

Contents list available at [www.jurnal.unimed.ac.id](http://www.jurnal.unimed.ac.id)

**CESS**  
**(Journal of Computing Engineering, System and Science)**

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



**Peramalan Kasus COVID-19 di Indonesia, Malaysia, Filipina, dan Vietnam  
Menggunakan ARIMA dan LSTM**

**Forecasting COVID-19 Cases in Indonesia, Malaysia, Philippines, and Vietnam  
Using ARIMA and LSTM**

**Marina Wahyuni<sup>1\*</sup>, Fergyanto E. Gunawan<sup>2</sup>**

<sup>1,2</sup> BINUS Graduate Program - Master of Industrial Engineering, Bina Nusantara University  
Jakarta, Indonesia 11480

email: <sup>1</sup>[marina.wahyuni@binus.ac.id](mailto:marina.wahyuni@binus.ac.id), <sup>2</sup>[fgunawan@binus.edu](mailto:fgunawan@binus.edu)

**ABSTRAK**

COVID-19 merupakan pukulan besar bagi perekonomian global, termasuk bagi negara-negara anggota ASEAN. Berbagai perencanaan dan strategi masih tetap diperlukan pada masa transisi pandemi-ke-epidemi untuk meminimalkan risiko penularan COVID-19. Penelitian ini berfokus pada jumlah total kasus terkonfirmasi COVID-19 di Indonesia, Malaysia, Filipina, dan Vietnam, yang merupakan negara-negara ASEAN dengan jumlah kasus tertinggi di Asia Tenggara. Negara-negara tersebut memiliki kesamaan budaya, di mana berkumpul dengan teman dan keluarga merupakan bagian penting dari kehidupan sosial. Penelitian ini mengevaluasi kemampuan ARIMA dan LSTM dalam memprediksi kasus COVID-19 di masing-masing negara, berdasarkan data harian dari 23 Januari 2020 hingga 22 Oktober 2022, menggunakan dataset yang diterbitkan oleh Universitas Johns Hopkins (JHU) dan Our World in Data (OWID) yang dapat diakses melalui Github. Dibandingkan dengan hasil permodelan ARIMA dengan nilai  $R^2$  0,8883 untuk Indonesia, 0,8353 untuk Malaysia, 0,97291 untuk Filipina, dan -3,105 untuk Vietnam, hasil permodelan LSTM mampu memberikan prediksi yang lebih akurat di keempat sampel negara ASEAN, dengan  $R^2$  0,9996 untuk Indonesia, 0,9707 untuk Malaysia, 0,7727 untuk Filipina, dan 0,9200 untuk Vietnam.

**Kata Kunci:** COVID-19, ASEAN, Prediksi, Peramalan, Time Series, LSTM

**ABSTRACT**

COVID-19 has severely impacted the global economy, including ASEAN countries. Various plans and strategies are still needed during the pandemic-to-epidemic transition period to minimize the risk of COVID-19 transmission. The research focuses on the total number of confirmed cases of COVID-19 in Indonesia, Malaysia, the Philippines, and Vietnam, which are among the ASEAN countries with the highest number of cases in Southeast Asia. Those

\*Penulis Korespondensi:  
email: [marina.wahyuni@binus.ac.id](mailto:marina.wahyuni@binus.ac.id)

countries have cultural similarities, where gathering with friends and family is an important part of social life. This research evaluates the ability of ARIMA and LSTM to predict COVID-19 cases in each country, using daily data from January 23, 2020 to October 22, 2022. Datasets published by Johns Hopkins University (JHU) and Our World in Data (OWID) are used, which are accessible through Github. Compared to ARIMA with  $R^2$  of 0,8883 for Indonesia, 0,8353 for Malaysia, 0.97291 for the Philippines, and -3.105 for Vietnam, LSTM model can predict better in the four sampled ASEAN countries, with an  $R^2$  of 0.9996 for Indonesia, 0.9707 for Malaysia, 0.97291 for the Philippines, and 0.9200 for Vietnam.

**Keywords:** COVID-19, ASEAN, Prediction, Forecasting, Time Series, LSTM

## 1. PENDAHULUAN

Berawal dari Wuhan, Provinsi Hubei di Cina, dan kemudian menyebar ke berbagai negara di seluruh dunia, pandemi COVID-19 telah menimbulkan ancaman keamanan non-tradisional yang signifikan bagi dunia. Tidak hanya menyebabkan angka kematian yang sangat tinggi di seluruh dunia, pandemi COVID-19 juga menimbulkan tantangan bagi stabilitas politik, ekonomi dan sosial, pertumbuhan berkelanjutan, dan bahkan ketahanan pangan. Di sisi lain, pandemi ini juga telah mempercepat laju transformasi digital dan membawa perubahan dalam pekerjaan, pembelajaran, ritel, dan bisnis. Hal ini juga telah menimbulkan tantangan baru, seperti perlunya penguatan keamanan dan privasi digital serta mengurangi ketimpangan akses terhadap teknologi [1]. Pandemi ini juga telah mempengaruhi negara-negara ASEAN, terutama pada sektor manufaktur dan pariwisata [2].

Data hingga 22 Oktober 2022 menunjukkan jumlah kasus terkonfirmasi COVID-19 di negara-negara ASEAN mencapai 6% dari total keseluruhan kasus di dunia, sementara total kasus kematian yang disebabkan oleh COVID-19 di negara-negara ASEAN mencapai 5% dari total keseluruhan kasus di dunia. Hal ini menunjukkan bahwa jumlah kasus COVID-19 dan tingkat kematian yang diakibatkannya di negara-negara ASEAN cukup tinggi, sehingga perlu mendapat perhatian dalam upaya penanganan pandemi. Jumlah kasus terkonfirmasi, kematian, dan mortality rate negara-negara ASEAN dapat dilihat di tabel (1).

