

Perbandingan ARIMA dan *Artificial Neural Networks* dalam Peramalan Jumlah Positif Covid-19 di DKI Jakarta *

Tri Wahyuni¹, Indahwati^{2‡}, Kusman Sadik³

¹²³Department of Statistics, IPB University, Indonesia
[‡]corresponding author: indahwati@apps.ipb.ac.id

Copyright © 2021 Tri Wahyuni, Indahwati, and Kusman Sadik. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

DKI Jakarta is the center of the spread of Covid-19. This is indicated by the higher cumulative number of Covid-19 positive in DKI Jakarta compared to other provinces. The high number of cases in DKI Jakarta is a concern for all groups, so it is necessary to do forecasting to predict the number of Covid-19 positive in the next period. Accurate forecasting is needed to get better results. This study compares the Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN) methods in predicting the number of Covid-19 positive in DKI Jakarta. Forecasting accuracy is calculated using the value of Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and correlation. The results show that the best model for forecasting the number of Covid-19 positive in DKI Jakarta is ARIMA(0,1,1) with drift, with a MAPE value of 15.748, an RMSE of 268.808, and the correlation between the forecast value and the actual value of 0.845. Forecasting using ARIMA(0,1,1) with drift and BP(3,10,1) models produces the best forecast for the long forecasting period of the next six weeks.

Keywords: ANN, ARIMA, Covid-19, forecast.

* Received: Aug 2021; Reviewed: Sept 2021; Published: Sept 2021

1. Pendahuluan

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-Cov-2) merupakan virus jenis baru yang menyebabkan penyakit *Coronavirus Disease 2019* (Covid-19). *Coronavirus disease* jenis baru ditemukan pada manusia sejak kejadian luar biasa di Wuhan, China pada 31 Desember 2019. Angka kematian Covid-19 kurang dari 5% lebih kecil dibandingkan SARS sebesar 9,6%, tetapi penyebaran Covid-19 lebih luas dan cepat ke beberapa negara dibanding dengan SARS (Kemenkes 2020). Virus Covid-19 saat ini sangat mudah menular dan menyebabkan kematian akibat susah bernafas. Terlepas dari peningkatan fasilitas pemerintah, pengujian, dan pemberian vaksinasi, jumlah kasus positif terus mengalami penambahan setiap harinya.

Indonesia adalah negara dengan kasus jumlah positif Covid-19 terbanyak ke-20 pada 18 Desember 2020. Indonesia pertama kali mengonfirmasi kasus positif Covid-19 pada 2 Maret 2020 sebanyak dua orang. DKI Jakarta menjadi episentrum penyebaran Covid-19 di Indonesia. Menurut Satuan Tugas Penanganan Covid-19, pada 18 Desember 2020 DKI Jakarta menyumbang 24,6% kasus positif Covid-19 dari jumlah kasus terkonfirmasi nasional. Jumlah kumulatif kasus terkonfirmasi positif Covid-19 di DKI Jakarta terbanyak dibandingkan provinsi lain. Tingginya penambahan kasus Covid-19 di DKI Jakarta menjadi perhatian bagi semua kalangan, sehingga perlu dilakukan peramalan untuk mengetahui gambaran penambahan kasus pada periode selanjutnya.

Hasil peramalan yang baik dihasilkan jika metode peramalan yang digunakan sesuai dengan kondisi data. Data jumlah positif Covid-19 di DKI Jakarta memiliki *trend* naik dan tidak stasioner. Salah satu metode yang dapat digunakan untuk peramalan data yang tidak stasioner adalah metode *Autoregressive Integrated Moving Average* (ARIMA). Model ARIMA disebut juga dengan model Box-Jenkins yang mengasumsikan fungsi linear dari beberapa pengamatan dimasa lalu. Asumsi kestasioneran menjadi hal yang harus dipenuhi pada model ARIMA. Namun, ketika model linear menghasilkan tingkat akurasi peramalan yang kecil dan kesalahan peramalan yang besar, kemungkinan model taklinear (*nonlinear*) mampu menjelaskan dan meramalkan dengan lebih baik dibandingkan dengan model linear (Zhang 2003). Selain itu, dalam dunia nyata banyak sekali ditemukan data yang taklinear, sehingga metode ARIMA mungkin kurang sesuai untuk menggambarkan data tersebut. *Artificial Neural Networks* (ANN) merupakan model yang mampu menjelaskan permasalahan yang kompleks dengan hubungan yang taklinear untuk peramalan jangka panjang (Adeyinka dan Muhajarine 2020). ANN adalah salah satu representasi buatan dari otak manusia yang menyimulasikan proses pembelajaran melalui komputasi. Adebiyi et al. (2014) menunjukkan bahwa model ARIMA dan ANN dapat mencapai ramalan yang baik dalam penerapannya pada masalah kehidupan nyata.

Penelitian sebelumnya, Sahai et al. (2020) melakukan peramalan dengan pemodelan ARIMA di lima negara yang mengalami dampak terbesar Covid-19 dan menghasilkan MAPE kurang dari 5%. Wiczorek et al. (2020) menunjukkan metode ANN memiliki nilai akurasi yang lebih tinggi dan mencapai akurasi maksimum dalam iterasi yang lebih sedikit dibandingkan RNN dalam peramalan penyebaran Covid-19. Oleh karena itu, penelitian ini dilakukan dengan pengujian beberapa parameter untuk

mengidentifikasi nilai parameter terbaik dari model linear ARIMA dan model taklinear ANN dalam peramalan jumlah positif Covid-19 di DKI Jakarta.

