

PEMODELAN DAN *FORECASTING* KEBUTUHAN AIR BERSIH DI PROPINSI DIY MENGGUNAKAN *AUTOREGRESSIVE INTEGRATED MOVING AVERAGE* (ARIMA)

Padrul Jana^{1*}, Nendra MS Dwipa²

¹Universitas PGRI Yogyakarta

²Universitas PGRI Yogyakarta

*padrul.jana@upy.ac.id

Abstract

The need of clean water will be something that is expensive in the future. Because of that reason, the focus of this research is the forecasting of clean water need in DIY. The purpose of this research is to find the best model and suitable with the data that will be used in the forecasting. The data of forecasting can be used by the people as the alternative solution to anticipate the possibility and the based reason to make regulation. The model that would be employed is ARIMA with the step of preprocessing, model identification, model estimation and diagnostic check and forecasting. The best model after data analyzing is ARIMA (1,1,0) with the Sum squared resid of 0,330933, Akaike info criterion of 1.050808, Schwarz info criterion of -0.951235, that best model will be employed to conduct forecasting some steps ahead. The result of clean water need forecasting in the next 5 years is 2016: 138.840,9, 2017: 138.000,8, 2018:138.400,6, 2019: 138.209,9 and 2020: 138.300,8 in thousand meter cubic.

Keywords : ARIMA, Forecasting, Clean Water Need, SSR, AIC.

A. PENDAHULUAN

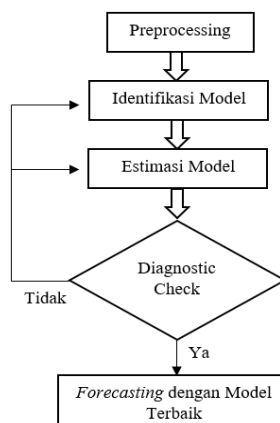
Kondisi terkini, manusia sangat dimanjakan oleh teknologi dengan kemajuan teknologi segala aktifitas dan kegiatan manusia bertumpu padanya. Masa yang akan datang kecerdasan buatan manusia yang bernama teknologi akan mampu menggantikan fungsi manusia. Efek kemajuan tersebut dunia akan menjadi bank data dan sekarang adalah baru permulaan dari masa periode data. Generasi yang menang adalah generasi yang menguasai data dan mampu menggunakan data. Dampak dari semua itu teknologi akan semakin murah, sedangkan barang yang menjadi mahal adalah kebutuhan manusia salah satunya air. Sumber air menjadi kotor oleh perbuatan manusia sendiri. Sungai telah tercemar, danau yang tidak lagi mengalirkan air bersih, semua itu dikarenakan pabrik – pabrik yang sembarangan membuang limbah – limbah kotor tersebut. Dulu berita mengenai air kotor di suatu negara adalah hal yang jarang, namun pada tahun 2013 sebuah berita mengenai krisis air bersih sudah menjadi topik keseharian. Dapat dibayangkan pada tahun 2023?.

Berkaca pada kondisi tersebut, sangat penting untuk mengetahui kondisi kebutuhan air bersih diwaktu yang akan datang. *Forecasting* kebutuhan air bersih menjadi mutlak perlu, untuk meramalkan kondisi ke depan. Dengan begitu, pemerintah atau pihak terkait dapat mempersiapkan lebih dini.

Salah satu metode untuk *forecasting* data adalah menggunakan ARIMA, karena ARIMA merupakan teknik mencari pola yang paling cocok dari data masa lalu dan digunakan untuk *forecasting* jangka pendek. Model ARIMA banyak digunakan dalam *forecasting* data seperti *forecasting* mengenai popularitas pencarian judul artikel dengan ARIMA berbasis *Backpropagation Neural Networks* (BPNNs) (Omar, Hoang, & Liu, 2016). *Forecasting* untuk tingkat polutan udara sekitar (O₃, NO, NO₂ dan CO), hasilnya model ARIMA cukup memuaskan dan menjadi prosedur *forecasting* yang disarankan untuk dapat digunakan secara efektif yang bertujuan memberi peringatan kualitas udara jangka pendek (Kumar & Jain, 2010). ARIMA juga akurat dalam *forecasting* kunjungan pasien rawat inap di rumah sakit yang bermanfaat untuk pengalokasian sumber daya kesehatan untuk memenuhi tuntutan medis (Luo, Luo, Zhang, & He, 2017). Penerapan lain dari metode ini untuk memprediksi tren kejadian *Hand, foot and mouth disease* (HFMD) dan memberikan manfaat untuk pencegahan penyakit *Hand, foot and mouth disease* serta kontrol atas penyakit tersebut (LIU, LUAN, YIN, ZHU, & LÜ, 2016).

