

Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)
Volume 10, No. 4 (2021), hal 379 – 388.

PREDIKSI DATA JUMLAH PENUMPANG KERETA DENGAN EFEK VARIASI KALENDER PADA MODEL SARIMAX

Hayati, Naomi Nesyana Debataraaja, Shantika Martha

INTISARI

Kereta api menjadi salah satu transportasi alternatif masyarakat untuk melakukan aktivitasnya dalam hal pekerjaan maupun pariwisata untuk jarak tempuh yang jauh. Untuk menjaga kenyamanan penumpang, PT Kereta Api Indonesia (KAI) harus terus meningkatkan sarana dan prasarananya. Oleh karena itu, diperlukan prediksi jumlah penumpang untuk membantu PT KAI mempersiapkan fasilitas-fasilitas serta mengantisipasi kenaikan jumlah penumpang. Penelitian ini bertujuan memprediksi jumlah penumpang kereta dengan model SARIMAX. Data yang digunakan yaitu data jumlah penumpang kereta pada tahun 2014 hingga 2018. Data tersebut memiliki pola musiman pada hari raya keagamaan dan libur akhir tahun. Hari raya keagamaan termasuk efek variasi kalender, karena selalu bergeser setiap tahunnya. Tahapan pengerjaan dimulai dari membagi data menjadi in-sample dan out-sample. Selanjutnya dilakukan uji stasioner untuk mengetahui apakah data sudah stasioner atau tidak. Pada data yang sudah stasioner, dilakukan pemodelan SARIMA. Kemudian, ditambahkan variabel dummy pada model SARIMA yang telah diperoleh. Uji signifikansi dilakukan pada variabel dummy. Setelah itu, dilakukan uji estimasi parameter dan uji diagnostik pada model SARIMA yang sudah ditambahkan variabel dummy. Model yang diperoleh untuk memprediksi data jumlah penumpang kereta wilayah Sumatera adalah model SARIMAX (1,1,1)(1,1,0)¹² dengan nilai MAPE sebesar 3,54%. Ini berarti model yang digunakan dalam kategori sangat baik.

Kata Kunci : Kereta, SARIMAX, runtun waktu

PENDAHULUAN

Kebutuhan transportasi merupakan kebutuhan masyarakat yang terus berkembang seiring berjalannya waktu dan sejalan dengan semakin meningkatnya taraf kehidupan ekonomi masyarakat. Salah satu alat transportasi darat yang banyak dibutuhkan adalah kereta api. Kereta api menjadi salah satu transportasi pilihan masyarakat untuk melakukan aktivitasnya dalam hal pekerjaan maupun pariwisata untuk jarak tempuh yang jauh. Kereta api dipilih juga karena pergerakan dari tempat satu ke tempat lainnya sangat nyaman, cepat, dapat membawa penumpang lebih banyak serta harga yang terjangkau.

PT Kereta Api Indonesia (KAI) merupakan Badan Usaha Milik Negara yang menyediakan dan mengurus jasa angkutan kereta api di Indonesia. PT KAI dituntut untuk terus meningkatkan sarana dan prasarananya demi menjaga kenyamanan penumpang salah satu caranya yaitu dengan memprediksi jumlah penumpangnya. Data jumlah penumpang kereta biasanya terjadi pelonjakan ketika masyarakat mudik menjelang Hari Raya Idul Fitri atau liburan akhir tahun setiap bulan Desember.

Data jumlah penumpang kereta merupakan salah satu data dalam analisis runtun waktu. Hal ini dikarenakan data dikumpulkan menurut urutan waktu. Analisis runtun waktu merupakan sekumpulan data yang dibutuhkan untuk memprediksi di masa yang akan datang dengan menggunakan observasi waktu-waktu sebelumnya. Data yang digunakan dapat berupa jam, harian, mingguan, bulanan, serta tahunan [1]. Penelitian ini menggunakan data bulanan yang berfluktuasi berulang. Pola data yang berfluktuasi berulang dalam kurun waktu tertentu disebut dengan pola musiman. Metode yang dapat digunakan adalah *Seasonal Autoregressive Integrated Moving Average* (SARIMA) [2].

Pada pendekatan runtun waktu, ada variabel lain yang mempengaruhi pergerakan data. Metode yang digunakan pada data musiman dengan menambahkan faktor X adalah *Seasonal Autoregressive Integrated Moving Average with Exogenous* (SARIMAX) [3]. Metode SARIMAX dapat digunakan untuk menganalisis data yang terdeteksi efek variasi kalender. Variasi kalender terjadi karena adanya pergeseran pada data dari bulan ke bulan, tahun ke tahun seperti libur hari raya keagamaan atau

kebudayaan tertentu [4]. Data penumpang kereta yang digunakan berfluktuatif berulang pada bulan yang mengandung Hari Raya Idul Fitri. Oleh karena itu, peneliti melakukan penelitian untuk memprediksi data jumlah penumpang kereta dengan efek variasi kalender pada model SARIMAX.

