

ANALISIS DATA MAHASISWA UNTUK Mendukung STRATEGI PROMOSI PADA UNIVERSITAS MEDAN AREA MENGGUNAKAN ALGORITMA *K-MEANS CLUSTERING*

M. Rishakim Marajari, Khairuddin Nasution, Tasliyah Haramaini

Program Studi Teknik Informatika, Fakultas Teknik
Universitas Islam Sumatera Utara, Medan, Indonesia

rishakim1702@gmail.com; khairudin33@gmail.com; tazlie@gmail.com

Abstrak

Data mining merupakan proses analisa data untuk menemukan suatu pola dari kumpulan data tersebut. Data mining mampu menganalisa data yang besar menjadi informasi berupa pola yang mempunyai arti bagi pengambil keputusan. Dalam penelitian ini yang akan dibahas adalah mengenai implementasi data mining untuk mendukung strategi promosi pada Universitas Medan Area. Salah satu langkah yang dapat dilakukan untuk meningkatkan jumlah mahasiswa baru adalah dengan memanfaatkan data mahasiswa diintegrasikan dengan teknik data mining sehingga menghasilkan kesimpulan. Kesimpulan tersebut dapat digunakan untuk pengambilan keputusan dalam menentukan wilayah promosi yang tepat sasaran. Pengolahan data mahasiswa seharusnya dilakukan agar dapat menentukan wilayah promosi yang tepat sasaran sehingga dapat meningkatkan jumlah mahasiswa pada tahun berikutnya. Salah satu teknik yang ada pada data mining adalah clustering. Pada penelitian ini akan dibahas teknik clustering yang diterapkan untuk menemukan pola yang terjadi pada data-data mahasiswa yang ada di Universitas Medan Area terutama hal yang berkaitan dengan strategi promosi. Teknik clustering yang akan digunakan adalah K-Means Clustering. Metode ini dipilih karena mampu membuat suatu hasil pengelompokan dengan memanfaatkan data-data mahasiswa yang sudah ada untuk mendukung strategi promosi. Sehingga dapat dihasilkan suatu pola hubungan antara atribut-atribut yang mendukung strategi promosi tersebut. Metode K-Means Clustering ini bertujuan untuk membuat hasil pengelompokan terbaik berdasarkan atribut-atribut yang ada.

Kata-Kata Kunci : *Data Mining, K-Means Clustering, Strategi, Promosi*

I. PENDAHULUAN

Pada institusi pendidikan perguruan tinggi, data dapat diperoleh berdasarkan data historis, sehingga data akan bertambah secara terus menerus, misalnya data mahasiswa. Proses penerimaan mahasiswa baru dalam sebuah perguruan tinggi menghasilkan data yang berlimpah berupa profil dari mahasiswa baru tersebut. Tahap selanjutnya mahasiswa akan melakukan kegiatan belajar mengajar, sehingga dapat diketahui data prestasi mereka setiap akhir semester. Hal ini akan terjadi secara berulang pada sebuah perguruan tinggi. Penumpukan data mahasiswa secara menerus akan memperlambat pencarian informasi terhadap data tersebut.

Berdasarkan berlimpahnya data mahasiswa, informasi yang tersembunyi dapat diketahui dengan cara melakukan pengolahan terhadap data tersebut sehingga berguna bagi pihak universitas. Pengolahan data mahasiswa perlu dilakukan untuk mengetahui informasi penting berupa pengetahuan baru (*knowledge discovery*), misalnya informasi mengenai pengelompokan data mahasiswa berpotensi berdasarkan kota asal mahasiswa. Pengetahuan baru tersebut dapat membantu pihak universitas dalam menentukan strategi promosi terhadap calon mahasiswa baru dengan tepat sasaran. Salah satu langkah yang dapat dilakukan untuk meningkatkan jumlah mahasiswa baru adalah dengan memanfaatkan data mahasiswa diintegrasikan dengan teknik *data mining* sehingga menghasilkan kesimpulan. Kesimpulan tersebut dapat digunakan untuk

pengambilan keputusan dalam menentukan wilayah promosi yang tepat sasaran. Pengolahan data mahasiswa seharusnya dilakukan agar dapat menentukan wilayah promosi yang tepat sasaran sehingga dapat meningkatkan jumlah mahasiswa pada tahun berikutnya.

