

MODEL NEURAL NETWORK BERBASIS ADABOOST DALAM MEMPREDIKSI PENYAKIT JANTUNG



Bakhtiar Rifai

14000319

LATAR BELAKANG

- Industri kesehatan memerlukan keputusan yang efektif dalam mengambil keputusan
- Penyakit jantung adalah penyakit yang paling ditakuti manusia, dan penyakit jantung banyak diderita oleh masyarakat Indonesia
- Faktor gejala dalam mendiagnosa penyakit jantung diantaranya jenis sakit dada, tekanan darah tinggi, kolesterol, nilai tes EKG, denyut jantung dan kadar gula.
- Dibutuhkan suatu metode yang paling akurat dalam memprediksi pasien berpenyakit jantung apa tidak.

LATAR BELAKANG (2)

- Salah satu cara dengan menggunakan klasifikasi data mining
- Dengan tujuan yang diinginkan yaitu dapat memprediksi pasien penyakit jantung dengan lebih tepat
- Perlu dilakukan evaluasi dan alaisa terhadap data record pasien penyakit jantung

INDETIFIKASI MASALAH

- Teknik dalam mengindetifikasi penyakit jantung sangat diperlukan.
- Kurang akuratnya metode-metode lain dalam mendeteksi penyakit jantung.
- Bagaimana neural network berbasis adaboost lebih akurat dan lebih baik dalam mendeteksi penyakit jantung

TUJUAN PENELITIAN

- Bagaimana *neural network* berbasis adaboost untuk meningkatkan akurasi dalam prediksi penyakit jantung.

MANFAAT PENELITIAN

- Manfaat praktis : diharapkan digunakan untuk dokter sebagai alat bantu dalam mendeteksi penyakit jantung.
- Manfaat teoritis : memberikan sumbangsih bagi pengembangan teori yang berkaitan dengan prediksi penyakit jantung dengan menggunakan *neural network* berbasis *adaboost*.
- manfaat kebijakan sebagai penunjang dalam mendeteksi penyakit jantung.

RUANG LINGKUP PENELITIAN

- Penelitian ini dibatasi pada *penerapan Artificial Neural Network* berbasis *adaboost* untuk menghasilkan prediksi penyakit jantung yang lebih akurat, dengan atribut yang diolah adalah umur, jenis kelamin, Atribut umur, jenis kelamin, tekanan darah, kolestrol, kadar gula, elektrokardiografi, tekanan jantung, angina induksi, olpeak, segmen_st, flaurosopy, denyut jantung dan hasil (atribut prediksi).

TINJAUAN STUDI

- M. Anbarasi, E. Anupriya, dan Inyenger (2010) meneliti : peningkatan prediksi jantung dengan subset fitur seleksi dengan genetik algoritma. Dengan tiga belas atribut dengan 6 atribut dan di klasifikasika menggunakan *naïve bayes*, *decision tree* dan *classification via clustering*. dengan hasil decision tree lebih baik dalam penggunaan ggenetik algoritma

TINJAUAN STUDI (2)

- Penelitian yang dilakukan tanwei xing, jie wang dan shing zhao tahun 2007. Meneliti : kombinasi metode data mining dengan data medis baru untuk memprediksi hasil dari penyakit jantung koroner dengan 3 metode klasifikasi, SVM, ANN dan decision tree. Hasil svm lebih baik dalam memprediksi penyakit jantung koroner

TINJAUAN STUDI (3)

- Sellappan palaniappan, Rafiah awang tahun 2008. meneliti sistem cerdas prediksi penyakit jantung menggunakan teknik data mining. Dengan tiga algoritma decision tree, naïve bayes dan neural network. Dengan hasil penelitian naïve bayes lebih baik dalam memprediksi dari decision tree dan neural network sangat sulit memahami

TINJAUAN STUDI (4)

- G. subblakshmi, K. Remesh, M. Chinna Rao tahun 2011 dengan meneliti pendukung keputusan dalam sistem prediksi penyakit jantung menggunakan naïve bayes dengan memodelkan data mining dengan teknik bayes sederhana dengan hasil penelitian bahwa naïve beyes modle yang sangat efektif dalam memprediksi dan kemudahan interpresentasi model.

