

PENERAPAN *PARTIAL LEAST SQUARE REGRESSION* (PLSR) PADA PEMODELAN KALIBRASI SENYAWA AKTIF KURKUMIN

¹Vera Maya Santi

¹Jurusan Matematika FMIPA Universitas Negeri Jakarta

Abstract. *Partial Least Square Regression (PLSR) merupakan salah satu analisis regresi yang dapat diterapkan untuk mereduksi dimensi data dalam mengatasi masalah multikolinieritas dan mampu diterapkan dalam kalibrasi multivariat. Data senyawa aktif kurkumin yang berupa data spektrum biasanya berdimensi besar menyebabkan terjadinya multikolinieritas. Untuk mengatasi hal ini, penulis menerapkan PLSR dalam mereduksi dimensi data serta membangun suatu model kalibrasi pada data senyawa aktif kurkumin dengan data pencilon (outlier). Berdasarkan kriteria RMSECV diperoleh sebanyak 121 peubah bebas berpengaruh bagi model. Model PLSR yang diperoleh dengan melibatkan sebanyak 121 peubah bebas menghasilkan nilai RMSEP sebesar 0,063.*

Kata kunci : *PLSR, model kalibrasi, senyawa aktif kurkumin*

1. Pendahuluan

1.1. Latar Belakang

Penduduk Indonesia sejak zaman dahulu telah menggunakan banyak tanaman menjadi jamu dan obat-obatan tradisional. Budaya ini telah diwariskan dari nenek moyang mereka secara turun-temurun. Pada era globalisasi ini, telah banyak perusahaan farmasi dan jamu yang berkembang pesat dan juga memanfaatkan tanaman obat dalam produk-produknya. Seiring dengan penggunaan tanaman obat yang semakin meluas, tentu saja kualitas tanaman obat yang digunakan juga harus mendapat perhatian khusus, terutama bagi kalangan industri farmasi dan jamu, untuk menjamin agar produknya tetap dapat bersaing dan diterima di kalangan masyarakat. Salah satu indikator kualitas tanaman obat adalah konsentrasi senyawa aktifnya.

Salah satu metode alternatif yang dapat menyatakan hubungan antara konsentrasi senyawa aktif hasil pengukuran HPLC dengan persen transmittan (absorban) yang diukur dengan menggunakan FTIR adalah model kalibrasi. Menurut Naes *et al* (2002), kalibrasi didefinisikan sebagai sebuah disiplin ilmu yang fokus pada penemuan sebuah hubungan antara sekumpulan hasil pengukuran yang relatif murah atau mudah diperoleh dengan sekumpulan hasil pengukuran yang relatif mahal atau sulit diperoleh yang bertujuan untuk memprediksi konsentrasi senyawa aktif dengan akurasi yang tinggi dari nilai persen transmittan yang secara ekonomi lebih murah dan mudah diperoleh.

Beberapa peneliti telah mengembangkan model kalibrasi untuk kasus yang berbeda. Andras (2008) membandingkan beberapa model kalibrasi dari data NIRS antara lain PCA-MLR (*Principal Component Analysis-Multiple Linier Regression*), PLS, PCA-ANN (*Artificial Neural Network*) dan GA-ANN, sedangkan Markus (1996) membandingkan MLR, *Principal Component Regression* (PCR) dan PLS untuk mengevaluasi dan mendeteksi data gas kromatografi.

Beberapa peneliti di Institut Pertanian Bogor (IPB) juga telah mengembangkan model kalibrasi data spektroskopi. Erfiani (2005) mengembangkan model kalibrasi dengan pendekatan Bayes dimana reduksi peubahnya melalui pendekatan regresi terpenggal. Sunaryo (2005) menggunakan Regresi Komponen Utama dengan metode wavelet sebagai metode pra-pemrosesan. Atok (2005) menggunakan Jaringan Syaraf Tiruan (JST) dengan pra-pemrosesan PCA.

1.2. Tujuan

Adapun tujuan penelitian ini adalah membangun model kalibrasi PLSR pada pengukuran konsentrasi senyawa aktif kurkumin berdasarkan data persen transmisinya.

