

**IMPLEMENTASI ALGORITMA C4.5 UNTUK MENENTUKAN
UANG KULIAH TUNGGAL (UKT) MAHASISWA
(Studi Kasus : Mahasiswa UMRAH Tahun 2018)**

Farida Astuti, Martaleli Bettiza , Nurfalinda
faridaastutie@gmail.com

Jurusan Teknik Informatika, Fakultas Teknik, Universitas Maritim Raja Ali Haji

Abstract

Lecture is an advanced learning process to increase insight, knowledge, interests and talents as well as the ability of a person with the desired choice of majors or expertise. However, higher education is also needed for financial planning for tuition. C4.5 algorithm is a decision tree classification algorithm that is widely used because it has the main advantages of other algorithms. In this UKT study has 4 attributes such as father's occupation, mother's occupation, gender number of siblings. Data from UKT students are divided into 5 groups in the form of parental income. The results of this study obtained 4 patterns of UKT classification obtained from the C4.5 algorithm. This system produces an accuracy of 50,00% with 360 testing data. However, this system cannot cross validation because all data are as testing data.

Keywords : : Clasification Single Lecture Money, Tanjungpinang.

I. Pendahuluan

Perkuliah merupakan proses pembelajaran tingkat lanjut untuk meningkatkan wawasan, ilmu, minat dan bakat serta kemampuan seseorang dengan pilihan jurusan atau keahlian yang diinginkan. Tetapi, untuk mengenyam pendidikan tinggi diperlukan pula perencanaan keuangan untuk biaya perkuliahan. Setiap Perguruan Tinggi Negeri mempunyai tarif Uang Kuliah Tunggal (UKT) yang berbeda-beda, hal ini dipengaruhi oleh kemahalan wilayah serta program studinya (Permendikbud No.55,2013).

Mengacu kepada Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 55 Tahun 2013 tentang Biaya Kuliah Tunggal dan Uang Kuliah Tunggal pada Perguruan Tinggi Negeri di Lingkungan Kementerian Pendidikan dan Kebudayaan. Melihat pada pasal 1 ayat 3 yang berbunyi, "Uang kuliah tunggal merupakan sebagian biaya kuliah tunggal yang ditanggung setiap mahasiswa berdasarkan kemampuan ekonominya".

Berdasarkan hasil wawancara pada bagian akademik di universitas maritim raja ali haji, Penentuan golongan uang kuliah tunggal mahasiswa baru pada kampus tersebut mempunyai persentase pada tiap golongan nya, apabila sudah melebihi presentase tiap golongan, maka otomatis langsung ke golongan berikutnya. Untuk penentuan golongan ukt, mahasiswa mengisi formulir yang disediakan oleh pihak kampus pada saat pendaftaran ulang ini bertujuan untuk mengetahui besaran pendapatan orangtua mahasiswa, kemudian data direkap dan diverifikasi oleh fakultas masing-masing. Mengingat jumlah mahasiswa yang begitu banyak, pihak universitas memerlukan sebuah aplikasi untuk menentukan kategori yang tepat dalam penetapan UKT pada setiap mahasiswa. Ada beberapa faktor yang mempengaruhi penetapan UKT tersebut, antara lain jenis kelamin, pekerjaan ayah, pekerjaan ibu, jumlah saudara dan penghasilan orang tua. Oleh karena itu

