

Nonparametric Regression with Local Polynomial Kernel on Relationship Between Schooling Years and Unemployment Rate in Banten

Bunga Miftahul Barokah, Fadhilah Fitri*, Chairina Wirdiastuti

Departemen Statistika, Universitas Negeri Padang, Kota Padang, Indonesia

*Corresponding author: fadhilahfitri@fmipa.unp.ac.id

Submitted : 29 Mei 2025

Revised : 20 Juni 2025

Accepted : 06 Agustus 2025

ABSTRAK

The Open Unemployment Rate (OUR) is a key indicator in assessing the economic performance of Banten Province. One of the factors believed to influence OUR is education, which is measured by the average number of years of schooling. This study aims to analyze the relationship between the average years of schooling and TPT using the Local Polynomial Kernel Nonparametric Regression method for the period 2017–2024. This method was chosen for its flexibility in modeling nonlinear relationships without requiring strict assumptions about the data. The optimal bandwidth parameter for smoothing was determined using the Direct Plug-In (DPI) method through the $dpill$ function in the R software. The results show that the nonparametric model has a coefficient of determination (R^2) of 0.2841, which is higher than that of the Ordinary Least Squares (OLS) linear regression model, which only reached 0.1710. This indicates that the nonparametric approach is better at capturing the complex relationship between education and unemployment. However, the low R^2 values in both models indicate the presence of other factors that influence the unemployment rate, such as economic conditions, labor market structure, and education policy. Therefore, increasing the average years of schooling alone may not be sufficient to significantly reduce the unemployment rate. More comprehensive policies are needed, such as job skill enhancement, vocational training, and economic strategies focused on job creation. The findings of this study are expected to provide useful insights for policymakers in formulating more effective strategies to address unemployment in Banten Province.

Keywords: Average Years of Schooling, Direct Plug-In (DPI), Local Polynomial Kernel, Nonparametric Regression, Open Unemployment Rate.



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Pengangguran adalah suatu permasalahan sentral dalam konteks ekonomi yang memiliki dampak signifikan terhadap kesejahteraan sosial dan kondisi ekonomi masyarakat. Tingkat pengangguran yang tinggi menunjukkan bahwa ada ketidakseimbangan antara jumlah lapangan kerja yang tersedia dan jumlah tenaga kerja yang ada. Di Provinsi Banten, tingkat pengangguran terbuka (TPT) tetap menjadi permasalahan yang signifikan karena memengaruhi stabilitas sosial dan ekonomi di wilayah tersebut (BPS Banten, 2024). Tingginya Tingkat Pengangguran Terbuka (TPT) berpotensi menurunkan produktivitas, meningkatkan angka kemiskinan, dan memperparah ketimpangan sosial dalam masyarakat.

Salah satu faktor penting yang diperkirakan dapat berpengaruh terhadap tingkat pengangguran adalah rata-rata lama sekolah yang digunakan sebagai indikator tingkat pendidikan. Dalam meningkatkan kualitas sumber daya manusia dan menciptakan peluang pekerjaan yang layak maka pendidikan memiliki peran yang sangat penting terhadap itu (Ace, 2002). Berdasarkan teori modal manusia, peningkatan pendidikan diharapkan dapat meningkatkan keterampilan dan produktivitas individu, sehingga menurunkan angka pengangguran. Namun, hubungan antara pendidikan dan pengangguran tidak selalu bersifat linear. Beberapa penelitian menunjukkan bahwa peningkatan rata-rata lama sekolah tanpa diimbangi dengan penciptaan lapangan kerja yang memadai justru berpotensi meningkatkan angka pengangguran (Widiansyah, 2017).

Menurut data yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Banten untuk periode 2017 hingga 2024, rata-rata lama sekolah di Provinsi Banten berkisar antara 6,20 hingga 11,86 tahun dengan rata-rata 8,87 tahun. Tren data menunjukkan bahwa rata-rata lama sekolah meningkat dari tahun ke tahun. Di sisi lain, Tingkat Pengangguran Terbuka (TPT) berkisar antara 4,67% hingga 13,06% dengan rata-rata 8,50%. Visualisasi data memperlihatkan bahwa meskipun

Rata-rata lama sekolah mengalami tren peningkatan, terdapat fluktuasi dengan pola yang tidak konsisten yang ditunjukkan oleh Tingkat Pengangguran Terbuka (TPT). Tingkat Pengangguran Terbuka (TPT) yang tinggi masih ditemukan di beberapa wilayah dengan rata-rata lama sekolah yang tinggi. Hal ini menunjukkan bahwa hubungan antara pendidikan dan pengangguran mungkin bersifat non-linear, sehingga memerlukan pendekatan analisis yang lebih fleksibel.

