

PERAMALAN BANYAK PENUMPANG DARI PELABUHAN BELAWAN MENGGUNAKAN MODEL ARIMAX DENGAN EFEK VARIASI KALENDER

Arfiyansyah Rizki Effendi¹⁾, Etik Zukhronah²⁾, Sugiyanto³⁾

^{1) 2) 3)} Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Sebelas Maret
arfiyansyahrizki@gmail.com, etikzukhronah@staff.uns.ac.id,
sugiyanto61@staff.uns.ac.id

Abstrak

Peramalan merupakan salah satu upaya untuk memperkirakan apa yang akan terjadi pada masa mendatang berdasarkan data masa lalu. Banyak penumpang di Pelabuhan Belawan merupakan data runtun waktu. Pola data banyak penumpang di Pelabuhan Belawan sangat dipengaruhi oleh hari libur, sehingga cenderung mempunyai pola musiman. Mayoritas penduduk Indonesia beragama Islam yang mempunyai Hari Raya Idul Fitri. Penentuan Hari Raya Idul Fitri tidak mengikuti kalender masehi, tetapi berdasarkan kalender Hijriah. Perbedaan penggunaan kalender masehi dan hijriah pada suatu data runtun waktu menyebabkan adanya variasi kalender. Data yang mengandung variasi kalender dapat dianalisis menggunakan model ARIMAX. Penelitian ini bertujuan untuk memperoleh model ARIMAX terbaik dan meramalkan banyak penumpang dari Pelabuhan Belawan beberapa periode ke depan. Hasil penelitian menunjukkan bahwa model ARIMAX $(0,0,1) V_t, V_{t+1}, S_{s,t}; s=1,2,\dots,12$ merupakan model terbaik. Berdasarkan model tersebut dapat diramalkan dari bulan September 2018 sampai Agustus 2019 secara berturut turut adalah 5357, 2448, 1849, 5634, 12894, 4410, 4195, 3724, 4339, 10448, 13499, 4084 penumpang.

Kata kunci: ARIMAX, Pelabuhan Belawan, Peramalan, Variasi Kalender

1. PENDAHULUAN

Pelabuhan Belawan adalah pelabuhan yang terletak di Kota Medan, Sumatera Utara, Indonesia dan merupakan pelabuhan terpenting di pulau Sumatera. Cikal bakal lahirnya Pelabuhan Belawan adalah Labuhan Deli. Labuhan Deli dulunya merupakan pusat pemerintahan Kerajaan Deli yang kesohor di kawasan Sumatera Timur. Bandar Labuhan Deli terletak di tepi Sungai Deli.

Pada tahun 1915, pelabuhan Labuhan Deli dipindahkan ke Belawan yang terletak di tepi Sungai Belawan. Hal ini disebabkan Sungai Deli kian dangkal, sehingga menghambat kapal masuk alur Sungai Deli menuju Labuhan Deli. Lokasi pelabuhan adalah Belawan lama (saat ini masih berfungsi sebagai terminal kargo umum dan sebagian menjadi pangkalan TNI Angkatan Laut). Belanda membangun dermaga Belawan lama hingga mencapai panjang 602 meter dan lebar 9-20 meter. Oleh Belanda, dermaga Belawan lama dipergunakan untuk sandar berbagai jenis kapal, baik kapal kargo maupun kapal penumpang. Pada tahun 1938, Pelabuhan Belawan menjadi pelabuhan terbesar di wilayah Hindia Belanda (Airriess dkk., 1991).

Pola banyak penumpang di Pelabuhan Belawan sangat dipengaruhi hari libur, sehingga cenderung mempunyai pola musiman. Mayoritas penduduk

Indonesia beragama Islam yang mempunyai Hari Raya Idul Fitri. Penentuan Hari Raya Idul Fitri tidak mengikuti kalender masehi, tetapi berdasarkan kalender Hijriah. Perbedaan penggunaan kalender masehi dan hijriah pada suatu data runtun waktu menyebabkan adanya variasi kalender. Variasi kalender adalah penetapan Hari Raya Idul Fitri yang berubah-ubah pada kalender masehi karena pada kalender masehi Hari Raya Idul Fitri maju satu bulan setiap tiga tahun sekali.

