Received: June 17, 2023. Revised: August 21, 2023.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Forecasting LTE Network Traffic Using Hot-Winter's Multiplicative Seasonal Method and Rolling Forecasting for Telecommunication Company Investment Optimization in Indonesia

Endi Rizal Ferdiansyah[1]* **Adian Fatchur Rochim[1]** **Wahyul Amien Syafei[1]**

[1]*Information System, Postgraduate Study, Diponegoro University, Indonesia*
*Corresponding author's Email: endirizal.f@gmail.com

Abstract: The rapid growth of mobile internet usage, particularly in LTE networks, poses challenges for mobile operators in maintaining quality of service (QoS) and optimizing network design planning. Accurate traffic volume forecasting is crucial for network planning and effective resource allocation. This study aims to propose an enhancement to the accuracy of the holt's winter multiplicative seasonal (HWMS) method by applying the rolling forecast technique to achieve better forecasting results. Using a public dataset of 56 cells, the evaluation based on mean absolute percentage error (MAPE) reveals the following sequence of prediction errors: HWMS & rolling forecast (20.47%), HWMS only (30.06%), FbProphet (30.45%), and auto regressive integrated moving average (ARIMA) (31.52%). The cells are categorized as "Good" (59%), "Reasonable" (39%), and "Poor" (2%). The results indicate that the proposed method achieves a significantly higher percentage of cells in the "Good" category, with a 45% difference compared to HWMS without rolling forecast, which only achieved 14%. Moreover, the proposed method outperforms ARIMA by 50% and FbProphet by 37% in the same category. Furthermore, when applied to real data from a telecommunications company in Indonesia, the proposed method identified 11 cells that require solutions out of a total of 100 cells. In comparison, the ARIMA method identified 3 cells, FbProphet identified 12 cells, and HWMS without Rolling Forecast identified 9 cells. Thus, the company can provide solutions for the identified 11 cells without the need for excessive investment while still maintaining revenue potential.

Keywords: Traffic forecasting, LTE network, HWMS, Rolling forecast, ARIMA, FbProphet.

1. Introduction

The growth of mobile internet usage, especially on LTE networks, is increasing rapidly. In 2019, the average per capita volume of traffic usage reached 7.27GB per month [1] and it is expected to be 200 to 1,000 times in the following years [2]. Therefore, mobile operators face challenges in maintaining Quality of Service (QoS) by optimizing their network design planning as efficiently and quickly as possible [3–8]. Forecasting future traffic volume is crucial for successfully planning, managing, and developing network systems [9, 10].

In planning telecommunications networks, inaccurate forecasting of traffic growth can lead to issues. Forecasting results that underestimate the volume of traffic may result in bottlenecks that reduce the network's capacity to accommodate users, potentially leading to decreased revenue for the company [11]. On the other hand, forecasting errors that result in overly high LTE traffic predictions can lead to wasted resources invested by the company in low-demand areas. Therefore, an effective forecasting system is crucial to avoid such inefficiencies in resource allocation [10, 12].

Currently, there is extensive research and development related to traffic forecasting using conventional time series forecasting methods such as auto regressive integrated moving average (ARIMA), exponential smoothing, and machine learning techniques such as FbProphet, long short-term memory (LSTM), gated recurrent unit (GRU), and artificial neural network (ANN) [7, 9, 11–13] as well as combination methods such as FbProphet and K-means [14]. The aim of these studies is to achieve high accuracy in traffic forecasting that is in line with the requirements. Exponential smoothing method demonstrates high accuracy compared to other conventional methods like ARIMA, and it also has lower computational requirements compared to machine learning approaches [9, 15]. Moreover, Holt's winter multiplicative seasonal (HWMS) is a variant of exponential smoothing that is particularly suitable for forecasting traffic datasets compared to other variants of exponential smoothing [9].

However, the accuracy of the HWMS forecasting method may gradually decrease as the forecasted values move further away from the last data point in the training set [16]. Furthermore, if the test dataset is too short to be adequately observed, efforts are needed to improve the forecasting accuracy by conducting resampling techniques [17]. *Walk forward cross validation, also known as rolling forecast*, can maximize the accuracy of forecasting by retraining the model at each forecasting step. This is achieved by incorporating the observed test data into the training data, allowing for continuous reevaluation and improvement of the model's performance in subsequent forecasts [16, 18] . Holt's winter multiplicative seasonal (HWMS) method validated through rolling forecast is expected to resuslt in higher accuracy in forecasting LTE traffic volume than ARIMA, FbProphet, or HWMS without rolling forecast methods.

Most textbooks recommend the use of mean absolute percentage error (MAPE) as the recommended accuracy evaluation method [19]. A good forecasting system based on MAPE calculations can help companies plan network designs more accurately. This allows them to optimize investments such as bandwidth addition or implementation of new cells as needed. This traffic forecasting system can also be used for decision-making in network design engineerings, such as load balancing, resource allocation or relocation and traffic attack detection [20]. This research is expected to contribute knowledge as an alternative traffic forecasting system that is cost-effective in terms of cost-benefit analysis for telecommunications companies [21].

This paper is organized into sections in which section 1 describes the concept of forecasting LTE traffic, the challenges faced by telecommunications companies, then introduces the proposed method. Section 2 discusses related research developments that have previously been carried out in recent years. Section 3 describes the research method that will be carried out starting from the dataset used, the forecasting method studied, the evaluation process, and determining the cells that need a solution. Section 4 presents a comparison of forecasting results between methods and their application to the real company data in Indonesia. Section 5 is the final part of the presentation as a conclusion obtained from this work.

2. Related work

Several studies have been conducted to explore various methods in predicting network traffic. Generally, network traffic prediction techniques can be categorized into two approaches: statistical-based and machine learning-based [22]. Forecasting systems utilizing statistical methods such as exponential smoothing have been found to possess high accuracy, computational simplicity, and cost-effectiveness[9, 10]. This notion is further supported by a comparative study conducted by [15], which concluded that the accuracy of exponential methods is comparable to trending machine learning methods. The single, double, holt-winter's no seasonal, holt-winter's additive seasonal, and holt-winter's multiplicative seasonal (HWMS) methods have been tested , and the results indicate that HWMS, as a variant of Exponential Smoothing, outperforms other variants in forecasting daily network traffic [9, 10]. However, HWMS still faces challenges in handling outliers in the dataset and may exhibit poor performance and unstable predictions when applied to longer-term forecasting.

