Journal of Intelligent Systems and Business Intelligence

Volume 1, Nomor 2, September 2012
ISSN 2302-268X

Diterbitkan oleh:
The First Conference and Workshop on Intelligent Systems and Business Intelligence (COWISBI)
Program Pascasarjana Magister TeknikInformatika UDINUS Semarang

Journal of Intelligent System and Business Intelligence	Volume 1 Nomor 2	Halaman 1 - 284	Semarang September 2012	ISSN 2302-268X

DAFTAR ISI

Segmentasi Citra Warna dan Tekstur Menggunakan Fuzzy C-Means dan Filter Gabor
Yanuar Wicaksono 1

Prediksi Nilai Tukar Rupiah terhadap Dollar Amerika Menggunkan Artificial Neural Network
Fransisika Suryaning P 6

Perbandingan Kualitas Segmentasi Citra Tanda Tangan Off-Line dengan Parameter Psnr Pada Metode K-Means dan Fcm
Fransisika Suryaning P, Guntur Darmawan 9

Pengelompokan Suara Format Wav Berbasis K-Means
Guntur Darmawan 13

Pengaruh Fungsi Kalibrasi Menggunakan Dua Kamera (Hp dan Pocket) dan dalam Kondisi Yang Berbeda dalam Computer Vision
Andhen Priyono, Indaryanto, Puri sulisyawati 16

Penerapan Neural Network untuk Prediksi Churn Pada Perusahaan Telekomunikasi
Setyoningsih Wibowo, Endah Ekasanti Saputri, Retno Wahyusari 20

Image Enhancement Pada Deteksi Teks dengan Menggunkan Median Filter
Endah Ekasanti Saputri, Setyoningsih Wibowo, Retno Wahyusari 26

Ekstraksi Pola Venasi Menggunakan Trainable Segmentation
Lukmono Yogi Prayogi, Ahmad Akrom 31

Teknik Proteksi terhadap Serangan Sql Injection Menggunakan Konsep Amnesia pada Aplikasi Web
Syah Askar Akbar, Ajib Susanto 34

Analisa Pengaruh Perbedaan Kadar Garam dalam Air terhadap Kalibrasi Kamera dengan Menggunkan Metode Zhang
Eko Indarto 45

Fuzzy Subtractive Clustering untuk Pemetaan Industri di Kabupaten Brebes
Ar[illegible]ah Suryo ajie, Vincent Suhartono, M Arief Soeleman 48

Algoritma Naïve Bayaes Berbasis Adaboost untuk Evaluasi Kelayakan Pemberian Kredit Perumahan Rakyat
Hoerman Sitohang, Abdul Syukur, Romi Satrio Wahono 52

Discrete Wavelet Transform untuk Mempercepat Prediksi Produksi Susu Sapi Perah Berbasis Neural Network
Warto, Arief Soeleman, Abdul Syukur 57

Aplikasi Adaptive Tresholding untuk Deteksi Tepi Citra
Retno Wahyusari, Endah Ekasanti Saputri, Setyoningsih Wibowo 65

Fuzzy Inference System (Fis) untuk Penentuan Pemenang Tender
Veronica Agnes Abels, Edi Noersasongko, Romi Satrio Wahono 69

Deteksi Tanda Tangan untuk Mencegah Pemalsuan Dokumen Menggunakan Algoritma Sift
Dwi Puji Prabowo, Dimas Irawan 75

Simulasi Antrian Kendaraan Pada Persimpangan Jalan Berbasis Multi Agent Menggunakan Logika Fuzzy
Suhanda, Moch. Hariadi, M. Arief Soeleman 79

Ekstraksi Fitur Motif Pada Wayang Kulit untuk Pengenalan Karakter dan Pakem Berbasis PCA
Aris Tri Jaka 84

Content Based Image Retieval Berbasis Fourier Transform untuk Pencarian Image Ikan Berdasarkan Fitur Bentuk
Artoni 88

Sistem Pengambil Keputusan Menggunakan Data Mining Sebagai Acuan Penentuan Strategi Promosi
Nidia Rosmawanti 93

Model Neural Network Berbasis Adaboost untuk Prediksi Bisnis Forex
Abdul Syukur, Catur Supriyanto, R. Hadapiningradja Kusumodestoni 100

Pengelompokan Dokumen Ilmiah Menggunakan Metode K-Means Berdasarkan Judul, Kata Kunci dan Abstrak
Sumaryono, Abdul Syukur, Hendro Subagyo 105

