Predicting Dengue Outbreaks in Cagayan de Oro, Philippines using Facebook Prophets and the ARIMA Model for Time Series Forecasting

.

ABSTRACT

In terms of public health, dengue fever continues to be an important issue. In recent years, there has been a sharp rise in dengue-related cases all throughout the world. The number of reported cases was consistently rising, especially in the cities of the Philippines. Time series analysis is used in this study to forecast the frequency of dengue by modeling the weekly incidence of cases of dengue hemorrhagic fever (DHF) in Cagayan de Oro City, Philippines. This study uses a dataset of confirmed dengue cases that were downloaded from the Department of Health's (DOH) official website to compare the effectiveness and precision of Facebook Prophet's and ARIMA's forecasting models. The performance indicators of the Facebook Prophet and ARIMA approaches are contrasted on the same dataset to choose the best accurate forecast model. The period covered by the dataset chosen for this study runs from the first week of 2016 to the sixteenth week of 2021. The performance of the forecast models is then evaluated by comparing them to the last 66 weeks of actual data. The result of this study shows that Facebook Prophet model outperforms ARIMA model.

Keywords: Dengue Hemorrhagic Fever (DHF), Forecasting, ARIMA Model, Facebook Prophet Model

1. INTRODUCTION

Over a hundred tropical and subtropical nations throughout the world have endemic dengue infections. One of the four dengue viruses, DENV-1, DENV-2, DENV-3, and DENV-4, is conveyed by the primary vectors Aedes aegypti and Aedes albopictus, has been identified [1]. The illness was referred to in the Philippines as hemorrhagic fever or infectious acute thrombocytopenic purpura in the early 1950s [2]. Dengue cases increase globally from year to year. According to data generated by the World Health Organization (WHO), there were 4.2 million cases worldwide in 2019 compared to 504,430 cases in 2000. It demonstrates an eight-fold increase in dengue fever cases during the previous 20 years. According to the overall number of Dengue Hemorrhagic cases worldwide up until this point, Asia has the highest number of instances [3].

In the Philippines, an overall total of 309 dengue cases and one death were reported during epidemiological week 35 of 2022. When compared to the same period in 2021 (n=1,601), there are 81% fewer cases. However, given that cases are still being verified and authenticated, reported cases for week 35 of 2022 may change. A total of 145,650 cases and 462 deaths (CFR 0.3%) were reported from 1 January to 3 September 2022(week 35), which is an increase of 168% over the 54,298 cases reported during the same period in 2021 [4]. Dengue fever illnesses in Cagayan de Oro increased by 170% throughout a nearly eight-month period in 2021, continuing a trend that has been present this year. The province of Bukidnon, which has reported more than 2,400 cases since January 2022—a 335.9% increase compared to cases reported during the same period in 2021—ranked first in terms of the number of dengue cases in Northern Mindanao, ahead of the city. Cagayan de Oro now has the most dengue infections among Northern Mindanao cities, with 1,028 cases since January. The number could rise further, according to the Cagayan de Oro City Health Office, because of the regular rains that provide dengue-carrying Aedes mosquito breeding grounds [5].

The Department of Health (DOH), the country's official agency for health, is responsible for dengue prevention and control. To safeguard Filipinos from the threat of the disease, the DOH dengue program carries out ongoing surveillance, case management and diagnosis, integrated vector management, outbreak response, health promotion and advocacy, and research. Through the Philippine Integrated Disease Surveillance and Response (PIDSR), which is supported by the DOH Regional Offices and the Department of Entomology at the Research Institute for Tropical Medicine, outbreaks are monitored and reported [6].

In any case, there is currently no medication that can be used to treat Dengue Hemorrhagic Fever (DHF), and treatments only address clinical side effects. Despite a small number of applicants' antibodies counting live constricted mono, tetravalent definition inactivated entirety infection immunizations, and recombinant subunit antibodies [7].

There has been a significant amount of study done on machine learning's expansion into forecasting. Time series forecasting is now used in a wide variety of industries, including energy, retail, transportation, and finance, to project product demand, resource allocation, financial performance, preventive maintenance, and a host of other uses. Time series forecasting can alter company models and boost profits, but many businesses have yet to adopt its technology and make use of its advantages. Time series forecasting is a method for predicting future events by examining historical trends, with the underlying premise being that prior trends will continue to hold for the foreseeable future. To anticipate future values, forecasting includes fitting models to historical data. Time series forecasting is necessary for prediction problems with a time component since it offers a data-driven method for effective and efficient planning [8].

The collection of such observations is known as time series data. A time series is a group of observations, each of which was recorded at a certain time. The data is processed to draw out statistical details, data features, and output predictions. Timeseries analysis and related methods have made prediction easier since time series data may tend to follow a pattern, which makes it harder for the Machine Learning model to forecast accurately [9].

Based on data already collected from the first week of 2016 through the sixteenth week of 2022, the researcher hopes to predict the time series of dengue fever cases in Cagayan de Oro City and show the patterns of the disease's time series for the next four weeks. To do this, the study employs two different methodologies with the potential to yield results in the future. They are the Facebook Prophet forecasting method and the Auto-Regressive Integrated Moving Average (ARIMA) model. Popular ARIMA models have been used in a variety of industries, including the banking stock market and tourism demand [10]. However, Facebook Prophet can be classified as a unique technique because it was only developed five years ago, is well known for its straightforward but efficient model, and has been used in several studies.

