

## ESTIMASI HUBUNGAN KUANTITATIF STRUKTUR-AKTIVITAS (HKSA) MENGUNAKAN ARTIFICIAL NEURAL NETWORKS (ANN)

Supriyanto

Program Studi Matematika, Fakultas Sains dan Teknik, Universitas Jenderal Soedirman  
Purwokerto

### ABSTRACT

The Quantitative structure-Activity Relationship (QSAR) study has been performed on pattern of structure-molecule relationship. Artificial Neural Network (ANN) model used to estimate pattern of enzyme activity structure-molecule and atomic location in three-dimension for compound of flavonoid as the predictors. Value of determination coefficient used to compare between actual value and value of estimating by ANN models based on enzyme's wavelength, so resulting cross validating is obtained. We use Quasy-Newton algorithm with Broyden-Fletcher-Goldfarb-Shanno (BFGS) procedure.

Keywords : QSAR, PCR, ANN, BFGS procedure, Quasy-Newton.

### PENDAHULUAN

Sebuah enzim mengikat sebuah molekul substrat dalam sebuah daerah yang dinamakan sisi aktif. Substrat diikat pada sisi aktif enzim oleh keadaan kimia dan fisika asam-asam amino pada enzim, seperti interaksi hidrofobik, interaksi ionik antara gugus berlawanan pada enzim dan substrat, ikatan hidrogen antara asam amino yang hidrofilik dan substrat polar (Stock 1991, dalam Putro 2005). Karena bentuk 3 dimensi yang khas pada sisi aktif dan keadaan alamiah proses pengikatan substrat, enzim hanya mampu mengikat substrat dengan struktur yang cocok.

Hubungan Kuantitatif Struktur-Aktivitas (HKSA) adalah kajian yang mencoba menghubungkan struktur atau keadaan suatu senyawa dengan aktivitasnya (Bevan tanpa tahun). Kajian HKSA mencoba menemukan hubungan tersendiri antara keragaman dalam nilai keadaan molekul dan aktivitas hayati dari sederet senyawa sehingga hubungan tersebut dapat dijadikan alat untuk mengevaluasi senyawa baru.

*Chemometrics* (pemodelan dalam bidang kimia), seringkali melibatkan variabel yang cukup besar (banyak),

sehingga model-model umum seperti analisis regresi linear berganda (*multiple linear regression*) tidak lagi relevan karena tidak terpenuhinya asumsi, misalnya asumsi multikolinearitas. Disamping asumsi multikolinearitas, seringkali ukuran variabel lebih besar dari pengamatan. Oleh karena analisis klasik seperti analisis regresi berganda tidak mungkin untuk diterapkan secara langsung. Salah satu alternatif dalam pemecahan permasalahan ini adalah ANN.

Model ANN merupakan salah satu pemodelan *free distribution* yang sangat fleksibel dalam hal distribusi dan skala variabel. Model ini bisa digunakan untuk kasus-kasus model linier dan non linier. Selain itu juga model ini dapat diterapkan untuk kasus-kasus data yang melibatkan banyak variabel termasuk dalam hal mengatasi multikolinearitas.

Penelitian ini bertujuan untuk mengestimasi nilai (besarnya) aktivitas enzim dengan posisi atom pada koordinat kartesius 3 dimensi dalam keadaan stabil (teroptimasi) pada senyawa turunan flavonoid. Model yang baik adalah model yang mampu mem-*fit* data sehingga menghasilkan nilai estimasi

yang mendekati nilai yang sebenarnya atau nilai aktual.

## METODE PENELITIAN

### Materi dan Alat

Data yang digunakan pada penelitian ini adalah data panjang gelombang dari spektrum Ultra Violet - Visibel dari Isolat pada konsentrasi 0.01g / 10 mL methanol (Asih dan Setiawan, 2008). Analisis data dan pemodelan komputasinya menggunakan PC Pentium K5 2,6GB, RAM 1GB, Harddisk 180 GB dengan bantun software Matlab Ver 6.1.

### Tahapan Penelitian

Tahap awal adalah menyusun data dalam bentuk row data. Posisi relatif setiap atom dalam ruang tiga dimensi diperlakukan sebagai variabel bebas. Untuk setiap satu atom diperlakukan tiga variabel bebas, misalkan atom ke- $i$  terletak pada posisi  $x_{i1}, y_{i1}, z_{i1}$ . Jadi jika terdapat 15 atom, maka terdapat 45 variabel bebas. Dengan demikian diperoleh struktur data berukuran 55 x 46, yaitu 55 observasi, 45 variabel bebas, dan 1 variabel respon (besarnya nilai panjang gelombang).