Penerapan Fuzzy Multi Criteria Decision Making untuk Pemilihan Rumah
Taufiq, Abdul Syukur, Romi Satrio Wahono 110

Optimasi Algoritma Stemmer dengan Teknik Enhanced Suffix Stripping untuk Bahasa Indonesia
Mohamad Yakop 115

Algoritma Naïve Bayes Berbasis Particle Swarm Optimization untuk Diagnosis Penyakit Hepatitis
Samsiatun Aisiah 120

Aplikasi Augmented Reality Sebagai Alat Pengukur Baju Wisudawan Wisudawati Di Universitas Dian Nuswantoro
Adi Purnomo, Hanny Haryanto 125

Decision Support System (Dss) Recruitment Frontliner dengan Metode Analytic Hierarchy Process (Ahp)
Dhimas Rahadhian, Stefanus S., Arif Soeleman 129

Deteksi Tingkat Keparahan Tuna Rungu – Wicara dengan Pendekatan Fuzzy Set untuk Membantu Slb Mengidentifikasi Siswa
Zohrahayaty, Vincent Suhartono, Stefanus Santosa 134

Deteksi Citra Tepi Batik
Moh Abu Choir 139

Deteksi Wajah Berbasis Segmentasi Warna Kulit Menggunakan Ruang Warna Ycbcr & Template Matching

Rony Wijanarko, Sri Sugianto, Ruri Suko Basuki 142

Diagnosis Penyakit Diabetes Menggunakan Algoritma K-Nearest Neighbor Berbasis Particle Swarm Optimization
Laily Hermawanti, Setyoningsih Wibowo, Vincent Suhartono 148

Ekstrasi Fitur Pada Iris Mata Menggunakan PCA dengan Penentuan Jarak Euclidean
Vita Kusvitaria, Ricardus Anggi Pramunendar 151

Menguatkan Daya Saing Pasar dengan Pengembangan Rancangan Tata Letak dengan Pendekatan Clustering Analysis dan Market Basket Analysis
Harwati, Yasser Azka, Reny Lituhayu Lagaida, Nur Riana 155

Model Prediksi Rentet Waktu Neural Network Berbasis Particle Swarm Optimization untuk Prediksi Harga Saham
Andri Pramuntadi 159

Optimasi Algoritma *Support Vector Machine* (Svm) Menggunakan Adaboost untuk Penilaian Resiko Kredit
Defri Kurniawan 164

Deteksi Tepi Robert Pada Gambar Daun Tembakau dengan Optimasi Ant Colony
Ali Muqoddas, Andhika Wira Utama 170

Penerapan Neural Network untuk Prediksi Data Time-Series Arus Lalu Lintas Jangka Pendek
Yuli Fitrianto, Abdul Syukur, Romi Satria Wahono 173

Pengenalan Motif Citra Batik Menggunakan Metode Pca (Principal Component Analysis) Berbasis Seleksi Eigenvector
RIFKI, Kusnandar Afrisia 177

Perbandingan Algoritma K-Means Dan Fuzzy C-Means Pada Segmentasi Citra
Karno, Nugroho, M. Fikri Hidayatullah 182

Prediksi Data Arus Lalu Lintas Jangka Pendek Menggunakan Optimasi Neural Network Berbasis Genetik Algorithm
Sucianna Ghadati Rabiha, Catur Supriyanto, Stefanus Santosa 186

Re-Design Pasar Tradisional Dengan Pendekatan Clustering Analysis Dan Market Basket Analysis Sebagai Penguatan Nilai Daya Saing Berbasis Ekonomi Kerakyatan
Harwati, Yasser Azka, Reny Lituhayu Lagaida, Nur Riana 192

Segmentasi Kulit Ras Manusia Pada Citra Digital Menggunakan Ruang Warna Hsv
Muhammad Fikri Hidayattullah, Nugroho, Kar no 197

Segmentasi Dan Eksstraksi Huruf Jawa Dengan Mennngunakan Metode Canny Edge Detection
Mustagfirin, Lukman Hakim 200

Analisa Penerimaan Sistem Informasi Kesehatan Studi Dinas Kesehatan Kabupaten Jepara
Buang Budi Wahono 204

Kombinasi Transformasi Dct-Dwt Berbasis *Pn Sequence* Untuk Meningkatkan Ketahanan *Image Watermarking* Dari *Image Compression*
Ferda Ernawan, Abdul Syukur, Ignatius Moses Setiadi D R 207

