

**IDENTIFIKASI LAMA STUDI BERDASARKAN  
KARAKTERISTIK MAHASISWA  
MENGUNAKAN ALGORITMA C4.5**

**(Studi Kasus Lulusan Fakultas Sains dan Matematika  
Universitas Diponegoro Tahun 2013/2014)**



**SKRIPSI**

**Oleh :**

**BRAMADITYA SWARASMARADHANA**

**24010210130068**

**JURUSAN STATISTIKA**

**FAKULTAS SAINS DAN MATEMATIKA**

**UNIVERSITAS DIPONEGORO**

**2014**

**IDENTIFIKASI LAMA STUDI BERDASARKAN  
KARAKTERISTIK MAHASISWA  
MENGUNAKAN ALGORITMA C4.5**

**(Studi Kasus Lulusan Fakultas Sains dan Matematika  
Universitas Diponegoro Tahun 2013/2014)**

**Oleh :**

**BRAMADITYA SWARASMARADHANA**

**24010210130068**

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar  
Sarjana Sains pada Jurusan Statistika**

**JURUSAN STATISTIKA  
FAKULTAS SAINS DAN MATEMATIKA  
UNIVERSITAS DIPONEGORO**

**2014**

## HALAMAN PENGESAHAN I

Judul : Identifikasi Lama Studi Berdasarkan Karakteristik Mahasiswa  
Menggunakan Algoritma C4.5  
(Studi Kasus Lulusan Fakultas Sains dan Matematika Universitas  
Diponegoro Tahun 2013/2014)

Nama : Bramaditya Swarasmaradhana

NIM : 24010210130068

Telah diujikan pada Sidang Tugas Akhir tanggal 29 September 2014 dan  
dinyatakan lulus pada tanggal 30 September 2014.

Semarang, Oktober 2014

Mengetahui,  
a.n. Ketua Jurusan Statistika  
Sekretaris Jurusan Statistika  
Fakultas Sains dan Matematika  
Undip



**Drs. Agus Rusgiyono, M.Si.**  
NIP. 196408131990011001

Panitia Penguji Ujian Tugas Akhir  
Ketua,



**Yuciana Wilandari, S.Si, M.Si.**  
NIP. 197005191998022001

## HALAMAN PENGESAHAN II

Judul : Identifikasi Lama Studi Berdasarkan Karakteristik Mahasiswa  
Menggunakan Algoritma C4.5  
(Studi Kasus Lulusan Fakultas Sains dan Matematika Universitas  
Diponegoro Tahun 2013/2014)  
Nama : Bramaditya Swarasmardhana  
NIM : 24010210130068

Telah diujikan pada Sidang Tugas Akhir tanggal 29 September 2014.

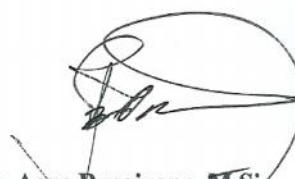
Semarang, Oktober 2014

Dosen Pembimbing I,

Dosen Pembimbing II,



**Moch. Abdul Mukid, S.Si, M.Si.**  
NIP. 197808172005011001



**Drs. Agus Rusgiono, M.Si.**  
NIP. 196408131990011001

## KATA PENGANTAR

Alhamdulillah, puji dan syukur penulis panjatkan ke hadirat Allah SWT karena berkat rahmat dan karunia-Nya penulis dapat menyelesaikan penulisan tugas akhir dengan judul “Identifikasi Lama Studi Berdasarkan Karakteristik Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus Lulusan Fakultas Sains dan Matematika Universitas Diponegoro Tahun 2013/2014)”. Tugas Akhir ini disusun sebagai salah satu mata kuliah yang wajib ditempuh untuk menyelesaikan studi jenjang S1 Jurusan Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro Semarang.

Penulis menyadari bahwa penyusunan Tugas Akhir ini tidak akan berjalan dengan baik tanpa adanya dukungan dan bantuan berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Ibu Dra. Hj. Dwi Ispriyanti, M.Si selaku Ketua Jurusan Statistika Fakultas Sains dan Matematika Universitas Diponegoro.
2. Bapak Moch. Abdul Mukid, S.Si, M.Si dan Bapak Drs. Agus Rusgiyono, M.Si selaku dosen pembimbing I dan II yang telah membimbing penulis hingga Tugas Akhir ini terselesaikan.
3. Bapak dan ibu dosen Jurusan Statistika Universitas Diponegoro.
4. Semua pihak yang telah memberikan bantuan yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa penyusunan Tugas Akhir ini jauh dari sempurna. Oleh karena itu kritik dan saran yang bersifat membangun sangat penulis harapkan. Semoga Tugas Akhir ini dapat berguna bagi semua pihak.

