

**KOMPARASI METODE KLASIFIKASI DATA MINING
UNTUK PREDIKSI KETEPATAN KELULUSAN MAHASISWA**



TESIS

Diajukan sebagai salah satu syarat untuk memperoleh gelar Magister Ilmu
Komputer (M.Kom)

HILDA AMALIA
14000320

PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
NUSA MANDIRI
2012

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

Nama : Hilda Amalia
NIM : 14000320
Program Studi : Magister ilmu komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Management Information System*

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul: “Komparasi Metode Klasifikasi *Data Mining* Untuk Prediksi Ketepatan Kelulusan Mahasiswa” adalah hasil karya sendiri, dan semua sumber baik yang kutip maupun yang dirujuk telah saya nyatakan dengan benar dan tesis belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa tesis yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri dicabut/dibatalkan.

Jakarta, 29 Februari 2012

Yang menyatakan,

Hilda Amalia

HALAMAN PENGESAHAN

Tesis ini diajukan oleh:

Nama : Hilda Amalia
NIM : 14000320
Program Studi : Magister ilmu komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Management Information System*
Judul Tesis : “Komparasi Metode Klasifikasi Data Mining
Untuk Prediksi Ketepatan Kelulusan Mahasiswa”.

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Magister ilmu Komputer (M.Kom) pada Program pascasarjana Magister ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).

Jakarta, 29 Februari 2012
Pascasarjana Magister ilmu Komputer
STMIK Nusa Mandiri
Direktur,

H. Mochammand Wahyudi, MM,M.Kom

DEWAN PENGUJI

Penguji I : Dr. Khamami Herusantoso

Penguji II : Sfenrianto, M.Kom

Penguji III : Romi Satria Wahono, M. Eng

KATA PENGANTAR

Puji syukur alhamdulillah, penulis panjatkan kehadiran Allah, SWT, yang telah melimpahkan rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan tesis ini tepat pada waktunya. Dimana tesis ini penulis sajikan dalam bentuk buku yang sederhana. Adapun judul tesis, yang penulis ambil sebagai berikut “Komparasi metode klasifikasi data mining untuk prediksi ketepatan kelulusan mahasiswa”.

Tujuan penulisan tesis ini dibuat sebagai salah satu untuk mendapatkan gelar Magister Ilmu Komputer (M.Kom) pada Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).

Tesis ini diambil berdasarkan data kelulusan mahasiswa pada objek penelitian di salah satu universitas negeri yang berada di Jakarta. Penulis juga lakukan mencari dan menganalisa berbagai macam sumber referensi, baik dalam bentuk jurnal ilmiah, buku-buku literatur, internet, dll yang terkait dengan pembahasan pada tesis ini.

Penulis menyadari bahwa tanpa bimbingan dan dukungan dari semua pihak dalam pembuatan tesis ini, maka penulis tidak dapat menyelesaikan tesis ini tepat pada waktunya. Untuk itu ijinilah penulis kesempatan ini untuk mengucapkan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Romi Satria Wahono, M.Eng selaku pembimbing tesis yang telah menyediakan waktu, pikiran dan tenaga dalam membimbing penulis dalam menyelesaikan tesis ini.
2. Bapak Irsal selaku kepala PUSKOM (Pusat Komputer) Universitas negeri Jakarta.
3. Orang tua dan suami tercinta yang telah memberikan dukungan material dan moral kepada penulis.
4. Seluruh staf pengajar (dosen) Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa

Mandiri yang telah memberikan pelajaran yang berarti bagi penulis selama menempuh studi.

5. Seluruh staf dan karyawan Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri yang telah melayani penulis dengan baik selama kuliah.

Serta semua pihak yang terlalu banyak untuk penulis sebutkan satu persatu sehingga terwujudnya penulisan tesis ini. Penulis menyadari bahwa penulisan tesis ini masih jauh sekali dari sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan karya ilmiah yang penulis hasilkan untuk yang akan datang.

