

Pemilihan Model Regresi Linier Multivariat Terbaik Dengan Kriteria *Mean Square Error* Dan *Akaike's Information Criterion*

Edriani Lestari¹, Rito Goejantoro², Memi Nor Hayati³

¹Laboratorium Statistika Terapan Program Studi Statistika, FMIPA, Universitas Mulawarman

²Program Studi Statistika FMIPA, Universitas Mulawarman

Corresponding Author: ritogoejantoro@yahoo.com

Abstrak Regresi linier multivariat merupakan salah satu metode analisis regresi yang melibatkan lebih dari satu variabel respon. Dalam penelitian ini dilakukan pemilihan model regresi Linier multivariat terbaik dengan kriteria *Mean Square Error* (*MSE*) dan *Akaike's information Criterion* (*AIC*) yang bertujuan untuk mengetahui faktor-faktor yang berpengaruh terhadap kualitas kertas di perusahaan PT. ITCI HUTANI MANUNGGAL pada tahun 2013-2014. Uji signifikansi parameter diperoleh *flow bottom* (X_1) dan *speed* (X_4) berpengaruh terhadap *basis weight* (Y_1) sedangkan *flow bottom* (X_1) dan *flow middle* (X_2) berpengaruh terhadap *caliper* (Y_2). Pada pemilihan model terbaik dengan kriteria *MSE* dan *AIC* diperoleh model terbaik $\hat{Y}_1 = 37,41 + 0,152X_1 + 0,691X_4$ dan

$$\hat{Y}_2 = 70,85 + 0,365X_1 + 0,278X_2$$

Kata-kata kunci *AIC*, Kualitas Kertas, *MSE*, Regresi Linier Multivariat

Pendahuluan

Analisis regresi linier multivariat merupakan bagian dari analisis regresi yang melibatkan tidak hanya satu variabel respon namun beberapa variabel respon. Pada model regresi linier multivariat, matriks Y maupun matriks ϵ diasumsikan berdistribusi normal multivariat [1]. *MSE* bisa dibilang kriteria yang paling penting yang digunakan untuk mengevaluasi kinerja prediktor atau estimator. *MSE* juga berguna untuk menyampaikan konsep bias, presisi, dan ketepatan dalam estimasi statistik. *AIC* merupakan kriteria *fit indices*. *AIC* digunakan dalam perbandingan model dimana nilai jumlah kegagalan model hanya dibandingkan dengan dengan nilai *AIC* dependen dan independen model. Selain itu *AIC* digunakan sebagai tolak ukur jumlah sampel yang kecil [2].

[3] pemilihan model terbaik pada regresi multivariat dengan menggunakan *Modified Akaike's Information Criterion* (*MAIC*) dan *Modified Cp Mallow's* (*MCP*). [4] model regresi linier multivariat untuk menentukan tingkat kesejahteraan kabupaten dan kota di Jawa Timur. [5] pemilihan model regresi linear multivariat terbaik dengan menggunakan kriteria *Mean Square Error* (*MSE*).

Metode analisis regresi dapat diterapkan untuk membuat kualitas produk dalam penelitian ini dengan menggunakan variabel prediktor berupa sistem kontrol proses pembuatan kertas. Terdapat korelasi antara

variabel yang digunakan sebagai kontrol dalam proses pembuatan kertas dengan spesifikasi produk yang akan dibuat. Kontrol tersebut berupa volume buburan pembuat kertas (lt/min) yang disebut *flow* dan kecepatan mesin *wire* (m/min) yang dikenal dengan *speed*. Selanjutnya dalam menentukan variabel respon, oleh karena spesifikasi produk memberi kemungkinan untuk dijadikan beberapa variabel respon, maka regresi multivariat dapat diterapkan dalam penelitian ini. PT. IHM melakukan pengendalian pada gramatur (*basis weight*) dan ketebalan (*caliper*) karena dianggap mampu mewakili sifat kertas lainnya. Sehingga pada penelitian ini, akan digunakan *basis weight* dan *caliper* sebagai variabel respon.

Tujuan penelitian ini adalah untuk mengetahui korelasi antar variabel respon, mengetahui kombinasi model yang berpengaruh, dan memperoleh model regresi linier multivariat terbaik.

