

PENDEKATAN REGRESI KUADRAT TERKECIL PARTIAL ROBUST DALAM MODEL KALIBRASI

Ismah¹, Aji Hamim Wigena², Anik Djuraidah²

¹Jurusan Akuntansi/Manajemen Fak. Ekonomi Univ. Satyagama

²Departemen Statistika FMIPA-IPB

E-mail : ¹ismah.fr@gmail.com

Abstrak

Aplikasi model kalibrasi di bidang kimia adalah pemodelan hubungan antara kandungan senyawa aktif yang ditentukan dari High Performance Liquid Chromatography (HPLC) dengan bentuk spektrum dari spektrometer Fourier Transform Infrared (FTIR). Permasalahan utama dalam kalibrasi adalah multikolinear dan pengamatan pencilan. Regresi Kuadrat Terkecil Parsial (RKTP) merupakan sebuah teknik prediktif yang mampu mengatasi masalah multikolinearitas.. SIMPLS (Straightforward Implementation PLS) adalah algoritma pendugaan RKTP yang tidak resisten terhadap pengamatan pencilan. Hubert and Brande (2003) mengemukakan algoritma RSIMPLS yang bersifat resisten terhadap pencilan. RSIMPLS dibentuk dari matriks ragam-peragam robust dan regresi linear robust. Pada penelitian ini dilakukan modifikasi fungsi bobot pada RSIMPLS dengan penduga-M Huber dimana setiap pengamatan akan diberikan nilai bobot kecil ($w_i \approx 0$) jika jarak robust dan jarak ortogonal pengamatan ke-i melebihi nilai batas yang ditentukan, dan $w_i = 1$ untuk lainnya. Dengan demikian besar w_i tidak hanya 0 dan 1, melainkan $0 < w_i \leq 1$. Hasil penelitian menunjukkan RMSEP (root mean square error) pada metode modifikasi bobot lebih kecil dibandingkan RSIMPLS

Kata kunci : calibration multivariate, SIMPLS, robust methods, ROBPCA, Huber estimator

PENDAHULUAN

Informasi khasiat dari tanaman obat dapat dicermati dari kandungan senyawa aktif yang terdapat pada tanaman obat tersebut. Oleh karena itu, kajian mengenai konsentrasi senyawa aktif tanaman obat atau pengisi sangat diperlukan. Penggunaan *High Performance Liquid Chromatography* (HPLC) untuk mengetahui kandungan senyawa aktif membutuhkan proses yang lama serta biaya yang mahal, sedangkan spektrometer *Fourier Transform Infrared* (FTIR) merupakan alat ukur dengan proses yang relatif lebih sederhana dan murah. Oleh karena itu, alternatif yang lebih murah, mudah dan cepat untuk memperoleh dugaan kandungan senyawa aktif adalah mengembangkan model kalibrasi (Setiawan 2007). Namun, ada beberapa masalah yang sering muncul dalam kalibrasi, diantaranya terdapat multikolinearitas dan pengamatan pencilan.

Regresi Kuadrat Terkecil Parsial (RKTP) merupakan sebuah tehnik prediktif yang mampu

mengatasi peubah bebas yang berdimensi besar, khususnya ketika terdapat masalah multikolinearitas. Dalam analisis kuadrat terkecil parsial klasik, perhitungannya didasarkan pada matriks peragam contoh (S) biasa dimana besar kemungkinan terinfeksi oleh pencilan. Untuk mengatasi masalah pencilan diperlukan suatu metode penduga yang resisten terhadap pencilan yang disebut sebagai metode *robust*. Sampai saat ini berbagai algoritma RKTP dengan metode *robust* terus berkembang dan digunakan dalam berbagai bidang, diantaranya adalah RSIMPLS yang dikenalkan oleh Hubert dan Branden (2003). Algoritma RSIMPLS mengkombinasikan antara metode matriks peragam *robust* dan regresi *robust*. Kedua algoritma tersebut dapat diterapkan untuk peubah multirespons. Dalam RSIMPLS besarnya bobot yang diberikan untuk setiap pengamatan yang diidentifikasi sebagai pencilan adalah nol, dengan kata lain bahwa pengamatan tersebut dihilangkan, hal ini sangat disayangkan mengingat bahwa sampel yang diperoleh dalam membentuk model

kalibrasi adalah mahal sehingga banyaknya jumlah sampel yang digunakan sangat terbatas. Oleh sebab itu, dibuat suatu modifikasi fungsi bobot berdasarkan konsep Huber, dimana setiap pengamatan akan diberikan nilai bobot kecil ($w_i \approx 0$) jika jarak robust dan jarak ortogonal pengamatan ke- i melebihi nilai batas yang ditentukan, dan $w_i = 1$ untuk lainnya. Sehingga besar w_i tidak hanya 0 dan 1, melainkan $0 < w_i \leq 1$.

Tujuan penelitian ini adalah menerapkan regresi kuadrat terkecil parsial robust dalam model kalibrasi untuk menentukan kadar senyawa aktif pada tanaman obat.

