Analisis Dan Implementasi Algoritma Neural Network Pada Classification Dalam Melakukan Prediksi Pelanggan Churn Telkom Speedy

Fauzan Azhim Umsohi[1], Jondri Drs., MT.[2], S iti S a' ad ah , ST., MT.[3]
[1,2,3] Departemen Informatika – Universitas Telkom
[1]fauz2117@gmail.com, [2]suryadi_joni@yahoo.com,[3] sitisaadah@telkomuniversity.ac.id

Penyedia layanan internet atau yang lebih dikenal sebagai *internet service provider* (ISP) saat ini semakin bertambah. Dari awalnya pelanggan di dominasi oleh satu *provider* sekarang sudah berganti dengan *provider* lain. Tentunya masih banyak lagi faktor lainnya, kejadian ini dinamakan *churn*. Untuk itu perlu dilakukan tindakan antisipasi untuk mengatasi hal tersebut dimana yang pada tulisan ini dengan melakukan klasifikasi prediksi *churn*.Data yang digunakan bersifat *non-linier* oleh karena itu pada tulisan ini digunakan metode *Single-layered Supervised Neural Network's* yang dikenal *SSNNs*. Dimana metode ini memang ditujukan untuk data semacam ini [1]. Dengan parameter-parameter optimal yang dihasilkan saat pengujian sistem, didapat akurasi prediksi sebesar 72,35%.

Kata Kunci: ***churn, Neural Network, klasifikasi, SSNNs***.

Abstract
Internet service provider, or better known as the ISP is currently increasing overtime. Initially ISP is dominated by one provider but now, it's already been replaced by the other providers. Of course provider turnover were caused by several other factors, such event was knows as churn. Of course given this kind of situation, churn prediction is necessary which is used this time. The data used contain non-linear data, for that the method used was Single-layered Supervised Neural Network's as known as SSNNs.This method was intended for this kind of data [1]. With the optimal parameters used in testing phase, the obtained accuracy from the predicted churn was 72,35%.

Keywords *: Churn, Neural Network, Classification, SSNNs*

BAB 1
PENDAHULUAN

1.1 Latar Belakang

Telkom merupakan salah satu perusahaan besar di Indonesia yang berkecimpung dibanyak bidang salah satunya Telkom memiliki layanan internet yang bernama Telkom Speedy. Dengan semakin berkembangnya infrastruktur teknologi telekomunikasi hadirlah *provider* lain di Indonesia yang menawarkan teknologi terbaru dengan kualitas yang sama bahkan lebih baik. Dengan bertambahnya saingan, Speedy mengalami penurunan jumlah pelanggan dengan dampak terbesar akibat ketatnya persaingan dengan *provider* lain lalu pemutusan layanan baik dari perusahaan maupun pengguna. Pemutusan layanan inilah yang dinamakan dengan *Churn.*Untuk mengantisipasi hal tersebut penulis pada tulisan ini melakukan prediksi *churn.*

Atribut - atribut yang digunakan dalam penelitian ini berupa data *demographic* pelanggan, data pemakaian, data gangguan, keluhan dan data layanan. Data tersebut akan diekspansi menjadi 55 atribut yang tentunya merupakan data yang tidak dapat dipisahkan secara linier. Untuk itulah penerapan algoritma kumpulan *Neural Network single layer* dengan metode pembelajaran *supervised* yang selanjutnya ditulis pada penelitian ini dengan *Single-layer Supervised Neural Network* (*SSNNs*) digunakan sebagaimana pengembangan tersebut mempertahankan kelebihan *Neural Network single layer* daripada *Neural Network* lain (*Neural Network multi layer*) yaitu kecepatannya dalam memproses data dan juga menutup kelemahan dari *Neural Network single layer* biasa yang berupa ketidak mampuannya untuk menangani data yang bersifat *non-linear* seperti pada data dari kasus yang digunakan pada tulisan ini [13].

Didalam melakukan prediksi *churn,* tentunya terdapat beberapa parameter utama yang digunakan, dimana dalam tulisan ini digunakan

learning rate dan *epoch*. Dengan melakukan pengujian-pengujian dengan parameter-parameter tersebut akan dicari akurasi yang peling optimal sehingga hasil prediksi inipun dapat berguna.

1.2 Perumusan Masalah

Rumusan masalah berdasarkan judul dan latar belakang :

1. Benarkah penerapan *SSNNs* dalam klasifikasi dapat menyelesaikan data *non-linear*?
2. Bagaimana cara kerja dari penerapan *neural network* untuk melakukan klasifikasi mengenai pengguna layanan agar dapat melakukan prediksi pelanggan yang berpotensi melakukan *churn* ?

1.3 Tujuan

Tujuan dilakukannya penilitian terhadap penggunaan neural network pada rule mining :

1. Melakukan analisa dari penerapan *neural network* dalam klasifikasi. Dimana hal ini dapat dilakukan dengan melakukan mencari nilai optimal dari tiap parameter yang selanjutnya digunakan untuk perhitungan akurasi dari prediksi yang dihasilkan.
2. Menghasilkan klasifikasi pengguna layanan yang nantinya berguna dan digunakan sebagai pertimbangan untuk melakukan tindakan antisipasi.

