Available online at https://journal.rescollacomm.com/index.php/ijeer/index

International Journal of Ethno-Sciences and Education Research

e-ISSN 2776-7590

Vol. 2, No. 4, pp. 152-160, 2022

Analysis of Average Length of Schooling of Indonesian Citizens in the Future Using Markov Chains

Riza Andrian Ibrahim[1*], Sivaperumal Sampath[2]

[1]*Doctoral of Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Bandung, Indonesia*

[2]*Faculty of Technology, VelTech University, India*

**Corresponding author email: riza17005@mail.unpad.ac.id*

Abstract

Based on data from the Central Bureau of Statistics of the Republic of Indonesia, the average length of schooling of Indonesian citizens is continuously increasing every year. This increase is predicted to continue in the future. However, it has yet to be discovered how the increasing rate in the average length of schooling will occur. Therefore, this study aims to analyze the increasing rate of the average length of schooling of Indonesian citizens in the future. The data used is the average length of schooling of Indonesian citizens from 2010 to 2021. The analytical method used is the discrete-time Markov chain. Furthermore, the states representing the increasing rate in the average length of schooling used are divided into two: the increasing rate in the average length of schooling that is smaller and larger than the average. Based on the analysis results, the probability that the increasing rate in Indonesian citizens' average length of schooling will be less than the average in the future is 0.4. In contrast, the probability that the increasing rate in Indonesian citizens' average length of schooling will be greater than the average in the future is 0.6. It indicates that Indonesian citizens will have a high education level in the future. The results can be used as the future education state projection of Indonesian citizens so that it can be accompanied by empowerment.

Keywords: the average length of schooling, Indonesia, the increasing rate, Markov chain, education level

1. Introduction

The Ministry of Education and Culture, Research and Technology of the Republic of Indonesia continues to make policies to increase the average length of schooling of Indonesian citizens. It is conducted to increase the human development index (HDI) in the education sector (Berliyanto and Santoso, 2018). The policies implemented include reducing the dropout rate, increasing access to and quality of education, increasing the competitiveness of higher education, and improving the quality of educators and education staff (Deffinika et al., 2022).

Visualization of the average length of schooling of Indonesian citizens from 2010 to 2021 based on data from the Central Bureau of Statistics of the Republic of Indonesia can be seen in Figure 1. Figure 1 shows that the average length of schooling of Indonesian citizens is continuously increasing. This increase is even predicted to continue in the future (Zulfa and Meutia, 2018). This increase is the policy results that have been implemented. In addition, the increasing rate in the average length of schooling for Indonesian citizens from 2010 to 2021 is also relatively high. Based on historical data, the average increasing rate in the average length of schooling from 2010 to 2021 is 0.0982. The visualization of the increasing rate in the average length of schooling from 2010 to 2021 is presented in Figure 2. Figure 2 shows the increasing rate in the average length of schooling is around the average line. It indicates that the increasing rate in the average length of schooling is stationary. In addition, most increasing rates are above the average. It is good news for the Indonesian government.

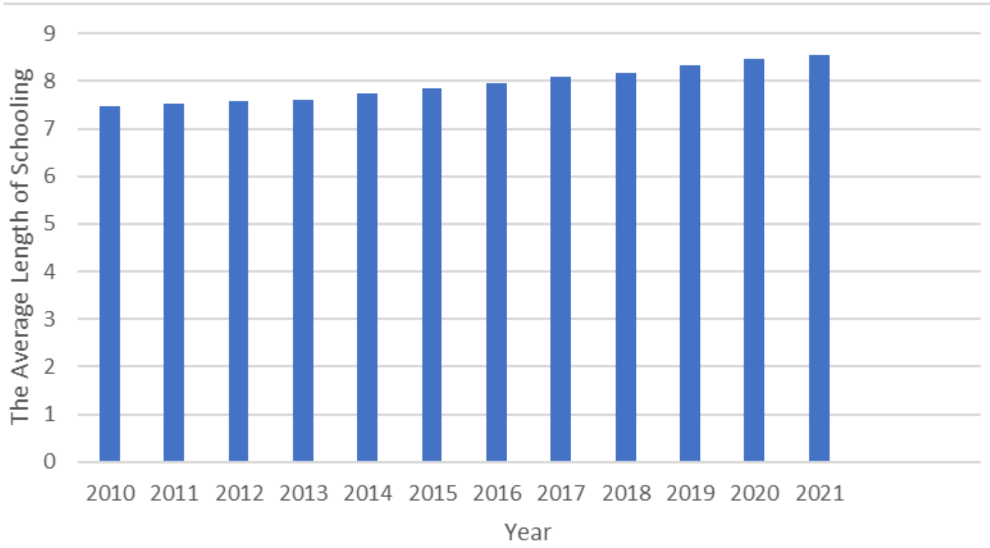

Figure 1: The Average Length of Schooling of Indonesian Citizens from 2010 to 2021

