

PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI JENIS PEKERJAAN ALUMNI DI UNIVERSITAS MUHAMMADIYAH YOGYAKARTA

Badrahini Masajeng Respati, Slamet Riyadi, Asroni
Teknik Informatika Universitas Muhammadiyah Yogyakarta

ABSTRACT

Perkembangan dunia pendidikan di Indonesia mengalami peningkatan yang sangat pesat. Salah satu hal yang menjadi tolak ukur keberhasilan mutu pendidikan suatu Universitas adalah jenis pekerjaan alumni yang didapatkan setelah lulus, tetapi *database* alumni yang ada masih berupa kumpulan data dengan atribut yang kompleks. Penelitian ini bertujuan untuk mengetahui faktor-faktor yang berpengaruh terhadap jenis pekerjaan alumni berdasarkan hasil pohon keputusan menggunakan metode klasifikasi dengan algoritma C4.5. Metodologi penelitian ini dimulai dengan melakukan studi pustaka, menentukan metode *data mining*, pengumpulan data, seleksi data, pengolahan data, pengujian data, dan membuat kesimpulan. Penelitian ini menggunakan beberapa atribut antara lain fakultas, tahun lulus, ipk, dan angkatan sebagai parameter untuk melakukan klasifikasi. Data alumni yang digunakan sebanyak 259, terdiri dari 3 fakultas yaitu Ekonomi, Kedokteran, dan Teknik angkatan 2001-2013 serta tahun lulus dari 2011-2016. Hasil penelitian yang telah dilakukan adalah Jika berasal dari fakultas Ekonomi, tahun lulus 2011 dan 2012 mayoritas bekerja di Swasta, Jika berasal dari fakultas Kedokteran dengan tahun lulus 2011 dan 2012 dengan ipk antara 3 sampai 3.5 mayoritas bekerja di Swasta, Jika berasal dari fakultas Teknik mayoritas bekerja di Swasta sehingga diperoleh atribut yang paling berpengaruh terhadap jenis pekerjaan alumni yaitu Fakultas karena sebagai akar dari pohon keputusan dan memiliki *gain ratio* paling tinggi di antara atribut lainnya dan algoritma C4.5 cocok diterapkan pada klasifikasi jenis pekerjaan alumni.

Kata Kunci : algoritma C4.5, *gain ratio*, pohon keputusan, *data mining*.

1. Pendahuluan

1.1 Latar Belakang

Perkembangan dunia pendidikan di Indonesia mengalami peningkatan yang sangat pesat. Salah satu hal yang menjadi tolak ukur keberhasilan mutu pendidikan suatu Universitas adalah jenis pekerjaan alumni yang didapatkan setelah lulus. Tentunya hal tersebut juga terjadi di Universitas Muhammadiyah Yogyakarta.

Banyaknya jumlah alumni di Universitas Muhammadiyah Yogyakarta memiliki peran penting terhadap peningkatan mutu pendidikan. Hal yang dapat dijadikan sebagai tolak ukurnya adalah jenis pekerjaan yang diperoleh alumni setelah lulus, tetapi masalahnya belum dilakukan pengolahan pada data alumni dengan menggunakan metode data mining.

Berdasarkan uraian diatas, klasifikasi jenis pekerjaan alumni belum pernah dilakukan sebelumnya, oleh karena itu penulis akan menggunakan algoritma C4.5 dalam melakukan klasifikasi jenis pekerjaan alumni dengan hasil berupa pohon keputusan yang dapat digunakan untuk mengidentifikasi dan melihat hubungan antara faktor-faktor yang mempengaruhi suatu kasus.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah mengimplementasikan metode klasifikasi menggunakan algoritma C4.5 dalam melakukan klasifikasi data jenis pekerjaan alumni fakultas Ekonomi, fakultas Teknik, dan fakultas Kedokteran di Universitas Muhammadiyah Yogyakarta. Dan mengetahui faktor-faktor yang berpengaruh pada jenis pekerjaan alumni berdasarkan hasil pohon keputusan dari algoritma C4.5 pada klasifikasi data jenis pekerjaan alumni UMY.