Penelitian ini akan melakukan pengelompokan menggunakan teknik *Data Mining*, berdasarkan pada data mahasiswa Universitas Medan Area. *Data mining* didefinisikan sebagai proses menemukan pola dalam data. Proses ini harus otomatis atau biasanya secara semi-otomatis. Pola yang dihasilkan harus berarti bahwa pola tersebut memberikan beberapa keuntungan. Pola tersebut diidentifikasi, divalidasi, dan digunakan untuk membuat sebuah prediksi. Data mining memiliki berbagai macam teknik analisa, diantaranya adalah teknik *Clustering*. *Clustering* merupakan data mentah yang sangat besar sulit untuk dianalisis dan dipahami, oleh karena itu perlu adanya pengelompokan/clustering. Dalam hal ini pengelompokan bertujuan untuk meningkatkan pemahaman terhadap data, untuk menilai kualitas dari data tersebut (Yahya, 2019).

Algoritma *K-Means* merupakan metode *non-hierarchical* yang 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 di dalam populasi data dan menandai komponen tersebut ke salah satu pusat *cluster* yang telah didefinisikan tergantung dari jarak minimum

antar komponen dengan tiap-tiap pusat *cluster* (Silvi Agustina, 2016).

Adapun atribut/variabel yang akan digunakan dalam penentuan pengelompokan wilayah untuk menentukan strategi promosi pada Universitas Medan Area adalah program studi mahasiswa, kota asal mahasiswa dan jenis kelamin. Hal-hal tersebut yang mendorong untuk melakukan penelitian ini dengan judul “Analisis Data Mahasiswa Untuk Mendukung Strategi Promosi Pada Universitas Medan Area Menggunakan *K-Means Clustering*”, diharapkan dengan dibuatnya skripsi ini dapat berguna untuk meningkatkan ketepatan promosi terhadap calon mahasiswa baru.

II. METODE PENELITIAN

2.1 Data Mining

Data adalah kumpulan informasi yang digunakan dalam proses pengambilan kesimpulan maupun pengambilan keputusan. Data merupakan bentuk jamak dari bentuk tunggal *datum* atau data-item. Data adalah kenyataan yang menggambarkan suatu kejadian dan kesatuan nyata (Eka Iswandy, 2016).

Data Mining merupakan suatu proses untuk mendapatkan informasi baru dari kumpulan data dengan menggunakan algoritma dan teknik yang melibatkan bidang ilmu statistik, mesin pembelajaran dan sistem manajemen *database*. *Data Mining* digunakan untuk ekstraksi informasi penting yang tersembunyi dari *dataset* yang besar. Dengan adanya *Data Mining* maka akan didapatkan suatu permata berupa pengetahuan di dalam kumpulan data yang banyak jumlahnya.

Banyak istilah yang digunakan untuk menunjukkan proses *Data Mining*. Untuk dapat memberikan pemahaman tentang *Data Mining*, berikut ini adalah beberapa fakta yang terjadi seperti banyaknya organisasi, baik dari dunia bisnis maupun pemerintah berurusan dengan sejumlah informasi dan juga pengelolaan basis data informasi tersebut dan bukan tidak mungkin termasuk didalamnya kebutuhan akan pembangunan *Data Warehouse* dalam skala besar. Seringkali data yang tersimpan tidak dapat secara langsung dianalisa dengan metode statistik standar, hal ini disebabkan karena adanya beberapa *record* yang hilang ataupun karena datanya dalam dimensi ukuran kualitatif bukan kuantitatif.

Data mining merupakan sebuah proses ekstraksi data menjadi informasi atau pengetahuan yang baru. Data yang diolah sebelumnya merupakan data yang bersifat implisit dan biasanya dianggap tidak berguna serta data dalam jumlah besar (Robi Yanto, 2017).

Terdapat empat tugas utama data mining yaitu sebagai berikut :

1. *Predictive Modelling*

Predictive modelling digunakan untuk membangun sebuah model untuk target *variable* sebagai fungsi dari *explanatory variable*. *Explanatory variable* dalam hal ini merupakan semua atribut yang digunakan untuk melakukan

prediksi, sedangkan variabel target merupakan atribut yang akan diprediksi nilainya. *Predictive modelling* dibagi menjadi dua tipe yaitu : *Classification* digunakan untuk memprediksi nilai dari target *variable* yang *discrete* (diskrit) dan *regression* digunakan untuk memprediksi nilai dari target *variable* yang *continue* (berkelanjutan).

2. *Association Analysis*

Association analysis digunakan untuk menemukan aturan asosiasi yang memperlihatkan kondisi-kondisi nilai atribut yang sering muncul secara bersamaan dalam sebuah himpunan data.

3. *Cluster Analysis*

Tidak seperti klasifikasi yang menganalisa kelas data obyek yang mengandung label. *Clustering* menganalisa objek data tanpa memeriksa kelas label yang diketahui. Label-label kelas dilibatkan di dalam data training. Karena belum diketahui sebelumnya. *Clustering* merupakan proses pengelompokkan sekumpulan objek yang sangat mirip.