Akhir kata semoga tesis ini dapat bermanfaat bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Jakarta, 29 Februari 2012

Hilda Amalia
Penulis

**SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH
UNTUK KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini, saya :

Nama : Hilda Amalia
NIM : 14000320
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Management Information System*
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, dengan ini menyetujui untuk memberikan ijin kepada pihak Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri) Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalti-Free Right*) atas karya ilmiah kami yang berjudul : “Komparasi Metode Klasifikasi *Data Mining* untuk Prediksi Ketepatan Kelulusan Mahasiswa.”

Dengan Hak Bebas Royalti *Non-Eksklusif* ini pihak STMIK Nusa Mandiri berhak menyimpan, mengalih-media atau bentuk-kan, mengelolaannya dalam pangkalan data (database), mendistribusikannya dan menampilkan atau mempublikasikannya di internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari kami selama tetap mencantumkan nama kami sebagai penulis/pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak STMIK Nusa Mandiri, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 25 Februari 2012

Hilda Amalia
Penulis

ABSTRAK

Nama : Hilda Amalia
Nim : 14000320
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Management Information System*
Judul : “Komparasi Metode Klasifikasi Data Mining untuk
Prediksi Kelulusan Mahasiswa”.

Perguruan tinggi adalah tempat bagi para mahasiswa untuk menimba ilmu sebelum terjun kedalam dunia persaingan kerja. Jumlah mahasiswa yang lulus tepat waktu menjadi *indicator* keberhasilan dari sebuah perguruan tinggi baik negeri dan swasta. Penelitian dalam hal memprediksi kelulusan mahasiswa dan siswa telah banyak dilakukan. Dalam penelitian ini dilakukan perbandingan metode *data mining* yaitu algoritma C4.5, *naïve bayes* dan *neural network* yang diaplikasikan pada data kelulusan mahasiswa baik yang lulus tepat waktu dan yang terlambat lulus. Algoritma C4.5 merupakan salah satu algoritma dalam metode *decision tree* yang merubah data menjadi pohon keputusan menggunakan rumus perhitungan entropi. *Naïve bayes* yaitu metode yang menghitung probabilitas antara kemunculan data yang satu dengan data yang lainnya. Neural network yaitu metode layaknya neuron manusia untuk mencari jalur terbaik. Dari hasil pengujian dengan mengukur kinerja keempat metode tersebut menggunakan metode pengujian *cross validation*, *confusion matrix* dan kurva *ROC* diketahui bahwa *neural network* memiliki nilai akurasi yang paling tinggi yakni sebesar 78,19% , diikuti oleh metode algoritma C4.5 dengan nilai akurasi 74,33%, kemudian metode *naïve bayes* dengan nilai akurasi 69,72%. Nilai *AUC* untuk metode *neural network* menunjukkan nilai tertinggi yakni 0,848 dan yang terendah adalah metode *naïve bayes* yaitu 69,72.

Kata Kunci:

Data Mining, algoritma C4.5, *naïve bayes* dan *neural network*,

ABSTRACT

Name : Hilda Amalia
Nim : 14000320
Study of Program : *Master of Computer Science*
Level : Strata Dua (S2)
Consentration : *Management Information System*
Title : “Komparasi Metode Klasifikasi Data Mining untuk
Prediksi Kelulusan Mahasiswa”.

College is a place for students to gain knowledge before jumping into the competitive world of work. The number of students who graduate on time become an indicator of college success both public and private. Research in terms of predicting student graduation and many students have done. In this study comparison of methods of data mining algorithms, namely C4.5, Naive Bayes and neural networks are applied to the data of students passing either the graduate on time and a late pass. From the test results to measure the performance of four methods using cross validation test method, confusion matrix and ROC curve is known that neural network has the highest accuracy value which is equal to 78.19%, followed by the C4.5 algorithm method with accuracy the value of 74.33 %, then the naive Bayes method with a value of 69.72% accuracy. AUC values for the neural network method shows the highest value of 0.848, and the lowest is the Naive Bayes method is 69.72.

