ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 4, Nomor 1, Tahun 2015, Halaman 123 - 132
Online di: http://ejournal-s1.undip.ac.id/index.php/gaussian

KLASIFIKASI LAMA STUDI MAHASISWA FSM UNIVERSITAS DIPONEGORO MENGGUNAKAN REGRESI LOGISTIK BINER DAN *SUPPORT VECTOR MACHINE* (SVM)

Sri Maya Sari Damanik[1], Dwi Ispriyanti[2], Sugito[3]
[1]Mahasiswa Jurusan Statistika FSM UNDIP
[2,3]Staf Pengajar Jurusan Statistika FSM UNDIP

ABSTRAK

Wisuda adalah hasil akhir dari proses kegiatan belajar mengajar selama mengikuti perkuliahan di perguruan tinggi. Dalam mencapai gelar S1 membutuhkan waktu normal yaitu selama empat tahun, tetapi ada banyak mahasiswa yang menyelesaikan studinya melebihi batas normal (lebih dari empat tahun) dan ada juga yang kurang dari empat tahun. Lama studi mahasiswa dapat dipengaruhi oleh banyak faktor antara lain Indeks Prestasi Kelulusan (IPK), jenis kelamin, jurusan, lama studi yang ditempuh, beasiswa, *part time*, organisasi, dan jalur masuk universitas. Pada penelitian ini, akan dilakukan klasifikasi berdasarkan status lama studi mahasiswa lebih dari empat tahun dan kurang dari sama dengan empat tahun. Metode yang digunakan untuk klasifikasi lama studi mahasiswa dengan jenis data nominal adalah Metode *Support Vector Machine* (SVM) dan akan dibandingkan dengan metode Regresi Logistik Biner. Berdasarkan hasil penelitian dengan metode regresi logistik biner, menunjukkan variabel yang berpengaruh terhadap lama studi mahasiswa adalah Jurusan dan IPK dengan ketepatan klasifikasi 70%. Sedangkan ketepatan klasifikasi dengan menggunakan SVM ketepatan klasifikasi tertinggi dengan menggunakan kernel linear, *Polynomial* dan RBF mencapai 90%.

Kata kunci : Lama studi, Regresi Logistik Biner, *Support Vector Machine* (SVM), Ketepatan Klasifikasi.

1. PENDAHULUAN

Universitas Diponegoro (UNDIP) adalah salah satu Universitas di Indonesia yang memiliki 11 (sebelas) fakultas. Fakultas Sains dan Matematika (FSM) adalah salah satu dari 11 Fakultas di Undip. FSM terdiri dari 7 (tujuh) jurusan dengan 6 (enam) program S1 yaitu Matematika, Biologi, Kimia, Fisika, Statistika, dan Teknik Informatika dan D3 Instrumentasi dan Elektronika. Setiap tahun UNDIP menyelenggarakan upacara wisuda dalam 4 periode yaitu periode Januari, April, Agustus, dan Oktober. Dalam 4 periode kelulusan jumlah lulusan dengan jumlah mahasiswa baru tidak sebanding. Lama studi mahasiswa kemungkinan dapat dipengaruhi oleh banyak faktor. Faktor-faktor yang kemungkinan mempengaruhi dalam hal kelulusan antara lain Indeks Prestasi Kelulusan (IPK), jenis kelamin, jurusan, lama studi yang ditempuh, beasiswa, *part time*, organisasi, dan jalur masuk universitas. Misalnya pada pendaftaran sidang yang sudah dijadwalkan, apabila mahasiswa tersebut terlambat sehari melakukan pendaftaran sidang dari batas yang ditentukan maka harus menunggu untuk sidang pada periode selanjutnya, dimana itu akan mempengaruhi lama studi mahasiswa tersebut. Oleh karena itu, peneliti ingin meneliti faktor-faktor yang mempengaruhi lama studi mahasiswa serta ingin mengklasifikasikan kelulusan mahasiswa ke dalam dua kategori yaitu lulus tepat waktu untuk mahasiswa yang menempuh pendidikan S1 kurang dari sama dengan 4 tahun (8 semester) dan lulus tidak tepat waktu untuk mahasiswa yang menempuh pendidikan lebih dari 4 tahun (8 semester).