Menurut Lee dkk. (2010) pemodelan pola data yang mempunyai efek variasi kalender dapat menggunakan metode *Autoregressive Integrated Moving Average Exogenous (ARIMAX)*. Variasi kalender dibagi menjadi dua yakni efek perdagangan dan efek liburan (Liu, 1986).

Pada penelitian ini digunakan model runtun waktu yang mengandung variasi kalender dengan efek liburan. Penelitian ini bertujuan untuk memperoleh model ARIMAX terbaik dan meramalkan banyak penumpang dari Pelabuhan Belawan. Pemodelan runtun waktu yang mengandung variasi kalender dapat dilakukan dengan menggunakan *ARIMAX*.

2. METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data banyak penumpang dari Pelabuhan Belawan dari bulan Januari 2009 sampai bulan Agustus 2018. Data penelitian ini merupakan data sekunder yang diperoleh dari website Badan Pusat Statistik (BPS) Indonesia.

2.2 Teori Penunjang

Analisis runtun waktu pertama kali diperkenalkan dan dikembangkan pada tahun 1970 oleh Box dan Jenkins. Runtun waktu adalah himpunan observasi terurut dalam waktu atau dalam dimensi lain.

Menurut sejarah nilai observasinya, runtun waktu dibedakan menjadi dua yaitu runtun waktu deterministik dan runtun waktu stokastik. Runtun waktu deterministik adalah suatu runtun waktu dimana keadaan yang akan datang dapat diramalkan secara pasti dan tidak perlu penyelidikan kembali. Runtun waktu stokastik adalah suatu runtun waktu dimana keadaan yang akan datang bersifat probabilistik, menurut observasi yang di masa lampau (Soejoeti, 1987).

Suatu runtun waktu dikatakan stasioner jika tidak terdapat kecenderungan peningkatan atau penurunan pada data tersebut dalam selang yang cukup panjang atau dengan kata lain fluktuasi data berada di sekitar rata-rata dan variansi konstan serta tidak tergantung pada waktu (Rosadi, 2010).

2.2.1 *Autoregressive (AR)*

Menurut Halim (2006) bentuk umum dari proses *autoregressive* tingkat p atau $AR(p)$ yaitu

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t$$

dengan

ϕ_p : parameter dari *autoregressive*

Y_t : runtun waktu ke- t

e_t : residu pada waktu ke- t

2.2.2 *Moving Average (MA)*

Menurut Halim (2006) bentuk umum dari proses *Moving Average* tingkat q atau $MA(q)$ yaitu

$$Y_t = \theta_1 Y_{t-1} - \theta_2 Y_{t-2} - \dots - \theta_q Y_{t-q} + e_t$$

dengan

θ_q : parameter dari *moving average*

2.2.3 *Autoregressive Moving Average (ARMA)*

Menurut Halim (2006) model umum untuk proses $ARMA(p, q)$ didefinisikan sebagai berikut.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 Y_{t-1} - \theta_2 Y_{t-2} - \dots - \theta_q Y_{t-q}$$

2.2.4 *Autoregressive Integrated Moving Average (ARIMA)*

Menurut Suhartono dkk. (2010) adapun model umum untuk $ARIMA$ nonmusiman adalah

$$\phi_p(B)Y_t(1-B)^d = \theta_q(B)e_t$$

dengan

$$\phi_p(B) = (1 - \phi_1 B^1 - \phi_2 B^2 - \dots - \phi_p B^p)$$

$$\theta_q(B) = (1 - \theta_1 B^1 - \theta_2 B^2 - \dots - \theta_q B^q)$$

p : orde untuk *autoregressive* (AR)

q : orde untuk *moving average* (MA)

d : banyaknya proses *differencing*

$(1 - B)^d$: operator *differencing* orde d

2.2.5 *Autoregressive Integrated Moving Average dengan Variabel Exogen (ARIMAX)*

Salah satu model runtun waktu yang dapat dipandang sebagai perluasan model runtun waktu $ARIMA$ adalah model $ARIMAX$. $ARIMAX$ adalah pemodelan pola data yang mempunyai efek variasi kalender.