2. Tinjauan Pustaka dan Landasan Teori

2.1 Tinjauan Pustaka

Menurut penelitian yang dilakukan oleh (Saefulloh & Moedjiono, 2013) yang berjudul "Penerapan Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Tepat Waktu". Metode yang digunakan dalam penelitian ini adalah algoritma C4.5, Naive Bayes, dan Neural Network untuk memperkirakan kelulusan tepat waktu mahasiswa dengan melihat pengaruh dari IMK dan IPK. Dari penelitian ini menunjukkan bahwa algoritma terbaik adalah algoritma yang paling tinggi tingkat accuracy pada model klasifikasi yaitu C4.5 dan Neural Network dengan tingkat accuracy 100% sedangkan Naive bayes 99.8878 %. Hasil data mining dari algoritma terpilih dalam penelitian ini menggunakan java engine yang dapat menampilkan prediksi kelulusan tepat waktu beserta jumlah kelulusan tepat waktu setiap Program Studi.

Menurut penelitian yang dilakukan oleh (Nugroho, 2014) yang berjudul "Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Kelulusan Mahasiswa Fakultas Komunikasi Dan Informatika Universitas Muhammadiyah Surakarta". Metode yang digunakan dalam penelitian ini adalah algoritma C4.5 untuk klasifikasi mahasiswa berdasarkan predikat kelulusannya. Berdasarkan hasil penelitian yang dilakukan dapat disimpulkan bahwa telah diperoleh klasifikasi predikat kelulusan mahasiswa Fakultas Komunikasi dan Informatika UMS. Variabel yang paling tinggi pengaruhnya terhadap predikat kelulusan adalah partisipasi mahasiswa menjadi asisten. Interpretasi hasil penelitian mengindikasikan bahwa variabel yang perlu digunakan sebagai pertimbangan bagi Fakultas Komunikasi dan Informatika UMS untuk memperoleh tingkat predikat

kelulusan yang maksimal adalah peran serta mahasiswa untuk menjadi asisten. Secara umum probabilitas predikat “Cumlaude” pada kelompok mahasiswa yang pernah menjadi asisten lebih tinggi dibandingkan dengan yang tidak pernah menjadi asisten jika berasal dari jurusan IPA semasa sekolah menengah atas memiliki probabilitas predikat kelulusan “Cumlaude” yang lebih tinggi dibandingkan dengan mahasiswa dari jurusan lainnya.

2.2 Landasan Teori

Data mining merupakan proses penggalian dan pertambangan pengetahuan dari sejumlah data yang besar, database atau repository database lainnya. Tujuan utama dari penambangan data ini untuk menemukan pengetahuan baru yang tersembunyi dari database tersebut (Elisa, 2017).

Klasifikasi adalah suatu teknik yang dilakukan untuk memprediksi class atau properti dari setiap instace data. Teknik ini dilakukan pada data baru dengan memanipulasi data yang ada dan telah diklasifikasikan menggunakan hasilnya untuk memberikan sejumlah aturan.

A. Konsep Entropy

Entropy (S) merupakan jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel S. Entropy dapat dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. Semakin kecil nilai entropy maka akan semakin baik entropy yang digunakan dalam mengekstrak suatu kelas. Entropy digunakan untuk mengukur ketidaksihan S.

$$Entropy (S) = \sum_{i=0}^n -p_i * \log_2(p_i) \dots\dots\dots(1)$$

Berikut keterangannya :

S : Himpunan kasus

n : Jumlah partisi

pi : Jumlah kasus pada partisi ke-i

Si : Jumlah kasus pada partisi ke- i

B. Information Gain

Information Gain adalah informasi yang didapatkan dari perubahan entropy di suatu kumpulan data, baik melalui observasi maupun disimpulkan dengan cara melakukan partisipasi terhadap suatu set data.

$$Gain (S, A) = Entropy (S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy (S_i) \dots\dots\dots(2)$$

Rumus (2) merupakan rumus yang digunakan dalam perhitungan information gain setelah melakukan entropy. Berikut keterangannya :

S : Himpunan kasus

n : Jumlah partisi atribut A

|Si| : Jumlah kasus pada partisi ke-i

|S| : Jumlah kasus dalam S

C. SplitInfo

SplitInfo merupakan rumus yang menyatakan informasi potensial atau entropy. dapat dilihat dalam rumus (3). Dan keterangannya :

$$SplitInfo(S, A) = -\sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \dots\dots\dots(3)$$

