

Sistem Pendukung Keputusan Prediksi Kecepatan Studi Mahasiswa Menggunakan Metode ID3

Ronny Ardi Giovani¹, Paulus Mudjihartono², Pranowo³

Program Studi Teknik Informatika, Universitas Atma Jaya Yogyakarta

Jl. Babarsari No. 43, Yogyakarta 55281, Indonesia

Email: stefanus_ronny@yahoo.com¹,

paul235@staff.uajy.ac.id², pran@staff.uajy.ac.id³

Abstract. *Decision Support System of Students' Study Speed Prediction Using ID3 Method.* Speed can be a decisive period of study a student taking a degree in sajana. In this study would be built applications that serve to speed decision making predictions Students study Computer Science University of Atma Jaya Yogyakarta. Students will be expected sooner or later than the period of study by taking one course or thesis that will be undertaken after a certain semesters. There are many methods of classification of one method of ID3 (Induction Decision 3 'Tree'). Development system in this study made use of intelligent systems-based applications. The results achieved after the system is formed, among others, sophisticated and intelligent system capable of storing past data is used as a reference for decision making, where students with certain criteria can know the travel time of their studies, and can refer to the database so the system can be more detailed and rigorous in determining the choice.

Keywords: *study period speed, Decision Support Systems, ID3, Intelligent Systems*

Kecepatan masa studi dapat menjadi penentu seorang mahasiswa dalam menempuh gelar sajana. Dalam penelitian ini akan dibangun aplikasi yang berfungsi untuk pengambilan keputusan prediksi kecepatan studi Mahasiswa Teknik Informatika Universitas Atma Jaya Yogyakarta. Mahasiswa akan diprediksi cepat atau lambat masa studi dalam menempuh mata kuliah maupun skripsi yang akan dijalani setelah semester tertentu. Ada banyak metode klasifikasi salah satunya metode ID3 (*Induction Decision 3 'Tree'*). Pembangunan sistem dalam penelitian ini dibuat menggunakan aplikasi berbasis sistem cerdas. Hasil yang dicapai setelah sistem ini terbentuk antara lain sistem canggih dan cerdas yang mampu menyimpan data masa lalu yang digunakan sebagai acuan pengambilan keputusan, dimana mahasiswa dengan kriteria tertentu dapat diketahui masa tempuh studi mereka, serta dapat mengacu pada database sehingga sistem dapat lebih detail serta teliti dalam menentukan pilihan.

Kata Kunci: kecepatan masa studi, Sistem Pendukung Keputusan, ID3, Sistem Cerdas

1. Pendahuluan

Di Era globalisasi seperti sekarang ini berdampak semakin banyaknya kebutuhan yang berhubungan dengan Teknologi Informasi. Hubungan antara manusia, pekerjaan dan Teknologi Informasi kini tidak dapat dipisahkan, Teknologi Informasi sangat membantu manusia untuk memudahkan pekerjaan. Disamping itu besarnya data dan keterbatasan waktu untuk menyelesaikan pekerjaan secara manual kurang memungkinkan tanpa adanya Teknologi Informasi karena akan memakan waktu yang lama dan keakuratan data kurang terjamin. Selain

itu penghematan biaya dan waktu dapat digunakan untuk melakukan hal lain yang lebih berguna sehingga dapat meningkatkan efektivitas pekerjaan. Salah satu cara yang ada dan dapat mempermudah pekerjaan adalah dengan metode klasifikasi. Penggunaan metode klasifikasi dapat diterapkan untuk menangani data alumni. Sehingga data dapat dipelajari dan diketahui polanya. Dengan metode ini diharapkan mahasiswa semester 4 dapat diprediksi kecepatan studinya, sehingga kepada mereka dapat diberikan *early warning* atau peringatan dini.

