

OPTIMASI NEURAL NETWORK MENGGUNAKAN GENETIC ALGORITHM UNTUK PREDIKSI PENYAKIT DIABETES

Hilda Amalia

Program Studi Manajemen Informatika
Akademi Manajemen Informatika dan Komputer Bina Sarana Informatika
AMIK BSI JAKARTA
JL.RS. Fatmawati no.24 Pondok Labu Jakarta Selatan
hilda.ham@bsi.ac.id

ABSTRACT

Diabetes is a dangerous disease that can lead to complications in the condition of the sufferer, leading to death . People suffering from this disease are unaware that they suffer from diabetes. early detection is required in order to avoid the risk of more severe complications and risk of death. Today many data mining applications are applied to solve problems in the medical world. The ability of data mining is very important that one of them is generating the prediction and classification, the ability to make a prediction and classification makes data mining became popular in the world of health. Results obtained from the processing of data mining can be used as new knowledge to predict adverse health events in the future that can be overcome. In this research will be to improve the accuracy of the method optimization value neural network using genetic algorithm optimization method. The value of diabetes dataset processing accuracy by using a neural network is 74.46% while the value of accuracy by using a neural network optimization algorithm genetic is 77.10%.

Kata Kunci: *Data Mining, diabetes, Neural Network, Genetic Algorithm*

I. PENDAHULUAN

Data mining menurut (Kolce dan Fansher, 2012) merupakan suatu teknik penemuan pola dari sekumpulan data, sekumpulan data ini biasanya hanya tersimpan dalam gudang data. Pada dasarnya data mining berarti melakukan ekstraksi atau "pertambangan" pengetahuan dari sejumlah besar data atau database. Data mining dapat digunakan dalam berbagai aspek kehidupan yaitu aspek ekonomi yaitu bisnis dan keuangan, aspek kesehatan dan bagi aspek ilmu pengetahuan yaitu bidang astronomi, geografis, biologi dan masih banyak lagi, salah satunya dapat digunakan dalam bidang kesehatan. Dalam bidang kesehatan Data mining telah banyak digunakan. Hal ini dikarenakan kemampuan dari data mining yang mampu menemukan pola atau aturan-aturan baru dari data-data yang selama ini hanya disimpan oleh pihak rumah sakit dan tercatat sebagai catatan kesehatan pasien. Data-data yang hanya tersimpan itu diolah dengan metode data mining untuk menghasilkan ilmu pengetahuan dalam bentuk prediksi atau penemuan pola penyakit. Prediksi dan penemuan pola yang dihasilkan dari metode data mining telah banyak membantu kehidupan manusia yaitu sebagai alat penunjang pengambilan keputusan. Sehingga bisa membantu untuk melakukan

penanganan dan deteksi dini terhadap berbagai penyakit-penyakit berbahaya salah satunya terhadap penyakit diabetes.

Menurut (Diabetes Federation International, 2012) Diabetes mellitus adalah kondisi serius yang berpotensi menghancurkan komplikasi yang mempengaruhi semua kelompok umur di seluruh dunia . Pada tahun 1985, sekitar 30 juta orang di seluruh dunia didiagnosis dengan diabetes pada tahun 2000, angka itu meningkat menjadi lebih dari 150 juta dan di 2012, Diabetes Federation International (IDF) memperkirakan bahwa 371 juta orang menderita diabetes Angka itu diproyeksikan meningkat menjadi 552 juta (1 dari 10 orang dewasa) pada tahun 2030, yang setara dengan 3 kasus baru per detik. Faktanya 378 Miliar orang menderita diabetes dan 46, 3% diantara tidak mengetahui menderita diabetes. Menurut (Tjokroprawiro, 2006), Risiko penyakit yang terjadi bagi penderita diabetes melitus jika dibandingkan dengan penderita yang tidak menderita diabetes melitus adalah dua kali lebih mudah 184 mengalami stroke, dua puluh lima kali lebih mudah mengalami buta, dua kali lebih mudah mengalami PJK (Penyakit Jantung Koroner), tujuh belas kali lebih mudah mengalami gagal ginjal kronik, dan lima kali lebih mudah mengalami selulitis.

Untuk itu diperlukan pendektisian dini terhadap kemungkinan seorang menderita diabetes atau tidak sangat diperlukan. Diagonis atau deteksi dini dapat menghindari penderita dari kemungkinan terburuk resiko penyakit diabetes yang dapat berakibat pada kematian. Jumlah kematian akibat diabetes menurut World Health Organization (WHO) dalam Perkumpulan Endokrinologi Indonesia (2006) meramalkan peningkatan jumlah penderita Diabetes Melitus (DM) pada tahun mendatang, khususnya di Indonesia yaitu WHO meramalkan kenaikan jumlah penderita dari 8,4 juta pada tahun 2000 menjadi \pm 21,3 juta pada tahun 2003. Hal ini juga dikemukakan oleh Pusat data dan informasi Kemenkes RI (2012) yaitu lebih dari dua per tiga (70%) populasi global akan meninggal akibat penyakit tidak menular seperti kanker, penyakit jantung, stroke dan diabetes.

