Infodemiological Approach to Disease Surveillance: A Case Study on the COVID-19 Pandemic

Mon Gabriel F. Lagustan and Maria Regina Justina E. Estuar

Department of Information Systems and Computer Science,
Ateneo de Manila University, Quezon City 1108, Philippines
mon.lagustan@student.ateneo.edu, restuar@ateneo.edu

Abstract. Conventional surveillance methods have encountered limitations in delivering timely insights on disease transmission prompting exploration on extending surveillance using alternative strategies. This study proposes an infodemiological approach in disease surveillance, specifically harnessing Google Trends data as an indicator of health seeking behavior pertaining to a disease. The study aims to provide a method for early detection of disease transmission using popular search terms as signals for the spread. This study uses the COVID-19 pandemic as a case study, with selected sites including National Capital Region (NCR), Region 4A (Calabarzon) and Region 3 (Central Luzon), as these specific regions contain the most active number of cases in the Philippines. An ARIMA model, enhanced with Google Trends data pertaining to English and Filipino search queries associated with COVID-19 symptoms and related terms were utilized and used as an exogenous variable in ARIMA to predict COVID-19 activity in the said regions. This study also compares the performance of the proposed model—ARIMAX, with traditional Autoregressive (AR), standard ARIMA models, and Long Short-Term Memory (LSTM) networks to evaluate efficacy in predicting cases. ARIMAX demonstrated lower score error and higher correlation in actual cases compared to other models. To assess the robustness of the ARIMAX model, rolling window cross-validation was applied over a two-year dataset. Thus, using Google Trends keywords related to COVID-19 symptoms in both Filipino and English improved the predictive accuracy of the ARIMA model. The case study proposes an additional disease surveillance tool incorporating online health seeking behavior into time series model for early detection of possible outbreaks. This study contributes to the broader discourse on innovative approaches to enhance disease surveillance and response strategies.

Keywords: Infodemiology · Syndromic Surveillance · COVID-19 · Google Trends · ARIMAX · ARIMA · Autoregressive · LSTM

1 Introduction

As the Philippines confronts the complex challenges posed by the virus, there arises a critical need for a comprehensive and adaptable methodology to monitor, predict, and manage the spread of COVID-19. To address these difficulties, keeping track of how people seek health related information through web searches has become a valuable method for promptly identifying health issues as they arise in specific regions and timeframes [1].

Studies have utilized Google Trends data to extract relevant information for health-seeking behaviors like symptoms of COVID-19 (i.e. cough, fever, sore throat, etc.) at the word level that could not have been otherwise detected. These investigations laid the groundwork for what is known as infodemiology (or information epidemiology), a methodology for analyzing data related to epidemiology that offers quicker access to information compared to traditional epidemiological studies and can reveal information that might otherwise go unnoticed [2,3,4,5].

Google Trends has emerged as a valuable resource for tracking and analyzing disease trends. The health officials can gain valuable insights into the spread and prevalence of various diseases by harnessing the vast amount of data generated by internet search queries. In 2008, Google launched Google Flu Trends (GFT), an internet-based surveillance tool that uses aggregated Google search data to estimate influenza activity in near-real time [17]. This pioneering project inspired later research, including a study that used Google Trends data with ARIMA models to predict influenza activity in Italy. The study showed significant correlations with influenza spread in Greece and Italy in 2011 and 2012 [16]. It demonstrated high precision in forecasting influenza peaks, aiding public health planning. GT's potential to provide timely and widespread surveillance of COVID-19 is comparable to other syndromic surveillance systems. However, it should complement traditional disease surveillance methods, prompting further investigation and collection of direct disease activity measures.[15].

ARIMA or autoregressive integrated moving average is a machine learning technique used for handling time series data. ARIMA is reported to have high accuracy with small datasets [6]. ARIMA models have gained popularity because of its ability to accurately do short-term forecasts [7,8]. In the domain of time series analysis, ARIMA models have received attention and application in forecasting epidemic diseases [10,11,12,13,14]. Notably, despite being frequently disregarded in pandemic forecasting due to perceived limitations in handling complex and dynamic scenarios, ARIMA models have shown potential in delivering favorable outcomes.

