

OPTIMASI JARINGAN SARAF TIRUAN UNTUK DIAGNOSIS PENYAKIT DIABETES INDIAN PIMA

Eva Yulia Puspaningrum¹⁾, Lailly Syifaul Qolby²⁾, dan Yisti Vita Via³⁾

^{1, 2, 3)}Teknik Informatika, UPN "Veteran" Jawa Timur

Jalan Raya Rungkut Madya No.1 Surabaya

e-mail: ephayulia@gmail.com¹⁾, laillyqolby19@gmail.com²⁾, yistivita@gmail.com³⁾

ABSTRAK

Untuk mengetahui kondisi seseorang menderita diabetes harus dengan melakukan beberapa tes pada laboratorium, US *National Institute of Diabetes* telah melakukan uji untuk penyakit Diabetes sesuai dengan kriteria Organisasi Kesehatan Dunia (*World Health Organization*, WHO) yang dilakukan pada sejumlah perempuan yang berusia 21 tahun, dari warisan Pima India dan tinggal di dekat Phoenix, Arizona sebanyak 768 objek. Jumlah data Diabetes Indian Pima yaitu sebanyak 768 data. Untuk percobaan ini, data tersebut dibagi menjadi dua yaitu 80% sebagai data *training* dan 20% sebagai data *testing*. Dengan menggunakan Jaringan Saraf Tiruan (JST) *Backpropagation*, data tersebut dikembangkan untuk diagnosa penyakit Diabetes. Hal ini diharapkan dapat digunakan untuk memprediksi potensi seseorang terserang Diabetes. Klasifikasi JST *Backpropagation* ini dioptimasi menggunakan metode Nguyen Widrow agar *rule* yang dihasilkan lebih signifikan atau *rule* yang dihasilkan dapat meningkatkan akurasi. Pengujian menggunakan data *testing* Diabetes dan inisialisasi Nguyen Widrow, maka dihasilkan tingkat akurasi sebesar 100%. Sedangkan jika menggunakan inisialisasi bobot *random*, maka dihasilkan tingkat akurasi sebesar 50%.

Kata kunci: *Backpropagation*, Diabetes, Jaringan Saraf Tiruan, Nguyen Widrow.

ABSTRACT

To determine the condition of a person suffering from diabetes need to do some tests in laboratories, the US *National Institute of Diabetes* has been test for Diabetes in accordance with the criteria of the *World Health Organization* conducted a number of women aged 21 years, from the legacy of Pima Indians and stay near Phoenix, Arizona as many as 768 objects. The amount of data Pima Indian Diabetes as many as 768 data. For this experiment, the data is divided into two: 80% as training data and 20% as a data testing. By using a neural network *Backpropagation*, the data developed for the diagnosis of Diabetes. It is expected-kan can be used to predict the potential of a person develops diabetes. Classification neural network *Backpropagation* is optimized using methods Nguyen Widrow that produced more significant rule or rule produced can improve accuracy. Diabetes testing using testing and initialization of data Nguyen Widrow, then the resulting accuracy rate of 100%. Whereas if you use random weight initialization, then produced a 50% accuracy rate.

Keywords: *Backpropagation*, Diabetes, Neural Network, Nguyen Widrow.

I. PENDAHULUAN

PENYAKIT Diabetes telah dikenal sejak berabad-abad lalu dan sampai sekarang masih menjadi masalah kesehatan di masyarakat. Diabetes adalah suatu kondisi yang mengakibatkan meningkatnya kadar gula di dalam darah. Selain itu, Diabetes adalah suatu kelainan reaksi kimia dalam hal pemanfaatan yang tepat atas karbohidrat, lemak, dan protein dari makanan karena tidak cukupnya pengeluaran atau kurangnya Insulin. Dengan kata lain, Diabetes terjadi ketika tubuh tidak dapat memanfaatkan beberapa makanan karena kekurangan produksi Insulin [1]. Jaringan Saraf Tiruan (JST) yang telah digunakan untuk menganalisis data kompleks dan untuk mengenali pola, dapat memberikan dukungan bagi pengambilan keputusan medis oleh dokter. Dengan melakukan analisis data penanganan pasien-pasien dengan permasalahan yang sejenis untuk mengambil keputusan, sehingga nantinya keputusan yang diambil merupakan keputusan yang terbaik untuk pasien.

