

**EVALUASI KINERJA AKADEMIK MAHASISWA
MENGUNAKAN ALGORITMA K-MEANS CLUSTERING**

SKRIPSI

Diajukan Kepada :

Fakultas Sains dan Teknologi

Universitas Islam Negeri Maulana Malik Ibrahim Malang

Untuk Memenuhi Salah Satu Persyaratan Dalam

Memperoleh Gelar Sarjana S.Kom

Oleh :

MOH. SHOHIBUL WAFA

NIM. 06550032

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2013**

**EVALUASI KINERJA AKADEMIK MAHASISWA
MENGUNAKAN ALGORITMA K-MEANS CLUSTERING**

SKRIPSI

Oleh :

MOH. SHOHIBUL WAFA

NIM. 06550032

Telah Diperiksa dan Disetujui untuk Diuji :

Tanggal : Januari 2013

Pembimbing I

Pembimbing II

Fatchurrochman, M.Kom

NIP. 197007312005011002

Syahiduzzaman, M.Kom

NIP. 197005022005011005

Mengetahui,

Ketua Jurusan Teknik Informatika

Ririen Kusumawati, M.Kom

NIP. 197203092005012002

EVALUASI KINERJA AKADEMIK MAHASISWA MENGUNAKAN ALGORITMA K-MEANS CLUSTERING

SKRIPSI

oleh :
MOH. SHOHIBUL WAFA
NIM. 06550032

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal : 26 Januari 2013

Susunan Dewan Penguji :	Tanda Tangan
1. Penguji Utama : <u>Ririen Kusumawati, M.Kom</u> NIP. 197203092005012002	()
2. Ketua Penguji : <u>A'la Syauqi, M.Kom</u> NIP. 197712012008011007	()
3. Sekretaris Penguji : <u>Fatchurrochman, M.Kom</u> NIP. 197007312005011002	()
4. Anggota Penguji : <u>Syahiduzaman, M.Kom</u> NIP. 197005022005011005	()

Mengesahkan,
Ketua Jurusan Teknik Informatika

Ririen Kusumawati, M.Kom
NIP. 197203092005012002

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini :

Nama : Moh. Shohibul Wafa

NIM : 06550032

Jurusan : Teknik Informatika

Fakultas : Sains dan Teknologi

menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau fikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, Januari 2013

Yang membuat pernyataan,

Moh. Shohibul Wafa
NIM. 06550032

MOTTO

“Jika kamu berada di sore hari jangan tunggu pagi hari, dan jika kamu berada di pagi hari jangan tunggu sore hari, gunakanlah kesehatanmu untuk (persiapan saat) sakitmu dan kehidupanmu untuk kematianmu.” (HR. Bukhori)



HALAMAN PERSEMBAHAN

"Sesungguhnya bersama kesulitan ada kemudahan"
Q.S. Al-Insyirah : 6

"Tugas kita bukanlah untuk berhasil. Tugas kita adalah untuk mencoba, karena di dalam mencoba itulah kita menemukan dan belajar membangun kesempatan untuk berhasil."

"Jalan kita masih panjang ! Tetap Semangat!"

... dari seorang anak, untuk kedua orang tuanya ...

Ayahanda Nur Ahmad & Ibunda Mabsusah

Yang selalu ada dan percaya. Untuk cinta dan semua yang telah kalian berikan sampai hari ini dan hari-hari berikutnya. Thank you for every pray and tears that dropped for me. Love you, more and more each day...

... dari seorang anak, untuk adik-adiknya ...

Adinda Nely Amalia Fadhila, Ana Raisa Masrurin

& Naila Khairil Mala

... dari seorang anak, untuk teman-teman seperjuangannya ...

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Tiada kata yang pantas terucap selain Alhamdulillah sebagai wujud rasa syukur kepada Allah SWT, yang mana hanya dengan rahmat, taufik, hidayah, dan inayah-Nya. Penulis telah dapat menyusun skripsi ini sebagai syarat untuk menyelesaikan pendidikan program S1 dalam bidang Teknik Informatika, pada Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Shalawat beserta salam tidak pernah lupa kami haturkan pada junjungan kita, Nabi besar Muhammad SAW yang karenanya kita mendapat pencerahan menuju jalan yang lurus, jalan yang diridhoi dan bukan jalan orang sesat yang dimurkai. Semoga Allah melimpahkan atas mereka rahmat yang sesuai dengan keutamaan mereka sebagai pahala atas amal perbuatan mereka.

Dalam penyusunan skripsi ini, penulis banyak mendapat bimbingan dan arahan dari berbagai pihak, oleh karena itu selayaknya penulis mengucapkan terima kasih dan penghargaan yang setinggi-tingginya kepada :

1. Bapak dan Ibu tercinta sebagai Orang tua serta Saudara-saudara kami yang selalu memberi do'a dan motivasi kepada kami. Perjuangan dan keikhlasan kalian membuat kami malu untuk tidak berprestasi dan berkarya.
2. Bapak Prof. Dr. H. Mudjia Rahardjo, M.Si, selaku Rektor Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang.
3. Bapak Prof. Dr. Sutiman Bambang Sumitro, SU., Dsc, selaku Dekan fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang.
4. Ibu Ririen Kusumawati, S.Si, M.Kom., selaku Ketua jurusan Teknik Informatika Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang

5. Bapak Fatchurrohman, M.Kom selaku Dosen wali dan Dosen pembimbing yang memberikan semangat kepada penulis selama menjalani perkuliahan serta telah bersedia meluangkan waktu, tenaga, pikiran, arahan dan masukan yang sangat berguna dalam menyelesaikan skripsi ini.
6. Bapak Syahiduzzaman, M.Kom selaku pembimbing agama yang telah bersedia memberikan pengarahan keagamaan dalam penyelesaian skripsi ini.
7. Seluruh Dosen jurusan Teknik Informatika fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang yang telah mengalirkan ilmu, pengetahuan, pengalaman, wacana dan wawasannya, sebagai pedoman dan bekal bagi penulis.
8. Teman-teman jurusan Teknik Informatika fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang yang telah memberi motivasi, informasi, dan masukannya pada penulis.
9. Segenap keluarga besar Lembaga Tinggi Pesantren Luhur Malang yang telah memberikan ilmu agama, do'a serta pengalaman yang berharga.
10. Kepada semua pihak yang telah membantu hingga terselesaikannya skripsi ini, baik secara langsung maupun tidak langsung yang tidak dapat penulis sebutkan satu-persatu.

Teriring doa dan harapan semoga apa yang telah mereka berikan kepada penulis, mendapatkan balasan yang lebih baik dari Allah SWT. Amin.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Malang, Januari 2013

Penulis

DAFTAR ISI

HALAMAN JUDUL	
HALAMAN PENGAJUAN	
HALAMAN PERSETUJUAN	
HALAMAN PENGESAHAN	
HALAMAN PERNYATAAN KEASLIAN TULISAN	
HALAMAN MOTTO	
HALAMAN PERSEMBAHAN	
KATA PENGANTAR	i
DAFTAR ISI	iv
DAFTAR TABEL	viii
DAFTAR GAMBAR	x
ABSTRAK	xii
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah	5
1.4 Tujuan dan Manfaat Penelitian	6
1.5 Metodologi Penelitian	7
1.6 Sistematika Penulisan	9
BAB II LANDASAN TEORI	
2.1 Konsep Evaluasi Kinerja Akademik Mahasiswa	10
2.2 Data Mining	11
2.2.1 Pengelompokan Data Mining	14
2.3 Algoritma Clustering	16
2.3.1 Karakteristik Clustering	19
2.4 Metode Pengelompokan	20
2.5 Distance Space	20

2.5.1	Euclidian Distance Space	21
2.6	Inner Product	22
2.7	K-Means	22
2.8	Normalisasi Data	25

BAB III ANALISIS DAN PERANCANGAN SISTEM

3.1	Analisis Data Mining	27
3.1.1	Sumber Data.....	27
3.1.2	Data yang Digunakan.....	31
3.1.3	Transformasi Data.....	33
3.1.3.1	Normalisasi Min-Max.....	34
3.1.3.2	Term Frequency.....	35
3.1.3.2.1	Term Frequency Jalur Masuk	38
3.1.3.2.2	Term Frequency Asal Sekolah.....	38
3.1.3.2.3	Term Frequency Pernah Mondok.....	39
3.1.4	Implementasi Algoritma K-Means.....	40
3.2	Analisis Sistem	48
3.3	Perancangan Proses	49
3.4	Perancangan Sistem	49
3.4.1	Flowchart Algoritma K-Means	50
3.4.2	Hasil Clustering	51
3.4.3	Data Flow Diagram Sistem	52
3.4.4	Desain Tabel	54
3.5	Desain User Interface	59
3.5.1	Desain Form Halaman Utama	59
3.5.2	Desain Form Data Mahasiswa	61
3.5.3	Desain Form Parameter	62
3.5.4	Desain Form Penilaian	62
3.5.5	Desain Form Pengaturan Skala Nilai	63
3.5.6	Desain Form Proses Cluster	64

BAB IV IMPLEMENTASI DAN PENGUJIAN

4.1 Implementasi Sistem	68
4.2 Kebutuhan Sistem	69
4.3 Implementasi dan Pengujian Sistem	70
4.3.1 Implementasi Perangkat Lunak	70
4.3.1.1 Class –class yang dibutuhkan	70
4.3.1.1.1 Class KmeansAkademik	70
4.3.1.1.2 Class Mahasiswa	75
4.3.1.1.3 Class Parameter	76
4.3.1.1.4 Class Penilaian	76
4.3.1.1.5 Class SkalaNilaiNumeric	76
4.3.1.1.6 Class SkalaNilaiNonNumeric	76
4.3.1.1.7 Class Klasterisasi	76
4.3.1.1.8 Class HalamanUtama	76
4.4 Pengujian Sistem	76
4.4.1 Pengujian Menu Utama	77
4.4.2 Pengujian Menu Mahasiswa	77
4.4.3 Pengujian Menu Parameter	78
4.4.4 Pengujian Sub Menu Skala Nilai Numeric	79
4.4.5 Pengujian Sub Menu Skala Nilai Non Numeric	79
4.4.6 Pengujian Menu Penilaian	80
4.4.7 Pengujian Menu Proses K-Means Clustering	80
4.5 Pengujian dan Analisis Program	86
4.6 Validasi Cluster	88
4.6.1 Confusion Matrix	88
4.7 Precision dan Recall	90
4.8 Kajian Agama	92
 BAB V PENUTUP	
5.1 Kesimpulan	96
5.2 Saran	97

DAFTAR PUSTAKA

DAFTAR TABEL

2.1 Perbedaan klasifikasi dan klasterisasi	18
3.1 Tabel Data Induk Mahasiswa	28
3.2 Tabel Data Induk Kelulusan	30
3.3 Indeks Prestasi Kumulatif Mahasiswa	34
3.4 Normalisasi Indeks Prestasi Kumulatif	35
3.5 Data Awal Parameter Mahasiswa	36
3.6 Frequency Jalur Masuk	38
3.7 Frequency Asal Jenis Sekolah	39
3.8 Frequency Pernah Mondok	39
3.9 Tabel Data Sample Mahasiswa	40
3.10 Tabel Term Frequency Atribut	41
3.11 Tabel Hasil Normalisasi Parameter	41
3.12 Posisi Cluster Pada Iterasi Pertama	43
3.13 Tabel Posisi Cluster Pada Iterasi Kedua	45
3.14 Hasil Pelabelan Cluster	47
3.15 Hasil Perhitungan K-Means Clustering	47
3.16 Daftar Tabel pada Database	55
3.17 Struktur Tabel Mahasiswa	56
3.18 Struktur Tabel Parameter	57
3.19 Struktur Tabel Normalisasi	57
3.20 Struktur Tabel Label Cluster	58
3.21 Struktur Tabel Hasil Cluster	58
4.1 Pembeding Hasil Aplikasi dengan data Sebenarnya	86
4.2 Rumus Confusion Matrix	88
4.3 Confusion Matrix	89

DAFTAR GAMBAR

Gambar 2.1. Data sebelum di klasterisasi	17
Gambar 2.2. Klasterisasi berdasarkan similaritas (kesamaan) warna	17
Gambar 2.3. Klasterisasi berdasarkan similaritas (kesamaan) bentuk	17
Gambar 2.4. Klasterisasi berdasarkan similaritas (kesamaan) jarak	18
Gambar 2.5. Karakteristik Klasterisasi	20
Gambar 2.6. Perhitungan Jarak Dua Titik	22
Gambar 3.1. Diagram Blog Sistem	49
Gambar 3.2. Flowchart K-Means Clustering	50
Gambar 3.3. Context Diagram Sistem	52
Gambar 3.4. Data Flow Diagram Level-0	54
Gambar 3.5. Desain Relasi Antar Tabel	55
Gambar 3.6. Desain Form Halaman Utama Aplikasi	60
Gambar 3.7. Desain Form Pengisian Data Mahasiswa	61
Gambar 3.8. Desain Form Pengisian Data Parameter	62
Gambar 3.9. Desain Form Pengisian Data Penilaian Parameter	63
Gambar 3.10. Desain Form Pengaturan Skala Nilai Numerik	64
Gambar 3.11. Desain Form Klasterisasi	65
Gambar 4.1. Pengujian Menu Utama	77
Gambar 4.2. Pengujian Form Mahasiswa	77
Gambar 4.3. Pengujian Form Parameter	78
Gambar 4.4. Pengujian Form Pengaturan Skala Numerik	79
Gambar 4.5. Pengujian Form Pengaturan Skala Non Numerik	79
Gambar 4.6. Pengujian Form Penilaian Parameter	80
Gambar 4.7. Pengujian Form Proses Cluster Tab Keterangan Mahasiswa	81
Gambar 4.8. Pengujian Form Proses Cluster Tab Keterangan Normalisasi	82

Gambar 4.9. Pengujian Form Proses Cluster – Load Data Mahasiswa	82
Gambar 4.10. Pengujian Form Proses Cluster – Inisialisasi Cluster Awal.....	82
Gambar 4.11. Pengujian Tab Proses Cluster – Iterasi Euclidean Distance.....	83
Gambar 4.12. Pengujian Form Proses Cluster - Hasil Cluster.....	83
Gambar 4.13. Pengujian Form Proses Cluster – Pelabelan Cluster	84
Gambar 4.14. Pengujian Form Proses Cluster Tab Hasil Cluster	84
Gambar 4.15. Pengujian Form Proses Cluster - <i>Pie Chart Diagram</i>	85
Gambar 4.16. Pengujian Form Proses Cluster - <i>Bar Chart Diagram</i>	85



ABSTRAK

Wafa, Moh.Shohibul. 2013. **Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Clustering**. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Pembimbing: (I) Fatchurrochman, M.Kom.

(II) Syahiduzzaman, M.Kom

Telah banyak model aturan yang dikembangkan sebagai sistem untuk mengevaluasi kinerja akademik mahasiswa menggunakan berbagai teknik pengelompokan (*Clustering*). *Clustering* adalah salah satu bidang penelitian yang paling penting dalam bidang data mining. *Clustering* adalah proses mengelompokkan atau penggolongan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan antar anggota satu *cluster* dan meminimumkan kesamaan antar *cluster*.

Kami mengajukan sebuah aplikasi data mining untuk mengevaluasi kinerja akademik mahasiswa menggunakan Algoritma K-Means Clustering. Hal ini memperkenalkan konsep dari Algoritma K-Means dan menggambarkan bagaimana konsep ini dapat diterapkan oleh pelaku akademik untuk mengevaluasi kinerja akademik mahasiswa.

Dari hasil uji coba, pola yang dibentuk mempunyai akurasi kecocokan mencapai 57,44%. dan nilai prosentase kesalahan sebesar 39,59% , hal ini disebabkan proses transformasi data yang belum optimal serta berbagai faktor lain yang mempengaruhi kinerja akademik mahasiswa.

Kata Kunci : K-Means Clustering, Evaluasi Kinerja Akademik

ABSTRACT

Wafa, Moh.Shohibul. 2013. Academic Performance Evaluation Using **K-Means Algorithm**. Thesis. Informatics Programme Faculty of Science and Technology The State of Islamic University Maulana Malik Ibrahim Malang.

Promotors: (I) Fatchurrochman, M.Kom.

(II) Syahiduzzaman, M.Kom

Many approaches have been developed for academic performance evaluation using various clustering techniques. Clustering is one of the most important research areas in the field of data mining. Clustering means creating groups of objects based on their features in such a way that the objects belonging to the same groups are similar and those belonging to different groups are dissimilar.

We have proposed a data mining application for student academic performance evaluation using K-Means Clustering Algorithm. It introduces the principles behind K-Means Algorithm and illustrates how these principles could be applied by educators to evaluating student academic performance.

From the test, a pattern that has formed the suitability reached 57.44% accuracy., And the value of the percentage error is 39.59%, this is due to the transformation of data that optimal not yet, and a variety of other factors that affect students' academic performance.

Keywords : K-Means Clustering, Academic performance Evaluation

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Perkembangan teknologi saat ini semakin meningkat dan memberi pengaruh yang besar hampir di setiap sektor kehidupan dan kenegaraan. Proses globalisasi yang terjadi di setiap negara di dunia saat ini juga mendukung perkembangan dan penggunaan teknologi. Hal ini menjadikan kebutuhan akan informasi yang akurat sangat penting dalam kehidupan sehari-hari. Namun kebutuhan informasi yang tinggi kadang tidak diimbangi dengan penyajian informasi yang memadai, sering kali informasi tersebut masih harus digali ulang dari data yang jumlahnya sangat besar. Kemampuan teknologi informasi untuk mengumpulkan dan menyimpan berbagai tipe data jauh meninggalkan kemampuan untuk menganalisis, meringkas dan mengekstrak pengetahuan dari data. Metode tradisional untuk menganalisis data yang ada, tidak dapat menangani data dalam jumlah besar.

Pemanfaatan data yang ada di dalam sistem informasi untuk menunjang kegiatan pengambilan keputusan, tidak cukup hanya mengandalkan data operasional saja, diperlukan suatu analisis data untuk menggali potensi-potensi informasi yang ada. Para pengambil keputusan berusaha untuk memanfaatkan gudang data yang sudah dimiliki untuk menggali informasi yang berguna membantu mengambil keputusan, hal ini mendorong munculnya cabang ilmu baru untuk mengatasi masalah penggalian informasi atau pola yang penting atau menarik dari data dalam jumlah

besar, yang disebut dengan *data mining*. Penggunaan teknik *data mining* diharapkan dapat memberikan pengetahuan-pengetahuan yang sebelumnya tersembunyi di dalam gudang data sehingga menjadi informasi yang berharga.

Perguruan tinggi saat ini dituntut untuk memiliki kemampuan bersaing dengan memanfaatkan semua sumber daya yang dimiliki. Selain sumber daya sarana, prasarana, dan manusia, sistem informasi adalah salah satu sumber daya yang dapat digunakan untuk meningkatkan kemampuan bersaing. Sistem informasi dapat digunakan untuk mendapatkan, mengolah dan menyebarkan informasi untuk menunjang kegiatan operasional sehari-hari sekaligus menunjang kegiatan pengambilan keputusan strategis.

Salah satu indikator perguruan tinggi yang mampu bersaing dapat dilihat dari keberhasilan mahasiswa dari perguruan tinggi tersebut dalam menempuh studinya. Keberhasilan seorang mahasiswa dapat dilihat dari indeks prestasi yang dicapainya. Indeks prestasi merupakan penilaian keberhasilan belajar mahasiswa yang dinyatakan dengan nilai kredit rata-rata yang merupakan satuan nilai akhir yang menggambarkan mutu bahwa mahasiswa telah menyelesaikan seluruh mata kuliah pada satu semester.

Banyak faktor yang menjadi penghalang bagi mahasiswa mencapai dan mempertahankan indeks prestasi tinggi yang mencerminkan usaha mereka secara keseluruhan selama masa kuliah di perguruan tinggi. Faktor tersebut dapat ditargetkan oleh pihak perguruan tinggi sebagai tindakan mengembangkan strategi untuk meningkatkan prestasi mahasiswa dan meningkatkan kinerja akademik

dengan cara memantau perkembangan kinerja mereka. Oleh karena itu, evaluasi kinerja akademik merupakan salah satu dasar untuk memantau perkembangan prestasi akademik mahasiswa didalam perguruan tinggi.

Berdasarkan isu ini, pengelompokan mahasiswa berdasarkan nilai akademik diperlukan untuk mengevaluasi kinerja akademik mahasiswa sebelum mereka lulus, agar nantinya mahasiswa berhasil melaksanakan studinya secara maksimal. Pengelompokan mahasiswa ini didasarkan pada predikat kelulusan yang akan mereka dapat berdasarkan nilai indeks prestasi selama kuliah.

Untuk memecahkan permasalahan tersebut, perlu adanya suatu sistem atau program yang khusus menangani hal tersebut agar pengelompokan mahasiswa dapat berjalan dengan optimal. Clustering merupakan suatu metode yang digunakan untuk mengelompokkan obyek-obyek berdasarkan kesamaan karakteristik diantara obyek-obyek tertentu. Algoritma K-means adalah salah satu dari metode clustering yang akan digunakan peneliti untuk memecahkan masalah pengelompokan mahasiswa untuk mengevaluasi kinerja akademik mahasiswa.

Seperti halnya dalam Al-Qur'an yang menjelaskan tentang pengelompokan umat manusia berdasarkan tingkat keimanannya sebagai berikut :

وَكُنْتُمْ أَزْوَاجًا ثَلَاثَةً ﴿٧﴾ فَأَصْحَابُ الْمَيْمَنَةِ مَا أَصْحَابُ الْمَيْمَنَةِ ﴿٨﴾ وَأَصْحَابُ الْمَشْأَمَةِ مَا أَصْحَابُ الْمَشْأَمَةِ ﴿٩﴾ وَالسَّابِقُونَ السَّابِقُونَ ﴿١٠﴾

Artinya :

(7) Dan kamu menjadi tiga golongan. (8) Yaitu golongan kanan, alangkah mulianya golongan kanan itu. (9) Dan golongan kiri, alangkah sengsaranya

golongan kiri itu. (10) Dan orang-orang yang paling dahulu (beriman), merekalah yang paling dahulu (masuk surga) (QS.Al-Waqi'ah : 7-10)

Berdasarkan tafsir *Al-Qarni*, yang menjelaskan firman Allah Ta'ala dalam surat Al-Waqi'ah: 7-10. “Kalian wahai manusia pada hari kiamat terbagi menjadi tiga golongan. Golongan pertama adalah golongan kanan yang merupakan pemilik derajat tinggi. Alangkah tingginya kedudukan mereka semasa hidup di dunia. Golongan kedua adalah golongan kiri yang merupakan pemilik derajat yang rendah. Alangkah sengsaranya kondisi mereka dan betapa ruginya usaha mereka karena buruknya amal perbuatan mereka semasa hidup di dunia. Golongan ke tiga adalah orang-orang yang paling dahulu memperoleh derajat tinggi di surga sebagai pahala atas perbuatan mereka”.

