

Decision Tree Learning Untuk Penentuan Jalur Kelulusan Mahasiswa

Winda Widya Ariestya¹, Yulia Eka Praptiningsih², Wahyu Supriatin³

Program Studi Sistem Informasi

Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Gunadarma

Jalan Margonda Raya 100, Depok 16424

Email : ¹winda_widya@staff.gunadarma.ac.id, ²yulia_eka@staff.gunadarma.ac.id,

³ayu_ws@staff.gunadarma.ac.id

ABSTRAK

Jalur kelulusan merupakan hal penentu seorang mahasiswa untuk memperoleh gelar jenjang pendidikan strata satu pada sebuah Perguruan Tinggi. Penelitian ini bertujuan untuk mengetahui pemanfaatan algoritma ID3 dalam penentu jalur kelulusan serta menghasilkan rule atau aturan pada penentuan jalur kelulusan mahasiswa tingkat akhir. Metode yang digunakan adalah Iterative Dichotomiser 3 (ID3) dengan menggunakan alat bantu perangkat lunak Rapidminer. Dari hasil penelitian diperoleh 7 rule atau aturan dalam penentuan jalur kelulusan yaitu 3 rule untuk jalur skripsi dan 4 rule untuk jalur non skripsi dan dapat dikatakan algoritma ID3 dapat dimanfaatkan dalam penentuan jalur kelulusan dengan nilai akurasi antara 0,85-1,00.

Kata Kunci: Decision Tree Learning, ID3, Jalur kelulusan

1. PENDAHULUAN

Decision tree learning adalah suatu metode belajar yang sangat populer dan banyak digunakan secara praktis. Metode ini merupakan metode yang berusaha menemukan fungsi-fungsi pendekatan yang bernilai diskrit dan tahan terhadap data-data yang terdapat kesalahan (noise data) serta mampu mempelajari ekspresi-ekspresi disjunctive. *Iterative Dichotomiser 3 (ID3)*, *Assistant* dan *C4.5* merupakan jenis dari *decision tree learning*. Dalam membangun *decision tree learning* dibutuhkan evaluasi semua atribut yang ada menggunakan suatu ukuran statistik untuk mengukur efektifitas suatu atribut dalam mengklasifikasikan kumpulan sampel data. Dalam hal ini information gain adalah yang paling banyak digunakan (Suyanto, 2011).

Jalur kelulusan merupakan tahap akhir yang harus dilalui seorang mahasiswa dalam menyelesaikan pendidikan di perguruan tinggi. Pada salah satu Perguruan Tinggi Swasta (PTS) di Jakarta, terdapat dua jalur kelulusan yang dapat dilalui, yaitu jalur skripsi dan jalur non-skripsi. Jalur skripsi merupakan proses kelulusan mahasiswa, dimana seorang mahasiswa diwajibkan menyusun karya tulis ilmiah berdasarkan penelitian lapangan dan/atau kepustakaan. Jalur non-skripsi merupakan proses kelulusan mahasiswa, dimana seorang mahasiswa tidak menyusun sebuah karya tulis ilmiah melainkan melalui pengujian secara lisan 3 mata kuliah yang berkaitan dengan bidang studi. Permasalahan yang terjadi adalah ketidaktepatan keputusan yang diambil pihak sekretariat dalam menentukan jalur kelulusan mahasiswa, seperti terjadinya seorang mahasiswa yang harusnya menempuh jalur skripsi tapi dinyatakan jalur non skripsi.

Dalam penelitian ini algoritma ID3 akan dibuat sebuah model yang akan menghasilkan sebuah pohon keputusan. Dimana pohon keputusan tersebut akan digunakan untuk menentukan jalur skripsi pada mahasiswa atau jalur non skripsi dalam jalur kelulusan mahasiswa tingkat akhir pada suatu Perguruan Tinggi Swasta.

2. TINJAUAN PUSTAKA

Beberapa penelitian terkait *decision tree learning* diantaranya; penelitian yang dilakukan oleh Rong Cao dan Lizhen Xu menggunakan Algoritma C4.5 untuk menganalisa penjualan. Lee (2010) menggunakan algoritma ID3 untuk penentuan penerima beasiswa. Wibowo (2009) menggunakan algoritma ID3 untuk membantu dalam pengambilan keputusan pada penentuan MVP di sebuah pertandingan bola basket.

