

Clustering and Forecasting of Covid-19 Data in Indonesia

Clustering dan Forecasting Data Covid-19 Di Indonesia

Diyah Astuti¹, Dyah Yunita hartanti², Susi Tri Nurhayanti³, Herlin Fransiska^{4,*}

Abstract

Indonesia reported its first case of Covid-19 in March 2020, which was suspected to have been infected by a foreigner who visited Indonesia. The distribution of cases that occurred in Indonesia has an uneven frequency considering that Indonesia is an archipelagic country, in the analysis of Covid-19 cases in Indonesia, there are many provinces and some have the same pattern of case characteristics. time series so that forecasting analysis can be used. So that clustering analysis and forecasting of Covid-19 data can be used in Indonesia. The analysis was carried out with 2 stages of analysis, namely clusters using the clustering hierarchy method and forecasting using the ARIMA method. By using 288 data from January 1, 2021 – October 15, 2021, the results show that the daily Covid-19 cases by province in Indonesia can be grouped into 2 clusters, in the forecasting analysis only one province is taken from each cluster used in determining the model, cluster 1 used data from the province of Banten and cluster 2 used data from the province of West Java. By using R software, a model for each cluster is obtained, namely ARIMA(0,1,1) for cluster 1 and ARIMA(2,1,2) for cluster 2. From the forecasting results obtained data until October 30, 2021 shows the number of cases tends to be constant.

Keywords: Covid-19, Clustering, Forecasting, Provinsi, Indonesia

Abstrak

Indonesia melaporkan kasus pertama Covid-19 pada Maret 2020, yang diduga tertular dari orang asing yang berkunjung ke Indonesia. Sebaran kasus yang terjadi di Indonesia memiliki frekuensi tidak merata mengingat Indonesia merupakan negara kepulauan, dalam analisis kasus Covid-19 di Indonesia banyaknya Provinsi yang ada dan sebagian memiliki pola karakteristik kasus yang sama maka analisis dapat dilakukan dengan menggunakan analisis *cluster*, data yang terbentuk berupa data deret waktu sehingga dapat digunakan analisis *forecasting*. Sehingga dapat digunakan analisis *clustering* dan *forecasting* data Covid-19 di Indonesia. Analisis dilakukan dengan 2 tahap analisis yaitu *cluster* menggunakan metode *hierarki clustering* serta *forecasting* menggunakan metode ARIMA. Dengan menggunakan 288 data dari 1 Januari 2021 – 15 Oktober 2021 diperoleh hasil bahwa kasus harian covid-19 berdasarkan provinsi di Indonesia dapat dikelompokkan menjadi 2 *cluster*, dalam analisis *forecasting* hanya diambil satu Provinsi dari masing-masing *cluster* yang digunakan dalam menentukan model, *cluster* 1 digunakan data provinsi Banten dan *cluster* 2 digunakan data Provinsi Jawa Barat. Dengan menggunakan *software* R diperoleh model untuk masing-masing *cluster* yaitu ARIMA(0,1,1) untuk *cluster* 1 dan ARIMA(2,1,2) untuk *cluster* 2. Dari hasil *forecasting* diperoleh data hingga 30 Oktober 2021 menunjukkan jumlah kasus cenderung konstan.

Kata kunci: Covid-19, *Clustering*, *Forecasting*, Provinsi, Indonesia

*Corresponding author

^{1,2,3,4}Program Studi Statistika, FMIPA, Universitas Bengkulu

Email address: diyahastuti8@gmail.com¹, dyahyunitahartanti@gmail.com²,

susitrinur866@gmail.com³, hfransiska@unib.ac.id



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)

1. PENDAHULUAN

Indonesia melaporkan kasus pertama Covid-19 pada maret 2020, yang diduga tertular dari orang asing yang berkunjung ke Indonesia. *Coronavirus disease 2019* (Covid-19) merupakan penyakit menular yang disebabkan oleh coronavirus jenis baru. Coronavirus merupakan *zoonosis* (ditularkan anatara hewan dan manusia). Masa inkubasi Covid-19 rata-rata 5-6 hari, dengan jarak antara 1 dan 14 hari namun dapat mencapai 14 hari. Orang yang terinfeksi dapat langsung menularkan sampai dengan 48 jam sebelum *onset* gejala (presimptomatik) dan sampai dengan 14 hari setelah *onset* gejala. Gejala Covid-19 yang paling umum adalah demam, rasa lelah, dan batuk kering. Beberapa pasien mungkin mengalami rasa nyeri dan sakit hidung tersumbat, pilek, nyeri kepala, konjungtivitis, sakit tenggorokan, diare, hilang penciuman dan pembauan atau ruam kulit.

