

PERAMALAN PENYEBARAN JUMLAH KASUS VIRUS EBOLA DI GUINEA DENGAN METODE ARIMA

Novita Eka Chandra¹ dan Sarinem²

¹ Universitas Islam Darul 'Ulum Lamongan, novitaekachandra@gmail.com

² Universitas Islam Darul 'Ulum Lamongan, ekosari0931@gmail.com

Abstract. Ebola virus can cause death. The spread number of cases of this virus is very rapidly, especially in the Guinea of West Africa. Based on the past data, the spread number of cases of ebola virus can be predicted by the method of time series namely ARIMA method. In this study the researcher used 63 cases of ebola virus. By using ARIMA method, it was found that an appropriate model for the spread of ebola virus cases is ARIMA(0,2,3). Based on the model, the spread number of cases of ebola virus can be predicted for the next 13 periods, with the result that the spread number of cases of ebola virus has decreased from period to period.

Keywords: *ebola virus, forecasting, time series, ARIMA method.*

Abstrak. Virus ebola dapat menyebabkan kematian. Penyebaran jumlah kasus virus ebola ini sangatlah pesat terutama di Guinea Afrika Barat. Berdasarkan data masa lalu, penyebaran virus ebola dapat diramalkan dengan metode time series, yaitu metode ARIMA. Pada penelitian ini, menggunakan 63 periode data kasus virus ebola. Dengan menggunakan metode ARIMA, diperoleh bahwa model yang sesuai untuk penyebaran jumlah kasus virus ebola ini adalah ARIMA(0,2,3). Berdasarkan model tersebut, penyebaran jumlah kasus virus ebola dapat diprediksi selama 13 periode kedepan, dengan hasil bahwa jumlah penyebaran jumlah kasus virus ebola mengalami penurunan dari periode ke periode.

Kata Kunci: *virus ebola, peramalan, time series, metode ARIMA.*

1 Pendahuluan

Peramalan adalah memperkirakan besarnya atau jumlah sesuatu pada waktu yang akan datang berdasarkan data pada masa lampau yang dianalisis secara ilmiah khususnya menggunakan metode statistika [6]. Peramalan dapat dilakukan dengan model deret waktu. Salah satu model deret waktu yang digunakan dalam penelitian ini adalah model ARIMA. *Emerging Disease* (EID) adalah merupakan penyakit yang dapat meningkat dengan pesat dan dapat menyebabkan kematian. Penyakit ini dapat disebabkan oleh virus, sebagai contoh virus ebola. Virus ebola pertama kali muncul pada tahun 1976 di Sudan dan Republik Congo [5]. *Center of Disease Control and Prevention* (CDC), sebuah Departemen Kesehatan dan Layanan Masyarakat Amerika Serikat menyebutkan bahwa pada tahun 2014 merupakan wabah ebola terbesar sepanjang sejarah yang muncul di Afrika Barat dimulai dari negara Guinea dan menjadi perhatian dunia [3]. Menurut data WHO (WHO: *Ebola Situation Report*) hingga 18 November 2015 telah terjadi 28.598 kasus dengan 11.299 meninggal [8]. Sampai saat ini masih belum ada vaksin atau pengobatan khusus

untuk kasus ebola ini. Dimungkinkan dapat terjadi penyebaran virus hingga ke Asia. Berdasarkan uraian di atas, penulis memprediksi penyebaran jumlah kasus virus ebola di Guinea dengan metode ARIMA.

2 Kajian Teori

2.1 Model Time Series

Analisis deret waktu (*time series*) diperkenalkan pada tahun 1970 oleh George E.P. Box dan Gwilym M. Jenkins melalui bukunya yang berjudul *Time Series Analysis: Forecasting and Control* [1]. Menurut [2], *time series* adalah suatu rangkaian pengamatan x_t , yang dilakukan pada waktu t . Bagian penting dalam analisis *time series* adalah pemilihan model probabilistik yang sesuai dengan data. Dalam memilih model *time series* perlu pula diperhatikan pola data. Pola data dibagi menjadi empat, yaitu pola horisontal, pola musiman, pola trend dan pola siklis.

