

Kriteria Pemilihan Model Peramalan Terbaik Berdasarkan Kriteria Informasi

¹Reni Agustini, ²Nusar Hajarisman, ³Siti Sunendiari

¹Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Bandung
Jl. Tamansari No. 1, Bandung 40116

e-mail : ¹agustinireni26@gmail.com, ²nusarhajarisman@yahoo.com, ³sunen_diari@yahoo.com

Abstract. This thesis discusses the selection of time series models commonly used for certain forecasts. The time series model observed here is a model that follows the autoregressive or AR process. There are six models considered in this thesis which are AR (1,0,0), AR (1,1,0), AR (1,2,0), AR (2,0,0), AR (2, 1,0), and AR (2,2,0). In statistical analysis there are a number of measures or criteria commonly used for the selection of the best model, where one of them is using the information criteria. The commonly used information criteria for selecting the best model are the Akaike (AIC) information criteria and Bayesian information criteria (BIC). However, in this thesis will be introduced a new criterion that is empirical information criterion (EIC). This method will be applied to data on the price of red pepper producers in Cianjur regency of West Java. The results obtained that by using AIC and BIC information criteria the best forecasting model is ARIMA (1,2,0) ie with value AIC = 1411,3 and BIC = 1415,12 whereas by using EIC information criteria the best forecasting model is ARIMA (2, 2,0) ie with the value of EIC = 1384.98. The result of forecasting is using ARIMA model (2,2,0) this is because the value of EIC work better. After forecasting, the price of red pepper in Kabupaten Cianjur experienced the highest increase occurred in April of 2017 and the lowest in May of 2016.

Key Word – AIC, BIC, EIC, *time series*,

Abstrak. Makalah ini membahas tentang pemilihan model deret waktu yang biasa digunakan untuk peramalan tertentu. Model deret waktu yang diamati disini adalah suatu model yang mengikuti proses autoregresif atau AR. Ada enam buah model yang dipertimbangkan yaitu AR (1,0,0), AR (1,1,0), AR (1,2,0), AR (2,0,0), AR (2,1,0), dan AR (2,2,0). Di dalam analisis statistika terdapat sejumlah ukuran atau kriteria yang biasa digunakan untuk pemilihan model terbaik, dimana salah satu diantaranya adalah menggunakan kriteria informasi. Kriteria informasi yang umum digunakan untuk memilih model terbaik adalah kriteria informasi Akaike (AIC) dan kriteria informasi Bayesian (BIC). Akan tetapi dalam makalah ini akan diperkenalkan suatu kriteria yang baru yaitu kriteria informasi empirik (EIC). Metode ini akan diaplikasikan pada data tentang harga produsen cabe merah di Kabupaten Cianjur Jawa Barat. Hasil yang diperoleh bahwa dengan menggunakan kriteria informasi AIC dan BIC model peramalan terbaik adalah ARIMA (1,2,0) yaitu dengan nilai AIC = 1411,3 dan BIC = 1415,12 sedangkan dengan menggunakan kriteria informasi EIC model peramalan terbaik adalah ARIMA (2,2,0) yaitu dengan nilai EIC = 1384,98. Hasil peramalan yang dilakukan adalah menggunakan model ARIMA (2,2,0) hal ini dikarenakan nilai EIC bekerja lebih baik. Setelah dilakukan peramalan, harga cabe merah di Kabupaten Cianjur mengalami kenaikan tertinggi terjadi pada bulan April tahun 2017 dan terendah pada bulan Mei tahun 2016.

Kata Kunci – AIC, BIC, EIC, *Data Deret Waktu*,

A. Pendahuluan

Deret waktu (*time series*) merupakan sekumpulan hasil observasi yang dapat diatur dan didapat menurut ukuran kronologis, biasanya dalam interval waktu yang sama. Deret waktu dianalisa untuk mendapatkan ukuran yang dapat digunakan dalam membuat keputusan masa kini, untuk peramalan, dan perencanaan operasional dimasa yang akan datang. Untuk dapat dianalisis terlebih dahulu harus diketahui pola dari deret waktu tersebut, kemudian dianalisa menggunakan metode yang sesuai sehingga menghasilkan ramalan yang tepat dimasa yang akan datang. Setelah diketahui pola maka akan dilakukan analisis deret waktu menggunakan metode-metode deret waktu. Metode deret waktu yang biasa digunakan antara lain: Metode Naïf, Metode rata-rata bergerak, Metode pemulusan eksponensial, Model regresi sederhana, dan Model ARIMA. Selanjutnya akan dilakukan evaluasi terhadap model yang telah ditentukan metode deret

