

BAB IV

HASIL DAN PEMBAHASAN

Pada penelitian ini data calon mahasiswa baru terdiri dari dua fakultas, yaitu Fakultas Kedokteran dan Ilmu Keperawatan (FKIK) dan Fakultas Ilmu Sosial dan Ilmu Politik (FISIPOL). Kedua fakultas ini memiliki peminat yang sangat banyak dan meningkat setiap tahunnya. Selain dari minat calon mahasiswa baru, faktor lain peneliti mengambil kedua fakultas ini berdasarkan jurusan pada masa Sekolah Menengah Atas (SMA). Fakultas Kedokteran dan Ilmu Keperawatan memperdalam Ilmu Pengetahuan Alam (IPA) dan Fakultas Ilmu Sosial dan Ilmu Politik mendalami Ilmu Pengetahuan Sosial (IPS). Kampus UMY memiliki peminat yang sangat banyak pada kedua fakultas ini, sehingga peneliti fokus pada pengambilan kasus di FKIK dan FISIPOL.

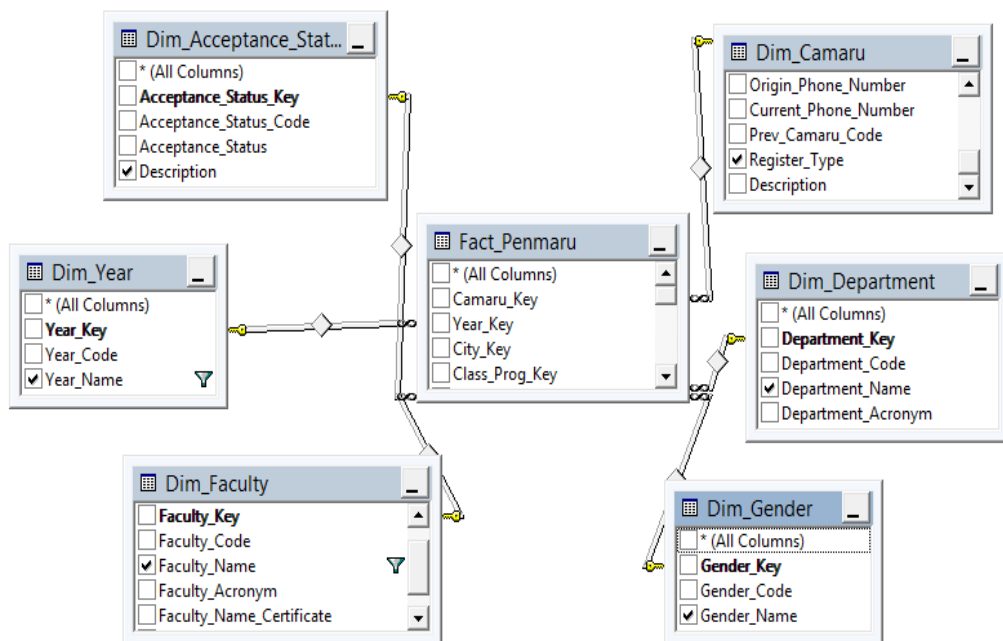
4.1 Pengumpulan Data

Sumber data utama yang digunakan dalam penelitian ini adalah *data warehouse* calon mahasiswa baru (penmaru) UMY pada FKIK dan FISIPOL tahun 2010, 2011, 2012, 2013 dan 2014. 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).

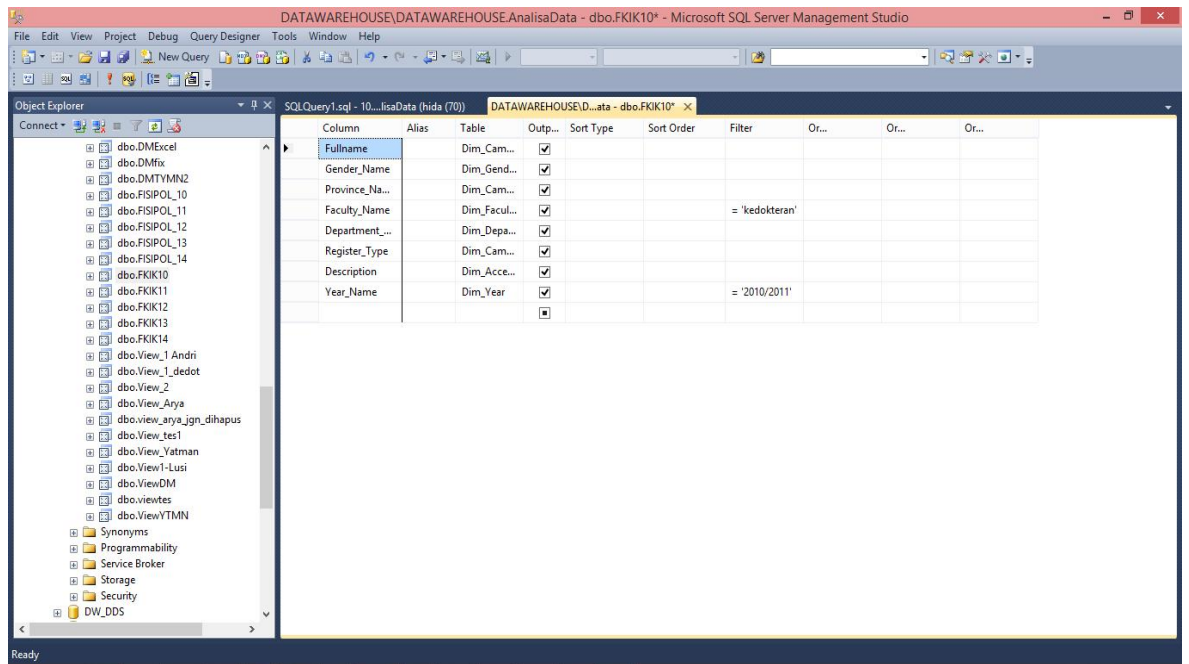
4.2 Penyeleksian Data

Data penmaru disaring terlebih dahulu dan diambil beberapa atribut dari tabel-tabel dimensi dan tabel *factual* untuk dianalisis. Pengambilan data dari *SQL Server Management* diolah dalam *SQL Server Data Tools for Visual Studio 2013* untuk dibuat

dalam *dataset* yang baru. Jumlah data calon mahasiswa baru FKIK dari tahun 2010 sampai 2014 adalah 25000 *record*, dan FISIPOL data calon mahasiswa baru berjumlah 17187 *record*. Terdapat lima atribut yang digunakan, yaitu kota asal, jenis kelamin, status lolos seleksi, jalur masuk dan jurusan. Pengambilan data berdasarkan atribut yang digunakan seperti pada gambar 4.1 dan penyeleksian data dari data asli (*data warehouse*) seperti pada gambar 4.2 sebagai berikut.



Gambar 4. 1 Tampilan *view* dari *data warehouse*.



Gambar 4. 2 Filter data dari data warehouse.

4.3 Preprocessing Data

Setelah data diseleksi sesuai dengan atribut yang akan digunakan maka dilakukan *preprocessing* data, agar tidak ada duplikasi data, tidak *missing value* dan memperbaiki kesalahan-kesalahan yang ada pada *dataset* baru dalam bentuk format *excel*. Pada tahap inilah data akan dilakukan *cleaning* atau pembersihan data, sehingga data tersebut dapat diolah dan dilakukan proses *data mining*. Data yang telah melewati tahap *preprocessing* akan disimpan dalam *dataset* yang baru menggunakan *Microsoft Office Excel* dengan format *.csv*. Perangkat lunak *WEKA* dapat membuka file dengan format *.csv*.

Pada data penmaru UMY Fakultas Ilmu Sosial dan Ilmu Politik terdapat *unknown* data yaitu pada atribut kota asal sebanyak 3264 data. Sedangkan pada Fakultas Kedokteran dan Ilmu Keperawatan *unknown* data pada atribut kota asal sebanyak 717 data.

4.4 Transformasi Data

Pada tahap ini dilakukan proses perubahan data, agar data dapat diolah dengan menggunakan algoritma *k-means clustering*. Data yang *non-numeric* dilakukan proses inisiasi ke dalam bentuk numerik. Pada penelitian ini seluruh atribut diubah ke bentuk numerik, yaitu kota asal, jenis kelamin, status lolos seleksi, jalur masuk dan jurusan. proses inisiasinya sebagai berikut:

a. Kota Asal

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

Tabel 4. 1 Data pada atribut kota asal pada FKIK.

Kota Asal	Frekuensi	Inisial
Jawa Tengah	6156	1
Daerah Istimewa Yogyakarta	4184	2
Jawa Barat	2533	3
Jawa Timur	1485	4
Kalimantan Timur	1134	5
Lampung	1073	6
Nusa Tenggara Barat (NTB)	981	7
Riau	730	8
No Name	717	9
Kalimantan Barat	594	10
Sumatera Selatan	561	11
Bengkulu	559	12

Kota Asal	Frekuensi	Inisial
Banten	523	13
Kalimantan Tengah	433	14
Kalimantan Selatan	421	15
Bangka Belitung	419	16
Jambi	335	17
DKI Jakarta	269	18
Kepulauan Riau	263	19
Maluku Utara	209	20
Papua	182	21
Sulawesi Tengah	165	22
Sulawesi Selatan	159	23
Sumatera Barat	150	24
Sumatera Utara	124	25
Maluku	103	26
Papua Barat	96	27
Nanggroe Aceh Darussalam	89	28
Nusa Tenggara Timur (NTT)	84	29
Sulawesi Tenggara	76	30
Bali	76	31
Sulawesi Utara	41	32
Gorontalo	39	33
Kalimantan Utara	23	34
Luar Negeri	7	35
Sulawesi Barat	7	36

Tabel 4. 2 Data pada atribut kota asal pada FISIPOL.

Kota Asal	Frekuensi	Inisial
Daerah Istimewa Yogyakarta	3319	1
No Name	3264	2
Jawa Tengah	2829	3
Jawa Barat	1310	4
Nusa Tenggara Barat (NTB)	684	5
Lampung	562	6
Jawa Timur	538	7
Kalimantan Timur	431	8
Bengkulu	428	9
Kalimantan Barat	341	10
Sumatera Selatan	312	11
Riau	289	12
Jambi	280	13
Kalimantan Tengah	251	14
Bangka Belitung	234	15
Banten	225	16
Nanggroe Aceh Darussalam	222	17
Kalimantan Selatan	199	18
Maluku Utara	185	19
Kepulauan Riau	173	20
Sulawesi Selatan	172	21
Sumatera Barat	160	22

Kota Asal	Frekuensi	Inisial
Sulawesi Tengah	103	23
DKI Jakarta	101	24
Sumatera Utara	79	25
Papua	68	26
Maluku	61	27
Sulawesi Tenggara	60	28
Bali	59	29
Sulawesi Utara	54	30
Papua Barat	43	31
Gorontalo	42	32
Luar Negeri	36	33
Sulawesi Barat	33	34
Nusa Tenggara Timur (NTT)	21	35
Kalimantan Utara	19	36

b. Jenis Kelamin

Inisiasi jenis kelamin perempuan diinisialkan 1 dan untuk laki-laki diinisialkan 2.

c. Status Lolos Seleksi

Atribut status lolos seleksi diinisiasi berdasarkan jumlah frekuensi dari terbesar hingga terkecil.

Tabel 4. 3 Data pada atribut status lolos seleksi pada FKIK.

Status Lolos Seleksi	Frekuensi	Inisial
Tidak Lolos	22172	1
Lolos Pilihan 1	2243	2
Lolos Pilihan 2	585	3

Tabel 4. 4 Data pada atribut status lolos seleksi pada FISIPOL.

Status Lolos Seleksi	Frekuensi	Inisial
Lolos Pilihan 1	9038	1
Tidak Lolos	6831	2
Lolos Pilihan 2	1318	3

d. Jalur Masuk

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

Tabel 4. 5 Data pada atribut jalur masuk pada FKIK.

Jalur Masuk	Frekuensi	Inisial
CBT	9330	1
RT	9212	2
PSB	3911	3
PMDK	1960	4
OT	275	5
PNUAN	162	6
KHSS	127	7
PBUP	23	8

Tabel 4. 6 Data pada atribut jalur masuk pada FISIPOL.

Jalur Masuk	Frekuensi	Inisial
PSB	5798	1
CBT	5208	2
PMDK	3457	3
RT	1761	4
OT	778	5
KHSS	114	6
PBUP	71	7

e. Jurusan

Pada atribut jurusan dilakukan inisiasi data berdasarkan frekuensi terbesar hingga terkecil.