B. METODE PENELITIAN

Metode deret waktu yang paling umum meliputi naïve, *moving average* (MA), *exponensial smoothing* (ES), dan *autoregressive integrated average moving* (ARIMA) (S. J. Wang, Huang, Wang, & Chen, 2010). Metode-metode yang tersedia harus sesuai dengan karakteristik data yang ada sehingga akurasi dari hasil *forecasting* semakin baik. *Forecasting* dapat dikelompokkan menjadi beberapa tipe, yaitu ekonometrika, deret berkala (*time series*), dan *forecasting* kualitatif (Jana, 2016). Proses model Box-Jenkins untuk menentukan model ARIMA didasarkan pada tiga tahap: (1) identifikasi model, (2) estimasi parameter dan (3) *Diagnostic check* (Kavousi-Fard & Kavousi-Fard, 2013), lebih detail di jelaskan pada gambar berikut ini:



Gambar 1. Diagram alir metodologi pemodelan Box-Jenkins

Preprocessing

Meliputi *ploting* digunakan untuk melihat data secara visual seperti adanya trend, komponen musiman, ketidastasioneran dalam variansi dan lain-lain. Teknik yang umum digunakan dalam *preprocessing* adalah membuang *outlier*, *filtering*, transformasi data, *differencing* dan lain-lain.

Identifikasi Model

Pada tahap kedua yaitu identifikasi model, adalah menentukan kemungkinan struktur persistensi dalam kumpulan data sampel, ACF dan PACF yang dapat digunakan (W. chuan Wang, Chau, Xu, & Chen, 2015). Identifikasi model melalui *ploting* sampel ACF/PACF dengan mempertimbangkan sifat-sifat fungsi ACF/PACF secara teoritis dari model ARMA (Rosadi, 2012). Pada tahap ini bertujuan untuk menemukan struktur model *autoregressive integrated average moving* (ARIMA) yang tepat yang akan menjadi model efisien serta membandingkan perkiraan untuk mendapatkan kesalahan terendah (Grigonytė, 2016).

Tabel 1. Rangkuman sifat-sifat ACF/PACF dari model ARMA

Proses	Sampel ACF	Sampel PACF
<i>White Noise</i>	Tidak ada yang melewati batas interval pada lag > 0	Tidak ada yang melewati batas interval pada lag > 0
AR(p)	Meluruh menuju nol secara eksponensial	Di atas interval maksimum sampai lag ke p dan di bawah batas pada lag > p
MA (q)	Di atas interval maksimum sampai lag ke p dan di bawah batas pada lag > q	Meluruh menuju nol secara eksponensial
ARMA(p,q)	Meluruh menuju nol secara eksponensial	Meluruh menuju nol secara eksponensial

Estimasi Model

Estimasi parameter dalam model pada koefisien dari model ARMA serta nilai variansi dari residual. Estimasi model ARMA dapat menggunakan *Maximum Likelihood Estimation* (MLE), *Least Square*, Hannan Rissanen dan lain-lain.

Diagnostic Check

Diagnostic check umumnya menggunakan uji *t* pada parameter hasil estimasi dan dilanjutkan dengan uji Q-Ljung-Box dan plot ACF/PACF untuk melihat korelasi residual. Jika uji asumsi tersebut terpenuhi maka model tersebut dapat dipertimbangkan menjadi model terbaik dan digunakan untuk *forecasting*.

