



<http://jurnal.universitasputrabangsa.ac.id/index.php/ijasta>  
e-ISSN: 2829-4858

---

---

**ARTICLE INFORMATION**

Received September 19<sup>th</sup> 2022  
Accepted September 30<sup>th</sup> 2022  
Published October 7<sup>th</sup> 2022

**DOI:**

<https://doi.org/xxx.xxx.xxx>

**OPTIMALISASI AKURASI DATA MINING DENGAN  
MENGGUNAKAN ALGORITMA PSO**

**Wahyuni Windasari**

Universitas Putra Bangsa

email: wahyuwindasari@gmail.com

---

---

**ABSTRAK**

Dewasa ini metode klasifikasi data mining telah banyak diterapkan di berbagai bidang. Salah satunya bidang kesehatan. Penelitian ini membandingkan akurasi kinerja dua metode klasifikasi data mining yaitu metode Naïve Bayes dan C4.5 untuk memprediksi tingkat resiko kehamilan berdasarkan faktor resikonya. Hasil penelitian menunjukkan bahwa akurasi algoritma C4.5 lebih baik dibandingkan algoritma Naïve Bayes. Selain itu, penambahan algoritma PSO pada design penelitian juga mampu menaikkan akurasi algoritma C4.5.

**Kata Kunci:** Data Mining; Klasifikasi; C4.5; Naïve Bayes; PSO.

**ABSTRACT**

*Data mining classification methods are widely applied in various fields. One of them is the health sector. This study compares the accuracy of the performance of two data mining classification methods, namely the Naïve Bayes method and C4.5 to predict the level of risk of pregnancy based on risk factors. The results showed that the accuracy of the C4.5 algorithm was better than the Naïve Bayes algorithm. In addition, the addition of the PSO algorithm to the research design is also able to increase the accuracy of the C4.5 algorithm.*

**Keywords:** Data Mining; Classification ; C4.5; Naïve Bayes; PSO.

**PENDAHULUAN**

Masa kehamilan merupakan salah satu periode penting bagi seorang perempuan. Pada periode ini, kesehatan dan kesejahteraan baik ibu maupun bayi dalam kandungannya sangat perlu diperhatikan. Infodatin (2014) menyatakan bahwa salah satu indikator yang peka dalam menggambarkan kesehatan masyarakat di suatu negara adalah Angka Kematian Ibu (AKI). Angka Kematian Ibu adalah banyaknya perempuan yang meninggal dari suite penyebab kematian terkait dengan gangguan kehamilan atau

penanganannya (tidak termasuk kecelakaan, bunuh diri atau kasus insidental) selama kehamilan, melahirkan, dan dalam masa nifas atau 42 hari setelah melahirkan (BPS, 2015). Semakin tinggi nilai AKI di suatu wilayah menunjukkan semakin rendahnya derajat kesehatan perempuan di wilayah tersebut. Hal ini juga dapat membawa dampak pada menurunnya perekonomian baik keluarga maupun negara.

Menurut data WHO (2017), kematian akibat komplikasi karena kehamilan maupun persalinan rata-rata mencapai 810 kasus setiap harinya. Bahkan di tahun 2019, Angka Kematian Ibu di dunia mencapai 303.000 jiwa. Salah satu faktor yang menyebabkan kematian perempuan baik selama kehamilan maupun postnatal adalah kurang optimalnya deteksi akan status resiko kehamilan. Resiko kehamilan memiliki tiga tingkatan yaitu kehamilan beresiko rendah, kehamilan beresiko tinggi, dan kehamilan beresiko sangat tinggi. Dengan mengetahui status resiko kehamilan lebih awal diharapkan baik ibu hamil maupun petugas pelayanan kesehatan dapat lebih optimal lagi dalam menjaga kesehatan dan keselamatan diri baik ibu hamil maupun buah hati dalam kandungannya.

