

THE EXPLORATION OF THE POTENCY OF FTIR SPECTROSCOPY AND CHEMOMETRICS FOR EXTRACT BIOACTIVITY PREDICTION AND RAPID ANALYSIS OF BIOACTIVE CONTENT OF MEDICINAL PLANTS

Darusman, LK^{1,3}, B. Susetyo², KA Notodiputro², AA Mattjik²,
R. Heryanto^{1,3} Erfiani² & Jajaung⁴

¹ Departemen Kimia FMIPA IPB; Jl Pajajaran Bogor 16151; Telp/Fax 0251-381949

² Departemen Statistika FMIPA IPB; Kampus IPB Darmaga Bogor

³ Pusat Studi Biofarmaka – LPPM IPB; Kampus IPB Taman Kencana Jl Taman Kencana No 3; Telp 0251-373561 Fax 0251-347525 Bogor 16151

⁴ Jurusan Matematika FMIPA Universitas Soedirman, Purwokerto

Abstract

An infrared spectra is a characteristic representation for a chemistry system. There are many researches have been conducted for the interpretation supported by chemometric method which aimed for metabolite fingerprinting. Nowadays, the application of IR spectroscopy and chemometric for quantitative analysis of mixture system is revealed to be increasing because of ease of use, speed, reliability, sample size, and calibration stability. The paper is about the use of FTIR spectroscopy technique and chemometric for building a model of bioactivity prediction of jatibelanda extract (extract inhibition to the lipase enzyme activity) and generating calibration model for determination gingerol in powder and ginger extract. Both of them were made by three steps, i.e. the data reduction for reducing the complexity of data structure of infrared spectra, the build of calibration model, and the model validation. The data reduction technique consisted of principal component analysis (PCA) dan fragmented regression. The generating calibration model and validation method used artificial neural network (ANN) dan Bayes approach. The ANN bioactivity model can make well-mannered of grouping of inhibition characteristic of jatibelanda extract and the calibration model of gingerol by the Bayes approach resulted high precise prediction.

