

## Support Vector Regression (SVR) Model for Forecasting Number of Passengers on Domestic Flights at Sultan Hasanudin Airport Makassar

### Model *Support Vector Regression* (SVR) untuk Peramalan Jumlah Penumpang Penerbangan Domestik di Bandara Sultan Hasanudin Makassar

Drajat Indra Purnama\*<sup>1</sup>, Siti Setianingsih\*<sup>2</sup>

#### Abstract

Sultan Hasanudin Airport is one of the largest airports in Indonesia, located in Makassar City. Its strategic location is the entrance of eastern Indonesia because it is a transit airport to other eastern regions of Indonesia. The number of airplane passengers at Sultan Hasanudin Airport has increased and decreased each time depending on certain moments. The increase in the number of passengers is closely related to the moments of religious holidays or year-end holidays. Whereas the decrease in the number of passengers was greatly influenced by the policy of rising plane ticket prices some time ago. Estimated number of passengers every month is needed in planning and making appropriate decisions from the government relating to fluctuations in the number of domestic flight passengers at Sultan Hasanudin Airport. Therefore, accurate forecasting techniques are needed to predict the number of passengers in the future. Because the data pattern of domestic flight passengers at Sultan Hasanudin Airport is not stationary, the ARIMA model can be used. However, the data on the number of passengers has a nonlinear data pattern, so we need a method that can overcome these problems. In this study the SVR model is used to overcome nonlinear patterns in the data. Compared to the ARIMA model, SVR has the advantage because it does not require stationary data assumptions as in ARIMA. The results of forecasting data on the number of domestic flight passengers at Sultan Hasanudin Airport using SVR show better accuracy or accuracy compared to the ARIMA model because it has a smaller MAPE value.

**Keywords:** ARIMA, SVR, Sultan Hasanudin Airport

#### Abstrak

Bandara Sultan Hasanudin merupakan salah satu bandara terbesar di Indonesia yang terletak di Kota Makassar. Letaknya yang strategis merupakan pintu masuk kawasan timur Indonesia karena menjadi bandara transit ke wilayah timur Indonesia lainnya. Jumlah penumpang pesawat di Bandara Sultan Hasanudin mengalami kenaikan dan penurunan setiap waktu tergantung pada momen-momen tertentu. Kenaikan jumlah penumpang erat kaitannya pada momen hari besar keagamaan ataupun liburan akhir tahun. Sedangkan penurunan jumlah penumpang sangat dipengaruhi kebijakan kenaikan harga tiket pesawat beberapa waktu yang lalu. Perkiraan jumlah penumpang setiap bulan sangat dibutuhkan dalam perencanaan dan pengambilan keputusan yang tepat dari pemerintah berkaitan dengan fluktuasi jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin. Oleh karena itu, dibutuhkan

\* *BPS Kabupaten Parigi Moutong, Sulawesi Tengah*

*Email : <sup>1</sup>drajatindrapurnama@bps.go.id, <sup>2</sup>Sitisetianingsih@gmail.com*

## Drajat Indra Purnama

teknik peramalan yang tepat untuk meramalkan jumlah penumpang di masa depan. Karena pola data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin yang tidak stasioner maka dapat digunakan model ARIMA. Akan tetapi data jumlah penumpang tersebut memiliki pola data nonlinear sehingga diperlukan metode yang dapat mengatasi permasalahan tersebut. Pada penelitian ini digunakan model SVR untuk mengatasi pola nonlinear pada data. Dibandingkan model ARIMA, SVR memiliki keunggulan karena tidak memerlukan asumsi data stasioner seperti pada ARIMA. Hasil peramalan data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin menggunakan SVR menunjukkan ketepatan atau akurasi yang lebih baik dibandingkan dengan model ARIMA karena memiliki nilai MAPE yang lebih kecil.

**Kata kunci:** ARIMA, SVR, Bandara Sultan Hasanudin

### 1. PENDAHULUAN

Transportasi telah menjadi kebutuhan dasar masyarakat oleh karenanya dari waktu ke waktu terjadi perkembangan sarana dan alat transportasi. 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. Banyaknya pilihan maskapai bagi masyarakat untuk bepergian menggunakan pesawat terbang membuat sarana transportasi ini makin diminati. Selain itu, kenyamanan dan akses ke Bandara yang semakin baik membuat pesawat terbang tetap menjadi favorit masyarakat di Indonesia.