Maksud yang tersirat dari tafsir di atas, bahwasanya tiap mahasiswa akan memperoleh hasil yang berbeda-beda pada saat kelulusan tergantung dari kesungguhan mereka selama menempuh pendidikan yang disimbolkan dengan predikat kelulusan mahasiswa.

Disamping itu, di dalam Al-Qur'an terdapat ayat-ayat yang menyebutkan tentang evaluasi, yakni :

يَتَأْتِيهَا الَّذِينَ ءَامَنُوا اتَّقُوا اللَّهَ وَلْتَنْظُرْ نَفْسٌ مَّا قَدَّمَتْ لِغَدٍ ۖ وَاتَّقُوا اللَّهَ ۚ إِنَّ اللَّهَ خَبِيرٌ بِمَا تَعْمَلُونَ



Artinya :

Hai orang-orang yang beriman, bertakwalah kepada Allah dan hendaklah Setiap diri memperhatikan apa yang telah diperbuatnya untuk hari esok (akhirat); dan bertakwalah kepada Allah, Sesungguhnya Allah Maha mengetahui apa yang kamu kerjakan. (QS. Al Hasyr: 18).

Berdasarkan masalah diatas maka penelitian ini akan memaparkan bagaimana menggunakan algoritma *K-Means Clustering* dalam mengelompokkan mahasiswa berdasarkan data akademik berupa predikat kelulusan berdasarkan nilai akademik yang hasilnya diharapkan dapat digunakan sebagai bahan untuk mengevaluasi kinerja akademik mahasiswa pada Jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.

1.2 RUMUSAN MASALAH

Berdasarkan uraian latar belakang masalah, maka dalam skripsi ini dapat dirumuskan permasalahan mengenai ”Bagaimana menggunakan metode *K-Means clustering* untuk mengevaluasi kinerja akademik Mahasiswa Jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.”

1.3 BATASAN MASALAH

Untuk memfokuskan pada permasalahan diatas, maka permasalahan dalam penelitian ini dibatasi sebagai berikut :

1. Data diambil dari mahasiswa angkatan tahun 2007 yang sudah lulus pada jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Variabel yang digunakan dalam penelitian yaitu, Indeks Prestasi Kumulatif selama empat semester awal, jenis asal sekolah, jalur masuk perguruan tinggi dan pernah di pesantren.

1.4 TUJUAN DAN MANFAAT PENELITIAN

a. Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah :

1. Mengimplementasikan metode *K-Means Clustering* pada data yang bersumber dari perguruan tinggi .
2. Membangun model aturan dalam mengevaluasi kinerja akademik mahasiswa jurusan teknik informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang. menggunakan algoritma *K-Means Clustering*.

b. Manfaat Penelitian

Adapun manfaat yang diharapkan dalam penelitian ini adalah :

1. Hasil penelitian ini dapat dijadikan sebagai model untuk mengevaluasi kinerja akademik mahasiswa jurusan teknik informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Penelitian ini dapat dijadikan masukan bagi pengambil keputusan di Jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang untuk meningkatkan prestasi akademik mahasiswa.
3. Penelitian ini dapat dijadikan sebagai referensi bagi peneliti berikutnya yang berkaitan dengan penggunaan metode *K-Means Clustering* di dunia akademik.

1.5 METODOLOGI PENELITIAN

Metodologi penelitian yang digunakan dalam penulisan skripsi ini adalah sebagai berikut :

1.5.1 Studi Pustaka

Bertujuan untuk mempelajari teori-teori dengan membaca beberapa buku dan literatur yang berhubungan dengan permasalahan yang di ambil. Khususnya kajian mengenai algoritma *K-Means Clustering*, bagaimana cara kerja algoritma *K-Means Clustering* dalam meng-cluster data, sistem informasi dan manajemen database yang digunakan.

1.5.2 Definisi dan Analisa Kebutuhan

Analisa kebutuhan yang dilakukan dalam penelitian ini adalah melakukan studi lapangan ke institusi. Studi lapangan ini dilakukan dengan tujuan untuk mengidentifikasi masalah yang dihadapi oleh setiap institusi. Dengan mengidentifikasi masalah maka kita nantinya bisa menentukan latar belakang masalah yang menjelaskan situasi dan hal-hal yang menyangkut bidang yang akan diteliti. Sehingga kita dapat merumuskan masalah yang akan diteliti secara tepat. Kemudian melakukan pengumpulan data dan data yang diambil adalah data mahasiswa jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.

1.5.3 Perancangan Arsitektur Sistem

Setelah mencari literatur dan memahami konsep-konsep dasarnya, dilakukanlah perancangan sistem terhadap konsep dasar tersebut. Analisa yang dilakukan adalah bagaimana keterkaitan antara konsep-konsep dasar tersebut dan bagaimana menyelesaikan klasterisasi data menggunakan algoritma *K-Means Clustering*. Perancangan sistem ini meliputi :

- a. Perancangan desain sistem aplikasi yang meliputi perancangan Data Flow Diagram yang menggambarkan aliran data dalam aplikasi serta perancangan model sistem database dan tabel yang digunakan untuk menyimpan data yang telah dikumpulkan,
- b. Perancangan proses pada aplikasi yang meliputi perancangan input data mahasiswa, perancangan model algoritma k- means untuk melakukan klasterisasi data mahasiswa dan Evaluasi kinerja akademik mahasiswa, perancangan model pelabelan klaster untuk menentukan label pada hasil klasterisasi KMeans dan analisa hasil klasterisasi Kmeans.
- c. Perencanaan desain user interface / form yang digunakan pada aplikasi untuk mengevaluasi kinerja akademik mahasiswa.

1.5.4 Pengujian dan Analisa

Metode ini dilakukan agar sistem aplikasi yang telah di buat sudah sesuai serta untuk mengetahui keberhasilan pembuatan sistem, apakah memenuhi spesifikasi yang direncanakan. Pada tahap ini dilakukan uji coba terhadap hasil implementasi yang telah dilakukan. Kemudian dievaluasi dan diperbaiki untuk mendapatkan hasil yang maksimal. Serta dilakukan pengkajian dan

analisis keluaran yang berasal dari perangkat lunak yang telah dibuat sebelumnya.

1.6 SISTEMATIKA PENULISAN

BAB I PENDAHULUAN

Berisi ulasan permasalahan sehingga dapat diangkat menjadi suatu kajian tugas akhir. Serta latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, metodologi, dan sistematika penulisan.

BAB II LANDASAN TEORI

Membahas teori dasar yang berhubungan dengan pembuatan sistem. Dasar teori dari literature dan referensi yang berkaitan dengan permasalahan yang diangkat sebagai bahan tugas akhir.

BAB III ANALISA DAN PERANCANGAN SISTEM

Menjelaskan mengenai cara kerja beserta analisa perencanaan sistem sesuai dengan teori-teori dasar yang ada.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Berisi tentang pembuatan program (*software*), pengoperasian serta pengujian sistem pengelompokan data mahasiswa untuk mengevaluasi kinerja akademik mahasiswa.

BAB V KESIMPULAN DAN SARAN

Merupakan kesimpulan dari pembahasan pada bab-bab sebelumnya, yang merupakan hasil akhir dari pembahasan masalah.

BAB II

LANDASAN TEORI

2.1 Konsep Evaluasi Kinerja Akademik Mahasiswa

Evaluasi dalam buku Pedoman Pendidikan BAB Sistem Akademik pada pembahasan Sistem Evaluasi dijelaskan sebagai suatu sistem penilaian yang dilakukan untuk mengukur kemampuan dan kecakapan mahasiswa dalam menerima, memahami, dan menalar bahan studi yang diberikan sesuai dengan kurikulum dan silabus yang telah ditetapkan, serta untuk mengetahui perubahan sikap dan ketrampilan mahasiswa.

Tujuan evaluasi adalah :

- a. Untuk menilai kemampuan dan kecakapan mahasiswa dalam rangka memahami dan menguasai bahan studi yang disajikan, perubahan sikap dan ketrampilan dalam waktu tertentu.
- b. Untuk mengetahui keberhasilan penyajian bahan studi oleh tenaga pengajar dan keberhasilan penyelenggaraan program pendidikan.
- c. Untuk mengetahui kedudukan seorang mahasiswa dalam suatu kelompok menurut kemampuan masing-masing.

Menurut Yaslis Ilyas (2002) Kinerja adalah penampilan hasil karya personel baik kuantitas maupun kualitas dalam suatu organisasi. Kinerja dapat merupakan penampilan individu maupun kelompok kerja personel. Penampilan hasil karya tidak

terbatas kepada personel yang memegang jabatan fungsional maupun structural, tetapi juga kepada keseluruhan jajaran personel didalam organisasi.

Deskripsi dari kinerja menyangkut 3 komponen penting yaitu : tujuan, ukuran, dan penilaian. Penentuan tujuan dari setiap organisasi merupakan strategi untuk meningkatkan kinerja. Tujuan ini akan memberikan arah dan mempengaruhi bagaimana seharusnya perilaku kerja yang diharapkan organisasi terhadap setiap personel. Walaupun demikian, penentuan tujuan saja tidaklah cukup, sebab itu dibutuhkan ukuran apakah seorang personel telah mencapai kinerja yang diharapkan. Untuk itu ukuran kuantitatif dan kualitatif standar kinerja untuk setiap tugas dan jabatan personel memegang peranan penting.

Aspek ketiga dari definisi kinerja adalah penilaian. Penilaian kinerja secara regular yang dikaitkan dengan proses pencapaian tujuan kinerja setiap personel. Tindakan ini akan membuat personel untuk senantiasa berorientasi terhadap tujuan dan perilaku kerja sesuai dan searah dengan tujuan yang hendak dicapai.

Berdasarkan pengertian-pengertian sistem evaluasi dan kinerja secara umum maka peneliti menyimpulkan bahwa di dalam penelitian ini padanan kata yang tepat digunakan adalah Evaluasi Kinerja Akademik Mahasiswa.

2.2 Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual (Pramudiono dalam Kusrini, 2009:3). *Data mining* adalah proses yang menggunakan teknik statistik, *matematika*, *kecerdasan buatan* dan *machine learning*

untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Turban, dkk dalam Kusri, 2009:3). *Data mining* merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran *mesin*, *pengenalan pola*, *statistik*, *database*, dan *visualisasi* untuk penanganan permasalahan pengambilan informasi dari *database* yang besar (Larose dalam Kusri, 2009:4). Istilah *data mining* dan *knowledge discovery in databases* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*. Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Fayyad dalam Kusri, 2009:6):

1. ***Data Selection***

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *knowledge discovery in databases* dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. ***Pre-Processing / Cleaning***

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*). Juga

dilakukan proses *enrichmen*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

3. *Transformation*

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data mining*

Data Mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Interpretation / Evaluation*

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

Berdasarkan beberapa pengertian tersebut dapat ditarik kesimpulan bahwa *data mining* adalah suatu teknik menggali informasi berharga yang terpendam atau

tersembunyi pada suatu koleksi data (*database*) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. Kata *mining* sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu *data mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, *statistik* dan *database* (Pramudiono dalam Kusriani, 2009:6).

2.2.1 Pengelompokan Data Mining

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Larose dalam Kusriani, 2009:10) :

1. Deskripsi (*Description*)

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. *Deskripsi* dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi (*Estimation*)

Estimasi hampir sama dengan klasifikasi, kecuali *variabel* target estimasi lebih ke arah *numerik* daripada ke arah *kategori*. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari *variabel* target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai

dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh, akan dilakukan estimasi tekanan darah *sistolik* pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, *indeks berat badan*, dan *level sodium darah*. Hubungan antara tekanan darah *sistolik* dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

3. **Prediksi (*Prediction*)**

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa datang. Contoh prediksi dalam bisnis dan penelitian adalah :

- Prediksi harga beras dalam tiga bulan yang akan datang.
- Prediksi persentase kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikkan.

Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. **Klasifikasi (*Classification*)**

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

5. **Pengklusteran (*Clustering*)**

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah kumpulan yang memiliki *record* kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam *cluster* lain.

Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya *variabel target* dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (*homogen*), yang mana kemiripan *record* dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

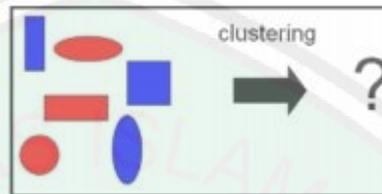
6. Asosiasi (*Assosiation*)

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

2.3 Algoritma Clustering

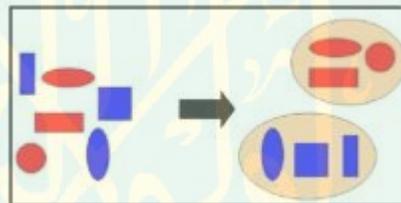
Cluster adalah sekumpulan objek yang mempunyai “*Kesamaan*” diantara anggotanya dan memiliki “*Ketidaksamaan*” dengan objek lain pada cluster lainnya, dengan kata lain sebuah *cluster* adalah sekumpulan objek yang digabung bersama karena persamaan atau kedekatannya. *Clustering* adalah proses membuat

pengelompokkan sehingga semua anggota dari setiap partisi mempunyai persamaan berdasarkan matrik tertentu. Gambar 2.1 berikut menunjukkan contoh data yang akan dilakukan klasterisasi :



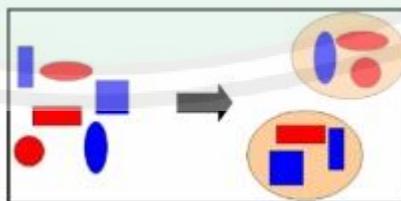
Gambar 2.1 Data sebelum di klasterisasi

Jika data dilakukan *clustering* (pengelompokkan) berdasarkan warna, maka pengelompokkannya seperti yang terlihat pada gambar 2.2



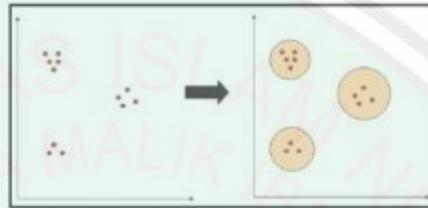
Gambar 2.2 Klasterisasi berdasarkan similaritas (kesamaan) warna

Jika data dilakukan *clustering* (pengelompokkan) berdasarkan bentuk, maka pengelompokkannya dapat dilihat seperti gambar 2.3 :



Gambar 2.3 Klasterisasi berdasarkan similaritas (kesamaan) bentuk

Selain dengan menggunakan *similaritas* (kesamaan) berdasarkan bentuk dan warna, clustering juga bisa dilakukan dengan menggunakan *similaritas* berdasarkan jarak, artinya data yang memiliki jarak berdekatan akan membentuk satu *cluster*, contohnya seperti dapat dilihat pada gambar 2.4 :



Gambar 2.4 Klasterisasi berdasarkan similaritas (kesamaan) jarak

Ada beberapa perbedaan antara metode klasifikasi dan metode *clustering*, dimana pada dasarnya terdapat tiga poin perbedaan yaitu : data, label dan analisa hasil. Perbedaan tersebut dapat ditabelkan seperti tabel 2.1 berikut:

Tabel 2.1 Perbedaan klasifikasi dan klasterisasi

Perbedaan	Klasifikasi	Klasterisasi
Data	Supervised	Unsupervised
Label	Ya	Tidak
Analisa Hasil	Error Ratio	Variance

Data *supervised* pada klasifikasi artinya data melalui pembelajaran terbimbing, sedangkan data *unsupervised* pada klasterisasi artinya data tidak melalui pembelajaran terbimbing. Analisa hasil pada klasterisasi dinyatakan dengan *variance* yang menunjukkan variansi data dalam satu *cluster*, sedangkan klasifikasi analisa

hasil diukur menggunakan rasio kesalahan (*error ratio*). Pada dataset yang digunakan oleh klasifikasi terdapat satu *attribut (label)* yang berfungsi sebagai *attribut* target, sedangkan dataset pada klasterisasi tidak terdapat *attribut (label)* sebagai *attribut* target.

2.3.1 Karakteristik Clustering

Ada beberapa karakteristik dari *clustering*, masing-masing akan dijelaskan berikut ini :

1. *Partitioning Clustering*

- a. Disebut juga *exclusive clustering*
- b. Setiap data harus termasuk dalam *cluster* tertentu
- c. Memungkinkan bagi setiap data yang termasuk cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke cluster yang lain.

Contoh : *K-Means, residual analysis.*

2. *Hierarchical Clustering*

- a. Setiap data harus masuk ke dalam *cluster* tertentu
- b. Suatu data yang masuk kedalam *cluster* tertentu pada suatu tahapan proses, tidak dapat berpindah ke *cluster* lain.

Contoh: *Single Linkage, Centroid Linkage, Complete Linkage*

3. *Overlapping Clustering*

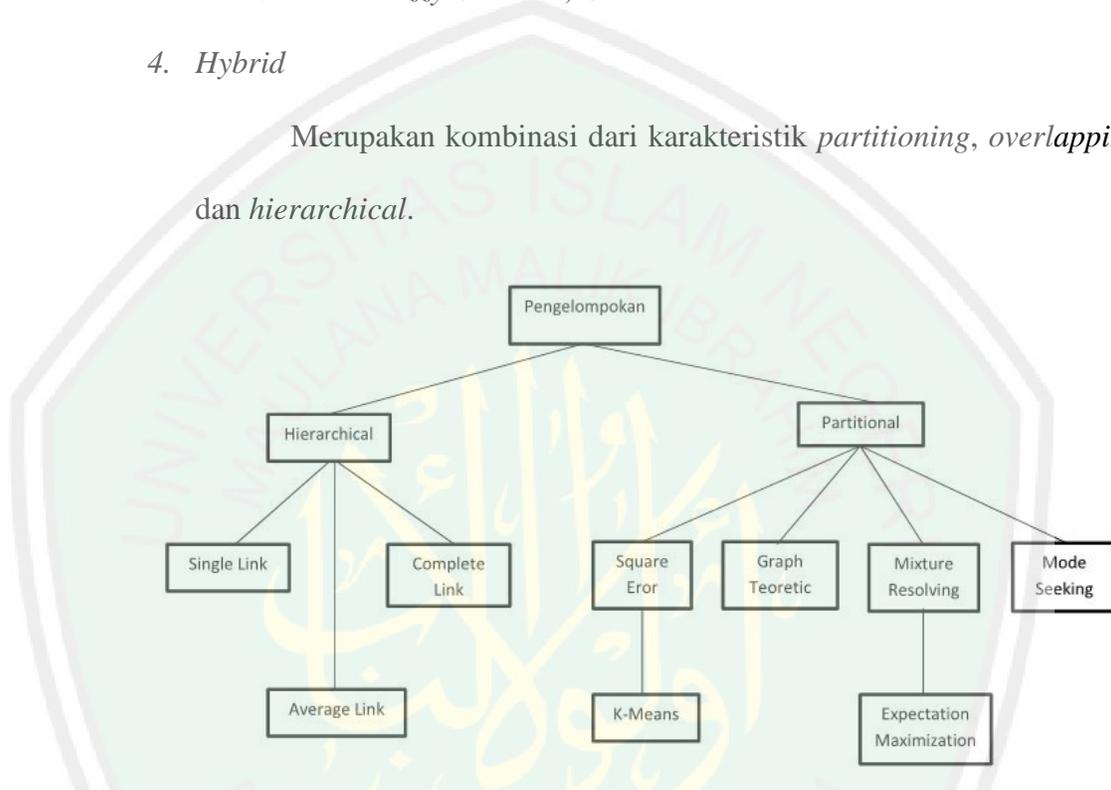
- a. Setiap data memungkinkan termasuk ke beberapa cluster

- b. Data mempunyai nilai keanggotaan (membership) pada beberapa cluster.

Contoh : *Fuzzy C-means, Gaussian Mixture*

4. Hybrid

Merupakan kombinasi dari karakteristik *partitioning*, *overlapping* dan *hierarchical*.



Gambar 2.5 Karakteristik Klasterisasi

2.4 Metode Pengelompokan

Metode pengelompokan pada dasarnya ada dua, yaitu metode pengelompokan Hirarki (*Hierarchical Clustering Method*) dan metode non Hirarki (*Non Hierarchical Clustering Method*). Metode pengelompokan hirarki digunakan apabila belum ada informasi jumlah kelompok yang akan dipilih. Sedangkan metode pengelompokan Non Hirarki bertujuan untuk mengelompokkan n objek kedalam k kelompok ($k < n$), dimana nilai k telah ditentukan sebelumnya. Salah satu prosedur

pengelompokan pada Non Hirarki adalah dengan menggunakan metode *k-means*. Metode ini merupakan metode pengelompokan yang bertujuan untuk mengelompokkan objek sedemikian hingga jarak tiap-tiap objek ke pusat kelompok didalam suatu kelompok adalah minimum.