Menurut Arunraj dkk. (2016), secara umum model $ARIMAX$ (p, d, q) dapat diberikan dengan persamaan berikut

$$\Phi_p(B)Y_t(1 - B)^d = \theta_q(B)e_t + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \dots + \alpha_k X_{k,t}$$

$X_{k,t}$: variabel independen ke- k pada saat t dengan $k = 1, 2, 3, \dots, k$

2.2.6 Model Regresi dengan Variabel *Dummy*

Regresi dalam konteks runtun waktu mempunyai bentuk umum yang sama dengan regresi linier. Pada analisis regresi, variabel respon sering dipengaruhi tidak hanya oleh variabel kuantitatif tapi juga oleh variabel kualitatif. Variabel kualitatif dengan sifat tertentu dapat diubah menjadi kuantitatif yaitu dengan cara mengkonstruksi variabel buatan bernilai 1 atau 0. Nilai 1 mengindikasikan adanya sifat dan 0 mengindikasikan tidak adanya sifat. Variabel yang mengasumsikan nilai 1 atau 0 dinamakan variabel *dummy*. Menurut Gujarati (2004) persamaan regresi variabel *dummy* tanpa konstanta dapat ditulis

$$Y_t = \beta_1 D_{1,t} + \beta_2 D_{2,t} + \beta_3 D_{3,t} + \dots + \beta_m D_{m,t} + \varepsilon_t$$

dengan β_i adalah parameter variabel *dummy* dan $D_{i,t}$ adalah variabel *dummy* efek liburan dengan $i = 1, 2, \dots, m$. Sedangkan persamaan regresi variabel *dummy* dengan konstanta dapat ditulis

$$Y_t = \beta_0 + \beta_1 D_{1,t} + \beta_2 D_{2,t} + \beta_3 D_{3,t} + \dots + \beta_{m-1} D_{m-1,t} + \varepsilon_t$$

dengan β_j adalah parameter variabel *dummy* dan $D_{j,t}$ adalah variabel *dummy* efek liburan dengan $j = 1, 2, \dots, m$

2.2.7 Uji Signifikasi Parameter Model Regresi Variabel *Dummy*

Untuk mengetahui apakah hasil estimasi parameter model variasi kalender signifikan atau tidak, dilakukan pengujian signifikasi parameter. Pengujian ini digunakan untuk mengetahui apakah setiap variabel yang digunakan berpengaruh pada Y_t atau tidak. Menurut Johnson dan Battacharya (1996) hipotesis yang digunakan adalah sebagai berikut

i. Uji Hipotesis:

$$H_0: \beta_i = 0 \text{ (parameter tidak signifikan)}$$

$$H_1: \beta_i \neq 0, i = 1, 2, \dots, k \text{ (parameter signifikan)}$$

ii. Daerah Kritis:

H_0 ditolak jika $|t_{hitung}| > t_{tabel}$, atau H_0 ditolak jika nilai $P_{value} < \alpha$ dengan α adalah tingkat signifikansi.

iii. Statistik Uji:

$$t = \frac{\hat{\phi}_j}{SE(\hat{\phi}_j)}$$

iv. Kesimpulan

2.2.8 Pemilihan Model Terbaik

Pada analisis data runtun waktu sering didapatkan beberapa model yang telah signifikan dengan memenuhi asumsi semua parameter, seperti sesatan *white noise* dan residu yang berdistribusi normal. Untuk menentukan model terbaik dalam peramalan data runtun waktu dapat digunakan kriteria *Root*

Mean Square Error (RMSE). Menurut Wei (2006) nilai RMSE untuk data *in-sample* dan *out-sample* dihitung dengan rumus seperti berikut.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n}}$$

dengan n menyatakan banyaknya data. Pemilihan model terbaik adalah model dengan RMSE yang terkecil.

