

## PEMODELAN UNTUK PENGIRIMAN BARANG DENGAN MEMANFAATKAN JASA KERETA APIDI JAWA DAN SUMATERA

**Sartono**

Badan Pendidikan dan Pelatihan Keuangan, Kementerian Keuangan RI

[sartono82@kemenkeu.go.id](mailto:sartono82@kemenkeu.go.id)

### *Abstrak*

*Business on delivery sector has been growth recently, especially using railway. The development of infrastructures and facilities must be managed efficiently considering to an accurate information about the future need based on a historical data analysis. This article tried to develop models which can be used by decision makers to determine a reasonable plan to improve the profit. The Box-Jenkins method was employed to develop a model of historical data of goods distribution using train for Java and Sumatera area. The residuals was maintained using the GARCH model. This model was compared to the Trend and Seasonal Linear Model (TSLM) by evaluating the mean of absolute errors (MAE) and the root of mean squared errors (RMSE). The result shows that the ARIMA+GARCH model did better than ARIMA, SARIMA, and TSLM in predicting one and three month ahead for Java area, while the TSLM was more suitable for six and nine month forecasting than the others. On the other hand, the linear model provided the least values of MAE and RMSE for one, three, six, and nine month ahead prediction for Sumatera area.*

*Kata Kunci: ARIMA, linear model, shipping by train, time series analysis, trend and seasonal.*

### 1. PENDAHULUAN

Usaha jasa pengiriman barang mulai berkembang akhir-akhir ini. Hal ini didukung oleh perkembangan di bidang usaha kecil dan menengah khususnya bisnis online. Menurut hasil penelitian yang dikelola oleh TNS (Taylor Nelson Sofres), Asosiasi E-commerce Indonesia (idEA), dan Google Indonesia pada tahun 2013 nilai pasar *e-commerce* di Indonesia mencapai US\$8 dan diprediksi akan naik tiga kali lipat menjadi US\$25 pada tahun 2016 (JPNN, 2015). Perkembangan bisnis ini didukung oleh kebijakan Kementerian Komunikasi dan Informasi yang mempermudah pendirian bisnis online, direncanakan untuk membuka usaha *e-commerce* pengusaha hanya cukup dengan mendaftar dan memproses akreditasi saja tidak perlu sertifikasi (Kominfo, 2016). Seiring berkembangnya *e-commerce* pertumbuhan bisnis pengiriman barang pun mengalami peningkatan. Salah satu jasa pengiriman yang dimanfaatkan oleh pelaku bisnis penjualan melalui internet adalah jasa kereta api.

Pengguna jasa kereta api barang, selain penjual produk online, adalah perusahaan yang memproduksi barang dalam skala besar atau barang berat, misalnya pabrik semen, batubara, dan sebagainya. Berdasarkan data dari PT KAI (KAI, 2016) saat ini setidaknya sudah ada 5 perusahaan batu bara, 1

perusahaan bahan bakar minyak, 3 perusahaan semen, 5 perusahaan petikemas yang menjadi pelanggan, selain usaha jasa pengiriman barang habis pakai dan parcel.

Pertumbuhan kebutuhan jasa distribusi barang dengan menggunakan kereta api perlu dikelola oleh pengambil kebijakan baik internal KAI maupun pemerintah pada umumnya. Kemampuan untuk memprediksi jumlah barang yang akan diangkut dengan menggunakan kereta barang diperlukan guna memberikan data yang akurat untuk pengambilan keputusan yang tepat. Apabila prediksi ini terlalu jauh melenceng dari kenyataan maka dapat menimbulkan inefisiensi dalam penganggaran khususnya belanja barang dan modal.

Referensi mengenai pemodelan distribusi barang melalui kereta api sangat terbatas. Makalah ini diharapkan dapat menjadi salah satu rujukan bagi penelitian-penelitian selanjutnya dalam konteks yang serupa. Teknik-teknik peramalan yang selama ini digunakan untuk bidang lain diadopsi dalam penelitian ini untuk membuat suatu landasan yang nantinya dapat dikembangkan oleh peneliti yang tertarik mendalami bidang ini. Teknik analisa yang disajikan oleh Box-Ljung dimanfaatkan dalam artikel ini mengingat teknik ini selama ini yang sangat banyak diterapkan oleh para peneliti dari berbagai disiplin keilmuan.

