

Prediksi Nilai Ekstrem Jumlah Kasus Harian Positif COVID-19 di Provinsi Jawa Timur dengan Model *Vector Autoregressive Moving Average* (VARMA)

Tasyalva Aulia I¹, Aniq Atiqi Rohmawati², Indwiarti³

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

¹tasyalva@students.telkomuniversity.ac.id,

²aniqatiqi@telkomuniversity.ac.id,³indwiarti@telkomuniversity.ac.id

Abstrak

COVID-19 telah menyebar ke seluruh dunia dan belum mereda. Pandemi COVID-19 muncul pertama kali pada akhir tahun 2019. Hal ini menjadi ancaman global bagi kehidupan manusia dan masyarakat. Kasus COVID-19 mulai berkembang sejak 2 Maret 2020 di Indonesia. Hingga saat ini, kasus positif COVID-19 sangat berfluktuasi. Oleh karena itu, diperlukan adanya estimasi nilai ekstrem penambahan kasus COVID-19 di Jawa Timur dengan melibatkan tingkat kesembuhan pasien di provinsi tersebut berdasarkan pada data *time series*. Penelitian ini bertujuan untuk mengantisipasi lonjakan kasus dan membantu membuat rekomendasi penanganan COVID-19. Pada penelitian ini pemodelan dilakukan dengan model *Vector Autoregressive Moving Average* (VARMA), kemudian untuk mengestimasi lonjakan kasus yang terjadi pada rentang tanggal 1 Juli 2020 sampai 28 Februari 2021 digunakan *Block Maxima*. Estimasi pelanggaran nilai ekstrem menggunakan *Block Maxima* dengan model terbaik VARMA(5,1) dan hasil MAPE 0.15 memberikan forecasting yang cukup baik untuk pelanggaran nilai dengan nilai 0.2983, sehingga jika terjadi lonjakan maka 70.17% mencapai 573 kasus.

Kata kunci : COVID-19, VARMA, Jawa Timur, estimasi batas atas, nilai ekstrem.

Abstract

COVID-19 has spread around the world and has not subsided. The COVID-19 pandemic first appeared at the end of 2019. It has become a global threat to human life and society. COVID-19 cases began to develop since March 2, 2020 in Indonesia. Until now, positive cases of COVID-19 fluctuated greatly. Therefore, it is necessary to estimate extreme value addition COVID cases-19 in East Java by involving the patient cure rate in the province based on the time series data. This study aims to anticipate a spike in cases and help make regulations for handling COVID-19. In this study, the modeling was carried out using *Vector Autoregressive Moving Average* (VARMA) model, then to estimate the spike in cases that occurred in the range from July 1, 2020 to February 28, 2021, *Block Maxima* was used. Estimation of extreme value violations using *Block Maxima* with the best model VARMA(5,1) and 0.15 MAPE results provide a fairly good forecasting for value violations with a value of 0.2983, so that if there is a spike then 70.17% will reach 573 cases.

Keywords: COVID-19, VARMA, East Java, upper limit estimate, extreme value.

1. Pendahuluan

Latar Belakang

Sejak ditemukannya COVID-19 yang berasal dari Wuhan, China pada akhir 2019, wabah ini telah menjadi perhatian berbagai pihak karena berdampak besar terhadap kehidupan manusia, perlambatan ekonomi dan kematian. Pandemi COVID-19 dinyatakan sebagai ancaman global terhadap kesehatan masyarakat dan kehidupan sosial oleh World Health Organization (WHO). Beberapa strategi pencegahan yang disarankan oleh WHO adalah menggunakan masker, mencuci tangan, menjaga jarak, menghindari kerumunan publik, dll. Pada tanggal 2 Maret 2020, dinyatakan terdeteksi 2 orang positif COVID-19 di Indonesia. Penambahan jumlah kasus positif COVID-19 perhari tidak dapat diperkirakan. Sekitar 20% manusia yang terinfeksi tidak menunjukkan gejala apapun [8] atau biasa disebut Orang Tanpa Gejala (OTG). Wabah ini lebih berbahaya untuk orang tua berusia di atas 60 tahun yang sebelumnya memiliki riwayat penyakit jantung, kanker, imunitas lemah, dll [6].

Kasus COVID-19 di Indonesia telah memicu diskusi risiko karena terjadinya lonjakan kasus yang ekstrem. Mengingat bahwa Indonesia merupakan negara dengan kepadatan dan mobilitas penduduk yang cukup tinggi. Selain itu, provinsi Jawa Timur masuk kedalam lima besar jumlah kasus COVID-19 tertinggi di Indonesia. Berdasarkan permasalahan diatas, maka penulis melakukan estimasi batas atas tingkat kesembuhan terhadap jumlah harian kasus positif COVID-19 Provinsi Jawa Timur dengan model *Vector Autoregressive Moving Average* (VARMA). Model VARMA merupakan kombinasi dari metode *Vector Autoregressive* (VAR) dan metode *Vector*

Moving Average (VMA). Model ini merupakan model yang efektif untuk data multi-variabel. Model VARMA



mempunyai syarat bahwa data harus stasioner [7]. Analisis pada penelitian ini menggunakan stasioneritas lemah karena hanya mengecek jarak setiap data mempunyai rata-rata yang sama atau tidak. Model digunakan untuk meramalkan data multivariat dan memperoleh estimasi penambahan kasus positif COVID-19 di Jawa Timur.

