

BAB IV

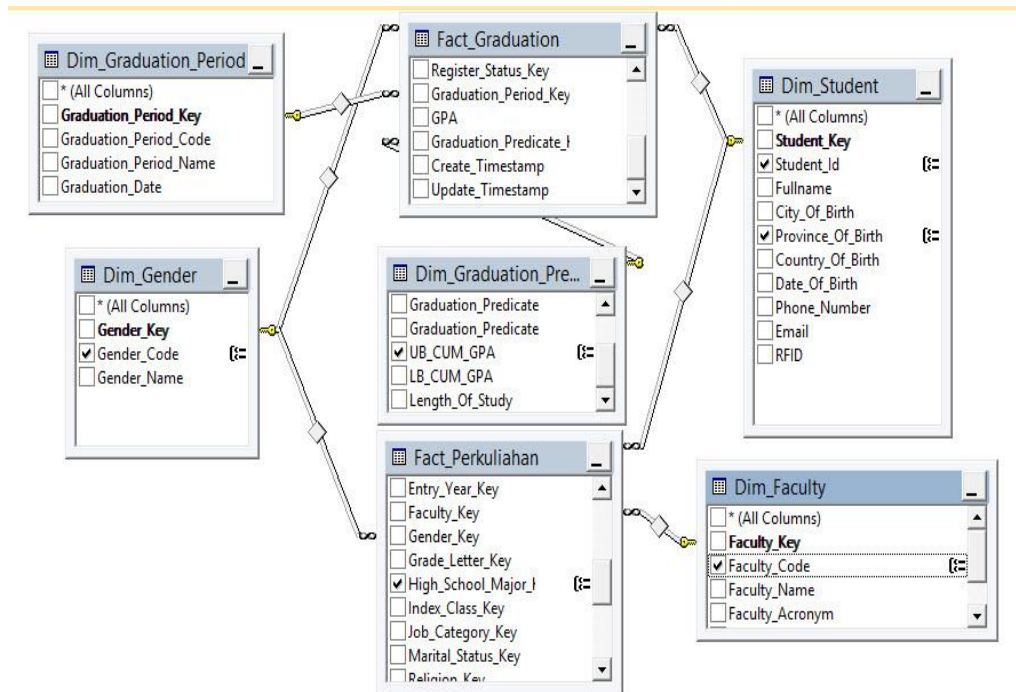
PEMBAHASAN

Pada penelitian ini menggunakan data mahasiswa Fakultas Teknik alumni 2013-2015, rentan waktu itu di jadikan sebagai bahan penelitian karena cukup relevan dengan tahun sekarang, faktor lain peneliti mengambil data alumni 2013-2015 karena data yang masuk di biro sistem informasi UMY belum terlalu lengkap untuk data tahun-tahun sebelumnya.

4.1 Pengumpulan Data

Sumber data utama yang digunakan dalam penelitian ini adalah *data warehouse* mahasiswa Fakultas Teknik UMY alumni tahun 2013, 2014, dan 2015. Data yang diterima dalam bentuk tabel *factual* dan dimensi, dapat diakses menggunakan *SQL Server Management 2014*. Agar dapat mengakses *data warehouse*, peneliti harus mempunyai hak akses ke database *server* kampus UMY di bawah wewenang Biro Sistem Informasi (BSI).

Data yang diambil dari *data warehouse* ini menggunakan dua tabel *factual* yaitu *fact_grad_high_major*. Pengambilan data dilakukan berdasarkan *uation* dan *fact_perkuliahan* sedangkan untuk tabel *dimensional* memiliki 5 tabel yaitu: *dim_gender*, *dim_graduation_periode*, *dim_department*, *dim_faculty*, dim atribut yang akan digunakan sebagai bahan penelitian, penyeleksian data dari data asli (*data warehouse*) dapat dilihat di **Gambar 4.1** sebagai berikut.



Gambar 4.1 Tampilan view dari data warehouse.

4.2 Penyeleksian Data (*Data selection*)

Data mahasiswa Fakultas Teknik UMY di *filter* terlebih dahulu dan diambil beberapa atribut dari tabel-tabel dimensi dan tabel *factual* untuk dianalisa. Dalam proses filter data ini, penulis mengambil data lulusan Fakultas Teknik. Pengambilan data dari *SQL Server Management* diolah dalam *SQL Server Data Tools for Visual Studio 2013* untuk dibuat dalam *dataset* yang baru. Jumlah data alumni mahasiswa Fakultas Teknik dari tahun 2013 sampai 2015 adalah 466 *record*. Terdapat empat atribut yang digunakan, yaitu, *student*, *high school major*, *gender* dan *province of birth*. Penyeleksian data dari data asli (*data warehouse*) seperti pada **Gambar 4.2** sebagai berikut.

Column	Alias	Table	Outp...	Sort Type	Sort Order	Group By	Filter
Student_Id		Dim_Stud...	<input checked="" type="checkbox"/>			Group By	
Gender_Code		Dim_Gend...	<input checked="" type="checkbox"/>			Group By	
UB_CUM_GPA		Dim_Grad...	<input checked="" type="checkbox"/>			Group By	
Province_Of_Birth		Dim_Stud...	<input checked="" type="checkbox"/>			Group By	
High_School_Major_Key		Fact_Perk...	<input checked="" type="checkbox"/>			Group By	
Faculty_Code		Dim_Facul...	<input checked="" type="checkbox"/>			Group By	= 1
Graduation_Period_Key		Fact_Grad...	<input checked="" type="checkbox"/>			Group By	IN ('20151','20132','20132','20133','20141','20142','20143')
			<input type="checkbox"/>				

Gambar 4.2 Tampilan *filter* data dari *data warehouse*

Setelah semua data lulusan mahasiswa Fakultas Teknik tahun 2013 sampai 2015 didapatkan dengan jumlah 466 *record*, penulis kembali melakukan proses *filter* data, hal ini dilakukan karena dari beberapa data tersebut terdapat data yang kurang lengkap, data yang kurang lengkap ini perlu dihilangkan karena akan mempengaruhi proses hasil penelitian.

Proses *filter* data yang kurang lengkap ini dilakukan dengan cara menghapus semua data-data yang dianggap tidak memiliki atribut yang lengkap. Setelah dilakukan *filter* maka total data yang layak di jadikan bahan penelitian adalah sebanyak 254 *record*.

4.3 Pembersihan Data

Setelah semua data lulusan mahasiswa Fakultas Teknik tahun 2013 sampai 2015 didapatkan dengan jumlah 466 *record*, setelah data diseleksi sesuai dengan atribut yang akan digunakan maka dilakukan *cleaning data* agar tidak ada duplikasi data, tidak *missing value* dan memperbaiki kesalahan-kesalahan yang ada pada *dataset* baru dalam bentuk format *excel* dari data yang diambil terdapat *unknown* data atau data yang tidak lengkap sebagaimana pada yang di tunjukan pada **Gambar 4.3**. Pada tahap inilah data akan dilakukan *cleaning* atau pembersihan

data. Data yang kurang lengkap ini perlu dihilangkan karena akan mempengaruhi hasil penelitian, proses *cleaning* data yang kurang lengkap ini dilakukan dengan cara menghapus semua data-data yang dianggap tidak memiliki atribut yang lengkap. Setelah dilakukan *cleaning* maka total data yang layak di jadikan bahan penelitian adalah sebanyak 254 *record*, sebagaimana terlihat pada **Gambar 4.4**.