Sabri(Shabri & Samsudin, 2014) menguji kehandalan model linear dibandingkan dengan ARIMA dan GARCH dalam memprediksi harga minyak dunia. Reza (Reza, 2015) menggunakan ARIMA untuk memprediksi waktu tempuh dan variasi relatif waktu tempuh pada jalan bebas hambatan di Charlotte, North Carolina. Ardebili (Ardebili et al., 2015) memanfaatkan ARIMA dalam rangka pemilihan portofolio saham di bursa efek Teheran. Holens(Holens, 1997) membandingkan antara ARIMA dan *neural network* untuk menganalisis data *future options*. ARIMA juga diterapkan oleh Sterba (Sterba et al., 2012) dalam peramalan konsumsi air di Mediterania.

Makalah ini menganalisis data berkala jumlah barang yang diangkut dengan kereta api di Pulau Jawa dan Sumatera serta mencari model yang paling rasional sesuai data tersebut. Beberapa model akan digunakan kemudian dibandingkan dengan menggunakan perbandingan terhadap *error* masing-masing model dalam peramalan untuk jangka waktu satu, tiga, enam, dan sembilan bulan.

## 2. METODE PENELITIAN

Makalah ini merupakan jenis analisa data berkala dan dalam prosesnya mengikuti tahapan; (1) pengumpulan data, (2) analisa data, (3) pembentukan dan pemilihan model, (4) validasi model. Semua perhitungan, pengujian hipotesis, maupun penggambaran grafik dilakukan dengan menggunakan *R-software*.

### a. Pengumpulan Data

Penelitian ini menganalisis data berkala distribusi barang dengan memanfaatkan jasa kereta api yang diperoleh dari Badan Pusat Statistik (BPS). Data tersebut merupakan data bulanan tentang jumlah barang yang didistribusikan melalui kereta api untuk wilayah Jawa dan Sumatera. Data tersebut memuat informasi dari tahun 2006 sampai dengan tahun 2015 dengan satuan ribu ton.

### b. Analisa Data

Terdapat 119 observasi untuk masing-masing wilayah, Jawa dan Sumatera. Data berkala dimulai dari hasil observasi pada bulan Januari 2006 sampai dengan observasi pada bulan November 2015. Analisa untuk wilayah Jawa dan wilayah Sumatera dilakukan secara terpisah. Untuk masing-masing lokasi, data tersebut akan dibagi menjadi dua bagian. Bagian pertama adalah data dari Januari 2006 sampai dengan Desember 2014. Bagian ini digunakan dalam pembentukan model. Sedangkan bagian kedua yakni hasil observasi pada tahun 2015 digunakan untuk validasi model.

### c. Pembentukan dan Pemilihan Model

Pembangunan model yang sesuai untuk data distribusi barang untuk area Jawa dan Sumatera dimulai dengan mengamati distribusi data melalui grafik. Grafik yang menyajikan data secara utuh maupun dekomposisi data dapat bermanfaat untuk menentukan model yang mampu menggambarkan kondisi yang relevan. Apabila grafik dekomposisi menunjukkan adanya tren atau gerakan musiman, maka model yang melibatkan tren atau siklus musiman dapat menjadi pilihan. Metode yang ditawarkan oleh Box-Jenkins (Box, Jenkins, & Reinsel, 1994) atau lebih umum dikenal dengan metode ARIMA dapat pula menjadi pilihan untuk membentuk model yang sesuai dengan sifat-sifat data. Secara umum model tersebut dituliskan dalam bentuk  $ARIMA(p, d, q)$ , dengan  $p$ ,  $d$ , dan  $q$  secara berturut-turut menyatakan order untuk komponen *autoregressive*, *differencing*, dan *moving average*.

Hal pertama yang dilakukan dalam menerapkan metode Box-Jenkins adalah dengan menginvestigasi data tersebut *stationary* atau tidak. Grafik yang diperoleh pada langkah di atas dapat digunakan untuk mengamati secara kasar stasionaritas data, namun tes formal diperlukan untuk memberikan hasil yang akurat. *Augmented Dickey-Fuller Test* (Dickey & Fuller, 1979) dapat digunakan dalam hal ini. Dengan memperhatikan model berikut

