

Perbandingan Algoritma *Fuzzy C-Means (FCM)* Dan Algoritma *Mixture* Dalam Penclusteringan Data Curah Hujan Kota Bengkulu

Herlina Latipa Sari
Dosen Tetap Program Studi Teknik Informatika
Universitas Dehasen Bengkulu
Email : herlinalatipasari@ymail.com

Dewi Suranti
Dosen Tetap Program Studi Teknik Informatika
Universitas Dehasen Bengkulu
Email : Mrs.Dewisuranti@gmail.com

Abstrak—Penelitian ini dilakukan untuk mendesain *Fuzzy Clustering* menggunakan algoritma *C-Means* dan algoritma *Mixture* dalam penclusteringan data curah hujan Kota Bengkulu, membandingkan algoritma *C-Means* dan Algoritma *Mixture* dalam menghasilkan performansi algoritma *C-Means* dan algoritma *Mixture* dalam menghasilkan tingkat keakuratan lokasi perkiraan curah hujan bulanan stasiun Klimatologi Pulau Baii Bengkulu. Data curah hujan Kota Bengkulu dianalisis menggunakan algoritma software matlab. Dari hasil pengujian menggunakan software matlab dibantu menggunakan software SOCR diperoleh hasil pengujian dengan menggunakan *Fuzzy C-Means* karena pengelompokan datanya berdasarkan dengan derajat keanggotaan sehingga pusat cluster yang dihasilkan dalam mencapai fungsi sasaran mencari solusi terbaik untuk cluster-cluster dapat dilakukan lebih cepat ditunjukkan dengan proses penghentian iterasi. Output dari *fuzzy C-means* bukan merupakan *Fuzzy Inference System*, namun merupakan deretan cluster dan beberapa derajat keanggotaan untuk tiap titik data. Informasi ini dapat digunakan untuk membangun suatu *fuzzy inference system*.

Kata Kunci—*Algoritma Fuzzy C-Means; Algoritma Mixture.*

I. PENDAHULUAN

Konvensi Internasional di seluruh dunia menyatakan bahwa curah hujan mempunyai peran yang sangat penting, untuk mendukung sektor penerbangan, diberikan layanan jasa meteorologi untuk meningkatkan keamanan dan keselamatan penerbangan. Indonesia juga telah menerapkan hal ini diantaranya dalam peraturan Pemerintah No. 3 tahun 2001 tentang keselamatan penerbangan.

Berdasarkan data prakiraan curah hujan tahun 2014 pada Kota Bengkulu yang dilakukan pada empat pos pengamatan yang digunakan dalam proses prakiraan curah hujan yang terjadi dari masing-masing pos pengamatan didapat prakiraan curah hujan dengan rata-rata 200mm - 400mm masuk dalam pengelompokan sifat prakiraan curah hujan normal.

Stasiun Klimatologi Pulau Baii Bengkulu adalah salah satu Unit Pelaksana Teknis Badan Meteorologi Klimatologi dan Geofisika yang mempunyai tugas pokok untuk melaksanakan pengamatan, pengolahan, dan penyebaran data unsur-unsur cuaca / iklim (angin, hujan, suhu, tekanan udara, *visibility* dan lainnya) sehingga data curah hujan dapat dikelompokkan disesuaikan dengan kebutuhan masyarakat.

Dari Pengelompokan data curah hujan dalam satu tahun atau 12 bulan sesuai dengan kebutuhan pengelompokan data

berdasarkan sifat hujan, dimana sifat hujan dibagi menjadi 3 (tiga) kategori yaitu Atas Normal (AN) jika nilai curah hujan lebih dari 115% terhadap rata-ratanya, Normal (N) jika nilai curah hujan antara 85%-115% terhadap rata-ratanya dan Bawah Normal (BN) jika nilai curah hujan kurang dari 85% terhadap rata-ratanya. Sehingga curah hujan yang terekam pada empat pos pengamatan kota Bengkulu akan dikelompok berdasarkan sifat hujan dengan menggunakan *Algoritma Fuzzy C-Means* dan *Algoritma Mixture*.

Metode *fuzzy clustering*, telah banyak diaplikasikan untuk mengelompokkan suatu data berdasarkan kesamaan/kemiripan yang dimiliki oleh suatu wilayah. Terdapat berbagai macam teknik *fuzzy clustering* misalnya *Fuzzy Cluster-Means (FCM)* dan *Mixture Modelling*. Untuk *Fuzzy C-Means (FCM)* ini adalah untuk meminimalisasikan *objective function* yang disertakan dalam proses *clustering*, yang ada pada umumnya berusaha meminimalisasikan variasi di dalam suatu cluster dan memaksimalkan variasi antar cluster. *Mixture Modelling* merupakan salah satu jenis data *clustering* dimana dalam pemodelannya, data dalam suatu kelompok diasumsikan terdistribusi sesuai dengan salah satu jenis distribusi statistik yang ada.

Beberapa contoh pemanfaatan Algoritma *Clustering* diantaranya oleh The How Liong dan PM Siregar (2003) dalam peranan pengelompokan samar dalam prediksi kekeringan di Indonesia, oleh Muhamad Aqil, Firmansyah, Abi Prabowo dan Moses (2000) memanfaatkan *klustering tingkat pemakaian pompa air tanah menggunakan model fuzzy clustering*, Susanto dan Ernawati (2006) juga memanfaatkan pembagian kelas peserta kuliah berdasarkan *fuzzy clustering* dan [9] *fuzzy C-Means* untuk *clustering data (Studi kasus : data performance mengajar dosen)*.

Menurut [1], Tujuan penggunaan algoritma *C-Means* dan algoritma *Mixture* adalah untuk langkah awal mengelompokkan record-record data yang dianalisis sehingga terkelompok ke dalam interval-interval kelas yang lebih sedikit yang diharapkan dapat mempertinggi tingkat akurasi yang dihasilkan. Dari data yang dihasilkan dari kedua algoritma ini nantinya akan dibandingkan sehingga dapat disimpulkan dari kedua algoritma ini akan dapat satu algoritma dimana hasilnya nantinya mendekati akurasi data. Oleh karenanya akan sangat menarik bagaimana menggunakan algoritma *C-Means* dan algoritma *Mixture* sehingga nantinya dapat dijadikan sebagai salah satu teknik dalam pengclusteringan data curah hujan.

Perumusan masalah dari penelitian ini adalah sebagai berikut :

1. Bagaimana menerapkan algoritma *C-Means* dalam pengclusteran data curah hujan ?
2. Bagaimana menerapkan algoritma *Mixture* dalam pengclusteran curah hujan ?
3. Membandingkan cluster data curah hujan dengan Algoritma *C-Means* dan Algoritma *Mixture* ?
4. Bagaimana menentukan algoritma yang sesuai digunakan untuk pengclusteran berdasarkan data curah hujan?

