

Pemodelan Faktor-faktor yang Mempengaruhi Prestasi Mahasiswa Pasca Sarjana ITS dengan Regresi Logistik dan *Neural Network*

Wijdani Anindya Hadi dan Suhartono

Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111

E-mail: w.anindyahadi@gmail.com⁽¹⁾ suhartono@statistika.its.ac.id⁽²⁾

Abstrak—Aspek yang secara langsung dapat dilihat ketika mahasiswa yang baru saja lulus adalah lama studi yang ditempuh mahasiswa tersebut apakah tepat waktu atau melebihi waktu yang ditentukan juga Indeks Prestasi atau yang lebih dikenal dengan IP. Metode yang digunakan untuk memodelkan faktor-faktor yang mempengaruhi prestasi mahasiswa Pasca Sarjana ITS dengan respon kategorik adalah regresi logistik dan *Neural Network*. Regresi logistik pada variabel respon IPK faktor yang mempengaruhi adalah TOEFL, Usia, IPK saat S1, Lama Tunggu, Status Perkawinan, dan Asal perguruan Tinggi saat S1. Sedangkan untuk variabel respon lama tempuh studi yaitu skor TPA, kesesuaian bidang, IPK S1, status perkawinan dan asal perguruan tinggi S1. Pada metode *Neural Network* neuron pada *hidden layer* dengan seluruh variabel untuk kategori IPK dan lama tempuh studi masing-masing jumlahnya sebanyak 8 dan 10 neuron. Sedangkan dengan menggunakan variabel yang signifikan baik IPK maupun lama tempuh studi *hidden layer* yang terbentuk sebanyak 8 neuron. Ketepatan klasifikasi yang tertinggi dengan menggunakan metode *Neural Network*.

Kata Kunci—Pasca Sarjana, Regresi Logistik Biner, *Neural Network*, neuron

I. PENDAHULUAN

EVALUASI kemajuan belajar setiap mahasiswa Pasca Sarjana di Institut Teknologi Sepuluh Nopember (ITS) Surabaya dilakukan setiap akhir semester yang dicapai mulai semester awal masa studi. Untuk dapat melanjutkan studi, seorang mahasiswa harus dapat memenuhi kriteria yang telah diterapkan, dan pada dasarnya harus dapat menyelesaikan masa studinya dalam jangka waktu dua tahun. Tetapi, tidak semua mahasiswa Pasca Sarjana dapat menyelesaikan studinya tepat waktu.

Pada penelitian ini, metode yang digunakan untuk memodelkan faktor-faktor yang mempengaruhi prestasi mahasiswa Pasca Sarjana ITS dengan respon kategorik adalah regresi logistik biner karena hanya mempunyai 2 kategori di masing-masing responnya. Selain regresi logistik biner, klasifikasi juga dapat dilakukan menggunakan jaringan saraf tiruan (*Neural Network*). *Neural Network* adalah salah satu representasi buatan dari otak manusia yang selalu mencoba untuk mensimulasikan proses pembelajaran pada otak manusia tersebut [1]. Dalam beberapa kasus, *Neural Network*

memberikan ketepatan hasil klasifikasi yang lebih baik jika dibandingkan dengan regresi logistik, namun hal tersebut belum dapat berlaku secara umum jika diterapkan pada data simulasi.

II. LANDASAN TEORI

A. Mahasiswa

Mahasiswa dalam peraturan pemerintah RI No.30 tahun 1990 adalah peserta didik yang terdaftar dan belajar di perguruan tinggi tertentu.

B. Lama Tempuh Studi

Pada Peraturan Akademik BAB VII tentang Evaluasi Masa Studi bagian ketiga Program Magister yang tertera pada Pasal 22 menyatakan bahwa masa studi paling lama untuk mahasiswa program magister adalah 8 semester dimana evaluasi masa studi mahasiswa dilakukan setiap semester, dimulai pada semester dua.