1.3. Pembatasan Masalah

Pada penelitian ini dibatasi pada metode pra-pemrosesan yang digunakan untuk mereduksi dimensi data. Metode yang digunakan sebagai pra-pemrosesan pada penelitian ini adalah metode *Genetic Algorithm* (GA), sedangkan data yang dibatasi pada data konsentrasi senyawa aktif kurkumin dengan pencilan (*outlier*).

2. Landasan Teori

2.1. Kandungan Senyawa Aktif Kurkumin pada Temulawak

Menurut Sinambela (1985), komposisi rimpang temulawak dapat dibagi menjadi dua fraksi utama yaitu zat warna kurkuminoid dan minyak atsiri. warna kekuningan temulawak disebabkan adanya kurkuminoid. Kandungan utama kurkuminoid terdiri dari senyawa kurkumin, desmetoksikurkumin dan bisdesmetoksikurkumin. Rimpang temulawak segar, selain terdiri dari senyawa kurkuminoid dan minyak atsiri juga mengandung lemak, protein, selulosa, pati, dan mineral. Kadar masing-masing zat tersebut tergantung pada umur rimpang yang dipanen serta juga dipengaruhi oleh letak dan ketinggian tempat temulawak berada.

2.2. Partial Least Square Regression (PLSR)

Ide dasar dari metode PLS adalah menemukan beberapa kombinasi linier dari peubah-peubah X data asli dan menggunakan kombinasi linier tersebut dalam persamaan regresi (Wold *et al* 1984, Martens & Naes 1989, Martens & Martens 2001, Naes *et al* 2002).

Matriks dan vektor data yang digunakan dalam regresi PLS dinotasikan dengan X dan \underline{y} . Matriks X dan vektor \underline{y} keduanya diasumsikan *mean-centred*. Struktur model untuk metode PLS dinyatakan dalam dua persamaan yang sering disebut model biliner, yaitu:

$$\begin{aligned} X &= TP^t + E \\ \underline{y} &= T\underline{q} + F \end{aligned} \quad \dots\dots (1)$$

dimana :

P : matriks loading dan menjelaskan hubungan antar variabel di dalam T dengan matriks data asal X .

\underline{q} : vektor loading dan menjelaskan hubungan antar variabel di dalam T dengan matriks data asal Y .

T : matriks skor.

E dan \underline{f} : matriks dan vektor sisaan.

Pada metode PLS, peubah-peubah dalam matriks T merupakan kombinasi linier dari peubah-peubah data asal dalam matriks X. Loading dalam P dan \underline{q} diduga dengan meregresikan X dan \underline{y} ke matriks T. Selanjutnya, akan diperoleh matriks dan vektor sisaan dari komponen pertama PLS sebagai berikut:

$$\begin{aligned} E &= X - TP^t \\ \underline{f} &= \underline{y} - T\underline{q} \end{aligned} \quad \dots (2)$$

Arah dari komponen pertama PLS diperoleh dengan memaksimumkan kriteria kovarian yang dinotasikan dengan \underline{w}_1 . \underline{w}_1 merupakan vektor satuan yang biasanya disebut vektor pembobot loading yang merupakan vektor ciri pertama dari akar ciri pertama matriks X^tYY^tX dengan X^tY merupakan kovarians antara matriks X dan vektor \underline{y} . Skor \underline{t} dari matriks X untuk komponen pertama PLS dihitung dengan rumus $\underline{t}_1 = X\underline{w}_1$. Semua peubah dalam X selanjutnya diregresikan ke \underline{t}_1 sehingga menghasilkan vektor loading \underline{p}_1 dan koefisien regresi \underline{q}_1 merupakan hasil regresi dari \underline{y} ke \underline{t}_1 . Setelah faktor-faktor pertama direduksi akan menghasilkan matriks X dan vektor \underline{y} hasil reduksi sebagai berikut:

$$\begin{aligned} X_1 &= X - \underline{t} \underline{t}^t X \\ \underline{y}_1 &= \underline{y} - \underline{t} \underline{t}^t \underline{y} \end{aligned} \quad \dots (3)$$

Pada metode PLS, banyaknya akar ciri dari matriks X^tYY^tX dinyatakan dengan N dan banyak komponen dinyatakan dengan K.

