

**PENERAPAN ALGORITMA C4.5 PADA APLIKASI  
PREDIKSI KELULUSAN MAHASISWA  
PRODI TEKNIK INFORMATIKA**



**SKRIPSI**

**Disusun Sebagai Salah Satu Syarat  
Untuk Memperoleh Gelar Sarjana Komputer  
Pada Departemen Ilmu Komputer/ Informatika**

**Disusun oleh:**

**RATNA PUSPITA SARI PUTRI**

**24010311130068**

**DEPARTEMEN ILMU KOMPUTER/ INFORMATIKA  
FAKULTAS SAINS DAN MATEMATIKA  
UNIVERSITAS DIPONEGORO**

**2018**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini :

Nama : Ratna Puspita Sari Putri

NIM : 24010311130068

Judul : Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Teknik Informatika

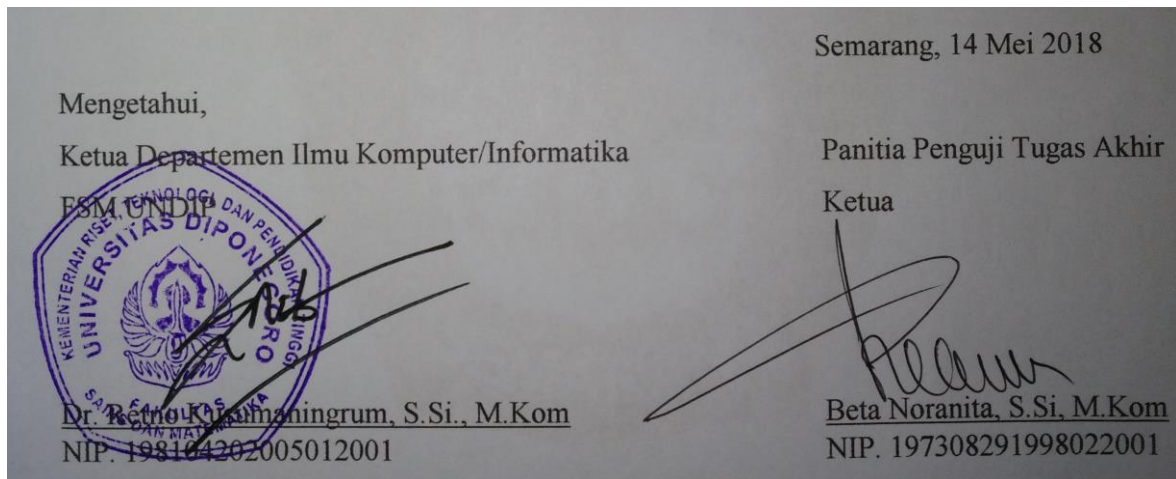
Dengan ini saya menyatakan bahwa dalam tugas akhir / skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.



## HALAMAN PENGESAHAN

Judul : Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa  
Prodi Teknik Informatika  
Nama : Ratna Puspita Sari Putri  
NIM : 24010311130068

Telah diujikan pada sidang tugas akhir pada tanggal 17 April 2018 dan dinyatakan lulus pada tanggal **17 April 2018**.



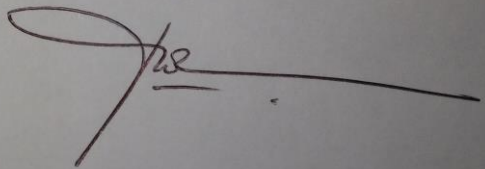
## HALAMAN PENGESAHAN

Judul : Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa  
Prodi Teknik Informatika  
Nama : Ratna Puspita Sari Putri  
NIM : 24010311130068

Telah diujikan pada sidang tugas akhir pada tanggal 17 April 2018.

