Indonesian Journal of Electrical Engineering and Computer Science
Vol. 34, No. 2, May 2024, pp. 1092~1101
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v34.i2.pp1092-1101

Comparison of ARIMA boost, Prophet boost, and TSLM models in forecasting Davao City weather data

Jamal Kay B. Rogers[1], Tamara Cher R. Mercado[1], Fredelino A. Galleto Jr.[2]
[1]College of Information and Computing, University of Southeastern Philippines, Davao City, Philippines
[2]Department of Electronics Engineering, College of Engineering and Information Technology, University of Southern Mindanao, Kabacan, Cotabato, Philippines

Article Info

Article history:

Received Nov 29, 2023
Revised Feb 18, 2024
Accepted Feb 20, 2024

Keywords:

ARIMA boost
Climatic data parameters
Metrics
Prophet boost
Time series forecasting
TSLM

ABSTRACT

The geography of the Philippines experiences climate variability thus, providing accurate and timely weather forecasts to the population is crucial. Climate forecasts, which are issued and disseminated by government agencies, serve as essential risk management tools. However, the country faces challenges in forecasting, further exacerbated by climate change. Thus, exploring the use of artificial intelligence has emerged as a strategy to enhance weather prediction accuracy. This research focuses on time series forecasting of rainfall, mean temperature, relative humidity, and wind speed weather data using a machine learning approach. Specifically, it aims to compare and identify the most beneficial forecasting models among autoregressive integrated moving average (ARIMA) boost, Prophet boost, and time series linear model (TSLM). It also seeks to evaluate the performance of these models using mean absolute error (MAE), mean absolute percentage error (MAPE), mean absolute scaled error (MASE), symmetric mean absolute percentage error (SMAPE), root mean squared error (RMSE), and R squared (RSQ) metrics. Results showed that the selection of the forecasting model varies based on the specific parameter under consideration, with no hyperparameter tuning in the analysis. For wind speed, ARIMA boost proves to be a favorable choice. At the same time, TSLM demonstrates effectiveness for relative humidity and mean temperature. Both ARIMA boost and TSLM exhibit strong performance for rainfall. Prophet boost consistently ranks as the least-performing model.

Corresponding Author:

Fredelino A. Galleto Jr.
Department of Electronics Engineering, College of Engineering and Information Technology
University of Southern Mindanao
Kabacan, Cotabato, Philippines
Email: fgalletojr@usm.edu.ph

1. INTRODUCTION

The variability in climate greatly influences the agriculture and natural resource sectors in the Philippines, considering its geographical location. Much of the seasonal climate variability in the country is attributed to El Niño Southern Oscillation (ENSO), leading to drier seasons in the western equatorial pacific and weak monsoon activity. The impact of seasonal variability is not only reflected through the effects of droughts and extreme floods but also in the formulation of conservative risk-averting farming approaches, which result in less efficient utilization of resources and hasten the degradation of natural resources [1]. The 1997/98 El Niño occurrence in the Philippines resulted in a 75% reduction in corn yield in the southern region, leading to an estimated loss of ₱3.2B in corn production [2]. The availability of and access to weather and climate forecasts offers a superior ability to manage risks caused by highly variable climatic conditions in the country.

The geographical location of the Philippines exposes the country to typhoons of various intensities, which carry heavy rainfall and strong winds, causing damage to crops, properties, and even human life. It is, therefore, imperative to deliver typhoon forecasts that are accurate, timely, and comprehensible to the local populace. Philippine Atmospheric, Geophysical, and Astronomical Services Administration (PAGASA) reports tropical cyclone warnings using signal numbers 1 to 5, with the extent of impacts increasing corresponding to the signal number. Climate and weather forecasts can serve as risk management tools and a basis for making decisions that will allow for the reduction of losses and maximizing of positive impacts of climate variability [3].

The Philippines faces several challenges in weather forecasting due to its geographical location and vulnerability to natural disasters such as tropical cyclones, earthquakes, and volcanoes [4]. The country is hit by an average of 20 tropical cyclones every year, and some of these weather disturbances are devastating [5]. As such, climate change exacerbates these challenges, making it difficult to predict extreme weather events such as stronger tropical cyclones and droughts. The accuracy of predictions depends on the reliability of historical data fed into the system [6]. The PAGASA is working to improve its forecasting capabilities by modernizing it’s equipment and facilities, developing impact-based forecasting and warning services, and partnering with other organizations to develop weather and climate capability in the Philippines [5], [7], [8]. The use of artificial intelligence is also being explored to improve the speed and accuracy of weather predictions. The accuracy of weather predictions is crucial for decision-makers to prepare and manage the impacts of meteorological hazards and for sectors such as agriculture to plan planting activities [6].

For climatic forecasting in the Philippines, where Davao City is located in the southern part of this country, this paper addresses the time series forecasting weather data using a machine learning approach. Specifically, to compare and determine the beneficial forecasting models among autoregressive integrated moving average (ARIMA) boost, prophet boost, and time series linear model (TSLM). Lastly, to utilize mean absolute error (MAE), mean absolute percentage error (MAPE), mean absolute scaled error (MASE), symmetric mean absolute percentage error (SMAPE), root mean squared error (RMSE), and R squared (RSQ) metrics in determining the performance of forecasting models.

