

Prediksi Penyebaran COVID-19 Harian di Jawa Timur Menggunakan Model *Vector Autoregressive Moving Average* (VARMA)

Zakiya Ainur Rohman¹, Aniq Atiqi Rohmawati², Indwiarti³

^{1,2,3} Universitas Telkom, Bandung

¹zakiyaainurrohman@students.telkomuniversity.ac.id, ²aniqatiqi@telkomuniversity.ac.id,

³indwiarti@telkomuniversity.ac.id

Abstrak

COVID-19 sampai saat ini belum mereda di seluruh negara, termasuk Indonesia. Akibatnya seluruh aktivitas masyarakat dibatasi yang mengakibatkan menurunnya perekonomian di Indonesia. Oleh karena itu, Peneliti melakukan penelitian untuk melakukan peramalan penyebaran kasus baru COVID-19 di Jawa Timur dengan menggunakan data kasus baru dan sembuh baru. Penelitian ini bertujuan untuk meramalkan jumlah kasus baru COVID-19 harian di Jawa Timur. Dalam penelitian ini, model yang digunakan untuk meramalkan kasus baru COVID-19 adalah model *Vector Autoregressive Moving Average* (VARMA). Hasil dari penelitian ini menunjukkan bahwa pada model VARMA(5,1) memiliki nilai MAPE sebesar 12.18%. Semakin kecil nilai MAPE, maka menunjukkan model atau hasil peramalan yang bagus.

Kata kunci : COVID-19, kasus baru, peramalan, VARMA, jawa timur, MAPE

Abstract

COVID-19 has not yet subsided in all countries, including Indonesia. As a result, all community activities are limited which results in a decline in the economy in Indonesia. Therefore, the researcher conducted a study to forecast the daily spread of new COVID-19 cases in East Java by using data on new cases and newly recovered cases. This study aims to predict the daily number of new COVID-19 cases in East Java. In this study, the model used to predict new cases of COVID-19 is the *Vector Autoregressive Moving Average* (VARMA) model. The results of this study indicate that the VARMA(5,1) model has a MAPE value of 12.18%. The smaller the MAPE value, the better the model or forecasting results.

Keywords: COVID-19, new case, forecast, VARMA, east java, MAPE

1. Pendahuluan

Latar Belakang

Dunia saat ini sedang dilanda pandemi yang sangat serius, wabah pandemi COVID-19 [1]. COVID-19 ini pertama kali dilaporkan di Wuhan, China pada 31 Desember 2019 [2]. Hampir semua daerah di Indonesia terjangkit virus ini. Di Indonesia, pandemi COVID-19 ditemukan pada awal Maret, lebih tepatnya pada 2 Maret 2020 [1]. Saat itu dua orang dinyatakan positif COVID-19, sehingga diberi nama Kasus 1 dan 2. [1]. Dari Kasus tersebut, hampir setiap hari mengalami lonjakan yang signifikan [1]. Akibatnya, pengadaan layanan, fasilitas, dan tenaga medis tidak dapat diprediksi [1].

Jawa Timur sendiri merupakan peringkat 1 provinsi dengan jumlah kasus positif COVID-19 di Indonesia pada tanggal 1 Juli 2020 dengan total kasus 12321 dan jumlah kasus aktif mencapai 7196 jiwa. Pada tanggal 26 November 2020 total kasus COVID-19 di Jawa Timur mencapai 60190 jiwa. Dari total kasus tersebut terdapat kasus positif (aktif) sebanyak 2784 jiwa dan tingkat kesembuhan mengalami kenaikan yaitu sebanyak 53131 jiwa. Dari kasus tersebut, maka Peneliti melakukan penelitian peramalan kasus baru COVID-19 di Jawa Timur.

Model VARMA adalah model yang menggambarkan hubungan antara pengamatan sebelumnya terhadap variabel yang sama dan hubungan dengan pengamatan variabel lain selama periode sebelumnya [3]. Seperti, penelitian (Ulya, 2019) yang melakukan Peramalan Harga Saham Penutupan, pada penelitian tersebut menggunakan Metode *Vector Autoregressive Moving Average* (VARMA). Penelitian tersebut menggunakan dua data yaitu data harga saham penutupan atau *closing price* dari PT. Kimia Farma Tbk dan PT. Indo Farma Tbk. Peramalan tersebut menggunakan model VARMA(1,1) yang menghasilkan nilai MAPE kurang dari 10% [3]. Pada penelitian lain, (Ciptaningtyas, 2017) yang melakukan Penerapan Model *Seasonal Vector Autoregressive Moving Average*. Penelitian tersebut menggunakan data deret waktu curah hujan dan suhu rata-rata di Stasiun Meteorologi Juanda, Sidoarjo. Dari data tersebut menghasilkan peramalan curah hujan sebesar 41.14% dan peramalan suhu rata-rata sebesar 45.4% [4]. Selain itu, terdapat penelitian lain (Wiguna et al., 2020) yang melakukan Prediksi Penyebaran COVID-19 di Jakarta, pada penelitian tersebut menggunakan Metode *Autoregressive Integrated Moving Average* (ARIMA). Penelitian tersebut menggunakan data COVID-19 di Jakarta yaitu data meninggal harian, data positif harian dan data sembuh harian mulai dari 1 Maret 2020 hingga 5 Juli 2020 yang digunakan untuk memprediksi kasus positif 14 hari ke depan. Dari hasil prediksi tersebut menghasilkan nilai MAPE sebesar 20.97% dan RMSE sebesar 42.415 [5]. Dalam penelitian ini dilakukan peramalan kasus baru COVID-19

menggunakan model VARMA pada data COVID-19 di Jawa Timur dengan menggunakan data kasus baru yang melibatkan sembuh baru.

Topik dan Batasannya

Topik dan batasan masalah dalam penelitian ini yaitu data yang digunakan yaitu dataset COVID-19 di Jawa Timur dari 1 Juli 2020 hingga 28 Februari 2021 dengan melibatkan variabel kasus baru dan sembuh baru yang diambil dari website resmi <https://kawalcovid19.id/>.

Tujuan

Tujuan dalam penelitian ini adalah untuk memodelkan kasus baru COVID-19 di Jawa Timur menggunakan model VARMA. Tujuan yang kedua yaitu untuk menganalisis performansi model VARMA digunakan nilai MAPE.

