



Algoritma C4.5 Berbasis Particle Swarm Optimization untuk Pola Klasifikasi Penerimaan Siswa Jurusan Teknik Komputer Jaringan

Lina Listiani¹, Awit Marwati Sakinah², Nova Agustina³

¹Manajemen Informatika, STMIK DCI, Jawa Barat, Indonesia

²Teknik Informatika, STMIK DCI, Jawa Barat, Indonesia

Correspondence: E-mail: linalistiani20@gmail.com

ABSTRACT

Sekolah Menengah Kejuruan merupakan jenjang pendidikan formal yang melatih siswa untuk mempelajari keterampilan khusus sebelum memasuki dunia kerja. Pemilihan jurusan yang tidak berdasarkan dengan minat dan bakat mempengaruhi proses belajar mengajar, dan berpengaruh terhadap kemampuan siswa ketika lulus. Teknik komputer jaringan merupakan jurusan yang paling banyak diminati. Proses seleksi masih dilakukan manual yaitu dengan mengolah rata-rata nilai raport dan hasil tes seleksi sehingga kurang efektif. Data PPDB yang dimiliki oleh pihak sekolah dapat digunakan untuk menggali informasi yang berguna menemukan pola klasifikasi penentuan jurusan siswa menggunakan teknik data mining. Teknik yang digunakan adalah klasifikasi menerapkan algoritma C4.5 berbasis Particle Swarm Optimization (PSO) menggunakan data PPDB sebanyak 701 record. Atribut yang digunakan antara lain nomor peserta, nilai raport bahasa Inggris, bahasa Indonesia, matematika, IPA, tes olahraga, tes kesehatan dan hasil. Penerapan algoritma C4.5 yang dioptimalkan dengan PSO terbukti meningkatkan akurasi sebesar 0,86%, terlihat dari perbandingan akurasi penerapan algoritma C4.5 menghasilkan akurasi sebesar 94,01%, dan apabila disandingkan PSO menghasilkan akurasi sebesar 94,87%.

ARTICLE INFO

Article History:

Submitted/Received 26 Juni 2025

First Revised 27 Juni 2025

Accepted 28 Juni 2025

First Available online 1 Juli 2025

Publication Date 1 Juli 2025

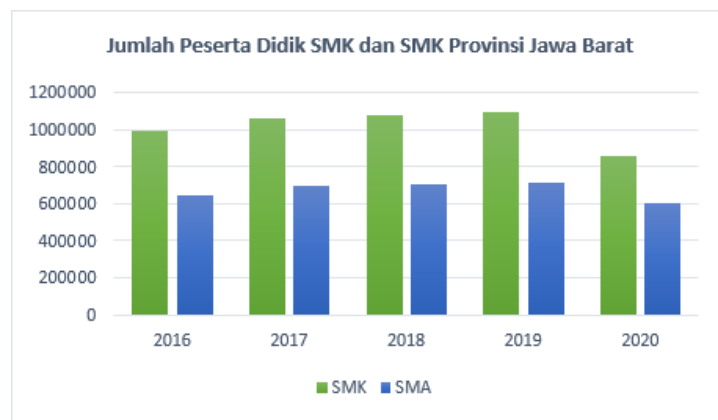
Keyword:

Algoritma C4.5,
Particle Swarm Optimization,
Pola Klasifikasi,
Sekolah Menengah Kejuruan.

1. INTRODUCTION

Sekolah Menengah Kejuruan merupakan jenjang pendidikan formal yang menerapkan sistem pembelajaran dengan melatih siswa untuk mempelajari ketrampilan khusus berdasarkan bidang konsentrasinya yang bertujuan agar siap memasuki dunia kerja dan dapat melanjutkan Pendidikan sesuai dengan bidang konsentrasinya (Rizmayanti, dkk. 2021). Berdasarkan Undang-Undang Sisdiknas No.20 Tahun 2003, tertuang tujuan pendidikan formal adalah berupaya melakukan pengembangan melalui proses belajar mengajar untuk menggali potensi siswa yang bertujuan untuk melatih keterampilan dan mengembangkan pengetahuan (Pemerintah Republik Indonesia, 2023).

Tingginya minat dari lulusan Sekolah Menengah Pertama untuk melanjutkan jenjang Pendidikan di Sekolah Menengah Kejuruan dapat dilihat pada portal data peserta didik nasional yaitu dapo.dikdasmen.kemendikbud.go.id Provinsi Jawa Barat pada **Gambar 1** berikut.



