

**PERBANDINGAN MODEL NEURAL NETWORK DAN REGRESI LOGISTIK
PADA KASUS MASA STUDI
MAHASISWA JURUSAN PENDIDIKAN MATEMATIKA FMIPA UNY**

Dhoriva Urwatul Wutsqa¹ dan Sri Rezeki²

¹Jurusan Pendidikan Matematika, FMIPA Universitas Negeri Yogyakarta

²Program Studi Pendidikan Matematika, FKIP Universitas Islam Riau

Abstrak

Penelitian ini bertujuan untuk membandingkan model *Neural Network (NN)* dengan regresi logistik dalam memodelkan masalah klasifikasi pada kasus masa studi mahasiswa Jurusan Pendidikan Matematika FMIPA UNY, dan mengetahui faktor-faktor yang mempengaruhi masa studi mahasiswa Jurusan Pendidikan Matematika FMIPA UNY. Subyek penelitian ini adalah mahasiswa Jurusan Pendidikan Matematika yang lulus sejak Agustus 2001 hingga November 2005. Jumlah responden dalam penelitian ini sebanyak 319 orang. Variabel respon (Y) dalam penelitian ini adalah ketepatan masa studi. Variabel penjelasnya yaitu jenis kelamin (X_1), IPK tahun pertama (X_2) dan program studi (X_3). Variabel penjelas dipilih berdasarkan ketersediaan data. Untuk membandingkan kedua model digunakan kriteria ketepatan klasifikasi, yaitu rasio antara banyaknya data yang terprediksi secara tepat, dengan keseluruhan data. Faktor-faktor yang mempengaruhi masa studi mahasiswa Jurusan Pendidikan Matematika FMIPA UNY ditentukan berdasarkan uji signifikansi terhadap parameter pada model logistik. Hasil perbandingan kedua model tersebut untuk kasus masa studi mahasiswa di UNY menunjukkan bahwa model regresi logistik lebih baik digunakan karena mampu memberikan tingkat ketepatan klasifikasi yang lebih tinggi dibandingkan *NN*. Diantara faktor-faktor jenis kelamin, IPK pada tahun pertama, dan program studi, IPK pada tahun pertama merupakan faktor yang secara signifikan mempengaruhi masa studi mahasiswa.

Kata kunci : Model *NN*, model regresi logistik, ketepatan klasifikasi, masa studi.

PENDAHULUAN

Kinerja suatu perguruan tinggi diukur dari beberapa aspek antara lain rata-rata IPK mahasiswa, masa studi, dan masa tunggu alumni untuk mendapatkan pekerjaan. Masa studi mahasiswa pada jenjang strata-1 (S1) di Jurusan Pendidikan Matematika adalah delapan semester. Mahasiswa dapat memperpanjang masa studi maksimal 14 semester. Kenyataannya sangat jarang mahasiswa yang dapat menyelesaikan studinya tepat delapan semester. Sebagaimana hasil selama setahun terakhir (2003-2004), rata-rata masa studi mahasiswa Jurusan Pendidikan Matematika adalah 5,17 tahun. Mengingat

masa studi merupakan salah satu ukuran kinerja jurusan atau program studi di suatu perguruan tinggi, maka klasifikasi ketepatan masa studi berdasarkan variabel-variabel lain yang relevan menjadi hal yang penting.

Ada beberapa faktor yang mempengaruhi lamanya waktu yang dibutuhkan oleh mahasiswa dalam menyelesaikan studi, diantaranya kemampuan awal, minat, motivasi, gaya belajar, jenis kelamin dan program studi yang dipilih. Sebagaimana diungkapkan oleh Munthe dalam Adam (1996) bahwa jenis kelamin merupakan salah satu faktor nonintelektual yang mempengaruhi keberhasilan mahasiswa dalam proses pendidikan. Hasil penelitian di IPB oleh Setyowati (1998) menunjukkan bahwa jenis kelamin mempengaruhi keberhasilan mahasiswa, dengan ditemukannya adanya kecenderungan mahasiswa perempuan tidak berhasil dalam studinya. Di samping itu, hasil penelitiannya juga mengungkapkan bahwa NEM atau nilai UAN dan program studi yang dipilih mempengaruhi keberhasilan mahasiswa dalam studinya.

