

KLASIFIKASI DAN KLASTERING MAHASISWA INFORMATIKA UNIVERSITAS MUHAMMADIYAH SURAKARTA

Yusuf Sulisty Nugroho

Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta

email: Yusuf.Nugroho@ums.ac.id

Abstract

Data in an organization which are currently increasingly more and accumulate, will not lead to the use of data become optimal. Informatics Department in UMS that has been established since 2007 is one of the study programs that have a large data. The amount of this data will only be a pile of data if it is not processed into strategic information with certain methods, such as classification and clustering. This study was done to take advantage of the abundant data as a source of strategic information for the department to classify the students' length of study and the students' degree of excellence and also to cluster them using data mining techniques. Students' classification was done by using Decision Tree from 223 graduated students' data, while the clustering was conducted by using K-Means algorithm from 209 active students' data. Attributes used in this study consists of high school majors, gender, high school region, the average number of credits hour per semester, and students' participation as an assistant which were set as an independent variable. While the length of study and the degree of excellence were set as the dependent variable. Informatics students classification and clustering shows that the most significant variable influencing on the length of study is the average of credit hours taken per semester by students, while the variables that most influence on students' degree of excellence is student participation as an assistant. The result interprets that the variables that need to be used as consideration for the department to obtain the effective rate of the length of study is the average of credit hours, while the variable as consideration to obtain the maximum degree of excellence is the student participation as an assistant.

Keywords: *classification, clustering, data mining, decision tree, degree of excellence, length of study*

1. PENDAHULUAN

Kemajuan teknologi informasi telah menyebabkan banyak orang dapat memperoleh data dengan mudah bahkan cenderung berlebihan. Data tersebut semakin lama semakin banyak dan terakumulasi, akibatnya pemanfaatan data yang terakumulasi tersebut menjadi tidak optimal (Nugroho, 2014). Sebagai contoh perusahaan retail yang akan memberikan brosur penawaran barang-barang yang dijual ke pelanggan sesuai basis data pelanggan yang mereka punya. Jika perusahaan retail tersebut mempunyai satu juta data pelanggan dan masing-masing pelanggan tersebut dikirimkan sebuah brosur penawaran dimana biaya pengiriman brosur tersebut adalah dua ribu rupiah, maka biaya yang akan

dikeluarkan oleh perusahaan tersebut adalah dua juta rupiah per bulan. Dari penggunaan dana tersebut mungkin hanya sepertiganya atau bahkan 8% saja yang secara efektif membeli penawaran tersebut (Buhlman, 2002).

Penumpukan data juga bisa terjadi dalam dunia pendidikan. Salah satu lembaga pendidikan yang cukup besar adalah Universitas Muhammadiyah Surakarta, yang memiliki salah satu program studi dengan jumlah mahasiswa yang cukup banyak yaitu Program Studi Informatika yang berdiri sejak tahun 2007. Jumlah mahasiswa yang diterima setiap tahunnya minimal sebanyak 150 mahasiswa. Dengan demikian, data yang dimiliki oleh program studi hingga kini telah melimpah. Banyaknya data ini hanya akan

menjadi tumpukan data jika tidak diolah menjadi informasi yang strategis dengan metode-metode tertentu, misalnya klasifikasi dan klustering.

Namun, kegiatan klasifikasi dan klustering jika dilakukan oleh manusia masih memiliki keterbatasan, terutama pada kemampuan manusia dalam menampung jumlah data yang ingin diolah. Selain itu bisa juga terjadi kesalahan akibat ketidakteelitian yang dilakukan. Salah satu cara mengatasi masalah ini adalah dengan menggunakan teknik Data Mining (DM) yang bisa digunakan untuk pengolahan data menjadi sumber informasi strategis dengan metode klasifikasi dan klustering. Data mining dapat membantu sebuah organisasi yang memiliki data melimpah untuk memberikan informasi yang dapat mendukung pengambilan keputusan (Kiron et al, 2012).

Dalam bidang analisis perusahaan dan manajemen resiko, data mining digunakan untuk merencanakan keuangan dan evaluasi aset, merencanakan sumber daya (resources planning) dan memonitor persaingan (Anggraini, 2009).

Berdasarkan pada latar belakang yang diuraikan, peneliti bermaksud untuk melakukan analisis data mahasiswa yang dapat menjadi sumber informasi bagi Program Studi Informatika terkait dengan rencana strategis di masa mendatang untuk menentukan calon mahasiswa baru berdasarkan kriteria-kriteria tertentu guna memperoleh mahasiswa yang unggul dari waktu ke waktu. Dengan demikian, hal ini bisa berdampak terhadap peningkatan nilai akreditasi bagi program studi.

