

## APLIKASI PREDIKSI WAKTU KELULUSAN MAHASISWA MENGUNAKAN ALGORITMA K-MEANS CLUSTERING

Silmi Nur Addini<sup>[1]</sup>, Dewanto Rosian Adhy<sup>[2]</sup>

<sup>[1]</sup>Sekolah Tinggi Teknologi Ybs Internasional

Email :

silminur@sttybsi.ac.id <sup>[1]</sup>, dewanto72@gmail.com <sup>[2]</sup>

### Abstrak

Aplikasi prediksi waktu kelulusan mahasiswa ini bertujuan memberikan sarana untuk melihat perkiraan waktu lulus mahasiswa dengan melihat mahasiswa-mahasiswa mana saja yang masuk kedalam suatu *cluster* tertentu berdasarkan parameter IPK dan kehadiran. Aplikasi ini menerapkan algoritma *k-Means*.

Algoritma *K-Means* merupakan model *cendroid*. Model *cendroid* adalah model yang menggunakan *cendroid* untuk membuat *cluster*. *K-Means* merupakan salah satu metode *clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*. Metode ini mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda di kelompokkan ke dalam *cluster* yang lain.

Aplikasi ini diharapkan dapat membantu pihak kampus maupun mahasiswa untuk memprediksi tingkat kelulusan tepat waktu dan untuk meningkatkan reputasi bagi pihak kampus itu sendiri dan kelulusan tepat waktu bagi mahasiswa supaya kelulusan mereka tidak terlambat, selain itu pihak kampus bisa melakukan hal-hal yang perlu dilakukan apabila mereka di prediksi lulus tidak tepat waktu seperti dengan melakukan bimbingan dan hal lainnya.

Kata kunci : Aplikasi prediksi, Algoritma *K-Means*

### Abstract

*This student graduation time application aims to provide a means to see when students graduate by seeing which students enter a particular cluster based on GPA and attendance parameters. This application implements the k-Means algorithm.*

*K-Means algorithm is a cendroid model. Cendroid model is a model that uses cendroid to create clusters. K-Means is a non-hierarchical clustering method that tries to partition existing data in the form of one or more clusters. This method partitions data into clusters so that data with the same characteristics are in the same cluster and data with different characteristics are grouped into other clusters.*

*This application is expected to help campuses and students predict the graduation rate on time and can improve the reputation of the campus itself and timely graduation for students so that their graduation is not late, besides that the campus can do things that need to be done when predictions pass. not on time like when doing guidance and other things.*

*Keywords: Prediction Application, K-Means Algorithm*

## I. Pendahuluan

### 1.1. Latar belakang

Kelulusan mahasiswa adalah satu *moment* yang sangat dinanti-nanti oleh setiap mahasiswa dan lulus tepat pada waktunya adalah suatu hal yang diharapkan. Selain itu setiap kampus juga pasti menginginkan mahasiswanya

untuk lulus tepat waktu, karena hal ini bermanfaat bagi pihak kampus untuk meningkatkan reputasi dan akreditasinya. Tapi pada kenyataannya para mahasiswa tidak bisa memprediksinya, mereka hanya mengikuti alur perkuliahan, sehingga ujung-ujungnya waktu kelulusan mereka menjadi terlambat. hal ini

terlihat dari tidak seimbangnya presentase kelulusan mahasiswa dari tahun ke tahun ataupun perbandingan jumlah yang sudah lulus dan belum lulus pada beberapa angkatan.

Salah satu metode yang digunakan untuk memprediksi waktu kelulusan mahasiswa ini adalah dengan menggunakan metode *K-Means Clustering*, yaitu dengan mengelompokkan n buah objek kedalam k kelas berdasarkan jaraknya dengan pusat *cluster* [1]. Analisis *cluster* merupakan teknik multivariat yang mempunyai tujuan utama untuk mengelompokkan objek - objek berdasarkan karakteristik yang dimilikinya. Analisis *cluster* mengklasifikasi objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam *cluster* yang sama. Analisis *cluster* merupakan salah satu alat analisis yang berguna sebagai peringkas data. Dalam meringkas data ini dapat dilakukan dengan jalan mengelompokkan objek-objek berdasarkan kesamaan karakteristik tertentu diantara objek-objek yang hendak diteliti.

