

OPTIMASI JUMLAH MAHASISWA ASIA MELALUI PREDIKSI MASA STUDI MENGGUNAKAN METODE INDUKSI DECISION TREE

Jaenal Arifin¹⁾, Puji Subekti²⁾ Widya Adharyanty³⁾

Teknik, STMIK ASIA Malang (Jaenal Arifin)

email : jaenalarifin@asia.ac.id

Teknik, STMIK ASIA Malang (Puji Subekti)

Email : pujisubekti88@gmail.com

Teknik, STMIK ASIA Malang (Widya Adharyanty)

Email : widyariyant@gmail.com

Abstract

The accuracy of the study period can be determinant the students to pursued undergraduate degree. The available data indicated that only 30% of students can graduate <5 years, the rest graduate > 5 years and became non-active students. To overcome of low graduation rate is needed the system to determine the relationship between the master's students with study period that taken by the students. The use of data mining techniques in this system is expected to provide insights that were previously hidden in the data warehouse to be valuable information. By utilizing data mining techniques particularly algorithm ID3, then the researchers made an application to find a pattern that can predict the future of a student's study period based on the data from students and academic score. The student data, scores, and the study period integrated into the data training. The data training is processed into a decision tree based on the calculation of the gain and entropy. From that tree made a rule that can predict a student's study period. From 140 data training and 20 data testing with 6 kinds of attributes input and 2 kinds of target attributes, can be obtained by the accuracy of the ID3 prediction result for 85%, while the error rate prediction results for 15%.

Keywords : Optimasi, Jumlah Mahasiswa Asia, Prediks, Induksi Decision Tree

PENDAHULUAN

1. LATAR BELAKANG

Tingkat kelulusan yang masih terbilang rendah disebabkan oleh beberapa faktor, diantaranya adalah kurangnya strategi mahasiswa dalam belajar, adanya perbedaan kemampuan setiap mahasiswa, kurangnya pengawasan orang tua, atau adanya keterbatasan kemampuan finansial. Untuk mengatasi masalah tersebut diperlukan sistem yang dapat mengetahui hubungan antara data mahasiswa dengan masa studi yang ditempuh mahasiswa.

Penggunaan teknik *data mining* dalam sistem ini diharapkan dapat memberikan pengetahuan yang sebelumnya tersembunyi di dalam gudang data sehingga menjadi informasi yang berharga. Dengan memanfaatkan teknik *data mining*, khususnya algoritma ID3, dibuat sebuah aplikasi yang dapat

menemukan pola untuk memprediksi masa studi mahasiswa berdasarkan data induk mahasiswa dan nilai akademik. Mahasiswa akan diprediksi cepat atau lambatnya masa studi yang akan dijalani selama masa perkuliahan.

Rumusan Masalah

Permasalahan yang dibahas adalah bagaimana optimasi jumlah mahasiswa STMIK ASIA melalui prediksi masa studi menggunakan metode induksi decision tree

Batasan Masalah

Batasan masalah sesuai dengan ruang lingkup pembahasan diantaranya adalah: - Sistem ini dibuat berdasarkan data-data mahasiswa STMIK Asia angkatan 2006 s/d 2008 jurusan Teknik Informatika dan asal sekolah adalah SMU atau sejenisnya.

- Variabel yang digunakan adalah Usia Saat Masuk, Nilai IP Semester 1, Nilai IP Semester
 - Tempat Tinggal Orang Tua, Pendidikan Orang Tua, dan Pekerjaan Orang Tua.
- a. Data yang dipakai tidak termasuk data mahasiswa yang:
 - a. Non Aktif
 - b. Cuti akademik
 - c. Pindah dari perguruan tinggi negeri/swasta yang sejenis
 - d. Pindah ke perguruan tinggi lain
 - e. Mahasiswa tugas belajar
 - f. Putus Kuliah (Drop Out)
 - g. Meninggal dunia
 - h. Pemberhentian tetap/ sementara sebagai mahasiswa STMIC Asia diatur dan ditetapkan dengan keputusan Ketua STMIC Asia setelah memperhatikan pertimbangan dari Ketua Jurusan dan Pembantu Ketua Jurusan dan Pembantu Ketua Bidang Akademik STMIC Asia. - Sistem ini menggunakan metode klasifikasi *Induksi Decision Tree (ID3)*.
 - Hasil dari proses ini berupa prediksi yang berbentuk kategori, yaitu mahasiswa lulus tepat waktu (< 5 tahun) dan lulus dengan waktu yang lebih lama (> 5 tahun).