Data yang telah melewati tahap *cleaning* akan disimpan dalam *dataset* yang baru menggunakan *Microsoft Office Excel* dengan format *.csv*. karena perangkat lunak Weka hanya dapat membuka file dengan format *.csv*.

	A	B	C	D	E	F	G	H
	Provinsi	Gender	Fakultas	PredikatKelulusan	periodeKelulusan	AsalSekolah	RegisterStatus	
1								
2	2	Laki-Laki	TEKNIK	Sangat Memuaskan	Periode III 2013/2014 (Kamis, 12 Juni 2014)	IPA	Mahasiswa Baru	
3	2	Laki-Laki	TEKNIK	-	Periode I 2013/2014 (19 Oktober 2013)	IPA	Mahasiswa Baru	
4	1	Laki-Laki	TEKNIK	Memuaskan	Periode I 2015/2016 (Sabtu, 17 Oktober 2015)	KEJURUAN	Mahasiswa Baru	
5	1	Laki-Laki	TEKNIK	-	Periode I 2013/2014 (19 Oktober 2013)	[Tidak Ada]	Mahasiswa Baru	
6	3	Laki-Laki	TEKNIK	Sangat Memuaskan	Periode II 2014/2015 (Sabtu, 14 Februari 2015)	[Tidak Ada]	Mahasiswa Baru	
7	1	Laki-Laki	TEKNIK	-	Periode II 2013/2014 (15 Pebruari 2014)	[Tidak Ada]	Mahasiswa Baru	
8	2	Laki-Laki	TEKNIK	-	Periode I 2013/2014 (19 Oktober 2013)	[Tidak Ada]	Mahasiswa Baru	
9	2	Laki-Laki	TEKNIK	-	Periode II 2013/2014 (15 Pebruari 2014)	[Tidak Ada]	Mahasiswa Baru	
10	4	Laki-Laki	TEKNIK	Memuaskan	Periode III 2014/2015 (Selasa, 16 Juni 2015)	KEJURUAN	Mahasiswa Baru	
11	1	Laki-Laki	TEKNIK	-	Periode II 2013/2014 (15 Pebruari 2014)	IPA	Mahasiswa Baru	
12	1	Laki-Laki	TEKNIK	-	Periode I 2013/2014 (19 Oktober 2013)	IPA	Mahasiswa Baru	
13	5	Laki-Laki	TEKNIK	-	Periode I 2013/2014 (19 Oktober 2013)	IPA	Mahasiswa Baru	
14	3	Laki-Laki	TEKNIK	Memuaskan	Periode III 2014/2015 (Selasa, 16 Juni 2015)	IPS	Mahasiswa Baru	
15	1	Laki-Laki	TEKNIK	Sangat Memuaskan	Periode II 2014/2015 (Sabtu, 14 Februari 2015)	KEJURUAN	Mahasiswa Baru	
16	2	Laki-Laki	TEKNIK	Sangat Memuaskan	Periode III 2014/2015 (Selasa, 16 Juni 2015)	KEJURUAN	Mahasiswa Baru	
17	1	Laki-Laki	TEKNIK	-	Periode II 2013/2014 (15 Pebruari 2014)	IPS	Mahasiswa Baru	
18	6	Laki-Laki	TEKNIK	-	Periode I 2013/2014 (19 Oktober 2013)	IPS	Mahasiswa Baru	
19	3	Laki-Laki	TEKNIK	-	Periode I 2013/2014 (19 Oktober 2013)	IPA	Mahasiswa Baru	
20	1	Laki-Laki	TEKNIK	-	Periode I 2013/2014 (19 Oktober 2013)	IPS	Mahasiswa Baru	
21	3	Laki-Laki	TEKNIK	-	Periode I 2013/2014 (19 Oktober 2013)	KEJURUAN	Mahasiswa Baru	
22	2	Laki-Laki	TEKNIK	-	Periode II 2013/2014 (15 Pebruari 2014)	[Tidak Ada]	Mahasiswa Baru	
23	3	Laki-Laki	TEKNIK	-	Periode II 2013/2014 (15 Pebruari 2014)	IPA	Mahasiswa Baru	

Gambar 4.3 data yang belum dibersihkan

	A	B	C	D	E
	PredikatKelulusan	Gender	Provinsi	AsalSekolah	
1					
2	No Cumlaude	Laki-laki	Jawa Bali	KEJURUAN	
3	No Cumlaude	Laki-laki	kalimantan	KEJURUAN	
4	No Cumlaude	Laki-laki	Sumtera Kepri	IPS	
5	No Cumlaude	Laki-laki	Jawa Bali	KEJURUAN	
6	No Cumlaude	Laki-laki	Jawa Bali	IPA	
7	No Cumlaude	Laki-laki	Jawa Bali	IPA	
8	No Cumlaude	Laki-laki	kalimantan	IPS	
9	No Cumlaude	Laki-laki	Nusa Tenggara, Maluku, Papua	IPS	
10	No Cumlaude	Laki-laki	Sumtera Kepri	KEJURUAN	
11	No Cumlaude	Laki-laki	Jawa Bali	IPA	
12	No Cumlaude	Laki-laki	Sumtera Kepri	IPS	
13	No Cumlaude	Laki-laki	Jawa Bali	IPA	
14	No Cumlaude	Laki-laki	kalimantan	IPS	
15	No Cumlaude	Laki-laki	Jawa Bali	KEJURUAN	
16	No Cumlaude	Laki-laki	Jawa Bali	KEJURUAN	
17	No Cumlaude	Laki-laki	Sumtera Kepri	IPS	
18	No Cumlaude	Laki-laki	Jawa Bali	IPS	
19	No Cumlaude	Laki-laki	Jawa Bali	KEJURUAN	
20	No Cumlaude	Laki-laki	Sumtera Kepri	KEJURUAN	
21	No Cumlaude	Laki-laki	Nusa Tenggara, Maluku, Papua	IPA	
22	No Cumlaude	Laki-laki	kalimantan	IPA	
23	No Cumlaude	Laki-laki	Sumtera Kepri	IPA	

Gambar 4.4 Data telah melalui proses *cleaning*

4.4 Transformasi Data

Pada tahap ini, penulis melakukan proses perubahan data, agar data dapat diolah dengan menggunakan algoritma *K-Means clustering*. Data yang *non-numeric* dilakukan proses inisiasi ke dalam bentuk *numeric*. Pada penelitian ini seluruh atribut diubah kebentuk *numeric*, yaitu, *High School Major*, *Gender*, dan *Province Of Birth*. Proses inisiasinya sebagai berikut:

a. Daerah Asal

Pada kota asal calon mahasiswa baru dilakukan pengurutan data dari frekuensi terbesar hingga yang terkecil

Tabel 4.1 Data atribut daerah asal

Daerah Asal	Frekuensi	Inisial
Jawa Bali	140	1
Sumatra Kepri	69	2
Kalimantan	26	3
Nusa Tenggara, Maluku, dan Papua	13	4
Sulawesi	5	5

b. *Gender*

Pada jalur masuk transformasi data dilakukan berdasarkan frekuensi terbesar hingga terkecil.

Tabel 4.2 Data atribut *gender*

Gender	Frekuensi	Inisial
Laki-laki	218	1
Perempuan	35	2

c. Asal Sekolah

Pada jalur masuk transformasi data dilakukan berdasarkan frekuensi terbesar hingga terkecil.