Gambar 1. Grafik Perbandingan Siswa yang Memilih SMA dan SMK

Memasuki tahun ajaran baru dilakukan proses penerimaan calon siswa untuk diseleksi. Calon siswa menentukan jurusan untuk dipilih disesuaikan dengan minat serta bakat yang dimiliki. Terbatasnya jumlah siswa yang diterima maka dilakukan tes seleksi untuk menentukan siswa yang memenuhi standar untuk diterima. Seleksi tidak hanya menggunakan nilai raport saja tetapi calon siswa harus mengikuti serangkaian tes diantaranya tes kesehatan dan tes olahraga. Hasil tes seleksi dan rata-rata nilai raport digunakan untuk menentukan jurusan siswa. Calon memilih dua jurusan ketika mendaftar, apabila nilai memenuhi akan diterima dijurusan pilihan pertama, apabila tidak memenuhi akan disarankan masuk jurusan kedua.

Terdapat beberapa pilihan jurusan yang diada di sekolah menengah kejuruan. Siswa dapat memilih jurusan yang ada sesuai dengan keahlian yang dimiliki. Permasalahan yang terjadi adalah siswa cenderung tidak mengetahui bidang keahlian yang dimiliki, sehingga pengambilan Keputusan untuk jurusan yang diambil berdasarkan jurusan yang paling banyak diminati disekolah tersebut. Dampak yang ditimbulkan adalah pada prestasi siswa dan proses belajar mengajar karena siswa akan kesulitan mengikuti mata Pelajaran yang kurang dikuasai sehingga akan berpengaruh terhadap prestasi yang dimiliki siswa.

Perkembangan teknologi yang semakin maju dan penggunaan teknologi hampir disemua sektor menjadi salah satu faktor yang mendukung jurusan Teknik Komputer Jaringan (TKJ) menjadi jurusan dengan peminat yang tinggi atau favorit dibandingkan dengan jurusan lain. Konsentrasi dari jurusan teknik komputer jaringan adalah mempelajari tentang hardware, perakitan komputer dan instalasi komputer (Sofian, 2016; Suharna, 2023). Penjurusan siswa

berpedoman pada hasil seleksi dari ujian penjurusan. Pengetahuan, ketrampilan yang dimiliki oleh siswa ketika lulus serta untuk melanjutkan ke jenjang pendidikan lebih lanjut (Roghib, 2024). Fakta yang terjadi dimasyarakat pemilihan jurusan tidak berdasarkan minat dan bakat tetapi berdasarkan jurusan yang paling diminati atau favorit. Aktualnya pemilihan jurusan siswa sebaiknya sesuai dengan minat serta bakat yang dimiliki oleh siswa karena akan berpengaruh terhadap kemampuan atau ketrampilan siswa ketika lulus. Selain itu pemilihan jurusan yang tidak sesuai akan berpengaruh terhadap proses belajar mengajar karena siswa siswa tidak menguasai dasar materi yang diajarkan.

Sistem seleksi calon siswa yang diterapkan oleh pihak sekolah kurang efektif dan melewati banyak proses maka dibutuhkan sistem baru yang memudahkan proses penentuan jurusan siswa. Data PPDB yang dimiliki oleh pihak sekolah dapat digunakan untuk menggali informasi yang bermanfaat pada proses penentuan jurusan menggunakan salah satu teknik data mining. Data PPDB diklasifikasikan menggunakan algoritma C4.5 menghasilkan suatu pola atau aturan yang digunakan untuk menentukan siswa yang diterima di jurusan TKJ.

Algoritma C4.5 merupakan Teknik pada data mining yang digunakan untuk mengklasifikasikan atau mengelompokkan secara prediktif. Algoritma C4.5 mampu mengklasifikasikan data numerik maupun kategorikal. Algoritma C4.5 akan dioptimalkan dengan algoritma PSO untuk pembobotan atribut karena tidak semua atribut pada suatu data berpengaruh dan penggunaan atribut yang terlalu banyak akan berpengaruh terhadap waktu yang dibutuhkan untuk proses seleksi, oleh karena itu diterapkan Particle Swarm Optimization (PSO) merupakan metode optimasi yang pada implementasinya diterapkan untuk menangani masalah optimasi dan seleksi fitur. Penelitian ini menggunakan PSO untuk pembobotan dan seleksi fitur untuk meningkatkan nilai akurasi yang dihasilkan. Penerapan PSO menghasilkan bobot masing-masing atribut sehingga diketahui atribut yang berpengaruh.

2. METHODS

2.1. Pengumpulan Data

Penelitian menggunakan data sebanyak 701 record menggunakan 8 field sebagai atribut yang berpengaruh. Atribut yang digunakan terdiri dari nomor peserta, jenis kelamin, nilai raport bahasa inggris, bahasa Indonesia, matematika, IPA, tes olahraga, tes kesehatan dan hasil. Atribut merupakan target yang menentukan apakah calon siswa diterima atau tidak di jurusan TKJ berdasarkan nilai raport, tes olahraga dan tes kesehatan.

