

# Peramalan Penumpang Angkutan Laut Rute Surabaya-Jayapura di PT Pelayaran Nasional Indonesia (Persero) Cabang Surabaya

Eliya Ainul Farri dan Irhamah

Departemen Statistika, Fakultas Matematika Komputasi dan Sains Data,

Institut Teknologi Sepuluh Nopember (ITS)

*e-mail:* irhamah@statistika.its.ac.id

**Abstrak**—Laut Indonesia memiliki peranan penting bagi kehidupan bangsa dan negara. Salah satunya dalam hal transportasi. Indonesia memiliki beberapa pelabuhan utama, salah satunya adalah Pelabuhan Tanjung Perak. Sejak tahun 2010 Pelabuhan Tanjung Perak menaungi 2 perusahaan kapal, salah satunya adalah PT Peln (Persero). Sebesar 41% dari total penumpang angkutan laut di Indonesia pada tahun 2017 menggunakan jasa PT Peln (Persero). Tingginya kebutuhan pelayanan di PT Peln (Persero) mengharuskan perusahaan memiliki database yang baik. Namun dikarenakan human error, data perusahaan tahun 2017 hilang. Pada penelitian ini akan didapatkan model terbaik untuk peramalan jumlah penumpang angkutan laut rute Surabaya-Jayapura di PT Peln (Persero) untuk tahun 2017 dengan menggunakan metode ARIMA Box-Jenkins dan regresi time series efek variasi kalender dengan residual data dilanjutkan dengan ARIMA Box-Jenkins. Efek variasi kalender digunakan karena pergeseran bulan terjadinya Hari Raya Idul Fitri. Hasil penelitian menunjukkan bahwa peramalan dengan menggunakan regresi time series memberikan hasil yang lebih baik dibandingkan dengan metode ARIMA Box-Jenkins karena nilai MAPE dan AIC yang dihasilkan cenderung lebih kecil.

**Kata Kunci**—ARIMA Box-Jenkins, Efek Variasi Kalender, Regresi Time Series.

## I. PENDAHULUAN

INDONESIA merupakan negara kepulauan yang memiliki banyak pulau dan negara maritim yang memiliki laut yang luas serta garis pantai yang panjang. Laut memiliki peranan penting bagi kehidupan bangsa dan negara. Sektor maritim dapat menjadi sangat strategis bagi Indonesia karena sumber daya perikanan, mineral, energi terbarukan, transportasi, pariwisata, dan keanekaragaman hayati sangat melimpah serta menunggu untuk menjadi sumber positioning daya saing Indonesia dalam persaingan global. Namun, sektor maritim masih kurang mendapatkan perhatian dibandingkan dengan sektor daratan.

Pelabuhan Tanjung Perak merupakan salah satu pelabuhan utama di Indonesia yang terletak di wilayah Jawa Timur dan merupakan pelabuhan paling sibuk kedua di Indonesia setelah Pelabuhan Tanjung Priok. Data publikasi BPS (2018) menunjukkan bahwa Pelabuhan Tanjung Perak cenderung mendominasi jumlah penumpang angkutan laut dari tahun 2006 hingga tahun 2017 dibandingkan dengan Pelabuhan Tanjung priok. Penumpang angkutan laut pada tahun 2017 di Indonesia mencapai angka 8,27 juta penumpang dan sebesar 3,43 juta penumpang dilayani oleh PT Pelayaran Nasional Indonesia (Persero) [1]. Harapan dari pe-

nelitian ini adalah memberikan informasi mengenai jumlah penumpang angkutan laut di PT Pelayaran Nasional Indonesia (Persero) rute Surabaya-Jayapura pada tahun 2017 karena data seluruh pelayaran di PT Pelayaran Nasional Indonesia (Persero) pada tahun 2017 telah hilang karena adanya human error. Data pada tahun 2017 penting karena digunakan oleh pihak manajemen PT Pelayaran Nasional Indonesia (Persero) untuk melakukan prediksi jumlah penumpang kapal laut untuk beberapa tahun ke depan. Prediksi ini penting karena arus mudik menggunakan transportasi laut akan semakin meningkat seiring dengan naiknya harga tiket pesawat terbang. Selain itu, penelitian ini juga dimaksudkan untuk mengetahui pola penumpang angkutan laut jika terdapat momen perayaan tertentu seperti Hari Raya Idul Fitri secara statistik.

Penelitian ini menggunakan metode ARIMA Box-Jenkins dan regresi *time series* dengan efek variasi kalender dengan residual data dimodelkan dengan ARIMA menggunakan data jumlah penumpang angkutan laut dari beberapa tahun sebelumnya. Jumlah penumpang angkutan laut diperkirakan akan meningkat cukup tinggi pada momen perayaan tertentu seperti Bulan Ramadhan dan Hari Raya Idul Fitri, kemudian kembali normal pada periode selanjutnya. Bergeser bulan terjadinya Bulan Ramadhan dan Hari Raya Idul Fitri mengindikasikan adanya variasi kalender. Penelitian dengan menggunakan ARIMA *Box-Jenkins* diantaranya pernah dilakukan oleh Dwijantari (2017) [2], Fanani (2017) [3], dan Endartyana (2017) [4].

## II. TINJAUAN PUSTAKA

Tinjauan pustaka pada penelitian ini diantaranya adalah sebagai berikut.

### A. Metode Time Series

Deret Waktu (Time series) adalah serangkaian pengamatan terhadap suatu variabel yang diambil dari waktu ke waktu dan dicatat secara berurutan menurut urutan waktu kejadian dengan interval waktu yang tetap [5]. Data time series dikatakan stasioner apabila dalam mean dan varians berfluktuasi secara konstan dari waktu ke waktu. Untuk memeriksa kestasioneran data dapat menggunakan plot Box-Cox dan time series plot ACF. Stasioneritas terhadap varians adalah apabila data berfluktuasi di sekitar nilai varians yang konstan, artinya data memiliki keragaman yang homogen.

