

Peramalan Penyebaran Jumlah Kasus Virus Covid-19 Provinsi Jawa Tengah dengan Metode Arima

Ariska Kurnia Rachmawati¹, Seftina Diah Miasary²

¹UIN Walisongo Semarang, ariskakurnia@walisongo.ac.id

²UIN Walisongo Semarang, seftina.diyah.m@walisongo.ac.id

DOI 10.31102/zeta.2021.6.1.11-16

ABSTRACT

The Covid-19 virus disease, a name given by WHO for patients with novel coronavirus infection, was first reported from the city of Wuhan, China at the end of 2018. The spread occurred rapidly and created a new pandemic threat. The purpose of this study was to analyze the development of a pandemic in Indonesia, especially in Central Java province. The method used is ARIMA modeling, to determine the percentage of Covid19 development in Central Java. By using the ARIMA method, the ARIMA model (1,1,1) is obtained as a model that corresponds to the smallest MAPE value, namely 4.83821 which can be used to predict the number of daily positive cases of Covid-19. Based on this model, with the results that the number of spread of the number of positive cases can be predicted over the next 10 periods, with the result that the number of spread of the number of cases of the Covid-19 virus has decreased from period to period.

Keywords: **mathematical model, Arima Modeling, Covid-19**

ABSTRAK

Virus Covid-19 sebuah nama yang diberikan oleh WHO bagi pasien dengan infeksi virus novel corona 19 yang pertama kali dilaporkan dari kota Wuhan, China pada akhir tahun 2018. Penyebaran terjadi secara cepat dan membuat amcaman pandemic baru. Tujuan penelitian ini adalah untuk menganalisa perkembangan pandemic di wilayah Indonesia khususnya provinsi Jawa Tengah. Metode yang digunakan adalah pemodelan ARIMA, untuk menentukan prosentase perkembangan covid19 di Jawa Tengah. Dengan menggunakan metode ARIMA, diperoleh model ARIMA(1,1,1) sebagai model yang sesuai dengan nilai MAPE terkecil yaitu 4,83821 dapat digunakan untuk memprediksi jumlah kasus positif harian covid-19. Berdasarkan model tersebut, dengan hasil bahwa jumlah penyebaran jumlah kasus positif dapat diprediksi selama 10 periode kedepan, dengan hasil bahwa jumlah penyebaran jumlah kasus virus covid-19 mengalami penurunan dari periode ke periode.

Kata Kunci: **mathematical model, Arima Modeling, Covid-19**

1. PENDAHULUAN

Penyakit virus Covid-19 sebuah nama baru yang diberikan oleh World Health Organization (WHO) bagi pasien dengan infeksi virus novel corona 2019 yang pertama kali dilaporkan dari kota Wuhan, Cina pada akhir 2019. Penyebaran terjadi secara cepat dan membuat ancaman pandemi baru. Pada tanggal 10 Januari 2020, etiologi penyakit ini diketahui pasti yaitu termasuk dalam virus *ribonucleic acid (RNA)* yaitu virus *corona* jenis baru, betacoronavirus dan satu kelompok dengan virus corona penyebab *Severe Acute Respiratory Syndrome (SARS)* dan *Middle East Respiratory Syndrome (MERS Cov)*. Menularnya Covid-19 membuat dunia menjadi resah termasuk Indonesia. Di Indonesia telah diambil kebijakan yaitu *social distancing* atau pembatasan sosial karena apabila menerapkan kebijakan lockdown seperti negara yang telah memberlakukannya akan berdampak pada perekonomian.

