

Optimasi PSO untuk Meningkatkan Performa Algoritma C4.5 dalam Memprediksi Risiko Kesehatan Kehamilan

Khaerul Ma'mur^{1*}, Asep Erlan Maulana²

^{1,2}Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No.46, Buaran, Serpong, Tangerang Selatan, Banten, 15310

e-mail: ¹kemunk.alfatih@gmail.com, dosen02716@gmail.com

*Corresponding author

Submitted Date: December 19th, 2024

Reviewed Date: December 27th, 2024

Revised Date: Desember 28th, 2024

Accepted Date: Desember 31th, 2024

Abstract

Nowadays, many diseases are feared to threaten the health of the body in pregnant women and their fetuses, so there is a need for early prevention. One's knowledge in predicting pregnancy health risks is important for prevention. However, precise prediction of such risks can be challenging, given the involvement of considerable and complex data. The C4.5 algorithm is widely used in data analysis. Unfortunately, its performance sometimes does not produce good accuracy. So, it needs to be improved by optimizing its structure parameters. For this reason, the purpose of this research involves the use of PSO (Particle Swarm Optimization) to find optimal parameters for the C4.5 algorithm that can increase the accuracy of pregnancy health risk prediction. The results show that the C4.5 algorithm model that has been optimized with PSO has an accuracy rate of 71.65%, and the standard C4.5 algorithm model only achieves an accuracy rate of 67.49%. There is a difference of 4.16%, which shows the superiority of the PSO optimization approach in improving prediction accuracy in the C4.5 algorithm. So, the application of the C4.5 algorithm optimized with PSO can be used as a positive implication in improving the health care of pregnant women and making more accurate medical decisions. Ultimately, this research illustrates the potential of PSO in optimizing the C4.5 data classification algorithm for health knowledge.

Keywords: Analysis, Prediction, Algoritma C4.5, Particle Swarm Optimization, Pregnancy

Abstrak

Sekarang ini, banyak penyakit yang dikhawatirkan mengancam kesehatan tubuh pada wanita hamil dan janinnya, sehingga perlu ada pencegahan sejak dini. Pengetahuan seseorang dalam memprediksi risiko kesehatan kehamilan penting sebagai upaya pencegahan. Namun, prediksi yang tepat tentang risiko tersebut dapat menjadi tantangan tersendiri, mengingat keterlibatan data yang cukup banyak dan kompleks. Algoritma C4.5 banyak digunakan secara luas dalam analisis data. Sayangnya performanya terkadang belum menghasilkan akurasi yang baik. Sehingga perlu ditingkatkan dengan mengoptimalkan parameter strukturnya. Untuk itu, tujuan penelitian ini melibatkan penggunaan PSO (*Particle Swarm Optimization*) untuk menemukan parameter-parameter optimal terhadap algoritma C4.5 yang dapat menaikkan akurasi prediksi risiko kesehatan kehamilan. Hasilnya menunjukkan, model algoritma C4.5 yang telah dioptimalkan dengan PSO tingkat akurasi sebesar 71,65% dan model algoritma C4.5 standar hanya mencapai tingkat akurasi sebesar 67,49%. Terdapat selisih sebesar 4,16% yang menunjukkan keunggulan pendekatan optimasi PSO dalam meningkatkan akurasi prediksi pada algoritma C4.5. Sehingga penerapan algoritma C4.5 yang dioptimasi dengan PSO dapat digunakan sebagai implikasi positif dalam meningkatkan perawatan kesehatan wanita hamil dan pengambilan keputusan medis yang lebih akurat. Hingga akhirnya penelitian ini menggambarkan potensi PSO dalam mengoptimalkan algoritma klasifikasi data C4.5 untuk pengetahuan kesehatan yang cukup penting.

Kata kunci: Analisis, Prediksi, Algoritma C4.5, Particle Swarm Optimization, Kehamilan



1 Pendahuluan

Kehamilan merupakan sebuah periode unik dalam kehidupan seorang wanita di mana berbagai perubahan fisiologis terjadi secara terkoordinasi, yang dipicu oleh perubahan hormon. Proses ini bertujuan untuk mempersiapkan tubuh wanita dalam mengakomodasi perkembangan dan pertumbuhan janin, serta untuk mempersiapkan tubuhnya saat persalinan dan menyusui. Kehamilan pada wanita umumnya terjadi dengan durasi rata-rata sekitar 9 bulan atau sekitar 37 hingga 42 minggu, mengharuskan perempuan membawa janin di dalam rahimnya (Putri & Ismiyatun, 2020).

Selama kehamilan, respons tubuh terhadap zat kimia dan risiko kesehatan tertentu dapat meningkat. Berbagai macam risiko kesehatan pada saat periode kehamilan dapat terjadi bila kurang memperhatikan asupan nutrisi yang sesuai. Deteksi dini kehamilan dapat membantu mengidentifikasi dan menangani risiko tinggi pada ibu hamil secara lebih efektif (Putri & Ismiyatun, 2020). Kehamilan dengan risiko tinggi dapat meningkatkan kemungkinan kematian pada ibu, karena kondisi ini dapat membahayakan nyawa ibu dan bayi, terutama saat proses persalinan (Ratnaningtyas & Indrawati, 2023).

