

OPTIMASI JARINGAN SYARAF TIRUAN BACKPROPAGATION MENGUNAKAN PARTICLE SWARM OPTIMIZATION UNTUK DETEKSI PENDERITA PENYAKIT JANTUNG

Hargianti Henni Oktawandari

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang
Jl. Nakula 1 No. 5-11, Semarang, 50131, Telp : (024) 351 7261, Fax (024) 352 0165
E-mail : hargiantiokt@gmail.com

Abstrak

Menurut data dari Himpunan Mahasiswa Epidemiologi Fakultas Kesehatan Masyarakat Universitas Hasanuddin (HIMAPID), di Indonesia, dengan jumlah penduduk sekitar 237 juta jiwa, terjadi kelahiran 6,6 juta bayi setiap tahunnya dan 48.800 di antaranya sudah menyandang penyakit jantung. Untuk mengklasifikasikan apakah seseorang terindikasi penyakit jantung, dalam penelitian kali ini diterapkan Artificial Neural Network atau Jaringan Syaraf Tiruan model Backpropagation dan Jaringan Syaraf Tiruan model Backpropagation yang bobotnya dioptimasi menggunakan Particle Swarm Optimization dengan tujuan meminimalkan mean squared error (MSE). Dari kedua metode tersebut akan dicari nilai akurasi tertinggi melalui Confusion Matrix dan penurunan nilai RMSE untuk kemudian diterapkan pada Graphical User Interface (GUI). Data yang digunakan berasal dari UCI Machine Learning Respository dengan 270 hasil observasi partisipan terdiri dari 13 variabel yang ditentukan untuk melakukan deteksi. Hasil pengujian menunjukkan pada percobaan kedua nilai akurasi BPNN+PSO lebih tinggi dari BPNN saja, yakni 87.7% dan nilai RMSE menurun dari 1.0862 menjadi 0.3403.

Kata Kunci: backpropagation, PSO, RMSE

Abstract

According to data from the Student Association of Epidemiology School of Public Health, Hasanuddin University (HIMAPID), in Indonesia, with a population of around 237 million people, there were 6.6 million babies born each year and 48,800 of them are already bearing heart disease. To classify whether a person is indicated for heart disease, this current research applied Artificial Neural Network model of Backpropagation and Backpropagation Neural Network models that its weight are optimized using Particle Swarm Optimization with the aim of minimizing the mean squared error (MSE). Both of those methods will be sought through the highest accuracy using Confusion Matrix and the impairment of RMSE for then applied to the Graphical User Interface (GUI). The data that used are from UCI Machine Learning respository with the observation of 270 participants consisting of 13 variables were determined to do the detection. The test results show the accuracy of BPNN + PSO is higher than BPNN, ie, 87.7% and RMSE values decreased from 1.0862 becomes 0.3403.

Keywords: backpropagation, PSO, RMSE

1. PENDAHULUAN

Menurut data dari Himpunan Mahasiswa Epidemiologi Fakultas Kesehatan Masyarakat Universitas Hasanuddin (HIMAPID), jumlah penderita hipertensi di Indonesia lebih dari 75 juta orang dimana setiap empat

orang dewasa terdapat satu orang yang berisiko terserang penyakit jantung. Di Indonesia, dengan jumlah penduduk sekitar 237 juta jiwa, terjadi kelahiran 6,6 juta bayi setiap tahunnya dan 48.800 di antaranya sudah menyandang penyakit jantung [1].

Dilihat dari berbahaya dan besarnya jumlah penderita penyakit jantung, penulis bermaksud membuat penelitian mengenai deteksi apakah seseorang menderita penyakit jantung atau tidak berdasarkan variabel yang telah ditentukan. Dalam penelitian kali ini akan dilakukan klasifikasi dengan menerapkan *Artificial Neural Network* atau Jaringan Syaraf Tiruan model *Backpropagation*. Jaringan Syaraf Tiruan dapat digunakan untuk melakukan estimasi berdasar kemampuan jaringan syaraf tiruan untuk mengingat dan membuat generalisasi dari data yang sudah ada sebelumnya [2]. Dengan *backpropagation*, rekaman data dipakai sebagai data pelatihan untuk mencari bobot yang optimal.