2.2 Eksperimen

Eksperimen menggunakan tool rapidminer dengan mengujikan dataset yang telah dibuat sebelumnya, kemudian dipecah menjadi dua bagian 80% data training dan 20% data testing. Dataset akan diujikan menggunakan metode yang telah ditentukan yaitu algoritma C4.5 yang dioptimasi dengan seleksi fitur yaitu Particle Swarm Optimization (PSO).

Eksperimen menggunakan tool rapidminer dengan mengujikan dataset yang telah dibuat sebelumnya dengan membagi dataset ke dalam dua bagian yaitu 80% menjadi data training dan 20% menjadi data testing. Dataset akan diujikan pada metode yang telah ditentukan yaitu algoritma C4.5 yang dioptimasi dengan seleksi fitur yaitu Particle Swarm Optimization (PSO).

2.3 Evaluasi

Pada tahap ini dataset yang dihasilkan dari tahap data preparation akan diujikan pada model yang digunakan. Evaluasi terhadap penerapan metode akan diketahui apakah model

yang dihasilkan mampu menyelesaikan permasalahan dipecahkan berupa nilai akurasi menggunakan model confusion matrix dan hasil klasifikasi divisualisasikan pada kurva ROC (Receiver Operating Characteristic). Setelah dihasilkan sebuah model pada tahap evaluasi maka ditentukan bagaimana model akan digunakan.

2.4. Algoritma C4.5

J. Ross Quiland pertama kali memperkenalkan Algoritma C4.5 pada tahun 1996 yang dikembangkan dari metode sebelumnya yaitu Algoritma ID3. Perbaikan yang dilakukan, Algoritma C4.5 mampu menangani tipe data numerik dengan melakukan pemotongan (*pruning*), decision tree, serta penurunan (*deriving*) rule set. Persamaan Algoritma C4.5 dan ID3, penentuan fitur yang digunakan untuk membagi node dengan menggunakan kriteria gain pada pohon yang akan diinduksi. Pemodelan akan menghasilkan output dalam bentuk pohon keputusan dari eksplorasi data (Hidayatulloh, 2022). Kelebihan Algoritma C4.5 adalah bentuk visualisasi dalam bentuk pohon keputusan sehingga hasil klasifikasi mudah diamati dan dimengerti (Larose, 2005). Decision Tree terbentuk menyerupai pohon yang terbalik yaitu akar (*root*) yang terletak diatas dan daun (*leaf*) yang terletak dibagian bawah. Dalam pembentukan pohon keputusan dengan langkah-langkah sebagai berikut:

Langkah 1. Mempersiapkan data training.

Langkah 2. Menentukan akar (*root*) pada pohon. Akar yang terbentuk pada pohon keputusan diperoleh dengan menghitung nilai entropy pada setiap atribut menggunakan persamaan:

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

Di mana S merupakan himpunan kasus, A merupakan atribut/fitur, N adalah jumlah partisi S, dan P_i merupakan proporsi dari S_i terhadap S.

Langkah 3. Setelah didapatkan nilai *entropy* untuk setiap atribut yang terpilih, selanjutnya dilakukan perhitungan nilai information gain dari masing-masing atribut. Pada pohon keputusan penentuan akar dipilih berdasarkan nilai information gain tertinggi. Persamaan yang digunakan adalah:

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

Di mana S merupakan himpunan kasus, A merupakan fitur, n merupakan jumlah partisi atribut A, $|S_i|$ adalah proporsi S_i terhadap S, $|S|$ adalah jumlah kasus dalam S.

Langkah 4. Ulangi Langkah 2 sampai semua record terpartisi.

Langkah 5. Proses partisi pada pohon keputusan akan berhenti apabila:

- Semua *record* pada *node* N dalam kelas yang sama
- Tidak terdapat atribut dalam *record* untuk dilakukan dipartisi lagi
- Tidak terdapat *record* di cabang yang kosong

