



IDENTIFIKASI MAHASISWA YANG MEMPUNYAI KECENDERUNGAN LULUS TIDAK TEPAT WAKTU PADA PROGRAM STUDI MMT-ITS DENGAN MENGGUNAKAN ALGORITMA C4.5

IDENTIFICATION OF STUDENTS WHO HAVE GRADUATED NOT IN TIME ON MMT-ITS USING C4.5 CLASSIFICATION ALGORITHM

Amelia Halim¹⁾ dan Joko Lianto Buliali²⁾

*1) Program Studi Magister Manajemen Teknologi, Institut Teknologi Sepuluh Nopember,
Jl. Cokroaminoto 12A, Surabaya, 60264, Indonesia
email:Amell.friend@gmail.com*

*2)Program Studi Teknik Informatika, Institut Teknologi Sepuluh Nopember
email : joko@its-sby.edu*

ABSTRAK

Setiap tahun Universitas menerima mahasiswa dari berbagai program, lokasi yang berbeda, latar belakang pendidikan yang beraneka ragam dan berbagai nilai dalam ujian masuk. Selain itu juga berasal dari universitas dengan akreditasi yang berbeda, masing-masing universitas memiliki kurikulum pelajaran yang berbeda dan tingkat kedalaman mata pelajaran yang diberikan juga berbeda. Proses menganalisis masa lalu kinerja mahasiswa akan memberikan perspektif yang lebih baik dari kemungkinan kinerja mahasiswa di masa depan. Penelitian ini menggunakan data pendidikan berupa sampel data alumni dari mahasiswa lulusan MMT – ITS dari angkatan 2009-2012. Metode yang digunakan adalah algoritma C4.5 yang merupakan salah satu teknik pengklasifikasian dalam data mining. Penelitian ini akan melakukan analisis untuk memperoleh informasi tentang kecenderungan mahasiswa lulus tepat waktu atau lulus tidak tepat waktu. Penelitian ini akan diuji menggunakan akurasi, recall dan precision. Uji coba dilakukan dalam 3 macam skenario. Skenario I mempunyai rata-rata nilai precision, recall dan akurasi yaitu 46.27%, 31.1 %, dan 60.35%. Skenario II mempunyai rata-rata nilai precision, recall dan akurasi yaitu 47.26%, 32.05%, dan 61.25%. Skenario II mempunyai rata-rata nilai precision, recall dan akurasi yaitu 47.61%, 35.39%, dan 61.50%.

Kata kunci: Algoritma C4.5, Mahasiswa, Lulus tepat waktu.

ABSTRACT

Every year, educational institutes admit students under various courses from different locations, educational background and with varying merit scores in entrance examinations. Moreover, schools and junior colleges may be affiliated to different boards, each board having different subjects in their curricula and also different level of depths in their subjects. Analyzing the past performance of admitted students would provide a better perspective of the probable academic performance of students in the future. This study uses the data in the form of education alumni data sample of graduate students MMT - ITS of force from 2009 to 2012. The method used is a C4.5 algorithm which is one of classification techniques in data mining. This study will conduct an analysis to obtain information about the tendency of students to graduate on time or do not pass on time. This study will be tested using the recall and precision. The trial was conducted in 3 kinds of scenarios. Scenario I have an average value of



precision, recall and accuracy is 46.27%, 31.1%, and 60.35%. Scenario II has an average value of precision, recall and accuracy is 47.26%, 32.05%, and 61.25%. Scenario II has an average value of precision, recall and accuracy is 47.61%, 35.39%, and 61.50%.

Keyword: C4.5 Algorithm, college student, graduate ontime.

PENDAHULUAN

Kualitas perguruan tinggi, khususnya program studi di Indonesia diukur berdasarkan akreditasi yang dilaksanakan oleh Badan Akreditasi Nasional Perguruan Tinggi atau BAN PT. Menurut BAN PT (BAN PT., 2011) kualitas tersebut diukur berdasarkan 7 standar utama, salah satunya adalah Mahasiswa dan Lulusan. Khusus mengenai evaluasi standar mahasiswa dan lulusan, komponen yang dinilai adalah: sistem rekrutmen mahasiswa baru, dan lulusan (rata-rata masa studi dan IPK). Berdasarkan uraian diatas, maka dapat diambil kesimpulan bahwa kualitas sebuah perguruan tinggi salah satunya ditentukan oleh rekrutmen mahasiswa baru dan lama masa studi mahasiswa.

Salah satu permasalahan utama dari institusi perguruan tinggi adalah untuk meningkatkan kualitas pendidikan untuk mahasiswa dan untuk meningkatkan kualitas dari keputusan manajerial institusi. Salah satu cara untuk mencapai kualitas level mutu tertinggi dari sistem perguruan tinggi adalah dengan menggali pengetahuan dari data bidang pendidikan sebagai atribut pembelajaran utama yang mempengaruhi pencapaian mahasiswa (Abu., et al, 2012). Data-data bidang pendidikan pada umumnya bisa berupa data profile mahasiswa, mata kuliah, KRS (kartu rencana studi), data alumni, dan sebagainya, yang biasanya tersimpan dalam database Sistem Informasi Akademik Kampus (SIAK) dalam jumlah yang besar, data bidang pendidikan tersebut dapat digunakan untuk menggali sebuah informasi.