Diharapkan dengan adanya prediksi waktu kelulusan mahasiswa ini, apabila mereka diprediksi lulus tidak tepat pada waktunya, pihak kampus dapat melakukan kegiatan-kegiatan yang dapat meminimalisir jumlah mahasiswa yang diprediksi lulus tidak tepat waktu, seperti dengan mengadakan bimbingan belajar, mengadakan semester pendek ataupun kegiatan-kegiatan bermanfaat yang dapat membantu meningkatkan prestasi belajar mahasiswa.

Berdasarkan latar belakang diatas, maka penulis tertarik untuk mengangkat permasalahan ini untuk dilakukan penelitian dengan judul “**Aplikasi Prediksi Waktu Kelulusan Mahasiswa Menggunakan Algoritma K-Means Clustering**”.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalahnya adalah :

1. Apa cara yang dilakukan dalam memprediksi kelulusan mahasiswa ?
2. Bagaimana proses Algoritma *K-Means* didalam mengelompokkan data mahasiswa ?

3. Bagaimana kemampuan Algoritma *K-Means* dalam memprediksi kelulusan mahasiswa ?

## 1.3. Tujuan Penelitian

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk :

1. Mengklasterisasi mahasiswa dengan menggunakan algoritma *k-Means*.
2. Mengklaster data-data mahasiswa berdasarkan parameter yang ada yaitu IPK dan kehadiran.
3. Melihat sejauh mana kemampuan algoritma *K-Means Clustering* dalam memprediksi waktu kelulusan mahasiswa serta memberikan sarana untuk melihat perkiraan waktu lulus.

## 2. Kajian Pustaka

### 2.1 Algoritma K-Means

Pada artikel yang diterbitkan Elsevier “*Data clustering: 50 years beyond K-means*” [2] diungkapkan bahwa pengorganisasian data ke dalam suatu kluster merupakan suatu model yang paling mendasar untuk pemahaman dan pembelajaran. Analisis kluster adalah studi formal untuk mengelompokkan, atau clustering benda-benda sesuai dengan karakteristik yang diukur berdasarkan kemiripan satu sama lain. *Clustering* adalah pengelompokan menggunakan teknik *unsupervised learning* dimana tidak diperlukan pelatihan pada metode tersebut atau dengan kata lain, tidak ada fase *learning* serta tidak menggunakan pelabelan pada setiap kelompok.

### 2.2 Clustering

Metode *clustering* mempartisi data ke dalam kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama [3]. Tujuan dari *clustering* ini adalah untuk meminimalisasi fungsi tujuan yang ditetapkan dalam proses *clustering*, yang umumnya berusaha meminimalisasi variasi dalam suatu *cluster* dan memaksimalkan variasi antar *cluster*.

*Clustering* atau analisis *cluster* adalah:

- a. Proses pembentukan kelompok data (*cluster*) dari himpunan data yang tidak

diketahui kelompok-kelompok atau kelas-kelasnya.

- b. Proses menentukan data-data termasuk ke dalam *cluster* yang mana.

Bisa dibilang, *clustering* merupakan proses untuk mengetahui kelas-kelas taksonomi atau batryologi, atau analisis topologi dari data-data yang ada. Dilihat dari kaca mata data mining, *clustering* bukanlah proses klarifikasi. Karena dalam proses klarifikasi, data dikelompokkan ke dalam kelas-kelas yang telah diketahui sebelumnya.

### 3. Metode Penelitian

#### 3.1. Analisis Kebutuhan

Pada analisa kebutuhan ini membahas beberapa kebutuhan atau persyaratan yang terkait dalam penganalisisan prediksi kelulusan mahasiswa. Berdasarkan data yang diperoleh hasil analisa kebutuhan tersebut adalah *system* yang akan dibangun membutuhkan input diantaranya NIM, nama, jenis kelamin, jurusan, IPK dan kehadiran sebagai data akademik atau data mahasiswa. Dan untuk kebutuhan proses, proses yang digunakan untuk mengolah data input adalah teknik *clustering* menggunakan algoritma *k-means*. Yang terakhir adalah kebutuhan *output* yang diharapkan adalah analisa dari hasil *clustering* sehingga dapat diketahui informasi yang dibutuhkan. Berikut adalah istilah-istilah didalam algoritma *k-means clustering* :

