

Penentuan Penyakit Peradangan Hati Dengan Menggunakan Neural Network Backpropagation

Rudianto

AMIK Bina Sarana Informatika Jakarta
rudianto.rdt@bsi.ac.id

Abstract - There is a medical record that had kept the symptoms of inflammation of the liver disease patients and the diagnosis of the disease. This sort of thing is certainly very useful for health professionals. They can use a medical record that has been there as an aid to making decisions about the diagnosis of the patient's disease. In this study will be used backpropagation neural network method for determination of inflammatory liver disease. From the test results to measure the performance of the method is to use testing methods confusion matrix and ROC curve, it is known that the method of back propagation neural network has a value of 84,62% accuracy and AUC value of 0.801. This shows that the model produced, including the classification is quite good because it has the AUC values between 0.8-0.9.

Keywords: Confusion Matrix, Neural Network Backpropagation, ROC Curva

Abstrak - Terdapat catatan rekam medis yang telah menyimpan gejala-gejala penyakit peradangan hati pasien dan diagnosis penyakit tersebut. Hal semacam ini tentu sangat berguna bagi para ahli kesehatan. Mereka dapat menggunakan catatan rekam medis yang telah ada sebagai bantuan untuk mengambil keputusan tentang diagnosis penyakit pasien. Pada penelitian ini akan digunakan metode *neural network backpropagation* untuk penentuan penyakit peradangan hati. Dari hasil pengujian untuk mengukur performa dari model menggunakan metode pengujian *Confusion Matrix* dan Kurva ROC, diketahui bahwa metode *neural network backpropagation* memiliki tingkat akurasi 84,62% dan dengan nilai area under the curva (AUC) sebesar 0,801. Hal ini menunjukkan bahwa model yang dihasilkan termasuk kategori klasifikasi cukup baik karena memiliki nilai AUC antara 0.8-0.9.

Kata kunci: Confusion Matrix, kurva ROC, Neural Network Backpropagation

1. PENDAHULUAN

Saat ini dalam dunia kedokteran, penentuan penyakit peradangan hati menjadi hal yang tidak mudah dilakukan. Tetapi terdapat catatan rekam medis yang telah menyimpan gejala-gejala penyakit pasien dan diagnosis penyakit peradangan hati. Hal semacam ini tentu sangat berguna bagi para ahli kesehatan. Mereka dapat menggunakan catatan rekam medis yang telah ada sebagai bantuan untuk mengambil keputusan tentang diagnosis penyakit pasien.

Hepatitis atau penyakit peradangan hati merupakan salah satu dari banyaknya jenis penyakit hati, yang lainnya seperti pembengkakan hati (*fatty liver*) dan kanker hati (*cirrhosis*). Di Indonesia, pada tahun 2007 penyakit hati merupakan salah satu dari sepuluh besar penyakit penyebab kematian terbesar di Indonesia (Departemen Kesehatan RI, 2009).

Penyakit hati merupakan salah satu penyakit yang menjadi masalah nasional di semua negara, baik di negara - negara berkembang seperti Indonesia maupun di negara - negara maju. Penyakit ini dapat terjadi pada semua golongan umur, mulai dari anak-anak, remaja, dewasa sampai orang tua. Data statistik menunjukkan bahwa penyakit hati adalah salah satu penyebab kematian yang utama, baik di Amerika Serikat maupun di seluruh dunia. Penderita penyakit hati sulit untuk dideteksi, terutama pada tahap awal penyakit. Hal ini dikarenakan pasien tidak merasakan gejala penyakit dan seakan - akan hati berfungsi secara normal, padahal sebagian hati sudah mengalami kerusakan (Hannan et al, 2010).

Teknik analisa konvensional secara manual yang selama ini digunakan tidak lagi efektif digunakan untuk mendiagnosa. Seiring dengan

perkembangan sistem berbasis pengetahuan medis, tuntutan akan adanya penggunaan sistem pengetahuan berbasis komputer sebagai teknik analisa dalam mendiagnosa penyakit menjadi semakin penting. oleh karenanya, saat inilah waktu yang tepat untuk mengembangkan sistem pengetahuan berbasis komputer yang modern, efektif dan efisien dalam mendiagnosa penyakit (Neshat, and Yaghoobi, 2009).