4. *Anomaly Detection*

Anomaly detection merupakan metode pendeteksian suatu data dimana tujuannya adalah menemukan objek yang berbeda dari sebagian besar objek lain. *Anomaly* dapat dideteksi dengan menggunakan uji statistik yang menerapkan model distribusi atau probabilitas untuk data.

Data Mining merupakan ilmu baru yang berakar dari berbagai bidang ilmu seperti kecerdasan buatan (*Artificial Intelligence*), mesin *learning*, statistik dan *database*. Oleh karena itu dalam menggali informasi, *Data Mining* menggunakan beberapa teknik, antara lain :

1. *Association Discovery*

Association Discovery adalah teknik mempelajari sekumpulan data dan untuk menunjukkan hubungan antara kemunculan atribut-atribut dalam data. Teknik ini mencoba untuk menyiapkan nilai-nilai yang muncul pada saat bersamaan dalam setiap barisnya dan menampilkan hasil keluaran yang disimpulkan dalam sebuah *rule*.

2. *Clustering*

Clustering adalah proses pengumpulan data yang serumpun dari sebuah *dataset* yang lebih besar. Teknik ini menyingkapkan sejumlah kelompok-kelompok yang digunakan sebagai masukan datanya. Dengan *Clustering* kelompok minoritas yang tersebar dikelompokkan dalam sebuah kelompok besar yang memiliki kemiripan entitas. *Clustering* dapat juga digunakan untuk mendeteksi secara otomatis *cluster* dari *record-record* yang berdekatan dengan pengertian tertentu di dalam keseluruhan variabel-variabel.

3. *Sequential Discovery*

Sequential Discovery adalah teknik mencari pola-pola diantara peristiwa-peristiwa yang muncul dalam periode waktu. Metode ini dapat digunakan untuk mencari pola komoditas yang

terjadi berulang kali. Teknik ini terkonsentrasi pada kebiasaan yang sama yang sering muncul dikemudian hari. Menurut laporan tersebut diungkapkan bahwa algoritma yang dipakai memiliki kriteria sebagai berikut :

- a. *One Scan* : algoritma ini membutuhkan paling sedikit satu kali penelusuran basis data untuk membuat *cluster* yang diinginkan.
- b. *Anytime Algorithm* : algoritma ini selalu dapat menyediakan jawaban yang terbaik setiap saat selama komputasi dijalankan.
- c. *Interruptable and Incremental* : algoritma ini dapat ditunda, dapat dihentikan dan dijalankan lagi, dan hasil sementara yang muncul dapat disimpan untuk kelanjutan komputasi data baru lebih lanjut.
- d. *Limited RAM Requirement* : algoritma ini mampu bekerja dalam kondisi memori yang terbatas.
- e. *Forward-only Cursor* : algoritma ini mampu untuk beroperasi pada basis data dengan kursor (proses) bergerak maju hingga akhir data.

4. Classification

Classification adalah proses pengumpulan data bersama-sama yang didasarkan atas sekumpulan kesamaan yang awalnya telah ditentukan oleh seorang analis sebelum analisa dimulai. Teknik ini memeriksa data yang telah diklasifikasikan dan dikumpulkan dalam grup bersama-sama sesuai dengan aturan keanggotaannya. Aturan keanggotaan bisa mempunyai komponen waktu, komponen geografis, komponen kuantitatif. Proses klasifikasi ini dapat dibagi menjadi tiga fase:

- a. *Learning* : algoritma yang mencari sejumlah *record* dari *training set* dan menciptakan sebuah deskripsi tentang model klasifikasi. Model ini dibuat kecil dan persis pada saat yang sama.
- b. Model yang sudah dibuat harus diuji dengan serangkaian uji coba dalam sebuah basis data. Sebagai *training set*, *record-record* uji coba harus merupakan kumpulan yang sudah lebih dahulu diklasifikasikan. *Record* untuk uji coba harus berbeda dengan *training set*. *Model classification* akan selalu bekerja dengan sempurna dalam *training set* yang digunakan.
- c. *Classification*, model *iterasi* yang digunakan untuk membuat klasifikasi pada sisa *record-record* dalam basis data yang ada.

5. Neural Network

Neural Network merupakan sebuah metode khusus untuk pengendalian identifikasi pola yang digunakan pada *trend* perkiraan berdasarkan kebiasaan yang telah diketahui sebelumnya. Suatu *trend* atau kecenderungan dapat diidentifikasi sebagai pergerakan yang ditujukan berdasarkan pada kebiasaan yang lama. Inti dari pemrosesan didasarkan pada pengolahan data yang meniru dari fungsi sistem saraf tubuh manusia. Pengetahuan dapat dipelajari dari suatu kumpulan data yang

terpisah dan kompleks. Ada tiga lapisan (*layers*) pada *network*, lapisan kiri yang menerima masukan (*input*), lapisan tersembunyi (*hidden layer neurons*) yang menjalankan fungsi kerja pemrosesan datanya, dan lapisan kanan yang menampilkan analisa hasil keluaran. Ketika *network* dilatih (*trained*) dengan informasi yang dimasukkan melalui masukan (*input*) akan diproses oleh lapisan tengah (*neurons*) yang menjadikannya cerdas dalam sejumlah elemen data yang ada dan mengeluarkan hasil pada lapisan keluarannya.

