

Penerapan Algoritma C4.5 Berbasis *Particle Swarm Optimization* (PSO) Untuk Deteksi Kanker Payudara

Muhammad Haris Luthfi ^{*1}, Chairani ²

^{1,2}IIB Darmajaya ; Jl. ZA. Pagar Alam No.93, Gedong Meneng, Kec. Rajabasa, Kota Bandar Lampung, Lampung 35141, telp (0721) 787214

Jurusan Teknik Informatika, IIB Darmajaya

e-mail: ^{*1}harisluthfi95@gmail.com, ²chairani@darmajaya.ac.id

Abstrak

Penelitian ini dilatarbelakangi oleh tantangan yang sering muncul terkait dengan rendahnya akurasi dan efektivitas dalam model klasifikasi saat menangani data yang kompleks. Untuk mengatasi tantangan ini, dibangun penelitian ini yang bertujuan untuk menilai dan memperbaiki performa model klasifikasi dengan memadukan metode *Particle Swarm Optimization* (PSO) dengan *Decision Tree C4.5* dalam platform *RapidMiner*. Pendekatan ini melibatkan eksperimen di mana PSO diterapkan untuk mengoptimalkan parameter *Decision Tree C4.5*, diikuti dengan evaluasi performa model yang dihasilkan. Hasil eksperimen menunjukkan adanya peningkatan yang signifikan, dengan akurasi model mencapai 99,34%, *precision* hingga 99,65%, *recall* sebesar 99,30%, dan *Area Under the Curve* (AUC) sebesar 0,997. Temuan ini menunjukkan bahwa integrasi PSO dan *Decision Tree C4.5* dapat secara signifikan meningkatkan efektivitas klasifikasi, menjadikannya metode yang layak untuk aplikasi pengolahan data yang memerlukan tingkat akurasi yang tinggi.

Kata Kunci: *Particle Swarm Optimization* (PSO), *Decision Tree C4.5*, *RapidMiner*, akurasi, *precision*, *recall*.

Abstract

This research is motivated by the challenges associated with low accuracy and effectiveness in classification models when dealing with complex data. To address these challenges, the study aims to assess and improve the performance of classification models by integrating *Particle Swarm Optimization* (PSO) with *Decision Tree C4.5* in *RapidMiner*. The approach involves conducting experiments where PSO is applied to optimize the parameters of *Decision Tree C4.5*, followed by evaluating the performance of the resulting model. The experimental results show a significant improvement, with model accuracy reaching 99.34%, *precision* up to 99.65%, *recall* at 99.30%, and *Area Under the Curve* (AUC) at 0.997. These findings demonstrate that the combination of PSO and *Decision Tree C4.5* can significantly enhance classification effectiveness, making it a viable method for data processing applications requiring high accuracy.

Keywords: *Particle Swarm Optimization* (PSO), *Decision Tree C4.5*, *RapidMiner*, accuracy, *precision*, *recall*.

1. PENDAHULUAN

Kanker payudara merupakan salah satu jenis kanker yang paling sering ditemui pada wanita di seluruh dunia. Pentingnya deteksi dini kanker payudara tidak dapat diabaikan, karena hal ini dapat meningkatkan kemungkinan keberhasilan pengobatan dan memperpanjang harapan hidup pasien. Meski demikian, proses deteksi kanker payudara sering kali menjadi tantangan dan membutuhkan teknik analisis data yang canggih untuk mencapai tingkat akurasi yang optimal. Dalam ranah ilmu data dan kecerdasan buatan, algoritma C4.5 telah terbukti mampu membuat keputusan berdasarkan data yang kompleks. Namun, kinerja algoritma ini masih bisa dioptimalkan lebih lanjut dengan menggunakan metode optimasi seperti *Particle Swarm Optimization (PSO)*[1][2].

PSO adalah sebuah teknik optimasi yang terinspirasi dari perilaku sosial kelompok burung atau ikan dalam mencari makanan. Dengan mengintegrasikan algoritma C4.5 dan *PSO*, diharapkan mampu meningkatkan akurasi dan efisiensi dalam mendeteksi kanker payudara. Penelitian ini bertujuan untuk menerapkan kombinasi algoritma C4.5 dan *PSO* guna mendeteksi kanker payudara dengan lebih akurat dan efisien, serta untuk membandingkan kinerja metode ini dengan metode deteksi kanker payudara lainnya[3][4]. Selama beberapa dekade terakhir, terjadi peningkatan signifikan pada kasus kanker payudara, menjadikannya salah satu tantangan kesehatan utama di dunia. Pentingnya deteksi dini dan diagnosis yang tepat sangat krusial dalam menurunkan angka kematian akibat kanker payudara. Oleh karena itu, diperlukan pendekatan yang mampu menganalisis data medis dengan efektif dan menghasilkan hasil yang akurat.