Dewasa ini, *medical diagnosis* dapat dilakukan dengan memanfaatkan ketersediaan data rekam medis. Salah satunya dengan menggunakan algoritma klasifikasi data mining. Melalui algoritma klasifikasi, data rekam medis digunakan sebagai input model yang selanjutnya dari data tersebut dapat digunakan untuk mengembangkan model terbaik yang dapat dimanfaatkan untuk mengklasifikasikan data pengujian. Metode klasifikasi yang cukup banyak diterapkan di bidang kesehatan adalah metode Naïve Bayes dan C4.5. Ningsih dan Noranita (2018) menggunakan algoritma C4.5 untuk menentukan status persalinan. Dari hasil penelitian tersebut diperoleh hasil bahwa algoritma C4.5 memberikan akurasi prediksi penentuan status persalinan yang tinggi yaitu sebesar 97,08%. Selanjutnya, Arga dan Zuhri (2020) menggunakan metode klasifikasi Naïve Bayes untuk memprediksi kondisi kelahiran bayi. Dari hasil penelitiannya diperoleh nilai akurasi, precision, dan recall dari metode Naïve Bayes masing-masing sebesar 92.73 %, 83,33%, dan 100%. Di sisi lain, penelitian terkait perbandingan kinerja akurasi antar metode klasifikasi data mining juga telah banyak dikembangkan. Hal ini tidak lain untuk mendapatkan model terbaik untuk mengurangi kesalahan prediksi di bidang kesehatan. Ade dan Azhari (2017) melakukan perbandingan kinerja algoritma C4.5 dan Naïve Bayes pada dataset penyakit jantung dan hepatitis. Dari hasil nilai akurasi, recall, dan precision diperoleh algoritma Naïve Bayes memberikan nilai yang lebih besar dibandingkan dengan algoritma C4.5. Hal ini menunjukkan bahwa algoritma Naïve Bayes lebih baik digunakan untuk klasifikasi dataset penyakit jantung dan hepatitis dibandingkan algoritma C4.5. Berbeda halnya dengan penelitian Kohsasih dan Situmorang (2022) yang berjudul Analisis perbandingan algoritma C4.5 dan Naïve Bayes dalam memprediksi penyakit Cerebrovascular menemukan bahwa algoritma C4.5 memberikan tingkat akurasi yang lebih tinggi dibandingkan dengan algoritma Naïve Bayes. Selain itu, optimalisasi kinerja metode data mining juga dapat dilakukan dengan mengaplikasikan algoritma *Particle Swarm Optimization* (PSO) seperti yang dilakukan oleh Siswa dan Prihandoko (2018). Pada penelitiannya *Particle Swarm Optimization* (PSO) berhasil meningkatkan kinerja akurasi C4.5 dan Naïve Bayes untuk mendeteksi kanker payudara.

Atas dasar latar belakang di atas maka pada penelitian ini akan dilakukan pemilihan metode terbaik untuk klasifikasi dataset status resiko kehamilan dengan melalui dua tahap yaitu (1) Melakukan komparasi optimalisasi kinerja metode C4.5 dan Naïve Bayes pada dataset status resiko kehamilan, (2) Melakukan pengoptimalan akurasi dengan menggunakan algoritma *Particle Swarm Optimization* (PSO) pada metode terpilih sebelumnya.

## METODE

Penelitian ini menggunakan data sekunder yang bersumber dari *UCI Machine Learning Repository* berupa dataset *Maternal Health Risk*. Dataset ini terdiri dari 1014 record data dengan 6 variabel atribut dan 1 variabel label.

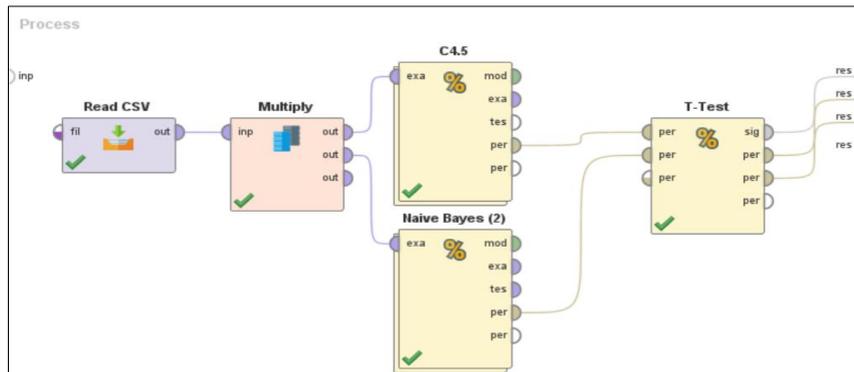
Tabel 1. Atribut Penelitian

No	Atribut	Keterangan
1	Age	Usia perempuan ketika hamil
2	Systolic Blood Pressure	Tekanan tertinggi pada saat otot jantung berkontraksi
3	Diastolic Blood Pressure	Tekanan darah di arteri saat jantung beristirahat
4	Blood Sugar	Kadar glukosa darah
5	Body Temperature	Suhu tubuh
6	Heart Rate	Denyut jantung pada saat istirahat normal per menit
7	Risk Level	Tingkat resiko kehamilan

Adapun metode yang digunakan pada penelitian ini adalah metode C4.5, Naïve Bayes, serta optimasi model terpilih menggunakan PSO.

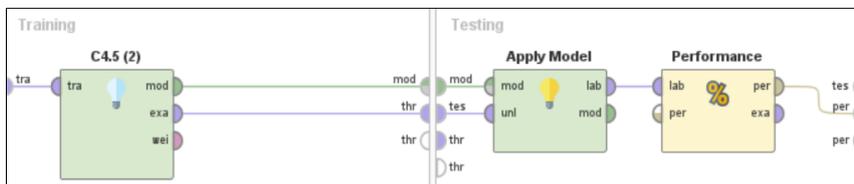
### HASIL DAN PEMBAHASAN

Design *main process* pada penelitian ini menggunakan software Rapidminer yang disajikan pada Gambar 1 berikut.



Gambar 1. Design Model Penelitian pada Rapidminer

### Pemodelan dengan Algoritma C4.5



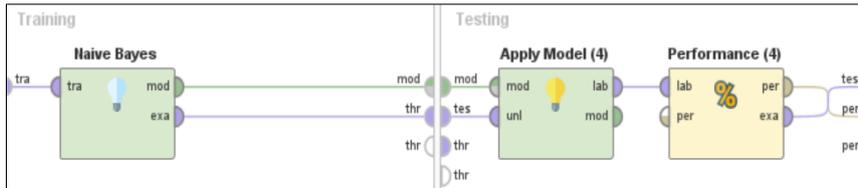
Gambar 2. Model C4.5 pada Rapidminer

Gambar 2 di atas merupakan design model untuk analisis dataset menggunakan algoritma C4.5. Pada penelitian ini menggunakan *performance 10-fold Cross Validation*. Adapun hasil output Rapidminer diberikan pada Gambar 3.

accuracy: 70.61% +/- 3.49% (micro average: 70.61%)				
	true high risk	true low risk	true mid risk	class precision
pred. high risk	225	9	36	83.33%
pred. low risk	14	368	177	65.83%
pred. mid risk	33	29	123	66.49%
class recall	82.72%	90.64%	36.61%	

Gambar 3. Output Model C4.5 pada Rapidminer

**Pemodelan dengan Algoritma Naïve Bayes**



Gambar 4. Pemodelan Naïve Bayes dengan Menggunakan Rapidminer

Sama seperti pemodelan dengan C4.5, pada pemodelan dengan algoritma Naïve Bayes ini juga menggunakan 10-fold Cross Validation. Dari hasil pengujian dan validasi diperoleh nilai akurasi sebesar 60,14 % seperti yang ditampilkan pada Gambar 5 berikut.

accuracy: 60.14% +/- 5.31% (micro average: 60.16%)				
	true high risk	true low risk	true mid risk	class precision
pred. high risk	185	8	32	82.22%
pred. low risk	35	348	227	57.05%
pred. mid risk	52	50	77	43.02%
class recall	68.01%	85.71%	22.92%	

Gambar 5. Hasil dari Pemodelan Naïve Bayes dengan Rapidminer

**Perbandingan Akurasi C4.5 dan Naïve Bayes**

Setelah dilakukan pengujian dan validasi diperoleh performa untuk setiap algoritma yang selanjutnya dapat dilakukan perbandingan. Nilai akurasi algoritma C4.5 tergolong dalam klasifikasi lepas yaitu sebesar 70,61 % sedangkan nilai akurasi dari algoritma Naïve Bayes termasuk dalam klasifikasi miskin dengan nilai akurasi sebesar 60,14%. Dari hasil perbandingan nilai akurasi dapat disimpulkan bahwa C4.5 lebih baik untuk memprediksi status resiko kehamilan dibandingkan dengan algoritma Naïve Bayes. Hal ini juga didukung dari hasil uji statistik t.

```

Pairwise t-Test
Probabilities for random values with the same result:
----- 0.000
-----

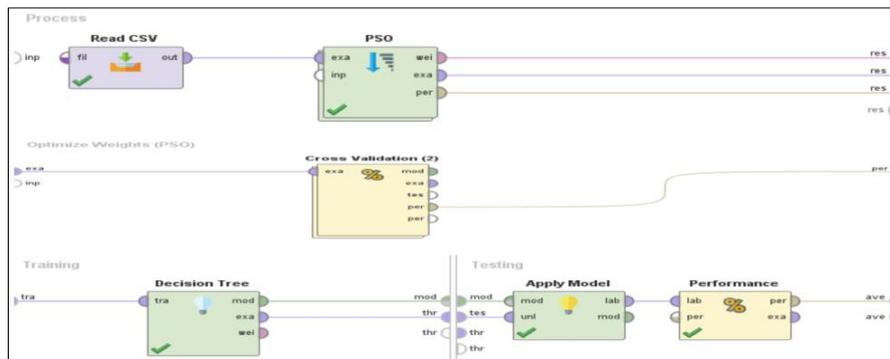
Values smaller than alpha=0.050 indicate a probably significant difference between the mean values!
List of performance values:
0: 0.706 +/- 0.035
1: 0.601 +/- 0.053
    
```

Gambar 6. Hasil Uji Statistik t-test.