Semarang, 14 Mei 2018

Pembimbing



Indra Waspada,ST,MTI  
NIP. 197902122008121002

## ABSTRAK

Data tentang mahasiswa yang lulus merupakan sebuah data yang penting baik bagi departemen, fakultas maupun universitas karena data tersebut digunakan dalam proses akreditasi. Data tentang mahasiswa yang lulus terus bertambah ditiap tahunnya dan menumpuk seperti data yang terabaikan karena jarang digunakan. Data tentang mahasiswa yang lulus dapat memberikan informasi yang berguna jika dimanfaatkan dengan maksimal. Maka dari itu penelitian ini akan memanfaatkan data tentang mahasiswa yang lulus dengan mengolahnya menggunakan *data mining* untuk mendapatkan informasi berupa prediksi kelulusan mahasiswa. Metode yang akan digunakan adalah metode pohon keputusan yang dibangun dengan algoritma C4.5 disertai dengan algoritma *error-based pruning* untuk proses pemotongan pohon keputusan. Kriteria yang akan digunakan adalah jenis kelamin, asal daerah, IPK, dan TOEFL. Dalam penerapannya, algoritma C4.5 dapat digunakan untuk menghasilkan prediksi kelulusan dengan nilai rata-rata *precision* 63.93%, *recall* 60.73%, dan akurasi 60.52%. Setelah pohon keputusan dipotong dengan menggunakan metode *error-based pruning*, didapatkan hasil yang lebih baik. Pohon yang dipotong dengan menggunakan nilai *confidence* 0,4 menghasilkan *precision* 70.70%, *recall* 50.65%, dan akurasi 61.57%. Sedangkan pohon yang dipotong dengan menggunakan nilai *confidence* 0,25 menghasilkan *precision* 73.77%, *recall* 48.84%, dan akurasi 62.44%.

Kata kunci: *Data Mining*, Kelulusan Mahasiswa, Pohon Keputusan, C4.5, *Error-based Pruning*

## ABSTRACT

Data of student who have graduated is essential data for department, faculty and university because it used in the process of validating the existence of a university. Data of student who have graduated continues to grow each year and accumulate as neglected data because it is rarely used. Data of student who have graduated can provide useful information if it is utilized to the fullest. Therefore, this research used data of student who have graduated by processing them using data mining and getting information in the form of students' graduation prediction. The method used in this research is decision tree method built with C4.5 algorithm with error-based pruning algorithm for decision tree pruning. The criteria that will be used are gender; regional origin; GPA; and TOEFL. In its implementation, the C4.5 algorithm can be used to predict students' graduation with an average 63.93% of precision, an average 60.73% of recall, and an average 60.52% of accuracy. After the decision tree pruned using the error-based pruning method, better results are obtained. Pruned tree with 0.4 confidence value gives result an average 70.70% of precision, an average 50.65% of recall, and an average 61.57% of accuracy. While the pruned tree with 0.25 confidence value gives result an average 73.77% of precision, an average 48.84% of recall, and an average 62.44% of accuracy.

Keywords: Data Mining, Students' Graduation, Decision Tree, C4.5, Error-based Pruning

## KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas rahmat yang diberikan sehingga penulis dapat menyelesaikan laporan tugas akhir yang berjudul “Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Teknik Informatika”. Laporan ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana komputer pada Departemen Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro.

Dalam penyusunan laporan ini penulis mendapat banyak bimbingan dan bantuan dari berbagai pihak. Untuk itu, pada kesempatan ini penulis menyampaikan rasa hormat dan ucapan terima kasih kepada:

1. Dr. Retno Kusumaningrum, S.Si, M.Kom selaku Ketua Departemen Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro.
2. Helmie Arif Wibawa, S.Si, M.Cs, selaku Koordinator Tugas Akhir Departemen Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro.
3. Indra Waspada, ST, MTI selaku dosen pembimbing yang telah membantu dalam membimbing dan mengarahkan penulis hingga selesainya laporan tugas akhir ini.
4. Pihak Prodi Teknik Informatika yang telah membantu dan memberikan informasi guna penyelesaian laporan tugas akhir ini.
5. Semua pihak yang telah membantu kelancaran dalam penyelesaian laporan tugas akhir ini, yang tidak dapat disebutkan satu per satu.

Penulis menyadari bahwa dalam laporan ini masih banyak kekurangan baik dari segi materi ataupun dalam penyajiannya karena keterbatasan kemampuan dan pengetahuan penulis. Oleh karena itu, kritik dan saran sangat penulis harapkan. Semoga laporan ini dapat bermanfaat bagi pembaca pada umumnya dan penulis pada khususnya.

