

Penerapan Algoritma C4.5 dengan Optimasi *Particle Swarm Optimization* untuk Prediksi Kelulusan Mahasiswa

Mercy Hermawati

Teknik Informatika, Universitas Indraprasta PGRI, Indonesia

Article Info

Article history:

Received Apr 18, 2023

Revised Jul 05, 2023

Accepted Sep 18, 2023

Keywords:

Data Mining

C4.5 Algorithm

Particle Swarm Optimization

Prediction

Graduation

ABSTRACT

College is a place for students to pursue higher education. Both state and private universities compete to be the best universities to produce the best graduates. The number of student graduates is an indicator of the success of a higher education institution, which will have an impact on government accreditation and public assessment. The aim of this research is to predict student graduation in order to know whether they will graduate on time or late by applying data mining techniques, namely classification using the C4.5 algorithm to obtain patterns of student graduation delays and the particle swarm optimization (PSO) algorithm to increase the accuracy of the C4 algorithm. 5. Testing uses cross-validation tests, confusion matrix, and ROC curve. The results of this research are that the C4.5 algorithm with particle swarm optimization (PSO) has an accuracy value of 86.72%, which is better than the C4.5 algorithm, whose accuracy is 82.05% and the difference between them is 4.67%. The difference between the AUC value of 0.033 was obtained from the C4.5 algorithm model, which had an AUC value of 0.870 with a good classification diagnostic level, and the C4.5 algorithm with PSO had an AUC value of 0.903 with an excellent classification diagnostic level. IPS3 is the attribute that most influences the accuracy of student graduation. The results of the C4.5 algorithm rule with PSO can be applied to create applications for GUI-based student graduation predictions.

Copyright © 2023 Universitas Indraprasta PGRI.
All rights reserved.

Corresponding Author:

Mercy Hermawati,

Teknik Informatika,

Universitas Indraprasta PGRI,

Jl. Nangka No. 58 C, Tanjung Barat, Jagakarsa, Jakarta Selatan.

Email: mercy.hermawati@gmail.com

1. PENDAHULUAN

Perguruan tinggi merupakan salah satu tempat untuk menempuh jenjang pendidikan yang lebih tinggi. Baik perguruan tinggi negeri dan swasta bersaing untuk menjadi perguruan tinggi terbaik untuk menghasilkan lulusan terbaik. Berdasarkan Undang-Undang RI Nomor 20, perguruan tinggi diharapkan dapat memberikan pendidikan yang berkualitas kepada mahasiswa sehingga dapat menghasilkan sumber daya manusia yang berkualitas, kompeten dan kreatif.

Buku VI matriks penilaian instrument akreditasi program studi BAN PT menerangkan bahwa persentase mahasiswa lulus tepat waktu merupakan salah satu elemen penilaian akreditasi universitas[1]. Dengan demikian, dianggap bahwa salah satu faktor penentu mutu perguruan tinggi adalah persentase

mahasiswa yang menyelesaikan studinya tepat waktu. Akreditasi perguruan tinggi juga dapat menjadi bahan promosi untuk menarik calon mahasiswa barunya.

Untuk mencapai *goals* tersebut diperlukan antisipasi penanganan dini terhadap mahasiswa yang berpotensi lulus tidak tepat waktu dengan suatu teknik yang dapat mendeteksi kemungkinan hal itu terjadi. Perguruan tinggi memiliki *database* yang menyimpan data akademik dan biodata mahasiswa. Data tersebut apabila digali dengan tepat maka akan dapat diketahui pola atau pengetahuan (*knowledge*) untuk mengambil keputusan. Serangkaian proses mendapatkan pengetahuan atau pola dari kumpulan data disebut dengan *data mining*[2].

Penelitian yang telah dilakukan dengan teknik *data mining* pohon keputusan algoritma C4.5 yaitu untuk mengatasi kuota mahasiswa yang meningkat setiap tahunnya namun tidak semua dapat lulus tepat pada waktunya berakibat pada akumulasi jumlah siswa[3]. Untuk *monitoring* dan pengambilan keputusan BAAK terhadap mahasiswa aktif, apakah mereka dapat lulus tepat waktu, cepat, terlambat atau *drop out*[4]. Penelitian dewi dkk, menggunakan *neural network* dan *particle swarm optimization* untuk membantu pihak manajemen dalam mengambil keputusan dan mendapatkan parameter terbaik dari metode yang digunakan untuk dapat memprediksi kelulusan mahasiswa[5]. Penelitian rusda wajhillah, menggunakan algoritma C4.5 dengan PSO untuk prediksi penyakit jantung, dari 13 atribut pada dataset *UCI Machine Learning Data Repository* setelah diseleksi hanya menjadi sembilan atribut yang digunakan dalam menentukan prediksi penyakit jantung[6]. Penelitian elah nurlelah, menerapkan algoritma *particle swarm optimization* untuk mengeliminasi atribut input pada metode algoritma C4.5, untuk meningkatkan akurasi prediksi penyakit liver. Sehingga diperoleh akurasi diagnosis penyakit liver yang lebih baik dibanding dengan menggunakan metode individual algoritma C4.5[7]. Penelitian santoso dkk, algoritma C4.5 memiliki kelemahan yaitu pada pembacaan data yang berjumlah besar. Sehingga untuk meningkatkan kinerja klasifikasi algoritma C4.5 ditambahkan algoritma lain yaitu *Particle Swarm Optimization* (PSO) sebagai fitur seleksi atribut pada dataset lulusan santri Ma'hadul Qur'an untuk meningkatkan akurasi dalam klasifikasi lama menghafal Al-Qur'an[8]. Berdasarkan latar belakang diatas maka diambil rumusan masalah sebagai berikut:

1. Bagaimana mendapatkan pola keterlambatan kelulusan mahasiswa menggunakan algoritma C4.5 berbasis *Particle Swarm Optimization* (PSO)?
2. Bagaimana akurasi algoritma C4.5 berbasis *Particle Swarm Optimization* (PSO) dalam memprediksi ketepatan kelulusan mahasiswa?
3. Bagaimana penerapan algoritma C4.5 berbasis *Particle Swarm Optimization* (PSO) untuk prediksi ketepatan kelulusan mahasiswa di Universitas Pakuan Bogor?

2. METODE

2.1. Metode Penelitian

Jenis penelitian yang dilakukan pada penelitian ini adalah jenis penelitian eksperimen. Metode ini menguji kebenaran sebuah hipotesis dengan statistik dan menghubungkannya dengan masalah penelitian[9]. Tujuan dari penelitian ini adalah menerapkan algoritma C4.5 untuk klasifikasi kelulusan mahasiswa dan algoritma *particle swarm optimization* (PSO) untuk pemberian bobot atribut guna meningkatkan akurasi dari algoritma C4.5 dalam segi seleksi atribut. Penelitian ini menggunakan data primer yang diperoleh dari Pusat Teknologi Informasi dan Komunikasi Universitas Pakuan Bogor.

2.2. Metode Pemilihan Sample

1. Populasi

Jenis penelitian yang dilakukan pada penelitian ini adalah jenis penelitian eksperimen. Metode ini menguji kebenaran. Populasi adalah keseluruhan subyek penelitian. Adapun populasi dalam dalam penelitian ini adalah *dataset* mahasiswa Strata 1 Reguler Ilmu Komputer Universitas Pakuan Bogor yang sudah dinyatakan lulus.

2. Sampel

Sampel adalah sebagian atau wakil populasi yang diteliti. Sampel pada penelitian ini dihitung dengan metode Slovin untuk menentukan jumlah sampel yang digunakan sebagai data pelatihan dan data pengujian proses *data mining*. Metode slovin ditunjukkan pada persamaan berikut.

$$n = \frac{N}{1 + N e^2}$$

Keterangan:

n = jumlah sampel

N = jumlah keseluruhan data/populasi

e = galat kesalahan (ditentukan sebesar 5%)

Data mahasiswa Prodi Ilmu Komputer program regular angkatan 2011 dan 2012 yang sudah dinyatakan lulus seluruhnya berjumlah (N) 1105 mahasiswa. Sesuai dengan rumus Slovin, dapat ditentukan jumlah sampel yang diambil yaitu sebanyak 293 data mahasiswa.

2.3. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah:

1. Observasi,
2. Wawancara,
3. Studi Literatur.

2.4. Instrumentasi Penelitian

Dalam penelitian ini ada beberapa instrumen yang digunakan yaitu:

1. Penelitian ini menggunakan data primer berupa data kelulusan mahasiswa angkatan 2011 dan 2012 program Strata 1 regular Ilmu Komputer yang sudah dinyatakan lulus diperoleh dari Pusat Teknologi Informasi dan Komunikasi Universitas Pakuan Bogor.
2. Data disajikan dalam bentuk tabulasi model dan variabel masing-masing sebanyak 293 data.
3. Perangkat lunak yang digunakan untuk menganalisis adalah *Rapid Miner*.

2.5. Teknik Analisis Data

Berdasarkan arsitektur *data mining* CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang terdiri dari 6 tahap[10], maka dilakukan tahap-tahap sebagai berikut:

1. *Business/Research Understanding Phase*
2. *Data Understanding Phase*
3. *Data Preparation Phase*
4. *Modelling Phase*
5. *Evaluation Phase*
6. *Deployment Phase*

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data dan Analisis Data

Atribut yang diambil dari data yang dimiliki dari PUTIK Universitas Pakuan Bogor yaitu: jenis kelamin, umur, asal sekolah, IPS 1, IPS 2, IPS 3, IPS 4 IPK dan lama studi.

