

Penerapan Data Mining untuk Mengcluster Kelulusan Mahasiswa Menggunakan Metode K-Means di Universitas Katolik Santo Thomas

Tiurma Wati Purba

^{1,2} Universitas Katolik Santo Thomas Medan, Jl. Setiabudi No. 479 F Tanjungsari, Medan, Indonesia
Email : tiurmawatipurba@gmail.com

Abstrak

Perguruan tinggi yang berakibatkan waktu kelulusan yang terlambat sehingga merugikan mahasiswa dan juga perguruan tinggi. Ketepatan kelulusan mahasiswa merupakan aspek untuk penilaian kelayakan Program Studi sebagai unit pelaksanaan Pendidikan perguruan tinggi, maka untuk mengatasi permasalahan tersebut diperlukan suatu solusi untuk menentukan atau mengcluster kelulusan mahasiswa. Pada penelitian ini dapat diklasifikasikan dengan menggunakan Teknik data mining dengan metode K-Means, sebelum dilakukan pengolahan data dilakukan proses normalisasi pada data, yang kemudian diolah menjadi beberapa kluster. Data yang telah diklusterisasi tersebut akan menghasilkan kategori lulus tidak lulusnya mahasiswa dalam jangka waktu 4 tahun. Dengan adanya penelitian ini pihak perguruan tinggi dapat mengetahui hasil dari pengelompokan kelulusan mahasiswa dan mampu memberikan solusi atau dapat memberikan tindakan untuk mengatasi keterlambatan kelulusan mahasiswa.

Kata Kunci : Ketepatan waktu kelulusan, Data mining, Clustering, K-Means.

Abstract

Universities which result in late graduation times which are detrimental to students and also universities. The accuracy of student graduation is an aspect for assessing the suitability of the Study Program as a unit for implementing higher education, so to overcome this problem a solution is needed to determine or cluster student graduation. This research can be classified using data mining techniques with the K-Means method, before data processing is carried out a normalization process is carried out on the data, which is then processed into several clusters. The clustered data will produce a pass or fail category for students within a 4year period. With this research, universities can find out the results of student graduation groupings and can provide solutions or take action to overcome delays in student graduation.

Keywords: Timeliness of graduation, Data mining, Clustering, K-Means.

PENDAHULUAN

Mahasiswa merupakan salah satu aspek penilaian keberhasilan atau kelayakan suatu program studi dalam penyelenggaraan perguruan tinggi. Aspek penilaian yang melibatkan mahasiswa diantaranya yaitu kelulusan mahasiswa yang tepat waktu yang akan menguntungkan kedua belah pihak. Setiap mahasiswa pasti menginginkan kelulusan yang tepat waktu begitu juga dengan perguruan tinggi, untuk mahasiswa semakin cepat lulus bisa melakukan star lebih awal di dunia kerja, lebih menikmati

hidup, menghemat biaya Pendidikan, membanggakan orang tua, dan minim tekanan. Sedangkan untuk perguruan tinggi kelulusan mahasiswa dapat meningkatkan kualitas, dan reputasi yang sangat berpengaruh pada akreditasi perguruan tinggi dan keberhasilan atau kelayakan suatu program studi di perguruan tinggi tersebut.

Normalnya kelulusan mahasiswa (Strata-1) yaitu dalam jangka waktu 4 tahun. Namun mahasiswa tidak selalu dapat menyelesaikan masa studinya dalam jangka waktu yang tepat. Untuk mengatasi masalah tersebut diperlukan analisa data set dan pembuatan sistem berbasis Web untuk pengelompokan data mahasiswa berdasarkan clustering kelulusan. Ketepatan kelulusan mahasiswa dapat diklasifikasikan dengan menggunakan Teknik Data Mining. Teknik Data Mining dalam penelitian ini adalah Clustering Algoritma K-Means.

Kriteria kelulusan Mahasiswa untuk jenjang S1 yaitu dengan Nilai Indeks Prestasi dari semester 1 sampai semester 7, Jumlah SKS, dan Masa Studi. Indeks Prestasi persemester yang dinyatakan lulus dengan nilai IPS 3.00 disetiap semester, sedangkan nilai IP semester yang dinyatakan tidak lulus dengan IP persemesternya ada nilai dibawah 3.00. Tetapi ada fakultas yang meluluskan mahasiswanya yang memiliki IP Semester dibawah 3.00 persemester di karenakan masa studi dan jumlah sks yang dibawa atau sudah diambil sudah selesai. Ada juga mahasiswa yang IP nya persemester di atas 3.00, jumlah sks yang diambil sudah selesai dan masa studinya sudah memenuhi kriteria, tetapi mahasiswa tersebut belum lulus, dikarenakan ada beberapa faktor yang mempengaruhi kelulusan mahasiswa antara lain, ada kendala dalam mengerjakan tugas akhir atau skripsi, ada fakultas yang mengharuskan matakuliah tertentu selesai terlebih dahulu baru sehingga bisa mengambil mata kuliah tugas akhir atau skripsi.

Clustering atau analisis cluster merupakan proses pembentukan kelompok data dari himpunan data yang tidak diketahui kelompok-kelompok atau kelas-kelasnya. Metode yang digunakan untuk mengcluster kelulusan mahasiswa yaitu dengan menggunakan metode K-Means, yaitu dengan mengelompokkan 'n' buah objek kedalam 'k' kelas berdasarkan jarak nya dengan pusat cluster. Analisis cluster mengklasifikasikan objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam cluster yang sama. Analisis cluster merupakan salah satu alat analisis yang berguna sebagai peringkasan data. Dalam meringkas data ini dapat dilakukan dengan cara mengelompokkan objek-objek berdasarkan kesamaan karakteristik tertentu diantara objek-objek yang hendak diteliti.

Penelitian ini fokus kepada penggunaan algoritma K-Means untuk mengcluster kelulusan mahasiswa. Diharapkan dengan adanya cluster kelulusan mahasiswa, apabila banyaknya mahasiswa dicluster lulus dan tidak lulusnya dalam jangka waktu 4 tahun, pihak kampus harus lah mampu melakukan kegiatan-kegiatan yang dapat meminimisasi jumlah mahasiswa yang dicluster tidak lulus tepat waktu yang dapat membantu mahasiswa meningkatkan prestasi belajar. (*Hendro Priyatma, Fahmi Sajid, Dannis Haldivany*).

METODE

Metode penelitian yang digunakan adalah metode penelitian kualitatif. Teknik pengumpulan data yang dipakai untuk mengumpulkan data adalah :

a. Observasi

Penulis mengamati secara langsung tempat obyek yang akan diteliti tepatnya di Universitas Katolik Santo Thomas. Sehingga penulis mendapatkan hasil secara lengkap.