Several methods enable one to predict how the future will appear if certain presumptions are true. The Auto-Regressive Integrated Moving Average (ARIMA) model is one of these methods, which is more accurately referred to as a forecasting model. This approach has been in use for many years, making it a tried-and-true method that can generate incredibly accurate projections. In addition to ARIMA and some other well-known models, such as the

Exponential Smoothing, new competitors have entered the forecasting fray, largely as a result of the recent data science, machine learning, deep learning, etc. One of these new rivals is Facebook Prophet, a time series model developed by the blue giant Facebook and used by them in their forecasting processes. It has the benefit of accounting for substantial seasonal effects, missing data, outliers, and changes in trends and was made freely available to the public as one of their open-source initiatives.

In 2017, Facebook released and made publicly accessible the package of Facebook prophet, an open-source forecasting tool in Python and R. It is a time series prediction algorithm. When compared to other forecasting tools, Prophet has experienced tremendous growth. The Prophet package has been downloaded 21,819,762 times as of this writing. According to research, Prophet also comes in first place among all other Python time series packages in terms of monthly downloads [11]. Prophet requires the time series variables y (target) and ds (DateTime). Therefore, data from several seasons and time series with large seasonal impacts are best for use with the Prophet model. Additionally, the Prophet model has no restrictions on measurement spacing regularity and can extract trends and periodic signals across a wider range of time scales than conventional exponential smoothing techniques. As a result, the Prophet model makes several time series studies simpler [12]. Prophet's ability to automatically identify seasonal trends with a collection of data and provide simple-to-understand parameters makes it a big advantage over other time series forecasting tools. This implies that even a non-statistician can begin using it and produce effective results on par with the professionals.

2. MATERIALS AND METHODS

2.1 Dataset Description and Preparation

<u>2.1.1 Data Collection</u>

The official website of the Department of Health (DOH) provided weekly reports of dengue hemorrhagic incidents from January 2016 to April 2022. There are 328 rows and 2 columns in the dataset. The dengue case and the morbidity week are these data's two primary characteristics.

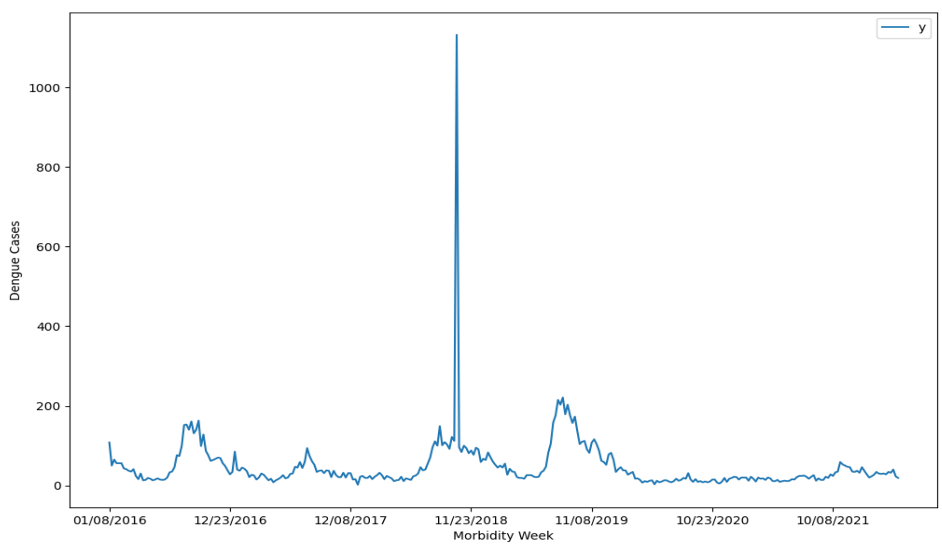

Figure 1. Cagayan de Oro City Weekly Dengue Cases

2.1.2 Data Analysis

The methodology for data analysis includes reading the .csv file, retrieving the attributes, computing the shape of the dataset, and visualizing the information of the dataset. The data for Cagayan de Oro City will be obtained, and to plot a graph displaying the number of new cases every week, as shown in Figure 1. In the current state of the world, the visualization and modeling of dengue cases are essential and very helpful.

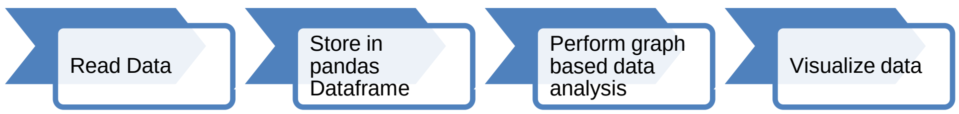

Figure 2. Model of the data analysis phase

The steps for data analysis are as follows:

Step1: The initial step in the data analysis phase is to load the essential libraries, such as NumPy, Pandas, Datetime, and matplot, for studying the dataset.
Step 2: These libraries aid programmers with data analysis. Import the dataset because it is in.csv format.
Step 3: The dataset is initially read, then it is then stored in a pandas data frame.
Step 4: The data is now plotted on the per-week statistics graph.

Details on how to implement the models are provided in Table 1.