Langkah kedua dilakukan dengan cara yang sama dengan langkah pertama, tetapi menggunakan sisaan/*residu* setelah mengurangi komponen pertama dengan data asal seperti pada persamaan (2). Proses ini dilakukan secara iteratif hingga menghasilkan sejumlah A komponen. Jika $N > K$ maka proses akan terus berlanjut hingga $A = K$.

Setelah proses penentuan komponen PLS selesai, maka diperoleh matriks loading P dan vektor loading \underline{q} . Selanjutnya, vektor koefisien regresi yang digunakan dalam pendugaan PLS linier dapat dihitung dengan menggunakan persamaan:

$$\underline{\hat{b}} = \underline{\hat{W}} (\underline{\hat{P}}^t \underline{\hat{W}})^{-1} \underline{\hat{q}} \quad \dots (4)$$

dimana $\underline{\hat{W}}$ merupakan matriks dari vektor pembobot loading. Vektor koefisien regresi $\underline{\hat{b}}$ untuk selanjutnya dapat digunakan pada pendugaan model regresi PLS sebagai berikut:

$$\underline{\hat{y}} = \underline{\hat{b}}_0 + X^t \underline{\hat{b}} \quad \dots (5)$$

2.3. Validasi Model

Salah satu jenis validasi model yaitu dengan menghitung nilai *Root Mean Squared Error of Prediction* (RMSEP) dengan rumus:

$$RMSEP = \sqrt{\frac{1}{n_p} \sum_{i=1}^{n_p} (\hat{y}_i - y_i)^2} \quad \dots (6)$$

dimana:

- y_i = nilai pengamatan ke-*i* pada kelompok data validasi.
- \hat{y}_i = nilai dugaan pengamatan ke-*i*.
- n_p = banyak sampel yang digunakan dalam model validasi.

2.4. Genetic Algorithm (GA)

Hal utama dari GA adalah populasi dari makhluk hidup dan fungsi *fitness*. Pada bahasa komputer, makhluk hidup dinyatakan sebagai himpunan dari variabel. Setiap variabel dinyatakan dalam kromosom yang dicirikan dengan himpunan dari gen. Biasanya kromosom dikodekan dengan bilangan biner yang disebut bit, sehingga setiap bit berkorespondensi dengan gennya.

Pada penelitian ini, fungsi *fitness*nya adalah meminimumkan nilai *Root Mean Square Error of Cross Validation* (RMSECV). Pada prosesnya, n data pengamatan dibagi menjadi dua bagian, bagian yang pertama digunakan untuk membangun model dan bagian kedua digunakan untuk menguji model (validasi). Adapun rumus RMSECV adalah:

$$RMSECV = \sqrt{\frac{1}{N_{cal}} \sum_{n=1}^{N_{cal}} (\hat{y}_{cal,n} - y_{cal,n})^2} \quad \dots\dots (7)$$

dimana :

$y_{cal,n}$ = nilai pengamatan sebanyak N_{cal} sampel kalibrasi.

$\hat{y}_{cal,n}$ = nilai dugaan pengamatan yang diperoleh dengan prosedur *leave-one-out* yaitu menghilangkan sebanyak n sampel dari sekumpulan data kalibrasi.

N_{cal} = banyak sampel yang digunakan dalam model kalibrasi.

