

PEMODELAN *B-SPLINE* DAN *MARS* PADA NILAI UJIAN MASUK TERHADAP IPK MAHASISWA JURUSAN DISAIN KOMUNIKASI VISUAL UK. PETRA SURABAYA

I Nyoman Budiantara¹⁾, Fredi Suryadi²⁾, Bambang Widjanarko Otok³⁾, Suryo Guritno⁴⁾

¹⁾Lecturer at Department of Statistics, ITS, Surabaya

²⁾ Staf. at Petra Christian University, Surabaya
e-mail: fsuryadi@petra.ac.id

³⁾ Ph.D Student at Department of Mathematics, UGM; Lecturer at Department of Statistics, ITS, Surabaya
e-mail: otok_bw@yahoo.com

⁴⁾ Lecturer at Department of Mathematics UGM, Yogyakarta
e-mail: suryoguritno@ugm.ac.id

ABSTRAK

Analisa regresi digunakan untuk melihat pengaruh variabel independen terhadap variabel dependent dengan terlebih dulu melihat pola hubungan variabel tersebut. Hal ini dapat dilakukan dengan melalui dua pendekatan. Pendekatan yang paling umum dan seringkali digunakan adalah pendekatan parametrik. Pendekatan parametrik mengasumsikan bentuk model sudah ditentukan. Apabila tidak ada informasi apapun tentang bentuk dari fungsi regresi, maka pendekatan yang digunakan adalah pendekatan nonparametrik. (Haerdle, 1990). Karena pendekatan tidak tergantung pada asumsi bentuk kurva tertentu, sehingga memberikan fleksibilitas yang lebih besar. Tujuan penelitian ini adalah mendapatkan model terbaik mengenai nilai ujian masuk terhadap nilai IPK (Indek Prestasi Kumulatif) mahasiswa jurusan Disain Komunikasi Visual tahun 1999 di Universitas Kristen Petra Surabaya dengan analisis regresi, baik parametrik maupun nonparametrik. Pendekatan regresi parametrik menggunakan regresi linear sederhana, kuadratik dan kubik, sedangkan regresi nonparametrik digunakan *B-Spline* dan *Multivariate Adaptive Regression Splines* (MARS). Secara keseluruhan, model terbaik dipilih berdasarkan koefisien determinasi terbesar. Namun demikian untuk MARS, model terbaik dipilih berdasarkan pada GCV, minimum MSA dan koefisien determinasi terbesar.

Kata kunci: regresi nonparametrik, *B-Spline*, *MARS*, koefisien determinasi.

ABSTRACT

*Regression analysis is constructed for capturing the influences of independent variables to dependent ones. It can be done by looking at the relationship between those variables. This task of approximating the mean function can be done essentially in two ways. The quiet often use parametric approach is to assume that the mean curve has some prespecified functional forms. Alternatively, nonparametric approach, .i.e., without reference to a specific form, is used when there is no information of the regression function form (Haerdle, 1990). Therefore nonparametric approach has more flexibilities than the parametric one. The aim of this research is to find the best fit model that captures relationship between admission test score to the GPA. This particular data was taken from the Department of Design Communication and Visual, Petra Christian University, Surabaya for year 1999. Those two approaches were used here. In the parametric approach, we use simple linear, quadric cubic regression, and in the nonparametric ones, we use *B-Spline* and *Multivariate Adaptive Regression Splines* (MARS). Overall, the best model was chosen based on the maximum determinant coefficient. However, for MARS, the best model was chosen based on the GCV, minimum MSE, maximum determinant coefficient.*

Keywords: nonparametric regression, *B-Spline*, *MARS*, determinant coefficients.

1. PENDAHULUAN

Analisis regresi memperlihatkan hubungan dan pengaruh variabel prediktor terhadap variabel respon. Misalnya y adalah variabel respon dan t adalah variabel prediktor, untuk n buah pengamatan, secara umum hubungan antara y dan t dapat ditulis sebagai berikut:

$$y_i = g(t_i) + \varepsilon_i; \quad i = 1, 2, \dots, n \quad (1)$$

dengan ε adalah sesatan random dan $g(t_i)$ merupakan kurva regresi.

Jika kurva regresi merupakan model parametrik maka disebut sebagai regresi parametrik dan apabila model yang diasumsikan ini benar, maka pendugaan parametrik sangat efisien, tetapi jika tidak, menyebabkan interpretasi data yang menyesatkan (Haerdle, 1990). Pendekatan parametrik mengasumsikan bentuk model sudah ditentukan. Apabila tidak ada informasi apapun tentang bentuk $g(t_i)$, maka pendekatan yang digunakan adalah pendekatan nonparametrik. Karena pendekatan tidak tergantung pada asumsi bentuk kurva tertentu, sehingga memberikan fleksibilitas yang lebih besar. Dalam hal ini diasumsikan $g(t_i)$ termuat dalam ruang fungsi (Eubank, 1988).

Ada beberapa teknik estimasi dalam regresi nonparametrik antara lain pendekatan histogram, estimator *Spline*, estimator kernel, estimator deret orthogonal, analisis wavelet dan lain-lain. Pendekatan estimator *Spline* ada bermacam-macam antara lain *Spline* original, *Spline* type M, *Spline* relaxed, *Spline* terbobot dan lain-lain. Pendekatan *Spline* mempunyai suatu basis fungsi. Basis fungsi yang biasa dipakai antara lain *Spline truncated* dan *B-Spline*. (Lyche dan Morken, 2004).

