Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Application of the Autoregressive Integrated Moving Average (ARIMA) Method in Forecasting the Number of Young Construction to Support Human Resource Planning in the Construction Sector

Vincentius Danu Bona Arta Nadeak[1]*, Daniel Isar Valentino Limbong[2], Glenn Desmon Sirait[3], Johannes Antonius Kristian Sitorus[4], Jaya Tata Hardinata[5]

[1, 2, 3, 4, 5]*HKBP Nommensen Pematangsiantar University*
nadeakvincentius@gmail.com[1*], daniellimbong216@gmail.com[2], siraitglenn1@gmail.com[3], johanzxy50@gmail.com[4], jayatatahardinata@uhnp.ac.id

Abstract
The construction sector in Indonesia faces a significant challenge in managing the demand and supply of skilled workers, particularly young construction workers, due to fluctuating market conditions, economic factors, and government policies. This imbalance creates difficulties in effectively planning training programs and allocating resources, ultimately hindering the sector's growth. To address this issue, this study applies the ARIMA (Autoregressive Integrated Moving Average) method to predict the future number of young skilled construction workers in Indonesia from 2012 to 2022. Using data from the Badan Pusat Statistik (BPS) across 34 provinces, the study compares three different ARIMA models: ARIMA(1,1,1), ARIMA(0,1,1), and ARIMA(1,1,0)—to identify the model that provides the most accurate predictions. The results indicate that the ARIMA(0,1,1) model performs best, with the smallest Mean Squared Error (MSE) and the lowest Mean Absolute Percentage Error (MAPE) of 2.5%, offering valuable insights for more effective workforce planning and resource distribution in the Indonesian construction sector.

Keywords: *forecasting, skilled construction workers, Indonesia, time series, prediction.*

1. Introduction

The construction sector plays a very important role in the Indonesian economy, as the main driver of infrastructure development and the fulfillment of basic community needs. With continued economic growth, this sector requires a skilled and competent workforce to support major projects throughout Indonesia. Skilled construction workers are a key element in carrying out high-quality and timely construction projects [1]. One of the main challenges in this sector is ensuring the availability of sufficient, skilled, and qualified workers to meet industry needs. In Indonesia, the construction sector experiences significant fluctuations in the number of skilled workers, which is influenced by various factors such as development policies, economic conditions, and market demand [2].

In recent years, the Autoregressive Integrated Moving Average (ARIMA) method has continued to be widely used in forecasting research due to its ability to capture trend patterns and historical relationships in time series data. For example, research by Dieny et al. (2024) [3]applied the ARIMA model to forecast tilapia production in Bangladesh, demonstrating the effectiveness of ARIMA in predicting commodity production with long-term datasets and robust metric evaluations. Another study by Nafisah et al. (2025) [4] used ARIMA in the context of raw material planning and production forecasting in the traditional songkok industry, showing how ARIMA was integrated with operational approaches to improve business planning accuracy. Furthermore, research by Ruslana (2024) [5] applied ARIMA in predicting the maximum monthly rainfall in Semarang for 2023, where the model was tested and evaluated based on MAPE, showing the level of prediction error that could be used as a basis for decision-making in weather and climate planning. These three examples show that the ARIMA model continues to be used and validated in various applied fields from hydrology, economics, to meteorology.

Young skilled construction workers, in particular, play a vital role in maintaining the sustainability of this sector. However, imbalances between labor supply and demand often occur, either due to a lack of adequate training, changes in development policies, or a mismatch between the skills needed and those available. Given the construction sector's dependence on skilled labor, it is important for the government and industry players to have accurate predictions about the number of workers needed in the future. This will help in planning more effective training programs, optimal resource distribution, and more targeted development policies [2].

In this context, forecasting becomes a very useful tool to help predict labor demand in the construction sector. Labor forecasting requires not only a deep understanding of labor market trends, but also models that can capture the complex dynamics in historical data. One approach that has proven effective in time series forecasting is the Autoregressive Integrated Moving Average (ARIMA) method. The ARIMA model is used to analyze time series data and predict future values based on patterns in historical data [6].
This study aims to apply the ARIMA method in predicting the number of young construction workers in Indonesia for the period 2012 to 2022. The data used was obtained from the Badan Pusat Statistik(BPS), which provides complete information on the number of skilled construction workers in 34 provinces in Indonesia. The data includes variations in the number of workers from year to year, providing a clear picture of the fluctuations that have occurred in the construction sector over the last decade. By using ARIMA, it is hoped that patterns can be found that can be used to predict future labor needs [1].

The ARIMA modeling process involves several important stages, namely identification, parameter estimation, and model diagnostics. In the identification stage, time series data is analyzed to determine the most appropriate ARIMA model sequence [7]. Once the model has been identified, the parameter estimation stage is carried out to obtain the optimal parameter values. Next, diagnostic tests are performed to ensure that the model built is adequate in explaining data variability and has good predictive capabilities. Model performance is evaluated using error metrics such as Mean Squared Error (MSE) [8] and Mean Absolute Percentage Error (MAPE). These metrics help measure the accuracy of the model's predictions. MSE calculates the average of the squared differences between the predicted values and the actual values, where a lower MSE indicates better model performance. On the other hand, MAPE measures the average percentage difference between the predicted and actual values, with a lower MAPE indicating that the model's predictions are closer to the actual data. Both MSE and MAPE are commonly used to assess the effectiveness of time series models, ensuring that the model can make accurate predictions for future data. [9].