Support Vektor Machine (SVM) merupakan salah satu bagian dari Data Mining yang digunakan untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi (Santosa, 2007). Menurut Criastinini dan Shawe (2000) dalam Supriyanto (2013), konsep SVM dapat dijelaskan dengan cara sederhana sebagai usaha untuk mencari fungsi pemisah (*hyperplane*)

terbaik dari berbagai alternatif garis pemisah yang mungkin. Fungsi pemisah yang paling baik adalah dengan memaksimalkan nilai margin yaitu jarak dari batas pemisah (*separating hyperplane)* ke masing-masing kelas dan posisi ini tercapai jika garis pemisah itu terletak tepat di tengah-tengahnya, memisahkan antar kelas positif dan kelas negatif. Menurut Supriyanto (2013), prinsip dasar SVM adalah klasifikasi yang bersifat linier, dan selanjutnya dikembangkan agar dapat bekerja pada problem *non-linier* dengan menggunakan fungsi kernel (fungsi yang memudahkan proses pengklasifikasian data).

2. TINJAUAN PUSTAKA

2.1. Peraturan Akademik

Berdasarkan Peraturan Rektor UNDIP Tentang Peraturan Akademik Bidang Pendidikan UNDIP Pendidikan Program Sarjana (S1) mempunyai beban studi sekurang-kurangnya 144 (seratus empat puluh empat) sks dan sebanyak –banyaknya 160 (seratus enam puluh) sks yang dijadwalkan untuk 8 (delapan) semester atau 4 (empat) tahun dan dapat ditempuh selama-lamanya 14 (empat belas) semester atau 7 (tujuh) tahun. Lulusan terbaik dikatakan apabila memiliki Indeks Prestasi Kumulatif (IPK) 3,51- 4,00 dengan pujian *Cumlaude* dan masa studi tidak lebih dari sama dengan 4 (empat) tahun. Berdasarkan buku wisuda ke-130, ke-131, ke-132, dan ke-133 peserta wisuda program sarjana (S1) di Fakultas Sains dan Matematika masih banyak menempuh studi lebih dari 4 tahun.

2.2. Klasifikasi

Salah satu pengukur kinerja klasifikasi adalah tingkat akurasi. Sebuah sistem dalam melakukan klasifikasi diharapkan dapat mengklasifikasi semua set data dengan benar, tetapi tidak dipungkiri bahwa kinerja suatu sistem tidak bisa 100% akurat. Umumnya, pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi. Matriks konfusi merupakan tabel pencatat hasil kerja klasifikasi (Prasetyo, 2012).

Tabel 1. Matriks Konfusi

Hasil Observasi	Kelas hasil prediksi	
	Kelas = 1	Kelas = 0
Kelas = 1	f_{11}	f_{10}
Kelas = 0	f_{01}	f_{00}

Akurasi hasil klasifikasi yang dapat dihitung dengan formula, sebagai berikut:

$$Akurasi = \frac{Banyaknya\ data\ yang\ diprediksi\ dengan\ benar}{Banyaknya\ data\ prediksi} = \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}}$$

2.3. Metode Regresi Logistik Biner

Model regresi logistik biner digunakan untuk menganalisa hubungan antara satu variabel respon (variabel tak bebas) dan beberapa variabel bebas, dengan variabel responnya berupa data kualitatif dikotomi yaitu bernilai 1 untuk menyatakan keberadaan sebuah karakteristik dan bernilai 0 untuk menyatakan ketidakberadaaan sebuah karakteristik. (Agresti, 2002). Jika diketahui Y variabel dependen bernilai 0 dan 1, maka

$P = (Y = 1|X = x_i) = \pi(x_i)$ dan $P = (Y = 0|X = x_i) = 1 - \pi(x_i)$
Sehingga diperoleh model regresi logistik :

$$\pi(x_i) = \frac{e^{g(x)}}{1+e^{g(x)}}$$

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

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}$$

Menurut Agresti (2002) untuk menentukan estimasi parameter digunakan metode masksimum Likelihood yang membutuhkan turunan pertama dan turunan kedua dari fungsi likelihood.

