

Modifikasi *Gain Ratio* Pada Algoritma C4.5 dengan Nilai Koefisien Determinasi untuk Prediksi Kelulusan Mahasiswa (Studi Kasus : Universitas Islam Madura)

Muhsi¹, Suprpto²

¹Program Studi Sistem Informasi Universitas Islam Madura, Jln. Komplek PP. Miftahul Ulum Bettet, Kab. Pamekasan, 69317, Indonesia

²Pendidikan Teknik Elektronika dan Informatika Universitas Negeri Yogyakarta, Jln. Colombo Yogyakarta No.1, Karang Malang, Kab. Sleman, Daerah Istimewa Yogyakarta 5528, Indonesia

Info Artikel

Riwayat Artikel:

Received 2023-10-09

Revised 2025-02-15

Accepted 2025-01-31

Abstract – C4.5 is a decision tree algorithm that can be used for making predictions. The stages start from forming a decision tree through splitting attributes, pruning and extracting rules or knowledge to then be used for prediction. However, one of the weaknesses of the C4.5 algorithm is the occurrence of overfitting and misclassification costs which result in low prediction performance. The development of the C4.5 algorithm has been carried out in terms of split attributes such as the imprecise info-gain ratio (Credal-C4.5) method using Imprecise Probability Theory, bosting gain ratio (C5.0) and average gain. This research applies the termination coefficient value (R^2) as a method for modifying the gain ratio in selecting attributes as decision tree nodes which is then implemented to predict student graduation on time using a case study at the Universitas Islam Madura (UIM). Testing of the decision tree model rule for predicting student graduation on time at UIM shows that the performance values of accuracy, precision and recall are 70.49%, 77.14% and 72.97%. This performance is higher compared to the C4.5 algorithm without making modifications to the coefficient of determination, especially in accuracy and recall performance, while the precision is lower but the difference is below 1%. The difference in performance values was 11.48% (positive) for accuracy and 27.03% (positive) for recall. Meanwhile, precision performance has a difference of -0.13% (negative). The application of the Knowledge Model Rule for student graduation on time at SIMAT UIM shows very good results because it displays a prediction results page.

Keywords: Coefficient of Determination; C4.5 algorithm; Decision Tree; Graduation Prediction; Universitas Islam Madura.

Corresponding Author:

Muhsi

Email: muhsi@uim.ac.id



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstrak – Salah satu algoritma untuk memprediksi yang dapat digunakan adalah algoritma pada decision tree berupa C4.5. Tahapannya dimulai dari pembentukan pohon keputusan melalui split atribut, pruning dan melakukan ekstraksi aturan (rule) atau pengetahuan untuk kemudian digunakan untuk klasifikasi dan prediksi. Namun demikian diantara kelemahan dari algoritma C4.5 adalah terjadinya overfitting dan misclassification cost yang berakibat pada rendahnya kinerja prediksi. Pengembangan algoritma C4.5 telah banyak dilakukan dalam hal split atribut seperti metode imprecise info-gain ratio (Credal-C4.5) menggunakan teori Imprecise Probability Theory, bosting gain ratio (C5.0) dan average gain. Penelitian ini melakukan penerapan nilai koefisien determinasi (R^2) sebagai metode untuk modifikasi gain ratio dalam pemilihan atribut sebagai node pohon keputusan yang selanjutnya diimplementasikan untuk memprediksi kelulusan mahasiswa tepat waktu dengan studi kasus di Universitas Islam Madura (UIM). Pengujian terhadap rule model pohon keputusan prediksi kelulusan mahasiswa tepat waktu di UIM menunjukkan bahwa nilai performa akurasi, presisi dan recall adalah 70.49%, 77.14% dan 72.97%. Performa tersebut lebih tinggi dibandingkan dengan algoritma C4.5 tanpa melakukan modifikasi dengan koefisien determinasi khususnya pada performa akurasi dan recall, sedangkan untuk presisi lebih rendah namun selisihnya di bawah 1%. Selisih nilai performanya adalah 11.48% (positif) untuk akurasi dan 27.03% (positif) untuk recall. Sedangkan pada performa presisi memiliki selisih -0.13% (negatif). Rule model pengetahuan kelulusan mahasiswa tepat waktu diintegrasikan di SIMAT UIM menunjukkan hasil yang sangat baik karena menampilkan halaman hasil prediksi dengan memberikan batasan tentang waktu yang dapat memprediksi sesuai hasil penelitian yaitu masih semester 6

Kata Kunci: Algoritma C4.5, Decesion Tree, Koefisien Determinasi, Prediksi Kelulusan, Universitas Islam Madura

I. PENDAHULUAN

Algoritma *decision tree* yang biasa digunakan untuk memprediksi memiliki tantangan tersendiri yang berhubungan dengan akurasi dan skalabilitasnya [1]. Salah satunya adalah algoritma C4.5 yang banyak digunakan di penelitian klasifikasi data mining karena mudah untuk diinterpretasikan [2], [3]. Kelemahan yang sering ditemui pada algoritma C4.5 adalah di aspek overfitting sehingga bagus dari sisi training namun lemah ketika diimplementasikan terhadap unseen data yang biasanya terjadi pada saat proses split atribut [1], [4], [5]. Penelitian ini bertujuan untuk

melakukan modifikasi gain ratio sebagai dasar penentuan atribut menjadi *root* atau daun di algoritma C4.5 menggunakan nilai dari koefisien determinasi yang digabungkan untuk peningkatan performa dari kinerja model yang dihasilkan nantinya.

Telah banyak penelitian pengembangan dari algoritma decision tree yang berhubungan dengan *gain ratio*. Seperti penggunaan metode gain yang mengukur probabilitasnya dengan nilai bits dari algoritma basis 2 sebagai minus untuk selanjutnya seluruh class yang ada dijumlahkan dengan frekuensinya. Selain itu juga terdapat peneliti mengembangkan metode *Info Gain Ratio* untuk mengatasi kelemahan dari metode gain sebelumnya yang tidak bisa diterapkan pada variabel kontinu dan untuk mengatasi missing value dari data yang kemudian dikenal dengan nama algoritma C4.5 [6], [7]. Sementara itu *Abellan (2014)* menggunakan *Imprecise Info Gain (IIG)* dimana dalam melakukan split atribut terhadap set data menghitung nilai *imprecise probability* dan *uncertainty measure* [8]. *Rahayu (2015)* menggunakan metode *Average Gain*, *Threshold Pruning* dan *Cost Complexity Pruning* untuk split atribut pada algoritma C4.5 [1]. *Sarailidis (2023)* juga melakukan penelitian untuk mengatasi *over-fitting* menggunakan teknik *pruning* [9]. Sementara terdapat peneliti lain, yaitu *Chawla (2023)* mengusulkan metode untuk *smoothing* yang diberi nama *m-estimation* dimana penggunaan metode ini dimaksudkan untuk memperoleh nilai perkiraan terbaik dari probabilitas [10].

Penelitian ini merupakan terapan dari penelitian dasar yang dilakukan penulis sebelumnya [11]. Pada penelitian tersebut melakukan modifikasi terhadap nilai *gain ratio* yang sebelumnya digunakan sebagai dasar penentuan split atribut. Namun pada penelitian tersebut nilai *gain ratio* dikalikan dengan nilai koefisien determinasi atau *R-square* (R^2) berupa nilai hasil penghitungan dari formula regresi sebagai dasar penentuan split atribut. Nilai tersebut digunakan untuk melihat kemampuan model regresi dalam menerangkan seberapa besar pengaruh variabel independen terhadap variabel dependen [12]. Koefisien determinasi juga digunakan untuk mengetahui sejauh mana nilai kontribusi variabel bebas dalam model regresi memiliki kemampuan dalam menjelaskan variasi dari variabel terikatnya [13]. Dalam perhitungan statistik regresi suatu nilai koefisien determinasi akan ditemui dengan melihat nilai *R-square* (R^2) berupa range angka 0 – 1. Atas dasar tersebut nilai R^2 setiap atribut data yang akan diolah menggunakan algoritma C4.5 digabungkan dengan nilai *gain ratio* untuk penentuan split data dengan tujuan meningkatkan performa dari model yang dihasilkan [11].

