

PENERAPAN METODE NEURAL NETWORK BERBASIS PARTICLE SWAM OPTIMIZED UNTUK PREDIKSI KESUBURAN PADA PRIA

Hilda Amalia

Program Studi Manajemen Informatika
Akademik Manajemen Informatika dan Komputer Bina Sarana Informatika
AMIK BSI Jakarta
Jl. RS. Fatmawati no.24, Jakarta Selatan
hilda.ham@bsi.ac.id

ABSTRACT

The fertility rate is to be considered, especially for adult men is because infertility is a global problem that occurs in couples. Infertility can result in couples unable to conceive. Infertility in one partner may terminate especially male lineage. Its high level of infertility in today's society can be caused by several things between environmental factors and lifestyle of today's society. Previous research on fertility has been done is by using artificial neural network and produce 82% accuracy. In this paper will be improving the accuracy of neural network method for predicting male fertility through semen using particle swam optimized (PSO). The survey results revealed that the use of the optimization method can improve the accuracy of the method used in this study was obtained accuracy of new methods for neural network improved its performance using PSO to 92%.

Keyword: *Neural Network, Optimized and Fertility*

I. PENDAHULUAN

Ketidaksuburan merupakan masalah global, yang mempengaruhi rata-rata 8-12 persen pasangan di seluruh dunia. Meskipun jarang diakui secara sosial, ketidaksuburan pada laki-laki merupakan kontribusi untuk setidaknya setengah dari semua kasus di seluruh dunia dan ketidaksuburan merupakan hal yang paling sulit untuk mengobati (Inhorn, 2002). Beberapa hal yang mempengaruhi tingkat kesuburan pada pria yaitu dapat disebabkan karena penyakit ataupun juga efek lingkungan dan pekerjaan serta gaya hidup masyarakat (Gilera dkk, 2012). Untuk mengevaluasi ketidaksuburan pada laki-laki, dokter menggunakan data yang diperoleh dari analisis sperma. Analisis air mani adalah prediktor yang baik dari potensi kesuburan pria (Kolettis, 2003). *As-sisted Reproductive Technology* (ART) adalah salah satu dari banyak perawatan yang tersedia Pilihan untuk menyembuhkan ketidaksuburan. Namun, efisiensi pengobatan ART masih tidak memadai (Mileski dkk, 2013). Proses analisis data medis memerlukan canggih teknik untuk pengolahan, penyimpanan dan pengaksesan informasi data. Teknik tradisional tidak cukup mampu untuk menghasilkan hasil yang optimal dari lengkap atau data yang

berlebih melalui proses analisis (Durairaj dan Sathyavathi, 2013).

Penggunaan informasi yang efektif dan teknologi sangat penting untuk perawatan kesehatan. Ha ini dibutuhkan oleh organisasi yang bergerak dibidang kesehatan untuk tetap kompetitif pada saat ini. Tantangan yang dihadapi ketika mencoba untuk memahami besar, beragam, dan sering Sumber data yang kompleks dari cukup. Dalam upaya untuk mengubah informasi menjadi pengetahuan, organisasi perawatan kesehatan yang menerapkan teknologi data mining dapat digunakan untuk membantu mengendalikan biaya perawatan dan meningkatkan efektivitas perawatan pasien.

Data mining dapat digunakan untuk membantu memprediksi perilaku pasien masa depan dan untuk meningkatkan program pengobatan. Data mining telah digunakan dalam sejumlah domain nonmedis (Sivanandam, 2006). Dengan mengidentifikasi pasien berisiko tinggi, dokter akan lebih baik memberikan perawatan pasien hari ini sehingga tidak menimbulkan masalah yang berarti dikemudian hari (Sivanandam, 2006).

