

## Peramalan Kasus Positif COVID-19 di Jawa Timur menggunakan Metode Hybrid ARIMA-LSTM

Rowan<sup>1</sup>, Lailil Muflikhah<sup>2</sup>, Imam Cholissodin<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya  
Email: <sup>1</sup>albertusro1@student.ub.ac.id, <sup>2</sup>lailil@ub.ac.id, <sup>3</sup>imamcs@ub.ac.id

### Abstrak

COVID-19 (*Coronavirus Disease 2019*) adalah jenis penyakit baru yang terkait dengan keluarga virus yang sama dengan *Severe Acute Respiratory Syndrome* (SARS) dan beberapa jenis virus flu biasa. Seiring meningkatnya kasus positif COVID-19, sumber daya yang dibutuhkan dalam penanganan kasus COVID-19 juga bertambah. Untuk mengatasi permasalahan ini diperlukan tindakan antisipasi agar sumber daya yang dibutuhkan dalam penanganan COVID-19 seperti tenaga kesehatan dan obat-obatan dapat tersedia sebelum kasus positif melonjak. Pada penelitian ini metode yang digunakan adalah metode *hybrid Autoregressive Integrated Moving Average* (ARIMA) dan *Long Short-Term Memory* (LSTM). Model ARIMA-LSTM dibangun dengan menggabungkan model ARIMA (2,1,2) dengan model LSTM yang memiliki 4 *hidden state* dan 1 *layer*. Model ARIMA bertugas untuk meramalkan nilai *trend* dari data deret waktu sedangkan model LSTM bertugas untuk melengkapi peramalan model ARIMA dengan meramalkan nilai *residual* dari deret waktu. Berdasarkan hasil pengujian, model ARIMA-LSTM memiliki akurasi yang tinggi terutama pada peramalan jangka pendek dengan MAPE sebesar 1,8 persen untuk kasus peramalan 3 hari ke depan.

**Kata kunci:** peramalan, arima, lstm, COVID-19

### Abstract

COVID-19 (*Coronavirus Disease 2019*) is a new type of disease related to the same virus family as *Severe Acute Respiratory Syndrome* (SARS) and several strains of the common cold virus. Along with the increase of positive cases, the resources needed in handling COVID-19 cases also increase. To overcome this problem, anticipatory measures are needed so that the resources needed in handling COVID-19 such as health workers and medicines will be available before positive cases spike. In this study, the method used is the hybrid *Autoregressive Integrated Moving Average* (ARIMA) and *Long Short-Term Memory* (LSTM) method. The ARIMA-LSTM model is built by combining the ARIMA (2,1,2) model with the LSTM model which has 4 hidden states and 1 layer. ARIMA model is used to predict the trend value from time series data while LSTM model is used to complete the ARIMA model forecasting by predicting the time series residual value. Based on testing, the ARIMA-LSTM model achieved high accuracy, especially in short-term forecasting with an error rate of 1.8 percent for forecasting cases for the next 3 days.

**Keywords:** forecasting, arima, lstm, COVID-19

### 1. PENDAHULUAN

COVID-19 (*Coronavirus Disease 2019*) adalah jenis penyakit baru yang terkait dengan keluarga virus yang sama dengan *Severe Acute Respiratory Syndrome* (SARS) dan beberapa jenis virus flu biasa. Penyakit ini pertama kali ditemukan pada Desember 2019 di Wuhan, Ibukota Provinsi Hubei China. Virus ini dapat dengan mudah ditularkan dari manusia ke

manusia dengan media droplet yang dikeluarkan manusia ketika batuk ataupun bersin. Pada 09 Juli 2020, *World Health Organization* menyatakan wabah COVID-19 sebagai pandemi yang secara global telah mengakibatkan 559.694 kematian dan 10.509.505 kasus yang dikonfirmasi. Indonesia pertama kali mengkonfirmasi kasus COVID-19 pada senin 2 Maret 2020.