The application of ARIMA in the context of construction labor is highly relevant given the nature of data that has seasonal patterns or long-term trends. Many external factors affect the construction sector, such as changes in government policy, economic developments, and natural disasters, which can create fluctuations in the number of workers. Therefore, proper modeling and the use of advanced forecasting techniques such as ARIMA are essential for planning labor requirements more effectively [10].

One of the main challenges in this study is ensuring that the data used is sufficiently representative and does not contain many gaps or anomalies that could interfere with the accuracy of the model. Therefore, the first step in this study is the data pre-processing stage, which includes checking for missing values and outliers. In addition, to ensure that the ARIMA model works properly, the data used must be stationary, meaning that there are no trends or unstable fluctuations in the data. If the data is not stationary, a differentiation process will be performed to make it stationary [11].

After the pre-processing stage is complete, the ARIMA model will be applied to predict the number of young construction workers in the future, focusing on 34 provinces in Indonesia. This process involves testing various ARIMA model sequences to find the model that is most effective in capturing historical data patterns. Once the best model is found, predictions will be made for longer periods, providing an overview of possible future trends [12].

The benefits of this research are enormous for the construction industry and public policy. With accurate predictions of the number of young construction workers, the government and industry players can plan better training programs, allocate resources efficiently, and reduce dependence on foreign workers. In addition, good predictions will help address the imbalance between labor supply and demand, which is often a major problem in this sector [13]. Overall, this study is expected to make a significant contribution to the development of the construction sector in Indonesia. Using the ARIMA method, which is known to be effective in handling time series data, this study not only provides insight into current labor trends, but also provides a strong foundation for data-driven decision making in the future. The accuracy of these predictions will depend heavily on the selection of the appropriate model and careful data processing, so that the results can be optimally utilized in planning future labor needs for Indonesia's construction sector [14].

2. Literature Review

2.1. ARIMA (Autoregressive Integrated Moving Average)

ARIMA is a statistical model used to analyze and predict time series data. This model consists of three main components, namely Autoregressive (AR), Integrated (I), and Moving Average (MA), each of which has a function to handle certain aspects of time series data [15].

Table 1: Main of ARIMA Model

Autoregressive (AR)	This component explains the relationship between current values and previous values in the data. In labor force forecasting, AR can be used to see the effect of past labor force values on the amount of labor force needed in the future.
Integrated (I)	This component aims to make data stationary by reducing or eliminating trends in the data. This process is usually done by differentiation, which is calculating the difference between data values in a certain time period and the previous period. Differentiation helps eliminate instability in the data that can interfere with prediction accuracy
Moving Average (MA)	This component addresses random fluctuations in the data by taking into account the average of past prediction errors. MA is used to adjust the model so that it can capture unexpected variations and fluctuations that occur in the data.

2.2. Forecasting

Forecasting is a statistical technique used to predict future values or events based on data observed in previous periods. In the context of the labor sector, labor forecasting aims to estimate the amount of labor needed in the future based on patterns that emerge in historical data [16]. The forecasting process can be carried out using various techniques, ranging from simple to complex methods. Time series forecasting models are one of the most commonly used approaches, especially for time-based sequential data. Workforce forecasting can cover various aspects, such as predicting the number of workers needed in the construction sector at a given time, training needs, and budget allocation for the workforce sector. The forecasting techniques used must be able to capture the variability and seasonal patterns or trends in labor data so that the prediction results can provide accurate insights for policymakers [17].

Forecasting plays a crucial role in planning and decision-making processes across various sectors, including the labor market. By analyzing historical data and identifying underlying trends, patterns, and cyclical behaviors, forecasting allows organizations to anticipate future demands and allocate resources effectively. In the labor sector, this means predicting not only the required workforce numbers but also understanding shifts in labor demand due to economic, technological, or demographic changes. Advanced forecasting models, such as those incorporating machine learning or artificial intelligence, can further improve accuracy by adapting to new data inputs and refining predictions over time. Accurate labor forecasts can significantly reduce risks and improve efficiency, helping organizations better prepare for future challenges and optimize workforce management strategies.

2.3. Time Series

A time series is a series of data observed or collected in a specific time sequence, such as monthly or annual data . In the context of labor force forecasting, a time series could be data on the number of skilled construction workers collected annually in various provinces. Time series data is typically used to analyze trends and seasonal patterns that occur over a long period of time [18].

It is important to note that time series data generally has three main components: trend, seasonality, and random fluctuations. Trend refers to long-term changes in the data, such as an increase in the workforce due to infrastructure development policies. Seasonality refers to patterns that repeat at certain intervals, such as an increase in labor demand during certain seasons. Meanwhile, random fluctuations are components that are difficult to predict because they are influenced by unexpected external factors, such as economic crises or natural disasters [19].

