

Penerapan *Data Mining* untuk Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Algoritma C4.5: Studi Kasus STMIK Primakara

**Dewa Made Aryadi Mertha Sanjaya, A. A Istri Ita Paramitha,
Nengah Widya Utami**

STMIK Primakara, Indonesia

Email corresponding author: dewasanjaya73@gmail.com

ABSTRAK

Mahasiswa non-aktif merupakan mahasiswa yang tidak melakukan registrasi awal atau tidak melakukan perkuliahan selama satu semester. Adanya mahasiswa non-aktif di suatu perguruan tinggi akan berdampak menurunnya nilai perguruan tinggi tersebut. Penelitian ini bertujuan untuk memprediksi mahasiswa yang berpotensi non-aktif di kampus STMIK Primakara. Metode yang digunakan dalam penelitian ini adalah *Knowledge Discovery from Data* (KDD) dengan lima tahapan yaitu: *selection*, *preprocessing*, *transformation*, *data mining*, dan *interpretation/evaluation*. Metode pengumpulan data dilakukan dengan wawancara dan permintaan data riwayat perkuliahan mahasiswa kepada kampus STMIK Primakara. Hasil dari penelitian menunjukkan bahwa mahasiswa dengan nilai IPK 0 - 1.99 atau 2.00 - 2.75, Jenis Tinggal Kontrakan atau Kos dan Jarak ke kampus cukup dekat, jauh dan sangat jauh diprediksi mahasiswa yang berpotensi non-aktif. Adapun atribut yang paling mempengaruhi mahasiswa berpotensi non-aktif di kampus STMIK Primakara adalah IPK sedangkan yang tidak terlalu mempengaruhi adalah status pembayaran SPP.

Kata kunci: *data mining*, *decision tree*, mahasiswa non-aktif

ABSTRACT

Non-active students are students who do not register early or do not lecture for one semester. The existence of non-active students in a university will have an impact on decreasing the value of the college. This study aims to predict students who may be active on the STMIK Primakara campus. The method used in this research is Knowledge Discovery from Data (KDD) with five stages, namely: selection, preprocessing, transformation, data mining, and interpretation/evaluation. The method of data collection was carried out by interviewing and requesting data for lecture activities on the STMIK Primakara campus. The results showed that students with GPA values of 0 - 1.99 or 2.00 - 2.75, Type of Rented Living or Boarding House and Distance to campus were quite close, far and highly predictable for students

who might not be active. The attribute that most affects students who may not be active on the STMIK Primakara campus is the GPA, while the one that does not really affect is the status of tuition payments.

Keywords: *data mining, decision tree, Non-active student*

PENDAHULUAN

Sebagai bangsa yang besar yang memiliki impian menjadi bangsa yang cerdas dan sejahtera, hal terpenting yang harus dimiliki rakyat Indonesia adalah pendidikan [1]. Tidak sedikit juga yang menempuh pendidikan ke tingkatan lebih tinggi seperti melanjutkan ke sekolah atau perguruan tinggi. Sekolah atau perguruan tinggi memiliki tingkatan tertinggi yang mempunyai sistem pembelajaran yang berbeda dengan jenjang sebelumnya yaitu kebebasan dan keinginan belajar. Motivasi dan keinginan belajar untuk menjadi lebih baik sangat mempengaruhi keberhasilan mahasiswa di tingkat perguruan tinggi. Tingkat keberhasilan dan keaktifan mahasiswa tersebut mempengaruhi predikat kesuksesan dan keunggulan suatu perguruan tinggi. Begitupun sebaliknya, mahasiswa yang tidak aktif dan kualitas mahasiswa yang rendah menjadi faktor ketidakmampuan perguruan tinggi dalam menyelenggarakan proses pendidikan tinggi.

Status mahasiswa yang *drop out* dan non-aktif merupakan hal yang berkaitan dengan gagalnya mahasiswa dalam menempuh perkuliahan [2]. Mahasiswa yang tidak melakukan pendaftaran akademik pada awal semester ataupun mahasiswa yang tidak mengikuti perkuliahan selama satu semester disebut sebagai mahasiswa non-aktif.

Kampus STMIK Primakara merupakan kampus swasta yang berdiri pada tahun 2012 dibawah naungan Yayasan Primakara. Perguruan tinggi yang mencetak sumber daya profesional merupakan komitmen dari Kampus STMIK Primakara itu sendiri. STMIK Primakara tergolong kampus yang muda tetapi telah meluluskan banyak sekali mahasiswa tetapi disamping capaian yang baik tersebut masih ada mahasiswa STMIK Primakara yang gagal studi maupun lulus tidak tepat waktu dalam studinya yang dikarenakan mahasiswa yang non-aktif di semester tertentu. Dari data terbaru yang disampaikan oleh bidang BAAK kampus STMIK Primakara persentase mahasiswa non-aktif dari angkatan 2017 hingga angkatan 2020 di STMIK Primakara mencapai 10,64%.