waktunya. Untuk mengevaluasi peramalan maka ada dua cara yang akan digunakan yaitu ketepatan metode peramalan melalui nilai kesalahan peramalan atau error dan kriteria informasi. Nilai kesalahan adalah selisih antara nilai yang diramalkan dengan nilai yang sebenarnya terjadi. Ada beberapa ukuran ketepatan peramalan yang sering digunakan adalah *MSE*, *MAD*, *MAPE*, dan *MPE*. Sedangkan untuk kriteria informasi yang biasa digunakan adalah *Akaike's information criterion (AIC)*, dan *Bayesian information criterion (BIC)*. Disini akan dibahas kriteria informasi baru yaitu *empirical information criterion (EIC)*. Adapun tujuan dari makalah ini adalah Memilih model tentatif berdasarkan plot ACF dan PACF, memilih model terbaik menggunakan metode kriteria informasi Akaike's (AIC), kriteria informasi Bayesian (BIC), memilih model terbaik menggunakan metode kriteria informasi empirik (EIC), melakukan peramalan dengan menggunakan metode autoregresif

B. Landasan Teori

Metode Data Deret Waktu

Metode deret waktu merupakan metode yang mempelajari tentang peramalan suatu data. Peramalan deret waktu adalah penggunaan model untuk memprediksi nilai di waktu mendatang berdasar peristiwa yang telah terjadi. Di dunia bisnis, data deret waktu digunakan sebagai bahan acuan pembuatan keputusan sekarang, untuk proyeksi, maupun untuk perencanaan pada masa depan. Ada beberapa asumsi penting yang harus dipenuhi agar data deret waktu dapat digunakan dalam keperluan peramalan (Yanti, 2008). Beberapa diantaranya adalah adanya ketergantungan antara kejadian masa mendatang terhadap masa sebelumnya atau lebih dikenal dengan istilah adanya [autokorelasi](#) antara Z_t dan Z_{t-k} , data yang digunakan adalah data yang stasioner, dan asumsi berikutnya adalah aktivitas pada masa depan mengikuti pola yang terjadi pada masa lalu dan hubungan atau keterkaitan pada masa lalu dapat ditentukan dengan pengamatan atau penelitian

Proses Autoregresif Ordo Pertama

Autoregresif merupakan suatu observasi pada waktu t dinyatakan sebagai fungsi linier terhadap p waktu sebelumnya ditambah dengan sebuah residual acak a_t yang independen dan berdistribusi normal dengan rata-rata 0 dan varian konstan σ_a^2 . Untuk persamaan dari metode autoregressif ordo pertama (Wei, 2006) adalah

$$\hat{Z}_t = \phi_1 \hat{Z}_{t-1} + \alpha_t \quad 1$$

Proses Autoregresif Ordo Kedua

Untuk autoregresif ordo ke 2, mempunyai persamaan

$$\hat{Z}_t = \phi_1 \hat{Z}_{t-1} + \phi_2 \hat{Z}_{t-2} + \alpha_t \quad 2$$

ACF dan PACF

ACF dan PACF merupakan tahapan penting dalam menentukan model dalam metode ARIMA. Model ARIMA berdasarkan ACF dan PACF dapat dilihat dalam tabel berikut (Yanti, 2008)

Tabel 1 Model ARIMA Menurut Plot ACF dan PACF

No	ACF	PACF	Model
1.	Menurun secara eksponensial atau mengikuti gelombang sinus	Mempunyai spike pada lag 1,2,...,p dan turun dengan cepat setelah lag p	Model AR
2.	Mempunyai spike pada lag 1,2,...,p dan turun dengan cepat setelah lag p	Menurun secara eksponensial atau mengikuti gelombang sinus	Model MA
3.	Mempunyai spike pada lag 1,2,...,p dan turun dengan cepat setelah lag Q	Mempunyai spike pada lag 1,2,...,p dan turun dengan cepat setelah lag p	<ul style="list-style-type: none"> • Jika PACF turun lebih cepat daripada ACF maka model AR • Jika ACF turun lebih cepat daripada PACF maka model MA