2.3 Metode Analisis

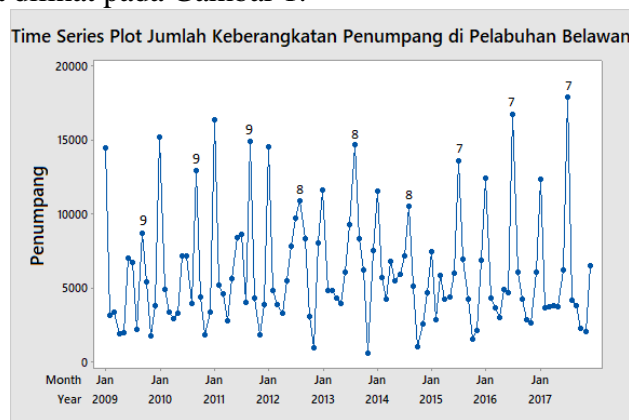
Langkah-langkah dalam penelitian ini dapat diuraikan sebagai berikut.

1. Membagi data menjadi data *in-sample* dan *out-sample*. Data *in-sample* digunakan sebagai pembentukan model, sedangkan data *out-sample* digunakan sebagai evaluasi model setelah model yang sesuai telah diperoleh. Dalam penelitian ini terdapat 116 data bulanan banyak penumpang dari bulan Januari 2009 sampai dengan bulan Agustus 2018. Data *in-sample* menggunakan 108 data observasi awal yang dimulai dari bulan Januari 2009 sampai bulan Desember 2017. Sedangkan data *out-sample* menggunakan 8 observasi yang dimulai dari bulan Januari 2018 sampai bulan Agustus 2018.
2. Menentukan variabel *dummy* untuk efek variasi kalender yaitu variabel *dummy* untuk efek libur Hari Raya Idul Fitri dan variabel *dummy* untuk efek musiman yang terdiri dari dua belas bulan mulai dari bulan Januari sampai bulan Desember.
 Nilai variabel *dummy* untuk efek Hari Raya Idul Fitri adalah V_t bernilai 1 untuk bulan yang terdapat Hari Raya Idul Fitri dan V_t bernilai 0 untuk bulan yang lainnya. Selanjutnya V_{t+1} bernilai 1 untuk bulan berikutnya yang terdapat Hari Raya Idul Fitri dan V_{t+1} bernilai 0 untuk bulan lainnya. Variabel *dummy* untuk efek musiman adalah $S_{1,t}$ yang bernilai 1 untuk bulan Januari dan bernilai 0 untuk bulan yang lainnya, $S_{2,t}$ yang bernilai 1 untuk bulan Februari dan bernilai 0 untuk bulan yang lainnya, dan seterusnya hingga $S_{12,t}$ yang bernilai 1 untuk bulan Desember dan bernilai 0 untuk bulan yang lainnya. Model regresi variabel *dummy* yang akan digunakan adalah model regresi variabel *dummy* tanpa konstanta dan model regresi variabel *dummy* dengan konstanta
3. Menghilangkan efek variasi kalender dan efek musiman pada data dengan melakukan regresi variabel *dummy* menggunakan metode *Ordinary Least Square* (OLS) dan uji signifikansi parameter menggunakan uji t. Apabila ada parameter yang tidak signifikan maka dilakukan estimasi ulang dengan tidak melibatkan parameter yang tidak signifikan tersebut sehingga diperoleh model regresi yang mempunyai parameter yang signifikan.
4. Melakukan pemodelan ARIMAX menggunakan plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) pada data residu model regresi.
5. Melakukan uji signifikansi parameter

6. Memilih model ARIMAX yang memiliki nilai RMSE *in-sample* dan RMSE *out-sampel* terkecil.
7. Melakukan uji asumsi *white noise* dengan menggunakan uji Ljung-Box dan uji asumsi distribusi normal menggunakan uji Kolgomorov-Smirnov pada residunya
8. Melakukan peramalan banyaknya keberangkatan penumpang di Pelabuhan Belawan menggunakan model terbaik yang telah didapatkan.