**Tabel 1.** Jumlah Kasus Terkonfirmasi, Kematian, dan *Mortality Rate* di Negara-Negara ASEAN

Country	Confirmed	Deaths	Mortality Rate (per 100)
Vietnam	11.496.829	43.159	0,38
Indonesia	6.469.276	158.416	2,45
Malaysia	4.880.005	36.444	0,75
Thailand	4.687.281	32.882	0,70
Philippines	3.993.148	63.742	1,60
Singapore	2.059.709	1.660	0,08
Myanmar	630.445	19.475	3,09
Brunei	234.582	225	0,10
Laos	216.177	758	0,35
Cambodia	137.979	3.056	2,21

Penelitian ini akan berfokus pada empat negara anggota ASEAN yaitu Indonesia, Malaysia, Filipina, dan Vietnam. Keempat negara ini dipilih karena memiliki jumlah kasus terkonfirmasi

dan kematian akibat COVID-19 yang tinggi di antara negara-negara ASEAN. Jumlah kasus terkonfirmasi di Indonesia, Malaysia, Filipina, dan Vietnam mencapai 77% dari total kasus di negara-negara ASEAN, sementara jumlah kematian di Indonesia, Malaysia, Filipina, dan Vietnam mencapai 84% dari total kasus di negara-negara ASEAN.

Penelitian ini menggunakan data kumulatif kasus harian COVID-19 di Indonesia, Malaysia, Filipina, dan Vietnam selama periode 23 Januari 2020 hingga 22 Oktober 2022. Penelitian ini penting untuk memantau tren perkembangan angka terkonfirmasi COVID-19 sehingga dapat berguna bagi pemerintah dan organisasi kesehatan untuk mengidentifikasi pola dan tren dalam data dan mengambil tindakan yang tepat untuk mengatasi penyakit ini. Meski sudah hampir tiga tahun sejak COVID-19 pertama kali teridentifikasi, COVID-19 masih merupakan masalah kesehatan global yang serius dan perlu adanya rencana dan strategi serta upaya bersama untuk memperkuat sistem kesehatan, menstabilkan ekonomi, menguatkan kerja sama internasional, meningkatkan daya tahan masyarakat, dan menjaga ketersediaan pangan.

Informasi ini juga bermanfaat bagi masyarakat untuk memahami risiko yang terkait dengan COVID-19 dan mengambil langkah-langkah preventif yang diperlukan. Meskipun beberapa negara telah mengubah status COVID-19 menjadi endemi, protokol kesehatan masih perlu diterapkan untuk meminimalkan risiko penularan. Penelitian ini bertujuan untuk mengetahui tingkat akurasi yang bisa dicapai oleh metode prediksi ARIMA dan LSTM serta metode mana yang lebih akurat untuk memprediksi kasus COVID-19 di Indonesia, Malaysia, Filipina, dan Vietnam.

## **2. TINJAUAN LITERATUR**

COVID-19 merupakan salah satu topik yang paling banyak dibahas dalam tiga tahun terakhir ini. Oleh karena itu, sudah banyak penelitian di seluruh dunia yang mengangkat topik peramalan terkait kasus COVID-19, namun hanya sedikit yang secara khusus membandingkan kasus di Indonesia, Malaysia, Filipina, dan Vietnam.

Penelitian ini akan menerapkan metode statistik ARIMA dan metode berbasis pembelajaran mesin LSTM untuk peramalan data deret waktu. Dua metode ini dipilih karena keakuratan yang terbukti dalam literatur. Penelitian terdahulu telah membandingkan metode ARIMA dan LSTM untuk peramalan COVID-19 di berbagai negara dan mendapatkan hasil yang berbeda. Kumar et al. (2022) menemukan bahwa ARIMA lebih efektif untuk peramalan kasus COVID-19 di India, Italia, Jepang, Spanyol, Inggris, dan Amerika Serikat. Kirbaş et al. (2022) menemukan LSTM lebih efektif untuk peramalan kasus di Denmark, Belgia, Jerman, Prancis, Inggris, Finlandia, Swiss, dan Turki. Sementara itu, ArunKumar et al. (2022) menunjukkan bahwa ARIMA dan LSTM masing-masing memiliki keunggulan dan kekurangan dan metode terbaik bervariasi di berbagai negara. Ini menunjukkan bahwa akurasi dan efektifitas dari metode peramalan tergantung pada data yang tersedia [3]. Secara umum, ARIMA dan LSTM dapat digunakan untuk memprediksi data deret waktu. Namun, pilihan terbaik akan bergantung pada sifat data yang akan diprediksi dan tujuan penggunaan model.

## **3. METODE PENELITIAN**

### **3.1 ARIMA**

ARIMA (Autoregressive Integrated Moving Average) adalah model prediksi deret waktu yang populer digunakan dalam analisis data. Model ARIMA didasarkan pada dua metode prediksi, AR (Autoregressive) dan MA (Moving Average), yang dikombinasikan menjadi satu

model untuk meningkatkan akurasi prediksi. Metode AR didasarkan pada asumsi bahwa nilai data saat ini dipengaruhi oleh nilai data di masa lalu, sedangkan MA didasarkan pada asumsi bahwa nilai data saat ini dipengaruhi oleh rata-rata nilai data di masa lalu. Model ARIMA menggunakan tiga parameter,  $p$ ,  $d$ , dan  $q$ , untuk mengatur komponen AR dan MA dalam model. Model ini berguna untuk mengidentifikasi pola-pola yang terjadi dalam data deret waktu dan memprediksi bagaimana pola-pola tersebut akan berkembang di masa depan. Secara umum ARIMA ( $p,d,q$ ) dimodelkan sebagai berikut [3]:

$$y'_t = \Delta^d y_t, \quad (1)$$

$$y'_t = \beta_0 + \sum_{i=1}^p \beta p_i y'_{t-1} + \sum_{i=1}^q \beta q_i \varepsilon_{t-1} + \varepsilon_t, \quad (2)$$

di mana observasi series  $y$  pada periode  $t$ , sedangkan  $\beta_0$ ,  $\beta_p$ , dan  $\beta_q$  adalah parameter *intercept* koefisien AR dan MA,  $\varepsilon_t$  adalah errors dan  $\Delta$  adalah tingkat differencing.