Berdasarkan laporan analisis data Covid-19 Indonesia yang dilakukan oleh satuan tugas penanganan Covid-19 menuliskan *update* per Juni 2021 jumlah kumulatif kasus di Indonesia sebesar 2.115.305 dimana 5 Provinsi dengan kasus tertinggi yaitu DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur dan Kalimantan Timur. Dengan kenaikan kasus tertinggi berdasarkan Provinsi yaitu DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur dan Banten sedangkan jumlah kasus terendah berdasarkan Provinsi yaitu Maluku Utara, Sulawesi Barat, Gorontalo, Maluku dan Bengkulu. Sebaran kasus yang terjadi di Indonesia memiliki frekuensi tidak merata mengingat Indonesia merupakan negara kepulauan, sebaran yang berbeda-beda dengan frekuensi dari tinggi, sedang, hingga rendah maka wilayah-wilayah yang terpapar kasus Covid-19 dapat dikelompokkan (*cluster*) sehingga memudahkan dalam melakukan analisis.

Analisis perlu dilakukan, untuk memecahkan suatu masalah sehingga dapat ditemukan suatu solusi melihat jumlah kasus yang semakin meningkat dan berdampak tidak hanya dalam aspek kesehatan tetapi sudah merambah ke aspek lainnya seperti ekonomi. Dalam analisis kasus Covid-19 di Indonesia banyaknya Provinsi yang ada dan sebagian memiliki pola karakteristik kasus yang sama maka analisis dapat dilakukan dengan menggunakan analisis *cluster*, data yang terbentuk berupa data deret waktu sehingga dapat digunakan analisis *forecasting*. Sehingga dapat digunakan analisis *clustering* dan *forecasting* data Covid-19 di Indonesia.

Penelitian mengenai analisis *cluster* dan *forecasting* telah banyak dilakukan sebelumnya. Beberapa penelitian terdahulu yang menggunakan analisis *cluster* dan *forecasting* yaitu penelitian yang dilakukan oleh Raditya Novidianto dan Andrea Tri Rian Dani (2020) dalam Jurnal Aplikasi Statistika & Komputasi Statistik Vol. 12, No. 2. Mereka menjelaskan bahwa Gambaran umum pola penyebaran jumlah kasus aktif Covid-19 sepanjang 21 Maret hingga 17 Juli 2020 menunjukkan adanya pola naik turun yang cenderung berbeda di setiap Provinsinya yang berarti setiap daerah memiliki karakteristik masing-masing. Pengelompokan kasus Covid-19 berdasarkan Provinsi di Indonesia menggunakan metode *agglomerative hierarchical clustering* menghasilkan 3 kelompok *cluster*. Penelitian yang dilakukan oleh Hutomo Atman Maulana (2018) dalam Jurnal Matematika, Statistika & Komputasi Vol. 15, No. 1, 50-63. Menjelaskan bahwa analisis deret waktu pada dasarnya digunakan untuk melakukan analisis data yang mempertimbangkan pengaruh waktu. Dari analisis diperoleh hasil bahwa dengan menggunakan model ARIMA hasil prediksi (*forecasting*) lebih mendekati data asli (sebenarnya).

Berdasarkan penjelasan di atas penelitian ini mempunyai tujuan yaitu: mengelompokkan Provinsi di Indonesia berdasarkan kasus covid-19 dan memprediksi kasus Covid-19 di Indonesia untuk masa yang akan datang.

2. METODE PENELITIAN

2.1 Data

Data yang digunakan pada penelitian ini berupa data harian kasus positif Covid-19 di Indonesia setiap provinsi dari 1 Januari 2021 sampai 15 Oktober 2021 dengan jumlah

keseluruhan sebanyak 288 data untuk setiap provinsi, yang diperoleh secara sekunder dari situs resmi satuan tugas Covid-19 pemerintah Indonesia.

2.2 Pre-processing data

Dalam penelitian kali ini sebelum dilakukan analisis data perlu dibagi menjadi 2 bagian yaitu data *training* sebagai data yang akan di analisis serta data *testing* untuk melakukan validasi dari model.

2.3 Analisis Cluster

Langkah-langkah analisis yang dilakukan dalam analisis cluster secara lebih lengkap adalah sebagai berikut:

1) Pengukuran kemiripan

Diperlukan perhitungan jarak dalam menggunakan metode hierarki, pada penelitian ini perhitungan jarak yang digunakan adalah *Dynamic Time Warping* (DTW). Jika diberikan dua data time series $Q = q_1, q_2, q_3, \dots, q_m$, dimana dengan ukuran m dan $C = c_1, c_2, c_3, \dots, c_n$ berukuran n , sehingga terbentuk matriks mxn . Setiap elemen (i, j) merupakan jarak kumulatif dari jarak (i, j) dan nilai minimum dari tiga elemen yang berdekatan dengan elemen (i, j) , dimana $0 < i \leq n$ dan $0 < j \leq m$.