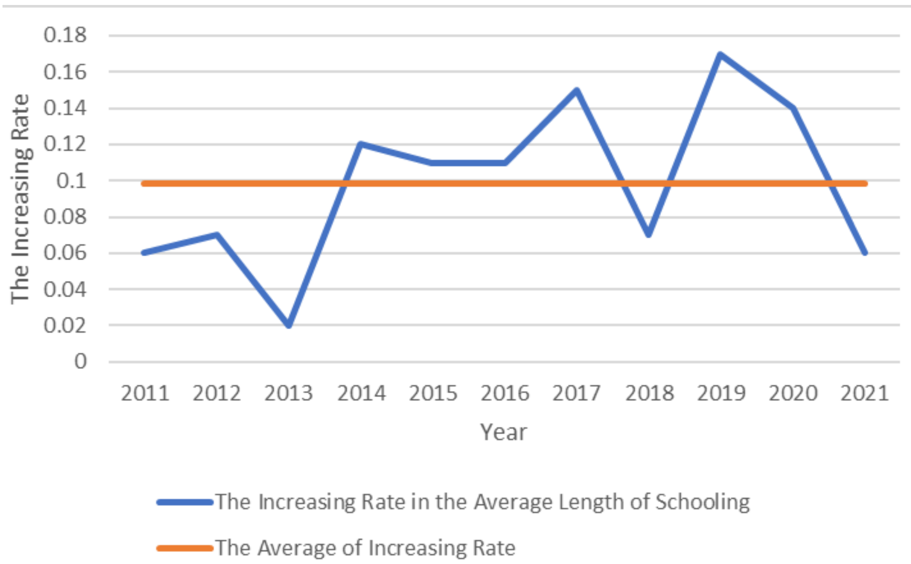

Figure 2: The Increasing Rate in the Average Length of Schooling of Indonesian Citizens from 2010 to 2021

Several articles have examined the average length of schooling in Indonesia. Zulfa and Meutia (2018) analyze the effect of economic growth and HDI on West Java Province, Indonesia. Kustandi et al. (2021) applied the K-Means clustering method to classify Indonesia's average schooling length. Then, Juned and Yusra (2021) used the fuzzy C-means clustering method to classify HDI in Aceh Province, Indonesia. Finally, Saepudin (2018) examines the development of the electrification ratio with HDI in West Java Province, Indonesia.

Although it is predicted that the average length of schooling of Indonesian citizens will continue to increase in the future, the increasing rate of it has yet to be discovered. Based on the previous articles and presented introduction, this is the first article to study this. Therefore, this study aims to analyze the increasing rate in the average length of schooling of Indonesian citizens in the future. The method used to analyze this is the discrete-time Markov chain. The discrete-time Markov chain is used since this method can describe the probability of a situation occurring in the future. The states of increasing rate in the average length of schooling used for Markov chain analysis are divided into two: the increasing rate in the average length of schooling that is smaller and larger than the average. This research is expected to provide an overview for the Indonesian government in projecting the average length of schooling for its population so that they can prepare policies regarding it precisely.

2. Materials and Methods

2.1. Materials

The data used in this study is data on the average length of schooling for Indonesian citizens from 2010 to 2021, accessed on September 20, 2022. This data can be accessed openly on the Central Bureau of Statistics of the Republic of Indonesia website as follows: https://www.bps.go.id.

2.2. Methods

In this section, an explanation of the supporting theory is briefly explained. The supporting theories are as follows: discrete-time Markov chain analysis, transition probability matrix, transition probability diagrams, irreducible Markov chains, recurrent Markov chains, aperiodic Markov chains, positive recurrent Markov chains, Ergodic Markov chains, stationary distributions, and the Chapman-Kolmogorov equation.

