

Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5

Chairun Nas

Program Studi Manajemen Informatika, Universitas Catur Insan Cendekia, Cirebon, Indonesia
e-mail: chairun.nas@cic.ac.id

Abstrak

Kota Cirebon memiliki banyak perguruan tinggi swasta yang beragam dan menawarkan berbagai macam jurusan. Dikarenakan banyaknya perguruan tinggi swasta yang ada, maka persaingan antar perguruan tinggi untuk menarik minat calon mahasiswa baru sangatlah ketat. Maka untuk dapat bersaing dengan perguruan tinggi tersebut, pihak kampus perlu melakukan prediksi terhadap minat calon mahasiswa baru dengan mengetahui faktor apa yang menjadi pendorong calon mahasiswa baru memilih perguruan tinggi. Tujuan penelitian ini adalah untuk memprediksi minat calon mahasiswa dalam memilih perguruan tinggi. Dalam penelitian ini, data yang diolah adalah hasil survei yang dilakukan pada calon mahasiswa baru pada Program Studi D3-Manajemen Informatika. Terdapat 25 data sampel sebagai training dataset dengan 5 data penilaian. Data tersebut akan diolah menggunakan proses Data Mining dengan Metode Klasifikasi. Selanjutnya dengan algoritma C4.5 pada metode Klasifikasi diperoleh iterasi sampai kepada node 4 dengan 3 data penilaian yang menjadi faktor dalam penentuan minat calon mahasiswa terhadap program studi. Selanjutnya hasil pengolahan data menggunakan metode klasifikasi dengan algoritma C4.5 tersebut, dilakukan pengujian dengan menggunakan Software Rapid Miner, dimana diperoleh tingkat akurasi 100% dengan 120 testing dataset. Hasil akhir penggunaan Data Mining dengan algoritma C4.5 ini telah mampu melakukan prediksi terhadap minat calon mahasiswa baru berdasarkan faktor penilaian dalam memilih perguruan tinggi.

Kata kunci: *Prediksi, Klasifikasi, Data Mining, Algoritma C4.5, Pohon Keputusan*

Abstract

The city of Cirebon has many diverse private universities and offers a variety of majors. Due to a large number of existing private universities, the competition between universities to attract prospective new students is very tight. So to be able to compete with these universities, the campus needs to predict the interest of prospective new students by knowing what factors are driving prospective new students to choose colleges. The purpose of this study was to predict the interest of prospective students in choosing a college. In this study, the data processed are the results of a survey conducted on prospective new students in the D3-Informatics Management Study Program. There are 25 sample data as training datasets with 5 assessment data. The data will be processed using the Data Mining process with the Classification Method. Furthermore, with the C4.5 algorithm in the Classification method, iterations are obtained up to node 4 with 3 assessment data which are factors in determining the interest of prospective students in the study program. Furthermore, the results of data processing using the classification method with the C4.5 algorithm were tested using Rapid Miner Software, which obtained a 100% accuracy rate with 120 testing datasets. The final result of using Data Mining with the C4.5 algorithm has been able to predict the interest of prospective new students based on assessment factors in choosing universities.

Keywords: *Predictions, Classification, Data Mining, C4.5 Algorithm, Decision Tree*

1. Pendahuluan

Perkembangan sebuah perguruan tinggi, salah satunya dilihat berdasarkan jumlah mahasiswa yang diperoleh setiap tahunnya. Pada perguruan tinggi terutama swasta,

perolehan mahasiswa setiap tahunnya menjadi faktor utama dalam berkembangnya perguruan tinggi tersebut. Dikarenakan, mahasiswa menjadi sumber daya utama bagi perguruan tinggi swasta untuk mendapat dana dalam menjalankan operasional yang ada di perguruan tinggi. Semakin banyaknya mahasiswa, maka akan semakin besar pemasukan bagi perguruan tinggi, sehingga perguruan tinggi mudah dalam menjalankan operasional kampus tanpa keterbatasan dana. Saat ini Program Studi Manajemen Informatika, Universitas Catur Insan Cendekia dari tahun ke tahun mengalami penurunan jumlah mahasiswa. Dari tahun 2018-2020 terjadi penurunan jumlah mahasiswa yang mendaftar sebesar 20%. Maka saat ini Program Studi berusaha semaksimal mungkin untuk mendapatkan calon mahasiswa baru untuk dapat melanjutkan pendidikan di Program Studi Manajemen Informatika, Universitas Catur Insan Cendekia.

Dalam memilih perguruan tinggi, seorang mahasiswa biasanya akan mencari informasi mengenai perguruan tinggi yang dituju. Informasi yang dibutuhkan tersebut antara lain adalah biaya kuliah, lokasi perguruan tinggi, fasilitas, perpustakaan, laboratorium, kegiatan kemahasiswaan, akreditasi (*Brand*) serta Program Studi apa saja yang terdapat pada perguruan tinggi tersebut [1]. Selain dari itu, terdapat juga faktor-faktor yang mempengaruhi calon mahasiswa dalam memilih perguruan tinggi, antara lain faktor pengaruh orang tua, teman, pengaruh kerabat, beasiswa, peluang kerja dan lain-lain [2]. Dengan faktor-faktor tersebut, pihak perguruan dapat memprediksi faktor apa yang menjadi pendorong utama calon mahasiswa memilih perguruan tinggi, sehingga perguruan tinggi dapat membuat keputusan terbaik untuk dapat merekrut calon mahasiswa sebanyak-banyaknya.

Informasi dan faktor tersebut dapat di olah untuk mengetahui seberapa besar minat calon mahasiswa memilih sebuah perguruan tinggi. Minat dapat diartikan suatu kecenderungan dan kegairahan yang tinggi atau keinginan yang besar terhadap sesuatu. Minat dapat timbul karena daya tarik dari luar dan juga datang dari hati sanubari [3]. Untuk mengetahui faktor yang mendukung minat calon mahasiswa, maka perlu dilakukan analisis dan pengolahan data dengan menggunakan *Data Mining*. Tujuan dari penelitian ini adalah untuk mengetahui minat calon mahasiswa baru terhadap sebuah perguruan tinggi. Selain dari itu bagaimana menerapkan proses *Data Mining* dengan algoritma C4.5 untuk memprediksi minat calon mahasiswa berdasarkan faktor penilaiannya menggunakan.

Data Mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan [4]. *Data mining* merupakan bagian dari *Knowledge Discovery in Database (KDD)* yang terdiri dari beberapa tahapan seperti pemilihan data, pra pengolahan data, transformasi, *data mining* dan evaluasi hasil [5]. Dengan menggunakan *data mining* maka informasi berharga di dalam kumpulan data dapat digali.

