
PEMODELAN ATURAN DALAM MEMPREDIKSI PRESTASI AKADEMIK MAHASISWA POLITEKNIK POLIPROFESI MEDAN DENGAN KERNEL K-MEANS CLUSTERING

A. Karim

Politeknik Poliprofesi Medan
Jalan Sie Batanghari No. 3 & 4 Medan
E-Mail : mas_karim@yahoo.com

Abstrak

Jurnal ini mengusulkan sebuah model aturan dalam memprediksi prestasi akademik mahasiswa di Jurusan Teknik Informatika Politeknik Poliprofesi Medan. Hingga saat ini memprediksi prestasi akademik mahasiswa masih menjadi perdebatan yang hangat di institusi-institusi pendidikan tinggi. Faktor-faktor yang berpengaruh secara dominan terhadap prestasi akademik mahasiswa masih belum dapat ditentukan secara pasti. Saat ini manajemen Politeknik Poliprofesi Medan masih menggunakan cara manual dalam memprediksi prestasi akademik mahasiswa. Sehingga sangat mungkin terjadi kesalahan dalam memprediksi prestasi akademik. Hal ini akan berpengaruh terhadap hasil keputusan yang akan diambil oleh pihak manajemen Politeknik Poliprofesi Medan. Untuk itu sangat penting dibuat sebuah model aturan untuk memprediksi prestasi akademik mahasiswa yang dapat digunakan pihak manajemen sebagai sistem pendukung dalam pengambilan keputusan. Data yang digunakan dalam penelitian ini berasal dari database akademik mahasiswa jurusan Teknik Informatika Politeknik Poliprofesi Medan tahun ajaran 2011-2012. Dalam penelitian ini algoritma Kernel K-Means Clustering telah digunakan untuk mendapatkan suatu model aturan prediksi prestasi akademik mahasiswa di jurusan Teknik Informatika Politeknik Poliprofesi Medan. Model aturan yang diperoleh menunjukkan bahwa predikat dengan pujian dapat diperoleh jika nilai rata-rata teori, nilai rata-rata praktek dan kehadiran semakin tinggi.

Kata kunci: kernel k-means clustering, model aturan, kelulusan

Abstract

The journal proposes a model rule in predicting academic achievement of students in the Department of Informatics Politechnic Poliprofesi field. Until now predicting students' academic achievement is still a warm perdebatan in higher education institutions. Factors that affect predominantly on student academic achievement still can not be determined with certainty. Currently Poliprofesi Polytechnic management field is still using the manual method in predicting students' academic achievement. So it is very possible that errors in predicting academic achievement. This will affect the results of the decision to be taken by the management of the Polytechnic Poliprofesi field. It is very important to be made a role model to predict the students' academic achievement that can be used as a management decision-making support system. The data used in this study comes from an academic database Polytechnic student majoring in Computer Science field Poliprofesi 2011-2012 school year. In this study the algorithm Kernel K-Means Clustering was used to obtain a model of student academic achievement prediction rules in Department of Information Engineering Polytechnic Poliprofesi field. The model shows that the rules obtained with praise predicate can be obtained if the average value theory, the average value of the practice and the higher attendance.

Keywords: kernel k-means clustering, the model rules, graduation

I. Pendahuluan

Dengan kemajuan teknologi informasi sekarang ini, kebutuhan akan informasi yang akurat sangat dibutuhkan dalam kehidupan sehari-hari. Terkadang kebutuhan informasi yang tinggi tidak diimbangi dengan penyajian informasi yang memadai, sehingga sering kali informasi tersebut masih harus di gali ulang dari kumpulan data yang jumlahnya sangat besar. Kemajuan teknologi untuk

mengumpulkan dan menyimpan berbagai jenis 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. Para pembuat keputusan berusaha untuk memanfaatkan kumpulan data yang sudah dimiliki untuk menggali 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 adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. 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.

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

Teknik data mining secara garis besar dapat dibagi dalam dua kelompok yaitu verifikasi dan discovery. Metode verifikasi umumnya meliputi teknik-teknik statistik seperti goodness of fit, dan analisis variansi. Metode *discovery* lebih lanjut dapat dibagi atas model prediktif dan model deskriptif. Teknik prediktif melakukan prediksi terhadap data dengan menggunakan hasil-hasil yang telah diketahui dari data yang berbeda. Model ini dapat dibuat berdasarkan penggunaan data historis lain. Sementara itu, model deskriptif bertujuan mengidentifikasi pola-pola atau hubungan antar data dan memberikan cara untuk mengeksplorasi karakteristik data yang diselidiki

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.