Tabel 4. 7 Data pada atribut jurusan pada FKIK.

Jurusan	Frekuensi	Inisial
Pendidikan Dokter	12447	1
Kedokteran Gigi	6963	2
Farmasi	3117	3
Ilmu keperawatan	2473	4

Tabel 4. 8 Data pada atribut jurusan pada FISIPOL.

Jurusan	Frekuensi	Inisial
Ilmu Hubungan Internasional	7500	1
Ilmu Komunikasi	5269	2
Ilmu Pemerintahan	4418	3

4.5 Pengolahan Data

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 pengelompokkan data yang memiliki kesamaan karakteristik pada setiap data.

Atribut pada data penmaru terdiri dari data *non-numeric*, data tersebut dapat dianalisis dengan menggunakan perangkat lunak *WEKA*. Data penmaru untuk FKIK

memiliki 25000 *record* data. Gambar 4.3 adalah data penmaru FKIK dengan format .CSV yang diakses melalui *software* WEKA. Sedangkan data penmaru FISIPOL memiliki 17187 *record* data. Gambar 4.4 adalah data penmaru FISIPOL dengan format .CSV.

FKIK-FIX.csv - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K
1	Jenis Kelamin	Kota Asal	Jurusan	Jalur Masuk	Status Kelulusan						
2	Laki-Laki	Nusa Tenggara Barat (NTB)	Ilmu Keperawatan	RT	Lolos Pilihan 1						
3	Laki-Laki	Nusa Tenggara Barat (NTB)	Farmasi	RT	Tidak Lolos						
4	Laki-Laki	Nusa Tenggara Barat (NTB)	Ilmu Keperawatan	CBT	Tidak Lolos						
5	Laki-Laki	Nusa Tenggara Barat (NTB)	Ilmu Keperawatan	RT	Lolos Pilihan 1						
6	Perempuan	Jawa Tengah	Pendidikan Dokter	CBT	Tidak Lolos						
7	Perempuan	Jawa Tengah	Pendidikan Dokter	RT	Tidak Lolos						
8	Perempuan	Jawa Tengah	Pendidikan Dokter	CBT	Tidak Lolos						
9	Perempuan	Jawa Tengah	Pendidikan Dokter	RT	Tidak Lolos						
10	Laki-Laki	Nusa Tenggara Barat (NTB)	Pendidikan Dokter	PSB	Tidak Lolos						
11	Laki-Laki	Nusa Tenggara Barat (NTB)	Pendidikan Dokter	RT	Tidak Lolos						
12	Laki-Laki	Nusa Tenggara Barat (NTB)	Farmasi	CBT	Tidak Lolos						
13	Laki-Laki	Nusa Tenggara Barat (NTB)	Kedokteran Gigi	CBT	Tidak Lolos						
14	Laki-Laki	Nusa Tenggara Barat (NTB)	Pendidikan Dokter	CBT	Tidak Lolos						
15	Laki-Laki	Jawa Timur	Pendidikan Dokter	CBT	Tidak Lolos						
16	Laki-Laki	Banten	Pendidikan Dokter	CBT	Tidak Lolos						
17	Laki-Laki	Jawa Tengah	Kedokteran Gigi	CBT	Tidak Lolos						
18	Laki-Laki	Jawa Tengah	Kedokteran Gigi	RT	Lolos Pilihan 2						
19	Laki-Laki	Jawa Tengah	Kedokteran Gigi	RT	Tidak Lolos						
20	Laki-Laki	Jawa Tengah	Pendidikan Dokter	CBT	Tidak Lolos						
21	Laki-Laki	Jawa Tengah	Pendidikan Dokter	RT	Tidak Lolos						
22	Laki-Laki	Nusa Tenggara Barat (NTB)	Pendidikan Dokter	PSB	Tidak Lolos						
23	Laki-Laki	Jawa Timur	Pendidikan Dokter	CBT	Tidak Lolos						
24	Perempuan	Jawa Barat	Pendidikan Dokter	PSB	Tidak Lolos						
25	Laki-Laki	Jawa Tengah	Farmasi	RT	Tidak Lolos						
26	Laki-Laki	Jawa Tengah	Ilmu Keperawatan	RT	Lolos Pilihan 1						
27	Laki-Laki	Banten	Kedokteran Gigi	CBT	Tidak Lolos						
28	Laki-Laki	Banten	Kedokteran Gigi	RT	Tidak Lolos						

Gambar 4. 3 Data penmaru FKIK format .CSV.

FISIPOL-FIX.csv - Microsoft Excel

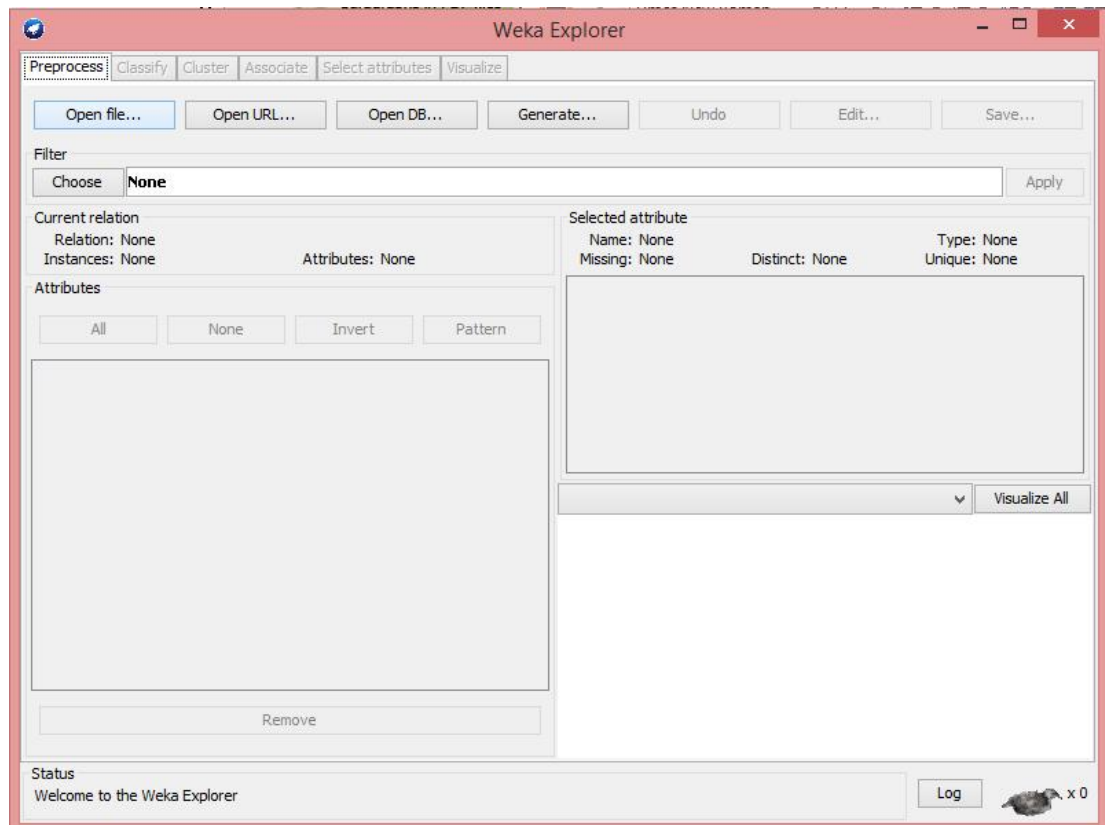
	A	B	C	D	E	F	G	H	I	J
1	Kota Asal, Jenis Kelamin, Jalur Masuk, Status Kelulusan, Jurusan									
2	Daerah Istimewa Yogyakarta, Laki-Laki, RT, Lolos Pilihan 1, IKom									
3	Jawa Tengah, Perempuan, RT, Lolos Pilihan 1, HI									
4	Jawa Tengah, Perempuan, RT, Tidak Lolos, HI									
5	Jawa Timur, Laki-Laki, RT, Lolos Pilihan 1, IKom									
6	Jawa Tengah, Perempuan, RT, Lolos Pilihan 1, HI									
7	Daerah Istimewa Yogyakarta, Perempuan, RT, Lolos Pilihan 1, HI									
8	Jawa Tengah, Laki-Laki, RT, Tidak Lolos, IKom									
9	Daerah Istimewa Yogyakarta, Perempuan, RT, Lolos Pilihan 1, IKom									
10	Daerah Istimewa Yogyakarta, Perempuan, RT, Tidak Lolos, HI									
11	Jawa Tengah, Perempuan, RT, Lolos Pilihan 1, IP									
12	Jawa Tengah, Perempuan, RT, Tidak Lolos, HI									
13	Jawa Tengah, Laki-Laki, RT, Lolos Pilihan 1, IP									
14	Daerah Istimewa Yogyakarta, Perempuan, RT, Tidak Lolos, IKom									
15	Daerah Istimewa Yogyakarta, Perempuan, RT, Lolos Pilihan 1, HI									
16	Papua Barat, Laki-Laki, RT, Lolos Pilihan 1, HI									
17	Daerah Istimewa Yogyakarta, Perempuan, RT, Lolos Pilihan 1, HI									
18	Daerah Istimewa Yogyakarta, Perempuan, RT, Tidak Lolos, IKom									
19	Jawa Tengah, Laki-Laki, RT, Lolos Pilihan 1, IP									
20	Daerah Istimewa Yogyakarta, Perempuan, RT, Lolos Pilihan 1, HI									
21	Kalimantan Selatan, Perempuan, RT, Tidak Lolos, IKom									
22	Daerah Istimewa Yogyakarta, Laki-Laki, RT, Tidak Lolos, HI									
23	Maluku Utara, Laki-Laki, RT, Lolos Pilihan 1, HI									
24	Maluku Utara, Laki-Laki, RT, Tidak Lolos, IP									
25	Jawa Timur, Laki-Laki, RT, Lolos Pilihan 1, HI									
26	Jawa Tengah, Laki-Laki, RT, Lolos Pilihan 1, HI									
27	Maluku Utara, Laki-Laki, RT, Tidak Lolos, IP									
28	Sulawesi Selatan, Laki-Laki, RT, Tidak Lolos, HI									

Gambar 4. 4 Data penmaru FISIPOL format .CSV.

4.5.1 Pengujian dengan *Software WEKA*

Atribut yang digunakan sebagai *class* adalah jurusan. Data akan dianalisa berdasarkan jurusan yang diminati atau yang dipilih calon mahasiswa baru dari Fakultas Kedokteran dan Ilmu Keperawatan dan Fakultas Ilmu Sosial dan Ilmu Politik.

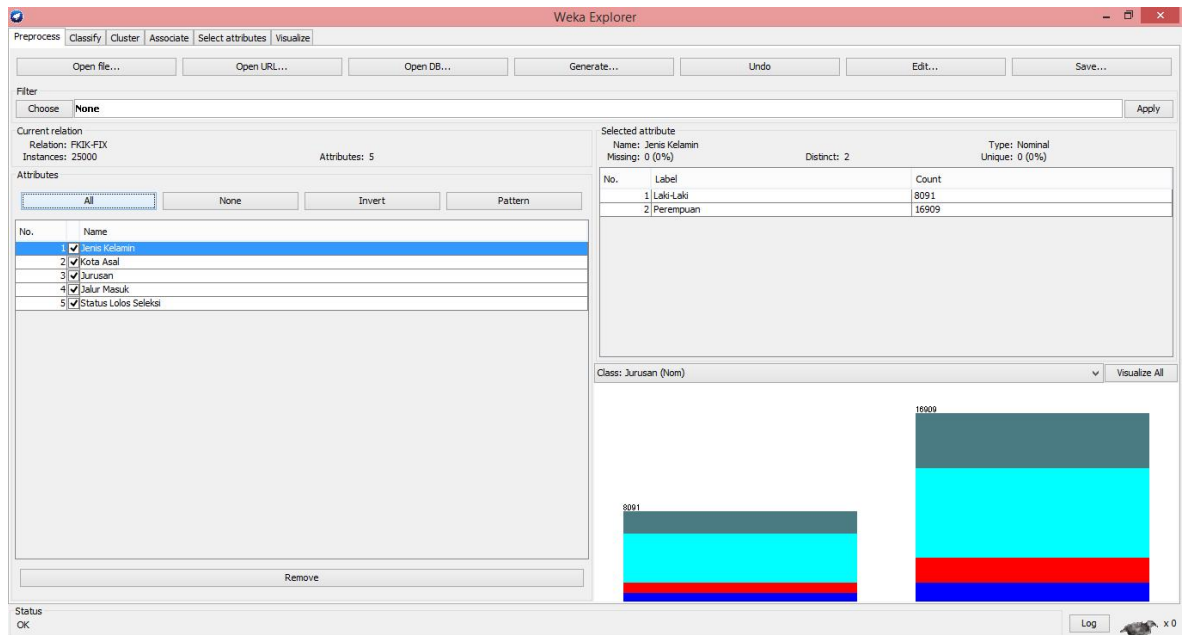
Data penmaru Fakultas Kedokteran dan Ilmu Keperawatan dibuat dalam *dataset* baru dengan format *.xlsx (excel) convert* ke format *.CSV* agar dapat dibuka menggunakan *software WEKA* seperti pada gambar 4.5 sebagai berikut.



