

DATA MINING UNTUK MENGANALISA PREDIKSI MAHASISWA BERPOTENSI NON-AKTIF MENGGUNAKAN METODE *DECISION TREE C4.5*

Dwi Untari A11.2010.05410
Program Studi Teknik Informatika –S1
Fakultas Ilmu Komputer
Universitas Dian Nuswantoro, Jl. Nakula 1 No. 5-11. Semarang
dwi.unthaa@gmail.com

ABSTRAK

Keberhasilan dan kegagalan studi mahasiswa merupakan cerminan kualitas suatu perguruan tinggi. Mahasiswa non-aktif adalah mahasiswa yang tidak melakukan registrasi pada awal semester atau tidak mengikuti perkuliahan sekurang-kurangnya satu semester. Keberadaan mahasiswa non-aktif ini tentunya berpengaruh terhadap mahasiswa lulus tidak tepat waktu, semakin banyak mahasiswa non-aktif maka semakin banyak juga mahasiswa lulus tidak tepat waktu selain itu mahasiswa non-aktif dapat meningkatkan jumlah mahasiswa berpotensi *drop out*. Lulus tepat waktu merupakan elemen penilaian akreditasi perguruan tinggi. Oleh karena itu untuk mengatasi masalah tersebut diterapkan *data mining* dengan metode C.4.5 untuk mencari karakteristik mahasiswa berpotensi non-aktif. Desain penelitian ini menggunakan CRISP-DM dan penelitian menggunakan data mahasiswa S1 Fakultas Ilmu Komputer Universitas Dian Nuswantoro. Proses validasi yang digunakan adalah *split validation*, dan untuk pengujian model menggunakan *confusion matrix*. Hasil menunjukkan nilai akurasi terbaik adalah 97.60% dengan *ratio data training* 90%. Berdasarkan percobaan yang dilakukan, mahasiswa dengan IPS Semester 3 adalah dibawah 2.60, SKS semester 3 yang diambil adalah <20 sks, SKS semester 4 adalah kosong dan IPS semester 4 adalah kosong lebih besar berpotensi menjadi mahasiswa non-aktif.

Kata Kunci : Mahasiswa non-aktif, *decision tree C4.5*, data mining, data mining klasifikasi.

I. PENDAHULUAN

Tingginya tingkat keberhasilan mahasiswa dan rendahnya tingkat kegagalan mahasiswa merupakan cermin kualitas dari suatu perguruan tinggi. Kegagalan mahasiswa dan faktor-faktor penyebabnya merupakan topik yang menarik untuk diteliti.

Saat ini intitusi perguruan tinggi berada dalam lingkungan yang sangat kompetitif. Setiap perguruan tinggi berusaha untuk terus memperbaiki menejemennya untuk meningkatkan mutu pendidikan dan meningkatkan akreditasi. Salah satu elemen penilaian akreditasi perguruan tinggi adalah lulus tepat waktu[1]. Selain itu wisuda tepat waktu merupakan isu yang penting karena tingkat kelulusan sebagai dasar efektifnya suatu perguruan tinggi[2].

Salah satu persoalan yang masih menjadi bahan pembicaraan tentang kegagalan mahasiswa adalah mengenai mahasiswa berhenti studi (*drop out* /

non-aktif)[3] dan kelulusan. Mahasiswa non-aktif (mangkir) adalah mahasiswa yang tidak melakukan registrasi pada awal semester atau tidak mengikuti perkuliahan sekurang-kurangnya satu semester[4]. Universitas Dian Nuswantoro merupakan salah satu perguruan tinggi terbesar di Jawa Tengah yang memiliki 13.416 mahasiswa dan 4.138 berstatus non-aktif [5]. Keberadaan mahasiswa non-aktif ini tentunya berpengaruh terhadap mahasiswa lulus tidak tepat waktu, semakin banyak mahasiswa non-aktif maka semakin banyak juga mahasiswa lulus tidak tepat waktu[5]. Menurut [4], mahasiswa yang memiliki status non-aktif selama empat semester berturut-turut adalah mahasiswa *drop out*[4].

Berdasarkan uraian diatas, mahasiswa non-aktif berpotensi menimbulkan masalah yaitu kelulusan tidak tepat waktu dan berpotensi *drop out* yang dapat

menurunkan mutu pendidikan dan akreditasi perguruan tinggi. Tidak adanya informasi terkait mahasiswa non-aktif dan data yang cukup menyebabkan tidak diketahui karakteristik mahasiswa non-aktif. Oleh karena itu, diperlukan analisa terhadap faktor-faktor penyebab mahasiswa non-aktif sehingga diketahui karakteristik mahasiswa non-aktif.