Dengan menggunakan *data mining*, data-data minat calon mahasiswa baru dapat diolah dengan sebuah algoritma. Didalam *data mining* terdapat algoritma-algoritma yang mampu menganalisis sebuah data. Dalam penelitian ini, algoritma yang digunakan adalah algoritma C4.5. Algoritma C4.5 merupakan metode klasifikasi dan prediksi yang digunakan untuk membentuk *Decision Tree* berdasarkan data *training* [6]. *Decision tree* adalah struktur *flowchart* yang memiliki *Tree* (Pohon), dimana setiap simpul internal menandakan suatu tes atribut, setiap cabang merepresentasikan hasil tes dan simpul daun merepresentasikan kelas atau distribusi kelas [7].

Pada Penelitian Sebelumnya, dalam memprediksi minat mahasiswa digunakan pendekatan *Theory Of Planned Behavior*. Pendekatan ini mengukur minat mahasiswa dalam berwirausaha dengan mengutamakan 3 determinan faktor keinginan untuk berwitausaha, yaitu Sikap Terhadap Perilaku, Norma Subjektif, dan Kontrol Keperilakuan. Hasil dari

pendekatan ini adalah sebuah kesimpulan dimana sikap terhadap perilaku tidak mempengaruhi minat mahasiswa sedangkan 2 faktor lainnya mempengaruhi [8].

Algoritma lain yang digunakan dalam penelitian sebelumnya untuk prediksi minat siswa SMK untuk masuk Perguruan Tinggi adalah algoritma *Naive Bayes*. Data yang digunakan adalah data lulusan tahun 2018 dan 2019 dimana terdapat 158 data. Selanjutnya data diolah dengan menggunakan algoritma *Naive Bayes* dengan 9 atribut, yaitu keahlian, nilai rapor, nilai UN, pekerjaan orang tua, dan lain-lain. Data tersebut diuji dengan menggunakan aplikasi *RapidMiner*. Hasilnya dari pengolahan tersebut diperoleh prediksi atribut pendapatan orang tua, nilai rapor, dan keinginan siswa, yang mempengaruhi minat siswa untuk melanjutkan pendidikan ke perguruan tinggi dengan tingkat akurasi sebesar 92,96% [9].

Pada penelitian sebelumnya, algoritma C4.5 telah digunakan untuk memprediksi faktor penyebab mahasiswa mengulang sebuah matakuliah. Pada penelitian tersebut, ditentukan variabel berupa Jumlah Semester, IPK, Nilai, Keadaan Ekonomi dan Status. Selanjutnya dilakukan pengolahan data dengan algoritma C4.5 sebanyak 141 data record sebagai data *training*. Setelah dilakukan pengolahan data menggunakan algoritma C4.5, selanjutnya dilakukan pengujian dengan menggunakan aplikasi *WEKA*. Hasil dari penelitian ini adalah informasi berupa *rule* yang dapat memprediksi mahasiswa yang mengulang sebuah matakuliah [10].

Selanjutnya algoritma C4.5 juga digunakan untuk Klasifikasi prediksi tingkat keberhasilan mahasiswa pada AMIK Tunas Bangsa. Dalam penelitian ini, beberapa variabel atribut ditentukan seperti Jenis Kelamin, Kehadiran, Sesi Perkuliahan, Nilai rata-rata dan asal sekolah. Selanjutnya dilakukan pengolahan data dengan menggunakan algoritma C4.5. Setelah dilakukan pengolahan data dengan menggunakan algoritma C4.5, dilakukan pengujian hasil dengan menggunakan *Rapid Miner*. Hasil dari pengolahan aplikasi *Rapid Miner* ditemukan bahwa hasil keputusan yang diperoleh berdasarkan perhitungan manual dan menggunakan *Rapid Miner* menghasilkan akurasi sebesar 92% terhadap prediksi [11].

Dari penelitian sebelumnya yang telah dijelaskan, penggunaan pendekatan *Theory Of Planned Behavior* belum optimal dalam membantu pengambil keputusan. Hal ini disebabkan tingkat akurasi prediksi tidak akurat, sedangkan algoritma *Naive Bayes* dibandingkan dengan algoritma C4.5 memiliki tingkat akurasi yang rendah, dan juga tidak menggambarkan bentuk sebuah keputusan dengan *Decision Tree*. Penggunaan aplikasi *WEKA* belum optimal dalam menguji algoritma C4.5. Maka untuk itu dalam penelitian ini algoritma C4.5 akan diuji dengan menggunakan aplikasi *Rapid Miner* dengan tingkat akurasi yang tinggi. Maka *data mining* dengan menggunakan algoritma C4.5 dan diuji dengan menggunakan aplikasi *Rapid Miner* cocok digunakan untuk memprediksi minat calon mahasiswa baru dalam memilih perguruan tinggi sehingga pihak perguruan tinggi dapat mengambil keputusan terbaik dalam menarik minat calon mahasiswa baru tersebut.

2. Kajian Pustaka

2.1 Data Mining

Data Mining merupakan suatu kandungan ikhtisar dari sekelompok data yang bertujuan untuk dapat menggali informasi yang bermanfaat didalam data tersebut, dimana informasi yang berharga tersebut tidak diketahui sebelumnya [12]. *Data Mining* juga dapat diartikan suatu keilmuan yang mempelajari metode dan algoritma untuk pengekstrakan pengetahuan atau menemukan pola dari suatu data [32]. *Data Mining* banyak digunakan untuk mengelola sekumpulan data, sehingga hasil pengelolaan tersebut membantu para

pengambil keputusan untuk membuat sebuah keputusan terbaik. *Data Mining* juga merupakan bagian dari proses *Knowledge Discovery in Database (KDD)* yang merupakan tahapan dalam proses penggalian data untuk mendapatkan informasi yang bermanfaat.

2.2 Metode Klasifikasi

Metode Klasifikasi merupakan suatu metode yang menjelaskan proses penemuan model (atau fungsi) yang mendeskripsikan dan membedakan kelas data dengan tujuan untuk memprediksi kelas dari objek yang data label kelasnya tidak diketahui atau samar [14]. Dalam klasifikasi terdapat 4 komponen, yaitu *class*, *Predictor*, *training dataset*, dan *testing dataset*. Selain dari itu didalam metode Klasifikasi terdapat beberapa algoritma yang dapat digunakan dan salah satunya adalah algoritma C4.5.

2.3 Algoritma C4.5

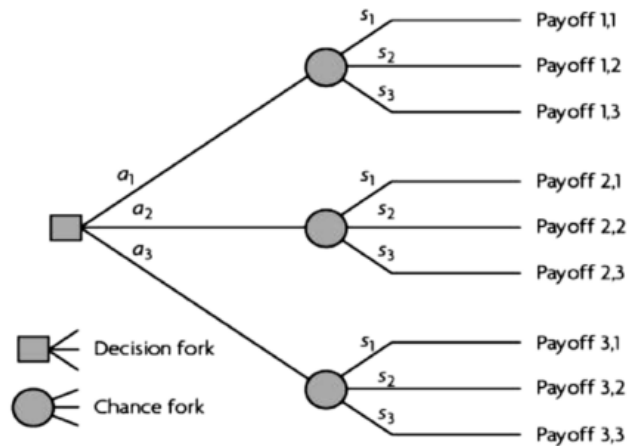
Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang melakukan perbaikan untuk mengatasi *missing data*, data *kontinu* dan *pruning* [15]. Algoritma C4.5 mengkonstruksi pohon keputusan dari data yang berupa kasus-kasus atau *record-record* (tupel) dalam data [16]. Maka dapat disimpulkan bahwa algoritma C4.5 menghasilkan sebuah keputusan dalam bentuk pohon keputusan yang diambil berdasarkan pengolahan data yang telah dilakukan.