Pada kondisi ini, Virus Covid-19 bukanlah wabah yang dapat diabaikan begitu saja. Bagi para analis kesehatan dan kedokteran, virus ini sangat berbahaya dan mematikan. Saat ini di tahun 2020, perkembangan penularan virus cukup signifikan karena penyebarannya sudah mendunia termasuk Indonesia. Selama hampir 1 tahun, pandemi telah mewabah di Indonesia, dan provinsi Jawa Tengah termasuk daerah yang tingkat penyebaran virus tertinggi di Indonesia. Selain tingkat penyebarannya yang sangat tinggi dan mudah menulari orang lain, Virus Covid-19 dapat menyebabkan krisis ekonomi termasuk di Indonesia khususnya Jawa Tengah yang notabene penduduknya paling padat diantara provinsi lain. Kegiatan peramalan dilakukan dengan tujuan memprediksi perkembangan Virus Covid-19 dimana peramalan / forecasting ini merupakan suatu kegiatan estimasi tentang apa yang akan terjadi di masa mendatang. Peramalan memerlukan data runtun waktu yang dikumpulkan, dicatat atau diobservasi sepanjang waktu secara berurutan.

Peramalan (*forecasting*) digunakan untuk memprediksi sesuatu yang kemungkinan besar akan terjadi di masa mendatang sehingga tindakan yang tepat dapat dilakukan. Metode data runtun waktu Box-Jenkins (ARIMA) adalah salah satu metode peramalan di mana dalam melakukan analisis ARIMA digunakan prosedur Box-Jenkins dimana tahap awal yang dilakukan adalah melakukan identifikasi data untuk mengetahui stasioneritas data sebagai asumsi awal yang harus dipenuhi sebelum melakukan uji lanjut. Akan tetapi pada prakteknya, terkadang penganalisa data mengalami kesulitan dalam melakukan uji stasioneritas

data. Sehingga dibutuhkan suatu software yang dapat memudahkan hal tersebut.

Analisis data runtun waktu memungkinkan untuk mengetahui perkembangan suatu kejadian serta hubungan maupun pengaruhnya terhadap kejadian lainnya (Setiawan, 2013: 172). ARIMA merupakan penggabungan dari metode *moving average* dan metode *autoregressive* yakni suatu metode peramalan data runtun waktu yang memanfaatkan data historis dan data sekarang untuk menghasilkan peramalan jangka pendek yang akurat. Hutasuht, dkk (2014) mengatakan bahwa metode ARIMA adalah metode yang fleksibel karena mengikuti pola data yang ada dan memiliki akurasi tinggi serta cenderung memiliki nilai error yang kecil karena prosesnya yang terperinci.

Dalam penelitian ini akan dicari perkembangan dari Virus Covid-19 di provinsi Jawa Tengah dengan menggunakan data mulai 1 Agustus sampai 13 Desember 2020 dengan melakukan simulasi menggunakan metode peramalan ARIMA. Metode ARIMA yang digunakan dalam penelitian ini dapat digunakan untuk memprediksi perkembangan dari data Virus Covid-19 untuk kedepannya dikarenakan setiap perkembangan mengalami perbedaan jumlah di setiap harinya dan memiliki kapasitas peningkatan yang pesat.

2. TINJAUAN PUSTAKA

2.1 Model Time Series

Analisis deret waktu (*time series*) diperkenalkan pada tahun 1970 oleh E.P. Box dan Gwilym M. Jenkins melalui bukunya yang berjudul *Time Series Analysis : Forecasting and Control* (Aswi dan Sukarna, 2006). Menurut Brockwell dan Davis (2002), *Time Series* adalah suatu rangkaian pengamatan x_t , yang dilakukan pada waktu t . Periode waktu t dapat berbentuk tahun, kuartal, bulan, minggu dan beberapa kasus dapat juga hari atau jam. Dalam memilih model *time series* perlu memperhatikan pola data dimana dibagi menjadi empat bagian, yaitu pola horizontal, pola musiman, pola trend dan pola siklis.

