

## PEMODELAN PRODUKSI PADI MENGGUNAKAN REGRESI SEMIPARAMETRIK KERNEL

**Tiani Wahyu Utami<sup>1</sup>, Indah Manfaati Nur<sup>2</sup>, Endah Suryaningsih<sup>3</sup>**

<sup>1,2,3</sup> Program Studi Statistika, FMIPA, Universitas Muhammadiyah Semarang  
Alamat e-mail : tianiutami@unimus.ac.id

### ABSTRAK

Negara Indonesia merupakan negara agraris dengan salah satu bahan makanan pokok di Indonesia adalah padi. Produksi padi di provinsi Jawa Timur merupakan tertinggi di Indonesia. Tujuan penelitian ini adalah melakukan estimasi parameter parametrik dan nonparametrik serta memodelkan produksi padi di 38 kabupaten/kota provinsi Jawa Timur menggunakan regresi semiparametrik kernel. Regresi semiparametrik merupakan gabungan dari regresi parametrik dan nonparametrik. Regresi parametrik memiliki kurva yang berpola, misal linier, dll. Regresi nonparametrik memiliki kurva yang smooth yang tidak diketahui polanya, sehingga dalam hal ini diperlukan teknik *smoothing* yang digunakan untuk menghaluskan kurva yaitu salah satunya adalah kernel dengan estimator parameter Nadaraya-Watson dan metode estimasi parameter menggunakan *Weighted Least Square* (WLS), serta pemilihan *bandwidth* (h) optimal dengan metode *Generalized Cross Validation* (GCV). Variabel yang digunakan dalam penelitian yaitu produksi padi sebagai variabel respon, sedangkan variabel prediktornya yaitu luas panen, produktivitas, dan luas kerusakan. Data yang digunakan merupakan data sekunder dari website resmi Badan Pusat Statistika (BPS) provinsi Jawa Timur. Berdasarkan hasil estimasi menunjukkan bahwa regresi semiparametrik kernel mendapatkan nilai *bandwidth* (h) optimalnya 0,006 dengan GCV=15058865548. Koefisien determinasi ( $R^2$ ) sebesar 92,46% dan MSE sebesar 5797077303.

**Kata kunci :** Produksi Padi, Regresi Semiparametrik, GCV, Nadaraya-Watson, WLS

### PENDAHULUAN

Negara Indonesia merupakan negara agraris. sehingga memiliki hasil pangan yang melimpah. Hasil pangan di Indonesia berupa tanaman padi, jagung, dll. Tanaman pangan yang utama di Indonesia yaitu tanaman padi. Orang Indonesia lebih memilih mengonsumsi hasil olahan padi sebagai makanan pokok dibandingkan yang lainnya. Meskipun demikian, bahan makanan pokok lainnya tetap dikonsumsi sebagai produk sampingan.

Indonesia sebagai negara agraris seharusnya memiliki hasil produksi

padi yang melimpah, sehingga tidak perlu mengadakan impor beras dari negara tetangga. Jawa Timur merupakan provinsi yang paling banyak menghasilkan produksi padi di Negara Indonesia pada tahun 2015 [2]. Meskipun pada tahun 2015 provinsi tersebut berada diposisi pertama di Indonesia mengenai produksi padinya. Meskipun produksi padinya tertinggi produksinya, harus memiliki simpanan padi untuk kebutuhan pangan dimasa yang akan datang guna memenuhi konsumsi sehari-harinya.

Faktor-faktor yang mempengaruhi produksi padi antara lain yaitu luas

panen, produktivitas, dan luas puso/kerusakan[7]. [3] mendefinisikan luas panen merupakan luas lahan sawah yang biasa diambil hasilnya. Produktivitas merupakan hasil yang diperoleh tiap satuan luas. Luas rusak adalah jika tanaman mengalami serangan organisme pengganggu tumbuhan, bencana alam, sedemikian rupa sehingga hasilnya kurang dari 11% keadaan normal.