Keyword:

Data Mining, the algorithm C4.5, Naive Bayes and neural network,

DAFTAR ISI

	Halaman
HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
HALAMAN PERNYATAAN ORISIONALITAS	iii
HALAMAN PENGESAHAN	iv
KATA PENGANTAR	v
SURAT PESETUJUAN PUBLIKASI.....	vii
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABLE	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN.....	1
BAB 1. PENDAHULUAN	1
1.1. Latar Belakang	3
1.2. Rumusan Masalah	3
1.2.1 Pernyataan Masalah.....	3
1.2.2 Pertanyaan Penelitian	3
1.3. Tujuan Penelitian	3
1.4. Manfaat Penelitian	4
1.5. Ruang Lingkup.....	4
1.6. Sistematika Penulisan.....	4
BAB 2. LANDASAN TEORI	6
2.1. Tinjauan Studi.....	7
2.2. Tinjauan Pustaka	7
2.2.1. Kelulusan Mahasiswa	7
2.2.2. <i>Data Mining</i>	7
2.2.3. Metode Klasifikasi Data Mining	7
2.2.3.1. Algortima C4.5.....	14
2.2.3.2. <i>Neural Network</i>	20
2.2.3.3. <i>Naïve Bayes</i>	29
2.2.4. Pengujian <i>K-Fold Cross Validation</i>	32
2.2.5. Evaluasi dan Validasi Metode Klasifikasi	33
2.4. Kerangka Pemikiran.....	37
BAB 3. METODOLOGI PENELITIAN.....	38
3.1. Desain Penelitian.....	38
3.2. Pengumpulan Data	40
3.3. Pengolahan Data Awal	41
3.4. Metode yang Diusulkan.....	43
BAB 4. HASIL ANALISA DAN PEMBAHASAN.....	44

4.1. Hasil	48
4.1.1. Algoritma C4.5	48
4.1.2. <i>Naïve Bayes</i>	52
4.1.3. <i>Neural Network</i>	62
4.2. Evaluasi dan Validasi Hasil.....	67
4.2.1. Algoritma C4.5	68
4.2.2. <i>Naïve Bayes</i>	70
4.2.3. <i>Neural Network</i>	76
4.2.4. Analisa Hasil Komparasi	77
BAB 5. PENUTUP	64
5.1. Kesimpulan	64
5.2. Saran	64
DAFTAR REFERENSI	

DAFTAR TABEL

	Halaman
Tabel 2.1. Model <i>Confusion Matrix</i>	20
Tabel 2.2. Data Cuaca	22
Tabel 2.3. Perhitungan Entropi dan <i>Gain</i> Semua Atribut	24
Tabel 2.4. Perhitungan Entropi dan <i>Gain</i> untuk Atribut Cuaca Cerah	25
Tabel 2.5. Perhitungan Entropi dan <i>Gain</i> untuk Atribut Cuaca Hujan	25
Tabel 2.6. Insiliasi Nilai input dan Nilai Bias Awal	27
Tabel 2.7. Nilai Bias dan Bobot baru	29
Tabel 3.1. Tabel Data Mahasiswa	38
Tabel 3.2. Ilustrasi Missing Data Pada Data Training	39
Tabel 3.3. Data Training setelah Dilakukan Teknik Validasi	40
Tabel 3.4. Tabel Atribut yang Digunakan	41
Tabel 4.1. Nilai entropy dan gain untuk menentukan akar.....	44
Tabel 4.2. Tabel Nilai <i>Entropy</i> dan <i>Gain</i> untuk menentukan simpul 1.1	45
Tabel 4.3. Perhitungan Nilai Probabilitas <i>Prior</i>	50
Tabel 4.4. Atribut X yang akan diprediksi	51
Tabel 4.5. Nilai Bobot Akhir Untuk <i>Hidden Layer</i>	54
Tabel 4.6. Nilai Bobot Akhir Untuk <i>Output Layer</i>	55
Tabel 4.9 Tabel <i>Confusion Matrix</i> Metode Algoritma C4.5	60
Tabel 4.10 Tabel <i>Confusion Matrix</i> Metode <i>Naïve Bayes</i>	61
Tabel 4.11 Tabel <i>Confusion Matrik</i> Metode <i>Neural Network</i>	62
Tabel 4.12. Perbandingan <i>Performance</i> Metode	64