2. Landasan Teori

2.1. Evaluasi Studi Mahasiswa

Evaluasi keberhasilan studi mahasiswa dilakukan melalui pemberian tugas, penyelenggaraan ujian, dan sejenisnya. Sumber yang dipergunakan dalam menentukan evaluasi keberhasilan studi adalah sebagai berikut: (1) Kegiatan Akademik Terstruktur (KAT), yaitu penilaian terhadap kegiatan seorang mahasiswa selama mengikuti kegiatan pembuatan tugas, hadir dan aktif dalam responsi, tes kecil, dan praktikum. (2) Ujian Tengah Semester (UTS), yaitu ujian yang diadakan pada pertengahan semester yang berjalan dan terjadwal dalam kalender akademik. (3) Ujian Akhir Semester (UAS), yaitu ujian yang diadakan pada setiap akhir semester dan terjadwal dalam kalender akademik.

2.2. Sistem Penilaian

Keberhasilan studi mahasiswa dinilai menurut huruf mutu seperti tercantum dalam tabel 1.

Tabel 1. Tabel bobot nilai

Huruf Mutu	Arti	Bobot Nilai
A	Istimewa	4
B	Baik	3
C	Cukup	2
D	Kurang	1
E	Buruk	0

2.3. Indeks Prestasi

Indeks Prestasi (IP) adalah nilai kredit rata-rata yang merupakan satuan nilai akhir yang menggambarkan mutu proses belajar mengajar tiap semester atau dapat diartikan juga sebagai besaran/angka yang menyatakan prestasi (keberhasilan dalam proses belajar mengajar) mahasiswa pada suatu semester. Indeks Prestasi (IP) dihitung jumlah angka mutu dibagi jumlah kredit pada satu semester dan dinyatakan dalam bilangan dengan dua angka desimal di belakang koma. Indeks Prestasi Kumulatif (IPK) adalah ukuran yang menunjukkan prestasi mahasiswa mulai semester pertama sampai semester terakhir yang telah ditempuh secara kumulatif.

2.4. Pengertian Klasifikasi

Klasifikasi merupakan salah satu teknik dalam data mining. Klasifikasi didefinisikan sebagai berikut: (1) Meramalkan kategori label kelas (diskrit atau nominal). (2) Mengklasifikasikan (membuat suatu model) berdasarkan himpunan pelatihan dan nilai-nilai (label kelas) dalam suatu atribut klasifikasi dan menggunakannya didalam mengklasifikasikan data baru.

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu

objek yang labelnya tidak diketahui. Model itu sendiri bisa berupa aturan “jika-maka”, berbentuk pohon pengambilan keputusan (*decision tree*), formula matematis seperti *Bayesian* dan *Support Vector* atau biasa juga berupa jaringan seperti neural network.

Proses klasifikasi biasanya dibagi menjadi dua fase: *learning* dan *test*. Pada fase *learning*, sebagian data telah diketahui kelas datanya diumpangkan untuk membentuk model prediksi. Karena menggunakan data yang telah diberikan label terlebih dahulu oleh ahli dibidang itu sebagai contoh data yang benar maka klasifikasi sering juga disebut sebagai metoda diawasi (*supervised method*). Kemudian pada fase *test*, model yang sudah terbentuk diuji dengan sebagian data lainnya untuk mengetahui akurasi dari model tersebut. Bila akurasinya mencakupi model ini dapat dipakai untuk prediksi kelas data yang belum diketahui.