Algoritma C4.5 merupakan salah satu algoritma pembelajaran mesin yang sering digunakan untuk membangun pohon keputusan. Algoritma ini memiliki kemampuan untuk menangani data yang kompleks dan menghasilkan interpretasi yang mudah dipahami. Namun, seperti algoritma lainnya, C4.5 memiliki keterbatasan, terutama dalam hal optimasi parameter dan kemampuan mencapai solusi global yang optimal[5][6]. *Metode Particle Swarm Optimization (PSO)* dikenal sebagai teknik optimasi yang sangat efektif dan efisien dalam mencari solusi optimal di berbagai bidang. *PSO* bekerja dengan meniru perilaku sosial kelompok hewan, di mana setiap partikel dalam kelompok tersebut mencari posisi terbaiknya berdasarkan pengalaman individu serta informasi yang diperoleh dari partikel lain.

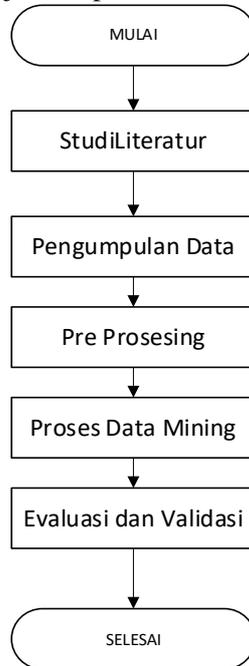
Penelitian ini bertujuan untuk mengintegrasikan algoritma C4.5 dengan *PSO* guna meningkatkan performa dalam deteksi kanker payudara. Dengan memanfaatkan dataset medis yang relevan, penelitian ini akan mengukur kinerja kombinasi kedua algoritma tersebut dalam hal akurasi, efisiensi, dan keandalan. Diharapkan bahwa hasil dari penelitian ini akan memberikan kontribusi yang signifikan dalam deteksi kanker payudara, serta membantu para profesional medis dalam pengambilan keputusan yang lebih baik untuk pasien mereka[7].

Penelitian terdahulu mengenai deteksi, klasifikasi, dan diagnosis kanker payudara telah melibatkan perbandingan algoritma *Decision Tree*, *Naive Bayes*, dan *K-Nearest Neighbors*. Hasil penelitian tersebut mengungkapkan bahwa *K-Nearest Neighbors* menunjukkan performa akurasi yang sangat baik dibandingkan dengan *Naive Bayes* dan *Decision Tree*, dengan tingkat akurasi mencapai 98% menggunakan *metode Hold-Out* dan 96% dengan *metode K-Fold*[8]. Selanjutnya, penelitian yang memberikan hasil yang baik: *KNN* mencapai akurasi sebesar 92,31%, sedangkan *SVM* mencapai akurasi 95,65%. Penelitian lain, "Analysis of Decision Tree and *K-Nearest Neighbor Algorithm* in the Classification of Breast Cancer," menunjukkan bahwa algoritma *Decision Tree* memiliki akurasi sebesar 91,23%, sementara *KNN* memiliki akurasi sebesar 95,61%[9].

Penelitian lainnya membandingkan tingkat akurasi antara model *Random Forest* dan *KNN* dalam mendiagnosis kanker payudara, dengan hasil akurasi masing-masing sebesar 96,5% untuk *KNN* dan 97,2% untuk *Random Forest*[10]. Penelitian ini berfokus pada eksplorasi potensi algoritma *KNN* dan *Decision Tree C4.5* dalam deteksi dini kanker payudara, serta melakukan perbandingan performa keduanya, meskipun keduanya telah sukses diterapkan dalam berbagai bidang berdasarkan penelitian sebelumnya.

2. METODE PENELITIAN

Pada bagian ini, dijelaskan langkah-langkah dalam proses penelitian yang akan dilakukan. Untuk menganalisis dan menemukan pola data yang akan digunakan sebagai dataset guna mempermudah pelaksanaan penelitian secara sistematis dan mencapai tujuan, dibuatlah alur tahapan penelitian sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Studi Pustaka

Penelitian ini didasarkan pada kajian literatur yang melibatkan pengumpulan dan rujukan berbagai sumber informasi, termasuk buku, tesis, skripsi, jurnal, prosiding, serta artikel daring yang relevan. Selain itu, proses ini juga mencakup tinjauan mendalam terhadap dokumen penelitian terkait, seperti jurnal, prosiding, dan publikasi ilmiah lainnya.

