Integration of Instagram Social Media Data in Forecasting Dengue Fever Cases Using Hybrid ARIMAX-LSTM

Wiwik Anggraeni
Dept. of Information Systems,
Faculty of Intelligent Electrical and Informatics Technology (ELECTICS),
Institut Teknologi Sepuluh Nopember,
Surabaya, Indonesia
wiwik@is.its.ac.id

Muhammad Taqi Akmaluddin
Dept. of Information Systems,
Faculty of Intelligent Electrical and Informatics Technology (ELECTICS)
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
mtaqiakmaluddin@gmail.com

Aparna Kumari
Department of Computer Science and Engineering
Nirma University
Gujarat, India
aparna.kumari@nirmauni.ac.in

Edwin Riksakomara
Dept. of Information Systems,
Faculty of Intelligent Electrical and Informatics Technology (ELECTICS),
Institut Teknologi Sepuluh Nopember,
erk@is.its.ac.id

Radityo Prasetianto Wibowo
Dept. of Information Systems,
Faculty of Intelligent Electrical and Informatics Technology (ELECTICS)
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
radityo.p.w@gmail.com

Reza Fuad Rachmadi
Dept. of Computer Engineering,
Faculty of Intelligent Electrical and Informatics Technology (ELECTICS)
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
fuad@its.ac.id

***Abstract*—Dengue Fever (DF) is a major health concern in Indonesia, with the total number of cases periodically increasing. Since the COVID-19 pandemic, many people have been hesitant to consult hospitals and tend to seek health-related information via social media. Social media has emerged as a common platform for sharing information. Over time, social media has been recognized for its potential in understanding disease outbreaks. This study utilizes Instagram social media data to forecast the number of dengue fever cases in various regions of Indonesia. The forecasting model employs a hybrid ARIMAX-LSTM approach, incorporating data from DF case studies, as well as climate data such as rainfall, humidity, and wind speed, along with Instagram data related to dengue fever. Experimental results suggest that incorporating exogenous variables, particularly climate and social media data, improves forecasting performance. Climate data generally yield better forecasting results with lower SMAPE values on a daily basis, while the combination of climate and social media data often results in lower SMAPE values than when either is used alone. The largest SMAPE difference observed was 9.18%, and the RMSE was 21.64.**

Keywords—Social Media, Instagram, Forecasting, Disease, ARIMAX, LSTM

I. Introduction

Dengue Fever is an infectious illness caused by the Dengue virus. The Dengue infection, transmitted by the Aedes Aegypti mosquito, is one of the fastest-spreading viral infections among humans [1]. Dengue Fever frequently occurs in tropical countries, influenced by climate and surrounding environmental factors [2]. It is one of the diseases that often causes outbreaks and leads to fatalities. Dengue Fever cases showed an increase during the COVID-19 pandemic in 2020, with even more cases reported in 2024. These cases were spread across 465 regencies/cities in 34 provinces, with fatalities occurring in 203 regencies/cities in 29 provinces.

As a result of the COVID-19 pandemic, many people were reluctant to seek medical consultation at hospitals due to fears of contracting the virus and being diagnosed with COVID-19. A survey involving 110 respondents was conducted found that about 71.8 percent of respondents had not visited a hospital or clinic since the onset of COVID-19. They typically sought information about their illnesses online, particularly through social media. Unfortunately, this behavior has persisted to this day, making it difficult for healthcare facilities like Community Health Centers or hospitals to collect accurate data on the number of cases, despite the fact that Dengue Fever cases continue to exist.

Social media has become a platform that enables people to share and seek information. It can be useful in understanding disease outbreaks [3]. To date, social media has emerged as a common platform for sharing information [4], especially on health-related topics. This behavior was also prevalent during the COVID-19 pandemic [5]. Instagram is a social networking platform that allows users to share photos and videos for free, with around two billion people using it globally. Throughout the pandemic, Instagram was extensively utilized for seeking and sharing health information. Instagram plays a crucial role in spreading health information [6].

Based on the current conditions as a long-term impact of COVID-19 and the fact that social media, especially Instagram, has become a medium for exchanging and seeking various health-related information, this study attempts to incorporate Instagram social media data to forecast the number of dengue fever cases in different regions of Indonesia. The forecasting model for dengue fever cases was developed using a hybrid approach that combines deep learning and statistical methods, involving data on the number of cases per region over time, as well as climate data such as rainfall, humidity, and wind speed, along with Instagram data related to dengue fever.