Table 1. Implementation of the specified model

Data Used	Dengue Case from DOH official website (January 2016 to April 2022)
Models used	ARIMA and Prophet
Language	Python
System Software	Jupyter
Libraries	Matplotlib, Facebook Prophet, ARIMA, Math, DateTime

The models to be used are Facebook Prophet and ARIMA. The data frame in the Prophet framework makes working with time series and seasonality data simple. The data frame must have two essential columns. One of these columns, "ds," keeps track of the date-time series. The "y" column in the time series data frame contains the pertinent values [13].

Time series data can be analyzed and predicted using statistical models like ARIMA. It provides a straightforward yet effective method for producing accurate time series forecasts by expressly catering to several typical time series data types. It integrates the idea of integration and is a more advanced form of the fundamental autoregressive moving average [14].

2.2 ARIMA Model

This is a linear regression model and is known as the "autoregressive integrated moving average". For predicting time sequences, this is also helpful. To predict future values based on historical data, it uses statistical models. This model's notation is as follows: (p, d, q). The autoregression element of the parameter p shows that it helps with the inclusion of previous ideals into the model by incorporating the amount of difference in time series. Then, q is the running average part, this assists in determining the model's inaccuracy. This is provided by Prabhakaran, 2022 and Vincent, 2017 [15] [16],

$$ARIMA\ (p, d, q)(P, D, Q)\ m,$$

where p denotes the order of the AR term and the number of lags used as a predictor; d denotes the number of differences needed to make the time series data stationary; if $d = 0$, then stationary; q denotes the order of the MA term and the number of lagged predicted errors that are used in the ARIMA model; m denotes each season of the number of periods, and (P, D, Q) denotes the periodic section of the time (p, d, q).

AR (Autoregression): Using historical data from earlier time locations, the AR, MA, ARMA, and ARIMA models are utilized to forecast the observation at $(t + 1)$. However, it is crucial to guarantee that the time series stays stationary over historical information for the observation period. A sort of model called autoregression (A.R.) estimates the series' present or future values by calculating the regression of historical time series. The moving average (M.A.) model determines the present or future values in the series after computing the residuals or errors of the previous time series. The ARMA model combines the AR and MA models. This model takes residuals and the effects of prior lags into account when projecting future data points in the time series [17] .

Below is the mathematical formula for the ARIMA model [15]:

$$y_t = \alpha + \beta_1 y_1 + \beta_2 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t \qquad (2.1)$$

where y_t is the stationary time series, α is the intercept term, β_p are parameters of the autoregressive model, and ε_t are the residual time (t).

MA (Moving Average): The mathematical equation is provided by [15],

$$y_t = \alpha + \epsilon_1 + \varphi_1 \epsilon_{t-1} + \varphi_2 \epsilon_{t-2} \ldots + \varphi_q \epsilon_{t-q} \qquad (2.2)$$

where ϵ_t are errors.

The ARIMA model equation is generally represented as [15],

Predicted value y_t= constant value + linear consolidation lags of Y + linear consolidation of lagged forecast errors.

The ACF (autocorrelation function) graph and PACF (partial autocorrelation function) are used to approximate the ARIMA model parameter [18].

The equation yields ARIMA, often known as the Box-Jenkins model, which displays future value predictions in a linear form [19],

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \epsilon_t - \theta_0 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q} \qquad (2.3)$$

where y_t is the actual data point at a given time (t), φ and θ denote the coefficient model, p and q denote the autoregressive and moving average functions, and integer value ϵ_t denotes an error.

2.3 Facebook Prophet Model

The Prophet method is an additive regression model with four primary components: a piecewise linear logistic growth curve trend; a seasonal component that is modeled annually using the Fourier series; a seasonal component that is modeled weekly using dummy variables; and a user-provided list of significant holidays. It is written [11],

$$y_t = g(t) + s(t) + h(t) + \epsilon_t \qquad (2.4)$$

where, $g(t)$ represents the trend and the objective is to capture the general trend of the series, it provides two possible trend model: saturating growth model and piecewise linear model, which can be seen in equation (3.5) and (3.9) respectively, $s(t)$ is the seasonality component, $h(t)$ the holidays component. Note that holidays vary between years, countries, etc. and therefore the information needs to be explicitly provided to the model, error term ϵ_t stands for random fluctuations that cannot be explained by the model. As usual, it is assumed that ϵ_t follows a normal distribution $N\,(0, \sigma^2)$ with zero mean and unknown variance σ^2 that has to be derived from the data.

The logistic growth function is used to model the nonlinear saturation trend function and is mathematically written as:

$$g(t) = \frac{c(t)}{1+e^{-k(t-m)}} \qquad (2.5)$$

where $c(t)$ is a time-varying capacity that indicates the maximum number of data added per day, k stands for a fluctuating growth rate, and m is the offset parameter. Additionally, by including five changepoints in the model, the effects of the interventions on the trend of growth are explicitly assessed. Assuming S changepoints are defined over time, the following formula defines a vector of the rate of change adjustments:

$$\delta \in R^s$$

The rate of change at any time t corresponds to the base rate plus the rate change of modifications made up until that time and is expressed as follows where δ_j indicates the rate of change at the time s_j:

$$\delta_t = k + \sum_{j:t>s_j} \delta_j \qquad (2.6)$$

Creating a vector that can be represented as:

$$a_j(t) = \begin{cases} 1, if\ t\ \geq s_j \\ 0, otherwise \end{cases} \qquad (2.7)$$

Consequently, the rate of change at time t. Then by modifying k and the offset parameter m, the proper adjustment at the changepoints j can be written as:

$$\gamma_j = \left(s_j - m - \sum_{l<j} \gamma l\right)\left(1 - \frac{k+\sum_{l<j}\delta_l}{k+\sum_{l\leq j}\delta_l}\right) \qquad (2.8)$$

Now, a logistic model that is piecewise can be observed:

$$g(t) = \frac{C(t)}{1+e^{(-(k+a(t)^T\delta)))}} \qquad (2.9)$$

The expected capacity of the system at any given time is referred to in this model as $C(t)$. The value of the capacity $C(t)$ is fixed taking into account the city's existing growth pattern to analyze the city's future growth pattern because this problem deals with everyday cases.

