PENERAPAN ALGORITMA *FUZZY C-MEANS* DALAM PENGELOMPOKKAN DATA TINGKAT BANJIR LIMPASAN

SKRIPSI

Untuk memenuhi sebagian
persyaratan untuk mencapai gelar Sarjana Komputer

Disusun oleh :
MUFLIHANA UTAMI JULY ERNAWATI
NIM. 0910963095

KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
UNIVERSITAS BRAWIJAYA
PROGRAM TEKNOLOGI INFORMASI DAN ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA DAN ILMU KOMPUTER
MALANG

2014

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

LEMBAR PERSETUJUAN

PENERAPAN ALGORITMA *FUZZY C-MEANS* DALAM PENGELOMPOKKAN DATA TINGKAT BANJIR LIMPASAN

SKRIPSI

Disusun oleh :

MUFLIHANA UTAMI JULY ERNAWATI

NIM. 0910963095

Telah diperiksa dan disetujui oleh :

Dosen Pembimbing I,	Dosen Pembimbing II,
Candra Dewi, S.Kom., M.Sc.	**Drs. Marji, M.T.**
NIP. 19771114 200312 2 001	**NIP. 19670801 199203 1 001**

LEMBAR PENGESAHAN

PENERAPAN ALGORITMA *FUZZY C-MEANS* DALAM PENGELOMPOKKAN DATA TINGKAT BANJIR LIMPASAN

SKRIPSI

Diajukan untuk memenuhi persyaratan memperoleh gelar Sarjana Komputer

Disusun Oleh:

MUFLIHANA UTAMI JULY ERNAWATI

NIM.0910963095

Skripsi ini telah diuji dan dinyatakan lulus pada tanggal 7 Januari 2014

Penguji I

Suprapto, S.T., M.T.
NIP.19710727 199603 1 001

Penguji II

Nurul Hidayat, S.Pd., M.Sc.
NIP. 19680430 200212 1 001

Penguji III

Dian Eka Ratnawati, S.Si., M.Kom.
NIP. 19730619 200212 2 001

Mengetahui

Ketua Program Studi Teknik Informatika

Drs. Marji, M.T.
NIP. 19670801 199203 1 001

PERNYATAAN

ORISINALITAS SKRIPSI

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah SKRIPSI ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata didalam naskah SKRIPSI ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia SKRIPSI ini digugurkan dan gelar akademik yang telah saya peroleh (SARJANA) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku. (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, 7 Januari 2014

Mahasiswa,

Muflihana Utami July Ernawati

NIM. 0910963095

repository.ub.ac.id
UNIVERSITAS BRAWIJAYA

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT, karena atas segala rahmat dan limpahan hidayah-Nya, Skripsi yang berjudul **"Penerapan Algoritma Fuzzy C-Means dalam Pengelompokkan Data Tingkat Banjir Limpasan"** ini dapat disusun dengan baik. Skripsi ini disusun dan diajukan sebagai syarat untuk memperoleh gelar sarjana pada Program Studi Informatika / Ilmu Komputer, Program Teknologi Informasi dan Ilmu Komputer, Universitas Brawijaya.

Dalam penyelesaian skripsi ini, penulis telah mendapat begitu banyak bantuan baik moral maupun materiil dari banyak pihak. Atas bantuan yang telah diberikan, penulis ingin menyampaikan penghargaan dan ucapan terima kasih yang sedalam-dalamnya kepada:

1. Candra Dewi, S.Kom., M.Cs., selaku dosen pembimbing utama yang telah meluangkan waktu untuk memberikan pengarahan dan masukan bagi penulis.
2. Drs. Marji, MT., selaku pembimbing kedua yang telah banyak memberikan bimbingan dalam penulisan skripsi ini.
3. Drs. Marji, MT., selaku Ketua Program Studi Teknik Informatika Program Teknologi Informasi & Ilmu Komputer Universitas Brawijaya.
4. Reza Andria Siregar, S.T., selaku Dosen Penasehat Akademik.
5. Ir. Sutrisno, MT, selaku Ketua Program Teknologi Informasi & Ilmu Komputer Universitas Brawijaya.
6. Segenap Bapak dan Ibu dosen yang telah mendidik dan mengajarkan ilmunya kepada Penulis selama menempuh pendidikan di Program Teknologi Informasi & Ilmu Komputer Universitas Brawijaya.
7. Segenap staf dan karyawan di Program Teknik Informatika Program Teknologi Informasi & Ilmu Komputer Universitas Brawijaya yang telah banyak membantu Penulis dalam pelaksanaan penyusunan skripsi ini.
8. Kedua orang tua, suami dan saudara penulis, terima kasih atas semua doa, kasih sayang dan perhatian yang tulus serta dukungan yang telah diberikan.

9. Rekan-rekan Program Studi Informatika/Ilmu Komputer yang telah memberikan dukungannya kepada penulis.
10. Semua pihak yang telah terlibat baik secara langsung maupun tidak langsung yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa skripsi ini tentunya tidak terlepas dari berbagai kekurangan dan kesalahan. Oleh karena itu, segala kritik dan saran yang bersifat membangun sangat penulis harapkan dari berbagai pihak demi penyempurnaan penulisan skripsi ini.

Akhirnya penulis berharap agar skripsi ini dapat memberikan sumbangan dan manfaat bagi semua pihak yang berkepentingan.

Malang, 5 Desember 2013

PENULIS

ABSTRAK

Muflihana Utami July Ernawati. 2014. Penerapan Algoritma *Fuzzy C-Means* Dalam Pengelompokkan Data Tingkat Banjir Limpasan.
Dosen Pembimbing: Candra Dewi, S.Kom., M.Sc. dan Drs. Marji, M.T.

Penelitian ini membahas penerapan algoritma *Fuzzy C-Means* (FCM) dalam pengelompokkan data tingkat banjir limpasan. *Clustering* merupakan metode mengelompokkan suatu objek ke dalam sejumlah kelompok (*cluster*) yang sesuai. Prinsip dari *clustering* adalah meminimumkan kesamaan antar anggota satu *cluster* dan memaksimalkan kesamaan antar anggota *cluster* yang berbeda. *Fuzzy C-Means* merupakan (FCM) merupakan teknik pengklusteran dimana tiap-tiap data ditentukan oleh derajat keanggotaannya. Semakin besar nilai derajat keanggotaan data dalam suatu *cluster* maka semakin besar pula data tersebut menjadi anggota *cluster* tersebut. Hasil *clustering* kemudian diproses atau dievaluasi menggunakan nilai *F-measure* dimana sebelumnya harus diketahui nilai *precission* dan nilai *recall* terlebih dahulu. Pengujian dilakukan pada dua jenis sebaran data yaitu sebaran data yang sama dan sebaran data yang berbeda dengan jumlah dataset yaitu 30 dataset, 90 dataset dan 120 dataset. Dari beberapa pengujian yang dilakukan menunjukkan bahwa nilai *F-measure* tertinggi cenderung berada pada *cluster* 3. Sehingga bisa dilihat bahwa pada pengelompokkan data tingkat banjir limpasan ini memiliki nilai terbaik pada jumlah *cluster* sebanyak 3.

Kata kunci : pengelompokkan data, limpasan, *fuzzy c-means*

ABSTRACT

Muflihana Utami July Ernawati. 2014. Application of Fuzzy C-Means Algorithm In a Runoff Flood Level Data Grouping
Advisor: Candra Dewi, S.Kom., M.Sc. and Drs. Marji, M.T.

This study discussed the application of Fuzzy C-Means algorithm (FCM) to cluster data of runoff flood rate. Clustering is a method of classifying objects into appropriate groups. The principle of clustering is to minimize the similarity between members of the cluster and maximize the similarity between members of different clusters. A Fuzzy C-Means (FCM) is a clustering technique where each of the data is determined by the degree of membership. The greater the degree of membership of data within a cluster, the greater the data become a member of the cluster. Clustering results are then evaluated using F-measure values. Testing is performed on two types of data distribution which is the one have same data distribution and the other haven't. The number of data for each distribution is 30 , 90 and 120. Based on the several tests have been performed showed that the F-measure values tend to be highest in cluster 3. Therefore, it can be concluded that for the clustering of flood runoff rate data, the number of clusters which give the best value is 3.

Keywords: data classification, runoff, fuzzy c-means

DAFTAR ISI

LEMBAR PERSETUJUAN i
LEMBAR PENGESAHAN ii
PERNYATAAN ORISINALITAS SKRIPSI iii
KATA PENGANTAR iv
ABSTRAK vi
ABSTRACT vii
DAFTAR ISI viii
DAFTAR GAMBAR xi
DAFTAR TABEL xii
DAFTAR LAMPIRAN xiii

BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 2
1.3 Batasan Masalah 3
1.4 Tujuan 3
1.5 Manfaat 3
1.6 Sistematika Penulisan 3

BAB II KAJIAN PUSTAKA 5
2.1 Data 5
2.2 Banjir 5
2.3 Limpasan 6
2.3.1 Penutupan Lahan 6
2.3.2 Kemiringan Lereng 7
2.3.3 Infiltrasi Tanah 8
2.3.4 Pola Aliran 8
2.4 *Clustering* 8
2.5 *Fuzzy C-Means* 11
2.6 Evaluasi 14

BAB III METODOLOGI DAN PERANCANGAN 16
3.1 Deksripsi Data 17
3.2 Deskripsi Umum 17
3.3 Perancangan 17
3.3.1. *Clustering Fuzzy C-Means* 18
3.3.1.1 Proses Pembentukan Matriks Awal U 20
3.3.1.2 Proses Perhitungan Pusat *Cluster* 21
3.3.1.3 Proses Perhitungan Fungsi Objektif 22
3.3.1.4 Proses Perubahan Matriks Partisi U 23
3.3.2. Pengujian *F-Measure* 24
3.4 Perhitungan Manual 25
3.5 Rancangan Antarmuka 34
3.6 Rancangan Uji Coba 36

BAB IV IMPLEMENTASI 37
4.1 Lingkungan Implementasi 37
4.1.1 Lingkungan Implementasi Perangkat Keras 37
4.1.2 Lingkungan Implementasi Perangkat Lunak 37
4.2 Implementasi Program 37
4.2.1 Implementasi metode *Fuzzy C-Means* 37
4.2.1.1. Penentuan Bilangan Random 38
4.2.1.2. Penentuan Nilai Derajat Keanggotaan Awal 38
4.2.1.3. Perhitungan Pusat Cluster 39
4.2.1.4. Perhitungan Fungsi Objektif 42
4.2.1.5. Perhitungan Nilai Perubahan Matriks Partisi 46
4.2.1.6. Proses Pengelompokkan 47
4.2.1.7. Evaluasi 48
4.3 Implementasi Antar Muka 52
4.4 Implementasi Pengujian 55

BAB V PENGUJIAN DAN ANALISIS 56
5.1 Skenario Evaluasi 56

5.2 Hasil Pengujian 56
5.2.1 Pengujian dengan Sebaran Data yang Sama 57
5.2.1.1 Pengujian Terhadap 30 Data 57
5.2.1.2 Pengujian Terhadap 90 Data 58
5.2.1.3 Pengujian Terhadap 120 Data 59
5.2.2 Pengujian dengan Sebaran Data yang Berbeda 60
5.2.2.1 Pengujian Terhadap 30 Data 60
5.2.2.2 Pengujian Terhadap 90 Data 61
5.2.2.3 Pengujian Terhadap 120 Data 62
5.3 Analisa *F-Measure* terhadap Jumlah *Cluster* 63
5.4 Analisa Jumlah Data terhadap Nilai *F-Measure* 64

BAB VI PENUTUP 66
6.1 Kesimpulan 66
6.2 Saran 66

DAFTAR PUSTAKA 67

LAMPIRAN 71
Lampiran 1. Hasil Akurasi *F-Measure* 71
Lampiran 2. Contoh Data Uji 76
Lampiran 3. Contoh Hasil Pengelompokkan pada 30 Data 77

DAFTAR GAMBAR

Gambar 2.1 Dendogram pada *Hierarchical Clustering*..... 10
Gambar 2.2 (a) Pengelompokkan *Hard K-Means*..... 11
Gambar 2.2 (b) Pengelompokkan *Fuzzy C-Means*..... 11
Gambar 3.1 Diagram Alur Penelitian 16
Gambar 3.2 Alur Perancangan 18
Gambar 3.3 *Flowchart Fuzzy C-Means Clustering*..... 19
Gambar 3.4 *Flowchart* Pembentukan Matriks Partisi Awal U 20
Gambar 3.5 *Flowchart* Perhitungan Pusat Cluster 22
Gambar 3.6 *Flowchart* Perhitungan Fungsi Objektif 23
Gambar 3.7 *Flowchart* Perubahan Matriks Partisi U 24
Gambar 3.8 *Flowchart* Perhitungan F-Measure..... 25
Gambar 3.9 Rancangan Antarmuka Nilai *Input*..... 34
Gambar 3.10 Rancangan Antarmuka *Fuzzy C-Means*..... 35
Gambar 4.1 *Source Code* Penentuan Bilangan Random 38
Gambar 4.2 *Source Code* Penentuan Nilai Derajat Awal..... 38
Gambar 4.3 *Source Code* Perhitungan Pusat *Cluster* 39
Gambar 4.4 *Source Code* Perhitungan Fungsi Objektif 43
Gambar 4.5 *Source Code* Perhitungan Derajat Keanggotaan Baru..... 46
Gambar 4.6 *Source Code* Nilai Derajat Baru..... 47
Gambar 4.7 *Source Code* Proses Pengelompokkan..... 47
Gambar 4.8 *Source Code* Evaluasi *F-Measure* 48
Gambar 4.9 Implementasi Antarmuka Nilai *Input*..... 52
Gambar 4.10 Implementasi Antarmuka Perhitungan *Fuzzy C-Means* 53
Gambar 4.11 Implementasi Antarmuka *Result* 54
Gambar 5.1 Grafik Pengaruh Nilai Akurasi dengan Jumlah *Cluster*
(Jumlah Sebaran Data yang Sama) 60
Gambar 5.1 Grafik Pengaruh Nilai Akurasi dengan Jumlah *Cluster*
(Jumlah Sebaran Data yang Berbeda)..... 63

DAFTAR TABEL

Tabel 2.1 Tabel Kelas Penutupan Lahan 7
Tabel 2.2 Tabel Penilaian Kelas Lereng dan Faktor LS 7
Tabel 2.3 Tabel Kategori Nilai Pola Aliran beserta *Skoring* 8
Tabel 2.4 Tabel Matriks *Confusion* 14
Tabel 3.1 Tabel Sampel Dataset 25
Tabel 3.2 Tabel Sampel Dataset setelah diindexkan 26
Tabel 3.3 Tabel Bilangan *Random* dan Nilai Qi 27
Tabel 3.4 Tabel Contoh Perhitungan Pusat *Cluster* 1 28
Tabel 3.5 Tabel Detail Perhitungan Fungsi Objektif 29
Tabel 3.6 Tabel Perhitungan Perbaikan Matriks 29
Tabel 3.7 Tabel Hasil *Clustering* 31
Tabel 3.8 Tabel Pencocokan *Cluster* dan *Class* 32
Tabel 3.9 Tabel *classID* paling Dominan 32
Tabel 3.10 Rincian *True Positive*, *False Negative*, *False Positive* tiap *Cluster* ... 33
Tabel 3.11 Tabel Hasil Perhitungan *Precission*, *Recall* dan *F-Measure* 34
Tabel 3.12 Tabel Pengujian Evaluasi 36
Tabel 5.1 Hasil Pengujian Pertama 56
Tabel 5.2 Tabel Pengujian Akurasi *F-Measure* terhadap Jumlah *Cluster* (30 data) 57
Tabel 5.3 Tabel Pengujian Akurasi *F-Measure* terhadap Jumlah *Cluster* (90 data) 58
Tabel 5.4 Tabel Pengujian Akurasi *F-Measure* terhadap Jumlah *Cluster* (120 data) 59
Tabel 5.5 Tabel Pengujian Akurasi *F-Measure* terhadap Jumlah *Cluster* (30 data) 61
Tabel 5.6 Tabel Pengujian Akurasi *F-Measure* terhadap Jumlah *Cluster* (90 data) 61
Tabel 5.7 Tabel Pengujian Akurasi *F-Measure* terhadap Jumlah *Cluster* (120 data) 62

DAFTAR LAMPIRAN

Lampiran 1 Hasil Akurasi *F-Measure* 71
Lampiran 2 Contoh Data Uji 76
Lampiran 3 Contoh Hasil Pengelompokkan pada 30 Data 77

BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi yang pesat sekarang ini membawa banyak perubahan dalam pola kehidupan masyarakat. Sekarang teknologi informasi juga menunjang dalam penyelesaian pekerjaan secara cepat, efektif dan efisien. Misalnya saja pada pengelompokkan data atau disebut juga dengan *Clustering* [DWI - 2012].

Clustering merupakan metode mengelompokkan suatu objek ke dalam sejumlah kelompok (*cluster*) yang sesuai. Menurut [HAN - 2001] prinsip dari clustering adalah memaksimalkan kesamaan antar anggota satu cluster dan meminimumkan kesamaan antar anggota cluster yang berbeda.

Terdapat dua jenis metode *clustering* yang sering digunakan dalam proses pengelompokkan data yaitu *hierarchical* (jumlah kelompok belum ditentukan) dan *non-hierarchical* (jumlah kelompok sudah ditentukan). Proses pengelompokkan secara *hierarchical* digunakan apabila belum ada informasi jumlah kelompok sedangkan metode *non-hierarchical* digunakan untuk mengelompokkan beberapa objek ke dalam suatu kelompok. Pengelompokkan *non-hierarchical* yang telah banyak digunakan adalah *C-Means*. Berdasarkan pengalokasian data ke dalam kelompok (*cluster*), *C-Means* dibedakan menjadi *Hard C-Means* dan *Fuzzy C-Means* [AGU - 2007]. *Hard C-Means* mengalokasikan data kembali ke pusat *cluster* secara tegas, sedangkan *Fuzzy C-Means* merupakan suatu teknik pengelompokkan yang mana keberadaan tiap titik data dalam suatu kelompok (*cluster*) ditentukan oleh derajat keanggotaannya. Penempatan posisi data pada *cluster* dilakukan dengan perbaikan penentuan pusat *cluster* awal dan nilai keanggotaan secara berulang [BEZ - 1983].

Pada skripsi ini dengan memanfaatkan metode untuk pengelompokkan data, akan dikembangkan untuk membangun aplikasi pada pengelompokkan data tingkat banjir limpasan pada suatu DAS (Daerah Aliran Sungai). Pengelompokkan data ini menggunakan variabel-variabel yang ada seperti kemiringan lereng,

indeks jenis tanah, penutupan lahan, dan pola aliran. Adanya kesimpulan data yang berupa tingkatan limpasan banjir akan memudahkan dalam penentuan evaluasi hasil cluster.

Pengelompokkan data banjir limpasan belum banyak diterapkan dalam suatu metode. Namun berbagai riset menggunakan *Fuzzy C-Means* telah diimplementasikan menggunakan beberapa macam kasus data. Salah satunya penelitian Roma Dhona [DHO - 2010] yang meneliti perbandingan metode *Fuzzy C-Means* dengan *Hard C-Means* yang diterapkan pada analisis gerombol. Hasil penelitian didapatkan bahwa *Fuzzy C-Means* memiliki tingkat kesalahan lebih kecil dalam mengelompokkan data daripada *Hard C-Means.* Selain itu terdapat penelitian Neny Septiana [SEP - 2013] yang menerapkan algoritma *Fuzzy C-Means* dalam pengelompokkan berita berbahasa Indonesia. Hasil penelitian didapatkan bahwa *cluster* terbaik yang dihasilkan mencapai akurasi *f-measure* sebesar 0,63.

Pada penelitian ini mencoba mengimplementasikan metode *Fuzzy C-Means* terhadap sejumlah data tingkat banjir limpasan yang didapat dari BP DAS Brantas Jawa Timur. Kecenderungan data untuk memasuki sebuah *cluster* ditentukan oleh besarnya nilai derajat keanggotaan data tersebut pada setiap *cluster*. Semakin besar derajat keanggotaan data tersebut maka semakin besar kecenderungan data tersebut memasuki *cluster*. Input yang digunakan dalam pengelompokkan adalah dataset tingkat banjir limpasan. Perhitungan nilai akurasi dihitung dengan mengubah nilai inputan jumlah kelompok (*cluster*) setelah didapatkan nilai yang *konvergen*, dimana kondisi fungsi objektif cenderung untuk tidak berubah-ubah.

Berdasarkan latar belakang yang telah dipaparkan, maka judul yang diambil pada skripsi ini adalah ***"Penerapan Algoritma Fuzzy C-Means dalam Pengelompokan Data Tingkat Banjir Limpasan"***.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka rumusan masalah dalam penelitian ini adalah :

1. Bagaimana menerapkan metode *Fuzzy C-Means* untuk pengelompokkan data tingkat banjir limpasan?

2. Berapa banyaknya *cluster* (kelompok) terbaik yang terbentuk pada pengelompokkan dengan *Fuzzy C-Means*?

1.3 Batasan Masalah

Agar pembahasan penelitian ini tidak menyimpang dari apa yang telah dirumuskan, maka diperlukan batasan-batasan. Batasan-batasan dalam penelitian ini adalah :

1. Data parameter yang digunakan pada perhitungan berupa parameter karakteristik DAS (Daerah Aliran Sungai) antara lain : kemiringan lereng, indeks jenis tanah, penutupan lahan dan pola aliran.
2. Data yang digunakan untuk pengelompokkan adalah data parameter karakteristik DAS (Daerah Aliran Sungai) BP DAS Brantas Jawa Timur.
3. Indeks kekaburan *fuzzy* (w) yang digunakan dalam pengujian adalah 2.

1.4 Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah :

1. Untuk menerapkan metode *Fuzzy C-Means* dalam pengelompokkan data tingkat banjir limpasan.
2. Untum menghitung jumlah cluster terbaik dan melakukan analisa dari proses penerapan algoritma *Fuzzy C-Means* pada pengelompokkan data tingkat banjir limpasan.

1.5 Manfaat

Manfaat yang bisa diambil dari penyusunan skripsi ini adalah menghasilkan program yang menerapkan algoritma *Fuzzy C-Means* yang dapat memudahkan dalam proses pengelompokkan data tingkat banjir limpasan.