Dewasa ini *data mining* berkembang digunakan untuk menyelesaikan masalah menyangkut pendidikan. Beberapa penelitian terkait *data mining* digunakan sebagai upaya meningkatkan akreditasi perguruan tinggi [6][7][5]

Data mining adalah teknik bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis data yang disimpan[8]. Salah satu teknik *data mining* adalah teknik klasifikasi. Teknik klasifikasi adalah teknik pembelajaran untuk prediksi suatu nilai dari target variabel kategori[9]. Karena dalam prediksi mahasiswa non-aktif ini memiliki dua variable target yaitu aktif dan non-aktif maka penelitian ini akan menggunakan *data mining* teknik klasifikasi.

Algoritma klasifikasi yang cukup populer adalah *decision tree*, *naïve bayes*, dan *neural network*[10]. Beberapa algoritma klasifikasi *data mining* telah digunakan untuk memprediksi perilaku mahasiswa yang berpotensi *non-aktif* diantaranya *decision tree*, *neural network*, *naïve bayes*, *instance-based learning*, *logistic regression*, *support vector machine*, *K-Nearest Neighbor* dan jaringan saraf tiruan [6][5][3]. Penelitian yang dilakukan [5] dengan melakukan uji komparasi algoritma klasifikasi yaitu *logistic regression*, *naïve bayes*, *neural network*, dan *decision tree* menunjukkan bahwa *decision tree* memiliki performansi yang paling tinggi. *Decision tree* terbukti memiliki performansi yang bagus dibanding algoritma klasifikasi lain. Algoritma *decision tree* bekerja dengan membentuk pohon keputusan yang dapat disimpulkan aturan-aturan klasifikasi tertentu, salah satunya adalah C4.5. Algoritma *decision tree* ini memiliki beberapa kelebihan diantaranya adalah[10] :

1. Mudah untuk diinterpretasikan.

2. Tidak membutuhkan biaya yang mahal saat membangun algoritma ini
3. Dapat menggunakan data numerik dan kategorikal
4. Proses pengambilan keputusan dapat dipahami dengan mudah
5. Stabil dan cepat apabila digunakan dalam data set yang besar.

Oleh karena itu dalam penelitian ini akan dilakukan analisa prediksi mahasiswa non-aktif menggunakan metode klasifikasi dengan algoritma *decision tree* C4.5.

II. METODE YANG DIUSULKAN

A. Decision Tree

Decision tree merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode *decision tree* mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami.

Proses pada *decision tree* adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule [15].

Saat menyusun sebuah *decision tree* pertama yang harus dilakukan adalah menentukan atribut mana yang akan menjadi simpul akar dan atribut mana yang akan menjadi simpul selanjutnya. Pemilihan atribut yang baik adalah atribut yang memungkinkan untuk mendapatkan *decision tree* yang paling kecil ukurannya. Atau atribut yang bisa memisahkan obyek menurut kelasnya. Secara heuristik atribut yang dipilih adalah atribut yang menghasilkan simpul yang paling "purest" (paling bersih). Ukuran purity dinyatakan dengan tingkat impurity, dan untuk menghitungnya, dapat dilakukan dengan menggunakan konsep Entropy, Entropy menyatakan impurity suatu kumpulan objek. Jika diberikan sekumpulan objek dengan label/output y yang terdiri dari objek berlabel 1, 2 sampai n , Entropy dari objek dengan n kelas ini dapat dihitung dengan rumus berikut.

$$Entropy(y) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 \dots - p_n \log_2 p_n$$

1. Kemudian setelah menghitung Entropy, hitung

Information Gain :

2. *Information gain* adalah kriteria yang paling populer untuk pemilihan atribut. *Information gain* dapat dihitung dari output data atau variabel dependent y yang dikelompokkan berdasarkan atribut A , dinotasikan dengan gain (y,A) . *Information gain*, gain (y,A) , dari atribut A relatif terhadap output data y adalah :

Dimana nilai (A) adalah semua nilai yang mungkin dari atribut A , dan y_c adalah subset dari y dimana A mempunyai nilai c .