C. Regresi Logistik Biner

Regresi logistik merupakan suatu metode analisis data yang mendeskripsikan antara variabel respon dengan satu atau lebih variabel prediktor. Regresi logistik biner variabel responnya yang bersifat dikotomis yang terdiri dari dua kategori yaitu 0 dan 1, sehingga variabel respon akan mengikuti distribusi Bernoulli dengan fungsi probabilitas sebagai berikut [2].

$$f(y_i) = \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \text{ dengan } y_i = 0,1$$

Berdasarkan [2] model regresi logistik adalah sebagai berikut.

$$\pi(x) = \frac{\exp(g(x))}{1 + \exp(g(x))} \quad (1)$$

Metode *Maximum Likelihood Estimator* (MLE) adalah metode yang digunakan untuk menduga parameter-parameter yang terdapat dalam model regresi logistik. Metode ini menduga β dengan memaksimalkan fungsi *likelihood*.

Fungsi *likelihood* yang dimaksimalkan adalah

$$L(\beta) = \ln(l(\beta)) = \sum_{j=0}^p \left[\sum_{i=1}^n y_i x_{ij} \right] \beta_j - \sum_{i=1}^n \ln \left[1 + \exp \left(\sum_{j=0}^p \beta_j x_{ij} \right) \right] \quad (2)$$

Persamaan dideferensialkan terhadap β , setelah dideferensialkan terhadap β kemudian disamakan dengan nol, namun cara ini sering kali diperoleh hasil yang implisit sehingga dilakukan metode iterasi Newton Rhapsion untuk memaksimalkan fungsi likelihood [3].

Pengujian parameter dalam regresi logistik biner dilakukan baik secara serentak maupun individu. Statistik uji yang digunakan dalam uji serentak adalah statistik uji G atau *likelihood ratio test*. Sedangkan statistik uji yang digunakan dalam uji parsial adalah statistik uji Wald [4].

Salah satu ukuran yang digunakan untuk menginterpretasi koefisien variabel prediktor adalah *Odds ratio*. Tabel 2.1 menunjukkan nilai model regresi logistik dengan variabel prediktor bersifat dikotomis.

Tabel 1.
Nilai Model Regresi Logistik dengan Variabel Prediktor Bersifat Dikotomis.

Variabel Respon (Y)	Variabel Prediktor (X)	
	$x = 1$	$x = 0$
$y = 1$	$\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$
$y = 0$	$1 - \pi(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}}$	$1 - \pi(0) = \frac{1}{1 + e^{\beta_0}}$
Total	1	1

Statistik uji yang dipakai untuk uji kesesuaian model adalah statistik *Hosmer-Lemeshow* (\hat{C})

D. *Neural Network*

Neural Network biasa digunakan dalam masalah klasifikasi (di mana *outputnya* adalah variabel kategoris) atau regresi (*outputnya* kontinu). *Neural Network* dimulai dengan *layer input*, dimana tiap simpul berkorespondensi dengan variabel prediktor. Simpul-simpul *input* ini terhubung ke beberapa simpul dalam *hidden layer*. Tiap simpul *input* terhubung dengan tiap simpul dalam *hidden layer*. Simpul dalam *hidden layer* bisa jadi terhubung ke simpul lain dalam *hidden layer*, atau ke *output layer*. *Output layer* terdiri dari satu atau beberapa variabel respon.

Tiga hal yang membedakan antara *Neural Network* yang satu dengan yang lain adalah arsitektur jaringan, pelatihan jaringan dan fungsi aktivasi yang digunakan [3]. *Neural Network* merupakan suatu bentuk arsitektur yang berdistribusi paralel dengan sejumlah besar *node* dan hubungan antara tiap *node* tersebut. Titik hubung antara *node* satu dengan *node* yang lain mempunyai nilai yang diasosiasikan dengan bobot.

Fungsi aktifasi yang ada pada *Neural Network* antara lain :

- a. Fungsi Identitas
 $f(x) = x$, untuk semua nilai x
- b. Fungsi Sigmoid
 $f(x) = \frac{1}{1 + e^{-x}}$
- c. Fungsi Threshold
 $f(x) = \begin{cases} 1, & \text{jika } x \geq 0 \\ 0, & \text{jika } x < 0 \end{cases}$
- d. Fungsi Tangent Hiperbola
 $f(x) = \tanh(x)$

e. Fungsi *softmax*

$$f(x) = \frac{e^{x_j}}{\sum_k e^{x_k}}$$

E. *Evaluasi Performansi Model*

Evaluasi performansi model yang digunakan pada penelitian ini adalah *classification accuracy*, *sensitivity*, dan *specificity*. *Sensitivity* merupakan ukuran ketepatan dari suatu kejadian yang diinginkan. *Specificity* merupakan suatu ukuran yang menyatakan persentase kejadian-kejadian yang tidak diinginkan.

