# **BAB 1**

**PENDAHULUAN**

## **Latar Belakang**

*Clustering* merupakan proses pengelompokan data sehingga dapat memberikan informasi yang lebih bermanfaat. *Clustering* juga dapat dijelaskan sebagai pengelompokan data berdasarkan informasi dari observasi yang menjelaskan objek observasi tersebut dan hubungannya (Tan, Steinbach, & Kumar, 2006). Metode *clustering* itu sendiri memiliki beberapa tipe yang dapat diklasifikasikan sebagai berikut *hierarchical versus partitional* , *exclusives versus overlapping versus fuzzy*, dan *complete versus partial* (Tan, Steinbach, & Kumar, 2006).

Salah satu pengklasifikasian *clustering* yang telah disebutkan adalah *exclusives versus overlapping versus fuzzy* (Tan, Steinbach, & Kumar, 2006). *Exclusives clustering* bisa diartikan sebagai sebuah objek hanya dapat masuk kedalam sebuah *cluster*. *Overlapping clustering* sendiri memiliki pemahaman yang berlawanan dengan *exclusive clustering*. Pada *overlapping* *clustering* suatu objek dapat tergabung dalam beberapa *cluster*. *Fuzzy clustering* sendiri mengikuti konsep *fuzzy* pada matematika dimana setiap objek merupakan anggota setiap *cluster* namun memiliki nilai keanggotaan 0 (bukan anggota cluster secara absolut) hingga 1 (anggota cluster secara absolut) di tiap *cluster* (Hoppner, Klawonn, Kruse, & Runkler, 2000).

*Fuzzy Clustering* memiliki keunggulan dalam pengelompokan. Kemunculan *fuzzy clustering* dipicu dengan adanya ketidakpastian dari *clustering* biasa yang dipengaruhi oleh *noise* (data-data yang terlihat sumbang atau pencilan) dan *fuzzy clustering* bertujuan membuat *clustering* lebih alami (Hoppner, Klawonn, Kruse, & Runkler, 2000). *Fuzzy clustering* sudah dikembangkan dan menghasilkan beberapa algoritme. Hoppner dkk (2000) juga menyebutkan 3 jenis *fuzzy* *clustering* klasik yaitu *fuzzy c-means algorithm*, *Gustafson-Kessel algorithm*, *Gath-Geva algorithm*.

FCM sendiri memang cukup populer digunakan dalam berbagai bidang aplikasi seperti MRI (Rajendran & Dhanasekaran, 2012), *image recognition* (Hoppner dkk, 2000), dan *data clustering* (Nayak dkk, 2014). Namun kelemahan dari FCM adalah sering kali terperangkap dalam optimum lokal dan inkonsistensi akibat pengambilan matriks keanggotaan awal yang dilakukan secara acak (Nayak dkk, 2000). Keadaan optimum lokal sendiri merupakan keadaan dimana *centroid* terjebak di sebuah area yang sempit dan tidak menyebar, hal ini mengakibatkan *clustering* menjadi tidak baik dan kurang efisien (dibutuhkannya centroid yang menyebar dan mengglobal sehingga menghasilkan *clustering* yang baik) (Nayak dkk, 2014).

Pada penelitian sebelumnya telah banyak dikembangkan metode-metode untuk mengatasi kekurangan FCM. Penelitian-penelitan tersebut mengembangkan FCM dengan menambahkan metode optimasi. FCM *based on* *Particle Swarm Optimization* (PSOFCM) (Kang & Zhang, 2011), *Subtractive Fuzzy C-Means* (SFCM) (Haqiqi, 2015), FCM *based on* *Forest Optimization* (FOFCM) (Ayed, Halima, & Alimi, 2016), *Firefly Algorithm* FCM (FAFCM) (Nayak dkk, 2014), dan lainnya adalah hasil dari penelitian untuk mengatasi sifat kecendrungan FCM terjebak dalam optimum local dan inkonsistensi. Jika dibandingkan dengan dengan FCM sendiri semua FCM yang dioptimasi memang terbukti lebih baik.

*Fuzzy C-Means based on Firefly Algorithm* (FAFCM) merupakan salah satu metode optimasi pada FCM dengan algoritme *Firefly*. *Firefly Algortihm* (FA) sendiri merupakan algoritme yang dapat melakukan *clustering* dimana mengikuti perilaku dari kunang-kunang. Kunang-kunang memiliki cara untuk menarik perhatian kunang-kunang lainnya menggunakan cahaya yang mereka pancarkan. Cahaya yang dipancarkan ini membuat kunang-kunang lainnya akan mendekat dan berkumpul, begitu juga dengan teknik *clustering* menggunakan FA (Yang, 2010). Pada FA akan terdapat titik observasi yang memiliki intensitas (diilustrasikan sebagai cahaya pada kunang-kunang) dan titik-titik lainnya akan mendekat sesuai dengan intensitas tersebut (titik-titik observasi yang mendekat akan tergabung dalam satu *clustering*) (Yang, 2010). Dengan adanya penggabungan FCM dan FA, FAFCM mampu menangani masalah inkonsistensi dan optimum lokal. FAFCM juga menunjukan hasil yang lebih baik dalam jumlah iterasi dan nilai yang lebih baik dibandingkan dengan FCM dan PSOFCM (Nayak dkk, 2014).