Analisa Tingkat Akurasi Metode Rabin Karp Dan Lsa Dalam Mendeteksi Kesamaan Dokumen Teks Bahasa Inggris
Sindhu Rakasiwi, Abdul Syukur, Catur Supriyanto 216

Perancangan Arsitektur Enterprise Dengan Menggunakan Togaf Di Pemerintah Kabupaten Blora
Setyo Wahyu Saputro 222

Sistem Informasi Manajemen Rantai Pasok Pada Sistem Informasi Manajemen Korporasi
Tony Hariono, Suprapedi, M. Sidiq 233

Peningkatan Akurasi Klasifikasi *Grade* Daun Tembakau Sebagai Bahan Baku Rokok Menggunakan *Backpropagation Neural Network*
Meila Izzana, Ricardus Anggi Pramunendar, Stefanus Santosa, Fajrian Nur Adnan 238

Prediksi Data Arus Lalu Lintas Jangka Pendek Menggunakan Optimasi *Neural Network Berbasis Genetik Algorithm*
Sucianna Ghadati Rabiha, Catur Supriyanto, Stefanus Santosa 244

Algoritma Klasifikasi Data Mining Naive Bayes Berbasis Paticle Swarm Optimization Untuk Deteksi Penyakit Jantung
NurAeni Widiastuti, Ade Yusupa, Catur Supriyanto, Stefanus Santosa 249

Peningkatan Segmentasi Padaidentifikasi Plat Nomor Kendaraan Indonesia Menggunakan Metode *Otsu* Dengan *Gaussian*
Erwin Dwika Putra, Stefanus Santosa, Ricardus Anggi Pramunendar 252

The Role Of University In New & Renewable Energy
Rohani Jahja Widodo 259

Membaca Hasil Jawaban Lembar Jawab Multiple Choice Manual Dengan Algoritma PCA Dan DTW
Masruhan Mufid 264

Closed Form Matting Pada Natural Image Menggunakan Image Enhancement Dan Alpha Smoothing Berbasis Lo Smoothing
Alexander Agung Nugroho, Dhar Heri Arimaya 268

Analisa Tingkat Akurasi Metode Rabin Karp Dan Lsa Dalam Mendeteksi Kesamaan Dokumen Teks Bahasa Inggris
sindu rakasiwi 267

Signature Recognition Menggunakan Metode Principal Component Analysis (Pca) Berbasis Euclidean Distance
Eko Adi Saputro, Puri Sulistiyawati 277

Implementasi Protokol Radius Untuk Ieee 802.11 Wireless
R. Rhoody Setiawan, Anteng Widodo 280

Journal of Intelligent Systems and Business Intelligence
Volume 1, Nomor 2, September 2012 | ISSN 2302-268X

PENERAPAN *NEURAL NETWORK* UNTUK PREDIKSI *CHURN* PADA PERUSAHAAN TELEKOMUNIKASI

Setyoningsih Wibowo
Magister Teknik Informatika
Universitas Dian Nuswantoro
Semarang, Indonesia
budinink_1623@yahoo.com

Endah Ekasanti Saputri
Magister Teknik Informatika
Universitas Dian Nuswantoro
Semarang, Indonesia
endah.ekasanti@gmail.com

Retno Wahyusari
Magister Teknik Informatika
Universitas Dian Nuswantoro
Semarang, Indonesia
wahyusari_r@yahoo.co.id

Abstrak - Perkembangan industry telekomunikasi saat ini sangat berkembang pesat. Perkembangan ini menjadikan semakin banyak jumlah industri telekomunikasi baik penyedia layanan GSM (*Global System Mobile*) maupun CDMA (*Code Devision Multiple Access*). Tiap perusahaan berlomba-lomba untuk menarik pelanggannya dengan berbagai macam cara misalnya dengan paket layanan hemat, tersedianya konten-konten menarik sampai dengan persaingan harga. Dalam sebuah perusahaan telekomunikasi terutama perusahaan telekomunikasi seluler, bahwa asset perusahaan yang paling bernilai adalah *existing customer* atau pelanggan setia yang mereka miliki. Dari permasalahan diatas maka muncullah suatu fenomena terjadinya *churn*. *Churn* adalah berpindahnya pelanggan dari satu provider ke provider yang lain. Terjadinya *churn* ini harus diwaspadai oleh perusahaan-perusahaan telekomunikasi karena hal inilah yang menyebabkan menurunnya penghasilan. Saat ini banyak perusahaan-perusahaan besar sudah memikirkan strategi bagaimana mengurangi *churn* pelanggannya, salah satunya dengan pendekatan data mining yaitu untuk memprediksi pelanggan yang akan *churn*. Dalam penelitian ini, mengusulkan penerapan *neural network* yang bertujuan untuk mengetahui tingkat akurasi dari prediksi ini.

