International Journal of Applied Engineering Research ISSN 0973-4562 Volume 18, Number 4 (2023) pp. 256-261

Forecasting the Number of Patients Visits with Diabetes Mellitus Diagnosed Using ARIMA at Mitra Sehat Clinic Sukoharjo Indonesia

Y.D. Sumanto[1], Susilo Hariyanto[2*] and Auliya Andriyati[3]

[1]Department of Mathematics, Diponegoro University, Semarang 50275, Indonesia
[2]Department of Mathematics, Diponegoro University, Semarang 50275, Indonesia
[3]Mitra Sehat Clinic, Regional Public Hospital Ir. Soekarno,Sukoharjo 57511, Indonesia
**Corresponding Author: susilohariyanto@lecturer.undip.ac.id*

Abstract
In providing health services, Clinic Mitra Sehat Sukoharjo has become a forefront health facility to maintain the health of society. During the last four years, the number of patients visited with diabetes mellitus diagnosed has reached 11,594 people. With that large number of patients visiting with Diabetes mellitus diagnosed, providing everything to serve patients well is necessary. This study will offer estimation data results that can be used by the Mitra Sehat Clinic, which previously only made preparations without data. This study will estimate the Number of Patients with Diabetes mellitus diagnosed that visit Mitra Sehat clinic Using ARIMA in the next few months. ARIMA, or Autoregressive Integrated Moving Average, is a forecasting method that will be used to predict the number of patient visits with specific criteria at the next time so that in the future, Mitra Sehat Clinic can prepare matters related to serving the visiting patients optimally. The steps for using the ARIMA method are using data from the required sample from patients visiting with Diabetes mellitus diagnosed in January 2019 – June 2023, 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 1652,96.

Keywords Arima, Forecasting, Patient Visits, Diabetes Mellitus, Minitab, Sukoharjo Indonesia

1. Introduction

Information technology has had many impacts on various sectors in today's life. The health sector is no exception, where information technology has started to be used in providing services both before, during, and after services. Information and communication technology development is a positive condition that will greatly support the development of Health Information Systems. Therefore, implementing information and communication technology in implementing the Health Information System is the wisest solution that must be taken. Although it is realized that information systems are not synonymous with computerization, today's information and communication technology developments significantly contribute to implementing information systems more professionally. [1]

At the end of 2021, the International Diabetes Federation (IDF), in its 10th edition of Atlas, confirmed that Diabetes is one of the fastest-growing global health emergencies in the 21st century. In 2021, more than half a billion people from all over the world will be living with Diabetes, or 537 million people to be precise, and this number is projected to reach 643 million in 2030 and 783 million in 2045. In addition to many people with Diabetes, it is estimated that the number of people with blood glucose levels starting to increase or are in the prediabetes phase, namely impaired glucose tolerance, in 2021 is around 541 million. Diabetes in this population also has a consequent high death rate associated with Diabetes, estimated at more than 6.7 million in the group of adults aged between 20–79 years. [2]

Diabetes does not only affect adults but also children and adolescents aged up to 19 years, and the number of people with Diabetes in this group is increasing every year. By 2021, it is estimated that more than 1.2 million children and adolescents will have diabetes mellitus [2]

Basic Health Research (RISKEDAS) carried out by the Indonesian Ministry of Health in 2018 collected data on diabetes mellitus sufferers in people aged over 15 years. The Diabetes Mellitus criteria in this study referred to the Indonesian Endocrinology Association (PERKENI) consensus by adopting the American Diabetes Mellitus Association (ADA) standards. [3]

RISKEDAS 2018 results show that the prevalence of Diabetes Mellitus in Indonesia based on a doctor's diagnosis in patients over 15 years of age is 2%. This figure shows an increase compared to the prevalence of Diabetes mellitus in 2015 at 1.5%, and according to blood sugar examination results, it increased from 6.9% in 2015 to 8.5% in 2018. [3]

Based on this background, this research aims to analyze and predict the number of patients visiting the Mitra Sehat Clinic with Diabetes Mellitus 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 because it has the highest number of BPJS registered patients in the district area [4]. 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 Diabetes Mellitus 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 Diabetes Mellitus Diagnosed at Mitra Sehat Clinic. Therefore, the sample was patients visiting with that criteria from January 2019 to June 2023.

