PERAMALAN JUMLAH PENGUNJUNG MAKAM BALAKAN DI SUKOHARJO MENGGUNAKAN SARIMA DAN SARIMAX

Lindha Wulandari[1], Etik Zukhronah[2] dan Sri Sulistijowati Handajani[3]
[1]Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret
email : lindhawulandari2@student.uns.ac.id
[2]Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret
email : etikzukhronah@staff.uns.ac.id
[3]Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret
email : rr_ssh@staff.uns.ac.id

ABSTRAK

Makam Balakan merupakan salah satu wisata di daerah Sukoharjo yang sering dikunjungi karena tertarik dengan budaya dan kepercayaan di tempat tersebut. Banyak pengunjung yang datang pada saat acara kirab pulung langse yang dilaksanakan pada bulan Muharram. Penetapan tahun baru pada kalender Hijriah selalu mengalami pergeseran atau maju sebelas hari jika dilihat dari kalender Masehi. Hal tersebut mengakibatkan adanya variasi kalender. Adanya efek variasi kalender tersebut maka data jumlah pengunjung Makam Balakan dapat dimodelkan dengan SARIMA dan SARIMAX. Tujuan dari penelitian ini untuk memperoleh model terbaik dan untuk meramalkanjumlah pengunjung Makam Balakan. Penelitian ini menggunakan data pengunjung Makam Balakan dari bulan Januari 2010 sampai bulan Desember 2018 sebagai data in-sample dan bulan Januari 2019 sampai bulan September 2019 sebagai data out-sample. Hasil penelitian menunjukkan bahwa model SARIMAX$(0,0,0)(1,0,0)^{12}V_t,S_{6,t}$adalah model terbaik dengan RMSE out-sample sebesar 381,13. Peramalan pengunjung Makam Balakandari bulan Oktober 2019 hingga bulan Mei 2020 adalah 1576, 1551, 1523, 1583, 1605, 1578, 1600, dan 1578 pengunjung.

***Kata Kunci**: pengunjung Makam Balakan, RMSE, SARIMA, SARIMAX, variasi kalender.*

1. PENDAHULUAN

Indonesia merupakan sebuah negara yang kaya akan bahasa, pulau, wisata alam, dan juga budayanya. Beberapa tempat wisata di Indonesia masih kental akan budaya dan kepercayaan yang diyakini oleh masyarakat sekitar. Oleh karena itutempat tersebut dipenuhi pengunjung dengan maksud tertentu dan pada waktu tertentu seperti musim libur lebaran dan acara keagamaan daerah setempat.Di Kabupaten Sukoharjo terdapat tempat wisata religi yaitu Makam Balakan atau Makam Kyai Ageng Balak. Makam ini tepatnya terletak di Desa Mertan Kecamatan Bendosari Kabupaten Sukoharjo. Kyai Ageng Balak dipercayai oleh masyarakat setempat sebagai keturunan dari raja Majapahit yang dikenal sebagai manggala yuda peperangan dengan memiliki berbagai macam kemampuan. Pengunjung yang datang ke makam tersebut rata-rata bertujuan untuk memohon agar dirinya berhasil dalam menjalankan usaha maupun dalam bersaing mendapatkan posisi atau jabatan. Beragam bentuk permohonan para pengunjung di makam tersebut seperti mandi atau kungkum dan berdoa hingga menjelang fajar.

Di Makam Kyai Ageng Balak terdapat acara rutin yang dilakukan masyarakat yaitu kirab pulung langse pada bulan Suro atau bulan Muharram, tepatnya dilaksanakan setelah tahun baru Islam. Ritual tersebut berupa acara mengganti kain penutup makam kemudian mencuci di sungai Ranjing dekat dengan makam Kyai Ageng Balak. Ritual pulung langse ini juga disertakan acara gunungan berupa buah dan hasil panen yang ditumpuk dan diperebutkan oleh para pengunjung makam tersebut. Oleh karena itu pada saat acara pulung langse diadakan, makam

Kyai Ageng Balak dipenuhi pengunjung dari berbagai daerah sehingga terjadi kenaikan pada setiap bulan tahun baru Islam. Perbedaan perhitungan di dalam kalender Hijriah dan kalender Masehi mengakibatkan bulan Muharram atau tahun baru Islam mengalami pergeseran setiap tahunnya. Hal tersebut mengakibatkan adanya variasi kalender.Variasi kalender dapat disebabkan karena adanya efek hari kerja dan efek liburan seperti libur karena hari besar agama ataupun kebudayaan tertentu dari bulan ke bulan hingga tahun ke tahun (Lee *et al.*, 2010).