Secara umum, algoritma C4.5 dalam membangun sebuah keputusan melakukan langkah-langkah sebagai berikut [15]:

- 1) Pilih atribut sebagai akar.
- 2) Buat Cabang untuk tiap-tiap nilai.
- 3) Bagi kasus dalam cabang.
- 4) Ulangi proses setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

2.4 Decession Tree

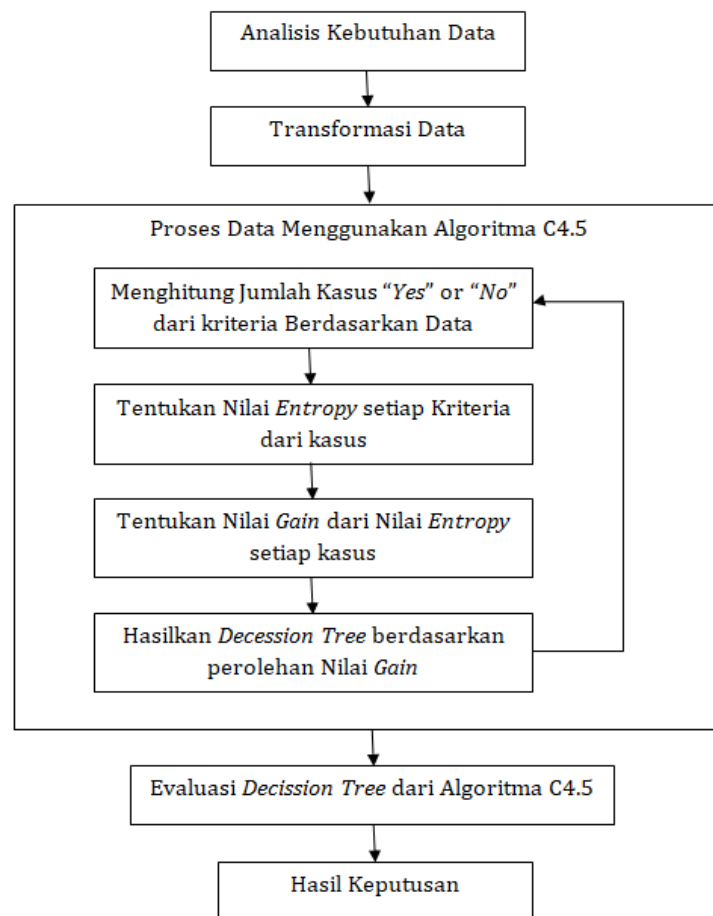
Decession Tree adalah sebuah diagram alir yang berbentuk seperti struktur pohon yang mana setiap *internal node* menyatakan pengujian terhadap suatu atribut, setiap cabang menyatakan *output* dari pengujian dan *leaf node* (bagian daun) menyatakan kelas-kelas atau distribusi kelas [17]. *Node* paling atas dari *Decession Tree* disebut sebagai *rootnode* atau *node* akar, dari *node* akar ini lah lahir *leaf node* yang merupakan hasil akhir yang mewakili label kelas dari kombinasi atribut sehingga terbentuk menjadi *rule* (aturan). Alur pada *Decession Tree* selalu ditelusuri dari *node* akar ke *node* daun yang memegang prediksi, adapun bentuk *Decession tree* secara umum dapat dilihat pada Gambar 1.



Gambar 1. Bentuk Umum *Decession Tree*

3. Metode Penelitian

Pada penelitian ini, dijelaskan tahapan-tahapan dalam proses mendapatkan keputusan terbaik dengan menggunakan algoritma C4.5 berdasarkan data yang telah diperoleh. Adapun tahapan yang dilakukan dapat dilihat pada Gambar 2.



Gambar 2. Tahapan Proses Penelitian

Dari Gambar 2, maka dapat dijelaskan tahapan penelitian sebagai berikut:

3.1 Analisis Kebutuhan Data

Kebutuhan data yang diperoleh dalam penelitian ini adalah data-data berupa data hasil survei terhadap mahasiswa dan data kriteria yang menjadi faktor dalam penentuan calon mahasiswa berminat untuk kuliah di kampus tersebut. Terdapat 5 kriteria beserta subkriteria yang menjadi faktor penentu yang dapat dilihat pada Tabel 1.

Tabel 1. Kriteria Faktor Penentu Minat Calon Mahasiswa Memilih Kampus

No	Kriteria	Subkriteria		
1	Uang Kuliah	- Murah	- Terjangkau	- Mahal
2	Fasilitas	- Sangat Baik	- Memadai	- Cukup
3	Akreditasi	- Sangat Baik	- Baik	- Cukup
4	Pelayanan	- Sangat Baik	- Baik	- Cukup
5	Lokasi	- Dekat	- Terjangkau	- Jauh

Alasan pemilihan kriteria Akreditasi, Pelayanan dan fasilitas dilihat dari segi kebutuhan mahasiswa terhadap perguruan tinggi. Dari segi Uang Kuliah dikarenakan banyaknya keluarga di kota Cirebon memiliki pendapatan perbulan menengah ke bawah dan segi lokasi, banyak mahasiswa yang datang dari luar daerah kota Cirebon. Berdasarkan kriteria faktor penentu di atas, maka dilakukan survei terhadap beberapa mahasiswa berdasarkan kriteria faktor penentu tersebut sehingga diperoleh data survei minat mahasiswa yang memilih dan tidak memilih untuk kuliah di kampus tersebut. Selanjutnya data hasil survei dilakukan Transformasi Data untuk dilakukan pengolahan.