2.3 Data Collection Methode

The data used in this research were obtained from secondary data from patients's visits with Diabetes Mellitus diagnosed in Mitra Sehat Clinic from January 2019 – June 2023 [5] accessed on the Pcare BPJS web page [5].

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 [11]. 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 [9]. Using this model, we could estimate the daily number of confirmed cases or patient visitation for the next month [9]. 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 [10]. 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. [13].

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 [11]. Data can be said to be stationary in variance if the data structure from time to time has a constant or constant fluctuation [11]. 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 [6]. 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 [13]. 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 [13]. 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 [12,13] and the Ljung-Box P-value is above the error tolerance limit (α) 5% or 0.05 [13]. In addition, based on the significance test, the best model can be determined with the smallest error value from each resulting model [13].

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 [8]. 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 [11]. The residual model is normally distributed if the residual plot probability result exceeds the error tolerance limit (α) of 5% or 0.05 [13].

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 [5], 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 Diabetes Mellitus 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 Diabetes Mellitus Diagnoses 2019 – 2023

No	Month	2019	2020	2021	2022	2023
1	January	220	202	273	249	258
2	February	155	213	243	220	229
3	March	214	166	280	256	255
4	April	198	165	281	219	241
5	May	203	165	259	248	278
6	June	227	248	253	296	244
7	July	383	282	181	236	243
8	August	188	292	205	238	-
9	September	250	317	281	247	-
10	October	236	298	292	268	-
11	November	176	287	235	253	-
12	December	224	274	265	233	-
Total		**2.674**	**2.909**	**3.048**	**2.963**	**1.748**

The number of patient visits with Diabetes Mellitus diagnosed in Mitra Sehat in the range of 2019 – 2023 increased in 2020 and 2021 and decreased in 2022. Forecasting is necessary to see an increasing trend for patient visits with Diabetes Mellitus 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 Diabetes Mellitus Diagnoses. The ARIMA method will use these data from 2019 – June 2023 to predict the number of patient visits with Diabetes Mellitus Diagnoses in Mitra Sehat Clinic.

4.2 Data Plotting

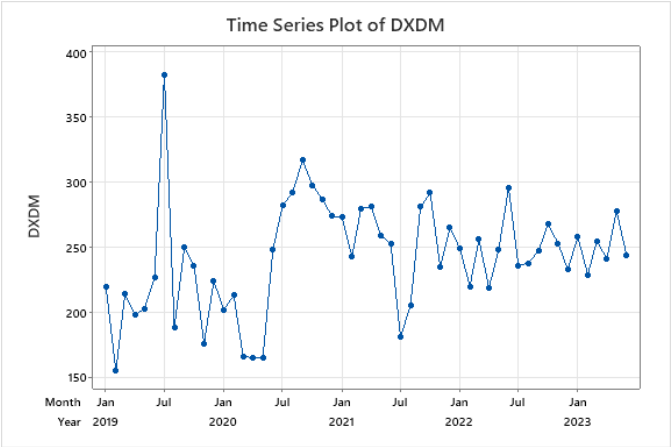

Figure 1. Time Series Plot of Patient Visits with Diagnoses of Diabetes Mellitus 2019 – 2023