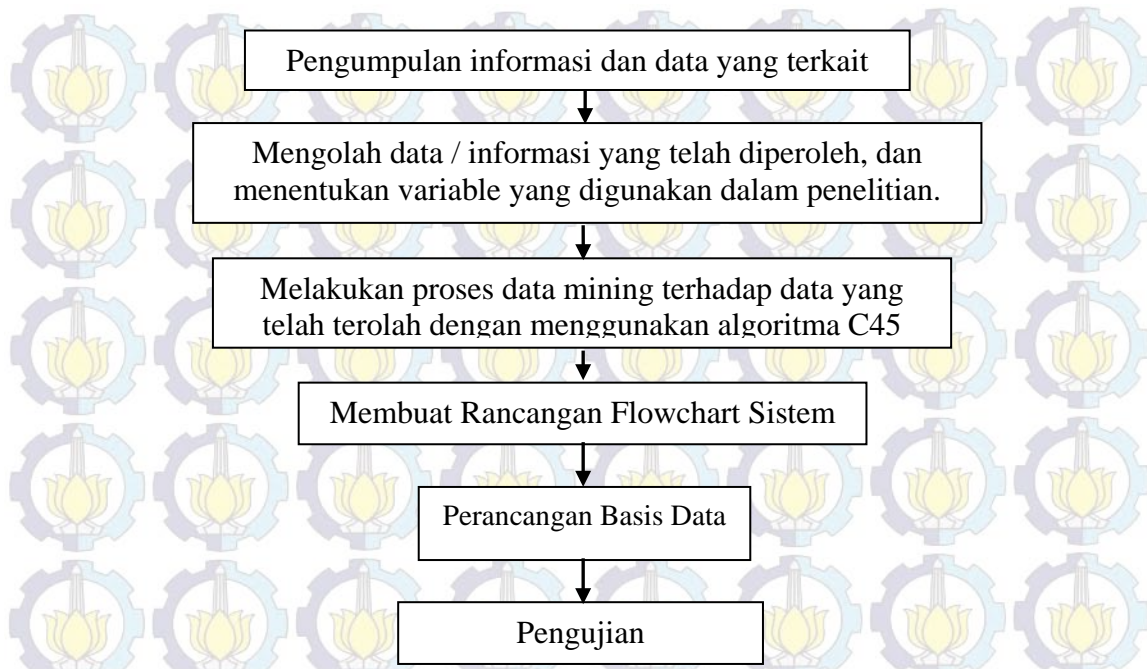
Setiap tahun, lembaga pendidikan menerima mahasiswa dalam berbagai program dari lokasi yang berbeda, latar belakang pendidikan dan dengan berbagai nilai dalam ujian masuk. Selain itu juga berasal dari universitas dengan akreditasi yang berbeda, masing-masing universitas memiliki berbeda pelajaran dalam kurikulum dan tingkat juga berbeda pada kedalaman mata pelajaran yang diberikan. Menganalisis masa lalu kinerja mahasiswa akan memberikan perspektif yang lebih baik dari kemungkinan kinerja mahasiswa di masa depan. Sebuah lembaga pendidikan perlu memiliki pengetahuan awal perkiraan siswa yang terdaftar untuk memprediksi kecenderungan lulus tidak tepat waktu. Pihak Universitas akan bisa mengambil kebijakan untuk lebih memperhatikan dan mengontrol mahasiswanya yang mempunyai kecenderungan lulus tidak tepat waktu.

Penelitian ini menggunakan sampel data alumni dari mahasiswa lulusan MMT – ITS dari angkatan 2009-2012. Metode yang digunakan adalah algoritma C4.5. Penelitian ini ditujukan untuk membangun aplikasi yang memberikan warning kepada dosen wali terhadap kecenderungan seorang mahasiswa lulus tidak tepat waktu pada awal masa studi mahasiswa. Diharapkan dari penelitian yang dilakukan terhadap sampel data alumni tersebut dapat diperoleh suatu informasi yang bisa membantu pihak institusi pendidikan untuk merancang strategi meningkatkan kualitas perguruan tingginya.

METODE

Berikut adalah proses penelitian yang digunakan untuk pengidentifikasi mahasiswa yang lulus tidak tepat waktu. Penelitian akan dilaksanakan sesuai dengan prosedur dan langkah berikut.

ISBN :



Gambar 1. Metodologi Penelitian

Langkah awal adalah mencari informasi dari pihak MMT ITS bagian akademik mengenai data histori mahasiswa. Informasi tersebut didapatkan dengan mengumpulkan data – data mahasiswa yang pernah mengenyam pendidikan di MMT ITS. Data – data tersebut dikumpulkan menjadi satu untuk dijadikan bahan penelitian. Setelah data terkumpul, langkah selanjutnya adalah mengolah data tersebut menjadi bentuk data yang dibutuhkan, membersihkan data, dan menyeleksi data yang akan digunakan sebagai inputan proses data mining. Selain itu, dibutuhkan juga untuk menentukan variable yang akan digunakan sebagai inputan proses data mining. Variabel – variabel tersebut yang akan digunakan untuk membangun pohon keputusan yang dibentuk oleh algoritma C45.

Pada Penelitian ini digunakan data sebanyak 178 data mahasiswa lulusan MMT – ITS dari angkatan 2009-2012. Dari proses pengumpulan data diperoleh 19 variabel meliputi NRP, Nama Perguruan Tinggi S1, Jurusan S1, IPK S1, Tahun Lulus S1, Tahun Masuk S2, Nama Perusahaan Tempat Bekerja, Pekerjaan, Nilai GMAT, Nilai Toefl, Nilai Materi Bidang, Nilai Wawancara, Nilai Skor Akhir, Program Studi S2, Bidang Keahlian S2, IPS Semester 1, IPS Semester 2, Sumber dana S2, dan Lama Studi. Variabel NRP digunakan sebagai Identifier karena sifatnya yang unik. Variable lama studi dalam satuan semester. Variable lama studi digunakan untuk menghasilkan class/Label dengan ketentuan lama studi ≤ 4 semester maka tergolong class Tepat Waktu, dan lama studi > 4 semester maka tergolong class Tidak Tepat Waktu. Variabel yang numerik dikonversi menjadi kategorikal dengan ketentuan pada tabel 1.

