

KLASIFIKASI PROSES BUSINESS DATA MAHASISWA UNIVERSITAS KANJURUHAN MALANG MENGGUNAKAN TEKNIK DATA MINING

Moh Ahsan
Universitas Kanjuruhan Malang
ahsan@unikama.ac.id

ABSTRAK. Universitas Kanjuruhan Malang adalah salah satu universitas swasta yang menyelenggarakan proses perkuliahan. Salah satu hal yang terpenting dalam proses penyelenggaraan perkuliahan adalah element masyarakat yang dituju dalam hal ini adalah mahasiswa. Bagaimana menganalisa data mahasiswa yang telah terkumpul sampai saat ini untuk menjadikan sebuah hasil yang dapat bermanfaat dikemudian hari. Kegunaan menganalisa data mahasiswa tersebut dapat dijadikan bahan pertimbangan pula bagi fakultas dan prodi untuk mendapat jatah promosi dalam memperoleh mahasiswa baru. Penerapan data mining dapat membantu menganalisa data mahasiswa yang diperoleh dari bagian tiap prodi. Metode yang digunakan yakni dengan clustering atau proses pengelompokan. Algoritma yang digunakan adalah metode K-Means, Informasi yang ditampilkan berupa nilai centroid dari tiap cluster dan kelompok fakultas yang layak mendapatkan promosi beserta sasaran sekolahnya.

Kata Kunci: *Data Mining; Metode Clustering; Algoritma K-Means; Data Mahasiswa.*

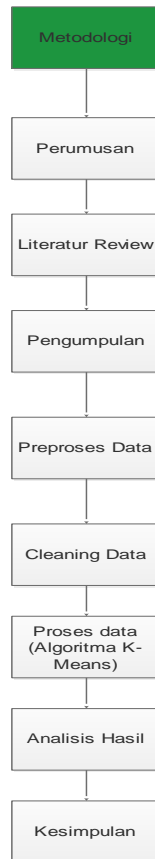
PENDAHULUAN

Bertambah atau berkurangnya mahasiswa setiap tahunnya yang mendaftar ke universitas membuat pengolahan data mahasiswa perlu melakukan yang berguna untuk mengetahui informasi penting berupa pengetahuan baru (*Knowledge Discovery*), misalnya informasi mengenai pengklasifikasian data mahasiswa berdasarkan data akademik. Terdapat banyak informasi yang tersembunyi dalam data mahasiswa diantaranya prediksi banyaknya mahasiswa yang akan datang, prediksi kelulusan mahasiswa tepat waktu atau tidak, estimasi waktu tempuh studi mahasiswa dan lain sebagainya. Hal tersebut perlu melakukan pengolahan data mahasiswa yang akan berguna bagi pihak Universitas. Pengetahuan baru tersebut dapat membantu pihak universitas untuk melakukan klasifikasi jumlah mahasiswa yang berasal dari berbagai daerah sebagai pendukung proses business dalam rangka memperoleh mahasiswa baru untuk tahun yang akan datang dengan tujuan untuk menentukan strategi promosi memperoleh mahasiswa baru untuk tahun berikutnya.

Jumlah mahasiswa Universitas Kanjuruhan Malang (UNIKAMA) mengalami peningkatan pada tahun 2013 sebanyak 2120 mahasiswa baru dan pada tahun 2014 bertambah menjadi 2342. Pada tahun 2015 jumlah mahasiswa baru mengalami penurunan dari sebelumnya yang berjumlah 2001 mahasiswa baru. Banyaknya mahasiswa menimbulkan penumpukan terhadap data mahasiswa sehingga mempengaruhi pencarian informasi terhadap data tersebut. Penelitian ini bertujuan untuk melakukan klasifikasi terhadap data mahasiswa Universitas Kanjuruhan Malang angkatan 2015 strata satu (S1) dengan memanfaatkan proses data mining dengan menggunakan teknik clustering. Metode yang digunakan adalah K-Means dengan melalui proses *business understanding*, data *understanding*, data *preparation*, *modeling*, *evaluation* dan *deployment*. Algoritma yang digunakan untuk clustering adalah K-Means. Algoritma K-Means akan mengelompokan data – data yang memiliki jarak antar pusat cluster. Semakin kecil jarak centroid dengan pusat cluster maka data termasuk dalam cluster tersebut. Atribut data yang digunakan untuk membantu menemukan nilai yang akurat meliputi adalah NIM, Nama, Jenjang, Progd, Provinsi Asal, Jenis Kelamin, SKS, IPK, dan Tahun Lulus. Hasil dari penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan kebijakan oleh Fakultas.