Tabel 4.3 Data atribut asal sekolah

High School Major	Frekuensi	Inisial
IPA	145	1
Kejuruan	79	2
IPS	28	3
Bahasa	1	4

4.5 Implementasi

Pada tahap ini dilakukan pemodelan data, adapun metode yang dipakai pada penelitian ini adalah metode *clustering* dengan menggunakan algoritma *K-means*. Data yang sudah dikumpulkan, diseleksi dan ditransformasi akan dilakukan pengolahan dari data tersebut dengan menggunakan metode *clustering*. Metode ini bekerja dengan melakukan pengelompokan data yang memiliki kesamaan karakteristik pada setiap data.

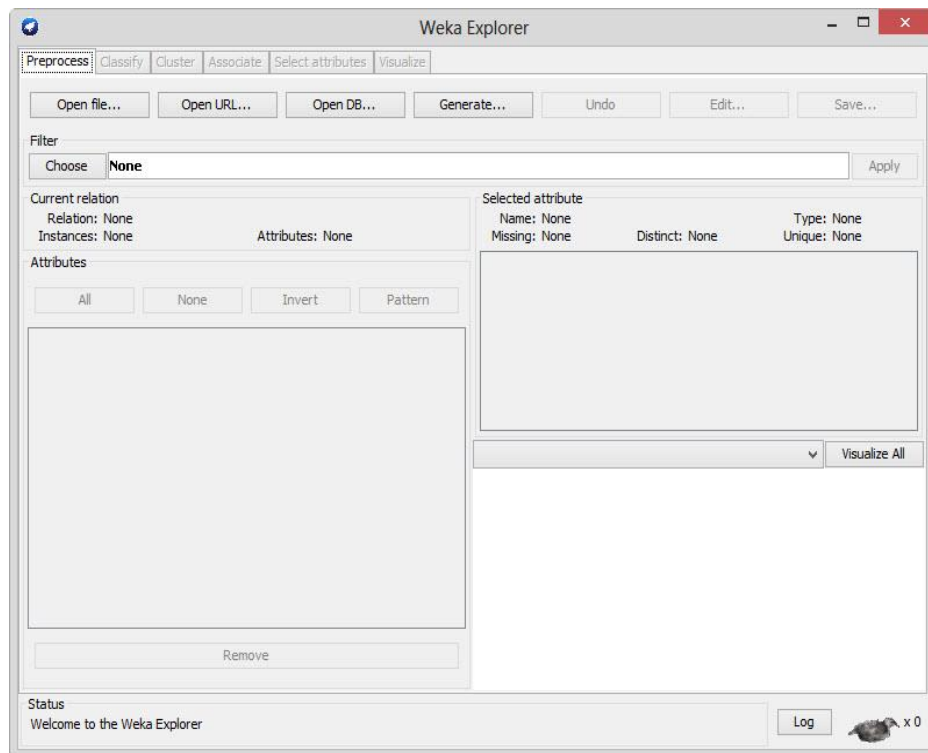
Atribut pada data mahasiswa *cumlaude* Fakultas Teknik periode 2013 - 2015 terdiri dari data *non-numeric*, data tersebut dapat dianalisa dengan menggunakan perangkat lunak *weka*. Data lulusan mahasiswa Fakultas Teknik 2013-2015 memiliki 466 data. **Gambar 4.5** di bawah ini adalah data lulusan mahasiswa Fakultas Teknik dengan format *.CSV* yang diakses melalui *software* *weka*.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	PredikatKelulusan	Gender	Provinsi	AsalSekolah													
2	No Cumlaude	Laki-laki	Jawa Bali	KEJURUAN													
3	No Cumlaude	Laki-laki	kalimantan	KEJURUAN													
4	No Cumlaude	Laki-laki	Sumtera Kepri	IPS													
5	No Cumlaude	Laki-laki	Jawa Bali	KEJURUAN													
6	No Cumlaude	Laki-laki	Jawa Bali	IPA													
7	No Cumlaude	Laki-laki	Jawa Bali	IPA													
8	No Cumlaude	Laki-laki	kalimantan	IPS													
9	No Cumlaude	Laki-laki	Nusa Tenggara, Maluku, Papua	IPS													
10	No Cumlaude	Laki-laki	Sumtera Kepri	KEJURUAN													
11	No Cumlaude	Laki-laki	Jawa Bali	IPA													
12	No Cumlaude	Laki-laki	Sumtera Kepri	IPS													
13	No Cumlaude	Laki-laki	Jawa Bali	IPA													
14	No Cumlaude	Laki-laki	kalimantan	IPS													
15	No Cumlaude	Laki-laki	Jawa Bali	KEJURUAN													
16	No Cumlaude	Laki-laki	Jawa Bali	KEJURUAN													
17	No Cumlaude	Laki-laki	Sumtera Kepri	IPS													
18	No Cumlaude	Laki-laki	Jawa Bali	IPS													
19	No Cumlaude	Laki-laki	Jawa Bali	KEJURUAN													
20	No Cumlaude	Laki-laki	Sumtera Kepri	KEJURUAN													
21	No Cumlaude	Laki-laki	Nusa Tenggara, Maluku, Papua	IPA													
22	No Cumlaude	Laki-laki	kalimantan	IPA													
23	No Cumlaude	Laki-laki	Sumtera Kepri	IPA													

Gambar 4.5 Data mahasiswa dalam *Format .csv*

4.5.1 Pegujian data menggunakan *software Weka*

Data mahasiswa lulusan Fakultas Teknik UMY tahun 2013-2015 dibuat *dataset* dengan *format .xlsx (excel) convert ke format .CSV* agar dapat dibuka menggunakan *software Weka*. Seperti pada **Gambar 4.6** di bawah ini.

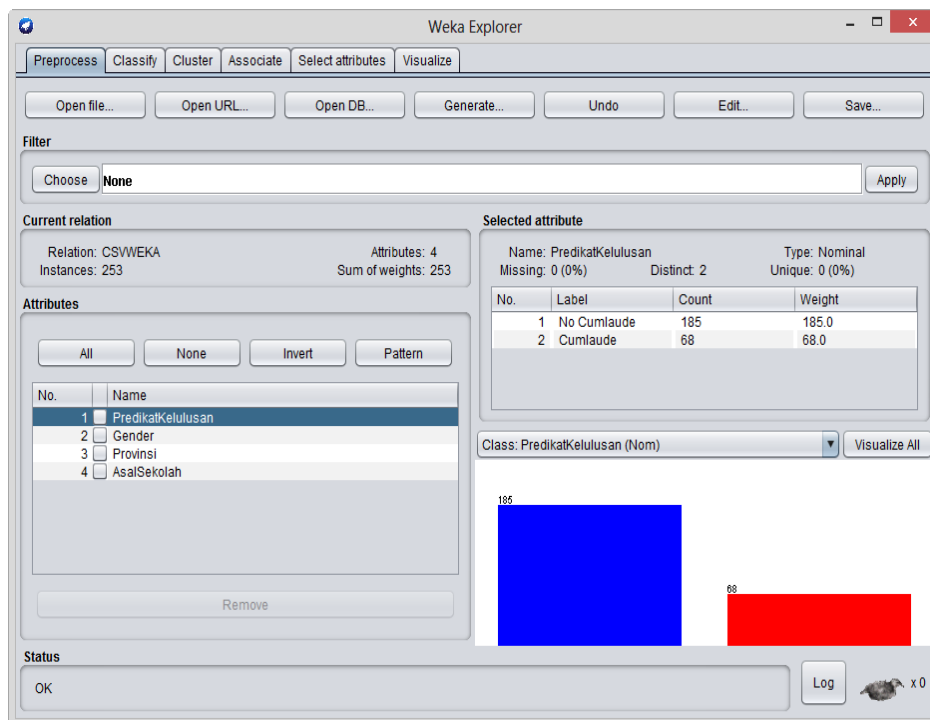


Gambar 4.6 Weka explorer.

