

# Peramalan Jumlah Penumpang Berangkat Melalui Transportasi Udara di Sulawesi Tengah Menggunakan *Support Vector Regression (SVR)*

Drajat Indra Purnama<sup>1\*</sup>, Oki Prasetya Hendarsin<sup>2</sup>

<sup>1</sup>Badan Pusat Statistik, Kabupaten Parigi Moutong  
Jl. Pakabata Kompleks Perkantoran Bambaemo, Parigi Moutong 94471, Sulawesi Tengah, Indonesia

<sup>2</sup>Ussama Multi Inovasi  
Jl. Purbaleuwi B No. 42, Kota Bandung 40293, Jawa Barat, Indonesia

\* Penulis Korespondensi. Email: [drajatindrapurnama@bps.go.id](mailto:drajatindrapurnama@bps.go.id)

## ABSTRAK

Sulawesi Tengah memiliki tujuh bandara sebagai akses transportasi udara keluar atau masuk. Jumlah penumpang berangkat menggunakan transportasi udara melalui ketujuh bandara tersebut mengalami fluktuasi setiap bulannya. Oleh karena itu, dibutuhkan teknik peramalan yang tepat untuk melihat fluktuasi dan meramalkan jumlah penumpang di masa depan. Hasil pengujian data jumlah penumpang berangkat melalui transportasi udara di Sulawesi Tengah menunjukkan bahwa data memiliki pola nonlinear sehingga diperlukan metode peramalan yang dapat mengatasi permasalahan pola data nonlinear. Dalam artikel ini digunakan model SVR. Hasil peramalan data jumlah penumpang berangkat melalui transportasi udara di Sulawesi Tengah menggunakan SVR menunjukkan akurasi peramalan yang baik dengan nilai MAPE 7,28 persen untuk data *training* dan 18,67 persen untuk data *testing*.

### Kata Kunci:

SVR; Transportasi Udara; Penumpang; Nonlinear

## ABSTRACT

Central Sulawesi has seven airports as access to air transportation in or out. The number of passengers departing via air transportation through the seven airports is experiencing fluctuations every month. Therefore, accurate forecasting techniques are needed to see fluctuations and predict the number of passengers in the future. The results of testing the number of passengers departing via air transportation in Central Sulawesi show that the data has a nonlinear pattern so a forecasting method is needed that can overcome the problem of nonlinear data patterns. This study using the SVR model. The results of forecasting the number of passengers departing via air transportation in Central Sulawesi using SVR shows good forecasting accuracy with a MAPE value of 7.28 percent for training data and 18.67 percent for testing data.

### Keywords:

SVR; Air Transportation; Passengers; Nonlinear

**Format Sitasi:**

D. I. Purnama and O. P. Hendarsin, "Peramalan Jumlah Penumpang Berangkat Melalui Transportasi Udara di Sulawesi Tengah Menggunakan *Support Vector Regression (SVR)*," *Jambura J. Math.*, vol. 2, no. 2, pp.49-59, 2020.

**1. Pendahuluan**

Transportasi merupakan kebutuhan utama bagi masyarakat untuk menunjang aktivitas ekonomi, sosial, dan sebagainya. Hal ini menyebabkan terjadinya perkembangan sarana dan alat transportasi dari waktu ke waktu. Salah satu sarana transportasi yang menjadi favorit masyarakat di era modern sekarang ini adalah pesawat terbang. Pesawat terbang masih dilihat sebagai sarana transportasi paling efisien untuk menghemat waktu, pergi ke daerah yang jaraknya jauh dengan kondisi yang nyaman. Banyaknya pilihan maskapai bagi masyarakat untuk bepergian menggunakan pesawat terbang membuat sarana transportasi ini makin diminati. Hal ini juga terjadi di Provinsi Sulawesi Tengah dimana pesawat terbang menjadi sarana transportasi yang disukai masyarakat untuk bepergian jauh baik di dalam area Sulawesi Tengah maupun keluar area Sulawesi Tengah.