While existing works have utilized the ARIMA model for prediction, this research advances the field by incorporating Filipino keywords related to COVID-19 symptoms, capturing the health-seeking behavior of local Filipinos during pandemics. This study presents an infodemiological approach to syndromic surveillance for COVID-19 in the Philippines. Leveraging Google search data and ARIMA, this approach provides early warning signals and insights into the spread of COVID-19 across different regions of the Philippines. The findings can inform strategies for future health activities, guiding proactive measures for a resilient healthcare system. The model helps predict infection increases by identifying regions with elevated COVID-19 search activity, aiding targeted interventions, surveillance, and public health campaigns to curb the virus spread.

2 Methodology

This study demonstrates the potential of infodemiology for syndromic surveillance of COVID-19 in various regions of the Philippines by incorporating quantified Google search data in the ARIMA model (kindly refer to Fig. 1).

Fig. 1: Framework

2.1 Dataset Preparation

Google Trends data were gathered from National Capital Region (NCR), Region 4A (Calabarzon) and Region III (Central Luzon) in the Philippines, covering the period from February 2020 to December 2023. This timeframe encompasses the period prior to the pandemic's peak until the point where the number of cases stabilized.

Relative Search Volumes (RSV) of COVID-19 symptoms and related terms were used as keywords ("cough + ubo", "fever + lagnat", "covid symptoms + sintomas ng covid", "flu + trangkaso", and "swab test") for obtaining the Google Trends data across different regions in the Philippines.

COVID-19 cases data was collected for the mentioned time period using the data provided by FASSSTER (Feasibility Analysis of Syndromic Surveillance Using Spatio-Temporal Epidemiological ModeleR) team and Department of Health (DOH) in the Philippines. Data preprocessing was done in Python to get the daily confirmed cases from the said time period in the National Capital Region (NCR), Calabarzon Region, and Central Luzon Region.

2.2 Choosing the Exogneous variable for the ARIMA model

Pearson's r correlation was used to determine which keywords from the list including: "cough + ubo", "fever + lagnat", "covid symptoms + sintomas ng covid", "flu + trangkaso", and 'swab test" will be used in the model. The "+" character was used to aggregate data from multiple searches using a logical OR, allowing the Relative Search Volume (RSV) of each keyword to be summed. The correlations were computed using the RSV of each keyword and the actual COVID-19 case counts recorded before the peak. The keyword demonstrating the strongest correlation with the preceding month's covid data will be selected as the exogenous variable to forecast the subsequent month's case volume. This addition aims to account for the influence of new possible infections and dynamics on the time series being forecasted, offering a more comprehensive and contextually relevant forecasting framework. The ARIMA model, incorporating an exogenous variable, will be trained every 24 hours using the current month's daily confirmed COVID-19 cases. This model will then predict the following month's confirmed cases, continually updating with the latest daily data.

2.3 Comparative Performance Assessment of the models

The forecasting of COVID-19 cases was achieved using various Python libraries for data analysis, modeling, and evaluation. Data preprocessing was performed with pandas and numpy, while matplotlib was used for visualization. The statsmodels library facilitated ARIMA model training, with pmdarima automating the selection of optimal parameters (p, d, q). The scipy library was used to calculate Pearson correlations. For machine learning, scikit-learn provided data scaling and evaluation metrics, and keras was used to build and train LSTM models. Evaluation metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) were calculated to assess model performance, with math being used for additional operations. This integrated approach of traditional statistical methods and advanced machine learning models enabled accurate forecasting of COVID-19 case trends.

The performance of the models ARIMAX (Autoregressive Integrated Moving Average with exogenous variable) and standard ARIMA were assessed to evaluate the impact of incorporating external factors on the model's predictive accuracy. Furthermore, other models were compared such as AutoRegressive (AR) and LSTM (Long Short-Term Memory).

