

## Regresi Smoothing Spline pada Pemodelan Pengaruh Curah Hujan terhadap Harga Cabai di Kota Bandung

*Elin Paulina<sup>1\*</sup>, Siti Sarah Sobariah<sup>2</sup>*

*<sup>1</sup>Program Studi Akuntansi Universitas Nusa Putra*

*<sup>2</sup>Program Studi Akuntansi Universitas Nusa Putra*

*\*Elin.paulina@nusaputra.ac.id*

**Abstrak:** Kenaikan harga cabai sangat tergantung pada musim panen dan musim tanam serta pengaruh iklim. Pada awal musim hujan, Oktober 2018 misalnya, harga cabai tercatat menjadi penyumbang inflasi. Dari tingkat inflasi 0,28% saat itu, andil cabai mencapai 0,09%. Sebaran pola hubungan curah hujan dengan harga cabai, tidak mengikuti pola tertentu sehingga tidak dapat diidentifikasi. Pendekatan regresi nonparametrik dengan smoothing splines memberikan solusi dalam pemodelan, dengan  $\lambda$  optimum senilai 0.0005 dengan nilai RMSE sebesar 9733.39 dan diperoleh knot sebanyak 48 maka kurva dengan pemulusan  $\lambda$  optimum dapat dikatakan dapat menggambarkan pola pergerakan data dengan baik.

**Kata kunci:** *Smoothing Splines, Lambda Optimum, GCV*

**Abstract:** *The increase in chili prices is highly dependent on the harvest and planting seasons as well as climate influences. At the beginning of the rainy season, October 2018, for example, chili prices were recorded as a contributor to inflation. From the inflation rate of 0.28% at that time, and chili reached 0.09%. The distribution pattern of the relationship between rainfall and chili prices does not follow a certain pattern so that it cannot be identified. The nonparametric regression approach with smoothing splines provides a solution in modeling, with an optimum of 0.0005 with an RMSE value of 9733.39 and 48 vertices obtained, the curve with optimum smoothing can be said to describe the pattern of data movement well.*

**Keyword:** *Smoothing Splines, Lambda Optimum, GCV*

### PENDAHULUAN

Hujan memang diperlukan dalam proses produksi bahan pangan di sentra-sentra Pertanian. Namun, jika curah hujan terlalu tinggi, panen bisa terganggu, produk buah dan sayur pun mudah rusak. Dilansir dari katadata.com pola ini sudah tampak selama bertahun-tahun. Pada awal musim hujan, Oktober 2018 misalnya, harga cabai tercatat menjadi penyumbang inflasi. Dari tingkat inflasi 0,28% saat itu, andil cabai mencapai 0,09%.

Kenaikan harga cabai sangat tergantung pada musim panen dan musim tanam serta pengaruh iklim. Disamping itu kenaikan harga juga berkaitan dengan kegiatan pemasaran. Bila dibandingkan dengan harga di daerah konsumen harga cabai di daerah produsen lebih rendah. Beberapa faktor yang mempengaruhi diantaranya faktor angkutan, rendahnya daya tahan cabai, dan daya beli masyarakat yang rendah (Santika dalam Fajri Rahmatul 2017).

Rendahnya daya tahan cabai dapat disebabkan karena musim yang ekstrem, ketahanan cabai terhadap air yang kurang, dapat mempercepat pembusukan cabai tersebut. Sehingga ketersediaan barang berkurang, sebagaimana hukum penawaran dan permintaan. Jika permintaan meningkat dan penawaran tetap tidak berubah, maka itu mengarah pada harga keseimbangan yang lebih tinggi dan kuantitas yang lebih tinggi. Pendekatan dengan menggunakan pemodelan nonparametrik dilakukan karena, pemodelan menggunakan regresi linier dalam metode parametrik tidak dapat digunakan karena residual model tidak berdistribusi normal.