Penelitian ini bertujuan membentuk model SARIMAX dengan efek variasi kalender dan menentukan hasil prediksi jumlah penumpang kereta api wilayah Sumatera tahun 2019. Langkah pertama dimulai dari membagi data menjadi *in-sample* dan *out-sample*. Selanjutnya dilakukan uji stasioner untuk mengetahui apakah data sudah stasioner atau tidak. Apabila data tidak stasioner terhadap variansi maka dilakukan transformasi. Sedangkan, data yang tidak stasioner terhadap rata-rata (*mean*) dilakukan *differencing*. Identifikasi model SARIMA dapat dilakukan dengan melihat *plot* ACF dan PACF. Kemudian, model SARIMA dipilih berdasarkan AIC terkecil.

Selanjutnya, dilakukan estimasi parameter pada model SARIMA untuk mengetahui apakah parameter model signifikan atau layak digunakan. Kemudian, dilakukan uji diagnostik pada model SARIMA untuk mengetahui apakah nilai residual *white noise* atau tidak. Setelah diperoleh model SARIMA maka dilakukan penambahan variabel *dummy*. Kemudian, dilakukan kembali estimasi parameter dan uji diagnostik pada model SARIMAX. Prediksi jumlah penumpang kereta wilayah Sumatera dilakukan dengan model SARIMAX yang sudah didapat. Kemudian, evaluasi hasil peramalan dengan menggunakan nilai *Mean Absolute Percentage Error* (MAPE) untuk mengetahui apakah peramalan yang dihasilkan dalam kategori baik atau tidak.

PROSES AUTOREGRESIF

Runtun waktu adalah data yang dikumpulkan dari waktu ke waktu untuk menggambarkan perkembangan dari suatu kegiatan. Runtun waktu jika digambarkan oleh suatu diagram akan memperlihatkan suatu fluktuasi yang merupakan gerakan naik turun dan dapat digunakan sebagai dasar penarikan tren atau garis yang menunjukkan arah perkembangan dari suatu data [5]. Runtun waktu merupakan data yang dikumpulkan dari waktu ke waktu yang memiliki korelasi antar pengamatannya [6].

Runtun waktu memiliki dua macam kondisi stasioner yaitu stasioner terhadap rata-rata (*mean*) dan stasioner terhadap variansi (*varians*). Bila data tidak stasioner terhadap variansi maka dilakukan transformasi menggunakan transformasi Box-Cox. Transformasi Box-Cox sebagai berikut [7]:

$$T(Y_t) = Y_t^\lambda = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln Y_t, & \lambda = 0 \end{cases} \quad (1)$$

dimana Y_t adalah data runtun waktu ke- t dan λ adalah parameter transformasi.

Ketika data runtun waktu dalam kondisi tidak stasioner terhadap rata-rata (*mean*), maka menggunakan uji *Augmented Dickey Fuller* (ADF). Uji ADF untuk mengetahui apakah di dalam model terdapat akar unit atau tidak.

Hipotesis yang digunakan yaitu:

$$H_0: \phi = 0 \text{ (Terdapat akar unit)}$$

$$H_1: \phi < 0 \text{ (Tidak terdapat akar unit)}$$

Hipotesis digunakan untuk statistik uji dalam ADF[7]:

$$t = \frac{\hat{\phi}}{SE(\hat{\phi})} \quad (2)$$

Apabila nilai t hitung lebih besar dari t tabel atau nilai p - *value* $< \alpha$ ($\alpha = 0,05$), maka H_0 ditolak yang artinya tidak terdapat akar unit atau data bersifat stasioner.

MODEL SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE DENGAN VARIABEL EXOGEN (SARIMAX)

Model SARIMAX adalah bentuk model SARIMA dengan variabel independen. Model SARIMAX $(p, d, q)(P, D, Q)^S(X)$ di mana, X merupakan variabel independen. Variabel independen pada model SARIMAX dapat dimodelkan dengan persamaan linier berganda. Bentuk umum model SARIMAX $(p, d, q)(P, D, Q)^S(X)$ dinyatakan[8]:

$$Y_t = \beta_0 + \beta_1 V_{1,t} + \beta_2 V_{2,t} + \dots + \beta_i V_{i,t} + \omega_T \tag{3}$$

dengan:

- Y_t = data pada waktu ke- t
- $V_{j,t}$ = variabel *dummy* ke- j pada saat t dengan $i = 1,2,3, \dots, j$
- β_0, β_k = koefisien regresi dari variabel *dummy*
- ω_t = residual pada waktu ke- t dari proses runtun waktu

Residual ω_t dinyatakan dalam model ARIMA [8]:

$$\omega_t = \frac{\theta_q(B)\Theta_Q(B^S)e_t}{\Phi_P(B^S)\phi_p(B)(1-B)^d(1-B^S)^D} \tag{4}$$

