

ANALISIS LAMA STUDI MAHASISWA FSM UNIVERSITAS DIPONEGORO MENGUNAKAN REGRESI LOGISTIK BINER

Diah Budiati¹, Yuciana Wilandari²

¹Mahasiswa Jurusan Statistika FSM Undip

²Staf Pengajar Jurusan Statistika FSM Undip

Abstrak

Kelulusan atau wisuda adalah hasil akhir dari proses kegiatan belajar mengajar selama mengikuti perkuliahan di perguruan tinggi. Dalam pencapaian gelar sarjana S1 membutuhkan waktu normal dalam menempuh perkuliahan yaitu selama empat tahun bahkan ada yang kurang dari empat tahun, tetapi dalam praktiknya masih banyak mahasiswa yang tidak dapat menuntaskan studinya selama waktu normal tersebut. Banyak faktor yang menyebabkan ketidaktepatan waktu kelulusan mahasiswa. Dalam penelitian ini faktor-faktor yang diduga mempengaruhi lama studi mahasiswa FSM Universitas Diponegoro adalah Indeks Prestasi Kumulatif (IPK), jurusan dan jalur masuk universitas. Untuk mengetahui faktor-faktor apa saja yang mempengaruhi lama studi mahasiswa dapat menggunakan analisis regresi logistik biner. Estimasi parameter model menggunakan fungsi maksimum likelihood, yang dilanjutkan dengan iterasi Newton Raphson. Untuk menguji signifikansi dari parameter digunakan uji rasio likelihood dan uji wald. Setelah dilakukan pengujian secara keseluruhan maupun secara individu variabel IPK dan jurusan signifikan mempengaruhi lama studi dengan ketepatan klasifikasi 80,8.

Kata Kunci: Lama Studi, Regresi Logistik Biner

1. Pendahuluan

Kelulusan atau wisuda adalah hasil akhir dari proses kegiatan belajar mengajar selama mengikuti perkuliahan di perguruan tinggi. Perguruan tinggi merupakan salah satu pendidikan formal yang disertai tugas dan tanggung jawab untuk mempersiapkan mahasiswa sesuai dengan sistem pendidikan nasional. Selain itu perguruan tinggi dituntut pula untuk menghasilkan lulusan yang berkualitas, yang mana perguruan tinggi akan mempersiapkan calon-calon sarjana yang handal dan mempunyai keahlian khusus dibidangnya. Dalam pencapaian gelar sarjana (S1) membutuhkan waktu normal dalam menempuh perkuliahan yaitu selama empat tahun bahkan ada yang kurang dari empat tahun, tetapi dalam praktiknya masih banyak mahasiswa yang tidak selalu dapat menuntaskan studinya selama waktu normal tersebut. Banyak faktor yang menyebabkan ketidaktepatan waktu kelulusan mahasiswa, faktor-faktor tersebut dapat bersumber dari faktor eksternal maupun internal.

Fakultas Sains dan Matematika Universitas Diponegoro terdiri dari enam jurusan (S1) yaitu Statistika, Matematika, Informatika, Biologi, Fisika, dan Ilmu Komputer. Dari enam jurusan tersebut pada tiap periode wisuda masih banyak mahasiswa Fakultas Sains dan Matematika Universitas Diponegoro yang waktu kelulusannya tidak tepat waktu (lebih dari empat tahun), oleh sebab itu penulis tertarik untuk menganalisis faktor-faktor apa saja yang mempengaruhi lama studi dalam kasus kelulusan mahasiswa Fakultas Sains dan Matematika Universitas Diponegoro pada periode wisuda tahun akademik 2012/2013. Pada penelitian ini faktor-faktor yang diduga mempengaruhi lama studi mahasiswa FSM Universitas Diponegoro adalah Indeks Prestasi Kumulatif (IPK), jurusan dan jalur masuk universitas. Karena lama studi disini dikategorikan menjadi 2 yaitu kurang atau sama dengan 4 tahun dan lebih dari empat tahun, maka analisis yang digunakan adalah model regresi logistik biner. Sehingga tujuan dari penelitian ini adalah untuk mengetahui faktor-faktor apa saja yang mempengaruhi lama studi dalam kasus kelulusan mahasiswa Fakultas Sains dan Matematika Universitas Diponegoro pada periode wisuda tahun akademik 2012/2013 dengan menggunakan regresi logistik biner.