This paper contributes to the use of Instagram social media in forecasting the spread of dengue fever cases, whereas most current models rely solely on conventional hospital data. Additionally, it examines the impact of social media factors on the performance of disease spread forecasting. Furthermore, this paper also contributes to the modeling approach used, which employs a hybrid LSTM and ARIMA.

II. Related Works

As of early 2024, research related to Dengue Fever (DF) has primarily relied on conventional data. Over the past four years, the conventional data used has included environmental data [7], population mobility, and meteorology data [8], climate data [9], larva-free rates [10], which are combined with climate and population density data [11]. On the other hand, modern data more frequently used includes Google Trends [12], internet queries [13], and Twitter [14]. However, it is still very rare, or even non-existent, to involve Instagram data as proposed in this paper.

The involvement of Google Trends data in relation to disease spread has been explored by several researchers. [13] examined the impact of Google Trends and Twitter data on forecasting the number of dengue fever cases in regions with different levels of internet penetration. The influence of Google Trends data on the DF case numbers has also been studied by [15]. Google Trends has also been used to observe disease behavior [16]. In addition, Google Trends has been employed for predicting outbreaks [17], forecasting the number of disease cases, including dengue fever [12] , Lyme disease [18], and Hand, Foot, and Mouth Disease [19].

Regarding the approach used, several recent methods that are widely utilized include LSTM [9][20]. Although LSTM is considered effective, it still has some weaknesses for forecasting Dengue Fever (DF). Therefore, LSTM has been developed in combination with other hybrid methods [13]. Another method is CNN, which has been applied to time series data, though it is less effective for medium- and long-term forecasting. Additionally, there are approaches such as Seq2Seq [14] and Wavelet-Neural Network. Beyond deep learning, many studies still employ mathematical approaches [7][8][10]. Mathematical approaches, when combined with deep learning, can enhance performance compared to when they stand alone. Thus, in this paper, the weaknesses of LSTM will be addressed by the strengths of ARIMA to improve forecasting performance.

III. Methods

The research stages in this paper are shown in Figure 1.

A. Data Acquisition

The data used in this study includes the number of dengue fever cases, as well as climate data such as temperature, humidity, rainfall, and wind speed, along with Instagram social media data. The dengue fever case data for Malang Regency were obtained from the Malang Regency Health Office. Climate data from dataonline.bmkg.go.id. Social media data was obtained through scraping from Instagram accounts with keywords related to dengue fever. The data collected includes dates and descriptions from some accounts. All data are presented in daily form for the period from January 1, 2022, to June 30, 2023.

B. Data Pre-processing

This stage produces an output dataset divided into training and testing sets. Data pre-processing includes checking for missing values and filling in any gaps. Additionally, social media data is selected based on keywords. The social media data obtained from Instagram initially includes all data. In this study, only data containing one or more of the keywords dengue fever, DF, Aedes, Aegypti, and dengue were selected. The data selection was checked based on the captions extracted.

C. Correlation Test

The correlation results will be assessed to determine if there is a statistically significant relationship between dengue fever case data and social media data. The level of the relationship between variables is assessed based on the Pearson correlation test value [21].

D. Train-Test Data Partition

The collected data is divided into training and testing sets, with 80% allocated for training and 20% reserved for testing. [21].

E. Modeling and Forecasting by ARIMAX

The steps in ARIMAX modeling and forecasting include:

E.1 Stationarity Test

The stationarity test is conducted using the Augmented Dickey-Fuller (ADF) test [12]. Data is considered stationary if the p-value is <0.05.

Additionally, by examining the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) graphs based on the visible lag patterns, a dataset can be considered stationary if there is a significant decrease in ACF and PACF lags. The ACF and PACF plots are also used to determine the p and q orders that will be used in the ARIMAX model development [12].