METODE PENELITIAN

Langkah-langkah yang dilakukan yaitu: perumusan masalah, penentuan teknik *clustering* yang akan dipergunakan, preproses data, transformasi data dengan teknik *clustering*, analisa hasil *clustering*, dan penarikan kesimpulan. Berikut digambarkan diagram tahapan penelitian yang digunakan.



Gambar 1. Diagram tahapan dalam penelitian

HASIL DAN PEMBAHASAN

Data yang di ambil yaitu data dari mahasiswa Universitas Kanjuruhan Malang untuk melihat pola pengelompokan serta mendapatkan hasil untuk bahan promosi fakultas yang terbaik.

1. Preproses Data
Preproses data merupakan tahapan pengumpulan data – data mahasiswa yang akan di inputkan kedalam sistem. Data berupa format excel yang berisikan seluruh data mengenai mahasiswa.
2. Cleaning Data
Proses cleaning data yaitu membersihkan data – data mahasiswa yang kurang valid informasinya seperti alamat, atau asal sekolah mahasiswa yang tidak terisi dalam file data mahasiswa. Perlunya cleaning data dalam datamining agar data yang masuk ke dalam sistem merupakan data yang valid dan benar – benar bisa di pertanggungjawabkan isi dan keabsahannya.
3. Proses Data (Algoritma K-Means)
Proses data yaitu mengolah data yang telah masuk kedalam database untuk dijadikan bahan olah guna menentukan pola data dari data mahasiswa sehingga pihak universitas dapat menentukan promosi pada fakultas mana dan sekolah mana yang akan dituju. Pengolahan data menggunakan beberapa langkah metode K-Means. Berikut langkah - langkahnya :
 - A. Pilih K buah titik *centroid* secara acak
 $K1 = 2.5 ; 50$
 $K2 = 3.85 ; 180$

$$K3 = 2.75 ; 112$$

Pusat cluster

	CLUSTER	IPK	SKS_TEMPUH
▶	1	2.5	50
	2	3.85	150
	3	2.75	112
*			

Cluster :

IPK :

SKS tempuh :

Gambar 2. Memilih titik *Centroid*

- B. Kelompokkan data sehingga terbentuk K buah *cluster* dengan titik *centroid* dari setiap *cluster*, merupakan titik *centroid* yang telah dipilih sebelumnya.

$$d_{Euclidean}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots\dots$$

$$D11 = \sqrt{(2.68 - 2.5)^2 + (112 - 50)^2}$$

$$= 62.00$$

$$D12 = \sqrt{(2.68 - 3.85)^2 + (112 - 150)^2}$$

$$= 38.02$$

$$D13 = \sqrt{(2.68 - 2.75)^2 + (112 - 112)^2}$$

$$= 0.07$$

Dari hasil perhitungan data mahasiswa pertama dengan tiap pusat cluster maka dapat dikatakan bahwa data mahasiswa pertama tergolong dalam cluster ke 3 karena jarak perhitungannya yang paling terkecil. Berikut seluruh data hasil perhitungan awal :