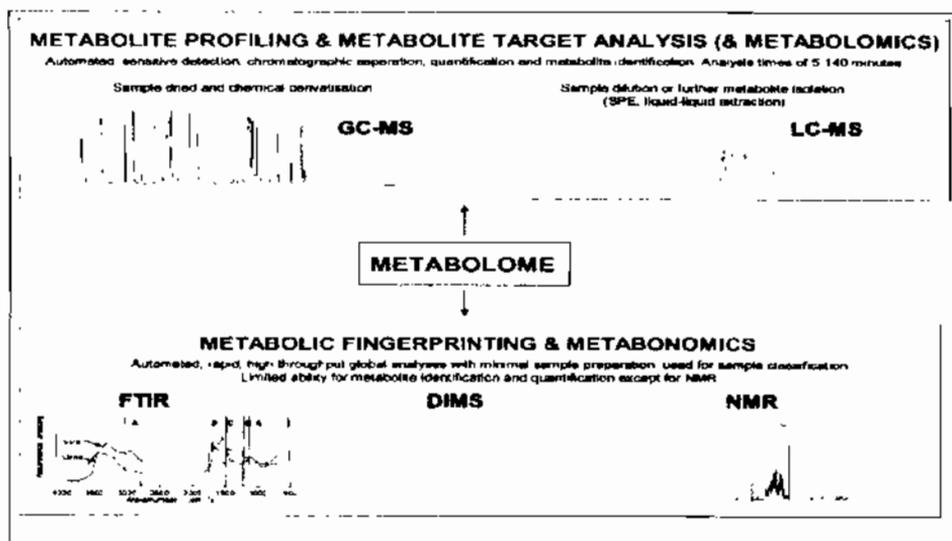
Abstrak

Spektra inframerah merupakan representasi khas sistem kimia yang diuji. Interpretasinya dengan bantuan metode kemometrik untuk tujuan *metabolite fingerprinting* dalam suatu sistem kimia telah banyak dilakukan. Fungsi lain dari kombinasi spektroskopi inframerah dan kemometrik untuk penentuan kuantitatif suatu komponen kimia di dalam sistem campuran mulai banyak mendapat perhatian karena preparasi sampelnya mudah, cepat dan tingkat keterulangannya yang baik. Makalah ini mengemukakan penggunaan teknik spektroskopi FTIR dan kemometrik untuk membentuk model penduga bioaktivitas berbagai ekstrak jatibelanda (inhibisi ekstrak terhadap aktivitas enzim lipase) dan model kalibrasi penentuan gingerol dari serbuk dan ekstrak jaha. Baik model penduga bioaktivitas maupun model kalibrasi dibentuk melalui tiga tahap proses yaitu reduksi data untuk mengurangi kompleksitas struktur data spektra inframerah, pembentukan model, dan validasi model. Teknik reduksi data yang digunakan terdiri dari analisis komponen utama (Principal component analysis, PCA) dan regresi terpenggal, sedangkan untuk metode pembentukan dan validasi model digunakan metode sistem jaringan syaraf tiruan (Artificial Neural Network, ANN) dan pendekatan bayes. Model ANN penduga bioaktivitas dapat mengelompokkan sifat inhibisi ekstrak jatibelanda dengan baik dan model kalibrasi gingerol dengan pendekatan Bayes menghasilkan ketepatan pendugaan relatif tinggi.

Pendahuluan

Metabolisme tumbuhan merupakan serangkaian proses yang dapat menghasilkan beragam komponen kimia yang dapat dimanfaatkan untuk pangan, papan maupun obat-obatan. Penggambaran keterkaitan antara proses metabolisme, kekhasan dari komponen kimia yang dihasilkannya serta fungsinya dilakukan oleh sains 'omics' yang dikembangkan pasca era genomic. Salah satu sains 'omics' yang bertujuan untuk mempelajari dan menentukan baik secara kualitatif maupun kuantitatif seluruh senyawa berbobot molekul rendah (metabolit) yang dibutuhkan untuk memelihara dan menjaga sel untuk tumbuh dan berfungsi secara normal (*metabolome*) disebut sebagai *metabolomics*.

Terdapat beberapa strategi dalam metabolomics untuk mempelajari metabolome dalam suatu sistem biologis, baik itu untuk kuantifikasi metabolit maupun hanya untuk klasifikasi sampel (Gambar 1). Strategi tersebut adalah *metabolite profiling*, *metabolite target analysis*, *metabolite fingerprinting* dan *metabonomics*. Dua strategi pertama utamanya digunakan untuk indentifikasi dan kuantifikasi metabolit sedangkan dua yang terakhir terutama digunakan untuk klasifikasi sampel.



Gambar 1. Ringkasan strategi kajian metabolomics (Trends in Analytical Chemistry, Vol. 24, No. 4, 2005)

Keempat strategi ini diimplementasikan melalui beberapa teknik analisis seperti kromatografi, spektroskopi, spektrometri massa dan kombinasi diantara ketiganya. Salah satu teknik spektroskopi yang umum digunakan dalam kajian metabolomics adalah FTIR (Fourier Transform Infrared). FTIR menggunakan karakteristik vibrasi dalam molekul untuk menghasilkan suatu spektra sidik jari yang fitur-fiturnya didefinisikan dari gugus fungsi yang ada dalam sampel. Adanya kekhasan dalam spektra FTIR dari suatu sampel yang disertai dengan kemudahan, kecepatan serta reproduibilitas data yang dihasilkannya, menjadikan FTIR sebagai metode pilihan dalam klasifikasi sampel berdasarkan asal atau sifat biologisnya.

Dalam makalah ini akan dicoba diketengahkan penggunaan teknik spektroskopi FTIR untuk *metabolite target analysis* dan *metabolite fingerprinting*. Karakteristik sidik jari spektra FTIR ekstrak ekstrak jatibelanda akan dikaitkan dengan bioaktivitasnya (inhibisi ekstrak terhadap

aktivitas enzim lipase) dan spektra FTIR jahe akan digunakan untuk memprediksi kandungan gingerolnya.

Spektra FTIR untuk *metabolite target analysis* dan *metabolite fingerprinting*

Salah satu ciri dari sains 'omics' adalah adanya suatu seri data yang kompleks yang diambil (merupakan respons) dari seluruh komponen kimia dalam sampel sehingga menjadi gambaran kondisi utuh suatu sampel. Demikian pula data yang terdapat dalam spektra FTIR yang diambil dari sel tumbuhan utuh, Spektra FTIRnya merupakan suatu set data multivariat yang besar. Sehingga untuk dapat digunakan sebagai pola sidik jari ataupun untuk kuantifikasi metabolit, data spektra yang ada tidak dapat hanya diinterpretasikan secara visual. Diperlukan serangkaian teknik kemometrik untuk bisa mengubah spektra FTIR menjadi informasi yang diinginkan.