Forecasting

Model terbaik yang diperoleh, digunakan untuk *forecasting* sifat-sifat data di waktu yang akan datang. *Time series* seringkali membagi data menjadi dua jenis yaitu data *in-sample* adalah data yang dipakai memilih model terbaik, jenis satunya adalah data *out-sample* data yang dipakai untuk validasi keakuratan dari model terbaik. Ukuran yang biasa digunakan untuk mengukur kesesuaian model adalah *Mean Square Error* (MSE), *Root of MSE*, *Mean Absolute Deviation* (MAD). MSE, RMSE dan MAD untuk data *in-sample* dirumuskan sebagai berikut:

$$MSE = \frac{\sum_{i=1}^m (X_i - \hat{X}_i)^2}{m}, m < n, RMSE = \sqrt{\frac{\sum_{i=1}^m (X_i - \hat{X}_i)^2}{m}}, m < n$$

$$MAD = \frac{\sum_{i=1}^m |X_i - \hat{X}_i|}{m}, m < n$$

(Rosadi, 2012)

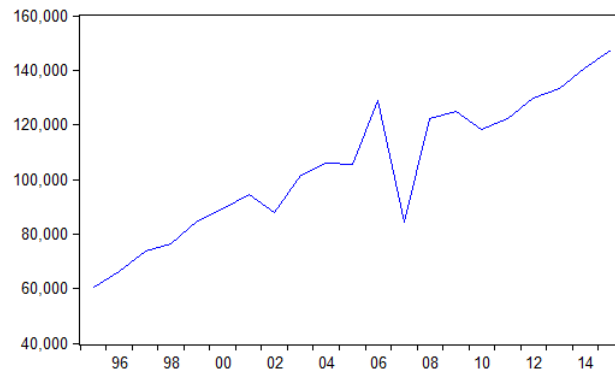
Sumber Data

Sumber data dalam artikel ini adalah data kebutuhan air bersih di propinsi DIY dalam kurun waktu tahun 1995-2015. Data diambil dari Badan Pusat Statistik (BPS), data *time series* tersebut merupakan kebutuhan air bersih dalam ribu meter kubik pada tiap tahunnya.

C. HASIL DAN PEMBAHASAN

Ploting Data

Langkah awal untuk mengetahui karakteristik data mengenai adanya trend atau tidak adalah dengan *ploting data*. *Ploting* berfungsi juga mengecek perkiraan data secara kasar dari bentuk model yang sesuai.

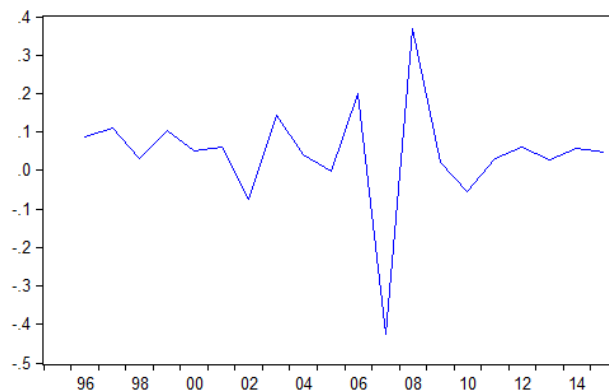


Gambar 2. *Ploting* data kebutuhan air propinsi DIY tahun 1995-2015

Dari *ploting data*, terlihat adanya trend sehingga data belum stasioner secara mean dan variansi. Fakta tersebut dikonfirmasi dengan uji *Augmented Dickey-Fuller* yaitu nilai *Augmented Dickey-Fuller test statistic* sebesar -3.350193 tidak lebih negatif dari *Test critical values* 5% sebesar -3.733200. Karena data masih mengandung trend, maka perlu melakukan transformasi awal agar data lebih stasioner baik secara mean maupun variansi.

Transformasi Data dan Plot ACF/PACF

Berdasarkan hasil *ploting data* kebutuhan air masih mengandung trend sehingga perlu transformasi agar data lebih stasioner secara mean. *Differencing* adalah salah satu cara untuk membuang trend, sedangkan dengan transformasi log (basis e) agar data stasioner dalam variansi. Hasil *Differencing* dan transformasi log terlihat pada gambar di bawah ini.

