

# METODE PLS UNTUK MENGATASI KOLINEARITAS DALAM KALIBRASI GANDA

Aji Hamim Wigena<sup>1</sup> dan Aunuddin<sup>1</sup>

## RINGKASAN

*Metode PLS (Partial Least Square), yang termasuk ke dalam pemodelan 'lunak' (soft modelling), digunakan untuk mengatasi masalah kolinearitas dalam model kalibrasi ganda. Model kalibrasi dalam makalah ini digunakan untuk menduga kandungan protein dalam gandum. Hasil pendugaan metode PLS dibandingkan dengan hasil metode Bayes bersyarat yang dikembangkan oleh du Plessis & van der Merwe (1995) dan metode klasik Brown (1982).*

## PENDAHULUAN

Dalam pendugaan model regresi ganda, kolinearitas merupakan masalah yang akan mengakibatkan penduga modelnya kurang tepat. Masalah ini sering terjadi pada model kalibrasi ganda (*multivariate calibration*) sehingga diperlukan suatu metode untuk pendugaan model kalibrasi dengan mengatasi kolinearitas. Metode ini diperlukan dalam pengembangan instrumen analitik modern (du Plessis & van der Merwe, 1995). Disamping itu Martens dan Naes (1989) menyatakan bahwa model kalibrasi ganda dapat juga meningkatkan kapasitas analitik dan keterandalan instrumen lama sehingga hal ini akan memperluas analisis kimia kuantitatif dalam pengendalian proses industri dan pemantauan polusi secara on-line.

Metode regresi *ridge* dan komponen utama dapat digunakan untuk mengatasi kolinearitas (Myers, 1990). Martens & Naes (1989) dan Young (1994) mengemukakan bahwa metode regresi *ridge*, komponen utama, NIPALS dan PLS dapat digunakan untuk pendugaan model kalibrasi. Herwindiati (1997) menyimpulkan bahwa metode PLS menghasilkan pendugaan kandungan lemak ikan lebih baik dari pada metode regresi *ridge* dan komponen utama. Dengan data yang sama, hasil metode PLS juga lebih baik dari metode NIPALS (Wigena & Aunuddin, 1997). Disamping itu, du Plessis dan van der Merwe (1995) menunjukkan bahwa metode Bayes bersyarat lebih baik dari

metode klasik Brown (1982) dalam menduga kandungan protein dalam gandum.

Dalam makalah ini dibahas tentang metode PLS untuk pendugaan kandungan protein dalam gandum dan pembandingannya dengan metode Bayes bersyarat dan tak bersyarat yang dikembangkan oleh du Plessis dan van der Merwe (1995) dan metode klasik Brown (1982).

## MODEL KALIBRASI DAN KOLINEARITAS

Secara umum kalibrasi menggunakan suatu fungsi matematik dengan data empirik dan pengetahuan untuk menduga informasi pada  $y$  yang tidak diketahui berdasarkan informasi pada  $X$  yang tersedia (Martens dan Naes, 1989). Dalam bidang kimia, model kalibrasi merupakan suatu fungsi hubungan antara absorbansi ( $X$ ) dari spektrometer dengan konsentrasi ( $y$ ) larutan unsur atau senyawa yang akan dianalisis (Nur dan Adijuwana, 1989). Dengan kalibrasi, konsentrasi larutan contoh dapat diketahui berdasarkan absorbansinya.

Pendugaan model kalibrasi tergantung pada jenis spektrometer yang digunakan. Spektrometer UV-VIS menghasilkan spektrum yang berbentuk satu puncak absorpsi, sehingga model kalibrasinya berupa garis lurus. Spektrometer NIR (*near infra red*) menghasilkan spektrum dengan banyak puncak absorpsi, sehingga akan terbentuk suatu model

<sup>1</sup> Dosen pada Jurusan Statistika FMIPA-IPB

kalibrasi peubah ganda. Model kalibrasi suatu senyawa lebih tepat menggunakan spektrum dengan banyak puncak daripada satu puncak absorpsi (Nur dan Adijuwana, 1989). Tetapi dalam pendugaan model kalibrasi ganda sering timbul masalah kolinearitas di antara peubah absorban (Naes, 1985), sehingga sering tidak memberikan solusi yang unik sedangkan pemilihan suatu absorban dari sejumlah absorban tidak mudah dilakukan.