1.4 Hipotesa

Neural Network sudah dikenal dengan kemampuannya dalam menemukan pola secara efisien dalam data yang sangat banyak sekalipun yang berarti *SSNNs* yang diterapkan pada klasifikasi dapat berjalan dan mengklasifikasikan dengan baik. Hasil *output* dari sistem ini dapat menunjukan pelanggan mana yang berpotensi melakukan *churn*. Hal ini tentunya akan berguna dan tentunya dapat digunakan untuk melakukan tindakan antisipasi.

BAB II TINJAUAN PUSTAKA

2.1 Churn

Churn merupakan aktfitas pengguna layanan dari suatu *provider* untuk berhenti menggunakan layanan baik memang pengguna tersebut berhenti atau pindah menggunakan *provider* lain. Hal tersebut akan sangat merugikan bagi *provider* untuk itulah hal ini harus bisa diantisipasi.

Beberapa penyebab terjadinya *churn* secara umum dapat disebabkan oleh *provider* maupun oleh pengguna. Dari sisi *provider* layanan dapat dihentikan bila pengguna layan tidak mengikuti ketentuan – ketentuan yang telah ditetapkan oleh *provider* misalnya pengguna telat terlalu banyak dalam melakukan pembayaran. Sedangkan dari sisi pengguna seperti yang telah disebutkan sebelumnya yaitu ketika pengguna pindah *provider* atau pengguna pindah tempat tinggal.

2.2 Data Mining

Terdiri dari *Data* dan *Mining*, dimana data sendiri berarti kumpulan fakta yang belum memiliki arti sehingga harus melewati proses pengolahan agar dapat memiliki arti tersendiri bagi yang menggunakan data tersebut. Sedangkan *mining* sendiri merupakan penggalian yang merupakan sebuah proses menggali dari sesuatu. Sehingga *Data Mining* adalah proses menggali kumpulan data atau fakta untuk memberikan informasi yang berguna bagi pengguna data tersebut. Tentunya hasil penggalian tersebut mendapatkan pola dan pola inilah yang digunakan sebagai informasi penting.

Data mining sendiri [5] erat keterkaiyannya dengan bidang ilmu – ilmu lain, seperti *database system*, *data warehousing*, statistik, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Oleh karena itu pada penelitian ini diterapkan pada *neural network*. *Data mining* yang didefinisikan sebagai proses menemukan pola-pola dalam data merupakan suatu proses yang bersifat otomatis atau semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar. Berikut ini merupakan karakteristik dari data mining [5]:

- *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- *Data mining* biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.

- *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi.

Dapat ditarik kesimpulan dari beberapa keterangan diatas bahwa [5] *data mining* adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (*database*) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. Kata mining sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu *data mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan *database*. Beberapa metode yang sering disebut-sebut dalam literatur *data mining* antara lain *clustering, classification, association rules mining, neural network, genetic algorithm* dan lain-lain.

2.3 Klasifikasi Data

Merupakan suatu pengelompokkan dari data. Berdasarkan Kamus Besar Bahasa Indonesia klasifikasi merupakan penyusunan bersistem dalam kelompok atau golongan menurut kaidah atau standar yang ditetapkan. Bila dikaitkan dengan data maka klasifikasi dapat berupa proses pengelompokkan data dengan atribut yang sama, dan juga membedakan data dengan atribut yang tidak sama.

2.4 Neural Network

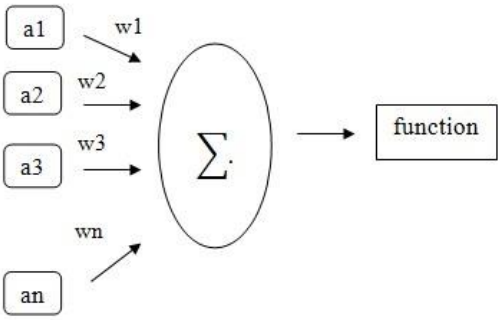
Neural Network [14] merupakan permodelan dari sistem syaraf otak manusia dimana hal ini dibuat berdasarkan kemampuan otak manusia dalam mengatur sel–sel yang dikenal sebagai *neuron*, sehingga dapat digunakan untuk menyelesaikan tugas yang diberikan seperti mengenali pola.

Pada strukturnya sendiri [14] *neural network* memiliki struktur paralel yang tersebar dan mempunyai kemampuan untuk belajar, dimana hal ini membuat *neural network* dapat menghasilkan generalisasi agar tercipta *output* yang benar untuk *input* yang baru atau belum dilakukan *training*. Hal ini membuat *neural network* dapat menyelesaikan masalah yang kompleks sekalipun.

