ORIGINAL RESEARCH PAPER
Volume - 11 | Issue - 06 | June - 2022 | PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

MODELING DIABETES PATIENTS' ATTENDANCE AT BAGAMOYO DISTRICT HOSPITAL USING AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODELS.

Statistics

Julius Moinget Loibor Marian University College (A Constituent College of St. Augustine University of Tanzania) Department of Mathematics and Computer Science, Marian University College, Tanzania

Mfano Mohamedi Kisoma* Marian University College (A Constituent College of St. Augustine University of Tanzania) Department of Mathematics and Computer Science, Marian University College, Tanzania*Corresponding Author

ABSTRACT

This study presented the use of Autoregressive Integrated Moving Average (ARIMA) technique to modeling the diabetes patients' attendance at Bagamoyo district hospital using monthly time series data. The data used in the analysis of this study are monthly reading of diabetes patients data covered the period from January 2014 to December 2021. The data were retrieved from the hospital electronic health management information system. The Autoregressive Integrated Moving Average (ARIMA) approach was applied to the diabetes patients' data through the model identification, estimation, diagnostic checking, stationary and forecasting in R statistical software. The study identified Autoregressive Integrated Moving Average ARIMA (0, 1, 1) model to be the best one to fit for the monthly total number of diabetes patients' attendance hospital outpatient department for seven years of monthly data. This was verified by Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) of the model selection criteria, Autoregressive Integrated Moving Average ARIMA (0, 1, 1) model shown the smallest values, hence this was selected as an adequate model to represent the Bagamoyo district hospital diabetes patients data. The forecast values indicate clearly that diabetes patients' attendance at Bagamoyo district hospital would be on an increase rate per month during the year 2022. Therefore, the government needs to put in place more and to engage necessary requirements for a satisfying healthcare system by increasing the medical supplies to the Bagamoyo district hospital.

KEYWORDS

Modeling, ARIMA model, Diabetes patients, hospital, Bagamoyo

1.0 Introduction

Growing diabetes becomes a big challenge in Tanzania specifically in Bagamoyo district, for that reason studying of this phenomenon becomes an important issue. Diabetes is a common disease around the world, which can cause various systemic diseases and high mortality (Al Zahrani et al., 2020). It is a disease which categorize by high sugar levels in the blood and urine (Al Zahrani et al., 2020). It is usually diagnosed by means of glucose tolerance test. There are three types of diabetes mellitus (Classification of Diabetes 2019). The first type of diabetes mellitus results from the body's failure to produce sufficient insulin. It is often occurs among young children. The second type diabetes mellitus results from resistance to the insulin, often initially with normal or increased levels of circulating insulin (Classification of Diabetes, 2019).Gestational diabetes, the third type is Gestational diabetes which happens when pregnant women without a previous history of diabetes develop a high blood glucose level. Diabetes is a major health challenge (Al Zahrani et al., 2020). Worldwide the estimated number diagnosed with Diabetes is roughly 463 million people per year and mortality is 4.2 million deaths per year (Classification of Diabetes 2019). Diabetes prevalence in Tanzania was estimated diabetes prevalence of 9.1% among adults aged between 25 and 64 years (Mwangome et al., 2017), and a significant increase from the estimated 2.5% diabetes prevalence in 1984 among persons aged 20 years and above (Stanifer et al., 2016) among 3145 Tanzanian's living in three geographical locations in north western part of the country. There is inadequate data on the trend of the prevalence of diabetes in Tanzania (Kinimi E., et al., 2017). Diabetes prevalence (% of population ages 20 to 79) in Tanzania was reported at 5.7 % in 2019. According to the World Bank collection of development indicators, compiled from officially recognized sources. Tanzania - Diabetes prevalence (% of population ages 20 to 79) - actual values, historical data, forecasts and projections were sourced from the World Bank on February of 2022. In this current study Autoregressive Integrated Moving Average (ARIMA) of time series models were applied to data representing diabetes patients in Bagamoyo district hospitals with the objective of finding out an appropriate models provide accurate prediction to diabetes patients in the Bagamoyo district hospital based accuracy measurements such as Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC).