TINJAUAN PUSTAKA

Fourier Transform Infrared (FTIR)

Cara kerja FTIR secara umum dapat digambarkan sebagai berikut: sampel discan, yang berarti sinar infra-merah akan dilalukan ke sampel. Gelombang yang diteruskan oleh sampel akan ditangkap oleh detektor yang terhubung ke komputer yang akan memberikan gambaran spektrum sampel yang diuji. Struktur kimia dan bentuk ikatan molekul serta gugus fungsional tertentu sampel yang diuji menjadi dasar bentuk spektrum yang akan diperoleh dari hasil analisa. Dengan demikian alat ini dapat digunakan untuk pengujian secara kualitatif dan kuantitatif.

Kromatografi Cair Kinerja Tinggi (HPLC)

HPLC adalah alat yang sangat bermanfaat dalam analisis. HPLC secara mendasar merupakan perkembangan tingkat tinggi dari kromatografi kolom. Selain dari pelarut yang menetes melalui kolom dibawah grafitasi, didukung melalui tekanan tinggi sampai dengan 400 atm, hal ini membuatnya lebih cepat. HPLC memperbolehkan penggunaan partikel yang berukuran sangat kecil untuk material terpadatkan dalam kolom yang mana akan memberi luas permukaan yang lebih besar berinteraksi antara fase diam dan molekul-molekul yang melintasinya. Hal ini memungkinkan pemisahan yang lebih baik dari komponen-komponen dalam campuran.

Perkembangan yang lebih luas melalui kromatografi kolom mempertimbangkan metode pendeteksian yang dapat digunakan. Metode-metode ini sangat otomatis dan sangat peka. Cara kerja HPLC dapat dilihat pada diagram alir HPLC pada Gambar 1.

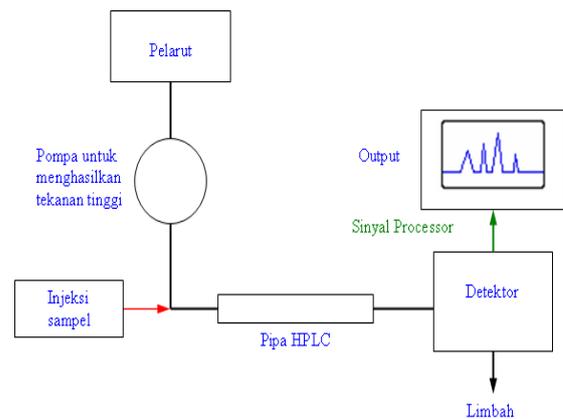
Algoritma SIMPLS

Metode SIMPLS mengasumsikan peubah-peubah x dan y dihubungkan dalam model bilinear seperti berikut ini :

$$\mathbf{x}_i = \bar{\mathbf{x}} + P_{p,k} \tilde{\mathbf{t}}_i + \mathbf{g}_i \quad (1)$$

$$\mathbf{y}_i = \bar{\mathbf{y}} + A'_{q,k} \tilde{\mathbf{t}}_i + \mathbf{f}_i \quad (2)$$

Dalam model tersebut, $\bar{\mathbf{x}}$ dan $\bar{\mathbf{y}}$ merupakan rata-rata dari peubah x dan y . $\tilde{\mathbf{t}}_i$ adalah skor berdimensi k , dengan $k \ll p$ dan $i = 1 \dots n$. $P_{p,k}$ adalah matriks loading x , sedangkan sisaan dalam model ini dinotasikan dengan \mathbf{g}_i dan \mathbf{f}_i . Matriks $A_{k,q}$ direpresentasikan sebagai matriks slope model regresi \mathbf{y}_i dalam $\tilde{\mathbf{t}}_i$.



Gambar 1 : Diagram alir HPLC

Struktur model bilinear (1) dan (2) mengimplikasikan sebuah algoritma 2 langkah. Setelah data dipusatkan, langkah yang pertama SIMPLS yaitu menghitung vektor skor $\tilde{\mathbf{t}}_i$ yang merupakan kombinasi linier rata-rata peubah \mathbf{x} yang telah dipusatkan $\tilde{\mathbf{t}}_{ia} = \tilde{\mathbf{x}}'_i \mathbf{r}_a$ atau sama dengan $\tilde{\mathbf{T}}_{n,k} = \tilde{\mathbf{X}}_{n,p} R_{p,k}$ dengan $R_{p,k} = (\mathbf{r}_1, \dots, \mathbf{r}_k)$, \mathbf{r}_a dan \mathbf{q}_a merupakan vektor bobot SIMPLS untuk setiap $a = 1, \dots, k$. \mathbf{r}_a dan \mathbf{q}_a diperoleh dengan memaksimumkan peragang antara komponen-komponen x dan y .

$$\max_{\|\mathbf{r}_a\|=1, \|\mathbf{q}_a\|=1} \text{cov}(\tilde{\mathbf{Y}}_{n,q} \mathbf{q}_a, \tilde{\mathbf{X}}_{n,p} \mathbf{r}_a) = \max_{\|\mathbf{r}_a\|=1, \|\mathbf{q}_a\|=1} \mathbf{q}'_a S_{yx} \mathbf{r}_a$$