Accuracy of forecast was evaluated in using the following measures: Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). These metrics provide quantitative measures of the differences between predicted values and actual values in forecasting models. RMSE calculates the square root of the average squared differences between predicted and actual values, providing a measure of the model's forecasting accuracy while penalizing large errors more heavily. MAE, on the other hand, computes the average of the absolute differences between predicted and actual values, offering a simpler measure of forecasting accuracy that is less sensitive to outliers. Pearson correlation was also computed to compare if the model was able to predict the trend of the COVID-19 cases especially during the peak of cases.

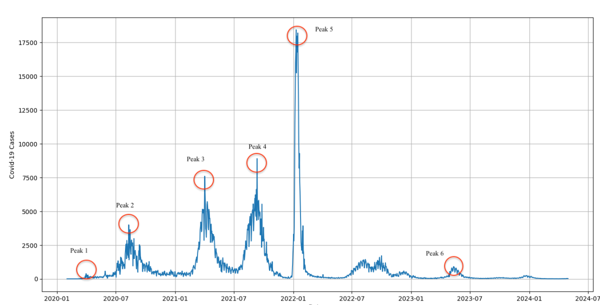

Fig. 2: NCR Confirmed Cases from 2020-2023

3 Results and Discussion

This section presents the results of the study in the three selected regions, namely: This covers results on National Capital Region (NCR), Calabarzon or Southern Tagalog Mainland (Region 4A), and Central Luzon Region (Region 3), as these regions contained the highest number of COVID-19 reported cases across the Philippines. The peak was determined every year by choosing the timeframe that contains most number of confirmed cases in historical data starting from 2020 until 2023 refer to Figure 2.

The model was tested during peaks of COVID-19 confirmed cases to test the model's capability on detecting COVID-19 outbreaks. ARIMAX, ARIMA, AR, and LSTM were compared by using different metrics such as Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Pearson correlation to test its accuracy with actual historical confirmed cases data.

Table 1: NCR Summary of Results

NCR		Timeline	ARIMAX	ARIMA	AR	LSTM	GT
Peak 1		Feb-Mar 2020					"cough + ubo"
	MSE		1410.44	1582.00	2946.11	2948.04	
	MAE		16.00	17.23	25.50	25.51	
	RMSE		37.55	39.77	54.27	54.29	
	Corr.		0.6903	0.6714	nan	-0.7096	
	p-value		4.2e-07	1.1e-6	nan	nan	
Peak 2		May-Aug 2020					"swab test"
	MSE		270536.62	269372.59	603576.91	1383239.49	
	MAE		355.58	352.29	489.93	819.88	
	RMSE		520.13	519.01	776.90	1176.11	
	Corr.		0.8167	0.8164	0.7209	0.7852	
	p-value		6.2e-17	6.5e-17	8.6e-12	5.9e-15	
Peak 3		Jan-Apr 2021					"covid symptoms"
	MSE		560509.08	584486.99	7979761.13	7721422.73	
	MAE		459.92	463.30	2154.47	2100.20	
	RMSE		748.67	764.51	2824.84	2778.74	
	Corr.		0.9290	0.9253	0.9315	0.9197	
	p-value		1.8e-28	8.2e-28	6.2e-29	7.1e-27	
Peak 4		Jul-Sep 2021					"flu + trangkaso"
	MSE		862196.63	867538.07	2433700.87	7246897.46	
	MAE		678.83	676.95	1192.25	2255.19	
	RMSE		928.54	931.41	1560.03	2692.00	
	Corr.		0.8500	0.8495	0.7913	0.8803	
	p-value		4.2e-16	4.6e-16	1.0e-24	1.7e-18	
Peak 5		Dec-Jan 2022					"cough + ubo"
	MSE		2340127.83	2394961.96	59650170.99	81549832.54	
	MAE		831.80	844.59	4766.17	5543.65	
	RMSE		1529.74	1547.56	7723.35	9030.49	
	Corr.		0.9777	0.9771	0.9831	0.8986	
	p-value		1.3e-22	2.0e-22	1.8e-24	1.2e-12	
Peak 6		Mar-May 2023					"covid symptoms"
	MSE		10355.86	10693.41	175937.24	109034.61	
	MAE		62.19	62.96	281.30	216.16	
	RMSE		101.76	103.40	419.44	330.20	
	Corr.		0.9305	0.9277	-0.9442	0.9424	
	p-value		2.4e-24	6.5e-24	9.3e-27	2.1e-26	

