

## KOMPORASI METODE KLASIFIKASI DATA *MINING* UNTUK PREDIKSI PENYAKIT JANTUNG

**Abdul Rohman**

Fakultas Teknik Universitas Pandanaran  
Jl. Banjarsari Barat No.1 Semarang  
email: abdulrohman@unpand.ac.id

### **ABSTRAK**

Dalam dunia kesehatan, penyakit jantung merupakan penyakit yang mendorong angka kematian yang cukup tinggi, sehingga banyak penelitian yang dilakukan sebelumnya untuk memprediksi penyakit jantung. Dalam penelitian ini dilakukan perbandingan metode *data mining* yaitu algoritma *neural network*, *k-nearest neighbor* dan *C4.5* yang diaplikasikan pada data pasien baik yang sehat dan sakit. Dari hasil pengujian dengan mengukur kinerja ketiga metode tersebut menggunakan *cross validation*, *confusion matrix* dan kurva *ROC*, diketahui bahwa *neural network* memiliki nilai akurasi yang paling tinggi yakni sebesar 86,06%, diikuti oleh metode algoritma *C4.5* dengan nilai akurasi 82,92%, kemudian metode *k-nearest neighbor* dengan akurasi 77,58%. Nilai *AUC* untuk metode *neural network* menunjukkan nilai tertinggi yakni 0,913 dan yang terendah adalah metode *C4.5* yaitu 0,857.

Kata Kunci: *data Mining*, algoritma *neural network*, *k-nearest neighbor* dan *C4.5*

### **PENDAHULUAN**

Dalam dunia kesehatan saat ini, sudah banyak mengembangkan data-data pasien untuk keperluan tertentu, terutama dalam pengambilan keputusan. Dengan pengambilan keputusan atas dasar dan informasi yang akurat akan menghasilkan keputusan dan prediksi penyakit yang akurat juga.

Penyakit jantung di Indonesia merupakan penyakit nomor satu yang mendorong angka kematian yang cukup tinggi, sehingga diperlukan untuk diprediksi dengan menggunakan metode klasifikasi *data mining*. Dengan itu praktisi kesehatan dalam pengambilan keputusan bisa lebih tepat dan akurat.

Banyak penelitian prediksi penyakit jantung dengan menggunakan metode klasifikasi *data mining*, diantaranya penelitian yang dilakukan oleh Bakhtiar Rifai, bahwa algoritma *neural network* dalam pengujian terhadap dataset penyakit jantung sebanyak 572 *record* menghasilkan tingkat akurasi 91,45%. (Rifai, 2013)

Penelitian yang dilakukan oleh Mei Lestari, algoritma *K-Nearest Neighbor (K-NN)* dalam mendeteksi penyakit jantung menghasilkan nilai akurasi 70% dengan dataset 100 *record*. (Lestari, 2014)

Selain itu juga penelitian yang dilakukan oleh Adi Suwondo, Dian Asmarajati dan Heri Surahman, algoritma *C4.5 (Decision Tree)*

dalam memprediksi penyakit koroner dengan dataset sebanyak 31 menghasilkan nilai akurasi 91,38% (Suwondo, Asmarajati, & Surahman, 2013)

Dari beberapa penelitian tentang pemecahan masalah penyakit jantung tersebut diatas dengan menggunakan metode klasifikasi *Data Mining* memiliki tingkat akurasi yang berbeda-beda dan dataset serta jumlah *recordnya* yang berbeda pula.

Untuk mengetahui tingkat akurasi yang akurat dari ketiga algoritma yaitu *neural network*, *Nearest Neighbor (K-NN)* dan *C4.5 (Decision Tree)* dengan dataset dan jumlah data yang sama, maka dalam penelitian ini akan melakukan komporasi dari ketiga Algoritma tersebut. (Jasoni dan Steinbrunn, 2011)

### **KAJIAN PUSTAKA**

#### **Penyakit Jantung**

Penyakit jantung merupakan penyakit yang terjadi karena terjadinya gangguan pada jantung terutama pemompa darah sehingga dapat mengakibatkan kematian.