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

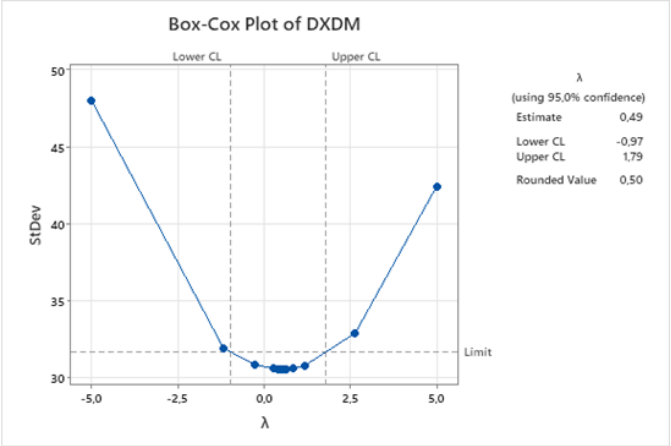

Figure 2. 1St Box-Cox Plot Data Transformation

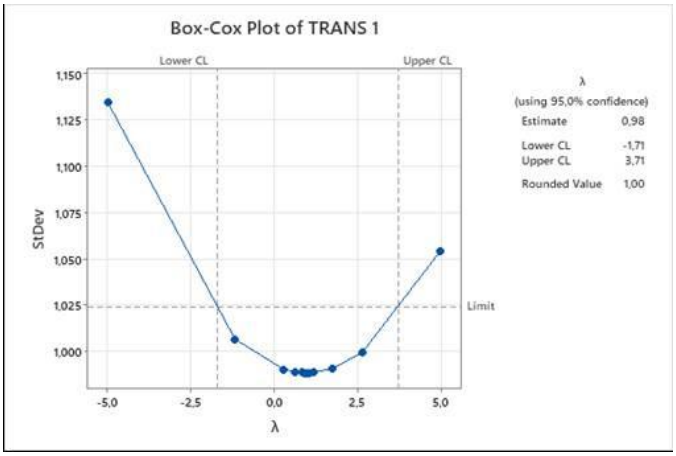

Figure 3. 2nd Box-Cox Plot Data Transformation

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.

