

**PEMODELAN REGRESI NONPARAMETRIK MENGGUNAKAN PENDEKATAN
 POLINOMIAL LOKAL PADA BEBAN LISTRIK DI KOTA SEMARANG**

Suparti¹, Alan Prahutama²

^{1,2} Departemen Statistika, Universitas Diponegoro

e-mail: suparti702@gmail.com

DOI: 10.14710/medstat.9.2.85-93

Abstract

Semarang is the provincial capital of Central Java, with infrastructure and economic's growth was high. The phenomenon of power outages that occurred in Semarang, certainly disrupted economic development in Semarang. Large electrical energy consumed by industrial-scale consumers and households in the San Francisco area, monitored or recorded automatically and presented into a historical data load power consumption. Therefore, this study modeling the load power consumption at a time when not influenced by the use of electrical load $(t-1)$ -th. Modeling using nonparametric regression approach with Local polynomial. In this study, the kernel used is a Gaussian kernel. In local polynomial modeling, determined optimum bandwidth. One of the optimum bandwidth determination using the Generalized Cross Validation (GCV). GCV values obtained amounted to 1425.726 with a minimum bandwidth of 394. Modelling generate local polynomial of order 2 with MSE value of 1408.672.

Keywords: *electrical load, local polinomial, gaussian kernel, GCV.*

1. PENDAHULUAN

Semarang merupakan ibu kota provinsi Jawa Tengah, dengan infrastruktur dan pertumbuhan ekomi yang tinggi. Fenomena pemadaman listrik yang terjadi di Semarang, tentunya mengganggu perkembangan ekonomi yang ada di Semarang. Hal ini dikarenakan investor enggan untuk berinvestasi di kota Semarang dikarenakan sering terjadi pemadaman listrik. Oleh karena itu, Kota Semarang harus memiliki suplai daya yang memadai untuk memenuhi kebutuhan energi listrik di areanya. Besar energi listrik yang dikonsumsi oleh konsumen skala industri maupun rumah tangga di area Semarang, terpantau atau tercatat secara otomatis dan tersaji menjadi data historis beban pemakaian listrik harian per 30 menit selama 24 jam ataupun menjadi data beban puncak pemakaian listrik.

Menurut Mujiman dan Priyosusilo (2012), data beban puncak listrik adalah data beban pemakaian energi listrik maksimal yang tercatat berdasarkan waktu yaitu, harian, mingguan, maupun bulanan. Beban puncak ini biasanya terjadi pada pukul 10.00 dan 19.00. Beban puncak terjadi ketika kebutuhan listrik konsumen menanjak ke titik yang paling tinggi di satu waktu tertentu, baik dalam rentang waktu jam, hari, minggu, bulan, hingga tahun. Pemodelan beban listrik diperlukan sebagai dasar untuk prediksi nilai beban listrik di kota Semarang.

Salah satu analisis statistika yang digunakan untuk pemodelan adalah analisis regresi. Analisis regresi merupakan analisis hubungan antara variabel variabel respon (Y) dengan prediktor (X). Pendekatan regresi dibedakan menjadi dua yaitu pendekatan secara parametrik dan pendekatan nonparametrik. Pendekatan parametrik merupakan pemodelan regresi yang terikat dengan asumsi-asumsi dalam regresi. Asumsi-asumsi tersebut antara lain multikolinieritas, residual normalitas, homokedastisitas residual, dan nonautokorelasi. Sedangkan pendekatan regresi nonparametrik tidak ada asumsi-asumsi yang harus dipenuhi dalam pemodelan. Regresi parametrik dilakukan apabila bentuk kurva regresinya diketahui. Sedangkan regresi nonparametrik dilakukan jika bentuk kurva regresinya tidak diketahui. Regresi semiparametrik digunakan jika sebagian bentuk kurva regresinya tidak diketahui sedangkan yang sebagian lainnya diketahui. Kurva regresi nonparametrik diasumsikan *smooth* (mulus/halus) yang termuat dalam suatu ruang fungsi tertentu misalnya ruang *sobolev* (Eubank *et al*, 2004). Pada prosedur regresi nonparametrik, data akan mencari sendiri bentuk kurva regresinya tanpa dipengaruhi oleh subjektivitas peneliti. Beberapa model regresi nonparametrik yang telah dikembangkan antara lain *Penalized Spline* (Kadiri *et al*, 2010), *Smoothing Spline* (Eubank *et al*, 2004), Regresi Spline Multirespon (Lestari *et al*, 2010), Regresi Menggunakan Kernel (Hu *et al*, 2004), *Kernel of Smoothing Spline* (Lin *et al*, 2004) dan Polinomial Lokal (Wu dan Zhang, 2006)

