ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 12, Nomor 1, Tahun 2023, Halaman 20 - 29
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

PERBANDINGAN MODEL ARIMA DENGAN MODEL NONPARAMETRIK POLINOMIAL LOKAL DAN *SPLINE TRUNCATED* UNTUK PERAMALAN HARGA MINYAK MENTAH *WEST TEXAS INTERMEDIATE* (WTI) DILENGKAPI GUI R

Salsabila Rizkia Gusman[1*], Suparti[2], Agus Rusgiyono[3]
[1,2,3] Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
***e-mail** : rgsalsabil@gmail.com

DOI: 10.14710/j.gauss.12.1.20-29

Article Info:
Received: 2022-09-06
Accepted: 2022-11-11
Available Online: 2023-05-4

Keywords:
Crude oil; WTI; Local Polynomial; Spline Truncated; ARIMA; GUI.

Abstract: Crude oil as one of the most important natural resources experiences price fluctuations from time to time, even the spot price of West Texas Intermediate (WTI) world crude oil on 20th April 2020 reached -36,98 USD/barrel due to the Covid-19 pandemic. WTI oil price data was modeled using the ARIMA method, local polynomial, and spline truncated nonparametric regression then compared and obtained the best model and formed R Graphical User Interface (GUI). The ARIMA model and nonparametric time series models can be used to model time series data, but in the ARIMA model there are assumptions that must be fulfilled. Nonparametric time series models, which include local polynomial model and truncated spline do not need to fulfill these assumptions. The ARIMA model obtained did not fulfilled the assumptions of normality and residual homoscedasticity, so the modeling was stopped and modeling was only carried out using nonparametric regression methods. Based on the minimum MSE criteria, the best nonparametric model was obtained, namely nonparametric truncated spline model with degrees 3 and 3 knot points which was categorized as a strong model based on R-squared in sample value and having a very good forecasting ability based on MAPE out sample value.

1. PENDAHULUAN

Minyak mentah sebagai salah satu sumber daya alam yang sangat penting berperan sangat besar pada sektor perdagangan. Pencapaian dari perdagangan minyak mentah mencapai 10% dari total perdagangan di dunia (Faozi dan Sulistijanti, 2016). Minyak mentah dunia mengalami fluktuasi harga seiring berjalannya waktu bahkan minyak mentah *West Texas Intermediate* (WTI) mencapai -36,98 USD/barel pada 20 April 2020 karena pandemi Covid-19. Pemodelan dari harga minyak mentah dunia perlu dilakukan karena fluktuasinya dapat memberikan imbas yang signifikan bagi perekonomian suatu negara.

Metode ARIMA dan metode regresi nonparametrik dapat digunakan untuk melakukan pemodelan dengan data *time series*. Metode parametrik ARIMA harus memenuhi asumsi yaitu stasioneritas data, normalitas, nonautokorelasi, dan homoskedastisitas residual. Model *time series* nonparametrik yang diantaranya terdapat model polinomial lokal dan *spline truncated* tidak harus memenuhi asumsi-asumsi tersebut. Menurut Suparti *et al*. (2018), regresi nonparametrik cocok digunakan untuk data yang berpola naik turun atau fluktuatif dan tidak membentuk pola hubungan tertentu.

Penelitian ini bertujuan untuk mendapatkan model terbaik untuk memodelkan data *time series* harga minyak mentah WTI dari metode-metode yang digunakan yaitu metode parametrik ARIMA dan dua metode *time series* nonparametrik yaitu polinomial lokal dan *spline truncated*. Pembentukan *Graphical User Interface* (GUI) R dilakukan dengan *package* Shiny untuk pemodelan nonparametrik polinomial lokal dan *spline truncated* sehingga aplikasi pemodelan menjadi lebih mudah dan praktis dengan tampilan web yang interaktif.

2. TINJAUAN PUSTAKA

Minyak mentah sebagai sumber energi utama kendaraan merupakan salah satu produk penggerak perekonomian yang paling penting di suatu negara. *West Texas Intermediate* (WTI) adalah salah satu jenis minyak mentah yang dijadikan sebagai tolak ukur dari penentuan harga minyak mentah dunia. Minyak mentah WTI ringan dan manis sehingga memiliki kualitas yang bagus untuk diolah menjadi bensin. Kualitas minyak mentah WTI yang tinggi ini menjadikannya sebagai salah satu harga standar minyak dunia (Ramadani, 2021).