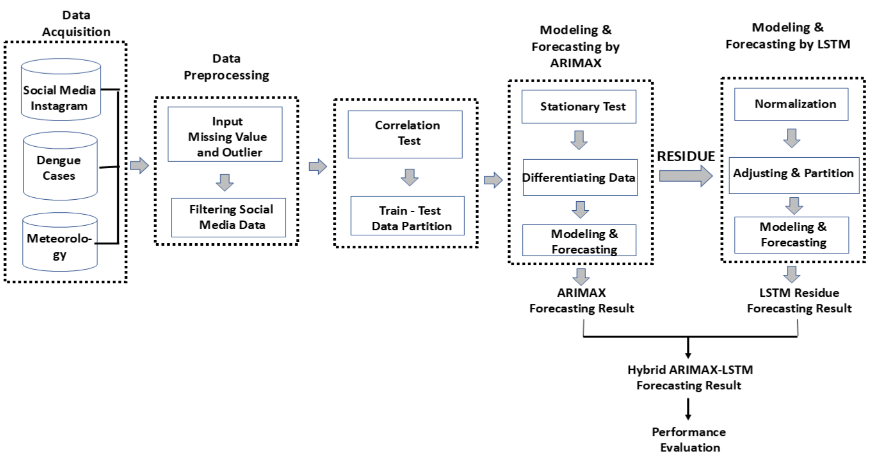

Fig. 1. Framework of the research stages conducted in this study

E.2 Differencing Data

A non-stationary time series must be transformed into stationary data through the differencing process. If the data does not yet meet the stationarity criteria, the differencing process can be repeated until stationary data is obtained [12].

E.3 Forecasting with ARIMAX

The ARIMAX forecasting process uses five scenarios. The first scenario involves using the developed model to perform ARIMAX forecasting on the dengue fever case data from the Health Office, using climate data as an exogenous variable. The second scenario forecasts using dengue fever case data from the Health Office with social media data as an exogenous variable. The third scenario involves forecasting dengue fever case data using data from the Health Office, incorporating both climate data and social media data as exogenous variables. The fourth scenario forecasts are based solely on the dengue fever case data from the Health Office. The fifth scenario involves forecasting using only social media data. Additionally, forecasting for the next four months will be carried out. This stage aims to produce forecasts of dengue fever cases and the residuals from the ARIMAX method, which will serve as inputs for the LSTM method.

F. *Modeling and Forecasting with LSTM*

The next step is to apply the hybrid method using the residual or error values generated from the ARIMAX forecasting as input for LSTM. The steps involved include:

F.1 Normalization, Adjustment, and Data Partitioning

The normalization of ARIMAX residual data is carried out using the MinMax Normalization method [9]. After normalizing the residual data, the next step is to transform it into a format that is suitable for LSTM prediction. This is done using the sliding window method. The residual data is divided into several time windows, each consisting of several adjacent residual data points.

F.2 LSTM Model Design

The Long Short-Term Memory (LSTM) model utilizes the Rectified Linear Unit (ReLU) activation function. ReLU is selected due to its effectiveness in addressing non-linear problems, leading to favorable outcomes. Additionally, other parameters are based on the best model from the study [13], such as the optimization function using Adaptive Moment Estimation (Adam), with a units parameter set to 64, epochs set to 200, and a batch size of 64.

F.3 Forecasting with LSTM

In this stage, the residual data obtained from ARIMAX is forecasted using the LSTM model. Additionally, residual forecasting is conducted for the next four months. The results of the LSTM forecasting are then denormalized or returned to their original values. The purpose of this stage is to produce residual data forecasts using the Long Short-Term Memory (LSTM) method. These forecasts will be used for the next stage of forecasting with the Hybrid ARIMAX-LSTM method.

G. *Forecasting with Hybrid ARIMAX-LSTM*

This process is performed using equation (1). Additionally, forecasts are made for the next 120 periods (4 months).

$$Forecast\ Hybrid\ ARIMAX - LSTM = Forecast\ ARIMAX + Forecast\ Residual\ ARIMAX - LSTM \quad (1)$$

H. *Forecasting Performance*

The effectiveness of the Hybrid ARIMAX-LSTM forecasting model is assessed by calculating the SMAPE (Symmetric Mean Absolute Percentage Error), RMSE (Root Mean Squared Error), and MAE (Mean Absolute Error) values [12][21].