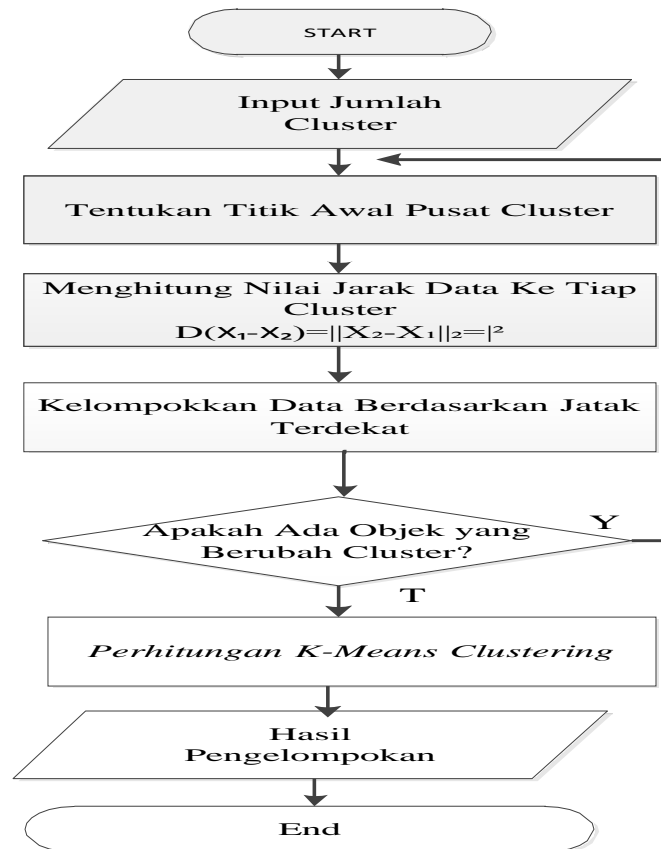
- b. Studi Pustaka
Penelitian yang menggunakan dan mempelajari buku-buku, atau Referensi-referensi maupun literatur yang berhubungan dengan penelitian ini yang dilandasi sebagai teori.
- c. Wawancara
Untuk mendapatkan data yang baik, dalam hal ini peneliti melakukan Pengumpulan data secara lisan kepada Lembaga Pengembangan Sistem Informasi (LPSI) universitas Katolik Santo Thomas terkait hasil studi mahasiswa Angkatan (Stambuk) 2019.

Tabel 1 Data Mahasiswa

No	NPM	NAMA	JK	INDEKS PRESTASI SEMESTER							SKS	MS
				1	2	3	4	5	6	7		
1	190600001	ensiana Runuh	P	3,85	0	3,87	0	0	3,9	4	116	4
2	190600002	enni Jupita Hutabarat	P	3,7	4,00	3,78	3,89	3,52	3,88	3,8	141	4
3	190600003	ndri Manusun P. Siagian	L	2,63	0	2,89	0	0	0	0	0	0
4	190600004	ika Clarensia	P	2,63	0	2,14	0	0	0	0	0	0
5	190600005	ulus Zai	L	3,68	3,68	3,63	3,4	3,25	2,63	0,9	147	4
6	190600006	Mayer Palentino Gultom	L	3,35	3,78	3,59	3,58	3,22	3,43	2,9	141	4
7	190600007	osafat Benedictus	L	3,15	3,9	3,74	3,21	3,29	2,26	3,5	141	4
8	190910001	vitania Br Sinuraya	P	3,5	3,81	3,44	3,69	3,48	4	4	144	4
9	190910002	anrein Aritonang	L	3,71	3,73	3,79	3,96	3,69	3,91	4	144	4
10	190910003	Nopita Br Tarigan	P	3,5	0,54	0	0	0	0	0	35	4
11	190910005	osepa Brtarigan	P	3,48	3,52	3,44	3,63	3,63	4	4	144	4
12	190910006	Marta Sr Silaban	P	3,83	3,92	3,83	4	4	4	4	144	4
13	190910007	mpuan Lidyawati Hasibuan	P	3,69	3,79	3,96	3,81	4	4	4	144	4
14	190910008	esi Sartika Sinurat	P	3,83	3,73	3,71	3,88	3,54	4	4	144	4
15	190920001	adar Rina Br Ginting	P	3,33	3,92	3,73	3,5	3,41	2,83	4	136	4
16	190920002	ignes S W Simatupang	P	3,25	3,46	3,86	3,08	3,18	3,15	4	134	4
17	190930001	iska Julinda Yani Zega	P	3,42	3,76	3,74	3,79	3,77	3,76	4	140	4
18	190930002	olasta Simanungkalit	P	3,5	3,81	3,7	3,75	3,82	2,72	4	140	4
19	190940002	ipita Cyndi C Simarmata	P	3,36	3,81	3,59	3,79	3,55	4	4	129	4
20	190940003	Martina Abigail I. Sihotang	P	4	4	3,96	3,96	3,84	3,82	4	134	4
21	190420001	oni Aman Tarigan	L	3,47	2,9	3	3,13	3,42	3,33	2,7	139	4
22	190420002	Niko Antonio Sitanggang	L	3,16	3,08	3,3	0	3,68	3,5	2,6	139	4
23	190430001	osi Pratama Simbolon	L	3,66	3,53	3,43	3,6	3,37	3,23	3,2	142	4
24	190430002	ramudya Sebastian Tarigan	L	2	3,28	2,8	2,82	0	2,53	2,9	142	4
25	190410001	dvnt Reuni Rajagukguk	P	3,56	3,86	3,96	3,59	3,36	3,83	3,7	142	4
26	190410002	Widya Asdelina Manurung	P	3,19	3,42	3,85	3,52	3,94	3,74	3,1	142	4
27	190310002	oel Yudika P. Aritonang	L	3,43	2,28	2,81	1,95	0,38	0,38	0,3	116	4
28	190310003	nselmus Delpin Manao	L	3,15	2,97	2,91	2,24	2,59	2,52	1,3	135	4
29	190320001	iderman Ndruru	L	3,55	2,43	0,67	0	3,1	0	0	103	4
30	190320002	idia Hotnauli Lumban Gaol	P	3,35	2,87	3,14	2,64	2,52	2,36	0,8	135	4

Flowchart dari Metode Penyelesaian

Berikut ini adalah *flowchart* dari proses prediksi kelulusan mahasiswa pada Universitas Santo Tomas Medan dalam satu tahun dengan menggunakan algoritma K-Means yaitu sebagai berikut:



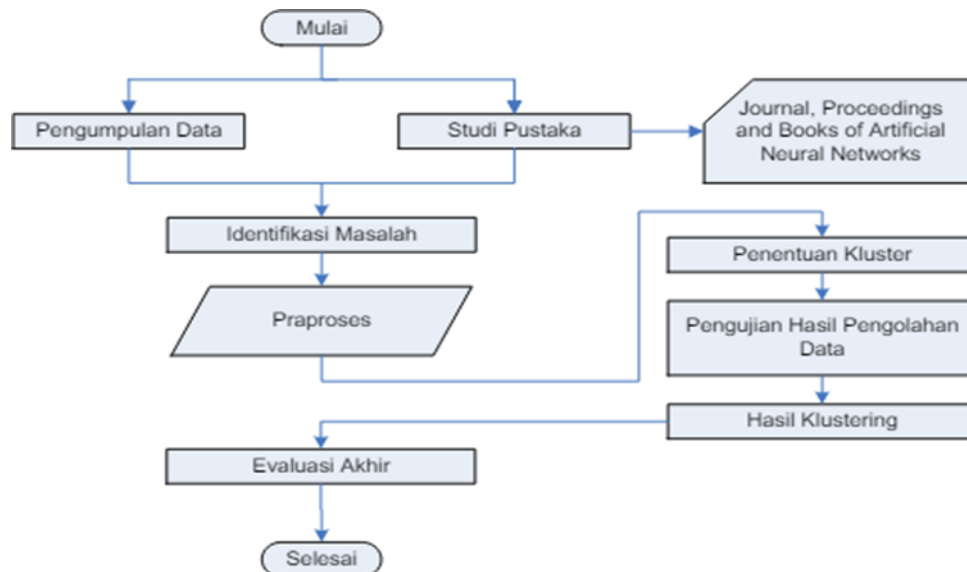
Gambar 1. Flowchar Penyelesaian Metode K-Means Clustering

Langkah-langkah metode K-Means dapat dijelaskan sebagai berikut :

1. Tentukan jumlah Cluster (k) pada data set.
2. Tentukan nilai pusat (Centroid). Penentuan nilai Centroid pada tahap awal dilakukan secara random
3. Pada masing-masing record, hitung jarak terdekat dengan Centroid. Ada beberapa cara yang dapat digunakan untuk mengukur jarak data ke pusat kelompok, diantaranya Euclidean, Manhattan/City Block, dan Minkowsky. Setiap cara memiliki kelebihan dan kekurangan masing-masing. Pada penelitian ini, jarak Centroid yang digunakan adalah Euclidean Distance,
4. Kelompokkan objek berdasarkan jarak ke Centroid terdekat 5. Ulangi langkah ke-3 hingga langkah ke-4, lakukan iterasi hingga Centroid bernilai optimal.