Decision tree merupakan teknik yang membantu dalam pengambilan keputusan yang menyerupai bentuk pohon atau berhirarki [8]. Model yang dihasilkan nantinya berupa prosedur rekursif, dimana dalam satu set suatu unit statistik secara progresif dibagi menjadi beberapa kelompok berdasarkan aturan pembagian yang bertujuan untuk memaksimalkan homogenitas atau kemurnian ukuran variabel respon di setiap kelompok yang diperoleh [14]. Beberapa algoritma yang dapat digunakan dalam *decision tree*, adalah *CHAID (Chi-squared Automatic Interaction Detection)*, *CART (Classification And Regression Tree)*, *ID3 (Iterative Dichotomiser 3)*, dan C4.5 yang merupakan pengembangan dari *ID3*. Kemudian dikembangkan oleh beberapa peneliti menjadi *Credal DT* dan *Credal C4.5* [8], [14]. Algoritma tersebut dapat diimplementasikan di teknik data mining berupa klasifikasi terhadap data yang sangat banyak sekali dengan maksud mengambil pengetahuan melalui pemahaman terhadap karakteristik yang ada di dalamnya.

Ekstraksi pengetahuan dari kumpulan data yang sangat banyak dan dilakukan dengan berbagai metode atau algoritma merupakan konsep dari data mining [15], [16], [17]. Teknik data mining untuk ekstraksi pengetahuan tersebut dapat berbentuk deskripsi, estimasi, prediksi, klasifikasi, klusterisasi dan asosiasi [1]. Masing-masing teknik tersebut memiliki algoritma atau metode tersendiri sesuai dengan karakteristik masing-masing seperti untuk prediksi dan estimasi dapat menggunakan algoritma *decision tree*. Implementasi algoritma *decision tree* telah banyak dilakukan di berbagai bidang seperti bidang kesehatan berupa pengklasifikasian terhadap penyakit kanker [18], [19], bidang bisnis perbankan berupa analisa resiko yang dilakukan di bank *Union-27* Eropa [20]. Sementara di bidang pendidikan juga telah dilakukan untuk memprediksi aktifitas siswa dalam mengikuti pelajaran setiap hari [3], [21], [22], [23].

II. METODE

A. Data

Kelulusan tepat waktu yang dimaksud dalam penelitian ini adalah mahasiswa UIM yang lulus maksimal pada semester 8, sedangkan bagi mahasiswa yang lulus di atas 8 semester dianggap lulus tidak tepat waktu. Data yang digunakan dalam penelitian ini berupa *datasets private*, yaitu data kelulusan mahasiswa di Universitas Islam Madura (UIM) tahun 2022. Data diperoleh dari system informasi manajemen akademik terpadu (SIMAT) UIM dengan data pembandingan dari pangkalan data Pendidikan tinggi (PDDIKTI). Data yang diambil berupa indeks prestasi semester (IPS) mulai dari semester 1 sampai dengan semester 8 dan kelulusan dari mahasiswa tahun masuk 2018. Untuk lebih jelasnya informasi dari dataset yang digunakan seperti ditunjukkan Tabel 1.

Pada Tabel 1 terlihat bahwa karakteristik dataset yang akan digunakan sesuai dengan formula algoritma C4.5 dan Regresi Liner yang akan digunakan nantinya. Dataset pada Tabel 1 digunakan dalam menyusun *rule model* untuk

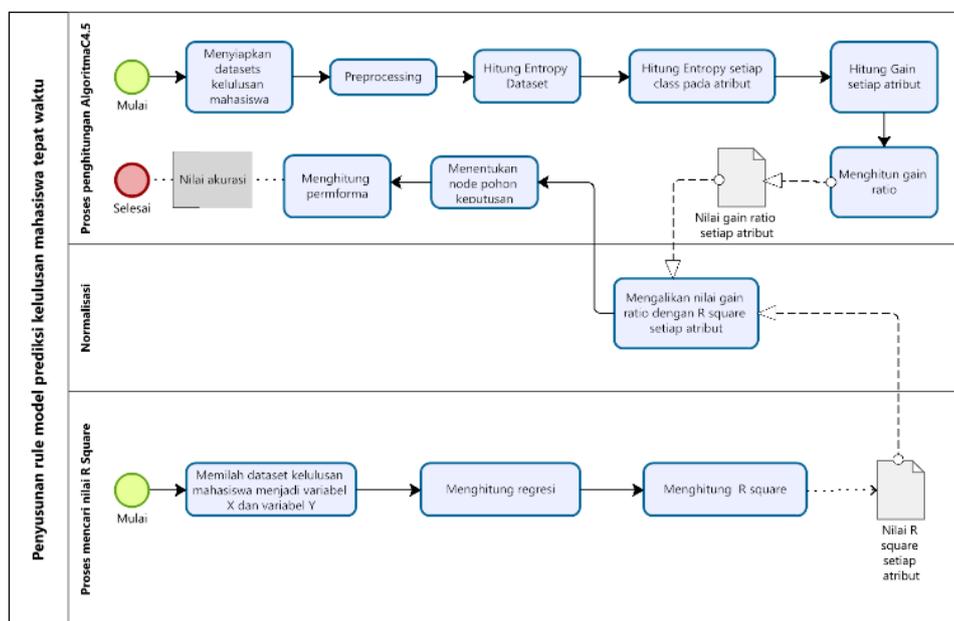
memprediksi kelulusan tepat waktu mahasiswa dan evaluasi dari kinerja *rule mode* tersebut sehingga dataset akan dipilah menjadi dua, yaitu data *training* sebanyak 80% dan data *testing* sebanyak 20%.

TABEL 1
INFORMASI DAN DESKRIPSI DATASET

| Informasi | Deskripsi |
|----------------|---------------|
| Karakter | Multivariate |
| Tipe | Real, Integer |
| Jumlah data | 306 |
| Jumlah atribut | 9 |
| Tahun data | 2022 |

B. Tahapan Penelitian

Penelitian ini merupakan terapan dari penelitian sebelumnya [10] berupa penggabungan antara algoritma C4.5 yang dikembangkan oleh *Quinlan* [7] dengan nilai koefisien determinasi (R^2) dari atribut dikorelasikan dengan kelas label. Penggunaan nilai R^2 dimaksudkan untuk modifikasi *gain ratio* sebagai dasar penentuan atribut sehingga *node* pohon keputusan yang didapat dengan maksud peningkatan performa dari *rule model*. Tahapan penelitian terlihat pada Gambar 1.