$$P(y_i = 1) = \binom{n}{1}\{\pi(x_i)\}^{y_i}\{1 - \pi(x_i)\}^{n-y_i}$$
$$= \frac{n!}{1!(n-1)!}\{\pi(x_i)\}^{y_i}\{1 - \pi(x_i)\}^{n-y_i}$$

Maka fungsi likelihoodnya adalah

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

untuk memperoleh dugaan maksimum likelihood bagi β dengan algoritma digunakan iterasi *newthon raphson* sebagai berikut :

1. Dipilih taksiran awal untuk β, misal $\hat{\beta} = 0$
2. Dihitung **X'(Y- $\boldsymbol{\pi(x_i)}$)** dan **X'VX,** selanjutnya dihitung invers dari **X'VX**
3. Pada setiap i+1 dihitung taksiran baru yaitu
   $\hat{\beta}_{i+1} = \hat{\beta}_I + \{\ \mathbf{X'VX}\ \}^{-1}\ \{\mathbf{X'(Y-\pi(x_i))}\}$ **,**
   iterasi berakhir jika diperoleh $\hat{\beta}_{i+1} \cong \hat{\beta}_I$

Pengujian Signifikansi Parameter

Untuk menguji signifikansi dari parameter dalam model digunakan uji rasio likelihood, uji wald dan uji kesesuain model

1. Uji Rasio Likelihood
   - Hipotesis
     $H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$
     H_1 : paling sedikit ada satu $\beta_j \neq 0$ dengan j = 1,2,…,p
   - Statistik uji rasio likelihood adalah
     $G = -2 \ln\left(\frac{likelihood\ tanpa\ variabel\ bebas}{likelihood\ dengan\ variabel\ bebas}\right)$
   - Kriteria uji :
     H_0 ditolak jika $G > \chi^2_{(\alpha,p)}$
2. Uji Wald
   - Hipotesis
     $H_0 : \beta_j = 0$ dengan j= 1,2, …, p
     $H_1 : \beta_j \neq 0$ dengan j = 1,2,…,p
   - Statistik uji wald

$$W_j = \left\{ \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \right\}^2$$

- Kriteria uji :
  H_0 ditolak jika $W_j > \chi^2_{(\alpha,1)}$

3. Uji Kesesuain model
   - Hipotesis
     H_0 = Model sesuai
     H_1 = Model tidak sesuai
   - Statistik Uji :
     $$\hat{C} = \sum_{k=1}^{g} \frac{(o_k - n\ \bar{\pi}_k)^2}{n_k \pi_k (1-\pi_k)}$$
     dengan $o_k = \sum_{j=1}^{nk} y_j$; $\bar{\pi}_k = \sum_{j=1}^{nk} \frac{m_j \hat{\pi}_j}{nk}$
   - Kriteria uji :
     Tolak H_0 jika $\hat{C} > \chi^2_{(\alpha,g-2)}$

