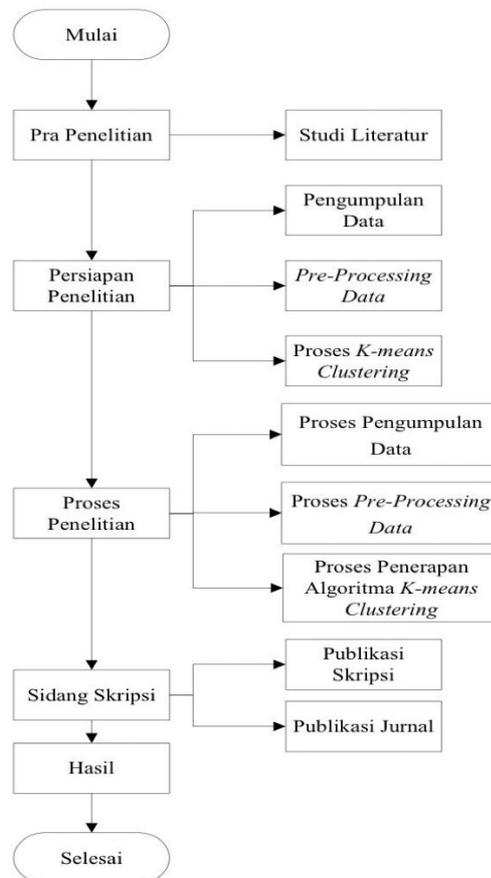


BAB III METODE PENELITIAN

Metode yang digunakan dalam penelitian ini yaitu menerapkan metode *Data Mining* dengan algoritma *K-means Clustering* pada data mahasiswa baru Program Studi Teknik Informatika untuk dianalisa dan dikelompokkan berdasarkan parameter jenis sekolah, asal kota/kabupaten dan jenis kelaminnya.

3.1 Alur Penelitian

Alur penelitian adalah konsep atau gambaran dari penelitian yang akan dilakukan. Penjabaran alur penelitian tersebut dapat dilihat pada gambar 3.1 berikut :



Gambar 3.1 Diagram Alur Penelitian

3.2 Pra Penelitian

Tahapan ini menjelaskan tentang proses awal penelitian yang dilakukan, yaitu studi literatur. Studi literatur digunakan untuk mencari literature dan referensi seperti buku dan jurnal sebagai bahan acuan dari penelitian ini, diantaranya :

- a. Mempelajari penelitian terdahulu yang memiliki topik serupa dengan penelitian yang sedang dilakukan.
- b. Mempelajari jenis algoritma penelitian, metode penelitian, sebagai pedoman dalam memahami jenis penelitian yang sedang dilakukan
- c. Memahami alur PRISMA Flowchart untuk analisis data.

3.3 Persiapan Penelitian

Tahapan ini menjelaskan tentang persiapan penelitian yang dilakukan, ada beberapa tahapan dalam persiapan penelitian ini, yaitu : Pengumpulan Data, *Pre-Processing Data* dan Proses *K-means Clustering*.

3.3.1 Pengumpulan Data

Untuk melakukan proses *K-means Clustering* tentu membutuhkan sebuah data yang cukup banyak dan sesuai dengan yang dibutuhkan, di dalam penelitian ini penulis menggunakan data mahasiswa baru Program Studi Teknik Informatika berdasarkan data 3 tahun terakhir yaitu data mahasiswa baru pada tahun 2020 , 2021 , 2022. Dataset ini dapat dilihat dalam lembar (Lampiran Tabel Dataset Mahasiswa Baru).

Pengambilan data di PMB IIB Darmajaya ini cukup dengan melampirkan surat izin penelitian dari dosen pembimbing ke bagian PMB IIB Darmajaya. Setelah mendapat balasan dari pihak PMB IIB Darmajaya, data dikirim melalui chat whatsapp dimana file tersebut ber-ekstensi xls, data yang didapat tidak sesuai dengan kebutuhan yang sudah tertulis di dalam surat izin penelitian, dimana data yang diterima kurang dari data yang sebenarnya ada, Seperti data mahasiswa baru tahun 2020, mahasiswa baru Program Studi Teknik Informatika tahun 2020 berjumlah 124 mahasiswa sedangkan data yang didapat hanya 22 mahasiswa, hal ini

dikarenakan pihak PMB IIB Darmajaya terdapat kendala saat mengekspor data sehingga data yang didapat tidak lengkap.

3.3.2 Pre-Processing Data

Tahap pre-processing data adalah tahap dimana data yang sudah didapatkan, dipilah, dan dipisahkan agar mendapatkan data yang dibutuhkan untuk proses selanjutnya. Tahapan ini mempunyai beberapa proses dimana setiap proses tersebut saling berhubungan satu sama lainnya. Proses dalam tahapan preprocessing adalah sebagai berikut:

1. Data Reduction

Data reduction adalah proses mereduksi atau pengurangan ukuran, atribut atau beberapa bagian informasi yang tidak diperlukan dalam file data. Reduksi data sangat berguna untuk mendapatkan atribut dan data yang digunakan dalam penelitian ini.

2. Data Cleaning

Data cleaning adalah proses pada tahap pre-processing yaitu mengisi data kosong atau blank sebanyak-banyaknya, menggandakan data, mengoreksi data yang tidak konsisten atau kesalahan ketik seperti huruf yang hilang dan huruf yang berlebih, mengubah dan memodifikasi data sehingga menjadi data.

3. Data Transformation

Karena metode *K-means Clustering* adalah metode yang bisa dilakukan apabila data yang dipakai adalah data berupa angka, maka proses transformation ini sangat dibutuhkan. Proses transformation adalah tahap untuk mengubah data atribut yang selain angka ke dalam nilai angka agar data tersebut dapat diolah menggunakan algoritma *K- Means Clustering*.

4. Data Integration

Data integration adalah suatu proses untuk menggabungkan atau mengintegrasikan data dari beberapa file sumber. Data integration hanya dilakukan jika data berasal dari tempat yang

berbeda-beda (sumber data tidak hanya dari 1 tempat). Langkah yang dilakukan antara lain mengintegrasikan skema, mengidentifikasi masalah entitas, dan mendeteksi sekaligus menyelesaikan konflik pada nilai data.

3.3.3 Proses *K-means Clustering*

Setelah data melalui tahap pre-processing dan tahap transformation maka data sudah siap untuk diolah menggunakan metode *K-means Clustering*. Metode *K-means Clustering* adalah proses untuk mengelompokkan data ke dalam sebuah cluster dengan titik pusat yang berbeda beda setiap cluster. Dalam penelitian ini penulis sudah menentukan untuk menggunakan 2 cluster. Adapun proses *K-means Clustering* tersebut meliputi 5 proses, yaitu :

1. Menentukan Jumlah Cluster (K)

Menentukan k untuk jumlah cluster yang ingin dibentuk adalah langkah awal untuk proses *K-means Clustering*.

2. Menentukan Titik Pusat Cluster

Setelah ditentukannya k dalam jumlah cluster yang ingin dibentuk kemudian tahap selanjutnya adalah menentukan titik pusat cluster, fungsi proses ini untuk menentukan titik awal sebagai patokan untuk mencari jarak antara data ke cluster yang sudah ditentukan. Titik awal pusat cluster disebut juga dengan centroid. Untuk menentukan titik pusat setiap cluster bisa dilakukan dengan mencari rata-rata dari data yang akan diolah ataupun sesuai keinginan.

3. Menghitung Jarak Data ke Setiap Cluster

Setelah menentukan titik pusat di setiap cluster proses selanjutnya adalah menghitung jarak antara data ke setiap cluster yang sudah dibentuk. Rumus untuk mencari jarak (distance) dari satu cluster adalah :

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

Keterangan :

D = titik dokumen

P = data record

Q = data centroid

Rumus tersebut adalah rumus untuk menentukan jarak dari satu baris data ke satu cluster tertentu. Dalam penelitian ini penulis menggunakan 2 cluster, maka perhitungannya pun menjadi 2 kali dengan titik pusat cluster yang berbeda di setiap cluster.

4. Menentukan Titik Pusat Cluster Baru

Setelah mengalokasikan data ke dalam cluster yang terbentuk maka proses selanjutnya adalah menentukan titik pusat cluster baru dengan cara yang sama yaitu mencari rata-rata di setiap atribut data. Tetapi dalam perhitungan kali ini sedikit berbeda dari yang pertama, apabila proses pertama adalah mencari rata-rata dari semua atribut data maka untuk menentukan titik pusat cluster baru ini data yang digunakan sesuai dengan clusternya masing-masing. Apabila data yang masuk ke cluster satu adalah data pertama, ketiga, keempat, dan keenam, maka rata-rata yang dicari hanya menggunakan data pertama, ketiga, keempat, dan keenam.

5. Memverifikasi Titik Pusat Cluster

Setelah mendapatkan titik pusat cluster baru maka proses selanjutnya adalah memverifikasi titik pusat cluster baru tersebut dengan titik pusat cluster yang lama. Apabila hasil titik pusat cluster baru yang didapat sama dengan titik pusat cluster yang lama, maka proses K-Means sudah selesai dan hasil dari proses

K-means Clustering sudah didapatkan dan data yang diklasifikasi sudah tidak bisa berubah-ubah lagi. Tetapi jika hasil titik pusat cluster baru yang didapat berbeda dari titik pusat cluster yang lama, maka proses K- Means tetap dilanjutkan dan mulai lagi dari proses kedua atau menghitung jarak data ke setiap cluster.

3.4 Proses Penelitian

Tahapan ini dilakukan setelah data yang diperoleh sudah lengkap. Pengolahan data dilakukan dengan menggunakan perhitungan algoritma K-Means setelah proses transformasi, data mahasiswa baru bisa diolah menggunakan metode *K-Means Clustering*.

3.4.1 Proses Pengumpulan Data

Pengumpulan data ini terdapat dari file excel berekstensi xls dari bagian PMB IIB Darmajaya . File yang didapatkan berupa data mentah yang belum diolah sama sekali. Jumlah data yang didapatkan berjumlah 356 data mahasiswa baru program studi teknik informatika 3 tahun terakhir, yaitu pada tahun 2020 berjumlah 22 data mahasiswa baru, pada tahun 2021 berjumlah 154 data mahasiswa baru, dan pada tahun 2022 berjumlah 180 data mahasiswa baru. Dataset ini terdiri dari beberapa kolom atau beberapa atribut yaitu :

1. Nama Mahasiswa
2. Program Pilihan
3. Jenis Kelamin
4. Kewarganegaraan
5. Alamat
6. Propinsi
7. Kota
8. Kecamatan
9. Kelurahan
10. Dusun
11. RT

12. RW
13. Kode Pos
14. No HP
15. Nama Ayah
16. Alamat Ayah
17. No HP Ayah
18. Nama Ibu
19. Status
20. Nama Sekolah
21. Email

Data tersebut ialah atribut lengkap yang sama sekali belum dipisahkan, dipilah dan belum dilakukan reduksi data, cleaning data dan transformasi data sehingga data ini belum bisa dilanjutkan untuk proses mining. Dataset mahasiswa baru sebelum dilakukan reduksi data dapat dilihat pada gambar 3.2 berikut :

No.	Nama	Program	Pilihan 1	Jenis Kelamin	Kewarga negaraan	Alamat	Propinsi	Kota	Kecamatan	Kelurahan	Dusun	RT	RW	Kode Pos	No HP	Nama Ayah	Alamat Ayah	No HP Ayah	Nama Ibu	Status	Nama Sekolah	Email
19	DANDI PRATAM A	Reguler	Teknik Informatika	Laki-Laki	ID	JATI WANGI A RT07 RW012	Prop. Lampung	Kab. Lampung Selatan	Kec. Tanjung Bintang	JATI INDAH	JATI WANGI	7	12	35361	8,5E+10	UNTUNG	JATI WANGI A RT07 RW012	8,6E+10	RIANA MEGA SARI	Mahasiswa Baru	SMKS BINA KARVA MANDIRI BEKASI	dandoyr evege@gmail.com
27	SANDI HANIF AZZAM	Kelas Karyawan	Teknik Informatika	Laki-Laki	ID	Jalan gunung sindoro no 13	Prop. Jawa Tengah	Kab. Banyuwangi	Kec. Purwokerto Utara	Pabuara		2	4	53124	8,2E+10	Teguh waluyo	Jalan gunung sindoro no 13	0	Waryati	Pindahan		sandy1511010005@mail.darmajaya.ac.id
36	SAIPUL SIHOTAN	Kelas Karyawan	Teknik Informatika	Laki-Laki	ID	JALAN SOEKARNA HATTA NO2	Prop. Lampung	Kota Bandar Lampung	Kec. Kedamaian	KEDAMAIAN	KEDAMAIAN	14	14	35122	8,5E+10	AMAT SIHOTAN	JALAN SOEKARNA HATTA NO2	0		Pindahan	SMK Bhakti Utama 2	spardaban@gmail.com
37	WHISNU DWI PRATUDH	Kelas Karyawan	Teknik Informatika	Laki-Laki	ID	Jl. Perum Bukit Kencana Blok J No.12A Antasari	Prop. Lampung	Kota Bandar Lampung	Kec. Kedamaian	Kalibala Kencana	Kalibala Kencana	0	0	35122	8,2E+10	SUNARTO	Jl.A Yani, KM 4,5 Kemelak bindung langit,RT /RW 01/05 kel. kemelak bindung langit,ke c.BatuRaja Timur	0	KARLINA	Pindahan		whismudp.1511010131@mail.darmajaya.ac.id

Gambar 3.2 Dataset Mahasiswa Baru Sebelum Dilakukan Reduksi Data

3.4.2 Proses *Pre-Processing Data*

Pada tahap ini adalah tahapan dari Data Mining yaitu suatu proses atau tahapan yang dilakukan untuk mengolah data mentah menjadi data yang berkualitas atau inputan yang baik untuk dilanjutkan ke proses selanjutnya. Tahap pre-processing data ini adalah tahap yang sangat krusial atau sangat penting dan harus dilakukan dengan teliti karena proses Data Mining membutuhkan data yang konsisten dalam penulisan, benar dalam format nya, tidak ada data yang kosong, duplikasi data, dan lain lain. Data yang tidak berkualitas maka hasil dari proses Data Mining ini akan menghasilkan hasil yang tidak berkualitas juga.