2.5. Particle Swarm Optimization (PSO)

Particle Swarm Optimization adalah algoritma berbasis populasi yang diperkenalkan oleh dokter Kennedy dan Eberhart yang mengambil gagasan berdasarkan perilaku sekawanan

burung atau kelompok. Untuk menemukan solusi terbaik dilakukan secara acak pada ruang pencarian dalam menyelesaikan masalah optimasi. Setiap individu digambarkan sebagai particle bergerak dalam ruang pencarian dengan mengikuti posisi terbaik yang telah dicapai oleh individu (Abraham, 2006). Persamaan yang digunakan untuk menghitung perpindahan posisi dan kecepatan partikel adalah:

$$V_i = V_i(t-1) + C_1 r_1 [X_{pbest_i} - x_i(t)] + C_2 r_2 [X_{gbest} - x_i(t)] \quad (3)$$

$$X_i(t) = X_i(t-1) + V_i(t)$$

Di mana $V_i(t)$ adalah kecepatan partikel i saat iterasi t , $X_i(t)$ adalah posisi partikel i saat iterasi t , $c_1 c_2$ adalah *learning rates* untuk kemampuan individu (*cognitive*) dan pengaruh social (grup), $r_1 r_2$ adalah bilangan random yang berdistribusi uninormal dalam interval 0 dan 1, X_{pbest_i} adalah posisi terbaik partikel i , dan X_{gbest} adalah posisi terbaik global.

2.6. Confusion Matrix

Sebuah matrik prediksi yang digunakan untuk membandingkan antara kelas inputan dengan kelas sebenarnya atau matrik yang dihasilkan dari proses klasifikasi yang berisi mengenai nilai aktual dengan nilai prediksi (Maurina, 2011). *Confusion Matrix* digunakan untuk mengevaluasi kinerja dari metode yang diterapkan berdasarkan perhitungan hasil prediksi benar dan tidak benar baik masing-masing kelas positif serta negatif (Bramer, 2007). Di dalam *Confusion matrix* terdapat empat kelas yaitu kelas positif dan negatif yang terdiri dari empat komponen utama yaitu True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN). Evaluasi terhadap kinerja dalam melakukan klasifikasi menghasilkan keluaran yaitu akurasi, recall, presisi.

Akurasi dihitung dengan persamaan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Presisi dihitung dengan persamaan:

$$Precision = \frac{TP}{TP + FP}$$

Recall dihitung dengan persamaan:

$$Recall = \frac{TP}{TP + FN}$$

Adapun confusion matrik untuk klasifikasi kelas diperlihatkan pada **Tabel 1** sebagai berikut.

Tabel 1. Confusion Matrix

		<i>Predicated Class</i>	
		Yes	No
Actual Class	Yes	True Positive (TP)	False Negative (FN)
	No	False Positive (FP)	True Negative (TN)

2.7. Cross Validation

Cross Validation merupakan teknik validasi yang dilakukan dengan melakukan pembagian data kedalam jumlah k dan setiap bagian akan melewati proses klasifikasi. Pengambilan sample data secara random atau membagi data dengan jumlah yang sama dengan jumlah partisi yang ditentukan (Hand, 2007). Penentuan jumlah data atau partisi untuk mendapatkan estimasi kesalahan terbaik adalah 10-fold validation atau ten-fold validation. Data akan dibagi menjadi jumlah yang sama yang kemudian error rate dihitung pada setiap bagian, kemudian rata-rata rate error dihitung dari 10 bagian yang ditentukan.

3. RESULTS AND DISCUSSION

3.1. Pemodelan Algoritma C4.5

Tahap modeling disebut juga dengan tahap learning, data PPDB pada tahap ini akan diklasifikasikan menggunakan metode yang telah dipilih untuk menemukan pola atau aturan. Langkah awal pada pembentukan pohon keputusan dilakukan dengan menghitung jumlah class yang terdiri dari class DITERIMA dan class TIDAK DITERIMA dari keseluruhan data. Selanjutnya dihitung nilai entropy dari setiap class berdasarkan atribut yang digunakan. Setelah diketahui nilai entropy dari setiap kasus, Langkah selanjutnya adalah dengan melakukan perhitungan dari nilai information gain untuk setiap atribut yang nantinya digunakan untuk menentukan node yang akan dipecah. Nilai information gain terbesar akan digunakan sebagai akar (node 1). Adapun nilai entropy total dihitung dengan persamaan 1.

$$Entropy (Total) = \left(- \left(\frac{97}{701} \right) \times \log_2 \left(\frac{97}{701} \right) \right) + \left(- \left(\frac{604}{701} \right) \times \log_2 \left(\frac{604}{701} \right) \right)$$

$$Entropy (Total) = 0,394829823 + 0,185134095$$

$$Entropy (Total) = 0,579963918$$

Menghitung nilai entropy masing-masing kategori dari atribut Bahasa inggris menggunakan persamaan 1.

$$Entropy (A) = \left(- \left(\frac{15}{109} \right) \times \log_2 \left(\frac{15}{109} \right) \right) + \left(- \left(\frac{94}{109} \right) \times \log_2 \left(\frac{94}{109} \right) \right) = 0,57795618$$

$$Entropy (B) = \left(- \left(\frac{69}{535} \right) \times \log_2 \left(\frac{69}{535} \right) \right) + \left(- \left(\frac{466}{535} \right) \times \log_2 \left(\frac{466}{535} \right) \right) = 0,554612033$$

$$Entropy (C) = \left(- \left(\frac{12}{55} \right) \times \log_2 \left(\frac{12}{55} \right) \right) + \left(- \left(\frac{43}{55} \right) \times \log_2 \left(\frac{43}{55} \right) \right) = 0,756833632$$