$$\nabla X_t = \gamma X_{t-1} + \beta + \delta t + y_t,$$

dimana  $y_t$  adalah suatu regresi yang *stationary* berorder  $p$ , yakni

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \omega_t,$$

dimana  $\omega_t$  adalah suatu *white noise*, dilakukan pengujian. Tes ini menginvestigasi  $H_0: \gamma = 0$  (data berkala *stationary*), dengan alternatif hipotesis  $H_1: \gamma < 0$ . Apabila data tidak *stationary* maka melakukan *differencing* data dapat menjadi solusi untuk menuju *stationary*. Apabila suatu data *stationary* setelah dilakukan *differencing* tingkat satu, maka pada model  $ARIMA(p, d, q)$  diperoleh nilai  $d=1$ . Sedangkan nilai  $p$  dan  $q$  secara intuitif dapat diamati dari diagram ACF dan PACF.

Homogenitas dari varian juga perlu diperhatikan. Jika varian bersifat heteroskedastik maka transformasi data terkadang dapat membantu menjadikan varian konstan terhadap waktu. Namun demikian, Chatfield (Chatfield, 2000) menyarankan untuk sedapat mungkin menghindari transformasi data. Dia mengingatkan bahwa melakukan transformasi dapat mengakibatkan parameter menjadi bias. Selain daripada itu, data hasil transformasi terkadang menjadi tidak memiliki makna ketika dikembalikan pada masalah aslinya.

Langkah selanjutnya, setelah memperoleh beberapa model yang mungkin dapat mewakili perilaku data adalah pemilihan model. Model terbaik dapat diperoleh dengan menggunakan *Akaike Information Criterion* (AIC)(Akaike, 1973) yang didefinisikan dengan

$$AIC = 2 \times \log(\text{likelihood}) + 2 \times k,$$

dimana  $k$  adalah banyaknya parameter.

#### d. Validasi Model

Model yang terbaik berdasarkan AIC perlu divalidasi untuk mengukur seberapa jauh ia dapat secara tepat mewakili perilaku distribusi barang dengan kereta api untuk masing-masing wilayah. Residual dari model yang dihasilkan haruslah *independent and identically distributed* (IID), konsekuensinya residual akan bersifat *white noise*. Jika tidak, maka model perlu diperbaiki untuk mendapatkan model yang lebih baik.

Varian dari residual juga harus konstan. Untuk mengetahui independensi dari residual dapat digunakan *McLeod-Li test*(McLeod & Li, 1983). Cara lain adalah dengan menggunakan *Box-Ljung test*(Ljung & Box, 1978) yang memeriksaketepatan suatu model  $ARMA(p,q)$ . Apabila varian tidak konstan, maka residual perlu dimodelkan dengan menggunakan ARCH/GARCH. Hal ini dapat diketahui dari diagram ACF dan PACF dari kuadrat residualnya. Jika dari diagram ini menggambarkan adanya *serialcorrelation* dari residual, maka ini membuktikan adanya heteroskedastisitas(Cowpertwait & Metcalfe, 2009).

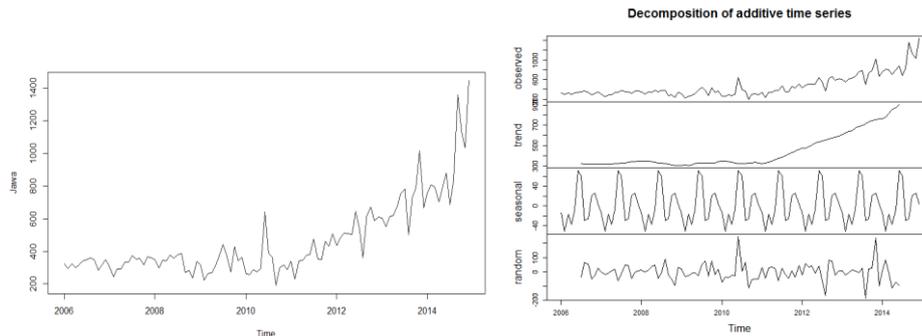
Setelah diperoleh model  $ARIMA(p, d, q)$  terbaik berdasarkan metode di atas, maka hasil yang diperoleh akan dibandingkan dengan model lain yang sesuai juga dengan perilaku data distribusi barang untuk wilayah Jawa dan Sumatera melalui kereta api tersebut. Dalam penelitian ini, model tersebut adalah *Trend and Seasonal Linear Model* (TSLM). Pada dasarnya model ini dapat dituliskan dengan  $X_t = m_t + s_t + \varepsilon_t$ , dimana  $m_t$  dan  $s_t$  masing-masing secara berurutan menyatakan tren dan gerakan musiman, dan  $\varepsilon_t$  adalah *error*.