2.2 Algoritma K-Means Clustering

K-Means adalah salah satu algoritma dalam *clustering* yang berulang-ulang. Algoritma *K-Means* dimulai dengan memilih secara acak K , K di sini merupakan banyaknya *cluster* yang ingin dibentuk. Kemudian menetapkan nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari *cluster* atau biasa disebut dengan *centroid*, *mean* atau “*means*”.

Clustering adalah studi formal untuk partisi atau mengelompokkan data dengan tidak menggunakan pelabelan kategori. *Clustering* bersifat *unsupervised learning* atau tidak mempunyai tahap pelatihan data, berbeda dengan klasifikasi. *Clustering* digunakan untuk mengelompokkan data secara alamiah berdasarkan kemiripan pada objek data dan sebaliknya meminimalkan kemiripan terhadap *cluster* lain (Beta Estri, 2018).

Menggunakan pendekatan *partitional clustering*. Tiap *cluster* dihubungkan dengan sebuah *centroid* (titik pusat). Tiap titik ditempatkan ke dalam *cluster* dengan *centroid* terdekat. Jumlah *cluster*, K , harus ditentukan. Berikut adalah langkah-langkah algoritma *K-Means* (Sulistiyowati, 2018) :

1. Pilih jumlah *cluster* k
2. Inisialisasi k pusat *cluster* bisa dilakukan dengan berbagai cara. Cara yang sering dilakukan adalah dengan cara acak. Pusat-pusat *cluster* diberi nilai awal dengan angka-angka *random*
3. Alokasikan semua data/objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean*.
4. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan *median* (nilai tengah) dari *cluster* tersebut. Jadi rata-rata (*mean*) bukan satu-satunya ukuran yang bisa dipakai.
5. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak

berubah lagi maka proses *clustering* selesai. Jika pusat *cluster* berubah maka kembali ke langkah nomor 3 lagi.

Pada langkah 3 lokasi *centroid* (titik pusat) setiap kelompok yang diambil dari rata-rata (*mean*) semua nilai data pada setiap fiturnya harus dihitung kembali. Jika M menyatakan jumlah data dalam sebuah kelompok, i menyatakan fitur ke- i dalam sebuah kelompok, dan p menyatakan dimensi data, untuk menghitung sentroid fitur ke- i digunakan formula :

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j$$

Formula tersebut dilakukan sebanyak p dimensi sehingga i mulai dari 1 sampai p .

Ada beberapa cara yang dapat digunakan untuk mengukur jarak data ke pusat kelompok, diantaranya *Euclidean*, *Manhattan/City Block* serta *Minkowsky*. Masing-masing cara mempunyai kelebihan dan kekurangan.

Pengukuran jarak pada ruang jarak (*distance space*) *Euclidean* menggunakan formula

$$D(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2}$$

D adalah jarak antara data x_2 dan x_1 , dan $|\cdot|$ adalah nilai mutlak. Pengukuran jarak pada ruang jarak *Manhattan* menggunakan formula

$$D(x_2, x_1) = \|x_2 - x_1\|_1 = \sum_{j=1}^p |x_{2j} - x_{1j}|$$

Pengukuran jarak pada ruang jarak *Minkowsky* menggunakan formula

$$D(x_2, x_1) = \|x_2 - x_1\|_\lambda = \sqrt[\lambda]{\sum_{j=1}^p |x_{2j} - x_{1j}|^\lambda}$$

λ adalah parameter *Minkowsky*. Secara umum, λ merupakan parameter penentu dalam karakteristik jarak. Jika $\lambda=1$, ruang jarak pada *Minkowsky* sama dengan *Manhattan*. Jika $\lambda=2$, ruang jaraknya akan sama dengan *Euclidean*; jika $\lambda=\infty$, ruang jaraknya akan sama dengan ruang jarak *Chebyshev*. Namun demikian, cara yang paling banyak digunakan adalah *Euclidean* dan *Manhattan*. *Euclidean* menjadi pilihan jika kita ingin memberikan jarak terpendek antara dua titik (jarak lurus), sedangkan *Manhattan* memberikan jarak terjauh pada dua data. *Manhattan* juga sering digunakan karena kemampuannya dalam mendeteksi keadaan khusus, seperti keberadaan *outlier*, dengan lebih baik.