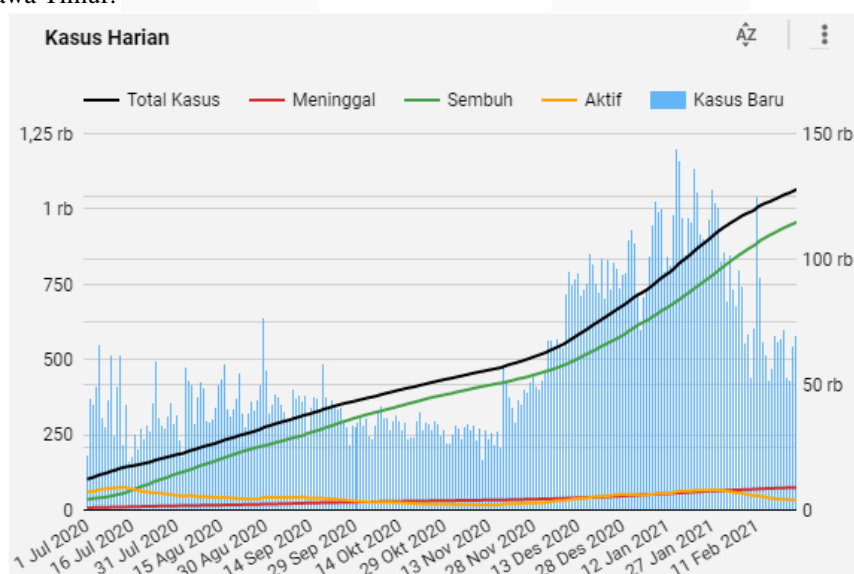
Organisasi Tulisan

Selanjutnya, setelah pendahuluan ini yaitu Bab 2 tentang studi terkait yang membahas COVID-19, kestasioneran data, model VARMA, *Pearson Correlation*, order pada model VARMA, MAPE. Setelah itu, Bab 3 tentang sistem yang dibangun yang membahas langkah-langkah meramalkan kasus baru COVID-19 di Jawa Timur dengan model VARMA. Pada Bab 4 tentang evaluasi yang membahas hasil dan analisis pengujian dan pada Bab 5 mengenai kesimpulan.

2. Studi Terkait

2.1. COVID-19

COVID-19 atau virus corona adalah penyakit yang disebabkan virus baru yaitu SARS-CoV2 [2]. Pada awal bulan Maret sampai dengan sekarang, virus corona ini masih menjadi penyakit yang serius. Di Jawa Timur sendiri merupakan provinsi dengan kasus COVID-19 terbanyak setelah Jakarta [6]. Berikut adalah grafik kasus harian COVID-19 di Jawa Timur.



Gambar 1. Grafik kasus harian COVID-19 di Jawa Timur mulai 1 Juli 2020 s/d 28 Februari 2021

Sumber : <https://kawalcovid19.id/>

Dari Gambar 1, terdapat grafik kasus harian COVID-19 di Jawa Timur pada bulan Juli 2020 sampai dengan bulan Februari 2021. Pada grafik total kasus dan sembuh dari awal bulan Juli 2020 terus mengalami kenaikan. Pada bulan Desember 2020 jumlah kasus baru mengalami kenaikan yang serius dan puncaknya pada tanggal 15 Januari 2021 dengan jumlah kasus baru 1198 orang.

2.2. Kestasioneran Data

Proses stasioner adalah bahwa rata-rata dan variansi dalam keadaan konstan dari waktu ke waktu [7]. Ketika tidak ada pertumbuhan atau penurunan pola yang signifikan, maka data tersebut dapat dianggap stasioner [3]. Data stasioner dapat dibagi menjadi dua kategori, yaitu stasioner lemah dan stasioner kuat. Pada proses stokastik Z_t jika distribusi gabungan dari $Z_{t_1}, Z_{t_2}, \dots, Z_{t_n}$ sama dengan distribusi gabungan dari $Z_{t_1-k}, Z_{t_2-k}, \dots, Z_{t_n-k}$, untuk semua t_1, t_2, \dots, t_n dan semua *lag* waktu k , maka distribusi tersebut dapat dikatakan stasioner kuat (*strictly stationary*) [8].

Stasioner data dalam rata-rata bahwa variansi data di sekitar *mean* adalah konstan dan tidak tergantung pada waktu dan variansi. Untuk mengatasi adanya data tidak stasioner dalam rata-rata dapat menggunakan *differencing* [3]. Berikut persamaan dari *differencing*:

$$W_t = y_t - y_{t-1} \quad (1)$$

dimana:

- W_t : data *differencing* pada waktu ke- t
 y_t : data asli pada waktu ke- t , dengan $t = 1, 2, 3, \dots, n$
 y_{t-1} : data asli tepat pada $t - 1$

Uji Kestasioneran dengan *Augmented Dickey-Fuller* (ADF)

Tujuan dari uji stasioneran data adalah untuk menentukan apakah data yang digunakan memenuhi asumsi tentang stabilitas rata-rata dan variansi. Dalam menentukan apakah data sudah stasioner atau belum menggunakan uji ADF dengan nilai signifikansi (α) = 5% [3]. Berikut adalah prosedur pengujian ADF [9]:

- Hipotesis
 H_0 : data tidak stasioner
 H_1 : data stasioner
- Statistik uji ADF

$$t = \frac{\delta}{SE(\delta)} \quad (2)$$

dimana δ adalah parameter AR, $SE(\delta)$ adalah standar *error* dari δ dan t adalah tes statistik.

- Kriteria penolakan ADF
Tolak H_0 jika $|t_\delta| \geq |t_{(n-1, \alpha)}|$ atau $p - value \leq \alpha$

2.3. Model *Vector Autoregressive Moving Average* (VARMA)

Analisis *Time Series* adalah serangkaian pengamatan data yang terjadi dalam suatu dimensi waktu secara terus-menerus pada interval waktu yang teratur [3]. Data yang digunakan adalah data yang berhubungan dengan waktu, dan data yang digunakan dalam deret waktu ini adalah tipe data sekarang dan masa lalu. Analisis deret waktu mencakup analisis yang melibatkan satu variabel atau *univariat*, dan analisis yang melibatkan banyak variabel, yang disebut *multivariat*, digunakan untuk memodelkan dan menggambarkan interaksi dan perilaku antar variabel [3].