3. Metodologi Penelitian

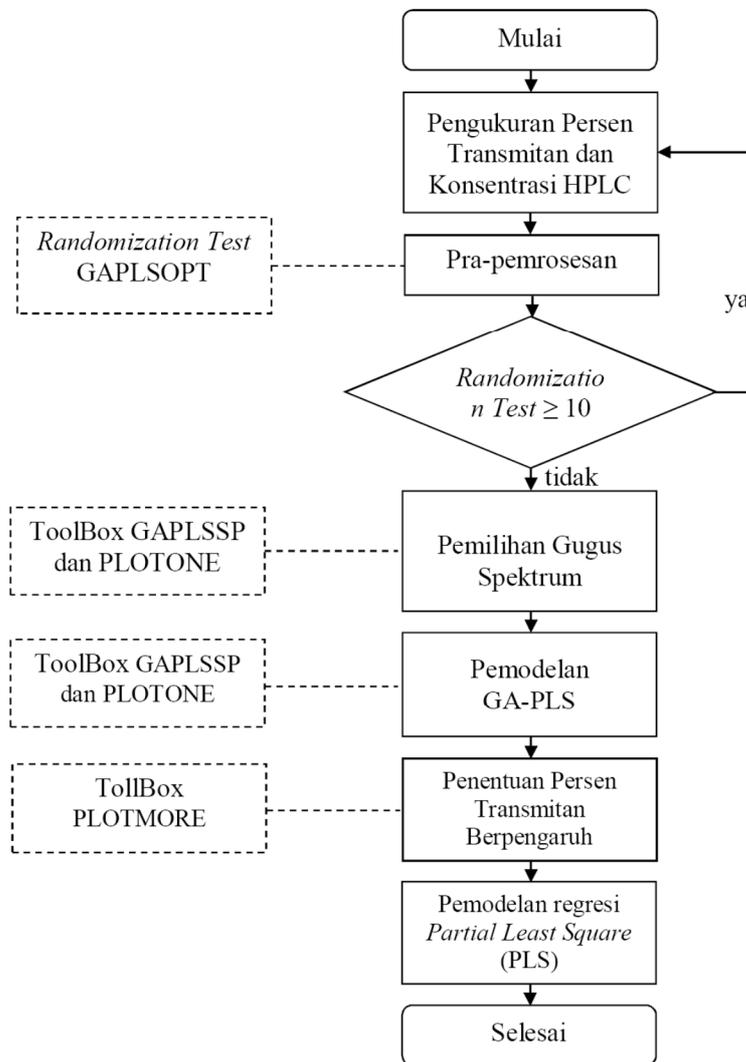
3.1. Bahan

Data yang digunakan dalam penelitian ini adalah data sekunder yang merupakan bagian dari data penelitian Hibah Pascasarjana tahun 2003-2005 hasil kerjasama antara Departemen Statistika IPB dengan Pusat Studi Biofarmaka LPPM IPB. Data yang digunakan adalah persen transmittan kurkumin dari serbuk temulawak hasil pengukuran 538spectrometer FTIR sebagai peubah bebas dan data konsentrasi senyawa aktif kurkumin yang diukur dengan menggunakan HPLC sebagai peubah respon. Temulawak yang dijadikan contoh diambil dari beberapa daerah sentra tanaman obat, yaitu: Bogor, Sukabumi, Kulon Progo, Karanganyar, Cianjur dan Balitro.

3.2. Metode Analisis

Penelitian ini terdiri atas 5 tahap analisis data, yaitu (1) tahap *Randomization Test* dan pra-pemrosesan, (2) tahap pemilihan gugus spektrum, (3) tahap pemilihan model GA-PLS, (4) tahap penentuan peubah bebas berpengaruh dan (5) tahap pemodelan PLSR.

Algoritma penelitian dapat dilihat pada Gambar 1 berikut :



Gambar 1. Diagram alur penelitian.

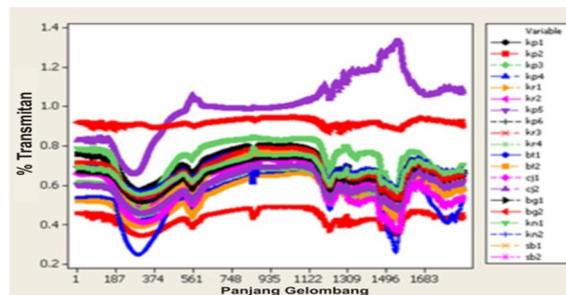
Pengolahan data pada tahap *preprocessing* dan tahap pemilihan model GA-PLS digunakan dalam bahasa pemrograman MATLAB, sedangkan pengolahan data pada tahap pemodelan PLSR digunakan dalam bahasa pemrograman SAS.

4. Hasil dan Pembahasan

4.1. Deskripsi Spektrum Kurkumin

Spektrum kurkumin diambil dari 20 sampel serbuk temulawak yang berasal dari berbagai daerah sentra tanaman obat dapat dilihat pada Gambar 2. Berdasarkan Gambar 2 spektrum kurkumin dari berbagai daerah tersebut sebagian besar memiliki pola yang hampir sama kecuali untuk beberapa spektrum yang menunjukkan pola yang agak berbeda. Terlihat bahwa spektrum kurkumin dari sampel serbuk temulawak yang diambil dari daerah Cianjur (sampel cj2) dan Bogor (sampel bg2) agak berbeda.