Gambar 1. Alur penelitian

Berdasarkan Gambar 1 dapat dijelaskan tahapan proses penyusunan *rule model* prediksi kelulusan mahasiswa tepat waktu di Universitas Islam Madura menggunakan algoritma C4.5 dengan modifikasi *gain ratio* dengan nilai koefisien determinasi adalah sebagai berikut :

1. Menyiapkan dataset kelulusan mahasiswa UIM.
2. *Preprocessing* data ada dua kali berupa preprocessing dengan hanya menyesuaikan kolom atribut dan baris.
3. Menghitung *entropy* untuk semua data dengan formula :

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

Dimana :

S : himpunan data set

n : banyaknya partisi S

pi : proporsi dari S_i terhadap S

4. Menghitung *entropy* setiap kelas menggunakan formula :

$$entropy(S_i) = \sum_{i=1}^n \frac{|S_i|}{|S|} \times entropy(S) \quad (2)$$

Dimana :
 S : himpunan data set
 n : banyaknya partisi atribut
 $|S_i|$: jumlah kasus pada partisi ke-i
 $|S|$: jumlah kasus dalam S

5. Menghitung *gain*

$$Gain(S, A) = entropy(S) - entropy(S_i) \quad (3)$$

Dimana :
 S : himpunan data set
 A : atribut
 n : banyaknya partisi atribut A
 S_i : jumlah sampel untuk atribut ke-i

6. Menghitung *info gain*

$$splitinfo(S, A) = \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (4)$$

Dimana :
 S : himpunan data set
 A : atribut
 n : banyaknya partisi atribut A
 S_i : jumlah sampel untuk atribut ke-i

7. Menghitung *gain ratio*

$$gainratio(A) = \frac{gain(S, A)}{splitinfo(S, A)} \quad (5)$$

Dimana :
 S : himpunan data set
 A : atribut
 $gain(S, A)$: *info gain* pada atribut A
 $splitinfo(S, A)$: *split info* pada atribut A

8. Memilah atribut dataset menjadi variabel *independen* (x) dan atribut label sebagai variabel *dependen* (y) untuk menyiapkan perhitungan regresi dan koefisien determinasi (R^2) menggunakan formula :

$$y = a + bx \quad (6)$$

Dimana:
 y : variabel terikat
 x : variabel bebas
 a : intercept
 b : slope

Dimana b didapat melalui formula :

$$b = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (7)$$

Karena $a = \bar{y} - b\bar{x}$ maka :

$$a = \frac{\sum_{i=1}^n y_i}{n} - b \frac{\sum_{i=1}^n x_i}{n} \quad (8)$$

9. Menghitung nilai koefisien determinasi (R^2) menggunakan formula :

$$R^2 = \frac{(n(\sum xy) - (\sum x)(\sum y))^2}{(n(\sum x^2) - (\sum x)^2)(n(\sum y^2) - (\sum y)^2)} \quad (9)$$

Dimana :

- R^2 : koefisien determinasi
- n : jumlah data
- x : variabel independen
- y : variabel dependen

10. Split atribut dilakukan dengan mengalikan nilai *gain ratio* yang paling tinggi pada masing-masing atribut dengan nilai R^2 setiap atribut. Atribut dengan nilai paling tinggi akan dipilih sebagai node akar. Demikian selanjutnya untuk setiap tahapan penentuan daun keputusan berikutnya dilakukan menggunakan proses 1 sampai 10. Proses tersebut berhenti sampai dengan atribut yang ada sudah tidak dapat lagi diturunkan menghasilkan daun keputusan yang lebih detail.
11. Menentukan pohon keputusan sebagai *rule model* untuk memprediksi kelulusan mahasiswa tepat waktu di UIM. *Rule model* tersebut kemudian dimasukkan ke dalam SIMAT sebagai fitur untuk mahasiswa dalam melakukan prediksi.
12. Menghitung performa *rule model* menggunakan *confusion matrix* untuk melihat nilai akurasi, presisi dan *recall* karena hanya memiliki dua kelas untuk masing-masing data. Di dalam penggunaan *confusion matrix* terdapat istilah yang akan digunakan untuk menghitung tingkat performa model yang telah dihasilkan. *True positif* (TP) digunakan untuk jumlah dari data yang diprediksi sama dengan kenyataan pada kelas tertentu sedang *True negatif* (TN) merupakan jumlah dari data yang diprediksi sama dengan kenyataan namun untuk kelas tertentu lainnya. *False positif* (FP) adalah jumlah data yang diprediksi tidak sama dengan kenyataan pada satu kelas sedangkan *false negatif* (FN) adalah jumlah data yang diprediksi tidak sama dengan kenyataan untuk kelas lainnya. Untuk lebih jelasnya seperti pada Tabel 2.

TABEL 2
CONFUSION MATRIX

| | Actually Positive (1) | Actually Negative (0) |
|-----------------------------------|--------------------------|--------------------------|
| Predicted Positive (1) | True Positive (TP) | False Positive (FP) |
| Predicted Negative (0) | False Negative (FN) | True Negative (TN) |

Nilai yang ada pada Table 2. kemudian dihitung menggunakan formula akurasi, presisi dan *recall*. Formula untuk masing-masing performa tersebut adalah sebagai berikut :

$$akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

$$presisi = \frac{TP}{TP+FP} \quad (11)$$

$$recall = \frac{TP}{TP+FN} \quad (12)$$

C. Perancangan Sistem

Rule model kelulusan mahasiswa tepat waktu yang dihasilkan dari penghitungan formula modifikasi *gain ratio* algoritma C4.5 dengan koefisien determinasi selanjutnya diintegrasikan ke dalam sistem informasi akademik terpadu (SIMAT). Tahapan perancangan fitur di SIMAT berupa prediksi kelulusan mahasiswa tepat waktu adalah sebagai berikut :

1) Perencanaan system

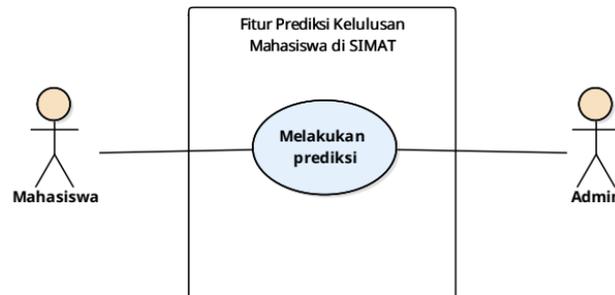
Penanaman *rule model* kelulusan mahasiswa tepat waktu di SIMAT berupa penambahan fitur sehingga perlu dirancang dengan baik. Fitur tersebut berupa menu yang akan ditambahkan di halaman mahasiswa, sehingga akan menambah kelengkapan menu yang dapat diakses oleh mahasiswa. Secara kelembagaan penambahan fitur di

SIMAT berupa prediksi kelulusan mahasiswa tepat waktu akan banyak memberikan manfaat terutama terhadap penjaminan mutu internal, fakultas maupun program studi.

2) *Analisa sistem*

a) *Use case diagram*

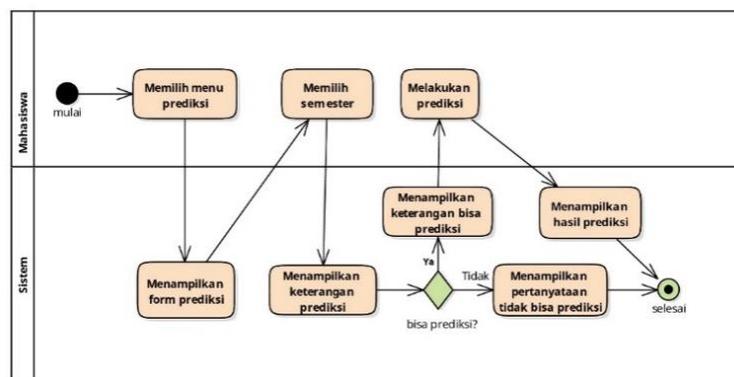
Pengguna terhadap fitur prediksi kelulusan mahasiswa tepat waktu di SIMAT UIM ada dua, yaitu mahasiswa dan admin. Pengguna tersebut akan dapat melakukan prediksi seperti terlihat pada Gambar 2.