Selanjutnya akan diterapkan pula algoritma *Particle Swarm Optimization*. Algoritma tersebut digunakan untuk mengoptimasi bobot pada pelatihan jaringan syaraf tiruan, tujuannya adalah meminimumkan *root mean squared error* (RMSE) dan meningkatkan akurasi.

Pramunendar, Dewi dan Asari (2013) melakukan prediksi awal penyakit jantung menggunakan *Backpropagation Neural Network* dengan metode Adaboost. Data yang digunakan diperoleh dari *database* UCI dengan 13 atribut dan sebuah keluaran. Dari penelitian tersebut didapatkan akurasi menggunakan BPNN sebesar 96.65% dan BPNN dengan Adaboost sebesar 99.29% [3]. Sedangkan Misbahuddin (2011) menerapkan algoritma PSO pada jaringan syaraf tiruan *backpropagation* untuk mengoptimasi bobot koneksi ketika proses pelatihan [4]. Kedua penelitian tersebut menunjukkan bahwa jaringan syaraf tiruan tepat digunakan untuk melakukan klasifikasi.

Pada penelitian ini diusulkan implementasi jaringan syaraf tiruan dengan metode pelatihan *backpropagation* yang bobotnya dioptimasi menggunakan *particle swarm optimization* guna mendapatkan hasil iterasi yang lebih baik untuk melakukan deteksi penderita jantung atau bukan berdasarkan variabel tertentu.

2. TINJAUAN PUSTAKA

2.1 Data Mining

Data mining adalah analisa terhadap data yang biasanya berukuran besar agar bermanfaat bagi pemilik data. Proses *data mining* dibagi menjadi tiga aktivitas yaitu:

- 1) Eksplorasi data, terdiri dari aktivitas pembersihan data, transformasi data, pengurangan dimensi, pemilihan ciri dan lain-lain.
- 2) Membuat model dan pengujian validasi model
- 3) Penerapan model dengan data baru untuk menghasilkan perkiraan dari kasus yang ada.

Aplikasi yang menggunakan *data mining* dimaksudkan untuk menyelesaikan permasalahan dengan membangun model dari data yang sudah digali kemudian diterapkan pada data yang lain [5]. Ada dua jenis aplikasi data mining:

- 1) Metode prediksi. Bermaksud memprediksi nilai yang akan datang berdasarkan data-data yang telah ada variabelnya seperti klasifikasi, regresi dan lainnya.
- 2) Metode deskriptif. Bermaksud membantu *user* agar mudah melihat pola-pola yang berasal dari data yang ada.

Klasifikasi merupakan bagian dari *data mining*. Klasifikasi adalah proses memetakan setiap sekumpulan atribut x (input) ke salah satu *class* y (target) yang telah didefinisikan sebelumnya [6]. Komponen utama dari proses klasifikasi antara lain:

- 1) Kelas, variabel tidak bebas yang merupakan label dari hasil klasifikasi.
- 2) Prediktor, merupakan variabel bebas yakni atribut data yang diklasifikasi.
- 3) Set data pelatihan, merupakan sekumpulan data berisi kelas dan predictor untuk dilatih agar model dapat mengelompokkan ke dalam kelas yang tepat.
- 4) Set data uji, berisi data-data baru yang dikelompokkan oleh model guna mengetahui akurasi model yang telah dibuat.

2.2 Jaringan Syaraf Tiruan

Jaringan Syaraf Tiruan (JST) adalah sistem pemroses informasi yang memiliki karakteristik mirip dengan jaringan syaraf biologi.