Metode Pengumpulan Data

Untuk memastikan terpenuhinya kebutuhan data dalam penelitian ini dan mencapai tujuan yang ditetapkan, beberapa teknik pengumpulan data telah diterapkan. Adapun teknik pengumpulan data yang digunakan dalam penelitian ini adalah sebagai berikut

2.2 Observasi

Metode observasi diterapkan untuk memahami dan mendapatkan gambaran yang mendetail tentang setiap aktivitas bisnis serta mengevaluasi dukungan yang ada guna memastikan bahwa aktivitas tersebut dapat dijalankan sesuai dengan tujuan organisasi[11].

2.3 Wawancara

Teknik ini digunakan untuk mendapatkan informasi mengenai aktivitas utama dalam sebuah organisasi. Selama proses ini, wawancara dilakukan dengan individu yang dianggap memiliki pengetahuan mendalam untuk mengumpulkan data serta informasi terkait dengan proses bisnis yang sedang berlangsung. Penelitian ini memanfaatkan wawancara sebagai metode pengumpulan data yang relevan tentang proses bisnis serta aliran input-proses-output, dengan tujuan mendukung pengembangan dan peningkatan profitabilitas organisasi[12].

2.4 *preprocessing*

Langkah ini dirancang untuk mengubah data input menjadi format yang sesuai untuk analisis selanjutnya. Pada tahap ini, data dari berbagai sumber digabungkan, dibersihkan untuk menghilangkan gangguan dan duplikasi, serta atribut yang relevan dipilih untuk proses data mining [13].

2.5 *Proses Datamining*

Langkah utama dalam penerapan metode melibatkan pengungkapan pengetahuan yang berharga dan tersembunyi dari data. Beragam teknik pengelompokan dalam data mining dapat digunakan untuk tujuan ini [14].

2.6 *Algoritma Particle Swarm Optimization (PSO)*

Optimasi Partikel Swarm (PSO) adalah metode optimasi yang didasarkan pada prinsip populasi dan terinspirasi oleh interaksi sosial dalam kelompok hewan, seperti burung yang terbang dalam kawanan atau ikan yang berenang dalam sekolah. Dalam PSO, setiap calon solusi direpresentasikan sebagai partikel di ruang solusi, dengan posisi dan kecepatan partikel yang diperbarui secara iteratif berdasarkan pengalaman masing-masing dan informasi dari seluruh populasi. Setiap partikel menilai kualitas solusi yang dihasilkan, dan melalui proses komunikasi dengan partikel lain, mereka memperbarui posisi mereka untuk mendekati solusi yang optimal. Proses ini melibatkan dua elemen utama: personal best (posisi terbaik individu) dan global best (posisi terbaik di seluruh populasi), yang membantu partikel dalam mengarahkan pencarian solusi ke area yang lebih baik dalam ruang solusi

2.7 *Algoritma Decision Tree C4.5*

Algoritma C4.5 adalah metode dalam pembelajaran mesin yang digunakan untuk klasifikasi dengan membangun model pohon keputusan berdasarkan atribut dari dataset. Proses kerja algoritma ini melibatkan pembagian dataset menjadi subset-subset yang lebih kecil berdasarkan nilai atribut, dengan tujuan untuk memaksimalkan informasi atau mengurangi ketidakpastian. Untuk memilih atribut yang paling relevan dalam membagi data, C4.5 memanfaatkan rasio informasi gain. Pohon keputusan yang dihasilkan memiliki cabang yang menunjukkan nilai-nilai atribut yang mungkin, dan daun pohon menunjukkan hasil kelas. Algoritma ini juga mencakup teknik pemangkasan pohon untuk menghindari *overfitting*, sehingga model tetap dapat diandalkan saat diterapkan pada data baru. C4.5 dikenal karena kemampuannya dalam mengelola data dengan atribut kategori dan numerik serta menghasilkan model yang mudah dipahami.

2.8 *Evaluasi*

Untuk menganalisis pola-pola yang signifikan dan mengintegrasikannya ke dalam Basis Pengetahuan yang telah ditetapkan, pengetahuan merujuk pada hasil yang diperoleh berupa informasi atau wawasan. [15].