Seiring meningkatnya kasus positif

COVID-19, sumber daya yang dibutuhkan dalam penanganan kasus COVID-19 juga bertambah. Hal ini menjadi salah satu tantangan dalam upaya pemerintah menangani pandemi ini, salah satu sumber daya yang paling dibutuhkan saat ini adalah sumber daya manusia yaitu tenaga kesehatan. Menurut Ketua Tim Mitigasi Ikatan Dokter Indonesia (IDI) Mohammad Adib Khumaidi, ditengah lonjakan kasus COVID-19 yang eksponensial, permasalahan utama tenaga kesehatan bukanlah kelelahan melainkan minimnya tenaga pengganti dikala banyak tenaga kesehatan lain turut terpapar (Newswire, 2021). Ketidakmampuan fasilitas kesehatan untuk menampung penderita, bukan hanya karena ruangan yang tidak mencukupi, permasalahan utamanya merupakan ada atau tidaknya tenaga kesehatan yang menjalankannya. Tidak hanya itu, tantangan yang lain yakni kapasitas sumber daya manusia yang belum menyeluruh di seluruh Indonesia.

Untuk mengatasi permasalahan ini diperlukan tindakan antisipasi agar sumber daya yang dibutuhkan dalam penanganan COVID-19 seperti tenaga kesehatan dan obat-obatan dapat tersedia sebelum kasus positif melonjak. Penelitian ini ditujukan untuk membantu meramalkan peningkatan kasus positif COVID-19 dengan menggunakan metode statistik. Terdapat banyak metode yang dapat digunakan dalam melakukan peramalan kasus positif COVID-19. Pada penelitian ini metode yang digunakan adalah metode *hybrid Autoregressive Integrated Moving Average* (ARIMA) dan *Long Short-Term Memory* (LSTM).

Metode ARIMA merupakan metode yang digunakan untuk melakukan peramalan pada data deret waktu dengan memanfaatkan variabel terikat dari data. Metode ini cocok digunakan untuk melakukan peramalan karena dianggap *flexible* (mengikuti pola data), memiliki akurasi peramalan yang tinggi, serta dapat diandalkan untuk peramalan jangka pendek (Putri & Anggraeni, 2018). Namun, jika digunakan dalam peramalan jangka panjang metode ARIMA akan mengalami penurunan akurasi. Untuk mengatasi hal ini, pada penelitian ini metode ARIMA akan dihibridisasi dengan metode LSTM. LSTM menjadi pilihan terbaik untuk mengatasi kekurangan dari metode ARIMA karena LSTM mengimplementasikan *cell state* yang membuat metode ini dapat menyimpan informasi dalam jangka waktu yang panjang. Kekurangan dari LSTM sama seperti sebagian besar *recurrent*

*neural network* adalah waktu pelatihan yang lama serta kombinasi parameter yang kompleks. Pada penelitian ini, LSTM akan memproses residual dari metode ARIMA sehingga beban kerja dari metode LSTM tidak terlalu berat dan akan mempersingkat waktu pelatihan serta kompleksitas parameter.

## 2. TINJAUAN PUSTAKA

### 2.1 COVID-19

COVID-19 (*Coronavirus Disease 2019*) adalah virus yang menyerang sistem pernapasan manusia. Virus ini dapat menyebabkan gangguan pernapasan ringan, infeksi paru-paru serius, hingga kematian. *Severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2) yang juga dikenal sebagai virus corona, adalah jenis baru dari coronavirus yang ditularkan ke manusia. Virus ini bisa menyerang siapa saja, mulai dari orang tua (lansia), dewasa, anak-anak hingga ibu hamil dan menyusui. COVID-19 pertama kali dikonfirmasi di kota Wuhan, China pada akhir bulan Desember tahun 2019. Virus ini memiliki tingkat penularan yang sangat cepat dan tersebar ke hampir seluruh negara, termasuk Indonesia hanya dalam waktu beberapa bulan (Alodokter, 2022).