1.6 Sistematika Penulisan

Sistematika penulisan dalam skripsi ini sebagai berikut :

BAB I Pendahuluan

Bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian serta sistematika penulisan skripsi.

BAB II Kajian Pustaka

Bab ini berisi uraian teori – teori yang berhubungan dengan banjir beserta paramernya dan *Fuzzy C-Means.*

BAB III Metodologi dan Perancangan

Bab ini berisi penjelasan mengenai metode dan perancangan yang digunakan untuk pengelompokkan data tingkat banjir limpasan serta terdapat contoh perhitungan manual dari proses-proses yang ada.

BAB IV Implementasi

Bab ini berisi hasil dari implementasi yang digunakan dalam penelitian ini.

BAB V Pengujian dan Analisis

Bab ini berisi pengujian dari implementasi yang telah dibuat serta menganalisa hasil yang dihasilkan oleh implementasi program.

BAB VI Penutup

Bab ini berisi kesimpulan dari seluruh rangkaian penelitian serta saran kemungkinan pengembangan.

BAB II
KAJIAN PUSTAKA

2.1 Data

Menurut [LEE - 2010], data dan informasi memiliki pengertian yang berbeda. Data dikatakan sebagai bahan mentah dari informasi. Data merupakan kumpulan kejadian nyata yang dapat dijamin kebenarannya. Data dapat dikelompokkan ke dalam 2 jenis, yaitu data kategori dan data numerik. Data kategori dan numerik dapat digunakan untuk kebutuhan statistika.

Data kategori adalah data yang mengambil suatu himpunan nilai tertentu. Sebagai contoh, warna mata, negara, merk mobil, dan segala tipe kategori yang tidak berupa numerik (data kategori tidak dapat dibandingkan lebih besar maupun lebih kecil). Ahli statistik biasa menyebut data kategori sebagai data kualitatif.

Data numerik adalah data yang berbentuk angka, seperti tinggi dan jumlah penduduk di suatu negara. Data numerik kebanyakan didefinisikan sebagai kuantitatif, dan dapat pula berupa "kontinyu", seperti waktu atau pendapatan, dapat pula berupa "diskrit". Amat memungkinkan mengubah data numerik menjadi data kategori. Sebagai contoh, jumlah anak dari suatu keluarga dapat digabung menjadi 3 kategori : tidak punya anak, 1-4, 5 atau lebih.

2.2 Banjir

Banjir adalah tinggi muka air melebihi normal pada sungai dan biasanya mengalir meluap melebihi tebing sungai dan luapan airnya menggenang pada suatu daerah genangan [HAD - 2011]. Selain itu, banjir menjadi masalah dan berkembang menjadi bencana ketika banjir tersebut mengganggu aktivitas manusia dan bahkan membawa korban jiwa dan harta benda [SOB - 2009].

Banjir di suatu tempat bisa berbeda-beda tergantung dari kondisi fisik wilayah. Dalam hal ini, ada yang mengalami banjir lokal, banjir kiriman (limpasan), maupun banjir rob [RIS - 2012].

Adapun penjelasan dari kejadian banjir tersebut antara lain :

1. Banjir Lokal

   Banjir lokal disebabkan oleh tingginya intensitas hujan dan belum tersedianya sarana drainase memadai.

2. Banjir Kiriman (Limpasan)

   Banjir kiriman disebabkan oleh peningkatan debit air sungai yang mengalir. Banjir ini diperparah oleh kiriman dari daerah atas.

3. Banjir Rob

   Banjir ini disebabkan oleh tingginya pasang surut air laut yang melanda daerah pinggiran laut atau pantai.

2.3 Limpasan

Salah satu komponen dalam siklus hidrologi adalah limpasan. Komponen limpasan dapat berupa *runoff* (aliran permukaan) ataupun aliran yang lebih besar seperti aliran di sungai. Limpasan akibat hujan ini dapat terjadi dengan cepat dan dapat pula setelah beberapa jam setelah terjadinya hujan. Lama waktu kejadian hujan puncak dan aliran puncak sangat dipengaruhi oleh kondisi wilayah tempat jatuhnya hujan. Makin besar perbedaan waktu kejadian hujan puncak dan debit puncak, makin baik kondisi wilayah tersebut dalam menyimpan air di dalam tanah [ANO - 2013].

Parameter-parameter yang digunakan untuk menghitung nilai limpasan berupa karakteristik fisik lahan dalam DAS (Daerah Aliran Sungai) yaitu penutupan lahan, kemiringan lereng, infiltrasi tanah dan pola aliran [MUR - 2010].

2.3.1 Penutupan Lahan

Penutup/penggunaan lahan merupakan salah satu informasi yang sering diturunkan dari data penginderaan jauh [MUR - 2010]. Kelas penutupan lahan adalah 23 katagorisasi areal terdiri dari 7 kelas hutan dan 15 kelas bukan hutan, serta kelas tertutup awan. Sebanyak 23 kelas yang digunakan merupakan standar kelas penutupan lahan kehutanan. Kelas ini, menghasilkan klasifikasi penafsiran yang telah terbangun dari puluhan tahun pengalaman memantau hutan [ANO - 2012]. Kelas penutupan lahan ditunjukkan pada tabel 2.1.

Tabel 2.1 Tabel Kelas Penutupan Lahan

Kode	Kelas Tutupan Vegetasi	Kelas Penutupan Lahan
0	Water	Rawa, Tambak, Air/Sungai/Danau/Waduk
1	Grassland/Cereal crops	Sawah, Ladang/Tegalan
2	Shrubland	Belukar/Semak
3	Broadleaf crop	Perkebunan, Ladang/Tegalan
4	Savannah	Belukar/Semak, Ladang/Tegalan, Lahan Terbuka
5	Broadleaf forest	Hutan primer, Hutan sekunder, Hutan mangrove, Hutan gambut, Hutan rawa
6	Needle leaf forest	-
7	Barren	Lahan Terbuka, Pemukiman kota, Pemukiman Kampung, Industri, Bandara, Pelabuhan

[SUW – 2010]

2.3.2 Kemiringan Lereng

Kemiringan lereng merupakan salah satu parameter yang memiliki peran besar cukup besar pada berbagai proses hidrologi [MUR - 2010]. Departemen Kehutanan dalam Rencana Teknik Lapangan, Rehabilitasi Lahan dan Konservasi Tanah telah menyusun penilaian kelas lereng dan faktor LS (panjang lereng) berdasarkan tabel 2.2 [PRA - 2011].

Tabel 2.2 Penilaian kelas lereng dan faktor LS

Kelas Lereng	Kemiringan Lereng(%)	LS
I	0-8	0,4
II	8-15	1,4
III	15-25	3,1
IV	25-40	6,8
V	>40	9,5

2.3.3 Infiltrasi Tanah

Infiltrasi tanah adalah salah satu faktor yang mempengaruhi nilai koefisien limpasan permukaan yang merupakan proses masuknya air ke dalam tanah secara vertikal. Data mengenai infiltrasi tanah sulit diketahui secara langsung melalui citra penginderaan jauh, namun hal ini dapat dilakukan dengan menggunakan pendekatan terhadap jenis tanah dan tekstur tanah [MUR - 2010].

2.3.4 Pola Aliran

Timbunan air di permukaan (pola aliran) terbentuk apabila laju presipitasi terus berlangsung dan melebihi laju infiltrasi. Timbunan air permukaan diperoleh berdasarkan interpretasi kerapatan aliran dengan asumsi bahwa semakin tinggi kerapatan aliran maka proses pengeringan semakin cepat dan juga sebaliknya, jika kerapatan aliran rendah makan daerah tersebut selalu mengalami genangan dan drainasenya digolongkan jelek. Apabila lapisan tanah telah dijenuhi oleh kelebihan air hujan makan akan membentuk aliran permukaan [MUR - 2010]. Kategori nilai pola aliran beserta *skoring*nya ditunjukkan pada tabel 2.3.

Tabel 2.3 Tabel kategori nilai pola aliran beserta skoring

Parameter	Besaran	Kategori Nilai	Skor
Timbunan di permukaan (pola aliran)	Selalu tergenang	Tinggi	5
	Dijumpai depresi permukaan, danau dan rawa	Normal	10
	Sistem saluran cukup baik	Rendah	15
	Pengeringan terlalu cepat	Diabaikan	20

Sumber : Ven. T. Chow, 1964

[SAN - 2011]

2.4 *Clustering*

Clustering adalah suatu metode pengelompokkan berdasarkan ukuran kedekatan (kemiripan). *Cluster* merupakan pola yang terbentuk dalam suatu proses pembagian sekelompok data ke dalam sejumlah *sub*-kelas. *Clustering* dapat digunakan untuk memberikan label suatu data yang belum diketahui

kelasnya, oleh karena itu clustering digolongkan sebagai *unsupervised learning* [HAN-2000].

Prinsip dari *cluster*ing adalah memaksimalkan kesamaan anggota dalam setiap kelompok (*cluster*) dan meminimalkan jarak antara pusat *cluster* dengan *cluster* lain. Perhitungan jarak tersebut digunakan untuk mengukur kemiripan data [ZAL - 2007].

Algoritma *Cluster*ing dibedakan menjadi lima, diantaranya yaitu [HAN - 2000] :

1. *Partitioning Algorithms (Non-Hierarchical)*

   Algoritma ini melakukan pemecahan pada sekelompok data dan menjadikannya beberapa kelompok kemudian dilakukan evaluasi dengan beebrapa kriteria.

2. *Hierarchy Algorithms*

   Algoritma ini digunakan untuk memproses pemecahan data dengan melakukan penggabungan kelompok-kelompok data yang kecil menjadi suatu kelompok besar sesuait tingkat kesamaan. Hasil dari algoritma ini berupa dendogram yang menggambarkan bagaimana data terelasi pada suatu *cluster*.

3. *Density Based*

   Algoritma *density based*, proses pembentukan *cluster* berdasarkan pada konektifitas dan kepadatan data. Sebuah *cluster* digambarkan dengan suatu area titik yang padat yang dipisahkan dengan area kepadatan yang rendah dari area kepadatan tinggi lainnya.

4. *Grid Based*

   Pembentukan *cluster* pada algoritma *Grid Based* berdasarkan pada struktur multiple *level granularity*.

5. *Model Based*

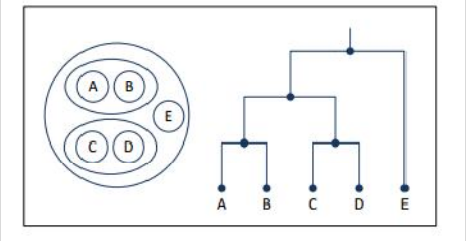
   Algoritma ini menggunakan suatu model sebagai analisa masing-masing *cluster* dan model yang baik dipilih model hipotesa tersebut.

Sedangkan berdasarkan hasil solusinya, algoritma *cluster*ing dibedakan menjadi *hierarchical* dan non *hierarchical clustering*. *Hierarchical clustering* adalah pengelompokkan dokumen menjadi kelompok-kelompok yang mempunyai suatu hirarki yaitu dari kelompok kecil menuju kelompok besar atau sebaliknya.

Sedangkan *non-hierarchical* adalah pengelompokkan dengan membaca seluruh dokumen kemudian mengelompokkan berdasarkan kedekatannya secara langsung [MAN - 2008]. Perbedaan keduanya adalah :

1. *Hierarchical Clustering*
   a. Jarak dihitung untuk semua variabel sehingga dapat menunjukkan antar variabel.
   b. Kurang handal dalam menangani jumlah data yang besar karena harus menghitung jarak satu persatu.
   c. Penggunaan waktu cukup banyak dengan jumlah anggota besar.
2. *Non Hierarchical Clustering*
   a. Algoritma lebih cepat karena tidak melakukan perhitungan jarak terhadap semua variabel.
   b. Tidak menunjukkan semua relasi antar variabel yang disebabkan jarak hanya dihitung dari titik ke pusat *cluster*.
   c. Lebih handal dalam menangani jumlah data besar.

Prinsip utama dalam *hierarchical cluster* adalah mengatur seluruh objek ke dalam sebuah pohon berdasarkan suatu kriteria tertentu. Pohon itu disebut dengan *dendogram* yang ditunjukkan pada gambar 2.1. Semakin tinggi level dari simpul pohon, maka semakin rendah similaritas antar objeknya. Objek yang memiliki kesamaan berada pada kelompok awal atau simpul pohon paling bawah.



Gambar 2.1 Dendogram pada Hierarchical Clustering

Sedangkan pengelompokkan *non-hierarchical clustering* atau *partitional clustering* digunakan untuk mengelompokkan sejumlah objek ke dalam suatu kelompok, sehingga dapat dikatakan jumlah kelompok pada metode ini ditentukan oleh pengguna. *Non-hierarchical clustering* mempunyai 2 algoritma :

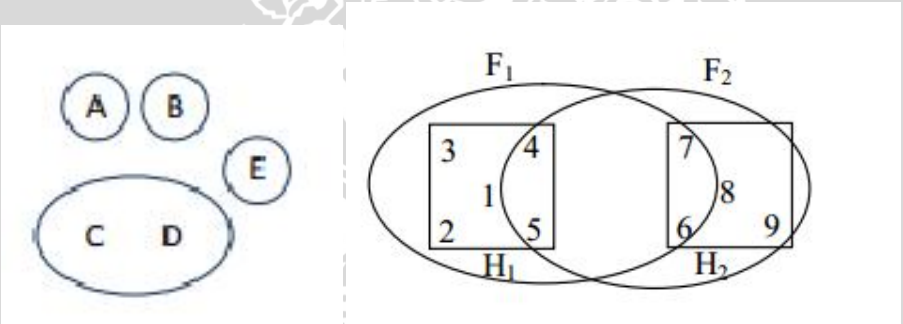
a. *Single Pass*

Algoritma ini melewatkan data dari suatu kumpulan data sekali saja untuk menciptakan *cluster*.

b. Metode *K-Means*

Melewatkan data berulanh kali untuk mengadakan pengujian terhadap pusat *cluster* dan menentukan pengelompokkan terbaik dengan mendefinisikan jumlah kelompok terlebih dahulu.

Salah satu metode pengelompokkan *non-hierarchical* adalah *k-means clustering* dimana metode ini bertujuan untuk mengelompokkan objek dengan meminumumkan jarak tiap-tiap objek ke pusat *cluster* secara tegas. Dalam perkembangannya terdapat pengelompokkan menggunakan *fuzzy clustering* yaitu dengan memperhatikan nilai dari tiap-tiap data dengan derajat keanggotaan. Perbedaan hasil pengelompokkan dengan menggunakan *non hierarchical clustering* dengan menggunakan *hard k-means* dan secara *fuzzy c-means* ditunjukkan pada gambar 2.2.



(a) (b)

Gambar 2.2 (a) Pengelompokkan *Hard K-Means* (b) Pengelompokkan *Fuzzy C-Means*

2.5 Fuzzy C-Means

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 [KUS-2010].

Metode *clustering* merupakan pengelompokkan beserta parameternya dalam kelompok-kelompok sesuai kecenderungan sifat masing-masing data tersebut

(kesamaan sifat). Ada beberapa algoritma *clustering* data, salah satu diantaranya adalah *Fuzzy C-Means*. *Fuzzy C-Means* adalah suatu teknik pengklusteran dimana keberadaan tiap-tiap titik data dalam cluster ditentukan oleh derajat keanggotaan [MIY - 2008].

Fuzzy C-Means (FCM) merupakan salah satu algoritma *fuzzy clustering*. *Fuzzy C-Means (FCM)* merupakan teknik pengklusteran dimana tiap-tiap data ditentukan oleh derajat keanggotaannya. Tujuan dari algoritma *Fuzzy C-Means* adalah untuk menemukan pusat *cluster* (*centroid*) dengan meminimumkan fungsi objektif [BEZ - 1983].

Konsep dasar FCM, pertama kali adalah menentukan pusat *cluster*, yang akan menandai lokasi rata-rata untuk tiap-tiap *cluster*. Pada kondisi awal, pusat *cluster* ini masih belum akurat. Tiap-tiap titik data memiliki derajat keanggotaan untuk tiap-tiap *cluster*. Dengan cara memperbaiki pusat *cluster* dan derajat keanggotaan tiap-tiap titik data secara berulang, maka akan dapat dilihat bahwa pusat *cluster* akan bergerak menuju lokasi yang tepat. Perulangan ini didasarkan pada minimisasi fungsi objektif yang menggambarkan jarak dari titik data yang diberikan ke pusat *cluster* yang terbobot oleh derajat keanggotaan titik data tersebut [KUS – 2010].

Output dari FCM bukan merupakan *fuzzy inference system*, melainkan deretan pusat *cluster* dan beberapa derajat keanggotaan untuk setiap titik data. Informasi ini dapat digunakan untuk membangun suatu *fuzzy inference system*. Algoritma *Fuzzy C-Means* (FCM) adalah sebagai berikut [KUS – 2010]:

1. *Input* data yang akan di*cluster* X, berupa matriks berukuran n x m (n= jumlah sampel data, m= atribut setiap data). Xij = data sampel ke-i (i=1,2,…,n), atribut ke-j (j=1,2,…,m).
2. Tentukan :
   - Jumlah *cluster* = c;
   - Pangkat = w;
   - Maksimum Iterasi = MaxIter;
   - *Error* terkecil yang diharapkan = ξ ;
   - Fungsi objektif awal = $P_0 = 0$;

- Iterasi Awal = t = 1;

3. Bangkitkan bilangan random μ_{ik}, i=1,2,…,n; k=1,2,…,c; sebagai elemen-elemen matriks partisi awal U. Kemudian hitung jumlah setiap kolom.

   Untuk menghitung jumlah setiap kolom digunakan persamaan 2.1 :

$$Qi = \sum_{k=1}^{c} \mu ik \tag{2.1}$$

   dengan j=1,2,…,n.

   Untuk menghitung nilai derajat keanggotaan awal digunakan persamaan 2.2 :

$$\mu ik = \frac{\mu ik}{Qi} \tag{2.2}$$

4. Hitung pusat *cluster* ke-k: Vkj, dengan k=1,2,…,c; dan j=1,2,…,m.

   Untuk menghitung pusat *cluster* digunakan persamaan 2.3 :

$$Vkj = \frac{\sum_{i=1}^{n} ((\mu ik)^{w} * Xij)}{\sum_{i=1}^{n} (\mu ik)^{w}} \tag{2.3}$$

5. Hitung fungsi objektif pada iterasi ke-t, P1.

   Untuk menghitung fungsi objektif digunakan persamaan 2.4 :

$$Pt = \sum_{i=1}^{n} \sum_{k=1}^{c} \left(\left[\sum_{j=1}^{m} (Xij - Vkj)^2 \right] (\mu ik)^{w} \right) \tag{2.4}$$

6. Hitung perbaikan matriks partisi.

   Untuk menghitung perbaikan matriks partisi digunakan persamaan 2.5 :

$$\mu ik = \frac{\left[\sum_{j=1}^{m} (Xij - Vkj)^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c} \left[\sum_{j=1}^{m} (Xij - Vkj)^2 \right]^{\frac{-1}{w-1}}} \tag{2.5}$$

7. Cek kondisi berhenti :
   - Jika (|Pt – Pt-1| < ξ) atau (t > MaxIter) maka berhenti;
   - Jika tidak: t = t+1, ulangi langkah ke-4.

2.6 Evaluasi

Pada proses peng*cluster*an, prosedur untuk mengevaluasi hasilnya dikenal sebagai validitas *cluster* dan dapat dilakukan dengan berbagai macam pengukuran yang disebut dengan pengukuran validitas *cluster* [PET - 2000].

Pegukuran validitas dapat dibagi menjadi dua kategori, dimana pembagian kategori tersebut berdasarkan pada ada tidaknya referensi pengetahuan yang dapat digunakan untuk membandingkan hasil pengelompokkan yang dilakukan oleh sistem. Kategori yang pertama adalah mengevaluasi seberapa bagus *cluster* dengan cara membandingkan kelompok-kelompok yang dihasilkan oleh sistem dengan data kelompok yang sudah diketahui kelas-kelasnya, dimana kategori ini disebut dengan pengukuran kualitas eksternal. Sedangkan kategori yang kedua adalah membandingkan sejumlah *cluster* tanpa adanya referensi pengetahuan yang disebut dengan pengukuran kualitas internal. Contoh pengukuran kualitas eksternal adalah *Entropy* dan *F-Measure* [STE - 2000].

Pengukuran validitas pada penelitian ini menggunakan *F-Measure* karena sudah ada data pembanding eksternal yang dapat digunakan dalam evaluasi. Untuk menguji akurasi dari hasil *F-Measure*, digunakan suatu standar yang disebut matriks *confusion*. Matriks *confusion* berisi informasi mengenai pengelompokkan yang sebenarnya dan prediksi pengelompokkan yang dilakukan oleh sistem [HAM - 2003]. Matriks *confusion* ditunjukkan pada tabel 2.4. [LEW – 1995].

Tabel 2.4 Matriks *Confusion*

		Pengklasteran Sebenarnya	
		Klaster Sama	**Klaster Beda**
Pengklasteran Sistem	**Kelas Sama**	*True Positive*	*False Negative*
	Kelas Beda	*False Positive*	*True Negative*

1. True Positive menunjukkan bahwa dokumen yang termasuk dalam hasil pengelompokkan oleh sistem memang merupakan anggota *cluster*.

2. False Positive menunjukkan bahwa dokumen yang termasuk dalam hasil pengelompokkan oleh sistem ternyata seharusnya bukan merupakan anggota *cluster*.
3. False Negative menunjukkan bahwa dokumen yang tidak termasuk dalam hasil pengelompokkan oleh sistem ternyata seharusnya merupakan anggota *cluster*.
4. True Negative menunjukkan bahwa dokumen yang tidak termasuk dalam hasil pengelompokkan oleh sistem ternyata seharusnya bukan merupakan anggota *cluster*.

Ada beberapa standar pengukuran yang digunakan dalam pengelompokkan dokumen, diantaranya adalah *recall*, *precision* dan *F-measure*.