Menurut data Badan Pusat Statistik Sulawesi Tengah [1], terdapat tujuh akses transportasi udara yakni melalui Bandara Mutiara Sis Aljufri (Palu), Syukuran Aminuddin Amir (Luwuk), Kasiguncu (Poso), Bungku (Morowali), Sultan Bantilan (Tolitoli), Pogogul (Buol) dan Tanjung Api (Ampana). Data jumlah penumpang transportasi udara di Provinsi Sulawesi Tengah melalui tujuh bandara tersebut mengalami fluktuasi setiap bulannya khususnya jumlah penumpang yang berangkat. Sehingga diperlukan suatu model peramalan yang tepat untuk meramalkan jumlah penumpang di masa yang akan datang. Hal ini dilakukan terkait pada antisipasi dan kebijakan untuk mendukung dan mengembangkan pariwisata, mobilisasi arus penumpang serta berpartisipasi dalam perdagangan dan industri di kawasan Sulawesi Tengah.

Metode klasik yang umum digunakan dalam model deret waktu adalah model *Autoregressive Integrated Moving Average (ARIMA)* [2]. Model ARIMA cukup fleksibel dalam memodelkan sebagian besar pola data deret waktu karena dapat menganalisis situasi yang acak, trend, musiman, bahkan bersifat siklis pada data deret waktu yang akan dianalisis. Meskipun metode ARIMA cukup fleksibel dalam memodelkan sebagian besar pola data deret waktu, namun metode ARIMA kurang cocok untuk sebagian besar permasalahan yang bersifat nonlinear [3]. Metode ARIMA cenderung mengalami penurunan tingkat keakuratan apabila digunakan pada data yang mengandung komponen nonlinear [4]. Oleh karena itu, dibutuhkan metode lain untuk menangani kasus nonlinear.

Berbeda dengan model ARIMA, model *Support Vector Regression (SVR)* termasuk model peramalan yang dapat digunakan untuk memprediksi data deret waktu nonlinear. Keunggulan SVR adalah kemampuan untuk mengatasi masalah data nonlinear dengan menggunakan fungsi kernel sehingga dapat mengatasi masalah *overfitting* dimana model yang dihasilkan hanya menghasilkan model yang baik untuk data *training* dan tidak untuk data *testing* [5]. Salah satu cara yang dapat digunakan untuk menentukan parameter dalam SVR adalah dengan metode *grid search*. *Grid*

*search* merupakan kombinasi parameter yang diujikan pada suatu model SVR untuk mencari nilai *error* dalam klasifikasi [6].

Dalam artikel ini, digunakan model SVR untuk memodelkan dan meramalkan data jumlah penumpang berangkat melalui transportasi udara di Provinsi Sulawesi Tengah. Hal ini didasarkan pada pengujian linearitas data jumlah penumpang berangkat melalui transportasi udara di Provinsi Sulawesi Tengah menggunakan uji *Ramsey RESET* menunjukkan bahwa data memiliki pola nonlinear. Artikel ini akan menunjukkan Model SVR terbaik yang diperoleh pada data jumlah penumpang berangkat melalui transportasi udara di Sulawesi Tengah, disertai dengan kriteria peramalan yang akurat berdasarkan nilai MAPE (*Mean Absolute Percentage Error*).

## 2. Metode

### 2.1. Uji Linearitas *Ramsey RESET*

Salah satu uji untuk mengetahui linear atau nonlinear suatu data deret waktu dapat menggunakan uji *Ramsey RESET* yang pertama kali diperkenalkan oleh Ramsey [7]. Jika terdapat variabel bebas  $X_t$  dan variabel prediktor  $Y_t$ , prosedur uji *Ramsey RESET* dapat dijelaskan sebagai berikut;

- a. Melakukan regresi  $Y_t$  pada  $X_t$  sehingga diperoleh model linear pada Persamaan (1),

$$Y_t = f_t + e_t, \text{ dengan } f_t = X_t \theta_t. \quad (1)$$

- b. Melakukan regresi  $e_t$  pada  $f_t^k$  dengan  $f_t^k$  adalah transformasi nonlinear dari  $f_t$ , sehingga diperoleh model pada Persamaan (2),

$$e_t = a_2 f_t^2 + \dots + a_k f_t^k + v_t, \text{ untuk } k \geq 2 \quad (2)$$

kemudian dengan memasukkan Persamaan (2) ke Persamaan (1) diperoleh model alternative pada Persamaan (3),

$$Y_t = X_t \theta_t + a_2 f_t^2 + \dots + a_k f_t^k + v_t, \text{ untuk } k \geq 2. \quad (3)$$

