

Penentuan Prediksi Awal Penyakit Jantung Menggunakan Algoritma Back Propagation Neural Network dengan Metode Adaboost

Ricardus Anggi Pramunendar¹, Ika Novita Dewi², Hasan Asari³

^{1,3}Program Pasca Sarjana Universitas Dian Nuswantoro Semarang

¹ricardus.anggi@research.dinus.ac.id, ³mas.acan2@gmail.com

²Fakultas Ilmu Komputer Universitas Dian Nuswantoro Semarang

¹ikadewi@research.dinus.ac.id

ABSTRAK

Penyakit jantung adalah terjadinya penyumbatan sebagian atau total dari suatu lebih pembuluh darah, akibatnya adanya penyumbatan maka dengan sendirinya suplai energi kimiawi ke otot jantung berkurang, sehingga terjadi gangguan keseimbangan antara suplai dan kebutuhan. Dari tahun ke tahun jumlah kasus pasien dengan penyakit jantung semakin meningkat. Sebagai upaya preventif dalam penanganan penyakit jantung perlu dilakukan usaha untuk melakukan prediksi lebih awal pasien dengan penyakit jantung. Hasil prediksi awal yang dilakukan dapat digunakan oleh para petugas medis sebagai alat bantu dalam penentuan penyakit jantung dan langkah awal penanganannya. Penerapan prediksi awal penyakit jantung dapat dilakukan dengan teknik komputasi cerdas menggunakan algoritma Back Propagation Neural Network (BPNN) dengan penambahan metode Adaboost. Hasil eksperimen menunjukkan bahwa nilai akurasi hasil prediksi menggunakan algoritma BPNN adalah 96,65 % dan algoritma BPNN dengan metode Adaboost menjadi 99,29 %, sehingga dapat dikatakan bahwa algoritma neural network dengan penambahan metode adaboost memiliki hasil prediksi yang baik dalam memprediksi penyakit jantung seorang pasien.

Kata kunci : Prediksi penyakit jantung, algoritma BPNN, metode adaboost

1. PENDAHULUAN

Jantung merupakan suatu organ otot berongga yang terletak dipusat dada yang menompa darah lewat pembuluh darah. Penyakit jantung di Indonesia merupakan penyakit nomor satu yang mendorong angka kematian yang cukup tinggi, sehingga sampai sekarang penyakit tersebut ditakuti oleh manusia. Penyakit jantung terjadi karena penyumbatan sebagian atau total dari suatu lebih pembuluh darah, akibat dari adanya penyumbatan maka dengan sendirinya suplai energi kimiawi ke otot jantung berkurang, sehingga terjadi gangguan keseimbangan antara suplai dan kebutuhan darah. Menurut The World Heart Federation, penyakit jantung adalah penyakit cardiovasular yang sering menjangkit pada anak-anak dan orang dewasa [1]. Penyakit jantung mempunyai beberapa jenis diantaranya penyakit jantung koroner, angina, serangan jantung dan gagal jantung [3]. Gejala-gejala pasien yang terdiagnosa penyakit jantung diantaranya adalah jenis sakit dada (*chest pain*), tekanan darah tinggi (*trebtps*), kolesterol (*chol*), nilai tes EKG (*resting electrodiagraphic*), denyut jantung (*thalach*) dan kadar gula FBS (*fasting blood sugar*) [2].

