Prediction of Illiteracy Rates in Indonesia Using Time Series

Muhamad Chamdani[1], Umi Mahmudah[2], Siti Fatimah[3]

[1,3]*Sebelas Maret University, Indonesia*
[2]*State Islamic Institute of Pekalongan, Indonesia*

muhamad_chamdani@staff.uns.ac.id[1]; u_mudah@yahoo.com[2]; stfatimah89@gmail.com[3]

Abstract

Illiteracy eradication will increase the quality of human resources; therefore, it should be a priority program for every country. This study applies time series analysis to predict illiteracy rates in Indonesia by using data from the Indonesia Statistic Centre from 2003 to 2017. As data on illiteracy rates are non-stationary, a differencing process is required. The results of this study indicate that the best model to predict illiteracy rates in Indonesia is ARIMA (3,3,1), which shows there are three processes of differencing to obtain stationary data. The results indicate that the point forecast of the illiteracy rate in 2025 is 1.51. Further, the results of the forecasting also reveal that over the next ten years there will a downward trend in illiteracy rates in Indonesia, with the average of the forecast points being 2.32 percent. This shows that continuity and commitment to the implementation of illiteracy eradication programs are required.

Keywords: ARIMA model, forecast, illiteracy, prediction, time series

INTRODUCTION

According to the 1945 Constitution of the Republic of Indonesia, Article 31 paragraph 1, every Indonesian citizen is entitled to education. This is continued in the National Education Law No. 20 of 2003 on the national education system, article 1 paragraph 1, which explains that education is a conscious and planned effort to create an atmosphere of learning, with learning processes that ensure learners are actively developing their potential for spiritual strength, self-control, personality and intelligence, noble character, as well as the skills that they needs, society, nation, and state. In addition to developing the potential of citizens, education is organized by developing a culture of reading, writing and counting for all citizens. A culture of reading and writing is expected to improve the literacy of citizens and reduce the illiteracy rate in Indonesia.

Based on data from the Ministry of Education and Culture from 2004, Indonesia had 15.41 million illiterate people, representing 10.2 percent of the adult population, of whom 64 percent were female (UNESCO, 2015). Reka et al. (2016) describe illiteracy as a person's lack of reading, writing and counting skills. It has many negative impacts, not only on personal development, but also in economic and social terms.

Ardila et al. (2010) state that illiteracy problems affect a significant proportion of the world's population. Illiteracy is not only related to cognition problems, but also to problems regarding knowledge of the world. There are two reasons why people may be illiterate: social factors (such as lack of schooling/absence of school facilities and poverty); and personal factors (such as learning difficulties, mental retardation, sensory issues). They also explained that illiteracy has a significant relationship with cognitive ability, especially in problem-solving. Sumardi (2012) states that the factors that affect illiteracy include the number of poor people; lack of access to education, especially in villages; low awareness of education, especially among women; and family economic needs. Life competition and a large number of family members cause children to drop out of school because school fees are unaffordable. To overcome the problem of illiteracy in Indonesia, the government launched a compulsory program stipulating at least 12 years' education (Republic of Indonesia government regulation No. 47 of 2008). This program strives for expansion and equalization of the opportunities for citizens to obtain quality education and to develop their potential to be able to live independently. The program plays a very important role in eradicating illiteracy in Indonesia.

The implementation of illiteracy eradication in Indonesia has been regulated in the regulation of the Minister of National Education No. 35 of 2006. It is an integrative and sustained government and community

movement, aimed at eliminating illiteracy at all levels of society, supporting the success of education in all programs, improving the ability and interest of the population to read and write, and supporting the quality of human resources. The eradication of illiteracy has become very important, since literacy is a basic right for everyone and is a key to opening up other basic rights. The problem of illiteracy is closely related to poverty, ignorance and helplessness. In addition, illiteracy affects the development of the nation; for example, low levels of community productivity; low awareness of educating children or the family; low ability to access information; difficulties in accepting innovation; and a low index of human development. To reduce illiteracy, in the strategy set forth in ministerial regulation No. 35 of 2006, the government states that there are three pillars of national education policy in the implementation strategy of the national movement for the acceleration of illiteracy eradication.