Dataset penyakit jantung yang dipublikasikan oleh UCI (*University of California Irvine*) terdapat 14 atribut yaitu umur, jenis kelamin, jenis sakit dada, tekanan darah, kolesterol, kadar gula, elektrokardiografi, tekanan darah, angina induksi, oldpeak, segmen\_st, dan denyut jantung. Dari data tersebut terdapat

kesimpulan apakah pasien dalam keadaan sehat atau sakit.

**Data Mining**

*Data mining* adalah serangkaian proses mendapatkan pengetahuan atau pola dari kumpulan data (Ian & Witten, 2011). *Data mining* akan memecahkan masalah dengan menganalisis data yang telah ada dalam basis data. *Data mining*, sering juga disebut *knowledge discovery in database (KDD)* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan pola keteraturan, pola hubungan dalam set data berukuran besar (Santoso, 2007). Hasil keluaran dari data mining ini dapat dijadikan untuk memperbaiki pengambilan keputusan di masa depan.

Dalam penelitian ini akan memanfaatkan *data mining* untuk mengklasifikasi data pasien penyakit jantung sehingga hasil keluarannya akan dimanfaatkan untuk keperluan prediksi penyakit jantung

a. Metode Klasifikasi Data Mining

Klasifikasi adalah proses penemuan model (fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bias digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Kamber, 2006).

Proses klasifikasi didasarkan pada empat komponen (Gorunescu, 2011):

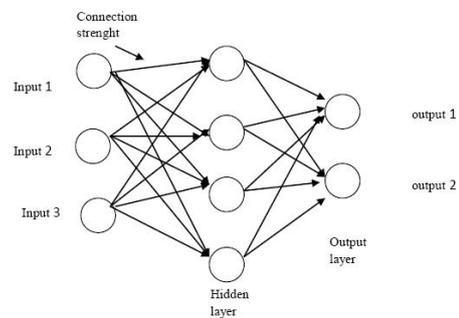
- 1) **Kelas.** Variabel dependen yang berupa kategorikal yang merepresentasikan "label" yang terdapat pada objek. Contohnya: resiko penyakit jantung, resiko kredit, *customer loyalty*, dan jenis gempa.
- 2) **Predictor.** Variabel independen yang direpresentasikan oleh karakteristik (atribut) data. Contohnya: merokok, minum alkohol, tekanan darah, tabungan, aset, dan gaji.
- 3) **Training dataset.** Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.
- 4) **Testing dataset.** Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi

Berikut ini adalah algoritma klasifikasi data mining yang paling populer yaitu adalah (Gorunescu, 2011):

- 1) *Decision/classification trees*
- 2) *Bayesian classifiers/ Naive Bayes classifiers*
- 3) *Neural networks*
- 4) *Statistical analysis*
- 5) *Rough sets*
- 6) *k-nearest neighbor classifier*
- 7) *Rule-based methods*
- 8) *Memory based reasoning*
- 9) *Support vector machines.*

b. Algoritma neural network

*Neural Network* adalah suatu usaha untuk meniru fungsi otak manusia. Otak manusia diyakini terdiri dari jutaan unit pengolahan kecil, yang disebut neuron, yang bekerja secara paralel (Skutla dkk., 2010). Neuron saling terhubung satu sama lain melalui koneksi neuron. Setiap individu neuron mengambil input dari satu set neuron. Kemudian memproses input tersebut dan melewati output untuk satu neuron. Keluaran dikumpulkan oleh neuron lain untuk diproses lebih lanjut. (Skutla dkk., 2010).



Gambar 1. Arsitektur *Neural Network*

Jenis-jenis desain *Neural Network* seperti *back-propagasi*, *recurrent network*, *selft organizing map*, *Bayesian network*.