Using the Fourier series, it is feasible to model the periodic impact of yearly seasonal variations. Consequently, a typical Fourier series that may be expressed as follows is related to an almost smooth seasonal effect:

$$s(t) = \sum_{n=1}^{N} a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \qquad (2.10)$$

Let P represent the predicted regular period for the time series (e.g., $P = 365.25$ for annual data or $P = 7$ for weekly data, when scaling a days-based time variable). However, the dataset is divided into 80:20 ratios with 80% of the data used to train the model and 20% used for testing and forecasting the data after minimizing variance and bias error.

While, $h(t)$ represents the effects of holidays. For every holiday i, D_i is the set of past and future dates of that holiday:

$$h(t) = [1(t \in D_1), \dots, 1(t \in D_L]k \quad (2.11)$$

t is throughout the holiday i, and assign k which represents the relevant predicted change.

The model's workflow diagram is provided below.

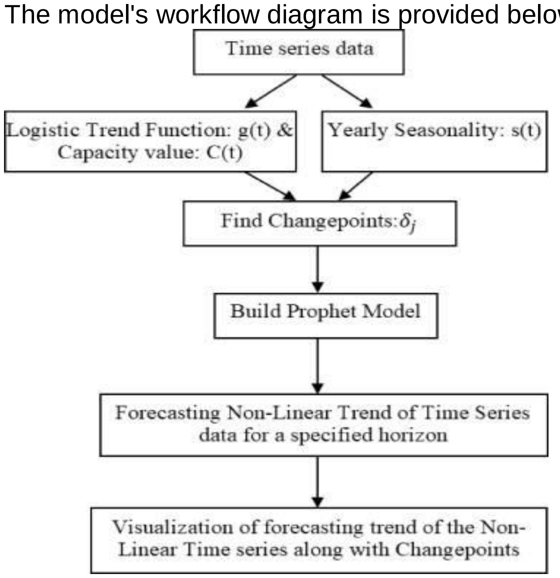

Figure 3. Prophet model workflow diagram

2.4 Performance Measures

The model-based forecast is evaluated to determine whether it will perform as predicted in terms of accuracy. Mean Absolute Percentage Error (MAPE) is typically the choice of the simplest model criterion that is widely used because of its value as a rate that makes it appropriate to be familiar with measuring the accuracy of a model. The statistical measure below is used to assess the prediction models. If a model's MAPE value is low, it is deemed to be the best model. Using the following equation, MAPE is used to calculate the percentage difference between the forecasted outcomes and the actual results.

Mean Absolute Percentage Error (MAPE):

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{X_t - \hat{X}_t}{X_t}\right| * 100\% \quad (2.12)$$

where n is the total number of observations, X_t is the actual value, and $\hat{X}_t$ is the predicted value. For evaluating forecasting, there is a MAPE standard category [20]. Table 2 shows these categories.

Table 2. Forecasting evaluation using the MAPE standard category

MAPE	Forecasting Criterion
<10%	Forecasting accuracy is very good
10-20%	Forecasting accuracy is good
20-50%	Forecasting accuracy is enough good
>50%	Forecasting accuracy is not good

3. RESULTS AND DISCUSSION

In this chapter, the use of the two forecasting models is presented. To evaluate the forecasting methods, MAPE (Mean Absolute Percentage Error) is employed as the performance metric. MAPE can be defined below:

Mean Absolute Percentage Error (MAPE):

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{X_t - \hat{X}_t}{X_t}\right| * 100\% \quad (3.1)$$

where n is the total number of observations, X_t is the actual value, and $\hat{X}_t$ is the predicted value.

3.1 Prediction Through ARIMA Model

The first step in time series forecasting is to check whether or not the time series data is stationary. Data must have a constant variance and mean in order to be considered stationary, in order for the pattern to remain the same across time. Data transformation may be required if it contains trend information. The de facto procedure in statistics to examine the stationarity of the data is the Augmented Dickey Fuller (ADF) Test. In order to assess whether or not our data is steady, ADF informs that the degree to which a null hypothesis can be rejected or not. The null hypothesis is either accepted or rejected based on the interpretation of this data using a threshold (0.05). The figure below illustrates the data's ADF results.

```
1. ADF:  -4.365773052281487
2. P-Value: 0.00034108581944242524
3. Num of Lags: 4
4. Num of Observations Used For ADF Regression and Critical Values Calculation: 323
5. Critical Values:
        1% : -3.4507587628808922
        5% : -2.870530068560499
        10% : -2.5715597727381647
```

Figure 4. ADF Tests of Dengue Cases in Cagayan de Oro City

From the findings, it can be concluded that:

• The p-value for the dengue cases data is lower than the threshold, which suggests that the null hypothesis must be rejected (the data is stationary).
• The test statistics for the data on dengue cases are close to or within the range of the critical values, demonstrating that the data are stationary.