Holt's trend exponential smoothing method has also been developed [23] by integrating it with K-means clustering to accommodate diverse data characteristics. Their findings suggest that the proposed model outperforms the holt's trend exponential smoothing method without clustering in terms of prediction accuracy. Similar to other variants of exponential smoothing, HWMS experiences a gradual decline in accuracy as the forecast results move further away from the last data point in the training series [16]. Therefore, further enhanced the previous forecasting model by integrating exponential smoothing, fuzzy C-means, long-short time memory (LSTM), and adaptive neuro-fuzzy inference system (ANFIS) to achieve higher accuracy compared to previous benchmark methods [5]. However, this enhanced model becomes computationally complex and lengthy due to the presence of LSTM with its numerous hyperparameters.

Traditional methods with lightweight computations can still be viable options for predicting cellular traffic, as demonstrated by [13]. This study employed the ARIMA method with parameter (0,0,6) to predict cellular traffic, which proved effective in supporting LTE bandwidth provisioning. Furthermore, traditional methods can be improved by incorporating additional techniques, as demonstrated by [11], who combined ARIMA with the disruptive formula to predict LTE traffic. This study combines statistical techniques with judgemental approaches to enhance prediction accuracy compared to using statistical methods alone. However, ARIMA is highly sensitive to the selection of autoregressive order (p), differencing order (d), and moving average order (q). The incorporation of the disruptive formula introduces additional parameters that need to be accurately determined. Optimal selection of disruptive formula parameters can be challenging in the presence of new changes or unexpected patterns, and errors in parameter selection can impact prediction accuracy.

LSTM and artificial neural network (ANN) have recently gained popularity as machine learning methods for network traffic prediction due to their claimed higher accuracy compared to conventional methods like ARIMA and simple moving average (SMA) [19]. The advantage of using LSTM lies in its complex architecture with memory units and gates, enabling it to capture multivariate datasets, non-linear patterns, and interactions among variables. LSTM is specifically designed to handle forecasting or prediction problems that involve long-term dependencies between data points. With its complex memory units, the model can retain important information from previous data points over longer time ranges. However, this becomes a drawback of LSTM in terms of time efficiency, as it requires longer time for observation.

To address the challenges associated with LSTM, [24] utilized the gated recurrent unit (GRU) method to reduce architectural complexity. The results showed that GRU still achieved good prediction accuracy for network traffic, with a simpler architecture compared to LSTM. GRU also excelled in minimizing training time and computational load. However, both GRU, LSTM, and ANN are susceptible to overfitting if their hyperparameter settings are not properly tuned.

The machine learning method FbProphet, developed by Facebook, has gained popularity as it aids mobile operators in efficient planning and cost-saving investments, especially when compared to LSTM [12]. FbProphet has been utilized by [12, 14, 25–29] for predicting cellular traffic growth in telecommunications networks in their respective studies. Not only for traffic, FbProphet has also been employed by [28] to predict revenue generated by telecommunications service providers in Ghana.

FbProphet's accuracy has been compared to ARIMA, ETS, and SNAIVE by [27], with results showing FbProphet achieving 90% accuracy on their dataset. Additionally, in the case of 5G traffic dataset, study [25] highlighted FbProphet's ability to work with incomplete data, outliers, sudden changes in the data series, and provide reasonable estimates. Study [14] proposed a hybrid model combining FbProphet and K-means, similar to [23]. FbProphet was used for volume traffic prediction, while K-means served as the algorithm for determining holiday dataset clusters. The combination resulted in improved performance compared to using FbProphet alone. In general, FbProphet is user-friendly machine learning and exhibits good accuracy. However, combining FbProphet with clustering algorithms may not always be relevant for all datasets. This approach should be based on the characteristics and patterns observed in the data. Sometimes, the FbProphet model can provide adequate predictions without the need for integration with K-means, as seen in the studies by [26, 28, 29]. Therefore, FbProphet is chosen as one of the comparative methods in this study.

2.1 Problem definition

The mentioned studies have contributed to the development of traffic prediction methods in telecommunication networks, with the majority showing improved prediction accuracy. The use of machine learning techniques such as LSTM, GRU, Fbprophet, and their combinations promises high accuracy. However, these methods often require extensive parameter tuning, long dataset observation time, and a sufficient length of historical data, which may be considered less efficient. On the other hand, the traditional HWMS method has proven to be effective in achieving good accuracy for univariate datasets and is computationally efficient. However, the forecasting accuracy of the HWMS method typically decreases as the forecasting distance from the last training data increases. Additionally, if the dataset is too short for observation, resampling efforts need to be made.

2.2 Research objective

This research focus on increasing the accuracy of HWMS in predicting LTE network traffic by employing the rolling forecast technique. This technique involves adding observation data from the test dataset into the training dataset to re-examine each subsequent forecasting step.

2.3 Scientific contribution

This research is expected to provide a significant contribution to the development of traffic forecasting methods, particularly in LTE networks, with several key points:

1. Introducing an alternative method that is more accurate compared to conventional methods such as ARIMA, HWMS without rolling forecast, and popular machine learning methods like FbProphet.
2. By improving the forecasting accuracy, companies can plan network designs that are more precise and efficient.

3. Method

3.1 Dataset

This research and development work uses the HWMS method and rolling forecast for LTE traffic volume forecasting. The study was conducted from March 2022 to November 2022 in Indonesia. The datasets used in this study include public datasets obtained from Kaggle.com and internal datasets from a telecommunications company in Indonesia.

The public dataset from Kaggle.com contains data from 56 cells from October 23, 2017, to October 22, 2018. The internal dataset from the telecommunications company includes data from 100 cells, covering the period from August 1, 2020, to July 31, 2021, with LTE traffic volume measured in mega bytes. The cells in the internal dataset were selected based on specific criteria, including having a physical resource block ranging from 60% to 70%. This criterion is considered close to the maximum threshold for LTE cells with 1800 Bandwidth and 20MHz handling user traffic [12].

3.2 Forecasting steps

The proposed forecasting method is HWMS using Rolling Forecast to compare its accuracy with HWMS, ARIMA, and FbProphet. Below is an explanation of each method that was performed. Essentially, the steps for forecasting traffic in each method are the same, as shown in Fig. 1.

All forecasting methods in this study will use the same data preprocessing to ensure fairness. The step-by-step process of performing the forecasts, as shown in Fig. 1, is as follows:

Step 1. Scaling the time period to daily and converting the volume unit to mega bytes. Cleaning the empty traffic data, removing any data points that are missing. Checking the length of the dataset series to ensure consistency among all cells.