Tiga tingkatan risiko kehamilan yang sering terjadi yakni kategori risiko rendah, sedang, dan tinggi (Windasari, 2022). Guna memprediksi risiko kesehatan kehamilan termasuk ke dalam kategori yang mana dapat digunakan algoritma *machine learning* berbasis *decision tree*, dalam ruang lingkup data mining. Data mining merupakan proses menggunakan komputer untuk menganalisis data dalam jumlah besar guna menemukan pola atau informasi bermakna. Metode ini membantu mengekstrak informasi yang bermanfaat dari data mentah dengan lebih efisien (Ma'mur, 2019). Penggunaan *data mining* lebih efektif dalam mengolah dan menganalisis data untuk mendapatkan hasil yang akurat dengan cara yang efisien (Reza & Muhammad Syaifur Rohman, 2024).

Salah satu pendekatan *data mining* yakni menggunakan algoritma C4.5 yang merupakan satu diantara bentuk struktur *decision tree* yang menyerupai flowchart yang sangat efektif untuk melakukan pengelompokan data (Saputra et al., 2021). Algoritma ini banyak digunakan secara luas untuk analisis dan prediksi di berbagai bidang. Hanya saja kinerja algoritma C4.5 sangat bergantung pada pengaturan parameter tertentu,

seperti kedalaman pohon keputusan dan kriteria pemisahan node. Penggunaan nilai kriteria gain pada algoritma C4.5 untuk memilih fitur yang paling sesuai sebagai pemisah node dalam pohon keputusan (Hayadi & Damanik, 2022).

Sayangnya sebuah prediksi seringkali masih kurang akurat sehingga penelitian ini menggunakan *Particle Swarm Optimization* (PSO) sebagai alat optimasi. PSO merupakan algoritma optimisasi yang inspirasinya diambil dari perilaku kelompok burung atau ikan dalam mencari solusi optimal melalui proses iteratif (Rizki Kurniawan et al., 2024) (Rizki & Nurlaili, 2021). PSO ditambahkan untuk melakukan seleksi atribut, yang bertujuan untuk meningkatkan tingkat akurasi hasil analisis (Fanani Rudi & Fikriah Katul, 2023).

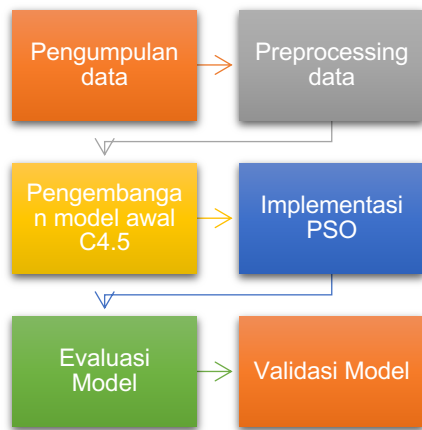
Pada tahap optimasi, dilakukan uji perbandingan untuk mengevaluasi tingkat akurasi yang dihasilkan dengan memvariasikan parameter *performance*. Hasil dari penelitian mencakup perbandingan nilai akurasi antara penggunaan algoritma C4.5 dan setelah penggunaan optimasi PSO terhadap algoritma C4.5 (Anggita & Ikmah, 2021). Akhirnya, model yang dihasilkan ada peningkatan yang cukup signifikan.

Tujuan penelitian ini selain dari manfaat kesehatan, juga memiliki implikasi dalam pengembangan bidang ilmu *data mining* dan *machine learning*. Dengan berfokus pada integrasi algoritma C4.5 yang cukup andal dalam klasifikasi data dengan teknik optimisasi PSO, penelitian ini memiliki potensi besar untuk memajukan pemahaman dan perawatan kesehatan kehamilan. Dengan kombinasi antara kecerdasan buatan dan analisis data, hasilnya dapat mencapai prediksi yang lebih tinggi, yang pada akhirnya akan membantu mengurangi risiko kesehatan kehamilan. Maka penelitian ini menjadi langkah penting dalam upaya untuk meningkatkan perawatan kesehatan wanita hamil dan menyelamatkan lebih banyak nyawa dalam prosesnya.

2 Metode Penelitian

Metode penelitian dalam hal penggunaan data mining mengacu pada serangkaian teknik dan pendekatan yang digunakan untuk mengekstraksi pengetahuan atau pola yang bermanfaat dari kumpulan data yang besar dan kompleks. Kemudian dirunut satu per satu untuk menghitung pola, model, dan evaluasinya. Pada penelitian ini digunakan metode penelitian sebagai alur atau

tahapan penyelesaian permasalahan penelitian yang relevan dengan topik penelitian, dengan alur sebagai berikut.



Gambar 1. Metode Penelitian

Secara umum tahapan pada metode penelitian seperti pada gambar 1 yang disusun untuk mempermudah alur penyelesaian penelitian. Berikut penjabaran per tahapannya.

