

PERAMALAN JUMLAH PENUMPANG DATANG MELALUI TRANSPORTASI UDARA DI SULAWESI TENGAH MENGGUNAKAN *SUPPORT VECTOR REGRESSION (SVR)*

Drajat Indra Purnama

Badan Pusat Statistik Kabupaten Parigi Moutong

Jalan Pakabata Kompleks Perkantoran Bambalemo, Parigi Moutong

drajatindrapurnama@bps.go.id

ABSTRACT

The number of air transportation passengers in Central Sulawesi shows an increase and decrease every month. For this reason, a forecasting method is needed to predict the number of air transportation passengers in the future. Because the pattern of data on the number of air transportation passengers in Central Sulawesi Province has a nonlinear data pattern, a forecasting method is needed that can overcome these problems where in this study using the SVR model. In this study, the SVR model uses the RBF kernel function to overcome nonlinear data patterns and uses the *grid search* method to obtain the optimal parameters of the model.

Keywords : Air transportation, *Grid Search*, RBF Kernel, SVR

ABSTRAK

Jumlah penumpang transportasi udara di Sulawesi Tengah menunjukkan kenaikan dan penurunan setiap bulan. Untuk itu diperlukan suatu metode peramalan untuk meramalkan jumlah penumpang transportasi udara di masa depan. Karena pola data jumlah penumpang transportasi udara di Provinsi Sulawesi Tengah memiliki pola data nonlinear, maka diperlukan metode peramalan yang dapat mengatasi permasalahan tersebut dimana dalam penelitian ini menggunakan model SVR. Penelitian ini, model SVR menggunakan fungsi kernel RBF untuk mengatasi pola data nonlinear dan menggunakan metode *grid search* untuk memperoleh parameter optimal model.

Kata kunci : Transportasi Udara, *Grid Search*, Kernel RBF, SVR

I. PENDAHULUAN

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 serta nyaman. Hal ini juga terjadi di Provinsi Sulawesi Tengah dimana pesawat terbang menjadi sarana transportasi yang disukai masyarakat untuk berpergian jauh baik di dalam area Sulawesi Tengah maupun keluar area Sulawesi Tengah. Menurut data Badan Pusat Statistik Sulawesi Tengah 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 kenaikan dan penurunan setiap bulannya khususnya jumlah penumpang yang datang. 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.

Model peramalan data deret waktu yang bisa digunakan adalah model *Autoregressive Integrated Moving Average* (ARIMA) yang diperkenalkan pertama kali oleh Box dan Jenkins (1976) [1]. ARIMA sangat baik ketepatannya untuk peramalan jangka pendek dan untuk data runtun waktu non stasioner pada saat linear karena prinsip utama dalam model ARIMA adalah bahwa model ini dibangun oleh proses linear. Akan tetapi, ARIMA memiliki kekurangan untuk sebagian besar permasalahan yang bersifat nonlinear. ARIMA cenderung mengalami penurunan tingkat keakuratan apabila digunakan pada data yang mengandung pola nonlinear [9]. Berbeda dengan model ARIMA, model *Support Vector Regression* (SVR) merupakan model peramalan yang dapat digunakan untuk memprediksi data deret waktu nonlinear. Menurut Ding [2], keunggulan SVR adalah kemampuan untuk mengatasi masalah data nonlinear dengan trik kernel dan dapat mengatasi masalah *overfitting* dimana model yang dihasilkan hanya menghasilkan model yang baik untuk data *training* dan tidak untuk data *testing*.

II. TINJAUAN PUSTAKA

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 [5]. Jika terdapat variabel bebas X_t dan variabel prediktor Y_t , prosedur uji *Ramsey RESET* dapat dijelaskan sebagai :

1. Melakukan regresi Y_t pada X_t sehingga diperoleh model linear

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

2. Melakukan regresi e_t pada f_t^k dengan f_t^k adalah transformasi nonlinear dari f_t , sehingga diperoleh

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

Kemudian dengan memasukkan persamaan (2) ke persamaan (1) diperoleh model alternatif

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

Uji statistik dilakukan untuk menguji hipotesis $H_0 : a_2 = \dots = a_k = 0$ (Data deret waktu tidak mengandung pola nonlinear). Jika $\mathbf{e}_t = (e_1, \dots, e_n)$ adalah vektor residual dari model linear pada persamaan (1) dan $\mathbf{v}_t = (v_1, \dots, v_n)$ adalah vektor residual dari model alternatif pada persamaan (3), maka statistik ujinya adalah

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

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)