First, the extension of access to education includes the expansion of cross-sectoral cooperation (for example, institutions/agencies concerned), both at central and regional levels, in the implementation of the program of educational equality. The movement also aims to strengthen cooperation in the implementation of literacy education programs between universities, the technical unit of non-formal education, and various social organizations such as religious, women's, professional, and other community organizations/institutions. The aim is to become a deep-rooted movement in society, utilizing various potential resources available to the community to support the implementation of literacy education programs, and organizing literacy education programs in stages with priority on areas with the highest illiteracy population.

The second pillar is improvement of the quality of literacy education through the development and stipulation of standards of literacy competence and literacy education content, ranging from basic literacy, to ongoing literacy and independent literacy. There should also be development and establishment of valid and reliable literacy assessment instruments for literacy and content standards literacy education, and the implementation of quality assurance of literacy education at the learning group level so that learning process quality can reach the standard of literacy competence. Quality assurance covers improvement in resources and learning processes, such as

education and education personnel, teaching materials, instructional suggestions, innovation in learning strategies, and the cost of learning, together with strengthening of literacy education programs integrated with life skills education, so that the learning process is interesting, literacy skills by providing Community Reading Park in the village/district declared completely literate.

Third, the governance and accountability of literacy education should include improvement in the population reporting mechanism. Based on a survey by the Indonesia Statistic Center in 2017, the number of illiterate people in Indonesia each year has been decreased. This means that government programs intended to eradicate illiteracy in Indonesia have had success.

In handling illiteracy, the government requires predictions from year to year. The results of these surveys and predictions can be taken into consideration by the government in their effort to eradicate illiteracy and increase literacy in the country. The eradication of illiteracy and development of a literacy culture have a close relationship. Therefore, the literacy culture of citizens will grow and increase when illiteracy has disappeared. Sumardi (2012) explains that literacy has a major impact on social, economic and cultural improvements. It can be used as a guide by the government to improve social, economic and cultural levels in the community. This is because the purpose of literacy education is to achieve skills, good understanding, and adaptation to overcoming the problems of life and work challenges. It also explains why literacy education programs are promoted in an effort to eradicate illiteracy.

The results of research by the Program for International Student Assessment (PISA) on culture literacy in 2012 ranked Indonesia 64th out of 65 countries. In addition, the reading position of Indonesian students is 57th out of 65 countries. PISA stated that none of the Indonesian students had achieved a literacy value at the fifth level, and only 0.4 percent had fourth-level literacy ability. The remainder were below level three, or even below level one. Their poor results indicate a serious impact on the quality of the existing resource development. Especially in the field of education, this has an impact on efforts to improve the quality of education in Indonesia. Table 1 shows statistical literacy data on Indonesia compiled by Education For All (EFA).

Table 1 Literacy Data in Indonesia

Program	Year					
	2008	**2009**	**2010**	**2011**	**2012**	**2013**
Basic Literacy	600,010	1,088,890	437,830	347.017	200,000	188,320
Self-Entrepreneurship Literacy	139,140	194,000	210,030	208,000	300,000	303,000
Life Skill Education		7,000	7,000	7,000	7,000	7,000
Entrepreneurship Literacy			1,400	1,400	1,800	600
Children's newspapers, mothers' newspapers. Literacy based on Folktales			1,800	1,800	2,000	1,400

(Source: *Report on Education for All (EFA) Republic of Indonesia*, 2015)

Table 1 shows that literacy in Indonesia has increased in certain programs. The increased literacy results certainly cannot be separated from the eradication of illiteracy movement promoted by the government.