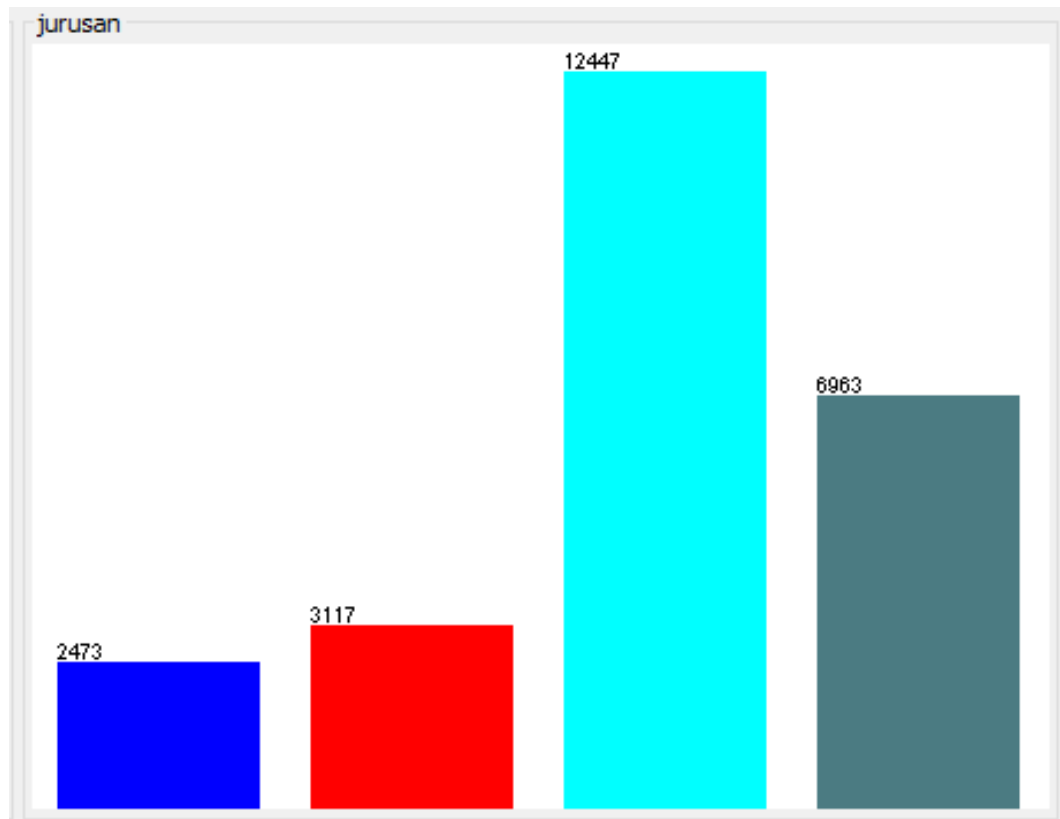
Gambar 4.5 *WEKA explorer*.

Setelah *open file* data penmaru yang akan dianalisis, maka selanjutnya *file* tersebut ditampilkan seperti pada gambar 4.6 di bawah ini.

Dengan menjadikan jurusan sebagai *class*, maka dalam *visualize* menampilkan empat warna yang berbeda-beda sebagai pembeda setiap jurusan di Fakultas Kedokteran dan Ilmu Keperawatan. Gambar 4.7 adalah *visualize* dari *class* jurusan.

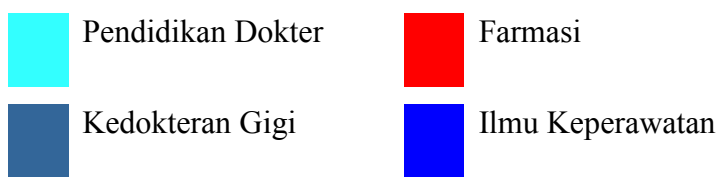


Gambar 4. 6 Data yang dianalisa berdasarkan atribut di *WEKA* pada FKIK.



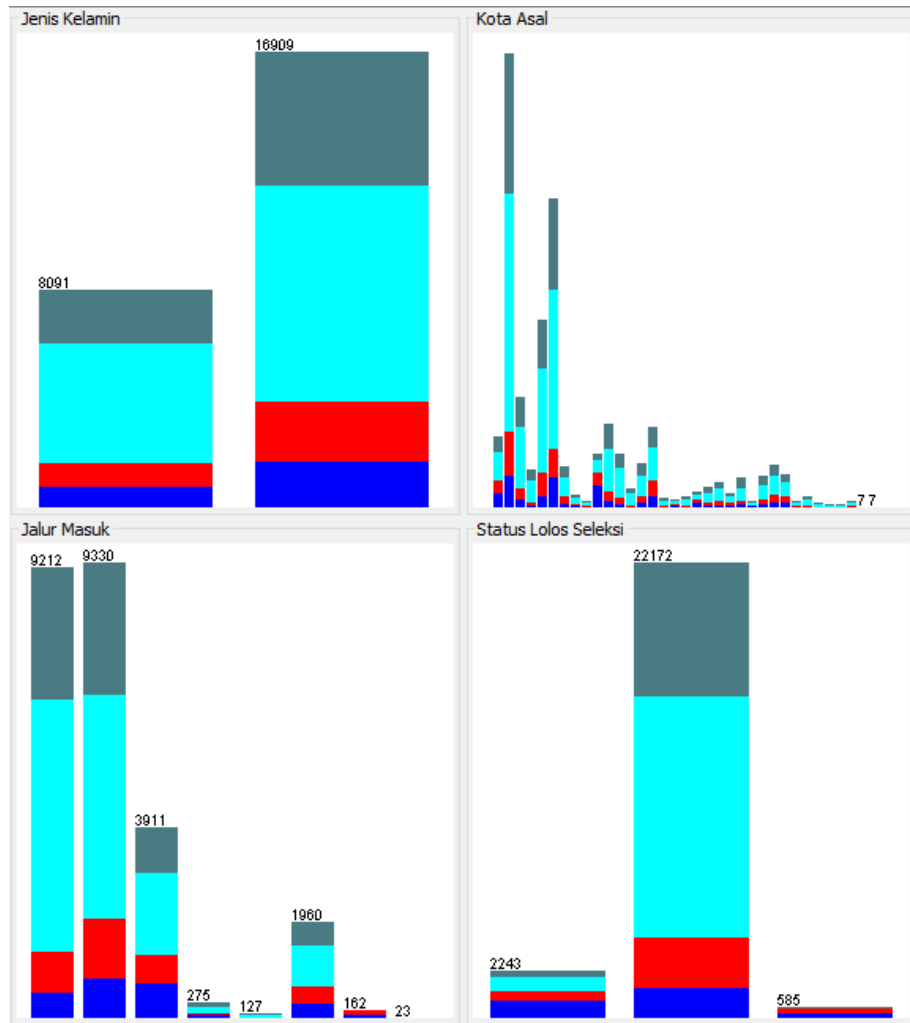
Gambar 4. 7 Visualize data berdasarkan *class* jurusan FKIK.

Keterangan :



FKIK terdiri dari empat jurusan yaitu Pendidikan Dokter, Kedokteran Gigi, Farmasi dan Ilmu Keperawatan. Dari grafik pada gambar 4.7 Pendidikan dokter memiliki 12447 *record* data, Kedokteran Gigi memiliki 6963 *record* data, jurusan Farmasi memiliki 3117 *record* dan jurusan Ilmu Keperawatan memiliki 2473 *record* data.

Atribut lainnya akan dianalisis berdasarkan dari *class* yang telah ditentukan yaitu *class* jurusan. Adapun hasil *visualize* keseluruhan atribut berdasarkan perangkat lunak *WEKA* dalam bentuk grafik pada gambar 4.8 adalah sebagai berikut.

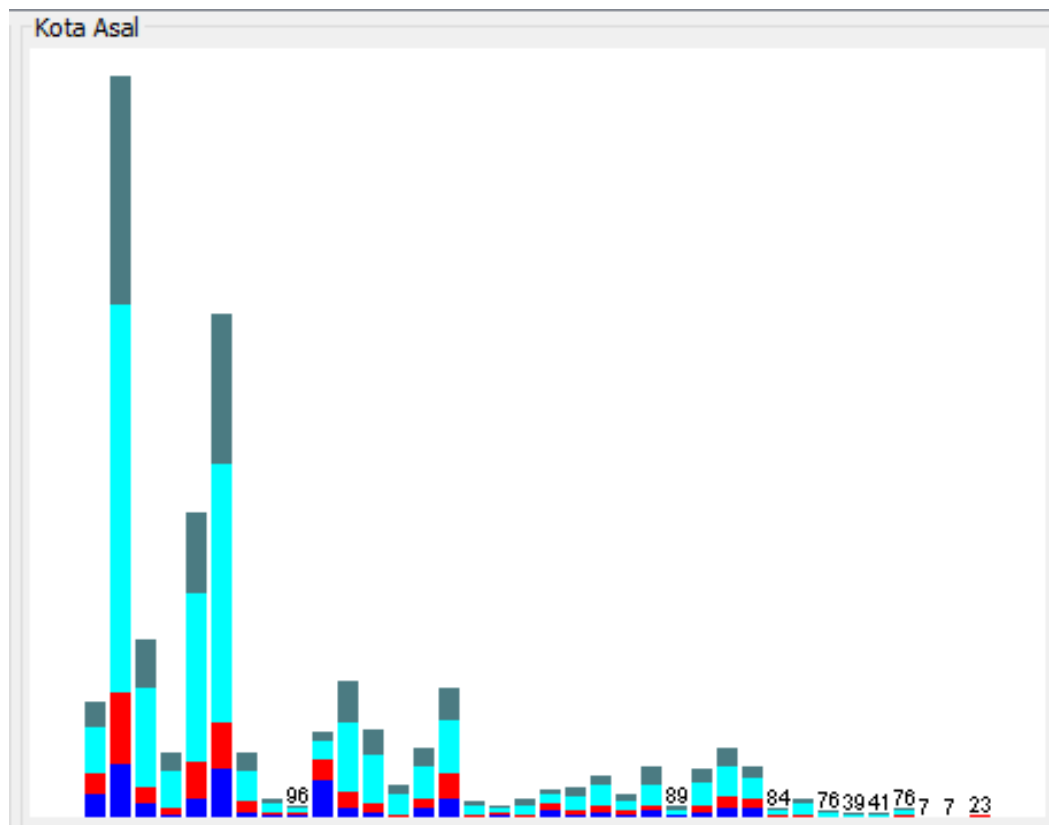


Gambar 4. 8 *Visualize* dari 4 atribut menggunakan *WEKA* pada FKIK.

