



Klasifikasi penyakit Demam Berdarah Dengue (DBD) menggunakan Algoritma C5.0 berbasis Binary Particle Swarm Optimization (BPSO)

Ani Rimadani ^{a,1}, Agustina Pradjaningsih ^{a,2,*}, I Made Tirta ^{a,3}

^a Universitas Jember, Indonesia;

¹ rimadaniani@gmail.com; ² agustina.fmipa@unej.ac.id; ³ itirta.fmipa@unej.ac.id

*Correspondent Author

KATAKUNCI

Klasifikasi
DBD
Algoritma C5.0
Algoritma BPSO

KEYWORDS

Classification
DHF
C5.0 Algorithm
BPSO Algorithm

ABSTRAK

Demam Berdarah Dengue atau DBD merupakan penyakit menular yang disebabkan oleh virus dengue. Penyakit yang berbahaya dan beresiko kematian harus ditangani dengan cepat dan tepat. Penentuan seseorang menderita DBD atau tidak dapat dilakukan dengan teknologi bidang klasifikasi yaitu metode decision tree algoritma C5.0. Seiring perkembangan waktu, jumlah fitur dalam klasifikasi semakin banyak. Reduksi fitur diperlukan untuk mendapatkan model yang baik yaitu dengan algoritma Binary Particle Swarm Optimization (BPSO). Penelitian ini dilakukan menggunakan 13 fitur yang terdiri dari 12 fitur bebas (x) dan 1 fitur terikat (y) yang memiliki 2 kelas yaitu Positif dan Negatif. Model yang lebih baik pada penelitian ini adalah klasifikasi algoritma C5.0 berbasis BPSO yang mampu mereduksi fitur dari 12 fitur menjadi 9 fitur dengan akurasi sebesar 86% dibandingkan klasifikasi dengan algoritma C5.0 saja yang menghasilkan akurasi sebesar 71%

Classification of Dengue Hemorrhagic Fever (DHF) using C5.0 algorithm based on Binary Particle Swarm Optimization (BPSO)

Dengue Hemorrhagic Fever, or DHF, is an infectious disease caused by the dengue virus. Diseases that are dangerous and at risk of death must be treated quickly and precisely. The determination that someone is suffering from DHF or not can be done with the classification technology, namely the decision tree method, the C5.0 algorithm. Over time, the number of features in the classification increases. Feature reduction is needed for a good model, namely the Binary Particle Swarm Optimization (BPSO) algorithm. This research used 13 features consisting of 12 independent features and one bound feature with two classes: Positive and Negative. The better model in this study is the classification C5.0 algorithm based on BPSO, which can reduce features from 12 to 9 features with an accuracy of 86% compared to classification with the C5.0 algorithm alone, which produces an accuracy of 71%.

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



Pendahuluan

Pola Hidup Bersih dan Sehat (PHBS) dalam setiap kegiatan harus selalu diterapkan, ancaman dari tidak diterapkannya PHBS adalah kesehatan bahkan nyawa setiap individu. Kesadaran akan PHBS bisa dimulai dari hal-hal sederhana seperti tidak membiarkan adanya genangan air yang dapat menjadi tempat perkembang biakan nyamuk. Genangan air ini memungkinkan bagi segala jenis nyamuk berkembang biak, salah satunya nyamuk aedes aegypti yang menjadi penyebab utama penyakit Demam Berdarah Dengue (DBD). Penyakit DBD merupakan salah satu penyakit mematikan dan sangat mudah menyebar melalui gigitan nyamuk. Sejak tahun 1968 hingga tahun 2009, Indonesia menjadi negara dengan kasus DBD tertinggi di kawasan Asia Tenggara [1].

Genangan air biasanya disebabkan oleh sisa hujan atau banjir. Kabupaten jember merupakan kabupaten dengan curah hujan cukup tinggi. Berdasarkan data BPS, total rata-rata curah hujan di Kabupaten Jember meningkat setiap tahunnya [2]. Kondisi ini menyebabkan daerah menjadi lembab dan banyak genangan sisa hujan. Genangan tersebut dapat menjadi tempat berkembang biak berbagai jenis nyamuk.

Salah satu bidang pendidikan yang banyak digunakan dalam penelitian adalah statistika. Ilmu statistika dapat diterapkan dalam berbagai hal, misalnya untuk melakukan klasifikasi, prediksi, dan lain-lain. Penelitian yang dilakukan oleh [3] dengan tujuan untuk memperoleh model persamaan regresi untuk memperkirakan curah hujan, hasilnya diperoleh bahwa variabel yang berpengaruh terhadap curah hujan adalah variabel lama penyinaran.

Penyakit yang berbahaya dan beresiko kematian seperti penyakit DBD harus ditangani dengan cepat dan tepat. Kesalahan dan keterlambatan penanganan dapat menyebabkan penyakit semakin parah dan kematian. Penentuan seseorang menderita penyakit DBD atau tidak dapat dilakukan dengan salah satu teknologi dalam data mining khususnya pada bidang klasifikasi. Penelitian sebelumnya dilakukan oleh [4] yang melakukan klasifikasi pada Analisis Sentimen aplikasi Maxim dengan metode KNN menggunakan perbandingan dari 5 jarak, didapatkan hasilnya bahwa 4 jarak memiliki nilai akurasi yang tinggi. Penerapan klasifikasi dalam berbagai penelitian telah banyak digunakan dan terbukti memiliki hasil yang sesuai. Penelitian yang dilakukan oleh [5] yang mengklasifikasikan penyakit DBD menggunakan Support Vector Machine (SVM) berbasis GUI Matlab dan menghasilkan nilai akurasi 96,4286%. Penelitian tentang klasifikasi penyakit DBD juga dilakukan [6] menggunakan metode Learning Vector Quantization (LVQ) dan menghasilkan akurasi 97,22%. Metode yang akan digunakan pada penelitian ini yaitu metode Decision Tree menggunakan algoritma C5.0.