Salah satu pendekatan untuk mengetahui nilai ekstrem adalah menggunakan *Block Maxima*. *Block Maxima* mengidentifikasi nilai ekstrem berdasarkan nilai maksimum data yang dikelompokkan sesuai periode tertentu. Pada penelitian-penelitian sebelumnya, nilai ekstrim telah diimplementasikan untuk mengetahui curah hujan, seperti pada jurnal [10], membahas tentang analisis curah hujan ekstrem dengan pendekatan *Block Maxima* pada data non-stasioner yang dibandingkan dengan data stasioner. Hasil penelitian tersebut menunjukkan bahwa hasil RSME data non-stasioner lebih kecil dari data stasioner. Selain itu, terdapat penelitian [3]. Menggunakan Metode Vector Autoregressive Moving Average (VARMA) melibatkan data makro ekonomi Indonesia dari Januari 1995 hingga Juli 2005, yaitu data Suku Bunga Sertifikat Bank Indonesia (SBI), data IDR/USD (nilai tukar rupiah terhadap dolar Amerika), dan IHSG (Indeks Harga Saham Gabungan). Hasil penelitian tersebut menunjukkan bahwa interval nilai analisis kesalahan di bawah 10% menyatakan model ini sangat baik untuk meramalkan data. Dalam penelitian ini dibahas tentang pemodelan VARMA untuk memprediksi banyaknya kasus harian di Jawa Timur, dan untuk mengestimasi lonjakan kasus yang dapat memungkinkan terjadi, dengan dilakukan prediksi batas atas dengan menggunakan *Block Maxima*.

Topik dan Batasannya

Topik dan batasan masalah dalam penelitian Tugas Akhir ini yaitu untuk mengetahui apakah model *Vector Autoregressive Moving Average* dapat mengestimasi batas atas penambahan kasus positif pada provinsi Jawa Timur. Penelitian ini terdapat variabel yang akan digunakan untuk dianalisis adalah jumlah kasus positif terhadap tingkat kesembuhan pada provinsi Jawa Timur mulai dari 1 Juli 2020 sampai 28 Februari 2021, yang diperoleh dari situs resmi <https://kawalcovid19.id/>. Estimasi batas atas (nilai ekstrem) dilakukan berdasarkan kedua data tersebut.

Tujuan

Adapun tujuan penelitian ini adalah memodelkan kasus positif COVID-19 di Jawa Timur terhadap tingkat kesembuhan menggunakan model VARMA, serta mengestimasi lonjakan kasus COVID-19 dengan pendekatan nilai ekstrem menggunakan *Block Maxima*.

2. Studi Terkait

2.1. COVID-19

Pada akhir Desember 2019, Coronavirus atau yang dikenal dengan COVID-19 pertama kali muncul di kota Wuhan, Cina. Pada tanggal 11 Maret 2020, Organisasi Kesehatan Dunia (WHO) mengumumkan bahwa virus corona yang saat ini telah merebak dikategorikan sebagai pandemi global. Mayoritas orang yang terinfeksi virus COVID-19 akan mengalami penyakit pernapasan ringan hingga sedang, seperti gejala flu, yang dapat sembuh tanpa perawatan khusus [11] sehingga menyebabkan virus ini sulit untuk dideteksi. Orang tua dan mereka yang menderita penyakit kardiovaskular, diabetes, penyakit pernapasan kronis, dan kanker lebih mungkin berkembang menjadi penyakit serius [11].

Walaupun vaksin COVID-19 sudah ditemukan, ternyata masih belum efektif karena berkembangnya virus corona varian delta. Namun vaksin COVID-19 tidak dapat mencegah seseorang tertular COVID-19 varian apa pun, setidaknya vaksin dapat mengurangi tingkat keparahan COVID-19 dan risiko kematian. Cara terbaik untuk memperlambat penyebaran adalah memahami sepenuhnya virus COVID-19, penyakit yang disebabkan, dan bagaimana penyebarannya. Saat ini, pencegahan yang dapat dilakukan untuk mengurangi penyebaran COVID-19 adalah dengan menjaga kesehatan, berdiam diri di rumah, mencuci tangan sebelum dan sesudah melakukan aktivitas, selalu menggunakan masker dan tidak berada di keramaian.

2.2. Metode Peramalan

2.2.1. Data Time Series

Data *time series* adalah serial dari kumpulan data dari waktu ke waktu. Deret waktu yang dimiliki oleh data *time series* bisa berupa jumlah data tahunan, bulanan, harian, jam, bahkan milidetik. Sewaktu data tersebut disimpan dalam urutan waktu, dapat disebut data *time series*. Tujuan dari analisis data *time series* adalah untuk memprediksi masa depan dari data *time series* itu sendiri [12]. Data *time series* juga dapat memberikan deskripsi statistik penting dari suatu data yang

dibutuhkan untuk membuat keputusan bisnis tanpa harus memprediksi apapun [12].



Secara garis besar karakteristik data *time series* dibagi menjadi dua, yaitu stasioner dan tidak stasioner. Jika nilai rata-rata dan varian data selalu konstan pada setiap periode, data dikatakan stasioner. Data stasioner tidak memiliki tren, yang berarti data tidak mengalami kenaikan atau penurunan. Sebaliknya, data yang tidak stasioner memiliki rata-rata dan varian yang tidak konstan. Seiring dengan bertambahnya jumlah data deret waktu, variansi menjadi semakin besar bila jumlah data runtut waktu diperluas dan autokorelasinya cenderung tidak menurun. Stasioneritas data mempunyai dua jenis stasioneritas yaitu *Strong Stationarity* dan *Weak Stationarity*. Proses *Weak Stationarity* memiliki sifat bahwa mean, variansi, dan autocovariance tidak berubah seiring berjalannya waktu. Pada analisis *time series* terdapat analisis yang mengandung satu variabel atau *univariate*, dan yang melibatkan beberapa variabel disebut *multivariate*.

