



# Peramalan Jumlah Penumpang Kereta Api Lodaya Jurusan Bandung-Solo Menggunakan Model Reg-ARIMA Dengan Variasi Kalender (Studi Kasus: PT. Kereta Api Indonesia)

Dyah Puspita Sari\*, Gumgum Darmawan, Soemartini

Departemen Statistika, FMIPA Universitas Padjadjaran

\*E-mail: dyahpuspitasari372@gmail.com

## Abstrak

Data *time series* dengan variasi kalender adalah data *time series* yang mengikuti pola pergeseran kalender Hijriyah. Pergeseran kalender Hijriyah terjadi karena penetapan perayaan hari yang menggunakan acuan kalender berbeda-beda, contohnya adalah penetapan perayaan Idul Fitri. Dalam kalender Hijriyah perayaan ini terjadi setiap tanggal 1 Syawal, tetapi dalam kalender Masehi perayaan Idul Fitri selalu berubah-ubah waktunya atau mengalami pergeseran setiap tahunnya. Perayaan Idul Fitri yang mengalami pergeseran ini kemudian memberi pengaruh berupa kenaikan atau penurunan data jumlah penumpang kereta api, inilah yang disebut dengan efek variasi kalender. Pada data jumlah penumpang kereta api, perayaan Idul Fitri berpengaruh terhadap kenaikan jumlah penumpang. Model yang bisa digunakan untuk meramalkan data *time series* dengan variasi kalender adalah Reg-ARIMA. Model Reg-ARIMA menggunakan metode regresi yang dikombinasikan dengan model ARIMA pada sisaan regresinya yang sesuai. Dalam penelitian ini, data jumlah penumpang kereta api Lodaya jurusan Bandung-Solo tahun 2011-2015 dimodelkan dengan menggunakan model Reg-ARIMA dengan bantuan program Win-X13. Tujuan penelitian ini adalah meramalkan jumlah penumpang kereta menggunakan model Reg-ARIMA. Hasil penelitian menunjukkan bahwa banyak penumpang pada bulan yang dipengaruhi variasi kalender yaitu Juli 2015 dan Juli 2016 adalah yang tertinggi apabila dibandingkan dengan bulan-bulan lain, yaitu sebesar 6,071 penumpang dan 6,355 penumpang.

*Kata Kunci: Reg-ARIMA, Variasi Kalender, Win- X13*

## 1. Pendahuluan

Dewasa ini, salah satu transportasi darat di Indonesia yang menjadi alternatif pemilihan para sektor pengguna diantaranya adalah kereta api. Masyarakat memilih kereta api karena masyarakat membutuhkan transportasi massal dengan waktu tempuh relatif lebih cepat karena terhindar dari kemacetan, harganya ekonomis serta tingkat keselamatan yang cukup tinggi. Dalam realitanya, tidak jarang permintaan akan angkutan penumpang kereta api jauh lebih besar dibandingkan dengan kapasitas *seat* yang disediakan, sehingga sering menimbulkan penumpukan penumpang di berbagai stasiun. Permasalahan tersebut sering dihadapi seluruh daerah operasi PT. Kereta Api Indonesia (Persero) khususnya pada hari libur sekolah dan hari libur perayaan hari-hari besar seperti hari Tahun Baru Masehi, hari Natal dan hari raya Idul Fitri, dimana jumlah *volume* penumpang kereta api mengalami kenaikan yang sangat signifikan dibandingkan hari-hari biasa.

Di Indonesia, presentase masyarakat muslim mencapai sekitar 85% dari seluruh penduduk Indonesia (Putra, 2016). Dimana setiap tahunnya pada tanggal 1 Syawal pada penanggalan Hijriyah masyarakat muslim di dunia, termasuk di Indonesia merayakan hari raya Idul Fitri. Umat muslim di Indonesia menjadikan hari raya Idul Fitri sebagai hari raya utama, momen untuk berkumpul kembali bersama keluarga. Dua minggu

sebelum terjadinya hari raya Idul Fitri, umat Islam di Indonesia mulai disibukkan dengan persiapan menyambut perayaan hari raya tersebut, salah satunya adalah mudik.