Wahba (1990) menunjukkan bahwa *Spline* memiliki sifat-sifat statistik yang berguna untuk menganalisis hubungan dalam regresi. *Spline* dalam regresi nonparametrik terus berkembang sampai pada model adaptive (Billier dan Fahrmeir, 2000) dan multivariate respon (Holmes dan Mallick, 2003). Untuk mengestimasi basis fungsi *Spline* telah dikembangkan beberapa metode seperti *monotonicity* (He dan Shi, 1998) dan *penalised* (Hall dan Opsomer, 2005).

Spline adalah salah satu jenis *piecewise* polinomial, yaitu polinomial yang memiliki sifat tersegmen. Sifat tersegmen ini memberikan fleksibilitas lebih dari polinomial biasa, sehingga memungkinkan untuk menyesuaikan diri secara lebih efektif terhadap karakteristik lokal suatu fungsi atau data.

Spline mempunyai kelemahan pada saat orde *Spline* tinggi, knots yang banyak dan knot yang terlalu dekat akan membentuk matrik dalam perhitungan yang hampir singular, sehingga persamaan normal tidak dapat diselesaikan. Basis lain yang dapat mengatasi kelemahan ini adalah basis *B-Spline*. Namun kesulitan dengan *B-Spline* karena basis ini hanya dapat didefinisikan secara rekursif dan karenanya tidak dapat dievaluasi secara langsung (Eubank, 1988) dan (Schuemaker, 1981). Salah satu penerapan *B-Spline* dilakukan dalam bidang *Fluid Dynamics* (Botella dan Shariff, 2003).

Dalam tulisan ini dilakukan studi perbandingan antara *Spline truncated* dan *B-Spline*. Sedangkan studi perbandingan *Spline* yang pernah dilakukan adalah studi perbandingan Reinsch dan Speckman *Spline* dalam regresi nonparametrik (Carter dan Silverman, 1992).

Pada penelitian ini dibahas mengenai estimasi *B-Spline* pada model regresi nonparametrik, dan melakukan simulasi untuk membandingkan MSE *B-Spline* dan *Spline truncated*. Selain itu juga penerapan *B-Spline* pada data riil.

2. B-SPLINE

Model regresi $y_i = g(t_i) + \varepsilon_i$; $i = 1, 2, \dots, n$, dengan ε_i merupakan residual dan $g(t_i)$ kurva regresi. Apabila digunakan pendekatan kurva *Spline truncated* dikatakan regresi nonparametrik, maka kurva regresi g dapat ditulis menjadi:

$$g(t) = \sum_{i=1}^m \alpha_i t^{i-1} + \sum_{j=1}^K \beta_j (t - u_j)_+^{m-1} \quad (\text{Eubank, 1988}) \tag{2}$$

dimana $u_j, j = 1, 2, \dots, K$ dengan $u_1 < u_2 < \dots < u_K$ adalah knot dan $m \in \mathbf{N}_0$ (integer non negatif). Nilai m menunjukkan derajat *Spline truncated*. Jika kurva regresi g didekati dengan fungsi *B-Spline* maka g dapat ditulis menjadi:

$$g(t) = \sum_{j=1}^{m+K} \gamma_j B_{j-m,m}(t) \tag{3}$$

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

Cara membangun fungsi *B-Spline* orde m dengan titik titik knot $a < u_1 < \dots < u_K < b$ adalah dengan terlebih dahulu mendefinisikan knot tambahan sebanyak $2m$, yaitu $u_{-(m-1)}, \dots, u_{-1}, u_0, \dots, u_{K+m}$ dimana $u_{-(m-1)} = \dots = u_0 = a$ dan $u_{k+1} = \dots = u_{K+m} = b$. Biasanya a diambil dari nilai minimum t dan b diambil dari nilai maksimum t . Fungsi *B-Spline* didefinisikan secara rekursif sebagai berikut:

$$B_{i,m}(t) = \frac{t - u_i}{u_{i+m-1} - u_i} B_{i,m-1}(t) + \frac{u_{i+m} - t}{u_{i+m} - u_{i+1}} B_{i+1,m-1}(t)$$

dimana

$$B_{j,1}(t) = \begin{cases} 1, & \text{jika } u_j < t \leq u_{j+1} \\ 0 & \text{jika } t < u_j \text{ atau } t \geq u_{j+1} \end{cases} \quad (\text{Botella dan Shariff, 2003}) \tag{4}$$

m adalah derajat dari *B-Spline*. Untuk $m=2$ memberikan fungsi *B-Spline* linear, $m=3$ memberikan fungsi *B-Spline* kuadratik dan $m=4$ memberikan fungsi *B-Spline* kubik. Untuk mengestimasi koefisien γ pada persamaan (3), didefinisikan matrik

$$B(\lambda) = (B_{j,m}(t_i))_{\substack{i=1, \dots, n \\ j=-(m-1), \dots, K}}$$

atau dapat ditulis sebagai berikut

$$B(\lambda) = \begin{pmatrix} B_{-(m-1),m}(t_1), B_{-(m-2),m}(t_1), \dots, B_{K,m}(t_1) \\ \vdots & \vdots & \vdots \\ B_{-(m-1),m}(t_n), B_{-(m-2),m}(t_n), \dots, B_{K,m}(t_n) \end{pmatrix}$$