Dari basis data mahasiswa di UNY dapat dilihat data masa studi mahasiswa dan juga data-data lain seperti program studi, jenis kelamin, nilai IPK per semester, tingkat sosial ekonomi, dan lain-lain. Mengacu pada hasil-hasil penelitian di atas, maka variabel yang relevan terhadap masa studi mahasiswa dari faktor non kognitif antara lain jenis kelamin dan program studi. Dari faktor kognitif, IPK pada tahun pertama dapat digunakan sebagai ganti nilai UAN, mengingat keduanya dapat dipandang sebagai kemampuan kognitif awal mahasiswa. Oleh karena itu nilai IPK tahun pertama jenis kelamin, dan program studi dapat dipilih sebagai variabel yang relevan untuk klasifikasi masa studi mahasiswa. Dari hasil pemodelan akan dapat diprediksi ketepatan masa studi seorang mahasiswa, berdasarkan IPK tahun pertama, jenis kelamin, dan program studi.

Sesuai dengan kurikulum 2002 mahasiswa Jurusan Pendidikan Matematika FMIPA UNY dikatakan lulus tepat waktu jika masa studinya 8 semester. Tetapi target tersebut masih jauh dari harapan, karena rata-rata masa studi yang dicapai selama lima tahun terakhir masih di atas 5 tahun, sehingga target diturunkan menjadi 5 tahun. Apabila masa studi tersebut dikategorikan menjadi dua kemungkinan yaitu lulus tepat waktu dengan masa studi kurang dari atau sama dengan 10 semester, dan lulus tidak tepat waktu dengan masa studi lebih dari 10 semester, maka akan diperoleh data biner. Model yang biasa digunakan untuk menganalisis data respon biner adalah model regresi logistik.

Model ini dapat menjelaskan hubungan antara variabel respon yang bersifat kategorik, dengan satu atau lebih variabel prediktor yang bersifat kontinu atau kategorik (Agresti, 1990). Selain itu, model regresi logistik dapat diterapkan untuk menyelesaikan masalah klasifikasi atau pemisahan objek. Dalam kasus ini, untuk mengelompokkan mahasiswa berdasarkan ketepatan masa studi dengan karakteristik tertentu.

Suatu pendekatan yang lebih fleksibel yaitu model neural network (*NN*) dapat digunakan untuk menyelesaikan berbagai masalah terapan, seperti *pattern recognition*, klasifikasi dan optimisasi (Stern, 1996). Menurut West dkk. (1997) *NN* mampu memberikan akurasi prediksi yang lebih baik dibandingkan regresi logistik. Sebagai model nonparametrik, *NN* tidak membutuhkan asumsi tentang bentuk hubungan fungsional antar variabel, sehingga tidak mengalami bias spesifikasi model. Beberapa penelitian yang menunjukkan bahwa model *NN* lebih unggul dibandingkan dengan model regresi logistik dapat dilihat pada Schumacher dkk. (1996), Zhou (1997), West (2000), McMillen (2000), dan Chiang dkk. (2004).

Berdasarkan uraian di atas, tulisan ini bertujuan untuk mengkaji hasil penerapan metode yang relatif baru, yaitu model *NN*, pada kasus masa studi mahasiswa Jurusan Pendidikan Matematika FMIPA UNY, kemudian membandingkan hasilnya dengan model regresi logistik sebagai model standar. Selain itu akan dikaji faktor-faktor yang secara signifikan mempengaruhi masa studi mahasiswa.

MODEL REGRESI LOGISTIK

Model regresi logistik digunakan untuk mencari hubungan antara variabel respon yang bersifat kategorik dengan variabel prediktor yang bersifat kontinu atau kategorik. Nilai dari variabel respon Y yang bersifat biner atau dikotomis dibedakan atas dua kategori, misalnya $Y = 0$ dan $Y = 1$. Misalkan terdapat p variabel $X' = (X_1, X_2, \dots, X_p)$ yang berpasangan dengan variabel respon Y . Peluang $Y = 1$ dinotasikan dengan $\pi(x)$. Fungsi regresi logistik $\pi(x)$ adalah:

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

dengan $g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ (Hosmer dan Lemeshow, 1989).