Decision tree atau pohon keputusan merupakan pohon terstruktur dari proses klasifikasi dengan tujuan memprediksi output-nya. Algoritma C5.0 telah banyak digunakan dalam penelitian bidang klasifikasi seperti yang dilakukan oleh [7] yang melakukan prediksi kredit macet pada koperasi dan memperoleh hasil bahwa algoritma C5.0 memiliki nilai akurasi yang lebih baik dari algoritma KNN.

Seiring perkembangan waktu, atribut atau fitur yang digunakan untuk klasifikasi semakin banyak. Fitur yang tidak diperlukan dapat menurunkan tingkat akurasi dari suatu algoritma. Penelitian yang dilakukan oleh [8] menyebutkan bahwa fitur yang berlebihan menyebabkan proses klasifikasi lebih lama dan hasilnya kurang efisien. Oleh karena itu, diperlukan adanya reduksi fitur untuk memaksimalkan akurasi. Penelitian yang dilakukan oleh [9] mengenai prediksi kemenangan pada permainan poker menunjukkan bahwa reduksi fitur menggunakan algoritma Weight Improved Particle Swarm Optimization (WIPSO) pada sistem klasifikasi menggunakan algoritma C5.0 memiliki nilai akurasi yang lebih tinggi jika dibandingkan dengan penggunaan algoritma C5.0 saja.

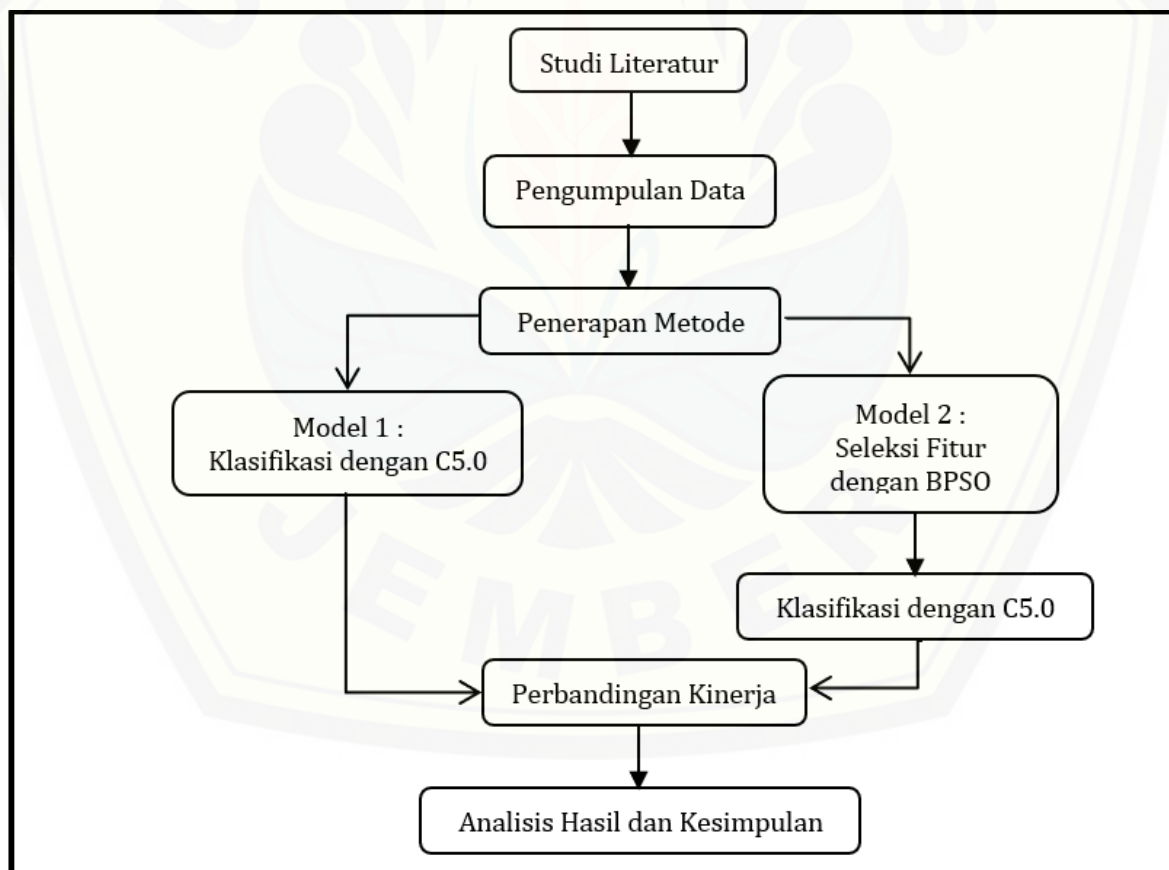
Salah satu algoritma yang dapat digunakan untuk mereduksi fitur yaitu Particle Swarm Optimization (PSO) yang saat ini memiliki pembaruan algoritma berupa Binary Particle Swarm Optimization (BPSO) [10]. PSO merupakan metode pendekatan metaheuristik yang terinspirasi dari sekumpulan serangga. Pada kasus seleksi fitur, representasi posisi

