

## **PENENTUAN *GENERALIZED CROSS VALIDATION* (GCV) SEBAGAI KRITERIA DALAM PEMILIHAN MODEL REGRESI *B-SPLINE* TERBAIK**

**Yuyun Yuniartika, Dadan Kusnandar, Muhlasah Novitasari Mara**

### **INTISARI**

*Regresi B-Spline merupakan salah satu model pendekatan nonparametrik yang fungsinya merupakan polinomial tersegmen atau terbagi pada suatu titik fokus yang disebut knot. Tujuan dari penelitian ini adalah menganalisis Generalized Cross Validation (GCV) yang digunakan sebagai kriteria dalam pemilihan titik knot yang optimal pada regresi B-Spline linier. Aplikasi regresi B-Spline diterapkan pada data pertumbuhan isolat Bacillus laterosporus. Hasil penelitian menunjukkan bahwa model regresi B-Spline linier terbaik dengan titik-titik knot yang optimal adalah tiga titik knot dengan nilainya masing-masing adalah  $k_1 = 6$ ,  $k_2 = 10$  dan  $k_3 = 32$  serta menghasilkan nilai GCV minimum sebesar 0,000000679.*

**Kata Kunci :** *B-spline, Generalized Cross Validation, Nonparametrik*

### **PENDAHULUAN**

Dalam kehidupan sehari-hari, sering kali ditemukan bahwa nilai suatu variabel dipengaruhi atau saling berhubungan dengan nilai variabel lain. Sebagai contoh, analisis hubungan antara pola konsumsi seseorang dalam suatu komoditas tertentu dalam hal ini kita sebut sebagai variabel dependen dengan penghasilan per bulan dalam hal ini disebut sebagai variabel independen. Dalam kasus tersebut, nilai-nilai dari variabel dependen bergantung kepada nilai-nilai dari variabel independennya.

Sasaran utama suatu percobaan ilmiah adalah menemukan hubungan antara variabel yang di analisis. Analisis regresi merupakan salah satu teknik statistik yang paling sering dipergunakan dalam membangun model-model matematis untuk menunjukkan hubungan antara variabel-variabel yang ada pada data percobaan [1]. Salah satu analisis regresi yang biasa digunakan adalah regresi nonparametrik. Regresi nonparametrik merupakan metode pendugaan model yang tidak terikat asumsi bentuk persamaan regresi tertentu, sehingga memberikan fleksibilitas yang lebih tinggi. Terdapat beberapa teknik estimasi dalam regresi nonparametrik antara lain estimator *Kernel*, estimator *Wavelet* dan estimator *B-Spline* [2].

Regresi *B-Spline* merupakan regresi nonparametrik yang menekankan pada suatu regresi ke arah *fitting* data dengan tetap memperhatikan kelulusan kurva. *B-Spline* merupakan model polinomial yang tersegmen atau terbagi pada suatu titik fokus yang disebut knot. Sifat tersegmen atau terbagi inilah yang memberikan fleksibilitas yang lebih baik daripada model polinomial biasa [3].

Ada tiga kriteria yang harus diperhatikan dalam membentuk model regresi *B-Spline* yaitu menentukan orde untuk model, banyaknya knot, dan lokasi penempatan knot. Orde untuk model dapat ditentukan berdasarkan pola umum yang terjadi pada data, sedangkan banyaknya knot dan lokasi knot ditentukan berdasarkan perubahan pola di daerah tertentu pada kurva [4]. Untuk memperoleh regresi *B-Spline* yang optimal maka perlu dipilih lokasi knot yang optimal pula. Ada beberapa kriteria yang dapat digunakan dalam pemilihan knot yang optimal yaitu fungsi resiko prediksi (P), *Cross Validation* (CV), dan *Generalized Cross Validation* (GCV). GCV merupakan salah satu pemilihan titik knot yang paling sering digunakan dalam penentuan model regresi terbaik dan dapat digunakan untuk semua

jenis *B-Spline* yang ada dikarenakan mempunyai perhitungan yang lebih sederhana dan cukup efisien [5].

Penelitian ini akan mengkaji penentuan GCV sebagai kriteria dalam pemilihan model regresi *B-Spline* terbaik. Penerapan regresi *B-Spline* akan digunakan pada data pertumbuhan isolat *Bacillus laterosporus* terhadap selang waktu inkubasi tertentu dengan variabel dependen yaitu kepadatan optik (*optical density*) media pertumbuhan dan variabel independen yaitu selang waktu inkubasi.