Semarang, Mei 2018

Penulis

## DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PENGESAHAN .....	iv
ABSTRAK .....	v
ABSTRACT .....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI .....	viii
DAFTAR TABEL .....	xii
DAFTAR GAMBAR.....	xv
DAFTAR ISTILAH DAN SINGKATAN.....	xvii
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang Masalah .....	1
1.2. Rumusan Masalah.....	2
1.3. Tujuan dan Manfaat .....	3
1.4. Ruang Lingkup .....	3
1.5. Sistematika Penulisan .....	3
BAB II LANDASAN TEORI.....	5
2.1. Tinjauan Pustaka.....	5
2.2. Kelulusan Mahasiswa .....	6
2.3. <i>Data Mining</i> .....	7
2.3.1 Tahapan <i>Knowledge Discovery in Databases</i> .....	7
2.3.2. Pengelompokan <i>Data Mining</i> .....	8
2.3.2.1. Deskripsi.....	8
2.3.2.2. Estimasi .....	9
2.3.2.3. Prediksi.....	9
2.3.2.4. Klasifikasi.....	9
2.3.2.5. Pengklusteran .....	9
2.3.2.6. Asosiasi .....	10
2.4. <i>Imbalance Data</i> .....	10
2.5. Pohon Keputusan .....	11



2.5.1.	Kebutuhan Pohon Keputusan .....	11
2.5.2.	Algoritma Pohon Keputusan .....	12
2.5.3.	Algoritma untuk Membangun Pohon Keputusan .....	14
2.5.3.1	<i>Chi-squared Automatic Interaction Detector (CHAID)</i> .....	14
2.5.3.2	<i>Classification and Regression Tree (CART)</i> .....	14
2.5.3.3	<i>Iterative Dichotomiser 3 (ID3)</i> .....	14
2.5.3.4	C4.5 .....	15
2.6.	Algoritma C4.5 .....	15
2.7.	<i>Pruning</i> .....	17
2.7.1.	<i>Pre-pruning</i> .....	18
2.7.2.	<i>Post-pruning</i> .....	18
2.7.2.1.	<i>Pessimistic Error Pruning</i> .....	19
2.7.2.2.	<i>Error-Based Pruning</i> .....	19
2.7.2.3.	<i>Reduced Error Pruning</i> .....	19
2.7.2.4.	<i>Cost-Complexity Minimization</i> .....	20
2.8.	<i>Error-Based Pruning</i> .....	20
2.9.	Metode <i>Waterfall</i> .....	21
2.10.	Pemodelan Analisis .....	22
2.10.1.	Pemodelan Data .....	23
2.10.1.1.	Obyek Data, Atribut, dan Relasi .....	23
2.10.1.2.	Kardinalitas dan Modalitas .....	24
2.10.1.3.	<i>Entity Relationship Diagram</i> .....	25
2.10.2.	Pemodelan Fungsional .....	25
2.10.2.1.	<i>Data Context Diagram</i> .....	26
2.10.2.2.	<i>Data Flow Diagram</i> .....	26
2.11.	Pengujian Perangkat Lunak.....	26
2.11.1.	<i>White Box</i> .....	27
2.11.2.	<i>Black Box</i> .....	27
2.12.	Pengukuran Kinerja.....	28
BAB III METODOLOGI PENELITIAN .....		30
3.1.	Garis Besar Penyelesaian Masalah .....	30
3.2.	Analisis Kebutuhan Aplikasi .....	31
3.2.1.	Kebutuhan Fungsional dan Non-Fungsional .....	31