Dengan mengambil nilai  $\alpha = 0.05$  dapat ditunjukkan bahwa nilai signifikansi model algoritma C4.5 sebesar  $0.000 < \alpha$ . Hal ini berarti kinerja dari model C4.5 secara nyata memberikan hasil prediksi yang lebih baik dibandingkan dengan algoritma Naïve Bayes.

**Optimalisasi C4.5 dengan Algoritma PSO**

Langkah selanjutnya adalah melakukan optimalisasi model terpilih yaitu algoritma C4.5 dengan menggunakan operator *Optimize Weights* (PSO). Adapun bentuk desainnya sebagai berikut :



Gambar 7. Design Algoritma C4.5 dan PSO

accuracy: 74.10%				
	true high risk	true low risk	true mid risk	class precision
pred. high risk	70	3	8	86.42%
pred. low risk	6	117	54	66.10%
pred. mid risk	6	2	39	82.98%
class recall	85.37%	95.90%	38.61%	

Gambar 8. Hasil Optimasi Kinerja C4.5 dengan PSO

Dari output Rapidminer diperoleh nilai akurasi hasil optimasi PSO mampu meningkatkan akurasi C4.5 sebesar 3,49%.

**SIMPULAN**

Berdasarkan penelitian di atas dapat disimpulkan bahwa :

1. Algoritma C4.5 memiliki akurasi yang lebih baik dibandingkan dengan algoritme Naïve Bayes dalam memprediksi status resiko kehamilan pada dataset *Maternal Risk UCI*.
2. Nilai kinerja algoritma C4.5 dapat ditingkatkan dengan menambahkan algoritma PSO.

Selanjutnya, adapun saran untuk penelitian selanjutnya adalah :

1. Untuk pengembangan penelitian selanjutnya dapat digunakan dataset dari sumber lainnya.
2. Dapat dikembangkan metode klasifikasi data mining lain seperti KNN dengan optimasi lainnya seperti menggunakan algoritma Genetika.

**REFERENSI**

Ade, A., & Azhari. (2017). *Perbandingan Algoritma C4.5 dan Naïve Bayes pada Kasus Hepatitis dan Jantung*. Universitas Gadjah Mada.

Arga, H., & Zuhri, S. (2020). Sistem Prediksi Kondisi Kelahiran Bayi Menggunakan Klasifikasi Naive Bayes. *Joined Journal*, 3(2), 1–9. <https://e-journal.ivet.ac.id/index.php/jiptika/article/view/1432/1095>

BPS. (2015). *Angka Kematian Ibu*. BPS. [https://www.bps.go.id/indikator/indikator/view\\_data/0000/data/1349/sdgs\\_3/1](https://www.bps.go.id/indikator/indikator/view_data/0000/data/1349/sdgs_3/1).

Infodatin. (2014). *Mother's Day*.

Kohsasih, K. L., & Situmorang, Z. (2022). Analisis Perbandingan Algoritma C4.5 dan Naïve Bayes Dalam Memprediksi Penyakit Cerebrovascular. *Jurnal Informatika*, 9(1), 13–17. <https://doi.org/10.31294/inf.v9i1.11931>

Ningsih, M. P. S. D., & Noranita, B. (2018). Status Proses Persalinan Menggunakan Algoritma C4.5. *Jurnal*

*Masyarakat Informatika*, 9(1), 1–13.

Siswa, T. A. Y., & Prihandoko. (2018). Analisis Penerapan Optimasi Perbandingan Kinerja Algoritma C4.5 Dan Naïve Bayes Berbasis Particle Swarm Optimization (Pso) Untuk. *Jurnal Bangkit Indonesia*, 7(2), 1. <https://doi.org/10.52771/bangkitindonesia.v7i2.48>

UCI Machine Learning Repository. (2020). *Maternal Health Risk Data Set Data Set*. <https://archive.ics.uci.edu/ml/datasets/Maternal+Health+Risk+Data+Set>

WHO. (2017). *Maternal Mortality*. <https://www.who.int/news-room/fact-sheets/detail/maternal-mortality>