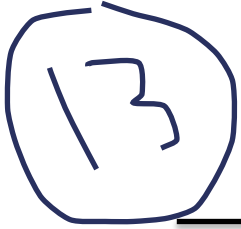




ANALISIS KOMPARASI PEMODELAN ALGORITMA DECISION TREE MENGUNAKAN METODE PARTICLE SWARM OPTIMIZATION DAN METODE ADABOOST UNTUK PREDIKSI AWAL PENYAKIT JANTUNG

Pareza Alam Jusia

Program Studi Teknik Informatika, STIKOM Dinamika Bangsa Jambi
Email: parezaalam@gmail.com



Abstrak

Kolesterol, penumpukan lemak, penggumpalan darah, dan unsur - unsur medis lainnya dapat menyebabkan aliran darah ke jantung menjadi terganggu sehingga kondisi medis ini menjadi darurat sehingga merusak dan menghancurkan otot jantung pada manusia yang mengakibatkan terjadi komplikasi bahkan kematian. Penelitian ini akan melakukan *improve Classification Accuracy / Ensemble Methods Techniques* dengan memodifikasi pemodelan algoritma klasifikasi *Decision Tree* yang ditambahkan dengan metode *Particle Swarm Optimization* dan metode *Adaboost* yang selanjutnya akan dilakukan analisis komparasi terhadap pemodelan tersebut untuk prediksi penyakit jantung. Dataset yang digunakan pada penelitian ini dari ekstraksi *public dataset* yang diambil dari archive *University of California Irvine (UCI)* sebanyak 270 record. Hasil Evaluasi dilakukan dengan tools *Rapid Miner 7* untuk menentukan nilai *confusion matrix* dan *ROC curve*, diketahui bahwa *Decision Tree* mempunyai nilai akurasi 79.26% dan untuk nilai AUC 0.889. Setelah dilakukan modifikasi pemodelan algoritma *decision tree* yang ditambahkan dengan metode *particle swarm optimization* menghasilkan nilai akurasi 82.59% dan untuk nilai AUC 0.916. Pemodelan Algoritma *decision tree* yang ditambahkan dengan metode *adaboost* menghasilkan nilai akurasi 79.26% dan nilai AUC 0.955.

Kata kunci: *Decision Tree*, klasifikasi, *Particle Swarm Optimization*, *adaboost*, jantung

Abstract

Cholesterol, fat accumulation, blood clots, and other medical elements can cause blood flow to the heart to be disrupted so that this medical condition becomes an emergency that destroys and destroys the heart muscle in humans resulting in complications and even death. This research will improve the Classification Accuracy / Ensemble Methods, Techniques by modifying the Decision Tree classification algorithm modeling, which is added by the Particle Swarm Optimization method and Adaboost method which then will be comparative analysis to the modification for heart disease prediction. The dataset used in this study from the extraction of public dataset taken from the archive of the University of California Irvine (UCI) of 270 records. Evaluation result done by Rapid Miner 7 tool to determine the confusion matrix and ROC curve value, it is known that Decision Tree has 79.26% accuracy and for AUC 0.889. After modifying the decision tree modeling algorithm added with the particle swarm optimization method resulted in an accuracy of 82.59% and for AUC 0.916. The decision tree decision algorithm added with the adaboost method yields an accuracy of 79.26% and an AUC value of 0.955.

Keywords: *Decision Tree*, classification, *Particle Swarm Optimization*, *adaboost*, heart

I. PENDAHULUAN

Jantung merupakan organ dalam tubuh yang mempunyai berongga dan terletak didaerah pusat dada. fungsi utama jantung adalah memberikan oksigen dan mengalirkan seluruh tubuh dari gas karbondioksida dengan siklus yang berulang-ulang secara terus menerus. Jika penurunan fungsi jantung, mengakibatkan kondisi aliran darah menuju ke jantung menjadi terhambat maka kinerja jantung yang menjadi abnormal dan menyebabkan penyakit jantung. Kolesterol, penumpukan lemak, penggumpalan darah, dan unsur - unsur medis lainnya dapat menyebabkan aliran darah ke jantung menjadi terganggu sehingga kondisi medis ini menjadi darurat sehingga merusak dan menghancurkan otot jantung pada manusia yang mengakibatkan terjadi komplikasi bahkan kematian. Berdasarkan survey data 10 penyebab kematian tertinggi di negara Indonesia jantung menempati posisi kedua dengan nilai 12,9 % angka kematian untuk laki-laki dan perempuan.