Model *Autoregressive Integrated Moving Average* (ARIMA) adalah model ARMA yang distasionerkan dengan melakukan *differencing* pertama atau lebih. Menurut Wei (2006), model ARIMA(p,d,q) adalah:

$$\phi_p(B)(1-B)^d Z_i = \theta_q(B)\varepsilon_i \tag{1}$$

Proses AR(p) atau ARIMA(p,0,0) dapat dituliskan sebagai berikut.

$$Z_i = \phi_1 Z_{i-1} + \phi_2 Z_{i-2} + \cdots + \phi_p Z_{i-p} + \varepsilon_i \tag{2}$$

Proses MA(q) atau ARIMA(0,0,q) dapat dituliskan sebagai berikut.

$$Z_i = \varepsilon_i - \theta_1 \varepsilon_{i-1} - \theta_2 \varepsilon_{i-2} - \cdots - \theta_q \varepsilon_{i-q} \tag{3}$$

dengan $\varepsilon_i \sim NID(0, \sigma^2)$.

Model subset ARIMA adalah bagian dari model ARIMA yang tergeneralisasi dan merupakan bagian dari model ARIMA dengan beberapa parameternya sama dengan nol (Tarno, 2013). Pemodelan ARIMA memerlukan terpenuhinya stasioneritas data dan juga terpenuhinya asumsi normalitas, nonautokorelasi, dan homoskedastisitas residual data.

Estimasi kurva dengan pendekatan nonparametrik memiliki fleksibilitas yang tinggi karena data mencari sendiri bentuk estimasi kurva regresinya (Eubank, 1999). Secara matematis, model regresi nonparametrik adalah sebagai berikut.

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1,2,\dots,n \tag{4}$$

dengan y_i merupakan variabel respon pada pengamatan ke-*i*, $f(x_i)$ adalah fungsi regresi nonparametrik yang tidak diketahui bentuk kurvanya, x_i adalah variabel prediktor pada pengamatan ke -*i*, dan ε_i adalah *error* pada pengamatan ke -*i* yang diasumsikan terdistribusi identik, independen, dengan *mean* 0 dan varians σ^2.

Regresi nonparametrik polinomial lokal merupakan salah satu metode regresi nonparametrik dengan menaksir fungsi regresi $m(x)$ dengan bentuk polinomial. Penggunaan metode *Weighted Least Square* (WLS) sebagai metode estimasi memerlukan fungsi Kernel sebagai pembobot. Berikut adalah beberapa jenis fungsi kernel, diantaranya (Hardle, 1991):

1. Kernel Gaussian

   $K(x) = \frac{1}{\sqrt{2}\pi} e^{-x^2/2}$, untuk $|x| < \infty$

2. Kernel Epanechnikov

   $K(x) = \begin{cases} 0,75(1-x^2), & |x| \le 1 \\ 0, & x \text{ lainnya} \end{cases}$

3. Kernel Segitiga

   $K(x) = \begin{cases} 1-|x|, & |x| \le 1 \\ 0, & x \text{ lainnya} \end{cases}$

4. Kernel Uniform

   $K(x) = \begin{cases} \frac{1}{2}, & \text{untuk } |\mathrm{x}| \le 1 \\ 0, & \mathrm{x} \text{ lainnya} \end{cases}$

Wu dan Zhang (2006) memberikan persamaan model regresi nonparametrik yang dapat didekati secara lokal oleh polinomial berderajat *m* oleh deret Taylor di sekitar $x = x_0$ dengan $\beta_r = \frac{f^{(r)}(x_0)}{r!}$, dan $r = 0,1,2,\dots,m$ sebagai berikut.

$$f(x) \approx \beta_0(x_0) + (x - x_0)\beta_1(x_0) + \cdots + (x - x_0)^m\beta_m(x_0) \quad (5)$$

Fan dan Gijbels (1996) mendefinisikan model regresi polinomial lokal derajat *m* sebagai berikut:

$$y_i = f(x_i) + \varepsilon_i \quad (6)$$

dengan $f(x_i) = \sum_{r=0}^{m} \beta_r\,(x_i - x_0)^r$ dan $i = 1,2,\dots,n$.
Persamaan (6) dalam bentuk matriks dapat dinyatakan sebagai berikut.