Metode *Backpropagation* merupakan salah satu metode JST yang dapat diterapkan dalam memprediksi suatu keadaan yang akan datang [2]. Dalam penelitian [2] menggunakan metode *backpropagation* untuk memprediksi cuaca, dan menemukan bahwa proses pelatihan dapat dilakukan dengan cepat. Hasilnya lebih akurat untuk memprediksi cuaca di masa depan ketika jumlah iterasi meningkat.

Optimasi nilai rata-rata untuk inisialisasi bobot awal *Neural Network* (Jaringan Saraf Tiruan, JST) dengan metode Nguyen Widrow lebih baik dibandingkan inisialisasi bobot awal *Neural Network* secara acak [3]. Nilai rata-rata standar deviasi untuk masing-masing metode hanya terpaut sekitar 0,005.

Metode *Backpropagation* yang telah dijelaskan di atas dapat diterapkan dalam memprediksi suatu keadaan yang akan datang, sedangkan metode Nguyen Widrow sebagai inisialisasi bobot awal terbukti menghasilkan rata-rata akurasi lebih baik. Pada penelitian ini akan dilakukan diagnosis penyakit Diabetes Indian Pima menggunakan JST *Backpropagation* yang dioptimasi dengan metode Nguyen-Widrow yang diharapkan dapat membantu mengurangi angka kematian dan dapat dilakukan tindakan antisipasi.

II. TINJAUAN PUSTAKA

A. Diabetes Indian Pima

Penyakit Diabetes adalah golongan penyakit kronis yang ditandai dengan peningkatan kadar gula dalam darah sebagai akibat adanya gangguan sistem metabolisme dalam tubuh, di mana organ pankreas tidak mampu memproduksi hormon insulin sesuai kebutuhan tubuh.

Terdapat 9 variabel Diabetes Indian Pima yaitu *number of times pregnant (time)*, *plasma glucose concentration a 2 hours in an oral glucose tolerance test (OGTT)*, *diastolic blood pressure (D)*, *triceps skin fold thickness (T)*, *2-Hour serum insulin (IPOST)*, *body mass index (BMI)*, *diabetes pedigree function (F)*, *Age (T)*, dan *Class variable (Diagnosa)*. Database Diabetes Pima India, disumbangkan oleh Vincent Sigillito. Data Diabetes India Pima adalah kumpulan laporan diagnostik medis dari 768 contoh-contoh dari populasi yang tinggal di dekat Phoenix, Arizona, Amerika Serikat [4] [5].

B. Jaringan Saraf Tiruan (JST)

Jaringan Saraf Tiruan (JST), *Artificial Neural Network* didefinisikan sebagai suatu sistem pemrosesan informasi yang mempunyai karakteristik menyerupai jaringan saraf manusia. Jaringan saraf sebagai sebuah kelompok pengolahan elemen dalam suatu kelompok yang khusus membuat perhitungan sendiri dan memberikan hasilnya kepada kelompok kedua atau berikutnya [6]. Setiap sub-kelompok menurut gilirannya harus membuat perhitungan sendiri dan memberikan hasilnya untuk subgrup atau kelompok yang belum melakukan perhitungan. Pada akhirnya sebuah kelompok dari satu atau beberapa pengolahan elemen tersebut menghasilkan keluaran dari jaringan [7].

C. Backpropagation

Backpropagation atau propagasi balik adalah salah JST *multilayer* yang mengubah bobot dengan cara mundur dari lapisan keluaran ke lapisan masukan. Tujuannya untuk melatih agar mendapatkan keseimbangan kemampuan untuk mengenali pola yang digunakan selama pelatihan.

Langkah-langkah dalam membangun algoritma *backpropagation* adalah sebagai berikut [8]:

Langkah 0: Inisialisasi bobot-bobot, konstanta laju pelatihan (α), toleransi *error* atau nilai bobot (bila menggunakan nilai bobot sebagai kondisi berhenti) atau set maksimal *epoch* (jika menggunakan banyaknya *epoch* sebagai kondisi berhenti).

Langkah 1: Selama kondisi berhenti belum dicapai, maka lakukan langkah ke-2 hingga langkah ke-9.

Langkah 2: Untuk setiap pasangan pola pelatihan, lakukan langkah ke-3 sampai langkah ke-8.