Menurut (Joseph L. Breaul, 2012) Penyakit diabetes dapat menghasilkan komplikasi yang mengerikan dari kebutaan, gagal ginjal, amputasi, dan kematian kardiovaskular dini, sehingga dokter dan regulator ingin tahu bagaimana meningkatkan hasil sebanyak mungkin. Untuk itu penggunaan metode data mining sebagai alat deteksi dini untuk penyakit diabetes adalah hal yang sangat tepat hal ini disebabkan data diabetes yang ada pada saat ini sudah menjadi gunung data, data yang diperoleh dari makin banyaknya penderita penyakit diabetes dan gudang data penyakit diabetes ini akan terus bertambah seiring dengan ketidakpedulian masyarakat untuk mencegah penyakit diabetes mellitus dan peningkatan populasi penduduk dunia serta gaya hidup modern saat ini.

Penelitian sebelumnya mengenai data diabetes ini telah dilakukan sebelumnya yaitu dengan menggunakan metode SVM pada oleh Sanakal DAN Jayakumari. Penelitian mengenai data diabetes juga dilakukan oleh L Josep dengan metode Rought Set pada tahun. Penelitian data diabetes juga dilakukan oleh kartihekeyani dkk dengan melakukan komparasi metode data mining yaitu SVM, K-NN, C45 dan lainnya metode lainnya. Dalam penelitian ini akan dilakukan pengolahan data diabetes dengan menggunakan metode *neural network* yang ditingkatkan kinerja metode nya dengan menggunakan metode optimas *genetic algorithm*(GA).

II. KAJIAN LITERATUR

2.1. Data Mining

Menurut (Sivandam, 2006) Data mining merupakan teknik yang berorientasikan

kepada penemuan hubungan yang tersembunyi dari suatu kumpulan besar data yang dapat digunakan untuk melakukan prediksi pada masa yang akan datang.

Dalam (Guronescu, 2011) menjabarkan mengenai akar dari cabang ilmu data mining. Data mining merupakan suatu cabang ilmu baru yang diturunkan dari cabang ilmu sebelumnya yaitu ilmu statistic, ilmu artificial intelegent (AI) dan ilmu sistem database.

Selama ini banyak yang menyamakan antara data mining dan *knowledge discovey of Database* (KDD) dalam (Han dan Kamber, 2007) menyebutkan bahwa Data mining merupakan salah satu tahapan yang harus ditempuh untuk mendapatkan *knowledge discovery of database* (KDD). Berikut adalah tahapan-tahapan yang ditempuh untuk mendapatkan pengetahuan atau *knowledge discovery of Database* (KDD):

1. *Data Cleaning* dilakukan untuk menghilangkan noise data dan data-data yang tidak konsisten
2. *Data Integration* dilakukan untuk mengintegrasikan beberapa data yang ada
3. *Data Selection* dilakukan untuk menyeleksi data-data yang diperlukan untuk penelitian
4. *Data Transformasi* dilakukan untuk mengubah data sehingga dapat digunakan
5. *Data mining*, tahapan paling penting menghasilkan model
6. *Pattern evaluation*, dilakukan untuk mengidentifikasi model yang dihasilkan pada tahapan data mining
7. *Knowledge presentation* dilakukan untuk memvisualisasikan pengetahuan kedalam bahasa yang mampu dipahami oleh pengguna.

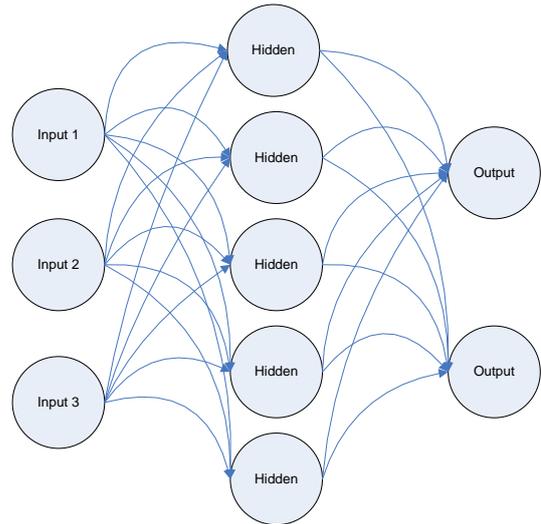
Menurut (Guronescu, 2011) data mining mempunyai dua tugas utama yaitu prediksi dan deskripsi. Hal-hal yang dapat dilakukan data mining dalam melakukan tugas prediksinya adalah klasifikasi, regresi dan melakukan pendekteksian, tujuan utama dari tugas prediksi ini adalah menghasilkan variabel yang dapat memprediksi. Sedangkan hal-hal yang dapat dilakukan data mining dalam tugasnya sebagai alat deksripsi yaitu clustering, penemuan aturan asosiasi dan penemuan pola terurut, tujuan dari melakukan tugas deskripsi adalah mengidentifikasi pola sehingga mudah dipahami oleh pengguna.