Tujuan

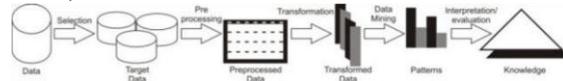
Tujuan yang ingin dicapai adalah menghasilkan rancangan aplikasi untuk memprediksi lama studi mahasiswa di STMIC Asia dengan menggunakan metode ID3

2. KAJIAN LITERATUR

Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, dimana pola yang ditemukan bersifat sah (*valid*), baru (*novel*), dapat bermanfaat (*potentially usefull*), dapat dimengerti (*ultimately understandable*). Proses KDD terdiri dari *Data Selection, Pre Processing/Cleaning,*

Transformation, Data Mining, dan Interpretation/ Evaluation (Fayyad, 1996).



Gambar 1 Tahapan proses KDD

Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Data mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya (**Pramudiono, 2006**).

Pohon Keputusan

Pohon Keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing-masing aturan pembagian, anggota himpunan hasil menjadi lebih mirip antara satu dengan yang lain (Linoff, 2004).

Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5 (Larose, 2005).

Data dalam Pohon Keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record*. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon. Misalkan untuk bermain tennis, kriteria yang diperhatikan adalah cuaca, angin, dan temperatur. Salah satu atribut merupakan atribut yang menyatakan data solusi per *item* data yang disebut kelas/target atribut. Atribut memiliki nilai-nilai yang dinamakan *instance*. Misalkan atribut cuaca memiliki *instance* berupa cerah, berawan, dan hujan. Proses pada Pohon Keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule* (Basuki, 2003).

Pohon keputusan merupakan himpunan aturan *IF...THEN*. Setiap *path* dalam *tree* dihubungkan dengan sebuah aturan, di mana premis terdiri atas sekumpulan node-node yang ditemui, dan kesimpulan dari aturan terdiri atas kelas yang terhubung dengan *leaf* dari *path* (Romansyah, 2009).

Induksi Decision Tree (ID3)

Secara ringkas, langkah kerja Algoritma ID3 dapat digambarkan sebagai berikut (Gorunescu, 2011):

- Menghitung nilai *Impurity* dan *Information gain* dari setiap atribut.

$$Entropy(i) = i_E(i) = - \sum_{j=1}^m f(i,j) \cdot \log_2 [f(i,j)]$$

$$Gain \Delta(S,i) = Entropy(S) - \sum_{i=1}^p \frac{n_i}{n} Entropy(i)$$

- Atribut yang memiliki nilai *information gain* terbesar akan dijadikan cabang pada pembentukan pohon keputusan.

- Membentuk simpul yang berisi atribut tersebut.
- Proses perhitungan *information gain* akan terus diulang sampai semua data telah termasuk dalam kelas yang sama.

METODE PENELITIAN

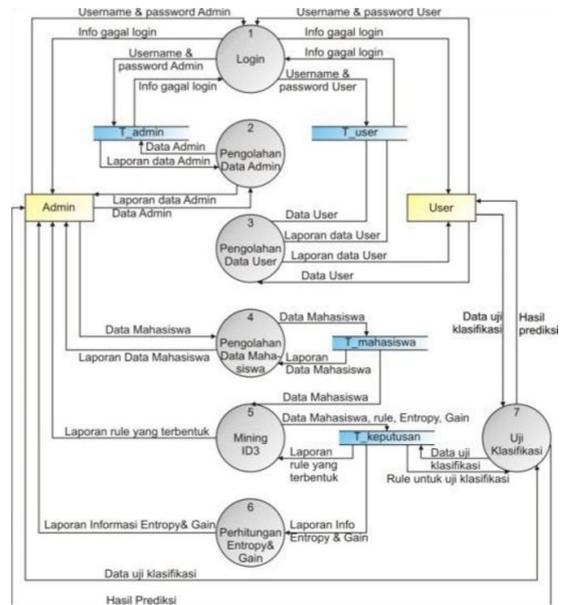
Context Diagram (CD)



