PENERAPAN DATA MINING MENGGUNAKAN METODE DECISION TREE C4.5 UNTUK PREDIKSI TINGKAT KELULUSAN MAHASISWA (Studi Kasus : STMIK WIT)

TESIS

Disusun sebagai salah satu syarat untuk
Memperoleh gelar Magister Komputer
dari Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI

Oleh :

Niken Adityas Lestari

NPM : 2016210015

PROGRAM STUDI PASCASARJANA
MAGISTER SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI
BANDUNG

2017

PENERAPAN DATA MINING MENGGUNAKAN METODE DECISION TREE C4.5 UNTUK PREDIKSI TINGKAT KELULUSAN MAHASISWA (Studi Kasus : STMIK WIT)

Oleh :

Niken Adityas Lestari

2016210015

Bandung, 9 Desember 2017

Menyetujui,

Dr. Djajasukma Tjahjadi, S.E., M.T.
Pembimbing

PROGRAM STUDI PASCASARJANA
MAGISTER SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI
BANDUNG
2017

Dipersembahkan Untuk Keluarga Tercinta :
Suamiku : Hernu Krisnandika, ST.
Anakku: Adzkia Hasnah Nabila
OrangTua, Mertua, dan adik -adikku

ABSTRAK

PENERAPAN DATA MINING MENGGUNAKAN METODE DECISION TREE C4.5 UNTUK PREDIKSI TINGKAT KELULUSAN MAHASISWA (Studi Kasus : STMIK WIT)

Niken Adityas Lestari
NPM : 2016210015

Setiap perguruan tinggi berusaha untuk terus memperbaiki manajemennya untuk meningkatkan mutu pendidikan dan meningkatkan akreditasi. Salah satu aspek penilaian akreditasi adalah mahasiswa dan lulusan. Tingkat kelulusan dan jumlah mahasiswa akan berpengaruh dalam proses akreditasi yang dilakukan oleh pemerintah. Database perguruan tinggi menyimpan berbagai data, dari data tersebut apabila digali dengan tepat maka dapat diketahui pola atau pengetahuan baru yang dapat dikembangkan untuk diterapkan pada perguruan tinggi diantaranya mengenai potensi mahasiswa lulus tepat waktu, lulus terlambat, dan drop out.

Penelitian ini menggunakan metode *Knowledge Discovery in Database* (KDD) untuk menganalisis data dalam penerapan data mining, mengekstrak pengetahuan apa yang dianggap sesuai dengan spesifikasi ukuran dan batas, menggunakan database bersama dengan *preprocessing* yang diperlukan, pengambilan sampel dan transformasi dari database. Teknik data mining dapat melakukan prediksi kelulusan tepat waktu, terlambat dan *drop out*, yaitu menggunakan *decision tree* atau pohon keputusan dengan algoritma C4.5 berdasarkan atribut IPK4, status pekerjaan, program studi, asal sekolah, jenis kelamin, gaji orangtua dan asal daerah. Algoritma C4.5 dapat mengkonstruksi pohon keputusan yang mampu mengatasi atribut bertipe kontinu, mengatasi nilai yang hilang dan dapat melakukan pemangkasan pohon yang kompleks. Data yang digunakan dalam proses prediksi adalah data mahasiswa tahun 2006 sampai dengan 2010, dan data yang digunakan untuk menguji pola prediksi adalah data mahasiswa tahun 2012.

Hasil algoritma C4.5 berupa pohon keputusan menunjukkan atribut IPK4 sebagai simpul akar, kemudian IPK 4 akan diuji dengan membuat model kedua yaitu *modelling* algoritma C4.5 berdasarkan IP semester 1,2 3 dan 4. Tujuan dari model kedua ini adalah untuk menemukan model terbaik dan menegaskan pola prediksi pada model pertama, apakah sudah tepat prediksi kelulusan dilakukan setelah mahasiswa menempuh perkuliahan sampai dengan empat semester. Setelah dilakukan pengujian, pohon keputusan model kedua menunjukkan IP semester 4 sebagai akar pohon keputusan. Artinya pola atau aturan pada model pertama sudah tepat untuk digunakan dalam memprediksi kelulusan. Evaluasi prediksi kelulusan menggunakan *confusion matrix,* model pertama memiliki nilai akurasi sebesar 84% sedangkan model kedua 77%. Proses *modelling* algoritma C4.5, pengujian dan evaluasi dibantu dengan tools Weka 3.9.

Kata Kunci : Data mining, *Knowledge discovery in database* (KDD), *decision tree*, algoritma C4.5, prediksi, lulus tepat waktu, lulus terlambat, *drop out*, *confusion matrix*, Weka 3.9.

ABSTRACT

APPLICATION OF DATA MINING USING METHOD DECISION TREE C4.5 FOR PREDICTION STUDENT GENERAL LEVELS (Case Study: STMIK WIT)

Niken Adityas Lestari
NPM : 2016210015

Each university strives to continuously improve its management to improve the quality of education and improve accreditation. One aspect of the accreditation assessment is students and graduates. Graduation rate and number of students will occur in the process of accreditation conducted by the government. The college database keeps a variety of data, from the data explored properly, it can be known that new patterns and knowledge can be developed to apply to universities about the potential of graduates on time, late passes, and drop outs.

This study uses Knowledge Discovery in Database (KDD) method to analyze data in the application of data mining, extracting what knowledge is considered in accordance with the size and limit specification, using the database together with the necessary preprocessing, sampling and transformation of the database. Data mining techniques can make timely, late and drop out predictions using decision tree or decision tree with C4.5 algorithm based on IPK4 attributes, job status, study program, origin of school, gender, parental salary and local origin. The C4.5 algorithm can construct a decision tree capable of overcoming continuous type attributes, overcoming missing values and can perform complex tree pruning. The data used in the prediction process is student data from 2006 to 2010, and the data used to test prediction patterns is student data of 2012.

The result of the C4.5 algorithm in the form of a decision tree shows the IPK4 attribute as the root node, then GPA 4 will be tested by creating a second model of modeling algorithm C4.5 based on IP semesters 1,2 3 and 4. The purpose of this second model is to find the best model and confirms the prediction pattern on the first model, whether it is appropriate predictions of graduation is done after the students take courses up to four semesters. After testing, the second model decision tree shows IP semester 4 as the root of the decision tree. This means that the pattern or rule in the first model is appropriate for use in predicting graduation. Evaluation of graduation prediction using confusion matrix, the first model has an accuracy value of 84% while the second model is 77%. The process modeling algorithm C4.5, testing and evaluation assisted with tools Weka 3.9.

Keywords: Data mining, decision tree, C4.5 algorithm, prediction, pass on time, late pass, drop out, confusion matrix, Weka 3.9.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Kuasa atas berkah, rahmat, dan karunia Nya, akhirnya penulis dapat menyelesaikan Tesis ini. Ucapan terima kasih yang setulusnya, disampaikan kepada semua pihak yang telah memberikan dukungan selama penyelesaian tesis ini, antara lain :

1. Bapak Dr. Djajasukma Tjahjadi, S.E., M.T. yang telah memberikan arahan dan bimbingan selama pembuatan tesis ini sehingga dapat diselesaikan dengan baik
2. Ibu Ketua STMIK WIT yang telah memberi dukungan moril maupun materil untuk penyelesaian Tesis ini
3. Staf, dan seluruh pengajar S-2 STMIK LIKMI, yang telah memberikan bekal pengetahuan hingga mampu menyelesaikan Tesis ini.
4. Seluruh staf STMIK WIT yang senantiasa memberi dukungan dan semangat dalam penyususnan tesis ini.
5. Rekan-rekan sejawat, yang telah memberikan kritik dan saran hingga Tesis ini dapat kami selesaikan.

Tesis ini disusun dengan menggunakan teori-teori yang diperoleh selama perkuliahan di STMIK LIKMI, dan data diperoleh dari tempat penulis bekerja yaitu STMIK WIT Cirebon. Oleh karena itu penulis berharap Tesis ini tidak hanya untuk memenuhi persyaratan kelulusan studi S-2 penulis di STMIK LIKMI, tetapi juga bermanfaat bagi institusi tempat penulis bekerja.

Penulis menyadari Tesis ini masih banyak kekurangan dan kelemahan. Oleh karena itu segala saran dan kritik konstruktif, untuk penyempurnaan Tesis ini sangat penulis harapkan.

Bandung, Desember 2017

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR RUMUS

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1 Latar Belakang

Beberapa tahun terakhir, data semakin heterogen dan kompleks dengan volume yang meningkat cepat secara eksponensial. Oleh karena itu, saat ini dikenal istilah *big data,* yang menggambarkan volume data yang sangat besar, terstruktur maupun tidak terstruktur, yang membanjiri dunia bisnis. *Big data* dapat dianalisis, sehingga organisasi atau perusahaan dapat mengambil keputusan strategis dengan lebih baik.

Dalam *big data,* tentu saja akan kesulitan dalam membaca dan mengetahui pola-pola dan relasi-relasi data jika dilakukan secara manual. Untuk itu dibutuhkan suatu teknik yang relatif cepat dan mudah untuk menemukan pengetahuan, pola dan relasi antar data secara otomatis yang disebut dengan Data Mining atau penambangan data.

"Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam database, data warehouse atau penyimpanan lainnya". (Meilani dan Susanti, 2014: 1-2)

Dalam prosesnya data mining akan mengekstrak informasi yang berharga dengan cara menganalisis adanya pola-pola ataupun hubungan keterkaitan tertentu dari data-data yang berukuran besar. Teknik data mining secara garis besar dapat dibagi dalam dua kelompok yaitu verifikasi dan discovery. Metode verifikasi umumnya meliputi teknik-teknik statistik seperti goodness of fit, dan analisis variansi. Metode *discovery* lebih lanjut dapat dibagi atas model prediktif dan model deskriptif. Teknik prediktif melakukan prediksi terhadap data dengan menggunakan hasil-hasil yang telah diketahui dari data yang berbeda. Model ini dapat dibuat berdasarkan penggunaan data historis lain. Sementara itu, model deskriptif bertujuan mengidentifikasi pola-pola atau hubungan antar data dan memberikan cara untuk mengeksplorasi karakteristik data yang diselidiki.

Bagian yang sangat penting dalam data mining adalah teknik klasifikasi, yaitu bagaimana mempelajari sekumpulan data sehingga dihasilkan aturan yang bisa mengklasifikasi atau mengenali data baru yang belum pernah dipelajari. Klasifikasi dapat

didefinisikan sebagai proses untuk menyatakan suatu objek data sebagai salah satu kategori (kelas) yang telah didefinisikan sebelumnya.

Decision tree adalah salah satu metode klasifikasi yang popular dan banyak digunakan secara praktis. Metode ini berusaha menemukan model klasifikasi yang tahan terhadap derau. Metode *decision tree* mengubah fakta yang sangat besar menjadi pohon keputusan yang memprediksikan aturan. Salah satu algoritma yang digunakan untuk membentuk pohon keputusan (*decision tree)* adalah C4.5.

Algoritma data mining C4.5 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokan dan bersifat prediktif. Dasar algoritma C4.5 adalah pembentukan pohon keputusan (*decision tree*). Cabang-cabang pohon keputusan merupakan pertanyaan klasifikasi dan daun-daunnya merupakan kelas-kelas atau segmen-segmennya.

Pada proses pengolahan data mining akan digunakan beberapa atribut sebagai parameter dalam pengklasifikasian data *training*. Atribut atau variabel menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon keputusan.

Untuk memprediksi kelulusan mahasiswa, faktor nilai dan ekonomi merupakan faktor utama dalam penentuan atribut, dari segi nilai digunakan atribuk IPK sedangkan dari segi ekonomi digunakan atribut penghasilan orangtua dan status pekerjaan mahasiswa. Setelah menentukan atribut yang paling utama dan penting, selanjutnya memilih atribut diluar faktor nilai dan ekonomi yaitu program studi, jenis kelamin, asal sekolah, asal daerah dan kelas.

Perguruan tinggi saat ini berada dalam lingkungan yang sangat kompetitif. Setiap perguruan tinggi berusaha untuk terus memperbaiki manajemennya untuk meningkatkan mutu pendidikan dan meningkatkan akreditasi. Salah satu aspek penilaian akreditasi adalah mahasiswa dan lulusan.

Berdasarkan hasil pengamatan STMIK WIT, ditemukan sebaran yang tak seimbang anatara jumlah mahasiswa yang masuk dan keluar karena telah menyelesaikan masa studi. Mahasiswa masuk dalam jumlah besar, namun jumlah mahasiswa yang lulus tepat

waktu jauh lebih kecil dibandingkan dengan jumlah mahasiswa yang masuk ke STMIK WIT.

Selama ini STMIK WIT belum memiliki pola-pola prediksi masa studi mahasiswa, lamanya waktu kelulusan mahasiswa tidak selalu dapat diprediksi secara dini oleh pihak mahasiswa maupun perguruan tinggi sehingga dapat berakibat pada waktu lulus yang terlambat dan merugikan kedua belah pihak.

Pengetahuan ini dapat digunakan dalam membantu pihak perguruan tinggi untuk lebih mengenal situasi para mahasiswanya dan dapat dijadikan sebagai pengetahuan dini dalam proses pengambilan keputusan untuk tindakan preventif dalam hal mengantisipasi mahasiswa yang berpeluang lulus terlambat dan *drop out*, untuk meningkatkan prestasi mahasiswa, meningkatkan proses kegiatan belajar dan mengajar dan banyak lagi keuntungan lain yang bisa diperoleh dari hasil penambangan data tersebut.

Berdasarkan uraian diatas, judul yang diambil dalam penelitian ini adalah "Penerapan Data Mining Menggunakan Metode *Decision Tree* C4.5 Untuk Prediksi Tingkat Kelulusan Mahasiswa"**.**

1.2 Rumusan Masalah

Dengan mengacu pada latar belakang masalah diatas, maka permasalahan yang dibahas dan diteliti adalah atribut apa saja yang dapat memprediksi kelulusan tepat waktu, terlambat dan *drop out* dengan menggunakan metode *decision tree* C4.5?

1.3 Tujuan Penelitian

Dari rumusan masalah diatas, tujuan yang akan dicapai dari tesis ini adalah:

Menggali pengetahuan tentang pola lulus tepat waktu, terlambat dan *drop out* dengan memperhatikan beberapa kondisi yaitu :

1. Hubungan kelulusan dengan program studi
2. Hubungan kelulusan dengan jenis kelamin
3. Hubungan kelulusan dengan asal daerah
4. Hubungan kelulusan dengan asal sekolah
5. Hubungan kelulusan dengan IPK semester 4

6. Hubungan kelulusan dengan status pekerjaan mahasiswa
7. Hubungan kelulusan dengan penghasilan orangtua
8. Hubungan kelulusan dengan kelas

1.4 Ruang Lingkup

Adapun batasan masalah yang akan dibahas adalah :

1. Data yang digunakan adalah data induk mahasiswa dan data akademik mahasiswa
2. Data mahasiswa yang digunakan dalam proses pembentukan *decision tree* atau pohon keputusan adalah data mahasiswa dari beberapa angkatan yaitu 2006 s.d 2010
3. Data mahasiswa yang digunakan dalam proses pengujian *decision tree* atau pohon keputusan yang telah terbentuk adalah data mahasiswa tahun 2011
4. Atribut yang akan diuji adalah program studi, jenis kelamin, asal daerah, asal sekolah, IPK semester 4, satus pekerjaan, gaji atau penghasilan orangtua, kelas dan status kelulusan
5. Target yang ingin dicapai dalam prediksi kelulusan ini adalah lulus tepat waktu, lulus terlambat dan *drop out*.

1.5 Metode Penelitian

Metode penelitian yang digunakan untuk mengolah data mining ini adalah metode *Knowledge Discovery in Database* (KDD). Proses KDD adalah proses menggunakan metode *data mining* untuk mengekstrak pengetahuan apa yang dianggap sesuai dengan spesifikasi ukuran dan batas, menggunakan database bersama dengan *preprocessing* yang diperlukan, pengambilan sampel dan transformasi dari database. Tahapan dari penelitian ini adalah seleksi data, integrasi data, *cleaning* data, transformasi data, mengolah data mining, evaluasi dan menerapkan pengetahuan.

1.6 Sistematika Penulisan

Untuk memberikan gambaran yang singkat mengenai pembahasan tesis, maka tesis ini dibagi menjadi 5 bab yang saling berhubungan. Adapun sistematika penulisan adalah sebagai berikut :

BAB I : PENDAHULUAN

Bab ini menjelaskan tentang latar belakang, identifikasi masalah, tujuan penelitian, ruang lingkup, metode penelitian dan sistematika penulisan.

BAB II : TINJAUAN PUSTAKA

Bab ini menguraikan dasar teori yang berkaitan dengan data mining dengan metode clustering.

BAB III : OBJEK DAN METODOLOGI PENELITIAN

Bab ini menjelaskan hasil analisis objek penelitian dan metodologi penelitian.

BAB IV : HASIL DAN PEMBAHASAN MASALAH

Bab ini menguraikan penyajian data penelitian, pengolahan terhadap data yang terkumpul dan hasil penelitian yang dicapai.

BAB V : KESIMPULAN DAN SARAN

Bab ini menjelaskan kesimpulan yang diperoleh dari hasil penelitian dan saran sebagai pemecahan masalah dan pencapaian yang lebih baik

BAB II

TINJAUAN PUSTAKA

2.1 Data Mining

Istilah data mining memiliki beberapa padanan, seperti knowledge discovery ataupun penemuan pengetahuan tepat digunakan karena tujuan utama dari data mining memang untuk mendapatkan pengetahuan yang masih tersembunyi di dalam bongkahan data. (Susanto dan Suryadi, 2010:2).

Data mining bukanlah suatu bidang yang baru. Dalam aplikasinya, *data mining* sebenarnya merupakan bagian dari proses *Knowledge Discovery in Database* (KDD), bukan sebagai teknologi yang utuh dan berdiri sendiri. Menurut (Al Fatta, 2007:13) menjelaskan bahwa :

Data mining adalah proses yang menggunakan teknik statistik, perhitungan, kecerdasan buatan dan machine learning untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar.

Secara sederhana, data mining dapat diartikan sebagai proses mengekstrak atau "menggali" pengetahuan yang ada pada sekumpulan data. Menurut (Brijesh Kumar Bhardwaj, 2010;1) menjelaskan bahwa :

Data Mining can be used in educational field to enhance our understanding of learning process to focus on identifying,extracting and evaluating variables related to the learning process of students as described.

Banyak orang yang setuju bahwa data mining adalah sinonim dari *Knowledge-Discovery in Database* atau yang biasa disebut KDD. Dari sudut pandang yang lain, data mining dianggap sebagai satu langkah yang penting didalam proses KDD

"Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam database, data warehouse, atau penyimpanan informasi lainnya". (Meilani dan Susanti, 2014:1-2)

Salah satu kesulitan untuk mendefinisikan data mining adalah kenyataan bahwa data mining mewarisi banyak aspek dan teknik dari bidang – bidang ilmu yang sudah mapan terlebih dahulu. Berawal dari beberapa disiplin ilmu, data mining bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani :

1. Jumlah data yang sangat besar
2. Dimensi data yang tinggi
3. Data yang heterogen dan berbeda sifat

Menurut (Kadir, 2013:240) menjelaskan bahwa :

Data mining atau penambangan data adalah perangkat lunak yang digunakan untuk menemukan pola-pola tersembunyi, tren, maupun aturan-aturan yang terdapat dalam basis berukuran besar dan mengahsilkan aturan-aturan yang digunakan utnuk memperkirakan perilaku di masa mendatang.

Ada beberapa macam pendekatan yang berbeda yang diklasifikasikan sebagai teknik pencarian informasi/pengetahuan dalam KDD. Ada pendekatan kuantitatif, seperti pendekatan probabilistik seperti logika induktif, pencarian pola, dan analisis pohon keputusan. Pendekatan yang lain meliputi deviasi, analisis kecenderungan, algoritma genetik, jaringan saraf tiruan, dan pendekatan campuran dua atau lebih dari beberapa pendekatan yang ada. Pada dasarnya ada enam elemen yang paling esensial dalam teknik pencarian informasi/pengetahuan dalam KDD yaitu:

1. Mengerjakan sejumlah besar data.
2. Diperlukan efesiensi berkaitan dengan volume data.
3. Mengutamakan ketetapan/keakuratan.
4. Membutuhkan pemakaian bahasa tingkat tinggi.
5. Menggunakan beberapa bentuk dari pembelajaran otomatis.
6. Menghasilkan hasil yang menarik.