Tahapan Penelitian

Adapun tahapan yang dilakukan untuk menyelesaikan masalah penelitian disusun dalam kerangka kerja berikut:



Gambar 2. Tahapan Penelitian

Tahapan Algoritma K-Means

Algoritma ini akan secara iterative meningkatkan variasi nilai dalam tiap-tiap cluster dimana obyek selanjutnya ditempatkan dalam kelompok yang terdekat, dihitung dari titik tengah kluster. Titik tengah baru ditentukan bila semua data telah ditempatkan dalam cluster terdekat. Proses penentuan titik tengah dan penempatan data dalam cluster diulangi sampai nilai titik tengah dari semua cluster yang terbentuk tidak berubah lagi (Han, Dkk, 2012)

- Langkah 1
Tentukan berapa banyak cluster k dari dataset yang akan dibagi.
- Langkah 2
Tetapkan secara acak data k menjadi pusat awal lokasi kluster.
- Langkah 3
Untuk masing-masing data, temukan pusat cluster terdekat. Dengan demikian berarti masing-masing pusat cluster memiliki sebuah subset dari dataset, sehingga mewakili bagian dari dataset. Oleh karena itu, telah terbentuk Cluster k : $C_1, C_2, C_3, \dots, C_k$. Dengan menggunakan rumus Euclidian distance
$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2}$$
- Langkah 4
Untuk masing-masing cluster k , temukan pusat luasan kluster, dan perbaharui lokasi dari masing-masing pusat cluster ke nilai baru dari pusat luasan.
- Langkah 5
Ulangi Langkah ke-3 dan ke-5 hingga data-data pada tiap cluster menjadi terpusat atau selesai.

Setelah semua data riset diproses inisialisasi, maka data-data tersebut telah dapat dikelompokkan dengan menggunakan algoritma *K-Means*. Untuk dapat melakukan pengelompokan data-data tersebut menjadi beberapa *cluster* perlu dilakukan beberapa langkah, yaitu:

1. Iterasi Ke-1

- Tentukan jumlah *cluster* yang diinginkan. Dalam penelitian ini data yang ada akan dikelompokkan menjadi tiga *cluster*.
- Tentukan titik pusat awal dari setiap *cluster*. Dalam penelitian ini titik

pusat awal ditentukan secara *random* dan didapat titik pusat dari setiap *cluster* dapat dilihat pada tabel 2

Tabel 2. Inisialisasi Titik Pusat (Centroid) Awal Setiap Cluster

Nama Mahasiswa	1	2	3	4	5	6	7	SKS	MS
MARTINA SIHOTANG	4	4	3,96	3,96	3,84	3,82	4	134	4
NOPITA BR TARIGAN	3,5	0,54	0	0	0	0	0	0	0

- c. Tempatkan setiap data pada *cluster*. Dalam penelitian ini digunakan metode *k-means* untuk mengalokasikan setiap data ke dalam suatu *cluster*, sehingga data akan dimasukan dalam suatu *cluster* yang memiliki jarak paling dekat dengan titik pusat dari setiap *cluster*. Untuk mengetahui *cluster* mana yang paling dekat dengan data, maka perlu dihitung jarak setiap data dengan titik pusat setiap *cluster* dengan menggunakan rumus:

$$D(X_1, X_2) = ||X_2 - X_1||_2 = \sqrt{\sum_{j=1}^p |x_2^j - x_1^j|^2} = 1, 2, 3 \dots n$$

Perhitungan Jarak dari data ke-1 terhadap pusat *cluster* adalah

$$\begin{aligned} D(1,1) &= (3,85 - 4)^2 + (0 - 4)^2 + (3,87 - 3,96)^2 + (0 - 3,96)^2 + (0 - 3,84)^2 + (0 - 3,96)^2 + \\ &+ (0 - 3,84)^2 + (3,92 - 3,82)^2 + (4 - 4)^2 + (116 - 134)^2 + (4 - 4)^2 \\ &= 3,70 \end{aligned}$$

$$\begin{aligned} D(1,2) &= (3,85 - 3,5)^2 + (0 - 0,54)^2 + (3,87 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (3,92 - 0)^2 + (4 - 0)^2 + \\ &+ (116 - 0)^2 + (0 - 0)^2 \\ &= 1,53 \end{aligned}$$

- Dan seterusnya dilakukan perhitungan jarak antara data ke-2 sampai data ke-30. Kemudian akan didapatkan hasil perhitungan jarak setiap data terhadap pusat cluster baru sebagai berikut:

Hasil Perhitungan Jarak Setiap Data Terhadap Pusat Cluster
pada Iterasi Ke-1

No	Maha siswa	1	2	3	4	5	6	7	SKS	MS	C1	C2	Jarak Terdekat	Kelompok
1	P01	3,85	0	3,87	0	0	3,9	4	116	4	3,70	1,53	1,53	C2
2	P02	3,7	4	3,78	3,89	3,52	3,88	3,8	141	4	0,48	7,34	0,48	C1
3	P03	2,63	0	2,89	0	0	0	0	0	0	7,03	3,07	3,07	C2
4	P04	2,63	0	2,14	0	0	0	0	0	0	7,18	2,37	2,37	C2
5	P05	3,68	3,68	3,63	3,4	3,25	2,63	0,9	147	4	0,99	6,72	0,99	C1
6	P06	3,35	3,78	3,59	3,58	3,22	3,43	2,9	141	4	1,07	6,83	1,07	C1
7	P07	3,15	3,9	3,74	3,21	3,29	2,26	3,5	141	4	1,28	6,82	1,28	C1
8	P08	3,5	3,81	3,44	3,69	3,48	4	4	144		0,87	6,95	0,87	C1
9	P09	3,71	3,73	3,79	3,96	3,69	3,91	4	144	4	0,46	7,34	0,46	C1
10	P10	3,5	0,54	0	0	0	0	0	35	4	7,64	0,00	0,00	C2
11	P11	3,48	3,52	3,44	3,63	3,63	4	4	144	4	0,96	6,86	0,96	C1
12	P12	3,83	3,92	3,83	4	4	4	4	144	4	0,28	7,63	0,28	C1
13	P13	3,69	3,79	3,96	3,81	4	4	4	144	4	0,43	7,54	0,43	C1
14	P14	3,83	3,73	3,71	3,88	3,54	4	4	144	4	0,51	7,19	0,51	C1

No	Maha siswa	1	2	3	4	5	6	7	SKS	MS	C1	C2	Jarak Terdekat	Kelom pok
15	P15	3,33	3,92	3,73	3,5	3,41	2,83	4	136	4	0,95	7,02	0,95	C1
16	P16	3,25	3,46	3,86	3,08	3,18	3,15	4	134	4	1,44	6,56	1,44	C1
17	P17	3,42	3,76	3,74	3,79	3,77	3,76	4	140	4	0,69	7,28	0,69	C1
18	P18	3,5	3,81	3,7	3,75	3,82	2,72	4	140	4	0,63	7,28	0,63	C1
19	P19	3,36	3,81	3,59	3,79	3,55	4	4	129	4	0,83	7,11	0,83	C1
20	P20	4	4	3,96	3,96	3,84	3,82	4	134	4	0,00	7,64	0,00	C1
21	P21	3,47	2,9	3	3,13	3,42	3,33	2,7	139	4	1,81	6,01	1,81	C1
22	P22	3,16	3,08	3,3	0	3,68	3,5	2,6	139	4	4,21	5,57	4,21	C1
23	P23	3,66	3,53	3,43	3,6	3,37	3,23	3,2	142	4	0,98	6,71	0,98	C1
24	P24	2	3,28	2,8	2,82	0	2,53	2,9	142	4	4,68	5,05	4,68	C1
25	P25	3,56	3,86	3,96	3,59	3,36	3,83	3,7	142	4	0,76	7,13	0,76	C1
26	P26	3,19	3,42	3,85	3,52	3,94	3,74	3,1	142	4	1,10	7,15	1,10	C1
27	P27	3,43	2,28	2,81	1,95	0,38	0,38	0,3	116	4	4,54	3,86	3,86	C2
28	P28	3,15	2,97	2,91	2,24	2,59	2,52	1,3	135	4	2,72	5,12	2,72	C1
29	P29	3,55	2,43	0,67	0	3,1					5,45	3,69	3,69	C2
30	P30	3,35	2,87	3,14	2,64	2,52	2,36	0,8	135	4	2,42	5,35	2,42	C1