Salah satu jalur darat di pulau Jawa yang digunakan oleh pemudik yaitu jalur selatan. Jalur yang berbelok-belok dan bisa menyebabkan mabuk darat ini menyebabkan sebagian besar pemudik memilih untuk menggunakan transportasi kereta api. Selain itu, jalur selatan juga digunakan penumpang kereta api sebagai jalur alternatif jika kuota untuk jalur utara sudah penuh. Adapun kereta api alternatif yang sering digunakan yaitu Lodaya. Hal inilah yang menjadi salah satu faktor terjadinya lonjakan penumpang pada kereta api Lodaya.

Pada data jumlah penumpang kereta api, perayaan hari raya Idul Fitri berpengaruh terhadap jumlah penumpang kereta api. Secara umum, hari raya Idul Fitri menyebabkan jumlah penumpang kereta api meningkat. Inilah yang disebut dengan *calender effect* dimana pergeseran pada kalender berpengaruh pada data *time series* (Liu, 1980). Penggunaan kalender yang bervariasi menyebabkan adanya pergeseran waktu pada perayaan hari raya tertentu. Pergeseran waktu ini yang kemudian menjadi masalah pada peramalan data *time series* musiman, karena model yang sering digunakan (ARIMA musiman) hanya sesuai untuk fenomena musiman dengan periode yang



relatif sama. Metode ini tidak bisa menangkap fenomena pergeseran musim sehingga mengakibatkan peramalan yang kurang tepat. Pada penelitian ini, peramalan jumlah penumpang kereta api Lodaya jurusan Bandung-Solo yang dipengaruhi oleh *calendar effect* diperlukan model *time series* dengan *calendar effect* Reg-ARIMA yang dapat digunakan untuk menjelaskan efek dari variasi kalender dan dapat menghasilkan ramalan yang tepat.

## 2. Metode

### 2.1 Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Wakil Presiden Pemasaran Penumpang Divisi Komersial Kantor Pusat PT. KAI. Diperoleh data bulanan selama 5 tahun dari bulan Januari 2011-Desember 2015.

Variabel yang akan diteliti pada penelitian ini difokuskan pada jumlah penumpang kereta api Lodaya jurusan Bandung-Solo selama bulan Januari 2011 hingga bulan Desember 2015.

### 2.2 Metode Analisis Data

Peramalan jumlah penumpang kereta api Lodaya untuk masa yang akan *dating* menggunakan metode peramalan variasi kalender dengan model Reg-ARIMA. Model ini menggunakan model regresi yang dikombinasikan dengan model ARIMA pada eror dari model regresinya. Selain itu, model Reg-ARIMA juga mempertimbangkan pergeseran musiman yang terjadi pada data.

Model variasi kalender pertama kali diperkenalkan oleh Bell dan Hillmer (1983) dengan bentuk umum sebagai berikut:

$$X_t = f(Y_t) + Z_t$$

Model variasi kalender ini menjadi model regresi ARIMA (Reg-ARIMA) yang didefinisikan sebagai berikut:

$$X_t = \sum_i \beta_i Y_{it} + Z_t$$

Berikut ini adalah prosedur pembentukan model *time series* dengan *calendar effect* menggunakan model Reg-ARIMA :

#### a. Menentukan Variabel Regresi dan Komponen Regresi

Dalam meramalkan menggunakan model Reg-ARIMA dibutuhkan variabel regresi untuk menghitung variasi kalendernya. Penentuan variabel regresi diawali dengan penentuan kapan hari raya Idul Fitri terjadi dan berapa banyaknya hari yang berpengaruh pada hari raya tersebut. Perayaan Idul Fitri mempengaruhi banyak penumpang kereta selama 2 minggu yaitu 7 hari sebelum Idul Fitri, saat Idul Fitri dan 6 hari setelah Idul Fitri. Berikut adalah tabel waktu terjadinya Idul Fitri tahun 2011-2015.