Gambar 2 Context Diagram (CD)

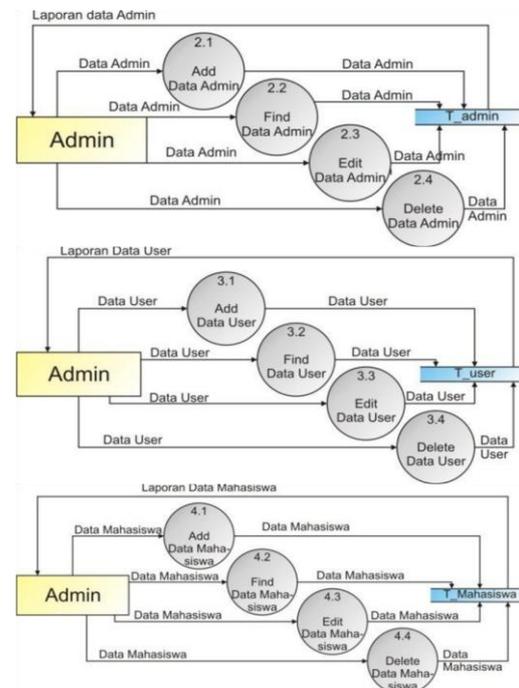
Context Diagram atau disebut juga dengan model sistem fundamental merepresentasikan seluruh elemen sistem sebagai sebuah *bubble* tunggal dengan data input output yang ditunjukkan oleh anak panah yang masuk dan keluar secara berurutan. Di sini tergambar *User* yang memakai sistem ada 2, yaitu Admin dan *User* lain.

Data Flow Diagram (DFD)



Gambar 3 Data Flow Diagram (DFD) Level 0

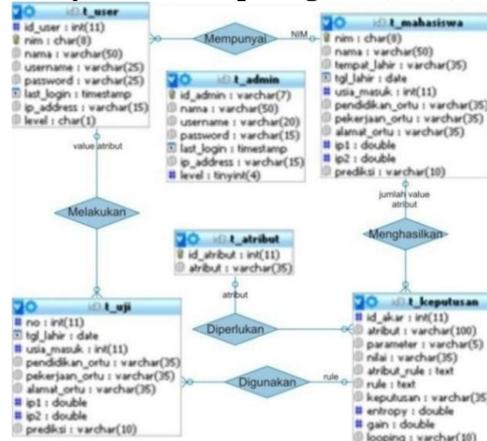
Pada *Data Flow Diagram* Level 0 terjadi 7 proses, yaitu: login, pengolahan data admin, pengolahan data user, pengolahan data mahasiswa, mining ID3, perhitungan gain & entropy dan uji klasifikasi.



Gambar 4 DataFlowDiagram (DFD) Level 1

Terdapat 3 buah DFD level 1, yaitu DFD level 1 Pengolahan Data Admin, DFD level 1 Pengolahan Data User, dan DFD level 1 Pengolahan Data Mahasiswa. Pada DFD level 1 ini hanya bagian Admin yang menggunakannya, yaitu berupa proses *Input, Edit, Delete*, dan *Find* baik untuk proses awal Pengolahan Data Admin, Pengolahan Data User, maupun Pengolahan Data Mahasiswa. Output dari proses ini adalah laporan data yang sudah diolah tersebut.