Human activities are largely shaped by weather and climatic conditions, with prominent influence over various economic undertakings agriculture, fishing, construction, aviation, and business [9], [10]. The inherent effects of climate variability are even worsened by the impact of climate change and meteorological disasters, posing severe challenges to human beings. Global warming has led to extreme weather and climate events and catastrophes in recent decades, causing socioeconomic losses on a global scale. A report by the United Nations showed that countries hit by disasters between 1998 and 2017 experienced economic damage amounting to US$ 2908 billion, of which 77% was attributed to climate-related disasters. Economic losses associated with weather and climate events rocketed by 151% compared to 1978-1997 [11]. Apart from its clear-cut impact on yield and market prices, the uncertainty in climate variability is a difficult task for management, as farmers and agriculture practitioners must make crucial, climate-sensitive decisions way before the impacts of climate strikes [12]. Farmers and decision-makers in agriculture are left unprepared for the impending weather conditions, and they formulate decisions based on their understanding and indigenous knowledge of general climate patterns for their locality. This typically leads to conservative approaches that sacrifice agricultural productivity to minimize the risk of losses in disfavored years [13]. In the face of climate uncertainty, farmers must make important decisions about resource allocation- land, labor, capital, which crop to plant, what area to plant, and levels of inputs to utilize [14]. Being situated off the eastern border of Asia, the Philippines is often and severely struck by extreme climate events, significantly affecting the country’s agricultural sector. Improper decisions relative to weather and climatic conditions, such as ill-timing of crop-ping and incorrect volume of imports, would be a waste of resources or a burden to the general consumers. In the absence of pertinent knowledge on climatic conditions and forecast systems, the farming sector suffers from reduced crop yield, shortage of produce, and price surge in agricultural commodities [3].

A time series is characterized as a sequence of values ordered chronologically and assessed over sometime [15]. Time is a factor measured on a continuous basis, and values in a time series are sampled at pre-determined intervals. Forecasts using time series aimed at predicting the future movement of a time-dependent phenomenon. A given phenomenon is depicted by a variable or a set of various factors, with measurements taken at different points or intervals [16]. Time-series forecasting is a complex discipline as it entails modeling techniques to define the correlation factor with time and simultaneously with previous observations. In particular, predicting and forecasting weather using time series has been a challenging field of research that has made deliberate advancements over the years.

The primary goal of different national and international climatic programs is to curtail the adverse while magnifying the positive socioeconomic impacts of weather and climatic variations through forecasting.

Climate information and advisory services are indispensable in the decision-making and management of climate-related risks. An essential human welfare implication of climate information and forecast involves its application in agriculture [9]. A run-through of its benefits includes increased farmers' preparedness, enhanced capability of farmers to deal with highly variable climatic conditions, better ability to manage risks, improved profits due to sounder crop decisions, reduced losses in poor climatic years and increased returns in good years [10], [12], [17], [18]. Predo *et al.* [2], which aimed to assess the effects of existing seasonal climate forecasts (SCF) in the Philippines and Australia, showed that SCF was valuable in deciding the planting season for corn in the Philippines. In an Australian context, the climate forecast was used as a valuable basis for the management of opportunity cropping systems. The research depicted that SCF aided the farmers in allocating farm inputs and resources for each season, improving their profits and minimizing potential losses. Domingo *et al.* [17] demonstrated that appropriate warnings and notifications through SCFs can assist farmers in making informed decisions on the proper choice of crop, the timing of the planting period, amounts of farming inputs, and the use of other mitigating actions. The use of seasonal climate forecasting in the Philippine Rice Policy was investigated [3]. The study demonstrated that SCF could provide a practical guide in the decisions of the National Food Authority (NFA) on rice importation. Integrating SCF in NFA's import decisions can potentially minimize the import cost of the government, providing a short-term resolution to the pressing issue of rice insufficiency in the country. Due to its expediency, it is essential to note the aspects of forecasting, including its reliability, accuracy, and timeliness, as they primarily affect its usability [17].

2. MATERIALS AND METHODS

The study gathered its data from the climatology and agrometeorology division of the PAGASA. The dataset encompasses daily climatic data parameters recorded from January 1, 2018, to December 31, 2020, totaling 1,096 observations. The information was obtained from the synoptic station in Davao City, located in region 11 of the Philippines. Specific coordinates of the device where the data was collected are depicted in Table 1. The dataset comprises daily entries for rainfall, mean temperature, relative humidity, and wind speed. An 8-inch standard rain gauge and a tipping bucket rain gauge were the instruments used in recording these data. The mean temperature used a thermometer from the averaged minimum and maximum recorded temperatures (units in degrees Celsius). The relative humidity is in percentages, and a psychrometric table was used. An Aerovane was used to measure the wind speed (in meters per second). Upon request, the raw climatological data mentioned in this study [19] can be provided. The PAGASA network incorporates meteorological observations, and its quality control procedures adhere to the best practices recommended by the World Meteorological Organization.