Tabel 1. Sampel Dataset

NPM	JK	Umur	Asal Sekolah	IPS 1	IPS 2	IPS 3	IPS 4	IPK	Lama Studi
065111170	L	22	2	3.25	2.82	2.38	2.48	2.73	10
065111175	L	23	2	2.57	3.16	3	3.1	2.96	11
065111176	L	23	2	3	3.18	2.68	1.91	2.87	10
065111177	L	22	2	3	3.5	3.75	3.5	3.44	8
065111178	L	24	2	2.7	3.36	3.44	3.57	3.27	8
065111179	L	23	2	3	3.11	3	3.57	3.17	8
065111180	L	23	1	4	4	3.7	3.67	3.84	8
065111184	L	23	2	2.8	2.9	2.92	2.8	2.86	9
065111191	L	22	1	3.63	3.25	3.25	3.18	3.33	8
065111194	P	22	2	3	2.86	3	2.73	2.90	9
065111197	L	24	1	2.71	2.65	2.69	2.64	2.67	10
065111200	L	22	1	3.4	3.18	2.75	2.27	2.90	10
065111201	L	22	2	3.25	3.05	3	2.5	2.95	9
065111202	P	22	1	2.8	3.13	3	2.14	2.77	11
065111205	P	22	1	3.6	3.5	2.89	2.85	3.21	8

Kelas data yang digunakan untuk *data mining* disiapkan (*preprocessing*) sehingga memiliki kelas *binominal* atau *polynomial* sesuai aturan yang telah dibuat berdasarkan nilai datanya. Tabel 4 merupakan pembagian *variabel* dan kelas data yang digunakan dalam analisis *data mining*.

Tabel 2. Pembagian Variabel dan Kelas Data

Variabel	Nama Field	Jenis Kelas Data	Kelas Data yang Digunakan
Y	Status Kelulusan	Binominal	Tepat Waktu, Terlambat
X1	Jenis Kelamin	Binominal	Laki-Laki, Perempuan
X2	Umur	Binominal	≤23, > 23
X3	Asal Sekolah	Binominal	Bogor, Luar Bogor
X4	IPS1	Polynomial	Kurang Memuaskan, Memuaskan, Sangat Memuaskan, <i>Cumlaude</i>
X5	IPS2	Polynomial	Kurang Memuaskan, Memuaskan, Sangat Memuaskan, <i>Cumlaude</i>
X6	IPS3	Polynomial	Kurang Memuaskan, Memuaskan, Sangat Memuaskan, <i>Cumlaude</i>
X7	IPS4	Polynomial	Kurang Memuaskan, Memuaskan, Sangat Memuaskan, <i>Cumlaude</i>
X8	IPK	Polynomial	Memuaskan, Sangat Memuaskan, <i>Cumlaude</i>

Potongan data hasil *preprocessing* sesuai dengan jenis kelas datanya dapat dilihat pada Tabel 5 dibawah ini.

Tabel 3: Potongan Hasil Preprocessing

JK	Umur	Asal Sekolah	IPS 1	IPS 2	IPS 3	IPS 4	IPK	Status Kelulusan
L	<=23	Bogor	Sangat Memuaskan	Sangat Memuaskan	Memuaskan	Memuaskan	Memuaskan	TERLAMBAT
L	<=23	Bogor	Memuaskan	Sangat Memuaskan	Sangat Memuaskan	Sangat Memuaskan	Sangat Memuaskan	TERLAMBAT
L	<=23	Bogor	Sangat Memuaskan	Sangat Memuaskan	Memuaskan	Kurang Memuaskan	Sangat Memuaskan	TERLAMBAT
L	<=23	Bogor	Sangat Memuaskan	Sangat Memuaskan	<i>Cumlaude</i>	Sangat Memuaskan	Sangat Memuaskan	TEPAT WAKTU
L	>23	Bogor	Memuaskan	Sangat Memuaskan	Sangat Memuaskan	<i>Cumlaude</i>	Sangat Memuaskan	TEPAT WAKTU
L	<=23	Bogor	Sangat Memuaskan	Sangat Memuaskan	Sangat Memuaskan	<i>Cumlaude</i>	Sangat Memuaskan	TEPAT WAKTU
L	<=23	Luar Bogor	<i>Cumlaude</i>	<i>Cumlaude</i>	<i>Cumlaude</i>	<i>Cumlaude</i>	<i>Cumlaude</i>	TEPAT WAKTU
L	<=23	Bogor	Sangat Memuaskan	TERLAMBAT				
L	<=23	Luar Bogor	<i>Cumlaude</i>	Sangat Memuaskan	Sangat Memuaskan	Sangat Memuaskan	Sangat Memuaskan	TEPAT WAKTU
P	<=23	Bogor	Sangat Memuaskan	Sangat Memuaskan	Sangat Memuaskan	Memuaskan	Sangat Memuaskan	TERLAMBAT
L	>23	Luar Bogor	Memuaskan	Memuaskan	Memuaskan	Memuaskan	Memuaskan	TERLAMBAT
L	<=23	Luar Bogor	Sangat Memuaskan	Sangat Memuaskan	Memuaskan	Memuaskan	Sangat Memuaskan	TERLAMBAT
L	<=23	Bogor	Sangat Memuaskan	Sangat Memuaskan	Sangat Memuaskan	Memuaskan	Sangat Memuaskan	TERLAMBAT
P	<=23	Luar Bogor	Sangat Memuaskan	Sangat Memuaskan	Sangat Memuaskan	Memuaskan	Sangat Memuaskan	TERLAMBAT
P	<=23	Luar Bogor	<i>Cumlaude</i>	Sangat Memuaskan	Sangat Memuaskan	Sangat Memuaskan	Sangat Memuaskan	TEPAT WAKTU

