

KLASIFIKASI TINGKAT KELULUSAN MAHASISWA PRODI TEKNIK ELEKTRO FST UNDANA MENGGUNAKAN ALGORITMA C4.5

Wenefrida T. Ina¹, Sarlince Manu², Molina Odja³, Evtaleny Mauboy⁴, Kalvein Rantelobo⁵, Amin Maggang⁶,
Stephanie Pella⁷

¹Program Studi Teknik Elektro, Universitas Nusa Cendana Kupang, Jl. Adi SuciptoPenfui Kupang

Email: wenefrida_ina@staf.undana.ac.id

^{2,3,4,5,6,7}Program Studi Teknik Elektro, Universitas Nusa Cendana Kupang, Jl. Adi SuciptoPenfui Kupang

ABSTRAK

Klasifikasi tingkat kelulusan mahasiswa penting dilakukan untuk mengetahui pola kelulusan mahasiswa sehingga bisa menjadi bahan rujukan bagi prodi dalam merencanakan bimbingan akademik yang relevan dalam meningkatkan mutu lulusan serta mempersingkat masa studi mahasiswa oleh karena itu perlu dilakukan penelitian yang bertujuan menghasilkan pola prediksi tingkat kelulusan mahasiswa Teknik Elektro FST Undana, dan salah satu algoritma yang digunakan adalah Algoritma C4.5, dimana Algoritma C4.5 merupakan salah satu algoritma terbaik (Ina, 2013) dalam pembentukan pohon keputusan dalam menghasilkan suatu pola untuk memprediksi suatu keputusan berdasarkan data-data masa lalu. Data-data yang diperlukan berupa Indeks Perstasi (IP) mahasiswa pada 5 (Lima) semester pertama perkuliahan dari mahasiswa angkatan tahun 2008-2010 sebagai data pelatihan untuk pembentukan pohon keputusan, kemudian data Indeks Perstasi (IP) mahasiswa pada 5 (Lima) semester pertama perkuliahan dari mahasiswa angkatan tahun 2011-2012 digunakan sebagai data uji. Data nilai tersebut akan diklasifikasikan ke dalam beberapa tingkat kelulusan yaitu cepat, sedang, dan lambat menggunakan Algoritma C4.5 pada *tools* WEKA sebagai software data mining. Data yang dapat diklasifikasi dengan baik sebanyak 91,1004%, dengan *Mean Absolute Error* 0,0982 dan *Root Relative Squared Error* 77,4171%. Klasifikasi yang diperoleh sebagai berikut: Jika **IP semester 3 Sedang** maka **Kelulusan Lama**, Jika **IP semester 3 Tinggi** maka **Kelulusan Sedang**, Jika **IP semester 3 Rendah** maka **Kelulusan Lama**. IP semester 3 (IP-3) menjadi penentu seorang mahasiswa Prodi Teknik Elektro – FST dapat menyelesaikan studi dengan Cepat/Sedang/Lama. Hal ini sejalan atau selaras dengan keberadaan matakuliah semester 3 pada Prodi Teknik Elektro yang sesungguhnya berisi materi-materi kuliah dasar keteknikan yang menjadi dasar ilmu bagi matakuliah – matakuliah selanjutnya di semester atas.

Kata Kunci : Klasifikasi, Kelulusan, Algoritma C4.5

Author : Wenefrida T. Ina, Sarlince Manu, Molina Odja, Evtaleny Mauboy, Kalvein Rantelobo, Amin Maggang, Stephanie Pella

1. PENDAHULUAN

Keberadaan mahasiswa pada suatu program studi (prodi) merupakan bonus point dalam hal tumbuh kembang prodi, sehingga prodi selalu melakukan berbagai upaya untuk meningkatkan jumlah mahasiswa, menekan angka *dropout* (DO), dan meningkatkan jumlah lulusan yang tepat waktu karena tingkat kelulusan mahasiswa merupakan salah satu unsur penunjang penilaian akreditasi suatu program studi, dimana semakin banyak mahasiswa yang menyelesaikan studi tepat waktu maka semakin tinggi nilai yang diperoleh dalam standar kelulusan mahasiswa. Agar mahasiswa dapat lulus tepat waktu maka prodi melakukan berbagai upaya untuk menunjang proses belajar mahasiswa serta menyediakan sarana konsultasi akademik melalui para dosen.