1. *Cluster* : *Cluster* adalah kelompok atau grup
2. *Cendroid* : *Cendroid* adalah titik tengah atau pusat
3. Iterasi : Iterasi adalah pengulangan

#### 3.2. Analisis Kebutuhan

Pada tahap ini dilakukan analisis data dari hasil pengclustoran menggunakan algoritma *k-means*. Data-data yang dibutuhkan untuk membangun aplikasi ini berupa parameter yang kita ambil sebagai acuan prediksi kelulusan mahasiswa yaitu IPK dan kehadiran. Dari dua parameter tersebut data diubah menjadi tipe data yang memudahkan untuk diproses. Tingkat kelulusan diukur dari IPK dan kehadiran. IPK dikategorikan berdasarkan predikat kelulusan

yang diatur dalam peraturan akademik BAB IV pasal 19 ayat 1 yang berbunyi “ predikat kelulusan program sarjana dan program diploma adalah sebagai berikut :”

- a. IPK memuaskan dengan IPK 2.00 – 2.75
- b. IPK sangat memuaskan dengan IPK 2.76 – 3.35
- c. IPK dengan pujian ( *cumlaude* ) dengan IPK 3.51 – 4.00

Berikut adalah pengkategorian data kelulusan berdasarkan kehadiran :

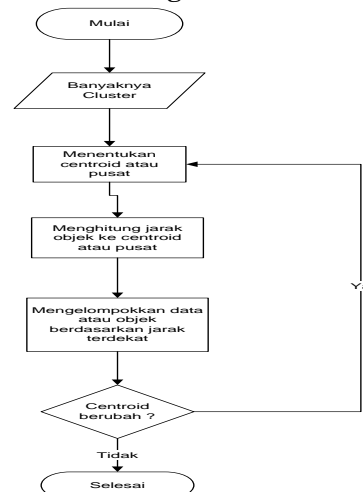
Tabel 1. Kategori Kehadiran

Keterangan Nilai Kehadiran		
86-100 %	A (5)	Rajin Sekali
71-85 %	B (4)	Rajin
56-70 %	C (3)	Cukup Rajin
41-55 %	D (2)	Malas
<40 %	E (1)	Sangat Malas

#### 3.3. Perancangan Sistem

Dalam bab ini, rancangan dituangkan dalam sebuah diagram alir dari algoritma *k-means*. Dalam perancangan ini dapat memberikan gambaran tentang sistem yang dibuat.

##### 3.3.1. Flowchart Algoritma K-Means



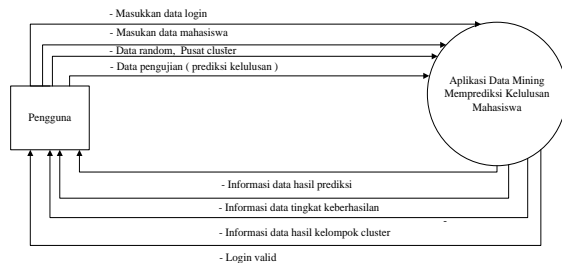
Gambar 1 Flowchart Algoritma Metode K-Means

##### 3.3.2. Diagram Konteks

Diagram konteks adalah suatu diagram yang menggambarkan keseluruhan system. Diagram ini menggambarkan masukan dan keluaran dari sebuah system yang berasal dari

dan untuk entitas yang terlihat dalam sebuah system.

Gambar berikut ini adalah diagram konteks dari aplikasi data mining dalam memprediksi kelulusan mahasiswa menggunakan *algoritma k-means clustering*.



Gambar 2 Diagram Konteks

### 3.4. Langkah-Langkah Yang Dilakukan Oleh Algoritma Metode K-Means

#### 1. Menentukan Jumlah Cluster

Dari data yang diperoleh, dapat ditentukan bahwa 100 objek tersebut memiliki 2 atribut (IPK dan kehadiran), dimana tiap-tiap mahasiswa mewakili 1 titik dengan 2 atribut (X, Y).