3.2. Model Algoritma C4.5

Untuk dapat membuat pohon keputusan, langkah-langkah klasifikasi data menggunakan algoritma C4.5 yaitu:

1. Siapkan data *training*. Data yang digunakan adalah 234 *sample* data *training*.
2. Hitung jumlah mahasiswa lulus tepat waktu dan lulus terlambat berdasarkan nilai tiap atribut.
3. Hitung nilai *entropy* total dimana jumlah mahasiswa lulus tepat waktu berjumlah 93 dan mahasiswa lulus terlambat berjumlah 141.

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^n - p_i \cdot \log_2 p_i \\
 &= (-93/234) \cdot \log_2 (93/234) + \\
 &\quad (-141/234) \cdot \log_2 (141/234) \\
 &= 0.9694
 \end{aligned}$$

4. Hitung nilai *gain* untuk masing-masing atribut. Kemudian tentukan nilai *gain* tertinggi. Atribut dengan nilai *gain* tertinggi maka atribut tersebut dijadikan sebagai akar. Disini, contoh hitung nilai *gain* untuk atribut jenis kelamin yaitu:

$$\begin{aligned}
 Gain(S,A) &= Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \\
 &= 0.9694 - ((182/234) * 0.95742 + \\
 &\quad (52/234 * 0.99573)) \\
 &= 0.003499
 \end{aligned}$$

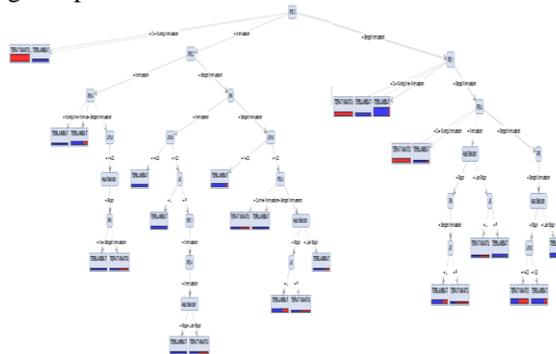
Hasil perhitungan nilai *entropy* dan *gain* semua atribut dilakukan untuk mendapatkan nilai *gain* tertinggi yang akan dijadikan sebagai akar. Hasil perhitungannya dapat dilihat pada tabel berikut:

Tabel 4. Hasil Perhitungan Nilai Entropy dan Gain Algoritma C4.5

Atribut	Entropy	Gain
Total	0.96943	

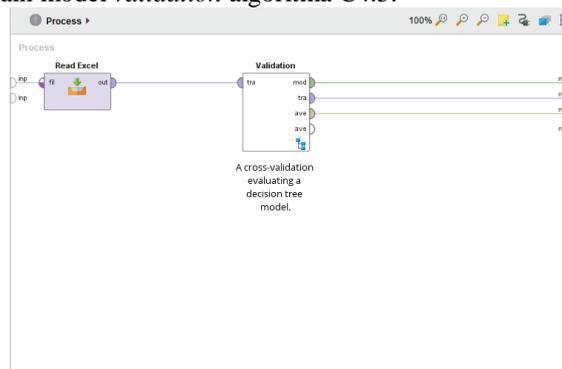
Jenis Kelamin		0.003499
Laki-Laki	0.95742	
Perempuan	0.99573	
Umur		0.056415
<=23	0.99892	
>23	0.80748	
Asal Sekolah		0.010537
Bogor	0.98523	
Luar Bogor	0.86673	
IPS1		0.245373
Kurang Memuaskan	0.00000	
Memuaskan	0.55351	
Sangat Memuaskan	0.98053	
Cumlaude	0.31600	
IPS2		0.229519
Kurang Memuaskan	1.00000	
Memuaskan	0.57135	
Sangat Memuaskan	0.90738	
Cumlaude	0.17926	
IPS3		0.969431
Kurang Memuaskan	0.00000	
Memuaskan	0.68404	
Sangat Memuaskan	0.94849	
Cumlaude	0.00000	
IPS4		0.901562
Kurang Memuaskan	0.32276	
Memuaskan	0.78713	
Sangat Memuaskan	0.97816	
Cumlaude	0.37123	
IPK		0.849831
Memuaskan	0.47435	
Sangat Memuaskan	0.98364	
Cumlaude	0.00000	