1. Pengumpulan Data

Pada tahap ini dikumpulkan dataset dari sumber data sekunder yang berasal dari *open data* uci *machine learning*. Data yang digunakan untuk penelitian ini sebelumnya pernah menjadi data penelitian dari sumber yang relevan dan kredibel. Di dalam data berisi data medis dan riwayat kesehatan wanita hamil. Ada 1014 sampel yang tersaji secara acak dengan kepemilikan 3 label *class* prediksi keputusan yakni *high risk* sebanyak 272, *mid risk* sejumlah 336, dan *low risk* sejumlah 406.

2. Preprocessing Data

Preprocessing data dalam data mining adalah serangkaian langkah persiapan yang dilakukan pada data mentah sebelum data tersebut dapat dianalisis oleh algoritma *data mining*. Tujuan dari preprocessing data adalah untuk membersihkan, mengorganisasi, dan mengubah data agar menjadi lebih sesuai digunakan untuk analisis. Pada tahap ini data dibersihkan menggunakan tahapan umum *preprocessing* data yang bertujuan agar data terhindar dari *missing values* (jika ditemukan), *outliers*, dan data yang tidak relevan serta melakukan normalisasi data jika diperlukan.

3. Pengembangan Model Awal C4.5

Pengembangan model awal merujuk pada proses pembuatan pohon keputusan atau model

keputusan berupa *decision tree*. Tahapan ini membangun model awal menggunakan algoritma C4.5 untuk memprediksi risiko kesehatan kehamilan berdasarkan atribut yang sesuai perhitungan terdapat dalam data.

Pengembangan model menggunakan *tools* RapidMiner untuk mengotomatiskan sebagian besar proses. Pada RapidMiner, peneliti menggunakan antarmuka grafis untuk memilih dan mengatur langkah-langkah dalam proses pengembangan model, termasuk pemrosesan data, pemilihan atribut, pembangunan model, dan evaluasi kinerja model.

4. Implementasi PSO

Tahapan ini kemudian mengintegrasikan algoritma *Particle Swarm Optimization* (PSO) untuk mengoptimalkan keakuratan model C4.5. Kemudian mengevaluasi performa model C4.5 setelah menambahkan optimalisasi PSO. Dalam penelitian ini, PSO digunakan untuk memilih atribut yang relevan guna meningkatkan kinerja algoritma C4.5. PSO ditempatkan setelah tahap *preprocessing* data, di mana algoritma ini mencari kombinasi atribut terbaik yang menghasilkan pembagian data optimal saat diterapkan pada pohon keputusan C4.5. Proses ini memastikan hanya atribut signifikan yang dipertahankan untuk meningkatkan keakuratan prediksi risiko kesehatan kehamilan, dengan implementasi dilakukan secara terintegrasi di RapidMiner sesuai alur yang dirancang.

5. Evaluasi Model

Evaluasi dilakukan sebagai cara melihat seberapa tingkat akurasi yang dihasilkan. Pada tahap evaluasi model digunakan metrik evaluasi seperti akurasi, sensitivitas, spesifisitas, dan nilai prediksi positif yang sering disebut *Confusion Matrix*. Tujuannya untuk menilai performa model yang telah dioptimalkan menggunakan metrik evaluasi tersebut dan mengukur sejauh mana model dapat memprediksi risiko kesehatan kehamilan dengan benar.

6. Validasi Model

Validasi model merujuk pada proses pengujian dan evaluasi model yang telah dikembangkan, yaitu model prediksi yang menggunakan algoritma C4.5 yang telah dioptimalkan dengan PSO. Validasi model merupakan tahapan dalam *Machine Learning* dan *Data Mining* guna memastikan

pemodelan mempunyai kinerja yang baik dan dapat digunakan untuk memprediksi risiko kesehatan kehamilan dengan tingkat akurasi yang dapat diandalkan. Validasi model menggunakan data independen atau menggunakan metode validasi silang (*cross-validation*) yang diterapkan sejak awal pembangunan model *decision tree* untuk memastikan generalisasi model. Kemudian ditambahkan validasi model T-test sebagai upaya penyamaan nilai kinerja rata-rata.

3 Hasil dan Pembahasan

3.1 Dataset

Seperti pada tahapan metode penelitian, contoh sampel data yang dijadikan bahan rujukan pembuatan dan analisis model ditunjukkan pada tabel 1.

Tabel 1. Dataset penelitian

Age	Systolic BP	Diastolic BP	BS	Body Temp	Heart Rate	Risk Level
25	130	80	15	98	86	high risk
35	140	90	13	98	70	high risk
29	90	70	8	100	80	high risk
30	140	85	7	98	70	high risk
35	120	60	6.1	98	76	low risk
23	140	80	7.0	98	70	high risk
23	130	70	7.0	98	78	mid risk
35	85	60	11	102	86	high risk
32	120	90	6.9	98	70	mid risk
42	130	80	18	98	70	high risk
23	90	60	7.0	98	76	low risk
19	120	80	7	98	70	mid risk
25	110	89	7.0	98	77	low risk
20	120	75	7.0	100	70	mid risk
48	120	80	11	98	88	mid risk
15	120	80	7.0	98	70	low risk
50	140	90	15	98	90	high risk
25	140	100	7.0	98	80	high risk
30	120	80	6.9	101	76	mid risk
10	70	50	6.9	98	70	low risk
...

