

Model Rule Identifikasi Atribut Yang Dapat Digunakan dalam Memprediksi Prestasi Akademik Mahasiswa dengan Teknik Kernel K-Means Clustering

Afen Prana Utama Sembiring

Program Studi S-1 Sistem Informasi, Universitas Mikroskil

E-mail: afen@mikroskil.ac.id

Info Artikel	Abstrak
Article History: Received: 05 Aug 2022 Revised: 15 Aug 2022 Accepted: 25 Aug 2022	<p><i>Usulan membuat model rule identifikasi atribut dalam memprediksi prestasi akademik mahasiswa menjadi satu hal menarik di institusi pendidikan tinggi. Untuk itu sangat penting dibuat sebuah model rule yang dapat digunakan pihak pengambilan keputusan atau membuat sebuah kebijakan. Dalam penelitian ini, penulis menggunakan algoritma Kernel K-Means Clustering yang telah digunakan untuk mendapatkan suatu model aturan dalam prediksi prestasi akademik mahasiswa di program studi manajemen Universitas Mikroskil. Model aturan yang diperoleh menunjukkan bahwa predikat dengan pujian dapat diperoleh jika nilai rata-rata teori, nilai rata-rata praktik dan kehadiran semakin tinggi.</i></p>
Keywords: <i>Model Rule, Kernel K-Means, Clustering, Prestasi</i>	

1. PENDAHULUAN

Informasi Saat ini sangat dibutuhkan dalam kehidupan sehari-hari. Dengan adanya kemajuan Teknologi Informasi organisasi dapat mengumpulkan dan menyimpan berbagai jenis data yang jumlahnya sangat besar. Perguruan Tinggi Negeri maupun swasta (PTN/PTS) telah memiliki data base sebagai sistem informasi yang dapat berupa data akademik, data mahasiswa, data sumber daya manusia, inventaris, keuangan dan sebagainya. Data sedemikian banyak dan tersimpan hanya bersifat historis saja.

Organisasi juga harus memiliki kemampuan untuk menganalisis, meringkas dan mengekstrak data yang dimiliki menjadi pengetahuan. Metode tradisional dalam hal menganalisis data yang ada, tidak dapat menangani sejumlah data yang jumlahnya besar. Pembuat keputusan, dalam hal ini Ketua program studi berusaha untuk memanfaatkan kumpulan data yang sudah dimiliki untuk menambang informasi yang berguna dalam mengambil keputusan. Hal ini mendorong munculnya cabang ilmu baru untuk mengatasi masalah penggalian informasi yang penting dari kumpulan data, yang disebut dengan *data mining*.

Data mining sering juga disebut *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari *data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan[1].

Kemampuan untuk membuat model *rule* memprediksi prestasi akademik mahasiswa sangat penting dalam institusi pendidikan. Dalam penelitian ini, digunakan teknik *data mining* yaitu algoritma Kernel K-Means *Clustering* untuk membuat model rule identifikasi atribut yang dapat digunakan Dalam Memprediksi Prestasi Akademik mahasiswa Dengan Teknik Kernel K-Means

Clustering. Data yang digunakan dalam penelitian ini adalah data mahasiswa pada program studi Manajemen Universitas Mikroskil Medan.

Penelitian ini dibatasi hanya pada analisis data mahasiswa pada data akademik maupun biodatanya, dan teknik yang digunakan Kernel K-Means.

2. KAJIAN PUSTAKA

2.1. *Data Mining : Knowledge Discovery Databases (KDD)*

Data mining adalah serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual [2]. *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 [3]. *Data mining* sering juga disebut *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari *data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan [1]. *Data mining* adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. *Data mining* berkaitan dengan bidang ilmu-ilmu lain, seperti *database system*, *data warehousing*, statistik, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, *data mining* didukung oleh ilmu lain seperti *neural network*, pengenalan pola, *spatial data analysis*, *image database*, *signal processing* [4]. *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 [5].

Masalah-masalah yang sesuai untuk diselesaikan dengan teknik *data mining* dapat dicirikan dengan [6]:

1. Memerlukan keputusan yang bersifat *knowlegde-based*.
 2. Mempunyai lingkungan yang berubah.
 3. Metode yang ada sekarang bersifat sub-optimal.
 4. Tersedia data yang bisa diakses, cukup dan relevan.
- Memberikan keuntungan yang tinggi jika keputusan yang ambil tepat.

2.2. *Tahapan Data Mining*

Untuk menemukan pola yang menarik dalam jumlah data yang banyak dibutuhkan metode analisis sehingga *data mining* berkaitan erat dengan bidang ilmu – ilmu lain, seperti *database system*, *data warehousing*, statistik, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Beberapa metode yang sering disebut-sebut dalam literatur *data mining* antara lain *clustering*, *classification*, *association rules mining*, *neural network*, *genetic algorithm* dan lain-lain[6].