\mathbf{q}_1 adalah vektor ciri dari $S_{yx} S_{xy}$ dan \mathbf{r}_1 adalah vektor ciri dari $S_{xy} S_{yx}$ dengan $(S_{xy} = S'_{yx})$. Selanjutnya sepasang vektor bobot SIMPLS $(\mathbf{r}_a, \mathbf{q}_a)$ dengan $2 \leq a \leq k$ adalah vektor ciri $S_{yx}^a S_{xy}^a$ dan $S_{xy}^a S_{yx}^a$.

$$S_{xy}^a = S_{xy}^{a-1} - \mathbf{v}_a (\mathbf{v}'_a S_{xy}^{a-1}) = (I_p - \mathbf{v}_a \mathbf{v}'_a) S_{xy}^{a-1} \quad (3)$$

dan $S_{xy}^1 = S_{xy}$. Proses ini diulang sampai jumlah komponen k ditentukan.

Salah satu teknik untuk menentukan jumlah komponen k yaitu dengan menghitung nilai *Root Mean Squared Error* (RMSE) untuk setiap k . Jumlah komponen yang optimal ditentukan dari komponen k yang memiliki nilai RMSE minimum. Langkah kedua, peubah-peubah respon diregresikan kedalam skor $\tilde{\mathbf{t}}_i$. Sehingga model regresi adalah,

$$\mathbf{y}_i = \boldsymbol{\alpha}_0 + \mathbf{A}'_{q,k} \tilde{\mathbf{t}}_i + \mathbf{f}_i \quad (4)$$

dengan $E(\mathbf{f}_i) = 0$ dan $\text{cov}(\mathbf{f}_i) = \Sigma_f$ yang merupakan performa dari regresi linier berganda.

Penduga regresi linier berganda diperoleh sebagai berikut :

$$\begin{aligned} \hat{\mathbf{A}}_{k,q} &= (S_t)^{-1} S_{ty} = (\mathbf{R}'_{k,p} S_x \mathbf{R}_{p,k})^{-1} \mathbf{R}'_{k,p} S_{xy} \\ \hat{\boldsymbol{\alpha}}_0 &= \bar{\mathbf{y}} - \hat{\mathbf{A}}'_{q,k} \tilde{\mathbf{t}} \\ S_f &= S_y - \hat{\mathbf{A}}'_{q,k} S_t \hat{\mathbf{A}}_{k,q} \end{aligned} \quad (5)$$

Karena $\tilde{\mathbf{t}} = 0$ maka intersept $\boldsymbol{\alpha}_0$ diduga dengan $\bar{\mathbf{y}}$. Dengan $\tilde{\mathbf{t}}_i = \mathbf{R}'_{k,p} (\mathbf{x}_i - \bar{\mathbf{x}})$ dari persamaan (2), diperoleh penduga parameter untuk model regresi linier yaitu :

$$\begin{aligned} \hat{\mathbf{B}}_{p,q} &= \mathbf{R}_{p,k} \hat{\mathbf{A}}_{k,q} \\ \hat{\boldsymbol{\beta}}_0 &= \bar{\mathbf{y}} - \hat{\mathbf{B}}'_{q,p} \bar{\mathbf{x}} \end{aligned}$$

penduga Σ_e yaitu S_e merupakan fungsi dalam parameter, $S_e = S_y - \hat{\mathbf{B}}'_{q,p} S_x \hat{\mathbf{B}}$.

Metode Least Trimmed Square (LTS)

Metode LTS menduga koefisien regresi dengan menggunakan metode kuadrat terkecil terhadap subhimpunan data berukuran h , dengan $\frac{n}{2} < h < n$.

Penentuan subhimpunan terbaik dilakukan dengan menggunakan algoritma resampling dari seluruh kemungkinan subhimpunan yang dapat dibentuk yaitu sebanyak $\binom{n}{h}$.

Metode Minimum Covariance Determinant (MCD)

Misalkan $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ merupakan suatu contoh dari n pengamatan dalam \mathbf{R}^k dan h , dengan $\frac{n}{2} < h < n$, cari subhimpunan \mathbf{J}^* berukuran h sedemikian hingga

$$\mathbf{J}^* = \min_{J \subset \{1, 2, \dots, n\}, |J|=h} \det \hat{S}_J$$

dengan \hat{S}_J adalah matriks peragam berdasarkan pada pengamatan x_i dengan $i \in J$,

$$\text{dengan } \hat{S}_{J^*} = \frac{1}{h} \sum_{i \in J^*} (x_i - \bar{x}_{J^*}) (x_i - \bar{x}_{J^*})'$$

$$\text{dan } \bar{x}_{J^*} = \frac{1}{h} \sum_{i \in J^*} x_i.$$

Metode MCD yang lebih efisien adalah dengan menggantikan rata-rata dan matriks peragam klasik dengan penduga pusat dan ragam MCD terboboti, dimana masing-masing x_i diberikan bobot $w_i, w_i = 1$ apabila $(x_i - \hat{\boldsymbol{\mu}}_0)' \hat{\Sigma}_0^{-1} (x_i - \hat{\boldsymbol{\mu}}_0) \leq \chi_{q,0.975}^2$ dan $w_i = 0$ untuk lainnya.