S : Himpunan kasus

A : Atribut

Si : Jumlah kasus pada partisi ke- i

D. Gain Ratio

Gain Ratio adalah modifikasi dari information gain yang digunakan untuk mengurangi bias atribut yang memiliki banyak cabang. Gain ratio memiliki sifat :

- 6. Bernilai besar jika data menyebar rata
- 7. Bernilai kecil jika semua data masuk ke dalam satu cabang

$$GainRatio (S, A) = \frac{Gain(S,A)}{SplitInfo(S,A)} \dots\dots\dots(4) \text{ dan}$$

Keterangannya sebagai berikut:

S : Himpunan kasus

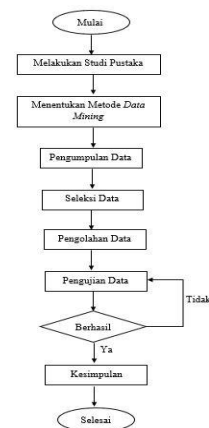
A : Atribut

Gain (S, A) : Information gain pada atribut A

SpiltInfo (S, A) : SplitInfo pada atribut A

3. Metodologi Penelitian

3.1 Prosedur Penelitian



Gambar 2 Diagram Alir Penelitian

1. Melakukan Studi Pustaka

Pertama hal yang dilakukan adalah melakukan studi pustaka yang berkaitan dengan penelitian-penelitian sebelumnya tentang penggunaan algoritma C4.5, serta untuk menguatkan penelitian ini berdasarkan teori yang digunakan. Dalam kasus ini peneliti ingin melakukan analisis pada klasifikasi jenis pekerjaan alumni di Universitas Muhammadiyah Yogyakarta dengan menggunakan algoritma C4.5 sekaligus untuk mengetahui pola dan cara kerja klasifikasi dalam *data mining* tersebut.

2. Menentukan Metode Data Mining

Setelah melakukan studi pustaka tahap selanjutnya adalah menentukan metode yang sesuai untuk digunakan dalam teknik klasifikasi, setelah melakukan pengamatan dan observasi peneliti memilih menggunakan algoritma C4.5. Karena algoritma C4.5 dapat lakukan transformasi atau mengubah nilai atribut data ke dalam bentuk data yang sesuai agar data dapat diproses menggunakan algoritma C4.5. Sehingga akan diperoleh dataset utuh yang digunakan untuk proses ke tahapan selanjutnya

3. Melakukan Tahapan Metode Klasifikasi

Berikut tahapan metode klasifikasi menggunakan algoritma C4.5 yang akan dilakukan peneliti untuk memperoleh hasil penelitian antara lain :

A. Melakukan Pengumpulan Data

Pada tahap ini peneliti melakukan pengumpulan data dan ini merupakan tahapan yang penting karena dapat berpengaruh terhadap hasil penelitian, sehingga dalam mengumpulkan data harus dilakukan dengan benar. Peneliti mendapatkan data ini dari *database server* BSI (Biro Sistem Informasi) Universitas Muhammadiyah Yogyakarta.

B. Melakukan Seleksi Data

Tahap ini dilakukan seleksi terhadap data *database excel*. Karena data yang diperoleh tidak semuanya digunakan, dipilih sesuai dengan atribut atau variabel yang dibutuhkan dalam penelitian dengan melakukan seleksi data sehingga menjadi *dataset*. Sebagai contoh pada penelitian ini atribut yang dipilih adalah fakultas, tahun lulus, ipk, angkatan, dan jenis pekerjaan. Pada tahap ini akan menghilangkan data yang *null*, data tidak valid, dan data yang ganda. karena data yang kosong ataupun tidak valid akan berpengaruh terhadap hasil yang diperoleh. Kemudian data akan diolah menggunakan *Microsoft Office Excel 2016*.

C. Melakukan Pengolahan Data

Setelah semua data yang diperlukan telah dipilih, maka tahap penelitian selanjutnya adalah pengolahan data. Pada tahap ini akan dilakukan transformasi atau mengubah nilai atribut data ke dalam bentuk data yang sesuai agar data dapat diproses menggunakan algoritma C4.5. Sehingga akan diperoleh dataset utuh yang digunakan untuk proses ke tahapan selanjutnya.