3. HASIL PENELITIAN DAN PEMBAHASAN

Plot runtun waktu banyak keberangkatan penumpang di Pelabuhan Belawan dapat dilihat pada Gambar 1.



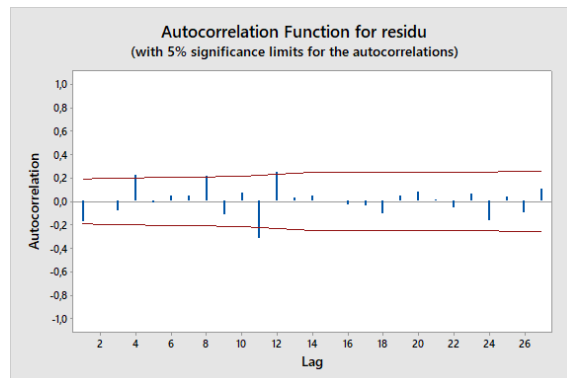
Gambar 1. Plot runtun waktu banyak keberangkatan penumpang di Pelabuhan Belawan

Gambar 1 menunjukkan bahwa bulan bulan tertentu pada setiap tahunnya mengalami kenaikan yang signifikan dibandingkan dengan bulan yang lain. Pada bulan Januari disetiap tahunnya mengalami kenaikan, hal ini dipengaruhi karena adanya liburan sekolah pada bulan Januari. Pada bulan September tahun 2009, 2010 dan 2011 juga mengalami kenaikan, selanjutnya pada bulan Agustus tahun 2012, 2013 dan 2014 mengalami kenaikan, kemudian diikuti bulan Juli tahun 2015, 2016 dan 2017 juga terjadi peningkatan banyak penumpang. Peningkatan yang terjadi pada bulan bulan tersebut terjadi karena bertepatan dengan adanya libur Hari Raya Idul Fitri. Adanya peningkatan yang signifikan dan berulang di tiga tahun ini karena Hari Raya Idul Fitri mengacu pada kalender hijriah yang banyak harinya selisih 11 hari dengan kalender masehi, sehingga jatuhnya bulan pada Hari Raya Idul Fitri berubah setiap tiga tahun sekali. Dapat dilihat juga bahwa data telah stasioner dalam rata rata sehingga tidak perlu dilakukan proses *differences*.

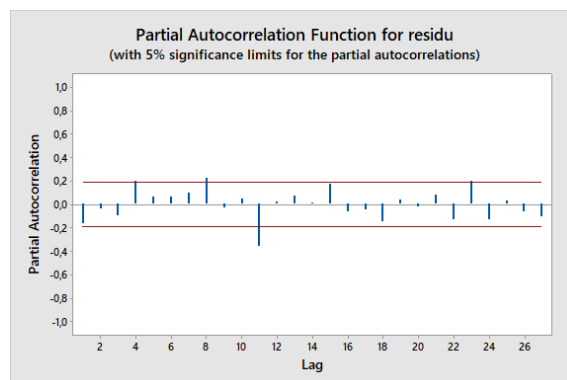
3.1 Model Regresi Variabel *Dummy* Tanpa Konstanta

Hasil regresi variabel *dummy* tanpa konstanta menunjukkan bahwa semua variabel yang digunakan telah signifikan, selanjutnya akan dilakukan pemodelan ARIMAX menggunakan plot ACF dan PACF residu regresi

variabel *dummy* tanpa konstanta. Pola ACF dan PACF data disajikan pada Gambar 2 dan Gambar 3.