IV. Results and Discussion

The conventional data obtained is shown in Figure 2. The data in Figure 2 exhibit fluctuations, but the range of values remains small, between 0 and 12, with only one period reaching a value of 26. This occurred in January, during the rainy season and an outbreak of Dengue Fever.

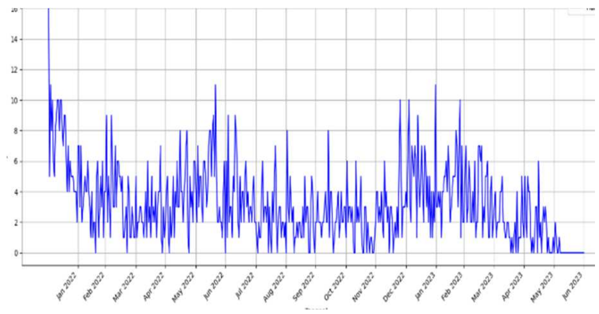

Fig. 2. Time-series distribution of reported conventional DF cases

The stationarity test using the Augmented Dickey-Fuller Test on the conventional Dengue Fever case data from the Health Office shows a p-value of 0.004, which is less than the specified alpha of 0.005, indicating that the data is stationary. Therefore, differencing is not necessary [12].

To obtain forecasting results, experiments were conducted in five scenarios as shown in Table I. The performance of the forecasting results for both training and testing data, using ARIMA/ARIMAX in each scenario, is shown in Table II. Table II shows that Scenario 5, which involves only social media data, has the best performance compared to the other scenarios that involve conventional data. Similarly, Scenarios 2 and 3, which include social media data as exogenous variables, also perform well, although the differences are not very significant. Forecasting using the ARIMAX model indicates that the data does not exhibit significant trends or seasonal movements. The forecast values remain constant throughout the forecasted time period, without showing dependence on actual values.

TABLE I. Experiment Scenarios For Model Evaluation

Sce-nario	Data Used for Dengue Fever (DF)	Exogenous Variables
1	Conventional DF Data from Health Office	Rainfall, humidity levels, and wind speed
2	Conventional DF Data from Health Office	Social media data
3	Conventional DF Data from Health Office	Rainfall, humidity levels, wind speed, and social media data
4	Conventional DF Data from Health Office	Previous period conventional DF data
5	Social media data	Previous period social media data

TABLE II. FORECASTING PERFORMANCE USING ARIMA/ARIMAX MODELS IN EACH SCENARIO

Model	Training			Testing		
	RM SE	SMAPE	MAE	RM SE	SMAPE	MA E
ARIMAX	2.19	26.78%	1.65	3.00	58.77%	2.73
ARIMAX	2.18	26.78%	1.64	3.02	58.90%	2.75
ARIMAX	2.19	26.60%	1.64	3.02	58.89%	2.75
ARIMA	2.18	26.77%	1.65	3.02	58.90%	2.75
ARIMA	0.51	11.54%	0.15	0.19	3.64%	0.04

TABLE III. RESIDUAL FORECASTING PERFORMANCE USING ARIMA/ARIMAX-LSTM MODEL IN EACH SCENARIO

Sce-nario	Model					
	ARIMAX			HYBRID ARIMAX-LSTM		
	RM SE	SMAPE	MAE	RM SE	SMAPE	MAE
1	3.01	58.77%	2.73	1.71	53.08%	1.53
2	3.02	58.90%	2.75	1.71	53.34%	1.55
3	3.02	58.89%	2.75	1.72	53.26%	1.55
4	3.02	58.90%	2.75	1.71	53.34%	1.55
5	0.19	3.64%	0.04	0.19	3.64%	0.04

Moreover, forecasting results using conventional dengue fever case data without exogenous variables do not yield better outcomes than forecasting scenarios using climate and social media data as exogenous variables. These results cannot yet be compared to others because no similar forecast has been made using Instagram data. The residual plots of the forecasting results using Scenarios 3 and 5 are shown in Figure 3.

The residuals appear irregular and noisy because there are nonlinear patterns or sudden events. The LSTM forecasting results using ARIMAX/ARIMA residual data for several scenarios are shown in Figure 4. The residual forecasting results using LSTM were then combined to obtain the final forecast. The hybrid ARIMA-LSTM forecasting results for the three scenarios are presented in Figure 5, while the performance is summarized in Table III. From Table III, it can be observed that the Hybrid ARIMAX-LSTM model provides lower SMAPE values than the ARIMAX model in all scenarios. Regarding the impact of using exogenous variables such as climate data and social media data, the smallest SMAPE value in the forecast results is found in Scenario 1, which uses climate data as the exogenous variable.