berdasarkan nilai acak tidak dapat mempresentasikan fitur yang terpilih. Maka dari itu, modifikasi dalam bentuk biner (BPSO) dapat menunjukkan fitur mana saja yang terpilih. Algoritma BPSO lebih cocok digunakan untuk mereduksi fitur pada klasifikasi menggunakan algoritma Decision Tree. Hal tersebut dibuktikan dalam penelitian sebelumnya bahwa menggunakan algoritma BPSO dengan algoritma C4.5 dan BPSO dengan algoritma boosted C5.0 hasilnya lebih akurat dibandingkan dengan yang tidak menggunakan algoritma BPSO untuk mereduksi fitur [11]. Penelitian serupa juga dilakukan oleh [12] yang menunjukkan bahwa penerapan BPSO+Decision Tree C4.5 pada deteksi kanker mampu menemukan fitur yang paling signifikan dan dapat meningkatkan akurasi.

Berdasarkan uraian di atas, peneliti memilih sistem klasifikasi dengan metode Decision Tree algoritma C5.0 berbasis Binary Particle Swarm Optimization (BPSO). Penelitian ini diharapkan mampu menunjukkan adanya pengaruh algoritma BPSO sebagai reduksi fitur untuk peningkatan akurasi pada algoritma C5.0 khususnya pada data pasien DBD di Kabupaten Jember.

Metode

Penelitian dilakukan di kampus Universitas Jember, dari bulan November 2022 – Februari 2023. Data sekunder yang digunakan diperoleh dari Rumah Sakit Daerah dr. Soebandi Kabupaten Jember. Tahapan penelitian yang akan dilakukan ditampilkan pada Gambar 1.



Gambar 1. Tahapan penelitian

Penjelasan mengenai tahapan penelitian pada Gambar 1 yaitu :

1. Studi Literatur

Peneliti mengumpulkan literatur mengenai kasus DBD, algoritma C50, algoritma BPSO, dan karya ilmiah yang berhubungan dengan penelitian.

2. Pengumpulan Data

Data pada penelitian ini adalah data pasien Demam Berdarah Dengue (DBD) yang diperoleh dari Rumah Sakit Daerah dr. Soebandi Kabupaten Jember. Terdapat 13 fitur yang digunakan dengan 12 fitur bebas (x) dan 1 fitur terikat (y). Penjelasan mengenai fitur dataset disajikan dalam Tabel 1.

Tabel 1. Daftar fitur dataset

	<i>Fitur</i>	<i>Deskripsi</i>	<i>Keterangan</i>
X1	Demam	Suhu tubuh pasien	1=suhu tinggi ($37^{\circ}\text{C} < x \leq 39$); 0=suhu normal ($35^{\circ}\text{C} \leq x \leq 37^{\circ}\text{C}$).
X2	Bibir Basah	Kondisi bibir basah	1=ya; 0=tidak
X3	Lemas	Kondisi lemas	1=ya; 0=tidak
X4	Mual	Mengalami mual	1=ya; 0=tidak
X5	Muntah	Mengalami muntah	1=ya; 0=tidak
X6	Pusing	Mengalami pusing	1=ya; 0=tidak
X7	Mimisan	Terjadi mimisan	1=ya; 0=tidak
X8	Nyeri Perut	Mengalami nyeri perut	1=ya; 0=tidak
X9	Sesak	Mengalami sesak	1=ya; 0=tidak
X10	Hipertermia	Mengalami hipertermia	1=ya; 0=tidak
X11	Gusi Berdarah	Terjadi perdarahan gusi	1=ya; 0=tidak
X12	Tinja Hitam	Tinja berwarna hitam	1=ya; 0=tidak
Y	Status Pasien	Status pasien	1=positif; 0=negatif

3. Penerapan Metode

Metode yang akan diterapkan pada penelitian ini adalah metode decision tree algoritma C5.0 dan algoritma BPSO untuk mereduksi fitur. Metode decision tree atau pohon keputusan berguna untuk mengeksplorasi data dan menemukan hubungan yang tersembunyi antara calon fitur input dan fitur target [13]. Algoritma C5.0 dipilih karena mampu melakukan klasifikasi dalam waktu yang lebih singkat, cara yang sederhana, dan hasil yang mudah dipahami. Aturan yang dihasilkan juga lebih sedikit dengan nilai akurasi tinggi [14]. Penelitian ini menghasilkan dua model. Model pertama yaitu pengklasifikasian dengan algoritma C5.0 saja. Langkah-langkah yang akan dilakukan adalah sebagai berikut [15].

- a. Menentukan *node* akar (*root*) yang diperoleh dari nilai *gain ratio* terbesar. Penentuan *gain ratio* diawali dengan perhitungan nilai *entropy* sesuai dengan persamaan berikut.

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

dengan: *Entropy* (S) = *entropy* dari himpunan kasus, n = jumlah partisi himpunan kasus, dan p_i = proporsi dari himpunan kasus ke-i terhadap himpunan kasus

- b. Menghitung nilai *gain* dari setiap fitur menggunakan persamaan berikut.

