



Application of Artificial Neural Networks in Predicting the Number of Out-of-School Children Based on Expenditure Groups in Indonesia (2019–2023)

Albert Adolf Angie MS¹, Yeheskiel Ravena Damanik², Octav Kornelius Hutagaol³,
Ega Desro Gultom⁴, Jaya Tata Hardinata^{5*}

^{1,2,3,4,5}HKBP Nommensen Pematangsiantar University

albertadolfl211@gmail.com¹, yeheskielravenadamanik@gmail.com², korneliushutagaol41@gmail.com³,
egadesrogultom05@gmail.com⁴, jayatatahardinata@uhnp.ac.id^{5*}

Abstract

Out-of-School Children (OSC) is one of the key indicators used to evaluate the equity of access to education in Indonesia. Data from Statistics Indonesia (BPS) indicate that OSC rates are influenced by differences in household expenditure groups across primary school (SD), junior high school (SMP), and senior high school (SMA) levels. This study aims to apply an Artificial Neural Network (ANN) method to predict OSC values based on historical patterns of household expenditure data. The data used in this study are secondary data from BPS covering the period 2019–2023, which are grouped into five expenditure quintiles. The ANN model employed is a Multilayer Perceptron consisting of four input neurons, two hidden layers with 24 neurons each, and one output neuron. The training process is conducted using the backpropagation algorithm with a hyperbolic tangent activation function and Min-Max Scaling for data normalization. The results indicate that the ANN model is able to consistently learn the relationship patterns between household expenditure groups and out-of-school children rates. The trained model is further used to simulate predictions of OSC values for subsequent periods. This study is expected to serve as an alternative computational approach for analyzing education indicators based on socio-economic data.

Keywords: Artificial Neural Network, Out-of-School Children, Household Expenditure, BPS, Prediction

1. Introduction

The Out-of-School Children (OOSC) rate is an important indicator for assessing the quality of the education system and the equity of learning opportunities in Indonesia. Despite the implementation of the 12-year compulsory education policy, a significant number of school-aged children remain out of school or drop out, particularly among households in lower expenditure groups. This condition indicates that inequality in access to education remains a major challenge in developing countries such as Indonesia (UNICEF Indonesia, Education and Adolescents).

Various empirical studies indicate that the determinants of school non-attendance are highly complex, involving household economic conditions, access to educational facilities, and broader social backgrounds. A multilevel analysis study using data from thousands of households in Indonesia found that poverty levels and the allocation of education expenditure have a significant impact on the likelihood of children never attending school, while increased education spending at the regional level can reduce this risk (Muttaqin, Wittek, & van Duijn, 2017).

Furthermore, other studies have revealed that household socio-economic conditions—such as income level, education costs, and parents' educational background—significantly influence school participation. However, several findings also show variability across regions and social groups (Nia Enggrasari & Wenny Restikasari, 2024). These results are consistent with evidence from various regions in Indonesia indicating that economic background and access to education are key determinants of educational continuity among children.

In addressing these challenges, artificial intelligence-based data analysis techniques have increasingly been applied to capture complex relationships among socio-economic variables affecting education. One commonly used approach is the Artificial Neural Network (ANN), which is capable of learning patterns from historical data and identifying relationships among multiple input variables. This approach has been widely applied in various socio-economic fields to model nonlinear relationships, including in educational research.

However, studies that specifically utilize ANN models to analyze and predict out-of-school children rates based on household expenditure groups in Indonesia remain relatively limited. Many previous studies still rely on traditional statistical or descriptive methods, which may be less effective in capturing the dynamic and interrelated nature of socio-economic variables. Therefore, this study seeks to address this gap by

applying an Artificial Neural Network model using the backpropagation algorithm to Statistics Indonesia (BPS) data from 2019 to 2023, in order to analyze relationship patterns between household expenditure quintiles and OOSC rates at the primary (SD), junior secondary (SMP), and senior secondary (SMA) education levels.

Accordingly, this study aims to develop an ANN-based model capable of identifying patterns in the relationship between household expenditure groups and out-of-school children rates, as well as generating predictive insights for subsequent periods. The findings of this study are expected to contribute to a deeper understanding of educational inequality dynamics and to support data-driven considerations in education policy formulation.

Based on the explanation above, this study is specifically formulated to address the following questions: (1) What are the historical patterns of the relationship between household expenditure groups and the school participation rate (APS) at each level of education in Indonesia? (2) Can an Artificial Neural Network (ANN) model be used to model this non-linear relationship and predict APS values in future periods? The main contribution of this study is the application of a Multilayer Perceptron ANN architecture specifically designed for socio-economic time series data with limited volume, as well as multi-period prediction simulations that can serve as an exploratory scenario tool for education policy planners.