2.0 Literature review

Inside available literature review, time series modeling usually was established and used in the area of econometrics. On the other hand, there are currently works in other areas such as biomedicine, health, etc. employing time series techniques for modeling and forecasting. For example, Juang W-C et al., (2017) had applied time series modeling and forecasting for emergency department visits in a Medical Centre in Southern Taiwan. The study had shown that the final suitable ARIMA (0, 0, 1) model was selected as the best fit and it has the smallest value of the MAPE 8.91%. Sukmak, Thongkam & Leejongpermpoon (2015) conducted a study using time series forecasting on visits of Anxiety Disorder outpatients at the Centre using Data mining. The main objective of their study was to forecast the number of anxiety disorder patients who are seeking treatment at an outpatient clinic in 2011 by comparing two Artificial Neural Network (ANN) models and selecting the most powerful model. The demand forecast model was constructed over the time series data from January 2007 to December 2010 and the model accuracy was evaluated via Mean Absolute Percentage Error (MAPE) and Radial Basis Function (RBF) was selected to be the best fit. Singye and Unhapipat (2018) conducted a study using time series forecasting on diabetes patients data, the monthly number of diabetes patients obtained from Jigme Dorji Wangchuk National Referral Hospital and they had fitted those data by using ARIMA model and also they had come up with ARIMA (0, 1, 1) was the best model to describe and predict the future trends of diabetes incidences on considering the simplest parsimonious lowest order model.

Hanaa Elgohari et al., (2019) conducted the study on forecasting the number of outpatient visits in tertiary hospital using time series based on Autoregressive Integrated Moving Average (ARIMA) and exponential smoothing and their study was concluded that the ARIMA (3, 1,3) model was more appropriate and more sensitive than exponential smoothing models of forecasting the outpatient visits in tertiary hospital. In Tanzania, few available literature reviews on diabetic study had relied on descriptive statistics and qualitative studies so, no one at all attempted to modeling and forecasting the number of patients for diabetes incidence using time series data with an ARIMA model approach. For example, Kinimi et al., (2017) conducted the study to assess prevalence of Diabetes Mellitus and associated risk factors among outpatients seeking medical care at Morogoro Regional Referral Hospital in Tanzania and the findings of their study indicated that an overall prevalence of diabetes was 10.08% for undiagnosed town dwellers outpatients.

Kibona and Mbago (2018) had applied time series Autoregressive Integrated Moving Average (ARIMA) model for forecasting

wholesale prices of maize in Tanzania. The major objective of their study was to modeling and forecasting whole prices of maize in Tanzania for data from February 2004 to August 2017 obtained from the Bank of Tanzania. Their study also employed mathematical modeling with the help of the Box- Jenkins method. Therefore, the results of the findings of their study concluded that the ARIMA (3, 1, 1) model was selected as the best model for maize wholesale prices based on the minimum value of AIC, BIC and the fitted model was found to be adequate using Ljung-Box test.

3.0 Data Collection and Methodology

3.1 Data collection

Data from the Bagamoyo District hospital were used as the basis for this study. The researchers were retrieved the diabetes patients data from hospital electronic health management information system. The monthly recorded diabetes patients data from January 2014 to December 2021 were used as the working out data on which the forecasting models were trained (estimating model parameters) base on the count of diabetes patients attendance at the Bagamoyo hospital.

3.2 Methodology

We used Box-Jenkins approach to forecast the number of monthly diabetes patients attending at Bagamoyo district hospital for the next six months by using the existing data from January 2014 until December 2021. The Box-Jenkins forecasting models are based on statistical concepts and principles as follows: Identification, Parameter estimation, diagnostic checking, Model selection and forecasting (Bowerman et al., 2005).

3.2.1 Identification

This is a step of determining necessity of differencing, to produce stationarity and defining the order of seasonal and non-seasonal Autoregressive (AR) and Moving average (MA) operators for the series (Singye and Unhapipat ,2018). Generally, in this step appropriate structure Autoregressive Integrated Moving Average (ARIMA) and the order of the model are specified (Singye and Unhapipat, 2018). The tools used for the identification are correlogram, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).

3.2.2 Model estimation

This is a second step in ARIMA modeling to estimate the coefficients of the model. Box- Jenkins methodology required that ARIMA model must satisfy the condition of stationary and invertible. In this current study that based on the selected ARIMA (0, 1, 1) model, the parameters were estimated for the diabetes patients attended at Bagamoyo district hospital.