Analisis regresi merupakan salah satu metode untuk mengetahui hubungan antara variabel respon (Y) dengan variabel prediktor (X). Analisis regresi terdapat tiga macam pendekatan, yaitu parametrik, nonparametrik, dan semiparametrik. Pendekatan parametrik yaitu pendekatan dengan variabel data yang polanya diketahui. Pendekatan nonparametrik yaitu pendekatan dengan variabel data dengan kurva *smooth* yang polanya tidak diketahui, sehingga data tersebut akan mencari bentuk kurvanya sendiri [5]. Regresi semiparametrik merupakan gabungan dari regresi parametrik dan nonparametrik [8]. Estimasi kurva sangat bergantung pada perilaku data, sehingga diperlukan teknik *smoothing*. Pendekatan semiparametrik yaitu gabungan dari pendekatan parametrik dan nonparametrik, sehingga pada pendekatan semiparametrik lebih tepat apabila diterapkan pada kasus yang mengandung komponen parametrik dan nonparametrik [9]. Pada umumnya, teknik *smoothing* yang sering digunakan dalam penelitian adalah kernel. Pada penelitian ini digunakan estimator Nadaraya-Watson dan fungsi kernel Gaussian.

Regresi kernel merupakan teknik statistika nonparametrik yang digunakan untuk mengestimasi fungsi  $m(x)$  yang terdapat dalam model regresi nonparametrik. Nadaraya dan Watson pada tahun 1994 mendefinisikan estimator regresi

kernel, sehingga disebut estimator Nadaraya-Watson [6]. Berikut merupakan estimator Nadaraya-Watson:

$$m(x) = \frac{\sum_{i=1}^n Kh \left( \frac{x-x_i}{h} \right) y_i}{\sum_{i=1}^n Kh \left( \frac{x-x_i}{h} \right)}$$

*Bandwidth* (h) adalah parameter pemulus (*smoothing*) yang memiliki fungsi untuk mengontrol kurva yang sedang diestimasi. *Generalized Cross Validation (GCV)* merupakan salah satu metode dalam analisis regresi untuk menentukan *bandwidth* yang optimal [10].

## METODE PENELITIAN

### Sumber Data dan Variabel Penelitian

Data yang digunakan bersumber dari website resmi Badan Pusat Statistika (BPS) di 38 kabupaten/kota provinsi Jawa Timur tahun 2015. Data yang digunakan yaitu sebagai berikut :

**Tabel 1.** Variabel Penelitian

Data	Variabel
Produksi Padi	Respon (y)
Luas Panen	Parametrik (x)
Produktivitas	Nonparametrik (t <sub>1</sub> )
Luas Puso	Nonparametrik (t <sub>2</sub> )

Variabel-variabel tersebut berpacu dari penelitian-penelitian terdahulu yaitu oleh penelitian [7] dan penelitian [1].

### Metode Analisis

Langkah-langkah yang digunakan dalam melakukan penelitian regresi semiparametrik kernel Nadaraya-Watson yaitu sebagai berikut :

1. Mengkaji estimasi model regresi semiparametrik kernel sebagai berikut :

a. Diberikan data observasi ( $y_i, x_i, t_i$ ) yang memenuhi model regresi semiparametrik :

$$y_i = \beta_0 + \beta_1 x_i + \sum_{j=1}^2 m_j(t_j), j=1,2,3, \dots, n$$

b. Dari langkah a tersebut dapat dinyatakan sebagai berikut :

$$y_i^* = \sum_{j=1}^2 m_j(t_j) + e_{ij}$$

Atau dalam bentuk matriks dapat ditulis sebagai berikut

$$y^* = M + \varepsilon \quad \text{dengan}$$

$$y^* = y - x\beta$$

c. Menyatakan  $m_j(t_j)$  dapat didekati dengan estimator Nadaraya-Watson :

$$m(t) = \frac{\sum_{i=1}^n K_h\left(\frac{t-t_i}{h}\right) y_i}{\sum_{i=1}^n K_h\left(\frac{t-t_i}{h}\right)}$$

d. Mengestimasi  $m_j(t_j)$  dengan meminimalkan kriteria *Weighted Least Square* (WLS) :

$$WLS = \sum_{i=1}^n (y_i^* - \sum_{j=1}^2 m_j(t_j))^2 K_h(t_i - t)$$

Dengan  $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$  fungsi pembobot kernel

e. Untuk mendapatkan matriks  $A(h)$  yang berukuran  $N \times N$  dengan menyelesaikan persamaan berikut :

$$M = A(h)y^* \quad \text{atau} \quad M = A(h)(y - x\beta)$$

f. Mensubstitusikan  $M$  untuk memperoleh estimasi  $\beta$  yaitu  $\hat{\beta}$