Di dalam pembangunan modelnya *neural network* memiliki 3 elemen dasar yaitu [14] :

1. Jalur hubungan yang memiliki bobot (w).
2. *Summing function* yang menjumlahkan input yang telah diberi bobot dari jalur hubungan (∑).
3. Fungsi aktivasi yang membatasi output dari *neuron* (*function).*

Diamana permodelan ini dapat digambarkan seperti dibawah :



Gambar 1 Komponen Dasar Neural Network

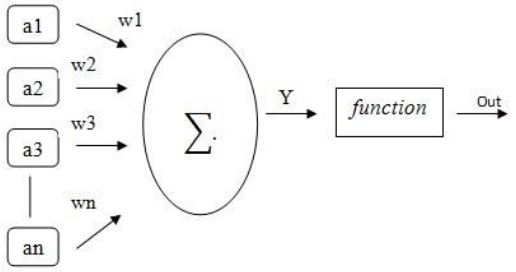
2.5 Supervised Learning

Seperti namanya *supervised* yang berarti diawasi dimana pada *neural netwrok* hal ini digambarkan dengan kumpulan sampel *input-output* yang memiliki pengetahuan tentang hasil yang diinginkan dari terbentukan *neural netwrok* yang dibangun. Hal ini menyebabkan metode *supervised learning* dapat memberikan respon seperti yang diinginkan oleh *neural netwrok*, dimana respon tersebut merepresentasikan aksi optimum dari suatu *neural netwrok.*

Di dalam metode ini [13] terdapat parameter-parameter jaringan yang berubah berdasarkan vektor latih dan sinyal kesalahan dimana skema ini akan dilakukan terus-menerus agan *neural netwrok* dapat memiliki kemampuan untuk mengawasi dengan baik. Hal ini tentunya akan membuat hasil dari sistem memiliki akurasi yang tinggi.

2.6 Single-layered Neural Network with Supervised Learning (SSNN)

SSNN (*Neural Network single layer* dengan metode pembelajaran *supervised*) memiliki model seperti berikut :



Gambar 2 Struktur Neural Network Single Layered

Dimana a merupakan data input w merupakan bobot yang diambil secara *random*. Dengan melihat dari gambar tersebut dapat diartikan dengan nilai input dikalikan dengan bobot yang hasilnya akan dijumlahkan menjadi y melalui persamaan [14] :

$$y = \sum_{\text{[illegible]}=1}^{n} \text{[illegible]}$$

Dimana y merupakan hasil dari jumlah hasil perkalian data dengan bobot ke-1 sampai data ke-n.
Hasil dari penjumlahan tersebut akan melalui fungsi yang merupakan fungsi sigmoid dengan persamaan sebagai berikut [14] :

$$\text{Out} = \frac{1}{1+\exp(-\text{[illegible]})}$$

Dimana y merupakan hasil dari persamaan sebelumnya dan out merupakan hasil klasifikasi yang menentukan *churn* atau tidaknya suatu data pelanggan.

Sehingga menghasilkan *SSNN* set yang telah terdapat hasil justifikasinya pada langkah selanjutnya. Berikut akan dijelaskan langkah–langkah algoritma learning *SSNN* dimana parameter inputnya yaitu :

T	: data *training*
Emin	: toleransi *error*
α	: *learning rate*
max_Epoch	: jumlah maksimal iterasi dalam proses *learning*

berikut merupakan penjelasan mengenai prosesnya

1. Mengambil data *training T* dengan jumlah atribut *n* dan jumlah baris *r*
2. M bobot w sejumlah n atribut yang nantinya akan menghasilkan *SSNN*
3. Lakukan pengulangan selama *error* > Emin dan iterasi < maxEpoch
4. *Update* bobot tiap iterasi dengan menggunakan fungsi

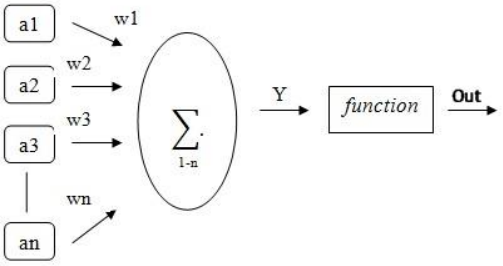
$$W^{j+1} = W^{j} + \alpha E^{j} X,$$
$$E^{j} = (t - y)$$

   Dimana W^j adalah bobot awal, t merupakan target *output*, y adalah *output* aktual, g merupakan fungsi aktivasi pada *neural network*, dan X adalah data input.
5. *Update* nilai *error*, lakukan kembali Step 3

2.7 SSNN Untuk Data Non-liniearly Separable (SSNNs)

Hanya *SSNN* saja tidak dapat memodelkan data yang memiliki hubungan *non-linearly separable*. Untuk itulah digunakan lebih dari satu *SSNN* untuk melakukan permodelan data yang memiliki hubungan yang bersifat *non-linearly separable* tersebut. Struktur permodelan ini dapat melakukan *mining* dengan cepat karena strukturnya yang merupakan implementasi paralel dari *SSNNs*, strukturnya simpel dan cepat dalam menanggapi perubahan dinamis.

Semakin banyak *SSNN* yang terbentuk maka semakin baik model yang dihasilkan dari seluruh data yang ada. Dimana banyaknya *SSNN* akan meng-*cover* seluruh data.