## METODOLOGI

### Regresi Nonparametrik

Regresi nonparametrik merupakan salah satu pendekatan yang digunakan untuk mengetahui pola hubungan antara variabel penjelas dan respon yang tidak diketahui kurva regresinya atau tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola data (Eubank, 1999). Secara visual bentuk pola yang diberikan oleh variabel prediktor  $x$  dan variabel respon  $y$  tidak mempunyai pola yang pasti. Secara umum regresi nonparametrik memiliki bentuk fungsi sebagai berikut:

$$y = f(x_i) + \epsilon_i \quad i = 1, 2, \dots, n \quad (1)$$

Dengan  $f(x_i)$  merupakan fungsi regresi dan  $\epsilon_i$  merupakan residual. Regresi nonparametrik tidak terikat asumsi tertentu seperti halnya pada regresi parametrik sehingga menjadikan pendekatan nonparametrik lebih fleksibel (Hardle, Wolfgang. 2004). Kurva regresi nonparametrik hanya diasumsikan mulus (smooth), dimana data akan mencari bentuk taksirannya sendiri

dari  $f(x_i)$  tanpa dipengaruhi subyektifitas (Eubank, 1999).

### Regresi Smoothing Spline

Spline merupakan salah satu potongan polinomial (*piecewise polynomial*) tersegmen yang memiliki sifat fleksibilitas sehingga kurva yang terbentuk dapat menyesuaikan dengan data dan tetap memperhitungkan kemulusan kurva. Hal ini yang membedakan dengan polinomial. Pendekatan spline mempunyai keunggulan dalam mengatasi pola data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik knot. Titik knot merupakan titik yang menunjukkan terjadinya perubahan-perubahan perilaku kurva pada interval – interval yang berbeda. Eubank (1999), menuliskan secara umum fungsi spline dengan ordo  $m$  sebagai berikut:

$$f(x) = \beta_0 + \sum_{m=1}^M \beta_m x^m + \sum_{l=1}^k \beta_{m+l}(x - K_l)^m + \epsilon \quad (2)$$

Penerapan dengan mempertimbangkan nilai  $k=3$  atau dapat dikatakan splines kubik merupakan splines yang paling umum. splines kubik adalah fungsi kubik piecewise yang kontinu, dan memiliki turunan pertama dan kedua kontinu. Perhatikan bahwa kontinuitas dalam semua turunan tingkat rendah membuat splines sangat mulus (Smooth). Persamaan splines kubik natural dapat dituliskan dengan:

$$f(x) \approx \beta_0 + \beta_1 x + \sum_{l=1}^k \beta_{1l}(x - K_l)_+^3 \quad (3)$$

Estimasi fungsi splines  $f(x_i)$  adalah meminimumkan fungsi Penalized Least Square (PLS), yaitu:

$$PLS = \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int_0^1 [f''(x)]^2 dx \quad (4)$$

Untuk:

- a)  $\sum_{i=1}^n (y_i - f(x_i))^2$  : Jumlah Kuadrat Residual  
 b)  $\lambda \int_0^1 [f''(x)]^2 dx$  : Ukuran kemulusan kurva dalam memetakan data (*roughness penalty*)  
 c)  $\lambda$  : Parameter pemulus

Peranan parameter pemulus ( $\lambda$ ) sangat penting dalam mengontrol atau mengendalikan ketepatan model dan kemulusan kurva. Dengan bagian (a) merupakan jumlah kuadrat error atau fungsi jarak antara data dan taksiran, bagian (b) merupakan roughness penalty, yaitu ukuran kemulusan kurva dalam memetakan data, Nilai  $\lambda$  adalah 0 sampai  $\infty$  merupakan parameter pemulus, yaitu pengontrol keseimbangan antara kecocokan terhadap data dan kemulusan kurva (penalty). Apabila nilai  $\lambda \rightarrow 0$  maka ketepatan model akan tinggi namun kurva kurang mulus. Kurva akan cenderung mengikuti pola data, sehingga terlihat lebih kasar. Jika  $\lambda \rightarrow \infty$  maka ketepatan model akan rendah namun kurva sangat mulus. Kelengkungan kurva sangat kecil, bahkan mendekati fungsi linier nilai. Kurva akan cenderung menjauhi pola data sehingga jarak antara data dan dugaan sangat jauh.