Algoritma C4.5 sendiri merupakan pengembangan dari algoritma ID3, dimana pengembangan dilakukan dalam hal: bisa mengatasi missing data, bisa mengatasi data kontinyu, pruning. Secara umum langkah algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

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

B. Proses Tahapan Data Mining

Istilah *data mining* dan *knowledge discovery databases* (KDD) sering digunakan secara bergantian untuk menjelaskan proses penggalian informasi dalam *database*. Secara garis besar proses KDD dapat dijelaskan sebagai berikut[9]:

1. *Data Selection*

Menyeleksi data dari sekumpulan data operasional sebelum dilakukan tahap penggalian informasi. Data hasil seleksi yang akan digunakan disimpan terpisah dari database operasional.

2. *Cleaning*

Proses *cleaning* ini mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

3. *Transformation*

Beberapa teknik *data mining* membutuhkan format data khusus sebelum

diaplikasikan oleh karena itu dilakukan transformasi sehingga data tersebut sesuai untuk proses *data mining*. Transformasi dan pemiliha data ini juga menentukan kualitas hasil dari *data mining*.

4. *Data Mining*

Dengan menggunakan teknik atau metode tertentu yang diaplikasikan pada data terpilih untuk mencari pola atau informasi dari data tersebut.

5. *Evaluation / Interpretation*

Pola atau informasi yang dihasilkan dari proses data mining ini disajikan dalam bentuk yang mudah dipahami. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya[9][9].

C. Confusion Matrix

Confusion matrix adalah *tool* yang digunakan untuk evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah [10]. Sebuah *matrix* dari prediksi yang akan dibandingkan dengan kelas yang asli dari inputan atau dengan kata lain berisi informasi nilai aktual dan prediksi pada klasifikasi.

Tabel 1. Tabel *Confusion Matrix* 2 kelas

<i>Classification</i>	Predicted class	
	Class = Yes	Class = No
Class = Yes	a (<i>true positive-TP</i>)	b (<i>false negative-FN</i>)
Class = No	c (<i>false positive-FP</i>)	d(<i>true negative-TN</i>)

Rumus untuk menghitung tingkat akurasi pada matrik adalah :

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} = \frac{A+D}{A+B+C+D}$$

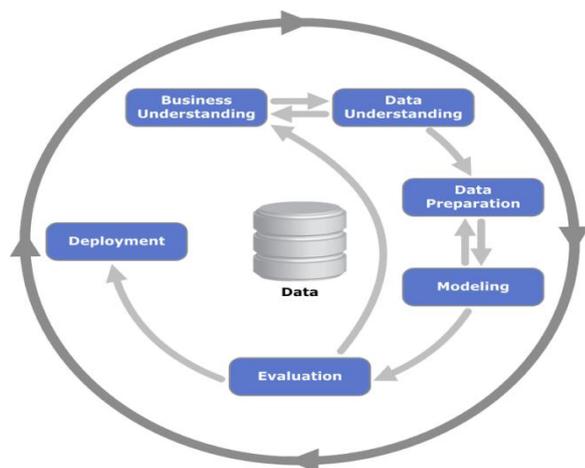
D. Split Validation

Split Validation adalah teknik validasi yang membagi data menjadi dua bagian secara acak, sebagian sebagai *data training* dan sebagian lainnya sebagai *data testing*. Dengan menggunakan *Split Validation* akan dilakukan percobaan *training* berdasarkan *split ratio* yang telah ditentukan sebelumnya, untuk kemudian sisa dari *split ratio data training* akan dianggap sebagai *data testing*. *Data training* adalah data yang akan dipakai dalam melakukan pembelajaran sedangkan *data testing* adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian kebenaran atau keakuratan hasil pembelajaran[16].

Tabel 2. Ilustrasi Split Validation

Training 90%	Test 10%
Training 80%	Test 20%
Training 70%	Test 30%
Training 60%	Test 40%
Training 50%	Test 50%
Training 40%	Test 60%
Training 30%	Test 70%
Training 20%	Test 80%
Training 10%	Test 90%

III. Desain Penelitian



Gambar 1. CRISP-DM

Penelitian ini adalah menggunakan model CRISP-DM (*Cross Industry Standard Process for Data Mining*), CRISP-DM adalah standar proses *data mining* sebagai strategi pemecahan masalah secara umum dari bisnis atau untuk penelitian[9]. Dalam standar ini proses *data mining* memiliki *life cycle* yang terbagi dalam enam fase.