Setelah proses perhitungan selesai, maka akan terbentuk table *cluster* baru dengan tiap-tiap anggota *cluster* seperti table dibawah ini:

1	2	3	4	5	6	7	SK	MS
3,7	4	3,78	3,89	3,52	3,88	3,8	141	4
3,68	3,68	3,63	3,4	3,25	2,63	0,9	147	4
3,35	3,78	3,59	3,58	3,22	3,43	2,9	141	4
3,15	3,9	3,74	3,21	3,29	2,26	3,5	141	4
3,5	3,81	3,44	3,69	3,48	4	4	144	4
3,71	3,73	3,79	3,96	3,69	3,91	4	144	4
3,48	3,52	3,44	3,63	3,63	4	4	144	4
3,83	3,92	3,83	4	4	4	4	144	4
3,69	3,79	3,96	3,81	4	4	4	144	4
3,83	3,73	3,71	3,88	3,54	4	4	144	4
3,33	3,92	3,73	3,5	3,41	2,83	4	136	4
3,25	3,46	3,86	3,08	3,18	3,15	4	134	4
3,42	3,76	3,74	3,79	3,77	3,76	4	140	4
3,5	3,81	3,7	3,75	3,82	2,72	4	140	4
3,36	3,81	3,59	3,79	3,55	4	4	129	4
4	4	3,96	3,96	3,84	3,82	4	134	4
3,47	2,9	3	3,13	3,42	3,33	2,7	139	4
3,16	3,08	3,3	0	3,68	3,5	2,6	139	4
3,66	3,53	3,43	3,6	3,37	3,23	3,2	142	4
2	3,28	2,8	2,82	0	2,53	2,9	142	4
3,56	3,86	3,96	3,59	3,36	3,83	3,7	142	4
3,19	3,42	3,85	3,52	3,94	3,74	3,1	142	4
3,15	2,97	2,91	2,24	2,59	2,52	1,3	135	4
3,35	2,87	3,14	2,64	2,52	2,36	0,8	135	4
3,43	3,61	3,58	3,35	3,34	3,42	3,1	141	4

Anggota Cluster 2

1	2	3	4	5	6	7	SKS	MS
3,85	0	3,87	0	0	3,9	4	116	4
2,63	0	2,89	0	0	0	0	0	0
2,63	0	2,14	0	0	0	0	0	0
3,5	0,54	0	0	0	0	35	4	4
3,43	2,28	2,81	1,95	0,38	0,38	0,116	4	4
3,55	2,43	0,67	0	3,1	0	0	0	0
3,27	0,88	2,06	0,33	0,58	0	0	0	0

Nilai pada baris *Average* merupakan hasil perhitungan rata-rata dari tiap-tiap kolom titik *cluster*. Nilai *average* ini nantinya menjadi titik pusat untuk *cluster 2* yang baru. Sehingga titik pusat *cluster* pada iterasi kedua yaitu :

Titik Pusat (Centroid) Baru Setiap Cluster Pada Iterasi 2

	1	2	3	4	5	6	7	SKS	MS
Cluster 1	3,43	3,61	3,58	3,35	3,34	3,42	3,1	141	4
Cluster 2	3,27	0,88	2,06	0,33	0,58	0	0	0	0

2. Iterasi Ke-2

Setelah dapat titik pusat yang baru dari masing-masing *cluster*, maka lakukan kembali dari langkah ketiga hingga titik pusat dari setiap *cluster* tidak berubah lagi dan tidak ada lagi data yang berpindah dari satu *cluster* ke *cluster* yang lain.

Penentuan Pusat Awal (centroid) Cluster awal

$$D(1,1) = (3,85 - 3,43)^2 + (0 - 3,61)^2 + (3,87 - 3,58)^2 + (0 - 3,35)^2 + (0 - 3,34)^2 + (3,92 - 3,34)^2 + (4 - 3,1)^2 + (116 - 141)^2 + (4 - 4)^2$$

$$= 5,97$$

$$D(1,2) = (3,85 - 3,27)^2 + (0 - 0,88)^2 + (3,87 - 2,06)^2 + (0 - 0,33)^2 + (0 - 0,58)^2 + (3,92 - 0)^2 + (4 - 0)^2 + (116 - 0)^2 + (4 - 0)^2$$

$$= 2,20$$

Dan seterusnya dilakukan perhitungan jarak antara data ke-2 sampai data ke-30. Kemudian akan didapatkan hasil perhitungan jarak setiap data terhadap pusat cluster baru sebagai berikut :

Tabel 3 Hasil Perhitungan Jarak Setiap Data Terhadap Pusat Cluster pada Iterasi Ke-2

No	Nama Mahasiswa	1	2	3	4	5	6	7	SKS	MS	C1	C2	Jarak Terdekat	Kelompok
1	P01	3,85	0	3,87	0	0	3,9	4	116	4	5,97	2,20	2,20	C2
2	P02	3,7	4	3,78	3,89	3,52	3,88	3,8	141	4	0,77	5,85	0,77	C1
3	P03	2,63	0	2,89	0	0	0	0	0	0	6,04	1,52	1,52	C2
4	P04	2,63	0	2,14	0	0	0	0	0	0	6,17	1,28	1,28	C2
5	P05	3,68	3,68	3,61	3,4	3,25	2,63	0,9	147	4	0,28	5,20	0,28	C1
6	P06	3,35	3,78	3,59	3,58	3,22	3,43	2,9	141	4	0,32	5,32	0,32	C1
7	P07	3,15	3,9	3,74	3,21	3,29	2,26	3,5	141	4	0,46	5,25	0,46	C1
8	P08	3,5	3,81	3,44	3,69	3,48	4	4	144	4	0,45	5,50	0,45	C1
9	P09	3,71	3,73	3,79	3,96	3,69	3,91	4	144	4	0,79	5,84	0,79	C1
10	P10	3,5	0,54	0	0	0	0	0	35	4	6,68	2,20	2,20	C2
11	P11	3,48	3,52	3,44	3,63	3,63	4	4	144	4	0,44	5,40	0,44	C1
12	P12	3,83	3,92	3,83	4	4	4	4	144	4	1,08	6,15	1,08	C1
13	P13	3,69	3,79	3,96	3,81	4	4	4	144	4	0,94	6,01	0,94	C1
14	P14	3,83	3,73	3,71	3,88	3,54	4	4	144	4	0,72	5,70	0,72	C1

