

ANALISIS DISKRIMINAN, REGRESI LOGISTIK, NEURAL NETWORK DAN MARS PADA PENGKLASIFIKASIAN DATA HBAT DAN IRIS

THOMAS PENTURY

Jurusan Matematika, FMIPA, UNPATTI
Jl. Ir. M. Putuhenam, Kampus Unpatti, Poka-Ambon
thomypentury@yahoo.com

ABSTRACT

The purpose of this research is to apply and compare the discriminant analysis, logistic regression, Neural Network (NN) and Multivariate Adaptive Regression Spline (MARS) at HBAT and IRIS data. Next will be the classification of fourth methods using the SPSS statistical software, MINITAB, MARS, and R. The results showed that the data HBAT predictor variables affected to the response variable which is the quality of the product (X_6), Complaint resolution (X_9) and Salesforce image (X_{12}), whereas all predictor variables on the IRIS data affect the response variable. A more precise method used in HBAT data classification is NN and discriminant analysis because the value of the resulting classification accuracy is greater, especially for testing. While a more precise method used in the IRIS data classification is discriminant analysis because the value of the resulting classification accuracy is greater

Keywords: *Discriminant Analysis, Logistic Regression, Neural Network, Multivariate Adaptive Regression Spline*

PENDAHULUAN

Perkembangan teknologi memungkinkan dilakukannya Ada beberapa metode statistika yang dapat digunakan untuk menyelesaikan permasalahan klasifikasi, antara lain analisis diskriminan dan regresi logistik. Pada analisis diskriminan akan diperoleh suatu fungsi yang dikenal dengan fungsi diskriminan yang dapat digunakan untuk mengelompokkan obyek. Sedangkan pada regresi logistik akan diperoleh suatu model logistik yang digunakan untuk menjelaskan hubungan antara prediktor dan respon (yang bersifat dikotomis atau ada dua kategori/kelompok), serta untuk mengelompokkan obyek ke dalam salah satu dari dua kategori respon. Dalam perkembangannya, regresi logistik dapat juga digunakan untuk respon kategori lebih dari dua kelompok, yang dikenal dengan regresi logistik polikotomis. Kedua metode statistika ini dalam beberapa *literature* klasifikasi sering disebut sebagai model klasik. Metode-metode ini mempunyai beberapa asumsi yang harus dipenuhi berkaitan dengan skala pengukuran prediktor, keterkaitan antara prediktor, dan distribusi bersama dari prediktor.

Salah satu metode klasifikasi yang berkembang dari kelompok *machine learning* adalah NN. Model ini tidak mensyaratkan skala pengukuran dan distribusi tertentu dari prediktor atau input dalam terminologi NN. Secara umum, ada dua kelompok besar dalam NN dikaitkan dengan ada tidaknya respon, yaitu *supervised* dan *unsupervised* NN. Dalam kasus analisis klasifikasi ini,

NN yang digunakan adalah termasuk dalam kelompok *supervised* NN, karena proses pembelajarannya (optimisasi fungsi) terawasi oleh suatu respon (output klasifikasi). Metode klasifikasi lain yang dikembangkan dari pendekatan nonparametrik, khususnya spline, adalah MARS. Dalam beberapa *literature* klasifikasi, kedua model ini seringkali disebut sebagai bagian dari model klasifikasi modern.

Huang dkk. (2003), membandingkan model *logistics regression, multivariate discriminan analysis, neural network* dan *Support Vector Machines*. Dari hasil penelitian tersebut model yang mempunyai *accuracy* ketepatan lebih tinggi adalah dengan model NN dan *Support Vector Machines*. Chikolwa (2007), telah berhasil membandingkan *predictive power* antara *ANN models* dan *ordinal regression models*, dan hasilnya ANN mempunyai *predictive power* yang lebih tinggi baik untuk *training* maupun *testing*.

Tujuan dari penelitian ini adalah menerapkan dan membandingkan analisis diskriminan, regresi logistik, NN dan MARS pada data HBAT dan IRIS. Selanjutnya akan dibandingkan performansi atau kebaikan hasil pemodelan, yang mencakup aspek interpretasi model dikaitkan dengan hubungan antara prediktor (input) dan respon (output), serta aspek ketepatan klasifikasi khususnya pada obyek baru yang tidak dimasukkan dalam pembentukan model (*data testing*).

TINJAUAN PUSTAKA

Pada bagian ini akan dijelaskan mengenai analisis diskriminan, regresi logistik, NN dan MARS.