Jadi $B(\lambda)$ adalah sebuah matrik berukuran $n \times (m+K)$

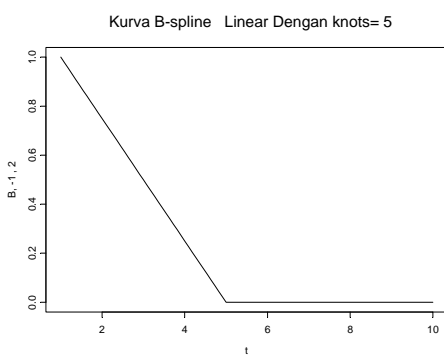
Sebagai gambaran untuk menjelaskan fungsi *B-Spline*, misalnya *B-Spline* linear ($m=2$), dengan satu titik knot, pada $t=5$, dengan nilai t minimum 1 dan nilai t maksimum 10. Langkahnya adalah menentukan knot tambahan sebanyak $2m$ yaitu diambil dari nilai minimum 1 dan maksimum 10, sehingga knots menjadi $u_{-1} = u_0 = 1, u_1 = 5, u_2 = u_3 = 10$, maka matrik yang akan dibentuk adalah $B(\lambda) = (B_{-1,2}(t_i), B_{0,2}(t_i), B_{1,2}(t_i))$, $i=1, 2, \dots, n$ yaitu sebuah matrik dengan ukuran $n \times 3$. Dari persamaan (4) $B_{-1,2}(t_i)$ dapat ditulis:

$$B_{-1,2}(t) = \frac{t-1}{1-1}B_{-1,1}(t) + \frac{5-t}{5-1}B_{0,1}(t)$$

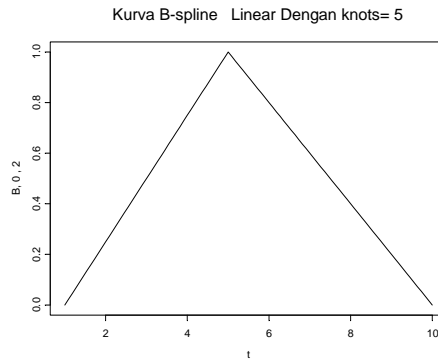
$B_{-1,1}(t)$ didefinisikan bernilai 0 karena $u_{-1} = u_0$ (Eubank,1988), sedangkan $B_{0,1}(t)$ akan bernilai 1 pada t yang bernilai $u_0 = 0$ sampai dengan $u_1 = 1$, dan bernilai 0 untuk yang lain, sehingga dapat ditulis sebagai berikut:

$$B_{-1,2}(t) = \begin{cases} \frac{5-t}{4}, & 1 < t \leq 5 \\ 0, & 5 < t \leq 10 \end{cases}$$

Bentuk kurva $B_{-1,2}(t)$ dan $B_{0,2}(t)$ ditunjukkan dalam gambar berikut:



Gambar 1. Kurva $B_{-1,2}(t)$



Gambar 2. Kurva $B_{0,2}(t)$,

Untuk basis $B_{0,2}(t)$ dengan menggunakan Persamaan (4) dapat ditulis:

$$B_{0,2}(t) = \frac{t-1}{5-1}B_{0,1}(t) + \frac{10-t}{10-5}B_{1,1}(t)$$

Selanjutnya dapat ditulis sebagai berikut:

$$B_{0,2}(t) = \begin{cases} \frac{t-1}{4}, & 1 < t \leq 5 \\ \frac{10-t}{5}, & 5 < t \leq 10 \end{cases}$$

Bentuk kurva $B_{0,2}(t)$ ditunjukkan pada Gambar 2.

Untuk basis $B_{1,2}(t_i)$ dengan menggunakan persamaan (4) dapat ditulis:

$$B_{1,2}(t) = \frac{t-5}{10-5}B_{1,1}(t) + \frac{10-t}{10-10}B_{2,1}(t) \quad B_{1,2}(t) = \frac{t-5}{5}B_{1,1}(t)$$

Selanjutnya dapat ditulis sebagai berikut:

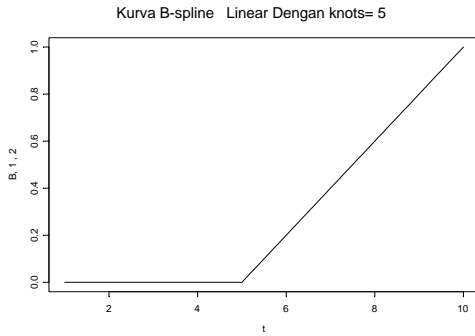
$$B_{1,2}(t) = \begin{cases} 0, & 1 < t \leq 5 \\ \frac{t-5}{5}, & 5 < t \leq 10 \end{cases}$$

Bentuk kurva $B_{1,2}(t)$ ditunjukkan pada Gambar 3.

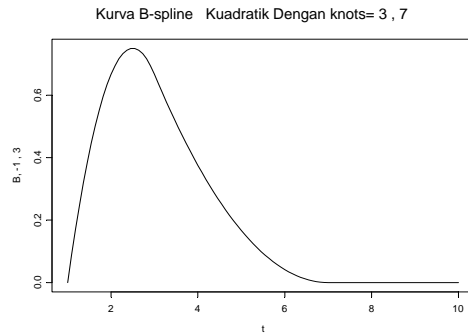
Untuk kurva *B-Spline* kuadratik dengan 2 titik knot misalnya pada $t=3$ dan $t=7$ dapat dicari dengan cara yang serupa, dengan hasil sebagai berikut:

$$B_{-1,3}(t) = \begin{cases} \frac{(t-1)(13-3t)}{12}, & 1 < t \leq 3 \\ \left(\frac{7-t}{24}\right)^2, & 3 < t \leq 7 \\ 0, & 7 < t \leq 10 \end{cases}$$

Bentuk kurva $B_{-1,3}(t)$ ditunjukkan pada Gambar 4.