Variabel penelitian

Variabel respon (terikat) dalam penelitian ini adalah:

Y_1 : *Basis Weight* (Berat Dasar) (gr/m²)

Y_2 : *Caliper* (Ketebalan) (μ m)

Variabel prediktor (bebas) dalam penelitian ini adalah :

X_1 : *Flow Bottom* (Aliran Bawah) (lt/min)

X_2 : *Flow Middle* (Aliran Tengah) (lt/min)

X_3 : *Flow Top* (Aliran Atas) (lt/min)

X_4 : *Speed* (Kecepatan) (m/min)

Metode Penelitian

Adapun teknik analisa data yang digunakan dalam penelitian ini adalah:

1. Mendeskripsikan variabel respon dan variabel prediktor.
2. Menguji korelasi antar variabel respon.
3. Mengestimasi parameter.
4. Pengujian signifikansi parameter:
 - a. Uji Simultan (Uji F)
 - b. Uji Parsial (Uji T)
5. Pengujian asumsi regresi linier multivariat:
 - a. Uji Kenormalan
 - b. Uji Kesamaan Matriks Varian dan Kovarian
6. Pemilihan model terbaik dengan kriteria *MSE* dan *AIC*.

Hasil dan Pembahasan

Tabel 1. Statistika Deskriptif pada data karakteristik produk kertas di Perusahaan PT. Itci Hutani Manunggal (IHM) pada tahun 2013-2014.

	n	Min	Maks	Rata-rata
Y_1	24	81,3	416,2	216,63
Y_2	24	67,6	393	214,79
X_1	24	82,9	442,2	233,33
X_2	24	96,1	422,2	210,98
X_3	24	6,8	29,81	19,05
X_4	24	33,8	393,9	194,9

Berdasarkan Tabel 1 dilihat bahwa pada *Basis Weight* (Y_1) data berjumlah 24, memiliki nilai minimum 81,3, nilai maksimum 416,2 dan rata-rata sebesar 216,63. Untuk *Caliper* (Y_2) data berjumlah 24, memiliki nilai minimum 67,6, nilai maksimum 393,0 dan rata-rata sebesar 214,79. Untuk *Flow Bottom* (X_1) data berjumlah 24, memiliki nilai minimum 82,9, nilai maksimum 442,2 dan rata-rata sebesar 233,33. Untuk *Flow Middle* (X_2) data berjumlah 24, memiliki nilai minimum 96,1, nilai maksimum 422,2 dan rata-rata sebesar 210,98. Untuk *Flow Top* (X_3) data berjumlah 24, memiliki nilai minimum 6,8, nilai maksimum 29,81 dan rata-rata sebesar 19,05. Untuk *Speed* (X_4) data berjumlah 24, memiliki nilai minimum 33,8, nilai maksimum 393,9 dan rata-rata sebesar 194,9.

Korelasi Variabel Respon

Pada tahap ini dilakukan pengujian untuk mengetahui apakah terdapat korelasi antar variabel respon.

Tabel 2. Korelasi Variabel Respon

	Y1	Y2
Korelasi	1	0,425
<i>P-Value</i>		0,039

Dapat dilihat pada Tabel 2 diperoleh *p-value* = 0,039 < α = 0,05 maka diputuskan menolak H_0 sehingga dapat disimpulkan bahwa antar variabel respon bersifat dependen, jadi data dapat digunakan pada analisis regresi linier multivariat karena variabel respon yaitu Y_1 adalah *Basis Weight* (gr/m^2) dan Y_2 adalah *Serta Caliper* (μm) bersifat dependen atau saling berkorelasi.

Estimasi Parameter

Setelah dilakukan pengujian korelasi antar variabel respon selanjutnya adalah mengestimasi parameter menggunakan *Least Square*. Nilai hasil estimasi parameter dapat dilihat pada Tabel 2.

Tabel 3. Estimasi Parameter

Variabel Respon	Variabel	β
Y_1	Konstan	18,816
	X_1	0,170
	X_2	0,155
	X_3	-0,080
	X_4	0,652
Y_2	Konstan	69,548
	X_1	0,348
	X_2	0,254
	X_3	-0,782
	X_4	0,129

Untuk model terbaik regresi linier multivariat untuk variabel respon *basic weight* (Y_1) dan *Caliper* (Y_2) adalah sebagai berikut:

$$\hat{Y}_1 = 18,816 + 0,170X_1 + 0,155X_2 - 0,080X_3 + 0,652X_4$$

$$\hat{Y}_2 = 69,548 + 0,348X_1 + 0,254X_2 - 0,782X_3 + 0,129X_4$$

Interpretasi model:

1. Untuk setiap penambahan satu l/min *flow bottom* maka akan menambahkan *basis weight* sebesar 0,170 (lt/min). Setiap penambahan satu l/min *flow middle* maka

akan menambahkan *basis weight* sebesar 0,155 (lt/min). Untuk setiap pengurangan satu l/min *flow top* maka akan mengurangi *basis weight* sebesar -0,080 (lt/min). Kemudian setiap penambahan satu m/min *speed* maka menambahkan *basis weight* sebesar 0,652 (m/min).