Metode PLS merupakan metode pemodelan 'lunak' yang dapat mengoptimalkan model kalibrasi peubah  $y$  dan  $X$ . Model ini berdasarkan peubah internal atau laten (*latent variable*). Model hubungan tersebut adalah (Anonim, 1982):

### 1. Hubungan internal

$$\eta = \beta_0 + \beta_1 \xi + \vartheta, \quad \dots (1)$$

dengan  $\beta_0$  dan  $\beta_1$  adalah koefisien model internal, dan  $\vartheta$  adalah faktor acak.

### 2. Hubungan eksternal

$$x = \pi_0 + \pi_1 \eta + \varepsilon$$

$$y = (\pi_0 + \pi_1 \beta_0) + \pi_1 \beta_1 \xi + (\varepsilon + \pi_1 \vartheta)$$

dengan  $\pi_0$  dan  $\pi_1$  adalah koefisien model eksternal bagi  $x$ , dan  $\varepsilon$  adalah faktor acak.

Hubungan internal yang digunakan oleh du Plessis dan van der Merwe (1995) serupa dengan (1) tetapi mengasumsikan bahwa  $\vartheta=0$ , sedangkan hubungan eksternalnya adalah sebagai berikut.

$$x = \xi + \delta$$

$$y = \eta + \varepsilon$$

dengan mengasumsikan bahwa  $\delta = 0$ .

## ALGORITMA PLS

Metode PLS digunakan untuk memodelkan hubungan antara peubah tak bebas  $y$  dengan satu kelompok peubah bebas  $X$  berdasarkan model (1) (Anonim, 1982). Algoritmanya disusun berdasarkan algoritma NIPALS yang menggunakan metode komponen utama dengan cara dekomposisi nilai singular (*singular-value decomposition*). Modelnya adalah sebagai berikut (Martens dan Naes, 1989):

$$X = t_1 p'_1 + t_2 p'_2 + \dots + t_A p'_A + E_A$$

dengan  $t_a$  adalah vektor berdimensi  $n$ , sebagai vektor skor (*score vector*), sedangkan  $p_a$  adalah vektor berdimensi  $k$ , sebagai vektor muatan (*loading vector*). Matriks  $E_A$  berukuran  $n \times k$  adalah matriks sisaan. Dasar metode PLS adalah bahwa hubungan antara  $X$  dan  $y$  melalui vektor-vektor tersebut, sehingga

$$y = t_1 q_1 + t_2 q_2 + \dots + t_A q_A + f_A$$

dengan  $q_a$  berupa skalar muatan dan  $f_A$  adalah vektor sisaan.

Pendugaan model hubungan  $y$  dengan  $X$  dan pendugaan nilai  $y$  tertentu menggunakan suatu algoritma. Proses penentuan model dilakukan secara iterasi di mana struktur ragam dalam  $y$  mempengaruhi komponen kombinasi linear dalam  $X$ , dan sebaliknya, struktur ragam dalam  $X$  berpengaruh terhadap kombinasi linear dalam  $y$  (Voght, 1990). Tahapan algoritma PLS adalah sebagai berikut (Martens dan Naes, 1989):

1. Ulangi tahap 1.1 s/d 1.5 untuk setiap komponen:
  - 1.1. Tentukan penduga vektor pembobot  $w_a$
  - 1.2. Tentukan penduga vektor skor  $t_a$
  - 1.3. Tentukan penduga vektor muatan  $p_a$
  - 1.4. Tentukan penduga vektor muatan  $q_a$
  - 1.5. Tentukan penduga  $E$  (=residual  $X$ ) dan  $f$  (=residual  $y$ )
2. Tentukan jumlah faktor ( $A$ ) dalam model
3. Tentukan penduga  $y$  berdasarkan  $A$  komponen.

## PENDUGAAN KANDUNGAN PROTEIN

Data yang digunakan untuk pendugaan kandungan protein diperoleh dari Brown (1982), yang digunakan juga oleh du Plessis dan van der Merwe (1995). Ada 21 larutan protein. Kandungan protein ( $y$ ) ditentukan di laboratorium dan absorbannya diukur dengan spektrometer NIR (*Near Infra Red*) pada empat panjang gelombang ( $x_1, x_2, x_3, x_4$ ). Lima belas larutan digunakan untuk menduga model kalibrasi, selainnya digunakan sebagai pembanding hasil dugaan. Data pembanding tercantum pada Tabel 1. Peubah  $x_1$  berkorelasi tinggi dengan  $x_2$  (0.87) dan peubah  $x_3$  berkorelasi tinggi dengan  $x_4$  (0.97). Korelasi antar peubah tercantum pada Tabel 2. Dengan demikian, diantara peubah  $X$  terjadi kolinearitas.