2. Metodologi

2.1 Data

Data yang digunakan pada penelitian ini adalah data sekunder. Data tersebut merupakan data harian jumlah positif Covid-19 di DKI Jakarta periode 2 Maret 2020 hingga 31 Desember 2020 sebanyak 305 data pengamatan. Data dibagi menjadi data latih dan data uji. Data latih dari 2 Maret 2020 hingga 31 Oktober 2020, sedangkan data uji dari 1 November 2020 hingga 31 Desember 2020.

2.2 Metode Penelitian

Tahapan-tahapan analisis yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Melakukan eksplorasi data untuk mengetahui gambaran dan mengidentifikasi pola data dengan cara melihat plot.
2. Membagi data menjadi dua bagian, yaitu data latih dan data uji. Adeyinka dan Muhajarine (2020), membagi data latih sebanyak 80% dan data uji sebanyak 20% berdasarkan pada prinsip Pareto.
3. Langkah-langkah metode ARIMA

- a. Pemeriksaan kestasioneran data dalam ragam dan rataan. Kestasioneran ragam diperiksa menggunakan plot Box-Cox, sedangkan kestasioneran rataan diperiksa melalui uji *Augmented Dickey Fuller* (ADF). Hipotesis uji ADF (Rusdi 2011):

$$H_0: \gamma = 0 \text{ (data tidak stasioner)}$$

$$H_1: \gamma < 0 \text{ (data stasioner)}$$

Statistik uji :

$$\tau = \frac{\hat{\gamma} - \gamma}{SE(\hat{\gamma})}$$

dengan $SE(\hat{\gamma})$ adalah galat baku dari $\hat{\gamma}$ dan $\hat{\gamma}$ adalah penduga kuadrat terkecil dari γ . Tolak H_0 apabila nilai kritis statistik-t kurang dari nilai kritis ADF atau nilai peluang dari statistik uji ADF kurang dari α . Data yang belum stasioner dalam ragam akan dilakukan transformasi, sedangkan data yang belum stasioner dalam rataan akan dilakukan pembedaan.

- b. Mengidentifikasi model ARIMA melalui plot ACF dan PACF. Berikut ini perilaku dari plot ACF dan PACF (Cryer dan Chan 2008):

Tabel 1 Identifikasi model AR, MA, dan ARMA

Model	ACF	PACF
AR(p)	<i>Tails off</i>	<i>Cuts off</i> setelah lag ke- p
MA(q)	<i>Cuts off</i> setelah lag ke- q	<i>Tails off</i>
ARMA(p, q), $p > 0$ dan $q > 0$	<i>Tails off</i>	<i>Tails off</i>

- c. Melakukan pendugaan parameter dan pemeriksaan signifikan parameter.
 d. Melakukan diagnostik model untuk mengetahui kelayakan model, melalui pemeriksaan kebebasan dan kenormalan sisaan. Uji Ljung-Box digunakan untuk memeriksa kebebasan sisaan. Hipotesis:

H_0 : tidak terdapat autokorelasi sisaan

H_1 : terdapat autokorelasi sisaan

Statistik uji (Montgomery *et al.* 2015):

$$Q_{LB} = T(T + 2) \sum_{k=1}^K \left(\frac{1}{T - k} \right) r_k^2$$

dengan T adalah banyaknya data, k adalah lag ke- k , K adalah lag maksimum, dan r_k^2 adalah koefisien autokorelasi antar sisaan pada lag ke- k . Tak tolak H_0 jika nilai statistik uji Q_{LB} lebih kecil dari nilai $\chi_{\alpha, K-p-q}^2$ atau nilai peluang lebih besar dari nilai α .

Uji Kolmogorov-Smirnov digunakan untuk memeriksa kenormalan sisaan. Hipotesis:

H_0 : sisaan menyebar normal

H_1 : sisaan tidak menyebar normal

Statistik uji:

$$D = \sup_x |F_n(x) - F(x)|$$

dengan $F_n(x)$ adalah sebaran frekuensi kumulatif contoh dan $F(x)$ adalah sebaran frekuensi kumulatif empirik. Tak tolak H_0 jika statistik uji D lebih kecil dari nilai kritis pada tabel Kolmogorov-Smirnov atau nilai peluang yang dihasilkan lebih besar dari α .

- e. Melakukan *overfitting* untuk dibandingkan dengan model tentatif awal. Model yang didapat pada *overfitting* akan dilakukan uji signifikan parameter dan periksa asumsi seperti pada tahap diagnostik model.
 f. Memilih model terbaik dari model-model ARIMA yang sudah terbentuk dengan cara memilih model yang memiliki nilai AIC dan BIC terkecil. Menurut Montgomery *et al.* (2015), persamaan AIC dan BIC adalah:

$$AIC = \ln \left(\frac{\sum_{t=1}^T e_t^2}{T} \right) + \frac{2p}{T}$$

$$BIC = \ln \left(\frac{\sum_{t=1}^T e_t^2}{T} \right) + \frac{p \ln(T)}{T}$$

dengan $\sum_{t=1}^T e_t^2$ adalah kuadrat sisaan, p adalah banyaknya parameter dalam model, dan T adalah banyaknya data

4. Melakukan peramalan sebanyak data uji untuk menghitung nilai RMSE dan MAPE. Nilai MAPE dan RMSE dapat dihitung dengan rumus sebagai berikut (Montgomery *et al.* 2015):

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}}$$

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{n}$$

dengan y_t adalah nilai pengamatan pada waktu ke- t , \hat{y}_t adalah nilai dugaan pada waktu ke- t , dan n adalah banyaknya pengamatan.