Bandara Sultan Hasanuddin sebagai pintu masuk utama Kota Makassar memegang peranan penting dalam perkembangan Kota Makassar. Selain itu, Bandara ini juga menjadi pintu gerbang udara di kawasan timur Indonesia karena sebagai bandara transit yang diarahkan untuk turut mendukung dan mengembangkan pariwisata, mobilisasi arus penumpang serta berpartisipasi dalam perdagangan dan industri di kawasan timur Indonesia. Untuk itu, penting untuk diketahui gambaran fluktuasi jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin di masa yang akan datang.

Secara umum pergerakan jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin cenderung tidak stasioner dalam rata-rata karena dalam beberapa waktu tertentu pergerakannya membentuk pola tren naik. Selain itu, jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin memiliki pola yang tidak teratur dengan besaran kenaikan dan penurunan yang bervariasi. Hal ini membuat pergerakan jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin memiliki hubungan yang sangat kompleks dan nonlinear. Secara umum, untuk meramalkan fluktuasi jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin pada periode mendatang dapat dilakukan dengan analisis deret waktu.

Analisis data deret waktu yang umum digunakan dalam peramalan data deret waktu adalah model *Autoregressive Integrated Moving Average* (ARIMA). Model ARIMA dapat digunakan untuk memprediksi data deret waktu nonstasioner pada saat linear. Kelebihan ARIMA adalah cukup fleksibel dalam memodelkan sebagian besar pola deret waktu karena dapat menganalisis situasi yang acak, trend, musiman, bahkan bersifat siklis. Namun demikian, ARIMA memiliki kekurangan yaitu kurang cocok untuk sebagian besar permasalahan yang bersifat non linear. ARIMA cenderung mengalami penurunan tingkat keakuratan apabila digunakan pada data yang mengandung komponen non linear [10].

## Drajat Indra Purnama

Model *Support Vector Regression* (SVR) merupakan model peramalan yang dapat digunakan untuk memprediksi data deret waktu nonlinear. SVR adalah modifikasi *Support Vector Machine* (SVM) yang digunakan untuk pendekatan regresi [2]. Keunggulan SVR adalah kemampuan untuk mengatasi masalah data nonlinear dengan trik kernel. Dalam SVR terdapat parameter *epsilon* ( $\epsilon$ ), *cost* ( $C$ ), dan *gamma* ( $\gamma$ ) yang ditentukan nilai optimalnya menggunakan metode *grid search*. *Grid search* merupakan kombinasi parameter yang diujikan pada suatu model SVR untuk mencari nilai *error* dalam klasifikasi [5].

Untuk itulah pada penelitian ini, digunakan model ARIMA dan model SVR untuk peramalan data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin yang mempunyai pola nonlinear. Harapannya bahwa model SVR dengan menggunakan trik kernel mampu mengatasi pola data nonlinear. Pada akhirnya, hasil peramalan menggunakan ARIMA dan SVR dibandingkan keakuratan peramalannya dengan melihat nilai *Mean Absolute Percentage Error* (MAPE). Model yang terbaik adalah model yang memiliki nilai MAPE lebih kecil.

## 2. METODE PENELITIAN

### 2.1 Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) yang merupakan data Jumlah penumpang keberangkatan penerbangan domestik di Bandara Sultan Hasanudin (Orang) [1]. Data merupakan data deret waktu periode Januari 2006 sampai dengan November 2019 atau sebanyak 167 data deret waktu.

### 2.2 Metode Analisis Data

Metode analisis yang digunakan dalam penelitian ini adalah menggunakan model ARIMA dan SVR. Adapun tahapan dalam melakukan pemodelan menggunakan ARIMA dan SVR sehingga pada akhirnya dapat membandingkan kedua model tersebut adalah sebagai berikut :

1. Melakukan uji linearitas data.
2. Melakukan uji stasioneritas data dan melakukan proses *differencing*.
3. Melakukan pemodelan jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin menggunakan model ARIMA.
  - a. Identifikasi model ARIMA dengan melihat plot ACF dan PACF.
  - b. Estimasi parameter model ARIMA menggunakan metode *Maximum Likelihood Estimation* (MLE).
  - c. Pengujian signifikansi parameter model ARIMA.
  - d. Pengujian diagnostik model ARIMA.
  - e. Pemilihan model ARIMA terbaik.
4. Melakukan pemodelan jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin menggunakan model SVR.
  - a. Menentukan *lag* berpengaruh menggunakan plot PACF.
  - b. Menentukan rentang nilai parameter.
  - c. Menentukan parameter optimal menggunakan metode *grid search*.
  - d. Memperoleh model SVR terbaik.
5. Membandingkan nilai MAPE dari model ARIMA dan SVR.

