iJournals: International Journal of Software & Hardware Research in Engineering
ISSN:2347-4890, ijournals.in/ijshre
Volume 12 Issue 9, September 2024

Forecasting the Number of Tuberculosis Patients Visiting Mitra Sehat Clinic With ARIMA Method

Author: Y.D. Sumanto[1]; Susilo Hariyanto[2*]; Auliya Andriyati[3]

Affiliation: Department of Mathematics, Diponegoro University, Semarang 50275, Indonesia[12]; Mitra Sehat Clinic, Sukoharjo 57511, Indonesia[3]

E-mail: susilohariyanto@lecturer.undip.ac.id

DOI: 10.26821/IJSHRE.12.9.2024.120905

ABSTRACT

Tuberculosis (TB) remains a significant public health concern in many parts of the world, including Indonesia. Accurate forecasting of the number of TB patients is crucial for health clinics to prepare resources and optimize patient care. This study aims to forecast the number of tuberculosis patients visiting the Mitra Sehat Clinic using the AutoRegressive Integrated Moving Average (ARIMA) method. The ARIMA model is widely used for time series forecasting due to its ability to capture the dynamics of various types of data. Data on patient visits to the clinic over a specified period were collected and analyzed to develop an appropriate ARIMA model. The steps for using the ARIMA method are using data from the required sample from patients visiting with Tuberculosis diagnosed in January 2021 – August 2024, determining the type of time series data pattern, then conducting a stationarity test, determining the ARIMA model, calculating and analyzing the accuracy of the model used, then forecasting the number of outpatient visits. Based on the calculation, the best ARIMA model for this forecasting is (1, 0, 1) with an error value of 42,0276. The results indicate that the ARIMA model can provide reliable forecasts, assisting healthcare providers in managing resources and improving service delivery for TB patients. This research contributes to the strategic planning of health services and highlights the importance of predictive analytics in healthcare.

Keywords: Arima, Forecasting, Patient Visits, Tuberculosis, Minitab, Sukoharjo Indonesia

1. INTRODUCTION

Tuberculosis (TB) is one of the leading infectious diseases worldwide, causing significant morbidity and mortality, especially in low- and middle-income countries. According to the World Health Organization [1], in 2022, approximately 10 million people were affected by TB globally, with Indonesia being among the top countries with the highest TB burden [1] . Managing and controlling TB effectively requires strategic planning and resource allocation, which depend on accurate predictions of the number of patients seeking medical care.

In the healthcare sector, forecasting plays a critical role in preparing for patient loads, managing clinic resources, and optimizing healthcare services. By predicting the number of patients who will seek treatment, clinics can better allocate staff, medicines, and other resources, thereby improving the quality of patient care. One of the widely recognized statistical techniques for forecasting is the AutoRegressive Integrated Moving Average (ARIMA) method, which has been successfully applied in various health-related time series predictions [2].

The ARIMA model is a robust forecasting tool that captures the temporal dependencies in time series data, making it suitable for predicting disease incidence rates and patient numbers [3]. In previous studies, the ARIMA model has been used to predict the spread of infectious diseases such as influenza [4] and COVID-19 [5], demonstrating its effectiveness in various public health contexts. However, few studies have focused on using ARIMA to forecast patient visits for TB specifically in clinic settings.

The Mitra Sehat Clinic in Sukoharjo, Indonesia, serves a significant number of patients diagnosed with TB. To enhance the clinic's preparedness and ensure the availability of necessary resources, this study aims to develop an ARIMA model to forecast the number of TB patients visiting the clinic. By analyzing historical patient visit data, we seek to provide a reliable prediction model that will aid in better planning and management of healthcare resources.

This study's findings are expected to contribute to the strategic planning of healthcare services at the Mitra Sehat Clinic and offer a methodological framework that other clinics in similar settings can adopt. The implementation of forecasting tools like ARIMA can potentially lead to more effective TB care by enabling clinics to proactively respond to patient needs and streamline their operations.

