

PREDIKSI SISWA LULUS TIDAK TEPAT WAKTU MENGUNAKAN BACKPROPAGATION NEURAL NETWORK

Agustian Noor

Program Studi Teknik Informatika, Politeknik Negeri Tanah Laut
agustiannoor@gmail.com

Abstrak

Banyak terjadi kasus siswa terlambat lulus sampai bertahun-tahun, ini sebuah masalah bagi suatu civitas perguruan tinggi dalam memajukan mutu dari kampus, penelitian ini dilakukan untuk menciptakan suatu sistem prediksi yang akurat yang mana nantinya akan membantu civitas yang mempunyai masalah seperti diatas. Data mining dapat membantu dalam memprediksi suatu sistem, sehingga dapat dilakukan pada penelitian ini agar prediksi lebih tepat dan akurat. Ada berbagai macam teknik dalam data mining, untuk penelitian ini teknik yang dipakai ialah Neural Network Backpropagation, terdapat beberapa tahap dalam penelitian ini yaitu tahap pengumpulan data, pengolahan data, model atau metode yang diusulkan, eksperimen pada model tersebut, evaluasi dan validasi hasil. Disimpulkan bahwa teknik data mining menggunakan neural network backpropagation dapat menghasilkan suatu prediksi yang tepat dan akurat dari penelitian sebelumnya untuk menentukan siswa lulus tidak tepat waktu pada Sekolah Menengah Kejuruan Negeri 1 Kertak Hanyar.

Kata Kunci: Prediksi Siswa Lulus Tidak tepat waktu, data mining, neural network backpropagation, MSE.

1. PENDAHULUAN

Semakin ketatnya persaingan dalam masing-masing perguruan tinggi dalam mencetak siswa yang berkompeten dan berkualitas sehingga menuntut perguruan tinggi tersebut untuk memperbaiki sistem dan kualitasnya menjadi lebih baik lagi, dimana salah satu masalah dalam sistem ialah apabila ada beberapa siswa yang terlambat lulus atau tidak tepat pada waktunya sehingga menjadi kendala untuk kemajuan suatu sistem untuk pengangkatan mutu dan kualitas dari suatu perguruan tinggi tersebut.

Akurasi memprediksi siswa berguna di banyak konteks yang berbeda di perguruan tinggi. Sebagai contoh, mengidentifikasi siswa luar biasa untuk beasiswa merupakan bagian penting dari penerimaan proses pendidikan selanjutnya, dan siswa yang lemah gagal mengidentifikasi untuk mengalokasikan sumber daya les terbatas (Nghe, Janecek, & Haddawy, 2007). Untuk memprediksi siswa tidak tepat waktu sangat tepat kiranya apabila menggunakan Backpropagation Neural Network karena

algoritma ini sudah teruji kualitasnya pada penelitian sebelumnya seperti prediksi tinggi muka air sungai barito menggunakan backpropagation.

Penelitian ini bertujuan untuk memprediksi siswa lulus tidak tepat waktu menggunakan Algoritma Backpropagation Neural Network yang lebih akurat.

Manfaat dari penelitian ini adalah :

- Manfaat praktis dari hasil penelitian ini adalah diharapkan agar dapat digunakan oleh civitas-civitas perguruan tinggi untuk memprediksi kelulusan dari siswa.
- Manfaat teoritis dari hasil penelitian ini diharapkan dapat menjadi referensi untuk pembelajaran dan penerapan algoritma neural network untuk memprediksikan suatu masalah.
- Manfaat kebijakan diharapkan agar algoritma neural network mampu sebagai alat pertimbangan untuk meningkatkan kegiatan mengajar dan pendekatan pembelajaran pada suatu lembaga pengajaran.

2. TINJAUAN PUSTAKA

2.1 Data Mining

Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, mengenal pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar (Larose).

