

Prediksi Harga Cabai dengan Pemodelan Time Series ARIMA

F.N Hadiansyah

School of Computing, Telkom University
Jl. Telekomunikasi Terusan Buah Batu Bandung 40257, Indonesia
nurhadiansyahfikri@gmail.com

Abstract

Chili price is one of the food commodities that can affect the value of inflation. The price of chili that is less erratic and even tend to continue to increase at some particular time will be bad for the state and society. ARIMA modeling is one of time series model that can be used to predict the possibility of the increasing of chili price. The ARIMA model uses previous chili price data to be used as a comparison so it can predict the price of chili in the future. The data will be analyzed and made into ARIMA model, then it will be determined which ARIMA model is possible so that an efficient model for chili price will be obtained. The results of this study yielded models of ARIMA (1 2 1), AR (1), ARI (1 2) and IMA (2 1) then there are tested with goodness fit to determine the best model by using RMSE, MAPE and R-square. Based on training and testing skenario, the best result by validation methods is ARIMA

Keywords: Time Series, ARIMA, Forecasting

Abstrak

Harga cabai merupakan salah satu komoditas pangan yang dapat berpengaruh terhadap nilai inflasi. Harga cabai yang kurang menentu dan bahkan cenderung terus mengalami kenaikan pada beberapa waktu tertentu akan berakibat buruk bagi negara serta masyarakat. Pemodelan ARIMA merupakan salah satu model time series yang dapat digunakan untuk memprediksi kemungkinan-kemungkinan terjadinya kenaikan harga cabai. Model ARIMA menggunakan data-data harga cabai sebelumnya untuk dijadikan sebagai perbandingan sehingga dapat memprediksi harga cabai di masa yang akan datang. Data akan dianalisis dan dibuatkan kedalam beberapa bentuk model ARIMA, kemudian ditentukan model ARIMA mana saja yang memungkinkan untuk dimodelkan sehingga akan didapatkan model yang efisien untuk pemodelan harga cabai ini. Hasil penelitian ini menghasilkan 4 model yaitu ARIMA (1 2 1), AR (1), ARI (1 2) dan IMA (2 1) yang kemudian dilakukan uji kebaikan untuk menentukan model terbaik dengan menggunakan RMSE, MAPE dan R-square. Berdasarkan skenario testing dan training, diperoleh model dengan performansi yang baik adalah ARIMA

Kata Kunci: *Time Series*, ARIMA, peramalan

I. PENDAHULUAN

Pemodelan time series banyak digunakan dalam bidang keuangan untuk peramalan atau *forecasting*, model time series mengamati perilaku observasi yang diambil dari waktu ke waktu secara berurutan (Wei, 2006). Pada penelitian-penelitian sebelumnya, model time series telah banyak diimplementasikan untuk memprediksi suatu observasi yang berkaitan dengan waktu, seperti Aridita (2010) yang membahas tentang prediksi penjualan baju di Boyolali dan Siana (2011) yang membahas tentang prediksi *time series* berbasis multivariat.

Pada penelitian ini, metode pemodelan time series digunakan untuk memodelkan harga cabai, sehingga dapat dilakukan peramalan harga cabai pada masa mendatang. Cabai merupakan salah satu komoditas pangan yang dibutuhkan dalam kehidupan sehari-hari. Faktanya kelangkaan produksi cabai menyebabkan semakin banyaknya permintaan sehingga harga cabai pun kian melonjak. Melonjaknya harga cabai yang tidak menentu

menyebabkan kesulitan dalam mempersiapkan cadangan produksi apabila terjadi kelangkaan (sumber: www.liputan6.com). Informasi naik turunnya harga cabai pada waktu-waktu sebelumnya, dapat menjadi variabel baru yang dipertimbangkan dalam pemodelan time series, yang dikenal sebagai volatilitas atau simpangan baku bersyarat terhadap waktu. Melihat permasalahan tersebut di harapkan pemodelan time series dapat digunakan dengan baik untuk memprediksi fluktuasi tren harga cabai di Indonesia, sebagai langkah antisipasi permintaan pasar.

II. TINJAUAN PUSTAKA

A. TIME SERIES MODEL ARIMA

Time Series merupakan kejadian-kejadian yang terjadi berdasarkan waktu-waktu tertentu secara berurutan. Forecasting dalam time series merupakan prediksi untuk memperkirakan kejadian-kejadian yang akan terjadi di masa yang akan datang berdasarkan data-data sebelumnya[1].