3.2 Transformasi Data

Transformasi Data adalah perubahan data dengan mengubah nilai atribut awalnya menjadi nilai atribut yang sesuai dengan kebutuhan data dalam pengolahannya [18]. Berdasarkan hasil survei minat calon mahasiswa dalam memilih kampus terhadap kriteria, maka selanjutnya dilakukan Transformasi Data. Untuk pengujian menggunakan algoritma C4.5 maka digunakan sebagai 25 data sampel sebagai *training dataset*. Adapun hasil Transformasi Data survei dapat dilihat pada Tabel 2:

Tabel 2. Transformasi Data Berdasarkan Data Survei

MHS	Akreditasi	Uang Kuliah	Fasilitas	Pelayanan	Lokasi	Minat
MHS 1	Baik	Terjangkau	Sangat Baik	Baik	Terjangkau	Yes
MHS 2	Cukup	Murah	Cukup	Cukup	Terjangkau	Yes
MHS 3	Cukup	Terjangkau	Cukup	Cukup	Jauh	No
MHS 4	Sangat Baik	Murah	Sangat Baik	Sangat Baik	Terjangkau	Yes
MHS 5	Baik	Murah	Memadai	Baik	Jauh	Yes
MHS 6	Sangat Baik	Terjangkau	Memadai	Baik	Terjangkau	Yes
MHS 7	Cukup	Terjangkau	Cukup	Cukup	Jauh	No
MHS 8	Baik	Terjangkau	Cukup	Baik	Dekat	Yes
MHS 9	Cukup	Mahal	Cukup	Cukup	Terjangkau	No
MHS 10	Cukup	Terjangkau	Cukup	Cukup	Jauh	No
MHS 11	Baik	Murah	Sangat Baik	Sangat Baik	Dekat	Yes
MHS 12	Cukup	Terjangkau	Sangat Baik	Cukup	Jauh	Yes
MHS 13	Cukup	Terjangkau	Memadai	Baik	Terjangkau	Yes
MHS 14	Cukup	Mahal	Memadai	Baik	Dekat	No

MHS	Akreditasi	Uang Kuliah	Fasilitas	Pelayanan	Lokasi	Minat
MHS 15	Baik	Terjangkau	Memadai	Cukup	Terjangkau	Yes
MHS 16	Baik	Mahal	Cukup	Baik	Terjangkau	No
MHS 17	Baik	Terjangkau	Sangat Baik	Baik	Dekat	Yes
MHS 18	Cukup	Murah	Cukup	Cukup	Terjangkau	Yes
MHS 19	Cukup	Terjangkau	Cukup	Cukup	Jauh	No
MHS 20	Cukup	Mahal	Cukup	Cukup	Terjangkau	No
MHS 21	Sangat Baik	Murah	Sangat Baik	Baik	Terjangkau	Yes
MHS 22	Baik	Terjangkau	Cukup	Sangat Baik	Jauh	Yes
MHS 23	Baik	Terjangkau	Cukup	Cukup	Dekat	Yes
MHS 24	Sangat Baik	Murah	Sangat Baik	Sangat Baik	Terjangkau	Yes
MHS 25	Cukup	Terjangkau	Cukup	Cukup	Terjangkau	No

Data yang terdapat pada Transformasi Data merupakan data sampel yang akan digunakan untuk pengujian dengan menggunakan Algoritma C4.5.

3.3 Proses Algoritma C4.5

Berdasarkan hasil Transformasi Data yang telah diperoleh, selanjutnya dilakukan proses algoritma C4.5. Algoritma C4.5 memiliki tahapan atau langkah-langkah dalam proses penyelesaian masalah sehingga sampai dalam tahapan hasil. Adapun tahapan dalam proses Algoritma C4.5 dapat dilihat sebagai berikut [19]:

- 1) Menghitung jumlah kasus “Yes” dan “No” berdasarkan data Tranformasi.
- 2) Menentukan nilai *Entropy* dari setiap kriteria berdasarkan kasus. Untuk menghitung nilai *Entropy* dapat menggunakan rumus:

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

Dimana:

S = himpunan kasus

n = jumlah partisi S

p_i = proporsi S_i terhadap S

- 3) Menentukan nilai *Gain* tertinggi berdasarkan perolehan nilai *Entropy* yang telah di dapatkan. Untuk menghitung nilai *Gain* dapat menggunakan rumus:

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

Dimana:

S = himpunan kasus

A = fitur

n = jumlah partisi AtributA

$|S_i|$ = proporsi S_i terhadap S

$|S|$ = jumlah kasus dalam S

- 4) Hasilkan *Decission Tree* untuk menghasilkan *node* pertama, lakukan proses 1 sampai 4 sampai tidak ada *record* didalam cabang *Decission Tree* yang kosong.

Berdasarkan rumus diatas dan perolehan data yang telah dilakukan transformasi, maka adapun tahapan proses algoritma C4.5 dalam penelitian ini dapat dijelaskan pada langkah-langkah berikut. Langkah pertama adalah menentukan jumlah “Yes” dan “No” berdasarkan data Tranformasi yang telah diperoleh, setelah dilakukan penentuan jumlah “Yes” dan “No” dalam kasus, selanjutnya dilakukan proses pencarian nilai *Entropy* dan nilai *Gain* dengan

menggunakan rumus yang telah dijelaskan diatas. Untuk *Node 1* hasil dari nilai *Entropy* dan *Gain* dapat dilihat pada Tabel 3.

Tabel 3. Hasil Perhitungan Nilai *Entropy* dan *Gain* pada *Node 1*

Kategori	Sub Kategori	Jumlah Kasus	Yes	No	Entropy	Gain
Total		25	16	9	0,9427	
Akreditasi						0,3207
	Sangat Baik	4	4	0	0	
	Baik	9	8	1	0,5033	
	Cukup	12	4	8	0,9183	
Uang Kuliah						0,4161
	Murah	7	7	0	0	
	Terjangkau	14	9	5	0,9403	
	Mahal	4	0	4	0	
Fasilitas						0,2985
	Sangat Baik	7	7	0	0	
	Memadai	5	4	1	0,7219	
	Cukup	13	5	8	0,9612	
Pelayanan						0,1972
	Sangat Baik	4	4	0	0	
	Baik	9	7	2	0,7642	
	Cukup	12	5	7	0,9799	
Lokasi						0,0594
	Dekat	5	4	1	0,7219	
	Terjangkau	13	9	4	0,8905	
	Jauh	7	3	4	0,9852	

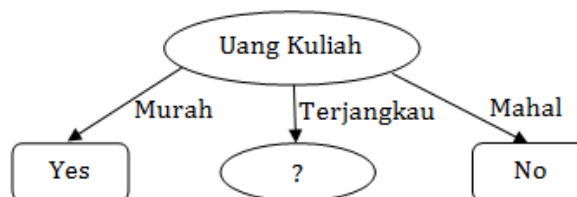
Perolehan nilai *Entropy* dari tabel diatas, sederhananya dapat dilihat sebagai berikut:

$$Entropy = \left(\frac{- Kasus YES}{Kasus Kategori} * Log_2 \frac{Kasus YES}{Kasus Kategori} \right) + \left(\frac{- Kasus NO}{Kasus Kategori} * Log_2 \frac{Kasus NO}{Kasus Kategori} \right)$$

Sedangkan nilai *Gain* dari dari tabel diatas, sederhananya dapat dilihat sebagai berikut:

$$Gain = Entropy Total - \left(\frac{Jumlah Kasus Kategori}{Jumlah Total Kasus} * Entropy Kategori \right)$$

Berdasarkan perolehan nilai *Entropy* dan *Gain* dari *Node 1*, maka diperoleh nilai *Gain* tertinggi sebesar 0,4161 pada kategori **Uang Kuliah** yang selanjutnya dapat dijadikan *Node* akar. Adapun bentuk *Decession Tree* sementara dari *Node 1* dapat dilihat pada Gambar 3.