Age	Systolic BP	Diastolic BP	BS	Body Temp	Heart Rate	Risk Level
55	120	90	18	98	60	high risk
35	85	60	19	98	86	high risk
43	120	90	18	98	70	high risk
32	120	65	6	101	76	mid risk

Dari 1014 data individu, disajikan sebagian untuk memberikan gambaran secara umum. Pada tabel di atas tertera enam atribut yang memiliki keterangan data berbeda-beda pada tiap individu. Setiap atributnya berperan dalam menunjang hasil akhir level risiko yang disimpulkan. Data disajikan secara sederhana untuk lebih mempermudah menafsirkan hasil. Setiap atribut dijelaskan secara singkat dan jelas dalam tabel 2 untuk memberikan pemahaman lebih detail.

Tabel 2. Penjelasan Atribut Dataset

No.	Atribut	Penjelasan
1	Age	Usia wanita hamil saat dilakukan pemeriksaan
2	SystolicBP	Nilai tekanan darah atas dalam mmHg
3	DiastolicBP	Nilai tekanan darah bawah dalam mmHg
4	BoldS	Kadar glukosa dalam darah
5	BodyTemp	Suhu tubuh
6	HeartRate	Denyut jantung
7	RiskLevel	Prediksi tingkat risiko kesehatan kehamilan

3.2 Preprocessing Data

Pada tahapan ini semua nilai pada atribut tidak ada yang terindikasi missing value, sehingga dataset yang disajikan tidak memerlukan proses *clearing*, *normalize*, atau pembetulan data. Jika terdapat indikasi *missing value*, maka dilakukan *Data Celaning* untuk menghapus data yang tidak lengkap. Salah satu alasan mengapa *preprocessing* data tidak dilakukan saat penelusuran adalah karena algoritma ini memiliki kemampuan bawaan yang kuat untuk menangani data mentah atau minim preprocessing.

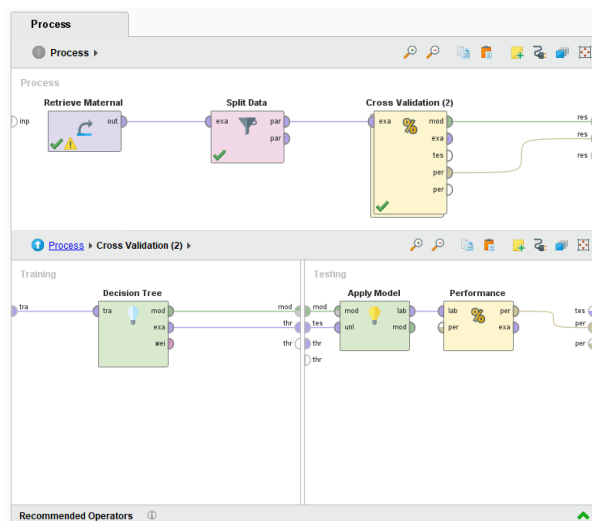
Algoritma C4.5 dapat secara efektif menangani masalah seperti nilai yang hilang, tipe data campuran, dan skala yang beragam tanpa



memerlukan langkah-langkah *preprocessing* yang kompleks (Khalimi, 2022). Dengan menggunakan data mentah langsung, dapat memanfaatkan keberagaman dan keseluruhan informasi yang terkandung dalam dataset tanpa risiko kehilangan detail atau mempengaruhi hasil analisis (www.kuliahkomputer.com, 2018). Sehingga dapat dilanjutkan untuk ke tahapan berikutnya.

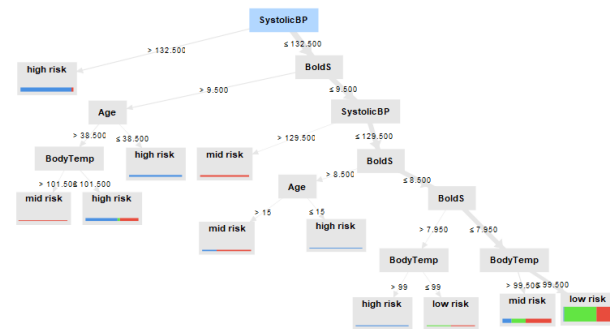
3.3 Pembangunan Model Awal

Tahap ini dilakukan pembuatan model berupa pohon keputusan yang ditentukan berdasarkan perhitungan entropy, information gain, dan gain ratio menggunakan *tool* Rapidminer. Masing-masing atribut dihitung total *information gain*, yang mana *gain* tertinggi akan menjadi node awal dalam pembentukan pohon keputusan. Langkah ini dilakukan guna menentukan jumlah *class* yang dinyatakan *low risk*, *mid risk*, dan *high risk*.



