

Kode / Nama Rumpun Ilmu : 458 / Teknik Informatika

**LAPORAN AKHIR
PENELITIAN DOSEN PEMULA**



**DATA MINING PADA PROSES PENENTUAN PREDIKSI
CHURN DAN SEGMENTASI PELANGGAN MENGGUNAKAN
BACKPROPAGATION NEURAL NETWORK BERBASIS
EVOLUTION STRATEGIES**

Oleh :

Junta Zeniarja, M.Kom, M.CS, 0606068802

Ardytha Luthfiarta, M.Kom, M.CS, 0625078504

**UNIVERSITAS DIAN NUSWANTORO SEMARANG
NOVEMBER 2014**

HALAMAN PENGESAHAN

Judul Kegiatan : Data Mining pada Proses Penentuan Prediksi Churn dan Segmentasi Pelanggan Menggunakan Backpropagation Neural Network berbasis Evolution Strategies

Peneliti / Pelaksana

Nama Lengkap : JUNTA ZENIARJA M.Kom
NIDN : 0606068802
Jabatan Fungsional :
Program Studi : Teknik Informatika
Nomor HP : 085727181701
Surel (e-mail) : junta@dsn.dinus.ac.id

Anggota Peneliti (1)

Nama Lengkap : ARDYTHA LUTHFIARTA M.Kom
NIDN : 0625078504
Perguruan Tinggi : Universitas Dian Nuswantoro

Institusi Mitra (jika ada)

Nama Institusi Mitra :
Alamat :
Penanggung Jawab :

Tahun Pelaksanaan : Tahun ke 1 dari rencana 1 tahun

Biaya Tahun Berjalan : Rp. 10.000.000,00

Biaya Keseluruhan : Rp. 15.000.000,00

Mengetahui
Dekan Fakultas Ilmu Komputer

(Dr. Abdul Syukur, MM)
NIP/NIK 0686.11.1992.017

Semarang, 29 - 10 - 2014,
Ketua Peneliti,

(JUNTA ZENIARJA M.Kom)
NIP/NIK 0686.11.2012.459

Menyetujui,
Kepala Pusat Penelitian

(Juli Ratnawati, S.E, M.Si)
NIP/NIK 0686.11.2000.193

DAFTAR ISI

BAB I. PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	2
1.3. Pertanyaan Penelitian	3
1.5. Batasan Masalah	4
1.6. Kontribusi Penelitian	4
1.7. Luaran yang Diharapkan	5
BAB 2. TINJAUAN PUSTAKA	6
2.1. Pengantar	6
2.2. Predksi Pelanggan Churn di CRM.....	6
2.3. Data Mining di CRM Kerangka	7
2.4. Artificial Neural Network (ANN)	11
2.5. Evolution Strategies (ES)	18
2.6. Evaluasi dan Validasi.....	19
2.7. Ringkasan.....	22
BAB 3. TUJUAN DAN MANFAAT PENELITIAN	23
3.1. Tujuan Penelitian.....	23
3.2. Manfaat Penelitian	23
BAB 4. METODOLOGI PENELITIAN	24
4.1. Jenis Penelitian.....	24
4.2. Desain Penelitian	25
4.3. Model yang diusulkan	27
4.4. Data Collection	27
4.5. Alat penelitian	29
4.6. Pengukuran Evaluasi	30
4.7. Ringkasan	30
BAB 5. HASIL DAN PEMBAHASAN	31
5.1 Pendahuluan.....	31
5.2 Evaluasi dan Validasi Model.....	31

5.3	Hasil Eksperimen standar BPNN	31
5.3.1	Training Cycles, Number and Size of Hidden Layer	31
5.3.2	Learning Rate dan Momentum	37
5.4	Hasil Perbandingan	40
5.5	Hasil Perbandingan	40
5.6	Ringkasan	43
BAB 6. RENCANA TAHAPAN BERIKUTNYA		44
BAB 7. KESIMPULAN DAN SARAN		45
DAFTAR PUSTAKA		46
LAMPIRAN 1 : TAMPILAN SAMPLE DATASET		48
LAMPIRAN 2 : SCRIPT CODING XML		49
LAMPIRAN 3 : BUKTI EMAIL PENERIMAAN JURNAL TECHNO COM.....		55
LAMPIRAN 4 : ARTIKEL DIMUAT DIJURNAL TECHNO COM.....		56
LAMPIRAN 5 : LAPORAN PENGGUNAAN DANA AKHIR		57

BAB I. PENDAHULUAN

1.1. Latar Belakang Masalah

Pelanggan adalah aset yang paling penting pada semua jenis bisnis. Prospek usaha hanya mungkin dapat dilakukan dengan adanya pelanggan yang puas yang selalu setia dan membangun hubungan mereka dengan perusahaan.Untuk alasan ini, perusahaan harus merencanakan dan menerapkan strategi untuk menciptakan pelanggan, umumnya dikenal sebagai Customer Relationship Management (CRM). Tsipitsis & Chorianopoulos (2010) mendefinisikan CRM sebagai strategi yang terkait dengan mempertahankan, mengelola, dan meningkatkan hubungan pelanggan setia dan tahan lama.

CRM terdiri dari serangkaian proses untuk mendukung strategi bisnis dalam rangka menciptakan dasar hubungan jangka panjang yang berharga dengan pelanggan. Salah satu dasar strategi CRM adalah Teknologi Informasi (TI). Perkembangan IT, khususnya telekomunikasi dan internet, memiliki dua efek. Pertama, telah meningkatkan peluang pasar, dan kedua perubahan pola manajemen dalam hubungan perusahaan-pelanggan.

Pelanggan Churn berarti kehilangan klien. Ini memiliki arti yang sama seperti gesekan pelanggan, pembelotan pelanggan, dan perputaran pelanggan. Churn pelanggan juga didefinisikan oleh Hung et al. (2006) di mana istilah ini dikenal dalam industri telekomunikasi yang umum digunakan untuk menyebut perpindahan pelanggan dari satu operator kepada operator yang lain(kompetitor).

Dengan mengacu pada perspektif bisnis cerdas, proses manajemen churn dalam kerangka CRM terdiri dari dua analisa utama yaitu pertama, prediksi pelanggan yang

cenderung untuk churn atau tidak churn dan yang kedua pendukung keputusan untuk membuat kebijakan program mempertahankan pelanggan loyal dan meningkatkan loyalitas pelanggan baru. Oleh karena itu, penelitian ini sangat terfokus pada pertimbangan pertama yaitu prediksi pelanggan churn.

Untuk memudahkan penelitian, teknologi informasi akan diterapkan dalam manajemen churn telekomunikasi. Salah satu teknologi informasi yang terkenal yang dilakukan dalam penelitian ini adalah dengan menggunakan teknik data mining untuk mencari model yang handal dari prediksi churn dari penyimpanan data. Tujuannya adalah untuk mencegah perpindahan pelanggan dan untuk meningkatkan daya saing.

Data mining adalah salah satu alat yang memiliki kemampuan untuk mencari solusi dari masalah manajemen pelanggan churn. Hal ini tepat terutama pada area prediksi churn. Beberapa penelitian sebelumnya, Liao (2007) menunjukkan teknik data mining secara pasti memiliki kemampuan dalam bidang layanan telekomunikasi untuk mengatasi masalah pelanggan churn. Ada beberapa teknik yang biasa digunakan dalam prediksi pelanggan churn. Diantaranya adalah Artificial Neural Network (ANN), Logistic Regression (LR), pohon keputusan (DT), dan Support Vector Machines (SVM).

1.2. Rumusan Masalah

Manajemen churn merupakan tugas penting bagi perusahaan atau perusahaan untuk mempertahankan pelanggan yang berharga. Riset pemasaran menunjukkan bahwa nilai rata-rata pelanggan yang churn atau pindah ke pesaing lain dari sebuah perusahaan operator seluler adalah sekitar 2% per bulan. Hung et al. (2006) menyatakan bahwa ada sekitar 24% dari pelanggan hilang setiap tahun. Berdasarkan

riset pasar, keadaan ini mendorong perusahaan untuk menyediakan biaya untuk dukungan penjualan, pemasaran, iklan, dan komisi untuk memperoleh pelanggan layanan mobile pelanggan baru. Dengan demikian, biaya untuk mendapatkan pelanggan baru jauh lebih tinggi daripada mempertahankan yang baru dan karenanya, kemampuan untuk memprediksi churn pelanggan adalah suatu keharusan.

Beberapa penelitian menggunakan Jaringan Syaraf Tiruan (JST) sebagai salah satu metode yang paling umum dalam prediksi data mining. Ia memiliki kemampuan untuk memprediksi untuk pemrosesan paralel dan fault tolerant. Berdasarkan Fei et al. (2009) ada beberapa kelemahan dari JST yang membatasi kepraktisan. Beberapa kelemahan diantaranya diperlukan data yang besar untuk pelatihan, over-fitting dari iterasi jaringan, lambat dalam konvergensi dan mudah terjebak dalam lokal optima. Untuk mengatasi kelemahan ini metode hybrid dalam data mining diusulkan.

Hybrid data mining adalah kombinasi dari dua atau lebih teknik dalam data mining untuk mencapai solusi terbaik(Tsai & Lu 2009). Ini terbukti dapat memberikan kemampuan yang lebih baik dibandingkan dengan teknik tunggal untuk beberapa domain masalah yang berbeda. Penelitian ini akan difokuskan pada model hybrid, yang menggabungkan algoritma JST untuk memecahkan prediksi churn, menggunakan Back Propagation Neural Network (BPNN) dan Evolution Strategi (ES) untuk pembobotan atributnya.

1.3. Pertanyaan Penelitian

Berdasarkan rumusan masalah yang dinyatakan dalam bagian sebelumnya, pertanyaan penelitian dari penelitian ini adalah sebagai berikut:

- a. Bagaimana teknik klasifikasi data mining dapat memecahkan masalah CRM?

- b. Bagaimana model hibrid dengan menggunakan BPNN berbasis ES, untuk mendapatkan akurasi yang lebih baik dibandingkan dengan BPNN dasar?
- c. Bagaimana akurasi kinerja ES-BPNN,dibandingkan dengan BPNN dasar?

1.4. Tujuan Penelitian

Tujuan penelitian ini sebagai berikut :

- a. Untuk menerapkan BPNN untuk prediksi churn CRM.
- b. Untuk mengusulkan sebuah model hibrida menggunakan BPNN dan mengoptimalkannya dengan ES untuk pembobotan atribut. Metode yang diusulkan bertujuan untuk memecahkan masalah prediksi churn pelanggan dengan menggunakan salah satu data set.
- c. Untuk membandingkan prediksi kinerja antara dasar BPNN dengan ES-BPNN.

1.5. Batasan Masalah

Beberapa lingkup yang diberikan di sini untuk menjaga agar studi ini fokus pada tujuan masalah yang telah ditetapkan pada bagian sebelumnya yaitu :

- a. Menggunakan BPNN dengan ES sebagai Hybrid pelatihan neural network diimplementasikan pada Churn Pelanggan Prediksi.
- b. Menggunakan ES untuk mengoptimalkan pelatihan BPNN.
- c. Eksperimen yang dilakukan menggunakan tools RapidMiner.

1.6. Kontribusi Penelitian

- a. Hybrid BPNN dengan ES yang dapat memperoleh hasil yang lebih baik dan error akurasi minimum pada prediksi pelanggan churn.

- b. Penggunaan ES untuk pengoptimalan bobot atribut pada training BPNN.
- c. Peningkatan kinerja untuk prediksi churn pelanggan dengan dataset UCI.

1.7. Luaran yang Diharapkan

- a. Penelitian ini memiliki target capaian luaran kegiatan berupa publikasi ilmiah dalam seminar atau conference nasional atau internasional, jurnal ilmiah local.

BAB 2. TINJAUAN PUSTAKA

2.1. Pengantar

Tinjauan pustaka ini difokuskan pada studi, ulasan, dan juga menguji pendekatan data mining dalam kerangka CRM. Bagian pertama dimulai dengan kajian umum CRM dan dilanjutkan dengan Data Mining. Bagian ini berlanjut dengan meninjau literatur yang ada dalam hal Neural Network dan kolaborasi dengan Algoritma Genetika. Akhirnya, bagian terakhir akan meninjau teknik Data mining yang akan digunakan dalam penelitian ini.

2.2. Prediksi Pelanggan Churn di CRM

Churn, juga dikenal sebagai gesekan, itu adalah istilah indikasi pelanggan meninggalkan layanan dari salah satu perusahaan di samping perusahaan lain (Tsiptsis & Chorianopoulos 2010). Penelitian sebelumnya dalam pemasaran membuktikan bahwa untuk mendapatkan pelanggan baru lebih mahal dibandingkan dengan mempertahankan pelanggan setia yang ada. Berdasarkan fakta ini, manajemen churn telah menjadi kegiatan yang penting bagi perusahaan untuk mempertahankan pelanggan setia. Skenario ini mencakup perusahaan Operator komunikasi seluler. Sebagai pasar layanan jaringan seluler semakin kompetitif, kemampuan untuk benar memprediksi churn pelanggan diperlukan.

2.3. Data Mining di CRM Kerangka

Data mining (DM) didefinisikan sebagai hasil penggalian informasi prediktif yang tersembunyi dari database besar (Liao et al. 2012). Ini adalah teknologi canggih dengan kemampuan besar yang dapat digunakan untuk membantu perusahaan dalam mendapatkan kecenderungan berharga dan penting dari database yang besar. Munculnya alat data mining mampu mengatasi pertanyaan-pertanyaan bisnis yang sudah diselesaikan secara konvensional dengan proses yang memakan waktu tinggi. Teknik data mining memungkinkan ekstraksi data mentah menjadi pengetahuan bisnis. Definisi lain dari data mining adalah kumpulan dari beberapa proses yang pemodelan, mengeksplorasi dan memilih sejumlah besar data yang bertujuan untuk mengetahui pola data baru yang tidak diketahui sebelumnya untuk keuntungan bisnis (Tsiptsis & Chorianopoulos 2010). Dari definisi ini, data mining terdiri dari analisis data dan algoritma pencarian untuk mengidentifikasi pola data untuk prediksi dan deskripsi.