Metode Pembedaan

Metode pembedaan adalah metode yang digunakan untuk menstasionerkan data yang tidak stasioner menjadi data yang stasioner. Meskipun banyak deret waktu bersifat nonstasioner karena beberapa faktor, namun sifat-sifat yang lainnya masih sama, kecuali perbedaannya pada tingkat rata-rata. Oleh karena itu, dengan membiarkan $\psi(B)$ menjadi pengganti autoregresif yang menggambarkan sifatnya, untuk itu persamaan yang di dapat adalah

$$Z_t = Z_{t-1} + a_t \tag{3}$$

Fungsi Kemungkinan Eksak

Baik fungsi kemungkinan bersyarat dan tanpa syarat adalah aproksimasi. Untuk menggambarkan turunan dari fungsi kemungkinan yang tepat untuk model time series, maka akan dilihat proses AR (1)

$$\dot{Z}_t = \phi \dot{Z}_{t-1} - a_t$$

dimana, \dot{Z}_t berdistribusi $N(0, \sigma_a^2 / (1 - \phi^2))$ dan a_t berdistribusi $N(0, \sigma_a^2)$ maka fungsi kemungkinan log-likelihood yang didapatkan adalah

$$\ln L(\dot{Z}_1, \dots, \dot{Z}_n | \phi, \mu, \sigma_a^2) = -\frac{n}{2} \ln 2\pi + \frac{1}{2} \ln(1 - \phi^2) - \frac{n}{2} \ln \sigma_a^2 - \frac{S(\phi, \mu)}{2\sigma_a^2} \tag{4}$$

Kriteria Pemilihan Model Peramalan Terbaik

Dalam menentuakn suatu model peramalan yang ingin digunakan maka dilakukan evaluasi terhadap model tersebut. Ada dua cara evaluasi model peramalan yang akan dilakukan yaitu menggunakan kriteria standar dan kriteria informasi. Pada bagian ini akan dijelaskan mengenai kedua evaluasi pemilihan model tersebut

Kriteria Standar

Dalam menentuakn suatu model peramalan tepat digunakan untuk meramalkan suatu kejadian adalah melalui nilai kesalahan peramalan (error). Ada beberapa nilai peramalan yaitu:

1. *Mean Square Error (MSE)*

$$MSE = \sum_{i=1}^n e_i^2 / n \tag{5}$$

2. *Mean Absolute Deviation (MAD)*

$$MAD = \sqrt{\sum_{i=1}^n e_i^n / (n-1)} \quad 6$$

3. *Mean Absolute Percentage Error (MAPE)*

$$MAPE = \sum_{i=1}^n \left| \left(\frac{X_i - F_i}{X_i} \right) (100) \right| / n \quad 7$$

4. *Mean Percentage Error (MPE)*

$$MPE = \sum_{i=1}^n \left(\frac{X_i - F_i}{X_i} \right) (100) / n \quad 8$$

Kriteria Informasi Akaike (AIC) dan Kriteria Informasi Bayesian (BIC)

Asumsikan bahwa model statistik parameter M digunakan pada data. Untuk menilai kualitas model yang baik, (Akaike 1973) memperkenalkan kriteria informasi. Kriteria tersebut telah disebut AIC (kriteria informasi Akaike) dan didefinisikan sebagai

$$AIC = -2 \log L(\hat{\theta}) + 2M \quad 9$$

Dimana M adalah jumlah parameter dalam model. Untuk model ARMA dan efektif jumlah observasi, fungsi log-likelihood adalah

$$\ln L = -\frac{n}{2} \ln 2\pi\sigma_\alpha^2 - \frac{1}{2\sigma_\alpha^2} s(\phi, \mu, \theta) \quad 10$$

Dengan memaksimalkan ϕ, μ, θ dan σ_α^2 , maka

Kelemahan dari AIC adalah cenderung melebih-lebihkan urutan autoregresif. (Akaike 1973) telah mengembangkan perpanjangan prosedur AIC Bayesian minimum, yang disebut kriteria informasi Bayesian (BIC), yang mengambil dari

$$BIC = -2 \log L(\hat{\theta}) + \log(n)M \quad 11$$

Kriteria Informasi Empirik

Misalkan kita mempunyai data deret waktu yang sama. Dengan menetapkan j menjadi observasi t th dimana j ($j = 1, \dots, m$ dan $t = 1, \dots, n_j$) maka akan didapatkan nilai $y = [y_1, \dots, y_m]$ dengan memasukkan model awal yang dipilih menggunakan kriteria informasi lainnya. EIC memiliki $f(n, q) = kq$. Persamaan untuk EIC adalah

$$EIC = -2 \log L(\hat{\theta}) + 2kq \quad 12$$

Model dengan EIC terkecil adalah model yang dipilih. Jika q adalah jumlah parameter untuk model maka k adalah pembobot. Setiap seri dalam set dibagi menjadi dua segmen yaitu segmen pertama terdiri dari $n_j^* = n_j - H$ observasi.