Pada bagian akhir, model-model tersebut akan dibandingkan dengan melihat pada nilai *MeanAbsolute of Errors*(MAE) dan *Mean of Root Squared Errors*(RMSE). Model terbaik adalah model dengan MAE dan RMSE terkecil. Untuk keperluan ini, masing-masing model akan diterapkan pada data distribusi barang pada tahun 2015 sebagai bentuk eksternal validasi. Peramalan dilakukan untuk jangka waktu satu bulan, tiga bulan, enam bulan, dan sembilan bulan.

### 3. HASIL PENELITIAN DAN PEMBAHASAN

#### a. Wilayah Jawa

Pola distribusi barang melalui kereta api untuk wilayah Jawa dapat diamati pada gambar berikut



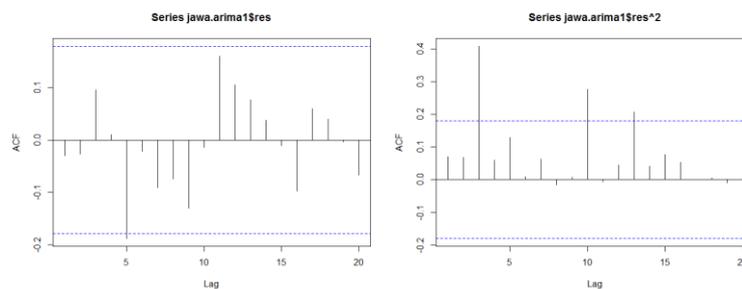
Gambar 1 Distribusi barang di Pulau Jawa dan dekomposisinya

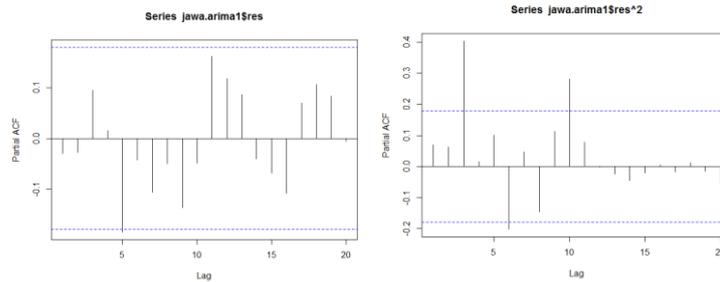
Dari gambar 1 dapat diamati ada semacam tren dan gerakan musiman pada data. Ini memberikan gambaran kasar untuk nantinya digunakan dalam menyusun model yang sesuai. Selain ARIMA, model linear dengan memasukkan unsur tren dan musiman kiranya dapat dimanfaatkan. Model ARIMA dengan memasukkan unsur musiman juga digunakan, yakni SARIMA( $p, d, q$ )( $P, D, Q$ ).

Tabel 1 AIC untuk pemodelan wilayah Jawa

No	ARIMA	AIC	No	ARIMA	AIC
1	(0,1,0)	1480,555	6	(1,1,2)	1458,961
2	(0,1,1)	1459,324	7	(2,1,0)	1458,654
3	(0,1,2)	<b>1457.189</b>	8	(2,1,1)	1460,024
4	(1,1,0)	1470,787	9	(2,1,2)	1460,955
5	(1,1,1)	1457,904			

Pertama-tama, penentuan model ARIMA. Nilai  $p=0.99$  pada *Augmented Dickey-Fuller Test* menunjukkan bahwa data *nonstationary*. Setelah melakukan *first differencing* maka nilai  $p=0.01$  yang menunjukkan data *stationary*. Dalam ARIMA model diperoleh  $d=1$ . Lebih lanjut, dengan memperhatikan nilai AIC pada table 1 diperoleh bahwa model terbaik adalah ARIMA(0,1,2). Memperhatikan grafik ACF dan PACF dari residual dan kuadrat residual model ARIMA(0,1,2) pada gambar 2





Gambar 2 ACF dan PACF residual dan kuadrat residual ARIMA(0,1,2)

nampak adanya autokorelasi diantara residual yang berarti residual perlu dimodelkan menggunakan GARCH. Hal ini didukung oleh Box-Ljung *test* yang memberikan nilai  $p=0.0001802$  yang berarti residual tidak independen. Dengan mempertimbangkan nilai AIC diperoleh model terbaik adalah GARCH(2,1). Jadi, diperoleh ARIMA(0,1,2) + GARCH(2,1) sebagai model yang terbaik diantara model ARIMA yang lain. Dengan mempertimbangkan faktor musiman, diperoleh SARIMA(0,1,1)(0,0,2).

