Jurnal UJMC, Volume 3, Nomor 2, Hal. 9 - 16
pISSN : 2460-3333 eISSN : 2579-907X

REGRESI NONPARAMETRIK MENGGUNAKAN METODE ROBUST DAN *CROSS-VALIDATION*
(STUDI KASUS MAHASISWA STIA MUHAMMADIYAH SELONG)

Ratna Yuniarti[1], Widya Hartati [2]

[1,2] STIA Muhammadiyah Selong , yuniarti.edu@gmail.com

Abstract This paper is about comparison of estimation regression function between nonparametric regression kernel using Nadaraya Watson estimator with simple linear regression. In addition, a long-time case study was given in obtaining the first job. The data used are fresh graduate student, of Public Administration departement in STIA Muhammadiyah Selong graduate 2016. In this case we will see the relation between Cumulative Achievement Index (*X*) with long waiting time to get the job (*Y*). The software statistics with the help of software applications R. For the selection of the best model is by Cross-Validation. In this paper, given the theory and prove about cross-validation method are described. And then, this paper describes the Robust method if there is outlier.

Keywords: *nonparametric regression kernel, estimator Nadaraya-Watson, cross validation, outlier, and Robust.*

Abstrak Penelitian ini tentang perbandingan hasil estimasi fungsi regresi antara model regresi nonparametrik kernel menggunakan estimator Nadaraya Watson dengan regresi linear sederhana. Selain itu diberikan studi kasus lama waktu tunggu dalam mendapatkan pekerjaan pertama. Data yang yang digunakan adalah data mahasiswa *fresh graduate* Program Studi Administrasi Publik STIA Muhammadiyah Selong angkatan 2016. Dalam hal ini akan dilihat hubungan antara Indeks Prestasi Kumulatif (*X*) dengan lama waktu tunggu mendapat pekerjaan (*Y*). Adapun *software* statistik dengan bantuan aplikasi *software R*. Untuk pemilihan model terbaik yaitu dengan metode *Cross-Validation* (Validasi Silang). Pada penelitian ini diuraikan teori dan pembuktian tentang metode *cross –validation.* Selain itu, pada paper ini dijelaskan tentang metode Robust jika terdapat pencilan.

Kata kunci: *regresi nonparametrik kernel, estimator Nadaraya-Watson, cross validation, outlier, dan Robust*

1 Pendahuluan

Analisis regresi dengan pendekatan parametrik digunakan apabila formulasi hubungan antara variabel prediktor X dan variabel respon Y diketahui. Namun, apabila formulasi hubungan antara variabel prediktor X dan variabel respon Y tidak diketahui maka alternatif lain yang dapat digunakan adalah dengan pendekatan nonparametrik [9]. Estimasi fungsi regresi nonparametrik dilakukan berdasarkan data pengamatan. Data akan mencari bentuk estimasinya sendiri tanpa dipengaruhi oleh subjektivitas dari peneliti, sehingga pendekatan regresi nonparametrik memiliki fleksibilitas yang tinggi [2]**.** Agar pendekatan nonparametrik ini menghasilkan estimasi m yang masuk akal, maka hal yang harus diperhatikan adalah asumsi bahwa m memiliki derajat kemulusan (*smoothing*), kurva regresi diasumsikan termuat dalam suatu ruang fungsi mulus yang berdimensi tak hingga yang mempunyai turunan yang kontinu atau dapat diintegralkan secara kuadrat. Pemilihan ruang fungsi tersebut biasanya dimotivasi oleh sifat kemulusan (*smoothness*) yang diasumsikan dimiliki oleh fungsi regresi. Ada beberapa teknik *smoothing* dalam regresi nonparametrik antara lain histogram, estimator spline, estimator kernel, deret fourier, deret orthogonal, k-NN, dan estimator wavelet [4]. Diantara metode-metode pendekaan nonparametrik tersebut, estimator kernel merupakan metode yang akan digunakan dalam penelitian ini. Estimasi dengan pendekatan kernel tergantung pada dua parameter yaitu bandwidth dan fungsi kernel. *Cross-validation* merupakan contoh metode yang dapat digunakan untuk mendapatkan model kurva regresi terbaik.