2.2. Neural Network

Menurut (Guronescu, 2011) *Artificial Neural Network* (ANN) pertama kali dihasilkan oleh seorang ahli neuron yaitu Warren McCulloch dan ahli logika Walter Pitts pada tahun 1943. Walaupun mereka memberikan teori dan asumsi berdasarkan model-model komputasi, sedikit dari mereka mengimplementasikan pada saat kondisi teknologi tersedia pada waktu dimana mereka belum dapat berbuat lebih.

Neural network telah menjadi teknik data mining yang populer dan dapat menyelesaikan berbagai bidang baik bisnis, industri dan ilmu pengetahuan. Saat ini neural network merupakan suatu standart tool data mining yang paling banyak digunakan. Menurut (Maimon dan Rokach, 2010) *Neural network* mampu melakukan klasifikasi pola, analisa time series, prediksi dan clustering. Secara spesifik neural network diibaratkan sebagai jaringan syaraf manusia yaitu sell neuron yang ada di otak manusia. menurut (Myatt, 2007) neural network merupakan mathematical model yang dapat membuat prediksi berdasarkan series dari input deskripsi variabel. Seperti model-model melakukan prediksi. *A neural network comprises of a series of independent processors or nodes. These nodes are connected to other nodes and are organized into a series of layers*

Pada perkembangan terdapat beberapa model dari neural network yang umum digunakan yaitu (SLP) dan *multi-layer perceptron* (MLP). SLP merupakan konsep dasar neuron, SLP memiliki satu node output, satu input terhubung dengan satu *hidden layer*. MLP memiliki beberapa layer dari unit komputasi atau neuron, dalam MLP output yang dihasilkan dalam beberapa node output. MLP menggunakan beberapa teknik salah satu yang terkenal adalah *back-propagation algorithm* (BP). Pada prinsipnya, nilai output dibandingkan dengan nilai sebenarnya dan error dikomputasi berdasarkan fungsi error E, berdasarkan hasil satu langkah kebelakang melalui jaringan untuk memperbaharui nilai bobot dalam rangka meminimalkan terjadinya error (Guronescu, 2011).



Sumber: Maimon & Rokach (2010)

Gambar 1. Arsitektur Neural Network

Langkah pembelajaran dalam algoritma *backpropagation* menurut (Myatt, 2010) adalah sebagai berikut :

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 dan 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)

- 3 Berdasarkan input dari langkah dua, selanjutnya membangkitkan output untuk simpul menggunakan fungsi aktifasi sigmoid:

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

- 4 Hitung Nilai Error antara nilai yang diprediksi dengan nilai yang sesungguhnya menggunakan rumus :

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

Keterangan:

Output_j = Output aktual dari simpul j

Target_j = Nilai target yang sudah diketahui pada data training

- 5 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 aktual dari simpul j

Error_k = error simpul k

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

- 6 Nilai Error yang dihasilkan dari langkah sebelumnya digunakan untuk memperbarui bobot relasi menggunakan rumus :

$$W_{ij} = W_{ij} + l . Error_j . Output_j$$

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 i

2.3. Genetic Algorithm

Genetic algorithm (GA) menurut (Guronescu, 2011) berdasarkan teori evolusi modern dikembangkan oleh Darwin. GA mewakili indentifikasi pendekatan optimasi dan masalah pencarian, menjadi kelas populer dari *evolutionary algorithm* (EA)

Algoritma Genetika biasa digunakan untuk klasifikasi dan juga masalah optimisasi. Dalam data mining, metode ini juga digunakan untuk melakukan evaluasi terhadap nilai fitness pada sebuah algoritma. Menurut (Whitcombe, 2006) Beberapa hal yang harus dilakukan dalam algoritma genetika adalah

- 1 Menentukan Populasi Awal.
Diperlukan Populasi awal sebelum optimasi dilakukan. Populasi awal dilakukan seperti melakukan pendeklarasian dalam sebuah program
- 2 Evaluasi Nilai Fitness
Nilai Fitness dievaluasi dengan tujuan agar setiap kromosom memiliki nilai baik atau tidak. Kemudian semua nilai fitness

ditentukan probabilitasnya masing – masing. Dari hasil probabilitas tertinggi, dihasilkan bahwa kromosom 1 mempunyai nilai fitness paling tinggi. Maka kromosom 1 juga mempunyai kesempatan paling besar dalam proses seleksi selanjutnya dengan *Roulette Wheel*.

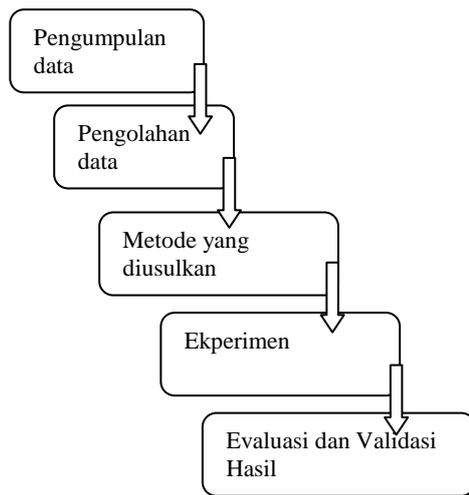
- 3 Seleksi Kromosom Induk
Proses dimana dipilih kromosom yang akan dijadikan kromosom induk dalam populasi yang akan dihitung. Proses seleksi kromosom yang digunakan dengan *Roulette Wheel*.