Data runtun waktu dapat dimodelkan dengan menggunakan *SARIMA*. Adanya variasi kalender yang terdeteksi pada data runtun waktu dapat dimodelkan dengan menggunakan *SARIMAX*. Cools *et al.* (2009) membandingkan *SARIMA*, *SARIMAX* dan *ARIMAX* pada data perhitungan lalu lintas harian dengan memperhatikan efek variasi kalender hari libur yang dirayakan di Belgia dan efek *day-of-week*.Arunraj *et al.* (2016) membandingkan *SARIMA* dengan *SARIMAX* pada data penjualan harian di industry ritel makanan yang memperhatikan efek variasi kalender hari libur di German dan efek musiman *day-of-week*. Pada paper ini dilakukan peramalan jumlah pengunjung Makam Balakan menggunakan *SARIMA* dan *SARIMAX*.

2. KAJIAN LITERATUR

2.1 Model *SARIMA*

SARIMA merupakan model *ARIMA* yang mengandung efek musiman atau *seasonal*. Berdasarkan Shumway dan Stoffer (2011), rumus *SARIMA(p, d, q)(P,D,Q)^S* adalah

$$\emptyset_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\Theta_Q(B^s)w_t$$

dengan

$$\emptyset_p(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p$$

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

$\emptyset_p(B)$:operator untuk parameter*autoregressive* non *seasonal* orde *p*
$\Phi_P(B^s)$:operator untuk parameter *autoregressive seasonal* orde *P*
$\theta_q(B)$:operator untuk parameter*moving average* non *seasonal* orde *q*
$\Theta_Q(B^s)$:operator untuk parameter *moving average seasonal* orde *Q*
$(1-B)^d$: operator untuk*differencing* non *seasonal* orde *d*
$(1-B^s)^D$:operator untuk*differencingseasonal* orde *D*
Z_t : runtun waktu ke-t
p : orde untuk *autoregressive*
q : orde untuk *moving average*

2.2 Uji Stasioner

Uji stasioner dapat dilakukan dengan menggunakan uji *Augmented* Dickey Fuller(ADF). Menurut Wei (2006) uji ADF menggunakan hipotesis

(1) $H_0 : \delta = 1$ (data tidak stasioner)
$H_1 : \delta \neq 1$ (data stasioner)
(2) α = 0,05
(3) Daerah kritis : H_0 ditolak jika $|t| \geq |t_{(n-1,\alpha)}|$ atau nilai p<α
(4) Statistik uji

$$t = \frac{\delta - 1}{SE(\delta)}, SE(\delta) = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n(n-1)}}$$

(5) Kesimpulan

2.3 *SARIMAX*

SARIMAX adalah perluasan dari model *SARIMA* yang mengandung variabel *exogenous* dengan bentuk umum *SARIMAX(p,d,q)(P,D,Q)*^{S}. Rumus *SARIMAX* dapat ditulis dengan (Arunraj *et al.*,2016)

$$Z_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_k X_{k,t} + \frac{\theta_q(B)\Theta_Q(B^s)}{\emptyset_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D} w_t$$

$X_{k,t}$ merupakan variabel *dummy*ke-*k* pada saat *t* dengan *k* = 0,1,2,3,...,*k*. β_k merupakan parameter dari variabel *dummy*.

2.4 Uji Signifikansi

Menurut Montgomery *et al.* (2015), hipotesis untuk menguji signifikansi parameter adalah

(1) $H_0 : \beta_j = 0$ (parameter tidak signifikan)
$H_1 : \beta_j \neq 0$ (parameter signifikan)
(2) α = 0,05
(3) Daerah kritis : H_0 ditolak jika $\left|t_{hitung}\right| \geq \left|t_{(n-p;\alpha/2)}\right|$atau nilai p<$\alpha$
(4) Statistik uji

$$t_{hitung} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}, SE(\hat{\beta}_j) = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}}$$

C_{jj} : elemen diagonal dari matrix (X'X)$^{-1}$
(5) Kesimpulan

2.5 Uji Autokorelasi

Uji autokorelasi dilakukan untuk mendeteksi ada tidaknya korelasi antar lag. Menurut Hanke dan Wichern (2014), uji autokorelasi menggunakan uji L-Jung Box dengan hipotesis seperti berikut