Gambar 3 Struktur Neural Network Single Layered Supervised Untuk Data Non-liniearly Separable (SSNNs)

2.8 Confusion Matrix

Sebagai alat ukur dari model klasifikasi ini diperlukan *confusion matrix*, yang merupakan alat ukur akurasi yang dapat mengukur dengan tepat walaupun dengan kelas yang berbeda [1]. Misal diketahui terdapat 2 kelas A dan B. Untuk mengukur model prediksi dari A akan tercipta matriks sebagai berikut :

Tabel 1 Sampel Confusion Matriks

		Prediksi	
		A	B
Data	A	25	1
	B	2	22

Matriks tersebut menunjukan bagaimana prediksi yang dibuat oleh model. Dimana baris menunjukkan kelas yang diketahui dan kolom menunjukkan kelas yang diprediksi. Nilai tiap anggota dalam matriks tersebut adalah jumlah prediksi yang dibuat oleh kelas yang berkaitan sebagai contoh dengan jumlah nilai yang benar. Nilai pada diagonal matriks tersebut menunjukan jumlah klasifikasi yang benar sebaliknya, selain yang terletak pada diagonal menunjukan bahwa jumlah tersebut adalah jumlah error yang dibuat. Untuk pembagiannya bisa dilihat pada tabel dibawah

Tabel 2 Struktur Confusion Matriks

		Prediksi	
		A	B
Data	A	TP	FN
	B	FP	TN

Dimana *TP* merupakan *True Positif*, *TN* adalah *True negatif*, *FP* adalah *False Positif*, dan *FN* adalah *False Negatif*.

Atribut dari matriks tersebut digunakan untuk menghitung nilai *precision* dan *recall* dimana :

$$Precision = tp/(tp+fp)$$

Sedangkan untuk *recall :*

$$Recall = tp/(tp+fn)$$

Nantinya hasil dari *precision* dan *recall* digunakan untuk menghitung *F-measure* yang merupakan perhitungan akurasi yang benar yang diketahui sebagai berikut :

$$F\text{-}measure = 2PR/(P+R)$$

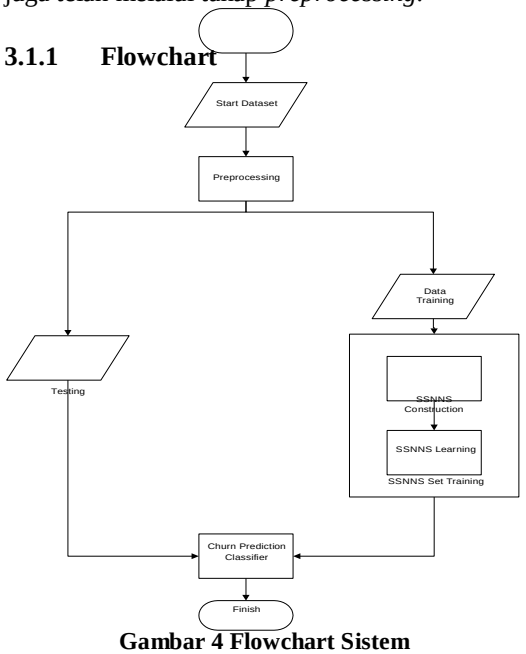
Dimana P merupakan *precision* dan R merupakan *recall.*

BAB III PERANCANGAN SISTEM

3.1 Gambaran Umum Sistem

Pada tahapan awal tugas akhir ini untuk melakukan klasifikasi dengan penambahan penerapan dari *neural network*, data yang digunakan akan melalui beberapa tahap yaitu *preprocessing* untuk menyesuaikan data sehingga bisa dimasukkan ke dalam sistem, *SSNNs Set Training* dimana akan dilakukan pembuatan *SSNNs* yang kemudian dilakukan pembelajaran terhadap data latih hasil *preprocessing* sebelumnya sehingga jadilah *SSNNs Set* yang selanjutnya akan menjadi *classifier* pada data *testing* yang sebelumnya juga telah melalui tahap *preprocessing*.

3.1.1 Flowchart



Gambar 4 Flowchart Sistem

3.1.2 Ilustrasi Preprocessing

Pada tahap preprocessing *data set* akan diolah dengan ilustrasi sebagai berikut dengan contoh dari data pelajar dalam 1 pelajaran:

Tabel 3 Contoh Data

	Sampel	**Kategori**
Nama	Tes	Nama
Duration	Short	Short,intermediate, long
Grade	Medium	Low, medium, high
Assignment	High	Low, medium, high
Quiz	Low	Low, medium, high

Akan dilakukan pengkategorian menjadi seperti berikut :

Tabel 4 Sampel Normalisasi Data

	Category 1	Category 2	Category 3	Category 4	Category 5	Category 6	Category 7	Category 8	Category 9	Category 10	Category 11	Category 12
Sampel\ Atribut	D:Short	D:Med	D:Long	G:Low	G:Med	G:High	A:Low	A:Med	A:High	Q:Low	Q:Med	Q:High
Tes	1	0	0	0	1	0	0	0	1	1	0	0

Maka untuk pelajar dengan nama tes dapat dikategorikan sebagai {1,5,9,10} yang akan menjadi salah satu *item* dalam suatu *itemset* yang nantinya akan diolah di dalam *Neural Network*.