Melihat banyaknya kasus pasien dengan penyakit jantung, maka perlu dilakukan suatu langkah dini sebagai upaya penanganan dan pencegahan penyakit jantung, seperti melakukan prediksi awal penentuan penyakit jantung. Prediksi awal ini perlu dilakukan karena sering kali keputusan klinis yang dibuat oleh petugas medis berdasarkan intuisi dan pengalaman dalam menangani pasien-pasien penyakit jantung. Hal ini bisa dihindari dengan memanfaatkan data-data pasien yang sudah tersimpan dalam basisdata untuk dibuat suatu pola penentuan penyakit jantung dengan teknik komputasi cerdas sehingga ketidaktepatan diagnosis dapat dihindari. Penerapan teknik komputasi cerdas ini sudah pernah dilakukan beberapa peneliti lain. Palaniappan dan Awang dengan melakukan perbandingan tiga algoritma prediksi yaitu *Naives Bayes (NB)*, *Decision Tree (DT)*, dan *Artificial Neural Network (ANN)* dengan total kasus 909 dan 15 atribut. Hasil dari penelitian tersebut menunjukkan bahwa prediksi dengan DT menghasilkan 89%, NB 86,53% dan ANN 85,53% [4]. Shukla dkk melakukan prediksi penyakit jantung dengan menggunakan parameter *age, sex, trestbps, chol, fbs, rest ecg, thalach, exang, oldpeak, slope, CA, dan thal* dan membandingkan algoritma *Back Propagation (BP)*, *Radial Basis Function Network (RBFN)*, *Learning Vector Quantization (LVQ)* dengan hasil akurasi masing-masing algoritma BPA 86,70%, RBF 78,33%, dan LVQ 86,70% [5]. Anbarasi dkk memprediksi kelangsungan hidup penyakit jantung dengan berdasarkan 909 kasus dengan menggunakan metode NB, DT dan Clasification Via Clustering dengan hasil akurasi prediksi DT 99,2%, NB 96,5% dan Clasication Via Clustering 88,3% [8]. Srinivas dkk melakukan deteksi penyakit jantung dengan menggunakan algoritma DT, *Multi-Layer Perceptron (MLP)*, *Baysian Model*, dan *Support Vevtor Machine (SVM)* dan parameter yang digunakan adalah *age, sex, chest pain type, fasting blood sugar, restecg, exercise, slope, CA, trest blood pressure, cholestrol, thalach, dan old peak*, dengan hasil akurasi DT 82%, MLP 89,2%, *Baysian Model* 82%, dan SVM 83,5% [6].

Dari beberapa hasil penelitian yang telah dilakukan sebelumnya sudah menunjukkan hasil akurasi yang baik, namun masih perlu ditingkatkan lagi agar performa suatu algoritma dapat menghasilkan prediksi dengan tepat dan akurasi yang tinggi. Penelitian ini akan melakukan prediksi awal penyakit jantung dengan menggunakan algoritma NN dan menambahkan metode Adaboost untuk mengoptimalkan performa NN sehingga menghasilkan tingkat akurasi prediksi yang tinggi. Metode Adaboost adalah teknik optimasi yang handal dengan mengkombinasikan beberapa pengklasifikasian dasar (*multiple base classifiers*) untuk menghasilkan suatu pengklasifikasian yang kuat (*strong classifier*) [7]. Metode adaboost melakukan *training* terhadap klasifikasi baru menggunakan dataset dengan koefisien bobot yang diatur berdasarkan performansi hasil *training* klasifikasi sebelumnya untuk meminimalkan fungsi kesalahan (*error function*) dengan memperbaiki (*update*) bobot data.

Penerapan algoritma NN dengan penambahan metode Adaboost dilakukan dengan melakukan perulangan (*iteration*) dan pemberian bobot pada tiap atribut (*attribute weighting*) untuk meningkat akurasi dalam prediksi awal penyakit jantung dengan menggunakan dataset yang digunakan berasal dari UCI *Machine Learning Repository*. Data set terdiri dari 14 atribut yaitu umur, jenis kelamin, jenis sakit dada, tekanan darah, kolestrol, kadar gula, elektrokardiografi, tekanan darah, angina induksi, oldpeak, segmen_st, flaurosopy, dan denyut jantung. Hasil prediksi awal yang didapatkan dapat digunakan oleh para petugas medis sebagai alat bantu dalam penentuan penyakit jantung dan langkah awal penanganannya.