Gambar 2. *Use case diagram*

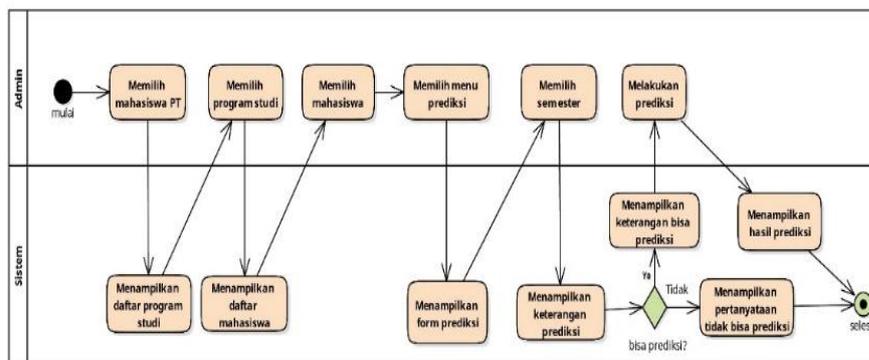
b) *Activity diagram*

Aktivitas yang dilakukan pengguna pada fitur prediksi di SIMAT UIM ada dua, yaitu prediksi yang dilakukan oleh mahasiswa dan prediksi yang dilakukan oleh admin. Untuk lebih jelasnya seperti terlihat pada Gambar 3 dan Gambar 4.



Gambar 3. *Activity diagram mahasiswa*

Pada Gambar 3 terlihat bahwa mahasiswa dapat langsung melakukan prediksi dengan memilih menu yang telah disajikan.

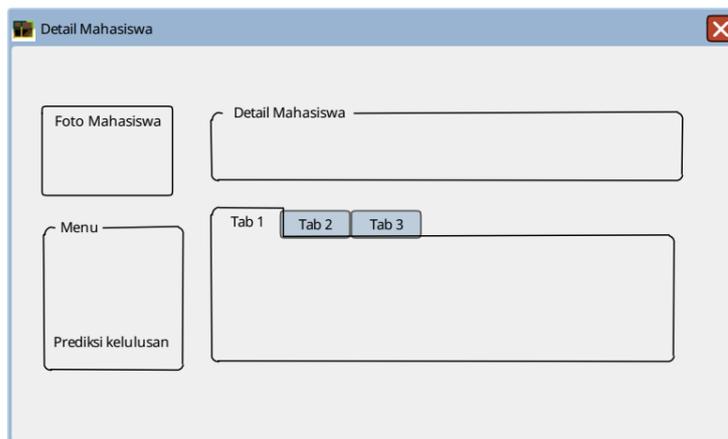


Gambar 4. *Activity diagram admin*

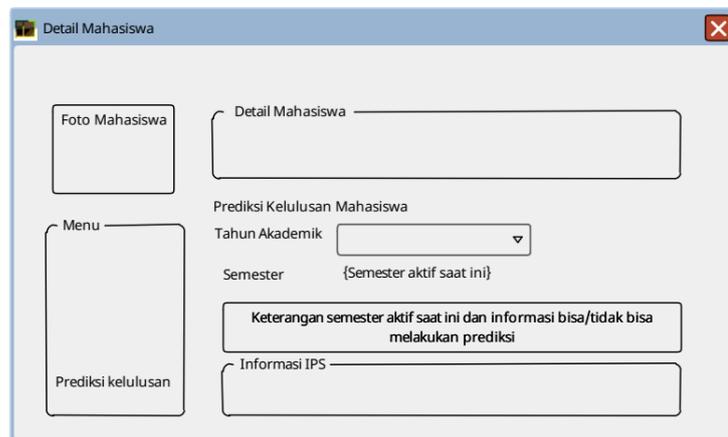
Pada Gambar 4 terlihat bahwa pengguna admin harus terlebih dahulu memasuki halaman mahasiswa sebelum melakukan prediksi.

3) *Desain sistem*

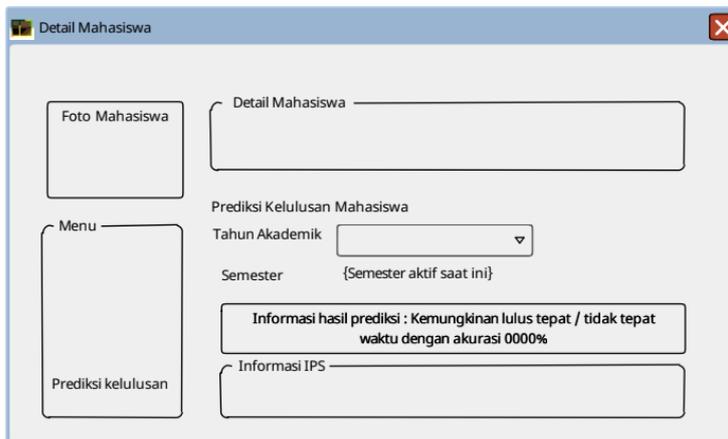
Desain sistem yang disusun berupa tampilan halaman detail mahasiswa (Gambar 5), halaman proses prediksi (Gambar 6) dan halaman hasil dari prediksi (Gambar 7). Pada gambar tersebut terlihat bahwa setiap tampilan sama-sama menyajikan informasi detail mahasiswa. Perbedaannya hanya di aspek menu untuk melakukan prediksi.



Gambar 5. Desain halaman detail mahasiswa



Gambar 6. Desain halaman prediksi



Gambar 7. Desain halaman hasil prediksi

Pada Gambar 5 nantinya akan berisi informasi detail mahasiswa berupa biodata, aktifitas, user dan lain-lain. Sedangkan pada Gambar 6 merupakan tampilan antar muka untuk mahasiswa untuk melakukan proses prediksi secara pribadi. Terdapat pilihan menu tahun akademik, informasi kondisi mahasiswa saat ini baik semester maupun IPS. Pada Gambar 7 merupakan tampilan rancangan antar muka saat menampilkan hasil prediksi.

4) *Implementasi system*

Pada tahapan ini yang dilakukan berupa *coding* untuk pembuatan sistem dan uji coba dari sistem tersebut. Terhadap uji coba sistem akan digunakan instrument berupa *functional suitability* yang berisi *test case* dari setiap halaman yang telah dibangun berupa prediksi kelulusan dan hasil prediksi kelulusan.

III. HASIL DAN PEMBAHASAN

A. *Penyiapan Data*

Data IPS mahasiswa UIM hasil download dari SIMAT dipadukan dengan data kelulusan mahasiswa di PDDIKTI untuk memastikan status dari mahasiswa tersebut. Proses selanjutnya adalah melakukan *preprocessing* dengan memilah data menjadi dua bagian berupa data training dan data testing dengan proporsi 80% dan 20% serta penentuan atribut label kelulusan dengan nilai 1 untuk lulus tepat waktu dan 0 untuk lulus tidak tepat waktu. Data IPS mahasiswa UIM seperti terlihat pada Table 3. Pemilahan data training dan testing dilakukan secara acak untuk memastikan data yang disiapkan untuk penyusunan model lebih konfident.

TABEL 3
 DATASET IPS MAHASISWA UIM

| Id | IPS1 | IPS2 | IPS3 | IPS4 | IPS5 | IPS6 | IPS7 | IPS8 | Kelulusan |
|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|------------------|
| 1 | 3.63 | 3.86 | 3.53 | 3.47 | 3.31 | 3.55 | 3.57 | 3.75 | 1 |
| 2 | 3.79 | 3.71 | 3.40 | 3.25 | 3.19 | 2.79 | 3.26 | 3.75 | 1 |
| 3 | 3.26 | 3.52 | 3.47 | 3.32 | 3.23 | 3.19 | 3.12 | 3.52 | 0 |
| 4 | 3.53 | 3.71 | 3.42 | 3.25 | 3.25 | 3.28 | 3.21 | 3.75 | 1 |
| 5 | 3.32 | 3.39 | 3.30 | 3.14 | 3.04 | 3.24 | 3.23 | 3.75 | 1 |
| 6 | 3.16 | 3.33 | 3.11 | 3.27 | 2.98 | 3.10 | 3.16 | 3.52 | 0 |
| 7 | 3.16 | 3.38 | 3.25 | 3.23 | 2.97 | 2.62 | 3.08 | 3.75 | 0 |
| 8 | 3.00 | 2.67 | 3.03 | 2.97 | 2.97 | 2.94 | 3.01 | 3.52 | 0 |
| 9 | 3.26 | 3.48 | 3.37 | 3.39 | 3.27 | 3.09 | 3.03 | 3.75 | 0 |
| 10 | 3.53 | 3.71 | 3.42 | 3.33 | 3.22 | 2.79 | 3.23 | 3.75 | 1 |
| 11 | 3.63 | 3.62 | 3.53 | 3.39 | 3.34 | 3.48 | 3.51 | 3.75 | 1 |
| 12 | 3.16 | 3.52 | 3.25 | 3.35 | 2.87 | 2.92 | 3.33 | 3.52 | 0 |
| .. | .. | .. | .. | .. | .. | .. | .. | .. | .. |
| .. | .. | .. | .. | .. | .. | .. | .. | .. | .. |
| 306 | 3.20 | 3.50 | 3.56 | 3.66 | 3.70 | 3.68 | 3.75 | 4.00 | 1 |