2.2.2. Vector Autoregressive Moving Average (VARMA)

Model *Vector Autoregressive Moving Average* (VARMA) adalah model *time series* yang dapat digunakan untuk mempelajari objek dengan dua variabel atau lebih, dimana variabel tersebut saling mempengaruhi. Model ini menjelaskan hubungan antara pengamatan sebelumnya terhadap variabel itu sendiri dengan pengamatan pada variabel tertentu [13]. Metode VARMA merupakan kombinasi antara orde p pada model *Vector Autoregressive* (VAR) dengan orde q pada model *Vector Moving Average* (VMA) sehingga membentuk model VARMA (p, q). Model VARMA (p, q) dengan K variabel mempunyai bentuk persamaan [4] :

$$y_t = w + \sum_{i=1}^p \varphi_i y_{t-i} + \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (3)$$

- y_t = data yang diamati pada waktu t tertentu dengan ukuran ($K \times 1$).
 φ_i = matriks koefisien autoregresi ke- i berukuran ($K \times K$).
 θ_i = matriks koefisien moving average ke- i dengan ukuran ($K \times K$).
 ε_t = vektor random pada waktu t berukuran $K \times 1$ yang memiliki mean 0 dan variansi 1.
 w = konstanta penyesuaian.

2.3. Uji Stasioneritas

Analisis deret waktu (*time series*) membutuhkan data stasioner sebelum melakukan analisis lebih lanjut dan syarat model VARMA bahwa data harus stasioner [7]. Untuk memenuhi syarat tersebut, maka uji stasioneritas perlu dilakukan. Uji stasioneritas dilakukan dengan pengujian akar unit (*Unit Root Tools*). Metode akar unit yang digunakan adalah *Dickey-Fuller test* (pengujian DF). Jika variabel Y_t sebagai variabel dependen, maka akan diubah menjadi :

$$Y_t = \rho Y_{t-1} + \varepsilon_t \quad (4)$$

Berdasarkan data, diperoleh nilai peluang (p -value), yang digunakan untuk menentukan kriteria pengujian hipotesis. Sedangkan hipotesis yang digunakan untuk menguji kestasioneran data adalah :

- $H_0 : \rho = 1$ (variabel Y tidak stasioner)
- $H_1 : \rho \neq 1$ (variabel Y stasioner)

Kriteria uji yang digunakan adalah : Tolak H_0 jika p -value $> \alpha$. Jika hipotesis H_0 ditolak, maka data *time series* Y_t stasioner. Jika p -value $\leq \alpha$, maka H_0 diterima, sehingga data Y_t tidak stasioner.

Jika koefisien $Y_{t-1}(\rho)$ adalah 1, yang berarti hipotesis diterima, maka variabel bersifat non-stasioner mengandung akar satuan (*unit root*). Untuk mengubah data non-stasioner menjadi stasioner dilakukan uji orde pertama (*first difference*). *Differencing* merupakan suatu metode yang digunakan untuk mengubah data non-stasioner menjadi data stasioner. Dalam penghitungan data *time series* diperlukan data berbentuk stasioner. Jika data *time series* tidak stasioner pada orde nol, maka kestasioneran data dapat dicari melalui orde berikutnya, sehingga diperoleh tingkat stasioneritas pada order ke- n *first difference*, *second difference* dan seterusnya. Rumus differencing dapat dinyatakan dalam rumus sebagai berikut [9] :

$$\Delta Y_t = Y_t - Y_{t-1} \quad (5)$$

ΔY_t = variabel *first difference* pada waktu t
 Y_t, Y_{t-1} = nilai variabel Y pada waktu t dan $t-1$
 ρ = koefisien *autoregressive*
 ε_t = error *white noise*

Re-Differencing merupakan suatu metode untuk mengubah kembali data *differencing* yang sebelumnya menjadi data non-stasioner. Rumus *re-differencing* dapat dinyatakan sebagai berikut :

$$Y_t = \Delta Y_t + Y_{t-1} \quad (6)$$

ΔY_t = variabel *first difference* pada waktu t
 Y_t = nilai *forecast* pada saat t
 Y_{t-1} = nilai variabel Y pada waktu $t-1$

2.4. Akaike's Information Criterion (AIC)

Untuk estimasi model VARMA diperlukan penentuan orde model. Menurut [5] model dengan kinerja terbaik ditentukan berdasarkan nilai AIC minimum. Rumus untuk mendapatkan nilai AIC adalah sebagai berikut [1] :

$$AIC = -2L + 2r \quad (7)$$

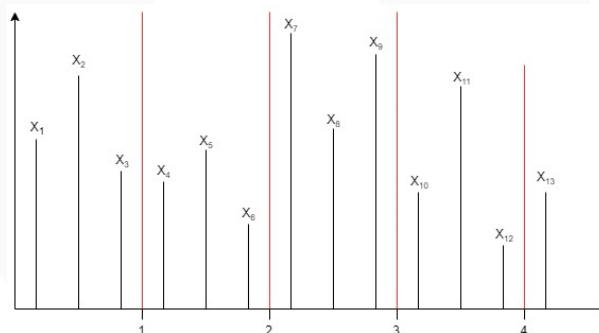
L = Log Likelihood

r = banyaknya parameter dalam model VARMA

2.5. Block Maxima

Block Maxima adalah salah satu pendekatan yang digunakan pada *Extrem Value Theory* untuk menentukan nilai ekstrem berdasarkan nilai maksimum dari data pengamatan yang dikelompokkan dalam periode tertentu. Langkah-langkah pengambilan sampel menggunakan metode *Block Maxima* adalah sebagai berikut [10] :