Penentuan nilai  $\lambda$  optimum dapat didasarkan berdasarkan dua kriteria yakni

melihat nilai MSE dan GCV terkecil nilai MSE ditulis sebagai berikut:

$$MSE(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2 \quad (5)$$

Nilai GCV diperoleh dari hasil perluasan pada CV sehingga apabila CV ditulis:

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \frac{y_i - \hat{f}(x_i)}{1 - h_{ii}} \quad (6)$$

Untuk,  $h_{ii}$  elemen diagonal ke- i dari matriks  $H = X(X^T X)^{-1} X^T$  maka, nilai GCV:

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \frac{y_i - \hat{f}(x_i)}{1 - h_{ii}} \quad (6)$$

Nilai  $\lambda$  dikatakan optimal ketika MSE dan GCV memiliki nilai yang minimum. Namun Ketika terjadi pilihan antara nilai GCV minimum atau MSE minimum, maka dipilih nilai GCV minimum sebagai dasar penentuan  $\lambda$  optimal karena nilai MSE berdasarkan eror sedangkan semakin kasar kurva, akan semakin kecil nilai error nya.

Dalam penelitian ini digunakan data sekunder curah hujan bulanan dan harga cabai eceran bulanan di Kota Bandung periode 2016 hingga 2019. Dengan operasionalisasi variabel sebagai berikut:

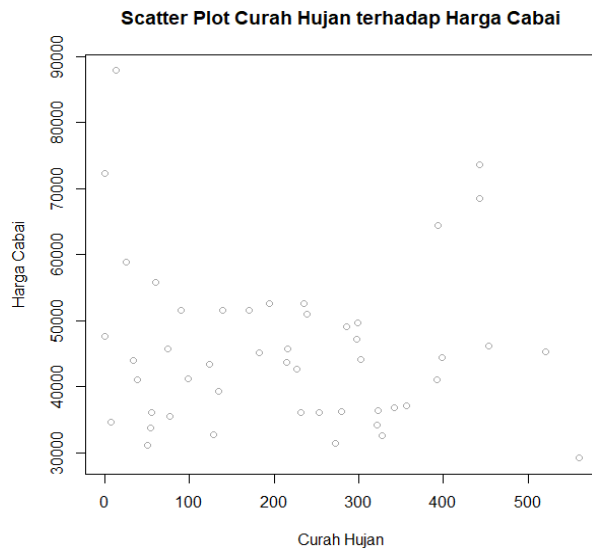
**Tabel 1. Operasionalisasi Variabel**

Variabel	Definisi Variabel	Sumber Data
Curah Hujan (X)	Curah Hujan adalah BMKG ketinggian air hujan yang terkumpul dalam penakar hujan pada tempat yang datar, tidak menyerap, tidak meresap dan tidak mengalir	
Harga cabai perbulan (Y)	Harga jual cabai di pasaran perkilogram	BPS

## HASIL DAN PEMBAHASAN

### Spesifikasi Model

Langkah awal yang dilakukan adalah menelaah spesifikasi model secara deskriptif terhadap pola sebaran datanya. Apabila pola sebaran data tidak dapat ditentukan fungsi slope dan intercept-nya, maka dapat digunakan pendekatan nonparametrik. Bentuk pola sebaran data kedua variabel tersebut dapat dilihat dengan scatter plot sebagai berikut:



**Gambar 1.** Pola Hubungan Data Curah Hujan terhadap Harga Cabai

Gambar 1 menunjukkan sebaran data menyebar secara acak, sehingga diduga analisis nonparametrik dapat digunakan. Untuk lebih jelasnya maka dilakukan uji normalitas residual dari regresi linier kedua variabel dengan hipotesis:

- Hipotesis:  
 $H_0$ : Residual data berdistribusi normal  
 $H_1$ : Residual data tidak berdistribusi normal
- Statistik uji: Shapiro Wilk

- Kriteria uji : tolak  $H_0$  ketika  $p - value < \alpha$

Shapiro-Wilk normality test

```
data: residu.1
W = 0.88518, p-value = 0.0002162
```

**Gambar 2.** Output R Uji Normalitas Residual

Hasil output pada Gambar 2, menunjukkan uji normalitas residual dengan menggunakan Shapiro Willk diperoleh p-value sebesar 0.0002162 dengan  $\alpha = 0.05$  maka,  $H_0$  ditolak sehingga dapat disimpulkan residual data tidak berdistribusi normal. Hal ini dapat memberikan alasan mengapa penelitian ini menggunakan pendekatan nonparametrik.