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \quad (7)$$

dengan

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix};\ \mathbf{X} = \begin{bmatrix} 1 & (x_1 - x_0) & (x_1 - x_0)^2 & \cdots & (x_1 - x_0)^m \\ 1 & (x_2 - x_0) & (x_2 - x_0)^2 & \cdots & (x_2 - x_0)^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (x_n - x_0) & (x_n - x_0)^2 & \cdots & (x_n - x_0)^m \end{bmatrix};\ \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix};\ \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Estimasi parameter $\boldsymbol{\beta}$ pada persamaan regresi nonparametrik polinomial lokal adalah:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X^T W X})^{-1}\mathbf{X^T W Y} \quad (8)$$

Sedangkan estimasi model regresi nonparametrik polinomial lokal adalah sebagai berikut.

$$\hat{\mathbf{Y}} = \hat{\mathbf{f}}(\mathbf{X}) = \mathbf{X}(\mathbf{X^T W X})^{-1}\mathbf{X^T W Y} = \mathbf{HY} \quad (9)$$

dengan $\mathbf{H} = \mathbf{X}(\mathbf{X^T W X})^{-1}\mathbf{X^T W}$ adalah matriks hat berukuran $(n \times n)$ untuk pemodelan polinomial dengan kuadrat terkecil (Takezawa, 2006).

Keakuratan dari estimasi kurva regresi polinomial lokal dipengaruhi oleh bandwidth *h*, derajat polinomial, dan bobot Kernel yang digunakan. Menurut Suparti *et al*. (2018), pemilihan fungsi Kernel dan derajat polinomial yang akan digunakan untuk pemodelan tidak terlalu penting dibandingkan pemilihan *bandwidth* karena jika terlalu besar estimasi kurva regresi yang diberikan akan menjadi *oversmooth* dan juga sebaliknya jika terlalu kecil akan menjadi *undersmooth*.

Spline merupakan sebuah fungsi berupa potongan (*truncated*) polinomial tersegmen yang kontinu. Titik knot tersebut adalah titik perpaduan bersama saat terdapat perubahan perilaku pada selang yang berlainan (Budiantara, 2005). Bentuk umum regresi *spline truncated* derajat *m* dengan titik knot $(\xi_1, \xi_2, \dots, \xi_k)$ satu variabel prediktor adalah:

$$y = \sum_{j=0}^{m} \beta_j x^j + \sum_{j=1}^{k} \beta_{j+m}(x - \xi_j)_+^m + \varepsilon \quad (10)$$

dengan $(a < \xi_1 < \xi_2 < \cdots < \xi_k < b)$, a adalah nilai minimum dari x dan b nilai maksimum dari x dan fungsi *truncated*: $(x - \xi_j)_+^m = \begin{cases} (x - \xi_j)^m & ;\ x - \xi_j \geq 0 \\ 0 & ;\ x - \xi_j < 0 \end{cases}$

Persamaan (10) dalam bentuk matriks dengan data amatan *n* dapat ditulis sebagai berikut.

$$\boldsymbol{Y} = \boldsymbol{X_1\delta_1} + \boldsymbol{X_2\delta_2} + \boldsymbol{\varepsilon} \quad (11)$$

dengan $\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$; $\boldsymbol{X_1} = \begin{bmatrix} 1 & x_1 & \cdots & x_1^m \\ 1 & x_2 & \cdots & x_2^m \\ 1 & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^m \end{bmatrix}$; $\boldsymbol{\delta_1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}$; $\boldsymbol{\delta_2} = \begin{bmatrix} \beta_{m+1} \\ \beta_{m+2} \\ \vdots \\ \beta_{m+k} \end{bmatrix}$; $\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

$$\boldsymbol{X_2} = \begin{bmatrix} (x_1 - \xi_1)_+^m & (x_1 - \xi_2)_+^m & \dots & (x_1 - \xi_k)_+^m \\ (x_2 - \xi_1)_+^m & (x_2 - \xi_2)_+^m & \dots & (x_2 - \xi_k)_+^m \\ \vdots & \vdots & \ddots & \vdots \\ (x_n - \xi_1)_+^m & (x_n - \xi_2)_+^m & \dots & (x_n - \xi_k)_+^m \end{bmatrix}$$

Estimasi parameter β adalah sebagai berikut.