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

dengan: *Gain* (S, A) = *gain* dari himpunan kasus S terhadap fitur A, *Entropy* (S) = *entropy* dari himpunan kasus, A = fitur, n = jumlah partisi fitur A, $|S_i|$ = jumlah kasus pada partisi ke-i, $|S|$ = jumlah kasus dalam S, dan *Entropy* (S_i) = *entropy* dari himpunan kasus partisi ke-i

- c. Menghitung nilai *gain ratio* dengan persamaan berikut.

$$Gain\ ratio = \frac{Gain(S,A)}{\sum_{i=1}^n Entropy(S_i)} \quad (3)$$

dengan: $Gain(S,A)$ = nilai $gain$ fitur dan $\sum_{i=1}^n Entropy(S_i)$ = jumlah nilai $entropy$ satu fitur

Fitur dengan nilai $gain\ ratio$ tertinggi akan dijadikan $node$ akar. Perhitungan nilai $entropy$, $gain$, dan $gain\ ratio$ dilakukan kembali untuk memperoleh $node$ selanjutnya. Proses perhitungan ini dilakukan secara terus menerus sampai tidak ada fitur yang tersisa.

- d. Menghitung akurasi dari model menggunakan persamaan berikut.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

dengan: TP = nilai $true$ positif, TN = nilai $true$ negatif, FP = nilai $false$ positif, dan FN = nilai $false$ negatif

Model kedua yaitu pengklasifikasian dengan algoritma C5.0 berbasis BPSO. Proses diawali dengan seleksi fitur dengan algoritma BPSO, kemudian fitur terpilih akan digunakan untuk klasifikasi menggunakan algoritma C5.0. Langkah-langkah yang dilakukan adalah sebagai berikut [16].

- a. Menentukan jumlah partikel, nilai c_1, c_2, r_1, r_2 , dan jumlah iterasi.
- b. Menentukan posisi awal secara acak kemudian merubah kedalam bentuk biner menggunakan fungsi sigmoid pada persamaan berikut.

$$Sigmoid = \frac{1}{1+e^{-v^{new}}} \quad (5)$$

dengan: $Sigmoid$ = hasil nilai $sigmoid$, dan $e^{-v^{new}}$ = eksponensial kecepatan partikel

- c. Menentukan posisi baru dari setiap partikel dengan cara membandingkan posisi sebelumnya dengan nilai fungsi sigmoid yang telah diperoleh sebelumnya. Persamaan yang digunakan untuk proses perbandingan adalah sebagai berikut.

$$x^{new} = \begin{cases} 1 ; & \text{jika } Sigmoid(S) > rand(0,1) \\ 0 ; & \text{lainnya} \end{cases} \quad (6)$$

dengan: x^{new} = posisi baru partikel x , $rand(0,1)$ = nilai random antara 0 dan 1, dan $Sigmoid(S)$ = nilai sigmoid pada $update$ kecepatan partikel

- d. Menghitung nilai $fitness$ menggunakan algoritma C5.0. Pada tahap ini fitur dengan nilai 1 menyatakan fitur terpilih dan nilai 0 menyatakan fitur tidak terpilih. Cara yang dilakukan adalah dengan melakukan klasifikasi dengan algoritma C5.0 seperti pada model 1 hingga mendapatkan nilai akurasi. Nilai akurasi ini yang akan digunakan sebagai nilai $fitness$.

- e. Menentukan nilai $pbest$ (posisi terbaik yang diperoleh masing-masing partikel) dan nilai $gbest$ (posisi terbaik yang diperoleh partikel secara keseluruhan).
- f. Melakukan $update$ kecepatan setiap partikel menggunakan persamaan berikut.

$$V^{new} = w \times v^{old} + c_1 \times r_1(pbest - X_i) + c_2 \times r_2(gbest - X_i) \quad (7)$$

dengan: V^{new} = kecepatan baru tiap partikel, w = bobot inersia, v^{old} = kecepatan partikel iterasi sebelumnya, X_i = posisi partikel, $pbest$ = $local\ best$ dari partikel ke- i , $gbest$ = $global\ best$ dari seluruh kawanan, c_1, c_2 = nilai $learning\ rate$, r_1, r_2 = nilai $random$ antara 0 sampai 1

- g. Mengulangi proses b-f sampai mencapai iterasi yang diinginkan atau mendapat fitur yang optimal.

- h. Menghitung nilai akurasi model 2 menggunakan Persamaan (4).

4. Perbandingan Kinerja

Perbandingan kinerja dilakukan dengan menghitung nilai akurasi, presisi, dan recall dari kedua model. Perhitungan nilai ini didasarkan pada tabel $confusion\ matrix$ berikut.

Tabel 2. Confusion matrix

Kelas Prediksi	Kelas Data Aktual	
	Positif (1)	Negatif (0)
Positif (1)	TP	FP
Negatif (0)	FN	TN

Empat istilah yang digunakan dalam *confusion matrix* adalah sebagai berikut.

True Positif (TP) adalah data yang sebenarnya positif dan oleh sistem diklasifikasikan sebagai data positif.

True Negatif (TN) adalah data yang sebenarnya negatif dan oleh sistem diklasifikasikan sebagai data negatif.

False Positif (FP) adalah data yang sebenarnya negatif dan oleh sistem diklasifikasikan sebagai data positif.

False Negatif (FN) adalah data yang sebenarnya positif dan oleh sistem diklasifikasikan sebagai data negatif.

Tahapan dalam perbandingan kinerja yaitu :

Perhitungan nilai akurasi menggunakan Persamaan (4)

Perhitungan nilai presisi menggunakan persamaan berikut.

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

dengan: TP = nilai *true* positif dan FP = nilai *false* positif

Perhitungan nilai *recall* menggunakan persamaan berikut.

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

dengan: TP = nilai *true* positif dan FN = nilai *false* negatif

Hasil dari perhitungan tiga nilai tersebut dapat digunakan untuk mengukur model yang telah dibuat baik atau tidak. Model dapat dikatakan baik apabila memiliki nilai akurasi, presisi, dan *recall* yang tinggi [17].