Gambar 3. Kurva $B_{1,2}(t)$



Gambar 4. Kurva $B_{-1,3}(t)$

Dengan cara yang serupa, dapat dibuat gambar kurva *B-Spline* dengan berbagai m dan beberapa titik knots.

3. MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS)

Spline adalah salah satu jenis potongan polinomial, yaitu polinomial yang memiliki sifat tersegmen. Sifat tersegmen ini memberikan fleksibilitas lebih dari polinomial biasa, sehingga memungkinkan untuk menyesuaikan diri secara lebih efektif terhadap karakteristik lokal dari suatu fungsi atau data. Secara umum, fungsi *Spline* berorde k adalah sembarang fungsi yang dinyatakan sebagai:

$$S(t) = \sum_{i=0}^{k-1} \alpha_i t^i + \sum_{j=1}^h \delta_j (t - u_j)_+^{k-1} \text{ dengan, } (t - u_j)_+^{k-1} = \begin{cases} (t - u_j)^{k-1}, & t \geq u_j \\ 0, & t < u_j \end{cases}$$

dimana:

- α dan δ : konstanta riil
- u_1, \dots, u_h : titik-titik knot

Fungsi *Spline* tersebut diatas menunjukkan:

- a) fungsi S merupakan potongan polinomial berorde k pada subinterval $[u_j, u_{j+1}]$
- b) fungsi S memiliki turunan kontinu tingkat $k-2$
- c) $S^{(k-1)}$ merupakan fungsi tangga dengan titik-titik lompatan u_1, \dots, u_h
- d) fungsi S adalah suatu polinomial dengan orde m di luar $[u_1, u_n]$

Recursive Partitioning Regression (RPR) merupakan pendekatan dari fungsi $f(t)$ yang tidak diketahui dengan:

$$\hat{f}(t) = \sum_{j=1}^S c_j(t)B_j(t) \tag{5}$$

dimana, $B_j(t) = I[t \in R_j]$, $I[\cdot]$ menunjukkan fungsi indikator yang mempunyai nilai 1 (satu) jika pernyataan benar ($t \in R_j$) dan 0 (nol) jika salah, $c_j(t)$ merupakan koefisien (konstanta) yang ditentukan dalam *subregion*.

Penentuan knots pada regresi *dummy* dilakukan secara manual, karena memiliki dimensi data yang rendah dan hal ini tidak akan mengalami kesulitan, sedangkan untuk data yang berdimensi tinggi terdapat kesulitan. Untuk mengatasi hal tersebut digunakan model *Recursive Partition Regression* karena penentuan knots tergantung (otomatis) dari data. Namun demikian model ini masih terdapat kelemahan yaitu model yang dihasilkan tidak kontinu pada knots, dan untuk mengatasinya digunakan model MARS.

Model MARS selain penentuan knots yang dilakukan secara otomatis dari data, juga menghasilkan model yang kontinu pada knots. Pemilihan knots pada MARS menggunakan algoritma *forward stepwise* dan *backward stepwise* yang salah satunya didasarkan nilai *Generalized Cross Validation* (GCV) minimum. Model MARS dapat ditulis sebagai berikut:

$$\hat{f}(t) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km} \cdot (t_{v(k,m)} - u_{km})] \tag{6}$$

di mana:

- a_0 = basis fungsi induk
- a_m = koefisien dari basis fungsi ke- m
- M = maksimum basis fungsi (*nonconstant basis fungsi*)
- K_m = derajat interaksi
- s_{km} = nilainya ± 1
- $t_{v(k,m)}$ = variabel independen
- u_{km} = nilai knots dari variabel independen $t_{v(k,m)}$

Penjabaran dari Persamaan (6) dapat disajikan sebagai berikut:

$$\begin{aligned} \hat{f}(t) = & a_0 + \sum_{m=1}^M a_m [s_{1m} \cdot (t_{v(1,m)} - u_{1m})] \\ & + \sum_{m=1}^M a_m [s_{1m} \cdot (t_{v(1,m)} - u_{1m})][s_{2m} \cdot (t_{v(2,m)} - u_{2m})] \\ & + \sum_{m=1}^M a_m [s_{1m} \cdot (t_{v(1,m)} - u_{1m})][s_{2m} \cdot (t_{v(2,m)} - u_{2m})][s_{3m} \cdot (t_{v(3,m)} - u_{3m})] \\ & + \dots \end{aligned} \tag{7}$$

dan secara umum Persamaan (6) dapat dituliskan sebagai berikut:

$$\hat{f}(t) = a_0 + \sum_{K_m=1} f_i(t_i) + \sum_{K_m=2} f_{ij}(t_i, t_j) + \sum_{K_m=3} f_{ijk}(t_i, t_j, t_k) + \dots \tag{8}$$

Persamaan (8), menunjukkan bahwa penjumlahan pertama meliputi semua basis fungsi untuk satu variabel, penjumlahan kedua meliputi semua basis fungsi untuk interaksi antara dua variabel, penjumlahan ketiga meliputi semua basis fungsi untuk interaksi antara tiga variabel dan seterusnya.