Additionally, the SMAPE value in scenarios using social media data as an exogenous variable does not yield better results than those in Scenario 1. This suggests that the inclusion of Instagram social media data does not significantly impact the SMAPE value when combined with climate variables. For Scenario 4, which involves forecasting using conventional dengue fever case data without exogenous variables, it is evident that the SMAPE value is not improved over that in Scenario 1, which utilizes climate data as the exogenous variable. This indicates that using climate data as an exogenous variable can provide better performance than Scenario 4. These findings are consistent with research conducted by [13] and [21]. Scenario 5 cannot be used as a forecasting reference because the correlation value between conventional dengue fever data and social media data is 0.11, indicating a weak positive relationship, making it unsuitable for use as a forecasting reference.

To assess the robustness of the model, experiments were also conducted on data with weekly and monthly periods in addition to daily data. The comparison of SMAPE across different time period scenarios on a weekly basis is shown in Table IV. Table IV shows that the Hybrid ARIMAX-LSTM model provides lower SMAPE values than the ARIMAX model across all scenarios for weekly data. The forecast with the smallest SMAPE value is found in Scenario 3, which uses climate data and social media data as exogenous variables. When comparing the SMAPE values between Scenario 1 and Scenario 2, Scenario 2 has a larger SMAPE value than Scenario 1, indicating that forecasting with climate data as an exogenous variable yields better performance compared to using social media data as an exogenous variable.

In Scenario 3, dengue fever cases were forecasted using exogenous variables that included both climate data and social media data. The results from Scenario 3 show a significant decrease in SMAPE value compared to Scenarios 1 and 2. This suggests that incorporating social media data as an exogenous variable yields improved performance when combined with climate data. For Scenario 4, which involves forecasting using conventional dengue fever case data without exogenous variables, the SMAPE value remains unchanged from that in Scenario 2. This suggests that using social media data as an exogenous variable does not yield better performance than when exogenous variables are included, which can enhance forecasting accuracy.

TABLE IV. RESIDUAL FORECASTING PERFORMANCE USING ARIMA/ARIMAX-LSTM ON WEEKLY DATA

Sce-nario	Model					
	ARIMAX			HYBRID ARIMAX-LSTM		
	RM SE	SMAPE	MAE	RM SE	SMAPE	MAE
1	51.63	47.43%	25.25	25.12	32.97%	11.06
2	14.78	41.99%	12.56	11.21	37.70%	9.37
3	46.25	46.95%	23.5	22.71	28.52%	9.94
4	15.44	42.61%	13.13	11.12	37.23%	9.12
5	0.9	77.08%	0.81	0.9	77.08%	0.81

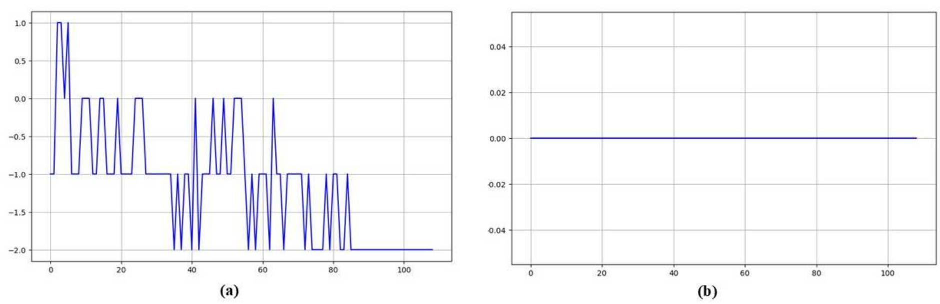

Fig. 3 Residual Forecast Plot for Dengue Fever Cases using ARIMAX Model for (a) Scenario 3, (b) Scenario 5

