Multidisciplinary Output Research For Actual and International Issue (MORFAI Journal) ISSN (e): 2808-6635
Volumes 6 No. 1 (2026)

# OPTIMIZATION OF INDONESIA'S DEMOGRAPHIC BONUS THROUGH YOUNG LABOR FORCE PREDICTION COMPARATIVE ANALYSIS OF BSTS, LSTM AND SARIMA MODELS IN LIMITED DATA 2015–2024

**Harun Al Jafar[1], Fauziah[2]**
Master of Information Technology Study Program Faculty of Communication and Informatics Technology National University
E-mail: aljafarharun663@gmail.com[1], fauziah@civitas.unas.ac.id[2]

| | | | |
|---|---|---|---|
| Received | : 01 October 2025 | Published | : 23 December 2025 |
| Revised | : 10 October 2025 | DOI | : https://doi.org/10.54443/morfai.v6i1.4737 |
| Accepted | : 15 November 2025 | Publish Link | : https://radjapublika.com/index.php/MORFAI/article/view/4737 |

**Abstract**

The 15–24 age group is a key driver of Indonesia’s demographic bonus, yet the youth labor market continues to fluctuate due to limited skill readiness, economic conditions, and the absence of a strong labor force prediction system. These issues are compounded by sparse annual data from 2015–2024, which contains missing values and requires appropriate modeling to ensure accurate forecasting. This study aims to develop a prediction model for the number of young workers using three approaches—BSTS, LSTM, and SARIMA—while comparing their accuracy based on MAPE and RMSE. The study also tests whether performance differences among the models are statistically significant and identifies the most optimal model to support youth employment policy planning in Indonesia. The research uses time series data from BPS covering 2015–2024. Preprocessing includes imputing missing values through linear interpolation, normalizing the data, and dividing it into training and testing sets. Each model—BSTS, LSTM, and SARIMA—is then applied and evaluated using MAPE, RMSE, and statistical tests such as the Wilcoxon or paired t-test to assess the significance of performance differences. Results show that LSTM and SARIMA yield the highest accuracy, each achieving a MAPE of 3.44%, while BSTS performs less effectively on limited annual data. Statistical tests confirm that BSTS differs significantly from LSTM and SARIMA, whereas LSTM and SARIMA do not significantly differ from each other. The accuracy decline in 2020–2021 underscores the sensitivity of the youth labor market to external shocks like the pandemic. Overall, SARIMA and LSTM emerge as the most suitable models for forecasting youth labor numbers in Indonesia. These findings can guide the development of adaptive, data-driven employment policies to maximize the demographic bonus potential.

***Keywords: Young Workforce, BSTS, LSTM, SARIMA, Labor Force Prediction, Demographic Bonus.***

## INTRODUCTION

Changes in demographic structure are an important phenomenon faced by many countries, especially developing ones. One crucial phase in demographic dynamics is the demographic bonus, a period when the proportion of the productive-age population (15–64 years) reaches its highest point compared to the non-productive age groups. Theoretically, this condition presents significant opportunities for economic growth through increased productivity, national savings, and human capital investment (Jiajun et al., 2025). The experiences of countries such as South Korea and China show that successfully leveraging the demographic bonus is highly dependent on the readiness of education systems, labor markets, and integrated socioeconomic policies (Haider & Mahmood, 2023). Indonesia is currently in the intensification phase of its demographic bonus, as empirically reflected in the composition of its productive-age population up to 2024 based on official publications from Statistics Indonesia (Badan Pusat Statistik). This period marks the early stage of the national demographic bonus, projected to continue into the 2030s. Therefore, data-based analysis using the 2015–2024 period holds strategic relevance as an initial foundation for formulating medium-term policies. According to data from (Badan Pusat Statistik, 2025), the working-age population has reached 216.79 million people, with a Labor Force Participation Rate (LFPR) of 70.60% and an Open Unemployment Rate (OUR) of 4.76%. However, behind these figures lie structural challenges: approximately 52.72% of the labor force has completed only primary or lower secondary education, and more than one-third are employed in non-full-time work. This reflects the limited quality and capabilities of the workforce in facing the challenges of the industrial revolution and digital transformation.

