

Pengaruh Normalisasi Data pada Jaringan Syaraf Tiruan Backpropagasi Gradient Descent Adaptive Gain (BPGDAG) untuk Klasifikasi

Nurul Chamidah

Jurusan Teknik Informatika
Fakultas MIPA UNS

Jl. Ir. Sutami No. 36 A Surakarta
nurulchamidah.2007@gmail.com

Wiharto

Jurusan Teknik Informatika
Fakultas MIPA UNS

Jl. Ir. Sutami No. 36 A Surakarta
wi_harto@yahoo.com

Umi Salamah

Jurusan Teknik Informatika
Fakultas MIPA UNS

Jl. Ir. Sutami No. 36 A Surakarta
umi_salamah@yahoo.com

ABSTRAK

Algoritma pelatihan pada multilayer perceptron menggunakan backpropagasi, biasanya fungsi aktivasi yang digunakan adalah fungsi sigmoid. Fungsi tersebut akan membawa nilai input dengan range yang tak terbatas ke nilai output yang terbatas, yaitu dalam sebuah range 0 sampai 1. Supaya dapat membawa range nilai output ke dalam range input, maka data input harus dilakukan normalisasi data ke dalam range 0 sampai 1, sehingga outputnya dapat di denormalisasi ke dalam range nilai input. Metode normalisasi ada bermacam-macam, masalahnya, metode mana yang paling efektif untuk kasus klasifikasi menggunakan algoritma pelatihan BPGD/AG. Penelitian ini akan menganalisa metode normalisasi mana yang paling efektif untuk mengklasifikasi data pada algoritma BPGD/AG. Eksperimen dilakukan pada data kanker payudara yang diklasifikasikan ke dalam kanker ganas dan kanker jinak. Hasil evaluasi menunjukkan bahwa metode normalisasi *Min-Max* memberikan akurasi tertinggi hingga 96.86% dengan rata-rata iterasi sebanyak 21 epoch.

Kata Kunci

Backpropagasi, Normalisasi, Denormalisasi, Min-max, Sigmoid.

1. PENDAHULUAN

Syaraf Tiruan (JST) merupakan salah satu metode yang dapat digunakan untuk mengklasifikasi data, dimana klasifikasi diperlukan untuk memprediksi kelas dari suatu data. JST efektif digunakan dalam klasifikasi data yang jumlah variabelnya cukup banyak dan memiliki korelasi variabel yang kompleks. Kelebihan JST salah satunya adalah kemampuannya dalam beradaptasi dan mampu belajar dari data masukan yang diberikan sehingga dapat memetakan hubungan antara masukan dan keluarannya. Selain itu JST mampu memprediksi keluaran berdasarkan masukan yang telah dilatihkan sebelumnya.

JST mempunyai banyak struktur jaringan, diantaranya multilayer perceptron. Untuk melakukan pelatihan pada struktur ini menggunakan algoritma backpropagasi. Algoritma pelatihan pada multilayer perceptron menggunakan backpropagasi, biasanya fungsi aktivasi yang digunakan adalah fungsi sigmoid. Fungsi tersebut akan membawa nilai input dengan range yang tak terbatas ke nilai output yang terbatas, yaitu dalam sebuah range 0 sampai 1 [1]. Supaya dapat membawa range nilai output ke dalam range input, maka data input harus dilakukan normalisasi data ke dalam range 0 sampai 1, sehingga outputnya dapat di denormalisasi ke dalam range nilai input. Metode normalisasi ada bermacam-macam, masalahnya, metode mana yang paling efektif untuk kasus klasifikasi.