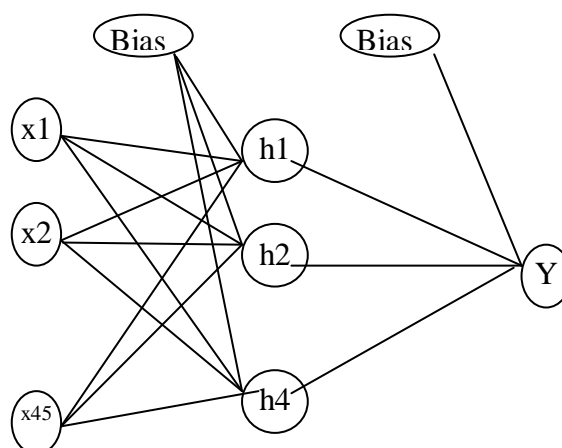
## HASIL DAN PEMBAHASAN

### Analisis Neural Networks (ANN)

Model yang handal mempunyai tingkat keakuratan tinggi dan tentunya mempunyai kesalahan (*error*) yang paling minimum. Pemodelan melalui ANN merupakan model yang estimasi parameternya (pembobot) dilakukan secara iterasi dan bentuk model ini umumnya dinyatakan dalam bentuk pemetaan (*mapping*). Algoritma pencarian estimasi parameter di sini menggunakan Quasi-Newton melalui prosedur Broyden-Fletcher-Goldfarb-Shanno (BFGS), karena prosedur ini paling akurat dalam mengestimasi parameter daripada prosedur lainnya. Model ANN juga dapat sekaligus

menangani pola hubungan yang non linier.

Dari 55 senyawa turunan flavonoid yang digunakan sebagai amatan, dengan 45 variabel bebas, yaitu posisi atom dalam ruang dimensi 3 dan besarnya nilai panjang gelombang sebagai variabel respon, diperoleh model dalam bentuk *mapping* (Gambar 1) dan koefisien (pembobot) dapat dilihat pada Tabel 1 dan Tabel 2.



Gambar 1. Model ANN regresi dengan 4 unit *hidden*

Keterangan :

- Bias 1 = intersep pada unit *hidden*
- Bias 2 = intersep pada unit output
- $X_i, H_j, Y$  = berturut-turut adalah unit input ke- $i$  (variabel bebas ke- $i$ ), unit *hidden* ke  $j$  dan unit output (variabel respon)
- $i = 1, \dots, 45$ ,
- $j = 1, \dots, 4$

Model ANN direpresentasikan dalam bentuk pemetaan dari unit-unit input yang mengandung variabel bebas ke unit *hidden*. Semua informasi yang ada dalam unit input disampaikan ke unit *hidden* melalui pembobot dari masing-masing variabel bebas yang dinotasikan dengan  $w_{ji}$ ,  $i = 1, \dots, 45$ ,  $j = 1, \dots, 4$  (Tabel 1). Keberadaan unit *hidden* ini merupakan ciri khas yang sangat penting dalam ANN. Selanjutnya semua informasi dari

unit *hidden* ini disampaikan ke unit output melalui pembobot  $w_{kj}$ ,  $k = 1, j = 1, \dots, 4$  (Tabel 2). Estimasi bagi respon pada observasi ke- $n$  dapat diperoleh dengan memasukan semua variabel  $x_1, x_2, \dots, x_{45}$ . Selanjutnya dibentuk kombinasi linier sebagai berikut :