2.4. Construction Experts Labor Force

Construction professionals refer to individuals who have technical skills and expertise in the field of construction that enable them to play a role in various stages of a construction project, from planning to implementation and supervision. These professionals consist of various categories of professions, such as civil engineers, architects, technicians, and trained field workers. They not only possess technical expertise, but also a deep understanding of the principles of safety, quality, and cost and time efficiency in every project [20].

In the Indonesian context, the need for young construction workers is very relevant because the construction sector plays a role as one of the driving forces of the rapidly growing economy. However, a common problem is the uncertainty of the number of skilled workers available to meet industry demand, given the fluctuating nature of projects, which are highly dependent on government policy, market conditions, and economic cycles. Therefore, predicting the number of skilled construction workers is crucial so that education and training plans can be tailored to market needs.

3. Research Method

This study uses a quantitative approach with a time series model-based forecasting method, namely ARIMA (Autoregressive Integrated Moving Average), to predict the number of young construction workers in 34 provinces in Indonesia during the period 2012 to 2022. The steps in this study include data collection, data pre-processing, ARIMA model identification and estimation, and model performance evaluation using forecasting error metrics. Overall, this study aims to provide accurate predictions of future skilled construction worker demand, which can be used to plan labor policies and develop the construction sector in Indonesia.

3.1. Data Collection

Data collection is the first step in the research process. The data used in this study is the number of young construction workers in 34 provinces in Indonesia from 2012 to 2022. This data source was taken from the Badan Pusat Statistik(BPS), which periodically publishes data related to labor in the construction sector. The data used is time series data that describes the number of workers each year.

This data contains crucial information regarding the distribution of skilled construction workers, which is influenced by various economic factors, development policies, and market demand. This data will serve as the basis for the ARIMA model to forecast future labor supply.

Table 2: Data Collection Number of Construction Experts by Province and Qualification (Persons)

Provinces	Number of Skilled Construction Workers by Province and Qualification (Persons)									
	Young									
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2022
ACEH	3056	140	322	320	6104	4437	4543	4058	3856	146
SUMATERA UTARA	3393	595	1309	700	2735	1635	2041	1663	1666	257

SUMATERA BARAT	1808	140	185	331	5954	5234	4373	3452	3099	152
RIAU	2437	910	2826	755	13013	14052	17320	16614	15505	115
JAMBI	545	70	256	135	731	519	880	1068	1094	665
SUMATERA SELATAN	877	105	90	294	1249	1089	1412	1766	2060	168
	…	…	…	…	…	…	…	…	…	…
MALUKU	751	140	178	130	634	505	731	892	900	498
MALUKU UTARA	536	70	22	297	233	247	143	127	219	351
PAPUA BARAT	718	70	124	41	486	450	587	572	501	395
PAPUA	1865	105	149	201	699	359	380	652	708	949

3.2. Splitting Data into Training and Test Data

After the data is collected, the next step is to divide the data into two main parts: training data and testing data. This dataset contains the annual number of certified construction experts (young qualifications) for each province in Indonesia. In this study, a subset of provinces was used to create time series: Gorontalo, West Sulawesi, and Maluku were used for training, while Maluku, North Maluku, and West Papua were used for testing to simulate future predictions for new provincial segments. This data will be divided with a 50:50 ratio, where 50% of the data is used for training and the remaining 50% is used for testing. The available years are 2012 to 2022, with one year missing (2021) in the provided file. Each province provides 10 observations. This time series is constructed by combining provincial observations chronologically based on the year in each province.

Table 3: Splitting Result Training and Test Data

Subset	Province	Observation
Training	Gorontalo; West Sulawesi; Maluku	3 provinces × 10 years = 30 points (20 supervised samples)
Testing	Maluku; North Maluku; West Papua	3 provinces × 10 years = 30 points (20 supervised samples)

Training data is used to train models or algorithms that will be used in research to make predictions. In the context of this study, the training data contains observations from three selected provinces, namely Gorontalo, West Sulawesi, and Maluku. Each of these provinces has data for 10 years (from 2012 to 2022, except for 2021, which is missing).

Number of Training Data Observations:

1. 3 Provinces × 10 Years = 30 Observations
2. However, only 20 supervised samples were used in model training. This means that even though there were 30 observations, only 20 samples were used to train the model, based on lag and target division.

Table 4: Training data

X	Training data GORONTALO-WEST SULAWESI-MALUKU										
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2022	Target
1	126	7	24	17	6522	6300	5983	2091	2583	581	667
2	7	24	17	6522	6300	5983	2091	2583	581	667	70
3	24	17	6522	6300	5983	2091	2583	581	667	70	29
4	17	6522	6300	5983	2091	2583	581	667	70	29	71
5	6522	6300	5983	2091	2583	581	667	70	29	71	234
6	6300	5983	2091	2583	581	667	70	29	71	234	338
	…	…	…	…	…	…	…	…	…	…	…
17	718	1069	1169	144	751	140	178	130	634	505	731
18	1069	1169	144	751	140	178	130	634	505	731	892
19	1169	144	751	140	178	130	634	505	731	892	900
20	144	751	140	178	130	634	505	731	892	900	498

Test data is used to measure the performance of models that have been trained on training data. After the model learns from the training data, it will be tested with data that it has never seen before to determine how accurate its predictions are. The test data consists of 3 provinces that are different from the training data, namely Maluku, North Maluku, and West Papua. Each province also has data for 10 years (from 2012 to 2022, except for 2021, which is missing).