2. Memodelkan produksi padi menggunakan regresi semiparametrik kernel :

a. Mengumpulkan data-data yang dibutuhkan dalam penelitian

b. Melakukan deskripsi data yang akan digunakan dalam penelitian dengan cara membuat *scatterplot* pada setiap variabel

c. Menentukan variabel komponen parametrik dan variabel komponen nonparametrik

d. Melakukan pemilihan *bandwidth* ( $h$ ) yang optimal dengan nilai *Generalized Cross Validation* (GCV) yang minimum

e. Memodelkan nilai *bandwidth* yang optimal

f. Menghitung nilai *Mean Square Error* (MSE)

## HASIL PENELITIAN

### Estimasi Regresi Semiparametrik Kernel

Diberikan data observasi ( $y_i, x_i, t_i$ ) sebanyak  $n$ , dengan  $y_i$  menyatakan variabel respon pengamatan untuk subjek ke- $i$ , sedangkan  $x_i$  dan  $t_i$  menyatakan variabel prediktor. Berikut bentuk model dengan regresi semiparametrik :

$$y = x\beta + TM + \varepsilon \quad \text{dengan}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{38} \end{bmatrix}, \varepsilon = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_{38} \end{bmatrix}, x = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ 1 & x_{31} \\ \vdots & \vdots \\ 1 & x_{381} \end{bmatrix},$$

$$M = \begin{bmatrix} m(t_{11}) + m(t_{12}) \\ m(t_{21}) + m(t_{22}) \\ m(t_{31}) + m(t_{32}) \\ \vdots \\ m(t_{381}) + m(t_{382}) \end{bmatrix}, T = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Persamaan model tersebut mengandung dua komponen yaitu komponen parametrik ( $x\beta$ ) dan komponen nonparametrik ( $TM$ ) yang dapat diasumsikan bahwa nilai  $\beta$  memiliki bentuk fungsi yang diketahui, sehingga dapat dibentuk sebagai berikut :

$$y^* = y - x\beta \quad \text{dengan}$$

$$y^* = TM + \varepsilon$$

Bentuk dari fungsi  $m(t_{ij})$  belum diketahui, sehingga didekati dengan estimator

Nadaraya-Watson yang dilakukan dengan cara meminimumkan kriteria *Weighted Least Square* (WLS). Metode *Weighted Least Square* (WLS) memiliki pembobot sebagai  $K_h$ , dengan  $K_h = \text{diag}(K_h(t_{i1}-t), K_h(t_{i2}-t), \dots, K_h(t_{i38}-t))$ . Matriks  $K_h$  merupakan matriks yang berisi fungsi pembobot dengan  $K(\cdot)$  merupakan fungsi kernel dan  $h$  yaitu *bandwidth*.  $K(\cdot)$  merupakan fungsi kernel Gaussian, yaitu

$$K(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{t_j - t}{h}\right)^2\right)$$

Metode estimasi *Weighted Least Square* (WLS) dalam bentuk matriks sebagai berikut :

$$\text{WLS} = (\mathbf{y}^* - \mathbf{TM})^T \mathbf{K}_h (\mathbf{y}^* - \mathbf{TM})$$

Dengan meminimumkan WLS tersebut, didapatkan matriks  $M$  sebagai berikut :

$$\hat{M} = (\mathbf{T}^T \mathbf{K}_h \mathbf{T})^{-1} \mathbf{T}^T \mathbf{K}_h \mathbf{y}^*$$

Sehingga, didapatkan matriks  $A(h)$  yaitu

$$\mathbf{A}(h) = (\mathbf{T}^T \mathbf{K}_h \mathbf{T})^{-1} \mathbf{T}^T \mathbf{K}_h$$

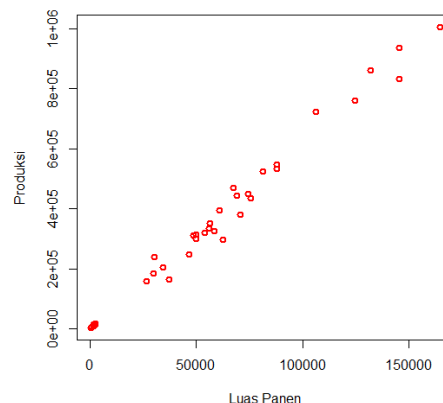
Dengan

$$\hat{M} = \begin{bmatrix} \hat{m}(t_{11}) + \hat{m}(t_{12}) \\ \hat{m}(t_{21}) + \hat{m}(t_{22}) \\ \hat{m}(t_{31}) + \hat{m}(t_{32}) \\ \vdots \\ \hat{m}(t_{381}) + \hat{m}(t_{382}) \end{bmatrix}$$