Fungsi regresi di atas berbentuk *curvilinear* sehingga untuk membuatnya menjadi fungsi linear dilakukan transformasi logit sebagai berikut (Agresti, 1990):

$$\text{Logit}[\pi(x)] = \log \left[\frac{\pi(x)}{1 - \pi(x)} \right] = g(x) \quad (2)$$

Pada model regresi linear diasumsikan bahwa suatu amatan dari variabel respon dinyatakan sebagai $y = E(Y|x) + \varepsilon$, dengan

$$E(Y|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

merupakan mean populasi dan ε adalah *error* yang merupakan komponen acak yang menunjukkan penyimpangan amatan dari rataannya. *Error* ε diasumsikan mengikuti distribusi normal dengan mean sama dengan nol dan variansi konstan. Pada model regresi logistik variabel respon dinyatakan sebagai $y = \pi(x) + \varepsilon$. Nilai ε mempunyai salah satu dari dua kemungkinan yaitu:

$$\varepsilon = \begin{cases} 1 - \pi(x) & \text{jika } y = 1 \\ -\pi(x) & \text{jika } y = 0 \end{cases} \quad (3)$$

sehingga distribusi *error* model regresi logistik akan mempunyai rata-rata sama dengan nol dan variansi $\{\pi(x) \cdot (1 - \pi(x))\}$ (Hosmer dan Lemeshow, 1989).

Prosedur penaksiran parameter dilakukan dengan metode *maximum likelihood*, yang secara lengkap dapat dilihat pada Hosmer dan Lemeshow (1989) dan Agresti (1990). Uji signifikansi parameter dilakukan dua tahap, yaitu uji serentak dilanjutkan dengan uji parsial. Uji serentak digunakan untuk mengetahui peran seluruh variabel prediktor dalam model secara serentak, dengan hipotesis:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{minimal ada satu } \beta_j \text{ yang tidak sama dengan nol, } j = 1, 2, \dots, p$$

Statistik uji yang digunakan adalah uji G (*Likelihood Ratio Test*):

$$G = 2 \left\{ \sum [y_i \ln(\hat{\pi}_i) + (1 - y_i) \ln(1 - \hat{\pi}_i)] - [n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n)] \right\}, \quad (4)$$

dengan:

n_0 = banyak observasi yang bernilai $Y = 1$

n_1 = banyak observasi yang bernilai $Y = 0$

$n = n_0 + n_1$

Hipotesis H_0 ditolak jika $P(G > \chi_{\alpha, \nu}^2)$ atau p-value $< \alpha$. Statistik uji G mengikuti distribusi *chi-square* dengan derajat bebas ν (banyaknya parameter dalam model).

Jika uji serentak menunjukkan adanya parameter yang signifikan maka dilanjutkan dengan uji parsial, yang merupakan pengujian β_i secara individual. Hasil pengujian akan menunjukkan apakah suatu variabel prediktor layak untuk masuk dalam model atau tidak. Hipotesis yang digunakan:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0, \quad j = 1, 2, \dots, p.$$

Statistik uji yang digunakan adalah uji Wald:

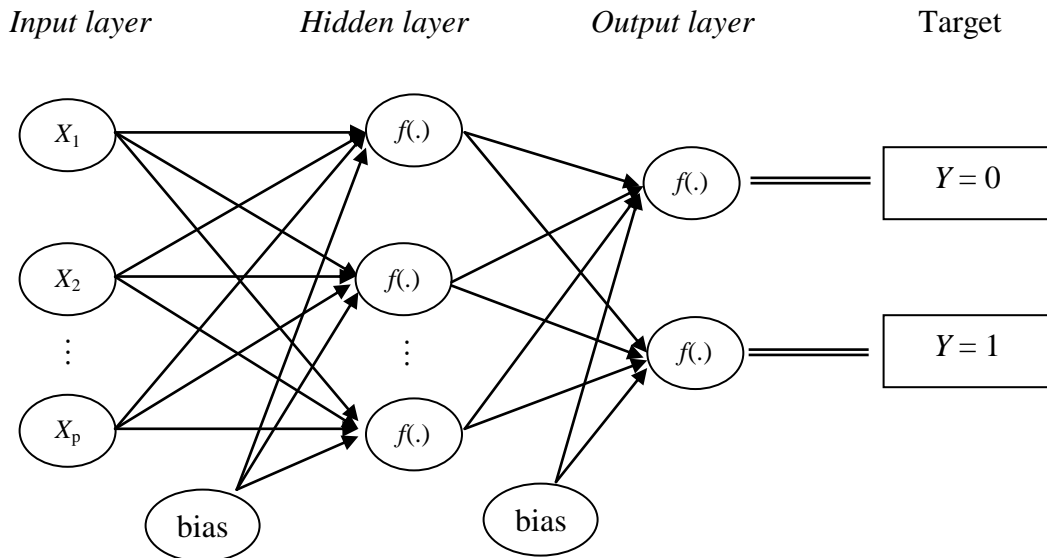
$$W = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \tag{5}$$