Gambar 1. Penyebab kematian Tertinggi di Indonesia
(sumber : <https://medtech.id/10-penyakit-kematian-tertinggi-di-indonesia>)

Berdasarkan pemaparan data tersebut khususnya untuk penyakit jantung yang menempati urutan kedua penyakit penyebab kematian tertinggi di Indonesia maka perlu melakukan sebuah langkah awal prediksi untuk penanganan pencegahan penyakit jantung. Prediksi ini dilakukan dengan memanfaatkan data-data pasien yang

tersimpan untuk membuat pola dengan komputasi cerdas. Pola-pola tersebut akan meminimalkan ketidaktepan petugas medis dalam melakukan diagnosa awal penyakit Jantung.

Beberapa penelitian tentang komputasi cerdas sudah pernah dilakukan sebelumnya diantaranya adalah. Hasil penelitian dari Sellapan dan Awang membandingkan tiga algoritma untuk prediksi penyakit jantung dengan sample 909 kasus menggunakan 15 atribut menghasilkan nilai akurasi *Artificial Neural Network* 85,53%, *Naive Bayes* 86,53%, dan *Decision Tree* 89% [1]. M.Anbarasi dkk melakukan prediksi penyakit jantung dengan sample 909 kasus menghasilkan nilai akurasi *Clasification Via Clustering* 88,3%, *Naive Bayes* 96,5% dan , *Decision Tree* 99,2% [2]. Prediksi juga dilakukan oleh Srinivas dkk, atribut yang digunakan adalah *age, sex, CA, chest pain type, fasting blood sugar, thalach, exercise, cholestrol, trest blood pressure, slope, restecg, dan old peak*, menghasilkan nilai akurasi untuk algoritma *Decision Tree* 82%, algoritma *Baysian Model* 82%, algoritma *Support Vector Machine* 83,5%, dan algoritma *Multi-Layer Perceptron* 89,2% [3].

Hasil penelitian sebelumnya sudah menunjukkan nilai akurasi yang baik, namun pada penelitian ini akan melakukan *improve Classification Accuracy / Ensemble Methods Techniques* dengan memodifikasi pemodelan algoritma klasifikasi *Decision Tree* yang ditambahkan dengan metode *Particle Swarm Optimization* dan metode *Adaboost* yang selanjutnya akan dilakukan analisis komparasi terhadap pemodelan tersebut untuk prediksi penyakit jantung.

II. KAJIAN LITERATUR

A. Decision Tree

Decision tree adalah *flowchart* seperti pohon yang setiap node menunjukkan suatu test pada suatu atribut, hasil



direpresentasikan pada tiap *branch* dari test tersebut, dan kelas-kelas atau distribusi kelas ditunjukkan oleh *leaf node*. *Decision tree* berguna untuk mengeksplorasi data yang sudah melewati tahap *preprocessing* dan menemukan model yang tersembunyi dari data dengan sebuah target variabel sehingga dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan *record* yang lebih kecil dengan memperhatikan variabel tujuannya, hal ini dinyatakan dalam sebuah referensi terkait [4]. Langkah awal dalam membuat pohon keputusan adalah mempersiapkan data training yang merupakan data yang telah dikelompokkan berdasarkan kategori atau kelas tertentu. Selanjutnya menghitung nilai *entropy* total dan *entropy* setiap class. *Entropy* digunakan untuk mengukur dari teori informasi karakteristik berdasarkan *impurity* dan *homogeneity* dari kumpulan data dengan rumus :

$$Entropy(S) = - \sum -P_j \log_2 P_j \quad (1)$$

Ket :