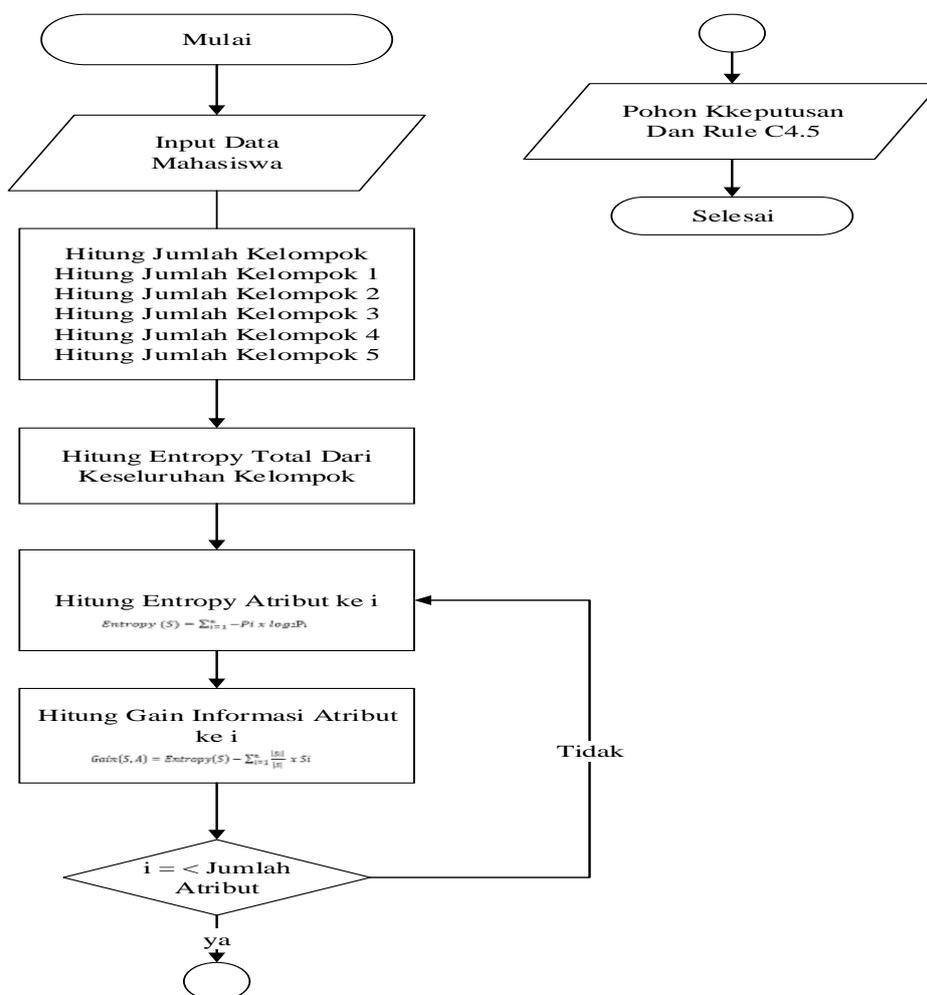
untuk pengolahan datanya, digunakan salah satu metode klasifikasi, metode yang sering digunakan adalah decision tree (Santoso, 2010).

Algoritma C4.5 merupakan algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki kelebihan utama dari algoritma yang lainnya. Kelebihan algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani atribut bertipe diskret dan dapat menangani atribut bertipe diskret dan numerik.

Oleh sebab itu peneliti melakukan penelitian dan membuat sebuah sistem yang dapat mendeteksi bagaimana menentukan besaran uang kuliah tunggal di Universitas Maritim Raja Ali Haji dengan menggunakan C4.5.

II. Metode Penelitian

Metode peramalan merupakan pengklasifikasian uang kuliah tunggal mahasiswa dengan cara mengelompokkan penghasilan orang tua berdasarkan pekerjaan ayah, pekerjaan ibu, jenis kelamin, jumlah saudara, data yang digunakan data mahasiswa UMRAH tahun 2018, dengan demikian metode pengklasifikasian diharapkan dapat memberikan objektivitas yang lebih besar. Selain pada itu metode tersebut, cara pengerjaannya yang teratur dan terarah, dengan demikian dapat dimungkinkannya penggunaan teknik analisis yang lebih maju.