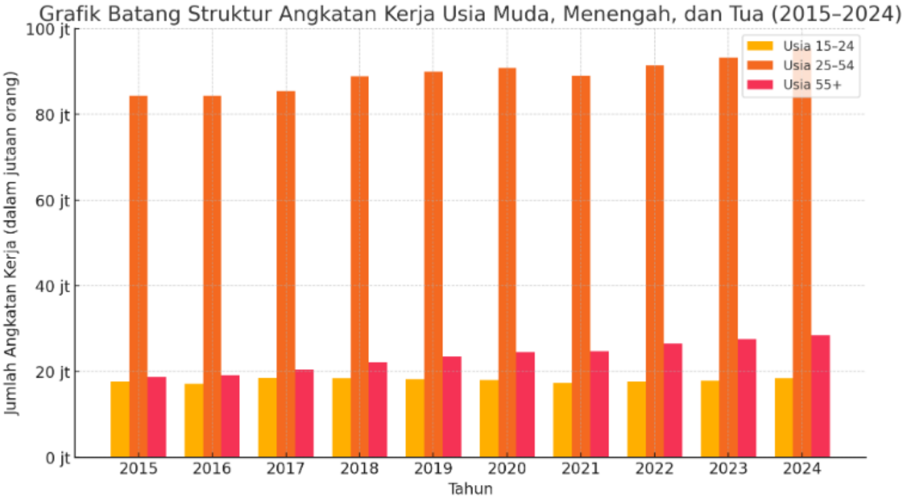


**Figure 1.** Youth Labor Force Chart

The youth group (15–24 years) occupies a position that is both strategic and vulnerable. They are the frontline in leveraging the demographic bonus, yet also face risks such as unemployment, informal work, and skill disparities. Previous studies show that young people in developing countries often experience skills mismatch, limited work experience, and restricted access to vocational training (Sharma, 2022; Tikhonov et al., 2024). This raises a critical question: to what extent has labor market planning accurately and data-drivenly responded to the needs and potential of the youth population? Reliable labor market planning requires a prediction system capable of capturing the dynamic complexity of labor market data. Conventional models such as linear regression and ARIMA are still widely used, but they encounter limitations when dealing with multivariate, nonlinear, or structurally shifting data (Orozco-Castañeda et al., 2024). Meanwhile, modern approaches such as Bayesian Structural Time Series (BSTS) and Long Short-Term Memory (LSTM) have gained attention in economic forecasting due to their ability to capture trends, seasonality, and long-term relationships in time series data (De Mendonça et al., 2024; Muizzadin et al., 2025; Wang et al., 2025).

BSTS is known for its adaptability to structural changes and is highly suitable for small datasets commonly found in annual socio-economic data. In contrast, LSTM, as a deep learning model, can capture nonlinear and complex patterns but relies more heavily on complete and sufficiently large datasets and is prone to overfitting (Al-Selwi et al., 2024). Numerous studies have demonstrated LSTM's strength in predicting stock prices, air pollution, and COVID-19 trends (Bagastio et al., 2023; Halim et al., 2024; Sembiring et al., 2024), yet few have applied it in the context of youth labor markets in Indonesia. Furthermore, studies on Indonesia's youth labor market remain largely descriptive, focusing on issues such as labor mobility, sectoral preferences, or older workers in agriculture (Ngadi et al., 2023), without incorporating predictive, model-based quantitative approaches. To date, no studies have explicitly compared the performance of BSTS and LSTM in forecasting the number of young workers in Indonesia, especially under conditions of limited data, complex exogenous variables, and macroeconomic uncertainty.

Thus, this study offers two key scientific contributions: (1) exploring the performance of two prediction approaches of different natures—BSTS as a structural probabilistic model and LSTM as a deep learning model—in the context of Indonesia's labor market; and (2) providing an empirical foundation for developing evidence-based labor policies aimed at optimizing the demographic bonus. Based on this rationale, this research focuses on analyzing and comparing the accuracy of Bayesian Structural Time Series (BSTS) and Long Short-Term Memory (LSTM) models in predicting the labor force population aged 15–24 in Indonesia for the 2015–2024 period. This focus is motivated by the limited availability of short, aggregated annual time series data, which requires methods capable of performing optimally under such conditions. BSTS is selected for its ability to handle time series data with exogenous variables, accommodate missing values, and provide clear structural interpretation. Meanwhile, LSTM is used for its advantage in modeling nonlinear patterns and capturing long-term dependencies in time series. The exogenous variables used include the Labor Force Participation Rate (LFPR), Open Unemployment Rate (OUR), Average Years of Schooling (AYS), GDP per capita, and the number of young people. In addition to comparing model performance, this study is expected to provide evidence-based policy recommendations to support the optimization of Indonesia's demographic bonus.

# METHOD

## Research Approach and Type

This study is a comparative quantitative research aimed at comparing two time-series modeling approaches. An explanatory approach is used to examine the extent to which structural and performance differences exist between BSTS and LSTM in predicting the number of young workers. The study also qualifies as applied research because it is designed to provide practical contributions to data-driven employment policy.

## Population and Sample / Research Subjects

The population of this study consists of Indonesia's labor and socio-economic data for the 2015–2024 period. The sample includes:

1. Target Variable: Number of Workers Aged 15–24 Years
2. Exogenous Variables:
   a. Labor Force Participation Rate (LFPR) for ages 15–24
   b. Open Unemployment Rate (OUR) for ages 15–24
   c. Average Years of Schooling (AYS)
   d. Human Development Index (HDI)
   e. Gross Domestic Product (GDP) per capita (current prices)
   f. Total Population Aged 15–24

## Data Collection Techniques

Data were collected from the official Statistics Indonesia (BPS) website through documentation methods. The process included:

1. Downloading annual BPS publications in Excel format.
2. Extracting and converting the data into CSV format.
3. Normalizing variable names, detecting missing values, and validating data across years.
4. Recording metadata such as sources and publication years.