Number of Test Data Observations:

1. 3 Provinces × 10 Years = 30 Observations
2. Similar to the training data, 20 supervised samples are used in model evaluation.

Table 5: Test data

X	Test Data MALUKU-NORTH MALUKU-WEST PAPUA										
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2022	Target
1	751	140	178	130	634	505	731	892	900	498	536
2	140	178	130	634	505	731	892	900	498	536	70
3	178	130	634	505	731	892	900	498	536	70	22
4	130	634	505	731	892	900	498	536	70	22	297
5	634	505	731	892	900	498	536	70	22	297	233
6	505	731	892	900	498	536	70	22	297	233	247
	…	…	…	…	…	…	…	…	…	…	…
17	143	127	219	351	718	70	124	41	486	450	587
18	127	219	351	718	70	124	41	486	450	587	572
19	219	351	718	70	124	41	486	450	587	572	501
20	351	718	70	124	41	486	450	587	572	501	395

The division between training data and test data is important to ensure that the model can learn patterns in the data in general and does not just memorize existing data (overfitting). By testing using previously unseen data, the model can be evaluated on its ability to make predictions on newer and more relevant data.

3.3. Data Preprocessing (Normalization)

During the pre-processing stage, normalization is performed on both data sets, namely the training data and testing data, to ensure that all values are on a uniform scale and the ARIMA model can process the data efficiently. Normalization is performed using the Min-Max Scaling technique, which converts data values into the range [0, 1] using the formula:

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

In the training data, normalization is calculated using the min and max values from the training data itself, which includes the provinces of Gorontalo, West Sulawesi, and Maluku. After that, the min and max values obtained from the training data are used to normalize the test data, which includes the provinces of Maluku, North Maluku, and West Papua, so that both have a uniform scale. This process helps prevent the model from prioritizing features with larger scales and speeds up the training process, while ensuring that the model can be applied consistently to previously unseen data.

Normalization of Training Data:

1. Training Data covers the provinces of Gorontalo, West Sulawesi, and Maluku with data from 2012 to 2020.
2. The minimum and maximum values are calculated based on this training data. For example, for Gorontalo province, data from 2012 to 2020 is used to find the minimum and maximum values for the number of skilled construction workers.
3. Normalization is performed on each province separately so that comparisons between provinces can be made on a uniform scale.

Table 6: Normal of Training Data

X	2012	2013	2014	2015	2016	2017	2018	2019	2020	2022	Target
1	0.103165	0.1	0.100452	0.100266	0.273294	0.267389	0.258957	0.155433	0.16852	0.115268	0.117556
2	0.1	0.100452	0.100266	0.273294	0.267389	0.258957	0.155433	0.16852	0.115268	0.117556	0.101676
3	0.100452	0.100266	0.273294	0.267389	0.258957	0.155433	0.16852	0.115268	0.117556	0.101676	0.100585
4	0.100266	0.273294	0.267389	0.258957	0.155433	0.16852	0.115268	0.117556	0.101676	0.100585	0.101702
5	0.273294	0.267389	0.258957	0.155433	0.16852	0.115268	0.117556	0.101676	0.100585	0.101702	0.106038
6	0.267389	0.258957	0.155433	0.16852	0.115268	0.117556	0.101676	0.100585	0.101702	0.106038	0.108804
	…	…	…	…	…	…	…	…	…	…	…
17	0.118912	0.128248	0.130908	0.103644	0.11979	0.103538	0.104548	0.103272	0.116678	0.113246	0.119258

18	0.128248	0.130908	0.103644	0.11979	0.103538	0.104548	0.103272	0.116678	0.113246	0.119258	0.12354
19	0.130908	0.103644	0.11979	0.103538	0.104548	0.103272	0.116678	0.113246	0.119258	0.12354	0.123753
20	0.103644	0.11979	0.103538	0.104548	0.103272	0.116678	0.113246	0.119258	0.12354	0.123753	0.11306

Normalization of Test Data:

1. Test Data covers the provinces of Maluku, North Maluku, and West Papua with data from 2021 to 2022.
2. Test data will be normalized using min and max parameters calculated from training data. This process is important to maintain consistency in how data is processed by models trained on training data.