Ada 3 macam variabel regresi yang digunakan yaitu REG1 (Menggunakan pembobotan satu variabel), REG2 (menggunakan pembobotan 2

variabel), REG3 (menggunakan pembobotan 3 variabel). Karena data penumpang kereta diasumsikan hanya dipengaruhi oleh hari raya Idul Fitri maka digunakan REG1 untuk menghitung variabel regresi.

Variabel regresi dihitung menggunakan dua kriteria (Shuja' *et al*, 2007), yaitu :

Kriteria 1: Jika Idul Fitri jatuh pada awal bulan yaitu dari tanggal 1-15.

$$REG1 = \begin{cases} \frac{n_1}{w}; & \text{untuk bulan terjadi Idul Fitri} \\ \frac{n_2}{w}; & \text{untuk bulan sebelum terjadi Idul Fitri} \\ 0; & \text{untuk bulan yang lainnya} \end{cases}$$

dengan:

$n_1$ : banyak hari yang berpengaruh pada bulan terjadi Idul Fitri

$n_2$ : banyak hari yang berpengaruh pada bulan sebelum terjadi Idul Fitri

w: total hari yang berpengaruh yaitu 14 hari

Kriteria 2: Jika Idul Fitri jatuh pada akhir bulan yaitu dari tanggal 16.-31

$$REG1 = \begin{cases} \frac{n_1}{w}; & \text{untuk bulan terjadi Idul Fitri} \\ \frac{n_2}{w}; & \text{untuk bulan sebelum terjadi Idul Fitri} \\ 0; & \text{untuk bulan yang lainnya} \end{cases}$$

dengan:

$n_1$ : banyak hari yang berpengaruh pada bulan terjadi Idul Fitri

$n_2$ : banyak hari yang berpengaruh pada bulan setelah terjadi Idul Fitri

w: total hari yang berpengaruh yaitu 14 hari

Setelah nilai variabel regresi diperoleh, selanjutnya dicari nilai komponen regresi. Komponen regresi digunakan untuk menemukan *series*  $Z_t$  yang akan dimodelkan menggunakan ARIMA.

Rumus untuk komponen regresi, yaitu:

$$komp = reg\beta$$

dimana *komp* merupakan nilai komponen regresi, *reg* $\beta$  merupakan nilai regresor pada periode  $t$  dan  $\beta$  merupakan parameter regresi.

#### b. Estimasi Parameter

Estimasi parameter model Reg-ARIMA dilakukan dengan menggunakan metode *maximum likelihood* dan *generalized least squares (GLS)*. Untuk mencari rumus umumnya digunakan persamaan Reg-ARIMA dengan asumsi  $e_t \sim (0, \sigma^2)$ . Estimasi model yang digunakan adalah menggunakan galat ARMA dengan parameter  $\theta$  diestimasi non linear *least square*, parameter  $\beta$  dan  $\phi$  diestimasi menggunakan regresi GLS secara terpisah. Penduga parameter model, yaitu  $\hat{\beta}, \hat{\phi}, \hat{\theta}$  diperoleh melalui persamaan berikut (Otto & Bell, 1987) :

$$\max_{\beta, \phi, \theta} L(\beta, \phi, \theta) = \max_{\phi, \theta} (\max_{\beta} L(\beta, \phi, \theta)) = \max_{\theta} (\max_{\phi} (\max_{\beta} L(\beta, \phi, \theta)))$$