### 2.2 ARIMA

Metode *Autoregressive Integrated Moving Average* (ARIMA) adalah salah satu model regresi linier paling populer untuk meramalkan deret waktu statis. ARIMA juga dikenal sebagai metode deret waktu Box-Jenkins. Metode ini dikembangkan oleh George Box dan Gwilym Jenkins pada tahun 1970. Metode ARIMA sangat baik digunakan untuk meramalkan data deret waktu dalam jangka pendek, sedangkan akurasiya menurun dalam jangka waktu yang cukup lama. Metode ARIMA tidak menggunakan variabel bebas dari data sama sekali, melainkan hanya variabel terikatnya saja. Model ARIMA biasanya dilambangkan ARIMA ( $p, d, q$ ) di mana  $p$  adalah urutan (jumlah penundaan) dari model autoregresif,  $d$  adalah jumlah *differencing* (berapa kali data telah dikurangi dari nilai masa lalu) dan  $q$  adalah nilai *moving-average* dari model. Persamaan 1 merupakan bentuk umum dari model ARIMA (Santoso, 2009).

$$\Delta y_t = \varphi_1 \Delta y_{t-1} + \varphi_2 \Delta y_{t-2} + \varphi_p \Delta y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_p e_{t-p} \quad (1)$$

Keterangan:

- $t$  = Koefisien waktu
- $Y_t$  = Nilai series pada waktu  $t$
- $Y_{t-1}, Y_{t-2}$  = Nilai lampau series
- $e_{t-1}, e_{t-2}$  = Nilai lampau residual
- $\theta_1, \theta_p, \varphi_1, \varphi_p$  = Koefisien model

### 2.3 Machine Learning & Deep Learning

*Machine learning* (ML) adalah mesin yang dikembangkan untuk belajar sendiri tanpa bimbingan secara langsung dari pengguna. Pembelajaran mesin dikembangkan di atas bidang lain seperti statistik, matematika, dan data mining sehingga mesin dapat belajar dengan menganalisis data tanpa perlu pemrograman ulang atau instruksi (Dicoding, 2020).

*Deep learning* adalah salah satu cabang dari *machine learning* yang terdiri dari algoritma pemodelan abstraksi tingkat tinggi pada data menggunakan sekumpulan fungsi transformasi non-linear yang ditata berlapis-lapis dan mendalam (Cholissodin et al., 2019). Pada dasarnya, deep learning adalah jaringan saraf dengan tiga atau lebih lapisan JST, memungkinkannya untuk belajar dan beradaptasi dengan sejumlah besar data dan untuk memecahkan berbagai masalah yang sulit dipecahkan oleh algoritma pembelajaran mesin lainnya (Dicoding, 2021).

### 2.4 LSTM

Long Short-Term Memory (LSTM) adalah evolusi dari arsitektur RNN, arsitektur ini pertama kali diperkenalkan oleh Hochreiter dan Schmidhuber pada tahun 1997. Termasuk dalam jenis RNN, LSTM memiliki modul berulang seperti rangkaian unit jaringan saraf (Faurina, Winduratna, & Nugroho, 2018). Arsitektur LSTM dikembangkan untuk menjadi solusi dari masalah *vanishing gradient* yang ditemukan pada RNN konvensional. *Vanishing gradient* adalah sebuah kondisi yang disebabkan oleh nilai gradien yang selalu mengecil hingga layer terakhir sehingga nilai bobot tidak berubah dan menyebabkan model tidak pernah memperoleh hasil yang lebih baik atau tidak konvergen. LSTM memiliki proses yang berbeda dengan RNN konvensional karena pada LSTM terdapat tambahan sinyal yang diberikan dari satu langkah waktu ke langkah waktu berikutnya, sinyal ini dinamakan *cell state* (Priyono, 2018).

Terdapat beberapa tahapan dalam

pemrosesan data pada LSTM:

#### 1. Forget Gate ( $f_t$ )

Langkah pertama dimulai melalui komponen *forget gate*. Pada bagian ini, informasi yang kurang penting atau kurang berarti untuk kasus yang sedang ditangani dihilangkan menggunakan fungsi sigmoid. Data  $x_t$  adalah input dan  $h_{t-1}$  adalah vektor *hidden state* dalam timestep sebelumnya.

Persamaan 2 merupakan perhitungan nilai *forget gate*.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

Keterangan:

- $W_f$  = bobot *forget gate*
- $h_{t-1}$  = vektor *hidden state* periode sebelumnya
- $x_t$  = data masuk
- $b_f$  = bias
- $f$  = *forget gate*

#### 2. Input Gate ( $i_t$ )

Langkah berikutnya, informasi diolah melalui komponen *input gate* ( $i_t$ ). Proses ini akan memilih dan menentukan informasi apa saja yang akan diperbarui ke bagian *cell state* menggunakan fungsi aktivasi sigmoid.