Tabel 1. Data Mahasiswa dalam cluster yang ke 3

ID	NPM	C1	C2	C3	TERMASUK_CLUSTER
1	110302010006	62	38.02	0.07	3
2	110302010014	68.01	32	6.04	3
3	120302020008	64	36.01	2.02	3
4	120302020012	68	32.01	6.01	3
5	126302020001	30	70.02	32	1
6	130302020003	48.01	52	14.01	3
7	140302020001	14.06	86	48.01	1
8	140302020002	14.01	86	48	1
9	140302020003	14.02	86	48	1
10	140302020004	14.01	86	48	1
11	140302020005	14.03	86	48	1
12	140302020006	14.04	86	48.01	1
13	140302020007	14.05	86	48.01	1
14	140302020008	14.02	86	48	1
15	140302020009	8.07	92	54.01	1
16	110303020004	65.01	35	3.07	3
17	120303010003	66.01	34	4.09	3
18	120303010004	66.01	34	4.08	3
19	120303010009	66.01	34	4.08	3
20	130303010001	44	56.02	18.01	3
21	130303010002	54.01	46	8.03	3
22	130303010006	54.01	46	8.03	3
23	130303010007	54.01	46	8.04	3
24	130303010008	54.01	46	8.03	3
25	130303010009	43	57.01	19	3
26	130303010012	47.01	53.04	15.04	3
27	110404020022	74	26.04	12	3
28	110404020024	98	2.13	36	2
29	110404020036	98	2.11	36	2
30	110404020049	98	2.07	36	2

C. Perbaharui nilai titik *centroid*.

$$C11 = \sum_{n=1}^{\infty} \left(\frac{\text{nilai ipk } c1}{\text{jumlah data } c1} \right)$$

$$C12 = \sum_{n=1}^{\infty} \left(\frac{\text{nilai sks_tempuh } c1}{\text{jumlah data } c1} \right)$$

$$C21 = \sum_{n=1}^{\infty} \left(\frac{\text{nilai ipk } c2}{\text{jumlah data } c2} \right)$$

$$C22 = \sum_{n=1}^{\infty} \left(\frac{\text{nilai sks_tempuh } c2}{\text{jumlah data } c2} \right)$$

$$C31 = \sum_{n=1}^{\infty} \left(\frac{\text{nilai ipk } c3}{\text{jumlah data } c3} \right)$$

$$C32 = \sum_{n=1}^{\infty} \left(\frac{\text{nilai sks_tempuh } c3}{\text{jumlah data } c2} \right)$$

Titik Pusat Cluster :

	CLUSTER	IPK	SKS_TEMPUH
▶	1	3.2	41.94
	2	3.15	150.89
	3	3.14	110.21
*			

Gambar 3. Pembaharuan nilai titik *Centroid*

- D. Ulangi langkah 2 dan 3 sampai nilai dari titik *centroid* tidak lagi berubah.
Perhitungan ke-1 :

Rasio iterasi data

Rasio iterasi

	ID	NPM	C1	C2	C3	TERMASUK_CLUSTER	PERHITUNGAN_KE_
▶	1	110302010006	70.06	38.89	1.85	3	1
	2	110302010014	76.06	32.89	7.8	3	1
	3	120302020008	72.06	36.89	3.79	3	1
	4	120302020012	76.06	32.89	7.79	3	1
	5	126302020001	38.07	70.89	30.22	3	1
	6	130302020003	56.06	52.89	12.21	3	1
	7	140302020001	22.07	86.89	46.22	1	1
	8	140302020002	22.06	86.89	46.21	1	1
	9	140302020003	22.06	86.89	46.21	1	1

Total Cluster 1 :

Total Cluster 2 :

Total Cluster 3 :

Titik Pusat Cluster :

	CLUSTER	IPK	SKS_TEMPUH
▶	1	3.2	41.86
	2	3.15	150.85
	3	3.14	109.98
*			

Jumlah iterasi:

Gambar 4. Mengulangi memilih titik *Centroid*

Perhitungan ke – 2 :

