

PEMODELAN REGRESI *SPLINE* (Studi Kasus: Herpindo Jaya Cabang Ngaliyan)

I Made Budiantara Putra^{§1}, I Gusti Ayu Made Srinadi², I Wayan Sumarjaya³

¹Jurusan Matematika, Fakultas MIPA - Universitas Udayana [Email: imadebudiantaraputra@yahoo.co.id]

²Jurusan Matematika, Fakultas MIPA - Universitas Udayana [Email: srinadiigustiayumade@yahoo.co.id]

³Jurusan Matematika, Fakultas MIPA - Universitas Udayana [Email: sumarjaya@gmail.com]

[§]Corresponding Author

ABSTRACT

Regression analysis is a method of data analysis to describe the relationship between response variables and predictor variables. There are two approaches to estimating the regression function. They are parametric and nonparametric approaches. The parametric approach is used when the relationship between the predictor variables and the response variables are known or the shape of the regression curve is known. Meanwhile, the nonparametric approach is used when the form of the relationship between the response and predictor variables is unknown or no information about the form of the regression function. The aim of this study are to determine the best spline nonparametric regression model on data of quality of the product, price, and advertising on purchasing decisions of Yamaha motorcycle with optimal knots point and to compare it with the multiple regression linear based on the coefficient of determination (R^2) and mean square error (MSE). Optimal knot points are defined by two point knots. The result of this analysis is that for this data multiple regression linear is better than the spline regression one.

Keywords: regression, nonparametric spline, point knot, purchase decision motorcycles Yamaha.

1. PENDAHULUAN

Analisis regresi merupakan metode analisis data yang menggambarkan hubungan antara variabel respons dan variabel prediktor (Hosmer dan Lemeshow [5]). Terdapat dua pendekatan untuk mengestimasi fungsi regresi yaitu pendekatan parametrik dan nonparametrik. Pendekatan parametrik yaitu pendekatan yang digunakan apabila bentuk hubungan antara variabel prediktor dan variabel respons diketahui atau bentuk dari kurva regresinya diasumsikan mengikuti pola tertentu. Pendekatan nonparametrik yaitu pendekatan yang digunakan apabila bentuk hubungan antara variabel respons dan prediktornya tidak diketahui atau tidak adanya informasi mengenai bentuk fungsi regresinya. Regresi parametrik memiliki asumsi yang ketat dan kaku seperti sisaan berdistribusi normal dan memiliki varians yang konstan. Selain itu, diketahuinya

karakteristik data dari penelitian sebelumnya sangatlah penting agar diperoleh model yang baik. Dalam model regresi parametrik, estimasi kurva regresi ekuivalen dengan estimasi terhadap parameter-parameter dalam model (Budiantara [2]). Untuk menghindari penggunaan asumsi-asumsi yang ketat maka salah satu alternatif yang dapat digunakan adalah dengan pendekatan regresi nonparametrik. Regresi nonparametrik adalah suatu metode pemodelan yang tidak terikat akan asumsi-asumsi dari persamaan regresi tertentu yang memberikan fleksibilitas yang tinggi dalam menduga sebuah model. Beberapa metode estimasi regresi nonparametrik yang dapat digunakan adalah deret Fourier, *spline*, kernel (Eubank [3]).

Regresi *spline* adalah suatu metode analisis regresi yang bersifat *piecewise polynomial*

yaitu suatu potongan-potongan polinom yang memiliki sifat tersegmen pada selang k yang terbentuk pada titik-titik *knot* (Wang & Yang [7]). Titik *knot* merupakan titik perpaduan bersama yang terjadi karena terdapat perubahan perilaku pola pada interval yang berlainan. *Spline* mempunyai keunggulan dalam mengatasi pola data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik *knot*, serta kurva yang dihasilkan relatif mulus (Härdle [4]). Estimator *spline* cenderung mencari sendiri estimasinya kemana pun data tersebut bergerak sehingga memperoleh model yang sesuai dengan bentuk data. Kriteria yang dapat digunakan dalam pemilihan *knot* yang optimal yaitu *Generalized Cross Validation (GCV)* (Budiantara [1]).