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

### 2.3 Uji Linearitas

Salah satu uji untuk mengetahui linear atau nonlinear suatu data deret waktu dapat menggunakan uji *Ramsey RESET* yang pertama kali diperkenalkan oleh Ramsey [6]. 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.

### 2.4 Stasioneritas Data

Data stasioner adalah data deret waktu yang tidak memiliki tren, pola musiman serta nilai rata-rata dan variansnya konstan atau homogen dari waktu ke waktu [3]. Data yang stasioner adalah data yang variansnya tidak terlalu besar dan mempunyai kecenderungan untuk mendekati nilai rata-ratanya. Pengujian stasioneritas dalam rata-rata data deret waktu dapat menggunakan uji *Augmented Dickey Fuller* (ADF) dengan langkah-langkah sebagai berikut

1. Hipotesis

$H_0 : \gamma = 1$  (Data tidak stasioner dalam rata-rata)

$H_1 : \gamma < 1$  (Data stasioner dalam rata-rata)

2. Statistik uji

$$T = \frac{\hat{\gamma} - 1}{se(\hat{\gamma})} \sim t_{(n-p-1)}$$

dengan  $se(\hat{\gamma})$  adalah standar *error* dari  $\hat{\gamma}$  dan  $p$  adalah banyaknya parameter

3. Kriteria uji

Tolak  $H_0$  jika nilai  $T > t_{tabel}$ , artinya data stasioner dalam rata-rata.

Ketidastasioneran data dalam rata-rata dapat diatasi dengan proses *differencing*. Proses *differencing* dapat dilakukan beberapa periode sampai data stasioner yaitu dengan cara mengurangi nilai data pada suatu periode dengan nilai data periode sebelumnya. Apabila data deret waktu tidak stasioner dalam rata-rata, maka data tersebut dapat dibuat lebih mendekati stasioner dengan melakukan proses *differencing* orde ke- $d$  dirumuskan sebagai

$$Z_t^d = (1 - B)^d Z_t$$

dengan

- $Z_t$  : observasi waktu  $t$   
 $Z'_t$  : variabel hasil *differencing*  $Z_t$   
 $d$  : orde *differencing*.  
 $B$  : operator *Backshift*  
 $(1 - B)^d$  : *differencing* dengan orde  $d$

## 2.5 Model Autoregressive Integrated Moving Average (ARIMA)

Model *Autoregressive Integrated Moving Average* (ARIMA) adalah model univariat yang umum digunakan dalam peramalan data deret waktu. Model ARIMA berasal dari model *Autoregressive Moving Average* (ARMA) yang merupakan gabungan antara model *Autoregressive* (AR) dan model *Moving Average* (MA). Model AR menyatakan bahwa pengamatan waktu sekarang dipengaruhi oleh pengamatan waktu sebelumnya dan kesalahan sekarang. Model MA menyatakan bahwa pengamatan waktu sekarang dipengaruhi oleh kesalahan sekarang dan kesalahan waktu sebelumnya. Prosedur ARIMA adalah melakukan *differencing* pada data deret waktu nonstasioner terlebih dahulu untuk menghasilkan data deret waktu yang stasioner, dan selanjutnya melakukan proses ARMA pada data hasil *differencing* tersebut.

Jika data stasioner pada proses *differencing* sebanyak  $d$  kali, dengan model dasar ARMA( $p, q$ ), maka model yang terbentuk menjadi ARIMA ( $p, d, q$ ). Menurut Wei [9], model ARIMA ( $p, d, q$ ) dituliskan sebagai

$$\phi_p(B)(1 - B)^d Z_t = \theta_q(B)\varepsilon_t$$

dengan

- $\phi_p$  : parameter AR orde  $p$   
 $\theta_q$  : parameter MA orde  $q$   
 $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$   
 $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$   
 $\varepsilon_t$  : residual yang menyebar normal dan independen  $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$

## 2.6 Model Support Vector Regression (SVR)

Ide dasar SVR adalah menentukan set data yang dibagi menjadi data *training* dan data *testing*. Kemudian dari data *training* 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 metode SVR linear dapat dituliskan sebagai berikut

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

dengan  $\mathbf{w}$  merupakan vektor bobot dan  $b$  merupakan bias.