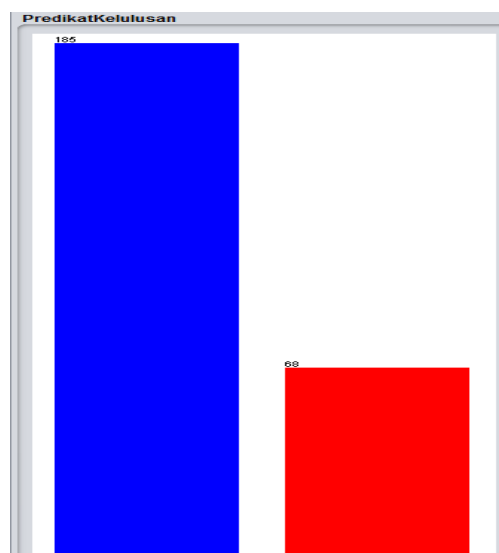
Setelah membuka data mahasiswa yang akan dianalisa, maka selanjutnya *file* tersebut ditampilkan seperti **Gambar 4.5** di bawah ini.

Dengan menjadikan predikat kelulusan sebagai *class*, maka dalam *visualize* menampilkan dua warna yang berbeda di setiap kategori predikat kelulusan.

Gambar 4.6 adalah *visualize* dari *class* predikat kelulusan.



Gambar 4.7 Data yang dialisis berdasarkan atribut *class* predikat kelulusan.



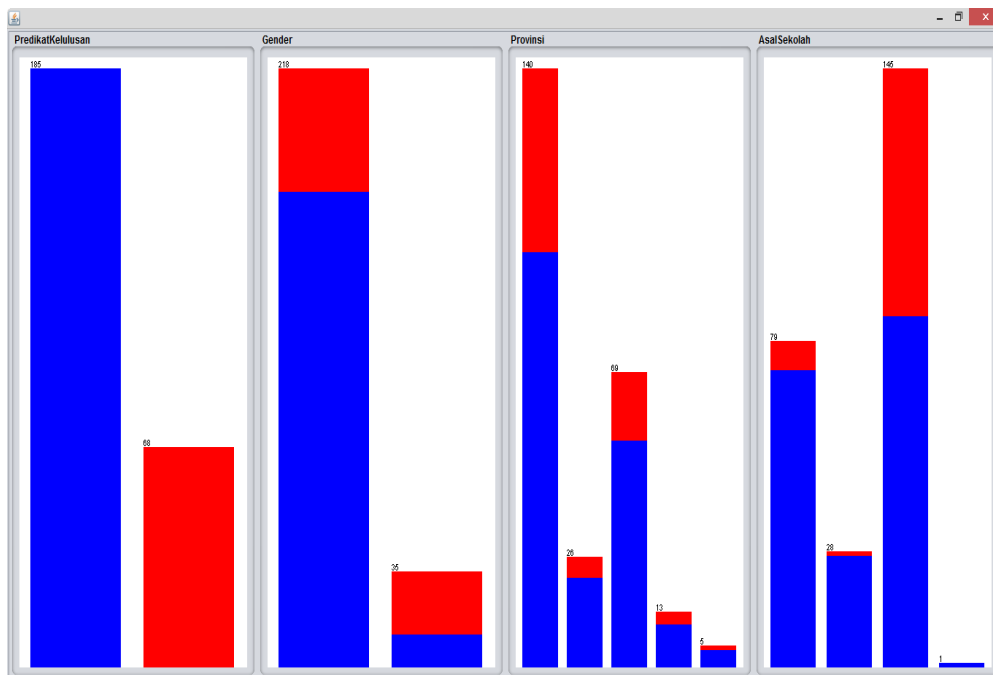
Gambar 4.8 Visualize data berdasarkan *class* predikat kelulusan.

Keterangan:



Berdasarkan grafik pada gambar 4.6 mahasiswa *cumlaude* memiliki *record* 68 data, dan *no cumlaude* memiliki *record* 185 data.

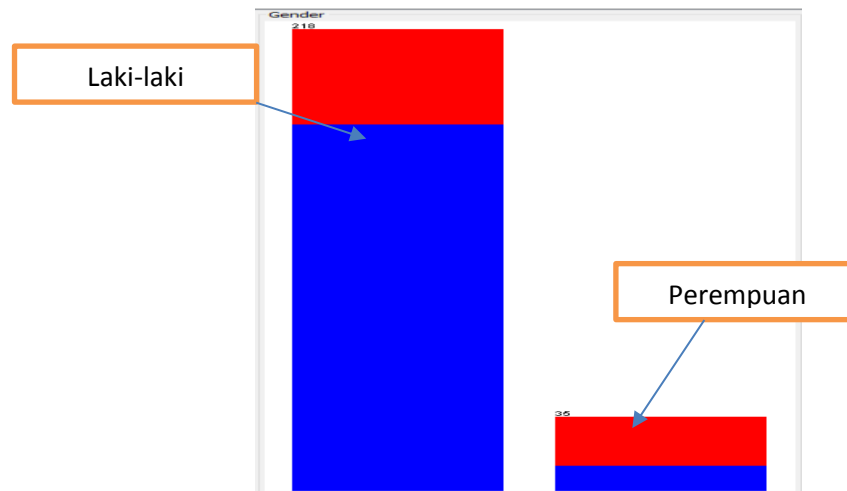
Atribut lainnya akan dianalisa berdasarkan dari *class* yang telah ditentukan yaitu *class* predikat kelulusan. Adapun hasil *visualize* keseluruhan atribut berdasarkan perangkat lunak Weka dalam bentuk grafik pada gambar 4.7 sebagai berikut.



Gambar 4.9 *Visualize* dari 4 atribut menggunakan Weka

Data *visualize*, data dapat dibaca dan ditentukan *knowledge* baru dari data tersebut. Untuk lebih jelasnya akan dijelaskan setiap *visualize* dari atribut yang digunakan.