Han dan Kamber (2001) mendefinisikan data mining (DM) sebagai penemuan pengetahuan dalam database, adalah proses mencari pola yang bermakna dalam database besar (Han & Kamber, 2001). Selanjutnya, Gorunescu (2011) menjelaskan bahwa data mining adalah sebuah aplikasi untuk memberikan keunggulan kompetitif untuk mencapai keputusan yang tepat. DM adalah proses investigasi dan kompleks yang melibatkan beberapa langkah berulang seperti tercantum di bawah :

- a. Pemahaman Bisnis(Business Understanding)

Merupakan tahap awal yaitu pemahaman penelitian, penentuan tujuan dan rumusan masalah data mining.

b. Pemahaman Data(Data Understanding)

Dalam tahap ini dilakukan pengumpulan data, mengenali lebih lanjut data yang akan digunakan.

c. Pengolahan Data(Data Preparation)

Tahap ini adalah pekerjaan berat yang perlu dilaksanakan secara intensif. Memilih kasus atau variable yang ingin dianalisis, melakukan perubahan pada beberapa variable jika diperlukan sehingga data siap untuk dimodelkan.

d. Pemodelan(Modeling)

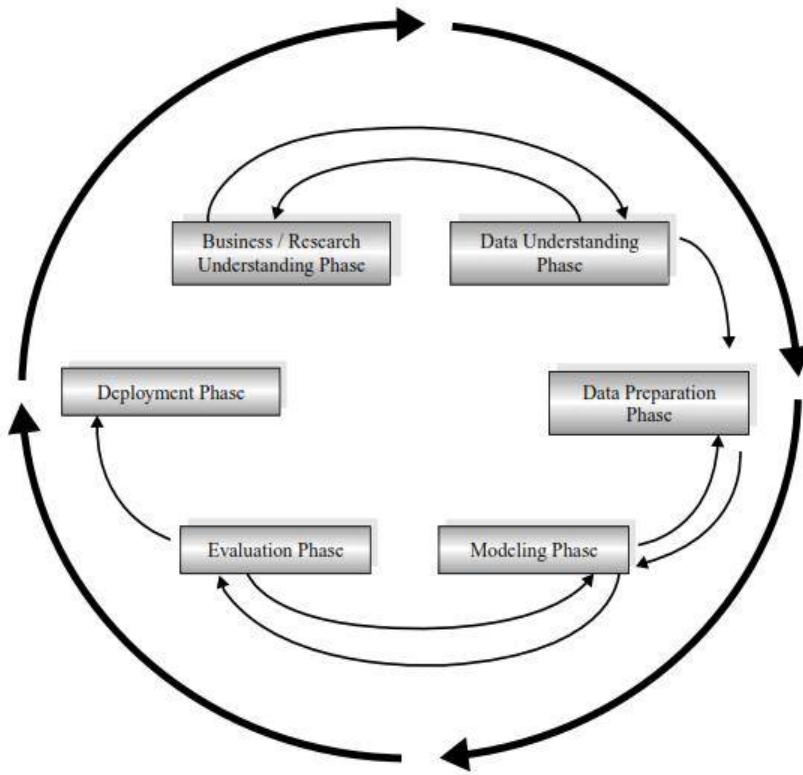
Memilih teknik pemodelan yang sesuai dan sesuaikan aturan model untuk hasil yang maksimal. Dapat kembali ke tahap pengolahan untuk menjadikan data ke dalam bentuk yang sesuai dengan model tertentu.

e. Evaluasi (Evaluation)

Mengevaluasi satu atau model yang digunakan dan menetapkan apakah terdapat model yang memenuhi tujuan pada tahap awal. Kemudian menentukan apakah ada permasalahan yang tidak dapat tertangani dengan baik serta mengambil keputusan hasil penelitian.

f. Penyebaran (Deployment)

Menggunakan model yang dihasilkan seperti pembuatan laporan atau penerapan proses data mining pada departemen lain.

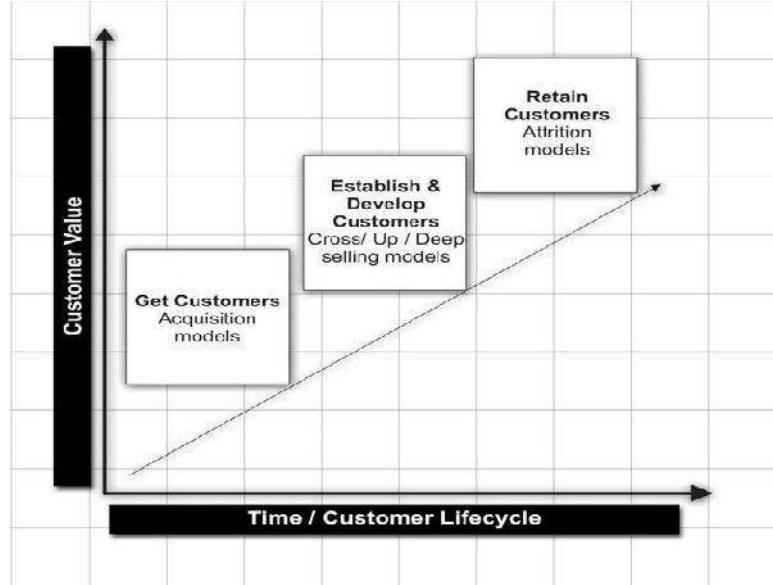


Gambar 2.1 Tahapan Data Mining

Data mining sangat mengeksplorasi pengetahuan tentang pelanggan yang signifikan untuk menciptakan strategi CRM yang efektif. Data mining mengarah untuk menyesuaikan hubungan dengan pelanggan sehingga akan meningkatkan kepuasan pelanggan dan mencapai hubungan pelanggan yang menguntungkan. Hal ini dapat digunakan untuk mempertahankan pelanggan yang dioptimalkan dan disesuaikan melalui tahapan siklus hidup pelanggan, mulai dari pencapaian korelasi yang tinggi untuk menghindari kehilangan pelanggan dan mendapatkan kembali pelanggan yang sudah hilang sebelumnya.

Mencapai pasar tinggi dan pangsa unggul pelanggan adalah tugas manajemen. Oleh karena itu, manajemen wajib untuk memperoleh, meningkatkan dan mempertahankan pelanggan. Model data mining dapat digunakan untuk manajemen

seluruh tujuan diatas. Gambar 2.2 digambarkan model data mining dalam kerangka CRM.



Gambar 2.2 Data mining dalam CRM

Ada banyak peluang implementasi teknologi data mining dapat CRM, selama ukuran database dan kualitas cukup (Tsiptsis & Chorianopoulos 2010). Dalam CRM, jaringan saraf tiruan (artificial neural network), pohon keputusan (decision tree), algoritma genetika (genetic algorithm), clustering, aturan induksi, dan asosiasi adalah contoh teknik data mining yang paling umum digunakan. Seperti ditunjukkan pada Tabel 2.1.

Tabel 2.1. Teknik Pertambangan dan aplikasi dalam CRM

Functionality	Technique	Application
Association	Set theory Statistics Bayesian Classification	Cross Sell

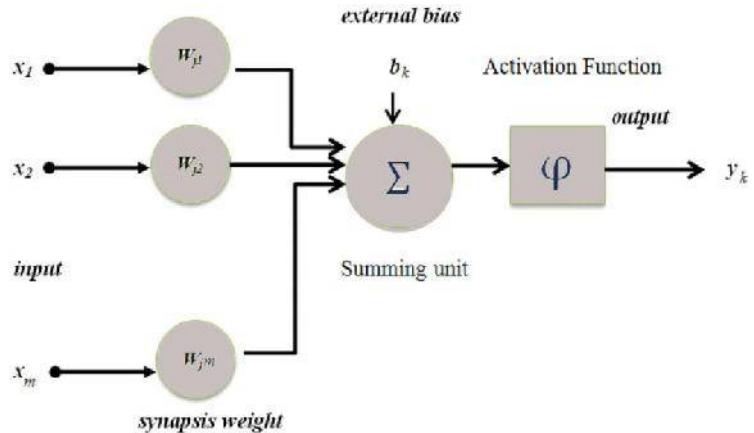
Estimation	Neural network	Exchange rate estimation	
	Statistics Time Series	Stock price estimation	
Classification	Decision tree Fuzzy Neural Network Genetic Algorithm	Credit embezzles Market segmentation	
Prediction	Regression Neural Network Decision tree	Churn prediction Fraudster prediction	
Segmentation	Neural network Statistics Genetic Algorithm Decision tree	Market segmentation	

2.4. Artificial Neural Network (ANN)

Artificial Neural Network (ANN) terinspirasi dari jaringan saraf makhluk hidup. Menurut Budi Santoso [11], kelebihan utama neural network adalah kemampuan memprediksi, kecepatan dan robust terhadap missing data.

Neuron adalah unit pemroses informasi dalam neural network yang terdiri atas:

- Set synapsis atau link penghubung yang ditandai dengan adanya bobot.
- Penambah, untuk menjumlahkan signal input yang diberi bobot yang disebut kombinasi linier.
- Fungsi aktivasi, untuk membatasi besarnya output dari suatu neuron.



Gambar 2.3 Model Neuron

Menghitung jumlah n signal input $x_{ij}=1,2,\dots,n$ yang diberi bobot dan menghasilkan nilai 1 bila jumlah diatas batas tertentu dan 0 bila dibawah batas. Secara matematis dapat dituliskan:

$$y = \varphi \left(\sum_j^n w_j x_j - u \right) \quad (1)$$

Dimana φ adalah fungsi aktivasi dan w adalah bobot sesuai dengan input ke-j. Dalam model neuron, bias dinyatakan sebagai b yang mempunyai fungsi menaikkan atau menurunkan net input untuk fungsi aktivasi. Neuron dinyatakan dengan k , dideskripsikan secara matematis:

$$u_k = \sum_{j=1}^m w_k x_j \quad (2)$$

dan

$$y_k = \varphi(u_k + b_k) \quad (3)$$

Dimana x_1, x_2, \dots, x_m adalah signal input dan w_1, w_2, \dots, w_m adalah bobot dari synapsis k. μ_k adalah kombinasi linier dari output yang dihasilkan signal. b_k adalah bias, φ adalah fungsi aktivasi dan y_k adalah signal output dari neuron yang bersangkutan. Pemakaian bias mempengaruhi *output neuron*

$$v_k = u_k + b_k \text{ atau}$$

$$v_k = \sum_{j=0}^m w_{kj} x_j \quad (4)$$

dan $y_k = \varphi(v_k)$. Dimana $f(*)$ adalah fungsi aktivasi dan b_k adalah bias. Sehingga fungsi aktivasi sigmoid didefinisikan sebagai berikut:

$$f(x) = \frac{1}{1 + e^{-ax}} \quad (5)$$

Dimana $a > 0$. Tujuan dari proses learning adalah menemukan bobot w dan bias b, sehingga network secara tepat menghasilkan output {-1,+1} untuk setiap data training yang dimasukkan. Untuk melatih perceptron adalah dengan mengawali nilai w dan b dengan nilai random dan akan memperbarui nilainya untuk setiap titik data bila nilai outputnya tidak sesuai yang diinginkan. Untuk mengupdate w_i pada input x_i dilakukan dengan cara:

$$w_{i+1} \leftarrow w_i + \Delta w_i \quad \text{dimana} \quad \Delta w_i = \eta(d - y)x_i$$

d adalah output dari data x_i , y adalah output dari perceptron, η adalah *learning rate* suatu bilangan positif. Untuk algoritma *backpropagasi*, mentraining *neural network* dengan banyak *neuron* digunakan metode *gradient descent*. Perbedaan antara *output* sebenarnya dengan *output perceptron* adalah E, yang bergantung pada nilai bobot yang ditemukan, maka didapatkan:

$$E(w) = \frac{1}{2} \sum_{t \in T} (d_t - y_t)^2 \quad (6)$$

dimana T adalah set sampel training, d_t adalah output yang diinginkan dan y_t adalah output yang didapat pada titik t . *Metode gradient descent* diterapkan untuk melakukan update vektor bobot w :

$$\nabla E(w) = \left[\frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w_1}, K, \frac{\partial E}{\partial w_m} \right] \quad (7)$$

Update w dilakukan dengan cara $w \leftarrow w + \Delta w$ dimana $w = -\eta E$. Dan untuk setiap komponen w bisa dituliskan $w_{i+1} \leftarrow w_i + \Delta w_i$ dimana $\Delta w_i = -\eta \frac{\partial E}{\partial w_i}$. Dari persamaan (6) bisa didapatkan turunan parsial dari E adalah :

$$\begin{aligned} \frac{\partial E}{\partial w_i} &= \frac{\partial}{\partial w_i} \frac{1}{2} \sum_{t \in T} (d_t - y_t)^2 \\ &= \frac{1}{2} \sum_{t \in T} \frac{\partial}{\partial w_i} (d_t - y_t)^2 \\ &= \sum_{t \in T} (d_t - y_t) \frac{\partial}{\partial w_i} (d_t - w \cdot x_t) \\ &= \sum_{t \in T} (d_t - y_t) (-x_{it}) \end{aligned}$$

Dengan hasil penurunan dapat dituliskan sebagai:

$$\Delta w_i = \eta \sum_{t \in T} (d_t - y_t)(-x_{it}) \quad (8)$$

Error adalah selisih antara target yang sebenarnya dan keluaran dari network pada unit output.

$$e_j(n) = d_j(n) - y_j(n) \quad (9)$$

Error untuk training adalah jumlah kesalahan keseluruhan error untuk semua unit output, contohnya j unit output:

$$E(n) = \frac{1}{2} \sum_j e_j^2(n) \quad (10)$$

Sedangkan menurut *gradient descent*, w diupdate dengan:

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t) \quad (11)$$

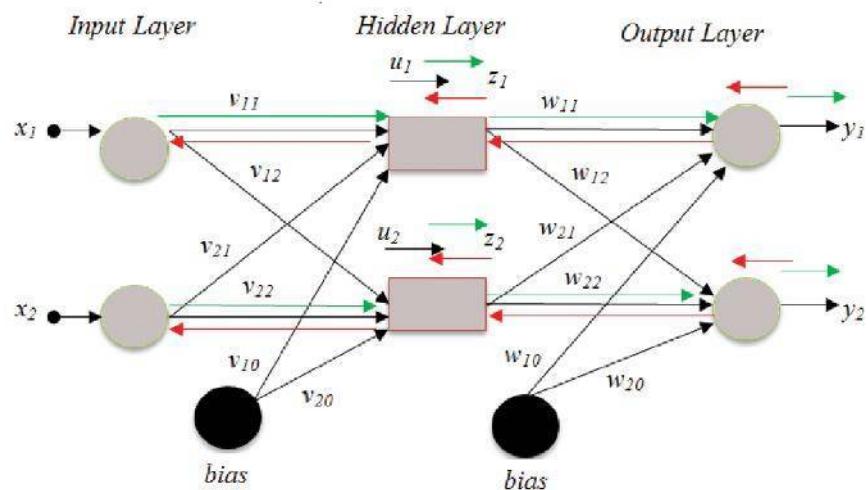
Sedangkan $\Delta w_{ij}(t)$ adalah :

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \quad (12)$$

dimana E adalah *error training* :

$$E(w) = \frac{1}{2} \sum_{k=1}^m (d_k - y_k)^2 \quad (13)$$

Untuk mekanisme backpropagasi dengan menggunakan multi layer perceptron, ditunjukkan pada gambar:

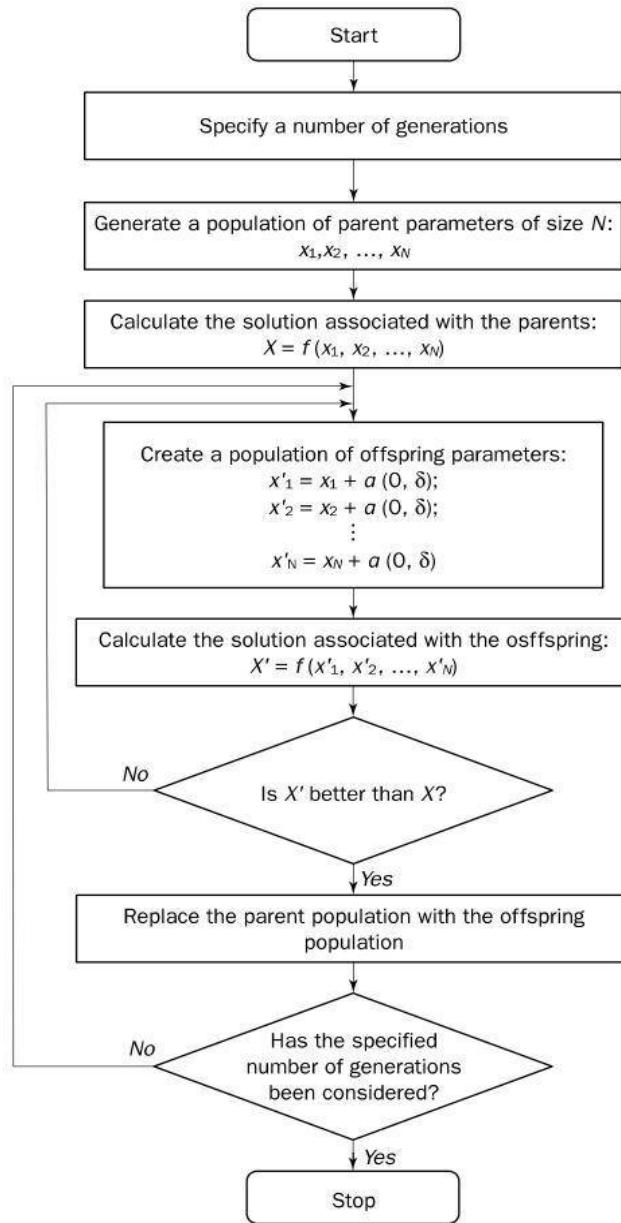


Gambar 2.4 Mekanisme backpropagasi pada multi layer perceptron.

2.5. Evolution Strategies (ES)

ES adalah pendekatan lain untuk mensimulasikan evolusi alami, yang telah diusulkan di Jerman pada awal 1960-an. Pendekatan ini dirancang untuk memecahkan masalah optimasi teknis. Ingo Rechenberg dan Hans-Paul Schwefel dari Technical University of Berlin Berdasarkan percobaan melelahkan dilakukan oleh mereka untuk menemukan solusi optimal, maka keduanya memutuskan untuk menggunakan perubahan acak dalam parameter berdasarkan pada mutasi alami. Oleh karena itu berdasarkan penelitian ini maka Evolution Strategies diciptakan (Negnevitsky, 2005).

Sampai saat ini, strategi evolusi biasanya diterapkan dalam masalah optimasi teknis karena tidak ada metode optimasi analitik atau konvensional ada. Tidak seperti GAs, strategi evolusi hanya beroperasi operator mutasi. Dalam bentuk yang paling sederhana, single parent mereproduksi keturunan tunggal untuk setiap generasi melalui penerapan mutasi terdistribusi normal. Langkah-langkah Evolution Strategies (Negnevitsky, 2005)



Gambar 2.5 Model Evolution Strategies (ES)

2.6. Evaluasi dan Validasi

Diperlukan cara yang sistematis untuk mengevaluasi kinerja suatu metoda.

Evaluasi prediksi didasarkan pada pengujian obyek benar dan salah. Dalam penelitian

ini menggunakan metode confusion matrix dan ROC Curve untuk mengukur hasil proses prediksi.

2.6.1 Confusion Matrix

Confusion matrix merupakan metode yang digunakan untuk mengetahui performansi algoritma. Dalam confusion matrix terdapat 4 sel yang harus ditentukan isinya. Kelas yang diprediksi ditampilkan dibagian atas matriks dan kelas yang diamati disisi kiri. Setiap sel berisi angka yang menunjukkan berapa banyak kasus yang sebenarnya dari kelas yang diamati untuk diprediksi.

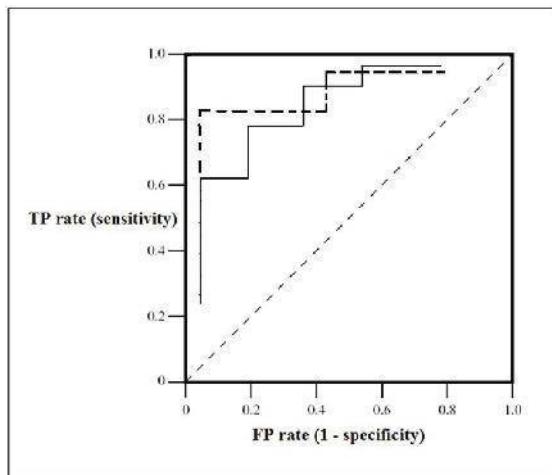
CLASSIFICATION	PREDICTED CLASS	
	Class = YES	Class = No
Class = YES	a (true positive-TP)	b (false negative-FN)
Class = No	c (true positive-TP)	d (false negative-FN)

Gambar 2.6 Confusion Matrix untuk 2 kelas.

Hasil klasifikasi dapat dihitung tingkat akurasinya berdasarkan kinerja matriks. Tingkat true positive (TP) adalah jumlah dari klasifikasi abnormal yang benar, TP juga disebut sensitivitas.

2.6.2 ROC Curve (AUC)

ROC curve digunakan dalam pembelajaran mesin dan penelitian data mining untuk menilai hasil prediksi.



Gambar 2.7 Contoh ROC Curve.

Kurva ROC dibagi dalam dua dimensi, dimana tingkat TP diplot pada sumbu Y dan tingkat FP diplot pada sumbu X. Sedangkan untuk merepresentasikan grafis yang menentukan klasifikasi mana yang lebih baik, digunakan metode yang menghitung luas daerah dibawah kurva ROC yang disebut AUC (Area Under the ROC Curve). AUC memiliki nilai antara 0,0 dan 1,0 dengan tingkat keakuratan klasifikasi sebagai berikut:

- $0,90 - 1,00 =$ sangat baik
- $0,80 - 0,90 =$ baik
- $0,70 - 0,80 =$ sama
- $0,60 - 0,70 =$ rendah
- $0,50 - 0,60 =$ gagal

2.7. Ringkasan

Dalam beberapa tahun terakhir, para peneliti telah dilakukan untuk mengusulkan berbagai teknik untuk memprediksi churn pelanggan. Model yang diusulkan termasuk pohon keputusan (Hung et al. 2006), jaringan syaraf tiruan (Tsai & Lu 2009), dukungan mesin vektor (Coussement & Dirk Van den Poel, 2008). Hybrid data mining adalah kombinasi dari dua atau lebih teknik dalam data mining untuk mendapatkan kinerja yang lebih baik daripada teknik tunggal melalui sejumlah masalah domain yang berbeda. Penelitian ini mengusulkan model hybrid dengan menggabungkan teknik jaringan saraf dan strategi evaluasi untuk masalah prediksi churn. Model jaringan saraf Terutama, penelitian ini menggunakan Back Propagation Neural Network (BPNN) dan model akan dioptimalkan menggunakan strategi Evolution untuk mendapatkan bobot atribut. Tabel 2.5 menunjukkan penelitian sebelumnya prediksi churn pelanggan.

BAB 3. TUJUAN DAN MANFAAT PENELITIAN

3.1. Tujuan Penelitian

Dalam penelitian ini bertujuan untuk :

- a. Menerapkan BPNN untuk prediksi churn CRM.
- b. Mengusulkan sebuah model hibrida menggunakan BPNN dan mengoptimalkannya dengan ES untuk pembobotan atribut. Metode yang diusulkan bertujuan untuk memecahkan masalah prediksi churn pelanggan dengan menggunakan salah satu data set.
- c. Membandingkan prediksi kinerja antara dasar BPNN dengan ES-BPNN.

3.2. Manfaat Penelitian

a. Manfaat Praktis

Hasil penelitian ini diharapkan untuk mendapatkan model data yang dapat digunakan untuk mengklasifikasi dan melakukan segmentasi pelanggan yang memiliki potensi churn. Sehingga model tersebut dapat dikembangkan menjadi sebuah Decision Support System.

b. Manfaat Akademis

Manfaat akademis yang didapatkan untuk memberikan kontribusi ilmiah dalam bidang komputasi menggunakan algoritma back propagation neural network dengan Evolution Strategis sebagai pembobot atribut memiliki kemampuan untuk memodelkan data dengan akurasi yang tinggi.

BAB 4. METODOLOGI PENELITIAN

Bab ini akan menyajikan rincian metodologi penelitian seperti jenis penelitian, desain penelitian, pengumpulan data dan metode yang diusulkan.

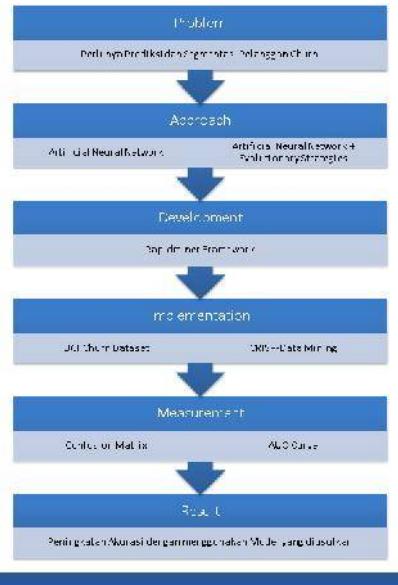
4.1. Jenis Penelitian

Ada empat metode penelitian yang umum digunakan, penelitian tindakan, eksperimen, studi kasus, dan survei (Dawson 2009). Penelitian eksperimen terdiri dari dua jenis, percobaan mutlak dan komparatif. Penelitian eksperimental umumnya dilakukan dalam pengembangan, evaluasi, dan pemecahan masalah proyek.

Penelitian ini akan mengadopsi metode penelitian eksperimen komparatif yang akan membandingkan dua model klasifikasi untuk prediksi churn pelanggan. Model pertama adalah Neural Network dasar, dan model kedua adalah Neural Network berbasis Evolution Strategies.

4.2. Desain Penelitian

Gambar 4.1 menggambarkan desain penelitian ini. Penelitian desgin terdiri dari enam langkah.



Gambar 4.1 Desain Penelitian

4.2.1. Masalah

Masalah yang memotivasi penelitian ini adalah tidak adanya algoritma klasifikasi mutlak dan akurat untuk prediksi churn pelanggan.

4.2.2. Pendekatan

Tujuannya adalah untuk mencapai tujuan, sedangkan ruang lingkup yang digunakan sebagai batas penelitian. Mendefinisikan keduanya akan membuat penelitian tetap fokus dan berada pada kisaran daerah penelitian. Tujuan dan ruang lingkup penelitian ini telah dinyatakan dalam bagian sebelumnya.

4.2.3. Pengembangan

Selama penelitian ini, percobaan dilakukan pada platform RapidMiner Software.

4.2.4. Implementasi

Percobaan dalam penelitian ini dilakukan dengan menggunakan UCI Churn set data, dan CRISP Data Mining.

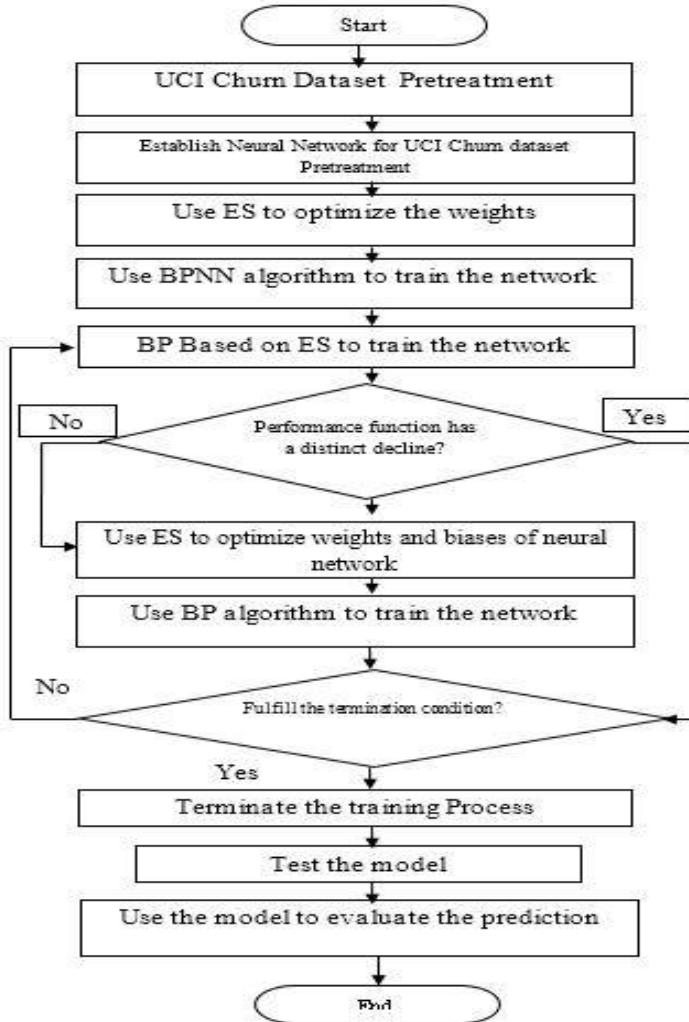
4.2.5. Pengukuran

Pengukuran adalah alat untuk mengevaluasi model. Dalam penelitian yang digunakan Kebingungan ini Matrix dan AUC digunakan untuk mengevaluasi model yang diusulkan.

4.2.6. Hasil

Hasil penelitian ini menunjukkan bahwa ES-BPNN mencapai akurasi yang lebih baik dibandingkan dengan dasar BPNN.