Polinomial lokal mempunyai beberapa kelebihan antara lain dapat mengurangi asimtotik bias dan menghasilkan estimasi yang baik (Welsh dan Yee, 2005). Estimasi Polinomial Lokal dapat menggunakan WLS (*Weighted Least Square*) dengan cara meminimumkannya (Takezawa, 2006). Dalam regresi polinomial lokal tingkat kemulusan fungsinya ditentukan bandwidthnya. Penentuan *bandwidth* optimal dapat menggunakan metode GCV (*Generalized Cross Validation*) (Wu dan Zhang, 2006). Pada penelitian ini akan dimodelkan beban listrik di Kota Semarang menggunakan model Polinomial Lokal. Pemodelan ini nantinya diharapkan bisa digunakan untuk peramalan beban listrik di Kota Semarang.

2. TINJAUAN PUSTAKA

2.1. Estimator Kernel

Salah satu metode estimasi pada polinomial lokal adalah menggunakan WLS (*Weighted Least Square*) sehingga diperlukan pembobotan. Salah satu pembobotan yang digunakan untuk mendapatkan estimasi adalah Fungsi Kernel (Eubank, 1988). Fungsi Kernel K dengan *bandwidth* h didefinisikan sebagai berikut:

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right); \quad -\infty < x < \infty \text{ dan } h > 0. \quad (1)$$

Sifat-sifat dari fungsi kernel adalah sebagai berikut:

1. $K(x) \geq 0$ untuk semua x
2. $\int_{-\infty}^{\infty} K(x) dx = 1$
3. $\int_{-\infty}^{\infty} xK(x) dx = 0$
4. $\int_{-\infty}^{\infty} x^2 K(x) dx = \sigma^2 > 0$

Sedangkan menurut Hardle (1990) terdapat beberapa jenis fungsi Kernel:

1. Kernel Uniform : $K(x) = \frac{1}{2}; I(|x| < 1)$
2. Kernel Segitga : $K(x) = (1 - |x|); I(|x| < 1)$
3. Kernel Eparichnikov : $K(x) = \frac{3}{4}(1 - x^2); I(|x| < 1)$
4. Kernel Kuadrat : $K(x) = \frac{15}{16}(1 - x^2)^2; I(|x| < 1)$
5. Kernel Twiweight : $K(x) = \frac{35}{32}(1 - x^2)^3; I(|x| < 1)$
6. Kernel Cosinus : $K(x) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}x\right); I(|x| < 1)$
7. Kernel Gaussian : $K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$

2.2. Polinomial Lokal

Model regresi nonparametrik dapat dinyatakan sebagai berikut:

$$y_i = \eta(x_i) + e_i, \quad i = 1, 2, \dots, n \quad (2)$$

dimana $\eta(x_i)$ adalah fungsi yang tidak diketahui bentuknya dengan prediktor x_i dan e_i adalah residual pengamatan ke- i . Salah satu pemodelan regresi nonparametrik menggunakan polinomial lokal. Pada dasarnya pemodelan polinomial lokal menggunakan prinsip deret Taylor, yang menyatakan bahwa setiap fungsi mulus dapat secara lokal didekati dengan polinomial dari beberapa derajat. Misalkan x_0 merupakan titik awal yang ditentukan dimana fungsi η akan diestimasi dengan estimator Kernel. Melalui deret Taylor, $\eta(x_i)$ pada persamaan (2) dapat didekati secara lokal oleh polinomial berderajat p sebagai berikut (Wu dan Zhang, 2006):

$$\eta(x_i) \approx \eta(x_0) + (x_i - x_0)\eta^{(1)}(x_0) + \dots + (x_i - x_0)^p \eta^{(p)}(x_0)/p! \quad (3)$$

Misalkan $\beta_r(x_0) = \eta^{(r)}(x_0)/r!$, $r = 0, 1, 2, \dots, p$, maka persamaan (3) dapat ditulis menjadi:

$$\eta(x_i) \approx \beta_0(x_0) + (x_i - x_0)\beta_1(x_0) + \dots + (x_i - x_0)^p \beta_p(x_0) \quad (4)$$