3.2.2.	Kebutuhan Non-Fungsional.....	31
3.3.	Pemodelan Analisis.....	31
3.3.1.	Pemodelan Data.....	31
3.3.2.	Pemodelan Fungsional.....	34
3.3.2.1.	<i>Data Context Diagram</i> (DCD).....	34
3.3.2.2.	<i>Data Flow Diagram</i> (DFD) .....	35
3.4.	Desain Aplikasi.....	38
3.4.1.	Desain Data.....	38
3.4.2.	Desain Antarmuka .....	40
3.5.	Tahapan KDD untuk Pembentukan Pohon Keputusan.....	44
3.5.1.	Pemahaman Domain dan Tujuan KDD .....	44
3.5.2.	Pemilihan dan Penambahan Data .....	45
3.5.3.	Pembersihan dan Pemrosesan Awal Data .....	46
3.5.4.	Transformasi Data .....	46
3.5.5.	<i>Data Mining</i> .....	48
3.5.5.1.	Memilih Tugas <i>Data Mining</i> .....	48
3.5.5.2.	Memilih Algoritma <i>Data Mining</i> .....	48
3.5.5.3.	Menerapkan Algoritma <i>Data Mining</i> .....	49
BAB IV	HASIL DAN PEMBAHASAN.....	60
4.1.	Tahapan KDD untuk Pembentukan Pohon Keputusan.....	60
4.1.1.	Interpretasi dan Evaluasi .....	60
4.1.1.1.	Interpretasi Pohon Keputusan .....	60
4.1.1.2.	Pengukuran Kinerja Pohon Keputusan .....	61
4.1.2.	Mengolah Pengetahuan.....	64
4.2	Implementasi Aplikasi .....	64
4.2.1	Lingkungan Implementasi .....	64
4.2.2	Implementasi Data.....	64
4.2.3	Implementasi Antarmuka .....	68
4.3	Pengujian Aplikasi .....	72
BAB V	KESIMPULAN DAN SARAN .....	74
5.1.	Kesimpulan .....	74
5.2.	Saran .....	74
DAFTAR	PUSTAKA.....	75

LAMPIRAN .....	77
Lampiran 1. Hasil Perhitungan Pembuatan Pohon Keputusan.....	78
Lampiran 2. Hasil Perhitungan Pemotongan Pohon Keputusan .....	87
Lampiran 3. Hasil Pohon Keputusan Tanpa Pemotongan .....	90
Lampiran 4. <i>Rules</i> dari Pohon Keputusan Tanpa Pemotongan .....	91
Lampiran 5. <i>Confusion Matrix</i> .....	93
Lampiran 6. Hasil Pengujian Aplikasi Prediksi Ketepatan Kelulusan Mahasiswa .....	95
Lampiran 7. Surat Keterangan Penelitian.....	98

## DAFTAR TABEL

Tabel 2.1 Daftar Referensi .....	5
Tabel 2.2 Tabel Notasi ERD .....	25
Tabel 2.3 Tabel Notasi DFD .....	26
Tabel 3.1 Kebutuhan Fungsional.....	31
Tabel 3.2 Kebutuhan Non-Fungsional .....	31
Tabel 3.3 Tabel <i>user</i> .....	38
Tabel 3.4 Tabel <i>lulusan_transform</i> .....	39
Tabel 3.5 Tabel <i>data_training</i> .....	39
Tabel 3.6 Tabel <i>data_testing</i> .....	39
Tabel 3.7 Tabel <i>data_prediksi</i> .....	40
Tabel 3.8 Tabel <i>pohon_keputusan</i> .....	40
Tabel 3.9 Data Lulusan Prodi Teknik Informatika.....	45
Tabel 3.10 Tabel Transformasi Data .....	47
Tabel 3.11 Hasil transformasi data lulusan Prodi Teknik Informatika.....	47
Tabel 3.12 Hasil perhitungan jumlah kasus <i>node</i> akar .....	52
Tabel 3.13 Hasil perhitungan <i>entropy node</i> akar.....	53
Tabel 3.14 Hasil perhitungan <i>information gain node</i> akar .....	53
Tabel 3.15 Hasil perhitungan <i>split information node</i> akar .....	54
Tabel 3.16 Hasil perhitungan <i>gain ratio node</i> akar .....	54
Tabel 3.17 Hasil perhitungan sub-pohon IPK-Sangat memuaskan.....	55
Tabel 3.18 Hasil perhitungan sub-pohon IPK-Dengan pujian .....	55
Tabel 3.19 Hasil perhitungan jumlah kasus sub-pohon IPK-Sangat memuaskan dan TOEFL-Dasar dan Asal daerah-Jateng.....	58
Tabel 3.20 Hasil perhitungan nilai <i>f</i> .....	58
Tabel 3.21 Hasil perhitungan <i>error estimate</i> sub-pohon IPK-Sangat memuaskan dan TOEFL-Dasar dan Asal daerah-Jateng dan Jenis Kelamin .....	59
Tabel 4.1 Detail pengukuran Kinerja .....	62
Tabel 4.2 Hasil perhitungan <i>precision, recall</i> , dan akurasi seluruh pengukuran .....	63
Tabel 4.3 Hasil perhitungan rata-rata <i>precision, recall</i> , dan akurasi.....	63
Tabel 4.4 Tabel spesifikasi perangkat keras dan lunak .....	64
Tabel 4.5 Rencana pengujian.....	72