Klasifikasi dan prediksi adalah dua bentuk analisis data yang dapat digunakan untuk mengekstrak menggambarkan model kelas data penting atau untuk memprediksi tren masa depan data. Analisis seperti ini dapat membantu memberikan kita pemahaman yang lebih baik dari data pada umumnya.

2.2 Neural Network

2.2.1 Pengertian Neural Network (Jaringan Syaraf Tiruan)

Pada jaringan syaraf tiruan, juga terdapat istilah *neuron* atau sering juga disebut dengan *unit*, *sel* atau *node*. Setiap *neuron* terhubung dengan *neuron-neuron* lain melalui layer dengan bobot tertentu. Ada 2 jenis pelatihan dalam jaringan syaraf, ialah proses belajar terawasi (*supervised learning*) dan proses belajar tak terawasi (*unsupervised learning*).

2.3 Proses Belajar Terawasi (Supervised Learning)

Proses belajar pada jaringan ini adalah dengan memberikan data-data yang disebut training data atau training vectors. Training data terdiri masukan dan keluaran atau input-output yang diharapkan dan merupakan *associative memory* ().

2.2.3 Proses Belajar Tak Terawasi (Unsupervised Learning)

Proses pada jaringan ini adalah jaringan hanya diberi data input (input vectors), tetapi tanpa target vectors. Jaringan akan memodifikasi bobot sehingga untuk input yang hamper sama, output yang dihasilkan

sama (cluster units). Jaringan akan menghasilkan contoh-contoh untuk setiap cluster yang terbentuk (M.T., 2007).

2.4 Fungsi aktivasi

Dalam fungsi aktivasi terdapat fungsi *linear* dan *non linear*, untuk backpropagation menggunakan yang non linear karena untuk hasilnya jika menggunakan *linear* hasilnya akan sama dengan jaringan *single layer* (M.T., 2007). Ada beberapa macam fungsi aktivasi dalam *neural network* yaitu Fungsi Undak Biner (*Hard Limit*), Fungsi Undak Biner (*Threshold*), Fungsi Bipolar (*Symetric Hard Limit*), Fungsi Bipolar (dengan *threshold*), Fungsi Linear (Identitas), Fungsi Saturating Linear, Fungsi Symetric Saturating Linear, Fungsi Sigmoid Biner, Fungsi Sigmoid Bipolar. Dalam penelitian ini menggunakan fungsi Sigmoid.

2.5 Fungsi Sigmoid Biner

Fungsi ini digunakan untuk jaringan syaraf yang dilatih dengan menggunakan metode backpropagation. Fungsi sigmoid biner memiliki nilai pada range 0 sampai 1. Oleh karena itu, fungsi ini sering digunakan untuk jaringan syaraf yang membutuhkan nilai output yang terletak pada interval 0 sampai 1. Namun, fungsi ini bisa juga digunakan oleh jaringan syaraf yang nilai outputnya 0 atau 1.

Fungsi sigmoid biner dirumuskan sebagai :

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

Dengan :

$$f'(x) = \sigma(x)[1 - f(x)]$$

2.6 Backpropagation

Dalam backpropagation terdapat beberapa langkah-langkah untuk menyelesaikan masalahnya, dimana ciri dari metode ini adalah meminimalkan error

pada output yang dihasilkan oleh jaringan (M.T., 2007).

Pelatihan dengan metode backpropagation terdiri dari tiga langkah, yaitu:

- Data dimasukkan ke input jaringan (*feedforward*)
- Perhitungan dan propagasi balik dari error yang bersangkutan
- Pembaharuan (*adjustment*) bobot dan bias

Secara detail, langkah-langkah pelatihan Backpropagation adalah sebagai berikut:

Langkah 0 : Inisialisasi bobot dan bias.

Baik bobot maupun bias dapat diset dengan sembarang angka(acak)dan biasanya angka disekitar 0 dan 1 atau -1(bias positif atau negatif).

Langkah 1 : Jika stopping condition belum dipenuhi, maka langkah 2-9 dijalankan.