Ketidakstasioneran terhadap varians diketahui dengan melakukan pengujian menggunakan pengujian Bartlett. Hipotesis yang digunakan dalam pengujian Bartlett yaitu

$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$  (data telah stasioner terhadap varians) dan  $H_1 : \exists \sigma_i^2 \neq \sigma_j^2, i \neq j$  (data tidak stasioner terhadap varians) dengan  $i, j = 1, 2, \dots, k$ . Statistik pengujian Bartlett disajikan pada persamaan (1) [6].

$$B = \frac{\left[ \sum_{i=1}^k v_i \times \ln \left( \frac{\sum_{i=1}^k v_i s_i^2}{\sum_{i=1}^k v_i} \right) \right] - \left[ \sum_{i=1}^k v_i \times \ln s_i^2 \right]}{1 + \frac{\sum_{i=1}^k \frac{1}{v_i} - \frac{1}{\sum_{i=1}^k v_i}}{3 \times (k-1)}}, \quad (1)$$

dengan  $s_i^2 = \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 / (n_i - 1)$ ,

$k$  = jumlah sampel,

$n_i$  = jumlah unit penelitian pada sampel ke- $i$ ,

$v_i = n_i - 1$ .

Daerah penolakan yang digunakan adalah hipotesis nol ditolak jika nilai statistik uji lebih besar dari nilai kritis atau jika nilai  $p$ -value kurang dari nilai taraf signifikan. Nilai kritis menggunakan  $\chi^2_{(k-1, \alpha)}$ .

Ketidastasioneran terhadap varians dapat diatasi dengan menggunakan transformasi Box-Cox yang ditunjukkan oleh persamaan (2) [5]

$$T(Z_t) = \frac{Z_t^\lambda - 1}{\lambda}, \text{ dimana } \lambda \neq 0 \quad (2)$$

dimana nilai  $\lambda$  merupakan parameter transformasi. Untuk pengujian stasioneritas terhadap mean dilakukan uji unit root menggunakan uji *Augmented Dickey-Fuller* (ADF). Hipotesis yang digunakan dalam pengujian tersebut adalah  $H_0 : \delta = 0$  dan  $H_1 : \delta < 0$  dengan nilai  $\delta = (\rho - 1)$  dan  $-1 < \delta < 1$ . Statistik uji yang digunakan ditunjukkan oleh persamaan (3).

$$\Delta Z_t = \beta_0 + \beta_1 t + \delta Z_{t-1} + \sum_{i=1}^m \alpha_i \Delta Z_{t-i} + a_t, \quad (3)$$

dengan  $a_t$  = komponen residual,

$Z_t$  = data series,

$\Delta Z_{t-i} = (Z_{t-1} - Z_{t-2})$ ,

$\Delta Z_{t-2} = (Z_{t-2} - Z_{t-3})$ , dan seterusnya,

$\beta_0$  = intersept,

$\beta_1$  = tren stokastik,

$\delta$  = koefisien dari  $Z_{t-1}$ ,

$\alpha$  = koefisien dari  $\Delta Z_{t-i}$ .

Apabila data tidak stasioner terhadap mean, maka perlu dilakukan differencing atau metode pembedaan [5].

**B. Autocorrelation Function dan Partial Autocorrelation Function**

*Autocorrelation Function* (ACF) adalah hubungan linier antara  $Z_t$  dengan  $Z_{t+k}$  pada suatu data time series. Fungsi ACF ditunjukkan oleh persamaan (4) [5].

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^n (Z_t - \bar{Z})^2}; k = 0, 1, 2, \dots \quad (4)$$

dengan  $\bar{Z} = \frac{1}{n} \sum_{t=1}^n Z_t$ ,

$Z_t$  = pengamatan pada waktu ke- $t$ ,

$Z_{t+k}$  = pengamatan pada waktu ke- $t$  dan pada lag ke- $k$ ,

$\hat{\rho}_k$  = taksiran autokorelasi data *insample* pada lag ke- $k$

*Partial Autocorrelation Function* (PACF) merupakan suatu fungsi untuk mengetahui nilai korelasi antara  $Z_t$  dengan  $Z_{t+k}$  setelah terdapat pengaruh dari variabel  $Z_{t+1}, Z_{t+2}, \dots, Z_{t+k-1}$  dihilangkan. Metode umum dalam menentukan fungsi autokorelasi parsial untuk setiap proses stasioner dengan fungsi autokorelasi sesuai dengan persamaan (5).

$$\hat{\rho}_{k+1, k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^{k-1} \phi_{kj} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^k \phi_{kj} \hat{\rho}_j}. \quad (5)$$

**C. ARIMA Box-Jenkins**

ARIMA Box-Jenkins adalah salah satu metode untuk meramalkan suatu data yang berbasis *time series*. Model autoregressive (AR) orde  $p$  menyatakan bahwa suatu model pada pengamatan waktu ke- $t$  merupakan kombinasi linear dari pengamatan sebelumnya dan didefinisikan pada persamaan (6).

$$\phi_p(B) \dot{Z}_t = a_t, \quad (6)$$

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \dots + \phi_p \dot{Z}_{t-p} + a_t,$$

dengan  $\phi_p$  = parameter AR pada lag ke- $j$  dengan  $j = 1, 2, \dots, p$ ,

$$\dot{Z}_t = Z_t - \mu,$$

$Z_t$  = data *time series* pada waktu ke- $t$ ,

$a_t$  = residual pada waktu ke- $t$ .

Model *moving average* (MA) orde  $q$  menyatakan bahwa suatu model pada pengamatan waktu ke- $t$  dipengaruhi oleh kesalahan masa lalu dan didefinisikan pada persamaan (7).

$$\dot{Z}_t = \theta_q(B) a_t, \quad (7)$$

$$\dot{Z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

dengan  $\theta_q$  = parameter MA pada lag ke- $l$  dengan  $l = 1, 2, \dots, q$ ,

$$\dot{Z}_t = Z_t - \mu,$$

$Z_t$  = data *time series* pada waktu ke- $t$ ,

$a_t$  = residual pada waktu ke- $t$ .