(1) $H_0 : \rho = 0$(tidak ada korelasi antar lag atau residu memenuhi asumsi non autokorelasi)
$H_1 : \rho > 0$ (ada korelasi antar lag atau residu tidak memenuhi asumsi non autokorelasi)
(2) α = 0,05
(3) Daerah kritis : H_0 ditolak jika Q> $X^2{}_{(\alpha,m)}$atau nilai p<α
(4) Statistik uji

$$Q = n(n+2)\sum_{k=1}^{m}(n-k)^{-1}r_k^2$$

r_k merupakan autokorelasi residu pada *lag* ke-k, dan m adalah jumlah maksimum *lag*
(5) Kesimpulan

2.6 Uji Normalitas

Metode yang digunakan untuk melakukan uji normalitas adalah uji Kolmogorov-Smirnov. Menurut DeGroot dan Schervish (2012), uji Kolmogorov-Smirnov menggunakan hipotesis seperti berikut

(1) $H_0 : F_n(x) = F(x)$ (data berdistribusi normal)
$H_1 : F_n(x) \neq F(x)$(data tidak berdistribusi normal)
(2) α = 0,05
(3) Daerah kritis : H$_0$ ditolak jika $D_{hitung} > D_{(1-\alpha,n)}$atau nilai p <$\alpha$
(4) Statistik uji

$$D_{hitung} = Max\,|F_n(x) - F(x)|$$

$F_n(x)$ adalah fungsi distribusi kumulatif normal dan $F(x)$ adalah fungsi distribusi kumulatif dari data
(5) Kesimpulan

2.7 Ukuran Akurasi

Dalam menentukan suatu model terbaik perlu dilihat juga ukuran akurasi atau seberapa besar kesalahan menggunakan model yang diuji. Untuk menentukan model terbaik dipilih ukuran akurasi yang paling kecil. Ukuran akurasi yang digunakan adalah *Root Mean Square Error (RMSE)*. Menurut Ermayanthi dkk. (2012), perhitungan *RMSE in-sample* dan *RMSE out-sample* adalah

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

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

3. METODE PENELITIAN

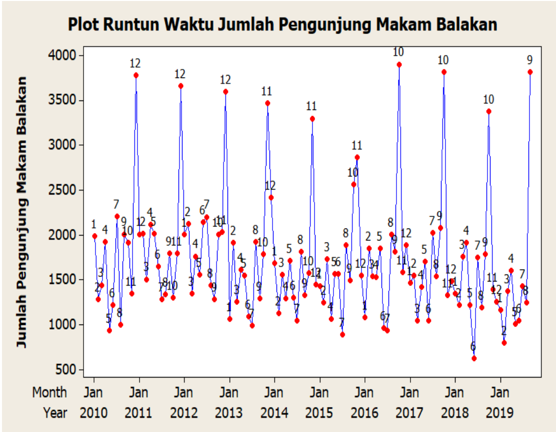
Data yang digunakan dalam penelitian ini adalah data jumlah pengunjung wisata Makam Balakan di Sukoharjo dari bulan Januari 2010 hingga September 2019 yang diperoleh dari Dinas Pendidikan dan Kebudayaan Sukoharjo dalam satuan ribu orang. Pada penelitian ini digunakan data *in-sample* sebanyak 109 data dari bulan Januari 2010 sampai bulan Desember 2018untuk menentukan model. Sedangkan data *out-sample* untuk menguji model menggunakan 9 data dari bulan Januari 2019 sampai bulan September 2019. Langkah-langkah yang dilakukan adalah sebagai berikut,

1. Melakukan uji stasioneritas untuk melihat apakah data sudah stasioner ataukah belum. Apabila belum stasioner perlu dilakukan *differencing*.