2. PREDIKSI MENGGUNAKAN ALGORITMA NEURAL NETWORK DENGAN METODE ADABOOST

2.1 Back propagation dan Multi-Layer Perceptron Neural Network

Ada dua proses learning dalam algoritma back propagation (BPNN) yaitu *feedforward* dan *proses backward*. Dalam proses *feedforward*, dataset diterapkan sebagai input dan jaringan menyebarkan pola input dari lapisan ke lapisan sampai output pola yang dihasilkan. Proses dalam lapisan input (*i*) ke lapisan hidden (*j*) ditunjukkan dengan

$$O_j = f(\text{net}_j) = \frac{1}{1 + e^{-\text{net}_j}} \quad (1)$$

$$\text{net}_j = \sum_j w_{ij} o_i + \theta_j \quad (2)$$

output diperoleh dari penjumlahan beban input dari sebuah node dan peta ke fungsi aktivasi jaringan. Proses pada lapisan hidden (*j*) menuju lapisan output (*k*) ditunjukkan dengan

$$O_k = f(\text{net}_k) = \frac{1}{1 + e^{-\text{net}_k}} \quad (3)$$

$$\text{net}_k = \sum_k w_{jk} o_j + \theta_k \quad (4)$$

Kesalahan dihitung untuk mengukur perbedaan antara yang output diinginkan dan output aktual yang telah dihasilkan dalam fase feedforward.

$$\text{error} = \frac{1}{2} (\text{output}_{\text{desire}} - \text{output}_{\text{actual}})^2 \quad (5)$$

backpropagation diterapkan dari output (*k*) ke hidden (*j*) seperti yang ditunjukkan

$$w_{kj}(t+1) = w_{kj}(t) + \Delta w_{kj}(t+1) \quad (6)$$

$$\Delta w_{kj}(t+1) = \eta \delta_k o_j + \alpha \Delta w_{kj}(t) \quad (7)$$

dengan

$$\delta_k = o_k(1 - o_k)(t_k - o_k) \quad (8)$$

dimana $w_{kj}(t)$ bobot dari node *k* ke node *j*, η merupakan learning rate, α merupakan momentum rate, t_k error node *k*, O_j actual network output untuk node *j*, O_k actual network output untuk node *k*, t_k adalah target output nilai untuk node *k*. Perhitungan hidden (*j*) ke input (*i*) dengan

$$w_{ji}(t+1) = w_{ji}(t) + \Delta w_{ji}(t+1) \quad (9)$$

$$\Delta w_{ji}(t+1) = \eta \delta_j o_i + \alpha \Delta w_{ji}(t) \quad (10)$$

dengan

$$\delta_j = o_j(1 - o_j) \sum_k \delta_k w_{kj} \quad (11)$$

selajutnya

$$O_k = f(\text{net}_k) = \frac{1}{1 + e^{-\text{net}_k}} \quad (12)$$

$$\text{net}_k = \sum_k w_{kj} o_j + \theta_k \quad (13)$$

Proses ini diulang secara berulang-ulang sampai konvergensi tercapai [9].

2.2 Metode Adaboost

Metode boosting dilakukan dengan memanfaatkan model hasil klasifikasi atau prediksi yang memiliki nilai bobot paling besar. Salah satu algoritma boosting yang populer adalah Adaboost. Pada permasalahan klasifikasi dua kelas, input vektor x_1, x_2, \dots, x_n dan target t_1, t_2, \dots, t_n dimana $t_n \in \{-1, 1\}$. Setiap data diberikan parameter bobot (nilai awal adalah $1/n$ untuk semua data). Pada proses pelatihan klasifikasi dasar menggunakan bobot data untuk fungsi $y(x) \in \{-1, 1\}$. Setiap tahapan algoritma, adaboost melatih klasifikasi baru menggunakan dataset dengan koefisien bobot yang diatur berdasarkan performansi hasil pelatihan pengklasifikasi sebelumnya, sehingga memberikan bobot besar data yang salah klasifikasi. Setelah pengklasifikasi dasar sudah melakukan proses pelatihan, selanjutnya dikombinasikan ke bentuk committee menggunakan koefisien dengan bobot berbeda ke pengklasifikasi dasar yang berbeda [9].

