalah data spasial, yang dapat diperoleh dari beberapa sumber antara lain :

- a. Peta Analog
- b. Data Sistem Penginderaan Jauh
- c. Data hasil pengukuran lapangan
- d. Data GPS (Global Positioning System)

3. METODOLOGI PENELITIAN

3.1 Jenis Data dan Sumber Data

Penelitian ini menggunakan data numerik dan data spasial. Untuk data numerik di gunakan data *Balance Scale*, *Sonar*, *Waveform*, *Vowel*. Dataset numerik ini di ambil dari *University of California Irvine Machine Learning Repository* atau Gudang data dari *Universitas California Irvine* yang adalah kumpulan database, teori domain, dan generator data yang digunakan oleh komunitas pelajar mesin untuk analisis empiris algoritma pembelajaran mesin. Arsip ini dibuat pada tahun 1987 oleh David Aha dan sesama mahasiswa pascasarjana di *UC Irvine*. Sejak saat itu, telah banyak digunakan oleh siswa, pendidik, dan peneliti di seluruh dunia sebagai sumber utama set data pembelajaran mesin. Untuk data spasial menggunakan data *Atlanta* dan data *Police*. Dataset spasial ini diambil dari *The Center for Spatial Data Science The University of Chicago* atau pusat data spasial dari universitas *Chicago*.

3.2 Metode Penelitian

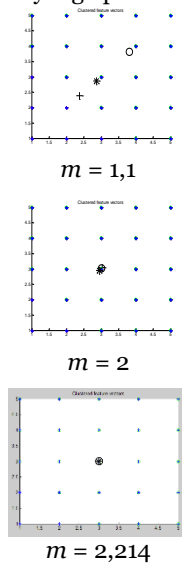
Dengan menggunakan Matlab dan Algoritma Fuzzy C-Means, penelitian ini dilakukan dengan dua step. Pertama, akan di estimasi interval dari parameter m pada data numerik dan data spasial. Parameter m akan di estimasi mulai dari angka 1,1 yang sesuai teori adalah batas bawah dari parameter m . Jika di mulai < 1,1 akan terjadi error pada program.

4. HASIL DAN PEMBAHASAN

4.1 Clustering Fuzzy C-Means pada data numerik

1. Balance Scale

Balance scale adalah dataset hasil eksperimen psikologi yang terdiri dari 3 *cluster* (kelompok) dan 208 *sample* data yang di ambil oleh Siegler R.S (1976). Akan di cari nilai m yang optimal.



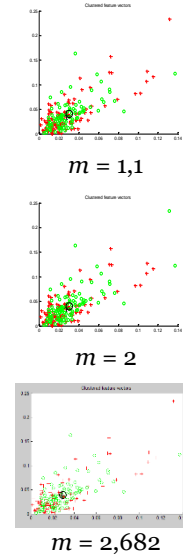
Gambar 8. Clustering *Balance Scale* dataset

Gambar 8 adalah hasil *clustering* fuzzy c-means untuk dataset *balance scale* dengan menggunakan matlab dan algoritma fuzzy c-means (lampiran 1) mendapatkan nilai m robust (kekar) di nilai 2,214.

Terlihat ketika $m = 1,1$ ketiga pusat *cluster* masih dalam keadaan semula tetapi ketika $m = 2,214$ ketiga pusat *cluster* ini menyatu.

2. Sonar

Sonar adalah dataset hasil eksperimen pemantulan sinyal sonar dari silinder logam di berbagai sudut dan kondisi yang di ambil dari koleksi Terry Sejnowski. Akan di cari nilai m yang *optimum*. *Sonar* terdiri dari 3 *cluster*.

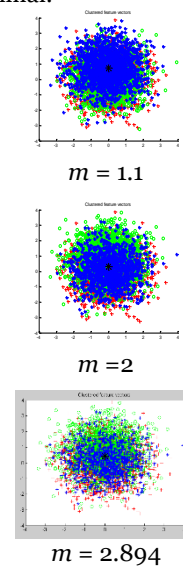


Gambar 9. Clustering *Sonar* dataset

Gambar 9 adalah hasil *clustering* fuzzy c-means untuk dataset *sonar* dengan menggunakan matlab dan algoritma fuzzy c-means (lampiran 1) mendapatkan nilai m robust (kekar) di nilai 2,682. Terlihat ketika $m = 1,1$ kedua pusat *cluster* masih dalam keadaan semula tetapi ketika $m = 2,682$ kedua pusat *cluster* ini menyatu.

3. Waveform

Waveform adalah dataset ombak yang terdiri dari 3 *cluster* yang di ambil dari Breiman L. Akan di cari nilai m yang optimal.



