

# Prototipe Komparasi Model *Clustering* Menggunakan Metode *K-Means* Dan FCM untuk Menentukan Strategi Promosi : Study Kasus Sekolah Tinggi Teknik-PLN Jakarta

Dine Tiara Kusuma<sup>#1</sup>, Nazori Agani<sup>#2</sup>

<sup>#</sup>*Program Studi Magister Ilmu Komputer, Universitas Budi Luhur*

*Jl. Raya Ciledug, Petukangan Utara, Jakarta Selatan 12260*

<sup>1</sup>dinetiarakusuma@gmail.com

<sup>2</sup>nazori.agani@gmail.com

**Abstraksi** —STT-PLN merupakan salah satu perguruan tinggi swasta yang mahasiswanya berasal dari seluruh provinsi yang ada di Indonesia, dimana mahasiswa STT-PLN mengetahui informasi mengenai STT-PLN dari berbagai media Informasi. Penggunaan media informasi ini tentunya sangat berkaitan erat dengan kegiatan promosi, dimana salah satu hal yang sangat penting diperhatikan dalam kegiatan promosi adalah efisiensi biaya. Penggunaan anggaran biaya promosi yang dialokasikan tanpa perencanaan khusus tidak akan optimal jika setiap wilayah kegiatan promosi yang dilakukan memiliki perlakuan yang sama. Agar kegiatan promosi dapat menjadi tepat sasaran maka diperlukan pengolahan data lebih lanjut dengan metode tertentu agar dapat menghasilkan sebuah sistem yang dapat membantu strategi promosi STT-PLN pada tahun berikutnya. Metode pengolahan data yang digunakan dalam penelitian ini adalah dengan menggunakan metode K-Means dan FCM. Hingga hasil penelitian ini adalah berupa sistem yang dapat membantu *Team* promosi Penerimaan Mahasiswa Baru (PMB) di STT-PLN untuk menganalisa hasil *clustering* data 4 tahun sebelumnya.

**Kata Kunci** : *K-Means, FCM, Clustering, Strategi Promosi, Cluster, STT-PLN*

**Abstract** -STT-PLN is a private college that students coming from all provinces in Indonesia, where students STT-PLN find information on STT-PLN from various information media. The use of information media is certainly very closely related to promotional activities, where one of the things that is very important to note in promotional activities is cost efficiency. The use of promotional costs allocated budget without specific planning would not be optimal if every region of promotional activities undertaken have the same treatment. In order for promotional activities can be targeted will require further data processing with a particular method in order to produce a system that can help STT-PLN promotion strategy in the next year. Data processing method used in this research is using the K-Means and FCM. As the result of this research is in the form of a system that can help the *Team* sale Admissions in the STT-PLN to analyze the results of the previous 4 years of data clustering .

**Keywords**: *K-Means, FCM, Clustering, Promotion Strategy, Cluster, STT-PLN*

## I. PENDAHULUAN

Informasi saat ini merupakan sesuatu hal yang telah menjadi sangat penting. Banyak institusi yang telah melakukan komputerisasi dengan salah satu tujuan agar data dari institusi tersebut tidak mudah hilang dan proses analisa dapat berjalan dengan cepat. Pengumpulan dan penyimpanan data dalam suatu institusi secara terus menerus dapat menyebabkan penambahan data yang berdampak pada terjadinya penumpukan data dalam skala yang besar. Apa yang terjadi ini tentu tidak akan memberikan suatu nilai tambah dan tidak akan menghasilkan suatu nilai yang dapat berguna jika data tersebut hanya tertumpuk begitu saja. Data tersebut tentunya memiliki informasi yang tersembunyi yang berharga jika diolah lebih lanjut.

Agar data tersebut dapat menghasilkan suatu yang berguna dan memiliki nilai tambah tersendiri, maka metode analisis dengan tujuan tertentu menjadi sangat dibutuhkan. Metode ini dapat mengelompokkan data tersebut sesuai dengan kelas maupun kelompok nya masing-masing. Salah satu model pengelompokan data berdasarkan kedekatan (kemiripan) di sebut dengan *clustering*.

STT-PLN merupakan salah satu perguruan tinggi yang swasta yang mayoritas mahasiswanya berasal dari berbagai wilayah di seluruh Indonesia. Mahasiswa di STT-PLN mengetahui STT-PLN dari berbagai media. Menurut data yang di peroleh informasi tentang STT-PLN mereka dapatkan dari berbagai macam media. Beda daerah beda media yang informasi yang mereka peroleh. Sehingga media yang ada ini tidak di ketahui secara jelas digunakan di daerah mana saja.

Dalam kegiatan promosi banyak hal yang tentunya harus diperhatikan. Salah satu nya adalah dari segi efisiensi biaya. Biaya yang telah dianggarkan untuk kegiatan promosi haruslah digunakan seoptimal mungkin. Penggunaan anggaran biaya promosi tidak akan optimal jika tanpa perencanaan pengalokasian biaya promosi tersebut kemana dan melalui apa. Oleh karena hal tersebut dibutuhkan suatu strategi khusus untuk melakukan kegiatan promosi agar tepat sasaran dan menjadi lebih efektif dan lebih efisien. Strategi yang dapat dilakukan adalah dengan pengolahan data lebih lanjut. Pengolahan data yang akan dilakukan adalah dengan mengklaster data-data yang diperlukan. Data yang di perlukan dalam pengolahan nya adalah berupa data berasal dari jurusan mana mahasiswa tersebut saat kuliah, data asal provinsi, dan data melalui media apa mahasiswa ini mengetahui STT-PLN. Semua data yang diperlukan ini tentunya memiliki keterkaitan satu sama lain. Hal ini dikarenakan dalam proses *cluster* data nanti nya di butuhkan beberapa *variable* data pendukung agar dapat membantu strategi promosi yang tepat. *Variable* data yang di butuhkan adalah data Jurusan saat mahasiswa tersebut telah kuliah, data asal provinsi mahasiswa tersebut dan data media tempat mereka mengetahui STT-PLN.

Data-data tersebut akan dikelompokkan berdasarkan metode *clustering* yang akan digunakan. Data-data yang dikelompokkan tersebut adalah data masa lalu yang telah ada sehingga hasil dari pengelompokan data tersebut akan menghasilkan sebuah system yang dapat membantu strategi promosi STT-PLN pada tahun berikutnya.