Therefore, it is important to investigate the illiteracy rates in Indonesia by supervising and controlling the movements. The purpose of this study is to predict the future rates of illiteracy in the country as reference material for overseeing this figure. In order to obtain the predicted values, the study applies the autoregressive integrated moving average (ARIMA) model of time series analysis, which is the most popular method of prediction analysis. However, this method requires stationary data to provide precise prediction results. Therefore, if the available data is not stationary, then a differencing process needs to be conducted so that the time series data meet the required criteria. Some studies on prediction analysis have used the ARIMA model. Floros (2005), Kurita (2010) and Nkwatoh (2012) employed this method to predict unemployment rates. Mahmudah (2017a) used it to forecast unemployment rates in Indonesia by using data from 1986 to 2015, with the results indicating that rates of unemployment would tend to decrease over the following ten years. Meanwhile, research on predicting illiteracy rate is very limited. As a matter of fact, it is very difficult to find references that focus on similar research. However, Jain and Mishra (2015) used multiple regression models to predict literacy rates in India. Their results provided very close predicted values to the actual literacy rate shown by the census of India.

METHOD

This study used the most well-known technique in forecasting time series data, which is the ARIMA model proposed by Box & Jenkins (1976), also known as the Box-Jenkins model. Stationary characteristics of data are required when using this method in order to obtain good prediction values. This means that the time series data are needed in prediction analysis laterally along the time axis. Stationary data indicate that fluctuations are not significant, meaning its values are always around the constant mean. According to Wei (2006), the ARIMA model can be written as follows:

$$y_t' = \varphi_1 y_{t-1}' + \cdots + \varphi_p y_{t-p}' + \theta_1 e_{t-1} + \cdots \theta_q e_{t-q} + e_t$$
$$(1)$$

where y_t' is the differenced series. Equation (1) indicates the ARIMA (p,d,q) model, where p represents the autoregressive order, q represents the moving average order and d represents the differencing process. It is important to point out that the autoregressive (AR) model requires a stationarity condition, while the moving average (MA) model needs invertibility conditions. Therefore, when these conditions are fulfilled, equation (1) can be rewritten in the following terms:

$$\left(1 - \varphi_1 B - \varphi_p B^p\right)(1 - B)^d y_t$$
$$= \left(1 + \theta_1 B + \cdots + \theta_q B^q\right)$$
$$+ e_t$$

$$(2)$$

where $\varphi_p(B)$ is a stationary AR operator and $\theta_q(B)$ is an invertible MA operator.

Generally, there are four steps in prediction analysis using the ARIMA model, which are explained as follows:

Step 1: Checking the stationary data
As stationary data are necessary in this model, it is important to identify whether the original data are stationary. The Augmented Dickey-Fuller test is usually used to check for such data.

Step 2: Differencing Process
If the Augmented Dickey-Fuller test fails, then a differencing process is required in order to obtain stationary data. This process continues until the desired characteristics are fulfilled.

Step 3: Choosing the best ARIMA model
This step is crucial because it greatly affects the final results of the prediction analysis in providing the future values. The best ARIMA model was selected from several different ones that may be applicable to predicting illiteracy rates in Indonesia. This study used the values of Akaike's Information Criterion (AIC), AIC-corrected (AICc) and the Bayesian Information Criterion (BIC) to determine the best ARIMA model which has the lowest values of the three criteria. However, it is important to note that the temporary ARIMA models that could be used are based on the values of the autoregressive order (*p*), the moving average order (*p*) and the differencing process (*d*).

Step 4: Forecasting
The best ARIMA model, which was determined in step 3, was used to predict illiteracy rates in Indonesia.

RESULTS AND DISCUSSION

This study uses 15 data of illiteracy percentage in Indonesia, based on the national socio-economic survey between 2003 and 2017 that was conducted by Statistics Indonesia. The study uses the R program to obtain the forecasting results of the illiteracy percentage in Indonesia using the ARIMA model. Table 3 shows the descriptive statistics of the data used to predict illiteracy rates. The data indicate the tendency of a decreasing percentage, even though there was an increase of 0.47 of a point in 2011. Moreover, illiteracy continued to decline significantly, with the highest decrease of 1.20 points occurring in 2014, and the lowest decrease in 2015 of 0.10 of a point. Based on the survey of the Indonesia Statistic Center, it can be explained that the smallest percentage of illiteracy is about 4.50, which occurred in 2017, whereas the highest percentage was 10.21 in 2003.