2. METODE PENELITIAN

A. Studi Literatur atau Kepustakaan.

Tahap studi dilakukan dengan menelusuri literatur serta menelaahnya untuk menggali teori-teori yang sedang berkembang, mencari metode penelitian yang digunakan terdahulu dan untuk memperoleh orientasi yang ada dalam permasalahan.

B. Pemilihan Obyek Penelitian.

Penelitian ini dilakukan untuk mengklasifikasikan dan klustering mahasiswa Program Studi Informatika UMS. Obyek penelitian ini sengaja dipilih dengan pertimbangan jumlah mahasiswa yang banyak dan mengalami peningkatan setiap tahun, sehingga data-data yang melimpah bisa dimanfaatkan untuk keperluan data mining.

C. Penentuan Variabel Data Mining.

Variabel variabel yang akan di-gunakan untuk proses data mining ini ditentukan berdasarkan tujuan penelitian. Masa studi dan predikat kelulusan mahasiswa digunakan sebagai variabel yang akan dicari pola pengelompok-annya.

Ada 2 (dua) jenis variabel yang ditentukan dalam proses data mining ini, yaitu:

a) Variabel dependen (Y)

Variabel dependen (Y) adalah variabel yang nilainya tergantung atau terikat berdasarkan nilai-nilai variabel lainnya. Variabel Y yang digunakan dibagi menjadi 2 macam yaitu :

- 1) Masa Studi Mahasiswa sebagai variabel Y1, dan
- 2) Predikat Kelulusan Mahasiswa yang nilainya ditentukan berdasarkan IPK sebagai variabel Y2.

b) Variabel independen (X)

Variabel independen (X) adalah variabel yang nilainya tidak ter-gantung dari nilai-nilai variabel lainnya. Variabel X yang diperlukan terdiri dari:

- 1) Jurusan Sekolah (SMA sederajat), sebagai X1
- 2) Jenis Kelamin, sebagai X2
- 3) Asal Sekolah, sebagai X3
- 4) Rerata Jumlah SKS per semester, sebagai X4
- 5) Pernah Menjadi Asisten, sebagai X5

D. Penentuan Nilai Class Variabel.

Berdasarkan variabel yang telah ditentukan, tahapan berikutnya adalah menentukan nilai-nilai class dari masing-masing variabel Y dan variabel X.

- a) Nilai *Class* Variabel Y
 - 1) Variabel Y1 memiliki 2 nilai *class* yang bertipe *label*, yaitu:
 - a. Tepat Waktu, jika $Y1 < 5$ tahun
 - b. Terlambat, jika $Y1 \geq 5$ tahun
 - 2) Variabel Y2 terdiri dari 3 nilai *class* yang bertipe *label*, yaitu:
 - a. Memuaskan, jika $2,00 \leq Y2 < 2,76$
 - b. Sangat memuaskan, jika $2,76 \leq Y2 < 3,51$
 - c. *Cumlaude*, jika $Y2 \geq 3,51$
- b) Nilai *Class* Variabel X
 - 1) Variabel X1 dibedakan menjadi 3 nilai *class* yang bertipe *polynomial*, yaitu: IPA (0), IPS (1) dan LAIN (2) (selain IPA dan IPS).
 - 2) Variabel X2 terdiri dari 2 nilai *class* dengan tipe *binomial*, yaitu PRIA (0) dan WANITA (1)
 - 3) Variabel X3 dibuat menjadi 2 nilai *class* yang bertipe *binomial*, yaitu SURAKARTA (1) (jika asal sekolah se-Karesidenan Surakarta) dan LUAR (0) (jika asal sekolah di luar Karesidenan Surakarta)
 - 4) Variabel X4 dibagi menjadi 2 nilai *class* yang bertipe *binomial*, yaitu $SKS \leq 18$ (1) dan $SKS > 18$ (0)
 - 5) Variabel X5 terdiri dari 2 nilai *class* dengan tipe *binomial*, yaitu YA (1)

(jika pernah menjadi asisten) dan TIDAK (0) (jika belum pernah menjadi asisten).

E. Pengumpulan data.

Sampel data mahasiswa Informatika diperoleh dari IT UMS dengan jumlah 223 mahasiswa yang sudah lulus dan 209 data mahasiswa aktif.