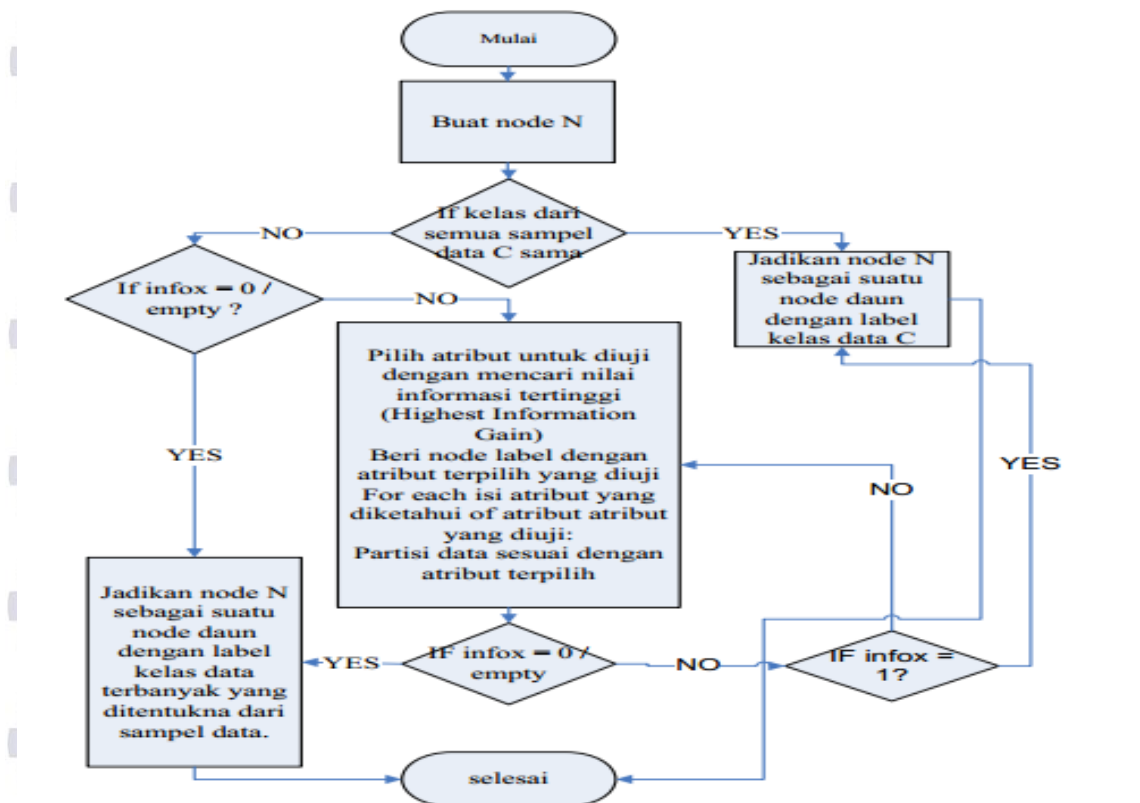
Tabel 1. Tabel Konversi IPK, GMAT, MBID, Wawancara, TOEFL, Sakhir

IPK S1		SAKHIR		GMAT		IPS S1	
Range	Kelas	Range	Kelas	Range	Kelas	Range	Kelas
4.00 - 3.79	A	100 - 92	A	100 - 86	A	4.00 - 3.68	A
3.78 - 3.58	B	91 - 84	B	85 - 72	B	3.67 - 3.36	B
3.57 - 3.37	C	83 - 76	C	71 - 58	C	3.35 - 3.04	C
3.36 - 3.16	D	75 - 68	D	57 - 44	D	3.03 - 2.72	D
< 3.16	E	< 68	E	< 44	E	< 2.72	E



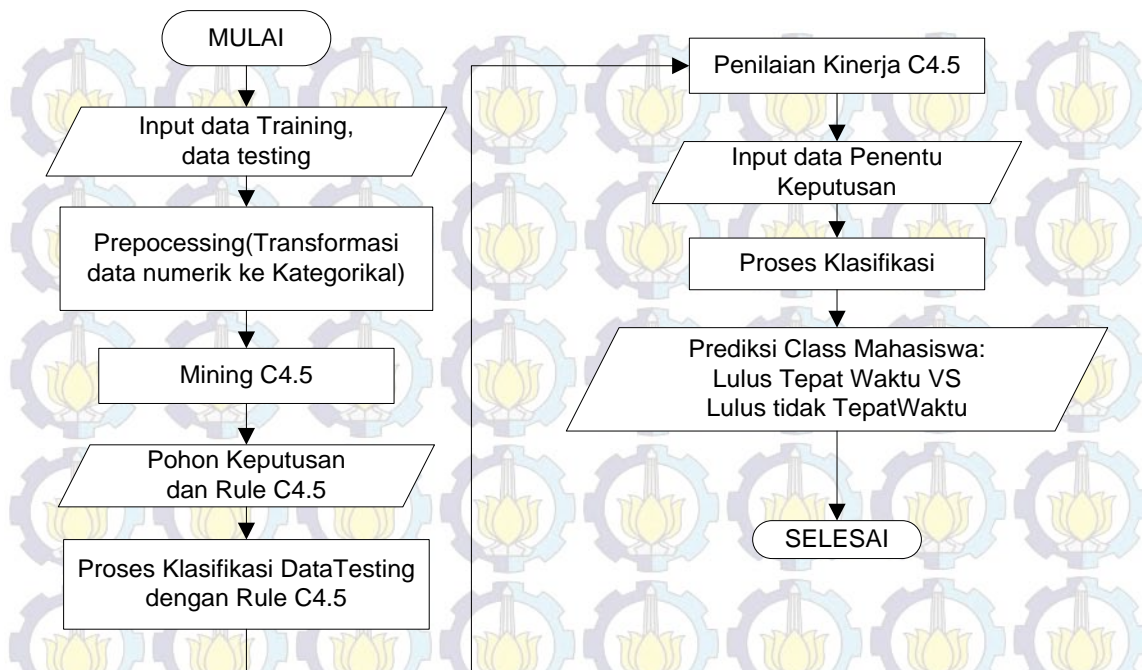
MBID		Wawancara		TOEFL		IPS S2	
Range	Kelas	Range	Kelas	Range	Kelas	Range	Kelas
100 - 86	A	100 - 92	A	677 - 604	A	4.00 - 3.6	A
85 - 72	B	91 - 84	B	603 - 531	B	3.5 - 3.2	B
71 - 58	C	83 - 76	C	530 - 457	C	3.1 - 2.8	C
57 - 44	D	75 - 68	D	456 - 384	D	2.7 - 2.4	D
< 44	E	< 68	E	< 384	E	< 2.4	E

Dari belasan variabel ada di atas, dipilih 14 variabel yang diikuti dalam proses klasifikasi yaitu Akreditasi Perguruan tinggi S1, Program Studi S2, Status Program Studi S1, Bidang Keahlian S2, Sumber Dana S2, IPK S1, IPS Semester 1, IPS Semester 2, Nilai GMAT, Nilai Toefl, Nilai Materi Bidang, Nilai Wawancara, Nilai Skor Akhir, Jarak S1-S2. Setelah melalui proses mengolah data, dilakukan mining data untuk membentuk suatu pohon keputusan. Pohon dibangun dengan cara membagi data secara rekursif hingga tiap bagian terdiri dari data yang berasal dari kelas yang sama menggunakan algoritma C4.5. Algoritma C4.5 digambarkan pada gambar 2.