SVR pada awalnya diusulkan oleh Vapnik [8]. SVR merupakan penerapan SVM yang digunakan untuk kasus regresi yang outputnya berupa bilangan riil atau kontinu. 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 sebagai

$$f(\mathbf{x}_i) = \mathbf{w} \cdot \mathbf{x}_i + b \quad (4)$$

dengan \mathbf{w} merupakan vektor bobot dan b merupakan bias. Koefisien \mathbf{w} dan b di sini berfungsi untuk meminimalkan fungsi resiko. Menurut Haykin [3], koefisien \mathbf{w} dan b pada persamaan (4) diestimasi dengan cara meminimalkan fungsi resiko (*risk function*) sebagai berikut

$$\min_{\mathbf{w}, \xi_i, \xi_i^*} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (5)$$

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*) mugkin. Konstanta C (*Cost*) > 0 adalah nilai tawar (*trade off*) antara ketipisan fungsi f dan batas atas deviasi lebih dari ε yang masih bisa ditoleransi [7]. Sedangkan ξ_i, ξ_i^* adalah variabel *slack* positif untuk mengatasi masalah pembatas yang tidak layak (*infeasible constrain*) dalam masalah optimasi.

Menurut Smola dan Scholkopf [7], untuk mendapatkan solusi persamaan (5) adalah menggunakan *lagrangian* dengan satu koefisien *lagrange* untuk setiap kendala sebagai berikut

$$L_p(\mathbf{w}, b, \xi_i, \xi_i^*, \alpha_i, \alpha_i^*, \eta_i, \eta_i^*) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i + f(\mathbf{x}_i) - y_i) - \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* - f(\mathbf{x}_i) + y_i) \quad (6)$$

dengan $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$ adalah koefisien *lagrange*. Untuk meminimumalkan *lagrangian* persamaan (6) maka dilakukan turunan parsial L_p terhadap \mathbf{w}, b, ξ_i , dan ξ_i^* untuk kendala optimasi sebagai berikut

$$\frac{\partial L_p}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \mathbf{x}_i \quad (7)$$

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

$$\frac{\partial L_p}{\partial \xi_i} = 0 \Rightarrow \alpha_i + \eta_i = C \quad (9)$$

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

Selanjutnya persamaan (7), (8), (9), dan (10) disubstitusikan ke dalam persamaan (6) menghasilkan persamaan dual untuk problem optimisasi dari SVR sebagai berikut :

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

dengan kendala

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

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

Berdasarkan persamaan dual pada persamaan diperoleh solusi optimal dari kendala estimasi parameter \mathbf{w} dalam bentuk koefisien *lagrange* α_i dan α_i^* yang dapat ditulis kembali 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 φ sehingga $\varphi: \mathbf{x} \mapsto \varphi(\mathbf{x})$ [6]. Sehingga fungsi regresi SVR nonlinear dituliskan sebagai

$$f(\mathbf{x}_i) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}) + b \quad (11)$$

Akan tetapi, kesulitan dalam pemetaannya adalah bahwa transformasi φ 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

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

dengan fungsi kernel tersebut mampu mendefinisikan secara implisit transformasi φ . Sehingga fungsi regresi pada persamaan (11) 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. Dalam penelitian ini, kernel yang digunakan pada model SVR adalah kernel *Radial Basis Function* (RBF) yaitu

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

Sehingga 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*), γ (*gamma*) dan ε (*epsilon*). Metode yang digunakan untuk mencari parameter optimal pada penelitian 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 [4]. 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* [6].

Proses untuk melakukan *grid search* secara lengkap memerlukan waktu yang sangat lama sehingga Hsu [4], menyarankan untuk melakukan *grid search* dengan dua tahap yaitu *loose grid* dan *finer grid*. *Loose grid* adalah tahapan dimana pemilihan nilai C dan γ dengan pangkat bilangan bulat, sedangkan *finer grid* adalah tahapan selanjutnya dari *loose grid* dimana

saat didapat nilai C dan γ dengan *error* terendah, maka *finer grid* menggunakan nilai di sekitaran C dan γ tersebut. Sejalan dengan hal tersebut, pada penelitian ini akan digunakan *grid search* dengan dua tahapan yaitu *loose grid* dan *finer grid* untuk menemukan nilai parameter optimal.

III. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Sulawesi Tengah yang merupakan data jumlah penumpang datang melalui transportasi udara di Sulawesi Tengah (Jiwa). Data merupakan data deret waktu bulanan periode Januari 2006 sampai dengan November 2019 tetapi minus data Bulan Agustus dan November 2018 dimana datanya tidak tersedia. Metode analisis yang digunakan dalam penelitian ini adalah menggunakan model SVR. Adapun prosedur dalam penelitian ini adalah Sebagai berikut :

1. Melakukan uji linearitas menggunakan uji *Ramsey RESET*
2. Menentukan *lag* berpengaruh menggunakan plot PACF.
3. Membentuk model SVR
 - a. Menentukan fungsi kernel
 - b. Menentukan rentang nilai parameter.
 - c. Menentukan parameter optimal menggunakan metode *grid search*.
 - d. Memperoleh model SVR terbaik.
4. Mengevaluasi model SVR

Pengolahan data pada penelitian ini menggunakan bantuan *software R* seri 3.6.1.

IV. HASIL DAN PEMBAHASAN

Tahapan awal penelitian adalah penentuan data *training* yang digunakan untuk pembentukan model dan data *testing* yang digunakan sebagai evaluasi model untuk peramalan. Penelitian ini menggunakan data jumlah penumpang datang melalui transportasi udara di Sulawesi Tengah (Jiwa) pada periode Januari 2011 sampai November 2019 (105 data). 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*.

3.1. Pengujian Linearitas

Pengujian linearitas menggunakan uji *Ramsey RESET* pada data jumlah penumpang datang melalui transportasi udara di Sulawesi Tengah selama periode Januari 2011 sampai dengan Desember 2018 diperoleh hasil yang disajikan pada Tabel 1.

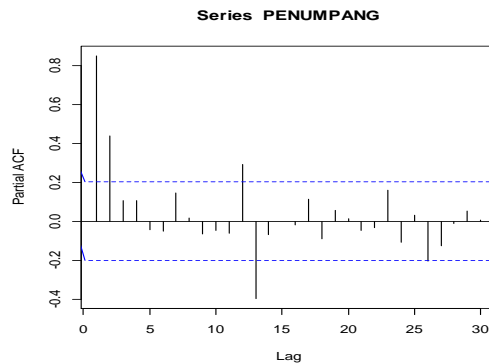
Tabel 1 : Uji *Ramsey RESET* Pada Data Jumlah Penumpang Datang Melalui Tranportasi Udara di Sulawesi Tengah

| Data | Statistik Uji (<i>RESET</i>) | <i>p-value</i> |
|---------------------|-----------------------------------|----------------|
| Penumpang Datang | 161,76 | 2,2e- 16 |

Berdasarkan Tabel 1 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 yang berarti bahwa data jumlah penumpang datang melalui tranportasi udara di Sulawesi Tengah mengandung pola nonlinear.

3.2. Penentuan *Lag* Berpengaruh Menggunakan Plot PACF

Sebelum masuk ke pemodelan menggunakan SVR, langkah yang dilakukan mengkonversi data jumlah penumpang datang melalui tranportasi udara di Sulawesi Tengah ke dalam bentuk *time lag* yang nantinya sebagai data input dalam model SVR. Penentuan banyaknya *time lag* berdasarkan pada plot PACF. Plot PACF data jumlah penumpang datang melalui tranportasi udara di Sulawesi Tengah dapat dilihat pada Gambar 1.



Gambar 1 : Plot PACF Data Jumlah Penumpang datang melalui tranportasi udara di Sulawesi Tengah

Berdasarkan plot PACF pada Gambar 1 dapat dilihat bahwa lag yang signifikan adalah lag 1, 2, 12 dan 13 sehingga jumlah *time lag* untuk input data SVR adalah 4 lag yaitu lag 1, 2, 12 dan 13.

3.3. Pemodelan data Jumlah penumpang datang melalui tranportasi udara di Sulawesi Tengah Menggunakan SVR

Tahapan awal dalam membentuk model SVR adalah melakukan penentuan parameter optimal menggunakan metode *grid search*. Pada penelitian ini dibatasi hanya menggunakan

fungsi kernel *Radial Basic Function* (RBF). Pada kernel RBF terdapat parameter C , γ dan ε yang harus ditentukan. Untuk mendapatkan nilai parameter yang optimal maka pada penelitian ini 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 γ dengan pangkat bilangan bulat. Sedangkan nilai ε ditentukan terlebih dahulu. Rentang nilai parameter yang digunakan untuk tahapan *loose grid* pada penelitian ini dapat dilihat pada Tabel 2.

Tabel 2 : Rentang Nilai Parameter Tahapan *Loose Grid* Metode *Grid Search*

| Parameter | Rentang Nilai |
|---------------|-----------------------------------|
| C | $2^{-3}, 2^{-1}, \dots, 2^7, 2^9$ |
| γ | $2^{-9}, 2^{-7}, \dots, 2^1, 2^3$ |
| ε | 0,01; 0,02; ...; 0,09; 0,1 |

Rentang nilai parameter pada Tabel 2 digunakan untuk proses penentuan parameter model SVR. Melalui proses *grid search* pada tahapan *loose grid* diperoleh nilai optimal dari parameter model SVR adalah $C = 2^3$, $\gamma = 2^{-3}$ dan $\varepsilon = 0,1$. 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 , γ dan ε yang diperoleh pada tahapan *loose grid*.