Selanjutnya, Persamaan 4 substitusikan ke dalam Persamaan 3. Model umum SARIMAX dapat ditulis [8]:

$$Y_t = \beta_0 + \beta_1 V_{1,t} + \beta_2 V_{2,t} + \dots + \beta_i V_{i,t} + \frac{\theta_q(B)\Theta_Q(B^S)e_t}{\Phi_P(B^S)\phi_p(B)(1-B)^d(1-B^S)^D} \tag{5}$$

$$\Phi_P(B^S)\phi_p(B)(1-B)^d(1-B^S)^D Y_t = \theta_q(B)\Theta_Q(B^S)e_t + \beta_0 + \beta_1 V_{1,t} + \beta_2 V_{2,t} + \dots + \beta_i V_{i,t} \tag{6}$$

dengan,

- p, d, q = orde AR, *differencing*, MA non musiman
- P, D, Q = orde AR, *differencing*, MA musiman
- S = musiman

$\Phi_P(B^S) = (1 - \Phi_1 B^{1S} - \dots - \Phi_P B^S)$ koefisien komponen AR musiman S dengan orde P

$\Theta_Q(B^S) = (1 - \Theta_1 B^{1S} - \dots - \Theta_Q B^S)$ koefisien komponen MA musiman S dengan orde Q

$(1 - B^S)^D$ = *differencing* musiman S dengan orde D

e_t = nilai *error* pada waktu ke- t

B = operator *backward shift* (operator langkah mundur)

VARIABEL DUMMY

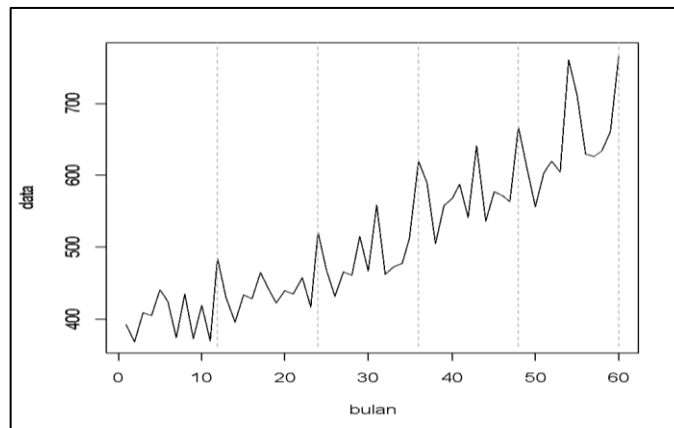
Pada analisis regresi, variabel dependen tidak hanya dipengaruhi oleh variabel-variabel kuantitatif, tetapi juga oleh variabel-variabel yang pada dasarnya bersifat kualitatif. Variabel kualitatif perlu dikonversi ke dalam bentuk kuantitatif sehingga dapat diaplikasikan pada analisis regresi dengan membentuk variabel *dummy*. Variabel tersebut diasumsikan menggunakan nilai 0 dan 1. Bentuk umum model regresi *dummy* sebagai berikut [1]:

$$Y_t = \beta_0 + \beta_1 D_{1t} + \beta_2 D_{2t} + \dots + \beta_m D_{mt} + \varepsilon_T \tag{7}$$

dengan β_0 menyatakan *intercept*, $\beta_1, \beta_2, \dots, \beta_m$ adalah koefisien parameter dari variabel *dummy* D_1, D_2, \dots, D_m serta ε_t adalah residual dari regresi *dummy*.

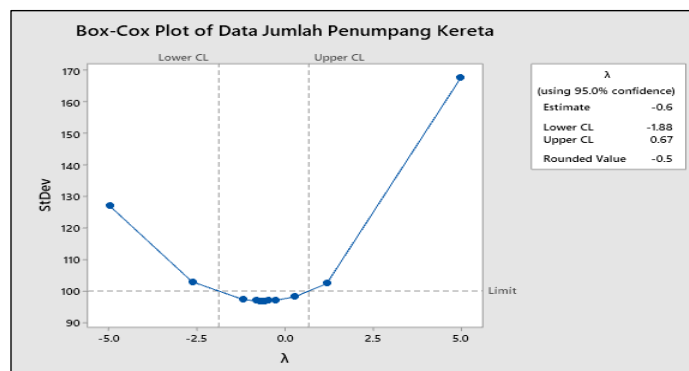
STUDI KASUS

Data yang digunakan pada penelitian ini adalah data jumlah penumpang kereta api wilayah Sumatera yang bersumber dari Badan Pusat Statistik (BPS). Data diambil dari periode bulan Januari 2014 sampai Desember 2018 sebanyak 60 datum. Data jumlah penumpang kereta api setiap tahunnya cenderung mengalami peningkatan pada bulan-bulan tertentu seperti pada bulan terjadinya Hari Raya Idul Fitri atau libur akhir tahun. Untuk mengetahui pergerakan dari data jumlah penumpang kereta ini dapat dilihat dalam Gambar 1.