Persamaan 3 merupakan perhitungan nilai *input gate*.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

Keterangan:

- $W_i$  = bobot *input gate*
- $h_{t-1}$  = vektor *hidden state* dalam timestep sebelumnya
- $x_t$  = data masuk
- $b_i$  = bias
- $i$  = *input gate*

#### 3. Kandidat Cell State ( $\hat{c}_t$ )

Pada langkah ini akan ada proses untuk membuat kandidat *cell state* baru menggunakan fungsi aktivasi tanh yang akan ditambahkan pada bagian *cell state*.

Persamaan 4 merupakan perhitungan nilai kandidat *cell state*.

$$\hat{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

Keterangan:

- $W_c$  = bobot *cell state*
- $h_{t-1}$  = vektor *hidden state* dalam timestep sebelumnya

$x_t$  = data masuk  
 $b_c$  = bias  
 $c$  = cell state

#### 4. Cell State ( $c_t$ )

Langkah selanjutnya adalah memperbarui nilai cell state lama menjadi cell state baru.

Persamaan 5 merupakan perhitungan memperbarui cell state.

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \hat{c}_t \quad (5)$$

Keterangan:

$f_t$  = nilai forget gate  
 $c_{t-1}$  = cell state lama  
 $i_t$  = nilai input gate  
 $\hat{c}_t$  = nilai kandidat cell state baru

#### 5. Output Gate ( $o_t$ )

Langkah terakhir terjadi pada output gate dengan menggunakan fungsi aktivasi sigmoid sehingga menghasilkan output pada hidden state serta memasukkan cell state pada tanh.

Persamaan 6 merupakan perhitungan output gate.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

Keterangan:

$W_o$  = bobot output gate  
 $h_{t-1}$  = vektor hidden state dalam timestep sebelumnya  
 $x_t$  = data masuk  
 $b_o$  = bias  
 $o$  = output gate

#### 6. Hidden State ( $h_t$ )

Setelah menghasilkan nilai output sigmoid dan nilai output tanh kedua hasil aktivasi tersebut dilakukan perkalian.

Persamaan 7 merupakan perhitungan hidden state.

$$h_t = o_t \cdot \tanh(c_t) \quad (7)$$

Keterangan:

$o_t$  = nilai output gate  
 $c_t$  = cell state

### 3. METODOLOGI PENELITIAN

#### 3.1. Dataset

Pengumpulan data dilakukan dengan melakukan request melalui API

data.covid19.go.id. Data yang didapatkan berupa teks dengan format json yang mengandung informasi daftar perkembangan COVID-19 di Jawa Timur. Data yang sudah dimiliki kemudian diubah formatnya menjadi data time series/deret waktu agar dapat digunakan pada proses peramalan.

Data yang digunakan adalah data harian COVID-19 di Indonesia khususnya daerah Jawa Timur yang didapatkan melalui API data.covid19.go.id. Data ini terdiri dari 9 atribut yaitu tanggal, kasus, meninggal, sembuh, dirawat or diisolasi, akumulasi kasus, akumulasi sembuh, akumulasi meninggal, dan akumulasi dirawat or diisolasi. Data yang akan digunakan dalam percobaan ini adalah data time series hasil preprocessing data harian COVID-19 Jawa Timur, atribut yang digunakan pada penelitian ini adalah atribut tanggal dengan tipe data datetime dan kasus dengan tipe data integer. Jumlah data adalah sebanyak 744 baris data.

#### 3.2 ARIMA-LSTM

ARIMA dan LSTM masing-masing memiliki kelebihan dan kekurangan. ARIMA mengalami kesulitan ketika memproses data time series yang bersifat non-linear. Sedangkan jaringan syaraf tiruan dapat bekerja dengan baik dengan data linear maupun non-linear. Namun dibutuhkan waktu pelatihan yang lama serta parameter yang sangat banyak untuk mencapai hasil maksimal. Berdasarkan faktor-faktor yang disebutkan sebelumnya, sebuah model gabungan akan dirancang sehingga kelebihan dari kedua model tersebut dapat dimanfaatkan sekaligus mengurangi/mengatasi kekurangan yang ada pada masing-masing model.

