Revenue Forecasting in Intelligent Water Management Systems using Arima Time Series Model

Coraina Y. Torar
Department of Computer Engineering, Manado State Polytechnic

Gloria Manggala
Department of Computer Engineering, Manado State Polytechnic

Meilani J. Ngantung
Department of Computer Engineering, Manado State Polytechnic

ABSTRACT

Smart Water Management System requires accurate revenue estimates for operational efficiency and financial planning. This study aims to implement the ARIMA (Autoregressive Integrated Moving Average) time-series model for revenue forecasting in the Smart Water System at the Manado State Polytechnic. The system provides four water volume packages (600 ml, 1,500 ml, 5,000 ml, and 19,000 ml) with three transaction methods: (1) online ordering - online pay (digital balance), (2) online order - pay cash, and (3) direct order - pay cash. The simulation data is generated for a period of 90 days (equivalent to three months of operation) by considering realistic transaction patterns based on literature studies and field observations on the Smart Water System. The stages of ARIMA modeling include stationary testing (Augmented Dickey-Fuller), parameter identification (ACF/PACF), model estimation, residual diagnostics, and forecasting for the next 30 days. The accuracy of the model was evaluated using MAPE (Mean Absolute Percentage Error). The results showed that the ARIMA(1,1,1) model provided the best compatibility with MAPE of 11.8%, which was categorized as good forecasting accuracy. This model successfully captured daily revenue patterns and seasonal trends. The revenue forecast for March 2026 shows an upward trend with a total projection of IDR 13,090,500. These findings show that ARIMA's time series modeling can effectively support financial planning and operational management of smart water systems.

General Terms

Time Series Analysis, Forecasting, Performance Evaluation

Keywords

ARIMA, Time Series Forecasting, Revenue Prediction, Smart Water Systems, Financial Planning, Water Management.

1. INTRODUCTION

The Smart Water Management System integrates water treatment technology, digital transaction systems, and IoT-based machine automation to provide efficient drinking water services. The system is designed to serve the general public around the campus, including the Manado State Polytechnic academic community by providing four water volume packages, namely 600 ml, 1,500 ml, 5,000 ml, and 19,000 ml. Three transaction methods are provided for the convenience of users: (1) online order - online pay (digital balance), (2) online order - pay cash, and (3) direct message - pay cash. The Manado State Polytechnic was chosen as the initial implementation location, with future development targets to reach a wider community. Forecasting is a science or method to predict events that will occur in the future using time series data through a mathematical model approach [1]. Accurate predictions allow managers to prepare adequate water stocks based on demand patterns, schedule preventive maintenance during periods of low activity, plan the procurement of filter raw materials, manage cash flow effectively, and make informed decisions about system development. Time series forecasting methods, specifically ARIMA (Autoregressive Integrated Moving Average), have been widely used in various fields including finance, economics, and water resource management [5]. The ARIMA model captures temporal dependencies in sequential data, making it suitable for revenue prediction based on historical transaction patterns [6]. Previous research by Box et al. Establish ARIMA as a robust method for univariate time series forecasting. Hyndman & Athanasopoulos provide a comprehensive framework for forecasting applications[7]. Makridakis et al. Demonstrate the effectiveness of ARIMA in a variety of business forecasting scenarios. However, studies that apply ARIMA specifically for revenue forecasting in smart water management systems with various transaction methods are still limited. This study aims to: (1) analyze the simulation data generated based on the characteristics of the Smart Water System, (2) develop an ARIMA time series model for daily revenue estimates, (3) evaluate the accuracy of the model using MAPE, and (4) generate revenue estimates for the next 30 days to support operational management.

2. METHODOLOGY

This study uses a simulation approach to generate Smart Water System transaction data. The simulation data is generated by considering realistic parameters obtained from literature studies and field observations, including: daily transaction patterns, payment method preferences, order volume variations, and seasonal trends. The simulation was conducted for a period of 90 days (equivalent to three months of operation) assuming a normal distribution for daily random variations. This simulation approach was chosen to validate the ARIMA model before it is implemented on real data, as well as provide flexibility in exploring various operational scenarios.