Berdasarkan uraian diatas pada penelitian ini akan dibahas mengenai teori dan pembuktian tentang *cross validation*. Selain itu, diberikan perbandingan hasil estimasi fungsi regresi menggunakan pendekatan parametrik, dan pendekatan nonparametrik. Serta bagaimana menggunakan regresi Robust apabila terdapat *outlier*.

2 Metode Penelitian

Penelitian tentang estimasi fungsi regresi nonparametrik dengan menggunakan metode Robust dan *cross-validation* ini dilakukan dengan studi kepustakaan. Penelitian ini dimulai dengan mencari dan menentukan jurnal yang akan dijadikan bahan acuan, mengumpulkan jurnal jurnal lain yang relevan dengan materi dalam jurnal acuan, mempelajari buku-buku pendukung yang berkaitan dengan topik permasalahan penelitian yaitu teori regresi nonparametrik, fungsi kernel dan sifat-sifatnya, estimator Nadaraya-Watson, estimasi fungsi dalam regresi nonparametrik, pemilihan *bandwidth*, metode *cross validation*, outlier, metode robust, regresi nonparametrik Robust. Bagian terakhir dari penelitian ini adalah studi kasus menggunakan software R. Data pada penelitian ini diperoleh dengan mengedarkan angket kepada alumni. Kuestioner/angket adalah teknik pengumpulan data yang dilakukan dengan cara memberi seperangkat pertanyaan atau pernyataan tertulis kepada responden untuk dijawabnya [5][7]. Kuestioner dapat berupa pertanyaan/pernyataan tertutup atau terbuka dapat diberikan kepada responden secara langsung atau dikirim melalui pos, atau internet.

3 Pembahasan

Metode *cross-validation* adalah metode untuk menyeleksi model berdasarkan kemampuan prediksi untuk memilih satu dari satu kelas model. Susunan data dipisahkan menjadi dua bagian. Untuk bagian pertama terdiri dari titik-titik data n_c yang digunakan untuk menyesuaikan satu model (mengkonstruksi model), sebaliknya bagian kedua terdiri dari $n_v = n - n_c$ yang tersedia untuk menilai kemampuan prediksi model (validasi model). Secara tegas dikatakan, validasi model diselesaikan tidak hanya dengan menggunakan n_v, tapi semua data $n = n_v + n_c$. Ada perbedaan $\binom{n}{n_v}$ cara untuk memisahkan susunan data. *Cross-*

validation, sesuai dengan namanya, memilih model dengan kemampuan prediksi rata-rata terbaik yang dihitung berdasarkan semua atau beberapa cara yang berbeda dari pemisahan data.

Dalam pemilihan bandwidth diperlukan suatu kriteria ukuran untuk menentukan bandwidth yang optimal, salah satu kriteria yang digunakan adalah *Mean Squared Error* (MSE) [1][6], sedangkan rata-rata integral dari MSE disebut *Mean Integrated Squared Error* (MISE). Dalam [8] dijelaskan

$$\begin{aligned} MISE[\widehat{m}(x), f(x)] &= \int_{-\infty}^{\infty} MSE\,[\widehat{m}(x) f(x)]dx \\ &= \int_{-\infty}^{\infty} E\,[(\widehat{m}(x) - m(x))^2] f(x)dx \\ &= E[\int_{-\infty}^{\infty} (\widehat{m}(x) - m(x))^2] f(x)dx \end{aligned}$$

dan rata-rata MSE adalah:

$$MISE[\widehat{m}(x)] = \frac{E[\sum_{i=1}^{n}[m(x_i - \widehat{m}(x))]^2]}{n} \quad (1)$$