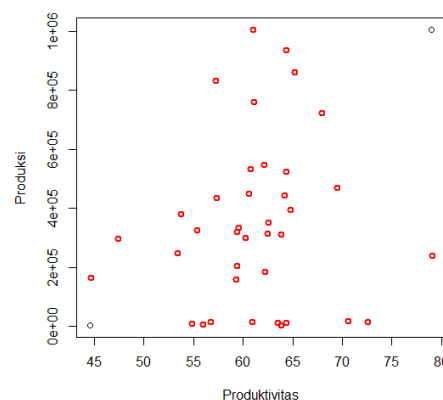
Atau dalam bentuk estimator Nadaraya-Watson sebagai berikut :

$$\hat{m}(t_{ij}) = \frac{\sum_{p=1}^{38} K\left(\frac{t_{ij} - t_{jp}}{h}\right) y_i^*}{\sum_{p=1}^{38} K\left(\frac{t_{ij} - t_{jp}}{h}\right)} = \sum_{i=1}^n W_i(t) (y_i - \sum_{k=1}^p \beta_k x_{ik})$$

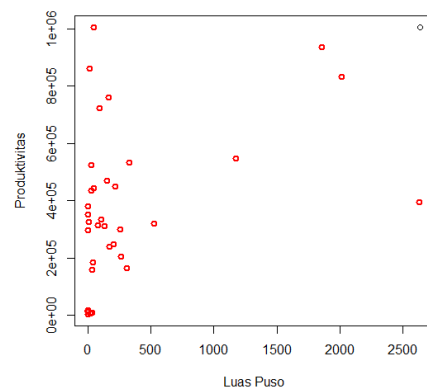
### Pola Hubungan Data Respon dan Prediktor



Gambar 1. Luas panen VS produksi



Gambar 2. Produktivitas VS produksi



Gambar 3. Luas Puso VS Produksi

Berdasarkan *scatterplot* diatas, dapat dikatakan bahwa gambar 1. hubungan produksi padi dengan luas panen termasuk komponen parametric, sedangkan variabel produksi padi dengan produktivitas serta hubungan produksi padi dengan luas puso termasuk komponen nonparametrik.

**Tabel 2.** Pemilihan *Bandwidth* (h) Optimal

No	<i>Bandwidth</i>	GCV
1	0,0001	15187967591
2	0,002	15187964224
3	<b>0,006</b>	<b>15058865548</b>
4	0,01	15779117764

Pemilihan *bandwidth* (h) optimal dilihat dari nilai GCV yang minimum, yaitu dengan nilai GCV 15058865548 terpilih *bandwidth* (h) optimalnya 0,006. Penentuan estimasi parameter  $\beta$  pada regresi parametrik yang diperoleh yaitu sebesar 6,20.

### Pemodelan Produksi Padi Menggunakan Regresi Semiparametrik Kernel

Estimasi masing-masing parameter yang sudah diperoleh, baik komponen parametrik maupun nonparametrik, maka dapat dibentuk pemodelan regresi semiparametrik kernel sebagai berikut :

$$\hat{y} = \mathbf{x}\hat{\beta} + \widehat{TM} + \varepsilon$$

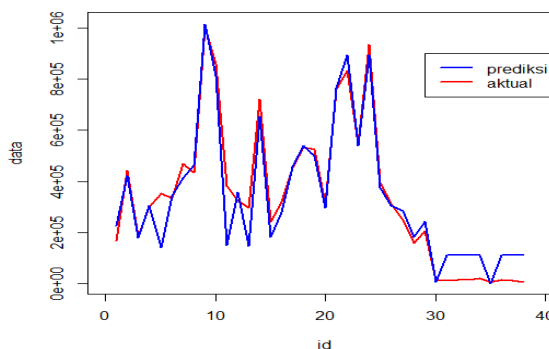
Berikut persamaan model semiparametrik kernel adalah sebagai berikut :

