

**IMPLEMENTASI METODE POHON KEPUTUSAN  
DENGAN ALGORITMA C4.5 DAN NAÏVE BAYES BERBASIS  
PARTICLE SWARM OPTIMIZATION DALAM MEMPREDIKSI  
PENYAKIT JANTUNG**

**SKRIPSI**

*Diajukan untuk memenuhi sebagian dari syarat memperoleh gelar Sarjana  
Matematika*



Disusun oleh:  
Novita Damayanti  
NIM 2001332

**PROGRAM STUDI MATEMATIKA  
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS PENDIDIKAN INDONESIA  
2024**

# LEMBAR PENGESAHAN

NOVITA DAMAYANTI

IMPLEMENTASI METODE POHON KEPUTUSAN  
DENGAN ALGORITMA C4.5 DAN *NAÏVE BAYES* BERBASIS  
*PARTICLE SWARM OPTIMIZATION* DALAM MEMPREDIKSI  
PENYAKIT JANTUNG

Disetujui dan disahkan oleh pembimbing:

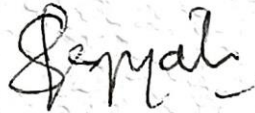
Pembimbing I,



**Dr. Khusnul Novianingsih S.Si., M.Si.**

**NIP. 197711282008122001**

Pembimbing II,



**Ririn Sispivati S.Si., M.Si.**

**NIP. 198106282005012001**

Mengetahui,

Ketua Prodi Matematika



**Dr. Kartika Yullanti, S.Pd., M.Si.**

**NIP. 198207282005012001**

**LEMBAR HAK CIPTA**

**IMPLEMENTASI METODE POHON KEPUTUSAN  
DENGAN ALGORITMA C4.5 DAN *NAÏVE BAYES* BERBASIS  
*PARTICLE SWARM OPTIMIZATION* DALAM MEMPREDIKSI  
PENYAKIT JANTUNG**

Oleh

Novita Damayanti

2001332

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar  
Sarjana Matematika pada  
Fakultas Pendidikan Matematika dan Ilmu pengetahuan Alam

©Novita Damayanti 2024

Universitas Pendidikan Indonesia

Agustus 2024

Hak Cipta dilindungi undang-undang

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian,  
dengan dicetak ulang, difotokopi, atau cara lainnya tanpa izin dari penulis.

## LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa skripsi dengan judul “**Implementasi Metode Pohon Keputusan dengan Algoritma C4.5 dan *Naïve Bayes* berbasis *Particle Swarm Optimization* dalam Memprediksi Penyakit Jantung**” ini beserta seluruh isinya adalah benar-benar karya saya sendiri. Saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini saya siap menanggung risiko/sanksi atau klaim dari pihak lain terhadap keaslian karya saya ini.

Bandung, 08 Agustus 2024

Peneliti



Novita Damayanti

NIM.2001332

## KATA PENGANTAR

Puji dan syukur kehadiran Allah SWT karena berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Implementasi Metode Pohon Keputusan dengan Algoritma C4.5 dan *Naïve Bayes* berbasis *Particle Swarm Optimization* dalam Memprediksi Penyakit Jantung” untuk memenuhi salah satu syarat memperoleh gelar sarjana Matematika di Universitas Pendidikan Indonesia.

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan motivasi dan dukungan dalam menyelesaikan skripsi ini. Bagi penulis segala bentuk motivasi dan dukungan sangat bermanfaat dan berarti. Penulis berharap, skripsi ini dapat bermanfaat baik bagi penulis ataupun pembaca.

Penulis menyadari penelitian ini terdapat kekurangan yang disebabkan oleh keterbatasan ilmu ataupun keterbatasan kemampuan penulis. Oleh karena itu penulis mengharapkan saran dan kritik yang membangun agar skripsi ini dapat lebih baik.

Bandung, 0 Agustus 2024



Penulis

**IMPLEMENTASI METODE POHON KEPUTUSAN  
DENGAN ALGORITMA C4.5 DAN NAÏVE BAYES BERBASIS  
PARTICLE SWARM OPTIMIZATION DALAM MEMPREDIKSI  
PENYAKIT JANTUNG**