2.4. Support Vector Machine (SVM)

2.4.1 Klasifikasi Linier Separable

Misalkan diberikan himpunan $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_n\}$, dengan $\boldsymbol{x}_i \in \mathfrak{R}^P$, dengan Telah diketahui $\boldsymbol{X}$ memiliki pola tertentu, yaitu apabila $\boldsymbol{x}_i$ termasuk dalam suatu kelas maka diberi label $y_i = +1$, jika tidak diberi label $y_i = -1$ untuk itu label masing-masing dinotasikan $y_i \in \{-1,+1\}$ sehingga data berupa pasangan $(\boldsymbol{x}_1,y_1)$, $(\boldsymbol{x}_2,y_2)$,… $(\boldsymbol{x}_i,y_i)$ dimana i= 1,2,….,n yang mana n adalah banyak data. Diasumsikan kedua kelas -1 dan +1 dapat terpisah secara sempurna oleh fungsi pemisah berdimensi *p*, yang didefinisikan : $\boldsymbol{w}^T\boldsymbol{x} + b = 0$ dimana $\boldsymbol{w}$ dan b adalah parameter model. Untuk mendapatkan fungsi pemisah terbaik adalah dengan mencari fungsi pemisah yang terletak ditengah-tengah antara dua bidang pembatas kelas dan untuk mendapatkan fungsi pemisah terbaik itu, sama dengan memaksimalkan margin atau jarak antara dua set objek dari kelas yang berbeda (Santosa, 2007). Selanjutnya, diformulasikan kedalam persamaan *quadratic programming* (QP), dengan meminimalkan invers persamaan $\frac{1}{2}\|w\|^2$, dimana $\|\boldsymbol{w}\|^2 = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}$ dengan syarat $y_i[(\boldsymbol{w}^T\boldsymbol{x}) + b] - 1 \geq 0$, $i = 1,2,3 \ldots, n$. Optimalisasi ini dapat diselesaikan dengan fungsi *Lagrange Multiplier* (Prasetyo, 2012) :

$$L(\boldsymbol{w}, b, \alpha) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} - \sum_{i=1}^{n} \alpha_i \{y_i[\boldsymbol{w}^T\boldsymbol{x}_i + b] - 1\}$$

Sehingga syarat optimal dari fungsi *lagrange multiplier* tersebut adalah

$$\frac{\partial L}{\partial b} = 0 \ \rightarrow \ \sum_{i=1}^{n} \alpha_i y_i = 0 \quad ; \qquad \frac{\partial L}{\partial w} = 0 \ \rightarrow \ \boldsymbol{w} = \sum_{i=1}^{n} \alpha_i \boldsymbol{x}_i y_i = 0$$

$$L(\boldsymbol{w}, b, \alpha) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} - \sum_{i=1}^{n} \alpha_i y_i (\boldsymbol{w}^T\boldsymbol{x}_i) - b\sum_{i=1}^{n} \alpha_i y_i + \sum_{i=1}^{n} \alpha_i$$

dimana $\boldsymbol{w}^T\boldsymbol{w} = \sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j (\boldsymbol{x}_i^T\boldsymbol{x}_j)$

maka, $L_d = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\, \alpha_j y_i\, y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$

$\max_\alpha L_d = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\, \alpha_j y_i\, y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$

dengan batasan, $\alpha_i \geq 0, \quad i = 1,2,\dots,n \quad$ dan $\quad \sum_{i=1}^{n} \alpha_i y_i = 0$

Dari hasil perhitungan ini diperoleh α_i kebanyakan bernilai positif. Data yang berkorelasi dengan $\alpha_i$yang positif disebut *support vector* (Vapnik,1995). Setelah solusi permasalahan *quadratic programming* ditemukan (nilai α_i), maka kelas dari data yang akan diprediksi atau data testing dapat ditentukan berdasarkan fungsi sebagai berikut :

$f(x_t) = \sum_{i=1}^{ns} \alpha_i y_i x_i . x_t + b$

2.4.2 Klasifikasi Linier Non-Separable

Metode SVM juga dapat digunakan dalam kasus *non-separable* dengan memperluas formulasi yang terdapat pada kasus linier. Masalah optimasi sebelumnya baik pada fungsi obyektif maupun kendala dimodifikasi dengan mengikutsertakan variabel *Slack* $\xi > 0$. Variabel *slack* merupakan sebuah ukuran kesalahan klasifikasi. Formulasinya sebagai berikut (Gunn, 1998).

$y_i[(\boldsymbol{w}^T \boldsymbol{x}_i) + b] \geq 1 - \xi_i\,, \;\; i = 1,2,\dots,n$

$\Phi(w,\xi) = \frac{1}{2} w^T w + C\sum_{i=1}^{n} \xi_i$ sehingga $\max_\alpha L_d = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\, \alpha_j y_i\, y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$

Dimana parameter *C* berfungsi untuk mengontrol hubungan antara variabel *slack* dengan margin. Semakin besar nilai *C*, maka semakin besar pula pelanggaran yang dikenakan untuk tiap klasifikasi (Prasetyo, 2012) dengan batas $0 \leq \alpha_i \leq C$, $i = 1,2,\dots, n$ dan $\sum_{i=1}^{n} \alpha_i y_i$.