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

a. Kota Asal

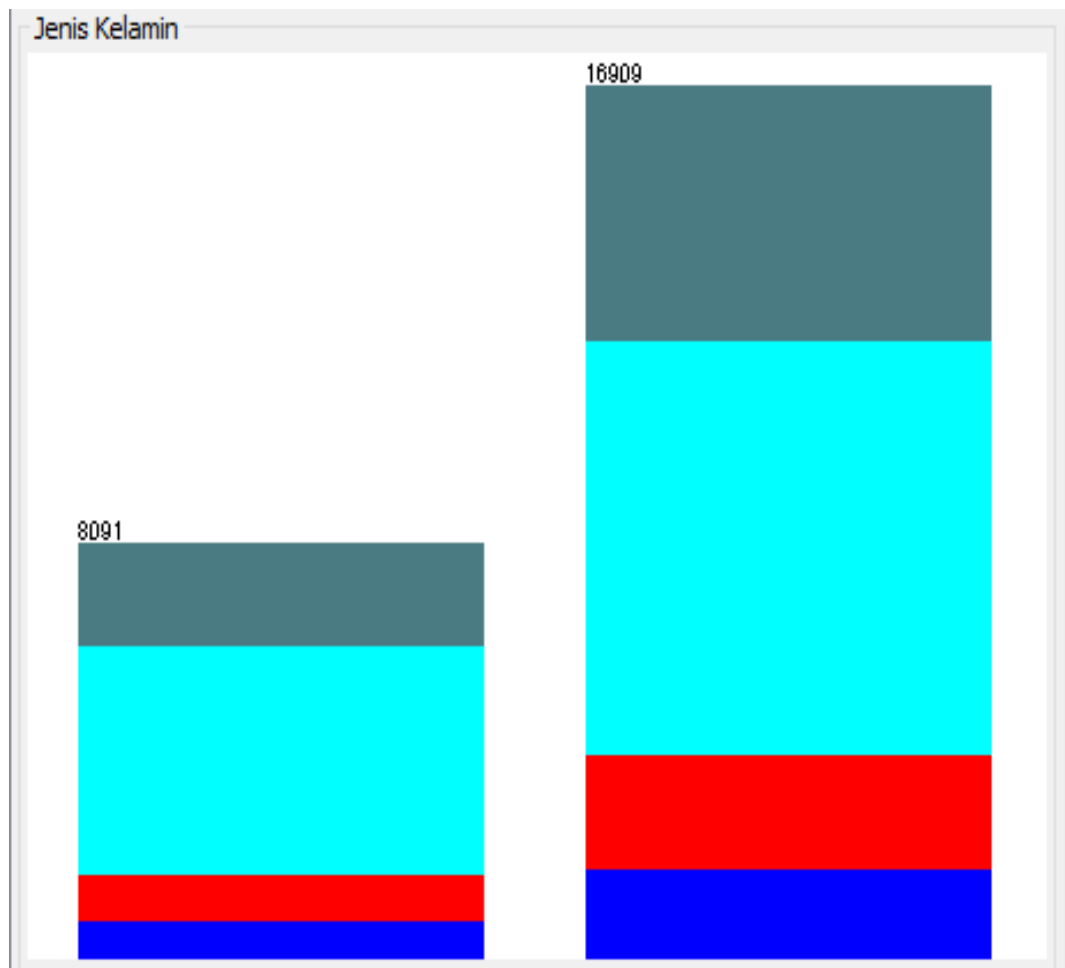
Hasil pengamatan *visualize* pada atribut kota asal, daerah yang paling banyak calon mahasiswanya berasal dari provinsi Jawa Tengah 6156 *record* data. *Visualize* kota asal dapat dilihat pada gambar 4.9 di bawah ini.



Gambar 4.9 Atribut kota asal dari *visualize WEKA* pada FKIK.

b. Jenis Kelamin

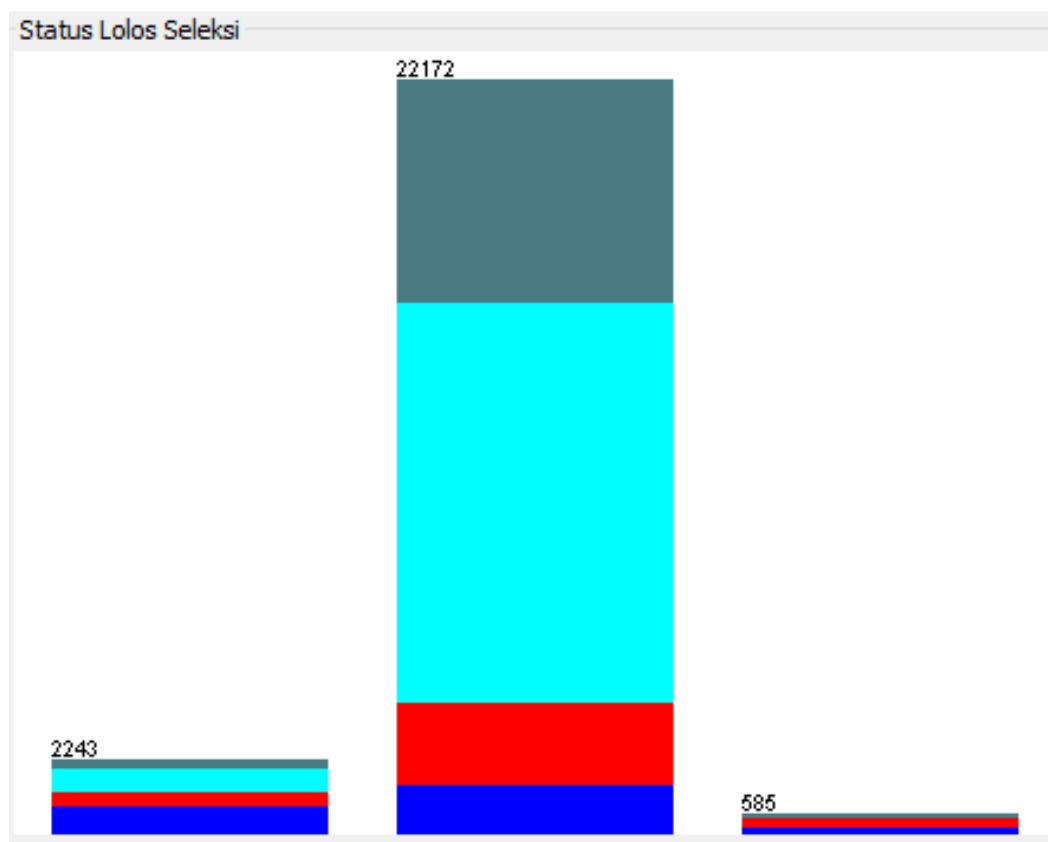
Dari hasil *visualize* untuk atribut jenis kelamin perempuan lebih banyak dari jenis kelamin laki-laki. Jumlah data jenis kelamin perempuan adalah 16909 *record* dan jenis kelamin laki-laki memiliki 8091 *record* data. Gambar 4.10 merupakan bentuk *visualize* atribut jenis kelamin.



Gambar 4. 10 Atribut jenis kelamin dari *visualize WEKA* pada FKIK.

c. Status Lolos Seleksi

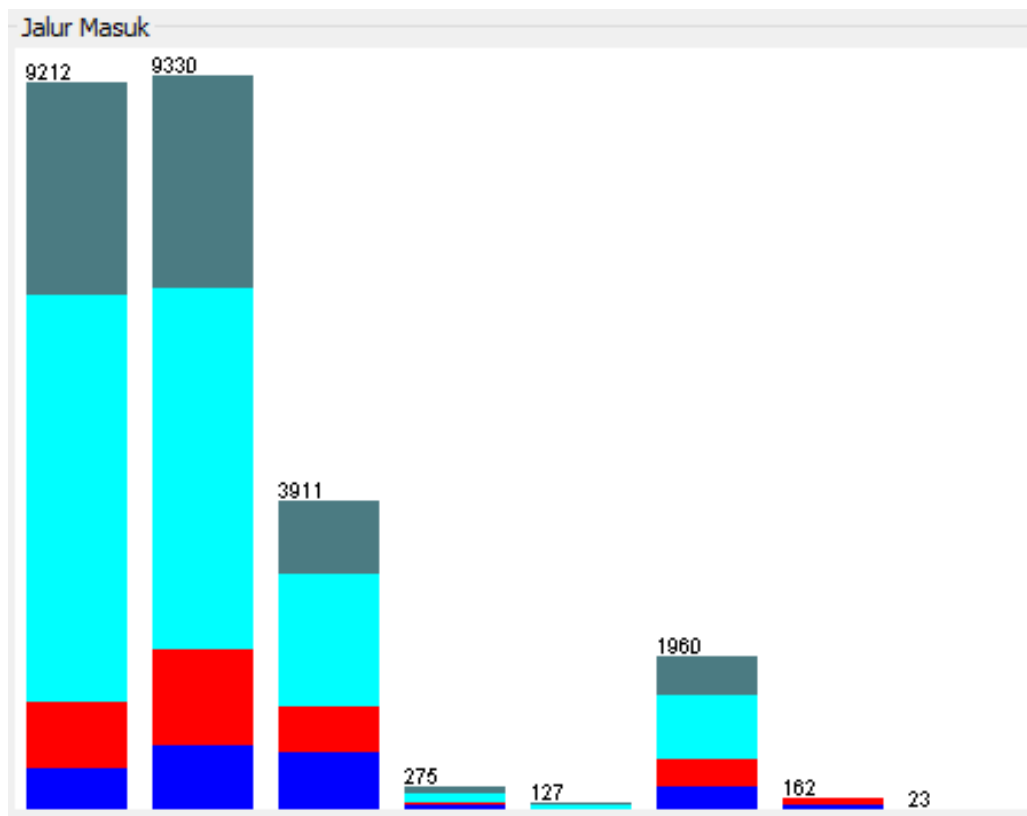
Pada atribut ini dapat dilihat dari *visualize* bahwa status lolos seleksi mahasiswa yang tertinggi yaitu tidak lolos dengan 22172 *record* data. Lolos pada pilihan 1 adalah 2243 *record* data dan lolos pada pilihan 2 adalah 585 *record* data.



Gambar 4. 11 Atribut status lolos seleksi dari *visualize WEKA* pada FKIK.

d. Jalur Masuk

Atribut jalur masuk memiliki 7 jalur pada FKIK, berdasarkan *visualize* dari data penmaru jumlah peminat yang paling banyak pada jalur CBT (*Computer Based Test*) yaitu 9330 *record* data. Selisih jalur CBT dan jalur RT (*Paper Based Test*) adalah 118 *record* data.

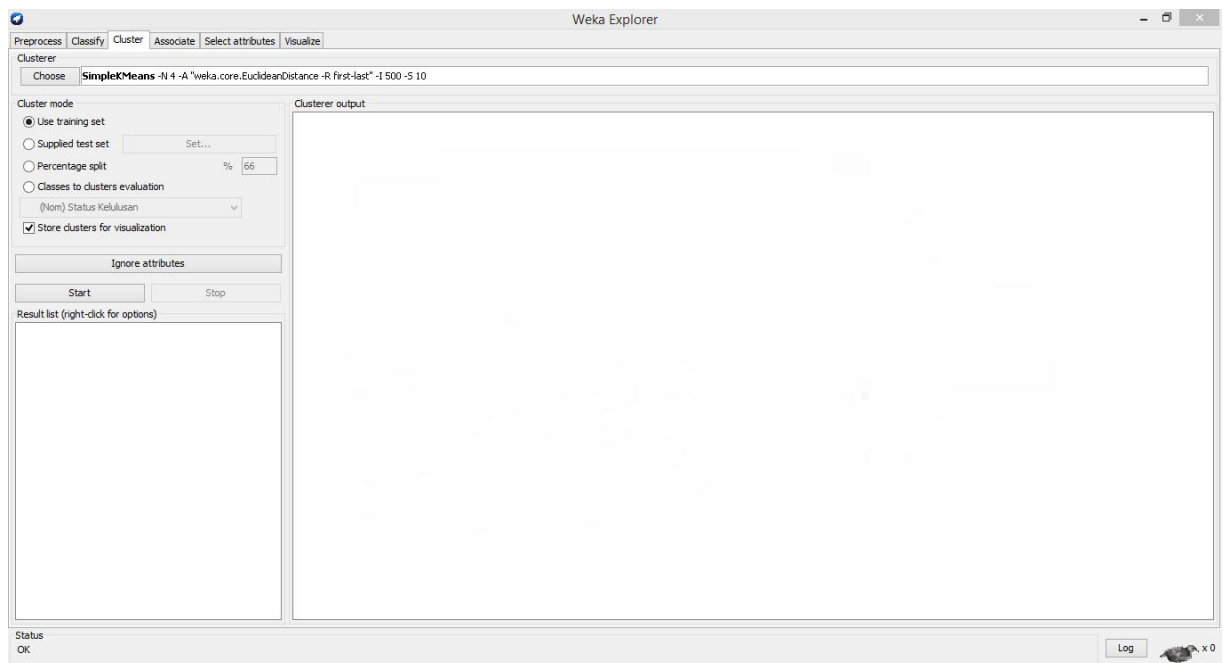


Gambar 4. 12 Atribut jalur masuk dari *visualize WEKA* pada FKIK.