Tabel 2.
Ketepatan Klasifikasi

Actual	Predicted	
	Positive = class 0	Negative = class 1
Positive = class 0	True Positive (TP)	False Negative (FN)
Negative = class 1	False Positive (FP)	True Negative (TN)

III. METODOLOGI

Data yang digunakan dalam penelitian kali ini merupakan data sekunder tentang prestasi yang meliputi lama tempuh dan status kelulusan dari mahasiswa Pasca Sarjana ITS periode lulusan 96-102. Data ini diperoleh dari Pasca Sarjana ITS.

Langkah-langkah analisis yang dilakukan pada penelitian ini adalah sebagai berikut.

1. Pengumpulan literatur yang berasal dari buku, jurnal serta *homepage* yang menunjang sumber ilmiah untuk penelitian. Tinjauan literatur yang digunakan terbagi menjadi dua antara lain dilihat dari aspek statistik dan aspek non-statistik. Pengumpulan data dari Pasca Sarjana ITS tentang prestasi mahasiswa Pasca Sarjana ITS periode kelulusan 96-102.
2. Membagi data menjadi dua yaitu *training* dan *testing* dimana data *training* digunakan untuk membentuk model dan data *testing* digunakan untuk melihat ketepatan model. Data dipartisi sebanyak lima kali yaitu 90:10, 80:20, 70:30, 60:40, dan 50:50 untuk masing-masing variabel respon dengan replikasi sebanyak 10 kali untuk masing-masing partisi data.
3. Melakukan pengolahan data dengan menggunakan Regresi Logistik dan *Neural Network*.

Regresi Logistik

- a. Melakukan pemodelan dengan regresi logistik dengan menggunakan seluruh variabel prediktor.
- b. Melakukan pengujian parameter model regresi logistik secara serentak maupun individu.
- c. Melakukan pengujian hipotesis untuk menentukan apakah model regresi logistik telah sesuai atau tidak.
- d. Melihat faktor-faktor apa saja yang mempengaruhi prestasi mahasiswa pasca sarjana ITS.
- e. Mengetahui hasil ketepatan klasifikasi berdasarkan model regresi logistik.
- f. Melakukan langkah a-e untuk variabel yang signifikan untuk masing-masing variabel respon.

Neural Network

- a. Melakukan *preprocessing* untuk mentransformasi data dengan rumus : $f = (z - \min) / (\max - \min)$.
- b. Menentukan banyak *input*. Pada *Neural Network* *input* merupakan variabel prediktor.

- c. Menentukan jumlah unit neuron pada *hidden layer*.
 - d. Menentukan optimasi *backpropagation*.
 - e. Melakukan pengolahan data dengan masing-masing variabel respon (*output*) dengan seluruh *input* kemudian memodelkannya.
 - f. Melakukan pengolahan data dengan masing-masing variabel respon (*output*) dengan *input* yang signifikan dari regresi logistik dan memodelkannya.
 - g. Melakukan pengolahan data dengan seluruh kategori pada variabel respon (*output*) dengan seluruh *input* kemudian memodelkannya.
 - h. Melakukan pengolahan data dengan seluruh kategori pada variabel respon (*output*) dengan *input* yang signifikan dari regresi logistik dan memodelkannya.
4. Membandingkan hasil ketepatan klasifikasi dari masing-masing metode yaitu Regresi Logistik dan *Neural Network*.

Variabel penelitian yang digunakan dalam mengevaluasi dan memprediksi prestasi mahasiswa Pasca Sarjana ITS periode lulusan 96-102 terdiri dari variabel respon (Y) dan variabel prediktor (X) berjumlah 1002 data yang terdiri dari jurusan yang mempunyai program Magister di ITS. Variabel pada penelitian ini ditunjukkan pada Tabel 3.