Semarang, September 2014

Penulis

## ABSTRAK

Berdasarkan peraturan akademik No. 209/PER/UN7/2012 masa studi mahasiswa Universitas Diponegoro ditetapkan dapat ditempuh selama 4 tahun. Dalam penelitian ini mahasiswa yang lulus dengan lama studi  $\leq 4$  tahun dikategorikan lulus tepat waktu dan mahasiswa dengan lama studi  $> 4$  tahun dikategorikan lulus tidak tepat waktu. Berlatarbelakang demikian, identifikasi mahasiswa untuk mengetahui mahasiswa yang lulus tepat waktu dan tidak tepat waktu dapat diketahui menggunakan Algoritma C4.5 berdasarkan jenis kelamin, jurusan, Indeks Prestasi Kumulatif (IPK), pengalaman berorganisasi, pengalaman kerja paruh waktu, beasiswa, asal mahasiswa dan jalur masuk. Algoritma C4.5 mengkonstruksikan pohon keputusan yang mampu mengatasi atribut bertipe kontinu, mengatasi nilai yang hilang dan dapat melakukan pemangkasan pohon yang kompleks. Hasil Algoritma C4.5 berupa pohon keputusan menunjukkan jurusan terpilih sebagai simpul akar, akurasi klasifikasi sebesar 84,475%, simpul yang terbentuk sebanyak 20 simpul dimana 13 diantaranya merupakan simpul daun. Profil mahasiswa yang teridentifikasi untuk kelulusan tepat waktu adalah mahasiswa Jurusan Fisika yang pernah mendapatkan beasiswa dan berjenis kelamin wanita; mahasiswa Jurusan Kimia dengan IPK  $> 3,06$ ; mahasiswa Jurusan Statistika dengan IPK  $> 3,43$  masuk universitas melalui jalur SNMPTN dan PSSB kemudian terakhir mahasiswa Jurusan Matematika dengan IPK lebih dari 2,96.

**Kata Kunci :** Lama Studi Mahasiswa, Algoritma C4.5, Pohon Keputusan.

## ABSTRACT

Based on academics regulation No. 209/PER/UN7/2012, the study period of student in Diponegoro University has been scheduled that is 4 years. In this thesis the graduation status of students that graduate under or equal to 4 years categorized as graduate on time, meanwhile students that graduate over 4 years categorized as graduate out of time. Thus background, students identification to classify students that graduate on time can be known using Algorithm C4.5 based on gender, majors, GPA, organizational experience, part time experience, scholarship, students origin and pathways scholar. Algorithm C4.5 constructs decision tree that can handle with missing values, continues attribute and pruning trees. Results of the Algorithm C4.5 that constructs a decision tree shows that majors is selected as a root node, the accuration of the Algorithm C4.5 is 84,475% and the number of the nodes are 20 nodes where 13 nodes are leaf nodes. The students profile that identified graduate on time are students of Physics who had received scholarship and a woman; students of Chemistry with GPA > 3,06; students of Statistics with GPA > 3,43 from SNMPTN also PSSB and students of Mathematics with GPA > 2,96.

**Keywords:** Study Period, Algorithm C4.5, Decision Tree.

# DAFTAR ISI

	<b>Halaman</b>
<b>HALAMAN JUDUL</b> .....	i
<b>HALAMAN PENGESAHAN</b> .....	ii
<b>KATA PENGANTAR</b> .....	iv
<b>ABSTRAK</b> .....	v
<b>ABSTRACT</b> .....	vi
<b>DAFTAR ISI</b> .....	vii
<b>DAFTAR TABEL</b> .....	ix
<b>DAFTAR GAMBAR</b> .....	x
<b>BAB I PENDAHULUAN</b>	
1.1. Latar Belakang .....	1
1.2. Permasalahan .....	3
1.3. Batasan Masalah .....	3
1.4. Tujuan Penulisan.....	4
<b>BAB II LANDASAN TEORI</b>	
2.1. Pendidikan Tinggi .....	5
2.2. Motivasi dalam Belajar .....	6
2.3. Algoritma C4.5 .....	6
2.3.1 Pembentukan Pohon Keputusan Algoritma C4.5 .....	7
2.3.2 Prosedur Pemilahan Algoritma C4.5 .....	10
2.3.3. Mengatasi Nilai yang Hilang ( <i>missing values</i> ) .....	11
2.3.4. Pemangkasan Pohon Keputusan .....	12