Table 2. Descriptive Statistics

Statistic	Illiteracy
N	15
Minimum	4.50
Maximum	10.21
Mean	7.16
Std. Deviation	1.86
Skewness	-0.09
Kurtosis	-1.04

Furthermore, Table 2 indicates that the data distribution tends to be on the left of the normal distribution due to negative value of skewness, with the value of kurtosis also negative, indicating that the distribution does not tend to peak. Figure 1 shows the illiteracy percentages in Indonesia from 2003 to 2017, clearly demonstrating a decrease over that period of time.
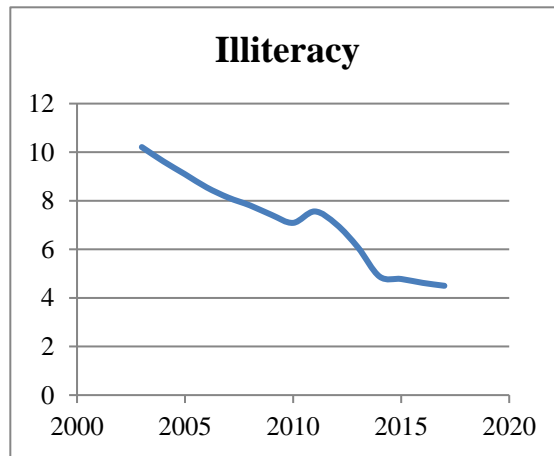

Figure 1. Percentage of Illiteracy in Indonesia

Figure 1 indicates that the original series of illiteracy percentages was clearly non-stationary, with a decreasing tendency clearly visible; there was a 0.59-point decrease in 2004 and a further decrease in 2005 of 0.53. The numbers continued to decrease until 2011, where the percentage of illiteracy in Indonesia increased by 0.47 of a point. Meanwhile, Figure 2 represents the plots of ACF and PACF of illiteracy data in Indonesia. Figure 2 shows that the ACF plot shows non-stationary properties, with the original data slowing down to zero.
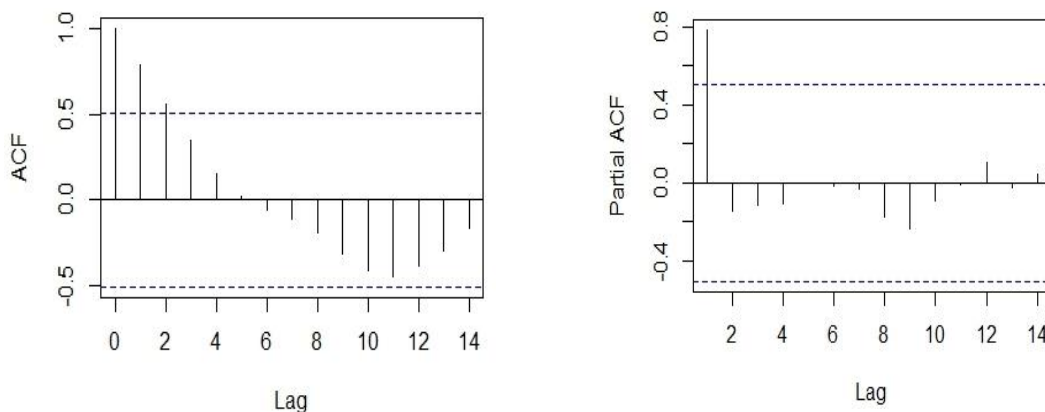