dengan $\hat{\beta}_j$ adalah penduga *maximum likelihood* dan $SE(\hat{\beta}_j)$ adalah *standard error* untuk $\hat{\beta}_j$. Hipotesis H_0 ditolak jika $P(|W| > Z)$ atau p-value $< \alpha$ (nilai yang ditetapkan peneliti). Statistik uji W merupakan variabel random yang mengikuti distribusi normal standar.

MODEL NEURAL NETWORKS (NN) UNTUK KLASIFIKASI DATA

Neural Networks dapat dipandang sebagai model regresi nonlinear dimana kompleksitas modelnya dapat diubah-ubah. Pada level kompleksitas yang paling rendah, *NN* hanya terdiri dari satu lapisan *input* dan satu lapisan *output*. *NN* memungkinkan untuk mengubah kompleksitas jaringan sehingga dapat mengakomodasi efek nonlinear, khususnya efek interaksi diantara variabel independen. Hornik dkk. (1987) telah menunjukkan bahwa *NN* dengan satu *hidden layer* mampu menghampiri sembarang fungsi pada himpunan kompak tanpa asumsi awal tentang fungsi yang dimodelkan.

Tipe *NN* yang digunakan dalam tulisan ini adalah *Feed Forward Neural Networks* (*FFNN*). Arsitektur *FFNN* untuk masalah klasifikasi dengan satu *hidden layer* disajikan dalam Gambar 1.



Gambar 1. FFNN dengan Satu *Hidden Layer* dan *Output Sigmoid*

Hidden layer dalam model berfungsi untuk mengakomodasi efek nonlinear, yang dinyatakan dalam fungsi aktivasi. Fungsi aktivasi pada *hidden layer* dan *output layer* yang digunakan dalam makalah ini adalah *Logistic sigmoid* (Logsig). Bentuk fungsi logsig untuk *hidden neuron* pada neuron ke i adalah:

$$f_i(\cdot) = \frac{1}{1 + \exp\left(-\left(\text{bias} + \sum_{j=1}^p w_{ij}x_j\right)\right)}, \quad i = 1, \dots, h,$$

dengan h adalah jumlah *hidden neuron*, w_{ij} adalah *weight* dari variabel ke j menuju *hidden neuron* ke i , $f_i(\cdot)$ adalah fungsi aktivasi pada *hidden neuron* ke i . Fungsi logsig menghasilkan nilai *output* terletak antara 0 dan 1. Arsitektur jaringan sebagaimana diberikan pada Gambar 1. dengan fungsi aktivasi logsig akan membentuk jaringan output sigmoid yang dapat dipandang sebagai generalisasi model logit binomial.

Parameter (*weight*) FFNN diestimasi dengan meminimalkan *mean square error* (*MSE*) yang dihasilkan oleh model. Metode yang paling umum digunakan untuk proses estimasi parameter atau proses training ini adalah *Back Propagation* (BP). Dalam metode ini, *weight* dan bias diatur untuk meminimalisasi nilai kuadrat beda antara *output* model

dan *output* taksiran yang disebut sebagai SSE (*Sum of Square Error*). Proses pengaturan biasanya berdasarkan pada algoritma sederhana *steepest descent* yang didasarkan pada turunan pertama fungsi *cost*

$$E = \frac{1}{2} \sum_{k=1}^Q [t_k - \underline{a}_k^\wedge]^T [t_k - \underline{a}_k^\wedge], \quad (6)$$

dengan Q banyak pasangan input (\underline{p}_k) output (t_k), dengan pasangan ke k dinotasikan sebagai (\underline{p}_k, t_k) , $k = 1, \dots, Q$, \underline{a}^\wedge adalah nilai *output* dari *NN* pada *layer* akhir. Penjelasan secara rinci tentang algoritma BP dapat dilihat pada Rumelhart dkk. (1986).

