

# PENERAPAN METODE CLUSTERING DENGAN ALGORITMA K-MEANS UNTUK PREDIKSI KELULUSAN MAHASISWA JURUSAN TEKNIK INFORMATIKA DI UNIVERSITAS MUHAMMADIYAH JEMBER

<sup>1</sup> Alfiyani Rindyatul Jannah(1210651237),<sup>2</sup>Deni Arifianto,M.Kom  
Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember  
Email : [alfyanirindi11@gmail.com](mailto:alfyanirindi11@gmail.com)

## Abstrak

Kelulusan mahasiswa yang tepat waktu akan menguntungkan pihak mahasiswa dan perguruan tinggi. Ada beberapa faktor yang mempengaruhi prediksi kelulusan mahasiswa yang sesuai dengan waktu studi, diantaranya : rata-rata IPK terakhir, jumlah SKS, keaktifan di organisasi, beasiswa, dan asal daerah.(Kondo, Ferry.2015)

Clustering merupakan suatu metode untuk pengelompokan dokumen dimana dokumen dikelompokkan dengan konten untuk mengurangi ruang pencarian yang diperlukan dalam merespon suatu query (Natalius, 2010). Dalam penelitian ini menggunakan tiga kategori kelulusan yaitu lulus cepat, tepat, dan lambat.

Hasil perhitungan dataset dengan metode K-Means didapatkan hasil Accuracy 60%, hal ini menunjukkan bahwa sistem dapat mengklasifikasikan data secara benar, namun dalam pengelompokannya belum optimal, karena terdapat data yang terklasifikasi benar masuk ke klasifikasi salah dan data terklasifikasi salah masuk ke klasifikasi benar.

Kata Kunci : Kelulusan, Clustering, K-Means

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Perguruan tinggi merupakan lembaga pendidikan tertinggi yang semakin banyak peminatnya di Indonesia. Menurut Kementerian Pendidikan Dan Kebudayaan (2015), perkembangan data statistik Perguruan Tinggi menunjukkan bahwa jumlah lembaga Perguruan Tinggi, mahasiswa baru, mahasiswa lulusan, dan dosen menunjukkan perkembangan meningkat tiap tahunnya.

Kelulusan mahasiswa yang tepat waktu akan menguntungkan pihak mahasiswa dan perguruan tinggi. Untuk mahasiswa, semakin cepat lulus maka kesempatan untuk mengikuti berbagai seleksi dalam mencari pekerjaan semakin banyak. Sedangkan bagi perguruan tinggi, kelulusan mahasiswa dengan tepat waktu dapat memajukan kualitas, meningkatkan reputasi, dan juga berpengaruh pada akreditasi perguruan tinggi tersebut. (Premashanti, Gita. 2015) Pada penelitian ini, digunakan teknik data mining untuk

menemukan pola kelulusan mahasiswa yang sudah lulus, kemudian dijadikan dasar untuk memprediksi kelulusan mahasiswa pada tahun berikutnya.

Data mining adalah proses menemukan hubungan dalam data yang tidak diketahui oleh pengguna dan menyajikannya dengan cara yang dapat dipahami sehingga hubungan tersebut dapat menjadi dasar klasifikasi data. Untuk mengetahui tingkat kelulusan mahasiswa dalam satu tahun ajaran dapat dilakukan suatu prediksi berdasarkan data-data mahasiswa pada tingkat atau tahun ajaran pertama. Ada beberapa faktor yang mempengaruhi prediksi kelulusan mahasiswa yang sesuai dengan waktu studi, diantaranya : rata-rata IPK terakhir, jumlah SKS.(Kondo, Ferry.2015) Dalam menentukan prediksi kinerja mahasiswa dengan akurasi yang tinggi bermanfaat untuk mengidentifikasi siswa dengan prestasi akademik (Suhartinah dan Emastuti, 2010).

Untuk menyelesaikan permasalahan pada penjelasan diatas adalah dengan

pemanfaatan algoritma K-Means Clustering. Algoritma K-Means adalah algoritma klustering yang paling sederhana dibanding algoritma klustering yang lain. Algoritma ini mempunyai kelebihan mudah diterapkan dan dijalankan, relatif cepat, mudah untuk diadaptasi, dan paling banyak dipraktikkan dalam tugas data mining.

Clustering merupakan suatu metode untuk pengelompokan dokumen dimana dokumen dikelompokkan dengan konten untuk mengurangi ruang pencarian yang diperlukan dalam merespon suatu query (Natalius, 2010).

Algoritma K-Means merupakan algoritma yang membutuhkan parameter input sebanyak  $k$  dan membagi sekumpulan  $n$  objek ke dalam cluster sehingga tingkat kemiripan antar anggota dalam satu cluster tinggi sedangkan tingkat kemiripan dengan anggota pada cluster lain sangat rendah. Kemiripan anggota terhadap cluster diukur dengan kedekatan objek terhadap nilai mean pada cluster atau disebut sebagai centroid cluster (Rismawan dan Kusumadewi, 2008). Data nilai mahasiswa hanya akan menjadi sekumpulan data yang tidak berguna jika tidak dilakukan penggalian data terhadapnya. Banyak informasi terpendam yang dapat diambil dari sekumpulan data tersebut sehingga dapat memberikan suatu pengetahuan untuk penentuan kebijakan. Penggalian data dapat dilakukan dengan cara pengelompokan data nilai mahasiswa menjadi beberapa kelompok, kelompok nilai baik dan nilai buruk.

Karena itu, pada penelitian ini akan dikembangkan suatu sistem yang dapat mengklustering kelulusan mahasiswa menggunakan algoritma K-Means dengan melihat pola kelulusan mahasiswa beberapa periode sebelumnya.

## 1.2 Tujuan

Adapun manfaat dalam penelitian yaitu :

1. Mampu memprediksi sistem mengenai kelulusan mahasiswa Jurusan Teknik

Informatika di Universitas Muhammadiyah Jember, apakah bernilai lulus cepat, lulus tepat dan lulus terlambat.

2. Mengetahui tingkat akurasi algoritma K-Means dalam memprediksi ketepatan kelulusan mahasiswa di Universitas Muhammadiyah Jember.