Recall adalah tingkat keberhasilan mengenali suatu *event* dari seluruh *event* yang seharusnya dikenali. Dari pengertian tersebut, *recall* berarti ukuran dari jumlah dokumen benar yang berhasil dikelompokkan oleh sistem dari keseluruhan dokumen yang seharusnya benar [TAL - 2003]. Cara mencari nilai *recall* ini adalah dengan membagi jumlah dokumen hasil pengelompokkan sistem yang benar dengan jumlah dokumen yang sebenarnya dalam *cluster* tersebut. Untuk menghitung nilai *recall* ditunjukkan pada persamaan 2.6.

$$recall = \frac{TP}{TP + FN} \tag{2.6}$$

[FUA – 2013]

Sedangkan *precision* adalah tingkat ketepatan hasil pengelompokkan terhadap suatu *event*. Artinya, dari seluruh dokumen hasil pengelompokkan, berapa persenkah yang dinyatakan benar [TAL – 2003]. Untuk menghitung nilai *precision* ditunjukkan pada persamaan 2.7.

$$precision = \frac{TP}{TP + FP} \tag{2.7}$$

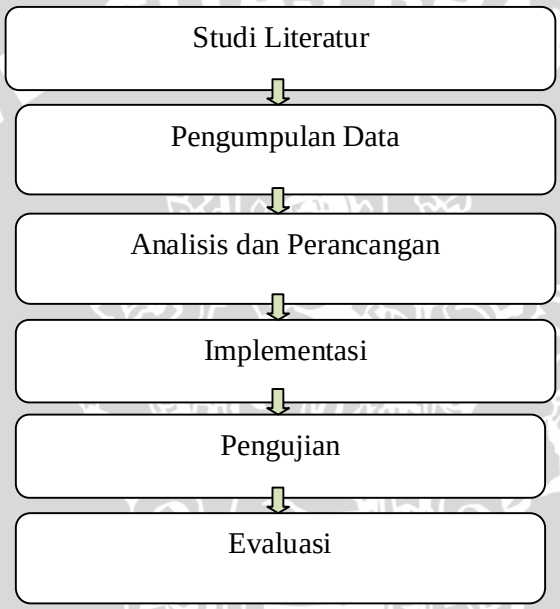
[FUA – 2013]

F-Measure merupakan gabungan antara *recall* dan *precision* yang didefinisikan dengan persamaan 2.8 [YAN - 1999] :

$$F = \frac{2xrecallxprecision}{recall + precision} \tag{2.8}$$

BAB III
METODOLOGI DAN PERANCANGAN

Pada bab metode penelitian ini berisi tentang pembahasan metodologi yang digunakan dalam penelitian yang berjudul "Penerapan Algoritma *Fuzzy C-Means* dalam Pengelompokan Data Tingkat Banjir Limpasan". Secara umum tahapan dari penelitian ini ditunjukkan pada gambar 3.1.



Gambar 3.1 Diagram Alir Penelitian

Penjelasan langkah-langkah pada gambar 3.1 adalah sebagai berikut :

1. Mempelajari berbagai literatur tentang Fuzzy C-Means dan parameter yang mempengaruhi tingkat banjir limpasan.
2. Pengumpulan data set yang digunakan untuk pengujian.
3. Melakukan analisa dan perancangan implementasi program menggunakan algoritma Fuzzy C-Means untuk pengelompokan data tingkat banjir limpasan.
4. Tahap dari proses implementasi program merupakan bagian dari pembuatan implementasi program.

5. Proses pengujian implementasi program yang telah dibuat, untuk memastikan bahwa implementasi program yang dibangun telah berjalan sesuai dengan yang diharapkan.
6. Melakukan evaluasi hasil yang diperoleh dari uji coba implementasi program.

3.1 Deskripsi Data

Dataset yang digunakan pada penelitian ini adalah data yang didapatkan dari BPDAS Brantas Jawa Timur. Dataset yang didapat memiliki 681 jumlah data dan 4 atribut berupa kemiringan lereng, jenis tanah, penggunaan lahan dan pola aliran. Dari 4 atribut tersebut didapatkan hasil tingkat banjir limpasan.

3.2 Deskripsi Umum

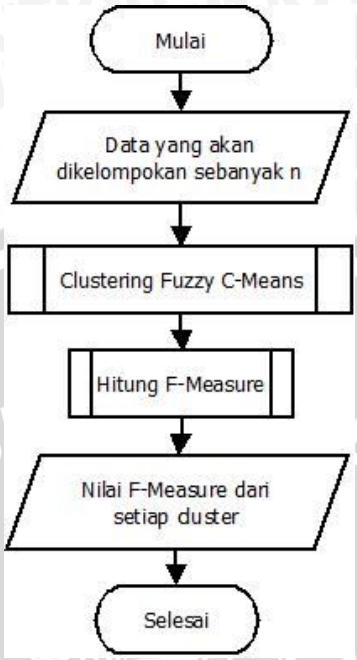
Merupakan implementasi program yang dapat mengelompokkan data menggunakan algoritma *Fuzzy C-Means*. Proses *cluster* data dilakukan dengan menentukan titik pusat suatu data serta menghitung jarak antar data berdasar kesamaan antar data, selain itu juga ditentukan seberapa besar nilai keanggotaan suatu data menjadi milik *cluster* tertentu.

Adapun proses yang dilakukan pada saat implementasi program antara lain :

1. User memasukkan data.
2. User memasukkan jumlah *cluster*, pangkat atau pembobot *fuzzy*, *error* terkecil dan maksimum iterasi.
3. Program melihat hasil *cluster*ing.

3.3 Perancangan

Perancangan dilakukan untuk menjelaskan bagaimana proses yang dikerjakan dalam pengelompokan data tingkat banjir limpasan. Data yang digunakan berupa format sql. Pada perancangan sistem ini terdiri dari 2 prosedur yaitu *cluster*ing *Fuzzy C-Means* dan hitung *F-Measure*. Rancangan alur pengelompokan secara umum dapat dilihat pada gambar 3.2.



Gambar 3.2 Alur perancangan

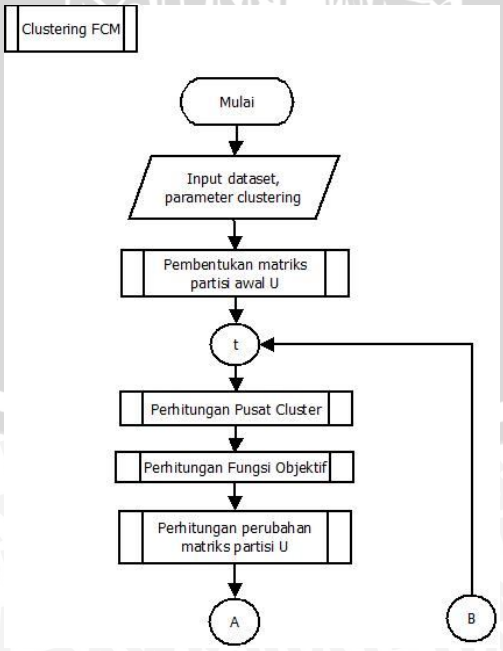
3.3.1 Clustering *Fuzzy C-Means*

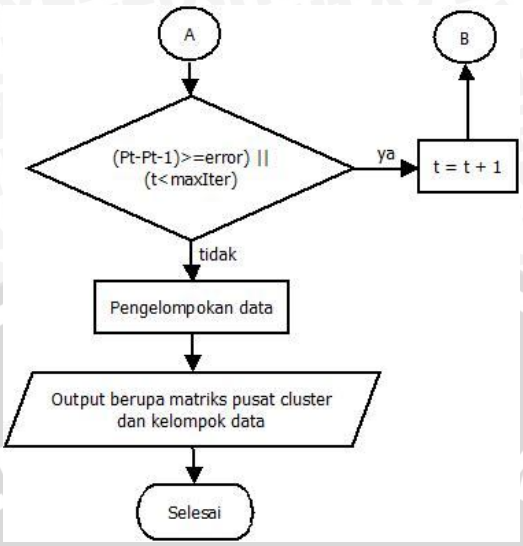
Proses *cluster*ing data terdiri atas beberapa *subproses*, diantaranya adalah pembentukan matriks partisi awal U, perhitungan pusat *cluster*, perhitungan fungsi objektif dan perhitungan perubahan matriks U. Input dari proses *cluster*ing berupa data tingkat banjir limpasan dan parameter *cluster*ing, sedangkan *output* proses ini berupa matriks pusat *cluster*, matriks partisi U dan kelompok data. Matriks pusat *cluster* berisi pusat data atribut pada setiap *cluster* (*Vkj*), matriks partisi U berisi derajat keanggotaan data pada semua *cluster* (μik) dan kelompok data berisi hasil pengelompokan data berdasarkan derajat keanggotaan data terhadap suatu *cluster*. Pengelompokan data dilakukan dengan memperhatikan nilai maksimum μik data pada setiap *cluster*. Pada gambar 3.3 dijelaskan alur dari keseluruhan proses pengelompokkan *Fuzzy C-Means*. Alur dari proses *cluster*ing data dengan menggunakan algoritma *Fuzzy C-Means* adalah sebagai berikut :

1. Menentukan *error* minimum (ξ), maksimum Iterasi (*maxIter*) dan nilai pangkat *cluster* (w).
2. Membentuk matriks partisi awal U.

3. Pembentukan matriks partisi awal dilakukan dengan cara membangkitkan bilangan random yang dihitung dengan persamaan 2.1 dan 2.2.
4. Selama memenuhi kondisi dimana nilai fungsi objektif lebih besar sama dengan error minimum yang diharapkan (Pt $\geq \xi$) atau iterasi lebih kecil sama dengan iterasi maksimum (t $\leq$ maxIter), maka dilakukan proses berikut secara berulang :
   a. Menghitung pusat *cluster* (*Vkj*) dengan menggunakan persamaan 2.3.
   b. Menghitung nilai dari fungsi objektif (Pt) dengan menggunakan persamaan 2.4.
   c. Menghitung perubahan matriks partisi U dengan menggunakan persamaan 2.5.
5. Pada tahap akhir dilakukan pengecekan kondisi berhenti. Alternatif kriteria penghentian adalah jika perubahan nilai *error* kurang dari *treshold* |Pn - Pn-1| < ε atau iterasi lebih besar dari iterasi maksimum (t > *maxIter*. Jika iterasi belum berhenti, kembali ke langkah 4.
6. Jika iterasi berhenti, ditentukan *cluster* dari tiap-tiap data. *Cluster* dipilih berdasarkan nilai matriks partisi terbesar.

Alur dari keseluruhan proses clustering *Fuzzy C-Means* ditunjukkan pada gambar 3.3.

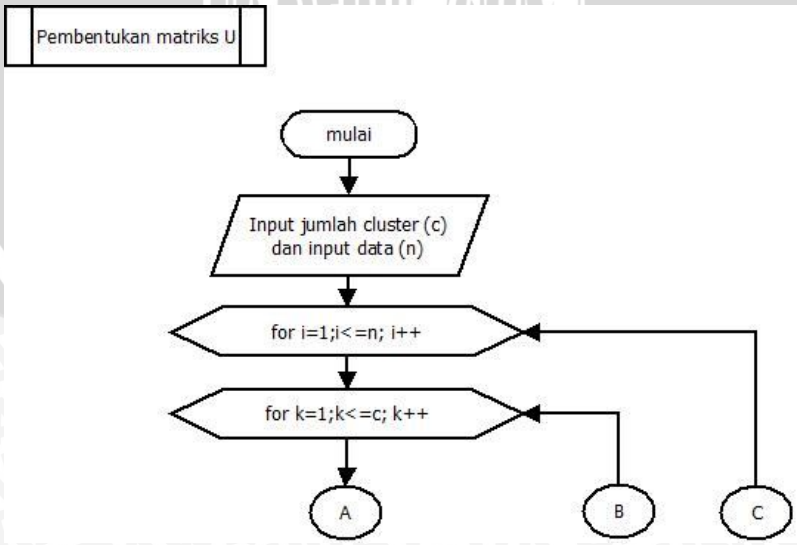


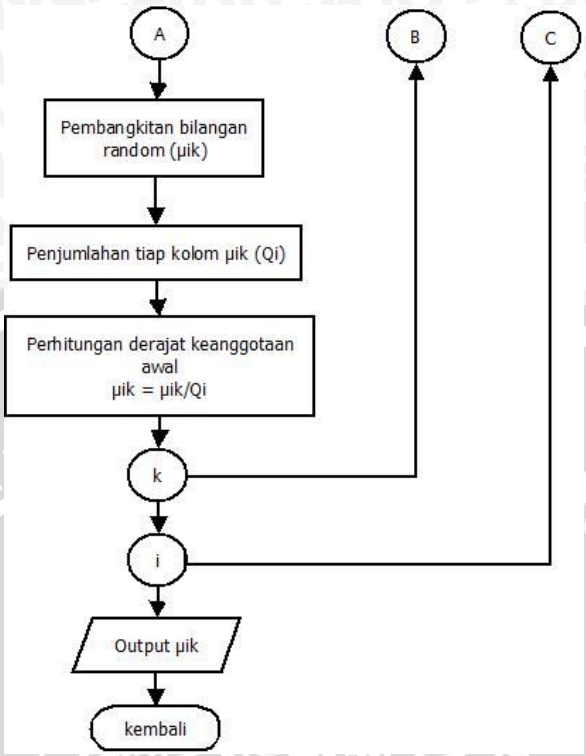


Gambar 3.3 Flowchart Fuzzy C-Means Clustering

3.3.1.1 Proses Pembentukan Matriks Awal U

Input yang diperlukan dalam proses pembentukan matriks partisi awal adalah nilai jumlah data (n) dan jumlah *cluster* yang akan dibentuk (c). Nilai *cluster* yang akan dibentuk merupakan suatu nilai *input*an, sehingga nantinya banyaknya *cluster* yang terbentuk sesuai nilai yang telah ditentukan oleh pengguna. *Output* dari proses ini adalah matriks derajat keanggotaan data terhadap *cluster* sesuai dengan banyaknya dimensi data. Alur dari keseluruhan proses pembentukan matriks U ditunjukkan pada gambar 3.4.





Gambar 3.4 Flowchart Pembentukan Matriks Awal U

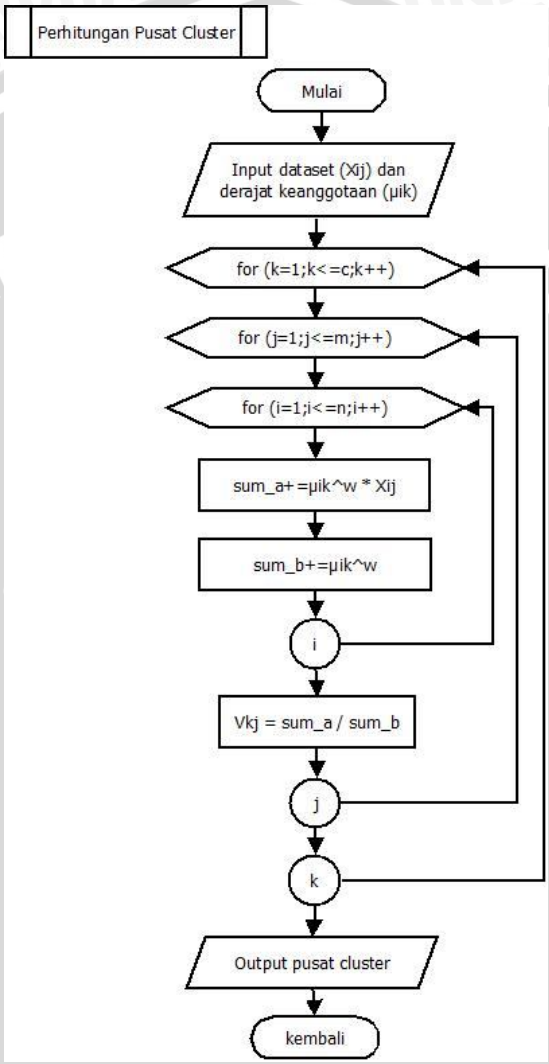
Tahapan proses pembentukan matriks partisi awal U adalah :

1. Membangkitkan bilangan random untuk derajat keanggotaan data ke-i dalam *cluster* ke-k. Rentang nilai random berkisar antara 0 hingga 1.
2. Menjumlahkan bilangan random μik pada setiap kolom iterasi ke-i (Qi = μ_{i1} + μ_{i2} + μ_{i3} + … + μ_{ik}) sehingga didapatkan Qi.
3. Perbaikan dari nilai elemen matriks derajat keanggotaan dihitung dengan menggunakan persamaan 2.5. Nilai perhitungan derajat keanggotaan dikatakan benar jika jumlah dari nilai derajat keanggotaan matriks U sebanyak cluster yang terbentuk bernilai sama dengan 1 (μ_{i1} + μ_{i2} + μ_{i3} + … + μ_{ik} = 1) jika nilai dari derajat keanggotaan tiap kolom tidak sama dengan 1 (μ_{i1} + μ_{i2} + μ_{i3} + … + $\mu_{ik} \neq 1$) maka nilai matriks keanggotaan salah.

3.3.1.2 Proses Perhitungan Pusat Cluster

Proses perhitungan pusat *cluster* memerlukan *input* berupa dataset limpasan (Xij) dan derajat keanggotaan hasil dari proses pembentukan matriks

partisi awal U (μ_{ik}). Perhitungan pusat *cluster* dilakukan menurut persamaan 2.3. Perhitungan ini menghasilkan matriks pusat *cluster* (V_{kj}) berukuran c x m (jumlah *cluster* x jumlah atribut). Alur proses perhitungan pusat cluster ditunjukkan pada gambar 3.5.

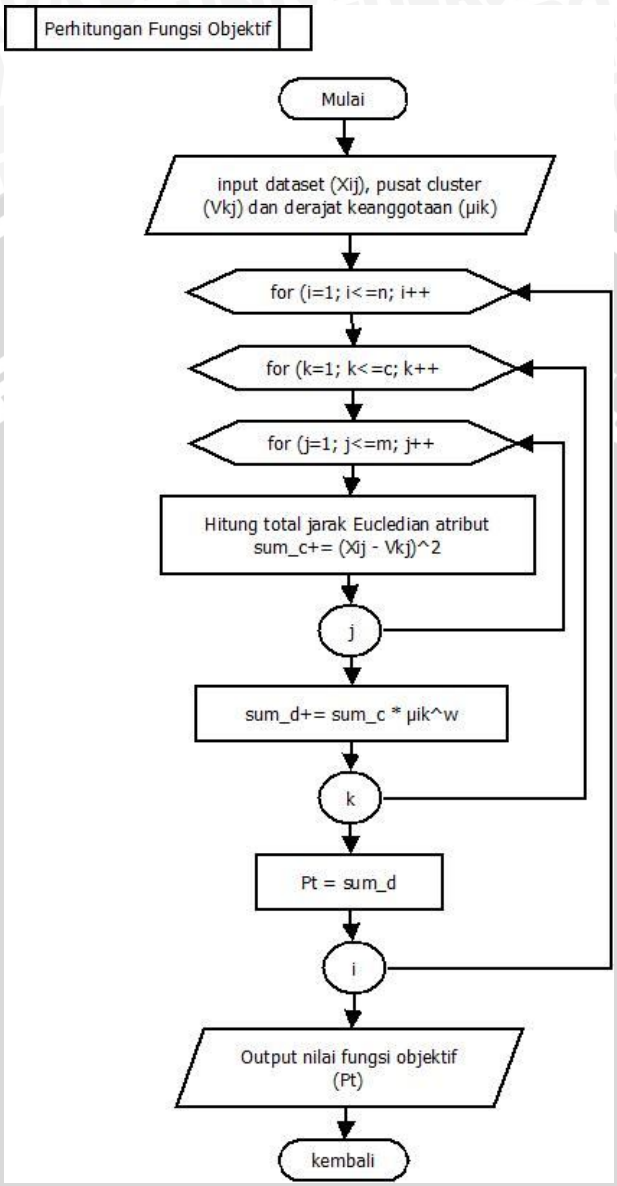


Gambar 3.5 Flowchart Perhitungan Pusat Cluster

3.3.1.3 Proses Perhitungan Fungsi Objektif

Proses perhitungan fungsi objektif melibatkan *input* dataset limpasan (Xij), pusat *cluster* (Vkj), derajat keanggotaan (μ_{ik}) serta input pangkat *cluster* (w). Proses ini akan menghasilkan *output* berupa nilai fungsi objektif iterasi ke-t

(Pt). Perhitungan fungsi objektif dilakukan dengan menggunakan persamaan 2.4. Alur proses perhitungan fungsi objektif ditunjukkan pada gambar 3.6.

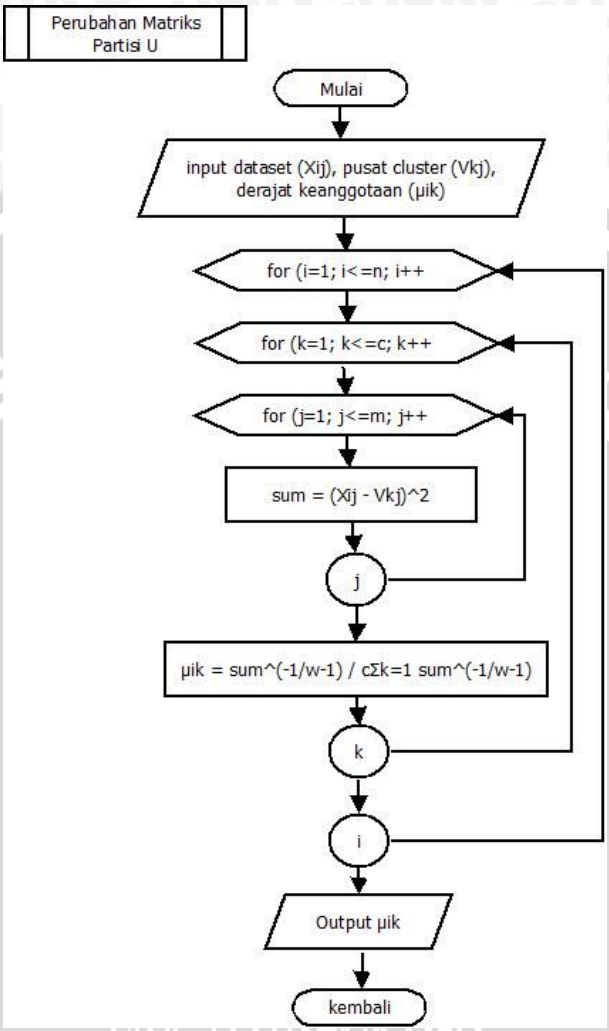


Gambar 3.6 Flowchart Perhitungan Fungsi Objektif

3.3.1.4 Proses Perubahan Matriks Partisi U

Seperti pada perhitungan fungsi objektif, proses perhitungan matriks partisi U memiliki *input* berupa data limpasan (Xij), pusat *cluster* (Vkj) dan derajat keanggotaan (μ_{ik}). *Output* proses ini adalah perbaikan derajat keanggotaan

μ_{ik} yang merupakan hasil perubahan matriks partisi U. Alur proses perubahan matriks partisi U ditunjukkan pada gambar 3.7.