Model dari autoregressive moving average pada orde  $p$  dan  $q$  dituliskan dengan persamaan (8).

$$\phi_p(B) \dot{Z}_t = \theta_q(B) a_t, \quad (8)$$

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \dots + \phi_p \dot{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

dengan  $\phi_p$  = parameter AR pada lag ke- $j$ ,

$\theta_q$  = parameter MA pada lag ke- $l$ ,

$$\dot{Z}_t = Z_t - \mu,$$

$Z_t$  = data *time series* pada waktu ke- $t$ ,

$a_t$  = residual pada waktu ke- $t$ .

Model ARIMA merupakan gabungan antara model *autoregressive* dan *moving average* dengan adanya pembedaan atau *differencing* ( $d$ ) yang disebabkan karena ketidastasioneran dalam *mean*. Sehingga model ARIMA

( $p, d, q$ ) dapat dituliskan dengan persamaan (9).

$$\phi_p(B)(1-B)^d \dot{Z}_t = \theta_q + \theta_q(B)a_t, \quad (9)$$

dengan  $\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ ,

$$\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p),$$

$\phi_0$  = koefisien tren deterministik,

$\phi_p$  = parameter AR pada lag ke- $j$ ,

$\theta_q$  = parameter MA pada lag ke- $l$ ,

$$\dot{Z}_t = Z_t - \mu,$$

$Z_t$  = data *time series* pada waktu ke- $t$ ,

$p$  = orde *Autoregressive* (AR),

$d$  = orde *differencing*,

$q$  = orde *Moving Average* (MA),

$a_t$  = residual pada waktu ke- $t$ .

Estimasi parameter pada model ARIMA menggunakan metode *Conditional Least Square*. Metode ini dilakukan dengan cara mencari nilai parameter dengan meminimumkan jumlah kuadrat error (SSE). Misalkan diterapkan pada model AR maka dapat dinyatakan sesuai dengan persamaan (10) [7].

$$\begin{aligned} \phi_p(B) \dot{Z}_t &= a_t \\ \dot{Z}_t &= \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + \dots + \phi_p \dot{Z}_{t-p} + a_t, \end{aligned} \quad (10)$$

dengan  $\phi_p$  = parameter AR pada lag ke- $j$  dengan

$$j = 1, 2, \dots, p,$$

$$\dot{Z}_t = Z_t - \mu,$$

$Z_t$  = data *time series* pada waktu ke- $t$ ,

$a_t$  = residual pada waktu ke- $t$ .

Hipotesis yang digunakan untuk melakukan uji signifikansi parameter model AR adalah  $H_0 : \phi_j = 0$  dan nilai  $H_1 : \phi_j \neq 0$ . Statistik uji yang digunakan dinyatakan pada persamaan (11).

$$t_{hitung} = \frac{\hat{\phi}_j}{SE(\hat{\phi}_j)}, \quad (11)$$

dengan  $SE(\hat{\phi}_j)$  = standar *error* estimasi parameter model AR. Hipotesis nol ditolak apabila nilai dari  $|t_{hitung}|$  lebih besar dari nilai  $t_{\alpha/2, (n-p)}$  dengan  $n$  merupakan banyak pengamatan dan  $p$  merupakan banyak parameter yang diestimasi.

#### D. Asumsi White Noise

Residual bersifat white noise jika terdapat korelasi antar re-sidual dengan mean adalah nol dan varians konstan. Plot residual dapat digunakan untuk melihat apakah varians residual konstan, sedangkan untuk melihat residual bersifat white noise dapat dilakukan dengan menggunakan pengujian Ljung-Box. Hipotesis yang digunakan dalam pengujian ini adalah  $H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0$  (residual pada lag ke- $k$  memenuhi asumsi *white noise*) dan  $H_1$  : minimal ada satu nilai  $\rho_k \neq 0$ . Statistik uji yang digunakan dalam pengujian ini dinyatakan pada persamaan (12)

$$Q^* = n(n+2) \sum_{k=1}^K (n-k)^{-1} \hat{\rho}_k^2, \quad (12)$$

dengan  $k = 1, 2, \dots, K$

$\hat{\rho}_k^2$  = ACF residual pada lag ke- $k$ ,

$K$  = besarnya lag maksimum,

$n$  = banyaknya pengamatan,

$p$  = orde *Autoregressive* (AR),

$q$  = orde *Moving Average* (MA),

$\alpha$  = nilai taraf signifikan.

Daerah penolakan yang digunakan adalah hipotesis nol ditolak jika nilai statistik uji lebih besar dari nilai kritis.

Nilai kritis menggunakan  $\chi^2_{(K-p-q, \alpha)}$ .

#### E. Asumsi Residual Distribusi Normal

Uji asumsi kenormalan residual digunakan adalah uji normalitas metode Kolmogorov-Smirnov [9]. Hipotesis yang digunakan dalam pengujian ini adalah

$H_0 : F(a_t) = F_0(a_t)$  (residual berdistribusi normal) dan

$H_1 : F(a_t) \neq F_0(a_t)$  (residual tidak berdistribusi normal).

Statistik uji yang digunakan disajikan pada persamaan (13).

$$D = \text{Sup} |F(\varepsilon_t) - F_0(\varepsilon_t)|, \quad (13)$$

dengan  $F_0(\varepsilon_t)$  = fungsi peluang kumulatif distribusi normal,

$F(\varepsilon_t)$  = fungsi distribusi yang belum diketahui,

$\text{Sup}$  = nilai maksimum dari  $|F(\varepsilon_t) - F_0(\varepsilon_t)|$ .

$\varepsilon_t$  = residual data ke- $t$ .

Daerah penolakan yang digunakan adalah hipotesis nol ditolak jika nilai statistik uji lebih besar dari nilai kritis. Nilai kritis menggunakan tabel *Kolmogorov-Smirnov* dengan derajat bebas adalah  $n$ , dimana  $n$  merupakan banyaknya observasi.