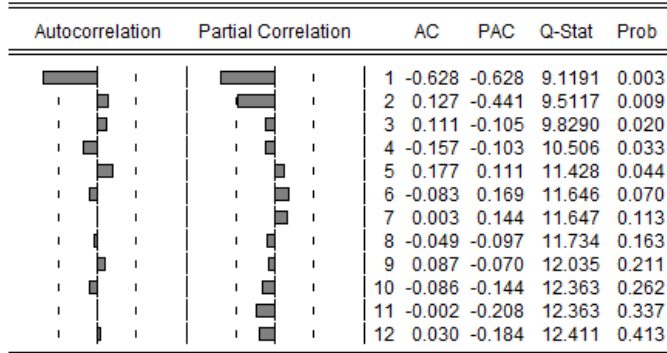


Gambar 3. *Ploting data* setelah *Differencing* ordo 1 dan transformasi log

Data relatif lebih stasioner baik dalam mean maupun variansi, setelah dilakukan *differencing* ordo pertama dan transformasi log. Hal ini diperkuat dengan uji *Augmented Dickey-Fuller* yaitu nilai *Augmented Dickey-Fuller test statistic* sebesar -6.299257 lebih negatif dari *Test critical values* 5% sebesar -3.690814.

Selanjutnya setelah data relatif lebih stasioner dilakukan identifikasi model *autoregressive integrated moving average*. Kandidat model yang sesuai dengan melihat hasil plot ACF dan PACF untuk data yang sudah stasioner, di bawah ini merupakan hasil plot ACF dan PACF.

Sample: 1995 2015
Included observations: 20

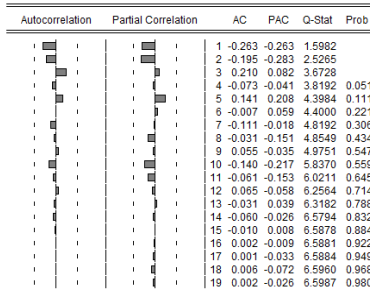


Gambar 4. Plot ACF/PACF dari $d(\log(\text{air}))$

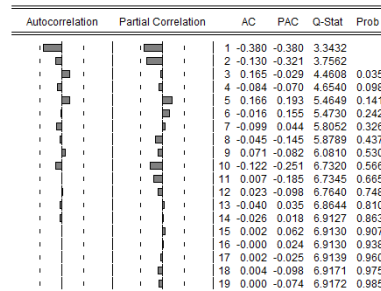
Dari gambar di atas terlihat ACF *cut off* pada lag 1 dan PACF *cut off* di lag 2. Nilai ACF menunjukkan parameter untuk *Moving Average* (MA) dan nilai PACF menunjukkan parameter untuk *Autoregressive* (AR) dan banyaknya berapa kali *differencing* menunjukkan *Integrated*. Data di atas di *differencing* sekali, sehingga model data sementara adalah ARIMA (2,1,1). Model yang relatif baik untuk memodelkan data dengan memperhatikan prinsip *parsimony* (kesederhanaan) dari pemodelan yang memiliki parameter yang sedikit.

Estimasi dan Diagnostic Checking Model

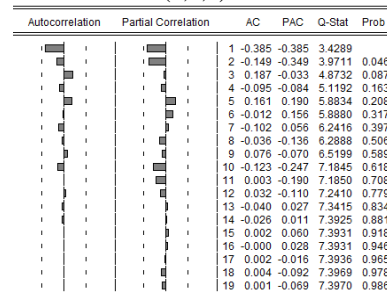
Model sementara dari data adalah ARIMA (2,1,1), sebagai model alternatif kombinasi model ARIMA adalah: ARIMA (1,1,0), ARIMA (0,1,1), ARIMA (1,1,1). Setelah menemukan calon model terbaik, model-model tersebut harus diuji menggunakan kriteria uji *t* untuk parameter hasil estimasi dan *diagnostic checking* dengan uji Q-Ljung-Box terakhir plot ACF/PACF untuk mendeteksi residu hasil estimasi model yang diperoleh.