Gambar 3.7 Flowchart Perubahan Matriks Partisi U

3.3.2 Pengujian F-Measure

Untuk menguji metode *Fuzzy C-Means* digunakan metode *f-Measure*. Langkah-langkah metode *f-Measure* adalah sebagai berikut :

1. Ambil Cluster[], classID[], classID adalah nilai kelas sebenarnya.
2. Cari classID paling dominan pada suatu *cluster*.
3. Jika :

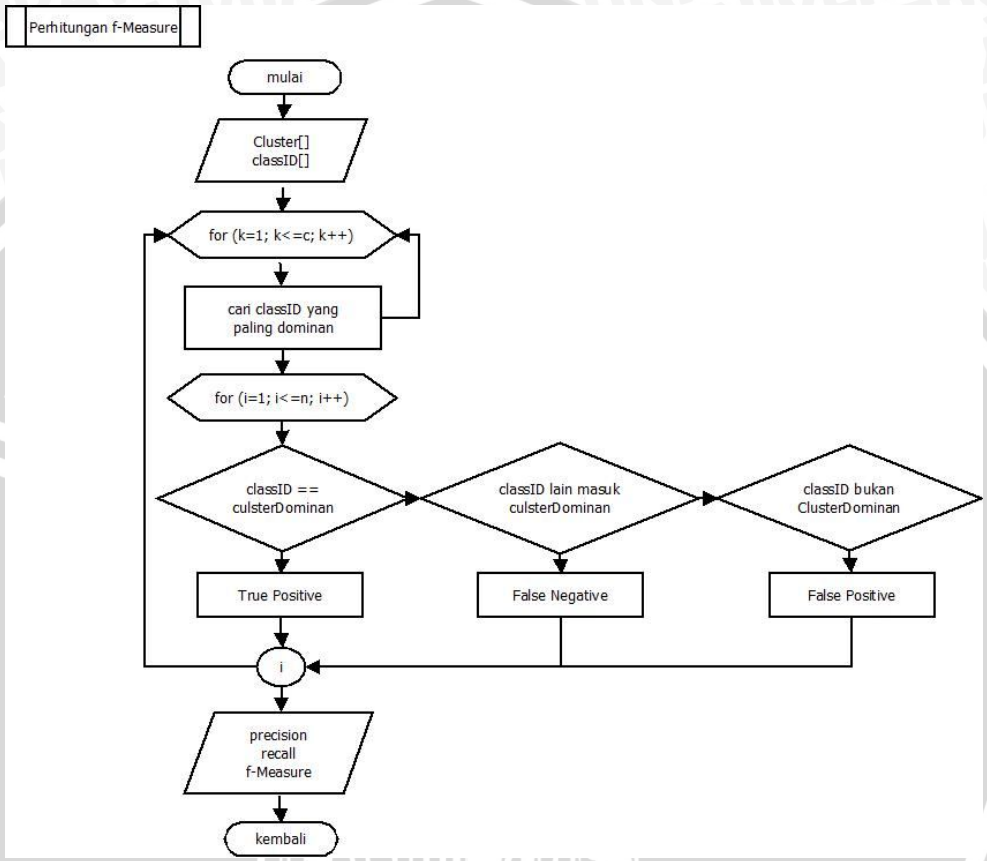
   3.1 ClassID = ClusterDominan → true positive

   3.2 ClassID cluster lain masuk ClusterDominan → false negative

3.3 ClassID bukan ClusterDominan → false positive

4. Lakukan perhitungan *precision*, *recall*, *f-Measure*.
5. Ulangi langkah 4 untuk semua *cluster*.

Alur proses *f-Measure* dapat dilihat pada gambar 3.8



Gambar 3.8 Flowchart Perhitungan f-Measure

3.4 Perhitungan Manual

Pada perhitungan manual ini digunakan sampel sebanyak 10 data dengan 4 atribut dan 3 kelas seperti yang ditunjukkan pada tabel 3.1 :

Tabel 3.1 Sampel Dataset

Lereng	Indeks Jenis Tanah	Penutupan Lahan	Pola Aliran	Limpasan
>40%	0.090	Hutan Lahan Kering Primer	Rendah	Ekstrem
25-40%	0.090	Pemukiman	Rendah	Ekstrem
15-25%	0.030	Perkebunan	Diabaikan	Ekstrem

0-8%	0.090	Pemukiman	Rendah	Tinggi
8-15%	0.160	Perkebunan	Diabaikan	Tinggi
15-25%	0.090	Hutan Tanaman	Rendah	Tinggi
0-8%	0.090	Sawah	Tinggi	Normal
0-8%	0.160	Laut/Tubuh Air	Normal	Normal
0-8%	0.160	Sawah	Tinggi	Normal
0-8%	0.160	Hutan Tanaman	Rendah	Normal

Dari data pada tabel 3.1, beberapa masih bertipe kategori. Namun masing-masing parameter yang masih bertipe kategori bisa diindexkan sesuai penjelasan di bab 2. Setelah di*index*kan / dikodekan sampel data menjadi seperti pada tabel 3.2.

Tabel 3.2 Sampel Dataset setelah diindexkan

Lereng	Indeks Jenis Tanah	Penutupan Lahan	Pola Aliran	Limpasan
9.50	0.090	5	15	Ekstrem
6.80	0.090	7	15	Ekstrem
3.10	0.030	3	20	Ekstrem
0.40	0.090	7	15	Tinggi
1.40	0.160	3	20	Tinggi
3.10	0.090	5	15	Tinggi
0.40	0.090	1	5	Normal
0.40	0.160	0	10	Normal
0.40	0.160	1	5	Normal
0.40	0.160	5	15	Normal

1. **Menentukan atribut**

Banyak *cluster* / c	3
Pembobot / w	2
Max iterasi / i	100
Eror terkecil yang diharapkan	0.00001
Fungsi Objektif awal / P0	0
Iterasi Awal / t	1

2. Membangkitkan bilangan random μ_{ik} dalam matriks berukuran (nxm).

Dengan menjumlah kolom setiap baris, makan didapat nilai Qi untuk masing-masing elemen ke-i.

Tabel 3.3 Bilangan random dan nilai Qi

Matriks Bilangan Random			Qi
0.78861	0.75142	0.4979744	2.03801
0.95898	0.22454	0.2906254	1.47415
0.38883	0.39906	0.6905513	1.47844
0.04971	0.60269	0.5376914	1.1901
0.67664	0.72469	0.9923285	2.39366
0.37784	0.32205	0.639584	1.33947
0.254	0.17508	0.8514168	1.2805
0.27712	0.07348	0.3904441	0.74104
0.2876	0.4063	0.8568715	1.55077
0.17235	0.54538	0.3315055	1.04924

Setiap bilangan random tersebut kemudian dibagi dengan Qi sehingga terbentuk matriks partisi awal U sebagai berikut :

$$\begin{vmatrix} 0.38695 & 0.3687 & 0.24434 \\ 0.65053 & 0.1523 & 0.19715 \\ 0.263 & 0.2699 & 0.46708 \\ 0.04177 & 0.5064 & 0.4518 \\ 0.28268 & 0.3028 & 0.41457 \\ 0.28208 & 0.2404 & 0.47749 \\ 0.19836 & 0.1367 & 0.66491 \\ 0.37396 & 0.0992 & 0.52689 \\ 0.18546 & 0.262 & 0.55255 \\ 0.16426 & 0.5198 & 0.31595 \end{vmatrix}$$

3. Menghitung Pusat Cluster

Hitung pusat *cluster* ke-k : Vkj, dengan k=1,2,3 dan j=1,2,…4. Pada iterasi pertama dilakukan perhitungan pusat *cluster* dengan menggunakan persamaan 2.3.

$$\frac{\sum_{i=1}^{10}((\mu ik)^2 * xij)}{\sum_{i=1}^{10}\mu ik^2}$$

Dengan persamaan 2.3 didapatkan 3 pusat *cluster*, Vkj dengan k=1,2,3 dan j=1,2,3,…4 sebagi berikut :

$$\begin{vmatrix} 4.66466 & 0.10488 & 4.5761 & 14.3378 \\ 2.10551 & 0.11631 & 4.8327 & 14.9006 \\ 1.35426 & 0.11285 & 2.7162 & 11.621 \end{vmatrix}$$

Untuk mendapatkan matriks pusat *cluster* seperti yang telah tercantum, berikut merupakan contoh perhitungan pusat *cluster* secara lebih detail. Tahapan pusat *cluster* dapat dilihat pada tabel 3.4 :

Tabel 3.4 Contoh perhitungan pusat *cluster* 1

Derajat keanggotaan cluster 1	Data Cluster			
μi1	Xi1	Xi2	Xi3	Xi4
0.38695	9.50	0.090	5	15
0.65053	6.80	0.090	7	15
0.263	3.10	0.030	3	20
0.04177	6.80	0.100	3	20
0.28268	3.10	0.030	5	15
0.28208	0.40	0.090	7	15
0.19836	1.40	0.160	3	20
0.37396	3.10	0.090	5	15
0.18546	6.80	0.100	1	5
0.16426	3.10	0.100	5	15

(lanjutan)

Derajat keanggotaan cluster 1					
μi1	$\mu i1^2$	$\mu i1^2$*xi1	$\mu i1^2$*xi2	$\mu i1^2$*xi3	$\mu i1^2$*xi4
0.38695	0.14973	1.422463118	0.013475966	0.748664799	2.245994397
0.65053	0.42319	2.877706447	0.038087291	2.962344872	6.347881869
0.263	0.06917	0.214418988	0.002075022	0.207502247	1.383348311
0.04177	0.00174	0.000697948	0.000157038	0.012214094	0.026173058
0.28268	0.07991	0.011187256	0.012785436	0.239726922	1.598179481
0.28208	0.07957	0.2466695	0.007161373	0.397854033	1.193562098
0.19836	0.03935	0.015738817	0.003541234	0.039347042	0.196735209
0.37396	0.13984	0.055937245	0.022374898	0	1.398431135
0.18546	0.03439	0.013757607	0.005503043	0.034394018	0.171970088
0.16426	0.02698	0.010793187	0.004317275	0.134914834	0.404744503
Σ	1.04388	4.869370114	0.109478576	4.77696286	14.96702015
$\Sigma[(\mu i1)^2$*Xij]/ $\Sigma(\mu i1)^2$		4.66466	0.10488	4.5761	14.3378

4. **Menentukan fungsi objektif.**

Fungsi objektif pada iterasi pertama P1 dapat dihitung menggunakan persamaan :

$$P1 = \sum_{i=1}^{10}\sum_{k=1}^{3}\left[\sum_{j=1}^{4}(Xij - V1j)^2\right](\mu i1)^2 = 146{,}2600396$$

Detail perhitungan fungsi objektif dapat dilihat pada tabel 3.5.

Tabel 3.5 Detail Perhitungan Fungsi Objektif

Kuadrat derajat keanggotaan data ke-i						
$\mu i1^2$	$\mu i2^2$	$\mu i3^2$	L1	L2	L3	L1+L2+L3
0.14973	0.13594	0.0597	3.59342	7.43833	4.95466	15.98640975
0.42319	0.0232	0.03887	4.60156	0.62055	2.30971	7.531818385
0.06917	0.07286	0.21816	2.55908	2.21192	16.0008	20.77182751
0.00174	0.25647	0.20413	0.04275	1.95341	6.26267	8.258831477
0.07991	0.09166	0.17186	4.39661	3.04565	12.3339	19.77616266
0.07957	0.05781	0.228	0.24401	0.0594	4.48738	4.790790949
0.03935	0.0187	0.4421	4.64969	2.16154	21.0855	27.89677829
0.13984	0.00983	0.27761	8.10365	0.49437	3.03091	11.62892986
0.03439	0.06864	0.30531	4.06448	7.93659	14.5617	26.56279472
0.02698	0.27018	0.09982	0.50751	0.79664	1.75154	3.05569603
						146.2600396

5. **Memperbaiki matriks partisi U dengan persamaan sebagai berikut**, persamaan ini merupakan penurunan dari persamaan 2.5.

$$\mu ik = \frac{\left[\sum_{j=1}^{4}(Xij - Vkj)^2\right]^{-1}}{\sum_{k=1}^{3}\left[\sum_{j=1}^{4}(Xij - Vkj)^2\right]^{-1}}$$

Tabel 3.6 Tabel Perhitungan Perbaikan Matriks

$\left[\sum_{j=1}^{4}(Xij - Vkj)^2\right]^{-1}$	$\left[\sum_{j=1}^{4}(Xij - Vkj)^2\right]^{-1}$	$\left[\sum_{j=1}^{4}(Xij - Vkj)^2\right]^{-1}$	$\sum_{k=1}^{3}\left[\sum_{j=1}^{4}(Xij - Vkj)^2\right]$	$\mu i1$	$\mu i2$	$\mu i3$
L1	L2	L3	LT=L1+L2+L3	L1/LT	L2/LT	L3/LT

0.04167	0.01828	0.01205	0.07199	0.578775	0.25385	0.167374
0.09197	0.03739	0.01683	0.14618	0.629121	0.255765	0.115114
0.02703	0.03294	0.01363	0.07360	0.367224	0.447528	0.185248
0.04081	0.13129	0.03259	0.20470	0.199387	0.641383	0.15923
0.01818	0.0301	0.01393	0.06220	0.292184	0.483808	0.224008
0.3261	0.97316	0.05081	1.35006	0.241544	0.720822	0.037634
0.00846	0.00865	0.02097	0.03808	0.222233	0.227137	0.55063
0.01726	0.01989	0.09159	0.12874	0.134046	0.154487	0.711467
0.00846	0.00865	0.02097	0.03808	0.222234	0.227141	0.550625
0.05317	0.33915	0.05699	0.44931	0.118332	0.754825	0.126843

Berdasarkan tabel 3.6, didapatkan perbaikan matriks partisi U sebagai berikut :

$$\begin{vmatrix} 0.57878 & 0.2539 & 0.16737 \\ 0.62912 & 0.2558 & 0.11511 \\ 0.36722 & 0.4475 & 0.18525 \\ 0.19939 & 0.6414 & 0.15923 \\ 0.29218 & 0.4838 & 0.22401 \\ 0.24154 & 0.7208 & 0.03763 \\ 0.22223 & 0.2271 & 0.55063 \\ 0.13405 & 0.1545 & 0.71147 \\ 0.22223 & 0.2271 & 0.55063 \\ 0.11833 & 0.7548 & 0.12684 \end{vmatrix}$$

Kemudian dilakukan pemeriksaan kondisi berhenti. Karena |P1 – P0| = 146.2600396 >> $\xi = {}^{10\text{-}5}$ dan iterasi = 1 < maxIter (=100), maka proses dilanjutkan ke iterasi 2 (t=2).

Pada iterasi ke-2, dapat dihitung kembali 3 pusat *cluster* sebagai berikut :

$$\begin{vmatrix} 5.55394 & 0.09303 & 4.954 & 15.0199 \\ 1.71746 & 0.1126 & 4.796 & 15.4656 \\ 0.52454 & 0.13624 & 1.0597 & 8.62134 \end{vmatrix}$$

Fungsi objektif pada iterasi P_2, didapatkan sebagai berikut :

$$P1 = \sum_{i=1}^{10}\sum_{k=1}^{3}\left[\sum_{j=1}^{4}(Xij - V1j)^2\right](\mu i1)^2 = 100{,}3645614$$

Kemudian nilai matriks partisi U diperbaiki :

$$\begin{vmatrix} 0.72998 & 0.1869 & 0.08312 \\ 0.80942 & 0.1503 & 0.04027 \\ 0.38529 & 0.5193 & 0.09542 \\ 0.16893 & 0.7627 & 0.06836 \\ 0.27479 & 0.6059 & 0.11935 \\ 0.25829 & 0.717 & 0.02476 \\ 0.07698 & 0.0873 & 0.83568 \\ 0.03636 & 0.0508 & 0.91283 \\ 0.07697 & 0.0873 & 0.8357 \\ 0.06765 & 0.9004 & 0.03197 \end{vmatrix}$$

Kemudian dilakukan pemeriksaan kondisi berhenti. Karena syarat berhenti belum terpenuhi, dimana |P2 – P1| = |100.3645614 – 146.2600396| = 45.8955 >> $\xi = 10^{-5}$ dan iterasi = 2 < maxIter (=100), maka iterasi diteruskan.

Setelah iterasi ke-16, salah satu syarat kondisi terpenuhi yaitu |P16 – P15| = |59.20250812 – 59.20251639| = 8.26315E-06 < $\xi = 10^{-5}$ sehingga diperoleh matriks cluster (matriks V).

$$\begin{vmatrix} 7.63206 & 0.08962 & 5.8618 & 15.144 \\ 1.24784 & 0.11302 & 4.597 & 16.8462 \\ 0.40784 & 0.13244 & 0.8064 & 6.16041 \end{vmatrix}$$

Selain pusat *cluster*, FCM juga memberikan hasil matriks derajat keanggotaan seperti pada tabel 3.7.

Tabel 3.7 Tabel Hasil *Clustering*

Data ke	Cluster ke-1	Cluster ke-2	Cluster ke-3	Cluster
1	0.92323	0.0548	0.022	1
2	0.94074	0.0472	0.01201	1
3	0.22034	0.7231	0.05661	2
4	0.14538	0.7877	0.06692	2
5	0.12741	0.8156	0.05702	2
6	0.23523	0.7161	0.04867	2
7	0.00762	0.0089	0.98353	3
8	0.10005	0.1647	0.73524	3
9	0.00761	0.0088	0.98354	3
10	0.07178	0.8884	0.0398	2

6. Perhitungan f-Measure

Ambil Cluster[] dan classID[], classID adalah nilai kelas sebenarnya.

Tabel 3.8 Tabel Pencocokan Cluster dan Class

Record	Cluster	ClassID
1	1	Ekstrem
2	1	Ekstrem
3	2	Ekstrem
4	2	Tinggi
5	2	Tinggi
6	2	Tinggi
7	3	Normal
8	3	Normal
9	3	Normal
10	2	Normal

7. Mencari classID paling dominan pada suatu *cluster*

Cluster 1 : 2 Ekstrem, 0 Tinggi, 0 Normal = Ekstrem Class

Cluster 2 : 1 Ekstrem, 3 Tinggi, 1 Normal = Tinggi Class

Cluster 3 : 0 Ekstrem, 0 Tinggi, 3 Normal = Normal Class

Jika :

7.1 ClassID = ClusterDominan → Trus Positive

7.2 ClassID cluster lain, masuk ClusterDominan → False Negative

7.3 ClassID, bukan ClusterDominan → False Positive

Tabel 3.9 Tabel classID paling dominan

Cluster 1	ClassID		Cluster 2	ClassID		Cluster 3	ClassID
1	Ekstrem		3	Ekstrem		7	Normal
2	Ekstrem		4	Tinggi		8	Normal
			5	Tinggi		9	Normal
			6	Tinggi			
			10	Normal			

Untuk *cluster* 1 yang merupakan *cluster* 'Ekstrem' mendapatkan nilai true positive dua didapat dari record 1 dan 2. False negative satu didapat dari record 3, sedangkan false positive nol.

Untuk *cluster* 2 yang merupakan *cluster* 'Tinggi' mendapatkan nilai true positive tiga didapat dari record 4,5,6. False negative nol, sedangkan false positive dua dari record 3 dan 10.

Untuk *cluster* 3 yang merupakan *cluster* 'Normal' mendapatkan nilai true positive tiga didapat dari record 7,8,9. False negative satu didapat dari record 10, sedangkan false positive nol.

Rincian nilainya dapat dilihat pada tabel 3.9.

Tabel 3.10 Rincian True Positive, False Negative, False Positif tiap *Cluster*

	True Positif	False Negatif	False Positif
Kluster1	2	1	0
Kluster2	3	0	2
Kluster3	3	1	0

Setelah didapat nilai true positive, false negative dan false positif, maka *precision*, *recall* dan *f-Measure* dapat dihitung. Perhitungannya sebagai berikut :

Cluster 1 :

$$precision = \frac{TP}{TP+FP} = \frac{2}{2+0} = \frac{2}{2} = 1$$

$$recall = \frac{TP}{TP+FN} = \frac{2}{2+1} = 0.67$$

$$f\text{-}measure = \frac{2xrecallxprecision}{recall + precision} = \frac{2*0.67*1}{0.67+1} = \frac{1.34}{1.67} = 0.8$$

Cluster 2 :

$$precision = \frac{TP}{TP+FP} = \frac{3}{3+2} = \frac{3}{5} = 0.6$$

$$recall = \frac{TP}{TP+FN} = \frac{3}{3+0} = \frac{3}{3} = 1$$

$$f\text{-}measure = \frac{2xrecallxprecision}{recall + precision} = \frac{2*1*0.6}{1+0.6} = \frac{1.2}{1.6} = 0.75$$

Cluster 3 :

$$precision = \frac{TP}{TP+FP} = \frac{3}{3+0} = \frac{3}{3} = 1$$

$$recall = \frac{TP}{TP+FN} = \frac{3}{3+1} = \frac{3}{4} = 0.75$$

$$f\text{-}measure = \frac{2xrecallxprecision}{recall + precision} = \frac{2*0.75*1}{0.75+1} = \frac{1.5}{1.75} = 0.86$$

Tabel 3.11 Tabel Hasil Perhitungan Precision, Recall, f-Measure

	Precision	Recall	F-Measure
Kluster1	1	0,67	0,80
Kluster2	0,6	1	0,75
Kluster3	1	0,75	0,86

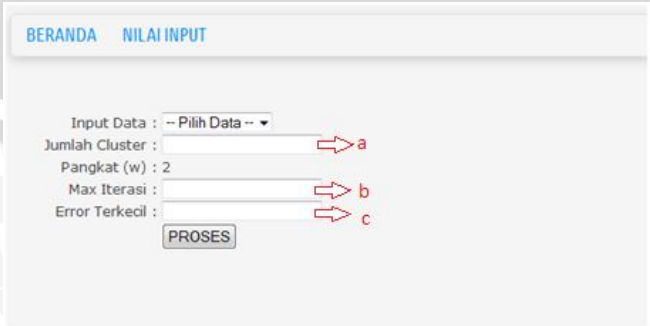
Lakukan rata-rata *f-Measure* untuk semua *cluster* : (0.80 + 0.75 + 0.86) / 3 = 0.80

Jadi hasil evaluasi clustering *Fuzzy C-Means* yang berupa *f-Measure* adalah 0.80

3.5 Rancangan Antar Muka

Perancangan antarmuka bertujuan untuk mewakili keadaan sebenarnya dari implementasi program yang akan dibangun. Implementasi program ini dibagi menjadi 2 halaman yaitu halaman nilai input dan halaman *fuzzy c-means*.

