ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Exploring Morbidity Cases in a Municipality in Zamboanga del Norte, Philippines: A Time Series Analysis and Forecasting Study

Urbano B. Patayon[1], Simon M. Arsola[2], Marcedita P. Ejan[3], Anselmo M. Dumaway[4] and Elidio C. Quiboyen Jr.[5]

ABSTRACT

Despite governmental efforts, the Tampilisan Rural Health Unit (RHU) needs comprehensive intervention plans for diverse health challenges, prompting the municipality to prioritize the development of sustainable solutions based on historical data and forward-thinking to safeguard residents' well-being. A time series analysis of morbidity cases in Tampilisan, Zamboanga del Norte, identified prevalent illnesses, including Upper Respiratory Tract Infection, Hypertension, Urinary Tract Infection, Post Traumatic Wound Infection, and Pneumonia, ranking as the most common. Notably, morbidity cases significantly declined from January 2020 to January 2022, likely influenced by the Covid-19 pandemic. The study employed the Prophet algorithm and fine-tuned hyperparameters during training and revealed that pneumonia had the highest R2 value (0.97) and the smallest RMSE (1.47), indicating a good fit. Conversely, the Gastritis dataset exhibited a lower R2 value (0.79) and the highest RMSE value (65.34), suggesting a weaker fit. Forecasting projected minimal uncertainty in morbidity rates for the six common illnesses, offering practical implications for local health authorities to formulate effective interventions for preventing and controlling these illnesses within the community.

Article information:

Keywords: Time Series Analysis, Morbidity Cases, Prophet Algorithm, Public Health, Zamboanga del Norte

Article history:
Received: June 28, 2023
Revised: October 19, 2023
Accepted: December 14, 2023
Published: December 23, 2023 (Online)

DOI: 10.37936/ecti-cit.2023174.253287

1. INTRODUCTION

Tampilisan, a municipality in Zamboanga del Norte, Philippines, faces significant health challenges. The high morbidity rate and prevalence of infectious diseases such as tuberculosis, measles, and malaria are attributed to the lack of access to safe drinking water and sanitation, poor nutrition, and inadequate healthcare infrastructure [1]. Despite efforts by the government to improve healthcare in rural areas, including Tampilisan, through initiatives such as the PhilHealth program, which provides low-income families with health insurance coverage, the Rural Health Unit (RHU) of Tampilisan still needs to develop intervention plans for lifestyle diseases, degenerative diseases, and emerging infectious diseases [2]. The recent COVID-19 pandemic has further highlighted the urgent need to strengthen the municipality's healthcare infrastructure and implement effective strategies to prevent and manage infectious diseases. Tampilisan must work towards sustainable solutions to address these health challenges and ensure the well-being of its residents. To provide sustainable solutions for addressing these health challenges and ensuring the well-being of its residents, it is necessary to base plans on historical facts and futuristic thinking, creating a comprehensive understanding of the population's health status [3].

Time-series analysis effectively studies data trends over time and is frequently utilized in future planning. It involves determining whether the relationships between dependent parameters are linear or nonlinear. While numerous linear and nonlinear techniques are available for time-series forecasting [4,5], using linear assumptions to forecast morbidity data may lead to issues with nonlinear variables. Therefore, nonlinear techniques are the preferred method for morbidity forecasting, as they allow the prediction of future occurrences of non-linear data, such as diseases

[1,2,3,4,5] The authors are with College of Computer Studies, Jose Rizal Memorial State University, Zamboanga del Norte, Philippines, E-mail:urbanopatayon@jrmsu.edu.ph, simonarsola12@gmail.com, ejanthata@gmail.com, anselmodumaway@gmail.com and elidioquiboyen@jrmsu.edu.ph

or health conditions.

Prophet is an additive model that excels when there are strong seasonal trends and an abundance of historical data spanning numerous seasons. It distinguishes itself from competing approaches by effectively handling missing data, adapting to changes in trends, and robustly managing outliers [6]. This ground-breaking instrument, known for its ease of use and potency, has been used in a variety of industries, including banking analysis of stock markets [7], criminal assessments and forecasts [8], modeling board examination volume trends [9], and epidemiological forecasting [10-14].

To effectively address health challenges in Tampilisan, a third-class municipality in the Philippines, it is crucial to devise sustainable solutions grounded in historical data and forward-thinking. The primary objective of this research is to develop a model trained on morbidity data to predict future trends and seasonal patterns in morbidity. Utilizing historical data and artificial intelligence, particularly the Prophet model, to forecast future morbidity rates is one technique that can aid local officials in decision-making. This study also proposes a straightforward method for creating forecasting models from the morbidity dataset using the Prophet algorithm. This algorithm is chosen for its flexibility, robustness to missing data and outliers, and scalability.

The study utilized recorded morbidity cases from 2017 to 2023 and employed the Prophet model to forecast future rates. The results of this study can be used to create mitigation plans and policies concerning community health, prevent disease outbreaks, reduce the impact of disease on communities, and improve public health outcomes.

2. RELATED RESEARCH

2.1 Applications of Times Series Analysis in Health

Time series analysis, a statistical technique used to analyze data collected over time, is crucial in uncovering patterns, trends, and dependencies for predicting future behavior [15]. Its application in medicine has transformed the field into a data-driven domain. Time series datasets, such as electronic health records, offer valuable insights into a patient's lifetime health information. Leveraging machine learning models, these datasets enable a deep understanding of individual health trajectories, including the progression of multiple conditions. Machine learning facilitates comprehensive, personalized care that adapts to patients' evolving contexts and medical histories by examining disease patterns and changes over time. The analysis of time series data contributes to unraveling insights about various diseases, supporting advancements in precision medicine [16].

Furthermore, wearable sensors and smart healthcare devices have revolutionized the collection of longitudinal medical data, requiring minimal effort from individuals. Consequently, abundant medical data is available for healthy individuals and those with medical conditions. These datasets serve as valuable resources for data analysis, allowing researchers to uncover intricate patterns and develop accurate forecasting models. Such forecasting capabilities are essential for identifying practical solutions for specific medical conditions [17].