Langkah pembelajaran dalam algoritma *back-propagasi (Neural Network)* adalah sebagai berikut (Myatt, 2007):

- 1) Inisialisasi bobot jaringan secara acak (biasanya antara -0.0 sampai 1.0)
- 2) Untuk setiap data pada *training*, hitung input untuk simpul berdasarkan nilai input dan bobot jaringan saat itu, menggunakan rumus:

$$input_j = \sum_{i=1}^n o_i w_{ij} + \theta_j$$

Keterangan:

$O_i$  = Output simpul  $i$  dari layer sebelumnya  
 $w_{ij}$  = Bobot relasi dari simpul  $i$  pada layer sebelumnya ke simpul  $j$

$\theta_j$  = bias

- 3) Berdasarkan input dari langkah dua, selanjutnya membangkitkan output untuk simpul menggunakan fungsi aktifitas sigmoid

$$Output = \frac{1}{1 + e^{-Input}}$$

- 4) Hitung nilai *Error* antara nilai yang diprediksi dengan nilai sesungguhnya menggunakan rumus:

$$Error_j = Output_j \cdot (1 - Output_j) \cdot (Target_j - Output_j)$$

Keterangan:

$Output_j$  = *Output* aktual dari simpul *j*

$Target_j$  = Nilai target yang sudah diketahui pada data *training*

- 5) Setelah nilai *Error* dihitung, selanjutnya dibalik ke *layer* sebelumnya. Untuk menghitung nilai *Error* pada *hidden layer*, menggunakan rumus:

$$Error_j = Output_j \cdot (1 - Output_j) \cdot \sum_{k=1}^n Error_k \cdot w_{jk}$$

- 6) Nilai *Error* yang dihasilkan dari langkah sebelumnya digunakan untuk memperbarui bobot relasi menggunakan rumus:

$$w_{ij} = w_{ij} + l \cdot Error_j \cdot Output_i$$

Keterangan:

$w_{ij}$  = bobot relasi dari unit *i* pada *layer* sebelumnya ke unit *j*

*l* = *learning rate* (konstanta, nilainya antara 0 sampai dengan 1)

$Error_j$  = *Error* pada output layer simpul *j*

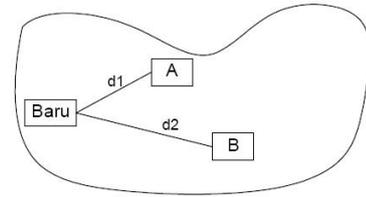
$Output_i$  = *Output* dari simpul *i*

- c. Algoritma *k-Nearest Neighbor (K-NN)*

Algoritma *K-Nearest Neighbord (K-NN)* merupakan suatu metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. *K-Nearest Neighbor (K-NN)* merupakan salah satu metode pengklasifikasian data berdasarkan similaritas dengan label data (Larose, 2006). Similaritas menggunakan matriks jarak dimana satuan jaraknya biasanya menggunakan Euclidean:

$$D(a, b) = \sqrt{\sum_{k=1}^d (a_k - b_k)^2}$$

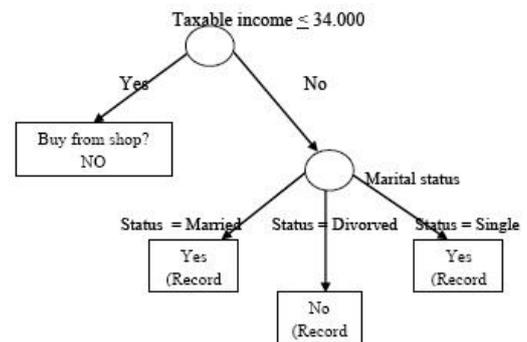
Dimana matriks  $D(a, b)$  adalah jarak skalar dari kedua vektor *a* dan *b* dari matriks dengan ukuran *d* dimensi.



Gambar 2. Ilustrasi Tetangga Terdekat

- d. Algoritma C4.5 (*Decision Tree*)

Algoritma C4.5 (*Decision Tree*) digunakan untuk membangun pohon keputusan yang mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar. (Gorunescu, 2011). *Decision Tree* atau pohon keputusan adalah model prediksi menggunakan struktur pohon atau hirarki dengan mengubah data menjadi pohon keputusan dan aturan-aturan keputusan.