Gambar 2. Flowchart Algoritma C4.5

Tahap berikutnya adalah perancangan *flowchart sistem*. Perancangan *flowchart* akan memudahkan pengembang untuk mengimplementasikan sistem ke dalam bahasa pemrograman, karena akan menjelaskan bagaimana cara kerja sistem dari awal hingga akhir. Pada tesis ini, *flowchart* berurutan yaitu input data Training dan data Testing, proses preprocessing yaitu transformasi data numeric ke categorical, Mining dengan algoritma C4.5, Pembentukan pohon keputusan dan Rule C4.5, Proses Klasifikasi Data Testing dengan rule C4.5, Penilaian Kinerja C4.5, Input data Penentu Keputusan, Proses Klasifikasi, dan Prediksi Class Mahasiswa : Lulus Tepat Waktu Vs Lulus tidak Tepat Waktu.



Gambar 3. Flowchart Sistem

Diagram Aliran Data / Data Flow Diagram (DFD) adalah sebuah teknis grafis yang menggambarkan aliran informasi dan transformasi yang diaplikasikan saat data bergerak dari input menjadi output. DFD dapat digunakan untuk menyajikan sebuah sistem atau perangkat lunak pada setiap tingkat abstraksi. DFD memberikan suatu mekanisme bagi pemodelan fungsional dan pemodelan informasi.

Pada tahap terakhir akan dilakukan pengujian sistem. Pengujian terhadap sistem dilakukan dengan berbagai variasi partisi data sample. Pengujian diukur dengan menggunakan perhitungan recall, precision dan accuracy. Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Sedangkan recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Accuracy didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual. Gambar 4 merupakan perumusan *recall*, *precision* dan *accuracy*.

		Nilai sebenarnya	
		TRUE	FALSE
Nilai prediksi	TRUE	TP (True Positive) Corect result	FP (False Positive) Unexpected result
	FALSE	FN (False Negative) Missing result	TN (True Negative) Corect absence of result

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Gambar 4. Perhitungan *Recall*, *Precision* dan *Accuracy*

HASIL DAN PEMBAHASAN

Pada Penelitian ini dilakukan ujicoba dengan 3 macam skenario. 3 macam skenario yaitu skenario pendeteksian saat mahasiswa baru diterima sebagai mahasiswa MMT ITS,



pendeteksian saat mahasiswa telah menyelesaikan semester 1 di MMT ITS, dan pendeteksian saat mahasiswa telah menyelesaikan semester 2 di MMT ITS.

Uji coba skenario I mengilustrasikan pada saat mahasiswa, baru diterima menjadi mahasiswa MMT ITS. Pengukuran hasil uji coba diukur dengan nilai precision, nilai recall dan nilai accuracy. Berikut akan disajikan dalam tabel nilai precision, recall dan accurasi dari hasil ujicoba yang telah dilakukan :

Tabel 2 Hasil Precision Uji Coba Skenario I

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	33.33%	37.00%	60.00%	28.57%	25.00%	66.67%	41.76%
II	44.83%	40.00%	50.00%	55.56%	50.00%	40.00%	46.73%
III	33.33%	28.57%	66.67%	75.00%	50.00%	100.00%	58.93%
IV	42.86%	30.00%	40.00%	50.00%	26.67%	50.00%	39.92%
V	27.27%	26.67%	50.00%	40.00%	20.00%	100.00%	43.99%
AVG	36.32%	32.45%	53.33%	49.83%	34.33%	71.33%	46.27%

Tabel 2 adalah Tabel Hasil Precision Uji Coba Skenario I. Pada skenario I ini dapat dilihat pada proses klasifikasi tanpa pruning, terjadi peningkatan persentase data training diikuti kecenderungan peningkatan nilai precision. Hal ini berarti peningkatan jumlah data training berpengaruh positif untuk peningkatan nilai precision. Sedangkan untuk klasifikasi menggunakan pre pruning, peningkatan jumlah data training tidak selalu meningkatkan nilai precision, terkadang malah membuat nilai precision menjadi lebih kecil. Hal ini diperkirakan disebabkan oleh keterbatasan data yang dimiliki peneliti untuk proses ujicoba menyebabkan pembentukan rule keputusan yang kurang sempurna pada proses klasifikasi dengan pre pruning. Pada Skenario ini, memiliki kecenderungan nilai precision menjadi lebih besar ketika dilakukan proses pre pruning.

Tabel 3 Hasil Recall Uji Coba Skenario I

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	31.58%	27.27%	37.50%	21.05%	18.18%	25.00%	26.76%
II	61.90%	46.15%	50.00%	23.81%	15.38%	33.33%	38.43%
III	12.00%	33.30%	25.00%	12.00%	12.50%	12.50%	17.88%
IV	50.00%	75.00%	40.00%	11.11%	50.00%	20.00%	41.02%
V	37.50%	30.77%	60.00%	12.50%	7.69%	40.00%	31.41%
AVG	38.60%	42.50%	42.50%	16.09%	20.75%	26.17%	31.10%

Tabel 3 adalah Tabel hasil Recall Uji Coba Skenario I. dari tabel tersebut dapat disimpulkan bahwa proses pre-pruning menyebabkan nilai Recall menurun. Pada proses Non Pruning maupun proses Pre Pruning, dapat dilihat bahwa bertambahnya data bisa menyebabkan nilai recall cenderung meningkat.