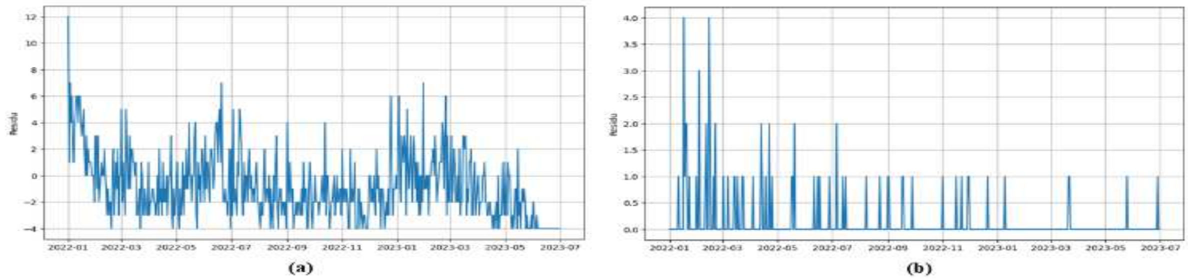

Fig. 4. Residual Forecast Plot for Dengue Fever Cases using LSTM Model for (a) Scenario 3, (b) Scenario 5

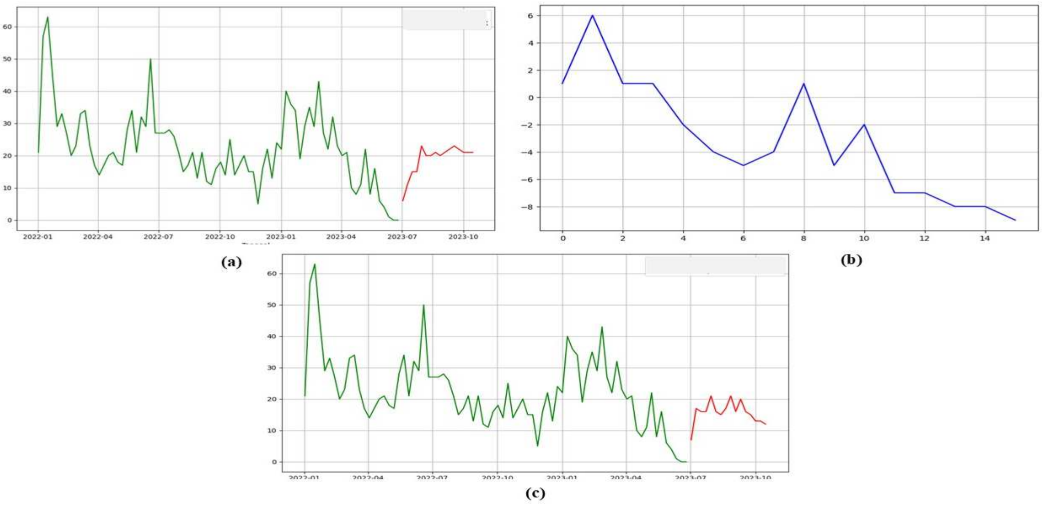

Fig. 5. Forecast Plot for Dengue Fever Cases using ARIMAX - LSTM model for (a) Scenario 2, (b) Scenario 3, (c) Scenario 5

TABLE V. RESIDUAL FORECASTING PERFORMANCE EVALUATION OF ARIMAX AND ARIMAX-LSTM HYBRID MODEL USING MONTHLY DATA

Sce-nario	Model					
	ARIMAX			HYBRID ARIMAX-LSTM		
	RMSE	SMAPE	MAE	RMSE	SMAPE	MAE
1	61.64	44.43%	57.50	48.35	38.12%	42.01
2	77.56	44.22%	65.75	64.62	37.83%	50.25
3	54.46	41.21%	48.25	42.98	35.49%	35.50
4	80.82	45.62%	70.50	64.65	35.96%	48.25
5	10.58	54.17%	10.50	10.22	45.83%	10.01

The performance of the hybrid model using monthly period data is presented in Table V. As shown in Table V, the Hybrid ARIMAX-LSTM model yields lower SMAPE values than the ARIMAX model across all scenarios. The forecast with the smallest SMAPE value is found in Scenario 3, which uses climate data and social media data as exogenous variables. When comparing the SMAPE values between Scenario 1 and Scenario 2, Scenario 1 has a larger SMAPE value than Scenario 2, indicating that forecasting using social media data as an exogenous variable yields better performance compared to using climate data as an exogenous variable. The results from Scenario 3 show a decrease in SMAPE value compared to Scenarios 1 and 2. This suggests that incorporating social media data as an exogenous variable yields improved performance when combined with climate data.