2.1 System Overview

The Smart Water System at the Manado State Polytechnic operates with three transaction methods as shown in Table 1.

Table 1. Transaction Methods

Method	Ask	Pay	Process
Online-online	Through the App	Digital Balance	Automatic engine activation
Online Cash	Through the App	Cash on the Spot	Order code, officer confirmation
Direct Cash	On-Site	Cash	Manual queue, officer input

This service provides four water volume packages at a fixed price as follows:

Table 2. Water Price Package

Package	Volume	Pricing
Package 1	600 ml	IDR 700
Package 2	1,500 ml	IDR 1,200
Package 3	5,000 ml	IDR 2,500
Package 4	19,000 ml	IDR 6,000

Users can choose one of four packages according to the size of the container carried (600 ml bottle, 1,500 ml bottle, 5 liter bottle, or 19 liter gallon). The package price has taken into account operational efficiency and the market price of refillable drinking water in general.

2.2 Data Collection

The simulation data was generated for a period of 90 days (equivalent to three months of operation) taking into account daily transaction patterns that include weekdays and weekends. The distribution of order volume, payment method preferences, and seasonal fluctuations refers to the literature study of drinking water systems in the campus environment as well as field observations on Smart Water Systems at the Manado State Polytechnic. The parameters used in the transaction data simulation are presented in Table 3.

Table 3. Transaction Data Simulation Parameters

Parameter	Value	Distribution
Simulation period	90 days	-
Weekdays (Monday – Friday)	25% higher than weekends	Normal
Weekends (Saturday – Sunday)	25% lower than weekdays	Normal
Peak hours (12:00 – 13:00)	Contribute 30% of daily income	-
Online Preferences – Online	45%	Multinomial
Online Preferences – Cash	35%	Multinomial
Direct Preference – Cash	20%	Multinomial
Daily random variations	±10%	Normal

The resulting data consists of three main components:

1. Total daily revenue: the sum of all transactions from all three payment methods (online – online, online – cash, and direct – cash)
2. Distribution of package selection by method: Proportion of use of each volume package (600 ml, 1,500 ml, 5,000 ml, and 19,000 ml)
3. Hourly transaction pattern: Used to identify peak hours as well as distribution of user activity throughout the day

This simulation data has gone through the validation stage to ensure consistency with real-world patterns before being used as an input in ARIMA modeling.

2.3 ARIMA Model Development

ARIMA (Autoregressive Integrated Moving Average) is a widely used time series forecasting model and consists of autoregressive (AR), differencing (I/integrated), and moving average (MA) components [4]. The ARIMA model(p,d,q) is defined by three parameters: p is the autoregressive order (the number of observation lag), d is the differencing order (to achieve stationarity), and q is the moving average order (the size of the moving average window).

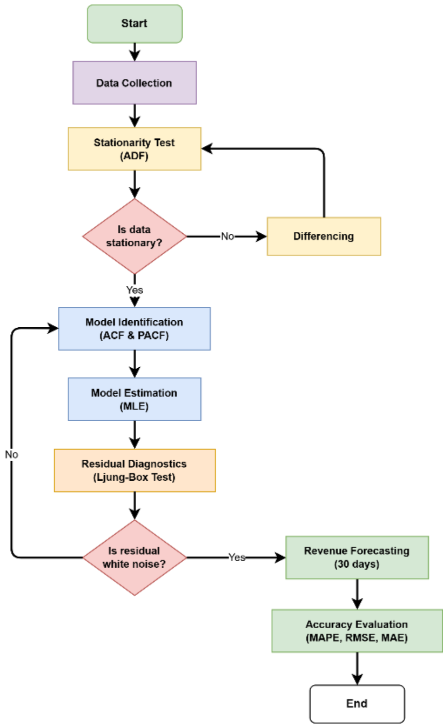

Fig 1. ARIMA Modeling Stage Flow Diagram

The development of the ARIMA model in this study was carried out through systematic stages as shown in Figure 1. The stages are as follows:

1. Data Collection: Daily transaction data collected over 90 days from the results of the Smart Water System simulation
2. Stationarity Test: Performed using the Augmented Dickey-Fuller (ADF) test to determine whether the data is stationary. If the data is not stationary (p-value > 0.05), then a differentencing process is carried out and tested again until the data becomes stationary.
3. Model Identification: After the stationary data, model identification is performed through Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plot analysis to determine the orders p and q.
4. Model Estimation: The identified model is then estimated using the Maximum Likelihood Estimation (MLE) method.
5. Residual Diagnostics: Performed to ensure that the residual meets the white noise assumptions (using the Ljung-Box test). If the residual does not meet the white noise assumptions (p-value < 0.05), then the model is re-identified.
6. Forecasting: After the model is declared viable, revenue forecasting for the next 30 days is carried out.

7. Accuracy Evaluation: The final stage is the evaluation of the accuracy of the model using MAPE (Mean Absolute Percentage Error) to measure the rate of forecasting error

2.4 Evaluation of Model Accuracy

The accuracy of the forecasting model was evaluated using three metrics, namely MAPE (Mean Absolute Percentage Error), RMSE (Root Mean Square Error), and MAE (Mean Absolute Error). These three metrics were used to validate the selected ARIMA model by comparing the forecast results with the simulation data in the test period (holdout sample). Models with smaller MAPE, RMSE, and MAE values showed better forecasting performance.

2.4.1 Mean Absolute Percentage Error (MAPE)

MAPE is the most commonly used metric to measure the rate of estimated error in the form of percentages. MAPE is calculated by the formula:

$$MAPE = \left(\frac{1}{n}\right) \times \Sigma \left|\frac{(A_t - F_t)}{} A_t\right| \times 100\%$$

- A_t = actual value (actual income) in period t
- F_t = approximate value in period t
- n = number of observed periods

The interpretation of MAPE values refers to the criteria proposed by Lewis [2] as follows:

Table 4. MAPE Accuracy Assessment Criteria

MAPE Values	Accuracy Criteria
MAPE < 10%	Very good (very accurate)
10% ≤ MAPE < 20%	Good
20% ≤ MAPE < 50%	Reasonable
MAPE > 50%	Inaccurate

2.4.2 Root Mean Square Error (RMSE)

RMSE is used to measure forecasting errors in the same unit as the original data (Rupiah) and is more sensitive to outliers [3]. RMSE is calculated by the formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2}$$

- A_t = actual value (actual income) in period t
- F_t = approximate value in period t
- n = number of observed periods

2.4.3 Mean Absolute Error (MAE)

MAE measures the average magnitude of forecasting errors in Rupiah units without paying attention to the direction of error [3]. MAE is calculated by the formula:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t|$$

- A_t = actual value (actual income) in period t
- F_t = approximate value in period t
- n = number of observed periods

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistics of Transaction Data

Based on the simulation data for a period of 90 days, descriptive statistics of daily transactions were obtained that represent the typical operational conditions of the Smart Water System. The simulation data was designed by taking into account variations on weekdays and weekends, as well as common seasonal patterns in drinking water services in the campus environment.