Tahap 1: Umpan Maju (*feedforward*)

Langkah 3: Setiap unit *input* x_i (dari unit ke-1 hingga unit ke- n pada lapisan *input*) mengirimkan sinyal *input* ke setiap *input* yang berada pada lapisan tersembunyi.

Langkah 4: Masing-masing unit di lapisan tersembunyi z_j (dari unit ke-1 hingga unit ke- p) dikalikan dengan dengan bobotnya dan dijumlahkan serta ditambahkan dengan biasnya, perhatikan Persamaan 1 dan Persamaan 2.

$$Z_{net_j} = V_{j0} + \sum_{i=1}^n X_i V_{ji} \quad (1)$$

$$Z_j = f(Z_{net_j}) = \frac{1}{1 + e^{-Z_{net_j}}} \quad (2)$$

Langkah 5: Masing-masing unit *output* y_k ($k = 1, 2, 3, \dots, m$) dikalikan dengan bobot dan dijumlahkan serta ditambahkan dengan biasnya, perhatikan Persamaan 3 dan Persamaan 4.

$$y_{net_k} = w_{k0} + \sum_{j=i}^p z_j k_j \quad (3)$$

$$y_k = f(y_{net_k}) = \frac{1}{1 + e^{-y_{net_k}}} \quad (4)$$

Tahap 2: Umpan Mundur (*Backward Propagation*)

Langkah 6: Masing-masing unit *output* y_k ($k = 1, 2, 3, \dots, m$) menerima pola target t_k sesuai dengan pola masukan/*input* saat pelatihan dan kemudian informasi kesalahan/*error* lapisan *output* (δ_k) dihitung. δ_k dikirim ke lapisan dibawahnya dan digunakan untuk menghitung besarnya koreksi bobot dan bias (ΔW_{jk} dan ΔW_{ok}) antara lapisan tersembunyi dengan lapisan *output*, perhatikan Persamaan 5.

$$\delta_k = (t_k - y_k) f'(y_{net_k}) = (t_k - y_k) y_k (1 - y_k) \quad (5)$$

Hitung suku perubahan bobot W_{jk} (yang akan digunakan untuk merubah bobot W_{jk}) dengan laju pelatihan $\alpha \Delta W_{jk} = \alpha \delta_k z_j$; $k = 1,2,3,\dots,m$; $j = 0,1,\dots,p$.

Hitung perubahan bias dengan Persamaan 6.

$$\Delta W_{ok} = \alpha \delta_k \quad (6)$$

Langkah 7: Pada setiap unit di lapisan tersembunyi (dari unit ke-1 hingga ke- p ; $i = l \dots n$; $k = l \dots m$) dilakukan perhitungan informasi kesalahan lapisan tersembunyi (δ_j). δ_j kemudian digunakan untuk menghitung besar koreksi bobot dan bias (ΔV_{ji} dan ΔV_{jo}) antara lapisan *input* dan lapisan tersembunyi, perhatikan Persamaan 7 dan Persamaan 8.

$$\delta_{net_j} = \sum_{k=1}^m \delta_k W_{kj} \quad (7)$$

$$\delta_j = \delta_{net_j} f'(\delta_{net_j}) = \delta_{net_j} z_j (1 - z_j) \quad (8)$$

Hitung suku perubahan bobot V_{ji} (yang digunakan untuk perbaikan bobot V_{ji}), perhatikan Persamaan 9.

$$\Delta V_{ji} = \alpha \delta_j x_i \quad (9)$$

Hitung perubahan bias (untuk memperbaiki V_{jo}), perhatikan Persamaan 10.

$$\Delta V_{jo} = \alpha \delta_j \quad (10)$$

Tahap 3: Update Bobot dan Bias

Langkah 8: Masing-masing unit *output*/keluaran ($Y_k, k = 1,2,3,\dots,m$) dilakukan *update* bias dan bobotnya ($j = 0,1,2,\dots,p$) sehingga menghasilkan bobot dan bias baru, perhatikan Persamaan 11.

$$W_{kj} (\text{baru}) = W_{kj} (\text{lama}) + \Delta W_{kj} \quad (11)$$

Demikian juga untuk setiap unit tersembunyi mulai dari unit ke-1 sampai dengan unit ke- p dilakukan *update* bobot dan bias, dengan Persamaan 12.

$$V_{ji} (\text{baru}) = V_{ji} (\text{lama}) + \Delta V_{ji} \quad (12)$$

Langkah 9: Uji kondisi berhenti (akhir iterasi).