## 2. TINJAUAN PUSTAKA

### 2.1 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban et al, 2005).

Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar (Larose, 2005). Tan et al, (2006) mendefinisikan data mining sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah data mining kadang disebut juga knowledge discovery.

### 2.2 Clustering

Clustering adalah suatu metode pengelompokan berdasarkan ukuran kedekatan. Perbedaan Clustering dengan grup, kalau grup berarti kelompok yang sama kondisinya kalau tidak ya pasti bukan kelompoknya. Tetapi kalau cluster tidak harus sama akan tetapi pengelompokannya berdasarkan kedekatan dari suatu

karakteristik sample yang ada, salah satunya dengan menggunakan rumus jarak euclidean. Aplikasi cluster ini sangat banyak, karena hampir banyak dalam mengidentifikasi permasalahan atau pengambilan keputusan selalu tidak sama persis akan tetapi cenderung memiliki kemiripan saja.

### 2.3 K-Means

K-Means merupakan algoritma untuk cluster objek berdasarkan atribut menjadi k partisi, dimana  $k < n$ . Secara Umum K-Means Clustering merupakan salah satu metode data Clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster atau kelompok.

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 dikelompokkan ke dalam cluster yang lain.

Istilah-istilah dalam K-Means :

1. N data : data set yang akan diolah sebanyak N data dimana N data tersebut terdiri dari atribut-atributnya
2. K centroid : Inisialisasi dari pusat cluster data adalah sebanyak K dimana pusat-pusat awal tersebut digunakan sebagai banyaknya kelas yang akan tercipta. Centroid didapatkan secara random dari N data set yang ada.
3. Euclidian Distance: merupakan jarak yang didapat dari perhitungan antara semua N data dengan K centroid dimana akan memperoleh tingkat kedekatan dengan kelas yang terdekat dengan populasi data tersebut.

Kelemahan K-Means :

1. Bila jumlah data tidak terlalu banyak, mudah untuk menentukan cluster awal.
2. Jumlah cluster, sebanyak K, harus ditentukan sebelum dilakukan perhitungan.

3. Tidak pernah mengetahui real cluster dengan menggunakan data yang sama, namun jika dimasukkan dengan cara yang berbeda mungkin dapat memproduksi cluster yang berbeda jika jumlah datanya sedikit.
4. Tidak tahu kontribusi dari atribut dalam proses pengelompokan karena dianggap bahwa setiap atribut memiliki bobot yang sama.

Langkah-langkah dalam Algoritma K-Means Clustering :

1. Menentukan jumlah cluster .
2. Menentukan nilai centroid. Dalam menentukan nilai centroid untuk awal iterasi, nilai awal centroid dilakukan secara acak. Sedangkan jika menentukan nilai centroid yang merupakan tahap dari iterasi, maka digunakan rumus sebagai berikut :
  - a. Menghitung jarak antara titik centroid dengan titik tiap objek
  - b. Pengelompokan objek untuk menentukan anggota cluster adalah dengan memperhitungkan jarak minimum objek.
  - c. Kembali ke tahap 2, lakukan perulangan hingga nilai centroid yang dihasilkan tetap dan anggota cluster tidak berpindah ke cluster lain.

## 3. METODOLOGI PENELITIAN

### 3.1 Studi Kasus

Berikut ini merupakan simulasi dari proses perhitungan K-Means terhadap data. Dalam perhitungan yang dilakukan, menggunakan data testing dan training, penggunaan data testing dan training bertujuan agar model yang diperoleh nantinya memiliki kemampuan generalisasi yang baik dalam melakukan klasifikasi data. Data training sebagai data sampel sedangkan data testing dipergunakan sebagai pembandingan dalam menghitung data training.

Diketahui dari 10 (sepuluh) data berikut terdapat 7(tujuh) data testing dan 3 (tiga) data training yang terdiri dari Lulus

Cepat (L.Cepat), Lulus Tepat (L.Tepat), dan Lulus Terlambat (L.Terlambat). Berikut ini adalah kesepuluh data tersebut :

Tabel 3.1 Data

NAMA	Observasi			Kategori Data
	IPK	SKS	Keterangan	
Person 1	3.4	144	L.CEPAT	Data Testing
Person 2	3.23	139	L.CEPAT	Data Training
Person 3	3.1	130	L.TEPAT	Data Training
Person 4	3.78	144	L.CEPAT	Data Training
Person 5	3.45	132	L.TEPAT	Data Testing
Person 6	2.71	120	L.LAMBAT	Data Training
Person 7	2.9	128	L.LAMBAT	Data Training
Person 8	3	110	L.LAMBAT	Data Testing
Person 9	3.12	131	L.CEPAT	Data Training
Person 10	3.34	142	L.CEPAT	Data Training

- Langkah pertama : Pilih *Centroid* secara acak  
Dari data yang ada sebelumnya, akan dipilih 2 ceteroid acak.

Tabel 3.2 *Centroid*

Person	Centroid	
	IPK	SKS
1	3.4	144
5	3.45	132
8	3	110

- Langkah kedua :  
Lakukan perhitungan jarak dengan *euclidean distance* dari masing-masing item dari *Centroid* (pusat) *cluster* dan tandai kembali setiap item berdasarkan kedekatan group. Jika item bergerak dari *initial configuration*, *Centroid* (pusat/*means*) *cluster* harus diupdate sebelum diproses. Kita hitung kwadrat jarak (*squared distance*) sbb:

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

Keterangan :

- Rumus *Euclidian distance* :  $d = |x - y| =$
- $x$  = pusat cluster
- $y$  = data