Before the actual peak of COVID-19 cases in March 2020 (first peak), the Google Trends (GT) keyword "cough + ubo" showed the highest correlation with the current daily confirmed cases (refer to Table 2). Integrating this variable into the ARIMA model (forming ARIMAX) improved prediction accuracy, where

Table 2: Correlation of GT keywords with actual cases before actual peaks

	Correlation		Correlation
Peak 1	**cough: 0.4701** fever: 0.3967 covid symptoms: 0.4304 flu: 0.2716 swab test: nan	Peak 4	cough: -0.2107 fever: -0.0410 covid symptoms: 0.6416 **flu: 0.6955** swab test: 0.5920
Peak 2	cough: 0.4284 fever: 0.5269 covid symptoms: 0.4849 flu: 0.2495 **swab test: 0.7087**	Peak 5	**cough: 0.8091** fever: 0.5483 covid symptoms: 0.5980 flu: 0.5241 swab test: 0.6982
Peak 3	cough: 0.8184 fever: 0.7260 **covid symptoms: 0.8836** flu: 0.5888 swab test: 0.7822	Peak 6	cough: 0.3662 fever: 0.6416 **covid symptoms: 0.7603** flu: 0.2247 swab test: nan

ARIMAX accurately predicted the trend outbreak peaks and got a lower error score and higher correlation with actual COVID-19 cases compared to standard ARIMA and other models (refer to Table 1).

In terms of visualization trends, a comparison of ARIMAX with AR and LSTM models reveals that ARIMAX demonstrates stronger trend correlation with actual cases (refer to Figure 4). While the visual similarity between standard ARIMA and ARIMAX may suggest comparable performance, a closer examination reveals distinctions in metric errors.

Before the second peak (May 15 - August 20, 2020), "swab test" emerged as the most correlated keyword (shown in Table 2). This surge in interest mirrored the ramp-up of testing in Metro Manila. Incorporating this keyword into the ARIMA model again enhanced prediction accuracy, demonstrating the model's adaptability to changing public concerns.

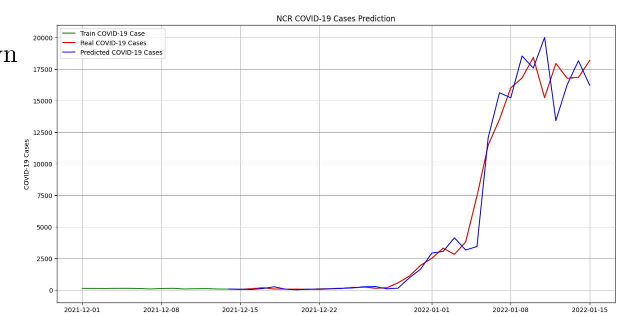

Fig. 3: Region NCR Peak 5 Confirmed Cases (ARIMAX)

In the third peak (January-April 2021), the resurgence of "cough + ubo" as a top keyword highlighted recurring symptoms associated with new COVID-19 variants. The ARIMA model with this keyword maintained high prediction accuracy (correlation 0.9290), underscoring the persistent relevance of specific symptoms.