Data time series dapat dipecah menjadi beberapa komponen yaitu trend dan residual, dimana trend adalah komponen linear dari time series dan residual adalah komponen non-linearnya. ARIMA akan digunakan untuk memproses komponen linear yaitu trend, lalu LSTM akan memproses residual.

Persamaan 8 merupakan perhitungan metode ARIMA-LSTM.

$$prediksi = ARIMA[trend] + LSTM[residual] \quad (8)$$

Proses pelatihan model gabungan ARIMA-LSTM dapat dibagi menjadi 2 bagian utama yaitu proses pelatihan model ARIMA, serta proses pelatihan model LSTM. Proses pelatihan model LSTM dilakukan setelah pelatihan model ARIMA selesai, hal ini dilakukan karena model

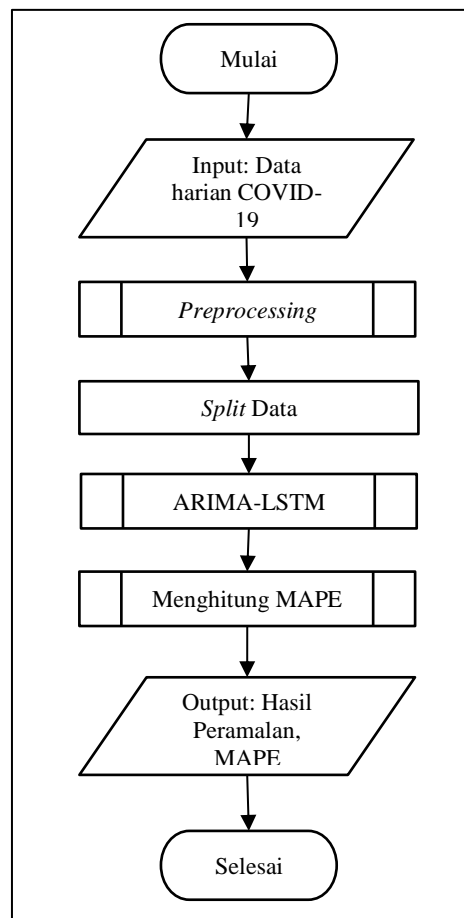
LSTM membutuhkan *residual error* dari hasil pelatihan model ARIMA sebagai data. Proses pelatihan model ARIMA-LSTM adalah sebagai berikut:

1. Melakukan proses *preprocessing* dan uji stasioneritas pada data yang akan dijadikan data masukan pada proses pelatihan model ARIMA.
2. Menganalisa grafik ACF dan PACF dari data untuk menentukan nilai parameter yang sesuai pada model ARIMA.
3. Melakukan proses pelatihan model ARIMA.
4. Mencari nilai *residual error* dari hasil pelatihan model ARIMA.
5. Melakukan *preprocessing* terhadap *residual* model ARIMA sehingga dapat dijadikan data masukan bagi pelatihan model LSTM.
6. Membangun dan melatih model LSTM menggunakan data *residual* yang sudah melewati tahap *preprocessing*.

Setelah model selesai melalui tahap pelatihan, maka tahap selanjutnya adalah peramalan dan pengujian menggunakan data uji. Proses peramalan dan pengujian model ARIMA-LSTM adalah sebagai berikut:

1. Melakukan peramalan kasus positif COVID-19 sebanyak  $k$  hari ke depan menggunakan model ARIMA.
2. Melakukan peramalan *residual* sebanyak  $k$  hari ke depan menggunakan model LSTM.
3. Menggabungkan hasil peramalan kedua model tersebut dengan cara menjumlahkan hasil peramalan model ARIMA dan LSTM.
4. Melakukan perhitungan MAPE antara data uji sebanyak  $k$  hari ke depan dengan hasil peramalan model ARIMA-LSTM.

Diagram alir proses perancangan metode ARIMA-LSTM dapat dilihat pada Gambar 1.



Gambar 1 Diagram Alir Perancangan ARIMA-LSTM

#### 4. HASIL DAN PEMBAHASAN

Terdapat 3 jenis skenario pengujian yaitu pengujian nilai parameter model ARIMA-LSTM, pengujian rentang data latih, dan pengujian jumlah peramalan.