2. Tinjauan Pustaka

2.1. Lama Tempuh Studi

Pada Peraturan Akademik Universitas Diponegoro Bab V Pasal 11 tentang Beban, masa studi dan penentuan mata kuliah, Pendidikan Program Sarjana mempunyai beban studi sekurang-kurangnya 144 sks dan sebanyak-banyaknya 160 sks yang dijadwalkan untuk 8 semester dan dapat ditempuh dalam waktu kurang dari 8 semester dan selama-lamanya 14 semester (Hadi, 2012).

2.2. Model Regresi Logistik Biner

Model regresi logistik biner digunakan untuk menganalisa hubungan antara satu variabel respon dan beberapa variabel prediktor, dengan variabel responnya berupa data kualitatif dikotomi yaitu bernilai 1 untuk menyatakan keberadaan sebuah karakteristik dan bernilai 0 untuk menyatakan ketidakberadaan sebuah karakteristik. Model regresi logistik biner digunakan jika variabel responnya menghasilkan dua kategori bernilai 0 dan 1, sehingga mengikuti distribusi Bernoulli (Agresti, 2002):

$$f(y_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad , y_i = 0, 1$$

dengan: π_i adalah peluang kejadian ke-1

y_i adalah peubah acak ke-i

Jika diketahui Y variabel respon bernilai 0 dan 1, maka

$$P = (Y = 1 | X = x_i) = \pi(x_i) \quad \text{dan} \quad P = (Y = 0 | X = x_i) = 1 - \pi(x_i)$$

dengan $i = 1, 2, \dots, p$

Sehingga diperoleh model regresi logistik:

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})}$$

dan logit dari $\pi(x_i)$ adalah :

$$\ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}$$

Untuk menentukan estimasi parameter digunakan metode iterasi Newton Raphson yang membutuhkan turunan pertama dan turunan kedua dari fungsi likelihood. y_i berdistribusi binomial, maka fungsi kepadatan peluangnya

$$p(y_i = 1) = \binom{n}{1} \{\pi(x_i)\}^{y_i} \{1 - \pi(x_i)\}^{n-y_i}$$

untuk $n=1$ maka $p(y_i = 1) = \{\pi(x_i)\}^{y_i} \{1 - \pi(x_i)\}^{1-y_i}$

Karena observasi-observasi saling bebas maka fungsi likelihoodnya adalah

$$l(\beta) = \prod_{i=1}^n \{\pi(x_i)\}^{y_i} \{1 - \pi(x_i)\}^{1-y_i}$$

maka:

$$\begin{aligned} L(\beta) &= \ln \left\{ \prod_{i=1}^n \{\pi(x_i)\}^{y_i} \{1 - \pi(x_i)\}^{1-y_i} \right\} \\ &= \sum_{i=1}^n [y_i \ln \{\pi(x_i)\} + (1 - y_i) \ln \{1 - \pi(x_i)\}] \\ &= \sum_{i=1}^n [y_i g(x_i) - \ln(1 + e^{g(x_i)})] \end{aligned}$$

dengan $g(x_i) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}$

Sehingga diperoleh turunan pertamanya :

$$\frac{\partial L(\beta)}{\partial \beta_0} = \sum_{i=1}^n \left[y_i - \frac{e^{g(x_i)}}{1 + e^{g(x_i)}} \right] = \sum_{i=1}^n [y_i - \pi(x_i)]$$

$$\frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^n \left[y_i - \frac{x_{1i} e^{g(x_i)}}{1 + e^{g(x_i)}} \right] = \sum_{i=1}^n x_{1i} [y_i - \pi(x_i)]$$

⋮

$$\frac{\partial L(\beta)}{\partial \beta_p} = \sum_{i=1}^n \left[y_i - \frac{x_{pi} e^{g(x_i)}}{1 + e^{g(x_i)}} \right] = \sum_{i=1}^n x_{pi} [y_i - \pi(x_i)]$$

Jika dinyatakan dalam bentuk matriks adalah $\mathbf{X}'(\mathbf{Y} - \boldsymbol{\pi}(\mathbf{x}))$

$$\text{dengan } \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdot & \cdot & \cdot & x_{1p} \\ 1 & x_{21} & \cdot & \cdot & \cdot & x_{2p} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ 1 & x_{n1} & \cdot & \cdot & \cdot & x_{np} \end{bmatrix} \quad \text{dan } \mathbf{Y} - \boldsymbol{\pi}(\mathbf{x}) = \begin{bmatrix} y_1 - \pi(x_1) \\ y_2 - \pi(x_2) \\ \cdot \\ \cdot \\ \cdot \\ y_n - \pi(x_n) \end{bmatrix}$$