Langkah 2 : Data training, lakukan langkah 3-8.

Umpan Maju (Feedforward)

Langkah 3 : Tiap unit input ($x_i, i=1, \dots, n$) menerima isyarat masukan x_i dan diteruskan ke unit-unit tersembunyi (hidden unit)

Langkah 4 : Tiap unit tersembunyi ($z_j, z=1, \dots, p$) menjumlahkan bobot sinyal input dan biasnya.

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

⇒ keluaran dari lapisan unit yang tersembunyi menerapkan fungsi aktivasi hitung:

$$z_j = f(z_in_j)$$

Langkah 5 : Tiap unit keluaran ($y_k, k=1, \dots, m$) menjumlahkan isyarat masukan berbobot dengan menerapkan fungsi aktivasi hitung :

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

Perambatan Galat Mundur (Backpropagation)

Langkah 6 : Tiap unit keluaran ($y_k, k=1, \dots, m$) menerima pola pelatihan masukannya. Hitung galat (error) informasinya:

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

Untuk menghitung koreksi bobot:

$$y_k = f(y_in_k)$$

dan biasnya:

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

Langkah 7 : Tiap unit tersembunyi ($z_j, z=1, \dots, p$) menjumlahkan delta masukannya (dari unit-unit yang berada pada lapisan atasannya).

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

Hitung galat (error) informasinya :

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

Hitung koreksi bobot dan biasnya :

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

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

Langkah 8 : Tiap unit keluaran ($y_k, k=1, \dots, m$) memperbaharui bobot dan bias ($j=0, 1, \dots, p$)

Tiap unit tersembunyi ($z_j, z=1, \dots, p$) memperbaharui bobot dan biasnya ($i=0, 1, \dots, n$)

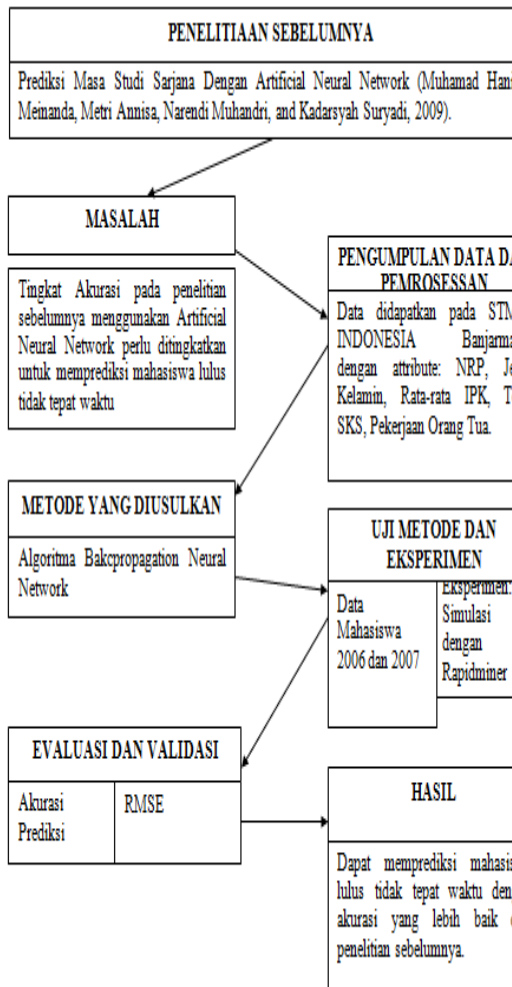
$$w_{jk}(\text{baru}) = w_{jk}(\text{lama}) + \Delta w_{jk}$$

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

Langkah 9 : Uji syarat berhenti

2.7 Kerangka Pemikiran

Penelitian ini dilakukan untuk mengamati dan memprediksi kelulusan siswa lulus tidak tepat waktu pada data angkatan 2006-2007 sebagai sample data nya. Berikut adalah kerangka pemikiran dalam bentuk gambar.