Piranti lunak *fuzzy* *clustering* berbayar maupun tidak berbayar dengan metode optimasi seperti yang sudah disebutkan masih belum dapat ditemui. Adapun pada penelitian-penelitian sebelumnya hanya diberikan *psudocode* (alur kode pemograman yang dituliskan secara sederhana). R merupakan salah satu piranti lunak yang mampu melakukan *fuzzy clustering*. R memiliki banyak *library* yang dapat digunakan untuk membantu dalam analisis data, namun *library* untuk melakukan *fuzzy* *clustering* sendiri masih belum ada. Pada piranti lunak R kita dapat memberikan tambahan *library* sendiri (pada kasus ini *library* *fuzzy* *clustering*). Penambahan *library* pada R dapat dimungkinkan karena piranti lunak ini berbasiskan *command line*, dimana kita dapat membuat kode pemograman sesuai dengan algoritme yang kita inginkan.

*Rcluzzy* merupakan aplikasi yang dikembangkan oleh Baiq Haqiqi Nurul sejak tahun 2014 di bidang *fuzzy clustering*. *Rcluzzy* merupakan sebuah aplikasi dengan bahasa pembangun *back end* adalah R. *Rcluzzy* memiliki bahasa *front end* adalah *C#*. Aplikasi ini memang bertujuan untuk memudahkan pengguna yang kesulitan melakukan pekerjaan dengan basis *command line*. Kesimpulannya *Rcluzzy* menggabungkan dua bahasa dimana mesin untuk melakukan analisiss adalah bahasa R (*back end*) dan untuk tampilan maupun *input* dan *output* menggunakan bahasa *C#* (*front end*).

*Rcluzzy* sendiri merupakan aplikasi yang memiliki alat-alat *fuzzy* *clustering* yang sudah mencakup fungsi-fungsi dasar, namun *Rcluzzy* masih belum terdapat metode optimasi seperti optimasi FCM dengan PSO, optimasi FCM dengan algoritme *forest*, optimasi FCM dengan algoritme *firefly*, dan lainnya. Mengetahui adanya keterbatasan tersebut penelitian ini akan memberikan penambahan modul FAFCM agar pengguna dapat menggunakan FAFCM dalam melakukan analisa data pada *Rcluzzy*.

Pada penelitian ini selain ditambahkan modul FAFCM ke dalam aplikasi *Rcluzzy*, akan dilakukan perbandingan antara hasil pengelompokan metode FCM dan beberapa FCM yang sudah dioptimasi (SFCM dan PSOFCM) dengan hasil pengelompokan menggunakan metode FAFCM. Dengan adanya perbadingan ini kita dapat melihat seberapa efisien metode FAFCM dibandingkan dengan metode FCM, SFCM, dan PSO-FCM apabila menggunakan data BPS (Pendidikan), maupun data bangkitan.

Data pendidikan menjadi begitu penting karena keberhasilan suatu bangsa dalam membangun pendidikan merupakan barometer tingkat kemajuan bangsa tersebut (Musyaddad, 2013). Pendidikan mampu memberikan kesejahteraan bagi rakyat. Rakyat yang berpendidikan akan dengan mudah untuk masuk dalam persaingan dunia pekerjaan. Semakin tinggi pendidikan seseorang dapat menunjukan seberapa tinggi kualitas pengetahuannya, maka dari itu pendidikan harus mendapatkan perhatian khusus oleh Negara (Musyaddad, 2013).

Penelitian ini akan dilakukan *clustering* pada sektor pendidikan di Indonesia khususnya di Pulau Jawa dengan tingkat kabupaten untuk menganalisa bagaimana bentuk pendidikan di Indonesia. Data ini diperoleh dari BPS dan akan dilakukan analisa data menggunakan metode FAFCM. Dengan adanya analisa data pendidikan ini akan dilihat dan disimpulkan bagaimana tingkat pendidikan yang ada di Pulau Jawa.

## **Rumusan Masalah**

Berdasarkan penjelasan dari latar belakang diatas, rumusan masalah dari penelitian ini dapat disimpulkan bahwa perlunya sebuah aplikasi yang mampu melakukan sebuah *clustering* pada kumpulan data dengan menggunakan teknik optimasi *firefly* pada FCM. Hal ini akan dilakukan dengan mengembangkan piranti lunak *Rcluzzy* dimana pada *Rcluzzy* sendiri sudah terdapat fitur FCM namun belum terdapat FCM dengan optimasi *firefly*. Pengembangan optimasi *firefly* pada FCM dapat menanggulangi kekurangan pada FCM seperti terperangkap dalam optimum local dan inkonsistensi. Penelitian ini akan dilakukan pada data pendidikan di Pulau Jawa, dimana hasil *clustering* data pendidikan akan menunjukan tingkat pendidikan suatu lokasi (Kabupaten/kota).

## **1.3 Batasan Masalah**

1. Belum adanya aplikasi *clustering* metode FAFCM.
2. *Rcluzzy* akan dikembangkan dengan menambahkkan fitur optimasi. Pada penelitian ini akan menambahkan fitur optimasi FCM dengan algoritme *Firefly* (FAFCM).
3. FAFCM akan menggunakan 2 jarak yaitu *euclidean* dan *mahalanobis*
4. Penggunaan data bangkitan dan data pendidikan dari BPS dalam *clustering* dan akan dilakukan perbandingan dengan metode FCM lainnya (FCM, SFCM, dan PSO-FCM).

## **1.4 Manfaat dan Tujuan Penelitian**

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Mengembangkan *Rcluzzy* dengan menambahkan optimasi *Firefly* pada *Fuzzy C-means* (FAFCM)
2. Menganalisis hasil *clustering* metode FAFCM menggunakan 2 fungsi jarak (*euclidean* dan *mahalanobis*) data bangkitan dan data Pendidikan dari BPS
3. Membandingkan hasil *clustering* metode FAFCM dengan FCM, SFCM, dan PSO-FCM pada data bangkitan dan data Pendidikan dari BPS.

