ISSN 2347 - 3983
Volume 8. No. 9, September 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter158892020.pdf
https://doi.org/10.30534/ijeter/2020/158892020

Bandwidth Provisioning for 4G Mobile Network Using ARIMA Based on Traffic Forecasting

Rahmat Fuadi Syam, Abba Suganda Girsang
Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, rahmat.fuady@binus.ac.id
Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, agirsang@binus.edu

ABSTRACT

Increasing the efficiency of telecommunications companies is currently very much needed to maintain the sustainability of the company, especially data transfer from each location. The problem that is often experienced is the lack of management of the amount of bandwidth needed so that there is often a shortage or excess of bandwidth allocation at each BTS site so that this problem can reduce the profits obtained by the company. From these problems, we need a system that can regulate and predict future bandwidth requirements. In this study, bandwidth requirements will be simulated using the Autoregressive Integrated Moving Average method, the ARIMA method using time series data from November 2019 to January 2020. After trying out several ARIMA models (1.0.0), (1.1.0), (1.0.1), (0.0.1) and (0.0.6), it was found that the ARIMA Model (0.0.6) has the best level of accuracy with an RMSE value of 419,139

Key words: Arima, Timeseries, Bandwith, Payload.

1. INTRODUCTION

In the field of technology, especially communication, the role of telecommunications companies is critical in maintaining the quality of the network that is owned where it requires a sufficient network planning foundation [1],[2], where sometimes the provision of bandwidth resources based on the amount of traffic measured at peak times is inefficient in terms of resource utilization [3], thus encouraging the use of technology, especially data mining processing. The process of searching for patterns in groups or called data mining [4]. In telecommunication companies, the high demand for mobile broadband services has made pre-deployment forecasting a significant step [5]. Supported by the amount of data that is owned is very large, but the amount of data is still not appropriately utilized, the data that the company has significantly varied, including having a period of time from a few seconds to minutes which can be used to monitor or as an alert when there is a disturbance on-site, while data that has a time period of time can be useful for engineering network traffic where telecommunications operators can change the Network flow from a busy line to a new line. Currently, the development of prediction techniques using time series data can be a consideration in solving problems in network planning to be more effective, but making predictions requires fulfilling some more precise prediction requirements and the desired time interval control [6]. The development of traffic prediction methods can be found that most of the techniques have introduced black-box modeling and structural modeling to solve prediction problems where the approach is based on the traditional time series prediction technique, called the Box-Jenkins approach. [7]. The ARIMA model has a better performance in predictions [8], because changes in the process of forecasts are measured consistent, long-term forecasts tend to have a less good level of accuracy because it tends to be flat for a reasonably long period. In the Integrated Moving Average (ARIMA) Autoregressive model wherein this model completely ignores the independent variables in making predictions [9], in this paper ARIMA is used to analyze time-series data where previously the data check is stationary or not stationary which is when the data is not stationary, the difference process will be carried out to stationary, then identify the values of p and q where these values are obtained by checking the autocorrelation function (ACF) and partial autocorrelation function (PACF) values of the time series before modeling with the Autoregressive Integrated Moving Average (ARIMA) method, in general, the ACF and PACF values can be identified based on Table 1

Table 1 :ACF and PACF identification

Model	**ACF**	**PACF**
AR (p)	Dies down	Cut-off after lag p
MA (q)	Cut Off after lag q	Dies down
ARMA (p,q)	Dies Down	Dies Down
AR(p) or MA(q)	Cut Off after lag q	Cut off after lag p

Source [10]

2. PREDICTION MODELS

2.1 Time Series

Time series is a sequence of observations obtained from observing one object over several time periods. Time series data which are analyzed according to their function can be modeled using the analytical method. Time series time is as follows [11].