Tabel L1.1 Hasil perhitungan sub-pohon IPK-Sangat Memuaskan dan TOEFL-Dasar .....	78
Tabel L1.2 Hasil perhitungan sub-pohon IPK-Sangat Memuaskan dan TOEFL-Dasar dan Asal Daerah-Jateng.....	78
Tabel L1.3 Hasil perhitungan sub-pohon IPK-Sangat Memuaskan dan TOEFL-Dasar dan Asal Daerah-Luar Jateng .....	78
Tabel L1.4 Hasil perhitungan sub-pohon IPK-Sangat Memuaskan dan TOEFL-Dasar dan Asal Daerah-Luar Jawa .....	79
Tabel L1.5 Hasil perhitungan sub-pohon IPK-Sangat Memuaskan dan TOEFL-Menengah bawah.....	79
Tabel L1.6 Hasil perhitungan sub-pohon IPK-Sangat Memuaskan dan TOEFL-Menengah bawah dan Asal Daerah-Jateng.....	80
Tabel L1.7 Hasil perhitungan sub-pohon IPK-Sangat Memuaskan dan TOEFL-Menengah bawah dan Asal Daerah-Luar Jateng .....	80
Tabel L1.8 Hasil perhitungan sub-pohon IPK-Sangat Memuaskan dan TOEFL-Menengah bawah dan Asal Daerah-Luar Jawa .....	81
Tabel L1.9 Hasil perhitungan sub-pohon IPK-Sangat Memuaskan dan TOEFL-Menengah Atas.....	81
Tabel L1.10 Hasil perhitungan sub-pohon IPK-Sangat Memuaskan dan TOEFL-Menengah atas dan Asal Daerah-Jateng.....	81
Tabel L1.11 Hasil perhitungan sub-pohon IPK-Sangat Memuaskan dan TOEFL-Menengah atas dan Asal Daerah-Luar Jateng .....	82
Tabel L1.12 Hasil perhitungan sub-pohon IPK-Sangat Memuaskan dan TOEFL-Mahir...	82
Tabel L1.13 Hasil perhitungan sub-pohon IPK-Dengan Pujian dan TOEFL-Dasar.....	83
Tabel L1.14 Hasil perhitungan sub-pohon IPK- Dengan Pujian dan TOEFL-Dasar dan Asal Daerah-Jateng .....	83
Tabel L1.15 Hasil perhitungan sub-pohon IPK- Dengan Pujian dan TOEFL-Dasar dan Asal Daerah-Luar Jateng.....	84
Tabel L1.16 Hasil perhitungan sub-pohon IPK- Dengan Pujian dan TOEFL-Menengah bawah.....	84
Tabel L1.17 Hasil perhitungan sub-pohon IPK- Dengan Pujian dan TOEFL-Menengah bawah dan Asal Daerah-Jateng.....	84
Tabel L1.18 Hasil perhitungan sub-pohon IPK- Dengan Pujian dan TOEFL-Menengah bawah dan Asal Daerah-Luar Jateng .....	85

Tabel L1.19 Hasil perhitungan sub-pohon IPK-Sangat Memuaskan dan TOEFL-Menengah Atas.....	85
Tabel L1.20 Hasil perhitungan sub-pohon IPK-Sangat Memuaskan dan TOEFL-Mahir...	86
Tabel L2.1 Hasil Perhitungan Pemotongan Pohon Keputusan .....	87
Tabel L5.1 <i>Confusion matrix</i> pengukuran ke-1 .....	93
Tabel L5.2 <i>Confusion matrix</i> pengukuran ke-2.....	93
Tabel L5.3 <i>Confusion matrix</i> pengukuran ke-3.....	93
Tabel L5.4 <i>Confusion matrix</i> pengukuran ke-4.....	93
Tabel L5.5 <i>Confusion matrix</i> pengukuran ke-5.....	93
Tabel L5.6 <i>Confusion matrix</i> pengukuran ke-6.....	93
Tabel L5.7 <i>Confusion matrix</i> pengukuran ke-7.....	93
Tabel L5.8 <i>Confusion matrix</i> pengukuran ke-8.....	94
Tabel L5.9 <i>Confusion matrix</i> pengukuran ke-9.....	94
Tabel L5.10 <i>Confusion matrix</i> pengukuran ke-10.....	94
Tabel L5.11 <i>Confusion matrix</i> pengukuran ke-11 .....	94
Tabel L5.12 <i>Confusion matrix</i> pengukuran ke-12.....	94
Tabel L5.13 <i>Confusion matrix</i> pengukuran ke-13.....	94
Tabel L5.14 <i>Confusion matrix</i> pengukuran ke-14.....	94
Tabel L5.15 <i>Confusion matrix</i> pengukuran ke-15.....	94
Tabel L6.1 Hasil Pengujian Aplikasi.....	95