MISE $[\widehat{m}(x)]$ mengandung $m(x)$ dimana $m(x)$ merupakan suatu fungsi tidak diketahui. Fungsi ini tidak bisa diestimasi dari data. Jadi, untuk menghitung rata-rata MISE $[\widehat{m}(x)]$ dibutuhkan suatu metode. Untuk itu suatu metode yang dapat digunakan untuk mengestimasi $m(x)$ menggunakan *Average Predictive Squared Error* (APSE) atau juga sering disebut *Prediction Mean Squared Error* (PMSE), didefinisikan sebagai berikut:

$$APSE[\widehat{m}(x_i)] = \frac{E[\sum_{i=1}^{n}(m(x_i) + \epsilon_i^* - \widehat{m}(x_i))^2]}{n} \quad (2)$$

dengan $y_i^*(1 \leq i < n$ nilai dari prediktor untuk data $(x_i\ y_i^*)$. Dengan nilai prediktor $(x_i\ y_i^*)$ sama dengan (x_i, y_i), maka y_i^* didefinisikan sebagai:

$$y_i^* = m(x_i) + \epsilon_i^* \quad (3)$$

dengan ϵ_i^* independen terhadap ϵ_i tetapi memiliki sifat-sifat yang sama. Dengan mensubstitusikan Persamaan (3) ke Persamaan (2) maka:

$$APSE[\widehat{m}(x)] = \frac{E[\sum_{i=1}^{n}(m(x_i) + \epsilon_i^* - \widehat{m}(x_i))^2]}{n}$$

$$\begin{aligned} &= \frac{E[\sum_{i=1}^{n}(m(x_i) - \widehat{m}(x_i))^2] + 2E[\sum_{i=1}^{n}(\epsilon_i^*(m(x_i) - \widehat{m}(x_i)))] + E[\sum_{i=1}^{n}(\epsilon_i^*)]}{n} \\ &= MISE'[\widehat{m}(x_i)] + 2E\left[\epsilon_i^*\left(m(x_i) - \widehat{m}(x_i)\right)\right] + E[(\epsilon_i^*)] \\ &= MISE'[\widehat{m}(x_i)] + 0 + \sigma^2 \\ &= MISE'[\widehat{m}(x_i)] + \sigma^2 \end{aligned} \quad (4)$$

dengan $\sigma^2 = E(\epsilon_i^*)$ adalah variansi ϵ_i^* sama seperti ϵ_i. Persamaan kedua pada baris ketiga terpenuhi dengan asumsi $E[(\epsilon_i^*)] = 0$ terpenuhi. Persamaan pada baris terakhir menunjukkan penurunan $m(x)$ dengan meminimumkan $MISE'[\widehat{m}(x_i)]$ ekuivalen dengan meminimumkan APSE$[\widehat{m}(x_i)]$. Dengan begitu, diperlukan suatu metode untuk mengestimasi prediksi *error* tersebut. Suatu metode yang dapat digunakan untuk mengestimasi APSE $[\widehat{m}(x_i)]$ adalah metode *cross-validation*.

Cross-validation adalah sebuah metode untuk mengevaluasi model dengan beberapa tahapan dan juga dengan beberapa cara. Ide dasar dari *cross-validation* adalah menemukan parameter penghalus yang merupakan estimator terbaik untuk mengestimasi data didasarkan pada observasi baru, serta mampu meberikan indikasi-indikasi seberapa baik yang akan dapat dilakukan untuk menentukan prediksi selanjutnya atau jika ada data terbaru yang diperoleh, sementara data tersebut belum pernah ada sebelumnya. Dan selanjutnya akan dipilih model terbaik. Didefinisikan

$$CV[\widehat{m}(x)] = \frac{\sum_{k=1}^{n}(y_k - \widehat{m}^{-k}(x_k))^2}{n} \tag{5}$$