Menghitung *Accuracy* = $(TP + TN) / (TP + TN + FP + FN)$

Menghitung *Precision* = $TP / (TP + FP)$

Menghitung *Recall* = $TP / (TP + FN)$

Menghitung *F1-Score* = $2 \times (Precision \times Recall) / (Precision + Recall)$

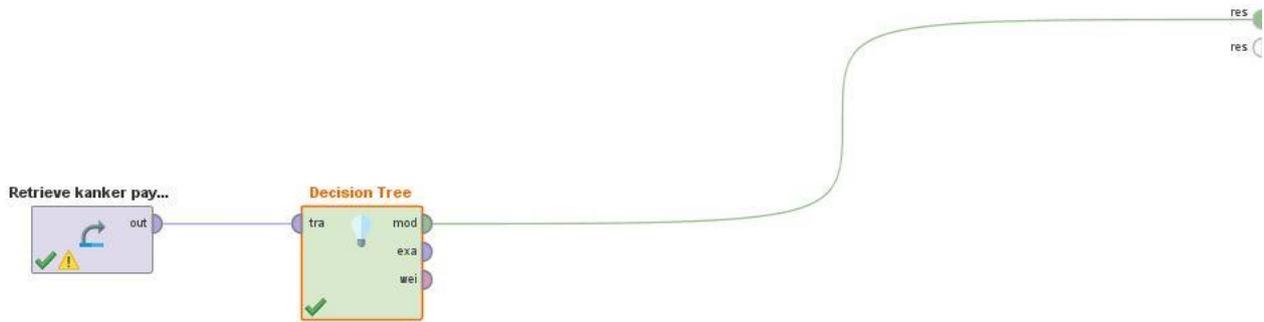
3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan dataset yang berfokus pada kanker payudara, yang diambil dari [https://archive.ics.uci.edu/ repository UCI Machine Learning](https://archive.ics.uci.edu/repository/UCI%20Machine%20Learning%20dengan%20nama%20Breast%20Cancer%20Wisconsin%20(Diagnostic)) dengan nama "Breast Cancer Wisconsin (Diagnostic)." Dataset ini mencakup 32 atribut yang terkait dengan analisis. Detail tentang variabel-variabel tersebut dapat ditemukan pada tabel.

Tabel 1 Atribut Dataset Penyakit Kanker

No	Atribut	Arti
1	<i>Id</i>	ID
2	<i>Diagnosis</i>	Diagnosis
3	<i>Radius_Mean</i>	Jari-jari Rata-rata
4	<i>Texture_Mean</i>	Tekstur Rata-rata
5	<i>Perimeter_Mean</i>	Keliling Rata-rata
6	<i>Area_Mean</i>	Luas Rata-rata
7	<i>Smoothness_Mean</i>	Kekasaran Rata-rata
8	<i>Compactness_Mean</i>	Kepadatan Rata-rata
9	<i>Concavity_Mean</i>	Ketidakkampuan Rata-rata
10	<i>Concave Points_Mean</i>	Titik Cekung Rata-rata
11	<i>Symmetry_Mean</i>	Simetri Rata-rata
12	<i>Fractal_Dimension_Mean</i>	Dimensi Fraktal Rata-rata
13	<i>Radius_Se</i>	Jari-jari Standar Erro
14	<i>Texture_Se</i>	Tekstur Standar Error
15	<i>Perimeter_Se</i>	Keliling Standar Error
16	<i>Area_Se</i>	Luas Standar Error
17	<i>Smoothness_Se</i>	Kekasaran Standar Error
18	<i>Compactness_Se</i>	Kepadatan Standar Error
19	<i>Concavity_Se</i>	Ketidakkampuan Standar Error
20	<i>Concave Points_Se</i>	Titik Cekung Standar Erro
21	<i>Symmetry_Se</i>	Simetri Standar Error
22	<i>Fractal_Dimension_Se</i>	Dimensi Fraktal Standar Error
23	<i>Radius_Worst</i>	Jari-jari Terburuk
24	<i>Texture_Worst</i>	Tekstur Terburuk
25	<i>Perimeter_Worst</i>	Keliling Terburuk
26	<i>Area_Se</i>	Luas Terburuk
27	<i>Smoothness_Worst</i>	Kekasaran Terburuk
28	<i>Compactness_Worst</i>	Kepadatan Terburuk
29	<i>Concavity_Worst</i>	Ketidakkampuan Terburuk
30	<i>Concave Points_Worst</i>	Titik Cekung Terburuk
31	<i>Symmetry_Worst</i>	Simetri Terburuk
32	<i>Fractal_Dimension_Worst</i>	Dimensi Fraktal Terburuk