F. Olah Data.

Olah data yang dilakukan meliputi pemisahan atribut-atribut yang diperlukan untuk proses data mining, standarisasi data (*preprocessing*), hingga perubahan data-data *real* menjadi data-data dengan tipe *binomial* maupun *polynomial* sesuai dengan kebutuhan data mining.

G. Analisis Data

Tahapan klasifikasi dilakukan dengan 2 tahap utama yaitu tahap analisis untuk menentukan klasifikasi mahasiswa berdasarkan masa studi dan klasifikasi untuk menentukan pola predikat kelulusannya. Klasifikasi tersebut dilakukan dengan metode *Decision Tree* dengan penentuan atributnya menggunakan *information gain* berdasarkan entropi dari masing-masing atribut yang telah ditentukan dengan persamaan (1) dan (2).

$$gain(y, A) = entropi(y) - \sum_{c \in nilai(A)} \frac{y_c}{y} entropi(y_c) \dots \dots \dots (1)$$

$$Entropi(y) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \dots - p_n \log_2 p_n \dots \dots \dots (2)$$

Tahap klastering mahasiswa dilakukan dengan algoritma K-Means. Metode ini mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam *cluster* yang lain.

Pembagian kelompok *cluster* menggunakan persamaan *Euclidean distance space* (persamaan 3) yang sering digunakan dalam perhitungan jarak, hal ini dikarenakan hasil yang diperoleh merupakan jarak terpendek antara dua titik yang diperhitungkan.

$$d_{ij} = \sqrt{\sum_{k=1}^p \{x_{ik} - x_{jk}\}^2} \dots \dots \dots (3)$$

3. HASIL DAN PEMBAHASAN

A. Data Collecting dan Preprocessing

Data-data mahasiswa Program Studi Informatika yang diperoleh dari hasil pengumpulan data untuk keperluan data mining masih bersifat transaksional dan tercatat dalam format Microsoft Excel (terlihat pada tabel 1 dan tabel 2).

Kelas data yang digunakan untuk data mining disiapkan (*preprocessing*) sehingga memiliki kelas *binomial* atau *polynomial* sesuai aturan yang telah dibuat berdasarkan nilai datanya. Tabel 3 merupakan pembagian variabel dan kelas data yang digunakan dalam analisis data mining. Potongan data hasil *preprocessing* sesuai dengan jenis kelas datanya dapat dilihat pada tabel 4.

Tabel 1. Potongan Data Mahasiswa Lulus

NIM	JUR SMA	GENDER	SMU	RATA SKS	ASIST	lama studi	IPK
L200070003	IPA	L	SMU Negeri 1 Ngrambe	19	TIDAK	5.00	2.89
L200070005	IPA	L	SMTA Lain-lain	21	YA	4.40	3.17
L200070006	IPS	L	SMU Negeri 1 Kebakkramat	18	TIDAK	5.90	2.85
L200070008	IPS	P	SMU Negeri 5 Surakarta	18	TIDAK	5.00	3.09
L200070009	IPA	L	SMTA Lain-lain	19	YA	4.90	3.10
L200070010	LAIN	L	SMK Negeri 2 Surakarta	19	TIDAK	5.00	2.89
L200070011	IPA	L	SMTA Lain-lain	17	TIDAK	5.90	2.63
L200070012	IPA	P	MA Negeri Sragen	17	TIDAK	4.40	3.06

Tabel 2. Potongan Data Mahasiswa Aktif

NIM	JUR SMA	GENDER	SMU	RATA SKS	ASIST
L200080051	Lainnya	L	SMU Veteran Sukoharjo	17	TIDAK
L200080052	Lainnya	L	SMK Negeri 2 Surakarta	16	TIDAK
L200080059	IPA	L	SMU Negeri 1 Nguter	18	TIDAK
L200080068	Lainnya	L	STM Bhineka Karya Surakar	20	TIDAK
L200080081	IPA	L	SMU Negeri 1 Kradenan	17	TIDAK
L200080088	IPS	P	SMU Muhammadiyah 1 Pemala	20	TIDAK
L200080137	IPS	L	SMU Negeri 1 Penengahan	19	TIDAK

Tabel 3. Pembagian Variabel dan Kelas Data

Variabel	Nama Field	Jenis Kelas Data	Kelas data yang digunakan
Y1	Masa Studi	Binomial	TEPAT, TERLAMBAT
Y2	Predikat Kelulusan	Polynomial	CUMLAUDE, SANGAT MEMUASKAN, MEMUASKAN
X1	Jurusan SMA	Polynomial	IPA, IPS, LAIN
X2	Gender	Binomial	PRIA, WANITA
X3	Asal SMA	Binomial	SURAKARTA, LUAR
X4	Rerata SKS	Binomial	SKS \leq 18, SKS $>$ 18
X5	Asisten	Binomial	YA, TIDAK