Langkah-langkah pengujian linearitas menggunakan uji *Ramsey RESET* adalah sebagai berikut;

- a. Hipotesis

$H_0 : a_2 = \dots = a_k = 0$  (Data deret waktu tidak mengandung pola nonlinear)

$H_1 : a_k \neq 0$  (Data deret waktu mengandung pola nonlinear)

- b. Statistik Uji

$$RESET = \frac{[(\mathbf{e}_t^T \mathbf{e}_t - \mathbf{v}_t^T \mathbf{v}_t)/(k - 1)]}{[(\mathbf{v}_t^T \mathbf{v}_t)/(n - k)]} \quad (4)$$

dengan

$\mathbf{e}_t = (e_1, \dots, e_n)$  : vektor residual dari model linear pada Persamaan (1)

$\mathbf{v}_t = (v_1, \dots, v_n)$  : vektor residual dari model alternatif pada Persamaan (3)

$t$  : pengamatan waktu ke- $t$

$n$  : banyak pengamatan  
 $k$  : banyak variabel prediktor dan respon.

c. Kesimpulan

Berdasarkan statistik uji *RESET*, maka kriteria ujinya yaitu menolak  $H_0$  jika statistik uji  $RESET > F_{(k-1, n-k)}$ , artinya data deret waktu mengandung pola nonlinear.

**2.2. Model Support Vector Regression (SVR)**

*Support Vector Regression* (SVR) pada awalnya diusulkan oleh Vapnik [8]. Ide dasar dari SVR yaitu dengan menentukan set data yang dibagi menjadi data *training* dan data *testing*. Kemudian dari data *training* tersebut ditentukan suatu fungsi regresi dengan batasan deviasi tertentu sehingga dapat menghasilkan prediksi yang mendekati nilai aktual.

Misalkan terdapat  $n$  set data *training*,  $(\mathbf{x}_i, y_i)$  dengan  $i = 1, 2, \dots, n$ . Sedangkan  $\mathbf{x}_i = \{x_1, x_2, \dots, x_p\}^T \in R^n$  merupakan vektor dalam *input space* dan  $y_i = \{y_1, y_2, \dots, y_n\} \in R$  merupakan nilai *output* berdasarkan  $\mathbf{x}_i$  yang bersesuaian. Fungsi regresi dari metode SVR linear dituliskan pada Persamaan (5),

$$f(\mathbf{x}_i) = \mathbf{w} \cdot \mathbf{x}_i + b \tag{5}$$

dengan  $\mathbf{w}$  merupakan vektor bobot dan  $b$  merupakan bias. Koefisien  $\mathbf{w}$  dan  $b$  di sini berfungsi untuk meminimalkan fungsi resiko.

Menurut Haykin [9], koefisien  $\mathbf{w}$  dan  $b$  pada persamaan (5) diestimasi dengan cara meminimalkan fungsi resiko (*risk function*) pada Persamaan (6),

$$\min_{\mathbf{w}, \xi_i, \xi_i^*} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \tag{6}$$

dengan kendala

$$\begin{aligned} y_i &\leq f(\mathbf{x}_i) + \varepsilon + \xi_i \\ y_i &\geq f(\mathbf{x}_i) - \varepsilon - \xi_i^* \\ \xi_i, \xi_i^* &\geq 0 \end{aligned}$$

dengan  $\|\mathbf{w}\|$  adalah reguralisasi yang merupakan fungsi yang diminimumkan agar membuat fungsi setipis (*flat*) mungkin. Konstanta  $C$  (*Cost*)  $> 0$  adalah nilai tawar (*trade off*) antara ketipisan fungsi  $f$  dan batas atas deviasi lebih dari  $\varepsilon$  yang masih bisa ditoleransi [10]. Sedangkan  $\xi_i, \xi_i^*$  adalah variabel *slack* positif untuk mengatasi masalah pembatas yang tidak layak (*infeasible constrain*) dalam masalah optimasi.