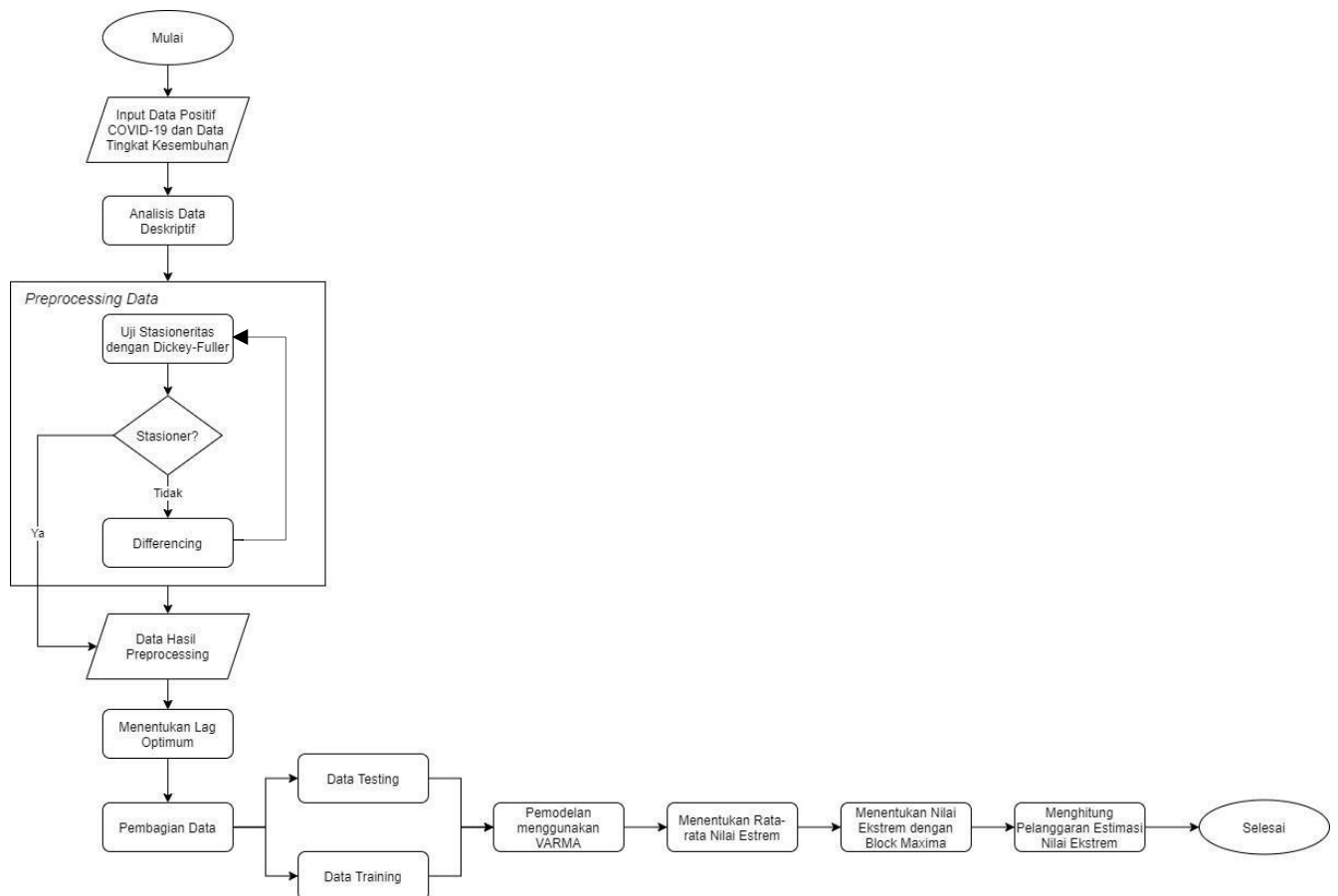
- Data dibagi dalam blok-blok dengan periode waktu tertentu, seperti bulanan, triwulan, atau tahunan.
- Setelah blok terbentuk, selanjutnya dipilih nilai ekstrem dengan menentukan nilai data tertinggi pada periode blok tersebut. Sampel dari nilai ekstrem ini yang akan digunakan untuk analisis.



Gambar 1. Ilustrasi pengambilan Sampel dengan Block Maxima

Dari Gambar 1 terlihat bahwa X_2 , X_7 , dan X_{11} adalah nilai ekstrem yang didapat dalam block pengamatan. Dari ilustrasi ini tergambar bahwa *Block Maxima* memerlukan data yang banyak sehingga memungkinkan teridentifikasi data nilai ekstrem yang cukup untuk dianalisis.