Data penelitian ini diambil dari penelitian yang dilakukan oleh Nugraha [6] untuk mengetahui faktor-faktor yang memengaruhi keputusan pembelian sepeda motor Yamaha pada Herpindo Jaya cabang Ngaliyan dengan menggunakan analisis regresi berganda dengan variabel respons yaitu keputusan pembelian (y) dan variabel prediktor yaitu kualitas produk (x_1), harga (x_2), dan iklan (x_3). Hasil analisisnya menyatakan bahwa ketiga variabel tersebut merupakan kriteria yang penting bagi pembeli sepeda motor Yamaha di Herpindo Jaya cabang Ngaliyan. Namun hasil dari *adjusted R²* diperoleh sebesar 0,555 artinya keputusan pembelian terhadap variabel kualitas produk, harga, dan iklan hanya dapat dijelaskan sebesar 55,5% dari model regresi berganda. Meninjau hasil penelitian yang dilakukan Nugraha [6], maka penulis ingin melanjutkan penelitian tersebut dengan menggunakan analisis regresi *spline*. Tujuan dari penelitian ini adalah menentukan model regresi *spline* terbaik pada data kualitas produk, harga, dan iklan terhadap keputusan pembelian sepeda motor Yamaha pada Herpindo Jaya cabang Ngaliyan lalu membandingkannya dengan analisis regresi linier berganda dengan kriteria koefisien determinasi (R^2) yang terbesar dan *mean square error (MSE)* yang terkecil.

2. TINJAUAN PUSTAKA

Analisis Regresi Parametrik

Analisis regresi merupakan sebuah alat statistika yang digunakan untuk melihat hubungan antara variabel respons dengan satu atau lebih variabel prediktor. Misalkan terdapat data berpasangan (x_i, y_i) untuk n pengamatan, maka hubungan antara variabel x_i dan variabel y_i dapat dinyatakan sebagai berikut:

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

dengan y_i adalah respons ke- i , $f(x_i)$ adalah fungsi regresi atau kurva regresi, serta ε_i adalah sisaan yang diasumsikan *independent* dengan nilai tengah nol dan variansi σ^2 .

Analisis Regresi Nonparametrik

Regresi nonparametrik merupakan suatu metode statistika yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dengan respons ketika tidak diperoleh informasi sebelumnya tentang bentuk fungsi regresinya atau tidak diketahui bentuk kurva regresinya. Fungsi dari model regresi nonparametrik dapat berbentuk apa saja, baik linear atau nonlinear. Misalkan variabel respons adalah y dan variabel prediktor adalah x untuk n pengamatan, model umum dari regresi nonparametrik adalah $y_i = f(x_i) + \varepsilon_i \quad ; \quad i = 1, 2, 3, \dots, n$ (2)

dengan y_i adalah variabel respons, x_i adalah variabel prediktor, $f(x_i)$ adalah fungsi regresi yang tidak diketahui bentuknya, dan ε_i adalah sisaan yang diasumsikan bebas dengan nilai tengah nol dan variansi σ^2 .

Analisis Regresi *Spline*

Spline merupakan model polinom yang tersegmen atau terpotong-potong yang mulus dan dapat menghasilkan fungsi regresi yang sesuai dengan data. Mengestimasi *spline* tergantung pada titik *knot*. Titik *knot* merupakan suatu titik perpaduan yang terjadi karena perubahan pola perilaku dari suatu fungsi pada selang yang berbeda. Fungsi *spline* pada suatu fungsi f dengan orde p dapat dinyatakan sebagai berikut:

$$f(x_i) = \sum_{j=0}^p \beta_j x_i^j + \sum_{l=1}^r \beta_{(p+l)} (x_i - k_l)_+^p \quad (3)$$

dengan k menyatakan banyaknya titik *knot* dan $(x_i - k_l)_+^p$ menyatakan fungsi potongan

(truncated) yang dapat djabarkan sebagai berikut:

$$(x_i - k_l)_+^p = \begin{cases} (x_i - k_l)^p, & x_i \geq k_l \\ 0, & x_i < k_l \end{cases} \quad (4)$$