"*Kegunaan data mining dapat dibagi menjadi dua: deskriptif dan prediktif. Deskriptif berarti data mining digunakan untuk mencari pola-pola yang dapat dipahami manusia dan menjelaskan karakteristik data. Sedangkan prediktif berarti data mining digunakan untuk mebentuk sebuah model pengetahuan yang akan digunakan untuk dilakukan prediksi*". (Suyanto, 2017:3)

Berdasarkan defenisi-defenisi yang telah disampaikan, hal penting terkait dengan Data Mining adalah:

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang besar.
3. Tujuan data mining adalah mendapatkan pola atau informasi yang akan mungkin memberikan indikasi yang bermanfaat di masa mendatang.

2.1.1 Pengelompokkan Data Mining

Menurut (Larose, 2005;11) *data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat di lakukan, yaitu :

1. Deskripsi

   Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecendrungan yang terdapat dalam data.

2. Estimasi

   Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

3. Prediksi

   Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

   Dalam klasifikasi, terdapat target variabel kategori. Sebagi contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, pendapatan rendah. Contoh lain klasifikasi adalah :

   a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau bukan.
   b. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
   c. Mendiagnosa penyakit seorang pasien termasuk kategori penyakit apa.

5. Pengklusteran

   Clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteriktik *(similarity)* antara satu data dengan data yang lain.

*Cluster*ing merupakan salah satu metode *data mining* yang bersifat tanpa arahan *(unsupervised)*. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, estimasi atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan, yang mana kemiripan dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok yang lain akan bernilai minimal.

Contoh pengklusteran dalam bisnis dan penelitian adalah :

a. Mendapatkan kelompok–kelompok konsumen untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
b. Untuk tujuan audita akuntansi, yaitu melakukan pemisahan terhadap perilaku financial dalam baik dan mencurigakan.
c. Melakukan pengklusteran dalam ekspresi dan gen, untuk mendapatkan kemiripan dari perilaku dari gen dalam jumlah besar.

6. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

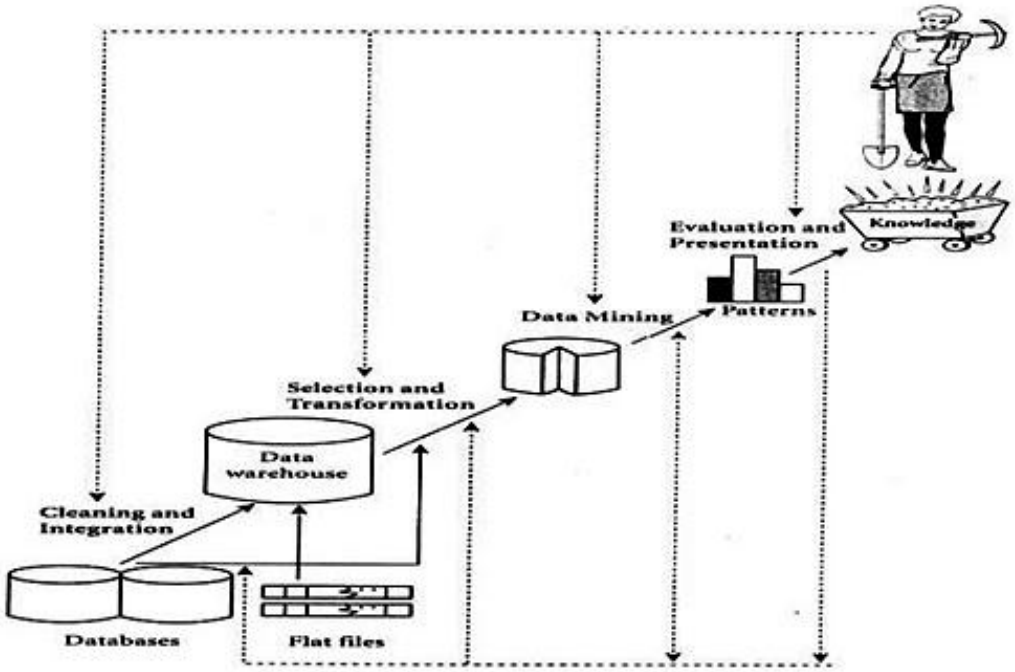
Contoh asosiasi dalam bisnis dan penelitian adalah :

a. Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respon positif terhadap penawaran upgrade layanan yang diberikan
b. Menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli secara bersamaan.

2.1.2 Metode *Knowledge Discovery in Database* (KDD)

Metode untuk menganalisis data dalam penerapan data mining ini adalah dengan menggunakan tahapan *Knowledge Discovery in Database* (KDD). Proses KDD adalah proses menggunakan metode *data mining* untuk mengekstrak pengetahuan apa yang dianggap sesuai dengan spesifikasi ukuran dan batas, menggunakan database bersama

dengan *preprocessing* yang diperlukan, pengambilan sampel dan transformasi dari database. Tahapan *Knowledge Discovery in Database* (KDD) terdiri dari beberapa tahapan. Tahapan *Knowledge Discovery in Database* (KDD) yang ditunjukkan pada Gambar 2.1.



Gambar 2.1.
Tahap penemuan *Knowledge* pada Data Mining (KDD) (Han & Kamber, 2006:6)

1. *Data cleaning*

   Untuk menghilangkan data *noise* (data yang tidak relevan/berhubungan langsung dengan tujuan akhir proses *datamining*, misal: *data mining* yang bertujuan untuk menganalisa hasil penjualan, maka data-data dalam kumpulan seperti ”nama pegawai”, ”umur”, dan sebagainya dapat di-*ignore*) dan tidak konsisten.

2. *Data integration*

   Untuk menggabungkan *multiple data source.*

3. *Data selection*

   Untuk mengambil sebuah data yang sesuai untuk keperluan analisa.

4. *Data transformation*

Untuk mentransformasikan data ke dalam bentuk yang lebih sesuai untuk di *mining. Data mining* Proses terpenting dimana metode tertentu diterapkan untuk menghasilkan *data pattern*.

5. *Pattern evaluation*

Untuk mengidentifikasi apakah benar interenting patterns yang didapatkan sudah cukup mewakili *knowledge* berdasarkan perhitungan tertentu.

6. Knowledge presentation

Untuk mempresentasikan *knowledge* yang sudah didapat dari *user.*

2.2 *Decision tree*

Decision tree atau pohon keputususan merupakan metode klasifikasi yang paling popular digunakan. Selain karena pembangunannya yang relatif cepat, hasil dari model yang dibangun mudah untuk dipahami.

Menurut (kusrini dan emha, 2009:13) menjelaskan bahwa :

Pohon keputusan atau decision tree merupakan metode klasifikasi dan prediksi yang sangat kuat dan tekenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat di ekspresikan dalam bentuk bahasa basis data seperti Structured Query Language untuk mencari record pada kategori tertentu.

Decision tree juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara jumlah calon variabel input dengan sebuah variabel target. Karena *decision tree* memadukan eksplorasi data dan pemodelan, dia sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain.

Pada *decision tree* terdapat 3 jenis *node*, yaitu:

1. *Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
2. *Internal Node*, merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
3. *Leaf node atau terminal node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.

Sebuah model pohon keputusan terdiri dari sekumpulan aturan untuk membagi sejumlah populasi yang heterogen menjadi lebih kecil, lebih homogeny dengan memperhatikan variabel tujuannya.

Variabel tujuan biasanya dikelompokkan dengan pasti dan model *decision tree* lebih mengarah pada perhitungan probabilitas dari tiap-tiap *record* terhadap kategori-kategori tersebut atau untuk mengklasifikasi *record* dengan mengekompokkannya dalam satu kelas.

Banyak algortima yang bisa digunakan dalam pembentukan *Decision tree* atau pohon keputusan, antara lain ID3, CART, dan C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID.

2.2.1 *Entropy*

Secara istilah entropy adalah keberbedaan atau keberagaman. Dalam data mining, entropy didefinisikan sebagai suatu parameter untuk mengukur heterogenitas (keberagaman) dalam suatu himpunan data. Semakin heterogen himpunan data, semakin besar pula nilai entropynya.(Suyanto, 2017:134)

$$Entropy(S) = \sum_{i}^{c} -p_i \log_2 p_i \quad \text{..... (1)}$$

c = jml nilai yang ada pada atribut target (jml kelas klasifikasi).

pi = jumlah proporsi sampel (peluang) untuk kelas i.

2.2.2 *Information Gain*

Secara istilah, information gain adalah perolehan informasi. Dalam data mining, information gain didefinisikan sebagai ukuran efektivitas suatu atribut dalam mengklasifikasikan data. .(Suyanto, 2017:136)

$$Gain\ (S, A) = \ entropy\ (S) - \sum_{i=1}^{n} \frac{|Si|}{S} * Entropy\ (Si) \quad \text{..... (2)}$$

Keterangan :

S = Himpunan Kasus

A = Fitur

n = jumlah partisi attribute A

|Si| = Proporsi Si terhadap S

|S| = jumlah kasus dalam S

2.3 Algoritma C4.5

Salah satu algoritma yang dapat digunakan untuk membuat pohon keputusan (*decision tree)* adalah algoritma C4.5.

C4.5 is a well-known algorithm used to generate a decision trees. It is an extension of the ID3 algorithm used to overcome its disadvantages. The decision trees generated by the C4.5 algorithm can be used for classification, and for this reason, C4.5 is also referred to as a statistical classifier. (Kalpesh Adhatrao, 2013;42)

Algoritma C4.5 merupakan algoritma yang sangat populer yang digunakan oleh banyak peneliti di dunia, hal ini dijelaskan oleh Xindong Wu dan Vipin Kumar dalam bukunya yang berjudul The Top Ten Algorithms in Data Mining. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang diciptakan oleh J. Rose Quinlan.

Menurut (kusrini dan emha, 2009:15) menjelaskan bahwa :

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

1. *Pilihlah atribut sebagai akar*
2. *Buat cabang untuk tiap-tiap nilai*
3. *Bagi kasus dalam cabang*
4. *Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama*

Menurut (Ade Putra, 2017:179) menjelaskan bahwa :

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. *Mempersiapkan data training. Data training merupakan data yang pernah terjadi sebelumnya atau disebut data masa lalu serta telah mengalami pengelompokan ke dalam kelas tertentu.*
2. *Menghitung akar dari pohon Akar akan diambil dari attribute yang akan terpilih dengan cara menghitung nilai gain dari setiap attribute kemudian nilai gain attribute yang tertinggi akan menjadi akar pertama atau node pertama. Sebelum menghitung nilai gain dari attribute hitung dahulu nilai entropy. Untuk menghitung nilai entropy digunakan persamaan 1 dibawah ini :*

$$Entropy\ (S) = \sum_{i=1}^{n} - pi\ \log_2 pi \qquad \text{..... (1)}$$

Keterangan :
S= Himpunan kasus
n = jumlah partisi S
Pi = proporsi Si terhadap S
Kemudian hitung nilai gain menggunakan persamaan 2 dibawah :

$$Gain\ (S, A) = \ entropy\ (S) - \sum_{i=1}^{n} \frac{|Si|}{S} * Entropy\ (Si) \qquad \text{..... (2)}$$

Keterangan :
S = Himpunan Kasus
A = Fitur
n = jumlah partisi attribute A

|Si| = Proporsi Si terhadap S
|S| = jumlah kasus dalam S

3. *Kemudian untuk tahap selanjutnya mengulangi proses 2 dan 3 hingga semua record terproses. Proses pohon keputusan akan berhenti saat :*
   a. Semua record dalam nilai N mendapat kelas yang sama.
   b. Tidak ada attribute di dalam record yang diproses lagi
   c. Tidak ada record yang menjadi cabang yang kosong

Sehingga akan diperoleh nilai Gain dari attribute yang paling tertinggi. Gain adalah salah satu attribute tiap node pada tree. attribute dengan nilai Gain tertinggi akan dipilih sebagai test attribute untuk node berikutnya.

Algoritma C4.5 merupakan algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki kelebihan utama dari algoritma yang lainnya.

2.4 *Waikato Environment Knowledge and Analisys* (WEKA)

Weka adalah sebuah perangkat lunak *data mining*/*machine learning* yang dibangun oleh *Department of Computer Science University of Waikato* di New Zealand. Weka dikembangkan mulai tahun 1999 sampai sekarang. Weka ditulis dengan bahasa pemprograman Java. Saat ini Weka didistribusikan dibawah lisensi publik GNU. *Software* ini banyak digunakan untuk riset, edukasi dan aplikasi pelengkap (komplemen) *data mining* yang dibuat oleh Witten dan Frank.

WEKA mudah digunakan dan diterapkan pada beberapa tingkatan yang berbeda. Tersedia implementasi algoritma-algoritma pembelajaran stateof-the-art yang dapat diterapkan pada dataset dari command line. WEKA mengandung tools untuk pre-processing data, klasifikasi, regresi, clustering, aturan asosiasi, dan visualisasi. User dapat melakukan preprocess pada data, memasukkannya dalam sebuah skema pembelajaran, dan menganalisa classifier yang dihasilkan dan performansinya – semua itu tanpa menulis kode program sama sekali. Contoh penggunaan WEKA adalah dengan menerapkan sebuah metode pembelajaran ke dataset dan menganalisa hasilnya untuk memperoleh informasi tentang data, atau menerapkan beberapa metode dan membandingkan performansinya untuk dipilih.(Putu Gede Surya Cipta Nugraha, 2016;37)

Hasil menerapkan *classifier* yang dipilih akan diuji sesuai dengan pilihan yang ditetapkan dengan mengklik pada kotak *Test Option*.

Ada empat mode tes:

1. *Use training set*

   Pengetesan dilakukan dengan menggunakan data *training* itu sendiri.

2. *Supplied test set*

   Pengetesan dilakukan dengan menggunakan data lain. Dengan menggunakan *option* inilah, bisa dilakukan prediksi terhadap data tes.

3. *Cross-validation*

Pada cross-validation, akan ada pilihan berapa fold yang akan digunakan. Nilai *default*-nya adalah 10. Mekanisme-nya adalah sebagai berikut :

Data *training* dibagi menjadi k buah *subset* (subhimpunan). Dimana k adalah nilai dari *fold*. Selanjutnya, untuk tiap dari *subset*, akan dijadikan data tes dari hasil klasifikasi yang dihasilkan dari k-1 *subset* lainnya. Jadi, akan ada 10 kali tes. Dimana, setiap *datum* akan menjadi data tes sebanyak 1 kali, dan menjadi data *training* sebanyak k-1 kali. Kemudian, *error* dari k tes tersebut akan dihitung rata-ratanya.

4. *Percentage split*

   Hasil klasifikasi akan dites dengan menggunakan k% dari data tersebut. K merupakan masukan dari *user*.

2.5 Contoh Kasus Pemanfatan *Decision tree* C4.5

Berikut ini contoh kasus pemanfaatan *decision tree* C4.5 dalam memprediksi kelulusan mahasiswa :

Penelitian Indah Puji Astuti yang berjudul “Prediksi Ketepatan Waktu Kelulusan dengan Algoritma Data Mining C4.5”. Target variabel pada penelitian ini adalah lulus tepat waktu dan lulus tidak tepat waktu. Langkah - langkah penelitian menggunakan model CRISP-DM *(Cross Industry Standard Process for Data Mining),* jumlah data set yang digunakan adalah sebanyak 100 *record.*

Atribut yang digunakan dalam penelitian ini adalah :

1. NIM

   Atribut NIM merupakan ID.

2. Jenis sekolah asal

   Atribut jenis sekolah asal terdiri dari 3 macam, yaitu “SMA”, “MAN”, dan “SMK”.

3. Asal daerah

   Atribut asal daerah terdiri dari 2 macam, yaitu “Ponorogo” dan “Luar Ponorogo”.

4. Pekerjaan orang tua

   Atribut pekerjaan orang tua terdiri dari 5 macam jenis pekerjaan, yaitu PNS, wiraswasta, petani, guru, dan polisi.

5. Kelas

Atribut kelas terdiri dari 2 macam, yaitu digunakan untuk membedakan bahwa mahasiswa berasal dari kelas Reguler atau kelas Khusus.

Tabel 2.1. Tabel Perhitungan *Node* Indah Puji Astuti, 2017; 8

Node			Jml data	TP	TTP	Entrophy	Gain
	Total		100	83	17	0.657704779	
	Jenis Sekolah Asal	SMK	46	40	6	0.558629373	0.007785513
		SMA	46	37	9	0.713146749	
		MAN	8	6	2	0.811278124	
	Asal Daerah	Ponorogo	76	66	10	0.561752608	0.021765324
		Luar Ponorogo	24	17	7	0.870864469	
	Pekerjaan Orangtua	PNS	22	20	2	0.439496987	0.095583014
		Wiraswasta	50	43	7	0.584238812	
		Petani	23	19	4	0.666578358	
		Guru	3	0	3	0	
		Polisi	2	1	1	1	
	Kelas	Reguler	87	71	16	0.688552168	0.00780273
		Khusus	13	12	1	0.391243564	

Sumber : Indah Puji Astuti, 2017; 8

Keterangan :

TP : Lulus Tepat Waktu

TTP : Lulus Tidak Tepat Waktu

Nilai atribut pekerjaan orangtua memiliki nilai *gain* tertinggi, maka atribut ini menjadi atribut *root* pada pohon keputusan, kemudian dilanjutkan dengan atribut daerah, dan jenis sekolah asal. Diakhiri oleh status yang menyatakan keterangan tepat dan tidak tepat yang berfungsi sebagai *leaf*. Maka dapat dikatakan bahwa, kasus pada data penelitian ini parameter penentu pertama tingkat kelulusan seorang mahasiswa pada waktu yang akan datang dilihat dari pekerjaan orang tua, daerah, dan jenis sekolah asal mahasiswa tersebut.

Akurasi klasifikasi didapatkan berdasarkan tabel confusion matrix. Confusion matrix dari data testing yang digunakan dengan output Tepat dan Tidak Tepat dapat dilihat pada Gambar 2.2.

```
=== Confusion Matrix ===

  a  b   <-- classified as
 79  4 |  a = TEPAT
 14  3 |  b = TIDAK TEPAT
```

Gambar 2.2
Evaluasi Menggunakan *Confusion Matrix* (Indah Puji Astuti, 2017; 8)

Tabel 2.2. Tabel Rumus Penilaian *accuracy* dengan konsep *confution matrix*

		Predict Class	
		Diidentifikasikan Tepat Waktu	Diidentifikasikan Tidak Tepat Waktu
Actual Class	Tepat waktu	a	b
	Tidak Tepat Waktu	c	d

Sumber : Indah Puji Astuti, 2017; 8

$$Accuracy = \frac{(a+d)}{(totalsample)} x\ 100\%$$

$$Accuracy = \frac{79+3}{100}\ x\ 100\ \%$$

$$= 0{,}82$$

Gambar 2.3
Perhitungan Rumus Akurasi (Indah Puji Astuti, 2017; 8)

Berdasarkan *confusion matrix* di atas didapatkan tingkat akurasi klasifikasi algoritma C4.5 sebesar 82%.

2.6 Evaluasi Menggunakan *Confusion Matrix*

Dalam tahap ini akan dilakukan pengukuran keakuratan hasil yang telah dicapai. Pengukuran terhadap kinerja suatu sistem klasifikasi merupakan hal yang penting. Kinerja sistem klasifikasi menggambarkan seberapa baik sistem dalam mengklasifikasikan data. *Confusion matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi.

"Confusion matrix merupakan sebuah tabel yang terdiri dari banyaknya baris data uji yang diprediksi benar dan tidak benar oleh model klasifikasi". (Liliana Swastina, 2013:95)

Hasil klasifikasi akan dihadirkan dalam betuk *confusion matrix*. Tabel ini terdiri dari *predict class* dan *actual class*. Model *confusion matrix* 3x3 ditunjukkan pada Tabel 2.3.