Based on this background, this research aims to analyze and predict the number of patients visiting the Mitra Sehat Clinic with Tuberculosis to be used to determine policies and prepare health services for patients. In this research, forecasters use the ARIMA method with a case study in Sukoharjo Regency, Central Java, Indonesia, with a specific location at the Mitra Sehat Clinic Health Facility. This research is structured in 6 parts: (1) Introduction, in which the rationale of the research is briefly introduced; (2) Materials and Methods; (3) Case Study, where the health facility Mitra Sehat in Sukoharjo Regency as a case problem is described; (4) Results and Discussion, which intensively analyzes the results; (5) Conclusions, which provides a summary of the main finding of this research.

2. MATERIALS AND METHODS

2.1 Research Design

This study uses the Autoregressive Integrated Moving Average (ARIMA) forecasting method to forecast the number of patients visiting with Tuberculosis Diagnosed at Mitra Sehat Clinic in Sukoharjo-Central Java, Indonesia.

2.2 Sample and Population

The target population in this study were patients who visited Mitra Sehat Clinic with Tuberculosis Diagnosed at Mitra Sehat Clinic. Therefore, the sample was patients visiting with that criteria from January 2021 to August 2024.

2.3 Data Collection Methode

The data used in this research were obtained from secondary data from patients's visits with Tuberculosis diagnosed in Mitra Sehat Clinic from January 2021 – Augusts 2024 [6] accessed on the Pcare BPJS web page [6] and internal data bank at Mitra Sehat Clinic.

2.4 Data Analysis Methode

A time series is a series of observations taken based on time sequences. Between adjacent and correlated observations, it is said that in a time series, each observation taken from a variable is correlated with the variable itself at the previous time [12]. The methods to analyze the data used to forecast the number of outpatient visits for the next months using the Autoregressive Integrated Moving Average (ARIMA) method. Compared to the other models, between the Autoregressive Model, Moving Average, ARMA, and ARIMA, ARIMA is the best prediction model for forecasting daily trends or monthly cases in this paper [10]. Using this model, we could estimate the daily number of confirmed cases or patient visitation for the next month [10]. ARIMA modeling is one of the best modeling techniques for forecasting a time series. ARIMA models are always represented with the help of some parameters, and the model is expressed as ARIMA (p, d, q). Here, p stands for the order of auto-regression, d signifies the degree of trend difference, and q is the order of moving average [11]. The first step is to determine the pattern type of the data before conducting advanced analysis. The first stage is inspecting the data's pattern by plotting data to see trends and patterns in data. [14] .

No	Month	2021	2022	2023	2024
1	January	3	6	17	16
2	February	7	9	14	13
3	March	2	3	25	18
4	April	10	19	14	16
5	May	9	14	29	13
6	June	2	19	28	11
7	July	6	8	17	9
8	August	6	33	29	5
9	September	5	29	25	
10	October	8	14	29	
11	November	5	20	15	
12	December	6	24	18	
Total		69	198	260	101

The stationary of the data also needs to be tested, which is already stationary or not to variance and mean or average. Data can be stationary in the average if the fluctuations in the data are around a constant average value, independent of the time and variance of these fluctuations [12]. Data can be said to be stationary in variance if the data structure from time to time has a constant or constant fluctuation [12]. The stationarity of the data can be obtained with time series plot graphs and plots of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). Data is said to be stationary if it is not affected by the change in time [7]. If the data is not stationary with variance, it is necessary to transform until the rounded value is 1.00.