2.2 Common Forecasting Algorithms

In the field of time series analysis, ARIMA (autoregressive integrated moving average) models are widely used to study and predict future trends based on past values [18]. These models offer flexibility and versatility when analyzing univariate time series data, as they can handle various patterns, such as linear or nonlinear trends, constant or varying volatility, and seasonal or non-seasonal fluctuations. ARIMA models are known for their simplicity and ease of implementation, requiring minimal parameters and assumptions. They provide reliable forecasts and confidence intervals based on statistical methods and theory [19]. However, it is essential to note that ARIMA models have limitations when analyzing multivariate time series data, as they are designed for univariate analysis. They cannot capture interactions between variables or consider external factors. ARIMA models require extensive data preprocessing and parameter tuning, and the assumptions of normal distribution and constant variance may not hold true for all datasets. These limitations should be considered when applying ARIMA models to real-world data [19].

On the other hand, SARIMA (Seasonal AutoRegressive Integrated Moving Average) models are an extension of ARIMA explicitly designed to capture seasonal patterns in time series data [20]. SARIMA models excel in handling data with seasonal patterns by incorporating seasonal differencing, autoregressive, and moving average components. They offer flexibility in capturing different trends, seasonality, and autocorrelation patterns. SARIMA models provide interpretable coefficients, offering insights into the influence of past observations. Known for their accurate forecasting capabilities, SARIMA models are suitable for both short-term and long-term predictions. However, SARIMA models also come with challenges and limitations. The selection of appropriate model orders can be time-consuming and requires expertise. SARIMA assumes stationarity, necessitating preprocessing for non-stationary data. Additionally, SARIMA models may have limitations in handling outliers and capturing sudden data changes or innovations [20].

In addition to ARIMA and SARIMA models, autoregressive (AR) models are commonly used in time series analysis [20]. Autoregressive models leverage the relationship between past and current values within the same series to predict future behavior. By employing linear regression techniques, the AR model uncovers the underlying dynamics of the time series and enables accurate predictions based on historical patterns. Autoregressive models assume that past values influence current values, particularly relevant in financial markets. However, the limitations of autoregressive models were evident during the 2008 financial crisis, as they failed to predict the sharp decline in stock prices caused by emerging risks. These highlight significant events' lasting impact on variables in autoregressive models [21]. While the AR model is a valuable tool for analyzing time series data, capturing autocorrelation, and offering simplicity and interpretability, it assumes linearity, which may not hold for data with complex patterns. Stationarity is also an assumption, and modifications are required for non-stationary data. Selecting the appropriate lag length can be challenging, and incorrect order selection can lead to biased estimates. The AR model does not inherently incorporate external factors, necessitating additional models or extensions for their inclusion. Moreover, the AR model is sensitive to outliers, which can impact parameter estimates and forecasts. Robust estimation techniques or alternative models can help mitigate the effects of outliers [22].

On the other hand, the Prophet algorithm is a time series forecasting algorithm designed to be an easy-to-use and effective tool for forecasting time series data with various patterns, such as trends, seasonality, and holiday effects. Prophet is particularly popular for its ability to handle data with missing values, outliers, and irregular sampling intervals, making it suitable for real-world datasets that are often messy and noisy. The algorithm considers the uncertainties in the data and provides uncertainty intervals for the forecast, helping users understand the confidence level in the predictions. It has found applications in various domains, including retail, finance, healthcare, and more [23, 24].

3. MATERIALS AND METHODS

3.1 Research Design

The research employed the input-process-outcome model as its research design. Figure 1 displays the model utilized to achieve the objectives of the research.

To begin the preliminary phase of objective number 1, researchers reviewed related studies, interviews, and documentary analysis at the RHU. During this phase, the data gathered from the RHU underwent a thorough cleaning, and unnecessary information irrelevant to the study's scope was removed. The cleaned data were used as the input for training models developed using the Prophet algorithm as part of objective number 2. During the training phase, fine-tuning was performed to achieve better results. The outputs of objective number 2 were different models trained for each morbidity case.

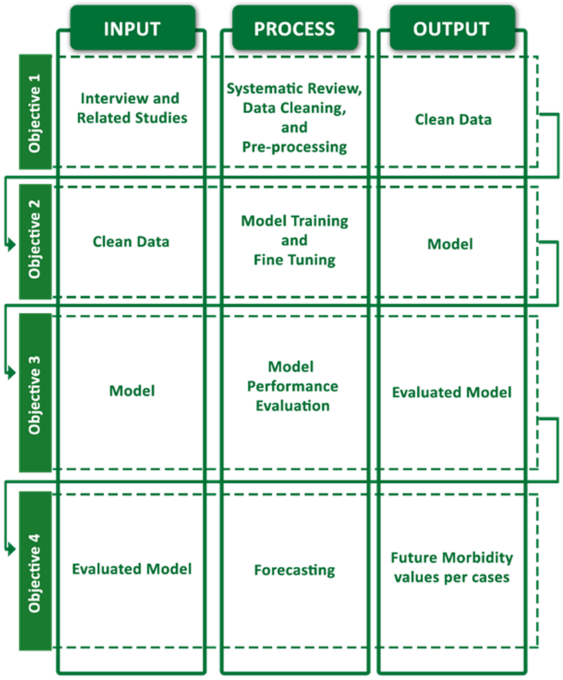


Fig.1: *Research design adopted in the study.*

Objective number 3 involved evaluating the models developed in objective number 2. Testing datasets were used to assess the performance of these models. The outputs of objective number 3 were the evaluated models.

In the preliminary phase of objective number 4, the models were utilized to forecast future morbidity values for each case. The results of this objective were the predicted future morbidity values for each case.

3.1.1 Experimental Set-up

The study involved collecting morbidity data from the Rural Health Unit in Tampilisan, Zamboanga del Norte. The data underwent thorough cleaning and validation before proceeding with model training and evaluation. The Prophet algorithm was used for training to estimate future morbidity levels and find the best model. This algorithm revealed underlying patterns in historical sequence data. Hyperparameters for each model were cautiously selected and adjusted. Evaluation metrics were then utilized to evaluate the performance of the candidate models. R2 evaluates how well the predicted values match the actual values, whereas RMSE examines how tightly the data points are clustered around the line of best fit. Models were used for forecasting after obtaining satisfactory evaluation results. A multi-step forward method of forecasting was used to calculate future morbidity rates. Figure 2 shows the experimental design used in the research.