1. Perancangan antarmuka halaman nilai input



Gambar 3.9 Rancangan Antarmuka Nilai Input

Rancangan antarmuka halaman nilai input pada gambar 3.9 memiliki keterangan sebagai berikut :

a. Digunakan untuk memasukkan banyaknya jumlah *cluster*.

b. Digunakan untuk memasukkan nilai maksimum iterasi yang diharapkan.

c. Digunakan untuk memasukkan nilai *error* minimum yang diharapkan.

2. Perancangan antarmuka halaman *fuzzy c-means*

Gambar 3.10 Rancangan Antarmuka *Fuzzy C-Means*

Rancangan antarmuka halaman *fuzzy c-means* pada gambar 3.10 memiliki keterangan sebagai berikut :

a. Bilangan Random

Digunakan untuk menampilkan bilangan random.

b. Derajat Awal

Digunakan untuk menampilkan nilai matriks derajat keanggotaan awal.

c. Pusat *Cluster*

Digunakan untuk menampilkan hasil perhitungan pusat *cluster*.

d. Fungsi Objektif

Digunakan untuk menampilkan hasil perhitungan fungsi objektif.

e. Derajat Baru

Digunakan untuk menampilkan nilai dari derajat keanggotaan baru.

f. Result

Digunakan untuk menampilkan data yang masuk pada tiap-tiap *cluster* serta menampilkan hasil perhitungan *precision*, *recall* dan *f-measeure.*

3.6 Rancangan Uji Coba

Setelah program selesai diimplementasikan, langkah selanjutnya adalah melakukan pengujian dan evaluasi terhadap sistem tersebut. Uji coba ini bertujuan mengevaluasi ketepatan hasil *cluster*ing yang dilakukan oleh program. Hal ini bertujuan mengetahui keefektifan program yang dibuat menggunakan standart *precision*, *recall* dan *f-measure*.

Pada pengujian program ini digunakan input dataset limpasan yang berupa format sql. Setelah data tersebut dikelompokkan dengan menggunakan *Fuzzy C-Means* maka dihitung nilai akurasi *precision*, *recall* dan *f-measure*. Dari hasil perhitungan akurasi inilah yang nantinya digunakan untuk mencari nilai *cluster* yang paling optimal. Penentuan saat mendapatkan nilai akurasi, maka terlebih dahulu dilakukan pengujian terhadap *input* parameter *error* minimum yang diharapkan (ξ) pada masing-masing *cluster* (c) dengan tujuan mendapatkan nilai *error* minimum (ξ) pada saat nilai fungsi objektif telah cenderung konvergen. Nilai *error* minimum (ξ) pada saat nilai telah konvergen lah yang nantinya digunakan untuk penentuan dalam menghitung akurasi dengan *f-measure*.

Perhitungan dalam akurasi yang didapat dihitung berdasarkan rata-rata *precision*, *recall* dan *f-measure* yang dilakukan pada masing-masing *cluster*. Rancangan tabel uji coba dalam penelitian ini ditunjukkan pada tabel 3.12.

Tabel 3.12 Tabel Pengujian Evaluasi

Cluster	Rata-Rata Precision	Rata-Rata Recall	Rata-Rata f-Measure

BAB IV
IMPLEMENTASI

4.1 Lingkungan Implementasi

Lingkungan implementasi yang akan dijelaskan dalam sub bab ini adalah lingkungan implementasi perangkat keras dan perangkat lunak.

4.1.1 Lingkungan Implementasi Perangkat Keras

Perangkat keras yang digunakan dalam penelitian ini adalah sebuah laptop dengan spesifikasi :

- CPU : Intel Core 2 Duo T6500
- Memory : 2 GB
- Harddisk : 300 GB

4.1.2 Lingkungan Implementasi Perangkat Lunak

Perangkat lunak yang digunakan dalam penelitian ini adalah sebuah laptop dengan spesifikasi :

- Operating System : Windows 7 Ultimate
- Bahasa Pemrograman : PHP
- Database : MySQL
- Web Service : Apache
- Text Editor : Macromedia Dreamweaver 8

4.2 Implementasi Program

Berdasarkan analisa dan perancangan yang telah dipaparkan pada BAB III, maka bab ini dijelaskan langkah-langkah implementasinya.

4.2.1 Implementasi metode *Fuzzy C-Means*

Ada beberapa operasi yang akan dijalankan pada *Fuzzy C-Means* antara lain menentukan bilangan random, perhitungan nilai derajat keanggotaan awal,

perhitungan pusat cluster, perhitungan fungsi objektif, perhitungan derajat baru serta evaluasi *f-measure*.

4.2.1.1 Penentuan Bilangan Random

Penentuan bilangan random untuk menentukan nilai derajat awal. Bilangan random yang diambil berkisar antara 0 hingga 1. Implementasi penentuan bilangan random ditunjukkan pada gambar 4.1.

```
1  // Mengisi array matriks bilangan random µik
2  if($z==1){
3    $hapus_random = mysql_query("delete from random");
4    for($i=0;$i<$cluster;$i++){
5      for($j=0;$j<$limit;$j++){
6        $random = number_format(rand(10000,100000) / 100000,10);
7        $µik[$i][$j] = $random;
8        $simpan_random = mysql_query("insert into random value
9        ('$j','$i','$random')");
10     }
11   }
12 }
```

Gambar 4.1 *Source Code* Penentuan Bilangan Random

4.2.1.2 Perhitungan Nilai Derajat Keanggotaan Awal

Perhitungan nilai derajat keanggotaan awal adalah dengan membagi nilai dari bilangan random dari masing-masing dokumen dengan jumlah bilangan random di setiap cluster. Implementasi penentuan nilai derajat keanggotaan awal ditunjukkan pada gambar 4.2.

```
1  // Mengisi array Qi
2  for($i=0;$i<$limit;$i++){
3    $temp = 0;
4    for($j=0;$j<$cluster;$j++){
5      $temp += $µik[$j][$i];
6    }
7    $µik_j[$i] = $temp;
8  }
9
```

```
10 // Mengisi array matriks partisi awal U
11 if ($z==1){
12   $hapus_derajat_awal      =      mysql_query("delete      from
13    derajatawal");
14   for($i=0;$i<$cluster;$i++){
15     for($j=0;$j<$limit;$j++){
16       $derajatawal = number_format($µik[$i][$j]/$µik_j[$j],5);
17       $m_p_a_U[$i][$j] = $derajatawal;
18       $simpan_derajat_awal    =    mysql_query("insert    into
19       derajatawal value ('$j','$i','$derajatawal')");
20     }
21   }
22 }
```

Gambar 4.2 *Source Code* Penentuan Nilai Derajat Awal

4.2.1.3 Perhitungan Pusat *Cluster*

Perhitungan pusat *cluster* didapat dari total hasil perkalian dari derajat keanggotaan awal dengan nilai masing-masing record dibagi dengan total jumlah derajat keanggotaan pada masing-masing cluster yang sudah dikuadratkan. Implementasi perhitungan pusat *cluster* ditunjukkan pada gambar 4.3.

```
1  Menghitung Pusat CLuster
2  <br>
3  ((u<sub>ik</sub>)w*X<sub>ij</sub>)
4  <?php
5    for($a=0;$a<$cluster;$a++){
6  ?>
7  <br>
8  <table border="1" width="976px">
9    <tr>
10     <td   align="center"   rowspan="<?php   echo   $limit+1   ;
11     ?>">K=<?php echo $a+1; ?></td>
12     <td align="center">j=1</td>
13     <td align="center">j=2</td>
14     <td align="center">j=3</td>
15     <td align="center">j=4</td>
16   </tr>
17 <?php
```

```
18    for ($i=0;$i<$limit;$i++){
19      $k_d[$a][$i]        = pow($m_p_a_U[$a][$i],2);
20      $m_p_c_j1[$a][$i]   = $k_d[$a][$i]*$lereng[$i];
21      $m_p_c_j2[$a][$i]   =
22      $k_d[$a][$i]*$indeks_jenis_tanah[$i];
23      $m_p_c_j3[$a][$i]   = $k_d[$a][$i]*$penutupan_lahan[$i];
24      $m_p_c_j4[$a][$i]   = $k_d[$a][$i]*$pola_aliran[$i];
25  ?>
26    <tr>
27      <td align="center"><?php echo
28      number_format($m_p_c_j1[$a][$i],5); ?></td>
29      <td align="center"><?php echo
30      number_format($m_p_c_j2[$a][$i],5); ?></td>
31      <td align="center"><?php echo
32      number_format($m_p_c_j3[$a][$i],5); ?></td>
33      <td align="center"><?php echo
34      number_format($m_p_c_j4[$a][$i],5); ?></td>
35    </tr>
36  <?php
37    }
38  ?>
39    </table>
40    <br><br>
41  <?php
42    }
43  ?>
44    <table border="1" width="976px">
45      <tr>
46        <td align="center" colspan="<?php echo $cluster;
47        ?>">Kuadrat Derajat</td>
48      </tr>
49      <tr>
50  <?php
51    for ($i=0;$i<$cluster;$i++){
52  ?>
53    <td align="center">U<?php echo $i+1; ?><sup>2</sup></td>
54  <?php
55    }
```

```
56  ?>
57    </tr>
58  <?php
59    for($a=0;$a<$limit;$a++){
60  ?>
61    <tr>
62  <?php
63    for ($i=0;$i<$cluster;$i++){
64  ?>
65    <td align="center"><?php echo
66    number_format($k_d[$i][$a],5); ?></td>
67  <?php
68    }
69  ?>
70    </tr>
71  <?php
72    }
73    for($i=0;$i<$cluster;$i++){
74      $temp_1 = 0;
75      $temp_2 = 0;
76      $temp_3 = 0;
77      $temp_4 = 0;
78      $temp = 0;
79
80      for($j=0;$j<$limit;$j++){
81        $temp_1 += $m_p_c_j1[$i][$j];
82        $temp_2 += $m_p_c_j2[$i][$j];
83        $temp_3 += $m_p_c_j3[$i][$j];
84        $temp_4 += $m_p_c_j4[$i][$j];
85        $temp += $k_d[$i][$j];
86      }
87
88      $p_c_1[$i] = $temp_1 / $temp;
89      $p_c_2[$i] = $temp_2 / $temp;
90      $p_c_3[$i] = $temp_3 / $temp;
91      $p_c_4[$i] = $temp_4 / $temp;
92    }
93  ?>
```

```
94    </table>
95    <br><br>
96    <table border="1" width="976px">
97      <tr>
98        <td align="center" colspan="4">Pusat Cluster</td>
99      </tr>
100 <?php
101   $hapus_pusat_cluster = mysql_query("delete from
102   pusatcluster");
103   for($i=0;$i<$cluster;$i++){
104   $simpan_pusat_cluster = mysql_query("insert into
105   pusatcluster value ('$i','1','$p_c_1[$i]')");
106   $simpan_pusat_cluster = mysql_query("insert into
107   pusatcluster value ('$i','2','$p_c_2[$i]')");
108   $simpan_pusat_cluster = mysql_query("insert into
109   pusatcluster value ('$i','3','$p_c_3[$i]')");
110   $simpan_pusat_cluster = mysql_query("insert into
111   pusatcluster value ('$i','4','$p_c_4[$i]')");
112 ?>
113   <tr>
114     <td align="center"><?php echo
115     number_format($p_c_1[$i],5); ?></td>
116     <td align="center"><?php echo
117     number_format($p_c_2[$i],5); ?></td>
118     <td align="center"><?php echo
119     number_format($p_c_3[$i],5); ?></td>
120     <td align="center"><?php echo
121     number_format($p_c_4[$i],5); ?></td>
122   </tr>
123 <?php
124   }
```

Gambar 4.3 *Source Code* Perhitungan Pusat *Cluster*

4.2.1.4 Perhitungan Fungsi Objektif

Pada perhitungan fungsi objektif melibatkan perhitungan nilai pusat *cluster*, nilai dari masing-masing atribut dan nilai kuadrat dari derajat

keanggotaan awal. Implementasi perhitungan fungsi objektif ditunjukkan pada gambar 4.4.

```
1   </table>
2   <br><br>
3   <table border="1" width="976px">
4     <tr>
5       <td colspan="5" align="center">
6         Menghitung Fungsi Obyektif<br>
7         (X<sub>ij</sub> - V<sub>kj</sub>)<sup>2</sup>
8       </td>
9     </tr>
10  <?php
11    for($i=0;$i<$cluster;$i++){
12  ?>
13    <tr>
14      <td   align="center"   rowspan="<?php   echo   $limit+1;
15      ?>">K=<?php echo $i+1; ?></td>
16      <td align="center">j=1</td>
17      <td align="center">j=2</td>
18      <td align="center">j=3</td>
19      <td align="center">j=4</td>
20    </tr>
21  <?php
22    for($j=0;$j<$limit;$j++){
23      $m_f_o_j1[$i][$j] = pow(($lereng[$j]              -
24      $p_c_1[$i]),2);
25      $m_f_o_j2[$i][$j]    =    pow(($indeks_jenis_tanah[$j]    -
26      $p_c_2[$i]),2);
27      $m_f_o_j3[$i][$j] = pow(($penutupan_lahan[$j]      -
28      $p_c_3[$i]),2);
29      $m_f_o_j4[$i][$j] = pow(($pola_aliran[$j]          -
30      $p_c_4[$i]),2);
31  ?>
32    <tr>
33      <td              align="center"><?php              echo
34      number_format($m_f_o_j1[$i][$j],5); ?></td>
35      <td              align="center"><?php              echo
36      number_format($m_f_o_j2[$i][$j],5); ?></td>
```

```
37      <td                    align="center"><?php                    echo
38      number_format($m_f_o_j3[$i][$j],5); ?></td>
39      <td                    align="center"><?php                    echo
40      number_format($m_f_o_j4[$i][$j],5); ?></td>
41    </tr>
42  <?php
43      }
44    }
45  ?>
46    </table>
47    <br><br>
48    <table border="1" width="976px">
49      <tr>
50        <td align="center" colspan="<?php echo $cluster;
51        ?>">E(X<sub>ij</sub> - V<sub>kj</sub>)<sup>2</sup> x
52        (u<sub>ik</sub>)<sup>w</sup></td>
53        <td align="center" rowspan="2">E(E(X<sub>ij</sub> -
54        V<sub>kj</sub>)<sup>2</sup> x
55        (u<sub>ik</sub>)<sup>w</sup>)</td>
56      </tr>
57      <tr>
58  <?php
59    for($i=0;$i<$cluster;$i++){
60  ?>
61      <td align="center">K=<?php echo $i+1; ?></td>
62  <?php
63    }
64  ?>
65      </tr>
66  <?php
67    for($i=0;$i<$limit;$i++){
68      $temp = 0;
69  ?>
70      <tr>
71  <?php
72      for($j=0;$j<$cluster;$j++){
73        $k_akhir[$i][$j]       =       ($m_f_o_j1[$j][$i]       +
74        $m_f_o_j2[$j][$i]       +       $m_f_o_j3[$j][$i]       +
```

```
75          $m_f_o_j4[$j][$i])*$k_d[$j][$i];
76          $temp += $k_akhir[$i][$j];
77  ?>
78      <td align="center"><?php echo
79      number_format($k_akhir[$i][$j],5); ?></td>
80  <?php
81    }
82    $jumlah_akhir[$z] += $temp;
83  ?>
84      <td   align="center"><?php   echo   number_format($temp,5);
85  ?></td>
86    </tr>
87  <?php
88   }
89  ?>
90    <tr>
91      <td align="center" colspan="<?php echo $cluster;
92      ?>">E(E(E(X<sub>ij</sub> - V<sub>kj</sub>)<sup>2</sup> x
93      (u<sub>ik</sub>)<sup>w</sup>))</td>
94      <td align="center"><?php echo
95      number_format($jumlah_akhir[$z],5); ?></td>
96    </tr>
97  <?php
98   if($z > 1){
99
100    $fobj = $jumlah_akhir[$z];
101    $fobj2 = $jumlah_akhir[$z-1];
102    $errors = abs($jumlah_akhir[$z] - $jumlah_akhir[$z-1]);
103    $simpan_fo = mysql_query("insert into fungsiobjektif value
104    ('$z','$fobj', '$fobj2','$errors')");
105
106    $error_t = abs($jumlah_akhir[$z] - $jumlah_akhir[$z-1]);
107  ?>
108    <tr>
109      <td   align="center"   colspan="<?php   echo   $cluster;
110      ?>">Error</td>
111      <td align="center"><?php echo number_format($error_t,5);
112      ?></td>
```

```
113     </tr>
114 <?php
115  }
116 ?>
```

Gambar 4.4 *Source Code* Perhitungan Fungsi Objektif

4.2.1.5 Perhitungan Nilai Perubahan Matriks Partisi

Perhitungan nilai perubahan matriks partisi digunakan untuk mendapatkan nilai dari derajat keanggotaan baru. Perhitungan matriks partisi diambil dari nilai fungsi objektif yang telah dipangkatkan dengan -1 ditunjukkan pada gambar 4.5 sedangkan perhitungan nilai perubahan matriks partisi ditunjukkan pada gambar 4.6.

```
1  //Perbaikan Matriks µik
2  if($z > 1){
3    for($i=0;$i<$cluster;$i++){
4      for($j=0;$j<$limit;$j++){
5        $perbaikan = pow(($m_f_o_j1[$i][$j] + $m_f_o_j2[$i][$j]
6        + $m_f_o_j3[$i][$j] + $m_f_o_j4[$i][$j]),-1);
7        $µik[$i][$j] = $perbaikan;
8      }
9    }
10 }
11
12 // Mengisi array matriks partisi awal U perbaikan
13 $hapus_derajat_awal = mysql_query("delete from
14 matriksperbaikan");
15 for($i=0;$i<$cluster;$i++){
16   for($j=0;$j<$limit;$j++){
17     $matriksperbaikan =
18     number_format($µik[$i][$j]/$µik_j[$j],5);
19     $m_p_a_U[$i][$j] = $matriksperbaikan;
20     $simpan_matriks_perbaikan = mysql_query("insert into
21     matriksperbaikan value ('$j','$i','$matriksperbaikan')");
22   }
23 }
```

Gambar 4.5 *Source Code* Perhitungan Derajat Keanggotaan Baru

```
// Mengisi array matriks partisi awal U perbaikan
$hapus_derajat_awal = mysql_query("delete from
matriksperbaikan");
for($i=0;$i<$cluster;$i++){
  for($j=0;$j<$limit;$j++){
    $matriksperbaikan =
    number_format($µik[$i][$j]/$µik_j[$j],5);
    $m_p_a_U[$i][$j] = $matriksperbaikan;
    $simpan_matriks_perbaikan = mysql_query("insert into
    matriksperbaikan value ('$j','$i','$matriksperbaikan')");
  }
}
```

Gambar 4.6 *Source Code* Nilai Derajat Baru

4.2.1.6 Proses Pengelompokkan

Perhitungan pengelompokkan dilakukan untuk mendapatkan record yang masuk dalam cluster yang telah ditentukan jumlahnya. Pengelompokkan dilihat dari nilai derajat keanggotaan baru. Penentuan record memasuki sebuah cluster ditentukan dari nilai derajat keanggotaan baru yang lebih besar. Implementasi kecenderungan data memasuki cluster ditunjukkan pada gambar 4.7.

```
 // Menentukan CLUSTER
 $mpa = number_format($m_p_a_U[$j][$i],5);
 if ($mpa > $temp_clust) {
   $no_clust = $j;
   $temp_clust = $mpa;
 }
}
 $idnya = $id[$i];
 $clusterdf = $no_clust+1;

// update
if ($DATANYA == 0) {
  $up_cluster = mysql_query("update dataa set cluster =
  '$clusterdf' where id='$idnya'");
} else if ($DATANYA == 1) {
  $up_cluster = mysql_query("update datauji1 set cluster =
```

```
  '$clusterdf' where id='$idnya'");
} else if ($DATANYA == 2) {
  $up_cluster = mysql_query("update datauji2 set cluster =
  '$clusterdf' where id='$idnya'");
}else if ($DATANYA == 3) {
  $up_cluster = mysql_query("update datauji3 set cluster =
  '$clusterdf' where id='$idnya'");
}
//update
?>
  <td  align="center"><?php  echo  "Cluster";  echo  $no_clust+1;
   ?></td>
 </tr>
<?php
```