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{\Pi}} = (\boldsymbol{X_\Pi^T X_\Pi})^{-1}\boldsymbol{X_\Pi^T Y} \quad (12)$$

Sedangkan fungsi penduga dari $f(x)$ adalah:

$$\hat{f}_\Pi = \boldsymbol{X}_\Pi \widehat{\boldsymbol{\beta}}_\Pi = \boldsymbol{X}_\Pi(\boldsymbol{X}_\Pi^T \boldsymbol{X}_\Pi)^{-1}\boldsymbol{X}_\Pi^T \boldsymbol{Y} = \boldsymbol{H}_\Pi \boldsymbol{Y} \quad (13)$$

dengan $\boldsymbol{H}_\Pi = \boldsymbol{X}_\Pi(\boldsymbol{X}_\Pi^T \boldsymbol{X}_\Pi)^{-1}\boldsymbol{X}_\Pi^T$ dan

$$\boldsymbol{X}_\Pi = \begin{bmatrix} 1 & x_1 & x_1^m & (x_1-\xi_1)_+^m & \cdots & (x_1-\xi_k)_+^m \\ 1 & x_2 & x_2^m & (x_2-\xi_1)_+^m & \cdots & (x_2-\xi_k)_+^m \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^m & (x_n-\xi_1)_+^m & \cdots & (x_n-\xi_k)_+^m \end{bmatrix}$$

Penentuan *bandwidth* optimum untuk model nonparametrik polinomial lokal maupun titik knot optimum untuk model nonparametrik *spline truncated* dapat menggunakan metode *Generalized Cross Validation* (GCV). Sedangkan *Mean Square Error* (MSE) digunakan untuk mendapatkan model terbaik dengan membandingkan model optimum dari ketiga model yang digunakan. Fungsi dari GCV adalah sebagai berikut.

$$GCV = \frac{MSE}{\left(\frac{1}{n} trace[\boldsymbol{I}-\boldsymbol{H}]\right)^2} \quad (14)$$

dengan

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (15)$$

dan n: banyak pengamatan; **I**: matriks identitas; $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$; y_i: data aktual ke-i; dan $\hat{y}_i$: data estimasi ke-i.

Pemodelan data *time series* dengan metode regresi nonparametrik berdasarkan lag paling signifikan pada plot *Partial Autocorrelation Function* (PACF), sehingga model regresi nonparametrik *time series* mengacu pada persamaan (2) dan (4) jika data signifikan pada lag ke-*p* dapat didefinisikan sebagai berikut.

$$y_i = f(y_{i-p}) + \varepsilon_i, \quad i = (p+1), (p+2), \dots, n \quad (16)$$

Koefisien determinasi (R^2) yang diaplikasikan pada data *in sample* metode regresi nonparametrik digunakan untuk mengukur seberapa besar proporsi variasi variabel dependen dijelaskan oleh variabel independen. Berikut adalah rumus koefisien determinasi.

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (17)$$

dengan $\bar{y}$ adalah rata-rata data aktual. Kriteria R^2 adalah jika R^2 < 0,30 maka model lemah, jika 0,30 ≤ R^2 < 0,60 maka model cukup kuat, dan jika R^2 ≥ 0,60 maka model kuat (Sanchez, 2013). Nilai dari koefisien determinasi yang kecil memberi arti kemampuan variabel independen dalam menjelaskan variasi variabel dependen sangat terbatas dan sebaliknya jika nilai koefisien determinasi mendekati 1 maka variabel independen dapat memberikan hampir semua informasi yang dibutuhkan dalam memprediksi variabel dependen.

MAPE (*Mean Absolute Percentage Error*) data *out sample* dapat memberikan petunjuk tentang ketepatan peramalan dari suatu model yang dirumuskan sebagai berikut.

$$MAPE = \sum_{i=1}^{n} \frac{\left|\frac{y_i - \hat{y}_i}{y_i}\right|}{n} \times 100\% \quad (18)$$

Kriteria MAPE adalah jika MAPE < 10% maka kemampuan peramalan sangat baik, jika 10% ≤ MAPE <20% maka baik, jika 20% ≤ MAPE < 50% maka cukup, dan jika MAPE ≥ 50% maka kemampuan peramalan buruk (Chen *et al*., 2003). Penggunaan MAPE diaplikasikan pada data *out sample* untuk melihat evaluasi kinerja model yang telah dibuat dengan data *in sample*.

Graphical User Interface (GUI) adalah pengembangan yang dilakukan untuk membuat R dapat diakses melalui menu grafis sehingga lebih ramah untuk pengguna dibandingkan dengan R yang berbasis *Command Line Interface* (CLI) yang lebih cocok digunakan untuk para pengembang. Aplikasi Shiny dilengkapi dengan berbagai *widget* sehingga dapat membentuk *user interface* yang interaktif dan juga cepat. Perancangan *default* dalam aplikasi Shiny efektif dan sangat mudah dalam hal pengembangan dan pengintegrasian aplikasi Shiny dengan konten web menggunakan HTML dan CSS (Beeley, 2013).

3. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari *website US Energy Information Administration* (www.eia.gov). Data yang diambil merupakan data harian harga minyak mentah *West Texas Intermediate* (WTI) mulai pada tanggal 4 Januari 2021 sampai dengan 31 Desember 2021. Data kemudian dibagi dengan proporsi 80:20 menjadi data *in sample* sebanyak 201 data (4 Januari 2021 - 19 Oktober 2021) dan *out sample* sebanyak 50 data (20 Oktober 2021 - 31 Desember 2021). Metode regresi nonparametrik menggunakan variabel respon data pada waktu ke-*i* Y_i dan variabel prediktor yang digunakan merupakan data pada waktu ke-(*i*-*p*) Y_{i-p} dengan p yang ditentukan berdasarkan lag paling signifikan dari plot *Partial Autocorrelation Function* (PACF) data *in sample*.

Pada penelitian ini, digunakan *software* RStudio dengan *package* Shiny untuk pembuatan *Graphical User Interface* (GUI). Langkah-langkah yang dilakukan dalam penelitian ini adalah sebagai berikut:

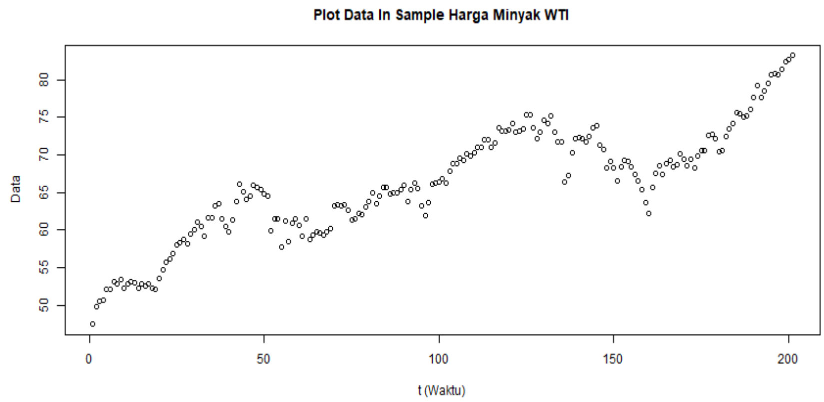
1. Membuat desain *user interface* (ui.R) dan menyusun sintaks pada *server* (server.R) pada *interface* GUI R.
2. Membagi data menjadi data *in sample* dan *out sample*.
3. Menganalisis statistik deskriptif dari data *in sample*.
4. Melakukan pemodelan dengan metode parametrik ARIMA dengan data *in sample*.
5. Pengujian asumsi model ARIMA. Jika semua asumsi terpenuhi dilanjutkan mencari model optimum dengan kriteria MSE dan jika asumsi tidak terpenuhi pemodelan ARIMA diselesaikan.
6. Menentukan variabel prediktor metode nonparametrik berdasarkan lag paling signifikan pada plot PACF data *in sample.* Jika plot PACF signifikan pada lag ke-*p* maka variabel prediktor adalah Y_{i-p}.
7. Membuat plot data Y_{i-p} dan Y_i dari penentuan variabel prediktor sebelumnya.
8. Mencari kombinasi derajat (antara 1, 2, 3, dan 4), *bandwidth* minimum dan maksimum untuk mencapai *bandwidth* optimum dengan kriteria GCV untuk model *time series* polinomial lokal optimum.
9. Mencari kombinasi derajat (antara 1, 2, 3, dan 4) dan titik knot (antara 1, 2, dan 3) dengan kriteria GCV untuk model *time series spline truncated* optimum.
10. Menghitung MSE dari model optimum yang didapatkan.
11. Melakukan perbandingan model optimum dengan kriteria MSE minimum sehingga diperoleh model terbaik.
12. Melakukan evaluasi dan pengukuran kebaikan model terbaik dengan MAPE *out sample* dan *R-squared in sample*.

4. HASIL DAN PEMBAHASAN

Data harga minyak mentah WTI dibagi dengan perbandingan 80:20 untuk data *in sample* dan *out sample*, data *in sample* sebanyak 201 data dari 4 Januari 2021 sampai dengan 19 Oktober 2021 dan data *out sample* sebanyak 50 data dari 20 Oktober 2021 sampai dengan 31 Desember 2021. Statistik deskriptif dan plot data *in sample* harga minyak mentah WTI terdapat pada Tabel 1 dan Gambar 1 berikut.