5. Analisis Hasil dan Kesimpulan

Setelah memperoleh data dan program selesai dibuat, selanjutnya dilakukan proses klasifikasi dengan algoritma C5.0 berbasis BPSO dan algoritma C5.0. Hasil yang didapatkan selanjutnya dianalisis dan digunakan untuk penarikan kesimpulan.

Hasil dan Pembahasan

1. Statistika Deskriptif

Data yang digunakan adalah data pasien yang terkonfirmasi positif dan negatif penyakit DBD sebanyak 67 data yang diperoleh dari RSD. dr. Soebandi Jember. Data terbagi menjadi dua kelas yaitu data positif sebanyak 31 data dan 36 data yang terkonfirmasi negatif DBD. Fitur yang akan digunakan sebanyak 13 yang terdiri dari 12 fitur bebas (x) dan 1 fitur terikat (y). Data diolah menggunakan software Python. Data yang akan digunakan disajikan dalam Tabel 3.

Tabel 3. Data penelitian

No.	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	Y
1.	37,8	1	1	1	1	1	0	1	1	1	0	0	1
2.	37	1	1	1	1	1	0	1	1	1	0	0	1
3.	36	0	1	1	1	0	1	1	1	0	1	0	1
4.	37	1	1	1	0	0	0	1	0	1	0	0	1
5.	37	1	1	1	1	0	0	0	1	0	0	1	1

:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:
67.	37	0	1	0	0	0	0	0	0	0	0	0	0

2. Pembagian Data Klasifikasi

Proses klasifikasi melalui dua tahap yaitu training dan testing. Penelitian ini menggunakan 3 jenis rasio perbandingan untuk data training dan data testing. Rasio perbandingan tersebut adalah 90%:10%, 80%:20%, dan 70%:30%. Dari ketiga perbandingan ini akan dipilih satu perbandingan dengan nilai akurasi tertinggi sehingga mampu menunjukkan bahwa reduksi fitur dengan BPSO mampu meningkatkan akurasi.

3. Penerapan Metode

a. Model 1 (Klasifikasi dengan algoritma C5.0)

1) Proses *Training*

Proses klasifikasi yang pertama adalah proses *training* dan hasilnya ditunjukkan pada **Tabel 4** berikut.

Tabel 4. Akurasi data training algoritma C5.0

Rasio Perbandingan	Akurasi
90%:10%	76%
80%:20%	77%
70%:30%	76%

Hasil *training* data menunjukkan bahwa akurasi terbaik diperoleh saat rasio perbandingan data sebesar 80%:20% yaitu 77%.

2) Proses *Testing*

Proses selanjutnya yaitu *testing* dengan 3 jenis rasio yang sama. Hasil proses *testing* sebagai berikut.

Tabel 5. Confusion matrix algoritma C5.0

Rasio Data	Kelas Prediksi	Kelas Data Aktual		Akurasi
		Positif (1)	Negatif (0)	
90%:10%	Positif (1)	1	1	71%
	Negatif (0)	1	4	
80%:20%	Positif (1)	6	1	79%
	Negatif (0)	2	5	
70%:30%	Positif (1)	12	1	81%
	Negatif (0)	3	5	

Nilai akurasi yang diperoleh ketika rasio data 90%:10% sebesar 71%, dengan nilai *true* negatif yang menunjukkan bahwa data aktual bernilai negatif dan terklasifikasi negatif oleh program memiliki nilai yang lebih besar yaitu 4. Rasio data kedua yaitu 80%:20% menghasilkan akurasi 79% yang memiliki nilai *true* positif paling tinggi sebesar 6. Nilai *true* positif merupakan nilai data aktual positif dan terklasifikasi positif oleh program. Pada rasio data selanjutnya yaitu 70%:30% juga memiliki nilai *true* positif yang lebih besar yaitu 12 sehingga nilai akurasinya sebesar 81%.

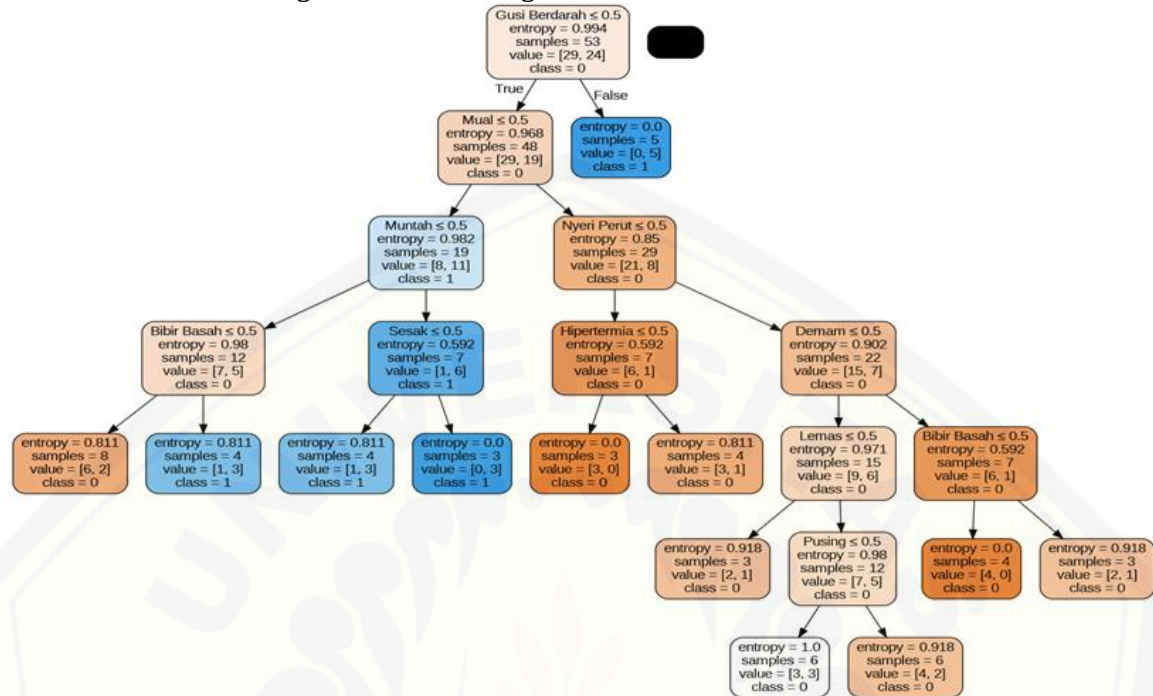
Evaluasi performa pada model 1 disajikan dalam **Tabel 6** berikut.