Penaksiran Penalti

Penaksiran penalti digunakan untuk mencari $f(n, k)$. Untuk nilai k yang akan digunakan adalah antara

$$-2 \log n < k < 2 \log n \quad 13$$

Langkah-langkah untuk memperkirakan k adalah sebagai berikut:

- 1) Gunakan n_j^* pengamatan untuk memperkirakan parameter masing-masing dari model N dengan estimasi likelihood maksimum
- 2) Catat kemungkinan log-likelihood yang dimaksimalkan
- 3) Untuk setiap uji coba k_1, \dots, k_N , dipilih model untuk setiap deret waktu dengan menggunakan EIC

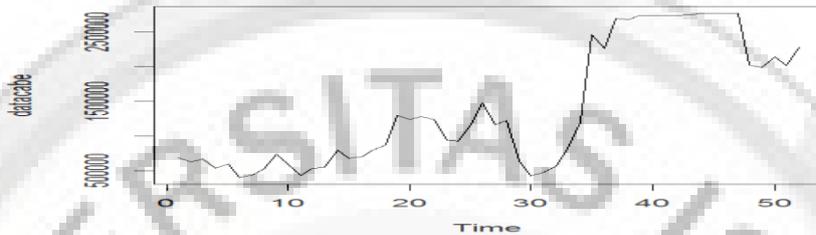
C. Hasil Penelitian

Deskripsi Data

Data yang digunakan adalah data harga produsen pedesaan cabe merah di kabupaten Cianjur dari tahun 2012 sampai 2016. Data diperoleh dengan cara sekunder dari Badan Pusat Statistik Jawa Barat.

Mengidentifikasi Model Runtun Waktu Data Harga Produsen Cabe Merah

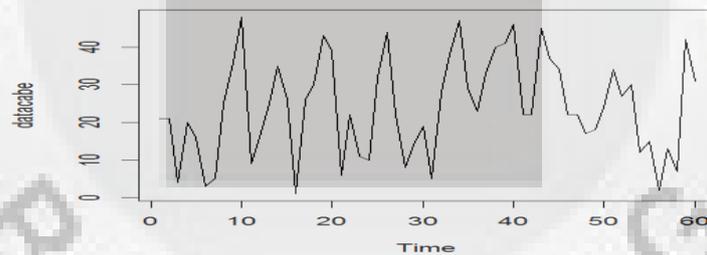
Membuat Plot data asli



Gambar 1 Plot Hasil Data Asli Harga Produsen Cabe Merah Kabupaten Cianjur Jawa Barat Tahun 2012-2016

Dari hasil plot pada data asli menunjukkan bahwa data tidak stasioner atau tidak berada pada rata-rata. Oleh karena itu maka akan dilakukan transformasi agar data tersebut stasioner dan dapat digunakan.

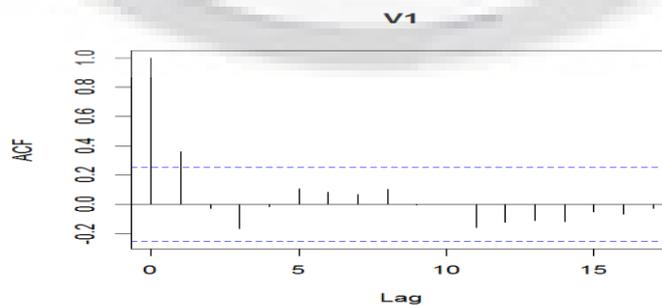
Plot data transformasi



Gambar 2 Plot Hasil Data transformasi Harga Produsen Cabe Merah Kabupaten Cianjur Jawa Barat Tahun 2012-2016