Menurut Haykin [4], koefisien  $\mathbf{w}$  dan  $b$  pada diestimasi dengan cara meminimalkan fungsi resiko (*risk function*) sebagai berikut

$$R(f(\mathbf{x}_i)) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n E_{\varepsilon} (y_i - f(\mathbf{x}_i))$$

atau dituliskan

## Drajat Indra Purnama

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n E_{\varepsilon}(y_i - f(\mathbf{x}_i)) \quad (4)$$

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 [7]. Sedangkan  $E_{\varepsilon}$  adalah  $\varepsilon$ -insensitive loss function yang didefinisikan sebagai

$$E_{\varepsilon}(y_i - f(\mathbf{x}_i)) = \begin{cases} |y_i - f(\mathbf{x}_i)| - \varepsilon, & \text{untuk } |y_i - f(\mathbf{x}_i)| \geq \varepsilon \\ 0, & \text{lainnya.} \end{cases}$$

dengan kendala

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

Fungsi regresi  $f$  diasumsikan merupakan fungsi yang dapat mengaproksimasi semua titik  $(\mathbf{x}_i, y_i)$  dengan presisi  $\varepsilon$ . Semua titik yang berada dalam rentang  $f \pm \varepsilon$  disebut *feasible*. Sedangkan titik yang berada diluar rentang  $f \pm \varepsilon$  disebut *infeasible*. Titik *infeasible* ini ditambahkan dengan variabel *slack* positif  $\xi_i, \xi_i^*$  untuk mengatasi masalah pembatas yang tidak layak (*infeasible constrain*) dalam masalah optimasi. Sehingga persamaan (4) dapat ditranformasikan ke dalam bentuk

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

dengan kendala

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

Menurut Smola dan Scholkopf [7], persamaan (5) menunjukkan bahwa penentuan parameter  $\mathbf{w}$  dan  $b$  menjadi masalah optimasi menggunakan *lagrangian* sebagai berikut

$$\begin{aligned} L_p = & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i + f(\mathbf{x}_i) - y_i) \\ & - \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* - f(\mathbf{x}_i) + y_i) \end{aligned}$$

dengan kendala dengan  $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$  adalah koefisien *lagrange*. *Lagrangian* menghasilkan solusi optimal parameter  $\mathbf{w}$  dalam bentuk koefisien *lagrange*  $\alpha_i$  dan  $\alpha_i^*$  sebagai berikut

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

Sedangkan estimasi akhir dari  $b$  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  $0 \leq \alpha_i^* \leq C$ . Sehingga fungsi regresi SVR dapat dituliskan sebagai

## Drajat Indra Purnama

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

SVR dapat digunakan untuk kasus nonlinear dengan pendekatan alternatif melalui pemetaan data  $\mathbf{x}$  dari *input space* ke *feature space* dengan dimensi yang lebih tinggi melalui suatu fungsi  $\varphi$  sehingga  $\varphi: \mathbf{x} \mapsto \varphi(\mathbf{x})$ . Sehingga fungsi regresi SVR nonlinear dituliskan sebagai

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

Transformasi  $\varphi$  pada umumnya tidak diketahui dan sulit dipahami. Menurut [8], masalah ini dapat diatasi dengan *kernel trick* yaitu  $K(\mathbf{x}_i, \mathbf{x}) = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x})$ . 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 nonlinear dengan kernel RBF adalah sebagai berikut

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

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 penelitian ini adalah metode *grid search*.

### 2.7 Metode Grid Search

Secara umum, *grid search* merupakan kombinasi parameter yang diujikan pada suatu model SVM untuk mencari nilai *error* dalam klasifikasi [5]. 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*. Pasangan parameter yang menghasilkan akurasi terbaik yang didapatkan dari uji *cross validation* merupakan parameter yang optimal.

Proses *grid search* secara lengkap memerlukan waktu yang sangat lama sehingga Hsu *et al.* [5], menyarankan untuk melakukan *grid search* dengan dua tahap yaitu *loose grid* dan *finer grid*. *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, maka *finer grid* menggunakan persekitaran nilai  $C$  dan  $\gamma$  tersebut.