Tabel 7: Normal of Test Data

X	2012	2013	2014	2015	2016	2017	2018	2019	2020	2022	Target
1	0.11979	0.103538	0.104548	0.103272	0.116678	0.113246	0.119258	0.12354	0.123753	0.11306	0.114071
2	0.103538	0.104548	0.103272	0.116678	0.113246	0.119258	0.12354	0.123753	0.11306	0.114071	0.101676
3	0.104548	0.103272	0.116678	0.113246	0.119258	0.12354	0.123753	0.11306	0.114071	0.101676	0.100399
4	0.103272	0.116678	0.113246	0.119258	0.12354	0.123753	0.11306	0.114071	0.101676	0.100399	0.107714
5	0.116678	0.113246	0.119258	0.12354	0.123753	0.11306	0.114071	0.101676	0.100399	0.107714	0.106011
6	0.113246	0.119258	0.12354	0.123753	0.11306	0.114071	0.101676	0.100399	0.107714	0.106011	0.106384
	…	…	…	…	…	…	…	…	…	…	…
17	0.103618	0.103192	0.105639	0.10915	0.118912	0.101676	0.103112	0.100904	0.112741	0.111783	0.115428
18	0.103192	0.105639	0.10915	0.118912	0.101676	0.103112	0.100904	0.112741	0.111783	0.115428	0.115029
19	0.105639	0.10915	0.118912	0.101676	0.103112	0.100904	0.112741	0.111783	0.115428	0.115029	0.11314
20	0.10915	0.118912	0.101676	0.103112	0.100904	0.112741	0.111783	0.115428	0.115029	0.11314	0.110321

3.4. Identifying ARIMA Models with ACF and PACF Analysis

The ARIMA model is identified using AutoCorrelation Function (ACF) and Partial AutoCorrelation Function (PACF) analysis to determine the appropriate parameters, namely p (order of Autoregressive), d (degree of differencing), and q (order of Moving Average). ACF is used to identify the MA (Moving Average) component by looking at the correlation between the current data value and the previous data value. The pattern obtained from ACF will help determine the q order, which is how much the past values affect the prediction error. On the other hand, PACF is used to determine the AR (Autoregressive) component, which measures the direct relationship between data values at a certain time and data values at a previous time, without the influence of other values in between. The results of PACF will help determine the order p, which is how much previous data influences the value being analyzed.

After the p and q values are determined through ACF and PACF analysis, the next step is to ensure that the data used is stationary. If the data is not stationary, a differentiation process is performed to eliminate trends and make the data more stable. This process produces an appropriate ARIMA model based on the parameters that have been determined, which is then applied to the training data. The identified model is then tested using test data to evaluate prediction accuracy using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Akaike Information Criterion (AIC), to ensure that the ARIMA model can accurately predict the number of skilled construction workers. The following table shows the results of ARIMA model identification based on ACF and PACF analysis:

Table 8: Results of ARIMA model identification based on ACF and PACF analysis

ARIMA Components	Identification Results
p (Autoregressive)	0 (A significant spike at lag 1 in the PACF analysis indicates AR(1))
d (Differencing)	1 (One differencing is applied to make the data stationary after the ADF test)
q (Moving Average)	1 (A significant correlation at lag 1 in the ACF analysis indicates MA(1))

Identified ARIMA Model: ARIMA(0,1,1)

1. p = 0: Indicates that one previous lag affects the current value.
2. d = 1: Indicates that one differentiation is required to make the data stationary.
3. q = 1: Indicates that errors in lag 1 affect the prediction of the current value.

This model is then used to train the training data and tested using the test data to evaluate its performance.

3.5. Model Estimation with Training Data

After the ARIMA model identification process is complete and the p, d, and q parameters have been determined, the next step is to estimate the ARIMA model using the training data. Model estimation is performed to obtain the ARIMA model coefficients that will be used to predict the number of skilled construction workers in the future.
Estimating the ARIMA model with training data involves two main things:

1. Training the ARIMA model on the processed training data.
2. Calculating the ARIMA model parameters based on the training data used, namely the coefficients for the AR (p), MA (q), and differentiation (d) components.

3.6. Model Estimation with Test Data

Estimating the model with test data is an important step to test how well the trained ARIMA model can predict values in previously unseen data. After the model is trained using training data, we apply the model to test data to predict the desired values, such as the number of skilled construction workers in the future. These predictions are then compared to the target values (original data) in the test data to evaluate the accuracy of the model. During this process, the MSE (Mean Squared Error) metric is used to measure the magnitude of the error between the predicted value and the original value, where a lower MSE value indicates better model performance. In addition, MAPE (Mean Absolute Percentage Error) is also used to evaluate the accuracy of the model, with models that have a lower MAPE value considered more optimal. The results of this estimation provide an overview of the model's ability to predict values that are not in the training data. The comparison between the original data and the predictions is presented in tabular form, making it easier to understand and analyze. In addition, graphical visualizations are also used to display prediction trends compared to actual data, providing a clearer picture of whether the ARIMA model is able to accurately capture patterns and trends in the data. By performing estimates on test data, we can assess the reliability and accuracy of the model, which is very important for further applications in future data forecasting.

4. Results

In time series analysis, selecting the right model is crucial for generating accurate predictions. One commonly used approach is the ARIMA (AutoRegressive Integrated Moving Average) model, which can handle data that has trends and seasonal patterns. In this study, three ARIMA models with different parameters were tested to predict the data. The three models were ARIMA(1,1,1), ARIMA(0,1,1), and ARIMA(1,1,0), which differed in their AR (AutoRegressive), I (differentiation), and MA (Moving Average) components. In this analysis, we included images of the results of the model analysis using Python, which show a comparison between the training data, test data, and prediction results generated by the three models.