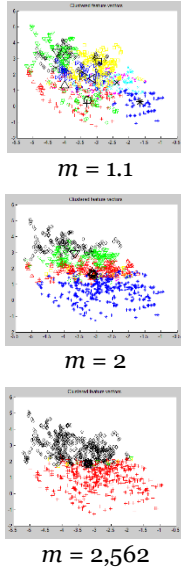
Gambar 10. Clustering *Waveform* dataset

Gambar 10 adalah hasil *clustering* fuzzy c-means untuk dataset *waveform* dengan menggunakan matlab dan algoritma fuzzy c-means (lampiran 1) mendapatkan nilai m robust (kekar) di nilai 2,894. Terlihat ketika $m = 1,1$ ketiga pusat *cluster* masih

dalam keadaan semula tetapi ketika $m = 2,894$ ketiga pusat *cluster* ini menyatu.

4. Vowel

Vowel adalah dataset sebelas *vocal* Bahasa Inggris British state. Data *vocal* ini diindeks dengan bilangan bulat 0-10. Untuk setiap ucapan, ada sepuluh nilai input *floating-point*, dengan indeks array 0-9. Data ini di teliti oleh Tony Robinson. Dataset *vowel* ini terdiri dari 11 *cluster*. Akan di cari nilai m yang *optimum*.



Gambar 11. Clustering *vowel* dataset

Gambar 11 adalah hasil clustering fuzzy *c-means* untuk dataset *vowel* dengan menggunakan matlab dan algoritma fuzzy *c-means* (lampiran 1) mendapatkan nilai m *robust* (kekar) di nilai 2,562. Terlihat ketika $m = 1,1$ kesebelas pusat *cluster* masih dalam keadaan semula tetapi ketika $m = 2,562$ kesebelas pusat *cluster* ini menyatu.

Tabel 3. Hasil Clustering FCM Data Numerik

Dataset	Jumlah sample	Jumlah features	Jumlah Cluster	Batas atas m
Balance Scale	625	4	3	2.214
Sonar	208	60	2	2.682
Waveform	5000	21	3	2.894
Vowel	990	10	11	2.562

Berdasarkan table 3 di atas, dilihat bahwa batas atas m yang diperoleh adalah 2,894 pada data *vowel*. Dengan demikian diperoleh interval parameter $m \in [1.1, 2.894]$.

4.2 Clustering Fuzzy C-Means pada data spasial

1. Atlanta

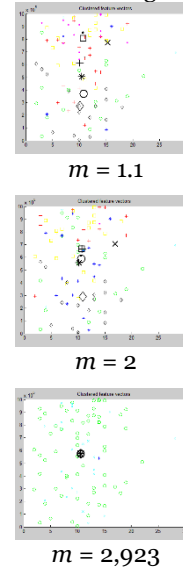
Atlanta adalah dataset spasial *Homicides and selected socio-economic characteristics for counties surrounding Atlanta, GA*. *Atlanta* memiliki jumlah *cluster* 7, 14 jumlah variable, dan 90 data orang yang diteliti.



Gambar 12. Data Spasial *Atlanta*

Gambar 12 disamping adalah pemetaan 90 data orang yang diteliti dari data *Atlanta*. Akan di cari nilai m yang optimum tetapi yang diambil hanya data numeriknya saja yang di convert ke format .txt (bisa dilihat di lampiran).

Berikut adalah hasil clustering data spasial *atlanta*.

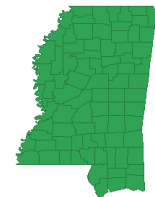


Gambar 13. Clustering spasial dataset *Atlanta*

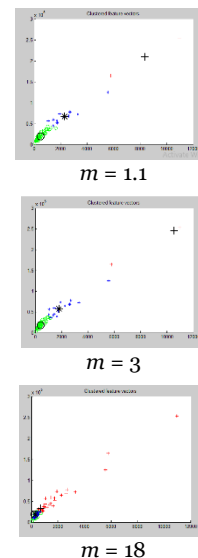
Gambar 13 adalah hasil clustering fuzzy *c-means* untuk dataset spasial *Atlanta* dengan menggunakan matlab dan algoritma fuzzy *c-means* (lampiran 1) mendapatkan nilai m *robust* (kekar) di nilai 2,923. Terlihat ketika $m = 1,1$ ketujuh pusat *cluster* masih dalam keadaan semula tetapi ketika $m = 2,923$ ketujuh pusat *cluster* ini menyatu.

2. Police

Police adalah dataset spasial *Police expenditures Mississippi counties*. *Police* memiliki jumlah *cluster* 3, 10 jumlah variabel dan 82 area yang diteliti. Gambar 14 disamping adalah pemetaan 82 area yang diteliti. Akan di cari nilai m yang optimal tetapi yang diambil hanya data numeriknya saja yang di convert ke format .txt (bisa dilihat di lampiran).



Gambar 14. Data Spasial *Police*



Gambar 15. Clustering spasial dataset *Police*

Gambar 15 adalah hasil *clustering* fuzzy c-means untuk dataset spasial Police dengan menggunakan matlab dan algoritma fuzzy c-means (lampiran 1) mendapatkan nilai *m robust* (kekar) di nilai $m = 18$. Tapi pusat cluster tidak menjadi satu *cluster* dan akan tetap seperti itu meskipun di input m sampai tak hingga. Ini adalah contoh data yang jelek.