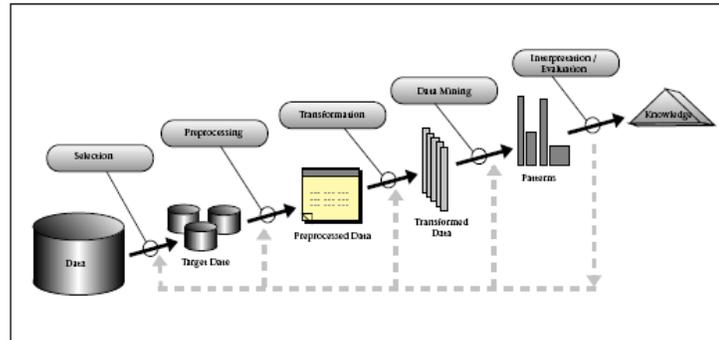
Menurut Davies ada tiga ciri atau karakteristik dari data mining, yaitu :

1. *Data mining* berkaitan dengan upaya menemukan sesuatu yang tersembunyi dan bentuk pola data tidak diketahui sebelumnya.
2. *Data mining* biasa menggunakan jumlah data yang sangat banyak untuk membuat hasil yang lebih dipercaya.
3. *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi[4].

2.3. *Tahap-Tahap Data mining*

Data yang ada, tidak dapat langsung diolah dengan menggunakan sistem *data mining*. Data tersebut harus dipersiapkan terlebih dahulu agar hasil yang diperoleh dapat lebih maksimal, dan

waktu komputasinya lebih minimal. Proses persiapan data ini sendiri dapat mencapai 60 % dari keseluruhan proses dalam *data mining*. Proses KDD secara garis besar dapat dijelaskan sebagai berikut [8].



Gambar 1 *Data Mining* : Proses KDD

Menurut Kusriani [9] proses KDD dapat diuraikan sebagai berikut :

- 1) Seleksi Data (*Data Selection*)
Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari *database* operasional.
- 2) Pra-pemrosesan / Pembersihan (*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 *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.
- 3) Transformasi (*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 *database*.
- 4) *Data mining*
Data mining adalah proses mencari pola atau informasi menarik dalam data yang 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) Interpretasi / Evaluasi (*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.

2.4. Algoritma Clustering (*Clustering Algorithm*)

Clustering (pengelompokan data) mempertimbangkan sebuah pendekatan penting untuk mencari kesamaan dalam data dan menempatkan data yang sama ke dalam kelompok-kelompok. *Clustering* membagi kumpulan data ke dalam beberapa kelompok dimana kesamaan dalam sebuah kelompok adalah lebih besar daripada diantara kelompok-kelompok [10].

Algoritma-algoritma *clustering* digunakan secara ekstensif tidak hanya untuk mengorganisasikan dan mengategorikan data, akan tetapi juga sangat bermanfaat untuk kompresi data dan konstruksi model. Melalui pencarian kesamaan dalam data, seseorang dapat merepresentasikan data yang sama dengan lebih sedikit simbol misalnya. Juga, jika kita dapat menemukan kelompok-kelompok data, kita dapat membangun sebuah model masalah berdasarkan pengelompokan-pengelompokan ini [12].

Clustering menunjuk pada pengelompokan *record*, observasi-observasi, atau kasus-kasus ke dalam kelas-kelas objek yang sama. *Cluster* adalah sekumpulan *record* yang adalah sama dengan satu sama lain dan tidak sama dengan *record* dalam *cluster* lain. *Clustering* berbeda dari klasifikasi dimana tidak ada variabel target untuk *clustering*. Tugas *clustering* tidak mencoba untuk mengklasifikasikan, mengestimasi, atau memprediksi nilai variabel target [3]. Bahkan, algoritma *clustering* berusaha menyegmentasikan seluruh kumpulan data ke dalam subkelompok-subkelompok atau *cluster-cluster* homogen secara relatif. Dimana kesamaan *record* dalam *cluster* dimaksimalkan dan kesamaan dengan *record* diluar *cluster* ini diminimalkan.

Clustering sering dilaksanakan sebagai langkah pendahuluan dalam proses pengumpulan data, dengan *cluster-cluster* yang dihasilkan digunakan sebagai input lebih lanjut ke dalam sebuah teknik yang berbeda, seperti *neural network*. Karena ukuran yang besar dari banyak *database* yang dipresentasikan saat ini, maka sering sangat membantu untuk menggunakan analisa *clustering* terlebih dahulu, untuk mengurangi ruang pencarian untuk algoritma-algoritma *downstream*. Aktivitas *clustering* pola khusus meliputi langkah-langkah berikut [12]:

- a. Representasi pola (secara opsional termasuk ekstraksi dan/atau seleksi sifat)
- b. Definisi ukuran kedekatan pola yang tepat untuk domain data
- c. *Clustering* pengelompokan
- d. Penarikan data (jika dibutuhkan), dan
- e. Pengkajian output (jika dibutuhkan).

2.5. Metode Kernel (*Kernel Methods*)

K-means adalah algoritma *unsupervised learning* yang membagi kumpulan data ke dalam sejumlah *cluster* yang dipilih dibawah beberapa ukuran-ukuran optimisasi. Sebagai contoh, kita sering ingin meminimalkan jumlah kuadrat dari jarak *Euclidean* antara sampel dan centroid. Asumsi di belakang ukuran ini adalah keyakinan bahwa ruang data terdiri dari daerah elliptical yang terisolasi. Meskipun demikian, asumsi tersebut tidak selalu ada pada aplikasi spesifik. Untuk menyelesaikan masalah ini, sebuah gagasan meneliti ukuran-ukuran lain, misalnya kesamaan kosinus yang digunakan dalam pencarian informasi. Gagasan lain adalah memetakan data pada ruang baru yang memenuhi persyaratan untuk ukuran optimisasi. Dalam hal ini, fungsi kernel merupakan pilihan yang baik.