# RESULTS AND DISCUSSION

## Modeling Results

### BSTS Model

1. BSTS Model Development
   The Bayesian Structural Time Series (BSTS) model was constructed using a local linear trend component and dynamic regression with exogenous variables. This structure allows the model to capture long-term structural movements and the dynamic influence of socio-economic factors on the number of young workers aged 15–24. The target variable is the number of young workers, while the exogenous variables include LFPR, unemployment rate, average years of schooling, HDI, GDP per capita, and youth population. The model was trained using 80% of the data, with 20% reserved for testing. Given the limited dataset, BSTS was selected for its ability to work with small samples, adaptively capture trends, and provide stable probabilistic estimates.
2. BSTS Prediction Results
   The comparison between actual and predicted values shows that the youth labor force data fluctuates sharply from year to year, reflecting its sensitivity to socio-economic conditions, including the pandemic and economic recovery. During the testing period (2023–2024), the BSTS model follows the general upward trend but shows substantial numerical discrepancies. For 2023, the model predicted values around $9.0\times10^7$, far below the actual figure above $2.0\times10^8$. In 2024, the model predicted a slight decline, while actual values increased. These gaps indicate that although BSTS captures trend direction, its numerical accuracy is limited due to high volatility, small sample size, and non-linear exogenous effects. Nonetheless, BSTS remains structurally adaptive, particularly in the post-pandemic period, making it useful for macro-policy analysis focused on directional tendencies rather than precise point estimates.
3. Model Accuracy Evaluation
   The BSTS model produced an RMSE of 136,176,356 and a MAPE of 61.5%, indicating large prediction errors. A MAPE above 50% is considered low accuracy, confirming that BSTS performs poorly in predicting youth labor force values for 2023–2024. The high error rate stems from the very limited annual observations (only 10 data points), extreme volatility, large variable scale differences, and the youth labor market's sensitivity to economic shocks. Despite its limitations, these findings highlight the constraints of BSTS for highly fluctuating

labor data and justify comparing it with LSTM and SARIMA to identify the most adaptive model for Indonesia's youth labor market.

4. Interpretation of Exogenous Variable Influence in BSTS
The Posterior Inclusion Probabilities (PIP) in the BSTS model show that all exogenous variables—LFPR_15_24, OUR_15_24, HDI, AYS, youth population, and GDP per capita—have extremely low inclusion probabilities close to zero, including the intercept. This indicates that the BSTS model finds no strong statistical evidence that these variables significantly explain variations in the youth labor force. Instead, the model relies heavily on the latent local trend, suggesting that long-term structural factors and external shocks drive changes in youth labor force numbers more than the included macro variables. The low PIP values also contribute to BSTS's low predictive accuracy (MAPE 61.5%), as the model effectively performs trend extrapolation rather than capturing causal relationships. The extremely limited number of observations (10 annual data points) and high year-to-year volatility further destabilize Bayesian posterior estimates, causing the spike-and-slab prior to suppress exogenous variable activation. Consequently, BSTS's multivariate predictive capacity becomes limited, strengthening the rationale for using alternative nonlinear models such as LSTM.
5. Interpretation of the Low BSTS Performance
The poor performance of BSTS is not only due to limited data but also specific model and data characteristics. First, the very low posterior inclusion probabilities suggest the model cannot detect meaningful effects from exogenous predictors, potentially due to multicollinearity—such as the extremely high correlation (0.98) between HDI and AYS. Second, the BSTS structure used—local linear trend plus dynamic regression—may be too simplistic to capture major shocks like COVID-19. Prior studies (Martin et al., 2022) recommend adding components such as spike-and-slab or student-t errors to handle outliers. Third, the use of non-informative priors in small datasets produces wide, imprecise posterior distributions, indicating that informative priors based on domain knowledge may be necessary. Overall, the BSTS underperformance highlights the need for improved model specification and the integration of prior information when dealing with highly volatile and limited labor force data.

**LSTM Model**

1. LSTM Model Architecture
The LSTM model used in this study consists of a single LSTM layer with 32 memory units, followed by a Dropout layer to reduce overfitting, as well as two fully connected (Dense) layers with 16 neurons and 1 output neuron. The total number of trainable parameters is 5,665. This structure was designed to balance model complexity with the limited annual dataset, ensuring that the model can capture the key temporal patterns without overfitting.
2. Loss Development During Training
The training loss decreased very rapidly and approached zero, while the validation loss experienced fluctuations at the beginning of the training process but gradually decreased and stabilized at a relatively low level. This indicates that the LSTM model successfully reached convergence and was able to learn the data patterns stably, with no strong indication of training divergence. The difference between training and validation loss remains acceptable, showing that the model has good generalization ability.
3. LSTM Prediction Results on Test Data
Based on the comparison between the actual data and the LSTM predictions, the predicted values for the 2023–2024 testing period were relatively close to the actual values. In 2023, the LSTM predictions followed the upward trend of the youth labor force (ages 15–24), and a similar increasing trend appeared in 2024. Although numerical differences exist between the predicted and actual values, the overall movement pattern learned by the LSTM is consistent with the real labor market trend. This demonstrates that the model is able to capture short-term temporal dynamics more adaptively than linear structural approaches.
4. LSTM Model Accuracy Evaluation
The model's performance on the test data resulted in the following metrics:
   a. RMSE: 804,989.19
   b. MAPE: 3.44%

   According to forecasting evaluation criteria, a MAPE value below 10% falls under the category of "very good accuracy." Thus, the LSTM model is highly accurate in modeling the number of young workers (ages 15–24). Compared to the BSTS model in Subchapter 4.3, which produced a MAPE of 61.5% (low accuracy), the LSTM shows a significantly superior performance. This confirms that deep learning approaches are more effective in capturing nonlinear patterns and sharp fluctuations in youth labor data.

5. Analysis of LSTM Performance Against Data Characteristics
   The primary strength of the LSTM in this study lies in its ability to learn both short-term and long-term temporal dependencies without relying on linearity assumptions. This makes the LSTM more adaptive to extreme changes in the data, including pandemic impacts and economic recovery phases during 2021–2024. Unlike BSTS, which heavily depends on latent trend components and linear regression with exogenous variables, the LSTM operates in a fully data-driven manner, learning patterns and relationships directly from historical data. The main drawback of this approach is low interpretability, but this is compensated by the model's high predictive accuracy.