Data-data mengenai mahasiswa tersebut telah tercatat di *database* dari kampus STMIK Primakara yang dimana *database* tersebut dapat dijadikan sumber pengetahuan baru yang digunakan untuk penambangan data atau *data mining*. Data mining merupakan langkah analisis terhadap proses penemuan pengetahuan

di dalam basis data atau knowledge discovery in databases yang disingkat KDD. Pengetahuan yang dimaksud dapat berupa pola data atau relasi antar data yang valid (yang tidak diketahui sebelumnya) [3]. Penggunaan *data mining* dalam mengolah data di kampus STMIK Primakara adalah untuk membantu dalam meningkatkan aktivitas pendidikan yang ada di kampus STMIK Primakara yang khususnya meminimalisir faktor yang dapat mempengaruhi kegagalan studi mahasiswa di STMIK Primakara. Dimana, dalam *data mining* terdapat algoritma *decision tree* yang termasuk dalam teknik klasifikasi dalam pengolahan datanya.

Teknik klasifikasi merupakan proses analisa sebuah data yang digunakan untuk membentuk suatu model yang menggambarkan kelas yang terdapat pada data tersebut [4]. Kelas yang terdapat pada metode pohon keputusan tersebut berbentuk suatu pohon. Metode *decision tree* biasa dipakai untuk mempelajari prediksi pada pola suatu data dan juga suatu klasifikasi dan menggambarkan suatu variabel atribut dan variabel target yang berbentuk sebuah pohon [5]. Kelebihan *Decision tree* yaitu daerah pengambilan keputusan yang rumit dan luas dapat diubah menjadi semakin spesifik dan juga sederhana, menghapus perhitungan yang tidak perlu dan *sample* yang diuji menggunakan kelas dan kriteria tertentu, dan algoritma *decision tree* ini fleksibel dalam memilih *features* dari *internal nodes* yang berbeda.

Algoritma C4.5 merupakan salah satu metode *decision tree* yang terkenal. C4.5 merupakan turunan dari ID3 yang telah dikembangkan dan mengalami peningkatan seperti pada penanganan suatu atribut yang berbentuk numerik, *missing value* beserta *noise* pada sebuah *dataset*, dan *rule-rule* yang didapatkan dari terbentuknya sebuah model yang berbentuk pohon [6]. Untuk pengujian algoritma C4.5 memiliki keunggulan dalam mengolah data dengan nilai nominal, orthogonal, dan kontinyu

Ada beberapa penelitian sebelumnya yang menggunakan metode *data mining* dalam memprediksi mahasiswa yang berpotensi non-aktif antara lain penelitian dari Irma Afliani dengan menggunakan pohon keputusan C4.5 dalam prediksi mahasiswa aktif dan non-aktif yang memperoleh *rules* yang dapat diajukan untuk menjadi acuan dalam pimpinan universitas dalam meningkatkan standar akademik Universitas Satya Negara Indonesia [7]. Lalu penelitian dari Nur Yanti Lumban Gaol pada tahun 2020 yang memprediksi mahasiswa yang berpotensi menjadi non-aktif yang menggunakan *decision tree* C4.5 [8]. Hasil yang didapatkan dari penelitian tersebut berupa *rules* yang memberikan suatu kriteria untuk menganalisa mahasiswa yang berpotensi menjadi non-aktif dimana kriteria-kriteria yang terbentuk tersebut adalah nilai gagal, nilai absensi, jadwal kuliah, dan pembayaran uang kuliah [9].

Karena adanya masalah tersebut penulis tertarik untuk melakukan penelitian Penerapan *Data Mining* untuk Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Algoritma C4.5 di Kampus STMIK Primakara. Penelitian ini bertujuan untuk mengetahui bagaimana hasil penerapan *data mining* dalam memprediksi mahasiswa yang berpotensi non-aktif.

METODE PENELITIAN

Tahapan - tahapan pada penelitian ini digunakan untuk memberi tujuan atau arah pada setiap tahap yang akan dilakukan yang meliputi:

1. Persiapan

Pada tahapan ini dilakukan identifikasi terhadap masalah yang ditemukan, mengutarakan tujuan masalah dan juga tinjauan pustaka.

2. Pengumpulan data

Pengumpulan data didasari oleh sumber data yang dimiliki dimana sumber data tersebut antara lain:

- 1) Studi Lapangan

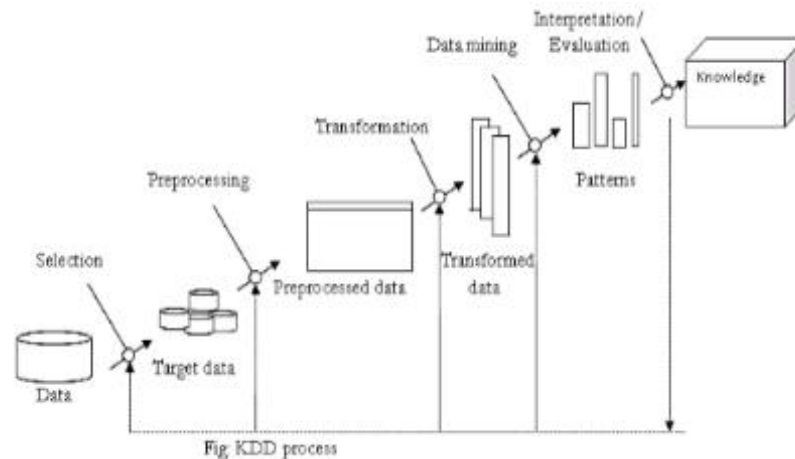
Studi lapangan dalam pengumpulan data ini dilakukan dengan cara wawancara terbuka. Wawancara yang dilakukan untuk mengetahui proses bisnis dari Kampus STMIK Primakara khususnya pada bidang akademik, dan data mahasiswa STMIK Primakara sebagai data yang diperlukan dalam penelitian ini dimana telah dibahas dalam instrumen penelitian.