## DAFTAR GAMBAR

Gambar 2.1 Tahapan <i>Knowledge Discovery in Databases</i> .....	7
Gambar 2.2 <i>Flowchart random over sampling (ROS)</i> .....	11
Gambar 2.3 Contoh pohon keputusan (Han, et al., 2012) .....	11
Gambar 2.4 Algoritma Pohon Keputusan .....	13
Gambar 2.5 <i>Pseudocode C4.5</i> (Ananda, 2014) .....	17
Gambar 2.6 Model <i>Waterfall</i> .....	22
Gambar 2.7 Struktur model analisis .....	22
Gambar 2.8 Relasi Satu ke Satu .....	24
Gambar 2.9 Relasi Satu ke Banyak .....	24
Gambar 2.10 Relasi Banyak ke Banyak .....	24
Gambar 2.11 <i>Confusion Matrix</i> .....	28
Gambar 3.1 <i>Flowchart</i> Aplikasi Prediksi Kelulusan.....	30
Gambar 3.2 ERD Aplikasi Prediksi Ketepatan Kelulusan Mahasiswa .....	33
Gambar 3.3 DCD Aplikasi Prediksi Ketepatan Kelulusan Mahasiswa.....	34
Gambar 3.4 DFD Level 1 .....	37
Gambar 3.5 DFD Level 2 Proses 2.....	37
Gambar 3.6 DFD Level 2 Proses 3.....	38
Gambar 3.7 Desain antarmuka data lulusan .....	42
Gambar 3.8 Desain antarmuka partisi data.....	42
Gambar 3.9 Desain antarmuka membuat pohon keputusan .....	42
Gambar 3.10 Desain antarmuka menampilkan pohon keputusan .....	43
Gambar 3.11 Desain antarmuka pengukuran kinerja .....	43
Gambar 3.12 Desain antarmuka prediksi .....	43
Gambar 3.13 Desain antarmuka hasil prediksi .....	44
Gambar 3.14 <i>Flowchart</i> tahap <i>data mining</i> .....	49
Gambar 3.15 <i>Flowchart</i> proses penanganan <i>imbalance data</i> .....	50
Gambar 3.16 <i>Flowchart</i> proses pembuatan pohon keputusan.....	51
Gambar 3.17 <i>Flowchart</i> metode EBP .....	57
Gambar 4.1 Pohon keputusan yang dipotong dengan <i>confidence</i> 0,4.....	61
Gambar 4.2 Implementasi tabel <i>user</i> .....	65
Gambar 4.3 Implementasi tabel <i>lulusan_transform</i> .....	66

Gambar 4.4 Implementasi tabel data_training.....	66
Gambar 4.5 Implementasi tabel data_testing .....	67
Gambar 4.6 Implementasi tabel data_prediksi .....	68
Gambar 4.7 Implementasi tabel pohon_keputusan .....	68
Gambar 4.8 Implementasi antarmuka data lulusan .....	70
Gambar 4.9 Implementasi antarmuka partisi data .....	70
Gambar 4.10 Implementasi antarmuka membuat pohon keputusan.....	71
Gambar 4.11 Implementasi antarmuka menampilkan pohon keputusan.....	71
Gambar 4.12 Implementasi antarmuka pengukuran kinerja.....	71
Gambar 4.13 Implementasi antarmuka prediksi.....	72
Gambar 4.14 Implementasi antarmuka hasil prediksi .....	72
Gambar L3.1 Pohon keputusan tanpa pemotongan .....	90