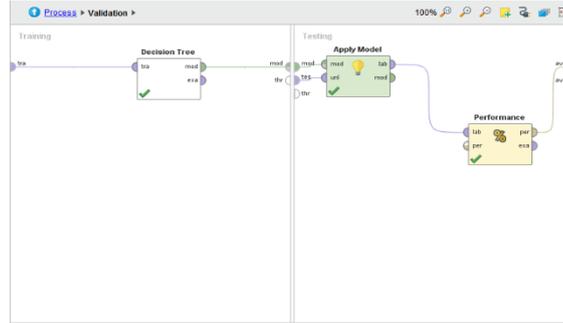
Terlihat bahwa atribut IPS3 memiliki nilai *gain* tertinggi yaitu 0.969431, sehingga atribut IPS3 merupakan simpul akar dari pohon keputusan. Lakukan *entropy* dan *gain* sampai pembentukan akar terakhir. Berikut adalah hasil *rule* yang didapatkan.



Gambar 1. Decision Tree Dengan Algoritma C4.5

Berikut adalah desain model *validation* algoritma C4.5.

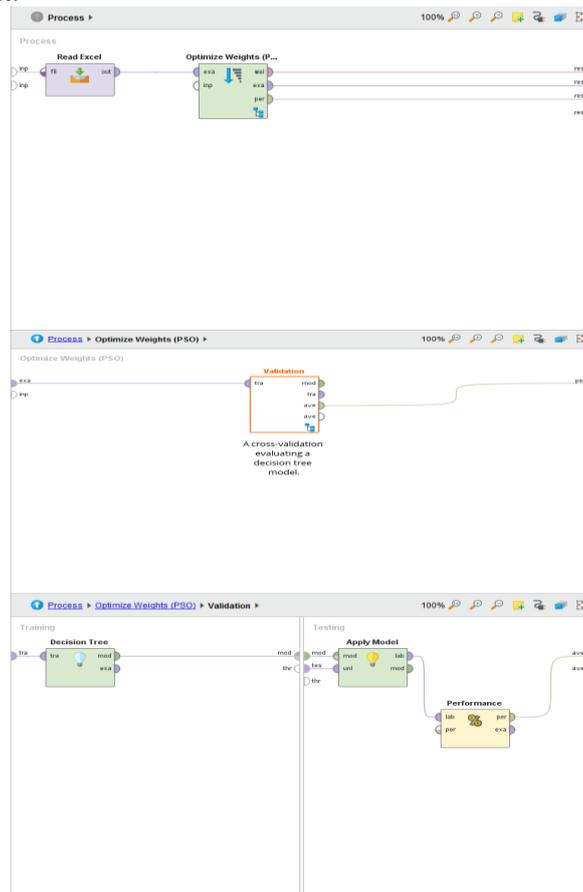




Gambar 2. Model Validation Algoritma C4.5

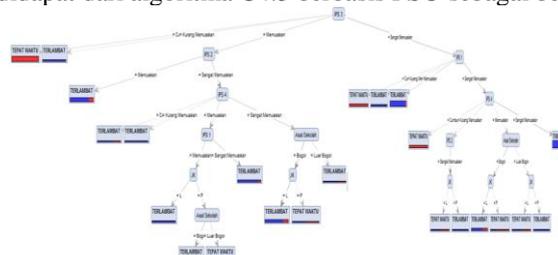
3.3. Model Algoritma C4.5 Berbasis PSO

Menentukan hasil pola keterlambatan mahasiswa menggunakan algoritma C4.5 berbasis PSO pada *RapidMiner* sebagai berikut:



Gambar 3. Model Validation Algoritma C4.5 Berbasis PSO

Adapun *rule* yang didapat dari algoritma C4.5 berbasis PSO sebagai berikut:



Gambar 4. Decision Tree Dengan Algoritma C4.5 Berbasis PSO

Tabel 7 menunjukkan hasil nilai *entropy* dan *gain* semua atribut dan nilai *gain* tertinggi yang akan menjadi simpul akar pohon keputusan.

Tabel 5. Hasil Perhitungan Nilai Entropy dan Gain Algoritma C4.5 Berbasis PSO

Atribut	Entropy	Gain
Total	0.96943	
Jenis Kelamin		0.003499
Laki-Laki	0.95742	
Perempuan	0.99573	
Asal Sekolah		0.010537
Bogor	0.98523	
Luar Bogor	0.86673	
IPS1		0.245373
Kurang Memuaskan	0.00000	
Memuaskan	0.55351	
Sangat Memuaskan	0.98053	
Cumlaude	0.31600	
IPS2		0.229519
Kurang Memuaskan	1.00000	
Memuaskan	0.57135	
Sangat Memuaskan	0.90738	
Cumlaude	0.17926	
IPS3		0.969431
Kurang Memuaskan	0.00000	
Memuaskan	0.68404	
Sangat Memuaskan	0.94849	
Cumlaude	0.00000	
IPS4		0.901562
Kurang Memuaskan	0.32276	
Memuaskan	0.78713	
Sangat Memuaskan	0.97816	
Cumlaude	0.37123	