Dalam penentuan model *time series*, diasumsikan bahwa data dalam keadaan stasioner. Sebuah *time series* $\{X_t, t = 0, \pm 1, \dots\}$ dikatakan stasioner, jika memiliki sifat statistik yang sama dengan deret waktu selanjutnya $\{X_{t+h}, t = 0, \pm 1, \dots\}$ untuk setiap bilangan bulat h . Stasioneritas dibagi menjadi dua, yaitu stasioner dalam mean dan stasioner dalam varian. Jika kondisi stasioner dalam rata-rata tidak terpenuhi, maka perlu dilakukan proses *differencing*. Selanjutnya, dilakukan beberapa proses pemeriksaan diagnostik, sebagai berikut:

1. Fungsi Autokorelasi (ACF)

Koefisien autokorelasi adalah suatu fungsi yang menunjukkan besarnya korelasi (hubungan linier) antara pengamatan pada waktu ke- t (dinotasikan dengan X_t) dengan pengamatan pada waktu-waktu sebelumnya (dinotasikan dengan $X_{t-1}, X_{t-2}, \dots, X_{t-k}$). Koefisien autokorelasi untuk *lag* k dari data runtun waktu dinyatakan sebagai berikut [4]:

$$r_k = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X}_t)(X_{t+k} - \bar{X}_t)}{\sum_{t=1}^n (X_t - \bar{X}_t)^2}.$$

Karena r_k merupakan fungsi atas k , maka hubungan koefisien autokorelasi dengan *lag*-nya disebut dengan fungsi autokorelasi dan dinyatakan dengan ρ_k .

2. Fungsi Autokorelasi Parsial (PACF)

Fungsi autokorelasi parsial (PACF) pada *lag* k adalah korelasi diantara X_t dan X_{t+k} setelah dependensi linier antara X_t dan X_{t+k} variabel antara $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$. Nilai ϕ_{kk} dapat diselesaikan dengan metode yang lebih efisien yang diperkenalkan oleh Durbin [4], persamaan yang dapat digunakan adalah

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1} \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1} \rho_{k-j}}.$$

Karena ϕ_{kk} merupakan fungsi atas k , maka ϕ_{kk} disebut fungsi autokorelasi parsial.

3. *White Noise*

Suatu proses $\{a_t\}$ disebut proses white noise (proses yang bebas dan identik) jika bentuk peubah acak yang berurutan tidak saling berkorelasi dan mengikuti distribusi tertentu. Untuk mengetahui apakah residu memenuhi proses white noise atau tidak, perlu dilakukan uji, salah satunya dengan Uji LjungBox.

4. Uji Asumsi Kenormalan

Selain memenuhi white noise, model juga harus memenuhi asumsi kenormalan [1]. Salah satu cara yang dapat ditempuh untuk melakukan uji kenormalan adalah uji Kolmogorov Smirnov.

2.2 Metode ARIMA

Metode ARIMA adalah singkatan dari *Autoregressive Integrated Moving Average*. Metode ARIMA atau biasa disebut juga sebagai metode Box-Jenkins merupakan metode yang secara intensif dikembangkan oleh George Box dan Gwilym Jenkins pada tahun 1960an [4]. Secara umum, model ARIMA mempunyai orde tertentu yang menyatakan adanya komponen *Autoregressive* (AR) orde p maupun *Moving Average* (MA) orde q , sehingga model ARIMA biasa dituliskan dalam bentuk ARIMA (p, d, q) . Orde d menyatakan orde *differencing* untuk membuat data menjadi stasioner [1]. Penentuan orde tersebut tergantung pada kondisi data. Pada penetapan model yang sesuai harus melalui beberapa tahap yaitu identifikasi model, pemeriksaan diagnostik, pemilihan model terbaik dan penerapan model ARIMA yang diperoleh untuk keperluan peramalan. Pemilihan model ARIMA mengikuti ketentuan berikut:

Model	ACF	PACF
AR	<i>dies down</i>	<i>cut off</i>
MA	<i>cut off</i>	<i>dies down</i>
ARMA	<i>diesdown</i>	<i>dies down</i>

2.3 Pemilihan Model Terbaik

Jika model yang diperoleh lebih dari satu, maka dilakukan pemilihan model terbaik dengan kriteria pemilihan menggunakan nilai MSE (*Mean Square Error*) yang minimum. Kriteria MSE dirumuskan sebagai berikut:

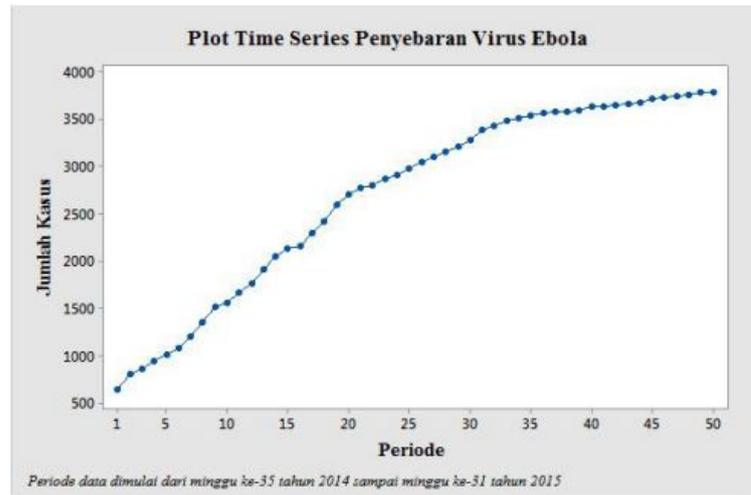
$$MSE = \frac{1}{n} \sum_{t=1}^n (X_t - \hat{X}_t)^2.$$

3 Hasil dan Pembahasan

3.1 Identifikasi Model

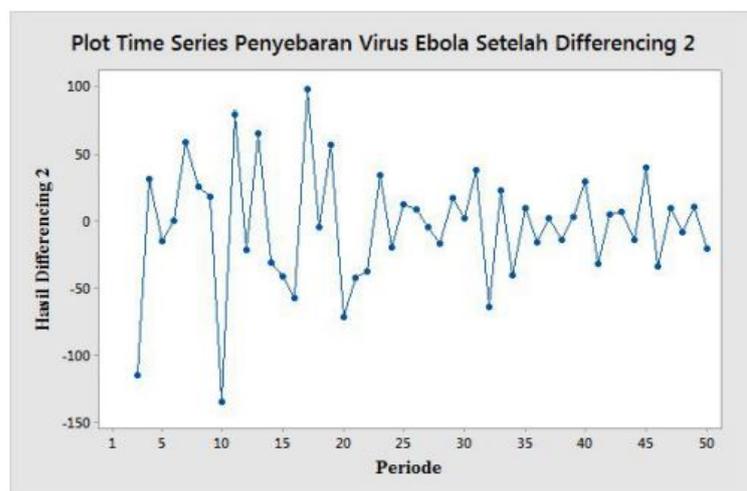
Data yang digunakan dalam penelitian ini merupakan data penyebaran virus ebola selama 63 periode [7]. Untuk membangun model yang baik, data *time series* dibagi menjadi dua bagian terlebih dahulu yaitu *inisialization set* dan *test set*. Prediksi akan dilakukan sejauh 13 periode kedepan, dengan demikian

test set terdiri dari 13 periode, sedangkan *inisialization set* terdiri dari 50 periode. Dengan bantuan software MINITAB 16, berikut ini plot *time series* data penyebaran virus ebola untuk 50 periode pertama.