No	Nama Mahasiswa	1	2	3	4	5	6	7	SKS	MS	C1	C2	Jarak Terdekat	Kelompok
15	P15	3,33	3,92	3,73	3,5	3,41	2,83	4	136	4	0,39	5,49	0,39	C1
16	P16	3,25	3,46	3,86	3,08	3,18	3,15	4	134	4	0,48	4,92	0,48	C1
17	P17	3,42	3,76	3,74	3,79	3,77	3,76	4	140	4	0,65	5,77	0,65	C1
18	P18	3,5	3,81	3,7	3,75	3,82	2,72	4	140	4	0,67	5,79	0,67	C1
19	P19	3,36	3,81	3,59	3,79	3,55	4	4	129	4	0,53	5,63	0,53	C1
20	P20	4	4	3,96	3,96	3,84	3,82	4	134	4	1,12	6,14	1,12	C1
21	P21	3,47	2,9	3	3,13	3,42	3,33	2,7	139	4	0,95	4,57	0,95	C1
22	P22	3,16	3,08	3,3	0	3,68	3,5	2,6	139	4	3,43	4,01	3,43	C1
23	P23	3,66	3,53	3,43	3,6	3,37	3,23	3,2	142	4	0,38	5,25	0,38	C1
24	P24	2	3,28	2,8	2,82	0	2,53	2,9	142	4	3,77	3,80	3,77	C1
25	P25	3,56	3,86	3,96	3,59	3,36	3,83	3,7	142	4	0,53	5,56	0,53	C1
26	P26	3,19	3,42	3,85	3,52	3,94	3,74	3,1	142	4	0,75	5,58	0,75	C1
27	P27	3,43	2,28	2,81	1,95	0,38	0,38	0,3	116	4	3,62	2,28	2,28	C2
28	P28	3,15	2,97	2,91	2,24	2,59	2,52	1,3	135	4	1,65	3,58	1,65	C1
29	P29	3,55	2,43	0,61	0	3,1	0	0	0	0	4,60	3,30	3,30	C2
30	P30	3,35	2,87	3,14	2,64	2,52	2,36	0,8	135	4	1,39	3,77	1,39	C1

Pada iterasi kedua data mengalami perubahan kelompok, maka dilakukan cluster baru yang akan dihitung berdasarkan data anggota tiap-tiap cluster pada iterasi ke-2 sesuai dengan rumus pusat anggota cluster sehingga didapatkan perhitungan sebagai berikut :

Anggota Cluster 1

1	2	3	4	5	6	7	SKS	MS
3,7	4	3,78	3,89	3,52	3,88	3,8	141	4
3,68	3,68	3,63	3,4	3,25	2,63	0,9	147	4
3,35	3,78	3,59	3,58	3,22	3,43	2,9	141	4
3,15	3,9	3,74	3,21	3,29	2,26	3,5	141	4
3,5	3,81	3,44	3,69	3,48	4	4	144	4
3,71	3,73	3,79	3,96	3,69	3,91	4	144	4
3,48	3,52	3,44	3,63	3,63	4	4	144	4
3,83	3,92	3,83	4	4	4	4	144	4
3,69	3,79	3,96	3,81	4	4	4	144	4
3,83	3,73	3,71	3,88	3,54	4	4	144	4
3,33	3,92	3,73	3,5	3,41	2,83	4	136	4
3,25	3,46	3,86	3,08	3,18	3,15	4	134	4
3,42	3,76	3,74	3,79	3,77	3,76	4	140	4
3,5	3,81	3,7	3,75	3,82	2,72	4	140	4
3,36	3,81	3,59	3,79	3,55	4	4	129	4
4	4	3,96	3,96	3,84	3,82	4	134	4
3,47	2,9	3	3,13	3,42	3,33	2,7	139	4
3,16	3,08	3,3	0	3,68	3,5	2,6	139	4
3,66	3,53	3,43	3,6	3,37	3,23	3,2	142	4
2	3,28	2,8	2,82	0	2,53	2,9	142	4
3,56	3,86	3,96	3,59	3,36	3,83	3,7	142	4
3,19	3,42	3,85	3,52	3,94	3,74	3,1	142	4
3,15	2,97	2,91	2,24	2,59	2,52	1,3	135	4
3,35	2,87	3,14	2,64	2,52	2,36	0,8	135	4
3,43	3,61	3,58	3,35	3,34	3,42	3,1	141	4

Anggota Cluster 2

1	2	3	4	5	6	7	SKS	MS
3,85	0	3,87	0	0	3,9	4	116	4

2,63	0	2,89	0	0	0	0	0	0
2,63	0	2,14	0	0	0	0	0	0
3,5	0,54	0	0	0	0	0	35	4
3,43	2,28	2,81	1,95	0,38	0,38	0,3	116	4
3,55	2,43	0,67	0	3,1	0	0	0	0
3,27	0,88	2,06	0,33	0,58	0	0	0	0

Anggota Cluster 2

	1	2	3	4	5	6	7	SKS	MS
Cluster 1	3,43	3,61	3,58	3,35	3,34	3,42	3,1	141	4
Cluster 2	3,27	0,88	2,06	0,33	0,58	0	0	0	0

Berdasarkan pengelompokan Cluster diatas maka data mahasiswa yang akan lulus akan dikonversi kedalam data awal sebagai berikut:

Anggota Cluster 1

No	NPM	NAMA	JK	INDEKS PRESTASI SEMESTER						
				1	2	3	4	5	6	7
2	190600002	Henni Jupita Hutabarat	P	3,7	4	3,78	3,89	3,52	3,88	3,8
5	190600005	Julius Zai	L	3,68	3,68	3,63	3,4	3,25	2,63	0,9
6	190600006	Mayer Palentino Gultom	L	3,35	3,78	3,59	3,58	3,22	3,43	2,9
7	190600007	Yosafat Benedictus	L	3,15	3,9	3,74	3,21	3,29	2,26	3,5
8	190910001	Evitania Br Sinuraya	P	3,5	3,81	3,44	3,69	3,48	4	4
9	190910002	Vanrein Aritonang	L	3,71	3,73	3,79	3,96	3,69	3,91	4
11	190910005	Yosepa Tarigan	P	3,48	3,52	3,44	3,63	3,63	4	4
12	190910006	Marta Sr Silaban	P	3,83	3,92	3,83	4	4	4	4
13	190910007	Ampuan Lidyawati Hasibuan	P	3,69	3,79	3,96	3,81	4	4	4
14	190910008	Desi Sartika Sinurat	P	3,83	3,73	3,71	3,88	3,54	4	4
15	190920001	Sadar Rina Br Ginting	P	3,33	3,92	3,73	3,5	3,41	2,83	4
16	190920002	Agnes S W Simatupang	P	3,25	3,46	3,86	3,08	3,18	3,15	4
17	190930001	Siska Julinda Yani Zega	P	3,42	3,76	3,74	3,79	3,77	3,76	4
18	190930002	Rolasta Simanungkalit	P	3,5	3,81	3,7	3,79	3,82	2,72	4
19	190940002	Elpita Cyndi C Simarmata	P	3,36	3,81	3,59	3,79	3,55	4	4
20	190940003	Martina Abigail I. Sihotang	P	4	4	3,96	3,96	3,84	3,82	4
21	190420001	Joni Aman Tarigan	L	3,47	2,9	3	3,13	3,42	3,33	2,7
22	190420002	Niko Antonio Sitanggang	L	3,16	3,08	3,3	0	3,68	3,5	2,6
23	190430001	Sosi Pratama Simbolon	L	3,66	3,53	3,43	3,6	3,37	3,23	3,2
24	190430002	Pramudya Sebastian Tarigan	L	2	3,28	2,8	2,82	0	2,53	2,9
25	190410001	Advent Reuni Rajagukguk	P	3,56	3,86	3,96	3,59	3,36	3,83	3,7
26	190410002	Widya Asdelina Manurung	P	3,19	3,42	3,85	3,52	3,94	3,74	3,1
28	190310003	Anselmus Delpin Manao	L	3,15	2,97	2,91	2,24	2,59	2,52	1,3
30	190320002	Lidia Hotnauli Lumban Gaol	P	3,35	2,87	3,14	2,64	2,52	2,36	0,8