As a result, the need to transform the data is no longer necessary. Choosing the optimal values for p, d, q, P, D, Q, is the next stage.

Next step is to forecast, Using auto_arima() function to get the best p, d, q, P, D, Q values. After splitting the data into test and training sets:

Dep. Variable:	DENGUE CASES	No. Observations:	262
Model:	ARIMA(1, 0, 1)	Log Likelihood	-1488.419
Date:	Thu, 16 Mar 2023	AIC	2984.838
Time:	20:37:11	BIC	2999.112
Sample:	01-08-2016	HQIC	2990.575
	- 01-08-2021		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
const	53.8969	46.594	1.157	0.247	-37.426	145.220
ar.L1	0.9080	0.072	12.635	0.000	0.767	1.049
ma.L1	-0.6899	0.084	-8.176	0.000	-0.855	-0.525
sigma2	5026.4797	129.008	38.963	0.000	4773.629	5279.330

Ljung-Box (L1) (Q):	0.19	Jarque-Bera (JB):	337570.80
Prob(Q):	0.66	Prob(JB):	0.00
Heteroskedasticity (H):	1.52	Skew:	11.95
Prob(H) (two-sided):	0.05	Kurtosis:	177.22

Figure 5. ARIMA Model best p,d,q values

The best ARIMA model was chosen as can be seen in the figure above by auto_arima() is ARIMA(1,0,1). The parameters are used to display the actual and predicted dengue cases in Cagayan de Oro City in the figure below. Using the best model, which is run by software, results a Mean Absolute Percentage Error (MAPE) value of 35.0%.

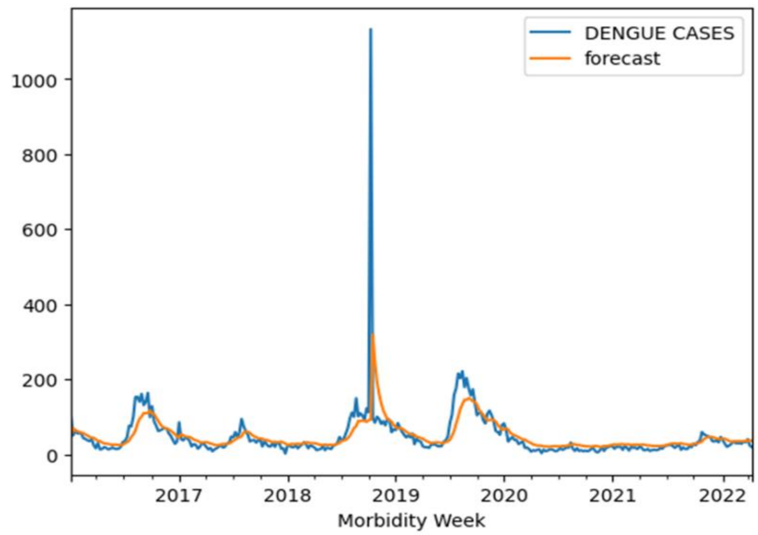

Figure 6. Actual vs Forecast for the dengue cases through ARIMA, where the x-axis is “dates” and the y-axis is “total cases”

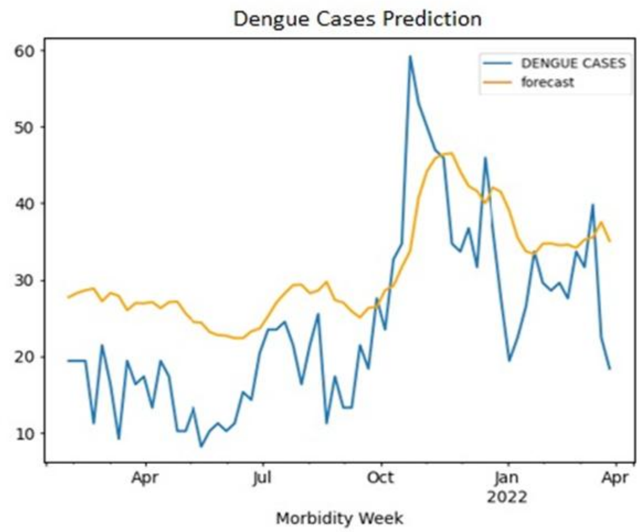

Figure 7. Actual and Predictive Values for the dengue cases through ARIMA of the Test Data, where the x-axis is “dates” and the y-axis is “total cases”