## 3. 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 penerbangan domestik di Bandara Sultan Hasanudin pada periode Januari 2006 sampai November 2019 (167 data). Data pada periode Januari 2009 sampai

## Drajat Indra Purnama

Februari 2017 (134 data) digunakan sebagai data *training* sedangkan data pada periode Maret 2017 sampai November 2019 (33 data) digunakan sebagai data *testing*.

### 3.1. Pengujian Linearitas

Pengujian linearitas menggunakan uji Ramsey RESET menggunakan hipotesis  $H_0$  adalah data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin tidak mengandung pola nonlinear dan  $H_1$  adalah data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin mengandung pola nonlinear. Hasil pengujian linearitas menggunakan uji *Ramsey RESET* pada data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin selama periode Januari 2006 sampai dengan Februari 2017 disajikan pada Tabel 3.1.

**Tabel 3.1.** Uji *Ramsey RESET* Pada Data Jumlah Penumpang Penerbangan domestik di Bandara Sultan Hasanudin

Data	Statistik Uji ( <i>RESET</i> )	<i>p-value</i>
Penumpang	8,4251	0,004344

Berdasarkan Tabel 3.1 dapat dilihat bahwa nilai statistik uji *Ramsey RESET* lebih besar dari  $F_{(0,05;1;131)} = 3,9134$  atau nilai *p-value* lebih kecil dari  $\alpha = 5\%$  maka dapat disimpulkan  $H_0$  ditolak yang berarti bahwa data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin mengandung pola nonlinear.

### 3.2. Stasioneritas Data

Syarat yang harus dipenuhi sebelum melakukan analisis data deret waktu adalah kestasioneran data. Dalam penelitian ini dilakukan uji stasioneritas data menggunakan uji *Augmented Dickey Fuller* (ADF) dengan menggunakan hipotesis  $H_0$  adalah data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin tidak stasioner dan  $H_1$  adalah data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin stasioner. Hasil uji ADF pada data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin disajikan pada Tabel 3.2.

**Tabel 3.2.** Pengujian ADF Data Jumlah Penumpang Penerbangan Domestik di Bandara Sultan Hasanudin Sebelum dan Sesudah *Differencing*

Data	Sebelum <i>Differencing</i>		Sesudah <i>Differencing</i> Pertama	
	<i>p-value</i>	Kesimpulan	<i>p-value</i>	Kesimpulan
Penumpang	0,1585	Data Tidak Stasioner	0,01	Data Stasioner

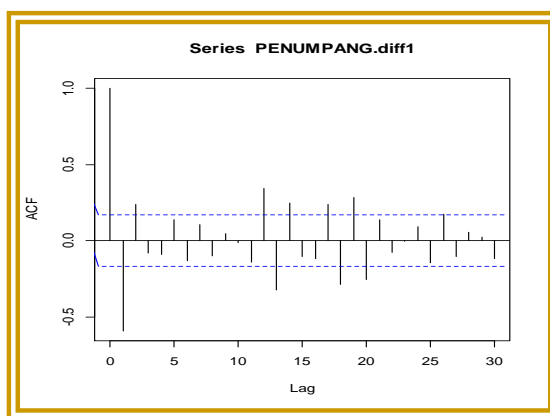
Berdasarkan Tabel 3.2, hasil uji ADF pada data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin menunjukkan nilai *p-value* lebih besar dari  $\alpha = 5\%$ , sehingga  $H_0$  diterima sehingga dapat disimpulkan data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin tidak stasioner. Oleh karena itu, salah satu cara yang dilakukan untuk membuat data stasioner adalah dengan *differencing*. Setelah *differencing* pertama, hasil uji ADF pada data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin hasil *differencing* pertama memiliki *p-value* lebih kecil dari  $\alpha = 5\%$ , sehingga  $H_0$  ditolak sehingga dapat disimpulkan data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin sudah stasioner.