3. Sistem yang Dibangun



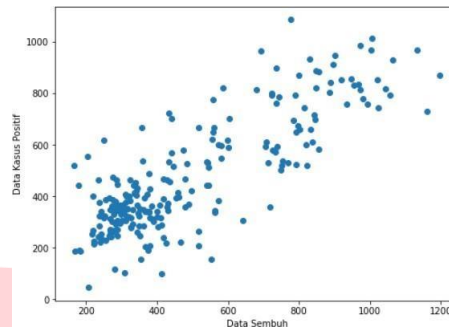
Gambar 2. Flowchart Gambaran Sistem

1. Input data, data yang digunakan adalah data time series dengan jumlah 2 variable, yaitu data kasus positif harian COVID-19 dengan melibatkan tingkat kesembuhan pasien pada Provinsi Jawa Timur mulai dari 1 Juli 2020 sampai 28 Februari 2021.
2. Melakukan analisis data secara deskriptif untuk melihat kriteria, struktur data, dan melihat korelasi antar data.
3. Melakukan preprocessing data untuk membuat data mentah menjadi data yang dapat digunakan untuk pengolahan selanjutnya. Dalam tahap preprocessing data dilakukan *missing value* pada data, pengujian stasioneritas data dengan pengujian *Dickey-Fuller* pada Persamaan (4), dan *differencing* yang terdapat pada Persamaan (5).
4. Data hasil preprocessing digunakan untuk menentukan *lag optimum* dan menentukan orde model terbaik menggunakan AIC pada Persamaan (7).
5. Melakukan pembagian data berupa data training dan data testing sebesar 90 : 10.
6. Pemodelan menggunakan metode VARMA, pada tahap ini menentukan parameter model VARMA untuk melakukan *forecasting* dan *re-differencing* data *forecast* pada Persamaan (6).
7. Menentukan rata rata nilai ekstrem dengan cara menjumlahkan data testing yang lebih besar dari data forecast, kemudian dibagi dengan banyaknya data tersebut.
8. Menentukan nilai ekstrem dengan *Block Maxima*, yaitu mengambil sampel nilai ekstrem dengan cara menentukan nilai tertinggi dari data pada setiap blok.
9. Menghitung pelanggaran estimasi nilai ekstrem dengan cara melihat berapa banyak data yang melebihi rata rata nilai esktrim dari keseluruhan data *testing*.

4. Evaluasi

4.1 Hasil Pengujian

Data yang digunakan untuk estimasi batas atas kasus harian positif COVID-19 adalah data kasus baru dan data sembuh, dengan menggunakan *Pearson Correlation* kedua data tersebut mempunyai korelasi sebesar 0.8314. Nilai korelasi berkisar antara 1 sampai -1, yang berarti hubungan kedua data cukup kuat karena mendekati nilai 1.



Gambar 3. Scatter Plot Data Kasus Positif dan Data Sembuh

4.1.1. Uji Stasioneritas

Untuk memastikan stasioneritas data, maka dilakukan pengujian *Dickey-Fuller*. Menggunakan Persamaan (4) dihasilkan bahwa data yang digunakan tidak stasioner, dengan nilai sebagai berikut:

Tabel 1. Hasil uji ADF pada dataset

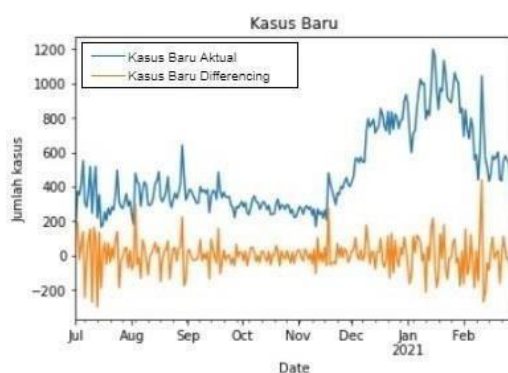
| Data | Nilai <i>p-value</i> | Nilai Signifikansi (α) | Kesimpulan |
|--|----------------------|---------------------------------|-----------------|
| Kasus Baru Positif COVID-19 Jawa Timur | 0.71 | 0.05 | Tidak stasioner |
| Sembuh Baru COVID-19 Jawa Timur | 0.45 | 0.05 | Tidak stasioner |

Tabel 2. Hasil uji ADF pada dataset dengan *First Differencing*

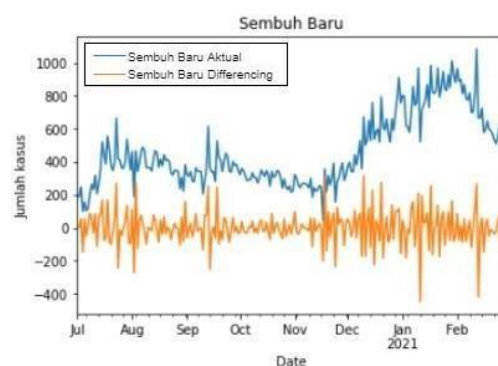
| Data | Nilai <i>p-value</i> | Nilai Signifikansi (α) | Kesimpulan |
|--|----------------------|---------------------------------|------------|
| Kasus Baru Positif COVID-19 Jawa Timur | 2×10^{-22} | 0.05 | Stasioner |
| Sembuh Baru COVID-19 Jawa Timur | 3×10^{-16} | 0.05 | Stasioner |

Dari Tabel 1, dapat disimpulkan bahwa $p - value > 5\%$, sehingga data asli yang digunakan tolak H_1 , dan menunjukkan bahwa hipotesis H_0 diterima, maka data tidak stasioner. Karena data yang diamati tidak stasioner maka perlu dilakukan *differencing*. Hasil pengolahan data yang sudah di *differencing* dengan Persamaan (6) dapat dilihat pada Tabel 2. Hasil menunjukkan bahwa kedua data yang diamati sudah stasioner.