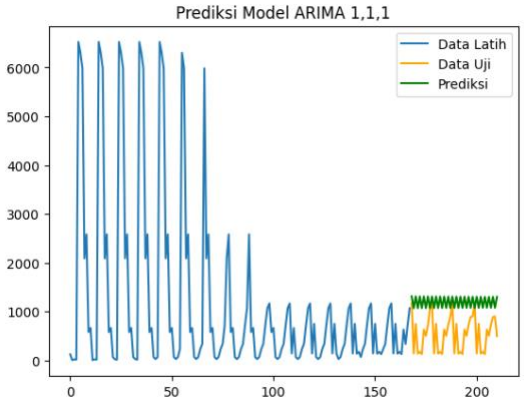

Fig 1: Result of Predict use ARIMA(1, 1, 1)

The figure shows the prediction results using the ARIMA(1,1,1) model. It can be seen that this model has difficulty following the test data pattern. Although this model has one AutoRegressive (AR) component, one differentiation (I), and one Moving Average (MA), the prediction results (green line) deviate significantly from the test data (orange line), especially at the end. This indicates that more complex models do not always produce more accurate predictions, and in this case, the model tends to overfit and fails to capture the test data pattern well.

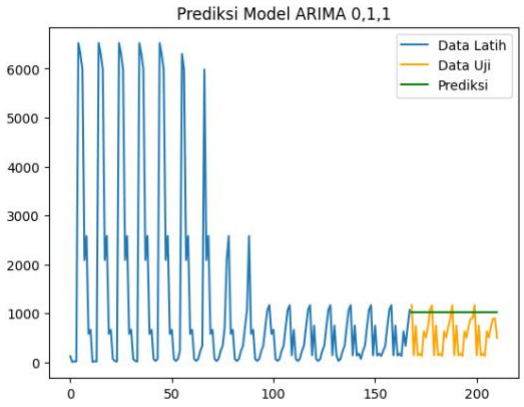

Fig 2: Result of Predict use ARIMA(0, 1, 1)

The figure shows the prediction results using the ARIMA(0,1,1) model. It can be seen that this model, which only uses MA components and one differentiation, provides better results. The prediction (green line) is closer to the test data (orange line) than the previous model. Although there are slight deviations, the prediction results are smoother and more accurate, indicating that a simpler model can be more effective in handling this data.

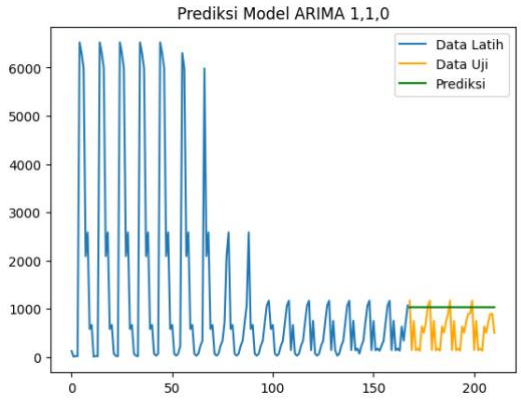

Fig 3: Result of Predict use ARIMA(1, 1, 0)

The third figure, which shows the prediction results using the ARIMA(1,1,0) model, also provides good results, although it is slightly more volatile than ARIMA(0,1,1). This model uses one AR component and one differentiation, without an MA component. The prediction (green line) appears to be quite accurate and better than the ARIMA(1,1,1) model, although slightly worse than the ARIMA(0,1,1) model.

In this study, three different ARIMA models were tested to predict data, namely ARIMA(1,1,1), ARIMA(0,1,1), and ARIMA(1,1,0). The ARIMA model is one of the popular methods in time series analysis used to model data that has trends and seasonal patterns. Each model has different parameters to handle the autoregressive (AR), differencing (I), and moving average (MA) components. These models were tested to see which one provided the best prediction results based on two main evaluation metrics: Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE).

Table 9: MSE and MAPE comparison table

Model	MSE	MAPE
ARIMA(1,1,1)	537546	2,9
ARIMA(0,1,1)	352795	2,5
ARIMA(1,1,0)	359098	2,5

Based on the results obtained, ARIMA(0,1,1) showed the best performance with the smallest MSE value of 352.795 and the lowest MAPE of 2.5%. This model relies on only one Moving Average (MA) component and one differentiation, which proved to be quite effective in handling the data with lower errors in both square and percentage forms. Meanwhile, ARIMA(1,1,1), although more complex with one

AutoRegressive (AR) component and one MA, actually produced a much higher MSE of 537.546 and the highest MAPE of 2.9%, indicating that adding components does not always contribute to improved prediction accuracy and can cause overfitting.