Prosedur umum dalam memanfaatkan spektra FTIR untuk *metabolite target analysis* dan *metabolite fingerprinting* diawali dengan mengembangkan suatu model kalibrasi berdasarkan teknik kemometrik yang dipilih. Spektra FTIR dikumpulkan dari setiap sampel yang dianalisis beserta dengan informasi rujukannya. Informasi rujukan yang bersifat kuantitatif ditentukan berdasarkan metode rujukan yang ada yang telah diakui, sedangkan informasi rujukan yang bersifat kualitatif dapat berasal dari analisis non FTIR sebelumnya atau didapat dari analisis spektra FTIR dengan teknik kemometrik pengenalan pola. Spektra yang terkumpul dibagi dua menjadi data uji dan data latihan. Model kalibrasi dibangun dari data latihan dan informasi rujukan yang ada. Tujuan dari dibentuknya model kalibrasi ini adalah untuk mejadi model penduga nilai kuantitatif atau sifat kualitatif suatu sampel lain diluar data yang digunakan untuk membangun model. Terakhir, untuk menguji keterandalan model kalibrasi dilakukan validasi menggunakan data uji yang tersedia. Garis besar proses pengembangan model ini ditunjukkan dalam Tabel 1.

Tabel 1 Pengembangan model kalibrasi berbasis spektra FTIR

Preparasi

- Mengumpulkan sampel yang representatif
- Melakukan analisis kuantitatif rujukan (dengan HPLC (*High Performance Liquid Chromatography*), GC (*Gas Chromatography*) dan lainnya) atau mengumpulkan Informasi kualitatif rujukan (asal geografis, kondisi lingkungan, kondisi nutrisi, bioaktivitas dan lainnya)
- Melakukan pengukuran spektra FTIR sampel
- Membagi data yang tersedia menjadi data kalibrasi (data latihan) dan data validasi (data uji)
- Meningkatkan kualitas spektra dengan melakukan *preprocessing* spektra menggunakan teknik *smoothing*, *derivatisasi*, *wavelet* dan lainnya

Kalibrasi

- Memilih metode pemodelan (PLS, ANN, Bayes dan yang lainnya)
- Membangun model dengan parameter terbaik (R^2 , RMSEP dan lainnya)
- Memprediksi nilai kuantitatif dan sifat kualitatif dengan model yang dibangun dan bandingkan dengan hasil analisis/info rujukan
- Mengidentifikasi pencilan dan membetulkan atau menghilangkannya
- Mengkalibrasi ulang model
- Mengevaluasi galat dan melakukan validasi silang

Validasi

- Memprediksi nilai kuantitatif dan sifat kualitatif dengan model yang dibangun dan bandingkan dengan hasil analisis/info rujukan
- Mengevaluasi galat dan melakukan validasi silang

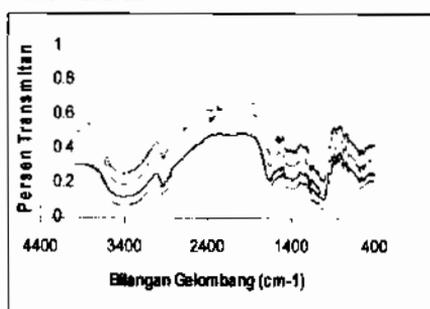
Model Kalibrasi Gingerol Menggunakan Pendekatan Bayes Hirarki

Proses penentuan konsentrasi senyawa aktif atau senyawa penciri yang dikandung oleh suatu tanaman obat perlu dilakukan secara cepat dan akurat. Untuk itu sangat diperlukan metode yang handal tetapi relatif mudah untuk dioperasikan. Secara kualitatif dan kuantitatif suatu senyawa aktif dapat diketahui antara lain melalui metode HPLC (*High Performance Liquid Chromatography*). Proses ini memerlukan waktu yang relatif lama dan biaya yang relatif mahal. Alternatif cara penentuan lain yang dapat dilakukan adalah dengan mengembangkan model kalibrasi yang menyatakan hubungan antara kandungan senyawa aktif atau penciri hasil pengukuran HPLC dengan data hasil pengukuran FTIR.