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \frac{\sum_{p=1}^{38} K_h \left( \frac{t_1 - t_{1p}}{h} \right) y_i^*}{\sum_{p=1}^{38} K_h \left( \frac{t_1 - t_{1p}}{h} \right)} + \frac{\sum_{p=1}^{38} K_h \left( \frac{t_2 - t_{2p}}{h} \right) y_i^*}{\sum_{p=1}^{38} K_h \left( \frac{t_2 - t_{2p}}{h} \right)}$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \frac{\sum_{p=1}^{38} K_h \left( \frac{t_1 - t_{1p}}{h} \right) (y_i - x_i \hat{\beta})}{\sum_{p=1}^{38} K_h \left( \frac{t_1 - t_{1p}}{h} \right)} + \frac{\sum_{p=1}^{38} K_h \left( \frac{t_2 - t_{2p}}{h} \right) (y_i - x_i \hat{\beta})}{\sum_{p=1}^{38} K_h \left( \frac{t_2 - t_{2p}}{h} \right)}$$

$$\hat{y} = -5021 + 6,20x_i + \frac{\sum_{p=1}^{38} K_{0,006} \left( \frac{t_1 - t_{1p}}{0,006} \right) (y_i - (-5021 - 6,20x))}{\sum_{p=1}^{38} K_{0,006} \left( \frac{t_1 - t_{1p}}{0,006} \right)} + \frac{\sum_{p=1}^{38} K_{0,006} \left( \frac{t_2 - t_{2p}}{0,006} \right) (y_i - (-5021 - 6,20x))}{\sum_{p=1}^{38} K_{0,006} \left( \frac{t_2 - t_{2p}}{0,006} \right)}$$

### Perbandingan Hasil Prediksi Produksi Padi dengan Data Aktual



**Gambar 4.** Perbandingan Prediksi Padi dengan Data Aktual

Berdasarkan grafik tersebut, dapat dikatakan bahwa data respon aktual dengan data respon prediksi memiliki pola data yang sama, artinya data prediksi mampu menutupi data aktual. Hal tersebut dapat dinyatakan bahwa pemodelan yang diperoleh telah sesuai untuk memprediksi produksi padi. Wilayah dengan nilai prediksi tertinggi yaitu berada di kabupaten Jember, sedangkan wilayah dengan nilai prediksi terendah berada di kota Mojokerto.

### Evaluasi Ketepatan Model

Evaluasi ketepatan model dapat dilihat pada nilai koefisien determinasi ( $R^2$ ) dan *Mean Square Error* (MSE). Semakin tinggi nilai koefisien determinasi maka model semakin bagus, serta semakin kecil nilai MSE maka model semakin baik. Hasil olahan program R menunjukkan bahwa nilai koefisien determinasi dan MSE yaitu sebesar 0.9246 dan 5797077303. Artinya, bahwa 92,46% variabel prediktor dapat menjelaskan variansi terhadap variabel respon, sedangkan sisanya yaitu 7,54% dipengaruhi oleh faktor lain.

**KESIMPULAN**

1. Hasil estimasi regresi Semiparametrik kernel

$$\widehat{\mathbf{M}} = (\mathbf{T}^T \mathbf{K} \mathbf{h} \mathbf{T})^{-1} \mathbf{T}^T \mathbf{K} \mathbf{h} \mathbf{y}^*$$

$$\widehat{\mathbf{M}} = \mathbf{A}(\mathbf{h}) \mathbf{y}^*$$

$$\widehat{\mathbf{M}} = \mathbf{A}(\mathbf{h})(\mathbf{y} - \mathbf{x}\beta)$$

sehingga,

$$\mathbf{A}(\mathbf{h}) = (\mathbf{T}^T \mathbf{K} \mathbf{h} \mathbf{T})^{-1} \mathbf{T}^T \mathbf{K} \mathbf{h}$$

atau dalam estimator Nadaraya-Watson adalah sebagai berikut :

$$\widehat{m}(t_{ij}) = \frac{\sum_{p=1}^{38} K\left(\frac{t_{ij}-t_{jp}}{h}\right) y_i^*}{\sum_{p=1}^{38} K\left(\frac{t_{ij}-t_{jp}}{h}\right)} = \sum_{i=1}^n W_i(t)(y_i - \sum_{k=1}^p \beta_k x_{ik})$$