2.5 Distance Space

Distance Space berfungsi untuk menghitung jarak antara data dan *centroid*. Ada beberapa macam *distance space* yang sudah diimplementasikan salah satunya adalah :

2.5.1 Euclidean distance space.

Euclidean sering digunakan karena penghitungan jarak dalam *distance space* ini merupakan jarak terpendek yang bisa didapatkan antara dua titik yang diperhitungkan. Jarak antara dua titik dapat dihitung dengan cara:

$$d = |x - y| = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$$

Dimana:

d : jarak

p : dimensi data

x : titik data pertama,

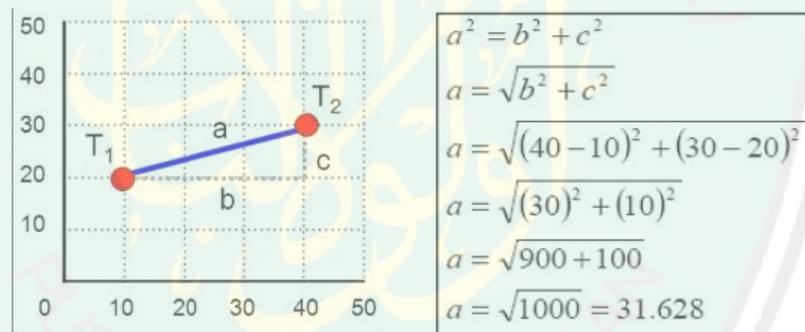
y : titik data kedua,

Dalam *Euclidean* perhitungan yang dilakukan merupakan jarak terpendek antara dua titik. Jika ada n titik pengamatan dengan p variabel,

maka sebelum dilakukan pengelompokan data atau objek, terlebih dahulu menentukan ukuran kedekatan sifat antar data. Ukuran data yang bisa digunakan adalah *euclidius (euclidian distance)*, antara dua titik dari p dimensi pengamatan. Jika antar titik $X (x_1, x_2, \dots, x_n)$ dan titik $Y=(y_1, y_2, \dots, y_n)$ di tentukan dengan rumus:

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$$

Penghitungan jarak dengan *Euclidian Distance* untuk dua titik seperti diilustrasikan pada gambar 2.6:



Gambar 2.6 Perhitungan Jarak Dua Titik

Dari penggambaran diatas dapat diartikan bahwa semakin kecil jarak atau nilai d , maka semakin besar keserupaan antar objek tersebut.

2.6 Inner Product

Inner Product yang merupakan istilah lain dari dot product atau dapat juga disebut sebagai scalar product adalah perkalian antara dua vector kolom, sehingga hasilnya berupa scalar. Rumus dasar dari inner product adalah sebagai berikut :

$$x \cdot y = \sum_{i=1}^n x_i \cdot y_i$$

$$a \cdot b = \sum_{i=1}^n a_i \cdot b_i + a_2 \cdot b_2 + \dots + a_n \cdot b_n$$

Contoh perhitungan dari persamaan di atas dapat dilihat pada contoh berikut :

$$[3 \ 1 \ 4] \cdot [2 \ 5 \ 1] = (3)(2) + (1)(5) + (4)(1) = 15$$

$$[1 \ 3 \ 2] \cdot \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} = (1)(3) + (3)(5) + (2)(2) = 22$$

2.7 K-Means

Algoritma *K-Means* adalah Metode *clustering* non hierarchical berbasis jarak yang membagi data kedalam cluster dan algoritma ini bekerja pada atribut numerik. Algoritma *K-Means* termasuk dalam *partitioning clustering* yang memisahkan data ke k daerah bagian yang terpisah. Algoritma *K-Means* sangat terkenal karena kemudahan dan kemampuannya untuk mengkluster data besar dan outlier dengan sangat cepat.

K-Means merupakan metode klusterisasi yang paling terkenal dan banyak digunakan di berbagai bidang karena sederhana, mudah diimplementasikan, memiliki kemampuan untuk mengkluster data yang besar, mampu menangani data *outlier*, dan kompleksitas waktunya *linear* $O(nKT)$ dengan n adalah jumlah dokumen, K adalah jumlah kluster, dan T adalah jumlah iterasi. Dalam algoritma *K-Means*, setiap data harus termasuk ke *cluster* tertentu pada suatu tahapan proses, pada tahapan proses berikutnya dapat berpindah ke *cluster* yang lain. Pada dasarnya penggunaan algoritma

K-Means dalam melakukan proses *clustering* tergantung dari data yang ada dan konklusi yang ingin dicapai. Untuk itu digunakan algoritma *K-Means* yang didalamnya memuat aturan sebagai berikut:

- a. Jumlah cluster yang perlu di inputkan
- b. Hanya memiliki atribut bertipe numerik

Algoritma *K-Means* pada awalnya mengambil sebagian dari banyaknya komponen dari populasi untuk dijadikan pusat *cluster* awal. Pada step ini pusat *cluster* dipilih secara acak dari sekumpulan populasi data. Berikutnya *K-Means* menguji masing-masing komponen didalam populasi data dan menandai komponen tersebut ke salah satu pusat cluster yang telah di definisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap pusat *cluster*. Posisi pusat *cluster* akan dihitung kembali sampai semua komponen data digolongkan kedalam tiap-tiap *cluster* dan terakhir akan terbentuk posisi *cluster* baru.

Algoritma *K-Means* pada dasarnya melakukan 2 proses yakni proses pendeteksian lokasi pusat *cluster* dan proses pencarian anggota dari tiap-tiap *cluster*.

Proses algoritma *K-Means* :

1. Tentukan k sebagai jumlah *cluster* yang ingin dibentuk
2. Bangkitkan k *centroids* (titik pusat *cluster*) awal secara *random*.
3. Hitung jarak setiap data ke masing-masing *centroids*.
4. Setiap data memilih *centroids* yang terdekat
5. Tentukan posisi *centroids* baru dengan cara menghitung nilai rata-rata dari data-data yang terletak pada *centroids* yang sama.

6. Kembali ke langkah 3 jika posisi *centroids* baru dengan *centroids* lama tidak sama.

Berdasarkan cara kerjanya, algoritma *K-Means* memiliki karakteristik:

1. *K-Means* sangat cepat dalam proses *clustering*
2. *K-Means* sangat sensitif pada pembangkitan *centroids* awal secara *random*
3. Memungkinkan suatu *cluster* tidak mempunyai anggota
4. Hasil *clustering* dengan *K-Means* bersifat tidak unik (Selalu berubah-ubah) – terkadang baik, terkadang jelek.

Adapun tujuan dari data *clustering* ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses *clustering*, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu *cluster* dan memaksimalkan variasi antar *cluster*.

2.8 Normalisasi Data

Normalisasi data merupakan sebuah metode untuk mengelompokkan range atau interval dari nilai-nilai yang berbeda ke dalam skala yang sama yang lebih kecil. Normalisasi penting digunakan untuk memberikan bobot yang sama terhadap nilai-nilai fitur yang berbeda dari hasil ekstraksi.

Normalisasi dapat dilakukan dengan berbagai cara, dan pada penelitian ini normalisasi yang dipergunakan adalah dengan metode *Min-Max Normalization* (Larose, 2005:36).

$$x' = \frac{(x - \min)}{(\max - \min)}$$

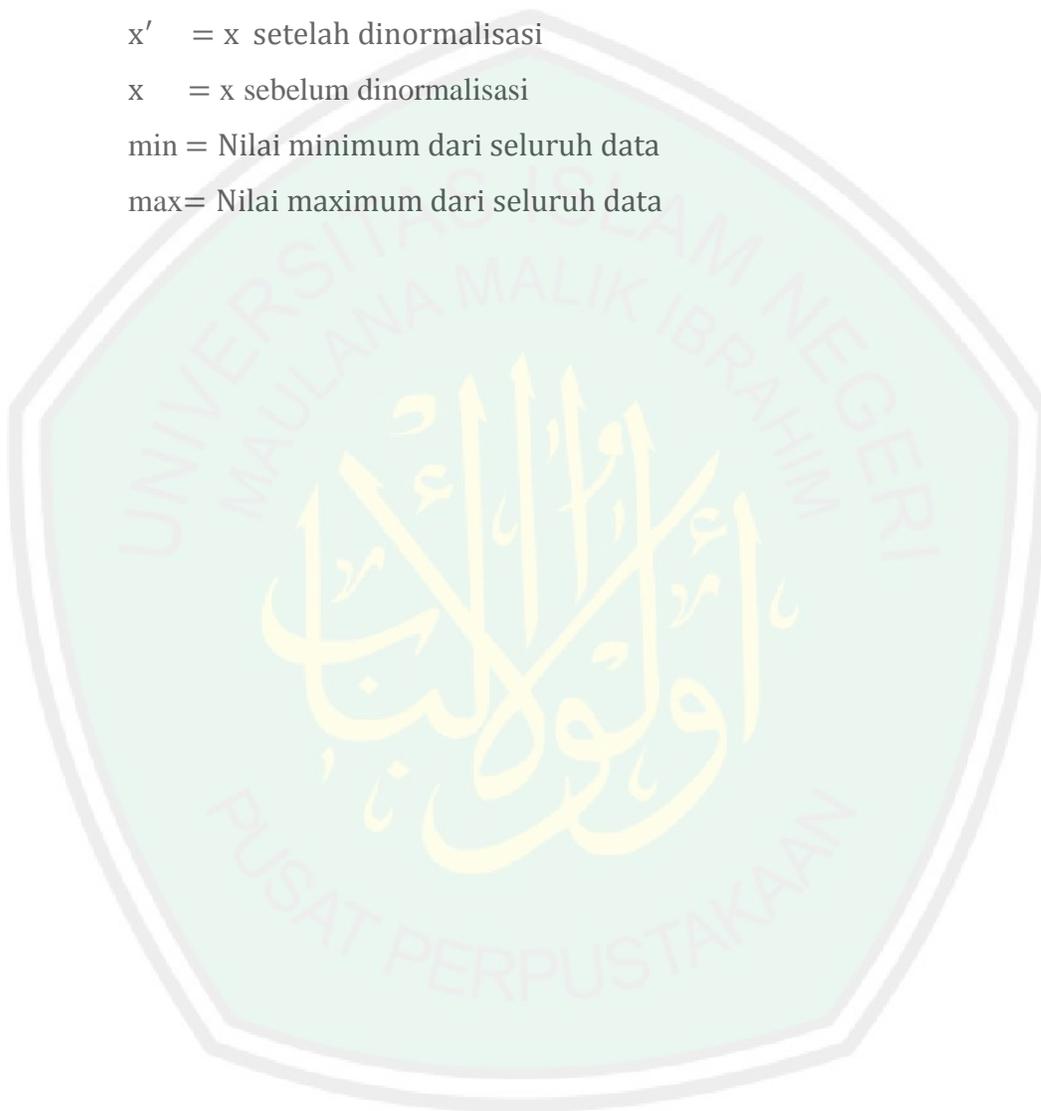
Dimana :

x' = x setelah dinormalisasi

x = x sebelum dinormalisasi

min = Nilai minimum dari seluruh data

max = Nilai maximum dari seluruh data



BAB III

ANALISIS DAN PERANCANGAN SISTEM

Bab ini menjelaskan tentang analisis dan perancangan dalam membangun aplikasi *data mining* untuk mengevaluasi kinerja akademik mahasiswa. Analisis meliputi analisis *data mining*, analisis lingkungan sistem serta analisis dalam membangun aplikasi.

3.1 Analisis Data Mining

Dalam penulisan skripsi ini akan dilakukan pengelompokan (*clustering*) mahasiswa menggunakan algoritma K-Means Clustering berdasarkan data-data mahasiswa untuk dicari kemiripannya. Dari cluster yang terbentuk dicari cluster terbaik untuk kemudian dilakukan pelabelan cluster. Tidak semua data mahasiswa akan dimasukkan dalam penghitungan sistem, hanya beberapa atribut yang kira-kira berguna dan sebarannya tidak terlalu acak. Karena data yang terlalu acak akan membuat proses *mining* memakan waktu lama dan tingkat akurasi pun rendah. Data mahasiswa yang akan digunakan sebagai atribut penghitungan antara lain jalur masuk, jenis asal sekolah, pernah tinggal di pesantren dan indeks prestasi kumulatif (IPK) empat semester awal.

3.1.1 Sumber Data

Data yang digunakan dalam penulisan skripsi ini terdiri dari dua sumber data, yaitu data induk Mahasiswa dan data kelulusan Mahasiswa.

1. Data Induk Mahasiswa

Data induk mahasiswa adalah data mahasiswa yang didata ketika mahasiswa pertama kali masuk perguruan tinggi setelah melakukan registrasi ulang. Data yang dicatat adalah identitas pribadi mahasiswa dan identitas sekolah asal mahasiswa.

Tabel 3.1 Tabel Data Induk Mahasiswa

Atribut	Keterangan
NIM	Nomor Induk Mahasiswa (NIM) adalah kode yang dimiliki mahasiswa sebagai nomor unik identitas di perguruan tinggi. Terdiri dari 8 digit yang mempresentasikan tahun masuk, jurusan dan nomor unik mahasiswa.
Nama	Merupakan nama lengkap mahasiswa yang bersangkutan
Jenis Kelamin	Merupakan jenis kelamin mahasiswa
Tempat, tanggal lahir	Merupakan kota kabupaten atau kotamadya tempat dan tanggal mahasiswa yang bersangkutan dilahirkan
Alamat Mahasiswa	Merupakan alamat asal lengkap mahasiswa
Jurusan	Jurusan studi dari mahasiswa yang

	bersangkutan
Jenis Asal Sekolah	Merupakan jenis asal sekolah menengah lanjutan mahasiswa yang bersangkutan
Jalur Masuk	Merupakan jalur masuk mahasiswa saat diterima di perguruan tinggi
Pernah Mondok	Merupakan keterangan pernah tinggal di pesantren dari mahasiswa yang bersangkutan
Nama Wali	Merupakan nama orang tua atau wali mahasiswa yang bersangkutan
Alamat Wali	Merupakan alamat orang tua atau wali mahasiswa yang bersangkutan
Pendidikan Wali	Merupakan pendidikan orang tua atau wali mahasiswa yang bersangkutan
IPK	Merupakan data nilai empat semester awal dari mahasiswa yang bersangkutan.

2. Data Kelulusan

Data kelulusan adalah data mahasiswa yang telah dinyatakan lulus. Data yang dicatat adalah identitas mahasiswa dan data kelengkapan kelulusan.

Tabel 3.2 Tabel Data Induk Kelulusan

Atribut	Keterangan
NIM	Nomor Induk Mahasiswa (NIM) adalah kode yang dimiliki mahasiswa sebagai nomor unik identitas di perguruan tinggi. Terdiri dari 8 digit yang mempresentasikan tahun masuk, jurusan dan nomor unik mahasiswa.
Nama	Merupakan nama lengkap mahasiswa yang bersangkutan
Tempat, Tanggal Lahir	Merupakan kota kabupaten atau kotamadya tempat dan tanggal mahasiswa yang bersangkutan dilahirkan
Jurusan	Jurusan studi dari mahasiswa yang bersangkutan
Tanggal lulus	Merupakan tanggal dari mahasiswa yang bersangkutan dinyatakan lulus
Indeks Prestasi Kumulatif (IPK)	Indeks Prestasi Kumulatif (IPK) adalah ukuran mahasiswa sampai pada waktu

	tertentu yang dapat dihitung berdasarkan jumlah (satuan kredit semester) SKS mata kuliah yang diambil sampai periode tertentu dikalikan dengan nilai bobot masing-masing mata kuliah dibagi dengan jumlah seluruh SKS mata kuliah
No Ijasah UIN	No Ijasah Kelulusan
Yudisium	Tanggal Yudisium
Judul Skripsi	Merupakan judul skripsi dari mahasiswa yang bersangkutan

Data yang diambil adalah data mahasiswa angkatan 2007 jurusan Teknik Informatika UIN Maulana Malik Ibrahim Malang. Hal ini didasarkan pada kebutuhan data yang akan dihubungkan dengan data kelulusan. Sedangkan data kelulusan yang diambil adalah data kelulusan mahasiswa angkatan 2007 yang telah menyelesaikan studinya.

3.1.2 Data Yang Digunakan

Dalam penulisan skripsi ini tidak semua data sumber digunakan, oleh karena itu perlu dilakukan pembersihan data agar data yang diolah benar-benar relevan dengan yang dibutuhkan. Pembersihan ini penting guna meningkatkan

performa proses K-Means. Cara pembersihan dilakukan dengan tidak memasukkan ke database data mentah yang tidak digunakan.

Atribut yang digunakan dalam data induk mahasiswa meliputi :

1. Atribut NIM sebagai *primary key* untuk menghubungkan dengan data kelulusan.
2. Atribut Nama sebagai informasi identitas nama mahasiswa
3. Atribut Jenis Kelamin sebagai informasi jenis kelamin mahasiswa
4. Atribut Tempat dan tanggal lahir sebagai informasi tempat dan tanggal lahir mahasiswa
5. Atribut Alamat sebagai informasi identitas alamat mahasiswa
6. Atribut Jurusan sebagai informasi identitas jurusan mahasiswa
7. Atribut Jalur masuk sebagai atribut pengukuran dalam proses K-Means Clustering
8. Atribut Asal jenis sekolah sebagai atribut pengukuran dalam proses K-Means Clustering
9. Atribut Pernah di pesantren sebagai atribut pengukuran dalam proses K-Means Clustering
10. Atribut IPK empat semester awal sebagai atribut pengukuran dalam proses K-Means Clustering

Atribut yang digunakan dalam data kelulusan meliputi :

1. NIM digunakan sebagai *primary key* untuk menghubungkan dengan data induk mahasiswa

2. Indeks Prestasi Kumulatif (IPK) digunakan sebagai pembanding hasil program dengan data kelulusan sebenarnya
3. Predikat Kelulusan sebagai pembanding hasil program dengan data kelulusan sebenarnya.

3.1.3 Transformasi Data

Transformasi data merupakan proses pengubahan atau penggabungan data ke dalam format yang sesuai untuk diproses data mining. Seringkali data yang digunakan dalam proses data mining mempunyai format yang belum langsung bisa digunakan, oleh karena itu perlu dirubah formatnya.

Dalam penulisan skripsi ini penulis berusaha melakukan pengelompokan berdasarkan data-data atribut mahasiswa untuk dicari kemiripannya berdasarkan jarak anggota ke pusat cluster yang dibentuk. Atribut pengukuran yang digunakan antara lain : jalur masuk, jenis asal sekolah, pernah mondok, dan indeks prestasi kumulatif empat semester awal untuk dibandingkan dengan hasil akhir kelulusan yang berupa predikat kelulusan berdasarkan IPK yang didapat pada saat lulus. IPK Dikategorikan berdasarkan predikat kelulusan yang diatur dalam peraturan sistem akademik BAB III pada pembahasan yudisium pada penjelasan nomor 4 yang berbunyi “predikat kelulusan terdiri dari 3 (tiga) tingkat, yaitu memuaskan, sangat memuaskan, dan dengan pujian yang dinyatakan pada transkrip nilai. Kemudian pada penjelasan nomor 5 yang berbunyi “Penentuan predikat kelulusan pada tingkat tertentu didasarkan pada Indeks Prestasi Kumulatif yang diperoleh mahasiswa yang bersangkutan sebagai mana dapat dilihat dalam tabel 3.3

Tabel 3.3 Indeks Prestasi Kumulatif Mahasiswa

No	Indeks Prestasi Kumulatif	Predikat
1	2,00 – 2,99	Memuaskan
2	3,00 – 3,50	Sangat Memuaskan
3	3,51 – 4,00	Dengan Pujian

Dari tabel di atas data kelulusan berdasarkan IPK dapat dikategorikan menjadi tiga yaitu :

1. Predikat kelulusan memuaskan dengan IPK 2,00 – 2,99
2. Predikat kelulusan sangat memuaskan dengan IPK 3,00 – 3,50
3. Predikat kelulusan dengan pujian dengan IPK 3,51 – 4,00

3.1.3.1 Normalisasi Min – Max

Normalisasi Min –Max bertujuan menyesuaikan data latih, data validasi, atau data uji sebelum masuk proses pelatihan. Data yang digunakan adalah data bertipe numerik. Dalam sistem ini atribut bertipe *numeric* yang di normalisasikan adalah data IPK empat semester awal. Median dari range nilai IPK pada masing-masing kategori Dengan Pujian, Sangat Memuaskan, Memuaskan dan Cukup dinormalisasi sehingga berada pada range [0.1,0.9] dengan menggunakan persamaan sebagai berikut :

$$x' = \frac{(x - \min)}{(\max - \min)}$$

Dimana :

x' = x setelah dinormalisasi

x = x sebelum dinormalisasi

min = Nilai minimum dari seluruh data

max = Nilai maximum dari seluruh data

Sehingga didapatkan hasil normalisasi tiap kategori Indeks Prestasi Kumulatif sesuai dengan yang telah diatur dalam peraturan akademik adalah sebagai berikut :

Tabel 3.4 Normalisasi Indeks Prestasi Kumulatif Mahasiswa

Range IPK	Median	Nilai Normalisasi
2.01 s/d 2.74	2.375	0.49
2.75 s/d 3.49	3.135	0.68
3.50 s/d 4.00	3.75	0.83

3.1.3.2 Term Frequency

Term frequency (TF) adalah algoritma pembobotan heuristik yang menentukan bobot dokumen berdasarkan kemunculan *term* (istilah). Semakin sering sebuah istilah muncul, semakin tinggi bobot dokumen untuk istilah tersebut, dan sebaliknya.

Bobot suatu *term* t dalam suatu himpunan dilambangkan dengan $w(t,d)$. Frekuensi kemunculan *term* t dalam himpunan d dilambangkan dengan $tf(t,d)$, Sedangkan banyaknya himpunan yang digunakan dalam uji coba dilambangkan dengan N sementara $n(t)$ adalah banyaknya himpunan yang mengandung *term* t . Dari formula tersebut diturunkan kembali formula untuk menormalkan *term* frekuensi sehingga didapatkan formula:

$$tf(t,d) = \frac{\text{freq } t}{\max \text{freq } (t,d)}$$

Dalam penulisan skripsi ini, selanjutnya *Algoritma term frequency* digunakan untuk mengukur bobot atribut jalur masuk, asal sekolah dan pernah mondok berdasarkan data yang ada pada tabel data awal.