Dari persamaan (4), dapat dinyatakan sebagai berikut :

$$\boldsymbol{\eta}(x_i) = \mathbf{X}\boldsymbol{\beta} \quad (5)$$

dengan
$$\mathbf{X} = \begin{bmatrix} 1 & (x_1 - x_0) & (x_1 - x_0)^2 & \dots & (x_1 - x_0)^p \\ 1 & (x_2 - x_0) & (x_2 - x_0)^2 & \dots & (x_2 - x_0)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (x_n - x_0) & (x_n - x_0)^2 & \dots & (x_n - x_0)^p \end{bmatrix}$$

$$\boldsymbol{\eta}(x_i) = [\eta(x_1) \quad \eta(x_2) \quad \dots \quad \eta(x_n)]^T \text{ dan } \boldsymbol{\beta} = [\beta_0(x_0) \quad \beta_1(x_0) \quad \dots \quad \beta_p(x_0)]^T$$

Untuk mendapatkan estimator $\hat{\boldsymbol{\beta}}$ dilakukan dengan meminimumkan kriteria *Weighted Least Square* (WLS) sebagai berikut:

$$\sum_{i=1}^n (y_i - x_i \beta)^2 K_h(x_i - x_0) . \quad (6)$$

$K_h(x_i - x_0) = \frac{K((x_i - x_0)/h)}{h}$ dengan K merupakan fungsi kernel dan h adalah sebuah *bandwidth*, sehingga kriteria WLS dapat ditulis sebagai berikut:

$$(\mathbf{y} - \mathbf{X}\beta)^T \mathbf{K}_h (\mathbf{y} - \mathbf{X}\beta) \quad (7)$$

dimana $\mathbf{K}_h = \text{diag}(K_h(x_1 - x_0), \dots, K_h(x_n - x_0))$, sehingga estimasi untuk $\hat{\beta}$ diberikan oleh:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{K}_h \mathbf{X})^{-1} \mathbf{X}^T \mathbf{K}_h \mathbf{y} \quad (\text{Wu dan Zhang, 2006}).$$

Hal yang harus diperhatikan dalam polinomial lokal. Pemilihan parameter *bandwidth* h memiliki peran penting dalam melakukan estimasi. Pemilihan *bandwidth* yang terlalu besar mengakibatkan plot hasil estimasi model akan menjauhi plot data awal sehingga menjadi sangat halus (*oversmoothing*). Ketika *bandwidth* terlalu besar maka bias pemodelannya akan besar dan keragaman akan kecil. Pemilihan *bandwidth* yang terlalu kecil mengakibatkan plot hasil estimasi model yang berliuk-liuk (*undersmoothing*). Ketika *bandwidth* terlalu kecil maka bias pemodelannya akan kecil dan keragaman akan besar. Oleh karena itu, harus dicari *bandwidth* yang optimal untuk menyeimbangkan bias dan keragaman agar diperoleh estimasi yang baik. Selain itu, pemilihan derajat pada polinomial lokal juga harus diperhatikan dimana derajat yang besar akan mengurangi bias pemodelan, tetapi akan menyebabkan keragaman yang besar.

2.3. Generalized Cross Validation (GCV)

Salah satu cara menentukan *bandwidth* yang optimal dengan menggunakan metode GCV (*Generalized Cross Validation*). Fungsi GCV diberikan sebagai berikut:

$$GCV(h) = \frac{MSE(h)}{\left(\frac{1}{n} \text{tr}[\mathbf{I} - \mathbf{A}(h)]\right)^2}$$

dengan $MSE(h) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$, dan $\mathbf{A}(h)$ diperoleh dari hubungan $\hat{\mathbf{y}} = \mathbf{A}(h)\mathbf{y}$ sehingga

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

Nilai GCV terkecil akan memberikan nilai *bandwidth* h yang optimal.