Data persen transmitan diperoleh dari pengukuran dengan menggunakan FTIR pada 1866 bilangan gelombang yang berkisar antara $4000 - 400 \text{ cm}^{-1}$. Pada indeks bilangan gelombang disekitar 1500 cm^{-1} ketika spektrum kurkumin serbuk temulawak dari sebagian besar sampel memiliki pola grafik yang cekung ke atas, temulawak yang diambil dari daerah Cianjur menunjukkan pola grafik yang cekung kebawah. Tampak juga bahwa spektrum kurkumin serbuk temulawak yang diambil dari daerah Bogor (sampel bg2) menunjukkan pola yang cenderung konstan di setiap bilangan gelombang. Sehingga layak diduga bahwa kedua spektrum kurkumin dari sampel cj2 dan bg2 merupakan spektrum pencilan (*outlier*). Pendeteksian adanya spektrum pencilan secara uji formal tidak akan dibahas pada bagian dari penelitian ini tetapi pada analisis data akan dilakukan terhadap data pengamatan tanpa pencilan yaitu dengan menggunakan 18 sampel serbuk temulawak.

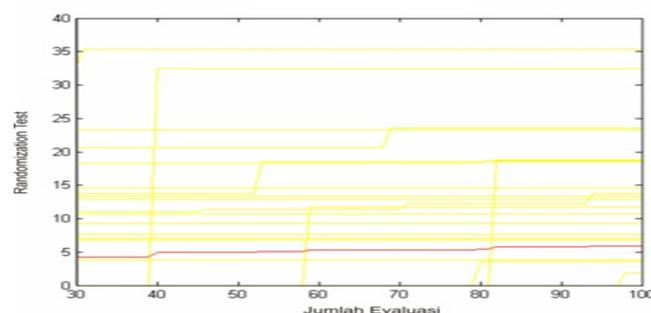


Gambar 2. Spektra kurkumin serbuk temulawak.

4.2. Randomization Test

Langkah awal yang dilakukan sebelum menggunakan metode GA adalah melakukan *randomization test*. Alasan utama mengapa perlunya dilakukan *randomization test* adalah untuk verifikasi apakah metode GA dapat digunakan atau tidak. Berdasarkan pengalaman Leardi (2000), jika hasil dari *randomization test* bernilai ≥ 10 , maka sangat tidak disarankan menggunakan metode GA karena akan mengakibatkan resiko untuk terjadinya *overfitting* yang cukup besar. Sehingga jika nilai dari *randomization test* ≥ 10 , maka disarankan untuk melakukan reduksi 'awal' peubah yaitu dengan menghitung rata-rata dari panjang gelombang yang saling berdekatan.

Pada proses ini, *randomization test* juga bertujuan untuk mengoptimalkan jumlah evaluasi pada setiap *running*. Gambar 3 menjelaskan bahwa hasil dari *randomization test* adalah sebesar 5.894 yang berarti nilainya kurang dari 10. Oleh karena itu, metode GA dapat diterapkan pada penelitian ini.

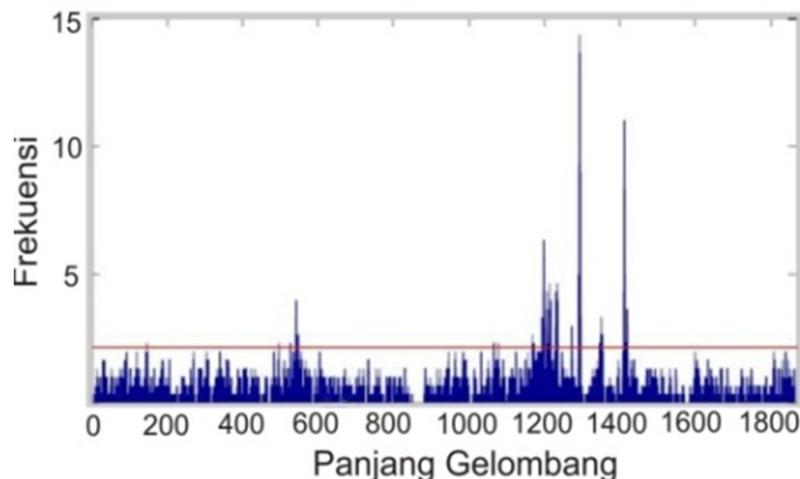


Gambar 3 Randomization Test.