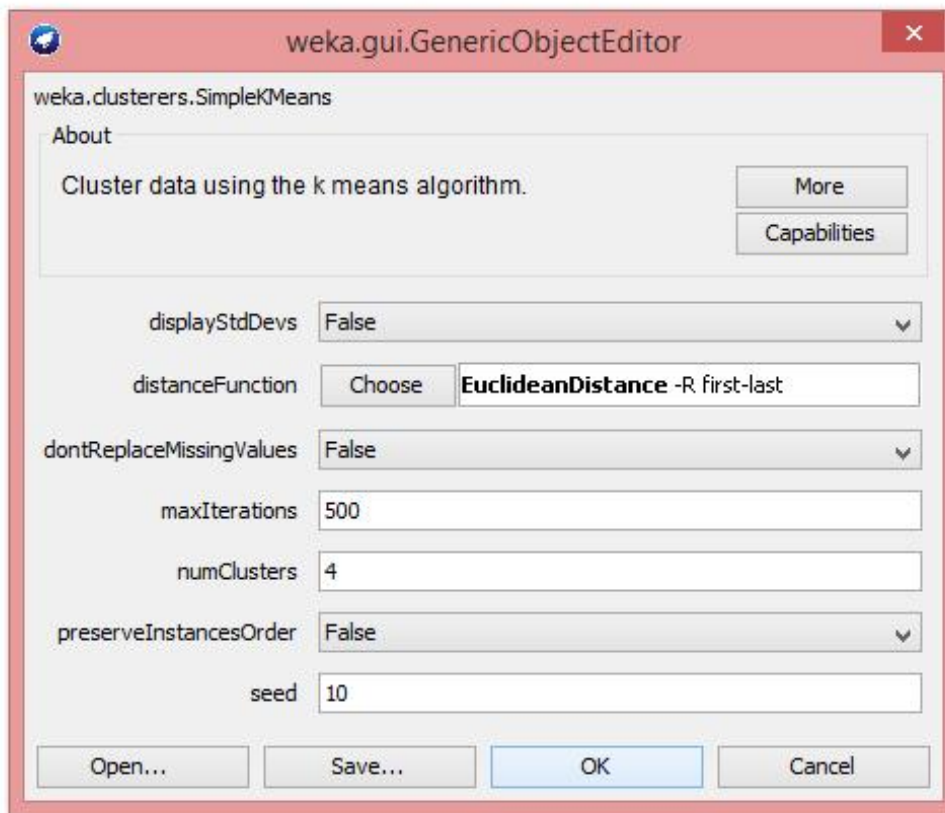
Setelah menganalisa data dari *visualize* pada perangkat lunak *WEKA*, maka dilakukan metode *clustering* dengan menggunakan algoritma *k-means*. Perangkat lunak *WEKA* menyediakan beberapa metode dan terdapat algoritma dari setiap metode tersebut.

Peneliti memasukkan jumlah *cluster* sesuai dengan jumlah kategori pada *class* jurusan yaitu 4 dan iterasi yang dilakukan adalah *default* dari *software* tersebut, dapat dilihat pada gambar 4.14.

Gambar 4.13 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*.

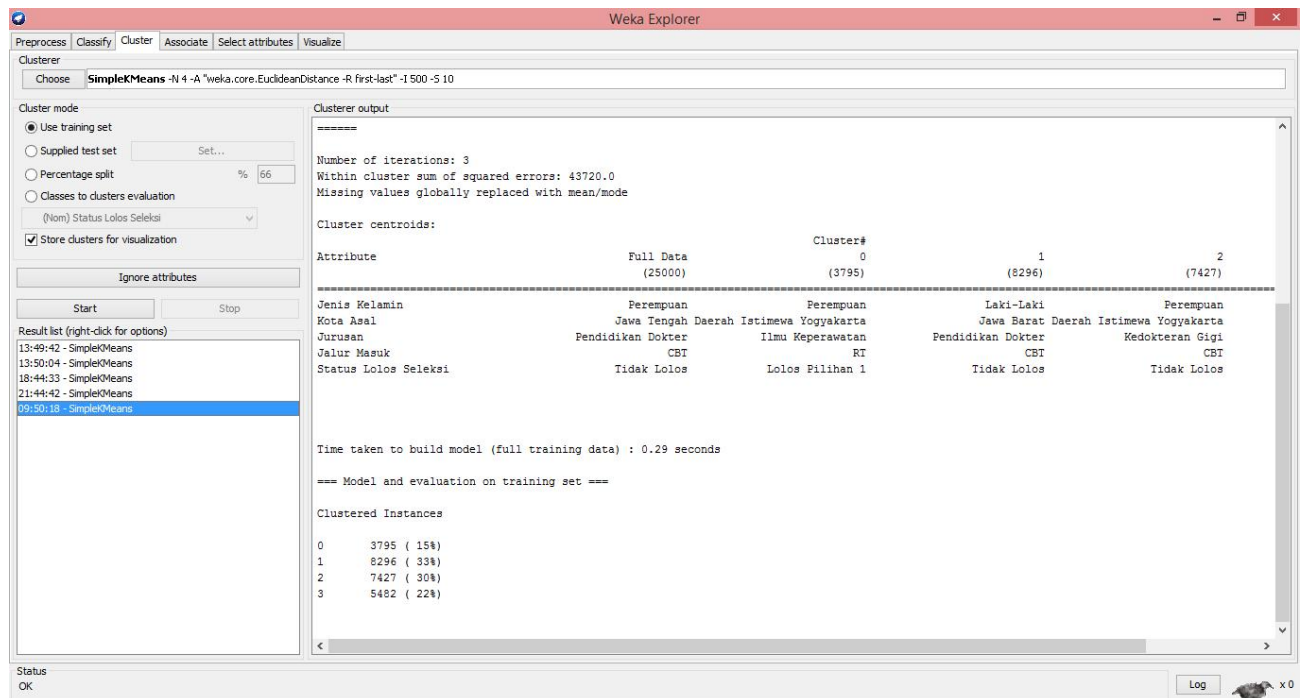


Gambar 4. 14 Cluster data menggunakan algoritma *k-means* FKIK.

Data yang sudah dapat dianalisa selanjutnya menggunakan algoritma *k-means* untuk mencari jumlah data yang masuk dari 4 *cluster*. Gambar 4.15 merupakan hasil dari algoritma *k-means* dengan 4 *cluster* yang digunakan sebagai pengelompokkan data calon mahasiswa baru berdasarkan jurusan yang dipilih dan hasil *cluster centroid* dengan menggunakan *WEKA* seperti pada gambar 4.16 sebagai berikut.

	Cluster#			
Full Data (25000)	0 (3795)	1 (8296)	2 (7427)	3 (5482)
Perempuan	Perempuan	Laki-Laki	Perempuan	Laki-Laki
Jawa Tengah Daerah Istimewa Yogyakarta	Jawa Tengah Daerah Istimewa Yogyakarta	Jawa Barat Daerah Istimewa Yogyakarta	Jawa Tengah	Jawa Tengah
Pendidikan Dokter	Ilmu Keperawatan	Pendidikan Dokter	Kedokteran Gigi	Pendidikan Dokter
CBT	RT	CBT	CBT	RT
Tidak Lolos	Lolos Pilihan 1	Tidak Lolos	Tidak Lolos	Tidak Lolos

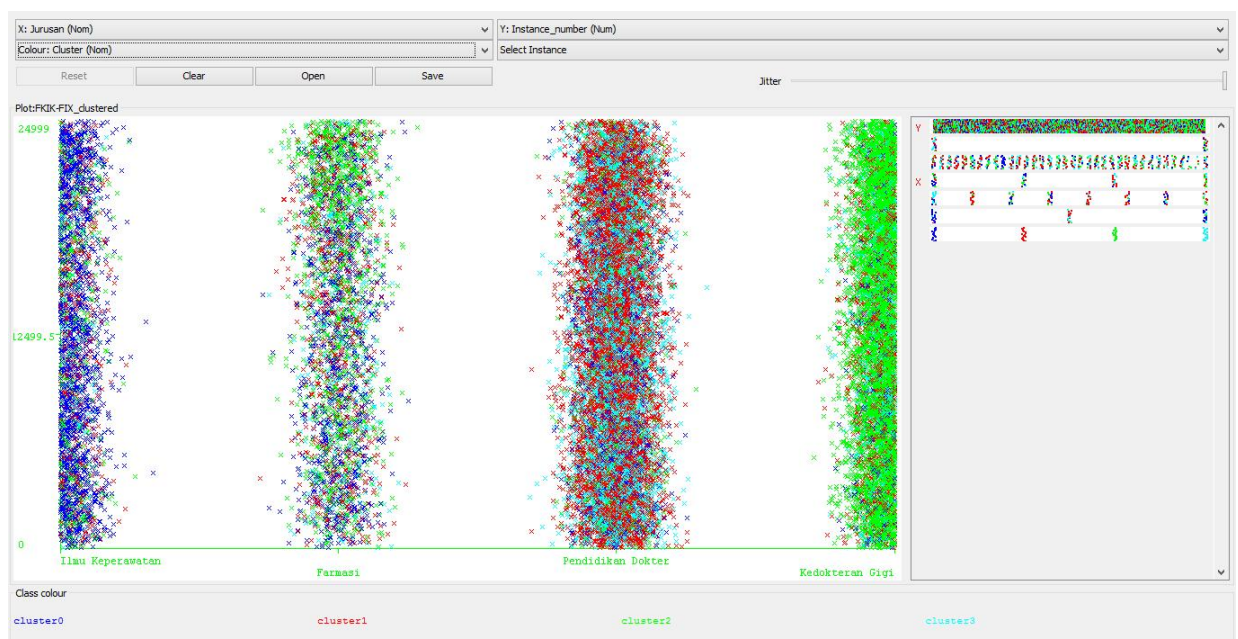
Gambar 4. 15 Cluster centroid pada WEKA FKIK.



Gambar 4. 16 Hasil cluster centroids dan clustered Instances dengan WEKA pada FKIK.

Grafik clustering dapat dilihat pada gambar 4.17, grafik tersebut memiliki banyak warna yang berbeda-beda. Warna tersebut merupakan warna setiap cluster, pada analisis ini terdapat 4 cluster maka warna setiap cluster berbeda-beda. Pada sumbu y kategori yang dipilih berdasarkan jurusan yang terdapat pada Fakultas Kedokteran dan Ilmu

Keperawatan. Dapat dilihat pada gambar 4.17 ada beberapa titik dengan warna yang berbeda-beda letaknya berada jauh dari tumpukan warna lainnya. Hal itu terjadi karena *cluster* tersebut akan mendekati ke jurusan yang dominan lebih dekat dengannya, berdasarkan atribut lain yang mendominasi. Oleh karena itu ada beberapa titik yang agak menjauh dari kumpulan titik warna lainnya.



Gambar 4. 17 Grafik *clustering* FKIK.

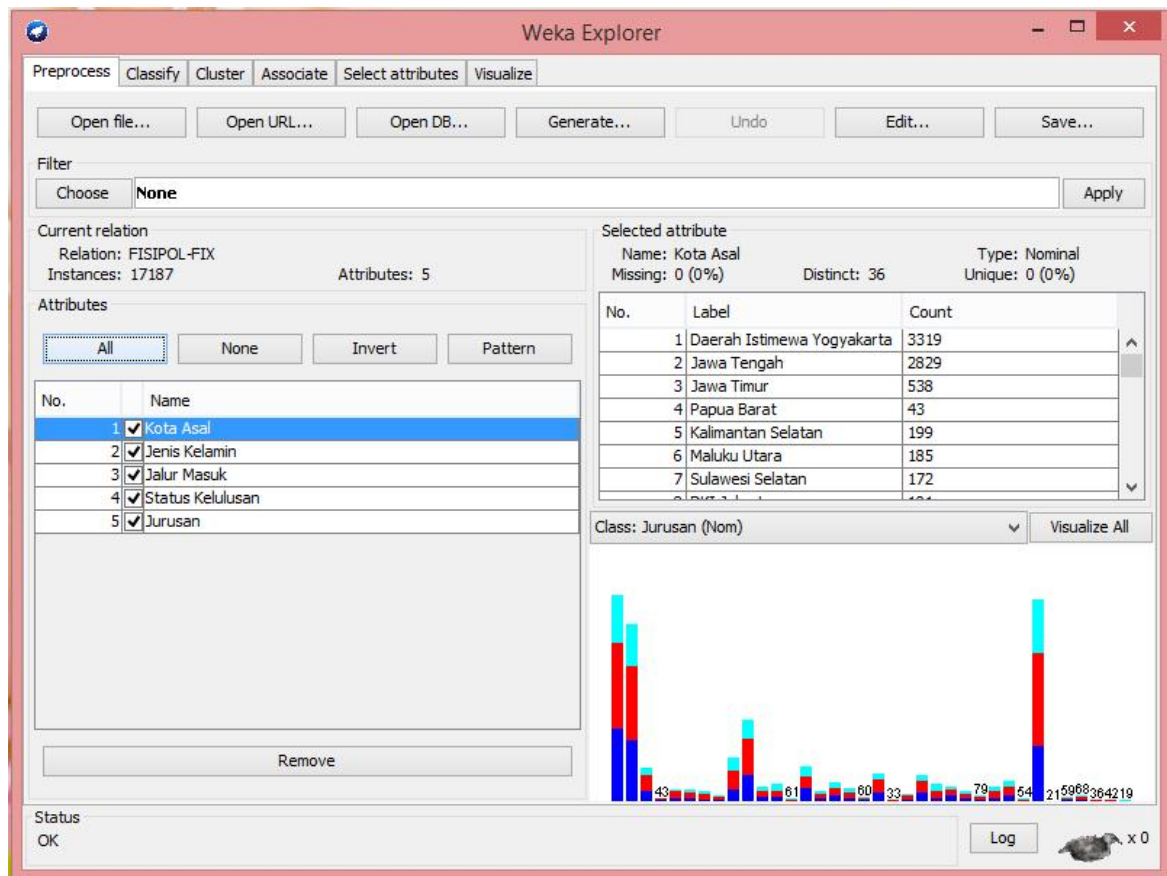
Keterangan:

- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3

Dari hasil *cluster instances* menggunakan *software WEKA* data penmaru Fakultas Kedokteran dan Ilmu Keperawatan pada setiap *cluster* adalah sebagai berikut.