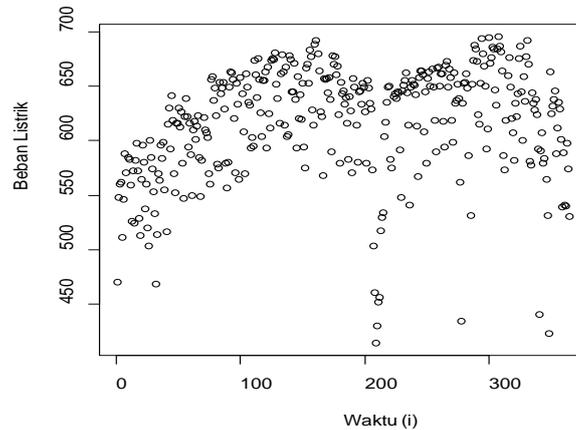
3. METODE PENELITIAN

Penelitian ini menggunakan data beban listrik di kota Semarang dari Januari 2014 sampai dengan Desember 2015. Pada penelitian ini data dibagi menjadi dua bagian yaitu data training dan data testing. Data training dari Januari 2014 sampai dengan November 2014, sedangkan data testing adalah pada bulan Desember 2014. Variabel respon pada penelitian ini adalah penggunaan beban listrik pada waktu ke- t (Y_t), sedangkan variabel prediktor adalah penggunaan beban listrik pada waktu ke- $(t-1)$ atau Y_{t-1} . Fungsi kernel yang digunakan dalam penelitian ini adalah fungsi kernel Gaussian.

4. HASIL DAN PEMBAHASAN

4.1. Statistika Deskriptif

Berdasarkan Gambar 1, terlihat bahwa beban listrik di kota Semarang dari Januari-November 2014 menyebar secara acak. Terlihat juga beberapa kejadian berada pada batas interval tertentu.



Gambar 1. Plot Data Beban Listrik Kota Semarang Tahun 2014

Statistika deskriptif beban listrik di Kota Semarang dari Januari-November 2014 disajikan pada Tabel 1.

Tabel 1. Statistika Deskriptif Data Beban Listrik di Kota Semarang Januari-November 2014

Minimum	Maksimum	Rata-Rata	Median	Variansi
414.9	695.4	616.7	628.6	2678.246

Data beban listrik Kota Semarang pada tahun 2014 mempunyai nilai minimum sebesar 414.9 dan nilai maksimum 695.4 dengan rata-rata 616.7 dan varian 2678.246. Data minimum terjadi pada tanggal 28 Juli dan data maksimum terjadi pada tanggal 4 November.

4.2. Pemodelan Beban Listrik di Kota Semarang Menggunakan Polinomial Lokal

Dalam regresi polinomial lokal, pemilihan *bandwidth* sangat penting untuk diperhatikan. *Bandwidth* menjadi salah satu faktor yang paling utama dalam penentuan estimasi parameter di regresi polinomial lokal. Selain *bandwidth*, titik lokal (x_0) dan orde polinomial juga merupakan hal yang penting. Metode yang digunakan dalam menentukan nilai optimal tersebut yaitu dengan *Generalized Cross Validation* (GCV). Untuk mendapatkan nilai x_0 yang optimal, dilakukan dengan *trial – error* dari sekumpulan nilai x_0 yang terdapat dalam interval tertentu yang dicobakan. Karena data minimum 414.9 dan data maksimum 695.4 maka nilai x_0 dicobakan pada interval 415 sampai dengan 695.

Sedangkan h dicoba-coba mulai dari 1 sampai diperoleh nilai optimal dan orde polinomial dicoba orde 2,3,4,5 dan 6. Sehingga estimasi polinomial lokal yang optimal merupakan kombinasi antara *bandwidth*, titik lokal (x_0) dan orde polinomial lokal. Setelah dilakukan beberapa trial dan error diperoleh GCV minimum terletak pada polinomial orde 2. Ada beberapa nilai X_0 dan h yang mempunyai GCV minimum. Namun peneliti memilih h terkecil dengan GCV minimum sebesar 1425.746 yaitu jatuh pada nilai $x_0 = 415$ dan $h = 394$.

Secara umum model polinomial lokal orde 2 adalah:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1(X_i - x_0)$$

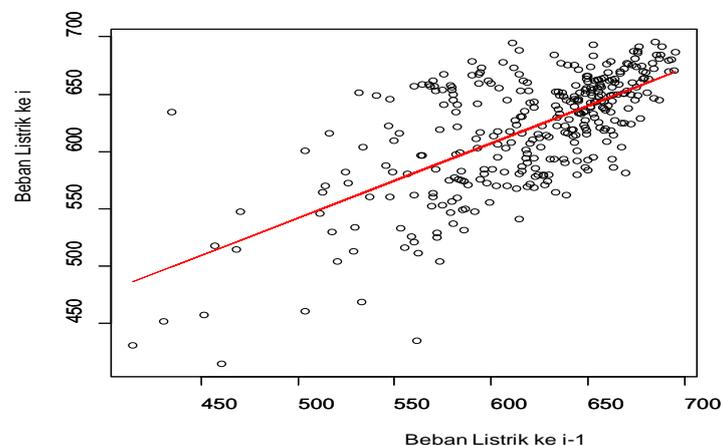
Tabel 2. Hasil Estimasi Pemodelan menggunakan Polinomial Lokal Orde 2 (Linier)