ARIMA(1,1,0), which combines one AR component and one differentiation without MA, provides results similar to ARIMA(0,1,1), with an MSE of 359.098 and a MAPE of 2.5%. Although ARIMA(1,1,0) has a slightly higher MSE than ARIMA(0,1,1), both models still produce a lower error rate than ARIMA(1,1,1). This shows that models with fewer components can be more effective in predicting relatively simple data, without the need to handle unnecessary complexity.

Based on the MSE and MAPE results, ARIMA(0,1,1) is the most effective model for this dataset, providing predictions with the smallest error. ARIMA(1,1,0) also provides good results, but is slightly higher in terms of MSE. Conversely, ARIMA(1,1,1) shows that more complex models are not always better, especially if the data does not require many components. Therefore, the ARIMA(0,1,1) model should be chosen for prediction applications on similar data due to its ability to produce more accurate predictions with smaller errors.

5. Conclusion

From the evaluation results, the ARIMA(0,1,1) model showed the best performance, with the smallest Mean Squared Error (MSE) of 352.795 and the lowest Mean Absolute Percentage Error (MAPE) of 2.5%. This indicates that a model with simpler components, relying only on one Moving Average (MA) and one differentiation (I), is more effective in handling the available data. This model is capable of providing smoother and more accurate predictions even though it only uses two components. This proves that a simpler approach can be more efficient in addressing labor forecasting in the construction sector.

Meanwhile, the ARIMA(1,1,1) model, which is more complex with one AutoRegressive (AR) component, one differentiation (I), and one Moving Average (MA), produced poorer results with an MSE of 537.546 and a MAPE of 2.9%. These results indicate that adding more components does not always improve prediction accuracy. In fact, more complex models such as ARIMA(1,1,1) actually show an increase in error, indicating the potential for overfitting, where the model adapts too closely to the training data and fails to capture more general patterns in the test data.

The ARIMA(1,1,0) model, which combines one AR and one differentiation (I) without MA, provides fairly good results with an MSE of 359.098 and a MAPE of 2.5%. Although slightly worse than ARIMA(0,1,1) in terms of MSE, both models produce almost similar prediction errors. This shows that although ARIMA(1,1,0) is slightly more complex than ARIMA(0,1,1), both models are still more effective than ARIMA(1,1,1) in predicting the number of young construction workers.

ARIMA(0,1,1) is the most effective and efficient model for predicting the demand for young construction workers in Indonesia. This model provides more accurate results with smaller prediction errors, both in terms of MSE and MAPE. The results of this study provide important insights for the construction industry and policymakers in Indonesia to plan more targeted training programs, resource distribution, and development policies, based on more accurate estimates of the number of workers needed in the future.

The evaluation results demonstrate that the ARIMA(0,1,1) model is the most effective and efficient for predicting the demand for young construction workers in Indonesia. With the smallest Mean Squared Error (MSE) of 352.795 and the lowest Mean Absolute Percentage Error (MAPE) of 2.5%, it provides smoother and more accurate predictions using only two components one Moving Average (MA) and one differentiation (I). This finding addresses the problem of labor forecasting in the construction sector by showing that a simpler model can outperform more complex ones. While the ARIMA(1,1,1) model, with its additional AutoRegressive (AR) component, led to higher prediction errors (MSE of 537.546, MAPE of 2.9%), indicating overfitting, the ARIMA(0,1,1) model proved that simplicity can lead to more efficient and reliable forecasts. The ARIMA(1,1,0) model, slightly more complex but close to ARIMA(0,1,1), also produced similar results but was less effective. Thus, ARIMA(0,1,1) offers the best balance of accuracy and simplicity, providing valuable insights for more accurate labor demand predictions. This conclusion addresses the key problem of forecasting labor needs in the construction sector, offering a robust model that can help the industry and policymakers in planning more effective training programs, resource distribution, and development policies.

6. Suggestions

Based on the results of this study, it is recommended that relevant parties in the construction industry and government policy makers use the ARIMA(0,1,1) model to plan future construction labor needs. In addition, it is important to continuously update data on a regular basis so that the model used remains relevant and can provide more accurate predictions. In the future, further research can be conducted by combining the ARIMA model with other techniques, such as Machine Learning or Artificial Intelligence, to improve the accuracy of predictions in more complex conditions, especially to deal with unexpected fluctuations in the construction sector.

References

[1] I. Basuki, “Tantangan Tenaga Kerja Konstruksi Dalam Infrastruktur Transportasi Berkelanjutan Menuju Indonesia Emas 2045,” 2024.
[2] F. Rachim, M. Tumpu, and Mansyur, “Research on Predicting Skilled Labor Availability to Enhance Sustainability Building Practices,” *International Journal of Sustainable Development and Planning*, vol. 19, no. 11, pp. 4183–4192, Nov. 2024, doi: 10.18280/ijsdp.191108.
[3] A. R. Dieny and T. F. Sutrisno, “Production Planning Forecasting using Seasonal and Non-Seasonal ARIMA Method with Minitab Applications (Study Case: DC Company),” *Journal of Economics and Management Scienties*, pp. 394–403, Jan. 2026, doi: 10.37034/jems.v8i2.259.
[4] S. Nafisah, A. R. E. Najaf, and P. K. F. Ananto, “Forecasting and Raw Material Planning in Traditional Songkok Production Using ARIMA and Simple Exponential Smoothing,” *JUSIFO (Jurnal Sistem Informasi)*, vol. 11, no. 1, pp. 31–42, Jun. 2025, doi: 10.19109/jusifo.v11i1.27833.
[5] Z. N. Ruslana, R. S. Prihatin, S. Sulistiyowati, and K. Nugroho, “Application of the Arima Method to Prediction Maximum Rainfall at Central Java Climatological Station,” *sinkron*, vol. 8, no. 4, pp. 2135–2141, Oct. 2024, doi: 10.33395/sinkron.v8i4.13984.