Kata kunci : *churn prediction, neural network*

I. PENDAHULUAN

Seiring dengan berjalannya waktu, perkembangan industry telekomunikasi saat ini sangat berkembang pesat. Perkembangan ini menjadikan semakin banyak jumlah industri telekomunikasi baik penyedia layanan GSM (*Global System Mobile*) maupun CDMA (*Code Devision Multiple Access*). Tiap perusahaan berlomba-lomba untuk menarik pelanggannya dengan berbagai macam cara misalnya dengan paket layanan hemat, tersedianya konten-konten menarik sampai dengan persaingan harga. Dalam sebuah perusahaan telekomunikasi terutama perusahaan telekomunikasi seluler, bahwa asset perusahaan yang paling bernilai adalah *existing customer* atau pelanggan setia yang mereka miliki. [1]. Dari permasalahan diatas maka muncullah suatu fenomena terjadinya *churn*. *Churn* adalah berpindahnya pelanggan dari satu provider ke provider yang lain [2]. Terjadinya *churn* ini harus diwaspadai oleh perusahaan-perusahaan telekomunikasi karena hal inilah yang menyebabkan menurunnya penghasilan. Saat ini banyak perusahaan-perusahaan besar sudah memikirkan strategi bagaimana mengurangi *churn* pelanggannya, salah satunya dengan pendekatan data mining yaitu untuk memprediksi pelanggan yang akan *churn*.

Penelitian-penelitian sebelumnya tentang churn prediction pada perusahaan telekomunikasi: Kim dan Jung mengusulkan pengenalan kombinasi klasifikasi *uniformly subsampled ensemble model of SVM (USE SVM)* dengan *PCA* tidak hanya mereduksi dimensi yang besar pada data tetapi juga meningkatkan keandalan dan keakuratan model kalibrasi pada dataset dengan kelas distro yang miring. Hasil dari penelitian ini menyatakan bahwa kinerja dari model USESVM dengan PCA adalah unggul untuk semua model dan jumlah komponen utama mempengaruhi akurasi model ensemble. [3]. Wang dan Xiao menggabungkan pembelajaran transfer dan beberapa klasifikasi ensemble dan mengusulkan sebuah model perpindahan ensambel untuk data tidak seimbang (TEMID, *Transfer Ensemble Model for Imbalance Data*), yang menghasilkan bahwa metode ini benar-benar dapat meningkatkan kinerja prediksi *churn*[4]. Zhang dkk memprediksi *churn* berbasis perilaku pelanggan dengan menggunakan pendekatan *neural network*, criteria yang berbeda dapat didefinisikan untuk pemilihan pelanggan sesuai dengan tujuan penyedia layanan telekomunikasi, hasil yang diperoleh menyatakan bahwa tidak semua variable dalam data dapat dipergunakan dalam penelitian ini [5]. Ghorbani dkk mengusulkan bahwa aplikasi dari algoritma *Locally Linear Model Tree* (LOLIMOT) sangat efektif untuk solusi masalah klasifikasi biner [6]. Tsa dan Lu membahas *churn* prediction dengan dua model hybrid [7]. Adem dan Dilek membahas GSM *churn* dengan menggunakan

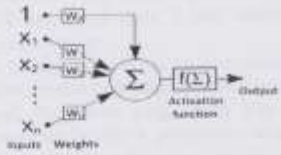
Fuzzy c-means dan *system neuro inferensi adaptif Fuzzy* [8]. Pan mencoba untuk memprediksi *churn* pelanggan dengan *C5.0*, *Logistic Regresion*, dan algoritma *neural network* [7]. Sebagian besar penelitian menekankan pada algoritma saat memprediksi *churn*. Ada sedikit penelitian berfokus pada atribut seleksi untuk *churn* prediksi perusahaan telekomunikasi. Jumlah tersebut dari atribut yang digunakan untuk memprediksi *churn* bervariasi dari tiga sampai lebih dari dua puluh. Sebagai contoh, dua puluh satu variable adalah digunakan sebagai input [8].