Tabel 3.
Variabel penelitian

Variabel Respon (Y)	
Y ₁ = Indeks Prestasi Kumulatif (IPK)	0 = < 3,50 1 = ≥ 3,50
Y ₂ = lama tempuh studi	0 = Tidak tepat waktu 1 = Tepat Waktu
Variabel Prediktor (X)	
Variabel	Keterangan
X1 = TPA	
X2 = TOEFL	
X3 = Usia	
X4 = Jenis Kelamin	0 = Perempuan 1 = Laki-laki
X5 = Kesesuaian Bidang	0 = Tidak Sesuai 1 = Sesuai
X6 = IPK saat S1	
X7 = Lama Tunggu	
X8 = Sumber Dana	0 = Biaya Sendiri 1 = Beasiswa
X9 = Status Perkawinan	0 = Belum Menikah 1 = Menikah 0 = Surabaya
X10 = Asal Daerah	1 = Luar Jawa Timur (tapi masih Pulau Jawa) 2 = Non Surabaya (Jawa Timur) 3 = Luar Jawa
X11 = Lulusan S1	0 = Lulusan ITS 1 = Lulusan Non-ITS Negeri 2 = Lulusan Non-ITS Swasta

IV. ANALISIS DAN PEMBAHASAN

A. Analisis Deskriptif

Mahasiswa Pasca Sarjana ITS dari seluruh jurusan yang mempunyai IPK ≥ 3,5 adalah sebanyak 515 orang sedangkan yang IPKnya <3,5 ada sebanyak 487 orang. Untuk mahasiswa yang lama tempuh studinya tepat waktu (≤ 4 semester) berjumlah 739 mahasiswa sedangkan untuk mahasiswa yang lama tempuh studinya tidak tepat waktu (>4 semester) berjumlah 263 orang. IPK mahasiswa Pasca Sarjana ITS yang < 3,5 adalah sebesar 3,29 dengan IPK minimum sebesar 2,68

dan maksimum sebesar 3,33. Sedangkan untuk rata-rata dari mahasiswa Pasca Sarjana yang IPK ≥ 3,5 adalah sebesar 3,68 dengan IPK minimum pada kategori ini adalah sebesar 3,51 dan IPK maksimum sebesar 3,98. Sedangkan lama tempuh studi dari kategori > 4 semester lama tempuh rata-ratanya adalah sebesar 5,5 semester sedangkan untuk lama tempuh ≤ 4 semester rata-rata lama tempuhnya adalah 3,78 semester.

B. Pemodelan Prestasi Mahasiswa Pasca Sarjana ITS dengan Regresi Logistik Biner

Sebelum melakukan analisis regresi logistik data dibagi menjadi *training* dan *testing* dengan partisi 50:50, 60:40, 70:30, 80:20 dan 90:10. Analisis dilakukan untuk partisi data 90:10 dengan melibatkan seluruh variabel dan variabel yang signifikan. Dengan α sebesar 5% diketahui bahwa *P-value*=0,000 menunjukkan bahwa *P-value* < α sehingga dapat dikatakan secara serentak koefisien yang didapatkan baik untuk IPK maupun lama tempuh studi yang melibatkan seluruh variabel ataupun variabel yang signifikan signifikan terhadap model regresi logistik biner.

Analisis regresi logistik parsial dengan menggunakan data partisi 90:10, Variabel yang digunakan dalam pembentukan model adalah seluruh variabel yang signifikan dari pemodelan awal dengan variabel respon IPK (y1) dan lama tempuh studi (y2).

Metode yang digunakan dalam pembentukan model adalah metode *enter* dengan memasukkan semua variabel prediktor. Dengan menggunakan partisi data *training* dan *testing* 90:10. Prestasi mahasiswa Pasca Sarjana ITS dilihat dari IPK (y1) dipengaruhi oleh skor TOEFL, usia, IPK saat S1, status perkawinan, dan perguruan tinggi saat lulus S1. Dengan *p-value* < α. Sehingga fungsi logit dari model tersebut adalah sebagai berikut :

$$\pi(x) = \frac{1}{1 + \exp(-(-3,1288 + 0,0070X_1 - 0,0822X_2 + 1,3346X_3 + 0,0726X_4 + 0,3883X_5))}$$

Prestasi mahasiswa Pasca Sarjana ITS dilihat dari lama tempuh studi (y2) dipengaruhi oleh skor TPA, kesesuaian bidang, IPK saat S1, status perkawinan dan asal perguruan tinggi saat S1. Dengan *p-value* < α model logitnya adalah sebagai berikut :

$$\pi(x) = \frac{1}{1 + \exp(-(-0,8280 - 0,0170X_1 + 0,6807X_2 + 0,3883X_3 + 0,4864X_4))}$$

