

**PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI PREDIKAT KELULUSAN
MAHASISWA FAKULTAS KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS MUHAMMADIYAH SURAKARTA**

Yusuf Sulistyono Nugroho¹

¹Jurusan Informatika, Fakultas Komunikasi dan Informatika, UMS

e-mail : yusuf.nugroho@ums.ac.id

ABSTRACT

The growth of database technology in educational system lead to sustainable and abundant students data continue to be generated. Meanwhile, the abundant data can be used for data mining as a source of strategic information in order to achieve better education management. Faculty of Communication and Informatics, Muhammadiyah University of Surakarta (FKI UMS) until the end of 2013 has had as many as 2358 students including those that have passed of approximately 700-800 students. If these data is only accumulated, it will become a burden database. This study was conducted to utilize the abundant data as strategic resources for faculty and department to classify the students' degree of excellence using data mining techniques. The students' degree of excellence was classified using the C4.5 algorithm. The number of samples was determined using the equation of Slovin. There are 341 students' data taken from the total 2358 of FKI students who have graduated as the data to be classified. Data processing was conducted on the separation of the attributes needed for data mining process, standardization of data (preprocessing), and the conversion of real data into nominal data. Attributes used consists of school major (equivalent to high school), gender, home schools, the average number of credits per semester, and assistant roles that are considered important in influencing students' degree of excellence. The result shows that the highest variable influencing students' degree of excellence is their participation as an assistant with the accuracy of 73.91%. The result of the study indicates that the variable to use as consideration for faculty to obtain maximum degree of excellence is student participation become an assistant.

Keywords: C4.5 algorithm, classification, data mining, decision tree, degree of excellence

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. Banyaknya data yang dimiliki oleh sebuah organisasi bisa menyebabkan kesulitan dalam pengklasifikasian data tersebut untuk kepentingan organisasi. Kegiatan pengklasifikasian yang dilakukan oleh manusia masih memiliki keterbatasan, terutama pada kemampuan manusia dalam menampung jumlah data yang ingin diklasifikasikan. Selain itu bisa juga terjadi kesalahan dalam pengklasifikasian yang dilakukan. Salah satu cara mengatasi masalah ini adalah dengan menggunakan *Data Mining* (DM) dengan teknik klasifikasi. Data mining dapat membantu sebuah organisasi yang memiliki data melimpah untuk memberikan informasi yang dapat mendukung pengambilan keputusan (Kiron et al, 2012).

Dalam dunia pendidikan, data yang berlimpah dan berkesinambungan mengenai siswa yang dibina dan alumni terus dihasilkan. Menurut Jing (2004) dan Merceron (2005) dalam Ayub (2007), data yang berlimpah membuka peluang diterapkannya data mining untuk pengelolaan pendidikan yang lebih baik dan data mining dalam pelaksanaan pembelajaran berbantuan komputer yang lebih efektif. Sementara itu, Luan (2002) menunjukkan bahwa data mining dapat digunakan untuk menyelesaikan siswa yang bermasalah dan membantu institusi menjadi lebih proaktif dalam mengidentifikasi dan merespon siswa tersebut. Luan menerapkan data mining sebagai cara untuk memprediksi ciri-ciri siswa yang akan dikeluarkan oleh sekolah dan kemudian kembali ke sekolah tersebut pada tahun berikutnya. Salah satu lembaga pendidikan yang cukup besar di Indonesia saat ini adalah Universitas Muhammadiyah Surakarta (UMS) yang memiliki 11 fakultas yang salah satunya adalah Fakultas Komunikasi dan Informatika.

Fakultas Komunikasi dan Informatika UMS sejak berdiri pada tahun 2006 hingga akhir tahun 2013 telah memiliki sebanyak 2358 mahasiswa termasuk yang sudah lulus sebanyak kurang lebih 700-800 mahasiswa. Dengan demikian data-data akademik mahasiswa yang ada juga cukup banyak. Jika data yang melimpah ini hanya dibiarkan menumpuk, maka hanya akan menjadi beban database yang dimiliki. Sementara itu, data-data yang melimpah ini sebenarnya bisa dimanfaatkan sebagai sumber informasi strategis bagi program studi untuk memprediksi masa studi dan predikat kelulusan mahasiswa dengan menggunakan teknik-teknik data mining. Hal ini tentunya selain dapat memberikan informasi yang bersifat strategis bagi fakultas dan program studi, juga dapat meningkatkan upaya untuk mendorong meningkatkan IPK mahasiswa. Sehingga selain dapat bermanfaat bagi mahasiswa sendiri, juga dapat meningkatkan nilai akreditasi bagi program studi.