Gambar 4.7 *Source Code* Proses Pengelompokkan

4.2.1.7 Evaluasi

Pada proses ini akan dilakukan evaluasi untuk mengetahui performa metode *Fuzzy C-Means*. Implementasi evaluasi *f-measure* ditunjukkan pada gambar 4.8.

```
// hitung cluster dominan
$hapus_clust_dominan      =      mysql_query("delete      from
dominancluster");
for($i=0;$i<$cluster;$i++){
  $clu = $i + 1;
    // update
    if ($DATANYA == 0) {
      $ce  =  mysql_query("select  *  from  dataa  where  cluster
      ='$clu' and limpasan = 'Ekstrem'");
      $jum_ce = mysql_num_rows($ce);
      $ct  =  mysql_query("select  *  from  dataa  where  cluster
      ='$clu' and limpasan = 'Tinggi'");
      $jum_ct = mysql_num_rows($ct);
      $cn  =  mysql_query("select  *  from  dataa  where  cluster
      ='$clu' and limpasan = 'Normal'");
      $jum_cn = mysql_num_rows($cn);
    } else if ($DATANYA == 1) {
```

```
18        $ce = mysql_query("select * from datauji1 where cluster
19        ='$clu' and limpasan = 'Ekstrem'");
20        $jum_ce = mysql_num_rows($ce);
21        $ct = mysql_query("select * from datauji1 where cluster
22        ='$clu' and limpasan = 'Tinggi'");
23        $jum_ct = mysql_num_rows($ct);
24        $cn = mysql_query("select * from datauji1 where cluster
25        ='$clu' and limpasan = 'Normal'");
26        $jum_cn = mysql_num_rows($cn);
27      } else if ($DATANYA == 2) {
28        $ce = mysql_query("select * from datauji2 where cluster
29        ='$clu' and limpasan = 'Ekstrem'");
30        $jum_ce = mysql_num_rows($ce);
31        $ct = mysql_query("select * from datauji2 where cluster
32        ='$clu' and limpasan = 'Tinggi'");
33        $jum_ct = mysql_num_rows($ct);
34        $cn = mysql_query("select * from datauji2 where cluster
35        ='$clu' and limpasan = 'Normal'");
36        $jum_cn = mysql_num_rows($cn);
37      }else if ($DATANYA == 3) {
38        $ce = mysql_query("select * from datauji3 where cluster
39        ='$clu' and limpasan = 'Ekstrem'");
40        $jum_ce = mysql_num_rows($ce);
41        $ct = mysql_query("select * from datauji3 where cluster
42        ='$clu' and limpasan = 'Tinggi'");
43        $jum_ct = mysql_num_rows($ct);
44        $cn = mysql_query("select * from datauji3 where cluster
45        ='$clu' and limpasan = 'Normal'");
46        $jum_cn = mysql_num_rows($cn);
47      }
48    //update
49    if ($jum_ce > $jum_ct AND $jum_ce > jum_cn){
50      $class = "Ekstrem";
51    }
52    if ($jum_ct > $jum_ce AND $jum_ct > $jum_cn){
53      $class = "Tinggi";
54    }
55    if ($jum_cn > $jum_ce AND $jum_cn > $jum_ct){
```

```
56      $class = "Normal";
57    }
58
59    // update dominan
60    if ($DATANYA == 0) {
61      $load = mysql_query("select * from dataa where cluster =
62      '$clu'");
63      while ($dataload = mysql_fetch_array($load)){
64        $idcl = $dataload['id'];
65        $edit_dom = mysql_query("update dataa set clust_dominan
66        = '$class' where id = '$idcl'");
67    }
68    } else if ($DATANYA == 1) {
69      $load = mysql_query("select * from datauji1 where cluster
70      = '$clu'");
71      while ($dataload = mysql_fetch_array($load)){
72        $idcl = $dataload['id'];
73        $edit_dom    =    mysql_query("update    datauji1    set
74        clust_dominan = '$class' where id = '$idcl'");
75    }
76    } else if ($DATANYA == 2) {
77      $load = mysql_query("select * from datauji2 where cluster
78      = '$clu'");
79      while ($dataload = mysql_fetch_array($load)){
80        $idcl = $dataload['id'];
81        $edit_dom    =    mysql_query("update    datauji2    set
82        clust_dominan = '$class' where id = '$idcl'");
83    }
84    }else if ($DATANYA == 3) {
85      $load = mysql_query("select * from datauji3 where cluster
86      = '$clu'");
87      while ($dataload = mysql_fetch_array($load)){
88        $idcl = $dataload['id'];
89        $edit_dom    =    mysql_query("update    datauji3    set
90        clust_dominan = '$class' where id = '$idcl'");
91    }
92    }
93    //update
```

```
94
95    // update
96    if ($DATANYA == 0) {
97      $tp  =  mysql_query("select  *  from  dataa  where  cluster
98      ='$clu' and limpasan = '$class'");
99      $jum_tp = mysql_num_rows($tp);
100     $fn  =  mysql_query("select  *  from  dataa  where  cluster
101     !='$clu' and limpasan = '$class'");
102     $jum_fn = mysql_num_rows($fn);
103     $fp  =  mysql_query("select  *  from  dataa  where  cluster
104     ='$clu' and limpasan != '$class'");
105     $jum_fp = mysql_num_rows($fp);
106   } else if ($DATANYA == 1) {
107     $tp = mysql_query("select  * from datauji1 where cluster
108     ='$clu' and limpasan = '$class'");
109     $jum_tp = mysql_num_rows($tp);
110     $fn = mysql_query("select  * from datauji1 where cluster
111     !='$clu' and limpasan = '$class'");
112     $jum_fn = mysql_num_rows($fn);
113     $fp = mysql_query("select  * from datauji1 where cluster
114     ='$clu' and limpasan != '$class'");
115     $jum_fp = mysql_num_rows($fp);
116   } else if ($DATANYA == 2) {
117     $tp = mysql_query("select  * from datauji2 where cluster
118     ='$clu' and limpasan = '$class'");
119     $jum_tp = mysql_num_rows($tp);
120     $fn = mysql_query("select  * from datauji2 where cluster
121     !='$clu' and limpasan = '$class'");
123     $jum_fn = mysql_num_rows($fn);
124     $fp = mysql_query("select  * from datauji2 where cluster
125     ='$clu' and limpasan != '$class'");
126     $jum_fp = mysql_num_rows($fp);
127   }else if ($DATANYA == 3) {
128     $tp = mysql_query("select  * from datauji3 where cluster
129     ='$clu' and limpasan = '$class'");
130     $jum_tp = mysql_num_rows($tp);
131     $fn = mysql_query("select  * from datauji3 where cluster
132     !='$clu' and limpasan = '$class'");
```

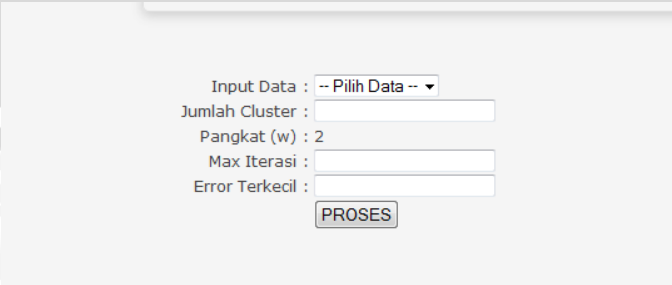
```
133     $jum_fn = mysql_num_rows($fn);
134     $fp = mysql_query("select * from datauji3 where cluster
135     ='$clu' and limpasan != '$class'");
136     $jum_fp = mysql_num_rows($fp);
137   }
138   //update
139
140   $precision = $jum_tp/($jum_tp+$jum_fp);
141   $recall = $jum_tp/($jum_tp+$jum_fn);
142   $fmeasure = (2*$recall*$precision)/($recall+$precision);
143
144   $simpan_dom = mysql_query("insert into dominancluster value
145   ('$clu','$jum_ce','$jum_ct','$jum_cn',
146   '$class','$jum_tp','$jum_fn','$jum_fp',
147   '$precision','$recall','$fmeasure')");
```

Gambar 4.8 *Source Code* Evaluasi *f-measure*

4.3 Implementasi Antar Muka

1. Nilai *Input*

Nilai *input* digunakan untuk memasukkan nilai dari banyaknya *cluster* yang dibentuk dan pengecekan kondisi berhenti yang meliputi masukan nilai maksimum iterasi dan *error* terkecil yang diharapkan. Sistem akan berhenti menghitung proses pengelompokkan jika memenuhi kondisi bahwa perhitungan mencapai nilai maksimum iterasi atau perhitungan nilai selisih fungsi objektif dari masing-masing iterasi mencapai nilai yang dimasukkan oleh pengguna. Implementasi perancangan antarmuka untuk nilai input ditunjukkan pada gambar 4.9.



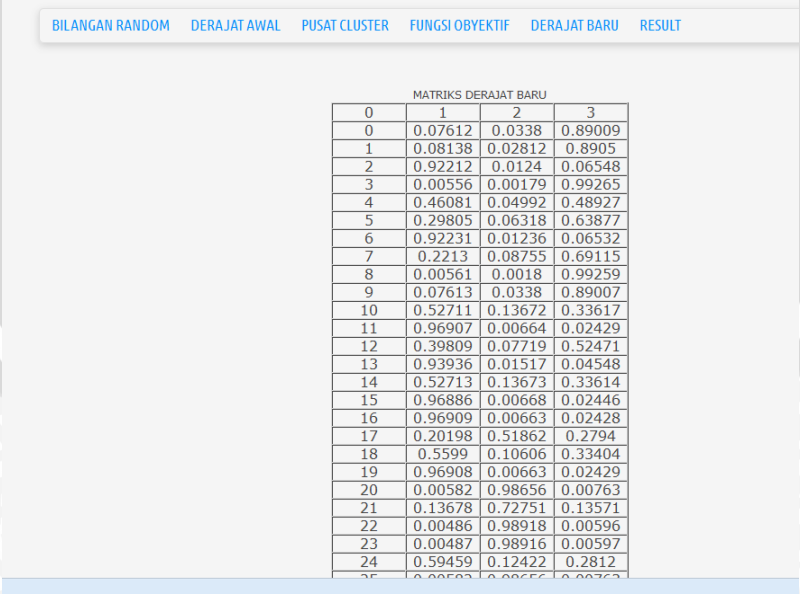
Gambar 4.9 Implementasi Antarmuka Nilai *Input*

Keterangan gambar 4.9 adalah sebagai berikut :

1. *Input Data* berbentuk suatu *option* yang digunakan untuk menentukan berapa banyaknya data yang akan dikelompokkan.
2. *Input jumlah cluster* berbentuk suatu *teks field* yang digunakan untuk memasukkan nilai dari banyaknya jumlah *cluster* yang dibentuk.
3. *Input pangkat fuzzy (w)* digunakan untuk menentukan nilai *w*. Nilai *input* pangkat fuzzy di default 2.
4. *Input Maksimum Iterasi* digunakan untuk memasukkan nilai dari banyaknya iterasi yang akan dilakukan.
5. *Input Error Terkecil* digunakan untuk memasukkan nilai terkecil yang diharapkan dari selisih fungsi objektif dari setiap iterasi.

2. Fuzzy C-Means

Pada halaman ini digunakan untuk menampilkan perhitungan dari *fuzzy c-means* hingga ditemukannya nilai dari derajat keanggotaan baru. Selain itu juga terdapat *tab result* yang digunakan untuk menampilkan hasil dari pengelompokkan dan perhitungan nilai evaluasi *cluster* dengan *precission*, *recall* dan *f-measure*. Implementasi perancangan antarmuka perhitungan *fuzzy c-means* ditunjukkan pada gambar 4.10 dan *result* ditunjukkan pada gambar 4.11.



MATRIKS DERAJAT BARU

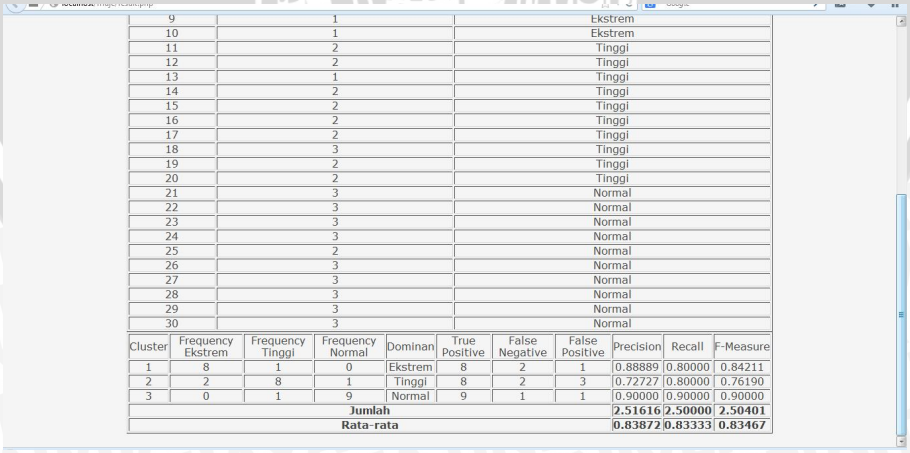
0	1	2	3
0	0.07612	0.0338	0.89009
1	0.08138	0.02812	0.8905
2	0.92212	0.0124	0.06548
3	0.00556	0.00179	0.99265
4	0.46081	0.04992	0.48927
5	0.29805	0.06318	0.63877
6	0.92231	0.01236	0.06532
7	0.2213	0.08755	0.69115
8	0.00561	0.0018	0.99259
9	0.07613	0.0338	0.89007
10	0.52711	0.13672	0.33617
11	0.96907	0.00664	0.02429
12	0.39809	0.07719	0.52471
13	0.93936	0.01517	0.04548
14	0.52713	0.13673	0.33614
15	0.96886	0.00668	0.02446
16	0.96909	0.00663	0.02428
17	0.20198	0.51862	0.2794
18	0.5599	0.10606	0.33404
19	0.96908	0.00663	0.02429
20	0.00582	0.98656	0.00763
21	0.13678	0.72751	0.13571
22	0.00486	0.98918	0.00596
23	0.00487	0.98916	0.00597
24	0.59459	0.12422	0.2812

Gambar 4.10 Implementasi AntarMuka Perhitungan *Fuzzy C-Means*

Keterangan gambar 4.10 adalah sbagai berikut :

1. *Tab Bilangan Random* digunakan untuk menampilkan hasil bilangan random.
2. *Tab Derajat Awal* digunakan untuk menampilkan nilai derajat keanggotaan awal.
3. *Tab Pusat Cluster* digunakan untuk menampilkan nilai dari perhitungan pusat cluster.
4. *Tab Fungsi Objektif* digunakan untuk menampilkan nilai dari perhitungan fungsi objektif.
5. *Tab Derajat Baru* digunakan untuk menampilkan nilai dari derajat keanggotaan baru.
6. *Tab Result* digunakan untuk menampilkan hasil pengelompokkan dan perhitungan nilai *f-measure*.

Hasil dari nilai derajat keanggotaan baru ini digunakan untuk menentukan kecenderungan sebuah data menempati *cluster*. Penentuan sebuah data menempati suatu kelompok dilihat dari nilai terbesar dari derajat keanggotaan pada setiap *cluster*. Kategori limpasan yang telah masuk dalam suatu *cluster* nantinya dilihat dan kemudian dievaluasi dengan menggunakan *f-measure* untuk dihitung nilai akurasi pada setiap *cluster*. Implementasi antarmuka *result* ditunjukkan pada gambar 4.11.



9	1	Ekstrem
10	1	Ekstrem
11	2	Tinggi
12	2	Tinggi
13	1	Tinggi
14	2	Tinggi
15	2	Tinggi
16	2	Tinggi
17	2	Tinggi
18	3	Tinggi
19	2	Tinggi
20	2	Tinggi
21	3	Normal
22	3	Normal
23	3	Normal
24	3	Normal
25	2	Normal
26	3	Normal
27	3	Normal
28	3	Normal
29	3	Normal
30	3	Normal

Cluster	Frequency Ekstrem	Frequency Tinggi	Frequency Normal	Dominan	True Positive	False Negative	False Positive	Precision	Recall	F-Measure
1	8	1	0	Ekstrem	8	2	1	0.88889	0.80000	0.84211
2	2	8	1	Tinggi	8	2	3	0.72727	0.80000	0.76190
3	0	1	9	Normal	9	1	1	0.90000	0.90000	0.90000
Jumlah								**2.51616**	**2.50000**	**2.50401**
Rata-rata								**0.83872**	**0.83333**	**0.83467**

Gambar 4.11 Implementasi Antarmuka *Result*

4.4 Implementasi Pengujian

Pada pengujian, data yang digunakan adalah data tingkat banjir limpasan. Pada pengujian ini menggunakan jumlah data yang berbeda yaitu 30 dataset, 90 dataset dan 120 dataset dengan 4 atribut. Penggunaan jumlah data yang berbeda digunakan untuk melihat apakah *cluster* terbaik yang dihasilkan dari perhitungan memiliki nilai yang sama atau tidak.

Proses pengujian pertama dilakukan untuk mengetahui pengaruh nilai fungsi objektif terhadap banyaknya *cluster* yang dibentuk. Proses pengujian tahap awal ini juga digunakan untuk melihat nilai fungsi objektif dan banyaknya iterasi yang digunakan, hal ini bertujuan untuk melihat *konvegenitas* (fungsi objektif cenderung tidak berubah-ubah). Banyaknya *cluster* yang digunakan pada proses pengujian ini ditentukan sebagai nilai inputan yaitu dari 2 hingga 9.

Pengujian kedua dilakukan untuk mengetahui banyaknya *cluster* optimum yang terbentuk yang dilakukan oleh sistem dilihat dari nilai akurasi *f-measure* yang dihitung dari rata-rata *precission* dan rata-rata *recall* pada masing-masing *cluster*.

BAB V
PENGUJIAN DAN ANALISIS

5.1 Skenario Evaluasi

Seperti yang telah dibahas pada bab 3, terdapat dua macam pengujian yang dilakukan. Pengujian yang pertama dilakukan untuk mendapatkan *error* minimum (ξ) pada saat nilai fungsi objektif telah cenderung konvergen (tidak berubah-ubah). Sedangkan pengujian kedua untuk mengetahui nilai rata-rata akurasi *f-measure* pada setiap cluster dengan error minimum yang telah dihasilkan dari pengujian pertama.

5.2 Hasi Pengujian

Pengujian pertama dilakukan untuk mengetahui nilai pada saat fungsi objektif telah mencapai kondisi cenderung konvergen (tidak berubah-ubah). Kondisi pada saat nilai telah mencapai konvergen inilah yang digunakan sebagai input parameter dalam menghitung nilai akurasi pada setiap *cluster*. Pada pengujian ini jumlah cluster yang digunakan adalah 2, iterasi maksimum adalah 100 dan error minimum adalah 0.0000000001. Hasil ujicoba ditunjukkan pada tabel 5.1 :

Tabel 5.1 Hasil Pengujian Pertama

No.	Iterasi	Fungsi Objektif	Error
1.	2	6804.3620282517	1002.1772445895
2.	3	6431.9489253965	372.41310285527
3.	4	4904.6570151757	1527.2919102208
4.	5	3609.2227963567	1295.434218819
5.	6	3508.6002345114	100.62256184535
6.	7	3505.3039074388	3.2963270725131
7.	8	3505.1292603971	$1.7464704172653 \times 10^{-1}$
8.	9	3505.1190060807	$1.0254316463033 \times 10^{-2}$
9.	10	3505.118395396	$6.1068466175129 \times 10^{-4}$
10.	11	3505.11835755	$3.7845942642889 \times 10^{-5}$

No.	Iterasi	Fungsi Objektif	Error
11.	12	3505.1183559608	$1.5892710507615 \times 10^{-6}$
12.	13	3505.1183558699	$9.0919911599485 \times 10^{-8}$
13.	14	3505.1183558675	$2.3328539100476 \times 10^{-9}$
14.	15	3505.1183558675	0

Pada hasil ujicoba yang dijelaskan pada tabel 5.1, nilai fungsi objektif konvergen (tidak berubah-ubah) ketika mencapai iterasi ke-14 dengan nilai minimum *error* terkecil di bawah 0.00000001 (1×10^{-8}). Maka untuk pengujian selanjutnya nilai yang digunakan sebagai nilai masukan yang digunakan pada saat proses perhitungan mencapai kondisi berhenti adalah pada saat nilai mencapai *error* minimum 0.00000001 dengan iterasi maksimum sebesar 100.

Pengujian kedua digunakan untuk mengetahui pengaruh jumlah *cluster* terhadap akurasi *f-measure* dengan menggunakan algoritma *fuzzy c-means*. Pengujian pada setiap *cluster* dilakukan sebanyak tiga kali secara berulang-ulang agar didapatkan nilai yang valid. Hasil ketiga ujicoba yang dilakukan kemudian dihitung nilai rata-rata *precision*, *recall* dan *f-measure* nya. Hasil pengujian dilakukan pada 2jenis sebaran data yaitu sebaran data yang sama dan sebaran data yang berbeda dengan jumlah dataset yaitu 30 dataset, 90 dataset dan 120 dataset.

5.2.1 Pengujian dengan Sebaran Data yang Sama

5.2.1.1 Pengujian Terhadap 30 Data

Hasil pengujian pada 30 dataset yang dikelompokkan oleh sistem didapatkan hasil yang ditunjukkan pada tabel 5.2 sebagai berikut :

Tabel 5.2 Tabel Pengujian Akurasi *F-Measure* Terhadap Jumlah *Cluster* (30 data)

No.	Nilai Cluster	Rata-rata Precission	Rata-rata Recall	Rata-rata F-Measure
1.	K=2	0.70000	0.95000	0.78333
2.	K=3	0.83872	0.83333	0.83467
3.	K=4	0.86458	0.65000	0.73786
4.	K=5	0.87912	0.52667	0.63543

No.	Nilai Cluster	Rata-rata Precission	Rata-rata Recall	Rata-rata F-Measure
5.	K=6	0.87705	0.43333	0.55129
6.	K=7	0.91791	0.38571	0.51137
7.	K=8	0.92609	0.33750	0.46032
8.	K=9	0.96040	0.31481	0.44610

Dimana :

K = jumlah *cluster*

Hasil yang ditunjukkan pada tabel 5.2 bahwa pengelompokkan dengan menggunakan dataset sebanyak 30 record yang dilakukan sebanyak tiga kali ujicoba dapat dinyatakan bahwa nilai akurasi tertinggi terletak pada *cluster* 3 dengan nilai akurasi *f-measure* sebesar 0.83467.

5.2.1.2 Pengujian Terhadap 90 Data

Hasil pengujian pada 90 dataset yang dikelompokkan oleh sistem didapatkan hasil yang ditunjukkan pada tabel 5.3 sebagai berikut :

Tabel 5.3 Tabel Pengujian Akurasi *F-Measure* Terhadap Jumlah *Cluster* (90 data)

No.	Nilai Cluster	Rata-rata Precission	Rata-rata Recall	Rata-rata F-Measure
1.	K=2	0.63158	0.91667	0.74354
2.	K=3	0.85855	0.83333	0.83055
3.	K=4	0.85588	0.63889	0.70151
4.	K=5	0.86026	0.51333	0.60132
5.	K=6	0.84643	0.43889	0.54003
6.	K=7	0.88783	0.38412	0.49643
7.	K=8	0.88224	0.33611	0.44507
8.	K=9	0.88192	0.29753	0.40817

Dimana :

K = jumlah *cluster*

Hasil yang ditunjukkan pada tabel 5.3 bahwa pengelompokkan dengan menggunakan dataset sebanyak 90 record yang dilakukan sebanyak tiga kali

ujicoba dapat dinyatakan bahwa nilai akurasi tertinggi terletak pada *cluster* 3 dengan nilai akurasi *f-measure* sebesar 0.83055.