3. DESAIN EKSPERIMEN

3.1 Dataset

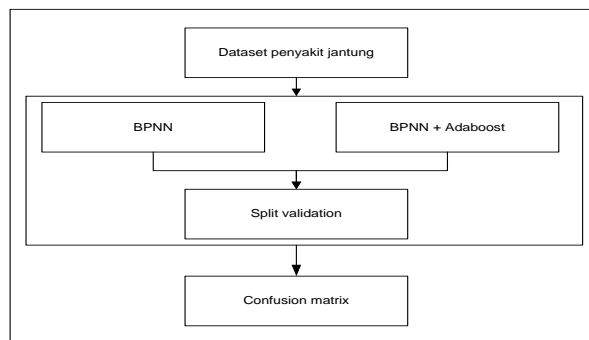
Dataset penyakit jantung berasal dari UCI *Machine Learning Repository* [10]. Data yang digunakan sebanyak 867 record, dengan menerapkan *data cleaning* didapatkan data dengan atribut kosong berjumlah 300 record sehingga data dengan atribut yang lengkap berjumlah 567 record yang akan digunakan dalam eksperimen dengan rincian atribut label sakit 256 record dan label sehat 311 record. Atribut-atribut yang digunakan adalah umur, jenis kelamin, jenis sakit dada, tekanan darah, kolestrol, kadar gula, elektrokardiografi, tekanan darah, angina induksi, oldpeak, segmen_st, flurosopy, denyut jantung dan atribut label yang terdiri atas sehat dan sakit.

Tabel 1: Dekripsi atribut

Atribut	Keterangan	Nilai	Atribut	Keterangan	Nilai
umur	umur dalam tahun	Continuous	tekanan jantung	Tekanan jantung mak.	Continuous
jenis kelamin	laki atau perempuan	1 = laki-laki 2 = Perempuan	angina induksi	Latihan angina induksi	0= no 1 = yes
jenis sakit dada	jenis nyeri dada	1 = typical type 1 2 = typical type angina 3 = non-angina pain 4 = asymptomatic	oldpeak	induksi depresi ST	Continuous
tekanan darah	Tekanan darah saat istirahat	Continuous mmhg	segmen_st	Kemiringan ST	1 = unsloping 2 = flat 3 = downsloping
kolestrol	Kolesterol dalam darah	Continuous mm/dl	flurosopy	floursopy	0-3 value
kadar gula	Kadar gula dalam darah	0 = normal 1 = having_ST_T wave abnormal 2 = left ventricular hypertrophy	denyut jantung	Jenis denyut jantung	3 = normal 6 = fixed 7 = reversible defect
elektrokardiogr afi	Eelektrokardiograp saat istirahat	1 ≥ 120 mg/dl 0 ≤ 120 mg/dl	Hasil	Kondisi Pasien (label)	0 = sick 1 = healthy

3.2 Langkah eksperimen

Langkah-langkah penerapan dataset untuk prediksi awal penyakit jantung dengan algoritma NN dan NN dengan metode Adaboost dapat dilihat dalam gambar 1.



Gambar 1 Desain Eksperimen

Eksperimen dilakukan dengan mengimplementasikan dataset penyakit jantung terhadap algoritma BPNN dan BPNN dengan menggunakan metode adaboost menggunakan 10 x-validation dan evaluasi menggunakan confusion matrix untuk mendapatkan keakurasian hasil prediksi.