Sedangkan turunan keduanya adalah

$$\frac{\partial^2 L(\beta)}{(\partial \beta_0)^2} = - \sum_{i=1}^n \pi(x_i) [1 - \pi(x_i)]$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_0 \partial \beta_j} = - \sum_{i=1}^n x_{ji} \pi(x_i) [1 - \pi(x_i)]$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_u \partial \beta_j} = - \sum_{i=1}^n x_{ui} x_{ji} \pi(x_i) [1 - \pi(x_i)]$$

$$\frac{\partial^2 L(\beta)}{(\partial \beta_j)^2} = - \sum_{i=1}^n (x_{ji})^2 \pi(x_i) [1 - \pi(x_i)]$$

untuk $u, j = 1, 2, \dots, p$

Jika dinyatakan dalam bentuk matriks adalah $-\mathbf{X}'\mathbf{V}\mathbf{X}$

dengan $\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdot & \cdot & \cdot & x_{1p} \\ 1 & x_{21} & \cdot & \cdot & \cdot & x_{2p} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ 1 & x_{n1} & \cdot & \cdot & \cdot & x_{np} \end{bmatrix}$

$$\mathbf{V} = \begin{bmatrix} \pi(x_i)[1 - \pi(x_i)] & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \pi(x_i)[1 - \pi(x_i)] & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & \pi(x_i)[1 - \pi(x_i)] \end{bmatrix}$$

Selanjutnya turunan pertama dan kedua dari fungsi maksimum likelihood tersebut digunakan untuk mencari estimasi parameter dengan menggunakan metode iterasi Newton-Raphson sebagai berikut:

1. Dipilih taksiran awal untuk β , misal $\hat{\beta} = 0$
2. Dihitung $\mathbf{X}'(\mathbf{Y} - \boldsymbol{\pi}(\mathbf{x}))$ dan $\mathbf{X}'\mathbf{V}\mathbf{X}$, selanjutnya dihitung invers dari $\mathbf{X}'\mathbf{V}\mathbf{X}$
3. Pada setiap $i+1$ dihitung taksiran baru yaitu

$$\hat{\beta}_{i+1} = \hat{\beta}_i + \{\mathbf{X}'\mathbf{V}\mathbf{X}\}^{-1} \{\mathbf{X}'(\mathbf{Y} - \boldsymbol{\pi}(\mathbf{x}))\}$$
4. Iterasi berakhir jika diperoleh $\hat{\beta}_{i+1} \cong \hat{\beta}_i$

(Hosmer, 2000)

Untuk menguji signifikansi dari parameter-parameter yang telah diperoleh digunakan dua uji, yaitu uji rasio likelihood dan uji Wald.

i. Uji Rasio Likelihood

Untuk mengetahui apakah secara bersama-sama variabel prediktor mempengaruhi model dilakukan uji Rasio Likelihood, dengan hipotesa sebagai berikut:

Hipotesis

$$H_0: \beta_0 = \beta_1 = \dots = \beta_p = 0$$

$$H_1: \text{salah satu dari } \beta_j \neq 0 \text{ dengan } j = 1, 2, \dots, p$$

Statistik uji : $G = -2 \ln \left(\frac{\text{likelihood tanpa variabel prediktor}}{\text{likelihood dengan variabel prediktor}} \right)$

Kriteria Uji : tolak H_0 jika nilai $G > \chi^2_{(a,p)}$

ii. Uji Wald

Uji wald bertujuan untuk mengetahui peranan masing-masing variabel prediktor terhadap variabel respon

Hipotesis

$$H_0: \beta_j = 0, \text{ dengan } j = 1, 2, \dots, p$$

$$H_1: \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, p$$

$$\text{Statistik uji : } W = \left[\frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \right]^2$$