Gambar 1. Flowchart Proses Utama

Keterangan

Berikut penjelasan alur kerja dari algoritma C4.5

1. Input seluruh data mahasiswa yang akan dihitung
2. Keseluruhan data akan diproses dan dihitung jumlah keseluruhan kelompok, jumlah kelompok 1, kelompok 2 kelompok 3 kelompok 4 kelompok 5
3. Menghitung *entropy* total dari keseluruhan jumlah kasus
4. Menghitung *entropy* dari tiap-tiap nilai atribut
5. Melakukan perhitungan *gain* terhadap tiap atribut
6. *Gain* tertinggi akan dijadikan sebagai *root* pada pohon keputusan.
7. Mengulangi proses 3 dan 4 untuk menentukan *internal node* dan *leafnode*. Proses akan terus dilakukan sampai semua atribut habis dieksekusi dan mencapai end of tree atau sampai diketahui golongan penghasilan ketika *leafnode* terisi dan terbentuk pohon keputusan C4.5 yang nantinya akan direpresentasikan ke dalam bentuk *rule* C4.5.

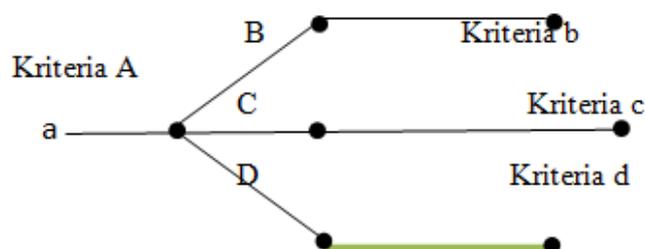
2.1 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. Secara sederhana, *data mining* atau penambangan data dapat didefinisikan sebagai proses seleksi, eksplorasi, dan pemodelan dari sejumlah besar data untuk menemukan pola atau kecenderungan yang biasanya tidak disadari keberadaannya. *Data mining* dapat dikatakan sebagai proses mengekstrak pengetahuan dari sejumlah besar data yang tersedia. (Pramadhani dan Setiadi, 2014)

2.2 Pohon Tree

Decision Tree atau pohon keputusan adalah sebuah struktur pohon, dimana setiap node pohon merepresentasikan atribut yang telah diuji, setiap cabang merupakan suatu pembagian hasil uji, dan node daun (*leaf*) merepresentasikan kelompok kelas tertentu. Level *node* teratas dari sebuah *Decision Tree* adalah node akar (*root*) yang biasanya berupa atribut yang paling memiliki pengaruh terbesar pada suatu kelas tertentu (Julianto dkk, 2014).

Pada umumnya *Decision Tree* melakukan strategi pencarian secara top-down untuk solusinya. Pada proses mengklasifikasi data yang tidak diketahui, nilai atribut akan diuji dengan cara melacak jalur dari node akar (*root*) sampai node akhir (*leaf*) dan kemudian akan diprediksi kelas yang dimiliki oleh suatu data baru tertentu.



Gambar 2. Susunan Pohon Keputusan Sumber: Julianto dkk (2014)

2.3 Algoritma C4.5

Algoritma C4.5 dan pohon keputusan merupakan dua model yang tak terpisahkan, karena untuk membangun sebuah pohon keputusan dibutuhkan Algoritma C4.5. Algoritma C4.5 merupakan generasi baru dari algoritma ID3 yang dikembangkan oleh J.Ross Quinlan pada tahun 1983. Algoritma ini mempunyai prinsip dasar kerja yang sama dengan algoritma ID3. Perbedaan utama C4.5 dari ID3 adalah (Luthfi, 2009) : C4.5 dapat menangani atribut kontinyu dan diskrit, C4.5 dapat menangani training data dengan missing value (dapat menangani data atribut yang hilang), C4.5 dapat menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain, Hasil pohon keputusan C4.5 akan dipangkas setelah dibentuk.

