

# KLASIFIKASI MASA STUDI MAHASISWA FAKULTAS KOMUNIKASI DAN INFORMATIKA UNIVERSITAS MUHAMMADIYAH SURAKARTA MENGUNAKAN ALGORITMA C4.5

*Yusuf Sulisty Nugroho<sup>1</sup>, Setyawan<sup>2</sup>*

<sup>1,2</sup>Jurusan Teknik Informatika Fakultas Komunikasi dan Informatika UMS

Jl.A.Yani Tromol Pos I Pabelan, Kartasura, Surakarta

Email : yusuf.nugroho@ums.ac.id

## ABSTRAK

Saat ini dalam dunia pendidikan data yang berlimpah dan berkelanjutan bisa dimanfaatkan untuk data mining dalam rangka pengelolaan yang lebih baik dan pelaksanaan pembelajaran yang lebih efektif. Salah satunya adalah Fakultas Komunikasi dan Informatika (FKI) UMS yang telah berdiri sejak tahun 2006 dan telah memiliki sebanyak 2358 mahasiswa termasuk yang sudah lulus sebanyak kurang lebih 600-700 mahasiswa. Penelitian ini dilakukan untuk memanfaatkan data-data yang melimpah tersebut sebagai sumber informasi strategis bagi fakultas untuk mengklasifikasi masa studi mahasiswa dengan menggunakan teknik data mining. Klasifikasi masa studi terhadap data lulusan mahasiswa FKI UMS menggunakan metode *Decision Tree* dengan algoritma C4.5. Dari total 2358 data diambil sebanyak 341 data mahasiswa yang sudah lulus. Atribut yang digunakan terdiri dari jurusan sekolah, jenis kelamin, asal sekolah, rerata jumlah SKS per semester, dan peran menjadi asisten. Hasil penelitian menunjukkan bahwa variabel yang paling tinggi pengaruhnya terhadap masa studi mahasiswa adalah rerata SKS per semester. Dengan demikian, interpretasi hasil penelitian mengindikasikan bahwa variabel yang perlu digunakan sebagai pertimbangan bagi fakultas untuk memperoleh tingkat masa studi yang efektif adalah rerata SKS yang diambil oleh mahasiswa.

**Kata kunci:** *data mining, decision tree, klasifikasi, masa studi*

## A. 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.

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. 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 telah memiliki sebanyak 2358 mahasiswa termasuk yang sudah lulus sebanyak kurang lebih 600-700 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 melakukan klasifikasi masa studi lulusan mahasiswa dengan menggunakan teknik *data mining*. Hal ini tentunya selain dapat memberikan informasi yang bersifat strategis bagi fakultas dan program studi, juga dapat meningkatkan upaya untuk mendorong dan mempercepat kelulusannya. Sehingga selain dapat bermanfaat bagi mahasiswa sendiri, juga dapat meningkatkan nilai akreditasi bagi program studi.

Berdasarkan pada latar belakang tersebut, maka penelitian ini perlu dilakukan untuk

menerapkan teknik *data mining* guna mengklasifikasi masa studi mahasiswa Fakultas Komunikasi dan Informatika UMS.

## B. METODOLOGI

### 1. Studi Literatur atau Kepustakaan.

Penelitian ini 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.

### 2. Pemilihan Obyek Penelitian.

Penelitian ini dilakukan untuk mengklasifikasikan masa studi 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*.

### 3. Penentuan Variabel *Data Mining*.

Variabel-variabel yang akan digunakan untuk proses *data mining* ini ditentukan berdasarkan tujuan penelitian. Masa studi mahasiswa digunakan sebagai variabel yang akan dicari pola pengelompokannya.

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 yaitu Masa Studi Mahasiswa.

b) Variabel independen (X)  
 Variabel independen (X) adalah variabel yang nilainya tidak tergantung 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

### C. 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 Y memiliki 2 nilai *class* yang bertipe *label*, yaitu:
  - a. Tepat Waktu, jika  $Y1 < 5$  tahun
  - b. Terlambat, jika  $Y1 \geq 5$  tahun

#### b) Nilai Class Variabel X

- 1) Variabel X1 dibedakan menjadi 3 nilai *class* yang bertipe *binomial*, yaitu: IPA, IPS dan LAIN (selain IPA dan IPS).
- 2) Variabel X2 terdiri dari 2 nilai *class* dengan tipe *polynomial*, yaitu PRIA dan WANITA
- 3) 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)

- 4) Variabel X4 dibagi menjadi 2 nilai *class* yang bertipe *polynomial*, yaitu  $SKS \leq 18$  dan  $SKS > 18$
- 5) Variabel X5 terdiri dari 2 nilai *class* dengan tipe *polynomial*, yaitu YA (jika pernah menjadi asisten) dan TIDAK (jika belum pernah menjadi asisten).