#### F. Regresi Time Series Efek Variasi Kalender

*Regression with Autocorelated Errors* sering disebut juga dengan regresi time series. Model regresi time series merupakan model yang menghubungkan variabel dependen  $Y_t$  dengan fungsi waktu. Model ini bermanfaat saat parameter yang menjelaskan peramalan time series berada pada kondisi konstan secara terus menerus [8]. Berbeda dengan model regresi konvensional, regresi time series merupakan fungsi antara satu variabel respon dengan satu atau lebih variabel prediktor dimana kedua variabel tersebut bersyarat pada waktu. Secara umum model regresi time series dapat ditulis sesuai persamaan (14) [5].

$$Y_t = \beta_1 X_{1,t} + \beta_2 X_{2,t} + \dots + \beta_i X_{i,t} + \varepsilon_t, \quad (14)$$

#### G. Pemilihan Model Terbaik

*AIC (Akaike's Information Criterion)* merupakan indikator kebaikan model yang pertama. *AIC* adalah suatu indikator pemilihan model terbaik yang diperkenalkan oleh Akaike pada tahun 1973 dengan mempertimbangkan banyaknya parameter dalam model. Kriteria *AIC* ditunjukkan pada persamaan (15) [8].

$$AIC = n \ln \left( \frac{SSE}{n} \right) + 2f + n(1 + \ln(2\pi)), \quad (15)$$

dengan  $n$  = banyaknya pengamatan,

$f$  = banyaknya parameter dalam model,

*SSE* = jumlah kuadrat *error*.

*MAPE (Mean Absolute Percentage Error)* merupakan indikator kebaikan model yang kedua. *MAPE* digunakan untuk data *outsample*. Nilai berkaitan dengan presentase residual [6]. Proses perhitungan *MAPE* ditunjukkan oleh

persamaan (16).

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Z_t - \hat{Z}_t|}{Z_t} \times 100, \quad (16)$$

dengan  $n$  = jumlah pengamatan,

$Z_t$  = data aktual pada periode ke- $t$ ,

$\hat{Z}_t$  = data ramalan pada periode ke- $t$ .

Nilai MAPE berkisar antara 0% hingga 100% sehingga nilai yang dihasilkan lebih mudah untuk diinterpretasi.

### III. METODOLOGI PENELITIAN

#### A. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari PT. Pelayaran Nasional Indonesia (Persero), yaitu data jumlah penumpang angkutan laut dari Januari 2012 hingga Desember 2016.

#### B. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri atas variabel respon dan variabel prediktor. Berdasarkan rute yang diuraikan tersebut maka variabel respon yang digunakan adalah sebagai  $Z_t$  = Jumlah penumpang angkutan laut rute Tanjung Perak – Jayapura. Variabel prediktor yang digunakan dalam penelitian ini merupakan variabel *dummy* yaitu *dummy* bulan dan Hari Raya Idul Fitri. *Dummy* variabel prediktor secara lengkap dijelaskan oleh Tabel 1.

Tabel 1.  
Variabel *Dummy*

Variabel Prediktor	Kategorik
<i>Dummy</i> Bulan	$M_{1,t}$ 1 = Bulan ke- $t$ adalah Bulan Januari. 0 = Lainnya.
	$M_{2,t}$ 1 = Bulan ke- $t$ adalah Bulan Februari. 0 = Lainnya.
	...
	$M_{12,t}$ 1 = Bulan ke- $t$ adalah Bulan Desember. 0 = Lainnya.
<i>Dummy</i> Bulan Hari Raya Idul Fitri	$R_{t-1}$ 1 = Bulan ke- $t-1$ adalah 1 bulan sebelum Idul Fitri. 0 = Lainnya.
	$R_t$ 1 = Bulan terjadi Idul Fitri. 0 = Lainnya.
	$R_{t+1}$ 1 = Bulan ke- $t+1$ adalah 1 bulan sesudah Idul Fitri. 0 = Lainnya.

#### C. Langkah Analisis

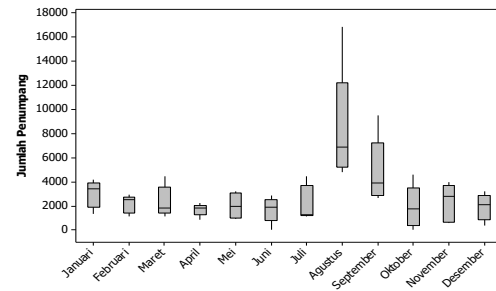
Data dibagi menjadi dua bagian, yaitu data insample dan outsample. Langkah-langkah yang digunakan adalah sebagai berikut.

1. Menganalisis statistika deskriptif dari data jumlah penumpang angkutan laut PT Pelayaran Nasional Indonesia (Persero).
2. Meramalkan data jumlah penumpang angkutan laut PT Pelayaran Nasional Indonesia (Persero) menggunakan metode ARIMA Box-Jenkins.
3. Meramalkan data jumlah penumpang angkutan laut PT Pelayaran Nasional Indonesia (Persero) menggunakan metode Regresi Time Series (TSR) dengan efek variasi kalender.
4. Membandingkan hasil ramalan jumlah penumpang angkutan laut PT Pelayaran Nasional Indonesia (Persero) tahun 2019 dari kedua metode.

### IV. ANALISIS DAN PEMBAHASAN

Analisis dan pembahasan pada penelitian ini diantaranya adalah eksplorasi data, peramalan dengan metode ARIMA, dan peramalan dengan menggunakan regresi *time series* dengan residual data diramalkan menggunakan ARIMA.

#### A. Karakteristik Data

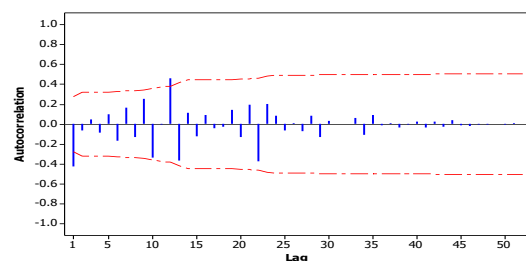


Gambar 1. Boxplot Data Kapal Penumpang Rute Surabaya-Jayapura.