Figure 2. ACF and PACF of Illiteracy in Indonesia

Since the original data on illiteracy in Indonesia were non-stationary, in order to use the ARIMA model correctly a differencing process was conducted. The first process did not obtain stationary data of illiteracy in Indonesia as the value of the Dickey-Fuller test was -2.996 with lag order = 2, and p-value = 0.193. Due to the required properties of stationary data not being fulfilled, a second differencing process was required, whose results also indicated that stationary data had not been obtained (Dickey-Fuller = -3.3446; lag order = 2; p-value = 0.085). Therefore, a third process was required, whose results indicated that this process had produced stationary data (Dickey-Fuller = -3.6265; lag order = 2, p-value = 0.048). Therefore, the possible model for predicting illiteracy in Indonesia was defined by the ARIMA (p,3,q) model. Figure 3 shows the third process of differencing which provided stationary data on the percentage of illiteracy, while Figure 4 indicates the plots of ACF and PACF from this third process.
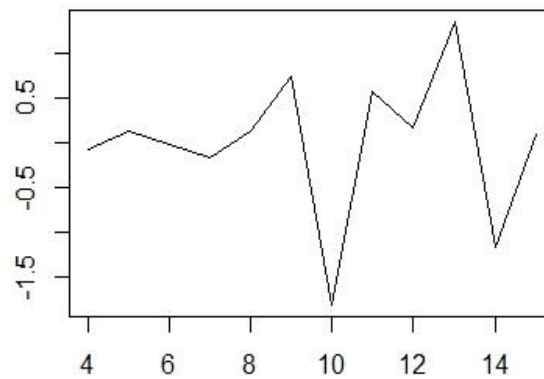
37

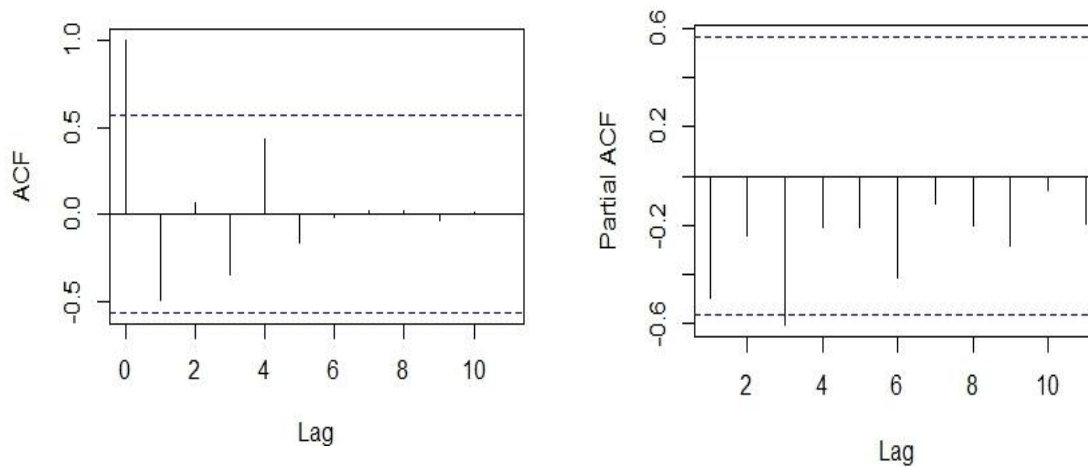**Figure 3. Percentage of Illiteracy (*d*=3)**

Figure 4. ACF and PACF of Illiteracy (*d*=3)

Figure 4 shows the PACF cutoff on lag 3, which indicates that the AR (3) model could be used for forecasting illiteracy. Furthermore, the ACF plot produces a cutoff in the first lag, so the MA (1) model could also possibly be used. Since the stationary data of illiteracy were determined through the third process of differencing, ARIMA (3,3,1) is a usable model. However, Table 4 provides temporary ARIMA models that may be applicable in this prediction analysis to determine the best ARIMA model. A summary of the results of these alternative models is presented in Table 3.