### D. PENGUMPULAN DATA.

Banyaknya data mahasiswa yang diperoleh dari IT UMS dihitung dengan metode Slovin untuk menentukan jumlah sampel yang digunakan sebagai data pelatihan dan data pengujian proses data mining.

Metode slovin dalam Nugroho (2009) ditunjukkan pada persamaan 1.

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

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.

### E. OLAH DATA

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

F. ANALISIS DATA

Tahapan analisis dilakukan untuk menentukan klasifikasi mahasiswa berdasarkan masa studi dan predikat kelulusannya.

1) **Klasifikasi Pohon Keputusan**

Klasifikasi mahasiswa berdasarkan masa studi dan predikat kelulusannya dilakukan dengan metode *Decision Tree* dengan penentuan atributnya menggunakan *information gain* berdasarkan entropi dari masing-masing atribut yang telah ditentukan dengan persamaan 2 dan 3.

$$gain(y, A) = entropi(y) -$$

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

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

G. HASIL DAN PEMBAHASAN

*Data Collecting dan Preprocessing*

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 Lulusan Mahasiswa

NIM	JUR SMA	GENDER	SMU	RATA SKS	ASIST	lama studi
L100090046	IPS	L	SMU Negeri 2 Sukoharjo	20	TIDAK	6,00
L100090054	IPS	L	SMU Negeri 1 Boyolali	20	TIDAK	6,00
L100090055	IPS	L	SMU Negeri 2 Sukoharjo	20	TIDAK	6,00
L100090144	IPA	P	SMU Negeri 3 Boyolali	20	TIDAK	6,00
L100090167	IPA	P	SMU Negeri 1 Magetan	20	TIDAK	6,00
L200070003	IPA	L	SMU Negeri 1 Ngrambe	19	TIDAK	5,00
L200070005	IPA	L	SMTA Lain-lain	21	YA	4,40
L200070006	IPS	L	SMU Negeri 1 Kebakkramat	18	TIDAK	5,90
L200070008	IPS	P	SMU Negeri 5 Surakarta	18	TIDAK	5,00
L200070009	IPA	L	SMTA Lain-lain	19	YA	4,90
L200070010	LAIN	L	SMK Negeri 2 Surakarta	19	TIDAK	5,00
L200070011	IPA	L	SMTA Lain-lain	17	TIDAK	5,90
L200070012	IPA	P	MA Negeri Sragen	17	TIDAK	4,40

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. Sedangkan potongan data hasil *preprocessing* sesuai dengan jenis kelas datanya dapat dilihat pada tabel 3.

Tabel 2. Pembagian Variabel dan Kelas Data

Var.	Nama Field	Jenis Kelas Data	Kelas data yang digunakan
Y	Lama Studi	Binomial	TEPAT, TERLAMBAT
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 3. Potongan Hasil Preprocessing Data Mahasiswa Lulus

JURUSAN	GENDER	ASAL_SEKOLAH	RERATA_SKS	ASISTEN	LAMA_STUDI
IPS	PRIA	SURAKARTA	SKS>18	TIDAK	TERLAMBAT
IPS	PRIA	SURAKARTA	SKS>18	TIDAK	TERLAMBAT
IPS	PRIA	SURAKARTA	SKS>18	TIDAK	TERLAMBAT
IPA	WANITA	SURAKARTA	SKS>18	TIDAK	TERLAMBAT
IPA	WANITA	LUAR	SKS>18	TIDAK	TERLAMBAT
IPA	PRIA	LUAR	SKS>18	TIDAK	TERLAMBAT
IPA	PRIA	SURAKARTA	SKS>18	YA	TEPAT
IPS	PRIA	SURAKARTA	SKS $\leq$ 18	TIDAK	TERLAMBAT
IPS	WANITA	SURAKARTA	SKS $\leq$ 18	TIDAK	TERLAMBAT
IPA	PRIA	SURAKARTA	SKS>18	YA	TEPAT
LAIN	PRIA	SURAKARTA	SKS>18	TIDAK	TERLAMBAT
IPA	PRIA	LUAR	SKS $\leq$ 18	TIDAK	TERLAMBAT
IPA	WANITA	SURAKARTA	SKS $\leq$ 18	TIDAK	TEPAT