Tabel 3.3 Perhitungan Tahap 1

Obyek Data	Cluster 1	
Person 1	$\sqrt{(3.4 - 3.4)^2 + (144 - 144)^2}$	= 0.000
Person 2	$\sqrt{(3.23 - 3.4)^2 + (139 - 144)^2}$	= 5.003
Person 3	$\sqrt{(3.10 - 3.4)^2 + (130 - 144)^2}$	= 14.003
Person 4	$\sqrt{(3.78 - 3.4)^2 + (139 - 144)^2}$	= 0.380
Person 5	$\sqrt{(3.45 - 3.4)^2 + (132 - 144)^2}$	= 12.000
Person 6	$\sqrt{(2.71 - 3.4)^2 + (120 - 144)^2}$	= 24.010
Person 7	$\sqrt{(2.90 - 3.4)^2 + (128 - 144)^2}$	= 16.008
Person 8	$\sqrt{(3.0 - 3.4)^2 + (110 - 144)^2}$	= 34.002
Person 9	$\sqrt{(3.12 - 3.4)^2 + (131 - 144)^2}$	= 13.003
Person 10	$\sqrt{(3.34 - 3.4)^2 + (142 - 144)^2}$	= 2.001
Obyek Data	Cluster 2	
Person 1	$\sqrt{(3.4 - 3.45)^2 + (132 - 144)^2}$	= 12.000
Person 2	$\sqrt{(3.23 - 3.45)^2 + (139 - 132)^2}$	= 7.003
Person 3	$\sqrt{(3.10 - 3.45)^2 + (130 - 132)^2}$	= 2.030
Person 4	$\sqrt{(3.78 - 3.45)^2 + (139 - 132)^2}$	= 12.005
Person 5	$\sqrt{(3.45 - 3.45)^2 + (132 - 132)^2}$	= 0.000
Person 6	$\sqrt{(2.71 - 3.45)^2 + (120 - 132)^2}$	= 12.023
Person 7	$\sqrt{(2.90 - 3.45)^2 + (128 - 132)^2}$	= 4.038
Person 8	$\sqrt{(3.0 - 3.45)^2 + (110 - 132)^2}$	= 22.005
Person 9	$\sqrt{(3.12 - 3.45)^2 + (131 - 132)^2}$	= 1.053
Person 10	$\sqrt{(3.34 - 3.45)^2 + (142 - 132)^2}$	= 10.001
Obyek Data	Cluster 3	
Person 1	$\sqrt{(3.4 - 3.0)^2 + (144 - 110)^2}$	= 34.002
Person 2	$\sqrt{(3.23 - 3.0)^2 + (139 - 110)^2}$	= 29.001
Person 3	$\sqrt{(3.10 - 3.0)^2 + (130 - 110)^2}$	= 20.000
Person 4	$\sqrt{(3.78 - 3.0)^2 + (139 - 110)^2}$	= 34.009
Person 5	$\sqrt{(3.05 - 3.0)^2 + (132 - 110)^2}$	= 22.005
Person 6	$\sqrt{(2.71 - 3.0)^2 + (120 - 110)^2}$	= 10.004
Person 7	$\sqrt{(2.90 - 3.0)^2 + (128 - 110)^2}$	= 18.000
Person 8	$\sqrt{(3.0 - 3.0)^2 + (110 - 110)^2}$	= 0.000
Person 9	$\sqrt{(3.12 - 3.0)^2 + (131 - 110)^2}$	= 21.000
Person 10	$\sqrt{(3.34 - 3.0)^2 + (142 - 110)^2}$	= 32.002

Dari ketiga *cluster* tersebut kemudian dibandingkan untuk memperoleh *ceteroid* baru,

Tabel 3.4 Perhitungan Tahap 2

Obyek Data	Cluster 1	Cluster 2	Cluster 3	Kelas C 1	Kelas C 2	Kelas C
Person 1	0.000	12.000	34.002	1	2	3
Person 2	5.003	7.003	29.001	1	2	3
Person 3	14.003	2.030	20.000	3	1	2
Person 4	0.380	12.005	34.009	1	2	3
Person 5	12.000	0.000	22.005	3	1	2
Person 6	24.010	12.023	10.004	3	2	1
Person 7	16.008	4.038	18.000	3	1	3
Person 8	34.002	22.005	0.000	3	2	1
Person 9	13.003	1.053	21.000	3	1	2
Person 10	2.001	10.001	32.002	1	2	3
Jumlah =				4	6	5

Dari perbandingan diatas didapat pusat *cluster* baru yaitu :

Tabel 3.5 Pusat *Cluster* Baru

<i>Centroid</i> 1	3.44	142.25
<i>Centroid</i> 2	3.14	130.25
<i>Centroid</i> 3	2.86	115.00

Maka koordinat dari pusat *cluster* ter-update sebagai berikut :