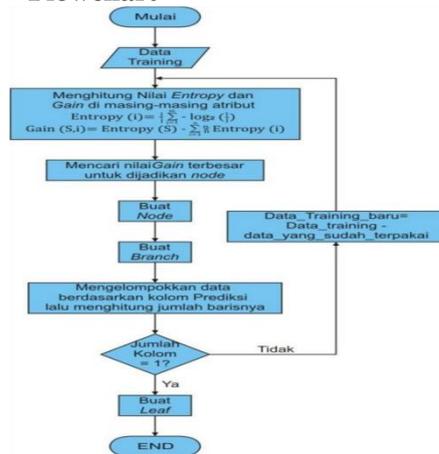
Entity Relationship Diagram (ERD)



Gambar 5 Entity Relationship Diagram (ERD)

Setiap mahasiswa d tabel T_Mahasiswa adalah *user* yang disimpan di tabel T_User yang dihubungkan dengan NIM. Jumlah semua *value* atribut dihitung untuk menghasilkan data di tabel t_keputusan yang membutuhkan tabel T_Atribut untuk menghasilkan *rule*. Rule yang terbentuk pada tabel T_Keputusan digunakan untuk melakukan pengujian yang dilakukan oleh *user* dengan cara memasukkan *value* atribut.

Flowchart



Gambar 6 Flowchart Pembentukan Tree

Pembentukan *tree* dimulai dengan memasukkan *data training*. Dari data training tersebut dihitung nilai *Entropy* dan *Gain*. Nilai *Gain* terbesar dijadikan akar (*root*). Pembentukan pohon dilanjutkan dengan membuat batang (*branch*). *Branch* berisi *value input atribut*. Jika hasil *target atribut* homogen (*pure*), maka dapat dibentuk *leaf*. Jika belum homogen, maka langkah menghitung nilai *Gain* diulang sampai mendapatkan hasil yang homogen.

Studi Kasus

Sebagai sample untuk membentuk *rule*, maka diambil 20 data mahasiswa acak. Kolom berwarna biru adalah atribut input, sedangkan kolom berwarna hijau adalah kolom atribut target. Kolom Usia berisi *record* usia mahasiswa saat masuk perguruan tinggi. Kolom TTOT berisi *record* Tempat Tinggal Orang Tua, jika tempat tinggal orang tua berada di wilayah Malang Raya maka tertulis “MR”, jika berada di wilayah Jawa Timur tetapi bukan Malang Raya maka akan tertulis “JT”, jika berada di wilayah pulau Jawa tetapi bukan di Jawa Timur maka akan tertulis “PJ”, jika selain tersebut di atas maka akan tertulis “LP” (Luar Pulau). Kolom Pd.OT berisi *record* Pendidikan Orang Tua. Kolom Pk.OT berisi *record* pekerjaan orang tua, jika pekerjaan adalah pensiunan maka akan tertulis “P”, jika pekerjaan adalah

Pegawai Negeri maka akan tertulis “PN”, jika pekerjaan adalah Pegawai Swasta maka akan tertulis “PS”, jika pekerjaan adalah petani maka akan tertulis “PT”, jika pekerjaan adalah Wiraswasta maka akan tertulis “WS”. Kolom IP1 dan IP2 berisi *record* nilai IP semester 1 dan nilai IP semester 2 tiap mahasiswa.

Tabel 1 Data Training Rekursi Level 0

Nama	Usia	TTOT	Pd.OT	Pk.OT	IP1	IP2	<5?
A	19	JT	SD	PT	3.3	3.27	ya
B	18	JT	S-1	P	3.15	2.68	tidak
C	19	LP	S-1	PN	2.68	2.95	tidak
D	18	MR	SD	WS	2.85	2.6	ya
E	20	LP	SD	WS	2.73	1.9	tidak
F	22	JT	SMA	PS	2.73	2.64	ya
G	18	JT	SD	PT	2.93	2.95	ya
H	18	JT	SMA	PN	3.23	3.21	tidak
I	19	JT	SMA	WS	2.78	2.52	ya
J	20	MR	S-1	PN	2.78	2.52	ya
K	19	JT	SD	WS	2.03	2.53	tidak
L	18	JT	SMA	PN	3.13	2.94	ya
M	19	JT	S-1	P	3.1	3.06	ya
N	20	LP	SD	PT	3.2	3.13	ya
O	18	JT	S-1	PN	2.68	2.55	tidak
P	19	JT	S-1	PN	2.08	2.22	tidak
Q	18	LP	SMA	PT	2.73	2.28	tidak
R	18	PJ	Diploma	P	2.93	2.12	ya
S	19	PJ	SMA	WS	2.6	2.57	tidak
T	21	PJ	SD	WS	3.1	3.15	ya