Tabel 3.5 Data Awal Parameter Mahasiswa

NIM	Predikat Kelulusan	Jalur Masuk	Jenis Asal Sekolah	Jenis Asal Sekolah
07650004	Sangat Memuaskan	PMDK	SMAN	TIDAK
07650006	Dengan Pujian	PMDK	SMAN	TIDAK
07650010	Memuaskan	PMDK	MAN	YA
07650012	Sangat Memuaskan	PMDK	MAN	TIDAK
07650014	Sangat Memuaskan	PMDK	SWASTA	TIDAK
07650015	Sangat Memuaskan	PMDK	MAN	YA
07650021	Sangat Memuaskan	PMDK	SWASTA	YA
07650022	Sangat Memuaskan	PMDK	SMAN	YA
07650025	Sangat Memuaskan	PMDK	SWASTA	TIDAK
07650026	Sangat Memuaskan	PMDK	MAN	TIDAK
07650030	Sangat Memuaskan	PMDK	SWASTA	YA
07650031	Sangat Memuaskan	PMDK	SMAN	TIDAK
07650036	Sangat Memuaskan	PMDK	SWASTA	YA
07650042	Sangat Memuaskan	PMDK	SMAN	TIDAK
07650047	Sangat Memuaskan	PMDK	SWASTA	YA
07650050	Dengan Pujian	SNMPTN	SMAN	TIDAK
07650055	Dengan Pujian	SNMPTN	SMAN	TIDAK
07650056	Dengan Pujian	SNMPTN	SMAN	TIDAK
07650064	Sangat Memuaskan	SNMPTN	SMAN	TIDAK
07650069	Sangat Memuaskan	REGULER	SMAN	TIDAK
07650073	Sangat Memuaskan	SNMPTN	SMAN	TIDAK
07650078	Dengan Pujian	SNMPTN	SMAN	TIDAK

07650079	Sangat Memuaskan	REGULER	SMAN	YA
07650080	Sangat Memuaskan	SNMPTN	SMAN	TIDAK
07650081	Sangat Memuaskan	REGULER	SMAN	TIDAK
07650082	Sangat Memuaskan	SNMPTN	SMAN	TIDAK
07650083	Sangat Memuaskan	SNMPTN	SMAN	TIDAK
07650086	Sangat Memuaskan	SNMPTN	MAN	TIDAK
07650087	Sangat Memuaskan	REGULER	SMK	YA
07650092	Sangat Memuaskan	SNMPTN	SWASTA	TIDAK
07650102	Sangat Memuaskan	REGULER	SMAN	TIDAK
07650107	Dengan Pujian	REGULER	MAN	YA
07650109	Sangat Memuaskan	REGULER	SWASTA	YA
07650114	Sangat Memuaskan	SNMPTN	SMAN	TIDAK
07650117	Sangat Memuaskan	SNMPTN	SMAN	TIDAK
07650120	Sangat Memuaskan	SNMPTN	SMAN	TIDAK
07650121	Sangat Memuaskan	SNMPTN	SMAN	TIDAK
07650124	Sangat Memuaskan	REGULER	SWASTA	YA
07650129	Sangat Memuaskan	REGULER	SWASTA	YA
07650130	Sangat Memuaskan	REGULER	MAN	YA
07650133	Sangat Memuaskan	REGULER	SMK	YA
07650136	Sangat Memuaskan	REGULER	SMAN	YA
07650142	Sangat Memuaskan	SNMPTN	SMAN	TIDAK
07650147	Sangat Memuaskan	REGULER	SWASTA	TIDAK
07650153	Sangat Memuaskan	REGULER	SWASTA	YA
07650154	Sangat Memuaskan	REGULER	SMAN	TIDAK
07650155	Sangat Memuaskan	REGULER	SMAN	TIDAK

Dari tabel 3.5 dapat diketahui *frequency* dari masing-masing atribut jalur masuk, asal jenis sekolah dan pernah mondok lebih jelasnya akan dibahas sebagai berikut :

3.1.3.2.1 Term Frequency Jalur Masuk

Dari tabel 3.6 dapat diketahui *frequency* dari masing-masing *term* jalur masuk pada tabel berikut :

Tabel 3.6 Frequency Jalur Masuk

Term Jalur Masuk	Frequency
SNMPTN	16
PMDK	15
REGULER	16
Total Freq	47

Dari tabel *frequency* jalur masuk dapat dihitung *term frequency* masing-masing atributnya adalah :

$$tf(SNMPTN, JALUR MASUK) = \frac{16}{47} = 0.34$$

$$tf(PMDK, JALUR MASUK) = \frac{15}{47} = 0.32$$

$$tf(REGULER, JALUR MASUK) = \frac{16}{47} = 0.34$$

3.1.3.2.2 Term Frequency Asal Jenis Sekolah

Dari tabel 3.7 dapat diketahui *frequency* dari masing-masing *term* asal jenis sekolah pada tabel berikut :

Tabel 3.7 Frequency Asal Jenis Sekolah

Term Asal Sekolah	Frequency
SMAN	26

MAN	7
SMK	2
SWASTA	12
Total	47

Dari tabel *frequency* asal jenis sekolah dapat dihitung *term frequency* masing-masing atributnya adalah :

$$tf(SMAN, ASAL SEKOLAH) = \frac{26}{47} = 0.553$$

$$tf(MAN, ASAL SEKOLAH) = \frac{7}{47} = 0.148$$

$$tf(SMK, ASAL SEKOLAH) = \frac{2}{47} = 0.042$$

$$tf(SWASTA, ASAL SEKOLAH) = \frac{12}{47} = 0.255$$

3.1.3.2.3 Term Frequency Pernah Mondok

Dari tabel awal dapat diketahui frequency dari masing-masing term pernah mondok pada tabel berikut :

Tabel 3.8 *Frequency* Pernah Mondok

Term Pernah Mondok	Frequency
YA	17
TIDAK	30
Total	47

Dari tabel *frequency* asal jenis sekolah dapat dihitung *term frequency* masing-masing atributnya adalah :

$$tf(YA, PERNAH MONDOK) = \frac{17}{47} = 0.361$$

$$tf(TIDAK, PERNAH MONDOK) = \frac{30}{47} = 0.638$$

3.1.4 Implementasi Algoritma K-Means

Berikut ini terdapat 10 sample data mahasiswa yang digunakan untuk implementasi perhitungan K-Means pada aplikasi evaluasi kinerja akademik mahasiswa.

Tabel 3.9 Tabel Data Sample Mahasiswa

NIM	SMA	Jalur Masuk	IPK	Pernah Mondok
07650006	SMAN	PMDK	3.17	TIDAK
07650012	MAN	PMDK	3.19	TIDAK
07650026	MAN	PMDK	3.16	TIDAK
07650030	SMA	PMDK	3.19	YA
07650042	SMAN	PMDK	2.9	TIDAK
07650047	SMA	PMDK	3.25	YA
07650050	SMAN	SNMPTN	3.24	TIDAK
07650055	SMAN	SNMPTN	3.73	TIDAK
07650056	SMAN	SNMPTN	3.54	TIDAK
07650078	SMAN	SNMPTN	3.6	TIDAK

Untuk pengelompokan data menggunakan algoritma K-Means. Proses ini diawali dengan melakukan normalisasi data dari database karena data parameter yang ada mengandung data non numerik dan terdapat nilai juga yang memiliki

range sangat jauh antara nilai terendah dengan nilai tertinggi sehingga perlu dilakukan normalisasi.

Atribut bertipe *numeric* dalam sistem ini adalah atribut IPK dengan batas nilai maximum 4 dan batas nilai minimum 0 dinormalisasikan dengan menghitung nilai median dari masing-masing *range* kategori Dengan Pujian, Sangat Memuaskan dan Memuaskan menggunakan rumus :

$$x' = \frac{(x - \min)}{(\max - \min)}$$

Dimana :

x' = x setelah dinormalisasi

x = x sebelum dinormalisasi

min = Nilai minimum dari seluruh data

max = Nilai maximum dari seluruh data

Kemudian data bertipe *non numeric* yang digunakan dalam sistem ini adalah atribut jalur masuk, jenis asal sekolah dan pernah mondok yang telah diketahui masing-masing bobotnya dengan menggunakan *algoritma term frequency* , jika dimisalkan masing-masing bobot nilainya adalah sebagai berikut :

Tabel 3.10 Tabel Bobot Nilai Atribut

Atribut	Term / Range	TF / Bobot
Jalur Masuk	SNMPTN	0.34
	PMDK	0.32
	REGULER	0.34
Asal Sekolah	SMAN	0.553
	MAN	0.148
	SMK	0.042

	SWASTA	0.255
Pernah Mondok	YA	0.361
	TIDAK	0.638
IPK	2.01 s/d 2.74	0.49
	2.75 s/d 3.49	0.68
	3.50 s/d 4.00	0.83

Kemudian dilakukan proses perhitungan normalisasi data parameter mahasiswa sehingga menghasilkan nilai normalisasi sebagai berikut :

Tabel 3.11 Tabel Hasil Normalisasi Parameter

NIM	SMA	Jalur Masuk	Pernah Mondok	IPK
07650006	0.553	0.32	0.638	0.68
07650012	0.32	0.32	0.638	0.68
07650026	0.32	0.32	0.638	0.68
07650030	0.25	0.32	0.361	0.68
07650042	0.553	0.32	0.638	0.68
07650047	0.25	0.32	0.361	0.68
07650050	0.553	0.34	0.638	0.68
07650055	0.553	0.34	0.638	0.83
07650056	0.553	0.34	0.638	0.83
07650078	0.553	0.34	0.638	0.83

Setelah data dinormalisasi maka data siap untuk dilakukan clustering. Data yang ada akan dikelompokkan menjadi 3 kelompok. Adapun langkah dari pengelompokkan data adalah sebagai berikut:

1. Tentukan pusat cluster secara acak, misalkan kita tentukan

Pusat Cluster [1]= (0.553,0.32,0.638,0.68)

Pusat Cluster [2]= (0.148,0.32,0.638,0.68)

Pusat Cluster [3]= (0.148,0.32,0.638,0.68)

Hitung jarak setiap data yang ada terhadap setiap pusat cluster dengan menggunakan rumus jarak euclidean distance. Jarak antara dua titik dapat dihitung dengan cara:

$$d = |x - y| = \sqrt{\sum_{i=1}^p (x_i - y_i)^2}$$

Dimana:

d : jarak

p : dimensi data

x : titik data pertama,

y : titik data kedua,

Misalkan untuk menghitung jarak data parameter mahasiswa pertama dengan pusat cluster pertama adalah :

$$C_{11} = \sqrt{(0.553 - 0.553)^2 + (0.32 - 0.32)^2 + (0.638 - 0.638)^2 + (0.7925 - 0.7925)^2} = 0$$

$$C_{12} = \sqrt{(0.32 - 0.553)^2 + (0.32 - 0.32)^2 + (0.638 - 0.638)^2 + (0.7925 - 0.7925)^2} = 0.405$$

$$C_{13} = \sqrt{(0.32 - 0.553)^2 + (0.32 - 0.32)^2 + (0.638 - 0.638)^2 + (0.7925 - 0.7925)^2} = 0.405$$

Tabel 3.12 Posisi Cluster Pada Iterasi Pertama

NIM	SMA	Jalur Masuk	Pernah Mondok	IPK	C1	C2	C3
07650006	0.553	0.32	0.638	0.68	0.0	0.405	0.405
07650012	0.32	0.32	0.638	0.68	0.405	0.0	0.0

07650026	0.32	0.32	0.638	0.68	0.405	0.0	0.0
07650030	0.25	0.32	0.361	0.68	0.4068	0.2969	0.2969
07650042	0.553	0.32	0.638	0.68	0.0	0.405	0.405
07650047	0.25	0.32	0.361	0.68	0.4068	0.2969	0.2969
07650050	0.553	0.34	0.638	0.68	0.0200	0.4045	0.4045
07650055	0.553	0.34	0.638	0.83	0.1513	0.4323	0.4323
07650056	0.553	0.34	0.638	0.83	0.1513	0.4323	0.4323
07650078	0.553	0.34	0.638	0.83	0.1513	0.4323	0.4323

2. Suatu data akan menjadi anggota dari suatu cluster yang memiliki jarak terkecil dari pusat clusternya. Misalkan untuk data pertama, jarak terkecil diperoleh pada *cluster* keempat, sehingga data pertama akan menjadi anggota dari cluster keempat.

3. Hitung pusat cluster baru. Untuk cluster pertama, ada 6 data yaitu data nim 07650006, 07650042, 07650050, 07650055, 07650056 dan nim 07650050 sehingga:

$$C_{11} = \left(\frac{0.553+0.553+0.553+0.553+0.553+0.553}{6} \right) = 0.553$$

$$C_{12} = \left(\frac{0.32+0.32+0.34+0.34+0.34+0.34}{6} \right) = 0.3333$$

$$C_{13} = \left(\frac{0.638+0.638+0.638+0.638+0.638+0.638}{6} \right) = 0.638$$

$$C_{14} = \left(\frac{0.68+0.68+0.68+0.83+0.83+0.83}{6} \right) = 0.755$$

Sehingga :

$$C_1 = (0.553, 0.3333, 0.638, 0.755)$$

Untuk cluster kedua, ada 4 data yaitu data nim 07650012, 07650026, 07650030 dan nim 07650047 sehingga:

$$C_{21} = \left(\frac{0.32+0.32+0.25+0.25}{4} \right) = 0.2014$$

$$C_{22} = \left(\frac{0.32+0.32+0.32+0.32}{4} \right) = 0.32$$

$$C_{23} = \left(\frac{0.638+0.638+0.361+0.361}{4} \right) = 0.4995$$

$$C_{24} = \left(\frac{0.68+0.68+0.68+0.68}{4} \right) = 0.68$$

Sehingga :

$$C_2 = (0.2014, 0.32, 0.4995, 0.68)$$

Untuk cluster ketiga tidak mempunyai anggota sehingga pusat cluster kedua adalah

$$C_3 = (0, 0, 0, 0)$$

Pusat cluster baru dari iterasi pertama :

$$\text{Pusat Cluster [0]} = (0.553, 0.3333, 0.638, 0.755)$$

$$\text{Pusat Cluster [1]} = (0.20149, 0.32, 0.4995, 0.68)$$

$$\text{Pusat Cluster [2]} = (0.0, 0.0, 0.0, 0.0)$$

Ulangi langkah 2 hingga posisi data sudah tidak mengalami perubahan.

Tabel 3.13 Tabel Posisi Cluster Pada Iterasi Kedua

NIM	SMA	Jalur Masuk	Pernah Mondok	IPK	C1	C2	C3
07650006	0.553	0.32	0.638	0.68	0.0761	0.3778	1.1303
07650012	0.32	0.32	0.638	0.68	0.4121	0.1484	0.9968
07650026	0.32	0.32	0.638	0.68	0.4121	0.1484	0.9968

07650030	0.25	0.32	0.361	0.68	0.4139	0.1484	0.8718
07650042	0.553	0.32	0.638	0.68	0.0761	0.3778	1.1303
07650047	0.25	0.32	0.361	0.68	0.4139	0.1484	0.8718
07650050	0.553	0.34	0.638	0.68	0.075	0.3783	1.1361
07650055	0.553	0.34	0.638	0.83	0.0752	0.4069	1.2318
07650056	0.553	0.34	0.638	0.83	0.0752	0.4069	1.2318
07650078	0.553	0.34	0.638	0.83	1.2318	0.4069	1.2318

Karena pada iterasi ke-1 dan ke-2 (Tabel 3.12 & 3.13) posisi cluster tidak berubah, maka iterasi dihentikan. Sehingga didapat pusat cluster terakhir dari perhitungan K-Means :

Pusat Cluster [0]=(0.553,0.3333,0.638,0.755)

Pusat Cluster [1]=(0.20149,0.32,0.4995,0.68)

Pusat Cluster [2]=(0.0,0.0,0.0,0.0)

Dari hasil *cluster* yang sudah terbentuk tersebut, selanjutnya dilakukan proses labelisasi cluster berdasarkan hasil perkalian matrix antara *centroid* terakhir dikalikan dengan nilai transposenya, dengan menggunakan Rumus sebagai berikut :

$$x \cdot y = \sum_{i=1}^n x_i \cdot y_i$$

$$a \cdot b = \sum_{i=1}^n a_i \cdot b_i + a_2 \cdot b_2 + \dots + a_n \cdot b_n$$

Contoh perhitungan dari persamaan di atas dapat dilihat pada contoh berikut :

$$[3 \ 1 \ 4] \cdot [2 \ 5 \ 1] = (3)(2) + (1)(5) + (4)(1) = 15$$

$$[1 \ 3 \ 2] \cdot \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} = (1)(3) + (3)(5) + (2)(2) = 22$$

Nilai terbesar dari hasil perkalian ini berarti diberi label Dengan Pujian, nilai tertinggi kedua diberi label Sangat Memuaskan dan nilai tertinggi ketiga diberi label Memuaskan sehingga diperoleh hasil pelabelan *cluster* sebagai berikut :

Tabel 3.14 Hasil Pelabelan Cluster

Cluster	Hasil Dot Product	Label
1	2.787978222	DENGAN PUJIAN
2	1.709805	SANGAT MEMUASKAN
3	0	CUKUP

Setelah proses pelabelan cluster dengan menghitung perkalian masing-masing pusat cluster dikalikan dengan transposenya, kemudian diurutkan. diperoleh hasil akhir pengelompokan data evaluasi kinerja akademik mahasiswa sebagai berikut :

Tabel 3.15 Hasil Perhitungan K-Means Clustering

NIM	Label Cluster
07650006	DENGAN PUJIAN
07650012	SANGAT MEMUASKAN
07650026	SANGAT MEMUASKAN
07650030	SANGAT MEMUASKAN

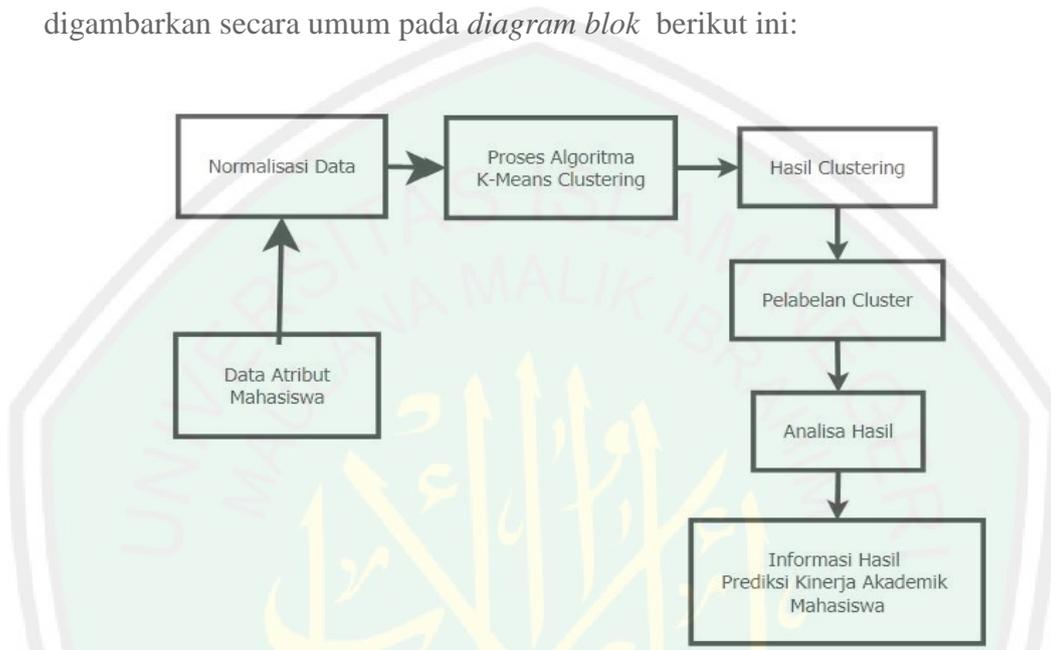
07650042	SANGAT MEMUASKAN
07650047	SANGAT MEMUASKAN
07650050	DENGAN PUJIAN
07650055	DENGAN PUJIAN
07650056	DENGAN PUJIAN
07650078	DENGAN PUJIAN

3.2 Analisis Sistem

Sistem perangkat lunak yang dikembangkan pada skripsi ini adalah untuk mengimplementasikan konsep segmentasi untuk mengevaluasi kinerja akademik mahasiswa berdasarkan data-data mahasiswa yang dijadikan sebagai parameter perhitungan yaitu, ipk selama empat semester awal, nilai ujian akhir SMA, asal sekolah, dan lama di pesantren. Data yang ada kemudian perlu dinormalisasikan karena data parameter ada yang berupa data non numerik, dan ada rentang yang jauh antara parameter satu dengan parameter lain, sehingga perlu dinormalisasikan. Data yang sudah dinormalisasi selanjutnya adalah proses segmentasi menggunakan algoritma K-Means Clustering. Kemudian dari parameter tersebut akan dikelompokkan ke dalam empat kelompok. Setiap kelompok yang terbentuk akan ditentukan pusat cluster / *centroid* yang berbeda. *Centroid* yang terbentuk ini akan dijadikan perhitungan dalam proses pelabelan cluster untuk menghasilkan kelompok mahasiswa dengan predikat Dengan Pujian, Sangat Memuaskan, Memuaskan dan Cukup.