5. Langkah-langkah metode *Artificial Neural Networks*
 - a. Menentukan *input* dan *output layer*. *Input layer* ditentukan berdasarkan *trail and error*, yaitu jumlah positif Covid-19 di DKI Jakarta beberapa hari sebelumnya, sedangkan *output layer* adalah jumlah Covid-19 di DKI Jakarta periode ke- t . Koefisien pembelajaran (*learning rate*) sebesar 0,01.
 - b. Melakukan normalisasi data, sehingga data *input* dan *output* berada pada selang nilai dari 0 sampai 1. Normalisasi data dilakukan dengan rumus sebagai berikut:

$$y'_t = \frac{y_t - y_{min}}{y_{maks} - y_{min}}$$

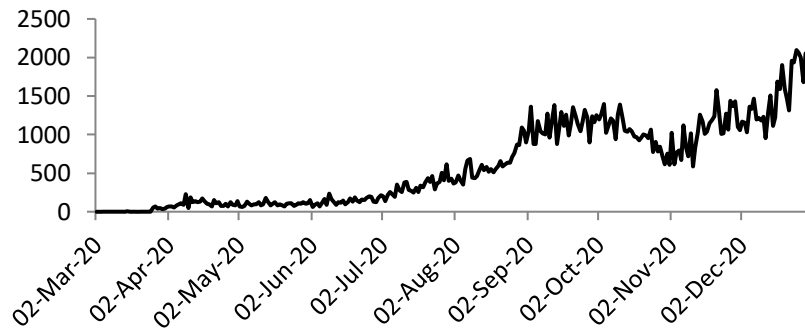
dengan y'_t adalah jumlah Covid-19 pada waktu ke- t yang telah dinormalisasi, y_t adalah jumlah Covid-19 pada waktu ke- t , y_{maks} adalah jumlah Covid-19 tertinggi, dan y_{min} adalah jumlah Covid-19 terendah.

- c. Membangun arsitektur ANN dari neuron *input* dan *output*, dengan terlebih dahulu menentukan banyaknya *hidden* dan jumlah neuron pada *hidden layer*. Neuron yang sudah ditentukan pada masing-masing lapisan akan diterapkan algoritma *backpropagation* secara berulang sampai didapatkan model yang diinginkan.
 - d. Melakukan proses validasi dengan menggunakan data uji setelah terbentuk model akhir untuk mendapatkan nilai RMSE dan MAPE.
6. Membandingkan hasil peramalan metode ARIMA dan ANN dengan melihat nilai RMSE dan MAPE terkecil serta hasil peramalan dari jumlah positif Covid-19 di DKI Jakarta.

3. Hasil dan Pembahasan

3.1 Eksplorasi Data

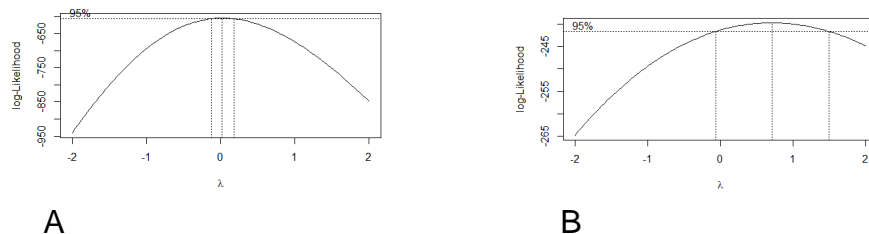
Eksplorasi data dilakukan untuk melihat gambaran dan mengidentifikasi data jumlah positif Covid-19 di DKI Jakarta yang disajikan dalam periode harian sebanyak 305 data dari 2 Maret 2020–31 Desember 2020. Sumbu x menunjukkan periode data dan sumbu y menunjukkan jumlah harian positif Covid-19 di DKI Jakarta. Gambar 1 memperlihatkan plot data deret waktu jumlah harian positif Covid-19 yang berfluktuasi dan cenderung memiliki *trend* naik. Data terlihat tidak stasioner baik dalam ragam maupun rata-rata. Total jumlah positif Covid-19 periode 2 Maret 2020–31 Desember 2020 adalah 183.246 kasus. Jumlah positif Covid-19 tertinggi terjadi pada 25 Desember 2020 sebanyak 2096 kasus. Hal tersebut terjadi bertepatan dengan adanya perayaan hari Natal, besarnya kenaikan kasus sebanyak 163 kasus dibandingkan tanggal 24 Desember 2020. Jumlah kasus terendah terjadi pada awal Maret 2020 sebanyak nol kasus. Rata-rata harian jumlah positif Covid-19 di DKI Jakarta periode 2 Maret 2020–31 Desember 2020 adalah 601 kasus. Rata-rata mingguan dan bulanan jumlah positif Covid-19 di DKI Jakarta periode tersebut masing-masing, yaitu 4385 dan 18.325 kasus. Jumlah positif Covid-19 di DKI Jakarta mengalami lonjakan tertinggi antara bulan November sampai Desember 2020.