2. Membuat plot *Autocorrelation Function(ACF)* dan plot *Partial Autocorrelation Function (PACF)* dari data yang telah stasioner untuk menentukan model *SARIMA*.
3. Menentukan variabel *dummy* untuk efek variasi kalender menggunakan efek perayaan tahun baru Islam. Nilai *dummy* untuk bulan yang terdapat perayaan tahun baru Islam (V_t) adalah 1 dan 0 untuk yang lainnya. Untuk variabel *dummy* efek musiman bernilai 1 untuk bulan Januari ($S_{1,t}$) dan 0 untuk yang lainnya begitu seterusnya sampai bulan Desember ($S_{12,t}$) bernilai 1 dan 0 untuk lainnya.
4. Melakukan pemodelan regresi variabel *dummy* dan estimasi parameter serta menguji signifikansi parameter.
5. Melakukan pemodelan *SARIMAX* menggunakan model *SARIMA* yang telah diperoleh pada langkah 2 dengan menambahkan variabel *dummy* yang telah signifikan.
6. Melakukan estimasi parameter dari model *SARIMA* dan *SARIMAX* serta dilakukan uji signifikansi model tersebut.
7. Uji diagnostik model dengan uji autokorelasi residu menggunakan uji L-Jung Box dan uji normalitas menggunakan uji Kolmogorov-Smirnov.
8. Memilih model *SARIMA* dan *SARIMAX* terbaik berdasarkan nilai *RMSE out-sample* terkecil.
9. Melakukan peramalan menggunakan model terbaik.

4. HASIL DAN PEMBAHASAN

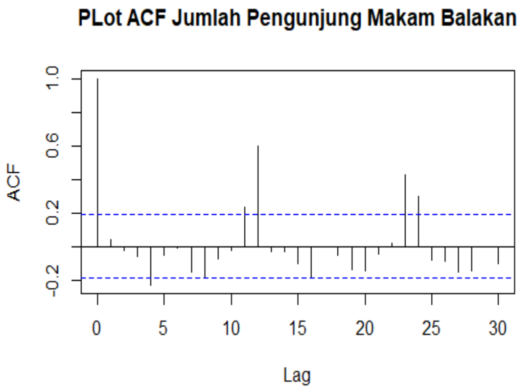
Plot data jumlah pengunjung Makam Balakan di Sukoharjo dari Januari 2010 sampai September 2019 dapat dilihat pada Gambar 1.



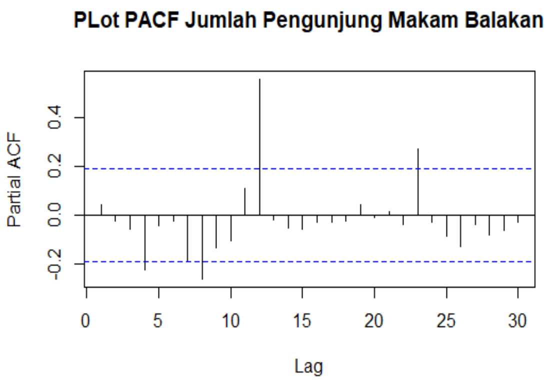
Gambar 1. Plot Runtun Waktu Jumlah Pengunjung Makam Balakan Sukoharjo

Gambar 1 menunjukkan bahwa terjadi kenaikan yang siginifikan pada bulan tertentu di setiap tahunnya. Kenaikan tersebut terjadi pada bulan Desember 2010, Desember 2011, Desember 2012, November 2013, November 2014, November 2015, Oktober 2016, Oktober 2017, Oktober 2018, dan September 2019. Bulan tersebut bertepatan dengan bulan Muharram dimana terjadinya Tahun Baru Islam. Adanya pergeseran bulan Muharram setiap tahunnya mengakibatkanterjadinya variasi kalender. Oleh karena itu data jumlah pengunjung Makam Balakan tersebut terdapat efek variasi kalender.

Gambar 1 terlihat bahwa plot cenderung membentuk pola horizontal atau stasioner karena data observasi berfluktuasi di sekitarrata-rata konstan. Kemudian dilakukan uji stasioner menggunakan uji *ADF* untuk membuktikan bahwa data tersebut stasioner. Hasil uji *ADF* menunjukkan bahwa nilai t adalah 5,8694lebih besar dari nilai $t_{(108;0,05)}$adalah 1,6449 dan nilai p adalah 0,01 kurang dari α, sehingga diperoleh kesimpulan bahwa data jumlah pengunjung Mkam Balakan di Sukoharjo telah stasioner dan tidak perlu dilakukan *differencing*. Setelah data stasioner kemudian membentuk plot *ACF* dan *PACF* dari data.