Sampai saat ini belum ada pengelompokan mahasiswa berdasarkan kemampuan akademik yang merujuk pada suatu pola tingkat kelulusan mahasiswa di Prodi Teknik Elektro FST Undana, sehingga para dosen penasehat akademik tidak mempunyai dasar yang kuat dalam merencanakan berbagai upaya pembimbingan akademik terhadap mahasiswa. Hal ini mengakibatkan kelulusan mahasiswa Prodi Teknik Elektro FST Undana masih semata-mata berdasarkan kemampuan akademik murni bawaan pribadi setiap mahasiswa. Dampak dari kondisi ini adalah prodi belum mampu mengendalikan masa studi mahasiswa sehingga jumlah kelulusan mahasiswa yang tepat waktu masih belum dapat diprediksi. Berdasarkan permasalahan yang ada maka diperlukan suatu penelitian yang bertujuan menghasilkan klasifikasi tingkat kelulusan mahasiswa agar membantu pihak prodi dalam merencanakan peningkatan mutu lulusan dan mempercepat waktu studi mahasiswa melalui berbagai upaya bimbingan oleh dosen penasehat akademik.

Untuk memenuhi kebutuhan prodi akan suatu pola klasifikasi kelulusan mahasiswa tersebut maka pada penelitian ini penulis menggunakan Algoritma C4.5 dalam membuat pohon keputusan. Algoritma C4.5 itu sendiri merupakan salah satu algoritma klasifikasi data terbaik dan sudah banyak digunakan pada berbagai kasus penelitian seperti Klasifikasi Penyakit Berdasarkan Kode Penyakit Internasional Menggunakan Algoritma C4.5 (Ina,2013), Prediksi Kompetensi Karyawan Menggunakan Algoritma C4.5 (Sunga, SENTIKA 2018).

Data yang digunakan adalah data Indeks Perstasi (IP) mahasiswa angkatan tahun 2008-2010 pada 5 (Lima) semester pertama perkuliahan sebagai data pelatihan untuk pembentukan pohon keputusan, kemudian data Indeks Perstasi (IP) mahasiswa pada 5 (Lima) semester pertama perkuliahan dari mahasiswa angkatan tahun 2011-2012 digunakan sebagai data uji. Data nilai tersebut akan diklasifikasikan ke dalam beberapa tingkat kelulusan yaitu cepat, sedang, dan lambat sesuai dengan aturan masa studi jenjang S1.

Salah satu teknik Data Mining yang umum digunakan adalah Teknik Klasifikasi. Klasifikasi adalah sebuah proses untuk menemukan model yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang kelasnya tidak diketahui (Tan et all, 2004). Di dalam klasifikasi diberikan sejumlah record yang dinamakan training set, yang terdiri dari beberapa atribut, atribut dapat berupa kontinyu ataupun kategoris, salah satu atribut menunjukkan kelas untuk record.

Algoritma C4.5 adalah algoritma klasifikasi data dengan teknik *decison tree* yang terkenal dan disukai karena memiliki kelebihan-kelebihan. Kelebihan ini misalnya: dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat di antara algoritma-algoritma yang menggunakan memori utama di komputer.