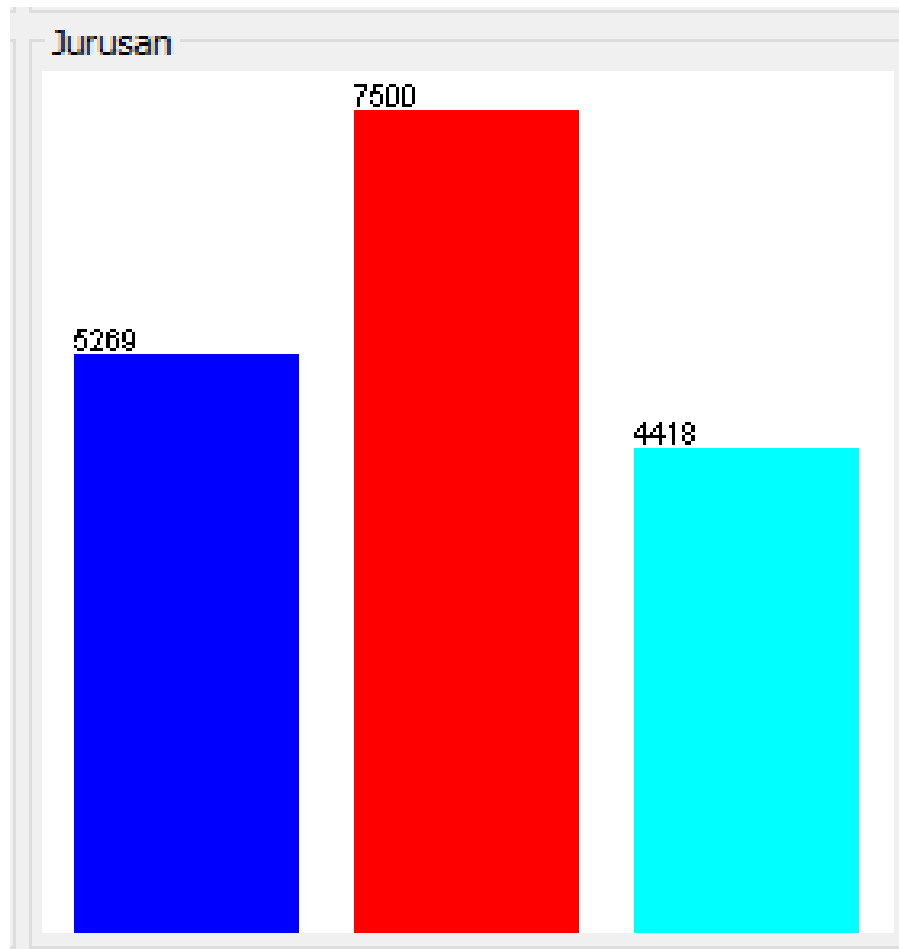
1. *Cluster* 0 dengan jurusan Ilmu Keperawatan, sebanyak 3795 pendaftar dari jumlah 25000 calon mahasiswa baru (15%).
2. *Cluster* 1 dengan jurusan Pendidikan Dokter, sebanyak 8296 pendaftar dari jumlah 25000 calon mahasiswa baru (33%).
3. *Cluster* 2 dengan jurusan Kedokteran Gigi, sebanyak 7427 pendaftar dari jumlah 25000 calon mahasiswa baru (30%).
4. *Cluster* 3 dengan jurusan Pendidikan Dokter, sebanyak 5482 pendaftar dari jumlah 25000 calon mahasiswa baru (22%).

Tahapan analisa dan penggunaan *software WEKA* pada Fakultas Ilmu Sosial dan Ilmu Politik sama halnya dengan tahapan sebelumnya. *Software WEKA* dapat diakses dengan menggunakan format .CSV, maka format *excel convert* ke dalam format .CSV. adapun data penmaru dalam format .CSV dapat diakses di *software WEKA* pada gambar 4.18 sebagai berikut.






Gambar 4. 18 Data yang dianalisa berdasarkan atribut di *WEKA* pada FISIPOL.

Berdasarkan data penmaru pada fakultas Ilmu Sosial dan Ilmu Politik atribut yang digunakan sebagai *class* adalah jurusan. *Visualize* data dapat dilihat pada gambar 4.19 dengan jurusan sebagai *class*.



Gambar 4. 19 Visualize data berdasarkan *class* jurusan FISIPOL.

Keterangan:

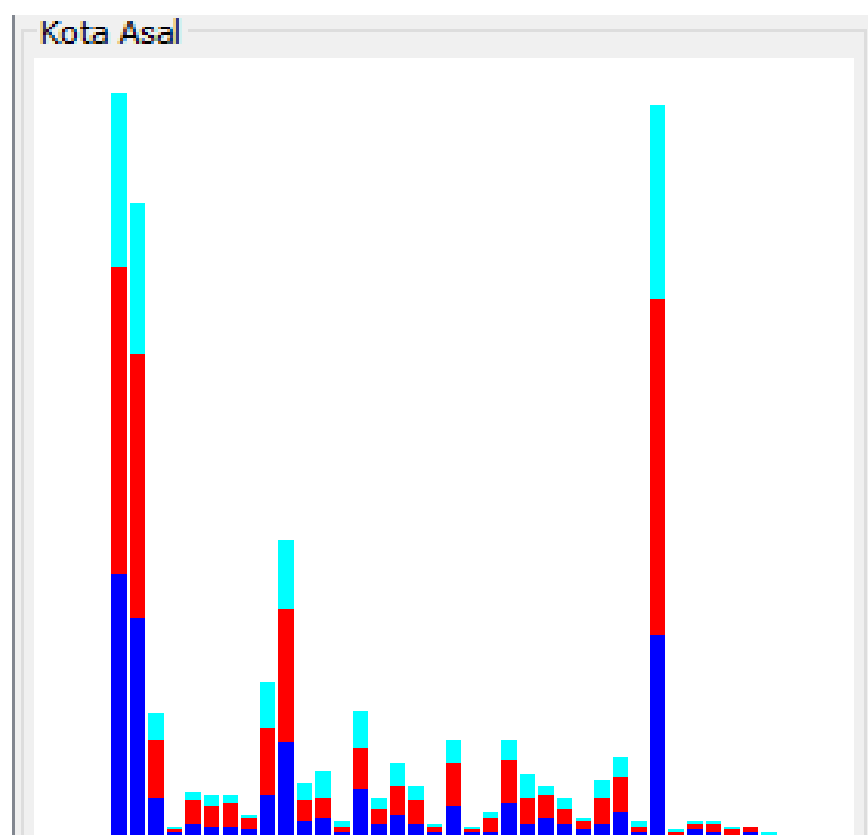
-  Ilmu Pemerintahan
-  Hubungan Internasional
-  Ilmu Komunikasi

Fakultas Ilmu Sosial dan Ilmu Politik memiliki tiga jurusan yaitu Ilmu Pemerintahan, Hubungan Internasional dan Ilmu Komunikasi. Berdasarkan data penmaru FISIPOL pada visualize di *WEKA*, jurusan Hubungan Internasional memiliki peminat terbanyak yaitu 7500 calon mahasiswa baru yang mendaftar, jurusan Ilmu Komunikasi memiliki 5269 calon mahasiswa baru dan jurusan Ilmu pemerintahan memiliki 4418 calon mahasiswa baru.

Adapun hasil *visualize* dari setiap atribut berdasarkan *class* Jurusan, berikut visualize sesuai dengan atributnya. Dari *visualize*, data dapat dibaca dan ditemukan *knowledge* baru dari data tersebut. Untuk lebih jelasnya akan dijelaskan setiap *visualize* dari atribut yang digunakan.

a. Kota Asal

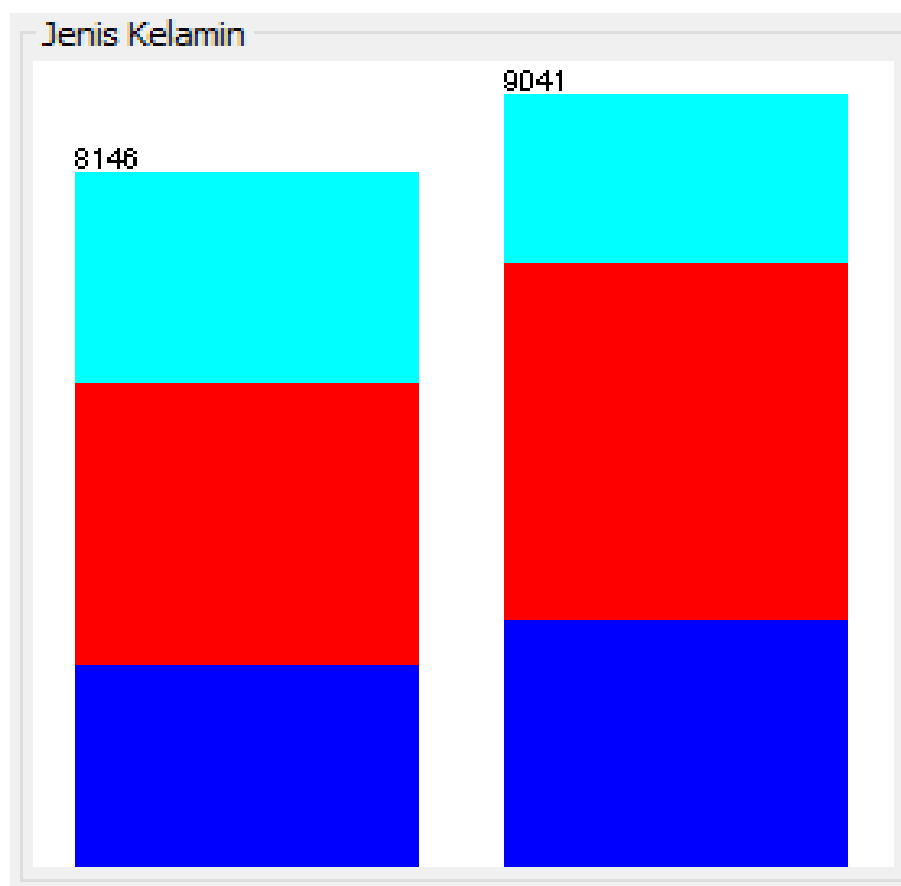
Hasil pengamatan *visualize* pada atribut kota asal, daerah yang paling banyak calon mahasiswanya berasal dari Daerah Istimewa Yogyakarta 3319 record data. *Visualize* kota asal dapat dilihat pada gambar 4.20.