1. Proses Reduksi Data

Data Reduction adalah proses untuk mereduksi atau mengurangi dimensi, atribut, ataupun sejumlah data yang tidak dibutuhkan. Pada data mahasiswa baru program studi teknik informatika 3 tahun terakhir, penulis mereduksi data yang tidak diperlukan seperti : nama mahasiswa, Program Pilihan, Kewarganegaraan, Alamat, Propinsi, Kecamatan, Kelurahan, Dusun, RT, RW, Kode Pos, No HP, Nama Ayah, Alamat Ayah, No HP Ayah, Nama Ibu, Status, Email. Dalam penelitian ini atribut yang dibutuhkan adalah kota/kabupaten asal mahasiswa, jenis sekolah mahasiswa, dan jenis kelamin mahasiswa, sehingga tahapan reduksi seperti pada tabel 3.1 :

Tabel 3.1 Dataset Setelah Tahapan Reduksi

No	Jenis Kelamin	Kota	Nama Sekolah
1	Perempuan	Kab. Lampung Timur	
2	Perempuan		
3	Perempuan		
4	Perempuan	Kab. Tulang Bawang Barat	
5	Laki-Laki	Kab. Brebes	
6	Perempuan	Kab. Way Kanan	
7	Laki-Laki	Kab. Tenggamus	

Tabel 3.1 Dataset Setelah Tahapan Reduksi (Lanjutan)

No	Jenis Kelamin	Kota	Nama Sekolah
8	Laki-Laki	Kota Bandar Lampung	
9	Perempuan		
10	Laki-Laki	Kota Metro	
11	Perempuan	Kab. Way Kanan	
12	Laki-Laki	Kab. Tenggamus	
13	Laki-Laki	Kab. Lampung Barat	
14	Laki-Laki	Kab. Lampung Selatan	
15	Perempuan		
....
356	Laki-Laki	Kota Bandar Lampung	SMA S XAVERIUS BANDAR LAMPUNG

2. Proses Data Cleaning

Data Cleaning adalah proses untuk mengisi missing value atau isian yang hilang atau kurang pas, mengoreksi data yang tidak konsisten, dan mendeteksi redudansi data yaitu penumpukan data atau duplikasi data. Pada tahap *pre-processing* pertama sudah melewati tahap reduksi yang terdapat kolom no, jenis kelamin, kota/kabupaten dan nama sekolah dengan data yang berjumlah 356 data mahasiswa baru. Kemudian di tahap ini penulis melakukan proses data cleaning dengan aplikasi Microsoft Excel dengan cara memfilter setiap kolom dan mencari data yang kosong atau missing. Pada kolom asal kota\kabupaten dan asal sekolah terdapat data yang tidak konsisten dalam penulisannya sehingga data harus dikoreksi dan diubah agar data tetap konsisten. Sehingga tahapan data cleaning seperti pada tabel 3.2 :

Tabel 3.2 Dataset Setelah Tahapan Data Cleaning

No	Jenis Kelamin	Kota/Kabupaten	Nama Sekolah
1	Laki-Laki	Kab. Tulang	SMKN 2 METRO

Tabel 3.2 Dataset Setelah Tahapan Data Cleaning (Lanjutan)

No	Jenis Kelamin	Kota/Kabupaten	Nama Sekolah
		Bawang Barat	
2	Perempuan	Kota Bandar Lampung	SMA N 1 GUNUNG AGUNG
3	Laki-Laki	Kab. Pasawaran	SMAN 12 BANDAR LAMPUNG
4	Laki-Laki	Kota Bandar Lampung	SMKS PENERBANGAN LAMPUNG
5	Laki-Laki	Kab. Lampung Barat	SMKS Yadika Natar
6	Laki-Laki	Kab. Pasawaran	MA NURUL IMAN KAB. LAMPUNG BARAT
7	Laki-Laki	Kab. Lampung Barat	SMAN 1 PADANG CERMIN
8	Laki-Laki	Kota Bandar Lampung	SMK NEGERI 1 WAY TENONG
9	Laki-Laki	Kota Bandar Lampung	SMAS PERINTIS 2 BANDAR LAMPUNG
10	Laki-Laki	Kota Bandar Lampung	SMAS TAMAN SISWA TELUK BETUNG UTARA
11	Laki-Laki	Kab. Lampung Utara	SMKN 4 Bandar Lampung Utara
12	Laki-Laki	Kota Bandar Lampung	SMAN 2 METRO
13	Laki-Laki	Kota Bandar Lampung	SMAN 7 BANDAR LAMPUNG
14	Laki-Laki	Kota Bandar Lampung	SMAN 7 BANDAR LAMPUNG

Tabel 3.2 Dataset Setelah Tahapan Data Cleaning (Lanjutan)

No	Jenis Kelamin	Kota/Kabupaten	Nama Sekolah
15	Laki-Laki	Kota Bandar Lampung	SMKN 6 BANDAR LAMPUNG
.....
279	Laki-Laki	Kota Bandar Lampung	SMA S XAVERIUS BANDAR LAMPUNG

3. Proses Data Transformation

Data transformation adalah proses mengubah suatu data supaya mendapatkan data yang lebih berkualitas atau sesuai dengan kebutuhan. Dalam tahapan ini proses yang pertama ialah penulis mengkonversi kolom nama sekolah ke jenis sekolah hal ini agar mempermudah untuk proses pengelompokkan pada transformasi data, sehingga tahapan konversi seperti pada tabel 3.3 :

Tabel 3.3 Konversi Nama Sekolah Ke Jenis Sekolah

Nama Sekolah	Jenis Sekolah
SMAN/ SMA Negeri.....	SMA Negeri
SMKN/ SMK Negeri...	SMK Negeri
SMAS/ SMA Swasta...	SMA Swasta
LAINNYA	LAINNYA
SMKS/ SMK Swasta	SMK Swasta

Setelah melakukan tahapan konversi untuk transformasi data sehingga data tersebut seperti pada tabel 3.4 :

Tabel 3.4 Tahapan Konversi Data

No	Nama Sekolah	Jenis Sekolah
1	SMKN 2 METRO	SMK Negeri
2	SMA N 1 GUNUNG AGUNG	SMA Negeri
3	SMAN 12 BANDAR LAMPUNG	SMA Negeri
4	SMKS PENERBANGAN LAMPUNG	SMK Swasta

Tabel 3.4 Tahapan Konversi Data (Lanjutan)

No	Nama Sekolah	Jenis Sekolah
5	SMKS Yadika Natar	SMK Swasta
6	MA NURUL IMAN KAB. LAMPUNG BARAT	Lainnya
7	SMAN 1 PADANG CERMIN	SMA Negeri
8	SMK NEGERI 1 WAY TENONG	SMK Negeri
9	SMAS PERINTIS 2 BANDAR LAMPUNG	SMA Swasta
10	SMAS TAMAN SISWA TELUK BETUNG UTARA	SMA Swasta
11	SMKN 4 Bandar Lampung	SMK Negeri
12	SMAN 2 METRO	SMA Negeri
13	SMAN 7 BANDAR LAMPUNG	SMA Negeri
14	SMAN 7 BANDAR LAMPUNG	SMA Negeri
15	SMKN 6 BANDAR LAMPUNG	SMK Negeri
.....
279	SMA S XAVERIUS BANDAR LAMPUNG	SMA Swasta

Setelah melewati proses data reduction dan data cleaning penulis sudah mendapatkan atribut sesuai dengan kebutuhan dalam penelitian ini tetapi isi data dari atribut tersebut masih belum memenuhi kondisi atau syarat yang cocok untuk dilanjutkan ke dalam proses algoritma *K-means Clustering*. Dalam proses algoritma *K-means Clustering* data yang bisa diolah hanya data yang berupa angka sedangkan data yang ada pada atribut kota/kabupaten asal mahasiswa, jenis sekolah mahasiswa dan jenis kelamin mahasiswa adalah data yang berupa text sehingga penulis harus mengubah data tersebut kedalam angka agar bisa dilanjutkan ke dalam proses *K-means Clustering*. Langkah pertama yang dilakukan penulis adalah mencari frekuensi atau jumlah data yang sama di setiap atribut lalu mengurutkannya dari frekuensi terbesar sampai ke terkecil dan memisahkannya kedalam tabel yang berbeda. Setelah itu penulis memberi inisial pada setiap data berdasarkan frekuensi terbesar ke terkecil tujuannya adalah frekuensi yang paling besar nilainya merupakan data yang paling banyak muncul maka inisial dimulai

dari frekuensi terbanyak. Pada atribut kota/kabupatten asal mahasiswa penulis mendapatkan 19 data asal kota/kabupaten mahasiswa dengan frekuensi dan inisial dapat dilihat pada tabel 3.5 berikut :

Tabel 3.5 Frekuensi Kota/Kabupaten Mahasiswa

Kota / Kabupaten	Frekuensi	Inisial
Kota Bandar Lampung	122	1
Kab. Lampung Tengah	27	2
Kab. Lampung Selatan	23	3
Kab. Pesawaran	21	4
Kab. Lampung Utara	11	5
Kab. Pringsewu	11	6
Kota Metro	9	7
Kab. Lampung Timur	9	8
Kab. Tulang Bawang	9	9
Kab. Tanggamus	8	10
Kab. Ogan Komering	8	11
Kab. Lampung Barat	6	12
Kab. Way Kanan	6	13
Kab. Pesisir Barat	3	14
Kab. Mesuji	2	15
Kab. Bogor	1	16
Kab. Rejang Lebong	1	17
Kab. Sarolangun	1	18
Kab. Tanggerang	1	19

Pada atribut jenis sekolah mahasiswa penulis mendapatkan 5 data jenis sekolah dengan frekuensi dan inisial dapat dilihat pada tabel 3.6 berikut :

Tabel 3.6 Frekuensi Jenis Sekolah Mahasiswa

Jenis Sekolah	Frekuensi	Inisial
SMA Negeri	120	1
SMK Negeri	77	2
SMA Swasta	30	3
LAINNYA	27	4

Tabel 3.6 Frekuensi Jenis Sekolah Mahasiswa (Lanjutan)

Jenis Sekolah	Frekuensi	Inisial
SMK Swasta	25	5

Pada atribut jenis kelamin mahasiswa penulis membagi 2 data jenis kelamin dengan frekuensi dan inisial dapat dilihat pada tabel 3.7 berikut :

Tabel 3.7 Frekuensi Jenis Kelamin Mahasiswa

Jenis Kelamin	Frekuensi	Inisial
Laki-Laki	220	1
Perempuan	59	2

4. Proses Data Integration

Data integration atau integrasi data adalah suatu proses untuk menggabungkan data dari beberapa file sumber.

3.4.3 Penerapan Algoritma *K-means Clustering*

Setelah data mentah yang didapat dari pihak PMB IIB Darmajaya diolah pada tahapan pre-processing dan sudah mendapatkan dataset yang siap untuk diolah maka tahap berikutnya adalah menerapkan algoritma *K-means Clustering* pada dataset tersebut. Dalam penelitian ini penulis menerapkan algoritma *K-means Clustering* dengan 10 sampel data mahasiswa baru menggunakan perhitungan manual aplikasi Microsoft Excel sampai dengan proses algoritma *K-means Clustering* berhenti, hal ini bertujuan untuk menerapkan algoritma *K-means Clustering* dalam mengetahui bagaimana proses algoritma ini berjalan. Kemudian penulis menguji percobaan seluruh data mahasiswa baru yang telah melalui tahap *pre-processing* pada aplikasi *Rapidminer* hal ini bertujuan untuk mengukur tingkat akurasi antara hasil dari aplikasi yang sudah dibuat dengan *Rapidminer*. Aplikasi *K-means Clustering* yang akan dibuat menggunakan bahasa pemrograman PHP. Penulis menguji percobaan dengan beberapa tahapan dalam melakukan algoritma *K-means Clustering* hal ini bertujuan

untuk membuktikan bahwa aplikasi *K-Means* menggunakan PHP yang dibuat oleh penulis mempunyai hasil yang sama dengan algoritma *K-means Clustering* yang diterapkan oleh *Rapidminer*.

1. Penerapan Algoritma Menggunakan Perhitungan Manual

Pada tahapan ini penulis menerapkan algoritma *K-means Clustering* dengan metode perhitungan manual. Tahapan ini menggunakan 10 sample data untuk perhitungan 10 data tersebut dapat dilihat pada tabel 3.8 berikut :

Tabel 3.8 10 Sample Data Untuk Perhitungan

No	Jenis Kelamin	Kabupaten/Kota	Jenis Sekolah
1	1	9	2
2	2	1	1
3	1	4	1
4	1	1	5
5	1	12	5
6	1	4	4
7	1	12	1
8	1	1	2
9	1	1	3
10	1	1	3

Untuk melakukan perhitungan algoritma *K-means Clustering*, ada beberapa tahapan proses yaitu sebagai berikut :

a. Menentukan Jumlah Cluster

Menentukan jumlah cluster yang digunakan pada data calon mahasiswa sebanyak 2 cluster yaitu diantaranya Wilayah yang berpotensi dan Wilayah yang kurang berpotensi peminatnya berdasarkan data mahasiswa baru program studi teknik informatika 3 tahun terakhir (2020,2021,2022).

b. Menentukan Nilai Titik Pusat Cluster

Titik pusat cluster atau bisa disebut centroid digunakan sebagai nilai pengurang untuk perhitungan jarak antara data ke setiap cluster atau disebut distance. Berikut titik pusat cluster yang sudah ditentukan secara acak.

Tabel 3.9 Titik Pusat Cluster Perhitungan Manual

Titik Awal Pusat (Centroid)	Jenis Kelamin	Kota / Kabupaten	Jenis Sekolah
Cluster 1	1	4	1
Cluster 2	1	4	4

c. Menghitung Jarak Data Ke Setiap Cluster

Setelah titik pusat awal cluster ditentukan, proses selanjutnya adalah menghitung jarak setiap cluster atau distance. Dalam menghitung nilai jarak euclidean distance dapat dilakukan dengan persamaan berikut :

$$d(P,Q) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

Keterangan :

P_x : Data pertama (diambil dari atribut pertama)

Q_x : Titik pusat cluster / centroid untuk atribut pertama

P_y : Data kedua (diambil dari atribut kedua)

Q_y : Titik pusat cluster / centroid untuk atribut kedua

P_z : Data ketiga (diambil dari atribut ketiga)

Q_z : Titik pusat cluster / centroid untuk atribut ketiga

Dari 10 data yang dijadikan sampel telah terpilih pusat awal cluster yaitu C1 (1,4,1) dan C2 (1,4,4), lalu dilakukan perhitungan jarak dari sisa sampel data dengan pusat cluster yang dimisalkan x,y,z, dimana x merupakan jenis kelamin, y merupakan kota/kabupaten asal dan z merupakan jenis

sekolah . Hal ini dilakukan agar cara perhitungan lebih mudah.

Tabel 3.10 10 Sample Data Menggunakan Permisalan

Mahasiswa	x	y	z
1	1	9	2
2	2	1	1
3	1	4	1
4	1	1	5
5	1	12	5
6	1	4	4
7	1	12	1
8	1	1	2
9	1	1	3
10	1	1	3

Menghitung *Euclidean distance* dari semua data ke setiap titik pusat pertama.

$$d(1,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1 - 1)^2 + (9 - 4)^2 + (2 - 1)^2} = 5,099$$

$$d(2,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(2 - 1)^2 + (1 - 4)^2 + (1 - 1)^2} = 3,162$$

$$d(3,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1 - 1)^2 + (4 - 4)^2 + (1 - 1)^2} = 0$$

$$d(4,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1 - 1)^2 + (1 - 4)^2 + (5 - 1)^2} = 5$$

$$d(5,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (12-4)^2 + (5-1)^2} = 8,944$$

$$d(6,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (4-4)^2 + (4-1)^2} = 3$$

$$d(7,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (12-4)^2 + (1-1)^2} = 8$$

$$d(8,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (1-4)^2 + (2-1)^2} = 3,162$$

$$d(9,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (1-4)^2 + (3-1)^2} = 3,605$$

$$d(10,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (1-4)^2 + (3-1)^2} = 3,605$$

Menghitung *Euclidean distance* dari semua data ke setiap titik pusat kedua.

$$d(1,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (9-4)^2 + (2-4)^2} = 5,385$$

$$d(2,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(2-1)^2 + (1-4)^2 + (1-4)^2} = 4,358$$

$$d(3,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (4-4)^2 + (1-4)^2} = 3$$

$$d(4,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (1-4)^2 + (5-4)^2} = 3,162$$

$$d(5,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (12-4)^2 + (5-4)^2} = 8,062$$

$$d(6,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (4-4)^2 + (4-4)^2} = 0$$

$$d(7,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (12-4)^2 + (1-4)^2} = 8,544$$

$$d(8,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (1-4)^2 + (2-4)^2} = 3,605$$

$$d(9,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (1-4)^2 + (3-4)^2} = 3,162$$

$$d(10,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (1-4)^2 + (3-4)^2} = 3,162$$

Setelah proses perhitungan jarak data ke setiap cluster selesai maka proses berikutnya adalah mengalokasikan data kedalam setiap cluster yang terbentuk.