3.3 Perancangan Proses

Diagram pembuatan sistem untuk evaluasi kinerja akademik ini digambarkan secara umum pada *diagram blok* berikut ini:

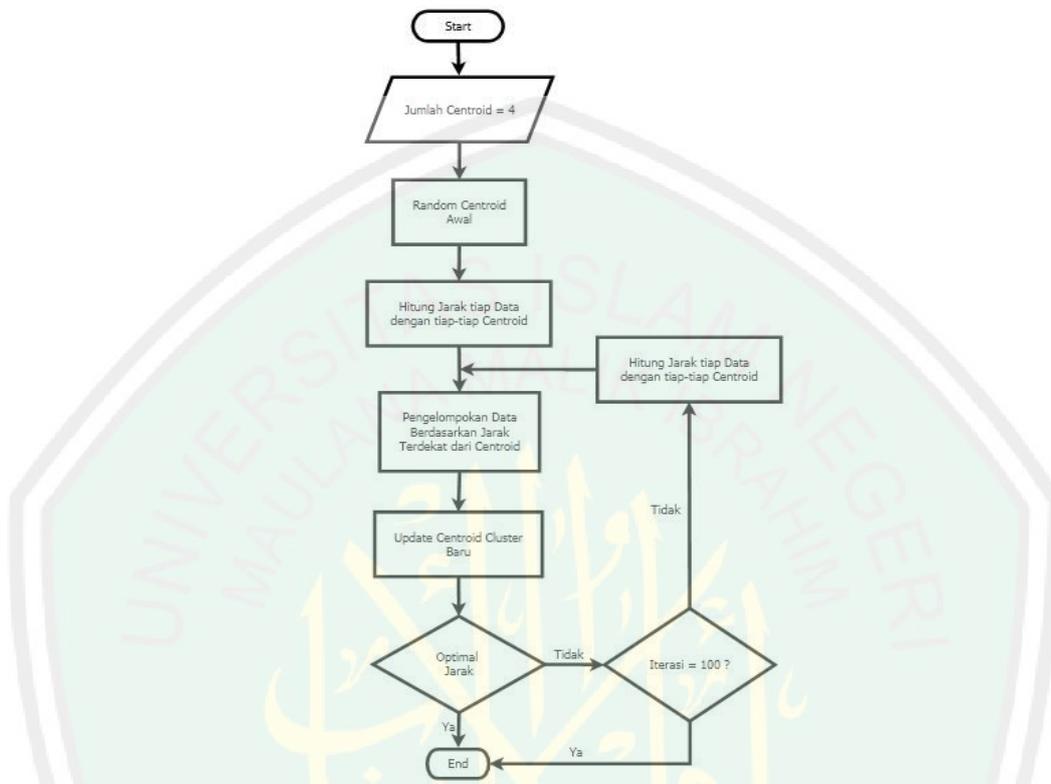


Gambar 3.1 Diagram Blok Sistem

3.4 Perancangan Sistem

Desain sistem pada aplikasi Evaluasi Kinerja Akademik Mahasiswa ini meliputi *Flowchart Algoritma K-Means* yang menjelaskan alur kerja algoritma K-Means Clustering, kemudian dibahas *Context Diagram System* yang menggambarkan lingkungan system secara umum, dan *Data Flow Diagram System* yang menggambarkan aliran data pada system beserta data item yang ada didalamnya dan Desain Tabel yang menggambarkan hubungan antar tabel pada database yang digunakan untuk menyimpan data atribut mahasiswa dan hasil klasterisasi K-Means.

3.4.1 Flowchart Algoritma K-Means



Gambar 3.2 Flow Chart K-Means Clustering

Pada tahap ini akan dilakukan proses utama yaitu pengelompokan data parameter yang diakses dari database yaitu sebuah metode algoritma K-Means Clustering. Gambar 3.2 merupakan diagram *flowchart* dari algoritma K-Means Clustering dengan asumsi bahwa parameter input adalah jumlah data set sebanyak n data dan jumlah inisialisasi centroid $K=3$ sesuai dengan jumlah pengelompokan kinerja akademik mahasiswa yaitu, Dengan Pujian, Sangat Memuaskan dan Memuaskan.

Pada diagram flowchart pada gambar 3.2 dapat dijelaskan beberapa langkah yang dilalui oleh algoritma K-Means Clustering memuat bagian-bagian sebagai berikut ini:

- N data : data set yang akan diolah sebanyak N data dimana N data tersebut terdiri dari atribut-atributnya N (nilai akhir ujian sekolah, ipk, , asal sekolah dan pernah di pesantren) yang berarti data N memiliki atribut sebanyak 4 atribut.
- K *centroid* : Inisialisasi dari pusat kluster data adalah sebanyak K dimana pusat-pusat awal tersebut digunakan sebagai banyaknya kelas yang akan tercipta. *Centroid* didapatkan secara random dari N data set yang ada.
- *Euclidian Distance*: merupakan jarak yang didapat dari perhitungan antara semua N data dengan K *centroid* dimana akan memperoleh tingkat kedekatan dengan kelas yang terdekat dengan populasi data tersebut.
- Pengelompokkan data: setelah sejumlah populasi data tersebut menemukan kedekatan dengan salah satu *centroid* yang ada maka secara otomatis populasi data tersebut masuk kedalam kelas yang memiliki *centroid* yang bersangkutan.
- Update *centroid* baru: tiap kelas yang telah tercipta tadi melakukan update *centroid* baru. Hal ini dilakukan dengan menghitung nilai rata-rata dari kelas masing-masing. Apabila belum memenuhi optimal hasil proses pengukuran *euclidian distance* dilakukan kembali.

3.4.2 Hasil Clustering

Hasil proses segmentasi dengan metode algoritma K-Means clustering ini akan diperoleh hasil berupa pengelompokan mahasiswa menjadi tiga kelompok.

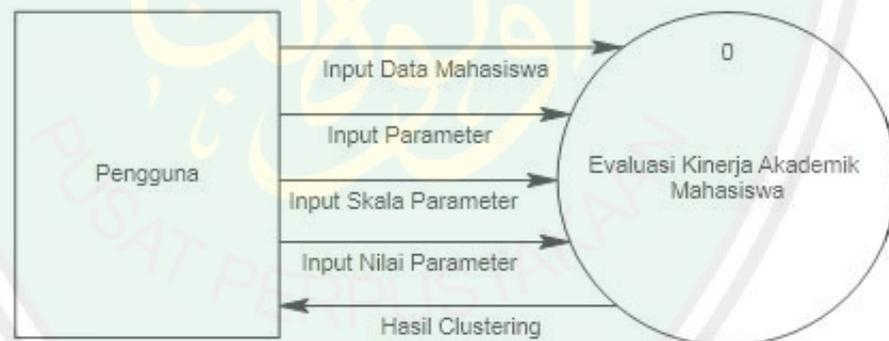
Yaitu kelompok mahasiswa dengan predikat Dengan Pujian, Sangat Memuaskan dan Memuaskan.

3.4.3 Data Flow Diagram Sistem

Secara umum, proses yang dilakukan sistem untuk mengevaluasi kinerja akademik mahasiswa menggunakan algoritma K-Means Clustering adalah seperti digambarkan dalam diagram-diagram berikut:

a. Context Diagram System

Context Diagram merupakan pendekatan terstruktur yang mencoba untuk menggambarkan sistem pertama kali secara garis besar (disebut *top level*). Context Diagram dibawah ini menggambarkan hubungan *input/output* antara sistem dengan pengguna.



Gambar 3.3 Context Diagram System

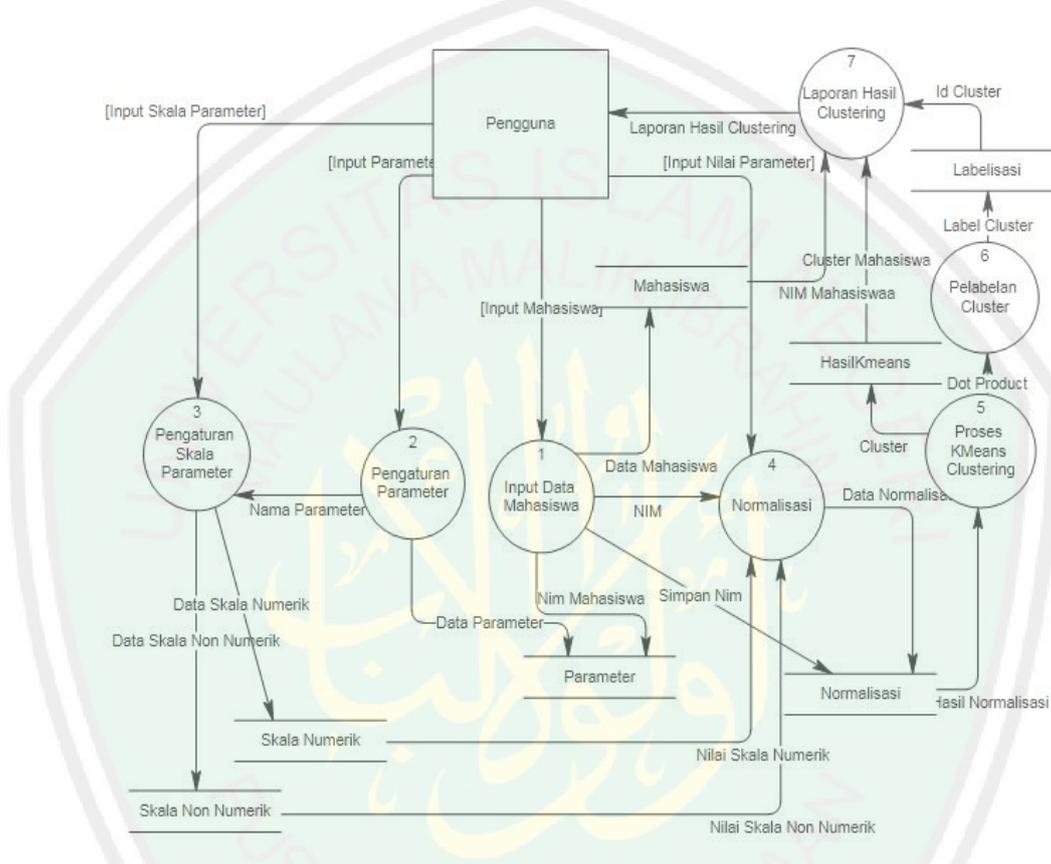
b. Data Flow Diagram level 0

Diagram ini menggambarkan proses-proses dalam sistem lebih rinci. Ada lima proses yang dilakukan dalam sistem yaitu :

- 1) Input data mahasiswa, pengguna memasukkan data-data mahasiswa yang berupa data identitas mahasiswa antara lain : nim, nama, jenis kelamin, tempat dan tanggal lahir, alamat, jurusan dan ipk lulus.
- 2) Pengaturan parameter, pengguna memasukkan parameter mahasiswa berupa ipk selama empat semester awal, nilai ujian akhir SMA , asal sekolah, dan pernah di pesantren.
- 3) Pengaturan skala parameter, pengguna memasukkan skala untuk parameter bertipe *numeric*, dan konversi nilai untuk parameter bertipe *non numeric*.
- 4) Normalisasi data atribut mahasiswa, pengguna melakukan input nilai parameter kemudian system mengambil data skala parameter numeric dan non numeric untuk selanjutnya dilakukan proses normalisasi skala nilai untuk data bertipe *numeric* dan konversi nilai untuk data bertipe *non numeric*.
- 5) Proses Algoritma K-Means, proses cluster data atribut mahasiswa menggunakan metode K-Means yang dilakukan oleh aplikasi.
- 6) Pelabelan cluster dihitung dengan cara menghitung dot product (perkalian centroid dari cluster iterasi terakhir dikalikan dengan transposnya) dan selanjutnya dilakukan pengurutan, nilai terbesar dari hasil perkalian ini berarti diberi label Dengan Pujian, nilai tertinggi kedua diberi label Sangat Memuaskan dan nilai tertinggi ketiga diberi label Memuaskan.

- 7) Laporan Hasil Evaluasi Akademik, mencetak laporan hasil pengelompokan untuk mengevaluasi kinerja akademik mahasiswa.

Gambar 3.4 merupakan *Data Flow Diagram* level 0 dari sistem

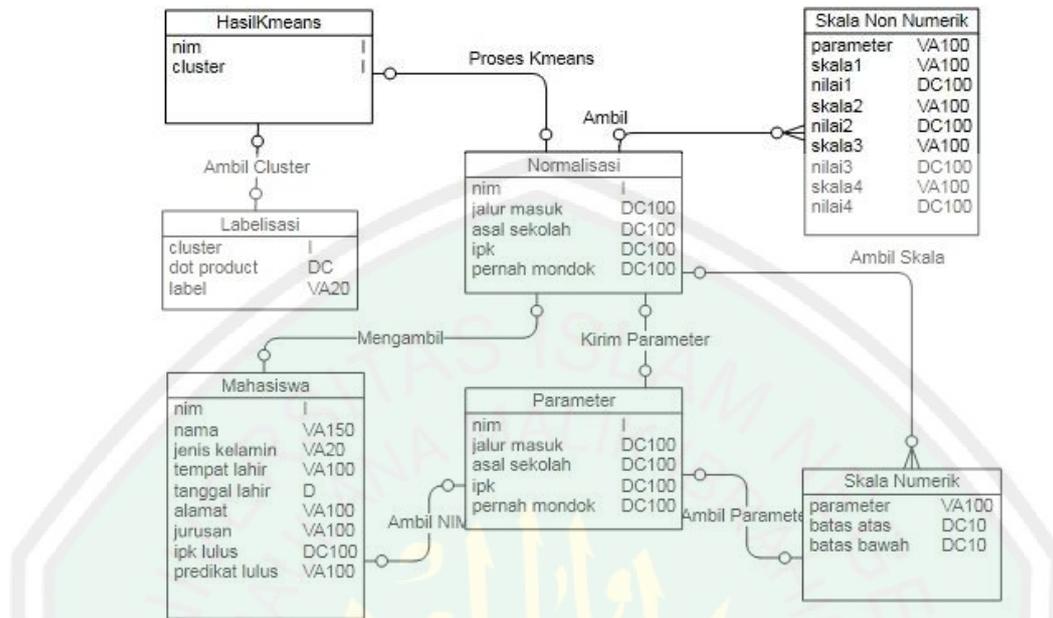


Gambar 3.4 Data Flow Diagram level 0

3.4.4 Desain Tabel

Sistem ini membutuhkan tabel sebagai tempat menyimpan data hasil klasterisasi dan atribut mahasiswa yang dimasukkan oleh pengguna.

Gambar 3.5 merupakan desain relasi antar tabel dari perangkat lunak.



Gambar 3.5 Desain Relasi Antar Tabel

Database yang digunakan dalam aplikasi evaluasi kinerja akademik ini terdiri dari beberapa tabel. Berikut ini adalah tabel yang digunakan yaitu :

Tabel 3.16 Daftar Tabel pada Database

No	Nama Tabel	Keterangan
1	Tabel Mahasiswa	Untuk menyimpan data keterangan mahasiswa
2	Tabel Parameter	Untuk menyimpan kolom parameter
3	Tabel Normalisasi	Untuk menyimpan data normalisasi
4	Tabel Labelisasi	Untuk pelabelan klaster
5	Tabel HasilKmeans	Untuk menyimpan hasil klaster
6	Tabel Skala Numerik	Untuk menyimpan nilai skala parameter <i>numeric</i>
7	Skala Non Numerik	Untuk menyimpan skala parameter <i>non numerik</i> .

Untuk melihat secara jelas tentang *field-field* yang ada pada masing-masing tabel dapat dilihat pada penjelasan tabel-tabel dibawah ini:

a. *Tabel Mahasiswa*

Tabel ini berfungsi sebagai penyimpanan data identitas mahasiswa yang diinputkan oleh user. Data yang disimpan antara lain : Nim, Nama, Tempat Lahir, Tanggal Lahir dan Jurusan.

Tabel 3.17 Struktur Tabel Mahasiswa

Nama Kolom	Type Data	Fungsi
Nim	Integer(11)	Sebagai primary key
Nama	Varchar (150)	Menyimpan data nama Mahasiswa
Jenis Kelamin	Varchar (50)	Menyimpan data jenis kelamin
Tempat_Lahir	Varchar (100)	Menyimpan data tempat lahir
Tanggal_lahir	Date	Menyimpan data tanggal lahir
Alamat	Varchar(150)	Menyimpan data alamat Mahasiswa
Jurusan	Varchar(100)	Menyimpan data jurusan Mahasiswa
Ipk Lulus	Double	Menyimpan Nilai IPK lulus
Predikat	Varchar (50)	Menyimpan data predikat lulus

b. *Tabel Parameter*

Tabel ini berfungsi untuk menyimpan data parameter, tabel parameter ini bersifat dinamis, sehingga pengguna dapat menambah atau menghapus parameter, disesuaikan dengan parameter yang dibutuhkan oleh aplikasi.

Tabel 3.18 Struktur Tabel Parameter

Nama Kolom	Type Data	Fungsi
Nim	Integer	Sebagai primary key
Jalur Masuk	Double	Menyimpan nilai jalur masuk
Asal Sekolah	Double	Menyimpan nilai asal sekolah
Ipk	Double	Menyimpan nilai ipk 4 semester awal
Pernah Mondok	Double	Menyimpan nilai pernah mondok

c. *Tabel Normalisasi*

Tabel ini berfungsi untuk menyimpan data atribut mahasiswa yang sudah dinormalisasi untuk pemrosesan klaster agar mendapatkan hasil yang optimal. Tabel normalisasi ini juga bersifat dinamis.

Tabel 3.19 Struktur Tabel Normalisasi

Nama Kolom	Type Data	Fungsi
Nim	Integer(11)	Sebagai primary key
Jalur Masuk	Double	Menyimpan nilai jalur masuk yang sudah dinormalisasi
Asal Sekolah	Double	Menyimpan nilai asal sekolah yang sudah dinormalisasi
Ipk	Double	Menyimpan nilai ipk 4 semester awal yang sudah dinormalisasi
Pernah	Double	Menyimpan nilai lama mondok yang

Mondok		sudah dinormalisasi
--------	--	---------------------

d. *Tabel Label_Cluster*

Tabel ini berfungsi untuk menyimpan identitas dan pelabelan klaster. Label cluster ini didapat setelah dilakukan perhitungan K-Means kemudian dihitung *dot product* dari cluster akhir dan hasilnya kemudian diurutkan dan diberi label.

Tabel 3.20 Struktur Tabel Label Cluster

Nama Kolom	Type Data	Fungsi
Id_Cluster	Integer(11)	Sebagai primary key pada table klaster
Dot product	Double	Menyimpan hasil dot product dari pusat cluster akhir
Label	Varchar(100)	Pemberian nama label cluster

e. *Tabel Hasil_Cluster*

Tabel ini berfungsi untuk menyimpan hasil proses cluster K-Means Clustering untuk mengevaluasi kinerja akademik Mahasiswa. Berikut deskripsi tabel hasil klaster :

Tabel 3.21 Struktur Tabel Hasil Cluster

Kolom	Type Data	Fungsi
Nim	Integer(11)	Sebagai primary key tabel hasil klaster

Cluster	Integer(11)	Menyimpan identitas klaster
---------	-------------	-----------------------------

3.5 Desain User Interface

Dalam pembuatan aplikasi evaluasi kinerja akademik ini, diperlukan desain layout atau antarmuka sebagai gambaran terhadap sistem yang dibuat. Pembuatan *user interface* memiliki beberapa *form* yang saling berhubungan antar form satu dengan lainnya. Berikut akan diuraikan secara lebih jelas pada masing-masing form.

3.5.1 Desain Form Halaman Utama

Halaman menu utama pada aplikasi ini merupakan pilihan bagi pengguna untuk memilih aktivitas yang ingin digunakan. Desain *form* menu utama :

Pada *form* menu utama seperti pada gambar 3.6, terdapat menu bar yang terdiri dari 4 menu:

- Menu “Aplikasi”, pada menu ini terdapat submenu “Keluar” digunakan untuk keluar dari sistem.
- Menu “Master”, pada menu ini terdapat submenu “Mahasiswa” yang berfungsi untuk memasukkan data identitas mahasiswa, submenu “Parameter” yang berfungsi untuk pengaturan parameter, parameter ini bersifat dinamis, kemudian submenu “Penilaian” yang berfungsi untuk pemberian nilai parameter, submenu “Skala Nilai” yang berfungsi untuk pengaturan skala nilai parameter untuk proses normalisasi, terdiri atas sub submenu Skala Nilai Numerik dan Skala Nilai Non Numerik.

- Menu “Clustering”, pada menu ini terdapat sub menu “Proses Cluster” yang berfungsi untuk memanggil class utama dari aplikasi. Form Proses Cluster ini menampilkan informasi hasil cluster yang ditampilkan dalam bentuk grafik dan tabel.
- Menu “Laporan”, pada menu ini terdapat sub menu laporan “Data Mahasiswa” yang berfungsi untuk mencetak laporan data mahasiswa, sub menu “Penilaian” yang berfungsi untuk mencetak data penilaian parameter mahasiswa dan sub menu laporan “Hasil Clustering Akademik” yang berfungsi untuk mencetak laporan hasil clustering akademik.
- Menu “Bantuan”, pada menu ini terdapat sub menu “Pedoman Penggunaan” yang berfungsi untuk menjelaskan penggunaan aplikasi dan sub menu “Tentang Pembuat” yang berfungsi untuk mengetahui informasi pembuat aplikasi.



Gambar 3.6 Halaman Utama Aplikasi

3.5.2 Desain Form Data Mahasiswa

Halaman ini digunakan untuk mengolah data identitas mahasiswa. Seperti menambah, mengedit, dan menghapus data identitas mahasiswa. Pada form ini terdapat empat fungsi tombol utama seperti pada gambar 3.7. Masing-masing tombol mempunyai fungsi sebagai berikut :

- Tombol “Edit” digunakan untuk melakukan perubahan data yang ada di database jika terjadi kesalahan atau ingin diperbaharui.
- Tombol “Hapus” digunakan untuk melakukan penghapusan data identitas mahasiswa dalam database.
- Tombol “Simpan” digunakan untuk menyimpan data baru identitas mahasiswa ke database.
- Tombol “Reset” digunakan untuk membersihkan *field – field* yang ada.

The screenshot shows a web application window titled "Form Mahasiswa" with a red header bar containing the text "DATA MAHASISWA". Below the header is a form titled "Input Data Mahasiswa" with several input fields: NIM (6550032), Nama (Wafa), Jenis Kelamin (radio buttons for Laki-Laki and Perempuan), Tempat Tgl Lahir (Nganjuk, 2012-11-07), Alamat (Nganjuk), and Jurusan (Teknik Informatika). Below the form are four buttons: "Tambah", "Simpan", "Edit", and "Hapus". At the bottom is a table with columns: nim, nama, alamat, tempat_jahir, tanggal_jahir, and jurusan. The table contains five rows of student data.

nim	nama	alamat	tempat_jahir	tanggal_jahir	jurusan
6550031	Syamsudin	Kediri	Kediri	2012-11-13	Teknik Informati...
6550032	Wafa	Nganjuk	Nganjuk	2012-11-07	Teknik Informati...
6550033	Tegar	Gresik	Gresik	2012-11-14	Teknik Informati...
6550034	Syahri	Mojokerto	Mojokerto	2012-11-12	Teknik Informati...
6550035	Richa	Malang	Malang	2012-11-20	Teknik Informati...