Salah satu jenis dari *decision tree learning* adalah algoritma ID3. Algoritma ID3 (*Iterative Dichotomiser 3*) merupakan algoritma *decision tree learning* (algoritma pembelajaran pohon keputusan) yang menggunakan strategi pencarian *hill-climbing*, yaitu dimulai dari pohon kosong, kemudian secara progresif berusaha menemukan sebuah pohon keputusan yang mengklasifikasikan sampel-sampel data secara akurat tanpa kesalahan. Pertumbuhan cabang-cabang pohon keputusan pada algoritma ID3 dilakukan sampai pohon tersebut mampu mengklasifikasikan

sampel data secara akurat dengan tingkat kebenaran 100 % sesuai dengan data latih (Suyanto, 2011). Adapun sample data yang digunakan oleh ID3 memiliki beberapa syarat menurut Setiawan (Setiawan, 2010), yaitu:

- 1) Deskripsi atribut-nilai. Atribut yang sama harus mendeskripsikan tiap contoh dan memiliki jumlah nilai yang sudah ditentukan.
- 2) Kelas yang sudah didefinisikan sebelumnya. Suatu atribut contoh harus sudah didefinisikan, karena mereka tidak dipelajari oleh ID3.
- 3) Kelas-kelas yang diskrit. Kelas harus digambarkan dengan jelas. Kelas yang kontinu dipecah-pecah menjadi kategori-kategori yang relatif, misalnya saja metal dikategorikan menjadi "hard, quite hard, flexible, soft, quite soft".
- 4) Jumlah contoh (example) yang cukup. Karena pembangkitan induktif digunakan, maka dibutuhkan test case yang cukup untuk membedakan pola yang valid dari peluang suatu kejadian.

Algoritma pada metode ini menggunakan konsep dari entropy informasi. Pemilihan atribut dengan menggunakan Information Gain. Pemilihan atribut pada ID3 dilakukan dengan properti statistik, yang disebut dengan information gain. Gain mengukur seberapa baik suatu atribut memisahkan training example ke dalam kelas target. Atribut dengan informasi tertinggi akan dipilih. Dengan tujuan untuk mendefinisikan gain, pertamanya digunakanlah ide dari teori informasi yang disebut entropy. Entropy mengukur jumlah dari informasi yang ada pada atribut dengan rumus :

$$Entropy(S) = -P_a \log_2 P_a - P_b \log_2 P_b$$

S adalah ruang (data) sample yang digunakan untuk training.

P_a adalah jumlah yang bersolusi positif (mendukung) pada data sample untuk kriteria tertentu.

P_b adalah jumlah yang bersolusi negative (tidak mendukung) pada data sample untuk kriteria tertentu.

Dari rumus entropy diatas dapat disimpulkan bahwa definisi entropy (S) adalah jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau-) dari sejumlah data acak pada suatu ruang sampel S. Entropy bisa dikatakan sebagai kebutuhan bit untuk meyakinkan suatu kelas. Semakin kecil nilai entropy maka semakin baik digunakan dalam mengekstraksi suatu kelas.

Pada algoritma ID3 pengurangan entropy disebut dengan informasi gain. Pembagian sample S terhadap atribut A dapat dihitung information gain dengan rumus:

$$Gain(S, A) = Entropy(S) - \sum_{v \in \text{nilai}(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Dimana :

A : atribut

V : suatu nilai yang mungkin untuk atribut A

Value (A) : himpunan yang mungkin untuk atribut A

$|S_v|$: Jumlah sampel untuk nilai v

$|S|$: jumlah seluruh sampel data

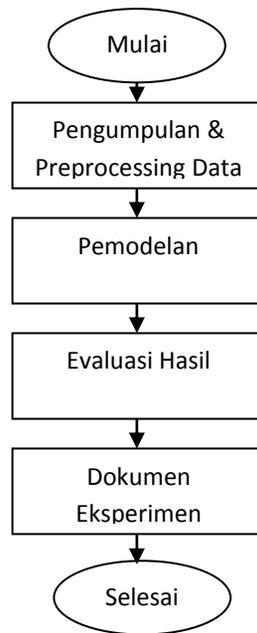
Entropy (S_v) : entropy untuk sampel-sampel yang memiliki nilai v

Secara ringkas, langkah kerja Algoritma ID3 dapat digambarkan sebagai berikut:

- 1) Penghitungan Information Gain dari setiap atribut
- 2) Pemilihan atribut yang memiliki nilai information gain terbesar,
- 3) Pembentukan simpul yang berisi atribut tersebut,
- 4) Ulangi proses perhitungan information gain akan terus dilaksanakan sampai semua data telah termasuk dalam kelas yang sama. Atribut yang telah dipilih tidak diikuti lagi dalam perhitungan nilai information gain.