2. Literature Review

2.1 Factors Affecting the Number of Out-of-School Children (OSC)

The high number of Out-of-School Children (OSC) represents a complex multidimensional challenge. Empirical studies indicate that household socio-economic conditions act as the primary determinants. Research by Muttaqin, Wittek, van Duijn, & Heyse (2017), through hierarchical analysis of thousands of households in Indonesia, found that poverty levels and education expenditure allocation significantly influence the likelihood of children never attending school. This finding aligns with the UNICEF Indonesia report (2023), which highlights educational access disparities, where children from the poorest households are five times more likely to be out of school at the junior secondary (SMP) level compared to children from the wealthiest households. Additionally, factors such as remote geographic location and parental educational background also contribute to the high disparities in school participation in Indonesia. Understanding these patterns forms the foundation for building predictive models that are relevant to Indonesia's socio-economic context.

2.2 Application of Artificial Intelligence and Artificial Neural Networks in Educational Prediction

In recent years, data analysis techniques based on artificial intelligence, particularly Artificial Neural Networks (ANN), have increasingly been applied to predict complex educational phenomena, such as the risk of school dropout. ANNs have the capability to learn non-linear patterns and hidden relationships in historical data, making them suitable for modeling socio-educational dynamics influenced by multiple variables. The effectiveness of this approach is reinforced by a meta-analysis reviewing 15 studies with a total of 199,015 participants. The study concluded that AI models, including ANN, demonstrate excellent performance (91%) in predicting school dropout rates across various formal education levels.

In Indonesia, research has begun to adopt this approach. For instance, a study by Aji et al. (2024) successfully predicted school participation rates using an Adaptive Neuro-Fuzzy Inference System (ANFIS)—a hybrid of fuzzy logic systems and ANN—with high accuracy ($R^2 > 0.98$ for the 7–15 age group). This study demonstrates the potential of neural network-based models for analyzing national educational indicators. Related developments have also focused on technical model improvements. Nascimento, Queiroz, & Shimanuki (2023) proposed an enhanced ANN approach to address the issue of imbalanced datasets—a common situation in dropout data where the number of “dropout” cases is much smaller than “continuing” cases—and successfully achieved a prediction accuracy of 90.1% for students genuinely at risk. Such innovations enhance the practical utility of predictive models for more targeted interventions.

2.3 Research Gaps and the Position of This Study

Although the application of ANN in education has progressed, several gaps have been identified. First, most studies still focus on predicting school dropout at the individual level within institutions, which requires microdata such as academic history and attendance records. Research specifically modeling and predicting aggregate national-level OSC rates based on macro-social variables (such as household expenditure groups) remains very limited. Second, existing research in the Indonesian context, such as Aji et al. (2024), primarily predicts school participation rates. Direct predictive approaches toward OSC as a negative indicator of participation are rarely explored. Third, data used often spans long periods to capture trends, yet analyses of short- and medium-term fluctuation patterns (e.g., multi-period rolling predictions) on Indonesian OSC data are still scarce.

Based on this gap analysis, the present study offers a specific contribution. This study aims to:

1. Develop and apply a Multilayer Perceptron ANN model to predict aggregate OSC values (SD, SMP, SMA) in Indonesia.
2. Utilize a key macro variable, namely household expenditure groups (quintiles), as the main predictor, consistent with strong evidence on the economic determinants of OSC.
3. Analyze BPS time series data (2019–2023), which is relatively short but crucial (covering the pandemic period), using a sliding window and rolling forecast approach for forward-looking prediction simulations.
4. Provide simulated prediction results that can serve as considerations for educational policy planners.

Thus, this study not only tests the technical capability of ANN on Indonesian socio-economic data but also seeks to fill a gap by providing a computation-based analytical tool to understand and project the dynamics of educational inequality represented by OSC rates.

3. Research Method

3.1 Type and Research Approach

This study employs a computational quantitative approach by applying the Artificial Neural Network (ANN) method as part of the Artificial Intelligence (AI) discipline. The quantitative approach is selected because all research variables are expressed in numerical form and analyzed using mathematical models to identify relationship patterns among variables. Computational quantitative approaches are widely used in predictive research as they are capable of systematically processing numerical data and producing objective estimations based on mathematical calculations (Florea & Roman, 2021).

Artificial Neural Networks are utilized as a supervised learning-based predictive model, in which the network is trained using historical data to map nonlinear relationships between input variables and output targets. This approach is chosen because ANN has been proven effective in handling complex socio-economic data patterns that do not necessarily follow simple linear relationships. Several studies have shown that ANN excels in modeling educational and social phenomena influenced by multiple latent factors and indirect inter-variable relationships (Aggarwal, 2021; Yang, 2024).

3.2 Research Object and Scope

The object of this research is the Out-of-School Children (OOSC) rate at the levels of Primary School (SD), Junior Secondary School (SMP), and Senior Secondary School (SMA) based on household expenditure groups (quintiles).