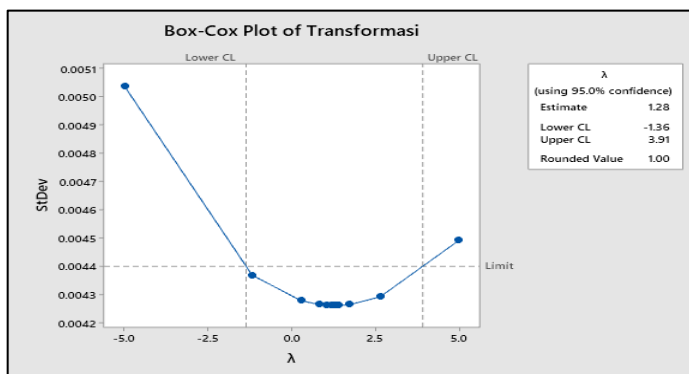
Gambar 1 Data Jumlah Penumpang Kereta Api Wilayah Sumatera

Gambar 1 terlihat bahwa peningkatan jumlah penumpang terjadi pada bulan Agustus 2014, Juli 2015, Juli 2016, Juli 2017 dan Juni 2018 yang dipengaruhi oleh perayaan Hari Raya Idul Fitri. Peningkatan jumlah penumpang ini juga dipengaruhi oleh libur akhir tahun yaitu di setiap bulan Desember. Sehingga dapat disimpulkan bahwa data berpola musiman dengan efek variasi kalender. Efek variasi kalender ini disebabkan karena adanya pergeseran pada perayaan Hari Raya Idul Fitri di setiap tahunnya. Selanjutnya, untuk mengetahui apakah data stasioner terhadap variansi dapat dilakukan transformasi Box-Cox. Hasil transformasi Box-Cox dapat dilihat pada Gambar 2.



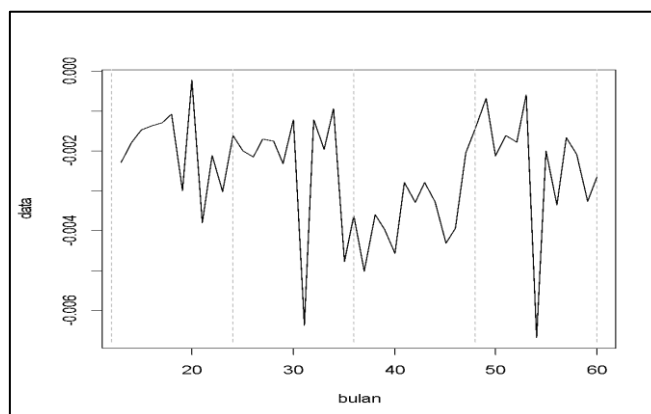
Gambar 2 Transformasi Pertama Data Jumlah Penumpang Kereta

Gambar 2 diperoleh hasil transformasi *Box-Cox* dengan nilai *Rounded Value* (λ) yaitu -0,50. Hal ini dapat disimpulkan bahwa data belum stasioner terhadap variansi sehingga perlu dilakukan transformasi kembali. Transformasi yang digunakan untuk nilai $\lambda = 0,50$ adalah $\frac{1}{\sqrt{z_t}}$. Hasil dari transformasi tersebut ditunjukkan pada Gambar 3.



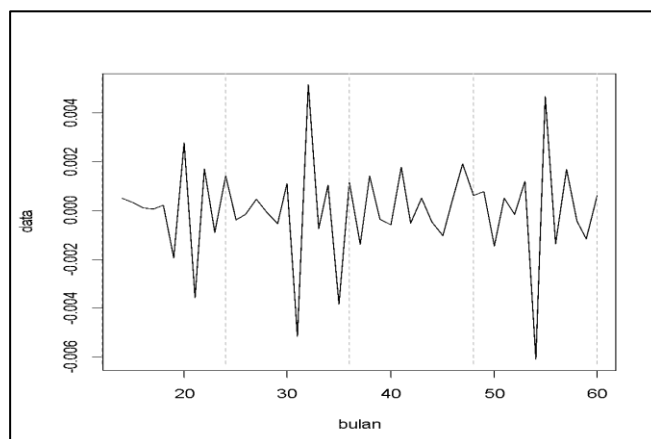
Gambar 3 Transformasi Kedua Data Jumlah Penumpang Kereta

Berdasarkan Gambar 3 dihasilkan nilai *Rounded Value* dari transformasi yang dilakukan adalah $(\lambda) = 1$, sehingga data dikatakan sudah stasioner dalam variansi. Data yang sudah diuji stasioner terhadap variansi, selanjutnya diuji kembali stasioner terhadap rata-rata dengan *differencing* musiman dan non-musiman. Hasil *differencing* musiman diperoleh seperti pada Gambar 4.



Gambar 4 Differencing Musiman Data Penumpang Kereta

Gambar 4 menunjukkan bahwa data dikatakan belum stasioner dalam rata-rata (*mean*). Untuk memvalidasi dugaan tersebut, dilakukan pengujian stasioner dengan uji *Augmented Dicky Fuller* (ADF). Hasil uji ADF menunjukkan bahwa nilai *p-value* yang diperoleh sebesar 0,3012. Nilai *p-value* ini lebih besar dari α (0,05), sehingga dapat dikatakan bahwa data belum stasioner terhadap *mean*. Selanjutnya, dilakukan *differencing* non-musiman menggunakan data hasil *differencing* musiman. Hasil *differencing* non-musiman ditunjukkan pada Gambar 5.