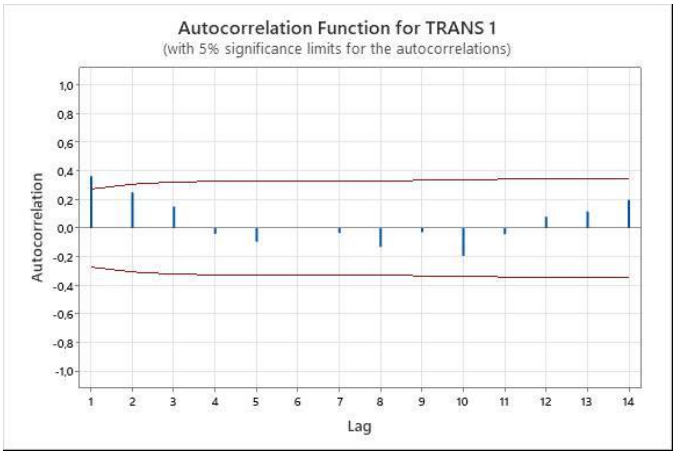

Figure 4. 1st ACF Graph Patient Data Transformation

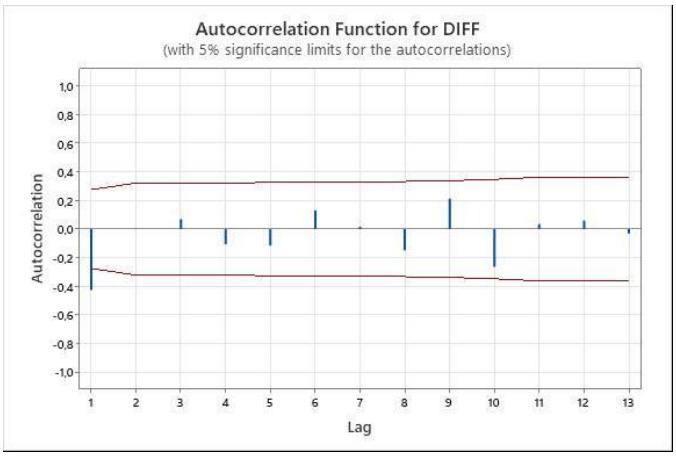

Figure 5. 2nd ACF Graph Patient Data Transformation

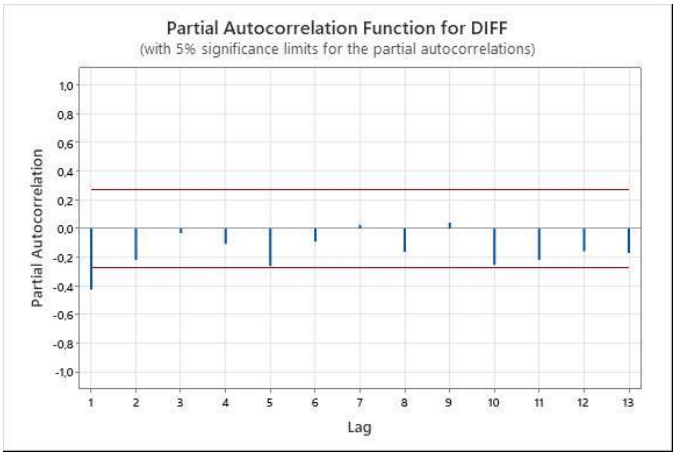

Figure 6. PACF Graph Patient Data Transformation

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 Parameters*				*P-Value Ljung-Box*				Decision
		AR 1	AR 2	MA 1	MA 2	12	24	36	48	
Transformation	1,1,0	0,001	-	-	-	0,263	0,526	0,287	0,474	Significant
	1,1,1	0,070	-	0,000	-	0,335	0,736	0,698	0,8	Rejected
	1,0,1	0,041	-	0,397	-	0,539	0,789	0,745	0,858	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	2000,96	< 0,010	Rejected
Transformation	1,0,1	1652,96	> 0,150	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,

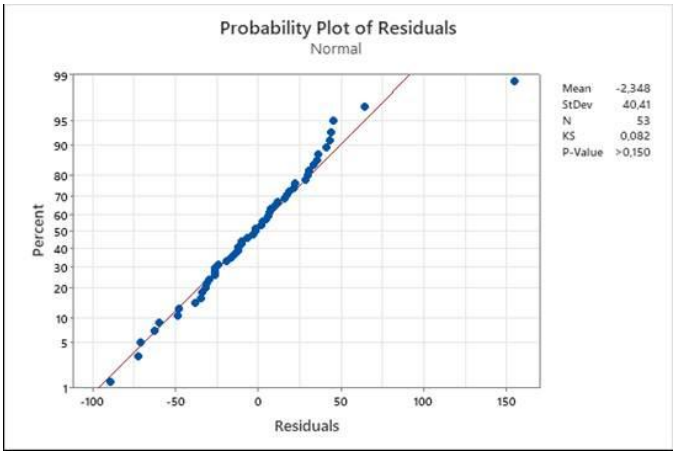

Figure 7. Residual Graph Model (1,0,1)

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 > 0.150, 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 1652,96. 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,

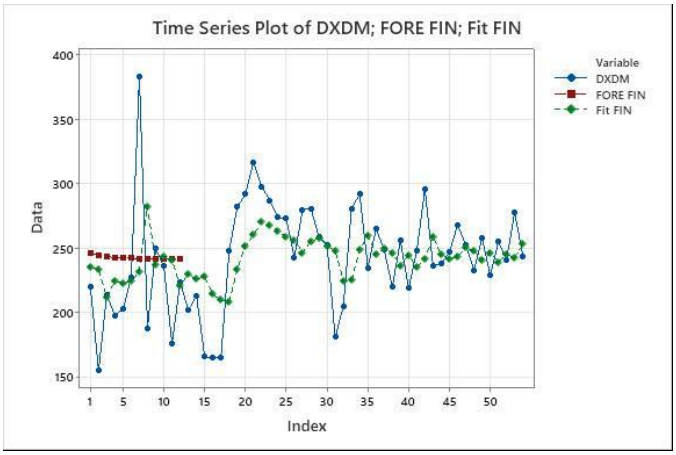

Figure 8. Time Series Plot Actual data, Fits, and Fore

4.5 Forecasting

The data to be forecasted is the original data. In this case, it is 54 data on the number of patients visiting Mitra Sehat Clinic with Diabetes Mellitus Diagnosed. The ARIMA model (1,0,1) with the error value (MS) 1652,96 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	Actual
55	246,047	166,344	325,750	243
56	244,473	160,995	327,951	
57	243,511	158,663	328,358	
58	242,922	157,567	328,276	
59	242,562	157,018	328,105	
60	242,341	156,727	327,955	
61	242,206	156,566	327,847	
62	242,124	156,474	327,774	
63	242,073	156,419	327,728	
64	242,043	156,387	327,698	
65	242,024	156,368	327,680	
66	242,012	156,356	327,668	

Based on Table 4, It can be seen that the highest forecast is in July 2023, with 246, and the lowest is at the end of June 2024, with 242. In the actual column, the exact number of patient visits with Diabetes Mellitus Diagnoses in the first month, July 2023, is 243, are also presented with results that are still within the tolerance threshold.