Dalam penentuan model *time series*, diasumsikan bahwa data dalam keadaan stasioner. Sebuah *time series* $\{X_t, t = 0, \pm 1\}$ dikatakan stasioner jika memiliki sifat statistik yang sama dengan deret waktu selanjutnya $\{X_{t+h}, t = 0, \pm 1\}$ untuk setiap bilangan bulat h . Stasioneritas dibagi menjadi 2 bagian yaitu stasioner dalam mean dan stasioner dalam varian. Jika stasioner dalam rata-rata tidak terpenuhi, maka perlu dilakukan differencing. Langkah selanjutnya dilakukan beberapa proses pemeriksaan diagnostic sebagai berikut :

1. Fungsi Autokorelasi (ACF)

Koefisien autokorelasi adalah suatu fungsi yang menunjukkan besarnya korelasi (hubungan linier) antara pengamatan pada waktu 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 :

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

Karena r_k merupakan fungsi atas k , maka hubungan koefisien autokorelasi dengan lagnya disebut dengan fungsi autokorelasi dan dinyatakan dengan ρ_k (Dedi Rosadi, 2016)

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} variable antara $X_{t-1}, X_{t-2}, \dots, X_{t-k-1}$.

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

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

3. White Noise

Suatu proses dikatakan sebagai proses *white noise* (proses yang bebas dan identik) jika bentuk variable random yang berurutan tidak saling berkorelasi dan mengikuti distribusi tertentu. Untuk mengetahui residu tersebut memenuhi proses *white noise* atau tidak perlu dilakukan uji yang salah satunya adalah uji Ljung Box.

4. Uji Asumsi Kenormalan

Selain memenuhi white noise, model juga harus memenuhi asumsi kenormalan (Dedi Rosadi, 2016). Salah satu cara yang dapat dilakukan adalah dengan melakukan uji kenormalan Komolgorov Smirnov.

2.2 Model Peramalan ARIMA

Model – model *Autoregressive integrated Moving Average* (ARIMA) telah dipelajari secara mendalam oleh George Box dan Gwilym Jenkins, dan nama mereka disinonimkan dengan proses ARIMA yang diterapkan untuk analisis *time series*, peramalan dan pengendalian. 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 stationer. Penentuan orde tersebut tergantung pada konsisi data. Pada penetapan model yang sesuai harus melalui beberapa tahap yaitu :

1. Tahap Identifikasi

Pada tahap ini, akan dilakukan pengidentifikasian jenis model yang dianggap sesuai.

2. Tahap Penaksiran dan Pengujian

Langkah ini dilakukan penaksiran terhadap parameter dalam model dan dilakukan diagnose checking untuk menyelidiki kelayakan model.

3. Tahap Penerapan

Setelah mendapat model yang sesuai, Langkah selanjutnya dalam analisis runtun waktu adalah melakukan peramalan. (Hendikawati, 2015)

2.3 Pemilihan Model Terbaik

Kriteria dalam pemilihan model terbaik menggunakan MSE (*Mean Square Error*) yang minimum dirumuskan sebagai berikut :

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

3. METODE PENELITIAN

3.1 Data dan Populasi

Adapun data penelitian yang digunakan adalah data sekunder yaitu kasus positif Virus Covid-19 di provinsi Jawa Tengah pada periode waktu 7 Agustus – 13 Desember 2020 yang diambil dari website covid19.go.id. Data diperoleh berupa data penambahan kasus positif per hari (data harian) di provinsi Jawa Tengah.

3.2 Teknik Analisis Data

Suatu model *time series* dikatakan baik jika telah sesuai dengan kenyataan. Dengan kata lain, apabila kesalahan model semakin kecil, maka model bisa dikatakan baik (Iriawan, 2006).

Analisis data dilakukan dengan menggunakan metode ARIMA dengan bantuan *software R*. Langkah-langkah penerapan metode ARIMA adalah sebagai berikut :

1. Pemeriksaan kestasioneran data.

2. Identifikasi model dalam ARIMA.

Melalui plot ACF dan PACF kita dapat menentukan model ARIMA yang bisa digunakan dalam prediksi.