Oleh karena itu, perlu dilakukan analisis data menggunakan teknik-teknik data mining. Data mining secara umum sudah diakui sebagai salah satu cara mengekstrak data menjadi

pengetahuan yang berharga. Penelitian mengenai tingkat kesuburan pada pria dengan menganalisa kualitas air mani sudah pernah dilakukan dengan metode artificial neural network (Girela dkk, 2012), komparasi metode bayesian network dan naïve bayes (Kahki dkk, 2013). Dalam penelitian ini metode data mining yang akan digunakan yaitu metode data mining. Dan peningkatan kinerja akurasi metode neural network menggunakan metode optimasi *particle swarm optimized* (PSO).

II. KAJIAN LITERATUR

2.1 Data Mining

Data mining adalah analisis data yang besar pada set data untuk menemukan hubungan terduga dan untuk meringkas data dalam cara baru yang sama-sama dimengerti dan berguna bagi pemilik data (Larose, 2007). data mining seharusnya lebih tepat bernama "pertambahan pengetahuan dari data". Banyaknya data, ditambah dengan kebutuhan untuk alat analisis data yang kuat, memiliki digambarkan sebagai situasi miskin kaya data tetapi informasi. Yang tumbuh cepat, luar biasa jumlah data, dikumpulkan dan disimpan dalam repositori data yang besar dan banyak, memiliki jauh melebihi kemampuan manusia untuk pemahaman tanpa alat yang kuat.

Banyaknya data, ditambah dengan kebutuhan untuk alat analisis data yang kuat, telah digambarkan sebagai kaya data tapi miskin informasi. Jumlah data yang tumbuh secara cepat, dikumpulkan dan disimpan dalam repositori data yang besar dan banyak, telah jauh melampaui kemampuan manusia untuk memahami data-data tersebut tanpa mampu mengolah data tersebut. Akibatnya, data yang dikumpulkan dalam repositori data yang besar menjadi "kuburan data" (Han dan Kamber, 2007).

2.2 Metode Neural Network

Neural network adalah suatu usaha untuk meniru fungsi otak manusia. Otak manusia diyakini terdiri dari jutaan unit pengolahan kecil, yang disebut neuron, yang bekerja secara parallel (Shukla dkk, 2010). Generalized Regression neural Network (GRNN) Generalisasi adalah kemampuan ANN untuk memberikan jawaban yang benar untuk input yang tidak diketahui. Generalisasi adalah ukuran dari seberapa baik sistem telah dilatih. Generalisasi diukur dengan kinerja dari sistem untuk mengatur data pengujian. Kemampuan generalisasi ANN itu maka harus setinggi mungkin. (Shukla dkk, 2010).

Penemuan algoritma *backpropagation* untuk *multilayer perceptron*, 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).

Multilayer perceptron (MLP) disebut juga *multilayer feedforward neural network* merupakan algoritma yang paling luas digunakan. Menurut Wong, Bodnovich dan Selvi (1997), sekitar 95% aplikasi bisnis yang menggunakan neural network, memakai algoritma ini (Vecellis, 2009).

Langkah pembelajaran dalam algoritma *backpropagation* adalah sebagai berikut (Myatt, 2007):

1. Inisialisasi bobot jaringan secara acak (biasanya antara -0.1 sampai 1.0).
2. Untuk setiap data pada data training, hitung input untuk simpul berdasarkan nilai input bobot jaringan saat itu, menggunakan rumus:

$$Input_j = \sum_{i=1}^n O_i w_{ij} + \theta_j$$

Keterangan:

O_i = output simpul i dari layer sebelumnya

w_{ij} = bobot relasi dari simpul i pada layer sebelumnya ke simpul j.

θ_j = bias (sebagai pembatas).

Berdasarkan input dari langkah dua, selanjutnya membangkitkan output untuk simpul menggunakan fungsi aktivasi sigmoid:

$$Output = \frac{1}{1 + e^{-Input}}$$

4. Hitung nilai error antara nilai yang diprediksi dengan nilai yang sesungguhnya menggunakan rumus:

$$Error_j = Output_j \cdot (1 - Output_j) \cdot (Target_j - Output_j)$$

Keterangan:

$Output_j$ = output actual dari simpul j .