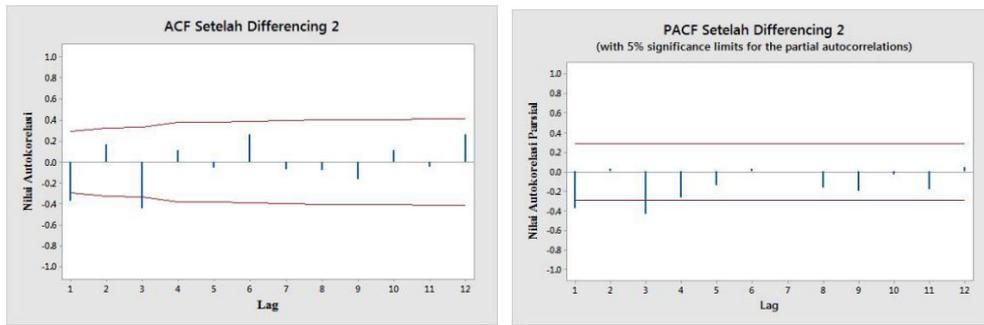
Gambar 1: Plot *Time Series* Penyebaran Virus Ebola

Dari Gambar 1 terlihat bahwa data penyebarannya dipengaruhi oleh pola trend (tidak stasioner dalam mean). Untuk menstasionerkan data dilakukan proses *differencing*. Data stasioner setelah dilakukan proses *differencing* sampai orde ke-2. Berikut plot hasil dari *differencing* pada *lag* 2.



Gambar 2: Plot *Time Series* Setelah *Differencing* pada *lag* 2

Terlihat pada Gambar 2 data sudah stasioner dalam mean, sehingga model ARIMA yang didapat akan mengandung *differencing* pada *lag* ke-2 (orde $d = 2$). Selanjutnya menentukan model sementara berdasarkan grafik ACF dan PACF berikut.



Gambar 3: ACF (kiri) dan PACF (kanan) *Time Series* Setelah *Differencing* Ke-2

Dari Gambar 3 menunjukkan bahwa ACF dan PACF *cut off* setelah lag ke-3. Dengan demikian, model yang diduga berdasarkan grafik ACF adalah model ARIMA (0,2,3) dan berdasarkan grafik PACF adalah model ARIMA (3,2,0).

3.2 Pemeriksaan Diagnostik

Berdasarkan kedua model yang diperoleh yaitu ARIMA (0,2,3) dan ARIMA (3,2,0) dilakukan pemeriksaan uji diagnostik yang meliputi uji signifikansi parameter, *white noise*, dan uji kenormalan. Berikut penjelasannya.

1. Uji signifikansi parameter

Pada Gambar 4 berikut menunjukkan bahwa untuk AR(3) pada model ARIMA (3,2,0) nilai t-hitung=-4,10 lebih kecil dari t-tabel= 2,01289, artinya parameter tidak signifikan digunakan. Akan tetapi, untuk MA(3) pada model ARIMA (0,2,3) nilai t-hitung=2,94 yang diberikan lebih besar dibandingkan t-tabel=2,01290, artinya parameternya signifikan.

Final Estimates of Parameters					
Type	Coef	SE Coef	T	P	
AR 1	-0.4126	0.1292	-3.19	0.003	
AR 2	-0.1674	0.1396	-1.20	0.237	
AR 3	-0.5210	0.1270	-4.10	0.000	
Constant	-4.525	5.126	-0.88	0.382	

Final Estimates of Parameters					
Type	Coef	SE Coef	T	P	
MA 1	0.6037	0.1630	3.70	0.001	
MA 2	0.0339	0.1610	0.21	0.834	
MA 3	0.4664	0.1586	2.94	0.005	
Constant	-1.8607	0.7161	-2.60	0.013	

Gambar 4: Hasil Uji Signifikan untuk Model ARIMA (3,2,0) (kiri) dan ARIMA(0,2,3) (kanan)

2. *White noise*

Pada Gambar 5 di bawah ini menunjukkan nilai *p-value* baik pada model ARIMA (3,2,0) dan ARIMA (0,2,3) kurang dari 0,05. Hal ini berarti bahwa residu kedua model tersebut memenuhi *white noise*.