4.1.2. Data Hasil *Preprocessing*



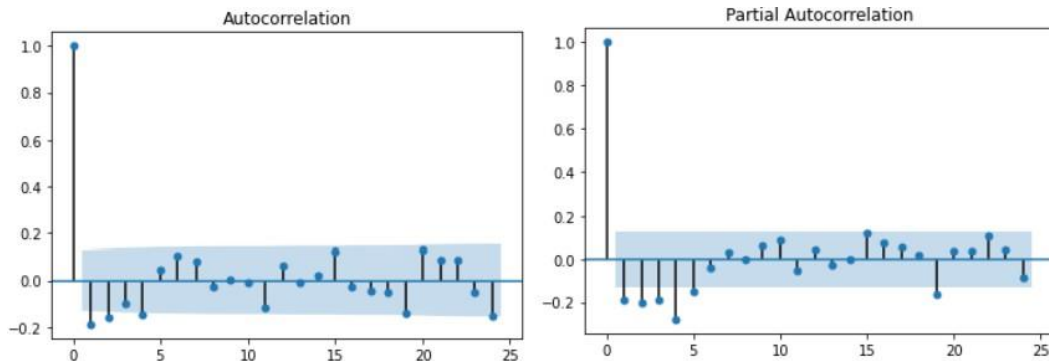
Gambar 4. Data *Differencing* Kasus Baru



Gambar 5. Data *Differencing* Sembuh Baru

Gambar 3 dan Gambar 4 menunjukkan transformasi data dari data non-stasioner menjadi data stasioner. Analisa data pada penelitian ini menggunakan stasioneritas lemah karena hanya mengecek apakah jarak setiap data mempunyai rata rata yang sama atau tidak. Data diharuskan untuk diubah ke dalam bentuk stasioner sebelum dilakukan analisis lebih lanjut.

4.1.3. Orde Model VARMA



Gambar 6. Grafik ACF

Gambar 7. Grafik PACF

Untuk menentukan panjang lag digunakan plot ACF dan PACF dari data *differencing* kasus baru dan sembuh baru. Plot ACF digunakan untuk mencari orde VAR dan plot PACF digunakan untuk mencari orde VMA. Berdasarkan plot ACF pada Gambar 6, *cut off* pertama pada lag ke-3 dapat diartikan bahwa data pada lag ke-4 dan seterusnya sudah tidak berkorelasi dengan data pada lag ke-1 hingga lag ke-2 dan data yang mempunyai jarak 3 periode tidak saling berkorelasi lagi. Sedangkan pada Gambar 7 terlihat plot PACF *cut off* pertama pada lag ke-6, menunjukkan bahwa data pada lag ke-7 dan seterusnya sudah tidak berkorelasi dengan data pada lag ke-1 hingga lag ke-5. Hal ini dapat dipertimbangkan untuk mencari lag optimal dengan nilai maksimal kombinasi orde VARMA (5,2) menggunakan AIC pada Persamaan (7).

Tabel 3. Nilai AIC

| Orde VARMA | Nilai AIC |
|-------------|-----------|
| VARMA (1,1) | 5699.03 |
| VARMA (1,2) | 5703.86 |
| VARMA (2,1) | 5694.58 |
| VARMA (2,2) | 5694.14 |
| VARMA (3,1) | 5693.94 |
| VARMA (3,2) | 5693.56 |
| VARMA (4,1) | 5690.29 |
| VARMA (4,2) | 5691.25 |
| VARMA (5,1) | 5679.05 |
| VARMA (5,2) | 5683.92 |

Tabel 3 menunjukkan pemilihan orde model VARMA, dimana model VARMA (5,1) akan digunakan pada penelitian ini karena memiliki nilai paling kecil dan merupakan model yang terbaik menurut AIC.

4.1.4. Pemodelan dengan VARMA

Dalam perhitungan kasus harian positif COVID-19, dilakukan pembagian data menjadi data *training* dan data *testing*. Karena mempertimbangkan data training sebanyak mungkin pada pemodelan VARMA, maka pembagian data *testing* sebesar 90:10, yaitu 24 data testing dari keseluruhan 243 data. Berdasarkan Tabel 3, model VARMA(5,1) yang terbentuk adalah :

$$\Delta y_t = w + \sum_{i=1}^5 \varphi_i \Delta y_{t-i} + \varepsilon_t - \theta \varepsilon_{t-1}$$

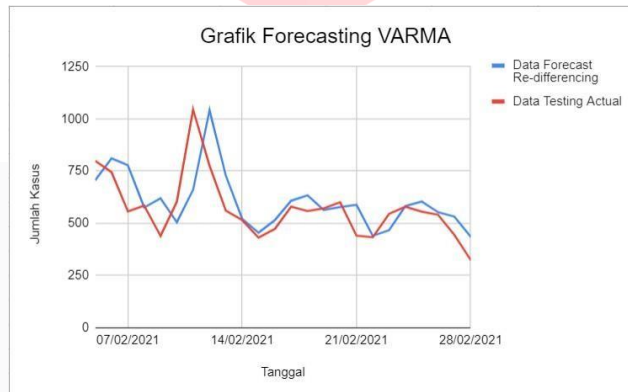
$$y_t = w + \sum_{i=1}^5 \varphi_i (y_{t-i} - y_{(t-i)-1}) + \varepsilon_t - \theta \varepsilon_{t-1} + y_{t-1}$$

Model VARMA(5,1) akan digunakan untuk menghasilkan koefisien variabel yang digunakan untuk *forecasting*, hasil tersebut dapat dilihat pada Tabel 4.