[6] L. M. Malihah and G. T. Meilania, "Perbandingan Model Peramalan Jumlah Pencari Kerja Menggunakan Arima Dan Double Exponential Smoothing," *Jurnal Litbang Sukowati : Media Penelitian dan Pengembangan*, vol. 7, no. 2, pp. 169–178, Nov. 2023, doi: 10.32630/sukowati.v7i2.441.

[7] A. Yusapra Salim *et al.*, "Analisis Deret Waktu Data Perencanaan Tenaga Kerja pada Perusahaan Manufaktur Menggunakan Model ARIMA Time Series Analysis of Man Power Planning Data at Manufacturing Company Using ARIMA Model," vol. 2024, no. 2, pp. 481–492, 2024, doi: 10.51132/teknologika.v14/2.

[8] J. T. Hardinata *et al.*, "Analisis Algoritma Fletcher-Reeves dalam Penentuan Model Terbaik untuk Prediksi Harapan Lama Sekolah di Sumatera Utara Analysis of the Fletcher-Reeves Algorithm in Determining the Best Model for Predicting School Life Expectancy in North Sumatra Article Info ABSTRAK," *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, vol. 2, no. 1, pp. 2828–9099, 2023, doi: 10.55123/jomlai.v2i1.1819.

[9] J. T. Tsoku, D. Metsileng, and T. Botlhoko, "A Hybrid of Box-Jenkins ARIMA Model and Neural Networks for Forecasting South African Crude Oil Prices," *International Journal of Financial Studies*, vol. 12, no. 4, Dec. 2024, doi: 10.3390/ijfs12040118.

[10] A. Pangestu, A. Irma Purnamasari, and I. Ali, "Analisis Peramalan Tingkat Pengangguran Terbuka di Jawa Barat: Pendekatan Time Series menggunakan Metode ARIMA," 2024. [Online]. Available: http://creativecommons.org/licences/by/4.0/

[11] A. S. AlSalehy and M. Bailey, "Improving Time Series Data Quality: Identifying Outliers and Handling Missing Values in a Multilocation Gas and Weather Dataset," *Smart Cities*, vol. 8, no. 3, Jun. 2025, doi: 10.3390/smartcities8030082.

[12] G. S. Osho, "A General Framework for Time Series Forecasting Model Using Autoregressive Integrated Moving Average-ARIMA and Transfer Functions," *Int. J. Stat. Probab.*, vol. 8, no. 6, p. 23, Sep. 2019, doi: 10.5539/ijsp.v8n6p23.

[13] M. F. Rizvi, "ARIMA Model Time Series Forecasting," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 12, no. 5, pp. 3782–3785, May 2024, doi: 10.22214/ijraset.2024.62416.

[14] M. Ma, V. W. Y. Tam, K. N. Le, and R. Osei-Kyei, "A systematic literature review on price forecasting models in construction industry," *International Journal of Construction Management*, vol. 24, no. 11, pp. 1191–1200, 2024, doi: 10.1080/15623599.2023.2241761.

[15] D. Gunawan and W. Astika, "The Autoregressive Integrated Moving Average (ARIMA) Model for Predicting Jakarta Composite Index," *Jurnal Informatika Ekonomi Bisnis*, Feb. 2022, doi: 10.37034/infeb.v4i1.114.

[16] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, "A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks," Aug. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/fi15080255.

[17] A. Fatkhudin, F. A. Artanto, F. Zamaroh, and V. A. Azarine, "Evaluasi Metode Exponential Smoothing dan Moving Average Untuk Peramalan Data Pengangguran di Indonesia," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 5, pp. 1227–1238, May 2025, doi: 10.52436/1.jpti.640.

[18] S. S. W. Fatima and A. Rahimi, "A Review of Time-Series Forecasting Algorithms for Industrial Manufacturing Systems," Jun. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/machines12060380.

[19] F. Damayanti, S. Sundari, and R. Liza, "Analisis Laju Pembelajaran pada Backpropagation dalam Memprediksi Bencana Alam Akibat Cuaca Ekstrim," vol. 16, no. 1, p. 2023.

[20] A. Yusapra Salim *et al.*, "Analisis Deret Waktu Data Perencanaan Tenaga Kerja pada Perusahaan Manufaktur Menggunakan Model ARIMA Time Series Analysis of Man Power Planning Data at Manufacturing Company Using ARIMA Model," vol. 2024, no. 2, pp. 481–492, doi: 10.51132/teknologika.v14/2.