Menurut Smola dan Schölkopf [10], untuk mendapatkan solusi persamaan (6) adalah menggunakan *lagrangian* dengan satu koefisien *lagrange* untuk setiap kendala sebagai ditunjukkan pada Persamaan (7).

$$L_p(\mathbf{w}, b, \xi_i, \xi_i^*, \alpha_i, \alpha_i^*, \eta_i, \eta_i^*) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i + f(\mathbf{x}_i) - y_i) - \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* - f(\mathbf{x}_i) + y_i) \quad (7)$$

dengan  $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$  adalah koefisien *lagrange*. Untuk meminimalkan *lagrangian* Persamaan (7), maka dilakukan turunan parsial  $L_p$  terhadap  $\mathbf{w}, b, \xi_i$ , dan  $\xi_i^*$  untuk kendala optimasi, sebagaimana ditunjukkan pada Persamaan (8)-(11),

$$\frac{\partial L_p}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \mathbf{x}_i \quad (8)$$

$$\frac{\partial L_p}{\partial b} = 0 \Rightarrow \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \quad (9)$$

$$\frac{\partial L_p}{\partial \xi_i} = 0 \Rightarrow \alpha_i + \eta_i = C \quad (10)$$

$$\frac{\partial L_p}{\partial \xi_i^*} = 0 \Rightarrow \alpha_i^* + \eta_i^* = C. \quad (11)$$

Selanjutnya Persamaan (8), (9), (10), dan (11) disubstitusikan ke dalam Persamaan (7) menghasilkan persamaan dual untuk problem optimisasi dari SVR sebagaimana ditunjukkan pada Persamaan (12),

$$L_d(\alpha_i, \alpha_i^*) = \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \mathbf{x}_i \mathbf{x}_j, \quad (12)$$

dengan kendala

$$\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0$$

$$0 \leq \alpha_i, \alpha_i^* \leq C.$$

Berdasarkan persamaan dual diperoleh solusi optimal dari kendala estimasi parameter  $\mathbf{w}$  dalam bentuk koefisien *lagrange*  $\alpha_i$  dan  $\alpha_i^*$  yang dapat ditulis sebagai,

$$\mathbf{w} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \mathbf{x}_i$$

sehingga fungsi regresi SVR untuk kasus linear adalah sebagai berikut,

$$f(\mathbf{x}_i) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) (\mathbf{x}_i, \mathbf{x}) + b$$

dengan nilai estimasi dari  $b$  yang diperoleh adalah  $b = y_i - \mathbf{w} \cdot \mathbf{x}_i - \varepsilon$  untuk  $0 \leq \alpha_i \leq C$  dan  $b = y_i - \mathbf{w} \cdot \mathbf{x}_i + \varepsilon$  untuk  $0 \leq \alpha_i^* \leq C$ .

Pada umumnya data dalam dunia nyata jarang yang bersifat *linear separable*, kebanyakan bersifat nonlinear. SVR untuk kasus nonlinear memberikan pendekatan alternatif dengan cara melakukan pemetaan data  $\mathbf{x}$  dari *input space* ke *feature space* dengan dimensi yang lebih tinggi melalui suatu fungsi  $\varphi$  sehingga  $\varphi: \mathbf{x} \mapsto \varphi(\mathbf{x})$  [11]. Dengan demikian, fungsi regresi SVR nonlinear dituliskan sebagai,

$$f(\mathbf{x}_i) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}) + b$$

Akan tetapi, kesulitan dalam pemetaannya adalah bahwa transformasi  $\varphi$  pada umumnya tidak diketahui dan sulit dipahami. Menurut Vapnik [8], masalah ini dapat diatasi dengan *kernel trick* yaitu perkalian skalar (*dot product*)  $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x})$  dalam *feature space* dapat digantikan fungsi kernel,

$$K(\mathbf{x}_i, \mathbf{x}) = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}).$$