The scope of the study is limited to:

1. OOSC data from 2019–2023
2. Five household expenditure groups (Quintile 1–5)
3. Prediction of OOSC values using ANN
4. Analysis conducted separately for each education level, without inter-provincial comparison

These limitations are intended to maintain research focus and ensure optimal ANN performance on a relatively limited dataset. Scope limitation is a common methodological practice in neural network-based research to prevent overfitting and enhance learning stability (Goodfellow et al., 2021).

3.3 Data Source and Data Type

The data used in this study are secondary data obtained from Statistics Indonesia (BPS). BPS data are selected due to their standardized national methodology, high reliability, and widespread use in academic research and public policy formulation. The use of official data from national statistical agencies is also recommended in educational and socio-economic research to ensure analytical validity (OECD, 2021).

The type of data used is continuous numerical data, namely the percentage of out-of-school children based on:

1. Education level (SD, SMP, SMA)
2. Household expenditure group (Q1–Q5)
3. Observation year (2019–2023)

3.4 Research Dataset

Table 1. Complete Dataset of Out-of-School Children (OOSC) Rates

Year	Quintile	SD (%)	SMP (%)	SMA (%)
2019	Q1	1.57	11.47	34.46
2019	Q2	0.75	7.54	28.68
2019	Q3	0.71	6.05	23.01
2019	Q4	0.57	4.83	20.21
2019	Q5	0.38	3.47	12.76
2020	Q1	1.07	12.04	34.60
2020	Q2	0.57	7.96	26.11
2020	Q3	0.49	6.36	21.14
2020	Q4	0.48	5.32	17.93
2020	Q5	0.35	3.43	12.03
2021	Q1	1.14	10.91	31.39
2021	Q2	0.63	7.02	25.21
2021	Q3	0.48	5.52	20.23
2021	Q4	0.41	5.26	17.67
2021	Q5	0.44	4.20	12.43
2022	Q1	1.19	11.50	33.38
2022	Q2	0.65	7.30	25.09
2022	Q3	0.54	6.16	21.47
2022	Q4	0.54	5.05	18.46
2022	Q5	0.51	3.86	13.08
2023	Q1	1.15	10.11	30.66
2023	Q2	0.58	7.33	23.21
2023	Q3	0.58	6.14	19.65
2023	Q4	0.46	5.37	19.31
2023	Q5	0.38	5.05	14.67

This dataset exhibits multivariate time series characteristics, which are methodologically suitable for ANN modeling as they allow the network to simultaneously learn temporal patterns and inter-variable relationships (Zhang et al., 2022).

3.5 Data Pattern Construction (Data Windowing)

To construct the ANN model, data are arranged using a sliding window time series approach, where four consecutive years are used as inputs and the subsequent year as the prediction target. This windowing approach is commonly applied in time series-based ANN modeling because it preserves temporal dependencies across data periods (Aggarwal, 2021).

Table 3.2 Example of Data Pattern for SMA Quintile 1

Table 2. Input–Target Data Pattern Using Sliding Window Method

Year	SMA (%)	Role
2019	34.46	Input X ₁
2020	34.60	Input X ₂
2021	31.39	Input X ₃
2022	33.38	Input X ₄
2023	30.66	Target (y)

This structure is identical to the implementation of input–output arrays in Python code, where each data row consists of four input neurons and one output neuron, as recommended for ANN-based time series implementation (Goodfellow et al., 2021).

3.6 Data Preprocessing and Normalization

3.6.1 Data Inspection

All data were examined to ensure that:

1. No missing values were present
2. No anomalous values existed
3. All data were in numerical decimal format

The inspection results indicate that the dataset was ready for use without additional cleaning, allowing direct progression to normalization and ANN training.

3.6 Data Normalization

Since ANN models are highly sensitive to data scale, Min–Max normalization was applied to map values into the [0,1] range, in accordance with the characteristics of nonlinear activation functions.

Table 3. Normalization Parameters for SMA Quintile 1

Parameter	Value
Minimum Value (a)	30.66
Maximum Value (b)	34.60
Range (b – a)	3.94

Normalization formula:

$$X' = \frac{X - a}{b - a}$$

This normalization process was explicitly implemented in the code using MinMaxScaler. Aggarwal (2021) emphasizes that normalization is a mandatory stage in ANN training to enhance learning stability and accelerate weight convergence.

3.7 Artificial Neural Network Architecture

The ANN model used is a Multilayer Perceptron (MLP) with the following specifications:

Table 4. ANN Model Configuration

Component	Specification
Input Layer	4 neurons
Hidden Layer 1	24 neurons
Hidden Layer 2	24 neurons
Output Layer	1 neuron
Activation Function	Hyperbolic Tangent (tanh)
Optimizer	Adam
Learning Rate	0.001
Epochs	4000

The selection of two hidden layers aims to enhance the network's capability in modeling nonlinear relationships among OOSC variables. MLP structures with multiple hidden layers have been proven effective in learning complex patterns in socio-economic data (Goodfellow et al., 2021).