Tabel 1. Pendugaan Kandungan Protein

Data	PLS	Bayes-b*	Bayes-tb*	Brown*
13.57	13.6503	13.5567	13.6875	13.6912
9.26	9.2531	9.2517	9.2517	9.2249
9.82	10.1773	10.1335	10.1673	10.1693
9.46	9.1474	9.1948	9.1215	9.1137
12.85	12.7341	12.6324	12.7535	12.7585
12.81	12.7815	12.8520	12.7993	12.7966
MSE	0.0410	0.0363	0.0433	0.0444

\*Sumber: du Plessis & van der Merwe (1995)

Bayes-b = Bayes bersyarat, Bayes-tb = Bayes tak bersyarat.

Tabel 2. Korelasi Peubah X

	x1	x2	x3
x2	0.87		
x3	0.29	0.42	
x4	0.20	0.24	0.97

Metode PLS menghasilkan model yang dapat digunakan sebagai penduga model kalibrasi kandungan protein ( $R^2=98.7$  dan  $s=0.14$ ). Berdasarkan model tersebut dilakukan pendugaan kandungan protein dan pembandingannya dengan nilai pengamatan, nilai dugaan metode Brown (Brown, 1982) dan metode Bayes bersyarat dan tak bersyarat (du Plessis dan van der Merwe, 1995), seperti tercantum pada Tabel 1. Ketiga nilai dugaan terakhir (Bayes bersyarat, Bayes tak bersyarat dan Brown) diambil dari du Plessis dan van der Merwe (1995). Berdasarkan nilai MSE (*Mean Square Error*) diperoleh bahwa metode Bayes bersyarat (0.0363) lebih baik dari metode PLS (0.0410), dan metode PLS sedikit lebih baik dari metode Bayes tak bersyarat (0.0433) dan metode Brown (0.0444).

Dalam proses pendugaan kandungan protein, metode Bayes bersyarat memperhatikan adanya kandungan air, sedangkan metode PLS hanya melibatkan konsentrasi proteinnya. Metode Bayes menggunakan metode peubah ganda di mana peubah respon y ada dua, yaitu kandungan protein dan air.

## PENUTUP

Metode PLS dapat digunakan untuk pendugaan model kalibrasi peubah ganda dengan satu peubah y dan memberikan hasil pendugaan yang relatif baik dari pada metode Brown, tetapi metode Bayes bersyarat masih lebih baik. Namun demikian metode PLS masih dapat digunakan sebagai metode alternatif untuk mengatasi masalah kolinearitas.

## DAFTAR PUSTAKA

- Anonim. 1982. *Encyclopedia of Statistical Sciences*, vol. VI. John Wiley & Sons. New York.
- du Plessis, J.L. & A.J. van der Merwe. 1995. A Bayesian approach to multivariate and conditional calibration. *Comp. Stat. and Data Analysis*, 19:539-552.
- Herwindiati, D.E. 1997. Pengkajian Regresi Komponen Utama, Regresi Ridge dan Regresi Kuadrat Terkecil Parsial. Tesis S2, Program Pascasarjana IPB, Bogor. (Tidak dipublikasikan).
- Martens, Harald dan Tormod Naes. 1989. *Multivariate Calibration*. John Willey & Sons. Chichester, England.
- Myers, Raymond H. 1990. *Classical and Modern Regression with Applications*. PWS-Kent Publishing Co. Boston.
- Naes, Tormod. 1985. Multivariate calibration when the error covariance matrix is structured. *Technometrics*, vol.27, no.3: 301-311.
- Nur, M.A. & H. Adijuwana. 1989. *Teknik Spektroskopi dalam Analisis Biologi*. Pusat Antar Universitas Ilmu Hayat, Institut Pertanian Bogor.
- Voght, N.B. 1990. Soft Modelling and Chemo-systematic (dalam *Chemometrics Tutorials*, 321-339). Elsevier Science Publ. Co. Amsterdam.
- Wigena, A.H. & Aunuddin. 1997. Suatu Kajian dan Terapan Metode PLS. Seminar Nasional Statistik IV di ITS Surabaya, 9-10 Desember 1997.
- Young, P.J. 1994. A reformulation of the partial least square regression algorithm. *SIAM J. SCL STAT Comput.*, vol.5, no.1 : 225-230.