Gambar 1 Plot data jumlah harian positif Covid-19 di DKI Jakarta

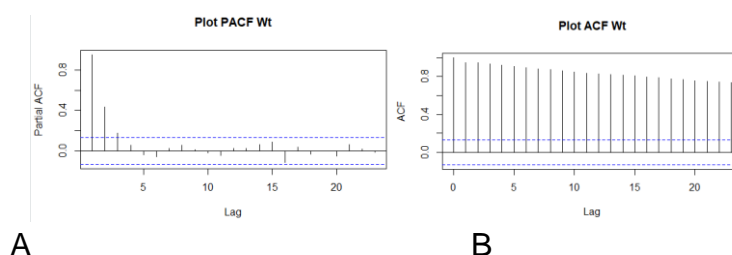
3.2 Kestasioneran Data

Data jumlah positif Covid-19 di DKI Jakarta perlu dilakukan pemeriksaan kestasioneran ragam dan rataan.



Gambar 2 Plot Box-Cox jumlah positif Covid-19 di DKI Jakarta yang belum stasioner (A) dan sudah stasioner (B) dalam ragam

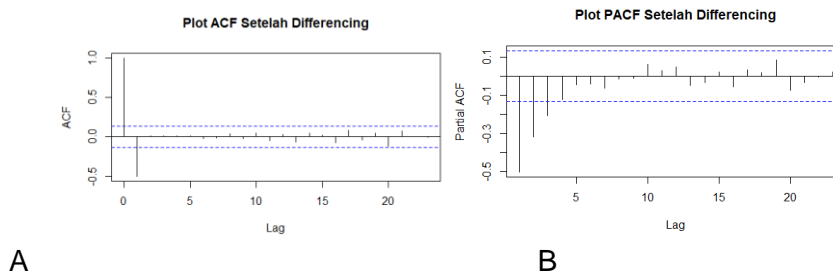
Gambar 2 menunjukkan hasil pemeriksaan kestasioneran dalam ragam. Gambar 2(A) menghasilkan *rounded value* (λ) sebesar 0,0223, yang menunjukkan data tersebut belum stasioner dalam ragam karena nilai λ yang dihasilkan tidak mendekati satu atau selang nilai λ tidak mencakup nilai satu, maka dilakukan transformasi Box-Cox sesuai dengan nilai λ yang diperoleh. Setelah dilakukan transformasi Box-Cox seperti pada Gambar 2(B), λ yang dihasilkan mendekati satu atau selang nilai λ mencakup nilai satu, artinya data sudah stasioner dalam ragam. Data yang sudah stasioner dalam ragam akan dilakukan uji kestasioneran rataan. W_t merupakan data hasil transformasi. Gambar 3 menunjukkan ketidakstasioneran data dalam rataan terlihat dari plot ACF yang memiliki pola menurun secara perlahan.



Gambar 3 Plot ACF (A) dan PACF (B) data jumlah positif Covid-19 di DKI Jakarta yang belum stasioner dalam rataan

Pemeriksaan kestasioneran data dalam rataan menggunakan uji ADF menghasilkan *p-value* sebesar 0,9232. Hasil tersebut menunjukkan bahwa data belum stasioner dalam rataan karena dihasilkan *p-value* yang lebih besar dari nilai taraf nyata 5%. Pembedaan dilakukan satu kali pada data hasil transformasi, kemudian dilakukan pemeriksaan kestasioneran kembali dengan melihat plot ACF.

Montgomery *et al.* (2015), jika plot ACF menurun secara cepat mendekati nol setelah *lag* satu atau dua maka data dalam keadaan stasioner dalam rata-rata. Gambar 4 menunjukkan bahwa data deret waktu jumlah positif Covid-19 di DKI Jakarta telah stasioner dalam rata-rata ditunjukkan dari plot ACF yang menurun secara cepat mendekati nol setelah *lag* satu.



Gambar 4 Plot ACF (A) dan PACF (B) data jumlah positif Covid-19 di DKI Jakarta yang sudah stasioner dalam rata-rata

Hasil uji ADF setelah proses pembedaan dilakukan menghasilkan *p-value* sebesar 0,01 lebih kecil dari nilai taraf nyata 5%, sehingga dapat disimpulkan bahwa data jumlah positif Covid-19 di DKI Jakarta sudah dalam keadaan stasioner dalam rata-rata. Data yang sudah stasioner dalam ragam dan rata-rata akan dilakukan identifikasi dan pendugaan parameter model tentatif ARIMA.

3.3 Identifikasi Model dan Pendugaan Parameter

Identifikasi model dapat dilakukan dengan melihat plot ACF dan PACF. Gambar 4 memperlihatkan plot ACF memiliki pola *cuts off* setelah *lag* satu dan PACF memiliki pola *tails off*. Hal ini menunjukkan adanya kemungkinan bahwa jumlah positif Covid-19 di DKI Jakarta dipengaruhi oleh data satu hari sebelumnya. Hasil pendugaan parameter dari kandidat model yang didapat pada tahap identifikasi model disajikan di Tabel 2.

Tabel 2 Nilai dugaan parameter model ARIMA untuk jumlah positif Covid-19 di DKI Jakarta

Model	Parameter	Koefisien parameter	<i>P-value</i>	AIC	BIC
ARIMA(0,1,1)	MA(1)*	-0,6721	<0,0001	66,84	73,610
ARIMA(1,1,0)	AR(1)*	-0,4994	<0,0001	97,67	104,45
ARIMA(2,1,0)	AR(1)*	-0,6550	<0,0001	77,29	87,460
	AR(2)*	-0,3124	<0,0001		
ARIMA(3,1,0)	AR(1)*	-0,7158	<0,0001	70,84	84,390
	AR(2)*	-0,4406	<0,0001		
	AR(3)*	-0,1950	0,0033		

*Parameter yang signifikan pada taraf nyata 5%

Berdasarkan Tabel 2 semua kandidat model memiliki nilai penduga parameter yang signifikan, karena dihasilkan *p-value* yang lebih kecil dari taraf nyata 5%.