Salah satu model dalam time series adalah ARIMA. ARIMA sering juga disebut sebagai runut waktu Box-Jenkins[4]. ARIMA biasanya digunakan untuk peramalan jangka pendek karena untuk peramalan jangka panjang, model ini memiliki ketepatan yang kurang baik karena cenderung akan datar untuk periode yang panjang[6]. Model Arima didefinisikan sebagai berikut [1]:

1. Autoregressive Model (AR)

Bentuk umum model *autoregressive* dengan orde p (AR(p)) atau model ARIMA (p,0,0) dinyatakan sebagai berikut:

$$Y_t = \alpha_0 + \theta_1 Y_{t-1} + \dots + \theta_p Y_{t-p} + e_t \quad (1)$$

Y_t = nilai observasi pada saat t

α_0 = konstanta

θ_p = parameter autoregresive ke p

e_t = nilai galat saat t

2. Moving Average (MA)

Bentuk umum model *moving average* orde q (MA(q)) atau ARIMA (0,0,q) dinyatakan sebagai berikut:

$$Y_t = \theta_0 + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} \quad (2)$$

θ_0 = konstanta

θ_q = parameter moving average ke q

e_{t-p} = nilai galat pada saat t - k

3. Proses Autoregressive Moving Average (ARMA)

Model umum untuk campuran proses AR (1) murni dan MA(1) murni, misalkan ARIMA (1,0,1) dinyatakan sebagai berikut:

$$Y_t = \alpha_0 + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} \quad (3)$$

4. Proses Autoregressive Integrated Moving Average (ARIMA)

Apabila non-stasioner ditambahkan pada campuran proses ARMA, maka memenuhi model umum ARIMA (p, d, q) terpenuhi. Persamaan untuk kasus sederhana ARIMA (p,1,q) adalah sebagai berikut:

$$Y_t = (1 + \alpha_1)y_{t-1} + \dots + (1 + \alpha_p)y_{p-1} + e_t + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} \quad (4)$$

B. ROOT MEAN SQUARE ERROR

RMSE merupakan untuk mencari kesalahan dari rata-rata error pada observasi (Willmott dan Matsuura 2005). RMSE dapat digunakan mencari tahu seberapa besar kesalahan pada data dari model yang digunakan. RMSE dapat dijadikan sebagai indikator ketidakcocokan dalam pemodelan. RMSE dapat dicari dengan menggunakan:

$$RMSE = \sqrt{\frac{\sum_i^n (\hat{x} - x_i)^2}{n}} \quad (5)$$

\hat{x} = nilai hasil forecast

x_i = nilai observasi ke - i

n = banyaknya data

C. MEAN ABSOLUTE PERCENTAGE ERROR

Pakaja (2012) menjelaskan bahwa Mean Absolute Percentage Error (MAPE) adalah kesalahan absolut pada tiap periode dibagi dengan nilai observasi yang nyata untuk periode itu. Kemudian dihitung rata-rata dari kesalahan persentase absolut tersebut. Sama seperti halnya RMSE, MAPE menghitung kesalahan dari data observasi dan prediksi yang dimunculkan dalam nilai persen. Nilai MAPE dapat dihitung dengan persamaan berikut :

$$MAPE = \left(\frac{100\%}{n} \right) \frac{\sum_i^n x_i - \hat{x}}{x_i} \quad (6)$$

D. R-SQUARED

Menurut A. Colin dan Frank (1995), R-squared adalah sebuah metode untuk model regresi linear untuk uji kebaikan dari sebuah mode. R-square bernilai 0-1, semakin nilai mendekati angka 1 maka model tersebut semakin baik. Untuk mencari nilai R-square dapat dicari dengan menggunakan

$$R^2 = 1 - \frac{\sum_i^n (\hat{x} - x_i)^2}{\sum_i^n (\hat{x} - \bar{x})^2} \quad (7)$$

III. FORECASTING

A. DATA

Data yang digunakan adalah data harga cabai periode 1 Mei 2016 – 14 Januari 2017 sebanyak 236 data yang diambil dari <http://infopangan.jakarta.go.id/publik/> yang ditunjukkan pada Gambar 1.