TINJAUAN STUDI (5)

- Anchana Khemphila dan Veera Boonjing tahun Penelitian ini menggunakan pendekatan Multi-Layer Perceptron (MLP) dengan Back-Propagation Learning Algorithm untuk mendeteksi penyakit jantung dengan mengurangi atribut yang biasa digunakan menggunakan 13 atribut menjadi 8 atribut dengan nilai akurasi yang diperoleh adalah 1.1% dan validasi dataset 0.82 %

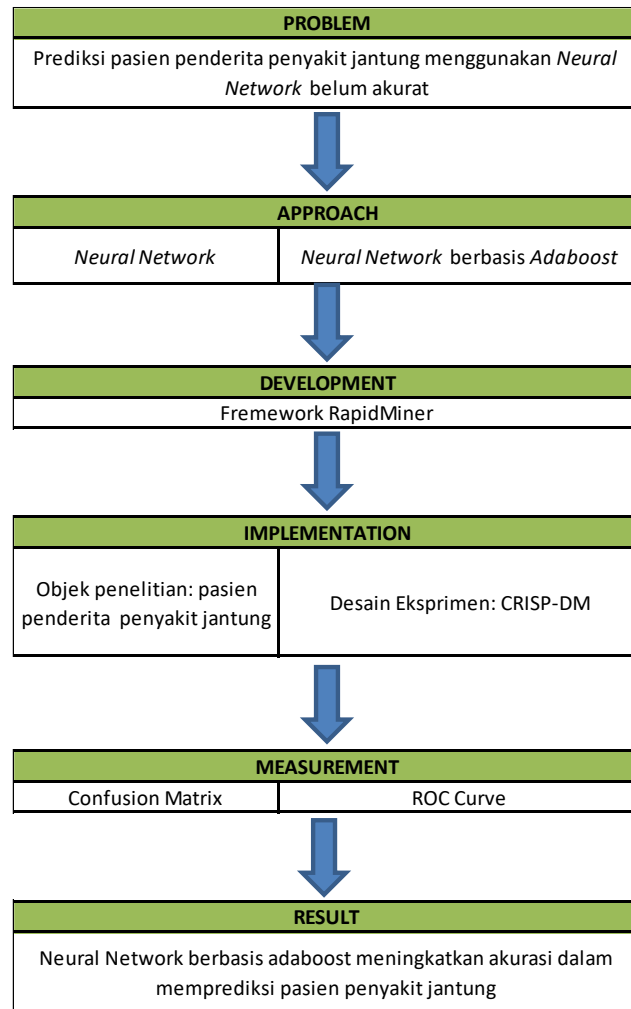
TINJAUAN PUSTAKA

- Dalam penulisan tesis ini penulis melakukan tinjauan studi dengan menggunakan buku dan jurnal yang berhubungan dengan tema yang dipilih.

ALGORITMA

- ***Neural Network*** yaitu suatu metode komputasi yang meniru sistem jaringan saraf. Metode ini menggunakan elemen perhitungan non-linear dasar yang disebut neuron yang diorganisasikan sebagai jaringan yang saling berhubungan.
- ***Adaboosts yaitu*** metode untuk meningkatkan ketelitian dalam proses klasifikasi dan prediksi dengan cara membangkitkan kombinasi dari sebuah model, tetapi hasil klasifikasi atau prediksi yang dipilih adalah metode yang memiliki nilai bobot besar.

KERANGKA PEMIKIRAN



PEMODELAN DATA MINING

- Model CRISP-DM dengan tahapan:
 - *Business Understanding*
 - *Data Understanding*
 - *Data Preparation*
 - *Modelling*
 - *Evaluation*
 - *Deployme*

BUSINESS UNDERSTANDING

- Dalam penelitian ini *business understanding* adalah data pemeriksaan penyakit jantung yang didapat sebanyak 573 orang yang diperiksa dan sebanyak 259 terdeteksi sakit dan 314 yang terdeteksi sehat. Untuk analisa penyakit jantung baik menggunakan *neural network* dengan tujuan hasil yang diberikan lebih akurat, maka penulis melakukan pengembangan dengan *neural network* dengan tujuan hasil yang didapat lebih baik dari *neural network*