3.2.3 Diagnostic checking

This is a third step in ARIMA modeling process. In this step is to identify the correct parameter of the model has been chosen. If the model is "good", then the residuals are expected to be random and close to zero. There are several ways of checking if a model is good. The common approach is to examine the residues. In this step of diagnostic checking, the ACF behavior of the residuals produced by estimation process is examined. For good model, the residuals time series should be close to an independent and identically distributed (Bowerman et al., 2005). Residual analysis of the ARIMA (0, 1, 1) model from the data of diabetes patients attended at Bagamoyo district hospital were carried out to verify the model.

3.2.4 Stationary and Non stationary of time series

To fit a time series data into ARIMA models, the data are required to be stationary (Bowerman et al., 2005). Therefore, the raw data need to be transformed to stationary first before next steps of modeling process are performed. In other words, the transformed data or original data have statistical properties that are constant through time. So in order to determine whether the data are stationary or, we can observe the ACF behavior of the series and also we can use ADF test for testing the unit root (Bowerman et al., 2005). Therefore, before forecasting the diabetes patients' attendance at Bagamoyo District Hospital from January 2014 to December 2021, the data were transformed and differenced so that the statinonrity of the ARIMA model is ensured before forecasting the data.

3.2.5 ARIMA model

ARIMA is one of the most traditional methods of non-stationary time series analysis (Rahman and Hasan, 2017). In contrast to the regression models, the ARIMA model allows time series to be explained by its past or lagged values and stochastic error terms (Rahman and Hasan, 2017). The models developed by this approach are usually called ARIMA models because they use a combination of autoregressive (AR), integration (I) - referring to the reverse process of differencing to produce the forecast and moving average (MA) operations (Judge et al., 1988).

The ARIMA model is represented by ARIMA (p, d, q), where "p" stands for the order of the auto regressive process, "d" is the order of the data stationary and "q" is the order of the moving average process. The general form of the ARIMA (p, d, q) can be written as (Judge et al., 1988)

$$\Delta^d y_t = \delta + \theta_1 \Delta^d y_{t-1} + \theta_2 \Delta^d y_{t-2} + \cdots + \theta_p y_{t-p} + e_{t-1}\alpha e_{t-1} - \alpha_2 e_{t-2}\alpha_q e_{t-2} \quad \ldots\ldots\ldots\ldots (1)$$

Where, Δ^d denotes differencing of order d, that is $\Delta y_t = y_t - y_{t-1}, \Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$ and so on, $y_{t-1}, \cdots, y_{t-p}$ are post observations, $\delta, \theta_1, \ldots, \theta_p$ are parameters of the model to be estimated similar to regression coefficients of the Autoregressive process AR of order p denoted by AR(p) and it be written as (Judge et al., 1988),

$$Y = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + , \ldots + \theta_p y_{t-p} + e_t \quad \ldots\ldots\ldots\ldots (2)$$

Where, e_t is forecast error, assumed to be independently distributed across time with mean θ and variance $\theta_2 e$, past forecast errors are $e_{t-1}, e_{t-2}, \ldots, e_{t-q}$, moving average MA coefficients are $\alpha_1, \ldots, \alpha_q$. While the moving average MA model of order q that is MA (q) can be written as (Judge et al., 1988),

$$Y_t = e_t - \alpha_1 e_{t-1} - \alpha_2 e_{t-2} - , \ldots , - \alpha_q e_{t-q} \quad \ldots\ldots\ldots\ldots (3)$$

3.2.6 Model selection criteria

Model selection criteria are criteria used to choose the best model to be used (Rahkmawati, Sumertajaya, & Aidi, 2019). There are several criteria that are often used in the selection of the ARIMA model (Rahkmawati, Sumertajaya, & Aidi, 2019). Bayesian Information Criteria (BIC) and Akaike Information Criteria (AIC) are the estimates of functions of the posterior probability of a model being true, under a certain both Bayesian and Akaike setup, So that a lower both values of BIC and AIC means that a model is more likely to be true model (Hanaa Elgohari, et al., 2019).

$$\text{BIC} = 2\log n - 2\log\left(L\left(\frac{\theta}{y}\right)\right) \quad \ldots\ldots\ldots\ldots (4)$$

Where n is the number of observation and θ is the number of parameters in the time series model. And also L is the maximized log-likelihood function.