Tabel 1. Statistik Deskriptif Data *In Sample*

Minimum	Maksimum	Rata-rata	Varian
47,47	83,19	66,02	55,12



Gambar 1. Plot Data *In Sample* Harga Minyak WTI

Pemodelan data dengan metode ARIMA diawali dengan pengecekan stasioneritas data dalam hal *mean* dan varian. Stasioneritas dalam hal varian belum dipenuhi maka dilakukan transformasi yang bersesuaian yaitu Z_i^2. Data hasil transformasi kemudian diuji kestasioneran dalam hal *mean* dan didapatkan data belum stasioner maka dilakukan *differencing* satu kali. Identifikasi model ARIMA dilakukan dari plot ACF dan PACF data yang telah stasioner dan didapatkan model yang terbentuk adalah model subset ARIMA([18],1,0), ARIMA([18],1,[26]), dan ARIMA(0,1,[26]). Ketiga model ARIMA tersebut kemudian diuji asumsi dan didapatkan hasil bahwa ketiga model tidak memenuhi asumsi normalitas dan homoskedastisitas residual yang diperlukan, maka pemodelan ARIMA dihentikan dan hanya dilakukan pemodelan dengan model *time series* nonparametrik.

Pemodelan dengan metode regresi nonparametrik diawali dengan menentukan variabel respon dan prediktor dari plot PACF data. Plot PACF data menunjukkan bahwa lag 1 memiliki korelasi yang kuat dan signifikan maka variabel prediktor menggunakan data pada waktu ke-(*i-1*) Y_{i-1} dan variabel respon menggunakan data pada waktu ke-*i* Y_i.

Pembentukan model *time series* nonparametrik polinomial lokal mengkombinasikan derajat polinomial, *bandwidth* minimum dan maksimum untuk mencapai *bandwidth* optimum dengan kriteria *Generalized Cross Validation* (GCV) dengan bantuan komputasi dari program RStudio. Pengolahan data menggunakan RStudio mencobakan titik lokal y_0 dengan interval (nilai minimum data + 1) sampai dengan (nilai maksimum data - 1) sedangkan fungsi kernel menggunakan kernel Gaussian, *bandwidth* dicobakan dari 1 sampai dengan 100, dan derajat yang dicobakan adalah derajat 1, 2, 3, dan 4.

Berdasarkan hasil pencarian model polinomial lokal dengan kriteria GCV, didapatkan model optimum dengan GCV sebesar 1,709126 adalah model *time series* nonparametrik polinomial lokal dengan *bandwidth* optimum 100, derajat 2, dan titik lokal 70,47. Berikut adalah estimasi model optimum polinomial lokal.

$$\hat{y}_i = 70{,}51397 + 1{,}002454(y_{i-1} - 70{,}47) + 0{,}001965(y_{i-1} - 70{,}47)^2 \qquad (19)$$

Pencarian model nonparametrik *spline truncated* diawali dengan mengkombinasikan derajat dan jumlah titik knot yang akan digunakan dengan bantuan komputasi pada GUI R. Kombinasi derajat yang digunakan adalah menggunakan derajat 1 sampai dengan 4 dan jumlah titik knot yang digunakan adalah 1 sampai dengan 3 titik knot. Berdasarkan kriteria GCV, didapatkan model optimum *spline truncated* GCV sebesar 0,7984775 adalah model dengan derajat 3 dan 3 titik knot di titik 53,08; 55,67; dan 64,55, berikut adalah hasil estimasi modelnya.

$$\begin{aligned}\hat{y}_i = {} & 3055{,}816 - 172{,}8674 y_{i-1} + 3{,}295137 y_{i-1}^2 - 0{,}02079916 y_{i-1}^3 \\ & + 0{,}02004963 (y_{i-1} - 53{,}08)_+^3 + 0{,}002012863 (y_{i-1} - 55{,}67)_+^3 \\ & - 0{,}00160743 (y_{i-1} - 64{,}55)_+^3\end{aligned} \qquad (20)$$

Pemodelan dengan metode parametrik ARIMA yang telah dilakukan menghasilkan kesimpulan bahwa model tidak memenuhi asumsi normalitas dan homoskedastisitas residual, melainkan hanya memenuhi asumsi nonautokorelasi. Pemodelan ARIMA diselesaikan dan hanya dilakukan pemodelan dengan metode regresi nonparametrik untuk data runtun waktu sebagai alternatif dari tidak terpenuhinya asumsi-asumsi yang ketat dan harus dipenuhi dalam pemodelan dengan metode parametrik ARIMA. Metode regresi nonparametrik yang digunakan adalah polinomial lokal dan *spline truncated.* Model optimum yang telah diperoleh yaitu model *time series* nonparametrik polinomial lokal dan *spline truncated* akan dilakukan perbandingan dengan kriteria MSE data *in sample*. Berikut adalah tabel perbandingan MSE dari model optimum yang didapatkan.