Tabel 4. Potongan Hasil Preprocessing Data Mahasiswa Lulus

JURUSAN	GENDER	ASAL_SEKOLAH	RERATA_SKS	ASISTEN	MASA_STUDI	PREDIKAT
IPA	PRIA	LUAR	SKS>18	TIDAK	TERLAMBAT	SANGAT MEMUASKAN
IPA	PRIA	SURAKARTA	SKS>18	YA	TEPAT	SANGAT MEMUASKAN
IPS	PRIA	SURAKARTA	SKS<=18	TIDAK	TERLAMBAT	SANGAT MEMUASKAN
IPS	WANITA	SURAKARTA	SKS<=18	TIDAK	TERLAMBAT	SANGAT MEMUASKAN
IPA	PRIA	SURAKARTA	SKS>18	YA	TEPAT	SANGAT MEMUASKAN
LAIN	PRIA	SURAKARTA	SKS>18	TIDAK	TERLAMBAT	SANGAT MEMUASKAN
IPA	PRIA	LUAR	SKS<=18	TIDAK	TERLAMBAT	MEMUASKAN
IPA	WANITA	SURAKARTA	SKS<=18	TIDAK	TEPAT	SANGAT MEMUASKAN
IPA	WANITA	LUAR	SKS<=18	YA	TEPAT	SANGAT MEMUASKAN
IPA	PRIA	SURAKARTA	SKS<=18	TIDAK	TERLAMBAT	SANGAT MEMUASKAN
IPA	PRIA	SURAKARTA	SKS>18	TIDAK	TEPAT	SANGAT MEMUASKAN

B. Klasifikasi Masa Studi Menggunakan Decision Tree

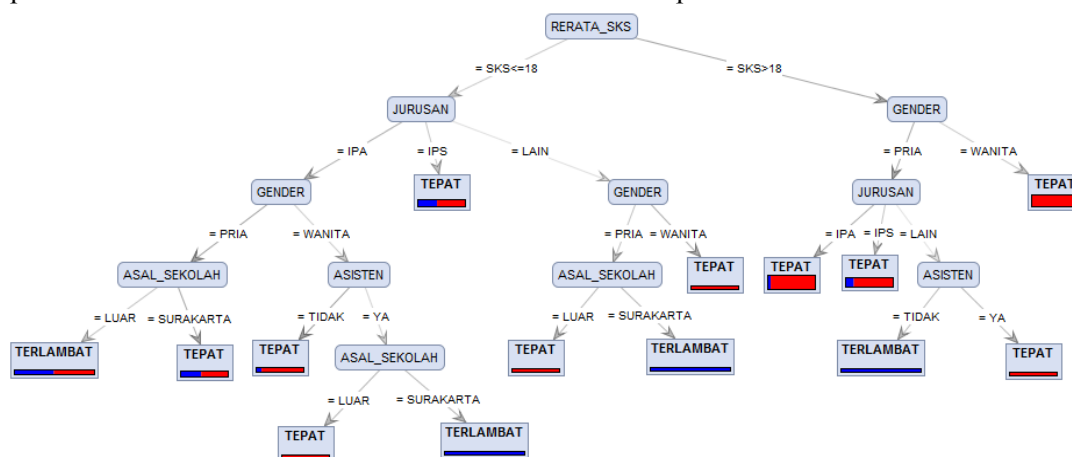
Hasil proses klasifikasi masa studi dari data mahasiswa dengan metode *Decision Tree* menggunakan aplikasi Rapid Miner 5 ditunjukkan pada gambar 1. Berdasarkan hasil pohon keputusan pada gambar 1, dapat dilihat bahwa atribut yang memiliki pengaruh paling tinggi untuk menentukan klasifikasi masa studi mahasiswa adalah Rerata SKS (X4). Hal ini ditunjukkan dengan variabel X4 yang menempati sebagai simpul akar (*root node*).