Table 3. Alternative ARIMA Models

ARIMA	SIGMA	LOG	AIC	ME	RMSE	MAE	MPE	MAPE
(3,3,0)	0.2361	-9.33	26.65	0.021	0.435	0.250	0.738	4.023
(3,3,1)	0.1484	-8.21	26.41	-0.035	0.345	0.212	-0.463	3.382
(2,3,0)	0.4343	-12.21	30.42	0.013	0.589	0.335	0.532	5.755
(2,3,1)	0.2385	-9.98	27.97	0.001	0.437	0.269	0.282	4.541
(1,3,0)	0.4659	-12.56	29.13	-0.004	0.611	0.387	0.179	6.762

In order to obtain the best ARIMA model, the study used the AIC criterion, which is the most commonly used accuracy measure in forecasting analysis. The lowest value of the AIC criterion was determined as the best model for predicting illiteracy in Indonesia. According to Table 3, ARIMA (3,3,1) has the lowest value of AIC, which is 26.41. Therefore, based on Mahmudah (2017a, 2017b) this model was determined as the best one to predict illiteracy rate in Indonesia because the lowest value of AIC tends to provide the best model of ARIMA. However, the values of other accuracy measures for prediction analysis are also presented in Table 3, namely the mean error (ME), root mean squared error (RMSE), mean absolute error (MAE), mean percentage error (MPE), and mean absolute percentage error (MAPE). Plotting the original data series and fitted model is one way to conduct the ARIMA model validation, but the most common method to determine the best model is simply by observing the values of the accuracy measures (Mahmudah, 2017b). Figure 5 shows the

fitted values and the original series of illiteracy data in Indonesia from the ARIMA (3,3,1) model.
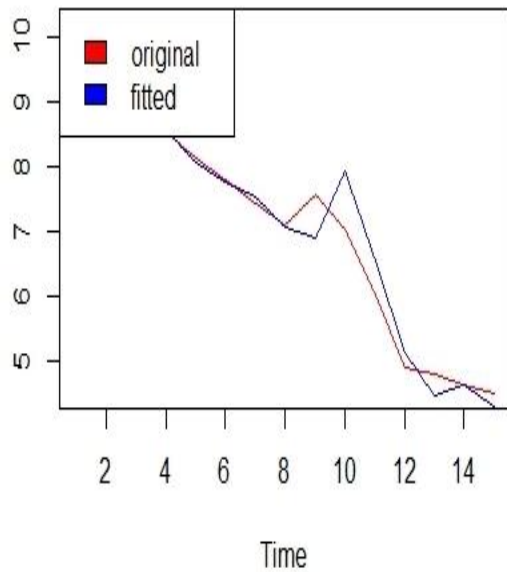

Figure 5. Original Series and Fitted Values

Figure 5 also indicates that the fitted values from ARIMA (3,3,1) always follow the original series, with the fitted model represented by the blue line, while the original data are represented by the red line.

Furthermore, Table 4 shows the forecasting results by using ARIMA (3,3,1) as the best model for predicting the percentage of illiteracy in Indonesia based on illiteracy data from previous years.

Table 4. Forecasting Results

Year	Point Forecast	Forecast Interval			
		Lo80	Hi80	Lo95	Hi95
2018	3.78	3.27	4.30	2.99	4.57
2019	3.32	2.28	4.36	1.73	4.91
2020	2.96	1.37	4.55	0.53	5.40
2021	2.86	0.79	4.93	-0.31	6.02
2022	2.51	-0.13	5.16	-1.53	6.56
2023	2.14	-1.20	5.48	-2.96	7.24
2024	1.74	-2.42	5.89	-4.63	8.10
2025	1.51	-3.50	6.53	-6.16	9.18
2026	1.28	-4.63	7.20	-7.76	10.33
2027	1.04	-5.83	7.92	-9.47	11.56

Based on data in Table 4, it is apparent that over the next 10 years there will be a continuous downward trend in the percentage of illiteracy in Indonesia, where the illiteracy rate has a tendency to decrease consistently. In other words, literacy rates tend to go up continuously. This trend will lead to good results, which indicates the reading and writing skills of Indonesian citizens are developing. These findings are consistent with what has been found in previous studies on predicting illiteracy rates.