Tabel 2. Perbandingan Model Optimum

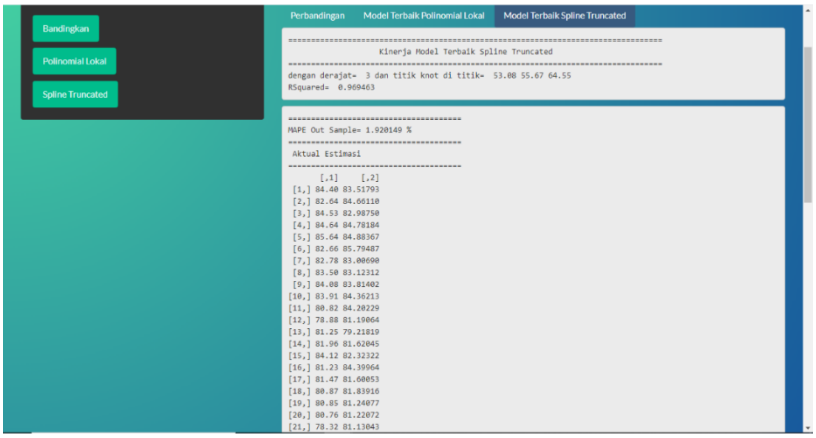
Model Optimum	MSE
Time series nonparametrik polinomial lokal	1,658237
Time series nonparametrik *spline truncated*	1,630406

Tampilan perbandingan model optimum polinomial lokal dan *spline truncated* pada GUI adalah sebagai berikut.

Gambar 2. Tab Perbandingan GUI untuk Pemodelan Harga Minyak Mentah WTI

Hasil *running* program RStudio untuk evaluasi kinerja dan pengukuran kebaikan model terbaik *spline truncated* yang outputnya terlihat pada GUI ditunjukkan pada Gambar 3.

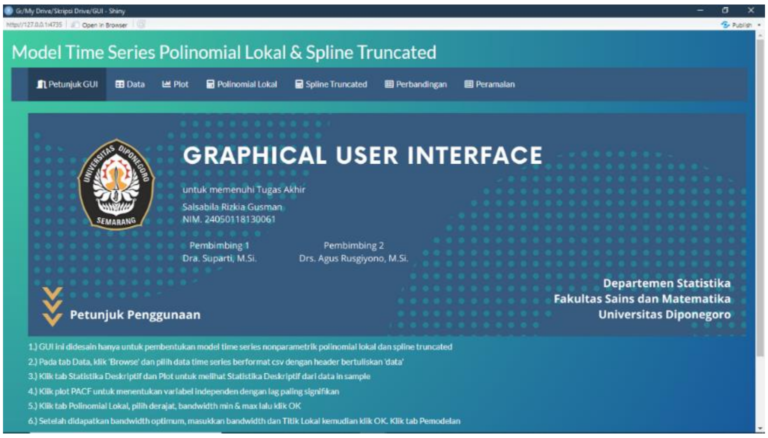
Output yang ditampilkan pada GUI R Gambar 3 menunjukkan *R-squared* data *in sample* untuk model terbaik nonparametrik *spline truncated* adalah sebesar 0,969463. Nilai *R-squared* sebesar 0,969463 yang terletak pada kategori $R^2 \geq 0{,}60$ sehingga dapat diinterpretasikan bahwa model kuat. Model terbaik tersebut juga menghasilkan MAPE *out sample* sebesar 1,920149% yang terletak pada kategori MAPE < 10% sehingga dapat diinterpretasikan bahwa model tersebut memiliki kemampuan peramalan yang sangat baik.



Gambar 3. Evaluasi Kinerja dan Pengukuran Kebaikan Model Terbaik *Spline Truncated*

Gambar 4. Plot Data *Out Sample* dan Estimasinya dengan Model Terbaik *Spline Truncated* Derajat 3 dan 3 Titik Knot di Titik 53,08; 55,67; dan 64,55

Penelitian ini dilengkapi dengan *Graphical User Interface* (GUI) yang dibuat dengan *package* Shiny pada *software* RStudio untuk pemodelan *time series* nonparametrik polinomial lokal dan *spline truncated*. *User Interface* (UI) sebagai komponen GUI dibagi menjadi enam tabPanel, diantaranya adalah petunjuk penggunaan GUI sebagai tampilan utama, tab data untuk menginput data dan statistika deskriptif, tab plot, tab pemodelan polinomial lokal, tab pemodelan *spline truncated*, dan tab perbandingan. Aplikasi GUI akan muncul ketika fungsi ‘shinyApp(ui,server)’ dijalankan dengan tampilan utama seperti pada Gambar 5 sebagai berikut.