Anggota Cluster 2

No	NPM	NAMA	JK	Indeks Prestasi Semester						
				1	2	3	4	5	6	7
1	190600001	Sensiana Runuh	P	3,85	0	3,87	0	0	3,9	4

No	NPM	NAMA	JK	Indeks Prestasi Semester						
				1	2	3	4	5	6	7
3	190600003	Andri Manusun P. Siagian	L	2,63	0	2,89	0	0	0	0
4	190600004	Rika Clarena	P	2,63	0	2,14	0	0	0	0
10	190910003	Nopita Br Tarigan	P	3,5	0,54	0	0	0	0	0
27	190310002	Joel Yudika P. Aritonang	L	3,43	2,28	2,81	1,95	0,38	0,38	0,3
29	190320001	Fiderman Ndruru	L	3,55	2,43	0,67	0	3,1	0	0

Berdasarkan data Mahasiswa diatas maka dapat disimpulkan sebagai berikut :

1. *Cluster* 1 adalah kelompok Mahasiswa yang akan lulus
2. *Cluster* 2 adalah kelompok Mahasiswa yang tidak akan lulus

Dengan adanya hasil data diatas diharapkan akan menjadi bahan evaluasi bagi kampus untuk mencari langkah apa yang harus dilakukan untuk meningkatkan niat belajar serta niat mahasiswa ingin cepat lulus.

Analisis Kebutuhan dan Perancangan

1. Analisis data

Metode ini merupakan metode yang digunakan untuk melakukan perhitungan atau pembobotan yang sering disebut dengan metode Clustering K-Means. Metode ini dapat menyelesaikan masalah secara kompleks yang mampu melakukan pendekatan perbandingan berpasangan (Pairwise Comparisons). Perbandingan berpasangan merupakan membandingkan satu elemen dengan elemen lainnya sehingga menghasilkan nilai elemen dalam bentuk kualitatif. Elemen dalam bobot tingkat tertingginya adalah pilihan keputusan yang benar-benar layak dipertimbangkan untuk diambil.

2. Analisis Kebutuhan

Dalam analisis ini mendeskripsikan beberapa kebutuhan yang terakait dalam penganalisisan prediksi kelulusan mahasiswa berdasarkan data yang diperoleh hasil analisis kebutuhan tersebut adalah system yang akan dibangun membutuhkan input seperti Nama, Npm, Jenis Kelamin, IP Semester, dan SKS sebagai data akademik untuk kebutuhan proses, proses yang digunakan untuk mengolah data input adalah Teknik clustering menggunakan algoritma k-means. Dan yang menjadi kebutuhan output diharapkan adalah analisis dari hasil clustering sehingga dapat diketahui informasi yang dibutuhkan. Maka berikut istilah-istilah algoritma *K-Means Clustering*:

- a. Cluster : kelompok, pengkelasan, atau group.
- b. Centroid : Titik tengah atau titik pusat.
- c. Iterasi : Pengulangan.

3. Desain Sistem

Dalam fase ini dibagi beberapa indikator atau elemen yaitu:

- (1) pemodelan sistem dengan menggunakan UML (*Unified Modelling Language*),
- (2) desain input, dan
- (3) desain output dari sistem pengelompokan yang akan dirancang dalam pemecahan masalah prediksi kelulusan mahasiswa

4. Pembangunan Sistem

Fase ini menjelaskan bagaimana melakukan pengkodean terhadap desain sistem Berbasis WEB yang dirancang baik dari sistem input, proses dan output menggunakan PHP MySQL

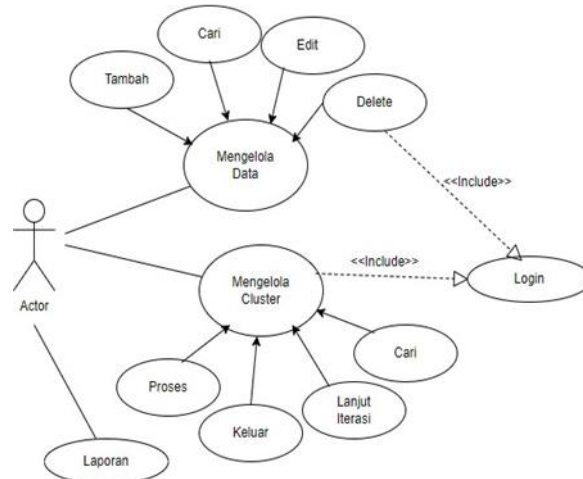
HASIL DAN PEMBAHASAN

Desain Sistem

Dalam proses perancangan sistem, tahap ini dilakukan dengan melakukan perancangan alur aktivitas yang akan diterapkan dalam sistem, rancangan *database*, serta tahap perancangan *interface*. Dalam penerapannya penulis menggunakan model UML (*Unified Modeling Language*), dimana tahap ini perancangan sistem dengan penggunaan *Use-caseDiagram*, *Activity Diagram*, dan *Class Diagram*.

1. Use Case Diagram

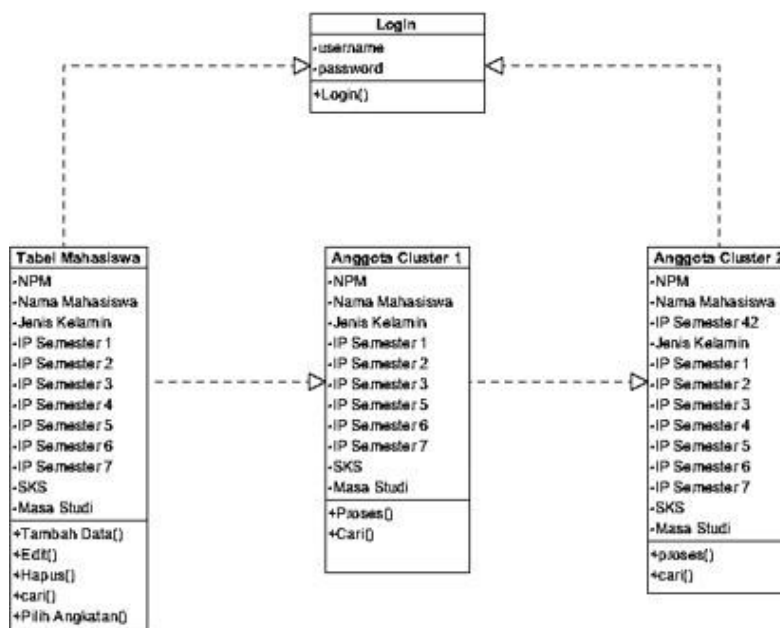
Use case diagram merupakan diagram yang menggambarkan hubungan interaksi antara aktor dan sistem.