Gambar 2. PLot *ACF* Data Jumlah Pengunjung Makam Balakan di Sukoharjo



Gambar 3. Plot *PACF* Data Jumlah Pengunjung Makam Balakan di Sukoharjo

Gambar 2 terlihat bahwa lag ke 0, 4, 12, dan 24 keluar dari pita konfidensi *ACF*dan pada gambar 3 terlihat bahwa lag ke 4 dan 12 keluar dari pita konfidensi *PACF*. Hal tersebut terlihat bahwa terdapat unsur musiman pada plot tersebut. Sehingga dugaan model *SARIMA* adalah *SARIMA(4,0,4)(1,0,2)*^{12}, *SARIMA(4,0,0)(1,0,2)*^{12}, *SARIMA(0,0,4)(1,0,2)*^{12}, *SARIMA(0,0,0)(1,0,2)*^{12}, *SARIMA(4,0,4)(1,0,0)*^{12}, *SARIMA(4,0,0)(1,0,0)*^{12}, *SARIMA(0,0,4)(1,0,0)*^{12}, *SARIMA(0,0,0)(1,0,0)*^{12}, *SARIMA(4,0,4)(0,0,2)*^{12}, *SARIMA(4,0,0)(0,0,2)*^{12}, *SARIMA(0,0,4)(0,0,2)*^{12},dan *SARIMA(0,0,0)(0,0,2)*^{12}.

Kemudian melakukan pemodelan regresi dari variabel *dummy* yang akan digunakan yaitu variabel *dummy*V_t untuk efek variasi kalender dan variabel *dummy*$S_{p,t}$dengan p : 1, 2, …, 12 untuk efek musiman. Setelah dilakukan uji signifikansi parameter variabel *dummy* diperoleh hasil bahwa tanpa memasukkan konstanta semua parameter variabel *dummy* yaitu $V_t$dan $S_{p,t}$dengan p : 1, 2, …, 12 telah signifikan sedangkan dengan memasukkan konstanta diperoleh hasil bahwa yang telah signifikan hanya parameter variabel *dummy*V_t, $S_{6,t}$, dan konstanta.

Untuk memperoleh model *SARIMAX* digunakan model *SARIMA* yang telah diperoleh sebelumnya dengan menambahkan variabel *dummy* yang telah signifikan. Kemudian melakukan uji signifikansi model *SARIMA* dan *SARIMAX*. Hasil menunjukkan bahwa model *SARIMA* dan *SARIMAX* yang telah signifikantanpa memasukkan konstanta adalah *SARIMA(0,0,0)(1,0,0)*^{12}, *SARIMA(0,0,0)(0,0,2)*^{12}, *SARIMA(4,0,0)(0,0,2)*^{12}, *SARIMA(0,0,4)(0,0,2)*^{12}, *SARIMAX(0,0,0)(1,0,2)*^{12} V_t, $S_{p,t}$, p : *1, 2, …, 12*, dan *SARIMAX(4,0,4)(0,0,2)*^{12} V_t, $S_{p,t}$, p : *1, 2, …, 12*. Sedangkan dengan memasukkan konstanta, model yang signifikan baik parameter maupun konstantanya adalah *SARIMA(0,0,0)(1,0,0)*^{12}, *SARIMA(0,0,0)(0,0,2)*^{12}, dan *SARIMAX(0,0,0)(1,0,0)*^{12} V_t, $S_{6,t}$.

Model yang baik adalah model yang residunya memenuhi asumsi non autokorelasi, berdistribusi normal serta memiliki nilai *RMSE* terkecil. Model yang residunya telah memenuhi asumsi non autokorelasi serta berdistribusi normal untuk tanpa memasukkan konstanta adalah *SARIMA(0,0,0)(1,0,0)*^{12}, *SARIMA(0,0,0)(0,0,2)*^{12}, *SARIMA(4,0,0)(0,0,2)*^{12}, *SARIMAX(0,0,0)(1,0,2)*^{12} V_t, $S_{p,t}$, p : *1, 2, …, 12*, dan *SARIMAX(4,0,4)(0,0,2)*^{12} V_t,

$S_{p,t}$, *p : 1, 2, …, 12*. Sedangakan dengan memasukkan konstanta, model yang residunya telah memenuhi asumsi non autokorelasi dan berdistribusi normal adalah *SARIMA(0,0,0)(1,0,0)*^{12}, *SARIMA(0,0,0)(0,0,2)*^{12}, *dan SARIMAX(0,0,0)(1,0,0)*^{12} V_t, $S_{6,t}$.