a. *Gender*



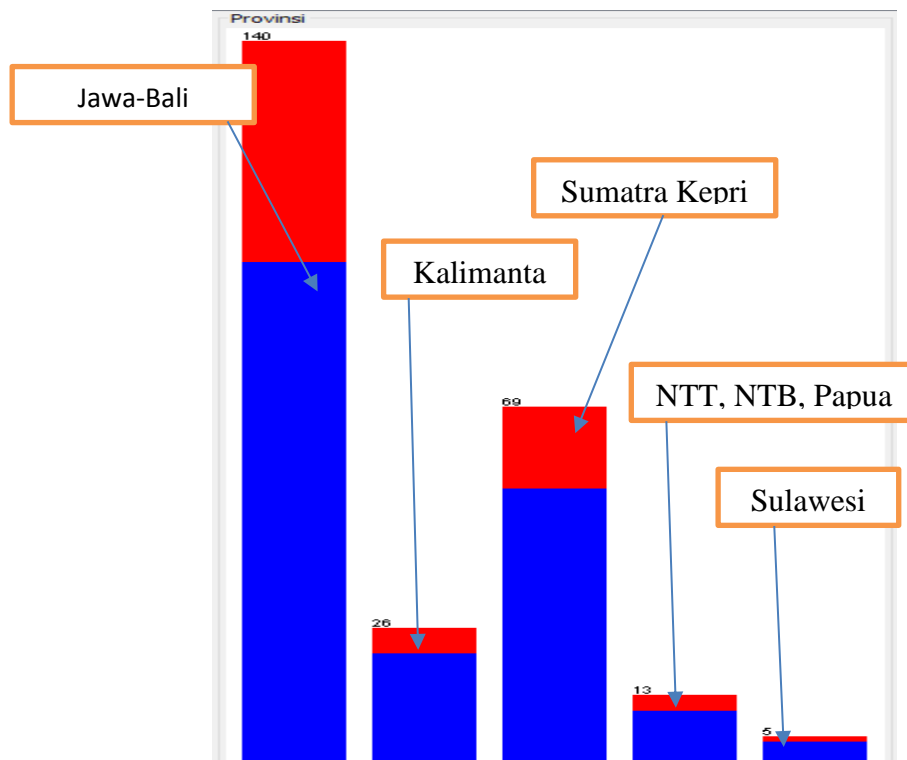
Gambar 4.10 Atribut *gender*

Keterangan:



Hasil pengamatan *visualize* pada atribut *Gender*, mahasiswa mayoritas berjenis kelamin laki-laki, dengan *record* 218 data, sementara perempuan terdapat *record* 35 data. Dalam hasil *visualize* itu juga kita dapat melihat bahwa mayoritas mahasiswa dengan predikat *cumlaude* merupakan laki-laki. Namun diantara jumlah mahasiswi perempuan yang hanya 35 tetapi mayoritas dari alumni perempuan meraih predikat *cumlaude*.

b. Provinsi



Gambar 4.11 Atribut provinsi

Keterangan:



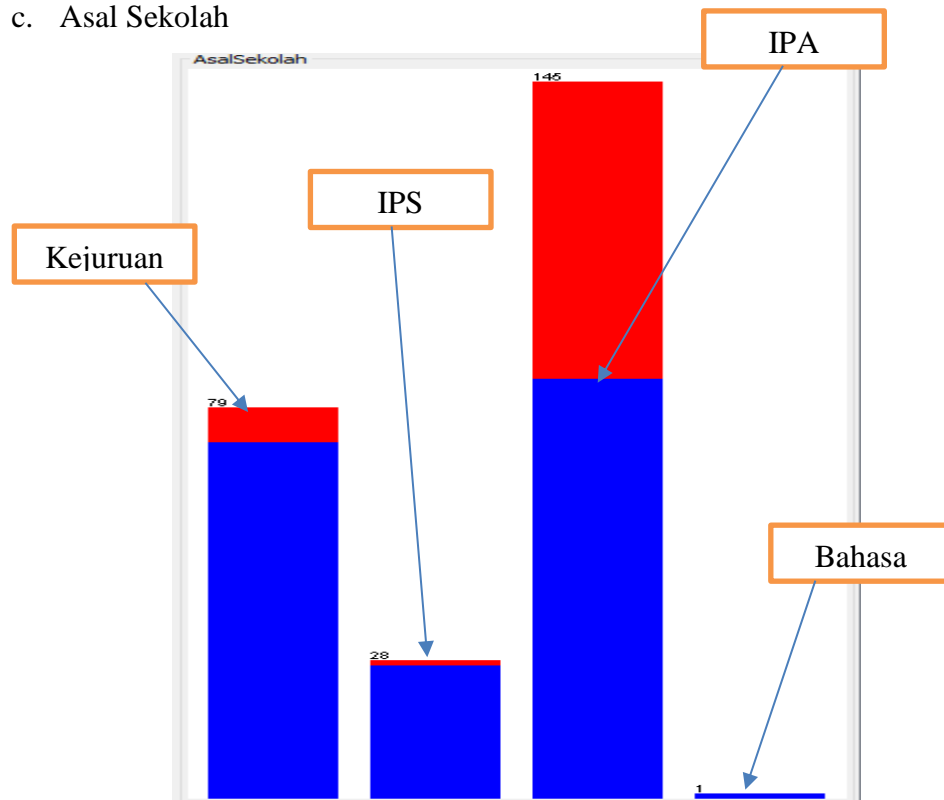
Dari hasil *visualize* untuk atribut asal provinsi, daerah yang paling banyak mahasiswanya berasal dari

1. Jawa Bali (140)
2. Sumatra Kepri (69)
3. Kalimantan (26)
4. NTT, NTB, Papua (13)
5. Sulawesi (5)

Dari seluruh mahasiswa yang berasal dari beberapa daerah, Jawa Bali merupakan daerah yang paling banyak meraih predikat *cumlaude*.

Kemudian untuk daerah Sulawesi selain paling sedikit meraih predikat *cumlaude*, namun juga sebagai daerah yang paling sedikit mahasiswanya.

c. Asal Sekolah



Gambar 4.12 Tampilan dari Asal Sekolah

Keterangan:



Berdasarkan hasil pengamatan tampilan pada atribut asal sekolah, banyak mahasiswa berlatar belakang jurusan IPA, berikut daftar jurusan asal mahasiswa ketika masih di bangku SMA.

1. IPA (145)
2. Kejuruan (79)

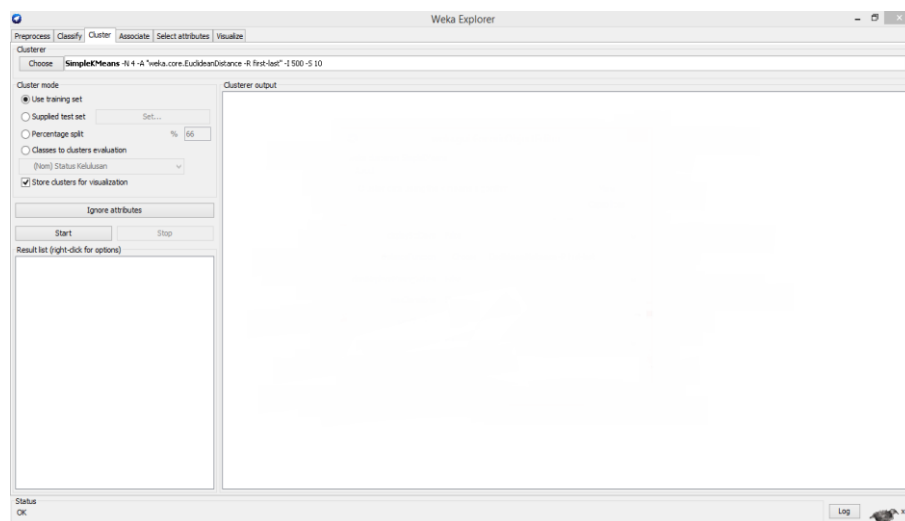
3. IPS (28)

4. Bahasa (1)

Jurusan IPA merupakan jurusan yang banyak meraih predikat *cumlaude* sementara jurusan Bahasa merupakan jurusan yang paling sedikit mahasiswanya dan tidak *cumlaude*.