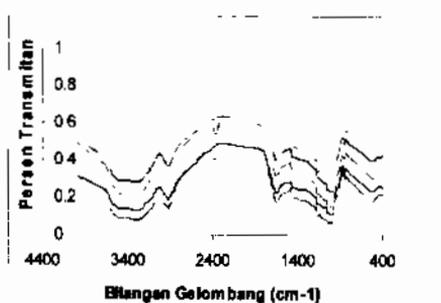
Pada penelitian ini dilakukan kajian penerapan pendekatan Bayes untuk pembentukan model kalibrasi gingerol. Pendekatan Bayes merupakan suatu alternatif untuk mengatasi masalah kekolinearitas karena dalam pendekatan ini informasi baru ditambahkan kedalam model dengan cara menganggap bahwa parameter model berasal dari sebaran tertentu sehingga tidak bersifat deterministik. Data yang digunakan adalah data pengamatan 20 sampel serbuk jahe. Kedua puluh sampel diukur spektra FTIRnya dan dianalisis kadar gingerolnya dengan metode HPLC.

Data keluaran FTIR berupa persen transmisi dan bilangan gelombang umumnya berupa matriks data yang berukuran besar. Keterbatasan perangkat lunak yang tersedia serta untuk mempercepat proses, data keluaran FTIR terlebih dulu direduksi menggunakan pendekatan regresi terpenggal. Model kalibrasi antara data hasil reduksi dengan data konsentrasi keluaran HPLC menggunakan pendekatan Bayes hirarki dengan perilaku parameter model β berhirarki dan σ acak. Pendekatan ini merupakan hasil terbaik kajian simulasi pendekatan Bayes untuk data dengan besaran ukuran contoh (n) jauh lebih besar dari jumlah peubah bebas (p) serta terdapat multikolinier antara peubah bebas. Ukuran kebaikan model yang digunakan adalah Jumlah Kuadrat Galat (JKG) dan *Root Mean Square Error of prediction* (RMSEP). Perangkat lunak yang digunakan adalah Winbugs14.

Pengukuran serbuk jahe menggunakan FTIR menghasilkan data transmisi setiap spektrumnya sebanyak 1866 titik. Gambar 2 menyajikan spektrum hasil keluaran FTIR untuk 20 contoh serbuk jahe yang diamati. Secara umum terlihat bahwa keseluruhan spektrum memiliki pola yang sama. Sehingga untuk keseluruhan contoh spektra serbuk jahe dapat dimodelkan dengan hanya satu model kalibrasi.



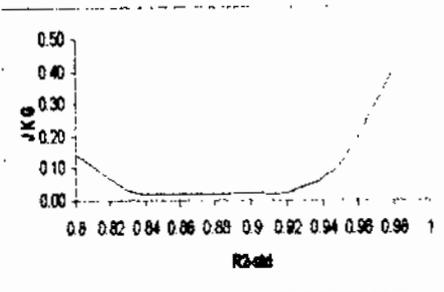
Gambar 2 Spektrum serbuk Gingerol 20 contoh



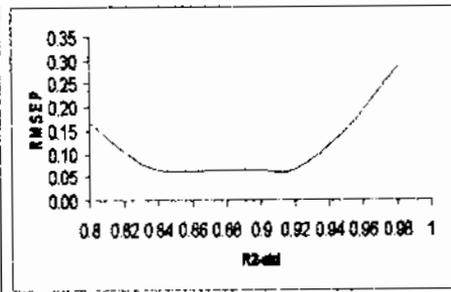
Gambar 3 Spektrum serbuk Gingerol 20 contoh hasil reduksi dengan Regresi Terpenggal $R^2_{std} = 0.999$

Sebelum dibentuk menjadi model kalibrasi, data spektra serbuk jahe direduksi dahulu dengan menggunakan pendekatan regresi terpenggal. Konsep pendekatan ini yaitu dengan membuat sekatan dengan setiap sekatan membentuk suatu persamaan regresi linier sederhana. Kriteria penentuan jumlah titik pada setiap partisi yaitu berdasarkan besaran koefisien Determinasi (R^2) yang diperoleh dengan menggunakan regresi linier sederhana tersebut. Jumlah titik yang berada

dalam satu partisi akan ditentukan oleh besaran R^2 yang ditetapkan (R^2_{std}). Semakin tinggi nilai penetapan (R^2_{std}), akan semakin kecil jumlah titik yang dihasilkan untuk setiap partisi. Gambar 2 spektrum hasil reduksi menggunakan pendekatan regresi terpenggal untuk R^2_{std} sebesar 0.999. Pendekatan ini mereduksi jumlah titik persen transmitan yang semula 1866 titik menjadi 108 titik. Salah satu keuntungan penggunaan pendekatan regresi terpenggal adalah hasil pereduksian tetap mempertahankan pola spektrum awal. Hal ini terlihat pada Gambar 3 yang memiliki pola relatif sama dengan Gambar 2. Penyusunan model menggunakan 15 contoh serbuk Gingerol, sedangkan 5 contoh lainnya digunakan untuk validasi model. Penentuan contoh untuk penyusunan model dan validasi model dilakukan secara acak.