Gambar 3. Contoh Konsep Keputusan Sederhana (Gorunescu, 2011)

Beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 (Gorunescu, 2011) adalah sebagai berikut:

- 1) Mempersiapkan data *training*, dapat diambil dari data *history* yang pernah terjadi sebelumnya dan sudah dikelompokan dalam kelas-kelas tertentu.
- 2) Menentukan akar dari pohon dengan menghitung nilai *gain* yang tertinggi dari masing-masing atribut atau berdasarkan nilai *index entropy* terendah. Sebelumnya dihitung terlebih dahulu nilai index entropy, dengan rumus:

$$Entropy(i) = \sum_{j=1}^m f(i, j) \cdot 2f[(i, j)]$$

- 3) Hitung nilai *gain* dengan rumus:

$$gain = - \sum_{i=1}^p \frac{n_i}{n} I E(i)$$

- 4) Untuk menghitung *gain ratio* perlu diketahui suatu term baru yang disebut *Split Inormation* dengan rumus:

$$Split\ Information = - \sum_{t=1}^c \frac{s1}{s} \log_2 \frac{s1}{s}$$

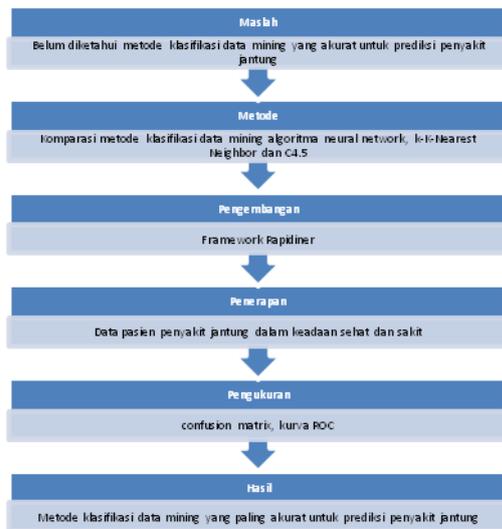
- 5) Selanjutnya menghitung *ratio*

$$Gainratio(S.A) = \frac{Gain(S.A)}{SplitInformation(S.A)}$$

- 6) Mengulangi langkah ke-2 hingga semua *record* terpartisi

e. Kerangka Pemikiran

Sesuai dengan permasalahan diatas, maka kerangka pemikiran yang dilakukan adalah:



Gambar 4. Kerangka Pemikiran

**Metode**

Dalam penelitian ini menggunakan data pasien yang melakukan pemeriksaan penyakit jantung yang didapat dari UCI (Universitas California, Invene) *Machine Learning Repository* (Jasoni dan Steinbrunn, 2011). Hasil yang didapat sebanyak 867 orang yang diperiksa dan sebanyak 364 pasien terdeteksi sakit, sehingga 503 pasient erdeteksi sehat. Dataset tersebuta adalah penggabunganantara dataset dari Cleveland yang terdiri dari 303 pasien, data dari statlog yang terdiri dari 270 pasien, dan data dari Hungaria terdiridari 294 pasien.

Penelitian ini adalah penelitian *experiment* yang melibatkan penyelidikan tentangperlakuan pada parameter dan variabel yang semuanya tergantung pada penelitiitu sendiri. *Software* dan *hardware* sebagai alat bantu dalam penelitian ini adalah sebagai berikut:

Tabel 1. Spesifikasi *Hardware* dan *Software*

Software	Hardware
Sistem Operasi: Windows 7 32 bit	CPU: AMD E-350 1,6 Ghz
Data Mining: RapidMiner Studio 7.3.1	Ram 2 GB, Hdd 320 GB

Data yang diperoleh dari UCI akan di *preprocessing* terlebih dahulu supaya data berkualitas dengan cara manual. Jadi data yang diolah dan diteliti sebanyak 567 dengan keadaan sakit sejumlah 257 orang dan keadaan sehat sejumlah 310 orang.