The fourth peak (July-September 2021) saw "flu + trangkaso" as the dominant keyword, reflecting public anxiety about flu-like symptoms during the spread of the Omicron variant. The model's high prediction accuracy (correlation 0.85) during this period illustrated its ability to capture the nuanced shifts in public health concerns. By the fifth peak (January 2022), "cough + ubo" resurfaced, correlating strongly with the highest number of confirmed cases. The ARIMA model with this keyword achieved an impressive correlation (0.9777), reinforcing the critical role of tracking symptom-related searches. The graph can be viewed in Figure 3. Lastly on year 2023, March 1 to May 15 2023, which was the sixth peak of COVID-19, the most correlated GT keyword circled back to "covid symptoms". Since this keyword has the strongest correlation with the confirmed cases, this keyword then was incorporated into the ARIMA model. It can be implied that people started to be curious again on what are the covid symptoms which may indicate possible infections. Overall, the consistent improvement of the ARIMAX model across different COVID-19 peaks highlights the value of incorporating real-time public search data. This approach not only enhances predictive accuracy but also provides valuable insights into the evolving nature of public health crises. Given these findings, the predictions can be utilized as an early warning signal, serving as a proxy for when people are likely to get infected in the upcoming month. The model forecasts the next month's infections by leveraging the knowledge from the previous month and incorporating Google Trends data from the current month to enhance the ARIMAX model's accuracy.

The same methodology was applied to the Calabarzon (Region 4A) and Central Luzon (Region 3), with results shown in Tables 3 and 4. In both cases, GT keywords improved the ARIMAX score, resulting in lower errors and higher correlation with actual cases during peak prediction.

Table 4: Region 3 Summary of Results

Region 3		Timeline	ARIMAX	ARIMA	AR	LSTM	GT
Peak 1		Feb-April-2021					"swab test"
	MSE		31220.21	32374.08	154862.52	198797.99	
	MAE		105.09	105.77	276.82	311.53	
	RMSE		176.69	179.92	393.52	445.86	
	Corr.		0.8770	0.8734	0.8734	0.9034	
	p-value		2.8e-15	4.9e-15	5.34e-15	2.0e-17	
Peak 2		July-Sept15-2021					"covid symptoms"
	MSE		172080.64	175080.22	1314219.14	1061726.89	
	MAE		324.62	326.23	988.33	879.40	
	RMSE		414.82	418.42	1146.39	1030.40	
	Corr.		0.8180	0.8157	0.7708	0.8728	
	p-value		4.3e-14	5.7e-14	9.2e-12	7.8e-18	
Peak 3		Dec-Jan-2022					"covid symptoms"
	MSE		118784.08	140065.92	4552417.71	2704966.53	
	MAE		193.76	210.80	1192.33	925.47	
	RMSE		344.65	374.25	2133.63	1644.67	
	Corr.		0.9739	0.9675	0.9781	0.9115	
	p-value		1.4e-21	4.5e-20	1.0e-22	1.6e-13	

Table 3: Region 4A Summary of Results

Region 4A		Timeline	ARIMAX	ARIMA	AR	LSTM	GT
Peak 1		June-Aug-2020					"swab test"
	MSE		66890	68102	206565	184867	
	MAE		148	150	307	279	
	RMSE		260	279	454	429	
	Corr.		0.709	0.701	0.654	0.739	
	p-value		1.8e-9	3.3e-9	8.0e-8	1.6e-10	
Peak 2		Jan-March-2021					"swab test"
	MSE		26128	27001	349565	309381	
	MAE		107.73	107.73	364.66	342.00	
	RMSE		162.64	164.32	591.24	556.22	
	Corr.		0.9479	0.9457	0.9442	0.9762	
	p-value		5.3e-27	1.4e-26	2.9e-26	5.1e-34	
Peak 3		July-Sept-2021					"flu + trangkaso"
	MSE		369981.68	373714.20	3925212.51	4876395.03	
	MAE		466.51	464.66	1667.36	1858.44	
	RMSE		608.26	611.32	1981.21	2208.25	
	Corr.		0.902	0.901	0.87	0.91	
	p-value		1.2e-20	1.5e-20	1.9e-18	1.0e-22	
Peak 4		December-Jan-2022					"cough + ubo"
	MSE		779767.41	786770.88	11054623.387	19977119.634	
	MAE		562.13	566.77	2172.73	2910.86	
	RMSE		883.04	887.00	3324.84	4469.57	
	Corr.		0.9678	0.9674	0.9674	0.8025	
	p-value		5.9e-22	7.2e-22	7.2e-22	3.9e-9	
Peak 5		March-May-2023					"fever + lagnat"
	MSE		2790.55	3020.64	33423.441	36191.64	
	MAE		34.19	34.55	114.60	118.17	
	RMSE		52.82	54.96	182.82	190.24	
	Corr.		0.9457	0.9400	0.9372	0.9304	
	p-value		4.7e-27	5.8e-26	1.9e-25	2.4e-24	