2.4.3 Klasifikasi Non-Linier

Data yang distribusi kelasnya tidak linier biasanya digunakan pendekatan kernel pada fitur data awal (Prasetyo, 2012). Proses pemetaan pada fase ini memerlukan perhitungan dot-product dua buah data pada ruang fitur baru yang dinotasikan sebagai $\Phi(x_i).\Phi(x_j)$. Trik komputasi ini sering dikenal dengan trik kernel, sebagai berikut: $K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \boldsymbol{\Phi}(\boldsymbol{x_i}).\boldsymbol{\Phi}(\boldsymbol{x_j})$
Dan prediksi pada data dengan dimensi fitur yang baru diformulasikan dengan

$$f(\Phi(z)) = sign(w.\Phi(z) + b) = sign(\sum_{i=1}^{l} \alpha_i y_i \Phi(x_i).\Phi(z) + b)$$

Dengan l adalah jumlah data yang menjadi *support vector*, x_i adalah *support vektor*, dan z adalah data uji yang akan diprediksi (Prasetyo, 2012).

Menurut (Kecman, 2005), fungsi kernel yang biasanya dipakai dalam literatur SVM:

- Linier : $\boldsymbol{K}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{x}.\boldsymbol{y}$
- *Polynomial* : $\boldsymbol{K}(\boldsymbol{x},\boldsymbol{y}) = (\boldsymbol{x}.\boldsymbol{y} + \boldsymbol{c})^{\boldsymbol{d}}$
- *Radial Basis Function* (RBF) : $\boldsymbol{K}(\boldsymbol{x},\boldsymbol{y}) = exp(-\gamma\|\boldsymbol{x} - \boldsymbol{y}\|^2)$, dengan $\gamma = \frac{1}{2\sigma^2}$

$\boldsymbol{x}$ dan $\boldsymbol{y}$ adalah pasangan dua data dari semua bagian data latih. Parameter $\sigma, c, d > 0$, merupakan konstanta.

3. METODOLOGI PENELITIAN

Jenis data yang digunakan dalam penelitian ini adalah data sekunder dan data primer. Data sekunder adalah data yang diperoleh atau dikumpulkan peneliti dari berbagai sumber yang telah ada. Pada penelitian ini data sekunder diperoleh dari buku Wisuda UNDIP ke-132, Wisuda UNDIP ke-133, Wisuda UNDIP ke-134, dan Wisuda UNDIP ke-135. Data yang diperoleh dari buku tersebut adalah Nama, No HP, lama studi mahasiswa, jenis kelamin, IPK, dan Jurusan. Sedangkan data primer adalah data yang diperoleh atau dikumpulkan oleh peneliti secara langsung dari sumber datanya. Teknik yang digunakan peneliti untuk mengumpulkan data primer antara lain wawancara secara langsung dan melalui media *elektronik* (*SMS* dan *By-Phone*) atau media *social* (*Watshapp* dan *Facebook*).

Pada penelitian ini, terdapat tujuh variabel independen (x), yang terdiri atas Jenis kelamin(x_1), Jurusan (x_2), IPK (x_3), Beasiswa (x_4), *Part time* (x_5), Organisasi (x_6), dan Jalur Masuk Undip (x_7). Data akan dikelompokkan ke dalam dua kelas yang merupakan variabel dependen (y), yaitu kurang dari sama dengan 4 tahun = 0 dan kelas lebih dari 4 tahun = 1.