- 4 Melakukan Crossover (Perkawinan Silang)
Dalam *crossover* juga melanjutkan ke langkah selanjutnya menggunakan bilangan acak R antara 0 sampai 1. Setelah melakukan pemilihan *parent* (Induk), proses selanjutnya adalah menentukan posisi *crossover*. Setelah didapatkan posisi *crossover* maka kromosom *parent* (Induk) akan dipotong mulai gen posisi *crossover* kemudian potongan gen tersebut saling ditukarkan antar *parent* (Induk).

- 5 Mutasi Kromosom
Jumlah kromosom yang mengalami mutasi dalam satu populasi ditentukan oleh persentase *p mutation*. Proses mutasi dilakukan dengan cara mengganti satu gen yang terpilih secara acak dengan suatu nilai baru yang didapat secara acak. Kromosom tersebut kemudian diuji bila belum sesuai tujuan, maka populasi ini belum memiliki kromosom yang ingin dicapai. Kromosom-kromosom pada populasi ini akan mengalami proses yang sama seperti generasi sebelumnya yaitu proses evaluasi, seleksi, *crossover* dan mutasi yang kemudian akan menghasilkan Kromosom-Kromosom baru untuk generasi yang selanjutnya. Proses ini akan berulang sampai sejumlah generasi yang telah ditetapkan sebelumnya

III. METODE PENELITIAN

Dalam penelitian ini data yang digunakan adalah data sekunder yang diambil dari *uci repository* sehingga dalam tahapan pengumpulan data dan pengolahan data awal telah diperoleh data dengan jumlah yang mewakili dataset diabetes. Dataset tersebut kemudian diolah dengan metode yang telah diusulkan dan dilakukan eksperimen terhadap dataset diabetes selanjutnya setelah diperoleh model dari hasil pengolahan dataset maka dilakukan evaluasi dan validasi dari hasil yang diperoleh dari eksperimen yang telah dilakukan. Penjabaran mengenai tahapan penelitian yang dilakukan disajikan dalam gambar berikut ini:



Sumber: Hasil Penelitian (2015)

Gambar 2 Tahapan Penelitian

- a. pengumpulan data
data yang digunakan dalam penelitian ini merupakan data sekunder yang telah tersedia yaitu dari sebuah web penyedia dataset untuk penelitian. Data sekunder didapat dari laman <https://archive.ics.uci.edu/ml/dataset> s Record yang ada sebanyak 768 record dengan jumlah atribut sebanyak 9 atribut dengan satu kelas didalamnya. Atribut yang digunakan yaitu berapa kali hamil, konsentrasi glukosa, serum insulin, massa tubuh, diabetes sisilah, umur dan satu label yaitu hasil ya diabetes atau tidak diabetes. Berikut tabel atribut yang digunakan:

Tabel 1 Atribut Yang Digunakan	
Nama Atribut	Nilai
Berapa Kali Hamil	Integer(0,1,2,3...etc

Hamil)
Konsentrat glukosa	Integer(10,...etc)
Tekanan darah	Integer
Lipatan	Integer
SerumInsulin	Integer
Diabetes sisilahFungs	Real
Masa Tubuh	Integer
Umur	Integer(21,22,...etc)
Kelas	binominal(Ya, Tidak)

Sumber: Hasil Penelitian (2015)

Berikut contoh dataset yang digunakan dalam penelitian ini:

Tabel 2 Dataset yang digunakan

Berapa Kali Hamil	Konsentrat Glukosa	Tekanand arah	Lipatanku lit	Serum Insulin	Masssa Tubuh	Diabetes Sisilah Fungsi	Umur	Kelas
6	14	148	35	0	33,6	0,627	50	Ya
1	85	66	29	0	26,6	0,351	31	Tidak
8	18	64	0	0	23,3	0,672	32	Ya
1	89	66	23	94	28,1	0,167	21	Tidak
0	13	40	35	168	43,1	2,288	33	Ya
5	11	74	0	0	25,6	0,201	30	Tidak
3	78	50	32	88	31,0	0,248	26	Ya
10	115	115	0	0	35,3	0,134	29	Tidak
2	19	70	45	543	30,5	0,158	53	Ya
8	12	96	0	0	0,0	0,232	54	Ya
4	11	92	0	0	37,6	0,191	30	Tidak
10	168	74	0	0	38,0	0,537	34	Ya
10	139	80	0	0	27,1	1,441	57	Tidak