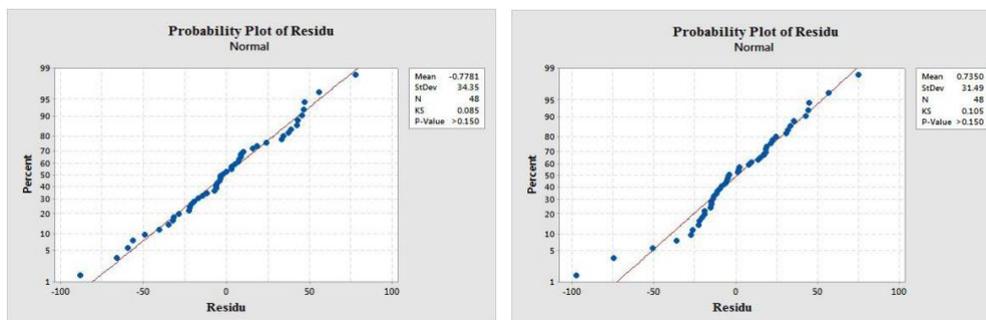
Modified Box-Pierce (Ljung-Box) Chi-Square statistic					
Lag	12	24	36	48	
Chi-Square	10.7	16.7	29.0	*	
DF	8	20	32	*	
P-Value	0.218	0.675	0.617	*	

Modified Box-Pierce (Ljung-Box) Chi-Square statistic					
Lag	12	24	36	48	
Chi-Square	12.2	20.0	26.6	*	
DF	8	20	32	*	
P-Value	0.143	0.461	0.737	*	

Gambar 5: Uji *Ljung-Box* untuk Model ARIMA (3,2,0) (kiri) dan ARIMA(0,2,3) (kanan)

3. Uji kenormalan

Pada Gambar 6 menunjukkan bahwa dengan menggunakan uji Kolmogorov Smirnov, nilai p-value dari kedua model ARIMA yaitu ARIMA (3,2,0) dan ARIMA (0,2,3) lebih besar dari 0,05. Artinya, residu kedua model tersebut mengikuti distribusi normal.



Gambar 6: Grafik *Normal Probability Plot* untuk Model ARIMA (3,2,0) (kiri) dan ARIMA (0,2,3) (kanan)

Tabel 1: Hasil Pemeriksaan Diagnostik Model ARIMA

Pemeriksaan Diagnostik	ARIMA (3,2,0)	ARIMA (0,2,3)
Uji signifikansi parameter	tidak signifikan	signifikan
<i>White noise</i>	memenuhi	memenuhi
Uji kenormalan	memenuhi	memenuhi

Berdasarkan Tabel 1, model yang akan digunakan pada proses peramalan adalah model ARIMA(0,2,3). Karena hanya ada satu model yang layak digunakan, maka pemilihan model terbaik tidak dilakukan.

3.3 Peramalan

Tahap terakhir pada metode ARIMA adalah peramalan. Pada penelitian ini akan dilakukan peramalan untuk 13 periode ke depan dari data awal sebanyak 63 periode. Model yang akan digunakan adalah ARIMA (0,2,3) yang secara matematis dinyatakan sebagai berikut:

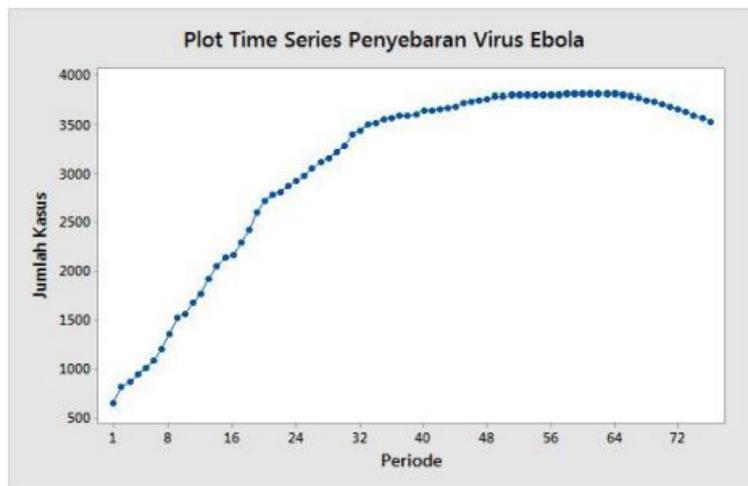
$$X_t = 2X_{t-1} - X_{t-2} + a_t - (0,5961)a_{t-1} - (-0,0552)a_{t-2} - (0,4537)a_{t-3}. \quad (1)$$

Dengan menerapkan model ARIMA(0,2,3) untuk 63 periode data, dan prediksi 13 periode kedepan, diperoleh data hasil ramalan sebagai berikut:

Tabel 2: Hasil Ramalan 13 Minggu Kedepan

Periode	Minggu Ke-	Tahun	Ramalan
1	45	2015	3800,68
2	46	2015	3791,07
3	47	2015	3777,20
4	48	2015	3761,28
5	49	2015	3743,31
6	50	2015	3723,28
7	51	2015	3701,19
8	52	2015	3677,05
9	53	2015	3650,86
10	1	2016	3622,61
11	2	2016	3592,30
12	3	2016	3559,94
13	4	2016	3525,52

Berdasarkan Tabel 2 di atas, dapat diketahui bahwa hasil ramalan untuk 13 minggu kedepan mengalami penurunan dari minggu ke minggu dengan persentase penurunan sebesar 7,24% dengan tingkat kepercayaan 95%. Dengan menggabungkan hasil ramalan, maka plot *time series* dari keseluruhan data sebagai berikut:



Gambar 7: Plot *Time Series* Keseluruhan Data

4 Kesimpulan

Model yang sesuai yang dapat diterapkan dalam prediksi penyebaran jumlah kasus virus ebola di Guinea Afrika Barat adalah ARIMA(0,2,3) pada Persamaan 1. Hasil prediksi untuk 13 minggu ke depan menunjukkan bahwa

penyebaran jumlah kasus virus ebola di Guinea Afrika Barat mengalami penurunan dari minggu ke minggu, dengan persentase penurunan sebesar 7,24% dan tingkat kepercayaan 95%.

Daftar Pustaka

- [1] Aswi dan Sukarna. 2006. *Analisis Deret Waktu: Teori dan Aplikasi*. Andira Publisher. Makassar.
- [2] Brockwell, P.J. dan Davis, R.A. 2002. *Introduction To Time Series and Forecasting*. Springer-Verlag New York, Inc., New York.
- [3] Center of Disease Control and Prevention. 2015. *Ebola (Ebola Virus Disease)*. CDC. US.
- [4] Mulyana. 2004. *Analisis Data Deret Waktu*. Universitas Padjajaran. Bandung.
- [5] Murphy, F. A. 2015. *Ebola Virus Disease-An Introduction*. WHO, Western Pacific Region.
- [6] Sudjana. 1989. *Metoda Statistik*. Tarsito. Bandung.
- [7] WHO. 2015. Ebola Data and Statistic. <http://apps.who.int/gho/data/view.ebola-sitrep.ebola-summary-20150211?lang=en>. Diakses pada 22 Desember 2015.
- [8] WHO. 2015. Situation Report-18 November (2015). http://apps.who.int/iris/bitstream/10665/134449/1/roadmapupdate18nov15_eng.pdf?ua=1. Diakses pada 1 Desember 2015.