Tabel 3.6 Hasil cluster

Obyek Data	Cluster 1
Person 1	$\sqrt{(3.4 - 3.44)^2 + (144 - 142.25)^2} = 1.750$
Person 2	$\sqrt{(3.23 - 3.44)^2 + (139 - 142.25)^2} = 3.257$
Person 3	$\sqrt{(3.10 - 3.44)^2 + (130 - 142.25)^2} = 12.255$
Person 4	$\sqrt{(3.78 - 3.44)^2 + (139 - 142.25)^2} = 1.783$
Person 5	$\sqrt{(3.85 - 3.44)^2 + (132 - 142.25)^2} = 10.250$
Person 6	$\sqrt{(2.71 - 3.44)^2 + (120 - 142.25)^2} = 22.262$
Person 7	$\sqrt{(2.90 - 3.44)^2 + (128 - 142.25)^2} = 14.260$
Person 8	$\sqrt{(3.0 - 3.44)^2 + (110 - 142.25)^2} = 32.253$
Person 9	$\sqrt{(3.12 - 3.44)^2 + (131 - 142.25)^2} = 11.254$
Person 10	$\sqrt{(3.34 - 3.44)^2 + (142 - 142.25)^2} = 0.268$
Obyek Data	Cluster 2
Person 1	$\sqrt{(3.4 - 3.14)^2 + (144 - 130.25)^2} = 13.752$
Person 2	$\sqrt{(3.23 - 3.14)^2 + (139 - 130.25)^2} = 8.750$
Person 3	$\sqrt{(3.10 - 3.14)^2 + (130 - 130.25)^2} = 0.254$
Person 4	$\sqrt{(3.78 - 3.14)^2 + (139 - 130.25)^2} = 13.765$
Person 5	$\sqrt{(3.85 - 3.14)^2 + (132 - 130.25)^2} = 1.777$
Person 6	$\sqrt{(2.71 - 3.14)^2 + (120 - 130.25)^2} = 10.259$
Person 7	$\sqrt{(2.90 - 3.14)^2 + (128 - 130.25)^2} = 2.263$
Person 8	$\sqrt{(3.0 - 3.14)^2 + (110 - 130.25)^2} = 20.251$
Person 9	$\sqrt{(3.12 - 3.14)^2 + (131 - 130.25)^2} = 0.750$
Person 10	$\sqrt{(3.34 - 3.14)^2 + (142 - 130.25)^2} = 11.752$
Obyek Data	Cluster 3
Person 1	$\sqrt{(3.4 - 2.86)^2 + (144 - 115.00)^2} = 29.005$
Person 2	$\sqrt{(3.23 - 2.86)^2 + (139 - 115.00)^2} = 24.003$
Person 3	$\sqrt{(3.10 - 2.86)^2 + (130 - 115.00)^2} = 15.002$
Person 4	$\sqrt{(3.78 - 2.86)^2 + (139 - 115.00)^2} = 29.015$
Person 5	$\sqrt{(3.85 - 2.86)^2 + (132 - 115.00)^2} = 17.010$
Person 6	$\sqrt{(2.71 - 2.86)^2 + (120 - 115.00)^2} = 5.002$
Person 7	$\sqrt{(2.90 - 2.86)^2 + (128 - 115.00)^2} = 13.000$
Person 8	$\sqrt{(3.0 - 2.86)^2 + (110 - 115.00)^2} = 5.002$
Person 9	$\sqrt{(3.12 - 2.86)^2 + (131 - 115.00)^2} = 16.002$
Person 10	$\sqrt{(3.34 - 2.86)^2 + (142 - 115.00)^2} = 27.004$

Dari ketiga *cluster* tersebut kemudian dibandingkan untuk memperoleh *ceteroid* baru,

Tabel 3.7 Hasil Cluster

Obyek Data	Cluster 1	Cluster 2	Cluster 3	Kelas C 1	Kelas C 2	Kelas C 3
Person 1	1.750	13.752	29.005	1	3	2
Person 2	3.257	8.750	24.003	1	2	3
Person 3	12.255	0.254	15.002	3	1	2
Person 4	1.783	13.765	29.015	1	2	3
Person 5	10.250	1.777	17.010	3	1	2
Person 6	22.262	10.259	5.002	3	2	1
Person 7	14.260	2.263	13.000	2	1	3
Person 8	32.253	20.251	5.002	3	2	1
Person 9	11.254	0.750	16.002	3	1	2
Person 10	0.268	11.752	27.004	1	2	3
		Jumlah =		4	4	2

Karena *ceteroid* pada iterasi 2 memiliki nilai yang sama dengan *threshold*, maka proses tidak dilanjut (proses berhenti). Pada penelitian ini, *threshold* (batasan/tolak ukur) yang digunakan adalah sebagai berikut :

Tabel 3.8 Centeroid

	IPK	SKS	Threshold Ceteroid
Cetroid 1	3.44	142.25	
Cetroid 2	3.14	130.25	
Cetroid 3	2.86	115.00	

Sehingga dapat dikelompokkan sebagai berikut berdasarkan data *training* yang telah ada adalah sebagai berikut :

Tabel 3.9 Hasil Klasifikasi Berdasarkan Data Training

Obyek Data	Cluster 1	Cluster 2	Cluster 3	Kelas C 1	Kelas C 2	Kelas C 3
Person 1	1.750	13.752	29.005	1		
Person 2	3.257	8.750	24.003	1		
Person 3	12.255	0.254	15.002		1	
Person 4	1.783	13.765	29.015	1		
Person 5	10.250	1.777	17.010		1	
Person 6	22.262	10.259	5.002			1
Person 7	14.260	2.263	13.000		1	
Person 8	32.253	20.251	5.002			1
Person 9	11.254	0.750	16.002		1	
Person 10	0.268	11.752	27.004	1		

Dari tael 3.9 diatas, dapat dikelompokkan sebagai berikut, sesuai dengan data *training* yang ada. Dimana nilai 1 adalah hasil dari klasifikasi yang telah dihasilkan, sehingga didapat data *training* yang termasuk pada lulus cepat terdapat 4 data (1 diantaranya adalah data *testing*), lulus tepat terdapat 4 data (1 diantaranya adalah data *testing*), dan lulus terlambat terdapat 2 data (1 diantaranya adalah data *testing*). Berikut ini adalah hasil klasifikasi tersebut :

Tabel 3.10 Hasil Perhitungan Data Testing

Obyek Data	Cluster 1	Cluster 2	Cluster 3	Keterangan
Person 1	1.750	13.752	29.005	L.CEPAT
Person 5	10.250	1.777	17.010	L.CEPAT
Person 8	32.253	20.251	5.002	L.LAMBAT.

Tabel 3.11 Hasil Perhitungan Data Training

Obyek Data	Cluster 1	Cluster 2	Cluster 3	Keterangan
Person 2	3.257	8.750	24.003	L.CEPAT
Person 3	12.255	0.254	15.002	L.TEPAT
Person 4	1.783	13.765	29.015	L.CEPAT
Person 6	22.262	10.259	5.002	L.LAMBAT
Person 7	14.260	2.263	13.000	L.TEPAT
Person 9	11.254	0.750	16.002	L.TEPAT
Person 10	0.268	11.752	27.004	L.CEPAT

### 3.1 Akurasi

Akurasi menunjukkan kedekatan nilai hasil pengukuran dengan nilai sebenarnya. Untuk menentukan tingkat akurasi perlu diketahui nilai sebenarnya dari besaran yang diukur dan kemudian dapat diketahui seberapa besar tingkat akurasinya. Tingkat akurasi berikut ini, diukur berdasarkan

pengelompokkan data testing yang ada dengan data yang sudah ada (data sebelum dilakukan perhitungan), sehingga didapat berikut :