4. HASIL EKSPERIMEN DAN PEMBAHASAN

4.1 Pengolahan data awal

Dataset penyakit jantung yang didapatkan dilakukan preprocessing untuk mendapatkan data yang berkualitas dengan menerapkan *data cleaning*, yaitu data yang nilai atributnya kosong akan dihapus agar data menjadi lebih akurat. Dataset penyakit jantung sebanyak 867 record, setelah proses data cleaning didapatkan data dengan atribut yang kosong berjumlah 300 record dan data dengan atribut lengkap berjumlah 567 record yang akan digunakan dalam proses eksperimen dan terdiri dari 256 record berlabel sakit dan 311 record berlabel sehat.

4.2 Pengujian model neuron

Dataset sebelum dan sesudah proses data cleaning akan diujikan terhadap BPNN untuk mendapatkan tingkat akurasi, dengan hasil perbandingan dapat dilihat pada tabel 2.

Tabel 2. Perbandingan akurasi data sebelum dan sesudah proses datacleaning

Jumlah dataset	Accuracy	AUC
867	84,44	0,914
567	94,00	0,957

Langkah berikutnya adalah menguji model neuron dengan menggunakan satu hidden layer dengan jumlah neuron 3 sampai dengan 19 dan menentukan nilai error menggunakan RMSE yang paling kecil. Hasil terbaik dengan satu hidden layer didapatkan pada hidden layer dengan jumlah neuron 15 dan RMSE yang dihasilkan sebesar 0.133. Hasilnya dapat dilihat pada tabel 3.

Tabel 3. Percobaan Penentuan Jumlah Size Hidden Layer.

RMSE	Jumlah Neuron
0,250	3
0,206	5
0,184	7
0,180	9
0,157	11
0,180	13
0,133	15
0,134	17
0,136	19

Selanjutnya menentukan nilai learning rate dan momentum dengan memasukkan nilai learning rate antara 0.10 sampai dengan 0.40 dan nilai momentum dengan range 0.80 sampai dengan 0.95 dengan menggunakan satu hidden layer dan 15 size neuron. Hasilnya terdapat pada tabel 4.

Tabel 4. Hasil dari model neuron dengan struktur 13-15-1

Momentum	Learning Rate	Sum of Square Error	Momentum	Learning Rate	Sum of Square Error
0,80	0,10	0,020	0,90	0,25	0,052
0,85	0,10	0,021	0,95	0,25	0,061
0,90	0,10	0,029	0,80	0,30	0,027
0,95	0,10	0,031	0,85	0,30	0,054
0,80	0,15	0,018	0,90	0,30	0,082
0,85	0,15	0,024	0,95	0,30	0,107
0,90	0,15	0,026	0,80	0,35	0,038
0,95	0,15	0,046	0,85	0,35	0,067
0,80	0,20	0,019	0,90	0,35	0,102
0,85	0,20	0,033	0,95	0,35	0,123
0,90	0,20	0,045	0,80	0,40	0,035
0,95	0,20	0,054	0,85	0,40	0,096
0,80	0,25	0,027	0,90	0,40	0,107
0,85	0,25	0,040	0,95	0,40	0,116

Berdasarkan hasil uji tabel 4, maka parameter NN yang dipilih adalah satu hidden layer dengan jumlah neuron 15, nilai momentum 0.80, learning rate 0.15 dan RMSE 0.018.

4.4 Pengujian hasil validasi

Pengujian validasi dilakukan dengan membandingkan dua cara, yaitu x-validation dan split validation. Hasil pengujian terdapat dalam tabel 5.