II. FUZZY CLUSTERING

Kemunculan *fuzzy clustering* dilatarbelakangi adanya masalah *curse of dimensionality*, yaitu jumlah rule yang begitu cepat membesar dengan bertambahnya jumlah variabel input FIS. Jumlah rule yang terlalu besar jelas akan membuat komputasi menjadi berat dan juga optimasi parameter-parameter rule menjadi lebih sulit.

Dengan *fuzzy clustering*, suatu data input output akan dikelompokkan dalam beberapa grup atau *cluster*. Informasi *cluster* ini akan membantu dalam FIS tipe Sugeno terbaik yang bisa memodelkan kelakuan hubungan data input-output dengan jumlah rule *minimum*. Definisi sebuah rule diasosiasikan dengan suatu *cluster* data. Pembangunan FIS demikian bisa secara otomatis dilakukan menggunakan fungsi *genfis2*[7].

Fuzzy clustering adalah salah satu teknik untuk menentukan cluster optimal dalam suatu ruang vektor yang didasarkan pada bentuk normal Euclidian untuk jarak antar vektor. Fuzzy clustering sangat berguna bagi pemodelan fuzzy terutama dalam mengidentifikasi aturan-aturan fuzzy.

Untuk pengelompokkan para pengambil keputusan menjadi kelompok-kelompok kecil, berdasarkan persamaan karakteristik, dibutuhkan suatu mekanisme tertentu. Pada proses pengclusteran (*clustering*) secara klasik, pembentukan partisi dilakukan sedemikian rupa sehingga setiap objek berada tepat pada satu partisi. Namun, adakalanya kita tidak dapat menempatkan suatu objek tepat pada satu partisi, karena sebenarnya objek tersebut terletak diantara dua atau lebih partisi yang lain. Pada logika fuzzy, ada beberapa metode yang dapat digunakan untuk melakukan pengelompokkan sejumlah data yang sering dikenal dengan nama *fuzzy clustering*. Pada kebanyakan situasi, *fuzzy clustering*, lebih alami jika dibandingkan dengan pengclusteran secara klasik. Suatu algoritma clustering dikatakan sebagai algoritma *fuzzy clustering* jika dan hanya jika algoritma tersebut menggunakan parameter strategi adaptasi secara *soft competitive* (Baraldi, 1998). Sebagian besar algoritma fuzzy clustering didasarkan atas optimasi fungsi obyektif atau midifikasi dari fungsi obyektif tersebut.

Pemilihan algoritma clustering yang tepat, sangatlah penting demi suksesnya proses clustering. Secara umum, algoritma pengclusteran dicirikan berdasarkan ukuran kedekatan dan kriteria pengclusteran (Vazirgiannis,2003).

Ukuran kedekatan menunjukkan seberapa dekat kedekatan fitur antara dua data; sedangkan kriteria pengclusteran biasanya diekspresikan dengan menggunakan fungsi biaya atau tipe aturan yang lainnya.

a. Partisi Klasik (*hard partition*)

Konsep partisi menjadi bagian yang sangat penting bagi proses pengclusteran. Tujuan proses pengclusteran pada partisi klasik adalah membagi himpunan data ke x ke dalam c kelompok (grup atau kelas) dengan asumsi bahwa c diketahui (Babuska, 2005). Dengan menggunakan teori himpunan klasik, partisi klasik x dapat didefinisikan sebagai suatu keluarga dari himpunan bagian- himpunan bagian $\{A_i | 1 \leq i \leq c\}$ $P(X), P(X)$ adalah *power set* dari X , dengan properti sebagai berikut (Bezdek,1981):

$$\bigcup_{i=1}^c A_i = X \dots\dots\dots (1)$$

$$A_i \cap A_j = \emptyset ; 1 \leq i \neq j \leq c \dots\dots\dots (2)$$

$$A_j \cap X = A_j ; 1 \leq j \leq c \dots\dots\dots (3)$$

Persamaan menunjukkan bahwa union himpunan bagian A_i berisi semua data. Himpunan bagian – himpunan bagian harus bersifat disjoint (persamaan 2), dan tidak boleh ada yang berupa himpunan kosong (persamaan 3). Dalam bentuk fungsi keanggotaan, suatu partisi dapat dipresentasikan sebagai matriks partisi $U = [u_{ik}]_{c \times n}$. Baris ke- i pada matriks tersebut berisi nilai keanggotaan μ_i pada himpunan bagian A_i . Berdasarkan persamaan 1 maka elemen-elemen pada matriks U harus memenuhi kondisi sebagai berikut :

$$u_{ik} \in \{0,1\}; 1 \leq i \leq c; 1 \leq k \leq n \dots\dots\dots (4)$$

$$\sum_{i=1}^c u_{ik} = 1; 1 \leq k \leq n \dots\dots\dots (5)$$

$$0 < \sum_{k=1}^n u_{ik} < n; 1 \leq i \leq c \dots\dots\dots (6)$$

Semua kemungkinan partisi matriks X disebut dengan *hard partition space* (Bezdek, 1981) Partisi Fuzzy (*fuzzy partition*)

Jika pada partisi klasik, suatu data secara eksklusif menjadi anggota hanya pada satu cluster saja, tidak demikian halnya dengan partisi fuzzy. Pada partisi fuzzy, nilai keanggotaan suatu data pada suatu cluster μ_{ik} terletak pada interval $[0,1]$. Matriks partisi pada partisi fuzzy memenuhi kondisi sebagai berikut :

$$u_{ik} \in [0,1]; 1 \leq i \leq c; 1 \leq k \leq n \dots\dots\dots (7)$$

$$\sum_{i=1}^c u_{ik} = 1; 1 \leq k \leq n \dots\dots\dots (8)$$

$$0 < \sum_{k=1}^n u_{ik} < n; 1 \leq i \leq c \dots\dots\dots (9)$$

Baris ke- i pada matriks partisi U berisi nilai keanggotaan data pada himpunan bagian fuzzy A_i . Jumlah

derajat keanggotaan setiap data pada semua cluster (jumlah setiap kolom) bernilai 1 (persamaan 8). Semua kemungkinan partisi dari matriks X disebut dengan *fuzzy partitioning space*, yang didefinisikan sebagai :

$$M_{hc} = U \left|_{ik}^{cn} \{0,1\}, i,k; \prod_{i=1}^c \sum_{k=1}^n u_{ik} = 1, k; 0 < \sum_{k=1}^n u_{ik} < n, i \right. \quad (8)$$

b. Partisi Possibilistik (possibilistic partition)