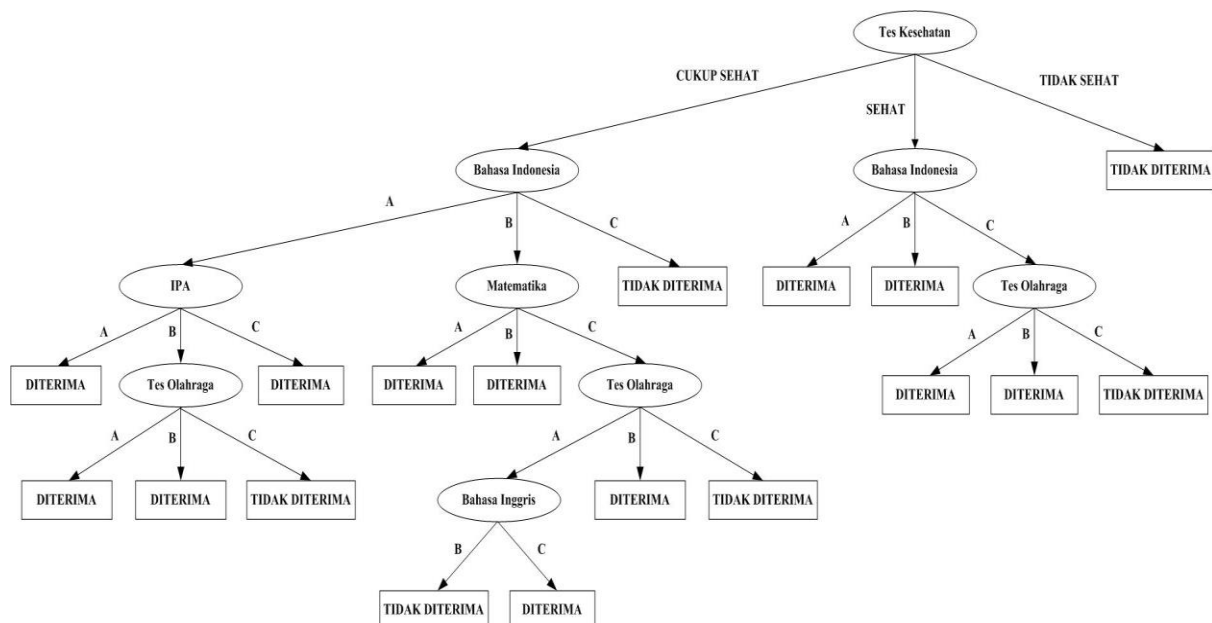
$$Entropy (D) = \left(- \left(\frac{1}{2} \right) \times \log_2 \left(\frac{1}{2} \right) \right) + \left(- \left(\frac{1}{2} \right) \times \log_2 \left(\frac{1}{2} \right) \right) = 0,756833632$$

Perhitungan entropy dan information gain pada node 1 menghasilkan data yang ditunjukkan pada Tabel 2 berikut.

Node	Atribut Total	Entropy	Information Gain
1	Bahasa Inggris	0,579964	
	A	0,577958	0,004585
	B	0,554612	
	C	0,756834	

Node	Atribut	Entropy	Information Gain
D		1,000000	
	Bahasa Indonesia		
A		0,532136	0,016533
B		0,552367	
C		0,998636	
D		0,000000	
	Matematika		
A		0,503258	0,008762
B		0,526910	
C		0,743170	
D		0,918296	
	IPA		
A		0,517583	0,003133
B		0,566190	
C		0,733538	
D		0,000000	
	Tes Kesehatan		
SEHAT			0,326028
		0,174135	(Information gain terbesar yang akan menjadi akar 1)
CUKUP SEHAT		0,979869	
TIDAK SEHAT		0,135036	
	Tes Olahraga		0,033443
A		0,423461	
B		0,480627	
C		0,850382	
D		0,918296	

Adapun pohon keputusan yang terbentuk berdasarkan hasil perhitungan diperlihatkan pada **Gambar 2** berikut.