Gambar 2. Proses pembuatan model algoritma C4.5 menggunakan Rapidminer

Seperti yang tertera dalam gambar 2, dataset yang telah dimasukkan ke dalam lembar kerja dihubungkan dengan *split data* dengan maksud agar data training dan data testing dapat dipisahkan secara otomatis dengan rasio tujuh dibanding 3. Kemudian dihubungkan kembali ke sebuah operator validasi yang disebut *cross validation*. Kemudian langkah selanjutnya pada kotak kerja *Training* dimasukkan operator modelling *Decision Tree* (C4.5). Setelahnya, memasukkan operator *scoring* yakni *Apply Model* dilanjutkan operator *Validation* yakni *Performance* di kotak *Testing* guna melihat berapa nilai akurasinya. Kemudian



Gambar 3. Model pohon keputusan yang dihasilkan menggunakan algoritma C4.5

klik ikon *play/Run*. Sehingga menghasilkan model pohon keputusan, *rule*, dan performa dari penggunaan algoritma C4.5 tanpa melibatkan penggunaan optimasi PSO. Hasil pohon keputusannya ada pada gambar 3.

Penggunaan algoritma *decision tree* bertujuan untuk menghasilkan aturan yang dapat mempermudah proses pengambilan keputusan ketika dihadapkan dengan data baru. (Azuaje, 2006). Dari pohon keputusan tersebut didapatkan aturan atau *rule* dalam memprediksi risiko kehamilan berdasarkan dataset yang telah diolah sebagai berikut.

1. *IF* SystolicBP > 132,5 *THEN* RiskLevel = high risk
2. *IF* SystolicBP ≤ 132,5 *AND* BoldS > 9,5 *AND* Age > 38,5 *AND* BodyTemp > 101,5 *THEN* RiskLevel = mid risk
3. *IF* SystolicBP ≤ 132,5 *AND* BoldS > 9,5 *AND* Age > 38,5 *AND* BodyTemp ≤ 101,5 *THEN* RiskLevel = high risk
4. *IF* SystolicBP ≤ 132,5 *AND* BoldS > 9,5 *AND* Age ≤ 38,5 *AND* BodyTemp ≤ 101,5 *THEN* RiskLevel = high risk
5. *IF* SystolicBP > 129,5 *AND* BoldS ≤ 9,5 *THEN* RiskLevel = mid risk
6. *IF* SystolicBP ≤ 129,5 *AND* BoldS > 8,5 *AND* Age > 15 *THEN* RiskLevel = mid risk
7. *IF* SystolicBP ≤ 129,5 *AND* BoldS > 8,5 *AND* Age ≤ 15 *THEN* RiskLevel = high risk
8. *IF* SystolicBP ≤ 129,5 *AND* BoldS ≤ 8,5 *AND* BodyTemp > 99 *THEN* RiskLevel = high risk

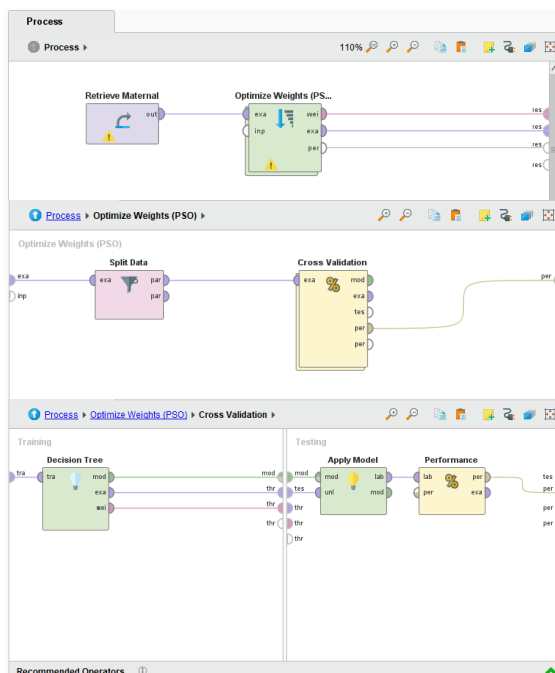


9. IF $SystolicBP \leq 129,5$ AND $BoldS \leq 8,5$ AND $BodyTemp \leq 99$ THEN $RiskLevel = low\ risk$
10. IF $SystolicBP \leq 129,5$ AND $BoldS \leq 7,95$ AND $BodyTemp > 99,5$ THEN $RiskLevel = mid\ risk$
11. IF $SystolicBP \leq 129,5$ AND $BoldS \leq 7,95$ AND $BodyTemp \leq 99,5$ THEN $RiskLevel = low\ risk$

Pada tahap ini juga dilihat persentase tingkat akurasi yang didapatkan menggunakan metode *cross validation* yang sudah disertakan pada saat pembentukan model pohon keputusan sebelumnya. Hal ini dilakukan sebagai penentu angka awal tingkat akurasi yang akan dibandingkan. Hasil tingkat akurasi dibahas pada poin evaluasi model.

3.4 Optimasi PSO

Pada tahap ini optimasi *particle swarm optimization* (PSO) ditambahkan ke dalam pemrosesan algoritma C4.5. Tahap ini dilakukan untuk memberikan pertambahan tingkat akurasi dalam penentuan risiko kesehatan kehamilan sesuai dataset yang ada. Pengoptimasian dilakukan dengan tetap menggunakan *tool* rapidminer karena kecepatan dan ketepatan dalam pembuatan model dengan menambahkan operator *Optimize Weight* (PSO) ke dalam area kerja. Berikut alur yang diterapkan pada proses optimasi PSO seperti yang ditunjukkan melalui gambar 4.