Pengalokasian data tersebut berdasarkan hasil jarak antara data ke setiap cluster, apabila nilai jarak antara data pertama ke cluster 1 lebih kecil daripada nilai jarak antara data pertama ke cluster 2 maka data pertama masuk kedalam cluster 1. Pengalokasian data ini dilakukan bertujuan agar bisa menentukan titik pusat cluster baru pada proses selanjutnya. Hasil dari proses pengalokasian data ini dapat dilihat pada iterasi pertama yang terdapat dalam tabel 3.11 berikut :

Tabel 3.11 Hasil Iterasi Ke-1

Mahasiswa	C1	C2
1	5,099	5,385
2	3,162	4,358
3	0	3
4	5	3,162
5	8,944	8,062
6	3	0
7	8	8,544
8	3,162	3,605
9	3,605	3,162
10	3,605	3,162

Tabel diatas merupakan hasil perhitungan jarak iterasi ke-1 adapun tabel diatas menunjukkan bahwa pada tabel yang berwarna kuning menghasilkan cluster yang didapat, Sehingga hasil yang didapat seperti berikut :

Anggota C1 = {Mahasiswa 1, Mahasiswa 2, Mahasiswa 3, Mahasiswa 7, Mahasiswa 8}

Anggota C2 = { Mahasiswa 4, Mahasiswa 5, Mahasiswa 6, Mahasiswa 9, Mahasiswa 10}

d. Menghitung Titik Pusat Cluster Baru

Setelah proses pengalokasian data sudah selesai maka tahapan selanjutnya adalah menentukan titik cluster baru.

Penentuan centroid baru ini menggunakan metode yang hampir sama dengan menentukan titik pusat cluster pada tahapan sebelumnya, yang membedakan dari kedua tahapan tersebut adalah dari jumlah data yang digunakan. Dalam tahapan ini centroid dapat ditentukan dengan cara menghitung nilai rata – rata sesuai dengan data yang sudah dikelompokkan kedalam setiap cluster masing - masing (tahapan pengalokasian data). Dalam proses ini penulis mencari nilai rata – rata dengan rumus :

$$C(i) = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\sum x}$$

Keterangan :

X_1 = Nilai data record ke = 1

X_2 = Nilai data record ke = 2

$\sum x$ = jumlah data record

Pada cluster 1 terdapat 5 data mahasiswa yang dapat dilihat pada tabel 3.12 berikut :

Tabel 3.12 Cluster 1 iterasi ke-1

Mahasiswa	x	y	z	C1
1	1	9	2	5,099
2	2	1	1	3,162
3	1	4	1	0
7	1	12	1	8
8	1	1	2	3,162

Maka dapat dihitung nilai rata-rata (avg) dari cluster 1 sebagai berikut :

$$\begin{aligned} & \text{avg}(\text{Mahasiswa1x}+\text{Mahasiswa2x}+\text{Mahasiswa3x}+ \\ & \text{Mahasiswa7x}+\text{Mahasiswa8x}) \\ & = \text{avg}(1+2+1+1+1) = 1,2 \end{aligned}$$

$$\begin{aligned} & \text{avg}(\text{Mahasiswa1y}+\text{Mahasiswa2y}+\text{Mahasiswa3y}+ \\ & \text{Mahasiswa7y}+\text{Mahasiswa8y}) \\ & = \text{avg}(9+1+4+12+1) = 5,4 \end{aligned}$$

$$\begin{aligned} & \text{avg}(\text{Mahasiswa1z}+\text{Mahasiswa2z}+\text{Mahasiswa3z}+ \\ & \text{Mahasiswa7z}+\text{Mahasiswa8z}) \\ & = \text{avg}(2+1+1+1+2) = 1,4 \end{aligned}$$

Perhitungan nilai rata-rata pada cluster 1 diatas didapatkan centroid cluster baru yang bisa dilihat pada tabel 3.13 berikut :

Tabel 3.13 Centroid Cluster Baru (Cluster 1)

Centroid Cluster Baru			
Centroid Cluster	x	y	z
1	1,2	5,4	1,4

Pada cluster 2 terdapat 5 data mahasiswa yang dapat dilihat pada tabel 3.14 berikut :

Tabel 3.14 Cluster 2 iterasi ke-1

Mahasiswa	x	y	z	C2
4	1	1	5	3,162
5	1	12	5	8,062
6	1	4	4	0
9	1	1	3	3,162
10	1	1	3	3,162

Maka dapat dihitung nilai rata-rata (avg) dari cluster 2 sebagai berikut :

$$\begin{aligned} & \text{avg}(\text{Mahasiswa4x}+\text{Mahasiswa5x}+\text{Mahasiswa6x}+ \\ & \text{Mahasiswa9x}+\text{Mahasiswa10x}) \\ & = \text{avg}(1+1+1+1+1) = 1 \end{aligned}$$

$$\begin{aligned} & \text{avg}(\text{Mahasiswa4y}+\text{Mahasiswa5y}+\text{Mahasiswa6y}+ \\ & \text{Mahasiswa9y}+\text{Mahasiswa10y}) \\ & = \text{avg}(1+12+4+1+1) = 3,8 \end{aligned}$$

$$\begin{aligned} & \text{avg}(\text{Mahasiswa4z}+\text{Mahasiswa5z}+\text{Mahasiswa6z}+ \\ & \text{Mahasiswa9z}+\text{Mahasiswa10z}) \\ & = \text{avg}(5+5+4+3+3) = 4 \end{aligned}$$

Perhitungan nilai rata-rata pada cluster 2 diatas didapatkan centroid cluster baru yang bisa dilihat pada tabel 3.15 berikut :

Tabel 3.15 Centroid Cluster Baru (Cluster 2)

Centroid Cluster Baru			
Centroid Cluster	x	y	z
2	1	3,8	4

Sehingga didapatkan centroid 1 dan centroid 2 yang baru sebagaimana bisa dilihat pada tabel 3.16 berikut :

Tabel 3.16 Centroid Cluster Baru Iterasi ke-1

Centroid Cluster	x	y	z
1	1,2	5,4	1,4
2	1	3,8	4

e. Memverifikasi Titik Pusat Cluster

Proses ini adalah tahap untuk memverifikasi titik pusat cluster. Proses tersebut adalah proses untuk memverifikasi antara titik pusat cluster baru dengan titik pusat cluster yang lama, apabila kedua titik pusat cluster tersebut mempunyai nilai centroid yang berbeda maka proses K-means Clustering masih berlanjut dan akan dimulai lagi pada proses ke-2 yaitu (menghitung jarak data ke setiap cluster) dengan menggunakan nilai titik pusat cluster yang baru. Sedangkan jika nilai kedua centroid tersebut bernilai sama maka proses K-means Clustering berhenti sampai di tahap tersebut. Proses verifikasi ini dilakukan dengan tujuan untuk menentukan apakah proses K-means Clustering sudah selesai atau masih perlu dilakukan perulangan proses kembali.

Hitung *Euclidean distance* untuk iterasi ke-2 dari semua data ke setiap titik pusat pertama guna memverifikasi data.

$$d(1,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1 - 1,2)^2 + (9 - 5,4)^2 + (2 - 1,4)^2} = 3,655$$

$$d(2,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(2 - 1,2)^2 + (1 - 5,4)^2 + (1 - 1,4)^2} = 4,489$$

$$d(3,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1 - 1,2)^2 + (4 - 5,4)^2 + (1 - 1,4)^2} = 1,469$$

$$d(4,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1 - 1,2)^2 + (1 - 5,4)^2 + (5 - 1,4)^2} = 5,688$$

$$d(5,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1 - 1,2)^2 + (12 - 5,4)^2 + (5 - 1,4)^2} = 7,52$$

$$d(6,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1 - 1,2)^2 + (4 - 5,4)^2 + (4 - 1,4)^2} = 2,959$$

$$d(7,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1 - 1,2)^2 + (12 - 5,4)^2 + (1 - 1,4)^2} = 6,615$$

$$d(8,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1 - 1,2)^2 + (1 - 5,4)^2 + (2 - 1,4)^2} = 4,445$$

$$d(9,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1 - 1,2)^2 + (1 - 5,4)^2 + (3 - 1,4)^2} = 4,686$$

$$d(10,1) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1 - 1,2)^2 + (1 - 5,4)^2 + (3 - 1,4)^2} = 4,686$$

Hitung *Euclidean distance* untuk iterasi ke-2 dari semua data ke setiap titik pusat kedua guna memverifikasi data.

$$d(1,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1 - 1)^2 + (9 - 3,8)^2 + (2 - 4)^2} = 5,571$$

$$d(2,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(2-1)^2 + (1-3,8)^2 + (1-4)^2} = 4,223$$

$$d(3,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (4-3,8)^2 + (1-4)^2} = 3,006$$

$$d(4,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (1-3,8)^2 + (5-4)^2} = 2,973$$

$$d(5,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (12-3,8)^2 + (5-4)^2} = 8,260$$

$$d(6,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (4-3,8)^2 + (4-4)^2} = 0,2$$

$$d(7,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (12-3,8)^2 + (1-4)^2} = 8,731$$

$$d(8,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (1-3,8)^2 + (2-4)^2} = 3,44$$

$$d(9,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (1-3,8)^2 + (3-4)^2} = 2,973$$

$$d(10,2) = \sqrt{(P_x - Q_x)^2 + (P_y - Q_y)^2 + (P_z - Q_z)^2}$$

$$\sqrt{(1-1)^2 + (1-3,8)^2 + (3-4)^2} = 2,973$$

Hasil perhitungan iterasi ke-2 dapat dilihat pada tabel 3.17 berikut :

Tabel 3.17 Hasil Iterasi Ke-2

Mahasiswa	C1	C2
1	3,655	5,571
2	4,489	4,223
3	1,469	3,006
4	5,688	2,973
5	7,52	8,26
6	2,959	0,2
7	6,615	8,731
8	4,445	3,44
9	4,686	2,973
10	4,686	2,973

Tabel diatas merupakan hasil perhitungan jarak iterasi ke-2 adapun tabel diatas menunjukkan bahwa pada tabel yang berwarna kuning menghasilkan cluster yang didapat, Sehingga hasil yang didapat seperti berikut :

Anggota C1 = {Mahasiswa 1, Mahasiswa 3, Mahasiswa 5, Mahasiswa 7}

Anggota C2 = {Mahasiswa 2, Mahasiswa 4, Mahasiswa 6, Mahasiswa 8, Mahasiswa 9, Mahasiswa 10}

Pada cluster 1 terdapat 4 data mahasiswa yang dapat dilihat pada tabel 3.18 berikut :

Tabel 3.18 Cluster 1 iterasi ke-2

Mahasiswa	x	y	z	C1
1	1	9	2	3,655

Tabel 3.18 Cluster 1 iterasi ke-2 (Lanjutan)

Mahasiswa	x	y	z	C1
3	1	4	1	1,469
5	1	12	5	7,52
7	1	12	1	6,615

Maka dapat dihitung nilai rata-rata (avg) dari cluster 1 sebagai berikut :

$$\begin{aligned} & \text{avg}(\text{Mahasiswa1x} + \text{Mahasiswa3x} + \text{Mahasiswa5x} + \text{Mahasiswa7x}) \\ & = \text{avg}(1 + 1 + 1 + 1) = 1 \end{aligned}$$

$$\begin{aligned} & \text{avg}(\text{Mahasiswa1y} + \text{Mahasiswa3y} + \text{Mahasiswa5y} + \text{Mahasiswa7y}) \\ & = \text{avg}(9 + 4 + 12 + 12) = 9,25 \end{aligned}$$

$$\begin{aligned} & \text{avg}(\text{Mahasiswa1z} + \text{Mahasiswa3z} + \text{Mahasiswa5z} + \text{Mahasiswa7z}) \\ & = \text{avg}(2 + 1 + 5 + 1) = 2,25 \end{aligned}$$

Perhitungan nilai rata-rata pada cluster 1 diatas didapatkan centroid cluster baru yang bisa dilihat pada tabel 3.19 berikut :

Tabel 3.19 Centroid Cluster Baru (Cluster 1)

Centroid Cluster Baru			
Centroid Cluster	x	y	z
1	1	9,25	2,25

Pada cluster 2 terdapat 6 data mahasiswa yang dapat dilihat pada tabel 3.20 berikut :

Tabel 3.20 Cluster 2 iterasi ke-2

Mahasiswa	x	y	z	C2
2	2	1	1	4,223
4	1	1	5	2,973
6	1	4	4	0,2
8	1	1	2	3,44
9	1	1	3	2,973
10	1	1	3	2,973

Maka dapat dihitung nilai rata-rata (avg) dari cluster 2 sebagai berikut :

$$\begin{aligned} & \text{avg}(\text{Mahasiswa}2x + \text{Mahasiswa}4x + \text{Mahasiswa}6x + \text{Mahasiswa}8x + \text{Mahasiswa}9x + \text{Mahasiswa}10x) \\ & = \text{avg}(2 + 1 + 1 + 1 + 1 + 1) = 1,16 \end{aligned}$$

$$\begin{aligned} & \text{avg}(\text{Mahasiswa}2y + \text{Mahasiswa}4y + \text{Mahasiswa}6y + \text{Mahasiswa}8y + \text{Mahasiswa}9y + \text{Mahasiswa}10y) \\ & = \text{avg}(1 + 1 + 4 + 1 + 1 + 1) = 1,5 \end{aligned}$$

$$\begin{aligned} & \text{avg}(\text{Mahasiswa}2z + \text{Mahasiswa}4z + \text{Mahasiswa}6z + \text{Mahasiswa}8z + \text{Mahasiswa}9z + \text{Mahasiswa}10z) \\ & = \text{avg}(1 + 5 + 4 + 2 + 3 + 3) = 3 \end{aligned}$$

Perhitungan nilai rata-rata pada cluster 2 diatas didapatkan centroid cluster baru yang bisa dilihat pada tabel 3.21 berikut :

Tabel 3.21 Centroid Cluster Baru (Cluster 2)

Centroid Cluster Baru			
Centroid Cluster	x	y	z
2	1,16	1,5	3

Sehingga didapatkan centroid 1 dan centroid 2 yang baru sebagaimana bisa dilihat pada tabel 3.22 berikut :

Tabel 3.22 Centroid Cluster Baru iterasi ke -2

Centroid Cluster	x	y	z
1	1	9,25	2,25
2	1,16	1,5	3

Setelah diketahui hasil iterasi ke-2 maka dapat dibandingkan titik pusat iterasi ke-2 dengan titik pusat sebelumnya untuk mengetahui kedua titik pusat cluster tersebut mempunyai nilai centroid yang berbeda atau sama. Apabila berbeda maka proses K-means Clustering masih berlanjut Sedangkan jika nilai kedua centroid tersebut bernilai sama maka proses K-means Clustering berhenti sampai di tahap tersebut. Perbandingan titik pusat iterasi ke-2 dengan sebelumnya dapat dilihat pada tabel 3.23 berikut :

Tabel 3.23 Perbandingan Titik Pusat Iterasi (1 dan 2)

Centroid Baru Iterasi Ke-1			
Centroid	x	y	z
1	1,2	5,4	1,4
2	1	3,8	4
Centroid Baru Iterasi Ke-2			
Centroid	x	y	z
1	1	9,25	2,25
2	1,16	1,5	3

Berdasarkan tabel 3.23 diatas tersebut, hasil dari perbandingan titik pusat iterasi 1 dan 2 berbeda, maka proses algoritma K-means Clustering akan tetap berjalan sampai dengan nilai kedua centroid tersebut bernilai sama.