Tabel 4 Hasil Accuracy Uji Coba Skenario I

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	52.83%	63.89%	61.11%	52.83%	58.33%	61.11%	58.35%
II	54.72%	55.56%	66.67%	62.26%	63.89%	61.11%	60.70%
III	47.17%	50.00%	61.11%	56.60%	66.67%	61.11%	57.11%
IV	60.38%	55.56%	66.67%	66.04%	58.33%	72.22%	63.20%
V	50.94%	44.44%	72.22%	67.92%	55.56%	83.30%	62.40%
AVG	53.21%	53.89%	65.56%	61.13%	60.56%	67.77%	60.35%



Tabel 4 adalah Tabel hasil Accuracy Uji Coba Skenario I. Pada skenario I ini, proses pre pruning bisa meningkatkan nilai Accuracy, terbukti bahwa hampir semua nilai Accuracy pada partisi data bersesuaian yang disajikan tabel 4 mengalami peningkatan ketika dilakukan pre pruning. Peningkatan jumlah data pada skenario I, memiliki kecenderungan diikuti oleh peningkatan nilai Accuracy. Hal ini dapat disimpulkan bahwa bertambahnya data training meningkatkan nilai Accuracy hasil uji coba.

Uji coba skenario II mengilustrasikan pada saat mahasiswa, telah menyelesaikan semester I di MMT ITS. Skenario ini terdapat 1 variabel mining tambahan yaitu Index Prestasi Semester 1. Pengukuran hasil uji coba diukur dengan nilai precision, nilai recall dan nilai accuracy. Berikut akan disajikan dalam tabel nilai precision, recall dan accurasi dari hasil ujicoba yang telah dilakukan:

Tabel 5. Hasil Precision Uji Coba Skenario II

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	36.84%	30.00%	71.43%	33.33%	25.00%	66.67%	43.88%
II	46.15%	41.67%	50.00%	55.56%	50.00%	40.00%	47.23%
III	40.00%	36.36%	66.67%	75.00%	50.00%	100.00%	61.34%
IV	44.00%	33.33%	33.30%	50.00%	26.67%	50.00%	39.55%
V	27.27%	28.57%	50.00%	40.00%	20.00%	100.00%	44.31%
AVG	38.85%	33.99%	54.28%	50.78%	34.33%	71.33%	47.26%

Tabel 5 adalah Tabel Hasil Precision Uji Coba Skenario II. Pada skenario II ini dapat dilihat pada proses klasifikasi tanpa pruning, terjadi peningkatan persentase data training diikuti kecenderungan peningkatan nilai precision. Sedangkan untuk klasifikasi menggunakan pre pruning, peningkatan jumlah data training tidak selalu meningkat nilai precision, terkadang malah membuat nilai precision menjadi lebih kecil.

Tabel 6 Hasil Recall Uji Coba Skenario II

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	36.84%	27.27%	62.50%	26.32%	18.18%	25.00%	32.69%
II	57.14%	38.46%	50.00%	23.81%	15.38%	33.33%	36.35%
III	16.00%	33.33%	25.00%	12.00%	16.67%	12.50%	19.25%
IV	61.11%	75.00%	40.00%	11.11%	56.00%	20.00%	43.87%
V	37.50%	30.77%	40.00%	12.50%	7.69%	40.00%	28.08%
AVG	41.72%	40.97%	43.50%	17.15%	22.78%	26.17%	32.05%

Tabel 6 adalah Tabel hasil Recall Uji Coba Skenario II. Dari tabel tersebut dapat disimpulkan bahwa proses pre-pruning menyebabkan nilai Recall menurun. Pada proses Non Pruning maupun proses Pre Pruning, dapat dilihat bahwa bertambahnya data bisa menyebabkan nilai recall bertambah maupun berkurang. Hal ini bisa jadi disebabkan oleh variasi rule yang dibentuk dari proses klasifikasi data training yang dirandom.

Tabel 7 Hasil Accuracy Uji Coba Skenario II

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	54.72%	58.33%	72.22%	54.72%	58.33%	61.11%	59.91%
II	56.60%	58.33%	66.67%	62.26%	63.89%	61.11%	61.48%
III	49.06%	58.33%	61.10%	56.60%	66.67%	61.11%	58.81%
IV	60.38%	61.11%	61.11%	66.04%	58.33%	72.22%	63.20%
V	50.94%	47.22%	72.22%	67.92%	55.56%	83.33%	62.87%
AVG	54.34%	56.66%	66.66%	61.51%	60.56%	67.78%	61.25%



Tabel 7 adalah Tabel hasil Accuracy Uji Coba Skenario II. Pada scenario II ini, proses pre pruning bisa meningkatkan nilai Accuracy, terbukti bahwa hampir semua nilai Accuracy pada partisi data bersesuaian yang disajikan tabel 7 mengalami peningkatan ketika dilakukan pre pruning. Peningkatan jumlah data pada skenario II, memiliki kecenderungan diikuti oleh peningkatan nilai Accuracy. Hal ini dapat disimpulkan bahwa bertambahnya data training meningkatkan nilai Accuracy hasil uji coba.

Uji coba skenario III mengilustrasikan pada saat mahasiswa, telah menyelesaikan semester I dan semester II di MMT ITS. Skenario ini terdapat 2 variabel mining tambahan yaitu Index Prestasi Semester 1 dan Index Prestasi Semester 2. Berikut akan disajikan dalam tabel nilai precision, recall dan accurasi dari hasil ujicoba yang telah dilakukan :

Tabel 8 Hasil Precision Uji Coba Skenario III

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	40.91%	9.09%	57.14%	33.33%	25.00%	57.14%	37.10%
II	48.15%	41.67%	42.86%	55.56%	50.00%	40.00%	46.37%
III	41.67%	45.95%	75.00%	75.00%	50.00%	100.00%	64.60%
IV	42.86%	31.58%	42.86%	50.00%	26.67%	50.00%	40.66%
V	31.82%	44.00%	60.00%	40.00%	20.00%	100.00%	49.30%
AVG	41.08%	34.46%	55.57%	50.78%	34.33%	69.43%	47.61%

Tabel 8 adalah Tabel Hasil Precision Uji Coba Skenario III. Pada skenario III ini dapat dilihat pada proses klasifikasi tanpa pruning, terjadi peningkatan persentase data training diikuti kecenderungan peningkatan nilai precision. Sedangkan untuk klasifikasi menggunakan pre pruning, peningkatan jumlah data training tidak selalu meningkat nilai precision, terkadang malah membuat nilai precision menjadi lebih kecil.