4.3. Model yang diusulkan



Gambar 4.2. Mengusulkan Model ES-BPNN

4.4. Data Collection

Untuk studi ini, kumpulan data churn yang digunakan adalah dari UCI Repository of Machine Learning Database di University of California, Irvine. Kumpulan data terdiri dari 20 variabel, dan berisi informasi sekitar 5000 pelanggan. Makalah ini mempertimbangkan 10-fold cross validasi, di mana 90% dari dataset tersebut akan digunakan sebagai model pelatihan dan sisanya sebagai model pengujian. Setiap

bagian akan dilatih dan diuji sepuluh kali untuk mendapatkan kinerja prediksi rata- rata. Indikasi apakah pelanggan yang churn atau tidak (meninggalkan perusahaan) juga termasuk dalam data. Rincian variabel ditunjukkan pada Tabel 4.1 dan Gambar 4.3, sebagai berikut:

Tabel 4.1. Dataset UCI Churn variabel

Tidak.	Nama Variabel	Jenis	Penjelasan
1.	Negara	Mutlak	Untuk 50 negara bagian dan District of Columbia
2.	Akun Panjang	Integer bernilai	Berapa lama akun telah aktif
3.	Kode Area	Mutlak	
4.	Nomor Telepon		Pada dasarnya pengganti untuk ID pelanggan
5.	Plan International	Mutlak	Dikotomis kategoris, ya atau tidak
6.	Rencana VoiceMail	Mutlak	Dikotomis kategoris, ya atau tidak
7.	Jumlah pesan suara	Integer bernilai	
8.	Jumlah menit sehari	Kontinu	Layanan pelanggan Menit digunakan siang hari
9.	Jumlah panggilan hari	Integer bernilai	
10.	Hari biaya total	Kontinu	Mungkin berdasarkan terdahulu dua variabel
11.	Jumlah menit malam	kontinu	Menit pelanggan menggunakan layanan selama malam
12.	Jumlah panggilan malam	Integer bernilai	
13.	Jumlah biaya malam	kontinu	Mungkin berdasarkan terdahulu dua variabel

14.	Jumlah menit malam	kontinu	
15.	Jumlah panggilan malam	Integer bernilai	
16.	Malam biaya total	kontinu	Mungkin berdasarkan terdahulu dua variabel
17.	Jumlah menit internasional	kontinu	Jasa menit pelanggan digunakan untuk membuat panggilan internasional
18.	Jumlah panggilan internasional	Integer bernilai	
19.	Muatan internasional Total	kontinu	Mungkin berdasarkan terdahulu dua variabel
20.	Jumlah panggilan ke layanan pelanggan	Integer bernilai	

4.5. Alat penelitian

Alat penelitian yang digunakan untuk mendukung rancangan percobaan disajikan dalam penelitian ini. Alat penelitian terdiri dari perangkat keras dan perangkat lunak persyaratan. Kebutuhan hardware yang digunakan untuk menjalankan penelitian ini terdiri dari:

Notebook: ASUS A42J Series

Sistem operasi: Windows 7 Ultimate 32-bit

Processor: Intel (R) Inti (TM) i3 CPU Quad

Memori Terpasang (RAM): Sodimm 4,0 GB DDR3 10600 Mhz

GPU: Ati Mobility Radeon HD 5470 dengan 512 MB

Persyaratan perangkat lunak yang digunakan untuk menjalankan penelitian ini adalah RapidMiner Version 5.2. Persyaratan untuk menggunakan RapidMiner adalah bahwa Java Runtime Environment (JRE) versi 1.5 (resmi Java 5.0) atau lebih tinggi, harus diinstal pada sistem.

4.6. Pengukuran Evaluasi

Keakuratan kinerja untuk memprediksi churn pelanggan dievaluasi oleh pengukuran evaluasi. Karena jaringan saraf adalah bagian dari teknik klasifikasi data mining, kebingungan matriks dan Area Under Curve ROC (AUC) akan digunakan untuk mengukur akurasi model yang diusulkan.

4.7. Ringkasan

Penelitian ini memiliki beberapa langkah. Langkah utama digambarkan seperti yang ditunjukkan pada Gambar 4.4, di mana percobaan dari model yang diusulkan dilakukan dalam perangkat lunak RapidMiner.

Langkah pertama adalah mengambil dataset, setelah data dikonversi dari binomial menjadi variabel numerik. Data tersebut kemudian diproses di Neural Network dan validasi pelatihan jaringan dilakukan.

Output dari proses terdiri dari hasil dari dua model dasar BPNN dan ES- BPNN. Kedua hasil tersebut dibandingkan untuk mengetahui model mana yang memperoleh kinerja yang lebih baik.

BAB 5. HASIL DAN PEMBAHASAN

5.1 Pendahuluan

Bab ini menjelaskan bagaimana percobaan yang dilakukan dan mendiskusikan hasil dari percobaan. Perbandingan antara model yang diusulkan, BPNN menggunakan ES, dengan standar BPNN. Perbandingan akurasi keduanya menggunakan Confusion Matrix dan AUC.

5.2 Evaluasi dan Validasi Model

Tujuan utama dari penelitian ini adalah untuk mengetahui akurasi algoritma klasifikasi yang digunakan dalam penelitian ini, BPNN menggunakan ES untuk pembobotan atribut, untuk menangani prediksi churn pelanggan. Sebagai perbandingan akurasi standar BPNN digunakan. Hal ini dilakukan untuk mengetahui mana yang memiliki akurasi yang lebih baik, apakah model pertama BPNN tanpa dioptimalkan atribut pembobotan oleh ES atau model kedua BPNN dengan ES untuk pembobotan atribut.

5.3 Hasil Eksperimen standar BPNN

5.3.1 Training Cycles, Number and Size of Hidden Layer

Nilai dari pelatihan neural network ditentukan oleh arsitektur jaringan neural network yang dipengaruhi oleh parameter training cycles, jumlah hidden layer dan banyaknya neuron dalam setiap layer hidden, serta momentum dan learning rate. Eksperimen dilakukan dengan memasukkan nilai untuk training cycles antara 50 sampai 1000. Metode yang sama juga digunakan untuk menentukan jumlah Hidden Layer sebagai parameter kedua. Nilai 0,3 digunakan untuk memberikan nilai momentum dan

0,2 untuk learning rate. Nilai-nilai yang mampu memberikan akurasi terbaik untuk BPNN dipilih.

Experimen pertama dengan melakukan training dengan beberapa percobaan beberapa kombinasi parameter BPNN dengan nilai Training Cycles 50 and Hidden Node Sizes antara 4, 8, 12, 16, 24, dan 32. Sehingga ada enam hasil percobaan eksperimen. Hasil percobaan dapat dilihat seperti pada gambar 5.1.

Table 5.1 Percobaan menggunakan 50 training cycles

Training Cycles	50					
Hidden Nodes Size	4	8	12	16	24	32
Accuracy	92.36	93.44	93.38	93.14	92.4	92.08
Precision	86.07	86.87	87.54	86.2	86.29	86.87
Recall	55.13	63.49	62.54	61.86	55.17	52.18
AUC	0.806	0.865	0.845	0.853	0.809	0.805
Time (s)	29	48	1.08	1.26	2.23	3.09

Dari table 5.1 diatas, hasil akurasi terbaik adalah pada percobaan 8 neuron pada layer hidden BPNN. Percobaan kemudian dikembangkan dengan 100 training cycles dan ternyata memberikan hasil akurasi yang lebih baik sebagaimana yang dapat dilihat pada Table 5.2.

Table 5.2 Training dengan 100 training cycles

Training Cycles	100					
Hidden Nodes	4	8	12	16	24	32
Accuracy	92.56	94.3	93.52	93.4	92.68	92.72
Precision	87.38	90.14	86.64	86.19	87.48	88.81
Recall	55.56	67.34	64.23	63.98	56.3	56.01
AUC	0.817	0.886	0.852	0.859	0.834	0.812
Time (s)	57	1.35	2.13	2.51	3.58	6.26

Percobaan penentuan parameter training cycles kemudian dikembangkan pada nilai, 150, 200, 250, 300, 350, 400, 450 and 500. Hasilnya bisa dilihat pada gambar 5.3 sd gambar 5.11

Table 5.3 Training dengan 100 training cycles

Training Cycles	150					
Hidden Nodes	4	8	12	16	24	32
Accuracy	92.46	95.04	93.58	93.56	92.82	92.74
Precision	87.48	91.59	87.37	85.16	86.12	88.36
Recall	55.55	71.27	64.09	65.88	58.71	56.43
AUC	0.808	0.89	0.855	0.863	0.842	0.821
Time (s)	1.27	2.23	3.23	4.17	6.54	9.17

Table 5.4 Training dengan 150 training cycles

Training Cycles	150					
Hidden Nodes	4	8	12	16	24	32
Accuracy	92.46	95.04	93.58	93.56	92.82	92.74
Precision	87.48	91.59	87.37	85.16	86.12	88.36
Recall	55.55	71.27	64.09	65.88	58.71	56.43
AUC	0.808	0.89	0.855	0.863	0.842	0.821
Time (s)	1.27	2.23	3.23	4.17	6.54	9.17

Table 5.5 Training dengan 200 training cycles

Training Cycles	200					
Hidden Nodes	4	8	12	16	24	32
Accuracy	92.6	95.12	93.7	93.6	93.16	93.04
Precision	87.21	92.02	87.32	86.17	87.33	89.18
Recall	56.4	72.01	65.23	65.95	60.4	58.12

AUC	0.806	0.891	0.855	0.862	0.854	0.824
Time (s)	1.58	3.17	4.32	5.35	8.38	12.46

Table 5.6 Training dengan 250 training cycles

Training Cycles	250					
Hidden Nodes	4	8	12	16	24	32
Accuracy	92.86	95.32	93.44	93.88	93.3	93.1
Precision	88.53	92.62	86.25	88.16	87.92	88.91
Recall	57.11	73.13	64.37	65.81	60.97	58.97
AUC	0.809	0.892	0.853	0.864	8.57	0.826
Time (s)	2.27	3.52	3.37	4.35	10.34	15.24

Table 5.7 Training dengan 300 training cycles

Training Cycles	300					
Hiden Nodes	4	8	12	16	24	32
Accuracy	92.78	95.26	93.46	93.82	93.36	93.16
Precision	88.66	92.27	86.84	87.65	88.22	89.01
Recall	56.26	72.99	63.67	65.81	61.11	59.25
AUC	0.811	0.892	0.852	0.864	88.22	0.826
Time (s)	1.54	3m	4.14	5.25	12.17	19.09

Table 5.8 Training dengan 350 training cycles

Training Cycles	350					
Hiden Nodes	4	8	12	16	24	32
Accuracy	92.62	95.22	93.54	93.88	93.26	93.14
Precision	87.35	92.04	87.4	87.32	87.57	89.08
Recall	56.26	72.71	64.09	66.37	60.96	58.97
AUC	0.811	0.892	0.852	0.869	0.857	0.827
Time (s)	2.07	5.39	4.56	6.2	15.41	21.05

Table 5.9 Training dengan 400 training cycles

Training Cycles	400					
Hiden Nodes	4	8	12	16	24	32
Accuracy	92.84	95.28	93.66	93.8	93.3	93.08
Precision	88.89	92.29	88.06	86.02	87.64	88.43
Recall	56.69	72.99	64.24	67.07	61.25	59.11
AUC	0.81	0.891	0.853	0.869	0.858	0.833
Time (s)	2m30s	6.41	5.38	7.13	19	13.56

Table 5.10 Training dengan 450 training cycles

Training Cycles	450					
Hidden Nodes	4	8	12	16	24	32
Accuracy	92.74	95.18	93.6	93.94	93.34	93.24
Precision	88.6	92.48	87.45	87.38	87.88	88.7
Recall	55.98	72	64.38	66.93	61.39	60.1
AUC	0.81	0.89	0.853	0.868	0.863	0.836
Time (s)	2.48	7.17	6.27	8.31	21.03	15.30m

Table 5.11 Training dengan 500 training cycles

Training Cycles	500					
Hidden Nodes	4	8	12	16	24	32
Accuracy	92.88	95.18	93.58	93.68	93.54	93.2
Precision	89.43	91.7	87.2	85.23	88.84	88.44
Recall	56.41	72.85	64.38	67.08	62.09	60.1
AUC	0.811	0.89	0.853	0.868	0.864	0.836
Time (s)	3.03	9.15	7.03	9.01	21.27	17.15

Hasil penelitian di atas menunjukkan bahwa dari berbagai percobaan nilai training cycle, hasil yang paling akurat dicapai dengan 8 ukuran neuron pada tiap hidden layer. Hasil akurasi terbaik adalah 95.32% didapatkan dengan kombinasi parameter

dengan konfigurasi 250 training cycles dengan jumlah hidden neuron adalah 8. Untuk memastikan konfigurasi tersebut adalah yang terbaik maka eksperimen dikembangkan dengan percobaan training cycles diatas nilai 500. ResUME hasil percobaan dapat dilihat pada table 5.12.

Table 5.12 Hasil Percobaan sampai dengan 500 training cycle

Hidden Layer	8 Hidden Layer									
	50	100	150	200	250	300	350	400	450	500
Accuracy	93.44	94.3	95.04	95.12	95.32	95.26	95.22	95.28	95.18	95.18
Precision	86.87	90.14	91.59	92.02	92.62	92.27	92.04	92.29	92.48	91.7
Recall	63.49	67.34	71.27	72.01	73.13	72.99	72.71	72.99	72	72.85
AUC	0.865	0.886	0.89	0.891	0.892	0.892	0.892	0.891	0.89	0.89
Time (s)	48	1.35	2.23	3.17	3.52	4.36	5.39	6.41	7.17	9.15

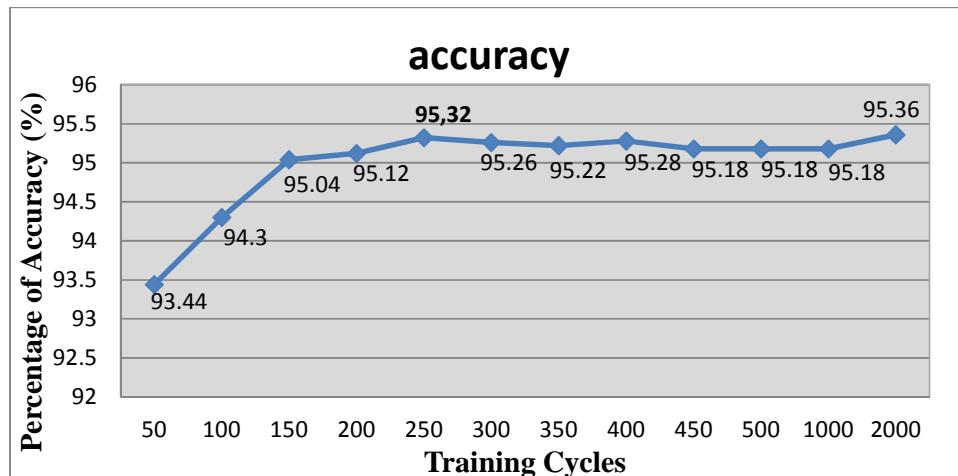
Percobaan kembali dikembangkan dengan nilai training cycles mencapai 2000. Nilai training cycle yang akan dikembangkan adalah 1000, 1500, and 2000. Jumlah hidden neuron yang dipilih adalah 8 sesuai hasil percobaan sebelumnya. Hasil percobaan dapat dilihat pada table 5.13.

Table 5.13 Testing with 500 – 2000 Training Cycles.