Pada iterasi-iterasi berikutnya sampai proses algoritma K-means Clustering ini berhenti bisa dilihat pada tabel-tabel berikut :

Tabel 3.24 Hasil Iterasi ke-3

Mahasiswa	C1	C2	Cluster
1	0,35355	7,56821	1
2	8,40387	2,22361	2
3	5,39676	3,2059	2
4	8,69626	2,06828	2
5	3,88909	10,6901	1
6	5,53399	2,69774	2
7	3,02076	10,6901	1
8	8,25379	1,13039	2
9	8,28402	0,52705	2
10	8,28402	0,52705	2

Tabel 3.25 Isi Data Cluster Sesuai Iterasi ke-3

Mahasiswa	Cluster 1			Cluster 2		
	x	y	z	x	y	z
1	1	9	2	0	0	0
2	0	0	0	2	1	1
3	0	0	0	1	4	1
4	0	0	0	1	1	5
5	1	12	5	0	0	0
6	0	0	0	1	4	4
7	1	12	1	0	0	0
8	0	0	0	1	1	2
9	0	0	0	1	1	3
10	0	0	0	1	1	3

Tabel 3.26 Centroid Cluster Baru Iterasi ke -3

Centroid Cluster Baru			
Centroid	x	y	z
1	1	11	2,666

Tabel 3.26 Centroid Cluster Baru Iterasi ke -3 (Lanjutan)

Centroid Cluster Baru			
Centroid	x	y	z
2	1,142	1,857	2,714

Tabel 3.27 Perbandingan Titik Pusat Iterasi (2 dan 3)

Centroid Iterasi Ke-2			
Centroid	x	y	z
1	1	9,25	2,25
2	1,16	1,5	3
Centroid Iterasi Ke-3			
Centroid	x	y	z
1	1	11	2,666
2	1,142	1,857	2,714

Berdasarkan tabel-tabel diatas tersebut, hasil dari perbandingan titik pusat iterasi 2 dan 3 berbeda, maka proses algoritma K-means Clustering akan tetap berjalan sampai dengan nilai kedua centroid tersebut bernilai sama. Pada iterasi-iterasi berikutnya sampai proses algoritma K-means Clustering ini berhenti bisa dilihat pada tabel-tabel berikut :

Tabel 3.28 Hasil Iterasi ke-4

Mahasiswa	C1	C2	Cluster
1	2,10819	7,1799	1
2	10,1871	2,09956	2
3	7,19568	2,74791	2
4	10,2686	2,44532	2
5	2,53859	10,3982	1
6	7,12585	2,50306	2
7	1,94365	10,2877	1
8	10,0222	1,12486	2
9	10,0056	0,91473	2
10	10,0056	0,91473	2

Tabel 3.29 Isi Data Cluster Sesuai Iterasi Ke-4

	Cluster 1			Cluster 2		
Mahasiswa	x	y	z	x	y	z
1	1	9	2	0	0	0
2	0	0	0	2	1	1
3	0	0	0	1	4	1
4	0	0	0	1	1	5
5	1	12	5	0	0	0
6	0	0	0	1	4	4
7	1	12	1	0	0	0
8	0	0	0	1	1	2
9	0	0	0	1	1	3
10	0	0	0	1	1	3

Tabel 3.30 Centroid Cluster Baru iterasi ke -4

Centroid Cluster Baru			
Centroid	x	y	z
1	1	11	2,666
2	1,142	1,857	2,714

Tabel 3.31 Perbandingan Titik Pusat Iterasi (3 dan 4)

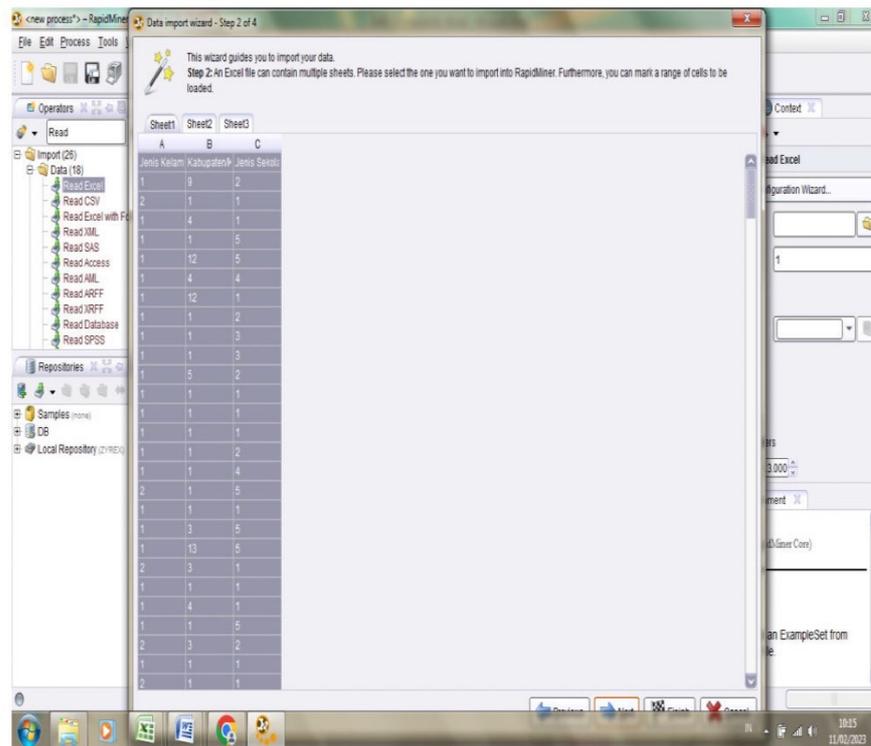
Centroid Iterasi Ke-3			
Centroid	x	y	z
1	1	11	2,666
2	1,142	1,857	2,714
Centroid Iterasi Ke-4			
Centroid	x	y	z
1	1	11	2,666
2	1,142	1,857	2,714

Berdasarkan tabel-tabel diatas tersebut, hasil dari perbandingan titik pusat iterasi 3 dan 4 sama, maka proses algoritma K-means Clustering ini di hentikan. Hasil dari perhitungan manual ini mendapatkan hasil 2 cluster,

dimana cluster 1 berjumlah 3 record data, cluster 2 berjumlah 7 record data pada tiap tiap cluster.

2. Penerapan Algoritma Menggunakan *Rapidminer*

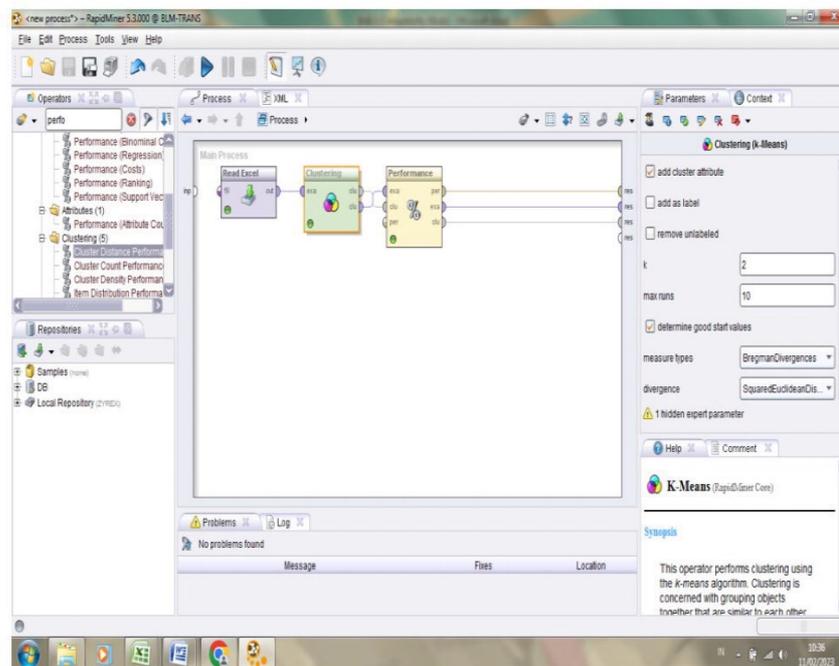
Rapidminer adalah perangkat lunak untuk pengolahan data dengan menggunakan prinsip dan algoritma data *mining*. *Rapidminer* mengekstrak pola-pola dataset yang besar dengan mengkombinasi metode statistika, kecerdasan buatan dan database. Data yang sudah melalui proses *pre-processing* data selanjutnya akan diolah menggunakan aplikasi *Rapidminer*. Pada gambar 3.3 berikut adalah proses melakukan import data ke aplikasi *Rapidminer*.



Gambar 3.3 Proses Impor Dataset Transformasi Data ke *Rapidminer*

Pada gambar 3.3 adalah proses import data ke *Rapidminer*, penulis menggunakan operator read excel dengan fitur import

configuration wizard. Fitur tersebut digunakan untuk memasukkan file data serta memilih tipe data atribut dan role. Selanjutnya tahap berikutnya ialah tahap permodelan desain clustering menggunakan algoritma *K-means Clustering*, tahap desain tersebut dapat dilihat pada gambar 3.4 berikut :



Gambar 3.4 Proses Tahap Desain Algoritma *K-means Clustering*

Proses pada gambar 3.4 diatas dapat dilihat bahwa proses pengujian dan percobaan clustering dimulai dengan operator read excel data transformasi data mahasiswa baru Program Studi Teknik Informatika yang bersumber dari PMB IIB Darmajaya, data tersebut berjumlah 279 data setelah tahap pre-processing data, data ini dibuka dengan fitur Import Configuration Wizard, fitur ini digunakan untuk memasukkan data file ke dalam *Rapidminer*. Setelah dataset itu di import kemudian dihubungkan untuk dilakukan pengujian dan percobaan menggunakan algoritma *K-means Clustering*. Setelah dihubungkan, selanjutnya dihubungkan lagi dengan operator cluster distance performance

untuk menghasilkan output dari algoritma clustering. Setelah pengujian berjalan maka didapatkan informasi lama waktu pemrosesan clustering seperti yang ditunjukkan pada gambar 3.5 berikut :



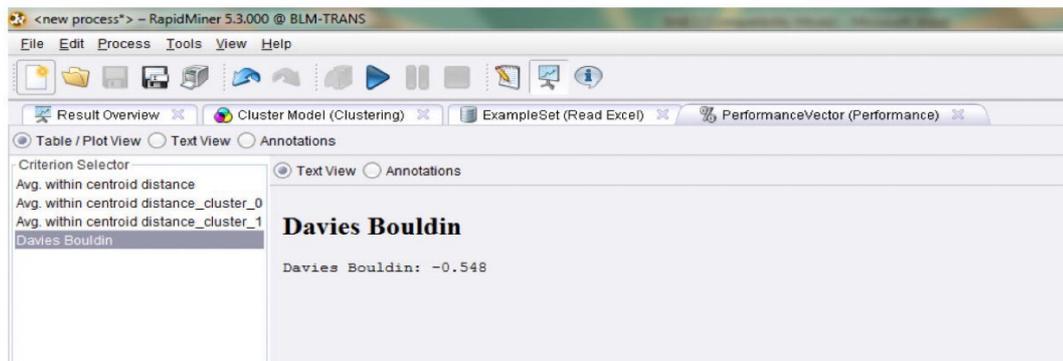
Gambar 3.5 *Result History* Algoritma *K-means Clustering*

Dalam proses pengujian dan percobaan ini penulis mencoba dengan jumlah $k=2$ dan $\text{max run} = 10$ sehingga dapat menghasilkan centroid akhir yang bisa dilihat pada gambar 3.6 berikut :

Attribute	cluster_0	cluster_1
Jenis Kelam	1.266	1.195
Kabupaten/h	10.562	2.093
Jenis Sekolah	2.172	2.130

Gambar 3.6 Centroid akhir menggunakan *Rapidminer*

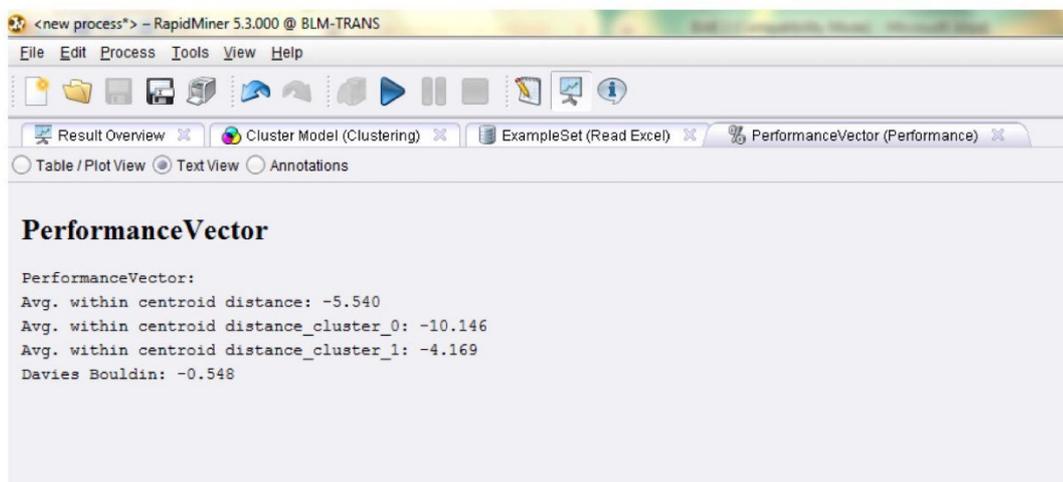
Dalam memvalidasi penelitian ini, penulis menggunakan Davies Bouldin Index (DBI) sehingga bisa dilihat pada gambar 3.7 berikut :



Gambar 3.7 DBI Algoritma *K-means Clustering* Menggunakan *Rapidminer*

Dari gambar 3.7 diatas menunjukkan bahwa hasil DBI tersebut -0,548 sehingga belum cukup baik karena nilai tersebut negatif ≤ 0 .