Table 1. Synoptic station location

Synoptic station	Latitude (N)	Longitude (E)	Elevation (m)
Davao City	7.13	125.65	17.29

2.1. Machine learning tools

Machine learning tools aided in the analysis of this research. R is the data science programming language used in the processing of climatic data parameters. The user interface used is RStudio IDE, a platform-independent, user-friendly interface for the R programming language designed to enhance productivity and ease the learning curve. R, an open-source software, is employed for data analysis, visualization, and modeling [20], [21]. Its high extensibility has led to a broad and dynamic community of users and developers. The software's versatility, comprehensive statistical and graphical features, along with the abundance of contributed packages render it an invaluable instrument for researchers and data analysts. Additionally, R's capability to manage and analyze spatial data further amplifies its usefulness across diverse research disciplines. Another tool is Tidyverse, a compilation of R packages dedicated to data manipulation, visualization, and analysis, which has played a transformative role in the realms of data wrangling and exploratory data analysis [22], [23]. The specific components of tidyverse employed include ggplot2 for graphic creation, dplyr for data manipulation, and tidyr for tidying simulated climatic data parameters. The machine learning framework used in this study is tidymodels, specifically utilized for model specification and data splitting. Tidymodels facilitates R users in efficiently orchestrating the creation and evaluation of predictive models, ensuring a tidy and well-organized approach [24]. Time series forecasting, which is the central concept of this paper, was realized using timetk, which is a package of R that is a part of the tidyverse ecosystem [25]. Timetk was used to plot time series data. Lastly, modeltime is a package for time series forecasting modeling. Models include ARIMA, Prophet, and TSLM from the modeltime package. The package for time series forecasting offers a range of specialized tasks, including specification, estimation, prediction, and diagnostic checking [26].

2.2. Time series forecasting

The time series forecasting framework is depicted in Figure 1. This section elaborates on the machine learning method applied to time series data. It is composed of three stages: identification of climatic data parameters, time series analysis of training data, and metrics performance of testing data.

Figure 1. Machine learning method applied to time series data

2.3. Climatic data parameters

The acquired data from the climatology and agrometeorology division of PAGASA were utilized in time series forecasting and in determining the beneficial model. The climatic data parameters processed were rainfall, mean temperature, relative humidity, and wind speed. Although various climatic variables can be utilized, these four parameters are the only available data that encompasses 1,096 daily recorded observations. Recorded rainfall data includes -1.0, which indicates trace values of less than 0.1mm. The highest recorded rainfall was 110 mm on December 29, 2018. The highest and lowest mean temperatures logged between January 1, 2018, and December 31, 2020, are 31.6 and 24.5 degrees Celsius, respectively. The lowest mean temperature was documented on November 19, 2018, while the highest mean temperature recorded was May 7, 2020. Relative humidity, which is expressed in percentage, has the highest (94%) and lowest (59%) logged data on September 11, 2019 and October 26, 2020, respectively. The slowest wind speed recorded is 1 meter per second, while the fastest wind speed is 1 meter per second.

2.4. Time series models

Every climatic data parameter is split into 80% training data and 20% testing data without randomization, which is the nature of time series data [27]. Prophet's automatic parameter tuning simplifies its use, particularly for individuals lacking advanced expertise in time series modeling and parameter selection. The combination of Prophet with XGBoost has gained popularity in time series forecasting. With automated parameter tuning, implementing Prophet becomes swift, diminishing the time and effort needed for model setup and configuration. Prophet exhibits versatility in handling various seasonal patterns, encompassing multiple and non-periodic seasonality, allowing it to capture intricate seasonality inherent in the data. Its resilience to missing data and outliers makes Prophet adept at managing time series datasets with irregular patterns or incomplete values. Furthermore, Prophet furnishes interpretable components such as trend, seasonality, and holiday effects, facilitating an understanding of the factors influencing the forecasted values [28], [29].

The concept behind merging ARIMA with XGBoost is to utilize ARIMA for short-term predictions and leverage XGBoost to capture more extended trends and patterns that ARIMA might overlook. ARIMA serves to forecast the immediate future steps, while the residuals (the disparities between actual and predicted values) from the ARIMA model are employed as features in an XGBoost model. Subsequently, the XGBoost model is trained to predict the remaining variability in the data that the ARIMA model might have missed. This hybrid strategy harnesses the strengths of both methods, leading to a more resilient and accurate time series forecasting model. It's crucial to recognize that the efficacy of such a combination relies on the specific characteristics of the analyzed time series data [30], [31].

A TSLM represents a statistical approach employed in the analysis and prediction of time series data. In the realm of time series analysis, this model presupposes that the connection between variables can be articulated through a linear equation. Specifically tailored for sequential data points where the order and timing are significant, the TSLM posits that the association between the dependent and independent variables is linear and remains consistent over time. Linear models find frequent application in time series analysis when the data manifests a relatively steady linear trend or when a straight line can adequately approximate relationships between variables. However, it is essential to acknowledge that real-world time series data often manifests more intricate patterns. In such instances, employing more advanced models like ARIMA, SARIMA, or machine learning methodologies such as XGBoost may be more fitting [32], [33]

2.5. Metrics performance evaluation

The forecasting errors align closely with the scale of the data. Consequently, accuracy measurements relying solely on scale are scale-dependent and unsuitable for comparing series employing different units. The two scale-dependent measurements that are most frequently utilized are based on absolute or squared errors. The MAE is a popular forecasting method due to its simplicity in interpretation and computation, mainly when applied to a single time series or multiple time series with identical units. A method minimizing the MAE yields forecasts of the median, while a method minimizing the RMSE yields forecasts of the mean. Despite being somewhat more challenging to interpret, the RMSE is also widely used [34]. RMSE is a method for determining the gap between values predicted by the machine learning model and actual values. Briefly defined, RMSE is the square root of variation or standard deviation [35].