### REGRESI NONPARAMETRIK

Regresi nonparametrik merupakan suatu metode untuk mengetahui pola hubungan antara variabel independen dengan variabel dependen yang tidak diketahui bentuk kurva regresinya. Sehingga model regresi nonparametrik dapat berbentuk fungsi apa saja, baik linier ataupun non linier. Model umum regresi nonparametrik adalah sebagai berikut:

$$y_i = f(x_i) + \varepsilon_i \quad \text{dimana } i = 1, 2, 3, \dots, n$$

dimana  $y_i$  adalah variabel dependen;  $x$  merupakan variabel independen;  $f(x_i)$  adalah fungsi regresi yang tidak diketahui bentuknya; sedangkan  $\varepsilon_i$  merupakan galat, faktor pengganggu yang tidak dapat dijelaskan oleh model [6].

Regresi nonparametrik memiliki fleksibilitas yang tinggi dan hanya diasumsikan mempunyai bentuk kurva yang mulus. Regresi nonparametrik sederhana biasa disebut *scatter plot smoothing* karena dalam penggunaannya adalah untuk menemukan kurva mulus melalui plot pencar  $Y$  terhadap  $X$ . Ada beberapa teknik untuk mengestimasi regresi dalam regresi nonparametrik, beberapa diantaranya dengan estimator *Wavelet*, *Kernel*, *B-Spline* dan lain sebagainya [7].

### REGRESI B-SPLINE

Jika diberikan pasangan data  $(x_i, y_i)$  dan hubungan antara variabel dependen dengan variabel independen tidak diketahui bentuknya, maka dapat digunakan regresi nonparametrik dengan model sebagai berikut:

$$y_i = f(x_i) + e_i \quad \text{dimana } i = 1, 2, \dots, n$$

dengan  $y_i$  adalah variabel dependen;  $e_i$  merupakan residual dan  $f(x_i)$  adalah fungsi regresi yang didekati dengan fungsi *B-Spline*. Fungsi *B-Spline*  $f(x_i)$  sebagai berikut:

$$f(x_i) = \sum_{j=1}^{m+k} \beta_j B_{j-m,m}(t)$$

dengan  $B_{j-m,m}$  merupakan basis *B-Spline* ke- $j$  berorde  $m$ .

### Pemilihan Model Regresi *B-Spline* yang Terbaik

Pemilihan titik knot  $k_1, k_2, \dots, k_m$  yang optimal sangat penting dalam regresi *B-Spline*. Titik knot merupakan suatu titik fokus, sehingga kurva yang dibentuk tersegmentasi pada titik tersebut. Oleh karena itu agar diperoleh regresi *B-Spline* yang terbaik perlu dipilih titik knot yang optimal. Jika titik knot yang optimal sudah diperoleh, maka akan memberikan model regresi *B-Spline* yang terbaik.

Salah satu metode pemilihan titik knot yang optimal adalah GCV. Model regresi *B-Spline* yang sesuai berkaitan dengan titik knot yang optimal didapat dari nilai GCV minimum. Fungsi GCV didefinisikan sebagai :

$$GCV(k_1, k_2, \dots, k_h) = \frac{MSE(k_1, k_2, \dots, k_h)}{\left( \frac{1}{n} \text{trace}[I - A(k_1, k_2, \dots, k_h)] \right)^2}$$

dengan

$n$  = jumlah data

$I$  = matriks identitas

$$\text{MSE} = n^{-1} \sum_{j=1}^p \left( y_j - \hat{y}(x_j) \right)^2$$

$k_1, k_2, \dots, k_h$  = titik knot

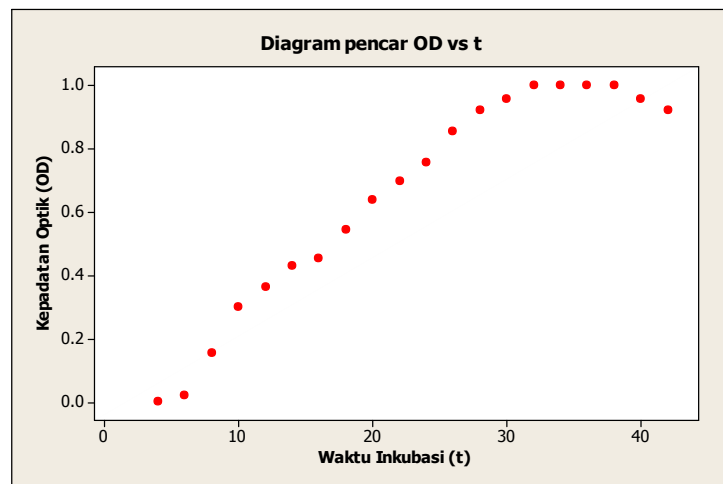
$$A(k_1, k_2, \dots, k_h) = T(T'T)^{-1}T'$$

$$T = \begin{bmatrix} 1 & x_1 & \dots & x_1^{m-1} & (x_1 - k_1)^{m-1} & \dots & (x_1 - k_h)^{m-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & x_n^{m-1} & (x_n - k_1)^{m-1} & \dots & (x_n - k_h)^{m-1} \end{bmatrix}$$

Nilai GCV dipakai karena aspek perhitungannya lebih sederhana dan cukup efisien. Selain itu, kriteria model regresi yang umumnya dipakai masih tetap dijadikan acuan pemilihan model regresi *B-Spline* terbaik. Model regresi *B-Spline* terbaik adalah model yang mampu menjelaskan pola hubungan antara variabel independen dengan variabel dependen.