Hidden Layer	8 Hidden Layer											
	50	100	150	200	250	300	350	400	450	500	1000	2000
Training Cycles	93.44	94.3	95.04	95.12	95.32	95.26	95.22	95.28	95.18	95.18	95.18	95.36
Accuracy (%)	86.87	90.14	91.59	92.02	92.62	92.27	92.04	92.29	92.48	91.7	91.1	91.91
Precision (%)	63.49	67.34	71.27	72.01	73.13	72.99	72.71	72.99	72	72.85	73.42	73.99
Recall (%)	0.865	0.886	0.89	0.891	0.892	0.892	0.892	0.891	0.89	0.89	0.893	0.894
AUC	48	1.35	2.23	3.17	3.52	4.36	5.39	6.41	7.17	9.15	16.46	37.41

Dari percobaan diatas dapat diketahui bahwa hasil terbaik ditunjukkan pada training cycle 2000, akan tetapi waktu yang dibutuhkan untuk training sangat lama jika

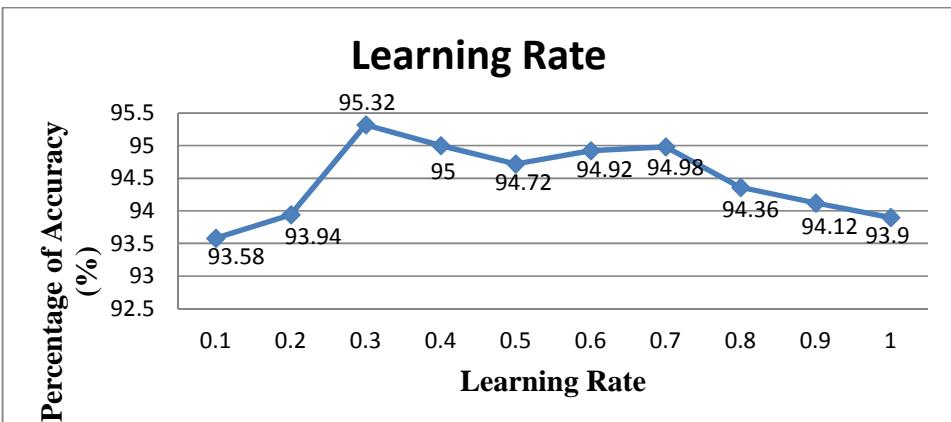
dibandingkan dengan training cycles 250. Oleh karena itu konfigurasi kombinasi parameter yang dipilih untuk kemudian dihibrid dengan algoritma ES adalah training cycle 250 dengan pertimbangan perbedaan akurasi yang sangat tipis hanya selisih 0.04%.



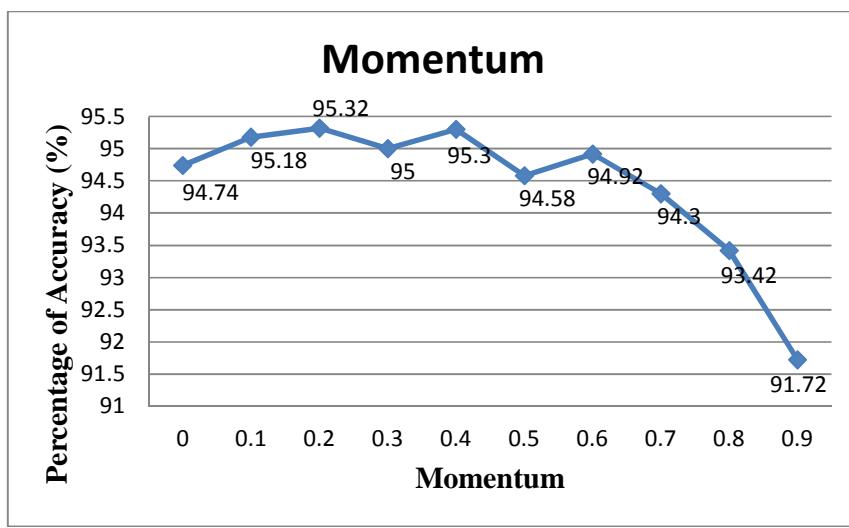
Gambar 5.2 Training Cycles Accuracy Chart

5.3.2 Learning Rate dan Momentum

Nilai Learning rate yang dipilih didasarkan nilai akurasi terbaik pada percobaan dengan menguji nilai mulai dari 0.1 sampai dengan 0.9. Untuk momentum awal dipilih 0.3. Sedangkan nilai training cycles 250 dan jumlah hidden neuron adalah 8 berdasarkan percobaan yang telah dilakukan . Hasil percobaan ditunjukkan seperti pada table 5.14 dan gambar 5.3. Dari hasil tersebut diketahui learning rate terbaik adalah 0.3. Sedangkan nilai momentum juga dilakukan dengan metode yang sama seperti pengujian learning rate.



Gambar 5.3 Learning Rate Accuracy Chart



Gambar 5.4 Momentum Accuracy Chart

Table 5.14 Testing learning rate with value 0.1 until 1

NeuralNet Structure	8 Hidden Layer 250 Epoch Momentum 0.2									
Learning Rate	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Accuracy	93.58	93.94	95.32	95	94.72	94.92	94.98	94.36	94.12	93.9
Precision	87.2	89.57	92.62	91.42	90.07	87.51	86.97	89.39	88.98	89.19
Recall	64.38	65.05	73.13	71.59	70.43	75.53	72	68.32	66.93	64.35
AUC	0.853	0.879	0.892	0.896	0.883	0.903	0.902	0.882	0.869	0.86
Time (s)	2.32	2.32	2.32	2.32	2.32	2.32	2.32	2.32	2.32	2.32

Table 5.15 Testing Momentum with value 0.0 until 0.9

ANN Structure	8 Hidden Layer 250 Epoch Learning Rate 0.3									
Momentum	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Accuracy	94.74	95.18	95.32	95	95.3	94.58	94.92	94.3	93.42	91.72
Precision	90.73	91.07	92.62	92	90.17	90.02	87.45	86.95	89.13	85.27
Recall	70.13	73.12	73.13	71.01	75.42	69.74	75.4	70.46	60.79	51.35
AUC	0.893	0.896	0.892	0.882	0.906	0.892	0.901	0.88	0.842	0.793
Time	2.32	2.32	2.32	2.32	2.32	2.32	2.32	2.32	2.32	2.32

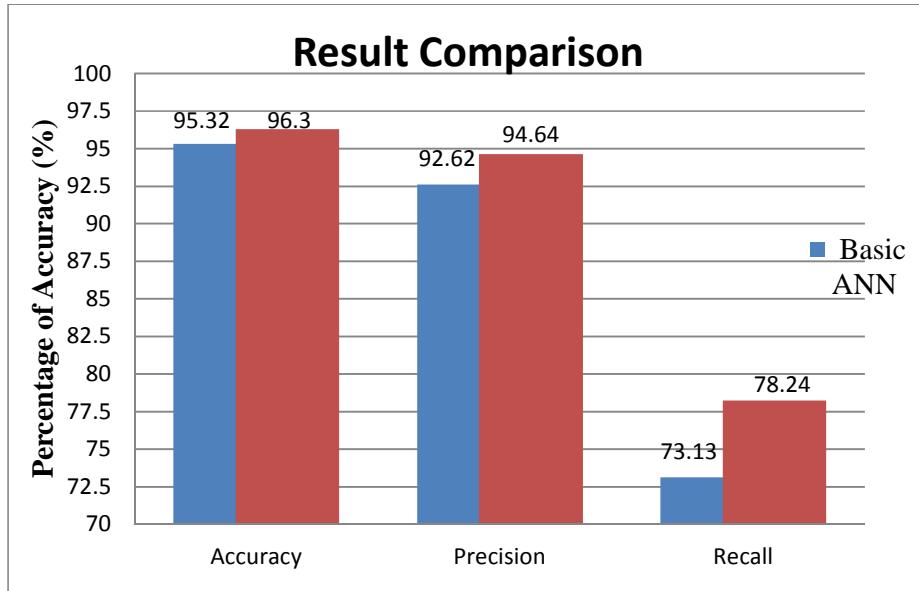
5.4 Hasil Perbandingan

Tabel 5.16 menunjukkan perbandingan akurasi antara BPNN dasar dan ES-BPNN. Hal ini menunjukkan bahwa ES-BPNN mengungguli di semua pengujian. Tapi waktu komputasi untuk ES-BPNN sangat panjang yaitu 5 jam 24 menit dan 06 detik, dibandingkan dengan BPNN dasar yang hanya membutuhkan waktu 52s 3m.

Table 5.16 Result of experiment

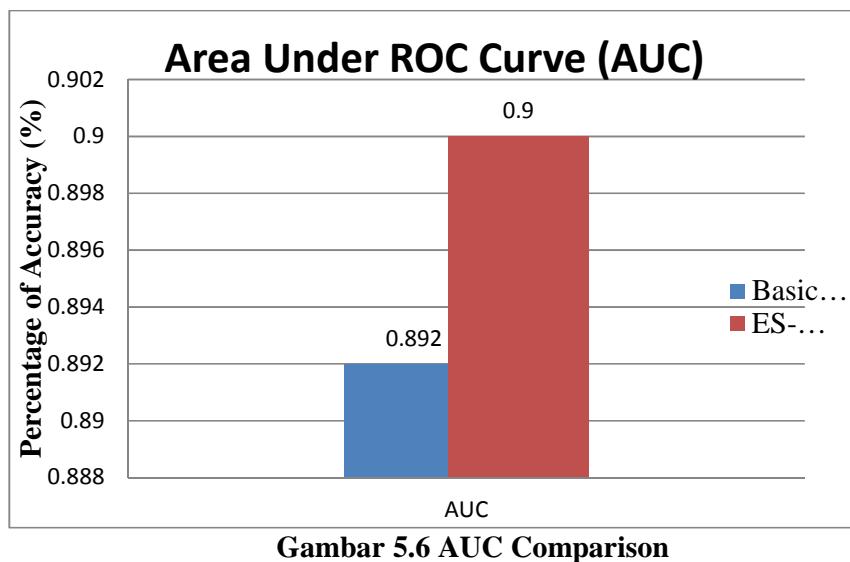
Comparison	Basic NN	ES-BPNN
Accuracy (Confusion Matrix)	95.32%	96.30%
Precision (Confusion Matrix)	92.62%	94.64%
Recall (Confusion Matrix)	73.13%	78.24%
Accuracy (AUC)	0.892	0.9
Execution time:	3 m. 52 s	5 h. 24m. 06s

5.5 Hasil Perbandingan



Gambar 5.5 Result Comparison

Dari Gambar 5.5 hasil eksperimen selama penelitian teknik ES-BPNN ini memperoleh akurasi yang lebih baik dengan 96,30% dibandingkan standar BPNN dengan 95,32%. Evaluasi dengan Confusion Matrix juga menunjukkan bahwa model yang diusulkan diperoleh presisi yang lebih baik (94,64%) dan recall (78,24%) dibandingkan dengan standar BPNN, yang memberikan hasil presisi dan recall masing-masing 92,62% dan 73,13%.



Evaluasi dengan AUC juga menunjukkan pada Gambar 5.6, bahwa model yang diusulkan lebih baik dengan 0,9 dari standar BPNN dengan 0,892.

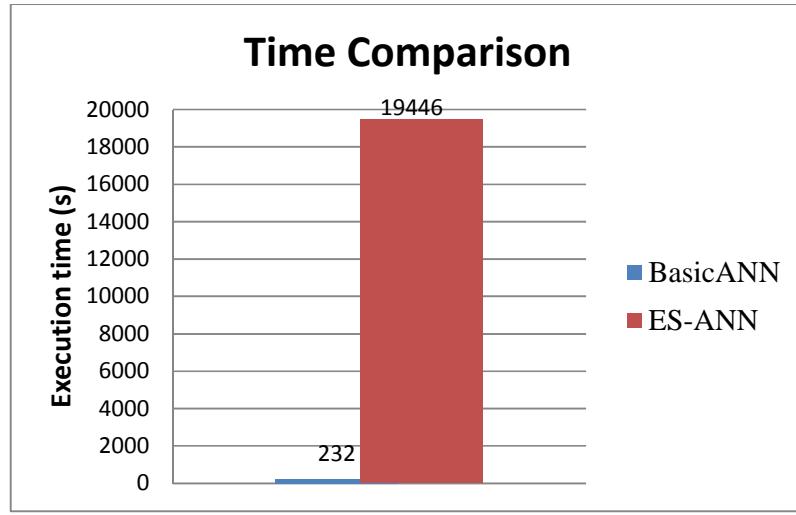


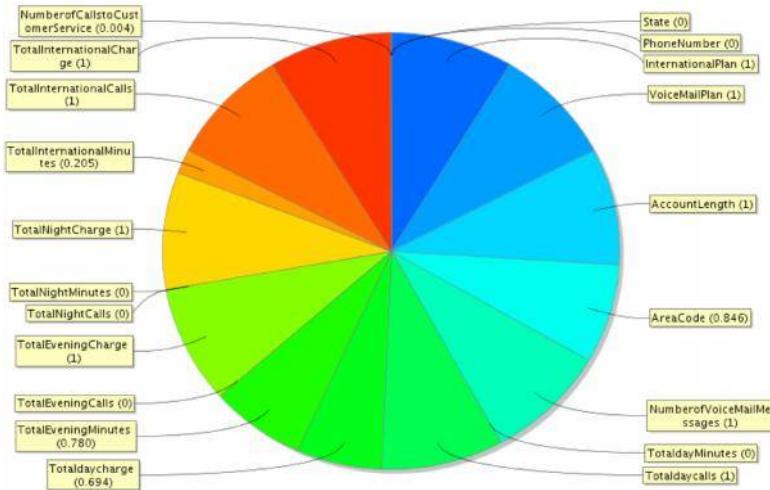
Figure 5.7 Execution time comparison

Sebagai perbandingan waktu eksekusi, BPNN standar hanya butuh 232 detik, sedangkan ES-BPNN perlu waktu lebih lama dengan 19.446 detik, bisa lihat pada Gambar 5.7 Dan hasil nilai baru koneksi berat badan yang membangun dengan ES-BPNN ditunjukkan pada Tabel 5.17.

Tabel 5.17 : Result of Attribute Weighting

Attribute	Weight
State	0.0
PhoneNumber	0.0
InternationalPlan	1.0
VoiceMailPlan	1.0
AccountLength	1.0
AreaCode	0.8456361867955082
NumberofVoiceMailMessages	1.0
TotaldayMinutes	0.0
Totaldaycalls	1.0
Totaldaycharge	0.6940721872082961
TotalEveningMinutes	0.7796641423521627
TotalEveningCalls	0.0
TotalEveningCharge	1.0
TotalNightMinutes	0.0
TotalNightCalls	0.0
TotalNightCharge	1.0
TotalInternationalMinutes	0.20536566789920074
TotalInternationalCalls	1.0
TotalInternationalCharge	1.0
NumberofCallsstoCustomerService	0.004033628904048681

Gambar 5.8 Menampilkan chart pie atribut pembobotan. Nilai dari setiap bobot dan perbandingan dengan jelas dapat dilihat pada Gambar 5.8 dan Gambar 5.9.