S = Himpunan Kasus

P_j = Proporsi dari S_j terhadap S

Setelah mendapatkan nilai *entropy* total dan *entropy* dari setiap class, langkah selanjutnya adalah menghitung nilai *information gain*. *Information gain* digunakan untuk mengukur nilai efektivitas atribut tertentu dalam sebuah klasifikasi data. nilai *information gain* dapat dirumuskan :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{S_i}{S_j} * Entropy(S_i) \quad (2)$$

Ket:

S : himpunan suatu kasus

A : atribut

S_i : jumlah kasus ke-i

S_j : jumlah pada dalam himpunan

$Entropy(S_i)$: Entropy pada nilai i

B. Metode Particle Swarm Optimization(PSO)

Algoritma *Particle Swarm Optimization* adalah teknik optimasi yang menyesuaikan

input dan proses untuk menghasilkan output yang maksimal. Algoritma PSO ini terinspirasi oleh perilaku sosial dan alami dari pergerakan burung atau ikan (*bird flocking* atau *fish schooling*). Prosedur pencarian pada metode *particle swarm optimization* ini berbasis pada populasi, ketika individu mengubah posisi terhadap waktu. Pada metode ini, individu akan melakukan pencarian multi dimensional (*multidimensional search space*) yang melakukan proses penyesuaian posisi berdasarkan pengalaman individu tersebut dan pengalaman individu lain di sekitarnya [5].

Contoh, individu ke-i dinyatakan sebagai: $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,d})$ dalam sebuah ruang. Posisi terbaik sebelumnya dari partikel ke-i disimpan dan dinyatakan sebagai $pbest_i = (pbest_{i,1}, pbest_{i,2}, \dots, pbest_{i,d})$. Indeks individu terbaik diantara semua individu dalam group dinyatakan sebagai $gbest_d$. Kecepatan individu ke-i dinyatakan sebagai : $v_i = (v_{i,1}, v_{i,2}, \dots, v_{i,d})$. Modifikasi dari kecepatan dan posisi tiap individu akan dihitung menggunakan kecepatan saat ini dan jarak $pbest_{i,d}$ ke $gbest_d$ seperti bentuk persamaan dibawah ini :

$$v_{i,m} = w.v_{i,m} + c_1 * R * (pbest_{i,m} - x_{i,m}) + c_2 * R * (gbest_{m,m} - x_{i,m})$$

$$x_{i,d} = x_{i,m} + v_{i,m} \quad (3)$$

ket :

n : jumlah individu pada suatu kelompok

d : dimensi

$v_{i,m}$: kecepatan individu ke-i pada iterasi ke-i

w : faktor bobot inersia

c_1, c_2 : konstanta akselerasi

R : angka random (antara 0 - 1)

$x_{i,d}$: posisi individu ke-i menuju iterasi ke-i

$pbest_i$: posisi terbaik sebelumnya dari individu ke-i

$gbest_m$: individu terbaik diantara kelompok.

Referensi menyatakan persamaan menghitung kecepatan baru untuk tiap individu berdasarkan pada kecepatan sebelumnya ($V_{i,m}$), lokasi individu ketika

nilai terbaik dicapai (*pbest*) dan lokasi populasi global (*gbest* untuk versi global, *lbest* untuk versi (*local*) atau *local neighborhood* pada algoritma versi *local* ketika nilai terbaik dicapai. Persamaan memperbaharui posisi tiap individu pada solusi. Dua bilangan acak *c1* dan *c2* dibangkitkan sendiri. Penggunaan berat memberikan performa yang meningkat pada sejumlah aplikasi dengan hasil perhitungan individu yaitu kecepatan individu diantara *interval* [0,1][6].