Tabel 3.12 Perbandingan Hasil Data

	Observasi				Keterangan
	IPK	SKS	Data Berdasarkan Data Testing	Data Sebelum dilakukan Penitlungan	
Person 2	3.23	139	L.CEPAT	L.CEPAT	TERKLASIFIKAN DENGAN BENAI
Person 3	3.1	130	L.TEPAT	L.TEPAT	TERKLASIFIKAN DENGAN BENAI
Person 4	3.78	144	L.CEPAT	L.CEPAT	TERKLASIFIKAN DENGAN BENAI
Person 6	2.71	120	L.LAMBAT	L.LAMBAT	TERKLASIFIKAN DENGAN BENAI
Person 7	2.9	128	L.TEPAT	L.LAMBAT	TIDAK TERKLASIFIKAN DENGAN BENAR
Person 9	3.12	131	L.TEPAT	L.CEPAT	TIDAK TERKLASIFIKAN DENGAN BENAR
Person 10	3.34	142	L.CEPAT	L.CEPAT	TERKLASIFIKAN DENGAN BENAI

Tabel 3.12 Perhitungan Nilai Accuracy

TP (True Positif)	FP (False Positif)
<b>5</b>	<b>2</b>
FN (False Negatif)	TN (True Negatif)
<b>0</b>	<b>0</b>

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \\
 &= \frac{5+0}{5+0+2+0} \\
 &= 0,71428 * 100 \% \\
 &= 71.428 \%.
 \end{aligned}$$

Dari perhitungan akurasi di dapatkan nilai TP yaitu data benar yang terklasifikasikan menjadi data benar sebesar 5, nilai FP yaitu data benar terklasifikasi salah sebesar 2, nilai FN data salah tidak di klasifikasikan salah sebesar 0 dan nilai TN adalah data yang tidak di klasifikasikan benar sebesar 0, sehingga menghasilkan nilai Accuracy sebesar 71.428 %.

### 3.2 Analisa

Berdasarkan hasil pembahasan tentang penerapan algoritma *K-Means* dapat mengelompokkan mahasiswa yang mendapat nilai rata-rata yang bagus dan yang kurang bagus. Namun, dari nilai yang

telah dihasilkan tidak menuntut kemungkinan bahwa mahasiswa dengan kategori lulus terlambat dapat menjadi mahasiswa lulus cepat, dan lulus tepat dapat menjadi mahasiswa lulus cepat. Sehingga keputusan akhir dari system ini hanyalah untuk membantu dalam pemilihan alternatif keputusan, sehingga dapat memberikan keputusan secara tepat. Dan keputusan akhir tetaplah berdasarkan pertimbangan pihak akademik jurusan Teknik Informatika Universitas Muhammadiyah Jember.

### 4.1 Data Pengujian

Data yang diolah pada tugas akhir ini adalah berupa data mahasiswa Jurusan Teknik Informatika angkatan 2011 dengan jumlah data 266 data.

Tabel 4.1 Data Mahasiswa

mhswid	mhswid	nama	ip	sks
1	1110651001	WAHYU FARABI FIRMANSAH	3.63	130
2	1110651002	ILMI AMALYAH RIZKY	3.33	110
3	1110651003	ZHENI THRIAN	0.81	75
4	1110651004	SYARIFAH FAJARWATI	3.48	125
5	1110651005	SUGENG HARI PURNOMO	3.45	139
6	1110651006	MIFTAHUL HUDA	3.7	110
7	1110651007	YASSIR ARAFAT	2.2	98
8	1110651008	ARIEF RAHMAN HAKIM	0.52	38
9	1110651009	ERFAN BAHTIAR	3.85	153
10	1110651010	AHMAD TAUFIQ	2.72	123
11	1110651011	NANANG CHAELANI	2.29	132
12	1110651012	ANDIKA YUDI PRASETYO	2.59	98
13	1110651013	DIANDRA ANGGARA PUTRA	1.58	78
14	1110651014	FADLI	2.75	115
15	1110651015	WA WAN WIDI KURNIAWAN	0.14	78
16	1110651016	ALFERO HENDI WIANTO SUTARMONO	0.73	98
17	1110651017	FAJAR WAHYUDI FARMADI	3.46	119
18	1110651018	IMAM ROMLI	2.58	101
19	1110651019	NURUL WEMI ISPRIYANI	3.24	110
20	1110651020	AGUS DWI PRAYOGO	0.53	108
21	1110651021	RIFKY ARDIANSYAH	3.33	119
22	1110651022	PRAYOGI DARMAWAN	0.97	101
23	1110651023	YODI HARYO TARUNO	0.44	98
24	1110651024	ANDI ARIF HIDAYATULLAH	1.79	101
25	1110651025	YULIA QURROTU AINI	2.77	115
26	1110651026	TAUFIQ HIDAYAT	2.64	121
27	1110651027	MUHAMMAD ARIF RAHMAN HAKIM	2.53	108
28	1110651028	A CHSAN DWILAKSONO	1.7	98
29	1110651029	INDRA CAHYA FEBRIYANTORO	2.87	128
30	1110651030	ALGA VIONITA AGUSTIN	2.39	78
31	1110651031	ERICK ADI NUGRAHA	3.46	115
32	1110651032	BAGUS ROHMADTULLOH	2.97	121
33	1110651033	IRA DWI AYUNITA	0.42	98
34	1110651034	MUHAMAD ARFIQ KHOIRON	3.62	131
35	1110651035	NIMATUL FITHRIYAH	2.92	129
36	1110651036	MOHAMAD TURMUDI	2.96	110
37	1110651037	ZAINUL MUHABAT	3.55	136
38	1110651038	DEWI INTAN MUSTIKASARI	3.69	131
39	1110651039	GAGAH PRAWONO SETYA PUTERA	3.65	143
40	1110651040	YENI WAHYUNING TIYAS	3.42	115
41	1110651041	NURCAHYA JUHANNI	1.97	78
42	1110651042	AKBAR PRATAMA	3.42	121
43	1110651043	MUHAMMAD HAFIDI	2.68	129

### 4.2 Proses K-Means pada Aplikasi

Aplikasi dibuat berdasarkan kasus yang ada, sehingga untuk setiap data yang

diinputkan akan dilakukan proses perhitungan untuk mengetahui klasifikasi kelulusan. Berikut ini adalah penerapan K-Means pada aplikasi :

1. Inputan Data Mahasiswa

Inputan data mahasiswa berisikan tentang data mahasiswa seperti nim, nama, dan prediksi kelulusan yang ditunjukkan pada gambar 4.7 sebagai berikut :

The screenshot shows a form titled "Tambah Data Sekolah". It contains four input fields: "NIM", "Nama Mahasiswa", "Tahun Angkatan", and "Status Kelulusan". At the bottom, there are "Submit" and "Cancel" buttons.