2.3 Unified Modeling Language (UML)

Untuk memberikan pemahaman secara jelas terhadap sistem yang akan dirancang, maka dapat digunakan *Unified Modeling Language* (UML). Dalam *Unified Modeling Language* (UML) ini akan terlihat dengan jelas aktifitas yang terjadi pada proses pembentukan aplikasinya dari awal hingga akhir. *Unified Modeling Language* (UML) adalah keluarga notasi grafis yang didukung oleh meta-model tunggal, yang membantu pendeskripsian dan desain sistem perangkat lunak, khususnya sistem yang dibangun menggunakan pemrograman berorientasi objek (Tofik Isa, 2017).

Unified Modeling Language (UML) merupakan bahasa standar yang digunakan untuk

mendokumentasikan, menspesifikasikan dan membangun perangkat lunak. UML merupakan metodologi dalam mengembangkan sistem berorientasi objek dan merupakan alat untuk mendukung pengembangan sistem.

III. METODOLOGI PENELITIAN

3.1 Analisis Permasalahan

Ada beberapa tahapan dalam membuat penyelesaian masalah dengan metode *K-Means Clustering* yaitu sebagai berikut:

1. Pilih jumlah *cluster* k .
2. Inialisasi k pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara *random*. Pusat-pusat *cluster* diberi nilai awal dengan angka-angka *random*.
3. Alokasikan semua data/objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$D_{(i,j)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

Di mana:

$D(i,j)$ = Jarak data ke i ke pusat *cluster* j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

4. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari *cluster* tersebut. Jadi rata-rata (*mean*) bukan satu-satunya ukuran yang bisa dipakai.
5. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi

Tabel 1. Inialisasi Titik Pusat Awal Setiap Cluster

Data Ke-	Nama Mahasiswa	Jenis Kelamin	Kota Asal	Prodi
06	SRI HARTINI	1	1	1
25	IMELDA BR TARIGAN	1	3	2

Data Ke-	Nama Mahasiswa	Jenis Kelamin	Kota Asal	Prodi
46	AHMAD SYUKUR	2	4	3

Sebagai contoh, dihitung jarak dari Mahasiswa pertama ke pusat *cluster* pertama:

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

Jarak data Mahasiswa pertama ke pusat *cluster* kedua:

$$D_{(1,2)} = \sqrt{(1-1)^2 + (2-3)^2 + (1-2)^2} = 1,41$$

Jarak data Mahasiswa pertama ke pusat *cluster* ketiga:

$$D_{(1,3)} = \sqrt{(1-2)^2 + (2-4)^2 + (1-3)^2} = 3,00$$

Kemudian Hitung Jarak dari data Mahasiswa Kedua ke pusat *cluster* pertama:

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

Jarak data Mahasiswa kedua ke pusat *cluster* kedua:

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

Jarak data Mahasiswa kedua ke pusat *cluster* ketiga:

$$D_{(2,3)} = \sqrt{(1-2)^2 + (3-4)^2 + (1-3)^2} = 2,45$$

Berdasarkan hasil ketiga perhitungan di atas dapat disimpulkan bahwa jarak data Mahasiswa Pertama yang paling dekat adalah dengan *cluster* 1, sehingga data Mahasiswa pertama dimasukkan ke dalam *cluster* 1. Untuk jarak data Mahasiswa Kedua yang paling dekat adalah dengan *cluster* 2, sehingga data Mahasiswa Kedua dimasukkan ke *cluster* 2.

Berdasarkan hasil pengelompokan data menggunakan metode k-means *clustering*, di dapatkan hasil *clustering* hingga iterasi ke-2, dimana titik pusat tidak lagi berubah dan tidak ada data yang berpindah antar *cluster*. Hasil dari *clustering* tersebut seperti pada tabel di bawah ini.

Tabel 2. Anggota Cluster 1 Hasil Iterasi 2

Nama	Jkel	Kota Asal	Prodi
TASYA NURUL HUDA	1	2	1
RAVITA RAMBE	1	1	1
SRI HARTINI	1	1	1
PUTRY AMALIA BR LUBIS	1	1	1
NADHILAH PUTRI	1	2	1
TITIN ASTIKA	1	1	1
ANNISA	1	2	1
NOVIKA PURNAMA SARI	1	1	2

Nama	Jkel	Kota Asal	Prodi
RIZKI KHAIRANI	1	1	2
ANNISA PUTRI BATUBARA	1	1	2
RESKA ELISIANA RITONGA	1	1	2
SULASNI	1	1	2
KURNIAWAN	2	1	3
RICARDO SEBAYANG	2	1	3
AYU SYAFIRA	1	1	3
RAMONTUA SINAGA	2	1	3
RIKARDO ZALUKHU	2	1	3