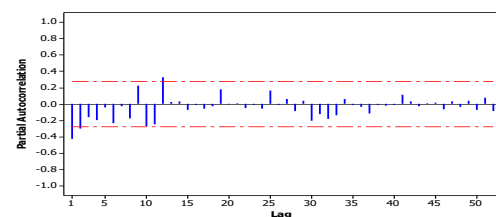
Gambar 1 menunjukkan bahwa median dari Bulan Januari hingga Desember tidak satupun yang memiliki median yang sama. Median tertinggi terdapat pada Bulan Agustus kemudian disusul oleh Bulan September. Terkait penyebaran data kapal penumpang rute Surabaya-Jayapura pada tahun 2012 hingga tahun 2016 dapat dilihat dari nilai interquartile range yang secara visual dapat dilihat dari panjang box pada boxplot. Box pada Bulan Agustus dan September memiliki rentang yang lebih besar dibandingkan dengan bulan lain. Hal tersebut menunjukkan bahwa data kapal penumpang rute Surabaya-Jayapura pada tahun 2012 hingga tahun 2016 di Bulan Agustus dan September lebih menyebar dibandingkan dengan bulan lain.

#### B. Pemodelan Kapal Penumpang Surabaya-Jayapura dengan Metode ARIMA

Pemodelan diawali dengan memeriksa asumsi stasioneritas terhadap varians dan *mean*. Berdasarkan analisis menggunakan pengujian Bartlett diperoleh kesimpulan bahwa data belum stasioner terhadap varians sehingga membutuhkan transformasi dan menggunakan pengujian Dickey-Fuller diperoleh kesimpulan bahwa data belum stasioner terhadap *mean* sehingga data perlu di *differencing*. Plot ACF dan PACF yang terbentuk disajikan pada Gambar 2.



(a)



(b)

Gambar 2. Plot ACF (a) dan PACF (b) Data Jumlah Kapal Penumpang Surabaya-Jayapura Setelah Proses Transformasi dan Differencing

Lag yang signifikan pada plot ACF adalah lag 1 dan 12, sedangkan pada plot PACF adalah lag 1, 2, dan 12. Time series plot pada Gambar 4.1 dan boxplot pada Gambar 4.2 mengindikasikan bahwa pada data jumlah penumpang kapal rute Surabaya-Jayapura memiliki pola musiman sehingga terdapat model dugaan yang mengandung model ARIMA seasonal. Beberapa model dugaan yang terbentuk antara lain adalah ARIMA ([12],1,1), ARIMA (2,1,[12]), ARIMA (0,1,1)(0,0,1)<sup>12</sup>, dan ARIMA (0,1,[1,12]). Hasil pengujian signifikansi parameter dari model-model dugaan tersebut disajikan pada Tabel 2.

Tabel 2. Pengujian Signifikansi Parameter ARIMA

Model	Parameter	Estimasi	p-value
ARIMA ([12],1,1)	$\phi_{12}$	0,49939	0,0007
	$\theta_1$	0,82965	<0,0001
ARIMA (2,1,[12])	$\phi_1$	-0,51245	0,0006
	$\phi_2$	-0,25445	0,0758
	$\theta_{12}$	-0,44457	0,0041
ARIMA (0,1,1)(0,0,1) <sup>12</sup>	$\theta_1$	0,84157	<0,0001
	$\Theta_1$	-0,43189	0,0049
ARIMA (0,1,[1,12])	$\theta_1$	0,73223	<0,0001
	$\theta_{12}$	-0,32679	0,0159

Hasil pengujian parameter pada Tabel 4.4 menunjukkan bahwa seluruh parameter pada model ARIMA ([12],1,1), ARIMA (2,1,[12]), ARIMA (0,1,1)(0,0,1)<sup>12</sup>, dan ARIMA (0,1,[1,12]) telah signifikan karena memiliki nilai p-value yang lebih kecil dari taraf signifikan sebesar 5%. Tahap selanjutnya adalah melakukan pengujian asumsi. Berdasarkan uji *white noise*, distribusi normal dan AIC yang terkecil maka ARIMA (0,1,1)(0,0,1)<sup>12</sup> dipilih sebagai model terbaik dengan model yang terbentuk disajikan pada persamaan (17).

$$Z_t^* = Z_{t-1} + a_t + 0,84157a_{t-1} + 0,43189a_{t-12} - 0,36346a_{t-13}, \quad (17)$$

dengan  $Z_t^* = (Z_t)^{1/2}$ .

C. Hasil Forecasting dengan Metode ARIMA

Model pada persamaan (17) kemudian digunakan untuk meramalkan jumlah penumpang kapal rute Surabaya-Jayapura pada tahun 2017. Hasil yang diperoleh disajikan pada Tabel 3.

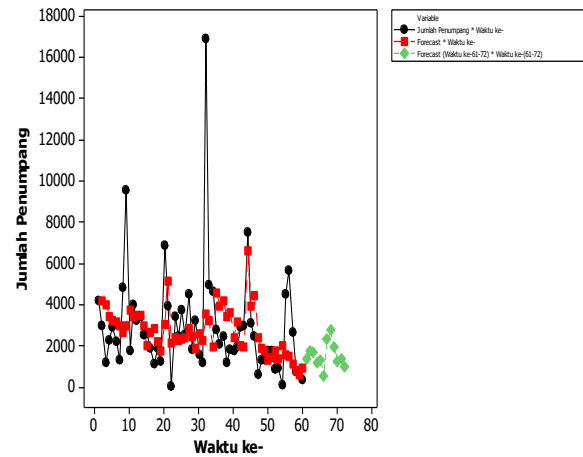
Tabel 3. Hasil Forecasting ARIMA

Bulan	t	Forecast	Real Forecast
Januari	61	37,3453	1394,7
Februari	62	42,1488	1776,5
Maret	63	41,4504	1718,1
April	64	34,4959	1190,0
Mei	65	36,0323	1298,3
Juni	66	22,5581	508,9
Juli	67	48,0977	2313,4
Agustus	68	52,5006	2756,3
September	69	44,0799	1943,0
Oktober	70	35,3770	1251,5
November	71	37,1733	1381,9
Desember	72	31,2771	978,3

Keterangan : Tanda tebal merupakan bulan terjadinya Hari Raya Idul Fitri.

Hasil peramalan dengan metode ARIMA menunjukkan bahwa jumlah penumpang kapal rute Surabaya-Jayapura pada tahun 2017 terbesar bukan berada pada bulan terjadinya Hari raya Idul Fitri ataupun Tahun Baru Masehi. Nilai AIC yang dihasilkan dari ramalan ini adalah sebesar

517,8454. Plot hasil ramalan disajikan pada Gambar 3.