**Aplikasi**

Untuk eksplorasi pola hubungan yang terjadi antara kepadatan optik media pertumbuhan ( $y$ ) dengan waktu ( $x$ ) maka pada Gambar 1 di bawah akan menampilkan *scatter plot* dari data pertumbuhan Isolat *Bacillus laterosporus*.



Gambar 1. Kurva Pertumbuhan Isolat *Bacillus laterosporus*

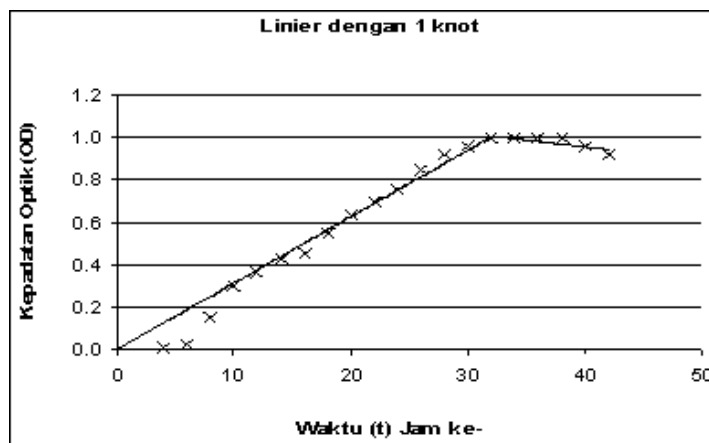
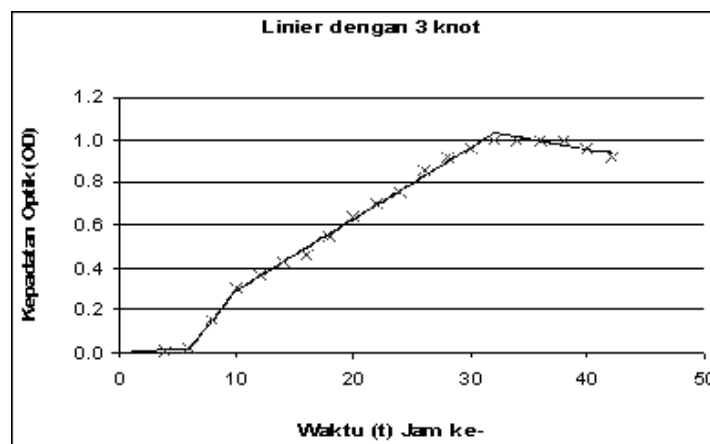
Data pertumbuhan isolat *Bacillus laterosporus* yang telah diperoleh akan dianalisis dengan menggunakan metode regresi *B-Spline*. Analisis data dengan metode regresi *B-Spline* dilakukan untuk memperoleh model regresi *B-Spline* yang tepat dengan GCV minimum. Pemilihan titik knot yang optimal terletak pada nilai GCV yang minimum. Nilai GCV yang minimum dari model regresi *B-Spline* linier dengan beberapa titik knot disajikan pada Tabel 1 berikut ini

Tabel 1. Nilai GCV dari beberapa Titik Knot

No	Knot	GCV( $\lambda$ )
1	28	0,000010264845
2	30	0,000008547129
3	32	0,000007537085
4	6; 28	0,000003166550
5	6; 30	0,000002569903
6	6; 32	0,000002901397
7	6; 8; 32	0,000001141906
8	6; 10; 30	0,000000995845
9	6; 10; 32	0,000000679713
10	6; 12; 32	0,000001031537

Pada Tabel 4.2 diperoleh nilai GCV minimum sebesar 0,000000679713 yang berada pada tiga titik knot yaitu  $k_1 = 6$ ,  $k_2 = 10$  dan  $k_3 = 32$ .