Dalam analisis statistik, pendekatan parametrik sering digunakan untuk mengeksplorasi hubungan antara dua variabel. Namun, metode ini memiliki keterbatasan karena mengharuskan data memenuhi asumsi seperti linearitas dan distribusi normal. Sebagai alternatif, metode nonparametrik menawarkan fleksibilitas lebih tinggi karena tidak bergantung pada bentuk distribusi tertentu (Silverman, 1992). Salah satu metode yang banyak digunakan adalah regresi nonparametrik polinomial lokal, yang mampu mengidentifikasi pola hubungan non-linear secara lebih akurat. Pendekatan ini digunakan dalam penelitian ini untuk menganalisis hubungan antara Rata-Rata Lama Sekolah dan Tingkat Pengangguran Terbuka (TPT) di Provinsi Banten. Metode ini dinilai sesuai karena mampu menangkap karakteristik hubungan yang kompleks antarvariabel tanpa terikat pada asumsi linearitas.

Penelitian serupa telah dilakukan oleh Fitri & Almuhyar (2025), yang membandingkan regresi linear dan regresi polinomial lokal dalam memodelkan prevalensi stunting berdasarkan rata-rata lama sekolah di Indonesia. Hasil penelitian tersebut menunjukkan bahwa regresi polinomial lokal memiliki koefisien determinasi (R^2) sebesar 43,44%, lebih tinggi dibandingkan model regresi linear sebesar 32,94%. Temuan ini memperkuat keunggulan metode regresi nonparametrik dalam menangkap hubungan yang tidak linier antarvariabel. Penggunaan fungsi $dpill$ dalam penelitian tersebut juga berhasil menghasilkan bandwidth optimal yang efektif dalam menggambarkan pola hubungan yang kompleks.

Metode Nonparametrik Polinomial Lokal tidak hanya unggul dalam hal fleksibilitas model, tetapi juga mampu mendeteksi pola-pola lokal yang mungkin tidak terlihat dalam pendekatan parametrik. Dengan menggunakan metode ini, diharapkan analisis pola hubungan antara Rata-Rata Lama Sekolah dan Tingkat Pengangguran Terbuka (TPT) di Banten dapat dilakukan secara lebih mendalam dan untuk hasil analisisnya dapat membantu para pembuat kebijakan dalam membuat rencana yang berguna untuk meningkatkan kualitas pendidikan dan menciptakan lapangan kerja yang lebih baik di wilayah ini.

Analisis pada tingkat regional, seperti di Banten, sangat penting dilakukan mengingat perbedaan karakteristik sosial-ekonomi dibandingkan dengan wilayah lain di Indonesia (Beribe & Budyanra, 2023). Oleh karena itu, penelitian ini berfokus pada menganalisis hubungan antara Rata-Rata Lama Sekolah dan Tingkat Pengangguran Terbuka (TPT) di Banten yang dilakukan dengan menggunakan metode Nonparametrik Polinomial Lokal Kernel. Metode ini dipilih karena dapat mengidentifikasi pola hubungan yang kompleks tanpa terikat pada asumsi linearitas. Penelitian ini diharapkan dapat meningkatkan pemahaman tentang pengaruh pendidikan terhadap tingkat pengangguran dan menjadi dasar bagi pembuat kebijakan dalam merumuskan strategi untuk menurunkan Tingkat Pengangguran Terbuka (TPT) di Banten.

II. METODE PENELITIAN

A. Data Penelitian

Penelitian ini menggunakan data sekunder dari Badan Pusat Statistik (BPS) Provinsi Banten. Data ini terdiri dari data Rata-Rata Lama Sekolah Menurut Kabupaten/Kota di Provinsi Banten dari Tahun 2017-2024 dan data Tingkat Pengangguran Terbuka (TPT) Menurut Kabupaten/Kota di Provinsi Banten dari Tahun 2017-2024. Adapun variabel yang dianalisis lebih lanjut dalam penelitian ini adalah :

Tabel 1. Variabel Penelitian

Nama Variabel	Skala	Deskriptif
Rata-Rata Lama Sekolah (Y)	Numerik	Rata-Rata Lama Sekolah di Provinsi Banten (Tahun)
Variabel Prediktor		
Tingkat Pengangguran Terbuka (X)	Numerik	Tingkat Pengangguran Terbuka di Provinsi Banten (Persen)