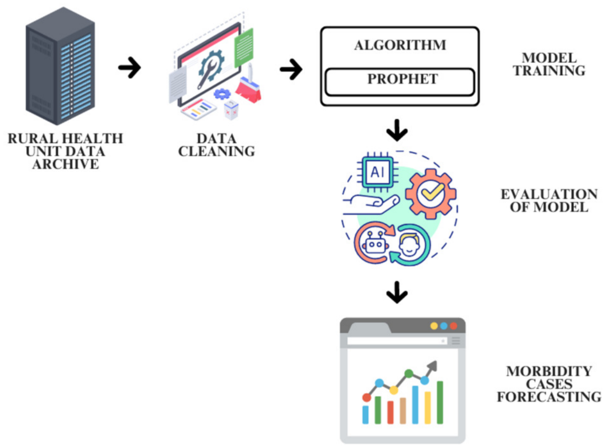


Fig.2: *Experimental setup used in the study.*

3.2 Data Collection and Pre-processing

The researchers used data from the Tampilisan Rural Health Unit's office in Tampilisan, Zamboanga del Norte, Philippines. A letter of request was submitted to ensure the legality of data gathering and to ensure the safekeeping of the records per the data privacy law. Data from the year 2017 to 2023 were extracted and utilized. The morbidity cases considered in this study include Upper Respiratory Tract Infection with 9701 cases, hypertension with 2523 recorded cases, Urinary Tract Infection with 1429 recorded cases, Post Traumatic Wound Infection with 770 recorded cases, Gastritis with 359 recorded cases, Myalgia with 352 cases, Superficial Injury with 232 cases, Noninfective Gastroenteritis with 312 recorded cases, Allergic Contact Dermatitis with 253 cases, Neuralgia with 140 recorded cases, Cutaneous Abscess with 227 cases, Impetigo with 254 cases, Osteoarthritis with 95 cases, Type 2 Diabetes with 171 cases, Pneumonia with 509 cases, Varicella with 121 cases, Allergic Urticaria with 77 cases, Bronchial Asthma with 159 cases, Muscle Pain with 34 cases, Acute Rhinitis with 146 cases, Open Wound with 64 cases, Tension-Type Headache with 114 cases, Acute Tonsilitis with 132 cases, Chicken Pox with 31 cases, Non-Infectious Diarrhea with 26 cases, Infected Wound with 130 cases, and Covid 19 with 238 cases. The 73 data points that constitute the final data have timestamps ranging from January 2017 to January 2023. Anchoring to the research by Chimulla & Zhang [25], the dataset is divided into two segments: an 80% training set comprising 58 data points for model training and a 20% testing set with 15 data points for model testing. Normalization was applied to address data inconsistencies.

3.3 Training

3.3.1 Hardware and Software

During the model training and testing period, the researcher utilized Google Colaboratory (Google Colab), which provides access to accessible GPUs and TPUs, powerful computing resources for training machine learning models, and processing large datasets.

The training process involved the use of two powerful tools: the Prophet forecasting algorithm developed by Facebook's Core Data Science team and Pystan, a Python package that interfaces with Stan, a probabilistic programming language for Bayesian inference [26, 27]. While several deep learning frameworks are available, including Theano, Caffe, Tensor Flow, Chainer, Microsoft Cognitive Toolkit or CNTK, MXNet, and OpenCL, the researchers opted for these specific tools for their suitability to the study's objectives. Pystan enabled the building, fitting, and analysis of Bayesian models using Stan, while the Prophet algorithm was used for forecasting. All of these tools were executed in the Python programming language and implemented through the Jupyter Notebook application.

3.4 FB Prophet

The Prophet algorithm was employed in the training and prediction stage, considering hyperparameters such as seasonality and growth. Prophet, an open-source algorithm for generating time-series models, combines traditional principles with innovative approaches. It is particularly effective in modeling time series with multiple seasonal patterns and addresses certain drawbacks in other methods [6, 28]. The algorithm's foundation lies in the summation of four-time functions, as shown in the equation below:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{1}$$

Where:

$g(t)$ = growth function
$s(t)$ = seasonality function
$h(t)$ = holiday/event Function
ϵ_t = error

The growth component represents the general trend of the data. Changepoints occur when the data changes its direction and serve as points where the trend can be altered or modified. There are three primary types of growth functions. The first is logistic growth, which is suitable for time series data that exhibit saturation and reach maximum or minimum values over time. Linear growth utilizes piecewise linear equations with varying slopes between changepoints and Flat growth, which indicates no significant overall change over time but still exhibits seasonality.

To capture the seasonality component, the Fourier Series was employed, taking into account the relationship between time and the seasonality function. The Fourier Series is composed of a combination of successive sine and cosine functions, with each term multiplied by a coefficient. In the case of Facebook Prophet, this summation was utilized to replicate the seasonal patterns observed in the data [6, 28] or approximate various types of curves. The Fourier series can be represented as follows:

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

Prophet automatically determines the optimal number of terms in the series, referred to as the Fourier order. The choice of Fourier order in Prophet depends on the specific characteristics of the dataset being analyzed. A higher order is selected when more terms are included in the series [6, 28].

Additionally, Prophet incorporates a holiday function that enables it to adjust predictions when holidays or significant events occur. By providing a list of dates, Prophet incorporates information from past data associated with those specific holiday dates. It then adjusts the predicted values by adding or subtracting from the growth and seasonality terms when each holiday date is encountered in the forecast [6, 28].

In this study, the researchers fine-tuned the algorithm's hyperparameters, such as seasonality, interval_width, and mcmc_samples, to achieve better accuracy.