METODE PENELITIAN

Penelitian ini dilakukan terhadap mahasiswa Jurusan Pendidikan Matematika FMIPA UNY yang lulus sejak Agustus 2001 hingga November 2005. Variabel respon dalam penelitian ini adalah masa studi mahasiswa ($Y=1$ jika masa studi ≤ 10 semester, $Y=0$ jika masa studi > 10 semester). Sedangkan variabel prediktor adalah jenis kelamin ($X_1=1$ jika laki-laki, $X_1=2$ jika perempuan), IPK tahun pertama (X_2), program studi ($X_3=1$ jika program studi Pendidikan Matematika, $X_3=2$ jika program studi Matematika). Untuk analisis deskriptif, data IPK dikategorikan menjadi IPK rendah ($IPK < 2,50$), IPK sedang ($2,5 \leq IPK \leq 2,99$) dan IPK tinggi ($IPK \geq 3,00$). Variabel penjelas dipilih berdasarkan ketersediaan data.

Untuk membandingkan model *NN* dengan regresi logistik dilakukan resampling sebanyak 50 kali. Sampling dilakukan secara acak dengan mengambil 50% dari keseluruhan data sebagai data training dan sisanya sebagai data testing. Data training digunakan untuk pembentukan model sedangkan data testing digunakan untuk validasi model. Proses analisis data dilakukan dengan software S-Plus 2000 untuk membandingkan kinerja model *NN* dan regresi logistik, dan software MINITAB versi 13 untuk estimasi serta interpretasi model regresi logistik.

Perbandingan model *NN* dan regresi logistik dilakukan dengan menggunakan kriteria ketepatan klasifikasi yang disebut dengan persentase *correct*, yaitu rasio antara banyaknya data yang terprediksi secara tepat, dengan keseluruhan data. Model terbaik

adalah model yang memiliki rata-rata ketepatan klasifikasi tertinggi pada data testing. Faktor-faktor yang mempengaruhi masa studi mahasiswa ditentukan dari model logistik.

HASIL DAN PEMBAHASAN

Deskripsi Responden

Responden dalam penelitian ini berjumlah 319 orang. Dari jumlah tersebut ada 186 (58,31%) mahasiswa yang mampu menyelesaikan studinya tepat waktu, sedangkan sisanya ada 133 (41,69%) yang dinyatakan tidak tepat waktu. Deskripsi responden berdasarkan variabel-variabel yang diteliti disajikan dalam Tabel 1. berikut.

Tabel 1. Persentase Ketepatan Masa Studi Mahasiswa

Variabel		Ketepatan Masa Studi		Jumlah
		Tepat	Tidak	
Jenis Kelamin	Laki-laki	51 (57,30%)	38 (42,70%)	89
	Perempuan	135 (58,70%)	95 (41,30%)	230
Total				319
IPK	Rendah	2 (22,22%)	7 (77,78%)	9
	Sedang	38 (35,85%)	68 (64,15%)	106
	Tinggi	146 (71,57%)	58 (28,43%)	204
Total				319
Program Studi	Pendidikan Matematika	98 (54,14%)	83 (45,86%)	181
	Matematika	88 (63,78%)	50 (36,23%)	138
Total				319

Berdasarkan Tabel 1., jika ditinjau dari jenis kelamin, maka persentase mahasiswa perempuan yang mampu menyelesaikan studi tepat waktu lebih tinggi dibandingkan dengan laki-laki. Namun perbedaan tersebut tidak terlalu besar. Perbedaan persentase yang cukup mencolok terjadi diantara mahasiswa yang memiliki IPK rendah, sedang dan tinggi. Dengan meningkatnya IPK, terlihat kecenderungan persentase yang lulus tepat waktu juga meningkat. Berdasarkan program studi yang diambil, mahasiswa program studi Matematika lebih banyak yang mampu menyelesaikan studi tepat waktu dibandingkan dengan mahasiswa program studi Pendidikan Matematika. Untuk mengetahui variabel mana yang berbeda secara signifikansi dapat dilihat pada hasil

inferensi dengan menggunakan uji Wald (5) yang akan diberikan pada pembahasan berikut.