3. METODE PENELITIAN

Metode penelitian dalam paper ini menggunakan metode eksperien dengan tahapan; pengumpulan data, pemodelan, evaluasi hasil dan dokumen eksperimen. Alur tahapan penelitian ini di tunjukkan pada gambar 1 di bawah ini.



Gambar 1 Alur Penelitian

Tahap pertama yang dilakukan adalah tahap pengumpulan dan preprocessing data, dimana pada tahap ini data mahasiswa dikumpulkan untuk selanjutnya dilakukan preprocessing yaitu penyaringan data mahasiswa dan mengklasifikasikan menjadi beberapa atribut. Tahapan kedua yang dilakukan adalah pemodelan yaitu penerapan algoritma ID3 dari data yang sudah tersedia. Tahapan ketiga adalah evaluasi hasil, yaitu tahapan dimana data yang telah diterapkan pada algoritma ID3 hasilnya di evaluasi tingkat akurasi dan yang terakhir adalah tahap Dokumen eksperimen.

4. PEMBAHASAN

4.1 Pengumpulan Dan Preprocessing Data

Sumber data utama yang digunakan dalam penelitian ini adalah dataset mahasiswa akuntansi angkatan 2012 pada sebuah PTS dengan format .xlsx. Dataset mahasiswa terdiri dari atribut NPM, Nama, IPK lokal, IPK utama, IPK total, SKS, MK utama, Tgl Lulus PI, Alamat, Telepon, HP, Tgl lahir, proposal dan kelas. Jumlah data pada atribut tersebut berjumlah 632 record, 282 data latih dan 350 data uji. Atribut yang akan digunakan pada penelitian ini adalah IPK total, jumlah SKS, MK Utama, PI, dan Proposal.

Beberapa atribut dilakukan proses klasifikasi data, diantaranya IPK total dikelompokkan menjadi 3 kategori (rendah, cukup, tinggi), SKS dikelompokkan menjadi 2 kategori (cukup, kurang), MK Utama dikelompokkan menjadi 2 kategori (memenuhi, belum memenuhi), PI dikelompokkan menjadi 2 kategori (lulus, belum lulus), Proposal dikelompokkan menjadi 2 kategori (diterima, ditolak). Klasifikasi atribut mahasiswa dapat dilihat pada tabel 1 di bawah ini.

Tabel 1 Range klaifikasi data mahasiswa

Atribut	Value	Range
IPK Total	Rendah	0 - 2,99
	Cukup	3,00 - 3,24
	Tinggi	≥ 3,25
SKS	Cukup	≥ 112
	Kurang	0 - 111
MK Utama	Memenuhi	≥ 11
	Belum memenuhi	0 - 10
PI	Lulus	-

	Belum Lulus	-
Proposal	Diterima	-
	Ditolak	-

Pada tahap preprocessing, data mahasiswa yang dipilih dipastikan layak untuk dilakukan proses pengolahan. Dari data mahasiswa dilakukan proses transformasi data dan reduksi data. Dalam proses transformasi, data ditransformasikan ke dalam bentuk yang sesuai untuk proses data mining. Selanjutnya data direduksi yaitu dengan melakukan penghilangan atribut yang tidak diperlukan sehingga ukuran dari database menjadi kecil dan hanya menyertakan atribut yang diperlukan dalam proses data mining. Dari dua proses yang telah dilakukan didapatkan hasil seperti tabel 2 berikut ini.

Tabel 2 Tabel data hasil preprocessing

IPK UTAMA	IPK TOTAL	SKS	MK UTAMA	PI	PROPOSAL	JALUR
Rendah	Cukup	Kurang	Belum memenuhi	Lulus	Ditolak	Non skripsi
Rendah	Tinggi	Kurang	Belum memenuhi	Lulus	Ditolak	Non skripsi
Rendah	Tinggi	Kurang	Belum memenuhi	Lulus	Ditolak	Non skripsi
Tinggi	Tinggi	Cukup	Memenuhi	Lulus	Diterima	Skripsi
Rendah	Cukup	Cukup	Memenuhi	Lulus	Ditolak	Non skripsi
Tinggi	Cukup	Cukup	Memenuhi	Lulus	Ditolak	Non skripsi
Cukup	Tinggi	Cukup	Memenuhi	Lulus	Diterima	Skripsi
Tinggi	Tinggi	Cukup	Memenuhi	Lulus	Diterima	Skripsi
.....