D. Melakukan Pengujian Data

Pada tahap pengujian hasil akan dilakukan pengujian data baik secara manual dengan algoritma C4.5 dan menggunakan *software RapidMiner*.

4. Membuat Kesimpulan Penelitian

Berdasarkan hasil pengujian maka dapat ditarik kesimpulan yang mengacu pada rumusan masalah dan tujuan penelitian. Dan saran yang digunakan untuk mengembangkan penelitian selanjutnya serta dimasukkan untuk meningkatkan kualitas penelitian.

4. Hasil dan Pembahasan

4.1. Pengumpulan Data

Pengumpulan data dilakukan dengan cara meminta data ke Biro Sistem Informasi Universitas Muhammadiyah Yogyakarta. Data yang diperoleh adalah data alumni Universitas Muhammadiyah Yogyakarta dengan jumlah data awal 688 data alumni. Di dalam data alumni terdapat beberapa atribut antara lain : id, nim, nama, tempat lahir, jenis kelamin, id prodi, bulan lulus, tahun lulus, ipk, judul skripsi, alamat, email, nomor hp, nomor telepon, id jenis pekerjaan, nama institusi, alamat institusi, bagian, masa tunggu, id salary, tanggal lahir, tanggal mulai kerja, relevansi pendidikan, dan jabatan.

4.2. Penyeleksian Data

Pemilihan atribut atau variabel untuk penelitian ini dilakukan dengan pertimbangan atribut yang dipilih sudah sesuai dengan atribut yang dibutuhkan untuk melakukan proses klasifikasi jenis pekerjaan alumni. Dari data yang tersedia diambil beberapa atribut antara lain : id prodi, tahun lulus, ipk, angkatan, dan id jenis pekerjaan. Dari jumlah seluruh data yang ada, data yang digunakan jumlahnya 259 data alumni yang diambil dari 3 fakultas yaitu fakultas Ekonomi, fakultas Teknik, dan fakultas Kedokteran dari angkatan 2001-2013 serta tahun lulus dari 2011-2016.

4.3 Pengolahan Data

Pada tahap ini dilakukan perubahan data atau transformasi data dengan mengubah nilai atribut yang awalnya bernilai nominal menjadi nilai-nilai atribut yang sesuai dengan data pada tabel dibawah ini. Dan data tidak ada *null*. Dan pada nilai atribut telah dirubah untuk memudahkan proses selanjutnya, ini dilakukan dalam proses transformasi data. Seperti gambar 2.

4.4 Pengujian Data

Pada tahap pengujian data dilakukan dua pengujian yaitu menggunakan *software RapidMiner* dan menggunakan Algoritma C4.5. Dibawah ini penjelasan dari hasil pengujian tersebut. Berikut adalah data yang akan dilakukan pengujian.