$$h_j = \sum_{i=1}^{45} w_{ji} x_i, j = 1, \dots, 4$$

Tabel 1. Koefisien (pembobot) dari unit input ke unit *hidden*

INT	Wj1	Wj2	Wj3	Wj4	Wj5	Wj6	Wj7	Wj8	Wj9	Wj10	Wj11	Wj12	Wj13	Wj14	Wj15
-0.41	-0.15	0.61	-0.61	0.02	-2.32	-0.31	-1.04	1.81	-0.73	-1.53	-1.41	3.47	-2.43	3.12	-1.49
-0.70	-1.81	1.35	0.53	-1.17	-3.12	0.15	-0.37	2.84	-0.67	-3.38	-1.81	4.37	-1.19	3.53	-2.42
-0.37	1.00	-0.25	-0.34	0.36	0.20	-0.55	-0.62	-0.03	-0.02	0.78	0.06	-0.29	-1.31	-0.53	0.52
-0.24	-1.61	2.60	-1.20	-1.15	-1.96	0.42	1.19	1.41	0.48	-1.85	-2.96	4.13	3.11	2.99	-0.90
Wj16	Wj17	Wj18	Wj19	Wj20	Wj21	Wj22	Wj23	Wj24	Wj25	Wj26	Wj27	Wj28	Wj29	Wj30	
2.86	1.90	-0.15	4.16	-2.21	-0.36	-2.68	2.72	2.53	1.05	-3.60	-2.10	0.75	-1.47	0.80	
4.55	0.56	-2.82	3.23	-4.43	-1.08	-1.28	5.53	2.00	1.19	-5.22	-2.73	0.18	-0.52	0.16	
-0.42	0.56	0.43	-0.41	-0.22	0.04	-0.30	-1.77	0.08	0.63	0.50	-0.60	0.24	-1.04	0.65	
2.89	-1.80	-0.15	5.48	-2.92	-2.41	-3.65	5.48	2.12	-0.72	-2.40	-1.83	-0.33	1.19	-0.75	
Wj31	Wj32	Wj33	Wj34	Wj35	Wj36	Wj37	Wj38	Wj39	Wj40	Wj41	Wj42	Wj43	Wj44	Wj45	
-0.61	-1.32	-3.28	-1.85	2.82	-1.26	-1.17	1.43	-3.04	0.20	-1.49	-0.13	1.40	-0.75	-2.49	
-0.73	-2.37	-1.03	-2.12	0.30	0.71	-1.45	2.00	-1.30	0.37	-1.15	0.37	1.60	0.80	-1.04	
0.13	0.55	-1.77	-0.69	1.20	-1.40	-0.74	0.33	-1.39	0.24	-0.56	-1.06	-0.25	-0.94	-1.10	
-0.02	-2.80	2.93	0.57	0.01	2.19	1.24	0.17	1.59	-0.69	0.13	4.35	-0.32	3.00	0.46	

Tabel 2. Koefisien pembobot dari unit *hidden* ke unit output

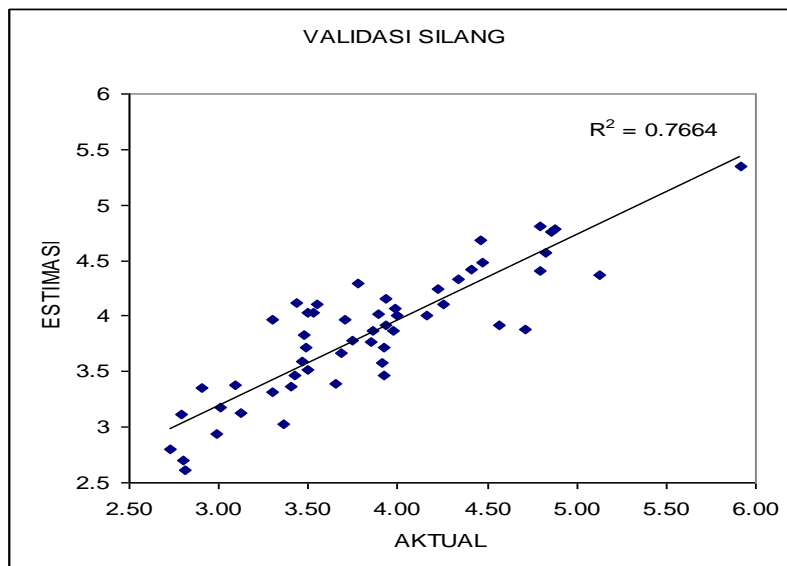
INTERSEP	wk1	Wk2	wk3	wk4
0.789441	-2.63394	-6.06591	2.220865	-3.2449

Berdasarkan koefisien atau pembobot seperti tertera pada Tabel 1 dan 2, selanjutnya untuk mengecek apakah model yang dihasilkan sesuai atau tidak, maka dilakukan validasi silang antara nilai panjang gelombang aktual

dan nilai panjang gelombang hasil estimasi. Nilai hasil ramalan model ditampilkan dalam Tabel 3 dan hasil validasi silang model ANN disajikan dalam gambar 2 di bawah ini.