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.

Secara umum penilaian atas prestasi mahasiswa untuk suatu mata kuliah terdiri dari berbagai komponen yang meliputi beberapa atau semua komponen berikut: absen kehadiran, tugas mandiri, praktikum, kuis, Ujian Tengah Semester (UTS) dan Ujian Akhir Semester (UAS). Evaluasi dan penilaian terhadap prestasi mahasiswa dilakukan dengan pemberian nilai oleh dosen pengajar kepada semua mahasiswa yang mengikuti mata kuliah yang diajarnya.

Indeks Prestasi Kumulatif (IPK) mencerminkan seluruh nilai yang diperoleh mahasiswa sampai semester yang sedang berjalan, yang menunjukkan prestasi akademik mahasiswa bersangkutan sampai semester tersebut. IPK diperoleh dengan cara menjumlahkan seluruh nilai mutu semua mata kuliah yang telah diambil dan membaginya dengan total sks (satuan kredit semester). Oleh karena itu, IPK masih tetap merupakan faktor yang paling umum digunakan oleh perencana akademis untuk mengevaluasi perkembangan di lingkungan akademik

Banyak faktor yang menjadi penghalang bagi mahasiswa mencapai dan mempertahankan IPK 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 merupakan salah satu dasar untuk memantau perkembangan prestasi akademik mahasiswa didalam perguruan tinggi

Salah satu topik perdebatan di pendidikan tinggi adalah prediksi keberhasilan di perguruan tinggi. Tidak ada kepastian jika ada prediktor akademis yang akurat untuk menentukan apakah mahasiswa akan menjadi jenius, drop out, atau berprestasi rata-rata. Tugas untuk mengembangkan prediktor efektif untuk keberhasilan akademis merupakan isu penting bagi para pendidik.

Berdasarkan isu ini, pengelompokan mahasiswa kedalam kategori yang berbeda sesuai dengan prestasi mereka menjadi tugas yang rumit. Dengan pengelompokan mahasiswa secara tradisional berdasarkan nilai rata-rata mereka, maka sulit untuk memperoleh pandangan yang menyeluruh mengenai keadaan prestasi mahasiswa.

Dengan bantuan teknik data mining, seperti algoritma clustering, yang memungkinkan untuk menemukan karakteristik-karakteristik dari prestasi mahasiswa dan menggunakan karakteristik mereka

untuk memprediksi prestasi di masa depan. Algoritma clustering yang baik idealnya menghasilkan kelompok dengan batasan non-overlapping yang berbeda, meskipun dalam praktek pemisahan yang sempurna biasanya tidak bisa dicapai.

Dalam Peraturan Akademik Politeknik Poliprofesi Medan seorang mahasiswa dinyatakan lulus dengan ketentuan $2,00 \leq \text{IPK} \leq 2,75$ dengan kriteria Memuaskan. Dengan mengetahui kategori karakteristik mahasiswa diharapkan para pengambil kebijakan dapat mendorong para mahasiswa untuk memperbaiki prestasi akademiknya sebelum masa studi berakhir.

Kemampuan untuk memprediksi prestasi akademik mahasiswa sangat penting dalam sistem lembaga pendidikan. Dalam penelitian ini, digunakan teknik data mining yaitu algoritma Kernel K-Means Clustering untuk membuat model aturan dalam memprediksi prestasi akademik mahasiswa. Data yang digunakan dalam penelitian ini adalah data mahasiswa pada jurusan Teknik Informatika Politeknik Poliprofesi Medan.

Algoritma K-Means mengklusterkan ukuran jarak Euclidean, di mana jarak dihitung adalah untuk mencari kuadrat dari jarak antara masing-masing nilai, menjumlahkan kuadrat dan menemukan akar kuadrat dari jumlah tersebut.

Kernel K-Means adalah pengembangan dari algoritma K-Means yang menggunakan metode Kernel untuk memetakan data yang berdimensi tinggi pada space yang baru sehingga dapat dipisahkan secara linear. Hal ini dilakukan untuk meningkatkan akurasi hasil klaster. Didalam Kernel K-Means diharapkan data bisa dipisahkan dengan lebih baik karena data yang overlap atau data outlier bisa menjadi linier di ruang dimensi baru.