3.4 Diagnostik Model

Model yang signifikan pada tahap identifikasi model, akan dilakukan diagnostik model menggunakan uji Ljung–Box dan Kolmogorov–Smirnov untuk memeriksa kelayakan model. Hasil uji Ljung–Box dari keempat model yang signifikan pada taraf nyata 5% disajikan di Tabel 3.

Tabel 3 Hasil uji Ljung-Box sisaan pada model ARIMA

Model	<i>P-value</i>	<i>Lag</i>
ARIMA (0,1,1)	0,8318	5
	0,9709	10
	0,9928	15
	0,9779	20
	0,9961	25
	0,9953	30
ARIMA (1,1,0)	0,00005*	5
	0,0016*	10
	0,0119*	15
	0,0207*	20
	0,0557	25
	0,0505	30
ARIMA (2,1,0)	0,0435*	5
	0,2047	10
	0,4261	15
	0,4152	20
	0,5908	25
	0,4574	30
ARIMA (3,1,0)	0,5705	5
	0,8479	10
	0,9298	15
	0,9106	20
	0,9765	25
	0,9544	30

*Parameter yang signifikan pada taraf nyata 5%

Tabel 3 menunjukkan bahwa model ARIMA(1,1,0) dan ARIMA(2,1,0) mempunyai autokorelasi pada sisaan karena beberapa *lag* memiliki *p-value* yang lebih kecil dari pada taraf nyata 5%. Sedangkan, model ARIMA(0,1,1) dan ARIMA(3,1,0) memiliki *p-value* yang lebih besar dari taraf nyata 5%, sehingga model tersebut tidak memiliki autokorelasi pada sisaannya. Setelah dilakukan pengujian kebebasan sisaan, selanjutnya dilakukan pemeriksaan kenormalan sisaan yang disajikan di Tabel 4.

Tabel 4 Pemeriksaan kenormalan sisaan

Model	<i>P-value</i>	Keterangan
ARIMA(0,1,1)	0,1280	Sisaan menyebar normal
ARIMA(3,1,0)	0,2496	Sisaan menyebar normal

Tabel 4 menunjukkan model ARIMA(0,1,1) dan ARIMA(3,1,0) memiliki sisaan yang menyebar normal karena *p-value* yang dihasilkan lebih besar dari taraf nyata 5%.

Tabel 5 Nilai AIC dan BIC dari model tentatif awal

Model	AIC	BIC
ARIMA(0,1,1)	66,84	73,61
ARIMA(3,1,0)	70,84	84,39

Berdasarkan Tabel 5, diperoleh model ARIMA(0,1,1) merupakan model yang memiliki nilai AIC dan BIC terkecil dibandingkan model ARIMA(3,1,0). Model terbaik ARIMA(0,1,1) akan dilakukan *overfitting* dengan cara menambah ordo pada model tentatif terbaik. Hasil *overfitting* beberapa kandidat model ARIMA dapat dilihat pada Tabel 6.

Tabel 6 Nilai dugaan parameter model ARIMA untuk jumlah positif Covid-19 di DKI Jakarta hasil *overfitting*

Model	Parameter	Koefisien parameter	P-value	AIC	BIC
ARIMA(0,1,1) <i>with drift</i>	MA(1)*	-0,7069	<0,0001	63,30	73,47
	Konstanta*	0,0139	0,0118		
ARIMA(0,1,2)	MA(1)*	-0,7460	<0,0001	66,94	77,11
	MA(2)	0,0940	0,1672		
ARIMA(1,1,1)	AR(1)	-0,1149	0,2048	67,26	77,43
	MA(1)*	-0,6186	<0,0001		
ARIMA(1,1,2)	AR(1)*	0,8716	<0,0001	65,98	79,54
	MA(1)*	-1,6164	<0,0001		
ARIMA(1,1,3)	MA(2)*	0,6812	<0,0001	67,30	84,25
	AR(1)*	0,9738	<0,0001		
	MA(1)*	-1,7290	<0,0001		
	MA(2)*	0,7951	<0,0001		
	MA(3)	-0,0467	0,5124		

*Parameter yang signifikan pada taraf nyata 5%

Tabel 6 memperlihatkan bahwa model dengan koefisien parameter yang signifikan pada taraf nyata 5%, yaitu ARIMA(0,1,1) *with drift* dan ARIMA(1,1,2), karena memiliki *p-value* yang lebih kecil dari taraf nyata 5%. Selanjutnya, akan dilakukan diagnostik model. Hasil uji Ljung-Box yang dihasilkan dari model yang signifikan tersebut disajikan pada Tabel 7.