Tabel 2.3 Tabel *Confusion Matrix*

		Predict Class		
		Class A	*Class* B	*Class* C
Actual Class	*Class* A	AA	AB	AC
	Class B	BA	BB	BC
	Class C	CA	CB	CC

Sumber : Derrick Iskandar, 2014;6

Hasil klasifikasi dapat dihitung tingkat akurasi dan *error rate* berdasarkan kinerja matriks. Untuk menghitung tingkat akurasi dan *error rate* pada matriks digunakan perhitungan rumus pada persamaan 3 dan 4 :

$$\text{Akurasi} = \frac{AA+BB+CC}{AA+AB+AC+BA+BB+BC+CA+CB+CC} \times 100\% \quad \text{..... (3)}$$

$$\textit{Error Rate} = \frac{AB+AC+BA+BC+CA+CB}{AA+AB+AC+BA+BB+BC+CA+CB+CC} \times 100\% \quad \text{..... (4)}$$

2.7 Penelitian terkait

Tabel 2.4 Tabel Penelitian terkait

No	Penulis	Judul	Variabel	Tahapan	Hasil
1.	A. Karim (2014)	*Pemodelan Aturan Dalam Memprediksi Prestasi Akademik Mahasiswa Politeknik Poliprofesi Medan dengan Kernel K-Means Clustering*	a. Nilai teori b. Nilai Praktek c. Kehadiran d. IPK e. Predikat	a. Membuat Kategorisasi data b. Mentransformasikan data c. Pengujian metode kernel K-means d. Menampilkan cluster berdasarkan prestasi akademik (niali teori, praktek, IPK dan kehadiran) e. Menampilkan cluster berdasarkan predikat f. Membuat model aturan untuk memprediksi prestasi akademik mahasiswa	Diperoleh model aturan untuk memprediksi prestasi akademik mahasiswa, yaitu mahasiswa yang memiliki predikat dengan pujian apabila nilai rata-rata teori tinggi(>70), nilai rata-rata praktek tinggi (>70) dan kehadiran baik (>93%)
2.	Ade Putra (2017)	Solusi Prediksi Mahasiswa *Drop Out* pada Program Studi Sistem Informasi Fakultas Ilmu	Nim, nama, program studi, IPK, sks, semester	a. Data yang diteliti adalah data lulusan program studi sistem informasi b. Melakukan Pembersihan data	Dengan memperhatikan bentuk dari pohon keputusan dapat diketahui bahwa semua *attribute* sudah masuk

No	Penulis	Judul	Variabel	Tahapan	Hasil
		Computer Universitas Bina Darma		c. Integrasi data d. Transformasi data e. Proses data mining menggunakan rapidminer, hasil dari rapidminer adalah *decision tree* prediksi mahasiswa drop out f. Perhitungan entrophy dan gain dari masing-masing atribut	kedalam class pada pohon keputusan, setelah pohon keputusan terbentuk dapat ditarik kesimpulan sebagai berikut : "JIKA Sks = Tidak Baik maka *Class* = " Drop Out " "JIKA Sks = Baik maka *Class* = " Tidak Drop Out " "JIKA Sks = Sangat Baik dan Semester = Baik maka *Class* =" Tidak Drop Out " "JIKA Sks = Sangat Baik dan Semester = Tidak Baik maka *Class* = " Drop Out "
3.	Asep saefulloh dan moedjion o (2013)	Penerapan Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Tepat Waktu	IPK (indeks prestasi kumulatif) dan IMK (indeks mutu kumulatif)	a. Penelitian ini didesain dengan menggunakan model CRISP-DM b. Komparasi Algoritma C4.5,*Naïve Bayes, Neural Network* c. Uji validitas menggunakan *Confusion Matrix* dan ROC (*Receveir Operating Characteristic*), pengukuran yang biasa digunakan adalah precision, recall dan accuracy d. Analisa hasil komparasi	Algoritma terbaik adalah algoritma yang paling tinggi tingkat *accuracy* pada model klasifikasi yaitu C4.5 dan *Neural Network* dengan tingkat *accuracy* 100% sedangkan Naïve Bayes 99.8878% . Ketiga algoritma tersebut termasuk klasifikasi sangat baik karena memiliki nilai AUC *(Area Under Curve)* antara 0.90-1.00 sehingga dapat dipergunakan untuk aplikasi prediksi.
4.	Diana Laily Fithri dan Eko darmanto (2014)	Sistem Pendukung Keputusan Untuk Memprediksi Kelulusan	Umur, alamat, jenis kelamin, status pekerjaan mahasiswa,	a. Melakukan transformasi data untuk mendapatkan atribut yang relevan dan sesuai	Tingkat akurasi 93%

No	Penulis	Judul	Variabel	Tahapan	Hasil
		Mahasiswa Menggunakan Metode Naïve Bayes	status pernikahan mahasiswa, rata-rata IPK, jumlah sks, status mahasiswa	dengan format input algoritma naïve bayes b. Penerapan metode naïve bayes c. Pengujian dilakukan dengan menggunakan sebagian data mahasiswa untuk training dan sebagian lagi dengan data testing d. Evaluasi dilakukan dengan mengamati hasil prediksi menggunakan algoritma naïve bayes.Validasi dilakukan dengan mengukur hasil prediksi dibandingkan dengan data asal.	
5.	David Hartanto Kamagi dan Seng Hansun (2014)	Implementasi Data Mining dengan algoritma C4.5 untuk memprediksi tingkat kelulusan mahasiswa	Jenis kelamin, asal sekolah, IPS 1 s.d. IPS 6	a. Perancangan dan pembangunan aplikasi dilakukan dengan menggunakan bahasa C# untuk aplikasi desktop dengan menerima masukan berupa file excel untuk data training dan data testing b. Implementasi aplikasi dilakukan untuk memprediksi tingkat kelulusan berdasarkan data yang telah diperoleh c. Uji coba terhadap aplikasi yang telah dibuat disertai dengan hasil evaluasi	a. Atribut yang paling berpengaruh adalah IPS semester 6 b. Aplikasi desktop berhasil memprediksi kelulusan mahasiswa dengan presentasi 87,5% dari 60 data training dan 40 data testing

No	Penulis	Judul	Variabel	Tahapan	Hasil
				d. Uji coba dilakukan untuk memperlihatkan apakah algoritma C4.5 bisa memprediksi tingkat kelulusan mahasiswa	
6.	Muhammad Syukri Mustafa dan I Wayan Simpen (20	Perancangan Aplikasi Prediksi Kelulusan Mahasiswa Baru Dengan Teknik Data Mining (studi kasus : STMIK Dipanegara Makasar)	NEM, Jenis Kelamin, Jurusan SMA, provinsi	a. pengujian dengan menerapkan algoritma KNN dan menggunakan data sampel alumni tahun wisuda 2004 s.d. 2010 untuk kasus lama dan data alumni tahun wisuda 2011 untuk kasus baru b. menyiapkan tabel kasus yang berisi data akademik alumni yang sudah menyelesaikan studi. c. menyimpan semua data kasus dan kedekatan kedalam variable array. Selanjutnya pengguna menginput data kasus baru untuk dilakukan perhitungan jarak antar antara atribut data testing terhadap data kasus. d. *Record* data kasus ke i akan dibandingkan terhadap seluruh data kasus. Hasil perhitungan jarak disimpan dalam array dengan menggunakan metode sort	penerapan algoritma KNN dapat diketahui hubungan kedekatan antara kasus yang baru dengan jasus yang telah ada dalam suatu gudang data (data warehouse). Hasil tingkat akurasi : 83,36%

No	Penulis	Judul	Variabel	Tahapan	Hasil
				maximum yang kemudian menjadi dasar dalam menentukan kasus mana yang memiliki nilai kedekatan tertinggi sehingga menjadi acuan dalam menentukan hasil prediksi apakah mahasiswa baru tersebut dapat menyelesaikan studi “tepat waktu “ atau “tidak” dengan mengacu pada kolom atribut kelulusan dari tabel kasus tersebut.	
7.	Hasbul Bahar (2014)	Prediksi Lulus Tepat dan Tidak Tepat Waktu Mahasiswa Menggunakan Algoritma K-Means	Jenis Kelamin,IPS 1, IPS 2, IPS 3, IPS 4, status kelulusan	a. Metode yang diusulkan pada penelitian ini berdasarkan *state of the art* tentang prediksi lulus tepat dan tidak tepat waktu mahasiswa menggunakan *Clustering Data Mining*. Metode yang diusulkan untuk pengolahan data mahasiswa adalah penggunaan algoritma *K-Means*. Data diolah dengan algoritma *K-Means* di implementasikan denganRapidMiner b. setelah diolah dan mengahasilkan model, maka terhadap model yang dihasilkan tersebut dilakukan pengujian menggunakan *kfold cross validation*, kemudian dilakukan evaluasi	Metode algoritma *K-Means* menghasilkan nilai akurasi yaitu 84.43%. Metode *neural network* menghasilkan nilai akurasi 90.41%. Dari evaluasi *confusion matrix* tersebut terlihat bahwa nilai akurasi tertinggi ada pada metode *neural network* dengan *Execution Time* 2 *second* untuk *K-Means dan* 15 *second* untuk *neural network.*

No	Penulis	Judul	Variabel	Tahapan	Hasil
				dan validasi hasil dengan *confusion matrix*. c. memandingkan hasil akurasi model *KMeans* dengan metode sebelumnya *Neural Network*, sehingga diperoleh model dari metode mana yang memperoleh nilai akurasi tertinggi	
8.	Nurjoko dan Hendra Kurniawan (2016)	Aplikasi Data Mining Untuk Memprediksi Tingkat kelulusan mahasiswa menggunakan algoritma *apriori* di ibi darmajaya bandar lampung	NPM, Nama, Kota asal, asal sekolah,kota sekolah, angkatan, tahun lulus, lama studi, jurusan, keterangan lulus, IPK, keterangan IPK	a. Tahapan dalam penelitian ini dimulai dengan proses *Extract, Transform, Load* (ETL) dengan mengambil data dari tabel induk mahasiswa dan tabel alumni yang kemudian dimasukkan ke dalam tabel_ETL. Setelah beberapa data penting berupa kolom-kolom tertentu didefinisikan dan tabel_ETL telah berisi data, maka langkah selanjutnya adalahmelakukan transformasi dengan mengubah beberapa atribut dalam tabel_ETL menjadi data yang sesuai dengan rancangan transformasi. b. Proses *mining* yang dilakukan menggunakan sebuah model *mining structure* dengan Input berupa Kota asal dan asal sekolah, kemudian satu kolom predict yaitu KetLulus. Model mining ini digunakan untuk mengetahui	Informasi yang ditampilkan berupa nilai *support* dan *confidence* hubungan antara tingkat kelulusan dengan data induk mahasiswa. Semakin tinggi nilai *confidence* dan *support* maka semakin kuat nilai hubungan antar atribut.

No	Penulis	Judul	Variabel	Tahapan	Hasil
				hubungan (*asosiasi*) antara asal mahasiswa dan asal sekolah dengan tingkat kelulusan.	

Kesimpulan dari beberapa penelitian diatas, dikemukakan beberapa hal seperti :

1. Atribut yang sering digunakan dalam prediksi kelulusan adalah NEM, asal sekolah, jenis kelamin dan IPK.
2. Target variabel dari prediksi kelulusan hanya 2 (dua) yaitu : adalah lulus tepat waktu dan lulus tidak tepat waktu. Tidak ada penelitian yang menggunakan tiga variabel
3. Penelitian tidak dapat secara langsung diimplementasikan karena perbedaan data yang digunakan.

Dari beberapa penelitian diperoleh bahwa antara satu model prediksi dengan model prediksi yang lain belum dapat dipastikan untuk dapat diimplementasikan dengan menggunakan database yang berbeda.

2.8 Perbedaan Dengan Penelitian Lainnya

Dalam penelitian ini, target variabel yang digunakan untuk prediksi kelulusan ada tiga yaitu : lulus tepat waktu, lulus terlambat dan *drop out.* Variabel yang digunakan dalam mencari model prediksi kelulusan adalah

1. Program Studi
2. Jenis Kelamin
3. Asal Daerah
4. Asal Sekolah
5. IPK semester 4
6. status pekerjaan mahasiswa
7. pekerjaan orangtua
8. Kelas

Atribut-atribut tersebut akan di cari atribut mana yang menjadi akar atau atribut utama prediksi kelulusan. Penelitian ini juga akan mencari di semester berapa prediksi dapat dilakukan dengan tepat.

Data *training* digunakan untuk menemukan pola prediksi kelulusan, data *testing* digunakan untuk menguji pola yang telah dihasilkan dari proses klasifikasi. Proses *learning* data *training* dan pengujian model menggunakan alat bantu yaitu tools Weka 3.9.

BAB III

OBJEK DAN METODOLOGI PENELITIAN

3.1 Profil STMIK WIT

Lembaga pendidikan yang mulai dirintis sejak tahun 1982 ini memulai perannya melalui pendidikan kursus atau LPK, yaitu Akuntansi, Bahasa Inggris dan Komputer. Mengingat begitu besarnya peran pendidikan dalam membangun masyarakat,Yayasan Web Informatika Teknologi membuka Sekolah Tinggi dengan biaya relatif ringan. Dengan SK No.09/D/O/2007, STMIK WIT (Sekolah Tinggi Manajemen Informatika dan Komputer Web Informatika Teknologi) sekarang sudah membuka jurusan S1 Teknik Informatika, S1 Sistem Informasi, D3 Manajemen Informatika, dan D3 Komputerisasi Akuntansi.

3.2 Visi, Misi dan Tujuan

Visi

Menjadikan WIT Sekolah Tinggi terkemuka tingkat nasional di tahun 2025, berbasis Teknologi Informasi, mengabdi kepada masyarakat untuk mencerdaskan bangsa, serta menghasilkan lulusan yang mampu bersaing dalam era globalisasi.

Misi

1. Menyelenggarakan pendidikan bermutu dengan biaya ringan.
2. Melakukan aktualisasi ilmu sesuai perkembangan global.
3. Membangun jaringan lapangan kerja dan mengembang kan jiwa kewirausahaan.

Tujuan

1. Mengabdi kepada Masyarakat untuk mencerdaskan kehidupan Bangsa.
2. Menyelenggarakan Pendidikan bermutu dengan biaya ringan, dengan melakukan aktualisasi ilmu sesuai perkembangan global dan membangun jaringan lapangan kerja serta mengembangkan jiwa kewirausahaan.
3. Menghasilkan sarjana dan tenaga ahli tingkat madya yang sanggup melaksanakan tugas-tugas secara profesional sesuai dengan bidang keahlian masing-masing.

3.3 Analisa Masalah

Pada data mahasiswa STMIK WIT, baik program sarjana ataupun diploma telah terdata banyak mahasiswa yang tidak lulus tepat waktu dan *dop out*. Artinya terjadi ketidak seimbangan antara jumlah mahasiswa aktif dan jumlah kelulusan.

Berdasarkan masalah tersebut, pihak akademik menganalisa adanya faktor – faktor yang dapat mempengaruhi kelulusan, dilihat dari segi akademik maupun non akademik. Dengan menggunakan data awal masuk sampai dengan data kelulusan mahasiswa yang tersimpan pada database kampus, dapat menjadi solusi dalam memecahkan masalah tersebut dengan mencari pola dan kecendrungan dari faktor – faktor yang mempengaruhi tepat atau tidaknya waktu kelulusan mahasiswa.

Dengan pola yang dihasilkan dari pengkombinasian data diharapkan dapat digunakan untuk memprediksi kelulusan calon mahasiswa, agar dapat menghimbau calon mahasiswa supaya lebih mempersiapkan diri untuk mengikuti proses perkuliahan sampai dengan selesai.

3.4 Analisa Kebutuhan Data

Data yang akan digunakan dalam membentuk suatu pengetahuan berupa pola kombinasi untuk memprediksi apakah calon mahasiswa lulus tepat waktu, terlambat atau *drop out* adalah data mahasiswa dari awal masuk sampai dengan lulus yang sudah tersimpan dalam database.

Alur pengolahan data mining pada penelitian ini dibagi menjadi dua proses yaitu proses training dan testing.

3.4.1 Proses *Training*

Data yang digunakan dalam proses ini disebut dengan data *training*, data training adalah data yang akan dipelajari dalam membetuk pohon keputusan. Data mahasiswa yang akan digunakan dalam mencari pola untuk membentuk pohon keputusan atau data *training* adalah data mahasiswa tahun 2006 s.d 2010. Data ini digunakan karena data mahasiswa tahun 2006 s.d 2010 telah selesai masa studinya. Proses *training* yaitu memasukkan data sampel ke dalam tabel yang dipersiapkan untuk perhitungan. Tabel tersebut meliputi atribut, jumlah data keseluruhan, jumlah data yang sudah terklasifikasi

berdasarkan target yang ditentukan, dalam penelitian ini adalah lulus tepat waktu, lulus terlambat dan *drop out*. Tahapan selanjutnya adalah menerapkan algoritma C4.5 yaitu menghitung *entropy* dan *gain* pada tiap-tiap atribut untuk dijadikan bentuk *tree* atau pohon keputusan. Pohon keputusan merupakan aturan klasifikasi yang akan diterapkan pada proses testing. Pada penelitian ini perhitungan algoritma C4.5 tidak dilakukan secara manual melainkan dibantu dengan tools WEKA 3.9.

3.4.2 Proses *Testing*

Data *testing* adalah data yang akan digunakan untuk pengujian pohon keputusan yang dihasilkan. Data *training* dan *testing* dalam penelitian ini merupakan data yang berbeda tetapi susunan atribut tetap sama.. Setelah pohon keputusan terbentuk, langkah selanjutnya adalah menerapkan pola atau *rule* kelulusan. Pada proses *testing*, data yang akan digunakan sebagai data *testing* (data yang akan diuji atau diprediksi) adalah data mahasiswa tahun 2012. Setiap data dalam atribut akan dibandingkan dengan aturan yang sudah terbentuk pada perhitungan data *training* sebelumnya. Selanjutnya data akan diklasifikasikan berdasarkan target yang akan dicapai dalam penelitian ini yaitu lulus tepat waktu, lulus terlambat dan *drop out*.

3.5 Pemilihan Atribut

Pada proses pengolahan data mining akan digunakan beberapa atribut sebagai parameter dalam pengklasifikasian data *training*. Atribut atau variabel menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon keputusan.

Atribut-atribut yang akan digunakan dalam proses data mining ditentukan berdasarkan tujuan dari penelitian. Status kelulusan mahasiswa sebagai atribut yang akan dicari pola pengelompokkannya dan sebagai atribut yang akan diprediksi bagi mahasiswa aktif di STMIK WIT.

Pada penelitian ini, atribut ditentukan berdasarkan tingkat prioritas, artinya atribut dipilih berdasarkan prioritas yang paling tinggi ke yang paling rendah atau dari yang paling penting sampai yang tidak penting.

Untuk menetukan kelulusan mahasiswa, faktor nilai dan ekonomi merupakan faktor utama dalam penentuan atribut, dari segi nilai digunakan atribuk IPK sedangkan dari segi

ekonomi digunakan atribut penghasilan orangtua dan status pekerjaan mahasiswa. Setelah menentukan atribut yang paling utama dan penting, selanjutnya memilih atribut diluar faktor nilai dan ekonomi yaitu program studi, jenis kelamin, asal sekolah, asal daerah dan kelas.

Atribut nilai yang digunakan dalam penelitian ini adalah IPK semester 4 (rata-rata IP semester 1 sampai dengan 4). Penentuan prediksi kelulusan mahasiswa akan dilakukan setelah mahasiswa menempuh perkuliahan sampai dengan empat semester.

Beberapa atribut yang diambil sebagai penentu keputusan dari permasalahan kelulusan tepat waktu, terlambat dan *drop out* dengan urutan berdasarkan tingkat prioritas tertinggi ke yang terendah atau atribut yang paling penting ke yang tidak penting di antaranya adalah sebagi berikut :

1. IPK semester 4

   Atribut ini digunakan untuk mencari pola hubungan kelulusan dengan IPK semester 4, untuk mengetahui pengaruh nilai IPK di semester terhadap kelulusan mahasiswa

2. Gaji orangtua

   Atribut gaji orangtua digunakan untuk mencari pola hubungan kelulusan dengan penghasilan orangtua, untuk mengetahui pengaruh kondisi ekonomi orangtua dengan kelulusan mahasiswa

3. Status pekerjaan mahasiswa

   Atribut status pekerjaan digunakan untuk mencari pola hubungan kelulusan dengan status pekerjaan mahasiswa, atribut ini terdiri dari dua macam yaitu bekerja dan tidak bekerja

4. Program Studi (Prodi)

   Atribut ini digunakan untuk mencari pola hubungan kelulusan dengan program studi yang ada si STMIK WIT, atribut ini terdiri dari empat macam prodi yaitu : Komputerisasi Akuntansi, Manajemen Informatika, Teknik Informatika dan Sistem Informasi

5. Kelas

   Atribut ini digunakan untuk membedakan mahasiswa berdasarkan kelas pagi atau kelas sore

6. Asal sekolah

   Atribut asal sekolah digunakan untuk mencari pola hubungan kelulusan dengan asal sekolah, atribut ini dipertimbangkan sebagai dasar penentuan wilayah-wilayah strategis yang akan digunakan untuk kegiatan promosi oleh pihak kampus untuk mencari bibit unggul sebagai calon mahasiswa di STMIK WIT. Atribut asal sekolah terdiri dari empat macam yaitu : SMA, SMK, MAN dan Paket C

7. Jenis kelamin

   Atribut jenis kelamin digunakan untuk mencari pola hubungan kelulusan dengan jenis kelamin, jenis kelamin dipilih karena dipandang dapat mempengaruhi tingkat kedisiplinan dan kepandaian seorang mahasiswa.