Gambar 2. Pohon Keputusan Hasil Penghitungan Algoritma C4.5

3.2. Evaluasi dan Validasi

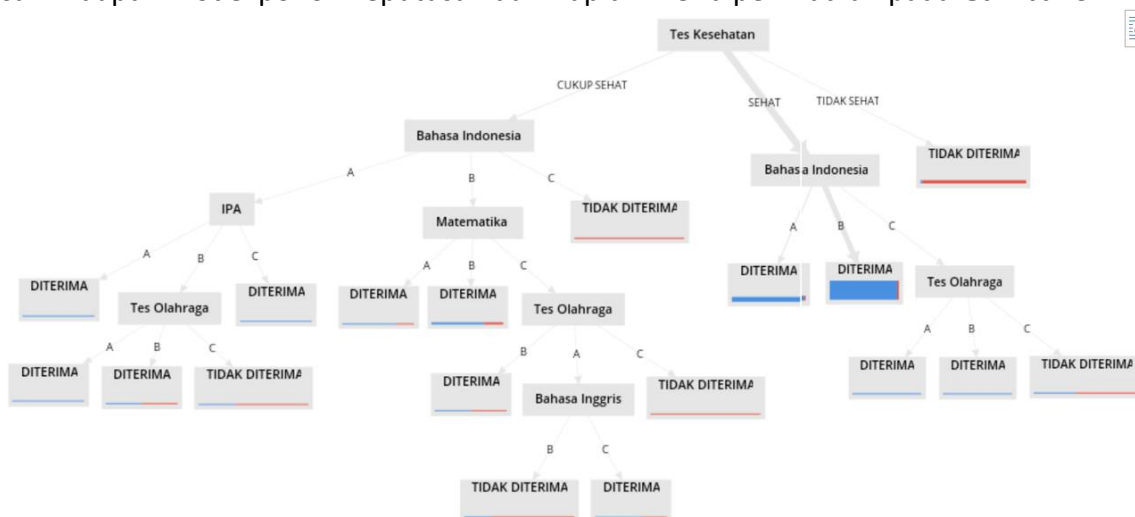
Pengujian metode menggunakan Framework Rapidminer 9.7 dengan mengatur konfigurasi dan penggunaan parameter yang mempengaruhi hasil akurasi dan penerapan metode. Konfigurasi Algoritma C4.5 pada rapidminer ditunjukkan pada **Tabel 3** sebagai berikut:

Tabel 3. Konfigurasi Parameter *Decision Tree* pada *Rapidminer*

Parameter	Algoritma C4.5
<i>Criteria</i>	Information Gain
<i>Maximal Depth</i>	10
<i>Minimal Gain</i>	0,01
<i>Minimal Leaf Size</i>	2
<i>Minimal Size for Split</i>	4
<i>Number of Prepruning</i>	3

Pada pengaturan konfigurasi kriteria diganti *information gain*, proses perhitungan *entropy* dan *information gain* harus memenuhi minimal kriteria gain 0,01 perhitungan tidak akan dilanjutkan apabila hasil perhitungan gain dibawah 0,01 maka tidak dilakukan perhitungan untuk *node* selanjutnya. Proses *node* harus mempunyai 2 *record* untuk percabangan dengan minimal 1 record pada setiap percabangan.

Validasi dan pengukuran akurasi pada penerapan algoritma C4.5 yang menggunakan *framework rapidminer*. Pengujian dilakukan menggunakan data yang sudah melewati tahap *preprocessing* untuk melihat hasil akurasi. Validasi model menerapkan *cross validation* dengan *tenfold validation* (10 partisi) dan *confusion matrix* digunakan pada evaluasi model yang divisualisasikan dalam kurva ROC. Pembentukan *rule* dari algoritma C4.5 pada *rapidminer* menggunakan *cross validation* menghasilkan pohon keputusan, akurasi, *presisi*, *recall*. Adapun model pohon keputusan dari Rapidminer diperlihatkan pada **Gambar 3**.



Gambar 3. Model Pohon Keputusan yang Dihasilkan Rapidminer

3.3. Algoritma C4.5 berbasis *Particle Swarm Optimization* (PSO)

Algoritma C4.5 dioptimalkan dengan *Particle Swarm Optimization* (PSO) untuk pembobotan atribut serta meningkatkan akurasi yang dihasilkan. Berikut merupakan konfigurasi algoritma PSO pada *Rapidminer* yang diperlihatkan pada **Tabel 4**.