C. Metode Adaboost

Metode *Adaboost* merupakan salah satu teknik dari *Ensemble Methods / boosting Methods* yang digunakan untuk menyeimbangkan dan mengkombinasikan *record – record* dalam sebuah kelas dengan cara memberikan koefisien bobot berdasarkan *performance* hasil training sebelumnya untuk meningkatkan nilai akurasi dengan meminimalkan *function error* pada klasifikasi. Penerapan metode ini dilakukan dengan melakukan *iteration* pada tiap atribut dengan tahapan metode [7] adalah sebagai berikut :

1. Inisialisasi bobot, $W_n(m)$ untuk $n = 1, 2, \dots, N$.
2. For $m = 1 \dots \dots \dots, M$.
 - a. Minimalkan fungsi kesalahan (*error function*) untuk training $Y_m(x)$ dengan rumus :

$$Jm = \sum_{n=1}^N W^{(m)}_n I(y_m(x_n) \neq t_n) \quad (4)$$

- b. Evaluasi nilai kesalahan

$$\epsilon_m = \frac{\sum_{n=1}^N w_n^{(m)} (y_m(x_n) \neq t_n)}{\sum_{n=1}^N w_n^{(m)}} \quad (5)$$

Lalu menggunakan evaluasi

$$a_m = \ln \left\{ \frac{1 - \epsilon_m}{\epsilon_m} \right\} \quad (6)$$

- c. Perbaiki bobot data

$$w_n^{(m+1)} = w_n^{(m)} \exp(a_m I(y_m(x_n) \neq t_n)) \quad (7)$$

3. Membuat prediksi menggunakan model terakhir sebagai berikut

$$Y_m(x) = \text{sign}(\sum_{m=1}^M a_m y_m(x)) \quad (8)$$

III. METODE PENELITIAN

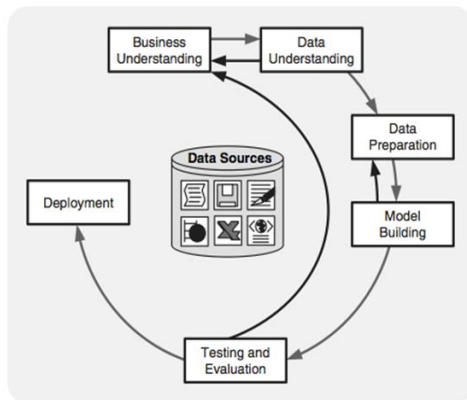
Jenis penelitian yang dilakukan adalah penelitian eksperimen. Metode ini menguji kebenaran dengan statistik dan menghubungkannya dengan masalah penelitian. Tujuan dari penelitian ini adalah melakukan komparasi dan evaluasi model pohon keputusan C4.5 sebagai algoritma terpilih dan pohon keputusan C4.5 berbasis metode *Particle Swarm Optimization (PSO)* dan metode *Adaboost* 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* dengan populasi yang sebanyak 270 *record* dari *dataset* penyakit jantung dibagi menjadi data *training* dan *testing* dengan software *Rapid Miner 7*. Kerangka konsep dari penelitian ini adalah



Gambar 2. Kerangka Konsep Penelitian

IV. HASIL DAN PEMBAHASAN

Proses *data mining* berdasarkan CRISP-DM terdiri dari 6 fase, yaitu : *Business Understanding, Data understanding, Data preparation, Model Building, Testing and Evaluation, Deployment*. Struktur dari metodologi ini akan membuat *data mining* dapat dilakukan dengan lebih cepat dalam prosesnya, lebih ekonomis pengerjaannya, dan mudah untuk diatur dalam pengelolaannya. [8]



Gambar 3. CRISP-DM process

1. Tahap *Business Understanding*, Kegiatan yang dilakukan antara lain :
 - a. Sasaran atau tujuan dari penelitian ini adalah menerapkan modifikasi pada algoritma *decision tree* menggunakan metode *particle swarm optimization* dan metode *adaboost* sebagai analisis komparasi dalam pemilihan parameter yang optimal, sehingga dapat meningkatkan akurasi hasil prediksi penyakit jantung.
 - b. Tujuan data mining untuk menemukan *knowledge* (pengetahuan) yang tersembunyi melalui pola – pola dari data yang berjumlah besar.
 - c. Perencanaan strategis dalam yang dilakukan direalisasikan dalam bentuk jadwal penelitian dilaksanakan
2. Tahap *Data Understanding*