Analisis Komponen Utama Robust

Salah satu pendekatan AKU *robust* dikembangkan oleh Hubert *et al* (2003) yang disebut ROBPCA. Metode ini mengkombinasikan dua pendekatan pursuit proyeksi dan penduga peragam *robust*. Langkah pertama melakukan prapemrosesan data dengan mereduksi ruang data kedalam subruang yang direntang dengan pengamatan n . Langkah kedua, menentukan keterpencilan untuk setiap pengamatan sebagai berikut

$$\text{outl}(\mathbf{x}_i) = \max_d \frac{|\mathbf{x}'_i d - \hat{\boldsymbol{\mu}}(\mathbf{x}'_i d)|}{\hat{\Sigma}(\mathbf{x}'_i d)} ; j = 1, \dots, n$$

dengan $\hat{\boldsymbol{\mu}}$ dan $\hat{\Sigma}$ adalah rata-rata dan simpangan baku *robust* dari MCD. Kemudian membentuk matriks peragam $\hat{\Sigma}_0$ dari titik-titik data h dengan keterpencilan yang paling kecil. $\hat{\Sigma}_0$ digunakan untuk memilih jumlah komponen k , selanjutnya data diproyeksi ke dalam subruang k . Langkah ketiga, metode MCD diaplikasikan untuk menduga pusat dan matriks ragam pada subruang data yang berdimensi k ($k < m$) dimana $m = p + q$. Dengan kata lain, pendugaan ini adalah transformasi balik untuk ruang data yang asli sehingga diperoleh penduga pusat robust $\hat{\boldsymbol{\mu}}_z$ dari $Z_{n,m} = (X_{n,p}, Y_{n,q})$ dan ragamnya $\hat{\Sigma}_z$.

Selanjutnya $\hat{\Sigma}_z$ dapat didekomposisi sebagai berikut :

$$\hat{\Sigma}_z = \begin{pmatrix} \hat{\Sigma}_x & \hat{\Sigma}_{xy} \\ \hat{\Sigma}_{yx} & \hat{\Sigma}_y \end{pmatrix} = \mathbf{P}^z \mathbf{L}^z (\mathbf{P}^z)'$$
 (6)

Dengan vektor ciri Z *robust* \mathbf{P}^z_{m,k_0} dan akar ciri Z adalah $\text{diag}(\mathbf{L}^z_{k_0,k_0})$.

Regresi Kuadrat Terkecil Parsial Robust (RSIMPLS)

Langkah awal dalam RKTP *robust* adalah menerapkan metode ROBPCA sehingga diperoleh

penduga pusat *robust* $\hat{\mu}_z$ dan ragamnya $\hat{\Sigma}_z$. Seperti pada tahapan SIMPLS diperoleh skor-skor *robust* berdimensi k yang kemudian diregresikan kedalam peubah respon

$$y_i = \alpha_0 + A'_{q,k} \mathbf{t}_i + \tilde{\mathbf{f}}_i \quad (7)$$

dimana penduga pusat $\boldsymbol{\mu}$ dan ragam Σ dari (t, y) yaitu rata-rata dan matriks peragam terboboti

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix} \hat{\boldsymbol{\mu}}_t \\ \hat{\boldsymbol{\mu}}_y \end{pmatrix} = \frac{\sum_{i=1}^n w_i \begin{pmatrix} \mathbf{t}_i \\ \mathbf{y}_i \end{pmatrix}}{\sum_{i=1}^n w_i} \quad \text{dan}$$

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_t & \hat{\Sigma}_{ty} \\ \hat{\Sigma}_{yt} & \hat{\Sigma}_y \end{pmatrix} = \frac{\sum_{i=1}^n w_i \begin{pmatrix} \mathbf{t}_i \\ \mathbf{y}_i \end{pmatrix} \begin{pmatrix} \mathbf{t}'_i & \mathbf{y}'_i \end{pmatrix}}{\sum_{i=1}^n w_i - 1}.$$

dengan $w_i = 1$ apabila pengamatan ke- i tidak diidentifikasi sebagai pencilan dengan metode ROBPCA dalam (x, y) dan $w_i = 0$ untuk lainnya.

Ketika ROBPCA diaplikasikan ada dua jenis pencilan yang dapat diidentifikasi. Jenis pencilan yang pertama dapat dengan mudah diidentifikasi apabila jarak *robust* pengamatan ke- i

$$D_{i(k_0)} = \sqrt{(\mathbf{t}_i^z)' (L^z)^{-1} \mathbf{t}_i^z} \text{ melebihi } \sqrt{\chi_{k_0, 0.975}^2}$$

dengan L^z diberikan pada persamaan (7). Jenis pencilan yang kedua, $w_i = 0$ apabila

$$OD_i > \sqrt{\hat{\mu}_{0d}^2 + \sigma_{0d}^2 z_{0.975}}$$

dengan $OD_i = \left\| (z_i - \boldsymbol{\mu}'_z) - P^z \mathbf{t}_i^z \right\|$ merupakan jarak orthogonal, $\hat{\mu}_{0d}^2$ dan σ_{0d}^2 adalah penduga pusat dan ragam dengan MCD. Setelah $\hat{\boldsymbol{\mu}}$ dan $\hat{\Sigma}$ diperoleh, proses selanjutnya menduga koefisien regresi diperoleh menggunakan MKT.