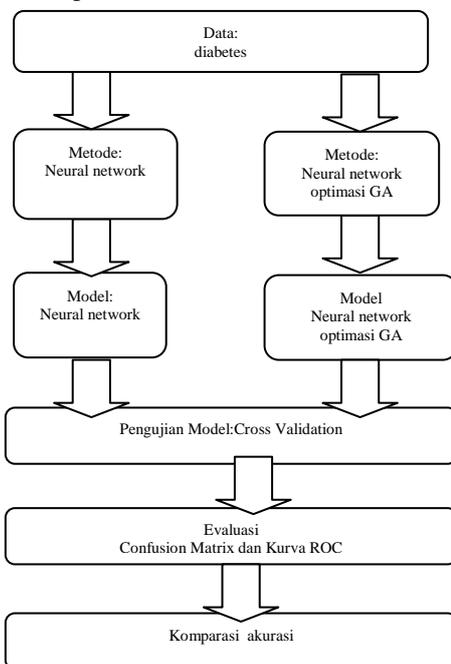
Sumber: Hasil Penelitian (2015)

- b. Pengolahan data awal
Kualitas dari input sangat mempengaruhi keakuratan dari suatu output. Untuk itu perlu dilakukan beberapa hal untuk memastikan bahwa data input yang akan digunakan dalam penelitian ini adalah data input yang baik. Tahapan ini dilakukan dengan beberapa cara yaitu *data validation, data integration dan transformation, data reduction and discretization*. Data validation digunakan untuk tujuan menghasilkan data yang bebas dari noise, tidak lengkap, tidak konsisten. Noise yaitu data yang nilainya tidak wajar, dalam penelitian ini dilakukan penghapusan record yang bernilai tidak lengkap atau bernilai nol. Data integration dan transformation digunakan untuk menyatukan dan merubah susunan tapi bukan merubah isi dari data. Hal ini dilakukan dengan tujuan menghilangkan atribut yang tidak diperlukan dalam

penelitian yang sedang dilakukan. *Data reduction and discretization* digunakan untuk mengurangi jumlah record atau atribut untuk memperoleh data set dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informatif dan lebih mewakili dataset (Vercalis, 2009).

c. Metode Yang Diusulkan

Metode yang diusulkan menggambarkan tahapan atau langkah-langkah yang ditempuh untuk mendapatkan model dari pengolahan dataset diabetes. Dalam penelitian ini akan dilakukan pengolahan dataset diabetes dengan menggunakan metode *neural network* kemudian ditingkatkan dengan menggunakan metode optimasi *genetic algorithm* (GA). Setelah itu diperoleh model dari masing-masing metode yakni model yang hanya menggunakan metode *neural network* dan metode yang dihasilkan dari penggunaan *neural network* dan *genetic algorithm* kemudian dilakukan pengujian model dengan menggunakan *cross validation*, kemudian dilakukan evaluasi model dengan menggunakan *confusion matrix* dan kurva ROC. Terakhir dilakukan perbandingan akurasi dan diperoleh nilai akurasi yang terbaik. Berikut model yang diusulkan dalam penelitian ini:



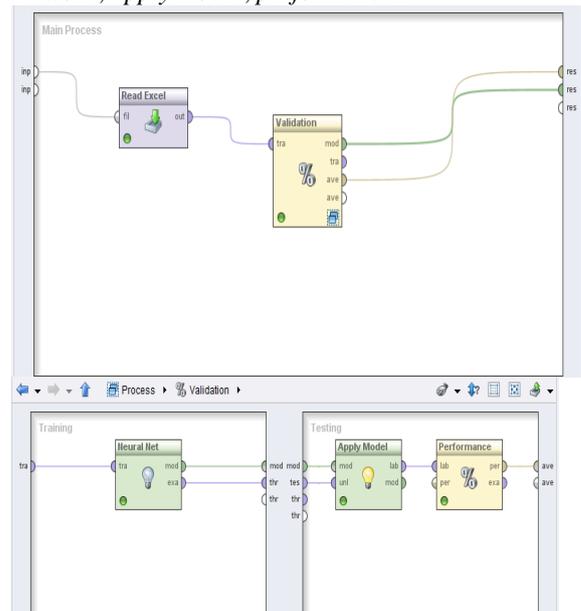
Sumber: Hasil Penelitian (2015)

Gambar 3. Model yang diusulkan

IV. PEMBAHASAN

4.1 Ekperimen dan Pengujian Metode

Dalam penelitian ini akan dilakukan perbandingan kinerja metode *neural network* dengan kinerja metode optimasi *genetic algorithm* menggunakan metode *neural network*. Pengolahan data pada penelitian ini menggunakan tools RapidMiner. Berikut ilustrasi penggunaan RapidMiner untuk metode *neural network*, modul yang digunakan adalah modul *read excel*, modul ini digunakan untuk memasukan dataset yang dimiliki dalam format *Ms. Excel* dihubungkan dengan modul *validation*, didalam modul *validation* terdapat modul metode *neural network, apply model, performance*::



Sumber:Hasil Penelitian (2015)