Untuk memecahkan masalah tersebut, pada penelitian ini akan digunakan metode klasifikasi *neural network backpropagation* untuk penentuan penyakit peradangan hati. Data yang digunakan dalam penelitian ini bersumber dari alamat web: <http://archive.ics.uci.edu/ml/>. Data yang diteliti merupakan hasil pemeriksaan terhadap 583 orang dari wilayah Andhra Pradesh, India.

Kelebihan metode neural network diantaranya melakukan generalisasi dan ekstraksi dari pola data tertentu, mampu mengakuisisi pengetahuan walau tidak ada kepastian, menyelesaikan masalah yang tidak terstruktur dan sulit didefinisikan, dan mampu melakukan perhitungan secara paralel sehingga proses lebih singkat.

2. KAJIAN LITERATUR

A. Data Mining

Data mining adalah rangkaian proses untuk menggali nilai tambah berupa informasi yang belum terekplorasi dari sebuah basis data, melakukan eksplorasi dengan cara-cara tertentu untuk memanipulasi data menjadi informasi yang lebih berharga dengan cara mengekstraksi dan mengenali pola penting dari basis data (Han & Kamber, 2006). Menurut Daryl Pregibons dalam (Gorunescu, 2011) "*Data mining* adalah perpaduan dari ilmu statistik, kecerdasan buatan, dan penelitian bidang *database*". Nama *data mining* berasal dari kemiripan antara pencarian informasi yang bernilai dari *database* yang besar dengan menambang sebuah gunung untuk sesuatu yang bernilai (Sumathi, 2006). Keduanya memerlukan penyaringan melalui sejumlah besar material, atau menyelidiki dengan cerdas untuk mencari keberadaan sesuatu yang disebut bernilai tadi.

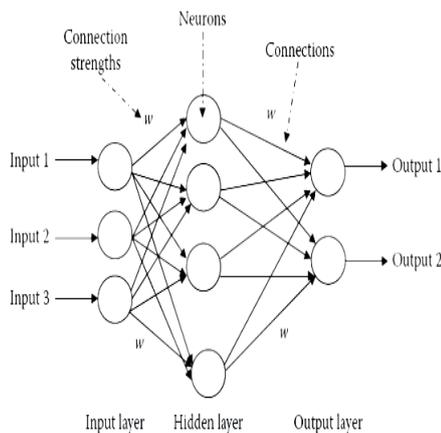
Data Mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Beberapa aplikasi data mining fokus pada prediksi, mereka meramalkan apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi di masa lalu (Witten, Frank, & Hall, 2011).

B. Neural Network

Neural network atau jaringan syaraf tiruan adalah merupakan salah satu representasi buatan dan otak manusia yang selalu mencoba mensimulasikan proses pembelajaran pada otak manusia. Istilah buatan disini digunakan karena jaringan syaraf ini diimplementasikan dengan menggunakan program komputer yang mampu menyelesaikan sejumlah proses perhitungan selama proses pembelajaran (Kusuma Dewi, 2010). *Neural network* adalah (Han, 2006) satu set unit *input/output* yang terhubung dimana tiap relasinya memiliki bobot. Hal yang perlu mendapat perhatian istimewa adalah bahwa jaringan syaraf tiruan tidak diprogram untuk menghasilkan keluaran tertentu. Semua keluaran atau kesimpulan yang ditarik oleh jaringan didasarkan pada pengalamannya selama mengikuti proses pembelajaran. Pada proses pembelajaran, kedalam jaringan syaraf tiruan dimasukkan pola-pola input (dan output) lalu jaringan akan diajari untuk memberikan jawaban yang bisa diterima (Diyah Puspitaningrum, 2006).

Neural Network dimaksudkan untuk mensimulasikan perilaku sistem biologi susunan syaraf manusia, yang terdiri dari sejumlah besar unit pemroses yang disebut *neuron*, yang beroperasi secara paralel (Alpaydin, 2010). *Neuron* mempunyai relasi dengan *synapse* yang mengelilingi *neuron-neuron* lainnya. Susunan syaraf tersebut dipresentasikan dalam *neural network* berupa graf yang terdiri dari simpul (*neuron*) yang dihubungkan dengan busur, yang berkorespondensi dengan *synapse*. Sejak tahun 1950-an, *neural network* telah digunakan untuk tujuan prediksi, bukan hanya klasifikasi tapi juga untuk regresi dengan atribut target kontinu (Vecellis, 2009).