Table 5: Rolling Window Cross-Validation: ARIMAX vs ARIMA

Model	**ARIMAX**	**ARIMAX**	**ARIMAX**	**ARIMAX**	**ARIMAX**	**ARIMA**
Google Trend	**ubo + cough**	**trangkaso + flu**	**sintomas ng covid + covid symptoms**	**lagnat + fever**	**swab test**	**-**
Overall Avg. MSE	247977.25	250066.75	240051.65	246508.12	246968.60	260848.01
Overall Avg. MAE	280.20	287.45	275.83	280.41	277.36	296.62
Overall Avg. RMSE	497.97	500.07	489.95	496.50	496.96	510.73
Correlation	0.943	0.942	0.945	0.943	0.943	0.940

To assess the robustness of the ARIMAX model (see Table 5), rolling window cross-validation was performed on a two-year dataset of daily COVID-19 cases, using a 30-day training period and a 60-day prediction window. The ARIMAX model included Google Trends data for keywords in English and Filipino: "cough + ubo," "flu + lagnat," "covid symptoms + sintomas ng covid," "fever + lagnat," and "swab test." The findings indicate that the ARIMAX model outperformed the ARIMA model, as reflected by its lower MSE, MAE, and RMSE values, along with a stronger correlation with actual cases. This suggests that the integration of search volume data enhances predictive accuracy and overall model performance.

4 Conclusion and Future Plans

Using Google Trends keywords related to COVID-19 symptoms in both Filipino and English improved the predictive accuracy of the ARIMA model. Keywords like "cough + ubo," "fever + lagnat," "covid symptoms + sintomas ng covid", "flu + trangkaso," and "swab test" enhanced the model's ability to forecast COVID-19 outbreaks, especially during peak periods. The ARIMAX model effectively captured case trends, providing critical insights for health officials' planning and response efforts. Analyzing the Relative Search Volumes (RSV) of COVID-19 symptoms revealed variations in health-seeking behaviors among Filipinos during different peaks. These keywords identified prevalent symptoms that can aid health officials in targeted response and mitigation strategies. The infodemiological approach provides timely data, enabling the prediction and monitoring disease transmission or outbreak across the Philippines using Google Trends as a proxy for COVID-19-related searches. This study integrates online health-seeking behavior into disease surveillance, providing a data-driven framework for monitoring COVID-19 and other infectious diseases. Future applications can make use of the ARIMAX model to enhance spatio-temporal approach in measuring risk of spread or in predicting cases. This concept is being implemented in a website application currently under development. The current progress can be viewed here.

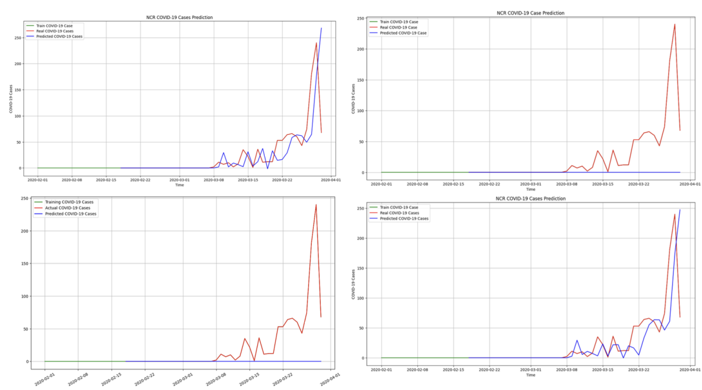

Fig. 4: ARIMAX (upperleft), AR (upper-right), LSTM (lowerleft), ARIMA (lowerright)