Jarak kumulatif terkecil atau jarak DTW pada (m, n) didefinisikan sebagai berikut:

$$d_{DTW}(Q, C) = \min_{W \in P} \left\{ \sqrt{\sum_{k=1}^K d_{wk}} \right\} \quad (2.1)$$

Dimana P adalah sekumpulan dari semua warping path yang mungkin, dan wk adalah elemen (ij) ke- k pada *warpinh path* dan K adalah panjang dari *warping path*.

2) Metode analisis cluster

Analisis *cluster* (analisis kelompok) adalah metode analisis yang digunakan untuk mengelompokkan obyek-obyek pengamatan menjadi beberapa kelompok sehingga akan diperoleh kelompok dimana obyek-obyek dalam satu kelompok tersebut mempunyai persamaan sedangkan dengan anggota kelompok lain mempunyai perbedaan. Metode hirarki (*hierarchical methods*) adalah salah satu teknik *clustering* dengan cara mengelompokkan dua atau lebih obyek yang mempunyai nilai kemiripan yang paling dekat. Teknik *cluster* hirarki yang digunakan adalah *average linkage*. Metode ini mencari nilai jarak rata-rata antar obyek. Dimulai dengan pencarian dua obyek yang paling dekat jaraknya dan kedua obyek tersebut membentuk cluster yang baru (pertama). Misalkan U dan V , sehingga diperoleh *cluster* (UV) . Selanjutnya jarak antara (UV) dan *cluster* W lainnya dapat diperoleh dengan cara sebagai berikut:

$$d_{(U,V),W} = \frac{[\sum_{i=1}^n \sum_{k=1}^m d_{ik}]}{N_{(UV)}N_W} \quad (2.2)$$

dimana:

d_{ik} : jarak antara objek i dalam kelompok (UV) dan objek k dalam kelompok W

$N_{(UV)}$: banyaknya objek-objek dalam kelompok (UV)

N_W : banyaknya objek-objek dalam kelompok W

3) Uji validitas cluster

Menurut Kaufman dan Rousseeuw (1990), salah satu metode evaluasi yang dapat digunakan untuk melihat kualitas dan kekuatan *cluster* adalah metode koefisien *silhouette*. Persamaan koefisien *silhouette* didefinisikan sebagai berikut:

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (2.3)$$

dengan

S_i : nilai koefisien *silhouette*

a_i : rata-rata jarak objek ke-i dengan semua objek pada satu *cluster* yang sama
 b_i : rata-rata jarak objek ke-i dengan semua objek pada satu *cluster* yang berbeda.

Hasil perhitungan nilai koefisien *silhouette* dapat bervariasi antara -1 hingga 1. Hasil *cluster* dikatakan baik jika nilai koefisien *silhouette* mendekati 1, yang berarti objek ke-i sudah berada dalam *cluster* yang tepat.

2.3 Forecasting

Langkah-langkah analisis yang dilakukan dalam *forecasting* secara lebih lengkap adalah sebagai berikut:

1) Kestasioneran

Stasioneritas merupakan salah satu prasyarat penting dalam analisis *time series* untuk data runtun waktu (*time series*). Data stasioner adalah data yang memiliki *mean*, varians dan autovarians (pada variasi *lag*) tetap sama pada waktu kapan saja data itu dibentuk atau dipakai. Pada penelitian ini digunakan uji akar-akar unit (ADF) sebagai uji stasioneritas.

2) Identifikasi Model

Identifikasi model awal yang dapat dilakukan yaitu dengan melihat pola dari plot ACF dan PACF dari data yang telah stasioner

3) Estimasi parameter model

4) Uji diagnostik

Model yang baik memiliki residual bersifat *random (white noise)*. Analisis residual dengan korelogram melalui ACF dan PACF. Jika koefisien ACF dan PACF secara individual tidak signifikan, residual bersifat *random*. Jika residual tidak *random*, pilih model yang lain. Pengujian Ljung-Box.

H_0 : antar sisaan tidak berkorelasi

H_1 : antar sisaan berkorelasi

$$Q^* = n(n + 2) \sum_{k=1}^K \left(\frac{\hat{r}_{e(k)}^2}{n - k} \right) \quad (2.4)$$

Dimana:

n = banyaknya data sisaan, \hat{e}_t

$\hat{r}_{e(k)}$ = autokorelasi \hat{e}_t dengan \hat{e}_{t-k}

5) Model Terbaik

Pengujian kelayakan model dengan mencari model terbaik. Model terbaik didasarkan kriteria AIC. Seleksi model merupakan suatu tahapan penting untuk memutuskan model yang terbaik., Akaike (1973,1974) memperkenalkan suatu kriteria informasi yang disebut AIC (*Akaike information criterion*).