	A	B	C	D	E
1	FAKULTAS	TAHUN LULUS	IPK	ANGKATAN	JENIS PEKERJAAN
2	Teknik	Lulusan 2011 dan 2012	IPK kurang dari 3	Angkatan 2001-2006	Swasta
3	Teknik	Lulusan 2011 dan 2012	IPK antara 3 sampai 3,5	Angkatan 2007-2013	Swasta
4	Kedokteran	Lulusan 2011 dan 2012	IPK antara 3 sampai 3,5	Angkatan 2007-2013	Swasta
5	Kedokteran	Lulusan 2011 dan 2012	IPK antara 3 sampai 3,5	Angkatan 2007-2013	Pemerintah
6	Kedokteran	Lulusan 2011 dan 2012	IPK antara 3 sampai 3,5	Angkatan 2007-2013	Swasta
7	Kedokteran	Lulusan 2011 dan 2012	IPK antara 3 sampai 3,5	Angkatan 2007-2013	Pemerintah
8	Kedokteran	Lulusan 2011 dan 2012	IPK kurang dari 3	Angkatan 2007-2013	Pemerintah
9	Kedokteran	Lulusan 2011 dan 2012	IPK antara 3 sampai 3,5	Angkatan 2007-2013	Pemerintah
10	Kedokteran	Lulusan 2011 dan 2012	IPK kurang dari 3	Angkatan 2007-2013	Swasta
11	Kedokteran	Lulusan 2011 dan 2012	IPK kurang dari 3	Angkatan 2007-2013	Swasta
12	Ekonomi	Lulusan 2011 dan 2012	IPK antara 3 sampai 3,5	Angkatan 2007-2013	Swasta
13	Teknik	Lulusan 2011 dan 2012	IPK antara 3 sampai 3,5	Angkatan 2001-2006	Swasta
14	Teknik	Lulusan 2011 dan 2012	IPK antara 3 sampai 3,5	Angkatan 2001-2006	Swasta
15	Teknik	Lulusan 2011 dan 2012	IPK antara 3 sampai 3,5	Angkatan 2001-2006	Swasta
16	Teknik	Lulusan 2011 dan 2012	IPK antara 3 sampai 3,5	Angkatan 2007-2013	Pemerintah
17	Teknik	Lulusan 2011 dan 2012	IPK lebih dari 3,5	Angkatan 2007-2013	Swasta
18	Teknik	Lulusan 2011 dan 2012	IPK lebih dari 3,5	Angkatan 2007-2013	Swasta
19	Ekonomi	Lulusan 2011 dan 2012	IPK lebih dari 3,5	Angkatan 2007-2013	Swasta
20	Ekonomi	Lulusan 2011 dan 2012	IPK antara 3 sampai 3,5	Angkatan 2007-2013	Swasta
21	Ekonomi	Lulusan 2011 dan 2012	IPK lebih dari 3,5	Angkatan 2007-2013	Swasta
22	Ekonomi	Lulusan 2011 dan 2012	IPK lebih dari 3,5	Angkatan 2007-2013	Swasta
23	Ekonomi	Lulusan 2011 dan 2012	IPK lebih dari 3,5	Angkatan 2007-2013	Pemerintah

Gambar 1 Data yang akan diuji

4.1.1. Pengujian Data Menggunakan Algoritma C4.5

1. Melakukan Pra Proses

Pada gambar 2 merupakan *dataset* yang digunakan untuk pra proses dari klasifikasi jenis pekerjaan alumni yang merupakan format data akhir yang didapat berdasarkan atribut yang telah dikelompokkan atau diklasifikasi.

2. Menghitung Entropy dan Gain Ratio

Tahap pembangunan pohon keputusan dilakukan setelah selesai menghitung nilai *entropy* total, nilai *entropy* dari tiap-tiap atribut, menghitung nilai *gain ratio*, menghitung nilai *split info*, dan menghitung nilai *gain ratio* serta selanjutnya menentukan nilai *gain ratio* tertinggi, hal ini dapat dilihat di Tabel 1 berikut ini :

Tabel 1 Node 1

Node	Atribut	Pemerintah	Swasta	Total	Entropy	Gain	Split Info	Gain Ratio
	Total	72	187	259	0.85270			
1	FAKULTAS			259		0.06549	1.55709	0.04206
	Ekonomi	18	81	99	0.68403			
	Teknik	10	53	63	0.63126			
	Kedokteran	44	53	97	0.99378			
2	TAHUN LULUS			259		0.01625	1.09948	0.01478
	2011-2012	42	84	126	0.91829		0.50571	
	2013-2014	30	99	129	0.78244		0.50084	
	2015-2016	0	4	4	0		0.09292	
3	IPK			259		0.00201	1.39906	0.00143
	<3	7	25	32	0.75784		0.37273	
	3-3.5	37	95	132	0.85586		0.49559	
	>3.5	28	67	95	0.87476		0.53073	
4	ANGKATAN			259		0.00208	0.33445	0.00624
	2001-2006	3	13	16	0.69621		0.2481	
	2007-2013	69	174	243	0.86077		0.08631	

Dengan menggunakan rumus (1), maka nilai *Entropy* dapat dihitung. Maka menghitung nilai *Entropy* Total dilakukan dengan cara menghitung jumlah keputusan "PEMERINTAH" dan "SWASTA" dari seluruh jumlah kasus yang ada.

$$Entropy (Total) = \left(\frac{-72}{259} \times \log_2 \left(\frac{72}{259} \right) \right) + \left(\frac{-187}{259} \times \log_2 \left(\frac{187}{259} \right) \right) = 0.85270$$