Berdasarkan pada latar belakang tersebut, maka penelitian ini dilakukan untuk menerapkan teknik data mining dengan algoritma C4.5 guna memprediksi dan mengklasifikasi predikat kelulusan mahasiswa Fakultas Komunikasi dan Informatika UMS.

METODE PENELITIAN

Tahap pertama penelitian ini adalah studi literatur atau kepustakaan yang 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.

Tahap kedua penelitian ini adalah pemilihan obyek penelitian yang dilakukan untuk memprediksi dan mengklasifikasikan indeks prestasi kumulatif mahasiswa Fakultas Komunikasi dan Informatika (FKI) 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.

Tahap ketiga adalah penentuan variabel data mining. Variabel-variabel yang akan digunakan untuk proses data mining ini ditentukan berdasarkan tujuan penelitian. Lama studi dan predikat kelulusan mahasiswa sebagai variabel yang akan dicari pola pengelompokannya dan sebagai variabel yang akan diprediksi bagi mahasiswa aktif di FKI UMS. Ada 2 (dua) jenis variabel yang ditentukan dalam proses data mining ini. Variabel pertama adalah variabel dependen (Y) yang nilainya tergantung atau terikat berdasarkan nilai-nilai variabel lainnya. Variabel Y yang digunakan adalah Predikat Kelulusan Mahasiswa yang nilainya ditentukan berdasarkan IPK. Variabel kedua adalah variabel independen (X) yang nilainya tidak tergantung dari nilai-nilai variabel lainnya. Variabel X yang diperlukan terdiri dari pertama Jurusan Sekolah (SMA sederajat) sebagai X1. Atribut jurusan sekolah dipertimbangkan sebagai salah satu faktor penting yang dapat mempengaruhi predikat kelulusan. Hal ini dilihat dari mata kuliah yang ada di jurusan komunikasi yang sebagian besar linier dengan jurusan IPS, sedangkan mata kuliah yang ada di jurusan informatika lebih linier dengan jurusan IPA. Kedua Jenis Kelamin sebagai X2, dimana atribut jenis kelamin dipandang dapat mempengaruhi tingkat kepandaian dan juga kedisiplinan seseorang. Sehingga atribut ini dapat mempengaruhi nilai IPK mahasiswa ketika lulus. Ketiga asal Sekolah sebagai X3, dimana atribut ini dipilih sebagai salah satu variabel independen karena akan dipertimbangkan sebagai dasar penentuan wilayah-wilayah strategis yang akan digunakan untuk kegiatan promosi oleh universitas untuk mencari bibit unggul sebagai calon mahasiswa di UMS. Keempat rerata jumlah SKS per semester sebagai X4, dimana atribut rerata jumlah SKS digunakan sebagai variabel X4 dengan pertimbangan untuk mengetahui seberapa besar pengaruhnya terhadap nilai IPK mahasiswa. Kelima pernah menjadi asisten sebagai X5, dimana partisipasi mahasiswa sebagai asisten juga digunakan sebagai variabel independen untuk mengetahui apakah memiliki pengaruh yang signifikan terhadap nilai IPK mahasiswa. Jika atribut ini memiliki pengaruh yang signifikan maka dapat dijadikan dasar untuk menentukan kebijakan strategis oleh fakultas terhadap mahasiswa.

Tahap keempat adalah penentuan nilai class variabel. Berdasarkan variabel yang telah ditentukan, tahapan berikutnya adalah menentukan nilai-nilai class dari masing-masing variabel Y dan variabel X. Nilai *class* variabel Y didasarkan pada Statuta Universitas Muhammadiyah Surakarta, maka variabel Y dibedakan menjadi 3 nilai *class* yang bertipe *label*, yaitu: memuaskan, jika $2.00 \leq Y1 < 2.76$, sangat memuaskan, jika $2.76 \leq Y1 < 3.51$, cumlaude, jika $Y1 \geq 3.51$. Sedangkan nilai *class* variabel X terdiri dari variabel X1 dibedakan menjadi 3 nilai *class* yang bertipe *binomial*, yaitu: IPA,

IPS dan LAIN (selain IPA dan IPS), variabel X2 terdiri dari 2 nilai *class* dengan tipe *polynomial*, yaitu PRIA dan WANITA, Variabel X3 dibuat menjadi 2 nilai *class* yang bertipe *polynomial*, yaitu SURAKARTA (jika asal sekolah se-Karesidenan Surakarta) dan LUAR (jika asal sekolah di luar Karesidenan Surakarta), Variabel X4 dibagi menjadi 2 nilai *class* yang bertipe *polynomial*, yaitu SKS ≤ 18 dan SKS > 18 , Variabel X5 terdiri dari 2 nilai *class* dengan tipe *polynomial*, yaitu YA (jika pernah menjadi asisten) dan TIDAK (jika belum pernah menjadi asisten).