Berdasarkan latar belakang tersebut diatas maka pada penelitian ini akan digunakan metode *clustering* untuk pengelompokan data mahasiswa baru STT-PLN, dengan judul: Prototipe Komparasi Model *Clustering* Menggunakan Metode K-Means dan FCM untuk menentukan Strategi Promosi. Hasil yang diharapkan nantinya dapat membantu para manager dan para pengambil keputusan untuk menentukan strategi promosi berdasarkan keadaan masing-masing wilayah asal mahasiswa tersebut.

## II. LANDASAN TEORI

Analisis model *clustering* adalah sebuah teknik dari analisis multivariabel yang digunakan untuk mengelompokkan objek-objek (*variabel* atau data) sehingga dapat menghasilkan suatu informasi untuk membantu pelaksanaan pengujian terhadap objek dan pada akhirnya dapat menyajikan suatu hipotesis berdasarkan relasi yang terjadi[1]. Prinsip yang digunakan adalah memaksimumkan homogenitas (kesamaan) dalam satu kelompok dan juga memaksimumkan heterogenitas (ketidaksamaan) antar kelompok[2]

*Clustering* adalah salah satu teknik *unsupervised learning* dimana kita tidak perlu melatih metode tersebut atau dengan kata lain, tidak ada *fase learning*. Tujuan dari metode *clustering* adalah untuk mengelompokkan sejumlah data atau objek kedalam klaster sehingga setiap klaster akan terisi data yang semirip mungkin[3]

Sri Kusumadewi dalam bukunya yang berjudul Aplikasi Logika Fuzzy Untuk Pendukung Keputusan, memberikan penjelasan tentang *fuzzy clustering* dalam menentukan pusat *cluster*, yang akan menandai lokasi rata-rata untuk tiap *cluster*[4].

Pada kondisi awal, pusat *cluster* masih belum akurat. Tiap titik data memiliki derajat keanggotaan untuk tiap *cluster*. Dengan cara memperbaiki pusat *cluster* dan derajat keanggotaan tiap titik data secara berulang maka akan dapat dilihat bahwa pusat *cluster* akan dapat dilihat bahwa pusat *cluster* akan bergerak menuju lokasi yang tepat.

### A. Macam-macam model clustering

Yudi Agusta dalam jurnalnya menjelaskan data *Clustering* merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis data *clustering* yang sering dipergunakan dalam proses pengelompokan data yaitu *hierarchical* (hirarki) data *clustering* dan *non-hierarchical* (nonhirarki) data *clustering*. *Fuzzy clustering* merupakan salah satu metode data *clustering non* hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan dari data *clustering* ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses *clustering*, yang pada umumnya berusaha meminimalisasikan *variasi* di dalam suatu *cluster* dan memaksimalkan *variasi* antar *cluster*.

Secara garis besar ada beberapa kategori algoritma *clustering* yang dikenal yaitu salah satu nya adalah metode partisi, dimana pemakai harus menentukan jumlah *k* partisi yang diinginkan lalu setiap data dites untuk dimasukkan pada salah satu partisi sehingga tidak ada data yang overlap dan satu data hanya memiliki satu *cluster*. Contohnya adalah algoritma K-Means, fuzzy C-means.

### B. Fuzzy C – Means (FCM)

Sebelum menjelaskan teory mengenai metode Fuzzy C-Means, sebelumnya akan di bahas terlebih dahulu teory tentang Fuzzy.

### C. Logika Fuzzy

Fuzzy secara bahasa diartikan sebagai kabur atau samar samar. Suatu nilai dapat bernilai besar atau salah secara bersamaan. Dalam fuzzy dikenal derajat keanggotaan yang memiliki rentang nilai 0 (nol) hingga 1(satu). Berbeda dengan himpunan tegas yang memiliki nilai 1 atau 0 (ya atau tidak). Logika Fuzzy merupakan suatu logika yang memiliki nilai kekaburan atau kesamaran (*fuzzyness*) antara benar atau salah. Dalam teori logika fuzzy suatu nilai bias bernilai benar atau salah secara bersama. Namun berapa besar keberadaan dan kesalahan suatu tergantung pada bobot keanggotaan yang dimilikinya. Logika fuzzy memiliki derajat keanggotaan

dalam rentang 0 hingga 1. Berbeda dengan logika digital yang hanya memiliki dua nilai 1 atau 0. Logika fuzzy digunakan untuk menterjemahkan suatu besaran yang diekspresikan menggunakan bahasa (*linguistic*), misalkan besaran kecepatan laju kendaraan yang diekspresikan dengan pelan, agak cepat, cepat, dan sangat cepat. Dan logika fuzzy menunjukan sejauh mana suatu nilai itu benar dan sejauh mana suatu nilai itu salah. Tidak seperti logika klasik (*scrisp*)/tegas, suatu nilai hanya mempunyai 2 kemungkinan yaitu merupakan suatu anggota himpunan atau tidak. Derajat keanggotaan 0 (nol) artinya nilai bukan merupakan anggota himpunan dan 1 (satu) berarti nilai tersebut adalah anggota himpunan. Logika fuzzy adalah suatu cara yang tepat untuk memetakan suatu ruang input kedalam suatu ruang output, mempunyai nilai kontinyu. Fuzzy dinyatakan dalam derajat dari suatu keanggotaan dan derajat dari kebenaran. Oleh sebab itu sesuatu dapat dikatakan sebagian benar dan sebagian salah pada waktu yang sama [4]. Logika Fuzzy memungkinkan nilai keanggotaan antara 0 dan 1, tingkat keabuan dan juga hitam dan putih, dan dalam bentuk linguistik, konsep tidak pasti seperti "sedikit", "lumayan" dan "sangat". Kelebihan dari teori logika fuzzy adalah kemampuan dalam proses penalaran secara bahasa (*linguistic reasoning*). Sehingga dalam perancangannya tidak memerlukan persamaan matematik dari objek yang akan dikendalikan.