Hasil *Entropy* total diperoleh dari menghitung nilai total keputusan PEMERINTAH dengan jumlah 72 kasus dan SWASTA dengan jumlah 187 kasus, sedangkan 259 adalah jumlah dari keseluruhan kasus. Kemudian menghitung nilai *Entropy* dari masing-masing nilai atribut yang ada. Dengan menggunakan rumus (2) dapat menghitung nilai *Information Gain* tiap-tiap atribut.

$$Gain (Total, Fakultas) = 0.085270 - \left(\left(\frac{99}{259} \times 0.68403 \right) + \left(\frac{63}{259} \times 0.63126 \right) + \left(\frac{97}{259} \times 0.99378 \right) \right) = 0.06549$$

$$Gain (Total, Tahun Lulus) = 0.085270 - \left(\left(\frac{126}{259} \times 0.91829 \right) + \left(\frac{129}{259} \times 0.78244 \right) + \left(\frac{4}{259} \times 0 \right) \right) = 0.01625$$

$$Gain Ratio (Fakultas) = \frac{0.06549}{1.55709} = 0.0426$$

$$Gain (Total, Angkatan) = 0.085270 - \left(\left(\frac{16}{259} \times 0.68403 \right) + \left(\frac{243}{259} \times 0.86077 \right) \right) = 0.00208$$

Selanjutnya menghitung nilai *SplitInfo* dari masing-masing nilai atribut dengan menggunakan rumus (3).

$$SplitInfo (Fakultas, Ekonomi) = \left(\frac{-99}{259} \times \log_2 \left(\frac{99}{259} \right) \right) = 0.53303$$

$$SplitInfo (Fakultas, Teknik) = \left(\frac{-63}{259} \times \log_2 \left(\frac{63}{259} \right) \right) = 0.49610$$

$$SplitInfo (Fakultas, Kedokteran) = \left(\frac{-97}{259} \times \log_2 \left(\frac{97}{259} \right) \right) = 0.53065$$

dan Total dari *SplitInfo* Fakultas adalah 1.55709

$$SplitInfo (Tahun Lulus, 2011-2012) = \left(\frac{-126}{259} \times \log_2 \left(\frac{126}{259} \right) \right) = 0.50571$$

$$SplitInfo (Tahun Lulus, 2013-2014) = \left(\frac{-129}{259} \times \log_2 \left(\frac{129}{259} \right) \right) = 0.50084$$

$$SplitInfo (Tahun Lulus, 2015-2016) = \left(\frac{-4}{259} \times \log_2 \left(\frac{4}{259} \right) \right) = 0.09292$$

dan Total dari *SplitInfo* Tahun Lulus adalah 1.09948

$$SplitInfo (IPK, <3) = \left(\frac{-32}{259} \times \log_2 \left(\frac{32}{259} \right) \right) = 0.37273$$

$$SplitInfo (IPK, 3-3.5) = \left(\frac{-132}{259} \times \log_2 \left(\frac{132}{259} \right) \right) = 0.49559$$

$$SplitInfo (IPK, >3) = \left(\frac{-95}{259} \times \log_2 \left(\frac{95}{259} \right) \right) = 0.53073$$

dan Total dari *SplitInfo* IPK adalah 1.39906

$$SplitInfo (Angkatan, 2001-2006) = \left(\frac{-16}{259} \times \log_2 \left(\frac{16}{259} \right) \right) = 0.24814$$

$$SplitInfo (Angkatan, 2007-2013) = \left(\frac{-243}{259} \times \log_2 \left(\frac{243}{259} \right) \right) = 0.08631$$

dan Total dari *SplitInfo* Angkatan adalah 0.33445

Selanjutnya menghitung semua *SplitInfo* dari nilai atribut yang ada seperti cara di atas, sampai mendapatkan nilai Total *SplitInfo* dari nilai atribut yang ada.

$$Gain (Total, IPK) = 0.085270 - \left(\left(\frac{32}{259} \times 0.75787 \right) + \left(\frac{132}{259} \times 0.85586 \right) + \left(\frac{95}{259} \times 0.87476 \right) \right) = 0.00201$$

$$Gain Ratio (Tahun Lulus) = \frac{0.01625}{1.09948} = 0.01478$$

$$Gain Ratio (IPK) = \frac{0.00201}{1.39906} = 0.00143$$

$$Gain Ratio (Angkatan) = \frac{0.00208}{0.33445} = 0.00624$$

Lakukan perhitungan sampai mendapatkan pohon keputusan.