3. Penentuan parameter p , d dan q dalam ARIMA.

4. Penentuan persamaan model ARIMA.

5. Validasi prediksi.

6. Prediksi.

Dengan menggunakan model terbaik dapat digunakan untuk prediksi penambahan jumlah kasus positif per hari pada bulan berikutnya.

4. HASIL PENELITIAN

4.1. Pemeriksaan Kestasioneran Data

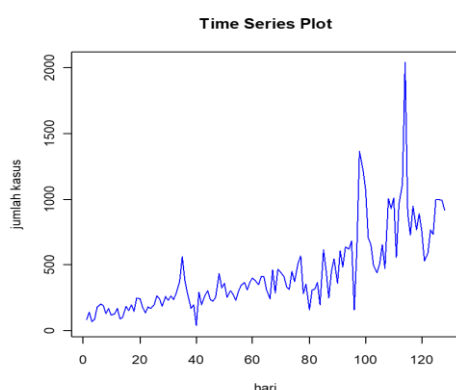
Data yang digunakan dalam penelitian ini

merupakan data penambahan kasus positif Covid-19 selama 128 hari. Data jumlah kasus Covid-19 tersaji dalam table berikut :

Tabel 1. Data Harian Kasus Virus Covid-19 Provinsi Jawa Tengah

No	Tanggal	Kasus positif harian
1	07/08/2020	232
2	08/08/2020	88
3	09/08/2020	140
4	10/08/2020	67
5	11/08/2020	86
6	12/08/2020	179
7	13/08/2020	204
⋮	⋮	⋮
128	13/12/2020	917

Plot data penambahan kasus positif selama 128 hari adalah sebagai berikut :

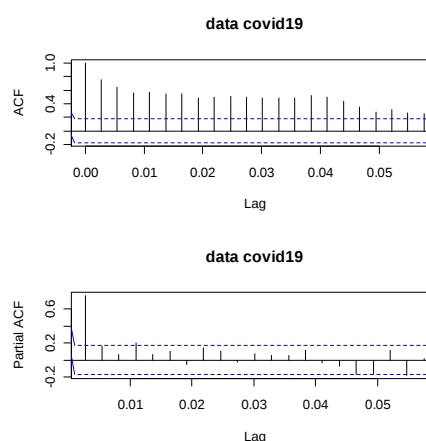


Gambar 1. Time Series Plot Kasus Covid-19

Dalam tahap awal, dilakukan identifikasi model runtun waktu yang digunakan untuk memodelkan sifat-sifat data. Identifikasi dilakukan dengan cara sederhana melalui plot dari data seperti Gambar 1 untuk melihat adanya trend, komponen musiman, nonstationeritas dalam variansi, dan lain lain. Terlihat data mengandung *trend* linier, yang selanjutnya dikonfirmasi menggunakan uji unit root dengan uji *Augmented Dickey-Fuller (ADF)* atau menggunakan ACF PACF sebagai berikut :

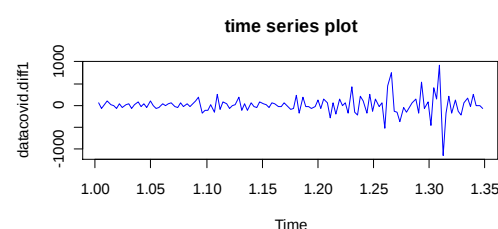
Tabel 2. Output Uji ADF

<i>Augmented Dickey-Fuller Test</i>
<i>data: datacovid</i>
<i>Dickey-Fuller = -3.5501, Lag order = 5,</i>
<i>p-value = 0.04062</i>
<i>alternative hypothesis: stationary</i>