- 2) Studi Pustaka

Studi pustaka yang dilaksanakan adalah mencari dan mempelajari jurnal, penelitian terdahulu ataupun buku yang berhubungan dengan topik yang diteliti. Studi pustaka ini dilaksanakan supaya penelitian ini memiliki dasar pengetahuan yang sesuai dengan tujuan penelitian yang akan dilakukan.

3. Implementasi *Knowledge Discovery from Data* (KDD)

Pengolahan data dalam penelitian ini didasari oleh proses *Knowledge Discovery from Data* (KDD). Berikut ini merupakan uraian proses di dalam KDD:



Gambar 1. Tahapan *Knowledge Discovery from Data*

1) *Selection*

Dalam tahapan ini data melalui proses pemilahan dimana data yang dipakai adalah data riwayat perkuliahan mahasiswa dari angkatan 2017 hingga 2020 dengan total mahasiswa hingga 653 mahasiswa. Dengan adanya proses ini maka proses pengolahan berlangsung lebih baik dan efektif sesuai dengan tujuan penelitian.

2) *Preprocessing*

Tahapan pembersihan data yang kosong, kurang maupun error. Data yang dibersihkan antara lain mengeliminasi data mahasiswa yang error dan kosong atau tidak lengkap dan mengurangi variabel yang tidak dipakai pada data mahasiswa yang diperoleh dari STMIK Primakara. Pada tahapan ini sebanyak 653 data mahasiswa yang disaring dimana mendapatkan 620 data mahasiswa yang dapat digunakan.

3) *Transformation*

Sebelum melakukan pengolahan *data mining*, data harus ditransformasi terlebih dahulu. Dimana data mahasiswa ditransformasi menjadi data yang sesuai dengan kriteria yang digunakan dalam perhitungan agar dapat diproses dalam *software orange*.

4) *Data Mining*

Proses ini merupakan proses penambangan data yang menggunakan metode klasifikasi algoritma C4.5. Langkah langkah dari C4.5 itu sendiri diawali dalam pemilihan atribut sebagai *root* selanjutnya pembuatan cabang pada setiap nilai setelah itu melakukan pembagian kasus dalam cabang dan dari proses tersebut diulang hingga seluruh kasus pada cabang memiliki kelas yang sama. Hasil yang didapatkan berupa *rule*.

5) *Interpretation/Evaluation*

Interpretation/evaluation merupakan proses pembahasan hasil yang didapat dari data mining menjadi bentuk yang mudah dimengerti dimana tahapan ini termasuk pemeriksaan pola informasi.

HASIL DAN PEMBAHASAN

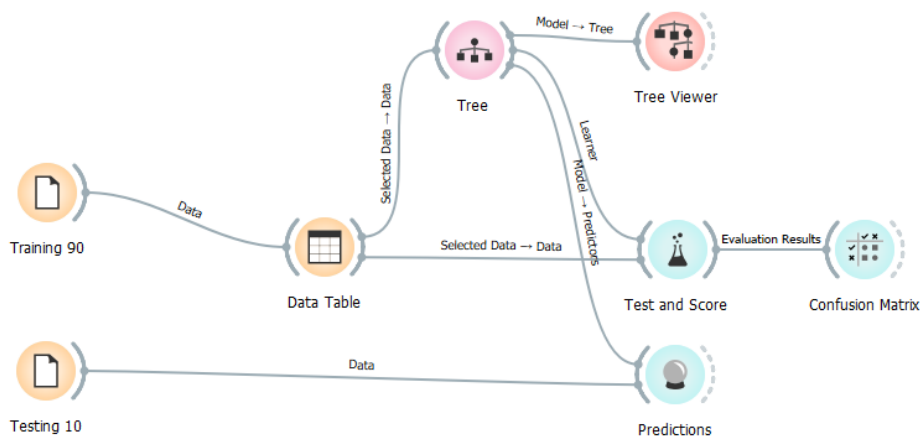
Pada tahap ini dilakukan pengujian metode yang digunakan yaitu menghitung dan mendapatkan *rules* yang ada pada algoritma yang digunakan yaitu C4.5. Pada tahapan pengumpulan data didapatkan *raw* data sebanyak 653 data mahasiswa yang disaring pada tahapan *data cleaning* menjadi sebanyak 620 data. 620 data mahasiswa tersebut, variabel pada setiap atribut data yang digunakan di transformasi menjadi kategori-kategori yang digunakan dalam proses *data mining*. Dibawah ini merupakan tabel transformasi variabel pada setiap atribut yang digunakan:

Tabel 1. Tabel Transformasi *Variable*

No	Atribut	Variabel Awal	Transformasi
1	Jenis Kelamin	Laki-laki	L
		Perempuan	P
2	Prodi	Sistem Informasi	SI
		Teknik Informatika	IF
		Sistem Informasi Akuntansi	SIA
3	Status Mahasiswa Baru	Mahasiswa Baru	Mahasiswa Baru
		Pindahan (Transfer)	Pindahan (Transfer)
		Transfer Alfaprima	Transfer Alfaprima
4	Umur	17-19 Tahun	<20 Tahun
		20 - 25 Tahun	20 - 25 Tahun
		26 - 39 Tahun	>25 Tahun
5	Jarak	0-5 KM	Dekat
		6-10 KM	Cukup Dekat
		11-15 KM	Cukup Jauh
		16-20 KM	Jauh
6	Jenis Tinggal	>20 KM	Sangat Jauh
		Rumah	Bersama Keluarga
		Kontrakan/Kos	Kontrakan/Kos
		0 - Rp 500.000	Sangat Rendah
7	Penghasilan	Rp 500.000 - Rp 999.999	Rendah
		Rp 1.000.000 - Rp 1.999.999	Cukup Rendah
		Rp 2.000.000 - Rp 4.999.999	Standar
		Rp 5.000.000 - Rp 20.000.000	Tinggi
		> Rp 20.000.000	Sangat Tinggi
8	IPK	0 - 1.99	Kurang Memuaskan
		2.00 - 2.75	Memuaskan
		2.76 - 3.50	Sangat Memuaskan
		3.51 - 4.00	Dengan Pujian

No	Atribut	Variabel Awal	Transformasi
9	Status Pembayaran SPP	Menunggak Tidak Menunggak	Menunggak Tidak Menunggak

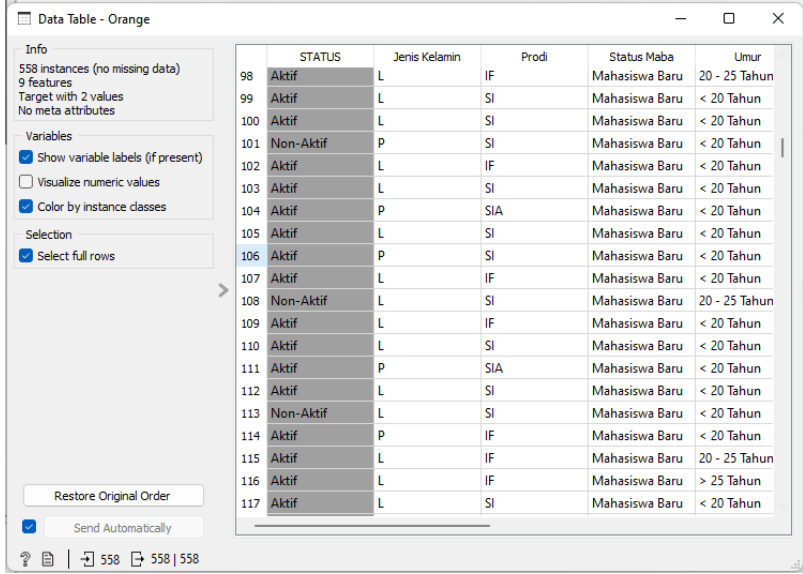
Setelah melalui proses transformasi selanjutnya adalah melakukan proses mining pada dataset yang tersedia. Tahapan *data mining* menggunakan algoritma yang telah ditentukan diawal penelitian dimana menggunakan algoritma *decision tree* untuk mencari prediksi mahasiswa yang berpotensi aktif dan non-aktif dalam perkuliahan di kampus STMIK Primakara. Proses *data mining* ini menggunakan *software* aplikasi *Orange*. Pada gambar dibawah merupakan alur dari proses *data mining* pada *software orange*.



Gambar 2. Tampilan Alur *Widget* pada Aplikasi *Orange*

Berdasarkan alur proses *data mining* menggunakan *software orange* dimana penjelasan dari setiap alurnya dapat dilihat dibawah ini:

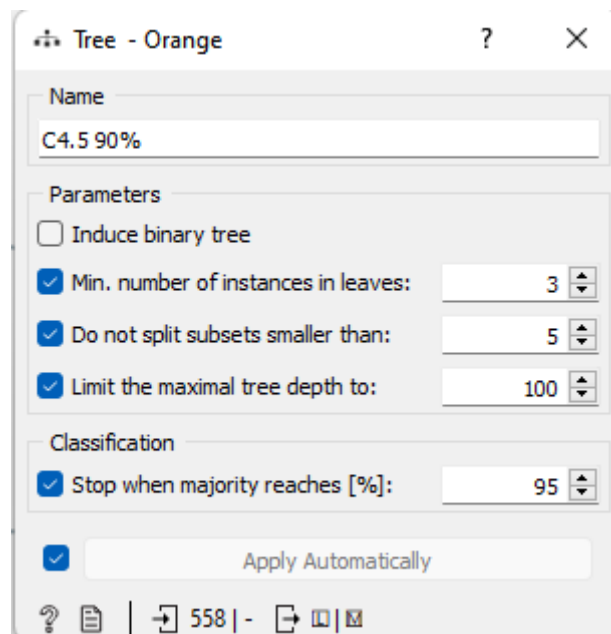
- 1) Proses *data mining* pada penelitian ini yang pertama adalah menentukan *data training* dan *data testing* pada *dataset* yang digunakan dimana *data training* yang digunakan pada penelitian ini adalah sebesar 90 % dari total *dataset* yaitu sebanyak 558 data mahasiswa dari 620 total *dataset* yang ada dan *data testing* yang digunakan adalah sebesar 10 % dari total *dataset* yaitu sebanyak 62 data mahasiswa dari 620 total *dataset* yang ada. Pada gambar diatas ditunjukkan pada *widget* "Training 90" dan "Testing 10" dan data yang digunakan pada proses *training* dapat dilihat pada *widget* "Data Table". Dibawah ini merupakan tampilan dari *widget* "Data Table".