Beberapa penelitian yang berkaitan dengan metode normalisasi sudah pernah dilakukan, diantaranya penelitian Jayalakshmi [2]. Penelitian tersebut membahas tentang pengaruh metode

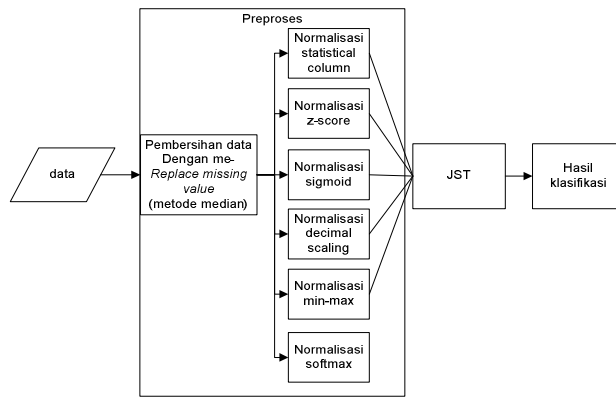
normalisasi pada algoritma backpropagasi dengan kasus diabetes mellitus. Algoritma pelatihan yang digunakan adalah algoritma backpropagasi momentum. Algoritma tersebut merupakan perbaikan dari algoritma gradient descent. Metode normalisasi yang digunakan belum mengakomodasi dua buah metode normalisasi yaitu decimal scaling dan softmax. Selain itu dalam penelitian tersebut mengabaikan missing value. Berdasarkan penelitian yang dilakukan Paulin dan Santhakumaran [3] pembuangan instance data yang mengandung missing value akan mempengaruhi akurasi dalam diagnosa.

Penelitian berkaitan dengan perbaikan algoritma gradient descent untuk pelatihan multi-layer perceptron telah banyak dilakukan. Perbaikan tersebut bertujuan untuk meningkatkan performa dari JST, baik dalam hal akurasi maupun kecepatan konvergensi. Selain dengan momentum, perbaikan algoritma dengan menambahkan parameter Gain yang nilainya adaptive, sehingga menjadi Backpropagasi *Gradient Descent* dengan *Adaptive Gain (BPGD/AG)* [7]. Algoritma BPGD/AG terbukti mampu meningkatkan efektifitas prediksi dan kecepatan training untuk prediksi dari data penyakit jantung [9].

Berangkat dari penelitian-penelitian tersebut, peneliti akan menganalisis efektifitas klasifikasi data dengan metode normalisasi softmax, min-max, z-score, sigmoid, decimal scaling dan statistical column. Data yang akan di klasifikasi sebelum dinormalisasi dilakukan pembuangan instance data yang mengandung missing value dengan metode median. Algoritma pembelajaran yang digunakan untuk pelatihan multilayer perceptron dipilih algoritma Backpropagasi *Gradient Descent* dengan *Adaptive Gain (BPGD/AG)* untuk mendapatkan hasil klasifikasi yang optimal pada kasus kanker payudara.

2. METODE PENELITIAN

Langkah kerja eksperimen dalam penelitian dilakukan melalui beberapa tahapan. Tahapan-tahapan tersebut ditunjukkan pada gambar berikut



Gambar 1 Langkah-langkah penelitian

1. Meload data
2. Mereplace missing value dengan metode median
3. Menormalisasi data dengan memvariasikan metode normalisasi agar range nilainya antara nol dan satu. Data ternormalisasi ini nantinya akan digunakan sebagai input dalam JST.
4. Membagi data menjadi dua bagian, 80% untuk training dan 20% untuk test.
5. Membuat arsitektur neural network dengan tiga layer yakni input layer, hidden layer, dan output layer. Dimana jumlah node input sama dengan jumlah input dari kasus, jumlah node hidden neuron pada hidden layer sebanyak enam neuron, dan satu node output.
6. Untuk setiap data ternormalisasi dengan metode yang berbeda, dilakukan langkah vii - ix
7. Melatih jaringan dengan data training dengan epoch tertentu.
8. Melakukan tes dengan data tes
9. Mengevaluasi keakuratan klasifikasi untuk data ternormalisasi.
10. Selesai

2.1 Data set

Klasifikasi ini diaplikasikan pada dataset kanker payudara yang diperoleh dari *UCI Machine Learning Repository* [14] sumbangan Rumah Sakit *University of Wisconsin Madison* dari Dr. William H. Wolberg [14] di <http://archive.ics.uci.edu/ml/>. Jumlah data 699 dengan 16 nilai atribut yang hilang, 458 (65,5%) tumor jinak dan 241 (34,5%) tumor ganas dengan deskripsi atribut sebagai berikut.