Tidak seperti halnya kedua partisi diatas, pada partisi possibilistic jumlah nilai keanggotaan suatu data pada semua cluster tidak harus 1, namun untuk menjamin suatu data menjadi anggota dari (paling tidak) suatu cluster, maka diharuskan ada nilai keanggotaan yang bernilai lebih dari 0. Matriks partisi pada partisi fuzzy memenuhi kondisi sebagai berikut (Krishnapuram, 1993):

$$u_{ik} \in [0,1]; 1 \leq i \leq c; 1 \leq k \leq n \dots \dots \dots (10)$$

$$u_{ik} > 0; k \dots \dots \dots (11)$$

$$0 < \sum_{k=1}^n u_{ik} < n; 1 \leq i \leq c \dots \dots \dots (12)$$

Semua kemungkinan partisi matriks X disebut dengan *Possibilistic partitioning space*, yang didefinisikan sebagai :

$$M_{pc} = U \left|_{ik}^{cn} \{0,1\}, i,k; \prod_{i=1}^c \sum_{k=1}^n u_{ik} > 0; k; 0 < \sum_{k=1}^n u_{ik} < n, i \right. \quad (9)$$

A. Algoritma C-Means

Fuzzy C-Means (FCM) adalah suatu teknik pengklusteran data yang mana keberadaan tiap-tiap titik data dalam suatu cluster ditentukan oleh derajat keanggotaan.

FCM bukan merupakan keanggotaan *fuzzy inference system*, namun merupakan deretan pusat cluster dan beberapa derajat keanggotaan untuk tiap-tiap titik data. Informasi ini dapat digunakan untuk membangun suatu fuzzy inference system [3].

Membership function untuk suatu data ke suatu cluster tertentu dihitung menggunakan rumus sebagai berikut :

$$u_{ik} = \frac{D(X_k, V_i)^{\frac{2}{m-1}}}{\sum_{j=1}^c D(X_k, V_j)^{\frac{2}{m-1}}} \dots \dots \dots (13)$$

dimana:

u_{ik} : *Membership function* data ke-k ke cluster ke-i

v_i : Nilai *centroid* cluster ke-i

m : *Weighting Exponent*

Membership function, u_{ik} , mempunyai wilayah nilai $0 \leq u_{ik} \leq 1$. Data item yang mempunyai tingkat kemungkinan yang lebih tinggi ke suatu kelompok akan mempunyai nilai *membership function* ke kelompok tersebut yang mendekati angka 1 dan ke kelompok yang lain mendekati angka 0.

1. Tentukan :

- Matriks X berukuran n x m, dengan n = jumlah data yang akan dicluster; dan m = jumlah variable (kriteria).
- Jumlah cluster yang akan dibentuk = C(≥2)
- Pangkat (pembobot) = w (>1).
- Maksimum iterasi
- Criteria penghentian = ξ (nilai positif yang sangat kecil)
- Iterasi awal, t=1, dan Δ =1;

2. Bentuk Matriks partisi awal, U^0 , sebagai berikut :

$$U = \begin{matrix} & \begin{matrix} 11(x_1) & 12(x_2) & \dots & 1n(xn) \end{matrix} \\ \begin{matrix} 21(x_1) & 22(x_2) & \dots & 2n(xn) \end{matrix} & \\ \vdots & \vdots & \dots & \vdots & \dots \\ \begin{matrix} c1(x_1) & c2(x_2) & \dots & cn(xn) \end{matrix} & \end{matrix} \dots$$

3. Hitung Pusat Cluster, V, untuk setiap cluster :

$$V_{ij} = \frac{\sum_{k=1}^n (u_{ik})^w \cdot x_{kj}}{\sum_{k=1}^n (u_{ik})^w} \dots \dots \dots (14)$$

4. Perbaiki derajat keanggotaan setiap data pada setiap cluster (perbaiki matriks partisi), sebagai berikut:

$$u_{ik} = \frac{d_{ik}^{2/(w-1)}}{\sum_{j=1}^c d_{jk}^{2/(w-1)}} \dots \dots \dots (15)$$

dengan :

$$d_{ik} = d(x_k, v_i) = \left(\sum_{j=1}^m (x_{kj} - v_{ij})^2 \right)^{1/2}$$

Tentukan criteria berhenti, yaitu perubahan matriks partisi pada iterasi sekarang dengan iterasi sebelumnya, sebagai berikut :

$$\Delta = \left\| U^t - U^{t-1} \right\| \dots \dots \dots (16)$$

Apabila $\Delta \leq \xi$, maka iterasi dihentikan, namun apabila $\Delta > \xi$, maka naikkan iterasi (t=t+1) dan kembali ke langkah 3. Pencarian nilai D dapat dilakukan dengan mengambil elemen terbesar dari nilai mutlak selisih antara $u_{ik}(t)$ dengan $u_{ik}(t-1)$.

B. Algoritma Mixture

Algoritma *Mixture* merupakan salah satu jenis data clustering dimana dalam pemodelannya, data dalam satu kelompok diasumsikan terdistribusi sesuai dengan salah satu jenis distribusi sesuai dengan salah satu jenis distribusi statistik yang ada. Algoritma *mixture* merupakan metode yang mempunyai optimasi yang sama dengan algoritma *C-Means* melalui proses optimization dan Maximization [1].

Distribusi statistik yang paling sering digunakan dalam data clustering menggunakan metode *mixture* adalah

distribusi *Gaussian/Normal*. Disamping karena kemudahan penurunan berbagai rumus yang diperlukan dengan distribusi *multivariate Gaussian* menghitung jarak dengan rumus sebagai berikut :

$$D_{Mahalanobis}(x_2, x_1) = \|x_2 - x_1\|_{mahalanobis} = (x_2 - x_1)^T \Sigma^{-1} (x_2 - x_1)$$

dimana:

$(x_2 - x_1)^T$: *Transpose* dari sebuah matriks

$(\Sigma)^{-1}$: *Inverse* dari sebuah matriks

Σ : *Variance Covariance* matriks

dengan *relative abundance* dari cluster yang bersangkutan seperti berikut ini

$$P_{ik} = p_i \cdot f_i(x_k | \theta_i) \dots \dots \dots (17)$$

dimana:

p_{ik} : Probabilitas data ke- k menjadi anggota cluster ke- i

p_i : *Relative abundances* cluster ke- i

$f_i(x_k | \theta_i)$: Distribusi probabilitas cluster ke- i

θ_i : Parameter yang tercakup di dalam distribusi yang diasumsikan untuk cluster ke- i

Untuk penelitian ini, dipaparkan metode *random search* yang memberikan nilai jumlah cluster secara random di awal setiap proses optimasi. Algoritma yang digunakan adalah sebagai berikut :

- a. Tentukan jumlah cluster
- b. Alokasikan data secara *random* ke masing-masing cluster yang telah ditentukan
 1. Hitung *means* (sama dengan *centroid* pada *K-Means*) dari masing-masing cluster
 2. Hitung standar deviasi/*variance covariance* dari masing-masing cluster
 3. Hitung nilai probabilitas masing-masing data ke masing-masing cluster
 4. Kembali ke Step b.1, apabila perubahan nilai probabilitas masih di atas nilai *threshold* yang ditentukan, atau apabila perubahan pada nilai *centroid* masih di atas nilai *threshold* yang ditentukan, atau apabila perubahan pada nilai *objective function* masih di atas nilai *threshold* yang ditentukan.
- c. Kembali ke Step a. apabila masih ada jumlah cluster yang ingin dianalisa.