Gambar 3.7 Form Pengisian Data Mahasiswa

3.5.3 Desain Form Data Parameter

Halaman ini digunakan untuk pengaturan parameter yang dibutuhkan pada proses utama. Pengaturan parameter ini bersifat dinamis sehingga pengguna dapat menambah parameter baru atau menghapus parameter yang sudah ada. Untuk menambah parameter baru User mengisi nama parameter dan type data parameter kemudian menekan tombol “Tambah” untuk menyimpan ke tabel parameter pada database. Untuk menghapus parameter yang ada, user mengisi nama parameter yang akan dihapus pada field nama parameter kemudian menekan tombol “Hapus” untuk menghapus parameter pada tabel parameter di database dan aplikasi.

nim	nem	ipk	asal_sekolah	lama_di_pesana...	pendapatan_ortu
6550018	3.0	2	3	1	3
6550021	4.0	3	3	1	3
6550025	3.0	3	3	1	4
6550026	4.0	3	3	1	4
6550031	3.0	3	4	1	2
6550032	3.0	3	4	3	4
6550033	2.0	2	4	1	3

Gambar 3.8 Form Pengisian Data Parameter

3.5.4 Desain Form Penilaian

Halaman ini digunakan untuk pemberian nilai parameter mahasiswa yang telah dibuat pada proses sebelumnya User memilih Nim mahasiswa dan parameter

yang akan diberi nilai. Kemudian mengisi field nilai parameter dan selanjutnya menekan tombol “Simpan” untuk menyimpan ke database.

Pada form ini terdapat tiga fungsi tombol utama seperti pada gambar 3.9. Masing-masing tombol mempunyai fungsi sebagai berikut :

- Tombol “Reset” digunakan untuk membersihkan *field – field* yang ada.
- Tombol “Tambah” digunakan untuk menambah nilai parameter mahasiswa dan menyimpan ke database.
- Tombol “Hapus” digunakan untuk menghapus data nilai parameter mahasiswa dari database.

nim	nem	ipk	asal_sekolah	lama_di_pesantren	pendapatan_ortu
6550018	3.0	2	3	1	3
6550021	4.0	3	3	1	3
6550025	3.0	3	3	1	4
6550026	4.0	3	3	1	4
6550031	3.0	3	4	1	2
6550032	3.0	3	4	3	4
6550033	2.0	2	4	1	3
6550034	3.0	2	3	1	4

Gambar 3.9 Form Pengisian Data Penilaian Parameter Mahasiswa

3.5.5 Desain Form Pengaturan Skala Nilai

Halaman ini digunakan untuk pengaturan skala nilai parameter mahasiswa. Skala nilai ini diperlukan untuk proses normalisasi data parameter.

Pada form ini terdapat tiga fungsi tombol utama seperti pada gambar

3.10. Masing-masing tombol mempunyai fungsi sebagai berikut :

- Tombol “Reset” digunakan untuk membersihkan *field – field* yang ada.
- Tombol “Simpan” digunakan untuk menambah nilai parameter mahasiswa dan menyimpan ke database.
- Tombol “Edit” digunakan untuk mengedit skala nilai parameter mahasiswa dan menyimpan ke database.
- Tombol “Hapus” digunakan untuk menghapus data nilai parameter mahasiswa dari database.

parameter	batas_atas	batas_bawah
ipk	4.0	0.0

Gambar 3.10 Form Pengaturan Skala Nilai Numerik

3.5.6 Desain Form Proses Cluster

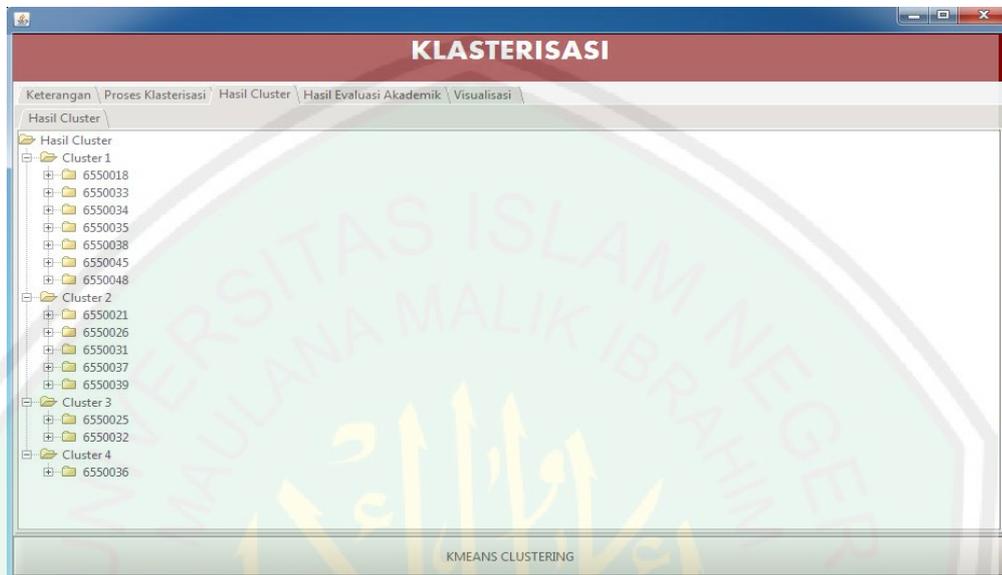
Halaman ini merupakan proses utama pada program aplikasi untuk mengevaluasi kinerja akademik mahasiswa. Form ini berupa tabular yang berisi masing-masing proses dan data. Pada form ini terdapat tab keterangan yang berisi data mahasiswa dan data normalisasi. Tab proses cluster dimana terdapat informasi perhitungan kluster dengan menggunakan K-Means. Tab hasil kluster yang digunakan untuk menampilkan hasil klusterisasi yang belum dilabelisasi yang ditampilkan dalam bentuk tree. Tab Hasil Evaluasi Akademik yang berisi informasi klusterisasi mahasiswa yang sudah dilabelisasi. dan Tab visualisasi yang berisi informasi hasil klusterisasi yang ditampilkan dalam *grid table* dan *grafik pie*. Form klusterisasi dapat dilihat sebagai berikut :

Pada gambar 3.11 adalah tab pertama digunakan untuk menampilkan keterangan Data Mahasiswa dan Data yang sudah dinormalisasi yang telah di masukkan kedalam aplikasi.

nim	nama	alamat	tempat_lahir	tanggal_lahir	jurusan
6550018	Juna	Gayo	Gayo	2012-11-22	Teknik Informatika
6550021	Dily	Malang	Malang	2012-11-02	Teknik Informatika
6550025	Jumhur	Batu	Batu	2012-11-16	Teknik Informatika
6550026	Roni	Medan	Medan	2012-11-05	Teknik Informatika
6550031	Syamsudin	Kediri	Kediri	2012-11-13	Teknik Informatika
6550032	Wafa	Nganjuk	Nganjuk	2012-11-07	Teknik Informatika
6550033	Tegar	Gresik	Gresik	2012-11-14	Teknik Informatika
6550034	Syahri	Mojokerto	Mojokerto	2012-11-12	Teknik Informatika
6550035	Richa	Malang	Malang	2012-11-20	Teknik Informatika
6550036	Arfi	Madiun	Madiun	2012-11-21	Teknik Informatika
6550037	Hari	Bondowoso	Bondowoso	2012-11-28	Teknik Informatika
6550038	Jamil	Bawean	Bawean	2012-11-09	Teknik Informatika
6550039	Lutfi	Lombok	Lombok	2012-11-11	Teknik Informatika
6550045	Eko	Samarinda	Samarinda	2012-11-16	Teknik Informatika
6550048	Fauzi	Pamekasan	Pamekasan	2012-11-22	Teknik Informatika

Gambar 3.11 Form Klusterisasi Tab Keterangan Mahasiswa

Pada gambar 3.12 adalah tab Hasil Cluster digunakan untuk melihat informasi hasil klasterisasi yang belum diberi label.



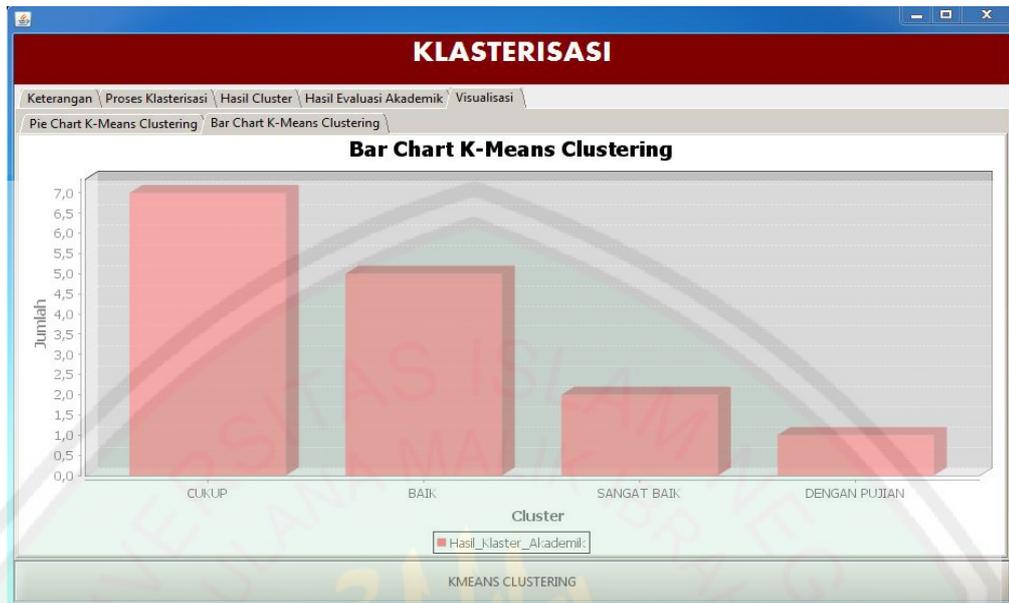
Gambar 3.12 Form Proses Cluster Tab Hasil Cluster

Pada gambar 3.13 adalah Tab Hasil Evaluasi merupakan tab untuk menampilkan informasi hasil klasterisasi yang sudah diberi label.

nim	nama	cluster	label
6550018	Juna	0	CUKUP
6550021	Dily	1	BAIK
6550025	Jumhur	2	SANGAT BAIK
6550026	Roni	1	BAIK
6550031	Syamsudin	1	BAIK
6550032	Wafa	2	SANGAT BAIK
6550033	Tegar	0	CUKUP
6550034	Syahri	0	CUKUP
6550035	Richa	0	CUKUP
6550036	Arri	3	DENGAN PUJIAN
6550037	Hari	1	BAIK
6550038	Jamil	0	CUKUP
6550039	Lutfi	1	BAIK
6550045	Eko	0	CUKUP
6550048	Fauzi	0	CUKUP

Gambar 3.13 Form Proses Cluster Tab Hasil Evaluasi Akademik

Pada gambar 3.14 adalah Tab Visualisasi berfungsi untuk menampilkan informasi hasil klasterisasi yang ditampilkan dalam *grid tabel* dan *grafik pie*.



Gambar 3.14 Form Proses Cluster Tab Visualisasi

BAB IV

IMPLEMENTASI DAN PENGUJIAN

Pada bab ini akan dibahas tentang implementasi dan pengujian sistem evaluasi kinerja akademik mahasiswa menggunakan algoritma K-Means Clustering. Sistem menyediakan form input/masukan. Melalui media ini, pengguna memasukkan perintah dan data. Di samping form input sistem juga menyediakan form keluaran yang merupakan respon dari sistem.

Dengan adanya program evaluasi kinerja akademik mahasiswa ini dapat memberikan informasi bagi dosen dan mahasiswa untuk mengevaluasi kinerja akademik mahasiswa, sehingga dapat menjadi bahan evaluasi dini bagi tiap mahasiswa untuk memperlancar studinya ke depan.

4.1 Implementasi Sistem

Tahap implementasi pengembangan perangkat lunak merupakan proses perubahan spesifikasi sistem menjadi sistem yang dapat dijalankan. Tahap ini merupakan lanjutan dari proses perancangan, yaitu proses pemrograman perangkat lunak sesuai dengan spesifikasi dan desain sistem.

Aplikasi evaluasi kinerja akademik mahasiswa ini menggunakan basis data MySQL yang berfungsi sebagai media penyimpanan data atau informasi yang terkumpul, yang terdiri dari beberapa tabel. Sedangkan untuk menjembatani antara informasi yang akan dibuat dengan basis data yang ada, digunakan Eclipse.

Setelah merancang keseluruhan sub-sistem selanjutnya dilakukan pengujian untuk mengetahui kerja dari masing-masing sub-sistem, apakah sudah sesuai dengan yang telah direncanakan. Sebelum pengujian sistem, harus mempersiapkan kebutuhan dari program yang akan diimplementasikan baik dari segi perangkat keras maupun perangkat lunak komputer.

4.2 Kebutuhan Sistem

Dalam proses penyelesaian skripsi ini diperlukan spesifikasi perangkat lunak dan keras yang mampu mendukung dalam pembuatan aplikasi dengan baik. Adapun spesifikasi perangkat keras dan perangkat lunak yang digunakan dalam pembuatan aplikasi ini adalah sebagai berikut :

1. Perangkat Lunak

- Windows 7 Ultimate sebagai sistem operasi.
- Java sebagai bahasa pemrograman yang digunakan untuk membangun aplikasi evaluasi kinerja akademik mahasiswa dengan menggunakan algoritma K-Means.
- Database MySql sebagai penyimpanan data hasil evaluasi kinerja akademik mahasiswa yang sudah terkelompok berdasarkan tiap cluster.

2. Perangkat Keras

- Processor intel Core 2 Duo
- Memory 2 GHz
- Harddisk 320 GB

4.3 Implementasi dan Pengujian Sistem

4.3.1 Implementasi Perangkat Lunak

Dalam implementasi algoritma K-Means Clustering pada perangkat lunak ini dibutuhkan class-class untuk menangani tiap – tiap. Tiap-tiap class mempunyai fungsi sendiri dan saling berkaitan.

4.3.1.1 Class – class yang dibutuhkan

1. Class KmeansAkademik

- **Method distributeSamples()**

Method ini digunakan untuk melakukan pengacakan nilai pusat kluster (centroid). Pengacakan ini mengambil data dari atribut mahasiswa di dalam database.

```
private void distributeSamples() {
    int i, pat, clustid, memberIndex;
    for (i=0; i<numClusters; i++) {
        cluster[i].numMembers=0;
    }
    for (pat=0; pat<numPatterns; pat++) {
        clustid=findClosestCluster(pat);
        memberIndex=cluster[clustid].numMembers;
        cluster[clustid].member[memberIndex]=pat;
        cluster[clustid].numMembers++;
        System.out.println("\n");
        System.out.println ("\npattern "+pat+"
ditugaskan di cluster"+clustid+", indeks
anggota:"+memberIndex);
    }
}
```

- **Method initClusters()**

Method ini digunakan sebagai inisialisasi proses klasterisasi pada klasterisasi kmeans. Jumlah klaster, jumlah atribut data dan beberapa variabel yang digunakan dalam proses perhitungan kmeans.

```
public void initClusters(){
    int i,j;
    System.out.println ("\Inisiasi Pusat Cluster:");
    for(i=0;i<numClusters;i++){
        cluster[i].member[0]=i;
        for(j=1;j<sizeVector;j++){
            cluster[i].center[j]=
                Double.parseDouble(pattern[i][j]);
        }
    }
    for(i=0;i<numClusters;i++){
        System.out.println("\n ");
        System.out.println("\nPusat Cluster ["+i+"]=
            ("+cluster[i].center[1]+",
            "+cluster[i].center[2]+",
            "+cluster[i].center[3]+",
            "+cluster[i].center[4]+",
            "+cluster[i].center[5]+")");
        System.out.println("\nIndeks Pusat Cluster :
            "+cluster[i].member[0]);
    }
}
```

- **Method eucNorm()**

Method ini digunakan untuk melakukan proses perhitungan jarak antara data dan titik pusat (centroid cluster).

```

private double eucNorm(int p,int c){
    double dist,x; int i; dist=0;
    for(i=1;i<sizeVector;i++){
        x=(cluster[c].center[i]-
            Double.parseDouble(pattern[p][i]))*
            (cluster[c].center[i]-
            Double.parseDouble(pattern[p][i]));
        dist+=x;}
    return dist;}

```

- **Method CalcNewClustCenters()**

Method ini digunakan untuk melakukan perhitungan pusat klaster(centroid cluster) terbaru pada proses cluster kmeans.

```

private int calcNewClusterCenter(){
    int convFlag,vectId,i,j,k;
    double [] tmp=new double[MAXVECTDIM];
    convFlag=TRUE;
    for(i=0;i<numClusters;i++){
        for (j=0;j<sizeVector;j++){
            tmp[j]=0.0;        }
        for(j=0;j<cluster[i].numMembers;j++){
            vectId=cluster[i].member[j];
            for(k=1;k<sizeVector;k++){
                tmp[k]+=
                    Integer.parseInt(pattern[vectId][k])
            }
        }
        for(k=1;k<sizeVector;k++){
            if(cluster[i].numMembers!=0)
                tmp[k]=tmp[k]/cluster[i].numMembers; else
                tmp[k]=0;
            if(tmp[k]!=cluster[i].center[k])
                convFlag=FALSE;
            cluster[i].center[k]=tmp[k];
        }return convFlag;}

```

- **Method findClosestClusters()**

Method ini digunakan untuk melakukan pengelompokan data kedalam kluster tertentu berdasarkan jarak terdekat antara data dan kluster.

```
private int findClosestCluster(int pat)
{
    int i, clustId;
    double minDist, d;
    minDist=9.9e+99;
    clustId=-1;
    for(i=0; i<numClusters; i++)
    {
        d=eucNorm(pat, i);
        if(d<minDist)
        {
            minDist=d;
            clustId=i;
        }
    }
    if(clustId<0)
    {
        System.out.println("Aaargh");
    }
    return clustId;
}
```

- **Method LoadData()**

Method ini digunakan untuk mengambil data parameter yang dibutuhkan untuk melakukan proses evaluasi kinerja akademik mahasiswa menggunakan algoritma Kmeans.

```

public int loadData() throws SQLException{
    try {
        String queryString=("select * FROM parameter");
        Statement stmt=conn.createStatement();
        rs=stmt.executeQuery(queryString);
        int i=0;
        while(rs.next()){
            for (int m = 0; m < sizeVector; m++) {
                pattern[i][m]=rs.getString(m+1);
            }
            i++;}
            numPatterns=i;}
        catch (Exception sqle){
            System.out.println("Unable to load
driver");}
        return SUCCESS;
    }
}

```

- **Method Labelisasi()**

Method ini digunakan untuk proses labelisasi dari data cluster yang sudah terbentuk

```

public void labelisasi() throws SQLException {
    int i;
    cent_transpos =new Double[numClusters];
    clus = new Integer[numClusters];
    Double temp;
    for( i = 0 ;i<cent_transpos.length;i++){
        temp=(cluster[i].center[1])+
            (cluster[i].center[2]+cluster[i].center[3])
            +(cluster[i].center[4]+cluster[i].center[5]);}
    clus[i]=i;
    cent_transpos[i]=temp;
    sort(clus,cent_transpos);
}

```

- **Method runKmeans()**

Method ini merupakan bagian utama dari proses klusterisasi kmeans, digunakan untuk memanggil method dan function yang lain dalam proses klusterisasi.

```

public int [][] runKmeans() throws
SQLException
{
    int converged,pass;
    if (loadData ()==FAILURE)
    {
        System.out.println("\Gagal!");
    }
    initClusters ();
    pass=1;
    converged=FALSE;
    while (converged==FALSE) {
        System.out.println ("ITERASI
KE="+pass);
        pass++;
        distributeSamples ();
        converged=calcNewClustCenters ();
        showCenters ();
    }
    return clusterSet;
}

```

2. Class Mahasiswa

Class ini digunakan untuk memasukkan data identitas mahasiswa ke dalam database.

3. Class Parameter

Class ini digunakan untuk menambah dan mengurangi parameter yang terdapat pada tabel parameter di dalam database.

4. Class Penilaian

Class ini digunakan untuk memasukkan nilai parameter mahasiswa.

5. Class SkalaNilaiNumeric

Class ini digunakan untuk normalisasi data parameter numerik

6. Class SkalaNilaiNonNumeric

Class ini digunakan untuk normalisasi data parameter non numerik

7. Class Klasterisasi

Class ini digunakan untuk menampilkan proses dan informasi hasil pengelompokan dari algoritma K-Means Clustering.

8. Class HalamanUtama

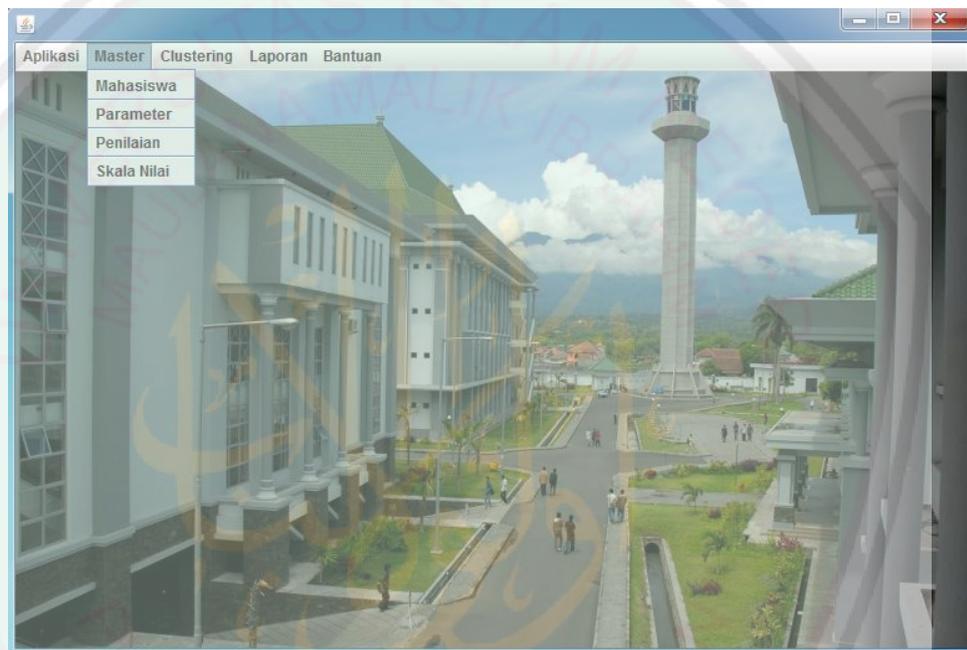
Pada class ini terdapat operasi `getFormMahasiswa()`, `getFormParameter()`, `getFormPenilaian()`, `getLaporanMahasiswa()`, `getLaporanHasilKmeans()`, `getFormProsesKlaster()`, `getFormPenggunaan()`, `getFormKeluar()`. Class ini digunakan untuk memanggil class – class lain yang terdapat pada aplikasi.