### 3.2 LSTM

LSTM (Long Short-Term Memory) adalah pengembangan dari jaringan syaraf tiruan rekursif yang dapat memecahkan masalah vanishing dan exploding gradient. LSTM lebih baik dalam membuat peramalan jangka panjang dan cocok untuk peramalan deret waktu. Metode LSTM memecahkan masalah RNN dengan menambahkan sel memori dengan kesalahan konstan sehingga kesalahan dapat direproduksi tanpa menemukan gradien yang hilang atau meledak. LSTM memiliki tiga gerbang: gerbang input, gerbang lupa, dan gerbang output. Setiap gerbang memiliki parameter yang disebut bobot dan bias. Proses komputasi LSTM dilakukan dengan tahapan sebagai berikut [4]:

Nilai input hanya dapat disimpan dalam status sel jika gerbang input mengizinkannya. Perhitungan nilai gerbang input dan kandidat dari keadaan sel dilakukan dengan menggunakan persamaan (3) dan (4).

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

dimana  $i_t$  adalah nilai input gate,  $W_i$  adalah bobot nilai input pada waktu  $t$ ,  $x_t$  adalah nilai input pada waktu  $t$ ,  $U_i$  adalah bobot nilai input dari waktu ke  $t - 1$ ,  $h_{t-1}$  adalah nilai input dari waktu ke  $t - 1$ ,  $b_i$  adalah bias pada input gate and  $\sigma$  adalah fungsi sigmoid.

$$C_t = \tanh(W_c x_t + U_c h_{c-1} + b_c), \quad (4)$$

dimana  $C_t$  adalah nilai kandidat cell state,  $W_c$  adalah bobot untuk nilai input pada cell ke  $c$ ,  $x_t$  adalah nilai input pada waktu ke  $t$ ,  $U_c$  adalah bobot untuk nilai output dari cell ke  $c - 1$ ,  $h_{c-1}$  adalah nilai output dari cell ke  $c - 1$  dan  $b_c$  adalah bias pada cell ke  $c$  dan  $\tanh$  adalah fungsi hyperbolic tangent. Kemudian nilai dari forget gate dihitung dengan menggunakan persamaan (5).

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f), \quad (5)$$

dimana  $f_t$  adalah nilai dari forget gate,  $W_f$  adalah bobot untuk nilai input pada waktu ke  $t$ ,  $x_t$  adalah nilai input pada waktu ke  $t$ ,  $U_f$  adalah bobot untuk nilai output dari waktu ke  $t - 1$ ,  $h_{t-1}$  adalah nilai output dari waktu ke  $t - 1$  dan  $b_f$  adalah bias pada forget gate dan  $\sigma$  adalah fungsi sigmoid. Selanjutnya memory cell state dihitung menggunakan persamaan (6).

$$C_t = i_t \odot C_t + f_t \odot C_{t-1}, \quad (6)$$

dimana  $C_t$  adalah nilai memory cell state,  $i_t$  adalah nilai dari input gate,  $C_t$  adalah nilai kandidat memory cell state,  $f_t$  adalah nilai forget gate dan  $C_{t-1}$  adalah nilai memory cell state pada cell sebelumnya, dan  $\odot$  adalah Produk Hadamard dari matriks (produk dari elemen-elemen pada posisi yang sama dari dua matriks). Setelah dihasilkan memory cell state yang baru, nilai dari output gate dapat dihitung dengan menggunakan persamaan (7).

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \quad (7)$$

dimana  $o_t$  adalah nilai dari output gate,  $W_o$  adalah bobot untuk nilai input pada waktu ke  $t$ ,  $x_t$  adalah nilai input pada waktu ke  $t$ ,  $U_o$  adalah bobot untuk nilai output dari waktu ke  $t - 1$ ,  $h_{t-1}$  adalah nilai output dari waktu ke  $t - 1$  dan  $b_o$  adalah bias pada output gate dan  $\sigma$  adalah fungsi sigmoid. Nilai output final dihitung dengan menggunakan persamaan (8).

$$h_t = o_t \odot \tanh(C_t), \quad (8)$$

dimana  $h_t$  adalah output final,  $o_t$  adalah nilai output gate,  $C_t$  adalah nilai memory cell state yang baru dan  $\tanh$  adalah fungsi hyperbolic tangent.

### 3.3 Model Evaluasi

Setelah model LSTM digunakan untuk memprediksi nilai-nilai pada data uji, hasil prediksi harus divalidasi dengan data asli untuk mengevaluasi seberapa akurat model. Studi ini menggunakan empat metrik evaluasi, yaitu MSE, MAE, RMSE, dan  $R^2$  yang dituliskan dalam persamaan (9), (10), (11), dan (12). Tujuannya adalah untuk mengukur seberapa jauh hasil prediksi model dari nilai sebenarnya dan seberapa baik model dalam menjelaskan variasi dari data uji.

$$MAE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i), \quad (9)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (11)$$

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}, \quad (12)$$

dimana  $n$  adalah jumlah obsevasi,  $y_i$  adalah observasi pada waktu  $i$ ,  $\bar{y}$  adalah nilai rata-rata dari  $y$ , dan  $\hat{y}$  adalah prediksi nilai  $y$ .