3.1 Implementasi Decision Tree yang telah dioptimalkan dilakukan menggunakan perangkat RapidMi

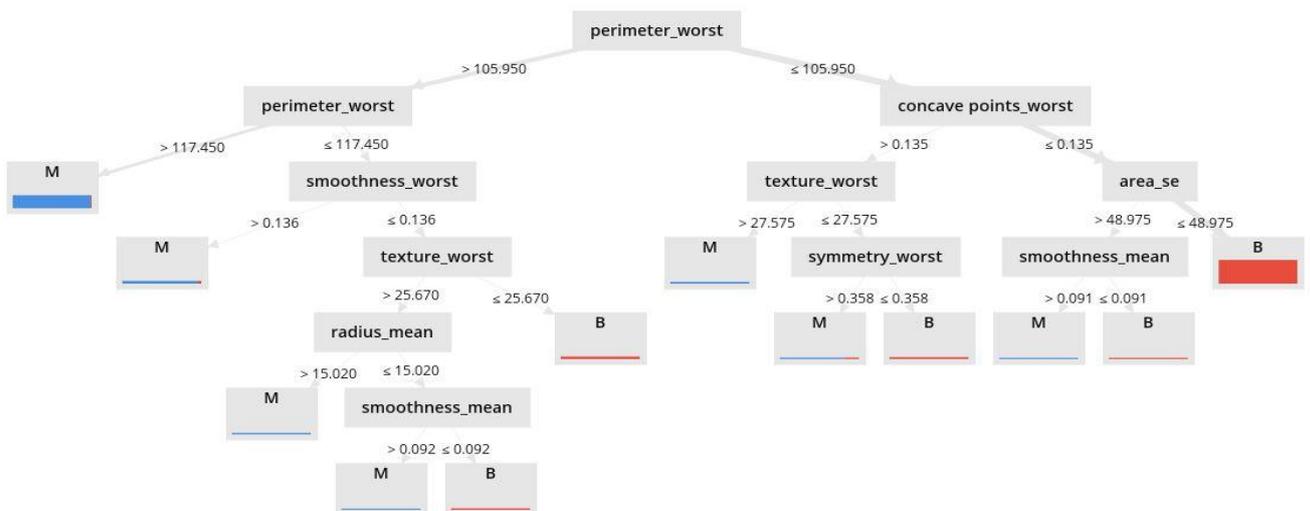


Gambar 2 Proses algoritma *Decision Tree*

Keterangan dari gambar :

1. *Retrieve kanker pay*: Langkah pertama dalam proses ini adalah mengambil data menggunakan operator "*Retrieve*". Nama lengkap dari dataset tidak terlihat sepenuhnya, namun kemungkinan dataset ini berhubungan dengan kanker payudara (breast cancer).
2. *Decision Tree*: Setelah data diambil, operator "*Decision Tree*" digunakan untuk membangun model pohon keputusan. Tiga koneksi input ke operator ini adalah:
 - *tra (training)*: Data yang digunakan untuk melatih model pohon keputusan.
 - *mod (model)*: Output model yang dihasilkan.
 - *exa (example)*: Contoh atau instance data yang digunakan untuk prediksi.
 - *wei (weight)*: Bobot dari masing-masing instance data.
3. *Output (res)*: Hasil dari model pohon keputusan dikirim ke dua output yang diberi label "*res*" (*result*). Kemungkinan besar ini adalah hasil evaluasi dan prediksi dari model pohon keputusan yang telah dibangun.

3.2 Tampilan pohon keputusan pada algoritma Decision Tree



Gambar 3 Pohon Keputusan algoritma *Decision Tree*

Selanjutnya kita dapat melihat deskripsi pohon keputusan diatas, dapat kita lihat seperti gambar 4.