Gambar 4. Alur penambahan optimasi PSO

Dalam lembar kerja awal seperti pada gambar dilakukan proses penghubungan antar operator dengan sumber dataset. Pada alur pertama, dataset dihubungkan dengan operator PSO sebagai *tool* optimasi. Kemudian di dalam operator PSO, akan membuka lembar kerja kedua yang diisi operaor *split data* yang sama perbandingan rasionya dengan tahap pemodelan awal. Kemudian dihubungkan juga dengan operator *cross validation*.

Setelah terkoneksi, akan tampil lembar kerja ketiga yang dibagi menjadi halaman *training* dan halaman *testing*. Pada halaman *training* diberikan operator model algoritma C4.5 yakni *Decision Tree*. Sedangkan pada halaman *testing* diberikan operator *apply model* dan *performance*. Hasil dari penambahan optimasi disajikan pada poin evaluasi model.

3.5 Evaluasi Model

Tahap evaluasi model digunakan untuk menguji performa model klasifikasi yang sudah dibangun dalam penelitian (Ilham et al., 2024). Pada tahapan ini pengujian dilakukan untuk mengamati hasil nilai akurasi. Pengujian ini bertujuan untuk mengevaluasi akurasi hasil klasifikasi menggunakan Algoritma C4.5 dan versi yang dioptimasi dengan PSO. Evaluasi dilakukan dengan menggunakan *confusion matrix* untuk menilai performa kedua algoritma tersebut.

Confusion matrix merupakan metode evaluasi yang dilakukan untuk menghitung nilai akurasi dalam data mining (Fadilah & Murnawan, 2021). *Confusion matrix* di RapidMiner digunakan untuk menilai sejauh mana model yang dilatih dapat mengklasifikasikan data dengan sesuai serta mengidentifikasi kesalahan klasifikasi.

Model *confusion matrix* pada tahapan ini menghasilkan matriks 3x3 sesuai jumlah *class label*. Data set yang telah dilakukan penghitungan performa sebelum penambahan optimasi PSO pada tahap sebelumnya ditampilkan berupa *confusion matrix* untuk melihat nilai akurasi dan tupel yang tersedia. Data yang dihasilkan ditampilkan pada gambar 5.

accuracy: 67.49% +/- 1.86% (micro average: 67.49%)

	true high risk	true low risk	true mid risk	class precision
pred. high risk	185	5	35	82.22%
pred. low risk	20	297	168	61.24%
pred. mid risk	13	23	66	64.71%
class recall	84.86%	91.38%	24.54%	

Gambar 5. Hasil confusion matrix yang didapatkan

Pada gambar 5, untuk menghitung nilai akurasi yang akan diperbandingkan menggunakan sebuah rumus atau persamaan. Akurasi diukur dengan membandingkan jumlah prediksi yang benar terhadap total prediksi yang dibuat, dihitung dalam bentuk persentase. Persamaannya yaitu:

$$Akurasi = \frac{\text{Jumlah prediksi benar}}{\text{Total prediksi}}$$

Oleh karena itu, nilai akurasi yang dapat dihitung berdasarkan *confusion matrix* di atas disajikan sebagai berikut:

$$Akurasi = \frac{185 + 297 + 66}{185 + 5 + 35 + 20 + 297 + 168 + 13 + 23 + 66}$$

$$Akurasi = \frac{548}{812}$$

$$Akurasi = 67,49$$

Hasil perhitungan di atas menunjukkan nilai akurasi algoritma C4.5 pada proses penelusuran data risiko kehamilan. Maka nilai akurasi yang didapatkan dengan menggunakan algoritma C4.5 tanpa optimasi PSO yaitu 67,49. Selanjutnya dihitung kembali nilai akurasi algoritma C4.5 menggunakan optimasi PSO sebagai pembanding dengan menampilkan hasil performa *confusion matrix*.

accuracy: 71.65% +/- 3.84% (micro average: 71.65%)

	true high risk	true low risk	true mid risk	class precision
pred. high risk	165	4	21	86.84%
pred. low risk	5	251	122	66.40%
pred. mid risk	20	29	92	65.25%
class recall	86.84%	88.38%	39.15%	

Gambar 6. Hasil confusion matrix algoritma C4.5 menggunakan optimasi PSO

Pada gambar 6, dihitung kembali nilai akurasinya dengan penambahan optimasi PSO menggunakan persamaan seperti sebelumnya. Dengan demikian maka perhitungan nilai akurasi berdasarkan hasil dari *confusion matrix* yang telah disajikan dapat diuraikan sebagai berikut:

$$Akurasi = \frac{165 + 251 + 92}{165 + 4 + 21 + 5 + 251 + 122 + 20 + 29 + 92}$$

$$Akurasi = \frac{508}{709}$$

$$Akurasi = 71,65$$

Berdasarkan hasil perhitungan di atas maka nilai akurasi yang dihasilkan pada prediksi risiko kehamilan menggunakan algoritma C4.5 dengan penambahan optimasi PSO sebesar 71,65.