Misalkan $\Lambda(\mathbb{W}) = \{h(\mathbb{K}^{\mathbb{W}})\}_{\mathbb{K}^{\mathbb{W}}}$ adalah himpunan dari variabel yang dihubungkan dengan basis fungsi B_m ke- m , maka setiap penjumlahan pertama pada Persamaan (8) dapat dinyatakan sebagai:

$$f_i(t_i) = \sum_{\substack{K_m=1 \\ i \in V(m)}} a_m B_m(t_i) \quad (9)$$

$f_i(t_i)$ merupakan penjumlahan semua basis fungsi untuk satu variabel x_i dan merupakan *Spline* dengan derajat $q=1$ yang merepresentasikan fungsi univariat. Setiap fungsi bivariat pada Persamaan (8) dapat ditulis sebagai:

$$f_{ij}(t_i, t_j) = \sum_{\substack{K_m=2 \\ (i,j) \in V(m)}} a_m B_m(t_i, t_j) \quad (10)$$

yang merepresentasikan penjumlahan semua basis fungsi dua variabel t_i dan t_j . Penambahan ini untuk menghubungkan kontribusi univariat, yang dituliskan sebagai berikut:

$$f_{ij}^*(t_i, t_j) = f_i(t_i) + f_j(t_j) + f_{ij}(t_i, t_j) \quad (11)$$

Untuk fungsi trivariat pada penjumlahan yang ketiga diperoleh dengan menjumlahkan semua basis fungsi untuk tiga variabel, yang dituliskan sebagai berikut:

$$f_{ijk}(t_i, t_j, t_k) = \sum_{\substack{K_m=3 \\ (i,j,k) \in V(m)}} a_m B_m(t_i, t_j, t_k) \quad (12)$$

Penambahan fungsi univariate dan bivariate mempunyai kontribusi dalam bentuk:

$$f_{ijk}^*(t_i, t_j, t_k) = f_i(t_i) + f_j(t_j) + f_k(t_k) + f_{ij}(t_i, t_j) + f_{ik}(t_i, t_k) + f_{jk}(t_j, t_k) + f_{ijk}(t_i, t_j, t_k) \quad (13)$$

Persamaan (8) merupakan dekomposisi dari analisis varians untuk table kontingensi, yang dikenal dengan dekomposisi ANOVA dari model MARS. Interpretasi model MARS melalui dekomposisi ANOVA adalah merepresentasikan variabel yang masuk dalam model, baik untuk satu variabel maupun interaksi antara variabel, selanjutnya merepresentasikan secara grafik. Penambahan aditif Persamaan (9) dapat ditunjukkan dengan membuat plot antara $f_i(t_i)$ dengan t_i sebagai salah satu model aditif. Kontribusi interaksi antara dua variabel dapat divisualisasikan dengan membuat plot antara $f_{ij}^*(t_i, t_j)$ dengan t_i dan t_j menggunakan kontur plot. Model dengan interaksi yang lebih tinggi dalam visualisasi dapat dibuat dengan menggunakan plot dalam beberapa variabel fixed dengan variabel komplemen.

Pada model MARS, pemilihan model menggunakan metode *stepwise* yang terdiri dari *forward* dan *backward*. *Forward stepwise* dilakukan untuk mendapatkan jumlah basis fungsi maksimum dengan kriteria pemilihan basis fungsi adalah meminimumkan *Average Sum Of Square Residual (ASR)*. Untuk memenuhi konsep parsimoni (model sederhana) dilakukan *backward stepwise* yaitu memilih basis fungsi yang dihasilkan dari *forward stepwise* dengan meminimumkan nilai *Generalized Cross-Validation (GCV)*. (Friedman and Silverman, 1989, Friedman, 1990, 1991).