Bentuk matematis dari fungsi *spline* pada persamaan (3), dapat dinyatakan bahwa *spline* adalah potongan-potongan polinom yang berbeda digabungkan bersama titik *knot* $k_1, k_2, k_3, \dots, k_r$ untuk menjamin sifat kontinuitasnya. Model regresi *spline* dapat pula disajikan dalam bentuk matriks yang dituliskan sebagai berikut:

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} 1 & x_1^1 & x_1^2 & \dots & x_1^p & (x_1 - k_1)_+^p & \dots & (x_1 - k_r)_+^p \\ 1 & x_2^1 & x_2^2 & \dots & x_2^p & (x_2 - k_1)_+^p & \dots & (x_2 - k_r)_+^p \\ 1 & x_3^1 & x_3^2 & \dots & x_3^p & (x_3 - k_1)_+^p & \dots & (x_3 - k_r)_+^p \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 1 & x_n^1 & x_n^2 & \dots & x_n^p & (x_n - k_1)_+^p & \dots & (x_n - k_r)_+^p \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \\ \beta_{(p+1)} \\ \vdots \\ \beta_{(p+r)} \end{bmatrix}$$

atau

$$f(x) = x\beta \quad (5)$$

Estimasi regresi nonparametrik *spline* diperoleh dengan menggunakan metode *Maximum Likelihood Estimator* (MLE). Jika *error* pada persamaan (2) diasumsikan berdistribusi normal, maka y_i juga berdistribusi normal dengan nilai tengah $f(x_i)$ dan varians σ^2 . Sehingga fungsi densitas peluang y_i menjadi

$$f(y; f(x), \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-f(x))^2}{2\sigma^2}\right], f(x) > 0, \sigma^2 > 0 \quad (6)$$

Fungsi *likelihood* dapat dituliskan sebagai berikut:

$$L(y, f) = \prod_{i=1}^n f(y_i; f(x_i), \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - f(x_i))^2\right] \quad (7)$$

Estimasi titik fungsi f diperoleh dengan memaksimumkan fungsi *likelihood* $L(y, f)$ yang dapat diuraikan sebagai berikut:

$$\max_f \{L(y, f)\} = \max_{\beta \in R^{p+r}} \left\{ (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\sum_{j=1}^p \beta_j x_i^j + \sum_{l=1}^r \beta_{p+l} (x_i - k_l)_+^p))\right)^2 \right\} \quad (8)$$

Apabila optimasi ini diselesaikan akan diperoleh estimator untuk β adalah:

$$\hat{\beta} = (x'x)^{-1}x'y. \quad (9)$$

Estimasi dari \hat{y} dapat dituliskan sebagai berikut:

$$\hat{y} = x\hat{\beta} \\ \hat{y} = x(x'x)^{-1}x'y = A(k)y \quad (10)$$

dengan $A(k)$ merupakan matriks yang digunakan untuk perhitungan pada rumus GCV dalam pemilihan titik *knot* optimal.

Pemilihan Titik Knot Optimal

Pemilihan estimator regresi *spline* terbaik diantara model-model yang didapatkan dilihat berdasarkan kriteria *Generalized Cross Validation* (GCV) yang minimum. Rumus untuk menghitung GCV adalah sebagai berikut:

$$GCV(k) = \frac{MSE(k)}{(n^{-1}tr[I-A(k)])^2} \quad (11)$$

dengan $MSE(k) = n^{-1} \sum_{i=1}^n (y_i - \hat{y})^2$, n adalah jumlah data, I adalah matriks identitas, k adalah titik *knot* $(k_1, k_2, k_3, \dots, k_n)$, dan $A(k) = x(x'x)^{-1}x'$.

3. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari penelitian Nugraha [6], penelitian tersebut membahas analisis pengaruh kualitas produk, harga, dan iklan terhadap keputusan pembelian sepeda motor Yamaha pada Harpindo Jaya cabang Ngalian. Variabel-variabel tersebut diukur menggunakan skala Likert. Penelitian yang dilakukan oleh Nugraha [6], untuk mendapatkan nilai pada setiap variabel dilakukan dengan menjumlahkan skor dari setiap pertanyaan yang menyusun masing-masing variabel. Pada penelitian ini, variabel-variabel yang digunakan adalah skor faktor dari indikator/pertanyaan penyusunnya.