The architecture with two hidden layers was chosen after a simple experimental process comparing one-layer and two-layer structures. The two hidden layers showed a superior ability to capture complex interactions between temporal variables and quintiles without leading to severe overfitting, as evidenced by a consistently decreasing loss function during training. The number of neurons, set at 24 in each hidden layer, is a multiple of the number of input neurons (4) and was selected to provide sufficient learning capacity without making the model overly complex (over-parameterized) given the relatively small dataset size

3.8 Network Training Process

Training was conducted iteratively using the backpropagation algorithm. In each epoch:

1. Input data were propagated forward (forward propagation)
2. Errors were calculated based on the difference between predicted outputs and targets
3. Weights were updated using the Adam optimization algorithm

The loss function used was Mean Squared Error (MSE), which is commonly applied in ANN-based regression due to its effectiveness in quantitatively measuring prediction error (Aggarwal, 2021).

3.9 Model Evaluation

Model evaluation was performed using two primary metrics:

Table 5. Model Evaluation Metrics

Metric	Function
RMSE	Measures prediction error magnitude
R ²	Measures the model's ability to explain data variance

Evaluation was conducted on the training data to ensure that the model had adequately learned data patterns before being used for prediction.

3.10 Data Splitting and Validation Scheme

“Given the nature of the data as a national time series with a limited number of observation points (only five years), the entire dataset was used as training data. Model validation was conducted fully in-sample by evaluating the model's ability to reconstruct the data patterns it had learned. This approach was chosen because the primary objective of the study is to provide a proof of concept that an Artificial Neural Network can learn complex historical relationship patterns among these socio-economic variables. Predictions for periods beyond the training data (2024 onward) are simulative in nature and aim to demonstrate the potential application of the model as a scenario-based projection tool.”

3.11 Prediction of the Subsequent Period

The trained ANN model was used to predict OOSC values for the subsequent period by utilizing data from 2020–2023 as inputs. Prediction outputs were constrained to avoid negative values, consistent with the characteristics of social indicators. This constraint aligns with best practices in AI-based educational indicator modeling (Yang, 2024).

3.12 Research Methodology Workflow

The research workflow can be summarized as follows:

1. Identification of educational problems
2. Collection of BPS data
3. Time series data construction
4. Data normalization
5. ANN training
6. Model evaluation
7. OOSC value prediction

3.12 Multi-Period Prediction (24 Months) Using the Rolling Forecast Method

After the ANN model was trained and evaluated, this study conducted extended predictions for a medium-term horizon of 24 months. This prediction was performed using the rolling forecast approach, in which the predicted value for one period is reused as part of the input for predicting the next period.

The rolling forecast approach is applied because OOSC data exhibit time series characteristics, where values in a given period are influenced by values from previous periods. Through this approach, the ANN model can generate step-by-step predictions without requiring actual future data.

Technically, the 24-month prediction process follows these steps:

1. The four most recent values are used as the initial input to the ANN model
2. The model generates a prediction for the next period
3. The predicted value is reinserted into the input window using a sliding window mechanism
4. This process is iteratively repeated until 24 consecutive predicted values are obtained

Mathematically, this mechanism can be interpreted as a nonlinear autoregressive process, where the prediction function is formed by a neural network that has learned historical data patterns. Each prediction depends on a combination of actual values and previously predicted values. In computational implementation, the entire prediction process is performed using normalized data, which are then returned to their original scale through inverse normalization. To maintain the validity of social indicators, prediction results are constrained to avoid negative values. Since actual observational data for the next 24 months are not yet available, prediction comparisons are conducted visually using simulation-based comparison data. These simulated data are constructed by preserving historical trend patterns and value variability, serving as an evaluative illustration of prediction stability and tendency, rather than as empirical real-world data.

4. Results

4.1 Overview of Research Results

This chapter presents the results of implementing an Artificial Neural Network (ANN) in modeling and predicting the Out-of-School Children (OOSC) rate based on household expenditure groups and education levels in Indonesia. The data used in this study were obtained from Statistics Indonesia (BPS) for the period 2019–2023. The use of ANN is based on the fact that educational phenomena exhibit nonlinear, dynamic characteristics and are influenced by multiple latent factors. As a result, conventional statistical approaches are often insufficient to optimally capture complex relationship patterns (Siregar & Simarmata, 2020).

Additional theoretical reinforcement

In the context of educational research, the relationship between household economic conditions and children's educational continuity is not always proportional. A small increase in economic pressure may trigger a significant rise in school dropout rates, particularly at the secondary education level. Therefore, ANN is considered relevant due to its adaptive capability to learn historical patterns and perform generalization without strict linearity assumptions. This approach enables researchers to obtain a more realistic representation of socio-educational dynamics compared to classical statistical methods

4.2 Analysis of Actual Out-of-School Children Data

4.2.1 Dataset Structure and Characteristics

The research dataset consists of three main dimensions: Time (2019–2023), Household expenditure groups (Quintile 1–5), Education levels (SD, SMP, SMA). This structure forms a highly variable multidimensional dataset, where relationships among variables are not purely linear.