JST ditentukan oleh tiga hal:

- a. Pola hubungan antar neuron (arsitektur jaringan)
- b. Metode untuk menentukan bobot penghubung (metode *training* atau *learning*)
- c. Fungsi aktivasi

2.2.1 Backpropagation

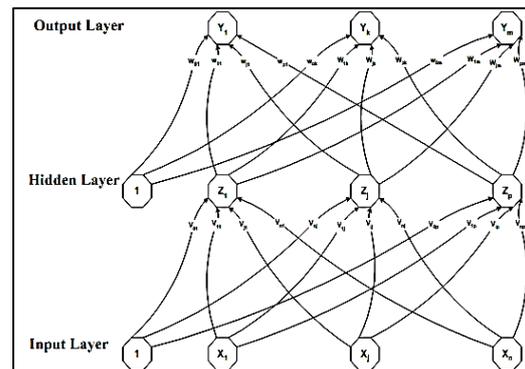
Backpropagation melatih jaringan untuk mendapatkan keseimbangan antara kemampuan jaringan untuk mengenali pola yang digunakan selama pelatihan serta kemampuan untuk memberikan respon yang benar terhadap pola masukan yang serupa dengan pola yang dipakai selama pelatihan.

Pola yang dibutuhkan dipengaruhi banyaknya bobot dalam jaringan serta

tingkat akurasi yang diharapkan. Jaringan dapat terus dilatih hingga semua pola pelatihan dapat dikenali dengan benar, namun tidak menjamin jaringan mampu mengenali pola pengujian dengan benar.

Umumnya data dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian. Perubahan bobot dilakukan berdasar pola pelatihan. Tetapi selama pelatihan, kesalahan yang terjadi dihitung berdasar semua data (pelatihan dan pengujian). Jika kesalahan menurun, pelatihan terus dijalankan. Jika kesalahan mulai meningkat, pelatihan dihentikan.

Pelatihan *backpropagation* terdiri atas tiga tahap: propagasi maju, propagasi mundur dan perubahan bobot [2]. Ketiga fase tersebut diulang hingga kondisi penghentian dipenuhi, yaitu menurut jumlah iterasi atau kesalahan. Dihentikan jika jumlah iterasi yang dilakukan melebihi jumlah maksimum iterasi yang ditetapkan, atau jika kesalahan yang terjadi sudah lebih kecil dari batas toleransi.



Gambar 1. Arsitektur Backpropagation

2.3 Particle Swarm Optimization

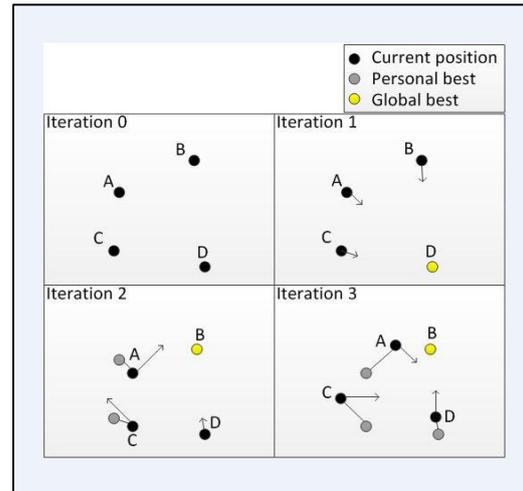
Particle Swarm Optimization (PSO) merupakan algoritma *swarm intelligence* yakni studi mengenai sistem komputasi yang terinspirasi oleh kecerdasan kolektif [7]. Kecerdasan

kolektif muncul berdasarkan populasi atau kerjasama homogen dalam sebuah lingkungan.

Dalam PSO, sebuah kawanan diasumsikan memiliki ukuran tertentu dengan setiap partikel memiliki posisi awal yang terletak acak di suatu lokasi dalam satu ruang. Setiap partikel diasumsikan memiliki dua karakter: posisi dan kecepatan [8]. Setiap partikel bergerak dalam ruang tertentu dan mengingat posisi terbaik yang pernah dilalui atau ditemukan terhadap sumber makanan. Setiap partikel menyampaikan informasi atau posisi terbaiknya kepada partikel lain.