Gambar 5.8 Pie chart of Attribute Weighting

5.6 Ringkasan

Eksperimen dibangun di atas RapidMiner. Penggunaan data mining teknik BPNN hibrid dengan ES untuk atribut bobot. Validasi model dilakukan dengan menggunakan 10 Fold Cross Validation dan evaluasi pengukuran selesai dengan menggunakan confusion matriks dan AUC.

Berdasarkan hasil percobaan, arsitektur terbaik dari BPNN adalah dengan menggunakan 250 siklus pelatihan, 8 neuron tersembunyi dengan satu lapisan tersembunyi. Dan nilai learning rate adalah 0,2 dan nilai momentum 0,2. Hasil percobaan menunjukkan bahwa Dasar BPNN memperoleh akurasi kinerja 95,30% dengan menggunakan pengukuran confusion matriks.

BPNN paramater kemudian digunakan dalam model kedua hyberid ES-BPNN. Hasil percobaan menunjukkan bahwa hybrid BPNN dengan ES memperoleh 96,32% dalam pengukuran matrix kebingungan.

Penelitian ini menunjukkan bahwa ES-BPNN mencapai kinerja yang lebih baik daripada standar BPNN.

BAB 6. RENCANA TAHAPAN BERIKUTNYA

Penelitian ini masih memiliki beberapa kegiatan dan target capaian yang belum terpenuhi. Kegiatan yang masih belum dilakukan dan akan dijalankan pada tahap berikutnya adalah sebagai berikut :

1. Revisi paper atau makalah yang telah registrasi di ICITACEE 2014.
2. Registrasi ulang dan unggah paper revisi ke situs resmi ICITACEE 2014.
3. Persiapan pembuatan slide presentasi untuk pemakalah ICITACEE 2014.
4. Melakukan presentasi makalah di ICITACEE 2014.
5. Menarik kesimpulan dari hasil penelitian ini dan mengidentifikasi topic potensial untuk penelitian selanjutnya yang relevan dengan penelitian ini.
6. Melakukan pelaporan akhir di SIMLITABMAS.
7. Penyerahan berkas laporan akhir penelitian ke LPPM UDINUS.

Kegiatan tersebut ditargetkan akan selesai pada awal bulan Desember 2014, sehingga dapat dilakukan penulisan laporan akhir pada bulan yang sama sebagai pertanggung jawaban kepada DIKTI selaku penyandang dana penelitian ini.

BAB 7. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil koordinasi dan diskusi yang dilakukan selama penelitian ini ada beberapa hal yang dapat disimpulkan dari proses penyusunan penelitian, yaitu :

1. Perlunya koordinasi yang rapi dan diskusi yang terencana baik.
2. Perlunya pembagian tugas dan job desk yang detil dan terstruktur.

Saran

Saran untuk menyelesaikan laporan tahap ke dua (30%), atau laporan akhir dengan baik adalah dengan memperhatikan faktor teknis, yaitu yang berkaitan dengan penelitian, juga memperhatikan faktor-faktor non teknis, yaitu yang tidak berhubungan langsung dengan penelitian, tetapi penting untuk di lakukan yaitu adanya koordinasi dan diskusi yang baik, rapi dan terstruktur dan pembagian tugas dan job des yang detil dan terstruktur.

DAFTAR PUSTAKA

- [1] Abbasimehr, H., 2011. A Neuro-Fuzzy Classifier for Customer Churn Prediction. , 19(8), pp.35-41.
- [2] Au, W.-ho, Chan, K.C.C. & Yao, X., 2003. A Novel Evolutionary Data Mining Algorithm With Applications to Churn Prediction. , 7(6), pp.532-545.
- [3] Basiri, J., Taghiyareh, F. & Moshiri, B., 2010. A Hybrid Approach to Predict Churn. *2010 IEEE Asia-Pacific Services Computing Conference*, pp.485-491.
- [4] Burez, J. & Van den Poel, D., 2009. Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36(3), pp.4626-4636.
- [5] Coussement, K. & Van den Poel, Dirk, 2008. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34(1), pp.313-327.
- [6] Dawson, C.W., 2009. *Projects in Computing and Information Systems A Student Guide*, Dong, Y.-jie, Wang, X.-hua & Zhou, J., 2009. CostBP Algorithm and its Application in Customer Churn Prediction. *2009 Fifth International Joint Conference on INC, IMS and IDC*, pp.794-797.
- [7] Fei, S.-wei, Miao, Y.-bin & Liu, C.-liang, 2009. Chinese Grain Production Forecasting Method Based on Particle Swarm Optimization-based Support Vector Machine. *Science*, pp.8-12.
- [8] Gorunescu, F., 2011. *Data Mining Concepts, Models and Techniques*, Huang, B., Kechadi, Mohand Tahar & Buckley, B., 2012. Customer churn prediction in telecommunications. *Expert Systems with Applications*, 39(1), pp.1414-1425.
- [9] Huang, Y., Huang, B.Q. & Kechadi, M. T., 2010. A new filter feature selection approach for customer churn prediction in telecommunications. *2010 IEEE*

International Conference on Industrial Engineering and Engineering Management, (2), pp.338342.

- [10] Hung, S.-Y., Yen, D.C. & Wang, H.-Y., 2006. Applying data mining to telecom churn management. *Expert Systems with Applications*, 31(3), pp.515-524.
- [11] Hur, Y. & Lim, S., 2005. Customer Churning Prediction Using Support Vector Machines in Online Auto Insurance Service. , pp.928-933.

LAMPIRAN 1 : TAMPILAN SAMPLE DATASET

1	2	3	4	5	6	7	8	9	10
KS	128,0	415,0	382-4657	no	yes	25,0	265,1	110,0	45,1
OH	107,0	415,0	371-7191	no	yes	26,0	161,6	123,0	27,5
NJ	137,0	415,0	358-1921	no	no	,0	243,4	114,0	41,4
OH	84,0	408,0	375-9999	yes	no	,0	299,4	71,0	50,9
OK	75,0	415,0	330-6626	yes	no	,0	166,7	113,0	28,3
AL	118,0	510,0	391-8027	yes	no	,0	223,4	98,0	38,0
MA	121,0	510,0	355-9993	no	yes	24,0	218,2	88,0	37,1
MO	147,0	415,0	329-9001	yes	no	,0	157,0	79,0	26,7
LA	117,0	408,0	335-4719	no	no	,0	184,5	97,0	31,4
WV	141,0	415,0	330-8173	yes	yes	37,0	258,6	84,0	44,0
IN	65,0	415,0	329-6603	no	no	,0	129,1	137,0	22,0
RI	74,0	415,0	344-9403	no	no	,0	187,7	127,0	31,9
IA	168,0	408,0	363-1107	no	no	,0	128,8	96,0	21,9
MT	95,0	510,0	394-8006	no	no	,0	156,6	88,0	26,6
IA	62,0	415,0	366-9238	no	no	,0	120,7	70,0	20,5
NY	161,0	415,0	351-7269	no	no	,0	332,9	67,0	56,6
ID	85,0	408,0	350-8884	no	yes	27,0	196,4	139,0	33,4
VT	93,0	510,0	386-2923	no	no	,0	190,7	114,0	32,4
VA	76,0	510,0	356-2992	no	yes	33,0	189,7	66,0	32,3
TX	73,0	415,0	373-2782	no	no	,0	224,4	90,0	38,2
FL	147,0	415,0	396-5800	no	no	,0	155,1	117,0	26,4
CO	77,0	408,0	393-7984	no	no	,0	62,4	89,0	10,6
AZ	130,0	415,0	358-1958	no	no	,0	183,0	112,0	31,1
SC	111,0	415,0	350-2565	no	no	,0	110,4	103,0	18,8
VA	132,0	510,0	343-4696	no	no	,0	81,1	86,0	13,8

LAMPIRAN 2 : SCRIPT CODING XML

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<process version="5.3.015">
    <context>
        <input/>
        <output/>
        <macros/>
    </context>
    <operator activated="true" class="process" compatibility="5.3.015" expanded="true"
              name="Root">
        <parameter key="logverbosity" value="init"/>
        <parameter key="random_seed" value="1977"/>
        <parameter key="send_mail" value="never"/>
        <parameter key="notification_email" value="" />
        <parameter key="process_duration_for_mail" value="30"/>
        <parameter key="encoding" value="SYSTEM"/>
        <process expanded="true">
            <operator activated="true" class="retrieve" compatibility="5.3.015" expanded="true"
                      height="60" name="Retrieve churn" width="90" x="45" y="75">
                <parameter key="repository_entry" value="//Local Repository/penelitian/churn"/>
            </operator>
            <operator activated="true" class="nominal_to_numerical" compatibility="5.3.015"
                      expanded="true" height="94" name="Nominal to Numerical" width="90" x="45" y="165">
                <parameter key="return_preprocessing_model" value="false"/>
                <parameter key="create_view" value="false"/>
                <parameter key="attribute_filter_type" value="all"/>
                <parameter key="attribute" value="" />
                <parameter key="attributes" value="" />
            </operator>
        </process>
    </operator>
</process>
```

```

<parameter key="use_except_expression" value="false"/>
    <parameter key="value_type" value="nominal"/>

<parameter key="use_value_type_exception" value="false"/>
    <parameter key="except_value_type" value="file_path"/>
        <parameter key="block_type" value="single_value"/>

<parameter key="use_block_type_exception" value="false"/>
    <parameter key="except_block_type" value="single_value"/>
        <parameter key="invert_selection" value="false"/>

<parameter key="include_special_attributes" value="false"/>
    <parameter key="coding_type" value="dummy coding"/>
    <parameter key="use_comparison_groups" value="false"/>
        <list key="comparison_groups"/>

<parameter key="unexpected_value_handling" value="all 0 and warning"/>
    <parameter key="use_underscore_in_name" value="false"/>
        </operator>

<operator activated="true" class="optimize_weights_evolutionary"
compatibility="5.3.015" expanded="true" height="94" name="Optimize Weights
(Evolutionary)" width="90" x="246" y="210">
    <parameter key="population_size" value="5"/>
    <parameter key="maximum_number_of_generations" value="30"/>
        <parameter key="use_early_stopping" value="false"/>
        <parameter key="generations_without_improval" value="2"/>
            <parameter key="normalize_weights" value="true"/>
            <parameter key="use_local_random_seed" value="false"/>
                <parameter key="local_random_seed" value="1992"/>
                <parameter key="show_stop_dialog" value="false"/>
                <parameter key="user_result_individual_selection" value="false"/>
                <parameter key="show_population_plotter" value="false"/>
                    <parameter key="plot_generations" value="10"/>
                    <parameter key="constraint_draw_range" value="false"/>

```

```

<parameter key="draw_dominated_points" value="true"/>
<parameter key="maximal_fitness" value="Infinity"/>
<parameter key="selection_scheme" value="tournament"/>
<parameter key="tournament_size" value="0.25"/>
<parameter key="start_temperature" value="1.0"/>
<parameter key="dynamic_selection_pressure" value="true"/>
<parameter key="keep_best_individual" value="false"/>
<parameter key="save_intermediate_weights" value="false"/>
<parameter key="intermediate_weights_generations" value="10"/>
<parameter key="mutation_variance" value="1.0"/>
<parameter key="1_5_rule" value="true"/>
<parameter key="bounded_mutation" value="false"/>
<parameter key="p_crossover" value="0.0"/>
<parameter key="crossover_type" value="uniform"/>
<parameter key="use_default_mutation_rate" value="true"/>
<parameter key="initialize_with_input_weights" value="false"/>
<process expanded="true">
<operator activated="true" class="x_validation" compatibility="5.3.015"
expanded="true" height="130" name="Validation" width="90" x="179" y="75">
<parameter key="create_complete_model" value="false"/>
<parameter key="average_performances_only" value="true"/>
<parameter key="leave_one_out" value="false"/>
<parameter key="number_of_validations" value="10"/>
<parameter key="sampling_type" value="stratified sampling"/>
<parameter key="use_local_random_seed" value="false"/>
<parameter key="local_random_seed" value="1992"/>
<process expanded="true">
<operator activated="true" class="neural_net" compatibility="5.3.015"
expanded="true" height="76" name="Neural Net" width="90" x="112" y="75">
<list key="hidden_layers"/>

```

```

<parameter key="training_cycles" value="500"/>
<parameter key="learning_rate" value="0.3"/>
<parameter key="momentum" value="0.2"/>
<parameter key="decay" value="false"/>
<parameter key="shuffle" value="true"/>
<parameter key="normalize" value="true"/>
<parameter key="error_epsilon" value="1.0E-5"/>
<parameter key="use_local_random_seed" value="false"/>
<parameter key="local_random_seed" value="1992"/>
</operator>

<connect from_port="training" to_op="Neural Net" to_port="training set"/>
<connect from_op="Neural Net" from_port="model" to_port="model"/>
<portSpacing port="source_training" spacing="0"/>
<portSpacing port="sink_model" spacing="0"/>
<portSpacing port="sink_through 1" spacing="0"/>
</process>
<process expanded="true">
<operator activated="true" class="apply_model" compatibility="5.3.015"
expanded="true" height="76" name="Apply Model" width="90" x="45" y="75">
<list key="application_parameters"/>
<parameter key="create_view" value="false"/>
</operator>
<operator activated="true" class="performance_classification"
compatibility="5.3.015" expanded="true" height="76" name="Performance" width="90"
x="179" y="30">
<parameter key="main_criterion" value="first"/>
<parameter key="accuracy" value="true"/>
<parameter key="classification_error" value="false"/>
<parameter key="kappa" value="false"/>
<parameter key="weighted_mean_recall" value="false"/>

```

```

<parameter key="weighted_mean_precision" value="false"/>
<parameter key="spearman_rho" value="false"/>
<parameter key="kendall_tau" value="false"/>
<parameter key="absolute_error" value="false"/>
<parameter key="relative_error" value="false"/>
<parameter key="relative_error_lenient" value="false"/>
<parameter key="relative_error_strict" value="false"/>
<parameter key="normalized_absolute_error" value="false"/>
<parameter key="root_mean_squared_error" value="false"/>
<parameter key="root_relative_squared_error" value="false"/>
<parameter key="squared_error" value="false"/>
<parameter key="correlation" value="false"/>
<parameter key="squared_correlation" value="false"/>
<parameter key="cross-entropy" value="false"/>
<parameter key="margin" value="false"/>
<parameter key="soft_margin_loss" value="false"/>
<parameter key="logistic_loss" value="false"/>
<parameter key="skip_undefined_labels" value="true"/>
<parameter key="use_example_weights" value="true"/>
<list key="class_weights"/>
</operator>

<connect from_port="model" to_op="Apply Model" to_port="model"/>
<connect from_port="test set" to_op="Apply Model" to_port="unlabelled data"/>
<connect from_op="Apply Model" from_port="labelled data" to_op="Performance"
        to_port="labelled data"/>
<connect from_op="Performance" from_port="performance" to_port="averagable
        1"/>
<portSpacing port="source_model" spacing="0"/>
<portSpacing port="source_test set" spacing="0"/>
<portSpacing port="source_through 1" spacing="0"/>

```