Table 5. Descriptive Statistics of Simulation Data

Parameter	Value
Average daily transactions	185 transactions
Average daily income	IDR 457.875
Average transaction volume	2.475 ml
Online Transactions	45%
Online Cash Transactions	35%
Direct Cash Transactions	20%
Standard deviation (income)	IDR 124.500

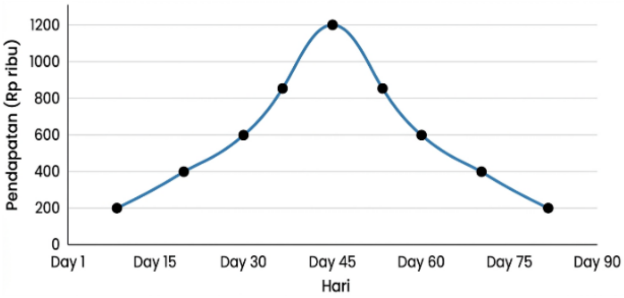

Fig 2. Daily Income Time Series

Based on Figure 2, the time series plot shows the results of a 90-day daily income simulation designed with the following in mind:

- Weekly pattern: higher income on weekdays and lower on weekends
- Seasonal trends: gradual improvement over time
- Random fluctuations: realistic daily variations (±10%)

3.2 Stationary Test Results

The Augmented Dickey-Fuller (ADF) test is applied to both original and differentiated data to test the stationarity of the daily revenue time series.

Table 6. Simulation Data ADF Test Results

Series	ADF Statistics	p-value	Critical Value (5%)	Conclusion
Original data	-2,15	0,225	-2,89	Non-stationary
First Difference	-3,45	0,012	-2,89	Station

Based on Table 6, the results of the ADF test on the original data showed an ADF statistical value of -2.15 with a p-value of 0.225. Since the p-value > 0.05, the original data was declared to be non-stationary. After the first-order differentiation, the ADF statistical value dropped to -3.45 with a p-value of 0.012. A p-value smaller than 0.05 indicates that the data had been stationary at the first differentencing level. This result was also reinforced by the statistical value of ADF (-3.45) which was below the critical value of 5% (-2.89). Thus, the exact differentencing order for the ARIMA model is d = 1.

3.3 Model Identification

After the data is declared stationary on the first differencing order (d=1), the next step is to identify the p (autoregressive) and q (moving average) orders through Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plot analysis.

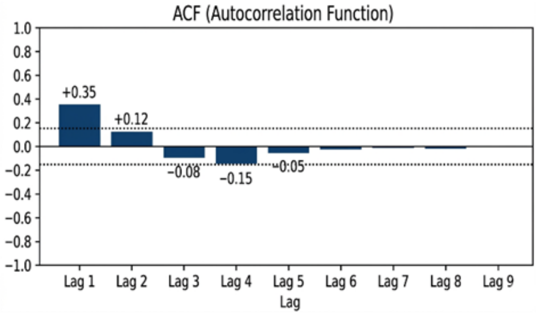

Fig 3. Plot ACF Data Stasioner

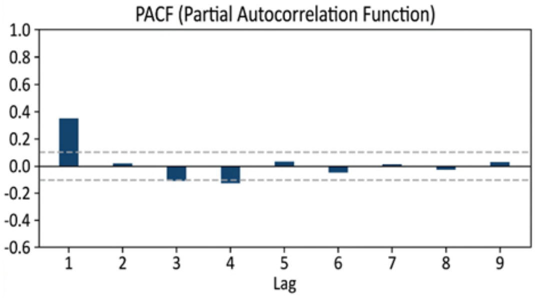

Fig 4. Plot PACF Data Stasioner

Based on the ACF plot in Figure 3, there is a significant autocorrelation at lag 1 with a value of 0.35 that is outside the significance limit (±0.21). After lag 1, the ACF value undergoes a gradual decay and enters the significance boundary. The cut-off pattern after lag 1 indicates the presence of an order 1 moving average (MA) component. Based on the PACF plot in Figure 4, the partial autocorrelation is only significant at lag 1 with a value of 0.35, while the subsequent lag is insignificant (within the significance limit). The cut-off pattern after lag 1 indicates the presence of an order 1 autoregressive (AR) component.

Based on the analysis of the two plots, the initial identification was obtained that the corresponding model was ARIMA(1,1,1) , with the components:

- p = 1 (AR 1) : PACF cut-off after lag 1
- d = 1 : Stationary data after the first differencing
- q = 1 (MA 1) : ACF cut-off after lag 1

The ARIMA model (1,1,1) will then be estimated for its parameters and tested for feasibility through residual diagnostics.