A scaled error is below one when it results from a superior forecast compared to the average one-step naive forecast computed in the sample. Conversely, it exceeds one if the forecast is inferior to the average one-step naive forecast computed in the sample [36]. MAPE is a metric used to evaluate the accuracy of a forecasting or prediction model, particularly in the context of time series analysis. It measures the average percentage difference between predicted and actual values, considering the relative magnitude of the errors [37]. SMAPE calculates the percentage difference between predicted and actual values in a symmetric manner, meaning it considers the magnitude of the values being compared [38]. R squared is a statistical metric that shows how much of the variation for a dependent variable in a regression model is explained by one or more independent variables [39].

3. RESULTS AND DISCUSSION

A machine learning framework was used to compare the performance of three model algorithms on the data set without tuning the hyperparameters, as depicted in Figure 2. Rainfall, mean temperature, relative humidity, and wind speed were forecasted following the framework using Prophet boost, ARIMA boost, and LTSM forecasting algorithms. The entire data analysis process was performed using the R programming language through the RStudio IDE. The tidyverse package extracts each weather parameter with the date variable to create four individual data sets containing only two variables. All plots during the analysis were generated using the timetk package.

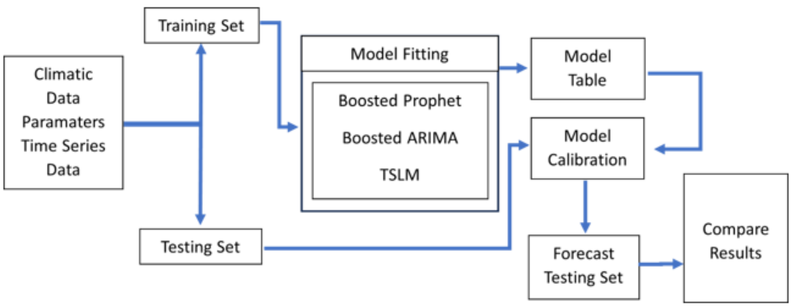

Figure 2. Framework for comparing time series model performance

3.1. Training and testing sets

Each data set is split using the tidymodels package with 80% training and 20% testing. Since the data is in time series format, splitting is not done at random, in contrast with the traditional machine learning process. Instead, the splitting is done chronologically, with the testing set making up 20% of the latest data recorded in the data set. The training set is later used for time series modeling, while the testing set is reserved for performance evaluations. Figures 3, 4, 5, and 6 depict the training and testing splits of the four weather parameters. The blue visual image corresponds to the training data, while the red visual image corresponds to the testing data.

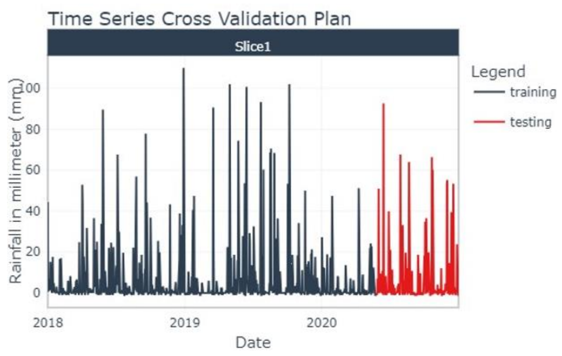

Figure 3. Rainfall training-testing split

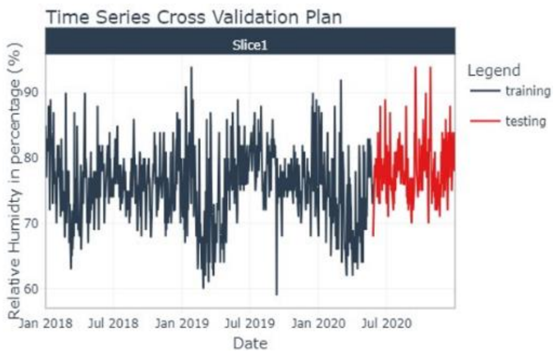

Figure 4. Relative humidity training-testing split

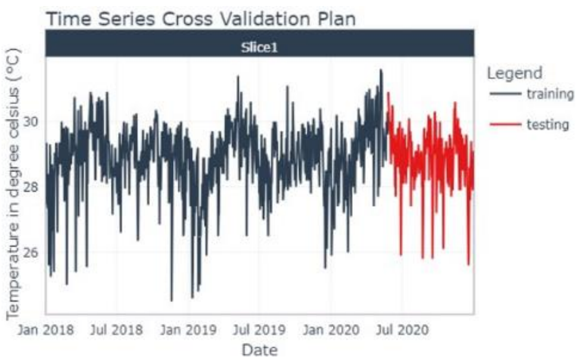

Figure 5. Mean temperature training-testing split

Figure 6. Wind speed training-testing split

3.2. Model fitting

Model fitting or training performed on the four climatic data parameter sets used the tidymodels package to leverage the three algorithms and their respective computational engines. It is the process of training a predictive model using a dataset. The goal is to create a model that can make accurate predictions on new, unseen data based on the patterns and relationships it learns from the training data. The process involves selecting a specific type of model, configuring its parameters, and optimizing its performance. The Prophet boost algorithm used the Prophet XGBoost, the ARIMA boost used the auto ARIMA XGBoost, and the LSTM used the linear model machine learning as the computational engine.

3.3. Model table, calibration, and forecast testing test

The modeltime package works alongside the tidymodels package and provides a workflow for time series modeling. It includes creating a table to wrap the three fitted models together, calibrating the testing set with the table to determine the confidence interval and metrics, and forecasting the testing set to evaluate the models. It introduces the concept of model tables, streamlining the comparison of different models, and incorporates features like calibration and time series cross-validation to enhance the accuracy of time series models. Forecasting the testing set evaluates the models' performance on the out-of-sample set, considering no hyperparameter tuning in the analysis of data. Figures 7, 8, 9, and 10 compare the three models' forecasted values on the test set's actual value for the four climatic data parameters. The actual data and the forecasted data using the models, as shown in the legend, depict various results. These images were generated in RStudio IDE after incorporating the Prophet boost algorithm using the Prophet XGBoost, the ARIMA boost using the auto ARIMA XGBoost, and the LSTM used the linear model machine learning as the computational engine.