5.2.1.3 Pengujian Terhadap 120 Data

Hasil pengujian pada 120 dataset yang dikelompokkan oleh sistem didapatkan hasil yang ditunjukkan pada tabel 5.4 sebagai berikut :

Tabel 5.4 Tabel Pengujian Akurasi *F-Measure* Terhadap Jumlah *Cluster* (120 data)

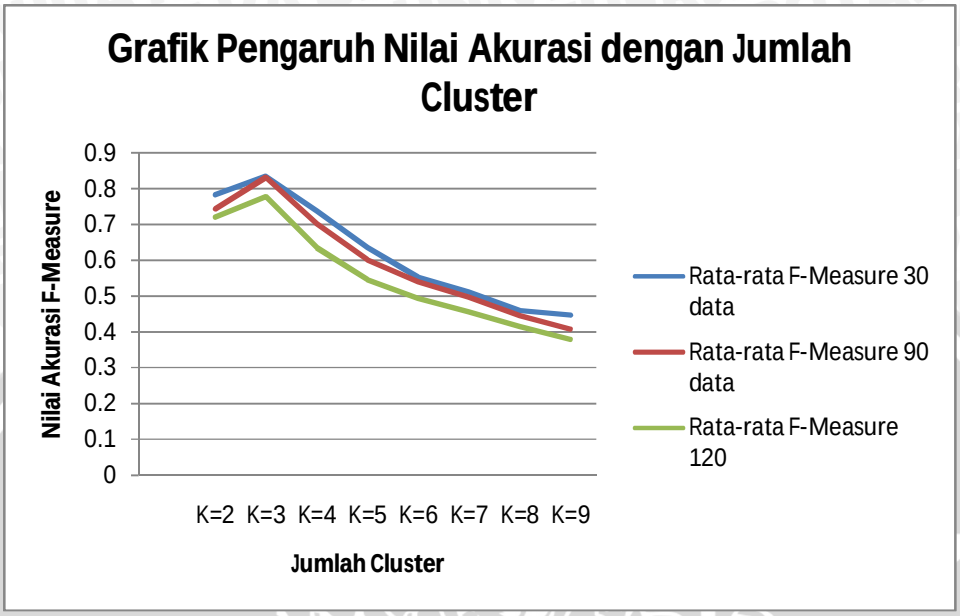
No.	Nilai Cluster	Rata-rata Precission	Rata-rata Recall	Rata-rata F-Measure
1.	K=2	0.60049	0.91092	0.71985
2.	K=3	0.77862	0.79112	0.77685
3.	K=4	0.75100	0.58172	0.63467
4.	K=5	0.75263	0.47625	0.54622
5.	K=6	0.77227	0.40434	0.49431
6.	K=7	0.79637	0.35513	0.45695
7.	K=8	0.79053	0.31087	0.41479
8.	K=9	0.79624	0.27637	0.37953

Dimana :

K = jumlah *cluster*

Hasil yang ditunjukkan pada tabel 5.4 bahwa pengelompokkan dengan menggunakan dataset sebanyak 120 record yang dilakukan sebanyak tiga kali ujicoba dapat dinyatakan bahwa nilai akurasi tertinggi terletak pada *cluster* 3 dengan nilai akurasi *f-measure* sebesar 0.77685.

Hasil ujicoba yang terdapat pada tabel 5.2, tabel 5.3 dan tabel 5.4 dapat digambarkan dalam gambar grafik 5.1 sebagai berikut :



Gambar 5.1 Grafik Pengaruh Nilai Akurasi dengan Jumlah *Cluster* (Jumlah Sebaran Data yang Sama)

Pada gambar grafik 5.1 dapat diketahui bahwa semakin banyak *cluster* yang didapat maka semakin kecil pula nilai dari akurasi *f-measure* dikarenakan kemungkinan suatu dataset masuk ke dalam suatu cluster akan semakin banyak. Nilai akurasi *f-measure* mengalami kenaikan hanya sampai *cluster* 3 dan cenderung mengalami penurunan pada saat *cluster* diatas 3. Selain itu semakin banyak jumlah dataset, nilai akurasi *f-measure* pada *cluster* terbaik semakin sedikit.

5.2.2 Pengujian dengan Sebaran Data yang Berbeda

5.2.2.1 Pengujian Terhadap 30 Data

Hasil pengujian pada 30 dataset yang dikelompokkan oleh sistem didapatkan hasil yang ditunjukkan pada tabel 5.5 sebagai berikut :

Tabel 5.5 Tabel Pengujian Akurasi *F-Measure* Terhadap Jumlah *Cluster* (30 data)

No.	Nilai Cluster	Rata-rata Precission	Rata-rata Recall	Rata-rata F-Measure
1.	K=2	0.76667	0.91667	0.82667
2.	K=3	0.75661	0.64444	0.69336
3.	K=4	0.74107	0.45833	0.54866
4.	K=5	0.85286	0.48667	0.60670
5.	K=6	0.89722	0.43889	0.56727
6.	K=7	0.93061	0.36190	0.48395
7.	K=8	0.94792	0.34167	0.47556
8.	K=9	0.93122	0.28148	0.40868

Dimana :

K = jumlah *cluster*

Hasil yang ditunjukkan pada tabel 5.5 bahwa pengelompokkan dengan menggunakan dataset sebanyak 30 record yang dilakukan sebanyak tiga kali ujicoba dapat dinyatakan bahwa nilai akurasi tertinggi terletak pada *cluster* 2 dengan nilai akurasi *f-measure* sebesar 0.82667.

5.2.2.2 Pengujian Terhadap 90 Data

Hasil pengujian pada 90 dataset yang dikelompokkan oleh sistem didapatkan hasil yang ditunjukkan pada tabel 5.6 sebagai berikut :

Tabel 5.6 Tabel Pengujian Akurasi *F-Measure* Terhadap Jumlah *Cluster* (90 data)

No.	Nilai Cluster	Rata-rata Precission	Rata-rata Recall	Rata-rata F-Measure
1.	K=2	0.73151	0.93333	0.81993
2.	K=3	0.90229	0.85556	0.86686
3.	K=4	0.90733	0.69167	0.74760
4.	K=5	0.91793	0.55833	0.66351
5.	K=6	0.91425	0.46528	0.57624
6.	K=7	0.91862	0.39167	0.51494
7.	K=8	0.90134	0.34583	0.46129

No.	Nilai Cluster	Rata-rata Precission	Rata-rata Recall	Rata-rata F-Measure
8.	K=9	0.91411	0.30741	0.42419

Dimana :

K = jumlah *cluster*

Hasil yang ditunjukkan pada tabel 5.6 bahwa pengelompokkan dengan menggunakan dataset sebanyak 90 record yang dilakukan sebanyak tiga kali ujicoba dapat dinyatakan bahwa nilai akurasi tertinggi terletak pada *cluster* 3 dengan nilai akurasi *f-measure* sebesar 0.86686.

5.2.2.3 Pengujian Terhadap 120 Data

Hasil pengujian pada 120 dataset yang dikelompokkan oleh sistem didapatkan hasil yang ditunjukkan pada tabel 5.7 sebagai berikut :

Tabel 5.7 Tabel Pengujian Akurasi *F-Measure* Terhadap Jumlah *Cluster* (120 data)

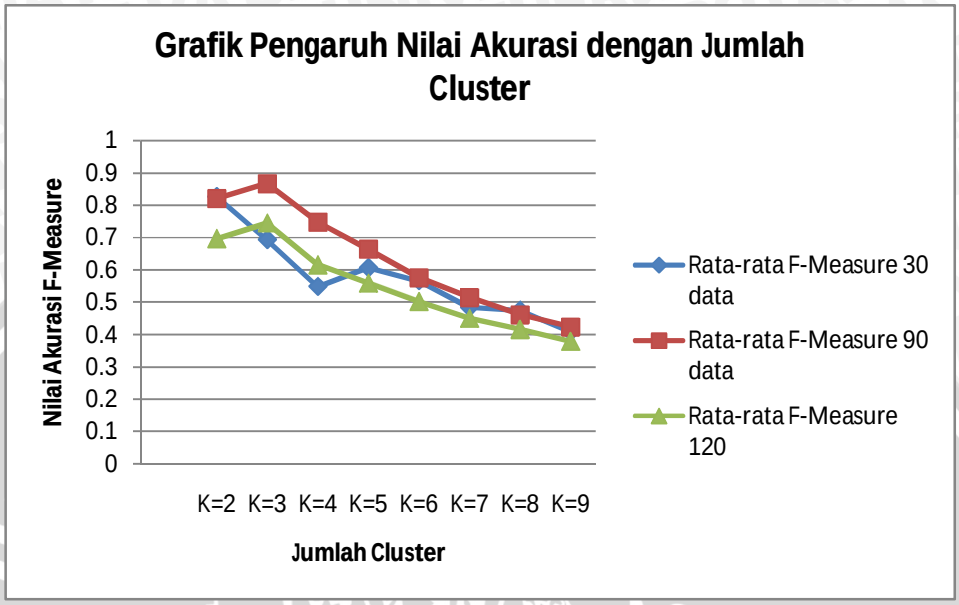
No.	Nilai Cluster	Rata-rata Precission	Rata-rata Recall	Rata-rata F-Measure
1.	K=2	0.56960	0.89683	0.69652
2.	K=3	0.73693	0.78060	0.74461
3.	K=4	0.71641	0.57804	0.61544
4.	K=5	0.76919	0.48466	0.55879
5.	K=6	0.79241	0.40697	0.50250
6.	K=7	0.78120	0.34883	0.45028
7.	K=8	0.81609	0.31217	0.41625
8.	K=9	0.79242	0.27584	0.37928

Dimana :

K = jumlah *cluster*

Hasil yang ditunjukkan pada tabel 5.7 bahwa pengelompokkan dengan menggunakan dataset sebanyak 120 record yang dilakukan sebanyak tiga kali ujicoba dapat dinyatakan bahwa nilai akurasi tertinggi terletak pada *cluster* 3 dengan nilai akurasi *f-measure* sebesar 0.74461.

Hasil ujicoba yang terdapat pada tabel 5.5, tabel 5.6 dan tabel 5.7 dapat digambarkan dalam gambar grafik 5.2 sebagai berikut :



Gambar 5.2 Grafik Pengaruh Nilai Akurasi dengan Jumlah *Cluster* (Jumlah Sebaran Data yang Berbeda)

Pada gambar grafik 5.2 terdapat perbedaan pola grafik pada 30 data. Untuk pola grafik dengan 30 data, semakin banyak nilai *cluster* yang dimasukkan nilai akurasi *f-measure* cenderung menurun, namun mengalami peningkatan dari *cluster* 4 ke *cluster* 5.

Sedangkan untuk grafik 90 data dan 120 data dapat diketahui semakin banyak *cluster* yang didapat maka semakin kecil pula nilai dari akurasi *f-measure* dikarenakan kemungkinan suatu dataset masuk ke dalam suatu cluster akan semakin banyak. Nilai akurasi *f-measure* mengalami kenaikan hanya sampai *cluster* 3 dan cenderung mengalami penurunan pada saat *cluster* diatas 3.

5.3 Analisa *F-Measure* terhadap Jumlah *Cluster*

Untuk hasil pengujian pada tabel 5.2, tabel 5.3 dan tabel 5.4, tiap-tiap jumlah data yang berbeda yaitu 30 data, 90 data dan 120 dengan sebaran data yang sama, tidak ada perubahan terhadap hasil akurasi *f-measure*. Akurasi *f-measure* mengalami kenaikan dari *cluster* 2 sampai *cluster* 3 dan cenderung

menurun ketika *cluster* di atas 3. Pada saat *cluster* berjumlah 3, akurasi *f-measure* mengalami kenaikan dikarenakan jumlah kelas yang ada pada data sesuai dengan jumlah *cluster* yaitu 3 sehingga penyebarannya merata untuk *cluster* dominan. Sedangkan pada saat jumlah cluster di atas 3 cenderung mengalami penurunan, dikarenakan semakin banyaknya jumlah *cluster* yang dimasukkan maka semakin banyak *cluster* yang memiliki nilai dominan yang sama dan otomatis hasil pengelompokkaannya kurang bagus. Untuk tiap-tiap jumlah data dengan sebaran data yang sama masing-masing memiliki cluster terbaik yang sama yaitu *cluster* 3.

Untuk hasil pengujian pada tabel 5.5, tabel 5.6 dan tabel 5.7, dapat dilihat pada jumlah 30 data dengan sebaran data berbeda, semakin banyak nilai *cluster* yang dimasukkan nilai akurasi *f-measure* cenderung menurun, namun mengalami peningkatan dari *cluster* 4 ke *cluster* 5. Sedangkan pada jumlah 90 data dan 120 data, akurasi *f-measure* mengalami kenaikan dari *cluster* 2 sampai *cluster* 3 dan cenderung menurun ketika *cluster* di atas 3. Untuk cluster terbaiknya, terdapat perbedaan *cluster* terbaik pada jumlah 30 data. *Cluster* terbaik pada 30 data dengan sebaran data yang berbeda, berada pada *cluster* 2 dikarenakan selain sebaran data yang tidak sama, data yang diuji lebih sedikit dibanding dengan jumlah data lainnya, yang menyebabkan adanya kelas dominan pada 30 data tersebut sehingga penyebaran *cluster* banyak didominasi oleh kelas yang dominan. Sedangkan pada jumlah 90 data dan 120 data, *cluster* terbaik berada pada cluster 3, dimana penyebarannya merata untuk setiap *cluster* dominan.

5.4 Analisa Jumlah Data terhadap Nilai *F-Measure*

Dapat dilihat pada gambar 5.1 bahwa semakin banyak jumlah data maka semakin sedikit nilai *f-measure* pada *cluster* terbaik. Hal ini dikarenakan semakin banyak data yang digunakan, maka penyebaran nilai dari *true positive*, *false positive* dan *false negative* akan semakin besar juga. Dari 3 parameter itulah yang nantinya akan digunakan untuk menghitung nilai *f-measure*. *False positive* dan *false negative* digunakan sebagai pembanding untuk mencari nilai *precision* dan *recall*. Apabila nilai dari *false positive* dan *false negative* sebagai pembanding semakin besar, maka akan didapatkan nilai *precision* dan *recall* yang semakin kecil daripada nilai *precision* dan *recall* pada jumlah data yang lebih sedikit. Oleh

sebab itu semakin banyak data yang digunakan akan semakin sedikit nilai *f-measure* yang dihasilkan, karena nilai *precision* dan *recall* yang lebih sedikit.

Pada 90 data dan 120 data, nilai dari *true positive* cenderung naik, namun besar kenaikan nilai *true positive* masih di bawah naiknya nilai *false positive* dan *false negative*, sehingga mempengaruhi turunnya nilai *f-measure* pada jumlah data 90 dan 120 bila dibandingkan dengan nilai *f-measure* pada jumlah 30 data.

BAB VI
PENUTUP

6.1 Kesimpulan

Kesimpulan yang didapatkan pada penelitian ini adalah sebagai berikut :

1. Metode *Fuzzy C-Means* dapat digunakan untuk mengelompokkan data yang memiliki nilai bermacam-macam, integer, double, maupun campuran antara keduanya. Dalam kasus kali ini digunakan dataset tingkat banjir limpasan. Hasil yang didapat dari pengelompokkan pada penelitian ini adalah nilai dari derajat keanggotaan dari masing-masing dokumen terhadap sejumlah *cluster*.
2. Berdasarkan pengujian yang telah dilakukan, 5 dari 6 pengujian data yang diujikan menunjukkan bahwa *cluster* terbaik berada pada *cluster* 3, adapun 1 pengujian data yang lain memiliki *cluster* terbaik yang berbeda yaitu *cluster* 2, dikarenakan oleh jumlah data yang diuji lebih sedikit dan sebaran data yang berbeda.
3. Dari beberapa data yang telah dilakukan pengujian, semakin besar jumlah *cluster* yang dimasukkan maka nilai *f-measure* yang dihasilkan cenderung semakin menurun.

6.2 Saran

Beberapa saran untuk pengembangan lebih lanjut dapat diberikan oleh penulis adalah :

1. Menggunakan data dengan nilai variabel yang lebih bervariasi sehingga meminimalkan kemiripan data dan data yang masuk pada satu *cluster* tingkat kemiripannya bisa lebih maksimal.
2. Untuk mendapatkan nilai *f-measure* yang lebih baik dalam penerapan *fuzzy c-means* terhadap pengelompokkan data disarankan agar dilakukannya modifikasi algoritma *fuzzy c-means* dengan algoritma lain.

DAFTAR PUSTAKA

[AGU – 2007] Agusta, Y. 2007. *K-Means- Penerapan, Permasalahan dan Metode Terkait*. Jurnal Sistem dan Informatika Vol.3 No. 47-60.

[ANO – 2012] Anonim. 2012. *Dokumen Pemantauan Hutan di Indonesia*. Overview of the forest monitoring activities by the Ministry of Forestry of Indonesia: Indonesia.

[ANO – 2013] Anonim. 2013. *Limpasan Hujan dan Hidrometri*. Universitas Hasanudin.

[BEZ – 1983] Bezdek, James C., Robert Ehrlich, William Full. 1983. *FCM: The Fuzzy C-Means Clustering Algorithm*.

[DHO – 2010] Dhona, R. 2010. *Perbandingan Metode Fuzzy C-Means dan K-Means pada Analisis Gerombol Bogor*. Skripsi Fakultas MIPA Institut Pertanian Bogor: Bogor.

[DWI – 2012] Dwi, Dela. 2012. *Penerapan Algoritma Improved K-Means Pada Pengelompokkan Data Tingkat Kesehatan Jantung*. Universitas Brawijaya: Malang.

[FUA – 2013] Fuadi, Taufik. 2013. *Accuracy Measure: Precision, Recall & F-Measure*. Bahan Kuliah Data Mining: FMIPA Universitas Syiah Kuala.

[HAD – 2011] Hadisusanto, N. 2011. *Aplikasi Hidrologi*. Jogja Mediautama: Jogjakarta.

[HAM - 2003] Hamilton, H dan Olive, W. 2003. *Confusion Matrix*. http://www2.cs.uregina.ca/dbd/cs831/notes/confusion_matrix/confusion_matrix.html.

[HAN - 2000] Han, J., & Kamber, M. 2000. *Data Mining Concept and Techniques Second Edition*. United Stated: Morgan Kaufman.

[HAN – 2001] Han, Jiawei and Khamber, Micheline. 2001. *Data Minig : Concepts and Techniques*. Morgan Kaufman Publishers, San Fransisco: USA.

[KUS – 2010] Kusumadewi, Sri, Hari Purnomo. 2010. *Aplikasi Logika Fuzzy Untuk Pendukung Keputusan*. Jakarta: Graha Ilmu.

[LEE - 2010] Lee, Finn dan Santana, Juan. 2010. *Data Mining : Meramalkan Bisnis Perusahaan*. PT. Elex Media Komputindo: Jakarta.

[LEW - 1995] Lewis, D. 1995. Evaluating and Optimizing Autonomous Text Classification Systems. AT&T Bell Laboratories Murray Hill, NJ 07974. USA. Proccedings of the Eighteenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, July, 1995, pp. 246-54 http://net.pku.edu.cn/~wbia/2005/public_html/papers/classification/.

[MAN – 2008] Manning, C. D., Raghavan, P., & Schutze, H. 2008. *An Introduction to Information Retrieval*. England: Cambridge University Press.

[MIY – 2008] Miyamoto, Sadaaki, Hidemoto Ichihashi, Katsuhiro Honda. 2008. *Algorithm for Fuzzy Clustering: Method in C-Means Clustering with Applications.* Springen-Verlag Berlin. Heindelberg.

[MUR – 2010] Murwibowo, Paramukti., Gunawan, Totok. 2010. *Aplikasi Penginderaan Jauh dan Sistem Informasi Geografis untuk Mengkaji Perubahan Koefisien Limpasan Permukaan Akibat Letusan Gunung Merapi Tahun 2010 di Sub Das Gendol Yogyakarta*. Yogyakarta.

[PET – 2000] Petridou, Koutsonikola, Vakali dan Papadimtriou. 2000. *A Divergence-Oriented Approach for Web Users Clustering.* Departement of Informatics Aristotle University: Greece.

[PRA – 2011] Prawijiwuri, Gitri. 2011. *Model Erosion Hazard untuk Pengelolaan Sub Daerah Aliran Sungai (DAS) Cisokan Provinsi Jawa Barat.* Universitas Diponegoro: Semarang.

[RIS – 2012] Ristya, Wika. 2012. *Kerentanan Wilayah Terhadap Banjir di Sebagian Cekungan Bandung*. Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia: Depok.

[SAN - 2011] Santoso, Harry. 2011. *Petunjuk Teknis Sistem Standar Operasi Prosedur (SSOP) Penanggulangan Banjir dan Tanah Longsor.* Kementrian Kehutanan: Indonesia.

[SEP - 2013] Septiana, Neni. 2013. *Pengelompokkan Dokumen Berita Berbahasa Indonesia Menggunakan Fuzzy C-Means.* Universitas Brawijaya: Malang.

[SOB – 2009] Sobirin, S. 2009. *Kajian Strategis Solusi Banjir Cekungan Bandung*. Seminar Nasional Teknik Sumber Daya Air: Bandung.

[STE - 2000] Steinbach, M., Karypis, G., dan Kumar, V. 2000. *A Comparison of Document Clustering Techniques*. Technical Report Departement of Computer Science and Engineering: University of Minnesota.

[SUW – 2010] Suwarsono. 2010. *Pengembangan Metode Penentuan Indeks Luas Daun Dari Data Satelit Penginderaan Jauh*. Pusat Pengembangan Pemanfaatan dan Teknologi Penginderaan Jauh Lembaga Penerbangan dan Antariksa Nasional (LAPAN): Jakarta.