Berbeda dengan Fuzzy C-means Clustering (FCM), atau dikenal juga sebagai Fuzzy ISODATA, merupakan salah satu metode *clustering* yang merupakan bagian dari metode *Hard K-Means*. FCM menggunakan model pengelompokan fuzzy sehingga data dapat menjadi anggota dari semua kelas atau *cluster* terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1. Tingkat keberadaan data dalam suatu kelas atau *cluster* ditentukan oleh derajat keanggotaannya. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981.

Dalam teori himpunan *fuzzy* akan memberikan jawaban terhadap sesuatu masalah yang mengandung ketidakpastian. Pada beberapa kasus khusus, seperti nilai keanggotaan yang kemudian akan menjadi 0 atau 1, teori dasar tersebut akan identik dengan teori himpunan biasa, dan himpunan *fuzzy* akan menjadi himpunan *crisp* tradisional. Ukuran *fuzzy* menunjukan derajat kekaburan dari himpunan *fuzzy*. Derajat/ indeks kekaburan merupakan jarak antara suatu himpunan fuzzy A dengan himpunan *crisp* C yang terdekat

*Fuzzy C-means (FCM) Clustering* adalah suatu teknik pengclusteran data yang mana keberadaan tiap-tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan FCM menggunakan model pengelompokan *fuzzy* dengan indeks kekaburan menggunakan *Euclidean Distance* sehingga data dapat menjadi anggota dari semua kelas suatu *cluster* yang terbentuk dengan derajat keanggotaan yang berada antara 0 hingga 1 [5].

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 data memiliki derajat keanggotaan untuk tiap-tiap *cluster*. Dengan cara memperbaiki pusat *cluster* dan nilai keanggotaan tiap-tiap data secara berulang, maka dapat dilihat bahwa pusat *cluster* akan menuju lokasi yang tepat. Perulangan ini didasarkan pada minimasi fungsi obyektif yang menggambarkan jarak dari titik data yang diberikan ke pusat *cluster* yang terbobot oleh derajat keanggotaan titik data tersebut.

Algoritma yang digunakan pada metode Fuzzy C-means adalah sebagai berikut:

- 1) Menginputkan data yang akan di *cluster* X, berupa matriks berukuran  $n \times m$ .

$n$  = jumlah sampel data

$m$  = atribut setiap data.

$X_{ij}$  = data sampel ke- $i$  ( $i=1,2,\dots,n$ ), atribut ke- $j$  ( $j=1,2,\dots,m$ ).

$$X = \begin{bmatrix} X_{11} & \dots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{nm} \end{bmatrix}$$

- 2) Menentukan:

Jumlah *cluster* =  $c$

Pangkat =  $w$

Maksimum Iterasi =  $\text{MaxIter}$

Error Terkecil yang diharapkan =  $\xi$

Fungsi Obyektif awal = 0

Iterasi awal =  $t = 1$

- 3) Membangkitkan bilangan random  $\mu_{ik}$ ,  $i=1,2,\dots,n$ ;  $k=1,2,\dots,c$ ; sebagai elemen-elemen matriks partisi awal U. Menghitung jumlah tiap kolom (atribut) :

$$Q_j = \sum_{k=1}^c \mu_{ik}$$

$$\text{Hitung : } \mu_{ik} = \frac{\mu_{ik}}{Q_j}$$

- 4) Menghitung pusat *cluster* ke -  $k$ :  $V_{kj}$ , dengan  $k=1,2,\dots,c$ ; dan  $j=1,2,\dots,m$

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w}$$

$$V = \begin{bmatrix} V11 & \cdots & V1m \\ \vdots & \ddots & \vdots \\ Vc1 & \cdots & Vcm \end{bmatrix}$$

5) Menghitung fungsi Obyektif pada iterasi ke=t, Pt :

$$Pt = \sum_{k=1}^n \sum_{i=1}^m \left( \left[ \sum_{j=1}^m ((x_{ik} - v_{ik})^2) \right] \mu_{ik}^w \right)$$

6) Menghitung perubahan matriks partisi :

$$\mu_{ik} = \frac{\left[ \sum_{j=1}^m (x_{ik} - v_{ik})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^c \left[ \sum_{j=1}^m (x_{ik} - v_{ik})^2 \right]^{\frac{-1}{w-1}}}$$

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

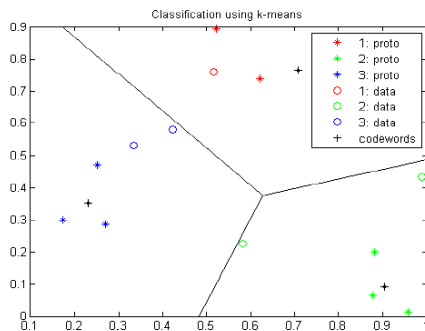
Cek kondisi berhenti :

Jika  $(|Pt - Pt-1| < \xi)$  atau  $(t > \text{MaxIter})$  maka berhenti;

Jika tidak :  $t=t+1$ , ulangi langkah ke - 4.

#### D. Algoritma K-means

Algoritma K-Means merupakan satu algoritma yang mudah dan kerap digunakan di dalam teknik pengelompokan karena melibatkan pengiraan yang efisien dan tidak memerlukan banyak parameter. K-Means menggunakan k kelompok yang telah ditetapkan (k kelompok pertama sebagai centroid) dan secara berterusan akan melalui proses pengiraan titik tengah (min) sehingga sesuatu fungsi kriteria dicapai (kelompok adalah tetap)[6]. Di dalam teknik pengelompokan, pengiraan untuk membedakan di antara kelompok dilakukan menggunakan satu algoritma yang dipanggil fungsi jarak yaitu tahap persamaan atau perbedaan.



Gbr. 1 Contoh Perolehan Titik Pusat Menggunakan Metode K-Means di Matlab

Dasar algoritma K-means adalah sebagai berikut :

- 1) Tentukan nilai k sebagai jumlah kluster yang ingin dibentuk.
- 2) Bangkitkan k *centroid* (titik pusat kluster) awal secara random.
- 3) Hitung jarak setiap data ke masing-masing centroid menggunakan rumus korelasi antar dua objek yaitu *Euclidean Distance* dan kesamaan *Cosine*.

