Arowana cultivation water quality monitoring and prediction using autoregressive integrated moving average

April Firman Daru, Susanto, Whisnumurti Adhiwibowo
Department of Informatics Engineering, Faculty of Information Technology and Communication, Universitas Semarang, Semarang, Indonesia

ABSTRACT

Decorative fish is a fish that humans keep for amusement. There are many decorative fish that exist in this world, one of them is known as the Arowana fish (Scleropages Formosus). This fish is known around Asia including in Indonesia. However, to ensure the Arowana is living well is not easy. The water quality inside a farm must follow a strict balance. The pH of the water must not exceed or below 7 pH. Meanwhile, the total dissolved solid (TDS) salt must not exceed 1000 parts per million. If the balance collapsed, the Arowana fish will not grow. Thus, the owner must monitor the water to make sure that the water is ideal. There were many approaches including internet of things (IoT) solutions. However, they have weaknesses with prediction. Because of this reason, this study designed pH and TDS monitoring with autoregressive integrated moving average (ARIMA) as the algorithm. To achieve the solution, this study used experiment methodology as the research fundamental from top to bottom. According to the evaluation, this study found that the accuracy of ARIMA model is 98.12% for pH and 98.86% for TDS. On the contrary, the seasonal autoregressive integrated moving average (SARIMA) model has an accuracy of 98.52% for pH and 99.89% for TDS.

This is an open access article under the CC BY-SA license.

Corresponding Author:
April Firman Daru
Department of Informatics Engineering, Faculty of Information Technology and Communication
Universitas Semarang
Soekarno Hatta street, Tlogosari Kulon, Pedurungan, Semarang City, Central Java, Indonesia
Email: firman@usm.ac.id

1. INTRODUCTION

The Arowana, scientifically known as Scleropages Formosus, serves as a highly sought-after ornamental fish, primarily found in the Asian region [1]. The appearance of their scales exhibits variations across different countries. For instance, the Indonesian Arowana displays silverish scales [2]. Despite these regional distinctions, the care and maintenance requirements for this species remain largely consistent. Arowana fish necessitate a specific level of water acidity and total dissolved salt. Additionally, Arowana are classified as freshwater fish, and thus, the total dissolved solid (TDS) concentration in the water should not exceed 1,000 parts per million. Neglecting water quality can result in severe issues for these fish, potentially leading to fatality [3], [4].

For this reason, it is imperative that Arowana fish owners or collectors exercise vigilance when monitoring water quality. When the levels reach a critical threshold, appropriate treatment becomes necessary. Typically, owners assess water quality manually employing specialized sensors. The ensuing figure demonstrates the process by which owners measure pH and the concentration of TDS salt within the water.
Figure 1 illustrates the process of water quality assessment employing a sensor. To determine the levels of acidity and TDS in the water, the owner places the sensor in contact with the water and awaits several moments. The measurement results typically appear on the display. Throughout the water quality assessment, the owner is present and observes the test's progression. However, this scenario can become problematic when the owner manages numerous aquariums or is absent from the fish farm. Even delegating this task to hired workers does not eliminate the possibility of forgetting the scheduled water quality assessment or called with human error [5]. In instances of oversight, a substantial number of fish may perish, resulting in significant losses for the owner. Given the relatively high value of these mature fish, the risk of financial loss becomes even more pronounced.

To mitigate the mortality of Arowana fish, various methods are at the disposal of the owner. The simplest approach involves scheduling an alarm to prompt the owner to conduct water quality checks. While this solution is straightforward to implement, it is equally prone to forgetfulness. When the notification device is either powered off or situated far from the owner, the notification can easily go unnoticed. The second available method involves the implementation of an internet of things (IoT)-based system for automated detection and notification [6]–[8]. This approach is cost-effective and highly recommended [6], as it eliminates the need for the owner's physical presence during water quality assessments.