2.2.1. The Discrete-Time Markov Chain Analysis

Markov chain analysis was introduced by a Russian mathematician named Andrei A. Markov in 1906. Via the Markov chain, information about the probability of a situation occurring in the future can be known (Susilo et al., 2019). This information can be used in decision-making (Duys and Headrick, 2004).

Based on the spaces of state and parameter, Markov chains are divided into two, namely discrete and continuous time Markov chains. A discrete-time Markov chain is a Markov chain with discrete spaces of state and parameter, while a continuous-time Markov chain is a Markov chain with continuous spaces of state and parameter. The discrete-time Markov chain is a particular form of a stochastic process $\{X_t, t = 0, 1, 2, \dots\}$ with a state space $k = \{0, 1, \dots, q\}$. The probability of the transition from $X_t = i$ to $X_{t+1} = j$, denoted by p_{ij}, is determined by the following equation:

$$p_{ij} = P\{X_{t+1} = j | X_0 = k_0, X_1 = k_1, \dots, X_{t-1} = k_{t-1}, X_t = i\} = P\{X_{t+1} = j | X_t = i\}, \quad (1)$$

for all j, t, i, k_1, k_2, …,$k_{t-1} \in k$. In detail, p_{ij} in equation (1) is referred to as the one-step transition probability from the state i to j (Osaki, 1972, p. 105). A verbal explanation of Equation (1) is given to make it easier to understand. The probability that state j will occur at time $(t + 1)$ is only affected by state i at time t. In other words, the probability that event j at time $(t + 1)$ is affected only by one previous step (Ogunnaike et al., 2018).

2.2.2. Transition Probability Matrix

The probability of each state transition can be expressed as a matrix. This matrix is called the transition probability matrix. The one-step transition probability matrix of the stochastic process X_t is expressed as follows (Ross, 1996, p. 163):

$$\boldsymbol{P} = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0q} \\ p_{10} & p_{11} & \cdots & p_{1q} \\ \vdots & \vdots & \ddots & \vdots \\ p_{q0} & p_{q1} & \cdots & p_{qq} \end{bmatrix}, \quad (2)$$

where $0 \le p_{ij} \le 1$ and $\sum_{j=0}^{q} p_{ij} = 1$. In general, t-steps transition matrix can be expressed as a one-step transition matrix raised to the power of t. Mathematically, it can be written as follows (Ross, 1996, p. 168):

$$\boldsymbol{P}^{(t)} = \left[p_{ij}^{(t)}\right] = \boldsymbol{P}^t, \quad (3)$$

where $\boldsymbol{P}^{(t)}$ represents t-steps transition matrix.

2.2.3. Transition Probability Diagrams

The probability of each state transition can be expressed in graph form. Vertices represent the state space, and each directed arrow weight represents each transition probability. This graph is called a probability transition diagram. Suppose that there are two state spaces of the Markov chain X_t, namely $k = \{0, 1\}$. The transition probability diagram of the X_t stochastic process can be seen in Figure 3.

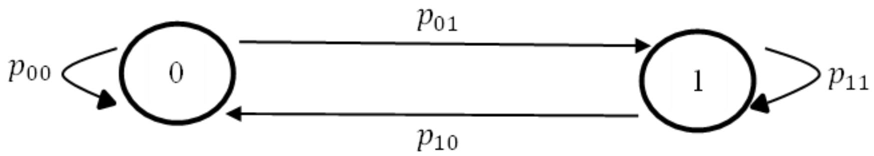

Figure 3: Transition Probability Diagrams of Markov Chain X_t with State Space $k = \{0, 1\}$

Figure 1 shows that the state space $k = \{0,1\}$ is represented as a vertex, while the transition probability between states is described as the weight of a directional arrow. This transition probability diagram can facilitate the Markov chain analysis process (Tsai et al., 2014).

2.2.4. Irreducible Markov Chain

State j can be reached from state i, denoted by $i \to j$, if there is a positive integer t so that the transition probability t-step from state i to state j is positive, $p_{ij}^t > 0$ (Ross, 1996, p. 168). In general, if state j can be reached from state i, and state i can be reached from state j, then state i and state j are called two states that communicate with each other. It is denoted as $i \leftrightarrow j$. If all states in a Markov chain communicate with each other, then the Markov chain is called an irreducible Markov chain.