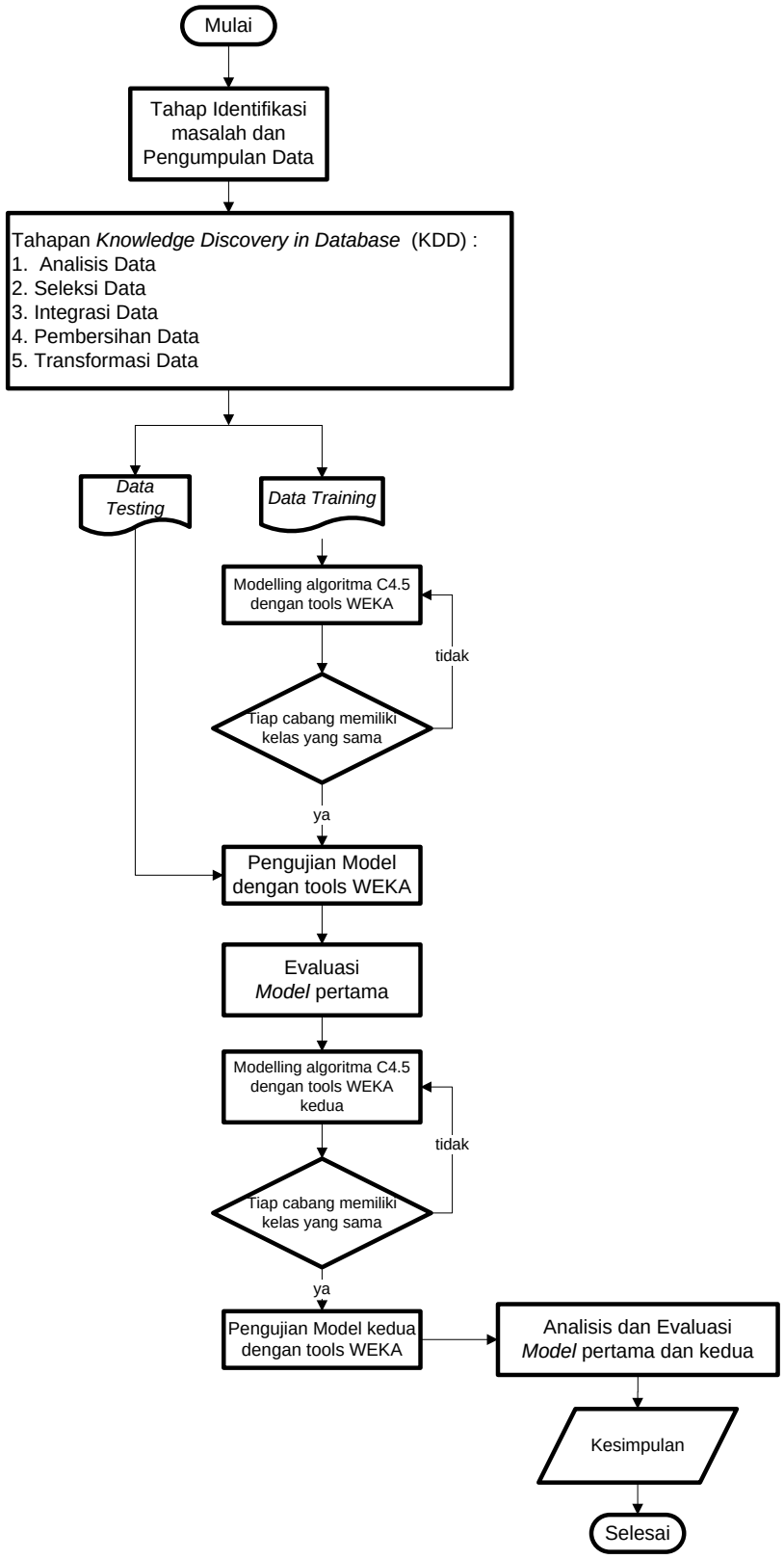
8. Asal daerah

   Atribut asal daerah digunakan untuk mencari pola hubungan kelulusan dengan asal daerah, atribut ini dipilih karena lingkungan dapat mempengaruhi kepribadian seseorang. Atribut asal daerah terdiri dari dua macam yaitu : Cirebon dan Luar Cirebon

Data mahasiswa tersebut akan diklasifikasikan berdasarkan target yang ditentukan dan dihitung menggunakan *Decision Tree* yaitu algoritma C4.5 untuk mencari nilai *Entropy* dan Informasi *Gain.* Setelah proses perhitungan selesai, maka akan menghasilkan kondisi atau rule yang digunakan dalam penentuan keputusan pada proses prediksi kelulusan. Target variabel dalam penelitian ini adalah lulus tepat waktu, terlambat dan *drop out.*

3.6 Metode Penelitian

Metode penelitian dapat dilihat pada Gambar 3.1 :



Gambar 3.1.
Metode Penelitian

Langkah – langkah penelitian dapat dijelaskan sebagai berikut :

1. Identifikasi Masalah

Pada Tahapan ini dilakukan analisa kebutuhan bisnis, analisa ini berfokus pada pemahaman tujuan kebutuhan berdasarkan penilaian bisnis. Kemudian pemahaman tersebut diubah menjadi sebuah rencana awal data mining yang dirancang untuk mencapai tujuan. Tujuan bisnis dari penelitian ini adalah mengenali pola mahasiswa untuk predikasi kelulusan.

2. Pengumpulan Data

Metode pengumpulan data yang tepat yaitu dengan mempertimbangkan penggunaannya berdasarkan jenis data dan sumbernya. Data yang objektif dan relevan dengan pokok permasalahan penelitian merupakan indikator keberhasilan suatu penelitian. Pengumpulan data penelitian ini dilakukan dengan cara sebagai berikut:

a. *Observasi*

   Merupakan metode pengumpulan data dengan cara mengadakan pengamatan langsung kepada objek penelitian.

b. Wawancara

   Merupakan teknik pengumpulan data dengan cara mengadakan tanya jawab atau wawancara langsung kepada bagian administrasi dan akademik STMIK WIT.

c. Studi Pustaka

   Studi pustaka, mengumpulkan data dengan mempelajari masalah yang berhubungan dengan objek yang diteliti serta bersumber dari buku- buku, literatur yang disusun oleh para ahli untu melengkapi data yang diperlukan dalam penelitian.

3. Analisa Data

Pada tahapan ini dilakukan analisis untuk memahami data yang didapatkan dari hasil pengumpulan data. Data yang akan dianalisa ini adalah data mahasiswa STMIK WIT jenjang D3 dan S1 pada tahun 2006 sampai dengan tahun 2010. Kemudian dilakukan eksplorasi dan analisa struktur tabel yang ada pada database STMIK WIT.

Output dari eksplorasi data ini adalah : tabel mahasiswa, tabel dosen, tabel mata kuliah, tabel FRS, rekap nilai dan tabel kelulusan.

4. Seleksi Data

Pada tahapan ini dilakukan rancangan set data yang sesuai dengan tujuan data mining. Pada tahap ini tabel yang berhubungan dipilih untuk mempermudah proses pemilihan data. Daftar tabel yang telah di eksplorasi adalah : tabel dosen, tabel mahasiswa, tabel mata kuliah, tabel FRS dan tabel kelulusan. Dari data tersebut, dipilih tabel yang berhubungan dengan penelitian ini, yaitu data mahasiswa, nilai dan kelulusan.

5. Integrasi Data

Proses ini adalah menggabungkan beberapa atribut dari data mahasiswa, rekap nilai dan kelulusan menjadi satu tabel.

6. Pembersihan Data

Pada tahapan ini akan dilakukan pembersihan data dengan membuang data yang tidak konsisten atau *noise,* duplikasi data, memperbaiki kesalahan data dan bisa diperkaya dengan data eksternal yang relevan.

Hasil evaluasi terhadap kualitas data adalah masih terdapat data yang rangkap dan bernilai null, sehingga jumlah dataset yang didapatkan dari proses pembersihan data ini adalah 977 *record,* seperti terlihat pada Tabel 3.1.

Tabel 3.1. *Dataset* jumlah mahasiswa STMIK WIT Angkatan 2006-2010

No	Program studi	Jenjang	Jumlah
1	Komputerisasi Akuntansi	D3	322
2	Manajemen Informatika	D3	296
3	Sistem Informasi	S1	142
4	Teknik Informatika	S1	217
Total			977

7. Transformasi Data

Pada tahap ini, akan dilakukan proses pengubahan data ke dalam format yang sesuai untuk diproses dalam data mining.

Data_mahasiswa_semester1-4 terdiri dari beberapa atribut antara lain jenis kelamin, asal daerah, asal sekolah, IPK semester 4, status pekerjaan mahasiswa, gaji orangtua. Pengkategorian data adalah sebagai berikut :

a. Atribut Program Studi

   Dikategorikan menjadi empat macam yaitu komputerisasi akuntansi, manajemen infomatika, Teknik informatika dan Sistem Informasi

b. Atribut jenis kelamin

   Dikategorikan menjadi dua yaitu P untuk jenis kelami perempuan dan L untuk jenis kelamin laki-laki

c. Atribut asal daerah

   Dikategorikan menjadi dua yaitu Cirebon dan luar Cirebon

d. Atribut asal sekolah

   Dikategorikan berdasarkan asal sekolah mahasiswa yaitu SMA, SMK, MAN dan PAKET C

e. Atribut Indeks Prestasi Kumulatif (IPK)

   Jenis data IPK diambil dari IPK semester 4. atribut dari IPK dikategorikan menjadi 3, seperti ditampilkan dalam Tabel 3.2.

Tabel 3.2 Kategori Indeks Prestasi Semester (IPS)

Kategori	Keterangan
Besar	IPS ≥ 3,00
Sedang	2,00 ≤ IPS ≤ 2,99
Kecil	IPS < 2,00

f. Variabel status pekerjaan mahasiswa

   Jenis data pada variabel status pekerjaan mahasiswa dikategorikan menjadi dua, seperti yang ditampilkan dalam Tabel 3.3

Tabel 3.3 kategori status pekerjaan

Kategori	Keterangan
Bekerja	Bekerja
Mahasiswa	Tidak Bekerja

g. Atribut Gaji ortu

atribut dari gaji orang tua dikategorikan menjadi tiga, terlihat pada Tabel 3.4

Tabel 3.4 Kategori Gaji Ortu

Kategori	Keterangan
Tinggi	>Rp. 3000.000
Sedang	Rp. 1.600.000 s.d Rp. 3000.000
Rendah	≤Rp. 1500.000

h. Atribut Kelas

Atribut ini dikategorikan menjadi dua yaitu pagi dan sore

Tabel 3.5 Cuplikan *Dataset* mahasiswa 2006-2010 yang belum ditransformasi

NIM	Nama Mahasiswa	PRODI	JK	Asal Daerah	Asal Sekolah	IPK 4	Gaji Ortu	Status Pekerjaan	KELAS	Status Kelulusan
060704001	Afiani Juniar	Komputerisasi Akuntansi	P	Cirebon	SMA	3,21	Rp 2.500.000	MAHASISWA	SORE	LULUS TEPAT WAKTU
060704002	Bobby Wahyu A.	Komputerisasi Akuntansi	L	Cirebon	MAN	1,11	Rp 2.500.000	BEKERJA	SORE	LULUS TERLAMBAT
060704003	Delia Kristianti	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	3,00	Rp 1.500.000	BEKERJA	PAGI	LULUS TERLAMBAT
060704004	Dian Lailatul Qadar	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	3,21	Rp 3.000.000	MAHASISWA	SORE	LULUS TEPAT WAKTU
060704005	Heni Yulianti	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	2,42	Rp 2.500.000	BEKERJA	PAGI	LULUS TERLAMBAT
060704006	Iman Antoni	Komputerisasi Akuntansi	L	Luar Cirebon	SMK	2,32	Rp 2.500.000	BEKERJA	PAGI	LULUS TERLAMBAT
060704007	Inon Febriyani	Komputerisasi Akuntansi	P	Luar Cirebon	PAKET C	2,63	Rp 750.000	BEKERJA	PAGI	LULUS TERLAMBAT
060704008	Juju Juhariah	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	2,32	Rp 1.760.000	BEKERJA	PAGI	DROP OUT
060704009	Kurniati	Komputerisasi Akuntansi	P	Cirebon	SMA	3,37	Rp 2.500.000	BEKERJA	PAGI	DROP OUT
060704010	Andhini Nur'aini	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	0,00	Rp 2.500.000	BEKERJA	PAGI	DROP OUT
060704011	Lilis Nurhayati	Komputerisasi Akuntansi	P	Cirebon	SMA	3,53	Rp 1.200.000	BEKERJA	PAGI	DROP OUT
060704012	Nazarudin	Komputerisasi Akuntansi	L	Luar Cirebon	SMA	3,05	Rp 800.000	BEKERJA	PAGI	DROP OUT
060704013	Puspa Indah	Komputerisasi Akuntansi	P	Luar Cirebon	SMK	3,00	Rp 2.500.000	MAHASISWA	PAGI	LULUS TEPAT WAKTU
060704014	Ronald Sumarko	Komputerisasi Akuntansi	L	Luar Cirebon	SMA	2,00	Rp 1.000.000	MAHASISWA	SORE	LULUS TERLAMBAT
060704015	Sri Mulyati	Komputerisasi Akuntansi	P	Luar Cirebon	SMK	2,53	Rp 1.000.000	BEKERJA	SORE	LULUS TERLAMBAT
060704016	Ade Rusmin	Komputerisasi Akuntansi	L	Luar Cirebon	SMK	1,84	Rp 1.000.000	BEKERJA	SORE	DROP OUT
060704017	Dwi Putri Mawar Dini	Komputerisasi Akuntansi	P	Luar Cirebon	SMK	2,16	Rp 1.000.000	BEKERJA	SORE	LULUS TERLAMBAT
060704018	Fitri Rayanti	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	2,47	Rp2.500.000	BEKERJA	PAGI	DROP OUT
060704019	Hidayat	Komputerisasi Akuntansi	L	Luar Cirebon	SMK	3,37	Rp2.500.000	BEKERJA	PAGI	DROP OUT
060704020	Noviani	Komputerisasi Akuntansi	P	Cirebon	MAN	2,32	Rp 3.000.000	BEKERJA	PAGI	LULUS TERLAMBAT

Tabel 3.6 Seleksi Atribut

Atribut	Detail Penggunaan	
NIM	x	ID
Nama	x	No
Program studi	✓	Nilai Model
Jenis Kelamin	✓	Nilai Model
Asal Daerah	✓	Nilai Model
Asal Sekolah	✓	Nilai Model
IPK4	✓	Nilai Model
Status Pekerjaan	✓	Nilai Model
Gaji Ortu	✓	Nilai Model

Atribut	Detail Penggunaan	
Kelas	✓	Nilai Model
Status	✓	Label Target

Tabel 3.6 menjelaskan atribut-atribut yang akan digunakan dalam penelitian, indikator "yes" (√) menandakan bahwa atribut bersangkutan akan digunakan dalam penelitian, sedangkan indikator "No" (x) menandakan atribut tersebut akan dieliminasi.

Tabel 3.7 Cuplikan *Dataset* mahasiswa 2006-2010 yang telah ditransformasi

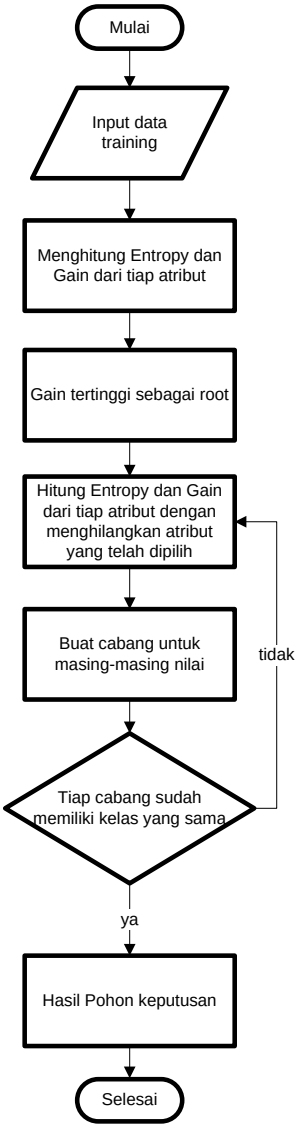
Program Studi	JK	Asal Daerah	ASAL SEKOLAH	IPK4	GAJI ORTU	STATUS PEKERJAAN	KELAS	STATUS
Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Sedang	Tidak Bekerja	SORE	Lulus Tepat Waktu
Komputerisasi Akuntansi	L	Cirebon	MAN	Kecil	Sedang	Bekerja	SORE	Lulus Terlambat
Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Besar	Rendah	Bekerja	PAGI	Lulus Terlambat
Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Besar	Sedang	Tidak Bekerja	SORE	Lulus Tepat Waktu
Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Sedang	Tinggi	Bekerja	PAGI	Lulus Terlambat
Komputerisasi Akuntansi	L	Luar Cirebon	SMK	Sedang	Sedang	Bekerja	PAGI	Lulus Terlambat
Komputerisasi Akuntansi	P	Luar Cirebon	PAKET C	Sedang	Sedang	Bekerja	PAGI	Lulus Terlambat
Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Sedang	Sedang	Bekerja	PAGI	Drop Out
Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Rendah	Bekerja	PAGI	Drop Out
Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Kecil	Sedang	Bekerja	PAGI	Drop Out
Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Rendah	Bekerja	PAGI	Drop Out
Komputerisasi Akuntansi	L	Luar Cirebon	SMA	Besar	Rendah	Bekerja	PAGI	Drop Out
Komputerisasi Akuntansi	P	Luar Cirebon	SMK	Besar	Sedang	Tidak Bekerja	PAGI	Lulus Tepat Waktu
Komputerisasi Akuntansi	L	Luar Cirebon	SMA	Sedang	Tinggi	Tidak Bekerja	SORE	Lulus Terlambat
Komputerisasi Akuntansi	P	Luar Cirebon	SMK	Sedang	Sedang	Bekerja	SORE	Lulus Terlambat
Komputerisasi Akuntansi	L	Luar Cirebon	SMK	Kecil	Rendah	Bekerja	SORE	Drop Out

8. *Modelling* dengan algoritma C4.5

Secara umum, fase algoritma C4.5 dalam membangun decision tree adalah sebagai berikut :

a. Pilih atribut sebagai root
b. Buat cabang untuk masing-masing nilai
c. Bagi kasus dalam cabang
d. Ulangi proses untuk masing-masing cabang sampai semua kasus memiliki kelas yang sama.

Langkah awal adalah mengelompokkan data training berdasarkan target yang ingin dicapapi yaitu lulus tepat waktu, lulus terlambat dan *drop out.* Lalu selanjutnya adalah menghitung *Entropy* dan *Gain* berdasarkan pengelompokkan data training. Nilai tertinggi dari beberapa atribut akan dijadikan root atau akar pada *decision tree,* dan sisa atribut yang ada akan dihitung kembali dengan rule yang sama yaitu dihitung dengan mencari nilai Gain tertinggi untuk dijadikan cabang dari akar pertama, rule perhitungan terus berlanjut sampai semua atribut dapat diketahui hasil akhirnya pada pola *tree* yang terbentuk.



Gambar 3.2.
Flowchart Algoritma C4.5

Pada penerapan *data mining* ini dibantu menggunakan aplikasi Weka 3.9. Weka merupakan sebuah perangkat lunak yang memiliki banyak algoritma *machine learning* untuk keperluan *data mining*.

9. Pengujian Model

Dalam tahap ini akan dilakukan pengujian model terhadap data *testing*, *rule* atau aturan pohon keputusan yang telah dihasilkan dari tahap sebelumnya akan diuji menggunakan data testing dengan bantuan tools Weka 3.9. Data *testing* dan data *training* merupakan data yang berbeda, tetapi harus memiliki atribut yang sama.

10. Evaluasi Hasil

Dalam tahap ini akan dilakukan pengukuran keakuratan hasil yang telah dicapai. Tahap ini digunakan untuk menguji kualitas dari data, apakah pola atau informasi yang ditemukan bersesuaian atau bertentangan dengan fakta sebelumnya. Evaluasi algoritma menggunakan *Confusion Matrix* berbentuk matrix 3x3.