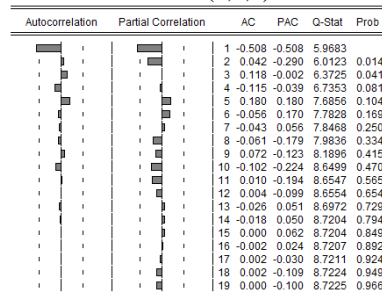
Gambar 5. Plot ACF/PACF Uji Q-LjungBox ARIMA (2,1,1)



Gambar 6. Plot ACF/PACF Uji Q-LjungBox ARIMA (1,1,1)



Gambar 7. Plot ACF/PACF Uji Q-LjungBox ARIMA (1,1,0)



Gambar 8. Plot ACF/PACF Uji Q-LjungBox ARIMA (0,1,1)

Berikut adalah rangkuman estimasi serta nilai-nilai statistik *diagnostic checking* terhadap model-model yang dimiliki.

Tabel 2. Rangkuman pemodelan dengan ARIMA

	ARIMA (2,1,1)	ARIMA (1,1,1)	ARIMA (1,1,0)	ARIMA (0,1,1)
a_1	0.416569 (0.1086)	-0.412671(0.3866)	-0.476629(0.0041)	-
a_2	0.583431 (0.3313)	-	-	-
b_1	-0.999998 (0.0000)	-0.089742(0.8934)	-	0.369505(0.1878)
SSR	0.254167	0.329369	0.330933	0.353799

	ARIMA (2,1,1)	ARIMA (1,1,1)	ARIMA(1,1,0)	ARIMA(0,1,1)
AIC	-1.007742	-0.955058	-1.050808	-0.989541
SBC/BIC	-0.808595	-0.805698	-0.951235	-0.889968
Jumlah Parameter	3	2	1	1
Q(12)	6.2564 (0.714)	6.7640 (0.748)	7.2410 (0.779)	8.6554(0.654)

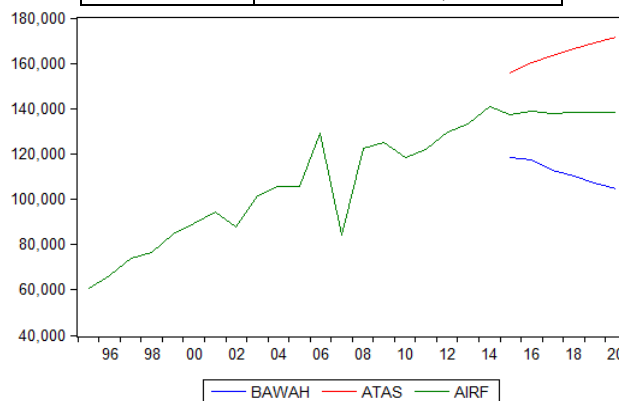
Analisis dari rangkuman pemodelan ARIMA di atas adalah pada model ARIMA (2,1,1), ARIMA (1,1,1), ARIMA (0,1,1) terlihat dari hasil uji t koefisien dari model tidak signifikan, meskipun hasil uji residual menunjukkan sudah tidak terdapat korelasi serial dalam data. Model ARIMA (1,1,0) hasil uji t menunjukkan signifikan juga hasil uji residual menunjukkan sudah tidak terdapat korelasi serial dalam data, model ARIMA (1,1,0) atau AR(1) dijadikan model untuk meramal data untuk beberapa tahun kedepan.