Gambar 3. *Decession Tree* Pada *Node 1*

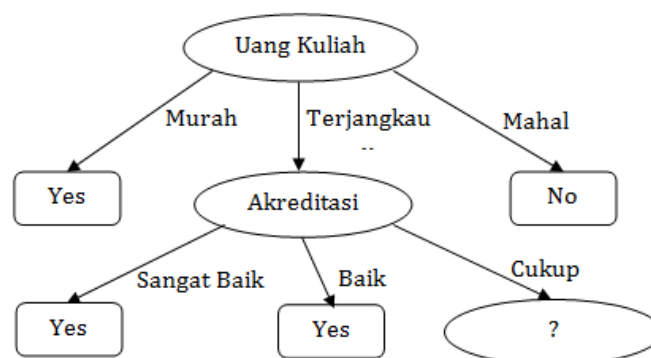
Berdasarkan *Decession Tree* diatas dijelaskan bahwa kriteria **Uang Kuliah - Murah** memiliki keputusan “Yes” bagi minat mahasiswa, sedangkan kriteria **Uang Kuliah - Mahal**

memilik keputusan “No” bagi minat mahasiswa. Langkah selanjutnya mencari nilai *Entropy* dan *Gain* berdasarkan kriteria **Uang Kuliah - Terjangkau** untuk menghasilkan *Node 2*. Adapun nilai *Entropy* dan *Gain* pada *Node 2* dapat dilihat pada Tabel 4.

Tabel 4. Hasil Perhitungan Nilai *Entropy* dan *Gain* pada *Node 2*

Kategori	Sub Kategori	Jumlah Kasus	Yes	No	Entropy	Gain
Uang Kuliah - Terjangkau		14	9	5	0,9403	
Akreditasi						0,5087
	Sangat Baik	1	1	0	0	
	Baik	6	6	0	0	
	Cukup	7	2	5	0,8631	
Fasilitas						0,3949
	Sangat Baik	3	3	0	0	
	Memadai	3	3	0	0	
	Cukup	8	3	5	0,9544	
Pelayanan						0,3949
	Sangat Baik	1	1	0	0	
	Baik	5	5	0	0	
	Cukup	8	3	5	0,9544	
Lokasi						0,2889
	Dekat	3	3	0	0	
	Terjangkau	5	4	1	0,7219	
	Jauh	6	2	4	0,9183	

Berdasarkan perolehan nilai *Entropy* dan *Gain* dari *Node 2*, maka diperoleh nilai *Gain* tertinggi sebesar 0,5087 pada kategori **Akreditasi** yang selanjutnya dapat dijadikan *Node* cabang dari *Node* akar. Adapun bentuk *Decession Tree* sementara dari *Node 2* dapat dilihat pada Gambar 4.



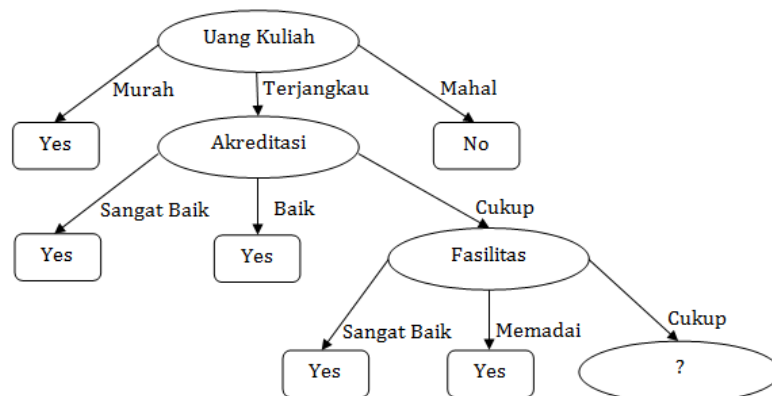
Gambar 4. *Decession Tree* Pada *Node 2*

Berdasarkan *Decession Tree* diatas dijelaskan bahwa kriteria Akreditasi – **Sangat Baik** dan **Akreditasi – Baik** memiliki keputusan “Yes” bagi minat mahasiswa. Langkah selanjutnya mencari nilai *Entropy* dan *Gain* berdasarkan Kriteria **Akreditasi - Cukup** untuk menghasilkan *Node 3*. Adapun nilai *Entropy* dan *Gain* pada *Node 3* dapat dilihat pada Tabel 5.

Tabel 5. Hasil Perhitungan Nilai *Entropy* dan *Gain* pada *Node 3*

Kategori	Sub Kategori	Jumlah Kasus	Yes	No	Entropy	Gain
Uang Kuliah – Terjangkau & Akreditasi Cukup		7	2	5	0,8631	
Fasilitas						0,8631
	Sangat Baik	1	1	0	0	
	Memadai	1	1	0	0	
	Cukup	5	1	5	0	
Pelayanan						0,3059
	Sangat Baik	0	0	0	0	
	Baik	1	1	0	0	
	Cukup	6	1	5	0,6500	
Lokasi						0,0617
	Dekat	0	0	0	0	
	Terjangkau	2	1	1	1	
	Jauh	5	1	4	0,7219	

Berdasarkan perolehan nilai *Entropy* dan *Gain* dari *Node 3*, maka diperoleh nilai *Gain* tertinggi sebesar 0,8631 pada kategori **Fasilitas** yang selanjutnya dapat dijadikan *Node* cabang dari *Node* akar. Adapun bentuk *Decession Tree* sementara dari *Node 3* dapat dilihat pada Gambar 5.



Gambar 5. *Decession Tree* Pada *Node 3*

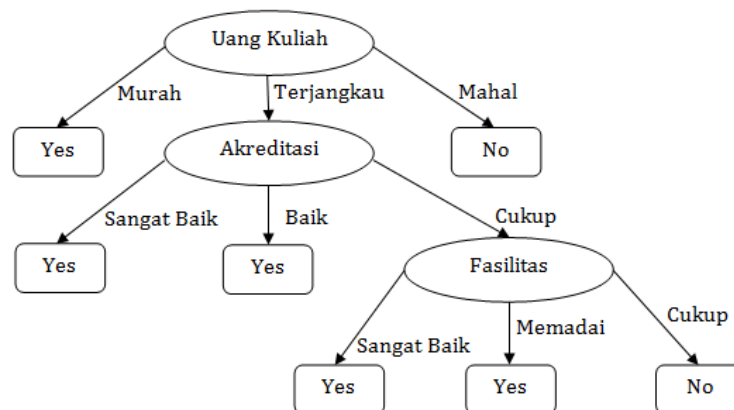
Berdasarkan *Decession Tree* diatas dijelaskan bahwa kriteria **Fasilitas – Sangat Baik** dan **Fasilitas – Memadai** memiliki keputusan “Yes” bagi minat mahasiswa. Langkah selanjutnya mencari nilai *Entropy* dan *Gain* berdasarkan kriteria **Fasilitas - Cukup** untuk menghasilkan *Node 4*. Adapun nilai *Entropy* dan *Gain* pada *Node 4* dapat dilihat pada Tabel 6.