Langkah 1: Menghitung nilai entropy, untuk menghitung nilai entropy digunakan rumus persamaan (1) sebagai berikut

$$Entropy(S) = \sum_{i=1}^n -P_i \times \log_2 P_i \quad (1)$$

Keterangan

S : Himpunan kasus

n : Jumlah partisi S

Pi : Proporsi dari Si terhadap S

Langkah 2: Menghitung nilai gain untuk masing masing atribut. Untuk menghitung nilai *gain* digunakan rumus persamaan (2)

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

Keterangan

S : Himpunan Kasus

A : Atribut

n : Jumlah partisi atribut A

|Si| : jumlah Kasus pada partisi ke-i

|S| : Jumlah kasus dalam S

Langkah 3: Atribut yang memiliki *gain* tertinggi dipilih menjadi akar (*root*) dan atribut yang memiliki nilai *gain* lebih rendah dari akar (*root*) dipilih menjadi cabang (*branches*).

Langkah 4: Menghitung lagi nilai *gain* tiap-tiap atribut dengan tidak mengikutsertakan atribut yang terpilih menjadi akar (*root*) di tahap sebelumnya. Sama halnya dengan langkah sebelumnya, sebelum mencari nilai *gain* terlebih dahulu mencari nilai *entropy*.

Langkah 5: Atribut yang memiliki *gain* tertinggi dipilih menjadi cabang (*branches*).

Langkah 6: Mengulangi langkah ke-4 dan ke-5 sampai dengan semua kasus pada cabang memiliki kelas yang sama atau tidak ada atribut didalam cabang yang kosong.

Adapula perhitungan dari beberapa atribut dilakukan secara terpisah karena menggunakan teknik *Laplacian Smoothing*. Menurut Cahyanti, dkk (2015) teknik *Laplacian Smoothing* merupakan metode yang biasa digunakan dalam klasifikasi yang biasa dikenal dengan nama *add one smoothing*,

2.4 Confusion Matriks Dan Akurasi

Menurut Han dan Kamber (2006), Confusion matrix adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Akurasi dalam klasifikasi adalah persentase ketepatan record data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi.

Tabel 1. Model Confusion Matrixs (Ganesan 2014)

		Kelas sebenarnya		
		Class A	Class B	Class C
Kelas Prediksi	Class A	<i>a</i> (TP A) (TN B) (TN C)	<i>b</i> (FP A) (FN B)	<i>c</i> (FP A) (FN C)
	Class B	<i>d</i> (FN A) (FP B)	<i>e</i> (TP B) (TN A) (TN C)	<i>f</i> (FN C) (FP B)
	Class C	<i>g</i> (FN A) (FP C)	<i>h</i> (FN B) (FP C)	<i>i</i> (TP C) (TN A) (TN B)

- a. TP (True Positive)
yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.
- b. TN (True Negative)
yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
- c. FN (False Negative)
yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.
- d. FP (False Positive)
yaitu jumlah data positif namun terklasifikasi salah oleh sistem.

Untuk menilai tingkat akurasi dari sistem maupun perhitungan manual, digunakan rumus persamaan berdasarkan Swastina (2013) ditunjukkan pada persamaan (3) sebagai berikut:

$$Akurasi = \frac{Jumlah\ data\ yang\ relevan}{Jumlah\ data\ testing} \times 100\% \quad (3)$$