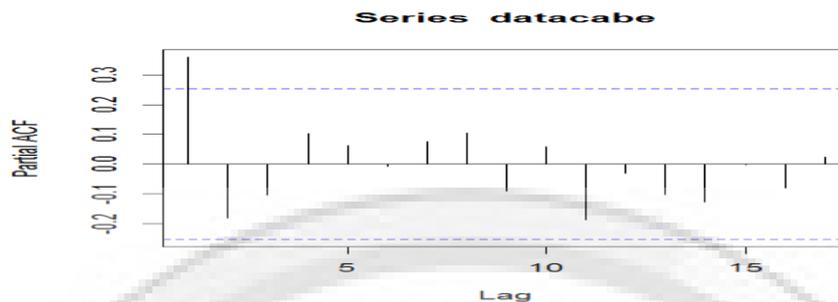
Dari Hasil dapat dilihat bahwa data sudah stasioner untuk itu selanjutnya adalah plot ACF dan PACF

Plot ACF



Gambar 3 acf Harga Produsen Cabe Merah Kabupaten Cianjur Jawa Barat Tahun 2012-2016

Dari hasil plot ACF maka hasilnya adalah model MA dengan lag 2
PACF



Gambar 4 Plot PACF Harga Produsen Cabe Merah Kabupaten Cianjur Jawa Barat Tahun 2012-2016

Dari hasil plot maka dapat dilihat PACF dari model adalah 2

Penaksiran Parameter Pada Model

Setelah satu atau beberapa model sementara yang diidentifikasi maka selanjutnya adalah melakukan penaksiran parameter pada model-model tersebut.

Tabel 2 Hasil Penaksiran Parameter dari Model ARIMA Harga Cabe Kabupaten Cianjur Jawa Barat

	ARIMA (1,0,0)	ARIMA (1,1,0)	ARIMA (1,2,0)	ARIMA (2,0,0)	ARIMA (2,1,0)	ARIMA (2,2,0)
AR1	0,9474	-0,054	-5,741	1,0223	0,0409	-0,686
AR2				-0,0786	0,2029	-0,193
Intercept	1434083,2			1432194,3		
MSE	7,4	7,6	9,72	7,39	7,3	9,33
Log likelihood	-725,79	-711,35	-703,65	-725,63	-710,26	-702,69
AIC	1457,58	1426,7	1411,3	1459,26	1426,52	1411,37
BIC	1463,43	1430,57	1415,122	1467,06	1432,32	1417,11

Menghitung Nilai Kriteria Informasi Empirik

Dalam persamaan kriteria informasi empirik terdapat suatu nilai pembobot atau k dimana nilai tersebut berada di antara $-2 \log n$ hingga $2 \log n$

Penaksiran Penalti

- 1) Gunakan n_j^* pengamatan untuk memperikaran model dimana $n_j^* = n - H$, dimana H adalah banyaknya pengamatan. Pada skripsi ini data yang digunakan adalah sebanyak 60 dan pengamatan yang akan dilakukan adalah 8, untuk itu nilai $n_j^* = 60 - 8 = 52$
- 2) Catat nilai kemungkinan log yang dimaksimalkan. Nilai log setiap tabel akan disajikan dalam tabel berikut.

Tabel 3 Nilai log likelihood dari Data Harga Cabe Kabupaten Cianjur Jawa Barat

Model	ARIMA (1,0,0)	ARIMA (1,1,0)	ARIMA (1,2,0)	ARIMA (2,0,0)	ARIMA (2,1,0)	ARIMA (2,2,0)
log likelihood	-725,79	-711,35	-703,65	-725,63	-710,26	-702,69

Nilai log-likelihood diatas akan digunakan untuk menghitung nilai EIC menggunakan persamaan 2.42

- 3) Penaksiran penalti adalah suatu metode yang yang digunakan untuk mencari nilai k atau pembobot pada kriteria informasi empirik. Langkah-langkah yang digunakan adalah uji nilai k pada setiap model menggunakan EIC. Nilai pembobot atau nilai k yang akan digunakan adalah $-2 \log n$ hingga $2 \log n$. Karena data yang digunakan sebanyak 60 maka nilai k berada pada $-2 \log 60$ hingga $2 \log 60$, untuk itu nilai k adalah -3,4 sampai 3,4 sehingga nilai kriteria informasi empirik yang didapatkan akan disajikan dalam tabel berikut

Pemilihan Model Terbaik

Dari hasil perhitungan di atas maka akan dilakukan pemilihan model terbaik menggunakan kriteria informasi empirik, kriteria informasi Akaike dan kriteria informasi Bayesian

Tabel 4 Perbandingan Nilai AIC, BIC, dan EIC dari Data Harga Cabe Kabupaten Cianjur Jawa Barat