Tabel 1 Atribut data kanker

No.	Atribut	Domain
1.	<i>Sample code number</i>	id number
2.	<i>Clump Thickness</i>	1 – 10
3.	<i>Uniformity of Cell Size</i>	1 – 10
4.	<i>Uniformity of Cell Shape</i>	1 – 10
5.	<i>Marginal Adhesion</i>	1 – 10
6.	<i>Single Epithelial Cell Size</i>	1 – 10
7.	<i>Bare Nuclei</i>	1 – 10
8.	<i>Bland Chromatin</i>	1 – 10
9.	<i>Normal Nucleoli</i>	1 – 10
10.	<i>Mitoses</i>	1 – 10
11.	<i>Class</i>	untuk kanker jinak, untuk kanker ganas

2.2 Preproses Data

Preproses data dilakukan dengan mereplace missing value dan menormalisasi data ke dalam rang 0 dan 1 dengan menggunakan beberapa metode. Berikut metode preproses yang digunakan dalam penelitian ini ;

a. *Mereplace Missing value*

Data yang hilang (*missing value*) di-replace dengan metode median. Yakni dengan mencari nilai tengah dan me-replace semua nilai yang hilang dengan nilai tengah tersebut. Metode ini merupakan metode *replacement* yang paling baik untuk kasus *Wisconsin breast Cancer Diagnosis* dengan Backpropagsi [3].

Nilai tengah untuk data ganjil :

$$Me = \frac{X_{\frac{n+1}{2}}}{2} \tag{1}$$

Nilai tengah untuk data genap :

$$Me = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} \tag{2}$$

b. Normalisasi

Metode normalisasi yang digunakan dalam penelitian ini adalah dengan mentransformasi data ke dalam *range* 0 dan 1. Metode - metode yang digunakan antara lain :

- *Decimal scaling*: digunakan ketika nilai berada dalam rentang logaritmik, misalnya rentang suatu nilai adalah antara 0 dan 1 sedangkan nilai lain pada rentang 0 dan 1000 [7].

$$s' = \frac{s}{10^n} \tag{3}$$

Dimana $n = \log_{10} \max\{s_k\}$

- *Sigmoid* : metode ini mengubah nilai kedalam *range* antara 0 dan 1 dengan fungsi sigmoid [1].

$$s' = \frac{1}{1 + e^{-s}} \tag{4}$$

- *Softmax* : metode ini mengubah nilai kedalam *range* antara 0 dan 1 menggunakan fungsi sigmoid dengan memanfaatkan mean dan standar deviasi [11].

$$s' = \frac{1}{1 + e^{-\frac{s-\mu}{\sigma}}} \tag{5}$$

Dimana μ adalah rata-rata dari tiap kolom dan σ merupakan standar deviasi.

- Min-max normalization: metode ini me-rescale data dari suatu range ke range baru lain. Data di skalakan dalam range 0 dan 1. Diberikan nilai yang bersesuaian (dalam satu kolom) $\{s_k\}$, $k=1,2,... n$. Maka nilai normalisasinya adalah [1] :

$$s' = \frac{s - \min\{s_k\}}{\max\{s_k\} - \min\{s_k\}} \tag{6}$$

- Statistical column : menormalisasi tiap sampel dengan sebuah nilai kolom ternormalisasi yakni dengan transformasi sqrt [13], yakni dengan menghitung normalisasi dari tiap kolom dengan mengakar kuadratkan kemudian setiap sampel dibagi dengan nilai normalisasi kolom tersebut dan dikalikan dengan nilai bias kecil [1].