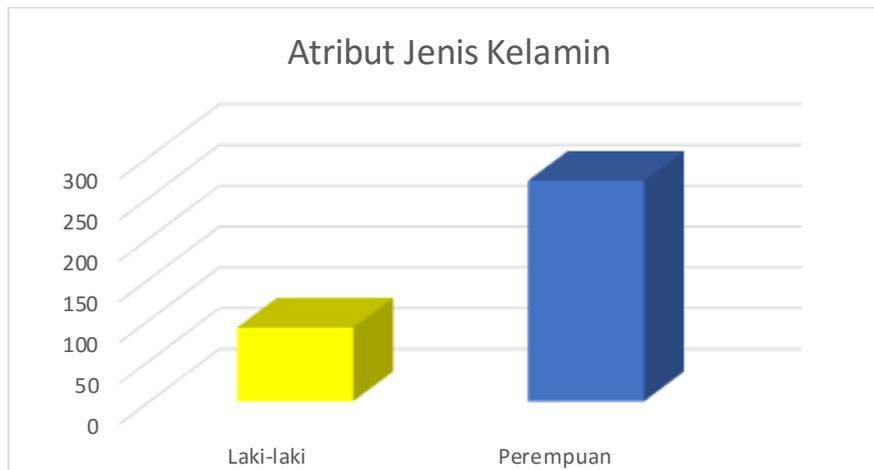
III. Hasil dan Pembahasan

3.1 Hasil Penelitian

Pembuatan pohon keputusan yang mana untuk menentukan *root* pada pohon keputusan dilakukan proses perhitungan *entropy* dan *gain information*. Atribut dengan *gain information* tertinggi inilah yang nantinya akan dijadikan *root/simpul* pada pohon keputusan. Selanjutnya untuk mengetahui predikat dari cabang atribut *gain information max* dilakukan proses perhitungan serupa terhadap cabang atribut *gain information max*, yaitu menghitung *entropy* dan *gain information*.

Setelah pohon keputusan terbentuk, selanjutnya sistem akan menampilkan hasil klasifikasi uang kuliah tunggal (UKT) mahasiswa umrah di tahun 2018 dapat menunjukkan visualisasi data secara keseluruhan yang dijabarkan sesuai atributnya.

1. Visualisasi Data Atribut Jenis Kelamin



Gambar 3. Visualisasi Data Atribut Jenis Kelamin

Keterangan gambar diatas menunjukkan bahwa:

1. Jenis kelamin laki-laki berjumlah 89 orang
2. Jenis Kelamin perempuan berjumlah 270 orang

2. Visualisasi Data Atribut Pekerjaan Ayah

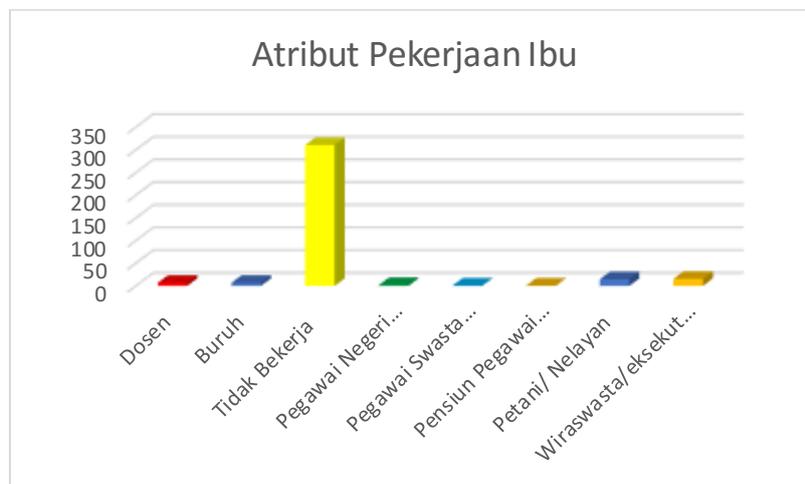


Gambar 4. Visualisasi Data Atribut Pekerjaan Ayah

Keterangan gambar diatas menunjukkan bahwa:

1. Jumlah pekerjaan ayah dosen/guru sebanyak 6 orang
2. Jumlah pekerjaan ayah tidak bekerja sebanyak 55 orang
3. Jumlah pekerjaan ayah pegawai negeri bukan dosen/guru sebanyak 28 orang
4. Jumlah pekerjaan ayah pegawai swasta bukan dosen/guru sebanyak 28 orang
5. Jumlah pekerjaan ayah pensiun pegawai negeri sebanyak 4 orang
6. Jumlah pekerjaan ayah petani/nelayan sebanyak 152 orang
7. Jumlah pekerjaan ayah profesional perorangan sebanyak 2 orang
8. Jumlah pekerjaan ayah wiraswasta/eksekutif/pedagang sebanyak 41 orang
9. Jumlah pekerjaan ayah buru sebanyak 60 orang

