

ANALISIS TINGKAT PENGANGGURAN TERBUKA (TPT) TERHADAP INDIKATOR KEMISKINAN DI JAWA TIMUR MENGUNAKAN REGRESI KERNEL

Nailul Mutammimah¹, Abdullah Hamid², Anik Nur Kholifah³

¹ UIN Sunan Ampel Surabaya, mimanailul@gmail.com

² UIN Sunan Ampel Surabaya, doelhamid@uinsby.ac.id

³ Dinas Komunikasi dan Informasi Kabupaten Gresik, anik.kholif25@gmail.com

Abstrak. Analisis regresi merupakan suatu metode analisis data yang dimanfaatkan untuk suatu hubungan antara dua atau lebih variabel. tujuan dari analisis regresi ini adalah untuk menyelidiki suatu pola hubungan antara variabel prediktor dan variabel responden. Estimator Regresi Nonparametrik Kernel merupakan estimator analisis yang mirip dengan estimator pada regresi nonparametric yang lain, bedanya pada estimator kernel ini lebih dikhususkan dalam masalah penggunaan metode bandwidth-nya. Tujuan dari penelitian ini yaitu untuk mengetahui pengaruh Regresi Kernel menggunakan bandwidth optimal pada tingkat pengangguran terbuka terhadap indikator kemiskinan di daerah Jawa Timur. Pada penelitian ini akan digunakan metode regresi kernel untuk memodelkan pengaruh indikator kemiskinan. Berdasarkan pada hasil pembahasan, maka dapat diperoleh kesimpulan bahwa nilai pemilihan bandwidth optimal diperoleh sebesar 0,65. model Regresi Nonparametrik Kernel terbaik terdapat pada pemilihan bandwidth optimal. sehingga Regresi kernel menggunakan bandwidth optimal berpengaruh pada tingkat pengangguran terbuka terhadap indikator kemiskinan di Jawa Timur

Kata kunci: Regresi Kernel, Kemiskinan, Tingkat Pengangguran Terbuka

Abstract. Regression analysis is a data analysis method that is used for a relationship between two or more variables. The purpose of this regression analysis is to investigate a pattern of relationship between predictor variables and respondent variables. Kernel Nonparametric Regression Estimator is an analytical estimator that is similar to the estimator in other nonparametric regressions, the difference is that this kernel estimator is more specialized in the problem of using its bandwidth method. The purpose of this study is to determine the effect of the open unemployment rate on poverty indicators in East Java. In this study, the kernel regression method will be used to model the effect of poverty indicators. Based on the results of the discussion, it can be concluded that the optimal bandwidth selection value is 0.65. The best nonparametric kernel regression model, the most important thing is the selection of the optimal bandwidth value, not in the selection of the kernel function, because the estimation results for the Gaussian kernel function using the optimal bandwidth are said to be very close together so it can be said that the use of the kernel function with the optimal bandwidth will result in a regression curve the same one.

Keywords: Regression, nonparametric kernel, Optimal Bandwidth.

1. Pendahuluan

Analisis regresi suatu metode analisis data yang dimanfaatkan untuk suatu hubungan antara dua atau lebih variabel. tujuan dari analisis regresi secara umum adalah untuk menyelidiki suatu pola hubungan antara variabel prediktor dan variabel responden. Untuk melakukannya dapat dilakukan dengan membuat diagram pencar, Mengestimasi nilai pada variabel respon berdasarkan nilai variabel prediktor yang dimiliki, dan Menyelidiki variabel prediktor yang mana saja yang berpengaruh secara signifikan terhadap variabel respon [1]. Terdapat dua pendekatan dalam mengestimasi fungsi regresi yaitu pendekatan parametrik dan pendekatan nonparametrik. Pendekatan parametrik adalah pendekatan yang digunakan untuk menggambarkan bentuk hubungan antara variable respons dan variabel prediktor yang telah diketahui atau bentuk dari kurva regresinya dapat diasumsikan mengikuti pola-pola tertentu. Contoh pendekatan parametrik yaitu linier, kuadratik, kubik, polinomial derajat p , eksponen dan lainnya. Pendekatan *non parametrik* merupakan pendekatan yang digunakan untuk menggambarkan bentuk suatu hubungan antara variable prediktor dan variabel responnya yang belum diketahui dan belum ada informasi atau keterikatan bentuk fungsi regresi [2]. Regresi Parametrik dan Nonparametrik juga mempunyai perbedaan yaitu dalam pendekatan parametrik data yang cenderung dipaksa untuk mengikuti suatu pola tertentu, sedangkan pada pendekatan regresi nonparametrik data akan diberi keleluasaan untuk mencari pola kurva regresinya sendiri sehingga sangat fleksibel dan obyektif. Beberapa model regresi nonparametrik yang banyak digunakan antara lain Kernel, Spline, Polinomial Lokal, Deret Ortogonal, Deret Fourier, kNN, Neural Network (NN), Wavelets dan lain-lain. Semua model regresi memiliki kelebihan dan kekurangan serta motivasi sendiri dalam memodelkan suatu pola [3]. Estimator Regresi Nonparametrik Kernel merupakan estimator analisis yang mirip dengan estimator pada regresi nonparametric yang lain, bedanya pada estimator kernel ini lebih dikhususkan dalam masalah penggunaan metode bandwidthnya [4].