3.6 Validasi Model

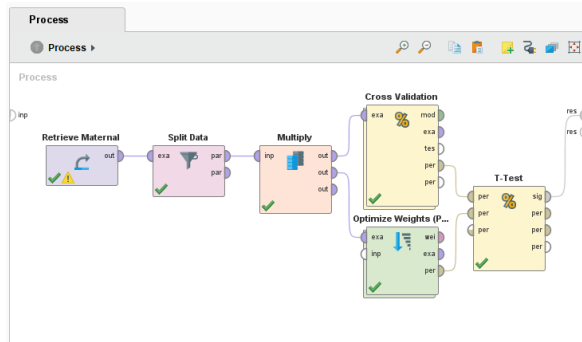
Bersasarkan hasil evaluasi model dan melihat hasil analisis *confusion matrix* menunjukkan bahwa algoritma C4.5 yang dioptimasi dengan *Particle Swarm Optimization* (PSO) memiliki akurasi yang lebih tinggi dibandingkan dengan algoritma C4.5 biasa. Algoritma C4.5 memiliki akurasi sejumlah 67,49%, sedangkan tingkat akurasi algoritma C4.5 berbasis PSO memiliki nilai 71,65%, dengan selisih perbedaan akurasi sebesar 4,16%, sebagaimana disajikan pada tabel 3.

Tabel 3. Hasil perbandingan pengujian algoritma klasifikasi C4.5

Algoritma	Akurasi
C4.5	67,49
C4.5 berbasis PSO	71,65

Untuk validasi atau pengujian lebih lanjut dan memastikan validitas hipotesis dalam penyelesaian penelitian, digunakan evaluasi dengan menggunakan metode statistik tradisional yang umum, yaitu uji t (t-Test) (Nurlelah & Mardiyanto, 2019). Tujuan pengujian ini adalah memeriksa apakah rata-rata semua nilai kinerja sama (hipotesis nol) menggunakan uji-t berpasangan (Rizquallah et al., 2024). T-test hanya berlaku untuk dua data, pengujian dilakukan pada semua pasangan data. Hasilnya adalah matriks signifikansi, yang menunjukkan hubungan antar data kinerja dengan tingkat signifikansi 0,05.

Proses atau alur validasinya menggunakan *tools RapidMiner* ditampilkan pada gambar 7.



Gambar 7. Model pengujian statistik t-test

Pada gambar 7, menggambarkan proses atau alur pemodelan saat menguji statistik t-test, untuk mengukur ketika pemodelan seperti pada gambar 7 di atas dijalankan, maka akan mendapatkan hasil perhitungan statistik seperti yang terlihat pada gambar 8.

Pairwise t-Test (T-Test)		
A	B	C
	0.674 +/- 0.026	0.701 +/- 0.030
0.674 +/- 0.026		0.049
0.701 +/- 0.030		

Gambar 8. Hasil pengujian statistik t-test

Berdasarkan gambar 8, menunjukkan hasil uji yang mengindikasikan bahwa algoritma berwarna oranye memiliki nilai alpha kurang dari 0,05 secara statistik. Sehingga dapat dikatakan bahwa algoritma C4.5 berbasis PSO menunjukkan perbedaan nilai yang cukup signifikan dengan probabilitas <0.05, yakni 0.049, dibandingkan dengan algoritma C4.5.

Hal ini menunjukkan bahwa algoritma C4.5 yang diterapkan dengan PSO lebih efektif dalam memprediksi risiko kehamilan dibandingkan dengan algoritma C4.5 biasa. Yang mana dengan nilai akurasi semakin tinggi tentu akan menghasilkan tingkat kemiripan atau keabsahan data yang kian akurat.

4 Kesimpulan

Berdasarkan evaluasi hasil pada pembahasan sebelumnya, didapatkan model algoritma C4.5 yang telah dioptimalkan dengan PSO mencapai tingkat akurasi sebesar 71,65%. Hasil ini menunjukkan peningkatan kinerja dibandingkan dengan model algoritma C4.5 standar yang hanya mencapai tingkat akurasi sebesar 67,49%.

Perbedaan angka sebesar 4,16% menunjukkan keunggulan dari pendekatan optimasi PSO dalam meningkatkan akurasi prediksi algoritma C4.5.

Pada studi kasus prediksi risiko kesehatan kehamilan, peningkatan akurasi ini cukup penting karena membantu sistem dalam mengidentifikasi tingkat risiko dengan lebih tepat berdasarkan data kesehatan ibu hamil, seperti usia, riwayat medis, dan faktor-faktor lain yang relevan. Hal ini menunjukkan bahwa penerapan teknik optimasi PSO tidak hanya efektif secara teknis tetapi juga memberikan kontribusi nyata dalam mendukung pengambilan keputusan medis yang lebih akurat.

