

PENERAPAN DATA MINING DALAM PENGELOMPOKAN KELAS MAHASISWA BERDASARKAN HASIL UJIAN SARINGAN MASUK DENGAN ALGORITMA K-MEANS

Trinanda Syahputra

Sistem Informasi, STMIK Triguna Dharma

Email: trinandasyahputra@gmail.com

Abstract: Processing student data needs to be done to find out important information in the form of new knowledge. Information about grouping student data based on test scores. The new knowledge can help the campus in determining new classes for prospective new students correctly. In classifying student classes with clustering methods that take the entrance examination exam based on the scores produced by each prospective student, the class will be determined according to the direction they take without any manipulation and the value information can be sent to prospective students.

Keywords: *Students, Clustering Method*

Abstrak: Pengolahan data mahasiswa perlu dilakukan untuk mengetahui informasi penting berupa pengetahuan baru. Informasi mengenai pengelompokan data mahasiswa berdasarkan nilai ujian. Pengetahuan baru tersebut dapat membantu pihak kampus dalam menentukan kelas baru bagi calon mahasiswa baru dengan tepat. Dalam pengelompokan kelas mahasiswa dengan metode *clustering* yang mengikuti ujian saringan masuk berdasarkan nilai yang dihasilkan setiap calon mahasiswa maka akan ditentukan kelas sesuai jurusan yang mereka ambil tanpa ada manipulasi dan informasi hasil nilai dapat di kirim kepada calon mahasiswa

Kata kunci: Mahasiswa, Metode Clustering

PENDAHULUAN

Pada industri pendidikan perguruan tinggi, data dapat diperoleh berdasarkan data historis, sehingga data akan bertambah terus menerus, misalnya data mahasiswa. Proses penerimaan mahasiswa baru dalam sebuah perguruan tinggi menghasilkan data yang berlimpah berupa profil dari mahasiswa baru tersebut[1]. Tahap selanjutnya mahasiswa akan melakukan kegiatan

belajar mengajar, sehingga dapat diketahui data prestasi mahasiswa 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 kampus. Pengolahan data mahasiswa perlu dilakukan untuk mengetahui informasi penting berupa pengetahuan baru (*knowledge discovery*), misalnya informasi mengenai pengelompokan data mahasiswa berdasarkan nilai ujian. Pengetahuan baru tersebut dapat membantu pihak kampus dalam menentukan kelas baru bagi calon mahasiswa baru dengan tepat[2].

K-means clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil[3].

Data Mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk menginteraksi mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terikat dari berbagai database besar[4].

METODE

Metode yang di gunakan meliputi metode analisis yang terdiri dari metode wawancara (lapangan) dan studi pustaka.

1. Metode Lapangan (*Field Research*)
 - a. Pengamatan

Dalam metode pengamatan ini di lakukan pengamatan secara langsung pada bagian inventory yang menangani segala kegiatan yang berhubungan dengan keluar masuk barang dan persediaan barang.

- b. Sampel
Sampel yang di gunakan ialah beberapa transaksi pengadaan dan peminjaman barang.
- c. Wawancara
Pengumpulan data atau informasi pada metode ini dapat di lakukan dengan wawancara atau mengajukan pertanyaan langsung pada bagian terkait.

2. Metode Pustaka (*Library research*)
Metode pustaka ialah suatu cara kepustakaan untuk mengumpulkan data berdasarkan buku-buku.

HASIL DAN PEMBAHASAN

Dalam kegiatan layanan dunia pendidikan untuk mempermudah seleksi dan penggolongan calon mahasiswa mengalami beberapa permasalahan yang menyangkut kebutuhan data dan informasi tentang batas nilai minimal, sehingga untuk melakukan kegiatan penerimaan mahasiswa baru mengalami beberapa masalah. Permasalahan itu antara lain, yaitu:

1. Sulitnya melakukan analisa yang efektif karena tidak adanya sistem yang dapat menyajikan data historis sehingga dapat memberikan *output* tentang berapa banyak jumlah calon mahasiswa yang dimiliki dan kelompok-kelompok calon mahasiswa yang mendaftar, karena data yang ada masih berbentuk data

manual dan belum dimanfaatkan secara maksimal.