Tabel 9 Hasil Recall Uji Coba Skenario III

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	47.37%	9.09%	50.00%	26.32%	18.18%	50.00%	33.49%
II	61.90%	38.46%	50.00%	23.81%	15.38%	33.33%	37.15%
III	20.00%	41.67%	37.50%	12.00%	16.67%	12.50%	23.39%
IV	50.00%	75.00%	60.00%	11.11%	56.00%	20.00%	45.35%
V	43.75%	61.54%	60.00%	12.50%	7.69%	40.00%	37.58%
AVG	44.60%	45.15%	51.50%	17.15%	22.78%	31.17%	35.39%

Tabel 9 adalah Tabel hasil Recall Uji Coba Skenario III. Dari tabel tersebut dapat disimpulkan bahwa proses pre-pruning menyebabkan nilai Recall menurun. Pada proses Non Pruning maupun proses Pre Pruning, dapat dilihat bahwa bertambahnya data bisa menyebabkan nilai recall bertambah maupun berkurang. Hal ini bisa jadi disebabkan oleh variasi rule yang dibentuk dari proses klasifikasi data training yang dirandom.

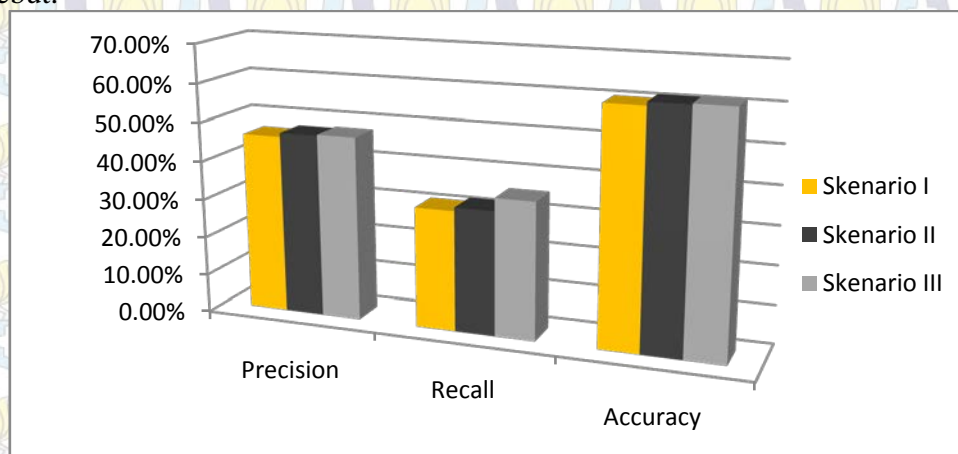
Tabel 10 Hasil Accuracy Uji Coba Skenario III

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	56.60%	44.44%	61.11%	54.72%	58.33%	61.11%	56.05%
II	58.49%	58.33%	61.11%	62.26%	63.89%	61.11%	60.87%
III	49.06%	63.89%	66.67%	56.60%	66.67%	61.11%	60.67%
IV	60.38%	58.33%	66.67%	66.04%	58.33%	72.22%	63.66%
V	54.72%	58.33%	77.78%	67.92%	55.56%	83.33%	66.27%
AVG	55.85%	56.66%	66.67%	61.51%	60.56%	67.78%	61.50%



Tabel 10 adalah Tabel hasil Accuracy Uji Coba Skenario III. Pada skenario III ini, proses pre pruning bisa meningkatkan nilai Accuracy, terbukti bahwa hampir semua nilai Accuracy pada partisi data bersesuaian yang disajikan tabel 10 mengalami peningkatan ketika dilakukan pre pruning. Peningkatan jumlah data pada skenario III, memiliki kecenderungan diikuti oleh peningkatan nilai Accuracy. Hal ini dapat disimpulkan bahwa bertambahnya data training meningkatkan nilai Accuracy hasil uji coba.

Pembeda tiga skenario tersebut adalah variable Index Prestasi Semester 1 (IPS1) dan Index Prestasi Semester 2(IPS 2). Pada skenario 1 IPS1 maupun IPS2 tidak ikut sebagai variable mining. Pada skenario II, variabel IPS1 ikut dalam proses mining karena mahasiswa telah menempuh studi di MMT ITS selama satu semester. Pada Skenario III, variabel IPS1 dan IPS2 ikut dalam proses mining karena mahasiswa telah menempuh studi di MMT ITS selama 2 semester yaitu pada semester 1 dan semester 2 sehingga mahasiswa telah memiliki kedua nilai tersebut.



Gambar 5. Perbandingan skenario I, skenario II, Skenario III

Pada gambar 5 dapat dilihat bahwa terjadi sedikit peningkatan Precision, Recall dan accuracy ketika variabel IPS1 dan IPS2 diikutkan dalam proses mining. Dugaan awal yang memperkirakan adanya peranan IPS1 dan IPS2 dalam pendeteksian lulus tidak tepat waktu adalah benar, walaupun pada percobaan ini peranan yang diberikan tergolong kecil. Hal ini membuktikan bahwa nilai IPS 1 dan nilai IPS2 mempunyai peranan yang cukup signifikan terhadap ketepatan lulus mahasiswa dalam menyelesaikan studinya di MMT ITS. Selain melakukan ujicoba dengan 3 skenario, selanjutnya dilakukan ujicoba dengan pemisahan program studi yaitu program studi MTI, MI dan MP. Pada Uji Coba program studi MTI menggunakan data training mahasiswa yang mengambil program studi MTI. Banyak data yang digunakan dalam uji coba ini adalah sebanyak 67 data.