Proses klasifikasi terdiri dari dua tahap, yaitu tahap belajar dari data pelatihan untuk menghasilkan model dan tahap klasifikasi yang menggunakan model untuk prediksi kelas. Pada tahap belajar dari data, algoritma C4.5 mengkonstruksi *decison tree* dari data pelatihan, yang berupa kasus-kasus atau record-record. Setiap kasus berisikan nilai dari atribut-atribut untuk sebuah kelas. Setiap atribut dapat berisidatadiskretataukontinyu(numerik). C4.5 juga menangani kasus yang tidak memiliki nilai untuk sebuah atau lebih atribut. Akan tetapi, atribut kelas hanya bertipe nominal dan tidak bolehkosong.

Sebuah obyek yang diklasifikasikan dalam pohon harus dites nilai entropinya. *Entropy* adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari *impurity*, dan *homogenity* dari kumpulan data. Entropi digunakan untuk memprediksi rata-rata jumlah bit per simbol yang diperlukan untuk mentransmisikan data, misalnya X, jika peluang kejadian setiap simbol pada data yang akan ditransmisikan diketahui (MacKay, D. J.C, 2003) dan (Moore, 2003). Dari nilai *entropy* tersebut kemudian dihitung nilai *Information gain* (IG) masing-masing atribut, rumus *entropy* didefinisikan pada persamaan (1)

$$E(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (1)$$

Keterangan :

- S = himpunan Kasus
- n = jumlah Pastisi S
- pi = proporsi dari Si terhadap S

Pada beberapa kasus, jika semua member dari set sample S memiliki nilai yang sama, maka nilai *entropy* adalah nol. Hal itu berarti tidak ada klasifikasi yang mempunyai ketidakpastian (uncertainty). Namun jika jumlah sample positif sama dengan jumlah sample negatif, maka *entropy* bernilai 1, berarti ketidakpastian bernilai maksimum. Ini berarti set sample tidak mempunyai ketidakpastian (keputusan dari *decision tree* sudah jelas). Jika jumlah sample positif tidak sama dengan jumlah sample negatif, maka nilai *entropy* akan berada diantara 0 dan 1. Dari rumus *entropy* diatas dapat disimpulkan bahwa definisi *entropy* (S) adalah jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada suatu ruang sample S. *Entropy* bisa dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. Semakin kecil nilai *entropy* maka semakin baik digunakan dalam mengekstraksi suatukelas.

Setelah mendapat nilai *entropy* untuk suatu kumpulan data, maka kitadapat mengukur efektivitas suatu atribut dalam mengklasifikasikan data. Ukuran efektifitas ini disebut *information gain*. Secara matematis, *infomation gain* dari suatu atribut A, rumusnya didefinisikan pada persamaan (2)

$$G(S, A) = E(S) - \sum_{i=1}^n \frac{S_i}{S} * E(S_i) \quad (2)$$

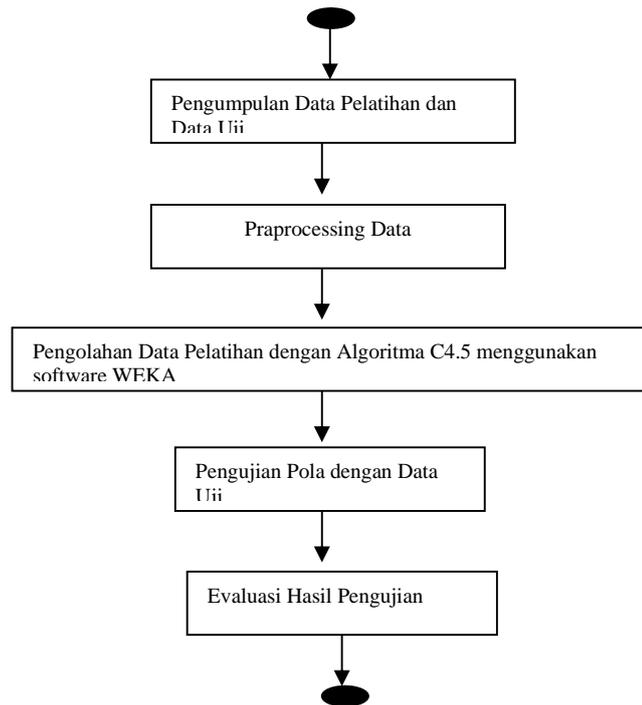
Keterangan :

- S = himpunan kasus
- A = atribut
- n = jumlah partisi atribut A
- |Si| = jumlah kasus pada partisi ke-i
- |S| = jumlah kasus dalam S

2. METODE PENELITIAN

Penelitian ini bersifat eksperimen dan analisis matematis dengan mengambil data sampel secara acak pada Prodi Teknik elektro FST Undana.