## 4. HASIL DAN PEMBAHASAN

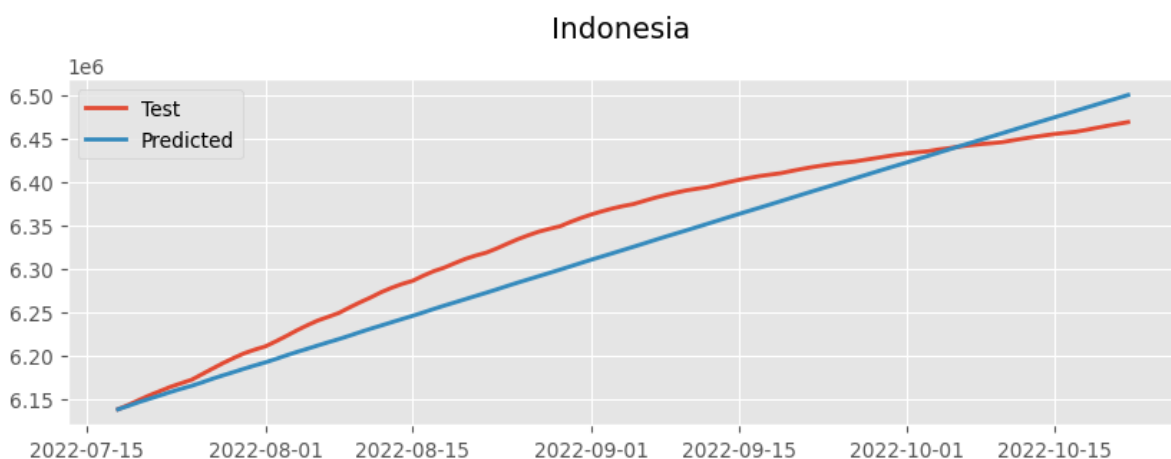
### 4.1 Hasil Pemodelan ARIMA

Database dibagi menjadi 90% data latih dan 10% data uji untuk masing-masing negara, yaitu 868 hari data latih dan 97 hari data uji di Indonesia, 901 hari data latih dan 101 hari data uji di Malaysia, 897 hari data latih dan 100 hari data uji di Filipina, serta 903 hari data latih dan 101 hari data uji di Vietnam. Model terlatih kemudian diuji pada data uji, dan ARIMA dan LSTM digunakan untuk memprediksi jumlah total kasus untuk setiap negara.

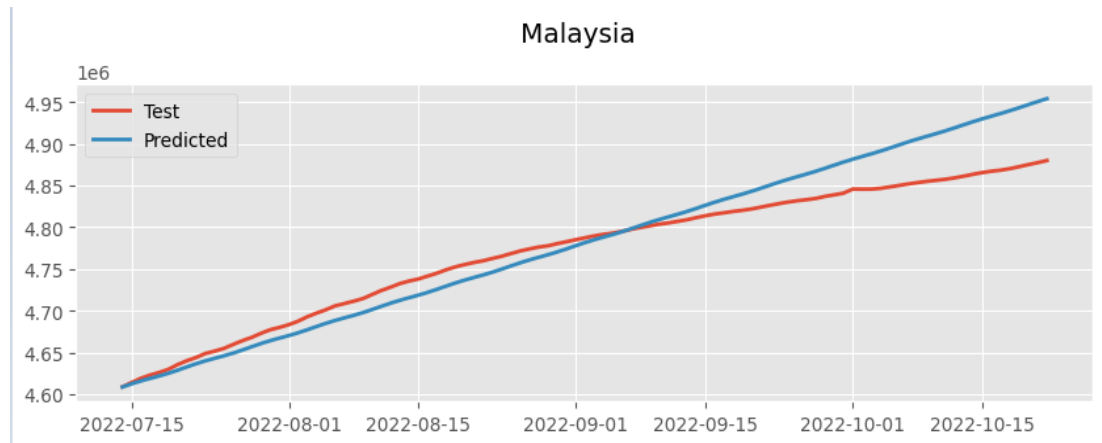
Dengan menggunakan library auto Arima, kita dapat menemukan model ARIMA terbaik. Penelitian ini menggunakan library auto ARIMA untuk menemukan model ARIMA terbaik dengan mencari model yang memiliki nilai Akaike Information Criterion (AIC) paling rendah. Library ini akan secara otomatis mencari kombinasi parameter ( $p, d, q$ ) yang tepat untuk model ARIMA, sehingga tidak perlu melakukan pencarian parameter secara manual. AIC adalah metode evaluasi model statistik, dimana nilai yang lebih kecil menunjukkan model yang lebih baik karena mampu memprediksi data dengan lebih baik dibanding model lain.

Untuk masing-masing negara, kombinasi parameter (5,1,5) memiliki nilai AIC paling rendah untuk Indonesia, parameter (3,2,2) untuk Malaysia, parameter (1,2,1) untuk Filipina, dan parameter (1,2,2) untuk Vietnam.