Dengan asumsi bahwa data terdistribusi secara normal, *means* cluster ke- i , μ_i dihitung dengan menggunakan rumus sama dengan metode *Fuzzy K-Means* dengan u_{ik} merupakan nilai probabilitas data tersebut termasuk di dalam cluster ke- i . Sedangkan standar deviasi/*variance covariance* cluster ke- i , Σ_i dihitung dengan menggunakan rumus sebagai berikut:

$$\mu_i = \sqrt{\frac{\sum_{k=1}^N (x_k - \mu_i)^2}{N - 1}} \dots \dots \dots (18)$$

$$\Sigma_i = \frac{\sum_{k=1}^N (x_k - \mu_i)(x_k - \mu_i)^T}{N - 1} \dots \dots \dots (19)$$

dimana:

N : Jumlah data

μ_i : *Means* cluster ke- i

sedangkan untuk menghitung nilai probabilitas data ke- k ke cluster ke- i digunakan rumus penghitungan probabilitas

C. Prakiraan Curah Hujan

Prakiraan adalah suatu proses yang memperkirakan sesuatu secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi di masa lalu dan sekarang yang dimilikinya agar kesalahan (selisih antara hasil pendugaan dengan kenyataannya) dapat diperkecil [10].

Curah Hujan (mm) merupakan ketinggian air hujan yang jatuh pada tempat yang datar dengan asumsi tidak menguap, tidak meresap dan tidak mengalir. Curah hujan 1 (satu) mm adalah air hujan setinggi 1 (satu) mm yang jatuh (tertampung) pada tempat yang datar seluas 1 m² dengan asumsi tidak ada yang menguap, mengalir dan meresap [10].

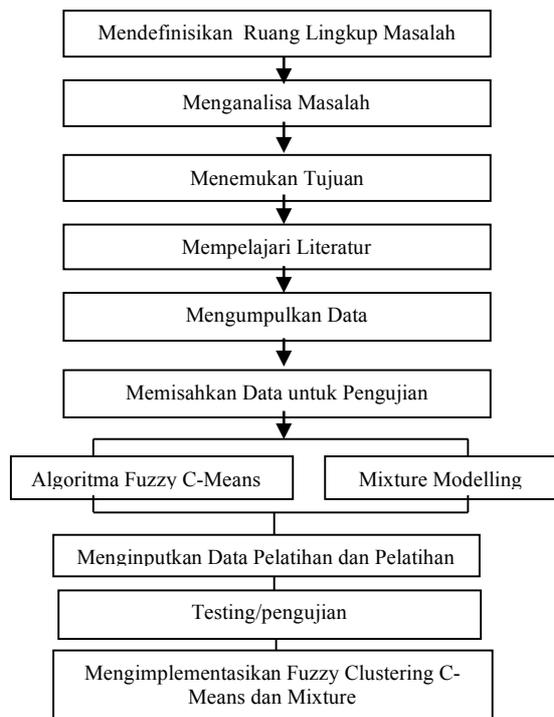
Curah hujan kumulatif 1 (satu) bulan adalah jumlah curah hujan yang terkumpul selama 28 atau 29 hari untuk bulan februari dan 30 atau 31 hari untuk bulan-bulan lainnya.

Sifat hujan merupakan perbandingan antara jumlah curah hujan selama rentang waktu yang ditetapkan (satu periode musim kemarau) dengan jumlah curah hujan normalnya (rata-rata selama 30 tahun (1971-2000) pada bulan dan tempat yang sama. Sifat hujan dibagi menjadi 3 (tiga) kategori, yaitu :

- a. Atas Normal (AN) : jika nilai curah hujan lebih dari 115% terhadap rata-ratanya.
- b. Normal (N) : jika nilai curah hujan antara 85%-115% terhadap rata-ratanya.
- c. Bawah Normal (BN) : jika nilai curah hujan kurang dari 85% terhadap rata-ratanya.

III. METODOLOGI PENELITIAN

Kerangka kerja (*frame work*) merupakan langkah-langkah yang dilakukan dalam penulisan. Adapun kerangka kerja yang digunakan dalam penulisan ini adalah seperti gambar 1.



Gambar 1. Kerangka kerja penelitian

Berdasarkan kerangka kerja pada gambar 1 maka masing-masing langkahnya diuraikan sebagai berikut :

1. Definisi Ruang lingkup masalah
Ruang lingkup masalah yang akan diteliti harus ditentukan terlebih dahulu karena tanpa mampu menentukan serta mendefinisikan rumusan dan batasan masalah yang akan diteliti, maka tidak akan pernah didapat solusi yang terbaik dari masalah tersebut. Jadi langkah ini adalah langkah awal yang terpenting dalam penulisan ini.
2. Analisa Masalah
Analisa masalah adalah langkah yang dilakukan untuk memahami masalah yang telah ditentukan ruang lingkup atau batasannya. Dengan menganalisa masalah yang telah ditentukan tersebut, maka diharapkan masalah dapat dipahami dengan baik.
3. Menentukan Tujuan
Berdasarkan pemahaman dari masala, maka ditentukan tujuan yang akan dicapai dari penelitian ini. Pada tujuan ini ditentukan target yang akan dicapai, terutama yang dapat mengatasi masalah-masalah yang ada.
4. Mempelajari Literatur
Untuk mencapai tujuan maka dipelajari beberapa literature-;iteratur yang dapat dijadikan dasar atau rujukan dalam penelitian ini.
5. Mengumpulkan Data
Dalam pengumpulan data dilakukan observasi yaitu pengamatan secara lansung ditempat penelitian sehingga permasalahan yang ada dapat diketahui dengan jelas. Kemudian dilakukan wawancara yang

bertujuan untuk mendapatkan informasi atau data yang dibutuhkan. Selain itu, juga dilakukan studi kepustakaan yaitu dengan membaca buku-buku yang menunjang dalam melakukan penganalisaan terhadap data dan informasi yang didapat.