4.4 Pengujian Sistem

Pada tahap ini, perancangan perangkat lunak direalisasikan sebagai serangkaian program atau unit program. Pengujian unit melibatkan verifikasi bahwa setiap unit telah memenuhi spesifikasinya.

4.4.1 Pengujian Menu Utama

Menu Utama merupakan tampilan awal saat pertama kali aplikasi dijalankan, Menu utama ini terdiri dari 5 menu yaitu : Menu Aplikasi, Menu Master, Menu Clustering, Menu Laporan dan Menu Bantuan. Tampilan menu utama sebagai berikut :



Gambar 4.1 Pengujian Menu Utama

4.4.2 Pengujian Menu Mahasiswa

Langkah pertama dalam menjalankan program adalah memasukkan data mahasiswa yang berisi identitas mahasiswa berupa NIM, Nama, Jenis Kelamin, Tempat dan Tanggal lahir, Alamat dan Jurusan.

nim	nama	jeniskelamin	alamat	tempat_lahir	tanggal_lahir	jurusan
7650004	YAYUK NOVITA	Perempuan	AMBUNTEN T...	SUMENEP	1989-06-22	TEKNIK INFO...
7650006	NUGROHO A...	Laki-Laki	JL. KARTINI G...	SUMENEP	1989-06-28	TEKNIK INFO...
7650010	DIYU ILHAM A...	Laki-Laki	JL. RAYA PAN...	PAMEKASAN	1989-08-26	TEKNIK INFO...
7650012	ANDI FIRDOUS	Laki-Laki	JL. SERSAN M...	PAMEKASAN	1990-01-16	TEKNIK INFO...
7650014	PURWADIAN ...	Laki-Laki	DS. KEJAPAN...	PASURUAN	1988-10-12	TEKNIK INFO...

Gambar 4.2 Pengujian Form Mahasiswa

4.4.3 Pengujian Menu Parameter

Langkah berikutnya adalah memasukan parameter yang akan dipakai dalam proses clustering. Parameter dalam aplikasi ini dibuat dinamis, parameter dapat ditambah atau dikurangi menyesuaikan parameter yang ada.

nim	jenis_sekolah	lama_mondok	jalur_masuk	pendapatan_or...	ipk
7650004	SMAN	0.0	PMDK	3500000.0	3.19
7650006	SMAN	0.0	PMDK	3500000.0	3.17
7650010	MAN	0.0	PMDK	2500000.0	2.52
7650012	MAN	0.0	PMDK	5500000.0	3.19
7650014	SWASTA	0.0	PMDK	2500000.0	3.0
7650015	MAN	6.0	PMDK	3500000.0	3.32
7650021	SWASTA	3.0	PMDK	2500000.0	2.26

Gambar 4.3 Pengujian Form Parameter

4.4.4 Pengujian Sub Menu Skala Nilai Numerik

Setelah parameter dimasukkan, perlu dilakukan pengaturan skala nilai untuk proses normalisasi data parameter.

parameter	batas_atas	batas_bawah
ipk	4.0	0.0

Gambar 4.4 Pengujian Form Pengaturan Skala Numerik

4.4.5 Pengujian Sub Menu Skala Nilai Non Numerik

Pengaturan skala parameter ini dibagi menjadi dua, yakni pengaturan skala parameter numeric dan pengaturan skala parameter non numerik. Pengaturan skala parameter ini sangat berpengaruh pada kinerja aplikasi.

parameter	skala1	nilai1	skala2	nilai2	skala3	nilai3	skala4	nilai4
asal_sekolah	SMAN	4.0	MAN	3.0	SMKN	3.0	MA	2.0

Gambar 4.5 Pengujian Form Pengaturan Skala Non Numerik

4.4.6 Pengujian Menu Penilaian

Setelah pengaturan skala selesai, selanjutnya dilakukan input penilaian parameter.

nim	jenis_sekolah	lama_mondok	jalur_masuk	pendapatan_orangtua	ipk
7650004	SMAN	0.0	PMDK	3500000.0	3.19
7650006	SMAN	0.0	PMDK	3500000.0	3.17
7650010	MAN	0.0	PMDK	2500000.0	2.52
7650012	MAN	0.0	PMDK	5500000.0	3.19
7650014	SWASTA	0.0	PMDK	2500000.0	3.0
7650015	MAN	6.0	PMDK	3500000.0	3.32
7650021	SWASTA	3.0	PMDK	2500000.0	2.26
7650022	SMAN	1.0	PMDK	3500000.0	2.91
7650025	SWASTA	0.0	PMDK	1000000.0	3.14
7650026	MAN	0.0	PMDK	5500000.0	3.16
7650030	SWASTA	0.0	PMDK	2500000.0	3.19
7650031	SMAN	0.0	PMDK	2500000.0	3.6

Gambar 4.6 Pengujian Form Penilaian Parameter

4.4.7 Pengujian Menu Proses K-Means Clustering

Setelah data mahasiswa, parameter, nilai skala parameter dan penilaian parameter selesai dilakukan, selanjutnya masuk ke proses menjalankan algoritma K-means Clustering. Proses ini terdapat pada menu Proses Cluster. Pada menu ini terdapat beberapa tab yang berisi keterangan sebagai berikut :

- a. Tab Keterangan, Berisi keterangan data mahasiswa dan data penilaian parameter yang sudah dinormalisasi. Berikut adalah tampilan jendela Klusterisasi, Tab Keterangan :

nim	nama	jeniskelamin	alamat	tempat_lahir	tanggal_lahir	jurusan
7650004	YAYUK NOVITA	Perempuan	AMBUNTEN TENG...	SUMENEP	1989-06-22	TEKNIK INFORMAT...
7650006	NUGROHO AGUNG...	Laki-Laki	JL. KARTINI Gg. IV ...	SUMENEP	1989-06-28	TEKNIK INFORMAT...
7650010	DIYU ILHAM AKBAR	Laki-Laki	JL. RAYA PANDEM...	PAMEKASAN	1989-08-26	TEKNIK INFORMAT...
7650012	ANDI FIRDOUS	Laki-Laki	JL. SERSAN MESRU...	PAMEKASAN	1990-01-16	TEKNIK INFORMAT...
7650014	PURWADIAN SUK...	Laki-Laki	DS. KEJAPANAN N...	PASURUAN	1988-10-12	TEKNIK INFORMAT...
7650015	NURUL MUFlichAH	Perempuan	JL. P. FLORES RT. 07...	BALIKPAPAN	1990-05-15	TEKNIK INFORMAT...
7650021	RIQQOTUL BADRIY...	Perempuan	PAYMAN SOLOKU...	LAMONGAN	1989-07-03	TEKNIK INFORMAT...
7650022	DESI LINAMAWATI	Perempuan	DSN. GEDONGSARI ...	NGANJUK	1988-12-08	TEKNIK INFORMAT...
7650025	MOH. MIFTAKHUR...	Laki-Laki	KRAJAN.A RT.002 R...	JEMBER	1989-01-05	TEKNIK INFORMAT...
7650026	FITRYA NADYA SA...	Perempuan	JL. YOS SUDARSO ...	TRENGGALEK	1989-06-01	TEKNIK INFORMAT...
7650030	TRI MUTAAHIROH	Perempuan	NGESTIBOGA 1 JAY...	LUBUK LINGGAU	1989-10-14	TEKNIK INFORMAT...
7650031	LAILATUS SOFIYAN...	Perempuan	ARCOPODO RT.05 ...	PASURUAN	1989-09-02	TEKNIK INFORMAT...
7650036	IMAM SARONI	Laki-Laki	JL. SUNAN GUNUN...	PONOROGO	1989-08-27	TEKNIK INFORMAT...
7650042	KURNIAWAN FARUK	Laki-Laki	JL. RAYA LETENG N...	SUMENEP	1988-12-19	TEKNIK INFORMAT...
7650047	LULUL MUKHASS...	Perempuan	ROWOGEMPOL NO...	PASURUAN	1989-04-14	TEKNIK INFORMAT...
7650050	CITRA ROSIANA LU...	Perempuan	RT 03 RW 01 SURU...	TRENGGALEK	1988-08-26	TEKNIK INFORMAT...
7650055	FAIZ AFIANI ROHM...	Perempuan	SIDOMULYO 3 NO3...	MOJOKERTO	1989-08-19	TEKNIK INFORMAT...
7650056	ANINCA DUCDA KIR...	Perempuan	JL. ALAKA NO.24 ...	MALANG	1989-01-32	TEKNIK INFORMAT...

Gambar 4.7 Pengujian Form Klasterisasi Tab Keterangan Mahasiswa

nim	jenis_sekolah	jalur_masuk	ipk	lama_mondok
7650004	3.0	2.75	3.0	2.5
7650006	3.0	2.75	3.0	2.5
7650010	2.75	2.5	2.5	2.5
7650012	2.75	2.75	3.0	2.5
7650014	2.5	2.75	3.0	2.5
7650015	2.75	2.75	3.0	2.75
7650021	2.5	2.75	2.0	2.75
7650022	2.5	2.75	3.0	2.75
7650025	2.5	2.75	3.0	2.5
7650026	2.75	2.75	3.0	2.5
7650030	2.5	2.75	3.0	2.5
7650031	3.0	2.75	3.5	2.5
7650036	2.5	2.75	2.5	3.0
7650042	3.0	2.75	2.5	2.5
7650047	2.5	2.75	3.0	3.0
7650050	3.0	3.0	3.0	2.5
7650055	2.75	3.0	3.5	2.5
7650056	2.0	2.0	2.5	2.5

Gambar 4.8 Pengujian Form Proses Cluster Tab Keterangan

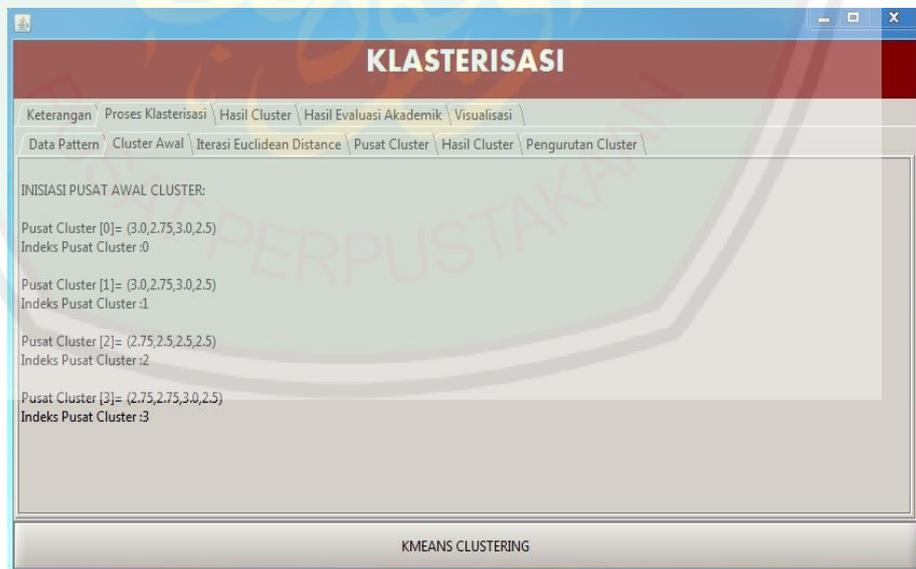
Normalisasi Parameter

- b. Tab Proses Cluster, berisi proses yang dijalankan oleh algoritma K-Means Clustering setelah menekan tombol proses k-means pada jendela Proses Cluster. Proses cluster ini terdiri atas tampilan data mahasiswa, inialisasi cluster awal, proses perhitungan jarak anggota

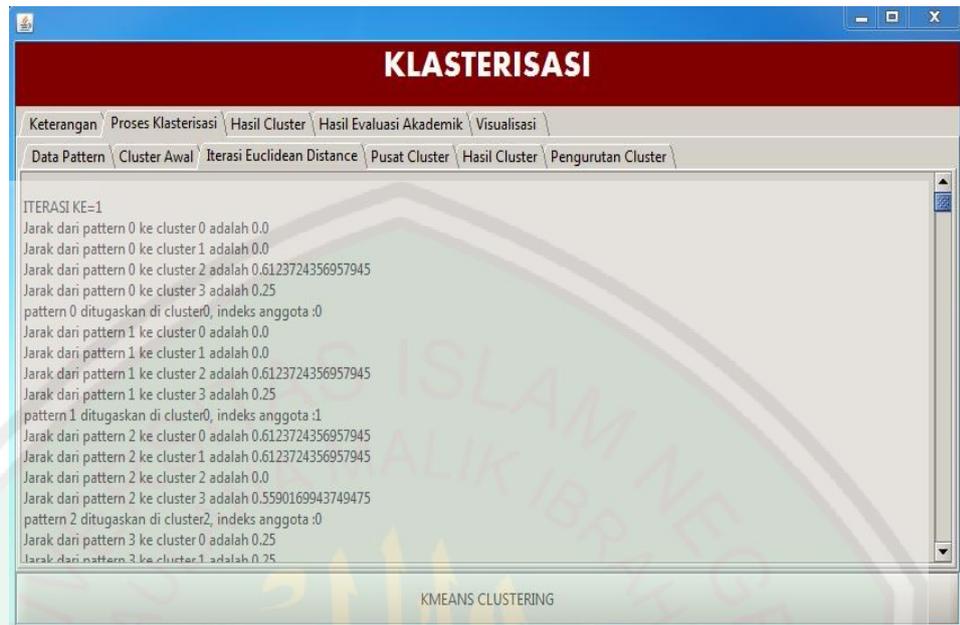
ke pusat cluster menggunakan euclidean distance, pusat cluster, hasil akhir cluster dan pelabelan cluster.



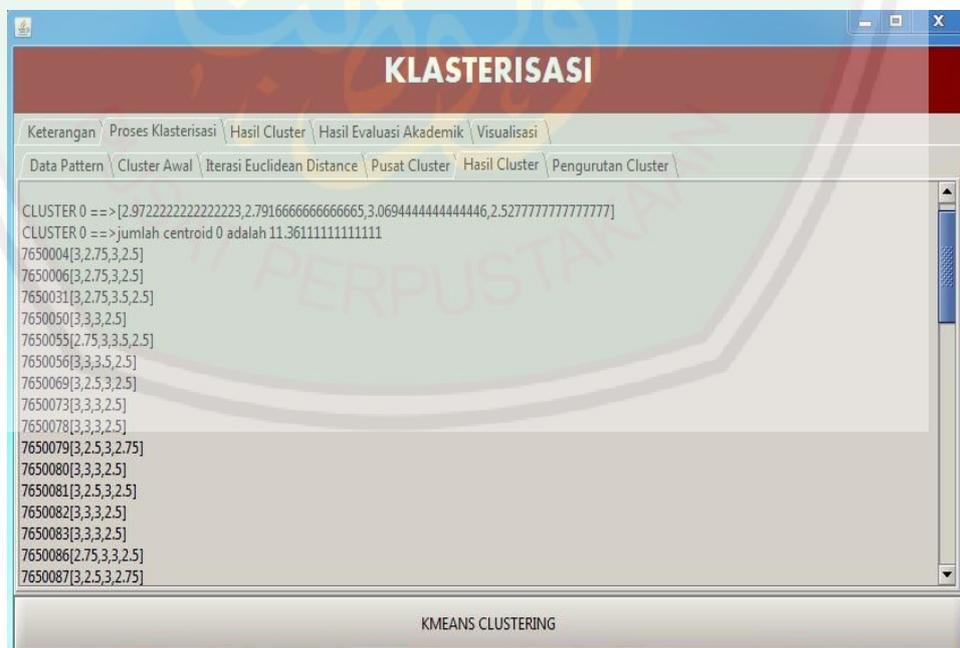
Gambar 4.9 Pengujian Form Proses Cluster Tab Proses Cluster –
Inisialisasi Cluster Awal



Gambar 4.10 Pengujian Form Proses Cluster Tab Proses Cluster –
Inisialisasi Cluster Awal



Gambar 4.11 Tab Proses Cluster – Iterasi Perhitungan Euclidean Distance



Gambar 4.12 Pengujian Form Proses Cluster Tab Proses Cluster – Hasil Cluster

Cluster	Dot Product	Label
1	0.0	CUKUP
2	27.2013888888889	BAIK
3	29.2446875	SANGAT BAIK
0	32.438657407405	DENGAN PUJIAN

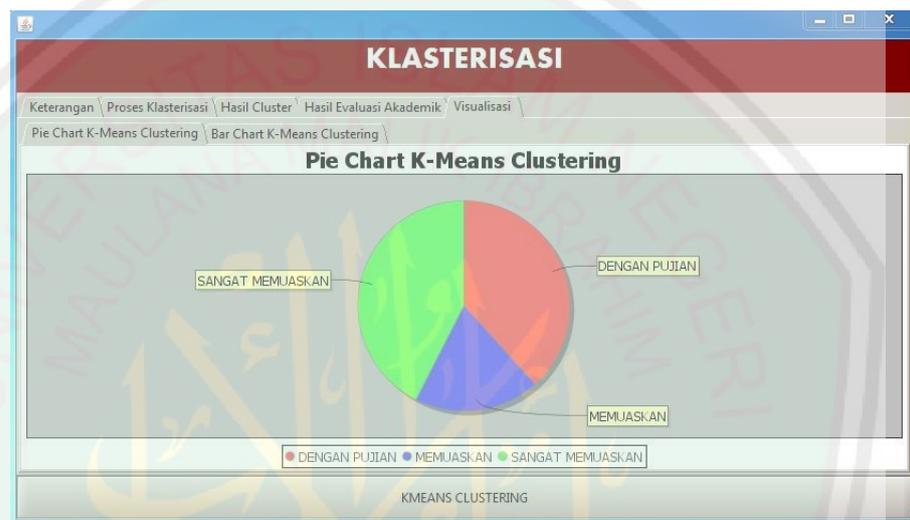
Gambar 4.13 Pengujian Form Proses Cluster Tab Proses Cluster –
Pelabelan Cluster

- c. Tab Hasil Evaluasi Akademik, pada bagian ini berisi laporan hasil akhir dari proses cluster program evaluasi kinerja akademik mahasiswa.

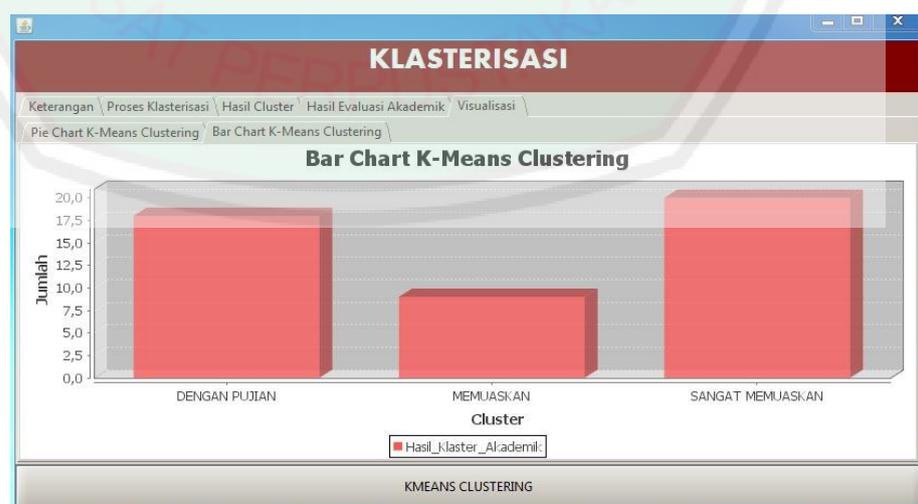
nim	nama	cluster	label
7650004	YAYUK NOVITA	0	DENGAN PUJIAN
7650006	NUGROHO AGUNG SUPRIYANTO	0	DENGAN PUJIAN
7650010	DIYU ILHAM AKBAR	2	MEMUASKAN
7650012	ANDI FIRDOUS	3	SANGAT MEMUASKAN
7650014	PURWADIAN SUKMANA	3	SANGAT MEMUASKAN
7650015	NURUL MUFLICHAH	3	SANGAT MEMUASKAN
7650021	RIQQOTUL BADRIYAH	2	MEMUASKAN
7650022	DESI LINAMAWATI	3	SANGAT MEMUASKAN
7650025	MOH. MIFTAKHUR ROKMAN	3	SANGAT MEMUASKAN
7650026	FITRYA NADYA SABYLA	3	SANGAT MEMUASKAN
7650030	TRI MUTAAHIROH	3	SANGAT MEMUASKAN
7650031	LAILATUS SOFIYANA	0	DENGAN PUJIAN
7650036	IMAM SARONI	2	MEMUASKAN
7650042	KURNIAWAN FARUK	2	MEMUASKAN
7650047	LULUUL MUKHASSONAH	3	SANGAT MEMUASKAN
7650050	CITRA ROSIANA LUTFI	0	DENGAN PUJIAN
7650055	FAIZ AFJANI ROHMA	0	DENGAN PUJIAN
7650056	AMIRCA RUSDA KIRAMA	0	DENGAN PUJIAN

Gambar 4.14 Pengujian Form Proses Cluster Tab Hasil Evaluasi
Akademik

- d. Tab Visualisasi, berisi visualiasi hasil akhir proses penghitungan dari program Evaluasi Kinerja Akademik Mahasiswa Menggunakan algoritma K-Means Clustering yang digambarkan dalam bentuk *pie chart diagram* dan *bar chart diagram*.



Gambar 4.15 Pengujian Form Proses Cluster Tab Visualisasi – *Pie Chart Diagram*



Gambar 4.15 Pengujian Form Proses Cluster Tab Visualisasi – *Bar Chart Diagram*

4.5 Pengujian Dan Analisis Program

Analisis program dilakukan dengan membandingkan data sebenarnya dengan hasil keluaran dari aplikasi ini. Pengujian dan analisa program ini juga dilihat dari kesamaan dan tingkat error dari aplikasi ini. Berikut adalah pengujian dengan mengambil contoh sebanyak 47 data :

Berikut adalah tabel pembanding antara hasil dari aplikasi dengan data sebenarnya.