Analisis Diskriminan

Analisis diskriminan merupakan metode statistika untuk mengelompokkan atau mengklasifikasikan sejumlah obyek ke dalam beberapa kelompok berdasarkan beberapa variabel. Setiap obyek yang diklasifikasikan akan menjadi anggota dari salah satu kelompok dan tidak ada obyek yang menjadi anggota lebih dari satu kelompok (Johnson & Wichern, 1992). Analisis diskriminan mengasumsikan data berdistribusi normal multivariat dan matriks kovarian yang homogen. Uji asumsi distribusi normal multivariat dapat dilakukan Chi-Square plot sedangkan asumsi homogenitas matriks kovarian diuji dengan uji Box's M. Analisis diskriminan adalah salah satu metode statistika yang digunakan untuk mengelompokkan objek pengamatan ke dalam kelompok yang telah didefinisikan berdasarkan sejumlah variabel prediktor (Dillon & Goldstein, 1984). Analisis diskriminan dapat menggambarkan perbedaan antar kelompok secara grafis atau melalui persamaan matematis yang disebut fungsi diskriminan. Fungsi diskriminan berfungsi sebagai aturan yang dapat membedakan objek pengamatan sekaligus mengelompokkannya ke dalam salah satu kelompok secara optimal (Johnson & Wichern, 2002). Kinerja suatu fungsi diskriminan diukur oleh total proporsi salah pengelompokkan objek pengamatan. Semakin kecil total proporsi salah pengelompokkan, semakin baik kinerja suatu fungsi diskriminan. Fungsi diskriminan tersebut mampu membedakan objek pengamatan secara optimal sehingga meminimumkan salah pengelompokkan objek pengamatan.

Regresi Logistik

Regresi logistik merupakan metode yang dipergunakan untuk menganalisis hubungan antara variabel respon dengan variabel prediktor. Variabel prediktor yang dipergunakan berupa data kategori atau kontinu dan variabel responnya berupa data dengan skala nominal atau ordinal (Agresti, 1990). Regresi logistik dapat dipergunakan untuk pengklasifikasian sejumlah objek ke dalam beberapa kelompok. Variabel respon Y yang bersifat random dan dikotomus, yakni bernilai 1 dengan probabilitas π dan bernilai 0 dengan probabilitas $1 - \pi$, disebut sebagai *point-binomial* (Le, 1998). Model regresi logistik dengan lebih dari satu variabel prediktor disebut juga model multivariat (Hosmer & Lemeshow, 1989).

Model regresi logistik dengan k variabel prediktor adalah :

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)} \dots\dots(1)$$

Jika model pada persamaan (1) ditransformasi dengan menggunakan transformasi logit, maka akan menghasilkan bentuk logit yaitu :

$$g(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \dots\dots\dots(2)$$

Dengan $g(x)$ merupakan fungsi linear dalam parameter-parameternya

Neural Network

Jaringan Syaraf tiruan adalah suatu sistem komputasi dengan arsitektur dan operasinya diilhami oleh pengetahuan tentang sel Syaraf biologis di dalam otak. Arsitektur ini merupakan salah satu representasi buatan dari otak manusia yang selalu mencoba menstimulasi proses pembelajaran pada otak manusia tersebut. NN dapat digambarkan sebagai model matematika dan komputasi untuk fungsi aproksimasi *non-linier*, klasifikasi data *cluster* dan regresi nonparametrik atau sebuah simulasi dari koleksi model Syaraf biologis (Hermawan, 2006). Seperti otak manusia, jaringan syaraf juga terdiri dari beberapa neuron dan neuron-neuron tersebut saling terhubung. Cara kerja neuron tiruan ini mirip dengan sel neuron biologis. Apabila input tersebut melewati suatu nilai ambang tertentu, maka neuron tersebut akan diaktifkan. Apabila neuron tersebut diaktifkan, maka neuron tersebut akan mengirimkan output melalui bobot-bobot outputnya ke semua neuron yang berhubungan dengannya.