A. The AR (Auto-Regressive) Model

In the Auto-Regressive Model, it can be seen from each observation within a time bracket which is connected in a consistent way and can be identified for one or more previous statements from the same series, and this model can be seen as Eq. (1)

$$Z_t' = \phi_1 Z'_{t-1} + \phi_2 Z'_{t-2} + \dots + \phi_p Z'_{t-p} + a_t \quad (1)$$

Where the AR model is symbolized by p with the parameter that shows the AR model is ϕ. n the AR process ACF will decrease following the exponential or sine waveform while in PACF the MA process will cut off after lag p can be seen with the structure $\phi_{kk} = 0$ where k > p

B. The MA (Moving Average) Model

The Moving Average model is the average form of the seasonal or cycles components in an adjusted time period which functions to smooth the original time series with the intermediate rolling subset of the original series elements [11], the MA model with the order q, where the parameters that show the model MA is ϕ. can be seen as Eq.(2)

$$Z_t = a_t - \phi_1 a_{t-1} - \phi_2 a_{t-2} - \dots - \phi_q a_{t-q} \quad (2)$$

The ACF structure in the MA model is $p_k = 0$ where k> q or cut off after laq q whereas, in PACF, it is dominated by a linear combination of the exponential form of a damped sine wave.

C. The ARMA**(**Auto Regressive Moving Average) Model

In the ARMA model which is a mixture of AR and MA models where AR is the p-value and MA is the q value so in general, the form of the equation of the mixed model AR and MA is shown in Eq. (3)

$$\phi_p(B)^d \; Z_t = \theta_q(B)\, a_1 \quad (3)$$

Where in the ARMA process (p, q) ACF is down after the lag to (q-p) and PACF will also decrease in lag (q-p)

C. The ARIMA(Auto-Regressive Integrated Moving Average) Model

In the ARIMA process, there are two approaches in modeling where if the series of objects is different and the time in achieving stationarity can be modeled with ARIMA (p, d, q) where I indicate integrated, which can be concluded that the ARIMA model is the most common in predicting time series which can be stationary with transformations such as differencing and logging [11] so that it can be modelled as Eq. (4)

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)\, a_t \quad (4)$$

3 METHOD

3.1 Forecasting framework

In the forecasting process, which significantly affects the time series data pattern generated, it requires an initial observation in making forecasts where this process will analyze the type of data considering that each statistical method has a different working phase, in Figure 1 shows the steps taken in the forecasting process

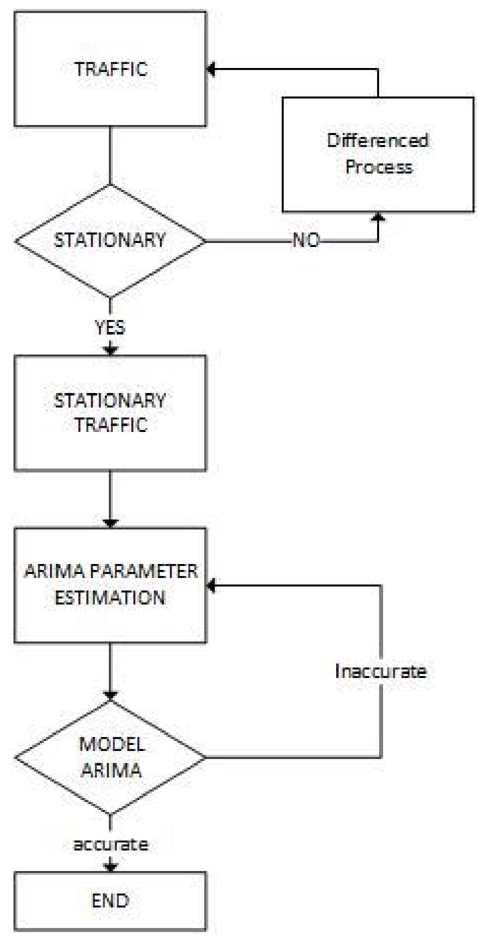

Figure 1: Process of ARIMA

Where the following steps [3]: first, it is to validate the data using a time series histogram. Stationary tests use the ADF unit root, which is a method for performing unit-roots testing [12]. Second, is when the data is not stationary, a differencing process is carried out until the data becomes stationary.
Third, modeling based on the analysis of ACF and PACF so that the p, d, q values are obtained. Fourth, then the

parameters that have been received are diagnostic with the RMSE evaluation so that the best RMSE value is obtained. And Fifth makes predictions using the ARIMA parameters and models that have been obtained.