Tabel 11 Hasil Uji Coba Program Studi MTI

SKN	Uji Coba	NON-PRUNING			PRE - PRUNING		
		Precision	Recall	Acc.	Precision	Recall	Acc.
I	VI	40.00%	100.00%	57.14%	50.00%	100.00%	71.43%
	VII	33.33%	100.00%	71.43%	25.00%	100.00%	57.14%
	VIII	33.33%	50.00%	57.14%	100.00%	50.00%	85.71%
	IX	33.33%	33.33%	42.86%	33.33%	33.33%	42.86%
	X	100.00%	25.00%	57.14%	50.00%	25.00%	42.86%
II	VI	33.33%	100.00%	42.86%	50.00%	100.00%	71.43%
	VII	33.33%	100.00%	71.43%	25.00%	100.00%	57.14%
	VIII	33.33%	50.00%	57.14%	100.00%	50.00%	85.71%
	IX	33.33%	33.33%	42.86%	33.33%	33.33%	42.86%



SKN	Uji Coba	NON-PRUNING			PRE – PRUNING		
		Precision	Recall	Acc.	Precision	Recall	Acc.
III	X	100.00%	25.00%	57.14%	50.00%	25.00%	42.86%
	VI	40.00%	100.00%	57.14%	66.67%	100.00%	85.71%
	VII	33.33%	100.00%	71.43%	25.00%	100.00%	57.14%
	VIII	100.00%	50.00%	85.71%	100.00%	50.00%	85.71%
	IX	33.33%	33.33%	42.86%	33.33%	33.33%	42.86%
	X	100.00%	25.00%	57.14%	50.00%	25.00%	42.86%
	AVG	52.00%	61.67%	58.09%	52.78%	61.67%	60.95%

Pada Tabel 11 dapat dilihat hasil uji coba menunjukkan dengan dikelompokannya berdasarkan program studi MTI maka nilai recall menjadi lebih besar dan cenderung mendekati sempurna, dan nilai Precisionnya menjadi lebih tinggi dibanding dengan percobaan sebelumnya yang melibatkan semua program studi. Ketika variabel nilai IPS1 dan IPS2 ikut dalam proses mining nilai akurasi cenderung meningkat, maka dapat disimpulkan bahwa pada mahasiswa program studi MTI nilai IPS 1 dan IPS 2 memberikan pengaruh yang cukup signifikan terhadap ketepatan lulus mahasiswa.

Pada Uji Coba program studi MI menggunakan data training mahasiswa yang mengambil program studi MI saja. Dengan partisi data yang digunakan dalam uji coba ini adalah 90:10 dan dilakukan perulangan uji coba sebanyak 3 kali dengan 3 macam Skenario. Skenario sama dengan skenario pada bab bab sebelumnya. Banyak data yang digunakan dalam uji coba ini adalah sebanyak 73 data.

Tabel 12. Hasil Uji Coba Program Studi MI

SKN	Uji Coba	NON-PRUNING			PRE – PRUNING		
		Precision	Recall	Acc.	Precision	Recall	Acc.
I	XI	33.33%	100.00%	71.43%	50.00%	100.00%	85.71%
	XII	25.00%	50.00%	42.86%	50.00%	100.00%	85.71%
	XIII	33.33%	25.00%	28,57%	100.00%	25.00%	57.14%
	XIV	50.00%	66.67%	57.14%	100.00%	33.33%	71.43%
	XV	100.00%	50.00%	85.71%	100.00%	50.00%	85.71%
II	XI	33.33%	100.00%	71.43%	50.00%	100.00%	85.71%
	XII	33.33%	50.00%	57.14%	100.00%	50.00%	85.71%
	XIII	50.00%	25.00%	42.86%	100.00%	25.00%	57.14%
	XIV	50.00%	66.67%	57.14%	100.00%	33.33%	71.43%
	XV	100.00%	50.00%	85.71%	100.00%	50.00%	85.71%
III	XI	25.00%	100.00%	57.14%	50.00%	100.00%	85.71%
	XII	25.00%	50.00%	42.86%	100.00%	50.00%	85.71%
	XIII	50.00%	25.00%	42.86%	100.00%	25.00%	57.14%
	XIV	50.00%	66.67%	57.14%	100.00%	33.33%	71.43%
	XV	100.00%	50.00%	85.71%	100.00%	50.00%	85.71%
	AVG	50.55%	58.33%	61.22%	86.67%	55.00%	77.14%

Pada Tabel 12 dapat dilihat hasil uji coba menunjukkan dengan dikelompokannya berdasarkan program studi MI maka nilai Precision menjadi lebih besar , dan nilai Recallnya menjadi lebih tinggi dibanding dengan percobaan sebelumnya yang melibatkan semua program studi. Sedangkan nilai Accuracy menjadi lebih baik. Ketika variabel nilai IPS1 dan IPS2 ikut dalam proses mining nilai akurasi menjadi turun, maka dapat disimpulkan bahwa pada mahasiswa program studi MI nilai IPS 1 dan IPS 2 tidak memberikan pengaruh yang signifikan terhadap ketepatan lulus mahasiswa.

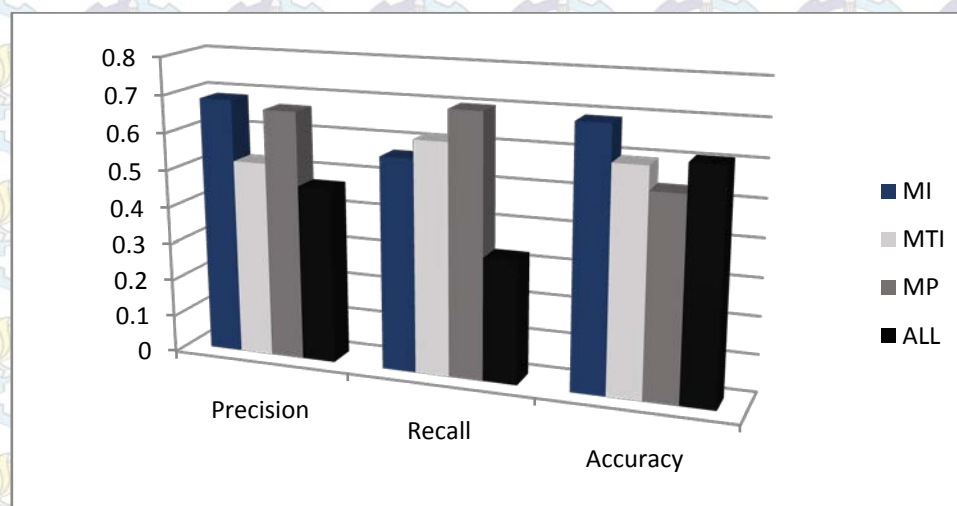


Pada Uji Coba program studi MP menggunakan data training mahasiswa yang mengambil program studi MP saja. Dengan partisi data yang digunakan dalam uji coba ini adalah 90:10 dan dilakukan perulangan uji coba sebanyak 3 kali dengan 3 macam Skenario. Skenario sama dengan skenario pada bab bab sebelumnya. Banyak data yang digunakan dalam uji coba ini adalah sebanyak 33 data.