2.2.5. Recurrent Markov Chain

A state in a Markov chain is called a recurrent state if it returns to its initial state when it transitions to any state. Suppose the Markov chain X_t has a state space $k = \{0,1,2\}$. An illustration of the recurrent state in the X_t Markov chain is given in Figure 4.

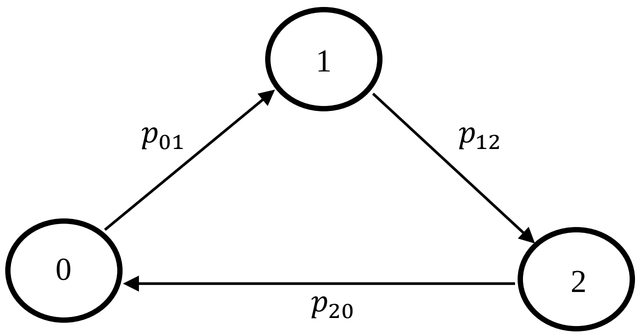

Figure 4: Recurrent State Illustration in Markov Chain

Based on Figure 2, if states 0, 1, or 2 transit anywhere, these states will eventually reach their original state. It shows that states 0, 1, and 2 are recurrent states. It is not recurrent if a state cannot reach its original state after making a transition. This situation is called a transient state. If every state in a Markov chain is recurrent, then the Markov chain is called a recurrent Markov chain (Osaki, 1972, p. 114).

2.2.6. Aperiodic Markov Chain

A state in a Markov chain is called aperiodic if all the arrows pointing to that state are multiples of one in length. Mathematically, state i is called aperiodic if the following equation applies (Ross, 1996, p. 169):

$$d(i) = \gcd\{t > 1 | p_{ii}^t > 0\} = 1, \quad (4)$$

where $d(i)$ represents of the period of state i, and $\gcd\{t > 1 | p_{ii}^t > 0\}$ represents *greet common divisor* of transition probability of state i to itself in t-steps, $t > 1$. If every state in the Markov chain is aperiodic, then the Markov chain is called an aperiodic Markov chain (Osaki, 1972, p. 115).

2.2.7. Positive Recurrent Markov Chain

A state is said to be positive recurrent if the average time it takes for the state to reach itself the first time has a finite value. Mathematically, state i is called positive recurrent if the following conditions apply (Ross, 1996, p. 173):

$$\mu_i = \sum_{t=1}^{\infty} t f_{ii}^t < \infty, \tag{5}$$

where μ_i represents the average time state i takes to reach itself the first time, and f_{ii}^t represents the probability that state i reaches itself for the first time in t-steps. If every condition in a Markov chain is positive recurrent, then the Markov chain is called a positive recurrent Markov chain (Osaki, 1972: p. 118).

2.2.8. Ergodic Markov Chain

To determine the long-run probability of a state in a Markov chain, the condition that must be met is that the Markov chain must be ergodic. A Markov chain is said to be ergodic if it is irreducible, aperiodic, and positive recurrent (Ross, 1996, p. 177).

2.2.9. The Stationary Distribution

The probability of a long-term transition from a state is the probability that that state will occur in the long term. The set of long-run probabilities for each state is called the stationary distribution of an ergodic Markov chain. Mathematically, the long-term probability of state j from the Markov chain X_t with state space $k = \{0, 1, 2, \dots, q\}$ is expressed as follows (Ross, 1996, p. 175):

$$\pi_j = \lim_{t\to\infty} p_{ij}^t;\ j = 0, 1, 2, \dots, q \tag{6}$$

where π_j represents the long-term probability of state j. $A = \{\pi_j, j = 0,1,2, \dots, q\}$ is called as stationary distribution of an ergodic Markov chain X_t.