Multivariate Adaptive Regression Spline

MARS merupakan sebuah pendekatan untuk memodelkan regresi non-parametrik multivariat yang dicetuskan pertama kali oleh Friedman (1991). MARS telah banyak diadopsi oleh bidang ilmu komputer sebagai kompetitor metode lain seperti jaringan syaraf tiruan (*Neural Networks*), *Generalized Additive Models* (Hastie & Tibshirani, 1990), dan *Classifications and Regression Trees* (Breiman dkk, 1993) yang semuanya bertujuan untuk menemukan model estimasi dengan pendekatan paling baik terhadap suatu fungsi sesungguhnya. Model MARS digunakan untuk mengatasi kelemahan RPR yaitu menghasilkan model yang kontinu pada knots. Penentuan knots secara otomatis pada MARS menggunakan algoritma *forward stepwise* dan *backward stepwise* yang didasarkan pada nilai GCV minimum.

METODOLOGI PENELITIAN

Data yang digunakan dalam penelitian ini terdiri dari 2 yaitu data HBAT dan data IRIS. Data HBAT merupakan data hasil survey pada suatu pabrik produsen kertas. Data ini terdiri dari 13 variabel prediktor yaitu kualitas produk (X_6), aktifitas E-Commerce (X_7), *technical support* (X_8), *Complaint resolution* (X_9), *advertising* (X_{10}), *product line* (X_{11}), *salesforce image* (X_{12}), kompetisi harga (X_{13}), Garansi dan klaim (X_{14}), produk baru (X_{15}), pemesanan dan penagihan (X_{16}), fleksibilitas harga (X_{17}) dan kecepatan pengiriman (X_{18}). Sedangkan variabel respon adalah variabel *customer intention* yaitu keinginan melanjutkan kerjasama di waktu yang akan datang (X_{23}). Sedangkan data IRIS terdiri dari 4 variabel prediktor yaitu panjang PETAL, lebar PETAL, panjang SEPAL dan lebar SEPAL bunga IRIS, serta variabel respon terdiri dari 3 jenis bunga yaitu SETOSA, VERSICOLOR, dan VIRGINICA.

Setiap data dibagi menjadi dua kelompok yaitu data untuk pemodelan (*training*) dan evaluasi (*testing*) dimana dalam penelitian ini perbandingan *training* dan *testing* adalah 80:20, 70:30, 60:40, 50:50 dan 40:60. Selanjutnya akan dilakukan pengklasifikasian dengan analisis diskriminan, analisis regresi logistik, MARS dan NN. Keempat metode tersebut diaplikasikan dengan menggunakan *software* statistika yaitu SPSS, MINITAB, MARS dan R

HASIL DAN PEMBAHASAN

Analisis diskriminan, regresi logistik, NN dan MARS akan diterapkan pada data HBAT dan IRIS. Masing-masing data dibagi menjadi dua kelompok yaitu data untuk pemodelan (*training*) dan evaluasi (*testing*) dimana perbandingan *training* dan *testing* adalah 80:20, 70:30, 60:40, 50:50 dan 40:60.

Analisis Diskriminan, Analisis Regresi Logistik, MARS dan NN Pada Data HBAT

Analisis Diskriminan

Analisis akan dilakukan untuk perbandingan *training* dan *testing* adalah 80:20. Dengan menggunakan *software* SPSS diperoleh 3 variabel prediktor yang berpengaruh terhadap variabel respon, sehingga diperoleh fungsi diskriminan yaitu :

$$\hat{y} = 0.524X_6 + 0.767X_9 + 0.538X_{12}$$

Dari fungsi di atas terlihat bahwa variabel X_9 memberikan kontribusi terbesar karena memiliki koefisien yang paling besar. Untuk perbandingan *training* dan *testing* lainnya, hasilnya adalah sebagai berikut :

Tabel 1. Koefisien Fungsi Diskriminan Data HBAT

Perbandingan Training & Testing	Koefisien		
	X_6	X_9	X_{12}
70:30	0.552	0.647	0.640
60:40	0.680	0.736	0.393
50:50	0.651	0.859	0.368
40:60	0.588	0.693	0.551

Dari hasil yang ditunjukkan dalam Tabel 1 terlihat bahwa untuk semua perbandingan *training* dan *testing*, variabel X_9 memberikan kontribusi terbesar.