AIC = -2 (maximized log-likelihood) + 2 (the number of parameters estimated). For the fitted time series models which will appear to have the smallest value of AIC is chosen to be the most appropriate time series model ARIMA for fitting the diabetes patients hospital data set.

3.2.7 Forecasting

Forecasting procedure is an approach to predict or estimate quantatively and qualitatively about what is going to occur in the future based on the relevant data from the past (Rahkmawati, Sumertajaya, & Aidi, 2019). In the decision making process, the forecasting will play an important role. In this current study forecasting involved predicting future number of diabetes patients attendance in Bagamoyo district hospital using the fitted selected ARIMA (0, 1, 1) model. In this present study next six months ahead monthly forecasts were generated. That is the forecasts from January to June 2022.

3.3 Ethical consideration.

With the help of an introduction letter from the Deputy Principal of Academic Affairs Office of Marian University College, the researchers were leaned to collect the secondary data of diabetes patients' attendance at Bagamoyo district hospital during period from January 2014 to December 2021. The Medical Officer in charge of Bagamoyo district hospital granted permission to carry out data collection from the respective hospital departments.

4.0 Results and Discussion

4.1.0 Discussion of the study

This part discusses the data analysis results of applying ARIMA models to data presenting the diabetes patients attendance at Bagamoyo District Hospital during period from January 2014 to

December 2021 through testing of stationary, identification, estimation, model selection, diagnostic checking and forecasting using statistical R software.

Before modeling the diabetes patients' data, the preliminary descriptive analysis of the diabetes patients' information obtained from Bagamoyo district hospital electronic health management information system was carried out first. The filtered diabetes patients' information obtained from the Bagamoyo district hospital health management information system holds information on the sex and age of the patients. The summary statistics on gender-wise frequency and percentage and age group of diabetes patients of Bagamoyo district hospital is illustrated in the table 1.0 below in the results section.

A quicker look at table 1.0 below indicating that the female patients were visited more in the diabetic outpatient department of Bagamoyo district hospital for the year 2014- 2021 were 2382 (58.15%) with ages between (5-60) years compared to males 1714 (41.85%) with ages between (5-60) years. The ratio is 2:1 for a female to male implies that females were more affected by diabetes diseases and required treatment on a daily basis. Also, again the results from table 1.0 below shows that most of the patients visited at the diabetic department for treatments in Bagamoyo district hospital for the year 2014- 2021 were of ages between (61+) years 1343(57.0%) are females while males were 1013(43.0%).

The sequence chart of the diabetes patients attended at Bagamoyo district hospital from the period of January 2014 to December 2021 fluctuates is shown in figure 1.0 below in the result section. It can be shown that number of diabetes patients fluctuates shows a slight increases. The time plot shows some sudden changes, particularly a big fluctuation from 2015 to 2016 and there is an increase in trend in the attendance from 2017 to 2020. Also, from 2020 to 2021 there is a big fluctuation whereby there are some months with few attendances and other with high number of attendances. There is an evidence of changing variance. So, we will do logarithmic transformation. The data are clear non-stationary as the series moves up and down for long periods. Therefore, we will take a difference of the data until we achieve stationarity.

In order to select appropriate model, the Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) were used to test different ARIMA models. The transformed and differenced diabetes patients' data has been generated. Several ARIMA models with Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) values were computed are tested as shown in table 3.0 below in order to select an appropriate ARIMA model for forecasting the diabetes patients' hospital time series data.

A quicker look at table 3.0 below can be shown that ARIMA (0, 1, 1) model has the smallest value of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) of model selection criteria. In this model, it is assumed that diabetes patients attended at Bagamoyo district hospital are subject to autoregressive of order 0, moving average of order 1, and difference of order 1. So, from the table 3.0 below, we conclude that the model (0, 1, 1) is the best, which gives us the lower values for both AIC and BIC respectively. Therefore, we will rely on this selected ARIMA (0, 1, 1) model to estimate the forecasts of the next months of the year 2022.