```

<portSpacing port="sink_averagable 1" spacing="0"/>
<portSpacing port="sink_averagable 2" spacing="0"/>

</process>
</operator>

<connect from_port="example set" to_op="Validation" to_port="training"/>

<connect from_op="Validation" from_port="averagable 1" to_port="performance"/>

<portSpacing port="source_example set" spacing="0"/>
<portSpacing port="source_through 1" spacing="0"/>
<portSpacing port="sink_performance" spacing="0"/>

</process>
</operator>

<connect from_op="Retrieve churn" from_port="output" to_op="Nominal to Numerical"
        to_port="example set input"/>

<connect from_op="Nominal to Numerical" from_port="example set output"
        to_op="Optimize Weights (Evolutionary)" to_port="example set in"/>

<connect from_op="Optimize Weights (Evolutionary)" from_port="example set out"
        to_port="result 1"/>

<connect from_op="Optimize Weights (Evolutionary)" from_port="weights"
        to_port="result 2"/>

<connect from_op="Optimize Weights (Evolutionary)" from_port="performance"
        to_port="result 3"/>

<portSpacing port="source_input 1" spacing="0"/>
<portSpacing port="sink_result 1" spacing="0"/>
<portSpacing port="sink_result 2" spacing="0"/>
<portSpacing port="sink_result 3" spacing="0"/>
<portSpacing port="sink_result 4" spacing="0"/>