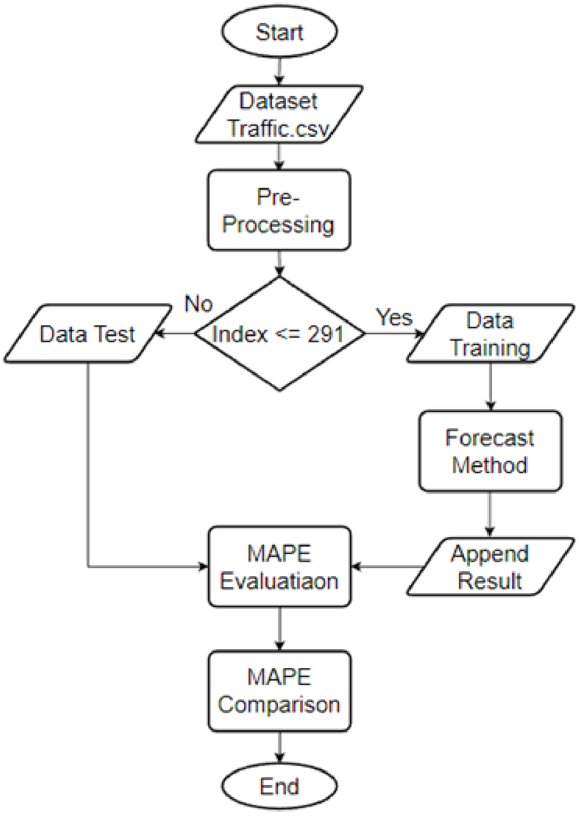

Figure. 1 Forecasting steps flowchart diagram

Step 2. The available dataset is divided into 80% for training data and 20% for testing data. The length of the dataset index is 365, so 80% of the index length is 292. These 292 indices will be observed by each method for forecasting purposes. The remaining 73 indices serve as the test data to evaluate the accuracy of the forecasted results.

Step 3. Perform the forecast and save it in a dataframe using the following steps for each method:

a. ARIMA

The ARIMA model is identified by three main parameters: p, d, and q. The parameter p (order of autoregressive term) indicates how many previous values will be used in the AR model. The parameter d (order of differencing) indicates how many times the time series needs to be differenced to become stationary in the integrated component I. The parameter q (order of moving average term) indicates how many previous prediction errors will be used in the MA model [13]. The $ARIMA(p, d, q)$ equation for modeling and predicting time series in this study is as follows :

$$ARIMA(p, d, q) = AR(p) + I(d) + MA(q) \quad (1)$$

Along with the details of its components:

Autoregressive components (AR):

$$AR(p) = \varphi^1 Y(t-1) + \varphi^2 Y(t-2) + \dots + \varphi_p Y(t-p) \quad (2)$$

Integrated components (I):

$$I(d) = (1-B)^d Y(t) \quad (3)$$

Moving average components (MA):

$$MA(q) = \theta^1 \varepsilon(t-1) + \theta^2 \varepsilon(t-2) + \dots + \theta\_q \varepsilon(t-q) \quad (4)$$

Where :
$Y(t)$: Value at time t in the time series.
$\varphi_1, \varphi_2, ..., \varphi_p$: Autoregressive coefficients.
B : Backshift operator, indicating a one-period shift backward.
p : Order of autoregressive term.
d : Level of differencing.
q : Order of moving average term.
$\theta_1, \theta_2, ..., \theta\_q$: Moving average coefficients.
$\varepsilon(t-1)$, $\varepsilon(t-2)$, ..., $\varepsilon(t-q)$: prediction errors at previous time points.

b. FbProphet

The open-source tool provided by Facebook in the form of a Python and R library combines the strengths of classical time series forecasting models and modern machine learning techniques [28, 30]. FbProphet's main features include handling multiple seasonality, non-linear trends, and holidays. The formulation of FbProphet is as follows:

$$y(t) = g(t) + s(t) + \in(t) \quad (5)$$

Where :
$y(t)$: Represents the time series
$g(t)$: Nonlinear saturating trend that models non-periodic changes in the series
$s(t)$: Component of annual periodic changes
$\in(t)$: Unusual changes that cannot be accommodated by the model.

This tool is user-friendly and flexible, with the ability to easily customize seasonal adjustments, making it ready to use with minimal tuning. FbProphet usually performs optimally when the dataset has the following characteristics:

- Hourly, daily, or weekly observations for at least several months
- Seasonal patterns directly related to human behavior
- Prediction disruptions occurring within predefined intervals
- Reasonable number of missing observations or extreme observations
- Historical trend changes
- Situations where the trend increases based on non-linear curves reaching its natural limits or saturation.

In this study, we only performed tuning on the interval forecasting aspect. The unique characteristics of LTE network traffic in different cell locations pose a challenge when using FbProphet.

c. HWMS

Holt-Winter's method essentially utilizes three smoothing constants: the level smoothing constant I_t, the trend smoothing constant B_t, and the seasonal smoothing constant S_t. There are two approaches that can be taken with Holt-Winter's method, namely multiplicative and additive. However, the emphasis in the research will be on the multiplicative approach because network LTE dataset highly fluctuating seasonal variations. The formula for the HWMS approach is as follows:

$$F_{t+m} = (S_t + B_t m) I_{t-L+m} \quad (6)$$

Along with its smoothing parameter values:
Overall smoothing *(level)*

$$S_t = \alpha \frac{X_t}{I_{t-L}} + (1-\alpha)(S_{t-1} + B_{t-1}) \quad (7)$$

Trend smoothing

$$B_t = \beta(S_t - S_{t-1}) + (1-\beta) B_{t-1} \quad (8)$$

Seasonal smoothing

$$I_t = \gamma \frac{X_t}{S_t} + (1-\gamma) S_{t-L} \quad (9)$$

Where:
S_t : Overall smoothing
B_t : Trend smoothing
I_t : Seasonal smoothing
S_{t-1} : The overall smoothing for the next period
B_{t-1} : The smoothing of the trend pattern for one period before
I_{t-L}: The smoothing of the seasonal pattern for the *L*-th period
X_t : The actual data for the *t*-th period

α, β, γ : Values for the parameters $0 \leq \alpha, \beta, \gamma \leq 1$
L : Seasonal length
F_{t+m} : The forecasted value for period m
m : The length of time.

Eq. (7) represents the overall smoothing level, while Eq. (8) represents the weighted average of the trend between the current level and the previous slope. Eq. (9) calculates the comparison of the latest seasonality with the last seasonality. The results of the first iteration are then stored in a data frame.

Step 4. The forecast results of each method, which have been stored in a data frame are evaluated for accuracy. This is done by comparing the forecasted values in the data frame with the actual test data using the Eq. (10):

$$MAPE = \frac{\frac{\sum|A_t - F_t|}{A_t}}{n} \times 100\% \quad (10)$$

Where:
$\sum|A_t - F_t|$: Total overall sum with absolute value.
A_t : Actual values
F_t : Forecasted values
n : Number of data points

MAPE calculates the difference between the actual and forecasted values, divided by the number of data points, and produces a positive value in percentage [31]. The forecasting method is considered better if the percentage of MAPE is smaller and vice versa. A detailed explanation of the evaluation process will be discussed in subsection 3-4 testing and system Evaluation.