Dengan menggunakan *software* SPSS dan MINITAB diperoleh ketepatan klasifikasi untuk masing-masing perbandingan *training* dan *testing* adalah sebagai berikut :

Tabel 2. Ketepatan Klasifikasi Analisis Diskriminan Data HBAT

Perbandingan Training & Testing	Ketepatan Klasifikasi	
	Training(%)	Testing(%)
80:20	80	80
70:30	78.6	80
60:40	85	77.5
50:50	82	82
40:60	70	86.7

Analisis Regresi Logistik

Analisis regresi logistik akan dilakukan untuk perbandingan *training* dan *testing* yaitu 80:20. Dengan menggunakan *software* SPSS diperoleh model logit adalah :

$$g(x) = -18.095 + 0.736X_6 + 1.315X_9 + 0.947X_{12}$$

Untuk perbandingan *training* dan *testing* lainnya, hasilnya adalah sebagai berikut :

Tabel 3. Koefisien Model Logit Data HBAT

Perbandingan Training & Testing	Koefisien			
	Konstan	X_6	X_9	X_{12}
70:30	-19.883	0.774	1.257	1.253
60:40	-20.177	0.981	1.479	0.756
50:50	-21.666	0.967	1.994	0.047
40:60	-17.265	0.751	1.220	0.873

Dengan menggunakan *software* SPSS dan MINITAB diperoleh ketepatan klasifikasi untuk masing-masing perbandingan *training* dan *testing* adalah sebagai berikut :

Tabel 4. Ketepatan Klasifikasi Analisis Regresi Logistik Data HBAT

Perbandingan Training & Testing	Ketepatan Klasifikasi	
	Training(%)	Testing(%)
80:20	81.3	80
70:30	85.7	76.7
60:40	78.3	75
50:50	78	78
40:60	75	85

Neural Network

Dengan menggunakan software R diperoleh ketepatan klasifikasi untuk masing-masing perbandingan training dan testing adalah sebagai berikut :

Tabel 5. Ketepatan Klasifikasi Neural Network Data HBAT

Perbandingan Training & Testing	Ketepatan Klasifikasi	
	Training(%)	Testing(%)
80:20	80	85
70:30	82.85	83.3
60:40	85	77.5
50:50	82	80
40:60	67.5	83.3

Multivariate Adaptive Regression Spline

Dengan menggunakan software MARS diperoleh ketepatan klasifikasi untuk masing-masing perbandingan training dan testing adalah sebagai berikut :

Tabel 6. Ketepatan Klasifikasi MARS Data HBAT

Perbandingan Training & Testing	Ketepatan Klasifikasi	
	Training(%)	Testing(%)
80:20	78.75	60
70:30	78.57	76.67
60:40	81.67	67.5
50:50	72	44
40:60	62.5	58.33

Analisis Diskriminan, Analisis Regresi Logistik, MARS dan NN Pada Data IRIS

Analisis Diskriminan

Analisis akan dilakukan untuk perbandingan training dan testing adalah 80:20. Karena variabel respon terdiri dari 3 kategori maka dengan menggunakan software SPSS diperoleh 2 fungsi diskriminan dengan semua variabel prediktor berpengaruh terhadap variabel respon, yaitu :

$$\hat{y}_1 = -0.361X_1 - 0.578X_2 + 0.893X_3 + 0.601X_4$$

$$\hat{y}_2 = 0.229X_1 + 0.558X_2 - 0.544X_3 + 0.681X_4$$

Dari fungsi di atas terlihat bahwa variabel X_4 memberikan kontribusi terbesar terhadap variabel respon karena memiliki koefisien yang paling besar.

Untuk perbandingan training dan testing lainnya, hasilnya adalah sebagai berikut:

Tabel 7. Koefisien Fungsi Diskriminan IRIS

Perbandingan Training & Testing	Koefisien			
	X_1	X_2	X_3	X_4
70:30	-0.643	-0.463	1.019	0.576
	-0.165	0.921	-0.031	0.294
60:40	-0.457	-0.568	1.003	0.874
	-0.231	0.437	-0.712	0.961
50:50	-0.320	-0.646	0.953	0.528
	0.034	0.630	-0.439	0.699
40:60	-0.619	-0.319	1.023	0.690
	-0.311	1.060	0.045	0.189

Dari hasil yang ditunjukkan dalam Tabel 7 terlihat bahwa untuk semua perbandingan training dan testing, variabel X_4 memberikan kontribusi terbesar.