B. *Penyusunan Rule Model Pohon Keputusan*

Penyusunan *rule model* pohon keputusan kelulusan tepat waktu mahasiswa UIM dilakukan mulai dari nilai IPS semester 3 dengan atribut prediktor adalah IPS 1 dan IPS 2 demikian juga bagi mahasiswa pada semester berikutnya menggunakan IPS semester sebelumnya. Sebelum penyusunan pohon keputusan dengan algoritma C4.5 menggunakan nilai koefisien determinasi terlebih dahulu dilakukan uji coba algoritma C4.5 tanpa nilai koefisien determinasi untuk melihat performa terbaik dari model pohon keputusan yang dihasilkan untuk prediksi setiap semester mahasiswa mulai semester 3. Berdasarkan penyusunan dan pengujian terhadap model yang dihasilkan dengan hanya menggunakan algoritma C4.5 didapat performa seperti ditunjukkan Tabel 4.

TABEL 4
 PERFORMA IMPLEMENTASI MODEL POHON KEPUTUSAN DENGAN ALGORITMA C4.5

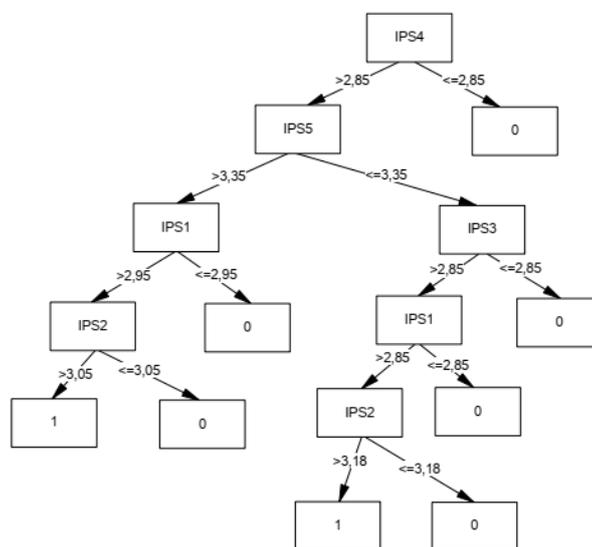
| Semester pengukuran | Akurasi | Presisi | Recall |
|----------------------------|----------------|----------------|---------------|
| 3 | 50.82% | 73.33% | 29.73% |
| 4 | 52.46% | 65.38% | 45.95% |
| 5 | 60.66% | 70.97% | 58.46% |
| 6 | 65.57% | 70.00% | 75.68% |
| 7 | 59.02% | 67.65% | 62.16% |
| 8 | 55.74% | 62.50% | 67.57% |

Pada Tabel 4 terlihat bahwa performa yang memiliki nilai presentase tertinggi adalah pada waktu pengukuran semester 6. Hasil tersebut digunakan sebagai proses penyusunan pohon keputusan dengan algoritma C4.5 berbasis nilai koefisien determinasi menggunakan atribut IPS1, IPS2, IPS3, IPS4 dan IPS5 dengan semester 6 sebagai waktu pengukuran.

Proses penyusunan pohon keputusan dengan algoritma C4.5 berbasis nilai koefisien determinasi dilakukan melalui tahapan seperti pada Gambar 1 sampai menghasilkan model pohon keputusan dengan rule yang nantinya akan diekstrak ke dalam SIMAT. Pohon keputusan kelulusan tepat waktu mahasiswa UIM menggunakan algoritma C4.5 berbasis nilai koefisien determinasi seperti ditunjukkan Gambar 8.

Berdasarkan pohon keputusan pada Gambar 8 maka *rule model* pengetahuan untuk prediksi kelulusan mahasiswa UIM adalah sebagai berikut :

- a. Jika $IPS4 \leq 2.85$ maka 0 (lulus tidak tepat waktu).
- b. Jika $IPS4 > 2.85$, $IPS5 \leq 3.35$ dan $IPS3 \leq 2.85$ maka 0 (lulus tidak tepat waktu).
- c. Jika $IPS4 > 2.85$, $IPS5 \leq 3.35$, $IPS3 > 2.85$ dan $IPS1 \leq 2.85$ maka 0 (lulus tidak tepat waktu).
- d. Jika $IPS4 > 2.85$, $IPS5 \leq 3.35$, $IPS3 > 2.85$, $IPS1 > 2.85$ dan $IPS2 \leq 3.18$ maka 0 (lulus tidak tepat waktu).
- e. Jika $IPS4 > 2.85$, $IPS5 \leq 3.35$, $IPS3 > 2.85$, $IPS1 > 2.85$ dan $IPS2 > 3.18$ maka 1 (lulus tepat waktu).
- f. Jika $IPS4 > 2.85$, $IPS5 > 3.35$ dan $IPS1 \leq 2.95$ maka 0 (lulus tidak tepat waktu).
- g. Jika $IPS4 > 2.85$, $IPS5 > 3.35$, $IPS1 > 2.95$ dan $IPS2 \leq 3.05$ maka 0 (lulus tidak tepat waktu).
- h. Jika $IPS4 > 2.85$, $IPS5 > 3.35$, $IPS1 > 2.95$ dan $IPS2 > 3.05$ maka 1 (lulus tepat waktu).



Gambar 8. Pohon keputusan kelulusan mahasiswa tepat waktu dengan algoritma C4.5 menggunakan nilai koefisien determinasi

C. Pengujian Rule Model Pohon Keputusan

Selanjutnya dari kedelapan rule tersebut dilakukan uji coba terhadap data tes pada Tabel 5 untuk melihat performanya. Hasil uji coba terhadap 61 data menggunakan *confusion matrix* didapat nilai seperti pada Tabel 6.

TABEL 5
 DATA TES UNTUK IMPLEMENTASI RULE MODEL POHON KEPUTUSAN KELULUSAN MAHASISWA TEPAT WAKTU

| No | IPS1 | IPS2 | IPS3 | IPS4 | IPS5 | Kelulusan |
|----|------|------|------|------|------|-----------|
| 1 | 3.59 | 3.07 | 3.00 | 3.39 | 3.33 | 0 |
| 2 | 3.68 | 3.24 | 3.00 | 3.65 | 3.58 | 1 |
| 3 | 3.66 | 2.51 | 3.00 | 3.18 | 3.31 | 0 |
| 4 | 3.61 | 2.40 | 3.00 | 3.18 | 3.31 | 0 |
| 5 | 3.77 | 3.43 | 3.00 | 3.60 | 3.68 | 1 |
| 6 | 3.02 | 2.86 | 3.00 | 3.03 | 2.98 | 0 |
| 7 | 3.07 | 2.93 | 3.00 | 3.00 | 3.33 | 1 |
| 8 | 3.05 | 2.86 | 3.00 | 3.08 | 2.93 | 0 |
| 9 | 3.11 | 2.93 | 3.00 | 2.98 | 3.35 | 1 |
| 10 | 3.11 | 2.88 | 3.00 | 3.08 | 3.00 | 0 |
| .. | .. | .. | .. | .. | .. | .. |
| 61 | 2.33 | 3.24 | 3.00 | 3.18 | 3.31 | 0 |

TABEL 6
CONFUSION MATRIX DARI HASIL PENERAPAN MODEL POHON KEPUTUSAN KELULUSAN MAHASISWA TEPAT WAKTU

| | True 1 | True 0 | |
|------------|--------|--------|-------|
| Prediksi 1 | 16 | 10 | 61.54 |
| Prediksi 0 | 8 | 27 | 77.14 |
| | 66.67 | 72.97 | |

Nilai *confusion matrix* pada Tabel 6 digunakan untuk menghitung performa dari model pohon keputusan prediksi kelulusan mahasiswa tepat waktu menggunakan formula 10 untuk akurasi, formula 11 untuk presisi dan formula 12 untuk *recall*. Hasil dari penggunaan ketiga formula tersebut untuk seperti terlihat pada Tabel 7.