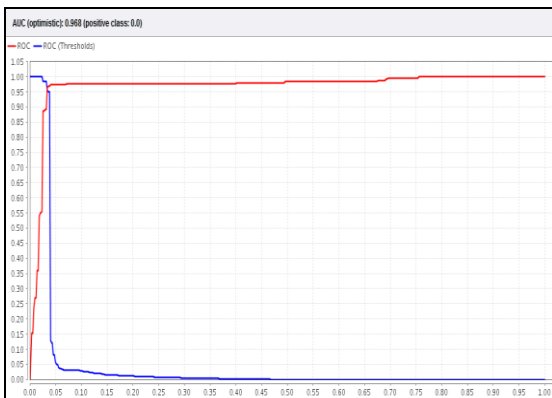
Tabel 5. Perbandingan *X Validation* dan *Split Validation*

Cross Validation <i>Sampling type : Stratified use local random seed</i>		Split Validation <i>Split : Relative Sampling type : shuffled sampling</i>	
Number Validation	Accuracy (%)	Split Data	Accuracy (%)
K = 3	82,19	60% (training) dan 40% (testing)	92,24
K = 5	82,36	70% (training) dan 30% (testing)	94,89
K = 7	83,60	80% (training) dan 20% (testing)	95,41
K = 10	84,30	90% (training) dan 10% (testing)	96,65

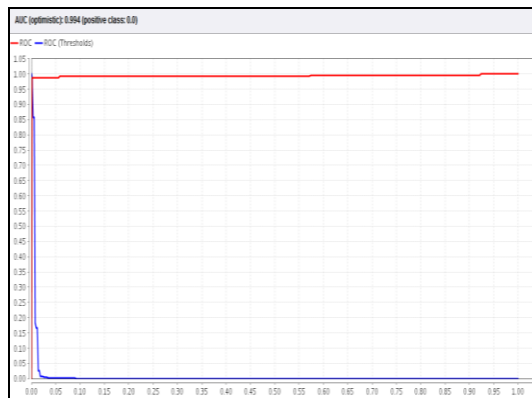
Dari perbandingan pengujian validasi antara x-validation dan split validation dengan mengganti variabel K-Fold dan split data diperoleh hasil bahwa split validation mempunyai tingkat akurasi paling tinggi, yaitu 96,65 %. Sehingga untuk pengujian validasi pada penelitian ini menggunakan split validation.

4.4 Evaluasi dan validasi hasil

Hasil pertama yang akan ditampilkan adalah hasil pengujian dataset penyakit jantung dengan algoritma BPNN. Hasilnya menunjukkan nilai *accuracy* 96.65%, *sensitivity* 97,40%, *specificity* 95,75%, *ppv* 96,46%, dan *npv* 96,88%, kurva ROC dapat dilihat pada gambar 2 dengan nilai AUC sebesar 0.967 dengan hasil prediksi termasuk dalam kategori *excellent classification*.



Gambar 2 Kurva ROC hasil prediksi dengan BPNN



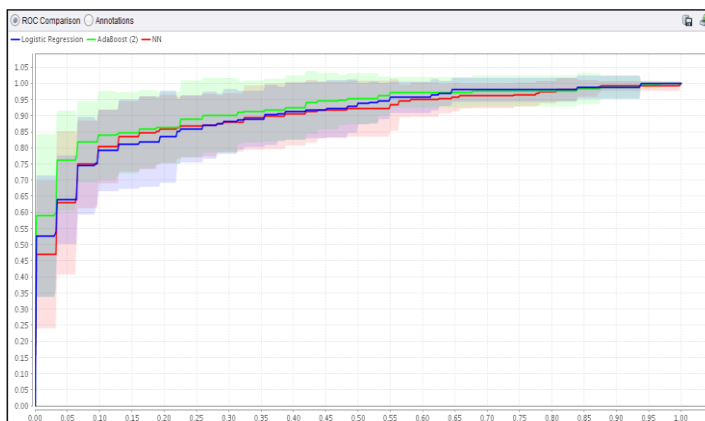
Gambar 3 Kurva ROC hasil prediksi dengan BPNN dengan Adaboost

Selanjutnya, akan ditampilkan hasil pengujian dataset penyakit jantung dengan algoritma BPNN dengan metode adaboost. Hasilnya menunjukkan nilai *accuracy* 99,29%, *sensitivity* 99,04%, *specificity* 99,61%, *ppv* 99,68%, dan *npv* 98,83%, kurva ROC dapat dilihat pada gambar 3 dengan nilai AUC sebesar 0.994 dengan hasil prediksi termasuk dalam kategori *excellent classification*.