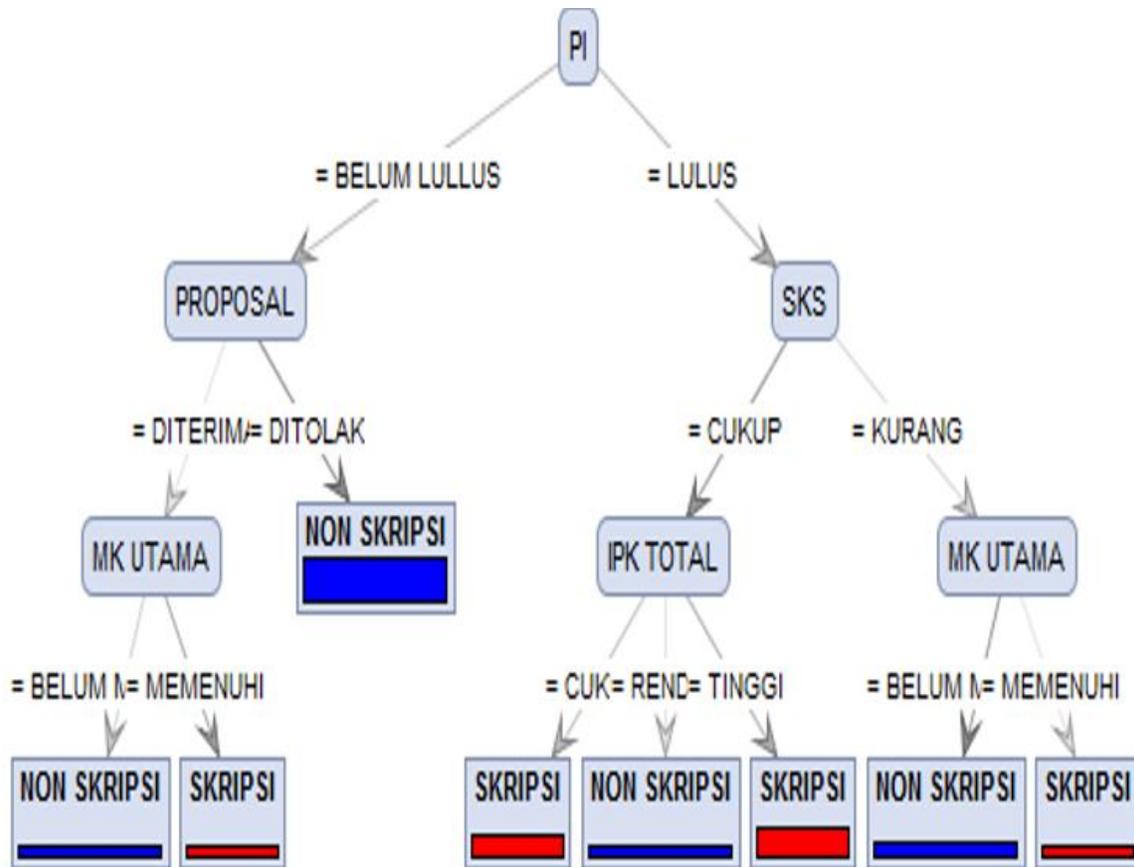
4.2 Pemodelan

Pemodelan merupakan tahap yang secara langsung melibatkan teknik data mining, yaitu dengan melakukan pemilihan teknik data mining dan menentukan algoritma yang akan digunakan. Algoritma yang digunakan dalam penelitian ini adalah algoritma.

ID3 dengan bantuan tool yang digunakan adalah RapidMiner Studio Basic versi 6.5.

Berikut merupakan pengolahan data dengan menggunakan algoritma ID3 pada RapidMiner :

Dengan menggunakan pemodelan ID3 seperti gambar 1 di atas, maka di dapatkan hasil dengan pohon keputusan yang terbentuk adalah sebagai berikut.



Gambar 2 Model pohon keputusan yang terbentuk

Aturan-aturan yang diperoleh dari pohon keputusan tersebut untuk menentukan jalur kelulusan berdasarkan atribut-atribut pada penelitian ini, yaitu:

IF PI = belum lulus and proposal = diterima and mk_utama = memenuhi then jalur kelulusan = skripsi

IF PI = lulus and SKS = kurang and mk_utama = memenuhi then jalur kelulusan = skripsi