Gambar 2-1 Kerangka Pemikiran

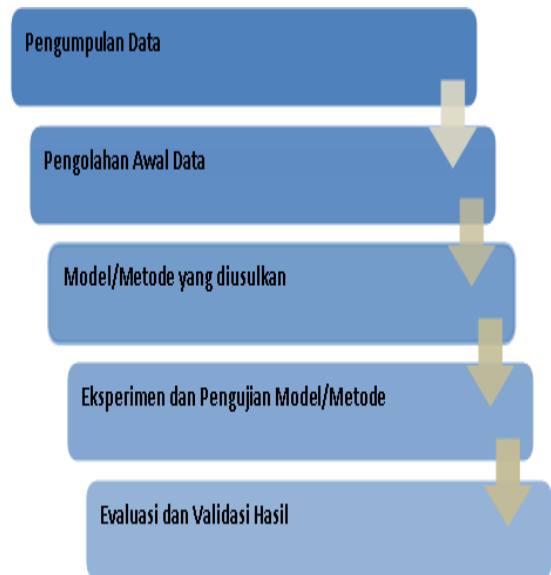
2.8 Sekolah Menengah Kejuruan

Sekolah Menengah Kejuruan adalah kelanjutan pendidikan menengah pertama yang diselenggarakan untuk menyiapkan peserta didik menjadi anggota masyarakat yang memiliki kemampuan akademik dan atau profesional yang dapat menerapkan, mengembangkan dan menciptakan ilmu pengetahuan.

Pendidikan profesional adalah pendidikan yang diarahkan terutama pada kesiapan penerapan keahlian tertentu dan dipersiapkan untuk memasuki tahap pendidikan selanjutnya seperti akademi, politeknik, sekolah tinggi, institut, dan universitas.

Beban studi sekurang-kurangnya 3 tahun untuk 6 (enam) semester.

3. METODE PENELITIAN



Gambar 3-1. Tahapan Metode Penelitian Eksperimen

3.1 Metode Pengumpulan Data

Data yang digunakan oleh peneliti dalam prediksi siswa lulus tidak tepat waktu:

a. Data Primer

Adalah data yang diperoleh secara langsung dari sumber, yaitu diperoleh secara langsung dari SMK Negeri 1 Kertak Hanyar yaitu dari data siswa angkatan 2006 dan 2007 dan mengambil 10% dari data awal sebagai data uji data sebanyak 363 record.

b. Data Sekunder

Selain data primer, peneliti juga menggunakan data sekunder yaitu data yang diperoleh secara tidak langsung, misalnya dari dokumentasi, literatur, buku, jurnal, dan informasi lainnya yang ada hubungannya dengan masalah yang diteliti.

3.2 Metode Pengolahan Awal Data

Data yang digunakan adalah data informasi akademik sekolah, data terdiri dari nomor induk siswa, tempat tanggal lahir, umur dari siswa, Nilai kumulatif