Tabel 7 Hasil uji Ljung-Box sisaan pada model ARIMA hasil *overfitting*

Model	P-value	Lag
ARIMA (0,1,1) <i>with drift</i>	0,9077	5
	0,9791	10
	0,9959	15
	0,9854	20
	0,9978	25
	0,9972	30
ARIMA (1,1,2)	0,9996	5
	0,9994	10
	0,9979	15
	0,9877	20
	0,9984	25
	0,9975	30

*Parameter yang signifikan pada taraf nyata 5%

Tabel 7 menunjukkan bahwa model ARIMA(0,1,1) *with drift* dan ARIMA(1,1,2) memiliki nilai-nilai yang tidak signifikan pada taraf nyata 5%, yang berarti kedua model tidak memiliki autokorelasi antar sisaan. Pengujian kenormalan sisaan menunjukkan *p-value* sebesar 0,2274 pada ARIMA(0,1,1) *with drift* dan 0,0946 pada ARIMA(1,1,2) yang lebih besar dari taraf nyata 5%, berarti sisaan menyebar normal. Berdasarkan kandidat model ARIMA yang terbentuk, ARIMA(0,1,1), ARIMA(0,1,1) *with drift*, dan ARIMA(1,1,2) merupakan model yang signifikan pada taraf nyata 5 %, memenuhi asumsi kebebasan sisaan, dan kenormalan sisaan. Model ARIMA yang terbentuk pada tahap identifikasi model akan dibandingkan dengan model hasil *overfitting* berdasarkan pada nilai AIC dan BIC terkecil diantara model-model tersebut. Tabel 8 memperlihatkan AIC dan BIC yang dimiliki model ARIMA(0,1,1), ARIMA(0,1,1) *with drift*, dan ARIMA(1,1,2).

Tabel 8 Perbandingan nilai AIC dan BIC terkecil pada model ARIMA

Model	AIC	BIC
ARIMA(0,1,1)	66,84	73,61
ARIMA(0,1,1) <i>with drift</i>	63,30	73,47
ARIMA(1,1,2)	65,98	79,54

AIC dan BIC yang dihasilkan dari model ARIMA(0,1,1) *with drift* lebih kecil dibandingkan ARIMA(0,1,1) dan ARIMA(1,1,2). Model ARIMA(0,1,1) *with drift* dipilih sebagai model ARIMA terbaik untuk meramalkan jumlah positif Covid-19 di DKI Jakarta. Persamaan model akhir untuk peramalan jumlah positif Covid-19 di DKI Jakarta adalah ARIMA(0,1,1) *with drift*:

$$y_t = (0,00031 + y_{t-1}^{0,0223} + 0,0223\varepsilon_t + 0,01576\varepsilon_{t-1})^{1/0,0223}$$

dengan y_t adalah jumlah positif Covid-19 di DKI Jakarta periode ke- t .

Berdasarkan persamaan diatas model ARIMA(0,1,1) *with drift*, jumlah positif Covid-19 di DKI Jakarta dipengaruhi oleh jumlah kasus pada satu hari sebelumnya. Jumlah kasus satu hari sebelumnya memiliki hubungan yang positif terhadap y_t yang berarti jumlah positif Covid-19 di DKI Jakarta pada periode satu hari sebelumnya memiliki hubungan yang searah dengan jumlah positif Covid-19 di DKI Jakarta pada periode yang akan datang.

3.5 Artificial Neural Networks

Penelitian ini menggunakan nilai masukan berupa *lag* data deret waktu ($y_{t-1}, y_{t-2}, y_{t-3}$), sementara keluaran merupakan data deret waktu ke- t (y_t). Penentuan banyaknya *input layer* dan *hidden layer* terbaik dilakukan dengan *trial and error*. Penelitian ini menggunakan neuron pada *hidden layer* sebanyak 1 hingga 10 neuron. Jaringan terbaik dipilih berdasarkan nilai RMSE terkecil yang diperoleh pada proses *trial and error*. Jaringan dibentuk dengan mengkombinasikan antara tiga neuron pada *input layer* (lapisan masukan), 10 neuron pada *hidden layer* (lapisan tersembunyi), dan satu neuron pada *output layer* (lapisan keluaran). Tabel 9 menunjukkan 30 jaringan yang terbentuk dari metode ANN algoritma *backpropagation*.

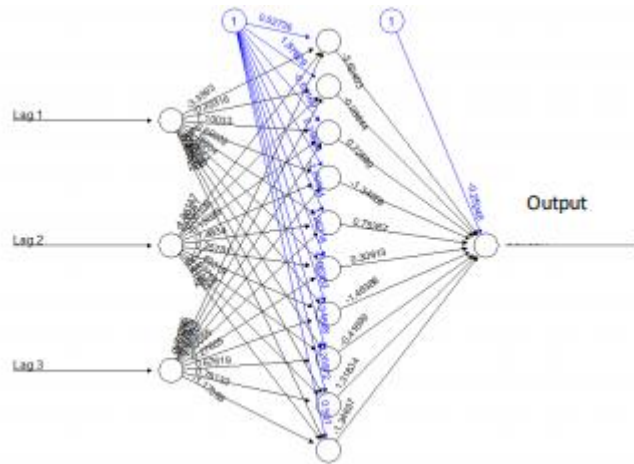
Tabel 9 Jaringan hasil *trial and error* metode ANN