- 4) Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya.
- 5) Tentukan posisi centroid baru (  $k \ C$  ) dengan cara menghitung nilai rata-rata dari data-data yang ada pada *centroid* yang sama.

$$C_k = \left( \frac{1}{n_k} \right) \sum d_i$$

Dimana  $k$  n adalah jumlah dokumen dalam *cluster* k dan  $i$  d adalah dokumen dalam *cluster* k.

- 6) Kembali ke langkah 3 jika posisi centroid baru dengan centroid lama tidak sama.

Adapun karakteristik dari algoritma K-Means salah satunya adalah sangat sensitif dalam penentuan titik pusat awal kluster karena K-Means membangkitkan titik pusat kluster awal secara random. Pada saat pembangkitan awal titik pusat yang random tersebut mendekati solusi akhir pusat kluster, K-Means mempunyai kemungkinan yang tinggi untuk menemukan titik pusat kluster yang tepat. Sebaliknya, jika awal titik pusat tersebut jauh dari solusi akhir pusat kluster, maka besar kemungkinan ini menyebabkan hasil pengklasteran yang tidak tepat. Akibatnya K-Means tidak menjamin hasil pengklasteran yang unik. Inilah yang menyebabkan metode K-Means sulit untuk mencapai optimum global, akan tetapi hanya minimum lokal. Selain itu, algoritma K-Means hanya bisa digunakan untuk data yang atributnya bernilai numeric.

Pengukuran persamaan atau jarak merupakan tugas yang penting di dalam proses analisa kelompok di mana hampir semua teknik pengelompokan menggunakan pengiraan matriks jarak (atau perbedaan)[7]. Algoritma K-Means juga menggunakan kaedah pengiraan ini bagi menjelaskan lagi persamaan bagi setiap corak kelompok. Matriks Jarak Euclidean merupakan salah satu matriks jarak yang sering digunakan di dalam algoritma K-Means.

Matriks Jarak Euclidean

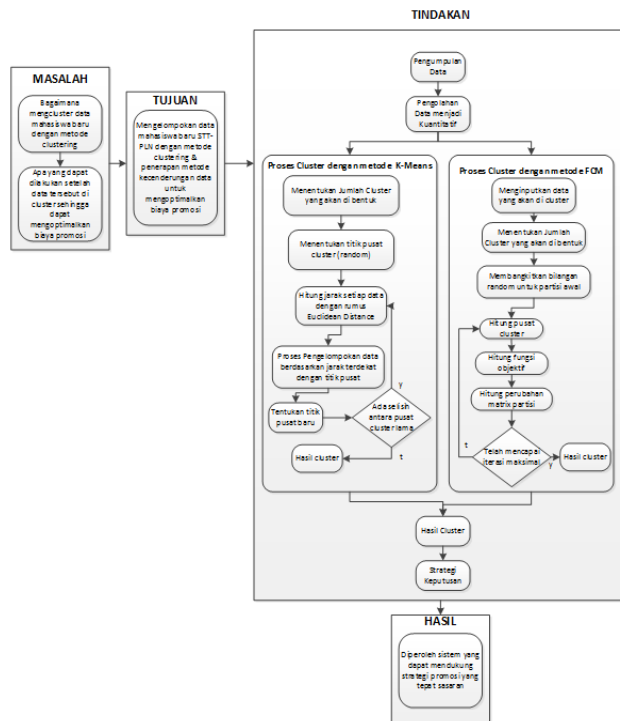
$$d(x,y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

$d(x,y)$  = jarak di antara x dan y

$y_i$  = nilai pembolehubah  $i$  bagi x

$x_i$  = nilai pembolehubah  $i$  bagi y

#### E. Kerangka Konsep



Gbr.2 Kerangka Konsep

#### F. Hipotesis

Berdasarkan kajian teoritis dan kerangka pikir yang telah dikemukakan diatas, maka dapat diduga hasil dari penelitian ini adalah proses pengelompokan data mahasiswa baru dengan menggunakan metode *clustering* dapat mengelompokkan data-data mahasiswa baru berdasarkan kemiripan karakteristik dari data tersebut sehingga dapat diperoleh beberapa himpunan *cluster* yang dapat digunakan untuk strategi promosi STT-PLN pada tiap masing-masing wilayah seluruh Indonesia.

### III. METODOLOGI

Metode penelitian yang digunakan dalam penelitian ini adalah dengan menggunakan metode penelitian deskriptif kuantitatif, dimana dalam melakukan penelitian ini metode deskriptif digunakan untuk menganalisis dengan mengemukakan data berdasarkan keadaan yang sebenarnya. Setelah data-data tersebut dikumpulkan kemudian dilakukan pengolahan data secara kuantitatif dengan melakukan proses penjumlahan, kemudian diimplementasikan sesuai dengan kebutuhan algoritma yang digunakan sehingga dapat dilakukan proses analisis terhadap data tersebut.

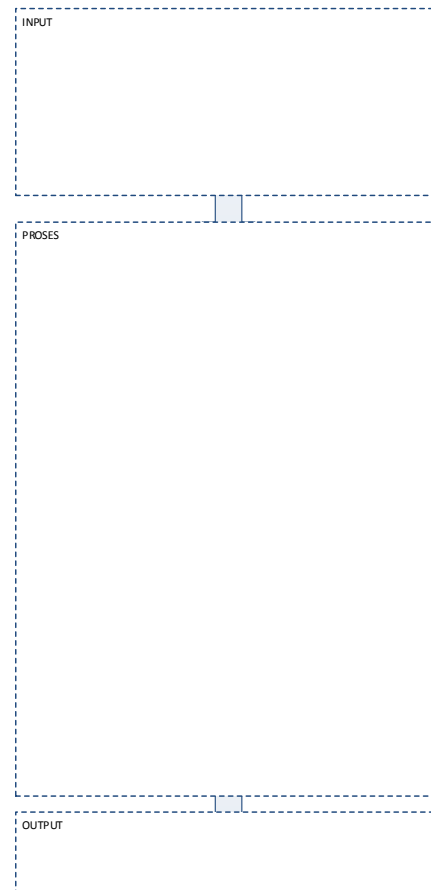
Data yang akan digunakan dalam penelitian ini adalah data masa lalu, diambil dari 4 tahun terakhir yaitu data mahasiswa baru STT-PLN pada tahun 2011, 2012, 2013, dan 2014 beserta dengan media perolehan informasi mereka mengenai STT-PLN. Data yang di peroleh ini adalah data skunder

dimana diperoleh dari dokumentasi sebuah lembaga yakni STT-PLN.