Tabel 4. Hasil Data Spasial

No	Data Spasial	Jumlah Cluster	Jumlah variable	Interval
1	Atlanta	7	14	[1.1, 2.923]
2	Police	3	10	[1.1, ∞]

Berdasarkan table 4 di atas, dilihat bahwa batas atas m yang diperoleh adalah 2,923 pada data *Atlanta* dengan data *Police* di buang karena terlihat pusat *cluster* yang tidak menyatu. Dengan demikian diperoleh interval parameter $m \in [1.1, 2.923]$.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang diperoleh maka dapat disimpulkan bahwa sesuai dengan hasil yang diperoleh yaitu

1. Data numerik dengan data *Balance Scale*, *Sonar*, *Waveform*, dan *Vowel* di peroleh nilai parameter $m \in [1.1, 2.894]$.
2. Data spasial dengan data *Atlanta* dan *Police* diperoleh nilai parameter $m \in [1.1, 2.923]$. Dengan data *Police* menunjukkan data yang pusat *clusternya* tidak akan menjadi satu meskipun nilai m sampai tak hingga.

5.2 Saran

1. Menggunakan teknik *clustering* lanjutan untuk mencapai akurasi yang lebih efisien dalam hasil
2. Menambah data spasial untuk mencapai akurasi yang lebih efisien.

REFERENSI

- [1] Kusumadewi, S., dan H. Purnomo. 2004. Aplikasi Logika Fuzzy untuk pendukung keputusan. Graha Ilmu, Yogyakarta
- [2] Pal, N.R., dan J.C. Bezdek. 1995. On cluster validity for fuzzy c-means model. *IEEE Trans on Fuzzy System Journal* 3(3):370-379.
- [3] Wu, K. 2011. *Analysis of parameter selections for fuzzy c-means*. Department of Information Management, Kun Shan University, Taiwan.
- [4] Ravichandran. 2009. *Hybrid Fuzzy C-Means Clustering Technique for Gene Expression Data*. Dept of Computer Science and Engineering Hindustan Institute of Tech, India.
- [5] Agusta, Y. 2015. *Clustering*. <https://yudiagusta.wordpress.com/clustering>. Diakses tanggal 8 Desember 2019.
- [6] Kusumadewi, S. 2006. *Fuzzy Multi-Attribute Decision Making*. Graha Ilmu, Yogyakarta.
- [7] Efiyah, U. 2014. Penerapan Algoritma Fuzzy C-Means untuk Pengelompokan Harga Gabah di Tingkat Penggilingan Berdasarkan Kualitas Gabah. Universitas Islam Negeri Maulana Malik Ibrahim, Malang.
- [8] Yu, J. Cheng., dan H. Q. Huang. 2004. *Analysis of the weighting exponent in the FCM*. Dept of Computer Science & Technol Northern Jiaotong Univ, Beijing.

- [9] Prahasta, E. 2009. Sistem Informasi Geografis: Konsep-konsep dasar (Perspektif Geodesi & Geomatika). Informatika, Bandung.
- [10] Gan, G., C. Ma., dan J. Wu. 2007. *Data Clustering*. The America Statistic Assocation, United States of America.
- [11] Huber, P.J. 1981. *Robust Statistics*. Wiley, Turkey.
- [12] Suganya, R., dan R. Shanthi. 2012. *Fuzzy C-Means Algorithm – A review*. Department of CS, Dr.SNS.Rajalakshmi College of Arts & Science, India.
- [13] Weku, W., dan M. Yang. 2017. *Robust GK Fuzzy clustering algorithms with cluster coreand analysis on parameter selection*. Chung Yuan Christian University, Taiwan.
- [14] Chaomurilige, Yu., dan Yang. 2015. *Analysis of parameter selection for Gustafson-Kessel fuzzy clustering using jacobian Matrix*. National Science Foundation Council and Ministry of Education of China, Taiwan.



Clinton H. Tambuwun

(tambuwunclinton@gmail.com)

Lahir di Motoling, 4 Juli 1995. Menempuh pendidikan tinggi Jurusan Matematika, FMIPA, Universitas Sam Ratulangi Manado. Tahun 2020 adalah tahun terakhir ia menempuh studi. Makalah ini merupakan hasil penelitian skripsinya yang dipublikasikan.



Altien J. Rindengan

(Altien@unsrat.ac.id)

Lahir pada tanggal 27 April 1974 . Memperoleh gelar Magister Komputer (M.Kom) dari Institut Pertanian Bogor. Menjadi pengajar tetap di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sam Ratulangi Manado.



Yohanes A. R. Langi

(varlangi@gmail.com)

Lahir pada tanggal 13 Juni 1970. Memperoleh gelar Magister Sains (M.Si) dari Institut Pertanian Bogor. Menjadi pengajar tetap di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sam Ratulangi Manado.