Tabel 3 : Rentang Nilai Parameter Tahapan *Finer Grid* Metode *Grid Search*

| Parameter | Rentang Nilai |
|---------------|--|
| C | $2^{2,25}, 2^{2,5}, 2^{2,75}, 2^3, 2^{3,25}, 2^{3,5}, 2^{3,75}$ |
| γ | $2^{-3,75}, 2^{-3,5}, 2^{-3,25}, 2^{-3}, 2^{-2,75}, 2^{-2,5}, 2^{-2,25}$ |
| ε | 0,1 |

Rentang nilai parameter yang digunakan pada tahapan *finer grid* dapat dilihat pada Tabel 3. Melalui proses *grid search* pada tahapan *finer grid* diperoleh nilai optimal dari parameter model SVR adalah $C = 2^{2,25}$, $\gamma = 2^{-3}$ dan $\varepsilon = 0,1$. Nilai parameter yang diperoleh menggunakan metode *grid search* pada tahapan *finer grid* merupakan parameter optimal yang digunakan pada model SVR. Sehingga model SVR terbaik yang diperoleh adalah model SVR menggunakan kernel *Radial Basic Function* (RBF) dengan parameternya adalah $C = 2^{2,25}$, $\gamma = 2^{-3}$ dan $\varepsilon = 0,1$.

Setelah diperoleh model SVR, langkah selanjutnya adalah melihat sejauh mana akurasi atau ketepatan model SVR untuk peramalan jumlah penumpang datang 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 5,01 persen. Menurut kriteria MAPE bahwa nilai MAPE dibawah 10 persen menunjukkan bahwa model memiliki kemampuan peramalan sangat baik. Sedangkan, hasil perhitungan nilai MAPE pada data *testing* model SVR diperoleh nilai MAPE sebesar 16,65 persen. Menurut kriteria MAPE bahwa nilai MAPE antara 10 sampai 20 persen menunjukkan bahwa model memiliki kemampuan peramalan baik.

V. KESIMPULAN

Model *Support Vector Regression* (SVR) merupakan salah satu model data deret waktu yang dapat digunakan pada data yang memiliki pola data nonlinear. Penelitian ini menerapkan model SVR pada data peramalan jumlah penumpang datang melalui transportasi udara di Sulawesi Tengah yang melalui uji Ramsey RESET memiliki pola data nonlinear. Model SVR terbaik yang diperoleh adalah model SVR menggunakan kernel *Radial Basic Function* (RBF) dengan parameternya adalah $C = 2^{2,25}$, $\gamma = 2^{-3}$ dan $\varepsilon = 0,1$. Hasil pemodelan menggunakan model SVR menunjukkan kriteria peramalan yang baik dilihat dari nilai MAPE data training sebesar 5,01 persen dan nilai MAPE data testing sebesar 16,65 persen.

DAFTAR PUSTAKA

- [1] Box, G.E., and Jenkins, G.M, *Time Series Analysis Forecasting and Control*, Holden Day, 1976, California.
- [2] Ding, Z. *Application od Support Vector Machine Regression in Stock Price Forecasting. Business, Economics, and Financial Sci., Manag., AISC 143*, 2012, 359-365.
- [3] Haykin, S., *Neural Networks and Learning Machines Third Edition*, Pearson Education, 2009, New Jersey.
- [4] Hsu, C. W, Chang, C.C., and Lin, C.J, *A Practical Guide to Support Vector Classificaton*, Department of Computer Science National Taiwan University, 2016, Taipei.
- [5] Ramsey, J. B, *Tests for Specification Errors in Classical Linear Least Squares Regression Analysis. Journal of the Royal Statistical Society, Series B*, 1969, 31 (2): 350–371.
- [6] Santosa, B, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis, Teori dan Aplikasi*, Graha Ilmu, 2007, Yogyakarta.
- [7] Smola, A.J. and Scholkopf, B, *A Tutorial On Support Vector Regression. Statistic and Computing 14*, 2004, 199-222.
- [8] Vapnik, V.N, *The Nature of Statistical Learning Theory*, Springer, 1995, New York.
- [9] Zhang, G.P, *Time Series Forecasting using a Hybrid ARIMA dan Neural Network Model. Neurocomputing 50*, 2003,159-175.