## H. KLASIFIKASI MASA STUDI

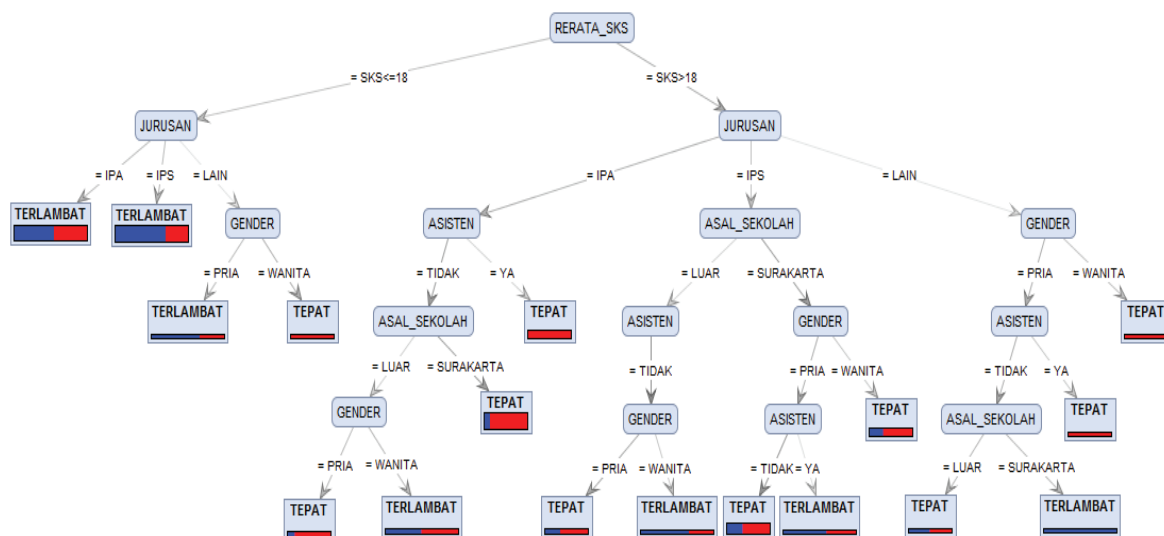
### MENGGUNAKAN DECISION TREE

Data yang digunakan untuk proses klasifikasi menggunakan *decision tree* hanya data sebanyak 341 data. Hal ini dikarenakan klasifikasi *decision tree* diproses untuk mengetahui pola atau tren mahasiswa FKI yang menyelesaikan masa studinya dengan tepat atau terlambat berdasarkan variabel-variabel yang diajukan.

Hasil proses klasifikasi masa studi 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

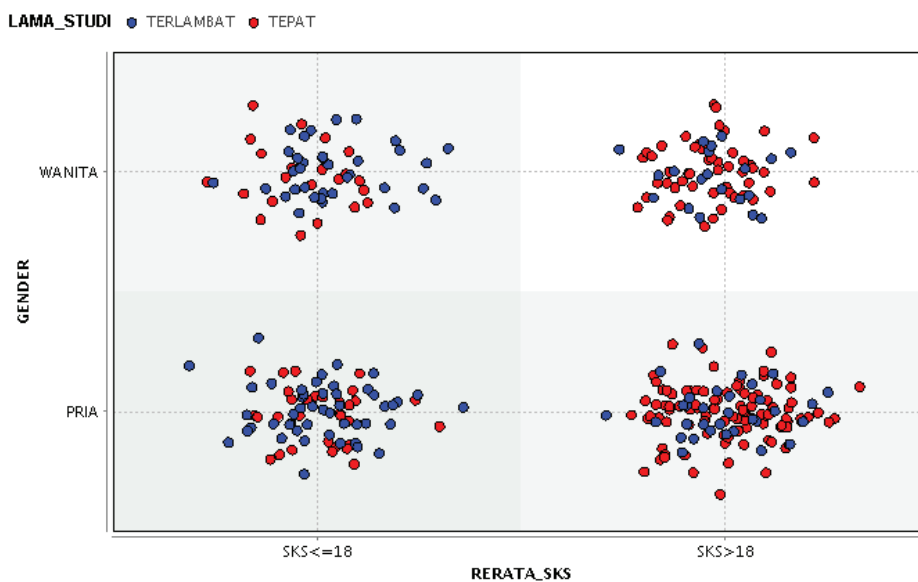
menentukan klasifikasi masa studi mahasiswa adalah Rerata SKS (X4). Hal ini ditunjukkan dengan variabel X4 yang menempati sebagai simpul akar (*root node*).