2.3.5. Pengukuran Ketepatan Hasil Klasifikasi.....	13
<b>BAB III METODOLOGI PENELITIAN</b>	
3.1. Sumber Data.....	14
3.2. Atribut Penelitian .....	14
3.3. Metode Penelitian .....	15
3.4. Diagram Alir Analisis .....	16
<b>BAB IV ANALISIS DAN PEMBAHASAN</b>	
4.1. Stastistika Deskriptif.....	17
4.2. Algoritma C4.5.....	28
4.2.1. Konstruksi Algoritma C4.5 .....	30
4.2.2. Pengukuran Ketepatan Hasil Klasifikasi Algoritma C4.5.....	36
4.2.3. Pemangkasan Pohon Keputusan.....	37
4.2.4. Pengukuran Ketepatan Hasil Klasifikasi Algoritma C4.5 Setelah Pemangkasan.....	44
4.2.5. Identifikasi Profil Mahasiswa.....	44
4.2.6. Pengujian Hasil Pohon Keputusan .....	45
<b>BAB V KESIMPULAN .....</b>	<b>47</b>
<b>DAFTAR PUSTAKA .....</b>	<b>48</b>
<b>LAMPIRAN.....</b>	<b>49</b>

## DAFTAR TABEL

	Halaman
Tabel 1. Matriks Konfusi.....	13
Tabel 2. Tabel Atribut Penelitian.....	14
Tabel 3. Tabel Status Lama Studi Mahasiswa Fakultas Sains dan Matematika Universitas Diponegoro Tahun 2013/2014 .....	18
Tabel 4. Tabel Status Lama Studi Berdasarkan Jenis Kelamin.....	19
Tabel 5. Tabel Status Lama Studi Berdasarkan Jurusan.....	20
Tabel 6. Tabel Indeks Prestasi Kumulatif Mahasiswa .....	21
Tabel 7. Tabel Status Lama Studi Berdasarkan Daerah Asal.....	22
Tabel 8. Tabel Status Lama Studi Berdasarkan Jalur Masuk .....	23
Tabel 9. Tabel Status Lama Studi Berdasarkan Beasiswa.....	24
Tabel 10. Tabel Status Lama Studi Berdasarkan Pekerjaan Paruh Waktu .....	25
Tabel 11. Tabel Status Lama Studi Berdasarkan Pengalaman Berorganisasi .	27
Tabel 12. Nilai <i>Information Gain</i> pada Simpul Akar .....	33
Tabel 13. Tabel Nilai Ambang Batas Atribut IPK Jurusan Statistika.....	35
Tabel 14. Nilai <i>Information Gain</i> Atribut Jurusan Statistika.....	35
Tabel 15. Matriks Konfusi Algoritma C4.5 .....	36
Tabel 16. Matriks Konfusi Algoritma C4.5 Setelah Pemangkasan .....	44
Tabel 17. Matriks Konfusi Sampel Pengujian .....	46

## DAFTAR GAMBAR

	Halaman
Gambar 1. Diagram Alir Penelitian .....	16
Gambar 2. Diagram Lama Studi Mahasiswa .....	18
Gambar 3. Diagram Lama Studi Berdasarkan Jenis Kelamin .....	19
Gambar 4. Diagram Lama Studi Berdasarkan Jurusan .....	20
Gambar 5. Diagram Lama Studi Berdasarkan Daerah Asal .....	22
Gambar 6. Diagram Lama Studi Berdasarkan Jalur Masuk .....	23
Gambar 7. Diagram Lama Studi Berdasarkan Beasiswa .....	24
Gambar 8. Diagram Lama Studi Berdasarkan Pengalaman Kerja Paruh Waktu .....	26
Gambar 9. Diagram Lama Studi Berdasarkan Pengalaman Berorganisasi	27
Gambar 10. Pohon Keputusan Tanpa Pemangkasan .....	29
Gambar 11. Pohon Keputusan Tingkat Pertama.....	33
Gambar 12. Pohon Keputusan pada Cabang Jurusan Statistika .....	36
Gambar 13. Kesalahan Terprediksi pada Simpul-simpul Jurusan Statistika .....	42
Gambar 14. Pohon Keputusan dengan Pemangkasan.....	43

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Pendidikan penting bagi setiap orang sebagai bekal untuk dapat melangsungkan kehidupannya. Pentingnya pendidikan bagi setiap orang di dalam sebuah negara akan memberikan pengaruh positif terhadap negara tersebut karena dengan pendidikan akan meningkatkan kualitas sumber daya manusia sehingga bagi negara tentu akan menambah daya saing terhadap negara lain. Menurut UU No. 12 tahun 2012, untuk meningkatkan daya saing bangsa dalam menghadapi globalisasi di segala bidang diperlukan pendidikan tinggi yang mampu mengembangkan ilmu pengetahuan dan teknologi serta menghasilkan intelektual, ilmuwan atau profesional yang berbudaya dan kreatif, toleran, demokratis, berkarakter tangguh, serta berani membela kebenaran untuk kepentingan bersama.