Model VARMA merupakan perpaduan dari dua model yaitu model VAR dan model VMA [1]. *Vector Autoregressive* (VAR) merupakan salah satu model prediktif untuk data *time series* multivariat dan sering digunakan karena lebih sederhana dan fleksibel dibandingkan model lainnya [3]. VAR adalah perpanjangan dari model autoregresif univariat dari data multivariabel periodik [10]. Sedangkan untuk model *Vector Moving Average* (VMA) merupakan rata-rata dari vektor. Berikut adalah rumus dari model VAR [11]:

$$y_t = v + A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t \quad (3)$$

dengan memberikan istilah *error* pada ε_t . Maka, ε_t merupakan *white noise* yang memiliki *mean* 0 dan variansi 1, dan memiliki representasi model *moving average* (MA) yang terbatas [11]. Berikut merupakan rumus dari VMA [11]:

$$\varepsilon_t = u_t + M_1 u_{t-1} + \dots + M_q u_{t-q} \quad (4)$$

karena model VARMA merupakan perpaduan dari model VAR dan VMA. Berikut adalah persamaan dari model VARMA(p,q) untuk k variabel [11]:

$$y_t = v + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t + M_1 u_{t-1} + \dots + M_q u_{t-q} \quad (5)$$

keterangan:

- y_t : vektor pengamatan berukuran $k \times 1$ berisi k variabel pada saat t
 v : vektor intercept
 u_t : vektor *error* berukuran $k \times 1$ yang memiliki *mean* 0 dan variansi 1
 A_i : matriks parameter koefisien VAR ke- i , dengan $i = 1, 2, \dots, p$ yang berukuran $k \times k$
 M_j : matriks parameter koefisien VMA ke- j , dengan $j = 1, 2, \dots, q$

Berikut ini merupakan contoh model VARMA(p,q) untuk 2-variabel:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} A_{11,1} & A_{12,1} \\ A_{21,1} & A_{22,1} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \dots + \begin{pmatrix} A_{11,p} & A_{12,p} \\ A_{21,p} & A_{22,p} \end{pmatrix} \begin{pmatrix} y_{1,t-p} \\ y_{2,t-p} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} + \begin{pmatrix} M_{11,1} & M_{12,1} \\ M_{21,1} & M_{22,1} \end{pmatrix} \begin{pmatrix} u_{1,t-1} \\ u_{2,t-1} \end{pmatrix} + \dots + \begin{pmatrix} M_{11,q} & M_{12,q} \\ M_{21,q} & M_{22,q} \end{pmatrix} \begin{pmatrix} u_{1,t-q} \\ u_{2,t-q} \end{pmatrix} \quad (6)$$

2.4. Pearson Correlation

Korelasi *Pearson* adalah sejenis uji korelasi, yang digunakan untuk mengetahui apakah hubungan antara dua variabel adalah interval normal, atau rasio distribusi rentang nilai dari -1 hingga 1, dan mengembalikan nilai koefisien korelasi. [12]. Berikut persamaan dari *Pearson Correlation* :

$$r_{ps} = \frac{n \sum_{i=1}^n p_i s_i - \sum_{i=1}^n p_i \sum_{i=1}^n s_i}{\sqrt{n \sum_{i=1}^n p_i^2 - (\sum_{i=1}^n p_i)^2} \cdot \sqrt{n \sum_{i=1}^n s_i^2 - (\sum_{i=1}^n s_i)^2}} \quad (7)$$

dengan r_{xy} adalah korelasi *pearson*, p merupakan variabel x , s merupakan variabel y , dan n adalah jumlah data.

2.5. Order model VARMA

ACF dan PACF

ACF (*Autocorrelation Function*) adalah hubungan antara data yang berhubungan dengan data pengamatan pada waktu t dan data $t + 1$ dengan simpangan waktu k [3]. Pada persamaan atau grafik ACF dapat digunakan untuk menentukan order dalam MA [3]. Berikut merupakan persamaan dari ACF [13]:

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \hat{Y})(Y_{t-k} - \hat{Y})}{\sum_{t=1}^n (Y_t - \hat{Y})^2} \quad (8)$$

dimana r_k adalah koefisien dari ACF, Y_t merupakan nilai variabel Y pada waktu ke- t dan Y_{t-k} adalah nilai variabel Y pada waktu $t - k$ serta \hat{Y} merupakan nilai rata-rata dari variabel Y dan n adalah jumlah data. Sedangkan untuk PACF (*Partial Autocorrelation Function*) adalah korelasi antara Y_t dan Y_{t-k} dengan menghilangkan hubungan linier variabel-variabel $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$ [3]. Pada persamaan atau grafik dari PACF digunakan untuk menentukan order dari AR. Berikut adalah persamaan dari PACF [13]:

$$r_{kk} = \text{Corr}(Y_t, Y_{t-k} | Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}) \quad (9)$$

dimana r_{kk} adalah korelasi dalam distribusi bivariat Y_t dan Y_{t-k} adalah *conditional* pada $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$.

Akaike Information Criteria (AIC)

Selain memeriksa grafik (plot) ACF dan PACF, ada cara lain untuk menentukan order VARMA yang baik yaitu dengan mempertimbangkan nilai *Akaike Information Criteria* (AIC) yang nilainya minimum [14]. Berikut persamaan dari AIC [15]:

$$AIC = -2 \ln(L) + 2p \quad (10)$$

dimana:

- L : *likelihood*
 p : jumlah parameter yang terlibat

2.6. Mean Absolute Percentage Error (MAPE)

MAPE digunakan untuk mengukur model perkiraan nilai kesalahan yang dinyatakan dalam modul rata-rata persentase absolut residual. Berikut merupakan persamaan dari MAPE:

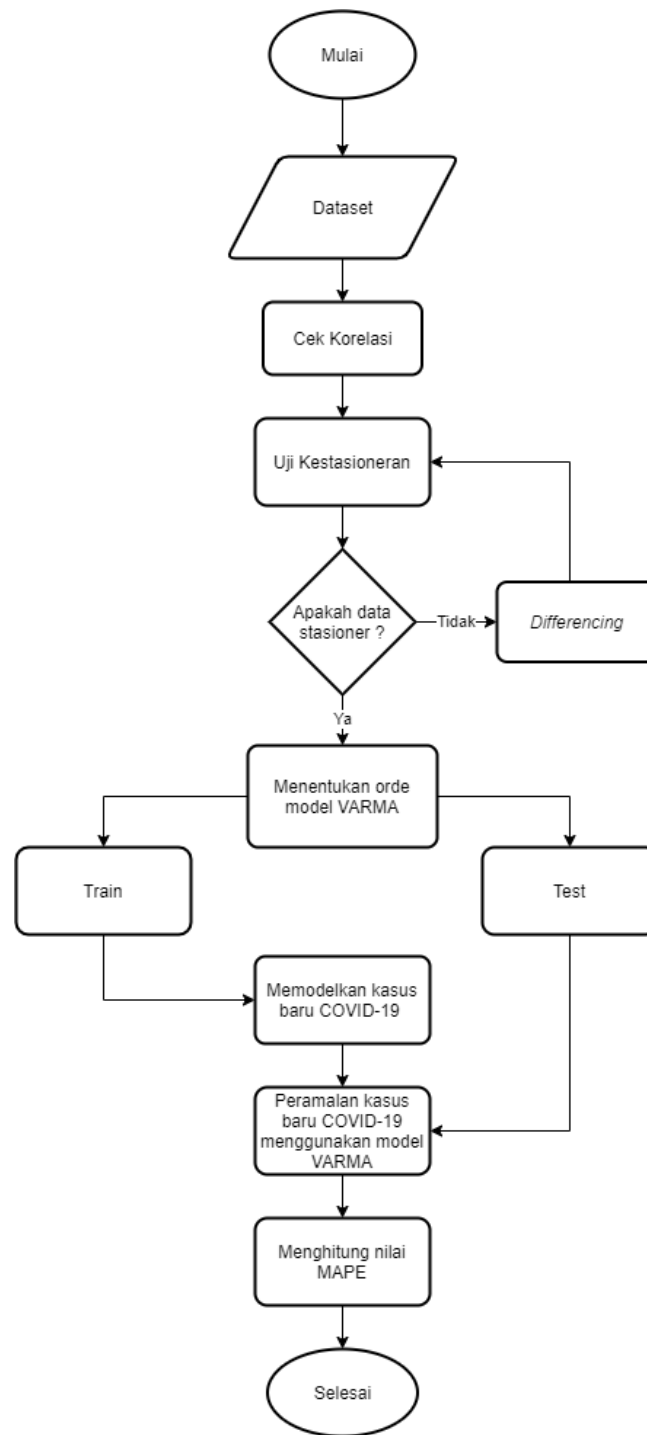
$$MAPE = \frac{\sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\%}{n} \quad (11)$$