Subsequently, the forecasting using conventional dengue fever case data without exogenous variables for Scenario 4 shows that the SMAPE value is not superior to that of Scenario 3. This suggests that the inclusion of exogenous variables can lead to better performance than Scenario 4, which does not incorporate exogenous variables. This finding is consistent with the results reported by [21]. It helps mitigate social media bias and data sparsity caused by unrepresentative users by integrating non-social media data and data augmentation techniques.

V. CONCLUSION

In general, the Hybrid ARIMAX-LSTM model demonstrates superior performance in forecasting dengue fever cases compared to the ARIMAX model alone. The use of exogenous variables generally improves forecast performance. Climate exogenous variables alone generally yield better performance than social media alone. However, when used together, they produce even better performance. This suggests that climate variables can enhance performance, and Instagram social media variables will further improve forecasting performance when combined with climate exogenous variables. The combined use of exogenous variables yields better performance than using them separately. It can help mitigate biases and data sparsity in social media analysis, which arise from unrepresentative user populations, dominance of certain groups, limited data volume, and uneven data distribution, by incorporating non-social-media data through data triangulation and data augmentation

For future development, the Hybrid ARIMAX-LSTM model and the combination of exogenous variables can be tested with more varied data training and testing proportions. This is necessary to examine the consistency of the findings. Additionally, other hybrid models can be used to further improve performance. The results of this experiment can be considered when using Instagram social media data for forecasting the spread of case numbers. Instagram social media data can serve as an alternative data source to complement the conventional data available from health offices or hospitals.

ACKNOWLEDGMENT

We extend our gratitude to the Ministry of Education, Culture, Research, and Technology for providing research funding under contract numbers 017/C3/DT.05.00/PL/2025, 1329/PKS/ITS/2025, and to the Public Health Office of Malang Regency for their support with the data.

REFERENCES

[1] L. Dey and A. Mukhopadhyay, "Compact Genetic Algorithm-Based Feature Selection for Sequence-Based Prediction of Dengue-Human Protein Interactions," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 4, pp. 2137–2148, 2022, doi: 10.1109/TCBB.2021.3066597.

[2] "Dengue and severe dengue." Accessed: Sep. 24, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue

[3] A. E. Wilson, C. U. Lehmann, S. N. Saleh, J. Hanna, and R. J. Medford, "Social media: A new tool for outbreak surveillance," *Antimicrob Steward Healthc Epidemiol*, vol. 1, no. 1, p. e50, 2021, doi: 10.1017/ash.2021.225.

[4] "Association between social media use and health promotion among individuals with depression and anxiety: Insights from the 2017–2020 Health Information National Trends Survey - ScienceDirect." Accessed: Aug. 12, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2950004423000068

[5] D. Kellner, M. Lowin, and O. Hinz, "Improved healthcare disaster decision-making utilizing information extraction from complementary social media data during the COVID-19 pandemic," *Decision Support Systems*, vol. 172, p. 113983, Sep. 2023, doi: 10.1016/j.dss.2023.113983.

[6] A. H. R. Jokar, S. Roche, and H. Karimi, "Stuttering on Instagram: What is the focus of stuttering-related Instagram posts and how do users engage with them?," *Journal of Fluency Disorders*, vol. 78, p. 106021, Dec. 2023, doi: 10.1016/j.jfludis.2023.106021.

[7] S. Yip, N. Che Him, N. I. Jamil, D. He, and S. K. Sahu, "Spatio-temporal detection for dengue outbreaks in the Central Region of Malaysia using climatic drivers at mesoscale and synoptic scale," *Climate Risk Management*, vol. 36, p. 100429, Jan. 2022, doi: 10.1016/j.crm.2022.100429.

[8] J. A. Gutierrez, K. Laneri, J. P. Aparicio, and G. J. Sibona, "Meteorological indicators of dengue epidemics in non-endemic Northwest Argentina," *Infectious Disease Modelling*, vol. 7, no. 4, pp. 823–834, Dec. 2022, doi: 10.1016/j.idm.2022.10.004.