Table 3. Actual Dengue Cases Values Vs Predicted ARIMA

Morbidity Week	Actual	ARIMA RMSE 27.1%
01/15/2021	20	26.81
01/22/2021	20	27.41
01/29/2021	20	27.83
02/05/2021	12	28.12
02/12/2021	22	26.56
02/19/2021	17	27.68
02/26/2021	10	27.36
03/05/2021	20	25.59
03/12/2021	17	26.57
03/19/2021	18	26.59
03/26/2021	14	26.82
04/02/2021	20	26.10
04/09/2021	18	26.92
04/16/2021	11	27.05
04/23/2021	11	25.60
04/30/2021	14	24.60
05/07/2021	9	24.56
05/14/2021	11	23.43
05/21/2021	12	23.09
05/28/2021	11	23.08
06/04/2021	12	22.84
06/11/2021	16	22.90
06/18/2021	15	25.34
06/25/2021	21	27.56
07/02/2021	24	29.60
07/09/2021	24	31.45
07/16/2021	25	33.15
07/23/2021	22	34.70
07/30/2021	17	36.11
08/06/2021	22	37.40
08/13/2021	26	38.58
08/20/2021	12	39.66
08/27/2021	18	40.65
09/03/2021	14	41.55
09/10/2021	14	42.37
09/17/2021	22	43.12
09/24/2021	19	43.80
10/01/2021	28	44.43
10/08/2021	24	45.00
10/15/2021	33	45.52
10/22/2021	35	46.00
10/29/2021	59	46.44
11/05/2021	53	46.83
11/12/2021	50	47.20
11/19/2021	47	47.53
11/26/2021	46	47.83
12/03/2021	35	48.11
12/10/2021	34	48.36
12/17/2021	37	48.59
12/24/2021	32	48.81
12/31/2021	46	49.00
01/07/2022	37	49.17
01/14/2022	28	49.34
01/21/2022	20	49.48
01/28/2022	23	49.62
02/04/2022	27	49.74
02/11/2022	34	49.85
02/18/2022	30	49.95
02/25/2022	29	50.05
03/04/2022	30	50.13
03/11/2022	28	50.21
03/18/2022	34	50.28
03/25/2022	32	50.35
04/01/2022	40	50.41
04/08/2022	23	50.46
04/15/2022	19	50.46

3.2 Prediction Through Facebook Prophet Model

The Facebook prophet package in Python (created by Facebook) is used, and the workflow is described in Figure 3 to develop a Prophet model. To prepare the dataset, a new data frame with the column’s, "ds" (date stamp) and "y" (forecasting measurement) is created. The forecasting period is chosen as 66 weeks, and applied in forecasting. The Prophet's automatically set default parameters are applied. The Prophet Model's trend, seasonality, and other additional variables are employed to make the prediction. A MAPE value of 28.0% is generated by the procedure, which is performed out by software.

Table 4. MAPE Value for ARIMA and Facebook Prophet Model

Model	MAPE
ARIMA	35.0%
Facebook Prophet	28.0%

Hereafter, Prophet model will be used to forecasts the number of dengue cases in Cagayan de Oro City. Plots showing actual and predictive data and forecasting data for the next 66 weeks from January 2021 to April 2022 by the Prophet model can be seen in Figure 8. The x-axis is the date in the forecast visualization, and the y axis is the number of dengue cases. The black dots are the actual number of cases in the testing dataset, and the blue line is the time series model prediction. The shaded area is the 95% prediction interval.

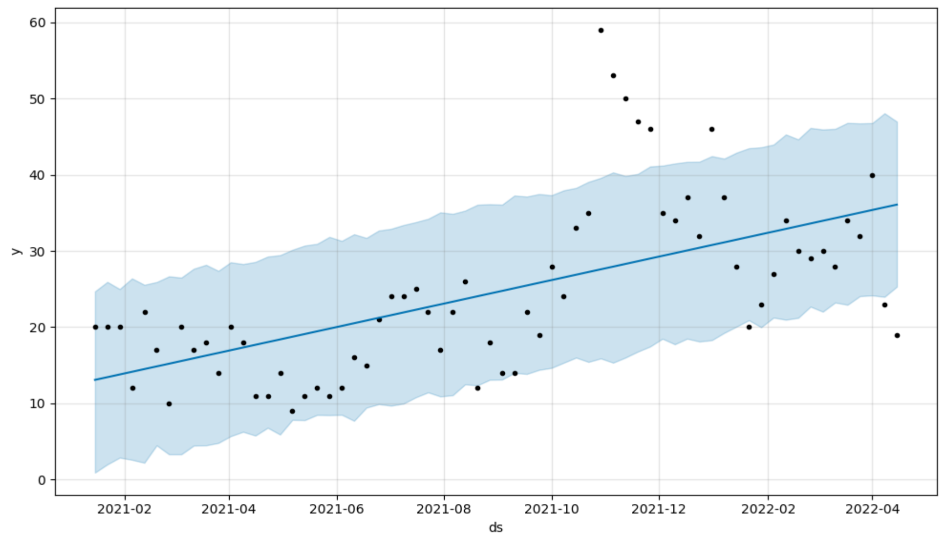

Figure 8. Predictions through Prophet for the number of dengue cases (for the next 66 weeks), where x-axis is “ds” and y-axis is “y” which is total number of cases

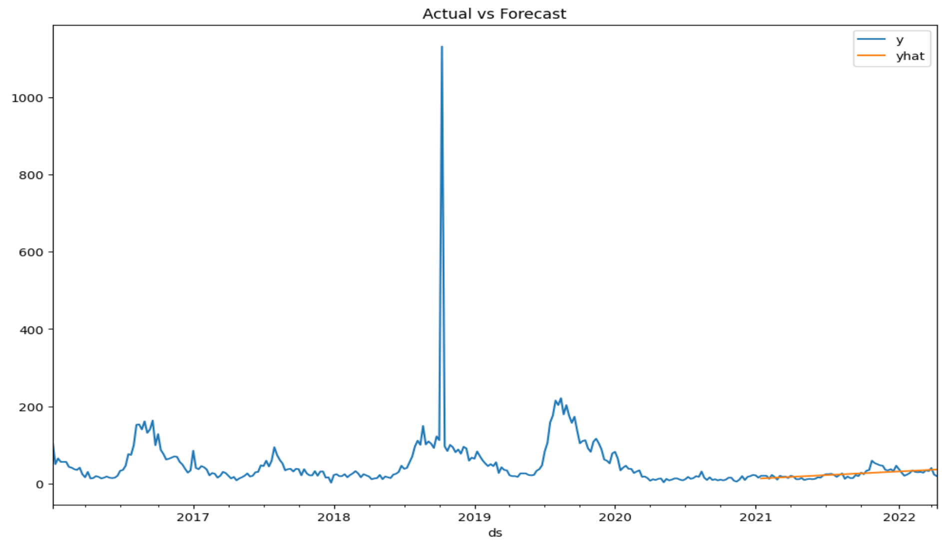

Figure 9. Actual vs Forecast for the dengue cases through Facebook Prophet, where the x-axis is “dates” and the y-axis is “total cases”