B. Teknik Analisis Data

Dalam penelitian ini, menggunakan dua pendekatan utama dalam analisis data, yaitu Regresi Kuadrat Terkecil (OLS) dan Regresi Nonparametrik dengan Penduga Kernel Polinomial Lokal. Analisis data yang dilakukan menggunakan *software R*. Berikut langkah-langkah analisisnya dalam penelitian ini :

1) Visualisasi Data

Langkah awal untuk memahami pola hubungan antara variabel bebas dan variabel terikat disebut dengan visualisasi data. Grafik scatter plot digunakan untuk melihat distribusi data serta indikasi pola linear maupun non-linear (Kurniawan et al., 2023). Dalam penelitian ini, variabel Tingkat Pengangguran Terbuka (TPT) dijadikan sebagai prediktor (X), sedangkan Rata-rata Lama Sekolah sebagai respons (Y). Visualisasi awal dari penelitian membantu dalam menentukan model analisis yang sesuai berdasarkan pola data.

2) Model Regresi Kuadrat Terkecil (OLS)

Regresi Kuadrat Terkecil (OLS) merupakan pendekatan awal yang digunakan untuk memodelkan hubungan antara variabel X dan Y dengan asumsi linearitas. Persamaan regresi linear sederhana didefinisikan sebagai:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad (1)$$

Dengan:

- Y_i : variabel respon pengamatan ke-i
- X_i : variabel prediktor pengamatan ke-i
- β_0 : intersep (konstanta)
- β_i : Koefisien regresi
- ε_i : residual pengamatan ke-i

Metode OLS mencari estimasi parameter dengan meminimalkan jumlah kuadrat galat (Sum of Squared Errors/SSE):

$$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2)$$

Evaluasi model OLS dilakukan melalui koefisien determinasi (R^2) untuk mengetahui seberapa besar variabilitas Y yang dapat dijelaskan oleh X.

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (3)$$

Dengan:

- Y_i : Nilai aktual variabel dependen pada observasi ke-i
- $\hat{Y}_i$: Nilai prediksi variabel dependen pada observasi ke-i
- $\bar{Y}$: Nilai rata-rata variabel dependen
- n : Jumlah observasi

Model dikatakan baik jika nilai koefisien determinasi (R^2) mendekati 1, yang menunjukkan bahwa model mampu menjelaskan sebagian besar variabilitas data (Sari & Kurniawan, 2022).

3) Regresi Nonparametrik Kernel Polinomial Lokal

Pendekatan pemulusan menggunakan teknik kernel sangat dipengaruhi oleh pemilihan jenis fungsi kernel serta nilai bandwidth yang diterapkan (Silverman, 1992).

Definisi 1. X merupakan variabel acak dengan distribusi kontinu $F(x)$ dan memiliki densitas $f(x)$ yang mana merupakan turunan dari $F(x)$, yaitu $f(x) = \frac{d}{dx} F(x)$. Untuk fungsi $f(x)$ penduga densitas kernelnya adalah: (Silverman, 1992)

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \quad (4)$$

Dengan

- X_i : variabel prediktor ke-i.
- x : angka spesifik yang nilainya tetap
- n : banyak data
- h : bandwidth
- K : fungsi pembobot

Kita nyatakan fungsi pembobot kernel yaitu $K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$ maka persamaan 4 dapat disederhanakan, menjadi

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(X_i - x) \quad (5)$$

Kedua persamaan diatas, K merupakan fungsi kernel yang bersifat terbatas, bernilai *real*, fungsi kontinu dan memenuhi $\int_{-\infty}^{\infty} K(x)dx = 1$. Bilangan positif yang disebut bandwidth dinyatakan dengan h . Ketika $K(u)$ merupakan fungsi dari densitas kernel, maka $K_h(u)$ juga termasuk fungsi densitas kernel (Guidom, 2015).

4) Penduga Kernel Polinomial Lokal

Model polinomial lokal didasarkan pada konsep Deret Taylor, dan dinyatakan bahwa setiap fungsi kontinu dapat didekatkan secara lokal dengan menggunakan polinomial dari derajat tertentu. Misalkan x berada dalam suatu interval yang disebut wilayah lokal, yang mana bentuk kurva fungsi $m(x_i)$ belum kita ketahui. Untuk menduganya, digunakan metode penduga polinomial lokal dengan pendekatan kernel. Dalam konteks ini, Deret Taylor digunakan sebagai asumsi dasar dalam estimasi $m(x_i)$, ialah:

$$m(x_i) \approx m(x) + (x_i - x) m^{(1)}(x) + \frac{(x_i - x)^2 m^{(2)}(x)}{2!} + \dots + \frac{(x_i - x)^p m^{(p)}(x)}{p!} \quad (6)$$