D. Nguyen-Widrow

Nguyen-Widrow adalah sebuah algoritma yang digunakan untuk inisialisasi bobot pada JST untuk mengurangi waktu pelatihan. Inisialisasi pembobot dari *neuron input* ke *neuron* tersembunyi bertujuan untuk meningkatkan kemampuan *neuron-neuron* tersembunyi untuk melakukan pembelajaran. Hal ini dilakukan dengan mendistribusikan pembobot dan bias awal sedemikian rupa sehingga dapat meningkatkan kemampuan lapisan tersembunyi dalam melakukan proses pembelajaran. Pada inisialisasi Nguyen-Widrow, inisialisasi acak tetap dipakai tetapi digunakan untuk menginisialisasi bias dan bobot dari unit tersembunyi ke unit *output*. Untuk bias dan bobot dari unit *input* ke unit tersembunyi digunakan bias dan bobot yang diskalakan agar jatuh pada *range* tertentu [8].

III. DESAIN PENELITIAN

Tujuan dari perancangan sistem adalah untuk menganalisa dan mengimplementasikan Jaringan Syaraf Tiruan-*Backpropagation* (JST-BP) dan metode Nguyen-Widrow untuk memprediksi penyakit Diabetes. Sistem dimulai dengan diprosesnya dataset Diabetes Indian Pima untuk dilakukan normalisasi pada masing-masing variabel sehingga nilainya diantara 0 hingga 1. Kemudian dataset dibagi menjadi data *training* dan data *testing*. Untuk data *training* dilakukan proses pelatihan hingga memperoleh bobot yang digunakan untuk proses pengujian dengan menggunakan data *testing*. Dari proses pengujian tersebut didapat hasil klasifikasi. Proses Keseluruhan dapat digambarkan seperti pada Gambar 1.

A. Dataset

Dataset yang digunakan pada penelitian ini adalah data Diabetes Indian Pima. Dataset Pima India ini didapatkan dari <http://archive.ics.uci.edu>. yang bersumber dari pemilik yang asli yaitu *National Institute of Diabetes and*

Digestive and Kidney Diseases yang diambil juga dari Vincent Sigillito (vgs '@' aplcen.apl.jhu.edu), *Research Center, RMI Group Leader, Applied Physics Laboratory, The Johns Hopkins University, Johns Hopkins Road, Laurel, dan MD 20707*. Dataset akan diujicobakan dalam dua tahap yaitu tahap pelatihan dengan menggunakan data *training* dan tahap pengujian dengan menggunakan data *training* dan data *testing*.

B. Preprocessing

Data yang menjadi masukan dalam pelatihan adalah seluruh data dataset Diabetes Indian Pima. Presentase yang digunakan yaitu dibagi menjadi 80% data untuk *training* dan 20% sisanya untuk *testing*. Data *input* ini kemudian dinormalisasi atau disebut dengan penskalaan data. Tujuan dari normalisasi data *input* dan *target* yaitu untuk mentransformasi data supaya kestabilan data dapat dicapai. Data tersebut dinormalisasi dalam interval $[0,1]$ karena terkait dengan fungsi aktivasi sigmoid biner.