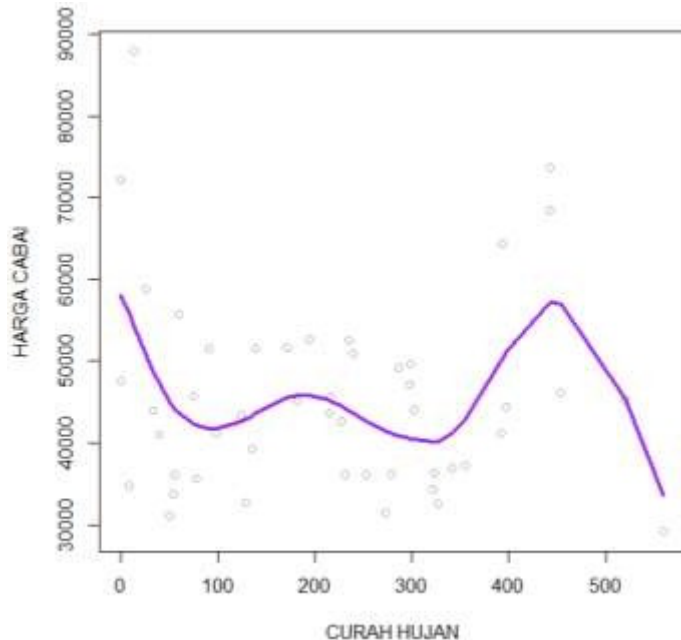
### Pemodelan Regresi Smoothig Splines

Pemodelan regresi Smoothing Splines dilakukan dengan menggunakan fungsi smooth.spline dengan kriteria Generalized Cross Validation (GCV). Sehingga diperoleh Nilai parameter  $\lambda$  optimum diperoleh secara default menggunakan kriteria GCV dengan knot sebanyak 50. Dari pemodelan yang dilakukan diperoleh parameter- parameter sebagai berikut:

**Tabel 2.** Parameter Regresi Smoothing Splines

Parameter	Nilai Parameter
Spar	0.8845879
$\lambda$	0.0005
Degree of Freedom	7.0281

Hasil dari Tabel 2, diperoleh secara default nilai parameter yang optimum sehingga nilai- nilai parameter tersebut dapat dilakukan untuk menentukan estimator pada model regresi smoothing spline. Dengan diperoleh nilai yang optimal yang dapat menggambarkan pola pergerakan data dengan baik seperti ditunjukkan pada gambar 3:



**Gambar 3.** Kurva Regresi Smoothing Spline

Dengan demikian, persamaan model regresi smoothing splines. Ditulis sebagai berikut:

$$\hat{Y} = 58024,15 + 580114,72 + 57288,80 (x-K_1)_+^3 + 56044,59 (x-K_2)_+^3 + 53677,86 (x-K_3)_+^3 + \dots + 3360,48 (x-K_{48})_+^3 + (7)$$

Seperti yang telah dijelaskan sebelumnya bahwa model regresi nonparametrik tidak dapat diinterpretasi.

### KESIMPULAN

Pendekatan regresi nonparametrik dengan smoothing splines memberikan solusi dalam pemodelan, dengan  $\lambda$  optimum senilai 0.0005 dan diperoleh knot sebanyak 48 maka kurva dengan pemulusan  $\lambda$  optimum dapat dikatakan dapat menggambarkan pola pergerakan data dengan baik.

Pada plot sebaran data terlihat jelas adanya outlier pada data, sehingga perlu dilakukan kajian lebih lanjut dalam mengatasi masalah outlier pada data.

### REFERENSI

- Eubank, R., (1999). Nonparametric Regression and Spline Smoothing. New York : Marcel Dekker
- Hardle, W., Muller, M., Sperlich, S., dan Werwatz, A. (2004). Nonparametric and Semiparametric Models, Springer-Verlag Berlin Heidelberg, New York.
- Fajri Rahmatul dkk. (2017). Nonparametric and Semiparametric Models, Springer-Verlag Berlin Heidelberg, New York.
- Mulyani, Sri. (2017). Pemodelan Analisis Faktor-Faktor Yang Mempengaruhi Harga Cabai Merah Di Kota Banda Aceh Jurnal Agribisnis Mahasiswa Pertanian Unsyiah Volume 2.