Terlihat atribut IPS3 mempunyai nilai *gain* tertinggi yaitu 0.969431 sehingga atribut IPS3 merupakan simpul akar pada pohon keputusan. Terjadi pembobotan dengan algoritma optimasi PSO terhadap atribut yang digunakan di dalam *dataset*. Bertujuan untuk mengetahui atribut terbaik dalam pengolahan data. Pembobotan atribut diperlukan karena tidak semua atribut berpengaruh terhadap hasil akurasi. Hasil dari pembobotan atribut disajikan pada Tabel 8.

Tabel 6. Hasil Pembobotan Atribut

atribute	weight
JK	1
Umur	0
Asal Sekolah	1
IPS 1	1
IPS 2	0.397
IPS 3	0.779
IPS 4	1
PK	0

Terdapat 8 atribut yang digunakan dan hasil dari pembobotan atribut yaitu 2 atribut dengan bobot bernilai 0, 4 atribut bobotnya bernilai 1, 1 atribut memiliki bobot 0,397 dan 1 atribut berbobot 0,779. Atribut yang berbobot 0 atau tidak berpengaruh terhadap akurasi yaitu umur dan ipk, yang jika dihilangkan dimungkinkan tidak akan mempengaruhi hasil akurasi.

3.4. Evaluasi dan Validasi Model

1. Pengujian Model Algoritma C4.5

Pengujian ditunjukkan pada tabel *confusion matrix* menggunakan *data training* dengan model algoritma C4.5. Hasil pemodelan yang telah diproses menggunakan *tools RapidMiner* sebagai berikut:

Tabel 7. Nilai Accuracy Algoritma C4.5

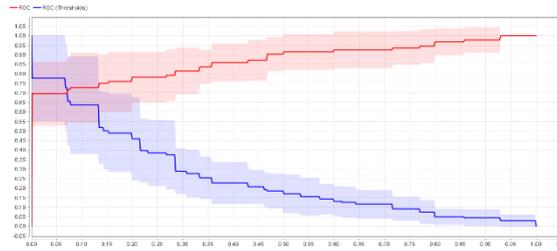
accuracy: 82.16% +/- 7.02% (min: 62.05%)

	true TERLAMBAT	true TEPAT WAKTU	class precision
pred. TERLAMBAT	125	26	82.76%
pred. TEPAT WAKTU	16	67	80.72%
class recall	88.65%	72.04%	

$$Accuracy: \frac{TP + TN}{TP + TN + FP + FN} = \frac{67 + 125}{67 + 125 + 16 + 26}$$

$$= \frac{192}{234}$$

$$= 0,8205 = \mathbf{82,05\%}$$



Gambar 5. Nilai AUC dalam Grafik ROC Algoritma C4.5

Hasil yang didapat dari pengolahan ROC untuk algoritma C4.5 dengan menggunakan data *training* yaitu sebesar 0.870 dengan tingkat diagnosa *Good Classification*.

2. Pengujian Model Algoritma C4.5 Berbasis PSO

Pertama, dilakukan uji coba C4.5 berbasis PSO dengan memberi nilai pada parameter bernilai *default population size: 5* dan *maximum number of generation: 30*. Nilai *population size* dan *maximum number of generation* pada PSO diubah-ubah untuk meningkatkan kinerja PSO yang berdampak pada peningkatan akurasi, yang terlihat pada Tabel 8.

Tabel 8. Hasil uji coba *population size* dan *maximum number of generation* secara berbeda

Parameter PSO	Accuracy	AUC	Waktu
<i>Default</i>	86.34%	0.876	12s
<i>Posize=10, generate=30</i>	85.53%	0.898	23s
<i>Posize=15, generate=30</i>	86.67%	0.898	1m 13s
<i>Posize=20, generate=30</i>	86,72%	0.903	50s
<i>Posize=10, generate=40</i>	85.85%	0.883	42s
<i>Posize=15, generate=40</i>	86.70%	0.886	1m
<i>Posize=20, generate=40</i>	86.72%	0.913	1m 22s
<i>Posize=5, generate=50</i>	86.68%	0.903	15s
<i>Posize=10, generate=50</i>	85.85%	0.883	38s
<i>Posize=15, generate=50</i>	86.70%	0.886	3m 58s

Berdasarkan pengujian *population size* dan *maximum number of generation*. Didapatkan hasil akurasi tertinggi dan waktu eksekusi terendah terjadi pada saat *population size* bernilai 20 dan *maximum number of generation* bernilai 30 dengan nilai akurasi sebesar 86,72% % dan AUC 0,903 dalam waktu 50 detik.