Untuk menghitung *Gain* di tiap atribut, maka *Entropy parent* harus dihitung terlebih dahulu. *Entropy parent* dihitung berdasarkan *value* pada atribut target. Dalam hal ini *value* “ya” adalah 11 dan *value* “tidak” adalah 9.

$$Entropy(i) = - \left[\frac{11}{20} \log_2 \frac{11}{20} + \frac{9}{20} \log_2 \frac{9}{20} \right] = 0,9927$$

Setelah menghitung *Entropy parent*, maka nilai *Gain* didapatkan dari hasil pengurangan *Entropy parent* dengan *Entropy* tiap *value* atribut. Untuk *value* atribut dengan kategori numerik, maka data harus diurutkan terlebih dahulu, lalu dihitung mediannya. Nilai *Gain* dihitung pada masing-masing titik median.

Tabel 2 Perhitungan Entropy dan Gain Rekursi level 0

Usia	≤ median usia		E	>median Usia		E	Gain
	Yes	No		Yes	No		
18	4	4	1	7	5	0.98	0.005
19	7	8	0.997	4	1	0.722	0.065
20	9	9	1	2	0	0	0.093
21	10	9	0.998	1	0	0	0.045
22							
IP1	≤ median IP1		E	< median IP1		E	Gain
	Yes	No		Yes	No		
2.03	0	1	0	11	8	0.982	0.06
2.08	0	2	0	11	7	0.964	0.125
2.6	0	3	0	11	6	0.937	0.197
2.68	0	5	0	11	4	0.837	0.365
2.73	1	7	0.544	10	2	0.65	0.385
2.78	3	7	0.881	8	2	0.722	0.191
2.85	4	7	0.946	7	2	0.764	0.129
2.93	6	7	0.996	5	2	0.863	0.043
3.1	8	7	0.997	3	2	0.971	0.002
3.13	9	7	0.989	2	2	1	0.002
3.15	9	8	0.998	2	1	0.918	0.007
3.2	10	8	0.991	1	1	1	8E04
3.23	10	9	0.998	1	0	0	0.045
3.3	11	9	0.993	0	0	0	0

Tabel 3 lanjutan Perhitungan Entropy dan Gain Rekursi level 0

IP2	≤ median IP2		E	> median IP2		E	Gain
	Yes	No		Yes	No		
1.9	0	1	0	11	8	0.982	0.06
2.12	1	1	1	10	8	0.991	8E04
2.22	1	2	0.918	10	7	0.977	0.024
2.28	1	3	0.811	10	6	0.954	0.067
2.52	3	3	1	8	6	0.985	0.003
2.53	3	4	0.985	8	5	0.961	0.023
2.55	3	5	0.954	8	4	0.918	0.06
2.57	3	6	0.918	8	3	0.845	0.115
2.6	4	6	0.971	7	3	0.881	0.067
2.64	5	6	0.994	6	3	0.918	0.033
2.68	5	7	0.98	6	2	0.811	0.08
2.94	6	7	0.996	5	2	0.863	0.043
2.95	7	8	0.997	4	1	0.722	0.065
3.06	8	8	1	3	1	0.811	0.031
3.13	9	8	0.998	2	1	0.918	0.007
3.15	10	8	0.991	1	1	1	8E04
3.21	10	9	0.998	1	0	0	0.045
3.27							

Pd.OT	yes	No	E	Gain
SD	5	2	0.863	0.115
SMA	3	3	1	
Dipl	1	0	0	
S-1	2	4	0.918	

Pk.OT	yes	No	E	Gain
PT	3	1	0.811	0.117
WS	3	3	1	
PN	2	4	0.918	
PS	1	0	0	
P	2	1	0.918	

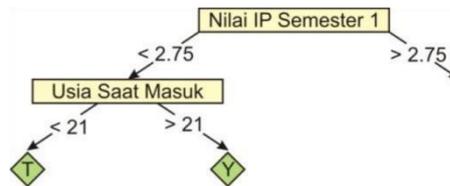
TTOT	yes	No	E	Gain
MR	2	0	0	0.146
IT	6	5	0.994	
PI	2	1	0.918	
LP	1	3	0.811	

lagi karena atribut *parent* tidak boleh dijadikan *node* selanjutnya.