Tabel 2. Dataset Pasien Penyakit Jantung

umur	Jenis kelamin	Jenis sakitdada	Tekanan darah	Kolesterol	Kadar gula	Elektrokardiografi	Tekanan jantung	Angi-na induksi	Old-peak	Segmenst	Fluoroso-py	Den-yut jan-tung	Hasil
63.0	male	typ_angina	145.0	233.0	t	left_vent_h yper	150.0	No	2.3	down	0.0	fixed_defect	healthy
67.0	male	asympt	160.0	286.0	f	left_vent_h yper	108.0	Yes	1.5	Flat	3.0	normal	sick
67.0	male	asympt	120.0	229.0	f	left_vent_h yper	129.0	Yes	2.6	Flat	2.0	reversable_defect	sick
37.0	male	non_anginal	130.0	250.0	f	normal	187.0	No	3.5	down	0.0	normal	healthy
41.0	female	atyp_angina	130.0	204.0	f	left_vent_h yper	172.0	No	1.4	Up	0.0	normal	healthy
56.0	male	atyp_angina	120.0	236.0	f	normal	178.0	No	0.8	Up	0.0	normal	healthy
62.0	female	asympt	140.0	268.0	f	left_vent_h yper	160.0	No	3.6	down	2.0	normal	sick
57.0	female	asympt	120.0	354.0	f	normal	163.0	Yes	0.6	Up	0.0	normal	healthy
63.0	male	asympt	130.0	254.0	f	left_vent_h yper	147.0	No	1.4	Flat	1.0	reversable_defect	sick

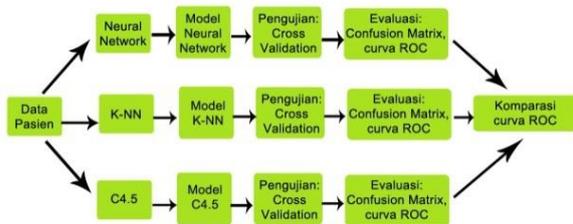
53.0	male	asympt	140.0	203.0	t	left_vent_h yper	155.0	Yes	3.1	down	0.0	reversable _defect	sick
57.0	male	asympt	140.0	192.0	f	normal	148.0	No	0.4	Flat	0.0	fixed_def ect	healthy
56.0	female	atyp_angina	140.0	294.0	f	left_vent_h yper	153.0	No	1.3	Flat	0.0	normal	healthy
56.0	male	non_anginal	130.0	256.0	t	left_vent_h yper	142.0	Yes	0.6	Flat	1.0	fixed_def ect	sick
44.0	male	atyp_angina	120.0	263.0	f	normal	173.0	No	0.0	Up	0.0	reversable _defect	healthy
52.0	male	non_anginal	172.0	199.0	t	normal	162.0	No	0.5	Up	0.0	reversable _defect	healthy
57.0	male	non_anginal	150.0	168.0	f	normal	174.0	No	1.6	Up	0.0	normal	healthy
48.0	male	atyp_angina	110.0	229.0	f	normal	168.0	No	1.0	down	0.0	reversable _defect	sick
54.0	male	asympt	140.0	239.0	f	normal	160.0	no	1.2	Up	0.0	normal	healthy
48.0	female	non_anginal	130.0	275.0	f	normal	139.0	no	0.2	Up	0.0	normal	healthy

Model yang diusulkan pada penelitian ini adalah menggunakan tiga metode klasifikasi data *mining*. Metode yang diusulkan untuk pengolahan dataset pasien adalah penggunaan algoritma *neural network*, *k-Nearest Neighbor (K-NN)* dan *C4.5*, setelah diolah dan menghasilkan model, maka dilanjutkan dengan pengujian menggunakan *k-fold cross validation*, kemudian dilakukan evaluasi dan validasi hasil dengan *confusion matrix* dan kurva *ROC*. Tahap selanjutnya adalah membandingkan hasil akurasi dan AUC dari tiap model sehingga diperoleh model dari metode klasifikasi yang memperoleh nilai akurasi dan AUC tertinggi.