$$AIC = e^{\frac{2k}{n}} \frac{\sum_{i=1}^n \hat{u}_i^2}{n} \quad (2.5)$$

dengan:

n = jumlah observasi

$e = 2,718281828$

u = residual

6) Forecasting

Menurut (Makridakis et al, 1988:519) peramalan/ *forecasting* merupakan prediksi nilai-nilai sebuah variabel berdasarkan kepada nilai yang diketahui dari variabel tersebut atau variabel yang berhubungan. Box dan Jenkis mempopulerkan metode peramalan data univariat runtun waktu yang terdiri dari 3 tahap dalam memilih model yang cocok untuk melakukan estimasi peramalan. Peramalan dengan menggunakan data runtun waktu secara sederhana sering dilakukan dengan analisis *trend*. Namun agar hasil peramalan

lebih akurat, terutama untuk data runtun waktu yang cukup panjang, maka data runtun waktu dapat dianalisis dengan mempertimbangkan berbagai perilaku data melalui berbagai model antara lain model *autoregressive* (AR) dengan persamaan

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t \quad (2.6)$$

moving average (MA) dengan persamaan

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_p e_{t-q} \quad (2.7)$$

dan *autoregressive moving average* (ARIMA) (Ghozali, 2007).

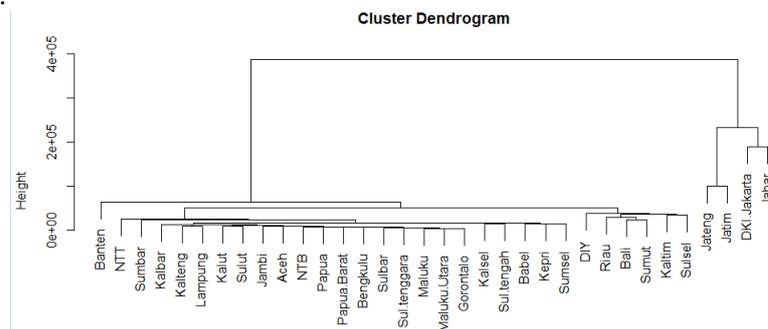
3. HASIL DAN PEMBAHASAN

3.1 Pre-Processing Data

Dari 288 data akan di gunakan 278 data sebagai data *training* dan 10 data sebagai data *testing* dari masing-masing provinsi. Data *testing* digunakan sebagai uji validasi hasil *forecasting* dari model yang terpilih. Metode ARIMA (*Autoregressive Integrated Moving Average*) adalah metode yang digunakan untuk peramalan jangka pendek, oleh karena itu hanya digunakan data *testing* sebesar 3% dan untuk meminimaslisir bias yang terjadi, jika diambil terlalu banyak data *testing* dikhawatirkan model yang terbentuk berdasarkan data *training* kurang mendekati data asli.

3.2 Analisis Cluster

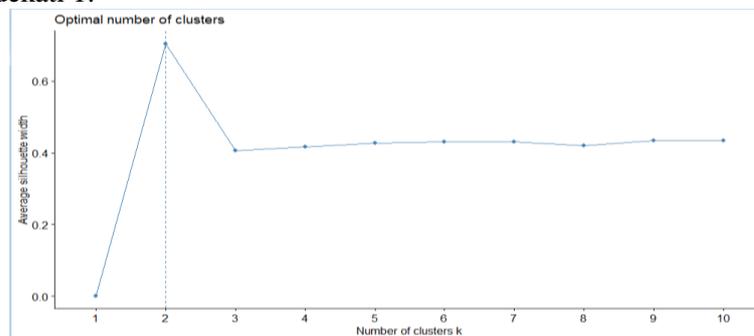
Pembentukan *cluster* dengan menggunakan jarak DTW dalam proses pengelompokkan menggunakan algoritma *average linkage* menghasilkan *cluster* yang jika dilihat secara garis besar berdasarkan hasil *cluster* dendrogram yang dapat dilihat pada gambar 3.1 terdapat 2 *cluster* yang terbentuk.



Gambar 3.1. Dendrogram jarak DTW

1. Uji Validitas Cluster

Pengujian validitas pada penelitian ini bertujuan untuk menghasilkan proses *clustering* yang optimal, artinya proses pembentukan *cluster* dengan algoritma *average linkage* didasarkan pada jumlah *cluster* yang optimal. Uji validitas *cluster* yang digunakan pada penelitian ini adalah metode koefisien *silhouette*. Jumlah *cluster* dikatakan optimal jika nilai koefisien *silhouette* mendekati 1.