Algoritma 1

Langkah 1 Menginisialisasikan pusat cluster, $c_i, i=1, \dots, c$.

Ini biasanya dilakukan melalui pemilihan secara acak point-point c diantara semua point-point data.

Langkah 2 Menentukan matriks keanggotaan U

Langkah 3 Menghitung fungsi biaya. Hentikan jika berada dibawah nilai toleransi tertentu atau perbaikannya terhadap iterasi sebelumnya adalah dibawah batas ambang tertentu.

Langkah 4 Perbaharui pusat-pusat cluster. Lanjutkan ke langkah 2.

Walaupun dapat dibuktikan bahwa prosedur tersebut akan selalu berakhir, algoritma k-means tidak perlu mencari konfigurasi yang paling optimal, yang sesuai dengan minimum fungsi

objektif global. Algoritma ini juga secara signifikan sensitif terhadap pusat-pusat *cluster* yang dipilih secara acak pada awalnya. Algoritma k-means dapat dijalankan beberapa kali untuk mengurangi efek ini.

Algoritma 2

Langkah 1 Pilih awal pusat $K: m_1, m_2, \dots, m_K$

Langkah 2 Menentukan setiap sample $x_i (1 \leq i \leq N)$ pada pusat terdekat, yang membentuk cluster K . Yaitu, menghitung nilai fungsi indikator $\delta(x_i, C_k), (1 \leq k \leq K)$.

$$1, D(x_i, m_k) < D(x_i, m_j) \text{ for all } j \neq k, \text{ otherwise } 0$$

Langkah 3 Hitunglah pusat baru m_k untuk setiap cluster C_k

$$m_k = \frac{1}{|C_k|} \sum_{i=1}^N \delta(x_i, C_k) x_i$$

Dimana $|C_k|$ adalah jumlah sampel dalam C_k

$$|C_k| = \sum_{i=1}^N \delta(x_i, C_k)$$

Langkah 4 Ulangi langkah 2 dan 3 hingga bertemu.

Langkah 5 Menghasilkan $m_k (1 \leq k \leq K)$

Isu utama yang memperluas ke adalah penghitungan jarak dalam ruang baru. Anggaplah $u = \Phi(x)$ dan u : -means tradisional kernel k-means adalah i menunjukkan transformasi i . Jarak Euclidean antara i

$$\begin{aligned} D^2(u_i, u_j) &= \|\Phi(x_i) - \Phi(x_j)\|^2 \\ &= \Phi^2(x_i) - 2\Phi(x_i)\Phi(x_j) + \Phi^2(x_j) \\ &= H(x_i, x_i) - 2H(x_i, x_j) + H(x_j, x_j) \end{aligned} \quad (2.15)$$

Anggaplah z_k adalah pusat *cluster* dalam ruang yang ditransformasikan dimana,

$$z_k = \frac{1}{|C_k|} \sum_{i=1}^N \delta(x_i, C_k) u_i \quad (2.16)$$

Dimana $\delta(x_i, C_k)$ adalah fungsi indikator. Jarak antara u_i dan z_k dinyatakan sebagai berikut:

$$D^2(u_i, z_k) = \|u_i - \frac{1}{|C_k|} \sum_{j=1}^N \delta(x_j, C_k) u_j\|^2 \text{ ditulis sebagai } = H(x_i, x_i) + f(x_i, C_k) + g(C_k) \quad (2.17)$$

Dimana,

$$f(x_i, C_k) = 2 \sum_{j=1}^N \delta(x_j, C_k) H(x_i, x_j) \quad (2.18)$$

$$g(C_k) = 2 \sum_{j=1}^N \delta(x_j, C_k) H(x_j, x_j)$$

Perbedaan utama antara kernel k-means dalam algoritma Kernel K-means yang tradisional berdasarkan algoritma : k-means dengan versi tradisional ada di langkah 5, Karena *cluster* dalam ruang transformasikan tidak dapat dinyatakan secara eksplisit harus memilih pseudo centre Dengan menggunakan k-means, diperoleh kernel

Algoritma 3

Langkah 1 Tentukan $(x, C_k) (1 \leq i \leq N, 1 \leq k \leq K)$ dengan nilai awal, yang membentuk cluster initial $K C_1, C_2, \dots, C_K$

Langkah 2 Untuk setiap cluster C_k , hitunglah $|C_k|$ dan $g(C_k)$.

Langkah 3 Untuk setiap sample latihan x_i dan cluster C_k , hitunglah $f(x_i, C_k)$ dan kemudian tentukan x_i pada cluster terdekat. $\delta(x_i, C_k)$

$$1, (x_i, C_k) + g(C_k) < f(x_i, C_j) + g(C_j) \text{ for all } j \neq k, \text{ otherwise } 0$$

Langkah 4 Ulangi langkah 2 dan 3 hingga bertemu.