keterangan:

- n : banyak periode peramalan
 Y_t : data asli pada saat t
 \hat{Y}_t : data hasil ramalan pada saat t



3. Sistem yang Dibangun



Gambar 2. Flowchart Perancangan Sistem

Flowchart pada Gambar 3 merupakan Langkah-langkah dalam melakukan peramalan penyebaran COVID-19 di Jawa Timur. Berikut adalah penjelasan dari setiap langkah-langkahnya:

1. Dataset

Mengumpulkan data yang akan digunakan untuk melakukan peramalan dan memasukkan data tersebut ke dalam program. Dataset yang digunakan yaitu dataset COVID-19 harian di Jawa Timur. Dataset tersebut diambil dari website resmi yaitu <https://kawalcovid19.id/> yang menampilkan data per provinsi, dan Peneliti mengambil provinsi Jawa Timur untuk di lakukan penelitian. Pada dataset COVID-19 di Jawa Timur terdapat dua variabel yang akan digunakan untuk melakukan peramalan penyebaran COVID-19 harian di Jawa Timur

yaitu data kasus positif COVID-19 harian (kasus baru) dan data sembuh baru. Dataset tersebut yang digunakan mulai bulan Juli 2020 s/d Februari 2021 atau 7 bulan.

2. Cek Korelasi

Pada tahap ini, Peneliti menghitung nilai korelasi antara variabel kasus baru dengan variabel sembuh baru COVID-19. Peneliti akan melihat apakah terdapat hubungan atau korelasi antara kedua variabel tersebut dengan menggunakan *Pearson Correlation*.

3. Uji Kestasioneran

Pada uji kestasioneran ini digunakan untuk melihat apakah data tersebut stasioner atau tidak dengan menggunakan tes ADF, jika data tersebut tidak stasioner, maka perlu dilakukan *differencing* yaitu dengan Persamaan 1.

4. Menentukan order model VARMA

Menentukan lag optimum dari model VARMA dengan langkah-langkah sebagai berikut. Pertama, menentukan order dari AR dan MA terlebih dahulu menggunakan plot ACF dan PACF. Dari plot ACF dan PACF akan mendapatkan order p dan q. Setelah itu, langkah kedua yaitu menentukan model terbaik berdasarkan order yang di dapat dari ACF dan PACF dengan melakukan kombinasi p dan q untuk mendapatkan model yang optimum dengan menggunakan *Akaike Information Criteria* (AIC), model yang dipilih yaitu model yang nilai AIC yang paling kecil.

5. Data Latih dan Data Uji

Data latih digunakan untuk memodelkan kasus baru, sedangkan untuk data uji digunakan untuk melakukan peramalan kasus baru. Dalam data *train* dan *test* terdapat tiga skenario pembagiannya yaitu yang pertama 95:5 dengan data *train* 95% dan *test* 5% dengan jumlah data *testing* 12 hari. Kedua yaitu 90:10 dengan jumlah data *testing* 24 hari. Dan yang terakhir 80:20 dengan jumlah data *testing* 48 hari.

Setelah membagi data, lalu data latih digunakan untuk melakukan pemodelan kasus baru COVID-19 dengan variabel sembuh baru menggunakan model VARMA. Kemudian, hasil dari pemodelan tersebut dilakukan peramalan kasus baru COVID-19 di Jawa Timur.

Pada peramalan kasus baru dengan model VARMA ini juga menggunakan data uji untuk mendapatkan peramalannya. Setelah mendapatkan model dari VARMA, lalu melakukan perhitungan untuk mendapatkan peramalan dari kasus baru.

6. Menghitung nilai MAPE

Dalam menghitung nilai MAPE ini dibutuhkan data hasil peramalan dan data *testing*/data aslinya. Kemudian, dilakukan perhitungan nilai MAPE.

4. Evaluasi

Dalam meramalkan kasus baru COVID-19 di Jawa Timur menggunakan model VARMA yang pertama dilakukan yaitu cek korelasi antar variabel. Pada cek korelasi ini menggunakan *Pearson Correlation* yang menghasilkan nilai korelasi antara variabel kasus baru terhadap sembuh baru yang menghasilkan nilai korelasi sebesar 0.831 atau 83.1%. Korelasi tersebut termasuk korelasi kuat dan bernilai positif, karena korelasinya mendekati nilai 1.

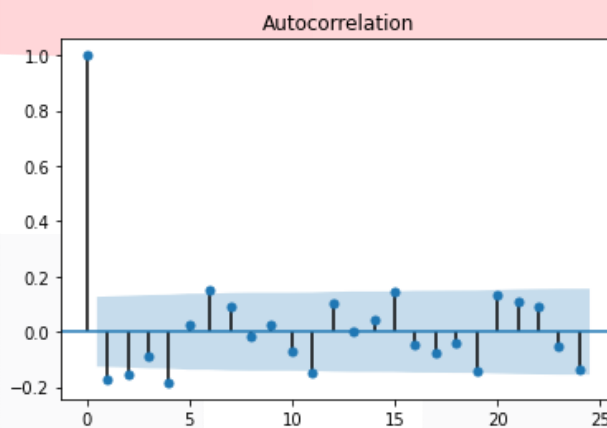
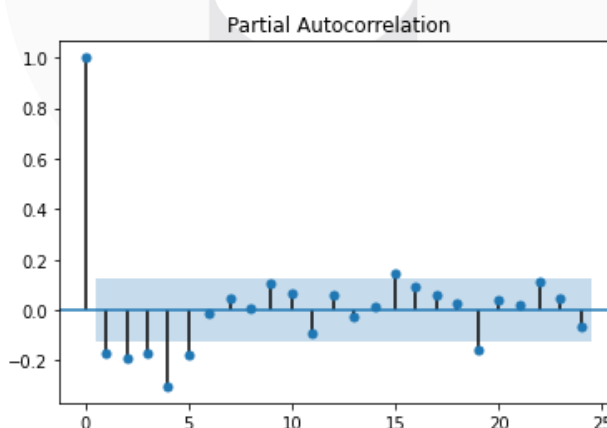
Tabel 1. Uji Kestasioneran menggunakan tes ADF

