

# PREDIKSI KELULUSAN MAHASISWA DENGAN METODE ALGORITMA C4.5

*Jajam Haerul Jaman*

Universitas Singaperbangsa Karawang  
Jl. H.S Ronggowaluyo Telukjambe Timur Karawang  
Jajam.haeruljaman@staff.unsika.ac.id

## ABSTRACT

*The world of education is a valuable asset for the development of science that continues to grow and constantly require appropriate analyzes as a current evaluation of materials for the development of education itself, in this case I tried to do research in predicting graduation with C45 algorithm method, while the research object is a student, the results obtained turned out its outline is the effect of gender on very high graduation, in this case the female gender who received a high percentage compared to male gender.*

**Keywords:** *Algoritma C45, Data Mining, Decision Tree*

## PENDAHULUAN

Dalam dunia pendidikan mahasiswa merupakan objek utama dari setiap penyelenggaraan pendidikan yang senantiasa mengedepankan kedisiplinan ilmu yang bermanfaat untuk mahasiswa itu sendiri ataupun orang lain, biasanya banyak hal yang terjadi kepada mahasiswa ketika proses pendidikan itu dilaksanakan, mulai ketidak patuhan akan aturan, ketidak nyamanan dalam belajar, ketidak seriusan dalam belajar, kemampuan dan bakat yang dimiliki yang tidak menunjang pada bidang yang diterimanya, kesenjangan sosial, ketidak mampuan secara *financial*, dan lain-lain, itu semua menyebabkan terjadinya ketidak konsistenan dalam proses belajar, sehingga dampak yang ditimbulkannya cukup kompleks, mulai dari kesiapan untuk menerima materi yang kurang, banyaknya mahasiswa yang mengulang, drop out dan berhenti sebelum lulus.

Dari dasar tersebutlah penulis mencoba melakukan sebuah penelitian dalam memprediksi kelulusan mahasiswa sebagai bahan pertimbangan dan bahan untuk melakukan kebijakan akademik, sehingga dapat menekan tingkat putus kuliah, drop out dan mengulang pada mahasiswa di universitas singaperbangsa karawang.

Data yang diperoleh penulis adalah data akademik mahasiswa dari tahun ketahunnya adalah mengambil dari data historis dan biodata awal penerimaan mahasiswa, dari data-data tersebutlah penulis melakukan prediksi dengan menggunakan algoritma C45 sebagai metode untuk mendapatkan hasil sesuai dengan keinginan

### Identifikasi Masalah

- Banyaknya mahasiswa yang putus kuliah, drop out dan mengulang
- Persentase kelulusan masih kurang dan perlu ditingkatkan

### Rumusan Masalah

Berdasarkan identifikasi masalah tersebut maka penulis mencoba melakukan prediksi kelulusan mahasiswa berdasarkan data awal yang diperoleh dengan menggunakan metode klasifikasi algoritma C45, sehingga dari metode klasifikasi tersebut dapat diperoleh klasifikasi mahasiswa yang menjadi objek penelitian ini.

### Tujuan Penelitian

Dengan menggunakan Metode Klasifikasi dengan konsep *Algoritma C4.5* diharapkan dapat meningkatkan keberhasilan dan keinginan lembaga untuk memberikan pikiran, pandangan, dan kebijakan baru kepada mahasiswa yang benar-benar membutuhkan bantuan tersebut, dengan kata

lain memaksimalkan kinerja lembaga kepada mahasiswa dalam upaya peningkatan persentase kelulusan mahasiswa.

### Manfaat Penelitian

Penelitian ini dapat berguna bagi peneliti sendiri khususnya, bagi lembaga penyelenggara pendidikan, maupun pihak-pihak lain yang berkepentingan dengan masalah yang ada.

- Bagi Peneliti  
Penelitian ini berguna untuk mengaplikasikan ilmu pengetahuannya untuk membantu stackholder dalam memahami permasalahan yang ada
- Bagi Lembaga Pendidikan  
Kegunaan penelitian bagi Lembaga Pendidikan adalah sebagai bahan untuk evaluasi kinerja Lembaga Pendidikan yang bersangkutan dalam menghadapi era globalisasi yang sangat kompleks dan penuh dengan persaingan.

## METODOLOGI PENELITIAN

Sebenarnya, desain riset adalah struktur konseptual di mana riset diperkenalkan; itu melembagakan koleksi untuk cetakbiru, pengukuran dan analisa data. dengan demikian desain riset termasuk di dalamnya satu garis besar dari apa yang peneliti itu akan cari hipotesisnya dan implikasi operasional kepada analisa akhir dari data. lebih tegasnya, keputusan-keputusan design menyangkut : akan melakukan studi mengenai apa ?, kenapa studi tersebut di lakukan ?, dimana studi baru itu akan dilaksanakan?, seperti apa data yang diperlukan?, dimana kebutuhan data tersebut ditemukan?, berapa periode waktu yang dibutuhkan?, apa yang merupakan rancangan contoh?, menggunakan teknik apa untuk mengkoleksi data?, bagai mana data itu nanti akan diteliti?, dengan gaya apa laporan tersebut dibuat? (C.R.Kothari, 2004)

Dari keadaan tersebut maka memang dibutuhkan tahap-tahap untuk menyusun sebuah metode penelitian yang sebagaimana diatas disebutkan, dalam tahap desain riset ini penulis akan menggunakan konsep eksperimen.