DAFTAR GAMBAR

Gambar 2.1.	Contoh <i>Decision Tree</i>	14
Gambar 2.2.	Akar Pohon Keputusan Data Cuaca.....	19
Gambar 2.3.	Pohon Keputusan Data Cuaca	20
Gambar 2.4.	Arsitektur <i>Neural Network</i>	21
Gambar 2.5.	Contoh <i>Neural Network Multilayer Feed-Forward</i>	26
Gambar 2.6.	Ilustrasi 10-Fold Cross Validation.....	33
Gambar 2.8	Kerangka Pemikiran.....	37
Gambar 3.1.	Tahapan Penelitian.....	36
Gambar 3.2.	Metode yang Diusulkan	43
Gambar 4.1.	Pohon Keputusan Data Kelulusan Mahasiswa	49
Gambar 4.2	Pengujian <i>K-Fold Cross Validation</i> Metode Algoritma C4.5	52
Gambar 4.3.	Pengujian <i>K-Fold Cross Validation</i> Metode <i>Naïve Bayes</i>	56
Gambar 4.4	Neural Net yang dihasilkan dari Metode <i>Neural Network</i>	57
Gambar 4.5.	Pengujian <i>K-Fold Cross Validation</i> Metode <i>Neural Network</i>	59
Gambar 4.5.	Grafik <i>Area Under Curve</i> Algoritma C4.5.....	60
Gambar 4.6.	Grafik <i>Area Under Curve Naïve Bayes</i>	61
Gambar 4.8.	Grafik <i>Area Under Curve Neural Network</i>	62
Gambar 4.9.	Design Model Komparasi Dengan <i>ROC's Curve</i>	63
Gambar 4.10.	Algoritma-algoritma dalam Modul <i>ROC's Curve</i>	63
Gambar 4.11.	Grafik Kompirasi <i>ROC Curve</i>	64

DAFTAR LAMPIRAN

	Halaman
Data Mahasiswa.....	67
Data Training.....	96

BAB I

PENDAHULUAN

1.1 Latar Belakang Penulisan

Mahasiswa merupakan elite masyarakat yang memiliki ciri intelektualitas yang lebih kompleks dibandingkan kelompok seusia mereka yang bukan mahasiswa, ataupun kelompok usia lain dibawah mereka. Ciri intelektualitas ini adalah kemampuan mahasiswa dalam menghadapi mahasiswa dalam menghadapi masalah dan mencari pemecahan masalahnya secara lebih sistematis (Azwar, 2004). Masa mahasiswa merupakan masa yang penuh tantangan, masa yang penting bagi pengembangan intelektual agar bisa bersaing dalam dunia kerja nantinya dan dalam masyarakat.

Perguruan tinggi merupakan satuan pendidikan yang menjadi terminal terakhir bagi seseorang yang berpeluang belajar setinggi-tingginya melalui jalur pendidikan sekolah (Nawawi & Martini, 1994). Perguruan tinggi saat ini dituntut untuk memiliki keunggulan bersaing dengan memanfaatkan semua sumber daya yang dimiliki. Selain sumber daya sarana, prasarana, dan manusia, sistem informasi adalah salah satu sumber daya yang dapat digunakan untuk meningkatkan keunggulan bersaing. Sistem informasi dapat digunakan untuk mendapatkan, mengolah dan menyebarkan informasi untuk menunjang kegiatan operasional sehari-hari sekaligus menunjang kegiatan pengambilan keputusan strategis.