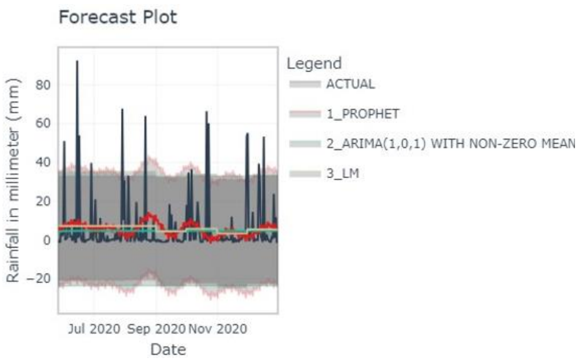

Figure 7. Rainfall forecasted values

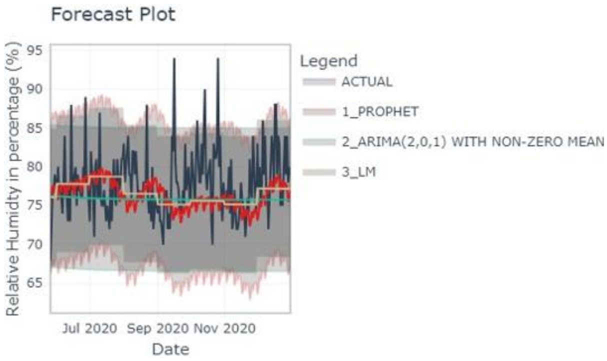

Figure 8. Relative humidity forecasted values

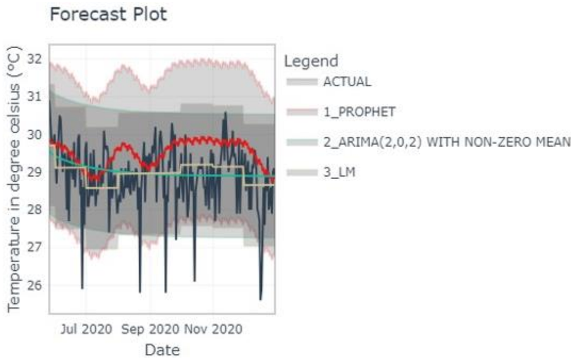

Figure 9. Mean temperature forecasted values

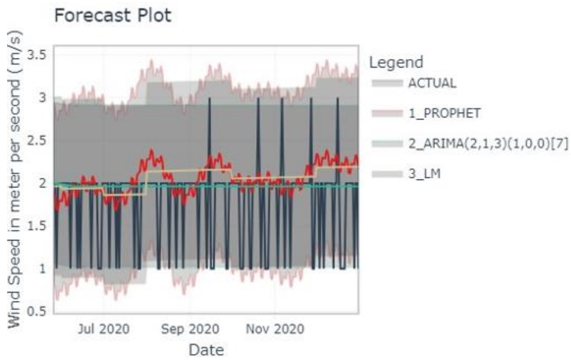

Figure 10. Wind speed forecasted values

3.4. Metric comparative results

Comparative results for each climatic parameter and forecasting algorithm are based on six metrics, namely, MAE, MAPE, MASE, SMAPE, RMSE, and RSQ, as shown in Tables 2, 3, 4, and 5. The different metrics highlight the different aspects of each model's performance. The values were generated using the code for the six metrics applicable to the three forecasting models in the RStudio IDE. The choice of metrics depends on the data's characteristics and the nature of the problem, which is not included in this study's scope. Also, the red values on these tables correspond to the lowest metric values, which makes for an easier comparison of the forecasting models. Please do note that in RSQ metric, the higher the value, the better the performance.

Table 2 summarizes the rainfall model results utilizing the six metrics. In interpreting these values, the smaller the value of the metric, the better the forecasting result compared to a higher value, except for RSQ, which is the opposite. The MAPE metric yielded an infinity (inf) value for the rainfall parameter due to division by zero and taking the logarithm of zero. In light of this issue's frequent occurrence, the authors are contemplating the exclusion of this metric for the rainfall parameter. The practical model for rainfall parameters is TSLM, as it mainly produced the lowest values (highest for RSQ). Although MAE and MASE rank second, the metric values of TSLM are closely the same as those from the ARIMA model. The relative humidity model results are depicted in Table 3, where TSLM emerged as the practical model. The same interpretation is also for the mean temperature parameter see in Table 4, where TSLM produced most of the lowest metric values, specifically for the MAE, MAPE, MASE, SMAPE, and RMSE metrics. Lastly, the wind speed climatic parameter has ARIMA as the interpretation for the practical forecasting model. This model generated the lowest MAE, MAPE, MASE, SMAPE, and RMSE metric values.