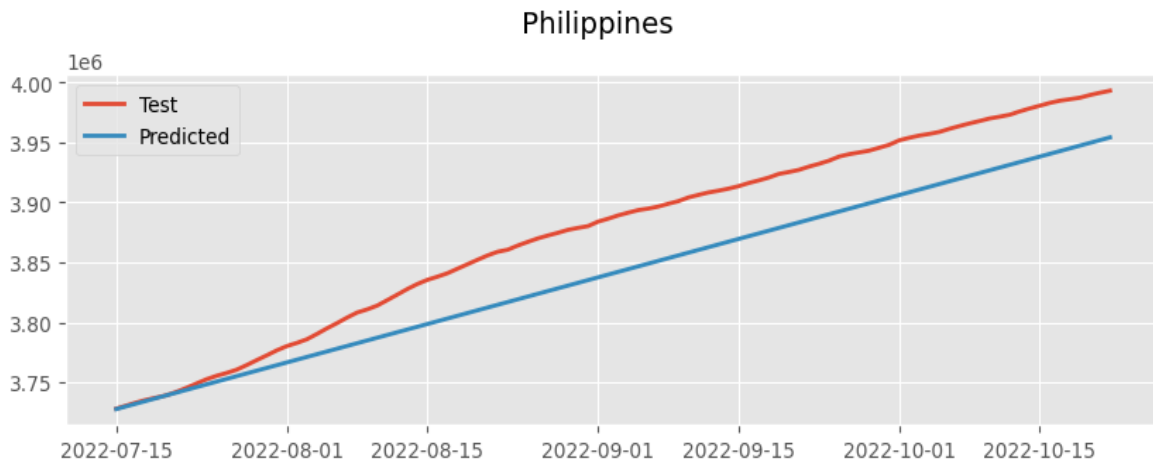
Setelah model ARIMA diperoleh, akan dilakukan prediksi dan dibuat plot perbandingannya dengan data uji. Fungsinya adalah untuk membantu kita melihat seberapa baik model kita dapat memprediksi data. Plot Perbandingan Data Uji & Hasil Prediksi untuk model ARIMA dapat dilihat pada gambar (1), (2), (3), (4).



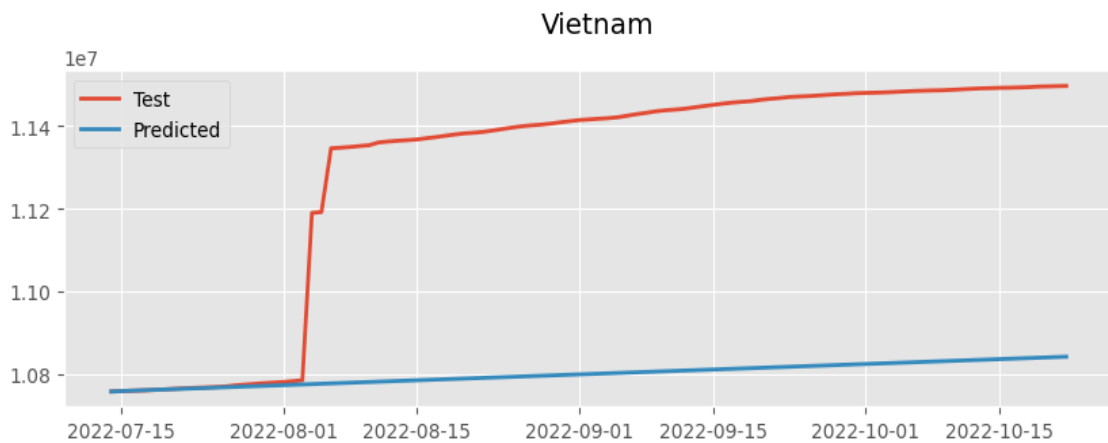
Gambar 1. Plot Perbandingan Data Uji & Hasil Prediksi – Indonesia



Gambar 2. Plot Perbandingan Data Uji & Hasil Prediksi – Malaysia



Gambar 3. Plot Perbandingan Data Uji & Hasil Prediksi – Filipina



Gambar 4. Plot Perbandingan Data Uji & Hasil Prediksi – Vietnam

Hasil evaluasi dengan menggunakan MAE, MSE, RMSE dan  $R^2$  untuk model ARIMA dirangkum dalam tabel (2).

**Tabel 2.** Rangkuman MAE, MSE, RMSE dan R<sup>2</sup> untuk model ARIMA

Country	MAE	MSE	RMSE	R-Squared
Indonesia	28.359,9605	1.070.783.273,5116	32.722,8250	0,8884
Malaysia	23.595,8270	963.120.330,7220	31.034,1800	0,8353
Philippines	34.202,7311	1.417.930.181,0634	37.655,4137	0,7728
Vietnam	492.997,1553	307.636.843.508,2610	554.650,1992	-3,1052

Berdasarkan tabel di atas, model ARIMA memiliki kinerja yang cukup baik untuk negara Indonesia, Malaysia, dan Filipina dengan nilai MAE, MSE, dan RMSE yang kecil, serta nilai R<sup>2</sup> yang paling tinggi. Ini menunjukkan bahwa model dapat menjelaskan variasi nilai sebenarnya dengan cukup baik. Namun model memiliki kinerja yang buruk pada Vietnam, dengan nilai MAE, MSE, dan RMSE yang besar, serta nilai R<sup>2</sup> yang negatif. Ini menunjukkan bahwa model tidak dapat memprediksi nilai dengan baik pada Vietnam.

#### 4.2 Hasil Pemodelan LSTM

Sebagai pembanding atas model ARIMA, akan dilakukan peramalan dengan model LSTM. Dua variable yaitu total kasus dan total kematian dijadikan variable input. Model dilakukan dengan parameter Adam optimizer, 96 batch size, satu layer LSTM, 440 epoch, dan 100 neuron pada hidden layer. Keluaran output hanya 1 yaitu prediksi total kasus.

Model yang sama akan kita aplikasikan ke dataset 4 negara yaitu Indonesia, Malaysia, Filipina, dan Vietnam dan dilakukan analisa secara terpisah. Ini akan memungkinkan kita untuk membandingkan hasil prediksi di setiap negara dan melihat apakah model kita dapat bekerja dengan baik untuk semua negara tersebut. Ringkasan model dan jumlah parameternya ditunjukkan pada tabel (3).