2. Metode Penelitian

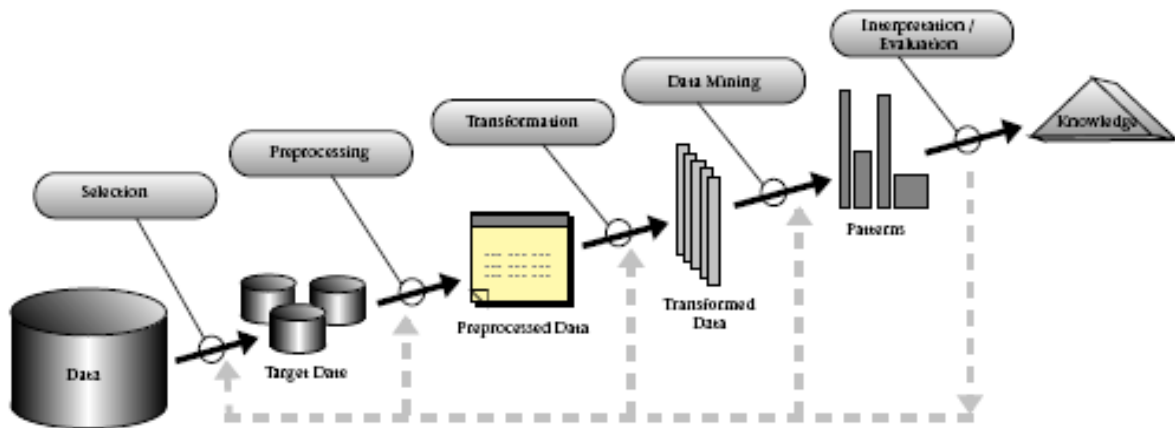
2.1 Data Mining : Knowledge Discovery Databases (KDD)

Institusi perguruan tinggi mencari cara mengelola dan mendukung prosedur pengambilan keputusan untuk membantu mereka dalam mengatur strategi dan rencana untuk manajemen yang lebih baik. Salah satu cara adalah dengan menggali pengetahuan (knowledge discovery) dan pengenalan pola (pattern recognition) terkait dengan proses pendidikan dan entitasnya. Pengetahuan dan pengenalan pola ini dapat diekstraksi dari data historis dan operasional yang berada di database organisasi pendidikan dengan menggunakan teknik data mining.

Data mining merupakan salah satu cara untuk menemukan informasi yang terkandung pada suatu data (*knowledge discovery*). Teknik Data mining dikembangkan untuk mencari data novel dan pola yang mungkin dapat digunakan pada database yang berskala besar.[1] *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. *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. Data mining secara formal adalah proses mengekstrak informasi yang valid, bermanfaat, tak dikenal, dan dapat dipahami dari data dan menggunakannya untuk membuat keputusan bisnis. *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 [2]

2.2 Tahapan 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 :



Gambar 1 *Data Mining* : Proses KDD

Menurut Kusri (Kusri & Emha, 2009:11) Proses KDD dapat diuraikan sebagai berikut :

- a. 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.
- b. 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.
- c. 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*.
- d. *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.
- e. 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.

2.3 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 [3]

Gagasan mengenai pengelompokan data, atau *clustering*, memiliki sifat yang sederhana dan dekat dengan cara berpikir manusia; kapanpun kepada kita dipresentasikan jumlah data yang besar, kita biasanya cenderung merangkumkan jumlah data yang besar ini ke dalam sejumlah kecil kelompok-kelompok atau kategori-kategori untuk memfasilitasi analisisnya lebih lanjut. Selain dari itu, sebagian besar data yang dikumpulkan dalam banyak masalah terlihat memiliki beberapa sifat yang melekat yang mengalami pengelompokan-pengelompokan natural[4].

Namun demikian, penemuan pengelompokan-pengelompokan ini atau upaya untuk mengkategorikan data adalah bukan sebuah tugas yang sederhana bagi manusia kecuali data memiliki dimensionalitas rendah (dua atau tiga dimensi paling banyak). Inilah sebabnya mengapa beberapa metode dalam *soft computing* telah dikemukakan untuk menyelesaikan jenis masalah ini. Metode ini disebut “Metode-metode Pengelompokan Data”

Algoritma-algoritma *clustering* digunakan secara ekstensif tidak hanya untuk mengorganisasikan dan mengkategorikan 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.