TABEL 7
NILAI PERFORMA DARI MODEL POHON KEPUTUSAN PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU

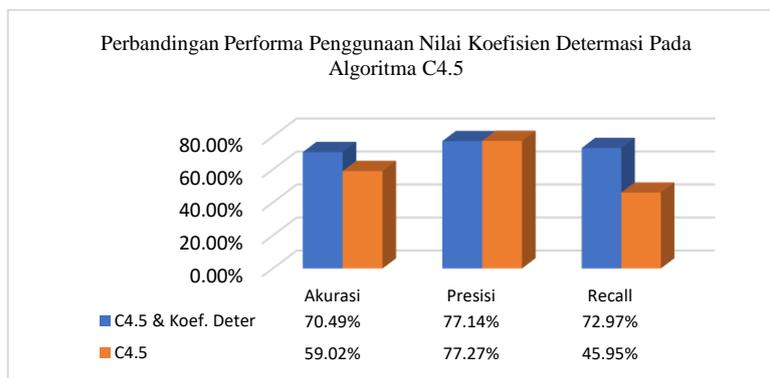
| Performa | Nilai |
|----------|--------|
| Akurasi | 70.49% |
| Presisi | 77.14% |
| Recall | 72.97% |

Pada Tabel 7 terlihat bahwa performa akurasi memiliki nilai tertinggi dibandingkan presisi dan *recall*. Hal ini menunjukkan bahwa model pohon keputusan yang dihasilkan memiliki kemampuan untuk memprediksi dengan baik kelulusan mahasiswa dengan kelas data lulus tepat atau tidak tepat waktu dari keseluruhan data yang diprediksi.

Performa *rule model* kelulusan tepat waktu dari penerapan algoritma C4.5 yang menggunakan koefisien determinasi (R^2) dibandingkan penelitian sebelumnya [11] yang tidak menggunakan R^2 untuk melihat tingkat perfforma dari keduanya. Hasil perbandingan menunjukkan bahwa performa dari penerapan *rule model* kelulusan tepat waktu dari pohon keputusan menggunakan algoritma C4.5 berbasis koefisien determinasi (R^2) memiliki nilai performa lebih tinggi untuk aspek akurasi dan *recall* sedangkan performa presisi lebih rendah seperti terlihat pada Tabel 8 dan Gambar 9

TABEL 8
PERBANDINGAN NILAI PERFORMA DARI MODEL POHON KEPUTUSAN PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU

| Performa | Nilai | | Selisih |
|----------|------------------------|----------------|---------|
| | Algoritma C4.5 & R^2 | Algoritma C4.5 | |
| Akurasi | 70.49% | 59.02% | 11.48% |
| Presisi | 77.14% | 77.27% | -0.13% |
| Recall | 72.97% | 45.95% | 27.03% |



Gambar 9. Grafik perbandingan performa algoritma C4.5 dengan dan tanpa nilai koefisien determinasi

Berdasarkan Tabel 8 dan Gambar 9 menunjukkan bahwa algoritma C4.5 yang menggunakan R^2 memiliki performa akurasi lebih tinggi, yaitu 70.49%, dibandingkan algoritma C4.5 yang tanpa R^2 , hanya 59.02%. Hal ini menunjukkan bahwa penggunaan nilai R^2 di algoritma C4.5 adalah bagus dalam memprediksi dengan benar status kelulusan mahasiswa, baik yang kemungkinan lulus tetap waktu maupun lulus tidak tepat waktu dari keseluruhan data

mahasiswa yang diprediksi. Demikian untuk performa *recall*, penggunaan nilai R^2 di algoritma C4.5 memiliki nilai lebih tinggi, yaitu 72.97%, dibandingkan tanpa R^2 yang hanya 45.95%. Hal juga menunjukkan bahwa algoritma C4.5 yang menggunakan nilai R^2 bagus dalam memprediksi benar kemungkinan mahasiswa lulus tepat waktu dibandingkan dengan keseluruhan data mahasiswa yang benar-benar lulus tepat waktu dari data lulusan mahasiswa yang diprediksi. Namun demikian untuk presisi, performanya lebih rendah algoritma C4.5 dengan nilai R^2 dibandingkan tanpa nilai R^2 dengan selisih yang sangat kecil, yaitu hanya 0.13%. Penggunaan koefisien determinasi memberikan nilai lebih *confident* karena melakukan normalisasi dari nilai *gain ratio* yang dihasilkan formula algoritma C4.5 secara umum. Karena factor inilah sehingga nilai performa akurasi dan recall lebih tinggi. Namun demikian data yang digunakan masih hanya sebatas dataset bertipe multivariat.

D. Penerapan Rule Model di SIMAT UIM

Rule model pengetahuan tentang kelulusan mahasiswa tepat waktu yang telah dihasilkan menggunakan metode integrasi nilai koefisien determinasi di algoritma C4.5 dalam pemilihan node. Selanjutnya diintegrasikan ke dalam sistem informasi akademik yang digunakan Universitas Islam Madura yang dikenal dengan nama SIMAT (Sistem Informasi Akademik Terpadu).

Modul prediksi kelulusan mahasiswa terdapat di halaman pengguna mahasiswa dan administrator SIMAT UIM. Berdasarkan penyusunan *rule model* sebelumnya bahwa mahasiswa yang dapat melakukan prediksi kelulusan adalah mahasiswa pada saat semester 6 sehingga bagi mahasiswa semester di bawahnya belum dapat melakukan prediksi. Demikian juga bagi mahasiswa dengan semester di atas semester 6 juga perlu melakukan pengaturan semester menjadi 6. Untuk lebih jelasnya seperti Gambar 10 dan Gambar 11.

| No | Thn. Akademik / Semester | Status MHS | IPS | IPK | SKS | |
|-----------|-------------------------------|------------|------|------|----------|-------|
| | | | | | Semester | Total |
| 1 | 2023/2024 Ganjil / Semester 1 | AKTIF | 3.23 | 3.23 | 19 | 19 |
| 2 | 2023/2024 Genap / Semester 2 | AKTIF | 3.42 | 3.33 | 23 | 42 |
| 3 | 2024/2025 Ganjil / Semester 3 | AKTIF | 3.50 | 3.39 | 22 | 64 |
| 4 | 2024/2025 Genap / Semester 4 | AKTIF | 0.00 | 2.49 | 23 | 64 |
| Total SKS | | | | | 23 | 64 |