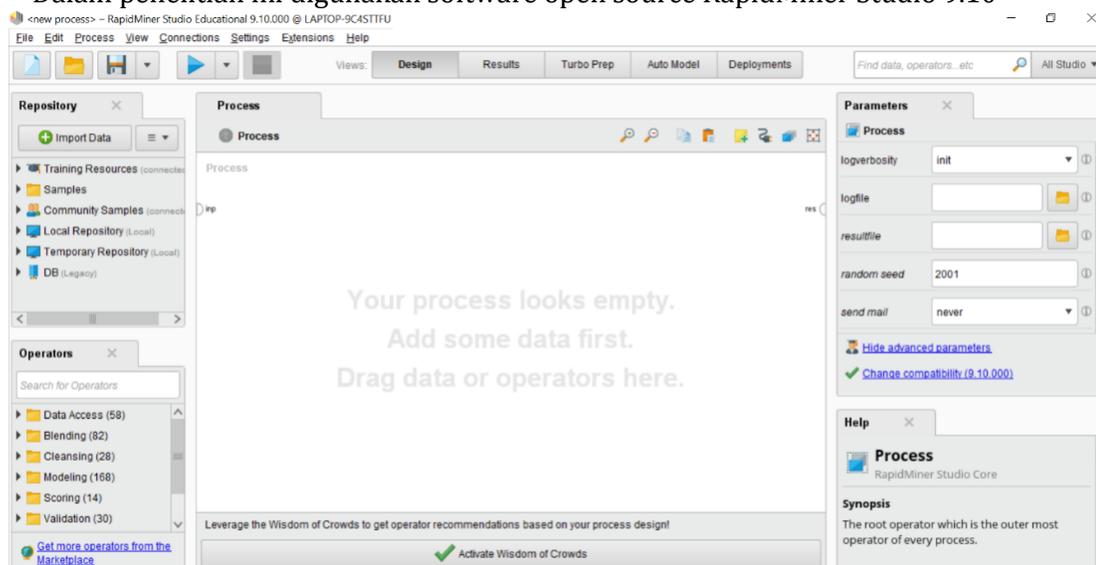
Langkah 5 Untuk setiap cluster C_k , pilih sample yang terdekat dengan pusat sebagai representatif dari C_k , $m_k = \text{Arg min } D(\Phi(x_i), z_k)$. X_i dimana $\delta(X_i, C_k) = 1$

faktor $H(x_i, x_i)$ diabaikan karena tidak berkontribusi untuk menentukan cluster terdekat.

Perbedaan utama antara kernel k-means dan versi tradisionalnya ada dalam langkah 5. K-Means sebagai berikut

2.6. Tools Analisis RapidMiner 9.10

Dalam penelitian ini digunakan software open source RapidMiner Studio 9.10



Gambar. 2 Tampilan RapidMiner

Gambar 2 merupakan tampilan RapidMiner Studio 9.10, yang digunakan pada penelitian ini untuk menampilkan cluster terhadap data dan memperlihatkan matrik yang tersebar pada tiap tiap kelompok. Algoritma Kernel K-Means clustering menggunakan alat-alat pengelompokan yang disediakan di perpustakaan Rapidminer.

3. METODE PENELITIAN

Langkah-langkah yang penulis lakukan dalam penelitian ini dapat diuraikan sebagai berikut:

1. Persiapan
2. Mengumpulkan data dari jurusan/program studi Manajemen. Data dikumpulkan dari database akademik mahasiswa angkatan 2018-2019.
3. Menyeleksi faktor penentu yang dapat dijadikan variabel penelitian atau faktor penentu yang diduga dominan berpengaruh, yakni: IPK, jenis kelamin mahasiswa, konsentrasi bidang minat, nilai rerata teori, nilai rerata praktek dan presensi kehadiran
4. Pengumpulan data dengan query data ke database SIPT
5. Seleksi data dilakukan untuk menyisihkan record yang tidak lengkap (pra pemrosesan data). Dilakukan untuk proses *cleaning* (pembersihan) menyisihkan record data yang tidak lengkap.
6. Analisis data. Analisis data dan pengujian dilakukan dengan menggunakan teknik Kernel K-Means Clustering.
7. Tools RapidMiner Studio 9.10, digunakan pada penelitian ini untuk menampilkan cluster terhadap data dan memperlihatkan matrik yang tersebar pada kelompok-kelompok.

8. Melakukan Interpretasi. Interpretasi dari hasil analisis data merupakan tahap yang sangat penting. Pada tahap ini akan dijelaskan *cluster* dari data yang akan dianalisis. Hasil analisis *cluster* untuk mendapatkan pengetahuan berupa model rule.

4. HASIL DAN PEMBAHASAN

4.1. Deskripsi Data

Data dalam penelitian ini adalah data mahasiswa tahun 2018-2019. Data tersebut diambil dari database SIPT Mikroskil dengan total records (baris) sebanyak 151 dan jumlah fields (kolom) sebanyak 74. Oleh karena tidak semua field dianggap relevan serta adanya field yang tidak lengkap sehingga dilakukan data cleaning sehingga jumlah field yang dipergunakan sebanyak 8 hal ini sesuai dengan tahapan ketiga pada tahapan data mining, yaitu seleksi data.