Gambar 5. Tampilan Utama GUI

5. KESIMPULAN

Pada pemodelan dengan menggunakan metode parametrik ARIMA untuk data harga minyak mentah *West Texas Intermediate* (WTI), model yang terbentuk yaitu ARIMA([18],1,0), ARIMA([18],1,[26]), dan ARIMA(0,1,[26]) tidak memenuhi asumsi normalitas dan homoskedastisitas residual yang diperlukan, maka pemodelan dialihkan menggunakan metode alternatif regresi nonparametrik yang tidak ketat akan terpenuhinya asumsi-asumsi seperti pada pemodelan dengan metode parametrik. Pemodelan minyak mentah WTI dengan menggunakan metode regresi nonparametrik polinomial lokal menghasilkan model optimum dengan bobot fungsi kernel Gaussian derajat 2 dengan *bandwidth* optimum 100 dan titik lokal 70,47. Pemodelan minyak mentah WTI dengan menggunakan metode regresi nonparametrik *spline truncated* menghasilkan model optimum dengan derajat 3 dan 3 titik knot di titik 53,08; 55,67 dan 64,55.

Perbandingan model-model optimum yang telah didapatkan yaitu model optimum dari regresi nonparametrik polinomial lokal dan *spline truncated*, diperoleh model terbaik dengan kriteria MSE adalah model time series nonparametrik *spline truncated*. Model tersebut termasuk kategori model kuat berdasarkan *R-squared* data *in sample* (0,969463) dan kategori model yang dapat memberikan kemampuan peramalan sangat baik berdasarkan MAPE *out sample* (1,92015%).

Pembangunan *Graphical User Interface* (GUI) dengan program RStudio dilakukan untuk tampilan pemodelan dan perbandingan metode regresi nonparametrik polinomial lokal dan *spline truncated*. GUI R yang telah dibuat dapat membuat pengguna lebih mudah dalam menggunakannya untuk mengolah dan menganalisis data dengan tampilan yang lebih menarik dan tidak perlu menuliskan sintaks kembali.

DAFTAR PUSTAKA

Beeley, C. (2013). *Web application with R using Shiny*. Birmingham: Packt Publishing Ltd.

Budiantara, I. N. (2005). *Model Keluarga Spline Polinomial Truncated dalam Regresi Semiparametrik*. Berkala Ilmiah MIPA, 15(3), 55–61.

Chen, R. J. C., Bloomfield, P., dan Fu, J. S. (2003). *An evaluation of alternative forecasting methods to recreation visitation*. Journal of Leisure Research, 35(4), 441–454.

Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing Second Edition*. New York: Marcel Dekker, Inc.

Fan, J., dan Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall.
Faozi, S., dan Sulistijanti, W. (2016). *Peramalan Harga Minyak Mentah Standar West Texas Intermediate dengan Pendekatan Metode ARIMA*. Seminar Nasional Pendidikan Sains dan Teknologi, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang, 308–316.
Hardle, W. (1991). *Smoothing Techniques With Implementation in S*. New York: Springer-Verlag.
Ramadani, K. (2021). *Pemodelan Harga Minyak West Texas Intermediate Menggunakan Model ARIMA, ARFIMA, Fuzzy Time Series Markov Chain dan Hybrid ARIMA-FTSMC*. PhD Thesis. Universitas Andalas Padang.
Sanchez, G. (2013). *PLS Path Modeling with R*. Berkeley: *R Package Notes*.
Suparti, Santoso, R., Prahutama, A., dan Devi, A. R. (2018). *Regresi Nonparametrik*. Ponorogo: Wade Group.
Takezawa, K. (2006). *Introduction to Nonparametric Regression*. New Jersey: John Wiley & Sons, Inc.
Tarno. (2013). *Kombinasi Prosedur Pemodelan Subset Arima Dan Deteksi Outlier Untuk Prediksi Data Runtun Waktu*. Prosiding Seminar Nasional Statistika Universitas Diponegoro, 1970, 583–592.
Wei, W. W. S. (2006). *Time Series Analysis Univariate and Multivariate Methods Second Edition*. New York: Pearson Education, Inc.
Wu, H., dan Zhang, J. T. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis.* New Jersey: John Wiley & Sons, Inc.