Gambar 4. Tampilan pengolahan data menggunakan neural network

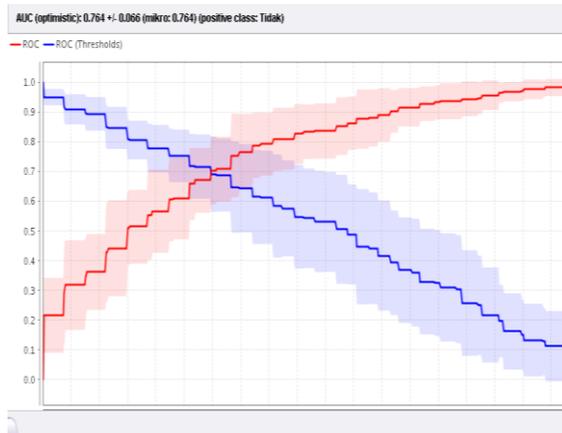
Dari hasil pengolahan data diabetes dengan menggunakan rapidminer diperoleh nilai akurasi 74,46%. Dengan nilai AUC 0,764. Berikut tampilan confusion matrix pengolahan dataset diabetes dengan metode *neural network*:

Tabel 3. Confusion Matrix neural network

accuracy: 74.46% +/- 4.83% (mikro: 74.46%)			
	true Ya	true Tidak	class precision
pred. Ya	139	67	67.48%
pred. Tidak	129	433	77.95%
recall recall	61.87%	66.60%	

Sumber:Hasil Penelitian (2015)

Berikut gambar kurva ROC dengan nilai AUC dari pengolahan dataset diabetes dengan menggunakan metode nueral network:



Sumber: RapidMiner

Gambar 5. Kurva ROC yang dihasilkan

Menurut Guronescu(2011:319) Klasifikasi akurasi merupakan suatu alat pengukuran mengenai seberapa baik suatu klasifikasi melakukan klasifikasi objectnya. *Confusion matrix* merupakan suatu alat pengukuran untuk melakukan klasifikasi berdasarkan object yang tepat dan object yang tidak tepat. *Confusion matrix* memberikan daftar mengenai klsifikasi yang salah. Berikut tampilan tabel confusion matrix

Tabel 4. Confusin Matrix

CLASSIFICATION	PREDICTED CLASS		
	Class = YES	Class = No	
OBSERVED CLASS	Class = YES	a (true positive-TP)	b (false negative)
	Class = NO	c (false positive-FP)	d (true negative)

Sumber: Guronescu (2011)

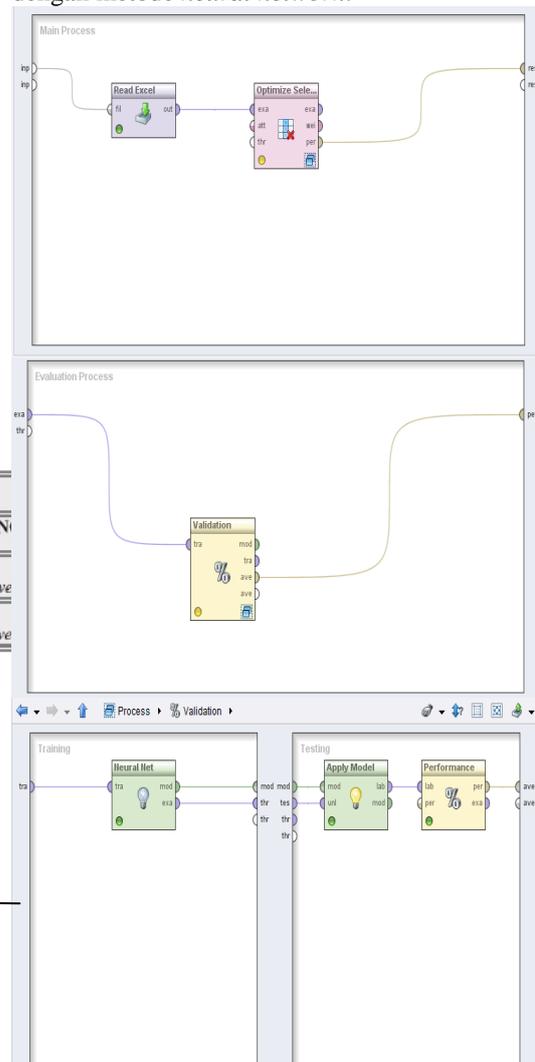
Menurut (Guronescu, 2011) Formula untuk melakukan perhitungan predited class:

$$Accuracy = \frac{a+d}{a+b+c+d} = \frac{TP+TN}{TP+TN+FP+FN}$$

Menurut (Guronescu, 2011) Kurva ROC atau *Reveicer Operating Characteristic Curve*, digunakan untuk menilai hasil dari prediksi (peramalan) yang telah dilakukan. ROC adalah suatu teknik untuk

memvisualisasi, organisasi dan klasifikasi terpilih berdasarkan kinerjanya.