Organisasi pendidikan tinggi berada dalam lingkungan kompetitif yang sangat tinggi dan bertujuan untuk mendapatkan keuntungan yang lebih kompetitif melalui kompetisi bisnis lainnya. Organisasi-organisasi ini harus meningkatkan kualitas layanan mereka dan memuaskan pelanggan mereka. Mereka menganggap mahasiswa dan dosen sebagai aset utama mereka dan mereka ingin terus meningkatkan indikator-indikator kunci mereka dengan menggunakan asset secara efektif dan efisien (Qudri

& Kalyankar, 2010). Dalam sistem pendidikan mahasiswa adalah asset penting bagi sebuah institusi pendidikan untuk itu perlu diperhatikan tingkat kelulusan mahasiswa tepat pada waktunya.

Administrator sekolah berebut untuk meningkatkan tingkat kelulusan untuk berbagai alasan. Pertama, adalah dalam misi masing-masing sekolah untuk mendidik siswa (yaitu menghasilkan lulusan) yang menjadi anggota produktif masyarakat dan berkontribusi terhadap kesejahteraan ekonomi bangsa. Selain itu, masing-masing sekolah tahu bahwa jumlah siswa yang putus diterjemahkan sebagai hilangnya pendapatan bagi lembaga (Karamouiz & Vrettos, 2008). Siswa atau mahasiswa yang lulus tidak tepat waktu akan mempengaruhi penilaian publik terhadap kredibilitas sekolah atau institusi pendidikan.

Wisuda, wisuda terutama tepat waktu merupakan isu kebijakan yang semakin penting. Tingkat kelulusan dianggap sebagai salah satu efektivitas kelembagaan (Qudri & Kalyankar, 2010). Sehingga memerhatikan jumlah kelulusan suatu perguruan tinggi menjadi hal penting. Penurunan tingkat kelulusan mahasiswa juga akan berpengaruh terhadap akreditasi perguruan tinggi tersebut. Tingkat penurunan kelulusan mahasiswa yang signifikan dan terus berkembang adalah sebuah masalah yang ada pada perguruan tinggi. Untuk itu perlu adanya pemantauan atau evaluasi terhadap kecenderungan mahasiswa lulus tepat waktu atau tidak.

Pemantauan kinerja melibatkan penilaian yang melayani peran penting dalam menyediakan informasi yang diarahkan untuk membantu siswa atau mahasiswa, guru atau dosen, administrator, dan pembuat kebijakan mengambil keputusan (Ogor, 2007). Dari penjelasan di atas, jelas bahwa memprediksi kelulusan adalah hal yang penting bagi institusi dan potensi besar bagi institusi untuk menentukan kebijaksanaan strategis bagi institusinya. Setelah identifikasi mahasiswa yang beresiko dengan mekanisme pendukung seperti orientasi, menasihati, *monitoring*, dan lain-lain dapat digunakan untuk meningkatkan kegigihan mahasiswa dan dengan demikian meningkatkan tingkat kelulusan. Tugas prediksi dapat dianggap sebagai menjadi dua

kelas yaitu “sukses” yakni mahasiswa yang lulus tepat waktu dan “gagal” bagi mahasiswa yang lulus terlambat.

Dalam hal pengolahan data siswa atau mahasiswa telah diselesaikan dengan beberapa metode yaitu menggunakan metode *neural network* (Karamouiz & Vrettos, 2008), *decision tree* (Qudri & kalyankar, 2010), *naïve bayes* (Suhartini & Ernastuti, 2010). Tetapi belum ada yang melakukan perbandingan metode-metode tersebut sehingga belum diketahui metode yang paling akurat.

Untuk itu dalam penelitian ini akan dilakukan perbandingan metode *Neural Network*, *Naïve Bayes*, dan Algoritma C4.5 sehingga diperoleh metode dengan akurasi prediksi kelulusan mahasiswa yang terbaik.