*Performance* keakurasian AUC (Gorunescu, 2011), dapat diklasifikasikan menjadi lima kelompok yaitu:

- a. 0.90 – 1.00 = *Exellent Clasification*
- b. 0.80 – 0.90 = *Good Clasification*
- c. 0.70 – 0.80 = *Fair Clasification*
- d. 0.60 – 0.70 = *Poor Clasification*
- e. 0.50 – 0.60 = *Failure*

Dalam tahapannya seperti gambar dibawah ini:

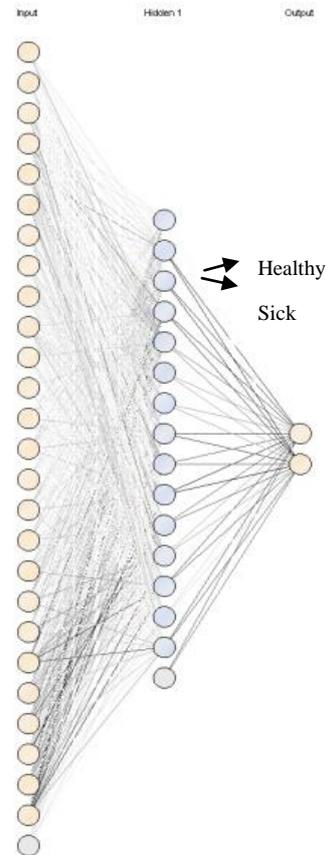


Gambar 5. Metode yang diusulkan

## HASIL DAN PEMBAHASAN

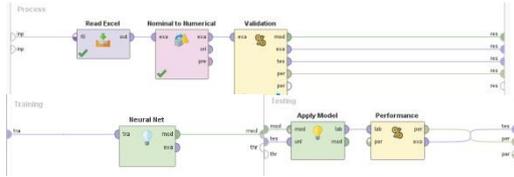
### Algoritma Neural Network

Setelah melakukan Preprocessing data, peneliti melakukan klasifikasi dengan menggunakan algoritma *Neural Network* dengan *training cycles: 500* dan *learning rate: 0,3* sehingga menghasilkan model seperti gambar dibawah ini



Gambar 6. Model Algoritma Neural Network

Tahap selanjutnya pengujian model algoritma *neural network* tersebut diatas dengan *Cross Validaton*, yaitu teknik validasi dengan membagi data secara acak ke dalam *k* bagian dan masing-masing bagian akan dilakukan proses klasifikasi (Kamber, 2006). Dengan menggunakan *cross validation* akan dilakukan percobaan sebanyak *k*. Data yang digunakan dalam percobaan ini adalah data *training* untuk mencari nilai *error rate* secara keseluruhan. Secara umum pengujian nilai *k* dilakukan sebanyak 10 kali untuk memperkirakan akurasi estimasi. Dalam penelitian ini nilai *k* yang digunakan berjumlah 10 atau *10-fold Cross Validation*.



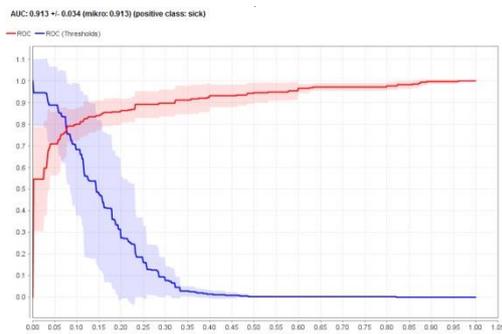
Gambar 7. Pengujian K-Fold Cross Validation Metode Neural Network

Hasil pengujian tersebut diatas diperoleh *accuracy* 86,06% dan *AUC* 0,913.

accuracy: 86.06% +/- 5.41% (mikro: 86.07%)

	true healthy	true sick	class precision
pred healthy	273	42	86.67%
pred sick	37	215	85.32%
class recall	98.06%	83.66%	

Gambar 8. Nilai Akurasi *Neural Network*



Gambar 9. Grafik Area Under Curva *Neural Network*

**Algoritma *K-Nearest Neigboard***

Setelah melakukan *Preprocessing dataset* penyakit jantung, peneliti melakukan klasifikasi dengan menggunakan algoritma *K-Nearest Naegboard (K-NN)*, dengan mengklaster *k=2* dari data 567, dengan keadaan sakit sejumlah 257 orang dan keadaan sehat sejumlah 310

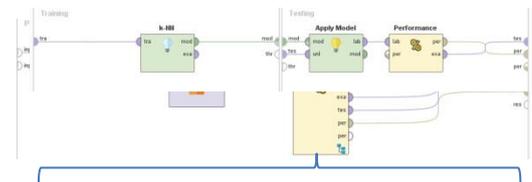
orang. Maka model yang diperoleh seperti hasil eksekusi dibawah ini.