##### 4.1 Pengujian Nilai Parameter ARIMA-LSTM

Parameter yang akan diuji pada proses ini adalah parameter  $p$ ,  $d$ , dan  $q$  pada ARIMA serta *hidden state*, dan jumlah *layer* model LSTM pada proses pelatihan. Kombinasi nilai yang akan digunakan pada proses pengujian ini adalah (1,2) untuk parameter  $p$  dan (0,1,2) untuk parameter  $q$ , sedangkan nilai *differencing*  $d$  akan tetap bernilai (1) untuk setiap pengujian dimana  $p$  adalah urutan (jumlah penundaan) dari model autoregresif,  $d$  adalah jumlah *differencing* (berapa kali data telah dikurangi dari nilai masa lalu) dan  $q$  adalah nilai moving-average dari model. Nilai *hidden state* LSTM yang akan diuji adalah (1,4,8) dan jumlah *layer* LSTM yang akan diuji adalah (1,2) dengan *epoch* sebanyak 200 untuk setiap pengujian. Berdasarkan

skenario tersebut, terdapat 36 kombinasi parameter. Tabel 1 menunjukkan hasil pengujian nilai parameter ARIMA-LSTM.

Tabel 1 Hasil Pengujian Nilai Parameter ARIMA-LSTM

<i>p</i>	<i>d</i>	<i>q</i>	<i>Hidden State</i>	<i>Layer</i>	MAPE (ARIMA)	MAPE (ARIMA-LSTM)
1	1	0	1	1	17,2425	3,03945
1	1	0	1	2	17,2425	6,78326
1	1	0	4	1	17,2425	11,3768
1	1	0	4	2	17,2425	11,3768
1	1	0	8	1	17,2425	11,3768
1	1	0	8	2	17,2425	11,3768
1	1	1	1	1	23,3452	1,86002
1	1	1	1	2	23,3452	6,80817
1	1	1	4	1	23,3452	10,9071
1	1	1	4	2	23,3452	10,9071
1	1	1	8	1	23,3452	10,9071
1	1	1	8	2	23,3452	10,9071
1	1	2	1	1	23,3692	1,83512
1	1	2	1	2	23,3692	6,83220
1	1	2	4	1	23,3692	11,0018
1	1	2	4	2	23,3692	11,0018
1	1	2	8	1	23,3692	11,0018
1	1	2	8	2	23,3692	11,0018
2	1	0	1	1	18,0024	3,69301
2	1	0	1	2	18,0024	5,63501
2	1	0	4	1	18,0024	10,0166
2	1	0	4	2	18,0024	10,0166
2	1	0	8	1	18,0024	10,0166
2	1	0	8	2	18,0024	10,0166
2	1	1	1	1	23,3932	1,81109
2	1	1	1	2	23,3932	6,85624
2	1	1	4	1	23,3932	11,0258
2	1	1	4	2	23,3932	11,0258
2	1	1	8	1	23,3932	11,0258
2	1	1	8	2	23,3932	11,0258
2	1	2	1	1	17,4166	9,83792
2	1	2	1	2	17,4166	2,08857
2	1	2	4	1	17,4166	1,86901
2	1	2	4	2	17,4166	1,86901
2	1	2	8	1	17,4166	1,86901
2	1	2	8	2	17,4166	1,86901

Berdasarkan hasil pengujian pada Tabel 1 dapat disimpulkan beberapa hal. Pertama, parameter model ARIMA yang memiliki nilai MAPE lebih tinggi tidak selalu memiliki nilai MAPE yang lebih tinggi juga setelah digabungkan dengan model LSTM, hal ini dapat disebabkan karena nilai *residual* model ARIMA yang tidak jauh berbeda antara setiap model uji sehingga tidak mempengaruhi performa model

setelah digabungkan dengan LSTM. Kedua, tingkat kompleksitas parameter model LSTM tidak berbanding lurus dengan akurasi dan nilai MAPE yang dihasilkan, hal ini disebabkan karena semakin kompleks sebuah model LSTM maka akan semakin besar kemungkinan terjadinya *overfitting* yang mengakibatkan nilai loss saat pengujian cenderung tinggi. Ketiga, berdasarkan data nilai MAPE pada Tabel 1 dapat dilihat bahwa parameter ARIMA yang memiliki performa paling konsisten dan optimal adalah ARIMA(2,1,2) dan konvergen pada nilai *hidden state* dan jumlah layer LSTM 4 dan 1.