Pada fase ini akan ditentukan atribut-atribut yang digunakan untuk membangun model. Data yang diolah merupakan *dataset* pasien jantung yang diambil dari *Repository UCI (University of California Irvine)* dengan rincian data hasil negatif diagnosa penyakit jantung sebanyak 120 *record* dan pasien diagnosa positif penyakit jantung sebanyak 150 *record*. *Dataset* di klasifikasikan kategori berdasarkan buku induk dari Departemen Kesehatan Indonesia tahun 2013. Terdiri dari *field* sebagai berikut :

Tabel 1. Kategorisasi Nilai Atribut

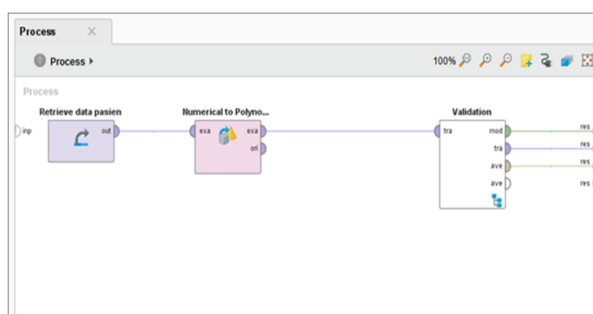
No	Atribut	Nilai	Kategori
1	age	0-5	balita
		5-11	kanak-kanak
		12-16	remaja awal
		17-25	remaja akhir
		26-35	dewasa awal
		36-45	dewasa akhir
		46-55	lansia awal
		56-65	lansia akhir
		>65	manula
2	sex	1	Pria
		2	Wanita
3	chest pain type	1	Typical Angina
		2	Atypica Angina
		3	Non Angina Pain
		4	Asymptomatic

4	resting blood pressure	< 90	Hipotensi
		90-119	Normal
		120-139	Prehipertensi
		140-159	Hipertensi Tingkat 1
		160-179	Hipertensi Tingkat 2
		≥ 180	Hipertensi Tingkat Darurat
5	serum cholestoral	<200	Normal
		200-239	Beresiko
		≥ 240	Kolesterol Tinggi
6	fasting blood sugar	1	Ya
		0	Tidak
7	resting electrocardiographic results	0	Normal
		1	Abnormality
		2	Left Ventricular Hypertrophy
8	maximum heart rate achieved	HRMAX > 220-Usia	Normal
		HRMAX < 220-Usia	Tidak Normal
9	exercise induced angina	0	Tidak
		1	Ya
10	oldpeak	< 1	0
		≥ 1 dan < 2	1
		≥ 2 dan	2

		< 3	
		≥ 3 dan < 4	3
		≥ 4	4
11	slope of the peak exercise ST segment	1	Unsloping
		2	Flat
		3	Downsloping
12	number major vessels	1	1
		2	2
		3	3
13	Thal	3	Normal
		6	Fixed defect
		7	Reversible defect

3. Tahap *Data Preparation*

Data preparation melakukan pemilihan *field* dan pemisahan data *training* dan *testing* secara otomatis dengan fasilitas *X validation*.