Forecasting dengan Model Terbaik

Forecasting dilakukan untuk data menggunakan data *in-sampel* terbaik, yaitu model ARIMA (1,1,0) atau AR(1) untuk beberapa tahun ke depan. Berikut hasil *forecasting* nya:

Tabel 3. Hasil *forecasting* pemodelan dengan ARIMA

Tahun	Ramalan
2016	138.840,9
2017	138.000,8
2018	138.400,6
2019	138.209,9
2020	138.300,8



Gambar 9. Plot data dan hasil *fitting* dengan ARIMA (1,1,0)

D. KESIMPULAN

Hasil serangkaian analisis dan pemilihan calon model yang akan digunakan untuk *forecasting* data kebutuhan air bersih di Propinsi DIY, didapatkan model terbaik adalah ARIMA (1,1,0) dengan *Sum squared resid* sebesar 0.330933, *Akaike info criterion* sebesar -1.050808, *Schwarz info criterion* sebesar -0.951235. Hasil *forecasting* kebutuhan air bersih di propinsi DIY untuk lima tahun berikutnya adalah pada tahun 2016: 138.840,9, 2017: 138.000,8, 2018: 138.400,6, 2019: 138.209,9 dan 2020: 138.300,8 dalam ribu meter kubik.

E. UCAPAN TERIMAKASIH

Penulis mengucapkan terimakasih kepada Universitas PGRI Yogyakarta yang telah memberi kemudahan, fasilitas penunjang, insentif dana bantuan penelitian sehingga luaran berupa artikel ilmiah ini dapat selesai. Lembaga Penelitian dan Pengabdian UPY yang membantu penyusunan laporan penelitian dan penyusunan luaran penelitian serta Badan Pusat Statistik yang menyediakan data sehingga peneliti dan mengunduh data sekunder dengan mudah.

F. DAFTAR PUSTAKA

- Grigonytė, E. (2016). Short-term wind speed forecasting using ARIMA model, (1), 45–55.
- Jana, P. (2016). Aplikasi Triple Exponential Smoothing untuk Forecasting Jumlah Penduduk Miskin. *Derivat*, 3(2), 76–81. Retrieved from <http://upy.ac.id/ojs/index.php/mtk/issue/view/81/showToc>
- Kavousi-Fard, A., & Kavousi-Fard, F. (2013). A new hybrid correction method for short-term load forecasting based on ARIMA, SVR and CSA. *Journal of Experimental and Theoretical Artificial Intelligence*, 25(4), 559–574. <https://doi.org/10.1080/0952813X.2013.782351>
- Kumar, U., & Jain, V. K. (2010). ARIMA forecasting of ambient air pollutants (O3, NO, NO2 and CO). *Stochastic Environmental Research and Risk Assessment*, 24(5), 751–760. <https://doi.org/10.1007/s00477-009-0361-8>

- LIU, L., LUAN, R. S., YIN, F., ZHU, X. P., & LÜ, Q. (2016). Predicting the incidence of hand, foot and mouth disease in Sichuan province, China using the ARIMA model. *Epidemiology and Infection*, *144*(1), 144–151. <https://doi.org/10.1017/S0950268815001144>
- Luo, L., Luo, L., Zhang, X., & He, X. (2017). Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. *BMC Health Services Research*, *17*(1), 1–13. <https://doi.org/10.1186/s12913-017-2407-9>
- Omar, H., Hoang, V. H., & Liu, D. R. (2016). A Hybrid Neural Network Model for Sales Forecasting Based on ARIMA and Search Popularity of Article Titles. *Computational Intelligence and Neuroscience*, *2016*. <https://doi.org/10.1155/2016/9656453>
- Rosadi, D. (2012). *Ekonometrika & Analisis Runtun Waktu Terapan (I)*. Yogyakarta: Andi Offset.
- Wang, S. J., Huang, C. T., Wang, W. L., & Chen, Y. H. (2010). Incorporating ARIMA forecasting and service-level based replenishment in RFID-enabled supply chain. *International Journal of Production Research*, *48*(9), 2655–2677. <https://doi.org/10.1080/00207540903564983>
- Wang, W. chuan, Chau, K. wing, Xu, D. mei, & Chen, X. Y. (2015). Improving Forecasting Accuracy of Annual Runoff Time Series Using ARIMA Based on EEMD Decomposition. *Water Resources Management*, *29*(8), 2655–2675. <https://doi.org/10.1007/s11269-015-0962-6>