a. Pemahaman Bisnis (*Bussiness Understanding*)

Saat ini institusi perguruan tinggi dituntut untuk memiliki keunggulan dalam bersaing dengan meningkatkan mutu dan akreditasinya. Dalam struktur pendidikan saat ini, mahasiswa memiliki peran penting bagi sebuah institusi pendidikan. Kegagalan studi mahasiswa menjadi tugas pihak perguruan tinggi untuk memperbaiki manajemen. Salah satu elemen penilaian akreditasi adalah lulus tepat waktu[1]. Mahasiswa non-aktif adalah mahasiswa yang tidak melakukan registrasi pada awal semester atau tidak mengikuti perkuliahan sekurang-kurangnya satu semester[4]. Semakin banyak mahasiswa non-aktif berarti semakin banyak pula mahasiswa lulus tidak tepat waktu yang berarti akan mempengaruhi akreditasi perguruan tinggi. Oleh karena itu perlu adanya monitoring serta evaluasi terhadap kecenderungan mahasiswa non-aktif sehingga nantinya dapat mengurangi jumlah mahasiswa non-aktif.

b. Pemahaman Data (*Data Understanding*)

Data yang digunakan dalam penelitian ini adalah sumber data sekunder. Data diperoleh dari database mahasiswa yang dimiliki oleh Universitas Dian Nuswantoro Semarang, yaitu melalui bagian Pusat Sistem Informasi (PSI) Universitas Dian Nuswantoro. Data yang dikumpulkan adalah data mahasiswa Fakultas Ilmu Komputer dengan program studi strata satu (S1) untuk tahun angkatan 2009-2011 dengan atribut nim (nomor induk mahasiswa), nama, program studi, umur, jenis kelamin, status marital, agama, status pekerjaan, ip (indeks prestasi) semester 1 sampai dengan ip semester 8, kota asal, sekolah asal, jalur masuk, sks semester 1 sampai dengan semester 8 dengan label keterangan aktif dan non-aktif.

c. Pengolahan Data (*Data Preparation*)

Pada tahap ini adalah pemilihan atribut data yang akan digunakan untuk pemodelan.

Tabel 3. Seleksi Atribut

Atribut	Detail Penggunaan	
NIM	√	ID
Progdi	√	No
Umur Masuk	√	Nilai Model
JK	×	No
Marital	√	Nilai Model
Status	√	Nilai Model
Kota Asal	√	Nilai Model
Agama	×	No
Asal Sekolah	√	Nilai Model
Jalur Masuk	√	Nilai Model
Status	√	Nilai Model
IPS1	√	Nilai Model
IPS2	√	Nilai Model
IPS3	√	Nilai Model
IPS4	√	Nilai Model
IPS5	×	No
IPS6	×	No
IPS7	×	No
IPS8	×	No
IPS9	×	No
SKS1	√	Nilai Model
SKS2	√	Nilai Model
SKS3	√	Nilai Model
SKS4	√	Nilai Model
KET	√	Label target

Tabel diatas menjelaskan mengenai atribut-atribut yang akan digunakan dalam penelitian, indikator yes (√) menandakan bahwa atribut bersangkutan akan digunakan dalam penelitian, sedangkan indikator no (×) menandakan atribut tersebut akan dieliminasi pada tahap *data preparation*.

Data dengan atribut yang telah dipilih kemudian dikonversikan untuk memudahkan proses *mining*, karena data akan diolah dengan *tools* bantu data mining. Berikut adalah atribut yang akan dilakukan proses konversi :

a. Umur

Atribut umur memiliki nilai tipe data integer dengan banyak nilai akan dikonversikan menjadi dua nilai kategori yaitu “ ≥ 20 ” dan “ <20 ”.

b. Kota Asal

Atribut kota asal awalnya berisikan nama kota asal mahasiswa dan akan dikonversikan menjadi dua nilai kategori yaitu “Semarang”, “Luar Semarang Jateng” dan “Luar Jateng”.

c. Asal Sekolah

Atribut asal sekolah sebelumnya berisikan alamat sekolah asal mahasiswa kemudian dikonversikan menjadi tiga nilai kategori yaitu “Negeri”, “Swasta” dan “Transfer”.