Manfaat dari penelitian ini adalah sebagai berikut:

1. Memberikan pengembangan metode *clustering* FAFCM pada Rcluzzy
2. Memudahkan pengguna yang tidak terbiasa dengan bahasa pemograman (*command line* R) dapat dengan mudah melakukan *fuzzy* *clustering* khususnya FAFCM menggunakan piranti lunak *Rcluzzy*
3. Dengan adanya penambahan optimasi FA pada FCM maka aplikasi Rcluzzy ini akan lebih fleksibel dalam melakukan analisa data.
4. Memberikan pegetahuan mengenai perbandingan FAFCM dengan optimasi lainnya (FCM, SFCM, dan PSO-FCM) sehingga dapat dilihat manakah algoritme optimasi yang lebih efisien.
5. Dengan studi kasus Pendidikan akan dapat terlihat *clustering* pada tingkat pendidikan.

## **1.5 Sistematika Penulisan**

Penulisan penelitian ini terdiri dari lima bab, dengan rincian dari masing

masing bab sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini dipaparkan mengenai latar belakang mengapa akan dilakukan penambahan optimasi pada Rcluzzy, rumusan masalah apa saja yang ditemui, batasan masalah yang menjadi focus dari penelitian ini, tujuan penulisan dan manfaat penelitian, serta sistematika penulisan.

BAB II KAJIAN PUSTAKA DAN KERANGKA PIKIR

Bab ini berisi tinjauan teori yang akan digunakan dalam penelitian ini dan

paparan kajian teori serta kerangka pikir bagaimana penelitian ini akan di lakukan.

BAB III METODOLOGI

Bab ini berisi metode penelitian dan alat-alat yang digunakan dalam

penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tahapan analisis, perancangan sistem yang akan diajukan,

implementasi sistem, uji coba sistem, hingga tahapan evaluasi sistem.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari hasil penelitian yang dilakukan, serta saran

yang bisa diberikan oleh peneliti

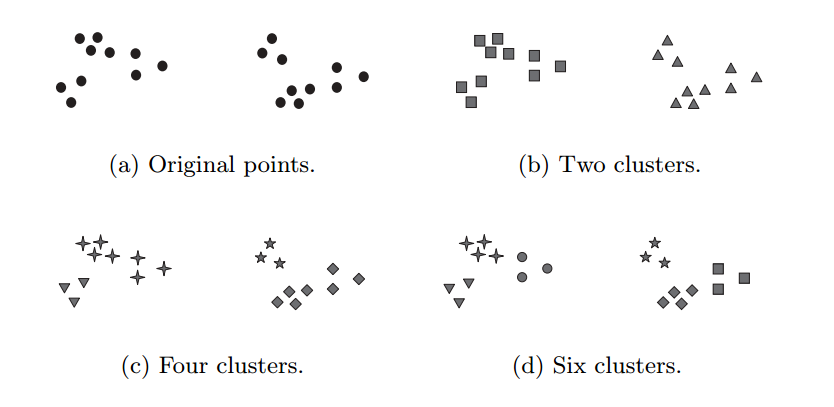
# **BAB 2**

**KAJIAN PUSTAKA DAN KERANGKA PIKIR**

## **2.1 Teori dan kerangka pikir**

***Clustering***

*Cluster analysis divides data into groups (clusters) that are miningfu, useful, or both* (Tan, Steinbach, & Kumar, 2006). *Clustering* adalah analisa data yang membagi data menjadi beberapa grup yang memiliki makna, bermanfaat, maupun keduanya. Dalam sebuah *cluster* setiap obyek yang terkumpul dalam sebuah *cluster* memiliki kemiripan karakteristik. Pada dasarnya manusia sendiri sering melakukan *clustering*, secara sederhana ketika anak kecil diberikan gambar sebuah kota maka dengan sendirinya anak kecil tersebut mampu membedakan yang mana namanya gedung, kendaraan, pepohonan, orang, dan hewan (Tan, Steinbach, & Kumar, 2006).



*Clustering* dibagi menjadi 3 yaitu *hierarchical versus partitional*, *exclusives versus overlapping versus fuzzy*, dan *complete versus partial* (Tan, Steinbach, & Kumar, 2006). Pada *hierarchical versus partitional* dapat dilihat pada *partitional* setiap objek hanya masuk dalam satu clauster, *hierarchical* sendiri adalah keadaan yang sama dengan *partitional* namun dalam sebuah *cluster* memiliki *cluster* (*nested cluster*). Klasifikasi *exclusives versus overlapping versus fuzzy* memiliki pemahaman berbeda dengan *hierarchical versus partitional*. *Exclusives* hanya mengijinkan satu objek masuk dalam satu *cluster*, berlawanan dengan *overlapping* dimana objek diijinkan untuk masuk kedalam lebih dari satu *cluster*. *Fuzzy* sendiri memiliki pemahahaman yang berbeda yaitu setiap objek merupakan anggota setiap *cluster* namun memiliki derajat keanggotaan dimana 0 (bukan anggota cluster secara absolut) hingga 1 (anggota cluster secara absolut). *Complete versus partial* sendiri memiliki paham keanggotaan yang sama yaitu setiap objek hanya diperbolehkan tergabung dalam satu *cluster*, namun pada *complete* setiap objek wajib tergabung dalam sebuah cluster sedangkan *partial* tidak.