3.1.3 Keterangan Dan Seleksi Atribut

Tabel 5 Pembagian Atribut

No	Nama Field	Keterangan
1	Line Cats (layanan)	Jenis Layanan Pelanggan (100 = KBU , 101 = rumah , 103 = perusahaan , 200 = sosial , 400 = bisnis, 701 =) (7 kategori)
2	PLBLCL (demographic)	Segmentasi Pelanggan (PL = Pelanggan Retail, BL = Pelanggan Bisnis, CL = Pelanggan Coorporate) (3 kategori)
3	CMDF (demographic)	Area Pelanggan (14 CMDF : CMDF Tegal lega, CMDF Bandung Barat, CMDF Geger Kalong, dll) (14 kategori)
4	Jenis Layanan Speedy (layanan)	Produk Speedy yang digunakan (Nama 10 Produk Speedy : Familia, Socialia, Instan,dll) (10 kategori)
5	Periode (data pemakaian)	Masa berlangganan pelanggan (3 kategori)
6	Average Usage (data pemakaian)	Rata-rata pemakaian pelanggan (3 kategori)
7	Gagguan Ulang (gangguan)	Gangguan yang terjadi dalam retang waktu berdekatan (1 kategori)
8	Jumlah Gangguan(gangguan)	Jumlah Gangguan yang dialami pelanggan (2 kategori)
9	TTR (gangguan)	waktu pembenahan gangguan (1 kategori)
10	Keluhan (gangguan)	Keluhan Gangguan (Jenis Keluhan Pelanggan : Buka Isolir, Intermitten, Tidak bias connect, dll) (11 kategori)

3.1.4 Input Data

Data yang digunakan meliputi data pengguna *churn* dan yang tidak *churn*. Data yang digunakan pada memiliki sampel sebagai berikut :

Tabel 6 Sampel Data

CATS	PLBLCL	Area Pelanggan	Nama Produk	Jumlah Hari	Avg Usage	Gaul	Jumlah Gangguan	TTR	Keluhan
200	PL	MDF BD.TURANGGA	Speedy SOCIALIA 1 MB - HSSP	623	16.3949	0	0	0	
100	PL	MDF BD.TURANGGA	Speedy SOCIALIA 1 MB - HSSP	662	54.2251	0	1	86400	Petugas diminta Datang
100	CL	MDF BD.UTARA/H.MANAH	Speedy Paket LOAD	1003	5.24427	0	0	0	
100	BL	MDF BD.TURANGGA	Speedy Paket BIZ	1195	90.1079	1	2	86400	Tidak Bisa Connect
100	BL	MDF BD.BARAT/R.WALI	Speedy Paket LOAD	1435	32.0209	1	2	100000	

3.1.5 Normalisasi Data

Pada tahapan ini data – data yang digunakan di dalam input akan dirubah menjadi bentuk yang sesuai dengan algoritma yang digunakan dalam penelitian ini yaitu *SSNN*s. *SSNN*s yang dibuat pada penelitian ini dapat menerima input berupa data yang bertipe boolean dimana untuk mendapatkan data bertipe *boolean* tersebut data akan di *expand* sesuai dengan kategori masing – masing. Untuk kasus – kasus tertentu seperti data numerik perlu dilakukan diskritisasi sebagai berikut :

a. Atribut Periode

Periode_1 : Periode < 180
Periode_2 : 180 < Periode < 730
Periode_3 : Periode > 730

b. Atribut Usage

1. Untuk Pelanggan BL

Usage_1 : Usage < 50
Usage_2 : 50 < Usage < 100
Usage_3 : Usage > 100

2. Untuk Pelanggan CL

Usage_1 : Usage < 75
Usage_2 : 75 < Usage < 125
Usage_3 : Usage > 125

3. Untuk Pelanggan PL

Usage_1 : Usage < 25
Usage_2 : 25 < Usage < 75
Usage_3 : Usage > 75

c. Atribut Jumlah Gangguan

Ggn_1 : ggn = 1
Ggn_2 : ggn > 2

d. Atribut Time To Repair

TTR_PL : time < 28800
TTR_BL : time < 28800
TTR_CL : time < 86400

Setelah dilakukan kategorisasi seperti diatas, data – data akan di *expand* sehingga mendapatkan hasil sebagai berikut :