Tree

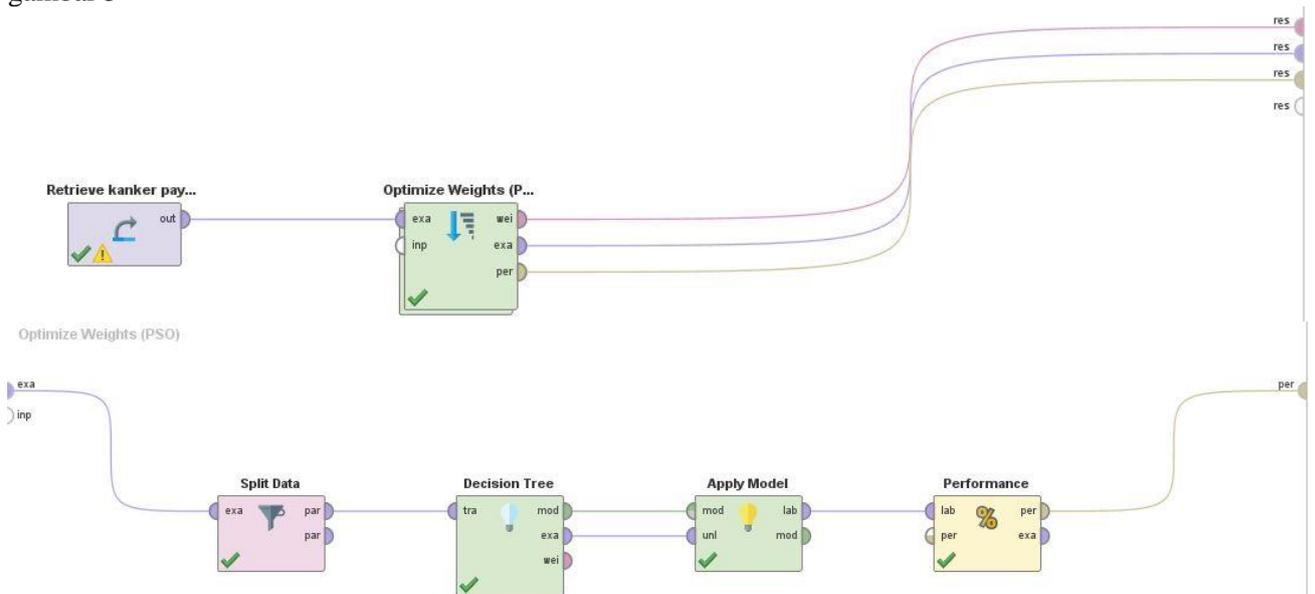
```

perimeter_worst > 105.950
|  perimeter_worst > 117.450: M {M=165, B=2}
|  perimeter_worst ≤ 117.450
|  |  smoothness_worst > 0.136: M {M=22, B=1}
|  |  smoothness_worst ≤ 0.136
|  |  |  texture_worst > 25.670
|  |  |  |  radius_mean > 15.020: M {M=6, B=0}
|  |  |  |  radius_mean ≤ 15.020
|  |  |  |  |  smoothness_mean > 0.092: M {M=2, B=0}
|  |  |  |  |  smoothness_mean ≤ 0.092: B {M=0, B=6}
|  |  |  |  |  texture_worst ≤ 25.670: B {M=0, B=20}
|  |  |  texture_worst ≤ 25.670: B {M=0, B=20}
perimeter_worst ≤ 105.950
|  concave points_worst > 0.135
|  |  texture_worst > 27.575: M {M=9, B=0}
|  |  texture_worst ≤ 27.575
|  |  |  symmetry_worst > 0.358: M {M=4, B=1}
|  |  |  symmetry_worst ≤ 0.358: B {M=0, B=11}
|  concave points_worst ≤ 0.135
|  |  area_se > 48.975
|  |  |  smoothness_mean > 0.091: M {M=2, B=0}
|  |  |  smoothness_mean ≤ 0.091: B {M=0, B=2}
|  |  area_se ≤ 48.975: B {M=2, B=314}
    
```

Gambar 4 Deskripsi Pohon Keputusan

3.3 Algoritma Decision Tree C4.5 Dengan Particle Swarm Optimization (PSO)

Untuk meningkatkan akurasi dalam klasifikasi, metode ini menggunakan teknik seleksi fitur dengan pembobotan atribut yang mengombinasikan PSO dan *Decision Tree C4.5*. Data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Penerapan metode ini dapat dilihat pada gambar 5



Gambar 5 Proses PSO dan Decision Tree C4.5

Dengan menerapkan PSO dan *Decision Tree C4.5* pada *RapidMiner*, terdapat peningkatan akurasi. Perbandingan akurasi, *precision*, *recall*, serta *AUC* sebelum dan sesudah penambahan metode PSO menunjukkan hasil yang lebih baik.

Hasil *accuracy* yaitu mendapatkan 99,34% seperti gambar 6.

accuracy: 99.34%			
	true M	true B	class precision
pred. M	169	2	98.83%
pred. B	1	284	99.65%
class recall	99.41%	99.30%	

Gambar 6 Nilai Akurasi PSO dan *Decision Tree C4.5*

Hasil *precision* yaitu mendapatkan 99,65% seperti gambar 7.

precision: 99.65% (positive class: B)			
	true M	true B	class precision
pred. M	169	2	98.83%
pred. B	1	284	99.65%
class recall	99.41%	99.30%	