Several studies contribute to monitoring Arowana's water quality with IoT-based technology. Many models use different approaches, such as low-cost models or predictions. The model was proposed in 2019, using green electricity to monitor the temperature of water and control fish feeding [9]. In the same year, the model was improved by adding pH and electric conductivity monitoring [10], [11]. In 2020, many monitoring models with different approaches exist. One model adds an ultrasonic sensor to monitor water volume changes. Thus, the owner can monitor easily via smartphone or computer [12]. In 2021, the improvement of the model continues. The proposed model adds new features such as water turbidity monitoring, feeding, and water level control to maintain water quality [13]. The next year (2022), the study develops water quality monitoring with a partnership for real implementation. Thus, the accuracy for the monitoring is high [14]. The most recent study that focuses on pH monitoring produces a model that is capable of detecting acidity within a 100% range and validates it with litmus paper for better accuracy [15].

Based on the aforementioned models, they exhibit common limitations. Most models primarily concentrate on either temperature or pH levels, consequently restricting the monitoring of salinity to electric conductivity. Another noteworthy concern pertains to their predictive capabilities, or rather, the lack thereof. Since none of these models are endowed with a prediction algorithm, their predictive capacity remains negligible. To address these issues, the primary objective of this study is to design an IoT-based model for monitoring pH and TDS salt levels. This model will be equipped with a prediction algorithm, such as autoregressive integrated moving average (ARIMA) and its seasonal autoregressive integrated moving average (SARIMA), enabling it to predict both parameters. Compared to other algorithms, ARIMA algorithm is the most suitable model to predict timeseries data in several sectors [16]–[19].

2. RESEARCH METHOD

In explicating the methodology, this meticulously describe the intricate process employed to architect the proposed model. Including the formulation intricacies, the intricacies of the communication
2.1. The proposed model design

In pursuit of a meaningful contribution to problem-solving, this study intricately crafted the proposed model. The design and implementation of the proposed model stand as an endeavour to provide comprehensive solutions to the identified problem, further reinforcing the study's commitment to impactful outcomes. The proposed model consists of several parts like sensors and processing board, as depicted in Figure 2.

Figure 2 represents the proposed model's design that consists of several parts such as ESP32 as the main board that has the job of collecting water quality data from the sensors. Using ESP32 in this situation is recommended due to the availability of an Analogue Pin [20]. Besides ESP32, the model is also equipped with PH-4502C (pH sensor) and a TDS sensor that is connected via analogue pins. The PH-4502C is a pH sensor that relies on a probe as a dipping part and a board to control the measurement. This kind of sensor is commonly used to monitor water acidity. Since this sensor is analogue, the board is equipped with a potentiometer to adjust the voltage output during the calibration process. The calibration process of this sensor is quite simple and only needs a pH meter and pH-specified liquid during the calibration process to ensure the result is similar. Besides that, the sensor used to detect the TDS salt is an electric conductivity-based reading to obtain salt content inside water [21]. The TDS sensor was also calibrated with the same process as the pH sensor but with different liquids and meters. All data stored to the online database through machine-to-machine communication through representational state transfer application programming interface (ReST API) [22]. The next part is the process flow of the model. This part is important to ensure that the model can monitor the water. The flow designed for this model is shown in Figure 3.

Figure 3. The proposed model's process flow
The process flow from Figure 3 starts with initializing the needed components such as wireless network connectivity and the sensors. After initializing all necessary components, the model reads the data from both pH and TDS sensors. The next step in data collection is the querying process to the online database. All data in the database is split into training for ARIMA modeling and test data for testing the trained ARIMA model.

The implementation of the ARIMA function in programming is straightforward, demanding no stringent parameters for its execution. The simplicity of the programming aspect belies the underlying complexity of the mathematical representation inherent in the ARIMA model, encapsulated succinctly in (1). This equation serves as the foundational expression guiding the model’s behaviour and predictive capabilities.

\[ Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} + \epsilon_t \]  

(1)

Where \( Y_t \) is the current value from timeseries at \( t \) time. Meanwhile, \( c \) is the constant. Variable \( \phi_1, \phi_2, \ldots, \phi_p \) are the coefficient for autoregressive. Variable \( \theta_1, \theta_2, \ldots, \theta_q \) are the coefficient for moving average. And the last variable in this equation is \( \epsilon_t \) contains the white noise error at \( t \) time.