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	4,621	11,150	215,214
Quadratic Trend Model	4,311	10,349	201,551
Growth Curve Model	4,619	11,171	215,905

5. Conclusions

Health issues in the current era have become an essential topic of discussion with various existing issues, so derivative health problems are also significant. In this research concerning Forecasting the Number of Outpatient Patient Visits with Diabetes Mellitus Diagnosed Using ARIMA at Mitra Sehat Clinic, it can be concluded that based on the forecasting results, the number of patient visits with diabetes mellitus diagnoses in 2023 and 2024 will tend to be stable until the rest of 2023 with various underlying factors, potentially be able to exceed the previous three years. Therefore, the ARIMA forecasting method must choose the best model with the smallest error value (MS). Several tests are needed to get the best model, such as the test stationarity, residual test, and normality test. After finding the best model, it can be used to estimate the number of visiting patients with Diabetes Mellitus Diagnosed in that Clinic.

With the potential for patient visits with diabetes mellitus diagnosed, advice that can be given to the Clinic is to prepare

all kinds of tools and needs to provide services to these patients. So, with this data, the Clinic can have a basis for preparing the equipment necessary and maintaining the quality of service to remain optimal.

Conflicts of Interest

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

Acknowledgments

This work was supported by the Department of Mathematics, Diponegoro University, Semarang, Indonesia. This research was funded by In addition to the state budget (besides APBN) of Science & Mathematics Faculty - Diponegoro University in 2023, SPK Number 22.A/UN7.F8/PP/II/2023. We thank the Dean of Science & Mathematics Faculty of Diponegoro University and all parties involved in this paper.

Funding Statement

This research was funded by In addition to the state budget (Selain APBN) of the Science & Mathematics Faculty - Diponegoro University in 2023, SPK Number 22.A/UN7.F8/PP/II/2023. We thank the Dean of Science & Mathematics Faculty of Diponegoro University and all parties involved in this paper.

Appendix A

A list of abbreviations used in this paper is shown below

Abbre-viation	Meaning
BPJS	Badan Penyelenggara Jaminan Sosial (Indonesian Health Social Security Agency)
ARIMA	Auto-Regressive Integrated Moving Average
ACF ; PACF	Autocorrelation Function; Partial Autocorrelation Function
AR; MA	Auto-regression; Moving Average
MS	Minimum Mean Square Error
MAPE	Mean Absolute Percentage Error
MAD	Mean Absolute Deviation
MSD	Mean as Squared Deviation

REFERENCES

[1] Resource Center - Ministry of Health Republic of Indonesia, "Sistem Informasi Kesehatan." Online available from https://rc.kemkes.go.id/sistem-informasi-kesehatan (accessed on 30 June 2023)

[2] Ministry of Health Republic of Indonesia, "Diabetes Melitus Adalah Masalah Kita.", Online available from https://yankes.kemkes.go.id/view_artikel/1131/diabetes-melitus-adalah-masalah-kita (accessed on 02 July 2023)

[3] Infodatin Pusat Data dan Informasi The Ministry of Health Republic of Indonesia, "Tetap Produktif Cegah, dan Atasi Diabetes Mellitus"." Online available from https://www.kemkes.go.id/downloads/resources/download/pusdatin/infodatin/Infodatin%202020%20Diabetes%20Melitus.pdf (accessed on 10 July 2023)

[4] Indonesian Health Social Security Agency (BPJS), "Jumlah Peserta Terdaftar." Online available from https://pcarejkn.bpjs-kesehatan.go.id/eclaim/ViewRekapPstPerPpk (Accessed on 01 July 2023)

[5] 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)

[6] 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.

[7] 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

[8] 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

[9] 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

[10] 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

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

[12] 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

[13] 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

[14] 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