1.2 Rumusan Masalah

1.2.1. Pernyataan Masalah

Berdasarkan permasalahan diatas maka rumusan masalah penelitian ini adalah belum diketahui metode data mining yang paling akurat untuk prediksi ketepatan kelulusan mahasiswa.

1.2.2. Pertanyaan penelitian

Bagaimana akurasi dari ketiga metode *data mining neural network*, *naïve bayes*, dan Algoritma C4.5 dalam memprediksi tingkat kelulusan mahasiswa sehingga diperoleh metode yang paling akurat.

1.3 Tujuan Penelitian

Tujuan Penelitian ini adalah mengetahui metode klasifikasi data mining yang paling akurat prediksi ketepatan kelulusan mahasiswa. Metode-metode yang digunakan adalah algoritma C4.5, *Naïve Bayes* dan *Neural Network*.

1.4 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Memberikan kontribusi keilmuan pada penelitian bidang klasifikasi data mining khususnya untuk prediksi kelulusan mahasiswa.
2. Membantu administrasi perguruan tinggi untuk memberikan peringatan dini dan pembimbingan awal bagi mahasiswa yang kemungkinan tidak lulus tepat waktu.
3. Membantu perguruan tinggi dalam membuat kebijakan untuk bisa meningkatkan kelulusan mahasiswa.

1.5 Ruang Lingkup

Ruang Lingkup penelitian ini terbatas pada penggunaan metode *neural network*, *naïve bayes*, serta algoritma C4.5 dalam memprediksi ketepatan lulus mahasiswa, dan melakukan perbandingan akurasi ketiga metode tersebut, sehingga diperoleh metode dengan akurasi yang tinggi dengan hasil “tepat” dan “terlambat”. Parameter yang diuji adalah fakultas, usia, jenis kelamin, IP (Indeks Prestasi) semester satu sampai IP semester empat.

1.6 Sistematika Penulisan

Bab I Pendahuluan

Pada Bab ini berisi latar belakang penulisan, permasalahan kelulusan mahasiswa, pemecahan masalah serta tujuan dari penelitian ini.

Bab II Landasan Teori

Pada bab ini berisi teori-teori mengenai mahasiswa, perguruan tinggi, data mining, algoritma C4.5, *naïve bayes*, *neural network* serta kerangka pemikiran dari penelitian, dan penjelasan mengenai objek penelitian.

Bab III Metodologi Penelitian

Pada bab ini berisi mengenai tahapan penelitian yang dilakukan, design penelitian dengan penggambaran analisa komparasi metode data mining yakni algoritma C4.5, *naïve bayes* dan *neural network*.

Bab IV Hasil dan Pembahasan

Pada bab ini berisi mengenai pembahasan, tata cara dan langkah-langkah algoritma data mining yaitu langkah-langkah pembentukan pohon keputusan dengan algoritma C4.5, langkah-langkah perhitungan metode *naïve bayes* dan hasil nilai dari *neural network*. Serta berisikan hasil akhir penelitian yaitu akurasi masing-masing metode *data mining*.

Bab V Penutup

Pada bab ini berisi kesimpulan dari penulisan pada bab-bab sebelumnya dan saran mengenai penelitian selanjutnya.

BAB II

LANDASAN TEORI

2.1. Tinjauan Studi

Literatur mengenai pembahasan prediksi kelulusan mahasiswa telah banyak dilakukan dengan beberapa metode. Berikut dijabarkan beberapa metode yang pernah digunakan untuk menyelesaikan masalah kelulusan mahasiswa.

Penelitian yang dilakukan oleh Karamouiz dan Vretoz pada tahun 2009 dengan judul *Sentivity Analysis of Neural Network for Identifying the Factors for Collage Students Success*. Permasalahan pada penelitian ini adalah tingkat kelulusan dianggap sebagai indikator efektivitas suatu lembaga institusi, metode yang digunakan adalah NN (*Neural Network*). Dari hasil data *training* yang dilakukan diperoleh presentasi hasil yang benar untuk kategori yang lulus adalah 86.04% dan *training* data untuk kategori yang tidak sukses yakni 68.21%, dan *error* yang diperoleh untuk kedua kategori tersebut adalah 0.18% (Karamouiz & Vrentoz, 2009).