Penduga-M (M-Estimator)

Pembobot dalam penduga-M bergantung pada residual dan koefisien. Fungsi penimbang untuk ketiga metode penduga-M yaitu, kuadrat terkecil, Huber dan Tukey (*biweight*) terdapat pada Tabel 1

Nilai k disebut tuning konstan, menurunkan nilai k akan menaikkan pembobot terhadap residual yang besar dan menaikkan nilai k akan menurunkan pembobot terhadap residual yang besar. Semakin besar k maka estimasi robust akan mendekati *least square*. Kuzmic *et al* (2004) menyebutkan penduga-M Huber efektif digunakan pada $\alpha=5\%$ dengan $k = 1.345$ sedangkan penduga-M Tukey dengan $k = 4.685$.

Tabel 1 fungsi pembobot pada pendugaan-M

Metode	Fungsi Pembobot
Kuadrat Terkecil	$w_{LS}(e) = 1$
Huber	$w_H(e) = \begin{cases} 1 & , \text{untuk } e \leq k \\ k/ e & , \text{untuk } e > k \end{cases}$
Tukey (Bisquare / Biweight)	$w_B = \begin{cases} \left[1 - \left(\frac{e}{k} \right)^2 \right]^2 & , \text{untuk } e \leq k \\ 0 & , \text{untuk } e > k \end{cases}$

Model Kalibrasi dan Validasi

Dalam validasi, data pengamatan sebanyak n buah dibagi dua, bagian yang pertama digunakan untuk membangun model dan yang kedua digunakan untuk menguji model (validasi). Untuk membangun model RKTP yaitu dengan memilih jumlah komponen yang optimal (k_{opt}). Nilai k_{opt} diperoleh dari nilai RMSECV *Robust* (R-RMSECV _{k}) minimum dari setiap k .

$$R\text{-RMSECV}_k = \sqrt{\frac{1}{n_c q} \sum_{i \in G_c} \sum_{j=1}^q (y_{ij} - \hat{y}_{-ij(k)})^2}$$

Masing-masing pengamatan dihilangkan dan G_c merupakan subset pengamatan dimana $c_{-i} = 1$ dengan $|G_c| = n_c$. $\hat{y}_{-ij(k)}$ adalah dugaan y_{ij} tanpa pengamatan ke- i .

Salah satu jenis validasi model yaitu dengan menghitung nilai *robust root mean squared error of prediction* (R-RMSEP),

$$\text{RMSEP}_{k(opt)} = \sqrt{\frac{1}{n_p q} \sum_{i=1}^{n_p} \sum_{j=1}^q (y_{ij} - \hat{y}_{ij(k)})^2}$$

dengan y_{ij} adalah nilai pengamatan ke- i dan dimensi ke- j pada kelompok data validasi, $\hat{y}_{ij(k)}$ adalah nilai dugaan pengamatan ke- i , dimensi ke- j dan komponen ke- k dengan menggunakan model yang dibangun dengan hanya melibatkan bagian data yang pertama.

Pencilan (Outlier)

Pencilan adalah pengamatan yang jauh dari pusat data yang mungkin berpengaruh besar terhadap koefisien regresi. Keberadaan data pencilan akan mengganggu dalam proses analisis data. Dalam kaitannya dengan analisis regresi, pencilan dapat menyebabkan hal-hal berikut :

- Nilai residual menjadi besar dari model yang terbentuk atau $E(e) \neq 0$
- Ragam pada data menjadi lebih besar

c. Taksiran interval memiliki rentang yang lebar
Terdapat beberapa cara untuk mengidentifikasi keberadaan pencilan.

a. Plot Diagnosa Regresi

Untuk mengidentifikasi pengamatan pencilan yaitu dengan mengamati persamaan regresi (8). Sebuah pengamatan diidentifikasi sebagai titik *good leverage* apabila posisi pengamatan tersebut berada dalam arah garis regresi atau subruang, dan nilai t -nya diidentifikasi sebagai pencilan. Sedangkan pengamatan yang diidentifikasi sebagai titik *bad leverage* apabila pengamatan jauh dari garis regresi. Dan pengamatan diidentifikasi sebagai pencilan vertikal apabila diidentifikasi sebagai pencilan dalam ruang y .

Untuk mengukur keterpencilan sebuah titik dalam ruang t , yaitu dengan menghitung jarak skor sebagai berikut.

$$SD_{i(k)}^2 = (\mathbf{t}_i - \hat{\boldsymbol{\mu}}_t)' \hat{\boldsymbol{\Sigma}}_t^{-1} (\mathbf{t}_i - \hat{\boldsymbol{\mu}}_t)$$

Jarak skor mencermati sumbu horizontal pada plot diagnosa regresi dan menyingkap titik *good* dan *bad leverage*. Pengamatan diidentifikasi sebagai titik *leverage* apabila jarak skor melebihi nilai batas $\sqrt{\chi_{k,0.975}^2}$.