Table 2. Rainfall model results

Model	Metrics					
	MAE	MAPE	MASE	SMAPE	RMSE	RSQ
Prophet	8.78	inf	0.86	154.65	14.90	0.002065
ARIMA	7.90	inf	0.78	150.92	14.48	0.001142
TSLM	8.41	inf	0.83	150.82	14.37	0.009583

Table 3. Relative humidity model results

Model	Metrics					
	MAE	MAPE	MASE	SMAPE	RMSE	RSQ
Prophet	3.59	4.50	0.96	4.60	4.83	0.001443
ARIMA	3.43	4.26	0.92	4.38	4.69	0.009103
TSLM	3.30	4.12	0.89	4.21	4.51	0.004019

Table 4. Mean temperature model results

Model	Metrics					
	MAE	MAPE	MASE	SMAPE	RMSE	RSQ
Prophet	0.81	2.85	1.12	2.79	1.06	0.01928
ARIMA	0.61	2.15	0.85	2.12	0.83	0.04797
TSLM	0.60	2.11	0.84	2.09	0.83	0.05656

Table 5. Wind speed model results

Model	Metrics					
	MAE	MAPE	MASE	SMAPE	RMSE	RSQ
Prophet	0.36	29.34	1.05	21.24	0.54	3.08 e-04
ARIMA	0.27	23.61	0.78	16.82	0.48	6.44 e-04
TSLM	0.34	28.65	1.00	20.42	0.53	2.57 e-05

4. CONCLUSION

The primary purpose of the analysis was to compare the Prophet boost, ARIMA boost, and TSLM forecasting models' performance on the PAGASA weather data for Davao City using MAE, MAPE, MASE, SMAPE, RMSE, and RSQ evaluation metrics. All these metrics except RSQ indicate better performance for smaller values. The best metric to focus on may depend on the specific requirement of the forecasting task. It is worth noting that no hyperparameters were tuned during the analysis. The descriptions of the evaluation metrics applied to the analysis are as follows. MAE measured the average absolute errors between the predicted and actual values. MAPE measured errors as a percentage of the actual values. MASE compared performance against a naïve baseline or a simple moving average. SMAPE is a symmetrical MASE version but less sensitive to extreme values. RMSE penalizes more significant errors more heavily than more minor errors. Lastly, RSQ indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. Based on the results of rainfall data, an intermittent time series is present, and the MAPE metric is not appropriate. The ARIMA boost outperforms the Prophet boost and TSLM in MAE and MASE, while the same is valid with TSLM in the SMAPE metric. In RMSE, the ARIMA boost is second to TSLM with only 0.01 difference. The TSLM outperforms the other two in RSQ. However, the choice between the three forecasting models may not bring a significant difference due to the slight differences in metric values, with only the MAE metric having a range of 0.88, which is closest to 1. On the relative humidity data, the TSLM model outperforms the other two models in all metrics except RSQ, which is second to the ARIMA boost with a slight 0.005 difference. However, all three models are close in all metrics, with a range of 0.29 for MAE, 0.38 for MAPE, 0.079 for MASE, 0.39 for SMAPE, 0.33 for RMSE, and 0.00766 for RSQ. The choice of forecasting model on this parameter may not necessarily bring a significant difference as the metric values are relatively close to each other. The mean temperature data is expected to have similar trend results to the relative humidity, given that temperature and humidity are correlated. However, the TSLM outperforms the other two in all metrics. Nevertheless, the results do not have significant differences. Similarly, the choice of model for this parameter may also not bring a significant difference. The wind speed had the most varying results among the four weather parameters. Although the ARIMA boost outperforms the other two in all metrics, there is a significant range difference in MAPE and SMAPE metrics. The range obtained for the MAPE is 5.7, while SMAPE obtained 4.4. Due to the significant differences in the two metrics, the more appropriate model for this parameter is the ARIMA boost.

The choice of the forecasting model varies from parameter to parameter. The ARIMA boost can be a good choice for wind speed. The TSLM can be a good choice for relative humidity and mean temperature. Both the ARIMA boost and TSLM are good choices for rainfall relevant to the metric used. In almost all cases, the Prophet boost is the worst performing model except for RSQ in rainfall and wind speed data, where it places second best or worst. Nevertheless, the Prophet boost can be eliminated, leaving only ARIMA boost and TSLM. The results are beneficial in time series forecasting of the four climatic data parameters, especially in Davao City, Philippines. For further studies, it is recommended to embed other time series forecasting models and tuning of hyperparameters for better comparison and have an optimized version of forecasting. Also, if data is available, other climatic parameters may be used for forecasting.

REFERENCES

[1] J. W. Hansen, “Realizing the potential benefits of climate prediction to agriculture: issues, approaches, challenges,” *Agricultural Systems*, vol. 74, no. 3, pp. 309–330, Dec. 2002, doi: 10.1016/S0308-521X(02)00043-4.

[2] C. Predo, J. Crean, P. Hayman, J. Mullen, “Profitable use of seasonal climate forecasts: farm management case studies in the Philippines and Australia,” *Philippine Journal of Crop Science (Philippines)*, 2007, [Online]. Available: https://www.researchgate.net/profile/Kevin_Parton/publication/265192178_Profitable_Use_of_Seasonal_Climate_Forecasts_Farm_Management_Case_Studies_in_the_Philippines_and_Australia/links/549098540cf225bf66a97c67.pdf

[3] C. M. Reyes, S. N. Domingo, A. Agbon, M. D. Olaguera, A. J. L. Umlas, K. M. C. Zuluaga, “Barriers to application of weather and climate information in smallholder vegetable farming in Benguet,” Philippine Institute for Development Studies, 2020.