Tabel 3. Anggota Cluster 2 Hasil Iterasi 2

Nama	Jkel	Kota Asal	Prodi
PUTRIE RIZKY ANANDA	1	3	1
CUT MUTIA SARI	1	4	1
AINUR RIZA	1	4	1
FARAH DILLA	1	4	1
YUSNITA SEMBIRING	1	4	1
BAGINDA ROMADHAN SIREGAR	2	3	2
HERLINA SIREGAR	1	4	2
ISRA HAYATI	1	2	2
NOVITA SARI	1	3	2
IMELDA BR TARIGAN	2	3	2
SIHOL SIHOMBING	1	2	2
ANNISA RITONGA	2	2	3
AHMAD ADJI ALAMSYAH	2	3	3
SUSI INDAH ROSLIA	2	4	3
FAKHRY HAMBALI	2	4	3

Tabel 4. Anggota Cluster 3 Hasil Iterasi 2

Nama	Jkel	Kota Asal	Prodi
ADE AULIA LESTARI	1	1	5
KHOIRUL FIKRI	2	1	4
ERISA SRJEKI MANURUNG	1	1	4
TRI WULANDARI	1	2	5
IBNU FAHMI	2	2	5
SYAFARUDDIN RITONGA	1	2	5
UMMY ZULFADLAH	1	2	5
M.FAHRUL RIDHO	2	3	5
M.DEDE ERIANDI	2	2	5
YOGI SAPUTRA	1	3	5
SANDI PRIMA NAIBAHO	2	3	4
JHON HORASMAN TURNIP	2	2	4
ARDI WANDIRA	2	2	4
NANDA HERIAMBANG	2	2	4
LAMEKI GOHAE	2	2	4
AHMAD SYUKUR	2	3	4
KHAIRIL HUDA	2	3	4

Dari hasil *clustering* di atas dapat diambil kesimpulan bahwa:

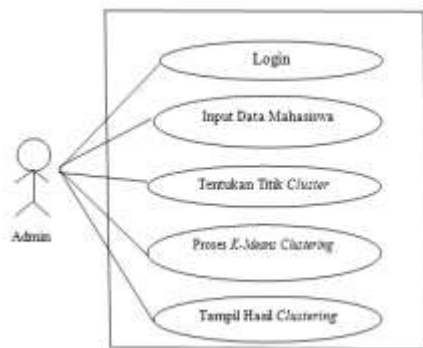
1. Karakteristik Data Mahasiswa pada *cluster* 1 mayoritas Kota Asal yaitu Medan dan Binjai, dengan jumlah data 17 data mahasiswa. Dengan mayoritas mahasiswa memilih Prodi Ilmu Hukum dan Agroteknologi. Berarti untuk

wilayah Medan lebih diprioritaskan untuk mempromosikan mengenai Ilmu Hukum dan Agroteknologi.

2. Karakteristik Data Mahasiswa pada *cluster* 2 mayoritas Kota Asal yaitu Aceh, dengan jumlah data 16 data mahasiswa. Dengan mayoritas mahasiswa memilih Prodi Agroteknologi. Berarti untuk wilayah Aceh lebih diprioritaskan untuk mempromosikan mengenai Prodi Agroteknologi.
3. Karakteristik Data Mahasiswa pada *cluster* 3 mayoritas Kota Asal yaitu Luar Kota 1 yaitu wilayah Berastagi, Kabanjahe, Kisaran, Siantar dan tebing Tinggi, dengan jumlah data 17 data mahasiswa. Dengan mayoritas mahasiswa memilih Prodi Agribisnis dan Administrasi Publik. Berarti untuk wilayah ini lebih diprioritaskan untuk mempromosikan mengenai Prodi Agribisnis dan Administrasi Publik.

3.2 Use Case Diagram

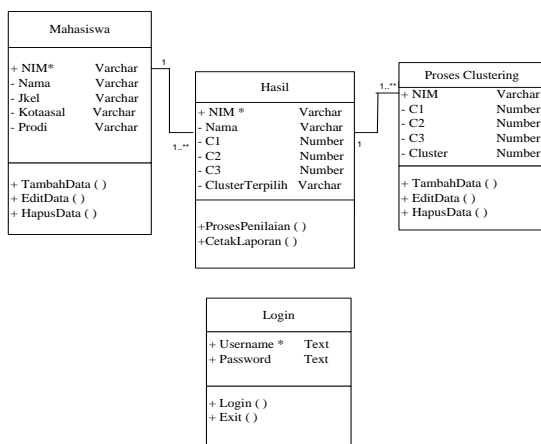
Use case menunjukkan hubungan interaksi antar aktor dengan *use case* di dalam suatu yang bertujuan untuk menentukan bagaimana aktor berinteraksi dengan sebuah sistem.