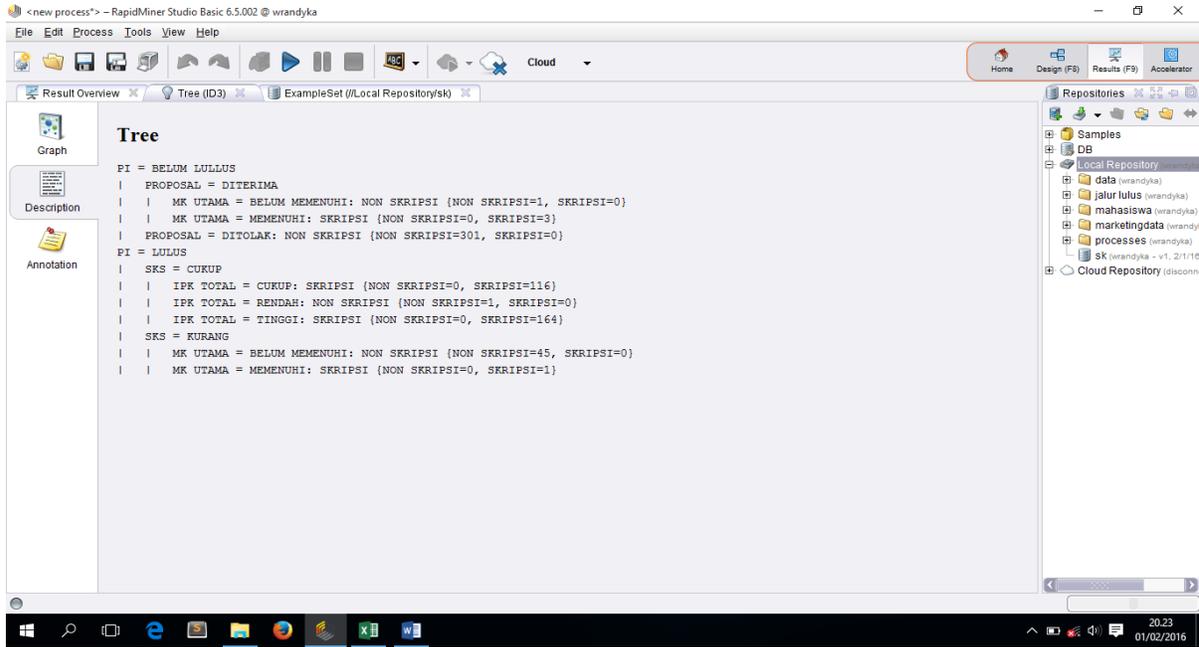
IF PI = lulus and SKS = cukup and IPK total = cukup or tinggi then jalur kelulusan = skripsi

IF PI = belum lulus and proposal = diterima and mk_utama = belum memenuhi then jalur kelulusan = non skripsi

IF PI = belum lulus and proposal = ditolak then jalur kelulusan = non skripsi

IF PI = lulus and SKS = cukup and IPK total = rendah then jalur kelulusan = non skripsi

IF PI = lulus and SKS = kurang and MK utama = belum memenuhi then jalur kelulusan = non skripsi



Gambar 3 Rule atau aturan yang diperoleh

4.3 Evaluasi Hasil

Komparasi nilai accuracy, precision dan recall yang diperoleh dari pengolahan data adalah sebagai berikut.

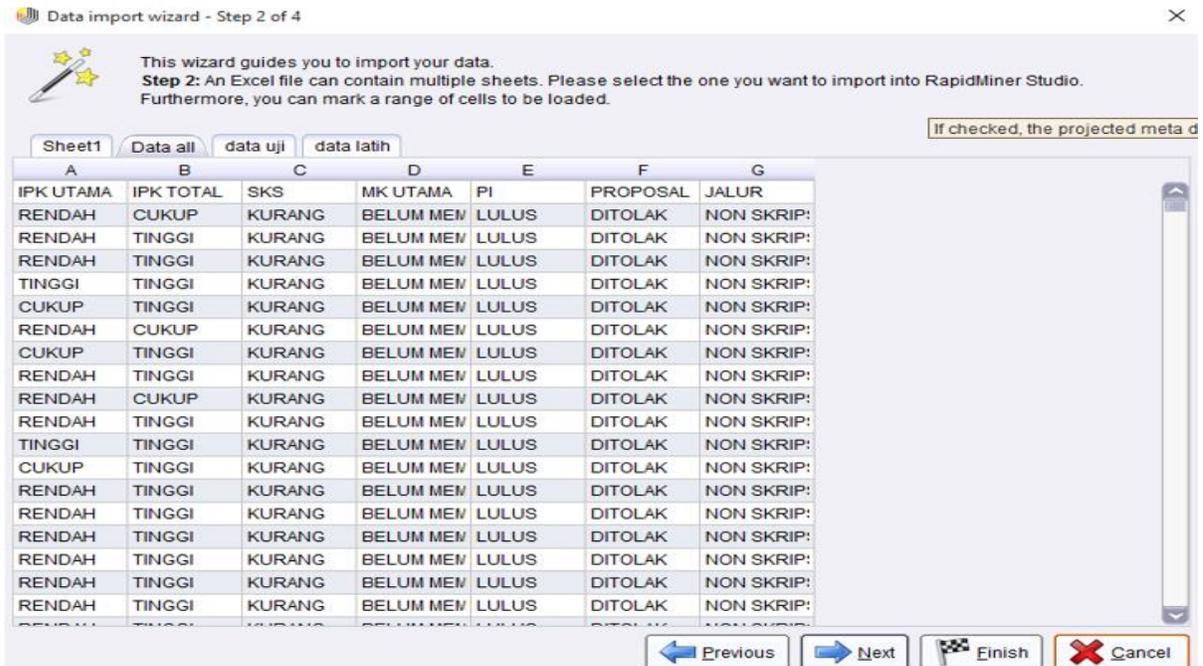
Tabel 3 Nilai accuracy, precision dan recall