Langkah-langkah Analisis

1. Menentukan variabel independen dan variabel dependen
2. Membuat rancangan penelitian yang meliputi tempat, waktu dan lokasi, sumber data, populasi, dan jumlah sampel peneltian serta metode pengambilan sampel.
3. Membagi data menjadi data *training* dan *testing* dengan proporsi 180:20 secara acak dengan bantuan paket program R 2.14.2.
4. Memodelkan dengan menggunakan Regresi Logistik Biner untuk mengetahui faktor-faktor yang mempengaruhi lama studi mahasiswa FSM Undip dengan menggunakan SPSS 16.
    a. Membuat model awal
    b. Melakukan uji rasio likelihood untuk mengetahui apakah variabel independen yang terdapat dalam model berpengaruh nyata atau tidak secara keseluruhan
    c. Melakukan uji wald untuk mengetahui variabel bebas yang mempunyai hubungan kuat dengan variabel respon
    d. Melakukan uji kesesuaian model
    e. Menentukan model akhir
    f. Menghitung ketepatan klasifikasi regresi logistik biner untuk data testing
5. Melakukan klasifikasi lama studi mahasiswa FSM Undip dengan menggunakan metode *Support Vector Machine* (SVM). Berikut algoritma metode SVM:
    a. Melakukan transformasi data sesuai dengan format software SVM yang akan digunakan
    b. Menentukan fungsi kernel permodelan
    c. Menetukan nilai-nilai parameter kernel dan parameter cost untuk optimasi
    d. Menentukan nilai parameter terbaik untuk optimasi data training untuk klasifikasi data testing
    e. Menghitung ketepatan klasifikasi
6. Membandingkan ketepatan klasifikasi yang diperoleh dari Regresi Logistik Biner dengan SVM
7. Kesimpulan

4. HASIL DAN PEMBAHASAN

4.1. Analisis Deskriptif

Analisis deskriptif digunakan untuk memperoleh gambaran data secara umum. Data yang digunakan pada penelitian ini adalah sebanyak 200 data alumni mahasiswa FSM periode ke-131 sampai dengan period eke-134 dengan persentase 61,5 % diantaranya adalah mahasiswa yang lama studi dengan waktu lebih dari 48 bulan (> 4 tahun), sedangkan sisanya yaitu 38,5% merupakan mahasiswa yang lama studinya kurang dari sama dengan 48 bulan (≤ 4 tahun).

4.2. Metode Regresi Logistik Biner

Pada analisis ini menggunakan data lama studi mahasiswa FSM dengan proporsi data *training* dan *testing* dengan perbandingan 90% dan 10%, yaitu data *training* sebanyak 180 dan data *testing* sebanyak 20. Pada penelitian ini variabel dependen yang digunakan adalah Lama studi mahasiswa kurang dari sama dengan 4 tahun (≤ 4tahun) dan lebih dari 4 tahun (> 4tahun). Variabel independen terdiri dari ipk, jenis kelamin, organisasi, beasiswa, *part time*, dan jalur masuk Undip.

Pada uji Rasio Likelihood, berdasarkan keputusan H_0 ditolak berarti bahwa variabel independen yang terdapat pada model berpengaruh nyata secara serentak. Pada uji wald, diperoleh hasil bahwa variabel yang tidak berpengaruh signifikan terhadap variabel dependen Lama Studi adalah variabel Jenis Kelamin, Beasiswa, Organisasi, *Part time*, dan Jalur Masuk. Sedangkan, variabel yang berpengaruh signifikan terhadap variabel dependen Lama Studi adalah Jurusan dan IPK.

Setelah dilakukan uji signifikansi terhadap model, baik secara keseluruhan maupun individual diperoleh hasil bahwa variabel Jurusan dan IPK berpengaruh signifikan terhadap variabel Lama Studi. Oleh karena itu, untuk memperoleh model akhir yang sesuai dilakukan analisis regresi logistik biner kembali dengan tidak mengikutsertakan variabel yang tidak berpengaruh signifikan. Pada uji Rasio Likelihood, bahwa variabel independen yang terdapat pada model berpengaruh nyata secara serentak. Pada uji wald, diperoleh hasil bahwa variabel Jurusan dan IPK berpengaruh signifikan terhadap variabel dependen Lama Studi. Berikut adalah model terbaik

$$\boldsymbol{\pi(x)} = \frac{exp(1.462-4.457x1(1)-1.832x1(2)-2.581x1(3)-1.770x1(4)-1.815x1(5)+4.165x2(1)+1.252x2(2)}{1+exp(1.462-4.457x1(1)-1.832x1(2)-2.581x1(3)-1.770x1(4)-1.815x1(5)+4.165x2(1)+1.252x2(2)}$$