Tabel 9. Nilai Accuracy Algoritma C4.5 Berbasis PSO

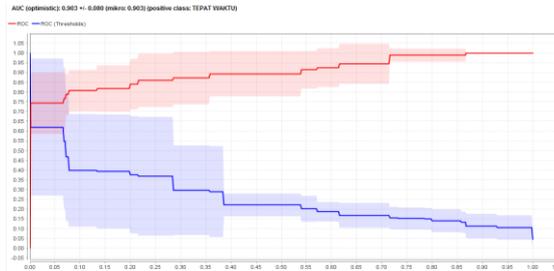
accuracy: 86.72% (+ 4.89% mikro: 86.75%)

	true TERLAMBAT	true TEPAT WAKTU	class precision
pred. TERLAMBAT	141	31	81.58%
pred. TEPAT WAKTU	0	62	100.00%
class recall	100.00%	66.67%	

$$\text{Accuracy: } \frac{TN + TP}{TN + FN + TP + FP} = \frac{141 + 62}{141 + 31 + 62 + 0}$$

$$= \frac{203}{234}$$

$$= 0,8675 = \mathbf{86,75\%}$$



Gambar 6. Nilai AUC dalam Grafik ROC Algoritma C4.5 Berbasis PSO

Hasil grafik ROC dengan nilai AUC (*Area Under Curve*) yaitu sebesar 0,903 dengan tingkat diagnosa *Excellent Classification*.

Berdasarkan pengujian model algoritma yang telah dilakukan, menunjukkan bahwa algoritma C4.5 dengan *particle swarm optimization* memiliki nilai akurasi lebih tinggi dibandingkan hanya menggunakan algoritma C4.5. Nilai akurasi untuk model algoritma C4.5 sebesar 82,05% dan nilai akurasi algoritma C4.5 berbasis PSO sebesar 86,72%, perbandingan hasil akurasi dan performa AUC dapat dilihat pada Tabel 9.

Tabel 10. Pengujian Algoritma C4.5 dan C4.5 Berbasis PSO

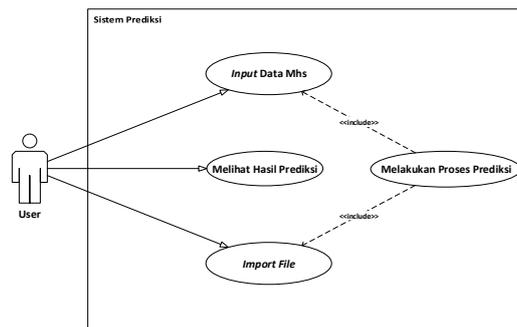
	C4.5	C4.5+PSO
<i>Accuracy</i>	82,05%	86,72%
AUC	0,870	0,903
<i>Execution time</i>	3 s	50 s

Diartikan bahwa algoritma C4.5 berbasis PSO memiliki nilai akurasi lebih baik dalam menentukan pola ketepatan kelulusan mahasiswa dilihat dari nilai akurasi sebesar 86,72% dan nilai nilai AUC 0,903 dengan tingkat diagnosa *Excellent Classification*.

3.5. Penerapan Algoritma

Rule hasil dari algoritma C4.5 dengan optimasi *particle swarm optimization* diterapkan kedalam pembuatan aplikasi untuk prediksi kelulusan mahasiswa dengan menggunakan bahasa pemrograman PHP dan MySQL sebagai *datasenyanya*.

1. Use Case



Gambar 7. Use Case Diagram

3.6. Design Graphical User Interface (GUI)

1. Layar Utama



Gambar 8. Tampilan Layar Utama

2. Layar Menu *Single Data (Input Data Mahasiswa)*

Gambar 9. Tampilan Layar Prediksi Berdasarkan NPM

3. Layar Hasil Prediksi *Single Data*

Gambar 10. Tampilan Hasil Prediksi

4. Layar Layar Menu *Multiple Data (Import File)*

Gambar 11. Tampilan Import File Data Mahasiswa

5. Layar Hasil Prediksi *Multiple Data*

No	NPM	IPS 1	IPS 2	IPS 3	IPS 4	Prediksi
1	005112303	3.41	2.65	3	2.75	Terlambat
2	005112306	2.78	3.09	2.93	2.71	Terlambat
3	005112308	3.35	3.77	3.73	3.35	Tepat Waktu
4	005112309	3.69	2.68	3	3.42	Tepat Waktu
5	005112312	3.69	3.75	3.82	3.5	Tepat Waktu
6	005112314	3.23	3.2	3.45	3.29	Terlambat
7	005112316	3.12	3	2.45	3.12	Terlambat
8	005112319	3.62	3.88	3.73	3.43	Tepat Waktu
9	005112321	2.91	3.36	3	3.57	Tepat Waktu
10	005112322	2.26	2.77	1.45	3.25	Terlambat
11	005112323	3.19	3.59	3.8	3.25	Tepat Waktu
12	005112331	3.38	3.44	3.14	3.43	Terlambat
13	005112332	3.35	3.3	3.1	3.35	Terlambat
14	005112342	3.43	3.7	3.73	3	Tepat Waktu
15	005112348	2.15	2.55	2.45	2.75	Terlambat