If the value of the rounded value or lambda (λ) is more than or equal to 1, the data is already stationary in variance [14] . If the data is not stationary to the average, it is necessary to do the differencing process, which is a process to eliminate trend elements and seasonal trends. After doing level one differencing, the data needs to be retested to determine whether it is stationary to the average by looking at the graph of the Autocorrelation Function (ACF) plot and Partial Autocorrelation Function (PACF). If not, level differencing is necessary, going to the second level until the data is stationary. The stationarity of ACF and PACF plots is also obtained from the lag outside the red line. If only one lag outside, it can be said that the data is stationer [14] . If the data is stationary, the next step is to identify temporary models based on ACF for MA values, PACF for AR values, and differencing for the value of d to determine the best forecasting model and measure accuracy by looking at the small error value.

The ARIMA model is said to be significant and feasible if it has a final P-value estimate of the parameter is below the error tolerance limit (α) 5% or 0.05 [13,14] and the Ljung-Box P-value is above the error tolerance limit (α) 5% or 0.05 [14] . In addition, based on the significance test, the best model can be determined with the smallest error value from each resulting model [14] .

The ARIMA method is used for short-term forecasting because it is accurate [7]. Autoregressive Integrated Moving Average (ARIMA) is a model that completely ignores the independent variables in forecasting. The values used by ARIMA for forecasting are past and present values of the dependent variable to produce accurate short-term forecasts [9]. The assumptions that must be met to use this method are data stationarity and error, whether white noise or not autocorrelated or normally distributed. White noise is a form of random variables that are not mutually correlated. The white noise process is determined by a constant average [12]. The residual model is normally distributed if the residual plot probability result exceeds the error tolerance limit (α) of 5% or 0.05 [14] .

3. CASE STUDY

Mitra Sehat Clinic, located in Sukoharjo, Central Java, Indonesia, has the highest number of registered BPJS patients. With more than 20,000 registered patients [6], sound planning is needed to provide good service to patients who visit for health consultations at these health facilities. Based on that, Mitra Sehat Sukoharjo is highlighted in this study. With that high number of patients, it is necessary to have a good plan to maintain optimal servicing. And this study can help analyze the good program the Clinic will prepare. This study will show the forecasting data for the Clinic and the prediction of patient visitation with Tuberculosis Diagnoses in the Clinic based on the data in the past. This study's findings may help the Clinic's policymakers to make a good policy or plan based on their understanding of their perspective to serve the patients in the future.

4. RESULTS AND DISCUSSION

4.1 Secondary Data

Table 1. Number of Patient Visits with Tuberculosis Diagnoses 2021 – 2024

The number of patient visits with Tuberculosis diagnosed in Mitra Sehat in the range of 2021 – 2024 increased in 2022 and 2023 and decreased in 2024. Forecasting is necessary to see an increasing trend for patient visits with Tuberculosis Diagnoses in Mitra Sehat. Mitra Sehat can use the forecast result to plan and run more specific programs based on these problems. Suppose the forecasting results indicate a predictive amount. In that case, Mitra Sehat can create a program to make a good plan to give optimal services to the Patient with Tuberculosis Diagnoses. The ARIMA method will use these data from 2021 – August 2024 to predict the number of patient visits with Tuberculosis Diagnoses in Mitra Sehat Clinic.

4.2 Data Plotting

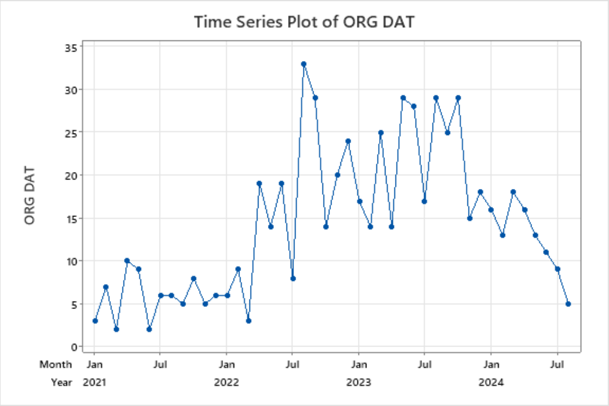

Figure 1. Time Series Plot of Patient Visits with Tuberculosis Diagnoses 2021 – 2024

The requirement in using the ARIMA model in analyzing data is to determine that the analyzed data is in a stationary condition. One way to determine the stationarity of the data is by plotting the data. Figure 1 shows that the data is not fluctuating. So, further analysis of seasonal patterns is necessary to determine whether the data is stationary in variance.