DATA UNDERSTANDING

- Data understanding dimulai dengan pengumpulan data awal penelitian dan kemudian dijelaskan dan di eksplorasi setiap atribut kemudian kualitas data diverifikasi. Untuk menentukan prediksi penyakit jantung dibutuhkan tiga belas atribut predictor dan satu atribut kelas

DATA PREPARATION

- Pada tahap ini dilakukan pemisahan data yang mengandung duplikasi, anomali atau inkonsisten data
- Untuk meningkatkan keakurasian data akan dilakukan replace missing

MODELLING

- Pada tahap modeling ini dilakukan pemrosesan data traning sehingga akan membahas metode algoritma yang di uji dengan memasukan data penyakit jantung kemudian di analisa dan dikomparasi

NEURAL NETWORK

- *Neural network* adalah metode algoritma untuk pelatihan supervised dan di desain untuk operasi pada *feed forward* multi lapis. Neural network bisa di deksripsikan sebagai berikut: ketika jaringan diberikan pola masukan sebagai pola pelatihan maka pola tersebut menuju ke unit-unit pada lapisan tersembunyi untuk diteruskan ke unit-unit lapisan terluaran

ADABOOST

- Pada algoritma *adaboost* pertama-tama nilai bobot untuk setiap tuple pada himpunan D diinisialisasi sebesar $1/d$ (d =jumlah anggota himpunan D) setelah suatu model M_i dibangunkan, kemudian dihitung tingkat kesalahan (*error rate*), dengan cara menjumlahkan nilai bobot dari setiap tuple yang *misclassified* (salah klasifikasi) pada M_i . Error rate ini dapat dihitung dengan.
- $error(M_i) = \sum_j^d W_j \times err(X_j)$

METODE EVALUASI

1. Cross Validation

- *Cross validation* (Redaelzadeh, Tang, & Liu, 2008) adalah metode statistic untuk mengevaluasi algoritma dan membandingkan dengan membagi data menjadi dua segmen, satu digunakan untuk *learning* dan satu lagi digunakan untuk memvalidasi model.

METODE EVALUASI

2. *Confusion matrix*

- *Confusion matrix* memberikan keputusan yang diperoleh dalam *training* dan *testing* (Bramer, 2007), *confusion matrix* memberikan penilaian *performance* klasifikasi berdasarkan objek dengan benar atau salah (Gorunescu, 2011)

Classification	Predicted Class		
Observed Class		Class = YES	Class = NO
	Class = YES	<i>a</i> (<i>true positive</i> – TP)	<i>b</i> (<i>false negative</i> – FN)
	Class = NO	<i>c</i> (<i>false positive</i> – FP)	<i>d</i> (<i>true negative</i> – TN)

METODE EVALUASI

- *Curva ROC*
- Untuk dapat melihat akurasi secara manual dilakukan perbandingan klasifikasi menggunakan *curva ROC* hasil eksperesi dari *confusion matrix*. ROC menghasilkan dua garis dengan *bentuk true positives* sebagai garis vertical dan *false positives* sebagai garis *horizontal*