2. Pemodelan Produksi Padi menggunakan Regresi Semiparametrik Kernel :

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i + \left( \frac{\sum_{p=1}^{38} K_h\left(\frac{t_1-t_{1p}}{h}\right) y_i^*}{\sum_{p=1}^{38} K_h\left(\frac{t_1-t_{1p}}{h}\right)} + \left( \frac{\sum_{p=1}^{38} K_h\left(\frac{t_2-t_{2p}}{h}\right) y_i^*}{\sum_{p=1}^{38} K_h\left(\frac{t_2-t_{2p}}{h}\right)} \right) \right)$$

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i + \left( \frac{\sum_{p=1}^{38} K_h\left(\frac{t_1-t_{1p}}{h}\right) (y_i - x_i \widehat{\beta})}{\sum_{p=1}^{38} K_h\left(\frac{t_1-t_{1p}}{h}\right)} + \left( \frac{\sum_{p=1}^{38} K_h\left(\frac{t_2-t_{2p}}{h}\right) (y_i - x_i \widehat{\beta})}{\sum_{p=1}^{38} K_h\left(\frac{t_2-t_{2p}}{h}\right)} \right) \right)$$

$$\widehat{y}_i = -5021 + 6,20x_i + \left( \frac{\sum_{p=1}^{38} K_{0,006}\left(\frac{t_1-t_{1p}}{0,006}\right) (y_i - (-5021 - 6,20x))}{\sum_{p=1}^{38} K_{0,006}\left(\frac{t_1-t_{1p}}{0,006}\right)} + \left( \frac{\sum_{p=1}^{38} K_{0,006}\left(\frac{t_2-t_{2p}}{0,006}\right) (y_i - (-5021 - 6,20x))}{\sum_{p=1}^{38} K_{0,006}\left(\frac{t_2-t_{2p}}{0,006}\right)} \right) \right)$$

Model regresi tersebut memiliki nilai koefisien determinasi ( $R^2$ ) sebesar 92,46% dan nilai MSE sebesar 5797077303. Hal tersebut menunjukkan bahwa variabel prediktor luas panen (x) dan variabel prediktor produktivitas ( $t_1$ ) dan luas puso ( $t_2$ ) memiliki kemampuan menjelaskan varians terhadap variabel respon (y) sebesar 92,46% dan 7,54% dijelaskan oleh variabel lainnya yang tidak termasuk dalam penelitian ini.

**UCAPAN TERIMKASIH**

Penulis mengucapkan terimakasih kepada DRPM Kemenristek Dikti atas Hibah Penelitian Dosen Pemula 2017-

2018 dengan no SK 014/UNIMUS.J/PJ/PG/2018.

**DAFTAR PUSTAKA**

[1] Arnanda, F, dan Karim, A. 2016. Pemodelan Produksi Padi di Provinsi Jawa Tengah Dengan Pendekatan Spatial Econometrics. *Jurnal Statistika*, 4(2): 20-27.

[2] Badan Pusat Statistik. 2015. Data BPS Provinsi Jawa Timur. BPS Jawa Timur. <http://jatim.bps.go.id>. 5 Desember 2017 (10:26).

[3] Badan Pusat Statistik. 2015. *Produksi Padi dan Palawija di Jawa Timur 2015-2016*. BPS Jawa Timur. Surabaya.

[4] Djiwandi, 1980. *Penyuluhan Pertanian*. Universitas Sebelas Maret. Surakarta

[5] Eubank, R. L. 1999. *Nonparametric Regression and Spline Smoothing*. New York: Marcel Dekker, Inc.

[6] Hardle, W. 1994. *Applied Nonparametric Regression*. Cambridge University Press. New York.

[7] Ishaq, M, dkk. 2016. Analisis Faktor-faktor yang Mempengaruhi Produksi Padi di Provinsi Padi di Jawa Timur Menggunakan Regresi Semiparametrik Spline. *Jurnal Sains dan Seni ITS* 5(2): 420-425.

[8] Ruppert, D., Wand, M. P., & Carrol, R. J. 2003. *Semiparametric Regression*. New York: Cambridge University Press.

[9] Utami, T. W. 2013. Estimasi Kurva Regresi Semiparametrik Pada Data Longitudinal Berdasarkan Estimator Polinomial Lokal. *Jurnal Statistika* 1(1): 30-36.

[10] Wu, H., dan Zhang, J.T. 2006. *Nonparametric Regression Methods for Longitudinal Data Analysis*. New York: John Wiley and Sons, Inc.