2.5. Sistem Pendukung Keputusan

Sistem pendukung keputusan (dalam istilah Inggris: *decision support systems* disingkat DSS) adalah bagian dari sistem informasi berbasis komputer (termasuk sistem berbasis pengetahuan (manajemen pengetahuan)) yang dipakai untuk mendukung pengambilan keputusan dalam suatu organisasi atau perusahaan. Dapat juga dikatakan sebagai sistem komputer yang mengolah data menjadi informasi untuk mengambil keputusan dari masalah semi-terstruktur yang spesifik. Beberapa pengertian sistem Pendukung keputusan menurut para ahli computer, Little(1970) mendefinisikan sistem pendukung keputusan sebagai sebuah himpunan/kumpulan prosedur berbasis model untuk memproses data dan pertimbangan guna membantu manajemen dalam mengambil suatu keputusan. Menurut Keen(1980) sistem pendukung keputusan adalah sistem berbasis computer yang dibangun lewat sebuah proses adaptif dari pembelajaran, pola-pola penggunaan dan evolusi sistem. Bonczek (1980) mendefinisikan sistem pendukung keputusan sebagai sebuah sistem berbasis komputer yang terdiri atas beberapa komponen antara lain komponen sistem bahasa (*language*), komponen sistem pengetahuan (*problem processing*) yang saling berinteraksi satu dengan lainnya. Hick(1993) menyebutkan sistem pendukung keputusan sebagai sekumpulan *tools* komputer yang terintegrasi untuk memungkinkan seorang *decision maker* untuk berinteraksi langsung dengan komputer guna menciptakan informasi yang berguna dalam membuat keputusan semi terstruktur dan tidak terstruktur yang tidak terantisipasi (Republik BM). Sedangkan menurut Alter (1990), apabila sistem pendukung keputusan dibandingkan dengan sebuah sistem pemrosesan data elektronik (EDP / *Electronic Data Processing*) maka akan terlihat 5 hal perbedaan seperti pada tabel 2. (Subakti, 2002).

Tabel 2. Tabel Perbandingan SPK dan EDP

Pembanding	SPK	EDP
Penggunaan	Aktif	Pasif
Pengguna	Manajemen	Operator/ Pegawai
Tujuan	Efektivitas	Efisiensi Mekanis
Time Horizon	Sekarang dan masa depan	Masa Lalu
Kelebihan	Fleksibilitas	Konsistensi

Berdasarkan definisi, dapat ditarik beberapa ciri/karakteristik umum dari sebuah sistem pendukung keputusan yang ideal yaitu: (1) SPK adalah sebuah sistem berbasis komputer dengan antarmuka antara mesin/komputer dan pengguna. (2) SPK ditujukan untuk membantu membuat keputusan dalam menyelesaikan suatu masalah dalam berbagai level manajemen dan bukan untuk mengganti posisi manusia sebagai pembuat keputusan. (3) SPK mampu member alternatif solusi bagi masalah semi/tidak terstruktur baik bagi perseorangan ataupun kelompok dan dalam berbagai macam proses pengambilan keputusan. (4) SPK menggunakan data, basis data dan analisa model-model keputusan. (5) SPK bersifat adaptif, efektif, interaktif, *easy to use* dan *fleksibel*.

Sistem pendukung keputusan terbangun dari beberapa subsistem, antara lain (Subakti, 2002): (1) Subsistem Manajemen Data. Subsistem ini meliputi basisdata yang relevan dengan keadaan yang ada, serta dikelola oleh sebuah sistem yang dikenal sebagai *database management system* (DBMS). (2) Subsistem Manajemen Model. Subsistem ini merupakan sebuah paket perangkat lunak yang berisi model-model *financial, statistic, management science* dan model kuantitatif lain yang menyediakan kemampuan analisis sistem dan manajemen perangkat lunak yang terkait. (3) Subsistem Manajemen Pengetahuan (*Knowledge*). Subsistem ini merupakan subsistem yang mampu mendukung subsistem lain atau berlaku sebagai komponen yang mampu berdiri sendiri. (4) Subsistem antarmuka Pengguna (User Interface). Subsistem ini merupakan media tempat komunikasi antara pengguna dan sistem pendukung keputusan serta tempat pengguna untuk memberikan perintah kepada sistem pendukung keputusan.