Rasio iterasi data

Rasio iterasi

ID	NPM	C1	C2	C3	TERMASUK_CLUSTER	PERHITUNGAN_KE_
8145	110302010006	70.14	38.85	2.07	3	2
8146	110302010014	76.14	32.85	8.03	3	2
8147	120302020008	72.14	36.85	4.02	3	2
8148	120302020012	76.14	32.85	8.02	3	2
8149	126302020001	38.15	70.86	29.99	3	2
8150	130302020003	56.14	52.85	11.98	3	2
8151	140302020001	22.15	86.85	45.98	1	2
8152	140302020002	22.14	86.85	45.98	1	2
8153	140302020003	22.14	86.85	45.98	1	2

Total Cluster 1 : 4415
 Total Cluster 2 : 1961
 Total Cluster 3 : 1768

Titik Pusat Cluster :

CLUSTER	IPK	SKS_TEMPUH
1	3.2	41.84
2	3.15	150.85
3	3.14	109.95

Jumlah iterasi: 2
 Lihat iterasi sebelumnya
 Lihat iterasi setelahnya

Gambar 5. Mengulangi memilih titik *Centroid*

Perhitungan ke – 3 :

Rasio iterasi data

Rasio iterasi

ID	NPM	C1	C2	C3	TERMASUK_CLUSTER	PERHITUNGAN_KE_
16289	110302010006	70.16	38.85	2.1	3	3
16290	110302010014	76.16	32.85	8.05	3	3
16291	120302020008	72.16	36.85	4.05	3	3
16292	120302020012	76.16	32.85	8.05	3	3
16293	126302020001	38.17	70.86	29.96	3	3
16294	130302020003	56.16	52.85	11.95	3	3
16295	140302020001	22.17	86.85	45.95	1	3
16296	140302020002	22.16	86.85	45.95	1	3
16297	140302020003	22.16	86.85	45.95	1	3

Total Cluster 1 : 4415
 Total Cluster 2 : 1961
 Total Cluster 3 : 1768

Titik Pusat Cluster :

CLUSTER	IPK	SKS_TEMPUH
1	3.2	41.84
2	3.15	150.85
3	3.14	109.93

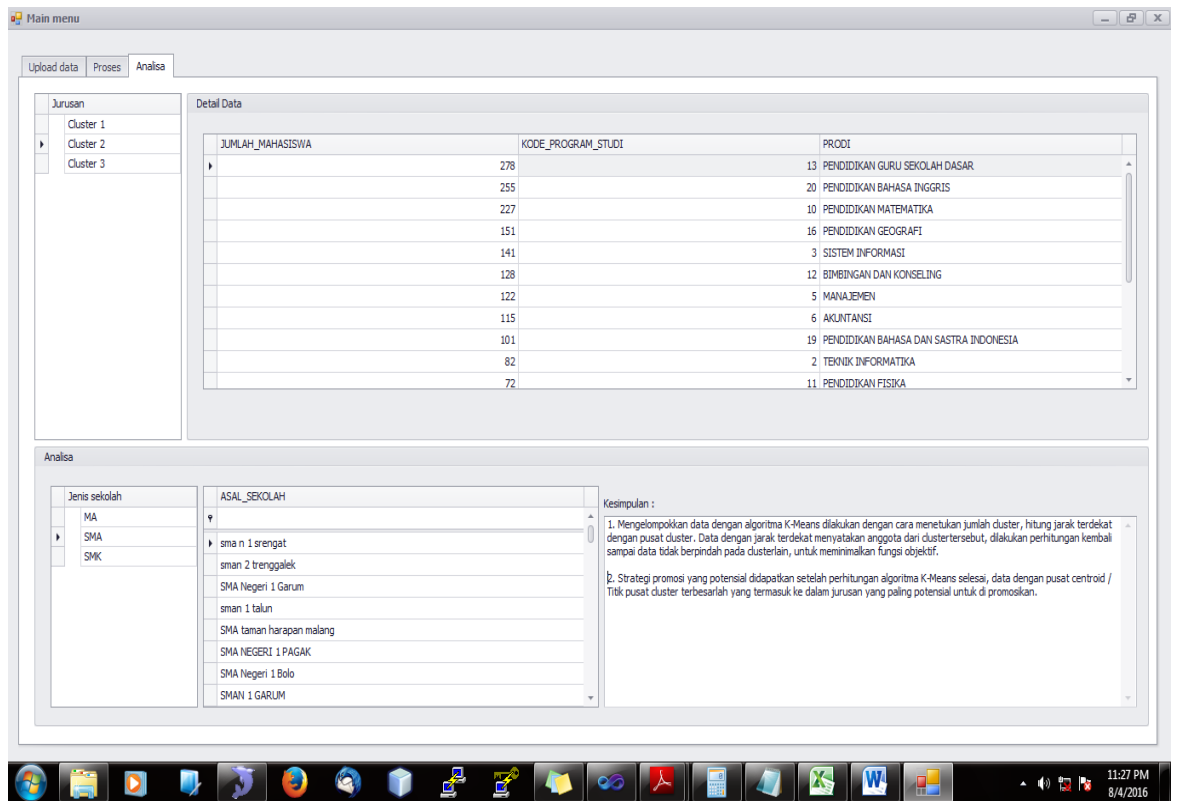
Jumlah iterasi: 3
 Lihat iterasi sebelumnya
 Lihat iterasi setelahnya