***Similarity measure* dalam *clustering***

*Clustering* data menggunakan ukuran jarak dalam melakukan pengelompokan, dimana objek akan dilihat kedekatannya dengan objek lainnya melalui ukuran jarak dan menjadi acuan dalam pengemlompokan. Ukuran jarak yang digunakan dalam penelitian ini adalah:

* ***Euclidean distance***

*Euclidean distance* merupakan ukuran jarak yang paling populer. Perhitungan Euclidean berdasarkan *hyperspherical cluster,* yaitu *cluster* yangpermukaan keanggotaannya konstan. ilustrasi ukuran jarak Euclidean dapat dilihat pada Gambar 1.

1. *Euclidean*

(b) *Euclidean* pada *hyperspherical cluster*

Gambar 1. Ilustrasi ukuran jarak Euclidean

Adapun persamaan ukuran jarak Euclidean adalah sebagai berikut :

(1)

Keterangan : : koordinat objek

: dimensi objek

* ***Mahalanobis distance***

Jarak Mahalanobis digunakan untuk mengatasi *hyperellipsoidal cluster*, yaitu *cluster* yang berbentuk elips(Abonyi & Feil, 2010). Ukuran jarak terdistorsi oleh adanya korelasi linear diantara fitur (variabel). Mahalanobis mengevaluasi ukuran jarak yang ada dengan menambahkan penimbang pada setiap variabel Jika tidak terdapat korelasi antara variabel, ukuran jarak yang dihasilkan akan sama dengan Euclidean. (Xiang, Nie, & Zhang, 2008).

Ilustrasi ukuran jarak Mahalanobis dapat dilihat pada Gambar 2. Adapun persamaan ukuran jarak Mahalanobis adalah sebagai berikut :

(2)

Keterangan : : invers dari kovarian matriks dan

: transpose dari vektor

(a) *Mahalanobis*

(b) *Mahalanobis* pada *hyperellipsoidal cluster*

Gambar 2. Ilustrasi ukuran jarak Mahalanobis. Jarak A sama dengan jarak B

***Fuzzy***

*Fuzzy sets are actually functions that map a value that might be a member of the set to a number between zero and one indicating its actual degree of membership*. *Fuzzy* set adalah fungsi pemetaan sebuah nilai dimana nilai tersebut tergabung dalam sebuah set dengan derajat keanggotaan mulai dari 0 hingga 1 (Cox, 1994)

***Fuzzy clustering***

*Fuzzy* *clustering* adalah salah satu cara menganalisa data dengan melakukan pengelompokan dimana sestiap objek merupakan anggota setiap *cluster* namun memiliki derajat keanggotaan dimana 0 (bukan anggota cluster secara absolut) hingga 1 (anggota cluster secara absolut) (Tan, Steinbach, & Kumar, 2006). *Fuzzy clustering* sudah dikembangkan dan menghasilkan beberapa algoritme. Hoppner dkk (2000) menyebutkan adanya 3 jenis *fuzzy* *clustering* klasik yaitu *fuzzy c-means algorithm*, *Gustafson-Kessel algorithm*, *Gath-Geva algorithm*.

***Fuzzy C-Means***

*Fuzzy C-Means* (FCM) merupakan algoritma *Fuzzy Clustering* yang dikembangkan oleh Bezdek pada tahun 1981 (Bezdek, 1981). FCM merupakan pengembangan dari *hard c-means* oleh Dunn pada tahun 1973 (Liu, Jeng, Yih, & Yu, 2009). Ukuran jarak yang digunakan adalah Euclidean. Adapun fungsi objektif dari FCM dapat dilihat pada rumus (5):

(3)

(4)

Dengan ketentuan :

(5)

,

; ;

Keterangan : : Fungsi Objektif FCM

: matriks keaggotaan data

: matriks pusat *cluster*

*:* matriksdata

: derajat keanggotaan objek pada *cluster*

: jumlah *cluster* pada data

: jumlah objek pada data

: jumlah atribut pada data

: tingkat kekaburan atau *fuzzifier*

: vektor pusat *cluster* ke-

: vektor baris data ke-

: jarak antara objek dengan

Proses pengelompokan berakhir saat ditemukan kondisi penghentian algoritma. Kondisi ini dapat berupa iterasi maksimal () maupun maksimal tingkat *error* ( yang diinginkan pengguna. Tingkat *error* () didapat dari selisih terbesar elemen pada iterasi terakhir dan iterasi sebelumnya ().

* **Proses algoritma FCM**

*Flowchart* perhitungan algoritma FCM dapat dilihat pada gambar berikut :

Flowchart FCM (1)

Gambar 3. *Flowchart* FCM jarak Euclidean

Adapun tahapan dalam melakukan perhitungan FCM standar sebagai berikut (Castillo, Rubio, Soria, & Naredo, 2012):

1. Menentukan parameter awal pengelompokan yaitu jumlah *cluster* , *fuzzifier* , dan syarat kondisi penghentian iterasi ( dan/atau .
2. Membangkitkan nilai acak matriks inisiasi dengan menggunakan rumus (4).
3. Menghitung pusat *cluster* menggunakan rumus (6).
4. Menghitung jarak antara data dengan seluruh pusat *cluster* menggunakan rumus (3).
5. Menghitung derajat keanggotaan seluruh objek pada setiap iterasi menggunakan rumus (7).