Data penelitian ini akan diambil dari faktor-faktor yang mempengaruhi indikator kemiskinan di provinsi Jawa Timur dengan menggunakan analisis regresi Kernel dengan variabel dependen yaitu indikator kemiskinan (Y) dan variabel independennya yaitu Tingkat Pengangguran Terbuka (X). Tujuan dari penelitian ini yaitu untuk mengetahui pengaruh Regresi Kernel menggunakan bandwidth optimal pada tingkat pengangguran terbuka terhadap indikator kemiskinan di daerah Jawa Timur.

Pada penelitian sebelumnya pada kasus strategi nasional dalam menanggulangi tingkat kemiskinan, dihasilkan bahwa dalam tingkat kemiskinan dapat dipengaruhi oleh beberapa faktor antara lain : tingkat pendapatan, pendidikan, tingkat kesehatan penduduk, akses penggunaan barang atau jasa, kondisi lingkungan serta pengamatan lokasi geografis [5].

2. Kajian Teori

2.1 Analisis Regresi Parametrik

Analisis regresi adalah alat statistika yang digunakan sebagai alat untuk melihat suatu hubungan antara variabel prediktor dengan satu atau lebih variabel respons. Sebagai contoh Misalkan terdapat data yang saling berpasangan (x_i, y_i) , maka hubungan dari

kedua variabel x_i dan variable y_i dapat dituliskan sebagai berikut [6]:

$$y_i = f(x_i) + \varepsilon_i ; i = 1,2,3, \dots, n \quad (1)$$

dengan diketahui y_i sebagai variabel respons ke- i , $f(x_i)$ merupakan fungsi regresi atau kurva regresi, dan ε_i merupakan sisaan dari asumsi *independent* dengan nilai tengah nol dan variansi σ^2 .

2.2 Analisis Regresi Nonparametrik

Analisis Regresi Nonparametrik merupakan salah satu metode statistika yang digunakan untuk membentuk atau mengetahui suatu pola dalam hubungan antara variabel prediktor dengan variable responden apabila sebelumnya belum diketahui bentuk kurva regresinya. Bentuk suatu factor dari regresi nonparametric dapat berbentuk apa saja, baik berupa linear maupun nonlinear. Misal diketahui bahwa variabel respons dalam pengamatan dapat disimbolkan sebagai y dan variabel prediktor disimbolkan sebagai x , dan untuk n adalah banyaknya pengamatan. Berikut adalah model umum dari analisis regresi nonparametrik [7]:

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

Dengan diketahui bahwa y_i merupakan respons dan variabel (x_i) merupakan variabel predictor. $f(x_i)$ dapat dikatakan sebagai fungsi yang belum atau tidak diketahui bentuk regresinya, dan untuk ε_i merupakan variabel sisa yang dapat diasumsikan sebagai variabel bebas dengan nilai tengahnya adalah nol dan variansi σ^2 .

2.3 Estimator Densitas Kernel

Dalam regresi nonparametric untuk mengestimasi nilai m dapat menggunakan regresi kernel sebagai fungsi suatu bobot. Diberikan n sampel random $X_i = 1,2,3,4, \dots, n$. Maka pada karakteristik dasar yang menggambarkan suatu sifat dari suatu variabel acak merupakan fungsi densitas f dari variabel acak tersebut.

Berdasarkan pada sampel acak tersebut akan diestimasi pada fungsi f yang belum diketahui dengan menggunakan pendekatan kernel. Regresi Kernel K dapat didefinisikan sebagai berikut [8].

$$K_h(x) = h^{-1} K\left(\frac{x}{h}\right) \quad (3)$$

Dengan K menyatakan fungsi kernel dan h adalah *bandwidth*.