3.5 Evaluation and Forecasting with Prophet

The results were evaluated using two metrics: the root mean square error (RMSE) and the coefficient of determination (R2). These statistical measures assess the performance and accuracy of the developed methods by comparing the target values with the corresponding output values. The RMSE and R2 generate scores that serve as indices to gauge the quality of the implemented approaches.

$$RMSE = \frac{1}{N} \sum (A - P)^2 \quad (3)$$

$$R^2 = \frac{N \sum (AP) - \sum A \sum P}{N \sum A^2 - \sum A^2 N \sum P^2 - \sum AP^2} \quad (4)$$

The coefficient of determination (R2) is a statistical measure that quantifies the extent to which differences in another variable can explain variations in one variable. In simple terms, a higher R2 value indicates a better forecasting model. On the other hand, the RMSE assesses the distribution of residuals, indicating how closely the data points are clustered around the best-fit line. Essentially, the RMSE provides insights into the overall dispersion of the data. As part of this study, a six-month forecast was conducted for both the overall data and selected morbidity cases.

4. RESULTS AND DISCUSSIONS

4.1 Morbidity Cases Data

The final dataset used in this study consisted of 73 data points, with the latest timestamp being January 2023 and the oldest being January 2017. Figure 3 shows the number of morbidity cases, while Figures 4 and 5 depict the trend in morbidity case data in Tampilisan, Zamboanga del Norte. The data reveals that among the various morbidities, Upper Respiratory Tract Infection emerges as the most prevalent, with a staggering 9701 recorded cases. This finding highlights the significance of respiratory health in the region and emphasizes the need for targeted interventions and preventive measures to address this condition effectively [29].

Following closely is hypertension, with 2523 recorded cases. The substantial number of cases underscores the importance of raising awareness about healthy lifestyle choices and implementing hypertension prevention and control strategies. Urinary Tract Infection (UTI) is another noteworthy morbidity, with 1429 recorded cases. The prevalence of UTI indicates that healthcare providers must emphasize preventive measures, promote proper hygiene practices, and ensure timely diagnosis and treatment to reduce the burden of this condition [29]. With 770 recorded cases, post-traumatic wound infections underscore the need for timely and appropriate wound management, along with education on wound care practices. This can minimize infections, promote faster healing, and improve overall health outcomes [29]. Additionally, with 509 recorded cases, pneumonia highlights the necessity for vaccination programs, better access to healthcare services, and enhanced public health campaigns to reduce the incidence and severity of pneumonia, especially among vulnerable populations such as children and the elderly [29].

On the other hand, Non-Infectious Diarrhea stands out as the morbidity with the lowest number of cases, recording only 26 instances. While non-infectious causes of diarrhea may be less common in the municipality of Tampilisan, it remains crucial to investigate and understand the underlying factors contributing to these cases. Identifying the root causes and implementing appropriate preventive measures can help maintain this low incidence rate.

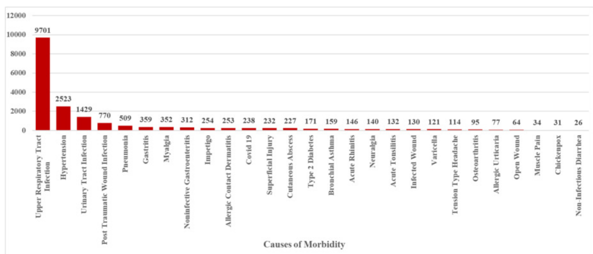


Fig.3: *Number of morbidity cases in the municipality of Tampilisan.*

The data also reveals an interesting trend in morbidity cases in Tampilisan, Zamboanga del Norte. From January 2017 to December 2019, higher trends of morbidity cases were recorded, indicating a more significant burden of illness during that period. However, a notable shift occurred from January 2020 to January 2023, with lower trends of morbidity cases. This change in morbidity trends can be attributed to various factors, including the impact of the COVID-19 pandemic. The first recorded case of COVID-19 in the Philippines was on January 30, 2020, while the Zamboanga Peninsula reported its first COVID-19 case on March 24, 2020 [30,31]. The fear and uncertainty surrounding the virus likely led to a reluctance among individuals to visit hospitals or rural health clinics for non-COVID-related health issues. This hesitancy resulted in a decrease in reported morbidity cases during the mentioned period. Factors such as the fear of exposure in healthcare settings, movement restrictions, and public health measures played a role in discouraging people from seeking medical attention. This phenomenon, known as the "pandemic effect" or "COVID-19 effect," has been observed worldwide, with non-COVID health services experiencing a decline in utilization [32, 33].

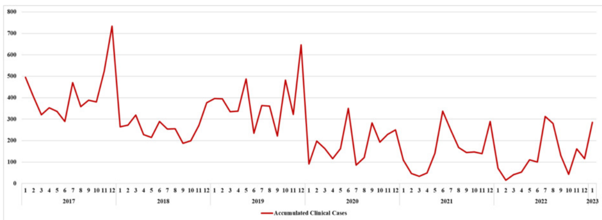


Fig.4: *Consolidated trend of morbidity cases in the municipality of Tampilisan, Zamboanga del Norte.*

Furthermore, the data analysis indicates a noticeable trend of decreasing morbidity cases from January 2020 to January 2022. However, there has been a gradual increase in morbidity cases from February 2022 to January 2023. Several factors may contribute to this upward trend. Implementing improved public health measures, such as increased testing capacity, contact tracing, and targeted interventions, may have led to better detection and reporting of morbidities. Additionally, the vaccination efforts against COVID-19 could have played a role in reducing the severity of the virus and consequently allowing individuals to seek healthcare for other conditions. As more people become vaccinated, the overall health-seeking behavior may have improved, increasing in reported morbidity cases. Furthermore, other COVID-19 prevention strategies, such as promoting hygiene practices, wearing masks, and social distancing, may have contributed to a decline in other infectious diseases. As a result, individuals might have been more inclined to seek medical attention for non-COVID-related morbidities, leading to the observed increase in morbidity cases during this period [34].