**ABSTRAK**

Penyakit jantung merupakan salah satu penyakit tidak menular yang mematikan. Penelitian ini bertujuan untuk memprediksi penyakit jantung dengan melihat indikasi-indikasi apa saja yang dapat menyebabkan seseorang terkena penyakit jantung menggunakan metode pohon keputusan dengan algoritma C4.5 dan *Naïve Bayes* berbasis *Particle Swarm Optimization*. Penelitian dilakukan dengan menggunakan metodologi CRISP-DM (*Cross Industry Standard Process for Data Mining*). Data yang digunakan merupakan data hasil wawancara lembaga CDC (*Centers of Disease Control and Prevention*). Pada tahap awal akan dilakukan pemilihan atribut yang berpengaruh menggunakan *Particle Swarm Optimization* (PSO). Selanjutnya atribut terpilih akan dimodelkan dengan metode pohon keputusan dengan algoritma C4.5 dan *Naïve Bayes*. Selanjutnya dilakukan proses validasi dan evaluasi untuk mengukur kinerja model yang telah dibangun. Hasil penelitian menunjukkan bahwa kedua model memiliki nilai akurasi yang sama. Akan tetapi, nilai *recall* dan *F1-Score* dari model pohon keputusan dengan algoritma C4.5 lebih besar dibandingkan model *Naïve Bayes*. Dalam penelitian ini, model pohon keputusan lebih baik dalam memprediksi penyakit jantung dibandingkan model *Naïve Bayes*.

**Kata Kunci:** Penyakit Jantung, Pohon Keputusan, Algoritma C4.5, *Naïve Bayes*, *Particle Swarm Optimization*.

***THE IMPLEMENTATION OF DECISION TREE WITH C4.5 ALGORITHM  
AND NAÏVE BAYES METHOD BASED ON PARTICLE SWARM  
OPTIMIZATION TO PREDICT HEART DISEASE***

**ABSTRACT**

*Heart disease is one of the deadly non-contagious diseases. This study aims to predict heart disease by looking at what indications can cause a person to develop heart disease using the decision tree method with the C4.5 Algorithm and Naïve Bayes based on Particle Swarm Optimization. The research was conducted using the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology. The data used is from interviews with CDC (Centers of Disease Control and Prevention) institutions. At the initial stage, the selection of influential attributes will be carried out using Particle Swarm Optimization (PSO). Furthermore, the selected attributes will be modeled with the decision tree method with C4.5 algorithm and Naïve Bayes method. Furthermore, the validation and evaluation process is carried out to measure the performance of the model that has been built. The results show that both models have the same accuracy value. However, the recall and F1-Score values of the decision tree model with the C4.5 algorithm are greater than the Naïve Bayes model. In this study, the decision tree model is better than Naïve Bayes model in predicting heart disease.*

***Keywords:*** *Hearth Disease, Decision Tree, C4.5 Algorithm, Naïve Bayes, Particle Swarm Optimization.*

## DAFTAR ISI

<b>LEMBAR PENGESAHAN .....</b>	<b>i</b>
<b>LEMBAR HAK CIPTA.....</b>	<b>ii</b>
<b>LEMBAR PERNYATAAN .....</b>	<b>iii</b>
<b>KATA PENGANTAR.....</b>	<b>iv</b>
<b>UCAPAN TERIMA KASIH .....</b>	<b>v</b>
<b>ABSTRAK .....</b>	<b>vii</b>
<b>ABSTRACT .....</b>	<b>viii</b>
<b>DAFTAR ISI .....</b>	<b>ix</b>
<b>DAFTAR GAMBAR .....</b>	<b>xi</b>
<b>DAFTAR TABEL .....</b>	<b>xii</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xiii</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	4
<b>BAB II KAJIAN PUSTAKA.....</b>	<b>5</b>
2.1 <i>Data Mining</i> .....	5
2.2 <i>Pohon Keputusan</i> .....	6
2.3 <i>Algoritma C4.5</i> .....	6
2.4 <i>Naïve Bayes</i> .....	8
2.5 <i>Penelitian Sebelumnya</i> .....	11
2.6 <i>Particle Swarm Optimization (PSO)</i> .....	12
2.7 <i>Confusion Matrix</i> .....	14
2.8 <i>Cross Validation</i> .....	16
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>18</b>
3.1 <i>Deskripsi Masalah</i> .....	18
3.2 <i>Tahapan Penelitian</i> .....	19
3.3 <i>Data Penelitian</i> .....	20
3.4 <i>Pengolahan Data</i> .....	22
3.4 <i>Algoritma PSO untuk Pembobotan Atribut</i> .....	24



3.5	Metode Pohon Keputusan dengan Algoritma C4.5 .....	30
3.6	Algoritma <i>Naïve Bayes</i> .....	32
3.7	<i>Evaluation</i> .....	33
3.8	<i>Perform Comparison</i> .....	34
3.9	Contoh Kasus.....	34
<b>BAB IV PEMBAHASAN .....</b>		<b>45</b>
4.1	Analisis Data Penelitian .....	45
4.2	<i>Preprocessing Data</i> .....	48
4.3	Pembobotan Atribut dengan PSO.....	51
4.4	Pemodelan Pohon Keputusan dengan Algoritma C4.5.....	52
4.5	Pemodelan dengan <i>Naïve Bayes</i> .....	55
4.6	Validasi dan Evaluasi .....	57
4.7	Perbandingan Performa Model Pohon Keputusan dan <i>Naïve Bayes</i> .....	61
<b>BAB V KESIMPULAN .....</b>		<b>63</b>
5.1	Kesimpulan.....	63
5.2	Saran .....	64
<b>DAFTAR PUSTAKA .....</b>		<b>65</b>
<b>LAMPIRAN .....</b>		<b>68</b>

## DAFTAR GAMBAR

Gambar 2. 1 Tabel <i>Confusion Matrix</i> . .....	15
Gambar 2. 2 Ilustrasi Proses <i>Cross Validation</i> . .....	16
Gambar 3. 1 Proses Klasifikasi. ....	19
Gambar 3. 2 Pohon Keputusan. ....	42
Gambar 4. 1 Distribusi <i>Hearth Disease or Attack</i> . .....	48
Gambar 4. 2 Pemeriksaan <i>Missing Value</i> . .....	48
Gambar 4. 3 Proses Transformasi Tipe Data. ....	49
Gambar 4. 4 Informasi Tipe Data Setiap Kolom. ....	49
Gambar 4. 5 Distribusi “ <i>Hearth Disease or Attack</i> ” Setelah Penyeimbangan. ....	50
Gambar 4. 6 Analisis Nilai Korelasi. ....	51
Gambar 4. 7 Menentukan <i>Personal Best</i> Awal. ....	51
Gambar 4. 8 Perhitungan <i>Velocity</i> . ....	51
Gambar 4. 9 Pencarian Nilai <i>Globalbest</i> Terbaik .....	52
Gambar 4. 10 Hasil PSO Atribut Terpilih. ....	52
Gambar 4. 11 Model Pohon Keputusan .....	53
Gambar 4. 12 Salah Satu Akar yang Ada pada Pohon Keputusan. ....	54
Gambar 4. 13 Algoritma <i>Naïve Bayes</i> .....	56
Gambar 4. 14 Evaluasi Model Pohon Keputusan. ....	57
Gambar 4. 15 Hasil <i>Confusion Matrix</i> Pohon Keputusan. ....	58
Gambar 4. 16 Nilai Performa Pohon Keputusan. ....	59
Gambar 4. 17 Evaluasi Model <i>Naïve Bayes</i> . ....	59
Gambar 4. 18 Hasil <i>Confusion Matrix Naïve Bayes</i> .....	60
Gambar 4. 19 Nilai Performa <i>Naïve Bayes</i> . ....	61

## DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu.....	11
Tabel 3. 1 Atribut-Atribut <i>Dataset</i> .....	21
Tabel 3. 2 Pengelompokan Data Berdasarkan PSO.....	25
Tabel 3. 3 Contoh Kasus PSO.....	34
Tabel 3. 4 Sampel Data.....	35
Tabel 3. 5 Nilai <i>Cost</i> .....	35
Tabel 3. 6 <i>Update</i> Kecepatan Iterasi 1.....	36
Tabel 3. 7 <i>Update</i> Posisi Iterasi 1.....	37
Tabel 3. 8 <i>Update</i> Nilai <i>Fitness</i> .....	37
Tabel 3. 9 <i>Update</i> Kecepatan Iterasi 2.....	37
Tabel 3. 10 <i>Update</i> Posisi Iterasi 2.....	38
Tabel 3. 11 <i>Update</i> Nilai <i>Fitness</i> Iterasi 2.....	38
Tabel 3. 12 Perbandingan Nilai <i>Fitness</i> Lama dan Baru.....	38
Tabel 3. 13 Bobot Akhir Tiap Atribut.....	39
Tabel 3. 14 Sampel Data dengan Atribut Terpilih.....	39
Tabel 3. 15 Perhitungan Nilai <i>Gain</i> .....	41
Tabel 3. 16 Contoh Data <i>Testing</i> .....	42
Tabel 4. 1 <i>Dataset</i> Penelitian.....	46
Tabel 4. 2 Tipe Data Setiap Variabel Penelitian.....	47
Tabel 4. 3 Hasil Prediksi Model Pohon Keputusan.....	53
Tabel 4. 4 Hasil Model <i>Naïve Bayes</i> .....	56
Tabel 4. 5 Perbandingan Performa.....	62

## DAFTAR LAMPIRAN

Lampiran 1 : Riwayat Hidup Penulis.....	68
Lampiran 2: <i>Coding Python</i> dalam Proses PSO. ....	69
Lampiran 3: <i>Coding Python</i> dalam Proses Penyeimbangan Data. ....	71
Lampiran 4: <i>Coding Python</i> dalam Proses Pemodelan Pohon Keputusan.....	72
Lampiran 5: Pohon Keputusan Lengkap.....	74
Lampiran 6: <i>Coding Python</i> dalam Proses Pemodelan <i>Naïve Bayes</i> .....	82
Lampiran 7: <i>Coding Python</i> dalam Proses Validasi dan Evaluasi Model Pohon Keputusan.....	84
Lampiran 8: <i>Coding Python</i> dalam Proses Validasi an Evaluasi Model <i>Naïve Bayes</i> ....	85

## DAFTAR PUSTAKA

- Achmad, B. D. M., & Slamet, F. (2012). Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree. *Jurnal IPTEK*, 16(1), 17–23.
- Berry, M. J., & Linoff, G. S. (2004). *Data Mining Techniques : For Marketing, Sales, and Customer Relationship Management Second Edition* (M. Bednarek, R. M. Elliott, & K. A. Malm, Ed.). Wiley Publishing Inc.
- Depari, D. H., Widiastiwi, Y., & Santoni, M. M. (2022). Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung. *Jurnal Informatik*, 18(3), 239–248.
- Dwiasnati, S., & Devianto, Y. (2019). Optimasi Prediksi Keputusan Calon Nasabah Potensial Berbasis Particle Swarm Optimization. *Jurnal Informatika*, 6(2), 286–292. <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- Handoko, M. R., & Neneng. (2021). Sistem Pakar Diagnosa Penyakit Selama Kehamilan Menggunakan Metode Naïve Bayes Berbasis Web. *Jurnal Teknologi dan Sistem Informasi (JTISI)*, 2(1), 50–58. <http://jim.teknokrat.ac.id/index.php/JTISI>
- Hanifah, W., Oktavia, W. S., & Nisa, H. (2021). Faktor Gaya Hidup dan Penyakit Jantung Koroner: Review Sistematis pada Orang Dewasa di Indonesia (Lifestyle Factors and Coronary Heart Disease: A Systematic Review Among Indonesian Adults). *The Journal of Nutrition and Food Research*, 44(1), 45–58.
- Imandasari, T., Irawan, E., Windarto, A. P., & Wanto, A. (2019). Algoritma Naïve Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 750–761.
- Jannah, S. S. N., Riska, N. F., Alamsyah, M., & Indrayatna, F. (2023). Pengklasifikasian Penyakit Jantung dengan Metode Decision Tree. *Seminar Nasional Statistika Aktuaria II*. <http://prosidingnsa.statistics.unpad.ac.id>
- Kusrini, & Emha, T. L. (2009). *Algoritma Data Mining*. Andi.
- Larassati, D., Zaidiah, A., & Afrizal, S. (2022). Sistem Prediksi Penyakit Jantung Koroner Menggunakan Metode Naïve Bayes. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 7(2), 533–546. [www.kaggle.com](http://www.kaggle.com)