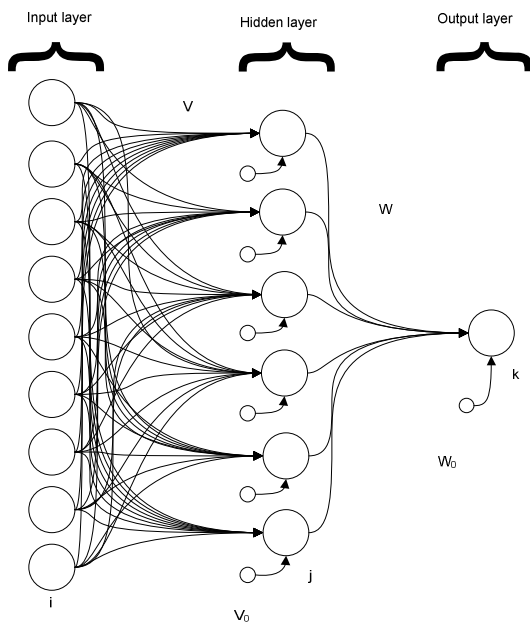
$$s' = \frac{s(i) - n(c_a)}{n(c_a)} \times 0.1 \tag{7}$$

- Z-score: teknik ini menggunakan rata-rata dan standar deviasi untuk menormalisasi tiap input [1].

$$s' = \frac{s - \mu}{\sigma} \tag{8}$$

2.3 Struktur Jaringan Syaraf Tiruan

Konsep algoritma backpropagasi adalah menyesuaikan bobot jaringan dengan mempropagasi error dari output ke input. Selama pelatihan, jaringan meminimisasi error dengan mengestimasi bobot dan berhenti pada Minimum Squared Error (MSE) 0.05 atau maksimal iterasi sebanyak 1000 epoch. Prosedur minimasi dilakukan dengan Backpropagasi Gradient Descent dengan Adaptive Gain (BPGD/AG) dan fungsi aktivasi sigmoid. Berikut struktur jaringan syaraf tiruan multilayer perceptron dengan satu hidden layer



Gambar 2 Struktur Jaringan syaraf Tiruan Multilayer perceptron

Arsitektur yang digunakan dalam BPGD/AG ini adalah satu input layer, satu hidden layer, dan satu output layer dengan formasi neuron 9-6-1. fungsi aktivasi yang dipakai adalah fungsi sigmoid dengan learning rate 0.01. Pada input layer, yang menjadi input neuron adalah input variabel data kasus dengan sembilan variabel yakni, Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, dan Mitoses. Pada hidden layer terdapat enam hidden neuron [10]. Pada Output layer terdapat satu hidden neuron untuk output, yakni hasil klasifikasi apakah kanker tersebut jinak atau ganas.

Algoritma Feed forward Backpropagation Gradient Descent dengan Adaptive Gain yang digunakan untuk pelatihan adalah sebagai berikut :

- Tahapan 0 : Pemberian inialisasi bobot, bias dan gain dengan nilai tertentu.
- Tahapan 1 : Menjalankan semua tahapan dari tahapan ke 2 sampai dengan tahapan ke 15 hingga kondisi iterasi terpenuhi.
- Tahapan 2 : Setiap data pelatihan melakukan langkah 3 sampai dengan langkah 14.

Feedforward

- Tahapan 3 : Untuk setiap unit masukan ($X_i = 1, \dots n$) menerima sinyal masukan dan diberi notasi X_i kemudian dikirimkan ke unit berikutnya yaitu unit hidden layer.
- Tahapan 4 : Pada masing-masing hidden neuron, input X_i dikalikan dengan penimbang dan dijumlahkan serta di tambah dengan biasnya.

$$Z_{in_j} = \left(\sum_{i=1}^n X_i v_{ij} \right) + v_{0j} \tag{9}$$

Kemudian melakukan fungsi pengaktifan menggunakan:

$$Z_j = f(Z_{in_j}) \tag{10}$$

Dalam bentuk Sigmoid fungsi pengaktifannya menjadi:

$$Z_j = \frac{1}{1 + e^{-c_{z_j} Z_{in_j}}} \tag{11}$$

Setelah melewati tahap pengaktifan maka keluaranya dikirim ke semua unit keluaran.