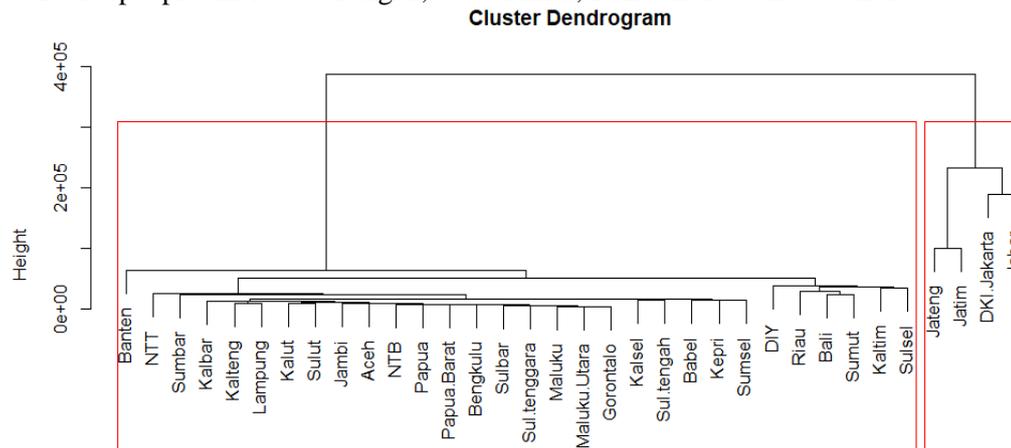


Gambar 3.2. Diagram koefisien *silhouette*

Berdasarkan Gambar 3.2, dapat diketahui bahwa nilai c-optimal dalam mengelompokkan data kasus positif covid-19 di Indonesia adalah 2 cluster.

2. Identifikasi Anggota Cluster

Berdasarkan uji validasi jumlah cluster yang terbentuk dengan menggunakan jarak DTW dan algoritma *average linkage* sebanyak 2 cluster yang dapat dilihat pada gambar 3.3, dengan cluster 1 terdapat Provinsi Banten, NTT, Sumatera Barat, Kalimantan Barat, Kalimantan Tengah, Lampung, Kalimantan Utara, Sulawesi Utara, Jambi, Aceh, NTB, Papua, Papua Barat, Bengkulu, Sulawesi Barat, Sulawesi Tenggara, Maluku, Maluku Utara, Gorontalo, Kalimantan Selatan, Sulawesi Tengah, Bangka Belitung, Kepulauan Riau, Sumatera Selatan. Sedangkan cluster 2 terdapat provinsi Jawa Tengah, Jawa Timur, DKI Jakarta dan Jawa Barat.



Gambar 3.3. Dendrogram jarak DTW 2 cluster

Dalam analisis selanjutnya hanya di ambil satu provinsi yang akan digunakan dalam menentukan model dari masing-masing cluster. Untuk cluster 1 akan di gunakan data provinsi banten dan cluster 2 akan digunakan data provinsi Jawa barat.

3.3 Analisis Menggunakan Metode ARIMA

1. Uji Stasioneritas

Pengujian stasioneritas ini bertujuan untuk melihat kestasioneran data setiap cluster agar dapat dilanjutkan pada analisis selanjutnya.

Tabel 3.1. Hasil R uji stasioneritas

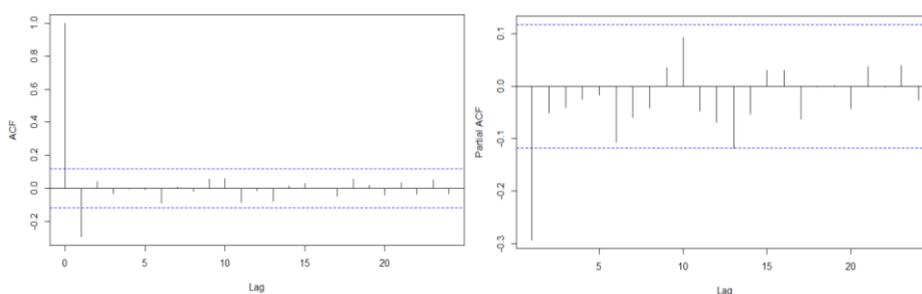
Cluster	Uji ADF	
	Data Awal	Setelah Differencing
Cluster 1	0,2633	0,01
Cluster 2	0,661	0,01

Dengan menggunakan uji ADF diperoleh hasil bahwa data pada setiap cluster tidak stasioner, kemudian dilakukan *differencing* 1 kali dan dengan $\alpha = 5\%$ diperoleh hasil data telah stasioner.

2. Identifikasi Model ARIMA

Identifikasi model dilakukan dengan melihat *plot* ACF PACF dari masing-masing cluster.