C. Pelatihan Jaringan Saraf Tiruan

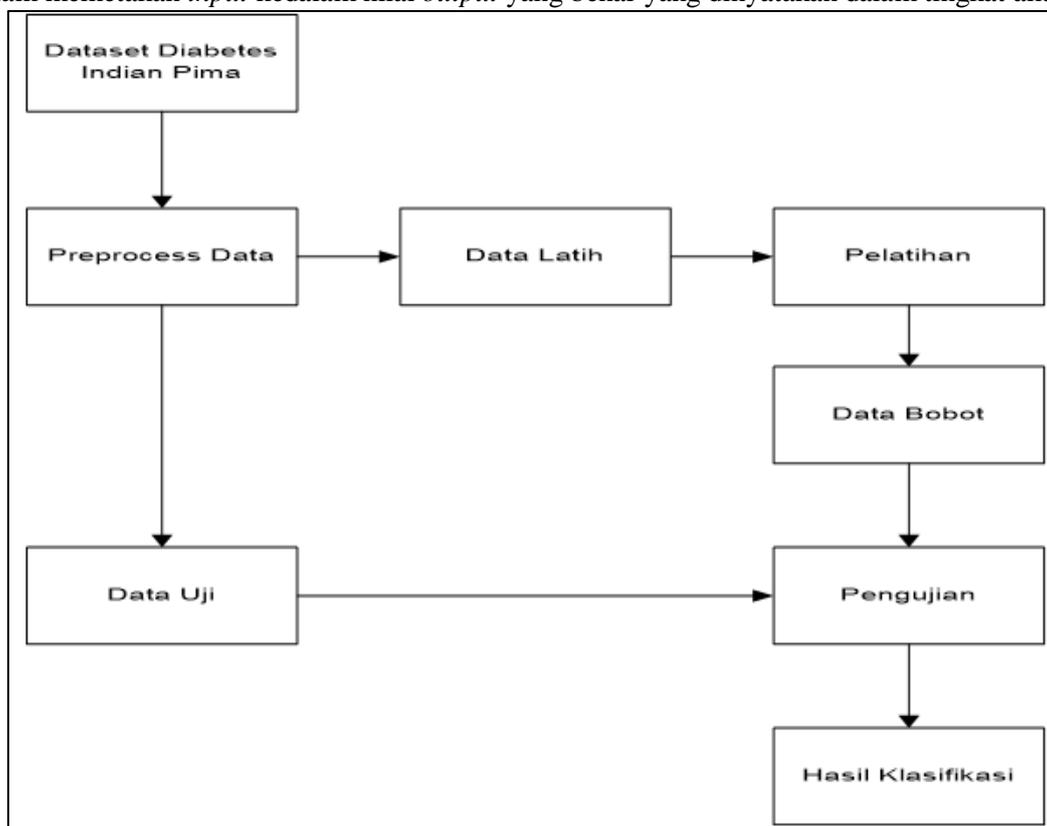
Algoritma JST yang digunakan adalah *backpropagation* dengan menggunakan fungsi sigmoid biner yang memiliki *range output* 0 sampai 1. Pada JST akan dilakukan pembelajaran hingga mencapai MSE atau telah mencapai jumlah iterasi maksimum yang telah ditentukan. Pelatihan JST dilakukan hingga mendapatkan bobot yang stabil pada jaringan. Sebuah bobot dikatakan stabil jika telah mencapai MSE atau telah mencapai jumlah iterasi maksimum yang telah ditentukan.

D. Optimasi Nguyen-Widrow

Langkah awal yaitu mendefinisikan faktor skala $\beta = \text{faktor skala} = 0.7(p)1/n = 0.7$. Nilai bobot harus diantara $-0,5$ hingga $0,5$. Langkah selanjutnya menghitung $||v_j|| = \sqrt{v_{j1}^2 + v_{j2}^2 + \dots + v_{jn}^2}$ dan menentukan bobot-bobot antara unit *input* ke unit tersembunyi, dan yang terakhir yaitu menentukan bias antara *input* ke unit tersembunyi dan terletak pada skala antara $-\beta$ dan β .

E. Pengujian Jaringan Saraf Tiruan

Pengujian terhadap data latih dilakukan dengan menggunakan keseluruhan data latih sebagai masukan bagi jaringan, yang kemudian dicari nilai keluarnya dengan menggunakan bobot, sehingga dapat diketahui kemampuan jaringan dalam memetakan *input* kedalam nilai *output* yang benar yang dinyatakan dalam tingkat akurasi.



Gambar 1. Flowchart Proses Keseluruhan

IV. HASIL DAN PEMBAHASAN

Penelitian dilakukan pada data *testing* Diabetes menggunakan inisialisasi Nguyen-Widrow. Uji coba berikut ini menggunakan data *testing* Diabetes, percobaan ini dilakukan dengan menggunakan nilai parameter *hidden node*

sebanyak 15, *learning rate* sebesar 0.75, momentum sebesar 0.75 dan iterasi maksimal sebanyak 1000000. Uji coba ini dilakukan sebanyak 20 kali percobaan. Hasil uji coba ditampilkan pada Tabel 1.