d. IPS 1-4

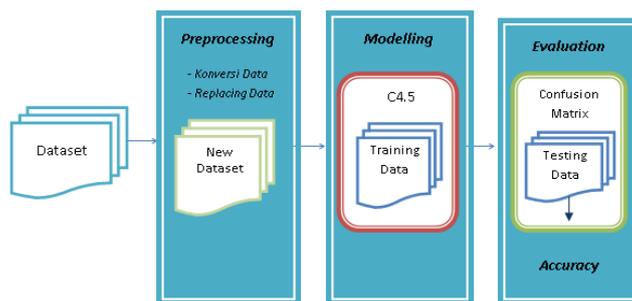
Atribut IPS1, IPS2, IPS3, IPS4 sebelumnya berisikan indek prestasi mahasiswa tiap semester dalam tipe data float dengan banyak nilai dan dikonversikan menjadi 3 nilai kategori yaitu “0 - 1.30”, “1.31 – 2.60” dan “2.61 - 4.00”.

e. SKS1-4

Atribut SKS1, SKS2, SKS3 DAN SKS4 sebelumnya berisikan satuan kredit semester yang diambil tiap semester oleh mahasiswa dalam tipe data integer dengan banyak nilai kemudian akan dikonversikan menjadi dua nilai kategori yaitu “ ≥ 20 ” dan “ <20 ”.

d. Pemodelan (*Modelling*)

Metode yang akan digunakan dalam penelitian ini adalah Algoritma C4.5. Untuk melakukan pengukuran akurasi dalam penelitian ini akan menggunakan *framework* RapidMiner Studio 6.0.003



Gambar 2. Model Penelitian

e. Validasi dan Evaluasi

Dalam tahapan ini akan dilakukan validasi serta pengukuran keakuratan hasil yang dicapai oleh model menggunakan beberapa teknik yang terdapat dalam *framework* RapidMiner Studio 6.0.003 yaitu *Confusion Matrix* untuk pengukuran tingkat akurasi model, dan *Split Validation* untuk validasi.

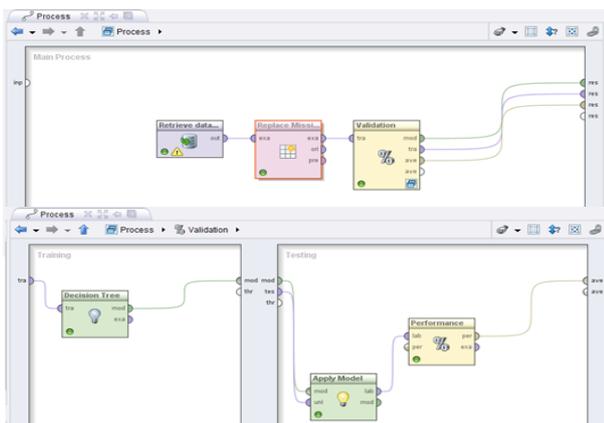
f. Penyebaran (Deployment)

Hasil dari penelitian ini berupa analisa yang mengarah ke *Decision Support System* (DSS), yang diharapkan dapat digunakan oleh institusi perguruan tinggi sebagai bahan pertimbangan dalam menentukan langkah guna mengatasi permasalahan mahasiswa non-aktif dan juga dapat digunakan sebagai bahan rujukan untuk penelitian selanjutnya. Selain itu hasil analisa ini juga akan digunakan sebagai dasar perancangan sebuah sistem pengambilan keputusan guna melakukan identifikasi mahasiswa non-aktif.

IV. Hasil dan Pembahasan

a. Validasi dan Evaluasi

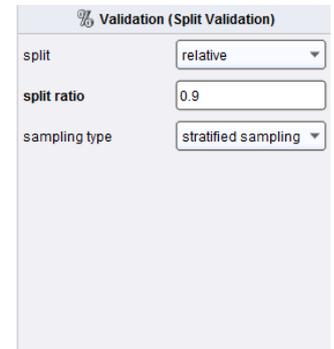
Tujuan utama dari penelitian ini adalah menganalisa prediksi mahasiswa berpotensi non-aktif dengan menerapkan teknik klasifikasi data mining dengan algoritma *decision tree C4.5*. Pada tahap pengujian model ini data yang digunakan telah melewati tahap *preprocessing*. Berikut adalah desain model yang akan digunakan.



Gambar 3. Desain Model Metode C4.5

1. *Retreiving Data* : Operator ini digunakan untuk mengimport *dataset* yang akan digunakan, pada penelitian ini data diimport dari *file excel*.
2. *Replace Missing Value* : Operator ini digunakan untuk mengganti nilai yang kosong yang ada pada *dataset*. Pada penelitian ini menggunakan model *average* yaitu mengganti nilai yang kosong dengan nilai rata-rata yang muncul.