2. Terlalu banyaknya calon mahasiswa yang mendaftar, sehingga diperlukan sebuah sistem yang bisa mendeteksi berapa jumlah calon mahasiswa.

Batas nilai minimal adalah suatu yang sifatnya dinamis artinya, setiap periode penerimaan mahasiswa baru memiliki standar tersendiri untuk jurusan-jurusan tertentu setiap tahunnya. Batas nilai minimal bukanlah standar kelulusan dalam saringan masuk mahasiswa baru tetapi keketapan yang bertujuan untuk menyeleksi yang baik dari yang terbaik, bukan menilainya. Adapun cara untuk menghitung batas nilai minimal dalam menjawab soal-soal yang diberikan pada saat seleksi. Contohnya dengan mengerjakan soal tahun lalu dan kemudian mencocokkan jawaban dengan kunci jawaban yang tersedia. Lalu hitung batas nilai minimal dengan **rumus** batas nilai minimal.

$$\text{Batas nilai minimal} = \frac{(B \times 4) - (S \times 1) \times 100}{JS \times 4}$$

Keterangan :

B = Jumlah jawaban benar

S = Jumlah jawaban salah

JS = Jumlah soal

Mahasiswa 1:

Onika Sihotang menjawab 50 soal dengan benar dan 18 soal yang salah 32 dan soal tidak dijawabnya.

$$\text{Batas nilai minimal :} \\ \frac{(50 \times 4) - (18 \times 1) \times 100}{100 \times 4} = 45.50$$

Mahasiswa 2:

Riski Tampubolon menjawab 27 soal dengan benar dan 4 soal yang salah dan 69 soal tidak dijawabnya.

$$\text{Batas nilai minimal :} \\ \frac{(27 \times 4) - (4 \times 1) \times 100}{100 \times 4} = 26$$

Mahasiswa 3:

Eva Veronika Br Ketaren menjawab 50 soal dengan benar dan 25 soal yang salah dan 5 soal tidak dijawabnya.

$$\text{Batas nilai minimal :} \\ \frac{(50 \times 4) - (25 \times 1) \times 100}{100 \times 4} = 43.75$$

Mahasiswa 4:

Vini Fionika menjawab 19 soal dengan benar dan 20 soal yang salah dan 61 soal tidak dijawabnya.

$$\text{Batas nilai minimal :} \\ \frac{(19 \times 4) - (20 \times 1) \times 100}{100 \times 4} = 14$$

Mahasiswa 5:

Srianta Oktaria menjawab 34 soal dengan benar dan 32 soal yang salah dan 34 soal tidak dijawabnya.

$$\text{Batas nilai minimal :} \\ \frac{(34 \times 4) - (32 \times 1) \times 100}{100 \times 4} = 26$$

Mahasiswa 6:

Mitra Arjuna Sitepu menjawab 17 soal dengan benar dan 21 soal yang salah dan 62 soal tidak dijawabnya.

$$\text{Batas nilai minimal :} \\ \frac{(17 \times 4) - (21 \times 1) \times 100}{100 \times 4} = 11.75$$

Mahasiswa 7:

Rotua Verawati menjawab 41 soal dengan benar dan 8 soal yang salah dan 51 soal tidak dijawabnya.

Batas nilai minimal :

$$\frac{(41 \times 4) - (8 \times 1) \times 100}{100 \times 4} = 39$$

Mahasiswa 8:

Revie Amalia Ginting menjawab 36 soal dengan benar dan 16 soal yang salah dan 32 soal tidak dijawabnya.

Batas nilai minimal :

$$\frac{(36 \times 4) - (16 \times 1) \times 100}{100 \times 4} = 32$$

Mahasiswa 9:

Dewi Afrida Manulang menjawab 50 soal dengan benar dan 32 soal yang salah dan 13 soal tidak dijawabnya.