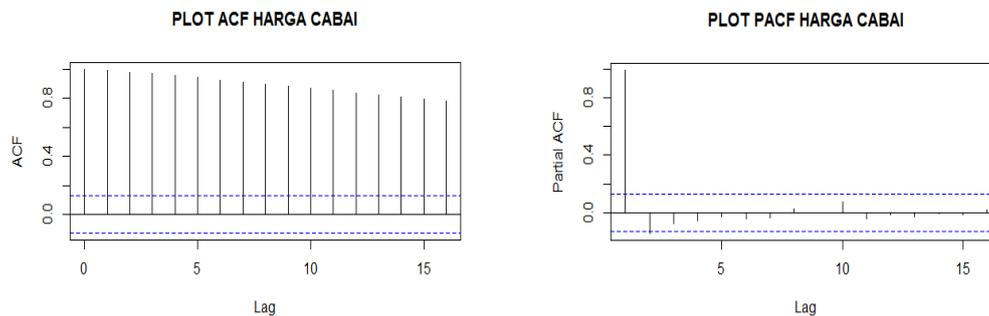


Gambar 1. Grafik time series harga cabai

Gambar merupakan grafik harga cabai dari yang diamati runtun waktu. Karakteristik pada data menunjukkan harga cabai cenderung mengalami tren kenaikan setiap periode dengan penurunan yang tidak signifikan. Dapat diperoleh perhitungan bahwa nilai variansi atau sebaran grafik tersebut tinggi, nilai variansi yang tinggi ekuivalen dengan risiko yang besar. Sehingga melakukan prediksi harga cabai di masa mendatang menjadi hal yang signifikan untuk mengantisipasi kebutuhan dan risiko pasar

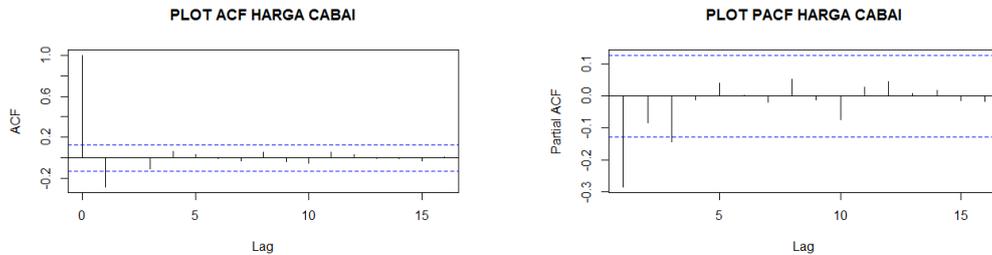
B. PENENTUAN MODEL ARIMA

Karakteristik pemodelan time series yang utama adalah adanya kebergantungan atau *dependency* observasi pada saat t terhadap oservasi di waktu sebelumnya. Sehingga perlu diperhitungkan dan ditentukan batasan waktu observasi yang mempengaruhi observasi berikutnya. Untuk menentukan kebergantungan observasi pada model time series yang digunakan, ACF (*Partial Autocorrelation Function*) dan PACF (*Partial Autocorrelation Function*) pada data sehingga dapat diketahui nilai korelasi antar observasi yang menjadi batasan model time series dalam prediksi harga cabai ini.



Gambar 2. Plot ACF (*Autocorrelation Function*) (kiri) dan Plot PACF (*Partial Autocorrelation Function*) (kanan) data harga cabai

Pada Gambar 2 terlihat bahwa pada plot PACF terjadi *cut off* pada lag pertama, artinya observasi harga cabai hanya dipengaruhi oleh harga cabai pada satu hari sebelumnya. Hal ini mengarahkan kita pada model time series AR(1) yang melibatkan observasi sebelumnya dengan selisih waktu satu lag. Pada plot ACF tidak terjadi *cut off* sehingga dilakukan normalisasi data dengan menggunakan fungsi logaritma, sehingga diperoleh plot ACF dan PACF untuk data normalisasi sebagai berikut.



Gambar 3. Plot ACF (Autocorrelation Function) (kiri) dan Plot PACF (Partial Autocorrelation Function) (kanan) data normalisasi harga cabai

Dari Gambar 3 terlihat pada plot ACF terdapat *cut off* pada lag ke-2 dan pada plot PACF terdapat *cut off* pada lag ke-1. Hasil tersebut mengarahkan kita pada model time series dengan faktor normalisasi yang dikenal dengan *integrated*, yaitu model ARI (1,2), IMA (2,1) dan ARIMA (1,2,1). Diperoleh empat model time series berdasarkan nilai ACF dan PACF, yang selanjutnya dapat di estimasi parameter model dan diuji untuk mendapatkan model prediksi yang baik untuk harga cabai.