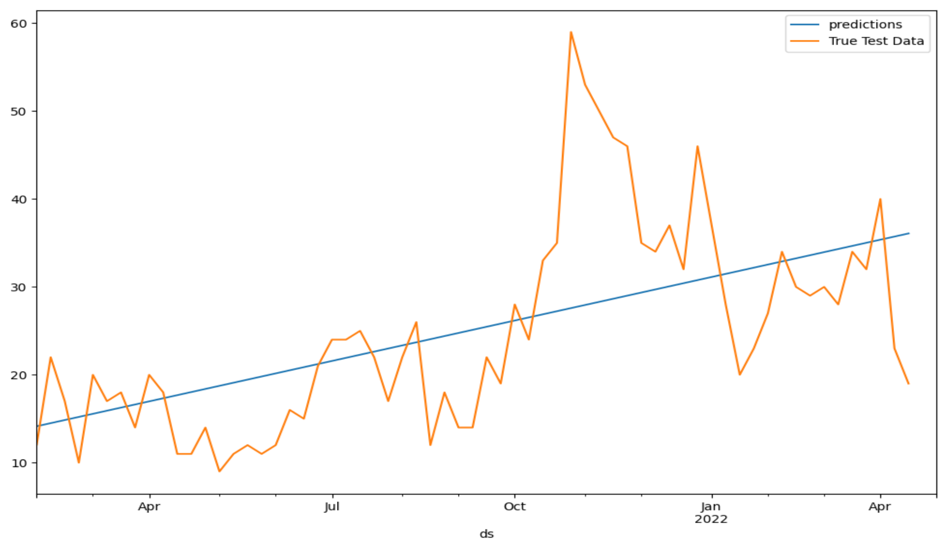

Figure 10. Actual and Predictive Values for the dengue cases through Facebook Prophet of the Test Data, where the x-axis is “dates” and the y-axis is “total cases”

Table 5. Actual Dengue Cases Values Vs Predicted Prophet

Morbidity Week	**Actual**	**FB Prophet** (**RMSE 9.01%**)
01/15/2021	20	13.08
01/22/2021	20	13.44
01/29/2021	20	13.69
02/05/2021	12	14.15
02/12/2021	22	14.50
02/19/2021	17	14.85
02/26/2021	10	15.21
03/05/2021	20	15.56
03/12/2021	17	15.91
03/19/2021	18	16.27
03/26/2021	14	16.62
04/02/2021	20	16.98
04/09/2021	18	17.33
04/16/2021	11	17.68
04/23/2021	11	18.04
04/30/2021	14	18.39
05/07/2021	9	18.74
05/14/2021	11	19.10
05/21/2021	12	19.45
05/28/2021	11	19.81
06/04/2021	12	20.16
06/11/2021	16	20.51
06/18/2021	15	20.87
06/25/2021	21	21.00
07/02/2021	24	21.57
07/09/2021	24	21.93
07/16/2021	25	22.28
07/23/2021	22	22.64
07/30/2021	17	22.99
08/06/2021	22	23.34
08/13/2021	26	23.70
08/20/2021	12	24.05
08/27/2021	18	24.40
09/03/2021	14	24.76
09/10/2021	14	25.11
09/17/2021	22	25.47
09/24/2021	19	25.82
10/01/2021	28	26.17
10/08/2021	24	26.53
10/15/2021	33	26.88
10/22/2021	35	27.23
10/29/2021	59	27.59
11/05/2021	53	27.94
11/12/2021	50	28.29
11/19/2021	47	28.65
11/26/2021	46	29.00
12/03/2021	35	29.36
12/10/2021	34	29.71
12/17/2021	37	30.06
12/24/2021	32	30.42
12/31/2021	46	30.77
01/07/2022	37	31.12
01/14/2022	28	31.48
01/21/2022	20	31.83
01/28/2022	23	32.19
02/04/2022	27	32.54
02/11/2022	34	32.89
02/18/2022	30	33.25
02/25/2022	29	33.60
03/04/2022	30	33.95
03/11/2022	28	34.31
03/18/2022	34	34.66
03/25/2022	32	35.02
04/01/2022	40	35.37
04/08/2022	23	35.72
04/15/2022	19	36.08

On the trend chart, an increasing trend can be seen from the beginning of 2021 to 2022.