Tabel 6. Evaluasi performa klasifikasi C5.0

	Akurasi	Presisi	Recall
90% : 10%	71%	0,80	0,80
80% : 20%	79%	0,83	0,71
70% : 30%	81%	0,83	0,62

Performa model pada penelitian ini ditentukan dengan membandingkan nilai akurasi, *precision*, dan *recall*. Rasio perbandingan yang memiliki nilai akurasi, *precision*, dan *recall* terbaik adalah pada rasio perbandingan 80% data *training*, dan 20% data *testing*.

Hasil *decision tree* algoritma C5.0 sebagai berikut



Gambar 2. Decision tree C5.0

Gambar 2 merupakan hasil pohon keputusan klasifikasi dengan algoritma C5.0. Hasil tersebut juga dapat dipahami menggunakan rules sebagai berikut.

1. *If* Gusi Berdarah = *False* *Then* Status = Positif
2. *If* Gusi Berdarah = *True* *and* Mual = *False* *Then* Status = Negatif
3. *If* Gusi Berdarah = *True* *and* Mual = *True* *and* Muntah = *False* *Then* Status = Positif
4. *If* Gusi Berdarah = *True* *and* Mual = *True* *and* Muntah = *True* *and* Bibir Basah = *False* *Then* Status = Positif
5. *If* Gusi Berdarah = *True* *and* Mual = *True* *and* Muntah = *True* *and* Bibir Basah = *True* *Then* Status = Negatif

b. Model 2 (Klasifikasi dengan algoritma C5.0 berbasis BPSO)

Sebelum dilakukan proses klasifikasi, terlebih dahulu dilakukan proses pemilihan fitur optimal menggunakan algoritma BPSO. Fitur terpilih selanjutnya digunakan untuk proses klasifikasi dengan algoritma C5.0.

1) Pemilihan Fitur dengan Algoritma BPSO

Proses pemilihan fitur menggunakan beberapa parameter yang telah ditentukan sebelumnya. Percobaan penggunaan parameter telah dilakukan sebanyak 5 kali dan diperoleh hasil terbaik menggunakan kombinasi parameter dengan iterasi sebanyak 100 kali, jumlah partikel 30, nilai konstanta 0,5, dan bobot inersia sebesar 0,9. Proses pemilihan fitur untuk mendapat fitur optimal ini memerlukan waktu 4-5 detik dan hasil fitur yang diperoleh disajikan pada **Tabel 7** berikut.

Tabel 7. Hasil pemilihan fitur algoritma BPSO

<i>Rasio Perbandingan</i>	<i>Fitur Terpilih</i>	<i>Fitur Tidak Terpilih (Ter-reduksi)</i>
90%:10%	Bibir basah, lemas, mual, muntah, pusing, nyeri perut, sesak, hipertermia, dan gusi berdarah.	Demam, mimisan, dan tinja hitam.
80%:20%	Bibir basah, lemas, mual, muntah, pusing, nyeri perut, sesak, hipertermia, gusi berdarah, dan tinja hitam.	Demam dan mimisan.
70%:30%	Demam, bibir basah, mual, muntah, mimisan, nyeri perut, sesak, hipertermia, dan gusi berdarah.	Lemas, pusing, dan tinja hitam

Pada rasio data 90%:10% dan 70%:30% algoritma BPSO mereduksi fitur dari 12 fitur menjadi 9 fitur dan rasio data 80%:20% mereduksi fitur dari 12 fitur menjadi 10 fitur.

2) *Proses Training*

Proses *training* pada algoritma C5.0 berbasis BPSO menghasilkan akurasi sebagai berikut.

Tabel 8. Akurasi data training algoritma C5.0 berbasis BPSO

<i>Rasio Perbandingan</i>	<i>Akurasi</i>
90%:10%	76%
80%:20%	71%
70%:30%	69%

Hasil *training* data dengan algoritma C5.0 berbasis BPSO menunjukkan bahwa akurasi terbaik diperoleh saat rasio perbandingan data sebesar 90%:10% yaitu 76%.

3) *Proses Testing*

Setelah dilakukan proses training, selanjutnya akan dilakukan pengujian data testing pada algoritma C5.0 menggunakan fitur optimal yang diperoleh dari algoritma BPSO. Hasil pengujian data testing dengan algoritma C5.0 berbasis BPSO disajikan dalam Tabel 9 berikut.