Salah satu teknik *cross-validation* yang diketahui adalah *leave one out cross-validation* (LOOCV) yang berarti bahwa meninggalkan satu untuk valiasi silang, yaitu dengan melibatkan sampel pengamatan tunggal dari sampel asli digunakan sebagai validasi data, dan sampel pengamatan yang tersisa digunakan sebagai data pelatihan. Hal ini dilakukan berulang pada setiap observasi dalam sampel yang digunakan sekaligus sebagai validasi data. Persamaan (5), $\widehat{m}^{-k}(x)$ fungsi regresi yang diperoleh dengan menghilangkan satu data (x_k, y_k) dari (x_i, y_i). Dengan demikian diperoleh:

$$E[CV(\widehat{m}(x))] \approx APSE[m(x)] \tag{6}$$

hubungan tersebut diperoleh dari:

$$\begin{aligned} E[y_k - \widehat{m}(x_k)^2] &= E[((y_k - \widehat{m}(x_k) - m(x_k) - \widehat{m}^{-k}(x_k))^2] \\ &= E[(y_k - \widehat{m}(x_k))^2] - 2E[\left(y_k - \widehat{m}(x_k)\right)(m(x_k) - \widehat{m}^{-k}(x_k)] + \\ &\quad E[((m(x_k) - \widehat{m}^{-k}(x_k))^2] \\ &= E[(y_k - m(x_k))^2] - 2E[(\epsilon_k \cdot m(x_k) — \widehat{m}^{-k}(x_k))] + E[((m(x_k) \\ &\qquad - \widehat{m}^{-k}(x_k))] \\ &= \sigma^2 + 0 + E[((m(x_k) - \widehat{m}^{-k}(x_k))] \\ &= \sigma^2 + MSE[\widehat{m}^{-k}(x_k)] \end{aligned} \tag{7}$$

Untuk lebih mudah mencari $CV[\widehat{m}(x)]$ pada Persamaan (3) estimator untuk $m(x)$ ditulis

$$\widehat{m}(x_i) = \sum_{i=1}^{n}[H_{ij}]\, y_j \tag{8}$$

dengan H (matriks ukuran $n \times n$) adalah hat matriks. Jika fungsi regresi dinyatakan seperti Persamaan (8), maka Persamaan (3) dapat direpresentasikan dengan elemen diagonal ([Hii]).

$$CV[\widehat{m}(x)] = \sum_{i=1}^{n}\frac{(y_i - \widehat{m}(x_i))^2}{n \cdot (1 - [H_{ii}])^2} \tag{9}$$

Jadi, $\widehat{m}^{-k}(x_k)$ merupakan jenis lain $\widehat{m}(x)$. Hubungan antara $\widehat{m}^{-k}(x_k)$ dan $\widehat{m}(x)$ dapat ditunjukkan sebagai berikut dengan memodifikasi $\widehat{m}^{k}(x_k) = \sum_{i=1, i\neq k}^{n}\frac{[H_{ki}]y_i}{1-[H_{kk}]}$ kemudian disubstitusikan ke Persamaan (5), maka:

$$\begin{aligned} CV[\widehat{m}(x)] &= \sum_{i=1}^{n}\frac{(y_i - \widehat{m}(x_i))^2}{n \cdot (1 - [H_{ii}])^2} \\ &= \frac{1}{n}\sum_{i=1}^{n}(y_k - \frac{\sum_{i=1,i\neq k}^{n} H_{ki}\, y_i}{1 - [H_{kk}]})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(\frac{y_k(1 - H_{kk}) - \sum_{i=1,i\neq k}^{n} H_{ki}\, y_i}{1 - [H_{kk}]})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(\frac{y_k - \sum_{i=1,}^{n} H_{ki}\, y_i}{1 - [H_{kk}]})^2 \end{aligned}$$

$$= \sum_{i=1}^{n} \frac{(y_k - \hat{m}(x_k))^2}{n \cdot (1 - [H_{kk}])^2}$$

Jadi, $\hat{m}(x)$ dan $\hat{m}^k(x_k)$ merupakan estimator untuk $m(x)$, hanya saja bedanya adalah $\hat{m}^k(x_k)$ menghilangkan data ke (xk; yk) pada $\hat{m}(x)$.