Neural network terdiri dari dua lapisan atau lebih, meskipun sebagian besar jaringan terdiri dari tiga lapisan : lapisan input, lapisan tersembunyi, dan lapisan output (Larose, 2005). Pendekatan *neural network* dimotivasi oleh jaringan saraf biologis. Secara kasar, *neural network* adalah satu set terhubung input/output unit, di mana masing - masing sambungan memiliki berat yang terkait dengannya. *Neural network* memiliki beberapa ciri yang membuat mereka populer untuk *clustering*. Pertama, *neural network* adalah arsitektur pengolahan *inheren paralel* dan terdistribusi. Kedua, *neural network* belajar dengan menyesuaikan bobot interkoneksi dengan data, Hal ini memungkinkan *neural network* untuk "menormalkan" pola dan bertindak sebagai fitur (atribut) *extractors* untuk kelompok yang berbeda. Ketiga, *neural network* memproses vektor numerik dan membutuhkan pola objek untuk diwakili oleh fitur kuantitatif saja (Gorunescu, 2011).



Gambar 1: Arsitektur Neural Network
Sumber: Shukla et al. (2010)

C. Algoritma Neural Network Backpropagation

Algoritma pelatihan *backpropagation* atau ada yang menterjemahkan menjadi propagasi balik pertama kali dirumuskan oleh Paul Werbos pada tahun 1974 dan dipopulerkan oleh Rumelhart bersama McClelland untuk dipakai pada *neural network*. Metode *backpropagation* pada awalnya dirancang untuk *neural network feedforward*, tetapi pada perkembangannya, metode ini diadaptasi

untuk pembelajaran pada model *neural network* lainnya (Astuti, 2009).

Backpropagation merupakan salah satu algoritma pembelajaran dalam jaringan syaraf tiruan. Proses pembelajaran dalam *backpropagation* dilakukan dengan penyesuaian bobot-bobot jaringan syaraf tiruan dengan arah mundur berdasarkan nilai *error* dalam proses pembelajaran (Kusrini, 2009). Penemuan algoritma *backpropagation* untuk *multilayer perceptron (MLP)*, merupakan metode yang sistematis untuk *training* sehingga bisa dilakukan dan lebih efisien. Algoritma *backpropagation* berasal dari *learning rule* Widrow dan Hoff, disusun oleh Werbos (1974), dibuat oleh Parker (1985), Rumelhart Hinton, Williams (Rumelhart dan Williams, 1986) dan peneliti lainnya (Maimon, 2005).

Algoritma pelatihan *backpropagation* terdiri dari dua tahapan yaitu *feedforward* dan *backpropagation* dari galatnya. Langkah pembelajaran dalam algoritma *backpropagation* adalah sebagai berikut (Kusumadewi, 2010):

- Inialisasi bobot (ambil bobot awal dengan nilai random yang cukup kecil).
- Kerjakan langkah-langkah berikut selama kondisi berhenti bernilai FALSE:
 1. Untuk tiap-tiap pasangan elemen yang akan dilakukan pembelajaran, kerjakan:

Feedforward:

- a. Tiap-tiap unit input ($X_i, i=1,2,3,\dots,n$) menerima sinyal x_i dan meneruskan sinyal tersebut ke semua unit pada lapisan yang ada di atasnya (lapisan tersembunyi).
- b. Tiap-tiap unit tersembunyi ($Z_j, j=1,2,3,\dots,p$) menjumlahkan sinyal-sinyal input terbobot:

$$z_in_j = v_0 + \sum x_i v_{ij}$$

Gunakan fungsi aktivitas untuk menghitung sinyal outputnya:

$$z_j = f(z_in_j)$$

Kirimkan sinyal tersebut ke semua unit dilapisan atasnya (unit-unit output)

- c. Tiap-tiap unit output ($Y_k, k=1,2,\dots,m$) menjumlahkan sinyal-sinyal input terbobot

$$y_in_k = w_0_k + \sum z_j w_{jk}$$

Gunakan fungsi aktivitas untuk menghitung sinyal output:

$$y_k = f(y_{in_k})$$

Dan kirimkan sinyal tersebut ke semua unit dilapisan atasnya (unit-unit output).