Adapun proses pelaksanaan penelitian digambarkan dalam diagram alir pada gambar 2.1 berikut:



Gambar 2.1 Tahapan Penelitian

1. Pengumpulan Data Pelatihan dan Data Uji
Pengumpulan data merupakan tahap awal penelitian. Data data yang dibutuhkan adalah data IP mahasiswa angkatan tahun 2008-2010 pada 5 (Lima) semester pertama perkuliahan sebagai data pelatihan untuk pembentukan pohon keputusan, kemudian data IP mahasiswa pada 5 (Lima) semester pertama perkuliahan dari mahasiswa angkatan tahun 2011-2012 digunakan sebagai data uji.
2. Praprocessing Data
Pada tahap ini dilakukan praprocessing data dimana data IP dibagi dalam kategori sangat baik, baik, cukup baik, kurang. Sedangkan masa studi dibagi dalam tingkat cepat, pas, cukup cepat, lambat sesuai lama tahun studi yaitu 3,5 tahun sampai 7 tahun. Data-data ini dibuat dalam file Excel dalam format .CSV
3. Pengolahan Data menggunakan Algoritma C4.5 dengan software WEKA
Proses utama terjadi pada tahapan ini, yakni data pelatihan dalam format.CSV dimasukkan kedalam software WEKA 3.8.2 untuk diproses menggunakan algoritma C4.5 agar menghasilkan pola pohon keputusan tingkat kelulusan mahasiswa
4. Pengujian Pola dengan Data Uji
Pada tahapan ini, pola pohon keputusan yang dihasilkan diuji menggunakan data uji.
5. Evaluasi Hasil Pengujian
Proses terakhir adalah mengevaluasi hasil pengujian, apakah pola yg dihasilkan Algoritma C4.5 pada tahap 3 sesuai dengan hasil pengujian menggunakan data uji. Jika hasil pengujian sesuai maka pola pohon keputusan yang dihasilkan dari penelitian ini dapat direkomendasikan kepada Prodi Teknik Elektro FST untuk dapat diterapkan.

3. HASIL YANG DIPEROLEH

a. Tahap Praprosesing Data

Data yang digunakan dalam penelitian ini berupa data mentah yang diambil oleh asisten peneliti pada Program Studi Teknik Elektro - FST – Undana. Data nilai IP mahasiswa yang diberikan oleh prodi kemudian dikategorikan dalam 3 kelas yaitu :

1. Nilai IP > 2,75 dikategorikan dalam Kelas Tinggi
2. Nilai IP 2,00 – 2,75 dikategorikan dalam Kelas Sedang
3. Nilai IP < 2,00 dikategorikan dalam Kelas Rendah seperti yang ditampilkan pada gambar 3.1 berikut :