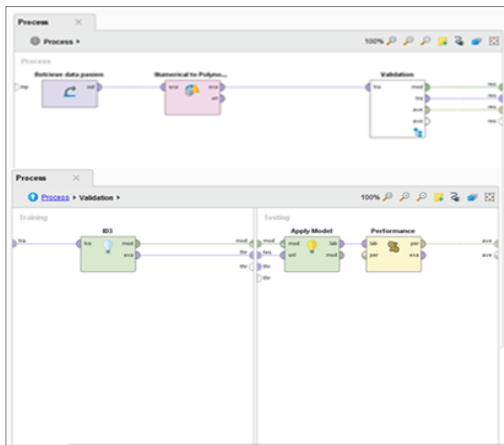


Gambar 4. Validasi data preparation

4. Tahap *Modelling*

Tahap *modelling* dilakukan untuk memilih dan menerapkan teknik pemodelan yang tepat, menentukan *tools* data mining yang digunakan, serta menentukan parameter dengan nilai yang optimal. Tahap

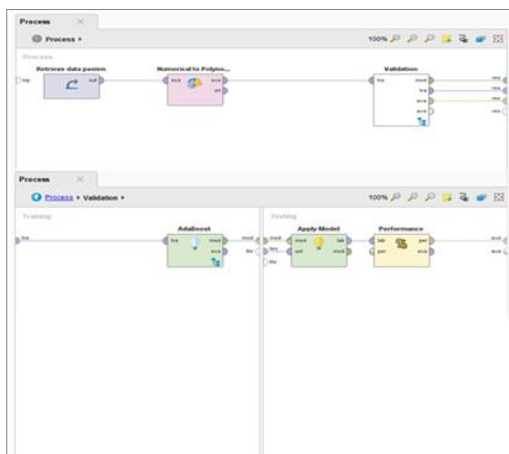
Modelling yang digunakan pada penelitian ini menggunakan tools Rapid Miner 7 dengan model sebagai berikut :



Gambar 5. Modelling Algorithm Decision Tree



Gambar 6. Modelling Algorithm Decision Tree modifikasi dengan metode Particle Swarm Optimization



Gambar 7. Modelling Algorithm Decision Tree modifikasi dengan metode Adaboost

5. Tahap Evaluation

a. Confusion Matrix

Confusion Matrix Evaluation digunakan untuk menentukan nilai kinerja pemodelan algoritma klasifikasi berdasarkan hasil perhitungan nilai akurasi prediktif. Untuk mendapatkan nilai akurasi di dapat dari jumlah prediksi jumlah kasus benar dan prediksi jumlah kasus yang salah. Rule prediksi tersebut adalah :

- (A) pred TIDAK – true TIDAK: jumlah data yang diprediksi tidak dan kenyataannya tidak
- (B) pred YA – true YA: jumlah data yang diprediksi ya dan kenyataannya ya
- (C) pred TIDAK – true YA: jumlah data yang diprediksi tidak tapi kenyataannya ya
- (D) pred YA – true TIDAK: jumlah data yang diprediksi ya tapi kenyataannya tidak

$$\text{Nilai Akurasi} = (A + B) / (A + B + C + D)$$

Nilai Accuracy Confusion Matrix dari pengujian algoritma DECISION TREE dan yang telah di modifikasi adalah sebagai berikut :

accuracy: 72.96% +/- 4.07% (mikro: 72.96%)

	true Ya	true Tidak	class precision
pred. Ya	83	36	69.75%
pred. Tidak	37	114	75.50%
class recall	69.17%	76.00%	

Gambar 8. Nilai Accuracy Algorithm decision tree Prediksi Penyakit Jantung

accuracy: 80.37% +/- 5.51% (mikro: 80.37%)

	true Ya	true Tidak	class precision
pred. Ya	91	24	79.13%
pred. Tidak	29	126	81.29%
class recall	75.83%	84.00%	

Gambar 9. Nilai Accuracy Algorithm decision tree menggunakan metode Particle Swarm Optimization Prediksi Penyakit Jantung

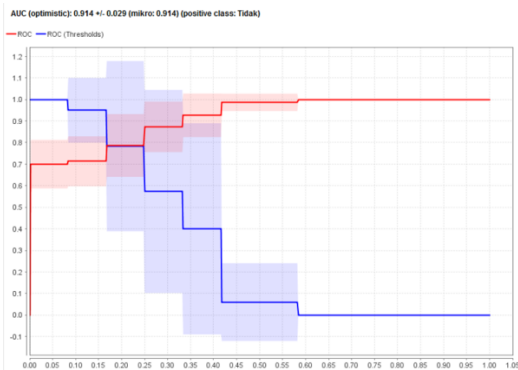
accuracy: 72.96% +/- 4.07% (mikro: 72.96%)

	true Ya	true Tidak	class precision
pred. Ya	83	36	69.75%
pred. Tidak	37	114	75.50%
class recall	69.17%	76.00%	