TABEL I
PENGUJIAN PADA DATA TESTING DIABETES

Percobaan ke -	MSE	Iterasi	Waktu
1	9.479×10^{-6}	3075	1.25
2	7.547×10^{-6}	3690	1.28
3	9.729×10^{-6}	6765	2.29
4	8.299×10^{-6}	4305	1.43
5	9.596×10^{-6}	3690	1.27
6	8.498×10^{-6}	5535	1.88
7	9.920×10^{-6}	3075	1.08
8	9.703×10^{-6}	3075	1.07
9	8.030×10^{-6}	3690	1.25
10	9.216×10^{-6}	3690	1.25
11	8.274×10^{-6}	4305	1.46
12	9.455×10^{-6}	7380	2.41
13	8.443×10^{-6}	4305	1.48
14	9.294×10^{-6}	5535	1.87
15	9.051×10^{-6}	4305	1.5
16	9.757×10^{-6}	4305	1.46
17	8.857×10^{-6}	4305	1.48
18	8.706×10^{-6}	3075	1.05
19	9.082×10^{-6}	5535	1.85
20	9.092×10^{-6}	3690	1.30
Rata - rata	9.0014×10^{-6}	4366.5	1.49

Dari hasil pengujian pada data testing Diabetes dengan menggunakan parameter *hidden layer* sebanyak 15, μ sebesar 0.75, α sebesar 0.75 dan dilakukan uji coba sebanyak 20 kali maka diperoleh nilai rata-rata MSE sebesar 9.0014×10^{-6} , iterasi sebanyak 4366.5 dan waktu selama 1.49 *sec*. Berdasarkan hasil dari pengujian yang telah dilakukan dan dihitung nilai akurasi maka didapatkan akurasi sebesar 100%. Hal ini membuktikan bahwa inisialisasi Nguyen-Widrow memberikan nilai *output* yang lebih baik.

V. KESIMPULAN

Dari hasil analisa terhadap pelatihan dan pengujian JST-BP dan Nguyen-Widrow. dapat diambil kesimpulan sebagai berikut:

1. Beberapa parameter yang mempengaruhi pencapaian target *error* pada JST-BP antara lain:
 - a) Semakin banyak jumlah *hidden node*. pada umumnya semakin sedikit jumlah iterasi yang dibutuhkan dan semakin lama waktu pengujian yang diperlukan.
 - b) Nilai momentum yang besar mengakibatkan jaringan lebih cepat mencapai minimum *error* jika dibandingkan dengan momentum yang kecil.
 - c) Nilai *learning rate* yang kecil mengakibatkan waktu pengujian yang lama untuk mencapai minimum *error*, akan tetapi perubahan MSE cukup stabil. Sedangkan *learning rate* yang besar mengakibatkan minimum *error* yang cepat tercapai namun jaringan kurang stabil.
2. Dengan menggunakan metode Nguyen-Widrow sebagai inisialisasi bobot awal hasil target yang didapatkan lebih optimal dibandingkan dengan inisialisasi bobot *random*.

VI. DAFTAR PUSTAKA

- [1] K. V. Narayan, J. P. Boyle, T. J. Thompson, S. W. Sorensen and D. F. Williamson, "Lifetime Risk for Diabetes Mellitus in the United States," *JAMA*, vol. 290, no. 14, pp. 1884-1890, 2003.

- [2] M. Kharola and D. Kumar, "Efficient Weather Prediction By Back-Propagation Algorithm," *IOSR Journal of Computer Engineering (IOSR-JCE)*, vol. 16, no. 3, pp. 55-58, 2014.
- [3] A. M. Zamani, B. Amaliah dan A. Munif, "Implementasi Algoritma Genetika pada Struktur Backpropagation Neural Network untuk Klasifikasi Kanker Payudara," *JURNAL TEKNIK ITS*, vol. 1, pp. A-222 - A-227, 2012.
- [4] M. P. M and D. W. Aha, *UCI Repository of Machine Learning Databases (Machine Readable Data Depository)*, California: Department of Information and Computer Science. University of California. Irvine. CA., 1995.
- [5] W. C. Knowler, P. H. Bennett, R. F. Hamman And M. Miller, "Diabetes Incidence And Prevalence In Pima Indians: A 19-Fold Greater Incidence Than In Rochester, MINNESOTA," *American Journal of Epidemiology*, vol. 108, no. 6, pp. 497-505, 1978.
- [6] B. R. Valluru and V. R. Hayagriva, *C++ Neural Networks And Fuzzy Logic*, MIS, 1995.
- [7] A. Hermawan, *Jaringan Saraf Tiruan Teori dan Aplikasi*, Yogyakarta: Andi, 2006.
- [8] P. Diyah, *Pengantar Jaringan Syaraf Tiruan*, Yogyakarta: Andi, 2006.