Keterangan :

X_1 = Jurusan ; X_2 =IPK

Dalam menghitung nilai ketepatan klasifikasi digunakan data *testing* sebanyak 20. Berikut adalah hasil ketepatan klasifikasi data *testing* regresi logistik biner sebagai berikut:

Tabel 2. Ketepatan Klasifikasi Regresi Logistik Biner

Kelas asli	Prediksi		Ketepatan klasifikasi
	≤ 4tahun	> 4tahun	70%
≤ 4tahun	3	3	
> 4tahun	3	11	

$$\text{akurasi} = \frac{f_{11}+f_{00}}{f_{11}+f_{10}+f_{01}+f_{00}}$$

$$= \frac{3+11}{3+3+3+11} = 0.70$$

Berdasarkan Tabel 2. menunjukkan hasil ketepatan klasifikasi dengan menggunakan rumus fungsi peluang regresi logistik biner bahwa secara keseluruhan hasil dari prediksi dengan menggunakan data *training* sebanyak 180 dan data *testing* sebanyak 20 menghasilkan ketepatan klasifikasi sebesar 70%.

4.3. Support Vector Machine (SVM)

Klasifikasi untuk data Lama Studi mahasiswa FSM Undip periode ke-131 sampai dengan periode ke-134 pada tahun 2013 menggunakan metode SVM, penelitian ini menggunakan fungsi kernel Linier, *Polynomial*, dan *Radial Basis Function* (RBF). Melakukan klasifikasi dengan menggunakan SVM, ada beberapa nilai C yang digunakan untuk mengetahui ketepatan klasifikasi terbaik yaitu 0.1, 1, 10, dan 100. Berikut adalah hasil eror klasifikasi dengan menggunakan data *training* dan *testing* sebesar 90% dan 10%.

Tabel 3. Eror Hasil Klasifikasi SVM untuk Penentuan Parameter Terbaik

Fungsi Kernel	Parameter Hyperplane Cost (C)		Eror Klasifikasi
Linier	0.1		0.327778
	1		0.344444
	10		0.327778
	100		0.327778
Polynomial	0.1	d= 2	0.3944444
		3	0.3055556
		4	0.3611111
		5	0.3277778
	1	d= 2	0.4055556
		3	0.3166667
		4	0.3611111
		5	0.3277778
	10	d= 2	0.4666667
		3	0.4
		4	0.3555556
		5	0.3277778
	100	d= 2	0.4388889
		3	0.4111111
		4	0.3555556
		5	0.3277778

Fungsi Kernel	Parameter Hyperplane Cost (C)		Eror Klasifikasi
RBF	0.1	γ = 0.003	0.3777778
		0.007	0.3777778
		0.015	0.3777778
		0.031	0.3777778
	1	γ = 0.003	0.3777778
		0.007	0.3777778
		0.015	0.3833333
		0.031	0.4055556
	10	γ = 0.003	0.3333333
		0.007	0.3111111
		0.015	0.3722222
		0.031	0.3333333
	100	γ = 0.003	0.3722222
		0.007	0.35
		0.015	0.3444444
		0.031	0.3333333

Untuk mengetahui hasil klasifikasi terbaik dilakukan pengolahan data dengan data *training* dan *testing* yang berbeda-beda, berikut adalah hasil klasifikasi dengan menggunakan fungsi kernel Linear, Polynomial, dan RBF .