Gambar 4. 20 Atribut kota asal dari *visualize WEKA* pada FISIPOL.

b. Jenis Kelamin

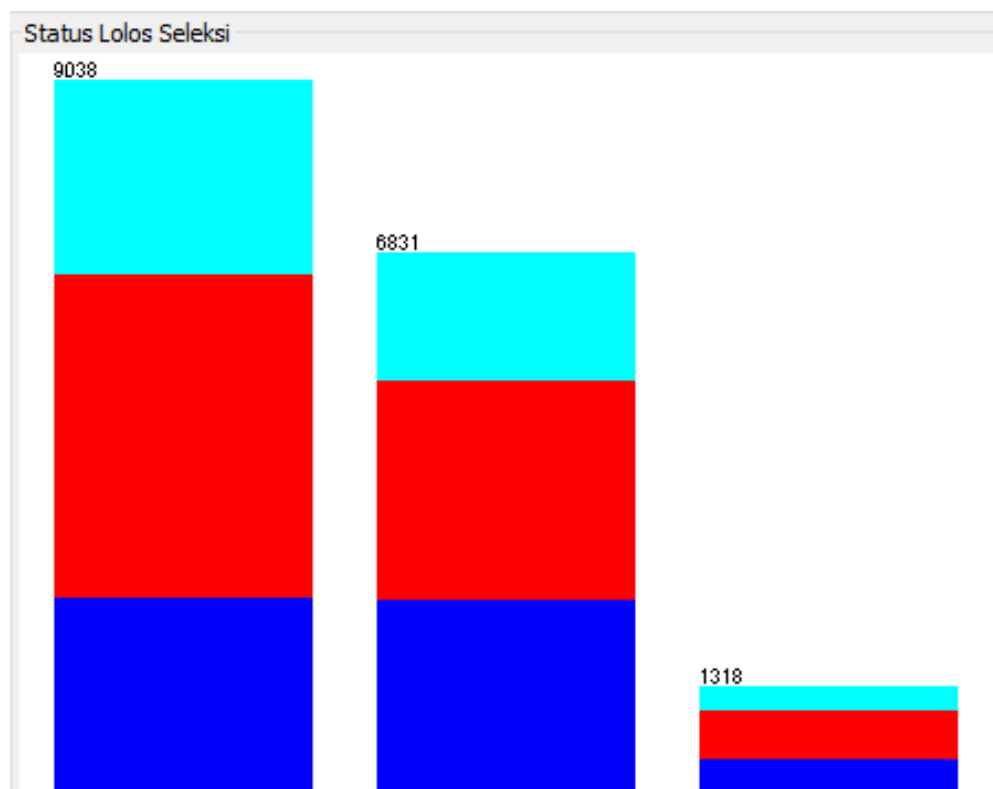
Dari hasil *visualize* untuk atribut jenis kelamin perempuan lebih banyak dari jenis kelamin laki-laki. Jumlah data jenis kelamin perempuan adalah 9041 *record* dan jenis kelamin laki-laki memiliki 8146 *record* data. Gambar 4.21 merupakan bentuk *visualize* atribut jenis kelamin.



Gambar 4. 21 Atribut jenis kelamin dari *visualize WEKA* pada FISIPOL.

c. Status Lolos Seleksi

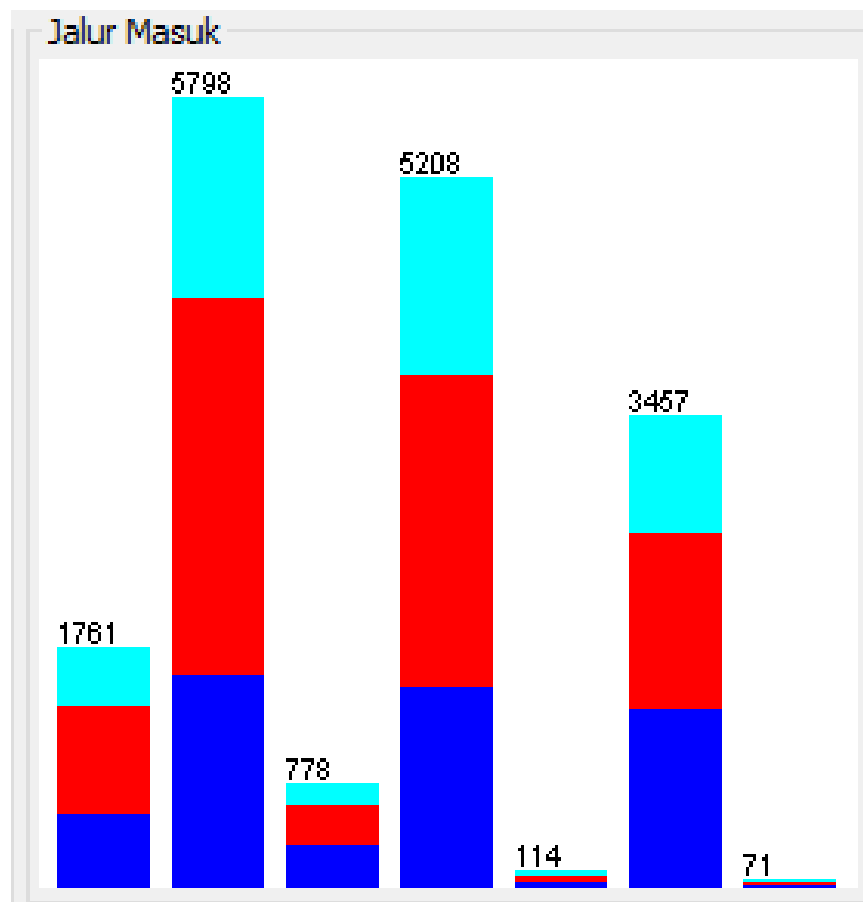
Pada atribut ini dapat dilihat dari *visualize* bahwa status lolos seleksi mahasiswa yang tertinggi yaitu lulus pada pilihan 1 dengan 9038 *record* data. Tidak lolos adalah 6831 *record* data dan lulus pada pilihan 2 adalah 1318 *record* data.



Gambar 4. 22 Atribut status lolos seleksi dari *visualize WEKA* pada FISIPOL.

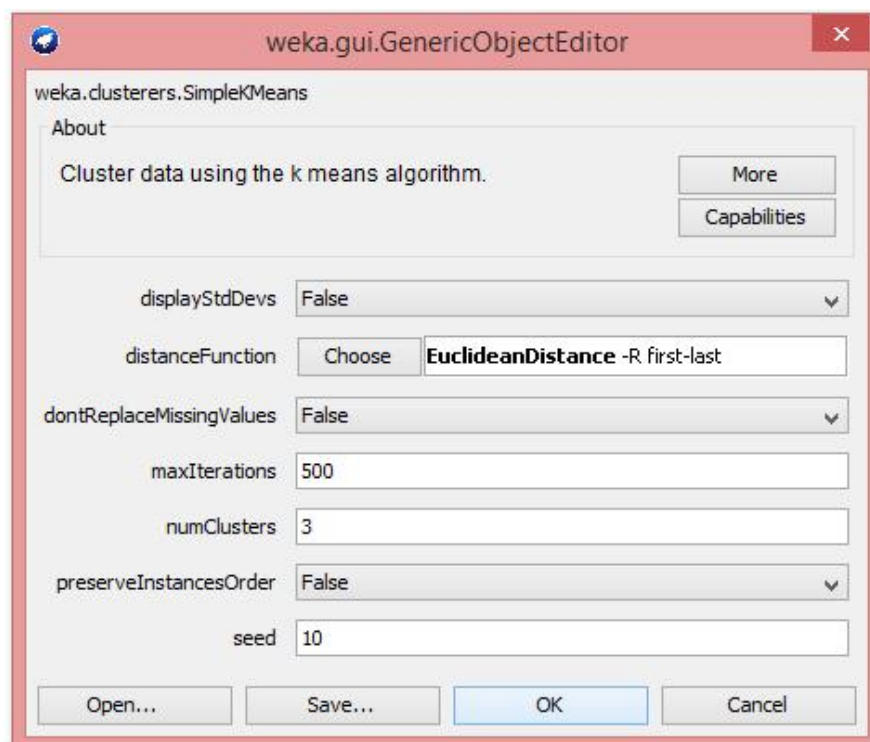
d. Jalur Masuk

Atribut jalur masuk memiliki 7 jalur pada FISIPOL, berdasarkan *visualize* dari data penmaru jumlah peminat yang paling banyak pada jalur PSB (Penerimaan Siswa Berprestasi) yaitu 5798 *record* data. Jalur CBT (*Computer Based Test*) yaitu 5208 *record* data.

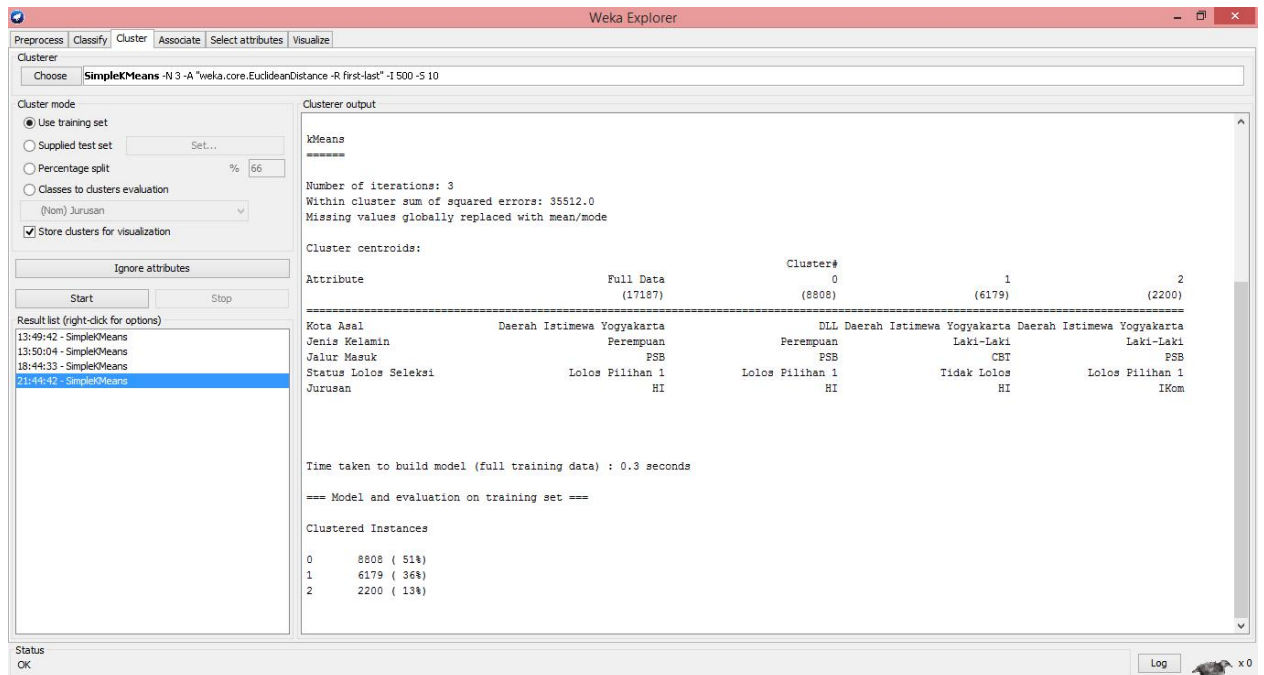


Gambar 4. 23 Atribut jalur masuk dari *visualize WEKA* pada FISIPOL.

Metode *clustering* dengan algoritma *k-means* akan menghasilkan 3 *cluster* berdasarkan *class* Jurusan pada FISIPOL. Hasil *cluster centroid* dan *cluster instance* dapat dilihat pada gambar 4.25. Pada data penmaru FISIPOL jumlah iterasi *default* dari *software WEKA* dengan menggunakan perhitungan jarak antar cluster menggunakan *euclidean distance* seperti pada gambar 4.24.