Gambar 12. Tampilan Import File Data Mahasiswa

Dilakukan perhitungan untuk mendapatkan nilai akurasi dari aplikasi dengan menggunakan data uji (*data testing*) yang dimasukkan ke dalam aplikasi. Maka nilai akurasi yang diperoleh dari aplikasi ini adalah:

$$\begin{aligned} \text{Accuracy:} \\ \frac{\text{Jumlah Pengujian Benar}}{\text{Jumlah Pengujian}} &= \frac{53}{60} = 0,8833 \\ &= \mathbf{88,33\%} \end{aligned}$$

$$\begin{aligned} \text{Error rate:} \\ \frac{\text{Jumlah Pengujian Salah}}{\text{Jumlah Pengujian}} &= \frac{7}{60} = 0,1167 \end{aligned}$$

=11,67%

4. PENUTUP

Kesimpulan yang dihasilkan berdasarkan penelitian ini sebagai berikut:

1. Algoritma C4.5 digunakan untuk membentuk klasifikasi kelulusan mahasiswa kemudian dioptimasi dengan algoritma *particle swarm optimization* (PSO) dalam segi seleksi atribut. Pola yang didapat dari algoritma C4.5 dengan *particle swarm optimization* (PSO) adalah ips 3 merupakan atribut yang paling mempengaruhi ketepatan kelulusan mahasiswa.
2. Validasi model menggunakan *k-fold cross-validation*. Evaluasi model menggunakan *confusion matrix* dan kurva ROC. Berdasarkan hasil evaluasi dan validasi, didapatkan bahwa algoritma C4.5 dengan optimasi *particle swarm optimization* (PSO) memiliki nilai akurasi sebesar 86,72% lebih baik jika dibandingkan dengan algoritma C4.5 yang akurasinya 82,05% sehingga didapatkan selisih diantaranya 4,67%. Didapatkan selisih nilai AUC 0,033 yaitu model algoritma C4.5 memiliki nilai AUC 0,870 dengan tingkat diagnosa *good classification* dan algoritma C4.5 dengan PSO memiliki nilai AUC sebesar 0,903 dengan tingkat diagnosa *excellent classification*.
3. *Rule* hasil dari algoritma C4.5 dengan *particle swarm optimization* diterapkan kedalam pembuatan aplikasi untuk prediksi kelulusan mahasiswa berbasis *graphical user interface* (GUI). Berdasarkan hasil pengujian menggunakan *black box testing* bahwasanya sistem diterima pengguna.

DAFTAR PUSTAKA

- [1] Badan Akreditasi Nasional Perguruan Tinggi, "Matriks Penilaian Instrumen Akreditasi Program Studi Sarjana," Jakarta, p. 2013, 2008.
- [2] M. A. H. Ian H. Witten, Eibe Frank, *Data Mining Practical Machine Learning Tools and Techniques*. Burlington: Elsevier Inc, 2011.
- [3] S. Mashlahah, M. A. Yaqin, and M. Faisal, "Prediction of Students Graduation Using Decision Tree Method with the Implementation of Algorithm C4.5," *Int. J. Sci. Technol.*, vol. 2, no. 2, pp. 1–8, 2013.
- [4] S. H. David Hartanto Kamagi, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *ULTIMATICS*, vol. VI, no. 1, pp. 15–20, 2014, doi: <https://doi.org/10.31937/ti.v6i1.327>.
- [5] W. W. W. dan M. R. A. Dewi Kusumawati, "Prediksi Kelulusan Mahasiswa Menggunakan Metode Neural Network Dan Particle Swarm Optimization," *Semin. Nas. Teknol. Inf. dan Multimed.*, pp. 37–40, 2015.
- [6] R. Wajhillah, S. Nusa, M. Sukabumi, and N. B. View, "Optimasi Algoritma Klasifikasi C4.5 Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Jantung," *SWABUMI*, vol. 1, no. September, 2014.
- [7] E. Nurlelah and M. S. Mardiyanto, "Pemilihan Atribut Pada Algoritma C4.5 Menggunakan Particle Swarm Optimization Untuk Meningkatkan Akurasi Prediksi Diagnosis Penyakit Liver," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 195–202, 2019, doi: 10.33480/pilar.v15i2.706.
- [8] F. Santoso, A. Syukur, and A. Z. Fanani, "Algoritma C4.5 Dengan Particle Swarm Optimization Untuk Klasifikasi Lama Menghafal Al-Quran Pada Santri Mahadul Quran," *J. Teknol. Inf.*, vol. 14, pp. 92–103, 2018.
- [9] C. R. Kothari, *Research Methodology Methods and Techniques*. New Delhi: New Age International, 2004.
- [10] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.