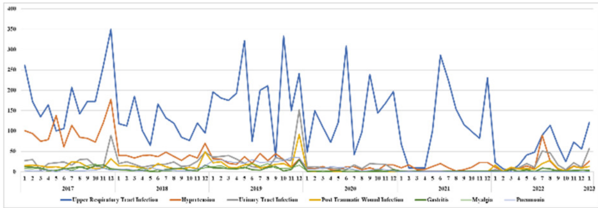


Fig.5: *Morbidity cases data trend in the municipality of Tampilisan, Zamboanga del Norte.*

4.1.1 Model Evaluation Result

The Prophet algorithm was applied to predict future morbidity cases in Tampilisan, Zamboanga del Norte. Table 1 presents the RMSE and R2 for various forecasting models' training and testing phases utilizing the morbidity cases dataset and Prophet. Notably, the pneumonia dataset, trained using the Prophet algorithm, had the highest R2 value of 0.97 and the lowest RMSE of 1.47 among the top 7 causes of morbidity. As evidenced by their higher R2 values and decreased RMSE, these results show that models developed using the pneumonia dataset and Prophet algorithm were highly concentrated within the best-fit line, suggesting a high level of accuracy. Other datasets, such as Post Traumatic Wound Infection, Myalgia, Urinary Tract Infection, Upper Respiratory Tract Infection, and Hypertension, also attained higher R2 values ranging from 0.87 to 0.95, along with smaller RMSE values ranging from 2.96 to 26.66 when trained using the Prophet time-series model generation algorithm indicating a strong relationship between the predicted and observed values. Similarly, the model trained using consolidated data also showed better results with a higher R2 value of 0.88 and a lower RMSE value of 49.74. This indicates that combining multiple datasets improved the model's performance, capturing the overall morbidity trend more effectively.

The model created using the Prophet algorithm and the Gastritis dataset exhibited the lowest R2 value of 0.79 and the lowest RMSE value of 65.34. These results indicate that the data points may not be closely clustered around the line of best fit, indicating a weaker relationship between the predicted and observed values. This could be due to various factors, such as the complexity of the Gastritis dataset and the presence of outliers that affect the model's accuracy.

The results during training and evaluation have proven that the Prophet algorithm is a valuable tool for predicting future morbidity cases in Tampilisan, Zamboanga del Norte. The accuracy and significant connections between the observed and projected values exhibited by the models trained on most of the datasets highlight the algorithm's potential for pre-

Table 1: *Results of the training and testing utilizing the morbidity case dataset and Prophet algorithm.*

Causes of Morbidity	RMSE	R2
Upper Respiratory Tract Infection	12.47	0.89
Hypertension	26.66	0.87
Urinary Tract Infection	6.84	0.90
Post Traumatic Wound Infection	2.96	0.95
Pneumonia	1.47	0.97
Gastritis	65.34	0.79
Myalgia	3.61	0.94
Consolidated Result	49.74	0.88

dicting morbidity cases. A comparison of the actual test data and prediction line is shown in Figure 6.

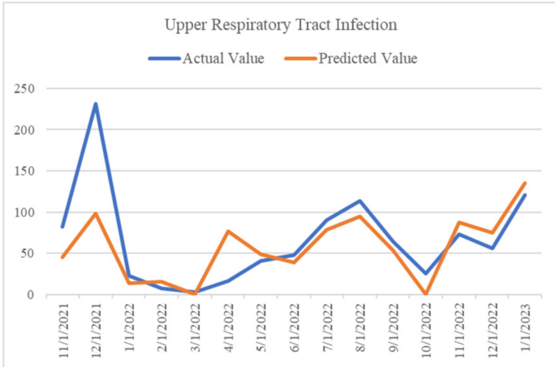
4.2 Forecasting

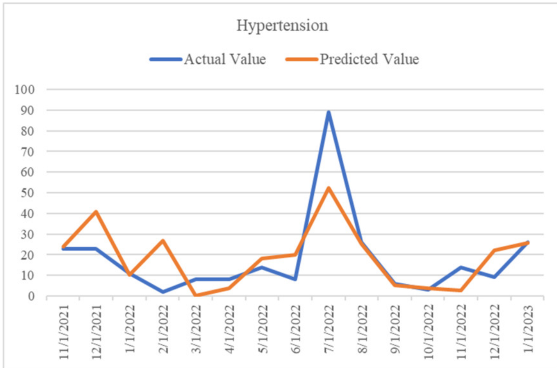
Table 2 showcases the predicted values for the models generated using morbidity case data and the Prophet approach for the upcoming six months. The table includes six causes of morbidity and their respective forecasted morbidity rates, upper and lower bounds, and the month and year of the forecast. The upper respiratory tract infection has a forecasted morbidity rate of 0 for February, March, May, and June 2023, while it has a morbidity rate of 552 in April 2023. Forecasted results also show that approximate upper respiratory tract infection cases for the six months will range from 0-42,369, as manifested in the upper and lower bounds given with a confidence interval of 95%.

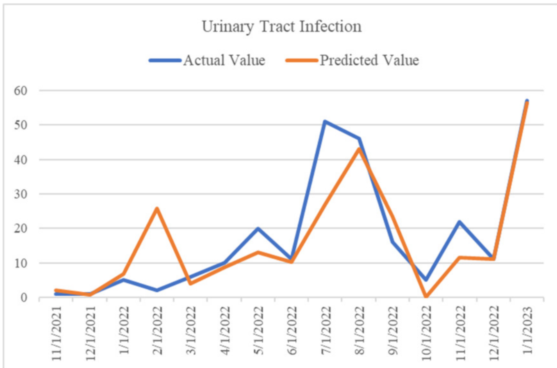
The forecasted morbidity rate for hypertension ranges from 0 to 148 from February to July 2023. The lower bound remains consistently at 0 for all months, while the upper bound ranges from 327 to 681. These variations indicate a high degree of uncertainty in the forecasted morbidity rates for hypertension. Urinary tract infection has a forecasted morbidity rate ranging from 0 to 3518 for February to July 2023, with a lower bound of 0 and an upper bound of 24,937.