Nama Variabel	<i>p – value</i>	Keterangan
Kasus Baru	0.7192	Tidak Stasioner
Sembuh Baru	0.4574	Tidak Stasioner

Tabel 2. Uji Kestasioner *first differencing* menggunakan tes ADF

Nama Variabel	<i>p - value</i>	Keterangan
Kasus Baru	0.0	Stasioner
Sembuh Baru	0.0	Stasioner

Pada uji kestasioneran ini menggunakan *Augmented Dickey Fuller* (ADF) dengan bantuan *library* pada *python* yaitu *Adfuller*. Dalam uji stasioner ini menggunakan stasioner lemah, karena hanya menggunakan dua karakteristik, yaitu nilai rata-rata dan variansi yang konstan. Dapat dilihat pada Tabel 1 yaitu uji kestasioneran yang terdapat nilai *p - value* pada kedua variabel. Dari tes ADF diperoleh *p-value* untuk variabel kasus baru adalah 0.7192 dan untuk variabel sembuh (baru) adalah 0.4574 pada tingkat signifikansi 5%. Karena *p - value* > 5% maka tolak H_1 yang artinya variabel tersebut tidak stasioner. Karena variabel tidak stasioner, maka dilakukan *first differencing*. Dari hasil *differencing* tersebut dilakukan uji kestasioneran menggunakan tes ADF. Pada Tabel 2 yaitu hasil tes ADF sesudah dilakukan *differencing*, dari hasil tersebut diperoleh *p - value* dari kedua variabel adalah 0.0 pada tingkat signifikansi 5%. Karena nilai *p - value* $\leq \alpha$ (0.05) sehingga tolak H_0 maka, kedua variabel tersebut stasioner.

**Gambar 3. Autocorrelation****Gambar 4. Partial Autocorrelation**

Setelah melakukan uji kestasioneran data, maka menentukan order pada model VARMA. Dalam menentukan order model VARMA ini menggunakan plot ACF dan PACF. Pada Gambar 3 terlihat plot ACF untuk menentukan order dari *Moving Average* (MA). Dari Gambar 3 tersebut dapat dilihat *cut off* terjadi pada lag-3 yang berada pada bayangan warna biru. Lag yang diambil merupakan lag yang maksimum atau sebelum terjadi *cut off* yaitu lag-2, maka order dari *Moving Average* yaitu MA(2). Kemudian, pada Gambar 4 merupakan plot PACF yang digunakan untuk menentukan order dari AR. Dari plot PACF tersebut, *cut off* terjadi pada lag-6 yang berada pada

bayangan warna biru artinya lag tersebut tidak memiliki korelasi. Lag yang diambil adalah lag yang maksimum yaitu lag-5 yang berada tepat sebelum *cut off*, maka lag-5 tersebut merupakan order dari AR(5). Dari plot ACF dan PACF menghasilkan order AR(5) dan MA(2). Setelah itu dilakukan kombinasi order AR(p) dan MA(q) tersebut dengan menggunakan AIC pada Tabel 3.

Tabel 3. Nilai AIC

Order VARMA(p,q)	AIC
VARMA(1,1)	5699.03
VARMA(1,2)	5703.86
VARMA(2,1)	5694.58
VARMA(2,2)	5694.13
VARMA(3,1)	5693.94
VARMA(3,2)	5694.55
VARMA(4,1)	5690.30
VARMA(4,2)	5691.26
VARMA(5,1)	5679.46
VARMA(5,2)	5683.92

Dapat dilihat pada Tabel 3 tersebut merupakan hasil kombinasi AR(p) dan MA(q) dengan $p = 5$ dan $q = 2$ menggunakan AIC. Dari kombinasi order AR(p) dan MA(q) dilihat nilai AIC yang terkecil untuk digunakan dalam model VARMA. Dilihat dari Tabel 3 tersebut nilai *Akaike Information Criteria* (AIC) yang terkecil yaitu pada order $p = 5$ dan $q = 1$, maka didapatkan order dari model VARMA adalah VARMA(5,1) yang terbentuk sebagai berikut:

$$y_t = v + A_1 y_{t-1} + A_2 y_{t-2} + A_3 y_{t-3} + A_4 y_{t-4} + A_5 y_{t-5} + u_t + M_1 u_{t-1}$$

dengan,

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} A_{11,1} & A_{12,1} \\ A_{21,1} & A_{22,1} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} A_{11,2} & A_{12,2} \\ A_{21,2} & A_{22,2} \end{pmatrix} \begin{pmatrix} y_{1,t-2} \\ y_{2,t-2} \end{pmatrix} + \begin{pmatrix} A_{11,3} & A_{12,3} \\ A_{21,3} & A_{22,3} \end{pmatrix} \begin{pmatrix} y_{1,t-3} \\ y_{2,t-3} \end{pmatrix} \\ + \begin{pmatrix} A_{11,4} & A_{12,4} \\ A_{21,4} & A_{22,4} \end{pmatrix} \begin{pmatrix} y_{1,t-4} \\ y_{2,t-4} \end{pmatrix} + \begin{pmatrix} A_{11,5} & A_{12,5} \\ A_{21,5} & A_{22,5} \end{pmatrix} \begin{pmatrix} y_{1,t-5} \\ y_{2,t-5} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} + \begin{pmatrix} M_{11,1} & M_{12,1} \\ M_{21,1} & M_{22,1} \end{pmatrix} \begin{pmatrix} u_{1,t-1} \\ u_{2,t-1} \end{pmatrix}$$

Dari persamaan model VARMA(5,1) diatas, terdapat y_t yang merupakan vektor pengamatan berukuran 2×1 karena dalam pengamatan ini menggunakan 2 variabel yaitu kasus baru dan sembuh (baru). Selain itu, terdapat juga v merupakan vektor intercept berukuran 2×1 yang nilainya diambil dari nilai koefisien parameter intercept yang terdapat pada model VARMA(5,1), untuk A_1, A_2, \dots, A_5 merupakan matriks dari parameter koefisien VAR yang terdapat model VARMA(5,1). Pada persamaan model VARMA(5,1) juga terdapat u_t yang merupakan vektor *error* yang berukuran 2×1 yang memiliki *mean* 0 dan variansi 1 yang merupakan bilangan acak (random) ke-i sebanyak 242 dengan rentang nilai 0 s/d 1 karena rata-ratanya yaitu 0 untuk batas bawahnya dan variansi 1 untuk nilai batas atasnya, untuk M_1 merupakan parameter koefisien VMA yang nilainya didapatkan dari model VARMA(5,1). Dari persamaan tersebut digunakan untuk peramalan kasus baru COVID-19 di Jawa Timur menggunakan model VARMA.