Tahap kelima adalah pengumpulan data. Banyaknya data mahasiswa yang diperoleh dari IT UMS dihitung dengan metode Slovin (terlihat pada persamaan 1) untuk menentukan jumlah sampel yang digunakan sebagai data pengujian proses data mining.

$$n = \frac{N}{1 + Ne^2} \dots\dots\dots (1)$$

Keterangan:

- n = jumlah sampel
- N = jumlah keseluruhan data / populasi
- e = galat kesalahan (ditentukan sebesar 5%)

Data mahasiswa Fakultas Komunikasi dan Informatika yang diperoleh dari IT UMS seluruhnya berjumlah (N) 2358 mahasiswa. Sesuai dengan rumus Slovin, dapat ditentukan jumlah sampel yang diambil yaitu sebanyak 341 data mahasiswa.

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

Tahap terakhir adalah melakukan analisis data. Klasifikasi mahasiswa berdasarkan predikat kelulusannya dilakukan dengan metode C4.5 berdasarkan entropi dari masing-masing atribut yang telah ditentukan dengan persamaan 2 dan 3. Lesmana (2012) menjelaskan bahwa *decision tree* merupakan salah satu metode belajar yang sangat populer dan banyak digunakan secara praktis. Metode ini merupakan metode yang berusaha menemukan fungsi-fungsi pendekatan yang bernilai diskrit dan tahan terhadap data – data yang memiliki kesalahan (*noisy data*) serta mampu mempelajari ekspresi – ekspresi disjunctive seperti ekspresi OR.

$$gain(y, A) = entropi(y) - \sum_{c \in nilai(A)} \frac{y_c}{y} entropi(y_c) \dots\dots\dots (2)$$

$$Entropi(y) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \dots - p_n \log_2 p_n \dots\dots\dots (3)$$

PEMBAHASAN

Data-data mahasiswa FKI yang diperoleh dari hasil pengumpulan data untuk keperluan data mining masih bersifat transaksional dan tercatat dalam format Microsoft Excel (terlihat pada Tabel 1).

Tabel 1. Potongan Data Mahasiswa Lulus

NIM	JUR SMA	GENDER	SMU	RATA SKS	ASIST	LAMA STUDI	IPK
L100090046	IPS	L	SMU Negeri 2 Sukoharjo	20	Tidak	6,00	3,33
L100090054	IPS	L	SMU Negeri 1 Boyolali	20	Tidak	6,00	3,51
L100090055	IPS	L	SMU Negeri 2 Sukoharjo	20	Tidak	6,00	3,43
L100090144	IPA	P	SMU Negeri 3 Boyolali	20	Tidak	6,00	3,73
L100090167	IPA	P	SMU Negeri 1 Magetan	20	Tidak	6,00	3,54
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 Sukarta	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

Kelas data yang digunakan untuk data mining disiapkan (*preprocessing*) sehingga memiliki kelas *binomial* atau *polynomial* sesuai aturan yang telah dibuat berdasarkan nilai datanya. Tabel 2 merupakan pembagian variabel dan kelas data yang digunakan dalam analisis data mining.

Tabel 2. Pembagian Variabel dan Kelas Data

Variabel	Nama Field	Jenis Kelas Data	Kelas data yang digunakan
Y	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 ≤ 18, SKS > 18
X5	Asisten	Binomial	YA, TIDAK