Kebenaran tentang suatu proses pengujian dari hipotesis statistik itu berkaitan dengan beberapa masalah riset dan itu dikenal dengan suatu eksperimen, sebagai contoh, kita dapat melakukan suatu eksperimen untuk menguji akurasi dari suatu kegunaan sistem yang sudah berjalan. Ada dua macam eksperimen yaitu:

- a. eksperimen komparatif
- b. eksperimen absolut.

Ketika kita ingin tahu mengenai dampak yang ditimbulkan dari internet terhadap anak maka itu merupakan eksperimen absolut, tetapi jika kita ingin menentukan perbandingan dampak yang ditimbulkan oleh internet dan juga media digital konvensional seperti Tv, Radio, dan Koran, maka itu disebut dengan eksperimen komparatif (C.R.Kothari, 2004).

## HASIL DAN PEMBAHASAN

### Algoritma C4.5

Dari tabel training diketahui bahwa jumlah kasus 46 kasus, mahasiswa yang termasuk kelas lulus berjumlah 22 Record, dan mahasiswa yang termasuk kelas tidak lulus berjumlah 24 record sehingga didapatkan *Entropy* nya sebagai berikut :

$$Entropy(S) = \sum_{i=1}^n - p_i \cdot \log_2 p_i$$

$$Entropy(S) = (-22/46 \cdot \log_2(22/46)) + (-24/46 \cdot \log_2(24/46))$$

$$Entropy(S) = 0,99864$$

Menghitung nilai *Gain* dari tiap-tiap atribut yang kemudian *Gain* tertinggi akan dijadikan sebagai akar dari pohon keputusan yang akan dibuat, sebagai contoh pada atribut gender diperoleh nilai *Gain* sebesar 0,058451942

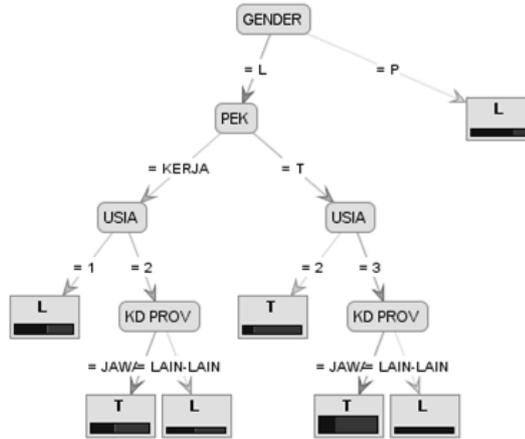
$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Tabel 1. Sample Pernghitungan Atribut untuk mendapatkan nilai gain

TEMPAT LAHIR							
NODE	JML	L	T	Entropy	Gain	L	T
BANDUNG	6	3	3	1		50%	50%
BEKASI	1	1	0	0		100%	0%
BENGKULU	1	0	1	0		0%	100%
BREBES	1	0	1	0		0%	100%
CIANJUR	2	1	1	1		50%	50%
CIREBON	1	0	1	0		0%	100%
JAKARTA	2	1	1	1		50%	50%
KARAWANG	24	13	11	0.994985		54%	46%
LAMPUNG TGH	1	0	1	0		0%	100%
PABUARAN	1	0	1	0		0%	100%
PAJAR BULAN	1	1	0	0		100%	0%
PEKALONGAN	1	1	0	0		100%	0%
PEMALANG	1	0	1	0		0%	100%
PURBALINGGA	1	0	1	0		0%	100%
TASIKMALAYA	2	1	1	1		50%	50%
	<b>46</b>	<b>22</b>	<b>24</b>		<b>0.21864388</b>	48%	52%
USIA							
NODE	JML	L	T	Entropy	Gain	L	T
1	7	4	3	0.985228		57%	43%
2	16	7	9	0.988699		44%	56%
3	23	11	12	0.998636		48%	52%
4	0	0	0			#DIV/0!	#DIV/0!
	<b>46</b>	<b>22</b>	<b>24</b>		<b>0.005496515</b>	48%	52%
GENDER							
NODE	JML	L	T	Entropy	Gain	L	T
L	40	17	23	0.983708		43%	58%
P	6	5	1	0.650022		83%	17%
	<b>46</b>	<b>22</b>	<b>24</b>		<b>0.058451942</b>	48%	52%
PROVINSI							
NODE	JML	L	T	Entropy	Gain	L	T
JAWA BARAT	38	18	20	0.998001		47%	53%
LAIN-LAIN	8	4	4	1		50%	50%
	<b>46</b>	<b>22</b>	<b>24</b>		<b>0.000287408</b>	48%	52%
PEKERJAAN							
NODE	JML	L	T	Entropy	Gain	L	T
BEKERJA	16	8	8	1		50%	50%

TIDAK	30	14	16	0.996792		47%	53%
	<b>46</b>	<b>22</b>	<b>24</b>		<b>0.000728378</b>	48%	52%

Dari hasil perhitungan nilai entropy dan gain dari tabel 1 maka didapatkan hasil gain tertinggi adalah gain gender dengan 0,58451942 oleh karena itu maka gain gender adalah gain untuk menentukan simpul akar berikutnya, dan setelah semua gain diperoleh maka hasil pohon keputusan seperti yang tampak pada gambar 1.