6. Memisahkan Data untuk pelatihan dan pengujian
Setelah dilakukan pengumpulan data, maka selanjutnya data yang diperoleh dipisahkan menjadi dua bagian. Bagian pertama dipergunakan untuk melatih Fuzzy Clustering dan bagian kedua digunakan untuk menguji algoritma Fuzzy C-Means dan Mixture.
7. Perancangan Aplikasi Fuzzy Clustering
Bagian ini merupakan perancangan aplikasi fuzzy clustering dimana dalam penelitian ini terdapat dua algoritma yang digunakan untuk mengujikan data curah hujan menggunakan algoritma fuzzy C-Means dan algoritma *mixture* berdasarkan dengan Algoritma yang dimiliki oleh masing-masing.
Dimana untuk algoritma C-Means melakukan urutan proses yang terdapat dalam algoritma dengan menggunakan rumus 14, 15, 16, 17 dan 18. sedangkan untuk algoritma *Mixture* melakukan urutan algoritma sama dengan algoritma C-Means menentukan jumlah cluster kemudian menjalankan rumus berdasarkan urutan algoritma
Mengumpulkan data untuk pelatihan Langkah ini dilakukan sebelum melakukan pelatihan agar dalam melakukan pelatihan data yang di gunakan lengkap.
8. Pelatihan
Langkah ini dilakukan untuk melatih kedua algoritma yaitu Fuzzy C-Means dan Mixture untuk menemukan akurasi data sehingga terjadi pengelompokan atau pengklusteran data curah hujan.
9. Pengujian / Testing
Bagian ini dilakukan untuk menguji dari kedua metode algoritma yang digunakan manakah yang menghasilkan cluster yang akurasi data menjadi cluster sesuai dengan cluster yang ditentukan.
10. Mengimplementasikan Fuzzy Clustering
Setelah fuzzy clustering dengan algoritma fuzzy c-means dan mixture diuji maka selanjutnya Fuzzy clusetring dengan kedua algoritma tersebut diimplementasikan.
Analisa sistem adalah penguraian dari suatu sistem utuh ke dalam bagian komponen-komponen dengan maksud untuk mengidentifikasi dan mengevaluasi permasalahan-permasalahan, hambatan-hambatan yang terjadi dan kebutuhan-kebutuhan yang diharapkan sehingga dapat diusulkan suatu perbaikan. Data yang digunakan adalah data curah hujan bulanan Kota Bengkulu selama lima tahun dimana data berasal dari alat pengukur curah hujan Ambrometer atau Type Hellman. Program bantu yang digunakan dalam pengclusteran data curah hujan Kota Bengkulu menggunakan Algoritma C-Means dan Algoritma Mixture adalah Matlab 6.5 yang mana dalam program Matlab versi 6.5 terdapat Toolbox Fuzzy Cluster yang

mempunyai kemampuan dalam mengelompokkan data sesuai dengan kelompok data yang telah ditentukan.

Kriteria yang digunakan untuk menghentikan proses iterasi dari algoritma C-Means adalah $\|U^t - U^{t-1}\|$, apabila $\Delta < (\text{toleransi})$ maka iterasi dihentikan. Namun apabila $\Delta > (\text{toleransi})$ maka naikkan iterasi ($t=t+1$) dan proses kembali ke menghitung nilai centroid.

Proses Pengujian yang dilakukan adalah penerapan algoritma C-Means untuk pengelusteran dengan data curah hujan, jumlah cluster, banyaknya iterasi, toleransi(ξ), dan pangkat/pembobot telah ditentukan. Data curah hujan, jumlah cluster, banyaknya iterasi, toleransi(ξ), dan pangkat/pembobot di input pada saat running program. Dalam proses ini akan dihitung Centroid dan Membership Function dengan menggunakan algoritma C-Means. Berikut ini akan dijabarkan langkah-langkah pengelusteran dalam model algoritma C-Means. Diberikan data sebagai berikut :

$$Z = \begin{matrix} 14 & 8 & 18 & 5 & 8 & 9 & 5 & 10 & 3 & 10 \\ 19 & 12 & 20 & 14 & 12 & 19 & 2 & 8 & 3 & 8 \end{matrix}$$

Akan dilakukan *Fuzzy Clustering* dengan FCM. Langkah 1

- Jumlah cluster yang diharapkan $c=3$
- Nilai pembobot $m=3$
- Toleransi penghentian $\xi = 1.10^{-6}$
- Maksimum iterasi = 100

Langkah 2 : Inisialisasi matriks partisi awal secara acak

$$U = \begin{pmatrix} 0.361 & 0.397 & 0.323 & 0.017 & 0.395 & 0.134 & 0.550 & 0.303 & 0.288 & 0.450 \\ 0.120 & 0.241 & 0.396 & 0.413 & 0.413 & 0.409 & 0.240 & 0.321 & 0.449 & 0.301 \\ 0.519 & 0.362 & 0.281 & 0.569 & 0.569 & 0.457 & 0.209 & 0.375 & 0.267 & 0.248 \end{pmatrix}$$

Iterasi 1 :

Langkah 3 hitung cluster center (*means*)

$$V = \begin{matrix} 8.890 & 8.457 & 9.129 \\ 9.533 & 11.762 & 13.985 \end{matrix}$$

Langkah 4 dan 5 hitung jarak dan perbaharui matriks partisi. Didapatkan matriks partisi sebagai berikut :

$$V = \begin{pmatrix} 0.318 & 0.151 & 0.351 & 0.333 & 0.151 & 0.297 & 0.289 & 0.902 & 0.195 & 0.902 \\ 0.604 & 0.839 & 0.544 & 0.598 & 0.839 & 0.628 & 0.142 & 0.090 & 0.105 & 0.090 \\ 0.078 & 0.009 & 0.078 & 0.009 & 0.105 & 0.009 & 0.075 & 0.568 & 0.008 & 0.699 \end{pmatrix}$$

$$\|U^i - U^{i-1}\| = 31.154 (>)$$

Iterasi 2 :

Langkah 3 : Hitung cluster center (*means*)

$$V = \begin{matrix} 9.986 & 9.529 & 4.099 \\ 9.674 & 14.799 & 3.127 \end{matrix}$$

Langkah 4 dan 5 hitung jarak dan perbaharui matriks partisi. Didapatkan matriks partisi baru sebagai berikut :

$$\|U^i - U^{i-1}\| = 0.05 (>)$$

Dan seterusnya, perhitungan akan dilaksanakan hingga $\|U^i - U^{i-1}\| < \text{terpenuhi}$ atau maksimum iterasi tercapai.

Kriteria yang digunakan untuk menghentikan proses iterasi dari algoritma Mixture adalah $P_{ik} = \mu_i \cdot f_i(x_k | \mu_i)$, apabila $P_{ik} < (\text{toleransi})$ maka iterasi dihentikan.