Sehingga didapatkan perbandingan hasil akurasi prediksi awal penyakit jantung yaitu BPNN dengan adaboost memiliki tingkat akurasi lebih tinggi dibandingkan dengan BPNN saja. Hasil perbandingannya terlihat pada tabel 6 dan kurva ROC pada gambar 4.

Tabel 6. Perbandingan tingkat akurasi dan AUC

Metode	Accuracy	AUC
neural network (NN)	96,65	0,968
neural network (NN) dengan adaboost	99,29	0,994



Gambar 4. Perbandingan kurva ROC BPNN dan BPNN dengan Adaboost

5. PENUTUP

Pengujian algoritma BPNN dan BPNN dengan Adaboost dilakukan untuk melakukan prediksi awal penyakit jantung. Hasil eksperimen menunjukkan bahwa nilai akurasi dengan BPNN adalah 96,65 % dan nilai AUC adalah 0.968, sedangkan nilai akurasi BPNN dengan adaboost adalah 99.29 % dan nilai AUC adalah 0.994. Dapat disimpulkan pengujian model prediksi awal penyakit jantung menggunakan BPNN dengan Adaboost lebih tinggi akurasinya, dengan peningkatan akurasi sebesar 2,64% dan peningkatan nilai AUC sebesar 0,026%. Penelitian selanjutnya direkomendasi untuk menyempurnakan lagi penelitian yang telah dilakukan ini, misalnya meningkatkan tingkat akurasi dengan mengurangi waktu proses yaitu melakukan penyelesaian dan penambahan atribut dataset dan mengoptimasi menggunakan algoritma Bagging, Stacking atau Metacost.

DAFTAR PUSTAKA

- [1] Hananta I Y and Muhammad H F, *Dietisien Deteksi Dini & Pencegahan* Use the "Insert Citation" button to add citations to this document. *7 Penyakit Penyebab Mati Muda*. Yogyakarta: Media Pressindo, 2011.
- [2] Mahmood M A and Kuppa M R, "Early Detection Of Clinical Parameters In Heart Disease By Improved Decision Tree Algorithm," *Second Vaagdevi Internasional Conference on Information Technology for Real World Problems*, pp. 24-28, 2010.
- [3] Anooj P.K., "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," *Journal of King Saud University Computer and Information Sciences*, pp. 27-40, November 2012.
- [4] Sellappan Palaniappan & Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques," *IJCSNS International Journal of Computer Science and Network Security*, vol. 8, Agustus 2008.
- [5] Shukla Anupam, Ritu Tawari, and Rahul Kala, *Real Life Application of Soft Computing*. New York, United State of America: CRC Press, 2010.
- [6] K. Srinivas, G. Raghavendra Rao, and Govardhan, "Survey On Prediction of Heart Morbidity using Data Mining Tehnique ," *International Journal of Data Mining & Knowledge Management Process*, vol. 1, Mei 2011.
- [7] Ethem Alpaydın, *Introduction to machine learning*, Thomas Dietterich, Ed. London, England: Massachusetts Institute of Technology, 2010.
- [8] E. Anupriya & Inyegar M. Anbarasi, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm," *International Journal of Engineering Science and Technology*, vol. 02, 2010.
- [9] Wu, Xingdong, *The Top Ten Algorithm in Data Mining*. Minnesota: Taylor & Francis Group, 2009
- [10] 11A Jasoni and W Steinbrunn, UCI MACHine Learning Repository.: Retrieved from UCI MACHine Learning Repository:<http://archive.ics.uci.edu/ml/datasets/Heart+Disease>, 2011.