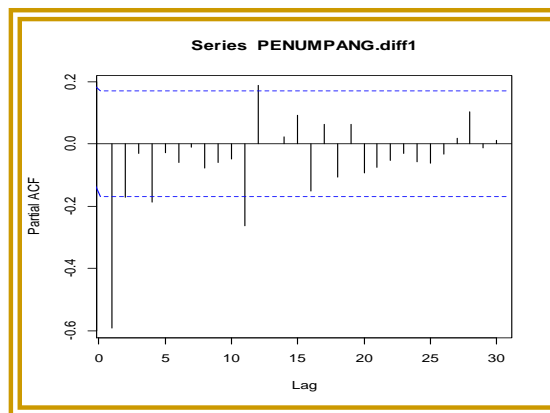
## Drajat Indra Purnama

### 3.3. Peramalan Jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin Menggunakan Model ARIMA

Identifikasi model deret waktu dilakukan untuk menentukan orde waktu untuk model ARIMA yang akan dibangun. Berdasarkan pengujian stasioneritas data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin merupakan data yang tidak stasioner sehingga perlu dilakukan *differencing* pertama. Grafik ACF dan PACF data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin setelah dilakukan *diferencing* pertama dapat dilihat pada Gambar 3.1 dan Gambar 3.2.



**Gambar 3.1.** Grafik ACF Data Jumlah Penumpang Penerbangan Domestik di Bandara Sultan Hasanudin *diferencing* pertama



**Gambar 3.2.** Grafik PACF Data Jumlah Penumpang Penerbangan Domestik di Bandara Sultan Hasanudin *diferencing* pertama

Berdasarkan Gambar 3.1 dan Gambar 3.2 terlihat bahwa ACF signifikan pada *lag* 1 sedangkan PACF signifikan pada *lag* 2 sehingga ditetapkan model sementara adalah ARIMA(1,1,2).

Tahap selanjutnya setelah diperoleh model ARIMA sementara adalah melakukan estimasi parameter model ARIMA yang mungkin untuk menggambarkan sifat data. Menurut prinsip

## Drajat Indra Purnama

*parsimony* dari pemodelan (model yang baik adalah model yang memiliki parameter yang sedikit). sedikit) dilakukan pengujian beberapa alternatif model ARIMA yang mungkin. Berdasarkan beberapa kemungkinan model ARIMA yang diestimasi parameternya terdapat empat model yang memiliki hasil uji parsial semua parameternya signifikan yaitu model ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1) dan ARIMA(0,1,2). Hasil uji parsial model ARIMA dapat dilihat pada Tabel 3.3.

**Tabel 3.3.** Hasil Uji Parsial Model SARIMA

No	Parameter	Estimasi Parameter	Standar Error	<i>t</i> -hitung	Kesimpulan	
1	ARIMA(1,1,0)	AR(1)	-0,6079	0,0702	-8,66	Signifikan
2	ARIMA(0,1,1)	MA(1)	-0,6535	0,0625	-10,46	Signifikan
3	ARIMA(1,1,1)	AR(1)	-0,3775	0,1407	-2,68	Signifikan
		MA(1)	-0,3736	0,1508	-2,48	Signifikan
4	ARIMA(2,1,0)	AR(1)	-0,7280	0,0874	-8,33	Signifikan
		AR(2)	-0,1953	0,0871	-2,24	Signifikan
5	ARIMA(0,1,2)	MA(1)	-0,7201	0,0818	-8,80	Signifikan
		MA(2)	0,1633	0,0844	1,93	Tidak Signifikan
6	ARIMA(2,1,1)	AR(1)	-0,0212	0,2032	-0,10	Tidak Signifikan
		AR(2)	0,2536	0,1601	1,58	Tidak Signifikan
		MA(1)	-0,7222	0,1699	-4,25	Signifikan
7	ARIMA(1,1,2)	AR(1)	-0,4696	0,2156	-2,18	Signifikan
		MA(1)	-0,2830	0,2228	-1,27	Tidak Signifikan
		MA(2)	-0,0811	0,1680	-0,48	Tidak Signifikan
8	ARIMA(2,1,2)	AR(1)	0,1324	0,3780	0,35	Tidak Signifikan
		AR(2)	0,2843	0,1467	1,94	Tidak Signifikan
		MA(1)	-0,8823	0,3792	-2,33	Signifikan
		MA(2)	0,0923	0,2166	0,43	Tidak Signifikan

Pengujian diagnostik model merupakan pemeriksaan asumsi dasar model deret waktu sudah terpenuhi. Dalam penelitian ini, dilakukan pengujian residual *white noise* dengan uji *Ljung-Box*. Hasil pengujian residual *white noise* dengan uji *Ljung-Box* ARIMA dapat dilihat pada Tabel 3.4.