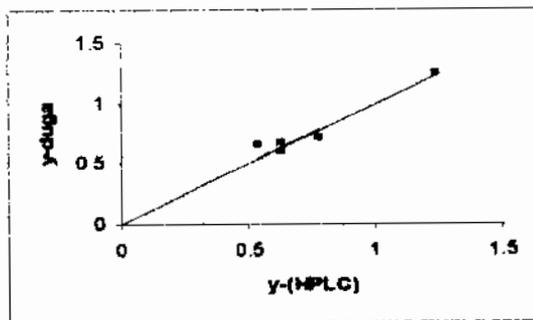
Gambar 4. Pola JKG pada seluruh model kalibrasi Gingerol



Gambar 5. Pola RMSEP pada seluruh model kalibrasi Gingerol

Gambar 4 dan Gambar 5 menyajikan besaran JKG dan RMSEP yang dihasilkan pada setiap R^2_{std} yang dicobakan. Hasil yang diperoleh menunjukkan bahwa model kalibrasi dengan besaran antara $R^2_{std} = 0.84$ sampai $R^2_{std} = 0.9$ memiliki besaran JKG dan RMSEP relatif kecil dibandingkan lainnya.

Hasil pendugaan 5 contoh menggunakan model kalibrasi dengan $R^2_{std} = 0.85$, nilai RMSEP yang dihasilkan sebesar 0.0622. Gambar 6 menampilkan plot antara nilai dugaan y (y -duga) dan y pengamatan keluaran HPLC (y -HPLC). Plot tersebut mendekati persamaan y -duga = y -HPLC, artinya model kalibrasi yang digunakan dapat melakukan pendugaan konsentrasi Gingerol dengan sangat baik.



Gambar 6 Plot y -duga dan y -(HPLC) Gingerol

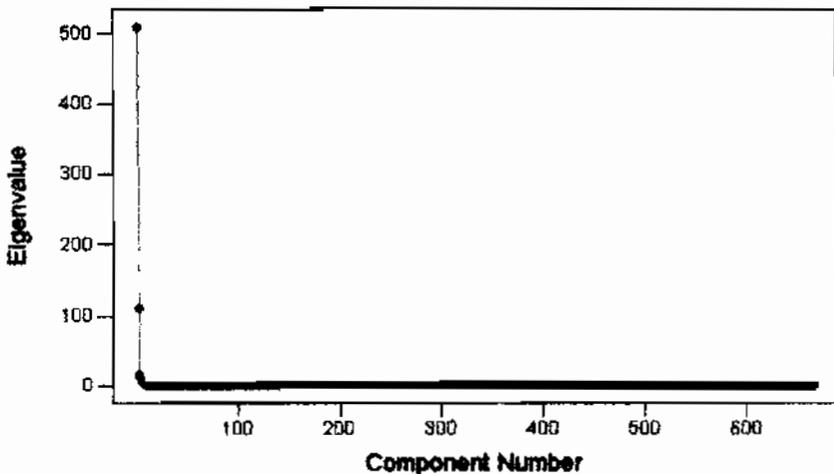
Model Prediksi Inhibisi Ekstrak Jati Belanda Terhadap Aktivitas Enzim Lipase

Bioaktivitas suatu ekstrak ditentukan oleh komposisi kimia penyusunnya. Gambaran dari komponen kimia penyusun ekstrak tersebut dapat diberikan oleh spektra FTIRnya. Bagaimana menghubungkan antara karakteristik spektra FTIR ekstrak dan sifat bioaktivitasnya dapat dilakukan dengan bantuan teknik kemometrik. Penelitian ini mencoba membentuk model yang mampu mengelompokkan beberapa ekstrak yang aktif terhadap kerja enzim lipase dari daun Jati Belanda berdasarkan spektra IR-nya sehingga ketika ada spektra baru dapat diduga kedekatan polanya dengan salah satu ekstrak yang terlebih dahulu sudah diketahui khasiatnya.