Dengan menggunakan software SPSS dan MINITAB diperoleh ketepatan klasifikasi untuk masing-masing perbandingan training dan testing adalah sebagai berikut :

Perbandingan Training & Testing	Ketepatan Klasifikasi	
	Training(%)	Testing(%)
80:20	98.3	97,3
70:30	97.1	100
60:40	97.8	98.3
50:50	98.7	97.3
40:60	100	97.8

Analisis Regresi Logistik

Analisis regresi logistik akan dilakukan untuk perbandingan training dan testing adalah 80:20. Dengan menggunakan software SPSS diperoleh 2 model logit yaitu :

$$g_1(x) = 32.491 + 10.002X_1 + 13.420X_2 - 24.857X_3 - 37.679X_4$$

$$g_2(x) = 40.486 + 0.916X_1 + 6.710X_2 - 7.394X_3 - 17.571X_4$$

Untuk perbandingan training dan testing lainnya, hasilnya adalah sebagai berikut :

Tabel 9. Koefisien Model Logit Data IRIS

Perbandingan Training & Testing	Koefisien				
	Konstan	X ₁	X ₂	X ₃	X ₄
70:30	5.844	2.173	28.312	-16.340	-36.266
	33.776	2.274	5.976	-8.357	-14.165
60:40	23.704	4.675	23.106	-21.456	-37.450
	9.357	3.128	10.124	-17.419	-30.238
50:50	136.119	34.258	57.991	-75.859	-99.405
	139.573	34.938	44.240	-67.609	-82.326
40:60	136.936	-3.188	136.820	-46.397	-199.144
	141.886	-12.721	134.898	-30.398	-187.903

Dengan menggunakan software SPSS dan MINITAB diperoleh ketepatan klasifikasi untuk masing-masing perbandingan training dan testing adalah sebagai berikut :

Tabel 10. Ketepatan Klasifikasi Analisis Regresi Logistik Data IRIS

Perbandingan Training & Testing	Ketepatan Klasifikasi	
	Training(%)	Testing(%)
80:20	98.3	97.5
70:30	98.1	97.8
60:40	97.6	95
50:50	100	93.3
40:60	100	96.3

Neural Network

Dengan menggunakan software R diperoleh ketepatan klasifikasi untuk masing-masing perbandingan training dan testing adalah sebagai berikut :

Tabel 11. Ketepatan Klasifikasi Neural Network Data IRIS

Perbandingan Training & Testing	Ketepatan Klasifikasi	
	Training(%)	Testing(%)
80:20	97.6	96.7
70:30	99	97.8
60:40	95	90.3
50:50	100	93.3
40:60	98.3	97.7

Multivariate Adaptive Regression Spline

Dengan menggunakan software R diperoleh ketepatan klasifikasi untuk masing-masing perbandingan training dan testing adalah sebagai berikut :

Tabel 12. Ketepatan Klasifikasi MARS Data IRIS

Perbandingan Training & Testing	Ketepatan Klasifikasi	
	Training(%)	Testing(%)
80:20	95	92.5
70:30	94.3	95.5
60:40	93.3	85
50:50	98.7	90
40:60	90.5	86.7

Perbandingan Hasil Klasifikasi

Berdasarkan analisis data yang telah dilakukan pada data HBAT maka ketepatan klasifikasi antara analisis diskriminan, regresi logistik, NN dan MARS diberikan pada Tabel 13. Dari Tabel 13 dapat diketahui bahwa perbedaan hasil klasifikasi antara analisis diskriminan, regresi logistik, NN dan MARS berbeda untuk setiap perbandingan. Bila dilihat dari nilai ketepatan klasifikasi pada *Training*, regresi logistik mempunyai ketepatan yang lebih tinggi untuk perbandingan 80:20, 70:30 dan 40:60. sedangkan untuk perbandingan 60:40 dan 50:50 analisis diskriminan dan NN mempunyai ketepatan yang lebih tinggi. Sedangkan bila didasarkan pada *Testing*, ketepatan klasifikasi analisis diskriminan lebih tinggi untuk perbandingan 60:40, 40:60 dan 50:50. sedangkan NN memiliki ketepatan klasifikasi lebih tinggi untuk perbandingan 60:40, 80:20 dan 70:30. Ketepatan klasifikasi pada *Testing* dijadikan dasar penentuan metode yang mempunyai ketepatan yang lebih tinggi. Hal ini dikarenakan *Testing* merupakan pengujian fungsi klasifikasi yang telah diperoleh dari *Training* sehingga tinggi atau rendah ketepatan pada *Testing* menunjukkan ketepatan fungsi klasifikasi yang dibentuk. Oleh karena itu, metode yang lebih tepat dipergunakan dalam data HBAT adalah analisis diskriminan dan NN karena nilai ketepatan klasifikasi yang dihasilkan lebih tinggi khususnya untuk *Testing*.