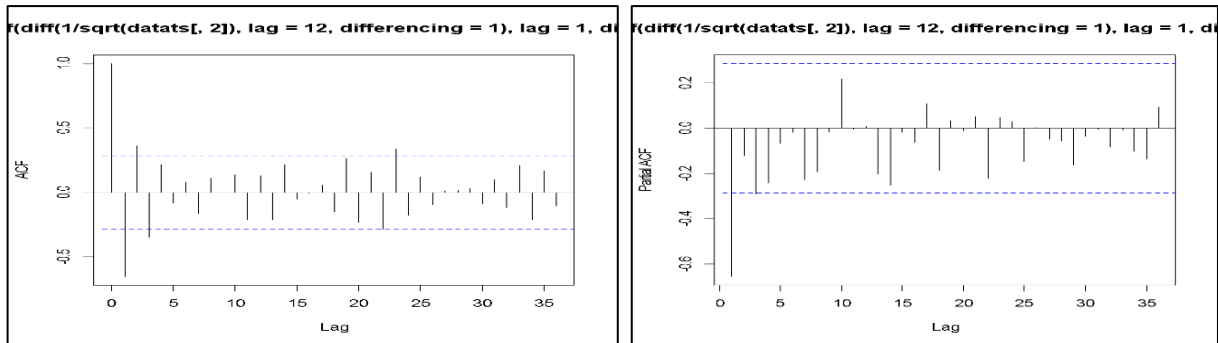


Gambar 5. Differencing Non-musiman Data Penumpang Kereta

Gambar 5 dapat dilihat bahwa data sudah dapat dikatakan stasioner terhadap rata-rata (*mean*). Selain itu, dilihat dari dari uji ADF dengan menggunakan bantuan *software R*. Hasil uji ADF diperoleh nilai *p-value* sebesar 0,01. Ini berarti nilai *p-value* kurang dari α (0,05). Sehingga dapat disimpulkan bahwa data penumpang kereta wilayah Sumatera sudah stasioner terhadap rata-rata (*mean*).

PEMODELAN SARIMA

Setelah data sudah stasioner, selanjutnya dilakukan identifikasi model SARIMA dengan melihat plot ACF dan PACF. Berikut plot ACF dan PACF yang diperoleh dengan menggunakan *software R* seperti pada Gambar 6.



Gambar 6. ACF dan PACF Data Penumpang Kereta

Berdasarkan Gambar 6 diketahui bahwa terdapat *lag* yang terpotong pada plot ACF yaitu *lag* 1, 2, 3, 23, dan 24. Sedangkan, pada plot PACF terdapat *lag* yang terpotong yaitu *lag* 1 dan 3. Sehingga dari Gambar 6 dipilih beberapa alternatif model dan dilakukan pemilihan model dari berdasarkan AIC terkecil. Model yang terpilih yaitu SARIMA (1,1,1)(1,1,0)¹².

Langkah selanjutnya adalah mengestimasi parameter dan uji diagnostik. Estimasi parameter digunakan untuk mengetahui apakah parameter model SARIMA signifikan atau tidak. Hasil uji signifikansi parameter disajikan pada Tabel 2.

Tabel 2 Estimasi Parameter Model SARIMA (1,1,1)(1,1,0)¹²

Parameter	Koefisien	Standar Error	P-value
AR (1)	-0,4191	0,1836	0,0261
MA (1)	-0,6828	0,1589	0,0001
SAR (1)	-0,3521	0,1714	0,0444

Berdasarkan Tabel 2 diketahui bahwa nilai *p – value* untuk masing-masing parameter AR (1), MA (1), dan SAR (1) kurang dari α (0,05). Ini berarti H_0 ditolak sehingga model (1,1,1)(1,1,0)¹² signifikan. Selanjutnya, dilakukan uji diagnostik untuk mengetahui apakah residual bersifat *white noise* atau tidak. Hasil uji diagnostik menggunakan *Q-Ljung Box* seperti pada Tabel 3.

Tabel 3. Hasil Uji Q-Ljung Box Model SARIMA (1,1,1)(1,1,0)¹²

	ACF	PACF	Q-Stats	P-Value
0	1.000000000	1.000000000	NA	NA
1	0.026451438	0.026451438	0.04411533	0.8336392
2	0.039746237	0.039073897	0.14543808	0.9298620
3	-0.186729485	-0.189194340	2.42102738	0.4897324
4	0.036204599	0.047389530	2.50810016	0.6431859
5	-0.039950021	-0.028470256	2.61604808	0.7589255
6	-0.088701967	-0.130192074	3.15806854	0.7887621
7	-0.069705882	-0.045011410	3.49910901	0.8353198
8	0.074905821	0.077881467	3.90050287	0.8659870
9	0.110346610	0.074357888	4.78866194	0.8523282
10	0.009312462	-0.020753418	4.79511405	0.9044375
11	-0.190867498	-0.184079046	7.56085066	0.7520180