(6)

(7)

1. Cek kondisi syarat penghentian algoritma. Jika belum terpenuhi, ulangi ke langkah 3-5.

***Fuzzy C-Means* – Common Mahalanobis**

FCM bekerja dengan baik hanya pada *spherical shaped cluster* karena fungsi objektif dihitung berdasarkan jarak Euclidean (Liu, Jeng, Yih, & Yu, 2009). Agar FCM dapat bekerja pada *hyperellipsoidal* *shape cluster*, jarak yang digunakan adalah Mahalanobis. Kovarians matriks yang digunakan mempertimbangkan derajat keanggotaan dan jarak ke pusat *cluster* dari masing-masing elemen data. Setiap *cluster* menggunakan kovarians matriks yang sama (*Common Mahalanobis*). Jarak antara elemen data dan pusat cluster dihitung dengan rumus berikut :

(8)

Jika ,

(9)

(10)

Keterangan : : kovarians Matriks pada iterasi ke-

: invers dari kovarians matriks

Dalam algoritma ini, pada data yang tingkat korelasi antar atributnya rendah, kovarians matriks akan bernilai sama dengan matriks identitas. Maka dari itu, hasil *clustering* ini akan sama dengan FCM jarak Euclidean. Formula yang digunakan yaitu sebagai berikut :

(11)

Jika atau ,

(12)

* **Proses algoritma FCM-CM**

Adapun tahapan dalam melakukan perhitungan FCM-CM hampir sama dengan FCM standar. Tahapannya adalah sebagai berikut:

1. Menentukan parameter awal pengelompokan yaitu jumlah *cluster* , *fuzzifier* , dan syarat kondisi penghentian iterasi ( dan/atau .
2. Membangkitkan nilai acak matriks inisiasi sesuai dengan rumus (4).
3. Mencari pusat masing-masing *cluster* menggunakan rumus (6).
4. Menghitung menggunakan rumus (8) dan menggunakan rumus (11). Kemudian bandingkan menggunakan rumus (12) untuk menghasilkan kovarians matriks.
5. Menghitung jarak antara titik dengan pusat *cluster* menggunakan rumus (9) dan (10).
6. Menghitung derajat keanggotaan seluruh objek pada setiap iterasi menggunakan rumus (7).
7. Cek kondisi syarat penghentian algoritma. Jika belum terpenuhi, ulangi ke langkah 3-8.

Flowchart FCM (1)

Gambar 4. *Flowchart* FCM jarak Mahalanobis

## **2.2 Penelitian Terkait**

Beberapa penelitian terkait yang menjadi dasar dari penelitian ini adalah :

1. An Improved Firefly Fuzzy C-Means (FAFCM) Algorithm for Clustering Real World Data Sets (Nayak dkk, 2014)

Pada penelitian ini dilakukan optimasi pada FCM dengan menggunakan algoritme *firefly*. Pengoptimasian ini dilakukan untuk mengatasi permasalahan pada FCM yaitu local optimum. Optimasi ini juga dibandingkan dengan FCM dan PSO-FCM. Hasil dari penelitian ini mengungkapkan bahwa FAFCM menghasilkan hasil yang lebih baik dari yang dihasilkan oleh PSO-FCM dan FCM.

1. Analisis Perbandingan Metode Fuzzy C-Means dan Subtractive Fuzzy C-Means (Haqiqi, 2015)

Pada penelitian ini dikembangkan aplikasi yang berisikan alat-alat fuzzy clustering yaitu *Rcluzzy*. Penelitian ini juga sudah dikembangkan lebih lanjut dan ditambahkan beberapa optimasi dalam aplikasi ini seperti *Subtracive Fuzzy C-Means* (SFCM), dan PSOFCM. *Rcluzzy* dapat dengan mudah dikembangkan karena menggunakan bahasa *back end* R, dengan tampilan (*front end*) menggunakan C#. Penelitian ini juga menunjukan kelebihan SFCM dibanding FCM. Penelitian ini menyarankan untuk menggunakan algoritma optimasi untuk menentukan nilai fuzzifier serta menggunakan ukuran jarak selain Euclidean yaitu Mahalanobis.

1. Fuzzy C-Means Algorithm Based on Standard Mahalanobis Distances (Liu, Jeng, Yih, & Yu, 2009)

Pada penelitian ini FCM dikembangkan dengan menggunakan ukuran jarak Mahalanobis. Penelitian ini membandingkan hasil pengelompokan dengan algoritma Gustafson-Kessel, Gath-Geva, dan FCM-Mahalanobis. FCM-Mahalanobis pada penelitian ini terdapat tiga metode yaitu adaptive *Mahalanobis distance* (FCM-M), *common Mahalanobis* (FCM-CM), dan *standard Mahalanobis* (FCM-SM). Pada penelitian ini dapat dilihat bahwa metode FCM-SM menunjukkan hasil yang paling baik. Namun hasilnya tidak berbeda jauh dengan FCM-CM.

## **2.3 Kerangka Pikir**

Kerangka penelitian ini berfokus pada pembangunan modul optimasi pada FCM dengan algoritme *firefly* pada aplikasi R-Cluzzy. FAFCM ini akan menggunakan dua ukuran jarak yaitu *Euclidean* dan *Mahalanobis*. Modul yang dibangun akan dilakukan uji coba pada data bangkitan (*iris* dan *lung cancer*) dan data studi kasus Pendidikan di pulau jawa berdasaran Kabupaten/Kota. Variabel yang menjadi dasar dalam pengelompokan data studi kasus adalah Index Pembangunan Manusia, Angka Melek Huruf, jumlah SD, jumlah SMP, jumlah SMA, Angka Partisipasi Murni untuk umur 7-12, 13-15, 16-18, rata-rata lama sekolah.

Pertama akan dilakukan pengelompokan dengan FAFCM menggunakan jarak Euclidean kemudian akan dibandingkan dengan hasil pengelompokan FCM dan PSO-FCM. FAFCM dengan jarak yang berbeda yaitu Mahalanobis, hasilnya akan dianalisis dengan FAFCM Euclidean. Perbandignan ini akan dilakukan dengan menggunakan indeks validitas yang telah ada di aplikasi. Dari hasil pengelompokan terbaik akan dianalisis variabel yang menjadi pembeda antar-cluster.