4.2. Analisis Data

Penelitian ini dimulai dengan mengumpulkan data dari prodi Manajemen Universitas Mikroskil Medan. Data dikumpulkan dari *database* akademik mahasiswa angkatan 2018-2019. variabel penelitian yang dikumpulkan adalah IPK selama 4 semester yakni: IPK, jenis kelamin mahasiswa, konsentrasi bidang minat, nilai rerata teori, nilai rerata praktek dan presensi kehadiran Data yang diperoleh dalam bentuk MDB dimodifikasi dan ditransformasi sehingga menjadi bentuk XLSX. Selanjutnya dilakukan pengujian terhadap data dengan menggunakan algoritma Kernel K-Means yang telah tersedia pada software *open source* RapidMiner Studio. Dari pengujian diperoleh *cluster* dari data yang telah diuji selanjutnya dilakukan analisis *cluster* untuk menganalisis dan mendapatkan model aturan.

Selanjutnya dipilih *field-field* yang akan dikategorisasi. *Field* yang akan dikategorisasi adalah IPK. IPK dikategorisasi menjadi 4 kategori yaitu, Dengan Pujian, Sangat Memuaskan, Memuaskan dan Buruk sebagaimana terlihat pada tabel 1 berikut :

Tabel 1 Kategorisasi IPK

No.	IPK	Predikat
1	< 2,00	Buruk
2	2,00 - 2,75	Memuaskan
3	2,76 - 3,50	Sangat Memuaskan
4	3,51 - 4,00	Dengan Pujian

Tabel 2 Kategorisasi Nilai Praktek dan Teori

No.	Praktek	Teori	Kategori
1	Praktek < 58	Teori < 58	Rendah
2	70 < Praktek < 58	70 < Teori < 58	Sedang
3	Praktek > 70	Teori > 70	Tinggi

Tabel 3. Kategorisasi Kehadiran

No.	KEHADIRAN	KATEGORI
1	Kehadiran > 93 %	Baik
2	86 % < Kehadiran < 93 %	Sedang
3	Kehadiran < 86 %	Kurang

Pada Kategorisasi nilai praktik dan teori dikategorikan menjadi 3 kategori yakni, kategori rendah, sedang dan tinggi, sebagaimana yang terlihat pada Tabel 2. Selanjutnya *field* yang akan dikategorisasi adalah GENDER. GENDER dikategorisasi menjadi 2 kategori yaitu Laki-laki dan Perempuan seperti pada tabel 3 berikut :

Tabel 3. Kategorisasi Gender

No.	Gender	Kategori
1	1	Laki-Laki
2	2	Perempuan

Selanjutnya *field* yang akan dikategorisasi adalah KEHADIRAN. KEHADIRAN dikategorisasi menjadi 3 kategori yaitu Baik, Sedang dan Kurang seperti pada tabel 4.

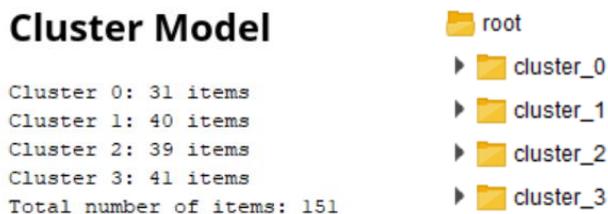
Tabel 4 Tampilan Kategorisasi Data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	PEMINATAN	RPRAKTEK	RTEORI	KEHADIRAN	KELAS	AGAMA	SKULIAH	IPK	GENDER	Peminatan	praktek	teori	hadir	Kelas	Agama	Shift Perkuliahan	Predikat	Jenis Kelamin
2	711	71	74	100	1	4	1	3.383	2	711	Tinggi	Tinggi	Baik	A	Islam	Pagi	Sangat Memuaskan	Perempuan
3	711	72	77	100	1	2	1	3.363	2	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Sangat Memuaskan	Perempuan
4	711	79	77	100	1	4	1	3.921	2	711	Tinggi	Tinggi	Baik	A	Islam	Pagi	Dengan Pujian	Perempuan
5	711	75	72	100	1	2	1	3.171	2	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Sangat Memuaskan	Perempuan
6	711	71	75	100	1	2	1	3.404	2	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Sangat Memuaskan	Perempuan
7	711	73	76	100	1	2	1	3.533	2	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Dengan Pujian	Perempuan
8	711	77	72	100	1	2	1	3.288	1	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Sangat Memuaskan	Laki-laki
9	711	73	72	100	1	2	1	3.525	1	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Dengan Pujian	Laki-laki
10	711	77	76	100	1	4	1	3.708	2	711	Tinggi	Tinggi	Baik	A	Islam	Pagi	Dengan Pujian	Perempuan
11	711	72	79	100	1	2	1	3.658	2	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Dengan Pujian	Perempuan
12	711	72	73	100	1	2	1	3.288	2	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Sangat Memuaskan	Perempuan
13	711	74	66	100	1	2	1	2.771	1	711	Tinggi	Sedang	Baik	A	Kristen Protestan	Pagi	Sangat Memuaskan	Laki-laki
14	711	71	72	100	1	3	1	3.458	2	711	Tinggi	Tinggi	Baik	A	Kristen Katholik	Pagi	Sangat Memuaskan	Perempuan
15	711	73	74	100	1	2	1	3.658	2	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Dengan Pujian	Perempuan
16	711	72	73	100	1	2	1	3.417	2	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Sangat Memuaskan	Perempuan
17	711	74	73	100	1	2	1	3.388	2	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Sangat Memuaskan	Perempuan
18	711	74	71	100	1	3	1	3.363	2	711	Tinggi	Tinggi	Baik	A	Kristen Katholik	Pagi	Sangat Memuaskan	Perempuan
19	711	71	74	100	1	2	1	3.4	2	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Sangat Memuaskan	Perempuan
20	711	72	75	100	1	2	1	3.321	2	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Sangat Memuaskan	Perempuan
21	711	75	75	100	1	3	1	3.863	2	711	Tinggi	Tinggi	Baik	A	Kristen Katholik	Pagi	Dengan Pujian	Perempuan
22	711	76	72	100	1	2	1	3.75	2	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Dengan Pujian	Perempuan
23	711	74	82	100	1	2	1	3.908	2	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Dengan Pujian	Perempuan
24	711	72	72	100	1	2	1	3.525	2	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Dengan Pujian	Perempuan
25	711	73	71	100	1	2	1	3.196	2	711	Tinggi	Tinggi	Baik	A	Kristen Protestan	Pagi	Sangat Memuaskan	Perempuan