[4] D. Faustino-Eslava, C. Dimalanta, G. Yumul, N. Servando, and N. Cruz, “Geohazards, tropical cyclones and disaster risk management in the philippines: adaptation in a changing climate regime,” *Journal of Environmental Science and Management*, vol. 16, no. 1, pp. 84–97, Jun. 2013, doi: 10.47125/jesam/2013_1/10.

[5] T. A. Cinco *et al.*, “Observed trends and impacts of tropical cyclones in the Philippines,” *International Journal of Climatology*, vol. 36, no. 14, pp. 4638–4650, Nov. 2016, doi: 10.1002/joc.4659.

[6] C. A. Jamandre and G. T. Narisma, “Spatio-temporal validation of satellite-based rainfall estimates in the Philippines,” *Atmospheric Research*, vol. 122, pp. 599–608, Mar. 2013, doi: 10.1016/j.atmosres.2012.06.024.

[7] F. Gallo *et al.*, “High-resolution regional climate model projections of future tropical cyclone activity in the Philippines,” *International Journal of Climatology*, vol. 39, no. 3, pp. 1181–1194, Mar. 2019, doi: 10.1002/joc.5870.

[8] F. A. Galleto, A. D. M. Africa, F. G. De Leon, and A. B. Culaba, “Fifth-generation networks: the adoption of technology in a developing country,” *SciEnggJ*, vol. 16, no. 2, pp. 310–320, 2023.

[9] R. M. Adams *et al.*, “Value of improved long-range weather information,” *Contemporary Economic Policy*, vol. 13, no. 3, pp. 10–19, Jul. 1995, doi: 10.1111/j.1465-7287.1995.tb00720.x.

[10] J. W. Mjelde, S. T. Sonka, D. S. Peel, “The socioeconomic value of climate and weather forecasting: a review,” *ISWS Miscellaneous Publication MP-112*, 1989.

[11] W. Huijun, “The weakening of the Asian monsoon circulation after the end of 1970’s,” *Advances in Atmospheric Sciences*, vol. 18, no. 3, pp. 376–386, May 2001, doi: 10.1007/BF02919316.

[12] F. J. Meza, J. W. Hansen, and D. Osgood, “Economic value of seasonal climate forecasts for agriculture: review of ex-ante assessments and recommendations for future research,” *Journal of Applied Meteorology and Climatology*, vol. 47, no. 5, pp. 1269–1286, May 2008, doi: 10.1175/2007JAMC1540.1.

[13] J. W. Jones, J. W. Hansen, F. S. Royce, and C. D. Messina, “Potential benefits of climate forecasting to agriculture,” *Agriculture, Ecosystems & Environment*, vol. 82, no. 1–3, pp. 169–184, Dec. 2000, doi: 10.1016/S0167-8809(00)00225-5.

[14] F. Hilario, R. De Guzman, D. Ortega, P. Hayman, and B. Alexander, “El Niño Southern Oscillation in the Philippines: impacts, forecasts, and risk management.,” *Philippine Journal of Development*, vol. 36, no. 1, pp. 9–34, 2009, [Online]. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:El+Ni?o+Southern+Oscillation+in+the+Philippines+:+Impacts+,+Forecasts+,+and+Risk+Management#0

[15] J. F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, and A. Troncoso, “Deep learning for time series forecasting: a survey,” *Big Data*, vol. 9, no. 1, pp. 3–21, Feb. 2021, doi: 10.1089/big.2020.0159.

[16] P. Chronis, G. Giannopoulos, and S. Athanasiou, “Open issues and challenges on time series forecasting for water consumption,” in *CEUR Workshop Proceedings*, 2016.

[17] C. M. Reyes, S. N. Domingo, C. D. Mina, and K. G. Gonzales, “Climate variability, seasonal climate forecast, and corn farming in Isabela, Philippines: A farm and household level analysis.” pp. 1–65, 2009.

[18] M. Radeny *et al.*, “Indigenous knowledge for seasonal weather and climate forecasting across East Africa,” *Climatic Change*, vol. 156, no. 4, pp. 509–526, Oct. 2019, doi: 10.1007/s10584-019-02476-9.

[19] M. Q. Villafuerte, J. C. R. Lambrento, C. M. S. Ison, A. A. S. Vicente, R. G. de Guzman, and E. L. Juanillo, “Climdatph: an online platform for Philippine climate data acquisition,” *Philippine Journal of Science*, vol. 150, no. 1, pp. 53–66, 2021, doi: 10.56899/150.01.05.

[20] S. Liverani, *R: A programming environment for data analysis and graphics*. Computer Science, Mathematics, 2008. [Online]. Available: https://api.semanticscholar.org/CorpusID:61979037

[21] M. Fernández Lizana, “Advantages of R as a tool for data analysis and visualization in social sciences,” *Revista Científica de la UCSA*, vol. 7, no. 2, pp. 97–111, Aug. 2020, doi: 10.18004/ucsa/2409-8752/2020.007.02.097.

[22] J. S. Racine, “RSTUDIO: a platform-independent IDE for R and sweave,” *Journal of Applied Econometrics*, vol. 27, no. 1, pp. 167–172, 2012, doi: 10.1002/jae.1278.

[23] J. Vernazi, “Getting started with RStudio,” *Journal of Chemical Information and Modeling*, vol. 53, no. 9, p. 92, 2011, [Online]. Available: https://api.semanticscholar.org/CorpusID:60539681

[24] X. Shen, H. Yan, C. Wang, P. Gao, C. H. Johnson, and M. P. Snyder, “TidyMass an object-oriented reproducible analysis framework for LC–MS data,” *Nature Communications*, vol. 13, no. 1, p. 4365, Jul. 2022, doi: 10.1038/s41467-022-32155-w.