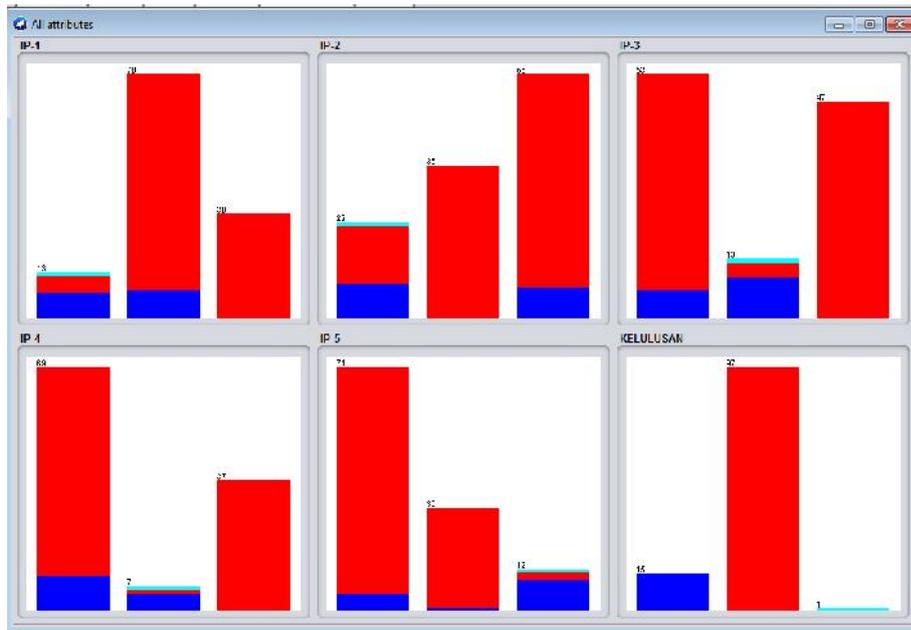
	A	B	C	D	E	F	G	H	I	J	K	L
1	IP-1	IP-2	IP-3	IP-4	IP-5	KETULUSAN						
2	tinggi	tinggi	sedang	sedang	sedang	sedang						
3	tinggi	sedang	sedang	sedang	rendah	sedang						
4	sedang	tinggi	tinggi	tinggi	sedang	sedang						
5	tinggi	sedang	sedang	sedang	sedang	lama						
6	tinggi	tinggi	tinggi	tinggi	tinggi	cepat						
7	rendah	rendah	sedang	sedang	sedang	lama						
8	rendah	rendah	rendah	sedang	sedang	lama						
9	rendah	rendah	sedang	sedang	sedang	sedang						
10	sedang	tinggi	sedang	sedang	tinggi	sedang						
11	sedang	sedang	rendah	sedang	rendah	lama						
12	sedang	sedang	sedang	rendah	sedang	lama						
13	sedang	sedang	tinggi	sedang	tinggi	sedang						
14	sedang	rendah	rendah	sedang	rendah	lama						
15	rendah	rendah	sedang	sedang	sedang	sedang						
16	rendah	sedang	rendah	sedang	sedang	lama						
17	rendah	rendah	sedang	sedang	sedang	lama						
18	tinggi	sedang	sedang	sedang	rendah	lama						
19	tinggi	sedang	tinggi	sedang	sedang	sedang						
20	sedang	sedang	rendah	rendah	sedang	lama						
21	sedang	sedang	sedang	rendah	sedang	lama						
22	sedang	sedang	rendah	sedang	sedang	lama						
23	sedang	sedang	rendah	rendah	sedang	lama						
24	sedang	tinggi	sedang	sedang	tinggi	sedang						
25	sedang	sedang	sedang	sedang	rendah	lama						

Gambar 3.1. Data IP mahasiswa berdasarkan kategori dalam penelitian

Tingkat kelulusan dikategorikan juga dalam 3 kelas yaitu :

1. Kelas Cepat untuk masa studi 8 semester
2. Kelas Sedang untuk masa studi 9 semester – 11 semester
3. Kelas Lama untuk masa studi 12 semester