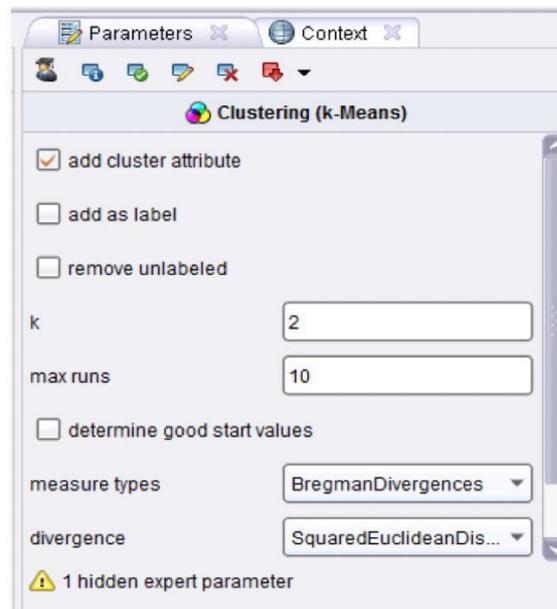
Untuk menunjukkan deskripsi *performance* yang menghasilkan *centroid distance* pada setiap *cluster* dapat dilihat pada gambar 3.8 berikut :



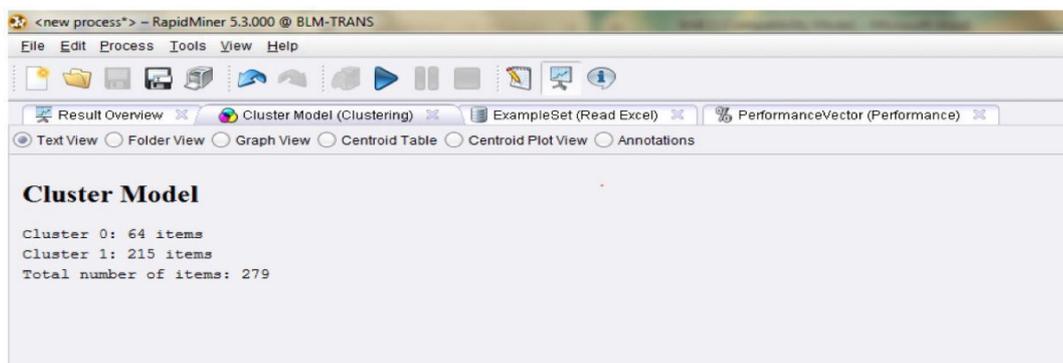
Gambar 3.8 Hasil DBI dan *Centroid Distance Performance Vector*

Pada gambar 3.10 menunjukkan bahwa penelitian ini cluster yang dihasilkan berjumlah 2 cluster , dimulai dari cluster 0 dan diakhiri dengan cluster 1 . Total dataset yang dicluster berjumlah 279 data dengan jumlah items pada tiap tiap cluster . Untuk

pengaturan jumlah cluster yang diinginkan dapat dilihat pada gambar 3.9 berikut :

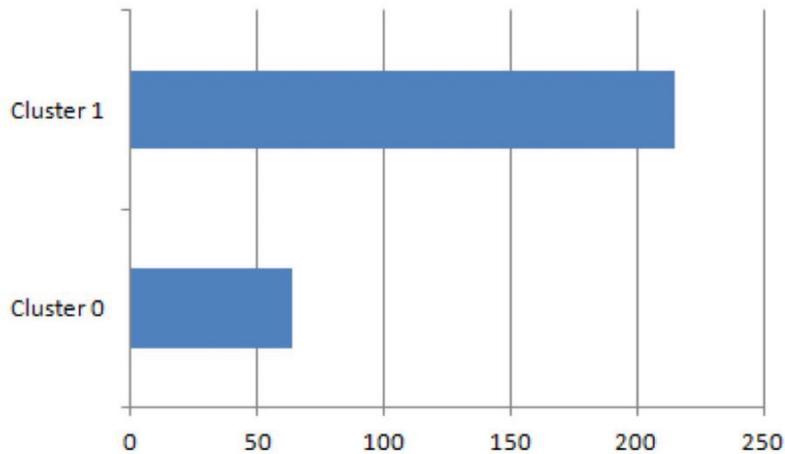


Gambar 3.9 Penentuan Jumlah Cluster



Gambar 3.10 Hasil Cluster Algoritma *K-means Clustering*

Pada gambar 3.10 menunjukkan bahwa penelitian menggunakan algoritma *K-means Clustering* menghasilkan cluster 0 berjumlah 64 items, jumlah cluster 1 berjumlah 215 items pada tiap tiap cluster. Berikut adalah hasil dari cluster model *K-means Clustering* yang digambarkan dalam bentuk chart bar yang dapat dilihat pada gambar 3.11 berikut :



Gambar 3.11 Tampilan *Chart Bar* Algoritma *K-means Clustering*

3. Penerapan Algoritma Menggunakan Aplikasi Berbasis Web

Setelah dilakukan proses manual menggunakan Aplikasi microsoft excel dan menerapkan algoritma *K-means Clustering* di aplikasi *Rapidminer*, maka penulis bermaksud untuk melakukan pengujian ulang dengan menggunakan aplikasi yang dibuat. Hal ini bertujuan sebagai perbandingan antara proses manual, proses penerapan *Rapidminer* dengan menggunakan aplikasi yang dibuat dan juga untuk pembuktian apakah algoritma *K- Means Clustering* telah diterapkan dengan benar pada proses manual ataupun sebaliknya. Dengan demikian akan terlihat apakah ada perbedaan antara proses manual dengan proses menggunakan aplikasi.

Penerapan algoritma *K-means Clustering* yang ketiga ini penulis menggunakan aplikasi yang dibuat dengan bahasa pemrograman PHP. Aplikasi ini adalah berbasis website dengan menggunakan database *MySQL* sebagai penyimpanan data. Sistem yang digunakan dalam penerapan ini menggunakan metode algoritma *K-means Clustering*. Sebelum aplikasi berbasis website ini dibuat, penulis melalui beberapa tahapan dalam pembuatannya mulai dari perancangan, percodingan sampai aplikasi ini berhasil menerapkan apa yang sudah dirancang.

Beberapa tahapan perancangan proses pembuatan aplikasi ini yaitu sebagai berikut:

a. **Perancangan Sistem Aplikasi *K-means Clustering***

Untuk membuat sebuah sistem aplikasi berbasis web penulis menggunakan aplikasi database MySQL sebagai tempat untuk menampung dan menyimpan sebuah data, aplikasi Sublime Text untuk proses pengcodingan dengan berbagai perintah, fungsi yang dibutuhkan untuk membangun sistem dengan algoritma *K-means Clustering*, dan sebuah browser untuk menampilkan hasil pengcodingan.

Perancangan sistem yang akan dibuat menggunakan UML (*Unified Modelling Language*) seperti : *Use Case Diagram*, *Activity Diagram*, *Class Diagram* dimana UML tersebut adalah suatu metode dalam pemodelan secara visual yang digunakan sebagai sarana perancangan sistem berorientasi objek. Adapun tahapan masing-masing perancangan sistem sebagai berikut :

1. *Use Case Diagram*

Use case diagram adalah gambaran sistem yang digunakan untuk merancang sistem aplikasi data mining menggunakan algoritma *K-means Clustering* berbasis web. Rancangan menggunakan use case diagram ini dapat dilihat pada gambar 3.12 berikut :



Gambar 3.12 Use Case Diagram

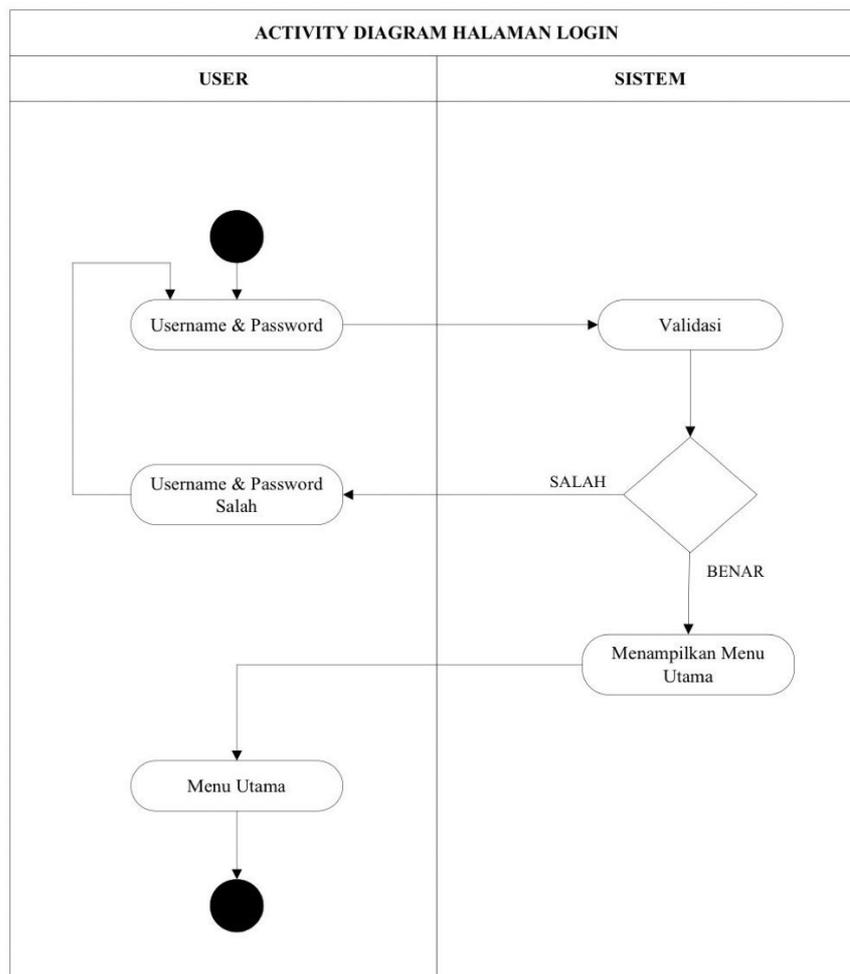
2. Activity Diagram

Activity Diagram yang dikembangkan antara lain adalah sebagai berikut :

a. Activity Diagram Login

- b. *Activity Diagram* Halaman Data Admin
- c. *Activity Diagram* Halaman Data Atribut
- d. *Activity Diagram* Halaman Data Cluster
- e. *Activity Diagram* Halaman Dataset Mahasiswa Baru
- f. *Activity Diagram* Halaman Dataset Setelah Tahapan *Pre-Processing*
- g. *Activity Diagram* Halaman Data Inisialisasi
- h. *Activity Diagram* Halaman Hasil

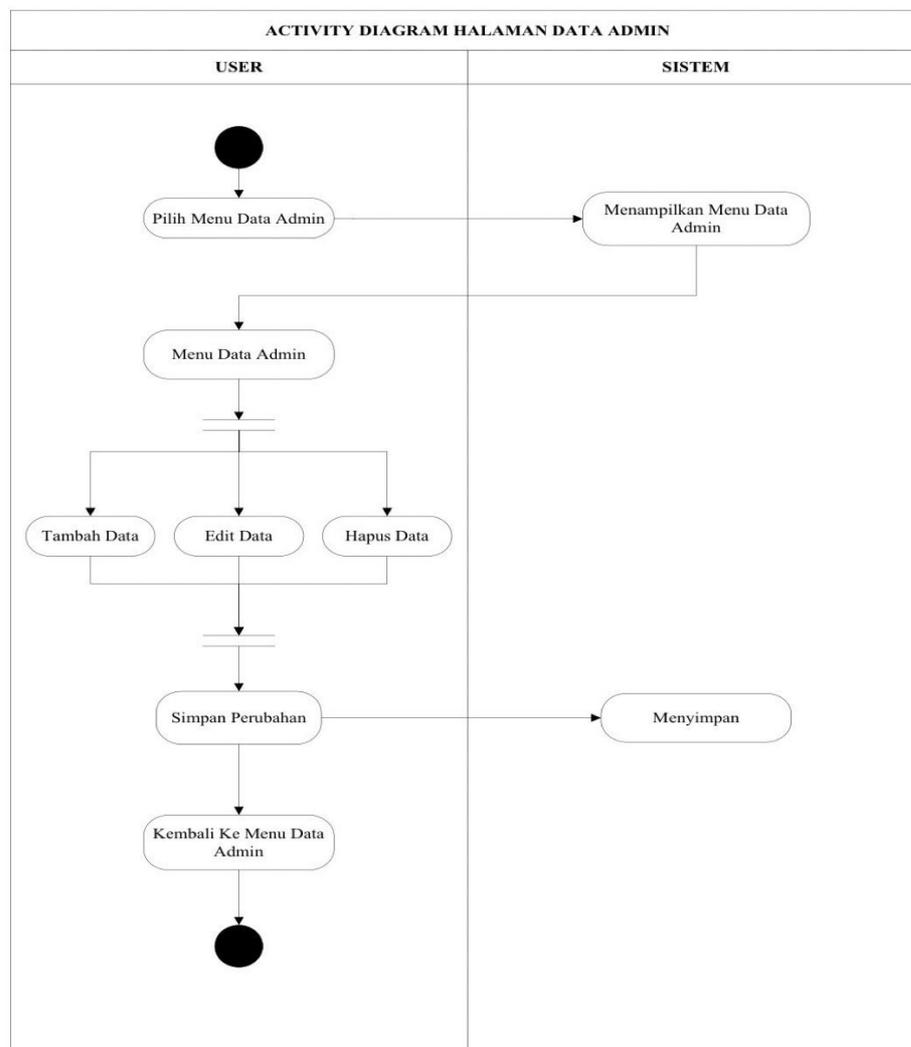
Activity Diagram tersebut bisa dilihat pada gambar- gambar berikut:



Gambar 3.13 *Activity Diagram* Login

Keterangan gambar 3.13 :

1. User memasukkan username dan password.
2. Sistem akan memvalidasi, jika benar sistem akan menampilkan menu utama, jika salah menampilkan informasi password salah dan user memasukkan kembali username dan password.

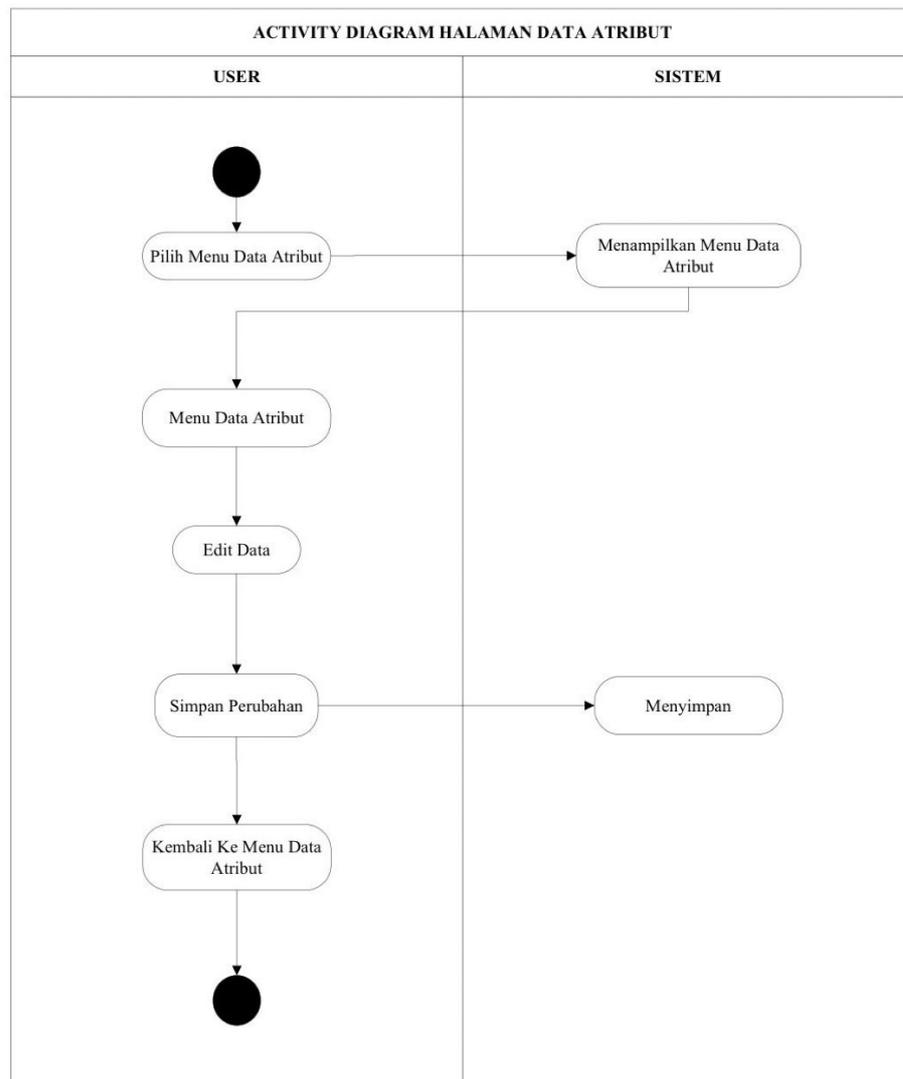


Gambar 3. 14 Activity Diagram Halaman Data Admin

Keterangan gambar 3.14 :

1. Admin memilih menu data admin.

2. Sistem akan menampilkan data admin.
3. Kemudian admin bisa melakukan edit data, tambah data dan menghapus data admin.
4. Setelah itu sistem akan menyimpan perubahan dan kembali menampilkan data admin.