Gambar 3. Plot Ramalan Data.

D. Pemodelan Kapal Penumpang Surabaya-Jayapura dengan Metode Regresi Time Series Efek Variasi Kalender

Pemodelan yang dilakukan adalah pemodelan jumlah penumpang kapal rute Surabaya-Jayapura dengan variabel *dummy*. Pemodelan dilakukan menggunakan regresi *time series* dengan model umum pada persamaan (18).

$$Z_t = \beta_0 + \beta_1 M_{1t} + \dots + \beta_{11} M_{11t} + \beta_{12} R_{t-1} + \beta_{13} R_t + \beta_{14} R_{t+1} + e_t \quad (18)$$

dengan  $e_t \sim IIDN$ . Tahap pertama adalah melakukan pengujian serentak dan diperoleh hasil yang disajikan pada Tabel 4.

Tabel 4. Pengujian Serentak Regresi Time Series Tahap Pertama

Source	df	SS	MS	F	P
Regression	14	248463445	17747389	6,00	0,000
Residual Error	39	115425694	2959633		
Total	53	363889139			

Pengujian serentak menunjukkan bahwa p-value yang terbentuk adalah sebesar 0,000. Berdasarkan hasil tersebut maka keputusan yang diperoleh adalah tolak  $H_0$ . Hal tersebut berarti bahwa minimal terdapat salah satu variabel *dummy* yang berpengaruh signifikan terhadap jumlah penumpang kapal rute Surabaya-Jayapura. Dilanjutkan pada pengujian secara parsial untuk memastikan hasil pada pengujian serentak. Hasil pengujian parsial disajikan pada Tabel 5.

Tabel 5. Pengujian Parsial Regresi Time Series Tahap Pertama

Prediktor	Koefisien	SE Koefisien	T	P
Konstan	2274,5	860,2	2,64	0,012
$M_{1t}$	754	1154	0,65	0,517
$M_{2t}$	-83	1154	-0,07	0,943
$M_{3t}$	90	1154	0,08	0,938
$M_{4t}$	-564	1154	-0,49	0,628
$M_{5t}$	-246	1154	-0,21	0,833
$M_{6t}$	-425	1281	-0,33	0,742
$M_{7t}$	-16	1820	-0,01	0,993
$M_{8t}$	5160	1855	2,78	0,008
$M_{9t}$	1068	1424	0,75	0,458
$M_{10t}$	-63	1216	-0,05	0,959
$M_{11t}$	421	1216	0,35	0,731
$R_{t-1}$	-318	1389	-0,23	0,820
$R_t$	-897	1683	-0,53	0,597
$R_{t+1}$	4083	1479	2,76	0,009

Keterangan : Tanda tebal merupakan parameter signifikan pada taraf 5%.

Pengujian parsial menunjukkan bahwa hanya terdapat beberapa variabel *dummy* yang berpengaruh signifikan ter-

hadap jumlah penumpang kapal rute Surabaya-Jayapura. Diperlukan pengujian regresi *time series* tahap kedua untuk memperoleh estimasi parameter pemodelan variasi efek kalender yang terbaik. Hasil pengujian diawali dengan pengujian serentak dan diperoleh hasil yang disajikan pada Tabel 6.

Tabel 6. Pengujian Serentak Regresi *Time Series* Tahap Kedua

Source	df	SS	MS	F	P
Regression	12	247604939	20633745	7,28	0,000
Residual Error	51	116284200	2836200		
Total	53	363889139			

Pengujian serentak menunjukkan bahwa *p-value* yang terbentuk adalah sebesar 0,000. Berdasarkan hasil tersebut maka keputusan yang diperoleh adalah tolak  $H_0$  Hal tersebut berarti bahwa minimal terdapat salah satu dari beberapa variabel *dummy* yang berpengaruh signifikan terhadap jumlah penumpang kapal rute Surabaya-Jayapura. Dilanjutkan pada pengujian secara parsial untuk memastikan hasil pada pengujian serentak. Hasil pengujian parsial disajikan pada Tabel 7.

Tabel 7. Pengujian Parsial Regresi *Time Series* Tahap Kedua

Prediktor	Koefisien	SE Koefisien	T	P
Konstan	2274,5	842,1	2,70	0,010
$M_{1t}$	754	1130	0,67	0,508
$M_{2t}$	-83	1130	-0,07	0,942
$M_{3t}$	90	1130	0,08	0,937
$M_{4t}$	-564	1130	-0,50	0,620
$M_{5t}$	-246	1130	-0,22	0,829
$M_{6t}$	-552	1130	-0,49	0,628
$M_{7t}$	-624	1191	-0,52	0,603
$M_{8t}$	4887	1331	3,37	0,002
$M_{9t}$	844	1331	0,63	0,530
$M_{10t}$	-63	1191	-0,05	0,958
$M_{11t}$	421	1191	0,35	0,726
$R_{t+1}$	4532	1191	3,81	0,000

Keterangan : Tanda tebal merupakan parameter signifikan pada taraf 5%.

Pengujian parsial menunjukkan bahwa *dummy* Bulan Agustus dan *dummy* 1 bulan setelah Hari Raya Idul Fitri yang berpengaruh signifikan terhadap jumlah penumpang kapal rute Surabaya-Jayapura. Hal tersebut dapat dilihat dari nilai *p-value* semua variabel yang lebih kecil dari nilai taraf signifikan sebesar 0,05. Pemodelan kapal penumpang dengan menggunakan regresi *time series* menghasilkan persamaan yang disajikan pada persamaan (19).

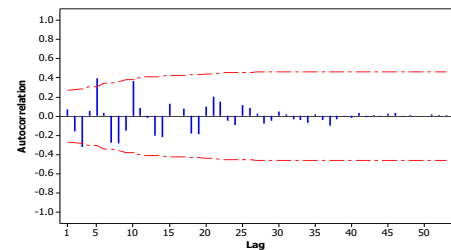
$$Z_t = 2274,5 + 754M_{1t} - 83M_{2t} + 90M_{3t} - 564M_{4t} - 246M_{5t} + \dots + (-552)M_{6t} - 624M_{7t} + 4887M_{8t} + 844M_{9t} - 63M_{10t} + \dots + 421M_{11t} + 4532R_{t+1} \quad (19)$$

Berdasarkan model yang terbentuk residual data diperiksa asumsi klasik dan memberikan hasil bahwa residual data tidak memenuhi asumsi residual identik dan independen namun memenuhi asumsi distribusi normal. Maka, residual data akan diramalkan dengan menggunakan ARIMA.