	STATUS	Jenis Kelamin	Prodi	Status Maba	Umur
98	Aktif	L	IF	Mahasiswa Baru	20 - 25 Tahun
99	Aktif	L	SI	Mahasiswa Baru	< 20 Tahun
100	Aktif	L	SI	Mahasiswa Baru	< 20 Tahun
101	Non-Aktif	P	SI	Mahasiswa Baru	< 20 Tahun
102	Aktif	L	IF	Mahasiswa Baru	< 20 Tahun
103	Aktif	L	SI	Mahasiswa Baru	< 20 Tahun
104	Aktif	P	SIA	Mahasiswa Baru	< 20 Tahun
105	Aktif	L	SI	Mahasiswa Baru	< 20 Tahun
106	Aktif	P	SI	Mahasiswa Baru	< 20 Tahun
107	Aktif	L	IF	Mahasiswa Baru	< 20 Tahun
108	Non-Aktif	L	SI	Mahasiswa Baru	20 - 25 Tahun
109	Aktif	L	IF	Mahasiswa Baru	< 20 Tahun
110	Aktif	L	SI	Mahasiswa Baru	< 20 Tahun
111	Aktif	P	SIA	Mahasiswa Baru	< 20 Tahun
112	Aktif	L	SI	Mahasiswa Baru	< 20 Tahun
113	Non-Aktif	L	SI	Mahasiswa Baru	< 20 Tahun
114	Aktif	P	IF	Mahasiswa Baru	< 20 Tahun
115	Aktif	L	IF	Mahasiswa Baru	20 - 25 Tahun
116	Aktif	L	IF	Mahasiswa Baru	> 25 Tahun
117	Aktif	L	SI	Mahasiswa Baru	< 20 Tahun

Gambar 3. Tampilan Widget Data Table

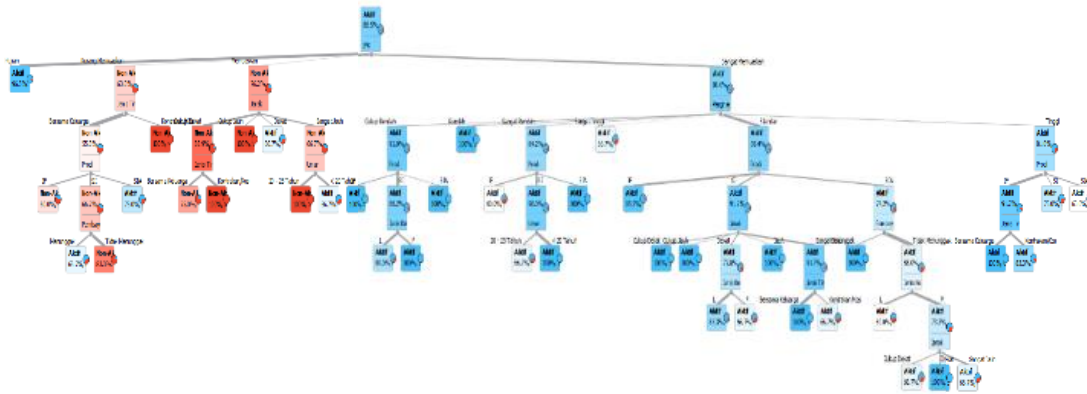
- 2) Selanjutnya adalah memasukkan algoritma yang digunakan pada proses *data mining* dimana menggunakan algoritma *decision tree C4.5* ditandai dengan widget "tree" pada gambar diatas. Tampilan pada widget "tree" dapat dilihat pada gambar dibawah ini:
- 3)



Gambar 4. Tampilan Widget Tree

- 4) Setelah memasukan widget "tree" proses selanjutnya adalah menampilkan tampilan *decision tree* atau pohon keputusan yang pada gambar 2 ditandai

dengan widget "Tree Viewer". Dibawah ini merupakan tampilan dari widget "Tree Viewer".



Gambar 5. Tampilan Widget Tree Viewer

Berdasarkan gambar diatas, interpretasi dari hasil *decision tree* dengan algoritma C4.5 adalah sebagai berikut:

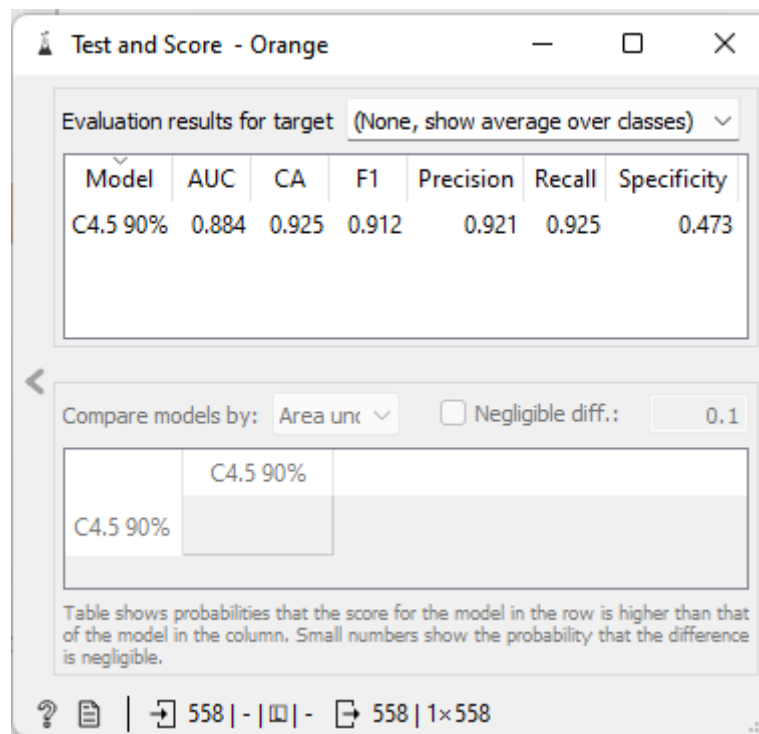
- 1) if IPK = Dengan Pujian, then Status = Aktif
- 2) if IPK = Kurang Memuaskan, and Jenis Tinggal = Kontrakan/Kos, then Status = Non-Aktif
- 3) if IPK = Kurang Memuaskan, Jenis Tinggal = Bersama Keluarga, and Prodi = IF, then Status = Non-Aktif
- 4) if IPK = Kurang Memuaskan, Jenis Tinggal = Bersama Keluarga, and Prodi = SIA, then Status = Aktif
- 5) if IPK = Kurang Memuaskan, Jenis Tinggal = Bersama Keluarga, Prodi = SI and, Pembayaran SPP = Menunggak, then Status = Aktif
- 6) if IPK = Kurang Memuaskan, Jenis Tinggal = Bersama Keluarga, Prodi = SI and, Pembayaran SPP = Tidak Menunggak, then Status = Non-Aktif
- 7) if IPK = Memuaskan and Jarak = Cukup Jauh, then Status = Non-Aktif
- 8) if IPK = Memuaskan and Jarak = Dekat, then Status = Aktif
- 9) if IPK = Memuaskan, Jarak = Cukup Dekat, and Jenis Tinggal = Bersama Keluarga, then Status = Non-Aktif
- 10) if IPK = Memuaskan, Jarak = Cukup Dekat, and Jenis Tinggal = Kontrakan, then Status = Non-Aktif
- 11) if IPK = Memuaskan, Jarak = Sangat Jauh, and Umur = 20 - 25 Tahun, then Status = Non-Aktif
- 12) if IPK = Memuaskan, Jarak = Sangat Jauh, and Umur = < 20 Tahun, then Status = Aktif

- 13) if IPK = Sangat Memuaskan and Penghasilan Ayah = Rendah, then Status = Aktif
- 14) if IPK = Sangat Memuaskan and Penghasilan Ayah = Sangat Tinggi, then Status = Aktif
- 15) if IPK = Sangat Memuaskan, Penghasilan Ayah = Cukup Rendah, and Prodi = IF, then Status = Aktif
- 16) if IPK = Sangat Memuaskan, Penghasilan Ayah = Cukup Rendah, and Prodi = SIA, then Status = Aktif
- 17) if IPK = Sangat Memuaskan, Penghasilan Ayah = Cukup Rendah, Prodi = SI, and Jenis Kelamin = L, then Status = Aktif
- 18) if IPK = Sangat Memuaskan, Penghasilan Ayah = Cukup Rendah, Prodi = SI, and Jenis Kelamin = P, then Status = Aktif
- 19) if IPK = Sangat Memuaskan, Penghasilan Ayah = Sangat Rendah, and Prodi = IF, then Status = Aktif
- 20) if IPK = Sangat Memuaskan, Penghasilan Ayah = Sangat Rendah, and Prodi = SIA, then Status = Aktif
- 21) if IPK = Sangat Memuaskan, Penghasilan Ayah = Sangat Rendah, Prodi = SI, and Umur = 20 - 25 Tahun, then Status = Aktif
- 22) if IPK = Sangat Memuaskan, Penghasilan Ayah = Sangat Rendah, Prodi = SI, and Umur = < 20 Tahun, then Status = Aktif
- 23) if IPK = Sangat Memuaskan, Penghasilan Ayah = Standar, and Prodi = IF, then Status = Aktif
- 24) if IPK = Sangat Memuaskan, Penghasilan Ayah = Standar, Prodi = SI, and Jarak = Cukup Dekat, then Status = Aktif
- 25) if IPK = Sangat Memuaskan, Penghasilan Ayah = Standar, Prodi = SI, and Jarak = Cukup Jauh, then Status = Aktif
- 26) if IPK = Sangat Memuaskan, Penghasilan Ayah = Standar, Prodi = SI, and Jarak = Jauh, then Status = Aktif
- 27) if IPK = Sangat Memuaskan, Penghasilan Ayah = Standar, Prodi = SI, Jarak = Dekat, and Jenis Kelamin = L, then Status = Aktif
- 28) if IPK = Sangat Memuaskan, Penghasilan Ayah = Standar, Prodi = SI, Jarak = Dekat, and Jenis Kelamin = P, then Status = Aktif
- 29) if IPK = Sangat Memuaskan, Penghasilan Ayah = Standar, Prodi = SI, Jarak = Sangat Jauh, and Jenis Tinggal = Bersama Keluarga, then Status = Aktif
- 30) if IPK = Sangat Memuaskan, Penghasilan Ayah = Standar, Prodi = SI, Jarak = Sangat Jauh, and Jenis Tinggal = Kontrakan/Kos, then Status = Aktif
- 31) if IPK = Sangat Memuaskan, Penghasilan Ayah = Standar, Prodi = SIA, and Pembayaran SPP = Menunggak, then Status = Aktif