Pengumpulan data dilakukan dengan melakukan :

- Wawancara :  
Ketua Sekolah Tinggi Teknik PLN  
Wakil Ketua Bidang Akademik  
Wakil Ketua Bidang Non-Akademik
- Kuisisioner  
Untuk data: kepada calon mahasiswa baru  
Untuk testing sistem : *Team* Penerimaan Mahasiswa Baru (PMB) STT-PLN
- Dokumentasi :  
Dari hasil kuisisioner tersebut maka akan terbentuk sebuah dokumentasi yang diperlukan untuk pengolahan data lebih lanjut untuk penelitian ini.  
Hasil dari dokumentasi tersebut akan direkapitulasi sehingga akan diperoleh data kuantitatif untuk di olah lebih lanjut.

Langkah-langkah penelitian yang dilakukan :



Gbr.3 Langkah-Langkah Penelitian



#### IV. ANALISA HASIL

##### A. Analisa Hasil K-Means

Data yang digunakan adalah data Mahasiswa Baru Tahun 2011, 2012, 2013, dan 2014, dimana berdasarkan hasil perhitungan jumlah mahasiswa baru aktif pada tahun-tahun tersebut adalah sebagai berikut :

TABEL 1.  
JUMLAH SAMPEL DATA

Tahun	Jumlah
2011	666
2012	557
2013	595
2014	862

Proses *cluster* dengan metode K-Means :

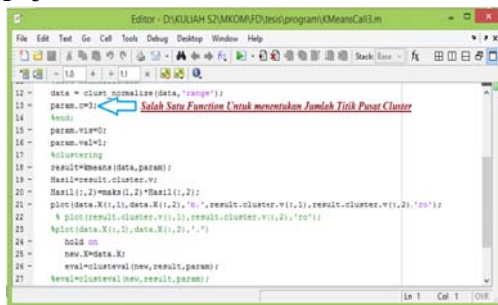
##### a. Tentukan jumlah *cluster* K

K = 2, atau

K = 3, atau

K = 4, atau

K = 5



Gbr.4 Potongan Code Untuk Menentukan Jumlah Cluser yang akan Dibentuk

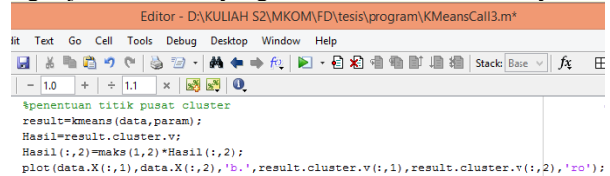
Pada code di atas tampak bahwa “param.c=3” ini adalah bagian dari code untuk menentukan jumlah *cluster* yang akan di bentuk. “param.c=3” artinya jumlah *cluster* yang akan dibentuk adalah 3. Jika ingin menentukan jumlah *cluster* sebanyak 2, 4 atau 5 dapat diganti dengan “param.c=2”, atau “param.c=4”, atau “param.c=5”

##### b. Tentukan titik pusat *cluster* (ditentukan secara Random)

Untuk memperoleh titik pusat *cluster* dibutuhkan *Matrix* inputan data dalam bentuk kuantitatif yang telah melewati proses akumulasi sebelumnya. Data-data tersebut kemudian diinput ke matlab menggunakan fasilitas input variable. Data-data tersebut di input dan dikelompokkan berdasarkan variable masing-masing tahun dengan format file \*.mat

Setelah data-data tersebut telah di input maka dengan menggunakan algoritma K-Means untuk pengacakan data sebagai titik pusat *cluster* digunakan salah satu potongan

program berikut maka titik pusat *cluster* akan terbentuk sesuai dengan jumlah *cluster* yang telah ditentukan sebelumnya.



Gbr.5 Potongan Program Untuk Menentukan Titik Pusat Cluster

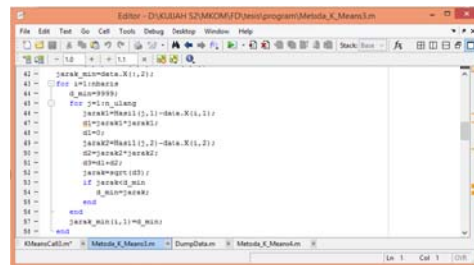
Titik pusat *cluster* yang di bentuk akan ditentukan secara random berdasarkan data yang telah diinput sebelumnya.

##### c. Menghitung jarak setiap data yang akan di *cluster*

Setiap data yang akan di *cluster* akan di tentukan jarak terdekat kelompok ke titik pusat objek nya yang telah di peroleh dengan menggunakan rumus *euclidiens distnce* berikut :

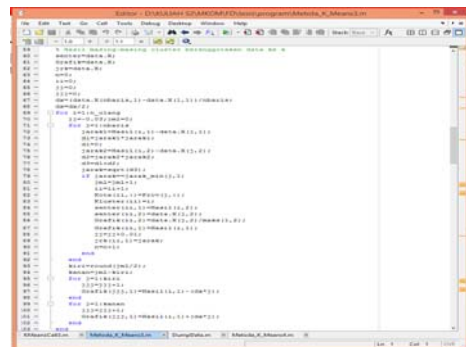
$$D_{ik}^2 = (x_k - v_i)^T (x_k - v_i), \quad 1 \leq i \leq c, 1 \leq k \leq N$$

Berikut potongan program di matlab untuk menentukan jarak tiap masing-masing data



Gbr.6 Potongan Program Untuk Menentukan Jarak Data

d. Kelompokkan data berdasarkan jarak titik pusat terdekat  
Proses selanjutnya adalah mengelompokkan data berdasarkan jarak terdekat dengan titik pusat *cluster*, berikut adalah code program di matlab pengelompokan data berdasarkan titik pusat terdekatnya



Gbr.7 Pengelompokan Data Berdasarkan Jarak Titik Pusat Terdekat

e. Tentukan posisi centroid baru ( $k$  C)

Dengan cara menghitung nilai rata-rata dari data-data yang ada pada centroid yang sama

$$C_K = \left( \frac{1}{n_k} \right) \sum d_i$$

Berikut adalah potongan program untuk menentukan titik pusat (*Centroid*) yang baru.