**Tabel 3.** Model LSTM

Model: "sequential"

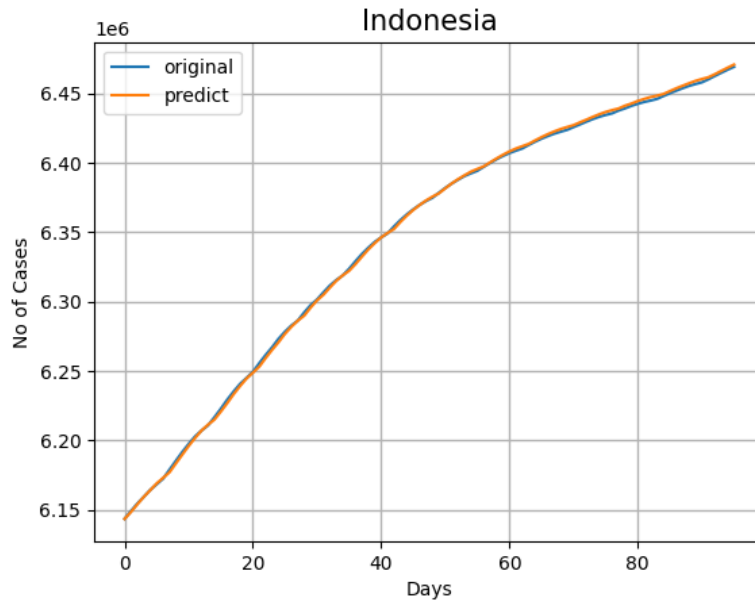
Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 100)	41200
dense (Dense)	(None, 1)	101
Total params: 41,301		
Trainable params: 41,301		
Non-trainable params: 0		

Tabel di atas menunjukkan arsitektur model yang disebut "sequential". Ini berarti model dibangun secara berurutan, lapisan demi lapisan. Model ini terdiri dari dua lapisan: LSTM (Long Short-Term Memory) dan Dense. Lapisan LSTM memiliki output berbentuk (None, 100), yang berarti lapisan ini akan menghasilkan output berupa array 2D dengan 100 kolom. Lapisan Dense memiliki output berbentuk (None, 1), yang berarti lapisan ini akan menghasilkan output

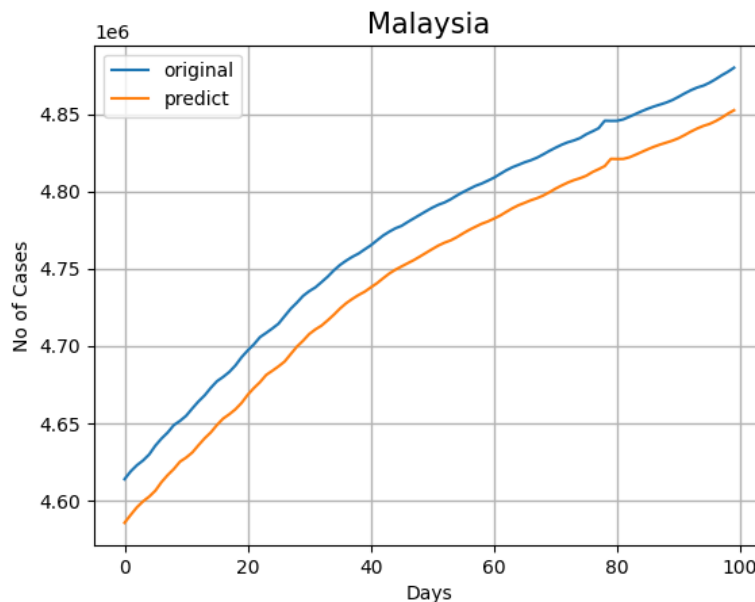


berupa array 2D dengan 1 kolom. Lapisan ini memiliki 101 parameter yang dapat dilatih untuk memprediksi output. Keseluruhan model memiliki 41.301 parameter yang dapat dilatih untuk memprediksi output. Semua parameter dapat dilatih untuk memperbaiki kinerja model. Namun, juga harus diperhatikan bahwa parameter yang terlalu banyak dapat menyebabkan overfitting.

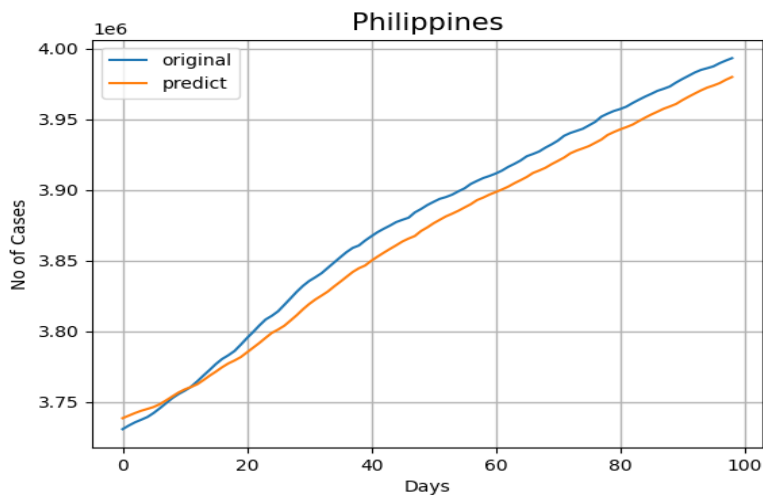
Setelah menentukan parameter yang sesuai, hasil prediksi ditampilkan pada Gambar (5), (6), (7), dan (8).