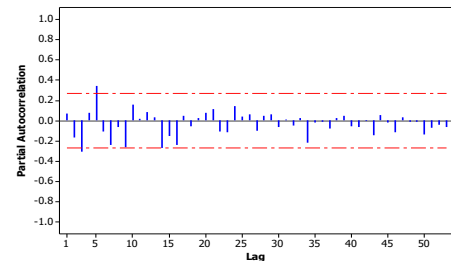
E. Pemodelan ARIMA Data Residual

Residual data hasil regresi *time series* dilakukan pengujian dengan menggunakan pengujian Bartlett diperoleh keputusan bahwa tidak stasioner terhadap varians sehingga perlu dilakukan transformasi dan setelah dilakukan pemeriksaan dengan menggunakan Box-Cox Transformasi diperoleh kesimpulan bahwa transformasi dilakukan dengan menggunakan nilai logaritma natural data. Namun, residual data telah stasioner terhadap *mean* sehingga tidak perlu dilakukan *differencing*. Plot ACF dan PACF yang terbentuk

disajikan pada Gambar 4.



(a)



(b)

Gambar 4. Plot ACF (a) dan PACF (b) Residual Data Hasil Pemodelan Efek Variasi Kalender

Identifikasi model residual data hasil pemodelan efek variasi kalender dilakukan berdasarkan plot ACF dan PACF yang disajikan pada Gambar 4. Lag yang signifikan pada plot ACF adalah lag 3 dan 5, sedangkan pada plot PACF adalah lag 3 dan 5. Beberapa model dugaan yang terbentuk antara lain adalah AR ([3,5]), AR([5]), MA ([3,5]), dan MA ([5]).

Tabel 8. Pengujian Signifikansi Parameter ARIMA Data Residual

Model	Parameter	Estimasi	p-value
AR([3,5])	$\mu$	6,38623	<0,0001
	$\phi_3$	-0,27193	0,0354
	$\phi_5$	0,36151	0,0059
AR([5])	$\mu$	6,43283	<0,0001
	$\phi_5$	0,40886	0,0024
MA([3,5])	$\mu$	6,37538	<0,0001
	$\theta_3$	0,32991	0,0162
	$\theta_5$	-0,21267	0,1244
MA([5])	$\mu$	6,39721	<0,0001
	$\theta_5$	-0,26126	0,0602

Hasil pengujian parameter menunjukkan bahwa seluruh parameter pada model AR ([3,5]) dan AR([5]) telah signifikan karena memiliki nilai *p-value* yang lebih kecil dari taraf signifikan sebesar 5%. Pengujian pada model MA ([3,5]) dan MA ([5]) tidak dilanjutkan karena salah satu parameter pada model tersebut tidak signifikan. Pengujian asumsi dilakukan pada model AR ([3,5]) dan AR([5]) dan hasil pengujian disajikan pada Tabel 9.

Tabel 9. Pengujian Asumsi ARIMA Residual Data

Model	White Noise		Normalitas	AIC
	Lag	p-value		
AR([3,5])	6	0,5616	>0,1500	163,5623
	12	0,3738		
	18	0,4377		
	24	0,3127		
AR([5])	6	0,3804	>0,1500	166,261
	12	0,0568		
	18	0,1033		
	24	0,0764		

Keterangan : Tanda tebal merupakan parameter signifikan pada taraf 5%.



Model terbaik yang terbentuk dari residual data model *time series* disajikan pada persamaan (20).

$$\begin{aligned} \dot{Z}_t^* &= -0,27193\dot{Z}_{t-3} + 0,36151\dot{Z}_{t-5} + a_t \\ Z_t^* &= 5,8141 - 0,27193Z_{t-3} + 0,36151Z_{t-5} + a_t \end{aligned} \quad (20)$$

dengan  $Z_t^* = \ln(Z_t)$ . Model AR ([3,5]) menghasilkan nilai MAPE untuk kesalahan ramalan data *outsample* adalah sebesar 13,82 %. Nilai ini lebih kecil dibandingkan dengan ramalan menggunakan model ARIMA (0,1,1)(0,0,1)<sup>12</sup> pada data jumlah penumpang kapal rute Surabaya-Jayapura. Sehingga model lengkap yang terbentuk dengan menggunakan regresi *time series* dengan efek variasi kalender ditunjukkan oleh persamaan (21).

$$\begin{aligned} Z_t &= 2274,5 + 754M_{1t} - 83M_{2t} + 90M_{3t} - 564M_{4t} - 246M_{5t} + \dots \\ &\dots + (-552)M_{6t} - 624M_{7t} + 4887M_{8t} + 844M_{9t} - 63M_{10t} + \dots \\ &\dots + 421M_{11t} + 4532R_{11t} + (5,8141 - 0,27193Z_{t-3} + 0,36151Z_{t-5} + a_t) \end{aligned} \quad (21)$$

Model yang terbentuk memberikan informasi bahwa apabila pelayaran kapal penumpang rute Surabaya-Jayapura terjadi pada Bulan Agustus maka jumlah penumpang kapal rute Surabaya-Jayapura akan naik sebesar 7161,5 atau 7162 penumpang dibandingkan dengan pelayaran pada Bulan Desember. Selain itu diketahui pula bahwa apabila pelayaran dilakukan pada 1 bulan setelah Hari Raya Idul Fitri maka jumlah penumpang kapal rute Surabaya-Jayapura akan bertambah sebesar 6806,5 atau 6807 penumpang.