Enam buah ekstrak yang berasal dari daun jati belanda (ekstrak FN Butanol, FE Asetat, Kloroform, Heksana, Triterpenoid dan Steroid) diukur spektra FTIRnya dan ditentukan aktivitasnya terhadap kerja enzim lipase. Reduksi data FTIR dilakukan dengan pengambilan data berselang. Pengelompokan ekstrak dilakukan dengan analisis komponen utama (AKU) dan pemodelan untuk analisis diskriminan dilakukan dengan jaringan syaraf tiruan (ANN).

Data yang digunakan untuk AKU dan ANN hanya digunakan 48 (80%) spektra IR untuk pembentukan model (biasanya diistilahkan dengan *training*), sisanya yakni 12 (20%) spektra IR digunakan untuk pengujian model (diistilahkan dengan *testing*). Matriks yang digunakan untuk AKU adalah matriks korelasi, ini dikarenakan antar peubah mempunyai nilai persen transmisi dengan tingkat keragaman yang cukup tinggi. Tingginya tingkat keragaman ini dapat dilihat dari penampakan plot persen transmisi (%T) dan bilangan gelombang.

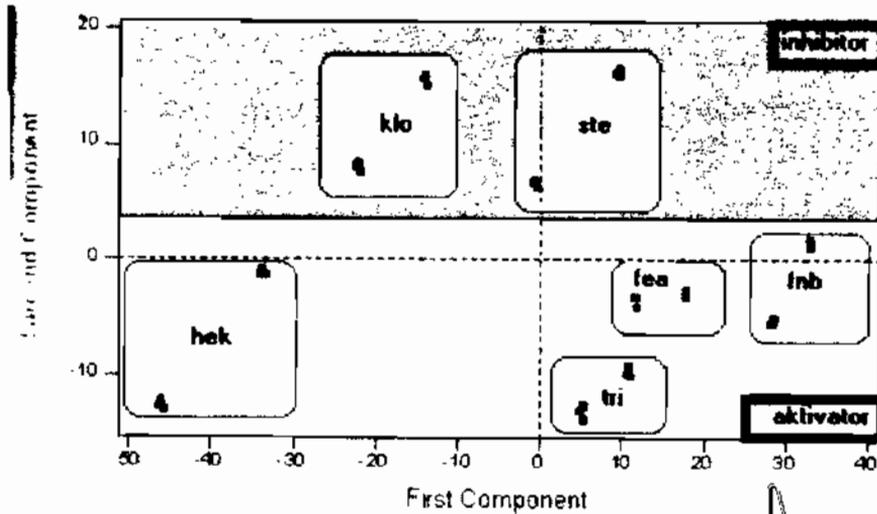
Hasil AKU dengan menggunakan matriks korelasi menunjukkan bahwa terjadi korelasi yang cukup tinggi diantara peubah persen transmisi (%T). Tingginya korelasi antar peubah dapat dilihat dari plot akar ciri dengan banyaknya komponen utama (Gambar 7). Plot akar ciri dan banyaknya komponen utama memperlihatkan garis yang menurun tajam (vertikal) dari akar ciri pertama ke akar ciri kedua dan dari akar ciri kedua ke akar ciri ketiga, pada akar ciri ketiga, keempat dan seterusnya, garis ini terlihat landai dan lurus mendatar.



Berdasarkan nilai-nilai akar ciri seperti terlihat pada Gambar 7, nilai-nilai akar ciri pertama dan kedua terlihat sangat tinggi yaitu masing-masing sebesar 551.78 dan 93.95 sedangkan nilai akar ciri ketiga, keempat dan seterusnya relatif sangat kecil dibandingkan dengan akar ciri pertama dan

Dengan demikian hanya digunakan dua komponen utama pertama untuk menggambarkan karakteristik dari spectra ke enam ekstrak yang diuji.

Kelompok-kelompok ekstrak melalui plot komponen utama 1 dan komponen utama 2 memperlihatkan adanya pemisahan ekstrak, yang bila dihubungkan dengan aktivitasnya terhadap enzim lipase terdapat dua kelompok besar yaitu kelompok ekstrak yang bersifat inhibitor dan kelompok yang bersifat aktivator (Gambar 8).