3. Visualisasi Data Atribut Jumlah Pekerjaan ibu



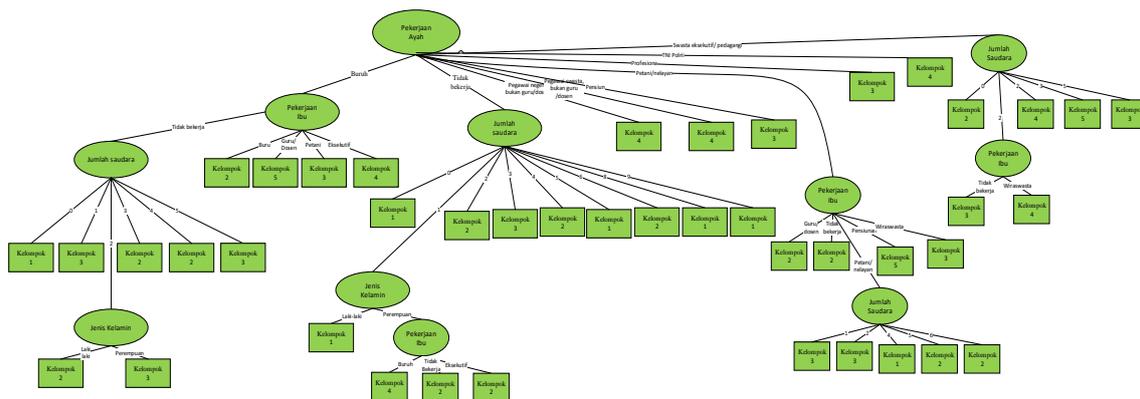
Gambar 5. Visualisasi Data Atribut Pekerjaan Ibu

Keterangan gambar diatas menunjukkan bahwa:

1. Jumlah pekerjaan ibu dosen/guru sebanyak 7 orang
2. Jumlah pekerjaan ibu tidak bekerja sebanyak 311 orang
3. Jumlah pekerjaan ibu pegawai negeri bukan dosen/guru sebanyak 2 orang
4. Jumlah pekerjaan ibu pegawai swasta bukan dosen/guru sebanyak 1 orang
5. Jumlah pekerjaan ibu pensiun pegawai negeri sebanyak 1 orang
6. Jumlah pekerjaan ibu petani/nelayan sebanyak 15 orang
7. Jumlah pekerjaan ibu wiraswasta/eksekutif/pedagang sebanyak 16 orang
8. Jumlah pekerjaan ayah buruh sebanyak 7 orang

3.3 Pohon Keputusan

Selanjutnya dari hasil penelitian dengan atribut pekerjaan ayah, pekerjaan ibu, jenis kelamin dan jumlah saudara yang menjadi root dari pohon keputusan adalah pekerjaan ayah.



Gambar 6. Hasil Pohon Keputusan

Pada pola klasifikasi yang dihasilkan menunjukkan bahwa terdapat beberapa *rule*/aturan uang kuliah (UKT) :

1. Jika Pekerjaan ayah buruh pekerjaan ibu buruh maka kelas UKT nya kelompok 2 atau 500.001-1.000.000
2. Jika Pekerjaan ayah buruh pekerjaan ibu guru/dosen maka kelas UKT nya kelompok 5 atau 3.400.001- 4.700.000
3. Jika Pekerjaan ayah buruh pekerjaan ibu tidak bekerja jumlah saudara 0 maka kelas UKT nya kelompok 1 atau 0 – 500.000
4. Jika Pekerjaan ayah buruh pekerjaan ibu tidak bekerja jumlah saudara 1 maka kelas UKT nya kelompok 1 atau 0 – 500.000
5. Jika Pekerjaan ayah buruh pekerjaan ibu tidak bekerja jumlah saudara 3 maka kelas UKT nya kelompok 2 atau 500.001 -1.000.000
6. Jika Pekerjaan ayah buruh pekerjaan ibu tidak bekerja jumlah saudara 4 maka kelas UKT nya kelompok 2 atau 500.001 -1.000.000 dst.