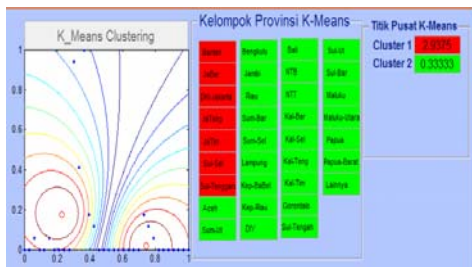
```

36 - ukuran=size(data.X);
37 - nbaris=ukuran(1,1);
38 - nbaris=nbaris;
39 - ukuran=size(Hasil);
40 - nulang=ukuran(1,1);
41 - data.X(:,2)=maks(1,2)*data.X(:,2);
42 - jarak_min=data.X(:,2);

```

Gbr.8 Potongan Program Untuk Menentukan Centroid Baru

Agar sistem ini dapat dipergunakan oleh pengguna dengan mudah maka dirancanglah suatu bentuk antar muka yang dapat memudahkan pengguna dalam melakukan proses analisa akhir dan membuat laporan sesuai dengan kebutuhan

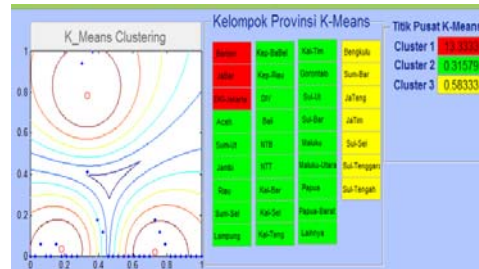


Gbr.9 Tampilan metode K-Means dengan 2 cluster

Pada tampilan diatas tampak bahwa hasil *clustering* pada tahun 2011 dengan media informasi Iklan / Surat kabar dengan metode K-Means adalah :

**Cluster 1 :** Banten, Jawa Barat, DKI Jakarta, Jawa Tengah, Jawa Timur, Sulawesi selatan, Sulawesi Tenggara.

**Cluster 2 :** Aceh, Sumatra Utara, Bengkulu, Jambi, Riau, Sumatra Barat, Sumatra Selatan, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DIY, Bali, NTB, Kalbar, Kalteng, Kaltim, Gorontalo, Sulawesi Tengah, Sulawesi Utara, Sulawesi Barat, Maluku, Maluku Utara, Papua, Papua Barat, Lainnya.



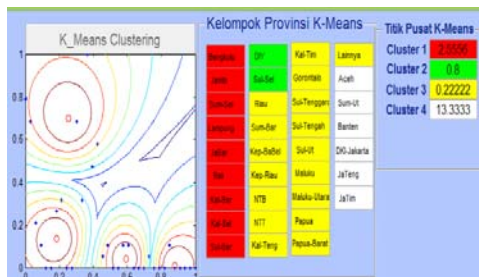
Gbr.10 Tampilan metode K-Means dengan 3 Cluster

Pada tampilan diatas tampak bahwa hasil *clustering* pada tahun 2011 dengan media informasi Iklan / Surat kabar dengan metode K-Means adalah :

**Cluster 1 :** Banten, Jawa Barat, DKI Jakarta

**Cluster 2 :** Aceh, Sumatra Utara, Jambi, Riau, Sumatra Selatan, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DIY, Bali, NTB, Kalbar, Kalteng, Kaltim, Gorontalo, Sulawesi Tengah, Sulawesi Utara, Sulawesi Barat, Maluku, Maluku Utara, Papua, Papua Barat, Lainnya

**Cluster 3 :** Bengkulu, Sumatra Barat, Jawa Tengah, Jawa Timur, Sulawesi selatan, Sulawesi Tenggara, Sulawesi Tengah



Gbr.11 Tampilan metode K-Means dengan 4 cluster

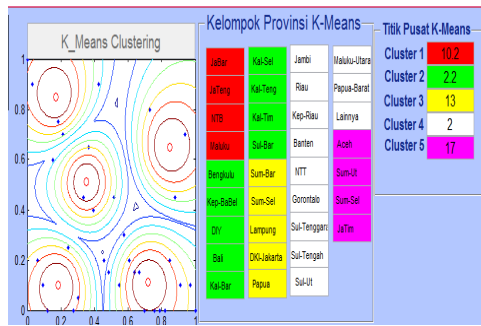
Pada tampilan diatas tampak bahwa hasil *clustering* pada tahun 2011 dengan media informasi Iklan / Surat kabar dengan metode K-Means adalah :

**Cluster 1:** Aceh, Sumatra Utara, Banten, DKI Jakarta, Jawa Tengah, Jawa Timur

**Cluster 2:** Bengkulu, Jambi, Sumatra Selatan, Lampung, Jawa Barat, Bali, Kalbar, Kalsel, Sulawesi Barat

**Cluster 3 :** DIY, Sulawesi Selatan

**Cluster 4 :** Riau, Sumatra Barat, Kepulauan Bangka Belitung, Kepulauan Riau, NTB, NTT, Kalimantan Tengah, Kalimantan Timur, Gorontalo, Sulawesi Tengah, Sulawesi Utara, Maluku, Maluku Utara, Papua, Papua Barat



Gbr.12 Tampilan metode K-Means dengan 5 cluster

**Cluster 1 :** Aceh, Sumatra Utara, Sumatra Selatan, Jawa Timur

**Cluster 2 :** Sumatra Barat, Sumatra Selatan, Lampung, DKI Jakarta, Papua

**Cluster 3:** Jawa Barat, Jawa Tengah, NTB, Maluku

**Cluster 4:** Bengkulu, Kepulauan Babel, DIY, Bali, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, Sulawesi Barat

**Cluster 5:** Jambi, Riau, Kepulauan Riau, Banten, NTT, Gorontalo, Sulawesi Tenggara, Sulawesi Tengah, Sulawesi Utara, Maluku Utara, Papua Barat, Lainnya

## B. Analisa Hasil FCM

Data yang digunakan adalah data Mahasiswa Baru Tahun 2011, 2012, 2013, dan 2014, dimana berdasarkan hasil perhitungan jumlah mahasiswa baru aktif pada tahun-tahun tersebut adalah sebagai berikut :