Referensi

- Anggita, S. D., & Ikmah, I. (2021). Implementasi Pso Untuk Optimasi Bobot Atribut Pada Algoritma C4.5 Dalam Prediksi Kelulusan Mahasiswa. *JIPi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 6(2), 416–423. <https://doi.org/10.29100/jipi.v6i2.2440>
- Azuaje, F. (2006). Data Mining: Practical Machine Learning Tools and Techniques 2nd edition. *BioMedical Engineering OnLine*, 5(1). <https://doi.org/10.1186/1475-925x-5-51>
- Fadilah, Z., & Murnawan. (2021). Performance Comparison of K-Nearest Neighbor and Decision Tree C4.5 by Utilizing Particle Swarm Optimization for Prediction of Liver Disease. *International Journal of Open Information Technologies*, 9(10), 9–15.
- Fanani Rudi, M., & Fikriah Katul, F. (2023). Seleksi Fitur PSO untuk Klasifikasi Jenis Kekerasan dengan Algoritma C4.5. *Smart Comp: Jurnalnya Orang Pintar Komputer*, 12(1). <https://doi.org/10.30591/smartcomp.v12i1.4407>
- Hayadi, B. H., & Damanik, A. R. (2022). Pendekatan Machine Learning Menggunakan Algoritma C4.5 Berbasis Pso Dalam Analisa Pemahaman Pemrograman Website. *Jurnal Informatika Dan Teknik Elektro Terapan*, 10(3).
- Ilham, M., Adnyani, N. L. S. S., & Suryadi, K. (2024). Pembangunan Model Pendeteksi Risiko Preeklamsia pada Ibu Hamil dengan Menggunakan Metode Data Mining. *Jurnal Teknik: Media Pengembangan Ilmu Dan Aplikasi Teknik*, 23(1), 50–60. <https://doi.org/10.55893/jt.vol23no1.543>
- Khalimi, A. M. (2022). *Teknik Perhitungan Pada*



- Atribut Data Numerik Algoritma C4.5. Pengalaman Edukasi.*
<https://www.pengalaman-edukasi.com/2022/12/cara-menentukan-split-point-pada-data.html>
- Ma'mur, K. (2019). Analisis Penerapan Algoritma ID3 dalam Mendiagnosis Kesuburan Pria. *Jurnal Informatika Universitas Pamulang*, 4(2), 35. <https://doi.org/10.32493/informatika.v4i2.2274>
- Nurlelah, E., & Mardiyanto, M. S. (2019). Pemilihan Atribut Pada Algoritma C4.5 Menggunakan Particle Swarm Optimization Untuk Meningkatkan Akurasi Prediksi Diagnosis Penyakit Liver. *Jurnal Pilar Nusa Mandiri*, 15(2), 195–202. <https://doi.org/10.33480/pilar.v15i2.706>
- Putri, I. M., & Ismiyatun, N. (2020). Deteksi Dini Kehamilan Beresiko. *JKM (Jurnal Kesehatan Masyarakat) Cendekia Utama*, 8(1), 40. <https://doi.org/10.31596/jkm.v8i1.565>
- Ratnaningtyas, M. A., & Indrawati, F. (2023). Karakteristik Ibu Hamil dengan Kejadian Kehamilan Risiko Tinggi. *HIGEIA (Journal of Public Health Research and Development)*, 7(3), 334–344. <https://doi.org/10.15294/higeia.v7i3.64147>
- Reza, A. A. R., & Muhammad Syaifur Rohman. (2024). Prediction Stunting Analysis Using Random Forest Algorithm and Random Search Optimization. *Journal of Informatics and Telecommunication Engineering*, 7(2), 534–544. <https://doi.org/10.31289/jite.v7i2.10628>
- Rizki, A. M., & Nurlaili, A. L. (2021). Algoritme Particle Swarm Optimization (PSO) untuk Optimasi Perencanaan Produksi Agregat Multi-Site pada Industri Tekstil Rumahan. *Journal of Computer, Electronic, and Telecommunication*, 1(2), 1–9. <https://doi.org/10.52435/complete.v1i2.73>
- Rizki Kurniawan, M., Nurul Sabrina, P., & Ilyas, R. (2024). Prediksi Customer Churn Pada Perusahaan Telekomunikasi Menggunakan Algoritma C4.5 Berbasis Particle Swarm Optimization. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(5), 3369–3375. <https://doi.org/10.36040/jati.v7i5.7476>
- Rizqullah, M. F., Raihana, N. T., & Jambak, M. I. (2024). Komparasi Penerapan Algoritma C4.5, K-Nearest Neighbor, dan Naïve Bayes untuk Keberlangsungan Pasien Gagal Jantung. *Klik: Kajian Ilmiah Informatika Dan Komputer*, 4(5), 2580–2587. <https://doi.org/10.30865/klik.v4i5.1788>
- Saputra, D., Irmayani, W., Purwaningtias, D., & Sidauruk, J. (2021). A Comparative Analysis of C4.5 Classification Algorithm, Naïve Bayes and Support Vector Machine Based on Particle Swarm Optimization (PSO) for Heart Disease Prediction. *International Journal of Advances in Data and Information Systems*, 2(2), 84–95. <https://doi.org/10.25008/ijadis.v2i2.1221>
- Wandasari, W. (2022). Optimalisasi Akurasi Data Mining Dengan Menggunakan Algoritma Pso. *Journal of Data Science Theory and Application*, 1(2), 66–71. <https://doi.org/10.32639/jasta.v1i2.189>
- www.kuliahkomputer.com. (2018). *Algoritma C4.5*. [Www.Kuliahkomputer.Com. https://www.kuliahkomputer.com/2018/07/algoritma-c45.html](https://www.kuliahkomputer.com/2018/07/algoritma-c45.html)