After identifying the best model, the process was followed by a stage of parameter estimation. ARIMA models are usually estimated which is necessarily stationary series. Parameter estimates along with corresponding standard errors of fitted ARIMA (0, 1, 1) model of monthly diabetes patients attendance at Bagamoyo hospital were shown in table 4.0 given below in section of the results. Results from Table 4.0 below indicates the estimated parameters for ARIMA (0, 1,1) selected as the best fit model for the forecasting process for diabetes patients hospital data. The fitted ARIMA (0, 1, 1) model is:

$$X_t = 0.101 - 0.708X_{t-1} \quad \ldots\ldots (5)$$

The negative estimate for the coefficient in ARIMA (0, 1, 1) model above means that lags it bear an inverse relationship with previous variables in previous periods.

Residual analysis of the ARIMA (0, 1, 1) model was carried out to verify the model. In the figure 3.0 below shows the Autocorrelation Function (ACF) plot of the diabetes patients' data. By observing the ACF plot shown in figure 3.0 below, we could see that the Autocorrelation Function (ACF) plot of the residuals from the ARIMA (0, 1, 1) model shows all correlations within the threshold limits indicating that the residuals are behaving like white noise. A Box-Ljung test returns a large p-value, also suggesting the residuals are white noise and also we could conclude that the model was adequate.

After we have identified the appropriate ARIMA model during the data analysis, then we are able to model Diabetes Meritus Attendance at Bagamoyo District Hospital using ARIMA (0, 1, 1). We have been able to predict the Diabetes Meritus Attendance for the next six months from January 2022 to June 2022. Table 5.0 below in the section of results shows the forecasted value of monthly diabetes patients' hospital data using the best selected fitted model ARIMA (0, 1, 1). Again, a quicker look on the results from table 5.0 below in the section of the results it was indicating that the forecasting values will increase for the Diabetes Patients attending to Bagamoyo district hospital per month during the year 2022.

4.2.0 Results of the study.

Table 1.0 Descriptive statistics of demographic characteristics of diabetes patients at Bagamoyo hospital (2014-2021).

Category variables	Gender wise			
	Male		Female	
	Freq.	%	Freq.	%
Age in years				
5-60	1714	41.85	2382	58.15
61+	1013	43.0	1343	57.0

Figure 1.0

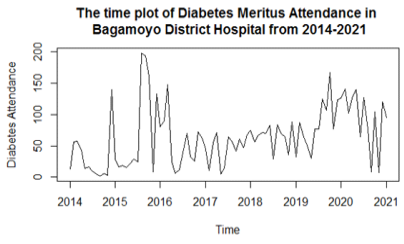


Table 2.0 ADF test results of diabetes patients' attendance at Bagamoyo district hospital

Unit root test/ Diabetes patients data		
	Test value	Prob
ADF test	-5.8374	0.01

Table 3.0 Model selection for diabetes patients' attendance at Bagamoyo district hospital (2014-2021)

Models	AIC	BIC
ARIMA(0, 1,0)	905.9835	908.4143
ARIMA(0, 1,1)	878.7499	883.6115
ARIMA(0, 1,2)	880.1269	887.4194
ARIMA(1, 1,0)	888.6497	893.5114
ARIMA(1, 1,1)	879.6863	886.9788
ARIMA(1, 1,2)	879.6516	889.3749

Table 4.0 Estimated parameters of ARIMA (0, 1, 1) model.

Parameter	Estimate
θ	-0.708
Standard error (s.e)	0.101
δ^2	1934

Table 5. Forecast of Diabetes Patients in Bagamoyo district hospital using ARIMA (0, 1, 1) model from January to June 2022.

Month	Forecasted values
January, 2022	144
February, 2022	148
March, 2022	152
April, 2022	156
May, 2022	161
June, 2022	165

Figure 2.0 The time plot of the transformed and difference diabetes patients' monthly data.

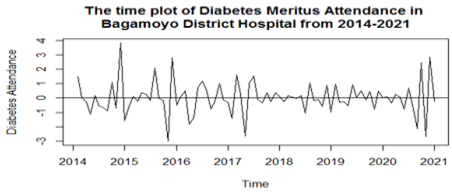
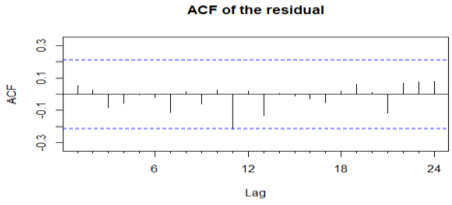


Figure 3.0 ACF of the Residual of ARIMA (0, 1, 1).