Gambar 1. Decision Tree Rapid Miner

```

Tree

GENDER = L
| PEK = KERJA
| | USIA = 1: L {L=4, T=3}
| | USIA = 2
| | | KD PROV = JAWA BARAT: T {L=3, T=4}
| | | KD PROV = LAIN-LAIN: L {L=1, T=1}
| | PEK = T
| | | USIA = 2: T {L=1, T=4}
| | | USIA = 3
| | | | KD PROV = JAWA BARAT: T {L=5, T=11}
| | | | KD PROV = LAIN-LAIN: L {L=2, T=0}
GENDER = P: L {L=3, T=1}
  
```

Gambar 2. Decision Tree Rapid Miner Text View

**KESIMPULAN**

Dalam penelitian ini penulis melakukan model penelitian *Algoritma C4.5* yang di analisis adalah data dari mahasiswa angkatan tahun 2008 Fakultas Ilmu Komputer Universitas singaperbangsa karawang, dari hasil analisis ternyata bahwa untuk kelulusan yang dilihat dari gender ternyata gender perempuan lebih besar persentasenya dibandingkan dengan gender laki-laki, kemudian nilai gain yang didapat dari usia pada mahasiswa yang bekerja lebih kepada mahasiswa dengan katagori usia 1, dan nilai gain yang didapat dari usia pada mahasiswa yang tidak bekerja lebih kepada mahasiswa dengan katagori usia 3.

**DAFTAR PUSTAKA**

Aurelie Harbelot. (2011). *Underspecified Quantification, Technical Report* Universiti Of Cambride United Kindom  
 C.R.Kothari. (2004). *Research Methology Methods and Techniques (Second Revised Edition)*. India: New Age International Limited

- Dian Oktafia dan D.I. Crispina Pardede. (2010) Perbandingan Kinerja *Algoritma Decision Tree* dan *Naive Bayes* dalam Prediksi Kebangrutan, Jakarta
- Firmansyah. (2011). Penerapan *Algoritma* Klasifikasi c4.5 Untuk Penentuan kelayakan Pemberian Kredit Koperasi, Jakarta
- Florin Gorunescu (2011) *Data Mining Concept, Model and Techniques, Intelligent System Reference Library*, Vol 12
- Gennady L. Andrienko dan Natalia V. Andrienko. (1999). *Data Mining with C4.5 and Interactive Cartographic Visualization*, Germany
- Han, j., dan Kamber, M. (2006). *Data Mining Concept and Techniques* . San fansisco: Morgan Kauffman
- Henny Leidiyana. (2011). Komparasi *Algoritma* Klasifikasi *Data Mining* Dalam Penentuan Resiko Kredit Kepemilikan Kendaraan Bemotor, Jakarta
- Ian H.Witten, Eibe Frank dan Mark A. Hall. (2011). *Data Mining Practical Machine Learning Tools and Techniques Third Edition*, Morgan Kaufmann
- Jiawei Han dan Jing Gao. (2009). *Research Challenges for Data Mining in Science and Engineering*, University Of Illinois at Urbana-Champaign
- Marco Aldenucci, Salvatore Ruggieri dan Massimo Torquati. (2011). *Porting Dcision Tree Building adn Pruning Algoritms to Multicore using FastFlow*, Dipartimento di Informatica
- Oded Maimon dan Lior Rokach. (2010). *Data Mining and Knowledge Discovery Handbook Second Edition*, Springer New York Dordrecht Heidelberg, London
- Olcay Taner Yıldız. (2011). *Univariate Decision Tree Introduction Using Maximum Margin Classification*, Intambul Turkey: Lain Stewart
- P.Bhargavi,M.Sc.,M.Tech dan Dr.S.Jyothi,M.Sc.,M.S,Ph.d. (2009). *Applying Naive Bayes Data Mining Technique for Classification of Agricultural Land Soils, International Journal of Computer Science and Network Security*, Vol 9 No. 8 Agustus 2009
- Veronika S Moertini. (2002). *Data Mining* Sebagai Solusi Bisnis, *Integral*, Vol. 7 No. 1 April 2002

#### BIODATA PENULIS

Jajam Haerul Jaman, SE., M.Kom. Lahir di Karawang, 10 November 1978, Pendidikan terakhir S2 Ilmu Komputer di STMIK Nusa Mandiri Jakarta.