2.2.10. Chapman-Kolmogorov Equation

The Chapman-Kolmogorov equation was introduced by British mathematician Sydney Chapman and Russian mathematician Andrey Kolmogorov. The Chapman-Kolmogorov equation calculates the probability of the t-step transition for each state in the Markov chain. Mathematically, the Chapman-Kolmogorov equation can be expressed as follows (Ross, 1996, p. 167):

$$p_{ij}^{(t)} = \sum_{k=1}^{\infty} p_{ik}^{(r)} p_{kj}^{(t-r)}, \tag{7}$$

where $p_{ij}^{(t)}$ represents the probability of the t-step transition from state i to state j, $p_{ik}^{(r)}$ represents the probability of the r-step transition from state i to state k, $p_{kj}^{(t-r)}$ represents the probability of the $(t-r)$-step transition from state k to state j, and $r < t$. If equation (7) is expressed in matrix form, then the following equation is obtained (Ross, 1996, p. 168):

$$\boldsymbol{P}^{(t)} = \boldsymbol{P}^{(r)}\boldsymbol{P}^{(t-r)} = \boldsymbol{P}^{r}\boldsymbol{P}^{t-r} = \boldsymbol{P}^{t}. \tag{8}$$

The stationary distribution $A = \{\pi_j, j = 0, 1, 2, \dots, q\}$ of the ergodic Markov chain X_t with the state space $k = \{0, 1, 2, \dots, q\}$ can be obtained by multiplying the transition probability matrices as much as possible until the convergence of the probability values for each column is obtained.

3. Results and Discussion

The process of Markov chain analysis on the average length of schooling data for Indonesian citizens is as follows:

(a) calculating the increasing rate in the average length of schooling each year from 2010 to 2021,
(b) descriptive statistical analysis of the data on the increasing rate in the average length of schooling,
(c) determination of the state of the Markov chain,
(d) determination of the matrix and diagram of the transition probability of the Markov chain,
(e) checking ergodicity of the Markov chain, and
(f) determining the stationary distribution or long-term probability of each state of the Markov chain using the Chapman-Kolmogorov equation.

3.1. Calculation of the Increasing Rate in the Average Length of Schooling in Every Year

The increasing rate in the average length of schooling in a year is the difference between the average length of schooling that year and the previous year. The calculating result of the increasing rate in the average length of schooling are given in Table 1.

Table 1: The Calculating Result of the Increasing Rate in the Average Length of Schooling

Year	the Increasing Rate in the Average Length of Schooling (Per Year)
2011	0.06
2012	0.07
2013	0.02
2014	0.12
2015	0.11
2016	0.11
2017	0.15
2018	0.07
2019	0.17
2020	0.14
2021	0.06

Table 1 shows the most significant increase rate in the average length of schooling occurred in 2017, which is 0.15 per year. Then, the smallest increase rate in the average length of schooling occurred in 2011 and 2021, which is 0.06 per year. The descriptive statistics are given in Table 2 to find out more about the increasing rate in the average length of schooling in Table 1.

Table 2: The Descriptive Statistics of the Increasing Rate in the Average Length of Schooling

Descriptive Statistics	Value
Average	0.0982
Variance	0.0021
Deviation Standard	0.0458

Table 2 shows that the average increasing rate in the average length of schooling is 0.0982 per year. Meanwhile, the average deviation from the increasing rate in the average length of schooling is 0.0458.

3.2. Determination of the States of the Markov Chain

The states used in the Markov chain analysis are the increasing rate in the average length of schooling, which is smaller and larger than the average. For example, the state with the increasing rate in the average length of schooling, which is smaller than the average, is declared state 0, and the other is state 1. The classification of the increasing rate in the average length of schooling based on 0 and 1 states is given in Table 3.

Table 3: The Classification of The Increasing Rate in the Average Length of Schooling based on 0 and 1 States

Year	The Increasing Rate in the Average Length of Schooling (Per Year)	State
2011	0.06	0
2012	0.07	0
2013	0.02	0
2014	0.12	1
2015	0.11	1
2016	0.11	1
2017	0.15	1
2018	0.07	0
2019	0.17	1
2020	0.14	1
2021	0.06	0

Table 3 shows that the transition frequency from state 0 to state 0 is two. Then, the transition frequency from state 0 to state 1 is one. Then, the transition frequency from state 1 to state 0 is two. Finally, the transition frequency from state 1 to state 1 is four.

3.3. Determination of the Matrix and Diagram of the Transition Probability of the Markov Chain

The transition frequency of all states to the other states is needed to determine the matrix and the transition probabilities. Regarding Table 3, the transition frequencies between states 0 and 1 are given in Table 4.