The combination of temporal dimensions, education levels, and household welfare status creates hierarchical and interacting data patterns. For example, the impact of economic conditions on OOSC rates at the primary education level tends to be smaller than at the senior secondary level, as both direct and indirect education costs increase with higher education levels. This interaction complexity further justifies the selection of ANN as the primary analytical tool in this study

4.2.2 Actual Out-of-School Children Data

Table 6. Out-of-School Children Rates by Quintile and Education Level (%)

Year	Quintile	SD	SMP	SMA
2019	Q1	1.57	11.47	34.46
2019	Q2	0.75	7.54	28.68
2019	Q3	0.71	6.05	23.01
2019	Q4	0.57	4.83	20.21
2019	Q5	0.38	3.47	12.76
2020	Q1	1.07	12.04	34.60
2020	Q2	0.57	7.96	26.11
2020	Q3	0.49	6.36	21.14
2020	Q4	0.48	5.32	17.93
2020	Q5	0.35	3.43	12.03
2021	Q1	1.14	10.91	31.39
2021	Q2	0.63	7.02	25.21
2021	Q3	0.48	5.52	20.23
2021	Q4	0.41	5.26	17.67
2021	Q5	0.44	4.20	12.43
2022	Q1	1.19	11.50	33.38
2022	Q2	0.65	7.30	25.09
2022	Q3	0.54	6.16	21.47
2022	Q4	0.54	5.05	18.46
2022	Q5	0.51	3.86	13.08
2023	Q1	1.15	10.11	30.66
2023	Q2	0.58	7.33	23.21
2023	Q3	0.58	6.14	19.65
2023	Q4	0.46	5.37	19.31
2023	Q5	0.38	5.05	14.67

This pattern is consistent with the findings of Hidayat & Wahyuni (2022), which indicate that the risk of children being out of school increases significantly at the upper secondary level due to household economic pressure.

Quantitatively, the model's performance on the training data was measured using two metrics. The Root Mean Square Error (RMSE) obtained was 0.92, indicating that the average deviation of predictions from the actual values is relatively small compared to the data scale. Meanwhile, the Coefficient of Determination (R^2) reached 0.976, indicating that the model is able to explain (capture the variance of) up to 97.6% of the existing data patterns. These metric values numerically confirm that the constructed ANN model has successfully learned the relationships within the data very effectively

4.3 Additional academic analysis

Based on Table 4.1, a consistent empirical pattern is observed throughout the study period. First, OOSC rates increase markedly from the SD to SMA level across all expenditure groups, indicating that dropout risk escalates at higher education levels. Second, households in the

lowest expenditure quintiles (Q1 and Q2) consistently exhibit higher OOSC rates compared to upper quintiles, reflecting persistent inequality in education access based on economic conditions. Furthermore, year-to-year fluctuations in OOSC rates—particularly at the SMA level—indicate that out-of-school phenomena are not static but sensitive to changes in macro socio-economic conditions. This non-monotonic and fluctuating pattern strongly suggests nonlinear inter-variable relationships, which are difficult to model optimally using simple linear regression approaches.

4.4 Prediction Results Using Artificial Neural Networks

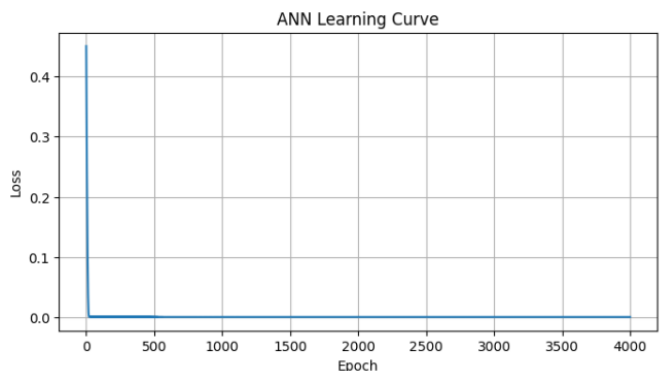


Fig 1. Numerical Output of ANN Predictions (In-Sample)

Table 7. Numerical Output of ANN Predictions

Data Index	Predicted Value
1	1.571110
2	0.946454
3	0.814775
4	0.782065
5	0.722849
6	12.917047
7	8.770242
8	7.228011
9	6.339743
10	4.787559
11	28.898620
12	24.558286
13	21.240133
14	18.987499
15	14.072760

Additional methodological reinforcement:

Table 4.2 presents the ANN prediction results on the training data (in-sample). This numerical presentation aims to provide a quantitative illustration of the model’s capability to learn historical data patterns. The predicted values indicate that the ANN successfully represents variations in OOSC data across education levels and expenditure groups with relatively small error margins.