- Tahapan 5 : Untuk setiap unit keluaran ($Y_k, k=1,2,\dots,m$), unit Z_j dikalikan dengan penimbang dan dijumlahkan serta ditambahkan dengan biasnya:

$$Y_{in_k} = \left(\sum_{j=1}^p W_{jk} Z_j \right) + W_{0k} \tag{12}$$

Kemudian dilakukan fungsi pengaktifan:

$$Y_k = f(Y_{in_k}) \tag{13}$$

Dalam fungsi Sigmoid menjadi :

$$Y_k = \frac{1}{1 + e^{-c_{-y_k} Y_{in_k}}} \quad (14)$$

Backpropagasi dan perubahan bobot

Tahapan 6 : Pada setiap unit keluaran berupa Y_k mendapatkan pola target seperti pola masukan saat pelatihan dan dihitung *error*nya:

$$E = \sum (t_k - Y_k)^2 \quad (15)$$

Tahapan 7 : Menghitung perubahan bobot pada *link* menuju *output*

$$\Delta W_{jk} = \eta \left(-\frac{\partial E}{\partial W_{jk}} \right) \quad (16)$$

Bentuk hasil turunan parsialnya

$$\Delta W_{jk} = \eta (t_k - Y_k) Y_k (1 - Y_k) c_{-y_k} Z_j$$

Tahapan 8 : Menghitung perubahan bias pada *link* menuju *output*

$$\Delta W_{0k} = \eta \left(-\frac{\partial E}{\partial W_{0k}} \right) \quad (17)$$

Bentuk hasil turunan parsialnya

$$\Delta W_{0k} = \eta (t_k - Y_k) Y_k (1 - Y_k) c_{-y_k} \quad (18)$$

Tahapan 9 : Menghitung perubahan bobot pada *link* menuju *hidden neuron*

$$\Delta V_{ij} = \eta \left(-\frac{\partial E}{\partial V_{ij}} \right) \quad (19)$$

Bentuk hasil turunan parsialnya

$$\Delta V_{ij} = \eta \left[\sum_k c_{-y_k} W_{jk} Y_k (1 - Y_k) (t_k - Y_k) \right] * tmp$$

$$tmp = c_{-z_j} Z_j (1 - Z_j) X_i \quad (20)$$

Tahapan 10: Menghitung perubahan bias pada *link* menuju *hidden neuron*

$$\Delta V_{0j} = \eta \left(-\frac{\partial E}{\partial V_{0j}} \right) \quad (21)$$

Tahapan 11 : Mengupdate semua bobot dan bias

$$W_{jk} = W_{jk} + \Delta W_{jk} \quad (22)$$

$$W_{0k} = W_{0k} + \Delta W_{0k} \quad (23)$$

$$V_{ij} = V_{ij} + \Delta V_{ij} \quad (24)$$

$$V_{0j} = V_{0j} + \Delta V_{0j} \quad (25)$$

Tahapan 12 : Menghitung perubahan *gain* pada *output*

$$\Delta c_{-y_k} = \eta \left(-\frac{\partial E}{\partial c_{-y_k}} \right) \quad (26)$$

dimana,

$$\frac{\partial E}{\partial c_{-y_k}} = -(t_k - Y_k) Y_k (1 - Y_k) (\sum W_{jk} Y_k + W_{0k})$$

Sehingga

$$\Delta c_{-y_k} = \eta (t_k - Y_k) Y_k (1 - Y_k) (\sum W_{jk} Y_k + W_{0k})$$

Tahapan 13 : Menghitung perubahan *gain* pada *hidden neuron*

$$\Delta c_{-z_j} = \eta \left(-\frac{\partial E}{\partial c_{-z_j}} \right) \quad (27)$$

dimana,

$$\frac{\partial E}{\partial c_{-z_j}} = \left[\sum_k c_{-y_k} W_{jk} Y_k (1 - Y_k) (t_k - Y_k) \right] *$$