Tabel 9. Confusion matrix algoritma C5.0 berbasis BPSO

<i>Rasio Data</i>	<i>Kelas Prediksi</i>	<i>Kelas Data Aktual</i>		<i>Akurasi</i>
		<i>Positif (1)</i>	<i>Negatif (0)</i>	
90%:10%	Positif (1)	2	0	86%
	Negatif (0)	1	4	
80%:20%	Positif (1)	7	0	86%
	Negatif (0)	2	5	
70%:30%	Positif (1)	13	0	86%
	Negatif (0)	3	5	

Nilai akurasi yang diperoleh dari algoritma C5.0 berbasis BPSO pada ketiga rasio data memiliki nilai yang sama yaitu 86%.

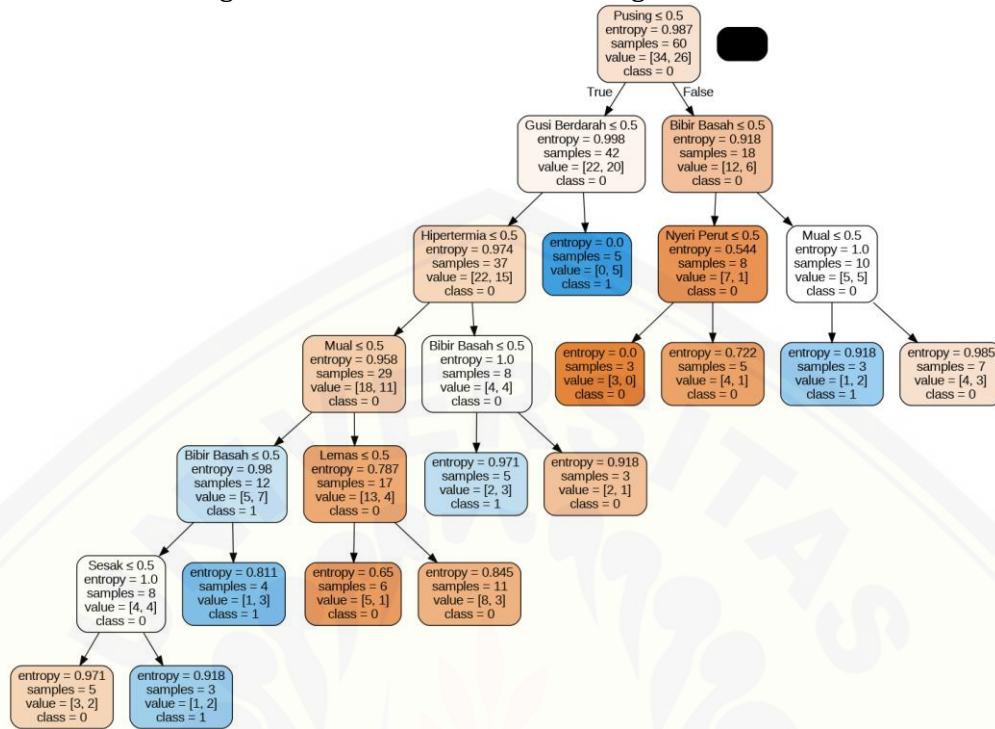
Evaluasi performa pada klasifikasi algoritma C5.0 berbasis BPSO meliputi nilai akurasi, presisi, dan *recall* sebagai berikut.

Tabel 10. Evaluasi performa klasifikasi C5.0 berbasis BPSO

	<i>Akurasi</i>	<i>Presisi</i>	<i>Recall</i>
90% : 10%	86%	1,00	0,80
80% : 20%	86%	1,00	0,71
70% : 30%	86%	1,00	0,62

Rasio perbandingan yang memiliki nilai akurasi, presisi, dan *recall* terbaik adalah pada rasio perbandingan 90% data *training*, dan 10% data *testing*.

Hasil *decision tree* algoritma C5.0 berbasis BPSO sebagai berikut.



Gambar 3. Decision tree C5.0 berbasis BPSO

Gambar 3 merupakan hasil pohon keputusan klasifikasi dengan algoritma C5.0 berbasis BPSO. Hasil tersebut juga dapat dipahami menggunakan *rules* sebagai berikut.

1. *If* Pusing = *False* and Bibir Basah = *False* and Mual = *False* Then Status = Negatif
2. *If* Pusing = *False* and Bibir Basah = *False* and Mual = *True* Then Status = Positif
3. *If* Pusing = *False* and Bibir Basah = *True* Then Status = Negatif
4. *If* Pusing = *True* and Gusi Berdarah = *False* Then Status = Positif
5. *If* Pusing = *True* and Gusi Berdarah = *True* and Hipertermia = *False* and Bibir Basah = *False* Then Status = Negatif
6. *If* Pusing = *True* and Gusi Berdarah = *True* and Hipertermia = *False* and Bibir Basah = *True* Then Status = Positif
7. *If* Pusing = *True* and Gusi Berdarah = *True* and Hipertermia = *True* and Bibir Basah = *True* Then Status = Positif
8. *If* Pusing = *True* and Gusi Berdarah = *True* and Hipertermia = *True* and Mual = *False* Then Status = Negatif
9. *If* Pusing = *True* and Gusi Berdarah = *True* and Hipertermia = *True* and Mual = *True* and Bibir Basah = *False* Then Status = Positif
10. *If* Pusing = *True* and Gusi Berdarah = *True* and Hipertermia = *True* and Mual = *True* and Bibir Basah = *True* and Sesak = *False* Then Status = Positif
11. *If* Pusing = *True* and Gusi Berdarah = *True* and Hipertermia = *True* and Mual = *True* and Bibir Basah = *True* and Sesak = *True* Then Status = Negatif

4. Perbandingan Kinerja

Setelah diperoleh hasil kinerja dari masing-masing model dalam proses *training* dan *testing*, maka didapatkan ringkasan sebagai berikut.