3.1 Pendekatan Robust pada Regresi Nonparametrik

Perhatikan persamaan regresi nonparametrik

$$y_i \quad = m(x_i) + \epsilon_i \quad i = 1, \dots, n \tag{10}$$

dengan fungsi regresi $m(.)$ dapat menjadi $\hat{m}(.)$ dengan dengan cara meminimumkan jumlah fungsi-fungsi dari residu yang monoton naik. Dengan demikian penghalus robust $\hat{m}(x)$ dari $m(x)$ dapat didefinisikan sebagai:

$$\frac{1}{n}\sum_{i=1}^{n} \rho\{y_i - \hat{m}(x_i) = min$$

dengan ρ merupakan fungsi monoton naik. Turunan pertama ρ adalah $\psi(.)$, Persamaan di atas disamakan dengan nol menjadi:

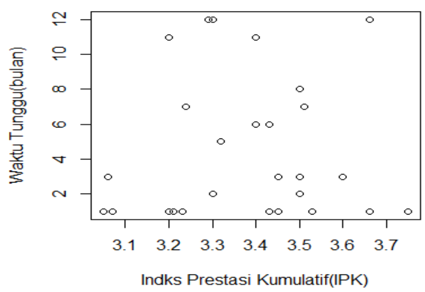
$$\sum_{i=1}^{n} \alpha_i(x)\rho\{y_i - \hat{m}(x_i)\} = min$$

dengan α_i adalah suatu fungsi bobot.

Definisi 1. Jika diberikan model regresi $y_i = m(x_i) + \varepsilon_i$, i *=1, 2, ...,n* maka $\sum_{i=1}^{n} \alpha_i(x)\rho\{y_i - \hat{m}(x_i)\} = min$ disebut kernel M-*smoother* $\hat{m}_h^M(x)$ untuk $m(.)$, dengan $\psi(.) = \rho^{'}$ [3]

4 Studi Kasus

Data yang digunaka pada penelitian ini adalah data alumni STIA Muhammadiyah program studi administrasi perkantoran angkatan 2016. Banyaknya alumni yang berhasil terdeteksi atau yang mengisi dan mengembalikan kuestioner adalah 30 orang dari ±100 orang. Berdasarkan data yang diperoleh terlihat bahwa 70% alumni merupakan mahasiswa non regular artinya mahasiswa tersebut rata-rata bekerja sambil kuliah. Sampel yang sesuai digunakan pada penelitian ini adalah data mahasiswa *fresh graduate* atau data mahasiswa regular. Oleh karena itu sampel yang digunakan adalah 30 orang. Untuk penelitian ini observasi yang dilakukan mengenai hubungan antara Indeks Prestasi Kumulatif (IPK) X dengan lama waktu tunggu mendapat pekerjaan (bulan) Y. Untuk lebih jelasnya berikut diberikan plot data.



Gambar 1. Data Indeks Prestasi Kumulatif dan Lama Waktu Tunggu Mendapat Pekerjaan Mahasiswa

4.1 Estimasi Fungsi Regresi dengan Menggunakan Pendekatan Parametrik

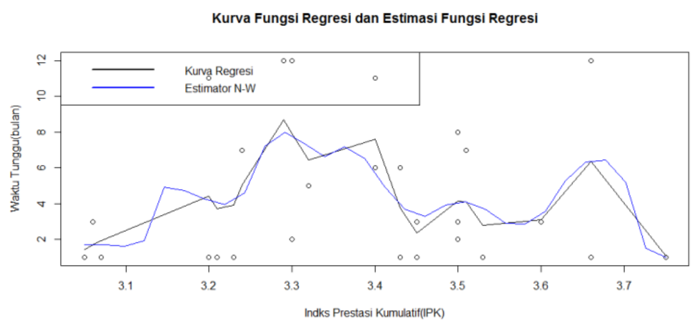
Berikut hasil estimasi model regresi dengan menggunakan metode kuadrat terkecil. Dengan menggunakan software R dipeorleh hasil model regresi dengan metode kuadrat terkecil yaitu:

$$y = 2{,}52 + 0{,}57x$$

dengan estimasi standar *error* sebesar 3,98.