Untuk mengetahui model yang paling baik diantara model yang telah memenuhi uji diagnostik model perlu dilakukan perbandingan nilai *RMSE*, tabel 1 merupakan perbandingan niali *RMSE* model.

Tabel 1. Perbandingan nilai RMSE model

Model	*RMSE in-sample*	*RMSE out-sample*
Tanpa Konstanta		
SARIMA(0,0,0)(1,0,0)^{12}	512.95	745.73
SARIMA(0,0,0)(0,0,2)^{12}	739.86	1054.43
SARIMA(4,0,0)(0,0,2)^{12}	537.81	846.34
SARIMAX(0,0,0)(1,0,2)^{12} V_t, $S_{p,t}$, *p : 1, 2, …, 12*	306.61	404.34
SARIMAX(4,0,4)(0,0,2)^{12}V_t, $S_{p,t}$, *p : 1, 2, …, 12*	286.01	471.73
Dengan Konstanta		
SARIMA(0,0,0)(1,0,0)^{12}	479.00	749.15
SARIMA(0,0,0)(0,0,2)^{12}	481.16	785.58
SARIMAX(0,0,0)(1,0,0 ^{12} V_t, $S_{6,t}$	356.07	367.50

Tabel 1 menunjukkan bahwa model *SARIMAX (0,0,0)(1,0,0)* ^{12} V_t, $S_{6,t}$memiliki nilai *RMSE out-sample* terkecil. Oleh karena itu model tersebut adalah model terbaik untuk meramalkan jumlah pengunjung Makam Balakan di Sukoharjo. Hasil peramalan jumlah pengunjung Makam Balakan untuk bulan Oktober 2019-Mei 2020 pada Tabel 2.

Tabel 2. Hasil Peramalan Pengunjung

Bulan	Jumlah Penumpang
Oktober 2019	1576
September 2019	1551
Desember 2019	1523
Januari 2020	1583
Februari 2020	1605
Maret 2020	1578
April 2020	1600
Mei 2020	1578

5. KESIMPULAN

Berdasarkan pada analisis yang telah dilakukan diperoleh model terbaik untuk meramalkan jumlah pengunjung Makam Balakan di Sukoharjo yaitu *SARIMAX(0,0,0)(1,0,0)* ^{12} V_t, $S_{6,t}$ dengan persamaan

$$Z_t = 1593{,}17 + 1859{,}82\,V_t - 329{,}08\,S_{6,t} + \frac{1}{(1 - 0{,}20289B^s)}w_t$$

Hasil peramalan pengunjung Makam Balakan dari bulan Oktober 2019 hingga bulan Mei 2020 menggunakan model terbaik adalah 1576, 1551, 1523, 1583, 1605, 1578, 1600, dan 1578 pengunjung

6. REFERENSI

Arunraj, N., Ahrens, D., and Fernandes, M. 2016. Applications of SARIMAX Model to Forecast Daily Sales in Food Retail Industry. *International Journal of Operations Research and Information Systems*. 7(1), 1-21. Germany.

Cools, M., Elke, M., and Wets, G. 2009. Investigating The Variability in Daily Traffic Counts Using ARIMAX and SARIMAX Models : Assessing Impact of Holidays on Two Divergent Site Locations. *Journal of Transportation Research Board.* 2136(1), 57-66, Canada.

DeGroot, M., and Schervish, M. 2012. *Probability and Statistics, 4th edition.* Addison-Wesley. Boston MA, US.

Ermayanthi, N. M. D., Widodo, D. A., dan Suhartono. 2012. Peramalan Penjualan Buah di Moena Fresh Bali dengan Menggunakan Model Variasi Kalender. *Jurnal Sains dan Seni ITS Vol.1 (Sept, 2012) ISSN : 2301-928X*. Surabaya, Indonesia.

Hanke, J. E. and Wichern, D. 2014. *Business Forecasting Ninth Edition*. Pearson Education Limited. United Kingdom.

Montgomery, D. C., Jennings, C. L., and Kulahci, M. 2015. *Introduction to Time Series Analysis and Forecasting: Second Edition.* John Wiley and Sons Inc. Canada.

Shumway, R. H. and Stoffer, D. S. 2011. *Time Series Analysis and Its Applications with R Examples: Third Edition*. Springer Science Business Media LLC. London.

Wei, W. W. S. 2006. *Time Series Analysis, Univariate and Multivariate Methods*. Addison Wesley Publishing Company. Canada