Hasil kategorisasi data akan digunakan untuk input data pada *software open source* RapidMiner. Kemudian dilakukan pengujian terhadap data menggunakan algoritma Kernel K-Means. Dari pengujian diperoleh cluster dari data yang telah diuji tersebut dan selanjutnya dilakukan analisis *cluster* untuk menganalisis dan mendapatkan model aturan yang digambarkan dari hasil *cluster*.

Cluster Model

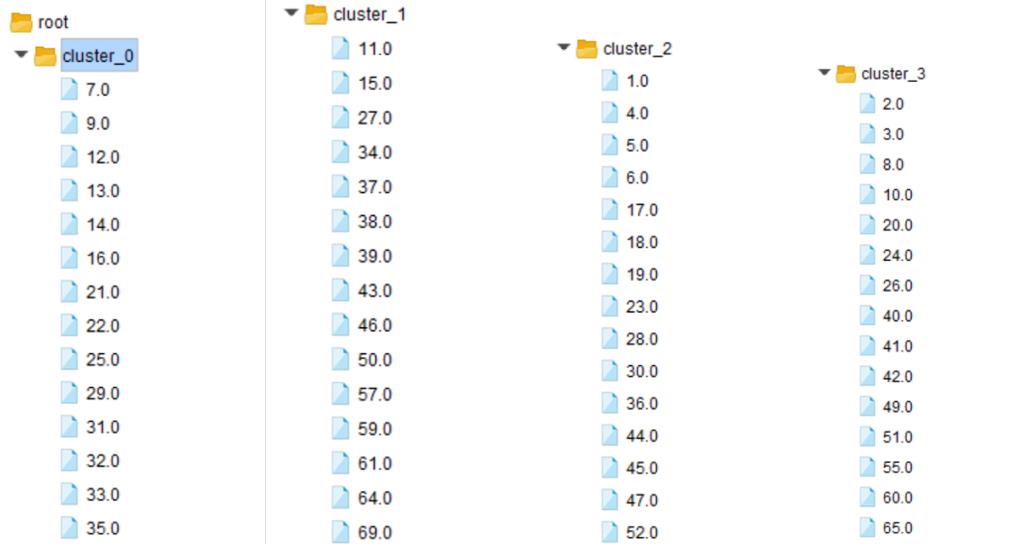
Cluster model yang diperoleh dari hasil pengujian terhadap data menggunakan metode Kernel K-Means seperti yang terlihat pada gambar 3 dibawah ini



Gambar 3 Cluster Model

Dari gambar 3 dapat dilihat *cluster* model yang dihasilkan terdiri dari *cluster* 0 sebanyak 31 item, *cluster* 1 sebanyak 40 item, *cluster* 2 sebanyak 39 item dan *cluster* 3 sebanyak 41 item dari total jumlah 151 item. Dalam bentuk folder root tampak bahwa root memiliki folder untuk 4 *cluster* yaitu

folder *cluster 0*, *cluster1*, *cluster 2* dan *cluster 3*. *Cluster* model juga dapat di gambar berupa sebuah *tree*. Induk adalah *root* yang memiliki *child* berupa 2, 3, 0 dan 1. Untuk memudahkan melihat anggota yang dimiliki oleh setiap folder *cluster*, gambar 4 menampilkan *membership* dari masing-masing *cluster*