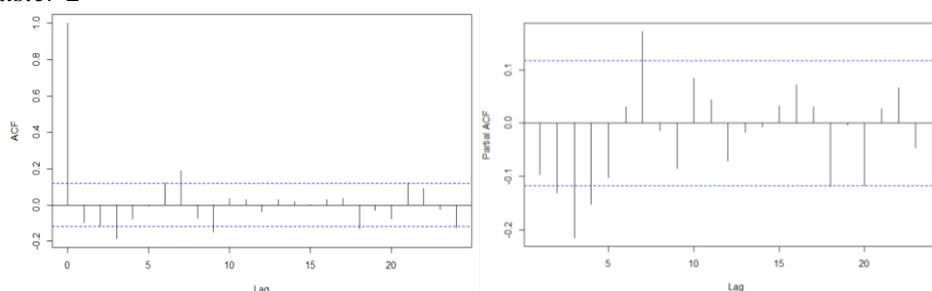
a. Cluster 1



Gambar 3.4. Plot ACF PCF cluster 1

Berdasarkan bentuk plot AFC yang *cut off* pada lag 1 dan plot PACF *dist down* yang terlihat pada gambar 3.4 maka model awal yang teridentifikasi yaitu ARI(1,1), IMA(1,1) dan ARIMA(1,1,1).

b. Cluster 2



Gambar 3.5. Plot ACF PCF cluster 2

Berdasarkan bentuk plot AFC yang *dist down* dan juga plot PACF yang *dist down* yang terlihat pada gambar 3.5 maka model awal yang teridentifikasi yaitu ARI(1,1), IMA(1,1) dan ARIMA(1,1,1).

3. Estimasi Parameter Model

Berdasarkan hasil uji signifikansi parameter dari dugaan model awal yang didapatkan serta menggunakan metode *underfitting* (menurunkan nilai parameter model) dan *overfitting* (menaikkan nilai parameter model) diperoleh hasil model-model yang signifikan pada $\alpha = 5\%$ untuk setiap *cluster*, karena memiliki parameter yang signifikan terhadap model sehingga dapat dijadikan sebagai model sementara. Hasil uji signifikansi parameter untuk setiap *cluster* adalah:

a. Cluster 1

Hasil uji signifikansi parameter *cluster 1* ditunjukkan pada tabel 3.2 berikut:

Tabel 3.2. Estimasi model *cluster 1*

Model	Parameter	Koefisien	Prob	AIC
ARIMA(1,1,0)	$\hat{\phi}_1$	-0.291680	3.654e-07	4026.95
ARIMA(0,1,1)	$\hat{\theta}_1$	-0.306643	9.989e-08	4026.05

Berdasarkan tabel di atas terlihat bahwa nilai probabilitas masing-masing model kurang dari $\alpha = 5\%$ yang berarti kedua model *cluster 1* ARIMA(0,1,1) dan ARIMA(1,1,0) signifikan.

b. Cluster 2

Hasil uji signifikansi parameter *cluster 2* ditunjukkan pada tabel 3.3 berikut:

Tabel 3.3. Estimasi model *cluster 2*

Model	Parameter	Koefisien	Prob	aic
ARIMA(3,1,0)	$\hat{\phi}_1$	-0.137958	0.0185609	4578.34
	$\hat{\phi}_2$	-0.154284	0.0081461	
	$\hat{\phi}_3$	-0.213522	0.0002531	

	$\hat{\phi}_3$			
ARIMA(4,1,0)	$\hat{\phi}_1$	-0.170842	0.003975	4573.89
	$\hat{\phi}_2$	-0.177886	0.002325	
	$\hat{\phi}_3$	-0.234462	5.676e-05	
	$\hat{\phi}_4$	-0.150741	0.010571	
ARIMA(0,1,2)	$\hat{\theta}_1$	-0.205356	0.0003891	4581.34
	$\hat{\theta}_2$	-0.206667	0.0001770	
ARIMA(0,1,3)	$\hat{\theta}_1$	-0.164113	0.005638	4576.1
	$\hat{\theta}_2$	-0.150876	0.009076	
	$\hat{\theta}_3$	-0.153266	0.006259	
ARIMA(2,1,2)	$\hat{\phi}_1$	1.09809	< 2.2e-16	4571.26
	$\hat{\phi}_2$	-0.68288	1.068e-09	
	$\hat{\theta}_1$	-1.24381	< 2.2e-16	
	$\hat{\theta}_2$	0.62937	1.406e-07	

Berdasarkan tabel di atas terlihat bahwa nilai probabilitas masing-masing model kurang dari $\alpha = 5\%$ yang berarti kedua model *cluster 2* ARIMA(3,1,0), ARIMA(4,1,0), ARIMA(0,1,2), ARIMA(0,1,3) dan ARIMA(2,1,2) signifikan.