Paper ini menjelaskan penerapan *neural network* yang bertujuan untuk mengetahui tingkat akurasi dari prediksi ini. Pada bab II akan menjelaskan metode yang diusulkan. Eksperimen dan hasil akan dibahas pada bab III dan kesimpulan pada bab IV.

II. METODE YANG DIUSULKAN

2.1. Neural Network

Neural Network adalah suatu arsitektur jaringan untuk memodelkan cara kerja system syaraf manusia (otak) dalam melaksanakan tugas tertentu. Pemodelan ini didasarkan oleh kemampuan otak manusia dalam mengorganisasi sel-sel penyusunnya (*neuron*), sehingga memiliki kemampuan untuk melaksanakan tugas-tugas tertentu khususnya pengenalan pola dengan efektifitas jaringan sangat tinggi [9].



Gambar 1. McCulloch dan Pitts

Gambar diatas memperlihatkan bahwa sebuah neuron memiliki tiga komponen:

1. *Synapse* (w1, w2,..., wn) T
2. Alat penambah (*adder*)
3. Fungsi aktivasi (f)

Dewasa ini *neural network* telah diaplikasikan di berbagai bidang. Hal ini dikarenakan *neural network* memiliki kelebihan-kelebihan sebagai berikut:

- Dapat memecahkan problem *non-linear* yang umum dijumpai di aplikasi
- Kemampuan memberikan jawaban terhadap pola yang belum pernah dipelajari (*generalization*)
- Dapat secara otomatis mempelajari data *numeric* yang diajarkan pada jaringan tersebut.

Backpropagation

Algoritma *backpropagation* adalah metode pelatihan terawasi, dan biasanya digunakan oleh perceptron dengan banyak lapisan untuk mengubah bobot-bobot yang terhubung dengan *neuron-neuron* yang ada pada lapisan tersembunyinya. Algoritma backpropagation menggunakan *error* output untuk mengubah nilai bobot-bobotnya dalam arah mundur (*backward*). Untuk mendapatkan *error* ini, tahapan perambatan maju (*forward propagation*) harus dikerjakan terlebih dahulu. Pada saat perambatan maju, *neuron-neuron* diaktifkan dengan menggunakan fungsi aktivasi *sigmoid*, yaitu:

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

Algoritma *backpropagation*:

A. Untuk tiap-tiap pasangan elemen yang akan dilakukan pembelajaran, kerjakan:

Feedforward:

a. Untuk tiap-tiap unit input (xi, i = 1,2,3,...,n) menerima sinyal xi dan meneruskan sinyal tersebut ke semua unit pada lapisan yang ada didepannya (lapisan tersembunyi).
b. Tiap-tiap unit tersembunyi (zj, j = 1,2,3,...,p) menjumlahkan sinyal-sinyal input terbobot:

$$z_in_j = v_{0j} + \sum_{i=1}^{n} x_i v_{ij}$$

gunakan fungsi aktivasi untuk menghitung sinyal outputnya:

$$z_j = f(z_in_j)$$

dan kirimkan sinyal tersebut ke semua unit di lapisan depannya (unit-unit output).

c. Tiap-tiap unit output (Yk, k = 1,2,3,...,m) menjumlahkan sinyal-sinyal input terbobot.

$$y_in_k = w_{0k} + \sum_{j=1}^{p} z_j w_{jk}$$

gunakan fungsi aktivasi untuk menghitung sinyal outputnya:

$$y_k = f(y_in_k)$$

dan kirimkan sinyal tersebut ke semua unit di lapisan depannya (unit-unit output)

Backpropagation

d. Tiap-tiap unit output (Yk, k = 1,2,3,...,m) menerima target pola yang berhubungan dengan pola input pembelajaran, hitung informasi errornya:

$$\delta_k = (t_k - y_k) f'(y_in_k)$$

kemudian hitung koreksi bobot (yang nantinya akan digunakan untuk memperbaiki nilai wjk):

$\Delta w_{jk} = \alpha \delta_k z_j$

hitung juga koreksi bias (yang nantinya akan digunakan untuk memperbaiki nilai w0k):

$\Delta w_{0k} = \alpha \delta_k$

kirimkan δ_k ini ke unit-unit yang ada di lapisan depannya.

e. Tiap-tiap input tersembunyi (zj, j = 1,2,3,...,p) menjumlahkan delta inputnya (dari unit-unit yang berada pada lapisan depannya):

$$\delta_in_j = \sum_{k=1}^{m} \delta_k w_{jk}$$

kalikan nilai ini dengan turunan dari fungsi aktivasinya untuk menghitung informasi *error*:

$\delta_j = \delta_in_j f'(z_in_j)$

kemudian hitung koreksi bobot (yang nantinya akan digunakan untuk memperbaiki nilai vij):

$\Delta v_{ij} = \alpha \delta_j x_i$

hitung juga koreksi bias (yang nantinya akan digunakan untuk memperbaiki bilai v0j):

$\Delta v_{0j} = \alpha \delta_j$

f. Tiap-tiap unit output (Yk, k= 1,2,3,...,m) memperbaiki bias dan bobotnya (j = 0,1,2,...,p):

$w_{jk}(baru) = w_{jk}(lama) + \Delta w_{jk}$

Tiap-tiap unit tersembunyi (zj, j = 1,2,3,...,p) memperbaiki bias dan bobotnya (I = 0,1,2,...n):

$v_{ij}(baru) = v_{ij}(lama) + \Delta v_{ij}$

B. Tes kondisi berhenti

Untuk kasus data yang peneliti gunakan, jaringan memiliki 12 unit *input layer*, 8 unit *hidden layer* dan 2 unit *output layer*.

Salah satu bidang dimana *backpropagation* dapat diaplikasikan dengan baik adalah bidang peramalan (*forecasting*). Peramalan yang sering kita dengar adalah peramalan besarnya penjualan, nilai tukar valuta asing, prediksi besarnya aliran sungai, prediksi *churn*, dan lain-lain.

Record data dipakai sebagai data pelatihan untuk pencarian bobot yang optimal. Sehingga perlu menetapkan besarnya periode dimana data berfluktuasi dan periode kita tentukan secara intuitif.

Perubahan bobot didasarkan atas gradien yang terjadi untuk pola yang dimasukkan saat ini. Modifikasi yang dapat dilakukan adalah melakukan perubahan bobot yang didasarkan atas arah gradien pola terakhir dan pola sebelumnya (yang disebut dengan momentum) yang dimasukkan. Jika tidak hanya pola masukan terakhir saja yang diperhitungkan.

Penambahan momentum dimaksudkan untuk menghindari perubahan bobot yang mencolok akibat adanya data yang sangat berbeda dengan yang lain. Apabila beberapa data terakhir yang diberikan ke jaringan memiliki pola serupa (berarti arah gradien sudah benar), maka perubahan bobot dilakukan secara cepat. Namun, apabila data terakhir yang dimasukkan memiliki pola yang berbeda dengan pola sebelumnya, maka perubahan dilakukan secara lambat. Penggunaan momentum ini sangat membantu saat pembobotan khususnya untuk data yang bervariasi.

2.2. *Churn Prediction*

Churn prediction adalah salah satu *task data mining* yang bertujuan untuk memprediksi pelanggan yang berpotensi untuk *churn*. Pelanggan yang *churn* dapat dibagi menjadi dua kelompok utama [10], yaitu:

a. *Voluntary churners*
   Adalah *churn* secara sukarela. Pelanggan ini lebih sukar ditentukan sebab pada pelanggan jenis ini terjadi ketika seorang pelanggan membuat keputusan secara sadar untuk mengakhiri layanan yang digunakan.

b. *Involuntary churners*
   Adalah *churn* tidak sukarela. Pelanggan ini lebih mudah untuk diidentifikasi, seperti pelanggan yang menggunakan jasa ditarik/dicabut dengan sengaja oleh perusahaan tersebut dikarenakan adanya beberapa alasan.

2.3. Dataset

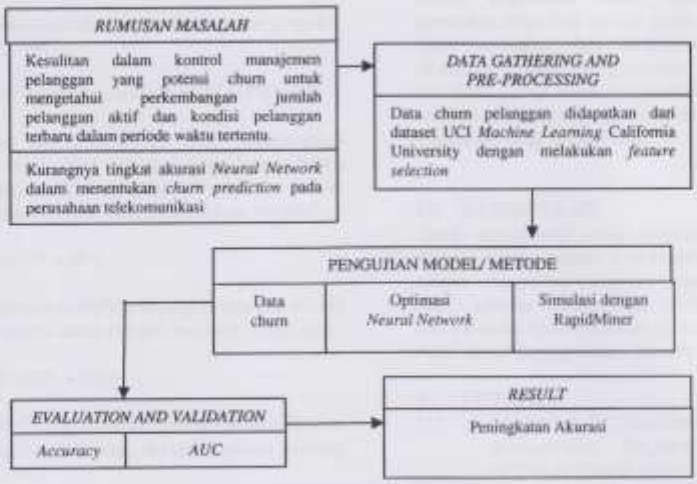
Churn dataset diambil dari database-UCI California University. Dalam dataset ini mendefinisikan *churn* per satu pelanggan seluler yang memberikan semua layanan dari satu perusahaan telekomunikasi, dalam jendela waktu tiga bulan terus menerus. Ada 19 karakteristik (8 dari variable kontinyu, 10 adalah variable diskrit dan 1 adalah kelas label). Terdiri dari 3333 data training dan 1667 data testing. Untuk data training 3333 sampel (2850 dari yang untuk *churn* dan 483 untuk *non-churn*) dan data testing 1667 sampel (1452 dari yang untuk *churn* dan 215 unutuk *non-churn*) dalam dataset.