Step 5. After determining the accuracy of each method using MAPE, the next step is to compare the accuracy of those methods with the proposed method.

3.3 Working prosses of HWMS and rolling forecast methods

The design of the HWMS and rolling forecast methods system is illustrated in a flowchart as shown in Fig. 2.

The sequence of steps in the HWMS and rolling forecast method in the system flowchart diagram above starts from:

Step 1. Performing pre-processing on the dataset by adjusting the time period scale.

Step 2. Dividing the dataset into 80% training data and 20% testing data. The difference from other methods is that the training data will always receive data updates from the dataset. In other words, HWMS will observe 100% of the dataset.

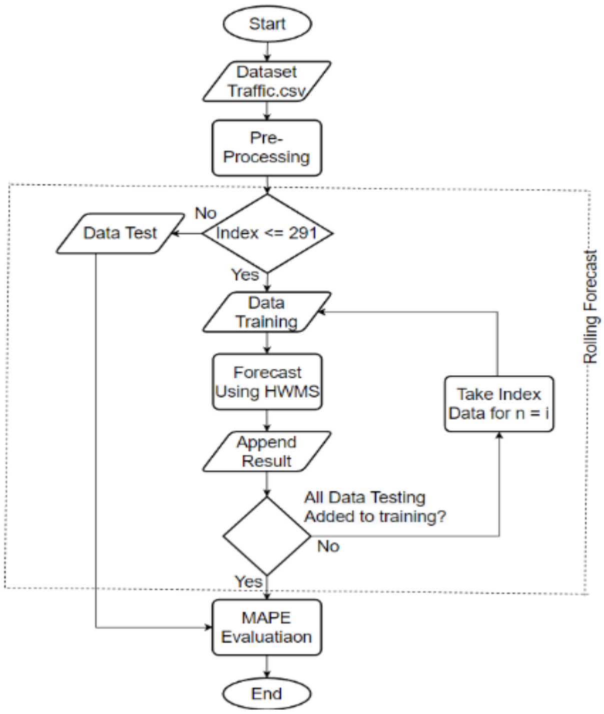

Figure. 2 HWMS and rolling forecast flow diagram

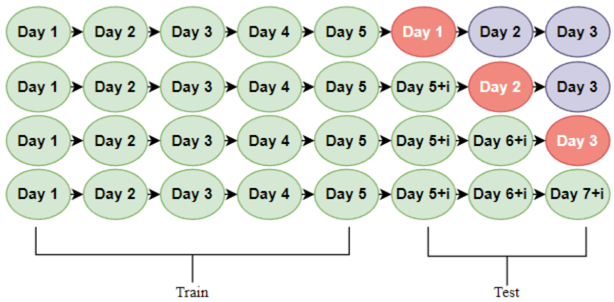

Figure. 3 Flow of data test to data train in subsequent iterations

Step 3. Conducting forecasting using HWMS in the first iteration using Eq. (6). The results of the first iteration forecasting are then stored in a data frame.

Step 4. After the HWMS performs the first iteration of forecasting, the system will update the data train by taking the data at index 0 from the data test and then performing observation again. Each time a forecast is made, the system will provide an updated data train from the data test used for error evaluation. As a result, the subsequent estimates will always be dynamic and include the most up-to-date actual data from the entire procedure. The illustration of the process of transitioning from data testing to data train can be seen in Fig. 3.

The flow from left to right represents the entire 100% data. The green ellipse represents the data train, followed by a red ellipse representing the index 0 of the data test used for error evaluation. The purple ellipses represent the remaining data test indices that have not been used for testing. The image's second row and subsequent rows represent the n-th iteration, where the red ellipse represents the data test changes to the data train. This process repeats until all the grey ellipses representing the data test are used. This means that, in this case, the iteration is performed for the entire length of the data test period, which is 73. This repetitive process is called rolling forecasting [16, 18].

Step 5. Similar to the methods explained above, the evaluation of HWMS and rolling forecast accuracy also uses MAPE.

3.4 Testing and evaluation

First, the system will be tested for its functionality to determine whether it can forecast LTE traffic volume for future periods based on the available dataset. This functionality test will check whether the system can take input data, run the implemented forecasting methods, and produce forecasting results that can be used to make predictions of LTE traffic volume in the future.

Second, after the system has been tested for its functionality, mean absolute percentage error (MAPE) will be calculated to measure the accuracy of the forecasting system for each of the 56 cells. MAPE (mean absolute percentage error) is commonly used as a performance metric in forecasting and predictive analysis due to several reasons [19]:

- Relative measure: MAPE provides a relative measure of forecast accuracy, allowing for the comparison of forecasting models or methods.
- Interpretability: MAPE is expressed as a percentage, making it easily interpretable for decision-makers.
- Scale-invariant: MAPE is not influenced by the magnitude of data, making it applicable to datasets with varying scales.
- Wide usage: MAPE has been widely adopted as a forecasting performance metric across industries and disciplines, enabling comparisons, benchmarking, and insights from past forecasting performance.

From the calculation of MAPE, the number of cells that can be forecasted by each method will be grouped into categories such as "Very Good," "Good," "Reasonable," and "Poor" based on the evaluation results. These categories will be determined based on predetermined thresholds or cutoffs to assess the accuracy level of the forecasting system [32]. A description of the criteria categories used in the calculation of MAPE is shown in Table 1:

Table 1. Category Table MAPE

Range MAPE %	**Category**
< 10	Very Good
10 - 20	Good
20 - 50	Reasonable
> 50	Poor

The forecasting system will be categorized as "Very Good" if the error rate of its forecasts is less than 10%. This means the system can make forecasts very close to the actual data. However, a system that does not fall into the "Very Good" category can still be used if the error rate of its forecasts is between 10% to 20% and falls into the "Good" category.

3.5 Determination of cells requiring solutions

Cell congestion is typically caused by fully utilized physical resource blocks (PRBs) reaching 100% capacity. The increase in PRBs results from the growing number of active users, reflected in the increased traffic volume in that cell. This congestion decreases the average throughput per user, resulting in suboptimal network service quality for that cell [12]. Consequently, the company may face potential revenue losses due to poor internet quality. To address this issue, the company needs to provide solutions such as redesigning the network or adding new cells to serve the area more effectively.