Proses Pengujian yang dilakukan adalah penerapan algoritma Mixture untuk pengelusteran dengan data curah hujan, jumlah cluster, banyaknya iterasi, toleransi(ξ), dan pangkat/pembobot telah ditentukan. Titik awal, banyaknya iterasi, toleransi(ξ), dan pangkat/pembobot di input pada saat running program. Dalam proses ini akan dihitung Centroid dan Membership Function dengan menggunakan algoritma Mixture. Berikut ini akan dijabarkan langkah-langkah pengelusteran dalam model algoritma Mixture, diberikan data pada tabel 1.

TABEL 1. DATA ALGORITMA MIXTURE

X1	X2	Y1	Y2
10.0	10.0	10.0	50.0
20.0	40.0	20.0	70.0
30.0	30.0	30.0	60.0
40.0	50.0	40.0	80.0
50.0	50.0	50.0	40.0
60.0	70.0	60.0	40.0
70.0	70.0	70.0	20.0
80.0	80.0	80.0	10.0

Akan dilakukan Fuzzy Clustering Dengan Gaussian Mixture Modelling (GMM).

Langkah 1 :

Tentukan jumlah cluster dimana Kernel/C = 2

Langkah 2 :

- Hitung Means (sama dengan centroid pada FCM) dari masing-masing cluster.

Inisialisasi matriks partisi acak dari data

$$U = \begin{pmatrix} 0.3233 & 0.5268 & 0.6171 & 0.2317 & 0.9771 & 0.6423 & 0.6233 & 0.6232 \\ 0.6767 & 0.4732 & 0.3829 & 0.7683 & 0.0229 & 0.3577 & 0.3767 & 0.3768 \end{pmatrix}$$

Maka didapat means :

$$V = \begin{matrix} 0.5980 & 0.6407 & 0.3637 & 0.2759 \\ 0.5615 & 0.6613 & 0.7555 & 0.3431 \end{matrix}$$

- Hitung Standar Deviasi/variance covariance dari masing-masing cluster

$$S_{XX} = 21.97402830259$$

$$S_{XY} = 83.81626342614$$

$$S_{YX} = 83.81628342614$$

$$S_{YY} = 8.133669424448$$

IV. SIMULASI DAN PENGUJIAN SISTEM

Percobaan dilakukan dengan menggunakan Bahasa Pemrograman Matlab 6.5, data yang digunakan

direpresentasikan dalam bentuk proses pengujian yang dilakukan dalam penerapan algoritma Fuzzy Clustering Means (FCM) dan Mixture. Dari hasil percobaan, parameter yang akan dihitung dalam kedua metode itu adalah matriks partisi, pusat cluster (centroid), standar deviasi, dan nilai probabilitas.

Dalam percobaan yang dilakukan dalam pengujian sistem ini dengan menggunakan data yang tersimpan dalam Microsoft Excel dengan nama DATA CURAH HUJAN.xls. Dimana data yang digunakan terdapat 12 sample data (dimana untuk data selama satu tahun/12 bulan), dalam dua variabel yaitu X1 : tempat pemeriksaan Diperta Padang Harapan dan X2: Stasiun Klomatologi Pulau Bali.

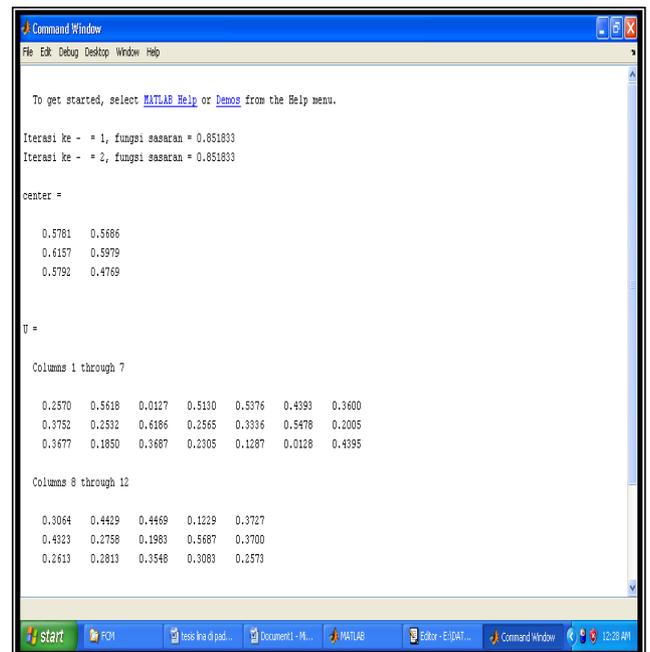
TABEL II. DATA CURAH HUJAN

DATA KE :	VARIABEL	
	X1	X2
1	225	174
2	163	205
3	407	475
4	310	226
5	129	84
6	34	79
7	61	72
8	189	227
9	129	190
10	164	295
11	580	757
12	717	794

Dimana data yang tabel ini nantinya yang akan diimplementasikan ke dalam algoritma Fuzzy Clustering Means (FCM) dan Algoritma Mixture.

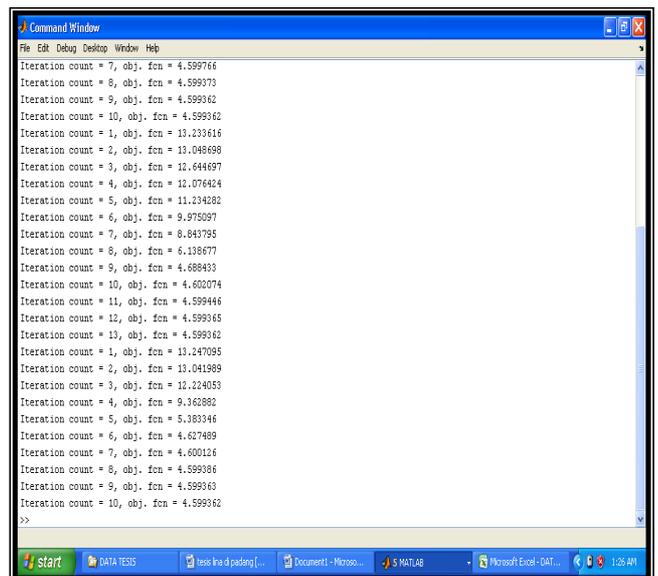
A. Pengujian Hasil Pengelusteran dalam Algoritma Fuzzy Clustering Means (FCM)

Pengujian hasil pengelusteran dalam algoritma Fuzzy clustering means (FCM) dimana dengan mengikuti algoritma yang sudah dibahas di bab empat. Dalam algoritma FCM terdapat Pengujian hasil matriks partisi dengan menginisialisasi matriks partisi awal secara acak dari data yang digunakan. Berikut langkah-langkah yang dilakukan dalam pengujian data tersebut dengan menggunakan Matlab 6.5 yang terdapat pada Matlab editor terdapat tiga pilihan, kita mengklik pilihan change MATLAB current directory. Maka muncul tampilan hasil running program.