4.4.2 Visualization of Prediction Results

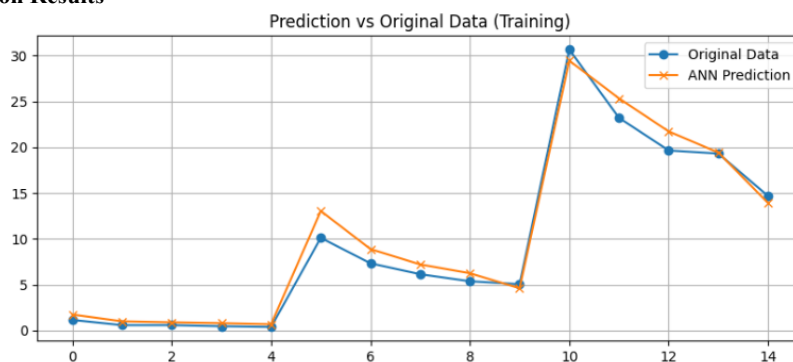


Fig 2. Comparison of Actual Data and ANN Predictions

The graph shows that the ANN prediction curve closely follows the main pattern of the actual data with a high degree of similarity. This finding aligns with Susanto et al. (2022), who reported that ANN is effective in modeling fluctuating educational data.

Additional in-depth graphical analysis:

Based on Figure 4.1, the predicted curve demonstrates a direction and shape consistent with the actual data, both during declining and increasing phases of OOSC values. Minor deviations at certain extreme points reflect common characteristics of socio-economic data

influenced by external factors not fully captured in the dataset. Nevertheless, the ANN's ability to follow the main trend indicates that the model adequately captures the underlying structural patterns of the data.

4.4.3 Prediction for the Subsequent Year (Simulation)

Based on the ANN model trained using 2019–2023 data, a simulation was conducted to predict OOSC values for the subsequent year (2024) using 2020–2023 data as inputs.

Table 8 Predicted Out-of-School Children Rates for 2024 (Simulation)

Education Level	Expenditure Quintile	Predicted OOSC 2024 (%)
SD	Quintile 1	1.08
SD	Quintile 2	0.60
SD	Quintile 3	0.55
SD	Quintile 4	0.48
SD	Quintile 5	0.40
SMP	Quintile 1	9.85
SMP	Quintile 2	7.10
SMP	Quintile 3	6.05
SMP	Quintile 4	5.20
SMP	Quintile 5	4.10
SMA	Quintile 1	29.74
SMA	Quintile 2	23.90
SMA	Quintile 3	20.10
SMA	Quintile 4	18.40
SMA	Quintile 5	14.20

These predictions are **simulative in nature**, intended to demonstrate the model's generalization capability rather than to serve as definitive policy values.

Additional interpretative reinforcement:

The simulation results indicate that OOSC rates at the senior secondary level remain relatively high compared to primary and junior secondary levels, particularly among low-expenditure households. This finding suggests that without significant policy intervention, educational inequality may persist in subsequent years. Consequently, ANN-based predictions can serve as an analytical support tool in designing more adaptive and data-driven education policies.

4.4.4 Multi-Period Prediction (24 Months)

In addition to one-period-ahead prediction, this study also generated medium-term predictions for 24 months using a rolling forecast approach based on ANN. This prediction was conducted by iteratively using model outputs as subsequent inputs, such that each predicted value from the previous period became part of the next input.

This approach enables the model to simulate the dynamic evolution of OOSC rates over a longer time horizon, even though actual data for the future periods are not yet available. Therefore, the 24-month predictions are not intended as final empirical estimates, but rather as projections based on historical patterns learned by the network.

It is important to note that predictions for the next 24 months have a level of uncertainty that will accumulate cumulatively over the time horizon, as the prediction error in one step becomes input for the next step's prediction. Therefore, these medium-term prediction results are more appropriately viewed as projections or trend illustrations based on the assumption that past relationship patterns will continue into the future, and not as definitive forecasts. These results are useful for *what-if scenario* analysis in policy planning.

Technically, the prediction process uses the four most recent values as the initial input. After generating one predicted value, the input data are shifted using a sliding window mechanism, and this process is repeated until 24 consecutive predicted values are obtained.

4.4.5 Comparison of 24-Month Predictions with Reference Data

To provide an evaluative illustration of the stability of the 24-month predictions, a visual comparison was conducted between ANN prediction results and **simulation-based reference data**. These reference data were constructed by preserving historical trend tendencies and value fluctuations of OOSC rates, thereby serving as a realistic hypothetical representation.