11. Kesimpulan

Setelah dilakukan evaluasi kemudian maka dapat ditarik kesimpulan apakah informasi yang ditemukan sesuai dengan fakta sebelumnya atau tidak, sehingga akan diperoleh pengetahuan baru. Dengan adanya pola untuk prediksi kelulusan mahasiswa, diharapkan dapat membantu STMIK WIT dalam proses pembuatan keputusan yang akan datang.

BAB IV

HASIL DAN PEMBAHASAN

4.1 *Modelling* Algoritma C4.5 Pertama

Jumlah data *training* yang akan digunakan untuk pengujian dalam pembentukan *decision tree* adalah sebanyak 977, sedangkan jumlah data testing adalah 130. Data *training* dan *testing* yang digunakan dalam *modelling* algoritma C4.5 ini merupakan data yang berbeda. Data *training* diambil dari data mahasiswa angkatan 2006 s.d 2010, sedangkan data *testing* diambil dari data mahasiswa tahun 2012.

Tabel 4.1 Cuplikan *Data training* yang telah melewati data *preprocessing*

Program Studi	JK	Asal Daerah	Asal Sekolah	IPK4	Gaji Ortu	Status Pekerjaan	Kelas	Status
Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Sedang	Tidak Bekerja	SORE	Lulus Tepat Waktu
Komputerisasi Akuntansi	L	Cirebon	MAN	Kecil	Sedang	Bekerja	SORE	Lulus Terlambat
Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Besar	Rendah	Bekerja	PAGI	Lulus Terlambat
Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Besar	Sedang	Tidak Bekerja	SORE	Lulus Tepat Waktu
Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Sedang	Tinggi	Bekerja	PAGI	Lulus Terlambat
Komputerisasi Akuntansi	L	Luar Cirebon	SMK	Sedang	Sedang	Bekerja	PAGI	Lulus Terlambat
Komputerisasi Akuntansi	P	Luar Cirebon	PAKET C	Sedang	Sedang	Bekerja	PAGI	Lulus Terlambat
Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Sedang	Sedang	Bekerja	PAGI	Drop Out
Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Rendah	Bekerja	PAGI	Drop Out
Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Kecil	Sedang	Bekerja	PAGI	Drop Out
Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Rendah	Bekerja	PAGI	Drop Out
Komputerisasi Akuntansi	L	Luar Cirebon	SMA	Besar	Rendah	Bekerja	PAGI	Drop Out
Komputerisasi Akuntansi	P	Luar Cirebon	SMK	Besar	Sedang	Tidak Bekerja	PAGI	Lulus Tepat Waktu
Komputerisasi Akuntansi	L	Luar Cirebon	SMA	Sedang	Tinggi	Tidak Bekerja	SORE	Lulus Terlambat
Komputerisasi Akuntansi	P	Luar Cirebon	SMK	Sedang	Sedang	Bekerja	SORE	Lulus Terlambat
Komputerisasi Akuntansi	L	Luar Cirebon	SMK	Kecil	Rendah	Bekerja	SORE	Drop Out
Komputerisasi Akuntansi	P	Luar Cirebon	SMK	Sedang	Sedang	Bekerja	SORE	Lulus Terlambat
Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Sedang	Sedang	Bekerja	PAGI	Drop Out
Komputerisasi Akuntansi	L	Luar Cirebon	SMK	Besar	Rendah	Bekerja	PAGI	Drop Out

Tabel 4.1 merupakan hasil dari data *preprocessing* yang kemudian disebut dengan *data training,* data ini akan dipelajari untuk membentuk pola atau pohon keputusan. Algoritma yang akan digunakan adalah C4.5.

Algoritma C4.5 akan menghasilkan *decision tree* atau pohon keputusan. Cara menentukan atribut mana yang akan menjadi akar atau *root*, dilihat pada nilai *gain* tertinggi dari atribut-atribut yang ada. Perhitungan *node* secara manual untuk menentukan *root* dapat dilihat pada Tabel 4.2.

Tabel 4.2 Tabel Perhitungan Node

Node			Jml data	TW	T	DO	Entrophy	Gain
	Total		977	162	276	539	1,418426933	
	Sekolah Asal	SMA	418	96	135	187	1,533206573	0,029342056
		SMK	470	55	118	297	1,281241536	
		MAN	64	10	17	37	1,383501191	
		PAKET C	25	1	6	18	1,021119189	
	IPK	Besar	222	108	90	24	1,380750421	0,537290897
		Sedang	363	54	182	127	1,438411124	
		Kecil	392	0	4	388	0,082143051	
	Gaji Ortu	Tinggi	9	3	6	0	0,918295834	0,014262044
		Sedang	633	92	179	362	1,380761634	
		Rendah	335	67	91	177	1,461439762	
	Status Pekerjaan	Bekerja	639	57	177	405	1,241016342	0,062901185
		Mahasiswa	338	105	99	134	1,572009507	
	Program Studi	KA	322	86	97	139	1,553329537	0,063116799
		MI	296	32	66	198	1,217755316	
		TI	217	14	55	148	1,13354731	
		SI	142	30	58	54	1,531905284	
	Jenis Kelamin	L	592	67	145	380	1,263406195	0,040403887
		P	385	95	131	159	1,554265062	
	Asal daerah	Cirebon	749	122	216	411	1,418893431	0,000454012
		Luar Cirebon	228	40	60	128	1,414948964	
	Kelas	Pagi	545	85	149	311	1,391444441	0,001390926
		Sore	432	77	127	228	1,449321664	

Keterangan :

TW : Lulus tepat waktu

T : Lulus terlambat

DO : *Drop Out*

Perhitungan untuk mendapatkan nilai *entropy* dan *gain* seperti terlihat pada Tabel 4.2, yaitu dengan menggunakan rumus berikut :

$$Entropy(S) = \sum_{i}^{c} -p_i \log_2 p_i \qquad \text{..... (1)}$$

Menghitung *entropy* total :

$$\text{Entropy total} = \sum_{i=1}^{n} \left(-\frac{162}{977} * \log_2\left(\frac{162}{977}\right)\right) + \left(-\frac{276}{977} * \log_2\left(\frac{276}{977}\right)\right) + \left(-\frac{539}{977} * \log_2\left(\frac{539}{977}\right)\right)$$

$$= 1{,}418426933$$

Menghitung *entropy* asal sekolah :

$$\text{Entropy (asal sekolah SMA)} = \sum_{i=1}^{n} \left(-\frac{96}{418} * \log_2\left(\frac{96}{418}\right)\right) + \left(-\frac{135}{418} * \log_2\left(\frac{135}{418}\right)\right) + \left(-\frac{187}{418} * \log_2\left(\frac{187}{418}\right)\right)$$

$$= 1{,}533206573$$

$$\text{Entropy (asal sekolah SMK)} = \sum_{i=1}^{n} \left(-\frac{55}{470} * \log_2\left(\frac{55}{470}\right)\right) + \left(-\frac{118}{470} * \log_2\left(\frac{118}{470}\right)\right) + \left(-\frac{297}{470} * \log_2\left(\frac{297}{470}\right)\right)$$

$$= 1{,}281241536$$

$$\text{Entropy (asal sekolah MAN)} = \sum_{i=1}^{n} \left(-\frac{10}{64} * \log_2\left(\frac{10}{64}\right)\right) + \left(-\frac{17}{64} * \log_2\left(\frac{17}{64}\right)\right) + \left(-\frac{37}{64} * \log_2\left(\frac{37}{64}\right)\right)$$

$$= 1{,}383501191$$

$$\text{Entropy (asal sekolah Paket C)} = \sum_{i=1}^{n} \left(-\frac{1}{25} * \log_2\left(\frac{1}{25}\right)\right) + \left(-\frac{6}{25} * \log_2\left(\frac{6}{25}\right)\right) + \left(-\frac{18}{25} * \log_2\left(\frac{18}{25}\right)\right)$$

$$= 1{,}021119189$$

Menghitung *entropy* IPK4 :

$$\text{Entropy (IPK besar)} = \sum_{i=1}^{n} \left(-\frac{108}{222} * \log_2\left(\frac{108}{222}\right)\right) + \left(-\frac{90}{222} * \log_2\left(\frac{90}{222}\right)\right) + \left(-\frac{24}{222} * \log_2\left(\frac{24}{222}\right)\right)$$

$$= 1{,}380750421$$

$$\text{Entropy (IPK sedang)} = \sum_{i=1}^{n} \left(-\frac{54}{363} * \log_2\left(\frac{54}{363}\right)\right) + \left(-\frac{182}{363} * \log_2\left(\frac{182}{363}\right)\right) + \left(-\frac{127}{363} * \log_2\left(\frac{127}{363}\right)\right)$$

$$= 1{,}438411124$$

$$\text{Entropy (IPK kecil)} = \sum_{i=1}^{n} \left(-\frac{0}{392} * \log_2\left(\frac{0}{392}\right)\right) + \left(-\frac{4}{392} * \log_2\left(\frac{4}{392}\right)\right) + \left(-\frac{388}{363} * \log_2\left(\frac{388}{363}\right)\right)$$

$$= 0{,}082143051$$

Menghitung *entropy* gaji ortu :

$$\text{Entropy (gaji tinggi)} = \sum_{i=1}^{n} \left(-\frac{3}{9} * \log_2\left(\frac{3}{9}\right)\right) + \left(-\frac{6}{9} * \log_2\left(\frac{6}{9}\right)\right) + \left(-\frac{0}{9} * \log_2\left(\frac{0}{9}\right)\right)$$

$$= 0{,}918295834$$

$$\text{Entropy (gaji sedang)} = \sum_{i=1}^{n} \left(-\frac{92}{633} * \log_2\left(\frac{92}{633}\right)\right) + \left(-\frac{179}{633} * \log_2\left(\frac{179}{633}\right)\right) + \left(-\frac{362}{633} * \log_2\left(\frac{362}{633}\right)\right)$$

$$= 1{,}380761634$$

$$\text{Entropy (gaji rendah)} = \sum_{i=1}^{n} \left(-\frac{67}{335} * \log_2\left(\frac{67}{335}\right)\right) + \left(-\frac{91}{633} * \log_2\left(\frac{91}{633}\right)\right) + \left(-\frac{177}{633} * \log_2\left(\frac{177}{633}\right)\right)$$

$$= 1{,}461439762$$

Menghitung *entropy* status pekerjaan :

$$\text{Entropy (status bekerja)} = \sum_{i=1}^{n} \left(-\frac{57}{639} * \log_2\left(\frac{57}{639}\right)\right) + \left(-\frac{177}{639} * \log_2\left(\frac{177}{639}\right)\right) + \left(-\frac{405}{639} * \log_2\left(\frac{405}{639}\right)\right)$$

$$= 1{,}241016342$$

$$\text{Entropy (status mahasiswa)} = \sum_{i=1}^{n} \left(-\frac{105}{338} * \log_2\left(\frac{105}{338}\right)\right) + \left(-\frac{99}{338} * \log_2\left(\frac{99}{338}\right)\right) + \left(-\frac{134}{338} * \log_2\left(\frac{134}{338}\right)\right)$$

$$= 1{,}572009507$$

Menghitung *entropy* program studi :

$$\text{Entropy (prodi KA)} = \sum_{i=1}^{n} \left(-\frac{86}{322} * \log_2\left(\frac{86}{322}\right)\right) + \left(-\frac{97}{322} * \log_2\left(\frac{97}{322}\right)\right) + \left(-\frac{139}{322} * \log_2\left(\frac{139}{322}\right)\right)$$

$$= 1{,}553329537$$

$$\text{Entropy (prodi MI)} = \sum_{i=1}^{n} \left(-\frac{32}{296} * \log_2\left(\frac{32}{296}\right)\right) + \left(-\frac{66}{296} * \log_2\left(\frac{66}{296}\right)\right) + \left(-\frac{198}{296} * \log_2\left(\frac{198}{296}\right)\right)$$

$$= 1{,}217755316$$

$$\text{Entropy (prodi TI)} = \sum_{i=1}^{n} \left(-\frac{14}{217} * \log_2\left(\frac{14}{217}\right)\right) + \left(-\frac{55}{217} * \log_2\left(\frac{55}{217}\right)\right) + \left(-\frac{148}{217} * \log_2\left(\frac{148}{217}\right)\right)$$

$$= 1{,}13354731$$

$$\text{Entropy (prodi SI)} = \sum_{i=1}^{n} \left(-\frac{30}{142} * \log_2\left(\frac{30}{142}\right)\right) + \left(-\frac{58}{142} * \log_2\left(\frac{58}{142}\right)\right) + \left(-\frac{54}{142} * \log_2\left(\frac{54}{142}\right)\right)$$

$$= 1{,}531905284$$

Menghitung *entropy* jenis kelamin :

$$\text{Entropy (Jenis kelamin L)} = \sum_{i=1}^{n} \left(-\frac{67}{592} * \log_2\left(\frac{67}{592}\right)\right) + \left(-\frac{145}{592} * \log_2\left(\frac{145}{592}\right)\right) + \left(-\frac{380}{592} * \log_2\left(\frac{380}{592}\right)\right)$$

$$= 1{,}263406195$$

Entropy (Jenis kelamin P) = $\sum_{i=1}^{n}\left(-\frac{95}{385}*\log_2\left(\frac{95}{385}\right)\right)+\left(-\frac{131}{385}*\log_2\left(\frac{131}{385}\right)\right)+\left(-\frac{159}{385}*\log_2\left(\frac{159}{385}\right)\right)$

= 1,554265062

Menghitung *entropy* asal daerah :

Entropy (asal Cirebon) = $\sum_{i=1}^{n}\left(-\frac{122}{749}*\log_2\left(\frac{122}{749}\right)\right)+\left(-\frac{216}{749}*\log_2\left(\frac{216}{749}\right)\right)+\left(-\frac{411}{749}*\log_2\left(\frac{411}{749}\right)\right)$

= 1,418893431

Entropy (asal Luar Cirebon) = $\sum_{i=1}^{n}\left(-\frac{40}{228}*\log_2\left(\frac{40}{228}\right)\right)+\left(-\frac{60}{228}*\log_2\left(\frac{60}{228}\right)\right)+\left(-\frac{128}{228}*\log_2\left(\frac{128}{228}\right)\right)$

= 1,414948964

Menghitung *entropy* kelas pagi:

Entropy (Kelas Pagi) = $\sum_{i=1}^{n}\left(-\frac{85}{545}*\log_2\left(\frac{85}{545}\right)\right)+\left(-\frac{149}{545}*\log_2\left(\frac{149}{545}\right)\right)+\left(-\frac{311}{545}*\log_2\left(\frac{311}{545}\right)\right)$

= 1,391444441

Entropy (Kelas Sore) = $\sum_{i=1}^{n}\left(-\frac{77}{432}*\log_2\left(\frac{77}{432}\right)\right)+\left(-\frac{127}{432}*\log_2\left(\frac{127}{432}\right)\right)+\left(-\frac{228}{432}*\log_2\left(\frac{228}{432}\right)\right)$

= 1,449321664

Menghitung nilai *Gain* pada atribut asal sekolah dihitung dengan menggunakan formula *gain* sebagai berikut:

$$Gain\ (S,A) = entropy\ (S) - \sum_{i=1}^{n}\frac{|Si|}{S} * Entropy\ (Si) \quad \text{..... (2)}$$

Gain (Total,asal sekolah)= 1,418426933-$\left(\begin{array}{l}\left(\frac{418}{977}*1{,}533206573\right)+\left(\frac{470}{977}*1{,}281241536\right)\\+\left(\frac{64}{977}*1{,}383501191\right)+\left(\frac{25}{977}*1{,}021119189\right)\end{array}\right)$

= 0,029342056

Gain (Total,IPK4) = 1,418426933-
$\left(\left(\frac{222}{977}*1{,}380750421\right)+\left(\frac{363}{977}*1{,}438411124\right)+\left(\frac{392}{977}*0{,}082143051\right)\right)$

= 0,537290897

Gain (Total,Gajiortu) = 1,418426933-

$\left(\left(\frac{9}{977}*0{,}918295834\right)+\left(\frac{633}{977}*1{,}380761634\right)+\right)\left(\frac{335}{977}*1{,}461439762\right)\Big)$

= 0,014262044

Gain (Total,status pekerjaan)= 1,418426933-$\left(\left(\frac{639}{977}*1{,}241016342\right)+\left(\frac{338}{977}*1{,}572009507\right)\right)$

$= 0{,}062901185$

$\text{Gain (Total, program studi)} = 1{,}418426933 - \left(\left(\frac{322}{977} * 1{,}553329537\right) + \left(\frac{296}{977} * 1{,}217755316\right) + \left(\frac{217}{977} * 1{,}13354731\right) + \left(\frac{142}{977} * 1{,}531905284\right)\right)$

$= 0{,}063116799$

$\text{Gain (Total, jenis kelamin)} = 1{,}418426933 - \left(\left(\frac{592}{977} * 1{,}263406195\right) + \left(\frac{385}{977} * 1{,}554265062\right)\right)$

$= 0{,}040403887$

$\text{Gain (Total, asal daerah)} = 1{,}418426933 - \left(\left(\frac{749}{977} * 1{,}418893431\right) + \left(\frac{228}{977} * 1{,}414948964\right)\right)$

$= 0{,}000454012$

$\text{Gain (Total, kelas)} = 1{,}418426933 - \left(\left(\frac{545}{977} * 1{,}391444441\right) + \left(\frac{432}{977} * 1{,}449321664\right)\right)$

$= 0{,}0011390926$

Dari Tabel 4.2, atribut dengan nilai *gain* tertinggi dipilih sebagai *node* pertama pada pohon keputusan. Pada *node* selanjutnya akan diisi oleh atribut-atribut yang bernilai *gain* lebih rendah dan begitu seterusnya. *Node* yang tidak memiliki percabangan akan menunjukkan *output* dari setiap cabangnya yang dikenal dengan nama *leaf* atau daun.

Pada Tabel 4.2 terlihat bahwa atribut IPK4 memiliki nilai *gain* tertinggi diantara atribut yang lainnya, maka atribut IPK4 akan menjadi *root* pohon keputusan. Kemudian dilanjutkan dengan atribut status pekerjaan, asal sekolah dan gaji orangtua. Diakhiri oleh status yang menyatakan keterangan lulus tepat waktu, lulus terlambat, *drop out*. Setalah semua *gain* diperoleh maka akan terbentuk pohon keputusan seperti yang terlihat pada Gambar 4.1.

Gambar 4.1
Decision Tree pada aplikasi Weka 3.9

Dari gambar pohon keputusan pada Gambar 4.1. Atribut IPK4 sebagai *root node*, sedangkan atribut lainnya sebagai *child node*. Atribut kelas tidak masuk dalam pohon keputusan, artinya atribut kelas tidak mempengaruhi pola kelulusan mahasiswa. Atribut yang membentuk pohon keputusan adalah IPK4, status pekerjaan, program studi, jenis kelamin, gaji ortu, asal sekolah dan asal daerah.