Fungsi kernel tersebut mampu mendefinisikan secara implisit transformasi  $\varphi$ , sehingga fungsi regresi SVR dituliskan menjadi,

$$f(\mathbf{x}_i) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b.$$

Jenis fungsi kernel mempengaruhi parameter kernel SVR. Fungsi kernel harus diatur dengan benar karena dapat mempengaruhi akurasi regresi. Kernel yang digunakan pada model SVR dalam makalah ini adalah kernel *Radial Basis Function* (RBF) yaitu

$$K(\mathbf{x}_i, \mathbf{x}) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2).$$

Menurut Borivikov [12], kernel RBF memberikan akurasi yang paling baik dibandingkan kernel linear atau polinomial. Oleh karena itu, fungsi regresi SVR untuk kasus nonlinear dengan kernel RBF adalah sebagai berikut,

$$f(\mathbf{x}_i) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2) + b$$

dengan  $\gamma > 0$ . SVR dengan fungsi kernel RBF memiliki tiga parameter yang harus ditentukan yaitu  $C$  (*cost*),  $\gamma$  (*gamma*) dan  $\varepsilon$  (*epsilon*). Metode yang digunakan untuk mencari parameter optimal pada makalah ini adalah metode *grid search*.

### 2.3. Metode Grid Search

Secara umum, *grid search* merupakan kombinasi parameter yang diujikan pada suatu model SVM untuk mencari nilai *error* dalam klasifikasi [6]. Tujuannya adalah mengidentifikasi parameter optimal dalam data *training*, sehingga model tersebut

mampu secara akurat memprediksi data *testing*. Salah satu pendekatan dalam *grid search* adalah prosedur *cross validation*. Prosedur *cross validation* adalah membagi secara acak data *training* menjadi  $n$  subset yang berukuran sama. Satu subset digunakan sebagai data *testing* dan  $n-1$  subset digunakan sebagai data *training* [11].

Proses untuk melakukan *grid search* secara lengkap memerlukan waktu yang sangat lama sehingga disarankan untuk melakukan *grid search* dengan dua tahap yaitu *loose grid* dan *finer grid* [6]. *Loose grid* adalah tahapan dimana pemilihan nilai  $C$  dan  $\gamma$  dengan pangkat bilangan bulat, sedangkan *finer grid* adalah tahapan selanjutnya dari *loose grid* dimana saat didapat nilai  $C$  dan  $\gamma$  dengan *error* terendah. Dengan demikian, *finer grid* menggunakan nilai di sekitaran  $C$  dan  $\gamma$  tersebut. Sejalan dengan hal tersebut, pada makalah ini digunakan *grid search* dengan dua tahapan yaitu *loose grid* dan *finer grid* untuk menemukan nilai parameter optimal. Parameter optimal adalah parameter yang memiliki akurasi pemodelan terbaik dengan nilai *error* tertendah.

#### 2.4. Kriteria Pemilihan Model Terbaik

Pemilihan model terbaik pada makalah ini menggunakan nilai *Mean Absolute Percentage Error* (MAPE). MAPE adalah gambaran ukuran ketepatan estimasi model maupun ramalan. Nilai MAPE pada data *training* digunakan untuk melihat ketepatan estimasi model sedangkan nilai MAPE pada data *testing* menggambarkan ketepatan ramalan model. MAPE dihitung dengan membagi persentase dari rata-rata harga mutlak residual pada tiap periode dengan nilai aktual. Menurut Wei [13], MAPE dirumuskan mengikuti persamaan,

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Z_t - \hat{Z}_t}{Z_t} \right| \times 100\%$$

dengan  $Z_t$  adalah nilai observasi pada waktu ke- $t$ ,  $\hat{Z}_t$  adalah nilai peramalan pada waktu ke- $t$  dan  $n$  adalah banyaknya observasi. Model yang baik memiliki nilai MAPE sesuai kriteria pada Tabel 1 [14].