Tabel 4. Hasil Klasifikasi data *testing* dengan menggunakan SVM

Fungsi Kernel	*Testing*		
	40	20	10
Linear	**0.675**	0.85	**0.9**
RBF	0.65	**0.9**	0.7
Polynomial	0.625	0.8	**0.9**

Dari Tabel 4. diperoleh informasi bahwa pada data *testing* yaitu 40 data, klasifikasi terbaik dengan menggunakan fungsi kernel Linier dengan akurasinya adalah 67.5%. Data testing yaitu 20 data, klasifikasi terbaik dengan menggunakan fungsi kernel RBF dengan akurasi yaitu 90%. Sedangkan pada data *testing* yaitu 10 data, klasifikasi terbaik dengan menggunakan fungsi kernel Linier atau *Polynomial* dengan akurasi 90%.

4.4. Perbandingan klasifikasi Metode Regresi Logistik Biner dan Support Vector Machine (SVM)

Dari evaluasi hasil klasifikasi pada analisis ini, membandingkan nilai ketepatan klasifikasi dengan melakukan beberapa percobaan untuk mengetahui klasifikasi terbaik menggunakan Regresi Logistik Biner atau metode *Support Vector Machine* diperoleh akurasi sebagai berikut:

Tabel 5. Perbandingan Klasifikasi dengan Regresi Logistik Biner dan SVM

Data Testing	Reg.Log.Biner	SVM		
		Linier	Polynomial	RBF
40	0.65	0.675	0.65	0.65
20	0.70	0.8	0.8	0.9
10	0.6	0.9	0.9	0.7

Dari Tabel 5 diperoleh informasi bahwa, klasifikasi terbaik adalah dengan menggunakan metode SVM dengan data *testing* 20 data, dengan fungsi kernel RBF, ketepatan klasifikasi yaitu 90% , dengan parameter C = 10 dan γ = 0.031.

5. KESIMPULAN

Berdasarkan analisis yang telah dilakukan, bahwa dengan menggunakan data Lama Studi mahasiswa FSM Undip tahun 2013 periode ke-131 sampai dengan periode ke-134 dengan menggunakan data *training* dan *testing* sebanyak 180 dan 20 diperoleh kesimpulan yaitu pada metode Regresi Logistik Biner variabel yang berpengaruh terhadap variabel Lama Studi adalah variabel Jurusan dan IPK dengan model terbaik adalah

$$\boldsymbol{\pi(x)} = \frac{e^{1.462-4.457JRS1-1.832JRS2-2.581JRS3-1.770JRS4-1.815JRS5+4.165IPK1+1.252IPK2}}{1+e^{1.462-4.457JRS1-1.832JRS2-2.581JRS3-1.770JRS4-1.815JRS5+4.165IPK1+1.252IPK2}}$$

dengan ketepatan klasifikasi adalah 70%. Sedangkan ketepatan klasifikasi dengan menggunakan metode SVM yaitu dengan fungsi kernel Linear, *Polynomial*, dan *Radial Basis Function* (RBF) ketepatan klasifikasi tertinggi dengan tiga kali percobaan mencapai 90%.

DAFTAR PUSTAKA

Agresti, A. 2002. *Categorical Data Analysis Second Edition.* Jhon Wiley & Sons, Inc: USA.

Cristanini, N and Jhon S. 2000. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambrige University Press. Cambridge.

Gunn, S. 1998. *Support Vector Machine for Classification and Regression.* University of Southampton: Southampton.

Hastie, T., Jerome, F., and Robert, T. 2008. *The Elements of Statistical Learning (2nd Ed) : Data Mining, Inference and Predition*. Stanford. California.

Prasetyo, E. 2012. *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*. C.V Andi Offset: Yogyakarta.

Santosa, B. 2007. *Data Mining: Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Graha Ilmu: Yogyakarta.

Supriyanto, H. 2013. Implementasi Support Vector Machines untuk Memprediksi Arah Pergerakan Harga Harian Valuta Asing (EUR/USD, GBP/USD, dan USD/JPY) dengan Metode Kernel Trick Menggunakan Fungsi Kernel Radial Basis Function. Jurnal Mahasiswa Statistik Vol 1, No1. Tersedia pada: http://statistik.studentjournal.ub.ac.id/index.php/statistik/article/view/7

Vapnik, V.N. 1999. *The Nature of Statistical Learning Theory Second Edition*. Springer: New York.