Gambar 8. Plot komponen utama 1 dan komponen utama 2 serta kaitan dengan aktivitasnya.

Untuk memperoleh fungsi yang dapat membedakan keenam jenis ekstrak dan pendugaan terhadap spektra ekstrak, dibuat suatu model ANN dengan menggunakan metode Quasi-Newton (QN). Dengan berdasarkan pada proporsi kumulatif dari akar ciri pertama dan kedua, maka untuk pembuatan model ANN hanya digunakan dua komponen utama, yaitu komponen utama 1 dan komponen utama 2.

Model dugaan dibentuk dengan menggunakan metode Quasi-Newton, menggunakan Software SAS 6.12 dengan memanfaatkan fasilitas Makro *Neural Networks*. Untuk kasus ini digunakan tingkat kesalahan 0.0001 dengan menggunakan metode QUANEW (Quasi-Newton) dan fungsi aktivasi yang digunakan adalah fungsi sigmoid. Untuk mencari model yang optimal dengan hanya menggunakan dua peubah bebas (komponen utama), dilakukan dengan menggunakan banyaknya unit *hidden* yang bervariasi, yaitu dari mulai 3 unit *hidden* sampai 10 unit *hidden*. Perbandingan Jumlah Kuadrat Galat (JKG), Kuadrat Tengah Galat (KTG) dan akar Kuadrat Kuadrat Galat (KTG) hasil ANN menggunakan 3 sampai 10 unit *hidden* disajikan pada Tabel 2. Berdasarkan Tabel 2, KTG dan akar KTG terkecil adalah model ANN dengan menggunakan 10 unit *hidden*, sehingga untuk model ANN ini dipilih 10 unit *hidden*.

Tabel 2. Perbandingan JKG, KTG dan akar KTG hasil ANN dengan menggunakan 3 sampai 10 unit *hidden*

Banyaknya unit <i>hidden</i>	JKG	KTG	Akar KTG
3	7.81221	0.030636	0.17503
4	5.33349	0.021681	0.14724
5	0.24855	0.0010403	0.032254
6	4.0000	0.017544	0.13245
7	7.81221	0.035672	0.18887
8	7.81219	0.037201	0.19288
9	0.26013	0.0012942	0.035975
10	0.00015497	0.0000008712	0.000894

Untuk menguji keterandalan dari model yang ada selanjutnya dicobakan dengan menggunakan data baru (data *testing*, yaitu data yang tidak dilibatkan dalam pembentukan model). Kesesuaian nilai-nilai aktual dan dugaan serta nilai sisanya berturut-turut dapat disajikan pada Tabel 3 dan Tabel 4.

Tabel 3. Kesesuaian nilai dugaan dan aktual keenam jenis ekstrak berdasarkan data *testing*

		Aktual					
		FN Butanol	FE Asetat	Kloroform	Heksana	Triterpenoid	Steroid
Dugaan	FN Butanol	0.9979	0.0017	0.0000	0.0000	0.0000	0.0000
	FE Butanol	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	FE Asetat	0.0002	0.9841	0.0000	0.0000	0.0000	0.0000
	FE Asetat	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
	Kloroform	0.0000	0.0000	1.0000	0.3682	0.0000	0.0000
	Kloroform	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
	Heksana	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
	Heksana	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
	Triterpenoid	0.0000	0.0000	0.0000	0.0041	0.9998	0.0000
	Triterpenoid	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
	Steroid	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
	Steroid	0.0000	0.0000	0.0773	0.0000	0.0000	0.8189

Tabel 4. Nilai sisaan (selisih aktual dan dugaan) keenam jenis ekstrak berdasarkan data *testing*

		Aktual					
		FN Butanol	FE Asetat	Kloroform	Heksana	Triterpenoid	Steroid
Dugaan	FN Butanol	0.0021	-0.0017	0.0000	0.0000	0.0000	0.0000
	FN Butanol	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	FE Asetat	-0.0002	0.0159	0.0000	0.0000	0.0000	0.0000
	FE Asetat	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Kloroform	0.0000	0.0000	0.0000	-0.3682	0.0000	0.0000
	Kloroform	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Heksana	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Heksana	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Triterpenoid	0.0000	0.0000	0.0000	-0.0041	0.0002	0.0000
	Triterpenoid	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Steroid	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Steroid	0.0000	0.0000	-0.0773	0.0000	0.0000	0.3811