$Target_j$ = nilai target yang sudah diketahui pada data training

- Setelah nilai error dihitung, selanjutnya dibalik ke layer sebelumnya (*backpropagated*). Untuk menghitung nilai *error* pada *hidden layer* menggunakan rumus:

$$Error_j = Output_j(1 - Output_j) \sum_{k=1}^n Error_k w_{jk}$$

Keterangan:

$Output_j$ = output actual dari simpul j

$Error_k$ = error simpul k .

w_{jk} = bobot relasi dari simpul j ke simpul k pada layer berikutnya

- Nilai error yang dihasilkan dari langkah sebelumnya digunakan untuk memperbaharui bobot relasi menggunakan rumus:

$$w_{ij} = w_{ij} + l \cdot Error_j \cdot Output_i$$

Keterangan:

w_{ij} = bobot relasi dari unit i pada layer sebelumnya ke unit j

l = learning rate (konstanta, nilainya antara 0 sampai dengan 1)

$error_j$ = error pada output layer

simpul j

$output_i$ = output dari simpul

2.3 Particle Swam optimized

Optimization Particle Swarm (PSO) adalah algoritma pencarian berbasis populasi dan diinisialisasi dengan populasi solusi acak yang disebut partikel (Abraham, 2006). PSO merupakan metode pencarian penduduk yang berasal dari studi tentang pergerakan sekelompok burung atau ikan. Serupa dengan algoritma genetika (GA), PSO melakukan pencarian menggunakan populasi (kawanan) individu (partikel) yang akan memperbaharui dari iterasi ke iterasi.

Cari solusi yang optimal, setiap partikel bergerak ke arah posisi terbaik sebelumnya dan posisi terbaik global. Sebagai contoh, partikel- i dinyatakan sebagai: $x_i = (x_{i1}, x_{i2}, \dots, x_{id})$ dalam ruang d -dimensi. Posisi sebelumnya terbaik dari partikel- i disimpan dan dinyatakan sebagai $pbest_i = (pbest_i, 1, pbest_i, 2, \dots, pbest_i, d)$. Partikel terbaik Index antara semua partikel dalam kelompok kawanan dinyatakan sebagai

$gbest_d$. Kecepatan partikel I -th dinyatakan sebagai: $v_i = (v_{i, 1}, v_{i, 2}, \dots, v_{i, d})$. Modifikasi kecepatan dan posisi setiap partikel dapat dihitung dengan menggunakan kecepatan arus dan jarak $pbest_i, d$ untuk $gbest_d$ seperti yang ditunjukkan dalam persamaan berikut:

- $v_{i, d} = w \cdot v_{i, d} + c_1 \cdot R \cdot (pbest_i, d - x_{i, d}) + c_2 \cdot R \cdot (gbest_d - x_{i, d})$

- $x_{id} = x_{i, d} + v_{i, d}$, d

dimana:

V_i, d = kecepatan partikel pada iterasi ke- i

w = faktor bobot inersia

c_1, c_2 = Konstan Percepatan (learning rate)

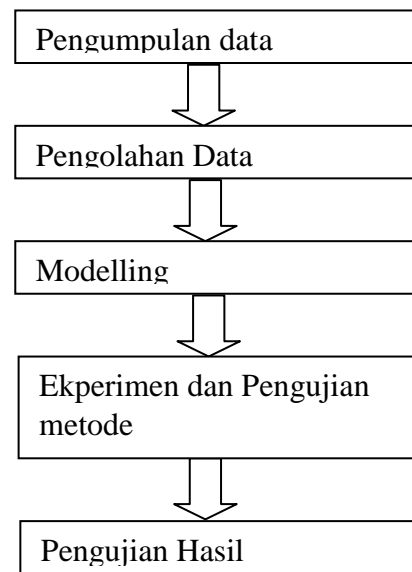
R = Nomor Acak (0-1)

x_i, d = posisi saat partikel pada iterasi ke- i

$pbest_i$ = posisi sebelumnya terbaik dari partikel ke- i (Abraham, 2009)