**KNNClassification**

2-Nearest Neighbour model for classification.  
The model contains 567 examples with 13 dimensions of the following classes:  
healthy  
sick

Gambar 9. Model Algoritma *K-Nearest Neigboard*

Tahap selanjutnya pengujian model algoritma *K-Nearest Neigboard* tersebut diatas dengan *Cross Validaton*.



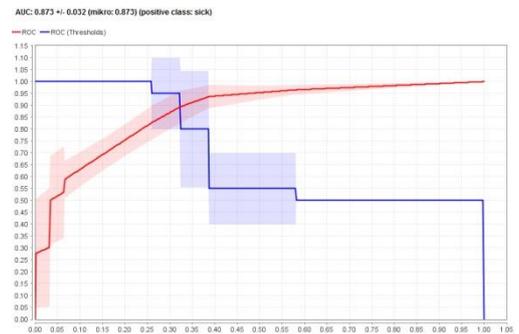
Gambar 10. Pengujian K-Fold Cross Validation Metode *K-Nearest Neigboard*

Hasil pengujian tersebut diatas diperoleh *accuracy* 77,58% dan *AUC* 0,873.

accuracy: 77.58% +/- 3.87% (mikro: 77.60%)

	true healthy	true sick	class precision
pred healthy	305	122	71.43%
pred sick	5	135	96.43%
class recall	98.39%	52.53%	

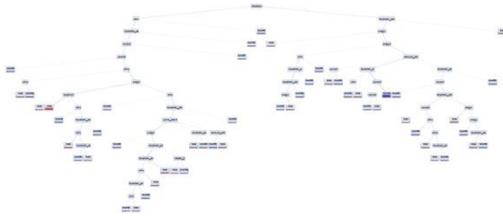
Gambar 11. Nilai Akurasi *K-Nearest Neigboard*



Gambar 12. Grafik Area Under Curva *K-Nearest Neigboard*

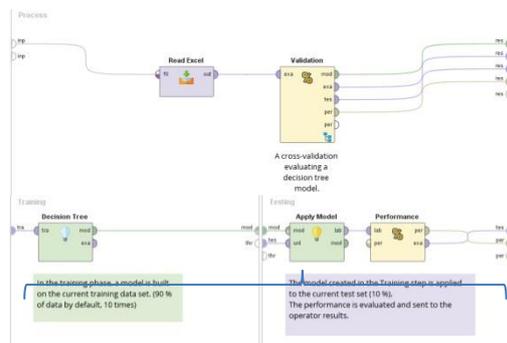
**Algoritma *C4.5 (Decision Tree)***

Dalam eksperimennya menggunakan algoritma *C4.5 (Decision Tree)* terhadap dataset penyakit jantung menggunakan *criterion* yaitu *gaint\_ration* dan *maximal depht* yaitu 20. Maka model yang dihasilkan seperti gambar dibawah ini



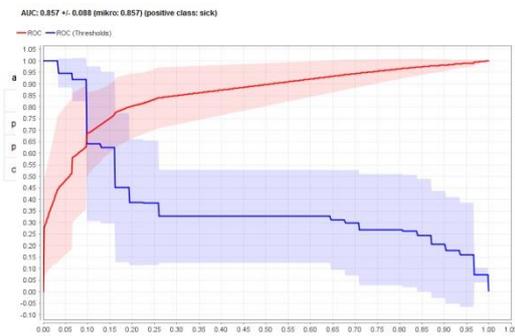
Gambar 13. Model Algoritma C4.5 (Decision Tree)

Dalam pengujian K-Fold Cross Validation Algoritma C4.5 dan Algoritma C4.5 berbasis Adaboost, peneliti juga menggunakan 10 kali percobaan dengan sampling type Stratified (bertingkat-tingkat).