3. *Validation* : Metode *validation* yang digunakan dalam penelitian adalah *Split Validation*, *validation* ini hanya membagi total dari keseluruhan *dataset* menjadi *data training* dan *data testing* berdasar ratio yang telah ditentukan.



Gambar 4. Penggunaan ratio data training

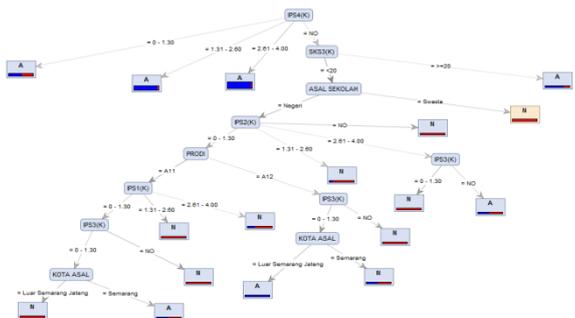
4. *Decision Tree* : Metode klasifikasi yang digunakan dalam penelitian ini.
5. *Apply Model* : Operator yang digunakan menjalankan metode yang digunakan dalam penelitian ini C4.5
6. *Perfoman ce* : Operator yang digunakan untuk mengukur performa akurasi dari model.

b. Hasil Percobaan
Percobaan 1

Tabel 4. Akurasi Percobaan 1

Percobaan 1	Ratio Data Training (%)								
	10	20	30	40	50	60	70	80	90
Akurasi(%)	86.03	90.01	88.45	89.67	91.56	90.81	91.92	92.38	92.79
Error rate (%)	13.97	9.99	11.55	10.33	8.44	9.19	8.08	7.62	7.21
Waktu	0 sec	0 sec	0 sec	1 sec	0 sec				

Tabel diatas menunjukkan hasil akurasi dari percobaan 1 dengan *ratio data training* yang berbeda. Pada percobaan 1 akurasi terbaik adalah 92.38% dengan ratio data training 90%.



Gambar 5. Pohon keputusan percobaan 1

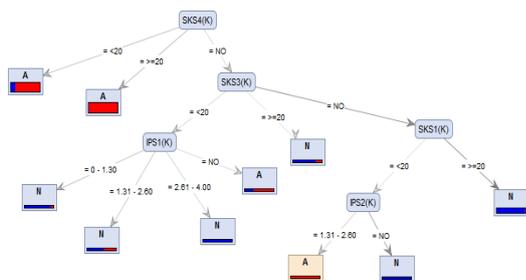
Gambar 5 merupakan pohon keputusan yang terbentuk dari percobaan 1.

Percobaan 2

Tabel 5. Akurasi Percobaan 2

Percobaan 2	Ratio Data Training (%)								
	10	20	30	40	50	60	70	80	90
Akurasi(%)	90.32	90.98	89.45	89.76	89.37	88.38	89.47	89.63	90.00
Error rate(%)	9.68	9.02	10.55	10.24	10.23	11.62	10.53	10.37	10.00
Waktu	0 sec	0 sec	0 sec	0 sec	0 sec	0 sec	0 sec	0 sec	1 sec

Tabel diatas menunjukkan hasil akurasi dari percobaan 2 dengan *ratio data training* yang berbeda. Pada percobaan 2 akurasi terbaik adalah 90.32% dengan ratio data training 20%.



Gambar 6. Pohon Keputusan percobaan 2

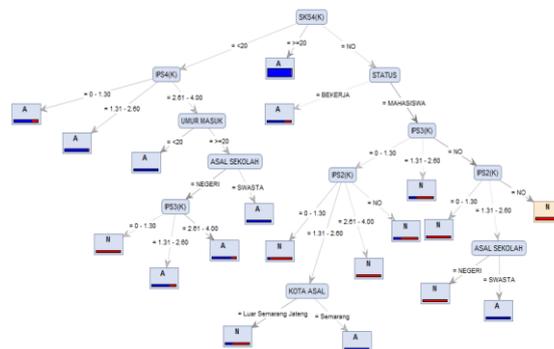
Gambar 6 merupakan pohon keputusan yang dihasilkan percobaan 2.