Setelah menganalisa data dari tampilan pada perangkat lunak Weka, maka dilakukan metode *clustering* dengan menggunakan algoritma *K-Means*. Perangkat lunak Weka menyediakan beberapa metode dan terdapat algoritma dari setiap metode tersebut. Peneliti memasukkan jumlah *cluster* sesuai dengan jumlah kategori pada *class* predikat kelulusan yaitu 2 dan iterasi yang dilakukan adalah *default* dari *software* tersebut, sebagaimana dapat dilihat pada **Gambar 4.10**.

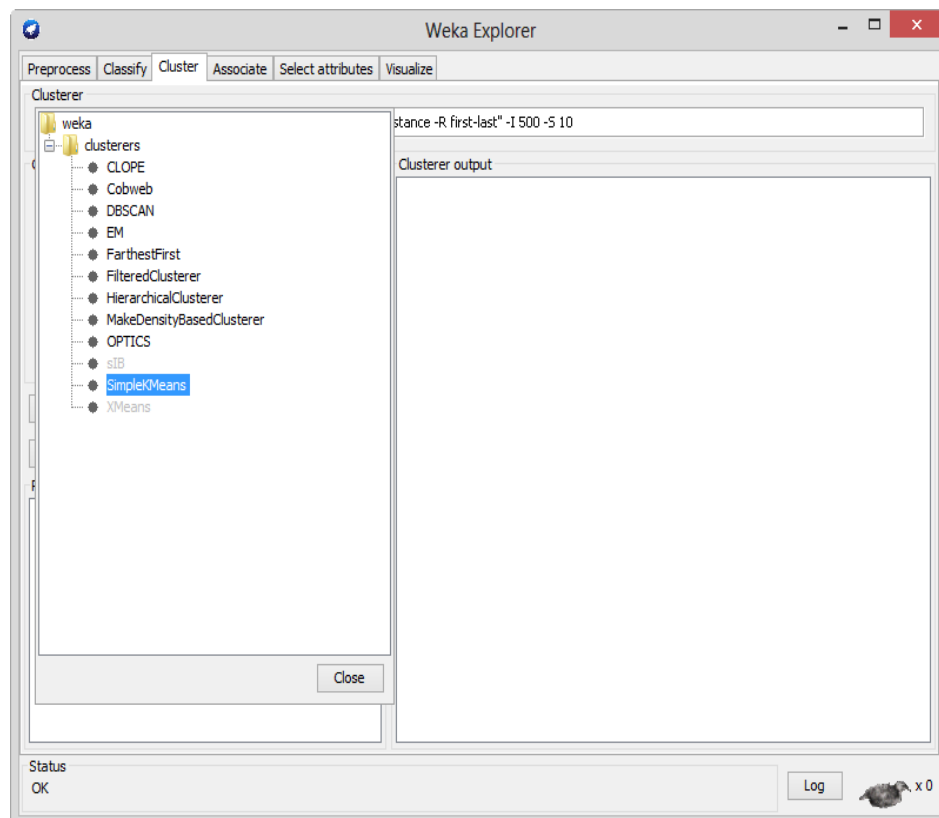
Gambar 4.11 merupakan tampilan *software* weka dengan menggunakan metode *clustering* dan pemilihan algoritma yang digunakan dapat dipilih dengan *click button Choose*.



Gambar 4.13 Tampilan weka metode *clustering*

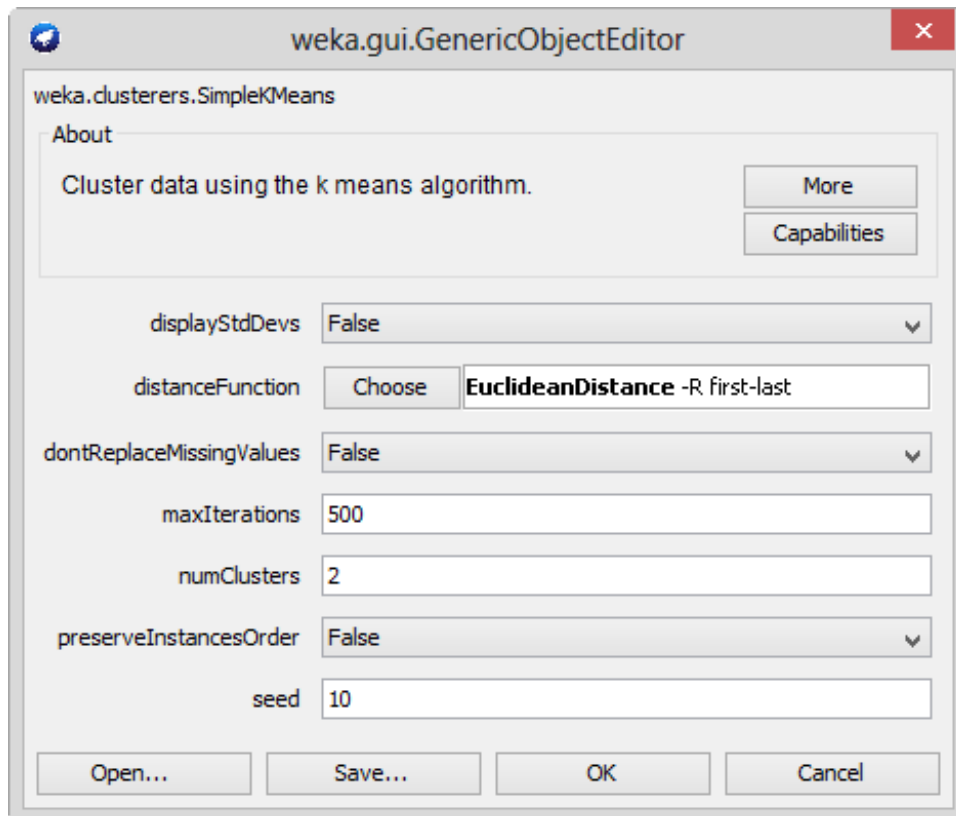
Pada form *Cluster* terdapat beberapa tipe algoritma *clusters*, penulis memilih *simple K-Means* sebagai algoritma yang ingin di pakai.

Sebagaimana terlihat pada **Gambar 4.14** di bawah ini.