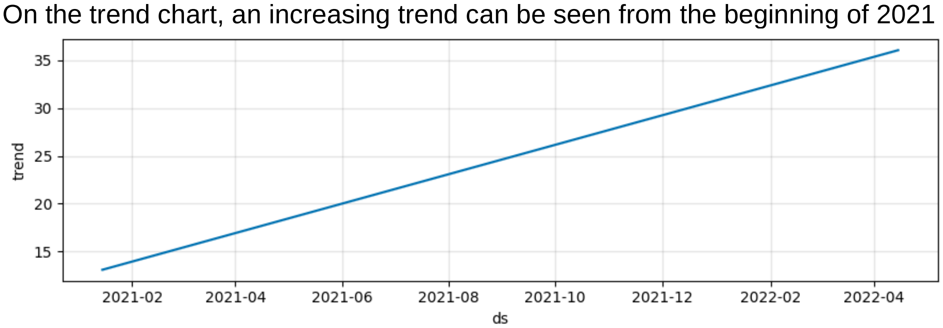

Figure 11. Time Series Decomposition

4. CONCLUSION

It can be inferred from the analysis performed to forecast the number of dengue cases in Cagayan de Oro City, Philippines that Facebook Prophet performs better than ARIMA. MAPE of 35.0% for the ARIMA Model and 28.0% for the Facebook Prophet Model. As a result, it may be concluded that both models have a sufficient level of accuracy because of its moderately low prediction error rate. The numbers of dengue cases predicted by the Facebook Prophet Model are compared to the actual value of the testing data shown in Table 5. This implies that the Facebook Prophet model will have a big impact on predicting the trend of dengue incidence over a specific time period. As a result, more research into the Facebook Prophet model is warranted. The study has ultimately come to the conclusion that the suggested strategy can help decision-makers better predict future dengue outbreaks and handle epidemics. Consequently, this will control resource use and decrease the effects of potential outbreaks.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the editor and referees for their corrections and improvement of this paper.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Chuang, T. W., Ng, K. C., Nguyen, T., & Chaves, L. Epidemiological Characteristics and Space-Time Analysis of the 2015 Dengue Outbreak in The Metropolitan Region of Tainan City, Taiwan. International Journal of Environmental Research and Public Health. 2018;15:396.
2. Juliano, L., Quintos, F., Reyes, A., Lim, L., & Lacson, P. Hemorrhagic fever observed among children in the Philippines. Herdinplus.
Available: https://www.herdin.ph/index.php?view=research&cid=25171#studydetails.

3. World Health Organization. The Work of WHO in the South-East Asia Region. 2011.
4. Dengue. World Health Organization, 10 July 2022. Available: https://www.who.int/westernpacific/emergencies/surveillance/dengue.
5. Lagsa, B. Cagayan de Oro dengue cases surge by 170%. 2022. Rappler. Available: https://www.rappler.com/nation/mindanao/dengue-cases-cagayan-de-oro-august-15-2022/.
6. Bravo, L., Roque, V. G., Brett, J., Dizon, R., & L'Azou, M. Epidemiology of Dengue Disease in the Philippines. PLoS Neglected Tropical Disease. 2014;8.
7. Wilder-Smith, A., Barrett, A. D., Carrijo, K., Cavaleri, M., de Silva, A., Durbin, A. P. et al., Clinical development and regulatory points for consideration for second-generation live attenuated dengue vaccines. Vaccine. 2018;36(24):3411-3417.
8. Georgiou, A. D. An introduction to time series forecasting. 2021. Infoworld. Available: https://www.infoworld.com/article/3622246/an-introduction-to-time-series-forecasting.html.
9. Bharatpur, A. A Literature Review on the Time Series Forecasting Methods. ResearchGate. 2022.
10. Almasarweh, M., & Wadi, S. A. ARIMA Model in Predicting Banking Stock Market Data. Modern Applied Science. 2018;12(11):309.
11. Taylor, S. J., & Letham, B. Forecasting at scale. 2017.
12. Belikov, D., Arshinov, M., Belan, B., Davydov, D., Fofonov, A., Sasakawa, M., et al., Analysis of the Diurnal, Weekly, and Seasonal Cycles and Annual Trends in Atmospheric CO2 and CH4 at Tower Network in Siberia from 2005 to 2016. Atmosphere. 2019;10(11):689.
13. Indhuja, M., & Sindhuja, P. P. Prediction of Covid-19 cases in India using prophet. International Journal of Statistics and Applied Mathematics. 2020;5(4):103-106.
14. Kulshreshtha, V., & Garg, N. K. Predicting the New Cases of Coronavirus [COVID-19] in India by Using Time Series Analysis as Machine Learning Model in Python. Journal of the Institution of Engineers (India). 2021;102(6):1303-1309.
15. Prabhakaran, S. ARIMA Model-Complete Guide to Time Series Forecasting in Python. Machine Learning Plus. 2022. Available: https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/.
16. Vincent, T. A Guide to Time Series Forecasting with ARIMA in Python 3. DigitalOcean Community. 2017. Available: https://www.digitalocean.com/community/tutorials/a-guide-to-time-series-forecasting-with-arima-in-python-3.
17. Shetty, C. Time Series Models-Towards Data Science. Medium. 2017. Available: https://towardsdatascience.com/time-series-models-d9266f8ac7b0.
18. Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. Application of the ARIMA model on the COVID-19 epidemic dataset. Data in Brief. 2020.
19. Mehrmolaei, S., & Keyvanpour, M. R. Time series forecasting using improved ARIMA. Artificial Intelligence and Robotics (IRANOPEN). 2016.
20. Lewis, C.D. Industrial and business forecasting methods: a practical guide to exponential smoothing and curve fitting. Butterworth Scientific. 1982.