4.2 Estimasi Fungsi Regresi dengan Menggunakan Pendekatan Nonparametrik

Estimasi fungsi regresi dengan pendekatan nonparametrik yaitu menggunakan estimator Nadaraya-Watson. Dari hasil pengolahan data dengan software R.3.0.2. Dengan *output* yang diperoleh sebagai berikut.



Gambar 2. Kurva Fungsi Regresi dan Estimasi Fungsi Regresi

Pada Gambar 2, kurva yang berwarna hitam merupakan kurva data yang akan diestimasi (kurva data sesungguhnya) dan kurva yang berwarna biru merupakan kurva regresi hasil estimasi dengan menggunakan estimator Nadaraya-Watson. Proses *smoothing* dilakukan dengan fungsi ”ksmooth”. Sebelum melakukan *smoothing* terlebih dahulu dipilih *bandwidth* optimal. Adapun bentuk estimasi model regresi nonparametrik dengan estimator Nadaraya-Watson menggunakan fungsi kernel Gaussian untuk data IPK terhadap lama waktu tunggu mendapatkan pekerjaan menggunakan *bandwidth* h diberikan oleh persamaan berikut:

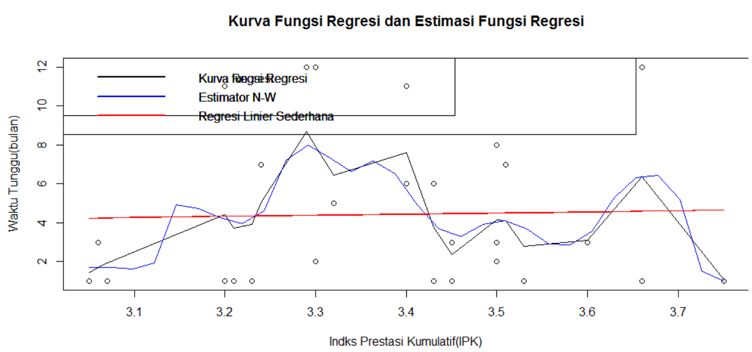
$$\hat{y} = \hat{m}(x_i) = \frac{\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x - X_i}{h}\right)^2\right) y_i}{\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x - X_i}{h}\right)^2\right)}$$

Sebagai contoh misalkan ingin diestimasi lama waktu tunggu seseorang mendapatkan pekerjaan jika IPK 3,06. Dengan menggunakan kernel Gaussian dan *bandwidth* sebesar 0,2 diperoleh:

$$\begin{aligned}
\hat{y} &= \hat{m}(x_i), i = 1, \ldots, 30 \\
&= \frac{\sum_{i=1}^{30} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x - X_i}{h}\right)^2\right) y_i}{\sum_{i=1}^{30} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x - X_i}{h}\right)^2\right)} \\
&= \frac{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{2.00 - 3.30}{0.2}\right)^2\right) + \cdots + \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{2.00 - 3.50}{0.2}\right)^2\right).4}{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{2.00 - 3.30}{0.2}\right)^2\right) + \cdots + \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{2.00 - 3.50}{0.2}\right)^2\right)} \\
&\approx 6
\end{aligned}$$

Jadi, lama waktu tunggu seseorang mendapatkan pekerjaan jika memiliki IPK 3.06 diprediksi sekitar 6 bulan. Kemudian estimasi untuk IPK sebesar 3,05, dan 3,20 diperoleh hasil estimasi lama waktu tunggu masing-masing adalah 3 bulan, dan 2 bulan. Dengan standar *error* 0,22. Hal ini menunjukkan IPK seseorang tidak memiliki hubungan dengan lamanya seseorang mendapatkan pekerjaan.