Beberapa tahapan untuk pembuatan sistem pendukung keputusan, antara lain: (1) Pendefinisian masalah. (2) Pengumpulan data atau elemen informasi yang relevan. (3) Pengolahan data menjadi suatu informasi baik dalam bentuk laporan grafik maupun tulisan. (4) Menentukan alternatif-alternatif solusi (bisa dalam bentuk persentase). Sedangkan tujuan dari sistem pendukung keputusan itu sendiri antara lain: (1) Membantu menyelesaikan masalah. (2) Mendukung manajer dalam mengambil suatu keputusan. (3) Mengingat efisiensi bukan efisiensi pengambilan keputusan.

Dapat terlihat jelas, bahwa sistem pendukung keputusan mempunyai beberapa manfaat penting yaitu: (1) Membantu pengambilan keputusan yang rasional, sesuai dengan jenis keputusan yang diperlukan. (2) Dapat membuat peramalan/*forecasting*. (3) Membandingkan alternatif tindakan. (4) Memuat model.

2.6. Algoritma Decision Tree

Decision Tree (pohon keputusan) adalah *flowchart* seperti struktur *tree*, dimana tiap internal node menunjukkan sebuah test pada sebuah atribut, tiap cabang menunjukkan hasil dari test, dan *leaf node* menunjukkan *class-class* atau *class distribution* (Han & Khamber, 2001).

Decision Tree adalah pohon terstruktur dari sekumpulan atribut untuk diuji dengan tujuan meramalkan *output*-nya. Pohon keputusan merupakan salah satu *tool* paling populer untuk klasifikasi karena hasilnya yang bisa dipahami dalam bentuk kaidah keputusan.

Untuk memutuskan atribut mana yang harus diuji pertama, yaitu atribut yang memiliki perolehan tertinggi.

Decision Tree berusaha menemukan atribut yang tepat untuk menyelesaikan dan menentukan kelas. Dengan kata lain, atribut yang berada di posisi paling atas dalam *decision tree* adalah atribut yang paling berpengaruh dalam menentukan hasil prediksi. Oleh karena itu, proses pembuatan model *decision tree* merupakan proses interative untuk menilai atribut paling berpengaruh yang akan menjadi percabangan dari struktur pohon ini. Percabangan ini juga sering disebut dengan istilah titik percabangan (*split point*).

Proses ini bisa dibagi menjadi dua tahap. Yang pertama adalah menghitung besarnya pengaruh dari setiap atribut. Tahap kedua adalah memilih atribut terbaik untuk menjadi titik percabangan dari struktur pohon. Untuk tahap pertama ada beberapa matrik yang bisa dipakai untuk menghitung besarnya pengaruh seperti *entropy, information gain* dsb.

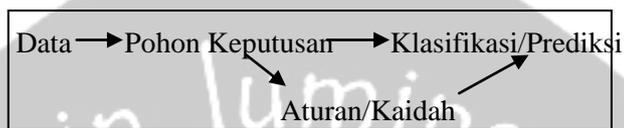
Algoritma *Decision Tree* yang sering digunakan adalah algoritma ID3 yang disebut juga algoritma *greedy*. Dalam algoritma ini sebuah pohon dibentuk dengan cara *top-down, recursive, divide-and-conquer*.

2.7. Induction Decision Tree (ID3)

ID3 merupakan sebuah metode yang digunakan untuk membangkitkan pohon keputusan. Input dari algoritma ini adalah sebuah database dengan beberapa variabel yang juga dikenal dengan atribut. Setiap masukan dalam database menyajikan sebuah objek dari domain yang disebut dengan variabel bebas (*independent variable*). Sebuah atribut didesain untuk mengklasifikasikan objek yang disebut dengan variabel tidak bebas (*dependent variable*).

Proses klasifikasi dilakukan dari node paling atas yaitu akar pohon(*root*). Dilanjutkan ke bawah melalui cabang-cabang sampai dihasilkan node daun (*leaves*) dimana node daun ini menunjukkan hasil akhir klasifikasi. Sebuah objek yang diklasifikasikan dalam pohon harus dities nilai entropinya. Entropi adalah ukuran dari teori informasi yang dapat mengetahui karakteristik *impurity* dan *homogeneity* dari kumpulan data. Dari nilai entropi tersebut kemudian dihitung nilai information gain (IG) masing-masing atribut *independent* terhadap atribut *dependent*-nya. IG merupakan nilai rata-rata entropi pada semua atribut.