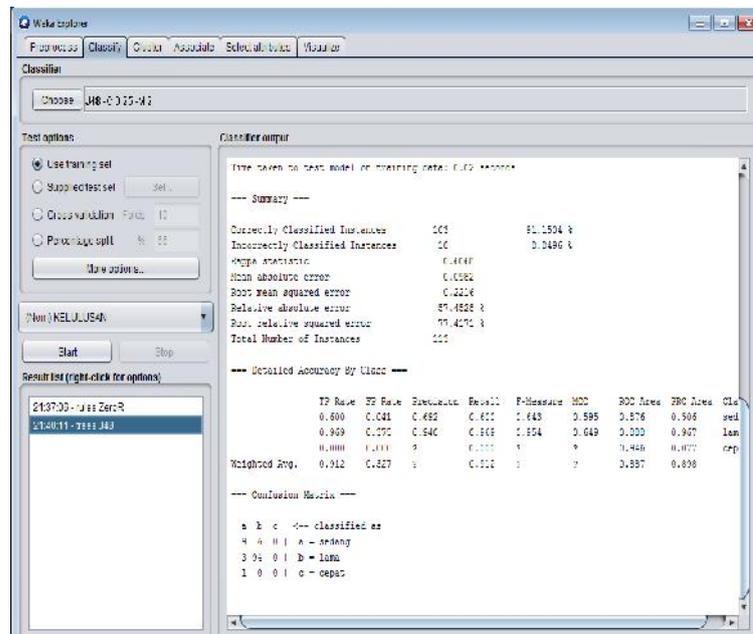
Visualisasi kondisi awal data menggunakan Software WEKA 3.8 ditampilkan pada gambar 3.2 berikut:



Dari gambar 3.2 dapat diketahui bahwa diagram batang berwarna Biru, Merah dan Biru Muda menunjukkan Kelas kelulusan dimana Warna Biru menunjukkan Kelas Lama, warna Merah menunjukkan Kelas Sedang dan warna Biru Muda menunjukkan Kelas Cepat dengan urutan dalam Tinggi, Sedang, Rendah dari masing-masing atribut IP yaitu IP-1, IP-2, IP-3, IP-4 dan IP-5

b. Tahap Pengolahan Data

Pada tahap ini data yang diperoleh dari praprosesing data diolah menggunakan Software WEKA 3.8 dengan menggunakan Algoritma C4.5 yang ada pada WEKA 3.8 untuk mendapatkan pohon keputusan sebagai hasil pengolahan data. Hal ini dapat ditampilkan pada gambar 3.3 berikut:



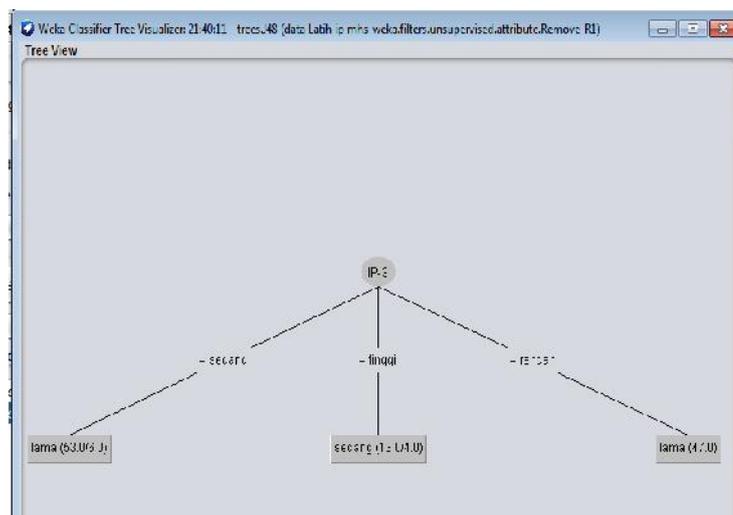
Gambar 3.3. Hasil Prosesing data pada WEKA menggunakan Algoritma C4.5

Dari gambar 3.3 dapat diketahui jumlah data keseluruhan sebanyak 113 data dimana data yang terklasifikasi dengan benar sebanyak 103 atau setara dengan 91,15 % data sedangkan 10 data dianggap sebagai data yang tidak terklasifikasi atau setara dengan 8,85% data. Nilai Relative Absolute Error sebesar 57,45% dan nilai Root Relative Squared Error sebesar 77,42%.