Tabel 4. Result for equation Kasus Baru menggunakan data *testing* 12 Hari

Parameter	coef	std err	z	P> z
Intercept (ν)	2.5807	5.631	0.991	0.322
L1.Kasus Baru ($A_{11,1}$)	-0.1522	0.324	-0.470	0.638
L1.Sembuh (baru) ($A_{12,1}$)	-0.0354	0.277	-0.156	0.876
L2.Kasus Baru ($A_{11,2}$)	-0.2979	0.137	-2.170	0.030*
L2.Sembuh (baru) ($A_{12,2}$)	0.0460	0.173	0.267	0.790
L3.Kasus Baru ($A_{11,3}$)	-0.2506	0.137	-1.827	0.068
L3.Sembuh (baru) ($A_{12,3}$)	-0.0723	0.130	-0.555	0.579
L4.Kasus Baru ($A_{11,4}$)	-0.2951	0.133	-2.615	0.009*
L4.Sembuh (baru) ($A_{12,4}$)	-0.1622	0.103	-1.575	0.115
L5.Kasus Baru ($A_{11,5}$)	-0.0924	0.126	-0.731	0.465
L5.Sembuh (baru) ($A_{12,5}$)	-0.0927	0.107	-0.865	0.387
L1.e(Kasus Baru) ($M_{11,1}$)	-0.2231	0.336	-0.665	0.506
L1.e(Sembuh (baru)) ($M_{12,1}$)	-0.0026	0.246	0.010	0.992

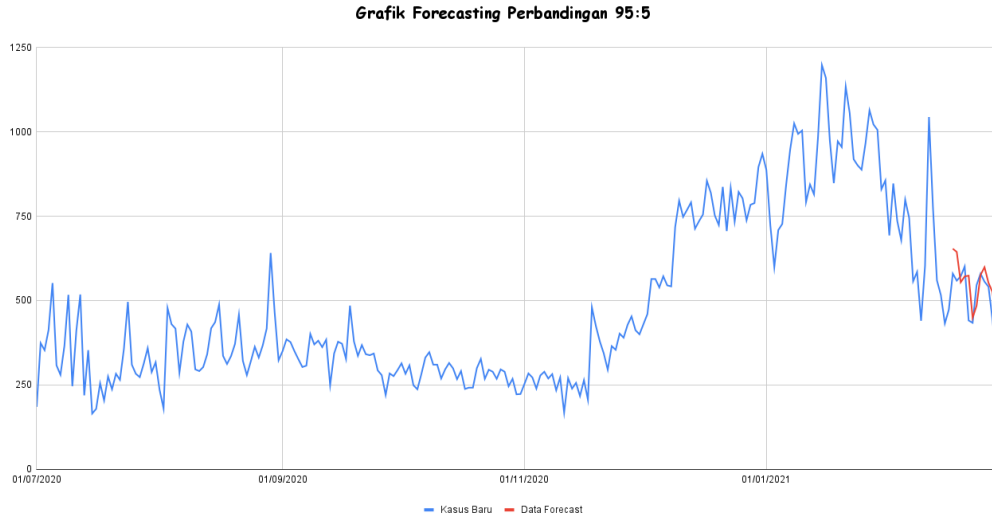
Tabel 5. Result for equation Kasus Baru menggunakan data *testing* 24 Hari

Parameter	coef	std err	z	P> z
Intercept (ν)	6.0013	4.982	1.205	0.228
L1.Kasus Baru ($A_{11,1}$)	-0.0350	0.303	-0.116	0.908
L1.Sembuh (baru) ($A_{12,1}$)	-0.0117	0.196	-0.060	0.952
L2.Kasus Baru ($A_{11,2}$)	-0.1743	0.148	-1.180	0.238
L2.Sembuh (baru) ($A_{12,2}$)	0.0443	0.160	0.277	0.782
L3.Kasus Baru ($A_{11,3}$)	-0.1983	0.108	-1.828	0.068
L3.Sembuh (baru) ($A_{12,3}$)	-0.1170	0.117	-0.996	0.319
L4.Kasus Baru ($A_{11,4}$)	-0.1883	0.102	-1.846	0.065
L4.Sembuh (baru) ($A_{12,4}$)	-0.1855	0.090	-2.054	0.040*
L5.Kasus Baru ($A_{11,5}$)	-0.0334	0.106	-0.315	0.752
L5.Sembuh (baru) ($A_{12,5}$)	-0.0927	0.116	-0.798	0.425
L1.e(Kasus Baru) ($M_{11,1}$)	-0.3703	0.316	-1.172	0.241
L1.e(Sembuh (baru)) ($M_{12,1}$)	-0.0118	0.213	-0.055	0.956

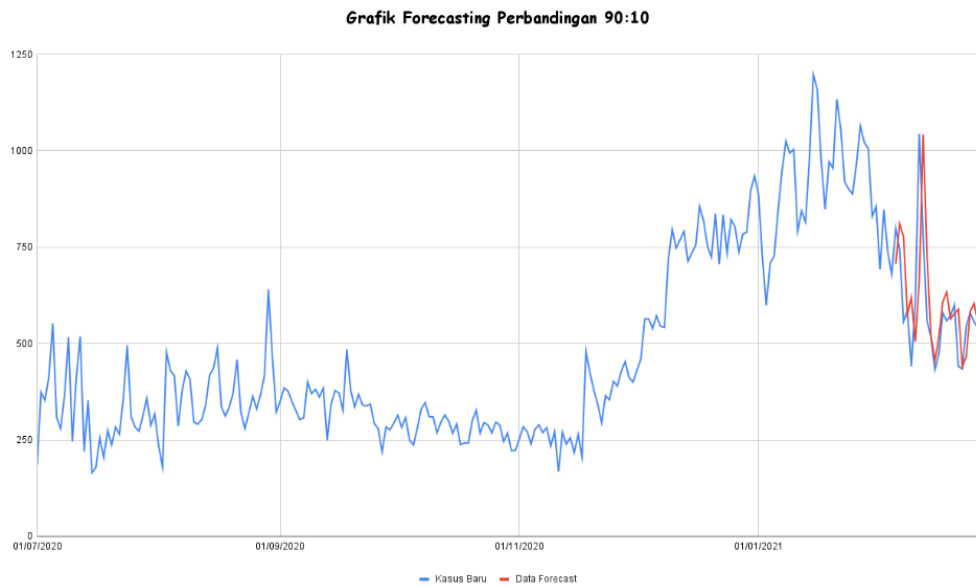
Tabel 6. Result for equation Kasus Baru menggunakan data *testing* 48 Hari