Misalkan :

$$\beta_r(x) = \frac{m^{(r)}(x)}{r!}$$

yang mana $r = 1, 2, \dots, p$ maka persamaan (6) menjadi:

$$m(x_i) \approx \beta_0(x) + (x_i - x)\beta_1(x) + \dots + (x_i - x)^p \beta_p(x) \quad (7)$$

Persamaan (7) diatas dinyatakan dengan cara berikut :

$$m(x) = X\beta \quad (8)$$

dimana X adalah matriks yang berukuran $n \times (p + 1)$, dengan

$$X = \begin{bmatrix} 1 & (x_1 - x) & (x_1 - x)^2 & \dots & (x_1 - x)^p \\ 1 & (x_2 - x) & (x_2 - x)^2 & \dots & (x_2 - x)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (x_n - x) & (x_n - x)^2 & \dots & (x_n - x)^p \end{bmatrix} \quad (9)$$

β merupakan matriks berukuran $p + 1 \times 1$ dengan

$$\beta = [\beta_0(x), \beta_1(x), \dots, \beta_p(x)]^T,$$

Dan

$$m(x) = [m(x_1), m(x_2), \dots, m(x_n)]^T$$

Misalkan

$$\hat{\beta} = [\hat{\beta}_0(x), \hat{\beta}_1(x), \hat{\beta}_2(x), \dots, \hat{\beta}_p(x)]^T$$

dinyatakan penduga β . Penduga ini didapatkan melalui metode *Weighted Least Squares* (WLS), yang mana fungsi kernel digunakan sebagai pembobot (Silverman, 1992).

Dalam penelitian ini digunakan derajat polinomial lokal sebesar 1 (lokal linear), yang merupakan pendekatan paling dasar dalam regresi kernel polinomial lokal. Pendekatan lokal linear dipilih karena bersifat stabil pada bagian tepi data (boundary) dan cukup fleksibel untuk menangkap pola non-linear sederhana yang mungkin ada dalam hubungan antara variabel.

5) Pemilihan Bandwidth h

Bandwidth (h) adalah parameter dalam teknik pemulusan yang digunakan untuk mengatur kemulusan (*smoothing*) bentuk kurva yang dihasilkan. *Bandwidth* menentukan seberapa besar jendela atau rentang data yang digunakan dalam regresi kernel, serta memberikan bobot pada setiap titik dalam jangkauan tersebut. Peran utama jendela ini adalah membantu fungsi kernel dalam menghasilkan kurva yang lebih halus. Jika *bandwidth* terlalu besar, hasil pemodelan akan memiliki bias yang tinggi, tetapi variansnya rendah. Sebaliknya, jika *bandwidth* terlalu kecil, bisa menjadi lebih rendah, tetapi variansnya meningkat. Oleh karena itu, memilih bandwidth yang tepat sangat penting agar diperoleh estimasi yang seimbang antara bias dan keragaman data (Silverman, 1992).

Agar hasil estimasi optimal, diperlukan bandwidth yang sesuai. Dalam penelitian ini, pemilihan bandwidth dilakukan menggunakan metode Direct Plug-In (DPI). Metode ini digunakan untuk mendapatkan bandwidth optimal secara teoritis tanpa melalui prosedur validasi silang eksplisit. Pendekatan DPI didasarkan pada teori asimtotik, dan dalam penelitian ini diimplementasikan melalui fungsi *dpill* dari paket *KernSmooth* pada perangkat lunak R.

Adapun formula umum estimasi bandwidth menggunakan metode Direct Plug-In (DPI) adalah sebagai berikut (Ruppert et al., 1995):

$$\hat{h}_{DPI} = C_1(K) \left[\frac{\hat{\sigma}_1^2 (\hat{\lambda}_{AMSE})(b-a)}{\hat{\theta}_{22}^{0.05} (\hat{g}_{AMSE})n} \right]^{\frac{1}{5}} \quad (10)$$

$$\hat{\lambda}_{AMSE} = C_3(K) \left[\frac{\hat{\sigma}_Q^4 (\hat{N})(b-a)}{\hat{\theta}_{22}^{0.05} (\hat{g}_{AMSE})^2 n^2} \right]^{\frac{1}{9}} \quad (11)$$

$$\hat{g}_{AMSE} = C_2(K) \left[\frac{\hat{\sigma}_Q^2 (\hat{N})(b-a)}{|\hat{\theta}_{24}^Q (\hat{N})| n} \right]^{\frac{1}{7}} \quad (12)$$

Dalam implementasinya di R, semua komponen tersebut dihitung secara internal oleh fungsi `dpill`, sehingga pengguna cukup menyediakan data X dan Y. Pendekatan DPI ini juga digunakan dalam studi oleh Fitri & Almuhyar (2025) untuk mengestimasi hubungan nonlinier antara prevalensi stunting dan rata-rata lama sekolah, dengan hasil yang menunjukkan keefektifan metode ini dalam menghasilkan model yang fleksibel dan akurat.