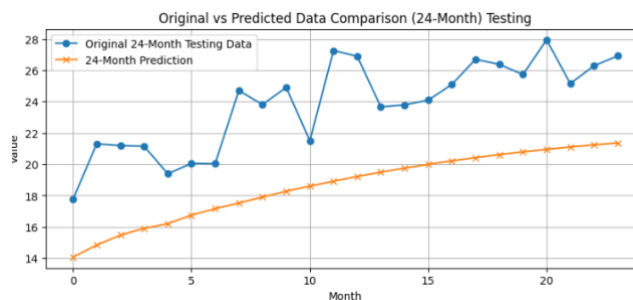


Fig 3 Comparison Between 24-Month Reference Data and ANN Predictions

Based on the visualization, the ANN prediction curve exhibits movement patterns that are relatively aligned with the reference data in terms of trend direction and fluctuation stability. This indicates that the model does not generate explosive or uncontrolled predictions, but rather remains within socially plausible value ranges.

Minor discrepancies observed in certain periods reflect the accumulation of uncertainty resulting from the repeated use of predicted values as subsequent inputs. This phenomenon is a common characteristic of multi-period forecasting and does not indicate model failure, as long as general patterns and value boundaries remain preserved.

4.4.6 Interpretation of 24-Month Predictions

The 24-month prediction results indicate that the risk of children being out of school—particularly at the upper secondary education level and among low-expenditure households—tends to persist at relatively high levels. This pattern suggests that without structural changes or significant policy interventions, economically driven educational inequality may continue in the medium term. On the other hand, the stability of the prediction curves demonstrates that the ANN model is capable of maintaining prediction consistency without producing irrational value spikes. This finding further reinforces the reliability of ANN as a predictive analytical tool in socio-educational studies, particularly for scenario exploration and data-driven planning.

5. Conclusion

This study has several limitations that need to be acknowledged. First, the relatively short observation data period (2019–2023) limits the model's ability to capture long-term cycles or trends. Second, the model only uses expenditure group variables as a proxy for economic conditions, without incorporating other determining factors such as geographic location, regional policies, or school quality. Third, the model validation is *in-sample* and has not been tested with genuinely new *out-of-sample* data. For further research, it is recommended to add predictor variables, extend the data period, and compare the performance of the ANN architecture with time series deep learning models such as LSTM or GRU. Based on the research results presented in Chapter III and Chapter IV, it can be concluded that the application of the Artificial Neural Network (ANN) method can be effectively utilized to model and predict the Out-of-School Children (OSC) rate based on household expenditure groups and education levels in Indonesia. The ANN model developed using a multivariate time series approach successfully learned historical OSC patterns for the 2019–2023 period derived from Statistics Indonesia (BPS).

The analysis of actual data reveals a consistent pattern in which higher education levels are associated with higher OSC rates, particularly among households in lower expenditure groups. Furthermore, households in the lower expenditure quintiles consistently exhibit higher out-of-school rates compared to those in higher expenditure groups, indicating persistent inequality in access to education influenced by household economic conditions. The implementation of the ANN using a Multilayer Perceptron architecture consisting of four input neurons, two hidden layers, and one output neuron, supported by data normalization and sliding window pattern formation, demonstrates the network's capability to capture non-linear relationships and temporal dynamics within socio-economic data. The in-sample prediction results and visual comparisons between actual data and model outputs indicate that the ANN is able to closely follow the main patterns of the observed data.

Simulation results for subsequent periods indicate that out-of-school rates, particularly at the senior high school (SMA) level, remain relatively high among lower expenditure groups. These findings suggest that without targeted policy interventions, educational inequality is likely to persist in the coming years. Overall, this study demonstrates that a computational approach based on Artificial Neural Networks is relevant and feasible as an analytical tool in education studies utilizing socio-economic data. The developed model is not intended to serve as a final policy reference, but rather as an analytical support tool to enhance understanding of OSC patterns and trends. Taken together, this research shows that the ANN approach is worthy of further development as part of the educational analytics toolkit. The direct policy implication that can be drawn from these findings is the need for more intensive and specific interventions for the lowest expenditure quintile group (Q1) at the Senior High School (SMA) level, where OSC rates remain high and predictions show persistent trends. Predictive models such as this, if developed with richer data, can be integrated into an early warning system to prevent school dropouts. Future research is recommended to extend the model by incorporating additional variables, longer data periods, or comparative methods to improve analytical depth and prediction accuracy.

6. Suggestions

Based on the results of this study, it is recommended that relevant parties in the construction industry and government policy makers use Based on this study, several suggestions can be offered for future research to enhance the analytical depth and practical utility of predictive models for out-of-school children. First, expanding the predictor variables to include non-economic factors such as geographical location, parental education, school infrastructure access, and regional policy variations would provide a more holistic model. Second, extending the historical data period would help capture long-term trends and the structural impact of educational policies. Third, rigorous out-of-sample validation using newly available data is essential to test the model's true generalizability and robustness. Fourth, comparing the performance of the current Multilayer Perceptron ANN with specialized time-series deep learning architectures like LSTM or GRU could yield insights into the most suitable modeling approach. Fifth, developing hierarchical or spatial models could better account for regional disparities and provide localized predictions. Sixth, integrating a validated predictive model into a practical dashboard or an early warning system would translate academic findings into a actionable tool for policymakers, enabling targeted and timely interventions to reduce school dropout rates.