[TAL – 2003] Tala, Fadillah Z. 2003. *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Master of Logic Project. Institute for Logic, Language and Computation. Universiteit van Amsterdam. The Netherlands.

[YAN – 1999] Yang, Yiming. 1999. *An Evaluation of Statistical Approaches to Text Categorization*. Journal of Information Retrieval I, pp 69-90, Kluwer Academic Publishers: Netherlands.

[ZAL - 2007] Zalilia, L. 2007. *Penerapan Data Mining untuk IDS*. http://www.budi.insan.co.id/courses/security/2006-2007/Report-Lia-Zalilia.pdf. Diakses tanggal 29 Mei 2013.

Lampiran 1.

Hasil Akurasi *F-Measure*

Dataset : 30

1. **Pengujian 1**

k	*precission*	*recall*	*f-measure*
2	0.70000	0.95000	0.78333
3	0.83872	0.83333	0.83467
4	0.86458	0.65000	0.73786
5	0.90952	0.54000	0.64879
6	0.86349	0.43333	0.55602
7	0.93061	0.38571	0.50623
8	0.93304	0.33750	0.46814
9	0.97778	0.32222	0.46275

2. **Pengujian 2**

k	*precission*	*recall*	*f-measure*
2	0.70000	0.95000	0.78333
3	0.83872	0.83333	0.83467
4	0.86458	0.65000	0.73786
5	0.90952	0.54000	0.64879
6	0.90417	0.43333	0.54185
7	0.91156	0.38571	0.51395
8	0.92262	0.33750	0.45641
9	0.94048	0.30000	0.42411

3. **Pengujian 3**

k	*precission*	*recall*	*f-measure*
2	0.70000	0.95000	0.78333
3	0.83872	0.83333	0.83467
4	0.86458	0.65000	0.73786

k	*precission*	*recall*	*f-measure*
5	0.81833	0.50000	0.60873
6	0.86349	0.43333	0.55602
7	0.91156	0.38571	0.51395
8	0.92262	0.33750	0.45641
9	0.96296	0.32222	0.45145

4. **Rata-Rata Pengujian**

k	*precission*	*recall*	*f-measure*
2	0.70000	0.95000	0.78333
3	0.83872	0.83333	0.83467
4	0.86458	0.65000	0.73786
5	0.87912	0.52667	0.63543
6	0.87705	0.43333	0.55129
7	0.91791	0.38571	0.51137
8	0.92609	0.33750	0.46032
9	0.96040	0.31481	0.44610

Dataset : 90

1. **Pengujian 1**

k	*precission*	*recall*	*f-measure*
2	0.63158	0.91667	0.74354
3	0.85855	0.83333	0.83055
4	0.83981	0.62500	0.69578
5	0.88232	0.50667	0.58994
6	0.84643	0.43889	0.54003
7	0.89033	0.38571	0.49550
8	0.86845	0.33333	0.44344
9	0.85132	0.29259	0.40233

2. **Pengujian 2**

k	*precission*	*recall*	*f-measure*
2	0.63158	0.91667	0.74354
3	0.85855	0.83333	0.83055
4	0.88802	0.66667	0.71299
5	0.82961	0.50000	0.58483
6	0.84643	0.43889	0.54003
7	0.88285	0.38095	0.49829
8	0.89137	0.33333	0.44388
9	0.91138	0.30370	0.41200

3. **Pengujian 3**

k	*precission*	*recall*	*f-measure*
2	0.63158	0.91667	0.74354
3	0.85855	0.83333	0.83055
4	0.83981	0.62500	0.69578
5	0.86885	0.53333	0.62921
6	0.84643	0.43889	0.54003
7	0.89033	0.38571	0.49550
8	0.88690	0.34167	0.44790
9	0.88307	0.29630	0.41020

4. **Rata-Rata Pengujian**

k	*precission*	*recall*	*f-measure*
2	0.63158	0.91667	0.74354
3	0.85855	0.83333	0.83055
4	0.85588	0.63889	0.70151
5	0.86026	0.51333	0.60132
6	0.84643	0.43889	0.54003
7	0.88783	0.38412	0.49643
8	0.88224	0.33611	0.44507

k	*precission*	*recall*	*f-measure*
9	0.88192	0.29753	0.40817

Dataset : 120

1. **Pengujian 1**

k	*precission*	*recall*	*f-measure*
2	0.60049	0.91092	0.71985
3	0.77862	0.79112	0.77685
4	0.75100	0.58172	0.63467
5	0.77619	0.48811	0.56536
6	0.76005	0.39160	0.48007
7	0.78450	0.35189	0.45423
8	0.79053	0.31087	0.41479
9	0.79348	0.27633	0.38488

2. **Pengujian 2**

k	*precission*	*recall*	*f-measure*
2	0.60049	0.91092	0.71985
3	0.77862	0.79112	0.77685
4	0.75100	0.58172	0.63467
5	0.74086	0.47033	0.53665
6	0.76354	0.41088	0.49524
7	0.78450	0.35189	0.45423
8	0.79053	0.31087	0.41479
9	0.83025	0.28127	0.38100

3. **Pengujian 3**

k	*precission*	*recall*	*f-measure*
2	0.60049	0.91092	0.71985
3	0.77862	0.79112	0.77685

k	*precission*	*recall*	*f-measure*
4	0.75100	0.58172	0.63467
5	0.74086	0.47033	0.53665
6	0.79324	0.41054	0.50762
7	0.82013	0.36163	0.46241
8	0.79053	0.31087	0.41479
9	0.76499	0.27151	0.37243

4. Rata-Rata Pengujian

k	*precission*	*recall*	*f-measure*
2	0.60049	0.91092	0.71985
3	0.77862	0.79112	0.77685
4	0.75100	0.58172	0.63467
5	0.75263	0.47625	0.54622
6	0.77227	0.40434	0.49431
7	0.79637	0.35513	0.45695
8	0.79053	0.31087	0.41479
9	0.79624	0.27637	0.37953

Lampiran 2.

Contoh Data Uji

Kemiringan Lereng	Jenis Tanah	Penggunaan Lahan	Pola Aliran	Limpasan
1.40	0.100	1	5	Normal
0.40	0.160	0	10	Normal
0.40	0.160	1	5	Normal
0.40	0.090	1	5	Normal
0.40	0.160	5	15	Normal
1.40	0.100	1	5	Normal
0.40	0.230	1	5	Normal
0.40	0.230	0	10	Normal
0.40	0.160	1	5	Normal
1.40	0.100	1	5	Normal
0.40	0.090	7	15	Tinggi
1.40	0.160	3	20	Tinggi
3.10	0.090	5	15	Tinggi
0.40	0.160	3	20	Tinggi
0.40	0.160	7	15	Tinggi
1.40	0.230	3	20	Tinggi
1.40	0.100	3	20	Tinggi
1.40	0.100	5	10	Tinggi
1.40	0.090	5	15	Tinggi
1.40	0.090	3	20	Tinggi
9.50	0.090	5	15	Ekstrem
6.80	0.090	7	15	Ekstrem
3.10	0.030	3	20	Ekstrem
6.80	0.090	5	15	Ekstrem
6.80	0.090	3	20	Ekstrem
9.50	0.090	4	20	Ekstrem
3.10	0.090	3	20	Ekstrem
6.80	0.090	1	5	Ekstrem
6.80	0.030	5	15	Ekstrem
9.50	0.030	5	15	Ekstrem

Lampiran 3

Contoh Hasil Pengelompokkan pada 30 Data

1. **Jumlah k = 2**

No.	Data ke-	Masuk *cluster* -	Kelas Limpasan
1	1	1	Ekstrem
2	2	1	Ekstrem
3	3	1	Ekstrem
4	4	1	Ekstrem
5	5	1	Ekstrem
6	6	1	Ekstrem
7	7	1	Ekstrem
8	8	1	Ekstrem
9	9	1	Ekstrem
10	10	1	Ekstrem
11	11	1	Tinggi
12	12	1	Tinggi
13	13	1	Tinggi
14	14	1	Tinggi
15	15	1	Tinggi
16	16	1	Tinggi
17	17	1	Tinggi
18	18	2	Tinggi
19	19	1	Tinggi
20	20	1	Tinggi
21	21	2	Normal
22	22	2	Normal
23	23	2	Normal
24	24	2	Normal
25	25	1	Normal
26	26	2	Normal

No.	Data ke-	Masuk cluster -	Kelas Limpasan
27	27	2	Normal
28	28	2	Normal
29	29	2	Normal
30	30	2	Normal

Cluster	E	T	N	Dominan	TP	FN	FP	P	R	F
1	10	9	1	Ekstrem	10	0	10	0,5	1,0	0,67
2	0	1	9	Normal	9	1	1	0,9	0,9	0,9
Jumlah								1,4	1,9	1,57
Rata-Rata								0,7	0,95	0,785

Keterangan :

E : Frekuensi Ekstrem TP : True Positive P : *Precision*

T : Frekuensi Tinggi FN : False Negative R : *Recall*

N : Frekuensi Normal FP : False Positive F : *F-Measure*

2. Jumlah k = 3

No.	Data ke-	Masuk *cluster* -	Kelas Limpasan
1	1	3	Ekstrem
2	2	3	Ekstrem
3	3	1	Ekstrem
4	4	3	Ekstrem
5	5	3	Ekstrem
6	6	3	Ekstrem
7	7	1	Ekstrem
8	8	3	Ekstrem
9	9	3	Ekstrem
10	10	3	Ekstrem
11	11	1	Tinggi
12	12	1	Tinggi

13	13	3	Tinggi
14	14	1	Tinggi
15	15	1	Tinggi
16	16	1	Tinggi
17	17	1	Tinggi
18	18	2	Tinggi
19	19	1	Tinggi
20	20	1	Tinggi
21	21	2	Normal
22	22	2	Normal
23	23	2	Normal
24	24	2	Normal
25	25	1	Normal
26	26	2	Normal
27	27	2	Normal
28	28	2	Normal
29	29	2	Normal
30	30	2	Normal

Cluster	**E**	**T**	**N**	**Dominan**	**TP**	**FN**	**FP**	**P**	**R**	**F**
1	2	8	1	Tinggi	8	2	3	0,72	0,8	0,76
2	0	1	9	Normal	9	1	1	0,9	0,9	0,9
3	8	1	0	Ekstrem	8	2	1	0,88	0,8	0,84
Jumlah								2,51	2,5	2,5
Rata-Rata								0,838	0,833	0,834

Keterangan :

E : Frekuensi Ekstrem
T : Frekuensi Tinggi
N : Frekuensi Normal

TP : True Positive
FN : False Negative
FP : False Positive

P : *Precision*
R : *Recall*
F : *F-Measure*

3. Jumlah k = 4

No.	Data ke-	Masuk *cluster* -	Kelas Limpasan
1	1	4	Ekstrem
2	2	4	Ekstrem
3	3	2	Ekstrem
4	4	4	Ekstrem
5	5	2	Ekstrem
6	6	4	Ekstrem
7	7	2	Ekstrem
8	8	4	Ekstrem
9	9	4	Ekstrem
10	10	4	Ekstrem
11	11	3	Tinggi
12	12	2	Tinggi
13	13	3	Tinggi
14	14	2	Tinggi
15	15	3	Tinggi
16	16	2	Tinggi
17	17	2	Tinggi
18	18	3	Tinggi
19	19	3	Tinggi
20	20	2	Tinggi
21	21	1	Normal
22	22	1	Normal
23	23	1	Normal
24	24	1	Normal
25	25	3	Normal
26	26	1	Normal
27	27	1	Normal
28	28	1	Normal
29	29	1	Normal

No.	Data ke-	Masuk *cluster* -	Kelas Limpasan
30	30	1	Normal

Cluster	E	T	N	Dominan	TP	FN	FP	P	R	F
1	0	0	9	Normal	9	1	0	1,0	0,9	0,947
2	3	5	0	Tinggi	5	5	3	0,625	0,5	0,556
3	0	5	1	Tinggi	5	5	1	0,833	0,5	0,625
4	7	0	0	Ekstrem	7	3	0	1,0	0,7	0,823
Jumlah								3,458	2,6	2,951
Rata-Rata								0,864	0,65	0,737

Keterangan :

E : Frekuensi Ekstrem TP : True Positive P : *Precision*

T : Frekuensi Tinggi FN : False Negative R : *Recall*

N : Frekuensi Normal FP : False Positive F : *F-Measure*

4. Jumlah k = 5

No.	Data ke-	Masuk *cluster* -	Kelas Limpasan
1	1	3	Ekstrem
2	2	3	Ekstrem
3	3	4	Ekstrem
4	4	3	Ekstrem
5	5	4	Ekstrem
6	6	3	Ekstrem
7	7	4	Ekstrem
8	8	3	Ekstrem
9	9	3	Ekstrem
10	10	3	Ekstrem
11	11	1	Tinggi
12	12	4	Tinggi
13	13	1	Tinggi

No.	Data ke-	Masuk *cluster* -	Kelas Limpasan
14	14	4	Tinggi
15	15	1	Tinggi
16	16	4	Tinggi
17	17	4	Tinggi
18	18	5	Tinggi
19	19	1	Tinggi
20	20	4	Tinggi
21	21	2	Normal
22	22	5	Normal
23	23	2	Normal
24	24	2	Normal
25	25	1	Normal
26	26	2	Normal
27	27	2	Normal
28	28	5	Normal
29	29	2	Normal
30	30	2	Normal

Cluster	E	T	N	Dominan	TP	FN	FP	P	R	F
1	0	4	1	Tinggi	4	6	1	0,8	0,4	0,533
2	0	0	7	Normal	7	3	0	1,0	0,7	0,823
3	7	0	0	Ekstrem	7	3	0	1,0	0,7	0,823
4	3	5	0	Tinggi	5	5	3	0,625	0,5	0,555
5	0	1	2	Normal	2	8	1	0.667	0,2	0,307
Jumlah								4,091	2,5	3,043
Rata-Rata								0,818	0,5	0,608

Keterangan :

E : Frekuensi Ekstrem TP : True Positive P : *Precision*

T : Frekuensi Tinggi FN : False Negative R : *Recall*

N : Frekuensi Normal FP : False Positive F : *F-Measure*

5. Jumlah k = 6

No.	Data ke-	Masuk *cluster* -	Kelas Limpasan
1	1	4	Ekstrem
2	2	4	Ekstrem
3	3	5	Ekstrem
4	4	4	Ekstrem
5	5	2	Ekstrem
6	6	2	Ekstrem
7	7	5	Ekstrem
8	8	4	Ekstrem
9	9	4	Ekstrem
10	10	4	Ekstrem
11	11	1	Tinggi
12	12	5	Tinggi
13	13	1	Tinggi
14	14	5	Tinggi
15	15	1	Tinggi
16	16	5	Tinggi
17	17	5	Tinggi
18	18	3	Tinggi
19	19	1	Tinggi
20	20	5	Tinggi
21	21	6	Normal
22	22	3	Normal
23	23	6	Normal
24	24	6	Normal
25	25	1	Normal
26	26	6	Normal
27	27	6	Normal

No.	Data ke-	Masuk *cluster* -	Kelas Limpasan
28	28	3	Normal
29	29	6	Normal
30	30	6	Normal

Cluster	E	T	N	Dominan	TP	FN	FP	P	R	F
1	0	4	1	Tinggi	4	6	1	0,8	0,4	0,533
2	2	0	0	Ekstrem	2	8	0	1,0	0,2	0,333
3	0	1	2	Normal	2	8	1	0,667	0,2	0,307
4	6	0	0	Ekstrem	6	4	0	1,0	0,6	0,75
5	2	5	0	Tinggi	5	5	2	0,714	0,5	0,588
6	0	0	7	Normal	7	3	0	1,0	0,7	0,823
Jumlah								5,180	2,6	3,336
Rata-Rata								0,863	0,433	0,556

Keterangan :

E : Frekuensi Ekstrem
T : Frekuensi Tinggi
N : Frekuensi Normal

TP : True Positive
FN : False Negative
FP : False Positive

P : *Precision*
R : *Recall*
F : *F-Measure*

6. Jumlah k = 7

No.	Data ke-	Masuk *cluster* -	Kelas Limpasan
1	1	4	Ekstrem
2	2	4	Ekstrem
3	3	2	Ekstrem
4	4	4	Ekstrem
5	5	3	Ekstrem
6	6	3	Ekstrem
7	7	2	Ekstrem
8	8	4	Ekstrem
9	9	4	Ekstrem

No.	Data ke-	Masuk *cluster* -	Kelas Limpasan
10	10	4	Ekstrem
11	11	7	Tinggi
12	12	2	Tinggi
13	13	6	Tinggi
14	14	2	Tinggi
15	15	7	Tinggi
16	16	2	Tinggi
17	17	2	Tinggi
18	18	6	Tinggi
19	19	6	Tinggi
20	20	2	Tinggi
21	21	1	Normal
22	22	5	Normal
23	23	1	Normal
24	24	1	Normal
25	25	7	Normal
26	26	1	Normal
27	27	1	Normal
28	28	5	Normal
29	29	1	Normal
30	30	1	Normal

Cluster	E	T	N	Dominan	TP	FN	FP	P	R	F
1	0	0	7	Normal	7	3	0	1,0	0,7	0,823
2	2	5	0	Tinggi	5	5	2	0,714	0,5	0,588
3	2	0	0	Ekstrem	2	8	0	1,0	0,2	0,333
4	6	0	0	Ekstrem	6	4	0	1,0	0,6	0,75
5	0	0	2	Normal	2	8	0	1,0	0,2	0,333
6	0	3	0	Tinggi	3	7	0	1,0	0,3	0,461

Cluster	E	T	N	Dominan	TP	FN	FP	P	R	F
7	0	2	1	Tinggi	2	8	1	0,667	0,2	0,307
Jumlah								6,380	2,7	3,597
Rata-Rata								0,911	0,385	0,513

Keterangan :

E : Frekuensi Ekstrem TP : True Positive P : *Precision*

T : Frekuensi Tinggi FN : False Negative R : *Recall*

N : Frekuensi Normal FP : False Positive F : *F-Measure*

7. Jumlah k = 8

No.	Data ke-	Masuk *cluster* -	Kelas Limpasan
1	1	6	Ekstrem
2	2	2	Ekstrem
3	3	1	Ekstrem
4	4	2	Ekstrem
5	5	1	Ekstrem
6	6	6	Ekstrem
7	7	1	Ekstrem
8	8	2	Ekstrem
9	9	2	Ekstrem
10	10	6	Ekstrem
11	11	3	Tinggi
12	12	8	Tinggi
13	13	5	Tinggi
14	14	8	Tinggi
15	15	3	Tinggi
16	16	8	Tinggi
17	17	8	Tinggi
18	18	5	Tinggi
19	19	5	Tinggi

No.	Data ke-	Masuk *cluster* -	Kelas Limpasan
20	20	8	Tinggi
21	21	4	Normal
22	22	7	Normal
23	23	4	Normal
24	24	4	Normal
25	25	5	Normal
26	26	4	Normal
27	27	4	Normal
28	28	7	Normal
29	29	4	Normal
30	30	4	Normal

Cluster	E	T	N	Dominan	TP	FN	FP	P	R	F
1	3	0	0	Ekstrem	3	7	0	1,0	0,3	0,461
2	4	0	0	Ekstrem	4	6	0	1,0	0,4	0,571
3	0	2	0	Tinggi	2	8	0	1,0	0,2	0,333
4	0	0	7	Normal	7	3	0	1,0	0,7	0,823
5	0	3	1	Tinggi	3	7	1	0,75	0,3	0,428
6	3	0	0	Ekstrem	3	7	0	1,0	0,3	0,461
7	0	0	2	Normal	2	8	0	1,0	0,2	0,333
8	0	5	0	Tinggi	5	5	0	1,0	0,5	0,667
Jumlah								7,75	2,9	4,079
Rata-Rata								0,968	0,362	0,509

Keterangan :

E : Frekuensi Ekstrem
T : Frekuensi Tinggi
N : Frekuensi Normal

TP : True Positive
FN : False Negative
FP : False Positive

P : *Precision*
R : *Recall*
F : *F-Measure*

8. Jumlah k = 9

No.	Data ke-	Masuk *cluster* -	Kelas Limpasan
1	1	3	Ekstrem
2	2	4	Ekstrem
3	3	7	Ekstrem
4	4	4	Ekstrem
5	5	8	Ekstrem
6	6	8	Ekstrem
7	7	7	Ekstrem
8	8	4	Ekstrem
9	9	4	Ekstrem
10	10	3	Ekstrem
11	11	6	Tinggi
12	12	7	Tinggi
13	13	2	Tinggi
14	14	7	Tinggi
15	15	6	Tinggi
16	16	7	Tinggi
17	17	7	Tinggi
18	18	1	Tinggi
19	19	2	Tinggi
20	20	7	Tinggi
21	21	5	Normal
22	22	9	Normal
23	23	5	Normal
24	24	5	Normal
25	25	2	Normal
26	26	5	Normal
27	27	5	Normal
28	28	9	Normal
29	29	5	Normal

No.	Data ke-	Masuk *cluster* -	Kelas Limpasan
30	30	5	Normal

Cluster	E	T	N	Dominan	TP	FN	FP	P	R	F
1	0	1	0	Tinggi	1	9	0	1,0	0,1	0,181
2	0	2	1	Tinggi	2	8	1	0,667	0,2	0,307
3	2	0	0	Ekstrem	2	8	0	1,0	0,2	0,333
4	4	0	0	Ekstrem	4	6	0	1,0	0,4	0,571
5	0	0	7	Normal	7	3	0	1,0	0,7	0,823
6	0	2	0	Tinggi	2	8	0	1,0	0,2	0,333
7	2	5	0	Tinggi	5	5	2	0,714	0,5	0,588
8	2	0	0	Ekstrem	2	8	0	1,0	0,2	0,333
9	0	0	2	Normal	2	8	0	1,0	0,2	0,333
Jumlah								8,380	2,7	3,806
Rata-Rata								0,931	0,3	0,422

Keterangan :

E : Frekuensi Ekstrem
T : Frekuensi Tinggi
N : Frekuensi Normal

TP : True Positive
FN : False Negative
FP : False Positive

P : *Precision*
R : *Recall*
F : *F-Measure*