Gambar 4.7 Inputan Mahasiswa

2. Inputan Nilai

Inputan nilai merupakan sebuah menu yang disediakan untuk mengisi nilai kriteria masing-masing mahasiswa, yaitu jumlah sks mahasiswa yang telah ditempuh selama 6 semester dan ipk terakhir pada semester ke-6, berikut ini adalah inputan nilai mahasiswa berdasarkan kriteria :

The screenshot shows a form titled "Tambah Nilai". It contains four input fields: "Nama", "Nilai", "IPK", and "SKS". At the bottom, there are "Submit" and "Cancel" buttons.

Gambar 4.8 Inputan Nilai Mahasiswa

3. Kriteria

Berisikan tentang kriteria yang digunakan, penulis memiliki 4 kriteria utama dengan penggunaan kriteria pada pengujian adalah 2

kriteria. Kriteria tersebut terdiri dari IPK, SKS, Beasiswa, dan Organisasi. Kriteria yang digunakan penulis adalah IPK dan SKS. Pada menu kriteria, penulis menyediakan sebuah mekanisme on dan off dimana keempat kriteria dapat diaktifkan ataupun dapat di non aktifkan. Berikut adalah tampilan dari kriteria :

ID	KRITERIA		
1	IPK	✓	✗
2	SKS	✓	✗
3	Beasiswa	✓	✗
4	Organisasi	✓	✗

Gambar 4.9 Tampilan Kriteria

The screenshot shows a form titled "Edit Kriteria". It contains two input fields: "Nama Kriteria" (with the value "Beasiswa") and "Status Kriteria" (with the value "Mati"). At the bottom, there are "Save" and "Cancel" buttons.

Gambar 4.10 Tampilan On/Off Kriteria

4. Menentukan Pusat Cluster

Penentuan pusat cluster dilakukan setelah data telah terinput secara keseluruhan. Data akan dipilih secara acak sesuai dengan kategorinya yang ditunjukkan pada gambar 4.11 sebagai berikut :

NO	NAMA	KRITERIA	
		IPK	SKS
1	WAHYU FARABI FIRMANSAH	3,63	130
2	ILMI AMALYAH RIDKY	3,33	110
3	ARIEF RAHMANN HAKIM	0,52	38

Gambar 4.11 Pusat Cluster

5. Jarak Cluster

Jarak cluster merupakan sebuah jarak terdekat iterasi dengan cluster, berikut ini implimentasi pada aplikasi yang ditunjukkan pada gambar 4.12 sebagai berikut :

Jarak Terhadap Pusat Cluster			
NAMA	CLUSTER		
	WAHYU FARABI FIRMANSAH	ILMI AMALIAH RIZKY	ARIEF RAHMAN HAKIM
WAHYU FARABI FIRMANSAH	0.000	20.002	92.053
ILMI AMALIAH RIZKY	20.002	0.000	72.055
ZHENI THIBAN	55.072	25.091	37.001
SYARIFAH FAJRHWATI	5.002	15.001	87.050
SUGENG HARI PURNOMO	9.002	29.000	101.042
MIFTAHUL HUDA	20.000	0.370	72.070
YASSIR ARAFAT	32.032	12.053	60.024
ARIEF RAHMAN HAKIM	92.053	72.055	0.000
ERFAN BAHTIAR	23.001	43.003	115.048
AHMAD TALUFIQ	7.059	13.014	85.028
NANANG CHAELANI	2.407	22.025	94.017
ANDIKA YUDI PRASETYO	32.017	12.023	60.036
DINDRA ANGGARA PUTRA	52.040	32.048	40.014
FAUZI	15.026	5.034	77.032

Gambar 4.12 Jarak Pusat Cluster

## 6. Hasil Klasifikasi

Berikut ini adalah hasil klasifikasi dari implementasi metode K-Means :

Hasil Akhir	
Keanggotaan C1	121 Data
Keanggotaan C2	135 Data
Keanggotaan C3	9 Data

Gambar 4.13 Hasil Klasifikasi

## 4.3 Hasil Akurasi Sistem

Hasil dari analisa menggunakan hasil keputusan dari 266 dataset yang terklasifikasikan ke dalam masing-masing cluster kemudian dilakukan kinerja rata-rata Accuracy sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual.

### 1. Akurasi Keanggotaan C1

Klasifikasi yang dihasilkan oleh sistem dengan menerapkan metode K-Means pada keanggotaan C1 sejumlah 108 mahasiswa, sedangkan klasifikasi yang dihasilkan oleh perhitungan K-Means sejumlah 33 mahasiswa. Sehingga diperoleh hasil akurasi sebagai berikut :

Tabel 4.5 Tabel Hasil Klasifikasi Keanggotaan C1 (Klasifikasi Lulus Cepat)