3.4 Akurasi

Pengukuran data dilakukan dengan *Confusion Matrix* untuk mengevaluasi hasil dari algoritma *Decision Tree C4.5*. Menurut Swastina (2013) *Confusion matrix* merupakan sebuah tabel yang terdiri dari banyaknya baris data uji yang diprediksi benar dan tidak benar oleh model klasifikasi. Tabel ini diperlukan untuk mengukur kinerja suatu model klasifikasi. Untuk lebih detail pada tabel berikut disajikan tabel *Confusion Matrix* sistem klasifikasi uang kuliah tunggal UKT mahasiswa UMRAH berdasarkan 5 kelompok.

Tabel 3. *Confusion Matrix* Sistem Klasifikasi Uang Kuliah Tunggal

		Prediksi Asli				
		K1	K2	K3	K4	K5
Predikat Hasil Klasifikasi	K1	<i>a</i> 10	<i>b</i> 2	<i>c</i> 5	<i>d</i> 1	<i>e</i> 1
	K2	<i>f</i> 23	<i>g</i> 81	<i>h</i> 49	<i>i</i> 18	<i>j</i> 1
	K3	<i>k</i> 9	<i>l</i> 17	<i>m</i> 53	<i>n</i> 8	<i>o</i> 5
	K4	<i>p</i> 0	<i>q</i> 9	<i>r</i> 11	<i>s</i> 31	<i>t</i> 9
	K5	<i>u</i> 0	<i>v</i> 0	<i>w</i> 2	<i>x</i> 0	<i>y</i> 5
Total	Data=360	K1=42	K2=109	K3=120	K4=58	K5=11

Perhitungan akurasi dihitung berdasarkan jumlah data yang sesuai antara data aktual dengan data hasil klasifikasi sistem pada tiap kelasnya dibagi dengan jumlah data keseluruhan kelas klasifikasi. Berdasarkan tabel diatas (Tabel *Confusion Matrix* Sistem Klasifikasi Uang Kuliah Tunggal UKT dengan 5 kelompok) sebagai berikut:

Akurasi

$$= \frac{10 + 81 + 53 + 31 + 5}{10 + 2 + 5 + 1 + 1 + 23 + 81 + 49 + 18 + 1 + 9 + 17 + 53 + 8 + 5 + 0 + 9 + 11 + 31 + 19 + 0 + 0 + 2 + 0 + 5} \times 100\%$$

$$= 50,00\%$$

Pada perhitungan diatas menunjukkan bahwa dengan menggunakan metode Algoritma C4.5 menunjukkan hasil akurasi sebesar 50,00%.

IV. Kesimpulan

Adapun kesimpulan yang diambil dalam penelitian ini sebagai berikut:

1. Pada sistem klasifikasi ini dapat diterapkan untuk mengetahui uang kuliah tunggal (UKT) mahasiswa dengan atribut pekerjaan ayah, pekerjaan ibu, jenis kelamin dan jumlah saudara dimana akurasi sistem sebesar 50,00% data yang gunakan sebanyak 360 data dan 44 *rule*/aturan yang dibentuk kedalam pohon keputusan.
2. Sistem ini dapat memberikan informasi uang kuliah tunggal mahasiswa pada tahun 2018 dan membantu pihak pihak kampus untuk mendata uang kuliah tunggal mahasiswa (UKT)