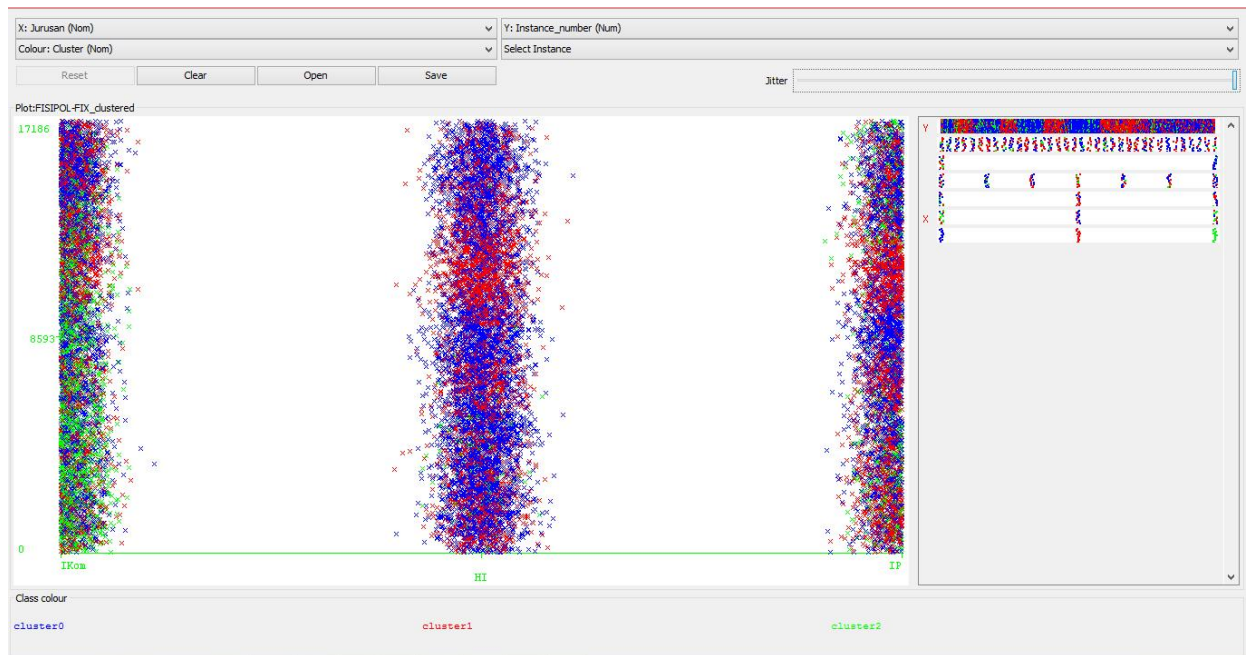


Gambar 4. 24 Cluster data menggunakan algoritma *k-means* FISIPOL.



Gambar 4. 25 Hasil *cluster centroids* dan *clustered Instances* dengan *WEKA* pada FISIPOL.

Grafik *clustering* dapat dilihat pada gambar 4.26, grafik tersebut dikelompokkan berdasarkan kategori jurusan pada sumbu y yang berada pada Fakultas Ilmu Sosial dan Ilmu Politik. Pada fakultas ini terdapat 3 jurusan, maka *cluster* yang dibentuk adalah 3 *cluster*. Berdasarkan hasil *cluster* yang dapat dilihat pada gambar 4.25, maka dari 3 *cluster* memiliki warna berbeda-beda yang masuk dan mendekati jurusan yang dipilih. Ada beberapa titik warna yang tidak masuk dalam kumpulan titik lainnya disebabkan karena pengelompokan data tersebut mencari data dari atribut yang mendekati ke kelompok jurusan yang dipilihnya.



Gambar 4. 26 Grafik *clustering* FISIPOL.

Keterangan:

- Cluster 0
- Cluster 1
- Cluster 2

Dari hasil *cluster instances* menggunakan *software WEKA* data penmaru Fakultas Ilmu Sosial dan Ilmu Politik pada setiap *cluster* adalah sebagai berikut.

1. *Cluster* 0 dengan jurusan Hubungan Internasional, sebanyak 8808 pendaftar dari jumlah 17187 calon mahasiswa baru (51%).
2. *Cluster* 1 dengan jurusan Hubungan Internasional, sebanyak 6179 pendaftar dari jumlah 17187 calon mahasiswa baru (36%).
3. *Cluster* 2 dengan jurusan Ilmu Komunikasi, sebanyak 2200 pendaftar dari jumlah 17187 calon mahasiswa baru (13%).

4.5.2 Algoritma *K-Means*

Dalam menggunakan algoritma *k-means* akan melakukan pengulangan tahapan hingga terjadi kestabilan. Peneliti melakukan satu kali iterasi dalam melakukan pengujian menggunakan algoritma *k-means* pada Fakultas Kedokteran dan Ilmu Keperawatan dengan tahapan sebagai berikut.

1. Menentukan jumlah *cluster* dan menentukan koordinat titik tengah *cluster*. Kelompok *cluster* yang dibuat adalah 5 kelompok berdasarkan jumlah atribut yang digunakan. Total jumlah inisiasi data dari setiap atribut tersebut untuk menentukan kelompok diambil dari frekuensi kurang, frekuensi sedang dan frekuensi padat secara acak dan hasilnya seperti pada tabel 4.9.
2. Penentuan nilai dari *cluster* untuk dijadikan acuan dalam melakukan perhitungan jarak objek ke *centroid*, perhitungan jarak mengacu pada rumus *euclidean*.

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

Rumus *euclidean* merupakan untuk perhitungan jarak antar *centroid*, perhitungan ini dilakukan di *excel* karena data yang jumlahnya sangat banyak. Lihat tabel 4.10 merupakan hasil perhitungan antar *centroid* dengan menggunakan rumus *euclidean distance*.

3. Setelah jarak antar *centroid* dihitung dengan menggunakan rumus *euclidean distance*, maka dilakukan pengelompokkan *centroid* sesuai dengan hasil dari jarak antar *centroid* tersebut seperti pada tabel 4.11. Hasil pada perhitungan jarak tersebut digunakan untuk penentuan kelompok *clustering*. Penentuan dalam pengelompokkan *centroid* adalah sebagai berikut.

- Jika jarak pada *centroid* 1 lebih kecil dari jarak *centroid* 2, *centroid* 3 dan *centroid* 4, maka termasuk kelompok *centroid* 1.
- Jika jarak pada *centroid* 2 lebih kecil dari jarak *centroid* 1, *centroid* 3 dan *centroid* 4, maka termasuk kelompok *centroid* 2.
- Jika jarak pada *centroid* 3 lebih kecil dari jarak *centroid* 1, *centroid* 2 dan *centroid* 4, maka termasuk kelompok *centroid* 3.
- Jika jarak pada *centroid* 4 lebih kecil dari jarak *centroid* 1, *centroid* 2 dan *centroid* 3, maka termasuk kelompok *centroid* 4.

Tabel 4. 9 Penentuan frekuensi dari data dalam menentukan jarak *centroid* pada FKIK.

Centroid					
Centroid 1	1.2	1.8	1.3	1.1	1.1
Centroid 2	1.5	16.7	2.2	3.6	1.9
Centroid 3	1.7	18.3	2.6	5.3	2.5
Centroid 4	1.9	34.5	3.8	7.7	2.8

Tabel 4. 10 Jarak antara *centroid* pada FKIK.

Centroid 1	Centroid 2	Centroid 3	Centroid 4
3.06	14.1	16.03	32.26
1.48	14.03	16.05	32.38
27.24	12.45	11.32	8.32
27.23	12.5	11.41	8.59
2.40	13.8	15.63	32.02
2.40	13.8	15.63	32.02
8.23	7.98	9.56	25.86
5.91	10.25	12.27	28.36
2.93	13.77	15.55	31.91
10.26	5.47	7.80	23.61
10.3	5.07	7.29	23.35
10.24	5.59	7.94	23.71
10.28	5.21	7.44	23.45
2.04	15.97	17.9	34.22
1.26	14.87	16.78	33.16
1.26	14.87	16.78	33.16
0.88	15.01	17	33.35
3.52	14.94	16.7	33
1.26	14.87	16.78	33.16
3.71	12.06	14.07	30.32
3.37	12.03	14.07	30.36
3.31	12.09	14.15	30.43
3.40	13.85	15.73	32.07
1.72	13.88	15.81	32.18
1.48	15.86	17.75	34.14
7.30	8.02	10.11	26.34
1.48	15.86	17.75	34.14
8.67	7.48	9.57	25.46
23.22	8.76	8.12	11.9
23.24	8.51	7.64	11.37
23.22	8.84	8.26	12.09
23.22	8.59	7.79	11.57
8.24	7.35	9.60	25.62
8.29	7.06	9.20	25.37
5.60	9.84	11.74	28.1

Tabel 4. 11 Pengelompokkan *centroid* pada FKIK.

Centroid 1	Centroid 2	Centroid 3	Centroid 4
✓			
✓			
			✓
			✓
✓			
✓			
	✓		
✓			
✓			
	✓		
	✓		
	✓		
	✓		
✓			
✓			
✓			
✓			
✓			
✓			
✓			
✓			

Centroid 1	Centroid 2	Centroid 3	Centroid 4
✓			
✓			
✓			
✓			
✓			
✓			
✓			
	✓		
		✓	
		✓	
		✓	
		✓	
	✓		
	✓		
✓			

Keterangan :

✓ = Kelompok *centroid*

Pada Fakultas Ilmu Sosial dan Ilmu Politik iterasi juga dilakukan satu kali dalam pengujian metode *clustering* dengan menggunakan algoritma *k-means*, adapun tahapan yang dilakukan adalah sebagai berikut.

1. Menentukan jumlah *cluster* dan menentukan koordinat titik tengah *cluster*. Kelompok *cluster* yang dibuat adalah 5 kelompok berdasarkan jumlah atribut yang digunakan. Total jumlah inisiasi data dari setiap atribut tersebut untuk menentukan kelompok diambil dari frekuensi kurang, frekuensi sedang dan frekuensi padat seperti pada tabel 4.12.
2. Penentuan nilai dari *cluster* untuk dijadikan acuan dalam melakukan perhitungan jarak objek ke *centroid*, perhitungan jarak mengacu pada rumus *Euclidean*.

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

Rumus *euclidean* merupakan untuk perhitungan jarak antar *centroid*, perhitungan ini dilakukan di *excel* karena data yang jumlahnya sangat banyak. Lihat tabel 4.13 merupakan hasil perhitungan antar *centroid* dengan menggunakan rumus *euclidean distance*.

3. Setelah jarak antar *centroid* dihitung dengan menggunakan rumus *euclidean distance*, maka dilakukan pengelompokkan *centroid* sesuai dengan hasil dari jarak antar *centroid* tersebut seperti pada tabel 4.14. Hasil pada perhitungan jarak tersebut digunakan untuk penentuan kelompok *clustering*. Penentuan dalam pengelompokkan *centroid* adalah sebagai berikut.

- Jika jarak pada *centroid* 1 lebih kecil dari jarak *centroid* 2 dan *centroid* 3, maka termasuk kelompok *centroid* 1.
- Jika jarak pada *centroid* 2 lebih kecil dari jarak *centroid* 1 dan *centroid* 3, maka termasuk kelompok *centroid* 2.
- Jika jarak pada *centroid* 3 lebih kecil dari jarak *centroid* 1 dan *centroid* 2, maka termasuk kelompok *centroid* 3.

Tabel 4. 12 Penentuan frekuensi dari data dalam menentukan jarak *centroid* pada FISIPOL.

Centroid					
Centroid 1	1.8	1.2	1.4	1.1	1.1
Centroid 2	17.3	1.5	3.2	1.9	1.8
Centroid 3	34.5	1.8	6.5	2.8	2.7

Tabel 4. 13 Jarak antara *centroid* pada FISIPOL.

Centroid 1	Centroid 2	Centroid 3
2.5	14.4	31.83
1.63	14.38	31.88
1.56	16.37	33.85
1.8	16.35	33.81
1.36	16.37	33.86
26.25	10.78	7.97
26.21	10.84	8.28
15.23	1.79	18.23
1.8	14.38	31.87

Centroid 1	Centroid 2	Centroid 3
18.32	3.35	15.29
27.22	11.8	7.4
15.44	2.1	18.09
15.24	1.55	18.18
10.29	5.46	22.98
10.22	5.58	23.09
2.46	16.39	33.8
1.56	16.37	33.85
19.23	4	14.37
19.23	4	14.37
2.46	14.43	31.87
29.23	13.79	6.02
29.24	13.76	5.8
14.24	2	19.14
1.86	14.36	31.84
2.15	16.42	33.86
6.36	9.39	26.91
2.97	13.43	30.88
1.8	16.35	33.81
21.22	5.96	12.59
6.78	9.53	26.89
6.78	9.53	26.89
11.31	4.49	21.99
5.93	10.5	27.86
2.29	16.42	33.85
17.27	2.15	16.17

Tabel 4. 14 Pengelompokkan *centroid* pada FISIPOL.

Centroid 1	Centroid 2	Centroid 3
✓		
✓		
✓		
✓		
✓		
		✓
		✓
	✓	
✓		
	✓	
		✓
	✓	
	✓	
	✓	
	✓	
✓		
✓		
	✓	
	✓	

Centroid 1	Centroid 2	Centroid 3
✓		
		✓
		✓
	✓	
✓		
✓		
✓		
✓		
✓		
	✓	
✓		
✓		
	✓	
✓		
✓		
	✓	

Keterangan :

✓ = Kelompok *centroid*