</process>
</operator>

</process>

```

LAMPIRAN 3 : BUKTI EMAIL PENERIMAAN JURNAL TECHNO COM

Konfirmasi Penerimaan Jurnal Techno COM

Konfirmasi Penerimaan Jurnal Techno COM Kotak Masuk x  

 Hanny Haryanto 1 Nov (1 hari yang lalu)  X  

ke saya ▼

Yth. Bapak Ibu Pengirim Artikel Jurnal Techno.Com

Bersama ini kami sampaikan bahwa artikel dengan judul "**PREDIKSI CHURN DAN SEGMENTASI PELANGGAN MENGGUNAKAN BACKPROPAGATION NEURAL NETWORK BERBASIS EVOLUTION STRATEGIES**" telah kami terima dan sedang kami review untuk penerbitan Jurnal Techno.COM, Vol. 14, No. 1, Februari 2015.

Demikian email kami, atas perhatian dan kerjasamanya kami ucapan banyak terimakasih.

Hormat kami,
Dewan Redaksi Jurnal Techno.Com

LAMPIRAN 4 : ARTIKEL DIMUAT DIJURNAL TECHNO COM

LAMPIRAN 5 : LAPORAN PENGGUNAAN DANA AKHIR

PREDIKSI CHURN DAN SEGMENTASI PELANGGAN MENGGUNAKAN BACKPROPAGATION NEURAL NETWORK BERBASIS EVOLUTION STRATEGIES

Junta Zeniarja¹⁾, Ardytha Luthfiarta²⁾

Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang

junta@dsn.dinus.ac.id¹⁾, ardytha.luthfi@dsn.dinus.ac.id²⁾

Abstract

Customer is an important part in ensuring the triumph and survival of a company. Therefore it is necessary to have a management system to ensure customers remain loyal and do not move to another competitor, known as churn management. Customer churn prediction is part of the churn management, that predicting customer behavior by classifying which customer is loyal and which one has tended to move to other competitors. The accuracy of these predictions is absolutely necessary because of the high level of customer migration to the company's competitors. This is important because the cost used to reach new customers is much higher as compared to maintain the existing customer loyalty. Even though a lot of study about customer churn prediction has been done, further research is still needed to improve the accuracy of the prediction. This study will discuss the use of data mining techniques Backpropagation Neural Network (BPNN) in hybrid with a Evolution Strategies (ES) for attribute weighting. Validation of the model was done by using the 10-Fold Cross validation and evaluation of measurement were done by using the confusion matrix and the Area Under the ROC Curve (AUC). The experimental results show that the hybrid BPNN with ES achieved better performance than the standard BPNN.

Keywords: *data mining, customer churn, prediction, backpropagation neural network, evolution strategies.*

1. PENDAHULUAN

Pelanggan adalah aset yang paling penting dari semua jenis bisnis. Prospek usaha hanya mungkin dapat dilakukan dengan kehadiran pelanggan yang puas yang selalu setia dan membangun hubungan mereka dengan perusahaan. Untuk alasan ini, perusahaan harus merencanakan dan menerapkan strategi untuk menciptakan pelanggan, umumnya dikenal sebagai Customer Relationship Management (CRM). K. Tsipitsis dan A. Chorianopoulos [1] mendefinisikan CRM sebagai strategi yang terkait dengan mempertahankan, mengelola, dan meningkatkan hubungan pelanggan setia dan langgeng.

Merujuk ke perspektif bisnis intelijen, proses manajemen *churn* dalam kerangka CRM terdiri dari dua bagian utama pemodelan analitis yang memprediksi bagi mereka yang cenderung *churn* atau tidak dan mendukung operator penyedia untuk membuat keputusan yang berharga dalam mempertahankan atau meningkatkan pelanggan baru. Oleh karena itu,

artikel ini difokuskan pada pertimbangan dalam prediksi pelanggan *churn*.

Prediksi *churn* pelanggan adalah bagian dari manajemen *churn*, yang memprediksi perilaku pelanggan dengan klasifikasi sebagai pelanggan setia dan mana yang cenderung untuk pindah ke kompetitor lain.

“Pelanggan *churn*” berarti kehilangan klien. Ini memiliki arti yang sama seperti gesekan pelanggan, pembelotan pelanggan, dan perputaran pelanggan. *Churn* pelanggan juga didefinisikan oleh Hung et al. (2006) di mana layanan nirkabel industri telekomunikasi yang umum digunakan dalam jangka gerakan pelanggan dari satu operator ke yang lain [2].

Keakuratan prediksi ini mutlak diperlukan karena tingginya tingkat migrasi pelanggan untuk perusahaan pesaing.

Manajemen *churn* merupakan tugas penting bagi perusahaan untuk mempertahankan pelanggan yang berharga. Riset pemasaran menunjukkan bahwa rata-rata

nilai pelanggan yang *churn* atau pindah ke pesaing lain dari perusahaan operator seluler adalah sekitar 2,2% per bulan. Hung et al. [2] menyebutkan bahwa ada sekitar 27% dari pelanggan hilang setiap tahun. Berdasarkan riset pasar, keadaan ini mendorong perusahaan untuk menyediakan biaya untuk dukungan penjualan, pemasaran, iklan, dan komisi untuk mendapatkan pelanggan layanan mobile dengan pelanggan baru adalah sekitar \$ 300 sampai \$ 600. Dengan demikian, biaya untuk mendapatkan pelanggan baru jauh lebih tinggi daripada mempertahankan yang baru dan karenanya, kemampuan untuk memprediksi *churn* pelanggan adalah suatu keharusan.

2. KAJIAN LITERATUR DAN PENGEMBANGAN HIPOTESIS

A. BPNN

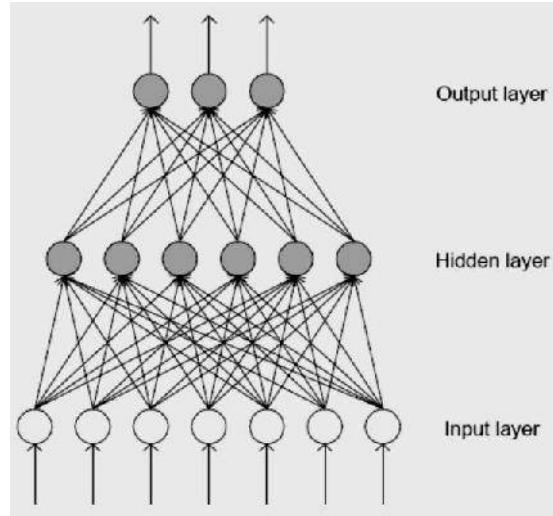
BPNN didefinisikan sebagai contoh dari suatu metode pembelajaran. Memiliki jaringan *feed forward* beberapa layer dengan bobot pola pas yang dapat digunakan untuk memodelkan beberapa diagram antara variabel set input dan output [3]. Arsitektur jaringan ditunjukkan pada Gambar 1 yang terdiri dari tiga unit output dan hidden layer tunggal, yang dapat dilatih menggunakan *back propagation*. Node yang diarsir pada gambar adalah unit pengolahan. Tanda panah yang menghubungkan input ke unit tersembunyi dan unit tersembunyi ke unit keluaran mewakili bobot.

B. ES

ES adalah pendekatan lain untuk mensimulasikan evolusi alam, yang telah diusulkan di Jerman pada awal 1960-an.

Ingo Rechenberg dan Hans-Paul Schwefel dari Technical University of Berlin adalah ES yang dikembangkan pertama kali. Keduanya bekerja terowongan angin dari Institut Teknik Aliran dalam penelitian. Berdasarkan penelitian melelahkan dilakukan oleh mereka untuk menemukan solusi optimal, maka keduanya memutuskan untuk menggunakan perubahan acak dalam parameter untuk mendefinisikan bentuk berdasarkan mutasi alam. Oleh karena itu berdasarkan penelitian ini maka strategi evolusi diciptakan [4].

Tidak seperti GAs, ES hanya beroperasi sebagai operator mutasi. ES biasanya diterapkan dalam masalah optimasi teknis.



Gambar 1. Arsitektur Back Propagation

3. METODE PENELITIAN

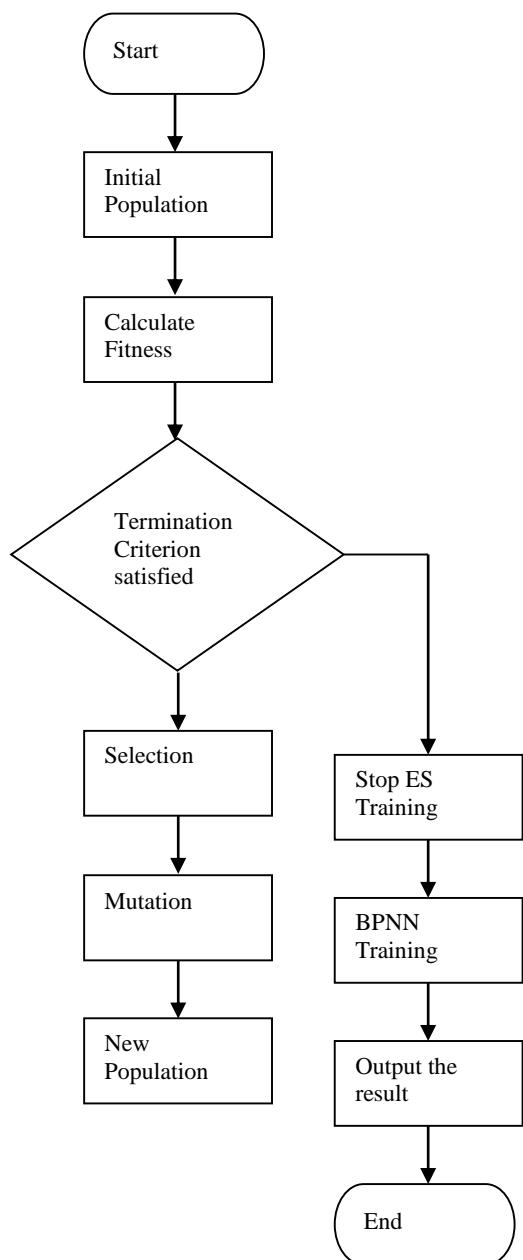
Ada empat metode penelitian yang umum digunakan yaitu penelitian tindakan, eksperimen, studi kasus, dan survey [5]. Penelitian eksperimen terdiri dari dua jenis, percobaan mutlak dan komparatif. Penelitian eksperimental umumnya dilakukan dalam memecahkan pengembangan, evaluasi, dan masalah proyek.

Penelitian ini akan mengadopsi metode penelitian eksperimen komparatif yang akan membandingkan dua model klasifikasi untuk prediksi *churn* pelanggan. Model pertama adalah dasar BPNN, dan model kedua adalah ES-BPNN.

A. Dataset

Untuk artikel ini, data set *churn* yang digunakan adalah UCI Repository dari Machine Learning Database di University of California, Irvine. Kumpulan data terdiri dari 20 variabel, dan berisi informasi dari sekitar 5000 pelanggan. Dari jumlah tersebut, 3333 diantaranya digunakan sebagai data training dan sisanya 1667 digunakan sebagai data uji. Indikasi apakah atau tidak pelanggan yang *churn* (meninggalkan perusahaan) juga termasuk dalam data.

B. Usulan Model



Gambar 2. Model ES-BPNN

C. Metode Evaluasi

Akurasi kinerja untuk memprediksi *churn* pelanggan dievaluasi oleh pengukuran evaluasi. Sejak BPNN adalah bagian dari teknik klasifikasi data mining, tabel confusion matrix dan ROC Curve (AUC) akan digunakan untuk mengukur akurasi model yang diusulkan [6].

D. Alat Penelitian

Alat penelitian yang digunakan untuk mendukung rancangan percobaan disajikan dalam artikel ini. Alat penelitian terdiri dari perangkat keras dan perangkat lunak.

Kebutuhan hardware yang digunakan untuk menjalankan penelitian ini terdiri dari:

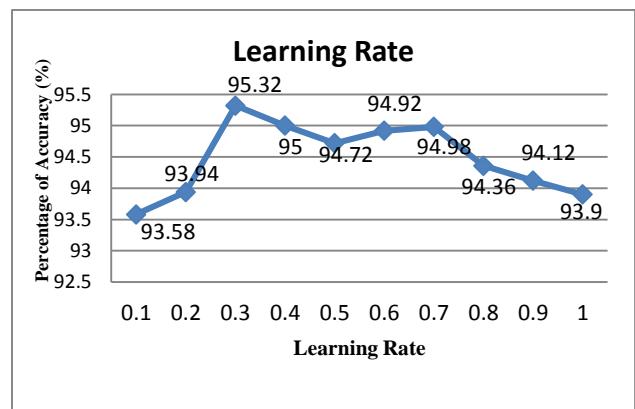
OS : Windows 7 Ultimate 32-bit
Processor : Intel® Core™ i3 370 Mobile
RAM : Sodimm 4 GB DDR3 10600 Mhz
GPU : ATI Mobility HD5470 512 MB

Software yang digunakan dalam penulisan ini adalah RapidMiner Versi 5.3. Persyaratan untuk menggunakan RapidMiner adalah Java Runtime Environment (JRE) versi 1.5 (resmi Java 5.0) atau lebih tinggi yang harus diinstal pada sistem.

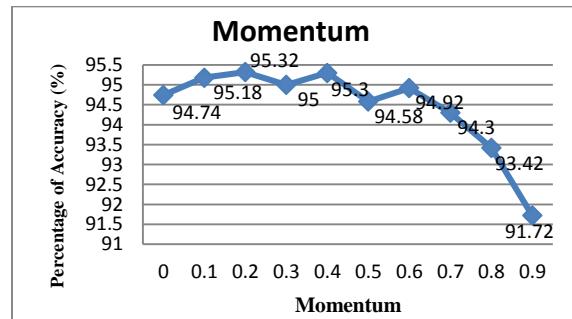
4. HASIL DAN PEMBAHASAN

A. Backpropagation Training Network

Nilai dari siklus pelatihan jaringan saraf ditentukan oleh trial and error, dalam hal ini dengan memasukkan nilai antara 50 sampai 1000. Metode yang sama juga digunakan untuk jumlah dan ukuran Hidden layer sebagai parameter kedua. Nilai default 0,3 dan 0,2 digunakan untuk tingkat pembelajaran dan momentum masing-masing. Nilai – nilai yang mampu memberikan akurasi terbaik untuk pemilihan parameter jaringan saraf berikutnya dipilih.



Gambar 3. Nilai Learning Rate



Gambar 4. Nilai Momentum

Nilai *Learning Rate* dipilih dengan memasukkan nilai *Learning Rate* antara 0,1 sampai 0,9 dengan standar momentum 0,3. Nilai dari *training cycle* dan *hidden layer* dipilih berdasarkan percobaan sebelumnya. Hasil percobaan ditunjukkan dalam gambar 3 dan Gambar 4. Dari hasil, akurasi terbaik dicapai dengan *Learning Rate* 0,3.

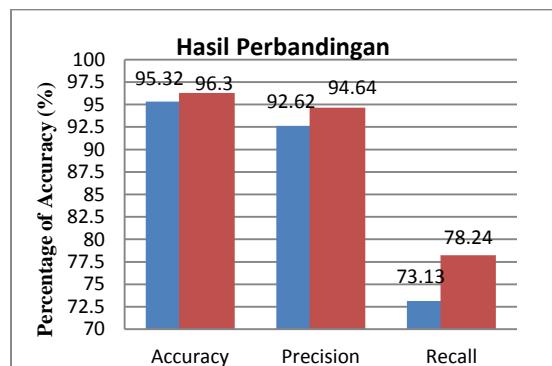
Metode yang sama diterapkan untuk memilih nilai momentum. Dengan memperbaiki nilai *Learning Rate* berdasarkan hasil percobaan untuk memilih harga *Learning Rate*, yang *Learning Rate*-nya adalah 0,3.

B. ES-BPNN

Tabel 1. Hasil Perbandingan

Comparison	Basic BPNN	ES-BPNN
Accuracy (Confusion Matrix)	95.32%	96.30%
Precision (Confusion Matrix)	92.62%	94.64%
Recall (Confusion Matrix)	73.13%	78.24%
Accuracy (AUC)	0.892	0.9
Execution time:	3 m. 52 s	5h. 24m. 06s

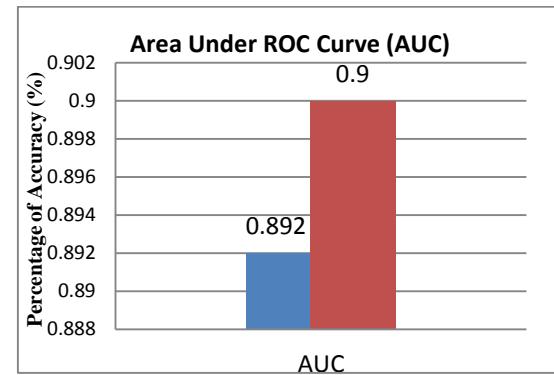
Pada tabel 1 di atas merupakan perbandingan akurasi antara *Basic BPNN* dan *ES-BPNN*. Hal ini menunjukkan bahwa *ES-BPNN* mengungguli di semua pengujian. Tapi waktu komputasi untuk *ES-BPNN* sangat panjang yaitu 5 jam 24 menit dan 06 detik, dibandingkan dengan *BPNN* dasar yang hanya membutuhkan waktu 3 menit 52 detik.



Gambar 5. Hasil Perbandingan Akurasi dari Basic BPNN dengan ES-BPNN

Gambar 5 menunjukkan hasil perbandingan nilai *accuracy*, *precision* dan *recall*. Dimana warna biru sebagai *Basic BPNN* dan warna merah sebagai *ES-BPNN*. Dari

diagram diatas menunjukkan *ES-BPNN* lebih baik dibandingkan dengan *Basic BPNN* karena nilai *accuracy*, *precision* dan *recall* yang lebih tinggi yaitu 96.3%, 94.64% dan 78.24%.



Gambar 6. Hasil Perbandingan kurva ROC (AUC) dari Basic BPNN dengan ES-BPNN

Gambar 6 menunjukkan hasil perbandingan kurva ROC (AUC). Dimana warna biru sebagai *Basic BPNN* dan warna merah sebagai *ES-BPNN*. Dari diagram diatas menunjukkan *ES-BPNN* lebih baik dibandingkan dengan *Basic BPNN* karena nilai *AUC* yang lebih tinggi yaitu 0.9.

5. KESIMPULAN

A. Kesimpulan

Model yang diusulkan dari *ES-BPNN* menghasilkan akurasi yang lebih baik dengan 96,30% dibandingkan dengan *Basic BPNN* dengan 95,32%. Tapi waktu komputasi untuk *ES-BPNN* sangat panjang yaitu 5 jam 24 menit 6 detik, dibandingkan dengan *Basic BPNN* yang hanya membutuhkan waktu 3 menit 52 detik.

Meskipun *ES-BPNN* mencapai hasil yang lebih baik, tetapi waktu eksekusinya lebih lama. Kita bisa menggunakan *ES-BPNN* untuk memprediksi klasifikasi secara offline yang hasilnya membutuhkan akurasi yang lebih tinggi tanpa mempertimbangkan waktu eksekusi. Selanjutnya jika kasus tersebut untuk prediksi klasifikasi seperti jaringan intrusion detection online, *Basic BPNN* lebih handal.

B. Future Work

Penelitian ini memiliki beberapa keterbatasan, maka ada beberapa masalah dapat dipertimbangkan untuk pekerjaan di masa depan antara lain :

- 1) Meskipun *ES-BPNN* mencapai hasil yang lebih baik tapi waktu eksekusi yang bersangkutan, untuk pekerjaan di masa depan dapat diterapkan *ES-BPNN* dengan waktu yang lebih pendek.
- 2) Langkah Preprocessing dalam data mining memiliki bagian penting untuk meningkatkan akurasi kinerja sehingga pekerjaan berikutnya seleksi fitur dapat diterapkan.
- 3) Penelitian ini membandingkan *Back-Propagation Neural Network* dengan kombinasi *Evolutionary Strategies - Backpropagation Neural Network*, untuk masa depan dapat menggunakan teknik klasifikasi data mining lainnya seperti *Decision Tree (DT)*, *Support Vector Machine (SVM)* dapat diterapkan untuk *Evolutionary Strategies* atau teknik optimasi lain seperti *Genetic Algorithm (GA)*, *Particle Swarm Optimization (PSO)*.
- 4) Penelitian ini menggunakan *UCI Churn Dataset* dari *learning repository UCI Machine*, untuk pekerjaan di masa depan dataset publik lainnya tentang *churn* pelanggan dapat diterapkan seperti data set dari *Terdata Duke University* atau *ACM Piala KDD Cup*.

6. REFERENSI

- [1] K. Tsipitsis and A. Chorianopoulos, Data Mining Techniques in CRM. Chichester, UK: John Wiley & Sons, Ltd, 2010.
- [2] S.-Y. Hung, D. C. Yen, and H.-Y. Wang, “Applying data mining to telecom churn management,” *Expert Systems with Applications*, vol. 31, no. 3, pp. 515–524, Oct. 2006.
- [3] I. Witten and E. Frank, “Data Mining: Practical machine learning tools and techniques,” 2011.
- [4] M. Negnevitsky, Artificial Intelligence : A Guide to Intelligent Systems. Pearson Education Limited, 2005.
- [5] W. Dawson, Projects in Computing and Information Systems A Student Guide. 2009.
- [6] Gorunescu, Data Mining Concepts, Models and Techniques. 2011.

LAMPIRAN 5 : LAPORAN PENGGUNAAN DANA AKHIR

Rekapitulasi Penggunaan Dana Penelitian

Judul	: Data Mining pada Proses Penentuan Prediksi Churn dan Segmentasi Pelanggan Menggunakan Backpropagation Neural Network berbasis Evolution Strategies
Skema Hibah	: Penelitian Dosen Pemula
Peneliti / Pelaksana	
Nama Ketua	: JUNTA ZENIARJA M.Kom
Perguruan Tinggi	Universitas Dian Nuswantoro
NIDN	: 0606068802
Nama Anggota (1)	: ARDYTHA LUTHFIARTA M.Kom
Tahun Pelaksanaan	: Tahun ke 1 dari rencana 1 tahun
Dana Tahun Berjalan	: Rp 10.000.000,00
Dana Mulai Diterima Tanggal	: 2014-06-27

Rincian Penggunaan

1. HONOR OUTPUT KEGIATAN				
Item Honor	Volume	Satuan	Honor/Jam (Rp)	Total (Rp)
1. Pajak Penghasilan (Pph21 5%)	2.00	orang	75.000	150.000
2. Pajak Pertambahan Nilai (PPN 10%)	1.00	orang	636.364	636.364
3. Honor Ketua Peneliti 70%	1.00	orang	1.260.000	1.260.000
4. Honor Anggota Peneliti 70%	1.00	orang	840.000	840.000
5. Honor Ketua Peneliti 30%	1.00	orang	540.000	540.000
6. Honor Anggota Peneliti 30%	1.00	orang	360.000	360.000
Sub Total (Rp)				3.786.364,00
2. BELANJA BAHAN				
Item Bahan	Volume	Satuan	Harga Satuan (Rp)	Total (Rp)
1. Belanja FlashDisk	1.00	paket	100.000	100.000
2. Belanja Cartridge Hitam	1.00	buah	185.000	185.000
3. Belanja Tinta Printer	1.00	buah	30.250	30.250
4. Refill Inkjet Cartridge	1.00	paket	33.000	33.000
5. Belanja FlashDisk	2.00	buah	110.000	220.000
6. ATK	1.00	paket	144.500	144.500
7. Jilid Laporan	1.00	paket	54.000	54.000
8. Cartridge dan Refill Tinta	1.00	paket	462.000	462.000
9. ATK	1.00	paket	74.300	74.300

				Sub Total (Rp) 1.303.050,00
3. BELANJA BARANG NON OPERASIONAL LAINNYA				
Item Barang	Volume	Satuan	Harga Satuan (Rp)	Total (Rp)
1. Konsumsi Rapat	1.00	paket	13.500	13.500
2. Pulsa dan Internet Bulan Mei	2.00	paket	200.000	400.000
3. Pulsa dan Internet Bulan Juni	2.00	paket	200.000	400.000
4. Pulsa dan Internet Bulan Juli	2.00	paket	200.000	400.000
5. Konsumsi Rapat	1.00	paket	158.000	158.000
6. Konsumsi Rapat	1.00	paket	55.500	55.500
7. Konsumsi Makan 1	1.00	paket	43.500	43.500
8. Konsumsi Makan 2	1.00	paket	211.000	211.000
9. Konsumsi Pulsa dan Internet Bulan Agustus	2.00	paket	200.000	400.000
10. Konsumsi Makan 1	1.00	paket	28.500	28.500
11. Konsumsi Makan 2	1.00	paket	35.000	35.000
12. Konsumsi Snak	1.00	paket	31.700	31.700
13. Konsumsi Minuman	1.00	paket	21.000	21.000
14. Konsumsi Pulsa dan Internet Bulan September	2.00	paket	200.000	400.000
15. Konsumsi Snak	1.00	paket	18.500	18.500
16. Konsumsi Makan 1	1.00	paket	31.000	31.000
17. Konsumsi Makan 2	1.00	paket	51.000	51.000
18. Konsumsi Makan 3	1.00	paket	55.500	55.500
19. Konsumsi Pulsa dan Internet Bulan Oktober	2.00	paket	200.000	400.000
20. Konsumsi Makan 1	1.00	paket	51.000	51.000
21. Konsumsi Minuman	1.00	paket	26.000	26.000
22. Konsumsi Snak	1.00	paket	23.500	23.500
23. Konsumsi Pulsa dan Internet Bulan Nopember	2.00	paket	200.000	400.000
			Sub Total (Rp) 3.654.200,00	

4. BELANJA PERJALANAN LAINNYA				
Item Perjalanan	Volume	Satuan	Biaya Satuan (Rp)	Total (Rp)
1. Biaya Bensin 1	1.00	kali	15.000	15.000
2. Biaya Bensin 2	1.00	kali	9.295	9.295
3. Biaya Bensin 1	1.00	kali	20.000	20.000
4. Biaya Bensin 2	1.00	kali	25.000	25.000
5. Biaya Bensin 3	1.00	kali	15.000	15.000
6. Biaya Bensin 4	1.00	kali	15.000	15.000
7. Biaya Bensin 1	1.00	kali	20.000	20.000
8. Transportasi JKT	2.00	kali	222.500	445.000
9. Biaya Bensin 2	1.00	kali	15.000	15.000
10. Biaya Bensin 3	1.00	kali	20.000	20.000
11. Biaya Bensin 4	1.00	kali	15.000	15.000
12. Biaya Bensin 5	1.00	kali	19.825	19.825
13. Biaya Bensin 6	1.00	kali	20.000	20.000
14. Biaya Bensin 1	1.00	kali	19.435	19.435
15. Biaya Bensin 2	1.00	kali	20.000	20.000
16. Biaya Bensin 3	1.00	kali	18.298	18.298
17. Biaya Bensin 4	1.00	kali	15.000	15.000
18. Biaya Bensin 5	1.00	kali	20.000	20.000
19. Biaya Bensin 6	1.00	kali	20.000	20.000
20. Biaya Bensin 7	1.00	kali	20.000	20.000
21. Biaya Bensin 8	1.00	kali	20.000	20.000
22. Biaya Bensin 9	1.00	kali	20.000	20.000
23. Biaya Bensin 1	1.00	kali	50.000	50.000
24. Biaya Bensin 2	1.00	kali	100.000	100.000
25. Biaya Bensin 3	1.00	kali	100.000	100.000
26. Biaya Bensin 4	1.00	kali	50.000	50.000
27. Biaya Bensin	3.00	kali	50.000	150.000
Sub Total (Rp)				1.276.853,00

Total Pengeluaran Dalam Satu Tahun (Rp) 10.020.467,00



Semarang, 2 - 11 - 2014
Ketua,

(JUNTA ZENIARJA M.Kom)
NIP/NIK 0686.11.2012.459