Perbandingan Model NN dan Model Regresi Logistik

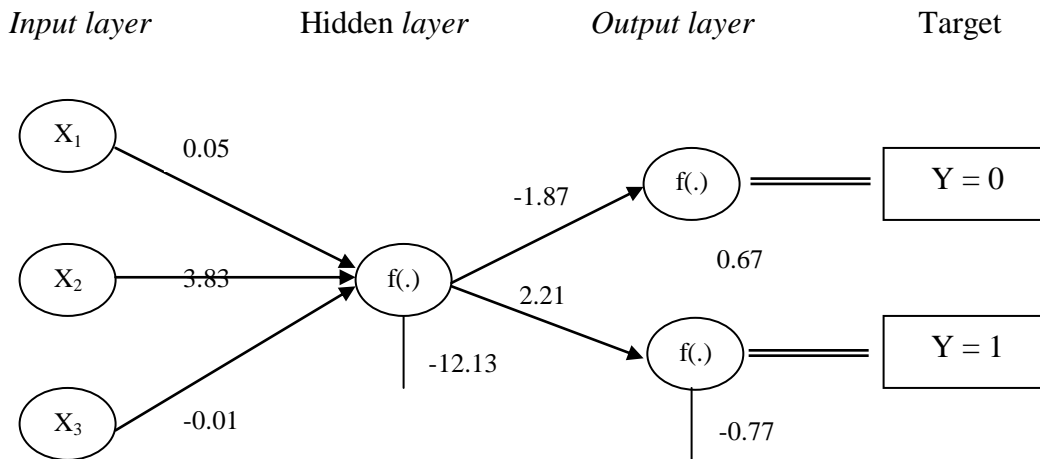
Untuk mendapatkan model NN, dalam penelitian ini digunakan arsitektur NN yang terdiri atas *input layer*, satu *hidden layer*, dan *output layer* sebagaimana diilustrasikan dalam Gambar 1. Jumlah neuron dalam *input layer* sama dengan jumlah variabel prediktor (tiga), sedangkan jumlah neuron pada *output layer* adalah dua, sesuai dengan banyak kategori pada variabel respon. Jumlah neuron pada *hidden* unit terletak pada rentang 1 sampai 5. NN (3-1-2) berarti model NN dengan 3 *input*, 1 neuron pada *hidden* unit dan 2 *output*. Hasil analisa data yang merupakan rata-rata ketepatan klasifikasi untuk semua model NN dan untuk model regresi logistik dirangkum dalam Tabel 2.

Tabel 2. Rata-rata ketepatan klasifikasi model NN dan regresi logistik

Model	Ketepatan klasifikasi	
	Training	Testing
Regresi Logistik	0,6947	0,6925
NN (3-1-2)	0,7010	0,6854
NN (3-2-2)	0,7077	0,6667
NN (3-3-2)	0,7175	0,6670
NN (3-4-2)	0,7312	0,6570
NN (3-5-2)	0,7400	0,6468

Berdasarkan hasil pada Tabel 2. pada data testing model NN memperlihatkan kecenderungan makin banyak jumlah neuroNNya, makin tinggi tingkat ketepatan klasifikasinya. Hal yang sebaliknya terjadi pada data testing, makin banyak jumlah neuron, makin rendah tingkat ketepatan klasifikasinya. Kejadian seperti ini sering disebut dengan *overfitting*. Model NN terbaik biasanya ditentukan berdasarkan pada prinsip *parsimony* dan hasil terbaik pada data testing

Diantara 10 model *NN* yang digunakan dalam penelitian ini, *NN* dengan jumlah neuron terkecil memiliki tingkat ketepatan klasifikasi yang tertinggi pada data testing. Oleh karena itu model *NN* dengan satu neuron (*NN* (3-1-2)) direkomendasikan sebagai model *NN* terbaik. Arsitektur *NN* (3-1-2) dan taksiran bobot-bobotnya (*weight*/parameter) dapat dilihat pada Gambar 5.



Gambar 5. Arsitektur *NN* (3-1-2)

Model matematis dari arsitektur tersebut adalah:

$$Y_{(0)} = \frac{1}{1 + \exp(-(0.67 - 1.87h_1))}$$

atau

$$Y_{(1)} = \frac{1}{1 + \exp(-(-0.77 + 2.217h_1))}$$

dengan

$$h_1 = \frac{1}{1 + \exp(-(-12.13 + 0.05X_1 + 3.83X_2 - 0.01X_3))}$$

Bobot-bobot pada model *NN* tidak dapat diinterpretasi dan diuji signifikansinya. Jadi model tersebut hanya dapat digunakan untuk keperluan prediksi. Contohnya: jika ada seorang mahasiswa perempuan, IPK tahun pertamanya adalah 3, dan memilih program studi Pendidikan Matematika, maka diperoleh nilai $h_1 = 0,34$, $Y_{(0)} = 0,49$ dan $Y_{(1)} = 0,51$.