Rule atau aturan yang dihasilkan dari 977 sampel *training* menghasilkan 8 *rule* lulus tepat waktu, 12 *rule* lulus terlambat dan 6 *rule drop out.* Aturan ini digunakan pada proses *testing* untuk memprediksi kelulusan, berikut keterangannya :

IPK4 = Besar
| STATUS PEKERJAAN = Tidak Bekerja
| | JK = P: Lulus Tepat Waktu (74.0/21.0)
| | JK = L
| | | ASAL SEKOLAH = SMA: Lulus Tepat Waktu (25.0/13.0)
| | | ASAL SEKOLAH = MAN: Lulus Tepat Waktu (1.0)
| | | ASAL SEKOLAH = SMK: Lulus Terlambat (15.0/5.0)
| | | ASAL SEKOLAH = PAKET C: Lulus Terlambat (0.0)
| STATUS PEKERJAAN = Bekerja
| | Program Studi = Komputerisasi Akuntansi
| | | JK = P: Lulus Terlambat (24.0/8.0)
| | | JK = L: Drop Out (9.0/4.0)
| | Program Studi = Manajemen Informatika: Lulus Tepat Waktu (32.0/17.0)
| | Program Studi = Teknik Informatika
| | | JK = P: Lulus Tepat Waktu (2.0)
| | | JK = L: Lulus Terlambat (13.0/6.0)
| | Program Studi = SISTEM INFORMASI
| | | JK = P: Lulus Terlambat (18.0/8.0)
| | | JK = L: Lulus Tepat Waktu (9.0/3.0)
IPK4 = Kecil: Drop Out (392.0/4.0)
IPK4 = Sedang

| STATUS PEKERJAAN = Tidak Bekerja

| | Program Studi = Komputerisasi Akuntansi: Lulus Tepat Waktu (44.0/19.0)

| | Program Studi = Manajemen Informatika

| | | GAJI ORTU = Sedang

| | | | Asal Daerah = Cirebon: Lulus Terlambat (11.0/2.0)

| | | | Asal Daerah = Luar Cirebon: Drop Out (4.0)

| | | GAJI ORTU = Rendah : Lulus Terlambat (0.0)

| | | GAJI ORTU = Tinggi : Lulus Tepat Waktu (6.0/1.0)

| | Program Studi = Teknik Informatika: Lulus Terlambat (35.0/11.0)

| | Program Studi = SISTEM INFORMASI: Lulus Terlambat (20.0/6.0)

| STATUS PEKERJAAN = Bekerja

| | Program Studi = Komputerisasi Akuntansi

| | | GAJI ORTU = Sedang : Lulus Terlambat (71.0/22.0)

| | | GAJI ORTU = Rendah : Drop Out (17.0)

| | | GAJI ORTU = Tinggi : Lulus Terlambat (1.0)

| | Program Studi = Manajemen Informatika

| | | JK = P: Lulus Terlambat (26.0/10.0)

| | | JK = L: Drop Out (58.0/31.0)

| | Program Studi = Teknik Informatika: Drop Out (37.0/16.0)

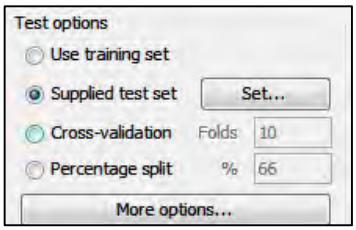
| | Program Studi = SISTEM INFORMASI: Lulus Terlambat (33.0/16.0)

4.1.1 Pengujian Model Pertama

Pengujian model menggunakan data *testing*, pengujian ini dilakukan untuk mengukur sejauh mana *classifier* melakukan klasifikasi dengan benar. Pengujian model menggunakan alat bantu yaitu tools Weka 3.9. Proses pengujian pada dasarnya membandingkan hasil prediksi dengan data sesungguhnya. Proses prediksi pada aplikasi Weka dilakukan melalui tahapan sebagai berikut :

1. Atribut yang ada pada data testing harus sama seperti yang ada pada data training, data *testing* disimpan dengan format .csv.

2. Buka aplikasi Weka, pilih menu Classify, ada 4 pilihan untuk melakukan uji *testing,* seperti yang terlihat pada Gambar 4.2, pada penelitian ini pengujian *data testing* mennggunakan pilihan *supplied test set,* kemudian pilih set untuk import data *testing* yang berformat .csv.



Gambar 4.2
Test Options pada aplikasi Weka 3.9

3. Setelah berhasil diinputkan, tiap *record* dari atribut data *testing* akan dicocokkan dengan *rule* atau aturan yang terbentuk saat proses perhitungan data *training.* Tampilan data *testing* yang telah diuji dapat dilihat pada Gambar 4.3. Kolom *predicted* status merupakan kolom hasil prediksi.

File Edit View

tes.arff

Relation: TRAININGSETPAKAI KELAS_predicted

No.	1: Program Studi	2: JK	3: Asal Daerah	4: ASAL SEKOLAH	5: IPK4	6: GAJI ORTU	7: STATUS PEKERJAAN	8: KELAS	9: prediction margin	10: predicted STATUS	11: **STATUS**
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Numeric	Nominal	Nominal
1	Komputerisasi Akuntansi	L	Cirebon	SMA	Besar	Sedang	Tidak Bekerja	PAGI	0.12	Lulus Tepat Waktu	Lulus Tepat Waktu
2	Komputerisasi Akuntansi	P	Cirebon	SMK	Kecil	Sedang	Bekerja	SORE	0.979592	Drop Out	Drop Out
3	Komputerisasi Akuntansi	P	Cirebon	SMA	Sedang	Sedang	Bekerja	PAGI	0.450704	Lulus Terlambat	Lulus Terlambat
4	Komputerisasi Akuntansi	L	Cirebon	SMA	Sedang	Tinggi	Bekerja	PAGI	1.0	Lulus Terlambat	Lulus Terlambat
5	Komputerisasi Akuntansi	P	Cirebon	SMA	Sedang	Rendah	Bekerja	PAGI	-1.0	Drop Out	Lulus Terlambat
6	Komputerisasi Akuntansi	L	Cirebon	SMA	Sedang	Sedang	Bekerja	PAGI	0.450704	Lulus Terlambat	Lulus Terlambat
7	Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Rendah	Bekerja	PAGI	0.416667	Lulus Terlambat	Lulus Terlambat
8	Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Rendah	Bekerja	PAGI	0.416667	Lulus Terlambat	Lulus Terlambat
9	Komputerisasi Akuntansi	P	Cirebon	SMK	Sedang	Sedang	Bekerja	PAGI	0.450704	Lulus Terlambat	Lulus Terlambat
10	Komputerisasi Akuntansi	P	Cirebon	SMA	Kecil	Sedang	Bekerja	PAGI	0.979592	Drop Out	Drop Out
11	Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Sedang	Tidak Bekerja	PAGI	0.459459	Lulus Tepat Waktu	Lulus Tepat Waktu
12	Komputerisasi Akuntansi	P	Cirebon	SMK	Besar	Rendah	Bekerja	PAGI	0.416667	Lulus Terlambat	Lulus Terlambat
13	Komputerisasi Akuntansi	P	Cirebon	SMK	Kecil	Sedang	Tidak Bekerja	PAGI	0.979592	Drop Out	Drop Out
14	Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Rendah	Bekerja	PAGI	0.416667	Lulus Terlambat	Lulus Terlambat
15	Komputerisasi Akuntansi	L	Luar Cirebon	MAN	Kecil	Sedang	Bekerja	PAGI	0.979592	Drop Out	Drop Out
16	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Besar	Sedang	Tidak Bekerja	SORE	-0.459459	Lulus Tepat Waktu	Lulus Terlambat
17	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Sedang	Rendah	Bekerja	SORE	-1.0	Drop Out	Lulus Terlambat
18	Komputerisasi Akuntansi	P	Cirebon	SMA	Kecil	Sedang	Tidak Bekerja	SORE	0.979592	Drop Out	Drop Out
19	Komputerisasi Akuntansi	L	Cirebon	SMK	Kecil	Rendah	Bekerja	SORE	0.979592	Drop Out	Drop Out
20	Komputerisasi Akuntansi	P	Cirebon	SMK	Sedang	Sedang	Bekerja	SORE	0.450704	Lulus Terlambat	Lulus Terlambat
21	Komputerisasi Akuntansi	L	Cirebon	SMA	Kecil	Sedang	Bekerja	PAGI	0.979592	Drop Out	Drop Out
22	Komputerisasi Akuntansi	P	Luar Cirebon	PAKET C	Kecil	Sedang	Tidak Bekerja	PAGI	0.979592	Drop Out	Drop Out
23	Komputerisasi Akuntansi	P	Luar Cirebon	SMK	Sedang	Sedang	Bekerja	PAGI	0.450704	Lulus Terlambat	Lulus Terlambat
24	Komputerisasi Akuntansi	L	Luar Cirebon	PAKET C	Sedang	Tinggi	Tidak Bekerja	PAGI	-0.295455	Lulus Tepat Waktu	Lulus Terlambat
25	Komputerisasi Akuntansi	P	Cirebon	SMK	Sedang	Rendah	Bekerja	PAGI	-1.0	Drop Out	Lulus Terlambat
26	Komputerisasi Akuntansi	P	Cirebon	SMK	Besar	Sedang	Tidak Bekerja	PAGI	-0.459459	Lulus Tepat Waktu	Lulus Terlambat
27	Komputerisasi Akuntansi	P	Luar Cirebon	SMK	Sedang	Rendah	Bekerja	PAGI	-1.0	Drop Out	Lulus Terlambat
28	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Sedang	Sedang	Tidak Bekerja	PAGI	-0.295455	Lulus Tepat Waktu	Lulus Terlambat
29	Komputerisasi Akuntansi	P	Cirebon	MAN	Kecil	Sedang	Tidak Bekerja	PAGI	0.979592	Drop Out	Drop Out
30	Manajemen Informatika	L	Cirebon	SMK	Sedang	Sedang	Tidak Bekerja	SORE	-0.818182	Lulus Terlambat	Lulus Tepat Waktu
31	Manajemen Informatika	P	Cirebon	SMK	Besar	Sedang	Tidak Bekerja	SORE	0.459459	Lulus Tepat Waktu	Lulus Tepat Waktu
32	Manajemen Informatika	P	Cirebon	SMK	Sedang	Tinggi	Tidak Bekerja	SORE	0.666667	Lulus Tepat Waktu	Lulus Tepat Waktu

Gambar 4.3
Hasil Pengujian Data *Testing* Pada Weka 3.9

Dari pengujian tersebut dapat diketahui tingkat kebenaran prediksi, terlihat pada Tabel 4.3.

Tabel 4.3 Tingkat Kebenaran Prediksi

Jumlah *Rule*	Jumlah Data *Testing*	Prediksi Benar	Prediksi Salah
14	130	109	21

4.1.2 Evaluasi Model Pertama

Evaluasi dilakukan dengan menganalisa hasil klasifikasi. Pengukuran data dilakukan dengan *confusion matrix*. Evaluasi pengukuruan ini membandingkan nilai akurasi dan *error rate* algoritma *decision tree* C4.5. Model *confusion matrix* 3x3 ditunjukkan pada Tabel 4.4.

Tabel 4.4 Model *Counfusion Matrix*

		Predict Class		
		Class A	*Class* B	*Class* C
Actual Class	*Class* A	AA	AB	AC
	Class B	BA	BB	BC
	Class C	CA	CB	CC

Dengan bantuan tools Weka, maka di dapatkan tabel *confusion matrix* untuk metode C4.5 seperti yang ditunjukkan oleh Gambar 4.4

```
=== Confusion Matrix ===

  a  b  c   <-- classified as
 28  7  0 |  a = Lulus Tepat Waktu
  8 42  5 |  b = Lulus Terlambat
  0  1 39 |  c = Drop Out
```

Gambar 4.4
Confusion Matrix Metode *Decision Tree* C4.5

Perhitungan akurasi dengan tabel *confusion matrix* adalah sebagai berikut:

$$\text{Akurasi} = \frac{AA+BB+CC}{AA+AB+AC+BA+BB+BC+CA+CB+CC} \times 100\% \quad \text{..... (3)}$$

$$\text{Akurasi} = \frac{28+42+39}{28+7+0+8+42+5+0+1+39} \times 100\%$$

$$= \frac{109}{130} \times 100\%$$

$$= 0{,}84 \times 100\%$$

$$= 84\%$$

$$\textit{Error Rate} = \frac{AB+AC+BA+BC+CA+CB}{AA+AB+AC+BA+BB+BC+CA+CB+CC} \times 100\% \quad \text{..... (4)}$$

$$\textit{Error Rate} = \frac{7+0+8+5+0+1}{28+7+0+8+42+50+1+39} \times 100\%$$

$$= \frac{21}{130} \times 100\%$$

$$= 0{,}16 \times 100\%$$

$$= 16\%$$

Tabel 4.5 Tabel Hasil Akurasi dan *Error Rate*

Dataset	Nilai Akurasi	Nilai *Error Rate*
Data *Testing*	84%	16%

4.2 *Modelling* Algoritma C4.5 Kedua

Pada uji *modelling* yang pertama, telah diketahui bahwa IPK4 merupakan *root* pohon keputusan karena memiliki nilai *gain* tertinggi diantara atribut-atribut lainnya. IPK 4 menjadi atribut utama yang digunakan dalam prediksi kelulusan mahasiswa.

Pada *modeling* yang kedua, proses prediksi kelulusan akan dilakukan berdasarkan Indeks prestasi per semester. Atribut IPK4 akan di *breakdown* menjadi empat kategori yaitu IP semester 1, 2, 3 dan 4. Atribut IP semester 1 sampai dengan 4 akan dijadikan atribut untuk mempredikasi kelulusan, sehingga akan diketahui IP semester berapa saja yang dapat memprediksi kelulusan dan IP semester berapa saja yang tidak masuk dalam pola prediksi kelulusan.

Tujuan dari proses *modelling* kedua ini adalah untuk menemukan model yang terbaik dan menegaskan hasil prediksi pada model pertama. IPK 4 diketahui merupakan atribut utama yang dapat memprediksi kelulusan, artinya prediksi dapat dilakukan setelah mahasiswa mengikuti perkuliahan sampai dengan empat semester. Pada model kedua akan dicari apakah sudah tepat bahwa prediksi dapat dilakukan di tahun kedua perkuliahan dan IP semester berapa saja yang harus di perhatikan untuk dapat memprediksi kelulusan, pola ini akan sangat membantu untuk mengurangi jumlah *drop out* dan lulus terlambat. Pola atau aturan yang dihasilkan dari model kedua ini akan digabung dengan pola yang terbentuk pada model pertama.

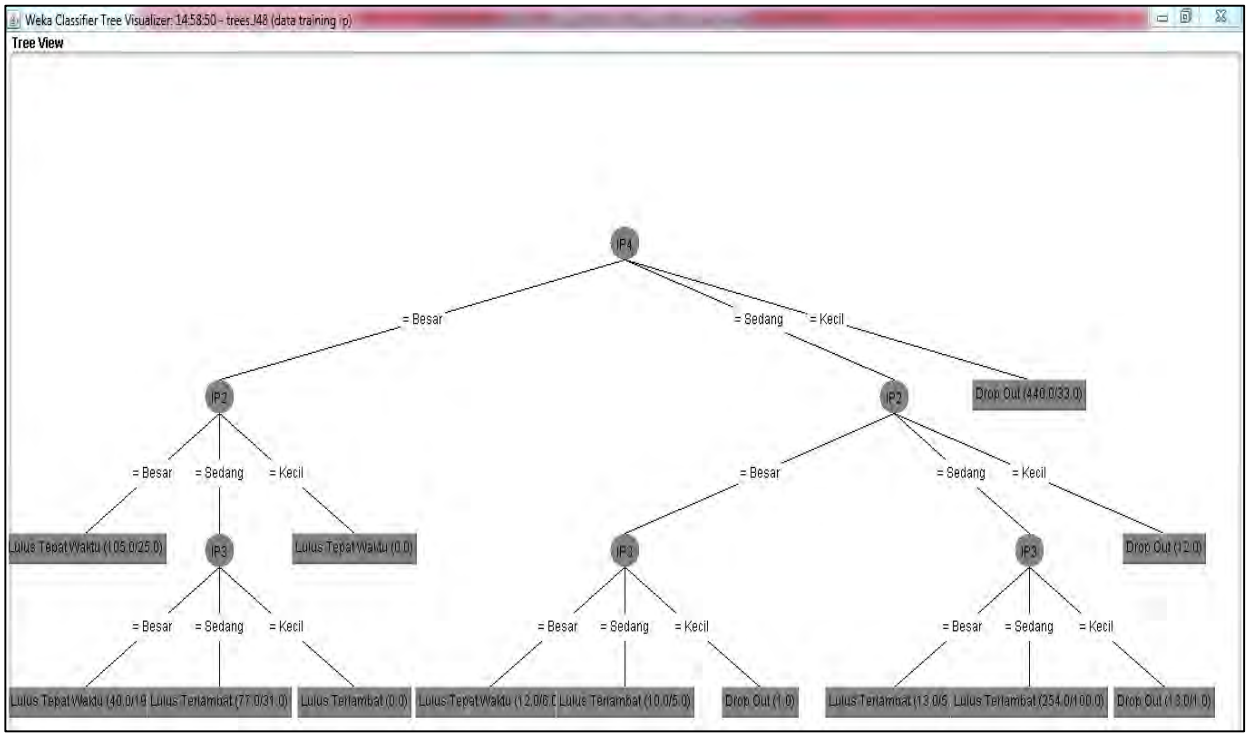
Record Data *training* yang digunakan pada proses *modelling* yang kedua merupakan data yang sama dengan data *training* sebelumnya (*modellingi* pertama),

hanya atributnya saja yang berbeda. Jumlah data *training* yang digunakan pada *modelling* kedua adalah sebanyak 977. Cuplikan data dapat dilihat pada Tabel 4.6.

Tabel 4.6 Cuplikan Data *Training* Untuk *Modelling* Yang Kedua

IPS1	IPS2	IPS3	IPS4	STATUS
Sedang	Besar	Besar	Besar	Lulus Tepat Waktu
Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
Sedang	Besar	Besar	Besar	Lulus Tepat Waktu
Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
Sedang	Sedang	Besar	Besar	Lulus Tepat Waktu
Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
Sedang	Sedang	Kecil	Kecil	Drop Out
Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
Sedang	Kecil	Kecil	Sedang	Drop Out
Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
Kecil	Kecil	Kecil	Kecil	Drop Out
Sedang	Sedang	Kecil	Kecil	Drop Out
Sedang	Sedang	Kecil	Kecil	Drop Out
Sedang	Sedang	Kecil	Kecil	Drop Out
Sedang	Sedang	Kecil	Kecil	Drop Out

Modelling algoritma C4.5 untuk data *training* diatas menggunakan alat bantu yaitu tools Weka 3.9. dengan menggunakan tools Weka maka dapat langsung terbentuk pohon keputusan seperti yang terlihat pada Gambar 4.5.



Gambar 4.5
Decision Tree C4.5 Pada *Modelling* yang Kedua menggunakan Weka 3.9

Dari gambar pohon keputusan pada Gambar 4.6, atribut IP4 (IP semester 4) sebagai *root node*, sedangkan atribut lainnya sebagai *child node*. Atribut IP1 tidak masuk dalam pohon keputusan, artinya atribut IP1 tidak mempengaruhi kelulusan mahasiswa. Urutan atribut yang membentuk pohon keputusan adalah IP semester 4, 2 dan 3.

Rule atau aturan yang dihasilkan dari 977 sampel *training* menghasilkan 8 *rule* lulus tepat waktu, 12 *rule* lulus terlambat dan 6 *rule drop out.* Aturan ini digunakan pada proses *testing* untuk memprediksi kelulusan, berikut keterangannya :

IP4 = Besar
| IP2 = Besar: Lulus Tepat Waktu (105.0/25.0)
| IP2 = Sedang
| | IP3 = Besar: Lulus Tepat Waktu (40.0/19.0)
| | IP3 = Sedang: Lulus Terlambat (77.0/31.0)
| | IP3 = Kecil: Lulus Terlambat (0.0)
| IP2 = Kecil: Lulus Tepat Waktu (0.0)
IP4 = Sedang
| IP2 = Besar
| | IP3 = Besar: Lulus Tepat Waktu (12.0/6.0)
| | IP3 = Sedang: Lulus Terlambat (10.0/5.0)
| | IP3 = Kecil: Drop Out (1.0)
| IP2 = Sedang
| | IP3 = Besar: Lulus Terlambat (13.0/5.0)
| | IP3 = Sedang: Lulus Terlambat (254.0/100.0)
| | IP3 = Kecil: Drop Out (13.0/1.0)
| IP2 = Kecil: Drop Out (12.0)
IP4 = Kecil: Drop Out (440.0/33.0)

4.2.1 Pengujian Model Kedua

Pengujian model kedua ini menggunakan data *testing* yang sama dengan model pertama, hanya atribut saja yang berbeda. Pengujian ini dilakukan untuk melihat nilai akurasi dan *error* dari aturan atau *rule* yang dihasilkan pada proses *modelling*

sebelumnya. Pengujian model menggunakan tools Weka 3.9. Hasil pengujian model kedua terlihat pada Gambar 4.6.