Sebuah pohon keputusan dapat dikonstruksi dengan melihat regurasi seperti pada gambar 1.



Gambar 1. Alur Diagram Pohon Keputusan

Proses Algoritma ID3 (Frank,2003): Input: Training samples, samples; Output: Decision Tree; Method: Create node N; (1) If samples are all of the same class, C then return N as a leaf node labeled with the class C; (2) If atribut-list is empty then Return N as a leaf node labeled with the most common class in samples; select test-atribut, atribut among attribute-list with the highest information gain; (3) Label node N with test-atribut; (4) For each known value a, of test-atribut: Grow a branch from node N for the condition test atribut ai; Let Si be the set of samples in samples for which test-atribut = ai; If Si is empty then Attach a leaf labeled with the most common class in samples; Else attach the node returned by Generate_decision_tree(Si,atribut-list-test-atribut).

Langkah menghitung information gain adalah sebagai berikut: (1) Hitung info kelas

$$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m P_i \cdot \log_2(P_i) \quad \text{dengan} \quad P_i = \frac{s_i}{s}$$

yang mempunyai class Ci. (2) Hitung Entropy attribute: adalah probabilitas dari sample

$$E(A) = \sum_{j=1}^v \frac{s_{1j} + \dots + s_{mj}}{s} I(s_1, s_2, \dots, s_m) \quad (3) \quad \text{Hitung gain:}$$

$$\text{Gain}(A) = I(s_1, s_2, \dots, s_m) - E(A)$$

Adapun sample data yang digunakan oleh ID3 memiliki beberapa syarat sebagai berikut: (1) Deskripsi atribut-nilai. Atribut yang sama harus mendeskripsikan tiap contoh dan memiliki jumlah nilai yang sudah ditentukan. (2) Kelas sudah didefinisikan sebelumnya, dan digambarkan dengan jelas. Untuk kelas yang kontinu dipecah-pecah menjadi kategori-kategori yang relative, misalnya saja metal dikategorikan sebagai "hard, quite hard, flexible, soft quite soft". (3) Jumlah contoh yang cukup. Karena dibutuhkan test case yang cukup membedakan pola yang valid dari peluang suatu kejadian.

Dengan pemodelan *decision tree*, persoalan klasifikasi yang cukup sederhana dapat dilakukan dengan akurat karena dalam mengklasifikasikan suatu objek, tentunya terdapat aturan-aturan yang menjadi penentu termasuk kategori manakah objek tersebut. Dengan algoritma ID3, pohon keputusan dibangun berdasarkan aturan-aturan tersebut, sehingga proses *learning* akan memiliki akurasi yang tinggi karena klasifikasinya dilakukan dengan melakukan inferensi terhadap aturan-aturan tersebut.

3. Metode Penelitian

Metode klasifikasi digunakan dalam penelitian ini, karena metode ini merupakan metode yang cocok digunakan untuk pembangunan sistem pendukung keputusan. Ada beberapa algoritma yang dapat digunakan untuk membantu dalam pengambilan keputusan antara lain