Jaringan	Lag input	Banyak parameter	RMSE
BP(1,1,1)	1	4	99,5199
BP(2,1,1)	1,2	5	97,1253
BP(3,1,1)	1,2,3	6	95,4314
BP(1,2,1)	1	7	99,8178
BP(2,2,1)	1,2	9	97,6610
BP(3,2,1)	1,2,3	11	95,2690
BP(1,3,1)	1	10	100,2443
BP(2,3,1)	1,2	13	97,3730
BP(3,3,1)	1,2,3	16	95,0067
BP(1,4,1)	1	13	100,3212
BP(2,4,1)	1,2	17	97,0065
BP(3,4,1)	1,2,3	21	94,6440
BP(1,5,1)	1	16	99,9412
BP(2,5,1)	1,2	21	97,5355
BP(3,5,1)	1,2,3	26	95,2173
BP(1,6,1)	1	19	100,2444
BP(2,6,1)	1,2	25	98,9347
BP(3,6,1)	1,2,3	31	95,3243
BP(1,7,1)	1	22	100,0494
BP(2,7,1)	1,2	29	98,1456
BP(3,7,1)	1,2,3	36	95,8256
BP(1,8,1)	1	25	100,6499
BP(2,8,1)	1,2	33	98,2619
BP(3,8,1)	1,2,3	41	96,8274
BP(1,9,1)	1	28	102,0901
BP(2,9,1)	1,2	37	97,6007
BP(3,9,1)	1,2,3	46	94,5901
BP(1,10,1)	1	31	100,4364
BP(2,10,1)	1,2	41	97,4893
BP(3,10,1)*	1,2,3	51	93,3417

*model ANN terbaik dari proses *trial and error*

Tabel 9 menunjukkan kemungkinan jaringan yang optimal untuk digunakan pada peramalan jumlah positif Covid-19 di DKI Jakarta. Berdasarkan 30 jaringan yang terbentuk BP(3,10,1) memiliki nilai RMSE terkecil, yaitu 93,3417. Jaringan BP(3,10,1) terpilih sebagai jaringan yang paling optimal dibandingkan dengan 29 jaringan lainnya. Jaringan BP(3,10,1) inilah yang digunakan untuk proses validasi data jumlah positif Covid-19 di DKI Jakarta.

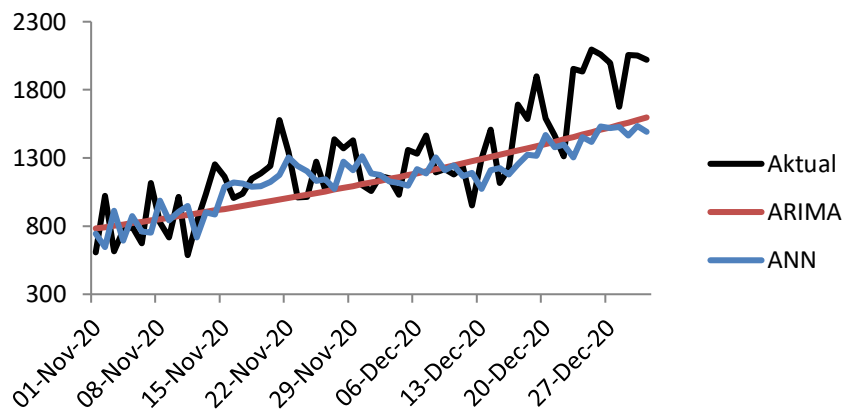
Tabel 9 juga menunjukkan bahwa penambahan neuron pada *input layer* dan *hidden layer* tidak menjamin jaringan yang terbentuk menjadi semakin optimal. Selain itu, penambahan neuron pada *hidden layer* tidak membuat nilai RMSE semakin mengecil. Gambar 5 memperlihatkan arsitektur BP(3,10,1) yang memiliki *input layer* sebanyak tiga, yaitu $y_{t-1}, y_{t-2}, y_{t-3}$. *Input layer* dan *output layer* dihubungkan oleh 10 neuron pada *hidden layer*.



Gambar 5 Arsitektur BP(3,10,1)

3.6 Perbandingan ARIMA dan ANN

Perbandingan model terbaik dilakukan dengan membandingkan nilai RMSE, MAPE, dan hasil peramalan jumlah positif Covid-19 di DKI Jakarta terhadap masing-masing metode terbaik yang dihasilkan pada proses pelatihan, yaitu model ARIMA(0,1,1) *with drift* dan BP(3,10,1).



Gambar 6 Hasil peramalan model ARIMA dan ANN dengan data uji

Gambar 6 memperlihatkan hasil peramalan jumlah positif Covid-19 di DKI Jakarta periode 1 November 2020 sampai 31 Desember 2020. Peramalan dengan metode ANN cenderung mengikuti pola data aktual. ANN mungkin lebih cocok untuk menggambarkan data yang memiliki *trend* taklinear, sedangkan model ARIMA cenderung memberikan hasil peramalan yang terarah (linear) untuk periode peramalan jangka panjang. Kriteria akurasi, yaitu RMSE, MAPE, dan korelasi antara data aktual dengan data hasil ramalan ditampilkan pada Tabel 10.

Tabel 10 Hasil validasi model-model terbaik dengan data uji

Model	Akurasi		
	RMSE	MAPE(%)	Korelasi
ARIMA	268,808	15,748	0,845
ANN	283,365	16,872	0,798

Tabel 10 menunjukkan bahwa metode ARIMA memiliki nilai RMSE, MAPE, dan korelasi yang lebih kecil dibandingkan dengan metode ANN. Berdasarkan selisih nilai RMSE, MAPE, dan korelasi terlihat bahwa metode ARIMA lebih unggul dibandingkan dengan metode ANN dalam kasus peramalan jumlah positif Covid-19 di DKI Jakarta dengan menggunakan data uji. Namun, kedua model tersebut memiliki kemampuan yang baik untuk peramalan jumlah positif Covid-19 di DKI Jakarta karena memperoleh MAPE yang lebih kecil dari 20%. Menurut Chang *et al.* (2007) model dikatakan baik jika memiliki nilai MAPE 10–20%.