Untuk sumbu vertikal dapat dicermati dari nilai jarak residual berikut.

$$RD_{i(k)}^2 = \mathbf{r}'_{i(k)} \hat{\boldsymbol{\Sigma}}_e^{-1} \mathbf{r}_{i(k)}$$

dengan $\mathbf{r}_{i(k)} = \mathbf{y}_i - \hat{\boldsymbol{\beta}}_0 - \hat{\mathbf{B}}' \mathbf{x}_i$.

Apabila peubah respon satu jarak residual disederhanakan menjadi residual yang distandarisasi yaitu $RD_{i(k)} = \frac{\mathbf{r}_{i(k)}}{\hat{\sigma}_e}$. Pengamatan diidentifikasi sebagai pencilan vertikal dan titik *bad leverage* apabila jarak residual pengamatan tersebut melebihi $\sqrt{\chi_{q,0.975}^2}$

b. Plot Diagnosa Skor

Klasifikasi pengamatan dapat dicermati dari persamaan AKU (1). Dalam sumbu horizontal yaitu dengan meletakkan jarak skor untuk masing-masing pengamatan. Sedangkan dalam sumbu vertikal, mencermati jarak ortogonal sebuah pengamatan terhadap ruang t .

$$OD_{i(k)} = \|\tilde{\mathbf{x}}_i - P_{p,k} \mathbf{t}_i\|$$

Kombinasi jarak skor dan jarak ortogonal menghasilkan jenis pengamatan seperti tertera pada Tabel 2

METODOLOGI PENELITIAN

Data

Data yang digunakan dalam penelitian ini merupakan bagian dari data penelitian Hibah Pascasarjana. Penelitian ini adalah hasil kerjasama

antara Departemen Statistika IPB dengan pusat studi Biofarmaka LPPM IPB yang didanai oleh DP3M, Dirjen Pendidikan Tinggi, Departemen Pendidikan Nasional. Banyaknya pengamatan yang digunakan untuk membangun model kalibrasi adalah 20 rimpang temulawak yang diukur menggunakan metode HPLC (*High Performance Liquid Chromatography*), mengenai konsentrasi senyawa aktif dalam rimpang temulawak yang disebut kurkuminoid sebagai peubah respon (Y). Dan data mengenai persen transmitan yang dihasilkan metode FTIR (*Fourier Transform Infrared*) pada 1866 titik di sepanjang kisaran bilangan gelombang $4000-400 \text{ cm}^{-1}$ sebagai peubah bebas (X).

Tahapan Analisis data

Analisis data dengan menggunakan metode RSIMPLS terbagi dalam dua tahap yaitu membangun model kalibrasi dan validasi data. Pada tahap 1 adalah membangun model kalibrasi dengan langkah-langkah sebagai berikut :

1. Mencari skor RSIMPLS
 - 1.1. Standarisasi peubah-peubah x dan y
 - 1.2. Tentukan matriks peragam robust menggunakan metode ROBPCA
 - 1.3. Hitung sepasang vektor bobot RSIMPLS yang pertama, \mathbf{r}_1 dan \mathbf{q}_1 .
 - 1.4. Hitung skor RSIMPLS T_1
 - 1.5. Hitung loading- x , $\mathbf{p}_j = (\mathbf{r}'_j \hat{\boldsymbol{\Sigma}}_x \mathbf{r}_j)^{-1} \hat{\boldsymbol{\Sigma}}_x \mathbf{r}_j$
 - 1.6. Langkah 5 dipenuhi apabila $\mathbf{p}'_j \mathbf{r}_a = 0$ untuk $a > j$.
 - 1.7. Hitung sebuah basis ortonormal $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{a-1}\}$ loading- x $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{a-1}\}$ untuk $(2 \leq a \leq k)$
 - 1.8. Hitung matriks peragam silang, $\hat{\boldsymbol{\Sigma}}_{xy}^a = \hat{\boldsymbol{\Sigma}}_{xy}^{a-1} - \mathbf{v}_{a-1} (\mathbf{v}'_{a-1} \hat{\boldsymbol{\Sigma}}_{xy}^a)$
 - 1.9. Hitung vektor bobot RSIMPLS \mathbf{r}_a dan \mathbf{q}_a ($2 \leq a \leq k$) sebagai vektor-vektor singular kiri dan kanan yang pertama $\hat{\boldsymbol{\Sigma}}_{xy}^a$
 - 1.10. Ulangi langkah 4 untuk $2 \leq a \leq k$
 - 1.11. Pilih k_{opt} berdasarkan pada nilai $R-RMSECV_k$ minimum.
2. Meregresikan peubah respon y terhadap skor RSIMPLS
 - 2.1. Hitung penduga koefisien regresi dengan metode kuadrat terkecil. Pada tahap ini koefisien regresi diperoleh berdasarkan penduga rata-rata dan penduga matriks peragam terboboti, dimana besarnya bobot sama dengan satu apabila pengamatan ke- i tidak diidentifikasi sebagai pencilan