Contoh penerapan klasifikasi pohon keputusan pada sebuah data berdasarkan gambar 1 sebagai berikut jika seorang mahasiswa yang mengambil rata-rata SKS tiap semester adalah lebih dari 18 SKS dan

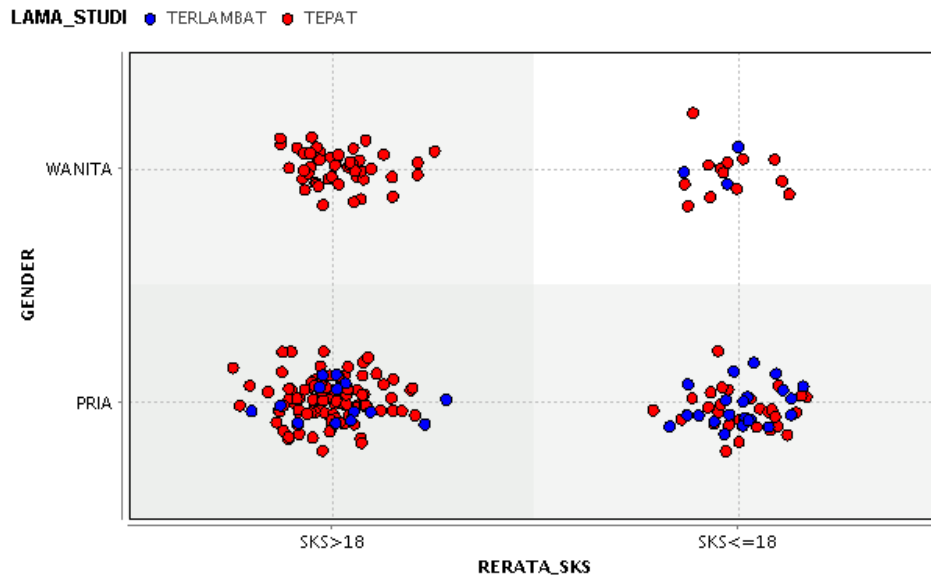
berjenis kelamin wanita, maka mahasiswa tersebut diklasifikasikan memiliki masa studi TEPAT WAKTU, tanpa mempedulikan asal sekolah, jurusan sekolah dan peran asisten.

Selain klasifikasi masa studi mahasiswa dapat dilihat menggunakan pohon keputusan, contoh pola distribusi antar atribut terhadap masa studi dapat disajikan menggunakan grafik *scatter plot* gambar 2.

Pola distribusi masa studi berdasarkan gambar 2 menunjukkan bahwa seorang mahasiswa yang mengambil rerata SKS lebih dari 18 SKS per semester dan berjenis kelamin wanita memiliki probabilitas masa studi tepat waktu paling tinggi dibandingkan dengan mahasiswa yang berjenis kelamin pria atau mengambil rerata SKS kurang dari 18 SKS per semester.



Gambar 1. Pohon Keputusan untuk Klasifikasi Masa Studi



Gambar 2. Distribusi Masa Studi Menggunakan Pohon Keputusan

C. Klasifikasi Predikat Kelulusan Menggunakan *Decision Tree*

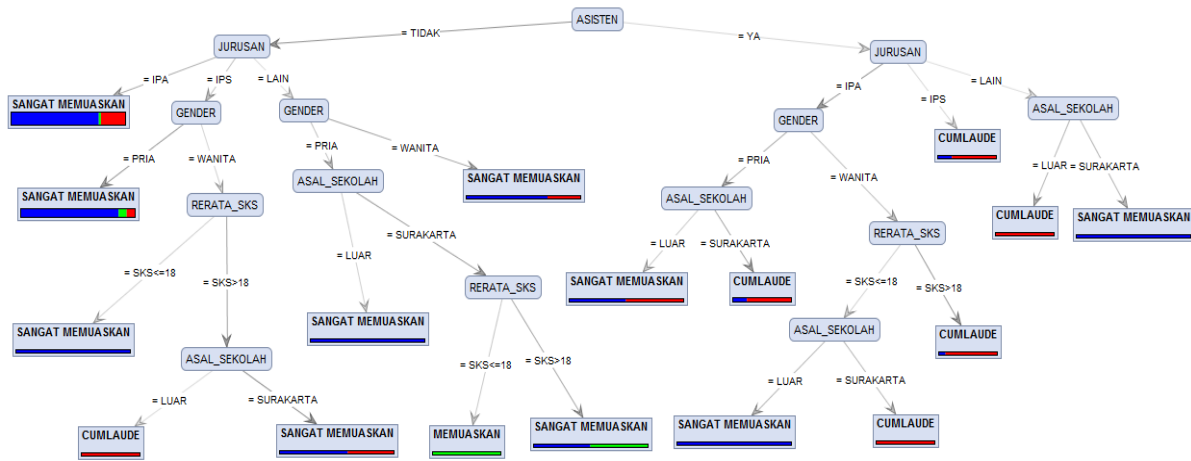
Hasil proses klasifikasi predikat kelulusan dengan metode *Decision Tree* ditunjukkan pada gambar 3. Berdasarkan hasil pohon keputusan pada gambar 3, dapat dilihat bahwa atribut yang memiliki pengaruh paling tinggi untuk menentukan klasifikasi predikat kelulusan mahasiswa adalah Asisten (X5). Hal ini ditunjukkan dengan variabel X5 menempati sebagai simpul akar (*root node*).