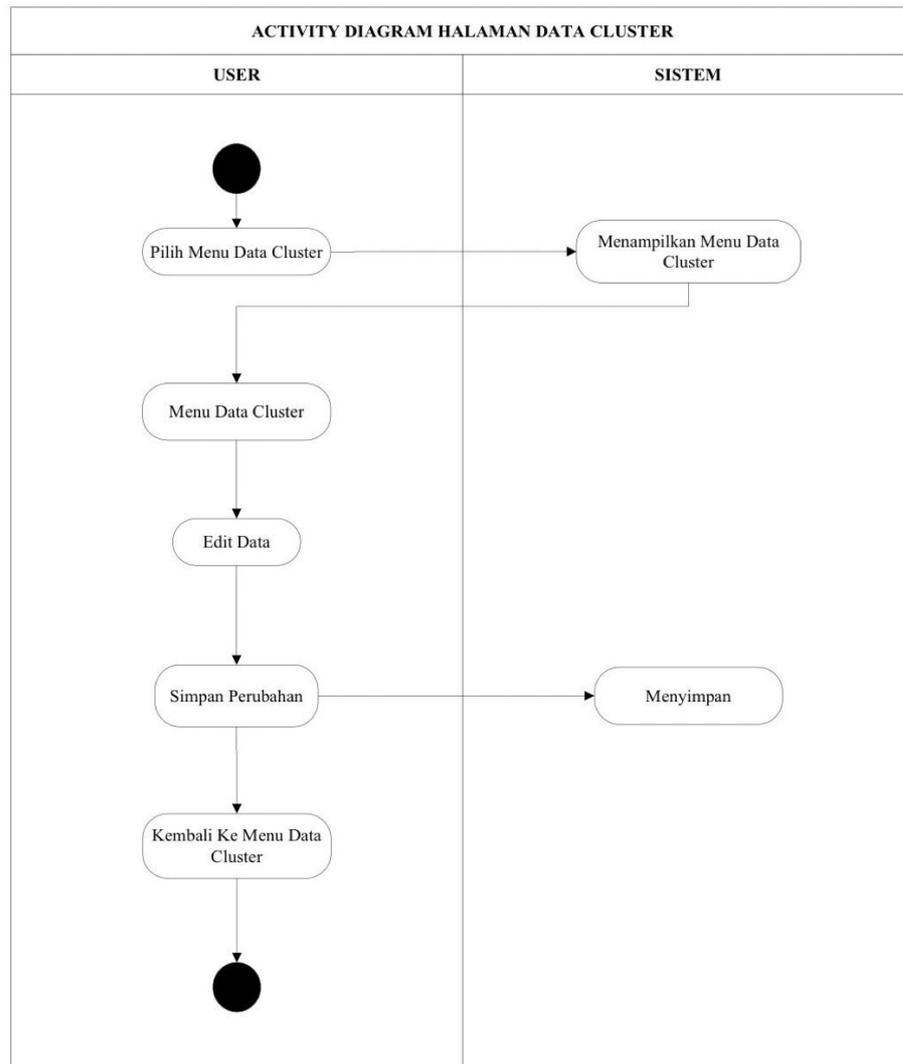


Gambar 3.15 Activity Diagram Halaman Data Atribut

Keterangan gambar 3.15 :

1. Admin memilih menu data atribut.
2. Sistem akan menampilkan data atribut.

3. Kemudian admin bisa melakukan edit data pada data atribut.
4. Setelah itu sistem akan menyimpan perubahan dan kembali menampilkan data atribut.

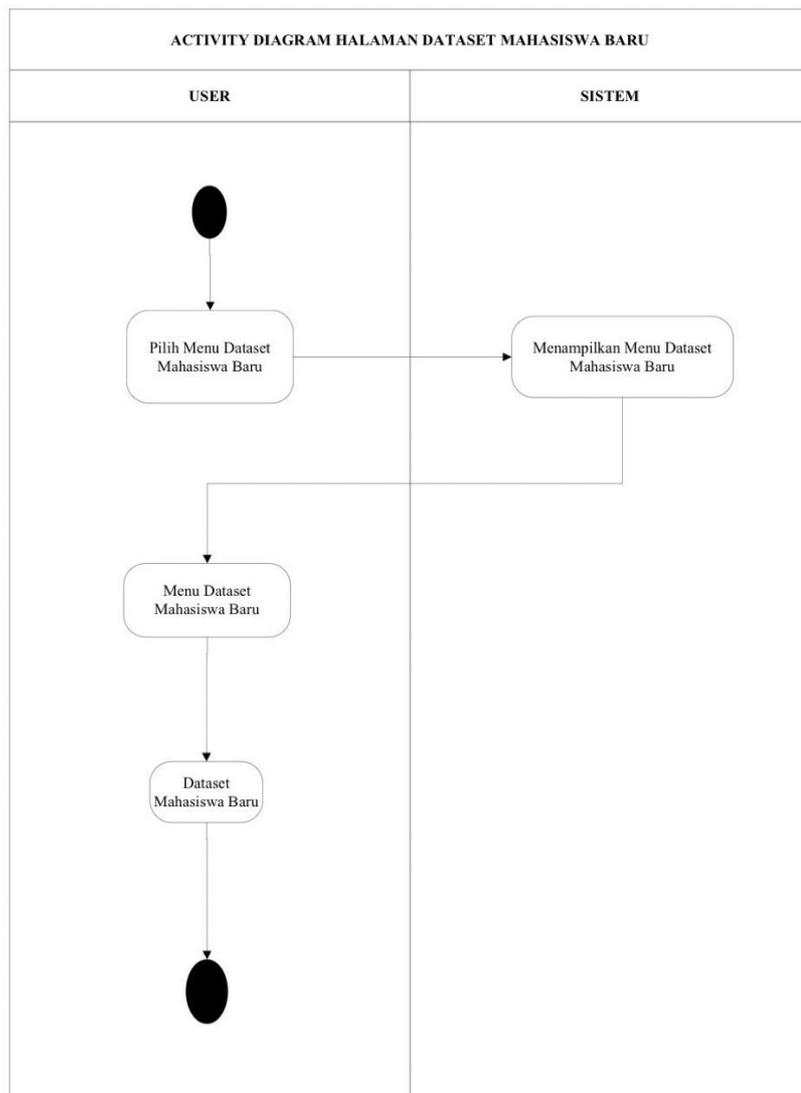


Gambar 3.16 Activity Diagram Halaman Data Cluster

Keterangan gambar 3.16 :

1. Admin memilih menu data cluster.
2. Sistem akan menampilkan data cluster.

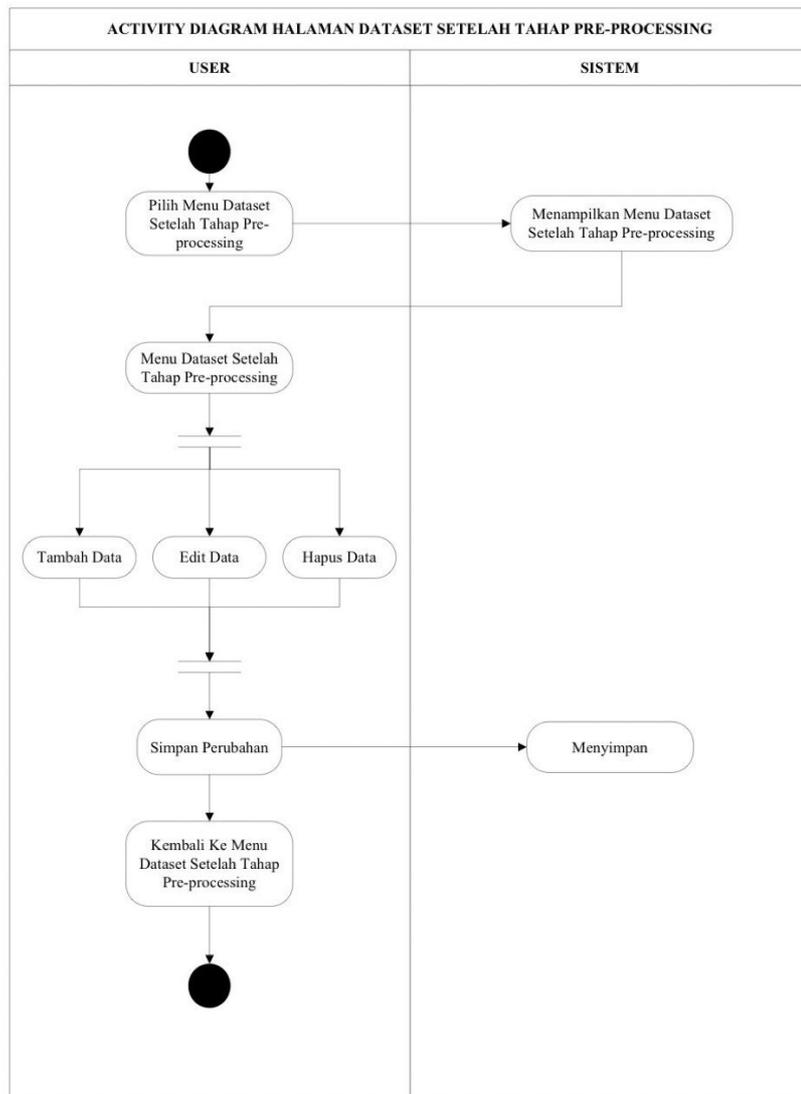
3. Kemudian admin bisa melakukan edit data pada data cluster.
4. Setelah itu sistem akan menyimpan perubahan dan kembali menampilkan data cluster.



Gambar 3.17 Activity Diagram Halaman Dataset Mahasiswa Baru

Keterangan gambar 3.17 :

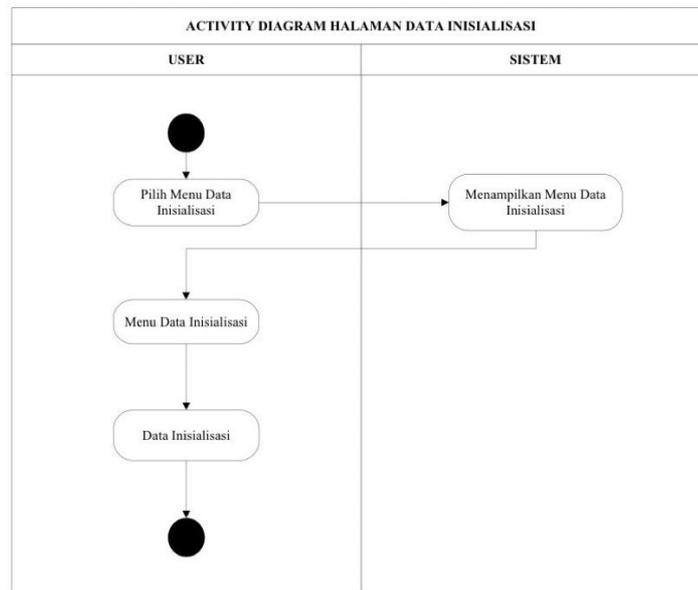
1. Admin memilih menu dataset mahasiswa baru.
2. Sistem akan menampilkan dataset mahasiswa baru.



Gambar 3.18 Activity Diagram Halaman Dataset Setelah Tahap *Pre-Processing*

Keterangan gambar 3.18 :

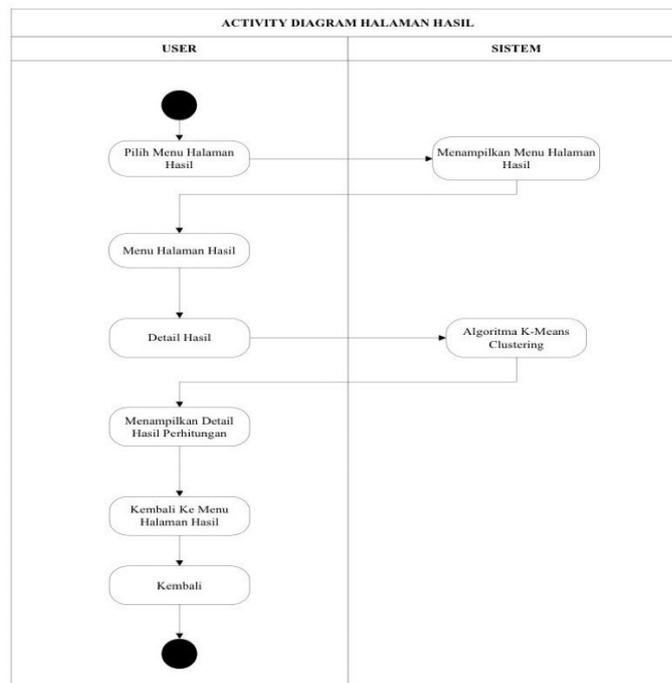
1. Admin memilih menu dataset setelah tahap pre-processing.
2. Sistem akan menampilkan dataset setelah tahap pre-processing.
3. Kemudian admin bisa melakukan edit data, tambah data dan menghapus data pada data mahasiswa baru.
4. Setelah itu sistem akan menyimpan perubahan dan kembali menampilkan data data mahasiswa baru.