Model regresi nonparametric kernel ini akan menggunakan penduga Nadaraya-Watson dengan diasumsikan bahwa peubah prediktor dan respon merupakan peubah acak. Misal $f(x)$ merupakan fungsi peluang peubah acak X , $f(y)$ merupakan fungsi peluang peubah acak Y , maka dinyatakan bahwa $f(x, y)$ merupakan fungsi peluang dari peubah acak (X, Y) . Jadi diketahui bahwa [9]:

$$m(x) = \frac{1}{f_x(x)} \int_{-\infty}^{\infty} y f(x, y) dy \quad (4)$$

$$\hat{f}_x(x) = \frac{1}{nh_x h_y} \sum_{i=1}^n K_x\left(\frac{x-X_i}{h_x}\right) K_y\left(\frac{y-Y_i}{h_y}\right) \quad (5)$$

Dimana $K_x\left(\frac{x-X_i}{h_x}\right)$ dan $K_y\left(\frac{y-Y_i}{h_y}\right)$ merupakan suatu fungsi dari kernel. Dan diperoleh penduga Nadaraya-Watson sebagai berikut [10].

$$\hat{m}(x) = \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h_x}\right) Y_i}{\sum_{i=1}^n K_i\left(\frac{x-X_i}{h_x}\right)} \quad (6)$$

$$\hat{m}(x) = \sum_{i=1}^n W_i(x)Y_i \quad (7)$$

2.4 Pemilihan Bandwidth Optimal

Pada pemilihan Bandwidth optimal pada fungsi regresi kernel sangat diperlukan sebagai fungsi penduga. Fungsi penduga ini akan undersmooth jika h mendekati nilai nol. Sebaliknya, fungsi penduga akan dikatakan oversmooth jika nilai h besar. Pada analisis data menggunakan regresi kernel akan memungkinkan menggunakan beberapa bandwidth (h) optimal dan memilih penduga yang akan digunakan dengan berdasarkan kriteria nilai MSE sebagai berikut [9].

$$MSE(\hat{m}) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{m}(x_i))^2 \quad (8)$$

Namun, pada penggunaan fungsi di atas bukan merupakan metode yang tepat dalam mengevaluasi model, karena semakin kecil nilai MSE maka bandwidth juga akan semakin kecil. Dalam hal ini dapat dilakukan pemulusan yang undersmooth yaitu dengan pendekatan selain MSE, dalam dugaan pada regresi nonparametric kernel dalam memilih bandwidth optimal adalah menggunakan *cross validation* (CV). CV merupakan metode penggunaan data yang menunjukkan apa yg harus dilakukan jika terdapat pengulangan pengamatan yang tersedia [11].

$$\hat{m}_{h,j}(x_j) = n^{-i} \sum_{i \neq j} W_{hi}(x_j)Y_i \quad (9)$$

Dengan diulang pada langkah pertama untuk $j = 1, 2, 3, \dots, n$, maka didapatkan fungsi :

$$cv(h) = n^{-i} \sum_{i=1}^n [Y_j - \hat{m}_{h,j}(x_j)]^2 \quad (10)$$

3. Metode Penelitian

Pada penelitian ini akan menggunakan data dari data sekunder pada tahun 2018 berupa data hasil Survei Sosial Ekonomi Nasional (Susenas) dan Survei Ketenagakerjaan Nasional yang telah diperoleh dari sumber data publikasi Badan Pusat Statistik (BPS) [12]. Variabel yang akan digunakan dalam penelitian ini terdiri dari variabel predictor dan variable respons. Pada variabel respons digunakan indikator kemiskinan yaitu indeks kedalaman kemiskinan (Y) dan variabel prediktornya adalah Tingkat Pengangguran Terbuka (X) pada tahun 2019.

Data yang diperoleh akan dianalisis dengan menggunakan metode analisis regresi nonparametric kernel dengan perhitungan menggunakan *Software* SPSS dan R Studio.