File Edit View

datatestingvalid.arff

Relation: data training ip_predicted

No.	1: IP1 Nominal	2: IP2 Nominal	3: IP3 Nominal	4: IP4 Nominal	5: prediction margin Numeric	6: predicted STATUS Nominal	7: **STATUS** Nominal
1	Besar	Besar	Besar	Besar	0.619048	Lulus Tepat Waktu	Lulus Tepat Waktu
2	Sedang	Sedang	Sedang	Sedang	-0.314961	Lulus Terlambat	Drop Out
3	Sedang	Sedang	Sedang	Sedang	0.314961	Lulus Terlambat	Lulus Terlambat
4	Sedang	Sedang	Sedang	Sedang	0.314961	Lulus Terlambat	Lulus Terlambat
5	Sedang	Sedang	Besar	Besar	-0.15	Lulus Tepat Waktu	Lulus Terlambat
6	Sedang	Sedang	Sedang	Besar	0.324675	Lulus Terlambat	Lulus Terlambat
7	Sedang	Sedang	Sedang	Besar	0.324675	Lulus Terlambat	Lulus Terlambat
8	Sedang	Sedang	Besar	Sedang	0.384615	Lulus Terlambat	Lulus Terlambat
9	Sedang	Sedang	Sedang	Besar	0.324675	Lulus Terlambat	Lulus Terlambat
10	Sedang	Kecil	Kecil	Kecil	0.859091	Drop Out	Drop Out
11	Sedang	Besar	Besar	Besar	0.619048	Lulus Tepat Waktu	Lulus Tepat Waktu
12	Sedang	Besar	Besar	Besar	0.619048	Lulus Tepat Waktu	Lulus Tepat Waktu
13	Sedang	Sedang	Kecil	Kecil	0.859091	Drop Out	Drop Out
14	Besar	Besar	Besar	Besar	0.619048	Lulus Tepat Waktu	Lulus Tepat Waktu
15	Sedang	Sedang	Kecil	Kecil	0.859091	Drop Out	Drop Out
16	Besar	Besar	Sedang	Besar	0.619048	Lulus Tepat Waktu	Lulus Tepat Waktu

Gambar 4.6
Hasil Uji Data *Testing* pada Model Kedua menggunakan Weka 3.9

Dari pengujian tersebut dapat diketahui tingkat kebenaran prediksi, terlihat pada Tabel 4.7.

Tabel 4.7 Tingkat Kebenaran Prediksi

Jumlah *Rule*	Jumlah Data *Testing*	Prediksi Benar	Prediksi Salah
11	130	100	30

4.2.2 Evaluasi Model Kedua

Evaluasi dilakukan dengan menganalisa hasil klasifikasi. Pengukuran data dilakukan dengan *confusion matrix*. Evaluasi pengukuruan ini membandingkan nilai akurasi dan *error rate* algoritma *decision tree* C4.5 dengan model *confusion matrix.*

Dengan bantuan tools Weka 3.9, maka di dapatkan tabel *confusion matrix* untuk metode C4.5 seperti yang ditunjukkan oleh Gambar 4.7.

```
=== Confusion Matrix ===

  a  b  c   <-- classified as
 31  6  0 |  a = Lulus Tepat Waktu
 13 38  2 |  b = Lulus Terlambat
  2  7 31 |  c = Drop Out
```

Gambar 4.7
Confusion Matrix Metode *Decision Tree* C4.5

Perhitungan akurasi dengan tabel *confusion matrix* adalah sebagai berikut:

$$\text{Akurasi} = \frac{AA+BB+CC}{AA+AB+AC+BA+BB+BC+CA+CB+CC} \times 100\% \quad \text{..... (3)}$$

$$\text{Akurasi} = \frac{31+38+31}{31+6+0+13+38+2+2+7+31} \times 100\%$$

$$= \frac{100}{130} \times 100\%$$

$$= 0{,}77 \times 100\%$$

$$= 77\%$$

$$\textit{Error Rate} = \frac{AB+AC+BA+BC+CA+CB}{AA+AB+AC+BA+BB+BC+CA+CB+CC} \times 100\% \quad \text{..... (4)}$$

$$\textit{Error Rate} = \frac{6+0+13+2+2+7}{31+6+0+13+38+2+2+7+31} \times 100\%$$

$$= \frac{30}{130} \times 100\%$$

$$= 0{,}23 \times 100\%$$

$$= 23\%$$

Tabel 4.8 Tabel Hasil Akurasi dan *Error Rate*

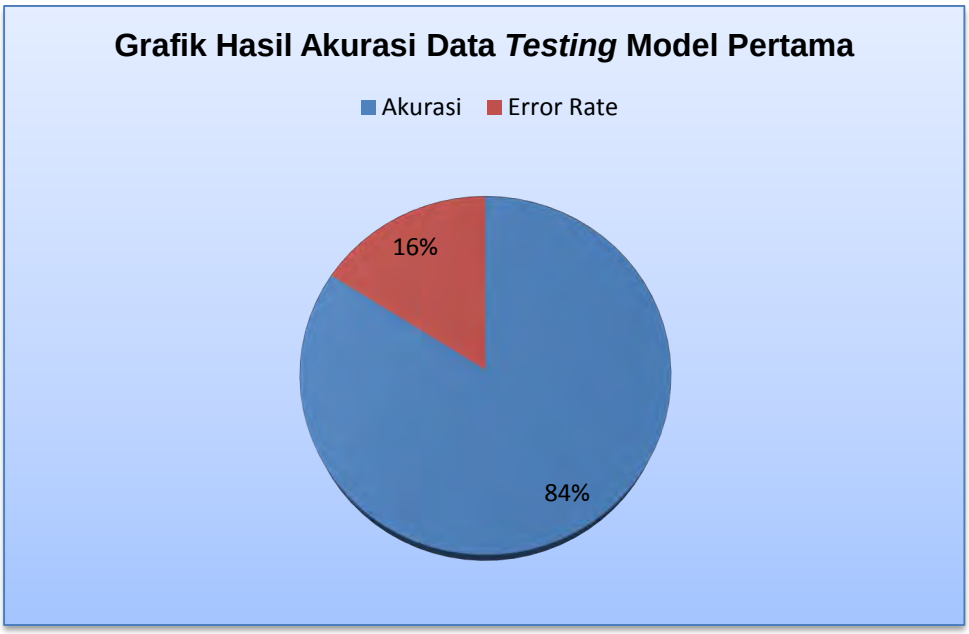
Dataset	Nilai Akurasi	Nilai *Error Rate*
Data *Testing*	77%	23%

4.3 Grafik Tingkat Akurasi

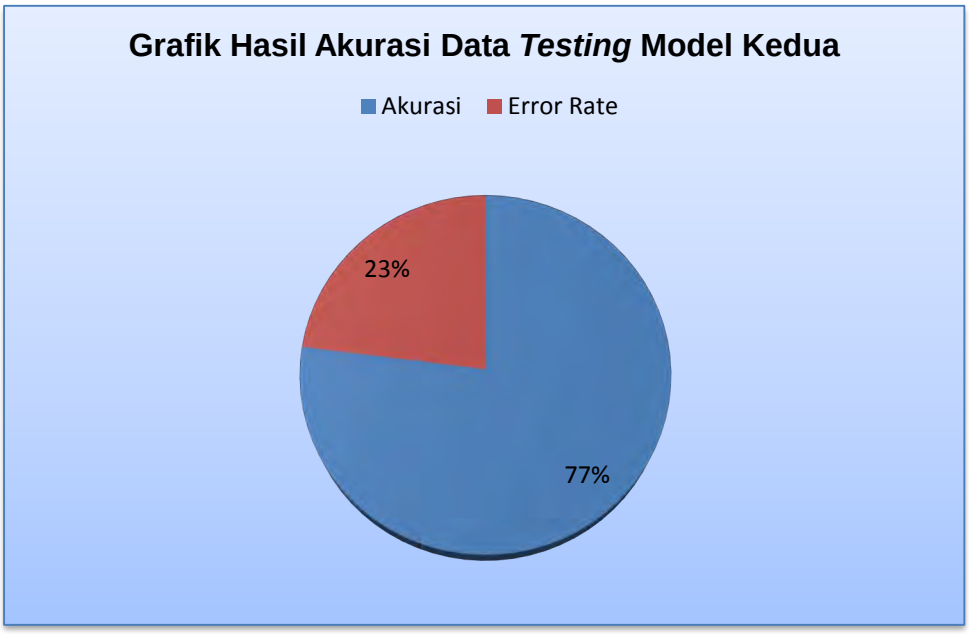
Grafik model pertama merupakan hasil uji dengan menggunakan atribut program studi, jenis kelamin, asal daerah, asal sekolah, IPK semester 4, gaji orangtua, status pekerjaan mahasiswa dan kelas.

Grafik model kedua merupakan hasil uji data *testing* dengan menggunakan atribut IP semester 1, 2, 3 dan 4. Masing-masing model menggunakan data *testing* yang sama.

Tampilan grafik hasil akurasi data *testing* model pertama dan kedua terlihat pada Gambar 4.8 dan 4.9.



Gambar 4.8
Grafik Tingkat Akurasi Data *Testing* Model Pertama



Gambar 4.9
Grafik Tingkat Akurasi Data *Testing* Model Kedua

4.4 Pola Pengetahuan

Pada model pertama, dari beberapa atribut yang diuji, ternyata IPK 4 adalah atribut utama dalam pola prediksi kelulusan. Selanjutnya pada model kedua, diketahui bahwa atribut utama pembentuk pohon keputusan adalah IP semester 4, sedangan IP semester 1 tidak masuk dalam pohon keputusan. Artinya pada model kedua IP yang sangat menentukan kelulusan adalah IP di semester 4.

Model pertama merupakan model yang akan digunakan untuk prediksi kelulusan karena setelah dilakukan pengujian pada model kedua dimana pengujian berdasarkan IP semester 1 sampai dengan 4 diketahui bahwa IP semester 4 yang menjadi atribut utama dalam pola prediksi kemudian diikuti oleh atribut IP semester 2 dan 3.

Dari uraian diatas, telah diperoleh pengetahuan baru dari pola prediksi kelulusan ini, yaitu untuk memprediksi lulus tepat waktu, lulus terlambat dan *drop out* tidak dapat dilakukan di tahun pertama perkuliahan, melainkan dapat dilakukan setelah mahasiswa menempuh perkuliahan sampai dengan semester 4. Nilai IPK semester 4 sangat tepat digunakan sebagai atribut utama dalam prediksi kelulusan, kemudian diikuti atribut lainnya yaitu : status pekerjaan, asal sekolah, jenis kelamin, program studi, gaji orangtua dan asal daerah. Berikut ini adalah pola prediksi kelulusan tepat waktu, terlambat dan *drop out* yang dihasilkan dari pohon keputusan pada model pertama :

Prediksi Lulus Tepat Waktu (TW)

1. Jika IPK 4 besar, tidak bekerja dan jenis kelaminnya perempuan maka lulus tepat waktu
2. Jika IPK 4 besar, tidak bekerja, jenis kelamin laki-laki dan asal sekolah SMA maka lulus tepat waktu
3. Jika IPK 4 besar, tidak bekerja, jenis kelamin laki-laki dan asal sekolah MAN maka lulus tepat waktu
4. Jika IPK 4 besar dan program studinya manajemen informatika maka lulus tepat waktu
5. Jika IPK4 besar, program studinya Teknik Informatika dan jenis kelaminnya perempuan maka lulus tepat waktu

6. Jika IPK 4 besar, program studinya sistem informasi dan jenis kelaminnya laki-laki maka lulus tepat waktu
7. Jika IPK 4 sedang, tidak bekerja dan program studinya komputerisasi akuntansi maka lulus tepat waktu
8. Jika IPK 4 sedang, tidak bekerja, program studinya manajemen informatika, gaji ortu tinggi maka lulus tepat waktu

Prediksi Lulus Terlambat (T)

1. Jika IPK 4 besar, tidak bekerja, jenis kelamin laki-laki dan asal sekolah SMK maka lulus terlambat
2. Jika IPK 4 besar, bekerja, program studinya komputerisasi akuntansi dan jenis kelaminnya perempuan maka lulus terlambat
3. Jika IPK4 besar, program studinya Teknik Informatika dan jenis kelaminnya laki-laki maka lulus terlambat
4. Jika IPK 4 besar, program studinya sistem informasi dan jenis kelaminnya perempuan maka lulus terlambat
5. Jika IPK 4 sedang, tidak bekerja, program studinya manajemen informatika, gaji ortu sedang dan asal daerah Cirebon maka lulus terlambat
6. Jika IPK 4 sedang, tidak bekerja, program studinya manajemen informatika, gaji ortu rendah maka lulus Terlambat
7. Jika IPK 4 sedang, tidak bekerja dan program studinya teknik informatika maka lulus terlambat
8. Jika IPK 4 sedang, tidak bekerja dan program studinya sistem informati maka lulus terlambat
9. Jika IPK 4 sedang, bekerja, program studinya komputerisasi akuntansi dan gaji ortu sedang maka lulus telambat
10. Jika IPK 4 sedang, bekerja, program studinya komputerisasi akuntansi dan gaji ortu tinggi maka lulus telambat
11. Jika IPK 4 sedang, bekerja, program studinya manajemen informatika dan jenis kelaminnya perempuan maka lulus terlambat

12. Jika IPK 4 sedang, bekerja, program studinya sistem informasi maka lulus terlambat

Prediksi *Drop Out* (DO)

1. Jika IPK 4 besar, bekerja, program studinya komputerisasi akuntansi dan jenis kelamin laki-laki maka drop out
2. Jika IPK 4 kecil maka drop out
3. Jika IPK 4 sedang, tidak bekerja, program studinya manajemen informatika, gaji ortu sedang dan asal daerah Luar Cirebon maka lulus drop out
4. Jika IPK 4 sedang, bekerja, program studinya komputerisasi akuntansi dan gaji ortu rendah maka drop out
5. Jika IPK 4 sedang, bekerja, program studinya manajemen informatika dan jenis kelaminnya laki-laki maka drop out
6. Jika IPK 4 sedang, bekerja, program studinya teknik informatika maka drop out

4.5 Hubungan Penelitian dengan STMIK WIT

Hasil penelitian yang telah dilakukan, memperlihatkan bahwa dengan adanya penerapan data mining untuk prediksi kelulusan mahasiswa ini nantinya akan memberikan perubahan dalam peningkatan mutu pendidikan. Berkaitan dengan hal tersebut maka diperlukan usaha yang nyata untuk mewujudkan perubahan yang lebih baik dalam dunia pendidikan khususnya di STMIK WIT.

Pada penelitian ini, penggunaan algoritma C4.5 mampu melakukan prediksi dengan baik (84%) terhadap masa studi mahasiswa yang tepat waktu, terlambat dan *drop out*. Pembentukan pohon keputusan (*Decision Tree*) dapat digunakan oleh pengelola akademik di dalam memetakan mahasiswa yang berpotensi mengalami keterlambatan masa studi dan *drop out* di masa mendatang. Hasi penelitian ini dapat digunakan sebagai langkah untuk menghindari penurunan kelulusan mahasiswa setiap tahunnya, penerapan data mining ini akan memberikan kemajuan dan kontribusi bagi STMIK WIT.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang dilakukan, dapat diambil kesimpulan sebagai berikut :

1 Data mining dengan algoritma C4.5 dapat diimplementasikan untuk memprediksi kelulusan mahasiswa STMIK WIT dengan tiga kategori yaitu lulus tepat waktu, lulus terlambat dan *drop out*. Variabel yang berpengaruh dalam hasil prediksi adalah IPK4, status pekerjaan, asal sekolah, asal daerah, jenis kelamin, gaji ortu dan program studi, sedangkan variabel yang tidak bepengaruh pada hasil prediksi adalah kelas.

2 Pada penelitian ini, penggunaan algoritma C4.5 mampu melakukan prediksi dengan baik (84%) terhadap masa studi mahasiswa yang tepat waktu, terlambat dan *drop out*. Pembentukan pohon keputusan (*Decision Tree*) dapat digunakan oleh pengelola akademik di dalam memetakan mahasiswa yang berpotensi mengalami keterlambatan masa studi dan *drop out* di masa mendatang. Penerapan *Educational Data Mining* (EDM) memberikan kemajuan dan kontribusi pada dunia pendidikan dan pada bidang riset data mining.

3 Dengan alat bantu Weka, hasil uji coba menggunakan 130 data *testing,* pola yang dibentuk pada model pertama memiliki nilai akurasi kecocokan sebesar 84%, sedangkan model kedua 77%.

4 Pola atau aturan yang dihasilkan dari model pertama dengan menggunakan 977 data *training* menghasilkan 8 *rule* lulus tepat waktu, 12 *rule* lulus terlambat dan 6 *rule drop out.*

5.2 Saran

Beberapa sara dari penulis untuk pengembangan penelitian lebih lanjut yaitu:

1. Untuk penelitian selanjutnya dapat menambah variabel lain, selain dari variabel yang dilakukan peneliti, seperti pendidikan orangtua dan jumlah anggota keluarga
2. Membandingkan algoritma C4.5 dengan algoritma lainnya seperti *k-nearest neighbor* dan *neural network.*
3. Pengklasifikasian terhadap data mahasiswa STMIK WIT sebaiknya dilakukan secara rutin setiap tahun sebagai langkah preventif untuk menghindari penurunan kelulusan mahasiswa setiap tahunnya.

DAFTAR PUSTAKA

Adhatrao, Kalpesh, Aditya Gaykar, Amiraj Dhawan, Rohit Jha and Vipul Honrao. *Predicting Students' Performance Using Id3 And C4.5 Classification Algorithms. International Journal of Data Mining & Knowledge Management Process.* 2013.

Al Fatta, Hanif. *Analisa dan Perancangan Sistem Informasi.* Andi Offset, Yogyakarta, 2007.

Bahar, Hasbul. Prediksi Lulus Tepat Dan Tidak Tepat Waktu Mahasiswa Menggunakan Algoritma K-Means. Jurnal Teknik Informatika, Vol 6. No. 02. 2014.

Bhardwaj, Brijesh Kumar & Saurabh Pal. *Data Mining: A prediction for performance improvement using classification. International Journal of Computer Science and Information Security.* 2011

Fithri, D.L. dan Eko Darmanto. Sistem Pendukung Keputusan Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Naïve Bayes. Prosiding Snatif ke-1. 2014.

Han,J. and Kamber, M., "Data Mining: Concepts and Techniques", 2nd edition. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor, 2006.

Indah Puji Astuti. Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4.5. Fountain Of Informatics Journal. Vol 2. No. 2. 2017

Iskandar, Derick dan Yoyon K Suprapto. Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan Antara Algoritma C4.5 Dan Naïve Bayes. Jurnal Ilmiah Nero Vo. xxx. No. xxx. 2014

Kadir, Abdul. *Pengantar Sistem Informasi edisi Revisi*. Andi Offset, Yogyakarta, 2013.

Kamagi, D. H. dan Seng Hansun. Implementasi Data Mining dengan Algoritma C4.5 Untuk Memprediksi Tingkat Kelulusan Mahasiswa. ULTIMATICS, Vol VI No. 1. 2014.

Karim, A. *Pemodelan aturan dalam memprediksi Prestasi akademik mahasiswa politeknik Poliprofesi medan dengan kernel k-means Clustering.* Jurnal Eksplora informatika vol. 3, no. 2, 2014.

Kusrini & Luthfi Taufiq Emha. Algoritma Data Mining. Andi Offset, Yogyakarta, 2009.

Larose T. Daniel, "Descovering Knowledge in Data: An Introducing to Data Mining", John Willey & Sons, Inc. 2005

Meilani, B.D. dan Susanti, Nofi. *Aplikasi Data Mining Untuk Menghasilkan Pola Kelulusan Siswa Dengan Metode Naïve Bayes.* Jurnal link vol 21 no. 2, 2014.

Mustafa, M.S. dan I Wayan Simpen. Perancangan Aplikasi Prediksi Kelulusan Mahasiswa Baru Dengan Teknik Data Mining (studi Kasus : STMIK Dipanegara Makasar). Citec Journal,Vol.1 No. 2. 2014.

Nugraha, P.G. Surya Cipta dkk. Penerapan Metode *Decision Tree*(Data Mining) Untuk Memprediksi Tingkat Kelulusan Siswa Smpn1 Kintamani. Seminar Nasional Vokasi dan Teknologi. 2016

Nurjoko dan Hendra Kurniawan. Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Apriori di Ibi Darmajaya Bandar Lampung. Jurnal Tim Darmajaya. Vol 02 No. 01. 2016.

Putra, Ade. Solusi Prediksi Mahasiswa Drop Out Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Bina Darma. Jurnal Simetris, Vol 8 No. 1. 2017.

Rusdiana dan Sam'ani. Pemodelan K-means Pada Penentuan Predikat Kelulusan Mahasiswa STMIK Palangkaraya. Jurnal Saintekom,Vol 6, No. 1. 2016

Saefulloh, Asep dan Moedjiono. Penerapan Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Tepat Waktu. Infosys Journal, Vol 2 No. 1. 2013

Santoso, Sani dan Dedy Suryadi. *Pengantar Data Mining.* Andi Offset,Yogyakarta, 2010.

Swastina, Liliana. Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa. Jurnal Gema Aktualita, Vol 2 No. 1. 2013.