Sedangkan dengan melibatkan variabel yang signifikan dari regresi logistik sebelumnya untuk variabel IPK (y1) model logit yang didapatkan adalah sebagai berikut :

$$\pi(x) = \frac{1}{1 + \exp(-(-3,1266 + 0,0082X_1 - 0,0732X_2 + 1,3334X_3 + 0,0693X_4))}$$

Untuk model logit yang terbentuk dari variabel yang signifikan dari regresi logistik sebelumnya untuk variabel lama tempuh studi (y2) model logit yang didapatkan adalah sebagai berikut :

$$\pi(x) = \frac{1}{1 + \exp(-(-1,5056 + 0,6704X_1 + 1,0591X_2 - 1,0962X_3))}$$

Berikut merupakan interpretasi koefisien parameter berdasarkan nilai *odds ratio* dengan menggunakan partisi data *training* dan *testing* 90:10 untuk variabel respon IPK :

Tabel 4.
Nilai Odds Ratio IPK

Variabel	odds ratio
TOEFL	1,08
Usia	0,93
IPK S1	1,96
Lama Tunggu	1,07
Status (1)	1,69
Lulusan S1 (1)	0,43
Lulusan S1 (2)	0,29

Berdasarkan pada Tabel 4 besarnya pengaruh masing-masing variabel prediktor yang signifikan dapat dijelaskan berdasarkan odds ratio sebagai berikut.

- 1) Skor TOEFL
Mahasiswa yang skor TOEFLnya meningkat sebanyak 10 satuan maka berpeluang untuk memiliki IPK $\geq 3,5$ meningkat sebesar 1,01 kali.
- 2) Usia
Sebesar 0,93 kali mahasiswa yang berusia lebih tua 1 tahun, berpeluang memiliki IPK $\geq 3,5$.
- 3) IPK S1
Mahasiswa yang IPK S1nya mengalami kenaikan sebanyak 0,5 satuan maka berpeluang untuk mendapatkan IPK $\geq 3,5$ meningkat 1,96 kali.
- 4) Lama Tunggu
Mahasiswa pasca sarjana yang lama tunggunya lebih besar satu satuan akan berpeluang memiliki IPK $\geq 3,5$ sebesar 1,07 kali.
- 5) Status Pernikahan
Mahasiswa Pasca sarjana yang berstatus belum menikah akan memiliki peluang sebesar 1,69 kali untuk mendapatkan IPK $\geq 3,5$ dibandingkan dengan mahasiswa yang telah menikah.
- 6) Perguruan Tinggi S1
Mahasiswa Pasca sarjana lulusan Perguruan Tinggi Negeri selain ITS memiliki peluang 0,43 kali untuk memiliki IPK $\geq 3,5$ dibandingkan mahasiswa ITS. Sedangkan mahasiswa yang berasal dari perguruan tinggi swasta memiliki peluang 0,29 kali untuk mendapatkan IPK $\geq 3,5$ dibandingkan dengan mahasiswa ITS.

Untuk interpretasi variabel respon lama tempuh studi adalah sebagai berikut :

Tabel 5.
Nilai Odds Ratio Lama Tempuh

Variabel	odds ratio
Kesesuaian Bidang ₍₁₎	1,96
IPK S1	1,69
Lulusan S1 ₍₂₎	0,33

- 1) Kesesuaian Bidang
Mahasiswa yang bidang sesuai akan memiliki peluang sebesar 1,96 kali untuk tepat waktu dalam menempuh lama studi dibandingkan dengan mahasiswa yang bidangnya tidak sesuai.

- 2) IPK S1
Mahasiswa yang IPK S1nya mengalami kenaikan sebanyak 10 satuan maka berpeluang untuk tepat waktu dalam menempuh studi meningkat 1,69%.
- 3) Perguruan Tinggi S1
Mahasiswa Pasca sarjana yang berasal dari perguruan tinggi swasta memiliki peluang 0,33 kali untuk menempuh lama studi lebih dari 4 semester atau tidak tepat waktu dibandingkan dengan mahasiswa ITS.