Fungsi likelihood untuk model regresi dengan galat ARMA adalah  $f(Y_1, \dots, Y_n) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{(Y-X\beta)'L'C'CL(Y-X\beta)}{2\sigma^2}\right)$

$$\frac{n-p}{2} |G'G|^{-1/2} \exp\left(-\frac{(Y-X\beta)'L'C'CL(Y-X\beta)}{2\sigma^2}\right)$$

dengan L :

$$L = \begin{bmatrix} -\theta_p & -\theta_{p-1} & \dots & -\theta_1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -\theta_p & \dots & -\theta_2 & -\theta_1 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & \dots & -\theta_p & \dots & -\theta_1 & 1 \end{bmatrix}$$

dan dugaan parameter  $\beta$  adalah sebagai berikut :

$$\hat{\beta} = (X' L' C' C L X)^{-1} X' L' C' C L Y$$

dengan

$$(\hat{\beta}) = (X' L' C' C L X)^{-1} \sigma^2$$

$C = (I - G(G'G)^{-1}G')$  dan karena  $(I - G(G'G)^{-1}G')$  adalah idempotent maka

$$C' C = H'(I - G(G'G)^{-1}G')H.$$

G adalah kolom-kolom q pertama pada matriks,

$$A^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \pi_1 & 1 & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ \pi_{n+q-1} & \dots & \pi_1 & 1 \end{bmatrix}$$

dan sisanya adalah H, dimana matriks A adalah sebagai berikut:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \theta_1 & \ddots & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 & 0 & 0 & 0 & 0 \\ -\theta_{q-1} & -\theta_{q-2} & \dots & -\theta_2 & -\theta_1 & 1 & 0 & 0 & 0 \\ -\theta_q & -\theta_{q-1} & \dots & -\theta_3 & -\theta_2 & -\theta_1 & 1 & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & \dots & -\theta_q & -\theta_1 & 1 & 0 & 0 \end{bmatrix}$$

$A^{-1}$  adalah deret berhingga dengan bobot  $\pi$  dari perluasan  $(B) = \sum_{j=0}^{\infty} \pi_j B^j = (B)^{-1}$ , yang mana  $\pi_0 = 1$ . Elemen-elemen  $G'G$  diperoleh dengan menghitung  $\pi$  yaitu dengan rekursi berikut:

$$(1 - \theta_1 B - \dots - \theta_q B^q)(\pi_0 + \pi_1 B + \dots + \pi_{n+q} B^{n+q}) = 1$$

Estimasi ini melibatkan estimasi nonlinear karena mengandung parameter yang nonlinear. Dikarenakan perhitungan manualnya sulit maka perhitungan numeriknya dibantu dengan penggunaan program *Win-X13*.

### c. Uji Signifikansi Parameter Model Reg-ARIMA

Apabila sudah didapatkan estimasi parameter model *Reg-ARIMA* dugaan model maka langkah selanjutnya adalah menguji apakah parameter yang didapatkan sudah signifikan atau belum. Uji signifikansi parameter dilakukan untuk menguji signifikansi model variasi kalender. Statistik uji yang dipakai untuk menguji masing-masing parameter, yaitu:

$$t = \frac{\hat{\theta}}{SE \hat{\theta}}$$

$$t = \frac{\hat{\beta}}{SE(\hat{\beta})}$$

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

Kriteria Uji:

Tolak  $H_0$  jika  $|t_{hitung}| > t_{0,025,n-p}$  dengan p merupakan banyaknya parameter, gagal tolak  $H_0$  dalam hal lainnya.

### d. Diagnostic Model

Pada diagnostik model *Reg-ARIMA* ada 2 pengujian yaitu uji *Ljung-Box-Pierce* untuk mengetahui *white noise* pada galat dan uji *Geary's* untuk mengetahui normal atau tidaknya galat.

#### 1. White Noise Residual

Residual dikatakan *white noise* apabila telah memenuhi dua sifat yaitu identik dimana variansnya konstan dan independen yaitu dimana residualnya tidak saling berkorelasi, dengan rata-rata nol. Statistik uji yang digunakan untuk menguji apakah residual *white noise* yaitu Statistik *Ljung-Box*.

Hipotesis pengujianya sebagai berikut:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0 \text{ (residual white noise)}$$

$$H_1: \text{minimal ada satu } \rho_k \neq 0$$

Statistik Uji:

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

dengan:

n : banyak pengamatan

$\hat{\rho}_k$  : ACF residual pada lag ke k

Kriteria Uji:

Tolak  $H_0$  jika  $Q > \chi_{\alpha, (K-p-q)}^2$  atau *p-value*  $< \alpha$ , gagal tolak  $H_0$  dalam hal lain. Dengan p merupakan orde PACF dan q adalah orde ACF.