Penelitian yang dilakukan oleh Qudri dan Kalanyar pada tahun 2010 dengan judul *Drop Out Feature of Student Data for Academic Performance Using Decision Tree techniques*. Masalah dalam penelitian ini adalah prestasi akademik siswa sangat penting bagi lembaga pendidikan karena program-program strategis dapat direncanakan untuk meningkatkan atau mempertahankan siswa kinerja selama periode mereka studi di lembaga. Kinerja akademik dalam penelitian ini diukur oleh indek kumulatif rata-rata setelah lulus. Hal ini penting untuk memahami faktor-faktor penentu tingkat penyelesaian yang sukses dan tepat waktu. Metode yang digunakan adalah *Decision Tree*, yakni algoritma J4.8. Hasil penelitian ini adalah sebuah pohon keputusan yang dapat dijadikan *rule* bagi prediksi siswa yang putus sekolah (Qudri & Kaylanyar, 2010).

Penelitian yang dilakukan oleh Suhartina dan Ernastuti pada tahun 2010 dengan judul *Graduation Prediction of Gunadarma University Students Using Algorithm and Naive Bayes C4.5 Algoritmh*. Masalah dalam penelitian ini adalah

banyaknya mahasiswa yang tidak lulus sesuai dengan waktu studi yang telah ditetapkan. Untuk mengetahui tingkat kelulusan mahasiswa dalam satu tahun ajaran dapat dilakukan suatu prediksi berdasarkan data-data mahasiswa pada tingkat atau tahun ajaran pertama. Algoritma yang digunakan adalah C45 dan *naïve bayes*. Hasil dari penelitian ini adalah akurasi dengan dua metode tersebut yakni akurasi untuk metode *naïve bayes* adalah 80,85% dengan presentasi kesalahan 19,05% Akurasi ketepatan hasil prediksi C4.5 85.7%, dan presentasi kesalahannya adalah 14,3% (Suhartina & Ernastuti, 2010).

Jadi dalam penelitian ini akan dilakukan komparasi metode klasifikasi data mining ketiga metode diatas yaitu algoritma C4.5, *naïve bayes* dan *neural network* untuk mengetahui metode klasifikasi yang paling akurat dalam prediksi ketepatan kelulusan mahasiswa.

2.2. Tinjauan Pustaka

Dalam penelitian ini menggunakan tinjauan pustaka didasarkan pada beberapa buku yang mendukung teoritis dari penelitian ini. Diantaranya:

2.2.1. Kelulusan Mahasiswa

Mahasiswa merupakan elite masyarakat yang mempunyai ciri intelektualitas yang lebih komplek dibandingkan kelompok seusia mereka yang bukan mahasiswa, ataupun kelompok usia dibawah dan diatas mereka. Ciri intelektualitas tersebut adalah kemampuan mahasiswa menghadapi, memahami dan mencari cara pemecahan masalah secara lebih sistematis (Azwar, 2004). Dalam setiap fakultas ataupun jurusan pada suatu universitas jarang sekali bahkan tidak pernah terjadi dimana jumlah mahasiswa yang diterima akan mengakhiri masa perkuliahannya pada waktu bersamaan dengan jumlah yang sama pula pada saat diterima di fakultas tersebut (Siregar, 2006).

Kelulusan mahasiwa adalah hal yang penting diperhatikan, karena penurunan jumlah kelulusan akan menghilangkan jumlah pendapatan institusi dan mempengaruhi penilaian pemerintah serta memperngaruhi status akreditasi institusi (Karamouiz & Vrettos, 2008). Beberapa faktor dapat mempengaruhi