- 32) if IPK = Sangat Memuaskan, Penghasilan Ayah = Standar, Prodi = SIA, Pembayaran SPP = Tidak Menunggak, and Jenis Kelamin = L, then Status = Aktif
- 33) if IPK = Sangat Memuaskan, Penghasilan Ayah = Standar, Prodi = SIA, Pembayaran SPP = Tidak Menunggak, Jenis Kelamin = P, and Jarak = Cukup Dekat, then Status = Aktif
- 34) if IPK = Sangat Memuaskan, Penghasilan Ayah = Standar, Prodi = SIA, Pembayaran SPP = Tidak Menunggak, Jenis Kelamin = P, and Jarak = Dekat, then Status = Aktif
- 35) if IPK = Sangat Memuaskan, Penghasilan Ayah = Standar, Prodi = SIA, Pembayaran SPP = Tidak Menunggak, Jenis Kelamin = P, and Jarak = Sangat Jauh, then Status = Aktif
- 36) if IPK = Sangat Memuaskan, Penghasilan Ayah = Tinggi and Prodi = SIA, then Status = Aktif
- 37) if IPK = Sangat Memuaskan, Penghasilan Ayah = Tinggi and Prodi = SI, then Status = Aktif
- 38) if IPK = Sangat Memuaskan, Penghasilan Ayah = Tinggi, Prodi = IF, and Jenis Tinggal = Bersama Keluarga, then Status = Aktif
- 39) if IPK = Sangat Memuaskan, Penghasilan Ayah = Tinggi, Prodi = IF, and Jenis Tinggal = Kontrakan/Kos, then Status = Aktif

Berdasarkan gambar hasil dari perhitungan pohon keputusan didapatkan sebanyak 39 *rules* yang mempengaruhi mahasiswa menjadi aktif maupun non-aktif. Hasil dari 9 Atribut yang diuji menunjukkan bahwa atribut yang paling mempengaruhi mahasiswa berpotensi non-aktif adalah IPK, dan diikuti oleh Jenis Tinggal dan juga Jarak dengan pola variabel yang mempengaruhi adalah pada IPK dengan IPK kurang Memuaskan (0 - 1.99) atau Memuaskan (2.00 - 2.75), lalu Jenis Tinggal Kontrakan/Kos, dan Jarak Cukup Dekat (6 - 10 KM), Cukup Jauh (11 - 15 KM), atau Sangat Jauh (>20 KM) menunjukkan mahasiswa berpotensi menjadi mahasiswa non-aktif.

- 5) Selanjutnya adalah melakukan uji akurasi dari algoritma *decision tree* terhadap *dataset* yang digunakan dimana ditandai dengan widget "*Test and Score*". Dibawah ini merupakan tampilan dari widget "*Test and Score*".



Gambar 5. Tampilan Widget Test and Score

Hasil dari proses pengujian akurasi algoritma C4.5 menggunakan widget "Test and Score" pada aplikasi orange didapatkan akurasi sebesar 92.5 %, F1 Score sebesar 91.2%, precision sebesar 92.1 %, recall 92.5 %, specificity 47.3 %, dan memperoleh nilai AUC (Area Under Curve) sebesar 88.4%. Nilai AUC adalah nilai perhitungan performa suatu indeks data klasifikasi semakin besar nilai AUC maka semakin baik performa indeks data yang digunakan. Dengan besar nilai AUC sebesar 0.884 atau 88.4% menunjukkan bahwa performa indeks data klasifikasi yang digunakan tersebut termasuk *Good Classification*.

6) *Confusion Matrix*

Proses selanjutnya adalah tahap evaluasi dengan menggunakan widget "confusion matrix" terhadap data mahasiswa STMIK Primakara. *Confusion matrix* ini menghasilkan *accuracy*, *F1 score*, *precision*, *recall*, *specificity*. Berikut merupakan hasil dari evaluasi menggunakan confusion matrix:

		Predicted		Σ
		Aktif	Non-Aktif	
Actual	Aktif	490	4	494
	Non-Aktif	38	26	64
Σ		528	30	558