HASIL PENELITIAN

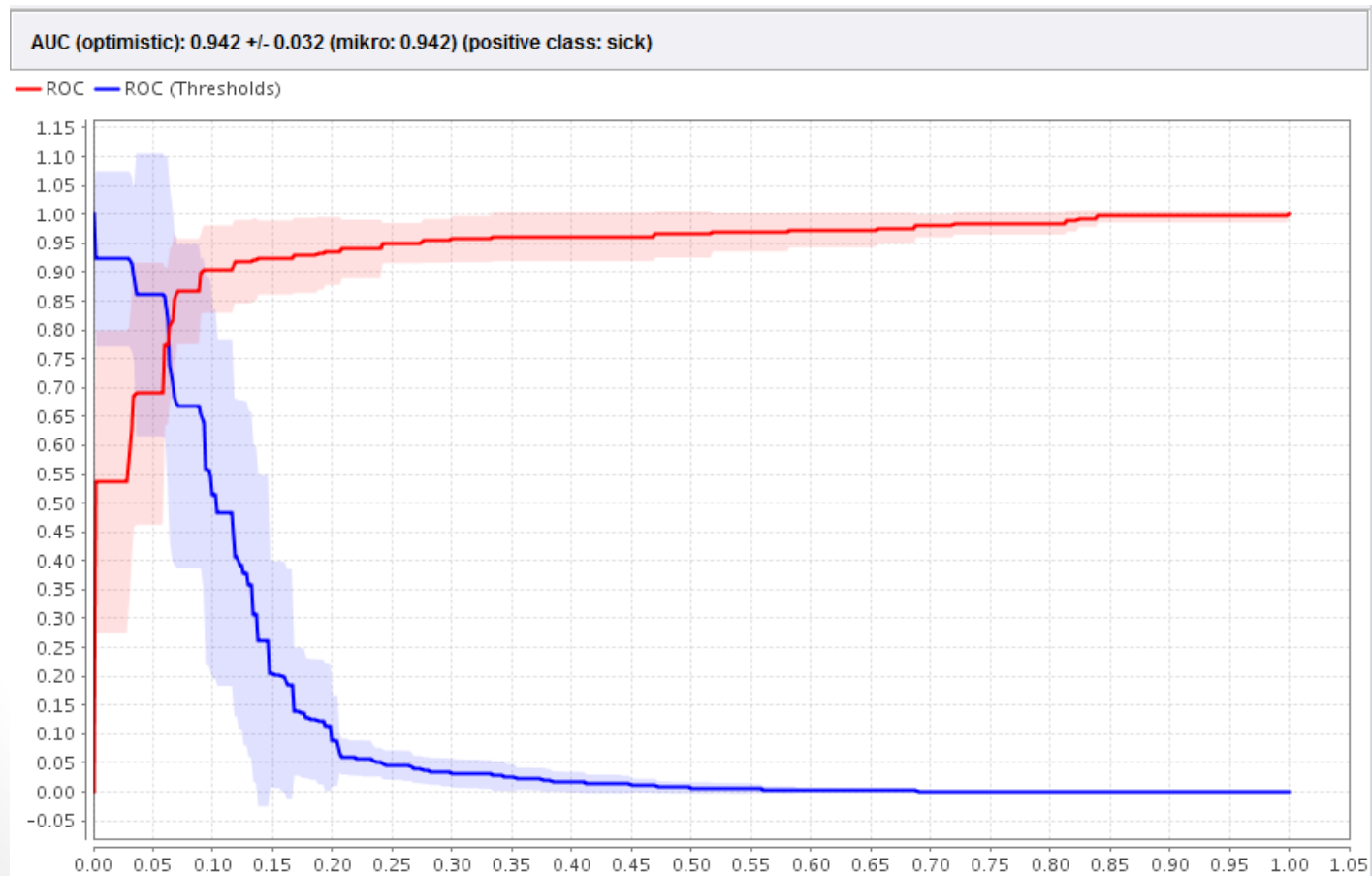
- Pengujian Neural Network

PerformanceVector

```
PerformanceVector:
accuracy: 91.64% +/- 3.91% (mikro: 91.62%)
ConfusionMatrix:
True:   healthy sick
healthy:    291    25
sick:    23    234
precision: 91.20% +/- 4.51% (mikro: 91.05%) (positive class: sick)
ConfusionMatrix:
True:   healthy sick
healthy:    291    25
sick:    23    234
recall: 90.61% +/- 6.95% (mikro: 90.35%) (positive class: sick)
ConfusionMatrix:
True:   healthy sick
healthy:    291    25
sick:    23    234
AUC (optimistic): 0.942 +/- 0.032 (mikro: 0.942) (positive class: sick)
AUC: 0.942 +/- 0.032 (mikro: 0.942) (positive class: sick)
AUC (pessimistic): 0.942 +/- 0.032 (mikro: 0.942) (positive class: sick)
```