Gambar 2. Plot ACF Regresi Variabel *Dummy* Tanpa Konstanta

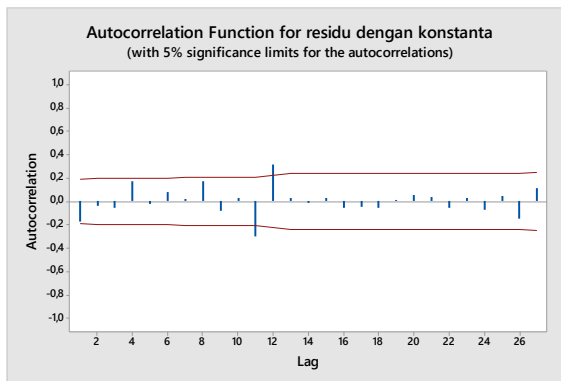


Gambar 3. Plot PACF Regresi Variabel *Dummy* Tanpa Konstanta

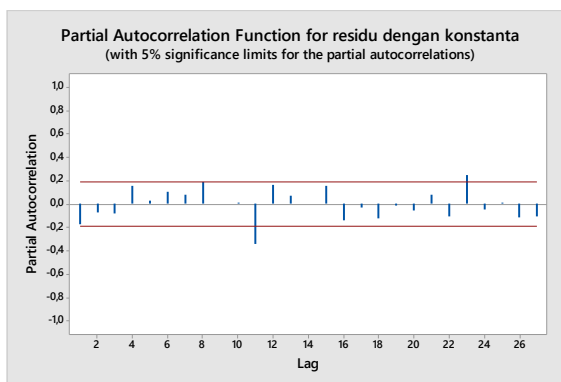
Gambar 2 menunjukkan bahwa lag ke 4, 8, 11 dan 12 keluar dari batas pita konfidensi, sedangkan pada Gambar 3 menunjukkan lag ke 4, 8, 11, 12 dan 23 keluar dari batas pita konfidensi. Berdasarkan hal tersebut dugaan model ARIMA sementara yang didapatkan adalah ARIMA (4,0,4), ARIMA (4,0,0), ARIMA (0,0,4). Hasil model ARIMAX yang diperoleh pada model regresi variabel *dummy* tanpa konstanta dengan parameter signifikan adalah model ARIMAX (1,0,0) $V_t, V_{t+1}, S_{s,t}; s=1,2,\dots,12$ dan ARIMAX (0,0,1) $V_t, V_{t+1}, S_{s,t}; s=1,2,\dots,12$

3.2 Model Regresi Variabel *Dummy* dengan Konstanta

Hasil regresi *dummy* dengan konstanta menunjukkan bahwa variabel yang signifikan adalah $V_t, V_{t+1}, S_{1,t}, S_{7,t}, S_{10,t}, S_{11,t}$ selanjutnya akan dilakukan pemodelan ARIMAX menggunakan plot ACF dan PACF residu regresi variabel *dummy* dengan konstanta. Pola ACF dan PACF data disajikan pada Gambar 4 dan Gambar 5.



Gambar 4. Plot ACF Regresi Variabel *Dummy* dengan Konstanta



Gambar 5. Plot PACF Regresi Variabel *Dummy* dengan Konstanta

Gambar 4 menunjukkan bahwa lag ke 11 dan 12 keluar dari batas pita konfidensi, sedangkan pada Gambar 5 menunjukkan lag ke 11 dan 23 keluar dari batas pita konfidensi. Berdasarkan hal tersebut dugaan model ARIMA sementara yang didapatkan adalah ARIMA (0,0,0) Hasil model ARIMAX yang diperoleh pada model regresi variabel *dummy* dengan konstanta dengan parameter adalah model ARIMAX (0,0,0) $V_t, V_{t+1}, S_{s,t}; s=1,7,10$ dan 11

Perbandingan nilai RMSE untuk setiap model disajikan pada Tabel 1.

Tabel 1 Perbandingan Nilai RMSE *In-sample* dan *Out-sample*

Model	RMSE <i>in-sample</i>	RMSE <i>out-sample</i>
ARIMAX (1,0,0) $V_t, V_{t+1}, S_{s,t}; s=1,2,\dots,12$	1865	1944
ARIMAX (0,0,1) $V_t, V_{t+1}, S_{s,t}; s=1,2,\dots,12$	1863	1943
ARIMAX (0,0,0) $V_t, V_{t+1}, S_{s,t}; s=1,7,10,11$	2012	1884

Tabel 1 menunjukkan bahwa berdasarkan nilai RMSE *in-sampel* dan *out-sample* terkecil, model ARIMAX terbaik untuk peramalan banyaknya

penumpang di Pelabuhan Belawan adalah model *ARIMAX* $(0,0,1) V_t, V_{t+1}, S_{s,t}; s=1,2,\dots,12$.