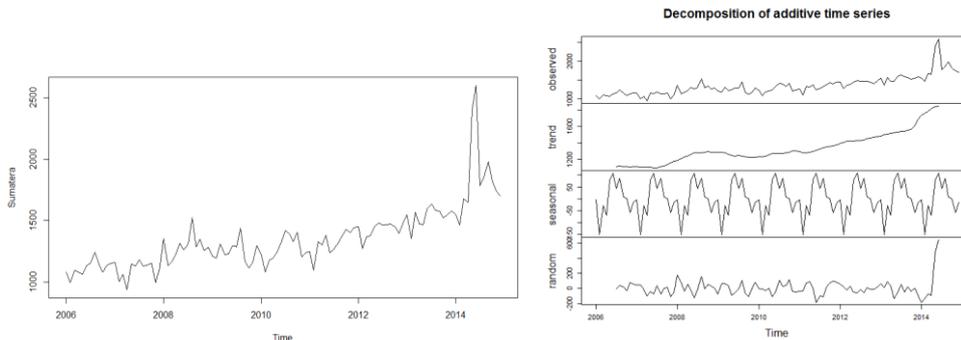
**b. Wilayah Sumatra**

Grafik distribusi barang untuk wilayah Sumatera dapat diamati pada gambar 3. Dapat diamati bahwa ada tren dan gerakan musiman pada data. Seperti halnya untuk wilayah Jawa, ARIMA dan SARIMA serta TSLM diterapkan untuk pemodelan data untuk wilayah Sumatera.

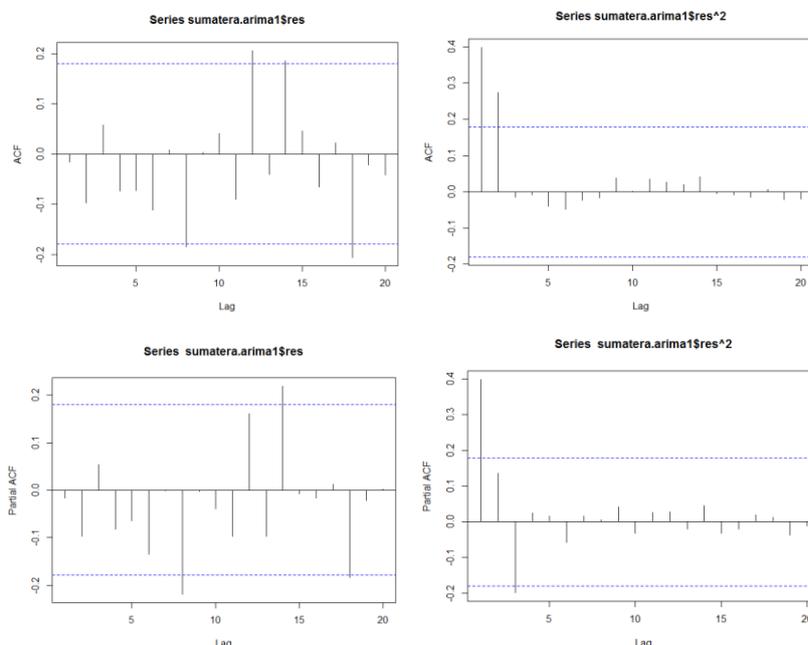
Pada pemodelan dengan ARIMA, menggunakan *Augmented Dickey-Fuller Test* diperoleh nilai  $p = 0.1075$  yang menunjukkan bahwa data *nonstationary*. Dengan *first differencing* dihasilkan nilai  $p=0.01$  yang menunjukkan data sudah *stationry* sekarang. Dari langkah ini diperoleh nilai  $d=1$  pada model ARIMA. Nilai AIC pada table 2 menunjukkan bahwa model terbaik adalah ARIMA(1,1,1).