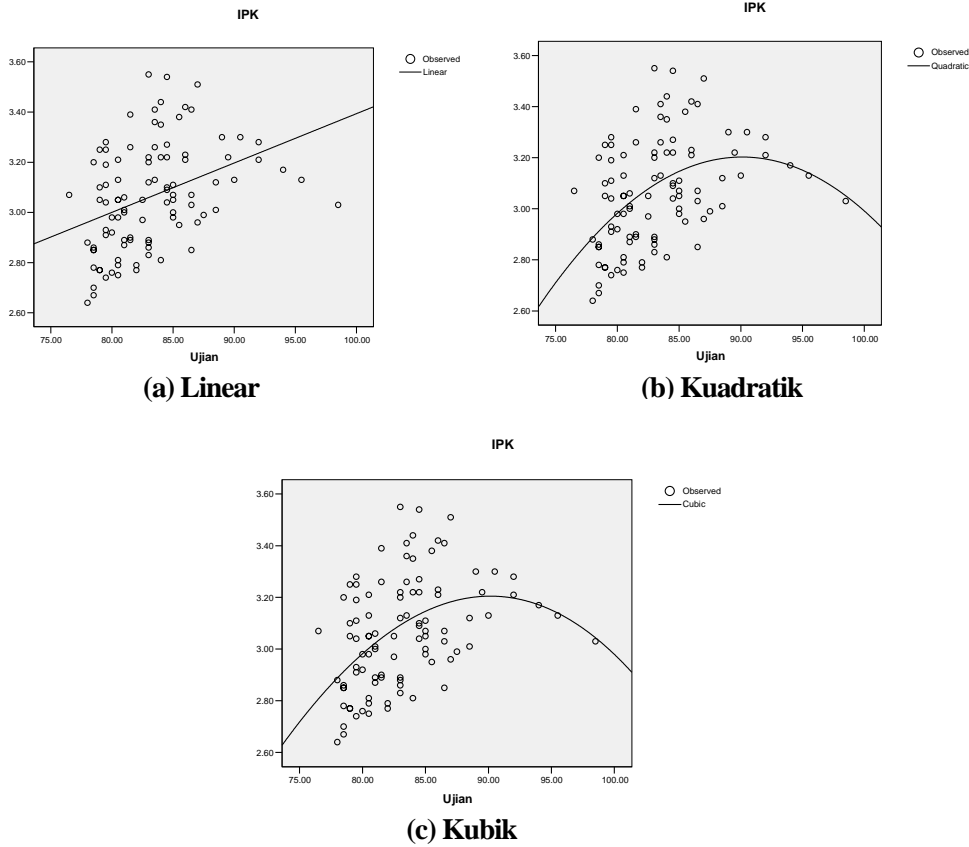
4. ANALISIS DATA DAN PEMBAHASAN

Pola Hubungan Nilai Ujian Masuk terhadap IPK dengan Regresi Linear

Salah satu variabel yang dipakai untuk menentukan seorang calon mahasiswa lulus atau tidak lulus dalam ujian masuk di Universitas Kristen Petra adalah nilai ujian tulis. Dasar itu dipakai karena diharapkan ada pola hubungan antara nilai ujian masuk tulis (t) dengan variabel Indeks

Prestasi Kumulatif kelulusan (y). Data yang dianalisis sebagai kasus adalah data nilai ujian masuk tulis dan IPK, 97 mahasiswa jurusan Disain Komunikasi Visual angkatan 1999.

Pola hubungan nilai ujian masuk tulis(t) dengan variabel Indeks Prestasi Kumulatif kelulusan (y) dengan analisis pada regresi linear, kuadratik dan kubik adalah sebagai berikut:



Gambar 5. Plot Regresi linear, kuadratik dan kubik

Sedangkan pengaruh nilai ujian masuk tulis (t) dengan variabel Indeks Prestasi Kumulatif kelulusan (y) masing-masing model, secara rinci disajikan pada Tabel berikut:

Tabel 1. Pengaruh Nilai Ujian Masuk Terhadap IPK dengan Regresi Linear

Model Regresi	Parameter	Koefisien Regresi	F-Hitung	Sig.	R-Square
Linear	β_0	1,427	16,466	0,000	0,148
	β_1	0,020			
Kuadratik	β_0	-14,31	12,936	0,000	0,216
	β_1	0,389			
	β_2	-0,002			
Kubik	β_0	-8,957	12,956	0,000	0,216
	β_1	0,202			
	β_2	0,000001			
	β_3	-0,000008			

Dari Tabel 1, menunjukkan bahwa nilai ujian masuk tulis (t) berpengaruh terhadap IPK (y) baik pada model regresi linear, kuadratik dan kubik. Hal ini dapat dilihat dari nilai Sig. Yang lebih kecil dari nilai alfa ($0,000 < 0,05$). Sedangkan nilai koefisien determinasi (R^2) pada model regresi kuadratik dan kubik memberikan nilai yang sama yaitu sebesar 0,216 dan pada model regresi linear sebesar 0,148. Hal ini menunjukkan bahwa model terbaik pada pemodelan regresi linear adalah model regresi kuadratik dan kubik. Namun demikian untuk kemudahan interpretasi dan kesederhanaan model dan juga didukung Gambar 5.b) dan 5.c) maka pola hubungan nilai ujian masuk tulis (t) dengan IPK (y), yang terbaik menggunakan model regresi kuadratik. Persamaan masing-masing model regresi linear adalah sebagai berikut:

Model Regresi Linear:

$$\hat{y} = 1,427 + 0,02t$$

di mana: t = nilai ujian

Model Regresi Kuadratik:

$$\hat{y} = -14,31 + 0,389t - 0,002t^2$$

di mana: t = nilai ujian

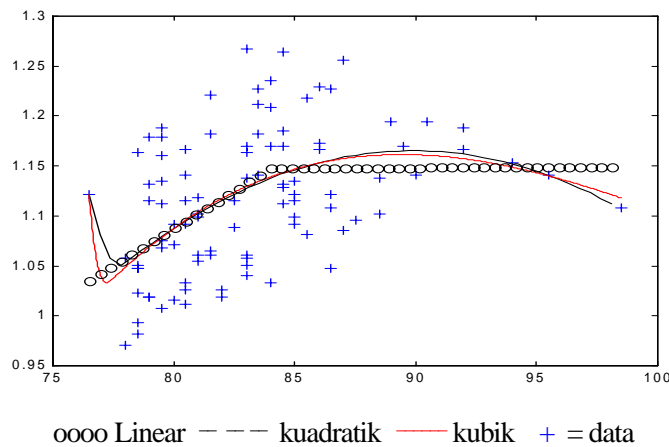
Model Regresi Kubik:

$$\hat{y} = -8,957 + 0,202t + 0,000001t^2 - 0,0000083t^3$$

di mana: t = nilai ujian

Pola Hubungan Nilai Ujian Masuk terhadap IPK dengan Pendekatan *B-Spline*

Pola hubungan nilai ujian masuk tulis (t) antara variabel Indeks Prestasi Kumulatif kelulusan (y) dengan analisis *B-Spline* yaitu *B-Spline* linear, kuadratik dan kubik adalah sebagai berikut:



Gambar 6. Plot *B-Spline* linear, kuadratik dan kubik

Sedangkan pengaruh nilai ujian masuk tulis (t) dengan variabel Indeks Prestasi Kumulatif kelulusan (y) masing-masing model, secara rinci disajikan pada tabel berikut:

Tabel 2. Pengaruh Nilai Ujian Masuk Terhadap IPK dengan *B-Spline*

Model <i>B-Spline</i>	Parameter	Koefisien Regresi	F-Hitung	Sig.	MSE	<i>R-Square</i>
Linear	γ_1	1,03532	10914,7	0,000	0,0037	0,997
	γ_2	1,14796				
	γ_3	1,14943				
Kuadratik	γ_1	1,12168	8230,5	0,000	0,0037	0,997
	γ_2	1,03838				
	γ_3	1,24666				
	γ_4	1,10664				
Kubik	γ_1	1,12168	6529,8	0,000	0,0037	0,997
	γ_2	1,02176				
	γ_3	1,20037				
	γ_4	1,17548				
	γ_5	1,11808				

Berdasarkan Tabel 2 di atas, ketiga model memberikan nilai MSE untuk model *B-Spline* linear, kuadratik maupun kubik masing masing mempunyai nilai yang mendekati sama yaitu 0,0037; 0,0037; 0,00371, begitu juga nilai koefisien determinasi (R^2) sebesar 0,997. Untuk kemudahan interpretasi dan kesederhanaan model dan juga didukung Gambar 6 maka pola hubungan nilai ujian masuk tulis (t) dengan IPK (y), yang terbaik menggunakan model *B-Spline* linear. Persamaan masing-masing model regresi linear adalah sebagai berikut:

Model *B-Spline* Linear:

$$\hat{y}^* = 1,03532B_0 + 1,14796B_1 + 1,14943B_2$$

dengan $B_0 = B_{-1,2}(t)$, $B_1 = B_{0,2}(t)$ dan $B_2 = B_{1,2}(t)$

Model *B-Spline* Kuadratik:

$$\hat{y}^* = 1,12168B_0 + 1,03838B_1 + 1,24666B_2 + 1,10664B_3$$

dengan $B_0 = B_{-2,3}(t)$, $B_1 = B_{-1,3}(t)$, $B_2 = B_{0,3}(t)$ dan $B_3 = B_{1,3}(t)$

Model *B-Spline* Kubik adalah:

$$\hat{y}^* = 1,12168B_0 + 1,02176B_1 + 1,20037B_2 + 1,17548B_3 + 1,11808B_4$$

dengan $B_0 = B_{-3,4}(t)$, $B_1 = B_{-2,4}(t)$, $B_2 = B_{-1,4}(t)$, $B_3 = B_{0,4}(t)$, $B_4 = B_{1,4}(t)$

Titik knot model *B-Spline* linear terletak pada nilai ujian masuk (t) sebesar 84, hal ini menunjukkan terjadi perubahan pola sebelum dan sesudah nilai ujian masuk (t) sebesar 84.

Pola Hubungan Nilai Ujian Masuk terhadap IPK dengan Pendekatan MARS

Pola hubungan antara nilai ujian masuk tulis (t) dengan variabel Indeks Prestasi Kumulatif kelulusan (y) pendekatan MARS, dengan variasi pada basis fungsi, interaksi antar variabel independen dan minimum observasi subregion secara lengkap tersaji pada Tabel 3. berikut.

Tabel 3. Pengaruh Nilai Ujian Masuk Terhadap IPK dengan MARS

Model MARS	BF	MO	F-Hitung	GCV	R ²	MSE
Model 1	10	10	14,874	0,041	0,996	0.0342872
Model 2	10	20	26,878	0,036	0,996	0.0348211
Model 3	15	10	14,874	0,041	0,996	0.0342872
Model 4	15	20	26,878	0,036	0,996	0.0348211
Model 5	20	10	26,878	0,036	0,996	0.0348211
Model 6	20	20	10,257	0,035	0,997	0.0342796
Model 7	30	10	26,878	0,036	0,996	0.0348211

Keterangan:

BF= basis fungsi, MI= maksimum interaksi, MO = minimum observasi pada setiap subregion

Berdasarkan Tabel 3 di atas, ternyata dengan kriteria GCV, R² dan MSE, maka model memberikan nilai MSE dan GCV yang kecil adalah Model 6, begitu juga pada Model 6 memberikan nilai R² yang paling besar yaitu 0,997. Sehingga model yang terbaik untuk pola hubungan nilai ujian masuk tulis (t) dengan IPK (y) pada pendekatan MARS adalah Model 6. Adapun persamaan Model 6 adalah sebagai berikut.

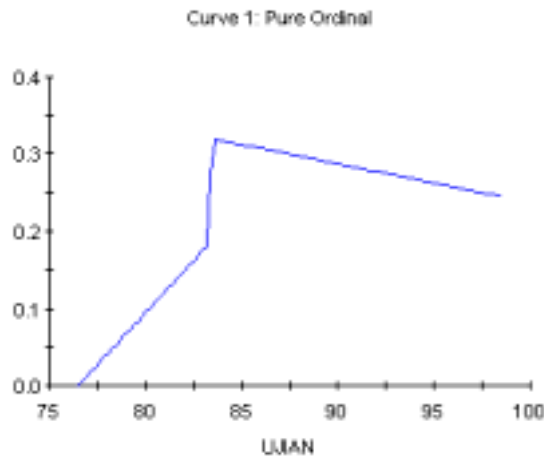
$$\hat{y} = 3,061 - 0,402BF1 - 0,027BF2 + 0,397BF3$$

dengan, BF1 = max(0, t - 83,5)

BF2 = max(0, 83,5 - t)

BF3 = max(0, t - 83,17)

Secara visual dapat dilihat pada gambar berikut:

**Gambar 7. Plot MARS, Piecewise Cubic dengan 3 Basis Functions**

Model 6 dan Gambar 7 menunjukkan bahwa terjadi perubahan pola pada $t=83,17$ dan $t=83,5$. Pola dari $t=78,5$ sampai nilai $t=83,17$ mempunyai kecenderungan naik secara tajam, dan begitu juga nilai dari $t=83,17$ sampai $t=83,5$ mempunyai kecenderungan naik secara tajam menuju nilai $t=83,5$, sedangkan nilai di atas $t=83,5$ mempunyai kecenderungan turun sampai IPK tertentu.