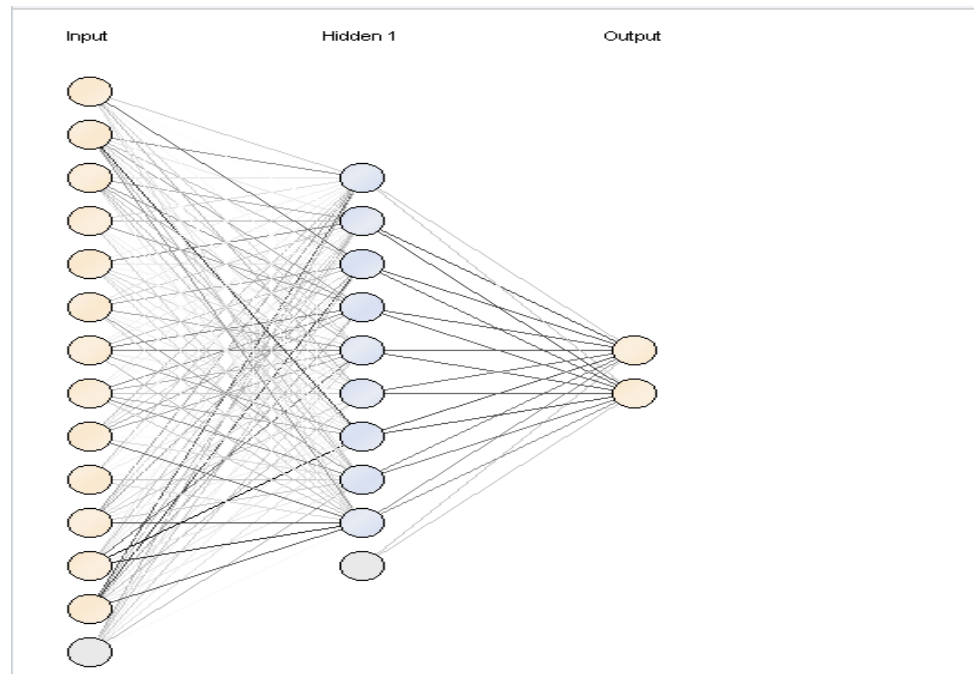
HASIL PENELITIAN

- Pengujian Neural Network AUC



HASIL PENELITIAN

- Neural net



Neural Network adaboost

PerformanceVector

PerformanceVector:

accuracy: 93.03% +/- 4.17% (mikro: 93.02%)

ConfusionMatrix:

True: healthy sick

healthy: 290 16

sick: 24 243

precision: 91.28% +/- 5.02% (mikro: 91.01%) (positive class: sick)

ConfusionMatrix:

True: healthy sick

healthy: 290 16

sick: 24 243

recall: 94.06% +/- 8.72% (mikro: 93.82%) (positive class: sick)

ConfusionMatrix:

True: healthy sick

healthy: 290 16

sick: 24 243

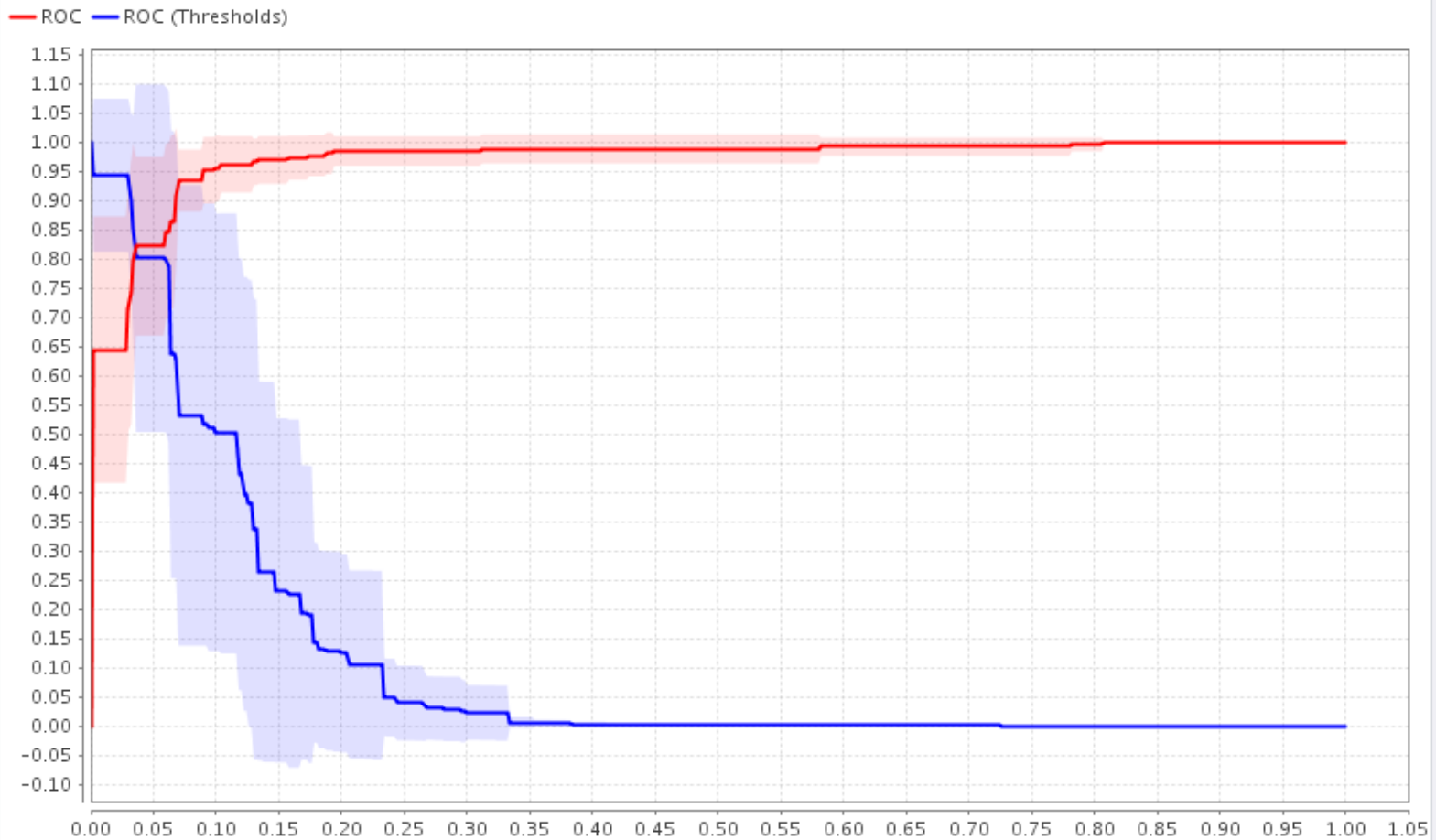
AUC (optimistic): 0.971 +/- 0.021 (mikro: 0.971) (positive class: sick)

AUC: 0.970 +/- 0.022 (mikro: 0.970) (positive class: sick)

AUC (pessimistic): 0.968 +/- 0.024 (mikro: 0.968) (positive class: sick)

Neural Network adaboost

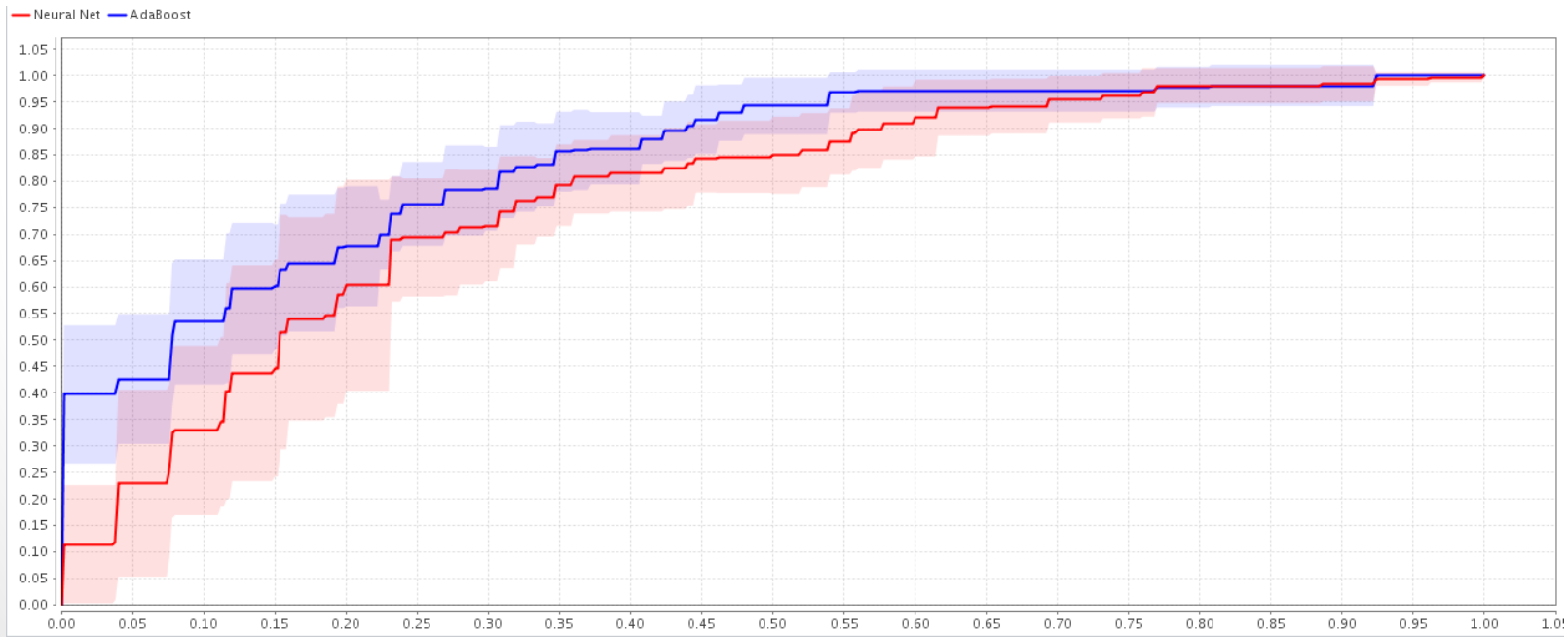
AUC (optimistic): 0.971 +/- 0.021 (mikro: 0.971) (positive class: sick)