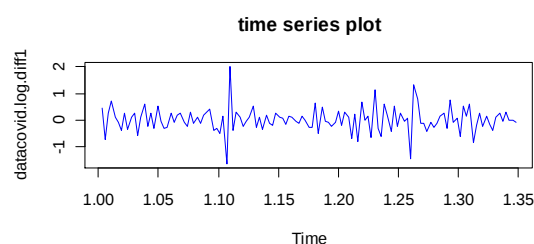
Gambar 2. Plot ACF Kasus Covid-19

Uji ADF menunjukkan bahwa adanya unit *root* dalam data (data tidak *stationer*) yang dipengaruhi oleh pola *trend* (tidak *stationer* dalam *mean*). Karena data kasus Covid-19 memiliki bentuk *trend*, diperlukan transformasi untuk membentuk data yang *stationer* (dalam *mean*) . Transformasi yang dilakukan untuk membuang trend adalah dengan melakukan *differencing* data. Hasilnya dapat dilihat dari gambar dibawah ini :



Gambar 3. Plot Hasil Differencing

Pada Gambar 3 , terlihat data belum stationer dalam variansi, harus dilakukan transformasi dari data sebelum dilakukan operasi *difference*. Bentuk transformasi yang digunakan adalah dengan transformasi Log (*base e*) atau secara umum dengan menggunakan transformasi Box-Cox. Hasil yang diperoleh adalah sebagai berikut :

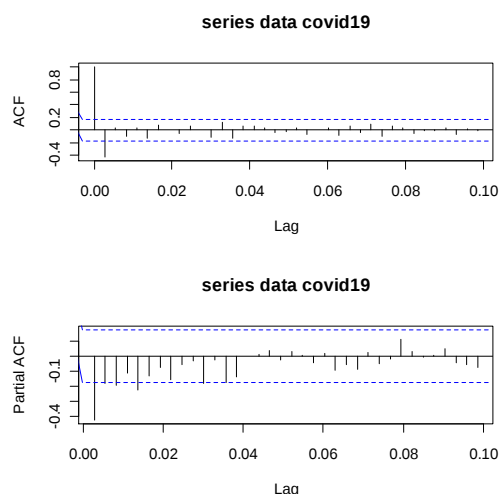


Gambar 4. Plot Hasil Differencing Tingkat 1

Data telah melalui proses *differencing* tingkat 1. Dari data tersebut dapat diamati adanya data yang sudah bersifat stationer. Proses *differencing* yang sudah dilakukan mengindikasikan bahwa nilai yang bisa dipakai adalah nilai $d = 1$.

4.2. Identifikasi Model Dalam ARIMA

Selanjutnya akan dilakukan identifikasi dari model ARIMA yang tepat untuk menggambarkan data hasil difference logaritma dengan menggunakan plot ACF PACF sebagai berikut :



Gambar 5. Plot ACF Data Kasus Covid-19 Hasil Difference

Berdasarkan Gambar 5, terlihat bahwa fungsi ACF/PACF signifikan pada *lag-1* dan meluruh menuju nol untuk lag yang lain. Apabila ACF dan PACF turun secara eksponensial atau secara cepat maka hal ini menunjukkan model ARIMA yang bisa digunakan adalah model ARIMA(p,d,q).

Untuk mengetahui nilai parameter ARIMA yang lebih terperinci dilakukan penentuan parameter p,d, dan q. Tingkat akurasi prediksi dapat diukur melalui nilai MAPE (*Mean Absolute Percentage error*) yang terkecil sehingga dapat diambil keputusan apabila suatu model layak digunakan.

Beberapa alternatif model ARIMA dengan $y_t = \Delta \log(\text{datacovid})$ adalah :

1. Model 1 AR(1) adalah model ARIMA(1,1,0)
2. Model 2 MA(1) adalah model ARIMA(0,1,1)
3. Model 3 ARMA(1) adalah model ARIMA(1,1,1)
4. Model 4 AR(2) adalah model ARIMA(2,1,0)
5. Model 5 MA(2) adalah model ARIMA(0,1,2).