Gambar 4.14 Pemilihan algoritma *simple K-Means* dalam weka

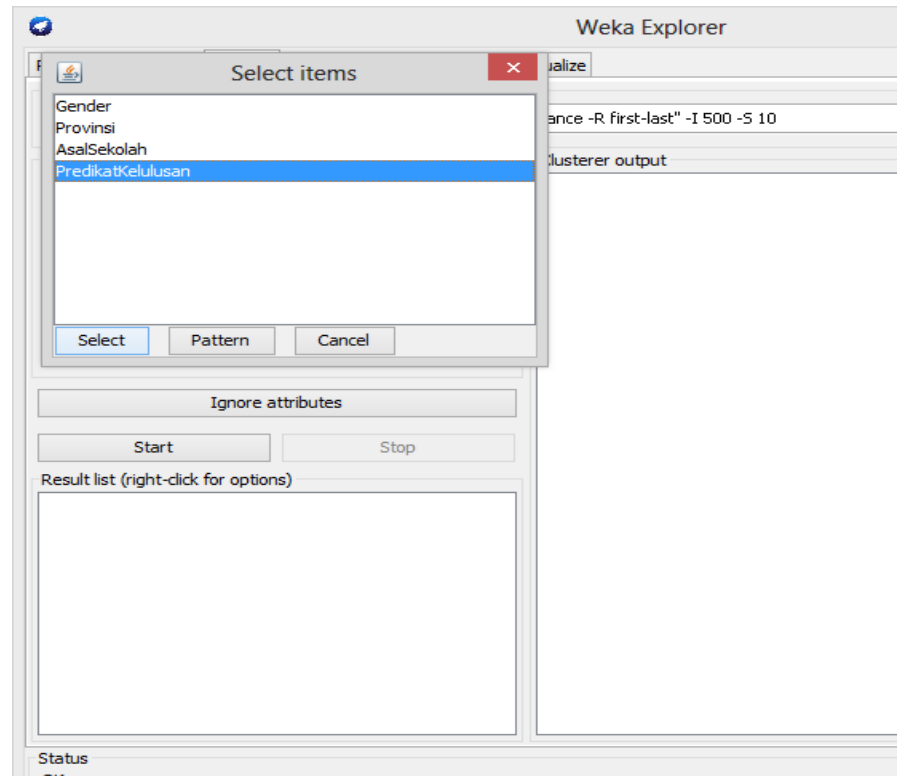
Sebelum metode ini dijalankan pada *form cluster*, maka data yang ada diatur terlebih dahulu sesuai *class* data yang diinginkan, *clustering* dengan algoritma *simple K-Means* akan di buat menjadi 2 *cluster* berdasarkan *comlaude* dan *non cumlaude*.



Gambar 4.15 Proses pengaturan *cluster* algoritma *K-Means*

Apa bila telah dilakukan prose pengaturan, maka selanjutnya adalah, atribut yang ingin dijadikan *class* perlu di centang sebagaimana contoh

Gambar 4.16.



Gambar 4.16 Proses mencentang *class* predikat kelulusan

Data yang sudah dapat dianalisa selanjutnya menggunakan algoritma *K-Means* untuk mencari jumlah data yang masuk dari 2 *cluster*. **Gambar 4.16** merupakan hasil dari algoritma *K-Means* dengan 2 *cluster* yang digunakan. Hasil *cluster centroid* dengan menggunakan weka seperti pada **Gambar 4.17** sebagai berikut.

```

Clusterer output

Cluster 0: Laki-laki,'Jawa Bali',IPS
Cluster 1: Perempuan,'Jawa Bali',KEJURUAN

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute          Full Data      Cluster#
                   (253.0)      (218.0)      (35.0)
=====
Gender             Laki-laki Laki-laki Perempuan
Provinsi           Jawa Bali Jawa Bali Jawa Bali
AsalSekolah       IPA        IPA        IPA

Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      218 ( 86%)
1      35 ( 14%)

```

Gambar 4.17 Cluster centroid pada weka

Dari hasil *cluster instances* menggunakan *software* weka data lulusan Fakultas Teknik adalah sebagai berikut.

1. Mahasiswa *no cumlaude* = Cluster 0 dengan *gender* laki-laki, asal Jawa Bali, jurusan IPA sebanyak 218 mahasiswa dari jumlah total 253 mahasiswa (86%).
2. Mahasiswa *cumlaude* = Cluster 1 dengan *gender* perempuan, asal Jawa Bali jurusan IPA sebanyak 35 dari jumlah 253 mahasiswa baru (14%).

Dari hasil di atas rata-rata mahasiswa *cumlaude* dan *non cumlaude* memiliki kesamaan, yaitu berasal dari Jawa Bali asal jurusan SMA adalah IPA, namun ada satu yang menjadi pembeda yaitu, *no cumlaude* dari *gender* laki-laki, dan *cumlaude* dari perempuan.

4.5.2 Perhitungan Algoritma *simple K-Means*

Dalam penelitian algoritma yang digunakan adalah algoritma *simple K-Means* dan untuk memudahkan penjelasan tentang algoritma *simple K-Means* dalam penelitian ini maka penulis membuat tabel data yang digunakan dalam penelitian. Data yang diambil merupakan data yang telah melewati proses filter, seperti **Table 4.5** berikut ini.

Tabel 4.5 *Sample data* alumni mahasiswa Fakultas Teknik UMY 2013-2015

No	Predikat Kelulusan	Gender	Provinsi	AsalSekolah
1	No Cumlaude	Laki-laki	Jawa Bali	KEJURUAN
2	No Cumlaude	Laki-laki	Kalimantan	KEJURUAN
3	No Cumlaude	Laki-laki	Sumtera Kepri	IPS
4	No Cumlaude	Laki-laki	Jawa Bali	KEJURUAN
5	No Cumlaude	Laki-laki	Jawa Bali	IPA
6	No Cumlaude	Laki-laki	Jawa Bali	IPA
7	No Cumlaude	Laki-laki	Kalimantan	IPS
8	No Cumlaude	Laki-laki	Nusa Tenggara, Maluku, Papua	IPS
9	No Cumlaude	Laki-laki	Sumtera Kepri	KEJURUAN
10	No Cumlaude	Laki-laki	Jawa Bali	IPA

Sebelum melakukan perhitungan secara manual menggunakan rumus

algoritma K-means

$$d(P, Q) = \sqrt{\sum_{j=1}^p (x_j(P) - x_j(Q))^2}$$

Maka sebaiknya semua data di inisiasi terlebih dahulu, seperti **Tabel 4.6** berikut ini:

Tabel 4.6 *Sample* hasil dari inisiasi setiap atribut

No	PredikatKelulusan	Gender	Provinsi	AsalSekolah
1	1	1	1	2
2	1	1	3	2
3	1	1	2	3
4	1	1	1	2
5	1	1	1	1
6	1	1	1	1
7	1	1	3	3
8	1	1	4	3

Dalam menggunakan algoritma *K-Means* akan melakukan pengulangan tahapan hingga terjadi kestabilan. Peneliti melakukan lebih dari satu kali iterasi dalam melakukan pengujian menggunakan algoritma *K-Means* dengan tahapan sebagai berikut.

1. Menentukan jumlah *cluster* dan menentukan koordinat titik tengah *cluster*.

Kelompok *cluster* yang dibuat adalah 4 kelompok berdasarkan jumlah atribut yang digunakan. Total jumlah dari setiap atribut tersebut untuk menentukan kelompok diambil dari frekuensi kurang, frekuensi sedang dan frekuensi padat seperti.

2. Penentuan nilai dari *cluster* untuk dijadikan acuan dalam melakukan perhitungan jarak objek ke *centroid*, perhitungan jarak mengacu pada rumus *Euclidean* yang dilakukan perhitungan di *excel*.
3. Pengelompokkan objek tersebut berdasarkan pada jarak yang dilakukan dengan menggunakan hasil pada perhitungan jarak yaitu pada proses langkah 2. Hasil pada perhitungan jarak tersebut digunakan untuk penentuan kelompok *clustering*.

Berikut merupakan inisialisasi k sebagai *centroid* yang dapat dibangkitkan secara random, penulis melakukan beberapa kali perhitungan dalam setiap perhitungan penulis memilih angka yang berbeda-beda.