The understanding that the percentage increase in traffic volume from the forecasting results is directly proportional to the rise in PRBs makes it a growth factor for determining which cells in which areas require solutions. This growth factor can be calculated using Eq. (1):

$$Growth\ Factor = \frac{A-B}{B} \times 100 \times P \quad (11)$$

Where:
A: Results of traffic volume forecasting (MB)
B: Traffic volume baseline (MB)
P: Physical resource block baseline (%)

Growth factors are calculated from the increase in the forecasting results of traffic volume compared to the baseline volume multiplied by the baseline PRB. This yields a percentage increase in PRB for each cell. The company has a threshold to determine the maximum percentage of load each cell can handle to

Received: June 17, 2023. Revised: August 21, 2023.

minimize losses. This threshold depends on the bandwidth of each cell. For example, an LTE 800MHz cell with a bandwidth of 10 MHz is considered capable of handling the load well if PRB usage is below 60%, while for an LTE 1800 cell with a bandwidth of 20 MHz, the threshold is less than 80% [12]. This is the reason why we added additional criteria to our internal dataset.

4. Results and discussion

Author names are to be centered beneath the title and printed in Times New Roman 11-point, boldface type. Author affiliations should be centered beneath author names and printed in Times New Roman 11-point, non-boldface type. Multiple authors may be shown in a two- or three- column format, with their affiliations italicized and centered below their respective names. Include addresses and the post code if possible. The corresponding author's email address should be centered in Times New Roman 10.5-point as shown here.

4.1 Methods comparison

Forecasting using the holt-winter's multiplicative seasonal and rolling forecast methods and other comparison methods implemented in Python programming to generate forecasts for LTE traffic volume 73 days ahead. In previous research, the accuracy of the forecasts was evaluated using root mean square error (RMSE), which only measures the error of the forecasts against the actual test data in the original unit of the observed data. This makes it difficult for telecommunications companies to assess the suitability of the forecasts or define their own suitability standards. Using mean absolute percentage error (MAPE) as the evaluation metric provides a solution, as it calculates the error in percentage, making it easier to understand and interpret on graph.

The calculation of MAPE has been conducted on 56 cells that were forecasted using all the investigated methods. This is the result of MAPE calculation for each method in the graph.

a. ARIMA

The accuracy of ARIMA is not too fluctuating for each cell, is depicted in Fig. 4.

The lowest prediction error is observed in cell_55 at 15.52%, while the highest error is found in cell_8 at 59.49%. The average MAPE for the total cells is 31.52%. Although there is one cell, cell_46 that failed to make accurate predictions. The ARIMA parameters used for each cell are different and have been automatically optimized using the pmdarima library in Python.

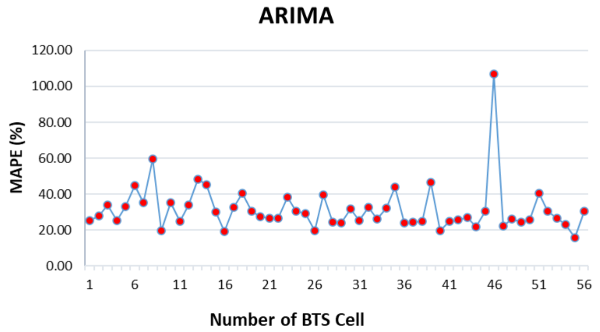

Figure. 4 ARIMA result

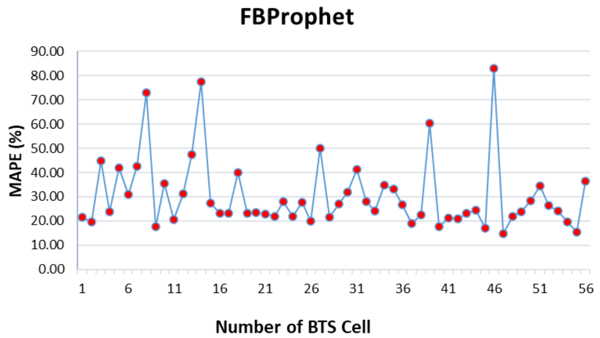

Figure. 5 FbProphet result

b. FbProphet

FbProphet has different accuracy compared to ARIMA in terms of percentage and patterns. FbProphet tends to exhibit significant fluctuation in accuracy across individual cells. However, when considering the overall average accuracy for all cells, it is reported to be approximately 30.45%, which is lower than that of ARIMA. The detailed accuracy of FbProphet can be seen in Fig. 5.

The lowest prediction error is observed in cell_55 at 15.50%, while the highest error is found in cell_46 at 83.16%. It is important to note that the FbProphet implementation we utilized has been automatically optimized using the Python library.

c. HWMS

HWMS exhibits a pattern similar to ARIMA, with relatively low fluctuation among cells. However, it should be noted that HWMS also failed to accurately predict the values for cell_46. Despite this, the average MAPE value for HWMS is approximately 30.06%, which is lower than both ARIMA and FbProphet. For a more detailed analysis of HWMS accuracy, please refer to Fig. 6:

The similarity between HWMS and ARIMA suggests that they share common characteristics in capturing and modeling the underlying patterns in the data. However, the specific reasons for HWMS's

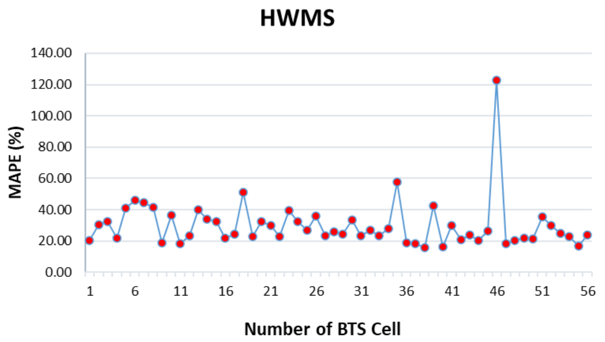

Figure. 6 HWMS result

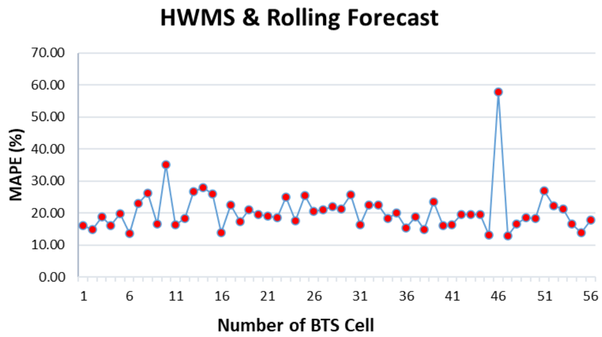

Figure. 7 HWMS and rolling forecast result

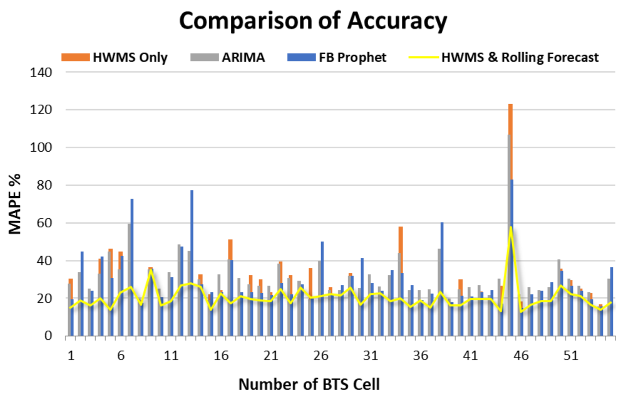

Figure. 8 Comparison of accuracy

failure to predict accurately for cell_46 should be further investigated.