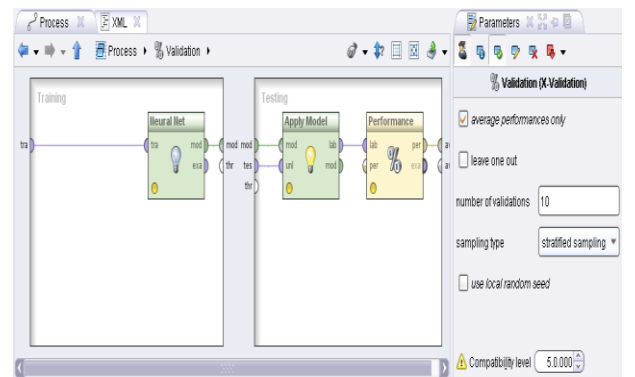
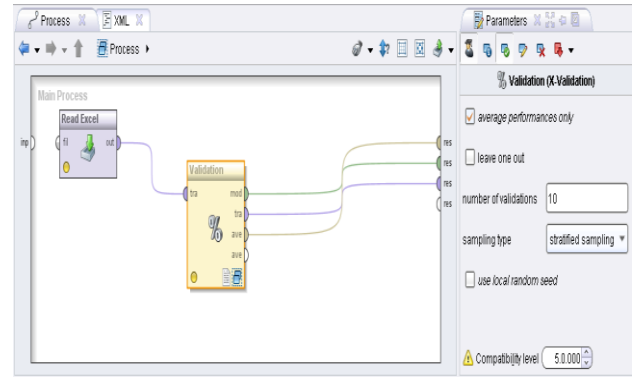
siswa semester 1 sampai semester 75, berapa jumlah SKS yang diambil, data diambil dari tahun 2006 dan 2007 sebanyak 363 Siswa dan diambil dari data awal 10 % menjadi 36 siswa untuk data uji. Dari data yang sudah ada di pecah lagi menjadi tabel diatas sehingga terdapat beberapa attribute yaitu NRP, jenis kelamin, umur, rata-rata raport, total SKS, pekerjaan orang tua, serta kelulusan.

3.3 Metode Yang Diusulkan

Metode yang diusulkan untuk penelitian ini ialah Neural Network dengan Algoritma Backpropagation, algoritma nanti akan di implementasikan menggunakan software Rapidminer 5.1.001. Untuk data pada siswa angkatan 2007 pada Neural Network Backpropagation didapat beberapa X1 sampai X5, dimana, Jenis Kelamin(1,2), Umur (1,2,3), Rata-rata Raport (1,2,3,4,5), Total SKS, Pekerjaan Orang Tua (1,2,3,4,5,6) dan Output yaitu Lulus (1,2) sehingga didapat untuk arsitektur neural network backpropagation X1 sampai dengan X5, terdapat ada 3 hidden layer untuk Z1,Z2,Z3, dan terdapat output Y1, Y2.

3.4 Eksperimen dan Pengujian Model/Metode dengan Matlab

Pelatihan dilakukan dengan data angkatan 2006 dan 2007 yang mana kedua angkatan tersebut digabung menjadi 1, untuk pengujian awal dilakukan menggunakan rapidminer, yang mana memasukkan data menggunakan file berformat excel, setelah itu menggunakan validasi dari rapidminernya untuk melakukan uji model menggunakan BPNN.



Gambar 3-2 Contoh untuk pengujian data menggunakan rapidminer

3.5 Evaluasi dan Validasi Hasil

Evaluasi dan validasi hasil dilakukan dengan menganalisa hasil dari prediksi kelulusan tepat waktu melalui backpropagation neural network. Pengukuran data menggunakan RMSE(Root Mean Square Error).

4. HASIL PENELITIAN DAN PEMBAHASAN

4.1 Hasil Eksperimen dan Pengujian Model

Data yang akan dieksperimenkan dan diuji ialah:

TABEL 4.1 TABEL HASIL PELATIHAN

NRP	JENIS KELAMIN	UMUR	Rata2 IPK 1-4	Total Sks	Pekerjaan Orang Tua	LULUS
06013060	1	1	3	72	6	1
06013061	1	2	2	80	3	2

Untuk pengujiannya menggunakan Software data mining yaitu Rapidminer, untuk uji pertama melalui data sample yaitu data angkatan 2006 dan 2007, pada bagian Nama, Tempat Lahir dan Tanggal Lahir akan dihilangkan untuk mendapatkan akurasi yang lebih tinggi, pada bagian Lulus atau keterangan untuk lulus ada 2 kategori yaitu tepat dan tidak tepat waktu atau terlambat.