**Tabel 1.** Kriteria MAPE

MAPE (%)	Tingkat Akurasi
$0 \leq MAPE < 10$	Sangat akurat
$10 \leq MAPE < 20$	Akurat
$20 \leq MAPE < 50$	Kurang akurat
$MAPE \geq 50$	Tidak akurat

### 3. Hasil dan Pembahasan

#### 3.1. Data Penelitian

Data yang digunakan merupakan data sekunder yang diperoleh dari Badan Pusat statistik (BPS) Provinsi Sulawesi tengah dengan variabel penelitian adalah data jumlah penumpang berangkat melalui transportasi udara di Provinsi Sulawesi Tengah (Jiwa). Data tersebut merupakan data deret waktu bulanan periode Januari 2006 sampai dengan November 2019 tetapi minus data Bulan Agustus dan November 2018 dimana datanya tidak tersedia. Data penelitian dibagi menjadi dua yaitu data *training* yang

digunakan untuk pembentukan model dan data *testing* yang digunakan sebagai evaluasi model untuk peramalan. Dasar pembagian data ini berdasarkan kriteria Woschnagg dan Cipan [15] bahwa untuk data *testing* dapat sejumlah 10 persen dari total keseluruhan data sehingga data pada periode Januari 2011 sampai Desember 2018 (94 data) digunakan sebagai data *training* sedangkan data pada periode Januari 2019 sampai November 2019 (11 data) digunakan sebagai data *testing*.

### 3.2. Uji Linearitas Ramsey RESET

Pengujian linearitas menggunakan uji *Ramsey RESET* pada data jumlah penumpang berangkat melalui transportasi udara di Sulawesi Tengah selama periode Januari 2011 sampai dengan Desember 2018 menggunakan statistik uji pada persamaan (4) diperoleh hasil yang disajikan pada Tabel 2.

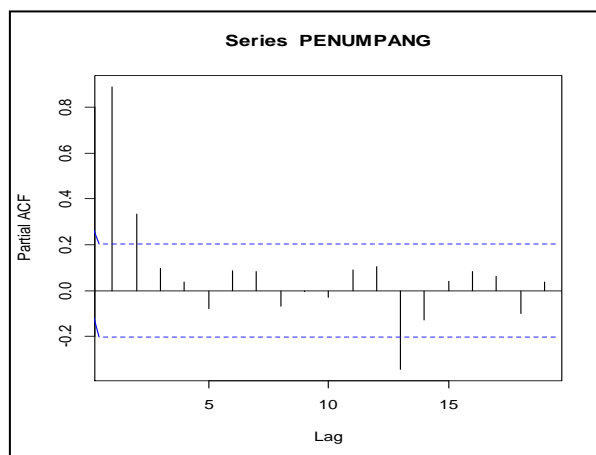
**Tabel 2.** Uji *Ramsey RESET* pada data jumlah penumpang berangkat melalui transportasi udara di Sulawesi Tengah

Data	Statistik uji ( <i>RESET</i> )	<i>p-value</i>
Penumpang Berangkat	98,145	3,996e-16

Berdasarkan Tabel 2 dapat dilihat bahwa nilai statistik uji *Ramsey RESET* lebih besar dari  $F_{(0,05;1;91)} = 3,9574$  atau nilai *p-value* lebih kecil dari  $\alpha = 5\%$  maka dapat disimpulkan  $H_0$  ditolak. Hal ini menunjukkan bahwa data jumlah penumpang berangkat melalui transportasi udara di Sulawesi Tengah mengandung pola nonlinear.

### 3.3. Penentuan Lag Berpengaruh Menggunakan Plot PACF

Sebelum masuk ke pemodelan menggunakan SVR, langkah pertama yang dilakukan adalah mengonversi data jumlah penumpang berangkat melalui transportasi udara di Sulawesi Tengah ke dalam bentuk *time lag* yang nantinya yang digunakan sebagai data input pada model SVR. Penentuan banyaknya *time lag* berdasarkan pada plot PACF. Plot PACF data jumlah penumpang berangkat melalui transportasi udara di Sulawesi Tengah ditampilkan pada Gambar 1.



**Gambar 1.** Plot PACF data jumlah penumpang berangkat melalui transportasi udara di Sulawesi Tengah



Berdasarkan plot PACF pada Gambar 1 dapat dilihat bahwa lag yang signifikan adalah lag 1, 2 dan 13 sehingga jumlah *time lag* untuk input data SVR adalah 3 lag yaitu lag 1, 2 dan 13.