C:\Users\joshj\Documents\Skripsi\NISSU\NISSU\Skripsi\Untitled Diagram (1).png

## **2.4 Hipotesis Penelitian**

Hipotesis penelitian ini adalah analisis clustering dengan menggunakan metode FAFCM menggunakan kedua jarak yaitu *Euclidean* dan *Mahalanobis* pada data dapat menghasilkan pengelompokan yang lebih baik dibandingkan FCM dan PSO-FCM.

# **BAB 3**

**METODOLOGI**

## **3.1 Ruang Lingkup**

Penelitian ini menggunakan 2 data, studi kasus dan data bangkitan. Berikut penjelasan mengenai data yang digunakan.

**Data bangkitan**

*Iris* merupakan data bangkitan yang akan dipakai dalam melakukan *clustering*. Iris berisikan data mengenai panjang lebar kelopak bunga (panjang lebar sepal dan petal). Data ini berjumlah 150 *record* dimana 50 record untuk ketiga spesies bunga (setosa, versicolor, virginica) (Edgar Anderson, 1935).

Penelitian ini juga menggunakan data bangkitan dengan jumlah data 20 dan 200. Setiap kelompok data (20 dan 200) akan memiliki jumlah variabel 2, 3, dan 4. Data akan dibuat menggunakan *software* R dengan fungsi ‘rnorm’ pada *package* stat ( norm{stat} ). Data akan dibentuk sebaran data *hyperellipsoidal* sehingga dapat terlihat perbedaanjarak *Euclidean* dan *Mahalanobis*.

**Data studi kasus**

Data studi kasus yang digunakan pada penelitian ini yaitu pada data *cross section* factor pendidikan di seluruh Indonesia pada tahun 2012/2013 per Kabupaten/Kota. Data diperoleh dari publikasi BPS di setiap provinsi. Adapun variabel yang dianalisis adalah sebagai berikut :

Tabel 3. Deskripsi variabel data studi kasus

|  |  |  |
| --- | --- | --- |
| Simbol | Deskripsi | Sumber |
| (1) | (2) | (3) |
|  | Index Pembangunan Manusia | (Badan Pusat Statistik, 2013) |
|  | Angka Melek Huruf, |
|  | Jumlah SD sederajat |
|  | Jumlah SMP sederajat |
|  | Jumlah SMA sederajat |
|  | Angka Partisipasi Murni untuk umur 7-12. |
|  | Angka Partisipasi Murni untuk umur 13-15. |
|  | Angka Partisipasi Murni untuk umur 16-18. |
|  | Rata – rata lama sekolah. |

## **3.2 Metode Analisis**

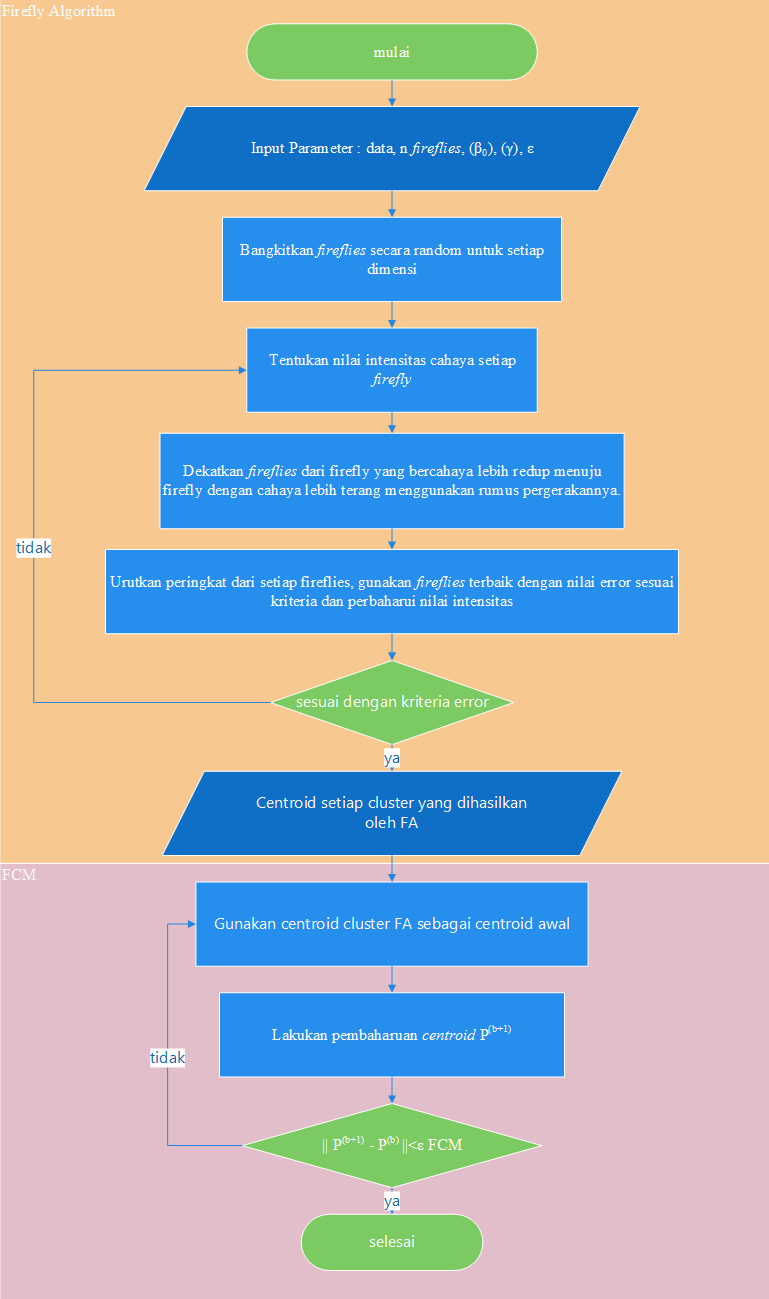
Penelitian ini menggunakan metode FAFCM untuk mengelompokan data bangkitan dan data studi kasus menggunakan dua jarak yang berbeda. Selain itu untuk melihat kefektifan dan kualitas hasil akan dilakukan perbandingan antara FAFCM, FCM, dan PSO-FCM. Cara membandingkan ketiga metode ini akan dilihat dari *Objective value mean* yang akan dibandingkan dengan *objective value best* serta akan digunakan jumlah iterasi setiap metode.