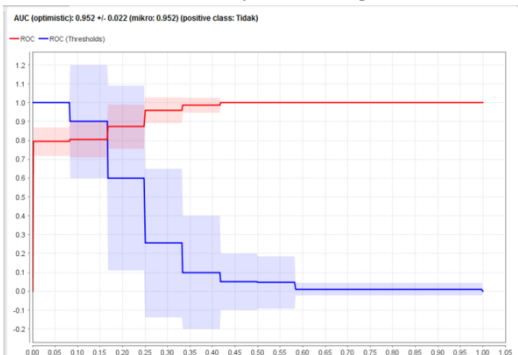
Gambar10. Nilai Accuracy Algorithm decision tree menggunakan metode Adaboost Prediksi Penyakit Jantung

b. ROC Curve Evaluation

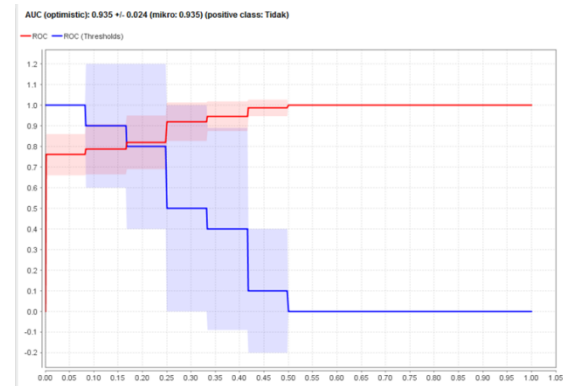
ROC (Receiver Operating Characteristic) curve Evaluation digunakan sebagai visualisasi dan mengelola klasifikasi berdasarkan kinerja menggunakan dua kelas keputusan *positif* atau *negative* yang dipetakan ke salah satu elemen dari himpunan pasangan [6]. AUC adalah singkatan untuk daerah di bawah kurva. Hal ini digunakan dalam analisis klasifikasi untuk menentukan model terbaik yang digunakan dalam memprediksi kelas. Probabilitas dengan variabel *sensitivitas* dan kekhususan (*specificity*) dengan nilai batas antara 0 hingga 1 digambarkan dengan kurva. Area di bawah kurva memberikan gambaran tentang keseluruhan pengukuran atas kesesuaian dari model yang digunakan. Nilai AUC (Area Under Curve) digunakan untuk mengukur akurasi secara umum. Semakin mendekati 1 maka akan semakin baik uji pemodelannya. Gambar berikut ini merupakan hasil dari Nilai AUC (Area Under Curve) dan ROC Curve.



Gambar 11. ROC Curve Algoritma *decision tree* Prediksi Penyakit Jantung



Gambar 12. ROC Curve Algoritma *decision tree* menggunakan metode *Particle Swarm Optimization*



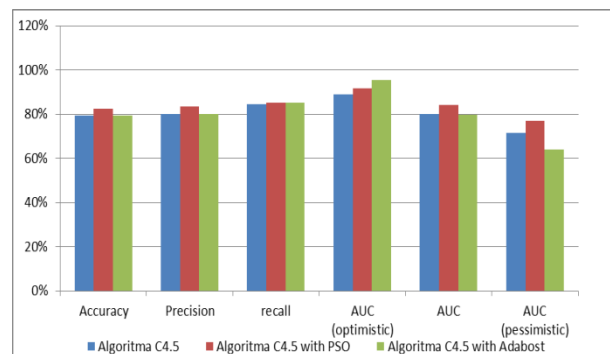
Gambar 13. ROC Curve Algoritma *decision tree* menggunakan metode *Adaboost*