Perhitungan 1

Cluster

<i>Cluster 1</i>	1.1	1.1	1.1	1.1
<i>Cluster 2</i>	1.5	1.5	4.5	3.5

Perhitungan 2

Cluster

<i>Cluster 1</i>	1	1	1	1
<i>Cluster 2</i>	2	2	5	4

Perhitungan 3

Cluster

<i>Cluster 1</i>	1.3	1.3	1.3	1.3
<i>Cluster 2</i>	1.6	1.6	4.6	3.6

Keterangan : *Cluster 1* merupakan nilai inisiasi terendah yang di tentukan secara random disetiap atribut

Cluster 2 merupakan nilai inisiasi tertinggi yang ditentukan secara random disetiap atribut

untuk perhitungan manual menggunakan rumus sebagai berikut

$$d(P, Q) = \sqrt{\sum_{j=1}^p (x_j(P) - x_j(Q))^2}$$

Menghitung perhitungan 1 untuk *cluster 1*

$$\begin{aligned} &= \sqrt{(1 - 1.1)^2 + (1 - 1.1)^2 + (1 - 1.1)^2 + (2 - 1.1)^2} \\ &= \sqrt{(-0.1)^2 + (-0.1)^2 + (-0.1)^2 + (-0.9)^2} \\ &= \sqrt{0.01 + 0.01 + 0.01 + 0.81} \\ &= \sqrt{0.84} \\ &= \mathbf{0,916} \end{aligned}$$

Menghitung perhitungan 1 untuk *cluster 2*

$$\begin{aligned} &= \sqrt{(1 - 1.5)^2 + (1 - 1.5)^2 + (1 - 4.5)^2 + (2 - 3.5)^2} \\ &= \sqrt{(0.5)^2 + (0.5)^2 + (-3.5)^2 + (1.5)^2} \\ &= \sqrt{0,25 + 0.25 + +12,25 + 2.25} \\ &= \sqrt{15} \end{aligned}$$

= 3,872

Tabel 4.7 Hasil perhitungan manual untuk iterasi pertama

No	Predikat Kelulusan	Gender	Provinsi	Asal Sekolah	Cluster 1				Cluster 2				
					1.1	1.1	1.1	1.1	1.5	1.5	4.5	3.5	
1	1	1	1	2	0.84				0.916515139	15			3.872983346
2	1	1	3	2	4.44				2.107130751	5			2.236067977
3	1	1	2	3	4.44				2.107130751	7			2.645751311
4	1	1	1	2	0.84				0.916515139	15			3.872983346
5	1	1	1	1	0.04				0.2	19			4.358898944
6	1	1	1	1	0.04				0.2	19			4.358898944
7	1	1	3	3	7.24				2.690724809	3			1.732050808
8	1	1	4	3	12.04				3.469870315	1			1

Tabel 4.8 Pengelompokan *centroid*

Centroid 1	Centroid 2
Yes	
Yes	
Yes	
Yes	
Yes	
Yes	
	Yes
	Yes
Yes	
227	26
89.72332016	10.27667984

Menghitung perhitungan 2 untuk *cluster* 1

$$\begin{aligned}
&= \sqrt{(1-1)^2 + (1-1)^2 + (1-1)^2 + (1-2)^2} \\
&= \sqrt{(0)^2 + (0)^2 + (0)^2 + (-1)^2} \\
&= \sqrt{0 + 0 + 0 + 1} \\
&= \sqrt{1} \\
&= \mathbf{1}
\end{aligned}$$

Menghitung perhitungan 2 untuk *cluster 2*

$$\begin{aligned}
&= \sqrt{(1-2)^2 + (1-2)^2 + (1-5)^2 + (2-4)^2} \\
&= \sqrt{(-1)^2 + (-1)^2 + (-4)^2 + (-2)^2} \\
&= \sqrt{1 + 1 + 16 + 4} \\
&= \sqrt{22} \\
&= \mathbf{4.690}
\end{aligned}$$

Tabel 4.9 Hasil perhitungan manual untuk iterasi ke dua

No	Predikat Kelulusan	Gender	Provinsi	Asal Sekolah	Cluster 1				Cluster 2					
					1	1	1	1	2	2	5	4		
1	1	1	1	2	1				1				22	4.69041576
2	1	1	3	2	5				2.236067977				10	3.16227766
3	1	1	2	3	5				2.236067977				12	3.464101615
4	1	1	1	2	1				1				22	4.69041576
5	1	1	1	1	0				0				27	5.196152423
6	1	1	1	1	0				0				27	5.196152423
7	1	1	3	3	8				2.828427125				7	2.645751311
8	1	1	4	3	13				3.605551275				4	2

Tabel 4.10 Pengelompokan *centroid*

Centroid 1	Centroid 2
Yes	
Yes	
Yes	
Yes	
Yes	
Yes	
	Yes
	Yes
Yes	
Yes	
Yes	
Yes	
Yes	
235	18
92.88537549	7.114624506

Menghitung perhitungan 3 untuk *cluster* 1

$$= \sqrt{(1 - 1.3)^2 + (1 - 1.3)^2 + (1 - 1.3)^2 + (2 - 1.3)^2}$$

$$= \sqrt{(-0.3)^2 + (-0.3)^2 + (-0.3)^2 + (0.7)^2}$$

$$= \sqrt{0.09 + 0.09 + 0.03 + 0.49}$$

$$=\sqrt{0.76}$$

$$= \mathbf{0.871}$$

Menghitung perhitungan 3 untuk *cluster 2*

$$=\sqrt{(1 - 1.6)^2 + (1 - 1.6)^2 + (1 - 4.6)^2 + (2 - 3.6)^2}$$

$$=\sqrt{(-0.6)^2 + (-0.6)^2 + (-3.6)^2 + (-1.6)^2}$$

$$=\sqrt{0.36 + 0.36 + 12.96 + 2.56}$$

$$=\sqrt{16.24}$$

$$= \mathbf{4.029}$$

Tabel 4.11 Hasil perhitungan manual iterasi ke tiga

No	Predikat Kelulusan	Gender	Provinsi	Asal Sekolah	Cluster 1				Cluster 2				
					1.3	1.3	1.3	1.3	1.6	1.6	4.6	3.6	
1	1	1	1	2	0.76				0.871779789	16.24			4.029888336
2	1	1	3	2	3.56				1.886796226	5.84			2.416609195
3	1	1	2	3	3.56				1.886796226	7.84			2.8
4	1	1	1	2	0.76				0.871779789	16.24			4.029888336
5	1	1	1	1	0.36				0.6	20.44			4.521061822
6	1	1	1	1	0.36				0.6	20.44			4.521061822
7	1	1	3	3	5.96				2.441311123	3.64			1.907878403
8	1	1	4	3	10.36				3.218695388	1.44			1.2

Table 4.12 Pengelompokan *centroid*

Centroid 1	Centroid 2
Yes	
Yes	
Yes	
Yes	
Yes	
Yes	
	Yes
	Yes
Yes	
Yes	
Yes	
Yes	
235	18
92.88537549	7.114624506

Setelah semua data telah dihitung maka akan melihat jarak antara *centroid* di setiap data masing-masing,

Perhitungan 1 merupakan angka yang paling mendekati hasil perhitungan yang dilakukan dengan menggunakan *software* Weka. Perhitungan Weka penulis mendapatkan hasil 86% untuk non *cumlaude* dan 14% untuk *cumlaude*, sementara dalam perhitungan manual penulis mendapatkan hasil 89.72% untuk non *cumlaude*

dan 10.276% untuk *cumlaude*, penulis melakukan beberapa kali perhitungan dengan kombinasi nilai yang berbeda-beda.