	Information Gain	Gain Ratio	Gini Index
Accuracy	90,00%	100%	85,00%
Precision	85,71%	100%	79,76%
Recall	93,33%	100%	83,33%

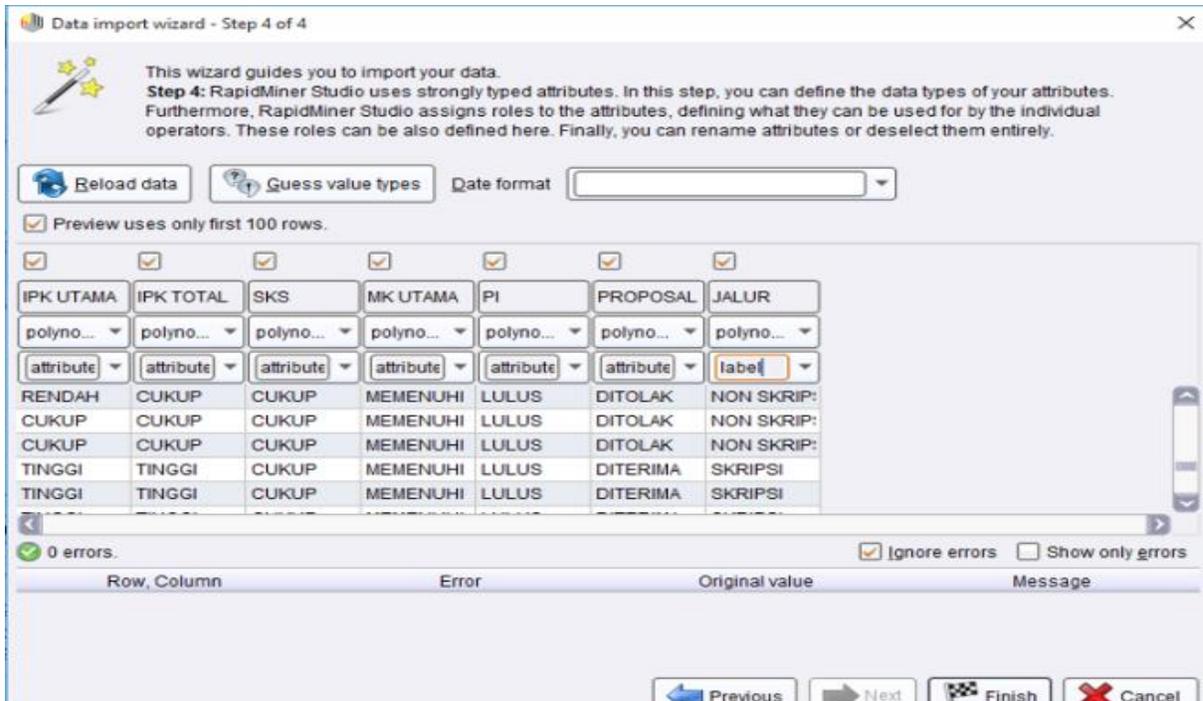
Model yang dihasilkan dengan criteria information gain. Gain ratio dan gini index diuji menggunakan metode confusion matrix, terlihat perbandingan nilai accuracy, precision dan recall pada table 3, untuk criteria gain ratio memiliki nilai accuracy, precision dan recall yang paling tinggi, diikuti dengan criteria information gain dan gini index.