3.2 Predicted Performance

In choosing the best model, we need an appropriate performance indicator in measuring the quality of the method or model, which can be assumed that the value of the original data in $y_{t,}$ and the value of the prediction results are $\hat{y}t.$

This paper uses two evaluation indicators, namely the MSE and RMSE values, which the equation can be as seen Eq. (5) and Eq. (6) [13], [14]

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2 \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{6}$$

MSE is very good at giving an idea of how consistent the model is built. By minimizing the MSE value, it means minimizing the model variance. Models that have small variants are able to provide relatively more consistent results for all input data compared to models with large variants, while RMSE is an alternative method for evaluating forecasting techniques used to measure the level of accuracy of the forecast results of a model.

4. BANDWITH FORECASTING

4.1 Data Source

The data used in this research is traffic from cells at the base transceiver station where data is taken from 5-11-2019 to 30-12-2019 with a time interval of 1 hour where the attribute data is shown in Table 2 and Figure 2.

Table 2: Data Attribute

No	Date	Cel l	Payload
1	5/11/2019 0:00	11	516.1843
2	5/11/2019 1:00	11	57.2152
3	5/11/2019 2:00	11	102.6939
4	………………..	…	……………
5	30/12/2019	11	1634.78

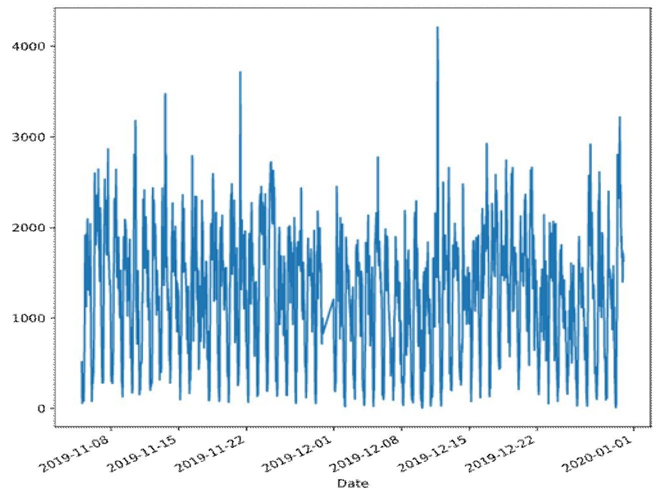

Figure 2: Traffic Cell 11

4.2 Experimental results and analysis

The first step is to validate the data to be used where at this stage it can be seen in Figure 3 that the histogram shows right tilt or positive tilt, this is because the tail is from the distribution point to the right and because it has a skewness value greater than 0 (positive)

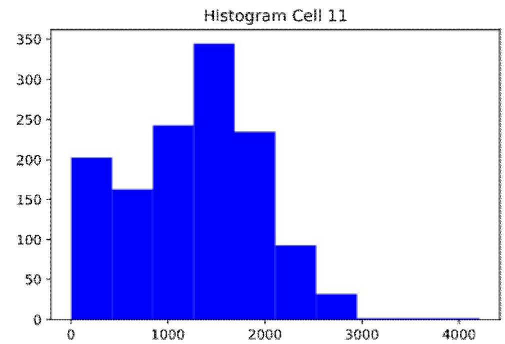

Figure 3: Histogram Traffic Cell 11

After knowing the data to be used shows positive direction, then the process of checking stationary data is carried out, in Table 3 shows that the data to be used is stationary where the p-value is less than 0.5 and the Critical value is 1% smaller than the ADF value so that the data can be concluded, which will be used stationary.

Table 3: Augmented Dickey-Fuller test

No	Variable		Value
1	ADF Statistic		-13.83748 7
2	P-Value		0.000000
3	Critical Value		
		1 %	-3.435
		5 %	-2.864
		10%	-2.568

Next, perform the analysis with Auto Correlation and Partial Auto Correlation Function, Auto-Correlation Function (ACF) as seen Figure 4calculates the correlation between data points in a series and data points in lags before data points. This information then helps in calculating the PACF (partial autocorrelation function) [15] as seen Figure 5.