Dalam penelitian ini, data yang telah diinputkan ke dalam *excel* kemudian perlu dibagi lagi menjadi 2 bagian. Yaitu, 1 bagian untuk *training* dan 1 bagian untuk *testing*. ukuran proporsi dari data mining untuk training dan testing set berada dalam range 50% training dan testing 50% sampai dengan 90% training dan 10% testing [11].

2.4. Gambaran Sistem

Dalam membuat system tersebut digunakan metode *neural network* dengan arsitektur *backpropagation*. Data yang ada tidak dapat langsung dipergunakan sebagai data input pada perangkat lunak *churn prediction* dengan *neural network* melainkan harus diproses terlebih dahulu. Alur proses yang terjadi adalah sebagai berikut:

tool yang baik untuk *data mining*, *data mining*



Gambar 2 Gambaran Umum

Reduksi data

Teknik reduksi data (*data reduction*) dapat diterapkan untuk mendapatkan representasi pengurangan data set yang lebih kecil dari volume, namun mempertahankan integritas data asli. Artinya, *mining* dikurangi *data set* seharusnya lebih efisien sebelum menghasilkan hasil analisis yang sama (atau hampir sama) [12].

Strategi untuk reduksi data adalah sebagai berikut [12]:

a. Agregasi kubus data (*data cube*), di mana operasi-operasi agregasi diterapkan pada data dalam konstruksi kubus data (*data cube*).
b. Atribut seleksi subset (*attribute subset selection*), di mana atribut-atribut tidak relevan, lemah atau redudansi atau dimensi dapat dideteksi dan dihapus.
c. Pengurangan dimensi, dimana mekanisme pengkodean (*encoding*) yang digunakan untuk mengurangi ukuran *data set*.
d. *Numerosity reduction*, di mana data diganti atau diperkirakan berganti-ganti, yang lebih kecil dari representasi data seperti model-model parametrik (perlu menyimpan hanya parameter-parameter model bukan data yang sebenarnya) atau metode-metode nonparametrik seperti *clustering*, *sampling*, dan penggunaan histogram.
e. Diskritisasi dan generasi hirarki konsep, di mana nilai-nilai data baris untuk atribut diganti dengan jarak atau tingkat konseptual yang lebih tinggi. Diskritisasi data adalah bentuk *numerosity reduction* yang sangat berguna bagi generasi otomatis dari hirarki konsep. Diskritisasi dan generasi hirarki konsep adalah pada abstraksi tingkat perkalian.

III. ANALISIS DAN PENGUJIAN

Diperlukan cara yang sistematis untuk mengevaluasi kinerja suatu metoda. Evaluasi klasifikasi didasarkan pada pengujian pada obyek benar dan salah [13]. Dalam penelitian ini menggunakan metode *confusion matrix* dan *ROC curve* untuk mengukur hasil proses klasifikasi.

3.1. Confusion Matrix

Confusion matrix merupakan metode yang digunakan untuk mengetahui performansi algoritma [13]. 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 3 *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. Untuk menghitung TP digunakan rumus:

True Positif Rate (Sensitivity) = TP / (TP + FN)

TN dikatakan senstifitas karena mengukur proporsi benar "negative".

True Negatif Rate (Specificity) = TN / (TN + FP)

Sensitivitas dan spesifisitas tidak memberikan informasi untuk nilai diagnose yang benar. Maka perlu adanya PPV (*Positif Predictive Value*) dimana proporsi kasus dengan hasil test "positif".

PPV = TP / (TP + FP)

Dan membutuhkan NPV (*Negatif Predictive Value*) dengan proporsi kasus dengan hasil tes "negative".