Variabel	Hasil Estimasi
$\hat{\beta}_0$	486.28
$\hat{\beta}_1$	0.653
x_0	415

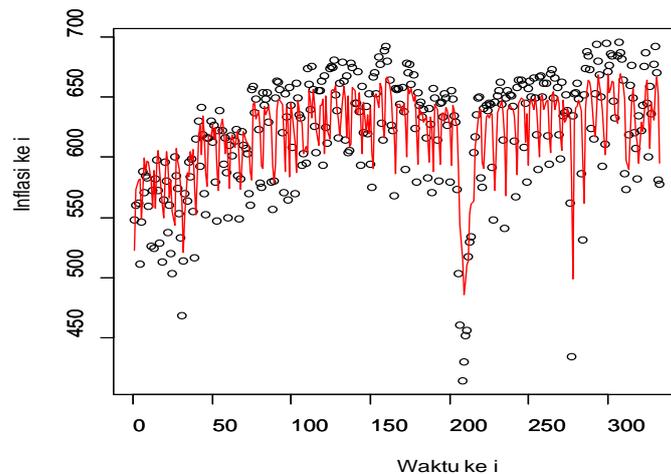
Berdasarkan Tabel 2, model polinomial yang terbentuk adalah:

$$\hat{y}_i = 486.28 + 0.653(X_i - 415)$$

Gambar plot data modifikasi dan estimasinya disajikan dalam Gambar 2. Setelah data dikembalikan dalam bentuk aslinya dan nilai estimasinya disajikan dalam Gambar 3.



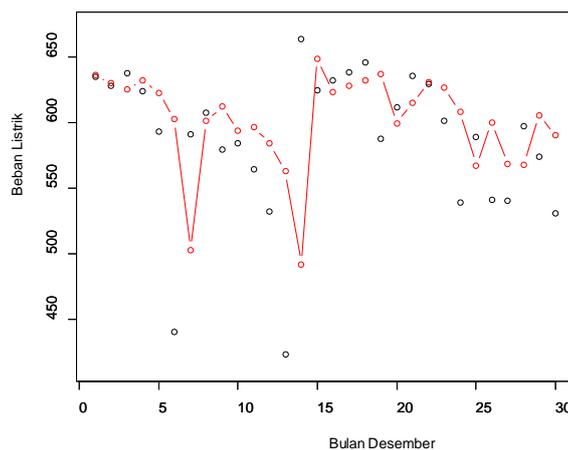
Gambar 2. Estimasi Model Data Modifikasi



Gambar 3. Estimasi Model Data Asli

Berdasarkan Gambar 3 dapat dilihat bahwa model polinomial lokal yang terpilih memiliki kemampuan yang cukup bagus dalam mengestimasi kurva regresi data aktual. Data hasil estimasi mengikuti pergerakan data aktual namun cenderung kurang fit. Namun, kemampuan yang cukup bagus dalam pemodelan belum bisa menjamin bahwa regresi polinomial lokal bagus ketika digunakan peramalan. Oleh karena itu, model perlu dicobakan pada data *out sampel* pada bulan Desember 2014.

Untuk melihat apakah model polinomial lokal orde 2 dengan *bandwidth* 394 dan titik lokal 215 bagus digunakan untuk peramalan, maka model tersebut perlu dicobakan terhadap data *out sampel* Bulan Desember dan dihitung MAPE nya.