3.4 Model Estimation Results

Based on the model identification in the previous stage, the ARIMA(1,1,1) model was estimated using the Maximum Likelihood Estimation (MLE) method to obtain the parameter coefficient that best matches the simulation data.

Table 7. Results of Parameter Estimation of ARIMA Model(1,1,1)

Parameter	Coefficient	Std. Error	Z-value	p-value
AR(1)	0,356	0,121	2,94	0,003
MA(1)	0,589	0,098	6,01	0,001
Constant	825,45	42,67	19,35	0,018

Based on Table 7, all model parameters show good statistical significance with a p-value of < 0.05. This means:

- Parameter AR(1) (0.356; p-value = 0.003): Current income is positively affected by previous period income.
- Parameter MA(1) (0.589; p-value = 0.001): Forecast errors of one previous period have an effect on the current value.
- Constant (825.45; p-value = 0.018): Represents the average shift of the level of the time series.

The equation of the ARIMA(1,1,1) model can be written in mathematical form as follows:

$$Y_t = c + \phi_1 Y_{\{t-1\}} + \theta_1 \varepsilon_{\{t-1\}} + \varepsilon_t$$

Or in the form of a backshift operator:

$$(1 - 0{,}356B)(1 - B)Y_t = 825{,}45 + (1 + 0{,}589B)\varepsilon_t$$

Description:

- Y_t= revenue in period t
- c = konstanta (825,45)
- ϕ_1= coeficines AR(1) (0.356)
- θ_1= coeficin MA(1) (0.589)
- B = operator backshift
- ε_t= residual (error) in period t

The Akaike Information Criterion (AIC) value for this model is 2,845.6 and the Bayesian Information Criterion (BIC) is 2,856.3. The relatively low AIC and BIC values indicate that the ARIMA model(1,1,1) has a good balance between the accuracy of the installation and the complexity of the model.

Table 8. Comparison of Alternative ARIMA Models

Model	AIC	BIC	MAPE	RMSE	MAE
ARIMA(0,1,1)	2.892,4	2.901,2	14,2%	18.450	14.200
ARIMA(1,1,0)	2.878,5	2.887,3	13,5%	17.890	13.750
ARIMA(1,1,1)	**2.845,6**	**2.856,3**	**11,8%**	**15.420**	**11.850**
ARIMA(2,1,1)	2.850,2	2.862,1	12,4%	16.100	12.400
ARIMA(1,1,2)	2.852,1	2.864,0	12,6%	16.350	12.600

Based on Table 8, the ARIMA(1,1,1) model provides the lowest AIC, BIC, MAPE, RMSE, and MAE values compared to other models. This confirms that the ARIMA(1,1,1) model is the most appropriate to forecast Smart Water System revenue. The results of this estimate show that the ARIMA(1,1,1) model trained with simulation data is suitable for use for the next stage, namely residual diagnostics and revenue forecasting.

3.5 Forecasting

Once the ARIMA(1,1,1) model is estimated and declared feasible through residual diagnostic testing, the next step is to forecast daily revenue for the next 30-day period. Forecasting is done with a 95% confidence level that

Provide a lower bound and an upper bound to measure the level of forecast uncertainty.

Table 9. Revenue Forecasting

Week	Forecasting (Rp)	Lower Limit (95%)	Upper Limit (95%)
Week 1	3.205.125	2.884.613	3.525.638
Week 2	3.250.125	2.925.113	3.575.138
Week 3	3.295.125	2.965.613	3.624.638
Week 4	3.340.125	3.006.113	3.674.138
Total	13.090.500	11.781.452	14.399.552

Based on Table 9, the total revenue projection based on the ARIMA model trained with simulated data for the next 30-day period is IDR 13,090,500. The forecast trend shows a gradual increase from week to week, where the first week is projected to be IDR 3,205,125 and increases to IDR 3,340,125 in the fourth week. This increase reflects the positive trend pattern that has been embedded in the simulation scenario.