Karena nilai probabilitas $Y_{(1)}$ lebih dari 0,5 maka mahasiswa tersebut diklasifikasikan kedalam kelompok mahasiswa yang akan lulus tepat waktu.

Hasil perbandingan model *NN* dan regresi logistik dapat dilihat pada rata-rata ketepatan klasifikasi antara model *NN* terbaik, yaitu *NN* (3-1-2) dengan regresi logistik. Tabel 2. menunjukkan bahwa model *NN* lebih unggul dibandingkan dengan model regresi logistik pada data training. Hasil ini sesuai dengan teori tentang *NN* yang dikemukakan oleh Hornik dkk. (1987) bahwa *NN* dengan satu *hidden layer* mampu menghampiri sembarang fungsi pada himpunan kompak tanpa asumsi awal tentang fungsi yang dimodelkan. Dari sini dapat disimpulkan bahwa pada proses estimasi, *NN* mampu memberikan hasil yang akurat.

Pada data testing, model regresi logistik lebih unggul dibandingkan dengan *NN*. Hasil ini tidak sejalan dengan beberapa hasil penelitian yang telah disebutkan dalam bab sebelumnya. Hal yang mungkin menjadi penyebabnya adalah kompleksitas data atau pola hubungan data. Untuk data-data yang tidak begitu kompleks ataupun data yang mempunyai pola hubungan linear, model parametrik seperti model logistik merupakan model yang lebih tepat. Model *NN* biasanya menjadi pilihan ketika data yang dihadapi mempunyai pola hubungan nonlinear atau data dengan bentuk hubungan yang kompleks.

Jadi, dalam kasus masa studi mahasiswa Jurusan Pendidikan Matematika FMIPA UNY, analisis data menggunakan model regresi logistik lebih sesuai dibandingkan dengan model *NN*. Hasil temuan dalam penelitian ini dapat memberikan gambaran perlunya kehati-hatian dalam pemilihan model. Hal ini mengingat semakin berkembangnya model-model yang tersedia untuk memecahkan permasalahan-permasalahan yang ada, sehingga perlu dipelajari terlebih dahulu kecocokan suatu model dengan data yang tersedia.

Faktor-Faktor yang Mempengaruhi Masa Studi

Analisis data menggunakan regresi logistik, menghasilkan nilai-nilai parameter model sebagai berikut:

Tabel 3. Pendugaan Parameter

Variabel	$\hat{\beta}$	S.E	z	Sigifikansi	Exp($\hat{\beta}$)
Jenis Kelamin	0,039	0,42	0,09	0,926	1,040
IPK	3,388	0,722	4,69	0,000	29,618
Jurusan	0,617	0,388	1,59	0,112	1,853
Konstanta	-10,27	2,194	-4,68	0,000	0

Uji serentak (4) menghasilkan nilai $G = 34,354$, dan $P\text{-Value} = 0,000$. Hal ini berarti minimal ada satu koefisien regresi yang tidak sama dengan nol. Nilai-nilai untuk uji parsial dapat dilihat pada Tabel 3 kolom z. Berdasarkan tabel tersebut, variabel yang secara signifikan berpengaruh terhadap ketepatan masa studi mahasiswa adalah variabel IPK. Pada variabel jenis kelamin, nilai eksponen ($\hat{\beta}$) = 1,040 berarti kesempatan mahasiswa perempuan tamat tepat waktu adalah 1,040 kali kesempatan yang laki-laki. Untuk variabel IPK, nilai eksponen ($\hat{\beta}$) = 29,618 berarti bahwa jika IPK meningkat 1 satuan maka kesempatan tamat tepat waktu meningkat sebesar 21,026 kali kesempatan untuk IPK sebelumnya. Interpretasi untuk variabel program studi yaitu mahasiswa program studi Matematika memiliki kesempatan tamat tepat waktu 1,853 kali kesempatan yang dimiliki oleh mahasiswa program studi Pendidikan Matematika. Model regresi logistik yang diperoleh adalah:

$$\text{Logit} [\pi(x)] = \log \left(\frac{\pi(x)}{1 - \pi(x)} \right) = -10.27 + 0.039X_1 + 3.388X_2 + 0.617X_3$$