[9] Mr. S. G. Shaikh, Dr. B. SureshKumar, and Dr. G. Narang, "Development of optimized ensemble classifier for dengue fever prediction and recommendation system," *Biomedical Signal Processing and Control*, vol. 85, p. 104809, Aug. 2023, doi: 10.1016/j.bspc.2023.104809.

[10] S. Ismail, R. Fildes, R. Ahmad, W. N. Wan Mohamad Ali, and T. Omar, "The practicality of Malaysia dengue outbreak forecasting model as an early warning system," *Infectious Disease Modelling*, vol. 7, no. 3, pp. 510–525, Sep. 2022, doi: 10.1016/j.idm.2022.07.008.

[11] M. Panja, T. Chakraborty, S. S. Nadim, I. Ghosh, U. Kumar, and N. Liu, "An ensemble neural network approach to forecast Dengue outbreak based on climatic condition," *Chaos, Solitons & Fractals*, vol. 167, p. 113124, Feb. 2023, doi: 10.1016/j.chaos.2023.113124.

[12] W. Anggraeni and L. Aristiani, "Using Google Trend data in forecasting number of dengue fever cases with ARIMAX method case study: Surabaya, Indonesia," in *2016 International Conference on Information Communication Technology and Systems (ICTS)*, Oct. 2016, pp. 114–118. doi: 10.1109/ICTS.2016.7910283.

[13] W. Anggraeni, *et al.*, "A Sparse Representation of Social Media, Internet Query, and Surveillance Data to Forecast Dengue Case Number using Hybrid Decomposition Bidirectional LSTM," *IJIES*, vol. 14, no. 5, pp. 209–225, Oct. 2021, doi: 10.22266/ijies2021.1031.20.

[14] B. Jang, Y. Kim, G. Il Kim, and J. Wook Kim, "Deep similarity analysis and forecasting of actual outbreak of major infectious diseases using Internet-Sourced data," *Journal of Biomedical Informatics*, vol. 133, p. 104148, Sep. 2022, doi: 10.1016/j.jbi.2022.104148.

[15] A. Husnayain, A. Fuad, and L. Lazuardi, "Correlation between Google Trends on dengue fever and national surveillance report in Indonesia," *Global Health Action*, vol. 12, no. 1, p. 1552652, Jan. 2019, doi: 10.1080/16549716.2018.1552652.

[16] R. A. Strauss, J. S. Castro, R. Reintjes, and J. R. Torres, "Google dengue trends: An indicator of epidemic behavior. The Venezuelan Case," *International Journal of Medical Informatics*, vol. 104, pp. 26–30, Aug. 2017, doi: 10.1016/j.ijmedinf.2017.05.003.

[17] M. Verma, K. Kishore, M. Kumar, A. R. Sondh, G. Aggarwal, and S. Kathirvel, "Google Search Trends Predicting Disease Outbreaks: An Analysis from India," *Healthc Inform Res*, vol. 24, no. 4, p. 300, 2018, doi: 10.4258/hir.2018.24.4.300.

[18] M. Kapitány-Fövény *et al.*, "Can Google Trends data improve forecasting of Lyme disease incidence?," *Zoonoses and Public Health*, vol. 66, no. 1, pp. 101–107, 2019, doi: 10.1111/zph.12539.

[19] Z. Du, L. Xu, W. Zhang, D. Zhang, S. Yu, and Y. Hao, "Predicting the hand, foot, and mouth disease incidence using search engine query data and climate variables: an ecological study in Guangdong, China," *BMJ Open*, vol. 7, no. 10, Oct. 2017, doi: 10.1136/bmjopen-2017-016263.

[20] J. Xu *et al.*, "Forecast of Dengue Cases in 20 Chinese Cities Based on the Deep Learning Method," *Int J Environ Res Public Health*, vol. 17, no. 2, Jan. 2020, doi: 10.3390/ijerph17020453.

[21] W. Anggraeni *et al.*, "A hybrid EMD-GRNN-PSO in intermittent time-series data for dengue fever forecasting," *Expert Systems with Applications*, vol. 237, p. 121438, Mar. 2024, doi: 10.1016/j.eswa.2023.121438.