#### 2. Pengesetan Nilai Awal Titik Tengah.

Untuk menentukan nilai awal *centroid* dilakukan secara acak. Untuk penentuan awal diasumsikan :

1. Sebagai pusat *cluster* pertama ( C1 ) : 3.13 ; 5
2. Sebagai pusat *cluster* kedua ( C2 ) : 3.18 ; 4
3. Sebagai pusat *cluster* ketiga ( C3 ) : 3.38 ; 3
4. Sebagai pusat *cluster* keempat ( C4 ) : 2.92 ; 2

#### 3. Menghitung Jarak Obyek Ke Centroid Dengan Menggunakan Rumus Jarak Euclid.

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Dimana :

$D_e$  adalah Enclidean Distance  
 $i$  adalah banyaknya objek  
 ( x, y ) merupakan koordinat objek  
 ( s, t ) merupakan koordinat *centroid*

Perhitungan jarak dari data ke 1 terhadap pusat cluster

$$C1 = D_e = \sqrt{(3.41 - 3.13)^2 + (3 - 5)^2} = 2.0195048898183$$

$$C2 = D_e = \sqrt{(3.41 - 3.18)^2 + (3 - 4)^2} = 1.0261091559868$$

$$C3 = D_e = \sqrt{(3.41 - 3.38)^2 + (3 - 3)^2} = 0.03$$

$$C4 = D_e = \sqrt{(3.41 - 2.92)^2 + (3 - 2)^2} = 1.1135977729863$$

Perhitungan jarak dari data ke 2 terhadap pusat cluster

$$C1 = D_e = \sqrt{(2.80 - 3.13)^2 + (4 - 5)^2} = 1.05304320899$$

$$C2 = D_e = \sqrt{(2.80 - 3.18)^2 + (4 - 4)^2} = 0.38$$

$$C3 = D_e = \sqrt{(2.80 - 3.38)^2 + (4 - 3)^2} = 1.1560276813295$$

$$C4 = D_e = \sqrt{(2.80 - 2.92)^2 + (4 - 2)^2} = 2.0035967658189$$

Perhitungan jarak dari data ke 3 terhadap pusat cluster

$$C1 = D_e = \sqrt{(3.06 - 3.13)^2 + (2 - 5)^2} = 3.0008165555395$$

$$C2 = D_e = \sqrt{(3.06 - 3.18)^2 + (2 - 4)^2} = 2.0035967658189$$

$$C3 = D_e = \sqrt{(3.06 - 3.38)^2 + (2 - 3)^2} = 1.0499523798725$$

$$C4 = D_e = \sqrt{(3.06 - 2.92)^2 + (2 - 2)^2} = 0.14$$

Perhitungan seterusnya sampai jarak dari data ke 100 terhadap pusat cluster.

#### 4. Langkah selanjutnya menghitung pusat cluster baru

Pusat *cluster* yang baru ditentukan berdasarkan pengelompokan anggota masing-masing *cluster*. Berdasarkan tabel 3.4, C1 untuk parameter IPK dan kehadiran mahasiswa memiliki 13 anggota, maka perhitungan *cluster* menjadi :

$$C1 = \frac{3.13+2.81+3.22+3.32+3.3+3.33+3.38+\dots}{5+5+5+5+5+5+5+5+5+5+5+5+5} = (3.1907692307692, 5)$$

Kemudian, karena *cluster* ke 2 mempunyai 32 anggota maka *cluster* baru menjadi :

$$C2 = \frac{1+3.04+3.25+3.04+3.16+3.03+3.06+3.08+3.09+\dots}{\frac{4+4+4+4+4+4+4+4+4+4+4+4+4+4+4+\dots}{32}}$$

C2 = (3.11125, 4)

Kemudian, karena *cluster* ke 3 mempunyai 17 anggota maka *cluster* baru menjadi :

$$C3 = \frac{3.41+3.38+2.83+2.81+3.01+3.13+3.2+3.13+3.37+2.87+3.21+3.19+3.1+3.1+2.99+2.92+2.79+2.76+2.99+3.35+2.96+3.19+2.87+3.05+3.35+3.02+3.03+\dots}{\frac{3+3+3+3+3+3+3+3+3+3+3+3+3+3+3+\dots}{31}}$$

C3 = (3.0874193548387, 3)

Kemudian, karena *cluster* ke 3 mempunyai 6 anggota maka *cluster* baru menjadi :

$$C4 = \frac{3.06+4+3.15+3.17+2.84+3.3+2.66+2.61+2.53+3.03+\dots}{\frac{2+2+2+2+2+2+2+2+2+2+2+2+2+2+\dots}{24}}$$