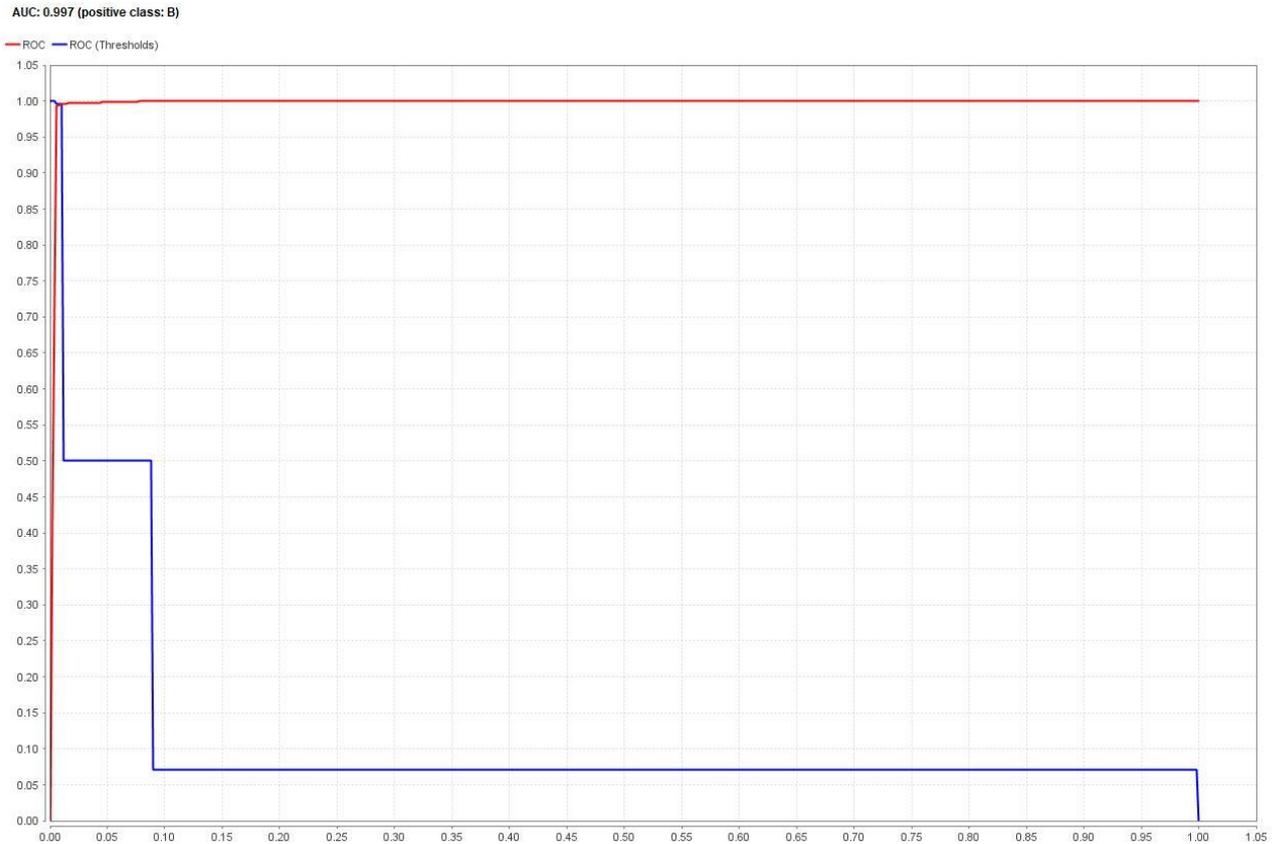
Gambar 7 Nilai precision PSO dan *Decision Tree C4.5*

Berikut adalah hasil *confusion matrix* yaitu mendapatkan 99.30% seperti gambar 8.

recall: 99.30% (positive class: B)			
	true M	true B	class precision
pred. M	169	2	98.83%
pred. B	1	284	99.65%
class recall	99.41%	99.30%	

Gambar 8 Nilai confusion matrix PSO dan *Decision Tree C4.5*

Berikut adalah hasil *kurva AUC* yaitu mendapatkan nilai 0,997 seperti gambar 9.



Gambar 9 Kurva AUC PSO dan *Decision Tree C4.5*

4. KESIMPULAN

Berdasarkan hasil dari penerapan *Decision Tree* dan *PSO*, dapat disimpulkan bahwa model tersebut menunjukkan kinerja yang sangat baik dengan akurasi sebesar 99,34%, *precision* sebesar 99,65%, *recall* sebesar 99,30%, dan *AUC* sebesar 0,997. Nilai-nilai ini mengindikasikan bahwa model memiliki kemampuan yang tinggi dalam mengklasifikasikan data dengan benar, sehingga dapat diandalkan untuk prediksi yang akurat.

5. SARAN

Penelitian selanjutnya, disarankan untuk mengeksplorasi variasi teknik optimasi dan algoritma pengklasifikasi lainnya guna membandingkan efektivitasnya dengan kombinasi *Decision Tree* dan *PSO*. Selain itu, penting untuk menguji model ini pada berbagai dataset untuk memastikan bahwa hasil yang diperoleh konsisten dan dapat diterapkan pada situasi nyata yang berbeda. Penelitian lebih lanjut juga bisa fokus pada pengoptimalan parameter *PSO* dan pengujian pada masalah klasifikasi yang lebih kompleks untuk mengevaluasi keberlanjutan dan adaptabilitas metode ini.