Tabel 4. Konfigurasi Algoritma PSO pada *Rapidminer*

Parameter	Algoritma PSO
<i>Population Size</i>	5
<i>Inertia Weight</i>	1,0
<i>Maximum Number of Generation</i>	30
<i>Local Best Weight</i>	1,0
<i>Global Best Weight</i>	1,0
<i>Minimal Weight</i>	0,0
<i>Maximal Weight</i>	1,0

Pembobotan atribut dari penerapan Algoritma PSO menghasilkan bobot setiap atribut sehingga diketahui atribut yang sangat berpengaruh dan tidak memiliki berpengaruh. Pembobotan atribut perlu dilakukan karena tidak semua atribut yang digunakan memiliki pengaruh yang signifikan terhadap akurasi yang dihasilkan. Nilai bobot yang dihasilkan dari penerapan PSO adalah berkisar 0-1. Nilai 1 menyatakan atribut yang berpengaruh sedangkan nilai 0 menyatakan atribut yang tidak berpengaruh. Hasil pembobotan atribut dari penerapan algoritma *Particle Swarm Optimization* diperlihatkan pada **Tabel 5** berikut.

Tabel 5. Hasil Pembobotan Atribut Metode PSO

Atribut	Bobot
Bahasa Inggris	1
Bahasa Indonesia	1
Matematika	0,343
IPA	1
Tes Olahraga	0
Tes Kesehatan	1

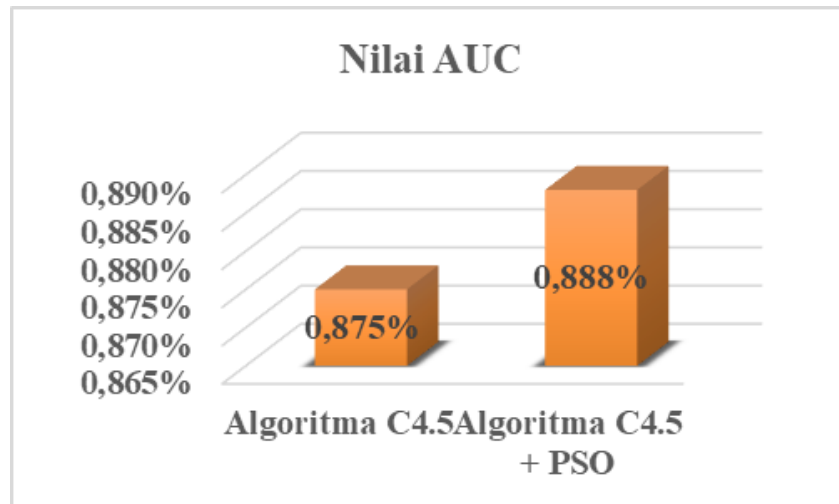
Atribut yang digunakan terdiri dari 7 atribut, terdapat 4 atribut dengan nilai bobot 1 yaitu bahasa inggris, bahasa Indonesia, IPA, dan tes kesehatan. Atribut matematika dengan nilai bobot 0,343 dan terdapat atribut dengan nilai bobot 0 yaitu atribut tes olahraga. Atribut dengan nilai bobot 0 dapat dihilangkan karena kemungkinan tidak berpengaruh terhadap nilai akurasi yang dihasilkan dan tidak berpengaruh terhadap proses klasifikasi. Kemudian atribut dengan nilai bobot 1 memiliki pengaruh yang signifikan.

Perbandingan nilai akurasi yang dihasilkan dari eksperimen dengan menerapkan Algoritma C4.5 dan Algoritma C4.5 PSO menunjukkan adanya peningkatan akurasi sebesar 0,86%. Selain akurasi, evaluasi kinerja dari penerapan algoritma menghasilkan nilai *recall* dan *presisi* yang juga menunjukkan peningkatan. Perbandingan nilai akurasi diperlihatkan pada **Tabel 6** berikut.

Tabel 6. Perbandingan Nilai Akurasi dari Penerapan Algoritma C4.5 dan C4.5 PSO

	C4.5	C4.5 PSO
Accuracy	94,01 %	94,87 %
Recall	98,68 %	98,84 %
Presisi	94,60 %	95,37 %
AUC	0,875 (Good)	0,888 (Good)

Evaluasi dari hasil kinerja algoritma dalam melakukan klasifikasi divisualisasikan dalam kurva ROC (*Receiver Operating Characteristic*) berupa nilai AUC (*Area Under Curve*) pada **Gambar 4**. AUC merupakan wilayah yang ada dibawah kurva dengan luas antara 0 dan 1.



Gambar 4. Grafik Perbandingan Nilai AUC Algoritma C4.5 dan C4.5 PSO

Nilai AUC (*Area Under Curve*) yang dihasilkan dari penerapan Algoritma C4.5 adalah sebesar 0,875, sedangkan nilai AUC (*Area Under Curve*) dari Algoritma C4.5 PSO sebesar 0,888. Nilai AUC yang dihasilkan dikatakan sempurna apabila nilai AUC yang dihasilkan mencapai 1.000 dan akurasi yang dihasilkan dikatakan buruk apabila nilai AUC yang dihasilkan dibawah 0.500. Dapat disimpulkan akurasi yang dihasilkan menunjukkan hasil yang bagus dengan nilai diatas 0.800. Penerapan *Particle Swarm Optimization* (PSO) pada Algoritma C4.5 menunjukkan kenaikan nilai AUC sebesar 0,013.