4.3. Pemilihan Gugus Spektrum untuk Data dengan Pencilan

Data persen transmitan diukur pada 1866 bilangan gelombang yang dalam pemodelan kalibrasi ini berperan sebagai peubah bebas. Tidak seperti banyak metode penyeleksian peubah bebas yang diterapkan pada sekumpulan data spektrum, metode GA merupakan metode yang sangat berguna dalam menangani pemilihan panjang gelombang khususnya pada kalibrasi peubah ganda yang dibentuk berdasarkan model PLS. Pada metode GA-PLS, peubah-peubah bebas dipilih berdasarkan frekuensi dan karakteristik gugus spektrum. Sehingga akan diperoleh suatu model dugaan yang lebih baik dibandingkan dengan model spektrum-penuh (*full-spectrum model*) atau dengan melibatkan keseluruhan dari panjang gelombang. Begitu juga jika hanya melibatkan peubah-peubah bebas pada daerah sidik jari (*finger print*) saja.

Gambar 4 menunjukkan histogram dari frekuensi terpilihnya masing-masing panjang gelombang untuk data dengan pencilan. Pada Gambar 4 terlihat panjang gelombang mana saja yang dipilih lebih sering dan panjang gelombang mana saja yang hampir tidak pernah terpilih. Garis hijau horizontal menunjukkan *cutoff* untuk model dengan nilai *Root-Mean-Square Error of Cross-Validation* (RMSECV) minimum.



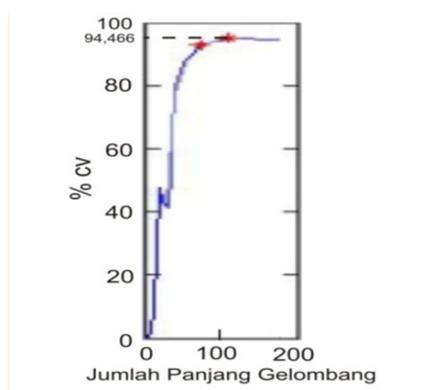
Gambar 4. Histogram frekuensi terpilihnya masing-masing peubah bebas dari data dengan pencilan.

Garis hijau juga menunjukkan *cutoff* untuk model dengan jumlah panjang gelombang paling sedikit diantara semua model dengan selisih RMSECV masing-masing model tersebut dengan RMSECV minimum tidak signifikan berdasarkan Uji F ($p < 0.1$). Selanjutnya, model dengan selisih RMSECV minimum ini disebut sebagai model terbaik berdasarkan kriteria Uji F.

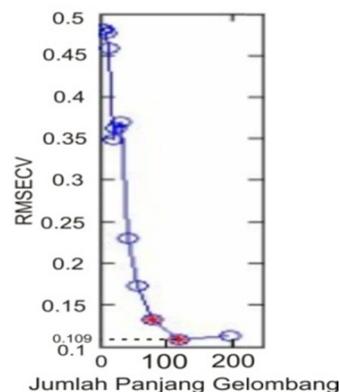
4.4. Pemodelan GA untuk Data dengan Pencilan

Pada penelitian ini, persentase CV yang menjelaskan keragaman sebuah fungsi dari sejumlah panjang gelombang yang disertakan dalam model dapat dilihat pada Gambar 5 berikut. Terlihat bahwa nilai CV meningkat seiring dengan banyaknya panjang gelombang yang ditambahkan dalam model. Pada akhirnya akan dihasilkan nilai CV maksimum dan selalu relatif konstan atau bahkan menurun. Bintang hijau mengindikasikan maksimum global dan bintang merah mengindikasikan nilai yang

berhubungan dengan uji F. Pada Gambar 5 terlihat bahwa nilai CV maksimum yaitu sebesar 94.466% yang berarti bahwa keragaman yang mampu dijelaskan sebesar 94.466% dari keragaman data asal.