4.3 Stationary Test of Variance and Means

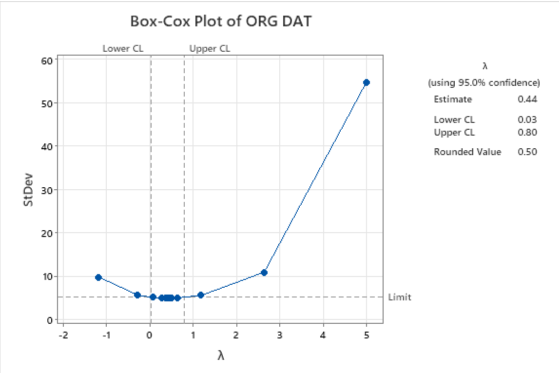

Figure 2. 1St Box-Cox Plot Data Transformation

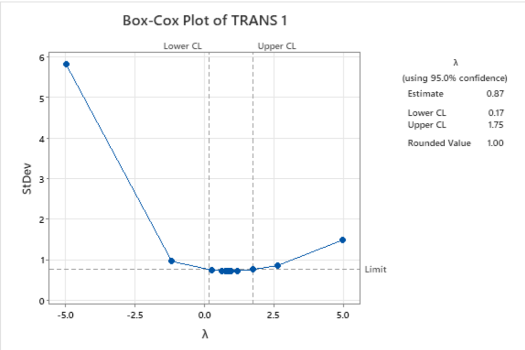

Figure 3. 2nd Box-Cox Plot Data Transformation

Figure 2 shows that the rounded value on that graph is not yet stationary, which means that the data must undergo a differencing or transforming process on the original data to get a rounded value at number 1, which means that the data is stationary.

Figure 3 shows that the rounded value on that graph is already 1. Hence, the data is already stationer in variance. Apart from observing Box-Cox plots, data stationarity can also be carried out by testing the graph of the auto-correlation function and auto-partial correlation with indicators of whether the data has been stationary in the mean. The following results are the graph autocorrelation plot and partial autocorrelation.