Tabel 13. Hasil Uji Coba Program Studi MP

SKN	Uji Coba	NON-PRUNING			PRE - PRUNING		
		Precision	Recall	Acc.	Precision	Recall	Acc.
I	XVI	100.00%	50.00%	66.67%	100.00%	50.00%	66.67%
	XVII	50.00%	50.00%	33.33%	50.00%	50.00%	33.33%
	XVIII	33.33%	100.00%	33.33%	33.33%	100.00%	33.33%
	XIX	100.00%	50.00%	66.67%	100.00%	50.00%	66.67%
	XX	50.00%	100.00%	66.67%	50.00%	100.00%	66.67%
II	XVI	100.00%	50.00%	66.67%	100.00%	50.00%	66.67%
	XVII	50.00%	50.00%	33.33%	50.00%	50.00%	33.33%
	XVIII	33.33%	100.00%	33.33%	33.33%	100.00%	33.33%
	XIX	100.00%	50.00%	66.67%	100.00%	50.00%	66.67%
	XX	50.00%	100.00%	66.67%	50.00%	100.00%	66.67%
III	XVI	100.00%	50.00%	66.67%	100.00%	50.00%	66.67%
	XVII	50.00%	50.00%	33.33%	50.00%	50.00%	33.33%
	XVIII	33.33%	100.00%	33.33%	33.33%	100.00%	33.33%
	XIX	100.00%	50.00%	66.67%	100.00%	50.00%	66.67%
	XX	50.00%	100.00%	66.67%	50.00%	100.00%	66.67%
AVG		66.67%	70.00%	53.33%	66.67%	70.00%	53.33%

Pada Tabel 13 dapat dilihat hasil uji coba menunjukkan dengan dikelompokannya berdasarkan program studi MP maka nilai Precision menjadi lebih besar, dan memiliki nilai recallnya meningkat. Sedangkan untuk nilai accuracy menjadi lebih rendah dari percobaan yang melibatkan semua program studi. Hal ini mungkin disebabkan data training untuk program studi MP kurang. Nilai Accuracy, Recall dan Precision tetap walaupun mengikutkan variabel IPS1 dan IPS2 maupun diproses Pre Pruning.



Gambar 6 . Perbandingan Program Studi MI, MTI, MP dan ALL

Pada gambar 6 nilai yang disajikan adalah nilai rata-rata, dapat dilihat bahwa perlakuan pengelompokan berdasarkan program studi dapat meningkatkan nilai Precision, Recall dan Accuracy. Ditinjau dari nilai precision, data mahasiswa yang mengambil program studi MI



dan mahasiswa yang mengambil program studi MP mempunyai rata-rata nilai presisi dan rata-rata nilai precision yang tinggi dibandingkan dengan data mahasiswa yang mengambil program studi MI. Hal ini berarti hasil prediksi mahasiswa yang memiliki kecenderungan tidak tepat waktu pada data mahasiswa yang mengambil program studi MI atau mahasiswa yang mengambil program studi MP lebih baik.

KESIMPULAN

Pada penelitian ini dapat ditarik beberapa kesimpulan antara lain:

- Algoritma C4.5 dapat digunakan untuk memudahkan dalam pengambilan keputusan dengan memproyeksikan data-data ke dalam bentuk pohon keputusan, berdasarkan nilai entropy dan gain yang dimiliki masing-masing atribut data.
- Untuk hasil prediksi yang lebih akurat dibutuhkan data dalam jumlah besar, artinya semakin besar jumlah data yang digunakan maka semakin akurat hasil prediksi yang dihasilkan.
- Dari hasil uji coba yang telah dilakukan, klasifikasi dengan prepruning confidence level 95%, membuat rule semakin pendek dan terkadang membuat recall, precision dan accuracy menjadi lebih jelek. Hal ini disebabkan karena kurangnya data training,
- Dengan jumlah data training dan jumlah data testing yang sama, beda variasi data dapat mengakibatkan perbedaan recall, precision dan accuracy, serta rule yang dibentuknya. Oleh karena itu, kualitas data training yang di inputkan untuk klasifikasi memberikan kontribusi yang besar dalam pembentukan rule pada algoritma C4.5, selain jumlah data.
- Dari percobaan global seluruh program studi yang dilakukan, nilai IPS1 dan IPS2 dapat meningkat nilai precision, recall dan accuracy .
- Nilai accuracy, precision, dan recall cenderung meningkat ketika dilakukan permissahan per program studi di klasifikasi sendiri. Nilai Precision meningkat bila dikelompokkan per program studi. Hal ini membuktikan bahwa mahasiswa pada program studi akan banyak memiliki kesamaan, mengelompok. Nilai Recall Meningkat bila dikelompokkan per program studi. Hal ini membuktikan bahwa ketika rule yang dipakai untuk uji coba khusus terhadap program studi itu maka jangkauan sistem dapat memprediksi lebih tepat.

DAFTAR PUSTAKA

- Witten, I. H., Frank, E., Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools and Techniques* (3rd ed). USA: Elsevier
- Han, J. and Kamber, M., (2006) *Data Mining: Concepts and Techniques*, Elsevier.
- Sunjana. *Klasifikasi Data Nasabah sebuah Asuransi Menggunakan Algoritma C4.5 Seminar Nasional Aplikasi Teknologi Informasi 2010 (SNATI 2010)*. Yogyakarta.
- MySQL – The world’s most popular open source database, <http://www.mysql.com/>
- RapidMiner, <http://rapid-i.com/content/view/181/190/>
- Witten, Ian H. “Text Mining”