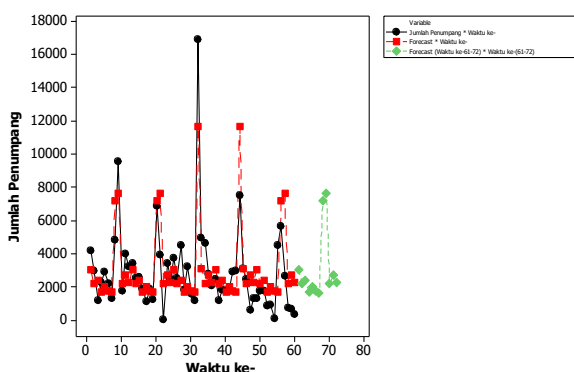
**F. Hasil Forecasting Menggunakan Regresi Time Series Efek Variasi Kalender**

Model pada persamaan (18) kemudian digunakan untuk meramalkan jumlah penumpang kapal rute Surabaya-Jayapura pada tahun 2017. Hasil yang diperoleh disajikan pada Tabel 10.

Tabel 10. Hasil Forecasting Regresi Time Series

Bulan	t	Forecast	Real Forecast
Januari	61	37,3485	3031,4
Februari	62	42,1265	2194,4
Maret	63	41,4321	2367,4
April	64	34,5145	1713,5
Mei	65	36,039	2031,5
Juni	66	22,543	1725,5
Juli	67	48,0771	1653,5
Agustus	68	52,4337	7164,5
September	69	44,0467	7653,5
Oktober	70	35,3637	2214,5
November	71	37,1628	2698,5
Desember	72	31,3028	2277,5

Keterangan : Tanda tebal merupakan bulan terjadinya Hari Raya Idul Fitri.



Gambar 5. Plot Ramalan Data Menggunakan Regresi Time Series Efek Variasi Kalender.

Hasil peramalan dengan metode ARIMA menunjukkan bahwa jumlah penumpang kapal rute Surabaya-Jayapura pada tahun 2017 terbesar bukan berada pada bulan terjadinya Hari raya Idul Fitri ataupun Tahun Baru Masehi. Nilai AIC yang dihasilkan dari ramalan ini adalah sebesar 181,9194. Plot hasil ramalan disajikan pada Gambar 5.

**G. Perbandingan Metode dengan Indikator Kebaikan Model**

Indikator kebaikan model dari kedua metode disajikan pada Tabel 11.

Tabel 11.  
Perbandingan Metode dengan Indikator Kebaikan Model

Model	MAPE Data Outsample	AIC
ARIMA (0,1,1)(0,0,1) <sup>12</sup>	62,155%	517,8454
Regresi Time Series dengan Pemodelan Residual Data Menggunakan AR ([3,5])	13,82%	181,9194

**V. KESIMPULAN DAN SARAN**

Kesimpulan yang bisa diambil dari penelitian ini adalah sebagai berikut.

1. Pengujian stasioneritas pada metode ARIMA memberikan hasil bahwa data membutuhkan transformasi dan differencing. Plot ACF dan PACF yang terbentuk menunjukkan bahwa lag yang signifikan pada plot ACF adalah lag 1 dan 12, sedangkan pada plot PACF adalah lag 1, 2, dan 12. Model terbaik yang terbentuk menggunakan model ARIMA (0,1,1)(0,0,1)<sup>12</sup>. Ramalan data *outsample* menghasilkan nilai MAPE sebesar 62,155%. Sedangkan AIC untuk data ramalan tahun 2017 yang dihasilkan adalah sebesar 517,8454.
2. Pemodelan regresi *time series* efek variasi kalender menunjukkan bahwa *dummy* Bulan Agustus dan *dummy* 1 bulan setelah Hari Raya Idul Fitri berpengaruh signifikan terhadap model. Pemeriksaan asumsi dari model yang terbentuk memberikan kesimpulan bahwa residual data tidak memenuhi asumsi identik dan independen namun memenuhi asumsi distribusi normal.
3. Pemodelan dilanjutkan dengan pemodelan residual data menggunakan metode ARIMA karena asumsi klasik pada regresi *time series* efek variasi kalender tidak terpenuhi. Pemodelan ARIMA menggunakan residual data hasil pemodelan efek variasi kalender memberikan hasil bahwa AR ([3,5]) merupakan model terbaik. Nilai MAPE untuk kesalahan ramalan data *outsample* adalah sebesar 13,82%. Sedangkan AIC untuk data ramalan tahun 2017 yang dihasilkan adalah sebesar 181,9194. Nilai MAPE dan AIC tersebut lebih kecil dibandingkan dengan ramalan menggunakan model ARIMA (0,1,1)(0,0,1)<sup>12</sup> pada data jumlah penumpang kapal rute Surabaya-Jayapura.

Saran dalam penelitian selanjutnya adalah mengambil rute pelayaran yang memiliki frekuensi lebih sering dalam 1 bulan. Selain itu, efek variasi kalender dapat dicoba menggunakan efek variasi kalender Hijriyah, bukan hanya Masehi.

**DAFTAR PUSTAKA**

[1] Badan Pusat Statistika, “Statistik Transportasi Laut 2017,” 2018.  
 [2] S. Dwijantari, “Peramalan Konsumsi Listrik Pada Segmen Rumah Tangga PT PLN (Persero) Distribusi Jawa Timur,” Institut Teknologi Sepuluh Nopember, 2017.  
 [3] N. Fanani, “Peramalan Jumlah Keberangkatan Penumpang

- Pelayaran Dalam Negeri dari Pelabuhan Tanjung Perak dengan ARIMA-Box Jenkins,” Institut Teknologi Sepuluh Nopember, 2017.
- [4] R. Endartyana, “Peramalan Jumlah Kedatangan Wisatawan Mancanegara di Indonesia Melalui Pintu Masuk Bandara Ngurah Rai Bali Menggunakan ARIMA Box-Jenkins,” Institut Teknologi Sepuluh Nopember, 2017.
- [5] W. Wei, *Time Series Analysis: Univariate and Multivariate Methods*. Pearson Addison Wesley, 2006.
- [6] R. Johnson and D. Wichern, *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hal, Inc, 2007.
- [7] J. Cryer and K. Chan, *Time Series Analysis with Application in R*, 2nd ed. New York: Springer Science Business Media, Inc, 2008.
- [8] H. Akaike, *Information Theory and an Extension of the Maximum Likelihood Principle*. 1978.