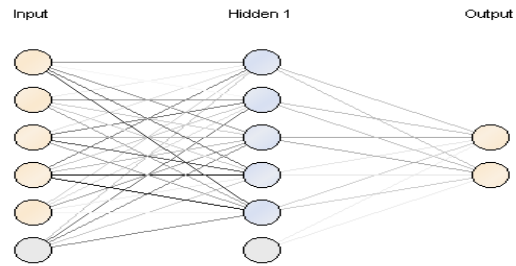
4.2 Uji Akurasi

Hasil Pengujian didapat prediksi kurs Rupiah terhadap US Dollar tahun 2010

NRP	L/P	UMUR	Rata-rata
6013055	1	1	2
6013056	1	1	2
6013057	1	1	2
6013059	1	1	2
6013060	1	1	3
6013061	1	2	2
6013062	1	1	2
6013063	1	1	2
6013064	1	1	2
6013066	1	1	2
6013067	2	1	3
6013068	2	1	2
6013069	1	1	2
6013070	2	1	2

Gambar 4-2. Uji akurasi dari data melalui software Rapidminer

Dimana diatas terdapat beberapa attribute yang mana masing – masing ialah NRP diberikan sebagai ID dan untuk nilainya integer, jenis kelamin sebagai attribute dan nilainya integer, umur sebagai attribute dan nilainya integer, rata-rata nilai raport sebagai attribute dan nilainya integer, Total SKS sebagai attribute, dan LULUS atau keterangan lulus sebagai Label yang mana keterangan lulus ini berposisi sebagai target, untuk uji akurasi yang mana pada software rapidminer akurasinya diberikan nilai binomial, binomial adalah suatu nilai dimana terdapat 3 atau lebih variasi untuk nilai variabelnya.



Gambar 4-2. Jaringan BPNN Pada data 2006/2007 di Rapidminer

Pada gambar diatas adalah bentuk dari arsitektur backpropagation dari data angkatan 2006 dan 2007 yang mana diuji dan dihasilkan oleh software dari rapidminer sehingga mendapatkan seperti pada gambar diatas, dalam gambar tersebut terdapat ada 3 layer yaitu layer input, layer hidden dan layer output, yang mana pada layer input terdapat 6 inputan termasuk untuk bias 1, pada layer hidden terdapat 6 hidden layer termasuk untuk bias 1 dan untuk output terdapat 2 yang mana untuk 2 output disana ialah keterangan dari lulus untuk Tepat dan tidak tepat atau Terlambat.

Gambar 4-3. Akurasi Prediksi siswa lulus tidak tepat waktu

accuracy: 79.08% +/- 3.67% (mikroc: 79.06%)			
	true 2	true 1	class precision
pred. 2	167	45	78.77%
pred. 1	31	120	79.47%
class recall	84.34%	72.73%	

Pada gambar didapatkan hasil akurasi sebesar 79.08% menggunakan software RapidMiner.

precision: 80.33% +/- 6.53% (mikroc: 79.47% (positive class: 1))			
	true 2	true 1	class precision
pred. 2	167	45	78.77%
pred. 1	31	120	79.47%
class recall	84.34%	72.73%	

Gambar 4.4. Arsitektur hasil terbaik dari metode Levenberg-Marquardt

Pada gambar untuk precisionnya didapatkan hasil sebesar 80.33% +- 6.53%.

Pada gambar ini menunjukkan recall 72.46% +- 10.55%, dan terdapat secara otomatis didapatkan kurva atau Area Under Curve (AUC) dimana AUC Optimistic 0.847 +- 0.048 dan AUC Premistic 0.843 +- 0.050.

4.3 Mean Square Error (MSE)

Pada RapidMiner untuk mendapatkan nilai dari MSE nya ialah dengan merubah status nilai variable pada keterangan Lulus yang mana posisinya sebagai Label dan nilai asalnya Polynomial menjadi nilai variable integer, untuk NRP tetap sebagai ID, dan yang lain tetap sebagai attribute, untuk keterangan lebih lengkap bisa dilihat pada gambar dibawah ini.