A similar result was obtained by Jain and Mishra (2015) who report literacy rate in India is predicted to rise continuously, where the predicted rate is very close to the actual rate. Other than that, UNESCO (2017) also reports literacy rates are increasing continuously from one generation to the next globally. For example, literacy rate increase in most of the Southern Asia.

Figure 6 shows the plot of forecasting results from the ARIMA (3,3,1) model.
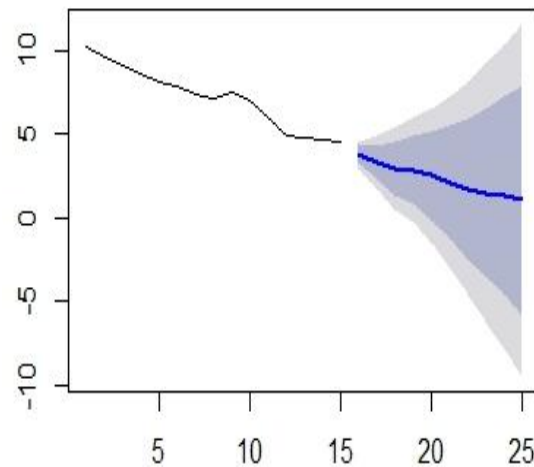
Forecasts from ARIMA(3,3,1)

Figure 6. Plot of Forecast Results

Figure 6 shows the forecast values, which are represented by the blue line, while the dark shaded area indicates the 85th prediction interval and the bright shaded area shows the 90th prediction interval. From the figure it can be clearly seen that the forecasting results decrease continuously. Further, Table 5 shows both the forecast values and forecast intervals for the next ten periods of illiteracy percentages in Indonesia. The prediction intervals play a very important role in forecasting the future values because a prediction analysis cannot yield the accuracy of the predicted values without being accompanied by these interval values of prediction.

In other words, the prediction intervals give an idea of the uncertainty of the outcome of the forecasting analysis, therefore we can see clearly the level of uncertainty of each prediction value. Two prediction intervals are commonly used in forecasting analysis, namely 80% and 95% even though any other prediction interval could also be applied. This study provides the fitted model from both of these prediction intervals, with the values of the prediction intervals presented in Table 5. It is important to note that when the prediction analysis yields higher uncertainty, the prediction intervals produce wider values. Table 5 also shows the lower and upper boundaries of both the 85% and 95% prediction intervals, with the ranges of both prediction intervals

tending to increase in line with the predicted time. In addition, the 95th prediction intervals produce wider ranges than the 80th ones. Figure 7 indicates the ranges for both prediction intervals.
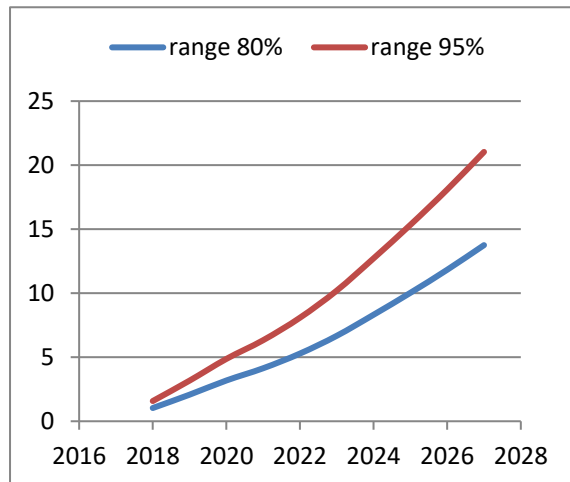

Figure 7. Range of Prediction Intervals

Furthermore, from Table 5 it can be seen that the highest percentage of illiteracy in Indonesia was expected at around 3.78 in 2018, with the lowest value of predicted percentage of 1.04 in 2027. The highest decrease occurred from 2018 to 2019, when the percentage of illiteracy in Indonesia was expected to fall by 0.46 percent, while the lowest decrease is expected to be 0.11 percent from 2020 to 2021. In general, although the results of forecasting illiteracy in Indonesia drop constantly, there are fluctuations in the values of the decline. Further, the average of the forecast points for the next ten years is 2.32 percent, with a standard deviation of 0.92 percent.