2.4 Kernel K-Means Clustering

K-Means merupakan algoritma yang umum digunakan untuk clustering dokumen. Prinsip utama K-Means adalah menyusun k prototype atau pusat massa (centroid) dari sekumpulan data berdimensi n” [5]. Sebelum diterapkan proses algoritma K-means, dokumen akan di preprocessing terlebih dahulu. kemudian dokumen direpresentasikan sebagai vektor yang memiliki term dengan nilai tertentu.

K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/kelompok. Metode ini mempartisi data ke dalam cluster/kelompok sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu cluster yang sama.

Pengorganisasian data ke dalam suatu kluster merupakan suatu model yang paling mendasar untuk pemahaman dan pembelajaran. Analisis kluster adalah studi formal untuk mengelompokkan, atau clustering benda-benda sesuai dengan karakteristik yang diukur berdasarkan kemiripan satu sama lain. Clustering adalah pengelompokan menggunakan teknik unsupervised learning dimana tidak diperlukan pelatihan pada metode tersebut atau dengan kata lain, tidak ada fase learning serta tidak menggunakan pelabelan pada setiap kelompok.

Pada survey yang dipublikasikan Springer, algoritma K-means ditempatkan pada posisi 2 (dua) sebagai algoritma paling banyak digunakan dalam data mining dan menjadi posisi pertama untuk algoritma clustering. Urutan Top Algoritma-nya adalah sebagai berikut :

- a) C4.5
- b) K-means
- c) SVM (Support Vector machines)
- d) Algoritma Apriori
- e) EM (Expectation Maximization)
- f) Algoritma PageRank
- g) Algoritma AdaBoost
- h) K-Nearest Neighbor
- i) Naive Bayes
- j) Classification and Regression Trees.

Metode kernel pertama dan barangkali yang paling tepat adalah *Support Vector Machine* (SVM)[6], yang mengoptimalkan kriteria margin maksimum dalam ruang fitur kernel. Algoritma k-means barangkali telah menjadi teknik *clustering* populer sejak diperkenalkan dalam era 1960an. Ini memaksimalkan jarak *Euclidean* kuadrat antara pusat-pusat *cluster*. Meskipun demikian, telah diketahui bahwa ini hanya optimal untuk (yang dapat dipisahkan secara linear) *cluster* terdistribusi Gaussian. Metode yang berbeda untuk melaksanakan algoritma ini dalam ruang kernel yakni kernel k-means telah diperoleh.

3. Hasil dan Pembahasan

Hasil kategorisasi data akan digunakan untuk input data pada software *open source* RapidMiner. Setelah itu dilakukan transformasi data dari format xls menjadi XML sehingga dihasilkan data pengujian dalam format XML. Dilakukan pengujian terhadap data yang sudah dalam bentuk XML menggunakan algoritma Kernel K-Means. Dari pengujian diperoleh *cluster* dari data yang telah diuji selanjutnya dilakukan analisis *cluster* untuk menganalisis dan mendapatkan model aturan yang digambarkan dari hasil *cluster*.