Tabel 11 Nilai akurasi pada berbagai panjang periode peramalan

Panjang periode peramalan	ARIMA(0,1,1) with drift		BP(3,10,1)	
	RMSE	MAPE(%)	RMSE	MAPE(%)
1 minggu	178,452	19,802	239,975	24,636
2 minggu	186,761	19,529	231,802	24,086
3 minggu	223,355	19,126	214,207	19,516
4 minggu	226,206	17,670	208,898	18,049
5 minggu	213,974	15,838	200,661	16,652
6 minggu	207,238	14,959	193,903	15,625
7 minggu	217,711	14,911	218,231	15,919
8 minggu	250,511	15,368	258,715	16,359
9 minggu	268,808	15,748	283,365	16,872

Tabel 11 menunjukkan pergerakan nilai RMSE dan MAPE dari model ARIMA(0,1,1) with drift dan BP(3,10,1). Tabel 11 memperlihatkan bahwa ARIMA(0,1,1) with drift memiliki nilai RMSE dan MAPE yang lebih kecil dibandingkan BP(3,10,1) pada panjang periode peramalan satu hingga sembilan minggu. Pola peramalan yang cenderung *trend* yang terarah (linear) pada ARIMA(0,1,1) with drift inilah yang menyebabkan nilai RMSE dan MAPE yang dihasilkan menjadi semakin kecil dibandingkan dengan model ANN. Peramalan dengan panjang periode enam minggu menggunakan model ARIMA(0,1,1) with drift dan BP(3,10,1) menghasilkan nilai RMSE dan MAPE yang paling kecil. Hal tersebut menunjukkan bahwa model ARIMA(0,1,1) with drift dan BP(3,10,1) dapat meramalkan dengan baik untuk panjang periode peramalan enam minggu ke depan pada kasus jumlah positif Covid-19 di DKI Jakarta.

5. Simpulan dan Saran

5.1 Simpulan

Berdasarkan hasil yang diperoleh, untuk data jumlah positif Covid-19 di DKI Jakarta model terbaik ARIMA didapatkan ARIMA(0,1,1) with drift, sedangkan yang didapatkan dari model ANN terbaik adalah BP(3,10,1). Pemodelan ARIMA yang terbentuk menunjukkan bahwa jumlah positif Covid-19 di DKI Jakarta dipengaruhi oleh jumlah kasus satu hari sebelumnya, sedangkan model ANN yang terbentuk menunjukkan bahwa jumlah positif Covid-19 di DKI Jakarta dipengaruhi oleh jumlah kasus satu hari, dua hari, dan tiga hari sebelumnya. Berdasarkan nilai RMSE, MAPE, dan korelasi yang diperoleh dapat ditunjukkan bahwa metode ARIMA meramalkan jumlah positif Covid-19 di DKI Jakarta lebih baik dibandingkan dengan metode ANN.

Peramalan yang dilakukan model ARIMA(0,1,1) *with drift* dan BP(3,10,1) menghasilkan ramalan terbaik pada panjang periode peramalan enam minggu.

5.2 Saran

Penelitian selanjutnya, diharapkan dapat mempertimbangkan peubah yang lainnya, karena penambahan jumlah positif Covid-19 di DKI Jakarta bukan hanya dipengaruhi oleh penambahan kasus pada periode sebelumnya tetapi bisa disebabkan oleh banyak faktor, seperti akses pengetahuan dan peralatan medis, populasi di wilayah tersebut, dan perilaku masyarakat di wilayah tersebut. Selain itu, pada penelitian ini belum adanya vaksinasi, sehingga untuk penelitian selanjutnya bisa ditambahkan variabel mengenai vaksinasi.

Daftar Pustaka

- Adebiyi AA, Aderemi AO, Ayo CK. 2014. Comparison of ARIMA and artificial neural networks models for stock price prediction. *Journal of Applied Mathematics*. 2014:1–7.
- Adeyinka DA, Muhajarine N. 2020. Time series prediction of under-five mortality rates for Nigeria: comparative analysis of artificial neural networks, Holt-Winters exponential smoothing and autoregressive integrated moving average models. *BMC Medical Research Methodology*. 20(292):1–11.
- Chang PC, Wang YW, Liu CH. 2007. The development of weighted evolving fuzzy neural network for PCB sales forecasting. *Expert System with Applications*. 32(2007):86–96.
- Cryer JD, Chan KS. 2008. *Time Series Analysis with Application in R*. Ed ke-2. New York(US):Springer.
- Montgomery DC, Jennings CL, Kulahci M. 2015. *Introduction to Time Series Analysis and Forecasting*. New Jersey(US):John Wiley & Sons.
- Rusdi. 2011. Uji akar-akar unit dalam model runtun waktu autoregresif. *Statistika*. 11(2):67–78.
- Sahai AK, Rath N, Sood V, Singh MP. 2020. ARIMA modeling & forecasting of Covid-19 in top five affected countries. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*. 14(2020):1419–1427.
- Satgas Penanganan Covid-19. [internet, diakses 2020 Desember 18]. Tersedia pada <https://covid19.go.id/peta-sebaran-covid19>.
- Wieczorek M, Silka J, Wozniak M. Neural network powered Covid-19 spread forecasting model. *Chaos, Solitons and Fractals*. 140(2020):1–15.
- Zhang GP. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*. 50(2003):159–175.
- [Kemenkes RI] Kementerian Kesehatan Republik Indonesia. [internet, diakses 2020 Desember 18]. Tersedia pada <https://www.kemkes.go.id/folder/view/full-content/structure-faq.html>.