Langkah-langkah yang dilakukan dalam penelitian ini adalah sebagai berikut: (a) memodelkan variabel respon dan prediktor dengan regresi nonparametrik *spline* dengan satu *knot*, dua *knot*, tiga *knot*, empat *knot*, dan lima *knot*, dan memilih titik *knot* optimal

berdasarkan kriteria GCV yang paling minimum, (b) melakukan perbandingan antara regresi *spline* dan regresi linier berganda berdasarkan kriteria R^2 dan MSE.

4. HASIL DAN PEMBAHASAN

Penentuan Model Regresi *Spline* yang Terbaik

Pemilihan model regresi *spline* yang terbaik dipilih berdasarkan lokasi dan banyaknya titik *knot*. Dalam memilih titik *knot* yang optimal sangatlah penting dalam menentukan model terbaik. Lokasi titik *knot* yang berbeda akan menghasilkan model yang berbeda. Titik *knot* yang optimal dilihat berdasarkan kriteria GCV yang paling minimum. Berikut ini merupakan tabel yang menunjukkan titik-titik *knot* yang optimal, GCV yang minimum, dan orde yang optimal pada setiap variabel.

Tabel 1. Nilai GCV Minimum untuk Masing-Masing Titik *Knot*

<i>knot</i>	Orde			GCV
	x_1	x_2	x_3	
1	2	2	2	0,771588
2	2	2	2	0,73093
3	2	2	2	0,73769
4	3	3	3	0,76741
5	3	3	3	0,80233

Berdasarkan Tabel 1, dapat dilihat bahwa nilai GCV minimum diperoleh untuk model dengan dua titik *knot*. Titik *knot* pada x_1 adalah -2,24862 dan 1,7902, pada x_2 adalah -2,087 dan 1,20372, yang terakhir x_3 adalah -2,21701 dan 2,01696. Estimasi model regresi *spline* dengan dua titik *knot* dapat dilihat pada Tabel 2.

Sehingga model terbaik untuk pengaruh kualitas produk, harga, dan iklan terhadap keputusan pembelian sepeda motor Yamaha menggunakan dua titik *knot*. Berikut merupakan persamaan model regresi nonparametrik *spline* dengan dua titik *knot*.

$$\hat{y} = 0,49562x_1 + 0,15367x_1^2 - 0,02524(x_1 - (-2,24862))^2_+ + 0,0319(x_1 - 1,7902)^2_+ + 0,34023x_2 - 0,11719x_2^2 + 0,07506(x_2 - (-2,087))^2_+ + 2,328(x_2 - 1,20372)^2_+ - 0,24392x_3 + 0,03027x_3^2 - 0,05981(x_3 - (-2,21701))^2_+ + 1,77059(x_3 - 2,01696)^2_+$$

Tabel 2. Estimasi Model Regresi *Spline* dengan dua Titik *Knot*

Parameter	Estimasi
β_1	0.49562
β_2	0.15367
β_3	-0.02524
β_4	0.0319
β_5	0.34023
β_6	-0.11719
β_7	0.07506
β_8	2.328
β_9	-0.24392
β_{10}	0.03027
β_{11}	-0.05981
β_{12}	1.77059

Perbandingan Antara Analisis Regresi Linier Berganda dengan Regresi *Spline*

Perbandingan antara regresi linier berganda dengan regresi *spline* dapat dilihat dari hasil nilai koefisien determinasi (R^2) yang lebih besar nilainya dan *Mean Square Error* (MSE) yang paling minimum. Berikut ini merupakan tabel perbandingan nilai koefisien determinasi (R^2) dan *Mean Square Error* (MSE) terhadap regresi berganda dan regresi *spline*.