Gambar 1. Use Case Diagram Sistem

3.3 Class Diagram

Class Diagram memperlihatkan hubungan antar kelas dan penjelasan detail tiap-tiap kelas di dalam model desain (dalam *logical view*) dari suatu sistem. Bentuk *Class Diagram* dari sistem yang dibangun dapat dilihat pada Gambar 2.



Gambar 2. Class Diagram

IV. HASIL DAN PEMBAHASAN

Implementasi sistem menjelaskan mengenai hasil sistem pendukung keputusan yang telah dibangun. Terdiri dari beberapa form input dan beberapa laporan. Berikut di bawah ini dijelaskan lebih detail..

1. Login

Pada awal aplikasi dijalankan akan menampilkan *form login*, dimana *user* diwajibkan untuk mengisi *username* dan *password* yang sudah terdaftar sebelumnya. Dan akan menampilkan pesan atau peringatan (*warning*) jika nama *user* atau *password* salah. Seperti pada Gambar 3.



Gambar 3. Form Login

2. Menu Utama

Tampilan selanjutnya adalah menu utama Aplikasi data *mining* setelah melakukan *login* terlebih dahulu. Menu utama terdiri dari beberapa menu yaitu Data Mahasiswa, Proses, Laporan dan Keluar. Setiap menu memiliki sub menu masing-masing, kecuali menu Keluar yang tidak memiliki sub menu. Karena Keluar hanya berfungsi untuk mengakhiri program. Gambar halaman utama aplikasi data *mining* ini dapat dilihat pada Gambar 4.



Gambar 4. Tampilan Form Menu Utama

3. Form Mahasiswa

Tampilan ini berisikan tentang data Mahasiswa yang berfungsi sebagai media dalam memasukan data mahasiswa baru dan juga mengedit serta menghapus data Mahasiswa. Tampilan *form* dirancang agar mudah untuk digunakan oleh *user*. Adapun tampilan *form* sebagai berikut:



Gambar 5. Tampilan Form Input Data Mahasiswa

4. Form Titik Cluster

Tampilan Form Titik Cluster ini berfungsi untuk menampilkan hasil inialisasi dan memilih 3 data sebagai titik pusat Cluster. Adapun cara penggunaannya dengan terlebih dahulu memilih (checklist) pada list yang tersedia pada halaman titik cluster. Tampilan form titik cluster dapat dilihat pada Gambar 6.



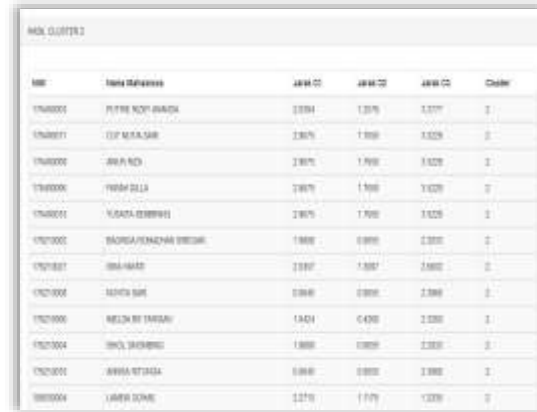
Gambar 6. Tampilan Form Titik Cluster

5. Form Proses Clustering

Tampilan Form Proses Clustering ini berfungsi untuk melakukan proses perhitungan jarak data ke titik Cluster menggunakan metode k-means clustering dan menampilkan hasil perhitungan. Adapun hasil perhitungannya tampil dalam bentuk listview. Klik tombol Proses untuk memulai perhitungan dengan metode K-Means Clustering. Tampilan form sebagai berikut :



Gambar 7. Tampilan Form Clustering (Cluster 1)



Gambar 8. Tampilan Form Clustering (Cluster 2)



Gambar 9. Tampilan Form Clustering (Cluster 3)

6. Laporan Hasil Clustering

Form Laporan ini berfungsi untuk melihat hasil perhitungan Cluster dengan metode K-Means Clustering beserta informasi lain mengenai Proses Titik Cluster tersebut. Adapun hasil Cluster akan tampil pada kolom Cluster. Tampilan preview dapat dilihat pada Gambar 10.



Gambar 10. Tampilan Preview Laporan Hasil Clustering

Laporan hasil clustering ini merupakan hasil perhitungan dari setiap data mahasiswa ke jarak tiap pusat cluster pada iterasi 2.