4.3 Conclusion.

This study aimed to model the diabetes patients' attendance at Bagamoyo district hospital during the period from January 2014 to December 2021 using Integrated Autoregressive Moving Average (ARIMA) model. The models were developed based on monthly diabetes patients' hospital time series data. On basis of the results obtained, finally the best model was selected. The study identified that the ARIMA (0, 1, 1) model has minimum values of both Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) so that this model will be appropriate to forecast the diabetes patients hospital data. In general, forecasting of diabetes patients attendance in Bagamoyo district hospital will increasing for the next months. From forecasting results in table 5.0 above it indicated that the forecasting values will increase for the Diabetes Patients attending the Bagamoyo district hospital per month during the year 2022. In addition to increase of number of diabetes patients' attendance in Bagamoyo district, therefore, the government needs to put in place more and to engage necessary requirements for a satisfying healthcare system by increasing the medical supplies to the Bagamoyo district hospital.

REFERENCES:

1. Al Zahrani, S., Al Rahman Al Sameeh, F., Musa, A.C.M and Shokeralla, A.A.A. (2020). Forecasting Diabetes Patients Attendance at Al- Baha Hospitals Using Autoregrssive Fractional Integrated Moving Average (ARFIMA) models. Data Analysis and Information processing, 8, 183-194.
2. Bowerman, B. L., O' Connell, R.T., Koehler, A. B. (2005). Forecasting, Time series, and Regression: An applied approach. Thomson Learning, Inc.
3. Hanaa Elgohari, Ashraf M. Bakri, Mohammed Abdul Majeed. (2019). Forecasting the number of outpatient visit in tertiary hospital using time series based on ARIMA and ES models. Australian Journal of Basic and Applied Science, 70-77.
4. Juang W-C, Huang F-D, et al. (2017). Application of time series analysis in modelling and forecasting emergency department visits in a medical center in Southern Taiwan.
5. Judge, G. G., Carter Hill, R., William, E. G and Helmut, I. (1988). Introduction to the Theory and Practice of Econometrics. 2^nd Edition,. New York: John Wiley and Son.
6. Kibona Shadrack Elia, & Mbago Maurice Chakusaga. (2018). Forecasting Whole sale Price of Maize in Tanzania Using Arima Model. General Letters in Mathematics vol.4, 131-141.
7. Kinimi E., Balthazary S., Kitua S., & Msalika S. (2017). Prevalence of diabetes mellitus in patients seeking medical care at Morogoro Regional Referral Hospital in Tanzania. Tanzania Journal of Health Research 19(4).
8. Mwangome M.Geubbels E., Klastser P. & Dieleman M. (2017). Perceptions on diabetes care provision among health providers in rural Tanzania; a qualitative study. Health Policy and Planning, 418-429.
9. Rahman, A. and Hasan, M. M. (2017). Modeling and Forecasting of Carbon Dioxide Emission in Bangladesh Using Autoregressive Integrated Moving Average (ARIMA) models. Open Journal of Statistics, 7, 560-566.
10. Stanifer, J. W., Cleland, C. R., Makuka, G. J, Egger, J. R.., Maro, V., Maro, H., Karia, F., Patel, U.D., Burton, M.J & Philippin, H. (2016). Prevalence, Risk factors, and Complications of Diabetes in the Kilimanjaro Region: A population- based study from Tanzania. PLOS One 11, 1310-1371.
11. Vatinee Sukmak, Jaree Thongkam, & Jintan Leejongpermpoon. (2015). Time Series Forecasting in Anxiety Disorders of Outpatient visits Using Data Mining. KKU Res 20(2), 241-253.
12. WHO. (2019). Classification of diabetes mellitus. Geneva.
13. Yeni Rahkmawati, I Made Sumertajaya & Muhammed Nur Aidi. (2019). Evaluation of Accuracy in Identification of ARIMA Models Based on Model Selection Criteria for Inflation Forecasting with the TSClust Approach. International Journal of Scientific and Research Publications, vol 9, 439-443.