6) Evaluasi Model dan Pemilihan Model Terbaik

Evaluasi dilakukan dengan membandingkan nilai Koefisien Determinasi (R^2) dari model OLS dan model Kernel Polinomial Lokal. Persamaan koefisien determinasi yang digunakan adalah:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (13)$$

Jika, R^2 mendekati 1 menunjukkan model yang semakin baik.

Model terbaik dipilih berdasarkan nilai Koefisien Determinasi (R^2) tertinggi. Kriteria pemilihan model yaitu sebagai berikut:

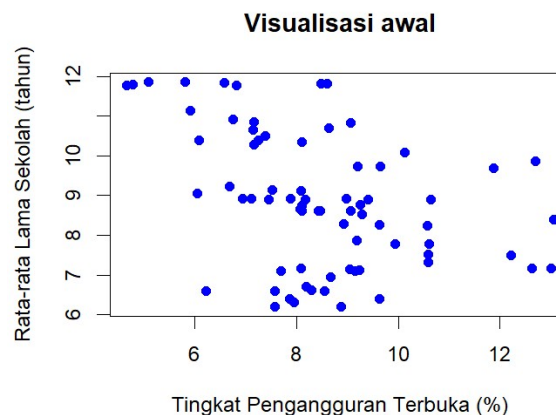
- Jika $R^2_{Kernel} > R^2_{OLS}$, maka model Kernel Polinomial Lokal lebih baik
- Jika $R^2_{Kernel} < R^2_{OLS}$, maka model OLS lebih baik

Selain nilai R^2 , pola hubungan yang terlihat pada grafik juga dijadikan pertimbangan. Model yang memberikan interpretasi lebih realistis terhadap hubungan antara dua variabel tersebut akan dipilih.

III. HASIL DAN PEMBAHASAN

A. Visualisasi Data

Langkah awal dalam memahami pola hubungan antara variabel bebas dan variabel terikat adalah melalui visualisasi data. Grafik scatter plot digunakan untuk melihat distribusi data serta mengidentifikasi indikasi pola linear maupun non-linear.



Gambar 1. Plot Visualisasi awal hubungan antara TPT dan Rata-rata Lama Sekolah dengan model OLS.

Grafik scatter plot di atas menunjukkan hubungan antara Tingkat Pengangguran Terbuka (TPT) dalam persen dan Rata-rata Lama Sekolah dalam tahun. Dari sebaran titik-titik data, terlihat bahwa tidak ada pola yang jelas yang menunjukkan hubungan kuat antara kedua variabel tersebut. Rata-rata lama sekolah bervariasi pada hampir setiap tingkat pengangguran, yang mengindikasikan bahwa semakin lama seseorang bersekolah tidak selalu berbanding lurus dengan rendahnya tingkat pengangguran. Beberapa titik data yang menyimpang juga menunjukkan kemungkinan adanya faktor lain yang mempengaruhi hubungan ini, seperti ketidaksesuaian keterampilan dengan kebutuhan pasar kerja, kualitas pendidikan, atau faktor ekonomi makro yang mempengaruhi ketersediaan lapangan kerja. Dengan demikian, berdasarkan visualisasi ini, dapat disimpulkan bahwa hubungan antara tingkat pengangguran terbuka dan rata-rata lama sekolah cenderung lemah atau tidak signifikan.

Dan Visualisasi awal ini menjadi langkah penting dalam menentukan model analisis yang sesuai, serta dapat diperkuat dengan analisis statistik lebih lanjut, di penelitian ini kita menggunakan analisis statistik dengan Regresi linier OLS dan Regresi Nonparametrik dengan Penduga Kernel Polinomial Lokal.