6. Komparasi Nilai

Nilai Akumulasi dari Algoritma *decision tree* dan modifikasi dari Algoritma *decision tree* pada table dibawah ini

Tabel 2. Hasil nilai yang dihasilkan

	Algoritma <i>decision tree</i>	Algoritma <i>decision tree</i> with PSO	Algoritma <i>decision tree</i> with <i>adaboost</i>
Accuracy	79.26%	82.59%	79.26%
Precision	79.94%	83.58%	79.94%
recall	84.67%	85.33%	85.33%
AUC (optimistic)	0.889	0.916	0.955
AUC	0.800	0.841	0.798
AUC (pessimistic)	0.714	0.771	0.641



Gambar 14. Grafik hasil nilai komparasi

V. KESIMPULAN DAN SARAN

Dalam evaluasi teknik data mining klasifikasi menggunakan dua indikator utama yaitu besarnya nilai *accuracy* dan ROC Curve / Area Under Curve (AUC). Berdasarkan hasil perhitungan yang



dilakukan untuk menganalisis modifikasi algoritma *decision tree* untuk prediksi penyakit jantung dengan menggunakan algoritma *Decision Tree* nilai tingkat akurasi 79.26% dan nilai AUC 0.889. Setelah dilakukan modifikasi algoritma *decision tree* dengan metode *particle swarm optimization* nilai tingkat akurasi 82.59% dan untuk nilai AUC 0.916. Modifikasi algoritma *decision tree* dengan metode *adabost* nilai akurasi yang dihasilkan sebesar 79.26% dan nilai AUC sebesar 0.955.

Penelitian ini menggunakan *dataset* bersifat *public dataset* dari *archive UCI (University of California Irvine) Repository* sebanyak 270 record data pasien penyakit jantung dengan rincian data hasil negatif diagnosa penyakit jantung sebanyak 120 *record* dan pasien diagnosa positif penyakit jantung sebanyak 150 *record*. Saran untuk penelitian selanjutnya agar dapat mengambil *dataset* yang dikumpulkan dari berbagai sumber di rumah sakit di Indonesia dan selanjutnya dapat menggunakan kombinasi *ensemble methods techniques* lainnya sebagai *boosting* untuk pemodelan algoritma klasifikasi.

REFERENSI

- S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," *2008 IEEE/ACS Int. Conf. Comput. Syst. Appl.*, vol. 8, no. 8, pp. 343–350, 2008.
- N. ch. s. n. iyenga. M. Anbarasi, e. Anupriya, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm," no. July, 2015.
- K. Srinivas and G. Raghavendra Rao, "Survey on Prediction of Heart Morbidity Using Data Mining Techniques," *Int. J. Data Min. Knowl. Manag. Process*, vol. 1, no. 3, pp. 14–34, 2011.
- S. A. Zega, "Penggunaan Pohon Keputusan untuk Klasifikasi Tingkat Kualitas Mahasiswa Berdasarkan Jalur Masuk Kuliah," *Semin. Nas. Apl. Teknol. Inf.*, pp. 7–13, 2014.
- A. Rosita and Y. Purwananto, "Implementasi Algoritma Particle Swarm untuk Menyelesaikan

Sistem Persamaan Nonlinear," *J. Tek. ITS Vol. 1, (Sept, 2012)*, vol. 1, pp. 1–5, 2012.

- R. Wajhillah, S. Nusa, M. Sukabumi, N. B. View, and T. M. View, "OPTIMASI ALGORITMA KLASIFIKASI C4 . 5 BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK PREDIKSI ...," no. September, 2014.
- X. Wu *et al.*, *Top 10 algorithms in data mining*, vol. 14, no. 1. 2008.
- I. Budiman, T. Prahasto, and Y. Christyono, "Data Clustering Menggunakan Metodologi Crisp-Dm Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma," vol. 2012, no. Snati, pp. 15–16, 2012.