Tabel 6. Hasil Perhitungan Nilai *Entropy* dan *Gain* pada *Node 4*

Kategori	Sub Kategori	Jumlah Kasus	Yes	No	Entropy	Gain
Uang Kuliah – Terjangkau & Akreditasi Cukup & Fasilitas - Cukup		5	0	0	0	
Pelayanan						0
	Sangat Baik	0	0	0	0	
	Baik	0	0	0	0	
	Cukup	5	0	5	0	

Kategori	Sub Kategori	Jumlah Kasus	Yes	No	Entropy	Gain
Lokasi						0
	Dekat	0	0	0	0	
	Terjangkau	1	0	1	0	
	Jauh	4	0	1	0	

Berdasarkan perolehan nilai *Entropy* dan *Gain* dari *Node* 4, maka untuk peroleh nilai *Gain* keseluruhan kategori adalah 0, maka untuk kategori **Pelayanan** dan **Lokasi** tidak menjadi faktor dalam penentuan minat mahasiswa dalam memilih kampus. Adapun bentuk *Decession Tree* akhir dapat dilihat pada Gambar 6.



Gambar 6. Hasil *Decession Tree* Pengolahan Data menggunakan Algoritma C4.5

Berdasarkan *Decession Tree* dijelaskan bahwa kriteria **Fasilitas - Cukup** memiliki keputusan “No” bagi minat mahasiswa. Maka *Decession Tree* diatas dijadikan sebagai keputusan akhir dalam prediksi penentuan minta calon mahasiswa dalam memilih perguruan tinggi.

3.4 Evaluasi Hasil *Decession Tree*

Setelah diperoleh hasil *Decession Tree* dari algoritma C4.5, selanjutnya dilakukan evaluasi terhadap hasil *Decession Tree* berdasarkan data Transformasi. Evaluasi bertujuan untuk melihat apakah terdapat kesalahan dari hasil *Decession Tree* yang diperoleh dan apakah perlu dilakukan pengujian kembali.

3.5 Hasil

Hasil dari *Decession Tree* selanjutnya diterjemahkan kedalam bentuk keputusan. Berdasarkan terjemahan keputusan tersebut, maka dapat menjadi alat bantu bagi seorang pengambil keputusan untuk memprediksi faktor kriteria apa yang mendukung minat calon mahasiswa dalam memilih perguruan tinggi yang di inginkan. Sehingga prediksi tersebut, pihak kampus dapat mengambil tindakan-tindakan yang tepat dalam menangani permasalahan ini.

4. Hasil dan Pembahasan

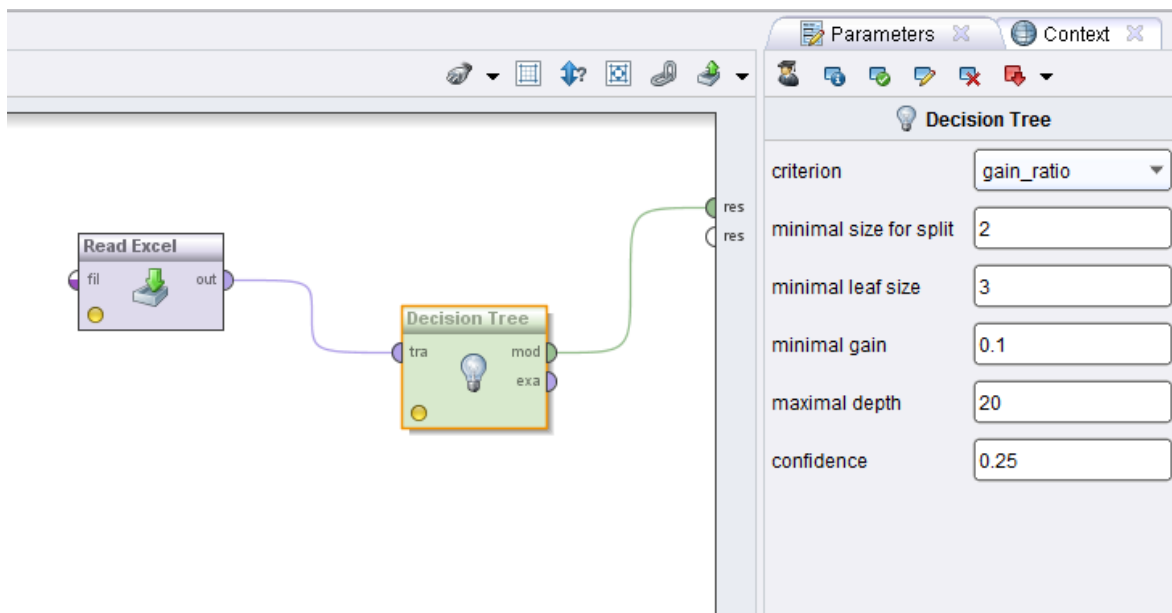
Berdasarkan proses algoritma yang telah dijelaskan diatas maka selanjutnya dilakukan pengujian dengan menggunakan *software Data Mining*, yaitu *Rapid Miner*. Untuk pengujian dengan menggunakan aplikasi *Rapid Miner* ini digunakan data *testing* sebanyak 120 data *testing*. Pengujian dilakukan dengan mengimport data pengujian dalam format excel ke *software Rapid Miner* seperti pada Gambar 7.

This wizard guides you to import your data.
Step 2: An Excel file can contain multiple sheets. Please select the one you want to import into RapidMiner. Furthermore, you can mark a range of cells to be loaded.