**Tabel 3.4.** Pengujian Diagnostik Model ARIMA

No	Model	<i>p</i> -value uji <i>Ljung-Box</i>	Kesimpulan
1	ARIMA(1,1,0)	0,1381	<i>White Noise</i>
2	ARIMA(0,1,1)	0,0756	<i>White Noise</i>
3	ARIMA(1,1,1)	0,8388	<i>White Noise</i>
4	ARIMA(2,1,0)	0,7878	<i>White Noise</i>

Pada Tabel 3.4 terlihat bahwa *p*-value uji *Ljung-Box* semua model ARIMA nilainya lebih dari  $\alpha = 5\%$ , dapat disimpulkan bahwa residual model ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1) dan ARIMA(2,1,0) bersifat *white noise*.

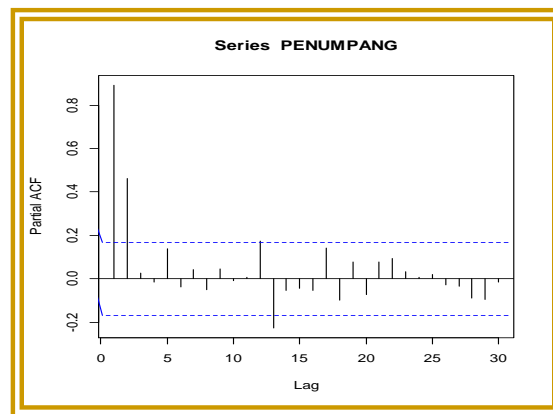
Pemilihan model ARIMA terbaik dapat dilakukan dengan melihat nilai *Akaike Information Criterion* (AIC) pada model ARIMA. Hasil perhitungan AIC dari model ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1) dan ARIMA(2,1,0) berturut-turut adalah 3.089,62, 3.089,89,

## Drajat Indra Purnama

3.085,94 dan 3.086,69. Model terbaik merupakan model dengan nilai AIC terkecil, dapat disimpulkan bahwa model terbaik yang diperoleh adalah ARIMA(1,1,1).

### 3.4. Peramalan Jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin Menggunakan Model SVR

Sebelum masuk ke pemodelan menggunakan SVR, langkah pertama yang dilakukan mengkonversi data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin 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 penerbangan domestik di Bandara Sultan Hasanudin dapat dilihat pada Gambar 3.3.



**Gambar 3.3.** Grafik PACF Data Jumlah Penumpang Penerbangan Domestik di Bandara Sultan Hasanudin

Berdasarkan plot PACF pada Gambar 3.3 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. Langkah selanjutnya 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  $C$ ,  $\gamma$  dan  $\varepsilon$  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  $\gamma$  dengan pangkat bilangan bulat. Sedangkan nilai  $\varepsilon$  ditentukan terlebih dahulu. Rentang nilai parameter yang digunakan untuk tahapan *loose grid* pada penelitian ini dapat dilihat pada Tabel 3.5.

**Tabel 3.5.** Rentang Nilai Parameter Tahapan *Loose Grid* Metode *Grid Search*

Parameter	Rentang Nilai
$C$	$2^{-7}, 2^{-5}, \dots, 2^1, 2^3$
$\gamma$	$2^{-3}, 2^{-1}, \dots, 2^5, 2^7$
$\varepsilon$	0,05; 0,06; ...; 0,1

## Drajat Indra Purnama

Rentang nilai parameter pada Tabel 3.5 digunakan untuk proses penentuan parameter model SVR. Melalui proses *grid search* pada tahapan *loose grid* diperoleh nilai optimal dari parameter model SVR adalah  $C = 2^1$ ,  $\gamma = 2^{-3}$  dan  $\varepsilon = 0,09$ . 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  $\varepsilon$  yang diperoleh pada tahapan *loose grid*.