Tabel 4.1 Tabel Pembanding Hasil Aplikasi dengan data Sebenarnya

No	NIM	Masuk Program		Keterangan
		Hasil Program	Data Sebenarnya	
1	07650004	DENGAN PUJIAN	3.28	Sudah lulus
2	07650006	DENGAN PUJIAN	3.58	Sudah lulus
3	07650010	SANGAT MEMUASKAN	3.10	Sudah lulus
4	07650012	SANGAT MEMUASKAN	3.31	Sudah lulus
5	07650014	SANGAT MEMUASKAN	3.09	Sudah lulus
6	07650015	SANGAT MEMUASKAN	3.33	Sudah lulus
7	07650021	SANGAT MEMUASKAN	2.99	Sudah lulus
8	07650022	DENGAN PUJIAN	3.15	Sudah lulus
9	07650025	SANGAT MEMUASKAN	3.29	Sudah lulus
10	07650026	SANGAT MEMUASKAN	3.26	Sudah lulus
11	07650030	SANGAT MEMUASKAN	3.25	Sudah lulus
12	07650031	DENGAN PUJIAN	3.45	Sudah lulus
13	07650036	SANGAT MEMUASKAN	3.01	Sudah lulus
14	07650042	DENGAN PUJIAN	3.27	Sudah lulus
15	07650047	SANGAT MEMUASKAN	3.27	Sudah lulus

16	07650050	DENGAN PUJIAN	3.58	Sudah lulus
17	07650055	DENGAN PUJIAN	3.78	Sudah lulus
18	07650056	DENGAN PUJIAN	3.66	Sudah lulus
19	07650064	DENGAN PUJIAN	3.17	Sudah lulus
20	07650069	DENGAN PUJIAN	3.33	Sudah lulus
21	07650073	DENGAN PUJIAN	3.20	Sudah lulus
22	07650078	DENGAN PUJIAN	3.52	Sudah lulus
23	07650079	DENGAN PUJIAN	3.24	Sudah lulus
24	07650080	DENGAN PUJIAN	3.35	Sudah lulus
25	07650081	DENGAN PUJIAN	3.38	Sudah lulus
26	07650082	DENGAN PUJIAN	3.45	Sudah lulus
27	07650083	DENGAN PUJIAN	3.11	Sudah lulus
28	07650086	SANGAT MEMUASKAN	3.17	Sudah lulus
29	07650087	SANGAT MEMUASKAN	3.30	Sudah lulus
30	07650092	SANGAT MEMUASKAN	3.44	Sudah lulus
31	07650102	SANGAT MEMUASKAN	3.21	Sudah lulus
32	07650107	SANGAT MEMUASKAN	3.52	Sudah lulus
33	07650109	SANGAT MEMUASKAN	3.32	Sudah lulus
34	07650114	SANGAT MEMUASKAN	3.00	Sudah lulus
35	07650117	DENGAN PUJIAN	3.19	Sudah lulus
36	07650120	DENGAN PUJIAN	3.26	Sudah lulus
37	07650121	DENGAN PUJIAN	3.38	Sudah lulus
38	07650124	SANGAT MEMUASKAN	3.19	Sudah lulus
39	07650129	SANGAT MEMUASKAN	3.46	Sudah lulus
40	07650130	SANGAT MEMUASKAN	3.46	Sudah lulus
41	07650133	SANGAT MEMUASKAN	3.32	Sudah lulus
42	07650136	DENGAN PUJIAN	3.43	Sudah lulus
43	07650142	DENGAN PUJIAN	3.17	Sudah lulus
44	07650147	SANGAT MEMUASKAN	3.03	Sudah lulus
45	07650153	SANGAT MEMUASKAN	3.10	Sudah lulus

46	07650154	DENGAN PUJIAN	3.10	Sudah lulus
47	07650155	DENGAN PUJIAN	3.10	Sudah lulus

Dari tabel tersebut dapat diambil kesimpulan, bahwa dari 47 data, ada 27 data yang memiliki kesamaan, dan 20 data yang tidak memiliki kesamaan. Sehingga prosentase tingkat error adalah $20/47 \times 100\% = 44,68\%$ dan Prosentase tingkat keberhasilan $27/47 \times 100\% = 57,44\%$.

4.6 Validasi Cluster

Validasi termasuk usaha yang dilakukan oleh peneliti untuk menjamin bahwa hasil cluster adalah representatif terhadap populasi secara umum, dan dengan demikian dapat digeneralisasi untuk obyek yang lain dan stabil untuk waktu tertentu.

Confusion Matrix

Confusion matrix merupakan sebuah tabel yang terdiri atas banyaknya tabel data uji yang diprediksi benar dan tidak benar oleh model klasifikasi, tabel ini diperlukan untuk menentukan kinerja suatu model klasifikasi.

Confusion Matrix menggambarkan prediksi yang benar dan salah dari sebuah klasifikasi.

Tabel 4.2 Tabel Rumus Confusion Matrix

		Predicted Class	
		+	-
Actual Class	+	$\int_{++}(\text{TP})$	$\int_{+-}(\text{FN})$
	-	$\int_{-+}(\text{FP})$	$\int_{--}(\text{TP})$

- True positive (TP), jika data sebenarnya positif sesuai dengan hasil prediksi.
- False negative (FN), jika data sebenarnya positif salah diprediksi negatif oleh hasil klasifikasi.
- False positive (FP), jika data sebenarnya negatif salah diprediksi positif oleh hasil klasifikasi.
- True negative (TN), jika data sebenarnya negatif diprediksi salah oleh hasil klasifikasi.

Tabel 4.3 Tabel Confusion Matrix

Hasil Cluster / Data Sebenarnya	MEMUASKAN	SANGAT MEMUASKAN	DENGAN PUJIAN
MEMUASKAN	0	0	0
SANGAT MEMUASKAN	0	22	19
DENGAN PUJIAN	0	1	5

Tabel confusion matrix merupakan tabel pengujian antara data sebenarnya dengan hasil cluster. Nilai 0 pada kolom MEMUASKAN dan Baris SANGAT MEMUASKAN menunjukkan bahwa ada 0 data yang sebenarnya yang masuk predikat SANGAT MEMUASKAN dengan hasil cluster MEMUASKAN. Nilai 22 pada baris SANGAT MEMUASKAN kolom SANGAT MEMUASKAN menunjukkan bahwa ada 22 data yang sama dengan data yang sebenarnya yang masuk predikat SANGAT MEMUASKAN dengan hasil cluster SANGAT

MEMUASKAN. Nilai 5 pada baris DENGAN PUJIAN kolom DENGAN PUJIAN menunjukkan bahwa ada 5 data sebenarnya yang masuk predikat DENGAN PUJIAN dengan hasil cluster DENGAN PUJIAN.

4.7 Precision dan Recall

Precision dan recall adalah dua pengertian yang banyak digunakan dalam aplikasi pemetaan yang mana dapat berhasil dalam menemukan satu golongan yang mempertimbangkan kepentingan lebih banyak daripada menemukan golongan yang lainnya. Definisi matrik tersebut dapat dinyatakan seperti di bawah ini:

$$\text{Precision, } p = \frac{TP}{TP + FP}$$

$$\text{Recal, } r = \frac{TP}{TP + FN}$$

Precision menentukan pecahan data yang sebenarnya mengeluarkan pernyataan positif di dalam kelompok yang dipisahkan lalu dilaporkan dalam suatu kelas yang positif. Lebih tinggi ketelitian, berarti menurunkan nomor yang tidak benar dan positif error dalam melakukan pengelompokan. Recall menentukan ukuran pecahan positif prediksi dengan tepat suatu pengelompokan. Pengelompokan dengan recall secara luas sangat sedikit kemungkinan yang positif dari kelompok negatifnya. Dalam pernyataan, nilai suatu recall adalah sama dengan rata-rata positif yang benar.

Berikut adalah hasil Precision dan Recall dari pengujian pada tabel

Confusion Hasil Cluster Data Matrix :

➤ **Precision**

$$\text{MEMUASKAN} = \frac{0}{0 + 0 + 0} = 0$$

$$\text{SANGAT MEMUASKAN} = \frac{22}{0 + 22 + 1} = 0.95$$

$$\text{DENGAN PUJIAN} = \frac{5}{0 + 19 + 5} = 0.21$$

$$\text{PRECISION} = \frac{0 + 0.95 + 0.21}{3}$$

$$= 0.39$$

$$= 39\%$$

➤ **Recall**

$$\text{MEMUASKAN} = \frac{0}{0 + 0 + 0} = 0$$

$$\text{SANGAT MEMUASKAN} = \frac{22}{0 + 22 + 19} = 0$$

$$\text{DENGAN PUJIAN} = \frac{5}{0 + 1 + 5} = 0.833$$

$$\text{RECALL} = \frac{0 + 0.536 + 0.833}{3}$$

$$= 0.45$$

$$= 45\%$$

4.8 Kajian Agama

a. Q.S al-Baqarah : 282

وَاتَّقُوا اللَّهَ وَيُعَلِّمُكُمُ اللَّهُ وَاللَّهُ بِكُلِّ شَيْءٍ عَلِيمٌ

Artinya :

“... Dan bertakwalah kepada Allah; Allah mengajarmu; dan Allah Maha Mengetahui segala sesuatu.”

Maksud dari ayat di atas ialah: Allah SWT memerintahkan manusia untuk bertakwa kepada-Nya karena dengan bertakwa, senantiasa Allah SWT akan memudahkan jalan dalam hidup. Hidup manusia selalu dilalui oleh proses belajar. Ketika manusia bertakwa maka Allah SWT akan membantu individu tersebut agar mampu menemukan ilmu dari hal yang sedang dipelajarinya.

Prestasi akademik sebagai salah satu tujuan dari proses belajar akan mampu diraih jika individu tersebut bertakwa sebab mahasiswa akan selalu berusaha semaksimal mungkin dengan disertai doa, dan dengan doa itulah Allah akan menunjukkan jalan menuju prestasi yang gemilang.

b. Q.S Al-‘Alaq : 1-5

أَقْرَأْ بِاسْمِ رَبِّكَ الَّذِي خَلَقَ ۝ خَلَقَ الْإِنْسَانَ مِنْ عَلَقٍ ۝ أَلَمْ يَكُنْ أَكْرَمُ ۝ الَّذِي عَلَّمَ بِالْقَلَمِ ۝ عَلَّمَ الْإِنْسَانَ مَا لَمْ يَعْلَمْ

Artinya:

Bacalah dengan (menyebut) nama Tuhanmu yang Menciptakan, Dia telah menciptakan manusia dari segumpal darah. Bacalah, dan Tuhanmulah yang Maha pemurah, Yang mengajar (manusia) dengan

perantaran kalam, Dia mengajar kepada manusia apa yang tidak diketahuinya.

Ayat di atas adalah ayat pertama yang diturunkan kepada Nabi Muhammad SAW. Kata pertama adalah “*bacalah*”, hal ini membuktikan bahwa pertama kali manusia diperintah untuk membaca (belajar) tetapi tetap dalam jalan yang benar (islam). Ketika ada perintah untuk membaca (belajar) maka secara otomatis manusia diperintah untuk berpretasi dalam belajar. Dengan prestasi belajar yang tinggi (belajar dengan orientasi dunia dan akhirat) maka manusia akan berhasil dalam hidupnya.

c. Q.S. Al-An’am : 160

مَنْ جَاءَ بِالْحَسَنَةِ فَلَهُ عَشْرُ أَمْثَالِهَا وَمَنْ جَاءَ بِالسَّيِّئَةِ فَلَا تُجْزَىٰ إِلَّا
مِثْلَهَا وَهُمْ لَا يُظْلَمُونَ ﴿١٦٠﴾

Artinya:

“Barangsiapa membawa amal baik, maka baginya (pahala) sepuluh kali lipat amalnya, dan Barangsiapa yang membawa perbuatan jahat Maka dia tidak diberi pembalasan melainkan seimbang dengan kejahatannya, sedang mereka sedikitpun tidak dianiaya (dirugikan)” (QS. Al-An’am : 160).

Dari ayat di atas, menjelaskan bahwa lipat sepuluh akan diberikan kepada setiap orang yang melakukan kebaikan, sedang kelipatan-kelipatan yang lebih dari itu, dan semua akan sesuai kehendak Allah berkaitan dengan keadaan-keadaan orang yang berbuat baik, yang Allah ketahui. Begitu pula sebaliknya, barangsiapa melakukan perbuatan buruk yang menjadi tabiat kekafiran dan diliputi oleh kekejian dan kemungkarannya, maka takkan diberi balasan kecuali hukuman yang buruk.

Berdasarkan pengertian di atas penilaian kinerja akademi mahasiswa sangatlah penting. Seorang mahasiswa haruslah melaksanakan tugasnya dengan sungguh-sungguh. Jika seorang mahasiswa melaksanakan tugas yang diberikan kepadanya dengan baik niscaya Allah akan memberikan balasan yang baik sesuai dengan amal yang dikerjakan, tapi jika seorang mahasiswa tersebut tidak melaksanakannya dengan baik, Allah akan memberikan balasan yang setimpal.

d. Hadits Rasulllah

مَنْ كَانَ يَوْمُهُ خَيْرًا مِنْ أَمْسِهِ فَهُوَ رَابِحٌ، وَمَنْ كَانَ يَوْمُهُ
مِثْلَ أَمْسِهِ فَهُوَ مَعْبُودٌ وَمَنْ كَانَ يَوْمُهُ شَرًّا مِنْ أَمْسِهِ فَهُوَ
مَلْعُونٌ (رواه الحاكم)

Artinya :

Barangsiapa yang harinya (hari ini) lebih baik dari sebelumnya, maka ia telah beruntung, barangsiapa harinya seperti sebelumnya, maka ia telah merugi, dan barangsiapa yang harinya lebih jelek dari sebelumnya, maka ia tergolong orang-orang yang terlaknat (HR. Hakim)

Hadits di atas memerintahkan kepada umat Islam bahwa dimanika kehidupan manusia dari hari ke hari harus terus berkembang menuju kepada sesuatu yang lebih baik, sesuatu yang lebih maju dari hari hari yang sudah berlalu, idealnya hari ini harus lebih baik dari hari kemarin, hari esok harus lebih baik dari hari ini dan seterusnya.

Mahasiswa sebagai para penuntut ilmu diperintahkan untuk selalu belajar dari waktu ke waktu agar kinerja akademiknya selalu mengalami peningkatan.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari hasil uji coba yang telah dilakukan oleh penulis pada akhir perancangan aplikasi evaluasi kinerja akademik mahasiswa tersebut dapat dirumuskan beberapa kesimpulan, sebagai berikut:

- Sistem untuk mengevaluasi kinerja akademik mahasiswa ini telah berhasil dibuat menggunakan menggunakan algoritma K-Means Clustering menggunakan bahasa pemrograman java dan database MySQL.
- Memberikan informasi bagi dosen dan mahasiswa tentang perkembangan mahasiswa di bidang akademik.
- Penentuan skala nilai parameter sangat berpengaruh besar pada algoritma *K-Means Clustering* dalam menentukan kelompok predikat kelulusan mahasiswa.
- Untuk mengetahui validasi cluster dilihat dari nilai Precision dan Recall yang dapat dihitung dari tabel Confusion Matrix. Pada pengujian data diketahui nilai Precision = 39 % dan Recall = 45 %. Semakin tinggi nilai Precision dan Recall, maka hasil cluster semakin baik.
- Dari hasil uji coba, pola yang dibentuk mempunyai akurasi kecocokan mencapai 57,44%. dan nilai prosentase kesalahan sebesar 39,59% , hal ini

disebabkan transformasi data atribut berbentuk nominal, ordinal yang belum optimal.

5.2 Saran

Adapun saran agar dapat meningkatkan aplikasi ini dimasa mendatang adalah sebagai berikut :

- Menambahkan variabel lain misal seperti nilai akhir ujian nasional, minat dan hasil test psikologis dalam menentukan evaluasi kinerja akademik mahasiswa.
- Menambahkan tampilan aplikasi yang lebih menarik dan interaktif.
- Mengoptimalkan proses transformasi data atribut.

DAFTAR PUSTAKA

- Agusta, Yudi. Pebruari 2007. “*K-Means Penerapan, Permasalahan dan Metode Terkait*”. Jurnal Sistem dan Informatika Vol.3 : 47-60.
- Barakbah Ridho Ali, *Clustering –Workshop Data Mining*, Politeknik Elektronika Negeri Surabaya.
- Fauziyah, Lilies, *Pendeteksian Serangan Pada Jaringan Komputer Berbasis IDS Snor Dengan Algoritma Clustering K-Means*, Tugas Akhir Mahasiswa Teknik Informatika Surabaya.
- Han, J. and Kamber, M. 2006. “*Data mining: Concepts and Techniques*”, 2nd *The Morgan Kaufmann series in Data Management System*, Jim Grey, series Editor.
- Ilyas, Y., 2001, *Kinerja, Teori, Penilaian dan Penelitian*, Jakarta: Pusat Kajian Ekonomi Kesehatan Fakultas Kesehatan Masyarakat
- Kurniawan, Ari - *Klasterisasi Kompetensi Guru Menggunakan Hasil Penilaian Portofolio Sertifikasi Guru Dengan Metode Data Mining*.
- Kusrini dan Emha Taufiq Luthfi, 2009, *Algoritma Data Mining*, Penerbit Andi Yogyakarta.
- Larose, Daniel T, 2009, *Discovering Knowledge in Data : An Introduction to Data Mining*. John Wiley and Son, Inc. , New York
- Oyelade, O. J, Oladipupo, O.O , Obagbuwa, I. C. *Application of k-Means Clustering algorithm for prediction of Students Academic Performance*. International Journal of Computer Science and Information Security (IJCSIS), Vol. 7, No.1, 2010
- Pramudiono, I. 2007. *Pengantar Data Mining : Menambang Permata Pengetahuan di Gunung Data*. <http://www.ilmukomputer.org/wp-content/uploads/2006/08/iko-datamining>. pada tanggal 27 Maret 2012 Pukul 11.34 WIB
- Purnomo Edi, *Penggunaan Metode Pengklasteran Untuk Menentukan Bidang Tugas Akhir Mahasiswa Teknik Informatika Pens Berdasarkan Nilai*.

Santosa Budi. 2007, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu, Yogyakarta.

Wijaya, Arim (2010) *Analisis Algoritma K-Means untuk Membantu Penjurusan Siswa di MAN Binong Subang*. Jurnal Universitas Komputer Indonesia, Bandung



Lampiran 1

Data Nilai Atribut Mahasiswa

No	NIM	ASAL Jenis Sekolah	Jalur Masuk	IPK	IPK Lulus	Pernah Mondok
1.	07650004	SMAN	PMDK	3.28	3.28	TIDAK
2.	07650006	SMAN	PMDK	3.58	3.58	TIDAK
3.	07650010	MAN	PMDK	3.10	3.10	YA
4.	07650012	MAN	PMDK	3.31	3.31	TIDAK
5.	07650014	SWASTA	PMDK	3.09	3.09	TIDAK
6.	07650015	MAN	PMDK	3.33	3.33	YA
7.	07650021	SWASTA	PMDK	2.99	3.03	YA
8.	07650022	SMAN	PMDK	3.15	3.15	YA
9.	07650025	SWASTA	PMDK	3.29	3.29	TIDAK
10.	07650026	MAN	PMDK	3.26	3.26	TIDAK
11.	07650030	SWASTA	PMDK	3.25	3.25	YA
12.	07650031	SMAN	PMDK	3.45	3.45	TIDAK
13.	07650036	SWASTA	PMDK	3.01	3.01	YA
14.	07650042	SMAN	PMDK	3.27	3.27	TIDAK
15.	07650047	SWASTA	PMDK	3.27	3.27	YA
16.	07650050	SMAN	SNMPTN	3.58	3.58	TIDAK
17.	07650055	SMAN	SNMPTN	3.78	3.78	TIDAK
18.	07650056	SMAN	SNMPTN	3.66	3.66	TIDAK
19.	07650064	SMAN	SNMPTN	3.17	3.17	TIDAK
20.	07650069	SMAN	REGULER	3.33	3.33	TIDAK
21.	07650073	SMAN	SNMPTN	3.20	3.20	TIDAK
22.	07650078	SMAN	SNMPTN	3.52	3.52	TIDAK
23.	07650079	SMAN	REGULER	3.24	3.24	YA
24.	07650080	SMAN	SNMPTN	3.35	3.35	TIDAK
25.	07650081	SMAN	REGULER	3.38	3.38	TIDAK

26.	07650082	SMAN	SNMPTN	3.45	3.45	TIDAK
27.	07650083	SMAN	SNMPTN	3.11	3.11	TIDAK
28.	07650086	MAN	SNMPTN	3.17	3.17	TIDAK
29.	07650087	SMK	REGULER	3.30	3.30	YA
30.	07650092	SWASTA	SNMPTN	3.44	3.44	TIDAK
31.	07650102	SMAN	REGULER	3.21	3.21	TIDAK
32.	07650107	MAN	REGULER	3.52	3.52	YA
33.	07650109	SWASTA	REGULER	3.32	3.32	YA
34.	07650114	SMAN	SNMPTN	3.00	3.02	TIDAK
35.	07650117	SMAN	SNMPTN	3.19	3.19	TIDAK
36.	07650120	SMAN	SNMPTN	3.26	3.26	TIDAK
37.	07650121	SMAN	SNMPTN	3.38	3.38	TIDAK
38.	07650124	SWASTA	REGULER	3.19	3.19	YA
39.	07650129	SWASTA	REGULER	3.46	3.46	YA
40.	07650130	MAN	REGULER	3.46	3.46	YA
41.	07650133	SMK	REGULER	3.32	3.32	YA
42.	07650136	SMAN	REGULER	3.43	3.43	YA
43.	07650142	SMAN	SNMPTN	3.17	3.17	TIDAK
44.	07650147	SWASTA	REGULER	3.03	3.00	TIDAK
45.	07650153	SWASTA	REGULER	3.10	3.10	TIDAK
46.	07650154	SMAN	REGULER	3.10	3.10	TIDAK
47.	07650155	SMAN	REGULER	3.10	3.10	TIDAK