3.3 Uji Asumsi *White Noise* dan uji asumsi distribusi normal pada residu model terbaik

Uji *white noise* dilakukan menggunakan uji Ljung-Box. Hasil uji Ljung-Box menunjukkan bahwa nilai-p pada lag ke 12, 24, 36 dan 48 secara berturut turut adalah 0,147 , 0,237 , 0,069 , 0,058 yang nilainya lebih besar dari $\alpha = 0,05$. Berdasarkan hasil tersebut dapat disimpulkan bahwa residu memenuhi asumsi *white noise*. Hasil uji asumsi distribusi normal menggunakan uji Kolgomorov-Smirnov diperoleh nilai-p sebesar 0,6024 yang nilainya lebih besar dari $\alpha = 0,05$. Berdasarkan hasil tersebut dapat disimpulkan bahwa residu memenuhi asumsi distribusi normal.

4 SIMPULAN

Berdasarkan hasil pembahasan dapat disimpulkan bahwa

1. Model variasi kalender untuk peramalan banyaknya penumpang dari Pelabuhan Belawan dengan pendekatan model *ARIMAX* adalah model *ARIMAX* $(0,0,1) V_t, V_{t+1}, S_{s,t}; s=1,2,\dots,12$ dengan persamaan

$$Y_t = 4327v_t + 4587v_{t+1} + 12894S_{1,t} + 4407S_{2,t} + 4196S_{3,t} + 3725S_{4,t} + 4340S_{5,t} + 6141S_{6,t} + 8855S_{7,t} + 4096S_{8,t} + 5371S_{9,t} + 2462S_{10,t} + 1842S_{11,t} + 5640S_{12,t} + (1 - 0,2031B)\epsilon_t$$

2. Nilai peramalan dengan model terbaik akan disajikan dalam Tabel 2.

Tabel 2. Nilai peramalan pada bulan September 2018 hingga Agustus 2019

Bulan	Penumpang
September	5357
Oktober	2448
November	1849
Desember	5634
Januari	12894
Februari	4410
Maret	4195
April	3724
Mei	4339
Juni	10448
Juli	13499
Agustus	4084

5 DAFTAR PUSTAKA

- Airriess, C. A. 1991. *Global economy and port morphology in Belawan, Indonesia*. Geographical Review 81.
- Arunraj, N. S., Ahrens D. and Fernandes M. 2016. Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry. *International Journal of Operations Research and Information Systems (IJORIS)*, vol. 7, issue 2, 1-21
- BPS. 2018. Diakses dari <https://bps.go.id/site/pilihdata>
- Gujarati. 2004. *Basic Econometrics*. Edisi 4. The Mc-Graw Hill Companies. USA.
- Halim, S. 2006. *Diktat Time Series Analysis*. Surabaya.
- Lee, M.H., Suhartono, and Hamzah, N. A. 2010. *Calender Variation Model Based on ARIMAX for Forecasting Sales Data with Ramadhan Effect*, Proceedings of the Regional Conference on Statistical Sciences, 349-361.
- Liu, L. 1986. Identification of Time Series Models in the Presence of calender Variation. *International Journal and Forecasting*, 357-372
- Rosadi, D. 2010. *Analisis Ekonometrika dan Runtun Waktu Terapan dengan R*. ANDI Yogyakarta: Yogyakarta.
- Soejoeti, Z. 1987. *Analisis Runtun Waktu*. Karunika: Jakarta.
- Suhartono, Lee, M.H., and Hamzah, N. A. 2010. *Calender Variation Model Based on Time Series Regression for Sales Forecast The Ramadhan Effects*, Proceedings of the Regional Conference on Statistical Sciences, 30-40.
- Wei, W. 2006. *Time Series Analysis: Univariate and Multivariate Methods*. Pearson Education inc. New York.