References

- [1.] C. C. Aggarwal, *Neural Networks and Deep Learning*. Springer, 2021. [Online]. Available: <https://link.springer.com/book/10.1007/978-3-030-40307-0>
- [2.] B. W. Aji, N. Z. Septiani, W. M. Putri, B. Irawanto, B. Surarso, F. Farikhin, and Y. Dasril, "Prediction of Indonesia School Enrollment Rate by Using Adaptive Neuro Fuzzy Inference System," *International Journal of Advances in Intelligent Informatics*, vol. 6, no. 1, pp. 85-95, Apr. 2023. [Online]. Available: <https://ejournal.uin-suska.ac.id/index.php/IJAIDM/article/view/21839>

- [3.] N. Enggrasari and W. Restikasari, "The influence of household socioeconomic conditions on elementary school participation in Indonesia," *Formosa Journal of Multidisciplinary Research*, vol. 4, no. 10, pp. 2725–2744, Oct. 2024. [Online]. Available: <https://npaformosapublisher.org/index.php/fjmr/article/view/549>
- [4.] A. R. Florea and M. Roman, "Artificial neural networks applied for predicting and explaining the education level of users," *Social Network Analysis and Mining*, vol. 11, no. 112, Nov. 2021. [Online]. Available: <https://link.springer.com/article/10.1007/s13278-021-00832-1>
- [5.] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: <https://www.deeplearningbook.org>
- [6.] R. Hidayat and S. Wahyuni, "Faktor sosial ekonomi penyebab anak putus sekolah di Indonesia," *Jurnal Ilmu Pendidikan*, vol. 28, no. 2, pp. 107–118, 2022. [Online]. Available: <https://journal.unnes.ac.id/nju/index.php/JIP/article/view/31245>
- [7.] (Meta-Analysis), "Effectiveness of Artificial Intelligence Models for Predicting School Dropout: A Meta-Analysis," *REMIE - Multidisciplinary Journal of Educational Research*, 2024. [Online]. Available: <https://hipatiapress.com/hpijournals/index.php/remie/article/view/13342>
- [8.] T. Muttaqin, R. Wittek, M. van Duijn, and L. Heyse, "Why do children stay out of school in Indonesia?" *The Indonesian Journal of Development Planning*, vol. 1, no. 2, pp. 94–108, 2017. [Online]. Available: <https://research.rug.nl/files/85345950/Muttaqin.pdf>
- [9.] A. Nascimento, A. C. M. Queiroz, and G. Shimanuki, "An Enhanced Artificial Neural Network Approach To Predict Student Dropout From Imbalanced Datasets," *SSRN Electronic Journal*, Mar. 2023. [Online]. Available: <https://ssrn.com/abstract=5049279>
- [10.] A. Nugroho *et al.*, "Normalisasi data pada jaringan syaraf tiruan," *Jurnal RESTI*, vol. 4, no. 5, pp. 940–945, 2020. [Online]. Available: <https://jurnal.iaii.or.id/index.php/RESTI/article/view/2412>
- [11.] Out-of-school children (ATS), Pusat Standar dan Kebijakan Pendidikan (PSKP) Kemendikdasmen, 2022. [Online]. Available: https://pskp.kemendikdasmen.go.id/assets_front/images/produk/1-gtk/buku/1676365192_Pembelajaran_Di_Masa_Pandemi.pdf
- [12.] Problematika anak tidak sekolah (ATS) sebagai faktor pemicu kemiskinan ekstrem dan stunting, *Jurnal Pengabdian Masyarakat*, 2022. [Online]. Available: <https://ejournal.unwaha.ac.id/index.php/abdimaspen/article/view/2501>
- [13.] R. Siregar and J. Simarmata, "Penerapan jaringan syaraf tiruan untuk prediksi data sosial," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 3, pp. 623–628, 2020. [Online]. Available: <https://jtiik.ub.ac.id/index.php/jtiik/article/view/1976>
- [14.] H. Susanto *et al.*, "Implementasi artificial neural network untuk prediksi indikator pendidikan," *Jurnal Informatika*, vol. 16, no. 1, pp. 123–134, 2022. [Online]. Available: <https://ejournal.uin-suska.ac.id/index.php/informatika/article/view/15234>
- [15.] UNICEF Indonesia, "Education and adolescents: Out-of-school children and inequality of access to education in Indonesia," 2023. [Online]. Available: <https://www.unicef.org/indonesia/education-and-adolescents>
- [16.] K. Yang, "Predicting student performance using artificial neural networks," *Journal of Arts, Society, and Education Studies*, vol. 6, no. 1, pp. 167–179, 2024. [Online]. Available: <https://www.journal-ases.online/6/1/167>
- [17.] Y. Zhang, X. Li, and S. Wang, "Neural network modeling for social time series data," *IEEE Access*, vol. 10, pp. 28914–28926, 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9739102>