Pada algoritma PSO, pencarian solusi dilakukan oleh suatu populasi yang terdiri dari beberapa partikel. Populasi dibangkitkan secara random dengan batasan nilai terkecil dan terbesar (batas bawah dan batas atas). Setiap partikel mencari solusi dengan melintasi ruang pencarian dengan cara melakukan penyesuaian terhadap posisi terbaiknya (*local best*) dan penyesuaian terhadap posisi partikel terbaik dari seluruh kawanan (*global best*) selama melintasi *search space*.

Dilakukan sejumlah iterasi untuk mencari posisi terbaik setiap partikel sampai didapatkan posisi yang relatif tetap atau mencapai batas iterasi yang ditetapkan. Pada tiap iterasi, setiap solusi (posisi partikel) dievaluasi performanya dengan cara memasukkan solusi tersebut ke dalam *fitness function*.



Gambar 2. Proses Partikel dalam Mencari Posisi Terbaik

PSO terbukti merupakan algoritma yang baik dan efektif untuk permasalahan optimasi karena kemudahannya dalam penerapan kode dan performanya yang konsisten [9]. Algoritma PSO yang digunakan untuk mengoptimasi bobot pada pelatihan jaringan syaraf tiruan tujuannya adalah meminimumkan *mean squared error* (MSE) [4].

3. METODE PENELITIAN

3.1 Kebutuhan Data

Dalam penelitian ini, digunakan dataset penyakit jantung yang didapat dari UCI *Machine Learning Respository*. Data tersebut memuat 13 atribut dan 2 variabel yang diprediksi serta terdapat sebanyak 270 observasi.

Atribut yang digunakan berdasarkan hasil pemeriksaan laboratorium, yaitu: *age*, *sex*, *chest pain type*, *resting blood pressure*, *serum cholestorol in mg/dl*, *fasting blood sugar >120mg/dl*, *resting electrocardiographic results*, *maximum heart rate achieved*, *exercise induced angina*, *oldpeak*, *the slope of the peak exercise ST segment*, *number of major vessels*, dan *thallium scan*. Keluaran dari atribut masukan yakni keputusan positif

atau negatif adanya indikasi penyakit jantung pada seseorang. Terdapat atribut yang merupakan bilangan real dan atribut berbentuk string yang kemudian dikonversi menjadi bilangan biner.

3.2 Teknik Analisis Data

Data yang digunakan dibagi menjadi data *train* dan data *test* untuk dua macam percobaan:

1. 270 data *train* dan 100 data *test* yang diambil dari data *train*.
2. 189 data *train* dan 81 data *test* yang berada diluar data *train*.

Percobaan 1 dilakukan untuk mengetahui apakah model yang dibentuk sudah cukup baik untuk diuji.

Sebelum dilakukan pelatihan dan pengujian menggunakan metode BPNN dan PSO, dilakukan normalisasi data agar data berada pada range [-1 1]. Hal ini perlu dilakukan guna menghindari data yang tidak konsisten, misalnya dua variabel dengan data yang berjarak cukup jauh. Konsistensi data merupakan salah satu syarat agar data dapat dinyatakan berkualitas untuk selanjutnya dapat diproses menjadi sebuah informasi yang lebih bermanfaat. Dilakukan normalisasi untuk masing-masing data pelatihan dan data uji.

Kedua kelompok data percobaan tersebut kemudian dilatih menggunakan jaringan syaraf tiruan *backpropagation* (BPNN) dan BPNN yang bobotnya dioptimasi menggunakan *particle swarm optimization* (PSO).

Berdasarkan hasil pengujian kedua metode tersebut akan ditentukan akurasi tertinggi. Terakhir, akurasi dan nilai *root mean squared error* (RMSE) dari masing-masing percobaan dibandingkan dan dipilih metode yang akan

diimplementasikan ke dalam *graphical user interface* (GUI).