Gambar 3.19 Activity Diagram Halaman Data Inisialisasi

Keterangan gambar 3.19 :

1. Admin memilih menu data data inisialisasi.
2. Sistem akan menampilkan data inisialisasi.



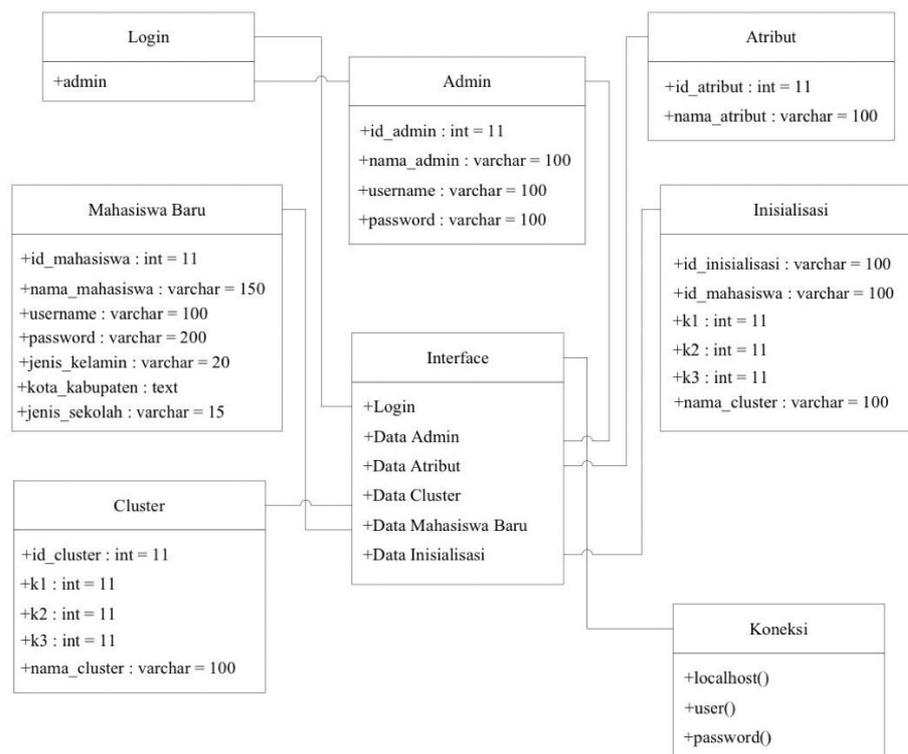
Gambar 3.20 Activity Diagram Halaman Hasil

Keterangan gambar 3.20 :

1. Admin memilih menu halaman hasil.
2. Sistem akan menampilkan menu halaman hasil.
3. Kemudian admin bisa melakukan melihat detail hasil.
4. Setelah itu sistem akan menerapkan algoritma K-means Clustering untuk menampilkan hasil.
5. Admin melihat tampilan detail hasil perhitungan.
6. Jika kembali maka admin kembali ke menu halaman hasil.

3. Class Diagram

Class Diagram pada perancangan sistem Algoritma *K-means Clustering* dapat dilihat pada gambar 3.21 berikut :



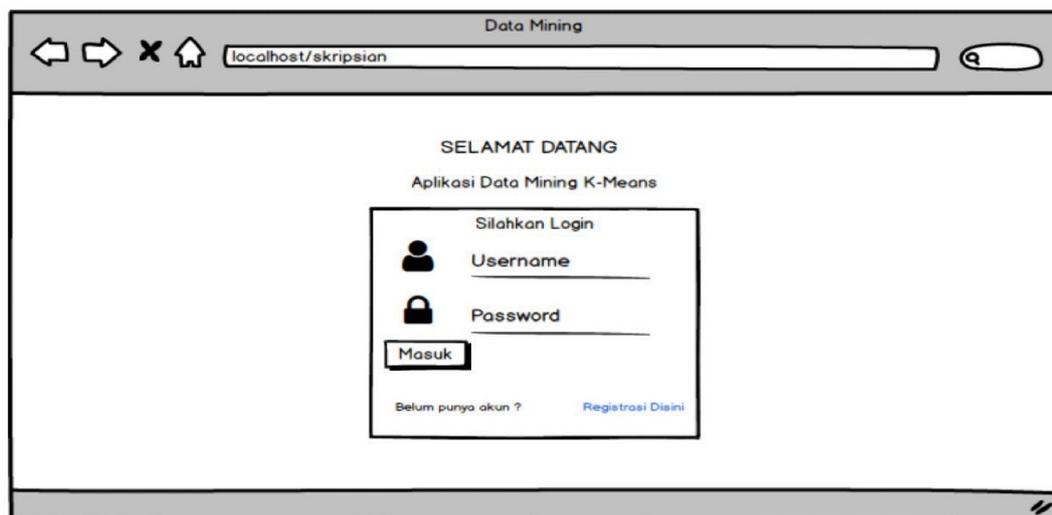
Gambar 3.21 *Class Diagram*

Setelah melakukan perancangan dengan UML (*Unified Modelling Language*) maka penulis melanjutkan perancangan *User*

Interface , dimana tahap ini adalah merancang tampilan website dan bahan yang dibutuhkan untuk membuat website tersebut. Adapun desain *interface* pada website ini dapat dilihat pada gambar gambar berikut :

1. Rancangan Form Login

Rancangan form login ini dimana user dapat memilih berbagai pilihan yang telah disediakan untuk kemudahan dalam menjalankan sistem. Rancangan form login tersebut bisa dilihat pada gambar 3.22 Berikut :



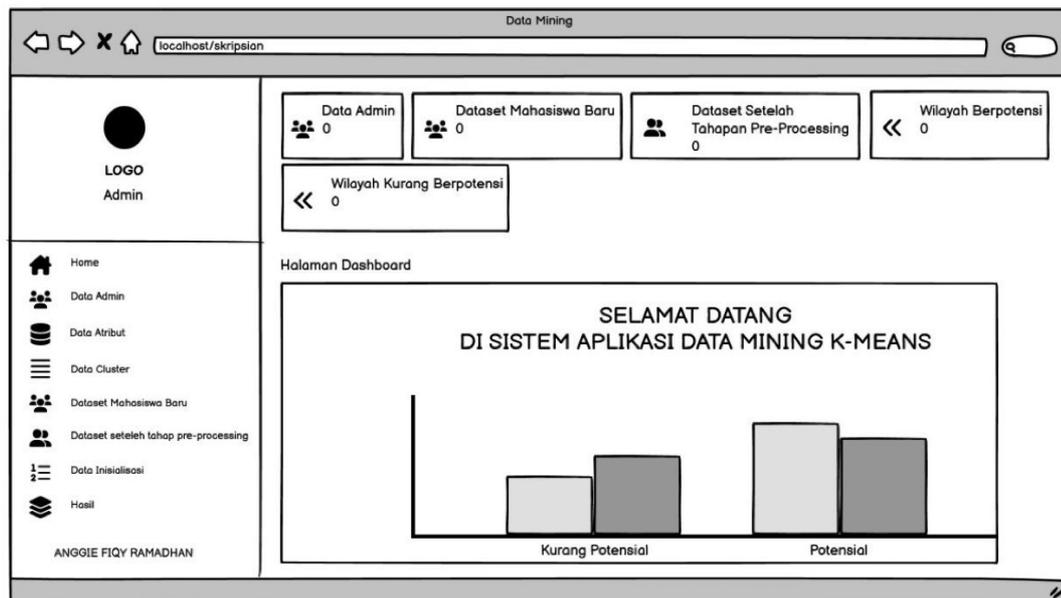
The image shows a web browser window with the title 'Data Mining'. The address bar contains 'localhost/skripsian'. The main content area displays a login form with the following elements:

- Header: SELAMAT DATANG
- Sub-header: Aplikasi Data Mining K-Means
- Form Title: Silahkan Login
- Username field: Username
- Password field: Password
- Login button: Masuk
- Footer: Belum punya akun? [Registrasi Di sini](#)

Gambar 3.22 *User Interface* Halaman Login

2. Rancangan Halaman Dashboard (home)

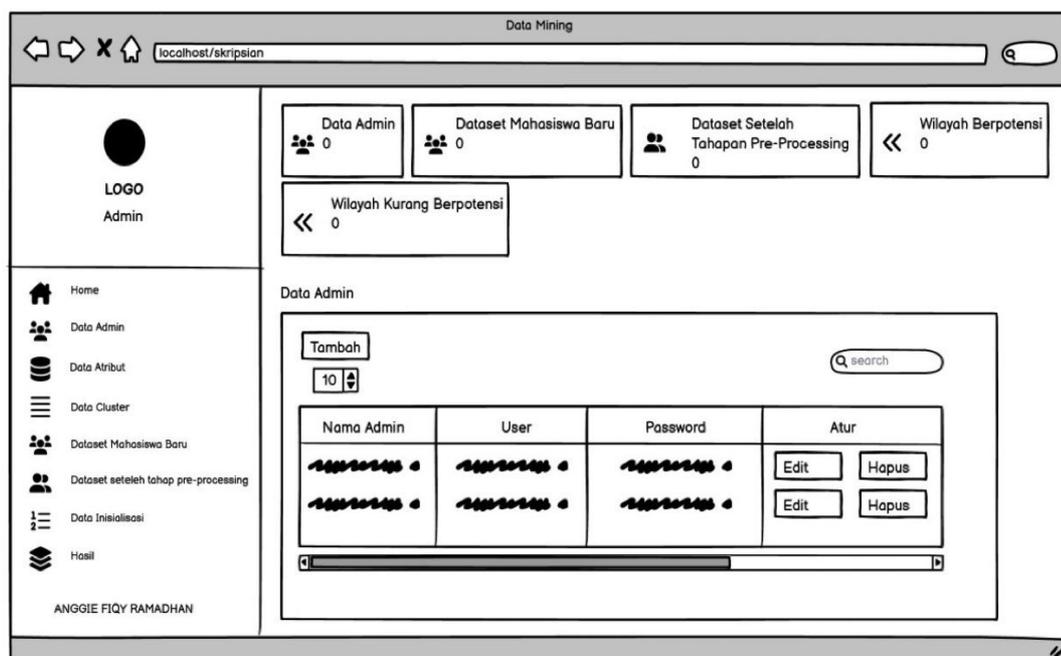
Rancangan halaman dashboard ini adalah halaman pertama kali dijalankan oleh program. Rancangan halaman dashboard ini bisa dilihat pada gambar 3.23 Berikut :



Gambar 3.23 *User Interface* Halaman Dashboard

3. Rancangan Data Admin

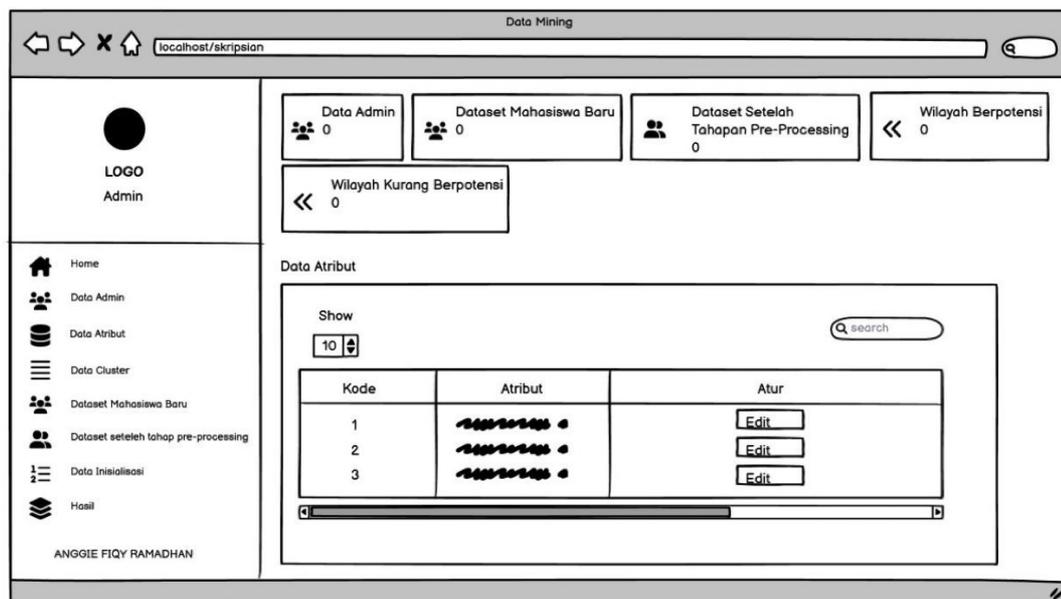
Rancangan data admin adalah dimana user (admin) bisa melihat data admin yang sudah masuk kedalam database. Data admin ini bisa dilihat pada gambar 3.24 berikut :



Gambar 3.24 *User Interface* Halaman Data Admin

4. Rancangan Data Atribut

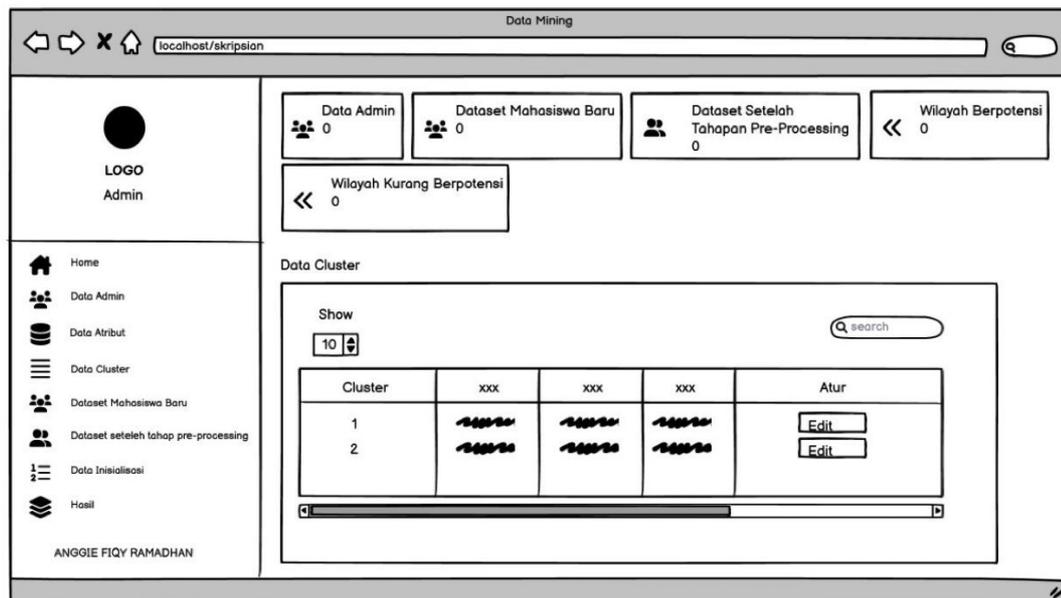
Rancangan data atribut adalah dimana user (admin) bisa melihat data atribut yang sudah masuk kedalam database. Data ini bisa diubah ubah tergantung kebutuhan . Data atribut ini bisa dilihat pada gambar 3.25 berikut :