#### 4.2 Pengujian Data Latih

Tahap pengujian ini dilakukan untuk mencari jumlah dan rentang data latih yang optimal digunakan untuk melatih model serta untuk mengetahui pengaruh dari jumlah data yang digunakan pada proses pelatihan. Berdasarkan perancangan pengujian pada bab sebelumnya, pengujian data latih dilakukan dengan beberapa variasi yaitu data dari bulan Maret 2020 sampai November 2021, data dari bulan Januari 2021 sampai Juni 2021, data dari bulan Agustus 2021 sampai Januari 2022, data dari bulan Maret 2020 sampai Februari 2022, dan data dari bulan Maret 2020 sampai Juni 2021. Jumlah hasil peramalan yang digunakan untuk tahap pengujian ini adalah sebanyak 3 hari ke depan. Pengujian ini dilakukan berdasarkan konsep *Nested Cross Validation* (NCV). Model ARIMA-LSTM yang digunakan pada tahap pengujian data latih adalah model yang dianggap memiliki hasil paling optimal berdasarkan tahap pengujian parameter model ARIMA-LSTM. Tabel 2 menunjukkan hasil pengujian data latih.

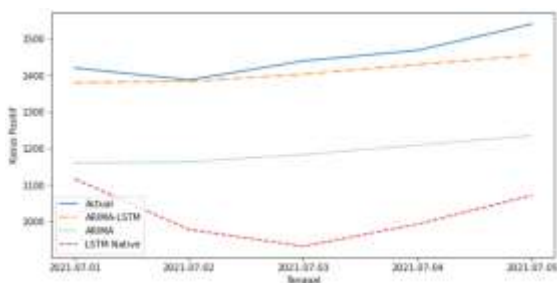
Tabel 2 Hasil Pengujian Data Latih

No.	Data Latih	MAPE
1	Maret 2020 - November 2021	8,18694
2	Januari 2021 - Juni 2021	14,16436
3	Agustus 2021 - Januari 2022	56,64748
4	Maret 2020 - Februari 2022	19,01145
5	Maret 2020 - Juni 2021	1,86901

Berdasarkan hasil pengujian data latih pada Tabel 2 dapat disimpulkan bahwa data latih dengan rentang dari bulan Maret 2020 sampai

Juni 2021 memiliki nilai MAPE yang jauh lebih kecil dibanding rentang data latih lainnya. Data latih dengan rentang bulan Agustus 2021 sampai Januari 2022 memiliki hasil evaluasi paling buruk, hal ini disebabkan karena fluktuasi data yang sangat tinggi pada rentang data tersebut.

Gambar 2 merupakan grafik perbandingan hasil peramalan metode ARIMA-LSTM dengan data aktual serta perbandingan hasil peramalan ARIMA-LSTM dengan model ARIMA secara individual. Model ARIMA-LSTM yang digunakan adalah model yang memiliki hasil optimal pada tahap pengujian. Data latih yang digunakan untuk melatih model adalah data dari bulan Maret 2020 hingga Juni 2021. Jumlah peramalan yang ditampilkan pada grafik adalah sebanyak 5 hari.



Gambar 2 Perbandingan Hasil Peramalan ARIMA-LSTM dengan Data Aktual

### 4.3 Pengujian Jumlah Peramalan

Proses pengujian akan dilakukan dengan variasi jumlah peramalan dari satu hingga sepuluh hari ke depan. Data latih yang digunakan pada proses pengujian yaitu data dari bulan Maret 2020 hingga Juni 2021. Model yang akan digunakan adalah model dengan hasil pengujian terbaik pada tahap sebelumnya. Tabel 3 merupakan tabel hasil pengujian jumlah peramalan.