Contoh penerapan klasifikasi pohon keputusan pada sebuah data berdasarkan gambar 1 sebagai berikut jika seorang mahasiswa yang berasal sekolah dari Surakarta, jurusan IPA dan memiliki rata-rata SKS yang diambil tiap semester adalah lebih dari 18 SKS, tetapi belum pernah menjadi asisten, maka mahasiswa tersebut tetap diklasifikasikan memiliki masa studi TEPAT WAKTU, tanpa mempedulikan jenis kelaminnya.



Gambar 1. Pohon Keputusan untuk Klasifikasi Masa Studi

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



Gambar 2. Distribusi Masa Studi Menggunakan Pohon Keputusan

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.

### I. RENCANA STRATEGIS BAGI FAKULTAS

Pengujian terhadap data lulusan mahasiswa menggunakan metode *decision tree* menghasilkan sebuah pohon klasifikasi. 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.

Berikut beberapa kriteria mahasiswa yang bisa diterapkan sebagai sebuah kebijakan strategis bagi Fakultas Komunikasi dan Informatika berdasarkan interpretasi hasil penelitian.

1. Jika Rerata SKS  $> 18$  per semester
  - 1) Jika Jurusan Sekolah = IPA
    - a) Mahasiswa tersebut harus menjadi asisten
    - b) Jika tidak menjadi asisten, maka perlu diprioritaskan bagi mahasiswa yang berasal sekolah dari Surakarta.
  - 2) Jika Jurusan Sekolah = IPS
    - a) Jika asal sekolah dari Surakarta, maka mahasiswa yang diprioritaskan adalah berjenis kelamin wanita
    - b) Jika asal sekolah dari luar Surakarta, maka mahasiswa yang diprioritaskan adalah berjenis kelamin pria.
  - 3) Jika Jurusan Sekolah = Lainnya
    - a) Mahasiswa yang diprioritaskan adalah berjenis kelamin wanita.
    - b) Jika mahasiswa berjenis kelamin pria, maka yang diprioritaskan adalah yang pernah menjadi asisten.

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

2. Jika Rerata SKS  $\leq 18$  per semester

Berdasarkan hasil penelitian, jika rerata SKS yang diambil adalah kurang dari 18 SKS per semester, maka mahasiswa yang bisa lulus tepat waktu berdasarkan klasifikasi adalah memiliki kriteria berasal dari jurusan sekolah selain IPA dan IPS, dan berjenis kelamin wanita. Jika seorang mahasiswa memiliki selain kriteria tersebut, maka akan memiliki masa studi yang terlambat. Sehingga mahasiswa tersebut memerlukan perhatian dan motivasi yang lebih tinggi dari fakultas atau program studi masing-masing.

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.

## J. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan maka dapat disimpulkan bahwa:

1. Telah diperoleh klasifikasi masa studi mahasiswa lulusan Fakultas Komunikasi dan Informatika UMS. Variabel yang paling tinggi pengaruhnya terhadap masa studi adalah rerata SKS yang diambil per semester oleh mahasiswa.
2. Interpretasi hasil penelitian mengindikasikan bahwa variabel yang perlu digunakan sebagai pertimbangan bagi Fakultas Komunikasi dan Informatika UMS untuk memperoleh tingkat masa studi yang efektif adalah rerata SKS.

## DAFTAR PUSTAKA

- Ayub, Mewati, 2007. *Proses Data Mining dalam Sistem Pembelajaran Berbantuan Komputer*. Jurnal Sistem Informasi Vol. 2 No. 1 Maret 2007 : 21-30
- 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. 2009. *Analisis faktor-faktor yang mempengaruhi tingkat daya beli konsumen terhadap listrik pada sektor rumah tangga:: Studi kasus Kota Salatiga*. Thesis, Universitas Gadjah Mada, Yogyakarta.
- Statuta Universitas Muhammadiyah Surakarta.
- Sunjaya. 2010. *Aplikasi Mining Data Mahasiswa dengan Metode Klasifikasi Decision Tree*. Seminar Nasional Aplikasi Teknologi Informasi 2010. Yogyakarta.