4.4 Dokumentasi Eksperimen

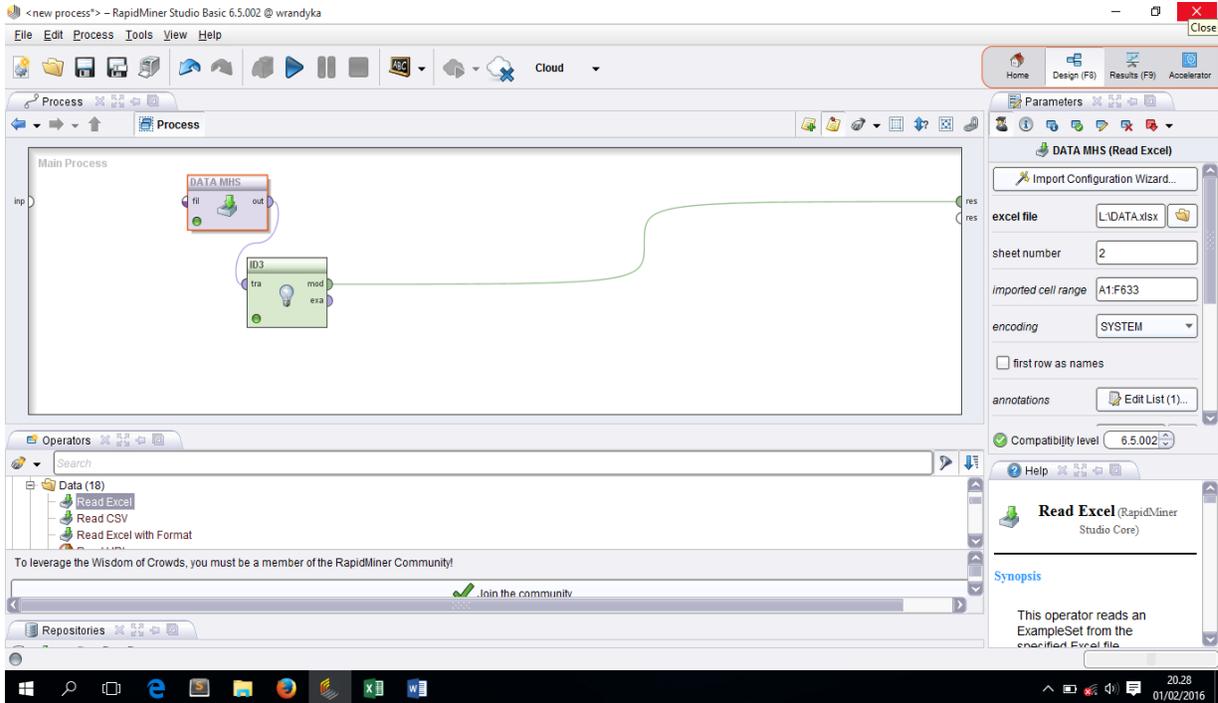
- a) Mengambil data dari file .xls yang merupakan dataset analisis mahasiswa tingkat akhir, adapun prosesnya dapat dilihat pada gambar 4. Pada gambar 4 di bawah menunjukkan data analisis mahasiswa tingkat akhir yang telah di *import* ke dalam perangkat lunak RapidMiner
- b) Menentukan id
 Pada Gambar 5 menunjukkan proses pengambilan id sebagai kata kunci utama (*primary key*) pada data analisis mahasiswa tingkat akhir.
- c) Penerapan Algoritma ID3
 Proses penerapan algoritma dilakukan setelah data diambil dari dataset dan ditentukan id. Gambar 6 menunjukkan penerapan algoritma ID3 dalam menentukan jalur kelulusan mahasiswa.
 Gambar 6 merupakan gambar yang menunjukkan penerapan algoritma ID3 dalam menentukan jalur kelulusan.
- d) Evaluasi hasil penerapan algoritma
 Gambar 7 menggambarkan proses mencari nilai akurasi dari pohon keputusan yang telah didapatkan.