Tabel 11. Perbandingan kinerja

	Model 1 (C5.0)	Model 2 (C5.0+BPSO)
90% : 10%	71%	86%
80% : 20%	79%	86%
70% : 30%	81%	86%

Ringkasan dari **Tabel 11** menunjukkan bahwa penggunaan algoritma BPSO mampu meningkatkan akurasi hingga 15% pada proses klasifikasi dengan rasio perbandingan data *training* dan data *testing* sebesar 90%:10%. Rasio perbandingan yang lain juga menunjukkan adanya peningkatan akurasi dari model 1 ke model 2. Artinya algoritma BPSO memiliki pengaruh untuk peningkatan akurasi dengan cara mereduksi fitur.

Simpulan

Berdasarkan hasil implementasi dan pengujian pada algoritma C5.0 dan algoritma C5.0 berbasis BPSO, dapat disimpulkan bahwa algoritma BPSO dapat digunakan untuk mereduksi fitur pada sistem klasifikasi algoritma C5.0. Pada model terbaiknya, algoritma BPSO mereduksi fitur dari 12 fitur menjadi 9 fitur optimal. Penggunaan algoritma BPSO dan reduksi fitur BPSO juga mampu meningkatkan akurasi menjadi 86%, dibandingkan tanpa menggunakan BPSO yang hanya menghasilkan akurasi sebesar 71%.

Daftar Pustaka

- [1] Kemenkes, Buletin Jendela Epidemiologi, Jakarta: Kementerian Kesehatan RI, 2010.
- [2] BPS, "jemberkab.bps.go.id," Badan Pusat Statistik, 26 October 2021. [Online]. Available: <https://jemberkab.bps.go.id/subject/151/iklim.html#subjekViewTab3>. [Accessed 15 November 2022].
- [3] R. Normelia, T. D. Fortuna, E. P. Putri and E. Widodo, "Penerapan Metode Regresi Linier Berganda untuk Memperkirakan Curah Hujan (Studi Kasus : Stasiun Geofisika Sleman)," Jurnal Ilmiah Matematika, vol. 9, no. 1, pp. 8-18, 2022.
- [4] Mujhid, A. Thobirin, S. N. Firdausy, Sugiyarto and L. A. Rahmadani, "Perbandingan 5 jarak K-Nearest Neighbor pada Analisis Sentimen," Jurnal Ilmiah Matematika, vol. 8, no. 2, pp. 84-97, 2021.
- [5] A. Hasibuan, M. A. Mukid dan A. Prahutama, "Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue (DBD) Menggunakan Support Vector Machine (SVM) Berbasis GUI MATLAB," Jurnal Gaussian, vol. 6, no. 2, pp. 171-180, 2017.
- [6] F. Tawakal dan A. Azkiya, "Diagnosa Penyakit Demam Berdarah Dengue (DBD) Menggunakan Metode Learning Vector Quantization (LVQ)," JISKa, vol. 4, no. 3, pp. 193-201, 2020.
- [7] T. Permana, A. M. Siregar, A. F. N. Masruriyah dan A. R. Juwita, "Perbandingan Hasil Prediksi Kredit Macet pada Koperasi Menggunakan Algoritma KNN dan C5.0," dalam Conference on Innovation and Application of Science and Technology, Malang, 2020.
- [8] D. Normawati dan S. Winiarti, "Seleksi Fitur Menggunakan Penambahan Data Berbasis Variable Precision Rough Set (VPRS) untuk Diagnosis Penyakit Jantung Koroner," Jurnal Ilmu Teknik Elektro Komputer dan Informatika, vol. 3, no. 2, pp. 100-109, 2017.
- [9] M. F. Januarsyah and Ermatita, "Prediksi Akurasi Kemenangan pada Permainan Poker Menggunakan Algoritma C5.0 dan WIPSO," Jurnal Teknik Informatika, vol. 13, no. 1, pp. 66-71, 2021.
- [10] Wiharto, Herianto and H. Kusnanto, "Evaluation of Performance Feature Selection on Data Mining Technique for Diagnosis Coronary Heart Disease," in Konferensi Nasional Teknologi Informasi dan Komunikasi, Yogyakarta, 2016.
- [11] E. Pashaei, M. Ozen and N. Aydin, "Improving Medical Diagnosis Reliability Using Boosted C5.0 Decision Tree empowered by Particle Swarm Optimization," in Proceedings of the Annual

- International Conference of the IEEE Engineering in Medicine and Biology Society, New Jersey, 2015.
- [12] C. Pradana, Adiwijaya and A. Aditsania, "Implementing Binary Particle Swarm Optimization and C4.5 Decision Tree for Cancer Detection Based on Microarray Data Classification," in The 2nd International Conference on Data and Information Science, Bristol, 2019.
- [13] J. Han, M. Kamber and J. Pei, *Data Mining Concepts and Techniques 3rd Edition*, USA: Morgan Kauffman, 2011.
- [14] Giri and P. Paul, *Applied Marketing Analytics using SPSS (Modeler, Statistics and AMOS Graphics)*, Delhi: PHI Learning Private Limited, 2021.
- [15] E. Bulolo, *Data Mining untuk Perguruan Tinggi*, Yogyakarta: Deepublish, 2020.
- [16] Lazinec, *Particle Swarm Optimization*, Austria: Intech, 2009.
- [17] Daqiqil, *Machine Learning : Teori, Studi Kasus dan Implementasi menggunakan Python*, Riau: UR Press, 2021.