- Hasil analisa dari penelitian neural network berbasis adaboost adalah dengan membuat iterasi sebanyak sepuluh kali iteransi dan menghasilkan 10 neural network dengan nilai w berbeda dari setiap *neural*. Dengan setiap neural adalah
- Model 1 [$w = 3.488$] (neural net)
- Model 2 [$w = 2.829$] (neural net)
- Model 3 [$w = 1.989$] (neural net)
- Model 4 [$w = 2.154$] (neural net)
- Model 5 [$w = 1.377$] (neural net)
- Model 6 [$w = 0.471$] (neural net)
- Model 7 [$w = 0.371$] (neural net)
- Model 8 [$w = 0.215$] (neural net)
- Model 9 [$w = 0.001$] (neural net)
- Model 10 [$w = 0.000$] (neural net)

PERFORMANCE KOMPARASI

	Neural Network	Neural Network Adaboost
Accuracy	91.64 %	93.03 %
Presicion	91.20 %	91.28 %
Recall	90.61 %	94.06 %
AUC	0.942	0.971



ANALISIS HASIL KOMPARASI

- dengan membandingkan nilai *accuracy* dan AUC terlihat bahwa algoritma *neural network* berbasis *adaboost* memiliki nilai *accuracy* dan nilai AUC yang lebih baik dibandingkan *neural network* dan dapat disimpulkan bahwa nilai AUC *neural network* dan *neural network* berbasis *adaboost* antara 0.90 – 1.00 termasuk klasifikasi sangat baik

KESIMPULAN

- Model yang dihasilkan diuji untuk mendapatkan nilai accuracy, precision, recall dan AUC dari setiap algoritma sehingga didapat pengujian dengan menggunakan neural network didapat nilai accuracy adalah 91.64 % dengan nilai precision 91.20 % dan nilai AUC adalah 0.942. sedangkan pengujian dengan menggunakan neural network berbasis adaboost didapatkan nilai accuracy 93.03 % dengan nilai precision 91.28 % dan nilai AUC adalah 0.971. maka dapat disimpulkan pengujian model penyakit jantung dengan menggunakan neural network dengan neural network berbasis adaboost didapat bahwa pengujian neural network berbasis adaboost lebih baik dari pada neural network sendiri.

SARAN

- Dari hasil pengujian yang telah dilakukan dan hasil kesimpulan yang diberikan maka ada saran atau usul yang di berikan antara lain:
 1. Menambah beberapa atribut atau mengurangi atribut yang ada dengan melakukan pengujian model kembali.
 2. Untuk dapat melihat lebih jauh tentang algoritma data mining, diperlukan perbandingan atau komparasi dengan metode algoritma yang lain seperti Support vector Machine, Naïve Bayes.
 3. Menggunakan metode seleksi atribut dengan menggunakan Chi-Square untuk ketepatan seleksi atribut