C. *Pemodelan Prestasi Mahasiswa Pasca Sarjana ITS dengan Neural Network*

Setelah memodelkan dengan regresi logistik biner selanjutnya data Prestasi mahasiswa Pasca Sarjana ITS akan dimodelkan dan dilihat ketepatan klasifikasinya dengan menggunakan metode *Neural Network*. Arsitektur yang digunakan adalah *Multilayer Perceptron* dan fungsi aktivasi yang digunakan adalah *softmax*. Pada *Neural Network* variabel prediktor akan digunakan sebagai *input* sehingga, *input* yang digunakan adalah keseluruhan variabel untuk masing-masing variabel respon sehingga terbentuk dengan sebelas *input*. Dengan partisi data 90:10 jumlah *neuron* pada *hidden layer* pada variabel IPK yang menunjukkan ketepatan klasifikasi yang tertinggi adalah sebanyak 8 *neuron* sedangkan pada lama tempuh studi ketepatan klasifikasi yang tertinggi dengan menggunakan 10 *neuron*.

Untuk Neural Network dengan menggunakan *input* yang signifikan pada partisi data 90:10 ternyata didapatkan ketepatan klasifikasi yang tertinggi baik untuk variabel IPK maupun lama tempuh studi dengan jumlah *neuron* pada *hidden layer* yaitu sebanyak 8 *neuron*. Berikut merupakan variabel-variabel yang signifikan dari seluruh partisi data.

Tabel 6.
Variabel yang Signifikan sebagai Input

Variabel	90:10		80:20		70:30		60:40		50:50	
	y1	y2	y1	y2	y1	y2	y1	y2	y1	y2
x1	x	v	x	v	x	v	x	v	x	x
x2	v	x	v	v	v	x	x	x	v	x
x3	v	x	v	x	v	x	v	v	v	v
x4	x	x	x	x	v	x	x	x	x	x
x5	x	v	x	v	x	v	x	v	x	x
x6	v	v	v	v	v	v	v	v	x	v
x7	v	x	v	x	x	x	x	x	v	v
x8	x	x	x	x	x	x	x	x	x	x
x9	v	v	v	x	v	x	x	x	v	x
x10	x	x	x	x	x	x	x	x	x	x
x11	v	v	v	v	v	v	v	v	v	v

v) variabel signifikan

D. *Perbandingan antara Regresi Logistik dengan Neural Network*

Setelah dilakukan pembahasan untuk masing-masing variabel respon yaitu IPK dan lama tempuh studi baik melibatkan seluruh variabel prediktor maupun dengan variabel yang signifikan kemudian akan dibandingkan ketepatan klasifikasi dari setiap metode. Adapun hasil perbandingan klasifikasinya adalah sebagai berikut:

Tabel 7.
Perbandingan Ketepatan Klasifikasi

partisi data	Regresi Logistik				Neural Network			
	seluruh variabel		variabel yang signifikan		seluruh variabel		variabel yang signifikan	
	y1	y2	y1	y2	y1	y2	y1	y2
90:10	74,25	80,19	73,26	73,27	72,27	79,2	76,23	80,19
80:20	68,15	75,12	67,16	74,62	64,67	71,05	70,14	76,85
70:30	65,11	72,86	67,1	59,68	61,12	71,85	69,1	73,81
60:40	71,25	74,31	65,5	73,81	64,75	75,27	65,75	78,65
50:50	67,46	73,85	69,44	73,85	65,26	67,46	69,67	74,23

Dari Tabel 5 dapat diketahui bahwa ketepatan klasifikasi yang terbesar untuk variabel IPK dan lama tempuh studi dengan melibatkan seluruh variabel prediktor adalah dengan menggunakan metode regresi logistik. Sedangkan untuk klasifikasi untuk variabel yang signifikan baik untuk variabel IPK maupun lama tempuh studi ketepatan klasifikasinya lebih baik dengan menggunakan metode Neural Network. Tetapi, untuk kasus ini dapat dikatakan Neural Network merupakan metode yang memiliki ketepatan klasifikasi yang tinggi.