Gambar 4. Use Case Diagram

2. Class Diagram

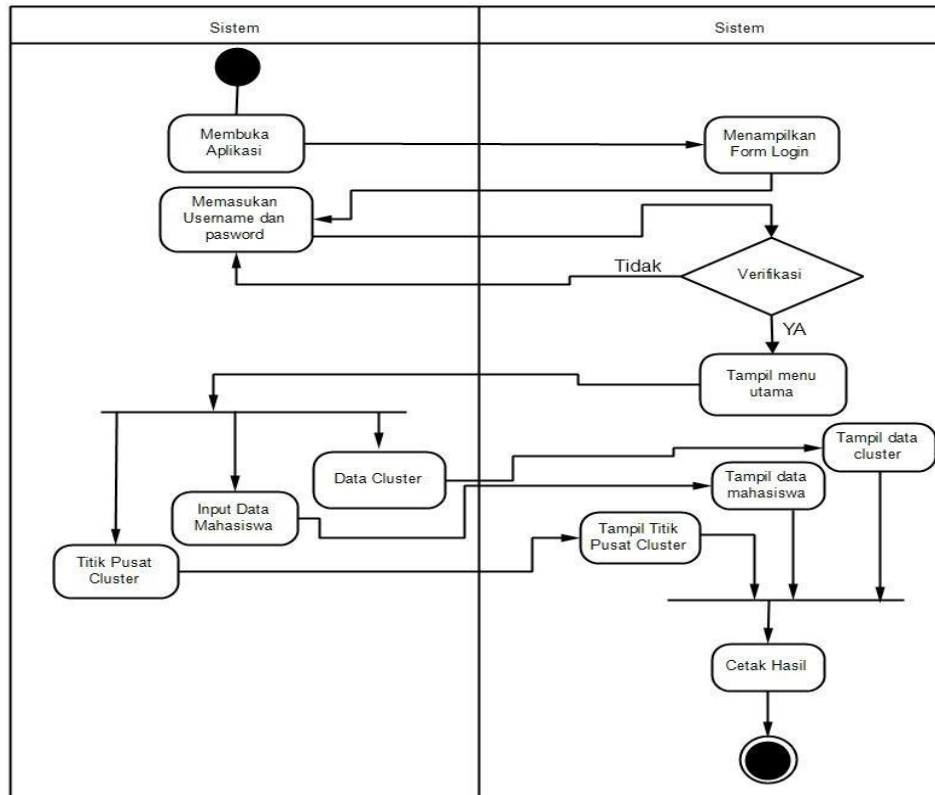
Class Diagram menggambarkan hubungan antar kelas dan penjelasan secara detail tiap-tiap kelas dari suatu sistem. Bentuk *class diagram* dari sistem yang dibangun dapat dilihat pada gambar di bawah ini



Gambar 5. Class Diagram

3. Activity Diagram

Activity Diagram adalah sebuah cara untuk memodelkan alur kerja (workflow) dari deskripsi Use case dalam bentuk grafik. Pada Activity Diagram berikut ini akan memperlihatkan aliran kendali dari suatu aktivitas ke aktivitas lainnya yang terjadi pada suatu sistem. Berikut ini merupakan tampilan Activity Diagram yang akan dilakukan dalam pengolahan data mahasiswa.

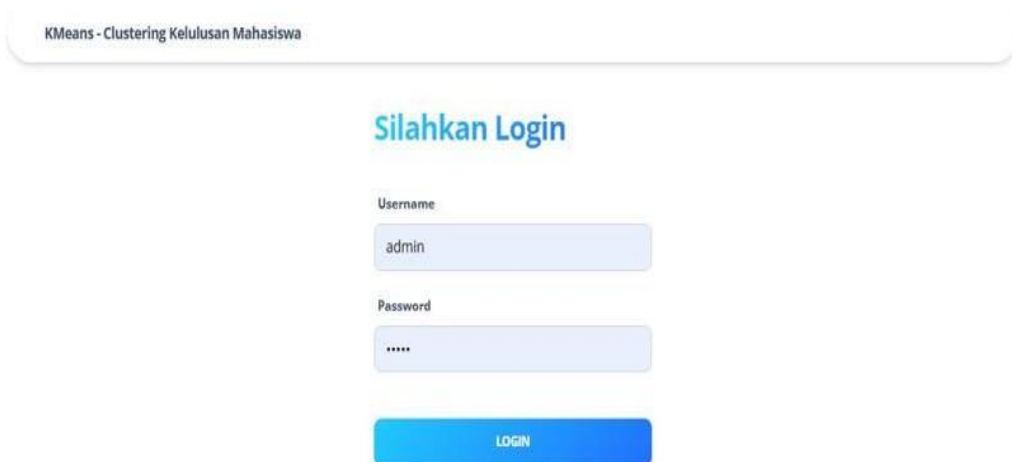


Gambar 6 . Activity Diagram

4. Tampilan Aplikasi

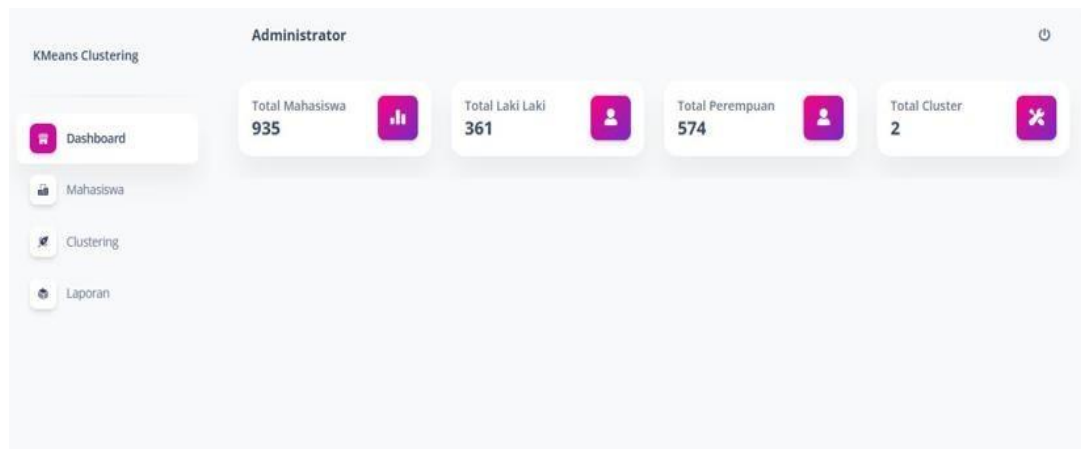
a. Halaman Login

Halaman ini merupakan halaman login yang dapat diakses oleh admin Lembaga Sistem Pengembangan Sistem Informasi Universitas ataupun dosen setiap fakultas dengan cara memasukan Username dan Password untuk maka dialihkan kehalaman lainnya. Bentuk rancangan tampilan halaman login.



b. Halaman Utama

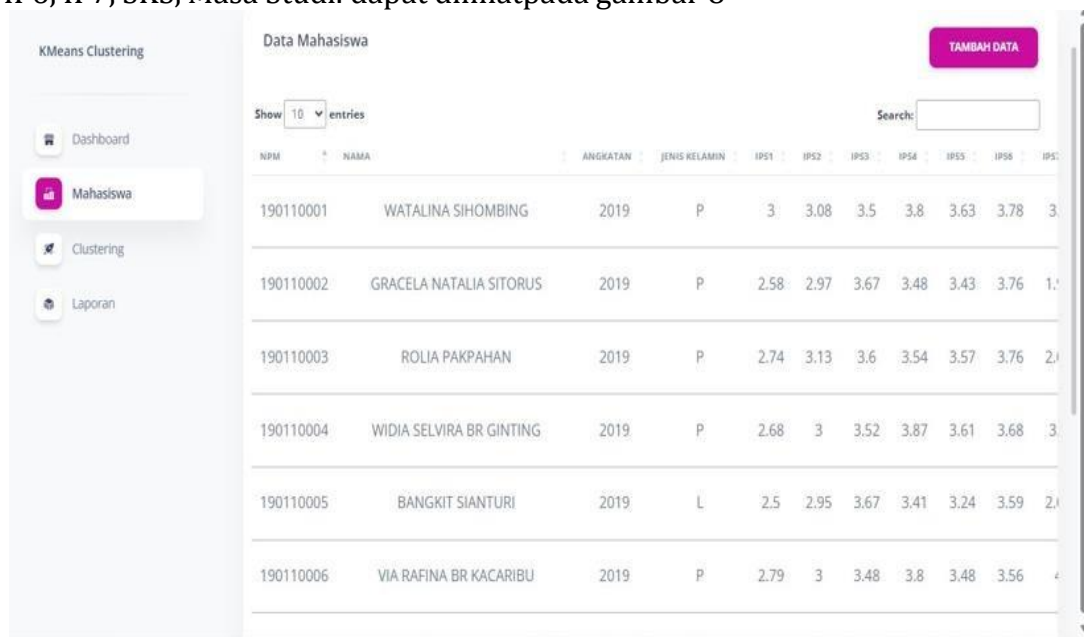
Pada halaman utama terdapat juga bar menu yang akan digunakan oleh admin.