Backpropagation

- d. Tiap-tiap unit output (Y_k , $k= 1,2, \dots, m$) menerima target pola yang berhubungan dengan pola input pembelajaran, hitung informasi errornya:

$$\delta_k = (t_k - y_k) f'(y - in_k)$$

Kemudian hitung koreksi bobot (yang nanti akan digunakan untuk memperbaiki nilai w_{jk}):

$$\Delta w_{jk} = \alpha \delta_k z_j$$

Hitung juga koreksi bias (yang nantinya akan digunakan untuk memperbaiki nilai w_{0k}):

Kirimkan δ_k ini ke unit-unit yang ada dilapisan bawahnya.

- e. Tiap-tiap unit tersembunyi ($Z_j, j=1,2,3, \dots, p$) menjumlahkan delta inputnya (dari unit-unit yang berada pada lapisan di atasnya):

$$\delta_{in_j} = \delta_k w_{kj}$$

Kalikan nilai ini dengan turunan dari fungsi aktivitas untuk menghitung informasi error:

$$\delta_j = \delta_{in_j} f'(z_{in_j})$$

Kemudian hitung koreksi bobot untuk memperbaiki nilai v_{ij} :

$$\Delta v_{ij} = \alpha \delta_j x_i$$

Hitung juga koreksi bias untuk memperbaiki nilai v_{0j} :

$$\Delta v_{0j} = \alpha \delta_j$$

- f. Tiap-tiap unit output (Y_k , $k = 1,2, \dots, m$) memperbaiki bias dan bobotnya ($j = 0,1,2, \dots, p$)

$$w_{jk} \text{ (baru)} = w_{jk} \text{ (lama)} + \Delta w_{jk}$$

Tiap-tiap unit tersembunyi (Z_j , $j = 1,2, \dots, p$) memperbaiki bias dan bobotnya ($i=0,1,2, \dots, n$).

$$v_{ij} \text{ (baru)} = v_{ij} \text{ (lama)} + \Delta v_{ij}$$

2. Tes kondisi berhenti

3. METODE PENELITIAN

A. Data Primer

Data primer adalah data yang dikumpulkan langsung oleh peneliti dari responden, dan bukan berasal dari pengumpulan data yang pernah dilakukan sebelumnya. Data primer adalah data yang diperoleh dari sumber-sumber asli. Sumber asli

disini diartikan sebagai sumber pertama darimana data tersebut diperoleh.

B. Data Sekunder

Data sekunder adalah data yang telah diolah dari data primer. Pada penelitian ini penulis menggunakan data sekunder yang bersumber dari alamat web: <http://archive.ics.uci.edu/ml/>. Data yang diteliti merupakan hasil pemeriksaan terhadap 583 orang dari wilayah Andhra Pradesh. Jenis data sekunder yang lain pada penelitian ini adalah semua pustaka dan laporan penelitian yang telah dilakukan baik dalam bentuk jurnal maupun kumpulan seminar nasional.

C. Evaluasi dan Validasi Model

Untuk mengukur akurasi model maka dilakukan evaluasi dan validasi menggunakan teknik:

1. *Confusion matrix*

Confusion Matrix adalah alat (*tools*) visualisasi yang biasa digunakan pada supervised learning. Tiap kolom pada matriks adalah contoh kelas prediksi, sedangkan tiap baris mewakili kejadian di kelas yang sebenarnya (Gorunescu, 2011). *Confusion matrix* berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada sistem klasifikasi.

2. Kurva ROC (*Receiver Operating Characteristic*)

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horisontal dan *true positives* sebagai garis vertikal (Vecellis, 2009). *The area under curve* (AUC) dihitung untuk mengukur perbedaan performansi metode yang digunakan. AUC dihitung menggunakan rumus: (Liao, 2007)

$$\theta^r = \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m \psi(x_i^r, x_j^r)$$

Dimana

$$\psi(X,Y) = \begin{cases} 1 & Y < X \\ \frac{1}{2} & Y = X \\ 0 & Y > X \end{cases}$$

Performance keakurasian AUC dapat diklasifikasikan menjadi lima kelompok yaitu (Gorunescu, 2011):
 0.90 – 1.00 = *Exellent Clasification*
 0.80 – 0.90 = *Good Clasification*
 0.70 – 0.80 = *Fair Clasification*
 0.60 – 0.70 = *Poor Clasification*
 0.50 – 0.60 = *Failure*