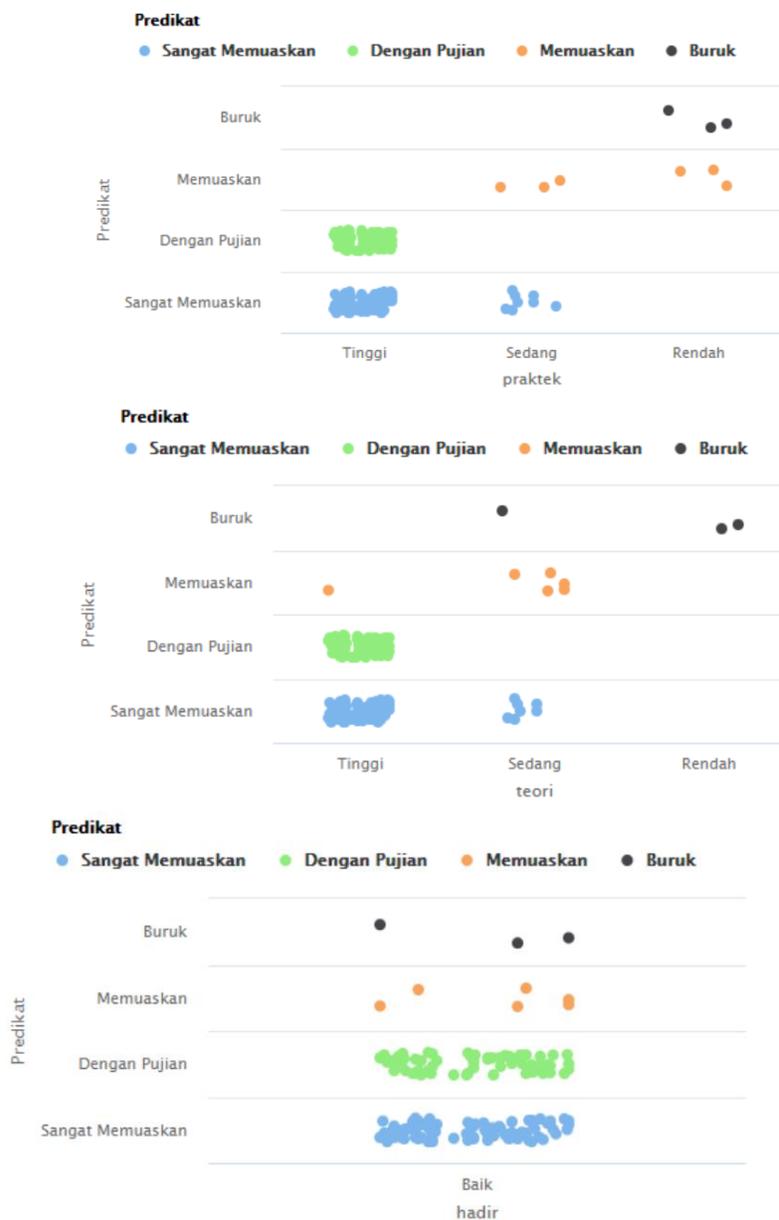
Gambar 4 Tampilan membership tiap cluster

Tabel 7 Hasil *Clustering* dalam Data View

Row No.	id	cluster	PEMINATAN	RPRAKTEK	RTEORI	KEHADIRAN	IPK	KELAS
1	1	cluster_2	711	71	74	100	3.383	1
2	2	cluster_3	711	70	77	100	3.362	1
3	3	cluster_3	711	79	88	100	3.921	1
4	4	cluster_2	711	75	72	100	3.171	1
5	5	cluster_2	711	71	75	100	3.404	1
6	6	cluster_2	711	73	76	100	3.533	1
7	7	cluster_0	711	77	72	100	3.288	1
8	8	cluster_3	711	73	72	100	3.525	1

Cluster Data Berdasarkan Predikat Prestasi Akademik

Cluster data dengan label predikat terhadap praktik, teori dan kehadiran, dapat dilihat pada gambar 5 berikut :



Gambar 5 Distribusi data antara predikat IPK dengan PRAKTEK, TEORI dan KEHADIRAN

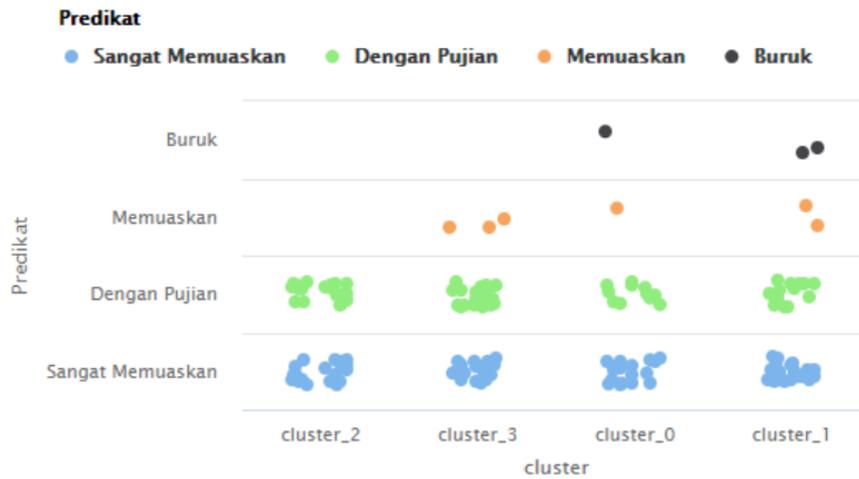
Gambar 5 merupakan gambar kondisi prestasi mahasiswa jurusan manajemen saat tahun ini. Warna biru menggambarkan predikat sangat memuaskan, warna hijau muda menggambarkan predikat dengan pujian, warna oranye menggambarkan predikat Memuaskan dan warna hitam menggambarkan predikat Buruk.