Gambar 7. Halaman Utama

c. Halaman Input Data Mahasiswa

Pada halaman input data mahasiswa, akan menampilkan Npm, Nama, IP1, IP2, IP3, IP4, IP5, IP6, IP7, SKS, Masa Studi. dapat dilihat pada gambar 8



Gambar 8. Halaman Input Data Mahasiswa

d. Halaman Cluster

Pada tahap ini terdapat dua fitur yaitu proses data dan refresh yang dimana pada proses data akan memunculkan hasil cluster yang mahasiswa yang lulus dan mahasiswa yang tidak lulus dan untuk fitur refresh digunakan untuk membersihkan data. Dapat dilihat pada gambar 9

Cluster Tidak Lulus (Jumlah Anggota : 394)

Show 10 entries

NPM	NAME	IPRAK C1	IPRAK C2	CLUSTER
19060001	SENSIANA RUNUH	5.46	4.93	2 (Tidak Lulus)
190110002	GRACELA NATALIA SITORUS	2.56	1.60	2 (Tidak Lulus)
190110003	ROLIA PAKPAHAN	2.30	1.45	2 (Tidak Lulus)
190110005	BANGKIT SIANTURI	2.55	1.44	2 (Tidak Lulus)
190110007	LAELA SITUMORANG	2.82	1.61	2 (Tidak Lulus)
190110009	JOIN SAROHA SIANTURI	2.24	1.56	2 (Tidak Lulus)
190110010	NEVANI KRISTIANI BR LUMBAN RAJA	2.50	1.44	2 (Tidak Lulus)
190110011	LESTARIA SIMANJUNTAK	3.18	1.92	2 (Tidak Lulus)
190110012	LUTHER FAJAR KRISTIAN WARUWU	2.27	1.35	2 (Tidak Lulus)
190110013	ROY PURNAMASARI SIMBOLON	2.83	1.79	2 (Tidak Lulus)

Showing 1 to 10 of 394 entries

Previous 1 2 3 4 5 ... 40 Next

Gambar 9. Halaman Cluster

e. Halaman Laporan hasil

Pada halaman ini akan menampilkan hasil prediksi kelulusan mahasiswa yang dimana terdapat tampilan diagram hasil prediksi kelulusan serta hasil prediksi bisa dicetak dalam bentuk file PDF dan EXEL.

Dashboard

Mahasiswa

Clustering

Laporan

Cluster 2 (Tidak Lulus)

Show 10 entries

NPM	NAME	IPRAK C1	IPRAK C2	IPRAK C3	IPRAK C4	IPRAK C5	IPRAK C6	IPRAK C7	IPRAK C8	IPRAK C9	IPRAK C10
19060001	SENSIANA RUNUH	L	3.8	4	3.87	0	0	3.9	4	1	
190110002	GRACELA NATALIA SITORUS	P	2.58	2.97	3.67	3.48	3.43	3.76	1.92	1	
190110003	ROLIA PAKPAHAN	P	2.74	3.13	3.6	3.54	3.57	3.76	2.08	1	
190110005	BANGKIT SIANTURI	L	2.5	2.95	3.87	3.41	3.24	3.58	2.07	1	
190110007	LAELA SITUMORANG	P	2.29	2.57	3.5	3.41	3.28	3.59	1.85	1	
190110010	NEVANI KRISTIANI BR LUMBAN RAJA	P	2.76	3.21	3.43	3.35	3.46	3.78	1.9	1	
190110011	LESTARIA SIMANJUNTAK	P	2.71	2.58	3.35	3.54	3.35	3.65	1.33	1	
190110012	LUTHER FAJAR KRISTIAN WARUWU	L	2.87	3.61	3.57	3.67	3.37	3.59	2	1	
190110013	ROY PURNAMASARI SIMBOLON	P	3.16	3.39	3.38	3.61	3.57	3.71	1.33	1	
190110014	DWITA MARSDA	P	3.18	3.24	3.54	3.54	3.55	3.74	1.75	1	

Showing 1 to 10 of 306 entries

Previous 1 2 3 4 5 ... 31 Next

Gambar 10. Laporan Hasil

KESIMPULAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa Penelitian ini dilakukan dengan mengimplementasikan model algoritma K-Means Clustering untuk mengcluster kelulusan Mahasiswa Universitas Katolik Santo Thomas Medan. Maka untuk memperoleh data yang efisien dan akurat perlu dilakukan preprocessing sebelum diterapkan ke dalam metode. Maka berdasarkan kesimpulan yang didapat pada penelitian ini adalah dengan menggunakan metode Algoritma K-Means untuk mengclustering kelulusan mahasiswa Angkatan 2019 dengan jumlah sampel data

mahasiswa sebanyak 30 mahasiswa, yang dinyatakan lulus sebanyak 24 mahasiswa dan 6 mahasiswa yang tidak lulus. Maka dari itu dalam menerapkan data mining untuk mengcluster kelulusan dengan metode K-Means (Angkatan 2019) sudah dapat menampilkan informasi cluster kelulusan mahasiswa, namun untuk menilai kemampuan dari real K-Means Clustering kelulusan tergantung pada mahasiswa itu sendiri.

DAFTAR PUSTAKA

- Roger, Richard J.& Geatz, Michael W. 2003 Data Mining a Tutorial – Based Primer, United State Of America: Pearson Education Inc
- Hutagalung & Sonata, 2021. Penerapan Metode K-Means untuk Menganalisis Minat Nasabah Asuransi.
- Tan, P., Steinbach, M., & Kumar, V., (2006). Introduction to Data Mining, Pearson Education.
- Santoso, B.2007. Data Mining Teknik Pemanfaatan untuk Keperluan Bisnis.Yogyakarta: Graha Ilmu
- Han, J., Kamber, M., Pei, J. 2012. Data Mining Concept and Techinques 3rd ed. Morgan Kauffman-Elsevier, Amsterdam.
- Jain. A.K 2009 Data Clustering: 50 Years Beyond K-Means. Pattern Recognition Letter. Hendro Priyatma, and Fahmi Fajid, and Dannis Haldivany, “Klasterisasi Menggunakan K-Means Clustering Untuk Mprediksi Waktu Kelulusan Mahasiswa.” *Jurnal Edukasi dan Penelitian Informatika* Vol.5.no.1 April 2019.
- Alfiyani Rindyantul Jannah, and Deni Arifianto, M. Kom, “Penerapan metode Clustering Dengan Algoritma K-Means Untuk Memprediksi Kelulusan Mahasiswa Jurusan Teknik Informatika Di Universitas Muhammadiyah Jember,”
- Rony Setiawan, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru,” vol. 3, no. 1 Mei 2016.
- Mardhia Safa, (2019), “Penerapan Data Mining Dalam Klasifikasi Tingkat Kelulusan Mahasiswa Di Universitas Muhammadiyah Jember,”.
- Silmi Nur Adsini and Dewanto Rosian Adhy, “Aplikasi Prediksi Waktu Kelulusan Mahasiswa Menggunakan Algoritma K-Means Clustering.”. *Jurnal Sainteasa* vol. 1. Edisi 1 28-10-2021.
- F. Handayani, (2022),” Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering Untuk Mengelompokan Mahasiswa Berdasarkan Gaya Belajar,” *Jurnal Teknologi dan Informasi* vol. 12. No.1 Edisi Maret 2022.
- Dwi Kartini and Muliadi, (2016), “Rancang Bangun Aplikasi K-Means Untuk Klasifikasi Kelulusan Siswa Sekolah Kepolisian Negara Daerah Kalimantan Selatan,” *JurnalPro TekInfo* vol.3 no. 1 September 2016.