3.3 Evaluasi Hasil

Hasil pelatihan dan pengujian diketahui berdasarkan nilai akurasi menggunakan *confusion matrix* serta nilai RMSE yang dihitung dengan rumus:

$$MSE = \frac{1}{M} \sum_{t=1}^M e_t^2 \quad (1)$$

dengan,

$$e_t = (Y_t - \hat{Y}_t) \text{ (taksiran sisa peramalan)}$$

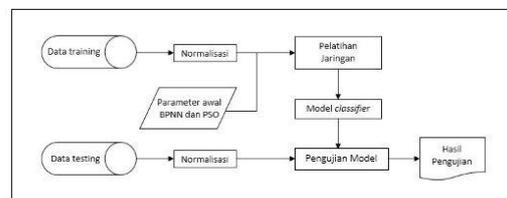
M = jumlah residual

RMSE merupakan akar dari MSE:

$$RMSE = \sqrt{MSE} \quad (2)$$

4. HASIL DAN PEMBAHASAN

Pelatihan jaringan menggunakan kedua metode dimulai dengan proses normalisasi data *train* kemudian parameter awal masing-masing metode ditentukan. Jaringan dilatih hingga membentuk model *classifier*. Selanjutnya model tersebut diuji menggunakan data *test* yang telah dinormalisasi dan menghasilkan hasil pengujian yang dapat diketahui berdasarkan nilai akurasi dan RMSE-nya.



Gambar 3. Proses Pelatihan dan Pengujian Menggunakan BPNN dan BPNN+PSO

Setelah melakukan proses pelatihan dan pengujian jaringan, didapatkan data berikut:

Tabel 1: Perbandingan Hasil Pengujian

Percobaan 1	BPNN		BPNN+PSO	
	Akurasi	RMSE	Akurasi	RMSE
	85.0%	0.9390	84.0%	0.3921
	95.0%	0.7046	80.0%	0.3820
	93.0%	0.6258	83.0%	0.3708
	90.0%	0.7839	82.0%	0.3704
	85.0%	0.8947	85.0%	0.3634
Percobaan 2	BPNN		BPNN+PSO	
	Akurasi	RMSE	Akurasi	RMSE
	74.1%	1.4043	84.0%	0.3818
	74.1%	1.2582	82.7%	0.3695
	75.3%	1.2177	86.4%	0.3723
	76.5%	1.2781	85.2%	0.3928
	82.7%	1.0862	87.7%	0.3403

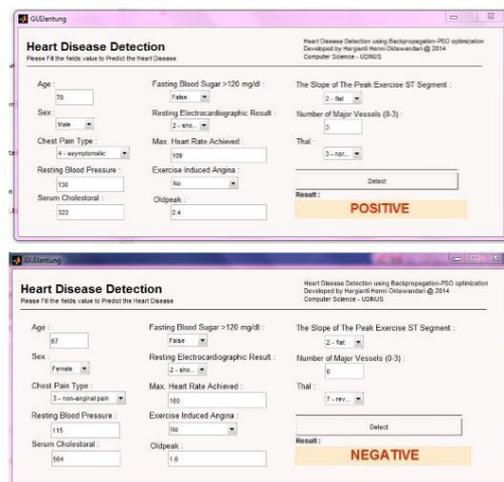
Pada percobaan 2 hasil pelatihan menggunakan BPNN mencapai akurasi 82.7% dengan nilai RMSE sebesar 1.0862 sedangkan pelatihan menggunakan BPNN dengan optimasi PSO menghasilkan akurasi sebesar 87.7% dan RMSE sebesar 0.3403. Tabel 1 menunjukkan bahwa pelatihan jaringan BPNN dengan bobot yang dioptimasi menggunakan PSO memiliki hasil yang lebih baik dibanding pelatihan jaringan BPNN tanpa PSO.

4.1 Implementasi pada GUI

Berdasarkan beberapa percobaan yang telah dilakukan, dapat dilihat bahwa dengan menerapkan PSO guna mengoptimasi bobot pada *backpropagation* nilai akurasi meningkat dan nilai RMSE menjadi lebih kecil. Dapat dinyatakan bahwa *backpropagation* yang bobotnya telah dioptimasi menggunakan metode PSO memiliki hasil iterasi yang lebih baik. Percobaan 2 dipilih untuk diimplementasikan pada GUI karena memenuhi dua parameter tersebut.