Gambar 6. Mengulangi memilih titik *Centroid*

Perhitungan hanya sampai ketiga dikarenakan titik pusat cluster sudah tidak berubah dan data sudah tidak ada yang berpindah cluster lagi.

4. Analisis

Analisis merupakan tahapan sistem dalam menampilkan hasil dari perhitungan dari algoritma K-Means.



Gambar 7. Hasil analisis perhitungan K-Means

KESIMPULAN

Dari penelitian yang telah dilakukan, maka dapat ditarik kesimpulan sebagai berikut : Mengelompokkan data dengan algoritma K-Means dilakukan dengan cara menentukan jumlah *cluster*, hitung jarak terdekat dengan pusat *cluster*. Data dengan jarak terdekat menyatakan anggota dari *cluster* tersebut, dilakukan perhitungan kembali sampai data tidak berpindah pada *cluster* lain, untuk meminimalkan fungsi objektif. Data pelanggan yang potensial didapatkan setelah perhitungan algoritma K-Means selesai, data dengan pusat *centroid* terbesar yang termasuk ke dalam fakultas/jurusan yang paling potensial untuk diberikan sasaran promosi.

DAFTAR PUSTAKA

- Azwar. (2004). "Penyusunan Skala Psikologi". Yogyakarta: Pustaka Pelajar.
- Arief Jananto, "Algoritma Naïve Bayes Untuk Mencari Perkiraan Waktu Studi Mahasiswa" Jurnal Teknologi Informasi DINAMIK, vol 18, no.1, Januari 2013.
- Afrisawati. "Implementasi data mining pemilihan pelanggan potensial menggunakan algoritma K-Means". vol 5, no.3, Desember 2013
- Eko Prasetyo, "Data Mining : Konsep dan Aplikasi menggunakan MATLAB", 1st ed. Yogyakarta, Indonesia: Andi, 2012
- Ian H. Witten, f. E. (2011). Data Mining: "Practical Machine Learning Tools and Techniques" (3 ed.). (A. S. Burlington, Ed.) United States of America: Morgan Kaufmann.
- Larose, Daniel T, "Data Mining Methods and Models". Hoboken New Jersey : Jhon Wiley & Sons, Inc, 2006.

Marselina S.S, Ernastuti, “*Graduation Prediction Of Gunadarma University Students Using Algorithm Naïve Bayes C4.5 Algorithm,*” Faculty Of Industrial Engineering, 2010

John F.S, “Data Mining Classification Untuk Prediksi Lama Masa Studi Mahasiswa Berdasarkan Jalur Penerimaan Dengan Metode Naïve Bayes,” Magister Teknik Informatika Universitas Atma Jaya Yogyakarta.

Santoso, B. (2007). “Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis” (1 ed.). Yogyakarta: Graha Ilmu.

Quinlan, J. (1993). C4.5: “*Programs for machine learning*”. Morgan Kaufmann.