Tabel 1 Tampilan Kategorisasi Data

D	E	F	G	H	I	J	K	L	M
RATA-RATA PRAKTEK	RATA-RATA TEORI	KEHADIRAN	PRODI	IPK	Praktek	Teori	Hadir	Program Studi	Predikat
70	79	97%	2	3.35	Tinggi	Tinggi	Baik	Teknik Informatika	Dengan Pujian
73	69	92%	2	2.59	Tinggi	Sedang	Sedang	Teknik Informatika	Memuaskan
69	75	91%	2	2.45	Tinggi	Tinggi	Sedang	Teknik Informatika	Memuaskan
73	71	95%	2	3.31	Sedang	Tinggi	Sedang	Teknik Informatika	Sangat Memuaskan
73	69	87%	2	2.76	Tinggi	Sedang	Baik	Teknik Informatika	Memuaskan
73	65	87%	2	2.60	Tinggi	Sedang	Sedang	Teknik Informatika	Memuaskan
69	63	87%	2	2.70	Sedang	Sedang	Sedang	Teknik Informatika	Memuaskan
72	69	87%	2	2.79	Tinggi	Sedang	Sedang	Teknik Informatika	Memuaskan
73	68	87%	2	2.72	Tinggi	Sedang	Sedang	Teknik Informatika	Memuaskan
74	71	93%	2	3.14	Tinggi	Tinggi	Sedang	Teknik Informatika	Sangat Memuaskan
73	64	93%	2	2.24	Tinggi	Tinggi	Sedang	Teknik Informatika	Sangat Memuaskan
71	71	87%	2	2.93	Tinggi	Tinggi	Sedang	Teknik Informatika	Sangat Memuaskan
73	71	93%	2	3.10	Tinggi	Tinggi	Sedang	Teknik Informatika	Sangat Memuaskan
78	75	92%	2	3.14	Tinggi	Tinggi	Sedang	Teknik Informatika	Sangat Memuaskan
73	75	93%	2	2.59	Tinggi	Sedang	Sedang	Teknik Informatika	Memuaskan
73	65	87%	2	2.60	Tinggi	Sedang	Sedang	Teknik Informatika	Memuaskan
69	63	87%	2	2.70	Sedang	Sedang	Sedang	Teknik Informatika	Memuaskan
72	69	87%	2	2.79	Tinggi	Sedang	Sedang	Teknik Informatika	Memuaskan
73	68	87%	2	2.72	Tinggi	Sedang	Sedang	Teknik Informatika	Memuaskan
74	71	93%	2	3.14	Tinggi	Tinggi	Sedang	Teknik Informatika	Sangat Memuaskan
73	64	93%	2	2.24	Tinggi	Tinggi	Sedang	Teknik Informatika	Sangat Memuaskan
73	69	92%	2	2.59	Tinggi	Sedang	Sedang	Teknik Informatika	Memuaskan
69	75	91%	2	2.45	Tinggi	Tinggi	Sedang	Teknik Informatika	Memuaskan
73	71	95%	2	3.31	Sedang	Tinggi	Sedang	Teknik Informatika	Sangat Memuaskan
73	69	87%	2	2.76	Tinggi	Sedang	Baik	Teknik Informatika	Memuaskan
73	65	87%	2	2.60	Tinggi	Sedang	Sedang	Teknik Informatika	Memuaskan

3.1 Hasil Transformasi Data

Data yang berasal dari format xls akan ditransformasi menjadi XML. Hal ini dilakukan untuk mendapatkan data pengujian yang akan digunakan sebagai data input pada proses clustering menggunakan algoritma Kernel K-Means. Data hasil transformasi dapat dilihat pada tabel 2 berikut :

Tabel 2. Tabel Transformasi Data

A	B	C	D	E	F	G	H	I	J
numeric	numeric	numeric	integer	numeric	polyno...	polyno...	polyno...	binomi...	polyno...
attribut	attribut	attribut	attribut	attribut	attribut	attribut	attribut	attribut	attribut
RATA-RATA	RATA-RATA	KEHADIRAN	PRODI	IPK	Praktek	Teori	Hadir	ProgramStudi	Predikat
70	79	0.97	2	3.35	Tinggi	Tinggi	Baik	Teknik Inforr	Dengan Puji
73	69	0.92	2	2.59	Tinggi	Sedang	Sedang	Teknik Inforr	Memuaskan
69	75	0.91	2	2.45	Tinggi	Tinggi	Sedang	Teknik Inforr	Memuaskan
73	71	0.95	2	3.31	Sedang	Tinggi	Sedang	Teknik Inforr	Sangat Mem
73	69	0.87	2	2.76	Tinggi	Sedang	Baik	Teknik Inforr	Memuaskan
73	65	0.87	2	2.60	Tinggi	Sedang	Sedang	Teknik Inforr	Memuaskan
69	63	0.87	2	2.70	Sedang	Sedang	Sedang	Teknik Inforr	Memuaskan
72	69	0.87	2	2.79	Tinggi	Sedang	Sedang	Teknik Inforr	Memuaskan
73	68	0.87	2	2.72	Tinggi	Sedang	Sedang	Teknik Inforr	Memuaskan
74	71	0.93	2	3.14	Tinggi	Tinggi	Sedang	Teknik Inforr	Sangat Mem
73	64	0.93	2	2.24	Tinggi	Tinggi	Sedang	Teknik Inforr	Sangat Mem
71	71	0.87	2	2.93	Tinggi	Tinggi	Sedang	Teknik Inforr	Sangat Mem
73	71	0.93	2	3.10	Tinggi	Tinggi	Sedang	Teknik Inforr	Sangat Mem
78	75	0.92	2	3.14	Tinggi	Tinggi	Sedang	Teknik Inforr	Sangat Mem