Sedangkan berdasarkan analisis data yang telah dilakukan pada data IRIS maka ketepatan klasifikasi antara analisis diskriminan, regresi logistik, NN dan MARS diberikan pada Tabel 14. Dari Tabel 14 dapat diketahui bahwa perbedaan hasil klasifikasi antara analisis diskriminan, regresi logistik, NN dan MARS berbeda untuk setiap perbandingan. Bila dilihat dari nilai ketepatan klasifikasi pada *Training*, analisis diskriminan dan regresi logistik mempunyai ketepatan yang lebih tinggi untuk perbandingan untuk beberapa perbandingan. Sedangkan bila didasarkan pada *Testing*, ketepatan klasifikasi analisis diskriminan lebih tinggi untuk semua perbandingan. Ketepatan klasifikasi pada *Testing* dijadikan dasar penentuan metode yang mempunyai ketepatan yang lebih tinggi. Hal ini dikarenakan *Testing* merupakan pengujian fungsi klasifikasi yang telah diperoleh dari *Training* sehingga tinggi atau rendah ketepatan pada *Testing* menunjukkan ketepatan fungsi klasifikasi yang dibentuk. Oleh karena itu, metode yang lebih tepat dipergunakan dalam data IRIS adalah analisis diskriminan karena nilai ketepatan klasifikasi yang dihasilkan lebih tinggi khususnya untuk *Testing*.

KESIMPULAN

Berdasarkan pembahasan maka kesimpulan dalam penelitian ini adalah:

1. Metode yang lebih tepat digunakan dalam klasifikasi data HBAT adalah analisis diskriminan dan NN karena nilai ketepatan klasifikasi yang dihasilkan lebih besar khususnya untuk *Testing*. Variabel yang signifikan adalah kualitas produk (X_6), *Complaint resolution* (X_9) dan *salesforce image* (X_{12})
2. Metode yang lebih tepat digunakan dalam klasifikasi data IRIS adalah analisis diskriminan karena nilai ketepatan klasifikasi yang dihasilkan lebih besar khususnya untuk *Testing*. Semua variabel prediktor berpengaruh terhadap variabel respon

DAFTAR PUSTAKA

- Agresti, A. (1990). *Categorical Data Analysis*. NY: John Wiley and Sons, Inc
- Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C. J. (1993). *Classification and Regression Trees*. CA: Wadsworth
- Chikolwa, B. (2007). An Empirical Analysis of Commercial Mortgage-Backed Securities Credit Rating. *Australian Evidence, 13th Pacific-Rim Real Estate Society Conference Premantle*. Western Australia
- Dillon, W.R., & Goldstein, M. (1984). *Multivariate Analysis Methods and Application*. NY: John Wiley and Sons, Inc
- Demaris, A. (1995). A Tutorial in Logistic Regression. *Journal of Marriage and the Family, 57*:956-968
- Friedman, J.H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, Vol. 19, No. 1, 1-67
- Hawkins, D.M., & McLachlan, G.J. (1997). High-Breakdown Linear Discriminant Analysis. *Journal of the American Statistical Association*, Vol. 92, No. 437, 136-143
- Hermawan, A. (2006). *Jaringan Syaraf Tiruan Teori dan Aplikasinya*. Yogyakarta: Andi
- Hosmer, D.W., & Lemeshow, S. (1989). *Applied Logistic Regression*. NY: John Wiley and Sons, Inc
- Huang, Z., Chen, H., Hsu, C.J., Chen, W.H., & Wu, S. (2003). Credit Analysis With Support Vector Machines and Neural Network : a Market Comparative Study, *Decision Support System 37. Elsevier*, 543-558
- Huberty, C. J. (1984). Issues in the Use and Interpretation of Discriminant Analysis. *Psychological Bulletin*, 95:156-171
- Iriawan, N., & Astuti, S.P. (2006). *Mengolah Data Statistik Dengan Mudah Menggunakan MINITAB 14*. Yogyakarta: Andi
- Johnson, R.A., & Wichern, D.W. (1992). *Applied Multivariate Statistical Analysis*. NJ: Prentice-Hall
- Le, C.T. (1998). *Applied Categorical data Analysis*. New York: John Wiley and Sons, Inc
- Morrison, D.G. (1969). On the Interpretation of Discriminant Analysis. *Journal of Marketing Research*, 6(2):156-163
- Sharma, S. (1996). *Applied Multivariate Techniques*. NY: John Wiley and Sons, Inc