C. ESTIMASI PARAMETER

Parameter untuk model-model yang telah didapatkan dicari dengan menggunakan *Maximum Likelihood Estimator*. Untuk penurunan rumusnya dapat dilihat pada (Wei, 2006),

Tabel 1. Nilai Parameter Model Time Series

AR(1)		ARI(1,1)		IMA(1,1)		ARIMA(1,1,1)		
α_0	α_1	α_0	α_1	θ_0	θ_1	α_0	α_1	θ_1
40.8998	0.996968	0.045	-0.556	0.198	0.994	0.603	0.444	0.556

Berdasarkan parameter pada Tabel 1, maka diperoleh model-model time series untuk harga cabai sebagai berikut,

1. AR (1):

$$Y_t = 0.00180461 + 0.996968Y_{t-1}$$

2. ARI (1,2)

$$Y_t = 0.0605 + 0.444Y_{t-1} - 0.556Y_{t-2}$$

3. IMA (2 1)

$$Y_t = 0.198 + 0.994Y_{t-1}$$

4. ARIMA (1 2 1)

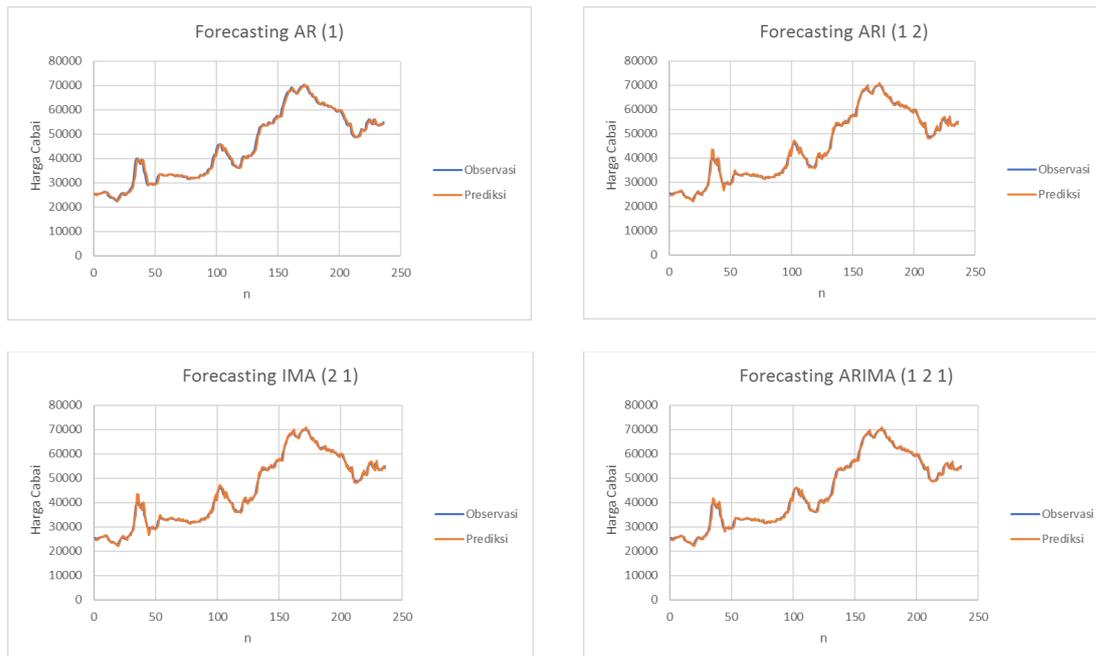
$$Y_t = 0.603 + 0.444Y_{t-1} - 0.556Y_{t-2}$$

dengan menggunakan model-model tersebut, dilakukan *forecasting* dan validasi kebaikan model, sehingga diperoleh model yang baik untuk memprediksi harga cabai di masa mendatang.

IV. HASIL DAN DISKUSI

A. FORECASTING DENGAN DATA *TRAINING*

Hasil dari penelitian ini adalah untuk mencari model ARIMA mana yang lebih cocok untuk digunakan pada data harga cabai ini. Dari model yang sudah didapatkan, data dijadikan sebagai *input* untuk masing-masing model sehingga nilai-nilai prediksi akan keluar sebagai *output* dari masing-masing model tersebut. Pada Gambar 4 di tunjukan hasil dari prediksi masing-masing model yang dibandingkan dengan data observasi yang asli.