4. Uji Diagnostik

Pemeriksaan diagnostik asumsi *white noise* dilakukan dengan menggunakan uji Ljung-Box untuk menguji apakah terdapat korelasi pada residual antar *lag* atau tidak dari setiap model.

a. Cluster 1

Hasil uji asumsi *white noise cluster 1* ditunjukkan pada tabel 3.4 berikut:

Tabel 4.4. Uji *white noise cluster 1*

Model	Lag	p-value	Keputusan
ARIMA(1,1,0)	12	0.4801	H ₀ diterima
	24	0.8384	H ₀ diterima
ARIMA(0,1,1)	12	0.5582	H ₀ diterima
	24	0.8792	H ₀ diterima

Berdasarkan uji independensi residual pada *cluster 1* didapat kesimpulan bahwa dengan $\alpha = 5\%$ semua model menerima H₀ artinya tidak terdapat korelasi pada residual antar *lag* (residual independen) sehingga asumsi *white noise* terpenuhi.

b. Cluster 2

Hasil uji asumsi *white noise cluster 2* ditunjukkan pada tabel 3.5 berikut:

Tabel 3.5. Uji *white noise cluster 2*

Model	Lag	p-value	keputusan
ARIMA(3,1,0)	12	0.02867	H ₀ ditolak
	24	0.02645	H ₀ ditolak
ARIMA(4,1,0)	12	0.0883	H ₀ diterima
	24	0.1184	H ₀ diterima
ARIMA(0,1,2)	12	0.004311	H ₀ ditolak
	24	0.002609	H ₀ ditolak
ARIMA(0,1,3)	12	0.05863	H ₀ diterima
	24	0.04601	H ₀ ditolak
ARIMA(2,1,2)	12	0.3708	H ₀ diterima
	24	0.4063	H ₀ diterima

Berdasarkan uji independensi residual pada *cluster 2* didapat kesimpulan bahwa dengan $\alpha = 5\%$ hanya 2 model yang menerima H₀ artinya tidak terdapat korelasi pada residual

antar *lag* (residual independen) sehingga asumsi *white noise* terpenuhi, sehingga model yang dapat digunakan untuk analisis selanjutnya yaitu ARIMA(4,1,0) dan ARIMA(2,1,2).

5. Model Terbaik

Model yang terpilih sebagai model terbaik yaitu model yang memenuhi uji asumsi tidak ada autokorelasi (*white noise*) serta nilai AIC terkecil. Sehingga dapat disimpulkan model yang terpilih untuk *cluster* 1 ARIMA(0,1,1) dan *cluster* 2 ARIMA(2,12).

6. Validasi Model

Dengan menggunakan model terbaik dari masing-masing *cluster* maka dapat dilakukan *forecasting* untuk melakukan uji validasi model untuk melihat kesesuaian hasil model dengan data asli.

a. *Cluster* 1

Hasil uji validasi *cluster* 1 ditunjukkan pada tabel 3.6 berikut:

Tabel 3.6. Uji validasi *cluster* 1

Obs	Data riil	Data Forecast	
279	33	37	0,12121
280	41	37	0,09756
281	39	37	0,05128
282	37	37	0
283	13	37	1,84615
284	17	37	1,17647
285	24	37	0,54167
286	34	37	0,08824
287	34	37	0,08824
288	26	37	0,42308
MAPE			0,44339

Berdasarkan hasil uji validasi di atas diperoleh nilai MAPE untuk *cluster* 1 sebesar 0,44339 atau 44,339% dengan model yaitu ARIMA(0,1,1).

b. *Cluster* 2

Hasil uji validasi *cluster* 2 ditunjukkan pada tabel 3.7 berikut:

Tabel 3.7. Uji validasi *cluster* 2

Obs	Data riil	Data Forecast	
279	104	115	0,105769
280	106	114	0,075472
281	107	106	0,009346
282	117	99	0,153846
283	63	95	0,507937
284	69	97	0,405797
285	127	101	0,204724
286	93	104	0,11828
287	102	105	0,029412
288	85	103	0,211765
MAPE			0,18223

Berdasarkan hasil uji validasi di atas diperoleh nilai MAPE untuk *cluster* 2 sebesar 0,18223 atau 18,223% dengan model yaitu ARIMA(2,1,2).

3.4 Forecasting

Tahap terakhir pada metode ARIMA yaitu melakukan peramalan (*forecasting*) dengan menggunakan model terbaik yang diperoleh pada tahap sebelumnya. Peramalan ini dilakukan untuk mengetahui jumlah kasus positif Covid-19 untuk 15 hari kedepan yang terlihat pada tabel 3.8 berikut.

Tabel 3.8. *Forecasting*

Tanggal	Jumlah Kasus	
	<i>Cluster</i> 1	<i>Cluster</i> 2
16/10/2021	28	84
17/10/2021	28	86
18/10/2021	28	89
19/10/2021	28	91
20/10/2021	28	91
21/10/2021	28	89
22/10/2021	28	88
23/10/2021	28	88
24/10/2021	28	88
25/10/2021	28	88
26/10/2021	28	89
27/10/2021	28	89
28/10/2021	28	89
29/10/2021	28	89
30/10/2021	28	88

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan, maka dapat ditarik kesimpulan sebagai berikut:

1. Data kasus Covid-19 di Indonesia dari 1 Januari 2021 – 15 Oktober 2021 dapat dikelompokkan menjadi 2 *cluster*. *Cluster* pertama terdapat Provinsi Banten, NTT, Sumatera Barat, Kalimantan Barat, Kalimantan Tengah, Lampung, Kalimantan Utara, Sulawesi Utara, Jambi, Aceh, NTB, Papua, Papua Barat, Bengkulu, Sulawesi Barat, Sulawesi Tenggara, Maluku, Maluku Utara, Gorontalo, Kalimantan Selatan, Sulawesi Tengah, Bangka Belitung, Kepulauan Riau, Sumatera Selatan. Sedangkan *cluster* 2 kedua terdapat provinsi Jawa Tengah, Jawa Timur, DKI Jakarta dan Jawa Barat.
2. Model yang diperoleh dari masing-masing *cluster* yaitu ARIMA(0,1,1) untuk *cluster* pertama dan ARIMA(2,1,2) untuk *cluster* kedua.
3. Hasil *forecasting* yang diperoleh selama 15 hari dari 16 Oktober 2021 sampai 30 Oktober 2021 menunjukkan bahwa jumlah kasus Covid-19 untuk masing-masing *cluster* cenderung konstan.

DAFTAR PUSTAKA

- [1] Dani, Andrea TR., Sri W. & Nanda AR., 2020. Pengelompokan Data Runtun Waktu Menggunakan Analisis Cluster. *Jurnal Eksponensial*, Vol. 11, No. 1.

- [2] Giusti, R & Batista, G.E.A.P.A. 2013, 'An empirical comparison of dissimilarity measures for time series classification' In Proceedings – 2013, *Brazilian Conference on Intelligent Systems, BRACIS 2013, IEEE Computer Society*, IEEE, pp. 82 – 88.
- [3] Hariadi, W., & Sulantari., 2021. Application of ARIMA Model for Forecasting Additional Positive Cases of Covid-19 in Jember Regency. *International Journal Of Statistics And Data Science*. Vol 1, Issue 1, pp. 20-27.
- [4] Hartati, 2017. *Penggunaan Metode Arima Dalam Meramal Pergerakan Inflasi*. <https://jurnal.ut.ac.id/index.php/jmst/article/download/163/146/>. [12 Agustus 2021]
- [5] Kartikasari, MD., 2021. Forecasting COVID-19 Cases in Indonesia using Hybrid Double Exponential Smoothing. *International Journal Of Statistics And Data Science*. Vol 1, Issue 2, pp. 53-57.
- [6] Kementerian Kesehatan RI, 2020. Pedoman Pencegahan Dan Pengendalian Coronavirus Disease (Covid-19). Kementerian Kesehatan RI. Jakarta.
- [7] Maulana, HA., 2018. Pemodelan deret waktu dan peramalan curah hujan pada dua belas stasiun di bogor. *Jurnal matematika statistika dan komputasi*. Vol. 15, No. 1.
- [8] Muhidin, A., 2017. Analisa Metode Hierarchical Clustering Dan K-Mean Dengan Model Lrfmp Pada Segmentasi Pelanggan. *Jurnal Teknologi Pelita Bangsa*, 7(1):81-88.
- [9] Niennattrakul, V & Ratanamahatana., 2007. On Clustering Multimedia Time and Dynamic Time Warping. *Computer Society*, hh. 733-738.
- [10] Novidianto, Raditya., & Andrea T., 2020. *Analisis Klaster Kasus Aktif Covid-19 Menurut Provinsi Di Indonesia Berdasarkan Data Deret Waktu*. <https://jurnal.stis.ac.id/index.php/jurnalasks/article/view/280/90>. [15 Agustus 2021]
- [11] Rencher & Alvin., 2002, *Methods of Multivariate Analysis Second Edition*, Wiley-Interscience, United States of America.
- [12] Satuan Tugas Penanganan Covid-19., 2021. *Analisis Data Covid-19 Di Indonesia*. <https://covid19.go.id/p/berita/analisis-data-covid-19-indonesia-update-27-jun-i-2021>. [15 Agustus 2021]
- [13] Yunita, Tasna., 2019. Peramalan Jumlah Penggunaan Kuota Internet Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA). *Journal of Mathematics: Theory and Applications*, Vol. 1, No. 2.
- [14] Zhang, Z., Diya L., Zhe Z., & Nicholas D, 2021. *A Time-Series Clustering Algorithm for Analyzing the Changes of Mobility Pattern Caused by COVID-19*. Proceedings of the 1st ACM SIGSPATIAL International Workshop On Animal Movement Ecology And Human Mobility. 13-17, 2021, https://scholar.google.com/scholar?hl=id&assdt=0%2C5&q=international+journal+clustering+time+series+covid+19&btnG=#d=gsqabs&u=%23p%3D_V_uW975wcxsJ. [02 Desember 2021]