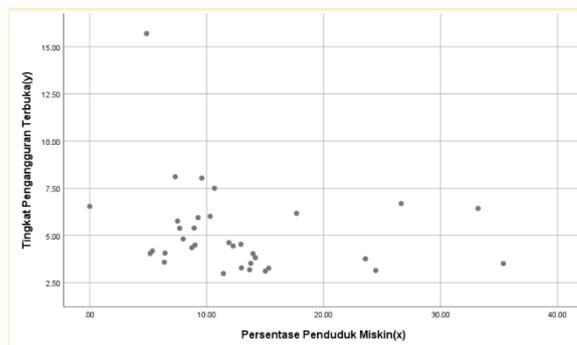
Pada penelitian ini dibutuhkan langkah-langkah sebagai berikut :

1. Membuat scatter plot antara variabel predictor dengan variable respons.
2. Menghitung analisis korelasi antara variabel prediktor dengan variable respons.
3. Menentukan bandwidth pengaruh terhadap kehalusan grafik yang diperoleh.
4. Menentukan Uji Parsial untuk mengetahui variabel respons yang berpengaruh terhadap variable predictor.
5. Menentukan estimasi dari variable predictor dengan masing-masing variabel respons.
6. Menentukan kurva dari data aktual dan data hasil estimasi.

4. Hasil dan Pembahasan

4.1 Pembuatan Scatter Plot antara variabel prediktor dan variabel respons.

Langkah pertama akan dilakukan proses scatter plot untuk mengetahui suatu hubungan antara variabel prediktor dan variabel respons. Berikut diagram scatter plot.



Gambar 1. Plot Kernel

Dilihat pada gambar 1. Dapat diketahui bahwa kedua variabel tidak membentuk pola tertentu, sehingga dalam hal ini akan lebih baik jika menggunakan metode regresi nonparametric karena dapat mengestimasi data yang tidak mempunyai pola tertentu dan cenderung mencari estimasi sendiri dari pola yang terbentuk.

4.2 Analisis korelasi

Dalam menghitung analisis korelasi ini akan dilakukan dengan perhitungan menggunakan Excel. Dan dari perhitungan diperoleh hasil analisis korelasi antara variabel x dan y adalah bernilai -0,065. Dan sedangkan pada nilai r tabel diperoleh nilai 0,278. Dikarenakan korelasi bernilai negatif(-) maka ditunjukkan bahwa hubungan antara kedua variabel bernilai negatif atau dengan kata lain setiap kali nilai pada variabel x maka nilai pada variabel y akan berkurang.

4.3 Pemilihan Bandwidth Optimal

Pada setiap pemilihan bandwidth akan berpengaruh terhadap kemulusan suatu grafik yang akan diperoleh. Dalam penelitian ini dalam menentukan bandwidth akan menggunakan CV. Dengan menggunakan *Software R-Studio* didapatkan hasil nilai bandwidth (h) yang optimal adalah sebesar 0,65.

4.4 Uji Parsial

Dalam Uji Parsial digunakan untuk mengetahui pengaruh di antara kedua variabel, yaitu variabel prediktor (y) dan variabel respons (x). dalam hal ini akan menggunakan software SPSS. Dan didapatkan hasil sebagai berikut.

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized	t	Sig.
		B	Std. Error	Coefficients		
1	(Constant)	5.996	.777		7.719	.000
	Persentase Penduduk Miskin(x)	-.068	.052	-.225	-1.306	.201

a. Dependent Variable: Tingkat Pengangguran Terbuka(y)

Gambar 2. Uji Parsial

Pada gambar 2 diperoleh bahwa Nilai t hitung adalah sebesar -1,306, sedangkan pada t table diperoleh nilai sebesar 1,690. Dan diperoleh nilai signifikan $0,201 > 0,05$, maka dapat disimpulkan bahwa tidak terdapat pengaruh antara variabel x terhadap y secara parsial.

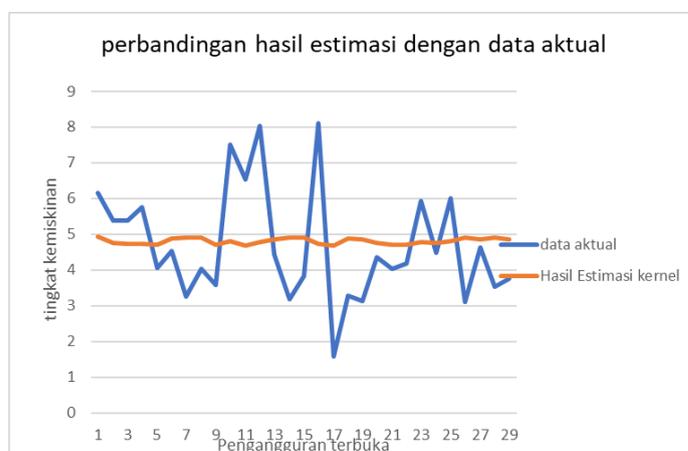
Estimasi Kernel

Pada penentuan ini dilakukan dengan menggunakan h senilai 0,65 dan didapatkan hasil estimasi sebagai berikut.