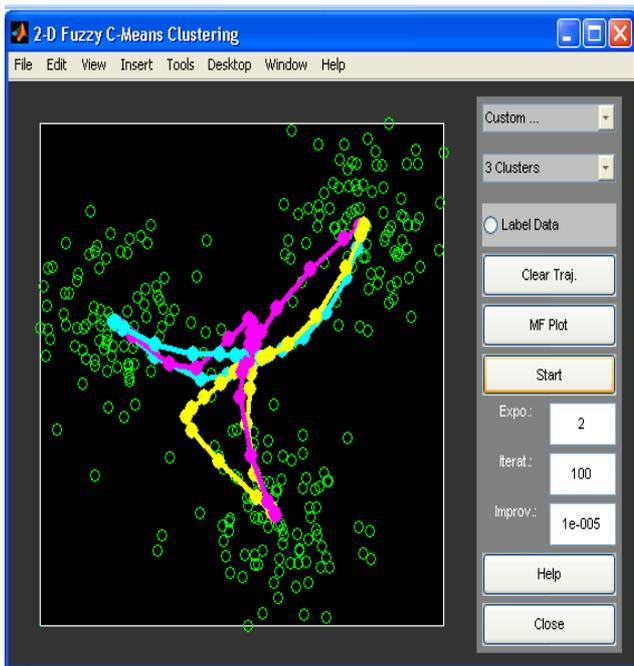
Gambar 2. Command Windows hasil running program

Setelah starnya dijalankan maka akan memulai dengan iterasi pertama, dan star kedua dijalankan maka akan meneruskan proses iterasinya dimana akan terlihat di command windows proses iterasi yang dilakukan.



Gambar 3. Command Windows untuk proses iterasi 2D FCM

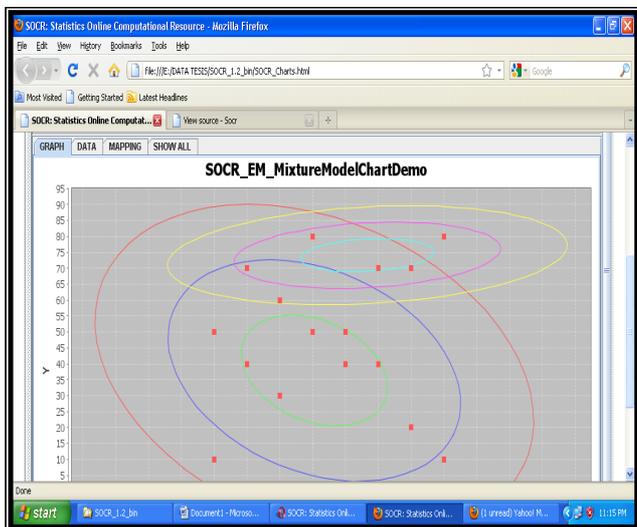
Dengan tampilan pada interface 2D FCM nya sebagai berikut :



Gambar 4. Hasil 2D FCM setelah dilakukan proses iterasi

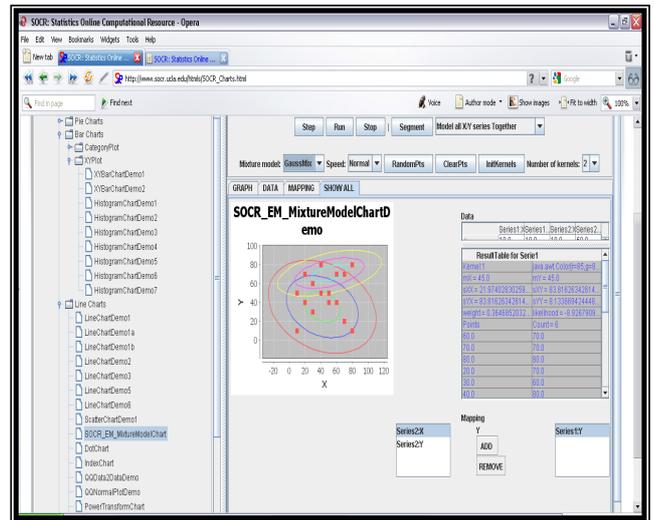
B. Pengujian Hasil Pengclusteran dalam Algoritma Mixture

Pengujian hasil pengclusteran dalam algoritma Mixture dimana dengan mengikuti algoritma yang sudah dibahas di bab sebelumnya. Dalam algoritma Mixture sebenarnya untuk input data sama dengan algoritma yang dibahas pada FCM, tetapi pada algoritma mixture ini terdapat nilai standar deviasi dan nilai probabilitas yang masuk dalam proses algoritma ini.



Gambar 5. Tampilan Graph dari Software SOCR

Chart Graph yang dihasilkan dari proses input data dan proses mapping yang sudah dilakukan dan juga dapat dengan fasilitas Show All karena karena pada tampilan ini akan menampilkan semua chart dengan tabel yang ada.



Gambar 6. Tampilan Show All pada Software SOCR

Dari hasil tampilan Show All terdapat result tabel for series dimana menampilkan informasi nilai means dari sumbu x dan y, nilai standar deviasi, bobot/weight, count/iterasi yang dilakukan berdasarkan dengan proses run yang dijalankan dan epsilon.

Hasil pengelompokan data *Gaussian Mixture Modelling* (GMM) menunjukkan pengelompokan data berdasarkan dengan distribusi statistik dengan *Gaussian*, pada algoritma ini langkah awal yang dilakukan berdasarkan dengan nilai *centroid* titik-titik data yang didapat mean baru dan probabilitas cluster tiap iterasi sehingga terjadi pergeseran data sehingga proses *run* yang dilakukan mendeteksi terjadinya overlap data dalam sebuah *cluster*.

C. Perbandingan Algoritma Fuzzy Clustering Means (FCM) dan Algoritma Mixture (Gaussian Mixture Modeling/GMM)

Data yang digunakan adalah data curah hujan tahun 2008 dengan dua pos pengamatan yaitu X1=Stasiun Klimatologi Pulau Baa dan X2=Diperta Padang Harapan. Dimana data diambil dari bulan januari sampai dengan desember 2008. Berdasarkan dengan penelusuran algoritma Fuzzy Clustering Means (FCM) dan Gaussian Mixture Modelling (GMM) dilakukan proses penginputan dan perhitungan data dengan menggunakan program Matlab 7.1 dengan jumlah cluster, matriks X, matriks partisi, centroid/means, standar deviasi, nilai probabilitas, pembobot/pangkat (w), maksimum iterasi dan kriteria penghentian. Hasil perhitungan dapat disajikan pada tabel 3.