d. HWMS and rolling forecast

In Fig. 7 below, it is evident that both HWMS and rolling forecast have a significant number of prediction errors below 20%. The average MAPE percentage for the total of all cells is approximately 20.47%. This indicates that the overall accuracy of HWMS and rolling forecast is relatively high, with a majority of predictions exhibiting a low level of error.

The ability of HWMS and rolling forecast to achieve such low MAPE values is a positive indication of their effectiveness in forecasting network traffic. The utilization of rolling forecasting, which incorporates the most up-to-date data for each iteration, allows for continuous adaptation to changing patterns and dynamics in the network traffic.

Table 2. Accuracy ranking based on prediction errors

No	Methods	MAPE %
1	HWMS & Rolling Forecast	20.47
2	HWMS Only	30.06
3	FB Prophet	30.45
4	ARIMA	31.52

e. Comparative analysis

The calculation results show that not all methods provide satisfactory percentage errors in forecasting as in Fig. 8.

HWMS and rolling forecast, represented by the yellow lines, consistently exhibit lower MAPE percentages compared to the other methods. In that graph it can be observed that the accuracy of HWMS using rolling forecast is generally below 20% in many cells. In cell_10, all methods show relatively similar MAPE percentages, suggesting comparable levels of accuracy. However, in cell _8, cell_14, cell_35, and cell_39, significant differences in MAPE values can be observed.

If sorted from the largest MAPE percentage to the smallest, the average accuracy of each tested method on 56 cells can be seen in Table 2.

HWMS with rolling forecast ranks at the top position, while ARIMA ranks at the bottom position. The sequence of prediction errors from the lowest to the highest is as follows: HWMS & rolling forecast (20.47%), HWMS only (30.06%), FB prophet (30.45%), and ARIMA (31.52%).

The calculated MAPE values vary due to the distinct time series patterns exhibited by each cell. These variations can be influenced by the location of the cell, such as being situated in a market, residential area, or office complex. The variations in MAPE percentages among the different methods reflect their respective strengths and weaknesses in capturing the unique patterns and dynamics of each cell. This emphasizes the importance of evaluating and comparing multiple forecasting methods to determine the most effective approach for different cell characteristics present in the dataset.

When grouped according to MAPE categories, the number of cells for each method can be seen in Table 3.

The table of MAPE calculation results is used as a reference to determine which category the system

Table 3. Number of cells within the MAPE calculation range

Methods	The number of cells in each MAPE category			
	Very Good	Good	Reason-able	Poor
ARIMA	0	5	49	2
FB Prophet	0	9	43	4
HWMS	0	8	45	3
HWMS & Rolling Forecast	0	33	22	1

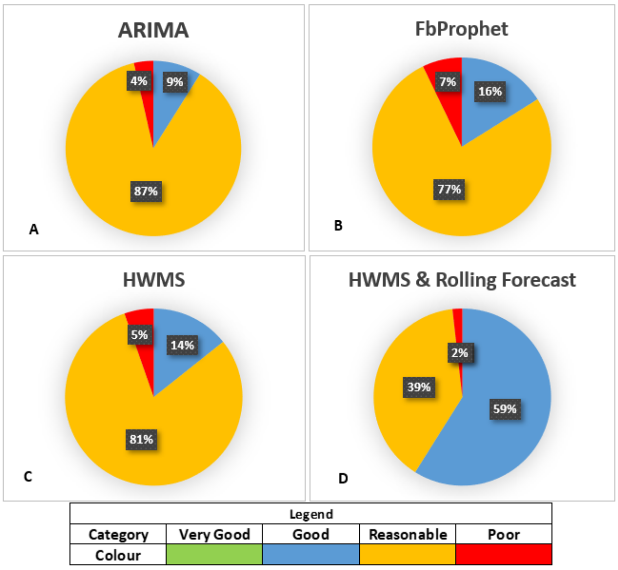

Figure. 9 Graph of portion of feasibility test method using MAPE

using HWMS and rolling forecast falls into for each cell's forecasting. Each cell has its own unique characteristics that affect how well a method can observe the available data and demonstrate the system's suitability.

ARIMA has the lowest number of cells categorized as "Good."FbProphet has the highest number of cells categorized as "Poor," with a total of 4 cells, followed by HWMS without rolling forecast with three cells, ARIMA with two cells, and HWMS and rolling forecast with the least number of cells, only one cell. The performance of HWMS without rolling forecast is not significantly different from FbProphet, with only a difference of 1 cell in the "Good" category. Overall, HWMS and rolling forecast perform relatively better than other methods. The order of the highest number of cells categorized as "Good" from highest to lowest is HWMS and rolling forecast with 33 cells, followed by FbProphet with nine cells, HWMS without rolling forecast with eight cells, and ARIMA with five cells.

Although overall, each method falls within the acceptable threshold for application as they have the majority of MAPE percentages categorized as "Reasonable." The proportion of feasibility based on MAPE categories for each method can be visualized in Fig. 9:

Fig. 9 represents the proportion of system accuracy using the HWMS and rolling forecast methods compared to other methods. None of the researched methods fall under the "Very Good" category. However, in the "Good" category, HWMS and rolling forecast have the highest percentages. HWMS and rolling forecast are able to forecast 59% or 33 cells out of the total 56 cells available. When compared in this category, the system using HWMS and rolling forecast has 45% more cells than the HWMS method without rolling forecast. When compared to ARIMA and FbProphet, HWMS and rolling forecast have a difference of 50% and 37%, respectively.