- Untuk setiap penambahan satu l/min *flow bottom* maka akan menambahkan *caliper* sebesar 0,348 (lt/min). Setiap penambahan satu l/min *flow middle* maka akan menambahkan *caliper* sebesar 0,254 (lt/min). Untuk setiap pengurangan satu l/min *flow top* maka akan mengurangi *caliper* sebesar -0,782 (lt/min). Kemudian setiap penambahan satu m/min *speed* maka menambahkan *caliper* sebesar 0,129 (m/min).

Penaksiran Parameter

1. Uji simultan (Uji F)

Menurut Hines dan Montgomery (1990) [6], pengujian terhadap kesesuaian model regresi yang dihasilkan disebut dengan uji simultan model regresi linear. Statistik uji yang digunakan adalah Uji F.

Tabel 4. Uji simultan (Uji F)

Var	F	P-value
Y_1	3,823	0,019
Y_2	2,972	0,046

Dari Tabel 4 dapat dilihat variabel Y_1 dan Y_2 memiliki nilai masing-masing *p-value* = 0,019 dan 0,0046 < α = 0,05 maka diputuskan untuk menolak H_0 sehingga dapat disimpulkan bahwa secara simultan minimal ada satu variabel bebas (*flow bottom*, *flow middle*, *flow top*, *speed*) yang berpengaruh terhadap *basis weight* dan *caliper*.

2. Uji Parsial (Uji T)

Uji t digunakan untuk mengetahui keberartian dari masing-masing penduga parameter secara parsial, apakah koefisien parsial yang diperoleh tersebut mempunyai pengaruh atau tidak dengan asumsi bahwa variabel tidak bebas lainnya konstan.

Tabel 5. Uji Parsial (Uji T)

Variabel	F	P-value
Konstan	0,362	0,701
X_1	2,674	0,096
X_2	1,325	0,291
X_3	0,044	0,957

X_4	4,934	0,020
-------	-------	-------

Pada Tabel 5 diperoleh nilai *p-value* = 0,096, 0,291, dan 0,0,957 untuk masing-masing variabel X_1 , X_2 , dan X_3 > α = 0,05 maka diputuskan untuk menerima H_0 sehingga dapat disimpulkan bahwa tidak ada pengaruh *flow bottom*, *flow middle*, dan *flow top* terhadap *basis weight* dan *caliper*. Sedangkan untuk variabel X_4 *p-value* = 0,020 < α = 0,05 maka diputuskan untuk menolak H_0 sehingga dapat disimpulkan bahwa *speed* berpengaruh terhadap *basis weight* dan *caliper*. Karena masih ada variabel yang tidak berpengaruh maka dilakukan pengujian ulang dengan mengeluarkan variabel yang tidak berpengaruh.

Tabel 6. Pengujian Ulang Hanya Pada Variabel Yang Berpengaruh

Variabel	F	P-value
Konstan	6,973	0,005
X_4	6,149	0,008

Pada tabel 6 diperoleh nilai *p-value* = 0,008 < α = 0,05 maka diputuskan untuk menolak H_0 sehingga dapat disimpulkan bahwa *speed* berpengaruh terhadap *basis weight* dan *caliper*. Karena setelah pengujian ulang dan tidak ada lagi variabel yang berpengaruh sehingga dapat dilanjutkan ke analisis selanjutnya.

Asumsi Regresi Multivariat

1. Uji Normal

Pengujian asumsi residual analisis regresi multivariat terpenuhi jika plot yang dihasilkan cenderung membentuk pola garis lurus. Selain itu, pada Tabel *Tests of Normality*.

Tabel 7. Uji Kenormalan

	P-value
Y_1	0,115
Y_2	0,714

Pada Tabel 7 dapat dilihat nilai untuk masing-masing variabel Y_1 dan Y_2 *p-value* = 0,115 dan 0,714 > α = 0,05 maka diputuskan untuk menerima H_0 sehingga dapat disimpulkan bahwa residual data yang digunakan berdistribusi normal dan dapat dilanjutkan ke uji asumsi selanjutnya.