NPV = TN / (TN + FN)

Tingkat FP (*False Positive*) adalah jumlah normal kasus yang kesalahannya diklasifikasikan sebagai kelas abnormal.

False Positive Rate = FP / (FP + TN)

Tingkat FN (*False Negative*) adalah jumlah kasus normal yang kesalahannya diklasifikasikan sebagai kelas normal.

False Negative Rate = FN / (FN + FP)

Akurasi paling banyak digunakan dalam klasifikasi yang menggunakan kinerja matriks. Dan untuk menghitung akurasi pada matriks:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

3.2. *ROC Curve*

ROC digunakan dalam pembelajaran mesin dan penelitian data mining untuk menilai hasil prediksi [13]. 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

Pada pengujian ini, data *churn* dibuat dalam dua dataset yaitu data *training* dan data *testing*. Data training digunakan oleh satu atau lebih skema belajar untuk pengklasifikasian. Kemudian data testing digunakan untuk menghitung tingkat kesalahan akhir dari metode yang digunakan.

Hasil evaluasi dari eksperimen yang telah dilakukan diukur menggunakan *confusion matrix* dan *ROC Curve* untuk mengetahui akurasinya. Dan validasi data digunakan untuk menentukan teknik klasifikasi yang paling baik.

IV. KESIMPULAN

Hasil eksperimen yang dilakukan pada *churn* dataset menggunakan penerapan *neural network* adalah didapatkan *accuracy* sebesar 71.99% dan AUC sebesar 0.680. Dari hasil ini disarankan, bahwa untuk meningkatkan *accuracy* maupun nilai AUC dapat menggunakan metode yang lainnya.

V. PUSTAKA

[1] Mattersion, R. "Telecom churn Management". Fuquary – Varina, NC : APDG Publising. (2001).

[2] John Hadden, Ashutosh Tiwari, Rajkumar Roy, and Dymitr Ruta, "Churn Prediction: Does Technology Matter?", ISSN 2006

[3] Kim., Lee., Jung., Kim., "A New Ensemble Model for Efficient Churn Prediction in Mobile Telecommunication". 45th Hawaii International Conference on System Sciences, (2012)., 1023 – 1029.

[4] Wang., Xiao., "Transfer Ensemble Model for Customer Churn Prediction with Imbalanced Class Distribution"., International Conference of Information Technology, Computer Engineering and Management Sciences, (2011)., 177 – 181.

[5] Zhang., Zheng., Berry., "Behavior-Based Telecommunication Churn Prediction with Neural Network Approach"., International Symposium on Computer Science and Society, (2011),307 – 310.

[6] Ghorbani., Taghiyareh., Lucas., "The Application of Locally Linear Model Tree on Customer Churn Prediction"., International Conference of Soft Computing and Pattern Recognition., (2009)., 472 – 477.

[7] Parag C. Pendharkar. "Genetic Algorithm Based Neural Network Approaches for Predicting Churn in Celluler Wireless Network Service". Expert System with Applications 36 (2009) 6714 – 6720.

[8] Adem Karahoca, Dilek Karahoka, "GSM Churn Management by Using Fuzzy c-Means Clustering and Adaptive Neuro Fuzzy Inference System"., Expert System with Applications 38 (2011) 1814 – 1822.

[9] Suyanto, "Soft Computing Membangun Mesin Ber-IQ Tinggi", Informatika, Bandung, (2008), 58.

[10] Cardell, Scott., Golovnya, Miklail., Steinberg, Dan., "Churn Modeling for Mobile Telecomunications". Salford System. California. (2003)

[11] Ian H Witten, Frank Eibe, and Mark A Hall, "Data Mining Practical Machine Learning Tools and Techniques", 3rd ed. Burlington, USA, Morgan Kaufmann Publishers, (2011).

[12] Jiawei Han and Micheline Kamber, *Data Mining Concepts And Techniques*, Elsevier, 2007.

[13] Gorunescu, F. (2011). *Data Mining: Concepts, Model and Techniques* (Vol. 12). (P. J. Kacprzyk, & P. L. Jain, Eds.) Berlin, Jerman: Springer.

[14] Daniel T Larose, "Data Mining Methods and Models", Hoboken, Ney Jersey, United State of America, John Wiley & Son, Inc., 2006.

[15] Kusumadewi, Sri. "Artificial Intellegence" (Teknik dan Aplikasi. Edisi pertama). Graha Ilmu. Jogjakarta. (2003) 279 -331.