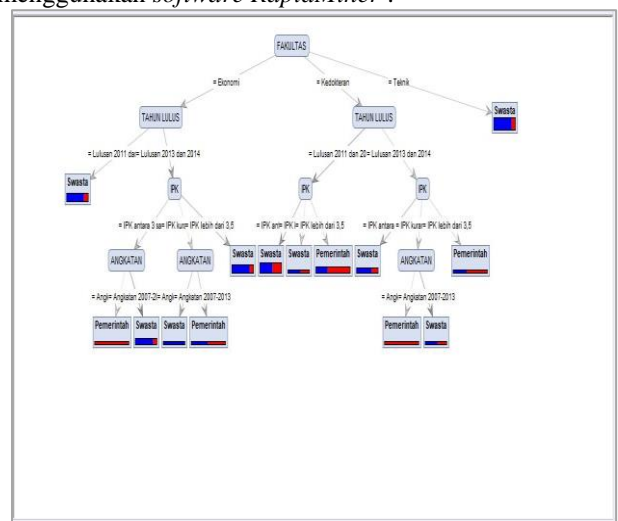
Adapun aturan atau rule yang terbentuk dari pohon keputusan (*decision tree*) terakhir berdasarkan gambar 3 adalah sebagai berikut :

1. *If* Fakultas = Ekonomi *And* Tahun Lulus = Lulusan 2011 dan 2012 *Then* Jenis Pekerjaan = Swasta
2. *If* Fakultas = Ekonomi *And* Tahun Lulus = Lulusan 2013 dan 2014, IPK = IPK antara 3 sampai 3,5, Angkatan = Angkatan 2001-2006 *Then* Jenis Pekerjaan = Pemerintah
3. *If* Fakultas = Ekonomi *And* Tahun Lulus = Lulusan 2013 dan 2014, IPK = IPK antara 3 sampai 3,5, Angkatan = Angkatan 2007-2013 *Then* Jenis Pekerjaan = Swasta
4. *If* Fakultas = Ekonomi *And* Tahun Lulus = Lulusan 2013 dan 2014, IPK = IPK kurang dari 3, Angkatan = Angkatan 2001-2006 *Then* Jenis Pekerjaan = Swasta
5. *If* Fakultas = Ekonomi *And* Tahun Lulus = Lulusan 2013 dan 2014, IPK = IPK kurang dari 3, Angkatan = Angkatan 2007-2013 *Then* Jenis Pekerjaan = Pemerintah
6. *If* Fakultas = Kedokteran *And* Tahun Lulus = Lulusan 2011-2012, IPK = IPK antara 3 sampai 3.5 *Then* Jenis Pekerjaan = Swasta

7. *If* Fakultas = Ekonomi *And* Tahun Lulus = Lulusan 2013 dan 2014, IPK = IPK lebih dari 3,5 *Then* Jenis Pekerjaan = Swasta
8. *If* Fakultas = Kedokteran *And* Tahun Lulus = Lulusan 2011-2012, IPK = IPK kurang dari 3 *Then* Jenis Pekerjaan = Swasta
9. *If* Fakultas = Kedokteran *And* Tahun Lulus = Lulusan 2011-2012, IPK = IPK lebih dari 3,5 *Then* Jenis Pekerjaan = Swasta
10. *If* Fakultas = Kedokteran *And* Tahun Lulus = Lulusan 2013-2014, IPK = IPK antara 3 sampai 3,5 *Then* Jenis Pekerjaan = Swasta
11. *If* Fakultas = Kedokteran *And* Tahun Lulus = Lulusan 2013-2014, IPK = IPK kurang dari 3, Angkatan = Angkatan 2001-2006 *Then* Jenis Pekerjaan = Pemerintah
12. *If* Fakultas = Kedokteran *And* Tahun Lulus = Lulusan 2013-2014, IPK = IPK kurang dari 3, Angkatan = Angkatan 2007-2013 *Then* Jenis Pekerjaan = Swasta
13. *If* Fakultas = Kedokteran *And* Tahun Lulus = Lulusan 2013-2014, IPK = IPK lebih dari 3,5 *Then* Jenis Pekerjaan = Pemerintah
14. *If* Fakultas = Teknik *Then* Jenis Pekerjaan = Swasta