B. Pemodelan Regresi Linier OLS

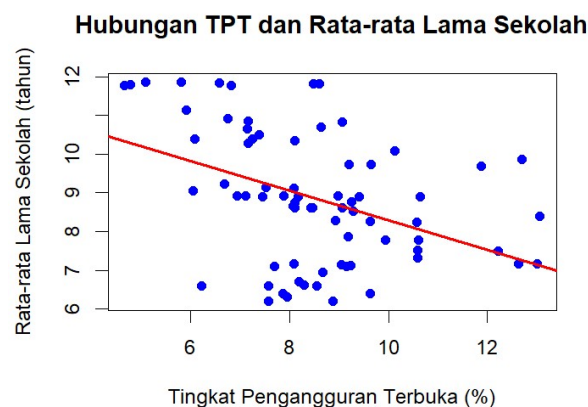
Model regresi linear sederhana yang diterapkan dengan menggunakan metode Regresi Kuadrat Terkecil (OLS) berguna untuk menganalisis hubungan antara Rata-Rata Lama Sekolah sebagai variabel respon (Y) dan tingkat pengangguran terbuka sebagai variabel prediktor (X). Berikut hasil estimasi model menunjukkan persamaan regresi sebagai berikut:

$$\hat{Y} = 12.0999 - 0.3804X \quad (14)$$

Hasil estimasi menunjukkan bahwa nilai intersep sebesar 12.0999 mengindikasikan bahwa jika Tingkat Pengangguran Terbuka adalah 0%, maka Rata-rata Lama Sekolah diprediksi sebesar 12.0999 tahun. Koefisien regresi sebesar -0.3804 ini menunjukkan bahwa setiap kenaikan 1% dalam Tingkat Pengangguran Terbuka akan menyebabkan penurunan Rata-rata Lama Sekolah sebesar 0.3804 tahun. Nilai ini cukup rendah, yang mengindikasikan bahwa terdapat faktor lain di luar pengangguran yang berperan lebih besar dalam menentukan lama pendidikan seseorang, seperti faktor ekonomi, kebijakan pendidikan, dan infrastruktur sekolah.

Berdasarkan data yang dianalisis, didapatkan nilai *Adjusted R-squared* sebesar 0.1592, ini menunjukkan bahwa hanya 15.92% variabilitas dalam Rata-rata Lama Sekolah yang dapat dijelaskan oleh Tingkat Pengangguran Terbuka. Hal ini dapat disimpulkan bahwa terdapat faktor lain yang mempengaruhi variabel respon/Rata-rata lama sekolah.

Hasil visualisasi model OLS yang ditampilkan pada **Gambar 1**, menunjukkan garis regresi cenderung menurun. Ini mengindikasikan bahwa terdapat hubungan negatif antara Tingkat Pengangguran Terbuka dan Rata-rata Lama Sekolah. Namun, ada beberapa titik data yang menyimpang dari garis regresi, yang menunjukkan bahwa model linier mungkin tidak sepenuhnya menangkap pola hubungan sebenarnya.



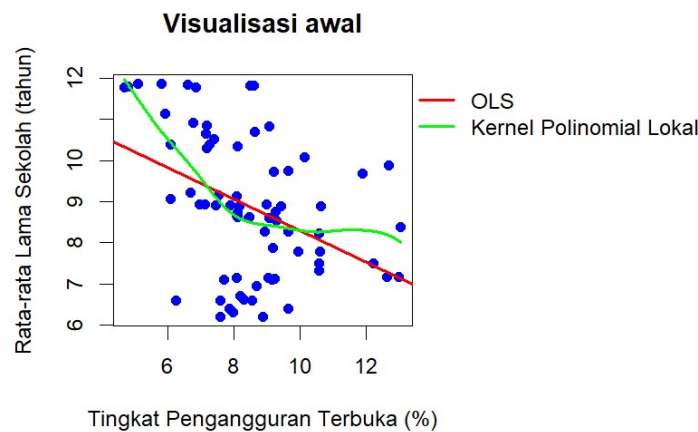
Gambar 2. Plot hubungan antara TPT dan Rata-rata Lama Sekolah dengan model OLS.

C. Pemodelan Regresi Nonparametrik dengan Penduga Kernel Polinomial Lokal

Untuk menangkap pola hubungan yang lebih kompleks antara kedua variabel tersebut, dapat menggunakan metode regresi nonparametrik dengan Penduga Kernel Polinomial Lokal. Metode ini lebih fleksibel dibandingkan regresi linier dikarenakan tidak mengasumsikan bentuk hubungan tertentu antara variabel. Pada analisis ini, nilai Bandwidth Optimal yang diperoleh adalah 0.8774, yang ditentukan menggunakan metode *Direct Plug-In (DPI)*. Model ini diterapkan menggunakan derajat polinomial 1 (lokal linear), sehingga setiap titik diperkirakan dengan fungsi linear

yang disesuaikan secara lokal berdasarkan data sekitar. Pemilihan derajat ini bertujuan untuk menghindari overfitting dan tetap mempertahankan stabilitas model, terutama di bagian tepi data. Dengan *bandwidth* ini, model regresi nonparametrik lebih baik dalam menangkap pola hubungan antara Rata-rata Lama Sekolah dan Tingkat Pengangguran Terbuka.