No	Hasil Sistem	No	Prediksi Awal	Keterangan
	Nama		Nama	
1	Wahyu Farabi Firmansah	1	Wahyu Farabi Firmansah	TP
2	Erfan Bahtiar	2	Erfan Bahtiar	TP
3	Muhamad Arfiq Khoiron	3	Muhamad Arfiq Khoiron	TP
4	Zainul Muhabat	4	Zainul Muhabat	TP
5	Dewi Intan Mustikasari	5	Dewi Intan Mustikasari	TP
6	Gagah Prawono Setya Putera	6	Gagah Prawono Setya Putera	TP
7	Laili Amalia Shofiana	7	Laili Amalia Shofiana	TP
8	Nur Anita	8	Nur Anita	TP
9	Firlian Okto Hari Pratama	9	Firlian Okto Hari Pratama	TP
10	Titis Suwardiningsih	10	Titis Suwardiningsih	TP
11	Fachry Husaini	11	Fachry Husaini	TP
12	Putri Megawati	12	Putri Megawati	TP
13	Laras Dwi Mawarni	13	Laras Dwi Mawarni	TP
14	Kunti Triantini	14	Kunti Triantini	TP
15	Fahmi Luthfillah	15	Fahmi Luthfillah	TP
16	Winda Alfa Mufida	16	Winda Alfa Mufida	TP
17	A ditya Angraini Yoga Kresna Wijayanti	17	A ditya Angraini Yoga Kresna Wijayanti	TP
18	Fauziah Ragilyah Prastikasari	18	Fauziah Ragilyah Prastikasari	TP
19	Lukie Perdanasari	19	Lukie Perdanasari	TP
20	Amelia Ayu Angraini	20	Amelia Ayu Angraini	TP
21	Imam Azhari	21	Imam Azhari	TP
22	Gufron Fauzi	22	Gufron Fauzi	TP
23	Risqi Albi Khainillah	23	Risqi Albi Khainillah	TP
24	Novan Agung Sucalhyo	24	Novan Agung Sucalhyo	TP
25	Geygi Neviansyah	25	Geygi Neviansyah	TP
26	Shadiqul Hasan Saifurrijal	26	Shadiqul Hasan Saifurrijal	TP
27	David Heriawanto	27	David Heriawanto	TP
28	Respati Bayu Adi	28	Respati Bayu Adi	TP
29	Dermawan Adi Wijaya	29	Dermawan Adi Wijaya	TP
30	Alifan Alfun Salim	30	Alifan Alfun Salim	TP
31	Faris Luthfir Rohman	31	Faris Luthfir Rohman	TP
32	Ibnati Nur Rohmaniyah	32	Ibnati Nur Rohmaniyah	TP
33	Ilmi Amaliah Rizky	33	Ilmi Amaliah Rizky	TP
34	Andika Yudi Prasetyo	34	Andika Yudi Prasetyo	TP
35	Alfero Hendi Wianto Sutarmono	35	Alfero Hendi Wianto Sutarmono	TP
36	Imam Romli	36	Imam Romli	TP
37	Nurul Wemi Ispriyani	37	Nurul Wemi Ispriyani	TP

Tabel 4.5 Tabel Hasil Klasifikasi Keanggotaan C1 (Klasifikasi Lulus Cepat)

No	Hasil Sistem	No	Prediksi Awal	Keterangan
	Nama		Nama	
38	Muhammad Arif Rahman Hakim	38	Muhammad Arif Rahman Hakim	TP
39	Alga Vionita Agustin	39	Alga Vionita Agustin	TP
40	Mohamad Turmudi	40	Mohamad Turmudi	TP
41	Dirta Andre Baskara	41	Dirta Andre Baskara	TP
42	Cahya Bagus Prianto	42	Cahya Bagus Prianto	TP
43	Ahmad Sony Listen	43	Ahmad Sony Listen	TP
44	Kunto Nashiruddin Ahmad	44	Kunto Nashiruddin Ahmad	TP
45	Bukhori	45	Bukhori	TP
46	Bahruallah	46	Bahruallah	TP
47	Bairotul Khoirul Anam	47	Bairotul Khoirul Anam	TP
48	Yogi Hermawan	48	Yogi Hermawan	TP
49	Ericko Dwi Yuliantoro	49	Ericko Dwi Yuliantoro	TP
50	Alfian Zulmi	50	Alfian Zulmi	TP
51	Muhammad Fachmy Rizal	51	Muhammad Fachmy Rizal	TP
52	Nishar Dhani Firmansyah	52	Nishar Dhani Firmansyah	TP
53	Farida Nurhidayah	53	Farida Nurhidayah	TP
54	Dimas Muharrom Rosdy Bahalwan	54	Dimas Muharrom Rosdy Bahalwan	TP
55	Ikbaril Catur Prasatya	55	Ikbaril Catur Prasatya	TP
56	Rofita Heru	56	Rofita Heru	TP
57	Mifahul Huda	57	Mifahul Huda	TP
58	Yassir Arafat	58	Yassir Arafat	TP
59	Junaedi Efendi	59	Junaedi Efendi	TP
60	Dimas Dwi Prasetyo	60	Dimas Dwi Prasetyo	TP
61	Ninis Kartika Sari	61	Ninis Kartika Sari	TP
62	Deny Kusuma Atmaja	62	Deny Kusuma Atmaja	TP
63	Christian Natanael	63	Christian Natanael	TP
64	Fery Dwi Septian	64	Fery Dwi Septian	TP
65	Wahyuning Tiyas Andarini	65	Wahyuning Tiyas Andarini	TP
66	Juwita Dwi Wahyudi	66	Juwita Dwi Wahyudi	TP
67	Dendy Yogaswara Anggaditya	67	Dendy Yogaswara Anggaditya	TP
68	Hamim Jazuli	68	Hamim Jazuli	TP
69	Isya Andika	69	Isya Andika	TP
70	Achmad Muhtar Shokheh	70	Achmad Muhtar Shokheh	TP
71	Desi Emita Putri	71	Desi Emita Putri	TP
72	Anugerah Riski Fardana	72	Anugerah Riski Fardana	TP
73	Muh Syamsudin Badari	73	Muh Syamsudin Badari	TP

LAMPIRAN ...

Tabel 4.6 Perhitungan Nilai Accuracy

TP (True Positif)	FP (False Positif)
-------------------	--------------------

<b>156</b>	<b>110</b>
FN (False Negatif)	TN (True Negatif)
<b>0</b>	<b>0</b>

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$= \frac{156 + 0}{156 + 0 + 0 + 110}$$

$$= 59 \%$$

Hasil dari perhitungan *Accuracy* didapat nilai akurasi sebesar 59%. Hal ini menunjukkan bahwa sistem dapat mengklasifikasi data dengan benar namun masih terdapat beberapa data yang terklasifikasikan pada kategori lain.