Table 4. Nilai Variabel Model VARMA

| Parameter | Hasil Persamaan Kasus Baru | Hasil Persamaan Sembuh Baru |
|-------------------|----------------------------|-----------------------------|
| Intercept | 6.0013 | 3.8373 |
| L1.Kasus Baru | -0.0350 | -0.3711 |
| L1.Sembuh Baru | -0.0117 | 0.1885 |
| L2.Kasus Baru | -0.1743 | -0.0769 |
| L2.Sembuh Baru | 0.0443 | 0.1984 |
| L3.Kasus Baru | -0.1983 | -0.1164 |
| L3.Sembuh Baru | -0.1170 | 0.0965 |
| L4.Kasus Baru | -0.1883 | -0.1287 |
| L4.Sembuh Baru | -0.1855 | -0.1285 |
| L5.Kasus Baru | -0.0334 | -0.1142 |
| L5.Sembuh Baru | -0.0927 | 0.1462 |
| L1.e(Kasus Baru) | -0.3703 | 0.4117 |
| L1.e(Sembuh Baru) | -0.0118 | -1.0047 |

4.1.4.1. Hasil *Forecast* Kasus Positif COVID-19



Gambar 8. Grafik Forecasting 90:10 VARMA(5,1)

Berdasarkan pengujian yang telah dilakukan menggunakan Persamaan (3) dengan variabel pada Tabel 4, didapatkan hasil *forecast* menggunakan model VARMA(5,1). Pada saat melakukan *forecasting* pada periode awal diasumsikan *history* data diambil dari data *training* bersesuaian dengan orde yang dilibatkan, yaitu orde VARMA (5,1). Pada grafik yang terdapat pada Gambar 8 terlihat bahwa data *forecast* yang sudah di *re-differencing* menggunakan Persamaan (6) tidak jauh berbeda dengan data *testing* aktual. Nilai MAPE pada *forecasting* 90:10, yaitu 0.155.

4.1.5. Nilai Ekstrem dengan *Block Maxima*

Pengambilan sampel nilai ekstrem kasus harian positif COVID-19 dengan metode Block Maxima dibagi menjadi 2 skenario dan menggunakan *history* data sebelumnya. Skenario pertama, membagi data ke dalam blok-blok tertentu. Data *forecast* yang digunakan sebanyak 238 data dan dibagi per-14 hari sehingga terdapat 17 blok. Skenario kedua, keseluruhan 238 data *forecast* tidak dibagi perblok jadi hanya mempunyai satu nilai ekstrem. Kemudian memilih data *forecast* yang lebih kecil dari data aktual *testing*. Setelah itu, hitung rata-rata data tersebut dengan cara :

$$\bar{x} = \frac{\sum(y_t < \hat{y}_t)}{n}$$

- \bar{x} = rata-rata data
- y_t = nilai *forecast* pada saat t
- \hat{y}_t = data aktual *testing*
- n = banyak data yang diamati

Berdasarkan perhitungan yang telah dilakukan, rata-rata nilai ekstrem dapat dilihat pada Tabel 5. Rata-rata nilai ekstrem digunakan untuk acuan apakah masih ada data dengan nilai yang melebihi nilai rata rata pada Tabel 5 atau tidak.

Tabel 5. Rata-rata Nilai Ekstrem Data Testing

| Skenario | Blok | Rata-rata Nilai Ekstrem |
|----------|-----------|-------------------------|
| 1 | 1 | 385 |
| | 2 | 332 |
| | 3 | 417 |
| | 4 | 411 |
| | 5 | 379 |
| | 6 | 389 |
| | 7 | 303 |
| | 8 | 296 |
| | 9 | 279 |
| | 10 | 308 |
| | 11 | 464 |
| | 12 | 745 |
| | 13 | 842 |
| | 14 | 950 |
| | 15 | 982 |
| | 16 | 776 |
| | 17 | 573 |
| 2 | Tidak ada | 533 |

4.1.5. Nilai Pelanggaran Estimasi Nilai Ekstrem

Hasil dari nilai ekstrem yang optimal harus lebih tinggi dari data kasus harian positif COVID-19. Data yang melewati rata-rata nilai ekstrem disebut pelanggaran. Hasil perhitungan dapat dilihat pada Tabel 6 dan Tabel 7. Rata-rata estimasi pelanggaran nilai ekstrem dari 238 data yang dibagi menjadi 17 blok adalah 0.2983, sedangkan rata-rata estimasi pelanggaran nilai ekstrem yang tidak dibagi perblok adalah 0.3403.

Tabel 6. Nilai Pelanggaran Skenario 1

| Blok | Jumlah Pelanggaran | Estimasi Pelanggaran Nilai Ekstrem |
|--|--------------------|------------------------------------|
| 1 | 3 | 0.2143 |
| 2 | 3 | 0.2143 |
| 3 | 6 | 0.4286 |
| 4 | 4 | 0.2857 |
| 5 | 4 | 0.2857 |
| 6 | 1 | 0.0714 |
| 7 | 6 | 0.4286 |
| 8 | 2 | 0.1429 |
| 9 | 5 | 0.3571 |
| 10 | 4 | 0.2857 |
| 11 | 4 | 0.2857 |
| 12 | 8 | 0.5714 |
| 13 | 3 | 0.2143 |
| 14 | 7 | 0.5 |
| 15 | 5 | 0.3571 |
| 16 | 3 | 0.2143 |
| 17 | 3 | 0.2143 |
| Rata-rata Pelanggaran Nilai Ekstrem | | 0.2983 |

Tabel 7. Nilai Pelanggaran Skenario 2

| Jumlah Pelanggaran | Estimasi Pelanggaran Nilai Ekstrem |
|--------------------|------------------------------------|
| 81 | $\frac{81}{238} = \mathbf{0.3403}$ |