Mode	ARIMA (1,0,0)	ARIMA (1,1,0)	ARIMA (1,2,0)	ARIMA (2,0,0)	ARIMA (2,1,0)	ARIMA (2,2,0)
AIC	1457,58	1426,7	1411,3	1459,26	1426,52	1411,37
BIC	1463,43	1430,57	1415,12	1467,06	1432,32	1417,11
EIC	1431,18	1409,1	1393,7	1424,06	14012	1384,98

Dari tabel diatas dapat dilihat bahwa model yang memiliki model terbaik berdasarkan AIC dan BIC adalah model ARIMA (1,2,0) dengan nilai AIC = 1411,3 dan BIC = 1415,12. Untuk model terbaik berdasarkan EIC adalah model ARIMA (2,2,0). Karena nilai EIC terkecil ada pada model ARIMA (2,2,0) maka model yang digunakan untuk peramalan selanjutnya adalah menggunakan model ARIMA (2,2,0).

Peramalan Menggunakan Model ARIMA (2,2,0)

Setelah melakukan pemilihan model terbaik maka selanjutnya akan dilakukan peramalan. Dengan menggunakan model ARIMA (2,2,0) maka akan dilakukan peramalan selama 12 bulan kedepan. Hasil dari peramalan akan disajikan dalam tabel berikut

Tabel 5 Nilai Prediksi dari Data Harga Cabe Kabupaten Cianjur Jawa Barat

Bulan	Z prediksi	Bulan	Z prediksi
Mei	2.324.734,00	November	2.993.724,00
Juni	2.445.649,00	Desember	3.103.299,00
Juli	2.558.793,00	Januari	3.212.679,00
Agustus	2.663.531,00	Februari	332.188,00
September	2.775.536,00	Maret	3.431.646,00
Oktober	2.884.177,00	April	3.541.114,00

Dari hasil peramalan dapat dilihat bahwa setiap tahunnya harga cabe merah di Kabupaten Cianjur mengalami kenaikan

D. Kesimpulan

1. Dari hasil plot pada data asli diketahui data belum stasioner sehingga perlu dilakukannya transformasi untuk mengstasionerkan data. Setelah dilakukan plot kembali dari hasil transformasi maka dapat dilihat bahwa data sudah stasioner.
2. Setelah memenuhi syarat yaitu data sudah stasioner maka selanjutnya akan di plot ACF dan PACF untuk menentukan model yang digunakan adalah model AR, MA atau ARIMA. Dari plot ACF dan PACF pada data diketahui bahwa plot PACF turun lebih cepat pada lag ke 2 sehingga model yang digunakan adalah model sementara yang digunakan adalah model ARIMA (1,0,0), ARIMA (1,1,0), ARIMA (1,2,0), ARIMA (2,0,0), ARIMA (2,1,0), dan ARIMA (2,2,0).
3. Setelah menentukan model maka akan dilihat model yang paling baik untuk digunakan. Dengan menghitung nilai kedua metode kriteria informasi dan nilai log-likelihood yang telah dijelaskan maka dapat ditentukan model yang paling baik digunakan adalah model ARIMA (1,2,0) nilai AIC = 1411,3 dan BIC = 1415,12. Setelah melakukan perhitungan nilai EIC maka dapat dilihat model yang paling baik digunakan adalah model ARIMA (2,2,0) yaitu dengan nilai EIC = 1384,98
4. Dari hasil peramalan menggunakan model ARIMA (2,2,0) dapat dilihat bahwa pada setiap bulannya harga cabe merah mengalami peningkatan. Harga cabe merah tertinggi terjadi pada bulan April 2017 dan terendah pada bulan Mei 2016.

Daftar Pustaka

- Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*. In: B.N. Petrow and Csaki, F. (Eds) *Second Internasional Symposium on Information Theory* (Budapest: Akademiai Kiado)
- https://jabar.bps.go.id/new/website/pdf_publicasi/Statistik-Harga-Produsen-Pedesaan-Jawa-Barat-2014.pdf. diakses 3 Agustus 2017
- https://jabar.bps.go.id/new/website/pdf_publicasi/Statistik-Harga-Produsen-Perdesaan-Jawa-Barat-2016.pdf . diakses 3 Agustus 2017
- Wei William, W. S. (2006). *Time Series Analysis Univariate and Multivariate Methods*. Edisi Kedua. United States of America: Pearson Education.
- Yanti Teti, S.(2008). *Analisis Deret Waktu*. Bandung: Program studi Statistika Universitas Islam Bandung.