Batas nilai minimal :

$$\frac{(50 \times 4) - (32 \times 1) \times 100}{100 \times 4} = 42$$

Mahasiswa 10:

Anita Natalia Marpaung menjawab 20 soal dengan benar dan 32 soal yang salah dan 18 soal tidak dijawabnya.

Batas nilai minimal :

$$\frac{(20 \times 4) - (32 \times 1) \times 100}{100 \times 4} = 12$$

Kasus yang diambil dalam perhitungan Batas nilai minimal adalah Batas nilai minimal dari beberapa PTS yang terletak di Kota Medan, Indonesia. Pembangunan perangkat lunak *Data Mining* dengan metode *Clustering* menggunakan algoritma *K-means* untuk pengelompokan mahasiswa dalam studi kasus ini, fungsi yang dipakai adalah fungsi untuk menentukan titik pusat yang berguna sebagai pusat-pusat kelompok mahasiswa. Fungsi yang dimaksud adalah:

1. Data awal yang digunakan dalam perhitungan ini berjumlah 20.

2. Data mahasiswa sebelum dilakukan perhitungan menggunakan metode *Clustering K-Means*.
3. Dan ketentuan tingkat Batas Nilai Minimal adalah pada tabel berikut :

Tabel 1. Batas Nilai Minimal

Batas Nilai Minimal Mahasiswa	Himpunan	Nilai Himpunan
0-20	Rendah	1
20.1 – 40	Cukup	2
> 40	Tinggi	3

Tabel 2. Data Awal

Calon Mahasiswa Ke	Batas Nilai Minimal (%)	Himpunan
1	45.5	3
2	26	2
3	43.75	3
4	14	1
5	26	2
6	11.75	1
7	39	2
8	32	2
9	42	3
10	12	1

Algoritma ini disusun atas dasar ide yang sederhana. Pada awalnya ditentukan berapa *cluster* yang akan dibentuk. Sebarang obyek atau elemen pertama dalam *cluster* dapat dipilih untuk dijadikan sebagai titik tengah (*centroid point*) *cluster*. Algoritma metode *K-Means* selanjutnya akan melakukan pengulangan langkah-langkah berikut sampai terjadi kestabilan (tidak ada obyek yang dapat dipindahkan) :

1. menentukan koordinat titik tengah setiap *cluster*,
2. menentukan jarak setiap obyek terhadap koordinat titik tengah,
3. mengelompokkan obyek-obyek tersebut berdasarkan pada jarak minimumnya.

Adanya langkah-langkah yang dilakukan oleh algoritma metode *K-Means* untuk kasus ini adalah sebagai berikut :

1. Penentuan nilai awal titik tengah. Langkah pertama, menentukan pusat *cluster* secara acak pada data awal yang ada di tabel 3.2. *Cluster* pertama yang akan digunakan dalam perhitungan adalah C1 (11.75:1) yang artinya Batas Nilai Minimum mahasiswa pada pengelompokan pertama adalah 10 dan himpunan ada 1. *Cluster* kedua C2 (26:2) dan *cluster* ketiga yang akan digunakan adalah C3 (42:3).
2. Menghitung jarak obyek ke *centroid* menggunakan rumus Euclidean yaitu :

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Dimana :

d = *Distance space*

x_1 = Nilai data pertama

pada *cluster* pertama

y_1 = Nilai data kedua pada *cluster* pertama

x_2y_2 = Nilai rata-rata disetiap *cluster*

Tabel 3. Posisi *cluster* setiap mahasiswa

Mahasiswa Ke	<i>Cluster</i>
1	C3
2	C2
3	C3
4	C1
5	C2
6	C1
7	C3
8	C2
9	C3
10	C1

3. Langkah selanjutnya menghitung pusat *cluster* baru.

- a. Pusat *cluster* yang baru ditentukan berdasarkan pengelompokan anggota masing-masing *cluster*. Berdasarkan tabel 3, *cluster* yang pertama untuk parameter nilai mahasiswa memiliki 3 anggota yaitu mahasiswa ke- 4, 7, dan 10. Pusat *cluster* baru untuk *cluster* yang pertama dihitung berdasarkan rata-rata koordinat ketiga anggota tersebut adalah:

$$C1_{\text{passing grade}(\text{baru})} = (1.55 + 0 + 1.03)/3 = 0.86$$

- b. *Cluster* yang kedua untuk parameter nilai mahasiswa memiliki 3 anggota yaitu mahasiswa ke- 2, 5, 8. Pusat *cluster* baru untuk *cluster* yang kedua dihitung berdasarkan rata-rata koordinat lima anggota tersebut adalah :

$$C2_{\text{passgrade}(\text{baru})} = (0 + 0 + 7.54)/3 = 2.51$$

- c. *Cluster* yang ketiga untuk parameter nilai mahasiswa memiliki 4 anggota yaitu mahasiswa ke- 1, 3, dan 7 dan 9. Pusat *cluster* baru untuk *cluster* yang ketiga dihitung berdasarkan rata-rata koordinat ke 3 anggota tersebut adalah :

$$C3_{\text{passgrade}(\text{baru})} = (6.55 + 1.75 + 4.66 + 0)/4 = 3.24$$

- d. Pusat *cluster* baru untuk passing grade adalah :

$$C1_{\text{passgrade}(\text{baru})} = 0.86 * 10 = 8.6$$

$$C2_{\text{passgrade}(\text{baru})} = 2.51 * 10 = 25.1$$

$$C3_{\text{passgrade}(\text{baru})} = 3.24 * 10 = 32.4$$

- e. Dari hasil perhitungan diatas, maka kelompok kelas mahasiswa baru adalah sebagai berikut :

Tabel 4. Hasil Pengelompokan Kelas Mahasiswa

No	Nama Mahasiswa	Nilai Minimum	C1	C2	C3
1	Onika Sihotang	45.5	*	**	***
2	Riski Tambuna	43.75	*	**	***
3	Eva Feronika	42	*	**	***
4	Vini Fionika	39	*	**	***
5	Srianta Oktabria	32	*	**	
6	Mitra Arjuna s	26	*	**	
7	Rotua Verawati	26	*	**	
8	Revie Amilia	14	*		
9	Dewi Afrida	12	*		
10	Anita natalia	11.75	*		

Ket : C1 = Mahasiswa yang memiliki nilai > 8.6 berada di Cluster Pertama.

C2= Mahasiswa yang memiliki nilai > 25.1 berada di Cluster Kedua.

C3 = Mahasiswa yang memiliki nilai >32.4 berada di Cluster Ketiga.

SIMPULAN

Dengan menerapkan *Metode Clustering* dalam pemeriksaan hasil ujian saringan akan mempermudah dan mempercepat kegiatan penentuan dan memberikan informasi secara akurat dan tepat mengenai pengelompokan

kelas mahasiswa. Pengembangan sistem ini dapat di lakukan dengan perancangan *user interface* yang *friendly* dan untuk penelitian selanjutnya dapat mengembangkan dengan menambahkan aktor di dalam sistem sesuai dengan kebutuhannya.

DAFTAR PUSTAKA

- [1] Hermawati, Astuti, Fajar.(2013). Data mining.(Edisi 1). Yogyakarta: Andi.
- [2] Sani Susanto, Ph. D. & Dedy Suryadi, ST., M.S.(2010). Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data. Yogyakarta : Penerbit ANDI
- [3] Sari, Venny Novita, Yupianti Yupianti, and Dewi Maharani. "Penerapan Metode K-Means Clustering Dalam Menentukan Predikat Kelulusan Mahasiswa Untuk Menganalisa Kualitas Lulusan." *JURTEKSI* 4.2 (2018): 133-140.
- [4] Sibuea, Mustika Larasati, and Andy Safta. "Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering." *JURTEKSI* 4.1 (2017): 85-92.