III. METODELOGI PENELITIAN

Pada penelitian ini data yang digunakan adalah data yang tersimpan pada uci repository yaitu sebanyak 101 data. Atribut yang digunakan sebanyak Sembilan atribut dengan satu label hasil. Dalam melakukan penelitian ini berikut tahapan penelitian yang dilakukan:



Sumber: Hasil Penelitian (2015)

Gambar 1. Tahapan Penelitian yang digunakan

a. Pengumpulan Data

Data diambil dari web yang menyediakan dataset yang dapat digunakan untuk keperluan penelitian yaitu dataset fertility yang ada uci repository dengan alamat <http://archive.ics.uci.edu/ml/datasets/Fertility>. Data yang diolah sebanyak 101 record dengan 9 parameter dan satu label hasil. Parameter

yang digunakan adalah musim saat analisis data, umur, penyakit saat kecil (misalnya cacar air, gondok, polio), kecelakan atau trauma yang dialami, pernah dibedah, demam tinggi pada tahun lalu, frekuensi konsumsi alcohol, kebiasaan merokok, jumlah jam yang dihabiskan untuk duduk perhari. Dan yang terakhir adalah parameter hasil yaitu output Normal (N), diubah (O). normal mengidentifikasi bahwa kondisi air mani tidak mengalami masalah dan kategori subur dan diubah berarti kondisi air mani bermasalah atau tidak subur.

b. Pengolahan Awal Data

Untuk mendapatkan data yang berkualitas, beberapa teknik yang dilakukan adalah sebagai berikut (vecellis, 2009):

- (1) *Data validation*, untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*).
- (2) *Data integration and Transformation*, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal.

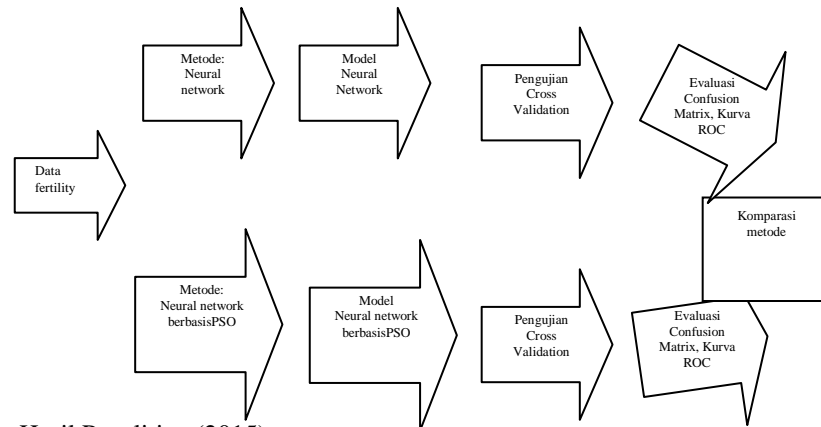
- (3) *Data size reduction and dicrtization*, untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informatif.

Dikarenakan data yang diperoleh dari repository sehingga data yang diperoleh telah melalui tahapan pengolahan data seperti diatas. Berikut tranformasi data yang dilakukan oleh pemilik data. Atribut musim terdapat empat kategori musim dingin, semi, panas dan hujan dikonversi menjadi -1,-0,0.33,1. Usia pada saat analisis diisi dari 18-36 dikonversi dari 0-1. Penyakit pada masa kecil 1 untuk ya yang mempunyai penyakit dan 2 untuk tidak punya penyakit

c. Metode yang diusulkan

Dalam penelitian ini akan dilakukan perbandingan nilai akurasi antara metode neural networkn dengan model yang dihasilkan oleh metode neural network yang ditingkatkan akurasinya dengan menggunakan *Particle Swan optimized(PSO)*.