Gambar 10. Mahasiswa semester 4 melakukan prediksi kelulusan

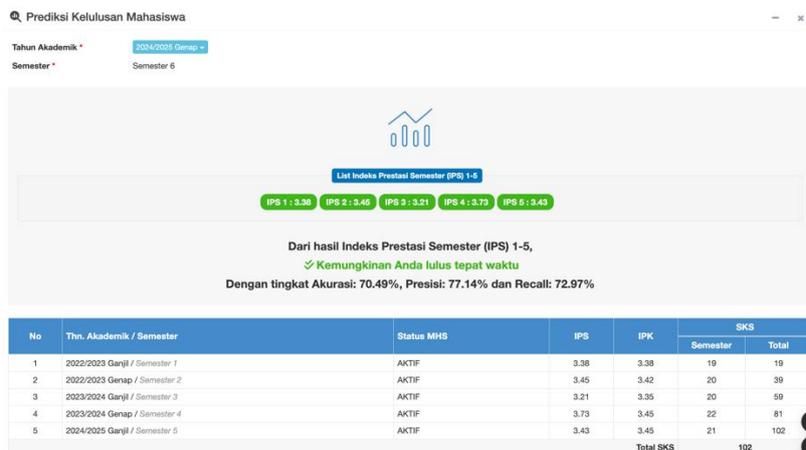
| No | Thn. Akademik / Semester | Status MHS | IPS | IPK | SKS | |
|-----------|-------------------------------|------------|------|------|----------|-------|
| | | | | | Semester | Total |
| 1 | 2022/2023 Ganjil / Semester 1 | AKTIF | 3.38 | 3.38 | 19 | 19 |
| 2 | 2022/2023 Genap / Semester 2 | AKTIF | 3.45 | 3.42 | 20 | 39 |
| 3 | 2023/2024 Ganjil / Semester 3 | AKTIF | 3.21 | 3.35 | 20 | 59 |
| 4 | 2023/2024 Genap / Semester 4 | AKTIF | 3.73 | 3.45 | 22 | 81 |
| 5 | 2024/2025 Ganjil / Semester 5 | AKTIF | 3.43 | 3.45 | 21 | 102 |
| Total SKS | | | | | 102 | 102 |

Gambar 11. Mahasiswa semester 6 melakukan prediksi kelulusan

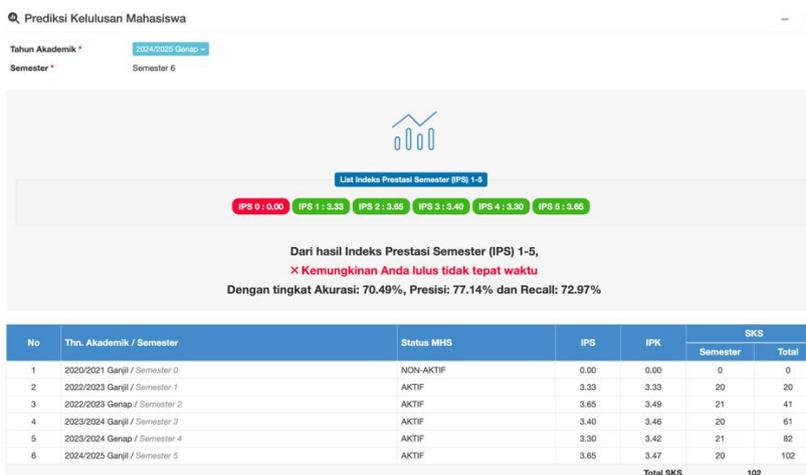
Pada Gambar 10 terlihat bahwa saat ini mahasiswa berada di semester 4 sehingga muncul pesan bahwa mahasiswa tersebut masih belum bisa melakukan prediksi. Sedangkan pada Gambar 11 terlihat bahwa mahasiswa saat ini ada di semester 6 sehingga muncul tombol sekaligus pesan untuk melakukan prediksi kelulusan tepat waktu.

Hasil prediksi kelulusan mahasiswa berdasarkan *rule model* pengetahuan dari pohon keputusan berupa kelulusan tepat waktu atau tidak tepat waktu. Dengan demikian hasil prediksi yang ditampilkan dari proses prediksi yang dilakukan mahasiswa akan menampilkan pesan yang menyatakan kemungkinan lulus tepat waktu atau kemungkinan tidak tepat waktu. Pernyataan pesan bagi yang kemungkinan lulus tepat waktu adalah “Kemungkinan Anda Lulus Tepat Waktu” dengan *font* warna hijau seperti ditunjukkan oleh Gambar 12, sedangkan pernyataan pesan bagi yang kemungkinan tidak lulus tepat waktu adalah “Kemungkinan Anda Lulus Tidak Tepat Waktu” dengan *font*

warna merah mencolok agar memudahkan pemahaman terhadap informasi yang disampaikan. Untuk lebih jelasnya seperti ditunjukkan oleh Gambar 13.



Gambar 12. Pernyataan pesan hasil prediksi kemungkinan lulus tepat waktu



Gambar 13. Pernyataan pesan hasil prediksi kemungkinan lulus tidak tepat waktu

E. Pengujian Sistem

Terdapat lima pengujian dari implementasi sistem prediksi kelulusan mahasiswa berdasarkan hasil perancangan. Hasil pengujian seperti terlihat pada Tabel 9.

TABEL 9
HASIL PENGUJIAN *FUNCTIONAL SUITABILITY* BERDASARKAN *SELF REVIEW*

| Test ID | Test case | Hasil yang diharapkan | Hasil aktual | Sukses/Gagal |
|---------|---|--|--|--------------|
| A1 | Menguji prediksi kelulusan mahasiswa dengan semester >6 | Menampilkan pesan untuk memilih semester 6 untuk melakukan prediksi | Halaman bagi mahasiswa dengan semester > 6 | sukses |
| A2 | Menguji prediksi kelulusan mahasiswa dengan semester =6 | Menampilkan halaman untuk melakukan analisa sekarang | Halaman bagi mahasiswa dengan semester = 6 | Sukses |
| A3 | Menguji prediksi kelulusan mahasiswa dengan semester <6 | Menampilkan pesan bahwa masih belum dapat melakukan prediksi dengan posisi semester sekarang karena prediksi dapat | Halaman bagi mahasiswa dengan semester < 6 | Sukses |

| <i>Test ID</i> | <i>Test case</i> | Hasil yang diharapkan | Hasil aktual | Sukses/ Gagal |
|----------------|--|--|---|--------------------------|
| | | dilakukan pada saat semester 6 | | |
| A4 | Menguji prediksi mahasiswa yang kemung-kinan lulus tepat waktu | Menampilkan pesan bahwa kemungkinan anda lulus tepat waktu berdasarkan IPS1 sd IPS5 dengan tingkat akurasi sekian persen | Menampilkan pesan kemungkinan lulus tepat waktu | sukses |
| A5 | Menguji prediksi mahasiswa yang kemung-kinan lulus tidak tepat waktu | Menampilkan pesan bahwa kemungkinan anda lulus tidak tepat waktu berdasarkan IPS1 sd IPS5 dengan tingkat akurasi sekian persen | Menampilkan pesan kemungkinan lulus tidak tepat waktu | Sukses |

Hasil pengujian terhadap semua *test case* didapat bahwa semua item tes atau 100% sukses menampilkan hasil yang diharapkan.