The finding show that the HWMS and rolling forecast methods perform better in terms of accuracy compared to other methods. These methods are able to forecast a larger number of cells in the "Good" category and have the highest ranking of accuracy when viewed from the average error among the methods tested. The significant difference observed when comparing HWMS with and without rolling forecast emphasizes the effectiveness of incorporating rolling forecast in improving the accuracy of forecasts. Furthermore, the HWMS and rolling forecast methods demonstrate higher accuracy compared to the ARIMA and FbProphet methods. This indicates that the HWMS and rolling forecast methods have a stronger ability to capture the complex patterns and dynamics present in network traffic datasets, resulting in more precise forecasts.

4.2 Implementation of proposed method in company

The forecasting results play a crucial role in determining the number of new cells required to meet the network's demand. The increase in traffic volume reflects the growing demand for internet usage, and this growth factor is utilized to estimate the required magnitude of physical resource blocks (PRB). The calculated growth factor is then added to the PRB baseline of each cell, serving as a reference to identify cells that necessitate solutions.

Due to the limited availability of bandwidth information in the dataset, which does not provide data for each cell, a decision is made based on a predetermined criterion. A cell is considered to

Received: June 17, 2023. Revised: August 21, 2023.

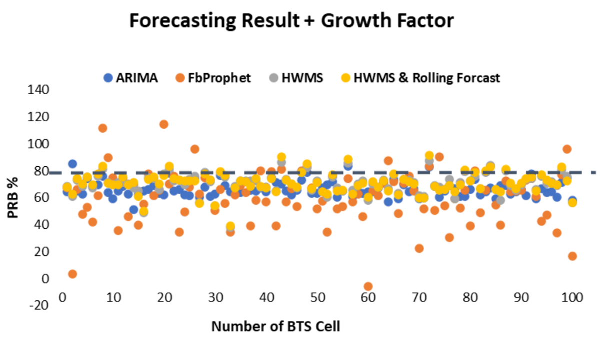

Figure. 10 Graph of forecasting result + growth factor

require a solution when it exceeds the maximum threshold set for LTE 1800 cells with a 20MHz bandwidth. Specifically, if the PRB utilization reaches or exceeds 80%, it is deemed to exceed the safe threshold [12].

Fig. 10 below illustrates a plot depicting the cells that surpass the 80% PRB threshold, providing visual insight into the specific cells requiring attention.The forecasting results each cell for the next 92 days using the internal dataset of 100 cells and the added growth factor will identify which cells require solutions.

Fig. 10 shows that the points above the dashed line represent cells that require solutions. The forecasting using the HWMS method and rolling forecast identified 11 cells that require solutions. On the other hand, using the HWMS method without rolling forecast identified 9 cells, ARIMA identified only 3 cells, and FbProphet identified 12 cells. The comparative analysis of these calculations indicates that the company may still face bottleneck issues if it relies solely on the forecasting results from the HWMS method without Rolling Forecast or ARIMA, even after expanding the locations that require solutions.

Furthermore, utilizing the calculation results from the FbProphet method has the potential for capital expenditure (CAPEX) wastage. By using the forecasting results from the HWMS method and rolling forecast, the company has the opportunity to save on implementing one new LTE location.

To mitigate bottleneck issues and optimize forecasting accuracy, it is recommended for the company to utilize the HWMS method with rolling forecast. By doing so, the company can address the identified cell issues more effectively. Additionally, the cost savings from avoiding unnecessary CAPEX expenditures can contribute to improved financial efficiency.

5. Conclusions

The forecasting of LTE traffic volume has been conducted in Indonesia using holt-winter's multiplicative seasonal (HWMS) method and rolling forecast. The results show that HWMS and rolling forecast can forecast LTE traffic volume with the "Good" MAPE category for 33 cells, "Acceptable" for 22 cells, and "Poor" for one cell. The MAPE evaluation for HWMS and rolling forecast is lower than the three other methods. Furthermore, HWMS and rolling forecast have more cells by 45% compared to HWMS without rolling forecast in the "Good" category. The number of HWMS and rolling forecast cells in this category is also higher by 50% compared to ARIMA and 37% compared to FbProphet. Based on these findings, the forecasting system using HWMS and rolling forecast methods has demonstrated their effectiveness in enhancing the accuracy of network traffic forecasting. This system is expected to assist telecommunication companies in obtaining relevant LTE traffic volume forecast data. The results of this LTE traffic forecasting analysis serve as a reference for determining network design to address bottleneck issues. They are used for CAPEX calculation in implementing cost-benefit analysis for telecommunication companies.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this research article.

Author contributions

Conceptualization, Endi Rizal Ferdiansyah and Adian Fatchur Rochim; methodology, Endi Rizal Ferdiansyah, Adian Fatchur Rochim and Wahyul Amien Syafei; software, Endi Rizal Ferdiansyah; validation, Endi Rizal Ferdiansyah, Adian Fatchur Rochim; formal analysis, Endi Rizal Ferdiansyah; investigation, Endi Rizal Ferdiansyah and Wahyul Amien Syafei; resources, Endi Rizal Ferdiansyah; data curation, Endi Rizal Ferdiansyah; writing—original draft preparation, Endi Rizal Ferdiansyah; writing—review and editing, Endi Rizal Ferdiansyah, Adian Fatchur Rochim, and Wahyul Amien Syafei; visualization, Endi Rizal Ferdiansyah; supervision, Adian Fatchur Rochim and Wahyul Amien Syafei;

References

[1] X. Zheng, W. Lai, H. Chen, S. Fang, and Z. Li, "A study of cellular traffic data prediction by kernel ELM with parameter optimization", *Applied Sciences (Switzerland)*, Vol. 10, No. 10, pp. 1–15, 2020.

[2] J. Singh and D. Kedia, "Improved precoding algorithm design for downlink large scale MU-

Received: June 17, 2023. Revised: August 21, 2023.

MIMO system", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 3, pp. 143–153, 2020, doi: 10.22266/ijies2020.0630.14.

[3] V. S. Kumar and A. Muthukumaravel, "Seasonal forecasting of mobile data traffic in GSM networks with linear trend", *Journal of Applied Science and Engineering*, Vol. 23, No. 3, pp. 469–474, 2020.

[4] F. W. Alsaade and M. H. A. Adhaileh, "Cellular Traffic Prediction Based on an Intelligent Model", *Mobile Information Systems*, Vol. 2021, pp. 1–15, 2021.

[5] T. H. H. Aldhyani, M. Alrasheedi, A. A. Alqarni, M. Y. Alzahrani, and A. M. Bamhdi, "Intelligent Hybrid Model to Enhance Time Series Models for Predicting Network Traffic", *IEEE Access*, Vol. 8, pp. 130431–130451, 2020.