Berikut ilustrasi penggunaan metode yang diusulkan dalam penelitian:



Sumber: Hasil Penelitian (2015)

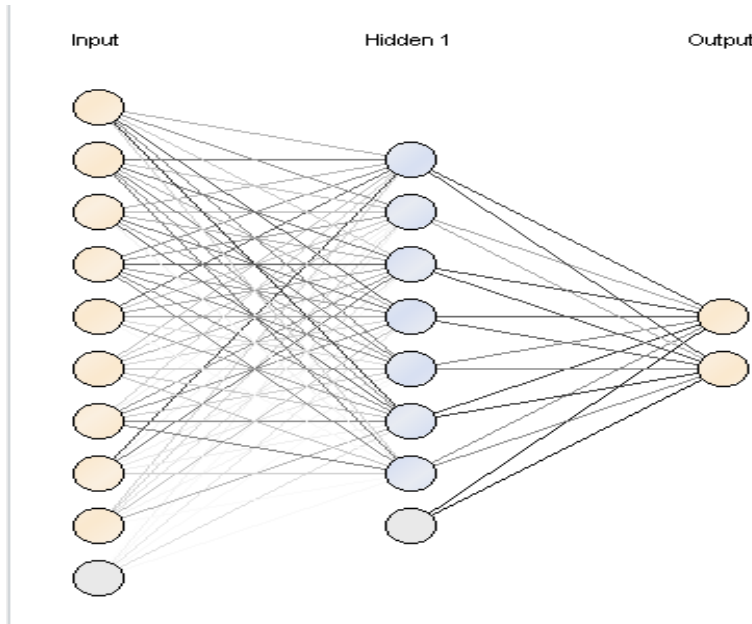
Gambar 2. Ilustrasi Penggunaan Metode yang Diusulkan

IV. PEMBAHASAN

4.1 Ekperimen dan Pengujian Metode

Dalam penelitian metode yang digunakan adalah metode Neural network dan tool yang

digunakan untuk mengolah data yang ada adalah RapidMiner. Dari hasil pengolahan tersebut diperoleh hasil seperti berikut:



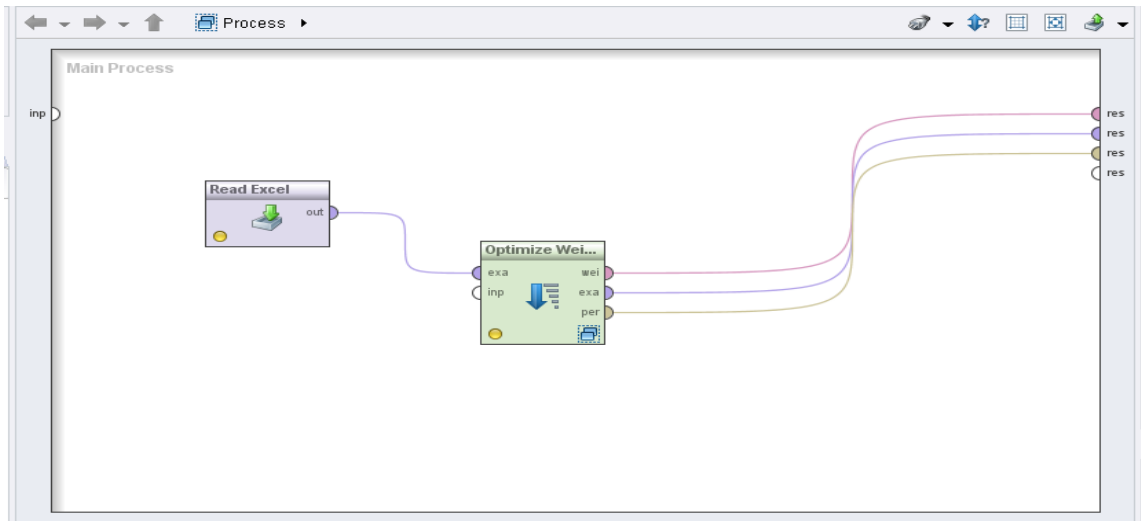
Sumber : RapidMiner

Gambar 3. Hasil Neural Network

Neural net yang dihasilkan adalah *multilayer proceptron*, pada pengolahan data neural net terdapat tiga layer. Pada lapisan pertama adalah *layer input*, lapisan kedua *hidden layer* dan lapisan layer ketiga adalah *layer output*. Pada lapisan *input layer* terdapat sembilan *node* dengan satu *node bias*, pada *hidden layer* terdapat tujuh *node hidden*

layer dengan satu *node bias*, dan pada *layer output* terdapat dua *node* yaitu ya dan tidak.

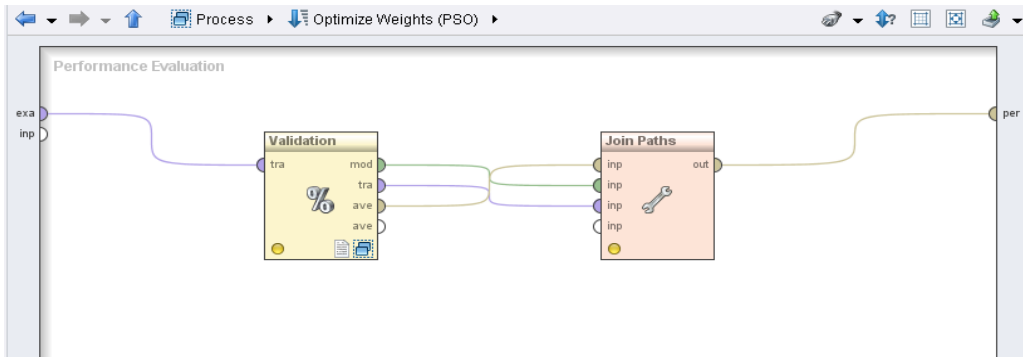
Berikut adalah langkah-langkah ekperimen yang dilakukan dengan menggunakan tool RapidMiner untuk pengolahan data dengan metode neural net yang ditingkatkan akurasinya menggunakan PSO:



Sumber: Rapidminer

Gambar 4. langkah pertama ekperimen

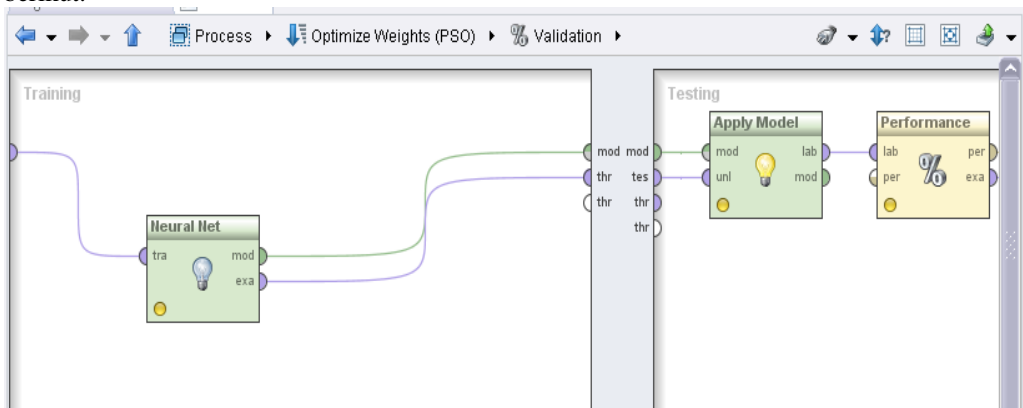
Double klik model Optimized maka akan muncul modul seperti berikut:



Sumber: Rapidminer

Gambar 5. Langkah kedua eksperimen

Kemudian didalam modul *validation* pastikan model yang digunakan adalah neural net, seperti berikut:



Sumber: Pengolahan data dengan Rapidminer

Gambar 6. Langkah ketiga eksperimen

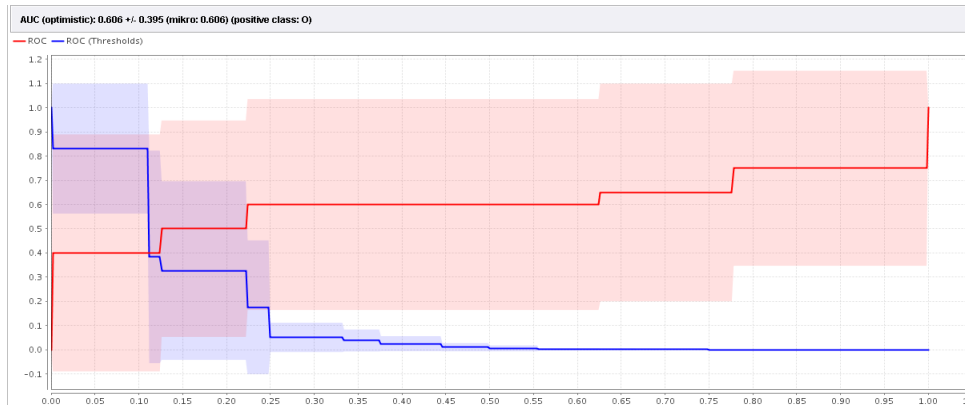
4.2 Pengujian Hasil

Dalam pengujian hasil eksperimen digunakan penilaian Ckurai menggunakan kurva ROC dan *Confussion Matrix*

Berikut adalah confusion matrix dan kurva ROC untuk metode neural network saja tanpa dioptimalkan dengan *particle swam optimized (PSO)*:

accuracy: 88.00% +/- 7.48% (mikro: 88.00%)			
	true N	true O	class precision
pred. N	84	8	91.30%
pred. O	4	4	50.00%
class recall	95.45%	33.33%	

Gambar 7. Confusion Matrix Metode Neural Network



Sumber: hasil pengolahan data dengan Rapidminer

Gambar 8. Kurva ROC Metode Neural network

Berikut adalah kurva ROC dan confusion matrix dari pengolahan data menggunakan neural network

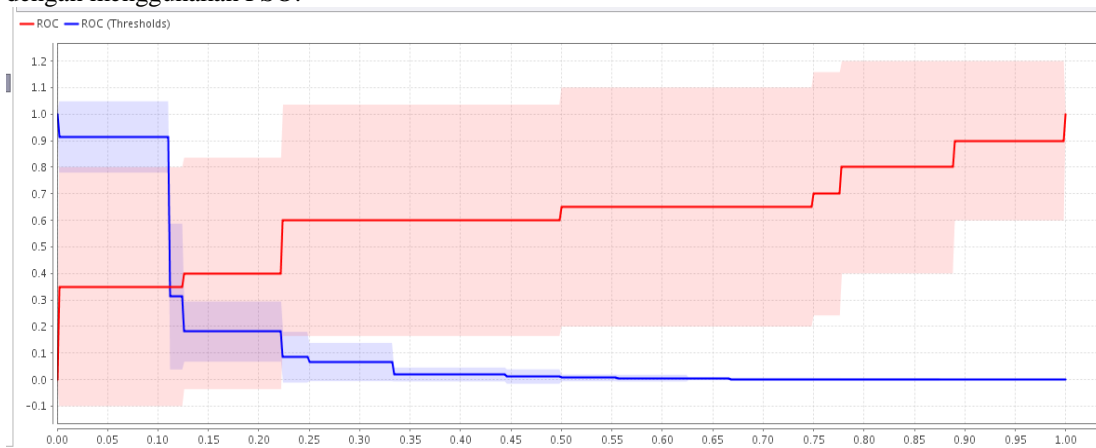
yang ditingkatkan kinerjanya dengan menggunakan PSO, berikut adalah gambarnya:

accuracy: 92.00% +/- 6.00% (mikro: 92.00%)			
	true N	true O	class precision
pred. N	88	8	91.67%
pred. O	0	4	100.00%
class recall	100.00%	33.33%	