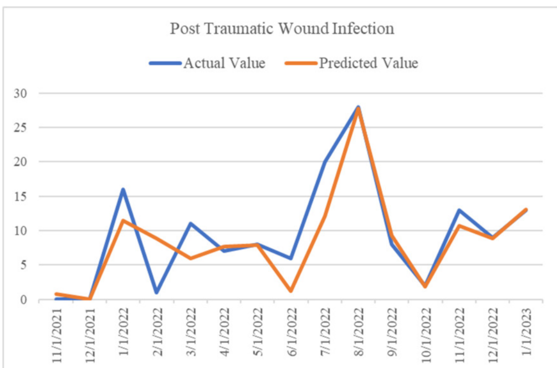
For post-traumatic wound infection, the forecasted morbidity rate ranges from 1 to 1178 for February and July 2023. The lower bound is 0, and the upper bounds are 15375 and 15381, respectively. As for Pneumonia, it has a forecasted morbidity rate of 0 for all months. However, an actual value can reach 44 cases, as indicated in the upper bound value.

The forecasted morbidity rate for pneumonia suggests consistently low values throughout the year. The zero forecasted value for pneumonia is attributed to the shallow recorded cases since January 2021. It is important to note that these exceptionally low recorded cases do not indicate a lack of pneumonia cases during that specific period but rather result from the ongoing pandemic. People in the municipality are hesitant to visit the Rural Health Unit (RHU) and seek medical attention due to the fear of being suspected of COVID-19 and subsequently subjected to quarantine.









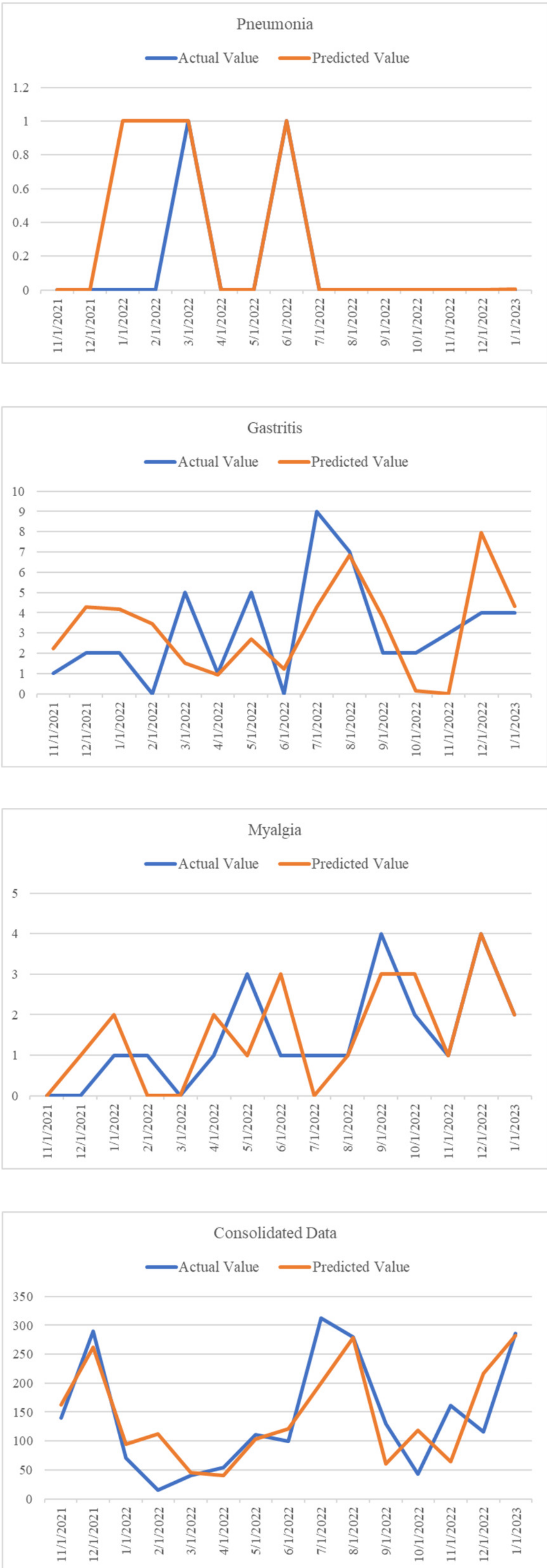


Fig.6: *Illustration of actual test data and prediction line.*

Table 2: *Six (6) months of the forecasted value of the models trained using morbidity cases data and prophet algorithm within 95 percent of the confidence interval.*

Causes of Morbidity	Month and Year	Forecasted Morbidity Rate	Lower Bound	Upper Bound
Upper Respiratory Tract Infection	February 2023	0	0	42369
	March 2023	0	0	189
	April 2023	552	0	1180
	May 2023	0	0	411
	June 2023	0	0	432
	July 2023	0	0	42233
Hypertension	February 2023	79	0	576
	March 2023	0	0	327
	April 2023	0	0	635
	May 2023	1	0	535
	June 2023	77	0	681
	July 2023	148	0	678
Urinary Tract Infection	February 2023	3518	0	24931
	March 2023	0	0	84
	April 2023	116	0	333
	May 2023	0	0	67
	June 2023	124	0	273
	July 2023	3502	0	24937
Post Traumatic Wound Infection	February 2023	1178	0	15375
	March 2023	18	0	58
	April 2023	59	5	118
	May 2023	1	0	48
	June 2023	87	47	129
	July 2023	1178	0	15381
Pneumonia	February 2023	0	0	44
	March 2023	0	0	60
	April 2023	0	0	25
	May 2023	0	0	27
	June 2023	0	0	41
	July 2023	0	0	32
Gastritis	February 2023	0	0	77
	March 2023	7	0	83
	April 2023	6	0	106
	May 2023	9	0	94
	June 2023	0	0	80
	July 2023	0	0	80
Myalgia	February 2023	15	0	72
	March 2023	19	0	76
	April 2023	0	0	48
	May 2023	0	0	41
	June 2023	29	0	87
	July 2023	14	0	68
Consolidated Result	February 2023	1085	0	2811
	March 2023	318	0	1851
	April 2023	0	0	1257
	May 2023	0	0	1308
	June 2023	342	0	2124
	July 2023	936	0	2655

Gastritis has a forecasted morbidity rate ranging from 0 to 9 for February to May 2023, with varying upper and lower bounds. Specifically, for April 2023, the forecasted morbidity rate is 6, with an upper bound of 106. In the case of Myalgia, the forecasted morbidity rate ranges from 0 to 29 for February to June 2023, with varying upper and lower bounds. Notably, for March 2023, the forecasted morbidity rate is 19, with an upper bound of 76.