**Tabel 3.6.** Rentang Nilai Parameter Tahapan *Finer Grid* Metode *Grid Search*

Parameter	Rentang Nilai
$C$	$2^{0,25}, 2^{0,5}, 2^{0,75}, 2^1, 2^{1,25}, 2^{1,5}, 2^{1,75}$
$\gamma$	$2^{-3}, 2^{-2,75}, 2^{-2,5}, 2^{-2,25}$
$\varepsilon$	0,09

Rentang nilai parameter yang digunakan pada tahapan *finer grid* dapat dilihat pada Tabel 3.6. Melalui proses *grid search* pada tahapan *finer grid* diperoleh nilai optimal dari parameter model SVR adalah  $C = 2^{0,5}$ ,  $\gamma = 2^{-3}$  dan  $\varepsilon = 0,09$ . 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  $C = 2^{0,5}$ ,  $\gamma = 2^{-3}$  dan  $\varepsilon = 0,09$ .

### 3.5. Perbandingan Model ARIMA dan Model SVR untuk peramalan Jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin

Perbandingan model ARIMA dan model SVR dilihat pada sejauh mana akurasi atau ketepatan model ARIMA dibandingkan model SVR dalam peramalan jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin. Penentuan keakuratan model dan peramalan dilakukan dengan membandingkan nilai MAPE. Hasil perhitungan nilai MAPE pada data training dan testing pada model ARIMA dan SVR dapat dilihat pada Tabel 3.7.

**Tabel 3.7.** Perbandingan MAPE Model ARIMA dan Model SVR

Model	MAPE (%)	
	<i>Training</i>	<i>Testing</i>
ARIMA	9,0551	12,3207
SVR	7,5755	11,3812

Tabel 3.7 menunjukkan bahwa nilai MAPE model SVR lebih kecil dibandingkan nilai MAPE model ARIMA baik untuk data *training* dan data *testing*. Hal tersebut menunjukkan bahwa model SVR secara signifikan mampu memberikan nilai peramalan jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin yang lebih akurat dibandingkan model ARIMA.

## 4. KESIMPULAN

Berdasarkan hasil analisis terhadap data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin dapat disimpulkan sebagai berikut :

## Drajat Indra Purnama

1. Data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin merupakan data yang berpola nonlinear.
2. Model ARIMA terbaik yang terbentuk adalah ARIMA(1,1,1). Sedangkan model SVR terbaik yang terbentuk adalah SVR dengan kernel RBF dan parameter optimalnya adalah  $C = 2^{0,5}$ ,  $\gamma = 2^{-3}$  dan  $\varepsilon = 0,09$ .
3. Perbandingan model ARIMA dan model SVR dalam memodelkan dan meramalkan data jumlah penumpang penerbangan domestik di Bandara Sultan Hasanudin menunjukkan bahwa model SVR lebih akurat dibandingkan dengan model ARIMA. Hal ini dilihat dari nilai MAPE untuk data *training* ataupun data *testing* model SVR lebih kecil dibandingkan model ARIMA.

### DAFTAR PUSTAKA

- [1] Badan Pusat Statistik. Diakses Pada Tanggal 11 Januari 2020 Melalui <https://www.bps.go.id/linkTableDinamis/view/id/812>
- [2] Boser, B.E., Guyon, I., and Vapnik, V. 1992. A Training Algorithm for Optimal Margin Classifiers. *Proceedings of the fifth annual workshop on Computational learning theory (COLT)* : 144-152.
- [3] Cryer, J.D. 1986. *Time Series Analysis*. Boston : PWS-KENT Publishing Company
- [4] Haykin, S. 2009. *Neural Networks and Learning Machines, Third Edition*. New Jersey : Pearson Education, Inc.
- [5] Hsu, C. W., Chang, C.C., and Lin, C.J. 2016. *A Practical Guide to Support Vector Classification*. Taipei : Department of Computer Science National Taiwan University.
- [6] Ramsey, J. B. 1969. Tests for Specification Errors in Classical Linear Least Squares Regression Analysis. *Journal of the Royal Statistical Society, Series B*. 31 (2): 350–371.
- [7] Smola, A.J. dan Scholkopf, B. 2004. A Tutorial On Support Vector Regression. *Statistic and Computing* 14, 199-222.
- [8] Vapnik, V.N. 1995. *The Nature of Statistical Learning Theory*. New York: Springer
- [9] Wei, W.W.S. 2006. *Time Series Analysis: Univariate and Multivariate Methods, Second Edition*. New York : Pearson Education.
- [10] Zhang, G.P. 2003. Time Series Forecasting using a Hybrid ARIMA dan Neural Network Model. *Neurocomputing* 50,159-175.