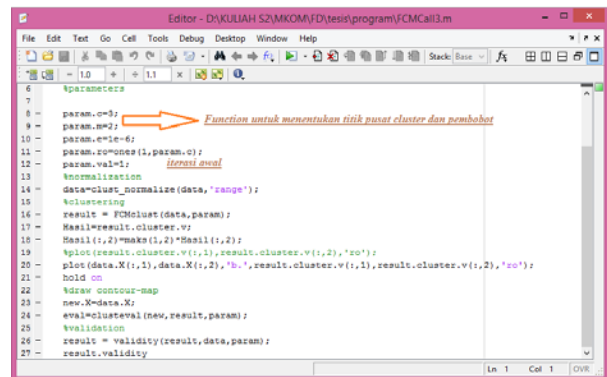
TABEL II.  
JUMLAH SAMPEL DATA

Tahun	Jumlah
2011	666
2012	557
2013	595
2014	862

- Menentukan :  
Jumlah cluster : 2 atau 3 atau 4 atau 5  
Pangkat (pembobot) = 2  
Maksimum iterasi = 100  
Error terkecil yang diharapkan = 0.01

Fungsi Objektif awal = 0

Iterasi awal = 1



Gbr.13 Code Untuk Menentukan Jumlah Cluser yang akan Dibentuk dan pembobot awal pada alg. FCM

- Membangkitkan bilangan random untuk *Matrix* partisi awal

```
%normalization
data=clust_normalize(data,'range');
%clustering
result = FCMclust(data,param);
```

Gbr.14 Potongan Code untuk membangkitkan bilangan Random FCM

- Menghitung pusat cluster

```
data.X(:,2)=maks(1,2)*data.X(:,2);
jarak_min=data.X(:,2);
```

Gbr.15 Potongan Program Untuk Menghitung Pusat Cluster

- Menghitung fungsi obyektif pada iterasi ke = t, Pt :

```
data=clust_normalize(data,'range');
```

Gbr.16 Potongan Program Untuk menghitung fungsi obyektif

- Menghitung perubahan matriks partisi

```
data.Xold=data.X;
if strcmp(method,'range')
    data.min=min(data.X);
    data.max=max(data.X);
    data.X=(data.X-repmat(min(data.X),size(data.X,1),1))./(repmat(max(data.X),...
    size(data.X,1),1)-repmat(min(data.X),size(data.X,1),1));
elseif strcmp(method,'var')
    data.X=(data.X-repmat(mean(data.X),size(data.X,1),1))./(repmat(std(data.X),size(data.X),
    data.mean=mean(data.X);
    data.std=std(data.X);
else
    error('Unknown method given')
end
```

Gbr.17 Potongan program untuk Menghitung Perubahan Matrix Partisi



The figure displays the results of Fuzzy Clustering (FCM) for Indonesian provinces. On the left, a 'Fuzzy Clustering' plot shows membership values on the y-axis (0 to 1) against province codes on the x-axis (0 to 1). The plot features several concentric contour lines representing different membership levels. On the right, a table titled 'Kelompok Provinsi FCM' lists the provinces assigned to two clusters. Cluster 1 (red background) includes Aceh, Sum-UT, Bengkulu, Jambi, Riau, Sum-Bar, Sum-Sel, Lampung, and Kep-Babel. Cluster 2 (green background) includes Kepulauan, DKI-Jakarta, Jawa-Tengah, Jawa-Bar, Jawa-Tim, Kalimantan, Papua, Papua-Barat, and Lampung. To the right of the table, a box titled 'Titik Pusat FCM' shows the cluster centroids: Cluster 1 at 0.83095 and Cluster 2 at 1.5371.

Kelompok Provinsi FCM			
Aceh	Kep-Rau	Korontalo	Banten
Sum-UT	DKI	Sul-Tengah	Jabar
Bengkulu	Bali	Sul-Lit	DKI-Jakarta
Jambi	NTB	Sul-Bar	Jawa-Teng
Riau	NTT	Kaluku	Jawa-Tim
Sum-Bar	Kal-Bar	Kaluku-Utara	Sul-Sel
Sum-Sel	Kal-Sel	Papua	Sul-Tenggara
Lampung	Kal-Teng	Papua-Barat	
Kep-Babel	Kal-Tim	Lampung	

**Titik Pusat FCM**

Cluster 1: 0.83095  
Cluster 2: 1.5371

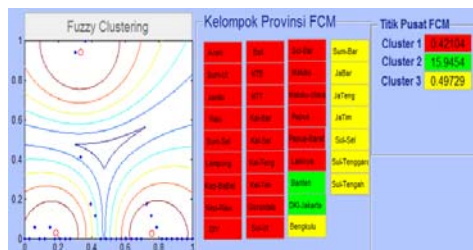
Kelompok	Provinsi	Kepala	Kabupaten
Cluster 1	Sumatera	Sumatera	Sumatera
Cluster 2	Jawa	Jawa	Jawa
Cluster 3	Sumatera	Sumatera	Sumatera
Cluster 4	Sumatera	Sumatera	Sumatera

Titik Pusat FCM

Kelompok	Titik Pusat FCM
Cluster 1	0.307
Cluster 2	0.289
Cluster 3	0.285
Cluster 4	1.046

*Cluster 4: Riau, Sumatra Barat, Kepulauan Bangka Belitung, Kepulauan Riau, NTB, NTT, Kalimantan Tengah, Kalimantan Timur, Gorontalo, Sulawesi Tengah, Sulawesi Utara, Maluku, Maluku Utara, Papua, Papua Barat*

*Cluster 2* : Aceh, Sumatra Utara, Bengkulu, Jambi, Riau, Sumatra Barat, Sumatra Selatan, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DIY, Bali, NTB, Kalbar, Kalteng, Kaltim, Gorontalo, Sulawesi Tengah, Sulawesi Utara, Sulawesi Barat, Maluku, Maluku Utara, Papua, Papua Barat, Lainnya.