Contoh penerapan klasifikasi pohon keputusan pada sebuah data berdasarkan gambar 3 sebagai berikut jika seorang mahasiswa yang tidak pernah menjadi asisten, berasal dari jurusan IPA, maka mahasiswa tersebut diklasifikasikan memiliki predikat kelulusan SANGAT MEMUASKAN tanpa mempedulikan rata-rata SKS yang diambil tiap semester, asal sekolah, dan jenis kelaminnya. Akan tetapi

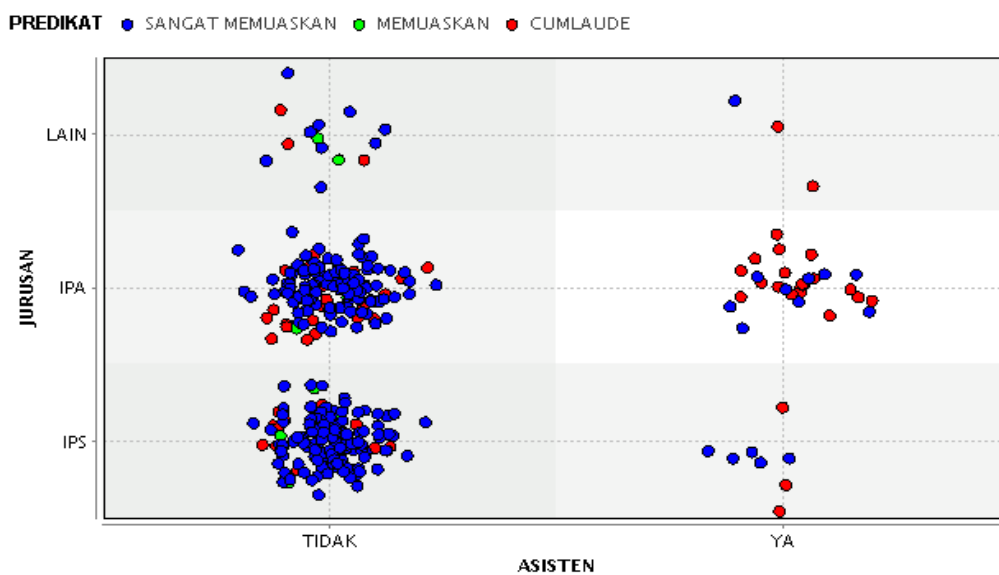
jika seorang mahasiswa yang pernah menjadi asisten dan berasal dari jurusan IPS, maka diklasifikasikan dalam kelompok mahasiswa yang berpredikat *Cumlaude*.

Selain klasifikasi predikat kelulusan mahasiswa dapat dilihat menggunakan pohon keputusan, contoh pola distribusi antar atribut terhadap predikat kelulusan dapat disajikan menggunakan grafik *scatter plot* gambar 4.

Pola distribusi masa studi berdasarkan gambar 4 menunjukkan bahwa seorang mahasiswa yang pernah menjadi asisten dan berasal dari jurusan IPA ketika masih sekolah memiliki probabilitas predikat kelulusan *Cumlaude* paling tinggi dibandingkan dengan mahasiswa yang tidak pernah menjadi asisten atau berasal dari jurusan selain IPA ketika masih sekolah.



Gambar 3. Pohon Keputusan untuk Klasifikasi Predikat Kelulusan



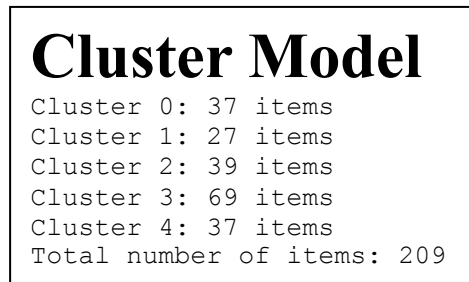
Gambar 4. Distribusi Predikat Kelulusan Menggunakan Pohon Keputusan