Tabel 2. Kombinasi jarak skor dan jarak ortogonal

Jarak	$SD \leq \sqrt{\chi_{k,0.975}^2}$	$SD > \sqrt{\chi_{k,0.975}^2}$
$OD_i \leq \sqrt{\hat{\mu}_{0d^2} + \sigma_{od^2}^2 z_{0.975}}$	Pencilan ortogonal	Titik <i>bad</i> AKU leverage
$OD_i > \sqrt{\hat{\mu}_{0d^2} + \sigma_{od^2}^2 z_{0.975}}$	Pengamatan regular	Titik <i>good</i> AKU leverage

dengan metode ROBPCA dalam dan besarnya bobot sama dengan nol untuk lainnya

2.2. Hitung koefisien regresi RSIMPLS untuk peubah-peubah asal.

Pada Langkah 2.1 untuk memperoleh penduga koefisien regresi t dan y diperoleh berdasarkan rataan dan matriks peragam terboboti, dimana bobot yang diberikan untuk setiap pengamatan yang diidentifikasi sebagai pencilan dimodifikasi berdasarkan konsep Huber, setiap pengamatan akan diberikan nilai bobot kecil ($w_i \approx 0$) jika jarak robust dan jarak ortogonal pengamatan ke- i melebihi nilai batas yang ditentukan, dan $w_i = 1$ untuk lainnya. Sehingga besar w_i tidak hanya 0 dan 1, melainkan $0 < w_i \leq 1$.

$$\bullet w_i = \begin{cases} 1 & \text{jika } D_i \leq \sqrt{\chi_{k_0,0.975}^2} \\ \frac{\sqrt{\chi_{k_0,0.975}^2}}{D_i} & \text{jika } D_i > \sqrt{\chi_{k_0,0.975}^2} \end{cases}$$

$$\bullet w_i = \begin{cases} 1 & \text{jika } OD_i \leq \sqrt{\hat{\mu}_{0d^2} + \sigma_{od^2}^2 z_{0.975}} \\ \frac{\sqrt{\hat{\mu}_{0d^2} + \sigma_{od^2}^2 z_{0.975}}}{OD_i} & \text{jika } OD_i > \sqrt{\hat{\mu}_{0d^2} + \sigma_{od^2}^2 z_{0.975}} \end{cases}$$

$$\bullet w_i(akhir) = \begin{cases} RD_i & \text{jika } RD_i \leq OD_i \\ OD_i & \text{jika } RD_i > OD_i \end{cases}$$

Pada tahap 2 dilakukan validasi model yaitu melakukan prediksi nilai y pada kelompok data validasi dengan menggunakan model yang dihasilkan pada tahap 1. Selanjutnya dilakukan validasi model dengan kriteria RMSEP.

HASIL DAN PEMBAHASAN

Analisis RSIMPLS dalam penelitian ini menggunakan nilai $h=17$ dan $h=15$ sebagai ukuran subhimpunan dengan $\frac{n}{2} < h < n$ dimana titik-titik

pengamatan yang berada pada subhimpunan tersebut memiliki nilai keterpencilan minimum, seperti yang telah dibahas dalam tinjauan pustaka mengenai ROBPCA dan tahapan analisis tahap 1 bagian 1.2. Masing-masing hasil analisis berdasarkan kedua nilai h tersebut disajikan pada Tabel 3 dan Tabel 4. Model yang dianalisis adalah :

1. RSIMPLS_Huber adalah semua kemungkinan jumlah sampel yang terambil untuk menguji model (15504) tanpa memodifikasi fungsi bobot.
2. RSIMPLS_new adalah semua kemungkinan jumlah sampel yang terambil untuk menguji model (15504) dengan memodifikasi fungsi bobot.
3. RSIMPLS_new(0) adalah semua kemungkinan jumlah sampel yang terambil untuk menguji model tidak terdapat lima pengamatan pencilan (3003) dengan memodifikasi fungsi bobot.
4. RSIMPLS_new(1) adalah semua kemungkinan jumlah sampel yang terambil untuk menguji model terdapat satu buah pengamatan pencilan (6825) dengan memodifikasi fungsi bobot.
5. RSIMPLS_new(2) adalah semua kemungkinan jumlah sampel yang terambil untuk menguji model terdapat dua buah pengamatan pencilan (4550) dengan memodifikasi fungsi bobot.
6. RSIMPLS_new(3) adalah semua kemungkinan dari banyaknya sampel yang digunakan untuk menguji model terdapat tiga buah pengamatan pencilan (1050) dengan memodifikasi fungsi bobot.
7. RSIMPLS_new(4) adalah semua kemungkinan dari banyaknya sampel yang digunakan untuk menguji model terdapat empat buah pengamatan pencilan (75) dengan memodifikasi fungsi bobot.
8. RSIMPLS_new(5) adalah semua kemungkinan jumlah sampel yang terambil untuk menguji model terdapat lima pengamatan pencilan (1) dengan memodifikasi fungsi bobot.