NRP	L/P	UMUR	Rata2 IPK 1-	Total Sks	Perkerjaan t	LULUS
integer	integer	integer	integer	integer	integer	integer
(id)	(attribute)	(attribute)	(attribute)	(attribute)	(attribute)	(label)
6013055	1	1	2	80	1	2
6013056	1	1	2	70	1	1
6013057	1	1	2	76	2	1
6013059	1	1	2	86	2	2
6013060	1	1	3	72	6	1
6013061	1	2	2	80	3	2
6013062	1	1	2	86	2	2
6013063	1	1	2	80	2	2
6013064	1	1	2	80	1	2
6013066	1	1	2	78	2	2
6013067	2	1	3	70	1	1
6013068	2	1	2	80	1	2
6013069	1	1	2	82	1	2
6013070	2	1	2	78	1	2
6013071	1	1	2	80	1	2

Gambar 4-8. Perubahan data menjadi integer

Dari proses diatas dalam rapidminer akan mendapatkan nilai untuk Mean Square Error(MSE) seperti akan terlihat pada gambar yang ada dibawah ini. Dari keterangan pada gambar diatas untuk errornya didapat sebesar 0.500 +/- 0.092 untuk data dari angkatan 2006 dan 2007.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

- Dengan penerapan *Neural Network* untuk prediksi siswa lulus tidak tepat

waktu berhasil dilakukan untuk tujuan mencari siswa lulus tidak tepat waktu.

- Pada penelitian ini pengelompokan data dan pengolahan awal data dapat sangat berpengaruh pada tingkat akurasi yang dihasilkan.
- Besarnya nilai *RMSE* dan kurang akuratnya prediksi yang dihasilkan dikarenakan data yang diperoleh kurang valid, faktor kesalahan manusia pada survei ke data siswa STMIK Indonesia banjarmasin terdapat banyak sekali data yang kurang atau data yang tidak ada (*missing value*).
- 363 siswa yang digabung menjadi 1 kelompok yaitu angkatan 2006 dan 2007 sangat sedikit, dikarenakan data-data yang ada terdapat banyak *missing value* sehingga sulit sekali untuk mendapatkan semua data persatu angkatan, data yang sedikit akan berpengaruh pada tingkat akurasi.
- Dalam hasil yang didapat pada penelitian siswa angkatan 2007 yang mana datanya sebanyak 363 siswa yang mana memiliki attribute sebanyak 5 attribute dan diuji melalui algoritma backpropagation neural network maka didapat tingkat akurasi 79.08% dan untuk uji kesalahan atau *Mean Square Error (MSE)* didapat 0.500 +/- 0.092.

5.2 Saran

- Untuk menghasilkan tingkat akurasi yang akurat dari penelitian ini, diperlukan pembersihan data (*data cleansing*) dari masukkan data yang tidak konsisten dan data yang rusak atau yang disebut dengan data sampah, pada tahap pengolahan awal data.
- Untuk attribut yang baik, lebih baik dilakukan *clustering data* dari data yang sudah ada, sehingga untuk prediksi dapat lebih maksimal dengan attribut data yang ada.
- Dan untuk menghasilkan akurasi yang lebih optimal diperlukan penelitian tahap berikutnya, yaitu dengan

mengoptimasikan *Neural Network* dengan pendekatan komputasi cerdas lainnya. Contoh optimasi backpropagation itu sendiri.

- Untuk saran akan lebih baik penelitian ini dikembangkan pada sebuah universitas yang mana terdapat siswa dari semua jurusan sehingga data yang didapat besar.

DAFTAR PUSTAKA

- Larose, D. T. *Discovering Knowledge in Data*.
M.T., P. D. (2007). *Paradigma Sistem Cerdas*.
jakarta: Ir. Kuswara Setiawan.
- Nghe, N. T., Janecek, P., & Haddawy, P.
(2007). A Comprative Analysis of
Techniques for Predicting Academic
Performance.
- www.dikti.go.id. (2007). Diambil kembali dari
www.dikti.go.id/Archive2007/kepmen23
2-2000.txt.