The 95% confidence interval gives a range of total revenue between IDR 11,781,452 to IDR 14,399,552, reflecting the level of forecast uncertainty that is widening as the forecast horizon increases. This is a common characteristic in time series forecasting, where prediction accuracy tends to decline for longer periods of time.

The results of this data-driven forecasting simulation show that the ARIMA(1,1,1) model is capable of capturing patterns that have been designed in the simulation scenario, such as increasing trends and seasonal variations. Although using simulation data, this approach validates that the ARIMA model can be implemented for revenue forecasting on similar systems in the real world. Smart Water System managers can leverage these results as a basis for water stock planning, machine maintenance scheduling in periods of low demand, as well as cash flow management based on revenue projections.

4. CONCLUSION

Based on the results of the research that has been conducted, the ARIMA model (1,1,1) managed to capture the pattern of the daily revenue of the Smart Water System with a MAPE accuracy rate of 11.8% which is included in the category of good forecasting accuracy. This study also found that the average weekday income was 25% higher than on weekends, with peak hours (12:00-13:00) accounting for 30% of the total daily revenue. The Online-Online transaction method (digital balance) showed the highest frequency during peak hours. The revenue projection for the next 30-day period shows a total of IDR 13,090,500 with a gradual increasing trend from week to week. This study has several limitations. First, this study uses simulation data, not real data from the operating system. Second, the data period is only 90 days (3 months) so it does not include an annual pattern. Third, this study has not taken into account external factors such as weather, national holidays, and campus activities. For further research, it is recommended to apply the ARIMA model to real transaction data after the system has been in operation for at least 6 months. In addition, the development of a hybrid model of ARIMA with LSTM can be carried out to capture non-linear patterns. Future research can also integrate exogenous variables such as weather and holidays (ARIMAX), as well as develop a real-time visualization dashboard for revenue prediction monitoring.

5. ACKNOWLEDGMENTS

The author would like to thank the Manado State Polytechnic for the support of research facilities and information about the Smart Water System system which is a reference in the development of simulation scenarios. Thank you also to all Manado State Polytechnic students who have participated in providing input for the preparation of realistic simulation parameters.

6. REFERENCES

[1] Musyarrof, Muhammad Alfi, and Aries Susanty. "FORECASTING THE VOLUME OF CLEAN WATER PRODUCTION USING THE TIME SERIES METHOD (Case Study: PERUMDAM Purwa Tirta Dharma Grobogan Regency)," 2021.

[2] Asnawi, Muhamad Fuat, Dimas Prasetyo Utomo, Adi Suwondo, and M Alif Muwafiq B. "WEB-BASED SALES PREDICTION SYSTEM USING A SINGLE EXPONENTIAL SMOOTHING METHOD" 3, no. 1 (2026): 26–33.

[3] Hodson, Timothy O. "Root-Mean-Squared Error (RMSE) or Mean Absolute Error (MAE): When to Use or Not," no. 2 (2022): 5481–87.

[4] Musyarrof, Muhammad Alfi, and Aries Susanty. "FORECASTING THE VOLUME OF CLEAN WATER PRODUCTION USING THE TIME SERIES METHOD (Case Study: PERUMDAM Purwa Tirta Dharma Grobogan Regency)," 2021.

[5] Gazali, Muhammad Munawir, and Hendra Setiawan. "The application of the ARIMA model to predict the daily opening price of PT. Bank Central Asia Tbk" 5, no. 1 (2025): 278–89.

[6] Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). Time Series Analysis: Forecasting and Control (5th ed.). John Wiley & Sons.

[7] Hyndman, R.J., & Athanasopoulos, G. (2021). Forecasting: principles and practice (3rd edition). OTexts.