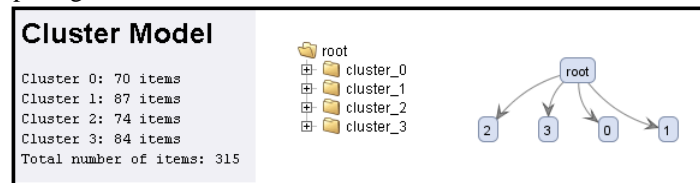
Pada tabel 2 diatas sebagai atribut adalah : RATAPRAKTEK dengan tipe numeric, RATA TEORI dengan tipe numeric, KEHADIRAN dengan tipe *numeric*, PRODI dengan tipe *integer*, IPK dengan tipe

numeric. Semua atribut ini merupakan data yang belum dikategorisasi. Sedangkan atribut yang sudah dikategorisasi adalah : praktek dengan tipe *polynomial*, teori dengan tipe *polynomial*, hadir dengan tipe *polynomial*, reigion dengan tipe *polynomial*, progstudi dengan tipe *polynomial*, jeniskel dengan tipe *binomial*, dan predikat dengan tipe *polynomial*.

Tipe *polynomial* merupakan tipe dengan banyak kategori. Dimana kategori untuk praktek dan teori terdiri dari rendah, sedang dan tinggi. Kategori untuk hadir terdiri dari baik, sedang dan kurang. Kategori untuk progstudi adalah Teknik Informatika, Manajemen Infomatika dan Teknik Komputer. Kategori untuk predikat adalah dengan pujian, sangat memuaskan, memuaskan dan buruk

3.2 Cluster Model

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



Gambar 2 Cluster Model

Dari gambar diatas dapat dilihat *cluster* model yang dihasilkan terdiri dari *cluster* 0 sebanyak 70 item, *cluster* 1 sebanyak 87 item, *cluster* 2 sebanyak 74 item dan *cluster* 3 sebanyak 84 item dari total jumlah 315 item. Dalam bentuk folder root tampak bahwa root memiliki folder untuk 4 *cluster* yaitu folder *cluster* 0, *cluster* 1, *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 berikut menampilkan *membership* dari masing-masing *cluster*.

Pada gambar 3 ditampilkan *membership* masing-masing *cluster* dimana setiap anggota wakili oleh nomor barisnya. *Cluster* 0 memiliki anggota mulai nomor baris 7.0, *cluster* 1 memiliki anggota mulai 11.0, *cluster* 2 memiliki anggota mulai nomor baris 1.0 dan *cluster* 3 memiliki anggota mulai nomor baris 2.0.

cluster_0	cluster_1	cluster_2	cluster_3
7.0	11.0	1.0	2.0
9.0	18.0	4.0	3.0
12.0	27.0	5.0	8.0
13.0	28.0	6.0	10.0
14.0	34.0	17.0	15.0
16.0	37.0	23.0	19.0
21.0	38.0	24.0	20.0
22.0	39.0	35.0	26.0
25.0	43.0	36.0	40.0
29.0	46.0	44.0	41.0
30.0	50.0	45.0	42.0
31.0	56.0	47.0	49.0
32.0	59.0	52.0	51.0
33.0	61.0	53.0	55.0
48.0	64.0	54.0	60.0
57.0	69.0	57.0	65.0
62.0	70.0	67.0	66.0
63.0	76.0	71.0	68.0
75.0	77.0	94.0	72.0
79.0	80.0	97.0	73.0
83.0	82.0	100.0	74.0
86.0	85.0	106.0	78.0

Gamabr 3 Tampilan Membership masing-masing Cluster

Untuk lebih jelasnya persentase jumlah anggota masing-masing *cluster* dapat dilihat pada tabel 3. berikut :

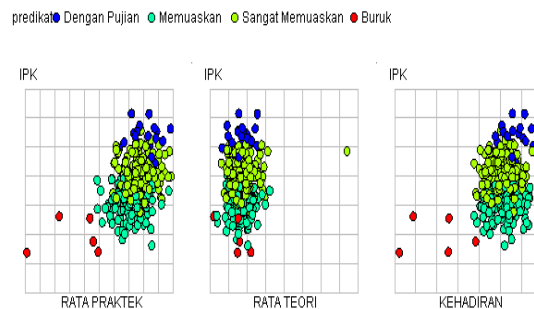
Tabel 3. Tabel Tabel *Cluster Model*

Cluster	Jumlah Anggota	Persentase
0	70	0,222
1	87	0,276
2	74	0,235
3	84	0,267

Performa sebuah *cluster* dapat dilihat dari persentase jumlah anggota dari masing-masing cluster hampir sama sebagaimana yang terlihat pada tabel diatas