The consolidated result of the forecasted morbidity rates for all causes of morbidity shows that the highest morbidity rate is 1085 for February 2023, while the lowest morbidity rate is 0 for April, May, and July 2023. The upper bounds for all months range from 1257 to 2655, while the lower bounds are all 0.

5. CONCLUSIONS AND RECOMMENDATIONS

Analyzing the historical morbidity cases in the municipality of Tampilisan, Zamboanga del Norte, provides valuable insights into the prevailing illnesses and their corresponding trends and patterns in the region. The data reveals that Upper Respiratory Tract Infection, Hypertension, Urinary Tract Infection, Post Traumatic Wound Infection, and Pneumonia are the most common morbidity cases in the municipality. At the same time, Non-Infectious Diarrhea has the lowest number of cases. Furthermore, the data indicates a noticeable trend of decreasing morbidity cases from January 2020 to January 2022, which could be attributed to the ongoing COVID-19 pandemic.

The training phase of the study involved utilizing the Prophet algorithm and fine-tuning hyperparameters such as seasonality, interval_width, and mcmc_samples. The model performed well, closely adapting to the data trends, indicating that the chosen hyperparameters were appropriate for this study. Furthermore, the pneumonia dataset exhibited the highest R2 value of 0.97 and the lowest RMSE of 1.47, suggesting that the models trained using the pneumonia dataset and the Prophet algorithm closely aligned with the best-fit line.

However, the model constructed using the Gastritis dataset and the Prophet algorithm yielded the lowest R2 value of 0.79 and the lowest RMSE value of 65.34. These results suggest that the data points might need to be more closely clustered around the line of best fit. In the forecasting phase, it was observed that the forecasted morbidity rates for each of the six common illnesses have minimal uncertainty, as evidenced by the upper and lower bounds of the confidence interval.

The study's results offer valuable insights into the morbidity cases in the municipality of Tampilisan, Zamboanga del Norte. Furthermore, it provides a more straightforward method for creating forecasts that address flexibility, missing data, outliers, and scalability. Based on these results, it is strongly recommended that a comprehensive analysis or research study of the various morbidity cases be undertaken, mainly focusing on those with higher recorded instances, to establish precise conclusions regarding the underlying causes of the observed increase. Furthermore, considering the enhancement and updating of the dataset by increasing the number of data points for more accurate forecasting is advisable.

References

[1] M. A. Salazar , R. Law, V. Winkler, "Health consequences of an armed conflict in Zamboanga, Philippines using a syndromic surveillance database," *International Journal of Environmental Research and Public Health*, vol. 15, no. 12:2690, Nov. 2018.

[2] R. C. Pangalangan, "The domestic implementation of the International Right to Health: The Philippine experience," *Advancing the Human Right to Health*, Oxford, pp. 143-158, Jul. 2013.

[3] S. T. Stewart , D. M. Cutler and A. B. Rosen, "Forecasting the effects of obesity and smoking on U.S. life expectancy," *New England Journal of Medicine*, vol. 361, no. 23, pp. 2252–2260, Dec. 2009.

[4] V. N. Vapnik, "An overview of statistical learning theory," in *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988-999, Sept. 1999.

[5] V. N. Vapnik, *The nature of statistical learning theory*, Springer, 2010.

[6] S. J. Taylor and B. Letham, "Forecasting at scale," 2017. [online]. Available:`https://doi.org/10.7287/peerj.preprints.3190v1.`

[7] Y. Xue, "Recursive neural network-based market demand forecasting algorithm for calligraphy practice products," *Journal of Mathematics*, vol. 2022, pp. 1–10, 2022.

[8] A. Thomas and N. V. Sobhana, "A survey on crime analysis and prediction," *Materials Today: Proceedings*, vol. 58, pp.310-315, 2022.

[9] A. S. Becker *et al.*, "Automatic forecasting of Radiology Examination Volume Trends for Optimal Resource Planning and allocation," *Journal of Digital Imaging*, vol. 35, pp. 1–8, Nov. 2021.

[10] M. Narmeen, M. U. Sattar, H. W. Khan, M. Fatima, M.-D. Azad and F. Ghani, "Impact of weather on COVID-19 in metropolitan cities of Pakistan: A data-driven approach," *International Journal of Computing and Digital Systems*, vol. 11, no. 1, pp. 905-915, Feb. 2022.

[11] C. B. Aditya Satrio, W. Darmawan, B. U. Nadia and N. Hanafiah, "Time series analysis and forecasting of coronavirus disease in Indonesia using Arima model and prophet," *Procedia Computer Science*, vol. 179, pp. 524-532, 2021.

[12] M. Kowalska, "Relationship between quality of ambient air and respiratory diseases in the Polish

population," *WIT Transactions on Ecology and the Environment*, vol. 207, pp. 195-202, 2016.

[13] J.-J. Zou, G.-F. Jiang, X.-X. Xie, J. Huang and X.-B. Yang, "Application of a combined model with seasonal autoregressive integrated moving average and support vector regression in forecasting hand-foot-mouth disease incidence in Wuhan, China," *Medicine*, vol. 98, no.6, pp. e14195, Feb. 2019.

[14] X. Song, J. Xiao, J. Deng, Q. Kang, Y. Zhang and J. Xu, "Time series analysis of influenza incidence in Chinese provinces from 2004 to 2011," *Medicine*, vol. 95, no. 26:e3929, 2016.

[15] R. Lund, "Time series analysis and its applications: With R examples," *Journal of the American Statistical Association*, vol. 102, no. 479, pp. 1079–1079, 2007.

[16] S. Aydin, "Time series analysis and some applications in medical research," *Journal of Mathematics and Statistics Studies*, vol. 3, no. 2, pp. 31–36, 2022.