Tabel 3. Hasil Uji Q-Ljung Box Model SARIMA (1,1,1)(1,1,0)¹² (Lanjutan)

	ACF	PACF	Q-Stats	P-Value
12	-0.047929970	-0.020886111	7.73889002	0.8051879
13	-0.212115205	-0.228743856	11.30002705	0.5856991
14	0.097949827	0.067866726	12.07590330	0.6002071
15	0.021859603	0.075935915	12.11540492	0.6702736
16	0.044575661	-0.054111301	12.28339586	0.7242518
17	0.008404330	-0.002437768	12.28950642	0.7822983
18	-0.004918661	-0.058053233	12.29164924	0.8318202
19	0.139438446	0.135666116	14.05575311	0.7804315
20	-0.138479941	-0.138869793	15.83918566	0.7265495
21	-0.105827675	-0.089901479	16.90744535	0.7167013
22	-0.167479980	-0.107966090	19.65334806	0.6047136
23	0.216493255	0.166279082	24.36561577	0.3838001
24	0.044284919	-0.071787422	24.56826835	0.4295156
25	-0.018387604	-0.051403280	24.60420398	0.4847214
26	-0.081459222	-0.055788376	25.33021721	0.5003506
27	-0.107501067	-0.173062709	26.63294761	0.4837267
28	-0.029174308	-0.069219806	26.73189267	0.5328805
29	-0.062488928	-0.040350413	27.20047661	0.5608897
30	-0.051417498	0.022038026	27.52830273	0.5954174
31	0.044775558	-0.049998965	27.78547667	0.6322172
32	0.053899944	-0.005911136	28.17145377	0.6608543
33	0.152135021	0.002895224	31.36032935	0.5488485
34	-0.045348562	-0.054716497	31.65456591	0.5830936
35	0.065832886	0.028145358	32.29946047	0.5991558
36	-0.024364322	0.121896449	32.39147160	0.6409611

Berdasarkan Tabel 3 hasil uji *Q-Ljung Box* diketahui bahwa nilai *p – value* pada semua *lag* lebih dari $\alpha = 0,05$. Ini berarti H_0 diterima sehingga residual sudah bersifat *white noise* dan model SARIMA (1,1,1)(1,1,0)¹² dapat digunakan.

PEMODELAN REGRESI VARIABEL DUMMY

Setelah diperoleh model SARIMA, langkah selanjutnya yaitu memasukkan variabel *dummy*. Nilai *dummy* yang terdapat bulan Hari Raya Idul Fitri dan libur akhir tahun (V_t) diberi nilai 1 dan 0 untuk lainnya. Untuk mengetahui apakah data jumlah penumpang kereta signifikan terhadap variabel variasi kalender, dilakukan uji signifikansi dengan menggunakan *software R*. Hasil pengujian disajikan dalam Tabel 4.

Tabel 4. Output Uji Signifikansi Variabel Variasi Kalender

	Estimate	Std.Error	t-value	Pr(> t)
<i>Dummy1</i>	0,04292	0,01889	2,272	0,0268 *
<i>Dummy2</i>	0,04083	0,01889	2,161	0,0348 *
Adjusted R-Squared	0,1155			

Berdasarkan Tabel 4 diketahui bahwa model SARIMA (1,1,1)(1,1,0)¹² pada data jumlah penumpang kereta signifikan terhadap variabel variasi kalender. Hal ini karena nilai *p-value* yang diperoleh kurang dari $\alpha (0,05)$. Selain itu, dapat dilihat dari nilai t_{hitung} diketahui bahwa nilai *t* hitung dari *Dummy1* (Hari Raya Idul Fitri) dan *Dummy2* (libur akhir tahun) lebih besar dibandingkan $t_{tabel(0,025;58)}$ yaitu 2,001. Adapula nilai *Adjusted R-Squared* sebesar 11,55% yang menunjukkan besar pengaruh variabel *dummy* terhadap model SARIMA dan sebesar 88,45% dipengaruhi oleh variabel lain.

PEMODELAN SARIMAX

Setelah diuji signifikansi variabel *dummy*, selanjutnya dilakukan pemodelan SARIMAX yaitu dengan mengestimasi parameter. Hasil estimasi parameter model SARIMAX disajikan pada Tabel 5.

Tabel 5. Estimasi Parameter Model SARIMAX (1,1,1)(1,1,0)¹²

Parameter	Koefisien	Standar Error	P-value
AR (1)	-0,4191	0,1838	0,000
MA (1)	-0,6828	0,1591	0,000
SAR (1)	-0,3521	0,1754	0,000
Dummy (1)	0,0429	0,01889	0,026
Dummy (2)	0,0408	0,01889	0,034

Tabel 5 diketahui bahwa nilai p – value untuk masing-masing parameter AR (1), MA (1), SAR (1), Dummy (1), dan Dummy (2) kurang dari α (0,05). Ini berarti H_0 ditolak sehingga model SARIMAX (1,1,1)(1,1,0)¹² signifikan atau layak digunakan. Selanjutnya, dilakukan uji diagnostik. Hasil uji diagnostik model SARIMAX disajikan pada Tabel 6.