Langkah berikutnya adalah melakukan pengolahan data dengan optimasi genetic algorithm pada metode *neural network*, modul-modul yang digunakan dalam optimasi ini yaitu modul *read excel* yaitu tempat data training diletakan, modul *optimize selection* merupakan modul pengolahan data dengan *genetic algorithm*, didalam modul *optimize selection* terdapat modul *validation* yang berfungsi untuk melakukan pengujian terhadap model yang akan dihasilkan oleh metode, dan didalam modul *validation* terdapat modul *neural network*, *apply model* dan *performace*. Berikut adalah ilustrasi dari pengolahan data optimasi genetic algoritmn dengan metode *neural network*:



Sumber:RapidMiner

Gambar 6. Pengolahan Data Optimasi Genetic Algorithm Dengan Menggunakan Neural Network

Pengolahan data *training* dilakukan sebanyak enam kali yaitu pengolahan data *training* diabetes dengan satu *hidden layer* sampai dengan enam *hidden layer*, cara ini digunakan pada masing-masing metode, yaitu yang dengan optimasi dan tanpa optimasi. Hal ini dilakukan untuk mengetahui nilai akurasi yang terbaik dari masing-masing *hidden layer*. Berikut tabel hasil pengolahan data *training* diabetes melalui *rapidminer*:

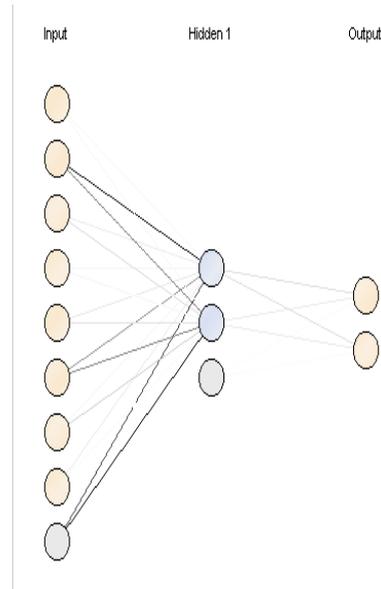
Tabel 5. Perbandingan Nilai Akurasi

momentum	Learning rate	Hidden layer	Accuracy neural network	Neural network dan GA
0,3	2	2	72,66%	77,10%
0,3	2	3	71,09%	76,43%
0,3	2	4	73,96%	76,54%
0,3	2	5	74,46%	75,78%
0,3	2	6	73,05%	76,05%

Sumber: Hasil Penelitian (2015)

Dari tabel diatas dapat dilihat bahwa nilai akurasi terbaik untuk metode *neural network* diperoleh dari nilai momentum 0,3 *learning rate* 2 dan *hidden layer* 5, sementara untuk optimasi *neural network* dan *genetic algorithm* nilai akurasi terbaik diperoleh dari nilai momentum 0,3 *learning rate* 2 dan *hidden layer* 2.

Berikut jaringan saraf tiruan yang terbentuk dari pengolaha data *training* diabetes dengan menggunakan dua *hidden layer*:



Sumber: Hasil Penelitian (2015)

Gambar 7. Hasil Jaringan Saraf Tiruan Dengan Dua Hidden Layer

Neural Network selain menghasilkan arsitektur jaringan syarat tiruan untuk dataset yang sedang diteliti juga menghasilkan nilai bobot untuk masing-masing node yang ada. Dibawah ini nilai node yang dihasilkan dari pengolahan data *training* diabetes dengan dua *hidden layer*:

Hidden 1

=====

Node 1 (Sigmoid)

- Berapa Kali Hamil: 0.839
- Konsentrasi Glukosa: 30.426
- Tekanandarah: -2.387
- Lipatankulit: -1.216
- Serum Insulin: -2.967
- Massa Tubuh: 13.914
- Diabetes Silsilah Fungsi: -1.151
- Umur: -3.134
- Bias: 27.973

Node 2 (Sigmoid)

- Berapa Kali Hamil: -1.074
- Konsentrasi Glukosa: -17.552
- Tekanandarah: -4.382
- Lipatankulit: -1.138
- Serum Insulin: -4.036
- Massa Tubuh: -16.088
- Diabetes Silsilah Fungsi: -5.708
- Umur: -0.511

Bias: -36.624

Output
=====

Class 'Ya' (Sigmoid)

Node 1: 4.188
Node 2: -2.573
Threshold: 0.301

Class 'Tidak' (Sigmoid)

Node 1: -4.188
Node 2: 2.573
Threshold: -0.30

Pada pengolahan data training diabetes dengan optimasi *genetic algorithm* diperoleh atribut-atribut yang berperan penting dan yang tidak berperan bagi penelitian ini. Atribut yang tidak berperan dapat dihilangkan pada penelitian selanjutnya, berikut tabel nilai atribut yang dihasilkan dari metode optimasi *genetic algorithm*:

Tabel 4. Tabel Atribut hasil pengolahan genetic algorithm

atribute	weight
Berapa Kali Hamil	0
Konsentrasi Glukosa	1
Tekanan Darah	1
Lipatan Kulit	0
Serum Insulin	1
Massa Tubuh	1
Diabetes Silsilah Fungsi	1
Umur	1

Sumber: Hasil Penelitian (2015)

Dari hasil yang diperoleh dari pengolahan dataset diabetes dengan menggunakan tools rapidminer untuk opetimasi genetic algorithm menggunakan metode neural network menghasilkan atribut apa saja yang berperan penting dan atribut apa saja yang tidak berperan penting dalam penelitian pengolahan dataset diabetes. Sehingga dapat diketahui bahwa terdapat enam atribut yang bernilai 1 dan dua atribut yang bernilai 0, atribut yang bernilai 1 merupakan atribut yang berperan penting terhadap penelitian ini dan atribut yang tidak penting untuk digunakan dalam penelitian ini bernilai 0.