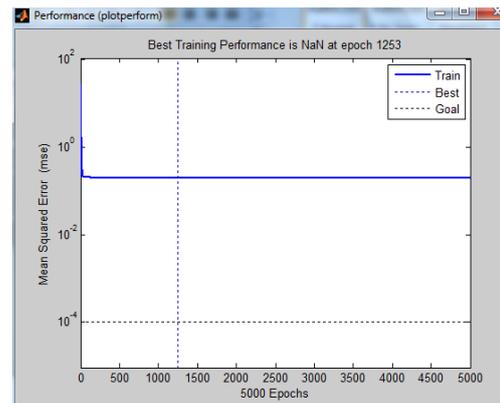
4. PEMBAHASAN

Data yang digunakan dalam penelitian ini bersumber dari alamat web: <http://archive.ics.uci.edu/ml/>. Data merupakan hasil pemeriksaan terhadap 583 orang dari wilayah Andhra Pradesh, India yang diperiksa dengan hasil 416 orang terdeteksi menderita penyakit hati dan 167 orang tidak terdeteksi menderita penyakit hati. Sumber data terdiri dari 441 orang berjenis kelamin laki-laki dan 142 orang berjenis kelamin perempuan. Dari data tersebut, 80% data (466 data) digunakan sebagai data training, dan 20% data (117 data) sebagai data testing. Variabel input pada penelitian ini terdiri dari sepuluh variabel, yaitu: 1. Age, 2. Gender, 3. TB, 4. DB, 5. Alkphos, 6. SGPT, 7. SGOT, 8. TP, 9. ALB, 10. AGRatio, Sedangkan variabel output adalah variabel Keputusan penyakit. Perangkat lunak yang digunakan untuk menganalisa adalah *matlab versi 7.7* dan *SPSS versi 17*.

Proses pelatihan jaringan syaraf tiruan menggunakan 80% dari total data, pada proses ini akan dilakukan pelatihan dengan arsitektur neural network yang berbeda-beda. Setiap arsitektur neural network akan menghasilkan bobot pelatihan yang terakhir yang akan digunakan sebagai bobot awal saat melakukan pengujian.

Hasil pelatihan oleh matlab versi 7.7 untuk arsitektur jaringan 10-15-1 diperlihatkan oleh gambar 2.

Mencari network terbaik dilakukan dengan cara mengubah jumlah neuron/node pada lapisan tersembunyi secara trial and error, maka di dapatkan konfigurasi terbaik. Berikut tabel Mean Square Error(MSE) dari beberapa variasi jaringan yang dilatih menggunakan matlab versi 7.7 pada data training dengan maximal epoch(iterasi) = 5000, learning rate (α) = 1, momentum = 0.8, dan target error = 0.0001:



Gambar 2. Mean Sequare Error dengan arsitektur Jaringan 10-15-1
 Sumber: Hasil Percobaan Menggunakan Matlab 7.7

Tabel 1. Hasil Pelatihan Arsitektur Jaringan

Arsitektur Jaringan	MSE
10-2-1	0,206
10-3-1	0,205
10-4-1	0,204
10-5-1	0,202
10-10-1	0,196
10-15-1	0,194
10-20-1	0,200
10-25-1	0,206
10-30-1	0,213

Sumber: Hasil Percobaan Menggunakan Matlab 7.7

Berdasarkan Tabel 1. maka didapatkan konfigurasi jaringan terbaik sebagai berikut:

1. Jumlah neuron lapisan input = 10
2. Jumlah neuron lapisan tersembunyi = 15
3. Jumlah neuron lapisan output = 1
 Dengan nilai MSE sebesar 0,194.

Arsitektur jaringan terbaik yang didapat saat pelatihan di atas akan digunakan sebagai arsitektur jaringan untuk mencari nilai keputusan penyakit hati pada data testing. Bobot awal pada data testing adalah bobot terakhir saat pelatihan dari arsitektur jaringan terbaik.