5. KESIMPULAN

Berdasarkan analisis data tentang pola hubungan nilai IPK sebagai variabel dependen (Y) dan nilai ujian masuk tulis sebagai variabel independen (t), dapat disimpulkan:

- a. Hasil analisis regresi linear, kuadratik dan kubik menghasilkan nilai R^2 yang cenderung sama dengan persamaan regresi kuadratik adalah sebagai berikut:

$$\hat{y} = -14,31 + 0,389t - 0,002t^2$$

dengan, t = nilai ujian

- b. Hasil analisis model *B-Spline* linear, kuadratik maupun kubik mempunyai nilai MSE yang mendekati sama. Model yang diperoleh untuk masing-masing orde adalah sebagai berikut : Model *B-Spline* Linear adalah:

$$\hat{y}^* = 1,03532B_0 + 1,14796B_1 + 1,14943B_2$$

dengan

$$B_0 = B_{1,2}(t) = \begin{cases} \frac{84-t}{7,5}, & 76,5 < t \leq 84 \\ 0, & 84 < t \leq 98,5 \end{cases} \quad B_1 = B_{0,2}(t) = \begin{cases} \frac{t-76,5}{7,5}, & 76,5 < t \leq 84 \\ \frac{98,5-t}{14,5}, & 84 < t \leq 98,5 \end{cases} \quad B_2 = B_{1,2}(t) = \begin{cases} 0, & \text{untuk yang lain} \\ \frac{t-84}{14,5}, & 84 < t \leq 98,5 \end{cases}$$

Pada Model *B-Spline* linear, terjadi perubahan pola pada $t=84$. Pola sebelum nilai 84 mempunyai kemiringan yang lebih tajam dibanding setelah nilai 84. Nilai setelah 84 mempunyai garis yang hampir mendatar, artinya nilai setelah 84, IPK kelulusannya cenderung sama

- c. Hasil analisis model MARS menghasilkan persamaan berikut:

$$\hat{y} = 3,061 - 0,402BF_1 - 0,027BF_2 + 0,397BF_3$$

dengan, $BF_1 = \max(0, t - 83,5)$

$$BF_2 = \max(0, 83,5 - t)$$

$$BF_3 = \max(0, t - 83,17)$$

terjadi perubahan pola pada $t=83,17$ dan $t = 83,5$. Pola dari $t=78,5$ sampai nilai $t=83,17$ mempunyai kecenderungan naik secara tajam, dan begitu juga nilai dari $t=83,17$ sampai $t=83,5$ mempunyai kecenderungan naik secara tajam menuju nilai $t = 83,5$, sedangkan nilai di atas $t = 83,5$ mempunyai kecenderungan turun sampai IPK tertentu.

DAFTAR PUSTAKA

1. Billier, C., and Fahrmeir, L., 2000, "Bayesian varying-coefficient models using adaptive regression Spline, Statistical Modeling", <http://citeseer.ist.psu.edu/biller00bayesian.html>.
2. Botella, O. and Shariff, K., 2003, "B-Spline Methods in Fluid Dynamics". *International Journal of Computational Fluid Dynamics*, 17 (2), 133-149.
3. Carter, C.K., and Silverman, B.W., 1992, "A comparison of the Reinsch and Speckman Splines". *Biometrika*, 79(1), pp 81-91.
4. Eubank, R.L., 1988. *Spline Smoothing and Nonparametric Regression*, Marcel Dekker: New York.
5. Friedman, J.H., 1990, *Estimating functions of mixed ordinal and categorical variables using Multivariate Adaptive Regression Splines*. Technical Report LCS 107, Statistics Department, Stanford University.

6. Friedman, J.H. and Silverman, B.W., 1989. "Flexible parsimony smoothing and additive modeling". *Technometrics*, **31**, 3 – 39.
7. Friedman, J.H., 1991. *Multivariate Adaptive Regression Splines (With Discussion)*. Stanford California 94309.
8. Hall, P., and Opsomer, J.D., 2005. "Theory for penalized *Spline* regression", *Biometrika*, 92,1. pg.105
9. Haerdle, W., 1990, *Applied Nonparametric Regression*, Cambridge University Press: New York.
10. He, X. and Shi P., 1998. "Monotone *B-Spline* Smoothing", *Journal of the American Statistical Association*; Jun 1998; 93, 442.
11. Holmes, C.C., and Mallick B.K., 2003, "Generalized Nonlinear Modelling With Multivariate Free-Knot Regression *Spline*", *Journal of the American Statistical Association*; 98,462.
12. Lyche, T., and Morken, K., 2004. *Spline* Methods Draft, (www.ub.uio.n/umn/english/index.html), down load tanggal 23 Feb 2005.
13. Schuemaker, L.L., 1981, *Spline Functions : Basic Theory*, John Wiley & Sons, Inc: Canada.
14. Wahba, G., 1990, "Spline Models For Observational Data". *SIAM, CBMS-NSF Regional Conference Series in Applied mathematics*, Philadelphia.