D. Klastering Mahasiswa

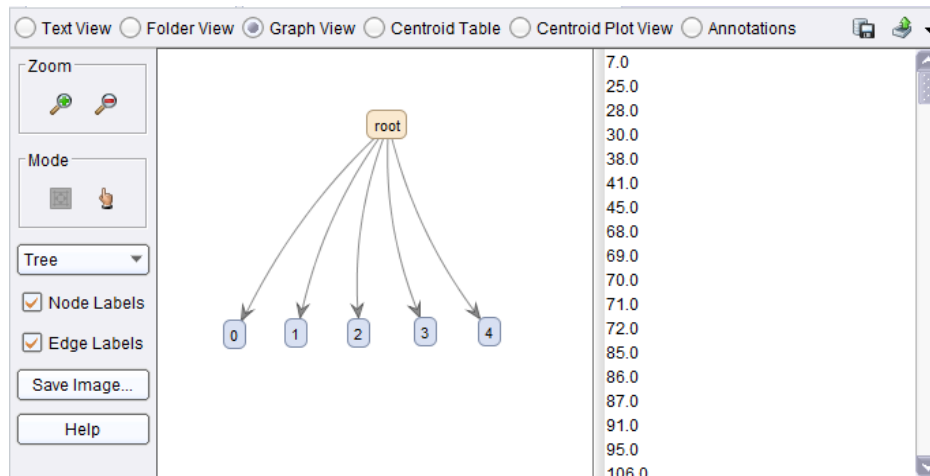
Proses klastering ini dilakukan untuk mengetahui pola kelompok mahasiswa informatika yang masih aktif berdasarkan nilai *Distance Performance* terhadap variabel-variabel yang diajukan. Klaster yang diinginkan dalam proses ini ditentukan sebanyak 5 klaster (nilai k). *Distance performance* dalam metode klastering dapat dihitung jika nilai pada setiap variabel

memiliki tipe numerik. Sehingga data yang digunakan dalam proses ini adalah data numerik.

Tahap ini menghasilkan sebuah model klaster (*cluster model*) untuk mengetahui kelompok-kelompok mahasiswa berdasarkan variabel-variabel bebas yang diajukan. Gambar 5 dan gambar 6 berikut adalah model klaster yang terbentuk.



Gambar 5. Cluster Model Text View Mahasiswa



Gambar 6. Cluster Model Graph View Mahasiswa

Gambar 5 menunjukkan bahwa kelompok mahasiswa berjumlah 5 kluster dengan masing-masing jumlah mahasiswa berbeda antara satu kluster dengan kluster yang lainnya. Sedangkan gambar 6 memperlihatkan contoh pembagian anggota setiap kluster yang dalam contoh ini adalah *Cluster_0*. Berdasarkan pembagian kluster tersebut dapat dilihat bahwa kluster yang memiliki anggota terbanyak adalah Cluster 3 yaitu sebanyak 69 dari 209 data, sedangkan yang memiliki anggota paling sedikit adalah Cluster 1 yaitu sebanyak 27 dari 209 data.

E. Rencana Strategis bagi Program Studi

Berikut beberapa kriteria mahasiswa yang bisa diterapkan sebagai sebuah kebijakan strategis bagi Program Studi Informatika berdasarkan interpretasi hasil penelitian.

A. Kriteria mahasiswa agar masa studi tepat waktu.

- 1) Jika Rerata SKS > 18 per semester
 - a) Gender = WANITA
 - b) Jika Gender = PRIA
 1. Mahasiswa tersebut harus berasal dari Jurusan IPA atau IPS
 2. Jika berasal dari Jurusan LAIN, maka perlu diwajibkan untuk menjadi ASISTEN.

Dengan demikian, program studi perlu memberikan perhatian atau motivasi yang lebih tinggi terhadap mahasiswa yang memiliki kriteria-kriteria selain tersebut di atas.

- 2) Jika Rerata SKS ≤ 18 per semester
 - a) Jika Jurusan = IPA, harus memiliki gender WANITA, pernah menjadi ASISTEN dan berasal sekolah dari LUAR SURAKARTA

- b) Jika Jurusan = LAIN, harus memiliki gender WANITA.

Jika seorang mahasiswa memiliki selain kriteria tersebut, maka dimungkinkan memiliki masa studi yang terlambat. Sehingga mahasiswa tersebut memerlukan perhatian dan motivasi yang lebih tinggi dari program studi.

B. Kriteria mahasiswa agar memiliki predikat kelulusan *Cumlaude*.

Hasil klasifikasi predikat kelulusan menunjukkan bahwa predikat kelulusan *Cumlaude* bisa diperoleh dengan syarat mahasiswa pernah menjadi seorang asisten. Jika tidak pernah menjadi asisten semasa kuliah, mahasiswa tersebut diklasifikasikan memiliki kemungkinan yang sangat kecil berpredikat *cumlaude*, kecuali berasal dari jurusan IPS, berjenis kelamin WANITA, memiliki rerata SKS > 18 per semester dan berasal sekolah dari LUAR SURAKARTA.