2. Uji Kesamaan Variansi

Pada Tabel *Levene's Test of Equality of Error Variances* yang terdapat di lampiran, maka hasil analisis sebagai berikut:

Tabel 8. Uji Kesamaan Variansi

Variabel	F	P-value
Y ₁	6,973	0,645
Y ₂	6,149	0,635

Pada Tabel 8 dapat dilihat nilai untuk masing-masing variabel Y₁ dan Y₂ *p-value* = 0,645 dan 0,635 > α = 0,05 maka diputuskan untuk menerima H₀ sehingga dapat disimpulkan bahwa variansi data hasil produksi kayu *caliper* pada *speed* adalah sama.

3. Uji Kesamaan Kovariansi

Pada Tabel *Box's Test of Equality of Covariance Matrices* yang terdapat di lampiran, maka hasil analisis sebagai berikut:

Tabel 9. Uji Kesamaan Variansi

Box's M	F	P-Value
12,505	3,758	0,01

Pada Tabel 9 dapat dilihat nilai untuk masing-masing variabel Y₁ dan Y₂ *p-value* = 0,645 dan 0,635 > α = 0,05 maka diputuskan untuk menerima H₀ sehingga dapat disimpulkan bahwa kovariansi variabel dependen pada setiap grup adalah sama.

Pemilihan Model Terbaik

1. Dengan Kriteria MSE

Prosedur pemilihan model terbaik dengan kriteria *MSE* adalah dengan menregresikan seluruh variabel X terhadap variabel respon. Oleh karena itu, dapat dikatakan bahwa model regresi linear multivariat terbaik adalah sebagai berikut:

Tabel 10. Menentukan Nilai MSE

Terhadap Y1		Terhadap Y2	
Prediktor	MSE	Prediktor	MSE
X ₁	9832,89	X ₁	5786,86
X ₂	9470,81	X ₂	6620,03
X ₃	10435,09	X ₃	7554,18
X ₄	6518,35	X ₄	7427,6
X ₁ ,X ₂	9448,43	X ₁ ,X ₂	5219,56
X ₁ ,X ₃	10298,71	X ₁ ,X ₃	6040,08
X ₁ ,X ₄	6350,27	X ₁ ,X ₄	5749,49
X ₂ ,X ₃	9858,95	X ₂ ,X ₃	6656,81

Terhadap Y1		Terhadap Y2	
Prediktor	MSE	Prediktor	MSE
X ₂ ,X ₄	6476,72	X ₂ ,X ₄	6758,9
X ₃ ,X ₄	6788,94	X ₃ ,X ₄	7496,41
X ₁ ,X ₂ ,X ₃	9917,2	X ₁ ,X ₂ ,X ₃	5452
X ₁ ,X ₂ ,X ₄	6406	X ₁ ,X ₂ ,X ₄	5339,14
X ₁ ,X ₃ ,X ₄	6667,75	X ₁ ,X ₃ ,X ₄	6016,74
X ₂ ,X ₃ ,X ₄	6763,55	X ₂ ,X ₃ ,X ₄	6819,75
X ₁ ,X ₂ ,X ₃ ,X ₄	6742,86	X ₁ ,X ₂ ,X ₃ ,X ₄	5593,15

Dari Tabel 10 dapat kita lihat model terbaik regresi linier multivariat terbaik dengan kriteria *MSE* adalah sebagai berikut:

$$\hat{Y}_1 = 37,41 + 0,152X_1 + 0,691X_4$$

$$\hat{Y}_2 = 70,85 + 0,365X_1 + 0,278X_2$$

Interpretasi Model:

1. Untuk setiap penambahan satu l/min *flow bottom* maka akan menambahkan *basis weight* sebesar 0,152 (lt/min). Kemudian setiap penambahan satu m/min *speed* maka menambahkan *basis weight* sebesar 0,691 (m/min).
2. Untuk setiap penambahan satu l/min *flow bottom* maka akan menambahkan *caliper* sebesar 0,365 (lt/min). Setiap penambahan satu l/min *flow middle* maka akan menambahkan *caliper* sebesar 0,278 (lt/min).