4. CONCLUSION

Penelitian dengan menerapkan Algoritma C4.5 yang dioptimaskan dengan PSO dapat disimpulkan bahwa:

1. Metode yang digunakan mampu mengklasifikasikan data PPDB dengan tingkat akurasi yang baik yaitu 94.01% untuk Algoritma C4.5 dan 94,87% untuk Algoritma C4.5 berbasis *Particle Swarm Optimization* (PSO). Penerapan Algoritma C4.5 yang optimasikan dengan PSO menunjukkan hasil akurasi yang lebih baik dengan peningkatan akurasi sebesar 0,86%.
2. Penerapan PSO pada Algoritma C4.5 mampu melakukan pembobotan pada setiap atribut yang digunakan. Hasil pembobotan dapat diketahui atribut yang berpengaruh dan tidak berpengaruh. Atribut dengan 0 adalah atribut yang tidak berpengaruh maka akan dihilangkan. Atribut dengan nilai bobot 0 adalah atribut tes olahraga.

3. Hasil klasifikasi berupa pola atau aturan dapat membantu bagi pihak sekolah khususnya panitia PPDB untuk menentukan siswa yang diterima di jurusan TKJ.

Adapun saran untuk penelitian selanjutnya adalah:

1. Pada prosesnya untuk mempermudah proses klasifikasi, sebaiknya pola aturan yang dihasilkan diimplementasikan kedalam sebuah perangkat lunak aplikasi.
2. Penelitian yang dilakukan hanya untuk klasifikasi siswa jurusan Teknik Komputer Jaringan (TKJ), maka penelitian dimasa mendatang sebaiknya untuk klasifikasi dilakukan pada semua jurusan.
3. Untuk tingkat akurasi yang lebih tinggi sebaiknya penelitian selanjutnya menggunakan data dengan jumlah yang lebih banyak dari beberapa sumber.
4. Untuk pembobotan atribut sebaiknya menggunakan metode optimasi lain untuk mengetahui tingkat akurasi yang dihasilkan.

5. AUTHORS' NOTE

Penulis menyatakan bahwa tidak ada konflik kepentingan terkait publikasi artikel ini. Penulis mengonfirmasi bahwa artikel ini juga bebas dari plagiarisme.

6. REFERENCES

- Abraham, R. A., & Grosan, C. (2006). *Swarm intelligence in data mining*. Springer.
- Bramer, M. (2007). *Principles of data mining*. Springer.
- Hand, D. J. (2007). *Principles of data mining*. <https://doi.org/10.2165/00002018-200730070-00010>
- Hidayatuloh, M. A., Kartika, K. P., & Permadi, D. F. H. (2022). Implementasi algoritma C4.5 untuk memprediksi capaian pembelajaran daring (Studi kasus siswa MAN 3 Blitar). *Jurnal Algoritma*, 3(1), 33–47. <https://doi.org/10.35957/algoritme.v3i1.3292>
- Larose, D. T. (2005). *Discovering knowledge in data: An introduction to data mining*. John Wiley & Sons.
- Maurina, D. M., & Fanani, A. Z. (2011). Penerapan data mining untuk rekomendasi beasiswa pada SMA Muhammadiyah Gubug menggunakan algoritma C4.5., 3.
- Pemerintah Republik Indonesia. (2003). *Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional*.
- Rizmayanti, A. I., Hidayati, N., Nugraha, F. S., & Gata, W. (2021). Penerapan data mining untuk memprediksi kompetensi siswa menggunakan metode decision tree (Studi kasus SMK Multicomp Depok). *Swabumi*, 9(1), 9–18. <https://doi.org/10.31294/swabumi.v9i1.8363>
- Roghib, M., Rahaningsih, N., & Dana, R. D. (2024). Penerapan algoritma C4.5 untuk seleksi penjurusan siswa baru pada sekolah menengah kejuruan (Studi kasus: SMK Plus Al-Hilal

Sofian, E. (2016). Algoritma K-Means dalam mengidentifikasi pekerjaan berdasarkan latar belakang, 2, 41–49.

Suharna, H. Y., Rosmalina, & Imamah, N. (2023). Pengembangan aplikasi pembelajaran jaringan komputer menggunakan PNETLab untuk sekolah menengah kejuruan jurusan teknik komputer dan jaringan dengan metode web-based learning. *Computing: Journal of Information System*, 10(1), 31–36. <https://doi.org/10.55222/computing.v10i01.1152>