LAMPIRAN 1

Objek Penelitian

LAMPIRAN 2

Tabel Data *Training* Model Pertama

No	Program Studi	JK	Asal Daerah	Asal Sekolah	IPK4	Gaji Ortu	Status Pekerjaan	Kelas	Status
1	Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Sedang	Tidak Bekerja	SORE	Lulus Tepat Waktu
2	Komputerisasi Akuntansi	L	Cirebon	MAN	Kecil	Sedang	Bekerja	SORE	Lulus Terlambat
3	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Besar	Rendah	Bekerja	PAGI	Lulus Terlambat
4	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Besar	Sedang	Tidak Bekerja	SORE	Lulus Tepat Waktu
5	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Sedang	Tinggi	Bekerja	PAGI	Lulus Terlambat
6	Komputerisasi Akuntansi	L	Luar Cirebon	SMK	Sedang	Sedang	Bekerja	PAGI	Lulus Terlambat
7	Komputerisasi Akuntansi	P	Luar Cirebon	PAKET C	Sedang	Sedang	Bekerja	PAGI	Lulus Terlambat
8	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Sedang	Sedang	Bekerja	PAGI	Drop Out
9	Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Rendah	Bekerja	PAGI	Drop Out
10	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Kecil	Sedang	Bekerja	PAGI	Drop Out
11	Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Rendah	Bekerja	PAGI	Drop Out
12	Komputerisasi Akuntansi	L	Luar Cirebon	SMA	Besar	Rendah	Bekerja	PAGI	Drop Out
13	Komputerisasi Akuntansi	P	Luar Cirebon	SMK	Besar	Sedang	Tidak Bekerja	PAGI	Lulus Tepat Waktu
14	Komputerisasi Akuntansi	L	Luar Cirebon	SMA	Sedang	Tinggi	Tidak Bekerja	SORE	Lulus Terlambat
15	Komputerisasi Akuntansi	P	Luar Cirebon	SMK	Sedang	Sedang	Bekerja	SORE	Lulus Terlambat
16	Komputerisasi Akuntansi	L	Luar Cirebon	SMK	Kecil	Rendah	Bekerja	SORE	Drop Out
17	Komputerisasi Akuntansi	P	Luar Cirebon	SMK	Sedang	Sedang	Bekerja	SORE	Lulus Terlambat
18	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Sedang	Sedang	Bekerja	PAGI	Drop Out
19	Komputerisasi Akuntansi	L	Luar Cirebon	SMK	Besar	Rendah	Bekerja	PAGI	Drop Out
20	Komputerisasi Akuntansi	P	Cirebon	MAN	Sedang	Sedang	Bekerja	PAGI	Lulus Terlambat
21	Komputerisasi Akuntansi	L	Cirebon	SMK	Sedang	Sedang	Bekerja	PAGI	Lulus Terlambat
22	Komputerisasi Akuntansi	P	Cirebon	SMK	Kecil	Sedang	Bekerja	PAGI	Drop Out
23	Komputerisasi Akuntansi	L	Cirebon	PAKET C	Kecil	Sedang	Bekerja	PAGI	Drop Out
24	Komputerisasi Akuntansi	L	Luar Cirebon	SMK	Sedang	Rendah	Bekerja	PAGI	Drop Out
25	Komputerisasi Akuntansi	L	Cirebon	SMK	Kecil	Rendah	Bekerja	SORE	Drop Out
26	Komputerisasi Akuntansi	L	Cirebon	SMA	Kecil	Sedang	Bekerja	SORE	Drop Out
27	Komputerisasi Akuntansi	L	Cirebon	MAN	Kecil	Sedang	Bekerja	SORE	Drop Out
28	Komputerisasi Akuntansi	L	Cirebon	SMA	Kecil	Sedang	Bekerja	PAGI	Drop Out
29	Komputerisasi Akuntansi	P	Cirebon	SMA	Kecil	Rendah	Bekerja	SORE	Drop Out
30	Komputerisasi Akuntansi	L	Cirebon	SMA	Kecil	Sedang	Bekerja	PAGI	Drop Out
31	Komputerisasi Akuntansi	P	Cirebon	SMA	Sedang	Rendah	Tidak Bekerja	SORE	Lulus Terlambat
32	Komputerisasi	L	Cirebon	SMK	Kecil	Sedang	Tidak	SORE	Drop Out

No	Program Studi	JK	Asal Daerah	Asal Sekolah	IPK4	Gaji Ortu	Status Pekerjaan	Kelas	Status
	Akuntansi						Bekerja		
33	Komputerisasi Akuntansi	L	Cirebon	SMK	Kecil	Rendah	Tidak Bekerja	SORE	Drop Out
34	Komputerisasi Akuntansi	P	Cirebon	SMK	Sedang	Tinggi	Tidak Bekerja	PAGI	Drop Out
35	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Sedang	Sedang	Bekerja	PAGI	Lulus Terlambat
36	Komputerisasi Akuntansi	P	Cirebon	SMA	Kecil	Sedang	Tidak Bekerja	PAGI	Drop Out
37	Komputerisasi Akuntansi	P	Cirebon	MAN	Kecil	Sedang	Bekerja	PAGI	Drop Out
38	Komputerisasi Akuntansi	L	Cirebon	SMA	Kecil	Sedang	Tidak Bekerja	PAGI	Drop Out
39	Komputerisasi Akuntansi	L	Cirebon	MAN	Sedang	Rendah	Tidak Bekerja	PAGI	Lulus Terlambat
40	Komputerisasi Akuntansi	L	Cirebon	SMA	Besar	Rendah	Bekerja	SORE	Drop Out
41	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Sedang	Rendah	Bekerja	SORE	Drop Out
42	Komputerisasi Akuntansi	L	Cirebon	SMK	Kecil	Sedang	Bekerja	SORE	Drop Out
43	Komputerisasi Akuntansi	P	Cirebon	SMK	Kecil	Sedang	Bekerja	SORE	Drop Out
44	Komputerisasi Akuntansi	L	Cirebon	SMK	Kecil	Sedang	Bekerja	SORE	Drop Out
45	Komputerisasi Akuntansi	P	Luar Cirebon	SMK	Sedang	Rendah	Bekerja	SORE	Drop Out
46	Komputerisasi Akuntansi	P	Cirebon	SMK	Kecil	Sedang	Bekerja	SORE	Drop Out
47	Komputerisasi Akuntansi	P	Cirebon	SMK	Kecil	Sedang	Bekerja	SORE	Drop Out
48	Komputerisasi Akuntansi	P	Luar Cirebon	SMK	Sedang	Rendah	Bekerja	PAGI	Drop Out
49	Komputerisasi Akuntansi	P	Cirebon	PAKET C	Sedang	Tinggi	Tidak Bekerja	SORE	Lulus Terlambat
50	Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Sedang	Tidak Bekerja	PAGI	Lulus Terlambat
51	Komputerisasi Akuntansi	P	Cirebon	SMK	Kecil	Sedang	Bekerja	PAGI	Drop Out
52	Komputerisasi Akuntansi	L	Cirebon	SMK	Kecil	Sedang	Bekerja	PAGI	Drop Out
53	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Sedang	Rendah	Bekerja	PAGI	Drop Out
54	Komputerisasi Akuntansi	P	Cirebon	SMK	Kecil	Rendah	Bekerja	PAGI	Drop Out
55	Komputerisasi Akuntansi	L	Luar Cirebon	SMK	Kecil	Rendah	Bekerja	PAGI	Drop Out
….									
….									
….									
….									
….									
….									
….									
….									
975	Sistem Informasi	P	Cirebon	SMA	Besar	Sedang	Tidak Bekerja	SORE	Lulus Tepat Waktu
976	Sistem Informasi	P	Cirebon	SMA	Besar	Sedang	Tidak Bekerja	SORE	Lulus Terlambat
977	Sistem Informasi	P	Cirebon	SMK	Besar	Sedang	Tidak Bekerja	SORE	Lulus Terlambat

LAMPIRAN 3

Tabel Data *Testing* Model Kedua

No	IPS1	IPS2	IPS3	IPS4	STATUS
1	Sedang	Besar	Besar	Besar	Lulus Tepat Waktu
2	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
3	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
4	Sedang	Besar	Besar	Besar	Lulus Tepat Waktu
5	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
6	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
7	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
8	Sedang	Sedang	Sedang	Sedang	Drop Out
9	Sedang	Sedang	Besar	Besar	Drop Out
10	Kecil	Kecil	Kecil	Kecil	Drop Out
11	Sedang	Sedang	Besar	Besar	Drop Out
12	Sedang	Sedang	Besar	Besar	Drop Out
13	Sedang	Sedang	Besar	Besar	Lulus Tepat Waktu
14	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
15	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
16	Sedang	Sedang	Kecil	Kecil	Drop Out
17	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
18	Sedang	Kecil	Kecil	Sedang	Drop Out
19	Besar	Besar	Besar	Besar	Drop Out
20	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
21	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
22	Kecil	Kecil	Kecil	Kecil	Drop Out
23	Sedang	Sedang	Kecil	Kecil	Drop Out
24	Sedang	Sedang	Kecil	Kecil	Drop Out
25	Sedang	Sedang	Kecil	Kecil	Drop Out
26	Sedang	Sedang	Kecil	Kecil	Drop Out
27	Sedang	Sedang	Kecil	Kecil	Drop Out
28	Kecil	Kecil	Kecil	Kecil	Drop Out
29	Sedang	Sedang	Sedang	Kecil	Drop Out
30	Sedang	Sedang	Kecil	Kecil	Drop Out
31	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
32	Kecil	Kecil	Kecil	Kecil	Drop Out
33	Sedang	Sedang	Kecil	Kecil	Drop Out
34	Sedang	Sedang	Sedang	Sedang	Drop Out
35	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
36	Sedang	Sedang	Kecil	Kecil	Drop Out
37	Sedang	Kecil	Kecil	Kecil	Drop Out
38	Sedang	Sedang	Kecil	Kecil	Drop Out
39	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat

No	IPS1	IPS2	IPS3	IPS4	STATUS
40	Sedang	Sedang	Sedang	Besar	Drop Out
41	Sedang	Sedang	Kecil	Sedang	Drop Out
42	Sedang	Sedang	Kecil	Kecil	Drop Out
43	Sedang	Sedang	Kecil	Kecil	Drop Out
44	Sedang	Sedang	Kecil	Kecil	Drop Out
45	Sedang	Sedang	Sedang	Sedang	Drop Out
46	Sedang	Kecil	Kecil	Kecil	Drop Out
47	Sedang	Sedang	Sedang	Kecil	Drop Out
48	Sedang	Sedang	Sedang	Sedang	Drop Out
49	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
50	Sedang	Sedang	Sedang	Besar	Lulus Terlambat
51	Sedang	Sedang	Kecil	Kecil	Drop Out
52	Sedang	Kecil	Kecil	Kecil	Drop Out
53	Sedang	Sedang	Kecil	Sedang	Drop Out
54	Sedang	Kecil	Kecil	Kecil	Drop Out
55	Sedang	Kecil	Kecil	Kecil	Drop Out
56	Sedang	Sedang	Sedang	Sedang	Drop Out
57	Kecil	Kecil	Kecil	Kecil	Drop Out
58	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
59	Besar	Besar	Sedang	Besar	Lulus Tepat Waktu
60	Sedang	Sedang	Sedang	Sedang	Drop Out
61	Sedang	Sedang	Besar	Sedang	Lulus Terlambat
62	Sedang	Kecil	Kecil	Kecil	Drop Out
63	Kecil	Kecil	Kecil	Kecil	Drop Out
64	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
65	Sedang	Sedang	Besar	Besar	Drop Out
66	Besar	Besar	Besar	Besar	Drop Out
67	Sedang	Sedang	Sedang	Besar	Lulus Terlambat
68	Sedang	Sedang	Besar	Sedang	Drop Out
69	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
70	Sedang	Besar	Besar	Sedang	Lulus Tepat Waktu
….					
….					
….					
….					
….					
….					
….					
975	Besar	Besar	Besar	Besar	Lulus Tepat Waktu
976	Besar	Besar	Besar	Besar	Lulus Terlambat
977	Besar	Besar	Besar	Besar	Lulus Terlambat

LAMPIRAN 4

Tabel Data *Testing* Model Pertama

No	Program Studi	JK	Asal Daerah	Asal Sekolah	IPK4	Gaji Ortu	Status Pekerjaan	Kelas	Status
1	Komputerisasi Akuntansi	L	Cirebon	SMA	Besar	Sedang	Tidak Bekerja	Pagi	Lulus Tepat Waktu
2	Komputerisasi Akuntansi	P	Cirebon	SMK	Kecil	Sedang	Bekerja	Sore	Drop Out
3	Komputerisasi Akuntansi	P	Cirebon	SMA	Sedang	Sedang	Bekerja	Pagi	Lulus Terlambat
4	Komputerisasi Akuntansi	L	Cirebon	SMA	Sedang	Tinggi	Bekerja	Pagi	Lulus Terlambat
5	Komputerisasi Akuntansi	P	Cirebon	SMA	Sedang	Rendah	Bekerja	Pagi	Lulus Terlambat
6	Komputerisasi Akuntansi	L	Cirebon	SMA	Sedang	Sedang	Bekerja	Pagi	Lulus Terlambat
7	Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Rendah	Bekerja	Pagi	Lulus Terlambat
8	Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Rendah	Bekerja	Pagi	Lulus Terlambat
9	Komputerisasi Akuntansi	P	Cirebon	SMK	Sedang	Sedang	Bekerja	Pagi	Lulus Terlambat
10	Komputerisasi Akuntansi	P	Cirebon	SMA	Kecil	Sedang	Bekerja	Pagi	Drop Out
11	Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Sedang	Tidak Bekerja	Pagi	Lulus Tepat Waktu
12	Komputerisasi Akuntansi	P	Cirebon	SMK	Besar	Rendah	Bekerja	Pagi	Lulus Terlambat
13	Komputerisasi Akuntansi	P	Cirebon	SMK	Kecil	Sedang	Tidak Bekerja	Pagi	Drop Out
14	Komputerisasi Akuntansi	P	Cirebon	SMA	Besar	Rendah	Bekerja	Pagi	Lulus Terlambat
15	Komputerisasi Akuntansi	L	Luar Cirebon	MAN	Kecil	Sedang	Bekerja	Pagi	Drop Out
16	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Besar	Sedang	Tidak Bekerja	Sore	Lulus Terlambat
17	Komputerisasi Akuntansi	P	Luar Cirebon	SMA	Sedang	Rendah	Bekerja	Sore	Lulus Terlambat
18	Komputerisasi Akuntansi	P	Cirebon	SMA	Kecil	Sedang	Tidak Bekerja	Sore	Drop Out
19	Komputerisasi Akuntansi	L	Cirebon	SMK	Kecil	Rendah	Bekerja	Sore	Drop Out
20	Komputerisasi Akuntansi	P	Cirebon	SMK	Sedang	Sedang	Bekerja	Sore	Lulus Terlambat
21	Komputerisasi Akuntansi	L	Cirebon	SMA	Kecil	Sedang	Bekerja	Pagi	Drop Out
22	Komputerisasi Akuntansi	P	Luar Cirebon	PAKET C	Kecil	Sedang	Tidak Bekerja	Pagi	Drop Out
23	Komputerisasi Akuntansi	P	Luar Cirebon	SMK	Sedang	Sedang	Bekerja	Pagi	Lulus Terlambat
….									
….									
….									
….									
….									
….									
129	Sistem Informasi	L	Cirebon	MAN	Kecil	Rendah	Tidak Bekerja	Sore	Drop Out
130	Sistem Informasi	L	Cirebon	SMA	Sedang	Rendah	Bekerja	Sore	Lulus Terlambat

LAMPIRAN 5

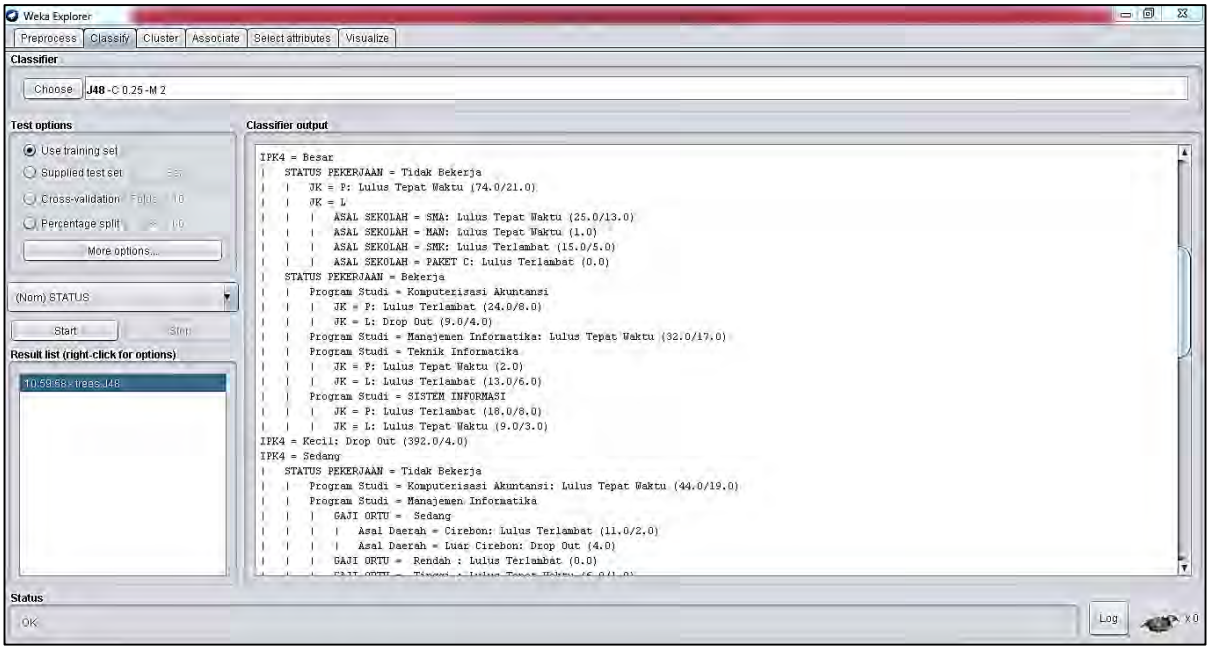
Tabel Data *Testing* Model Kedua

No	IP1	IP2	IP3	IP4	STATUS
1	Besar	Besar	Besar	Besar	Lulus Tepat Waktu
2	Sedang	Sedang	Sedang	Sedang	Drop Out
3	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
4	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
5	Sedang	Sedang	Besar	Besar	Lulus Terlambat
6	Sedang	Sedang	Sedang	Besar	Lulus Terlambat
7	Sedang	Sedang	Sedang	Besar	Lulus Terlambat
8	Sedang	Sedang	Besar	Sedang	Lulus Terlambat
9	Sedang	Sedang	Sedang	Besar	Lulus Terlambat
10	Sedang	Kecil	Kecil	Kecil	Drop Out
11	Sedang	Besar	Besar	Besar	Lulus Tepat Waktu
12	Sedang	Besar	Besar	Besar	Lulus Tepat Waktu
13	Sedang	Sedang	Kecil	Kecil	Drop Out
14	Besar	Besar	Besar	Besar	Lulus Tepat Waktu
15	Sedang	Sedang	Kecil	Kecil	Drop Out
16	Besar	Besar	Sedang	Besar	Lulus Tepat Waktu
17	Sedang	Sedang	Sedang	Besar	Lulus Terlambat
18	Sedang	Kecil	Kecil	Sedang	Drop Out
19	Sedang	Kecil	Kecil	Kecil	Drop Out
20	Sedang	Sedang	Sedang	Besar	Lulus Terlambat
21	Kecil	Kecil	Kecil	Kecil	Drop Out
22	Sedang	Kecil	Kecil	Sedang	Drop Out
23	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
24	Sedang	Sedang	Sedang	Sedang	Lulus Terlambat
25	Sedang	Sedang	Besar	Sedang	Lulus Terlambat
26	Sedang	Besar	Besar	Besar	Lulus Tepat Waktu
27	Sedang	Sedang	Besar	Besar	Lulus Terlambat
28	Sedang	Sedang	Besar	Besar	Lulus Terlambat
29	Sedang	Kecil	Kecil	Sedang	Drop Out
30	Sedang	Sedang	Sedang	Kecil	Lulus Terlambat
….					
….					
….					
….					
….					
….					
128	Sedang	Sedang	Kecil	Kecil	Drop Out
129	Kecil	Kecil	Kecil	Kecil	Drop Out
130	Kecil	Kecil	Kecil	Kecil	Drop Out

LAMPIRAN 6

Screenshot Classifier Output Weka 3.9

1. *Training* Model Pertama



2. *Training* Model Kedua