$$Z_j (1 - Z_j) \left(\left(\sum_i V_{ij} X_i \right) + V_{0j} \right)$$

Sehingga

$$\Delta c_{-z_j} = \eta \left[-\sum_k c_{-y_k} W_{jk} Y_k (1 - Y_k) (t_k - Y_k) \right] *$$

$$Z_j (1 - Z_j) \left(\left(\sum_i V_{ij} X_i \right) + V_{0j} \right)$$

Tahapan 14 : Mengupdate *gain*

$$c_{-y_k} = c_{-y_k} + \Delta c_{-y_k} \quad (28)$$

$$c_{-z_j} = c_{-z_j} + \Delta c_{-z_j} \quad (29)$$

Tahapan 15 : Iterasi berakhir

Kondisi berhenti

Epoch = maksimal iterasi

$$\text{Atau } MSE \leq \frac{\sum E}{\text{jumlahdata}} \quad (30)$$

3. PEMBAHASAN

Eksperimen ini dilakukan dengan dua proses, yakni proses training dan proses testing. Dari 699 record data, proses training menggunakan 80% dari data, yakni dengan merandom 80% record dari data kanker jinak dan 80% record dari data kanker ganas, sedangkan sisanya untuk proses testing yang dilakukan dengan menggunakan random 20% data, terdiri dari 20% record dari kanker jinak dan 20% dari kanker ganas.

Pada setiap satu variasi eksperimen, dilakukan seratus kali pengulangan. Setiap pengulangan pada proses training, diberikan nilai inisial bobot random dan didapatkan jumlah iterasi untuk mencapai konvergensi. Percobaan berhenti pada *MSE* atau pada maksimal iterasi dan diasumsikan telah mencapai konvergensi serta menghasilkan bobot yang akan digunakan untuk *testing*.

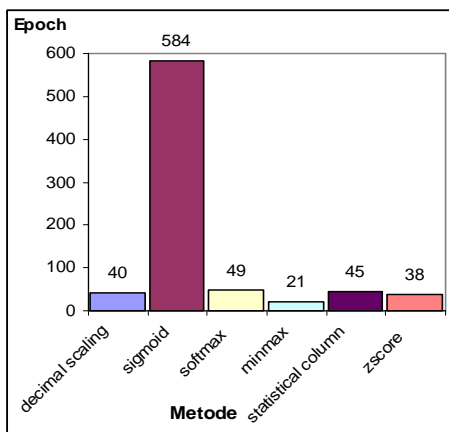
Setiap bobot hasil *training* yang telah konvergen dipakai untuk melakukan *testing*. Data klasifikasi hasil *testing* dibandingkan dengan data klasifikasi yang sebenarnya sehingga diperoleh jumlah data yang diprediksi benar dan yang diprediksi salah.

Evaluasi dari eksperimen dilakukan dengan mengambil rata-rata epoch dan akurasi dari seratus pengujian, hasil pengujian pada penelitian ini dapat dilihat pada Tabel 2.

Tabel 2 Perbandingan metode normalisasi

Metode	Kriteria rata-rata epoch	rata-rata akurasi (%)
decimal scaling	40	96.00
sigmoid	584	76.91
softmax	49	96.19
min-max	21	96.86
statistical column	45	95.91
zscore	38	95.68

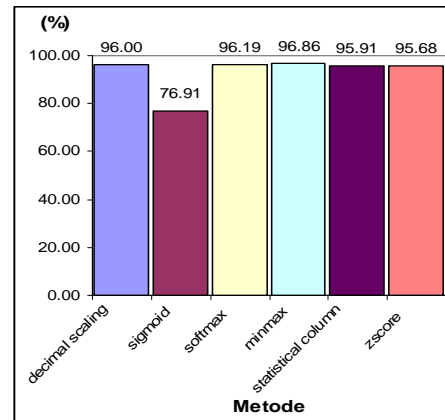
Perbandingan rata-rata epoch dari berbagai metode normalisasi ditunjukkan pada Gambar 3, dari gambar tersebut dapat dilihat bahwa minmax memiliki rata-rata epoch paling rendah yang berarti metode minmax memiliki waktu yang paling singkat untuk mencapai konvergensi.