Gambar 4. Data Aktual dan Data Prediksi *Out Sampel*

Tabel 3. Perbandingan Data Aktual *Out Sampel* dengan Data Estimasi *Out Sampel* Model Polinomial Lokal

Tanggal	Data Aktual <i>Out Sampel</i>	Data Estimasi <i>Out Sampel</i>	Tanggal	Data Aktual <i>Out Sampel</i>	Data Estimasi <i>Out Sampel</i>
2 Desember	634.5886	635.9049	17 Desember	632.1266	623.2457
3 Desember	628.0057	629.6593	18 Desember	637.9046	628.0518
4 Desember	637.7018	625.3611	19 Desember	645.5569	631.8245
5 Desember	623.8896	631.6920	20 Desember	587.7743	636.8209
6 Desember	592.7005	622.6735	21 Desember	611.6911	599.0925
7 Desember	440.4632	602.3090	22 Desember	635.608	614.7087
8 Desember	590.7902	502.9074	23 Desember	629.386	630.3249
9 Desember	607.5718	601.0617	24 Desember	601.5433	626.2623
10 Desember	579.3455	612.0190	25 Desember	539.1744	608.0828
11 Desember	584.0842	593.5890	26 Desember	588.9178	567.3597
12 Desember	564.6017	596.6831	27 Desember	540.9547	599.8391
13 Desember	531.9338	583.9622	28 Desember	540.2697	568.5222
14 Desember	423.7644	562.6321	29 Desember	597.4321	568.0749
15 Desember	663.3461	492.0041	30 Desember	574.2065	605.3984
16 Desember	624.7659	648.4362	31 Desember	530.8385	590.2335

Dari data *out sampel* dan estimasinya diperoleh nilai MAPE sebesar 7.47 % yang mengindikasikan model mempunyai kinerja yang sangat bagus.

5. KESIMPULAN

Pemodelan beban listrik di Kota Semarang menggunakan pendekatan polinomial lokal dengan fungsi kernel Gaussian didapat *bandwidth* optimal sebesar 394. Model tersebut menghasilkan nilai MSE model dengan data *in sample* sebesar 1408.672 dengan orde optimum polinomial lokalnya adalah orde 2 (linier). Pada data *out sample* menghasilkan nilai estimasi dengan MAPE yang hanya 7.47%.

PENGHARGAAN

Pada penelitian ini, penulis ingin mengucapkan terima kasih kepada DRPM Kemenristek Dikti sebagai pemberi dana dalam skema penelitian Fundamental yang didanai pada tahun 2016. Penulis juga ingin berterima kasih kepada Departemen Statistika Universitas Diponegoro Semarang.

DAFTAR PUSTAKA

- Eubank, R.L., 1988, *Nonparametric Regression and Spline Smoothing*, second edition, Marcell dekker, Inc., New York.
- Eubank, R.L., Huang, C., Maldonado, Y.M., Wang, N., Wang, S., dan Buchanan, R.J., 2004, Smoothing Spline Estimation in Varying-Coefficient Models, *J. R. Statist. Soc.*, 66, Part 3, pp. 653-667.

- Hardle, W, 1990, *Applied Nonparametric Regression*, Cambridge University, New York.
- Hu, Z., Wang, N., dan Carroll, R.J., 2004, Profile Kernel Versus Backfitting In The Partially Linier Models For Longitudinal Or Clustered Data, *Biometrika*, Vol. 91, No. 2, pp. 251-262.
- Kadiri, M. Al., Carroll R.J. dan Wand, M.P., 2010, Marginal Longitudinal Semiparametric Regression Via Penalized Spline, *Statistics and Probabbility Letters*, 80, pp. 1242-1252.
- Lestari, B., Budiantara, I.N., Sunaryo, S., dan Mashuri, M., 2010, Spline Estimator in Multi-Response Nonparametric Regression Model with Unequal Correlation of Errors, *Journal of Mathematics and Statistics*, Vol. 6, No. 3, hal. 327-332.
- Lin, X., Wang, N., Wels, A.H., dan Carroll, R.J., 2004, Equivalent Kernels Of Smoothing Splines In Nonparametrics Regression For Clustered/Longitudinal Data, *Biometrika*, Vol. 91, No. 1, pp. 177-193.
- Mujiman dan Priyosusilo, L., 2012, *Permodelan Beban Puncak Gardu Induk Waters dengan Program Aplikasi Microsoft Excel*, Prosiding Seminar Nasional Aplikasi Sains dan Teknologi (SNAST) Periode III, Yogyakarta.
- Takezawa, K., 2006. *Introduction to Nonparametric Regression*, John Wiley & Sons Inc.,New Jersey.
- Welsh, A.H dan Yee, T.Y. 2005, Local Regression for Vector Responses, *Journal of Statistical Planning and Inference*, Vol. 136, hal. 3007-3031.
- Wu, H. dan Zhang, J.T., 2006, *Nonparametric Regression Methods for Longitudinal Data Analysis*, A John-Wiley and Sons Inc. Publication, New Jersey.