Sebagai ilustrasi penggunaan model tersebut untuk prediksi, misalkan ada seorang mahasiswa perempuan, IPK tahun pertamanya adalah 3, dan memilih program studi Pendidikan Matematika, maka probabilitas mahasiswa tersebut lulus tepat waktu yaitu $P(Y=1) = \pi(x) = 0,64$. Nilai probabilitas ini lebih besar dari nilai yang dihasilkan pada model NV (3-1-2), namun dalam hal ini hasil klasifikasi kedua model tersebut adalah sama, yaitu mahasiswa tersebut diklasifikasikan kedalam kelompok mahasiswa yang akan lulus tepat waktu.

PENUTUP

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa perbandingan kedua model tersebut untuk kasus masa studi mahasiswa di UNY menunjukkan bahwa model regresi logistik lebih baik digunakan karena mampu memberikan tingkat ketepatan klasifikasi yang lebih tinggi dibandingkan *NN*. Hasil ini tidak sesuai dengan beberapa hasil penelitian sebelumnya. Oleh karena itu, disarankan untuk penelitian lebih lanjut mengkaji pola data yang sesuai dengan model *NN*, sehingga dapat menjadi pedoman bagi para peminat *NN* untuk dapat menerapkan model *NN* secara tepat.

Berdasarkan hasil pada model regresi logistik ditunjukkan bahwa diantara ketiga faktor, *IPK* pada tahun pertama merupakan faktor yang secara signifikan mempengaruhi masa studi mahasiswa.

DAFTAR PUSTAKA

- Adam, B.M. 1996. *Penerapan Analisis Diskriminan Non parametric untuk Menduga Keberhasilan Mahasiswa TPB IPB*. Skripsi. Bogor: Jurusan Statistika FMIPA IPB.
- Agresti, A. 1990. *Categorical Data Analysis*. New York: John Wiley and Sons, Inc.
- Chiang, W.K., Zhang, D. & Zhou, L. 2004. Predicting and Explaining Patronage Behavior toward Web and Traditional Stores Using Neural Networks: a Comparative Analysis with Logistic Regression, Decision Support System, xx.
- Fausett, L. 1994. *Fundamental of Neural Network: Architecture, Algorithms and Applications*. New Jersey: Prentice Halls International, Inc.
- Handayani, D. 1996. *Profil Mahasiswa TPB 1995/1996 yang Kurang Berhasil di IPB*. Skripsi. Bogor: Jurusan Statistika FMIPA IPB.
- Hornik, K., Stinchcombe, M. & White, H. 1989. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, 359-66.
- Hosmer, D.W. & Lemeshow, S. 1989. *Applied Logistic Regression*. New York: John Wiley & Sons Ltd.
- Lolombulan, J.H. 1990. *NEM SMA Sebagai Indikator Keberhasilan Belajar di Perguruan Tinggi*. Disertasi. Bogor: Program Pascasarjana IPB.
- McMillen, R. 2000. *Neural Networks as a Methodological Tools*. draft do not cite: Mississippi State University: Tracy Henly.
- Rumelhart, D. & McClelland, J. 1986. *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*. Vol. 1, Cambridge: MIT Press.

- Schumacher, M., Roßner, R. & Vach, W. 1996. Neural Network and Logistic Regression: Part I. *Computational Statistic & Data Analysis*, 21: 661-682.
- Setyowati, R. 1998. *Deskripsi Faktor Penyebab Kegagalan Mahasiswa Wanita TPB IPB 1996/1997*. Skripsi. Bogor: Jurusan Statistika FMIPA IPB.
- Stern, H.S. 1996. Neural Networks in Applied Statistics, *Technometrics*, Vol. 38, No. 3, 205-214.
- West, D. 2000. Neural Network Credit Scoring Model”, *Computer & Operations Research*, 27:131 – 1152.
- West, P.M., Brockett, P.L. & Golden, L.L. 1997. A comparative analysis of neural networks and statistical methods for predicting consumer choice. *Marketing Science*, 16, 370–391.
- Zhou, L., Ersheng, G. & Pihua, J. 1997. Comparison between the Logistic regression and Back Propagation Neural Network, Shanghai: Department of Health Statistic, Shanghai Medical University.