Untuk lebih jelas berikut diberikan kurva perbandingan hasil estimasi fungsi regresi yaitu dengan pendekatan parametrik menggunakan regresi linear sederhana dan pendekatan nonparametrik menggunakan regresi kernel dengan estimator Nadaraya-Watson.



Gambar 3. Kurva Perbandingan Fungsi Regresi

Berdasarkan Gambar 3, terlihat bahwa kurva yang berwarna hitam merupakan kurva dari data yang akan diestimasi. Hasil estimasi dengan regresi linear sederhana terlihat pada kurva yang berwarna merah.Terlihat bahwa hasil estimasi dengan pendekatan parametrik belum mendekati kurva dari data karena berupa garis lurus. Apabila diestimasi dengan pendekatan nonparametrik menghasilkan estimasi kurva regresi yang lebih tepat mendekati bentuk kurva datanya terlihat dari kurva yang berwarna biru. Adapun perbandingan standar error dari hasil estimasi di atas sebagai berikut:

Tabel 1. Perbandingan Standar *Error* Hasil Estimasi Fungsi Regresi

No	**Metode**	**Standar *error***
1	Regresi linier sederhana dengan metode kuadrat terkecil	3,98
2	Regresi nonparametrik kernel dengan estimator Nadaraya-Watson	0,22

6 Kesimpulan

Berdasarkan hasil yang diperoleh dapat disimpulkan bahwa :

1. Berdasarkan hasil *output* SPSS di atas hasil estimasi model regresi dengan pendekatan parametrik yaitu $y = 2,52 + 0,57x$ dengan estimasi standar *error* sebesar 3,98.
2. Estimasi hubungan antara Indeks Prestasi Kumulatif (IPK) dan lama waktu tunggu mendapat pekerjaan alumni STIA Muhammadiyah Selong program studi administrasi publik angkatan 2016 lebih cocok menggunakan regresi nonparametrik dengan kernel Gaussian dan bandwidth 0,2. Hal ini dapat dilihat dari SE yang dihasilkan sebesar 0,22.
3. Berdasarkan perbandingan ketiga kurva hasil estimasi terlihat bahwa pemodelan lama waktu tunggu mendapatkan pekerjaan pertama alumni STIA Muhammadiyah Selong program studi administrasi publik angkatan 2016 lebih mendekati jika menggunakan regresi nonparametrik kernel.

Daftar Pustaka

[1] Bain, L. J dan Engelhardt. 1991. *Introduction to Probability and Mathematical Statistics Second Edition*. Duxbury Press. USA.

[2] Eubank, R. L. 1988. *Spline Smoothing and Nonparametric Regression.* Marcell Dekker Inc. New York.

[3] Hardle, Wolfgang. 1990. *Smoothing Techniques With Implementation in S*. Springer-Verlag. New York.

[4] Hardle, W, Hall, P, dan Marron, J. 1992. Regression Smoothing Parameters that are not Far from Their Optimum. *J. Amer Statist. Assoc*. 87: 227-233.

[5] Putranto, R. T., dan Mashuri, M. 2012. *Analisis Statistik Tentang Faktor-Faktor yang Mempengaruhi Waktu Tunggu Kerja Fresh Graduate di Jurusan Statistika Institut Teknologi Sepuluh Nopember (ITS) dengan Menggunakan Regesi Logistik Ordinal*. *Jurnal Sains dan Seni ITS*. 1(1) :324 – 328.

[6] Roussas, George.G. 1997. *Acourse in Mathematical Statistics Second Edition.* Academic Press. USA.

[7] Sugiyono. 2007. *Metode Penelitian Kuantitati Kualitatif dan R and D*. Alfabeta. Bandung.

[8] Takezawa, Kunio. 2005. *Introduction to Nonparametric Regression.* Jhon Wiley. USA.

[9] Wand, M.P dan Jones, M.C. 1995. *Kernel Smoothing*. Chapman and Hall.London.