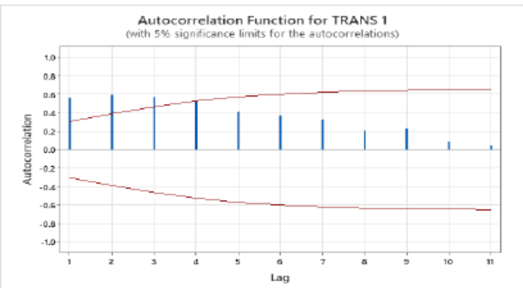

Figure 4. 1st ACF Graph Patient Data Transformation

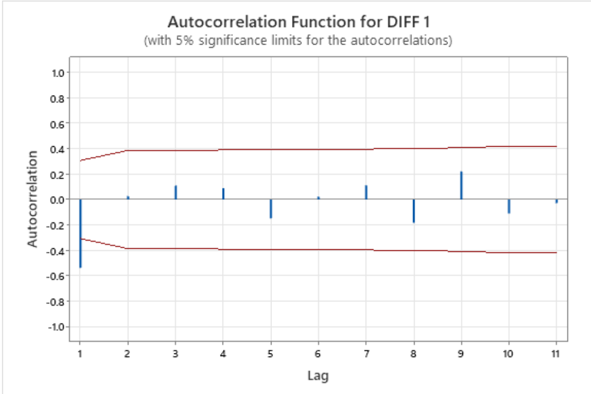

Figure 5. 2nd ACF Graph Patient Data Transformation

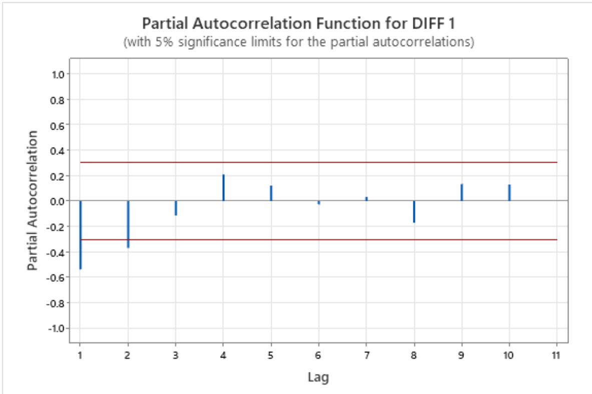

Figure 6. PACF Graph Patient Data Transformation

ACF plots in Figure 4 show that one lag (1st lag) is out from the red line, so it is necessary to differencing to get the upper lag in the red line.

After performing differencing ACF and PACF plots in Figures 5 and 6 shows that the data transformation already stationer in means due to the there is no coefficient auto-correlation and partial auto-correlation, which is outside the red line (Bartlett), and the data graph is not too dies down, so there is no need to do the differencing process. So, to determine the parameter model using ACF and PACF plots form the data transformation only. Temporary models which can be used to do advanced testing are ARIMA (1,1,0), ARIMA (1,1,1), and ARIMA (1,0,1).

4.4 ARIMA Models Estimation

The ARIMA model is significant and feasible if it has a P-value final estimate of the parameter below the error tolerance limit (α) 5% or 0.05 and the Ljung-Box P-value above the error tolerance limit (α) 5% or 0.05. In addition, based on the significance test, the best model can be determined with the smallest error value from each resulting model.

The temporary model is based on parameters testing with P-Value as follows,

Table 2. P-Value Analysis Report

	ARIMA Model	*P-Value Final Estimates of Paramaters*		*P-Value Ljung-Box*			Decision
		AR 1	MA 1	12	24	36	
Transformation	1,1,0	0,00		0,09	0,02	0,02	Significant
	1,1,1	0,32	0,03	0,61	0,29	0,76	Rejected
	1,0,1	0,00	0,00	0,55	0,16	0,55	Significant

The results of the analysis show that there are two ARIMA models with different significant parameters, namely (1, 1, 0) and (1, 0, 1) because it has a P-Value Final Estimates of Parameters value below the error tolerance limit (α) 5% or 0.05. Also, the Ljung-Box P-value exceeds the tolerance limit error (α) by 5% or 0.05.

And then the result of residual test results as the table follows,

Table 3. Residuals and Error Testing

	Models	MS	Probability Plot of Residuals	Decision
Transformation	1,1,0	47,7662	0,663	Rejected
Transformation	1,0,1	42,0276	0,053	Significant

The best model is ARIMA (1,0,1). Then, it is necessary to verify the model with the residual test consisting of white noise and normality test using Kolmogorov-Smirnov. The residual model is said to be normally distributed if the result of the residual plot probability is more than the error tolerance limit (α) of 5% or 0.05.

The residual graph of the ARIMA model (1,0,1) is as follows,

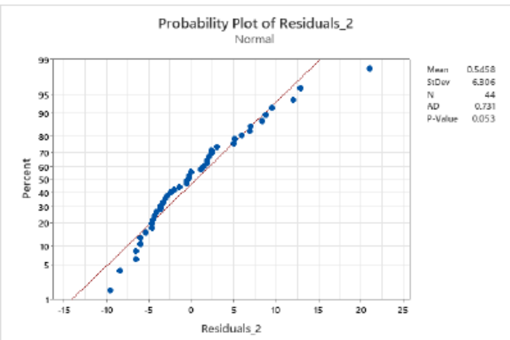

Figure 7. Residual Graph Model (1,0,1)