C4 = (3.04875, 2)

Tabel 2 Pusat *cluster* pada iterasi ke 1

Cluster ke	IPK	Kehadiran
C1	3.1907692307692	5
C2	3.11125	4
C3	3.0874193548387	3
C4	3.04875	2

**5. Pengulangan langkah ke 2 hingga posisi data tidak mengalami perubahan.**

Perhitungan jarak dari data ke 1 terhadap pusat cluster, dilakukan seperti perhitungan jarak pada iterasi 1 dengan pusat cluster ipk dan kehadiran sesuai dengan hasil pusat cluster pada iterasi 1. Dan hasil dari penghitungan pusat *cluster* pada iterasi ke 2 adalah :

Tabel 3 Pusat *cluster* pada iterasi ke 2

Cluster ke	IPK	Kehadiran
C1	3.1907692307692	5
C2	3.11125	4
C3	3.0874193548387	3

C4	3.04875	2
----	---------	---

Karena G1 = G2 dimana anggota yang sama, maka tidak perlu dilakukan iterasi lagi. Dan sampai disini hasil *clustering* sudah mencapai stabil atau konvergen. Dari hasil perhitungan didapatkan 2 hasil *cluster* yaitu :

1. *Cluster* 1 memiliki pusat *cluster* (3.1907692307692 ; 5 ) dengan anggota mahasiswa ke 9, 21, 27, 36, 42, 46, 73, 75, 77, 80, 82, 94, 99
2. *Cluster* 2 memiliki pusat *cluster* (3.11125 ; 4 ) dengan anggota mahasiswa ke 2, 6, 7, 13, 16, 18, 23, 25, 33, 37, 41, 43, 45, 54, 57, 59, 60, 61, 62, 65, 66, 67, 68, 70, 72, 78, 81, 90, 91, 93, 95, 97
3. *Cluster* 3 memiliki pusat *cluster* ( 3.0874193548387 ; 3 ) dengan anggota mahasiswa 1, 8, 12, 14, 15, 20, 22, 26, 30, 32, 35, 38, 40, 44, 48, 51, 52, 53, 56, 58, 69, 71, 74, 79, 85, 86, 87, 89, 92, 96, 100
4. *Cluster* 4 memiliki pusat *cluster* ( 3.04875 ; 2 ) dengan anggota mahasiswa 3, 4, 5, 10, 11, 17, 19, 24, 28, 29, 31, 34, 39, 47, 49, 50, 55, 63, 64, 76, 83, 84, 88, 98

**4. Pembahasan**

**4.1. Implementasi**

Setelah system dianalisis dan didesain, maka selanjutnya adalah menuju tahap implementasi. Implementasi merupakan tahap meletakkan system sehingga siap untuk dioperasikan. Implementasi bertujuan untuk mengkonfrmasi modul-modul perancangan, sehingga pengguna dapat memberikan masukan kepada pembangun program aplikasi.

**4.2. Pengujian**

Setelah dilakukan perancangan dan implementasi ke, selanjutnya akan dilakukan simulasi untuk pengujian aplikasi dengan tujuan memperoleh perbandingan antara hasil peragaan dengan hasil yang diharapkan. Dan metode yang digunakan dalam pengujian ini adalah menggunakan metode pengujian *black box*. Metode ujicoba *blackbox* memfokuskan pada keperluan fungsional dari software. Karna itu ujicoba *blackbox* memungkinkan pengembang

*software* untuk membuat himpunan kondisi input yang akan melatih

5. seluruh syarat-syarat fungsional suatu program, uji coba *black box* merupakan pendekatan yang melengkapi untuk menemukan kesalahan lainnya.

#### 4.2.1 Rencana Pengujian

Rencana pengujian dapat dilihat pada tabel berikut ini.