Hasil pelatihan disimpan dalam file bernama *psobpNew.mat* dipanggil kembali ke dalam *workspace* menggunakan perintah 'load' pada matlab. File berisi rekaman hasil

pelatihan jaringan tersebut digunakan dalam membangun *graphical user interface* (GUI) sebagai bentuk implementasi guna mendeteksi apakah seseorang menderita penyakit jantung atau tidak. Data yang diinputkan pada GUI merupakan data yang belum dinormalisasi, oleh karena itu data yang berasal dari GUI perlu dipreprosesing sebelum disimulasikan dengan jaringan *psobpNew.mat* yang di load pada *workspace*. Setelah disimulasikan, data hasil keluaran simulasi dengan jaringan tersebut di *post-processing* untuk melihat hasil klasifikasi apakah pasien terindikasi penyakit jantung atau tidak.



Gambar 4. Graphical User Interface

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan dapat ditarik kesimpulan bahwa hasil pelatihan menggunakan metode BPNN yang bobotnya dioptimasi dengan PSO nilai akurasinya mencapai 87.7%, lebih tinggi dibanding hasil pelatihan yang hanya menggunakan metode BPNN yaitu sebesar 82.7%. Hasil iterasi yang lebih baik juga ditunjukkan dengan nilai RMSE yang lebih kecil sebesar 0.3403 dibandingkan dengan tanpa

menggunakan PSO yang menghasilkan RMSE sebesar 1.0862.

5.2 Saran

Beberapa saran yang dapat diberikan untuk penelitian selanjutnya yaitu:

1. Dapat dilakukan perbaikan pada parameter yang digunakan dalam penelitian ini sehingga diperoleh hasil pelatihan yang lebih baik.
2. Perlu dilakukan pengujian validasi menggunakan *split validation* guna menemukan akurasi yang lebih baik dan nilai *error* yang lebih kecil.
3. Direkomendasikan untuk menyempurnakan penelitian yang telah dilakukan ini dengan menggunakan metode optimasi lainnya.

DAFTAR PUSTAKA

- [1] Indri, "ANTARANEWS.com," 5 July 2013. [Online]. Available: <http://www.antaraneWS.com/berita/383661/penyakit-jantung-penyebab-kematian-nomor-satu-di-jakarta>. [Accessed 14 April 2014].
- [2] J. J. Siang, Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan MATLAB, Yogyakarta: ANDI, 2009.
- [3] R. A. Pramunendar, I. N. Dewi and H. Asari, "Penentuan Prediksi Awal Penyakit Jantung Menggunakan Algoritma Backpropagation Neural Network dengan Metode Adaboost," *Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2013*, 2013.
- [4] Misbahuddin, "Optimasi Jaringan Syaraf Tiruan Berbasis Particle Swarm Optimization," *Dielektrika*, vol. 2, no. 1, 2011.
- [5] P. P. Widodo, R. T. Handayanto and Herlawati, Penerapan Data Mining dengan Matlab, Bandung: Rekayasa Sains, 2013.
- [6] Bertalya, "Konsep Data Mining (Klasifikasi : Pohon Keputusan)," Universitas Gunadarma, 2009.
- [7] J. Brownlee, "Clever Algorithms," 2013. [Online]. Available: <http://www.cleveralgorithms.com/nature-inspired/swarm.html>. [Accessed 4 April 2014].
- [8] B. Santosa, "Tutorial Particle Swarm Optimization," [Online]. Available: <http://oc.its.ac.id/ambilfile.php?idp=1221>. [Accessed 13 April 2014].
- [9] N. Dhanasaputra and B. Santosa, "Pengembangan Algoritma Cat Swarm Optimization (CSO) untuk Klasifikasi," *Jurnal ITS*, p. 3, 2010.