Tabel 6. Hasil Uji Q-Ljung Box Model SARIMAX (1,1,1)(1,1,0)¹²

	ACF	PACF	Q-Stats	P-Value
0	1.000000	1.000000	NA	NA
1	0.026451	0.026451	0.044115	0.8336392
2	0.039746	0.039074	0.145438	0.9298620
3	-0.186729	-0.189194	2.421027	0.4897324
4	0.036205	0.047390	2.508100	0.6431859
5	-0.039950	-0.028470	2.616048	0.7589255
6	-0.088702	-0.130192	3.158069	0.7887621
7	-0.069706	-0.045011	3.499109	0.8353198
8	0.074906	0.077881	3.900503	0.8659870
9	0.110347	0.074358	4.788662	0.8523282
10	0.009312	-0.020753	4.795114	0.9044375
11	-0.190867	-0.184079	7.560851	0.7520180
12	-0.047930	-0.020886	7.738890	0.8051879
13	-0.212115	-0.228744	11.300027	0.5856991
14	0.097950	0.067867	12.075903	0.6002071
15	0.021860	0.075936	12.115405	0.6702736
16	0.044576	-0.054111	12.283396	0.7242518
17	0.008404	-0.002438	12.289506	0.7822983
18	-0.004919	-0.058053	12.291649	0.8318202
19	0.139438	0.135666	14.055753	0.7804315
20	-0.138480	-0.138870	15.839186	0.7265495
21	-0.105828	-0.089901	16.907445	0.7167013
22	-0.167480	-0.107966	19.653348	0.6047136
23	0.216493	0.166279	24.365616	0.3838001
24	0.044285	-0.071787	24.568268	0.4295156
25	-0.018388	-0.051403	24.604204	0.4847214
26	-0.081459	-0.055788	25.330217	0.5003506
27	-0.107501	-0.173063	26.632948	0.4837267
28	-0.029174	-0.069220	26.731893	0.5328805
29	-0.062489	-0.040350	27.200477	0.5608897
30	-0.051417	0.022038	27.528303	0.5954174
31	0.044776	-0.049999	27.785477	0.6322172
32	0.053900	-0.005911	28.171454	0.6608543
33	0.152135	0.002895	31.360329	0.5488485
34	-0.045349	-0.054716	31.654566	0.5830936
35	0.065833	0.028145	32.299460	0.5991558
36	-0.024364	0.121896	32.391472	0.6409611

Berdasarkan Tabel 6 hasil uji Q-Ljung Box diketahui bahwa nilai $p - value$ pada semua lag lebih dari $\alpha (0,05)$. Ini berarti H_0 diterima sehingga residual sudah bersifat *white noise*. Berdasarkan hasil uji signifikansi parameter dan uji diagnostik, maka terbentuk model SARIMAX (1,1,1)(1,1,0)¹² sebagai berikut:

$$(1 - (-0,3521B^{12}))(1 - (-0,4191B))(1 - B)^1(1 - B^{12})^1 Y_t = (1 - (-0,6828B))e_t + 0,0429V_{1,t} + 0,0408V_{2,t}$$

$$(1 + 0,3521B^{12})(1 + 0,4191B)(1 - B)^1(1 - B^{12})^1 Y_t = (1 + 0,6828B)e_t + 0,0429V_{1,t} + 0,0408V_{2,t}$$

$$Y_t = 0,5809Y_{t-1} + 0,4191Y_{t-2} + 0,6479Y_{t-12} - 0,3764Y_{t-13} - 2715Y_{t-14}$$

$$+ 0,3521Y_{t-24} - 0,2045Y_{t-25} - 0,1476Y_{t-26} + (-0,6828e_{t-1})$$

$$+ 0,0429V_{1,t} + 0,0408V_{2,t}$$

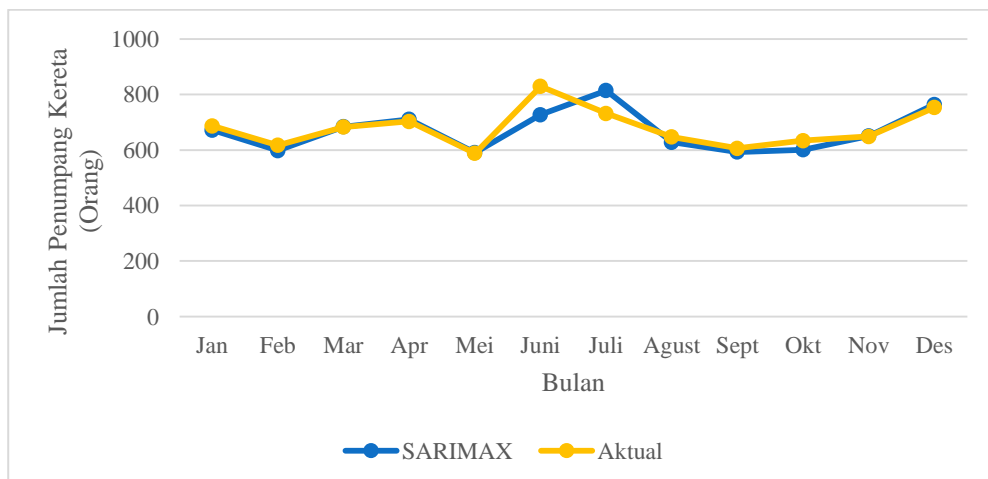
PERAMALAN DATA JUMLAH PENUMPANG KERETA

Setelah mendapatkan model, selanjutnya dilakukan peramalan untuk memprediksi data jumlah penumpang kereta wilayah Sumatera pada tahun 2019. Hasil peramalan disajikan pada Tabel 7.