Parameter	coef	std err	z	P> z
Intercept (v)	9.8760	10.457	0.944	0.345
L1.Kasus Baru ($A_{11,1}$)	-0.5559	0.393	-1.414	0.157
L1.Sembuh (baru) ($A_{12,1}$)	-0.4987	0.555	-0.899	0.369
L2.Kasus Baru ($A_{11,2}$)	-0.3283	0.199	-1.650	0.099
L2.Sembuh (baru) ($A_{12,2}$)	-0.3029	0.396	-0.764	0.445
L3.Kasus Baru ($A_{11,3}$)	-0.2268	0.116	-1.368	0.171
L3.Sembuh (baru) ($A_{12,3}$)	-0.2336	0.243	-0.961	0.337
L4.Kasus Baru ($A_{11,4}$)	-0.2003	0.127	-1.575	0.115
L4.Sembuh (baru) ($A_{12,4}$)	-0.3712	0.155	-2.393	0.017*
L5.Kasus Baru ($A_{11,5}$)	-0.0564	0.127	-0.0445	0.657
L5.Sembuh (baru) ($A_{12,5}$)	-0.3435	0.195	-1.764	0.078
L1.e(Kasus Baru) ($M_{11,1}$)	0.1522	0.407	0.374	0.708
L1.e(Sembuh (baru)) ($M_{12,1}$)	0.4657	0.574	0.811	0.418

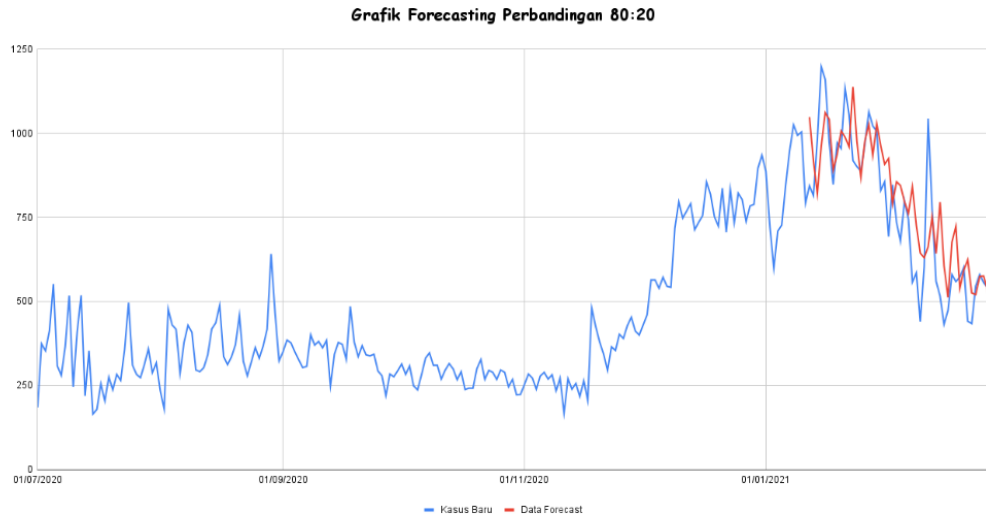
Dari order yang didapat dari nilai AIC terkecil tersebut, kemudian dilakukan pemodelan. Dalam memodelkan kasus baru ini menggunakan tiga skenario, yaitu skenario pertama perbandingan 95:5 dengan jumlah data *train* 230 hari dan data *test* 12 hari, lalu skenario kedua perbandingan 90:10 dengan jumlah data *train* 218 hari dan data *test* 24 hari, kemudian skenario yang ketiga perbandingan 80:20 dengan jumlah data *train* 194 hari dan data *test* 48 hari. Tabel 4 merupakan hasil persamaan kasus baru dari model VARMA dengan jumlah data *train* 230 hari. Dari hasil tersebut terdapat parameter L_i . Kasus Baru dan L_i . Sembuh (baru). L_i . Kasus Baru adalah parameter untuk banyaknya kasus baru di lag ke- i sedangkan L_i . Sembuh (baru) adalah banyaknya sembuh baru pada lag ke- i . Selain itu, terdapat juga *coef*, *std error*, z , dan $P > |z|$ atau p - value. *Coef* tersebut digunakan untuk peramalan kasus baru COVID-19 di Jawa Timur, sedangkan untuk $P > |z|$ digunakan untuk melihat parameter mana saja yang signifikan terhadap model VARMA. Dari Tabel 4 tersebut dapat dilihat parameter yang signifikan yaitu L2.Kasus Baru dan L4.Kasus Baru dengan nilai ($P > |z|$) < 5%. Selanjutnya untuk Tabel 5 merupakan hasil persamaan kasus baru dari model VARMA dengan jumlah data *train* 218 hari. Dari hasil tersebut dapat dilihat $P > |z|$ yang nilai kurang dari 0.05 (5%) yaitu pada parameter L4.Sembuh (baru), maka parameter L4.Sembuh (baru) signifikan terhadap model VARMA. Kemudian untuk Tabel 6 yaitu hasil persamaan kasus baru dari model VARMA yang menggunakan data *train* 194 hari. Dari hasil tersebut parameter yang signifikan terletak pada posisi yang sama dengan Tabel 5 yaitu L4.Sembuh (baru) dengan nilai $P > |z|$ yaitu 0.017 atau kurang dari 5%.



Gambar 5. Grafik forecasting dengan perbandingan 95:5 (12 Hari)



Gambar 6. Grafik forecasting dengan perbandingan 90:10 (24 Hari)



Gambar 7. Grafik forecasting dengan perbandingan 80:20 (48 Hari)

Setelah memodelkan kasus baru, didapatkan parameter dan koefisien dari Tabel 4, 5, dan 6. Kemudian dilakukan peramalan kasus baru menggunakan nilai *coef* dari parameter yang terdapat dari hasil pemodelan pada Tabel 4, 5, dan 6. Dari peramalan tersebut mendapatkan grafik peramalan dari 3 skenario. Pada grafik yang terdapat pada Gambar 5 yang merupakan skenario pertama, menunjukkan bahwa grafik antara *actual* dengan *forecast* mengalami kemiripan. Ketika grafik *actual* mengalami kenaikan, grafik dari *forecast* juga mengalami kenaikan dan sebaliknya. Pada Gambar 6 merupakan skenario kedua yang menunjukkan grafik dari data *actual* dengan data *forecast*. Dapat dilihat antara grafik *actual* dengan *forecast* mengalami kenaikan dan penurunan yang hampir sama, yang sangat membedakan pada tanggal 9 Februari 2021 yaitu data *actual* jumlah kasusnya 440 sedangkan untuk peramalannya 660 kasus. Dan untuk Gambar 7 merupakan skenario ketiga menunjukkan grafik antara *actual* dengan *forecast*, untuk grafik *forecast* banyak yang berada di atas grafik *actual*, tetapi untuk penurunan dan kenaikan grafik hampir sama.