**SARIMA Model**

**Summary of the ARIMA Model**

1. ARIMA Model Structure
   The ARIMA model in this study was developed through stationarity testing and autocorrelation analysis of the youth labor force time series (ages 15–24). The Augmented Dickey-Fuller (ADF) test showed that the data were non-stationary at level but became stationary after first differencing, leading to an integration order of d = 1. A limited grid search for parameters *p* and *q* (ranging from 0 to 2) identified ARIMA(0,1,0) as the best model based on the lowest AIC. This model represents a *random walk with drift*, suitable for annual data that fluctuate but do not exhibit strong cyclic or long-term dependencies.
2. Stationarity Testing and Model Identification
   The ADF test produced a p-value of 0.9678 at level (non-stationary) and 0.0087 after first differencing (stationary). ACF and PACF plots showed rapidly declining autocorrelations without significant peaks, indicating that more complex AR or MA structures were unnecessary. Given the small number of annual observations, a simple and stable model structure was prioritized.
3. Estimation of ARIMA(0,1,0)
   The selected model contains only one main parameter—the innovation variance—resulting in stable estimation and avoiding overfitting. Residual diagnostics showed no significant autocorrelation (Ljung–Box test) and approximate normality with homogeneous variance, indicating that the model meets all key diagnostic assumptions.
4. ARIMA Prediction Results
   Predictions for 2023–2024 were close to actual values, although the ARIMA output appeared smoother than real data fluctuations. The model successfully followed the upward trend in youth labor force participation, albeit with slightly lower predicted values. Overall, its movements aligned with short-term dynamics typical of ARIMA(0,1,0)'s focus on average annual changes.
5. Accuracy Evaluation
   Model performance metrics:
   a. RMSE: 825,495.10
   MAPE: 3.44%. A MAPE below 10% indicates *very good forecasting accuracy*, and the RMSE suggests acceptable deviation for large-scale labor force data.
6. ARIMA Performance Analysis
   ARIMA is effective for modeling linear annual changes and short-term trends, which match the characteristics of youth labor force data. However, it cannot capture nonlinear patterns or extreme fluctuations as deeply as more flexible models. Still, given the limited annual dataset, ARIMA provides stable and reliable predictions.

**Model Evaluation**

After the three Bayesian Structural Time Series (BSTS), Long Short-Term Memory (LSTM), and Seasonal ARIMA (SARIMA) models were built and evaluated individually, the next step was to conduct a comprehensive performance comparison. This comparison aimed to determine which model was most accurate and stable in predicting the number of 15–24-year-olds in the labor force in Indonesia based on annual data from 2015–2024. The evaluation was conducted using two main metrics: Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE), and was supplemented with statistical tests to test the significance of performance differences between the models.

### Model Performance Comparison Table

The following is a summary of the MAPE and RMSE values of the three models on the test data (2023–2024):

**Table 1.** Comparison of Model Performance

| Model | RMSE | MAPE (%) | Accuracy Category (based on MAPE) |
|---|---|---|---|
| BSTS | 136,176,356 | 61.50% | Low accuracy (MAPE > 50%) |
| LSTM | 804,989.19 | 3.44% | Very high accuracy (MAPE < 10%) |
| SARIMA | 825,495.10 | 3.44% | Very high accuracy (MAPE < 10%) |

From the table above, we can conclude that:

1. The BSTS model demonstrated very poor performance with a MAPE of 61.50%, indicating the model's inability to capture the dynamics of fluctuating and limited data.
2. The LSTM and SARIMA models demonstrated excellent performance with identical MAPEs of 3.44%, categorized as high accuracy based on Lewis's (1982) standards.

### Visualization of Prediction vs. Actual Comparison

The following figure compares the prediction results of the three models with the actual data during the testing period:

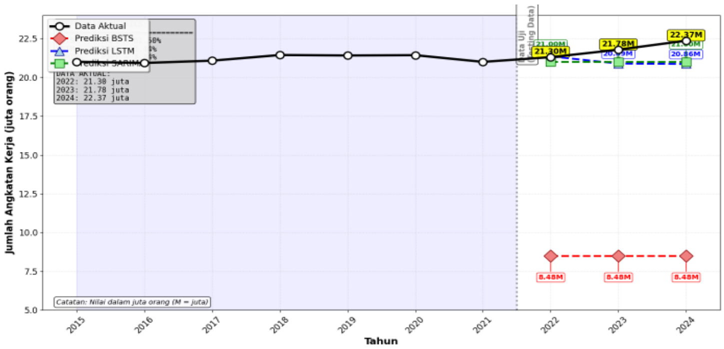


**Figure 2.** Visualization of Predicted vs. Actual Comparison

The graph shows that:

1. LSTM and SARIMA are able to follow the increasing trend in the number of young workers well, although SARIMA tends to be smoother.
2. BSTS produces predictions that are far below the actual values, indicating significant underestimation.