Hasil regresi Kernel Polinomial Lokal yang divisualisasikan dalam **Gambar 2**, menunjukkan bahwa model ini lebih akurat dalam menggambarkan pola hubungan yang kompleks dibandingkan OLS. Kurva regresi nonparametrik menunjukkan bahwa hubungan antara Rata-rata Lama Sekolah dan Tingkat Pengangguran Terbuka tidak sepenuhnya linier. Pada beberapa tingkat pengangguran tertentu, hubungan antara kedua variabel mengalami perubahan arah atau memiliki pola yang lebih kompleks. Berdasarkan hasil regresi Kernel Polinomial Lokal, didapatkan nilai Koefisien Determinasi (*R-squared*) sebesar 0.2841, lebih tinggi dibandingkan regresi OLS sebesar 0.17102. Hal ini dapat disimpulkan bahwa model nonparametrik lebih dapat menangkap pola hubungan antara Rata-rata Lama Sekolah dan Tingkat Pengangguran Terbuka yang tidak terdeteksi dalam model linier.



Gambar 3. Plot Hubungan TPT dan Rata-rata Lama Sekolah dengan model Kernel Polinomial Lokal

D. Evaluasi Model dan Pemilihan Model Terbaik

Kriteria koefisien determinasi dari regresi kuadrat terkecil/OLS dan regresi polinomial lokal kernel digunakan untuk memilih model terbaik. Dari penelitian ini didapatkan nilai koefisien determinasi yang diperoleh ialah:

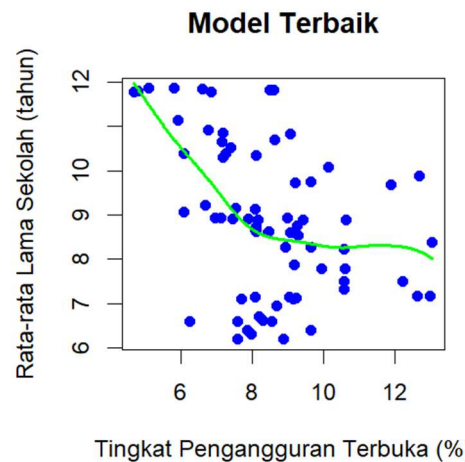
Tabel 2. Perbandingan Nilai Koefisien Determinasi Antar Model

Model	Koefisien Determinasi (R^2)
Regresi Kuadrat Terkecil (OLS)	0.1710
Kernel Polinomial Lokal (degree =1)	0.2841

Berdasarkan **Tabel 2**, Model Kernel Polinomial Lokal dengan derajat 1 memiliki nilai koefisien determinasi sebesar 0,2841, lebih tinggi dibandingkan model Ordinary Least Square (OLS) yang hanya sebesar 0,1710. Artinya, sekitar 28,41% variasi dalam Tingkat Pengangguran Terbuka (TPT) dapat dijelaskan oleh Rata-Rata Lama Sekolah pada model nonparametrik, sedangkan OLS hanya mampu menjelaskan sebesar 17,10%. Hal ini menunjukkan bahwa model regresi polinomial lokal lebih mampu menangkap pola hubungan antara kedua variabel secara lebih baik, terutama dalam konteks hubungan yang tidak bersifat linear.

Dengan demikian, regresi polinomial lokal kernel lebih baik dalam menangkap pola hubungan antara Rata-rata Lama Sekolah dan Pengangguran Terbuka dibandingkan dengan regresi kuadrat terkecil (OLS). Oleh karena itu, model Kernel Polinomial Lokal dipilih sebagai model terbaik dalam penelitian ini dan dapat dilihat di **Gambar 3** yang menunjukkan bahwa model nonparametrik lebih sesuai dalam menggambarkan pola data yang sebenarnya.

Selain itu, hasil ini menunjukkan bahwa tujuan penelitian telah tercapai, yakni menganalisis hubungan antara Rata-Rata Lama Sekolah dan Tingkat Pengangguran Terbuka (TPT) menggunakan pendekatan regresi nonparametrik. Model yang digunakan telah mampu menjelaskan variasi data secara lebih baik dan sesuai dengan karakteristik hubungan yang diteliti.