The illiteracy eradication improvement will have a good impact on the government in realizing an education program for all, enabling the government to increase community productivity and the human development index. This argument is as described in the Regulation of the Minister of National Education No. 35 of 2006 that illiteracy eradication is an integrative and sustained government and community movement, aimed at eradicating illiteracy at all levels of society, supporting the success of education for all programs, improving the ability and interest of the population to read and write, and supporting the quality of human resources.

Oyekunle (2018) explains that illiteracy eradication in order to realize a literate society will enable society to contribute to the psycho-economic and cultural development. Therefore, illiteracy eradication has become the focus of the government to give the rights to education to all people.

CONCLUSION AND RECOMMENDATIONS
The results indicate that the original series of illiteracy percentages has non-stationary properties; therefore, differencing processes were employed. The study needed a third process to obtain stationary illiteracy data. Further, the results also suggest that the best model for predicting the percentage of illiteracy in Indonesia was the ARIMA (3,3,1) model. The forecasting results show a decreasing tendency in the forecast values. A recommendation for further research is that more data used for predictive results will be more accurate.

REFERENCES
Ardila, A., et al. (2010). Illiteracy: The neuropsychology of cognition without reading. *Archives of Clinical Neuropsychology 25, 689-712.* https://www.ncbi.nlm.nih.gov/pubmed/21075867. doi: 10.1093/arclin/acq079.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control.* Revised edition. San Francisco: Holden-Day.

Education for All (EFA). (2015). *A report on Education for All (EFA) Achievement period 2000-2015.* Education for All Coordinator Forum Republic of Indonesia.

Floros, C. (2005). Forecasting the UK unemployment rate: Model comparisons. *International Journal of Applied Econometrics and Quantitative Studies*, *2*(4), 57-72.

Jain, S. & Mishra, N. (2015). Forecasting of literacy rate using statistical and data mining methods. *International Journal of Advanced Computational Engineering and Networking*, 3(8), 26-31.

Kurita, T. (2010). A forecasting model for Japan's unemployment rate. *Eurasian Journal of Business and Economics*, *3*(5), 127-134.

Mahmudah, U. (2017a). predicting unemployment rates in Indonesia. *Economic Journal of Emerging Markets*, 9(1), 20-28.

Mahmudah, U. (2017b). Autoregressive integrated moving average model to predict graduate unemployment in Indonesia. *Practice and Theory in System of Education*, *12*(1), 43-50. DOI 10.1515/ptse-2017-0005

Nkwatoh, L. S. (2012). Forecasting unemployment rates in Nigeria using univariate time series models. *International Journal of Business and Commerce*, *1*(12), 33-46.

Oyekunle, Y. 2018. Eradicating rural illiteracy through adult education programmes in Lagos State, Nigeria. *Journal of Educational Thought*, *7*(1): 136-150.

Statistics Indonesia. (2017). https://www.bps.go.id/subject/28/pendidikan.html#subjekViewTab3

Suwardi, K. (2012). Adult education through multiple method for poor rural illiterate women in Indonesia. *Journal of Education and Learning, 6*(6), 243-252.

UNESCO. (2015). *Indonesia literacy for life skills and entrepreurship.* UNESCO Bangkok Office.

UNESCO. (2017). *Literacy rates continue to rise from one generation to the next.* Fact Sheet No. 45 September 2017 FS/2017/LIT/45. http://uis.unesco.org/sites/default/documents/fs45-literacy-rates-continue-rise-generation-to-next-en-2017_0.pdf

Vagvolgyi, R. et al. (2016). *A Review about Funcional Illiteracy: Definition, Cognitive, Linguistic, and Numerical Aspects.*

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5102880/. doi: 10.3389/fpsyg.2016.01617

Wei, W. S. W. (2006). *Time series analysis unvariate and multivariate methods* 2nd Edition. Boston, United State of America: Pearson Addision-Wesley Publishing Company.