Tabel 3. Perbandingan Analisis Regresi Berganda dengan Regresi *Spline*

	R^2	MSE
Regresi Berganda	53,1%	0,469
Regresi <i>Spline</i>	36,06%	0,645

Berdasarkan Tabel 3 koefisien determinasi regresi *spline* nilainya lebih kecil dari regresi linier berganda dan juga nilai MSE dari regresi *spline* lebih besar dari regresi linier berganda. Pada data ini, model regresi berganda lebih baik dari pada regresi *spline* karena koefisien data regresi *spline* lebih kecil dari pada regresi linier berganda. Artinya keputusan pembelian

sepeda motor Yamaha dapat dijelaskan sebesar 53,1% oleh variabel kualitas produk, harga dan iklan dengan regresi linier berganda sedangkan 36,06% dengan regresi *spline*. Nilai MSE dari regresi linier berganda pada data ini lebih baik dari regresi *spline* karena nilai MSE regresi linier berganda lebih kecil dari regresi *spline*. Artinya tingkat kesalahan pada model regresi linier berganda lebih kecil dari regresi *spline* atau pengukuran yang dilakukan regresi linier berganda lebih teliti dari pada regresi *spline*.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil pembahasan, maka diperoleh kesimpulan bahwa model regresi nonparametrik *spline* yang terbaik adalah model regresi *spline* dengan dua titik *knot*. Nilai GCV yang dihasilkan adalah 0,73093. Model regresi nonparametrik *spline* adalah sebagai berikut.

$$\begin{aligned} \hat{y} = & 0,49562x_1 + 0,15367x_1^2 \\ & - 0,02524(x_1 - (-2,24862))_+^2 \\ & + 0,0319(x_1 - 1,7902)_+^2 \\ & + 0,34023x_2 - 0,11719x_2^2 \\ & + 0,07506(x_2 - (-2,087))_+^2 \\ & + 2,328(x_2 - 1,20372)_+^2 \\ & - 0,24392x_3 + 0,03027x_3^2 \\ & - 0,05981(x_3 - (-2,21701))_+^2 \\ & + 1,77059(x_3 - 2,01696)_+^2. \end{aligned}$$

Berdasarkan perbandingan nilai koefisien determinasi (R^2) dan *Mean Square Error* (MSE) antara analisis regresi *spline* dan regresi linier berganda pada data pengaruh kualitas produk, harga dan iklan terhadap keputusan pembelian sepeda motor Yamaha menunjukkan bahwa regresi linier berganda lebih baik dari regresi *spline* karena koefisien determinasi regresi linier berganda lebih besar dari regresi *spline* dan nilai MSE dari regresi berganda lebih kecil dari regresi *spline*.

Untuk penelitian selanjutnya saran yang dapat penulis sampaikan adalah sebuah model pada regresi *spline* multivariabel, orde pada setiap variabel tidak harus sama melainkan bisa berbeda-beda untuk setiap variabel. Dan apabila ingin menggunakan data penelitian ini disarankan untuk menambahkan

indikator lain yang belum dimasukkan dalam model.

DAFTAR PUSTAKA

- [1] Budiantara, I. N. 2000. Metode U, GML, CV, dan GCV dalam Regresi Nonparametrik Spline. *Majalah Ilmiah Himpunan Matematika Indonesia (MIHMI)*, Vol. 6, 285-290.
- [2] _____, 2009. *Spline dalam Regresi Nonparametrik dan Semiparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang. Pidato Pengukuhan Untuk Jabatan Guru Besar Dalam Bidang Ilmu Matematika Statistika dan Probabilitas, Pada Jurusan Statistika, Fakultas MIPA. Institut Teknologi Sepuluh Nopember, ITS Press: Surabaya.*
- [3] Eubank, R. 1988. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker. New York.
- [4] Härdle, W. 1990. *Applied Nonparametric Regression*. Cambridge University Press, New York.
- [5] Hosmer, D.W. and S. Lemeshow. 2000. *Applied Logistic Regression*, second edition. John Wiley and Sons, Inc. New York.
- [6] Nugraha. M. N. S. 2010. Analisis Pengaruh Kualitas Produk, Harga, dan Iklan Terhadap Keputusan Pembelian Sepeda Motor Yamaha pada Harpindo Jaya Cabang Ngaliyan. *Skripsi*. Jurusan Manajemen Fakultas Ekonomi Universitas Diponegoro, Semarang.
- [7] Wang, J. and Yang, L. 2009. Polynomial Spline Confidence Bands for Regression Curves. *Statistica Sinica*. 19: 325-342.