[25] F. Bischoff and P. P. Rodrigues, “tsmp: An R package for time series with matrix profile,” *R Journal*, vol. 12, no. 1, pp. 76–86, 2020, doi: 10.32614/rj-2020-021.

[26] F. X. Diebold, “Structural time series analysis and modelling package: a review,” *Journal of Applied Econometrics*, vol. 4, no. 2, pp. 195–204, Apr. 1989, doi: 10.1002/jae.3950040207.

[27] T. Schreiber, “Constrained randomization of time series data,” *Physical Review Letters*, vol. 80, no. 10, pp. 2105–2108, Mar. 1998, doi: 10.1103/PhysRevLett.80.2105.

[28] R. Gupta, A. K. Yadav, S. Jha, and P. K. Pathak, “Time series forecasting of solar power generation using Facebook prophet and XGBoost,” in *2022 IEEE Delhi Section Conference (DELCON)*, IEEE, Feb. 2022, pp. 1–5. doi: 10.1109/DELCON54057.2022.9752916.

[29] C. Junior, P. Gusmão, J. Moreira, and A. M. M. Tome, “Time series forecasting in retail sales using LSTM and prophet,” in *Handbook of Research on Applied Data Science and Artificial Intelligence in Business and Industry*, 2021, pp. 241–262. doi: 10.4018/978-1-7998-6985-6.ch011.

[30] M. Alim, G.-H. Ye, P. Guan, D.-S. Huang, B.-S. Zhou, and W. Wu, “Comparison of ARIMA model and XGBoost model for prediction of human brucellosis in mainland China: a time-series study,” *BMJ Open*, vol. 10, no. 12, p. e039676, Dec. 2020, doi: 10.1136/bmjopen-2020-039676.

[31] M. S. Rahman, A. H. Chowdhury, and M. Amrin, "Accuracy comparison of ARIMA and XGBoost forecasting models in predicting the incidence of COVID-19 in Bangladesh," *PLOS Global Public Health*, vol. 2, no. 5, p. e0000495, May 2022, doi: 10.1371/journal.pgph.0000495.
[32] A. Alsharef, K. Aggarwal, Sonia, M. Kumar, and A. Mishra, "Review of ML and autoML solutions to forecast time-series data," *Archives of Computational Methods in Engineering*, vol. 29, no. 7, pp. 5297–5311, Nov. 2022, doi: 10.1007/s11831-022-09765-0.
[33] R. P. Masini, M. C. Medeiros, and E. F. Mendes, "Machine learning advances for time series forecasting," *Journal of Economic Surveys*, vol. 37, no. 1, pp. 76–111, Feb. 2023, doi: 10.1111/joes.12429.
[34] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014, doi: 10.5194/gmd-7-1247-2014.
[35] N. Klaas and M. Faber, "Estimating the uncertainty in estimates of root mean square error of prediction: application to determining the size of an adequate test set in multivariate calibration," *Chemometrics and Intelligent Laboratory Systems*, vol. 49, no. 1, pp. 79–89, Sep. 1999, doi: 10.1016/S0169-7439(99)00027-1.
[36] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, Oct. 2006, doi: 10.1016/j.ijforecast.2006.03.001.
[37] A. de Myttenaere, B. Golden, B. Le Grand, and F. Rossi, "Mean absolute percentage error for regression models," *Neurocomputing*, vol. 192, pp. 38–48, Jun. 2016, doi: 10.1016/j.neucom.2015.12.114.
[38] P. Goodwin and R. Lawton, "On the asymmetry of the symmetric MAPE," *International Journal of Forecasting*, vol. 15, no. 4, pp. 405–408, Oct. 1999, doi: 10.1016/S0169-2070(99)00007-2.
[39] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE, and RMSE in regression analysis evaluation," *PeerJ Computer Science*, vol. 7, p. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.

BIOGRAPHIES OF AUTHORS

Jamal Kay B. Rogers is an Assistant Professor at the College of Information and Computing (CIC) of the University of Southeastern Philippines (USeP) in Davao City, Philippines. Mr. Rogers is registered electronics engineer with a Bachelor's in Electronics and Communication Engineering and a Master's in Engineering, majoring in Electronics and Communication Engineering. He is taking a Doctor of Philosophy in Education major in Information Technology Integration. His research interests are the internet of things (IoT), data science, machine learning, and educational data mining (EDM). He can be contacted at email: jamalkay.rogers@usep.edu.ph.

Tamara Cher R. Mercado is a professor at the College of Information and Computing (CIC) of the University of Southeastern Philippines (USeP) in Davao City, Philippines. Dr. Mercado holds degrees in Doctor of Philosophy in Information Technology, Master of Science in Information Science, and Bachelor of Science in Computer Engineering. Her research interests include information systems, knowledge management, and offline web services. She can be contacted at email: tammy@usep.edu.ph.

Fredelino A. Galleto Jr. is licensed electronics engineer in the Philippines. Currently, he is taking up Doctor of Philosophy in Electronics and Communications Engineering at De La Salle University, Manila. He has been a faculty of the Department of Electronics Engineering at the University of Southern Mindanao, Kabacan, Cotabato, Philippines, since October 2018. His current research interest involves data optimization, antenna and communications systems, and information and communications technology. He can be contacted at email: fgalletojr@usm.edu.ph.