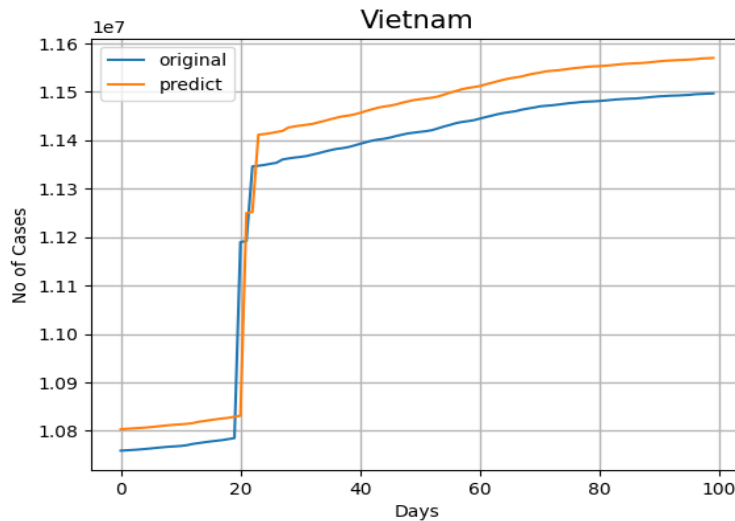
**Gambar 5.** Plot Data Aktual dan Data Prediksi COVID-19 – Indonesia



**Gambar 6.** Plot Data Aktual dan Data Prediksi COVID-19 – Malaysia



**Gambar 7.** Plot Data Aktual dan Data Prediksi COVID-19 – Filipina



**Gambar 8.** Plot Data Aktual dan Data Prediksi COVID-19 – Vietnam

Model dievaluasi berdasarkan MAE, MSE, RMSE, dan R<sup>2</sup>. Rangkuman hasil evaluasi dapat dilihat pada Tabel (4). Berdasarkan tabel tersebut, dapat disimpulkan bahwa model LSTM dapat memprediksi total kasus terkonfirmasi dengan baik di empat negara, dengan R<sup>2</sup> 0,9996 untuk Indonesia, 0,9707 untuk Malaysia, 0,9729 untuk Filipina, dan 0,9200 untuk Vietnam.

**Tabel 4.** Rangkuman MAE, MSE, RMSE dan R<sup>2</sup> untuk model LSTM

Country	MAE	MSE	RMSE	R-Squared
Indonesia	1.553,4531	3.258.696,8000	1.805,1860	0,9996
Malaysia	12.579,9850	165.235.660,0000	12.854,4020	0,9707
Philippines	12.123,7500	164.598.900,0000	12.829,6100	0,9729
Vietnam	69.045,1200	5.820.895.700,0000	76.294,8000	0,9200

#### 4.3 Perbandingan Hasil Prediksi ARIMA & LSTM

Hasil evaluasi peramalan model ARIMA dan LSTM menunjukkan bahwa LSTM memiliki akurasi yang lebih baik untuk keempat negara, Indonesia, Malaysia, Filipina, dan Vietnam. Model berhasil memprediksi kasus COVID-19 di setiap negara terlepas dari intervensi, kebijakan, dan kesediaan fasilitas kesehatan masing-masing negara.

**Tabel 5.** Rangkuman Perbandingan MAE, MSE, RMSE dan R<sup>2</sup> untuk model ARIMA & LSTM

Model	Country	MAE	MSE	RMSE	R-Squared
LSTM	Indonesia	1553,4531	3258696,8	1805,186	0,999647662
ARIMA(5, 1, 5)		28359,96055	1070783274	32722,82496	0,888368066
LSTM	Malaysia	12579,985	165235660	12854,402	0,970715057
ARIMA(3,2,2)		23595,82698	963120330,7	31034,18004	0,835329237
LSTM	Philippines	12123,75	164598900	12829,61	0,972936958
ARIMA(1, 2, 1)		34202,73113	1417930181	37655,4137	0,77276753
LSTM	Vietnam	69045,12	5820895700	76294,8	0,920038683
ARIMA(1, 2, 2)		492997,1553	3,07637E+11	554650,1992	-3,10516087

Berdasarkan tabel (5) dapat disimpulkan bahwa model LSTM lebih unggul dibandingkan model ARIMA dalam hal akurasi prediksi untuk negara Indonesia, Malaysia, Filipina, dan Vietnam. Hal ini ditunjukkan oleh nilai MAE, MSE, dan RMSE, yang lebih rendah serta R<sup>2</sup> yang lebih tinggi untuk model LSTM dibandingkan model ARIMA untuk negara-negara tersebut.

Nilai R<sup>2</sup> negatif pada prediksi untuk negara Vietnam yang menggunakan model ARIMA (1,2,2), menunjukkan bahwa model tersebut tidak dapat digunakan untuk memprediksi data untuk negara tersebut.

#### 5. KESIMPULAN

Penelitian ini menerapkan model ARIMA & LSTM untuk memprediksi kasus COVID-19 di Indonesia, Malaysia, Filipina, dan Vietnam. Tingkat akurasi yang bisa dicapai oleh metode prediksi ARIMA & LSTM bervariasi tergantung dari negara yang diteliti. Berdasarkan hasil evaluasi, metode LSTM menunjukkan tingkat akurasi yang cukup tinggi, dengan nilai R<sup>2</sup> berkisar antara 0,77 dan 0,999. Sedangkan metode ARIMA menunjukkan tingkat akurasi yang lebih rendah, dengan nilai R<sup>2</sup> berkisar antara -3,1 dan 0,888. Dapat disimpulkan bahwa untuk penelitian ini, metode LSTM lebih akurat dibandingkan metode ARIMA. Perlu diingat bahwa hasil penelitian ini mengabaikan faktor-faktor lain seperti intervensi, kebijakan, dan ketersediaan fasilitas kesehatan masing-masing negara. Hasil penelitian dapat berbeda dengan database yang berbeda untuk setiap negara dan rentang waktu.