Tabel 7 Ekspansi Atribut

Nomor	Atribut	Nomor	Atribut
1	LINECATS = 100	30	PROD = SPEEDY FAMILIA
2	LINECATS = 101	31	PROD = SPEEDY LOAD
3	LINECATS = 103	32	PROD = SPEEDY MAIL
4	LINECATS = 200	33	PROD = SPEEDY SOCIALIA
5	LINECATS = 400	34	PERIOD < 180
6	LINECATS = 700	35	180 <= PERIOD <= 730
7	JENIS PELANGGAN = PL	36	PERIOD > 730
8	JENIS PELANGGAN = BL	37	USAGE = LOW
9	JENIS PELANGGAN = CL	38	USAGE = MEDIUM
10	MDF =BD.BARAT/RAJAWALI.	39	USAGE = HIGH
11	MDF = BD.CENTRUM	40	IS_GANGGUAN
12	MDF = BD.CIJAURA	41	GAUL = 0
13	MDF = BD.CIMAHI	42	GAUL > 0
14	MDF = BD.DAGO	43	JUMLAH GANGGUAN <= 1
15	MDF = BD.GEGERKALONG	44	JUMLAH GANGGUAN > 1
16	MDF = BD.KOPO SAYATI	45	TTR = COMPLY
17	MDF = BD.NANJUNG	46	TTR NOT COMPLY
18	MDF = BD.PANGALENGAN	47	Keluhan = Isolir
19	MDF = BD.TEGALEGA	48	Keluhan = Gangguan IP
20	MDF = BD.TIMUR	49	Keluhan = Intermitten
21	MDF= BD.TURANGGA	50	Keluhan = Jumlah Tagihan
22	MDF= BD.UJUNGBERUNG	51	Keluhan = Download Lambat
23	MDF= BD.UTARA/HAGERMANAH.	52	Keluhan = Lambat
24	PROD = SPEEDY NULL	53	Keluhan = Petugas Diminta Datang
25	PROD = SPEEDY INSTAN	54	Keluhan = Sharing PC tidak jalan
26	PROD = SPEEDY OFFICE	55	Keluhan = Tidak bisa browsing
27	PROD = SPEEDY BIZ	56	Keluhan = Tidak bisa connect
28	PROD = SPEEDY CHAT	57	Keluhan = Tidak bisa remote IP
29	PROD = SPEEDY EXECUTIVE		

Setelah itu akan dilakukan persilangan atribut sesuai dengan *value* dari masing – masing atribut sehingga didapat hasil sebagai berikut dimana 1 – 23 merupakan nomor urut atribut hasil ketegorisasi yang diambil sebagai sampel dari yang asli yaitu 57 nomor sesuai dengan tabel hasil ekspansi di atas. Berikut contoh hasil dari konversi *boolean*.

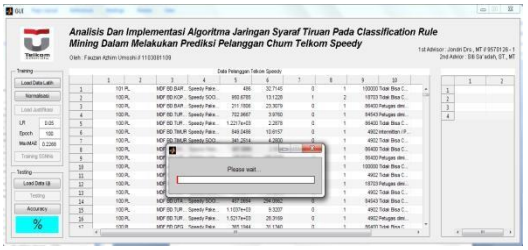
Tabel 8 Sampel Hasil Normalisasi

NoEA -> User	1	2	3	4	5	6	7	8	9	10	11	12	13	...	54	54	55	56	57
A	1	0	0	0	0	0	0	0	1	0	0	0	0	...	0	0	0	0	1
B	0	0	1	0	0	0	0	0	1	0	0	0	0	...	0	1	0	0	0
C	1	0	0	0	0	0	0	0	1	0	0	0	0	...	0	0	0	0	1
D	0	0	1	0	0	0	0	0	1	0	0	0	0	...	0	1	0	0	0
E	1	0	0	0	0	0	0	1	0	1	0	0	0	..	0	0	0	0	0
F	0	0	1	0	0	0	0	0	1	0	0	0	0	...	0	1	0	0	0
...																			

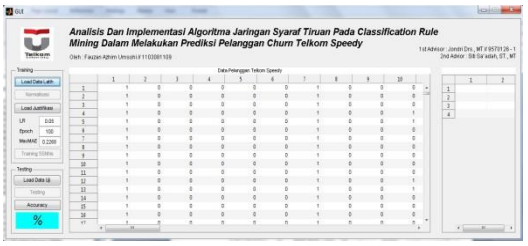
*NoEA : Nomor Ekspansi Atribut (1-57 pada tabel ekspansi)

3.1.6 Normalisasi Pada Sistem

Gambar 5 Load Data Sistem



Gambar 6 Normalisasi Pada Sistem

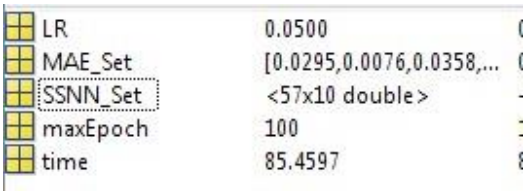


Gambar 7 Hasil Normalisasi Sistem

3.1.7 SSNNs Set Training

Melakukan training *SSNN*s dengan membangun dan malukan *learning* untuk melakukan testing *SSNN*s. Dimana pada tahap ini pengujian dengan parameter-parameter seperti *epoch* dan *learning rate* diujikan pada skema yang akan dibahas pada bab selanjutnya.

Gambar 8 Parameter Sistem



Gambar 9 SSNNs Pada Sistem

3.1.8 SSNNs Testing

Pada tahap ini akan dilakukan pengujian terhadap *data testing* dengan mengukur akurasi dari *classification rule mining* dengan cara menghitung menggunakan *confusion matrix.*
Dimana akan dilakukan penghitungan jumlah data True Positif, True Negatif, False Positif dan False Negatif untuk menentukan nilai *precision* dan *recall* sehingga dapat ditemukan hasil akurasi yang benar.