Tabel 3 Hasil Pengujian Jumlah Peramalan

No.	Jumlah Peramalan (Hari)	MAPE
1	1	2,81690
2	2	5,55408
3	3	1,86901
4	4	2,06592
5	5	2,75663
6	6	5,29645

7	7	10,37759
8	8	14,09707
9	9	16,86730
10	10	18,19354

Hasil pengujian jumlah peramalan pada Tabel 3 menunjukkan bahwa nilai MAPE cenderung meningkat seiring bertambahnya jumlah peramalan yang dilakukan oleh model, hasil ini sama seperti hasil penelitian terdahulu mengenai peramalan batas penglihatan menggunakan ARIMA (Salman & Kanigoro, 2021). Berdasarkan hasil pengujian, nilai MAPE terkecil adalah pada saat jumlah pengujian 3 hari dan 4 hari, lalu meningkat cukup drastis saat melakukan pengujian untuk 7 hari dan seterusnya. Berdasarkan data ini dapat disimpulkan bahwa model akurat bila digunakan untuk meramalkan data jangka pendek dalam kasus ini 3 sampai 4 hari ke depan.

## 5. KESIMPULAN

Berdasarkan hasil dan pengujian pada bab sebelumnya tentang peramalan kasus positif COVID-19 di Jawa Timur menggunakan metode ARIMA-LSTM didapatkan kesimpulan bahwa model ARIMA-LSTM paling optimal dibangun dengan menggabungkan model ARIMA (2,1,2) dengan model LSTM. Model ARIMA bertugas untuk meramalkan nilai trend dari data deret waktu sedangkan model LSTM bertugas untuk melengkapi peramalan model ARIMA dengan meramalkan nilai residual dari deret waktu. Model ARIMA-LSTM memiliki akurasi yang tinggi terutama pada peramalan jangka pendek dengan MAPE sebesar 1,8 persen untuk kasus peramalan 3 hari ke depan.

## 6. DAFTAR PUSTAKA

Alodokter. (2022). *Virus Corona*. Dipetik Maret 16, 2022, dari <https://www.alodokter.com/virus-corona>

Brockwell, P. J., & Davis, R. A. (1996). *Introduction to Time Series and Forecasting*. New York: Springer.

Cholissodin, I., Sutrisno, Soebroto, A. A., Hasanah, U., & Febiola, Y. I. (2019). *Buku Ajar AI, Machine Learning & Deep Learning*. Malang: Fakultas Ilmu

Komputer (FILKOM), Universitas Brawijaya (UB).

- Dicoding. (2020, Agustus 19). *Apa itu Machine Learning? Beserta Pengertian dan Cara Kerjanya*. Diambil kembali dari Dicoding Blog:  
<https://www.dicoding.com/blog/machine-learning-adalah/>
- Dicoding. (2021, Oktober 9). *Mengenal Deep Learning Lebih Jelas*. Diambil kembali dari Dicoding Blog:  
<https://www.dicoding.com/blog/mengenal-deep-learning/>
- Faurina, R., Winduratna, B., & Nugroho, P. (2018). Predicting Stock Movement Using Unidirectional LSTM and Feature Reduction: The Case of An Indonesia Stock. *International Conference on Electrical Engineering and Computer Science*, (hal. 180-185). Bali.
- Newswire. (2021). *Pak Menkes, Begini Cara Mengatasi Kekurangan Tenaga Kesehatan untuk Covid-19*. Dipetik Maret 08, 2022, dari  
<https://kabar24.bisnis.com/read/20210706/15/1414172/pak-menkes-begini-cara-mengatasi-kekurangan-tenaga-kesehatan-untuk-covid-19>
- Prijono, B. (2018, April 13). *Pengenalan Long Short Term Memory (LSTM) dan Gated Recurrent Unit (GRU) – RNN Bagian 2*. Diambil kembali dari IndoML:  
<https://indoml.com/2018/04/13/pengenalan-long-short-term-memory-lstm-dan-gated-recurrent-unit-gru-rnn-bagian-2/>
- Putri, M. C., & Anggraeni, W. (2018). Penerapan Metode Campuran Autoregressive Integrated Moving Average dan Quantile Regression QR) untuk Peramalan Harga Cabai Sebagai Komoditas Strategis Pertanian Indonesia. *Jurnal Teknik ITS*, 7.
- Salman, A. G., & Kanigoro, B. (2021). Visibility Forecasting Using Autoregressive Integrated Moving Average (ARIMA) Models. *Procedia Computer Science* 179, 252-259.
- Santoso, S. (2009). *Business Forecasting : Metode Peramalan Bisnis Masa Kini dengan Minitab dan SPSS*. Jakarta: Elex Media Komputindo.