Percobaan 3

Tabel 6. Akurasi Percobaan 3

Percobaan 3	Ratio Data Training (%)								
	10	20	30	40	50	60	70	80	90
Akurasi(%)	89.94	92.96	92.96	93.25	95.24	93.45	94.16	94.84	97.60
Error rate(%)	10.06	7.04	7.04	6.75	4.76	6.55	5.84	5.16	2.40
Waktu	1 sec	0 sec	0 sec	0 sec	0 sec	0 sec	0 sec	0 sec	0 sec

Tabel diatas menunjukkan hasil akurasi dari percobaan 3 dengan *ratio data training* yang berbeda. Pada percobaan 3 akurasi terbaik adalah 97.60% dengan *ratio data training* 90%.



Gambar 6. Pohon Keputusan percobaan 3

Gambar 6 merupakan pohon keputusan yang terbentuk dari percobaan 3.

V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan maka dapat diambil beberapa kesimpulan, antara lain:

1. *Ratio data training* yang digunakan mempengaruhi nilai akurasi pada setiap percobaan. Pada percobaan ke- 1 nilai akurasi terbaik adalah 92.79% dengan *ratio data training* 90%. Pada percobaan ke-2 nilai akurasi terbaik adalah 90.98% dengan *ratio data training* 20%. Pada percobaan ke-3 nilai akurasi terbaik adalah 97.60% dengan *ratio data training* 90%.
2. Variabel yang paling berpengaruh terhadap mahasiswa non-aktif adalah Indeks Prestasi Semester(IPS) dan Satuan Kredit Semester (SKS) yang diambil. Berdasarkan percobaan yang dilakukan mahasiswa dengan IPS Semester 3 adalah dibawah 2.60, SKS semester 3 yang diambil adalah <20 sks, SKS semester 4 adalah NO (kosong) dan IPS semester 4 adalah NO (kosong) adalah mahasiswa yang berpotensi non-aktif.

VI. DAFTAR PUSTAKA

- [1] BANPT, Buku I Naskah Akademik Akreditasi Instistusi Perguruan Tinggi. 2007.
- [2] M. N. Quadri and N. V Kalyankar, "Drop Out Feature of Student Data for Academic Performance Using Decision Tree Techniques," *Glob. J. Comput. Sci. Technol.*, vol. 10, no. 2, pp. 2–5, 2010.
- [3] M. M. Hidayat, D. Purwitasari, and H. Ginardi, "Analisis Prediksi DO Mahasiswa Dalam Educational Data Mining Mengguakan Jaringan Syaraf Tiruan," *J. IPTEK*, vol. 17, no. 2, 2013.
- [4] Keputusan Rektor Universitas Dian Nuswantoro nomor: 075/KEP/UDN-01/IV/2009 tentang Peraturan Akademik Universitas Dian Nuswantoro tahun akademik 2009/2010. .
- [5] K. Hastuti, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Mahasiswa Non-Aktif," no. Semantik, 2012.
- [6] I. Ernawati, "Prediksi Status Keaktifan Studi Mahasiswa Dengan Algoritma C5.0 dan K-Nearest Neighbor," 2008.
- [7] C. M. Vera, C. Romero, and S. Ventura, "Predicting School Failure Using Data Mining," *J. Educ. Data Min.*, 2011.
- [8] E. Prasetyo, Data Mining Konsep dan Aplikasi menggunakan MATLAB.pdf. Yogyakarta: Andi Offset, 2012.
- [9] Kusriani and E. T. Luthfi, Algoritma Data Mining. Yogyakarta: Andi Offset, 2009.
- [10] F. Gorunescu, *Data Mining Concept, Models and Techniques*. Verlag Berlin Heidelberg: Springer, 2011.
- [11] J. Han and M. Kimber, *Data Mining Concepts and Techniques 2nd edition*. San Fransisco: Mofgan Kaufhann, 2006.
- [12] B. Santoso, Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Surabaya: Graha Ilmu, 2007.
- [13] Larose D T, *Discovering Knowladge in Database*. New Jersey: John Wiley and Sons Inc, 2005.
- [14] M. J. A. Berry and G. S. Linoff, *Data Mining Techniques For Marketing, Sales, Customer Relationship Management Second Editon*. United States of America: Wiley Publishing Inc, 2004.
- [15] A. Basuki and I. Syarif, *Decision Tree*. Surabaya: Politeknik Elektronika Negeri Surabaya - ITS, 2003.
- [16] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining Practical Machine Learning Tools and Technique*. Burlington: Morgan Kaufmann Publisher, 2011.