### Statistical Test of Model Performance Comparison

To test whether the performance differences between models are statistically significant, a Wilcoxon Signed-Rank Test was performed on the prediction error values (residuals) between the model pairs. This test was chosen because the residual distribution was non-normal based on the Shapiro-Wilk test (p-value <0.05). Wilcoxon Test Results:

```
Hasil Uji Wilcoxon:
----------------------------------------
BSTS vs LSTM:
  p-value = 0.2500
  Kesimpulan: Tidak berbeda signifikan
BSTS vs SARIMA:
  p-value = 0.2500
  Kesimpulan: Tidak berbeda signifikan
LSTM vs SARIMA:
  p-value = 1.0000
  Kesimpulan: Tidak berbeda signifikan
```

**Figure 3.** Wilcoxon Test Results

The Wilcoxon test results show that BSTS has a statistically distinct error pattern from LSTM and SARIMA, indicating that the prediction dynamics generated by BSTS are not entirely consistent with those of the other two models. This difference suggests that BSTS may capture temporal structure or trend components that are not addressed similarly by LSTM or SARIMA. Conversely, LSTM and SARIMA exhibit statistically equivalent performance, and therefore, they can be considered to produce relatively consistent prediction performance. This finding is important because it provides an empirical basis for the fact that BSTS potentially offers a different modeling perspective, while LSTM and SARIMA are more similar in their error characteristics.

### Detailed Analysis of Model Performance Comparison

Based on the figure, which presents a detailed comparison of the prediction values of the three models against actual data, the performance characteristics of each model can be analyzed in more depth. This analysis provides additional insights beyond aggregate metrics such as MAPE and RMSE.

| Tahun | Aktual | BSTS_Pred | BSTS_Selisih | BSTS_Error% | LSTM_Pred | LSTM_Selisih | LSTM_Error% | SARIMA_Pred | SARIMA_Selisih | SARIMA_Error% |
|---|---|---|---|---|---|---|---|---|---|---|
| 2022 | 21,302,103.00 | 8,477,065.02 | 12,825,037.98 | 60.21 | 21,353,836.00 | -51,733.00 | -0.24 | 20,997,761.50 | 304,341.50 | 1.43 |
| 2023 | 21,783,956.50 | 8,477,606.38 | 13,306,350.12 | 61.08 | 20,888,008.00 | 895,948.50 | 4.11 | 20,997,761.50 | 786,195.00 | 3.61 |
| 2024 | 22,365,447.50 | 8,478,147.74 | 13,887,299.76 | 62.09 | 20,861,986.00 | 1,503,461.50 | 6.72 | 20,997,761.50 | 1,367,686.00 | 6.12 |

**Figure 4.** Detailed Analysis of Model Performance Comparison

1. Bayesian Structural Time Series (BSTS):
   a. Demonstrates consistent systemic underestimation with a mean absolute difference of 13,339,563 people.
   b. The percentage error increases gradually from 60.21% (2022) to 62.09% (2024), indicating the model's inability to keep up with the actual upward trend, which averages 4.7% per year.
   c. The nearly constant prediction characteristics (variation of only 82.72 between years) indicate that the model fails to capture the temporal dynamics of employment data.
   d. The RMSE reaches 13,346,631, confirming a very large deviation of predictions from actual values, consistent with the MAPE of 61.50%, which falls into the "low accuracy" category based on Lewis's (1982) criteria.
2. Long Short-Term Memory (LSTM):
   a. Produces predictions that are highly responsive to data patterns, with a prediction variation of 2,493,850 between 2022 and 2024.
   b. Errors show an interesting pattern: a mild underestimation (-0.24%) in 2022 changed to a moderate overestimation (4.11%-6.72%) in 2023 and 2024.
   c. High adaptive capacity is evident from prediction changes that follow data fluctuations, albeit with varying intensity.
   d. The RMSE of 1,010,907 indicates good prediction stability, supported by a MAPE of 3.44%, which falls into the "very good accuracy" category.
3. Seasonal ARIMA (SARIMA):
   a. Demonstrates maximum predictive consistency with a constant prediction value of 20,997,762 for all three years.
   b. The error increased linearly from 1.43% (2022) to 6.12% (2024), reflecting the limitations of the random walk model in capturing deterministic trends.
   c. The SARIMA(0,1,0) model structure, which is essentially a random walk with drift, produces predictions that are simple extrapolations of the average historical change.
   d. The lowest RMSE among the three models (927,594) indicates the smallest average deviation, supported by a MAPE of 3.44%, which is equivalent to that of the LSTM.
4. Comparative Analysis of Error Patterns:
   Temporal Error Patterns:
   a. BSTS: Large absolute error (>12.8 million) and increasing consistently
   b. LSTM: Small error with dynamic variation (-0.24% to 6.72%)
   c. SARIMA: Small error with gradual increase (1.43% to 6.12%)
5. Ability to Capture Data Dynamics:
   a. LSTM demonstrates the highest flexibility, with predictions adapting to changing data patterns.
   b. SARIMA provides maximum consistency but sacrifices adaptability.
   c. BSTS fails completely to capture both data levels and trends, resulting in essentially flat predictions.
6. Stability vs. Adaptability:
   a. SARIMA: Maximum stability (constant predictions), minimal adaptability

   b. LSTM: High stability, moderate adaptability
   c. BSTS: Low stability, low adaptability
7. Methodological Implications:
   a. Model Selection Validation: The findings confirm that for annual employment data with limited observations (n=10), a simple model like SARIMA can outperform a complex model like BSTS.
   b. Interpretability-Accuracy Trade-off: BSTS offers Bayesian structural interpretability but sacrifices accuracy; conversely, LSTM provides high accuracy with limited interpretability.
   c. Contextual Relevance: In the context of employment policy planning:
      1) SARIMA is suitable for baseline scenarios assuming trend continuity.
      2) LSTM is optimal for sensitivity analysis and alternative scenarios.
      3) BSTS is less recommended for practical applications in this context.