#### 2. Uji Kenormalan Residual

Uji kenormalan residual digunakan untuk memeriksa apakah galat berdistribusi normal atau tidak. Uji kenormalan juga bisa dilakukan dengan uji *Geary's* dengan

$$H_0: \text{galat berdistribusi normal}$$

$$H_1: \text{galat tidak berdistribusi normal}$$

$$\alpha: 0.05$$

Statistik Uji : menggunakan uji statistik *Geary's*

$$z = \frac{\alpha - 0.7979}{0.2123/\sqrt{n}}$$

Nilai 0,7979 dan 0,2123 adalah konstanta untuk mencapai kenormalan.

Kriteria Uji :

Tolak  $H_0$  jika nilai  $|z| > z_{\alpha}$ , gagal tolak  $H_0$  dalam hal lain. Artinya Galat yang dihasilkan tidak berdistribusi normal.



**e. Peramalan**

Peramalan merupakan kegiatan dalam memperkirakan apa yang akan terjadi pada masa yang akan datang. Peramalan adalah kegiatan mencoba menduga perubahan yang akan terjadi. Peramalan yang paling baik adalah yang menghasilkan galat paling kecil apabila dibandingkan dengan data aslinya. Peramalan *time series* menggunakan Reg-ARIMA bisa diterapkan pada data yang mengandung variasi kalender.

**3. Hasil dan Pembahasan**

Pembentukan model Reg-ARIMA diawali dengan menghitung regresor yang akan dipakai untuk mendapatkan nilai dari  $Z_t$ . Hasil perhitungan regresor dapat dilihat pada tabel 1, khususnya untuk bulan yang dipengaruhi efek Idul Fitri. Bulan-bulan lain yang tidak ada efek Idul Fitri regresornya bernilai nol. Regresor yang diperoleh kemudian diterapkan pada data banyak penumpang kereta menggunakan program Win-X13 untuk mendapatkan dugaan parameter regresi.

Tabel 1. Nilai Regresor

Tahun	Bulan	Regresor
2011	8	0.643
2011	9	0.357
2012	8	1
2013	8	1
2014	7	0.786
2014	8	0.214

Dugaan parameter regresi yang diperoleh ditunjukkan oleh tabel 2. Dugaan parameter ini nantinya akan digunakan untuk menghitung komponen regresi dan *series*  $Z_t$ .

Tabel 2. Hasil estimasi parameter regresi model Reg-ARIMA

Parameter	Dugaan	Standar Error	t-value
B	36429.02	6089.66	5.98

Berdasarkan tabel 2, nilai *t-value* parameter =  $5.98 > t_{(0,025,58)} = 2.001$  sehingga disimpulkan parameter regresi yang diperoleh telah signifikan. Kemudian parameter regresi dikalikan dengan nilai regresor untuk masing-masing bulan untuk memperoleh komponen regresi. Dari komponen regresi diperoleh  $Z_t$  yang dimodelkan dengan ARIMA, yaitu dengan mengurangi *series* data penumpang  $X_t$  dengan komponen regresi.

Tabel 3. Hasil estimasi dan uji signifikansi parameter model Reg-ARIMA (1 0 0)(0 1 1)<sup>12</sup>

Parameter	Dugaan	Standar Error	t-value
AR (1)	0.4466	0.12091	3.69
SMA (1)	0.6251	0.11667	5.35

Berdasarkan tabel 3,  $|t\text{-value}| > t_{(0,025,58)} = 2.001$  untuk masing-masing parameter sehingga disimpulkan bahwa parameter signifikan dan bisa digunakan dalam model.

Diagnostik model dilakukan untuk mengetahui kesesuaian model yakni galat memenuhi asumsi *white noise*. Cek diagnostik disajikan pada tabel 4.