Selain mahasiswa harus menjadi asisten, berikut 4 kriteria atau persyaratan lain yang harus dipenuhi oleh mahasiswa untuk memperoleh predikat kelulusan *cumlaude* berdasarkan hasil penelitian:

- 1) Berasal dari jurusan IPS.
- 2) Jika berasal dari jurusan IPA ketika masih sekolah, harus berjenis kelamin PRIA dan berasal sekolah dari SURAKARTA. Jika memiliki jenis kelamin WANITA maka harus memiliki rerata SKS > 18 per semester.
- 3) Jika berasal dari jurusan LAIN, maka mahasiswa tersebut berasal sekolah dari LUAR SURAKARTA.

Dengan demikian mahasiswa yang tidak memiliki kriteria tersebut di atas, perlu diberi perhatian yang lebih dan motivasi yang tinggi dari program studi.

4. SIMPULAN

Berdasarkan hasil penelitian yang dilakukan maka dapat disimpulkan bahwa:

1. Telah diperoleh klasifikasi masa studi dan predikat kelulusan mahasiswa Program Studi Informatika UMS. Variabel yang paling tinggi pengaruhnya terhadap masa studi adalah rerata SKS

yang diambil per semester oleh mahasiswa, sedangkan variabel yang paling berpengaruh terhadap predikat kelulusan mahasiswa adalah partisipasi mahasiswa menjadi asisten.

2. Interpretasi hasil penelitian mengindikasikan bahwa variabel yang perlu digunakan sebagai pertimbangan bagi Program Studi Informatika UMS untuk memperoleh tingkat masa studi yang efektif adalah rerata SKS, sedangkan variabel sebagai pertimbangan untuk memperoleh predikat kelulusan yang maksimal adalah peran serta mahasiswa untuk menjadi asisten.
3. Proses klustering telah dihasilkan sebanyak 5 kelompok klaster yang terdiri dari *Cluster* 0 sampai dengan *Cluster* 4 yang masing-masing terdiri dari 37, 27, 39, 69, dan 37 mahasiswa dengan total 209 data mahasiswa.
4. Klaster yang memiliki anggota terbanyak adalah *Cluster* 3, sedangkan yang memiliki anggota paling sedikit adalah *Cluster* 1.

5. REFERENSI

- Ayub, Mewati, 2007. *Proses Data Mining dalam Sistem Pembelajaran Berbantuan Komputer*. Jurnal Sistem Informasi Vol. 2 No. 1 Maret 2007 : 21-30
- Bhardwaj, Ankit, Sharma, Arvind, Shrivastava, V.K. 2012. *Data Mining Techniques and Their Implementation in Blood Bank Sector - A Review*. International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622, Vol. 2, Issue4, July-August 2012, pp.1303-1309
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, P. J. 1984. *Classification and Regression Tree*. Belmont, CA: Wadsworth International Group.
- Karlinger, Fred, N. 1973. *Foundation of Behavior Science Research*. Holt, Rinehart.

- Lesmana, Dody Putu. 2012. *Perbandingan Kinerja Decision Tree J48 dan ID3 Dalam Pengklasifikasian Diagnosis Penyakit Diabetes Mellitus*. Jurnal Teknologi dan Informatika, Vol. 2, no. 2.
- Lin, S. H. 2012. *Data Mining for Student Retention Management*. Journal of Computer Science. Coll, 27(4), 92-99.
- Luan, J. (2002). *Data Mining and Knowledge Management in Higher Education Applications*. Paper presented at the Annual Forum for the Association for Institutional Research, Toronto, Ontario, Canada. <http://eric.ed.gov/ERICWebPortal/detail?accno=ED474143>
- Nugroho, Yusuf Sulisty. 2014. *Klasifikasi Masa Studi Mahasiswa Fakultas Komunikasi dan Informatika*. Jurnal Komunikasi dan Teknologi Informasi (KomuniTi) ISSN: 2087-085X, Volume VI No. I Maret 2014.
- Statuta Universitas Muhammadiyah Surakarta.
- Sunjaya. 2010. *Aplikasi Mining Data Mahasiswa dengan Metode Klasifikasi Decision Tree*. Seminar Nasional Aplikasi Teknologi Informasi 2010. Yogyakarta.