Kesimpulannya, secara umum ARIMA dan LSTM sama-sama dapat digunakan untuk memprediksi kasus COVID-19 di berbagai negara. Pilihan terbaik antara kedua model tersebut akan tergantung pada sifat-sifat data yang akan diprediksi dan tujuan penggunaan model tersebut. Jika data yang akan diprediksi memiliki pola-pola yang tidak terlalu kompleks dan tidak memerlukan akurasi yang tinggi, maka model ARIMA dapat digunakan. Namun, jika data memiliki pola-pola yang kompleks dan memerlukan akurasi yang tinggi, maka model LSTM dapat digunakan.

Untuk meningkatkan keakuratan dalam penelitian kedepannya, dapat ditambahkan beberapa variabel tambahan yang mungkin akan mempengaruhi hasil prediksi. Misalnya, dapat ditambahkan informasi tentang tingkat kepadatan penduduk, tingkat mobilitas masyarakat, tingkat kepatuhan terhadap protokol kesehatan, tingkat pelaksanaan tes dan isolasi, dan efektivitas vaksinasi di wilayah yang diteliti. Dengan mempertimbangkan variabel-variabel ini, kita dapat meningkatkan kapasitas dan kinerja model untuk memprediksi perkembangan jangka panjang dari persebaran infeksi dengan lebih akurat dan membuat rekomendasi yang lebih tepat untuk upaya pencegahan dan pengendalian infeksi. Namun, perlu diingat bahwa menambahkan faktor-faktor ini juga dapat memperumit analisis dan dibutuhkan data yang lebih banyak dan lengkap untuk mengukurnya. Pada dasarnya pemilihan variabel tambahan yang tepat tergantung pada data yang tersedia dan tujuan penelitian yang ingin dicapai.

## REFERENSI

- [1] N. A. Yaacob, M. I. Yusof, S. M. Nuruddin, Z. M. Zain och N. A. Mustapa, "Non-Traditional Security Issues in Southeast Asia during COVID-19: Implications and Mitigation Strategies by ASEAN," *Proceedings*, vol. 82, nr 1, p. 90, 2022.
- [2] M. R. Ridzuan och N. A. S. A. Rahman, "The deployment of fiscal policy in several ASEAN countries in dampening the impact of COVID-19," *Journal of Emerging Economies & Islamic Research*, vol. 9, nr 1, pp. 16-28, 2021.
- [3] R. J. Hyndman och G. Athanasopoulos, *Forecasting: principles and practice*, 3rd edition, Melbourne: OTexts, 2021.
- [4] H. Chung och K.-s. Shin, "Genetic algorithm-optimized long short-term memory network for stock market prediction," *Sustainability*, vol. 10, nr 10, p. 3765, 2018.
- [5] R. Vega, L. Flores och R. Greiner, "SIMLR: Machine Learning inside the SIR model for COVID-19 Forecasting," *Forecasting*, vol. 4, nr 1, pp. 72-94, 2022.
- [6] P. Kumari och D. Toshniwal, "Real-time estimation of COVID-19 cases using machine learning and mathematical models-The case of India," *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, pp. 369-374, 2020.
- [7] H. R. Niazkar och M. Niazkar, "Application of artificial neural networks to predict the COVID-19 outbreak," *Global health research and policy*, vol. 5, nr 1, pp. 1-11, 2020.
- [8] M. Shawaqfah och F. Almomani, "Forecast of the outbreak of COVID-19 using artificial neural network: Case study Qatar, Spain, and Italy," *Results in Physics*, vol. 27, p. 104484, 2021.
- [9] M. d. B. Braga, R. d. S. Fernandes, G. N. d. S. Jr, J. E. C. d. Rocha, C. J. F. Dolácio, I. d. S. T. Jr, R. R. Pinheiro, F. N. Noronha, L. L. S. Rodrigues och R. Thia, "Artificial neural networks for short-term forecasting of cases, deaths, and hospital beds occupancy in the COVID-19 pandemic at the Brazilian Amazon," *PLoS ONE* 16(3): e0248161, p. <https://doi.org/10.1371/journal.pone.0248161>, 2021.
- [10] G. Toga, B. Atalay och M. D. Toksari, "COVID-19 prevalence forecasting using autoregressive integrated moving average (ARIMA) and artificial neural networks (ANN): Case of Turkey," *Journal of infection and public health*, vol. 14, nr 7, pp. 811-816, 2021.

- [11] N. Abidin, A. Ahmarofi och N. Zaibidi, "Prediction of Intensive Care Cases for COVID-19 Pandemic in Malaysia: An Artificial Neural Networks Approach," *ASM Science Journal* 16, p. <https://doi.org/10.32802/asmscj.2021.747>, 2021.
- [12] M. Kumar, S. Gupta, K. P. Kumar och M. Sachdeva, "Spreading of COVID-19 in India, Italy, Japan, Spain, UK, US: a prediction using ARIMA and LSTM model Digital Government: Research and Practice," *Digital Government: Research and Practice*, vol. 1, nr 4, pp. 1-9., 2020.
- [13] K. ArunKumar, D. V.Kalaga, C. S. Kumar, M. Kawaji och T. M.Brenza, "Comparative analysis of Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM) cells, Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA) for forecasting COVID-19 trends," *Alexandria Engineering Journal*, vol. 61, nr 10, pp. 7585-7603, 2022.
- [14] İ. e. a. Kirbaş, "Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches," *Chaos, Solitons & Fractals*, vol. 138, p. 110015, 2020.