Pendidikan adalah usaha untuk mewujudkan suasana belajar dan proses pembelajaran. Perguruan tinggi adalah satuan pendidikan yang menyelenggarakan pendidikan tinggi. Menurut Djamarah (2002), untuk memperoleh hasil belajar yang baik harus melalui proses tertentu yang dipengaruhi oleh faktor dari dalam diri individu dan di luar individu. Banyak faktor dari luar anak didik yang mempengaruhi proses belajar seperti lingkungan sosial dan lingkungan alami. Sedangkan faktor dari dalam yang dapat mempengaruhi anak didik seperti halnya minat, kecerdasan, dan motivasi, dimana motivasi dalam perguruan tinggi salah satunya berupa ijazah kelulusan.

Dalam perguruan tinggi, dimana sistem perkuliahan berbeda dengan tingkat pendidikan dibawahnya serta tidak memiliki patokan standar dari aspek jadwal maupun waktu, menjadikan setiap mahasiswa bebas serta memiliki lebih banyak waktu untuk melakukan aktivitas di luar perkuliahan (ekstrakurikuler). Hal tersebut dapat dimanfaatkan bagi mahasiswa untuk mengembangkan diri diluar aktivitas perkuliahan agar siap terjun ke masyarakat maupun dunia kerja.

Berbeda dengan jenjang pendidikan dasar dan menengah, pada tingkat perguruan tinggi terutama program sarjana S1 memiliki syarat kelulusan bagi setiap mahasiswa adalah telah menempuh minimal 144 sampai 160 SKS. Sedangkan untuk lama studi mahasiswa, dikarenakan penelitian ini mengambil studi kasus di Universitas Diponegoro maka dengan berpedoman pada peraturan akademik Universitas Diponegoro No.209/PER/UN7/2012, lama studi mahasiswa ditetapkan dapat ditempuh dalam kurun waktu 4 tahun atau 8 semester dengan batas maksimal adalah 7 tahun atau 14 semester.

Berdasarkan uraian tersebut penelitian tugas akhir ini membahas mengenai identifikasi lama studi mahasiswa menggunakan Algoritma C4.5. Algoritma C4.5 merupakan salah satu metode klasifikasi untuk mengkonstruksikan pohon keputusan (*decision trees*) yang terdapat pada data mining. Algoritma C4.5 merupakan pembaharuan metode ID3 oleh Quinlan. Kelebihan Algoritma C4.5 dari metode pohon keputusan sejenis adalah bahwa algoritma C4.5 mampu mengatasi atribut yang bersifat kontinu, mengatasi nilai yang hilang dan melakukan pemangkasan pohon yang kompleks. Dalam penelitian tugas akhir ini peneliti ingin membentuk pohon klasifikasi untuk mengidentifikasi mahasiswa yang lulus dengan lama studi  $\leq 4$

tahun yang akan dikategorikan tepat waktu dan mahasiswa yang lulus dengan lama studi  $> 4$  tahun yang dikategorikan tidak tepat waktu berdasarkan faktor jenis kelamin, jurusan, IPK, beasiswa, pengalaman berorganisasi, kerja paruh waktu, daerah asal dan jalur masuk. Penelitian ini mengambil sampel lulusan Fakultas Sains dan Matematika Universitas Diponegoro periode April 2013 sampai dengan Januari 2014. Hasil dari penelitian ini berupa pohon keputusan mengenai lama studi mahasiswa baik yang lulus tepat waktu dan mahasiswa yang lulus tidak tepat waktu.

### **1.2. Permasalahan**

Berdasarkan latar belakang yang telah diuraikan maka perumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana membuat klasifikasi antara mahasiswa yang lulus tepat waktu atau menempuh lama studi  $\leq 4$  tahun dan mahasiswa yang lulus tidak tepat waktu atau menempuh lama studi  $> 4$  tahun menggunakan algoritma C4.5 ?
2. Bagaimana mengidentifikasi karakteristik mahasiswa yang lulus tepat waktu dan tidak tepat waktu ?

### **1.3. Batasan Masalah**

Berdasarkan perumusan masalah yang telah dipaparkan, maka penelitian dalam penulisan tugas akhir ini dibatasi hanya menggunakan 8 atribut dalam pembuatan klasifikasi lama studi dengan pohon keputusan menggunakan algoritma C4.5. Penelitian ini dibatasi sampai dengan konstruksi pohon algoritma C4.5 dengan pemangkasan dan tidak melibatkan estimasi untuk nilai yang hilang.

#### **1.4. Tujuan Penelitian**

Tujuan dari penelitian dalam penulisan tugas akhir ini adalah:

1. Mengkonstruksikan pohon keputusan menggunakan Algoritma C4.5 dan melakukan pemangkasan kemudian menganalisis profil yang terbentuk.
2. Mengukur ketepatan hasil klasifikasi pohon keputusan.
3. Identifikasi profil mahasiswa yang lulus tepat waktu.