**Pengelompokan data studi kasus dan data bangkitan**

FCM memiliki kelemahan yaitu seringnya terjebak dalam optimum lokal, hal ini dapat diatasi dengan menggunakan metode FAFCM untuk melakukan *clustering*. Pada penelitian ini akan dilakukan proses penentuan titik dan jumlah *centroid* dengan menggunakan FA dimana hasil terbaik (*global best*) *centroid* akan digunakan sebagai titik cluster pada FCM, sehingga FCM tidak akan melakukan penentuan *centroid* secara acak. Berikut adalah tahapan melakukan analisis data menggunakan FAFCM:

1. Menginput parameter:
   1. Data yang akan dilakukan pengelompokan (data bangkitan, data iris, data Pendidikan)
   2. Jumlah *fireflies* yang diinginkan
   3. Nilai *attractiveness* (β0)
   4. Nilai *Light absorption coefficient* (γ)
   5. Kriteria error untuk menghentikan proses atau jumlah iterasi maksimal
2. Membangkitkan n *firefly* secara random dengan setiap “C” *cluster* dan “d” dimensi
3. Melakukan iterasi hingga kriteria error tercapai:
   1. Tentukan nilai intensitas cahaya dari setiap *fireflies* menggunakan fungsi tujuan dari FCM. Ii
   2. Hitung nilai ketertarikan dari setiap *fireflies*. β(r)
   3. Dekatkan *fireflies* dari firefly yang bercahaya lebih redup menuju firefly dengan cahaya lebih terang menggunakan rumus pergerakannya.
   4. Urutkan peringkat dari setiap fireflies, gunakan *fireflies* terbaik dengan nilai error sesuai kriteria dan perbaharui nilai intensitas cahaya untuk dilakukan pengulangan selanjutnya.
4. Setelah mendapatkan optimum global cluster dengan metode *firefly* maka dapat kita gunakan centroid setiap cluster yang akan kita gunakan sebagai centroid pada FCM. (anggap ini sebagai centroid pada iterasi b=0)
5. Lakukan iterasi dibawah ini dan hentikan iterasi jika || P(b+1) - P(b) ||<ε FCM :
   1. Hitung matriks partisi W(b)
   2. Lakukan pembaharuan *centroid* P(b+1)
6. Hasil akan berupa cluster.

Berikut adalah flowchart mengenai metode FAFCM:



## **3.2 Rancangan Penelitian**

Penelitian ini diharapkan dapat menghasilkan artefak berupa perangkat lunak analisis *clustering* sehingga metode penelitian yang digunakan pada riset ini adalah metode *Design Research.* Adapun tahapan yang dilakukan dalam penelitian ini sebagai berikut :

1. *Awareness of Problem*

Pada tahap ini dilakukan identifikasi terhadap masalah yang ada, dengan batasan masalah yang telah ditentukan. Informasi mengenai permasalahan yang ada pada *software* R-Cluzzy diperoleh berdasarkan kajian terhadap pustaka.

1. *Suggestion*

Pada tahap ini dilakukan pendalaman pemahaman terhadap masalah yang ada pada *software* R-Cluzzy, yaitu dengan melakukan analisis terhadap sistem yang sedang berjalan untuk menemukan masalah-masalah yang terdapat pada sistem berjalan serta mengidentifikasi fitur-fitur yang dibutuhkan oleh pengguna.

1. *Development*

Tahap ini dilakukan setelah melalui proses pemilihan solusi teknis alternatif. Teknik pencarian solusi yang digunakan dalam penelitian ini adalah *Easy Solution First* yaitu dengan mengutamakan solusi yang mudah terlebih dahulu. Kemudian akan dilakukan perancangan sistem usulan yaitu:

* Perancangan sistem usulan,
* Perancangan proses,
* Perancangan antarmuka (*user interface*).

1. *Evaluation*

Pada tahap ini dilakukan uji coba dan evaluasi langsung terhadap kemampuan sistem berdasarkan spesifikasi dan rancangan pada tahap *suggestion* oleh pengguna. Uji coba dan evaluasi dilakukan dengan menggunakan pendekatan uji *whitebox* dan uji *blackbox*.*.*

* Uji coba

Pada tahapan ini akan dilakukan uji coba pada aplikasi, yaitu uji coba dilakukan pada setiap komponen yang ada dalam aplikasi. Tujuannya agar dapat diketahui setiap komponen dalam aplikasi dapat berjalan sesuai dengan fungsinya atau tidak.

* Evaluasi

Evaluasi yang akan dilakukan terhadap sistem yang telah selesai dibangun atau dikembangkan memiliki tujuan utama sebagai berikut.