Sumber: Hasil pengolahan data dengan Rapidminer

Gambar 9. Confusion Matrix Neural Network berbasis PSO

Berikut tampilan untuk kurva ROC dari metode neural network yang ditingkatkan kinerjanya dengan menggunakan PSO:



Sumber: Hasil Pengolahan data dengan Rapidminer

Gambar 10. Kurva ROC neural network berbasis PSO

V. PENUTUP

Dari penelitian yang dilakukan dengan menggunakan data fertility yang diperoleh melalui web uci repository, kemudian data diolah menggunakan metode neural network dan metode neural network yang ditingkatkan kinerjanya menggunakan metode optimasi PSO. Maka diperoleh hasil akurasi yang lebih baik yaitu dengan

menggunakan neural network berbasis PSO yaitu nilai akurasinya sebesar 92%, jauh lebih tinggi dibandingkan jika hanya menggunakan metode neural network saja yaitu sebesar 88%. Sedangkan untuk. Sehingga dapat disimpulkan bahwa dengan melakukan peningkatan kinerja metode dengan menggunakan metode optimasi particle swam optimized (PSO) lebih baik kinerjanya daripada

hanya menggunakan metode neural networknya saja.

DAFTAR PUSTAKA

- Durairaj & Sathyavathi. (2013). Applying Rough Set Theory for Medical Informatics Data Analysis. *International Journal of Scientific Research in Computer Science and Engineering*, Vol-1, Issue-5 ISSN: 2320-7639
- Gilera et. (2012). Predicting seminal quality with artificial intelligence methods, *Expert Systems with Applications* 39 (2012) 12564–12573
- Han, J., & Kamber, M. (2007). *Data Mining Concepts and Techniques*. San Fransisco: Mofgan Kaufan Publisher
- Han, J., & Kamber, M. (2007). *Data Mining Concepts and Techniques*. San Fransisco: Mofgan Kaufan Publisher
- Inhorn. (2002). Global infertility and the globalization of new reproductive technologies: illustrations from Egypt, *Social Science & Medicine* 56 (2003) 1837–1851, Elsevier Science Ltd. All rights reserved.
- Kahki et. (2013). A model based on Bayesian Network for prediction of IVF Success Rate, 7thSASTech 2013, Iran, Bandar-Abbas. 7-8 March
- Kolettis, P. N. 2003. Evaluation of the subfertile man. *American Family Physician*, 67(10), 2165–2172
- Larose, D. T. (2005). *Discovering Knowledge in Databases*. New Jersey: John Willey & Sons Inc.
- Myatt, Glenn J. *Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining*. New Jersey: John Wiley & Sons, Inc.
- Mileski et. (2013). Comparison of Artificial Neural Networks and Logistic Regression Analysis in Pregnancy Prediction Using the In Vitro Fertilization Treatment, *STUDIES IN LOGIC, GRAMMAR AND RHETORIC* 35 (48), ISBN 978–83–7431–392–6 ISSN 0860-150X DOI: 10.2478/slgr-2013-0033
- Maimon. (2005). *Data Mining and Discovery Handbook Second Edition*, ISBN 978-0-387-09822-7 e-ISBN 978-0-387-09823-4, Springer, New York
- Sivanandam. (2006). *Introduction to Data Mining and its Applications*, Springer, New York.
- Shukla, A., Tiwari, R., & Kala, R. (2010). *Real Life Application of Soft Computing*. Taylor and Francis Groups, LLC.
- Vecellis. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*, Ltd. ISBN: 978-0-470-51138-1, United Kingdom, John Wiley & Sons