Gambar 3.25 User Interface Halaman Data Atribut

5. Rancangan Data Cluster

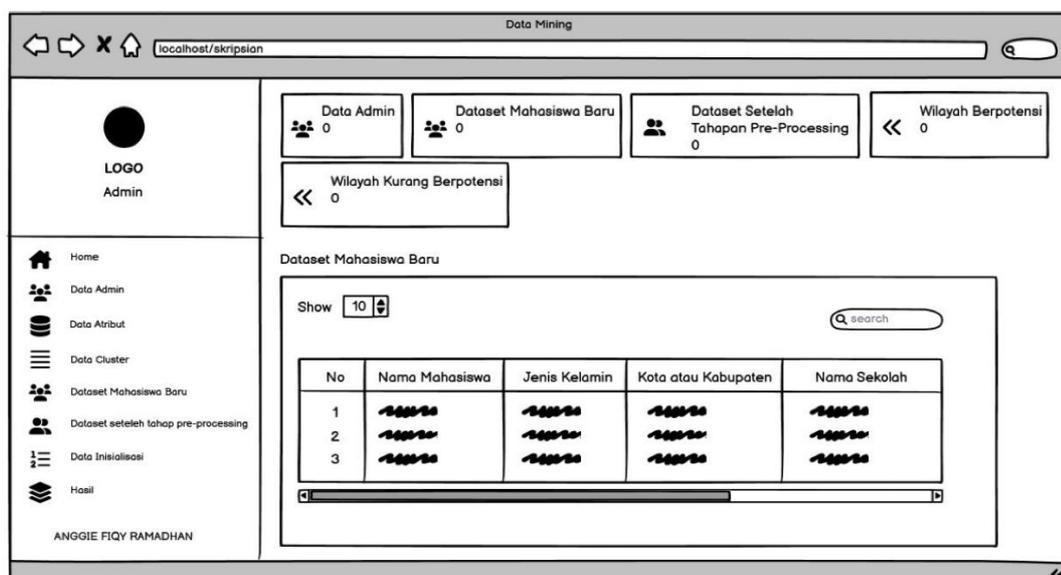
Rancangan data cluster adalah dimana user (admin) bisa melihat data cluster yang sudah masuk kedalam database. Data ini bisa diubah ubah tergantung kebutuhan . Data cluster ini bisa dilihat pada gambar 3.26 berikut :



Gambar 3.26 User Interface Halaman Data Cluster

6. Rancangan Data Halaman Dataset Mahasiswa Baru

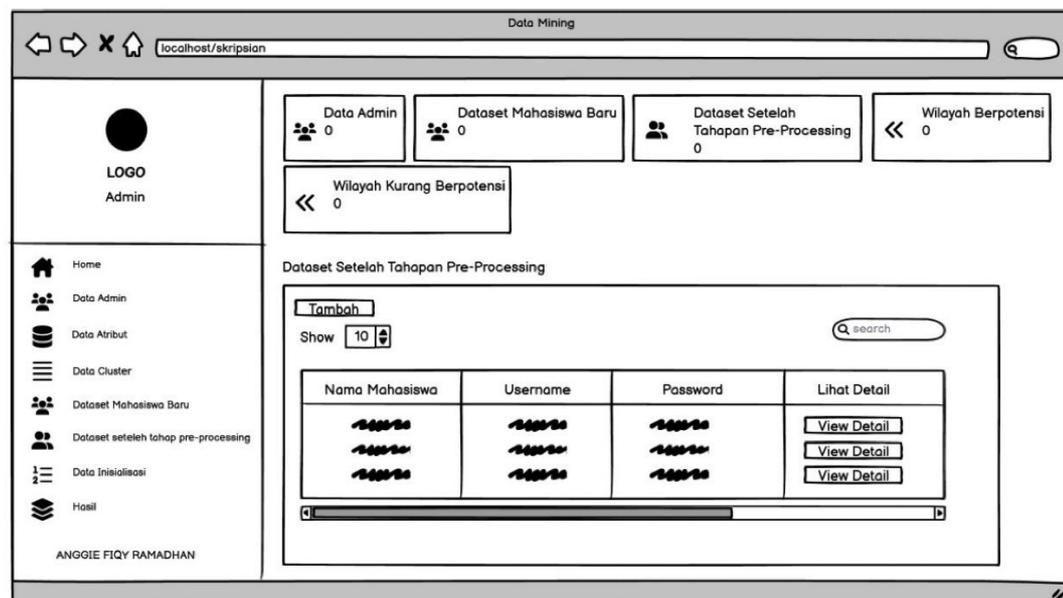
Rancangan data halaman dataset mahasiswa baru ini adalah dimana user (admin) bisa melihat dataset mahasiswa baru yang sudah masuk kedalam database. Dataset mahasiswa baru ini bisa dilihat pada gambar 3.27 berikut :



Gambar 3.27 User Interface Halaman Dataset Mahasiswa Baru

7. Rancangan Dataset Setelah Tahap *Pre-Processing*

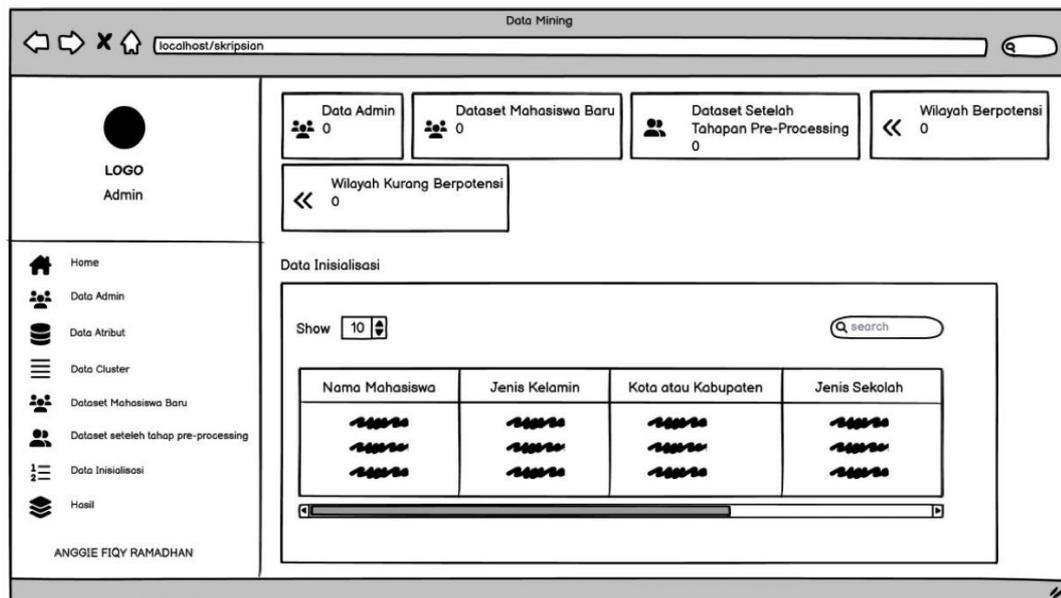
Rancangan data halaman dataset setelah tahap *pre-processing* ini adalah dimana user (admin) bisa melihat dataset setelah tahap *pre-processing* yang sudah masuk kedalam database. Data ini bisa diubah ubah tergantung kebutuhan . Dataset setelah tahap *pre-processing* ini bisa dilihat pada gambar 3.28 berikut :



Gambar 3.28 *User Interface* Halaman Dataset Mahasiswa Baru

8. Rancangan Data Inisialisasi

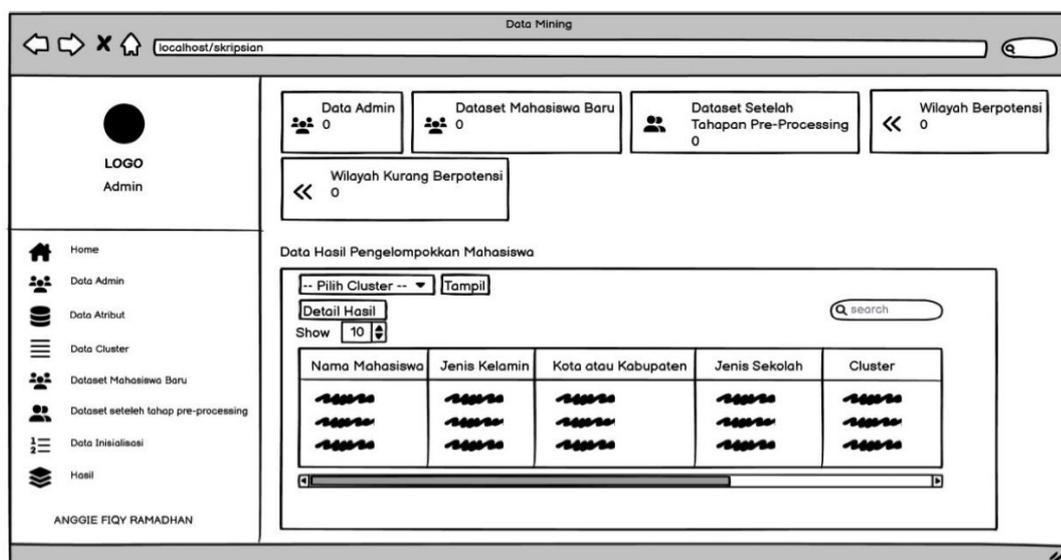
Rancangan data inisialisasi ini adalah dimana user (admin) bisa melihat data inisialisasi yang sudah masuk kedalam database. Data ini bisa diubah ubah tergantung kebutuhan . Data mahasiswa baru ini bisa dilihat pada gambar 3.29 berikut :



Gambar 3.29 User Interface Halaman Data Inisialisasi

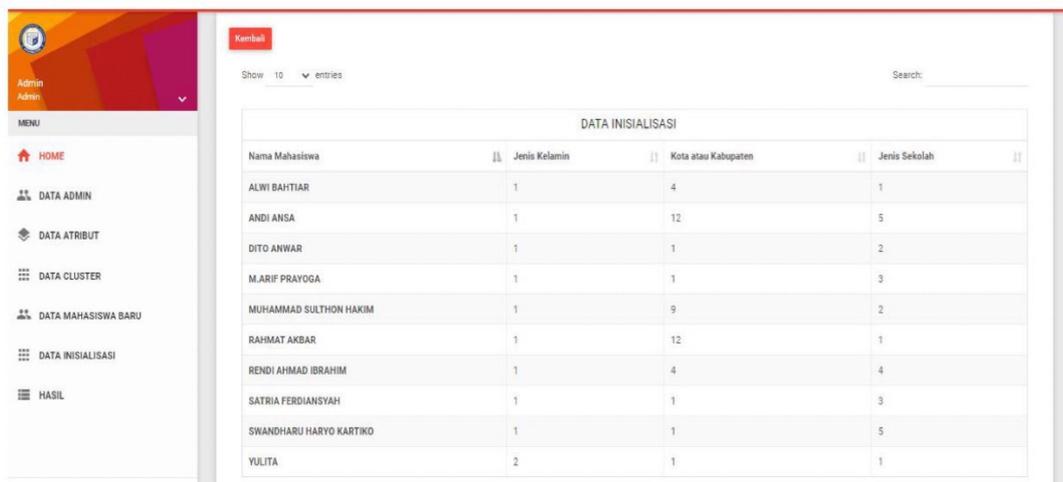
9. Rancangan Halaman Hasil

Rancangan data inisialisasi ini adalah dimana user (admin) bisa melihat data inisialisasi yang sudah masuk kedalam database. Data ini bisa diubah ubah tergantung kebutuhan . Data mahasiswa baru ini bisa dilihat pada gambar 3.30 berikut :



Gambar 3.30 User Interface Halaman Hasil

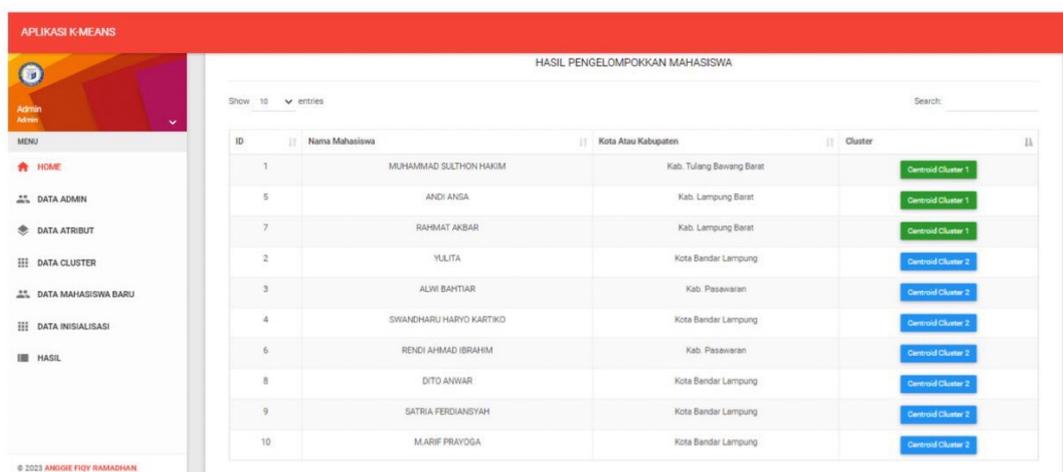
Setelah melakukan tahapan-tahapan perancangan dalam pembuatan aplikasi, maka pada tahapan berikutnya adalah menerapkan aplikasi yang dibuat dan dibangun dengan menggunakan metode algoritma *K-means Clustering*. Pada tahapan-tahapan penerapannya bisa dilihat pada gambar gambar berikut :



Nama Mahasiswa	Jenis Kelamin	Kota atau Kabupaten	Jenis Sekolah
ALWI BAHTIAR	1	4	1
ANDI ANSA	1	12	5
DITO ANWAR	1	1	2
M.ARIF PRAVOGA	1	1	3
MUHAMMAD SULTHON HAKIM	1	9	2
RAHMAT AKBAR	1	12	1
RENDI AHMAD IBRAHIM	1	4	4
SATRIA FERDIANSYAH	1	1	3
SWANDHARU HARYO KARTIKO	1	1	5
YULITA	2	1	1

Gambar 3.31 10 Sample Data Penerapan Sistem *K-means Clustering*

Gambar 3.31 diatas merupakan gambar 10 sample data penerapan sistem k-means, dalam hal ini penulis menguji percobaan apakah sistem ini berjalan dengan baik atau tidak.



ID	Nama Mahasiswa	Kota Atau Kabupaten	Cluster
1	MUHAMMAD SULTHON HAKIM	Kab. Tulang Bawang Barat	Centroid Cluster 1
5	ANDI ANSA	Kab. Lampung Barat	Centroid Cluster 1
7	RAHMAT AKBAR	Kab. Lampung Barat	Centroid Cluster 1
2	YULITA	Kota Bandar Lampung	Centroid Cluster 2
3	ALWI BAHTIAR	Kab. Pasewaran	Centroid Cluster 2
4	SWANDHARU HARYO KARTIKO	Kota Bandar Lampung	Centroid Cluster 2
6	RENDI AHMAD IBRAHIM	Kab. Pasewaran	Centroid Cluster 2
8	DITO ANWAR	Kota Bandar Lampung	Centroid Cluster 2
9	SATRIA FERDIANSYAH	Kota Bandar Lampung	Centroid Cluster 2
10	M.ARIF PRAVOGA	Kota Bandar Lampung	Centroid Cluster 2

Gambar 3.32 Hasil Pengelompokkan Data 10 Sample Data

Gambar 3.32 diatas merupakan gambar hasil pengelompokkan 10 sample data penerapan sistem *k-means*, setelah di uji percobaannya menggunakan 10 sample data yang sama halnya dengan perhitungan manual, maka penulis dapat menyimpulkan bahwa sistem yang dibuat telah berjalan dengan baik dan benar .

3.5 Sidang Skripsi

Tahapan ini adalah tahapan akhir dalam sebuah penelitian setelah tahap proses penelitian. Di tahap ini penulis mempresentasikan apa yang sudah dilewati dalam penelitiannya.

3.5.1 Publikasi Skripsi

Sebagai syarat kelulusan penulis diharuskan untuk mempublikasikan artikel ilmiah atau skripsi setelah menyelesaikan penelitiannya.

3.5.2 Publikasi Jurnal

Selain menjadi bukti utama orisinalitas penelitian, publikasi jurnal dapat menjadi modal rekam jejak penelitian. Semakin banyak jurnal ilmiah dikutip, maka akan meningkatkan reputasi penulis.