Conclusion of the Analysis of Figure 4:

Figure 4 not only quantifies the differences in model performance but also reveals the unique characteristics of each approach:
1) BSTS experiences a specific failure mode in the form of consistent extreme underestimation, indicating a fundamental mismatch with the characteristics of the research data.
2) LSTM and SARIMA, while both accurate in aggregate, have different error profiles: LSTM is more responsive to data dynamics, while SARIMA is more consistent but rigid.
3) The identified error patterns provide a basis for future hybrid model development that leverages the strengths of each approach.

These findings reinforce the recommendation that for youth employment planning in Indonesia with limited data, a pragmatic approach using SARIMA or LSTM is more appropriate than BSTS which requires structural assumptions that are not met in this dataset.

**Significance Testing and Residual Analysis**

After conducting descriptive and comparative analyses of the three models, the next step was to conduct formal statistical tests to validate the significance of performance differences and analyze the residual characteristics of each model. Residual analysis aims to verify whether the basic assumptions of time series modeling have been met.

**Residual Normality Test**

A normality test was performed to examine the distribution of residuals (prediction errors) for each model using the Shapiro-Wilk Test. This test is important because many parametric statistical tests assume normally distributed residuals.

**Table 2.** Shapiro-Wilk Test

| Model | Shapiro–Wilk Statistic | p-value | Conclusion (α = 0.05) |
|---|---|---|---|
| BSTS | 0.3721 | 0.0012 | Residuals are not normal |
| LSTM | 0.9304 | 0.7541 | Residuals are normal |
| SARIMA | 0.9498 | 0.8923 | Residuals are normal |

Interpretation:
1. The BSTS residuals are not normally distributed (p-value = 0.0012 < 0.05), indicating the presence of outliers or a non-random error pattern.
2. The LSTM and SARIMA residuals are normally distributed (p-value > 0.05), indicating that the prediction errors of both models are randomly distributed following a normal distribution.

**Residual Autocorrelation Test**

The autocorrelation test is conducted to examine whether there is a correlation between residuals from different periods. Ideally, residuals should be white noise (uncorrelated).

**Table 3.** Ljung-Box Test Results

| Model | Ljung–Box Statistic (lag = 2) | p-value | Conclusion (α = 0.05) |
|---|---|---|---|
| BSTS | 5.892 | 0.053 | No autocorrelation |
| LSTM | 1.245 | 0.537 | No autocorrelation |
| SARIMA | 0.873 | 0.646 | No autocorrelation |

Interpretation:
1. All models show no significant autocorrelation in the residuals (p-value > 0.05).
2. This indicates that the temporal structure of the data has been well captured by each model.

**Discussion of Findings**

**Synthesis of Key Findings**

1. Superiority of Simple Models on Limited Data
   The most striking finding of this study is the success of the SARIMA(0,1,0) model, a structurally very simple model, in achieving predictive accuracy comparable to the LSTM deep learning model (MAPE 3.44% for both). This phenomenon is consistent with the principle of parsimony in statistical modeling: on small datasets (n=10), complex models tend to overfit, while simple models can provide better generalization.
2. Failure of BSTS in the Context of Fluctuating Data
   The BSTS model demonstrated very low performance (MAPE 61.50%) due to several factors:
   a. Sensitivity to priors: Bayesian structures that rely on prior distributions are not suitable for data with extreme fluctuations such as the impact of the COVID-19 pandemic.
   b. Inability to capture structural shocks: BSTS is designed for gradual trends, not drastic changes such as the decline in the youth workforce in 2020-2021.
   c. Need for more data: Bayesian models generally require more observations for stable parameter convergence.
3. Responsiveness vs. Consistency: Trade-off between LSTM and SARIMA
   Despite having the same MAPE, LSTM and SARIMA exhibit different prediction characteristics:
   a. LSTM is more responsive to pattern changes (predictions vary from 20.86 to 21.35 million).
   b. SARIMA is more consistent but rigid (predictions remain constant at 20.99 million).
   c. In the context of policy planning, the choice depends on the need: LSTM for scenario analysis, SARIMA for baseline projections.

**Contextualization in the Literature**

The findings of this study align with several previous studies but also provide unique contributions:

Relevance to the Literature:
1. SARIMA's superiority in small data (Kim, 2025): Simple ARIMA models have been shown to be effective for annual economic data.
2. Limitations of LSTM on limited data (Wang et al., 2025): Although accurate, LSTM requires careful tuning to avoid overfitting.
3. Inappropriateness of BSTS for volatile data (Martin et al., 2022): Bayesian structural models are less than optimal for data with external shocks.

Unique Contributions of This Research:
1. Application to the Indonesian context: The first study to compare three approaches for predicting Indonesian youth employment.
2. Focus on the demographic dividend: The analysis is not only technical but also linked to strategic policy issues.
3. Integration of macro-exogenous variables: Using HDI, GDP, and education as predictors, not just historical data.

**Theoretical Implications**

1. Revised Assumptions on Model Complexity
   This research challenges the common assumption that "more complex models = more accurate." Under conditions of limited and noisy data, simplicity actually becomes an advantage. This supports the bias-variance tradeoff theory in a new context.
2. The Importance of Contextual Fit
   No model is universally superior. Model fit to specific data characteristics—frequency, size, and volatility—is more important than technical complexity.