Gambar 5. Persentase CV.



Gambar 6. Nilai RMSECV.

Pada Gambar 6 menunjukkan RMSECV sebagai sebuah fungsi dari jumlah panjang gelombang yang disertakan dalam model. Terlihat bahwa nilai RMSECV yaitu sebesar 0.109. Gambar 5 dan Gambar 6 secara bersama-sama digunakan untuk menentukan jumlah peubah bebas yang akan digunakan dalam model kalibrasi.

4.5. Pemilihan Model GA Terbaik untuk Data dengan Pencilan

Pada penelitian ini, kriteria evaluasi model GA terbaik adalah berdasarkan nilai RMSECV, jumlah gugus spektrum dan persentase CV yang nilainya dapat dilihat pada Tabel 2 dibawah ini. Dengan menganggap konsentrasi kurkumin yang merupakan hasil dari pengukuran HPLC sebagai peubah respon maka pemilihan model GA terbaik untuk data dengan pencilan dapat dilakukan.

Tabel 2. Matriks hasil dari *final stepwise* untuk data dengan pencilan

Jumlah Panjang Gelombang	Nilai % CV	Jumlah Gugus Spektral	RMSECV
1	0	1	0.482
3	0	1	0.482
4	0	3	0.482
31	41.123	8	0.369
43	77.046	10	0.231
58	87.102	7	0.173
79	92.512	15	0.132
121	94.466	14	0.109
196	94.288	15	0.113

Berdasarkan Tabel 2 dapat disimpulkan bahwa model yang relevan untuk pemodelan kalibrasi konsentrasi kurkumin untuk data dengan pencilan adalah model dengan jumlah panjang gelombang sebanyak 121 yang memberikan nilai CV sebesar 94.466%, nilai RMSECV sebesar 0.109 dan sebanyak 14 gugus spektrum. Nilai RMSECV tersebut bersesuaian dengan pengukuran persen transmittan pada bilangan gelombang 4000-400 cm^{-1} .

4.6. Pemodelan Regresi PLS

Pada bagian sebelumnya sudah dijelaskan bahwa model GA-PLS yang relevan bagi model kalibrasi konsentrasi kurkumin adalah model dengan kriteria persentase CV maksimum dan nilai RMSECV minimum berdasarkan karakteristik gugus spektrumnya. Selanjutnya dilakukan transformasi dari masing-masing jumlah panjang gelombang terpilih ke persen transmittan/absorban konsentrasi kurkumin yang bersesuaian dengan panjang gelombang sehingga akan menghasilkan sejumlah peubah bebas berpengaruh dalam model kalibrasi.

Implementasi regresi PLS dengan menggunakan sejumlah peubah bebas terpilih baik untuk data dengan semua panjang gelombang (*full-spectrum*), data dengan pencilan hasil dari transformasi ke panjang gelombang yang bersesuaian menghasilkan nilai RMSEP, berikut nilai RMSEP untuk semua model.

Tabel 3. Nilai RMSEP untuk semua model

Model Kalibrasi	20 Sampel (Dengan Pencilan)
1866 panjang gelombang (tanpa pra-pemrosesan GA)	0.713
GA <i>running</i> I	0.067 (121)
GA <i>running</i> II	0.059 (154)
GA <i>running</i> III	0.071 (117)
GA <i>running</i> IV	0.058 (143)
GA <i>running</i> V	0.061 (163)
Rata-rata GA	0.063

Keterangan: dalam tanda kurung menyatakan banyaknya peubah bebas yang disertakan dalam model kalibrasi

Berdasarkan Tabel 3 diperoleh nilai RMSEP dari model PLS tanpa menggunakan pra-pemrosesan GA untuk data dengan pencilan adalah 0.713. Nilai RMSEP rata-rata hasil metode GA untuk data dengan pencilan (20 sampel) yaitu sebesar 0.063. Terlihat bahwa nilai RMSEP hasil metode PLSR dengan pra-pemrosesan GA untuk data dengan pencilan jauh lebih baik jika dibandingkan dengan nilai RMSEP metode PLSR tanpa GA. Berdasarkan kriteria RMSEP dapat disimpulkan bahwa PLSR akan lebih baik dengan adanya pra-pemrosesan GA yang mampu mereduksi dan menyeleksi peubah bebas sehingga akan menghasilkan suatu model kalibrasi yang lebih baik jika dibandingkan dengan model PLSR tanpa pra-pemrosesan GA meskipun data mengandung pencilan (*outlier*).