Gambar 5. Plot Data observasi dengan prediksi AR (1) (kiri atas), plot data observasi ARI (1 2) (kanan atas), plot data observasi IMA (2 1) (kiri bawah) dan Plot Data observasi dengan prediksi ARIMA (1 2 1) (kanan bawah)

Terlihat Hasil *output* masing-masing model prediksi dibandingkan dengan data observasi sekilas dari ke empat model tersebut hampir sama sehingga harus dilakukan uji kebaikan untuk melihat model mana yang lebih cocok. Pada Tabel 5 di tunjukan hasil numerik dari masing-masing model yang didapatkan untuk nantinya digunakan sebagai *input* mencari model terbaik pada masing-masing teknik uji kebaikan yang sudah dijelaskan sebelumnya.

Tabel 2. Forecasting data training masing-masing model

NO	Tanggal	AR (1)		ARI (1 2)		IMA (2 1)		ARIMA (1 2 1)	
		Data	Forecast	Data	Forecast	Data	Forecast	Data	Forecast
1.	4-1-2017	50795	48909	50795	49069	50795	48771	50795	48988
2.	5-1-2017	51935	49414	51935	49801	51935	49541	51935	49680
3.	6-1-2017	51356	50770	51356	51857	51356	51396	51356	51425
4.	7-1-2017	51419	51907	51419	53157	51419	52750	51419	52469
5	8-1-2017	51682	51329	51682	51390	51682	51623	51682	51118
6	9-1-2017	54261	51392	54261	51260	54261	51608	54261	51468
7	10-1-2017	55067	51654	55067	51879	55067	51902	55067	51821
8	11-1-2017	55977	54226	55977	56025	55977	55413	55977	55445

9	12-1-2017	55617	55030	55617	56505	55617	56084	55617	55458
10	13-1-2017	54755	55938	54755	56855	54755	56954	54755	56416
11	14-1-2017	53972	55579	53972	55711	53972	56068	53972	55486
12	15-1-2017	56025	54719	56025	54076	56025	54690	56025	54396
13	16-1-2017	54244	53938	54244	53166	54244	53626	54244	53646
14	17-1-2017	53932	55986	53932	57079	53932	56627	53932	56975
15	18-1-2017	53700	54209	53700	53825	53700	53908	53700	53470
16	19-1-2017	53395	53898	53395	53105	53395	53607	53395	53815
17	20-1-2017	53886	53667	53886	53445	53886	53414	53886	53618
18	21-1-2017	54341	53363	54341	53121	54341	53103	54341	53279
19	22-1-2017	54093	53852	54093	54100	54093	53905	54093	54128
20	23-1-2017	55025	54306	55025	54814	55025	54533	55025	54567

Dari Tabel 2 terlihat bahwa masing-masing model memiliki hasil prediksi yang berbeda, terlihat dari masing-masing model terdapat beberapa nilai yang hampir mendekati nilai data training pada t tertentu namun terdapat juga yang perbedaannya nilainya cukup jauh dengan data training sehingga diperlukan perhitungan selanjutnya untuk mencari model yang paling efisien. Untuk menentukan model yang paling efisien harus dicari terlebih dahulu nilai kesalahan untuk masing-masing model sehingga dapat ditentukan model mana yang paling baik, maka dari itu digunakan RMSE, MAPE dan R-Square seperti yang diperlihatkan pada Tabel 3,

Tabel 3. Validasi Kebaikan Model

	AR (1)	ARI (1 2)	IMA (2 1)	ARIMA (1 2 1)
RMSE	1478.625	1052.968	1000.410	964.005
MAPE	1.947%	1.70%	1.656%	1.479%
R-Square	0.989	0.994	0.995	0.995

Pada Tabel 3 terlihat untuk masing-masing model memiliki nilai RMSE yang berbeda untuk masing-masing model dan memiliki selisih yang cukup jauh pada masing-masing model. Model ARIMA memiliki nilai RMSE yang paling baik adalah dengan nilai RMSE yang kecil. Sama seperti halnya seperti nilai RMSE, nilai MAPE seperti yang ditunjukkan pada Tabel 7 memiliki selisih yang cukup jauh untuk masing-masing model. Model ARIMA memiliki nilai MAPE kecil yang dipilih sebagai model yang paling efisien sama seperti RMSE. Masing-masing memiliki nilai R-square yang mendekati satu, hal ini menunjukkan model time series dapat menjelaskan variable harga cabai secara signifikan dengan melibatkan observasi masa lalu (histori).