Gambar 14. Pengujian K-Fold Cross Validation Metode C4.5

Hasil pengujian tersebut di atas diperoleh *accuracy* 82,92% dan *AUC* 0,857.



Gambar 15. Grafik Area Under Curva C4.5

Berdasarkan dari analisis pengujian masing-masing metode di atas maka dapat dirangkum hasilnya seperti Tabel 3 berikut:

Tabel 3. Perbandingan *Performance* Metode

	NN	K-NN	C4.5
<i>Accuracy</i>	86,06%	77,58%	82,92%
<i>AUC</i>	0,913	0,873	0,857

Melihat hasil perhitungan yang terangkum pada Tabel 3 di atas dengan menerapkan klasifikasi *performance* keakurasian *AUC* maka diperoleh hasil penelitian yaitu, terdapat satu metode yang merupakan kategori *Excellent Clasification* yaitu untuk metode *Neural Network* dengan nilai *AUC* 0,913, dan kategori *Good Clasification* terdapat 2 metode yaitu *K-Nearest Neigboard* dengan *AUC* 0,873 dan *C4.5* dengan nilai *AUC* 0,857.

Berdasarkan Tabel 3 di atas juga dapat dilihat bahwa nilai akurasi maka dapat diketahui metode yang terbaik adalah *Neural Network* dengan nilai akurasi 86,06%, yang kedua adalah algoritma *C4.5* dengan nilai akurasi 82,92%, dan yang terakhir adalah *K-Nearest Neigboard* dengan nilai 77,58%

**SIMPULAN**

Dari penelitian yang telah dilakukan menggunakan algoritma *neural network*, *k-nearest neigboard* dan *c4.5* menggunakan data pasien penyakit jantung untuk prediksi ketepatan penyakit jantung. Model yang dihasilkan dikomparasi untuk dapat diketahui metode terbaik dalam prediksi ketepatan penyakit jantung. Untuk mengukur kinerja model digunakan *confusion matrix* dan kurva *ROC*, dan diketahui bahwa Metode algoritma *neural network* menghasilkan nilai akurasi yaitu 86,06% dan nilai *AUC* 0,913, Metode algoritma *C4.5* dengan nilai akurasi 82,92% dan nilai *AUC* 0,857, dan yang terakhir adalah *K-Nearest Neigboard* dengan nilai 77,58% dan nilai *AUC* 0,873.

Dengan demikian metode *neural network* adalah metode yang terbaik untuk pemecahan masalah prediksi penyakit jantung.

**DAFTAR PUSTAKA**

Gorunescu, F., 2011, *Data mining concepts models and technique*, Springer, Berlin.

Ian and Witten, 2011, *Data Mining Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, United States of America.

Jasoni and Steinbrunn, 2011, *datasets Heart Disease*, Retrieved 2017, from UCI Machine Learning Repository: <https://archive.ics.uci.edu>

- Kamber, 2006, *Data Mining Concept and Tehniques*, Morgan Kauffman, San Fransisco.
- Larose, 2006, *Data Mining Methods and Models*, Joh Wiley & Sons.Inc., Canada.
- Lestari, M., 2014, Penerapan Algoritma Klasifikasi Nearest Neighbor (K-NN) untuk Prediksi Penyakit Jantung, *Faktor Exacta*, 366-371.
- Myatt, G., 2007,. *Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining*, Neural Network, New Jersey.
- Rifai, B., 2013, Algoritma Neural Network Untuk Prediksi Penyakit Jantung, *Techno Nusa Mandiri*, 1-9.
- Santoso, 2007, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu, Yogyakarta.
- Skutla, A., Tiwari, R., and Kala, R., 2010, *Real Life Application of Soft*, Taylor and Francis Groups, LLC.
- Suwondo, A., Asmarajati, D., and Surahman, H., 2013, Algoritma C4.5 Berbasis Adaboost Untuk Prediksi Penyakit Jantung Koroner, *Fastikom Unsiq*, 1-11.