IV. SIMPULAN

Implementasi teknik pemilihan atribut untuk menentukan node di algoritma C4.5 melalui modifikasi *gain ratio* menggunakan nilai koefisien determinasi (*R square*) terhadap data kelulusan mahasiswa tepat waktu di Universitas Islam Madura adalah performa yang dihasilkan lebih tinggi dibandingkan dengan pohon keputusan yang hanya dihitung dengan algoritma C4.5 tanpa R^2 terutama pada aspek performa akurasi dan *recall*. Perbandingan nilai performa tersebut adalah 70.49% dan 59.02% dengan selisih positif 11.48% untuk akurasi serta 72.97% dan 45.95% dengan selisih positif 27.03% untuk *recall*. Pada performa presisi lebih rendah yaitu 77.14% dan 77.27% dengan selisih negatif -0.13%. Implementasi *rule model* pengetahuan kelulusan mahasiswa Universitas Islam Madura di sistem informasi akademik SIMAT menunjukkan hasil yang sangat baik karena menampilkan hasil prediksi dan memberikan batasan tentang waktu yang dapat memprediksi sesuai hasil penelitian yaitu hanya di semester 6. Hal ini terlihat dari hasil pengujian *functional suitability* fitur prediksi di SIMAT semuanya sukses dieksekusi. Untuk tujuan prediksi kelulusan mahasiswa tepat waktu diharapkan perlu penambahan atribut seperti status mahasiswa bekerja/tidak bekerja atau nilai lain yang memiliki korelasi terhadap aktivitas perkuliahan mahasiswa. Penggunaan data *training* yang lebih banyak sangat disarankan untuk menghasilkan performa lebih baik. Proses perhitungan dari nilai *gain ratio* di algoritma C4.5 dengan nilai koefisien determinasi (*R square*) dilakukan secara tersistem sehingga memudahkan pada saat penambahan data *training*.

UCAPAN TERIMA KASIH

Ucapan terima kasih penulis samapaikan kepada pihak Biro Administrasi dan Perencanaan Sistem Informasi Universitas Islam Madura khususnya kepada tim pengembang SIMAT yang telah membantu dan berkenan melakukan uji coba fitur prediksi kelulusan mahasiswa tepat waktu di lingkungan Universitas Islam Madura.

DAFTAR PUSTAKA

- [1] E. S. Rahayu, R. Satria, and C. Supriyanto, "Penerapan Metode Average Gain, Threshold Pruning dan Cost Complexity Pruning Untuk Split Atribut Pada Algoritma C4.5," *Journal of Intelligent Systems*, vol. 1, no. 2, pp. 91–97, 2015.
- [2] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning," *Decision Analytics Journal*, vol. 3, p. 100071, Jun. 2022, doi: 10.1016/j.dajour.2022.100071.
- [3] M. Muhsi, "Model dan Analisa Faktor Eksternal Aktifitas Siswa Kelas X TKJ SMKN 1 Pakong Pamekasan Menggunakan Algoritma Decision Tree," *Jurnal Aplikasi Teknologi Informasi dan Manajemen (JATIM)*, vol. 2, no. 2, pp. 92–106, 2021, doi: 10.31102/jatim.v2i2.1239.
- [4] H. Bin Wang and Y. J. Gao, "Research on C4.5 algorithm improvement strategy based on MapReduce," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 160–165. doi: 10.1016/j.procs.2021.02.045.
- [5] J. Wang, "Application of C4.5 Decision Tree Algorithm for Evaluating the College Music Education," *Mobile Information Systems*, vol. 2022, 2022, doi: 10.1155/2022/7442352.
- [6] S. L. Salzberg, "C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993," *Mach Learn*, vol. 16, no. 3, pp. 235–240, 1994, doi: 10.1007/bf00993309.
- [7] J. R. Quinlan, *{C4}.5 - Programs for Machine Learning*. San Mateo, California: Morgan Kaufmann Publishers, 1993.

- [8] C. J. Mantas and J. Abellán, "Credal-C4.5: Decision tree based on imprecise probabilities to classify noisy data," *Expert Syst Appl*, vol. 41, no. 10, pp. 4625–4637, Aug. 2014, doi: 10.1016/j.eswa.2014.01.017.
- [9] G. Sarailidis, T. Wagener, and F. Pianosi, "Integrating scientific knowledge into machine learning using interactive decision trees," *Comput Geosci*, vol. 170, Jan. 2023, doi: 10.1016/j.cageo.2022.105248.
- [10] N. V. Chawla, "Many Are Better Than One: Improving Probabilistic Estimates from Decision Trees," in *Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment*, J. Quiñero-Candela, I. Dagan, B. Magnini, and F. d'Alché-Buc, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 41–55.
- [11] M. Muhsi, S. Suprpto, and R. Rofiuddin, "Node Selection Method for Split Attribute in C4.5 Algorithm Using the Coefficient of Determination Values for Multivariate Data Set," *Jurnal Penelitian Pendidikan IPA*, vol. 9, no. 7, pp. 5574–5583, Jul. 2023, doi: 10.29303/jppipa.v9i7.4031.
- [12] A. Yalçınkaya, İ. G. Balay, and B. Şenoğlu, "A new approach using the genetic algorithm for parameter estimation in multiple linear regression with long-tailed symmetric distributed error terms: An application to the Covid-19 data," *Chemometrics and Intelligent Laboratory Systems*, vol. 216, Sep. 2021, doi: 10.1016/j.chemolab.2021.104372.
- [13] A. H. AL-Marshadi, M. Aslam, and A. Alharbey, "Selecting the covariance structure for the seemingly unrelated regression models," *J King Saud Univ Sci*, vol. 34, no. 4, Jun. 2022, doi: 10.1016/j.jksus.2022.102027.
- [14] I. D. Mienye, Y. Sun, and Z. Wang, "Prediction performance of improved decision tree-based algorithms: A review," in *Procedia Manufacturing*, Elsevier B.V., 2019, pp. 698–703. doi: 10.1016/j.promfg.2019.06.011.
- [15] J. R. Saura, D. Palacios-Marqués, and D. Ribeiro-Soriano, "Using data mining techniques to explore security issues in smart living environments in Twitter," *Comput Commun*, vol. 179, pp. 285–295, Nov. 2021, doi: 10.1016/j.comcom.2021.08.021.
- [16] H. Thakkar, V. Shah, H. Yagnik, and M. Shah, "Comparative anatomization of data mining and fuzzy logic techniques used in diabetes prognosis," *Clinical eHealth*, vol. 4, pp. 12–23, 2021, doi: 10.1016/j.ceh.2020.11.001.
- [17] J. Kalezhi, M. Chibuluma, C. Chembe, V. Chama, F. Lungo, and D. Kunda, "Modelling Covid-19 infections in Zambia using data mining techniques," *Results in Engineering*, vol. 13, Mar. 2022, doi: 10.1016/j.rineng.2022.100363.
- [18] M. Patrício *et al.*, "Using Resistin, glucose, age and BMI to predict the presence of breast cancer," *BMC Cancer*, vol. 18, no. 1, 2018, doi: 10.1186/s12885-017-3877-1.
- [19] H. Sastypratiwi, Y. Yulianti, and H. Muhardi, "Uji Komparasi Algoritma Naïve Bayes dan Decision Tree Classification Menggunakan Covid-19 Dataset," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 8, no. 1, p. 1, 2022, doi: 10.26418/jp.v8i1.49841.
- [20] T. Kristóf and M. Virág, "EU-27 bank failure prediction with C5.0 decision trees and deep learning neural networks," *Res Int Bus Finance*, vol. 61, Oct. 2022, doi: 10.1016/j.ribaf.2022.101644.
- [21] M. S. Abubakari and S. Suprpto, "Educational Data Mining to Predict Students Performance Based on Deep Learning Neural Network," in *Proceeding International Conference On Health, Social Sciences And Technology*, 2021, pp. 13–16. Accessed: Aug. 03, 2022. [Online]. Available: <https://ojs.poltekkespalembang.ac.id/index.php/icohsst/article/view/697>
- [22] I. G. T. Isa and F. Elfaladonna, "Penilaian Kinerja Akurasi Metode Klasifikasi dalam Dataset Penerimaan Mahasiswa Baru Universitas XYZ," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 8, no. 2, p. 292, 2022, doi: 10.26418/jp.v8i2.54316.
- [23] A. Ermillian and K. Nugroho, "Perancangan Model Deteksi Potensi Siswa Putus Sekolah Menggunakan Metode Logistic Regression Dan Decision Tree," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 9, no. 3, pp. 281–295, Dec. 2024, doi: 10.30591/jpit.v9i3.8007.