Kriteria Uji : tolak H_0 jika nilai $W > \chi^2_{(\alpha,1)}$

Sedangkan untuk mengetahui kecenderungan suatu individu terhadap peristiwa (kategori) tertentu digunakan rasio kecenderungan (odds rasio), yaitu suatu ukuran yang berupa angka kecenderungan yang didefinisikan sebagai rasio antara jumlah individu yang mengalami kasus atau peristiwa tertentu dengan jumlah individu yang tidak mengalami kasus atau peristiwa tersebut, baik didalam sampel maupun populasi. Untuk peluang π adalah berhasil, maka nilai odds rasio didefinisikan sebagai berikut (Agresti, 2002):

$$\Omega_i = \frac{\pi_i}{1 - \pi_i}$$

Dalam tabel kontigensi 2x2 seperti terlihat pada Tabel 1, dalam baris i diketahui bahwa peluang sukses adalah $\Omega_i = \frac{\pi_i}{1 - \pi_i}$. Rasio kemungkinan Ω_1 dan Ω_2 yang kemudian

disebut odds rasio adalah sebagai berikut:

$$\psi = \frac{\Omega_1}{\Omega_2} = \frac{\pi_1/1 - \pi_1}{\pi_2/1 - \pi_2}$$

Tabel 1. Tabel Kontigensi 2x2

baris	kolom		total
	1	2	
1	π_{11}	π_{12}	π_{1i}
2	π_{21}	π_{22}	π_{2i}
total	π_{i1}	π_{i2}	1

Untuk distribusi peluang bersama π_{ij} nilai odds dalam baris ke i adalah $\Omega_i = \pi_{1i}/\pi_{2i}$ dengan $i = 1,2$. Sehingga persamaan odds rasio adalah sebagai berikut:

$$\psi = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}$$

(Kleinbaum, 2010)

3. Metodologi Penelitian

Dalam penelitian ini, data yang diperlukan adalah:

1. Data sekunder, yaitu data yang diperoleh dari sumber-sumber yang telah ada. Dalam hal ini penulis memperoleh data lulusan mahasiswa S1 Fakultas Sains dan Matematika Universitas Diponegoro tahun 2012/2013 dari bagian akademik Fakultas Sains dan Matematika Universitas Diponegoro.
2. Data primer, yaitu data yang diperoleh dengan melakukan wawancara tidak langsung dari subyek yang diteliti, yaitu mahasiswa S1 Fakultas Sains dan Matematika UNDIP yang lulus pada periode wisuda ke 127,128,129, dan 130. Peneliti melakukan wawancara tidak langsung untuk mendapatkan data tentang jalur masuk Universitas Diponegoro dari mahasiswa-mahasiswa tersebut.

Variabel penelitian yang dianalisis terdiri dari variabel respon dan variabel prediktor yang dijelaskan pada Tabel 2.

Tabel 2. Variabel yang Diteliti

Variabel	Definisi
Lama Studi (Y)	0 = ≤ 4 tahun 1 = > 4 tahun
IPK	variabel kontinu
Jalur masuk (JM)	1 = SPMB, SNMPTN 2 = UM 3 = PSSB, PMDK
Jurusan (JUR)	1 = Matematika 2 = Biologi 3 = Kimia 4 = Fisika 5 = Statistika 6 = Ilmu Komputer

4. Analisa dan Pembahasan

Jumlah lulusan sarjana S1 yang diwisuda pada tahun akademik 2012/2013 sebanyak 380 mahasiswa dengan perincian jumlah lulusan jurusan Matematika sebanyak 85 mahasiswa, Biologi 62 mahasiswa, Kimia 90 mahasiswa, Fisika 48 mahasiswa, Statistika 43 mahasiswa dan Ilmu Komputer 52 mahasiswa. Setelah dilakukan pengolahan data dengan menggunakan software SPSS diperoleh hasil sebagai berikut:

Tabel 3. Hasil Output Model Regresi Logistik Biner

Variabel	Koefisien Parameter	W _j	Sig	Odds Rasio
Konstanta	18,772	52,939	0,000	
IPK	-4,855	46,631	0,000	0,008
JM_1	-0,449	1,367	0,242	0,638
JM_2	-0,197	0,221	0,638	0,821
JUR_1	-3,733	44,013	0,000	0,024
JUR_2	-0,673	1,454	0,228	0,510
JUR_3	-1,097	4,137	0,042	0,334
JUR_4	-0,484	0,582	0,445	0,616
JUR_5	-2,637	21,724	0,000	0,072