Untuk mengukur ketepatan dan keakuratan model dilakukan pengujian dengan *confusion matrix* dan kurva ROC sebagai berikut:

1. Confusion Matrix

Berikut nilai dari confusion matrix model neural network backpropagation:

Tabel 2. Confusion Matrix

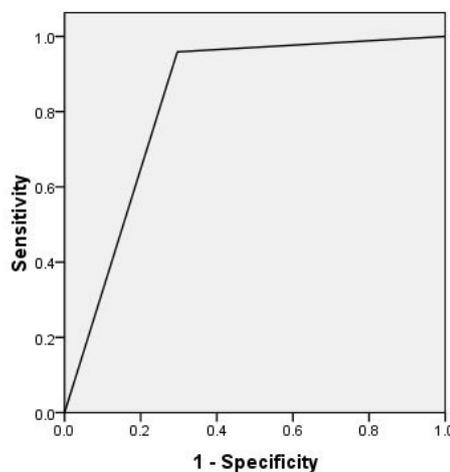
Accuracy : 84,62%			
	True Y	True N	Class Precission
Pred. Y	76	12	86,36%
Pred. N	6	23	79,31%
Class Recall	92,68%	65,71%	

Sumber: Hasil Pengolahan Menggunakan spss 17.0

Dari tabel 2. terlihat bahwa tingkat akurasi model sebesar 89%.

2. Kurva ROC

Berikut adalah grafik kurva ROC dengan menggunakan SPSS 17.



Gambar 3. Kurva ROC

Sumber: Hasil Pengolahan Menggunakan SPSS 17.0

Nilai AUC (Area Under the Curva) sebesar 0,801. Seperti pada Tabel 3

Tabel 3. Area Under the Curva (AUC)

Area under curva	Std. Error	Asymptotic 95% confidence interval	
		Lower bound	Upper bound
0,801	0,055	0,712	0,918

Sumber: Hasil Pengolahan Menggunakan SPSS 17.0

Model yang dihasilkan termasuk klasifikasi cukup baik karena memiliki nilai AUC antara 0.8-0.9, yaitu sebesar 0.801.

5. PENUTUP

Kesimpulan yang dapat diambil berdasarkan penelitian ini adalah bahwa performa model *neural network backpropagation* untuk penentuan penyakit peradangan hati yang dibentuk dari data training dan divalidasi pada data testing memberikan tingkat akurasi kebenaran sebesar 84,62% dengan nilai area under the curva (AUC) sebesar 0,801. Hal ini menunjukkan bahwa model yang dihasilkan termasuk katagori klasifikasi cukup baik karena memiliki nilai AUC antara 0.8-0.9.

DAFTAR PUSTAKA

- [1] Alpaydin, Ethem. (2010). *Introduction to Machine Learning*. London: The MIT Press.
- [2] Gorunescu, Florin (2011). *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer
- [3] Han, J., & Kamber, M. (2006). *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman.
- [4] Kusri, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Publishing.
- [5] Kusumadewi, Sri (2010). *Pengantar Jaringan Syaraf Tiruan*. Yogyakarta. Teknik Informatika FT Ull.
- [6] Larose, D. T. (2005). *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc.
- [7] Maimon, Oded & Rokach, Lior. (2005). *Data Mining and Knowledge Discovey Handbook*. New York: Springer
- [8] Mehdi Neshat, and Mehdi Yaghoobi. (2009, October, 20-22). "Designing a Fuzzy Expert System of Diagnosing the Hepatitis B Intensity Rate and Comparing it with Adaptive Neural Network Fuzzy System". Proceeding of the world congress on engineering and computer science 2009, Vol II, WCECS 2009, ISBN:978-988-18210-2-7. pp 1-6, October 20-22.
- [9] Puspitaningrum, Diyah (2006). *Pengantar Jaringan Syaraf Tiruan*. Yogyakarta: Andi Offset.
- [10] Shaikh Abdul Hannan, R. R. Manza, and R. J. Ramteke. (2010).

- “Generalized Regression Neural Network and Radial Basis Function for Heart Disease diagnosis”. *International Journal of Computer Application* (0975-8887). Vol. 7, No.13, pp. 7-13.
- [11] Shukla, A., Tiwari, R., & Kala, R. (2010). *Real Life Applications of Soft Computing*. United States of America on: Taylor and Francis Group, LLC.
- [12] Sumathi, & S., Sivanandam, S.N. (2006). *Introduction to Data Mining and its Applications*. Berlin Heidelberg New York: Springer
- [13] Vercellis, Carlo (2009). *Business Intelligent: Data Mining and Optimization for Decision Making*. Southern Gate, Chichester, West Sussex: John Willey & Sons, Ltd.
- [14] Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning and Tools*. Burlington: Morgan Kaufmann Publisher.