4.4.2. Pengujian Data Menggunakan Software RapidMiner

Berikut adalah hasil proses pengujian yang dilakukan menggunakan *software RapidMiner* :



Gambar 3. Pohon Keputusan

Berdasarkan pohon keputusan pada gambar di atas, dapat dilihat bahwa pengaruh yang dominan terhadap jenis pekerjaan alumni adalah Fakultas karena Fakultas dijadikan sebagai akar atau *root* dari pohon keputusan tersebut. Yang dapat menjadi sebuah akar pada pohon keputusan adalah atribut yang memiliki nilai *gain ratio* tertinggi dibandingkan dengan atribut yang lainnya.

Sedangkan pertimbangan terakhir yang berpengaruh pada jenis pekerjaan adalah Angkatan karena berada pada *node* terakhir. Gambar 5. menunjukkan deskripsi lengkap tentang pohon keputusan (*decision tree*) yang terbentuk dari 259 data alumni yang berasal dari 3 fakultas yaitu fakultas Ekonomi, fakultas Teknik, dan fakultas

Kedokteran untuk menentukan jenis pekerjaan pemerintah dan swasta. Berdasarkan hasil pengujian yang telah dilakukan pada data alumni telah diperoleh sebuah klasifikasi jenis pekerjaan alumni. Dan hasil dari pengujian tersebut dapat digunakan sebagai informasi strategis yang dapat dijadikan sebagai sebuah pengetahuan. Sehingga pengetahuan ini dapat dijadikan sebagai pendukung keputusan atau suatu kebijakan strategis untuk Universitas Muhammadiyah Yogyakarta.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan uraian pada bab-bab sebelumnya yang telah dibahas, maka dapat ditarik kesimpulan sebagai berikut :

1. Berdasarkan pohon keputusan (*decision tree*) atribut yang paling berpengaruh terhadap jenis pekerjaan alumni adalah Fakultas. Hal itu terjadi karena fakultas sebagai akar dari pohon keputusannya, dan memiliki nilai *gain ratio* yang paling tinggi di antara atribut yang lain, selain atribut fakultas masih ada atribut lain yang berpengaruh yaitu tahun lulus, dan ipk.

Sedangkan angkatan tidak terlalu berpengaruh terhadap jenis pekerjaan alumni.

2. Berdasarkan hasil penelitian penerapan metode algoritma C4.5 cocok diterapkan pada proses klasifikasi jenis pekerjaan alumni fakultas Ekonomi, fakultas Teknik, dan fakultas Kedokteran.
3. Pada *dataset* jenis pekerjaan alumni telah diterapkan salah satu metode *data mining* yaitu klasifikasi menggunakan algoritma C4.5 dan telah dilakukan pengujian menggunakan *Software RapidMiner* menghasilkan klasifikasi yang sama.

5.2 Saran

1. Sumber data terkait alumni belum dikelola dengan baik, karena banyak terdapat data yang kosong sehingga dapat berpengaruh terhadap proses *data mining*, diharapkan kedepannya data terkait alumni bisa jauh lebih lengkap.
2. Kepedulian alumni dalam mengisi data *tracer study* masih kurang, karena jumlah data yang ada relatif sedikit.
3. Berdasarkan analisis dari penelitian ini dapat dikembangkan menjadi sebuah *web mining*.
4. Pada penelitian selanjutnya dapat ditambahkan atribut masa studi.

Referensi

- Adhatrao, K., Gaykar, A., Dhawan, A., Jha, R., & Honrao, V. (2013). Predicting Students' Performance Using Id3 and C4.5 Classification Algorithms. *International Journal of Data Mining and Knowledge Management Process*, 3(5), 39–52. <https://doi.org/10.5121/ijdkp.2013.3504>
- Pambudi, R. H., & Setiawan, B. D. (2018). Penerapan Algoritma C4 . 5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(7), 2637–2643. Retrieved from <http://j-ptiik.ub.ac.id5>
- Saefulloh, A., & Moedjiono. (2013). Penerapan Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Tepat Waktu. *InfoSys Journal*, 2(1), 41–54. <https://doi.org/10.1017/CBO9781107415324.004>