V. KESIMPULAN DAN SARAN

Setelah dilakukan analisis dan pembahasan maka kesimpulan yang dapat diambil adalah sebagai berikut:

1. Sebanyak 51% mahasiswa pasca sarjana ITS periode lulusan 96-102 mempunyai IPK < 3,5 dengan rata-rata IPK sebesar 3,29 dan 49% yang memiliki IPK ≥ 3,5 dengan rata-rata 3,68. Untuk lama tempuh studi sebanyak 74% mahasiswa pasca sarjana ITS tepat waktu dalam menempuh studi sedangkan sisanya tidak tepat waktu.
2. Faktor-faktor yang mempengaruhi prestasi mahasiswa pasca sarjana ITS periode lulusan 96-102 pada partisi data 90:10 untuk variabel IPK adalah skor TOEFL, Usia, IPK saat S1, Lama Tunggu, Status Perkawinan, dan Asal perguruan Tinggi saat S1. Sedangkan untuk variabel lama tempuh faktor-faktor yang mempengaruhi skor TPA, kesesuaian bidang, IPK saat S1, status perkawinan dan asal perguruan tinggi saat S1 jika di tabelkan untuk seluruh partisi data adalah sebagai berikut:

Variabel	90:10		80:20		70:30		60:40		50:50	
	y1	y2	y1	y2	y1	y2	y1	y2	y1	y2
x1	x	v	x	v	x	v	x	v	x	x
x2	v	x	v	v	v	x	x	x	v	x
x3	v	x	v	x	v	x	v	v	v	v
x4	x	x	x	x	v	x	x	x	x	x
x5	x	v	x	v	x	v	x	v	x	x
x6	v	v	v	v	v	v	v	v	x	v
x7	v	x	v	x	x	x	x	x	v	v
x8	x	x	x	x	x	x	x	x	x	x
x9	v	v	v	x	v	x	x	x	v	x
x10	x	x	x	x	x	x	x	x	x	x
x11	v	v	v	v	v	v	v	v	v	v

v) faktor-faktor yang mempengaruhi

3. Dengan metode Neural Network dari partisi data 90:10 untuk variabel IPK dengan melibatkan seluruh prediktor membentuk ketepatan klasifikasi yang tertinggi dengan

jumlah neuron pada *hidden layer* sebanyak 8 neuron. Untuk variabel lama tempuh studi jumlah neuron pada *hidden layer* yang membentuk ketepatan klasifikasi yang tertinggi yaitu dengan jumlah 10 neuron. Untuk Neural Network yang melibatkan variabel yang signifikan baik untuk variabel IPK dan lama tempuh studi ketepatan klasifikasi yang tertinggi yaitu dengan jumlah neuron pada *hidden layer* sebanyak 8 neuron.

4. Jika dilihat dari ketepatan klasifikasi yang paling tinggi dari seluruh partisi data didapatkan sebesar 80,19% tepat diklasifikasikan baik dengan regresi logistik maupun neural network dengan partisi data 90:10. Ketepatan klasifikasi antara Regresi Logistik dan *Neural Network* dengan partisi data 90:10 untuk variabel IPK dan lama tempuh studi yang melibatkan seluruh prediktor yang paling besar adalah dengan menggunakan regresi logistik yaitu sebesar 74,25% dan 80,19%. Sedangkan jika melibatkan variabel yang signifikan ketepatan klasifikasi yang tertinggi yaitu dengan menggunakan metode Neural Network dengan ketepatan klasifikasinya sebesar 76,23% dan 80,19% untuk masing-masing kategori IPK dan lama tempuh studi. Pada kasus ini, Neural Network merupakan metode yang lebih baik untuk mengklasifikasikan data prestasi mahasiswa pasca sarjana ITS dibanding regresi logistik.

Berdasarkan hasil analisis dalam penelitian ini, telah ditunjukkan bahwa ketepatan klasifikasi Neural Network hanya satu kali memiliki ketepatan klasifikasi yang tinggi dari regresi logistik di variabel yang berbeda sehingga untuk penelitian selanjutnya dapat menggunakan arsitektur dan fungsi aktifasi yang berbeda.

DAFTAR PUSTAKA

- [1] S. Kusumadewi, *Membangun Jaringan Syaraf Tiruan Menggunakan MATLAB dan Excel Link*. Yogyakarta: Graha Ilmu (2004).
- [2] S. Manel, J.M. Dias and S.J. Ormerod, Comparing discriminant analysis, *Neural Networks* and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling* 120 (1999) 337-347.
- [3] D. Michie, D.J. Spiegelhalter, and C.C.Taylor, *Machine Learning, Neural and Statistical Classification* (1994).
- [4] M. Brahmantyo, *Analisis Faktor-Faktor yang Mempengaruhi Prestasi Akademik Mahasiswa Pasca Sarjana ITS*. Tugas Akhir Jurusan Statistika Institut Teknologi Sepuluh Nopember (2010).