Untuk menguji ketepatan model digunakan nilai $rmsep$ (*root mean square error of prediction*) dan $rmse$ (*root mean square error*). Pengamatan yang dideteksi sebagai pencilan yaitu pengamatan ke 2, 14, 16, 17 dan 18

Tabel 3. Hasil analisis beberapa model RSIMPLS dengan h=17

Model	rrmse	minrmsep	maxrmsep	rrmse	minrmse	maxrmse
RSIMPLS_Huber	1.240	0.165	8.310	-	-	-
RSIMPLS_new	0.884	0.127	5.811	0.545	0.180	1.480
RSIMPLS_new(0)	0.444	0.198	1.173	0.366	0.217	0.616
RSIMPLS_new(1)	0.810	0.127	2.805	0.569	0.180	1.480
RSIMPLS_new(2)	1.176	0.164	5.165	0.662	0.229	1.375
RSIMPLS_new(3)	1.294	0.174	5.811	0.403	0.190	1.178
RSIMPLS_new(4)	1.759	0.519	4.000	0.363	0.187	0.658
RSIMPLS_new(5)	1.429	1.429	1.429	0.319	0.319	0.319

Tabel 4. Hasil analisis beberapa model RSIMPLS dengan h=15

Model	rrmse	minrmsep	maxrmsep	rrmse	minrmse	Maxrmse
RSIMPLS_new	0.919	0.108	7.676	0.610	0.192	1.806
RSIMPLS_new(0)	0.537	0.122	1.780	0.814	0.308	1.806
RSIMPLS_new(1)	0.789	0.108	4.522	0.575	0.192	1.613
RSIMPLS_new(2)	1.105	0.156	6.069	0.533	0.207	1.606
RSIMPLS_new(3)	1.870	0.136	7.676	0.592	0.249	1.602
RSIMPLS_new(4)	3.311	0.462	7.224	0.655	0.396	1.274
RSIMPLS_new(5)	2.367	2.367	2.367	0.474	0.474	0.474

Keterangan untuk Tabel 3 dan Tabel 4.

rrmse = rata-rata nilai rmsep
 maxrmsep = nilai terbesar rmsep
 minrmse = nilai terkecil rmse

minrmsep = nilai terkecil rmsep
 rrmse = rata-rata nilai rmse
 maxrmse = nilai terbesar r rmse

KESIMPULAN

1. Nilai rmsep yang dihasilkan dengan metode RSIMPLS berdasarkan konsep Huber cenderung lebih besar dibandingkan metode RSIMPLS dengan memodifikasi fungsi bobot, dalam hal ini disebut dengan RSIMPLS_new.
2. Dari kedua tabel di atas terlihat bahwa semakin banyak pengamatan pencilan yang terdapat pada gugus sampel untuk menguji model kecenderungan nilai rmsep semakin besar. Dengan kata pencilan yang dideteksi mempengaruhi nilai rmse dan rmsep.

DAFTAR PUSTAKA

Debruyne M, Engelen S., Hubert M, Rousseeuw PJ. 2006. Robustness and Outlier Detection in Chemometrics. *Critical reviews in analytical chemistry*, 36(3): 22.

Engelen S, Hubert M, Branden, K, dan Verboven S. 2004, *Robust PCR dan Robust PLSR : a comparative study. Theory and applications of recent robust methods*. Basel, Birkhäuser : 105-117.

Hubert M & Branden KV. 2003. Robust methods for Partial Least Squares Regression, *Journal of Chemometrics*. 17 : 537-549.

Hubert M, Rousseeuw P J, Branden KV. 2003. ROBPCA: A New Approach to Robust Principal Component Analysis. *Technometrics*. 47(1): 64-79.

Hubert M, Rousseeuw PJ, Verboven S. 2002. A fast robust method for principal component with applications to chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 60 : 101-111.

Norliza. 2006. Comparing Three Methods of Handling Multicollinearity Using Simulation Approach [Tesis]. Universiti Teknologi Malaysia.

Rousseeuw PJ, Van Aelst S, Van Driessen K. 2004. Robust multivariate regression. *Technometrics*, 46(1): 293-305.

Setiawan. 2007. Pendekatan Regresi Kontinum Dalam Model Kalibrasi. [Disertasi]. Bogor : Sekolah Pascasarjana, Institut Pertanian Bogor.

Sundberg R. 1999. Multivariate Calibration-Direct and Indirect Regression Methodology. *Board of the Foundation of the Scandinavian Journal of Statistics* 26: 161-207.

Suryana. 2007. Analisis Data Outlier Pada Data
Pengeluaran Rumah Tangga Di Kota Kupang,
NTT Tahun 2005 Dengan Metode ROBPCA.
[Paper]. Surabaya : Program Studi Magister,

Jurusan Statistik, Fakultas MIPA, Institut
Teknologi Surabaya.
Verboven S Hubert, M. 2004. LIBRA: a MATLAB
Library for Robust Analysis. <http://www.wis.kuleuven.ac.be/stat/robust.html>. (1 Maret 2009).