Tabel 3 Data Training Rekursi Level 1 Iterasi Ke-1

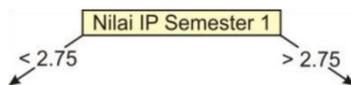
Nama	Usia	TTOT	Pd. OT	Pk. OT	IP1	IP 2	<5?
C	19	LP	S-1	PN	2.68	2.95	tidak
E	20	LP	SD	WS	2.73	1.9	tidak
F	22	JT	SMA	PS	2.73	2.64	ya
K	19	JT	SD	WS	2.03	2.53	tidak
O	18	JT	S-1	PN	2.68	2.55	tidak
P	19	JT	S-1	PN	2.08	2.22	tidak
Q	18	LP	SMA	PT	2.73	2.28	tidak
S	19	PJ	SMA	WS	2.6	2.57	tidak

Seperti pada rekursi sebelumnya, nilai *Entropy parent* dihitung lebih dahulu. dilakukan iterasi lagi.



Gambar 8 Pohon Keputusan Rekursi Level 1 Iterasi ke-1

IP1 dengan titik 2.75, sehingga IP1 harus dijadikan *root node* dengan *branch* ≤ 2.75 dan > 2.75 . Pohon sementara yang terbentuk adalah



Gambar 7 Pohon Keputusan Rekursi Level 0

Karena hasil masih heterogen, maka harus dilakukan perhitungan lagi dengan data yang sudah dikelompokkan menurut kolom Nilai IP Semester 1 dengan nilai ≤ 2.75 dan > 2.75 . Atribut Nilai IP Semester 1 tidak perlu dihitung

Pada rekursi level 1 iterasi ke-2, data yang dikelompokkan adalah nilai IP Semester 1 yang lebih dari 2.75.

password terlebih dahulu. Admin menambahkan *username* dan *password user* lain.



Gambar 11 Form Login

Jika *login* sebagai admin, maka akan muncul menu lengkap yang terdiri dari: Data Admin, Data User, Data Mahasiswa, Proses Mining Rule ID3, Informasi Gain dan Entropy, dan Uji Klasifikasi. Jika *login* sebagai *user*, maka hanya akan tampil satu menu, yaitu Uji Klasifikasi

2. Form Pengolahan Data Admin



Gambar 12 Form Pengolahan Data Admin

Form Data Admin berfungsi untuk menambah, mencari, mengubah, dan menghapus data Admin.

Form Pengolahan Data User



Gambar 13 Form Pengolahan Data User

Form Data User berfungsi untuk menambah, mencari, mengubah, dan menghapus data User.

4. Form Pengolahan Data Mahasiswa



Gambar 14 Form Pengolahan Data Mahasiswa

Form Data Mahasiswa berfungsi untuk menambah, mencari, mengubah, dan menghapus data Mahasiswa.

Form Mining ID3



Gambar 15 Form Mining ID3

Proses mining ID3 berfungsi menghitung Entropy dan Gain untuk menghasilkan rule yang akan digunakan untuk proses uji klasifikasi. Output dari proses ini adalah rule yang terbentuk dari proses perhitungan Gain dan Entropy.

Form Perhitungan Gain dan Entropy



Gambar 16 Form Perhitungan Gain dan Entropy

Proses dalam form ini sama seperti proses mining ID3. Yang membedakan adalah outputnya. Output di form ini adalah rule yang terbentuk beserta nilai entropy dan gain.

Form Uji Klasifikasi



Gambar 17 Form Uji Klasifikasi

Form ini digunakan oleh mahasiswa yang ingin memprediksi berapa lama masa studi mahasiswa. Pertanyaan di form ini didasarkan pada proses mining ID3 yang telah dilakukan sebelumnya. Output di form ini berupa hasil prediksi kelulusan kurang dari 5 tahun atau lebih dari 5 tahun.