Gambar 6. Tampilan Widget Confusion Matrix

Perhitungan *manual* dari nilai yang terdapat pada gambar 6:

$$Accuracy = \frac{490+26}{558} = 0.9247 = 92.5 \%$$

$$Precision = \frac{494}{(494+4)} = 0.991 = 99.1 \%$$

$$Recall = \frac{494}{(494+38)} = 0.928 = 92.8 \%$$

$$F_1 \text{ Score} = \frac{2(0.991 \times 0.928)}{0.991+0.928} = 0.958 = 95.8\%$$

$$Specificity = \frac{26}{(26+4)} = 0.866 = 86.6\%$$

Berdasarkan gambar diatas, dari 558 data mahasiswa yang diklasifikasikan dengan algoritma *decision tree C4.5*. Terdapat 494 data mahasiswa yang dikategorikan mahasiswa aktif dan terdapat 64 data mahasiswa yang dikategorikan mahasiswa non-aktif. 490 data mahasiswa sudah memiliki klasifikasi yang sesuai dalam kategori mahasiswa aktif sedangkan 4 data mahasiswa kategori aktif terprediksi non-aktif. Terdapat 38 data mahasiswa yang berkategori mahasiswa non-aktif terprediksi aktif sedangkan 26 data mahasiswa non-aktif yang telah sesuai dengan klasifikasi.

KESIMPULAN

Bersumber pada hasil penerapan *data mining* pada *dataset* mahasiswa STMIK Primakra untuk menghitung prediksi terhadap mahasiswa yang berpotensi status non-aktif memakai *algoritma C4.5*. Maka dapat disimpulkan bahwa Algoritma C4.5 dapat dipakai untuk proses pengklasifikasian data mahasiswa STMIK Primakara yang dimana split data training dan data testing yang digunakan adalah 90:10 yaitu sebanyak 558 *data training* dan 62 *data testing* dengan menggunakan aplikasi *orange*. Algoritma C4.5 mendapatkan hasil skor akurasi (CA) sebesar 92.5%, *F1 Score* sebesar 91.2%, *precision* sebesar 92.1%, *recall* sebesar 92.5%, *specificity* sebesar 47.3%, dan *AUC (Area Under Curve)* sebesar 88.4% atau dapat dikategorikan *good classification*. Hasil evaluasi dengan memakai *confusion matrix* pada C4.5 memperoleh skor akurasi (CA) sebesar 92.5%, *F1 Score* sebesar 95.8%, *Precision* sebesar 99.1%. *Recall* sebesar 92.8%, *specificity* sebesar 86.6%. Berdasarkan hasil pohon keputusan terdapat 39 rules yang terbentuk bahwa 39 *rules* tersebut menunjukkan bahwa atribut yang paling mempengaruhi mahasiswa berpotensi non-aktif adalah mahasiswa dengan nilai IPK 0 - 1.99 atau 2.00 - 2.75, jenis tinggal mahasiswa adalah kontrakan/kos dan Jarak dengan variabel cukup dekat (6 - 10 KM), cukup jauh (11 - 15 KM), atau sangat jauh (> 20 KM) sedangkan atribut yang tidak terlalu mempengaruhi mahasiswa berpotensi non-aktif adalah atribut status pembayaran SPP.

DAFTAR PUSTAKA

- Paramita, "Pengaruh Learning Cycle 5E Terhadap Hasil Belajar IPA Kelas V SD Pupuan," 2016, [Online]. Available: <http://ejournal.undiksha.ac.id/index.php/JJPGSD/article/view/6950/4740>.
- Dwi Untari, K. Hastuti, E. Y. Hidayat, Dwi Untari, N. Limão, and N. Y. L. Gaol, "Data Mining untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree C4.5," *Fak. Ilmu Komput. Univ. Dian Nuswantoro*, vol. 2013, no. November, pp. 31-48, 2010, doi: 10.37034/jidt.v2i1.22.
- A. A. I. I. P. Nengah Widya Utami, "Penerapan Data Mining Untuk Mengetahui Pola Pemilihan Program Studi Di Stmik Primakara Menggunakan Algoritma K-Means ...," *J. Teknol. Inf. dan ...*, vol. 3, pp. 456-463, 2021, [Online]. Available: <http://jurnal.undhirabali.ac.id/index.php/jutik/article/view/1540>.
- J. Pei and K. J. , Han Micheline, *Data Mining: Concepts and Techniques 3rd Edition*. 2011.
- N. Ye, *Data Mining Theories, Algorithms, and Examples*, vol. 16, no. 4. Taylor & Francis Group, LLC, 2014.
- T. Larasati, D. A. H. D., & Sutrisno, "Tourism Site Recommendation in Jakarta Using Decision Tree Method Based on Web Review," pp. 195-209, 2018.
- I. Afliani, "Implementasi Data Mining Untuk Prediksi Mahasiswa Aktif dan Non Aktif Dengan Algoritma (Decision Tree) C4.5 Studi Kasus : Universitas Satya Negara Indonesia," pp. 31-48, 2016.
- N. Y. L. Gaol, "Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam Decision Tree dan Algoritma C4.5," *J. Inf. Teknol.*, vol. 2, pp. 23-29, 2020, doi: 10.37034/jidt.v2i1.22.
- N. K. M. Astuti, N. W. Utami, and I. G. P. K. Juliharta, "Classification of Blood Donor Data Using C4.5 and K-Nearest Neighbor Methods (Case Study: Utd Pmi Bali Province)," *J. Pilar Nusa Mandiri*, vol. 18, no. 1, pp. 9-16, 2022, doi: 10.33480/pilar.v18i1.2790.