Table 4: The Transition Frequency of All States to the Other States

Transition between States	Frequency
Transition from State 0 to State 0	2
Transition from State 0 to State 1	2
Transition from State 1 to State 0	2
Transition from State 1 to State 1	4

After the transition frequencies between states are obtained in Table 4, the transition probabilities between states are determined. The transition probability between states is determined based on the division between the transition frequency from one state to a particular state and the transition frequency from one state to each state. The transition probabilities between these states are given in Table 5.

Table 5: The Transition Probabilities between States

The Transition Probabilities between States	Value
p_{00}	$\frac{1}{2}$
p_{01}	$\frac{1}{2}$
p_{10}	$\frac{2}{6}$
p_{11}	$\frac{4}{6}$

Table 5 shows that the most significant transition probability is the transition probability from state 1 to state 1. In contrast, the most negligible transition probability is the transition probability from state 0 to state 0 and from state 0 to state 1. Note that the sum of p_{00} and p_{01} is one, and the sum of p_{10} and p_{11} is also one. Via equation (2), the transition probability matrix in this study can be expressed in the following equation:

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{2}{6} & \frac{4}{6} \end{bmatrix}. \tag{9}$$

Visualization of the transition probability diagram is given in Figure 5.

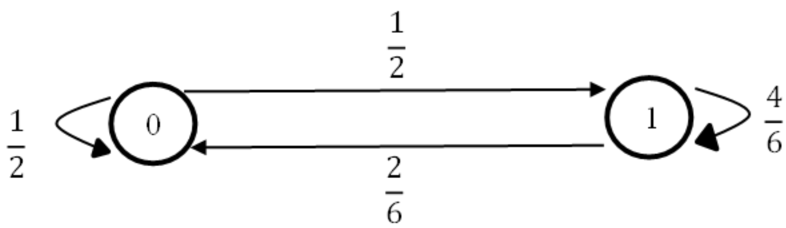

Figure 5: The Transition Probability Diagram in This Study

3.4. Checking the Ergodicity of the Markov Chain

3.4.1. Checking the Irreducibility of the Markov Chain

Since there is $t = 1$ so that p_{01} and p_{10} are positive, 0 and 1 states are said to communicate with each other. Since each state communicates with the other, the Markov chain in this study is said to be irreducible..

3.4.2. Checking the Aperiodicity of the Markov Chain

Based on equation (4), since there are $t = 2, 3, 4, ...$, such that p_{00}^t and p_{11}^t are positive, the period of states 0 and 1, denoted by $d(0)$ and $d(1)$, is 1. Then, states 0 and 1 are aperiodic. Since every state is aperiodic, the Markov chain in this study is said to be aperiodic.

3.4.3. Checking the Recurrent Positiveness of the Markov Chain

The average recurrent times of states 0 and 1, which are calculated by equation (5), are respectively as follows:

$$\mu_0 = \sum_{t=1}^{\infty} t f_{00}^{t} = 1\left(\frac{1}{2}\right) + 2\left(\frac{1}{2}\right)\left(\frac{2}{6}\right) + 3\left(\frac{1}{2}\right)\left(\frac{4}{6}\right)\left(\frac{2}{6}\right) + 4\left(\frac{1}{2}\right)\left(\frac{4}{6}\right)^2\left(\frac{2}{6}\right) + \cdots = \frac{1}{2} + \frac{1}{6}\sum_{m=2}^{\infty} m\left(\frac{2}{3}\right)^{m-2} \approx \frac{5}{2} \quad (10)$$

and

$$\mu_1 = \sum_{t=1}^{\infty} t f_{11}^{t} = 1\left(\frac{4}{6}\right) + 2\left(\frac{2}{6}\right)\left(\frac{1}{2}\right) + 3\left(\frac{2}{6}\right)\left(\frac{1}{2}\right)^2 + 4\left(\frac{2}{6}\right)\left(\frac{1}{2}\right)^3 + \cdots = \frac{2}{3} + \frac{1}{3}\sum_{m=2}^{\infty} m\left(\frac{1}{2}\right)^{m-1} \approx \frac{5}{3}. \quad (11)$$

Since the average recurrent time for states 0 and 1 is finite, states 0 and 1 are said to be positively recurrent. Since every state is positive recurrent, the Markov chain in this study is also said to be positive recurrent .

3.4.4. Checking the Ergodicity

Since the Markov chain in this study is irreducible, aperiodic, and positively recurrent, it is said to be ergodic.