3. Dengan Kriteria AIC

Pemilihan model terbaik dengan menggunakan *AIC* adalah dengan meregresikan seluruh variabel prediktor (X) terhadap variabel respon (Y). Untuk memperoleh nilai *AIC* terlebih dahulu harus mengetahui nilai dari rata-rata kuadrat error. Nilai rata-rata kuadrat error:

Tabel 11. Menentukan Nilai AIC

Terhadap Y1		Terhadap Y2	
Prediktor	AIC	Prediktor	AIC
X ₁	220,55	X ₁	207,83
X ₂	219,66	X ₂	211,06
X ₃	221,98	X ₃	214,23
X ₄	218,48	X ₄	213,82
X ₁ ,X ₂	212,69	X ₁ ,X ₂	206,24
X ₁ ,X ₃	222,55	X ₁ ,X ₃	209,74
X ₁ ,X ₄	210,95	X ₁ ,X ₄	208,56

ISBN: 978-602-72658-1-3

Terhadap Y1		Terhadap Y2	
Prediktor	AIC	Prediktor	AIC
X_2, X_3	221,5	X_2, X_3	212,08
X_2, X_4	211,42	X_2, X_4	212,44
X_3, X_4	212,55	X_3, X_4	214,93
X_1, X_2, X_3	222,47	X_1, X_2, X_3	208,11
X_1, X_2, X_4	211,98	X_1, X_2, X_4	207,61
X_1, X_3, X_4	212,95	X_1, X_3, X_4	210,48
X_2, X_3, X_4	213,29	X_2, X_3, X_4	213,49
X_1, X_2, X_3, X_4	213,98	X_1, X_2, X_3, X_4	20950

Dari Tabel 11 dapat kita lihat model terbaik regresi linier multivariat terbaik dengan kriteria AIC adalah sebagai berikut:

$$\hat{Y}_1 = 37,41 + 0,152X_1 + 0,691X_4$$

$$\hat{Y}_2 = 70,85 + 0,365X_1 + 0,278X_2$$

Interpretasi Model:

1. Untuk setiap penambahan satu l/min *flow bottom* maka akan menambahkan *basis weight* sebesar 0,152 (lt/min). Kemudian setiap penambahan satu m/min *speed* maka menambahkan *basis weight* sebesar 0,691 (m/min).
2. Untuk setiap penambahan satu l/min *flow bottom* maka akan menambahkan *caliper* sebesar 0,365 (lt/min). Setiap penambahan satu l/min *flow middle* maka akan menambahkan *caliper* sebesar 0,278 (lt/min).

Kesimpulan

Berdasarkan analisis yang telah dilakukan sebelumnya diperoleh kesimpulan sebagai berikut :

1. Dalam pengujian antar variabel respon bersifat *dependent* dan memiliki

hubungan yang cukup kuat karena mempunyai nilai korelasi sebesar 0,425.

2. Dari hasil pemilihan model *MSE* dan *AIC* diperoleh faktor-faktor yang berpengaruh terhadap kualitas kertas *basis weight flow bottom* dan *speed* sedangkan yang berpengaruh terhadap *caliper* adalah *flow bottom* dan *flow middle*.

Model regresi linear multivariat terbaik yang diperoleh untuk variabel respon *basis weight* (Y_1) dan *Caliper* (Y_2) adalah

$$\hat{Y}_1 = 37,41 + 0,152X_1 + 0,691X_4$$

$$\hat{Y}_2 = 70,85 + 0,365X_1 + 0,278X_2$$

Daftar Pustaka

- [1] Timm, N.H. 2002. *Applied Multivariate Analysis*. New York: Springer-Verlag New York Inc.
- [2] Hengky, Latan, 2012. *Structural Equation Modeling Konsep dan Aplikasi*. Bandung: Alfabeta.
- [3] Saleh, Salmin. 2010. *Pemilihan Model Terbaik Pada Regresi Multivariat Menggunakan MAIC dan MCP*. Surabaya.
- [4] Fadillah. Faris. 2013. *Model Regresi Multivariat Untuk Menentukan Tingkat Kesejahteraan Kabupaten dan Kota Di Jawa Timur*. Surabaya.
- [5] Aminuddin. 2013. *Pemilihan Model Regresi Linear Multivariat Terbaik Dengan Kriteria Mean Square Error*. Jurnal Gaussian, Volume 2 Nomor 1 Halaman 11-18.
- [6] Hines, W.W. dan Montgomery, D.C. 1990. *Probabilitas dan Statistik Dalam Ilmu Rekayasa dan Manajemen Edisi Kedua Terjemahan Rudiansyah*. UI. Press Jakarta.