Model regresi *B-Spline* ini disajikan dalam Gambar 2 di bawah ini:

Gambar 2.b Kurva *B-Spline* Linier dengan Satu Titik KnotGambar 2.b Kurva *B-Spline* Linier dengan Tiga Titik Knot

Gambar 2.a menunjukkan adanya perubahan pola kurva  $k_1 = 30$  dan dari gambar 2.b menunjukkan terjadi perubahan pola kurva sebelum dan sesudah titik knot pada  $k_1 = 6$ ,  $k_2 = 10$  dan  $k_3 = 32$ . Pola

kurva sebelum  $k_1 = 6$  cenderung mendatar dibanding setelah  $k_1 = 6$  yang mempunyai kecenderungan naik. Begitupun pada pola kurva setelah  $k_2 = 10$  yang cenderung naik. Namun nilai setelah  $k_3 = 32$  mempunyai garis yang cenderung menurun

Dengan menggunakan tiga buah knot, yakni yang dipilih adalah knot 6, 10 dan 32 kurva yang dibentuk dapat menangkap banyak titik amatan. Semakin banyak knot yang digunakan pada regresi *B-Spline* linier, maka kurva juga semakin menangkap lebih banyak titik pengamatan.

Dengan memperhatikan hasil yang telah diperoleh, dapat disimpulkan bahwa titik knot yang paling optimal dengan nilai GCV minimum adalah penggunaan tiga titik knot pada regresi *B-Spline* linier. Nilai GCV beberapa model regresi *B-Spline* dengan tiga beberapa titik knot ditunjukkan pada Tabel 2.

Tabel 2. Nilai GCV Minimum Regresi *B-Spline* dengan beberapa Titik Knot

Model	Jumlah Knot	Letak Titik Knot			Nilai GCV Minimum
		1	2	3	
Linier	1	32	-	-	0,000007537085
	2	6	30	-	0,000002569903
	3	6	10	32	0,000000679713

Berdasarkan Tabel diatas dapat disimpulkan bahwa model terbaik untuk untuk memprediksi nilai kepadatan optik pertumbuhan Isolat *Bacillus laterosporus* pada selang waktu adalah dengan tiga titik knot  $k_1 = 6$ ,  $k_2 = 10$  dan  $k_3 = 32$  yaitu dengan nilai GCV minimum 0,000000679713.

Model regresi *B-Spline* terbaik untuk data pertumbuhan Isolat *Bacillus laterosporus* dapat diartikan bahwa, untuk memprediksi nilai kepadatan optik pertumbuhan Isolat *Bacillus laterosporus* terhadap selang waktu, digunakan regresi *B-Spline* linier dengan tiga titik knot.

**PENUTUP**

Berdasarkan uraian dari penulisan ini dapat diambil kesimpulan sebagai berikut:

1. GCV (*Generalized Cross Validation*) merupakan salah satu kriteria yang dapat digunakan dalam pemilihan knot yang optimal dalam penentuan model regresi *B-Spline* yang terbaik.
2. Semakin banyak knot pada model regresi *B-Spline* linier maka nilai GCV akan semakin kecil.
3. Penentuan lokasi knot yang berbeda, akan menghasilkan model regresi *B-Spline* dengan nilai GCV yang berbeda pula.
4. Model regresi *B-Spline* yang terbaik dalam model regresi *B-Spline* linier adalah terletak pada titik knot  $k_1 = 6$ ,  $k_2 = 10$  dan  $k_3 = 32$  dengan nilai GCV minimum 0,000000679713.

**DAFTAR PUSTAKA**

[1] Green, P.J. dan Silverman, B.W. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. St Edmundsbury. Chapman & Hall. 1994.  
 [2] Hardle, W. *Applied Nonparametric Regression*. Cambridge University Press. 1990.  
 [3] Lyche, T. and Morken, K. *Spline Methods*. Department of Informatics, University of Oslo. 2004.  
 [4] Budiantara, I.N., Suryadi, F., Otok, B.W. dan Guritno, S. Pemodelan B-Spline dan MARS Pada Nilai Ujian Masuk terhadap IPK Mahasiswa Jurusan Disain Komunikasi Visual UK. Petra Surabaya, *Jurnal Teknik Industri*, **8**:1-13. 2006.  
 [5] Eubank, R. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker. New York. 1988.

- [6] John Fox. *Robust Regression Appendix to An R and S-Plus Companion to Applied Regression*. University of Oxford. 2002.
- [7] Suparti. *Perbandingan Estimator Regresi Nonparametrik Menggunakan Metode Fourier dan Metode Wavelet*. *Jurnal Matematika*, **8**:88-94. 2005.

YUYUN YUNIARTIKA : FMIPA UNTAN Pontianak, zahra\_alfarish@yahoo.com  
DADAN KUSNANDAR : FMIPA UNTAN Pontianak, dkusnand@yahoo.com  
MUHLASAH NOVITASARI MARA : FMIPA UNTAN Pontianak, noveemara@gmail.com

---