DAFTAR PUSTAKA

- [1] H. Oktavianto and R. P. Handri, "Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma Naive Bayes," *INFORMAL Informatics J.*, vol. 4, no. 3, p. 117, 2020, doi: 10.19184/isj.v4i3.14170.
- [2] I. Yulianti, H. Setyawan, and D. Sutiningsih, "Faktor-Faktor Risiko Kanker Payudara," *J. Kesehat. Masy.*, vol. 4, no. 4, pp. 401–409, 2021.
- [3] L. W. Astuti, I. Saluza, F. Faradilla, and M. F. Alie, "Optimalisasi Klasifikasi Kanker Payudara Menggunakan Forward Selection pada Naive Bayes," *J. Ilm. Inform. Glob.*, vol. 11, no. 2, 2021, doi: 10.36982/jiig.v11i2.1235.
- [4] R. T. P. Sudewo, Y. Pratama, and E. Yanti, "Analisis Data Mining Untuk Prediksi Kanker Payudara Menggunakan Algoritma Klasifikasi," *J. Pustaka Data (Pusat Akses Kaji. Database, Anal. Teknol. dan Arsit. Komputer)*, vol. 3, no. 2, pp. 62–69, 2023, doi: 10.55382/jurnalpustakadata.v3i2.656.
- [5] L. Hermawanti, "Penerapan Algoritma Klasifikasi C4.5 untuk Diagnosis Penyakit Kanker Payudara," *J. Tek. Unisfat*, vol. 7, no. 1, pp. 57–64, 2012.
- [6] H. Firda, R. Athiyah, and M. Ihsan Jambak, "Perbandingan Algoritma Klasifikasi Data Mining Model C4.5 Dan Naive Bayes Untuk Prediksi Keganasan Kanker Payudara," vol. 14, no. 3, pp. 150–233, 2024, [Online]. Available: <https://jurnal.umj.ac.id/index.php/just-it/index>
- [7] Y. E. Praptiningsih, W. W. Ariestya, I. Astuti, and S. Nurulita, "Data Mining untuk Klasifikasi Diagnosa Kanker Payudara Dengan Menerapkan Algoritma C4.5," *J. Ilm. FIFO*, vol. 15, no. 1, p. 37, 2023, doi: 10.22441/fifo.2023.v15i1.005.
- [8] M. Abdul Jabbar, E. Hasmin, C. Susanto, W. Musu, and I. Artikel, "Komparasi Algoritma Decision Tree, Naive Bayes, dan K-Nearest Neighbors dalam Klasifikasi Kanker Payudara Comparison of Decision Tree Algorithms, Naive Bayes, and K-Nearest Neighbors in Breast Cancer Classification," *Oktober*, vol. 14, no. 3, pp. 258–270, 2022, [Online]. Available: <https://www.doi.org/10.22303/csrid.14.3.2022.258-270>
- [9] F. K. Nasser and S. F. Behadili, "Breast Cancer Detection using Decision Tree and K-Nearest Neighbour Classifiers," *Iraqi J. Sci.*, vol. 63, no. 11, pp. 4987–5003, 2022, doi: 10.24996/ijs.2022.63.11.34.
- [10] C. W. Cahyana and A. Nurlayli, "Analisis Performa Logistic Regression, Naive Bayes, dan Random Forest sebagai Algoritma Pendeteksi Kanker Payudara," *Inser. Inf. Syst. Emerg. Technol. J.*, vol. 4, no. 1, pp. 51–64, 2023.
- [11] T. Sofa, A. Wardiyah, and Rilyani, "Faktor Risiko Kanker Payudara Pada Wanita," *J. Penelit. Perawat Prof.*, vol. 2, no. 5474, pp. 1333–1336, 2024.
- [12] A. Herawati, S. Rijal, A. S. F. Arsal, R. Purnamasari, and D. A. Abdi, "Karakteristik Kanker Payudara," *FAKUMI Med. J. J. Mhs. Kedokt.*, vol. 1, no. 1, pp. 44–53, 2021, doi: 10.33096/fmj.v1i1.8.
- [13] M. S. Firdaus, A. Primajaya, and A. Jamaludin, "Implementasi Algoritme C4.5 untuk Prediksi Penanaman Cabai Merah," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 6, no. 2, p. 158, 2021, doi: 10.30998/string.v6i2.10378.
- [14] F. Dwi Astuti and F. Nova Lenti, "Implementasi SMOTE untuk mengatasi Imbalance Class pada Klasifikasi Car Evolution menggunakan K-NN," *J. JUPITER*, vol. 13, no. 1, pp. 89–98, 2021.
- [15] A. Ardiansyah, J. Triloka, and Indera, "Evaluasi Kinerja Model YOLOv8 dalam Deteksi Kesegaran Buah," *JUPITER J. Penelit. Ilmu Dan Teknol. Komput.*, pp. 357–368, 2024.