Tabel 9 Confusion Matrix

Data\Prediksi	Churn	Not Churn

Churn	True Positif	FalseNegatif
Not Churn	False Positif	True Negatif

Setelah dilakukan perhitungan nilai atribut dari matriks di atas barulah dilakukan pencarian akurasi dengan mendapatkan *F-measure.*

BAB IV IMPLEMENTASI, PENGUJIAN, DAN ANALISIS SISTEM

4.1 Implementasi

Berikut ini merupakan spesifikasi dari *hardware* dan *software* yang digunakan untuk melakukan uji coba sistem

4.1.1 Hardware

Laptop :

- Intel Core i3-3110M (Processor)
- 8 GB RAM (Memory)
- 500 GB (Hard Disk)

4.1.2 Software

Program pendukung:

- Windows 7 Ultimate 64-bit (Sistem Operasi Laptop)
- Notepad++ (Development tool)
- Matlab R2009a
- MS Office

4.2 Pengujian

Pada penelitian ini dilakukan beberapa skema yang berbeda untuk menguji pengaruhnya terhadap sistem yang antara lain sebagai berikut

Skema 1

Menguji nilai *learning rate* dengan mengukur akurasi saat melakukan *testing.*

Skema 2

Dilakukan menggunakan hasil optimal dari skema awal yang dilanjutkan dengan mencari nilai *epoch* maksimal. Hal ini dilakukan dengan mencoba bergam nilai *epoch* dan mengukur akurasinya.

Skema 3

Menguji pengaturan jumlah *summing* (*SSNNs*) menggunakan parameter yang didapat.

Skema 4

Menggunakan parameter – parameter yang didapat melalui skema 1, skema 2, dan skema 3 lalu melakukan klasifikasi.

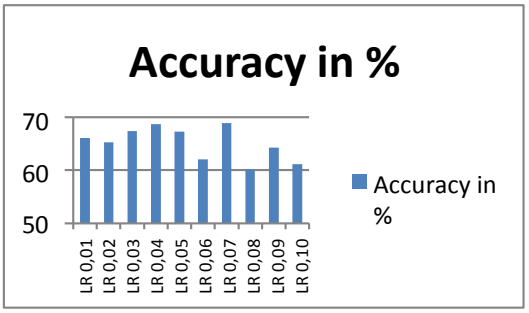
Skema 5

Mendapatkan akurasi dari parameter-parameter yang di dapat dari skema sebelumnya.

4.3 Analisis

Analisi yang dilakukan pada sistem dilakukan dengan perubahan parameter seperti yang sudah dijelaskan pada bagian sebelumnya pada bagian pengujian.

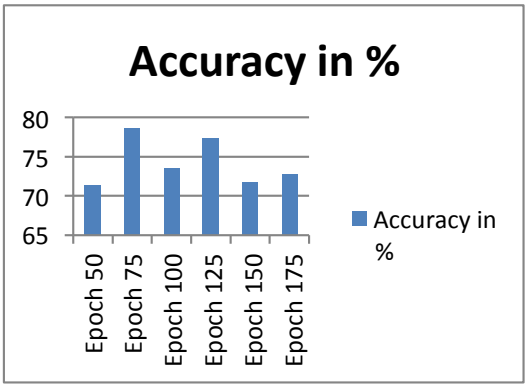
4.3.1 Pengaruh Learning Rate terhadap akurasi (Skema 1)



Grafik 1 Hasil Skema 1

Dari *chart* di atas, terlihat bahwa pengujian menggunakan nilai *Learning Rate* 0,07 memiliki hasil yang paling optimal. Maka dari itu untuk pengujian tahap selanjutnya digunakan *Learning Rate* 0,07.

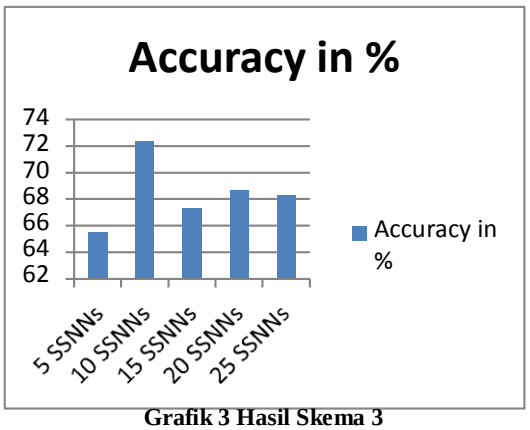
4.3.2 Pengaruh nilai Max Epoch terhadap akurasi



Grafik 2 Hasil Skema 2

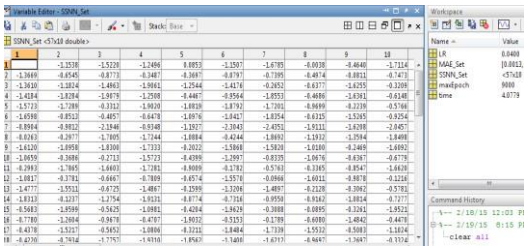
Dari hasil diatas terlihat bahwa pengujian menggunakan Epoch (perulangan *learning*) 75 memiliki hasil paling optimal dari yang lain.