Pengujian

Dari 140 buah data training yang telah diproses, maka akan dilakukan pengujian dengan 20 buah data uji yang diambil secara acak. Hasil sebenarnya pada data uji akan dibandingkan dengan hasil prediksi untuk mengetahui tingkat keakuratan klasifikasi.

Tabel 8 Perbandingan Masa Studi Pada Data Uji dengan Hasil Prediksi sebanyak 20 data.

No	Usia	TT OT	Pd.OT	Pk. OT	IP 1	IP 2	Masa Studi	Hasil Prediksi	Hasil Kebenaran
1	18	JT	SMA	WS	3.1	2.59	<5	<5	B
2	19	MR	S-1	PN	3.93	3.85	<5	<5	B
3	18	JT	SMP	WS	3.18	2.48	>5	>5	B
4	20	JT	SD	PT	2.7	2.69	>5	>5	B
5	19	MR	SMA	PT	2.78	2.71	>5	>5	B
6	18	MR	SMP	WS	2.85	2.79	>5	>5	B
7	18	MR	SMP	PS	2.58	1.98	>5	>5	B
8	20	MR	SMA	PN	2.45	0.93	>5	>5	B
9	18	MR	S-1	PN	3.43	2.8	<5	>5	S
10	18	JT	S-1	PN	3.3	2.73	>5	>5	B
11	20	MR	SMA	WS	3.35	2.48	>5	>5	B
12	18	JT	S-1	PN	2.5	2.38	>5	>5	B
13	18	JT	D	PN	2.88	3.1	>5	>5	B
14	18	JT	SD	PT	3.45	2.43	<5	<5	B
15	20	JT	D	PN	3.33	2.67	<5	>5	S
16	19	JT	D	WS	2.98	2.5	<5	<5	B
17	19	JT	S-2	PN	2.95	2.29	<5	<5	B
18	19	MR	SMA	WS	3.5	3.09	>5	>5	B
19	19	LP	SMA	PT	2.33	2.42	>5	>5	B
20	18	MR	SMA	WS	3.48	2.65	<5	>5	S

5. KESIMPULAN

Algoritma ID3 dapat digunakan untuk memprediksi masa studi

mahasiswa. Hal yang paling berpengaruh dalam masa studi mahasiswa adalah nilai IP semester 2. Dari 140 *data training* dan 20 *data testing* dengan 6 macam atribut input (Usia Saat Masuk, Tempat Tinggal Orang Tua, Pendidikan Orang Tua, Pekerjaan Orang Tua, Nilai IP Semester 1, Nilai IP Semester 2) dan 2 macam atribut target (Masa studi < 5 tahun dan masa studi > 5 tahun), dapat diperoleh tingkat keakuratan hasil prediksi ID3 sebesar 85%, sedangkan *error rate* hasil prediksi sebesar 15%.

6. REFERENSI

- [1] Basuki, Achmad dan Syarif, Iwan. Pohon keputusan. <http://lecturer.eepis-its.edu/~basuki/lecture/DecisionTree.pdf>. 2003.
- [2] Fayyad, Usama. Advances in Knowledge Discovery and Data Mining. MIT Press, 1996.
- [3] Gorunescu, Data Mining: Concepts, Models and Technique, Berlin: Springer, 2011.
- [4] Linoff, Gordon S. dan Berry, Michael J.A. Data Mining Techniques for Marketing, Sales, Customer Relationship Management. New York: Wiley Publishing, Inc, 2005.
- [5] Larose, Daniel T. Discovering Knowledge in Data: An Introduction to Data Mining. New York: John Wiley & Sons Inc, 2005.
- [6] Pramudiono, Iko, Pengantar Data Mining Menambang Permata Pengetahuan Di Gunung Data. Yogyakarta: Andi, 2006.
- [7] Romansyah, F., Sitanggang I. S., dan Nurdiati, S. Fuzzy Decision Tree dengan Algoritma ID3 pada Data Diabetes. *Internetworking Indonesia Journal* Vol. 1/No. 2 (2009). 1:

