

OPTIMASI ALGORITMA KLASIFIKASI C4.5 BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK PREDIKSI PENYAKIT JANTUNG

Rusda Wajhillah

Program Studi Manajemen Informatika

AMIK BSI Sukabumi

Jl. Cemerlang No. 8 Sukakarya, Sukabumi

<http://www.rusdaw.jimdo.com>

E-mail: rusda.rwh@bsi.ac.id

ABSTRACT

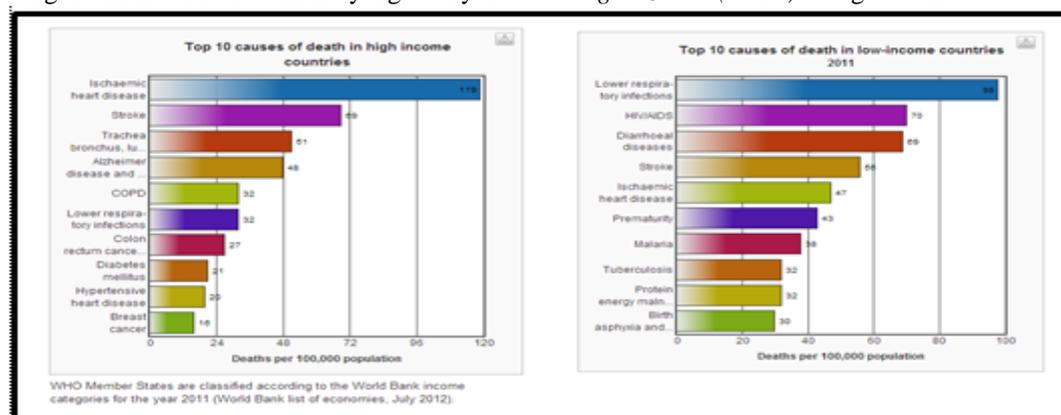
According to WHO, 17 million people die every year due to diseases of the heart and blood vessels around the world. Coronary Heart Disease (CHD) alone holds the title as the cause of death and disability in the world number one. It will be applied algorithm C4.5 decision tree feature selection based Particle Swarm Optimization (PSO) to find out how accurate the PSO feature selection to improve the accuracy of the C4.5 decision tree in predicting heart disease. The results for the accuracy of classification algorithm C4.5 worth 81,25%, whereas the accuracy for C4.5 classification algorithm based on PSO is worth 93,75% the value that is equal to 12,5% accuracy. While evaluation using ROC curve for both, the value of AUC by ROC curve for C4.5 classification algorithm is worth 0,718 with Fair diagnosis classification level, whereas for C4.5 classification algorithm based on PSO is worth 0,855 with Good Classification diagnosis rate, the difference in AUC values is 0,137. It can be concluded that the application of particle swarm optimization techniques can improve the accuracy of the algorithm C4.5.

Keywords: Heart Diseases, Algorithm C4.5, Particle Swarm Optimization

I. PENDAHULUAN

Jumlah orang yang terkena penyakit jantung sangat bervariasi di berbagai negara (Davidson, 2003). Penyakit ini berkaitan erat dengan gaya hidup, fakta menunjukkan di Inggris atau Skandinavia yang dikenal dengan menu Mediteranian yang banyak

mengandung sayuran segar, salad, buah, ikan dan sedikit daging atau hasil susu membantu mencegah timbulnya penyakit jantung (Davidson, 2003). Namun demikian beberapa pakar menjelaskan kemungkinan meningkatnya jumlah penderita jantung di Indonesia, berdasarkan data *World Health Organization (WHO)* sebagai berikut:



Gambar 1 Grafik 10 Penyakit Penyebab kematian pada Negara Maju dengan Pendapatan Tinggi dan Rendah

Sumber: WHO (Update Juli 2013)

Penelitian yang telah dilakukan untuk mendiagnosa pasien diantaranya dilakukan oleh Rajkumar & Reena, September (2010) menggunakan *tanagra tool* untuk membandingkan akurasi kinerja data mining untuk mendiagnosa penyakit jantung. Chen et al. (2011) yang mengembangkan sistem prediksi penyakit jantung menggunakan *artificial neural network* yang dapat membantu profesional medis dalam memprediksi status penyakit jantung berdasarkan data klinis pasien. Kumar, Sathyadevi, & Sivanesh (2011) menggunakan beberapa algoritma pohon keputusan untuk mengklasifikasi jenis penyakit dan membandingkan efektifitas diantara beberapa pohon keputusan. Alfiahrin (2012) membandingkan tiga metode klasifikasi untuk prediksi penyakit jantung, diperoleh pohon keputusan C4.5 yang paling akurat diantara Naive Bayes dan Neural Network.

Berdasarkan beberapa penelitian tersebut di atas, untuk menangani kelemahan-kelemahan yang masih ada maka akan diterapkan algoritma pohon keputusan C4.5 berbasis *feature selection Particle Swarm Optimization (PSO)* untuk mengetahui seberapa akurat *feature selection PSO* dalam meningkatkan akurasi pohon keputusan C4.5 dalam memprediksi penyakit jantung.

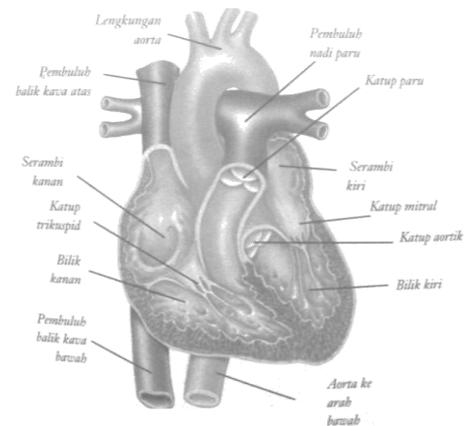
II. LANDASAN PEMIKIRAN

2.1 Tinjauan Pustaka

2.1.1 Penyakit Jantung

Jantung adalah organ berongga dan memiliki empat ruang yang terletak antara kedua paru-paru di bagian tengah rongga toraks (Sloane, 2004). Jantung merupakan salah satu organ vital pada tubuh yang berfungsi untuk memompa darah keseluruh tubuh melalui pembuluh darah. Jantung terdiri dari empat ruang, yang terdiri dari ruang atas dinamakan atrium kanan dan kiri sedangkan ruang bawah dinamakan ventrikel kanan dan kiri. Darah dari seluruh tubuh mengalir ke vena yang bermuara di atrium kanan, darah dari seluruh tubuh melewati katup jantung dan mencapai ventrikel kanan. Selama kontraksi dari ventrikel kanan darah didorong ke paru-paru untuk mengeluarkan karbondioksida (CO_2), mengambil oksigen dan kembali melalui vena pulmonari ke atrium kiri. Jika jantung tidak dapat bekerja dengan baik maka dapat menyebabkan penyakit jantung.

Struktur jantung bagian dalam dapat dilihat pada gambar seperti di bawah ini:



Gambar 2 Struktur Jantung Bagian Dalam

Sumber: Davidson, 2003

Penyakit Jantung Koroner (PJK) merupakan suatu kondisi jantung yang tidak dapat bekerja sebagaimana mestinya, karena otot jantung mengalami kerusakan akibat kekurangan oksigen (Prodia, 2013).

2.1.2 Data Mining

Data mining adalah proses menemukan korelasi baru yang bermakna, pola dan tren dengan memilah-milah sejumlah besar data yang tersimpan dalam repositori, menggunakan teknologi penalaran pola serta teknik-teknik statistik dan matematika (Larose, 2005). Data mining merupakan bagian dari *Knowledge Discovery Data (KDD)* yang merupakan proses ekstraksi informasi yang berguna, tidak diketahui sebelumnya dan tersembunyi dari data (Bramer, 2007).

Secara garis besar proses KDD tersebut dapat dijelaskan sebagai berikut (Kusrini & Luthfi, 2009):

1. *Data Selection*
2. *Pre-processing/Cleaning*
3. *Transformation*
4. *Data Mining*
5. *Interpretation/Evaluation*

2.1.3 Klasifikasi

Klasifikasi merupakan proses pembelajaran suatu fungsi tujuan (target) f yang memetakan

tiap himpunan atribut x ke satu dari label kelas y yang didefinisikan sebelumnya. Fungsi target disebut juga model klasifikasi (Hermawati, 2009). Hal tersebut dapat digambarkan pada blok diagram sebagai berikut:



Gambar 3 Blok Diagram Model Klasifikasi

Sumber: Hermawati, 2009

Berdasarkan gambar di atas, dapat juga dikatakan bahwa model dalam klasifikasi mempunyai arti yang sama dengan kotak hitam, dimana ada suatu model yang menerima masukan, kemudian mampu melakukan pemikiran terhadap masukan tersebut dan memberikan jawaban sebagai keluaran dari hasil pemikirannya (Prasetyo, 2012).

2.1.4 Algoritma C4.5

Algoritma C4.5 merupakan struktur pohon di mana terdapat simpul yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji (Widodo, 2013). Algoritma ini secara *rekursif* mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak bisa dibagi lagi. Konsep yang digunakan untuk memilih entropi yang optimal adalah dengan *information gain* atau *entropy reduction*.

Terdapat beberapa tahap dalam membuat pohon keputusan dengan Algoritma C4.5 (Larose, 2005), yaitu:

1. Mempersiapkan *datatraining*, *data training* biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menghitung *Total Entropy* sebelum dicari masing-masing *Entropy class*

$$H(T) = - \sum_j P_j \log_2 (P_j) \quad (2.1)$$

Keterangan:

H = Himpunan Kasus

T = Atribut

P_j = Proporsi dari H_j terhadap H

3. Hitung nilai *gain* dan *information gain* dengan rata-rata:

$$\text{Gain Average} = H(T) - H_{\text{aving}}(T) \quad (2.2)$$

Keterangan:

$H(T)$ = Total Entropy

4. Ulangi langkah ke-2 dan ke-3 hingga semua tupel terpartisi.
5. Proses partisi pohon keputusan akan berhenti pada saat:
 - a. Semua tupel dalam node N mendapat kelas yang sama
 - b. Tidak ada atribut di dalam tupel yang dipartisi lagi
 - c. Tidak ada tupel di dalam cabang yang kosong

2.1.5 Particle Swarm Optimization

Particle Swarm Optimization (PSO) adalah teknik optimasi stokastik berbasis populasi yang dikembangkan oleh Eberhart dan Kennedy pada tahun 1995, yang terinspirasi oleh perilaku sosial kawanan burung atau ikan (Hu, Shi, & Eberhart, 2004).

PSO menggunakan teknik perhitungan evolusioner (Hu et al, 2003):

1. PSO diinisialisasi dengan sekumpulan solusi acak.
2. PSO mencari solusi yang optimum dengan memperbaharui generasi
3. Perkembangan populasi berdasarkan pada generasi sebelumnya.

Dalam prosesnya, metode *PSO* dipengaruhi oleh sifat individu dan kelompok dalam solusi optimal. Sebagai pengganti operator evolusi, misalnya untuk persoalan optimasi d-variabel, akan disebar kawanan partikel (misalnya sebanyak n-partikel) dalam ruang d-dimensi secara acak. Masing-masing partikel merupakan kandidat solusi dan mempunyai nilai *fitness* tertentu. Kemudian tiap partikel akan bergerak dengan kecepatan tertentu yang dipengaruhi oleh pengalaman terbang sendiri ataupun pengalaman partikel lain.

Sebagai contoh, partikel ke-i dinyatakan sebagai: $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,d})$ dalam ruang d-dimensi. Posisi terbaik sebelumnya dari partikel ke-i disimpan dan dinyatakan sebagai $pbest_i = (pbest_{i,1}, pbest_{i,2}, \dots, pbest_{i,d})$. Indeks partikel terbaik diantara semua partikel dalam

kawanan *group* dinyatakan sebagai $gbest_d$. Kecepatan partikel ke- i dinyatakan sebagai: $v_i = (v_{i,1}, v_{i,2}, \dots, v_{i,d})$. Modifikasi kecepatan dan posisi tiap partikel dapat dihitung menggunakan kecepatan saat ini dan jarak $pbest_{i,d}$ ke $gbest_d$ seperti ditunjukkan pada persamaan berikut:

$$v_{i,m} = w \cdot v_{i,m} + c_1 * R * (pbest_{i,m} - x_{i,m}) + c_2 * R * (gbest_m - x_{i,m}) \quad (2.3)$$

$$x_{i,d} = x_{i,m} + v_{i,m} \quad (2.4)$$

Dimana:

- n : jumlah partikel dalam kelompok
- d : dimensi
- $v_{i,m}$: kecepatan partikel ke- i pada iterasi ke- i
- w : faktor bobot *inersia*
- c_1, c_2 : konstanta akselerasi (*learning rate*)
- R : bilangan *random* (0-1)
- $x_{i,d}$: posisi saat ini dari partikel ke- i pada iterasi ke- i
- $pbest_i$: posisi terbaik sebelumnya dari partikel ke- i
- $gbest$: partikel terbaik diantara semua partikel dalam satu kelompok atau populasi

Persamaan (2.3) menghitung kecepatan baru untuk tiap partikel (solusi potensial) berdasarkan pada kecepatan sebelumnya ($V_{i,m}$), lokasi partikel dimana nilai *fitness* terbaik telah dicapai ($pbest$) dan lokasi populasi global ($gbest$ untuk versi global, $lbest$ untuk versi *local*) atau *local neighborhood* pada algoritma versi *local* dimana nilai *fitness* terbaik telah dicapai.

Persamaan (2.4) memperbaharui posisi tiap partikel pada ruang solusi. Dua bilangan acak c_1 dan c_2 dibangkitkan sendiri. Penggunaan berat *inersia* telah memberikan performa yang meningkat pada sejumlah aplikasi (Hu et al, 2004). Hasil dari perhitungan partikel yaitu kecepatan partikel diantara interval [0,1] (Hu et al, 2004).

2.1.6. Evaluasi Confusion Matrix dan Receiver Operating Characteristic (ROC)

1. Evaluasi Confusion Matrix
Evaluasi kinerja model klasifikasi didasarkan pada pengujian objek yang diprediksi dengan benar dan salah, hitungan ini ditabulasikan confusion matrix (Gorunescu, 2011). *Confusion Matrix* adalah perangkat yang berguna untuk menganalisis seberapa baik *classifier* dapat mengenali tupel dari

kelas yang berbeda (Han & Kamber, 2006).

2. Evaluasi ROC Curve
Kurva *ROC* (*Receiver Operating Characteristic*) banyak digunakan untuk menilai hasil prediksi, kurva *ROC* adalah teknik untuk memvisualisasikan, mengatur, dan memilih pengklasifikasian berdasarkan kinerja mereka (Gorunescu, 2011). Kurva *ROC* adalah perangkat dua dimensi yang digunakan untuk menilai kinerja klasifikasi yang menggunakan dua *class* keputusan, masing-masing objek dipetakan ke salah satu elemen dari himpunan pasangan, positif atau negatif.

2.2 Tinjauan Studi Terdahulu yang Relevan

D. Senthil Kumar, G. Sathyadevi dan S. Sivanesh dalam penelitiannya menyatakan bahwa algoritma pohon keputusan adalah salah satu metode klasifikasi yang paling efektif. Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi sistem pendukung keputusan bagi dunia medis untuk perawatan pasien yang menderita penyakit jantung, diabetes dan hepatitis. Penelitian ini menggunakan beberapa algoritma pohon keputusan yaitu C4.5, ID3 dan CART untuk mengklasifikasi penyakit diabetes, hepatitis dan jantung. Dari perbandingan algoritma pohon keputusan didapat bahwa CART menunjukkan kinerja yang lebih baik dari pada algoritma yang lainnya dengan akurasi sebesar 83,2%, sedangkan algoritma C4.5 dan ID3 masing-masing menghasilkan akurasi sebesar 71,4% dan 64,8% (2011).

AH Chen, SY Huang, PS Hong, CH Cheng dan EJ Lin mengembangkan sistem prediksi penyakit jantung dengan menggunakan algoritma *neural network* yang dapat membantu profesional medis dalam memprediksi status penyakit jantung berdasarkan data klinis pasien. Sistem yang dibangun terdiri dari data klinis, tampilan kurva ROC dan bagian prediksi (eksekusi waktu, akurasi, sensitivitas, spesifisitas dan hasil prediksi). Keakuratan dari prediksi yang digunakan oleh Chen et al mencapai hasil mendekati 80% (Chen et al, 2011).

Sa'diyah Noor Novita Alfiahrin pada tesisnya melakukan komparasi algoritma pohon keputusan C4.5, Naive Bayes dan Neural Network untuk memprediksi penyakit jantung, dengan menggunakan 13 atribut yang

terdapat pada gejala penyakit jantung diperoleh dari *UCI Repository* diperoleh akurasi yang paling tinggi dalam memprediksi penyakit jantung adalah menggunakan metode C4.5 yaitu sebesar 96.15% (Alfisahrin, 2012).

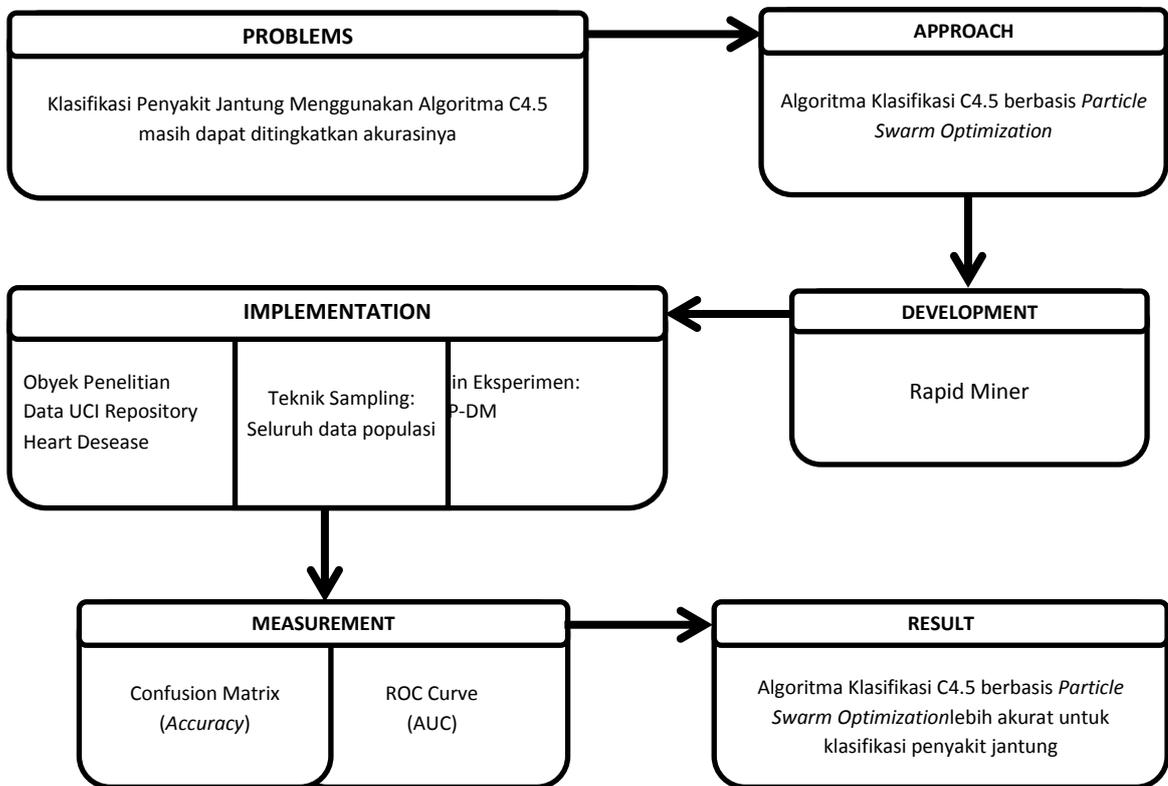
<http://archive.ics.uci.edu/ml/datasets/Statlog+project> sebagai subset dari dataset publik yang digunakan dalam proyek statlog eropa. Proyek tersebut melakukan perbandingan kinerja mesin pembelajaran, statistik dan algoritma jaringan syaraf tiruan pada dataset dari dunia nyata pada daerah industri termasuk dalam bidang kedokteran. Dataset yang digunakan terdiri dari 75 atribut yang kemudian direduksi menjadi 13 atribut.

2.3 Tinjauan Objek Studi

Pada penelitian ini objek yang dijadikan bahan penelitian adalah *statlog database* yang berasal dari

2.4 Kerangka Konsep Penelitian

Model kerangka yang digunakan pada penelitian ini dapat digambarkan sebagai berikut:



Gambar 4 Kerangka Konsep Penelitian

2.5 Hipotesis

Hipotesis merupakan dugaan awal terhadap sebuah penelitian yang dilakukan. Terdapat dua jenis hipotesis yang ditentukan dalam penelitian ini, yaitu hipotesis umum dan hipotesis khusus.

1. Hipotesis umum yang dijadikan aspek penelitian adalah:
Diduga algoritma klasifikasi C4.5 berbasis *Particle Swarm Optimization (PSO)* memiliki akurasi lebih tinggi dibandingkan algoritma C4.5 karena dapat mengoptimalkan atribut-atribut yang ada pada prediksi penyakit jantung.
2. Hipotesis khusus yang dijadikan aspek penelitian adalah:
Diduga *Graphical User Interface (GUI)* yang dihasilkan dapat diterima bagi pengguna.

III. METODE PENELITIAN

3.1 Jenis Penelitian

Jenis penelitian yang dilakukan pada penelitian ini adalah jenis penelitian eksperimen. Metode ini menguji kebenaran sebuah hipotesis dengan statistik dan menghubungkannya dengan masalah penelitian (Kothari, 2004). Tujuan dari penelitian ini adalah melakukan komparasi dan evaluasi model pohon keputusan C4.5 sebagai algoritma terpilih dan pohon keputusan C4.5 berbasis *Particle Swarm Optimization (PSO)* untuk mengetahui algoritma yang memiliki keakuratan lebih tinggi dalam memprediksi penyakit jantung. Penelitian ini menggunakan data sekunder yang diperoleh dari *University of California Irvine (UCI) Machine Learning Data Repository*.

3.2 Metode Pemilihan Populasi dan Sampel

3.2.1 Populasi

Pada penelitian ini populasi yang digunakan adalah 270 *record* dari dataset penyakit jantung yang diperoleh dari *University of California Irvine (UCI) Machine Learning Repository* yang bersifat publik yang akan dibagi menjadi data *training* dan *testing*.

3.2.2 Sampel

Penarikan contoh (*sampling*) merupakan teknik utama yang digunakan untuk seleksi data. Proses ini sering digunakan untuk persiapan penyelidikan dan analisis data akhir.

3.3 Metode Pengumpulan Data

Dalam penelitian ini metode pengumpulan data untuk mendapatkan sumber data yang digunakan adalah metode pengumpulan data sekunder. Data utama diperoleh dari *University of California Irvine machine learning data repository* yang dapat diperoleh melalui alamat web <http://archive.ics.uci.edu/ml/datasets/Statlog+project>, sedangkan data pendukung didapatkan dari buku, jurnal dan publikasi lainnya.

3.4 Instrumen Penelitian

Instrumen pada penelitian ini sebagai berikut:

1. Penelitian ini menggunakan data sekunder berupa data pasien yang menderita penyakit jantung ataupun tidak yang berasal dari dataset publik *UCI Repository*.
2. Data disajikan dalam bentuk tabulasi model dan variabel masing-masing sebanyak 270 pasien terdiri dari pasien positif jantung sebanyak 150 dan pasien tidak positif jantung sebanyak 120 pasien yang diambil pada tahun 2009, data dapat dilihat pada Lampiran.2.
3. Perangkat lunak yang digunakan untuk menganalisis adalah *Rapid Miner* dan *Graphical User Interface (GUI)* untuk menguji *rule* algoritma terpilih adalah *Java Netbeans Versi 7.3.1*

3.5 Metode Analisis dan Pengujian Data

Dalam penelitian ini pengujian data menggunakan metode eksperimen dengan model *Cross Industry Standard Process for Data Mining (CRISP-DM)* yang terdiri dari enam tahap (Larose, 2005), yaitu:

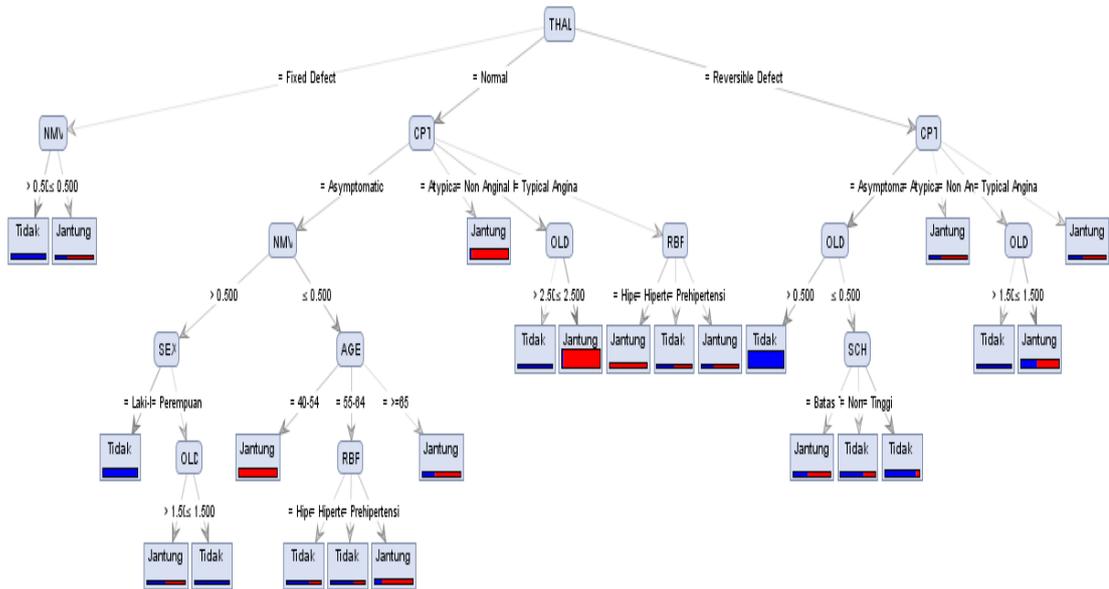
1. Tahap Business Understanding
2. Tahap Data Understanding
3. Tahap Data Preparation
4. Tahap Modelling

3.5.1 Model Algoritma Klasifikasi C4.5

Untuk dapat membuat pohon keputusan, langkah pertama adalah menghitung jumlah

class yang terkena penyakit jantung dan tidak dari masing-masing class berdasarkan atribut yang telah ditentukan dengan menggunakan

data training. Kemudian menghitung Entropy (Total) menggunakan persamaan. Berdasarkan perhitungan diperoleh rule sebagai berikut:

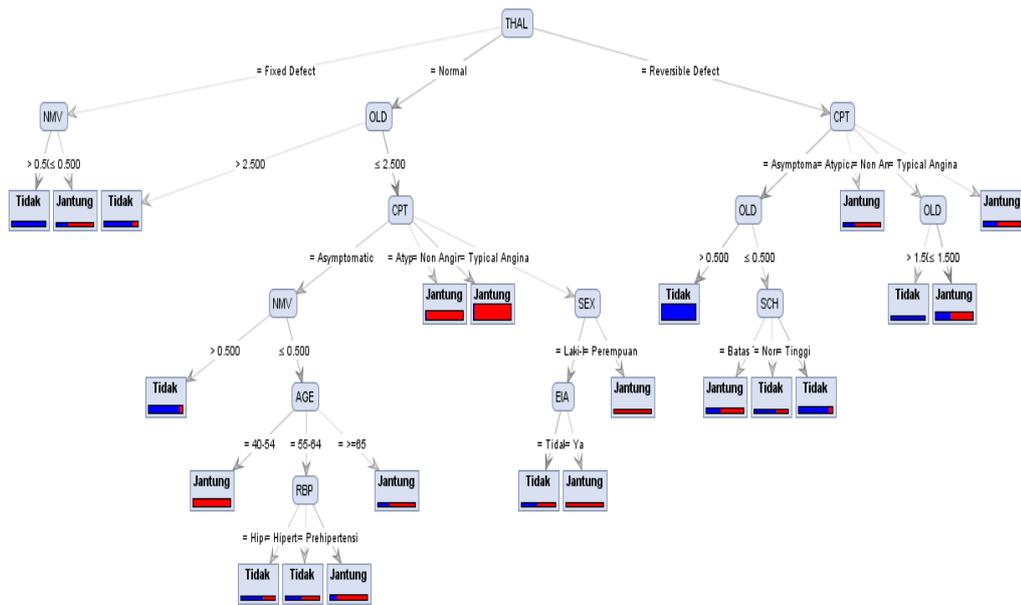


Gambar 5 Pohon Keputusan Klasifikasi Penyakit Jantung menggunakan algoritma C4.5

Terdapat 24 rule yang dihasilkan dari pohon keputusan algoritma klasifikasi C4.5, dengan jumlah class tidak sebanyak 11 rule dan 13 rule untuk class jantung.

3.5.2 Model Algoritma Klasifikasi C4.5 berbasis PSO

Rule untuk algoritma klasifikasi C4.5 berbasis PSO sebagai berikut:



Gambar 6 Pohon Keputusan Klasifikasi Penyakit Jantung menggunakan algoritma C4.5 berbasis PSO

dapat 22 rule yang dihasilkan dari pohon keputusan algoritma klasifikasi C4.5 berbasis PSO, dengan jumlah class tidak sebanyak 10 rule dan 12 rule untuk class jantung.

3. HASIL PENELITIAN DAN PEMBAHASAN

4.1 Pengukuran Penelitian

Pengukuran penelitian yang dilakukan pada penelitian ini menggunakan *tools, rapid miner versi 5.2*

4.1.1 Hasil Penelitian

Tujuan dari penelitian ini adalah mengembangkan model dengan cara mengoptimalkan salah satu algoritma yang sudah terbentuk dari algoritma terpilih yaitu pohon keputusan C4.5. Data dinalisa dengan melakukan dua perbandingan yaitu menggunakan algoritma klasifikasi C4.5 dan algoritma klasifikasi C4.5 berbasis *Particle Swarm Optimization (PSO)*.

4.1.2 Evaluasi dan Validasi Model

1. Hasil Pengujian Algoritma C4.5

Hasil Pengujian ditunjukkan pada tabel *Confusion Matrix* Algoritma Klasifikasi C4.5 pada data *testing* sebagai berikut:

accuracy: 81,25%			
	true Jantung	true Tidak	class precision
pred. Jantung	11	1	91,67%
pred. Tidak	2	2	50,00%
class recall	84,82%	66,67%	

Berdasar tabel diatas dari 53 *tuple* data *testing* terdapat kerangkapan data sehingga hanya terdapat 16 *tuple* dengan rincian jumlah *True Positive (TP)* 11, *False Negative (FN)* 2, *False Positive (FP)* adalah 1 dan *True Negative (TN)* 2.

Sedangkan hasil pengujian terhadap data *testing* untuk algoritma klasifikasi C4.5 terhadap nilai ROC diketahui pada gambar di bawah ini:



Gambar 7 Nilai AUC Algoritma C4.5 dalam Grafik ROC

Nilai AUC sebesar 0,718 untuk evaluasi algoritma C4.5 menunjukkan nilai *Fair Classification*.

2. Hasil Pengujian Algoritma C4.5 berbasis PSO

Hasil Pengujian ditunjukkan pada tabel *Confusion Matrix* Algoritma Klasifikasi C4.5 berbasis PSO pada data *testing* sebagai berikut:

Tabel 1 Tabel Hasil Pengujian Confusion Matrix

accuracy: 93,75%			
	true Jantung	true Tidak	class precision
pred. Jantung	11	1	91,67%
pred. Tidak	0	4	100,00%
class recall	100,00%	80,00%	

Berdasar tabel diatas dari 53 *tuple* data *testing* diperoleh jumlah *True Positive (TP)* 11, *False Negative (FN)* 4, *False Positive (FP)* adalah 1 dan *True Negative (TN)* 0.

Hasil pengujian terhadap data *testing* untuk algoritma klasifikasi C4.5 berbasis PSO terhadap nilai ROC diketahui pada gambar di bawah ini:



Gambar 8 Nilai AUC Algoritma C4.5 berbasis PSO dalam Grafik ROC

Nilai AUC sebesar 0,855 untuk evaluasi algoritma C4.5 berbasis PSO menunjukkan nilai *Good Classification*.

4.2 Analisis Evaluasi dan Validasi Model

Nilai akurasi untuk model algoritma klasifikasi C4.5 sebesar 81,25% dengan nilai AUC 0,718 dan nilai akurasi algoritma klasifikasi C4.5 berbasis PSO (*Particle Swarm Optimization*) sebesar 93,75% dengan AUC 0,855. Berdasarkan nilai tersebut diperoleh selisih akurasi sebesar 12,5% dan selisih AUC sebesar 0,137, dapat dilihat pada Tabel dibawah ini:

Tabel 2 Pengujian Algoritma klasifikasi C4.5 dan C4.5 berbasis PSO

	Accuracy	AUC
C4.5	81,25%	0,718
C4.5 berbasis PSO	93,75%	0,855

4.3 Hasil Pengukuran Uji Kehandalan Graphical User Interface (GUI)

Berdasarkan evaluasi pada optimasi algoritma klasifikasi C4.5 berbasis PSO diketahui bahwa hasilnya lebih tinggi dari pada algoritma klasifikasi C4.5, sehingga rule yang dihasilkan dari algoritma C4.5 berbasis PSO dijadikan sebagai rule untuk pembuatan *Graphical User Interface* guna membantu dokter, tim medis atau masyarakat awam dalam mendiagnosa penyakit jantung. Interface yang digunakan dalam penelitian ini dibuat dengan menggunakan *Java Netbeans Versi 7.3.1* seperti gambar di bawah ini:



Gambar 9 GUI Aplikasi Sistem Prediksi Penyakit Jantung

Untuk mendapatkan hasil yang maksimal dari *GUI* yang dibuat, dibutuhkan Jaminan terhadap kualitas perangkat lunak tersebut. Berikut ini adalah tabel pembobotan nilai *Software Quality Assurance (SQA)*, yang terdiri dari delapan kriteria yang dapat digunakan untuk mengukur kualitas sebuah perangkat lunak secara kuantitatif (Widodo: 2013):

Tabel 3 *Metric of Software Quality Assurance (SQA)*

No	Metrik	Deskripsi	Bobot
1	<i>Auditability</i>	Memenuhi standard atau tidak	0.10
2	<i>Accuracy</i>	Keakuratan komputasi	0.15
3	<i>Completeness</i>	Kelengkapan	0.10
4	<i>Error Tolerance</i>	Toleransi terhadap kesalahan	0.10
5	<i>Execution Efficiency</i>	Kinerja Eksekusi	0.10
6	<i>Operability</i>	Kemudahan untuk dioperasikan	0.15
7	<i>Simplicity</i>	Kemudahan untuk difahami	0.15
8	<i>Training</i>	Kemudahan pembelajaran fasilitas Help	0.15

Berdasarkan *metric* tersebut diimplementasikan pada kuesioner, hasil evaluasi metode *SQA* dapat dilihat pada tabel 4.7 dibawah ini:

Tabel 4 Hasil Evaluasi SQA

Responden	Skor Metrik								Skor
	1	2	3	4	5	6	7	8	
	0.10	0.15	0.10	0.10	0.10	0.15	0.15	0.15	
#1	90	90	80	80	90	95	85	80	86,5
#2	75	70	73	72	75	78	74	70	73,3
#3	75	80	80	70	75	85	85	85	80,25
#4	70	75	85	75	70	85	80	85	78,75
#5	71	72	74	75	72	75	75	72	73,3
#6	80	80	85	85	90	85	88	85	84,7
#7	80	85	85	75	77	85	90	85	83,45
Total *Bobot	54	82,8	56	53	55	88,2	88,6	84,3	112,05
Rata-Rata									80,04

Tabel di atas merupakan hasil angket yang dilakukan pada tujuh orang pengamat yang berperan sebagai pengujian program dan diambil secara acak.

Rumus:

$$\begin{aligned} \text{Skor} &= (541 \times 0,10) + (552 \times 0,15) + (562 \times 0,1) \\ &+ (532 \times 0,1) \\ &+ (549 \times 0,10) + (588 \times 0,15) + (577 \times 0,15) \\ &+ (562 \times 0,15) \\ &= \mathbf{80,04} \end{aligned}$$

Hasil pengujian kualitas tersebut menunjukkan kategori nilai yang baik terhadap perangkat lunak dimaksud.

4. KESIMPULAN

Hasil penelitian untuk nilai akurasi algoritma klasifikasi C4.5 senilai 81,25%, sedangkan untuk nilai akurasi algoritma klasifikasi C4.5 berbasis PSO sebesar 93,75% sehingga tampak selisih nilai akurasi yaitu sebesar

12,5%. Sedangkan evaluasi menggunakan *ROC curve* diperoleh hasil untuk algoritma klasifikasi C4.5 bernilai 0,718 dengan tingkat diagnosa *Fair Classification* dan algoritma klasifikasi C4.5 berbasis PSO bernilai 0,855 dengan tingkat diagnosa yang sama yaitu *Good Classification*, didapatkan selisih nilai AUC sebesar 0,137. Sehingga dapat disimpulkan bahwa penerapan teknik optimasi *particle swarm optimization* dapat meningkatkan nilai akurasi pada algoritma C4.5.

Berdasarkan hasil rekapitulasi kuesioner yang diberikan kepada tujuh responden terhadap *Graphical User Interface (GUI)* yang dibuat menggunakan *Java Netbeans 7.3.1* diperoleh skor 80,04. Artinya secara umum *GUI* yang dihasilkan dapat diterima oleh pengguna dalam membantu prediksi penyakit jantung. Dari 13 atribut yang terdapat pada dataset *UCI Machine Learning Data Repository*, kemudian selanjutnya diseleksi menjadi hanya sembilan atribut yang digunakan dalam menentukan prediksi penyakit jantung, atribut-atribut

tersebut yaitu : THAL, NMV, OLD, CPT, AGE, RBP, SEX, EIA dan SCH.

V. DAFTAR PUSTAKA

- [1] Bhuvaneswari, R., & Kalaiselvi, K. (2012). Naive Bayesian Classification Approach in Healthcare Applications. *International Journal of Computer Science and Telecommunications* Volume 3, Issue 1, January , 106-112.
- [2] Bramer, M. (2007). *Principles of Data Mining*. London: Springer.
- [3] Chen, AH., et al. (2011). HDPS: Heart Disease Prediction System. *Computing in Cardiology*. 557-560
- [4] Cho, YJ., et al.(2011). Optimization of Decision Tree for Classification Using Particle Swarm. *IEMS Vol. 10 No. 4 pp. 272-278*
- [5] Cohen, L., Manion, L., & Morrison, K. (2005). *Research Methods in Education 5th Edition*. London: The Taylor & Francis e-Library.
- [6] Davidson, Christopher. 2003. *Penyakit Jantung Koroner*. Jakarta: Dian Rakyat.
- [7] Eberhart, Russel C., Yuhui Shi. (2001). *Particle Swarm Optimization: Developments, Application and Resources*. 0-7803-6657-3/01. IEEE.
- [8] Gorunescu, Florin. (2011). *Data Mining Concept, Model and Technique*. Verlag Berlin Heidelberg: Springer.
- [9] Hermawati, Fajar Astuti. (2013). *DATA MINING*. Yogyakarta: Andi Offset.
- [10] Khan, M. G. (2005). *Encyclopedia of Heart Diseases*. New York: Academic Press.
- [11] Kothari, C. R. (2004). *Research Methology Methods and Techniques*. India: New Age International Limited.
- [12] Kumar, D. Senthil., Sathyadevi, G., & Sivanesh, S. (2011). Decision Support System for Medical Diagnosis Using Data Mining. *International Journal of Computer Science Issues* Vol.8, Issue 3, No 1, 147-153.
- [13] Kusriani, Emha Taufik Lutfi. (2009). *ALGORITMA DATA MINING*. Yogyakarta: Andi Offset.
- [14] Larose, T. Daniel. (2005). *DISCOVERING KNOWLEDGE IN DATA: An Introduction to DATA MINING*. New Jersey: John Wiley & Sons, Inc.
- [15] Larose, T. Daniel. (2006). *DATA MINING: METHOD AND MODELS*. New Jersey: John Wiley & Sons, Inc.
- [16] Nugroho, W. (2006). *Komunikasi Dalam Keperawatan Gerontik*. Jakarta: EGC.
- [17] Prasetyo, Eko. (2012). *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi Offset.
- [18] Prodia. Dikutip 26 Juni 2013, dari Penyakit Jantung koroner : <http://prodia.co.id/penyakit-dan-diagnosa/penyakit-jantung-koroner>.
- [19] Rajkumar, A., & Reena, G. S. (September 2010). Diagnosis Of Heart Disease Using Data Mining Algorithm. *Global Journal of Computer Science and Technology*, Vol. 10 Issue 10, 38-43.
- [20] Sloane, Ethel. (2004). *ANATOMI DAN FISILOGI: Untuk Pemula*. Jakarta: Penerbit Buku Kedokteran (EGC).
- [21] Sugiyanto. (2008). *Metode Penelitian Kuantitatif Kualitatif dan R&D*. Bandung:Alfabeta.
- [22] Tsai, MC., et al. (2012). An Application of PSO Algorithm and Decision Tree for Medical Problem. 2nd International Conference on Intelligent Computational System (ICS'2012) Oct. 13-14, 2012 Bali-Indonesia.
- [23] University of California Irvine Machine Learning Repository. Dikutip 25 Juni 2013, dari <http://archive.ics.uci.edu/ml/machine-learning-databases/statlog/heart/heart.dat>
- [24] Widodo, Pudjo Prabowo, et al. (2013). *Penerapan Data Mining dengan MATLAB*. Bandung: Rekayasa Sains.
- [25] Wu, X., & Kumar, V. (2009). *The Top Ten Algorithms in Data Mining*. New York: CRC Press.
- [26] World Health Organization. (2013). Dikutip 29 Juni 2013, dari The Top 10Causes of Death: <http://www.who.int/mediacentre/factsheets/fs310/en/index.html>
- [27] Zhan, ZH., et al (2010). Orthogonal Learning Particle Swarm Optimization. *IEEE Transaction On Evolutionary Computation*, 1089-778X