1. Untuk mengetahui apakah terdapat kesalahan pada detail perancangan*.*

Dalam hal ini dilakukan pengecekan dengan menggunakan analisis *whitebox.* Sehingga dengan menggunakan *whitebox*, akan dilakukan pengujian terhadap semua keputusan *logical*, dan seluruh struktur data internal yang menjamin validitas.

1. Untuk mengetahui apakah fungsionalitas yang dibangun sesuai dengan kebutuhan sistem.

Untuk tujuan ini, dilakukan setelah melakukan analisis *whitebox* yaitu dengan menggunakan analisis *blackbox* kepada pengguna sistem.

1. Untuk mengetahui sejauh mana penilaian pengguna terhadap aplikasi yang telah dibangun.

Untuk tujuan ini, dilakukan setelah melakukan analisis *blackbox* yaitu dengan pendekatan *usability testing*.

1. *Conclusion*

Kesimpulan dari hasil evaluasi sistem, apakah sistem telah berhasil atau belum dan apakah masih membutuhkan pengembangan dan perbaikan lebih lanjut.

# Refrensi

Abonyi , J., & Feil, B. (2010). *Cluster Analysis for Data Mining and System Identification.* Boston, Berlin, Jerman: Springer. Retrieved Februari 6, 2016

Alsmadi , M. K. (2014). A Hybrid Firefly Algorithm with Fuzzy-C Mean. *American Journal of Applied Sciences 11 (9): 1676-1691, 2014*, 16.

Ayed, A. B., Halima, M. B., & Alimi, A. M. (2016). Cluster Forests Based Fuzzy C-Means for Data Clustering. *International Joint Conference SOCO’16-CISIS’16-ICEUTE’16*, 564-573.

Badan Pusat Statistik. (2013). Retrieved from Badan Pusat Statistik: http://jatim.bps.go.id/linkTabelStatis/view/id/181

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithm.* New York: Plenum Press. doi: 10.1007/978-1-4757-0450-1

Castillo, O., Rubio, E., Soria, J., & Naredo, E. (2012, Maret). Optimization of the Fuzzy C-Means Algorithm using Evolutionary Methods. *Engineering Letters*, 61-67. Dipetik Januari 20, 2016, dari http://www.engineeringletters.com/issues\_v20/issue\_1/EL\_20\_1\_08.pdf

Cox, E. (1994). *The Fuzzy Systems Handbook .* New York: Academic Press.

Halgamuge, S. K., & Wang, L. (2005). *Classification and Clustering for Knowledge Discovery.* Berlin: Springer.

Haqiqi, B. N. (2015). *Analisis Perbandingan Metode Fuzzy C-Means dan Subtractive Fuzzy C-Means.* Jakarta: STIS.

Hoppner, F., Klawonn, F., Kruse, R., & Runkler, T. (2000). *Fuzzy Cluster Analysis : Method for Classification, Data Analysis, and Image Recognition.* New York: Wiley.

James, C. Bezdek; Ehrlich, Robert; Full, William;. (1984). FCM: The Fuzzy C-Means Clustering Algorithm. *Computers & Geosciences Vol. 10, No. 2-3, pp. 191-203, 1984*, 14.

Kang, J., & Zhang, W. (2011). Combination of Fuzzy C-Means and Particle Swarm Optimization for Text Document Clustering. *Advances in Intelligent and Soft Computing*, 247-252.

Kogan, J. (2007). *Introduction to ClusteringLlarge and High-Dimensional Data.* Cambridge, New York: Cambridge.

Liu, H.-C., Jeng, B.-C., Yih, J.-M., & Yu, Y.-K. (2009). Fuzzy C-Means Algorithm Based on Standard Mahalanobis Distances. *Proceedings of the 2009 International Symposium on Information Processing (ISIP'09)* (pp. 422-427). Huangshan, P. R. China: Academy Publisher. Retrieved Februari 7, 2016

Mordeson, J. N., & Nair, P. S. (2001). *Fuzzy Mathematics : An Introduction for Engineers and Scientists .* New York: Springer-Verlag Berlin Heidelberg .

Nayak, Janmenjoy; Nanda, Matrupallab; Nayak, Kamlesh; Naik, Bighnaraj; Behera, Himansu Sekhar;. (2014). An Improved Firefly Fuzzy C-Means (FAFCM) Algorithm for Clustering Real World Data Sets. *Advanced Computing and Informatics, Networking and Informatics – Volume 1 (ICACNI-2014)*, 354-363.

Rajendran, A., & Dhanasekaran, R. (2012). Brain Tumor Segmentation on MRI Brain Images with Fuzzy Clustering and GVF Snake Model. *Int J Comput Commun*, 530-539.

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining.* London: Pearson.

Wahyono, T. Y., Haryanto, B., Mulyono, S., & Adiwibowo, A. (2010, Agustus). Faktor-faktor yang Berhubungan dengan Kejadian Demam Berdarah dan Upaya Penganggulangannya di Kecamatan Cimanggis, Depok, Jawa Barat. (S. Pangribowo, A. Tryadi, & I. S. Indah, Eds.) *Buletin Jendela Epidemiologi - Demam Berdarah*, pp. 31-43. Retrieved Februari 14, 2016

Xiang, S., Nie, F., & Zhang, C. (2008, Mei 16). Learning a Mahalanobis Distance Metric for Data Clustering and Classification. *Pattern Recognition*, 3600 - 3612. doi:10.1016/j.patcog.2008.05.018

Yang, X.-S. (2010). Firefly Algorithm, Levy Flights and Global Optimization. *Research and Development in Intelligent Systems XXVI*, 209-218.