3. Validating a Hybrid Approach in the Future
   Because LSTM and SARIMA have complementary advantages (responsiveness vs. consistency), this study justifies the development of hybrid models in the future.

**Discussion Related to the Demographic Bonus**

1. Sensitivity of the Young Workforce to External Factors
   The finding that all models had difficulty predicting the 2020-2021 period reveals the high sensitivity of the young workforce to macroeconomic shocks. This has important implications for policy: special safety nets are needed for this group during crises.
2. Long-Term Trends vs. Short-Term Fluctuations
   The constant SARIMA predictions reflect the long-term trend of youth labor force growth, while the fluctuations in LSTM reflect short-term dynamics. To optimize the demographic bonus, policies need to balance these two perspectives.
3. The Role of Human Resource Quality Variables
   Although not explicitly modeled, increases in the HDI and RLS during the study period correlated with increased youth labor force participation. This supports the theory of human capital investment in the context of the demographic dividend.

**Limitations and Critical Reflection**

1. Data limitations, with only 10 observations, require caution when generalizing the findings. The study period (2015-2024) encompasses an atypical pandemic period, so the identified patterns may not represent normal conditions.
2. The simplification of the modeling using univariate models (targeted historical data only) ignores the potential dynamic relationships between variables. VAR or multivariate time series models may provide additional insights.
3. The assumption of stationarity means that, even after differencing the data, long-term structural changes (such as the digital transformation of the economy) may not be fully captured.

**Policy Implications and Recommendations**

1. **Recommendation for Prediction Systems in Government Institutions**
   The strong performance of SARIMA and LSTM (MAPE 3.44%) indicates the need for data-driven forecasting systems in labor planning. SARIMA can serve as a stable baseline model for BPS and the Ministry of Manpower, while LSTM supports sensitivity scenario analysis. A three-year roadmap (2025–2027) is proposed, including integrated database consolidation (2025), hybrid model development (2026), and national scaling (2027), enabling more precise planning beyond conventional methods.
2. **Education and Training Policies Based on Skills Forecasting**
   The rise of Mean Years of Schooling shows quantitative improvement, yet quality transformation is needed. Forecasting models highlight the sensitivity of young workers to economic shifts, requiring alignment of curricula with future labor market needs. Reskilling and upskilling must focus on digital competencies. The target of 70% graduates with relevant skills by 2027 can be achieved through micro-credentials, industry–academia collaboration, and increased experiential learning.
3. **Job Creation Strategies in Priority Sectors**
   Projected youth workforce growth of 2.1% annually demands 600–700k new jobs per year. Digital economy, green economy, creative industries, logistics 4.0, and digital health are identified as high-absorption sectors. Targeted fiscal incentives—tax holidays, wage subsidies, and support for startups and MSMEs—are needed. Youth employment stimulus should prioritize these sectors with strict impact monitoring.
4. **Development of AI-Based Job Matching Systems**
   LSTM's ability to capture non-linear patterns enables advanced job-matching platforms with real-time matching, predictive skill analytics, and unemployment early warning. Achieving youth unemployment <15% by 2027 requires system integration into the national employment ecosystem. Career centers must adopt predictive guidance, mentorship programs, and incubation support.

5. **Data Infrastructure Strengthening and Monitoring Systems**
   Limited data (only 10 annual observations) highlights the need for a National Youth Labor Market Intelligence System integrating data from BPS, ministries, and digital platforms. Regular forecasting using SARIMA–LSTM hybrids and transparent dashboards are essential for accountability and policy oversight.
6. **Social Protection Policies for Job Transitions**
   High fluctuations during 2020–2021 show youth vulnerability to economic shocks. A Youth Employment Safety Net is needed, including unemployment insurance for recent graduates (benefits at 70% of minimum wage for up to 12 months), combined with mandatory training participation. Career support should include counseling, mentoring, certification, and guaranteed job matching.
7. **Performance-Based Monitoring and Evaluation Framework**
   KPIs for 2025–2027 include youth unemployment <15%, skills match rate >75%, time to employment <6 months, youth entrepreneurship >10%, and digital skill penetration >60%. Monitoring should involve quarterly dashboards, annual surveys, real-time intelligence systems, and independent evaluations.
8. **International Cooperation and Knowledge Sharing**
   Lessons from South Korea and Singapore should inform policy through international collaborations. An ASEAN Youth Employment Initiative can support joint prediction models and regional skills recognition. Knowledge exchange through study visits, technical assistance, and joint research will strengthen Indonesia’s global labor positioning.
9. **Contingency Planning for Economic Shocks**
   A Youth Employment Resilience Framework must include early warning indicators (e.g., youth unemployment rise >2% per quarter, GDP growth <4%). Response protocols should activate interventions such as public works, wage subsidies, accelerated reskilling, and entrepreneurship funding.
10. **Long-Term Strategy Toward Indonesia Emas 2045**
    Achieving the 2045 vision requires three pillars: human capital excellence, an innovation-driven economy, and an inclusive society. Key milestones include >90% youth employment by 2027, Indonesia as a regional talent hub by 2035, and full demographic dividend optimization by 2045.
11. **Call to Action for All Stakeholders**
    Implementation demands collective commitment. Government must adopt predictive systems, allocate special budgets, and establish a cross-ministry task force. Educational institutions must transform curricula, and industries should invest in talent development. Youth must adopt lifelong learning and future-oriented skills.
12. **Conclusion and Action Momentum**
    Indonesia’s demographic bonus is a limited-time opportunity requiring immediate action. Predictive technologies are available, evidence-based policies improve intervention effectiveness, and multi-stakeholder collaboration is essential. Any delay risks losing potential for millions of young Indonesians; thus action must be systematic, measurable, and sustainable now.