Klasifikasi yang diperoleh dari pengolahan data menggunakan algoritma C4.5 pada WEKA 3.8 adalah Tingkat Kelulusan Lama sebanyak 100 data, Kelulusan Sedang 13 data sedangkan Kelulusan Cepat 0 data. Hal ini terjadi karena pada data yang diperoleh dari prodi, hanya 5 orang yang memiliki kategori lulusan cepat, sehingga persentasinya sangat kecil jika dibandingkan dengan jumlah keseluruhan data.

c. Hasil Pohon Keputusan

Dari proses pengolahan data yang sudah dilaksanakan menggunakan software WEKA 3.8 dengan algoritma C4.5 diperoleh pohon keputusan sebagai hasil penelitian ini, yang ditampilkan pada gambar 3.4 berikut:



Pada gambar 3.4 menunjukkan pohon keputusan sebagai hasil akhir dari penelitian ini.

Hasil Klasifikasi yang diperoleh sebagai berikut: Jika IP semester 3 Sedang maka Kelulusan Lama, Jika IP semester 3 Tinggi maka Kelulusan Sedang, Jika IP semester 3 Rendah maka Kelulusan Lama Hasil pola ketentuan atau klasifikasi yang diperoleh adalah IP semester 3 (IP-3) menjadi penentu seorang mahasiswa

Prodi Teknik Elektro – FST dapat menyelesaikan studi dengan Cepat/Sedang/Lama. Hal ini sejalan atau selaras dengan keberadaan matakuliah semester 3 pada Prodi Teknik Elektro yang sesungguhnya berisi materi-materi kuliah dasar keteknikan yang menjadi dasar ilmu bagi matakuliah – matakuliah selanjutnya di semester atas.

4. KESIMPULAN

Kesimpulan.

Berdasarkan hasil pengolahan data yang dilakukan peneliti menggunakan Algoritma C4.5 pada software WEKA 3.8 diperoleh hasil aturan klasifikasi tingkat kelulusan mahasiswa Prodi Teknik Elektro FST Undana sebagai berikut:

- Jika **IP semester 3 Sedang** maka **Kelulusan Lama**
- Jika **IP semester 3 Tinggi** maka **Kelulusan Sedang**
- Jika **IP semester 3 Rendah** maka **Kelulusan Lama**

Hasil penelitian ini memberikan gambaran kepada seluruh sivitas akademik Prodi Teknik Elektro FST undana bahwa perkuliahan pada semester 3 sangat menentukan hasil akhir studi mahasiswa.

Saran.

Adapun beberapa saran yang diberikan peneliti berdasarkan hasil penelitian ini sbb:

- Perlu adanya penelitian lanjutan dengan algoritma klasifikasi yang lain tanpa merubah data agar dapat membandingkan hasilnya dengan hasil yang sudah didapat oleh peneliti pada kesempatan ini.

- Hasil penelitian ini dapat dijadikan rujukan oleh para dosen Penasehat Akademik pada Prodi Teknik Elektro FST Undana dalam memberikan bimbingan dan arahan bagi mahasiswa binaannya sehingga dapat meningkatkan mutu lulusan dan juga mengurangi lama masa studi mahasiswa.

DAFTAR PUSTAKA

- Berry, Michael, dan Gordon. (2004). *Data Mining Techniques for Data Mining*. John Willey & Sons, Inc
- Bertalya. (2009), *Konsep Data Mining, Klasifikasi: Pohon keputusan*, Universitas Gunadarma, Jakarta
- Dwi, A.S.,Dimas., (2013), *Model Prediksi Tingkat Kelulusan Mahasiswa Dengan Teknik Data Mining Menggunakan Metode Decision Tree C4.5*, Skripsi. UNY, Yogyakarta
- Kursini dan Luthfi., (2009). *Algoritma Data Mining*, Andi Offset, Yogyakarta