Gambar 3 Rata-rata epoch

Sedangkan perbandingan rata-rata akurasi ditunjukkan pada Gambar 4 dimana metode minmax memiliki akurasi yang

paling tinggi untuk kasus klasifikasi kanker payudara. Hasil perbandingan menunjukkan bahwa normalisasi dengan metode minmax memberikan rata-rata akurasi tertinggi bila dibandingkan dengan metode lain serta memerlukan epoch yang paling sedikit,



Gambar 4 Rata-rata akurasi

4. PENUTUP

Penelitian ini menunjukkan pengaruh metode normalisasi data pada algoritma JST Backpropagasi *Gradient Descent* dengan *Adaptive Gain (BPGD/AG)* untuk klasifikasi yang diujikan dengan data kasus kanker payudara. Evaluasi dari eksperimen menunjukkan bahwa efektifitas JST *BPGD/AG* sangat dipengaruhi oleh metode normalisasi yang dipilih untuk mencapai akurasi terbaik, dan metode normalisasi yang memberikan efektifitas terbaik dalam hal akurasi dan kecepatan konvergensi adalah metode Minmax yang mencapai akurasi rata-rata hingga 96.86% dan memerlukan iterasi yang paling sedikit dengan rata-rata sebanyak 21 *epoch*.

5. DAFTAR PUSTAKA

- [1] Kusumadewi, Sri. Membangun Jaringan Syaraf Tiruan Menggunakan Matlab & Exel Link. 2004. Graha Ilmu. Yogyakarta. pp. 150-161
- [2] Jayalakshmi, T. and Santhakumaran, A. 2011. Statistical Normalization and Back Propagation for Classification. *International Journal of Computer Theory and Engineering*. III(1).
- [3] Paulin, F. and Santhakumaran, A. 2010. An Algorithm to Reconstruct the Missing Values for Diagnosing the Breast Cancer. *Global Journal of Computer Science and Technology*. X(2).
- [4] Bishop C.M., *Neural Networks for pattern recognition*. 1995:Oxford University Press.
- [5] Huang H.Y., A unified Approach to quadratically convergent algorithms for function minimization. *J. Optim. Theory Application*, 1970.5:p. 405-423.
- [6] M. R. Hestenes and E. Stiefel, *Methods of conjugate Gradient fr solving linear systems*. *J. Research NBS*, 1952. 49:p.409.
- [7] Nandakumar, K., Anil J., Arun R. 2005. Score Normalization in Multimodal Biometric Systems. *Pattern Recognition* 38. pp. 2270-2285
- [8] Nawi, N. M., M. R. Ransing, and R. S. Ransing. A New Improvement Of Back-Propagation Neural Network Learning Algorithms With Adaptive Gain. *Journal of Science and Technology*.

- [9] Nawi, N. M., R. Ghazali and M.N.M. Salleh. 2011. Predicting Patients with Heart Disease by Using an Improved Back-propagation Algorithm. *Journal of Computing*. III(2) : 53-58.
- [10] Paulin F. and Santhakumaran A. 2010. Back Propagation Neural Network by Comparing Hidden Neuron: Case Study on Breast Cancer Diagnosis. *International Journal of Computer Application*, II(4).
- [11] Priddy, Kevin L. and Paul E. Keller. 2005. *Artificial Neural Networks: An Introduction*, Washington : SPIE - The International Society for Optical Engineering.
- [12] R.A. Jacobs, Increased rates of convergence through learning rate adaptation. *Neural Networks*, 1988.1:p. 295-307.
- [13] Roberts, Seth. 2008. Statistics column : Transform your data. *Nutrition* 24. pp. 492-494
- [14] UCI Machine Learning Repository. [<http://archive.ics.uci.edu/ml/>]. Irvine, CA: University of California, Center for Machine Learning and Intelligent Systems
- [15] Wolberg W. H. And Mangasarian O.L, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology", in proceedings of the National Academy of Sciences, vol. 87, 9193-9196, U.S.A.,*(Dec. 1990)