B. FORECASTING DENGAN DATA TESTING

Untuk mendapatkan model yang paling baik, dilakukan peramalan dengan data testing, dimana telah dipisahkan 10 data yang tidak dimasukkan dalam pengolahan model diatas untuk mengecek tingkat akurasi dari model yang telah didapatkan seperti yang ditunjukkan pada Tabel 4,

Tabel 4. Forecasting harga cabai dengan data testing

No	Data Testing	Forecast AR (1)	Forecast ARI (1 2)	Forecast IMA (2 1)	Forecast ARIMA (1 2 1)
1.	32140	31977	32175	32175	32137
2.	31884	32163	32389	32389	32252
3.	32093	31908	31790	31790	31794
4.	32023	32116	32142	32142	32213
5.	32000	32047	32057	32057	32016
6.	32116	32024	31965	31965	32014

7.	32767	32139	32188	32188	32193
8.	33093	32789	33234	33234	33089
9.	32930	33114	33539	33539	33269
10.	32930	32951	32945	32945	32884
Validasi Keباikan Model dengan Data Testing					
RMSE	305.348	396.221	358.490	383.827	
MAPE	0.651%	0.885%	0.712%	0.672%	
R-SQUARE	0.635	0.405	0.513	0.537	

Sama halnya seperti hasil training, pada uji testing ini masing-masing model memiliki nilai yang berbeda dan terdapat beberapa nilai model yang hampir mendekati nilai data testing pada saat t tertentu. Maka dari itu, dari hasil testing ini perlu dilakukan juga uji keباikan model dengan teknik-teknik yang sama dengan training sebagai validasi perbandingan antara nilai training dengan testing apakah berbanding lurus atau berbanding terbalik.

V. KESIMPULAN

Dari hasil penelitian, untuk data prediksi harga cabai dengan menggunakan model *time series* ARIMA dengan data *training*, berdasarkan nilai RMSE dan MAPE, model yang paling baik menunjukkan pada model ARIMA (1 2 1) dengan nilai RMSE 964.005 dan nilai MAPE 1.479%. Sedangkan apabila dilihat dari nilai R-square, model yang paling baik menunjukkan pada model IMA (2 1) dan ARIMA (1 2 1). Untuk data *testing*, berdasarkan nilai RMSE, MAPE dan R-Squared, model yang paling baik menunjukkan pada model AR (1) dengan nilai RMSE 305.348, nilai MAPE 0.651% dan nilai R-square 0.635. Secara keseluruhan, baik testing dan training, model ARIMA memiliki performansi yang baik untuk memprediksi harga cabai di masa mendatang, sebagai langkahantisipasi permintaan pasar yang fluktuatif.

REFERENCES

- [1] Wei, W. S. 2006. Time Analysis Univariate and Multivariate Methods, New York : Addison Wesley Publishing Company, Inc
- [2] Halim, Siana dan Chandra, Arief. 2011. Pemodelan Time Series Multivariat secara Automatis. Jurnal Teknik Industri Vol 13
- [3] Perdana, Sukma. 2010. Perbandingan Metode Time Series Regression Dan Arimaxpada Pemodelan Data Penjualan Pakaian Di Boyolali. Surabaya
- [4] Huang, Theresa. 2013. The Box-Jenkins Methodology for Time Series Models.SAS Global Forum, 2013
- [5] Box, G.E.P., & Jenkins, G.M. 1976. Time Series Analysis, Forecasting and Control. HoldenDay: San Francisco.
- [6] Octora, Metta & Kunotoro. 2013. Perbandingan Metode ARIMA (Box Jenkins) dan Metode Winter dalam Peramalan Jumlah Kasus Demam Berdarah Dengue. Jurnal Biometrika dan Kependudukan, Vol. 2, No. 1 Juli 2013: 88–98
- [7] Willmott J Cort dan Matsuura Kenji. 2005. Advantages of the Mean Absolute Error (MAE) Over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. Department of Geography, University of Delaware. Newark USA.
- [8] Pakaja, F., Naba, A., Purwanto. (2012), Peramalan Penjualan Mobil Menggunakan Jaringan Syaraf Tiruan dan Certainty Factor, Jurnal EECCIS, Vol.6, No.1, Juni 2012.
- [9] A. Colin & A.G. Frank. 1995. An R-squared measure of goodness of fit for some common nonlinear regression models.
- [10] Febri, kiki; Hoyyi, Abdul & Mukid, Abdul. 2014. Verifikasi Model Arima Musiman Menggunakan Peta Kendali Moving Range. JURNAL GAUSSIAN, Volume 3, Nomor 4, Tahun 2014, Halaman 701 - 710 .