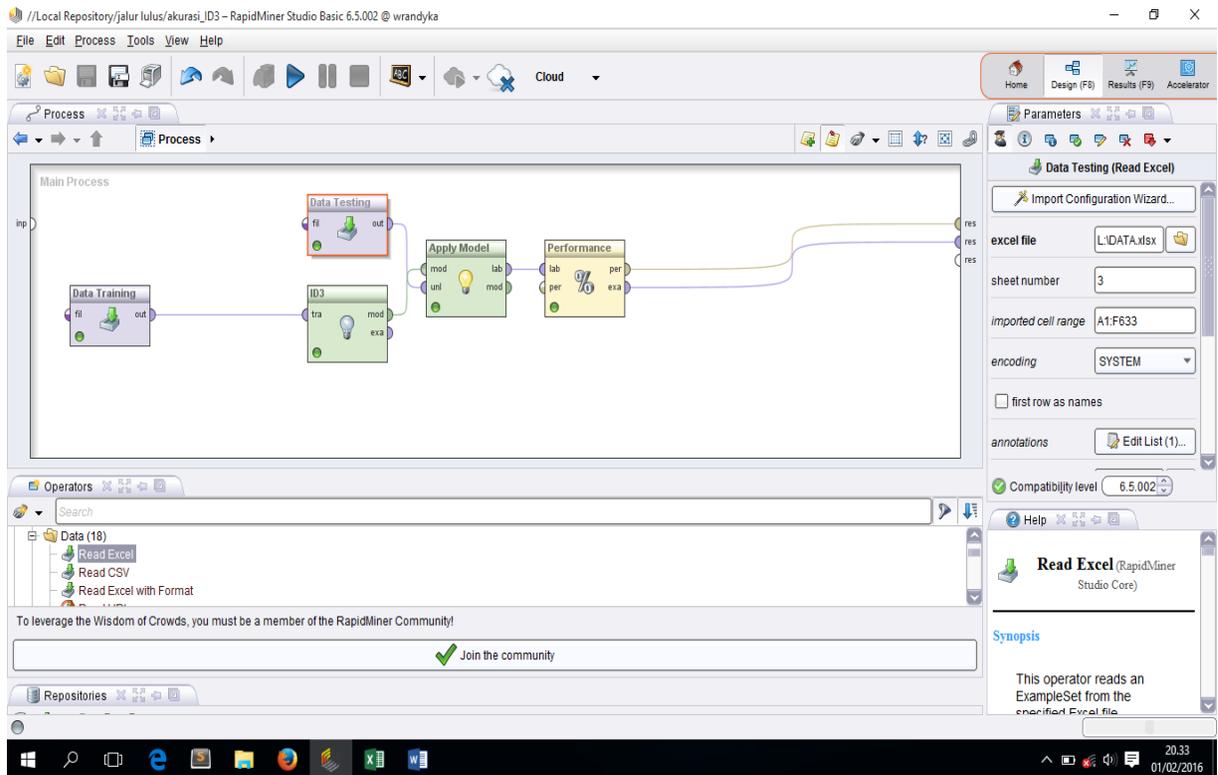
Gambar 4 Import data dari dataset



Gambar 5 Menentukan id



Gambar 6 Penerapan algoritma ID3



Gambar 7 Menentukan nilai akurasi

5. KESIMPULAN

Dalam penelitian ini dilakukan pembuatan model menggunakan algoritma ID3 menggunakan data mahasiswa tingkat akhir pada suatu PTS. Dari hasil pohon keputusan diperoleh 7 rule atau aturan yaitu 3 aturan untuk jalur skripsi dan 4 aturan untuk jalur non skripsi.

Model yang dihasilkan diuji keakuratannya dengan cara mengambil data untuk diuji/validasi dari data cleaning yang telah didapatkan dengan perangkat lunak RapidMiner dan sisanya sebagai data training. Model yang dihasilkan dapat dikategorikan klasifikasi yang sangat baik karena pada pengukuran kinerja kriteria information gain, gain ratio dan gini index memiliki nilai accuracy antara 0,85-1,00, sehingga dapat dikatakan pohon keputusan dengan algoritma ID3 dapat diterapkan dalam penentuan jalur kelulusan mahasiswa tingkat akhir.

DAFTAR PUSTAKA

- [1] Lee, Michael. 2010. *Perancangan Klasifikasi Penerimaan Beasiswa Menggunakan Algoritma ID3 (Iterative Dichtomizer Three) (Studi Kasus : Beasiswa Rutin UKSW Salatiga)*. Salatiga: FTI UKSW.
- [2] Rong Cao and Lizhen Xu. 2009. *Improved C4.5 Algorithm for the Analysis of Sales*. In 2009 Sixth Web Information Systems and Applications Conference.
- [3] Setiawan, Bambang. 2010. *Perancangan Sistem Pendukung Keputusan (Spk) Untuk Menentukan Kelaiklautan Kapal : Studi Kasus Di Kantor Administrasi Pelabuhan Klas Utama Tj. Perak Surabaya*. Master Thesis, Institut Teknologi Sepuluh Nopember: Surabaya.
- [4] Suyanto. 2011. *Artificial Intelegent (Cetakan kedua)*. Informatika: Bandung.
- [5] Wibowo, Bagus Ari. 2011. *Perancangan dan Implementasi Sistem Pendukung Keputusan untuk Jalan Menggunakan Metode ID3 (Studi Kasus BAPPEDA Kota Salatiga)*. Universitas Kristen Satya Wacana: Jawa Tengah.
- [6] URL <http://akademik.gunadarma.ac.id> diakses tanggal 10 Desember 2015.