Tabel 2 AIC untuk pemodelan wilayah Sumatera

No	ARIMA	AIC	No	ARIMA	AIC
1	(0,1,0)	1518.394	6	(1,1,2)	1504.874
2	(0,1,1)	1510.463	7	(2,1,0)	1509.463
3	(0,1,2)	1503.787	8	(2,1,1)	1504.975
4	(1,1,0)	1514.963	9	(2,1,2)	1505.409
5	(1,1,1)	<b>1503.148</b>			



Gambar 3 Distribusi barang di Pulau Sumatera dan dekomposisinya  
 Grafik ACF dan PACF dari residual dan kuadrat residual model ARIMA(1,1,1) pada gambar 4 memperlihatkan bahwa residual tidak independen sehingga perlu dimodelkan menggunakan GARCH.



Gambar 4 ACF dan PACF residual dan kuadrat residual ARIMA(1,1,1)

Lebih lanjut Box-Ljung *test* dengan nilai  $p=0,00002472$  mendukung hasil tersebut. Model terbaik adalah GARCH(2,2) dengan memperhatikan nilai AIC untuk model GARCH. Dengan demikian, ARIMA(1,1,1) + GARCH(2, 2) adalah model terbaik diantara model ARIMA yang lain. Untuk data wilayah Sumatera dengan mempertimbangkan efek musiman diperoleh SARIMA(0, 1, 2)(0,1,1) sebagai yang paling mewakili.

**c. Perbandingan Keandalan antar Model**

Dari bagian 3.1 dan 3.2 telah diperoleh model ARIMA, SARIMA, dan TSLM yang terbaik diantara model lain dalam kelompok mereka. Pada bagian ini, model-model tersebut diperbandingkan untuk mendapatkan model yang paling

efektif jika digunakan dalam peramalan pada data distribusi barang di Jawa dan Sumatera. Untuk wilayah Jawa, nilai MAE dan RMSE masing-masing model dapat diamati pada tabel 3.

Tabel 3. Perbandingan MAE dan RMSE masing-masing model untuk data wilayah Jawa.

Pembanding	Waktu (bulan)	ARIMA(0,1,2)	SARIMA(0,1,1) (0,0,2)	ARIMA + GARCH	TSLM
MAE	1	222.8617	187.46908	<b>23.47318</b>	278.3264
	3	379.8135	363.70967	<b>41.16356</b>	113.5116
	6	391.8848	405.75383	78.08178	<b>72.38426</b>
	9	421.2419	449.36725	94.49896	<b>82.71373</b>
RMSE	1	222.8617	187.46908	<b>23.47318</b>	278.3264
	3	395.7035	384.48015	<b>58.20200</b>	163.3407
	6	400.6379	417.53375	<b>118.2817</b>	<b>118.4946</b>
	9	433.3178	467.78656	132.9885	<b>126.6766</b>

Berdasarkan nilai MAE dan RMSE diperoleh bahwa model ARIMA+GARCH lebih baik untuk peramalan jangka pendek yakni satu dan tiga bulan, sedangkan TSLM lebih baik jika digunakan untuk meramalkan pada jangka waktu yang lebih panjang yakni dalam tabel 3 tersebut adalah untuk waktu enam dan sembilan bulan.

Tabel 4. Perbandingan MAE dan RMSE masing-masing model untuk data wilayah Sumatera.

Pembanding	Waktu (bulan)	ARIMA(1,1,1)	SARIMA(0,1,2) (0,1,1)	ARIMA + GARCH	TSLM
MAE	1	73.27304	95.34274	252.3072	<b>13.54861</b>
	3	133.3881	150.80327	227.1477	<b>47.33102</b>
	6	131.9417	236.80307	199.7646	<b>60.25926</b>
	9	169.372	182.43788	156.1048	<b>124.8418</b>
RMSE	1	73.27304	95.34274	252.3072	<b>13.54861</b>
	3	172.7072	158.9703	228.3676	<b>53.1016</b>
	6	166.4562	255.3461	204.7704	<b>71.67402</b>
	9	197.5281	213.4328	172.6288	<b>159.9581</b>

Berbeda dengan wilayah Jawa, tabel 4 menunjukkan bahwa, untuk wilayah Sumatera, TSLM selalu menjadi model yang lebih baik digunakan dalam peramalan untuk jangka waktu satu, tiga, enam, dan sembilan bulan dibandingkan ARIMA, ARIMA+GARCH, dan SARIMA.