Setelah mendapatkan peramalan kasus baru COVID-19, kemudian menghitung nilai MAPE untuk dilakukan evaluasi hasil. Dari skenario 95:5 mendapatkan nilai MAPE 12.18%, sedangkan untuk skenario 90:10 mendapatkan nilai MAPE 15.56% dan skenario 80:20 mendapatkan nilai MAPE 16.17%. Banyaknya data *train* yang digunakan berpengaruh pada nilai MAPE yang semakin kecil, karena dari data *train* tersebut memiliki variansi data yang lebih banyak, yang menjadikan hasil peramalan yang lebih baik. Dari ketiga skenario tersebut dapat di ambil kesimpulan bahwa skenario 95:5 merupakan hasil peramalan yang terbaik dibanding skenario 90:10 dan 80:20 karena skenario 95:5 mendapatkan nilai *error* (MAPE) yang terkecil.

5. Kesimpulan

Berdasarkan hasil pengujian dan analisis diatas, dapat diambil kesimpulan bahwa order model VARMA yang di dapat dari nilai AIC yang terkecil yang bernilai 5679.46 yang mendapatkan order model VARMA(5,1). Dari model VARMA(5,1) dilakukan pengujian menggunakan tiga skenario. Dari tiga skenario yang di uji, skenario 95:5 mendapatkan nilai MAPE yang terkecil yaitu 12.18% dengan menggunakan jumlah data latih 230 hari dan data uji 12 hari. Hal ini menunjukkan peramalan model VARMA(5,1) dengan perbandingan data 95:5 merupakan hasil peramalan yang terbaik dibandingkan dengan skenario 90:10 dan 80:20.

Untuk mengatasi lonjakan kasus baru COVID-19 bisa digunakan prediksi nilai ekstrim dengan menggunakan model *multivariate time series*. Selain itu, dapat mempertimbangkan pengambilan sampel data per hari untuk mendapatkan hasil peramalan yang baik.

REFERENSI

- [1] F. S. D. Arianto and P. Noviyanti, "Prediksi Kasus Covid-19 Di Indonesia Menggunakan Metode Backpropagation Dan Fuzzy Tsukamoto," *JurTI (Jurnal Teknol. Informasi)*, vol. 4, no. 1, pp. 120–127, 2020.
- [2] padk.kemkes.go.id, "HINDARI LANSIA DARI COVID 19," *padk.kemkes.go.id*, 2018. <http://www.padm.kemkes.go.id/article/read/2020/04/23/21/hindari-lansia-dari-covid-19.html> (accessed Nov. 15, 2020).
- [3] A. Ulya, *Peramalan harga saham penutupan menggunakan metode vector autoregressive moving average (VARMA)*. Universitas Islam Negeri Maulana Malik Ibrahim, 2019.
- [4] V. Ciptaningtyas, *Penerapan Model Seasonal Vector Autoregressive Moving Average (Seasonal VARMA) pada Data Curah Hujan dan Suhu Rata-rata di Stasiun Meteorologi Juanda, Sidoarjo*. Universitas Brawijaya, 2017.
- [5] H. Wiguna, Y. Nugraha, F. Rizka R, A. Andika, J. I. Kanggrawan, and A. L. Suherman, "Kebijakan Berbasis Data: Analisis dan Prediksi Penyebaran COVID-19 di Jakarta dengan Metode Autoregressive Integrated Moving Average (ARIMA)," *J. Sist. Cerdas*, vol. 3, no. 2, pp. 74–83, 2020, doi: 10.37396/jsc.v3i2.76.
- [6] A. S. Albana and S. Azhari, "Prediksi Penyebaran COVID-19 Kota Surabaya dengan Simulasi Monte Carlo," *J. Adv. Inf. Ind. Technol.*, vol. 2, no. 1, pp. 36–42, 2020, doi: 10.52435/jaiit.v2i1.40.
- [7] R. RINALDY, *ANALISIS PERAMALAN DATA RUNTUN WAKTU MENGGUNAKAN VECTOR AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (VARIMA) (Studi Kasus Data Tingkat Inflasi Indonesia dan BI Rate)*. FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM, 2016.
- [8] Y. Tunang, T. Manurung, and N. Nainggolan, "Penerapan Model Vector Autoregressive (VAR) untuk Memprediksi Harga Cengkeh, Kopra dan Pala di Sulawesi Utara," *d'CARTESIAN*, vol. 8, no. 2, pp. 100–107, 2019, doi: 10.35799/dc.8.2.2019.23967.
- [9] J. Rizal and S. Akbar, "Perbandingan Uji Stasioner Data Timeseries Antara Metode: Control Chart, Correlogram, Akar Unit Dickey Fuller, dan Derajat Integrasi," *GRADIEN J. Ilm. MIPA*, vol. 11, no. 1, pp. 1040–1046, 2015.
- [10] N. Hendajany and R. Wati, "Prediksi indikator makro ekonomi indonesia dengan model vector autoregressive periode 2019-2023," *J. Ekon. dan Bisnis*, vol. 23, no. 2, pp. 189–202, 2020, doi: 10.24914/jeb.v23i1.2878.
- [11] H. Lütkepohl, *New introduction to multiple time series analysis*. 2005.
- [12] Y. E. Windarto, "ANALISIS PENYAKIT KARDIOVASKULAR MENGGUNAKAN METODE KORELASI PEARSON, SPEARMAN DAN KENDALL," *J. Saintekom*, vol. 10, no. 2, pp. 119–127, 2020, doi: 10.33020/saintekom.v10i2.149.
- [13] K.-S. Cryer, Jonathan D and Chan, *Time series analysis: with applications in R*, vol. 2. Springer, 2008.
- [14] I. Muthahharah, "Pemodelan Harga Saham Negara ASEAN Menggunakan VARMA dan VARMAX," Institut Teknologi Sepuluh Nopember, 2015.
- [15] N. H. Miswan, R. M. Said, and S. H. H. Anuar, "ARIMA with regression model in modelling electricity load demand," *J. Telecommun. Electron. Comput. Eng.*, vol. 8, no. 12, pp. 113–116, 2016.