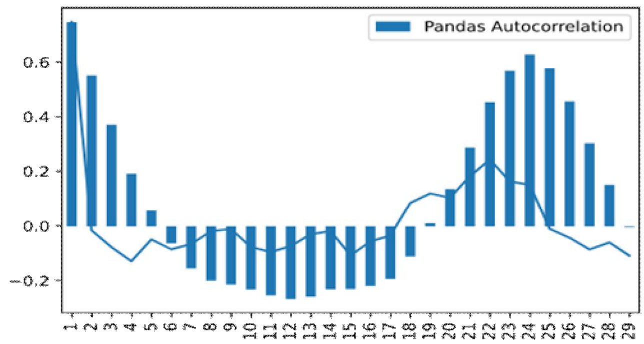

Figure 4: Auto Correlation Function

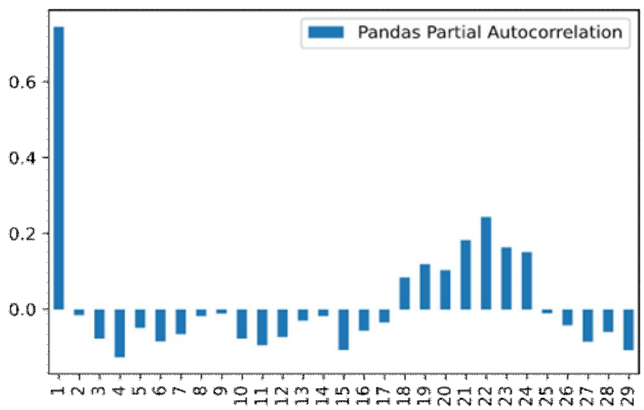

Figure 5: Partial Auto Correlation Function

In the last stage, testing with ARIMA, the previous data model is divided into 2, namely training data and testing data with a composition of 70% training data and 30% testing data, then testing the parameters p, d, q of the ARIMA model (1.0.0), (1.1.0), (1.0.1), (0.0.1) and (0.0.6) and the results show the ARIMA / MA model (0.0.6) based on table 4 shows the best value of the forecast. It can be seen on Figure 6 and Figure 7

Table 4: Comparison Model ARIMA

NO	AR	d	MA	MSE	RMSE
1	1	0	0	183603.937	428.490
2	0	0	1	254943.558	504.919
2	1	1	0	205939.640	453.805
3	1	0	1	183741.851	428.651
4	2	0	0	187357.031	432.847
5	0	0	6	175678.158	419.139