4.2 Analisis Hasil Pengujian

Setelah melakukan pengujian stasioneritas data dengan *Dickey-Fuller* pada tahap *preprocessing* data kasus harian positif COVID-19 di Jawa Timur, maka data hasil *preprocessing* dapat digunakan untuk menentukan *lag optimum* dengan menggunakan ACF dan PACF. Pemodelan dilakukan dengan model *Vector Autoregressive Moving Average* (VARMA), kemudian pemilihan orde model dengan AIC dimana model VARMA(5,1) memiliki nilai paling kecil dan merupakan model yang terbaik. Setelah itu, dilakukan pembagian data menjadi data *training* dan data *testing* sebesar 90:10, yaitu 24 data testing dari keseluruhan 243 data. Model VARMA(5,1) digunakan untuk menghasilkan koefisien variabel untuk peramalan. Sedangkan mengestimasi lonjakan kasus yang terjadi pada rentang tanggal 1 Juli 2020 sampai 28 Februari 2021 digunakan *Block Maxima*. Untuk menentukan rata-rata dan nilai ekstrem digunakan sebanyak 238 data peramalan yang sudah di *re-differencing*, karena mempertimbangkan data *testing* sebanyak mungkin pada metode *Block Maxima*. Estimasi pelanggaran nilai ekstrem menggunakan *Block Maxima* dengan model terbaik VARMA(5,1) dan hasil MAPE 0.15 memberikan forecasting yang cukup baik untuk pelanggaran nilai ekstrem skenario 1 dengan nilai 0.2983, sedangkan pada skenario 2 adalah 0.3403. Sehingga jika terjadi lonjakan pada skenario 1 maka 70% mencapai 573 kasus. Sedangkan pada skenario 2 jika terjadi lonjakan maka 66% mencapai 533 kasus.

5. Kesimpulan

Penelitian ini berfokus pada estimasi pelanggaran nilai ekstrem kasus positif harian COVID-19 di provinsi Jawa Timur dengan melibatkan tingkat kesembuhan pasien berdasarkan pada data *time series*. Pemodelan dilakukan dengan model *Vector Autoregressive Moving Average* (VARMA), kemudian untuk mengestimasi lonjakan kasus yang terjadi pada rentang tanggal 1 Juli 2020 sampai 28 Februari 2021 digunakan *Block Maxima*. Pada hasil penelitian dapat dilihat bahwa estimasi nilai pelanggaran skenario 1 lebih baik dari skenario 2, karena pengambilan blok setiap 14 hari lebih presisi dan secara umum seseorang yang terinfeksi COVID-19 sudah tidak menular setelah 10-14 hari sejak terdiagnosis positif COVID-19. Sehingga untuk mengantisipasi lonjakan kasus skenario 1 pada bulan Februari pemerintah atau *public health* harus mempersiapkan dan mempertimbangkan adanya 573 kasus positif pada bulan Februari.

Terdapat beberapa saran yang diberikan untuk penelitian selanjutnya adalah menambah ukuran data kasus positif COVID-19, dapat dioptimalkan dengan *machine learning* sehingga diperoleh akurasi model VARMA yang lebih baik. Dengan mengevaluasi hasil pelanggaran, dapat mempertimbangkan metode lain untuk nilai ekstrem, yaitu *Peak-Over Threshold* (POT) dan *Value at Risk* (VaR).

REFERENSI

- [1] AL-Nasser, A. H., & Abdullah, L. T. (2021). The Estimators of Vector Autoregressive Moving Average Model VARMA of Lower Order: VARMA (0,1), ARMA (1,0), VARMA (1,1), VARMA (1,2), VARMA (2,1), VARMA (2,2) with Forecasting. *Journal of Physics: Conference Series*.
- [2] Amalia, F., Saepudin, D., & Rohmawati, A. A. (2019). Ukuran Risiko Model Vector Autoregressive pada Harga Saham. *e-Proceeding of Engineering : Vol.6*, 2576.
- [3] Anggraeni, W., & Dewi, L. K. (2008). Peramalan Menggunakan Metode Vector Autoregressive Moving Average (VARMA). *Jurnal Ilmiah Teknologi Informasi*.
- [4] Athanasopoulos, G. (2006). A Complete VARMA Modelling Procedure based on SCMs. *PhD thesis, Monash University, Melbourne, Australia*.
- [5] Fotheringham, A. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. *International Journal of Geographical Information Science*.
- [6] Liu, K., Chen, Y., Lin, R., & Kunyuan, H. (2020). Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients. *Journal of Infection*.
- [7] Lutkepohl, H. (2004). Forecasting with VARMA Models. *Department of Economics, European University Institute, Firenze, Italy*.
- [8] Mallapaty, S. (2020). What the cruise-ship outbreaks reveal about COVID-19. *Nature*, 580(7801): 18, 2020 04.
- [9] Nasom, G. (2006). Stationary and non-stationary time series. (S. C. H. Mader, Ed.) *Statistics in Volcanology*, 129 - 142.
- [10] Rahmayani, D., & Sutikno. (2019). Analisis Curah Hujan Ekstrem Non-Stasioner dengan Pendekatan Block Maxima di Surabaya dan Mojokerto. *JURNAL SAINS DAN SENI ITS Vol. 8, No. 2*, 2337-3520.
- [11] Sunaryo, T. (2007). *Manajemen risiko finansial*. Penerbit Salemba. Retrieved from World Health Organization: <https://www.who.int/health-topics/>
- [12] Wijaya, C. Y. (2019, September 5). *Mengenal Time Series dan Struktur yang Membentuknya*. Retrieved from medium.com: www.medium.com
- [13] Wutsqa, U. D., & Suhartono. (2010). Peramalan Deret Waktu Multivariat Seasonal Pada data Pariwisata Dengan Model VAR-GSTAR. *Jurnal Ilmu Dasar*.