**Fuzzy Clustering**

**Kelompok Provinsi FCM**

Provinsi	Kelompok
Sul-Sel	Kep-Riau
Provinsi Irian	Banten
Lampung	Gorontalo
Provinsi Bangka	Sul-Tenggara
Sul-Jawa	Sul-Tengah
Bengkulu	Sul-Lit
Makulu	Lainnya
Provinsi	Sul-Sel
Kep-Babel	Sul-Sel
Aceh	Sul-Sel
Jambi	Sul-Sel
JaBar	Sul-Sel
JaTeng	Sul-Sel
Bali	Sul-Sel
DIY	Sul-Sel
NTT	Sul-Sel
NTB	Sul-Sel
Kali-Teng	Sul-Sel
Kali-Tim	Sul-Sel

**Titik Pusat FCM**

Cluster	1	2	3	4	5
Cluster 1	14.142				
Cluster 2	2.151				
Cluster 3	4.1805				
Cluster 4	15.343				
Cluster 5	1.1017				

**Cluster 5** : Jambi, Riau, Kepulauan Riau, Bnaten, NTT, Gorontalo, Sulawesi Tenggara, Sulawesi Tengah, Sulawesi Utara, Maluku Utara, Papua Barat, Lainnya

Pengujian sistem dilakukan untuk mengetahui bahwa sistem yang dibuat telah berjalan dengan baik. Pengujian dilakukan dengan :

## a. Alpha Test :

$$\text{Kriteria penilaian : } \frac{\text{Total Score Bobot}}{\text{Jumlah Butir Uji}} \times 100\%$$

$$\begin{aligned} \text{Total Score Bobot} &= \frac{P1+P2+P3+P4+P5+P6+P7+P8+P9+P10+P11+P12+P13+P14}{14} \times 100\% \\ &= \frac{1+1+1+1+1+1+1+1+1+1+1+1+1+1}{14} \times 100\% = 100\% \end{aligned}$$

Berdasarkan uji coba sistem oleh salah satu *Team PMB* yang merupakan Koordinator Programing Pengembangan Sistem dapat disimpulkan bahwa 100% program telah sesuai dengan algoritma K-Means dan tidak terdapat *error* pada program.

## b. User Acceptance Test

Metode yang digunakan untuk menguji sistem ini dari sisi *user* adalah dengan mengadaptasi karakteristik dari ISO 9126, sehingga dari pengolahan hasil dari uji coba sistem ini adalah sebagai berikut :

TABEL III.  
HASIL UJI SISTEM OLEH USER

Aspek	Skor Aktual	Skor Ideal	% Skor Aktual	Kriteria
<b>Functionality</b>	505	600	84.17%	Sangat Baik
<b>Reliability</b>	315	400	78.75%	Baik
<b>Usability</b>	262.5	300	87.50%	Sangat Baik
<b>Efficiency</b>	145	200	72.50%	Baik
<b>Total</b>	<b>1227.5</b>	<b>1500</b>	<b>81.83%</b>	<b>Baik</b>

Hasil dari penelitian ini rencana untuk di implementasikan bersamaan dengan kegiatan Penerimaan Mahasiswa Baru tahun 2015.

TABEL IV.  
RENCANA IMPLEMENTASI

No	Kegiatan	Bulan (2014-2015)					
		Sept	Oct	Nov	Des	Jan	Feb
1	Uji Coba Sistem					√	√
2	Perolehan hasil <i>cluster</i>					√	√
3	Penentuan hasil <i>cluster</i> dengan wilayah tertinggi peminat nya dan wilayah yang rendah peminat nya sesuai dengan hasil <i>cluster</i> data 4 tahun terakhir					√	√
4	Pelaporan hasil <i>cluster</i> dengan wilayah tertinggi peminat nya dan wilayah yang rendah peminat nya sesuai dengan hasil <i>cluster</i> data 4 tahun terakhir kepada <i>team PMB STT-PLN 2015</i>						√
5	Rapat strategi promosi oleh Para Pimpinan						√
6	Pemberian hasil keputusan strategi promosi						√
7	Implementasi promosi						√

## V. KESIMPULAN

Proses pengelompokan data mahasiswa baru di STT-PLN dapat dilakukan dengan menggunakan metode K-Means, dimana jumlah *cluster* dapat ditentukan sesuai dengan permintaan pengguna. Semakin banyak jumlah *cluster* yang maka semakin sedikit anggota himpunan *cluster* yang terbentuk, semakin sedikit jumlah *cluster* yang dibentuk maka semakin banyak anggota himpunan *cluster*nya. Hasil anggota himpunan *cluster* yang terbentuk dapat menjadi dasar penyesuaian terhadap anggaran promosi yang tersedia sehingga dapat membantu dalam penentuan strategi promosi dimasa yang akan datang.

Dari hasil *cluster* yang telah diperoleh, diambil anggota himpunan *cluster* yang selama 4 tahun tersebut berada di posisi *cluster* tertinggi yang memiliki peminat terbanyak dan ditentukan pula anggota himpunan *cluster* yang memiliki peminat rendah dimana selama 4 tahun tersebut tidak berada di dalam *cluster* tertinggi sehingga dapat dijadikan landasan oleh para pimpinan untuk menentukan strategi promosi ditahun yang akan datang.

## REFERENSI

- [1] Andreson and Dkk, *Multivariate Data Analysis with Readings*, 4th ed. New Jersey: Prentice-Hall, 1995.
- [2] R. A. Johnson , *Applied Multivariate Statistical Analysis*, 4th ed. 1998.
- [3] B. D. Santosa, *Pemeliharaan Variable Dengan Linier SVIM untuk Kasus Multikelas Bidang Data Mining*, Dikti, 2007.
- [4] Kusumadewi Sri, *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Yogyakarta: Graha Ilmu, 2004.
- [5] Luthfi, Emha Taufik, *FUZZY C-MEANS UNTUK CLUSTERING DATA*, Seminar Nasional Teknologi 2007, ISSN: 1978-9777, (November 2007) D1-D7
- [6] MacQueen, "Some Methodes for Classification and Analysis of Multivariate Observations," *Proc. 5-th Berkeley Symp. Math. Stat. Probab.*, vol. 5, pp. 281–297, 1967.
- [7] B. Mahadi, "Non-Euclidean Norms dan Data Normalization," *Pengelompokan Data Kaji Cuaca Menggunakan K-Means Bagi Peramalan Taburan Hujan*, p. 3, 2001.