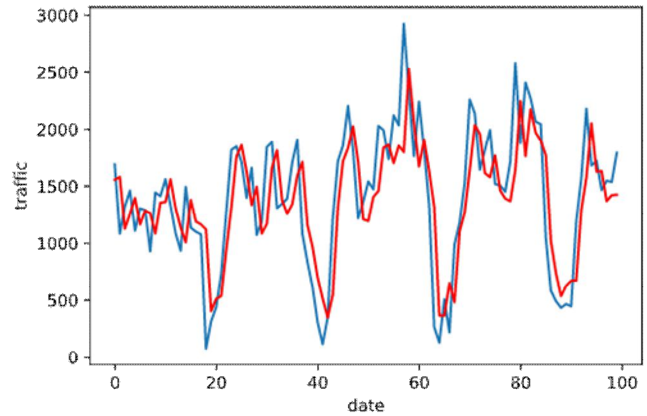

Figure 6: Arima (0.0.6) plot

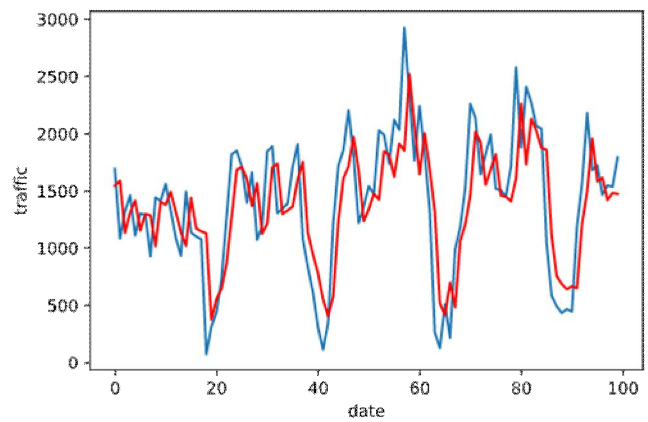

Figure 7: Arima (1.0.0) plot

5. CONCLUSION

This study uses the ARIMA method to predict the amount of traffic usage on the cell base transceiver station to identify the forecasting results of the ARIMA model using the MSE, RMSE method, where the test results show ARIMA (0.0.6) has a better value than other models. With an RMSE value of 419.139, which from the plot of the image offers the accuracy of the forecasting results close to the actual data, so that we will easily predict the network based on historical network traffic data on the cell transceiver station. But in the future, by collecting more data from historical data, it can further improve the accuracy of predict which many factors can affect the ups and downs of bandwidth requirements, the number of users, bandwidth per user, types of applications, users accessing time properties, etc. Our goal is to examine how these factors affect traffic from the network

REFERENCES

1 Y. Yu, M. Song, Y. Fu, and J. Song. **Traffic prediction in 3G mobile networks based on multifractal exploration**, *Tsinghua Sci. Technol.*, vol. 18, no. 4, pp. 398–405, 2013.

2 V. S. Kumar and A. Muthukumaravel. **Seasonal**

forecasting of mobile data traffic in GSM networks with linear trend, vol. 23, no. 3, pp. 469–474, 2020.

3 H. W. Kim, J. H. Lee, Y. H. Choi, Y. U. Chung, and H. Lee **Dynamic bandwidth provisioning using ARIMA-based traffic forecasting for Mobile WiMAX**, *Comput. Commun.*, vol. 34, no. 1, pp. 99–106, 2011.

4 D. T. Larose and C. D. Larose. **Data Mining and Predictive Analytics (Wiley Series on Methods and Applications in Data Mining)**, *Wiley Ser.*, p. 794, 2015.

5 M. Suryanegara, A. Ramadhan, A. K. Akbar, and M. Asvial. **The forecasting model of 4G LTE implementation in Indonesia**, *ICMIT 2014 - 2014 IEEE Int. Conf. Manag. Innov. Technol.*, no. 1, pp. 461–466, 2014.

6 P. Kuan Hoong. **BitTorrent Network Traffic Forecasting With ARMA**, *Int. J. Comput. Networks Commun.*, vol. 4, no. 4, pp. 143–156, 2012.

7 H. Z. Moayedi and M. A. Masnadi-Shirazi. **Arima model for network traffic prediction and anomaly detection**, *Proc. - Int. Symp. Inf. Technol. 2008, ITSim*, vol. 3, pp. 6–11, 2008.

8 B. Krithikaivasan, K. Deka, and D. Medhi. **Adaptive bandwidth provisioning envelope based on discrete temporal network measurements**, *Proc. - IEEE INFOCOM*, vol. 3, no. C, pp. 1786–1796, 2004.

9 S. Siami-namini and N. Tavakoli. **A Comparison of ARIMA and LSTM in Forecasting Time Series**, *2018 17th IEEE Int. Conf. Mach. Learn. Appl.*, pp. 1394–1401, 2018.

10 P. Purnawansyah, H. Haviluddin, R. Alfred, and A. F. O. Gaffar. **Network Traffic Time Series Performance Analysis Using Statistical Methods**, *Knowl. Eng. Data Sci.*, vol. 1, no. 1, p. 1, 2017.

11 S. Jung, C. Kim, and Y. Chung. **A prediction method of network traffic using time series models**, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3982 LNCS, pp. 234–243, 2006

12 Emeka and A. Kelvin. **A detailed document for ARDL**, *J. Stat. Econom. Methods*, vol. 5, no. 3, pp. 63–91, 2016.

13 M. F. Zhani, H. Elbiaze, and F. Kamoun. **Analysis and prediction of real network traffic**, *J. Networks*, vol. 4, no. 9, pp. 855–865, 2009.

14 L. Ni, X. Chen, and Q. Huang. **ARIMA model for traffic flow prediction based on wavelet analysis**, *2nd Int. Conf. Inf. Sci. Eng. ICISE2010 - Proc.*, vol. 0, no. 2, pp. 1028–1031, 2010.

15 A. T. Namel, M. A. Sahib, and S. M. Hasan. **Bandwidth Utilization Prediction in LAN Network Using Time Series Modeling**, *Iraqi J. Comput. Commun. Control Syst. Eng.*, pp. 78–89, 2019.