Hasil *cluster* tersebut dapat dilihat bahwa jumlah mahasiswa yang berpredikat Dengan Pujian sangat sedikit. Pengetahuan ini dapat dijadikan sebagai bahan pertimbangan kepada manajemen untuk mengambil langkah-langkah / tindakan dalam usaha meningkatkan prestasi mahasiswa manajemen.

Analisis Cluster

Anggota Cluster (Cluster Membership) Berdasarkan Predikat

Anggota *cluster* berdasarkan label predikat juga dapat ditampilkan dalam seperti gambar 6 berikut :



Gambar 6 Anggota cluster berdasarkan predikat

Gambar 6 merupakan gambar kondisi mahasiswa jurusan manajemen saat ini. Warna oranye berpredikat memuaskan untuk cluster 1 berjumlah 2, cluster 0 berjumlah 1, cluster 3 berjumlah 3. Warna hitam berpredikat buruk untuk cluster 1 berjumlah 2, cluster 0 berjumlah 1. Sedangkan hijau dan muda sangat mendominasi pada predikat Sangat Memuaskan dan dengan pujian.

Interpretasi Cluster

Hasil dari *cluster* yang terbentuk dapat dibuat sebuah model aturan. Tabel 8 merupakan model aturan untuk memprediksi prestasi akademik mahasiswa manajemen.

Tabel 8 Model aturan (*rule model*)

No.		Keterangan rule	Predikat
1	If	Rteori = Tinggi, Rpraktek = Tinggi dan Kehadiran Baik	Dengan Pujian
2	If	Rteori = Tinggi, Rpraktek = Tinggi dan Kehadiran Baik	Sangat Memuaskan
3	If	Rteori = Tinggi, Rpraktek = Sedang dan Kehadiran Baik	Memuaskan
4	If	Rteori = Sedang, Rpraktek = Tinggi dan Kehadiran Baik	Memuaskan
5	If	Rteori = Sedang, Rpraktek = sedang dan Kehadiran Baik	Memuaskan
6	If	Rteori = rendah, Rpraktek = rendah dan Kehadiran Baik	Buruk

5. KESIMPULAN DAN SARAN

Setelah dilakukan interpretasi terhadap hasil analisis data maka penulis dapat membuat

beberapa kesimpulan serta saran-saran bagi perguruan tinggi khususnya bagi pimpinan serta bagi para peneliti:

1. Diperoleh suatu model aturan yang dapat digunakan untuk memprediksi mahasiswa di jurusan manajemen sehingga dapat memberi manfaat dalam pengambilan keputusan.
2. Dalam *cluster* ini diperoleh mahasiswa yang memiliki predikat Dengan Pujian apabila nilai rata-rata teori tinggi (> 70), nilai rata-rata praktik tinggi (> 70) dan kehadiran baik (100 %).
3. Algoritma Kernel K-Means mampu mengklusterkan data campuran (numeric dan data kategorisasi) sehingga dihasilkan *cluster* mahasiswa dengan label predikatnya
4. Masih perlu dilakukan pengujian terhadap model aturan prediksi akademik mahasiswa
5. Masih perlu dilakukan pengujian clustering internal evaluation dan external evaluation.

DAFTAR PUSTAKA

- [1] Santosa Budi. 2007, Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis, Graha Ilmu, Yogyakarta.
- [2] Pramudiono, I. 2007. Pengantar Data Mining : Menambang Permata Pengetahuan di Gunung Data. <http://www.ilmukomputer.org/wp-content/uploads/2006/08/iko-datamining>.
- [3] Larose Daniel T., 2005, Discovering knowledge in data : an introduction to data mining, John Wiley & Sons, Inc.
- [4] Han, J. and Kamber, M. 2006. "Data mining: Concepts and Techniques", 2nd edition. The Morgan Kaufmann series in Data Management System, Jim Grey, series Editor.
- [5] Turban, E., dkk. 2005. Decision support Systems and Intelligent Systems, Yogyakarta, Andi Offset.
- [6] Piatetsky, G and Shapiro, 2006, An Introduction Machine Learning, data mining, and knowledge discovery, Course in data mining Kdnuggets.
- [7] Fayyad, U. M, 1996, Advances in Knowledge Discovery and Data Mining. Camberidge, MA: The MIT Press.
- [8] Kusri dan Emha Taufiq Luthfi, 2009, Algoritma Data Mining, Penerbit Andi Yogyakarta.
- [9] Rui Xu and Donald C. Wunsch II, 2009, Clustering, A John Wiley & Sons, Inc., Publication.
- [10] Hammouda, K. Karray, F. 2003, A Comparative Study of Data Clustering Techniques. Unpublished
- [11] Dubes R. C. and Jain, A.K., 1988, Algorithms for Clustering Data, Prentice-Hall.