## 5. Kesimpulan

Dari uji coba dan analisa yang telah dijelaskan dalam bab sebelumnya, maka dapat diambil beberapa kesimpulan sebagai berikut :

1. Hasil prediksi kelulusan dengan menerapkan metode K-Means dapat berjalan sesuai dengan harapan. Dengan sampling data pada pembahasan didapat 108 mahasiswa dengan kelulusan cepat, 107 mahasiswa dengan kelulusan tepat, dan 50 mahasiswa dengan kelulusan lambat.
2. Hasil perhitungan *dataset* dengan metode *K-Means* didapatkan hasil *Accuracy* 59%, hal ini dikarenakan tidak memfilter data ,dimana data set yang digunakan data yang belum diolah dengan data yang diolah oleh sistem (telah melalui proses perhitungan *K-Means* terhitung kembali) sehingga menghasilkan pengelompokkan (klasifikasi) yang kurang benar dan berakibat memiliki nilai *accuracy* rendah.

## 5.2 Saran

Dengan segala kelebihan yang terdapat pada proyek akhir ini, tidak terlepas dari kekurangan yang tentunya sangat diharapkan adanya saran-saran yang mendukung proses penyempurnaannya. Penulis ingin memberikan beberapa saran yang mungkin dapat membantu dalam pengembangan Tugas Akhir ini adalah sebagai berikut :

1. Untuk pengembangan selanjutnya, penulis menyarankan penggunaan metode lain seperti *Fuzzy*, *Topsis*, *K-means*, ataupun membandingkannya metode lainnya.
2. Menambah *variable* baru yang ditambah dalam. Semakin banyak *variable* yang digunakan, maka akan menghasilkan hasil keputusan yang lebih *valid*.

## DAFTAR PUSTAKA

- A. A. Mahdi, A. M. Razali and A. A. Salih, "The Diagnosis of Chicken Pox and Measles Using Fuzzy Relations," *Journal of Basic and Applied Scientific Research*, pp. 679-686, 2011.
- Abidin, Taufik Fuadi.2009. "Naive Bayesian Classifier", Jurusan Informatika Unsyiah.bahan kuliah Data Mining program study Informatika FMIPA-Unsyiah
- Angga, Asfan Candra. 2014. *Diagnosa Penyakit Sinusitis Menggunakan Metode Algoritma Genetika Dan Bayesian Berbasis Jsp*. Surabaya : ITS Surabaya
- Basuki,Ahmad. 2006, "Metode Bayes", PENSITS, Surabaya.
- Gershon AA. Meas'es. Dalam : *Mandell ed. Principles and practice of infectious diseases*; 5<sup>th</sup> ed. London : Churchill Livingstone, 2000; 132-4.
- Handayani, L dan Sutikno, T. 2008. *Sistem Pakar untuk Diagnosis Penyakit THT Berbasis Web dengan "e2gLite Expert System Shell"*. Jurnal

- Teknologi Industri, Volume 12, Nomor 1.
- Hariyanto, B, 2004, *Sistem Manajemen Basis Data*, Bandung : Penerbit Informatika.
- Haryanto.2012. *Penerapan Naïve Bayes Classifier Pada Sistem Pakar Untuk Klasifikasi Bakteri Escherichis Coli*. Medan : Universitas Sumatera Utara.
- J.Klein,“Rubella (German Measles),” 2012 <http://kidshealth.org/parent/infections/skin/germanmeasles>. Diakses 23 Maret 2016.
- Kondo, Ferry.2015. *Klasifikasi Variabel Penentu Kelulusan Mahasiswa FMIPA Unpati Menggunakan Metode CHAID*. Ambon : Universitas Pattimura Ambon.
- Krugman S, Katz SL, Gershon AA, Wilfert CIV!, eds. Measles (Rubeola). *Infectious Disease of Children*. St Louis: The Mosby Co, 1992; 223-45 .
- Krugman S, Katz SL, Gershon AA, Wilfert CIVI, eds. Measles (Rubeola). *Infectious Disease of Children*. St Louis: The Mosby Co, 1992; 223-45.
- Kusrini & Emha Taufiq Luthfi.2009. *Algoritma Data Mining*, Penerbit Andi Yogyakarta.
- Maryuni,Bekti.2015. *Aplikasi Android Untuk Mendeteksi Penyakit Saluran Kemih Menggunakan Algoritma Naive Bayes*. Jember : Politeknik Negeri Jember.
- Perdana, Yudi.2009. *Klasifikasi berita berbahasa Indonesia menggunakan naïve bayes classifier*. *Jurnal Nasional FPMIPA UPI*.
- Prasetyo, Eko.2014. *Perbandingan K-Support Vector Nearest Neighbor Terhadap Decision Tree Dan Naive Bayes*. Surabaya : Universitas Bayangkara Surabaya.
- Premashanti, Gita.2015. *Penerapan Metode Klustering Dengan Algoritma K-Means Untuk Prediksi Kelulusan Mahasiswa Pada Program Studi Teknik Informatika Strata Satu*. Semarang : Universitas Dian Nuswantoro .
- Rahayu,Sri.2013.*Sistem Pakar Untuk Mendiagnosa Penyakit Gagal Ginjal Dengan Menggunakan Metode Bayes*. Semarang: Universitas Dian Nuswantoro.
- Setiawan, I Made. *Penyakit Campak*. Jakarta : PT Sagung Seto; 2008.
- Whitten. 2004. *Decision Support Systems and Intelligent Systems. 6th edition*. Prentice Hall: Upper Saddle River, NJ.
- WHO, “Rubella Vaccines: WHO Position Paper,” *Weekly Epidemiological Record*, pp. 349-360, 15 July 2008.
- WHO. *World health Statistics 2011* ([http://www.who.int/csr/don/2011\\_04\\_21/en/](http://www.who.int/csr/don/2011_04_21/en/)) Diakses 23 Maret 2016.
- Yakub,Suardin.2008. *Sistem Pakar Deteksi Penyakit Diabetes Mellitus dengan Menggunakan Pendekatan Naïve Bayesian berbasis Web*. Medan : Universitas Sumatera Utara.