Saran

Adapun saran dari penelitian kedepannya adalah sebagai berikut

1. Penelitian selanjutnya diharapkan dapat memodelkan klasifikasi uang kuliah tunggal (UKT) mahasiswa UMRAH dengan metode klasifikasi lainnya.
2. Penelitian selanjutnya sebaiknya menambah variabel yang digunakan, tidak hanya menggunakan pekerjaan ayah, pekerjaan ibu, jenis kelamin, jumlah saudara namun juga menambah variabel-variabel yang mempengaruhi lainnya

V. Daftar Pustaka

- Adhatrao, K., Dhawan, A., Jha, R., & Honrao, V., 2015, Predicting Students' Performance Using Id3 And C4.5 Classification Algorithms, Vol. 3, No. 5 International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.3, No.5,.
- Arifin & Fitriyah., 2018, Penerapan Algoritma Klasifikasi C4.5 dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada, DOI: 10.22441/incomtech.v8i1.2198.
- Arista, N., 2015, Penerapan Algoritma ID3 Dan C4.5 Dalam Menemukan Hubungan Data Awal Masuk Mahasiswa Dengan Prestasi Akademik, *Skripsi*, Universitas Maritim Raja Ali Haji, Tanjungpinang.
- Ginting, S. L. Br., Zarman, W., & Hamidah, I., 2014, Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik, Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST), Yogyakarta. ISSN: 1979-911X
- Gorunescu, Florin. (2011). Data Mining: Concepts and Techniques. Verlag berlin Heidelberg: Springer
- Han, J. and Kamber, M, 2006, "Data Mining Concepts and Techniques Second Edition". Morgan Kauffman, San Francisco. ISBN 13: 978-1-55860-901-3
- Julianto, W., Yunitarini, R., & Sophan, M. K., 2014, Algoritma C4.5 Untuk Penilaian Kinerja Karyawan, SCAN VOL. IX NOMOR 2 JUNI 2014, ISSN : 1978-0087
- Kamagi, D. H., & Hansun, S., 2014, Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa, *ULTIMATICS*, Vol. VI, No. 1, Juni 2014.
- Kusrini & Luthfi, E. T, 2009. ALGORITMA DATA MINING, I ed., Theresia Ari Prabawati, Ed. Yogyakarta, Indonesia:
- Mahardhika, A. A., Saptono, R., & Anggrainingsih, R., 2015, Sistem Klasifikasi Feedback Pelanggan dan Rekomendasi Solusi Atas Keluhan Di UPT PUSKOM UNS dengan Algoritma Naïve Bayes Classifier dan Cosine Similarity, *JURNAL ITSMART*, Vol 4. No 1. Juni 2015 ISSN : 2301-7201.
- Pambudi, R. H., Setiawan, B. D, & Indriawati, T., 2018, Penerapan Algoritma C4.5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal, Vol. 2, No. 7 (2015) hlm. 2637-2643.
- Rahmayuni, I., 2014, Perbandingan Performansi Algoritma C4.5 Dan Cart Dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Komputer Politeknik Negeri Padang, *Jurnal TEKNOIF* , Vol. 2 No. 1 (2014) ISSN : 2338-2724
- Republik Indonesia. 2013. "Peraturan Menteri Pendidikan dan Kebudayaan No. 55 Tahun 2013 Tentang Biaya Kuliah Tunggal dan Uang Kuliah Tunggal pada Perguruan Tinggi Negeri di Lingkungan Kementerian Pendidikan dan Kebudayaan. Kementerian Pendidikan dan Kebudayaan Republik Indonesia. Jakarta
- Sari, S. I., 2019, Pola Penyakit Berdasarkan Profil Pasien Menggunakan Algoritma C4.5, *Skripsi*, Universitas Maritim Raja Ali Haji, Tanjungpinang.
- Sitorus, C. I., 2017, Prediksi Tingkat Keberhasilan Siswa Menggunakan Metode Decision Tree Algoritma C4.5, *Skripsi*, Universitas Maritim Raja Ali Haji, Tanjungpinang.
- Swastina, L., 2013, Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa, *Jurnal GEMA AKTUALITA*, Vol. 2 No. 1, Juni 2013.