[6] H. D. Trinh, L. Giupponi, and P. Dini, "Mobile Traffic Prediction from Raw Data Using LSTM Networks", In: *Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Bologna, Italy, pp. 1827–1832, 2018.

[7] C. J. Sagar, H. Gupta, J. S. Jadon, N. Arora, and S. Kumar, "Prediction of traffic volume in BTS sites using deep learning", In: *Proc. of 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)*, Noida, India, pp. 587–590, 2020.

[8] S. Mishra and N. Mathur, "Load Balancing Optimization in LTE/LTE-A Cellular Networks: A Review", *arXiv*: 1412.7273, Vol. 1, pp. 1-7, 2014, [Online]. Available: http://arxiv.org/abs/1412.7273.

[9] Q. T. Tran, L. Hao, and Q. K. Trinh, "Cellular Network Traffic Prediction Using Exponential Smoothing Methods", *Journal of Information and Communication Technology*, Vol. 1, No. 1, pp. 1–18, 2019.

[10] Q. T. Tran, L. Hao, and Q. K. Trinh, "A comprehensive research on exponential smoothing methods in modeling and forecasting cellular traffic", *Concurrency and Computation: Practice and Experience*, Vol. 32, No. 23, pp. 1–11, 2020.

[11] A. S. Arifin and M. I. Habibie, "The prediction of mobile data traffic based on the arima model and disruptive formula in industry 4.0: A case study in Jakarta, Indonesia", *Telkomnika (Telecommunication Computing Electronics and Control)*, Vol. 18, No. 2, pp. 907–918, 2020.

[12] D. Chmieliauskas and D. Gursnys, "LTE Cell Traffic Grow and Congestion Forecasting", In: *2019 Open Conference of Electrical, Electronic and Information Sciences, eStream 2019 - Proceedings*, Vilnius, Lithuania, pp. 1–5, 2019.

[13] R. F. Syam and A. S. Girsang, "Bandwidth provisioning for 4g mobile network using arima based on traffic forecasting", *International Journal of Emerging Trends in Engineering Research*, Vol. 8, No. 9, pp. 5869–5873, 2020.

[14] M. Xu, Q. Wang, and Q. Lin, "Hybrid holiday traffic predictions in cellular networks", In: *Proc. of IEEE/IFIP Network Operations and Management Symposium: Cognitive Management in a Cyber World, NOMS 2018*, Taipei, Taiwan, pp. 1–6, 2018.

[15] V. A. S. Makridakis and E. Spiliotis, "Statistical and Machine Learning forecasting methods: Concerns and ways forward", *PLoS One*, Vol. 13, No. 3, pp. 1–26, 2018, [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0194889.

[16] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. 2018.

[17] Q. Li, T. Bedi, C. U. Lehmann, G. Xiao, and Y. Xie, "Evaluating short-term forecasting of COVID-19 cases among different epidemiological models under a Bayesian framework", *Gigascience*, Vol. 10, No. 2, pp. 1–11, 2021.

[18] A. I. Almazrouee, A. M. Almeshal, and A. S. Almutairi, "Long-Term Forecasting of Electrical Loads in Kuwait", *Applied Science*, Vol. 10, No. 5627, pp. 1–17, 2020.

[19] 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, 2006.

[20] H. Yang, X. Li, W. Qiang, Y. Zhao, W. Zhang, and C. Tang, "A network traffic forecasting method based on SA optimized ARIMA–BP neural network", *Computer Networks*, Vol. 193, No. 108102, pp. 1–12, 2021.

[21] S. Ariyanti, "Comparison of Network Cost & Feasibility LTE Technology to Support Broadband Plan in Indonesia", *Buletin Pos dan Telekomunikasi*, Vol. 17, No. 1, pp. 1–16, 2019.

[22] I. Lohrasbinasab, A. Shahraki, A. Taherkordi, and A. Delia Jurcut, "From statistical- to machine learning-based network traffic prediction", *Transactions on Emerging Telecommunications Technologies*, Vol. 33, No. 4, pp. 1–20, 2022.

[23] T. H. H. Aldhyani and M. R. Joshi, "Intelligent Time Series Model to Predict Bandwidth Utilization", *International Journal of Computer Science and Applications*, Vol. 14, No. 2, pp. 130–141, 2017.

[24] E. P. L. Shiang, W. C. Chien, C. F. Lai, and H. C. Chao, "Gated Recurrent Unit Network-based Cellular Trafile Prediction", In: *Proc. of International Conference on Information Networking*, Barcelona, Spain, pp. 471–476, 2020.

[25] P. Chakraborty, M. Corici, and T. Magedanz, "A comparative study for Time Series Forecasting within software 5G networks", In: *Proc. of 2020 14th International Conference on Signal Processing and Communication Systems*, Adelaide, Australia, pp. 1–7, 2020.

[26] W. H. Rankothge, N. D. U. Gamage, H. G. H. Dewwiman, M. M. T. R. Ariyawansa, S. A. A. Suhail, and M. D. B. P. Senevirathne, "Network Traffic Prediction for a Software Defined Network based Virtualized Network Functions Platform", In: *Proc. of 2021 6th IEEE International Conference on Recent Advances and Innovations in Engineering, ICRAIE 2021*, Kedah, Malaysia, pp. 1–4, 2022.

[27] M. A. R. Shuvo, M. Zubair, A. T. Purnota, S. Hossain, and M. I. Hossain, "Traffic Forecasting using Time-Series Analysis", In: *2021 Proc. of the 6th International Conference on Inventive Computation Technologies (ICICT)*, Dhaka, Bangladesh, pp. 269–274, 2021.

[28] E. O. Kumih, M. A. Kissiedu, and E. A. Aboagye, "Historical Analysis and Time Series Forecasting of Internet Data Usage and Revenues in Ghana Using Machine Learning Based Facebook Prophet Model", *SSRN Electronic Journal*, 2022, [Online]. Available: doi: 10.2139/ssrn.4290081.

[29] P. Cembaluk and J. Aniszewski, "Forecasting the network traffic with PROPHET", In: *Proc. of 3rd Polish Conference on Artificial Intelligence*, Gdynia, Poland, pp. 215–218, 2022.

[30] S. J. Taylor and B. Letham, "Forecasting at Scale", *American Statistician*, Vol. 72, No. 1, pp. 37–45, 2018.

[31] U. Khair, H. Fahmi, S. A. Hakim, and R. Rahim, "Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error", *Journal of Physics: Conference Series*, pp. 1–6, 2017.

[32] R. K. Klimberg, G. P. Sillup, K. J. Boyle, and V. Tavva, "Forecasting performance measures - What are their practical meaning?", *Advances in Business and Management Forecasting*, Vol. 7, pp. 137–147, 2010.