	A	B	C	D	E	F	G
(a)	Akreditasi	Uang Kuliah	Fasilitas	Pelayanan	Lokasi	Minat	
MHS 1	Baik	Terjangkau	Sangat Baik	Baik	Terjangkau	Yes	
MHS 2	Cukup	Murah	Cukup	Cukup	Terjangkau	Yes	
MHS 3	Cukup	Terjangkau	Cukup	Cukup	Tidak Terjan	No	
MHS 4	Sangat Baik	Murah	Sangat Baik	Sangat Baik	Terjangkau	Yes	
MHS 5	Baik	Murah	Memadai	Baik	Tidak Terjan	Yes	
MHS 6	Sangat Baik	Terjangkau	Memadai	Baik	Terjangkau	Yes	
MHS 7	Cukup	Terjangkau	Cukup	Cukup	Tidak Terjan	No	
MHS 8	Baik	Terjangkau	Cukup	Baik	Dekat	Yes	
MHS 9	Cukup	Mahal	Cukup	Cukup	Terjangkau	No	
MHS 10	Cukup	Terjangkau	Cukup	Cukup	Tidak Terjan	No	
MHS 11	Baik	Murah	Sangat Baik	Sangat Baik	Dekat	Yes	
MHS 12	Cukup	Terjangkau	Sangat Baik	Cukup	Tidak Terjan	Yes	
MHS 13	Cukup	Terjangkau	Memadai	Baik	Terjangkau	Yes	
MHS 14	Cukup	Mahal	Memadai	Baik	Dekat	No	
MHS 15	Baik	Terjangkau	Memadai	Cukup	Terjangkau	Yes	
MHS 16	Baik	Mahal	Cukup	Baik	Terjangkau	No	
MHS 17	Baik	Terjangkau	Sangat Baik	Baik	Dekat	Yes	
MHS 18	Cukup	Murah	Cukup	Cukup	Terjangkau	Yes	
MHS 19	Cukup	Terjangkau	Cukup	Cukup	Tidak Terjan	No	
MHS 20	Cukup	Mahal	Cukup	Cukup	Terjangkau	No	
MHS 21	Sangat Baik	Murah	Sangat Baik	Baik	Terjangkau	Yes	
MHS 22	Baik	Terjangkau	Cukup	Sangat Baik	Tidak Terjan	Yes	
MHS 23	Baik	Terjangkau	Cukup	Cukup	Dekat	Yes	
MHS 24	Sangat Baik	Murah	Sangat Baik	Sangat Baik	Terjangkau	Yes	
MHS 25	Cukup	Terjangkau	Cukup	Cukup	Terjangkau	No	
MHS 26	Baik	Terjangkau	Sangat Baik	Baik	Terjangkau	Yes	
MHS 27	Cukup	Murah	Cukup	Cukup	Terjangkau	Yes	

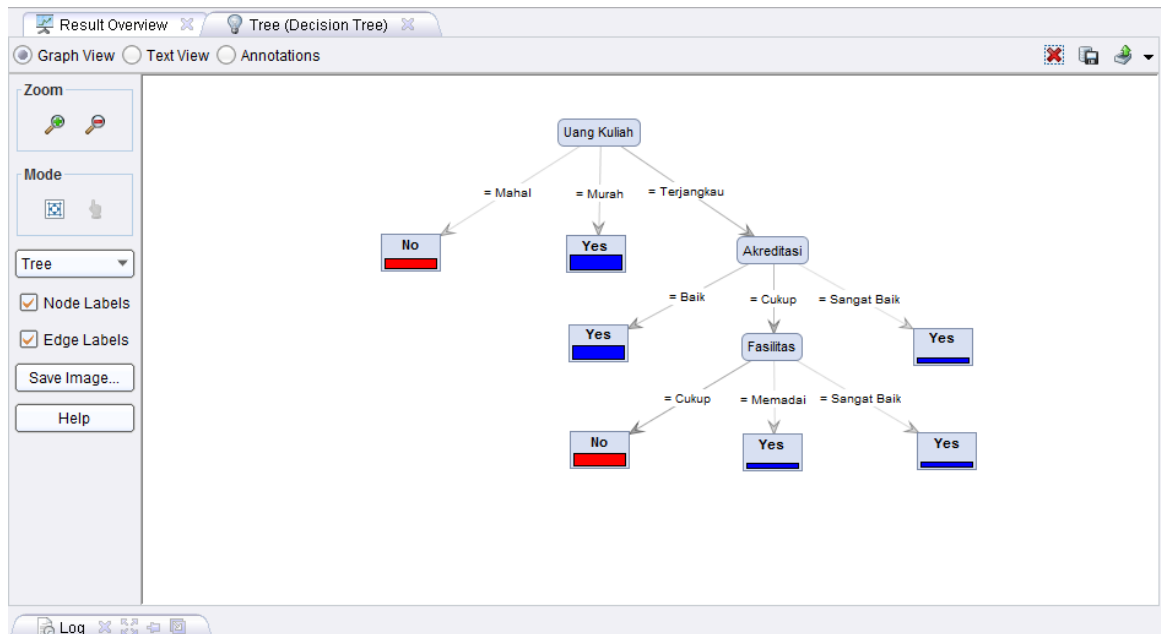
Gambar 7.Import Data Testing ke Software RapidMiner

Selanjutnya berdasarkan data yang telah di import, dilakukan *drag* dan *drop* tabel *learning dataset* kedalam *process view* dan mengatur operator *decession tree* yang dapat dilihat pada Gambar 8:



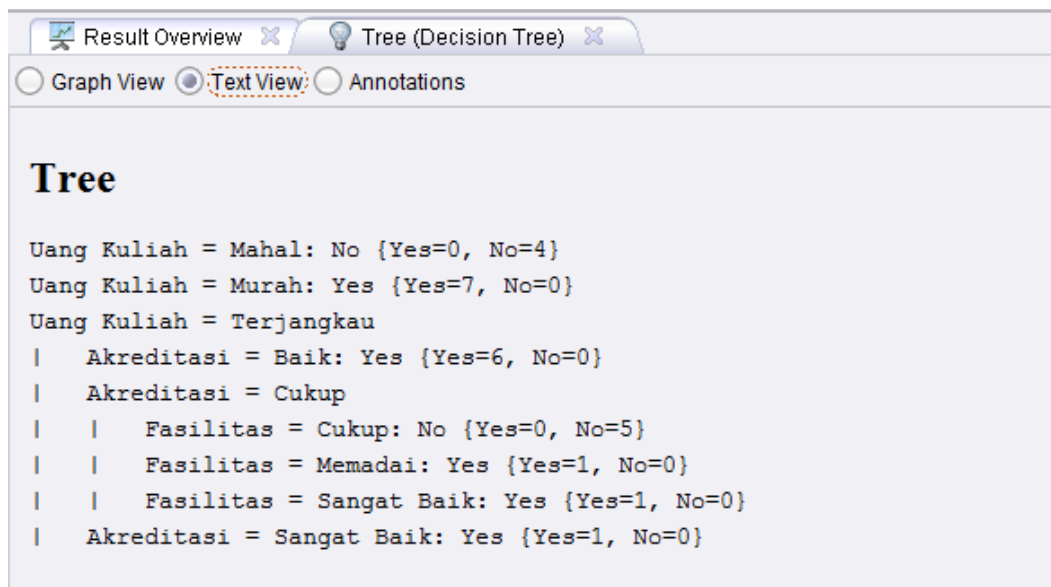
Gambar 8.Process View dalam Software RapidMiner

Berdasarkan *process view* yang telah dibuat, maka apabila dilakukan *running* akan menghasilkan hasil *decession tree* dari data *testing* yang dapat dilihat pada Gambar 8:



Gambar 7. *Decession Tree* Hasil Pengujian Dengan Menggunakan *Software RapidMiner*

Sedangkan bentuk keputusan yang dihasilkan melalui *software Rapid Miner* dapat dilihat pada Gambar 8:



Gambar 8. Hasil Keputusan Pengujian Berdasarkan *Software RapidMiner*

Berdasarkan pengujian yang telah dilakukan, maka dapat disimpulkan faktor utama yang mempengaruhi minat calon mahasiswa baru dalam memilih sebuah perguruan tinggi dengan menggunakan algoritma C4.5 sebagai berikut :

- 1) Jika **Uang Kuliah** biayanya “Murah”, maka minat calon mahasiswa “YES” untuk memilih perguruan tinggi dan jika **Uang Kuliah** “Mahal”, maka minat calon mahasiswa “NO” untuk memilih perguruan tinggi.
- 2) Jika **Uang Kuliah** biayanya “Terjangkau” dan **Akreditasi** kampus “Sangat Baik” atau “Baik”, maka minat calon mahasiswa “YES” untuk memilih perguruan tinggi.