[17] N. Tyagi, "5 applications of Time Series analysis." `https://www.analyticssteps.com/blogs/5-applications-time-series-analysis` (accessed Jul. 4, 2023).

[18] A. Hayes, "Autoregressive integrated moving average (ARIMA) prediction model," Investopedia. `https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp` (accessed ;Jul. 4, 2023).

[19] S. L. Ho and M. Xie, "The use of Arima models for reliability forecasting and analysis," *Computers & Industrial Engineering*, vol. 35, no. 1-2, pp. 213-216, Oct. 1998.

[20] Z. Xinxiang, Z. Bo and F. Huijuan, "A comparison study of outpatient visits forecasting effect between ARIMA with seasonal index and SARIMA," *2017 International Conference on Progress in Informatics and Computing (PIC)*, Nanjing, China, pp. 362-366, 2017.

[21] M. Valipour, "Long-Term Runoff Study Using Sarima and Arima models in the United States," *Meteorological Applications*, vol. 22, no. 3, pp. 592–598, Jul. 2015.

[22] H. Chhabra, "A comparative study of Arima and Sarima models to forecast lockdowns due to covid-19," *Research square*, 2022.

[23] B. Long, F. Tan and M. Newman, "Forecasting the monkeypox outbreak using Arima, Prophet, NeuralProphet, and LSTM models in the United States," *Forecasting*, vol. 5, no. 1., pp. 127-137, 2023.

[24] U. B. Patayon and R. V. Crisostomo, "Time Series Analysis on Enrolment Data: A case in a State University in Zamboanga del Norte, Philippines," *2022 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Depok, Indonesia, pp. 13-18, 2022.

[25] V. K. R. Chimmula and L. Zhang, "Time series forecasting of COVID-19 transmission in Canada using LSTM networks," *Chaos, Solitons & Fractals*, vol. 135, p. 109864, Jun. 2020.

[26] A. Shakeel, D. Chong and J. Wang, "Load forecasting of district heating system based on improved FB-prophet model," *Energy*, vol. 278, p.127637, Sep. 2023.

[27] M. Nishio, E. Ota, H. Matsuo,T. Matsunaga, A. Miyazaki and T. Murakami, "Usefulness of pystan and numpyro in Bayesian Item Response theory," *medRxiv*, 2023.

[28] J. Brownlee, "Time series forecasting with Prophet in python [Internet]." `https://machinelearningmastery.com/time-series-forecasting-with-prophet-in-python/2020` (Accessed Jun. 15, 2023).

[29] K. Timmis , W. E. Huang and J. Timmis, "Strategies to minimize preventable morbidity and mortality resulting from pandemics like covid-19," *Environmental Microbiology*, vol. 22, no. 10, pp. 4085-4092, 2020.

[30] X. Gregorio, "Philippines confirms first case of novel coronavirus." `https://www.cnnphilippines.com/news/2020/1/30/Philippines-coronavirus-case.html` (accessed: Jun. 15, 2023).

[31] GMA News Online, "Beta variant cause of covid-19 community transmission in Zamboanga City-mayor." GMA News Online. `https://www.gmanetwork.com/news/news/regions/791336/beta-variant-cause-of-covid-19-community-transmission-in-zamboanga-city-mayor/story.` (accessed Jun. 15, 2023).

[32] T. M. Aklilu *et al.*, "The impact of covid-19 on care seeking behavior of patients at tertiary care follow-up clinics: A cross-sectional telephone survey Addis Ababa, Ethiopia," *medRxiv - Health Systems and Quality Improvement*, 2020.

[33] H. B. Joseph, S. Kuppusamy, S. K. Mahalik, A. P. Shetty and K. Das, "Telemedicine – a boon to parents of children with health care needs during COVID-19 pandemic: A qualitative study from India," *Turkish Archives of Pediatrics*, vol. 57, no. 5, pp. 526–531, Sep. 2022.

[34] N. C. Cardenas, "Correspondence harnessing strategic policy on COVID-19 vaccination rollout in the Philippines." *Journal of Public Health*, vol. 44, no. 2, pp. e279-e280, Jun. 2022.

Urbano B. Patayon is an Instructor I at Jose Rizal Memorial State University (JRMSU). He is the program head of Computer Science Department of Jose Rizal Memorial State University-Tampilisan Campus. He obtained his Bachelor of Science in Computer Science degree at Jose Rizal Memorial State University as Academic Scholar and finished his Master of Science in Computer Science at Mindanao State University-Iligan Institute of Technology as CHED K-12 scholar. He is an educator and a researcher at heart. He has presented researches in national and international fora and had published scientific papers in SCOPUS indexed journals. His research interest focusses in artificial intelligence, learning science, internet of things, data science, and software engineering.

Simon M. Arsola is a Bachelor of Science in Computer Science graduate from Jose Rizal Memorial State University-Tampilisan Campus. He was a former member of the Computational Science Group at Jose Rizal Memorial State University-Tampilisan Campus. His research interests focus on machine learning and image processing.

Mercedita P. Ejan is a Bachelor of Science in Computer Science graduate from Jose Rizal Memorial State University-Tampilisan Campus. As a former member of the Computational Science Group at the university, she has honed her skills and expertise in the field. Mercedita's research pursuits are centered around the realms of machine learning and image processing, showcasing a keen interest in advancing these domains.

Anselmo M. Dumaway successfully earned a degree in Bachelor of Science in Computer Science from Jose Rizal Memorial State University-Tampilisan Campus. As a past contributor to the Computational Science Group at the same university, he dedicated his efforts to advancing the field. Anselmo's research pursuits primarily center around the intricate realms of machine learning and image processing, reflecting his focused interest in these areas.

Elidio C. Quiboyen Jr. holds the position of Assistant Professor I at Jose Rizal Memorial State University (JRMSU). He earned his Bachelor of Science in Computer Science degree from the University of San Jose – Recoletos and subsequently achieved his Master's degree in education from Saint Vincent's College. Rooted in a passion for education and research, Elidio is both an adept educator and a dedicated researcher. His research pursuits are centered on the realms of learning science and software development, showcasing a commitment to advancing knowledge in these areas.