#### 4. SIMPULAN

Penelitian ini mencoba menemukan model terbaik yang dapat mewakili perilaku dari data distribusi barang melalui kereta api untuk wilayah Jawa dan Sumatera. Dalam penelitian ini, ditemukan hal-hal berikut:

1. Untuk wilayah Jawa, menggunakan ARIMA+GARCH dalam peramalan akan memberikan hasil yang lebih akurat untuk meramal dalam jangka waktu satu dan tiga bulan dibandingkan model ARIMA lainnya serta *Trend and Seasonal Linear Model*.

2. *Trend and Seasonal Linear Model* lebih mampu memprediksi nilai distribusi barang di wilayah Jawa untuk enam dan sembilan bulan dibandingkan model-model yang sekelompok dengan ARIMA.
3. Distribusi barang di wilayah Sumatera dapat diramalkan lebih baik oleh *Trend and Seasonal Linear Model* untuk satu, tiga, enam, dan sembilan bulan.
4. Perilaku menghiraukan independensi dari residual akan membuat model kurang akurat digunakan dalam peramalan, sebaliknya, dengan memodelkan residual yang tidak independen menggunakan ARCH/GARCH dapat mengoptimalkan akurasi model dalam peramalan distribusi barang melalui kereta api baik untuk wilayah Jawa maupun Sumatera.

## 5. DAFTAR PUSTAKA

- Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving-average models. *Biometrika*, 60, 255-266.
- Ardebili, M., Hashemi, M., Shahabi, A., & Barough, M. (2015). Optimized selection of stock portfolio by using the fuzzy artificial neural networks web model, ARIMA & markowitz model in tehran stock exchange. *European Online Journal of Natural and Social Sciences*, 4(1), 831-844.
- Box, G., Jenkins, G., & Reinsel, G. (1994). Time Series Analysis, Forecasting and Control. *Prentice-Hall, Englewood Cliffs, NJ*.
- BPS. (2015). Data distribusi barang dengan menggunakan jasa kereta api untuk wilayah Jawa dan Sumatera.
- Chatfield, C. (2000). *Time-Series Forecasting*. New York: Chapman & Hall.
- Cowpertwait, P., & Metcalfe, A. (2009). *Introductory Time Series with R*. New York: Springer.
- Dickey, D., & Fuller, W. (1979). Distribution of the estimates for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* 74, 427--431.
- Holens, G. (1997). Forecasting and selling futures using ARIMA models and a neural network. *ProQuest*, (304378804).
- JPNN. (2015, April 27). *Perkembangan Bisnis e-Commerce di Indonesia Melesat*. Dipetik Januari 25, 2016, dari <http://www.jpnn.com/read/2015/04/27/300672/Perkembangan-Bisnis-e-Commerce-di-Indonesia-Melesat>
- KAI. (2016). *Pelayanan Angkutan*. Dipetik Januari 25, 2016, dari <http://kargo.kereta-api.co.id/>

- Kominfo. (2016, Januari 22). *Komitmen Kominfo Kembangkan e-Commerce*. Dipetik Januari 25, 2016, dari [http://kominfo.go.id/index.php/content/detail/6622/Komitmen+Kominfo+Kembangkan+%3Ci%3Ee-Commerce%3C-i%3E/0/berita\\_satker#.VqWcxllavnI](http://kominfo.go.id/index.php/content/detail/6622/Komitmen+Kominfo+Kembangkan+%3Ci%3Ee-Commerce%3C-i%3E/0/berita_satker#.VqWcxllavnI)
- Ljung, G., & Box, G. (1978). On a measure of lack of time series models. *Biometrika*, 553-564.
- McLeod, A., & Li, W. (1983). Diagnostic checking ARMA timeseries models using squared residual autocorrelations. *Journal of Time Series Analysis*, 4, 269-273.
- Reza, R. M. (2015). Forecasting travel time and variations in travel time due to vehicle accidents in spatio-temporal context along freeway. *ProQuest*, (1752635231).
- Shabri, A., & Samsudin, R. (2014). Crude oil price forecasting based on hybridizing wavelet multiple linear regression model, particle swarm optimization techniques, and principal component analysis. *The Scientific World Journal*, doi:<http://dx.doi.org/10.1155/2014/854520>.
- Sterba, J., Rublíková, E., & y Ocerín, J. (2012). Hybrid ARIMA-neural network model for prediction of water consumption aggregate. *Social Science Letters*, 2(2), 14-20.