V. KESIMPULAN

Dari hasil ekperimen dan pembahasan diperoleh nilai akurasi terbaik untuk pengolahan data *training* diabetes dengan menggunakan hanya metode *neural network* diperoleh hasil akurasi terbaik yaitu 74, 46%. Nilai terbaik untuk metode *neural network* diperoleh dari hasil momentum 0,3 *learning rate* 2 dan *hidden layer* lima. Sementara hasil pengolahan data *training* diabetes dengan optimasi *genetic algorithm* dengan menggunakan *neural network* diperoleh nilai akurasi tertinggi yaitu 77,10%. Nilai ini diperoleh dari nilai momentum 0,3 *learning rate* 2 dan *hidden layer* sebanyak dua. Dari penelitian ini dapat disimpulkan bahwa penggunaan metode optimasi *genetic algorithm* dapat meningkatkan akurasi dari metode neural network. Dari hasil pembahasan penelitian ini diperoleh atribut yang berperan peting bagi pengolahan dataset diabetes yaitu konsentrasi gula, tekanan darah, serum insulin, massa tubuh, diabetes silsilah fungsi dan umur, sedangkan atribut yang tidak berperan penting dan bisa dihilangkan serta tidak digunakan dalam penelitian pengolahan data training diabetes yaitu berapa kali hamil, lipatan kulit.

DAFTAR PUSTAKA

Elma, kolce(cela) dan Neki, Frasher. (2012). "A Literature Review of Data Mining Techniques used in Healthcare Databases", ICT Innovations Web Proceedings - Poster Session ISSN 1857-7288.

Han, J & Kamber. (2007). "Data Mining Concepts, Models and Techniques ", Second Edition, Morgan Kaufmann Publisher, Elsevier.

International Diabetes Federation. (2012). IDF Diabetes Atlas 5th ed, Brussels: International Diabetes Federation, www.idf.org/diabetesatlas. Accessed December 3, 2014.

Joseph L. Breault. (2012). "Data Mining Diabetic Databases:Are Rough Sets a Useful Addition".

Karthikeyani, V dkk (2012). "Comparative of Data Mining Classification Algorithm (CDMCA) in Diabetes Disease Prediction", International Journal of Computer Applications (0975 – 8887) Volume 60– No.12 hal 21-26.

Larose, Daniel T. (2005). "Discovering Knowledge in Data An Inntroduction Data

- Mining”, Wiley-Interscience, New Jersey, USA.
- Maimon, O., Rokach, L. (2010). “Data Mining and Knowledge Discovery Handbook 2nd Ed “, Springer, Berlin.
- Maimon, Oded & Rokach, Lior (2005), “Data Mining and Knowledge Discovery Handbook”, Springer, New York
- Myatt, Glenn J (2007), “ Making sense of data: A Practical Guide to Exploratory data analysis and Data Mining”, John Wiley & Sons Inc, New Jersey.
- Perkumpulan Endokrinologi Indonesia (PERKENI). (2006). “Diabetes Mellitus Konsensus Pengelolaan dan Pencegahan DM tipe 2 di Indonesia”, Jakarta
- Pusat data dan informasi Kemenkes RI. (2012). “Penyakit Tidak Menular, Buletin Jendela data dan Informasi Vol 2”, ISSN 2008-207x hal 1-13, Jakarta
- Sanakal, Ravi., Jayakumari, T Smt. (2014). “Prognosis of Diabetes Using Data mining Approach-Fuzzy C Means Clustering and Support Vector Machine”, International Journal of Computer Trends and Technology (IJCTT) – volume 11 number 2, ISSN: 2231-2803 hal 94-98.
- Sivandam S.N, Sumathi, S. (2006). “Intodution to Data Mining and Its Application”, Berlin, Springer.
- Tjokroprawiro, A. (2006). “Hidup Sehat dan Bahagia Bersama Diabetes Melitus”, Jakarta: PT Gramedia Pustaka Utama
- Vercellis,C. (2009). “Business Intelligence : Data Mining and Optimization for Decision Making, Wiley
- Whitcombe, J.M., Cropp, R.A., Braddock, R.D., Agranovski, I.E. (2006). ”The use of sensitivity analysis and genetic algorithms for the management of catalysemmissions from oil refi neries” Math. Comput. Model. 4 4, 430 e 438