Tabel 4. Hasil cek diagnostik autokorelasi residual

lag	Q	Df	$\chi^2_{0.05,df}$	p-value
6	6.75	4	9.4877	0.150
12	13.11	10	18.30	0.218
18	15.40	16	26.29	0.496
24	16.73	22	36.415	0.778

Berdasarkan tabel 6 terlihat bahwa nilai *p-value* lebih dari 0,05 atau nilai *Q* untuk semua lag kurang dari nilai sehingga autokorelasi galat tidak signifikan atau tidak terdapat korelasi antar lag, yang berarti bahwa asumsi *white noise* dipenuhi.

Uji kenormalan galat dilakukan dengan menggunakan dengan uji normalitas *Geary's* dan diperoleh nilai  $|z| = 0.7687 < Z_\alpha = 1.645$ , sehingga disimpulkan bahwa galat dalam model berdistribusi normal.

Model Reg-ARIMA(1 0 0)(0 1 1)<sup>12</sup> untuk data banyak penumpang kereta yang mengandung variasi kalender Idul Fitri adalah

$$X_t = 36429.02 Y_t + Z_{t-1} + Z_{t-12} - Z_{t-13} - 0.4466e_{t-1} - 0.6251e_{t-12} + 0.2792 e_{t-13}$$

Tabel 5. Hasil Peramalan Model Reg-ARIMA

Bulan	Jumlah Penumpang	
	2015	2016
Januari	3,641	4,045
Februari	3,109	3,414
Maret	3,618	3,795
April	3,720	3,795
Mei	4,753	5,084
Juni	4,006	4,742
Juli	6,071	6,355
Agustus	5,397	5,446
September	5,561	5,233
Oktober	4,989	5,008
November	4,894	4,831
Desember	5,422	5,162

Berdasarkan Tabel 5, peramalan jumlah penumpang mengalami kenaikan pada bulan Juli, dan Desember. Banyaknya penumpang pada bulan Juli dipengaruhi karena adanya perayaan Idul Fitri. Dari analisis tersebut maka model Reg-ARIMA baik untuk meramalkan karena mampu menangkap terjadinya pergeseran perayaan Idul Fitri.



#### 4. Kesimpulan

Prosedur peramalan banyak penumpang kereta menggunakan model *Reg-ARIMA* terdiri dari empat tahap, yaitu diawali dengan membuat daftar tanggal perayaan atau liburan yang mengandung variasi kalender yang efeknya digunakan untuk menentukan variabel regresi (*regressor*) dan komponen regresi. Tahap berikutnya estimasi parameter menggunakan metode *maximum likelihood*, dilanjutkan dengan uji signifikansi parameter, dan uji galat apakah memenuhi asumsi *white noise* dengan uji *Ljung-Box Pierce* dan asumsi normal dengan uji *Geary's*. Model peramalan yang diperoleh adalah *Reg-ARIMA*  $(1,0,0)(0,1,1)^{12}$ .

Hasil peramalan penumpang kereta menggunakan model *Reg-ARIMA* menunjukkan adanya kenaikan pada bulan Juli, hasil peramalan dengan model *Reg-ARIMA* tersebut mampu menangkap fenomena pergeseran musim pada perayaan Idul Fitri.

#### Daftar Pustaka

- Bell, W.R., and Hillmer, S.C. 1983. *Modelling Time Series with Calendar Variation*. *Journal of the American Statistical Association*, 78.
- Liu, L. M. 1980. *Analysis of Time series with Calendar Effects*. *Management Science*, 26(1):106-112
- Otto, M. C., Bell, W. R. & Burman, J. P. (1987). An Iterative GLS Approach to Maximum Likelihood Estimation of Regression Models with ARIMA Errors. *In Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 1-26
- Putra, Erik. 2016. "Persentase Umat Islam di Indonesia Jadi 85 Persen". [REPUBLICA.CO.ID](http://REPUBLICA.CO.ID). 09 Januari 2016
- Shuja' *et al.* (2007). *Moving Holiday Effects Adjustment for Malaysian Economic Time series*. *Journal of Department of Statistics Malaysia*, 1-50.