Dari Tabel 3, maka model regresi logistik binernya adalah sebagai berikut :

$$\pi(x_i) = \frac{e^{18,772-4,855IPK-0,449JM_1-0,197JM_2-3,733JUR_1-0,673JUR_2-1,097JUR_3-0,484JUR_4-2,637JUR_5}}{1 + e^{18,772-4,855IPK-0,449JM_1-0,197JM_2-3,733JUR_1-0,673JUR_2-1,097JUR_3-0,484JUR_4-2,637JUR_5}}$$

Selanjutnya untuk mengetahui apakah secara bersama-sama variabel prediktor mempengaruhi model dilakukan uji Rasio Likelihood, dengan hipotesa

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_8 = 0$$

$$H_1 : \text{paling sedikit satu } \beta_j \neq 0, j = 1, 2, \dots, 8$$

Statistik uji :

$$G = -2 \ln \left(\frac{\text{likeli hood tanpa variabel prediktor}}{\text{likeli hood dengan variabel prediktor}} \right)$$

$$= 469,228 - 320,530 = 148,698$$

Kriteria uji : tolak H_0 jika $G > \chi^2_{0,05;8}$

Karena $G = 148,698 > \chi^2_{0,05;8} = 15,507$, maka H_0 ditolak sehingga secara bersama-sama variabel prediktor mempengaruhi model

Sedangkan untuk mengetahui apakah masing-masing variabel prediktor mempengaruhi model digunakan uji Wald dengan hipotesa

$$H_0: \beta_j = 0, \text{ dengan } j = 1, 2, \dots, 8$$

$$H_1: \beta_j \neq 0, \text{ dengan } j = 1, 2, \dots, 8$$

Kriteria Uji : tolak H_0 jika nilai $W > \chi^2_{0,05;1}$

Karena $\chi^2_{0,05;1} = 3,841$, maka dari Tabel 3 terlihat variabel IPK dan jurusan mempengaruhi model. Sehingga diperoleh model regresi logistik biner sebagai berikut:

$$\pi(x_i) = \frac{e^{18,772-4,855IPK-3,733JUR_1-0,673JUR_2-1,097JUR_3-0,484JUR_4-2,637JUR_5}}{1 + e^{18,772-4,855IPK-3,733JUR_1-0,673JUR_2-1,097JUR_3-0,484JUR_4-2,637JUR_5}}$$

Dari odds ratio untuk IPK yaitu 0,008 terlihat bahwa dengan bertambahnya IPK kecenderungan seorang mahasiswa untuk lulus lebih dari 4 tahun adalah 0,008 kalinya seorang mahasiswa untuk lulus kurang dari 4 tahun, artinya dengan bertambahnya IPK peluang seorang mahasiswa untuk lulus lebih dari 4 tahun lebih kecil daripada lulus kurang dari 4 tahun.

Dari tabel klasifikasi diperoleh hasil sebagai berikut:

Tabel 4. Tabel Klasifikasi

		Data Prediksi Lama Studi		
		≤ 4 tahun	> 4 tahun	Persentase Ketepatan
Data Observasi Lama Studi	≤ 4 tahun	69	48	59
	> 4 tahun	25	238	90,5
				80,8

Sehingga ketepatan klasifikasi adalah $= \left(\frac{69+238}{380} \right) \times 100 = 80,8$

5. Kesimpulan

Dari hasil analisis yang telah dilakukan, maka dapat diperoleh kesimpulan sebagai berikut:

Faktor-faktor yang mempengaruhi lama studi mahasiswa FSM Undip adalah IPK dan jurusan. Hal ini sesuai dengan odds ratio untuk IPK 0,008, yang berarti dengan bertambahnya IPK, peluang seorang mahasiswa untuk lulus lebih dari 4 tahun lebih kecil daripada lulus kurang dari 4 tahun. Sedangkan ketepatan klasifikasi adalah 80,8

DAFTAR PUSTAKA

Agresti, A. 2002. *Categorical Data Analysis*, 2nd edition. John Willey & Sons. New York

- Hadi, S.P. 2012. *Peraturan Rektor Universitas Diponegoro No.209/PER/UN7/2012*. Semarang. Universitas Diponegoro.
- Hosmer, D.W. and Lemeshow, S. 2000. *Applied Logistic Regression*, 2nd edition. John Wiley & Sons. New York
- Kleinbaum, DG and Klein, M. 2010. *Logistic Regression A Self-Learning Text*. Third Edition. Springer. New York