Tabel 4. Rencana Pengujian Aplikasi Prediksi Kelulusan Mahasiswa

Item Uji	Detail pengujian	Jenis Uji
Login	Verifikasi data Login	<i>Black Box</i>
Pengolahan Data Mahasiswa	Tambah data mahasiswa Edit data mahasiswa Hapus data mahasiswa Simpan data mahasiswa	<i>Black Box</i>
Pengolahan Iterasi Awal	Proses iterasi	<i>Black Box</i>
Pengolahan Kecocokan Data	Proses untuk mengetahui tingkat kecocokan iterasi	<i>Black Box</i>
Pengolahan Hasil Prediksi	Simpan data prediksi untuk menyimpan hasil prediksi	<i>Black Box</i>

## 5. Simpulan

### 5.1. Kesimpulan

Setelah dilakukan pengujian dan analisa program, maka dapat diperoleh keimpulan sebagai berikut :

1. Pada kasus ini implementasi algoritma *k-means* dalam data mining sudah berhasil, dan bisa menampilkan informasi prediksi kelulusan mahasiswa, namun untuk melihat tingkat kemampuan *real k-means clustering*

dalam memprediksi waktu kelulusan tergantung pada mahasiswa itu sendiri

2. Program Aplikasi ini bisa dijadikan sarana untuk melihat perkiraan waktu lulus berdasarkan parameter yang digunakan walaupun dengan kesederhanaannya.
3. Program yang dibuat berhasil mengkluster atau mengelompokkan mahasiswa-mahasiswa yang mempunyai karakteristik yang sama dikelompokkan kedalam suatu *cluster* yang sama berdasarkan parameter yang digunakan.

### 5.2 Saran

Dari beberapa kesimpulan yang telah diambil, maka dapat dikemukakan saran-saran yang akan sangat membantu untuk pengembangan aplikasi ini selanjutnya

Untuk para peneliti yang akan melakukan penelitian serupa dengan ini, akan sangat menarik dan lebih baik apabila parameter yang digunakan ditambah lebih banyak lagi guna mendapatkan hasil yang lebih optimal dan akurat.

Untuk mengetahui pengembangan analisis dan aplikasi data mining yang lainnya serta untuk mendapat keputusan yang beragam, kita juga bisa menggunakan algoritma lain dalam data mining didalam melakukan penelitian seperti *C4.5*, Algoritma Apriori, *Naive Bayes* dan algoritma lainnya untuk mengetahui perbedaan dan cara kerja dari setiap masing-masing Algoritma dalam data mining tersebut.

### Daftar Pustaka

#### DAFTAR PUSTAKA

- [1] Al Fatta, Hanif. 2007. Analisis & Perancangan Sistem Informasi Untuk Keunggulan Bersaing Perusahaan & Organisasi Modern. Andi: Yogyakarta.
- [2] Herjanto, Eddy. Manajemen Operasi. Edisi Ketiga. Gramedia Widiasarana Indonesia. Jakarta
- [3] Jogiyanto. 2002. Analisis dan Desain Sistem Informasi. ANDI: Yogyakarta.

- [4] Larose, Daniel T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*.
- [5] Mcleod, Raymond. 2008. *Sistem Informasi Manajemen*. Edisi 10. Salemba Empat: Jakarta
- [6] Nursal. 2006. *Buku Panduan Visual Basic 1*. Dinamika Ilmu: Jakarta.
- [7] Pressman, Roger S. 2001. *Software Engineering A Practitioner's Approach*. McGraw – Hill: New York.
- [8] Santosa, Budi. 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu: Yogyakarta.
- [9] Widianti, Sri. 2009. *Pengantar Basis Data*. Lentera Ilmu Cendikia: Jakarta.
- [10] Wu, Xindong. 2009. *Algorithm In Data Mining*. Crc Pers: Francis.
- [11] Jain. 2009. *Data clustering: 50 years beyond K-means*
- [12] Refaat, M. 2007. *Data Preparation For Data Mining Using Sas*. Diane D Cerra. San Francisco.
- [13] Rismawan, Tedy dan Kusumadewi, Sri. 2008, *Aplikasi K-Means Untuk Pengelompokan Mahasiswa Berdasarkan Nilai Body Mass Index (BMI) dan Ukuran Kerangka*: Yogyakarta.
- [14] Syafrianto, Andri. *Perancangan Aplikasi K-Means Untuk Pengelompokan Mahasiswa Stmik Elrahma Yogyakarta Berdasarkan Frekuensi Kunjungan Ke Perpustakaan Dan Ipk*: Yogyakarta.
- [15] Velmurugan. 2010. *Performance Evaluation of K-means and Fuzzy C-Means Clustering Algorithms for Statistical Distributions of Input Data Points*