Based on Table 3 and Figure 7 above, it is obtained that the P-Value for the residual ARIMA model (1,0,1) is worth 0,053, which means that the value exceeds the value of the error tolerance limit (α) 5% or 0.05 in other words, the residuals from the ARIMA (1, 0, 1) model are already normal. From Table 3, It is also found that the ARIMA (1, 0, 1) model has a relatively small error value of 42,0276. So, in the end, it was determined that the best model for forecasting using the ARIMA method for this data is the ARIMA (1, 0, 1) model. The graph of the fits and fore is as follows,

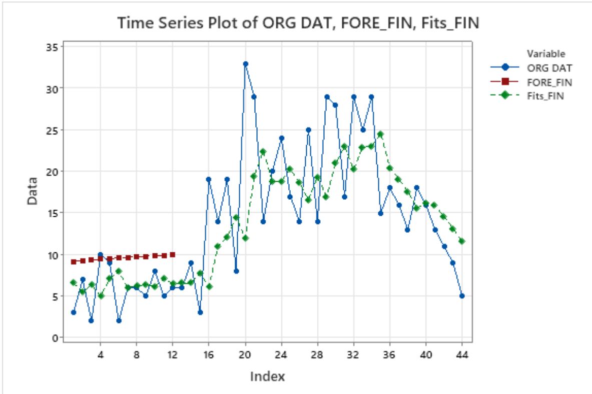

Figure 8. Time Series Plot Actual data, Fits, and Fore

4.5 Forecasting

The data to be forecasted is the original data. In this case, it is 44 data on the number of patients visiting Mitra Sehat Clinic with Tiberculosis Diagnosed. The ARIMA model (1,0,1) with the error value (MS) 142,0276 is to be used as the best ARIMA model for forecasting the original data. After the forecasting process for 12 months later, the following results are obtained,

Table 4. Forecasting Results

Forecasting for 12 Months			
Period	Forecast	95% Limits	
		Lower	Upper
45	9,17	-3,54	21,88
46	9,27	-4,24	22,77
47	9,35	-4,82	23,53
48	9,44	-5,32	24,20
49	9,52	-5,75	24,78
50	9,59	-6,12	25,30
51	9,67	-6,43	25,76

52	9,73	-6,71	26,17
53	9,80	-6,94	26,54
54	9,86	-7,15	26,87
55	9,92	-7,33	27,17
56	9,97	-7,49	27,44

Based on Table 4, It can be seen that the highest forecast is in August 2025, with almost 10, and the lowest is at Spetember 2024, with 9.

From the forecast results, an estimated error calculation will be carried out on the value of the forecast results to the estimated actual data value that will occur in the 12 months. The error calculation result of the actual data estimate is as follows,

Table 5. Error value of forecast data and actual data estimate

Actual Data Estimate	MAPE	MAD	MSD
Linear Trend Model	27,629	3,421	17,566
Quadratic Trend Model	29,226	2,981	11,801
Growth Curve Model	27,894	3,862	24,168

5. CONCLUSIONS

This study demonstrates the effectiveness of the ARIMA method in forecasting the number of tuberculosis (TB) patients visiting the Mitra Sehat Clinic in Sukoharjo, Indonesia. By analyzing historical patient visit data, the ARIMA model provided reliable predictions that can aid the clinic in strategic planning, resource allocation, and enhancing patient care services.

Implementing this forecasting approach can enable clinics to optimize staff scheduling, medication stock management, and overall operational efficiency. Furthermore, the insights gained from this study can guide other healthcare facilities in adopting similar predictive models to manage patient loads more effectively.

Future research can extend this work by incorporating additional variables, such as environmental factors, socioeconomic conditions, and public health interventions, to enhance the model's predictive power. Additionally, comparing ARIMA with other advanced forecasting techniques, such as machine learning models, may provide further opportunities for improving patient visit predictions in healthcare settings.