3.3 Cluster Data Berdasarkan Predikat Prestasi Akademik

Cluster data dengan label predikat terhadap praktek, teori dan kehadiran dapat dilihat pada gambar dibawah ini :



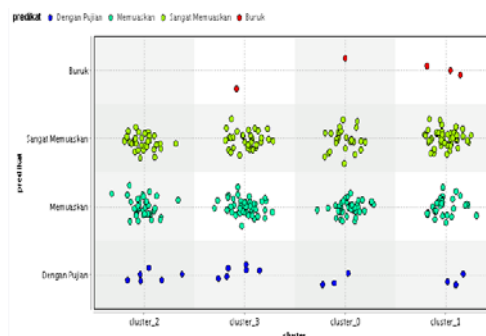
Gambar 4. Distribusi data antara IPK dengan PRAKTEK, TEORI dan KEHADIRAN

Gambar 4. merupakan gambar kondisi prestasi mahasiswa jurusan Teknik Informatika Politeknik Poliprofesii Medan saat ini. Warna biru menggambarkan predikat Dengan Pujian, warna hijau muda menggambarkan predikat Sangat Memuaskan, warna Biru muda menggambarkan predikat Memuaskan dan warna merah menggambarkan predikat Buruk.

Berdasarkan 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 jurusan Teknik Informatika Politeknik Poliprofesi Medan.

3.4 Anggota Cluster (*Cluster Membership*) Berdasarkan Predikat

Anggota *cluster* berdasarkan label predikat juga dapat ditampilkan dalam seperti gambar dibawah ini :

Gambar 5. Anggota *Cluster* berdasarkan Predikat

Gambar 5. diatas merupakan gambar kondisi mahasiswa jurusan Teknik Informatika saat ini. Warna biru berpredikat Dengan Pujian untuk *cluster* 2 berjumlah 6, *cluster* 3 berjumlah 6, *cluster* 0 berjumlah 3 dan *cluster* 1 berjumlah 3. Warna merah berpredikat Buruk untuk *cluster* 3 berjumlah 1, *cluster* 0 berjumlah 1 dan *cluster* 1 berjumlah 3. Sedangkan warna biru muda dan hijau muda sangat mendominasi pada predikat Sangat Memuaskan dan Memuaskan.

Jumlah anggota *cluster* pada gambar diatas dapat diuraikan dalam bentuk tabel seperti pada tabel dibawah ini

Tabel 4 Jumlah Anggota *Cluster* Terhadap Predikat

Predikat	Jumlah Anggota			
	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Dengan Pujian	3	3	6	6
Sangat Memuaskan	31	49	42	37
Memuaskan	29	29	33	43
Buruk	1	2	-----	1

3.5 Interpretasi *Cluster*

Berdasarkan hasil dari *cluster* yang terbentuk dapat dibuat sebuah model aturan. Tabel dibawah ini merupakan model aturan untuk memprediksi prestasi akademik mahasiswa jurusan Teknik Informatika Politeknik Poliprofesi Medan.

Tabel 5. Model Aturan (*Model Rule*)

No.		Keterangan <i>Rule</i>	Predikat
1.	If	Rata-rata Teori = Tinggi, Rata-rata Praktek = Tinggi dan Kehadiran = Baik	Dengan Pujian
2.	If	Rata-rata Teori = Tinggi, Rata-rata Praktek = Tinggi dan Kehadiran = Sedang	Sangat Memuaskan
3.	If	Rata-rata Teori = Tinggi, Rata-rata Praktek = Sedang dan Kehadiran = Sedang	Memuaskan
4.	If	Rata-rata Teori = Sedang, Rata-rata Praktek = Tinggi dan Kehadiran = Sedang	Memuaskan
5.	If	Rata-rata Teori = Tinggi, Rata-rata Praktek = Sedang dan Kehadiran = Baik	Memuaskan
6.	If	Rata-rata Teori = Sedang, Rata-rata Praktek = Tinggi dan Kehadiran = Baik	Memuaskan
7.	If	Rata-rata Teori = Sedang, Rata-rata Praktek = Sedang dan Kehadiran = Sedang	Memuaskan
8.	If	Rata-rata Teori = Sedang, Rata-rata Praktek = Sedang dan Kehadiran = Baik	Memuaskan
9.	If	Rata-rata Teori = Rendah, Rata-rata Praktek = Rendah dan Kehadiran = Kurang	Buruk