Gambar 4. Plot perbandingan model terbaik antara OLS dan Kernel Polinomial Lokal

IV. KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan, dapat disimpulkan bahwa terdapat hubungan negatif antara Rata-rata Lama Sekolah dan Tingkat Pengangguran Terbuka (TPT), sebagaimana ditunjukkan oleh model regresi OLS. Namun, hubungan ini tidak dapat dijelaskan sepenuhnya oleh model linier karena nilai koefisien determinasi yang rendah. Model regresi nonparametrik dengan Penduga Kernel Polinomial Lokal berderajat 1 terbukti lebih baik dalam menangkap pola hubungan yang kompleks, dengan nilai R-squared yang lebih tinggi dibandingkan model OLS.

Selain itu, hasil analisis menunjukkan bahwa Tingkat Pengangguran Terbuka bukan satu-satunya faktor yang mempengaruhi Rata-rata Lama Sekolah. Terdapat variabel lain seperti faktor ekonomi, kebijakan pendidikan, dan akses terhadap fasilitas pendidikan yang juga perlu diperhitungkan dalam analisis lebih lanjut. Oleh karena itu, model regresi nonparametrik dengan derajat 1 lebih direkomendasikan dalam analisis hubungan antara Tingkat Pengangguran Terbuka dan Rata-rata Lama Sekolah karena memberikan hasil yang lebih akurat dan fleksibel dibandingkan dengan regresi linier OLS. Hasil penelitian ini juga menunjukkan bahwa hubungan antara tingkat pengangguran dan lama pendidikan tidak selalu bersifat sederhana dan linier.

Untuk penelitian selanjutnya, disarankan agar menggunakan data panel atau data individu yang lebih rinci, serta mempertimbangkan variabel lain yang relevan seperti jenis pekerjaan, sektor ekonomi, keterampilan tenaga kerja, dan kondisi wilayah. Selain itu, pengujian dengan pendekatan nonparametrik lain atau hybrid modeling (misalnya kombinasi antara ARIMA dan machine learning) juga dapat dilakukan untuk memperoleh hasil analisis yang lebih komprehensif.

DAFTAR PUSTAKA

- Ace, S. (2002). *Pendidikan, investasi SDM, dan pembangunan: isu, teori, dan aplikasi* / Ace Suryadi (Cet. ke-2). Balai Pustaka.
- Beribe, K. E., & Budyanra. (2023). *Determinan kemiskinan multidimensi rumah tangga di Provinsi Banten tahun 2020*. 18(1). <https://doi.org/10.55981/jki.2023.1695>
- Dianty, M. AL, Suhartawan, B., Daawia, Mohzana, Suprpto, A., Agustini, P., Erlyn, P., H.Buan, C., Mataram, K. A., & Achmad. (n.d.). *STATISTIK DASAR* (P. T. Cahyono (Ed.)). Yayasan Cendikia Mulia Mandiri.
- Fan, J., & Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC.
- Fitri, F., & Almuhayar, M. (2025). Comparison of Linear Regression and Polynomial Local Regression in Modeling Prevalence of Stunting. *Journal of Mathematic and Mathematics Education*, 4(1).
- Guidom, A. (2015). Kernel Estimator and Bandwidth Selection for Density and its Derivatives. *The Kedd Package*.

- Kurniawan, J., Hartoto, Fahmi, A. Z., Ahyani, H., Hikmah, Ridwan, M., Amane, A. P. O., Afnarius, S., Priyanda, R., Arnita, Yudawisastra, H. G., Rosmawati, A., & Hozairi. (2023). Analisis Dan Visualisasi Data. In *Angewandte Chemie International Edition*, 6(11), 951–952. (Vol. 3, Issue 1). <https://medium.com/@arifwicaksanaa/pengertian-use-case-a7e576e1b6bf>
- Ruppert, D., Sheather, S. J., & Wand, M. P. (1995). An Effective Bandwidth Selector for Local Least Squares Regression. *Journal of the American Statistical Association*, 90(432), 1257–1270.
- Sari, F. P., & Kurniawan, R. (2022). Pengaruh Promosi, Harga dan Ulasan Produk Terhadap Keputusan Pembelian Pada Marketplace Shopee. *Simposium Manajemen Dan Bisnis (SIMANIS)*, 5.
- Silverman, B. (1992). *Density Estimation For Statistick and Data Analysis*. Chapman and Hall.
- Tingkat Pengangguran Terbuka (TPT) Menurut Kabupaten/Kota di Provinsi Banten (persen). (2024). <https://banten.bps.go.id/id/statistics-table/2/MTU3IzI=/tingkat-pengangguran-terbuka-tpt-menurut-kabupaten-kota-di-provinsi-banten.html>
- Widiansyah, A. (2017). *Peran Ekonomi dalam Pendidikan dan Pendidikan dalam Pembangunan Ekonomi*. XVII(2).