In conclusion, the ARIMA method offers a valuable approach to forecasting TB patient visits, contributing to more effective healthcare planning and resource management, ultimately leading to improved patient outcomes at the Mitra Sehat Clinic.

6. REFERENCES

[1]. World Health Organization (WHO). (2022). Global Tuberculosis Report.

[2]. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

[3]. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.

[4]. Zhu, X., et al. (2018). Influenza-like illness prediction using a local epidemic model with ARIMA. *Journal of Infectious Diseases*.

[5]. Benvenuto, D., et al. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief*

[6]. Indonesian Health Social Security Agency (BPJS), “Daftar Kunjungan Peserta Sakit.” Online available from https://pcarejkn.bpjs-kesehatan.go.id/eclaim/ViewKunjunganSakitByPpkTgl (Accessed on 1 February 2022)

[7]. Montgomery D. C., Jennings C. L., Kulahci M., "Time Series Analysis And Forecasting: Statistics Background For Forecasting," 2nd ed, John Wiley & Sons Inc., 2015, pp. 25–41.

[8]. Fadhilah Abd. Razak, Mahendran Shitan, Amir H. Hashim, Izham Z. Abidin, "Load Forecasting Using Time Series Models." Jurnal Kejuruteraan, pp. 53-62, 2009. DOI : dx.doi.org/10.17576/jkukm-2009-21-06

[9]. Nany Salwa, Nidya Tatsara, Ridha Amalia, Aja Fatimah Zohra, “Peramalan Harga Bitcoin Menggunakan Metode ARIMA (Autoregressive Integrated Moving Average).”, Journal of Data Analysis, Vol. 1, No. 1, pp. 21–31, 2018. DOI: https://doi.org/10.24815/jda.v1i1.11874

[10]. Saleh I Alzahrani, Ibrahim A Aljamaan, Ebrahim A Al-Fakih, "Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions." J Infect Public Health, National Library of Medicine, pp. 914-919, 2020. DOI : 10.1016/j.jiph.2020.06.001

[11]. Hiteshi Tandon, Prabhat Ranjan, Tanmoy Chakraborty, Vandana Suhag, "Coronavirus (COVID-19): ARIMA based time-series analysis to forecast near future." arXiv Cornell University, 2020. DOI: https://doi.org/10.48550/arXiv.2004.07859

[12]. Willian W.S. Wei, "Time Series Analysis, Univariate and Multivariate Methods," 2nd ed, Pearson Addison Wesley, 2006.

[13]. Ridha Maya Faza Lubis1, Zakarias Situmorang, Rika Rosnelly, "Autoregressive Integrated Moving Average (ARIMA-Box Jenkins) Pada Peramalan Komoditas Cabai Merah di Indonesia", Jurnal Media Informatika Budidarma, pp. 487-489, 2021. DOI: 10.30865/mib.v5i2.2927

[14]. Muhammad Bintang Pamungkas, Arief Wibowo, "Aplikasi Metode Arima Box-Jenkins Untuk Meramalkan Kasus Dbd Di Provinsi Jawa Timur", The Indonesian Journal of Public Health, pp. 183-186, 2018. DOI: https://doi.org/10.20473/ijph.v13i2.2018.183-196

[15]. Susilo Hariyanto, "Forecasting the Number of Outpatient Patient Visits Using ARIMA at Mitra Sehat Clinic: Case Study in Sukoharjo, Central Java Indonesia," Universal Journal of Public Health Vol. 10(5), pp. 482 – 491, 2022. DOI: 10.13189/ujph.2022.100506

[16]. Susilo Hariyanto, " Forecasting the Number of Patients Visits with Diabetes Mellitus Diagnosed Using ARIMA at Mitra Sehat Clinic Sukoharjo Indonesia," International Journal of Applied Engineering Research (IJAER) Volume 18, Number 4 (2023), pp: 256-261.