- Maisyaroh, S. (2016). Optimalisasi Penentuan Lurah Berprestasi Se-Kecamatan Bogor Timur Kota Bogor dengan Menggunakan Metode Decision Tree. *Jurnal Ilmiah Teknologi dan Informasi*, 6(1), 51–61.
- Muflikhah, L., Ratnawati, D. E., & Putri, R. R. M. (2018). *Data Mining* (Cetakan Pertama). UB Press.
- Muhamad, H., Prasajo, C. A., Sugianto, N. A., Surtiningsih, L., & Cholissodin, I. (2017). Optimasi Naïve Bayes Classifier dengan Menggunakan Particle Swarm Optimization Pada Data Iris. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 4(3), 180–184.
- Mutiara, E. (2020). Algoritma Klasifikasi Naïve Bayes Berbasis Particle Swarm Optimization untuk Prediksi Penyakit Tuberculosis (Tb). *Jurnal SWABUMI*, 8(1), 46–58.
- Muzakir, A., & Wulandari, R. A. (2016). Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree. *Scientific Journal of Informatics*, 3(1), 19–26. <http://journal.unnes.ac.id/nju/index.php/sji>
- Nasrullah, A. H. (2021). Implementasi Algoritma Decision Tree untuk Klasifikasi Produk Laris. *Jurnal Ilmiah Ilmu Komputer*, 7(2), 45–81. <http://ejournal.fikom-unasman.ac.id>
- Nofriansyah, D. (2014). *Konsep Data Mining Sistem Pendukung Keputusan* (Pertama). Deepublish.
- Pambudi, A., Abidin, Z., & Permata. (2023). Penerapan CRISP-DM Menggunakan MLR K-Fold pada Data Saham PT. Telkom Indonesia (Persero) Tbk (Tlkm) (Studi Kasus: Bursa Efek Indonesia Tahun 2015-2022). *JDMSI*, 4(1), 1–14.
- Prabowo, A. D. R., & Mujono. (2018). Prediksi Nasabah yang Berpotensi Membuka Simpanan Deposito Menggunakan Naive Bayes Berbasis Particle Swarm Optimization. *Jurnal Teknologi Informasi*, 17(2), 208–219. <https://doi.org/https://doi.org/10.33633/tc.v17i2.1648>
- Purnamasari, E., Rini, D. P., & Sukemi. (2020). Seleksi Fitur menggunakan Algoritma Particle Swarm Optimization pada Klasifikasi Kelulusan Mahasiswa dengan Metode Naive Bayes. *Jurnal Resti ( Rekayasa Sistem dan Teknologi Informasi)*, 4(3), 469–475.
- Putra, D., & Wibowo, A. (2020). Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes. *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, 2, 84–92.
- Ramdhani, L. S. (2016). Penerapan Particle Swarm Optimization (PSO) untuk Seleksi Atribut dalam Meningkatkan Akurasi Prediksi Diagnosis

- Penyakit Hepatitis dengan Metode Algoritma C4.5. *SWABUMI*, 4(1), 1–15.
- Sambani, E. B., & Nuraeni, F. (2017). Penerapan Algoritma C4.5 Untuk Klasifikasi Pola Penjurusan di Sekolah Menengah Kejuruan (SMK) Kota Tasikmalaya. *CSRID Journal*, 9(3), 149–157. <https://doi.org/10.22303/csrid.9.3.2017.149-157>
- Saputra, R. A., Agustina, C., Puspitasari, D., Ramanda, R., Warjiyono, Pribadi, D., Lisnawanty, & Indriani, K. (2020). Detecting Alzheimer's Disease by the Decision Tree Methods Based on Particle Swarm Optimization. *Journal of Physics: Conference Series*, 1641(1). <https://doi.org/10.1088/1742-6596/1641/1/012025>
- Sukron, M., Supriadi, A., Sulton, R., Nurul Jadid, U., & Artikel, R. (2021). Optimasi Metode Naïve Bayes Menggunakan Algoritma Particle Swarm Optimization (PSO) Untuk Prediksi Penyakit Diabetes Mellitus. *Jurnal Kecerdasan Buatan, Komputasi dan Teknologi Informasi*, 2(2), 18–24. <https://ejournal.unuja.ac.id/index.php/core>
- Suyanto. (2014). *Artificial Intelligence: Searching, Reasoning, Planning, Learning* (1 ed.).
- Tarigan, D. M. Br., Rini, D. P., & Sukemi. (2019). Particle Swarm Optimization-Based on Decision tree of C4.5 Algorithm for Upper Respiratory Tract Infections (URTI) Prediction. *Journal of Physics: Conference Series*, 1196(1). <https://doi.org/10.1088/1742-6596/1196/1/012077>
- Vanegas, C. E. D., Mejía, J. C. G., Agudelo, F. A. V., & Duran, D. E. S. (2023). A Representation Based on Essence for the CRISP-DM Methodology. *Computacion y Sistemas*, 27(3), 675–689. <https://doi.org/10.13053/CyS-27-3-3446>
- Weng, C. H., Huang, T. C. K., & Han, R. P. (2016). Disease Prediction with Different Types of Neural Network Classifiers. *Telematics and Informatics*, 33(2), 277–292. <https://doi.org/10.1016/j.tele.2015.08.006>
- Yates, L. A., Aandahl, Z., Richards, S. A., & Brook, B. W. (2022). Cross validation for model selection: A review with examples from ecology. *Ecological Monographs*, 93(1). <https://doi.org/10.1002/ecm.1557>