5. Penutup

5.1. Kesimpulan

Berdasarkan hasil penelitian, dapat diambil beberapa kesimpulan sebagai berikut:

1. Berdasarkan kriteria CV, RMSECV serta jumlah gugus spektrum, diperoleh sejumlah persen transmittansi atau 121 peubah bebas yang memberikan pengaruh paling besar terhadap model kalibrasi PLSR.
2. Pemodelan regresi PLS dengan pra-pemrosesan GA yang melibatkan peubah bebas atau persen transmittansi berpengaruh berdasarkan karakteristik gugus spektrum memberikan kecenderungan hasil yang lebih baik jika dibandingkan dengan tanpa menggunakan pra-pemrosesan GA meski dengan data mengandung pencilaan (*outlier*) maupun model kalibrasi dengan melibatkan seluruh panjang gelombang (*full spectrum*).

6. Daftar Pustaka

- Andras F, Jenő H. 2008. Comparison Of Calibration Models Based On Near Infrared Spectroscopy Data for The Determination Of Plant Oil Properties. Hungary: Department of Hydrocarbon and Coal Processing, *Journal of Chemical and Process Engineering*, University of Pannonia.
- Atok RM. 2005. Jaringan Syaraf Tiruan dalam Pemodelan Kalibrasi dengan Pra-pemrosesan Analisis Komponen Utama dan Transformasi Fourier Diskrit [Tesis]. Bogor: Program Pascasarjana, Institut Pertanian Bogor.
- Braz J. 2007. Cross-Validation for The Selection of Spectral Variables Using The Successive Projections Algorithm. *Journal of The Brazilian Chemical Society*. Vol.18 no.8.
- Chen T, Morris J, Martin E. 2007. Gaussian Process Regression for Multivariate Spectroscopic Calibration. *Chemometrics and Intelligent Laboratory Systems* 87: 85-97.
- Eko WP. 2009. Penerapan Beberapa Metode Eksploratif untuk Menguji Kesamaan Pola FTIR (Studi Kasus: Senyawa Aktif Kurkuminoid Temulawak). [Tesis]. Bogor: Program Pascasarjana, Institut Pertanian Bogor.
- Erfiani. 2005. Pengembangan Model Kalibrasi dengan Pendekatan Bayes (Kasus Tanaman Obat [Disertasi]. Bogor: Program Pascasarjana, Institut Pertanian Bogor.
- Leardi R. 2000. Application of Genetic Algorithm-PLS for Feature Selection in Spectral Data Sets. *Journal of Chemometrics Volume* 14, Pp: 643-655. Jhon Wiley & Sons.
- Markus L. 1996. Comparison of PLS, PCR and MLR for the Quantitative Determination of Foreign Oils and Fats in Butter Fats of Several European Countries by Their Triglyceride Composition. *Journal of Zeitschrift für Lebensmitteluntersuchung und -Forschung A*, Vol 202 pp 193-198.
- Naes T, Issackson T, Fearn T, Davies T. 2002. *User Friendly Guide to Multivariate Calibration and Classification*. United Kingdom: NIR Publication Chichester.
- Nur MA, Adijuwana H. 1989. *Teknik Spektroskopi dalam Analisis Biologi*. Bogor: Institut Pertanian Bogor.
- Otto M. 1999. *Chemometrics: Statistics and Computer Application in Analytical Chemistry*. Clarkson University.
- Sunaryo S. 2005. Model Kalibrasi dengan Transformasi Wavelet sebagai Metode Pra-Pemrosesan [Disertasi]. Bogor: Program Pascasarjana, Institut Pertanian Bogor.
- Sinambela JM. 1985. Fitoterapi, Fitostandar dan Temulawak. Prosiding Simposium Nasional Temulawak. Bandung, 17 september 1985. Lembaga Penelitian Universitas Padjajaran.