4.3. Penentuan persamaan model ARIMA terbaik

Harga estimasi dari koefisien, standar error dari koefisien dan harga-harga statistic untuk diagnostic checking bagi model model yang diamati diatas terangkum dalam table berikut :

Tabel 3. Rangkuman Output Hasil Estimasi Model ARIMA

	ARIMA (1,1,0)	ARIMA (0,1,1)	ARIMA (1,1,1)	ARIMA (2,1,0)	ARIMA (0,1,2)
a1	-0,4191 SE=0,0804		0,2183 SE=0,1081	-0,4890 SE=0,0872	
a2				-0,1693 SE=0,0877	
b1		-0,7234 SE=0,0892	0,8300 SE=0,0573		-0,6428 SE=0,0612
b2					-0,1300 ES=0,0762
RMSE	0.40911624	0.38734463	0.38166253	0.40316012	0.38298081
MAPE	5,21883	5,02071	4,88921	5,18712	4,91519
AIC	138.59	125.24	123.57	136.92	124.45
BIC	144.27	130.93	132.1	145.45	132.98

Dari table 2 terdapat 5 model yang digunakan untuk prediksi jumlah kasus positif Covid-19, berdasarkan teori semakin kecil nilai MAPE yang dihasilkan maka model semakin baik. Sehingga diketahui bahwa parameter $p=1$, $d=1$ dan $q=1$ atau ARIMA(1,1,1) dengan nilai MAPE terkecil yaitu 4,83821 dapat digunakan untuk memprediksi jumlah kasus positif harian covid-19.

4.4. Pemeriksaan Diagnostik

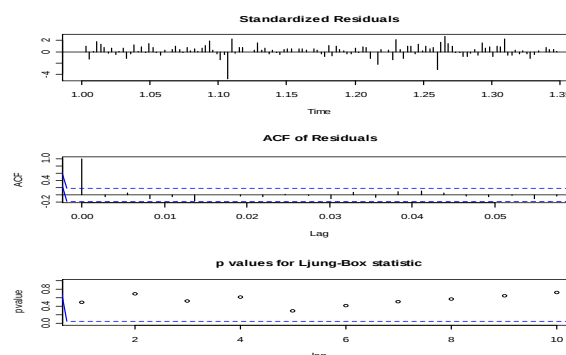
Berdasarkan model yang diperoleh yaitu ARIMA(1,1,1) dilakukan pemeriksaan uji diagnostic yang meliputi uji signifikansi parameter, *white noise*, dan uji kenormalan.

1. Uji signifikansi parameter

Pada gambar 6 menunjukkan bahwa untuk AR(1) pada model ARIMA(1,1,1) nilai *t-hitung* = 2,0194 dan MA(1) nilai *t-hitung* = 14,48 lebih besar daripada nilai *statistic table t* ($df = n-1 = 128 - 1 = 127$; $\alpha = 2.5\%$) = 1,97. Dengan demikian koefisien ini signifikan.

Tabel 4. Estimasi dari ARIMA(1,1,1)

Coefficient					
	AR1	MA1			
	0,2183	0,8300			
s.e	0,1081	0,0573			
Sigma^2 estimated as			0,1468		
Log likelihood			-57,78		
ME	RMSE	MAE	MPE	MAPE	MASE
0,7706	0,381662	0,27762	0,9374	4,8892	0,8615



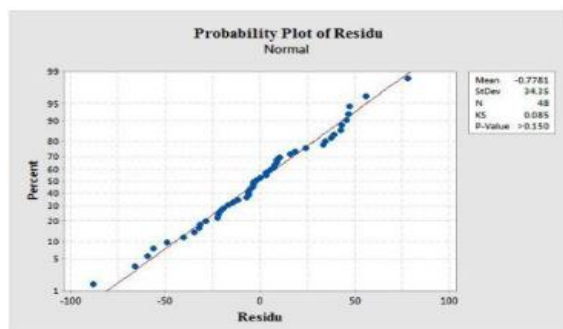
Gambar 6. Plot Diagnostik Model 3

2. White Noise

Pada gambar 6 diatas menunjukkan nilai p -value pada model ARIMA(1,1,1) kurang dari 0,05. Hal ini berarti bahwa residu dari model tersebut memenuhi *white noise*.