Tabel 3. Perbandingan nilai aktual dengan nilai ramalan model beserta simpangannya

Pereaksi Geser		Panjang Gelombang (nm)		
		Aktual	Model	Pergeseran
Sumbu X	MeOH	378.6	-	-
	MeOH + NaOH	375.6	378.2	-2.6
	MeOH + NaOH (setelah 5 menit)	376.8	377.1	-0.3
	MeOH + NaOAc	375.6	374.3	1.3
	MeOH + NaOAc + HeBO3	381.1	380.3	0.8
	MeOH + AlCl3	375.6	375.2	0.4
	MeOH + AlCl3 + HCl	375.6	377.1	-1.5
	Sumbu Y	MeOH	275.6	-
Sumbu Y	MeOH + NaOH	276.9	278.3	-1.4
	MeOH + NaOH (setelah 5 menit)	276	277.1	-1.1
	MeOH + NaOAc	276.2	278.4	-2.2
	MeOH + NaOAc + HeBO3	275.1	274.3	0.8
	MeOH + AlCl3	275.5	276.1	-0.6
	MeOH + AlCl3 + HCl	274.1	276.2	-2.1
	Sumbu Z	MeOH	245.1	-
Sumbu Z	MeOH + NaOH	243.3	244.4	-1.1
	MeOH + NaOH (setelah 5 menit)	243.7	245.3	-1.6
	MeOH + NaOAc	244.8	243.7	1.1
	MeOH + NaOAc + HeBO3	245.9	246.4	-0.5
	MeOH + AlCl3	245.1	246.9	-1.8
	MeOH + AlCl3 + HCl	244.7	245.5	-0.8



Gambar 2. Validasi silang model ANN

Berdasarkan plot antara nilai hasil estimasi dan nilai sebenarnya (aktual) terlihat bahwa sebagian besar mendekati pola garis linier. Dari hal tersebut menunjukkan bahwa selisih nilai panjang gelombang yang sebenarnya dan nilai hasil estimasi (sisaan) cukup kecil. Dengan nilai-nilai sisaan yang kecil ini maka akan mengakibatkan jumlah kuadrat sisaan (JKS) yang kecil pula dan tentunya akan menghasilkan nilai koefisien determinasi yang besar.

Berdasarkan Gambar 2, kesesuaian model dengan menggunakan validasi silang antara nilai-nilai aktual dan nilai estimasi sebesar 76,64%, yang ditunjukkan oleh nilai koefisien determinasi ( $R^2 = 0,7664$ ). Semakin besar nilai koefisien determinasi ( $R^2$ ) semakin baik model yang terbentuk. Dengan  $R^2 = 76,64\%$  artinya bahwa dengan analisis *Artificial Neural Networks* (ANN) dapat menjelaskan total variansi dari nilai aktivitas enzim yang sebenarnya sebesar 76,64% berdasarkan estimasinya.

## KESIMPULAN

Terdapat pola hubungan antara besarnya nilai aktivitas enzim dengan posisi atom-atom dalam ruang dimensi tiga pada senyawa turunan Flavonoid.

Ada kecenderungan bahwa pola hubungan antara besarnya nilai aktivitas enzim berdasarkan panjang gelombangnya dengan posisi atom-atom dalam ruang dimensi tiga pada senyawa turunan Flavonoid bersifat non linier.

## SARAN

Model ANN pada kasus ini hanya berlaku untuk senyawa turunan Flavonoid sehingga untuk memperluas cakupan, perlu dilibatkan senyawa lain.

## DAFTAR PUSTAKA

- Asih, I. A. R. Astiti dan I M. Adi Setiawan, 2008, *JURNAL KIMIA* 2 (2), JULI 2008 : 111-116
- Jajang, 2004, Penerapan Analisis Artificial Neural Networks (ANN) dalam Pengelompokan Ekstrak Daun Jati Belanda (*Guazuma ulmifolia* Lamk.), Thesis, Sekolah Pascasarjana IPB, Bogor.
- Janes A. Freeman and David M. Skapura, 1992, Neural Network Algorithms, Application and Programming Techniques, Addison-Wesley Publishing Company, California.
- Kun-Long Ho, Yuan-Yih Hsu and Chen Chuen Yang, 1992, "Short Term Load Forecasting Using A Multi Layer Neural Network With An Adaptive Learning Algorithm", IEEE Tr. On Power Sys. Vol. 7, No. 1.
- K. y. Lee, J. H. Park, 1992, "Short Term Load Forecasting Using An Artificial Neural Network", IEEE Tr. On power Sys., Vol. 7, No. 1.
- Laurence Fausett, 1994, Fundamental of Neural Network Architectures, Algorithms And Application, Prentice Hal, Englewood Cliffs, New Jersey.
- Shin-Tzo Chen, David C. Yu and A. R. Moghaddamlo, 1992, "Weather Sensitive Short Term Load Forecasting Using Non Fully Connected Artificial Neural Network", , IEEE Tr., On Power Sys. Vol. 7, No. 5, Feb.

- O. Mohammed, D. Park, R. Merchant, T. Dihn, C. Tong, A. Azeem, 1995, "Practical Experiences With An Adaptive Neural Network Short Term Load Forecasting System, IEEE Transaction On Power System, Vol. 10, No. 1, pp 254-265.