provinsi	Hasil Estimasi kernel
Pacitan	4,93
ponorogo	4,77
trenggalek	4,74
tulungagung	4,73
blitar	4,71
kediri	4,88
malang	4,92
lumajang	4,91
jember	4,71
banyuwangi	4,82
bondowoso	4,67
situbondo	4,78
probolinggo	4,87
paasuruan	4,90
sidoarjo	4,91
mojokerto	4,73
jombang	4,69
nganjuk	4,89
madiun	4,85
magetan	4,76
ngawi	4,70
bojonegoro	4,70
tuban	4,78
lamongan	4,77
gresik	4,81
bangkalan	4,92
sampang	4,86
pamekasan	4,90
sumenep	4,86

Gambar 3. Hasil Estimasi Kernel

Dan berikut merupakan perbandingan data aktual dan data nilai estimasi.



Gambar 4. Perbandingan data aktual dan estimasi

Berdasarkan hasil estimasi untuk fungsi kernel Gaussian dengan menggunakan bandwidth optimal, dikatakan sangat berhimpitan sehingga dapat dikatakan bahwa penggunaan fungsi kernel dengan bandwidth yang optimal akan menghasilkan kurva regresi yang sama.

5. Simpulan

Berdasarkan pada hasil pembahasan, maka dapat diperoleh kesimpulan bahwa nilai pemilihan bandwidth optimal diperoleh sebesar 0,65. bahwa nilai pemilihan bandwidth optimal diperoleh sebesar 0,65. model Regresi Nonparametrik Kernel terbaik terdapat pada pemilihan bandwidth optimal. sehingga Regresi kernel menggunakan bandwidth optimal berpengaruh pada tingkat pengangguran terbuka terhadap indikator kemiskinan di Jawa Timur.

Referensi

- [1] Astuti, Y. S., & Tresnawan, D. "Implementasi Cox Proportional Hazard Model Parametrik Pada Analisis Survival (Studi Kasus: Mahasiswa Universitas Internasional Batam)", *Unisda Journal of Mathematics and Computer Science (UJMC)*, Vol. 3, no. 1, 2017, 29-38.
- [2] B.P. Made, A.M.S. I Gusti, and S. I Wayan, "Pemodelan Regresi Spline (Studi Kasus: Herpindo Jaya Cabang Ngaliyan)", *E-Jurnal Matematika*. Vol. 4, no. 3, Agustus 2015, pp. 110-114.
- [3] Eubank, R.L. "Nonparametrik Regression and Spline Smoothing". Marcel Dekker, Inc. New York, 1999
- [4] Sudrajat, J., Rizki, S. W., & Perdana, H., "Perbandingan Model Regresi Parametrik Eksponensial dan Weibull Pada Data Survival Tersensor Interval.", *BIMASTER*, Vol. 7, no. 3, 2018.
- [5] Pratiwi DA, Budiantara IB, Wibowo W. Pendekatan Regresi Semiparametrik Spline Untuk Memodelkan Rata-Rata Umur Kawin Pertama (UKP) di Jawa Timur, *Jurnal Sains dan Seni ITS*. 2017; 6(1):129-136.
- [6] Sa'yidah, Y. H. dan Arianti, F. "Analisis Kemiskinan Rumah Tangga melalui Faktor-Faktor yang Mempengaruhinya di kecamatan Tugu Kota Semarang", *Diponegoro Journal Of Economics*. Vol.1, No. 1, pp.1-11, 2012.
- [7] Rohmaniah, S. A., & Danardono, D. "Pemodelan Regresi Cox dan Regresi Weibull Waktu Sembuh Diare Pada Balita.", *Unisda Journal of Mathematics and Computer Science (UJMC)*, Vol.2, no.1, 50–55, 2016.
- [8] Intan PK. " Comparison of Kernel Function on Support Vector Machine in Classification of Childbirth" , *Jurnal Matematika MANTIK*. Vol 5, no. 2, 2019, 2527-3159.
- [9] Esty. " Regresi Kernel Dengan Metode Nadaraya Watson", Skripsi. Universitas Negeri Yogyakarta, 2014
- [10] H. Toto, "Aplikasi Bootstrap Pada Analisis Regresi untuk Data Kecelakaan Kerja", *Academy of Education Journal*, Vol. 10 No. 1 Tahun 2019.
- [11] Takezawa, K. "Introduction to Nonparametric Regression", New Jersey: John Wiley & Sons, Inc.2006.
- [12] BPS (2018). Diakses dari www.jatim.bps.go.id