3. Uji kenormalan

Pada gambar 6 menunjukkan bahwa dengan menggunakan uji Kolmogorov Smirnov, dengan nilai p -value dari model ARIMA(1,1,1) lebih besar dari 0,05. Artinya residu dari model tersebut mengikuti distribusi normal.



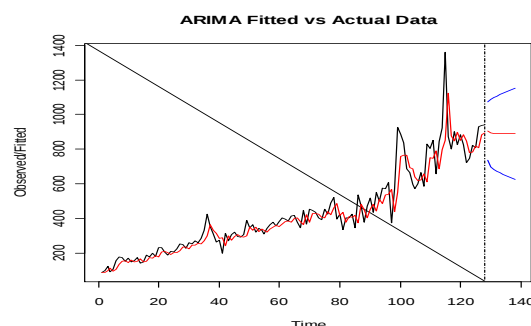
Gambar 7. Plot Distribusi Normal

4.5. Peramalan

Tahapan yang terakhir pada metode ARIMA adalah prediksi / peramalan. Pada penelitian ini akan dilakukan peramalan untuk 10 periode kedepan. Model yang akan digunakan adalah ARIMA(1,1,1) untuk 128 periode data, dan prediksi 10 periode kedepan, diperoleh hasil :

Tabel 5. Tabel Prediksi Kasus Covid-19 di Jawa Tengah

Hari Ke-	Hasil Prediksi	Batas bawah hasil prediksi	Batas atas hasil prediksi
129	905,96	737,95	1073,98
130	893,56	697,92	1089,21
131	890,23	682,23	1098,22
132	889,43	672,34	1106,521
133	889,26	663,99	1114,51
134	889,22	656,19	1122,23
135	889,21	648,70	1129,711
136	889,20	641,44	1136,96
137	889,21	634,40	1144,01
138	889,22	627,54	1150,87



Gambar 8. Plot Data Aktual dan Hasil Prediksi

5. KESIMPULAN

Berdasarkan hasil penelitian menunjukkan bahwa kasus positif harian di provinsi Jawa Tengah sangat fluktuatif dari jumlah 232 orang pada 7 Agustus 2020 sampai 917 pada tanggal 13 Desember 2020, sehingga dari data kasus yang telah disimulasikan diprediksi bahwa sampai bulan Desember 2020 yaitu pada hari ke 129 mengalami peningkatan jumlah kasus positif kemudian hari ke 138 akan mengalami penurunan jumlah kasus positif Virus Covid-19. Model *time series* yang sesuai yang dapat diterapkan dalam prediksi penularan virus Covid-19 di Jawa tengah adalah ARIMA(1,1,1) dengan hasil output diperoleh nilai RMSE 0.38166253; MAE 0.27761916; MPE 0.93741765; MAPE 4.88921623; MASE 0.86158633.

DAFTAR PUSTAKA

- Brockwell, P.J. dan Davis, R.A. (2002). *Introduction To Time Series and Forecasting*. Springer-verlag, New York, Inc. New York.
- Diah Handayani, dkk. (2019). *Penyakit Virus Corona 2019*. Jurnal Resporologi Indonesia Volume 40 No 2.
- Hendikawati, P. (2015). *Peramalan Data Runtun Waktu Metode dan Aplikasinya dengan Minitab & Eviews*. Semarang: FMIPA Unnes.
- Hutasuhut, dkk. (2014). *Pembuatan Aplikasi Pendukung Keputusan untuk Peramalan Persediaan Bahan Baku Produksi Plastik Blowing dan Inject Menggunakan Metode ARIMA (Auto Regressive Integrated Moving Average) di CV. ASIA*. Jurnal Teknik POMITS.
- <https://covid19.go.id/peta-sebaran-covid19>
diakses pada 13 Desember 2020.
- Rosadi, D. (2016). *Analisis Runtun Waktu dan Aplikasinya dengan R*. Gadjah Mada University Press. Yogyakarta.