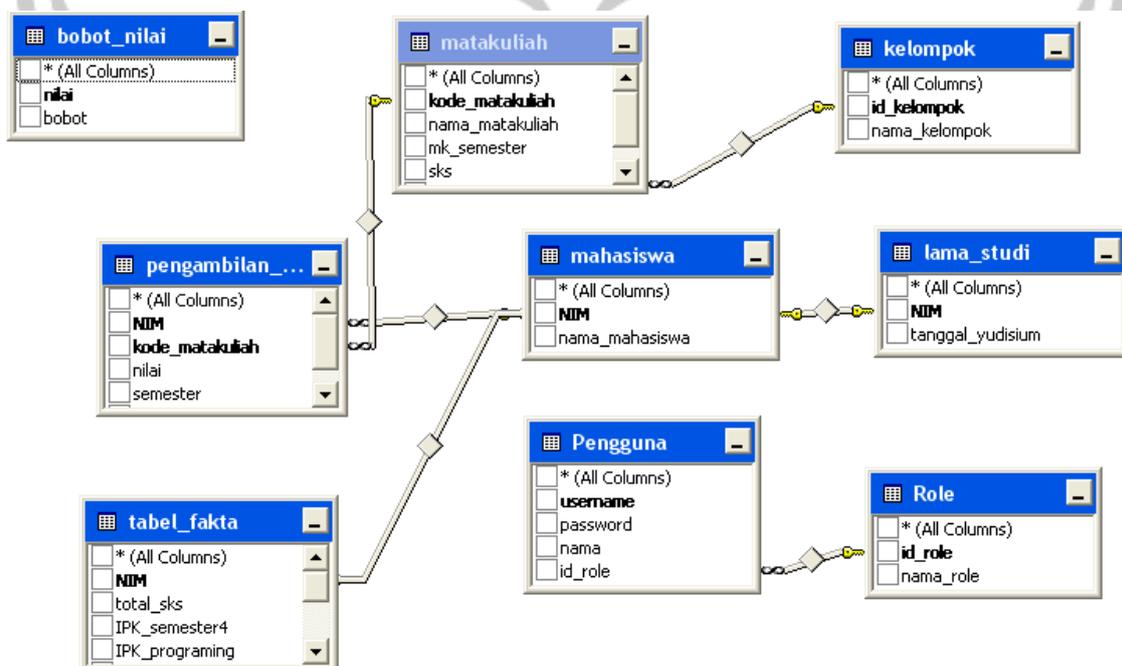
Algoritma Genetika, Logika Fuzzy, ID3. ID3 merupakan salah satu algoritma yang dapat digunakan untuk pengklasifikasian. Keunggulan dari algoritma ID3 adalah tahap belajar yang cepat, *time complexity* yang rendah, ketelitian klasifikasi yang tinggi. Tujuan algoritma ID3 adalah mendapatkan decision tree yang merupakan salah satu bentuk “*Classification Models*” yang terbaik. Salah satu masalah yang dapat dipecahkan menggunakan algoritma ini yakni upaya untuk mendapatkan *decision tree* terbaik yakni pendekatan seminimal mungkin yang konsisten dari sekumpulan data, termasuk dalam kategori algoritma *NP-Hard/Completeness*. Mekanisme Pembangunan algoritma ini yaitu dilakukan secara *top-down* kemudian diawali dengan pertanyaan “*Attribute* mana yang harus diperiksa pada *root* dari *decision tree*” kemudian pembentukan dilakukan dengan mempartisi *training examples*. Kekuatan utama pada algoritma adalah fungsi *heuristic information gain* untuk memilih atribut terbaik, mewujudkan *Greedy Heuristic Search* antara lain *Hill-Climbing* tanpa *Backtracing*.

Sistem informasi ini akan berbasis *client server* sehingga dapat diakses oleh beberapa *user* sekaligus baik aplikasi desktop maupun web. Data yang digunakan dalam pembangunan sistem informasi ini antara lain data mahasiswa, matakuliah dan data nilai. Keunggulan dari Sistem Informasi ini adalah data *driven, updated*, dan *time variant* karena keluaran akan selalu berubah berdasarkan data nilai dan matakuliah yang pernah diambil siswa dimasa lampau yang digunakan sebagai acuan cepat atau lambatnya masa studi mereka. Hasil yang dicapai dengan aplikasi sistem pendukung keputusan prediksi kecepatan studi mahasiswa menggunakan id3 antara lain dapat melakukan klasifikasi dan memberikan umpan balik berupa status lama studi mahasiswa.