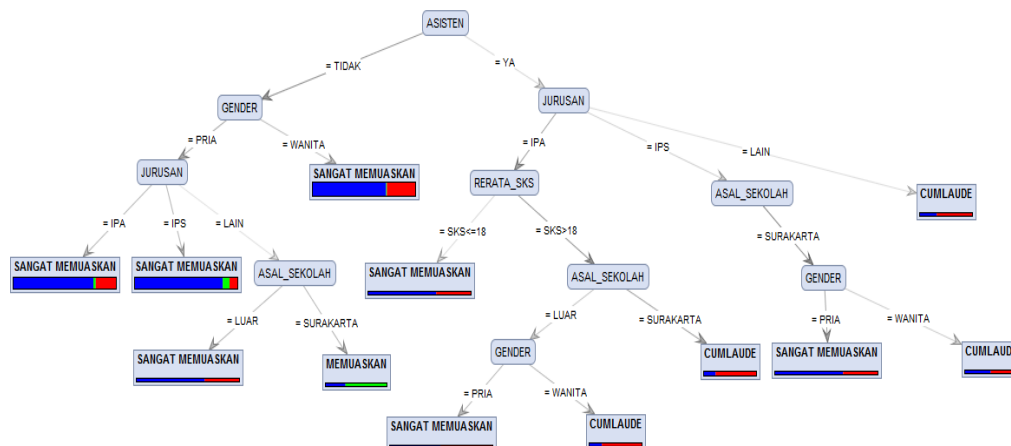
Potongan data hasil *preprocessing* sesuai dengan jenis kelas datanya dapat dilihat pada Tabel 3.

Tabel 3. Potongan Hasil Preprocessing Data Mahasiswa

JURUSAN	GENDER	ASAL SEKOLAH	RERATA SKS	ASISTEN	LAMA STUDI	PREDIKAT
IPS	PRIA	SURAKARTA	SKS >18	TIDAK	TERLAMBAT	SANGAT MEMUASKAN
IPS	PRIA	SURAKARTA	SKS >18	TIDAK	TERLAMBAT	SANGAT MEMUASKAN
IPS	PRIA	SURAKARTA	SKS >18	TIDAK	TERLAMBAT	SANGAT MEMUASKAN
IPA	WANITA	SURAKARTA	SKS >18	TIDAK	TERLAMBAT	CUMLAUDE
IPA	WANITA	LUAR	SKS >18	TIDAK	TERLAMBAT	CUMLAUDE
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

Data yang digunakan untuk proses klasifikasi predikat kelulusan menggunakan *decision tree* sebanyak 341 data untuk mengetahui pola atau tren mahasiswa FKI yang dapat menyelesaikan studinya dengan predikat kelulusan *cumlaude*, sangat memuaskan atau memuaskan berdasarkan variabel-variabel yang diajukan.

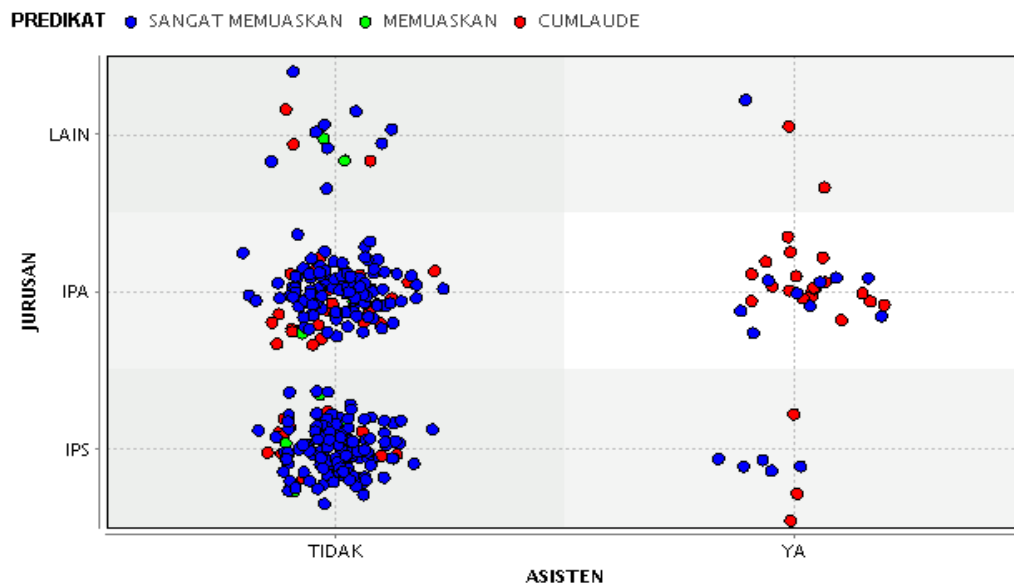
Hasil proses klasifikasi predikat kelulusan dengan metode *Decision Tree* ditunjukkan pada Gambar 1. Berdasarkan hasil pohon keputusan pada Gambar 1, dapat dilihat bahwa atribut yang memiliki pengaruh paling tinggi untuk menentukan klasifikasi predikat kelulusan mahasiswa adalah Asisten (X5). Hal ini ditunjukkan dengan variabel X5 menempati sebagai simpul akar (*root node*).