### 3.4. Pemodelan Data Jumlah Penumpang Berangkat Melalui Transportasi Udara di Sulawesi Tengah Menggunakan SVR

Tahapan awal dalam membentuk model SVR adalah melakukan penentuan parameter optimal menggunakan metode *grid search*. Pada makalah ini dibatasi hanya menggunakan fungsi kernel *Radial Basic Function* (RBF). Pada kernel RBF terdapat parameter  $C, \gamma$  dan  $\epsilon$  yang harus ditentukan. Untuk mendapatkan nilai parameter yang optimal maka digunakan metode *grid search* dengan dua tahapan yaitu *loose grid* dan *finer grid*.

Tahapan awal metode *grid search* adalah menentukan rentang nilai parameter. Pada tahapan *loose grid* digunakan nilai parameter  $C$  dan  $\gamma$  dengan pangkat bilangan bulat. Adapun nilai  $\epsilon$  ditentukan terlebih dahulu. Rentang nilai parameter yang digunakan untuk tahapan *loose grid* disajikan pada Tabel 3.

**Tabel 3.** Rentang nilai parameter tahapan *Loose Grid* Metode *Grid Search*

Parameter	Rentang Nilai
$C$	$2^{-5}, 2^{-3}, \dots, 2^5, 2^7$
$\gamma$	$2^{-7}, 2^{-5}, \dots, 2^3, 2^5$
$\epsilon$	0,03; 0,04; ...; 0,07; 0,08

Rentang nilai parameter pada Tabel 3 digunakan untuk proses penentuan parameter model SVR. Melalui proses *grid search* pada tahapan *loose grid* diperoleh nilai optimal dari parameter model SVR adalah  $C = 2^{-1}$ ,  $\gamma = 2^{-1}$  dan  $\epsilon = 0,03$ . Setelah diperoleh parameter optimal model SVR pada tahapan *loose grid*, langkah selanjutnya dilakukan *grid search* menggunakan tahapan *finer grid*. *Finer grid* merupakan tahapan *grid search* menggunakan persekitaran nilai  $C, \gamma$  dan  $\epsilon$  yang diperoleh pada tahapan *loose grid*. Rentang nilai parameter yang digunakan pada tahapan *finer grid* disajikan pada Tabel 4.

**Tabel 4.** Rentang nilai parameter tahapan *Finer Grid* Metode *Grid Search*

Parameter	Rentang Nilai
$C$	$2^{-1,75}, 2^{-1,5}, 2^{-1,25}, 2^{-1}, 2^{-0,75}, 2^{-0,5}, 2^{-0,25}$
$\gamma$	$2^{-1,75}, 2^{-1,5}, 2^{-1,25}, 2^{-1}, 2^{-0,75}, 2^{-0,5}, 2^{-0,25}$
$\epsilon$	0,03

Melalui proses *grid search* pada tahapan *finer grid* diperoleh nilai optimal dari parameter model SVR adalah  $C = 2^{-1,5}$ ,  $\gamma = 2^{-1,25}$  dan  $\epsilon = 0,03$ . Nilai parameter yang diperoleh menggunakan metode *grid search* pada tahapan *finer grid* merupakan parameter optimal yang digunakan pada model SVR. Dengan demikian, model SVR terbaik yang diperoleh adalah model SVR menggunakan kernel *Radial Basic Function* (RBF) dengan parameternya adalah  $C = 2^{-1,5}$ ,  $\gamma = 2^{-1,25}$  dan  $\epsilon = 0,03$ .

Setelah diperoleh model SVR, langkah selanjutnya adalah melihat sejauh mana akurasi atau ketepatan model SVR untuk peramalan jumlah penumpang berangkat melalui transportasi udara di Sulawesi Tengah dengan melihat akurasi keakuratan model dengan melihat nilai *Mean Absolute Percentage Error* (MAPE). Hasil pemodelan tersebut akan diukur ketepatan pemodelannya dengan menggunakan nilai MAPE pada data *training*, sedangkan untuk mengukur ketepatan peramalannya digunakan nilai MAPE pada data *testing*. Hasil perhitungan nilai MAPE pada data *training* diperoleh nilai sebesar 7,28 persen. Menurut kriteria MAPE pada Tabel 1, bahwa nilai MAPE dibawah 10 persen menunjukkan bahwa model memiliki kemampuan peramalan sangat akurat.