## DAFTAR ISTILAH DAN SINGKATAN

<b>DCD</b>	<i>Data Context Diagram</i>
<b>DFD</b>	<i>Data Flow Diagram</i>
<b>EBP</b>	<i>Error-based Pruning</i>
<b>ERD</b>	<i>Entity Relationship Diagram</i>
<b>FN</b>	<i>False Negatives</i>
<b>FP</b>	<i>False Positives</i>
<b>IPK</b>	Indeks Prestasi Kumulatif
<b>KDD</b>	<i>Knowledge Discovery in Databases</i>
<b>ROS</b>	<i>Random Over Sampling</i>
<b>TN</b>	<i>True Negatives</i>
<b>TOEFL</b>	<i>Test of English As A Foreign Language</i>
<b>TP</b>	<i>True Positives</i>

# BAB I

## PENDAHULUAN

Bab ini memaparkan latar belakang, rumusan masalah, tujuan dan manfaat, serta ruang lingkup tugas akhir mengenai aplikasi prediksi kelulusan mahasiswa Prodi Teknik Informatika.

### 1.1. Latar Belakang Masalah

Di era digital ini banyak instansi dan perusahaan yang telah menyimpan data mereka di dalam sebuah *database* yang terkomputerisasi. Dunia pendidikan pun tidak terlepas dari perkembangan teknologi ini. Universitas Diponegoro termasuk salah satu perguruan tinggi yang telah menyimpan datanya dalam *database* yang terkomputerisasi. Data tersebut merupakan data mahasiswa, data dosen, serta berbagai data lain yang berhubungan dengan Universitas Diponegoro.

Data tersebut tidak banyak memiliki kegunaan dan seolah-olah menjadi sekumpulan data terabaikan yang bertambah besar tiap tahunnya. Data tersebut hanya digunakan saat universitas membutuhkan suatu informasi tertentu atau saat proses akreditasi. Saat mahasiswa telah lulus maka data mereka akan semakin jarang digunakan. Padahal data tentang mahasiswa yang lulus merupakan data yang penting dan digunakan dalam proses akreditasi.

Data tentang mahasiswa yang lulus dapat memberikan informasi yang berguna bagi universitas jika dimanfaatkan dengan maksimal. Salah satu cara untuk memanfaatkan data tentang mahasiswa yang lulus ini adalah dengan mengolahnya menggunakan *data mining*. Dengan proses *data mining* ini dapat ditemukan pola atau aturan yang dapat digunakan untuk menghasilkan suatu informasi seperti prediksi kelulusan mahasiswa.

Prediksi kelulusan mahasiswa dapat digunakan lebih lanjut untuk membantu universitas dalam mengevaluasi dan memperbaiki sistem pembelajaran sehingga universitas dapat menghasilkan lulusan yang berkualitas.

Penelitian ini akan dilaksanakan pada Prodi Teknik Informatika. Prodi Teknik Informatika telah berdiri dari tahun 2004 dan memiliki sasaran untuk menjadi program studi unggulan. Oleh karena itu hasil prediksi kelulusan mahasiswa dapat membantu Prodi Teknik Informatika dalam mengambil langkah strategis.

Dalam penelitian yang berjudul *Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques*, Jadhav dan Channe membandingkan performa metode K-NN, Naïve Bayes, dan pohon keputusan dalam berbagai aspek dengan menggunakan berbagai *dataset*. Dari hasil penelitian tersebut dapat diketahui bahwa pohon keputusan merupakan metode yang paling cepat performanya dibandingkan dengan metode yang lain. Selain itu pohon keputusan lebih akurat dan memiliki *error rate* yang rendah (Jadhav & Channe, 2016).

Dalam penelitian lain yang berjudul *Comparative Analysis of Decision Tree Algorithms for The Prediction of Eligibility of A Man for Availing Bank Loan*, Mohankumar dkk membandingkan berbagai algoritma untuk membangun pohon keputusan dan algoritma C4.5 merupakan algoritma dengan performa tercepat dan memiliki akurasi yang paling tinggi (Mohankumar, et al., 2016).