Gambar 1. Pohon Keputusan untuk Klasifikasi Predikat Kelulusan

Salah satu interpretasi hasil penerapan klasifikasi pohon keputusan pada sebuah data berdasarkan Gambar 1 adalah jika seorang mahasiswa yang pernah menjadi asisten, berasal dari jurusan IPA, memiliki rata-rata SKS yang diambil tiap semester adalah lebih dari 18 SKS, dan berasal sekolah dari Surakarta, maka mahasiswa tersebut dapat diklasifikasikan memiliki predikat kelulusan *Cumlaude*, tanpa mempedulikan jenis kelaminnya.

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 2.



Gambar 2. Distribusi Predikat Kelulusan Menggunakan Pohon Keputusan

Pola distribusi predikat kelulusan berdasarkan Gambar 2 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.

Pengujian terhadap data pengujian menghasilkan sebuah klasifikasi predikat kelulusan mahasiswa. Hasil tersebut dapat dijadikan sebagai sebuah informasi strategis yang dapat diubah menjadi sebuah pengetahuan (*knowledge*). Pengetahuan inilah yang bisa digunakan sebagai pendukung suatu keputusan atau kebijakan strategis bagi suatu organisasi.

Hasil klasifikasi menunjukkan bahwa predikat kelulusan *Cumlaude* bisa diperoleh dengan syarat utama adalah pernah menjadi seorang asisten, meskipun menjadi asisten tidak menjamin berpredikat *cumlaude*. Jika tidak pernah menjadi asisten semasa kuliah, mahasiswa tersebut dapat diklasifikasikan berpredikat kelulusan sangat memuaskan atau bahkan hanya memuaskan.

Selain mahasiswa harus menjadi asisten, berikut 4 kriteria atau persyaratan lain yang harus dipenuhi oleh mahasiswa untuk memperoleh predikat kelulusan *cumlaude* berdasarkan hasil penelitian syarat pertama jika berasal dari jurusan IPA ketika masih sekolah menengah atas, rerata SKS yang diambil harus lebih dari 18 SKS per semester dan berasal sekolah dari Surakarta. Syarat kedua adalah jika syarat pertama terpenuhi namun berasal sekolah dari luar Surakarta, maka mahasiswa tersebut harus berjenis kelamin wanita. Syarat ketiga jika berasal dari jurusan IPS, maka mahasiswa tersebut berasal sekolah dari Surakarta dan berjenis kelamin wanita. Syarat keempat jika berasal dari jurusan selain IPA dan IPS, maka tidak perlu mempertimbangkan variabel lainnya. Karena mahasiswa tersebut diklasifikasikan memiliki predikat kelulusan *cumlaude* apapun nilai variabel lainnya.

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

KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan maka dapat disimpulkan bahwa telah diperoleh klasifikasi predikat kelulusan mahasiswa Fakultas Komunikasi dan Informatika UMS. Variabel yang paling tinggi pengaruhnya terhadap terhadap predikat kelulusan adalah partisipasi mahasiswa menjadi asisten. Interpretasi hasil penelitian mengindikasikan bahwa variabel yang perlu digunakan sebagai pertimbangan bagi Fakultas Komunikasi dan Informatika UMS untuk memperoleh tingkat predikat kelulusan yang maksimal adalah peran serta mahasiswa untuk menjadi asisten. Secara umum probabilitas predikat “Cumlaude” pada kelompok mahasiswa yang pernah menjadi asisten lebih tinggi dibandingkan dengan yang tidak pernah menjadi asisten. Seorang mahasiswa dari kelompok yang pernah menjadi asisten jika berasal dari jurusan IPA semasa sekolah menengah atas memiliki probabilitas predikat kelulusan “Cumlaude” yang lebih tinggi dibandingkan dengan mahasiswa dari jurusan lainnya.

DAFTAR PUSTAKA

- Ayub, Mewati, 2007, Proses Data Mining dalam Sistem Pembelajaran Berbantuan Komputer, *Jurnal Sistem Informasi*, Vol. 2 No. 1, Maret 2007, hal. 21-30.
- Kiron, D., Shockley, R., Kruschwitz, N., Finch, G., & Haydock, M., 2012, *Analytics: The Widening Divide*. MIT Sloan Management Review, 53(2), 1-22.
- 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.
- 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>.
- Statuta Universitas Muhammadiyah Surakarta.