Dari tabel diatas dapat dijelaskan bahwa jika Rata-rata Teori = Tinggi, Rata-rata Praktek = Tinggi dan Kehadiran = Baik maka diperoleh predikat Dengan Pujian. Jika Rata-rata Teori = Tinggi, Rata-rata Praktek = Tinggi dan Kehadiran = Sedang maka diperoleh predikat Sangat Memuaskan. Jika Rata-rata Teori = Tinggi, Rata-rata Praktek = Sedang dan Kehadiran = Sedang maka diperoleh predikat Memuaskan. Jika Rata-rata Teori = Sedang, Rata-rata Praktek = Tinggi dan Kehadiran = Sedang maka

diperoleh predikat Memuaskan. Jika Rata-rata Teori = Tinggi, Rata-rata Praktek = Sedang dan Kehadiran = Baik maka diperoleh predikat Memuaskan. Jika Rata-rata Teori = Sedang, Rata-rata Praktek = Tinggi dan Kehadiran = Baik maka diperoleh predikat Memuaskan. Jika Rata-rata Teori = Sedang, Rata-rata Praktek = Sedang dan Kehadiran = Sedang maka diperoleh predikat Memuaskan. Jika Rata-rata Teori = Sedang, Rata-rata Praktek = Sedang dan Kehadiran = Baik maka diperoleh predikat Memuaskan. Jika Rata-rata Teori = Rendah, Rata-rata Praktek = Rendah dan Kehadiran = Kurang maka diperoleh predikat Buruk.

4. Penutup

Algoritma Kernel K-Means *Clustering* dalam dunia pendidikan dapat digunakan untuk membangun model aturan dalam memprediksi prestasi akademik mahasiswa. Dari uraian yang telah dipaparkan dalam penelitian ini dapat dihasilkan kesimpulan sebagai berikut :

- a. Diperoleh suatu model aturan yang dapat digunakan untuk memprediksi predikat mahasiswa di jurusan Teknik Informatika Politeknik Poliprofesi Medan yang dapat memberi manfaat dalam pengambilan keputusan.
- b. Dalam *cluster* ini diperoleh mahasiswa yang memiliki predikat Dengan Pujian apabila nilai rata-rata teori tinggi (> 70), nilai rata-rata praktek tinggi (> 70) dan kehadiran baik ($> 93\%$).
- c. Algoritma Kernel K-Means mampu mengklusterkan data campuran (numeric dan data kategorisasi) sehingga dihasilkan *cluster* mahasiswa dengan label predikatnya.

Untuk penelitian yang lebih lanjut hasil dari penelitian ini dapat dilakukan pengujian terhadap model aturan prediksi prestasi akademik mahasiswa Politeknik Poliprofesi Medan dengan menggunakan metode klasifikasi misal SSVM (*Smooth Support Vector Machine*).

Untuk mendapatkan hasil yang lebih variatif mungkin penelitian ini dapat juga dikembangkan dengan teknik *data mining* yang lain seperti algoritma *hierarchical clustering*, algoritma *K-Nearest Neighbor* atau algoritma *Fuzzy Clustering* untuk mendapatkan model aturan dalam memprediksi prestasi akademik di Politeknik Poliprofesi Medan

Referensi:

- [1] Cahyo, dkk Clustering Kelompok Swadaya Masyarakat (KSM) dalam Menentukan Kebijakan Bantuan Badan Pemberdayaan Masyarakat di Kota Surabaya dengan Menggunakan Metode Self-Organizing Map (SOM) dan K-Means, JURNAL TEKNIK ITS Vol. 1, No. 1(Sept. 2012) ISSN: 2301-9271
- [2] Heribertus Yulianton, Data Mining untuk Dunia Bisnis, Jurnal Teknologi Informasi DINAMIK Volume XIII, No.1, Januari 2008 : 9-15
- [3] Rui Xu and Donald C. Wunsch II, 2009, *Clustering*, A John Wiley & Sons, Inc., Publication
- [4] Hammouda, K. Karray, F. (2003), *A Comparative Study of Data Clustering Techniques*. Unpublished
- [5] Dwi Noviati Nango, (2012) “Penerapan Algoritma K-Means Untuk Clustering Data Anggaran Pendapatan Belanja Daerah Di Kabupaten Xyz” Universitas Gorontalo
- [6] Burges, Christopher.(1998), *A Tutorial on Support Vector Machines for Pattern*