David Kamagi mengimplementasikan algoritma C4.5 dalam penelitian berjudul “Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa” dan menghasilkan prediksi dengan tingkat keakuratan yang tinggi, yaitu 87,5% (Kamagi, 2014). Kamagi menggunakan empat kelas target, yaitu lulus cepat, lulus tepat, lulus terlambat, dan *drop out*. Atribut yang digunakan adalah IPS, jenis kelamin, asal sekolah, tipe kelulusan, dan jumlah SKS. Dalam penelitian ini akan digunakan kelas target  $< 5$  tahun dan  $\geq 5$  tahun. Sedangkan atribut yang digunakan adalah IPK, TOEFL, asal daerah, dan jenis kelamin.

Berdasarkan latar belakang di atas akan dibangun sebuah aplikasi prediksi kelulusan mahasiswa Prodi Teknik Informatika dengan menggunakan metode pohon keputusan yang dibangun menggunakan algoritma C4.5.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan dapat dirumuskan suatu permasalahan, yaitu:

1. Bagaimana membangun sebuah aplikasi prediksi kelulusan mahasiswa Prodi Teknik Informatika menggunakan algoritma C4.5?
2. Bagaimana pengaruh pemotongan pohon pada algoritma C4.5 terhadap akurasi pohon keputusan yang dihasilkan?

### 1.3. Tujuan dan Manfaat

Tujuan yang ingin dicapai dalam penelitian ini adalah membangun sebuah aplikasi yang dapat memprediksi kelulusan mahasiswa Prodi Teknik Informatika menggunakan algoritma C4.5 serta mengetahui pengaruh pemotongan pohon pada algoritma C4.5 terhadap akurasi pohon keputusan yang dihasilkan.

Adapun manfaat yang diharapkan dari penelitian tugas akhir ini adalah membantu Prodi Teknik Informatika untuk mengambil langkah strategis dalam mengevaluasi dan memperbaiki sistem pembelajaran sehingga dapat menghasilkan lulusan yang lebih berkualitas.

### 1.4. Ruang Lingkup

Berikut ruang lingkup pengembangan aplikasi prediksi ketepatan kelulusan mahasiswa Prodi Teknik Informatika:

1. Aplikasi yang dibangun merupakan aplikasi berbasis *web*.
2. Aplikasi dibangun menggunakan bahasa pemrograman PHP dan HTML dengan sistem manajemen basis data MySQL.
3. Metode yang digunakan dalam pengembangan aplikasi adalah metode *waterfall*. Metode *waterfall* akan dilakukan sampai tahap pengujian.
4. Model yang digunakan untuk *data mining* adalah model *knowledge discovery in Databases* (KDD).
5. Metode yang digunakan adalah metode pohon keputusan yang dibangun dengan algoritma C4.5.
6. Metode yang digunakan untuk melakukan pemotongan pohon keputusan adalah metode *error-based pruning*.
7. Data yang digunakan adalah data tentang mahasiswa yang lulus pada Prodi Teknik Informatika tahun 2013-2017.

### 1.5. Sistematika Penulisan

Sistematika penulisan yang digunakan dalam pembuatan tugas akhir ini terbagi dalam beberapa pokok bahasan yaitu:

#### BAB I PENDAHULUAN

Berisi pendahuluan yang terdiri atas latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan.

## BAB II LANDASAN TEORI

Berisi dasar teori yang berhubungan dengan topik tugas akhir. Dasar teori yang digunakan dalam penyusunan tugas akhir ini meliputi tinjauan pustaka, ketepatan kelulusan mahasiswa, *data mining*, *imbalance data*, pohon keputusan, algoritma C4.5, *pruning*, *error-based pruning*, metode *waterfall*, pemodelan analisis, pengujian perangkat lunak, dan pengukuran kinerja.

## BAB III METODOLOGI PENELITIAN

Berisi tentang tahapan-tahapan dalam penyusunan tugas akhir serta perancangan aplikasi, yaitu tahapan KDD untuk pembentukan pohon keputusan, analisis kebutuhan, pemodelan analisis, dan desain aplikasi.

## BAB IV HASIL DAN PEMBAHASAN

Berisi hasil dan pembahasan dari penelitian yang telah dilakukan.

## BAB V PENUTUP

Berisi kesimpulan yang diambil berkaitan dengan aplikasi yang dibangun dan saran untuk pengembangan lebih lanjut.