### 3.5. Peramalan Jumlah Penumpang Berangkat Melalui Transportasi Udara di Sulawesi Tengah Pada Januari 2019 - November 2019

Berdasarkan nilai MAPE pada data *training* terbukti bahwa model SVR yang diperoleh memiliki kemampuan peramalan sangat akurat. Selanjutnya model SVR digunakan untuk melakukan peramalan pada data *testing*. Hasil peramalan jumlah penumpang berangkat melalui transportasi udara di Sulawesi Tengah pada Januari 2019 - November 2019 yang dapat dilihat pada Tabel 5.

**Tabel 5.** Peramalan jumlah penumpang berangkat melalui transportasi udara di Sulawesi Tengah pada Januari 2019 - November 2019

Bulan	Jumlah Penumpang (Jiwa)
Januari 2019	71.691
Februari 2019	74.115
Maret 2019	73.989
April 2019	77.346
Mei 2019	79.754
Juni 2019	77.741
Juli 2019	74.019
Agustus 2019	78.788
September 2019	76.849
Oktober 2019	75.530
November 2019	79.182

Hasil perhitungan nilai MAPE pada data *testing* model SVR diperoleh nilai MAPE sebesar 18,67 persen. Menurut kriteria MAPE bahwa nilai MAPE antara 10 sampai 20 persen menunjukkan bahwa model memiliki kemampuan peramalan akurat.

## 4. Kesimpulan

Model SVR terbaik yang diperoleh untuk data jumlah penumpang berangkat melalui transportasi udara di Sulawesi Tengah adalah model SVR menggunakan kernel *Radial Basic Function* (RBF) dengan parameternya adalah  $C = 2^{-1,5}$ ,  $\gamma = 2^{-1,25}$  dan  $\varepsilon = 0,03$ . Hasil pemodelan menggunakan model SVR menunjukkan kriteria peramalan yang akurat dilihat dari nilai MAPE data *training* sebesar 7,85 persen dan nilai MAPE data *testing* sebesar 18,24 persen.

## Referensi

- [1] G.A. Nasser, "Perkembangan Tingkat Penggunaan Sarana Akomodasi dan Transportasi," *Berita Resmi Statistik*, no. 42. BPS Provinsi Sulawesi Tengah, pp. 1-12, 2019.
- [2] G. E. Box and G. M. Jenkins, *Time Series Analysis Forecasting and Control*. California: Holden Day, 1976.
- [3] M. Khashei and M. Bijari, "An artificial neural network (p, d, q) model for timeseries forecasting," *Expert Syst. Appl.*, vol. 37, no. 1, pp. 479-489, 2010.
- [4] P. G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159-175, 2003.
- [5] D. Zhongxin, "Application of Support Vector Machine Regression in Ozone Forecasting," *Business, Econ. Financ. Sci. Manag.*, vol. 40, no. 4, pp. 359-365, 2019.
- [6] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," Taipei, 2016.
- [7] J. B. Ramsey, "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis," *J. R. Stat. Soc. Ser. B*, vol. 31, no. 2, pp. 350-371, 1969.
- [8] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 1995.
- [9] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. New Jersey: Pearson Education, Inc., 2009.
- [10] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 14, pp. 199-222, 2004.
- [11] B. Santosa, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis, Teori dan Aplikasi*. Yogyakarta: Graha Ilmu, 2007.
- [12] E. A. Borovikov, *An Evaluation of Support Vector Machines as a Pattern Recognition Tool*. California: arXiv preprint arXiv:1412.4186, 1999.
- [13] W. W. Wei, *Time Series Analysis: Univariate and Multivariate Methods*. New York: Pearson Education, Inc., 2006.
- [14] L. Latipah, S. Wahyuningsih, and S. Syaripuddin, "Peramalan Pendapatan Asli Daerah Provinsi Kalimantan Timur Menggunakan Model Grey-Markov (1,1)," *Jambura J. Math.*, vol. 1, no. 2, pp. 89-103, 2019.
- [15] E. Woschnagg and J. Cipan, *Evaluating Forecast Accuracy*. Austria: University of Vienna, 2004.