Tabel 7. Hasil Peramalan Jumlah Penumpang Kereta wilayah Sumatera

Bulan	Data Aktual	SARIMAX
Januari	687	672
Februari	617	598
Maret	683	684
April	703	710
Mei	588	591
Juni	829	727
Juli	732	815
Agustus	647	628
September	606	593
Oktober	634	601
November	649	650
Desember	753	764
MAPE	3,54%	

Berdasarkan Tabel 7 diketahui bahwa hasil peramalan jumlah penumpang kereta dengan model SARIMAX (1,1,1)(1,1,0)¹² menghasilkan nilai MAPE sebesar 3,54%. Nilai MAPE yang dihasilkan kurang dari 10% yang berarti hasil peramalan dengan model SARIMAX (1,1,1)(1,1,0)¹² sangat baik dan mendekati data aktualnya. Hal ini dapat dilihat secara visual pada Gambar 7.



Gambar 7 Perbandingan Data Aktual dan SARIMAX

KESIMPULAN

1. Berdasarkan hasil analisis diperoleh model SARIMAX (1,1,1)(1,1,0)¹² yang digunakan untuk memprediksi jumlah penumpang kereta yaitu:

$$\begin{aligned} (1 - (-0,3521B^{12}))(1 - (-0,4191B))(1 - B)^1 (1 - B^{12})^1 Y_t &= (1 - (-0,6828B))e_t + 0,0429V_{1,t} + 0,0408V_{2,t} \\ (1 + 0,3521B^{12})(1 + 0,4191B)(1 - B)^1 (1 - B^{12})^1 Y_t &= (1 + 0,6828B)e_t + 0,0429V_{1,t} + 0,0408V_{2,t} \\ Y_t &= 0,5809Y_{t-1} + 0,4191Y_{t-2} + 0,6479Y_{t-12} - 0,3764Y_{t-13} - 2715Y_{t-14} \\ &\quad + 0,3521Y_{t-24} - 0,2045Y_{t-25} - 0,1476Y_{t-26} + (-0,6828e_{t-1}) \\ &\quad + 0,0429V_{1,t} + 0,0408V_{2,t} \end{aligned}$$

2. Model SARIMAX (1,1,1)(1,1,0)¹² menghasilkan nilai MAPE sebesar 3,54%. Ini berarti peramalan yang dihasilkan menggunakan model SARIMAX (1,1,1)(1,1,0)¹² dalam kategori sangat baik karena nilainya mendekati data aktual.

DAFTAR PUSTAKA

- [1]. Gujarati, D. N., *Basic Econometrics*, New York: McGraw-Hill; 2003.
- [2]. Makridakis, S., Wheelwright, S. C., dan Hyndman, R. J., *Forecasting Methods and Applications*, John Wiley & Sons, 1999.
- [3]. Virati, M. Q., Pamanik, D. P., dan Pramana, S., Forecasting Number of Passengers of TransJakarta using Seasonal ARIMAX Method, *Journal of Data Science and Its Applications*, 2020; 3(1):31-37.
- [4]. Cools, M., Moons, E., dan Wets, G., Investigating the Variability in Daily Traffic Counts through use of ARIMAX and SARIMAX Models: Assessing the Effect of Holidays on Two Site Locations, *Transportation Research Record*, 2009; 2136(1):57-66.
- [5]. J. Supranto., *Statistik (Teori dan Aplikasi) Edisi Keenam*, Jakarta: Erlangga; 2000.
- [6]. Box, G. E., Jenkins, G. M., Reinsel, G. C., dan Ljung, G. M, *Time Series Analysis: Forecasting and Control*, Ed ke-5, John Wiley & Sons, 2016.
- [7]. Wei, W. W., *Time Series Analysis, Univariate and Multivariate Method*, Ed ke-2, New York: Pearson Education; 2006.
- [8]. Arunraj, N. S., Ahrens, D., & Fernandes, M., Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry, *International Journal of Operations Research and Information Systems (IJORIS)*, 2016; 7(2):1-21

HAYATI : Jurusan Matematika FMIPA UNTAN, Pontianak,
hayatihayati27@student.untan.ac.id

NAOMI NESSYANA DEBATARAJA : Jurusan Matematika FMIPA UNTAN, Pontianak,
naominessyana@math.untan.ac.id

SHANTIKA MARTHA : Jurusan Matematika FMIPA UNTAN, Pontianak,
shantika.martha@math.untan.ac.id