3.5. Determining the Stationary Distribution

Determination of the stationary distribution of the Markov chain in this study is carried out using equation (8). By multiplying the transition probability matrix $\boldsymbol{P}$ as much as possible, for example, the 500-step transition probability matrix, which is also a stationary distribution of the Markov chain in this study, is as follows:

$$\boldsymbol{P}^{(500)} = \begin{bmatrix} 0.4 & 0.6 \\ 0.4 & 0.6 \end{bmatrix}. \quad (12)$$

In other words, the stationary distribution of the Markov chain in this study is as follows:

$$A = \{\pi_0 = 0.4, \pi_1 = 0.6\}. \quad (13)$$

Verbally, the stationary distribution in Equation (13) shows that the long-term probability that the increasing rate in the average length of schooling will be less than the average is $0.4 = 40\%$. In comparison, the long-term probability that the increasing rate in the average length of schooling will be greater than the average is $0.6 = 60\%$. It can happen with the assumption that Indonesia's condition is stable. Another meaning of this is that the education level of Indonesian citizens in the future will be even higher.

4. Conclussion

This research examines how fast the increasing rate of Indonesian citizens' average length of schooling will be. This research needs to be conducted for the future education state projection of Indonesian citizens. This projection will later be used as a basis for preparation for issuing supporting policies.

The experimental data in this study is the average length of schooling of Indonesian citizens from 2010 to 2021. The data is analyzed using a discrete-time Markov chain. Furthermore, two states are applied, namely, the increasing rate in the average length of schooling, which is smaller and larger than the average.

The analysis results show that the probability that the increasing rate in Indonesian citizens' average length of schooling will be less than the average in the future is 0.4. In contrast, the probability that the increasing rate in Indonesian citizens' average length of schooling will be greater than the average in the future is 0.6. It indicates that Indonesian citizens will have a high level of education in the future. These results can be used as a future education state projection of Indonesian citizens so that a high level of public education can be accompanied by empowerment.

References

Berlianto, and Santoso, H. B. (2018). Indonesian Perspective on Massive Open Online Courses: Opportunities and Challenges. *Journal of Educators Online, 15*(1), 1-14.

Deffinika, I., Putri, I.W., and Angin, K. B. (2022). Higher Education and Training Towards Global Competitiveness and Human Development in Indonesia. *Geojournal of Tourism and Geosites*, *40*(1), 1280-1288.

Duys, D. K., and Headrick, T. C. (2004). Using Markov Chain Analyses in Counselor Education Research. *Counselor Education and Supervision*, *44*(2), 108-120.

Zulfa, A., and Meutia, R. (2018). The Influence of Economic Growth Per Capita and the Average Length Of Schooling on the Human Development Index in West Java. *Indian Journal of Public Health Research and Development*, *9*(12), 1826-1831.

Juned, V., and Yusra. (2021). Using Fuzzy C-Means Algorithm to Cluster Human Development Index. *Communications in Mathematical Biology and Neuroscience*, *2021*(5), 1-15.

Kustandi, C., Bhawika, G. W., Susanto, E., Van Harling, V. N., and Ekawati, A. D. Implementation of the K-Medoids Method in the Old School Expectations in Indonesia by Utilizing Educational Data Mining. *ARPN Journal of Engineering and Applied Sciences*, *16*(17), 1811-1816.

Ogunnaike, O. O., Bishop, S. A., Akinsulire, H., Kehinde, B. E., and Oreagba, O. T. (2018). Education For Sale: Markov Chain Analysis of Physical Qualities and Engineering Students Switching Behaviour. *International Journal of Mechanical Engineering and Technology*, *9*, 230-239.

Osaki, S. Applied Stochastic System Modeling. New York: Springer-Verlag.

Ross, S. M. (1996). Stochastic Processes, Second Edition. Toronto: John Wiley & Sons.

Susilo, A., Hisyam, F., and Wasis. (2019). Earthquake Analysis in East Java, Indonesia between 1960 - 2017 Using Markov Chain Model. *International Journal of Geomate*, *17*(63), 149-156.

Tsai, M.-C., Li, Y.-L., and Chung, P.-H. (2014). A Study on Predicting the Turnover of Nursing Staff of Different Education Backgrounds: Using the Absorbing Markov Chain Method. *Journal of Quality*, *21*(6), 501-520.