4. Pembahasan

4.1. Analisis dan Perancangan Database

Terdapat 9 tabel yang saling berhubungan satu dengan yang lain berdasarkan primary dan foreign key yang telah didefinisikan sejak awal. Analisis harus berfikir seperti manajer dalam membaca data dan menformulasikan paket informasi, primary key, foreign key, check schema dan role yang telah ditentukan pada saat pendefinisian database.



Gambar 2. Analisis dan Perancangan Database

4.2. Aturan Nilai dan Kecepatan Studi

Dalam penelitian ini penulis menggunakan beberapa aturan untuk menentukan nilai mahasiswa tersebut itu tinggi, sedang atau rendah dan aturan untuk menentukan kecepatan studi itu cepat ataupun lambat. Penentuan nilai dikelompokkan berdasarkan matakuliah yang telah ditentukan sejak awal.

Untuk nilai dapat di golongkan sebagai berikut: Tinggi jika nilai diatas 3.0, Sedang jika diatas 2.5 hingga 3.0, Rendah jika 0 hingga 2.5. Untuk kecepatan skripsi dapat digolongkan sebagai berikut: Cepat jika mahasiswa menyelesaikan skripsinya kurang atau sama dengan 12 bulan, Lambat jika mahasiswa menyelesaikan skripsinya lebih dari 12 bulan. Untuk kecepatan teori dapat digolongkan sebagai berikut: Cepat jika mahasiswa menyelesaikan studinya kurang atau sama dengan 36 bulan, Lambat jika mahasiswa menyelesaikan studinya lebih dari 36 bulan.

4.3. Pengolahan Data

Dari data yang ada dan dikelompokkan berdasarkan aturan yang telah dibuat maka dapat di hitung jumlah mahasiswa angkatan 2005 yang memiliki kecepatan skripsi cepat sebanyak 238 mahasiswa dan yang lambat sebanyak 40 mahasiswa. Sedangkan mahasiswa yang memiliki kecepatan teori cepat sebanyak 98 mahasiswa dan memiliki kecepatan teori lambat sebanyak 180 mahasiswa.

5. Kesimpulan

Kesimpulan yang dapat diambil oleh penulis terhadap Aplikasi Sistem Pendukung Keputusan Prediksi Kecepatan Studi Mahasiswa Dengan Menggunakan Metode ID3 adalah: (1) Aplikasi penKIK wangi telah berhasil di desain serta diimplementasikan menggunakan Microsoft Visual Studio .Net 2008 dengan baik. Dengan adanya penKIK wangi dapat memudahkan prediksi kecepatan studi mahasiswa. (2) Aplikasi penKIK wangi telah berhasil menentukan klasifikasi kecepatan studi mahasiswa dengan menggunakan metode ID3.

Referensi

- Baaz, M., Hajek, P., Montagna, F., and Veith, H., 2002, "Complexity of t-tautologies", *Annals of Pure and Applied Logic* 113: 3-11.
- Belnap, Jr., Nuel, D., 1996, "Agents in branching time", in *Logic and Reality: Essays on the Legacy of Arthur Prior*, Jack Copeland, ed., Oxford: Oxford University Press, 239-271.
- Frank, Vanden Berghen, 2003, *Classification Trees : C4.5*, University Libre de Bruxelles.
- Hanikova, Z., 2002, "A note on the complexity of propositional logics of individual t-algebras", *Neural Network World*, 21: 453-460.
- Jiawei Han, Micheline Kamber, 2001, *Data Mining Concept and Techniques*.
- Pyeat Larry D, Howe Addele E., 1998. "Decision Tree Function Approximation in Reinforcement Learning", Colorado State University.
- Sharma T, Silvescu A, Honavar V., 2000 "Learning Classification Trees from distributed Horizontally and vertically Fragmented Data Sets", Artificial Intellegence Reserch Laboratory Department of Computer Science Iowa State Univercity, Ames.