No	PREMISI	WILAYAH	0.141	1.163	1.943
01	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
02	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
03	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
04	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
05	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
06	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
07	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
08	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
09	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
10	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
11	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
12	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
13	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
14	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
15	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
16	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
17	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
18	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
19	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
20	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
21	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
22	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
23	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
24	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
25	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
26	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
27	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
28	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
29	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
30	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
31	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
32	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
33	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
34	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
35	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
36	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
37	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
38	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
39	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
40	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
41	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
42	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
43	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
44	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
45	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
46	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
47	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
48	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
49	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184
50	PREMISI	ALUNDA BIRANG	0.864	0.837	1.184

Gambar 11. Tampilan Preview Laporan Hasil Clustering (Bagian 2)

Dari hasil yang di atas, dapat dilihat bahwa dengan adanya aplikasi data mining mendukung strategi promosi pada Universitas Medan Area (UMA) menggunakan algoritma *K-Means Clustering* ini, yang dibuat dengan menggunakan perangkat-perangkat lunak (*software*) dapat menganalisa data mahasiswa sehingga menghasilkan laporan pengelompokan data mahasiswa pada tiap-tiap *cluster* yang dapat digunakan untuk kepentingan strategi promosi.

V. KESIMPULAN

- Berdasarkan pembahasan dan evaluasi, maka dapat ditarik beberapa kesimpulan sebagai berikut :
1. Hasil dari aplikasi data mining yang dibangun dapat mengadopsi metode *K-Means Clustering* yang digunakan untuk mendukung strategi promosi di Universitas Medan Area (UMA) secara tepat dan akurat.
 2. Hasil dari aplikasi data mining yang dibangun dengan menggunakan perangkat-perangkat lunak (*software*) dapat menganalisa data mahasiswa sehingga menghasilkan laporan pengelompokan data mahasiswa pada tiap-tiap *cluster* yang dapat digunakan untuk kepentingan strategi promosi.
 3. Setelah dilakukan pengujian dengan sampel data sebanyak 50 data mahasiswa, diperoleh hasil 3 (tiga) cluster. Dimana masing-masing cluster memiliki karakteristik yang berbeda-beda.

DAFTAR PUSTAKA

[1] Agustina, Silvia, dkk, 2016, *Clustering Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode K-Means*, *J. Informatika dan Teknologi*, vol. 2, no. 2.

[2] Estri, Beta, dkk, 2018, *Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering*, *J. JUTEI*, vol.2, no.1.

[3] Heriyanto, Yunahar, 2018, *Perancangan Sistem Informasi Rental Mobil Berbasis Web Pada PT.APM Rent Car*, *J. Intra-Tech*, vol. 2, no. 2, pp. 64–77.

[4] Iswandy, Eka, 2016, *Analisa Dan Perancangan Sistem Informasi Penagihan Purchasing Order Customer Studi Kasus Pada Cv. Vertical Cipta Relasi Padang Dengan Metode Centralized Data Processing*, *J. TEKNOIF*, vol. 4, no. Oktober, p. 14.

[5] Mega, Windha, 2015, *Clustering Menggunakan Metode K-Means Untuk Menentukan Status Gizi Balita*, *Jurnal Informatika.*, vol. 15, no. 2.

[6] Santoso dan R. Nurmalina, 2017, *Perencanaan dan Pengembangan Aplikasi Absensi Mahasiswa Menggunakan Smart Card Guna Pengembangan Kampus Cerdas (Studi Kasus Politeknik Negeri Tanah Laut)*, *J. Integr.*, vol. 9, no. 1, pp. 84–91.

[7] Suendri, 2018, *Implementasi Diagram UML (Unified Modelling Language) Pada Perancangan Sistem Informasi Remunerasi Dosen Dengan Database Oracle (Studi Kasus: UIN Sumatera Utara Medan)*, *J. Ilmu Komput. dan Inform.*, vol. 3, no. 1, pp. 1–9.

[8] Sulistyowati, dkk, 2018, *Analisa Segmentasi Konsumen Menggunakan Algoritma K-Means Clustering*, Institut Teknologi Adhi Tama Surabaya.

[9] Tofik Isa dan G. P. Hartawan, 2017, *Perancangan Aplikasi Koperasi Simpan Pinjam Berbasis Web (Studi Kasus Koperasi Mitra Setia)*, *J. Ilm. Ilmu Ekon.*, vol. 5, no. 10, pp. 139–151.

[10] Yahya dan Mahfuz, 2019, *Penggunaan Algoritma K-Means Untuk Menganalisis Pelanggan Potensial Pada Dealer SPS Motor Honda Lombok Timur Nusa Tenggara Barat*, *J. Informatika dan Teknologi*, vol. 2, no. 2.

[11] Yanto, Robi dan R. Khoiriah, 2017, *Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat*, *Creat. Inf. Technol. J.*, vol. 2, no. 2, p. 102.