TABEL III. HASIL PERHITUNGAN FCM DAN GMM

No	Fuzzy Clustering Means (FCM)	Gaussian Mixture Modelling (GMM)
1	Pengelompokan data berdasarkan dengan derajat keanggotaan	Pengelompokan data berdasarkan dengan distribusi statistik (<i>Gaussian Mixture Modelling</i>)
2	Pada hasil perhitungan nilai center/ pusat cluster : a. Untuk iterasi 1 : pada cluster 1 terdapat peningkatan pergeseran sebesar 0,07% b. Untuk iterasi 1 : pada cluster 2 terdapat peningkatan pergeseran 0,05% c. Untuk iterasi 2 : pada cluster 1 terdapat penurunan sebesar 0,02% d. Untuk iterasi 2 : pada cluster 2 terdapat penurunan sebesar 0,008% Dari pergeseran yang terjadi pada proses dua iterasi sudah mencapai titik pusat cluster yang mencapai fungsi sasaran (<i>object function</i>) pada koordinat C1:716.25554 dan 716.8897 dan C2:793.1132 dan 793.9201	Pada hasil perhitungna nilai centroid pada 10 iterasi untuk masing-masing cluster didapat nilai means dari masing-masing cluster yang mengandung perubahan data 0% nilai probabilitas 0.5000 untuk cluster 2, dan variences untuk cluster 1 dan 2 terdapat trend data naik turun. Dari koordinat yang terdapat dalam cluster 1 dan 2 semakin tinggi nilai standar deviasi penyimpangan yang terjadi emakin besar dengan kondisi buruk. Standar deviasi dari hasil kriteria yang pertama dan kedua yang dihasilkan matrik deviasi yang trendnya naik turun.
3	Iterasi akan terhenti apabila mencapai fungsi sasaran (<i>object function</i>)	Hasil iterasi berhenti sampai overlap data dan cluster yang ditentukan
4	Pada <i>Fuzzy C-Means</i> dalam penclusturan data membagi data dibagian tengah tanpa memikirkan komposisi dan keadaan data yang dimodel.	Pada <i>Gaussian Mixture Modelling</i> pembagian data dengan menyesuaikan pada keadaan data yang melihat sebaran dan distribusi data yang dianalisa.
5	Dalam <i>Fuzzy C-Means</i> mempunyai kemampuan pengalokasian ulang data ke cluster dan <i>objective function</i> yang digunakan.	<i>Gaussian Mixture Modelling</i> (GMM) mempunyai kemampuan untuk mendeteksi keberadaan suatu cluster yang <i>overlap</i> dengan cluster yang lain.
6	Pengalokasian data dalam setiap cluster tergantung pada pusat cluster.	Pengalokasian data berdasarkan dengan means masing-masing cluster.

Dari tabel 3 dapat disimpulkan bahwa kedua algoritma *Fuzzy Clustering Means* (FCM) dan *Gaussian Mixture Modelling* (GMM) pada data curah hujan Kota Bengkulu maka algoritma dapat dikembangkan dengan kasus ini adalah *Fuzzy C-Means* karena pengelompokan datanya berdasarkan dengan derajat keanggotaan sehingga memperbaiki pusat cluster yang dihasilkan dalam mencapai minimisasi fungsi sasaran yang menggambarkan jarak dari titik-titik data yang diberikan ke pusat cluster untuk mencari lokasi terbaik untuk cluster-cluster dapat dilakukan lebih cepat yang ditunjukkan dengan proses penghentian iterasi.

V. KESIMPULAN DAN SARAN

Perbandingan algoritma *Fuzzy Clustering Means* (FCM) dan *Gaussian Mixture Modelling* (GMM) untuk mendapatkan alternatif yang terbaik, dapat digunakan untuk memecahkan masalah dalam mengelompokkan data yang memiliki kesamaan jumlah data curah hujan yang sama atau mendekati, sehingga dapat digunakan sebagai pendukung pengambilan keputusan dalam mengelompokkan data.

Hasil analisa perbandingan antara algoritma *Fuzzy C-Means* (FCM) dan *Gaussian Mixture Modelling* (GMM) maka algoritma *Fuzzy C-Means* dapat dikembangkan dalam pengelompokkan data curah hujan Kota Bengkulu yang berdasarkan sifat hujan karena algoritma fuzzy *C-Means* dapat menentukan lokasi terbaik dalam cluster berdasarkan dengan proses iterasinya. Hasil penelitian ini Algoritma FCM dapat membantu Badan Meteorologi, Klimatologi dan Geofisika Stasiun Klimatologi Pulau Baii Bengkulu dalam mengelompokkan atau mengclusterkan data berdasarkan dengan sifat hujan.

DAFTAR PUSTAKA

- [1] Agusta Yudi, (2007), "*K-Means-Penerapan, Permasalahan dan Metode Terkait*", *Journal Sistem dan Informatika* Vol.3 hal 47-60.
- [2] Aqil M, Firmansyah, Prabowo A, Macalinao M, 2007, "*Klastering Tingkat Pemakaian Pompa Air Tanah Menggunakan Model Fuzzy*" *Jurnal Informatika Pertanian* Volume 16 No.1
- [3] Kusumadewi S, 2002, "*Analisis Desai Sistem Fuzzy menggunakan Toolbox Matlab*", Yogyakarta, Graha Ilmu.
- [4] Kusumadewi S dan Purnomo H, 2004, "*Aplikasi Logika Fuzzy Untuk Pengambilan Keputusan*", Yogyakarta, Graha Ilmu.
- [5] Kusriani dan Luthfi Taufik Emha, 2009, "*Algoritma Data Mining*", Yogyakarta, Penerbit Andi.
- [6] Kusumadewi S, Hartati, S, Harjoko A, Wardoyo R, 2006, "*Fuzzy Multi Atribut Decision Making (Fuzzy MADM)*" Yogyakarta, Graha Ilmu.
- [7] Naba Agus, 2009, "*Belajar Cepat Fuzzy Logic Menggunakan Matlab*"; "*Belajar Cepat Fuzzy Logic Menggunakan Matlab*" Yogyakarta, Andi
- [8] Susanto dan Ernawati, 2005, "*Pembagian Kelas Peserta kuliah ebrdasarkan dengan Fuzzy Clustering dan Partition Coefficient and Exponential Separation (PCAES) Index*".
- [9] Luthfi Taufik E, 2007, "*Fuzzy C-Means untuk Clustering data (Studi Kasus : Data Performance Mengajar Dosen)*", Seminar National Teknologi Tahun 2007, Yogyakarta.
- [10] Warsito B dan Sumiyati S, 2003, "*Prediksi Curah Hujan Kota Semarang dengan Feedward Neural Network Menggunakan Algoritma Quasi Newton BFGS dan Levenberg-Marquardt*", *Jurnal Program Studi Teknik Lingkungan FT Undip*.