References

1. Szilagyi, I. S., Ullrich, T., Lang-Illievich, K., Klivinyi, C., Schittek, G. A., Simonis, H., Bornemann-Cimenti, H.: Google trends for pain search terms in the world's most populated regions before and after the first recorded COVID-19 case: Infodemiological study. Journal of Medical Internet Research **23**(4), e27214 (2021)

2. Bernardo, T. M., Rajic, A., Young, I., Robiadek, K., Pham, M. T., Funk, J. A.: Scoping review on search queries and social media for disease surveillance: a chronology of innovation. Journal of medical Internet research **15**(7), e2740 (2013)
3. Kamiński, M., Łoniewski, I., Marlicz, W.: "Dr. Google, I am in Pain"—global Internet searches associated with pain: a retrospective analysis of Google Trends data. International journal of environmental research and public health **17**(3), 954 (2020)
4. Mavragani, A.: Infodemiology and infoveillance: scoping review. Journal of medical internet research **22**(4), e16206 (2020)
5. Mavragani, A., Ochoa, G., et al.: Google Trends in infodemiology and infoveillance: methodology framework. JMIR public health and surveillance **5**(2), e13439 (2019)
6. ArunKumar, K. E., Kalaga, D. V., Kumar, C. M. S., Kawaji, M., Brenza, T. M.: 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 **61**(10), 7585–7603 (2022)
7. Albeladi, K., Zafar, B., Mueen, A.: Time Series Forecasting using LSTM and ARIMA. International Journal of Advanced Computer Science and Applications (IJACSA) **14**(1) (2023)
8. Contreras, J.: ARIMA models to predict next-day electricity process. IEEE Trans. Power Syst. **19**(1), 366–374 (2004)
9. Feng, Y., Hao, W., Li, H., Cui, N., Gong, D., Gao, L.: Machine learning models to quantify and map daily global solar radiation and photovoltaic power. Renewable and Sustainable Energy Reviews **118**, 109393 (2020)
10. Leiva, V., Saulo, H., Souza, R., Aykroyd, R. G., Vila, R.: A new BISARMA time series model for forecasting mortality using weather and particulate matter data. *Journal of Forecasting*, 40(2), 346–364 (2021). Wiley Online Library
11. Jerez-Lillo, N., Álvarez, B. L., Gutiérrez, J. M., Figueroa-Zúñiga, J. I., Leiva, V.: A statistical analysis for the epidemiological surveillance of COVID-19 in Chile. *Signa Vitae*, 18(2) (2022).
12. Ospina, R., Leite, A., Ferraz, C., Magalhaes, A., Leiva, V.: Data-driven tools for assessing and combating COVID-19 outbreaks in Brazil based on analytics and statistical methods. *Signa Vitae*, 18(3) (2022).
13. Yousaf, M., Zahir, S., Riaz, M., Hussain, S. M., Shah, K.: Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan. *Chaos, Solitons & Fractals*, 138, 109926 (2020). Elsevier
14. de Araújo Morais, L. R., da Silva Gomes, G. S.: Forecasting daily Covid-19 cases in the world with a hybrid ARIMA and neural network model. *Applied Soft Computing*, 126, 109315 (2022). Elsevier
15. Jurić, T.: Google trends as a method to predict new COVID-19 cases. *medRxiv* (2021), 2021–03.
16. Samaras, L., García-Barriocanal, E., Sicilia, M.-A., et al.: Syndromic surveillance models using web data: the case of influenza in Greece and Italy using Google Trends. *JMIR Public Health and Surveillance*, 3(4), e8015 (2017).
17. Google.org: Google Flu Trends. Google.org. Available: www.google.org/flutrends (2009). Accessed 2011 July 28.

Acknowledgements This research acknowledges the support and resources from Ateneo Social Computing Science Lab, Department of Science and Technology (DOST-ERDT), Ateneo Center for Computing Competency and Research, and the FASSSTER Project.