4.3.3 Pengaruh Jumlah *SSNNs* terhadap akurasi



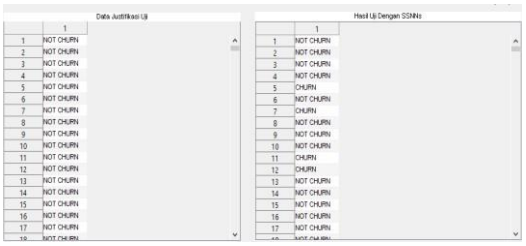
Grafik 3 Hasil Skema 3

4.4 Analisa Hasil Rule SSNNs (Skema 4 dan 5)



Gambar 10 Pembobotan SSNNs Pada Sistem

Berikut ini merupakan jumlah *SSNN* yang terbentuk dengan menggunakan parameter yang telah terpilih dari tes sebelumnya.



Gambar 11 Hasil Klasifikasi Prediksi

Terlihat bahwa hasil pengujian menunjukan adanya perubahan kelas dari pelanggan *not-churn* menjadi *churn*. Yang menandakan bahwa pelanggan tersebutlah yang berpotensi untuk melakukan *churn* berdasarkan hasil dari pengujian sistem ini. Hasil penerapan dari hasil parameter optimum *Learning Rate* 0,07, *Epoch* 75 adalah 72,35%.

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil yang diperoleh dari pengujian yang dilakukan, didapat bahwa algoritma *SSNNs* dapat diterapkan pada studi kasus *churn*. Dengan melakukan pengujian-pengujian parameter sesuai skema dimana pada skema 1 didapat hasil *Learning Rate* optimal 0,07 yang selanjutnya pada skema 2 digunakan untuk mencari *epoch* optimal yang didapat pada 75 dan diakhiri pada skema 3 dengan parameter jumlah *SSNN* 10.

Setelah didapat parameter-parameter tersebut dapat dilanjutkan keskema akhir yaitu skema 4 dan 5 dimana pada skema 4 didapat bahwa terdapat pergantian kelas pada data pelanggan dari *not churn* dan sebaliknya. Hal inilah yang digunakan pada skema 5 sebagai atribut pengukur akurasi dengan *F-measure.*Akurasi yang didapat adalah 72,35%.

5.2 Saran

Berikut ini merupakan saran yang dapat diberikan penulis setelah membuat penulisan ini :

a) Menerapkan metode Undersampling atau Oversampling agar hasil lebih akurat.

b) Menggunakan metode lain sebagai pembanding dalam klasifikasi.

DAFTAR PUSTAKA

[1] Anonym., Tanpa tahun. "*Evaluating a classification model – What does precision and recall tell me?*". *From* http://www.compumine.com/web/public/newsletter/20071/precision-recall, 2014.

[2] B. Huang, M. T. Kechadi dan B. Buckley., 2012, "*Customer churn prediction in telecommunications*," Expert Systems with Applications, vol. 39, p. 1414–1425.

[3] Clemente. M, Giner-Bosch. V, dan San Matias. S., 2012, *"Assessing Classification Methods for Churn Prediction by Composite Indicators,"*, Spain : Universitat Politecnica de Valencia.

[4] Dunham, Margaret H., 2003, "*Data Mining Introductory and Advanced Topics*," New Jersey: Prent ice Hall

[5] Huda, Nuqson ST., 2010, "Aplikasi Data Mining Untuk Menampilkan Informasi Kelulusan Mahasiswa", Indonesia : Universitas Diponegoro.

[6] J. Burez dan D. V. d. Poel., 2009, "*Handling Class Imbalance In Customer Chrun Prediction*," Expert Systems with Applications, vol. 36, pp. 4626-4636.

[7] J. Han, M. Kamber dan J. Pei.,2012. "*Data Mining Concepts and Techniques"*, Morgan Kaufmann Publishers.

[8] K. Clement, H. Li, C. Wilson, K. Hillary., 2013, "*Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining*," IJCSI, Vol. 10, Issue 2, No 1, March.

[9] Larose dan T. Daniel., 2005, "*Discovering Knowledge in Data: An Introduction to Data Mining*," John Willey & Sons. Inc.

[10] Moro, Sergio dan Laureano Raul M. S., 2012. "*Using Data Mining for Bank Direct Marketing : An Application of The Crisp-DM Methodology*," Lisboa : Instituto Universitario de Lisboa

[11] R. Agrawal dan R. Srikant., 1994, "*Fast algorithms for mining association rules*," Intl. Conf. on Very Large Databases, p. 487–499.

[12] R. Agrawal, T. Imielinski dan A. Swami., 1993, "*Mining associations between sets of items in massive databases*," Proc. ACM SIGMOD Intl. Conf. On Management of Data, p. 207–216.

[13] Rahman, Monzurur S. M., 2006, *Thesis : Data Mining Using Neural Network*, Australy : RMIT University.

[14] Suyanto. ST, MSC., 2007, *"Artificial Intellegence"*. Bandung: INFORMATIKA.