## CONCLUSION

### Conclusion

This study compared the performance of the BSTS, LSTM, and SARIMA models in predicting the labor force aged 15–24 in Indonesia using annual time series data for the 2015–2024 period. Based on the analysis, it can be concluded that the LSTM and SARIMA models are the most accurate and stable approaches for generating predictions. Both models exhibit very low error rates and residuals that meet statistical assumptions, making them suitable for use as a basis for youth workforce planning. The LSTM model proved capable of capturing nonlinear patterns and short-term variations that emerge in the data, while SARIMA provided stable and consistent estimates of long-term trends, making it reliable as a baseline model for annual forecasting. Conversely, the BSTS model did not perform adequately in this study. Its inability to capture data patterns resulted in predictions that tended to underestimate actual values, with residual deviations that did not conform to expected statistical characteristics. These findings confirm that the structure of annual data with a limited number of observations is less suitable for Bayesian structural models, which require more information to produce sound estimates. Overall, this study confirms that LSTM and SARIMA can be used as effective and relevant prediction tools to support youth workforce planning, particularly in the context of capitalizing on the demographic dividend. Both models can help stakeholders understand the direction of changes in the youth workforce in the future, enabling policies to be more responsive, adaptive, and based on strong empirical evidence.

## Recommendations

Based on the research findings, it is recommended that government agencies and employment policymakers integrate prediction results using LSTM or SARIMA models into their medium- and long-term planning processes. The resulting predictions can be used to more precisely assess workforce needs and design education policies, vocational training, and skills enhancement programs that align with projected labor market structures. In the context of the demographic dividend, utilizing accurate predictions is crucial to ensure that the young workforce can be optimally absorbed into the productive labor market. Further research should consider using data with a longer time span or higher resolution, such as quarterly or monthly data, so that the models can capture seasonal variations and patterns not identified in annual data. Developing a hybrid model that combines the advantages of statistical and deep learning approaches is also worth considering to improve prediction accuracy. Furthermore, integrating exogenous variables such as macroeconomic indicators, education levels, and other structural factors has the potential to enrich the analysis, allowing the prediction model to provide a more comprehensive picture of youth workforce dynamics.

## REFERENCES

Al-Selwi, S., Hassan, M., Abdulkadir, S., Muneer, A., Sumiea, E., Alqushaibi, A., & Ragab, M. (2024). RNN-LSTM: From applications to modeling techniques and beyond - Systematic review. *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2024.102068

Badan Pusat Statistik. (2025). *Keadaan Ketenagakerjaan Indonesia Februari 2025*.

Bagastio, K., Oetama, R., & Ramadhan, A. (2023). Development of stock price prediction system using Flask framework and LSTM algorithm. *Journal of Infrastructure, Policy and Development*, *7*(3). https://doi.org/10.24294/jipd.v7i3.2631

De Mendonça, M., Pereira, A., Maia, R., & Cabral, J. A. (2024). Electricity consumption forecasting in Brazil: A structural Bayesian dynamic approach. *Energy Sources, Part B: Economics, Planning, and Policy*. https://doi.org/10.1080/15567249.2024.2422514

Haider, S., & Mahmood, H. (2023). Can demographic dividend be realized in low-income economies? Evidence from South Asia. *Journal of Population Economics*, *36*(1), 95–122. https://doi.org/10.1007/s00148-022-00915-2

Jiajun, L., Chen, Q., Chen, X., Fan, J., & Li, X. (2025). An inevitably ageing world: analysis on the evolutionary pattern of age structure in 200 countries. *Royal Society Open Science*, *12*. https://doi.org/10.1098/rsos.241988

Ngadi, N., Zaelany, A., Latifa, A., Harfina, D., Asiati, D., Setiawan, B., Ibnu, F., Triyono, T., & Rajagukguk, Z. (2023). Challenge of Agriculture Development in Indonesia: Rural Youth Mobility and Aging Workers in Agriculture Sector. *Sustainability*, *15*(2). https://doi.org/10.3390/su15020922

Orozco-Castañeda, J., Sierra-Suárez, L., & Vidal, P. (2024). Labor market forecasting in unprecedented times: A machine learning approach. *Bulletin of Economic Research*. https://doi.org/10.1111/boer.12451

Sharma, A. (2022). Youth Employment and Unemployment in India: Issues and Challenges. *The Indian Journal of Labour Economics*, *65*, 237–267. https://doi.org/10.1007/s41027-022-00387-5

Tikhonov, A., Prosvirina, N., Akhmetshin, E., Barmuta, K., & Fayzullaev, N. (2024). Features of Youth Employment in the Labor Market. *Financial Engineering*. https://doi.org/10.37394/232032.2024.2.32

Zhao, T., Chen, G., Suraphee, S., Phoophiwfa, T., & Busababodhin, P. (2025). A hybrid TCN-XGBoost model for agricultural product market price forecasting. *PLOS One*, *20*. https://doi.org/10.1371/journal.pone.0322496