- 3) Jika **Uang Kuliah** biayanya “Terjangkau” dan **Akreditasi** kampus “Cukup” dan **Fasilitas** Kampus “Sangat Baik” Atau “Memadai”, maka minat calon mahasiswa “YES” untuk memilih perguruan tinggi, namun jika **Uang Kuliah** biayanya “Terjangkau” dan **Akreditasi** kampus “Cukup” dan **Fasilitas** kampus “Cukup”, maka minat calon mahasiswa “NO” untuk memilih perguruan tinggi.

5. Kesimpulan

Berdasarkan hasil penelitian yang diperoleh, disimpulkan bahwa proses *datamining* dengan menggunakan Algoritma C4.5 telah dapat menghasilkan sebuah keputusan untuk memprediksi minat calon mahasiswa baru dalam memilih perguruan tinggi berdasarkan faktor kriterianya. Berdasarkan hasil pengolahan data menggunakan Algoritma C4.5 dengan 25 data sampel, kriteria yang menjadi faktor utama terhadap minat calon mahasiswa baru dalam memilih perguruan tinggi adalah kriteria Uang Kuliah, Akreditasi dan Fasilitas, dimana Uang Kuliah Murah, Akreditasi Sangat Baik atau Baik dan Fasilitas Sangat Baik atau Memadai menarik minat calon mahasiswa baru dalam memilih perguruan tinggi. Sedangkan Uang Kuliah Mahal, Akreditasi Cukup dan Fasilitas Cukup tidak menarik minat calon mahasiswa baru dalam memilih perguruan tinggi. Berdasarkan pengujian dengan menggunakan software *Rapid Miner* dengan 120 data testing, diperoleh hasil *Decession Tree* dan keputusan antara pengujian manual adalah sama, dengan tingkat akurasi sebesar 100%. Sehingga penggunaan algoritma C4.5 telah menghasilkan faktor penilaian yang dapat memprediksi minat calon mahasiswa, serta pihak perguruan tinggi dapat memprediksi minat calon mahasiswa baru untuk kedepannya.

Daftar Pustaka

- [1] M. Ary, “Menentukan Prioritas Pilihan Mahasiswa Dalam Memilih Kampus Menggunakan Analisis CONJOINT,” *PARADIGMA*, vol. 18, no. 1, pp. 48–55, 2016.
- [2] M. Fakhri, A. Gilang, and N. Ratnayu, “Analisis Faktor Pembentukan Keputusan Pemilihan Perguruan Tinggi Swasta Universitas Telkom,” *Jurnal Ekonomi, Bisnis & Entrepreneurship*, vol. 11, no. 1, pp. 1–12, 2017.
- [3] Z. Masriah, M.N.Malay, and A. Fitriani, “Persepsi Mahasiswa Terhadap Jurusan Perguruan Tinggi Dan Konsep Diri Dengan Kesesuaian Minat Memilih,” *ANFUSINA: JOURNAL OF PSYCHOLOGY*, vol. 1, no. 1, pp. 61–76, 2018.
- [4] S. Haryati, A. Sudarsono, and E. Suryanan, “Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5,” *Jurnal Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
- [5] C. Nas, “Data Mining Pengelompokan Bidang Keahlian Mahasiswa Menggunakan Algoritma K-Means (Studi Kasus : Universitas CIC Cirebon),” *Syntax: Jurnal Informatika*, vol. 09, no. 2, pp. 1–14, 2020.
- [6] Rismayanti, “Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa Di STT Harapan Medan,” *Jurnal Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
- [7] Rismayanti, “Decission Tree Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika,” *QUERY: Jurnal Sistem Informasi*, vol. 02, no. 1, pp. 16–24, 2018.
- [8] R.L. Dewanti and T.B. Abad, “Prediksi Minat Mahasiswa Untuk Berwirausaha Menggunakan Pendekatan Theory Of Planned Behavior,” *Jurnal Telaah Bisnis*, vol. 15, no. 2, pp. 139–150, 2014.
- [9] Khoirunnisa, L. Susanti, I.T. Rokhmah and L. Stianingsih, “Prediksi Siswa SMK Al-Hidayah Yang Masuk Perguruan Tinggi Dengan Metode Klasifikasi,” *Jurnal Informatika*, vol. 8, no. 1, pp. 26–33, 2021.

- [10] N. Azwanti, "Algoritma C4.5 Untuk Memprediksi Mahasiswa Yang Mengulang Mata Kuliah (Studi Kasus AMIK Labuhan Batu)," *Jurnal SIMETRIS*, vol. 09, no. 1, pp. 11–22, 2018.
- [11] Y.S. Luvia, D. Hartama, D. Windarto, and Solikhun, "Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di AMIK Tunas Bangsa," *JURASIK*, vol. 01, no. 1, pp. 75–79, 2016.
- [12] K. Ahmed, and T. Jesmin, "Comparative Analysis of Data Mining Classification Algorithms in Type-2 Diabetes Prediction Data Using WEKA Approach," *IJSE*, vol. 07, no. 02, pp. 155–160, 2014.
- [13] H. Sulastri, and A.I. Gufroni, "Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia," *Jurnal Nasional Teknologi Dan Sistem Informasi*, vol. 03, no. 02, pp. 299–305, 2017.
- [14] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naïve Bayes," *ILKOM Jurnal Ilmiah*, vol. 10, no. 02, pp. 160–165, 2018.
- [15] E. Elisa, "Analisa Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT. Arupadhatu Adisesanti," *JOIN*, vol. 02, no. 01, pp. 36–41, 2017.
- [16] D.W.T. Putra, "Algoritma C4.5 Untuk menentukan Tingkat Kelayakan Motor Bekas Yang Akan Dijual," *Jurnal TEKNOIF*, vol. 04, no. 01, pp. 16–22, 2016.
- [17] P. Kasih, "Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara," *Innovation in Research of Informatics (INNOVATICS)*, vol. 01, no. 02, pp. 63–69, 2019.
- [18] Asroni, B.M. Respati, and S. Riyadi, "Penerapan Algoritma C4.5 Untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta," *SEMESTA TEKNIKA*, vol. 21, no. 2, pp. 158–165, 2018.
- [19] W.D. Septiani, "Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi penyakit Hepatitis," *Jurnal Pilar Nusa Mandiri*, vol. 13, no. 1, pp. 76–84, 2017.