Unlike ARIMA, the seasonal version has a different mathematical representation to provide seasonal behavior in time series dataset. The (2) represents the mathematical form of SARIMA.

\[ Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} + e_t - \Phi_1 Y_{t-s} - \Phi_2 Y_{t-2s} - \cdots - \Phi_p Y_{t-ps} + \Theta_1 e_{t-s} + \Theta_2 e_{t-2s} + \cdots + \Theta_q e_{t-qs} + \epsilon_t \]  

(2)

This equation has additional formula compared to ARIMA’s equation. In this case, \( Y_t \) is the value of the time series according to \( t \) time. The \( c \) variable is the constant for the equation. Variable \( \phi_1, \phi_2, \ldots, \phi_p \) are the coefficient for autoregressive. Variable \( \theta_1, \theta_2, \ldots, \theta_q \) are the coefficient for moving average. Variable \( s \) refers to the number of steps taken in each season. And the last is \( \epsilon_t \) contains the white noise error at \( t \) time.

2.2. The communication topology

The next design is the network topology used by the model to communicate with external services in the cloud system. In this case, the study used a real-time database from firebase Google as the provider of the service [23], [24]. Real-time database will store any data from the model in a key-pair shape for easier access and reading. The Figure 4 illustrates the topology used by the proposed model.

![Figure 4. Network topology setup](image)

Figure 4 illustrates the connectivity needs of the model. According to the figure, a wireless router with 2.4 GHz is needed to allow communication between the model and the router [25]. For internet access, this study uses existing wide-area network (WAN) connectivity available on the campus. This study uses IPv4 addressing to identify the proposed model within the wireless network. Since there is only one model, it is easier to assign the identity to the model [26].

While Figure 4 provides insight, it falls short of elucidating the data flow from the model to the database. In a bid to comprehensively expound on the service’s functionality, this study introduces an additional figure. This new illustration precisely delineates the intricate process by which the service retrieves data from the model and seamlessly uploads it to the cloud, offering a more nuanced understanding of the operational dynamics.

According to Figure 5, the external service has a task to receive the data from the model. In this case, the service receives both pH and TDS levels as the input. After receiving the input, the service processes the data using ReST API to the real-time database endpoint [27]. ReST API is a lightweight data
A communication protocol that enables communication between services or machine-to-service. If compared to another protocol like simple object access protocol (SOAP), ReST API is lighter and suitable for the IoT-based devices. The final data inside the database is the output of the service. This process repeats until the process from the model is interrupted by network timeout or power.

![Service operation](image)

**Figure 5. Service operation**

### 2.3. Evaluation scenario

The final step in this study will be to evaluate how the model will perform in real-world conditions. To assess its monitoring capabilities, this study will allow the model to run for more than 24 hours. The study will also configure the model to collect data at five-second intervals to ensure that it generates as much data as possible over the course of 24 hours. The testbed for the model will consist of a sample of fresh water obtained from a nearby river, known for its higher mineral content compared to drinking water. All the data generated in the database will then be processed using ARIMA (both normal and seasonal) to predict water quality.

### 3. RESULTS AND DISCUSSION

In this section, this study presented and discussed the outcomes of the proposed model. All detection results were stored in a table to facilitate the reader's understanding of the sample data. Subsequently, the results were complemented by graphical representations from ARIMA and SARIMA predictions to illustrate how the model predicted water acidity and salinity possibilities.

<table>
<thead>
<tr>
<th>Num</th>
<th>PH</th>
<th>TDS (ppt)</th>
<th>Salt (gr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.0276</td>
<td>4.4522</td>
<td>0.0178</td>
</tr>
<tr>
<td>2</td>
<td>7.0859</td>
<td>4.6271</td>
<td>0.0185</td>
</tr>
<tr>
<td>3</td>
<td>7.2785</td>
<td>4.5178</td>
<td>0.0181</td>
</tr>
<tr>
<td>4</td>
<td>7.2473</td>
<td>4.7802</td>
<td>0.0191</td>
</tr>
<tr>
<td>5</td>
<td>7.2134</td>
<td>4.7802</td>
<td>0.0191</td>
</tr>
<tr>
<td>6,335</td>
<td>7.3222</td>
<td>9.1977</td>
<td>0.0368</td>
</tr>
<tr>
<td>6,336</td>
<td>7.6646</td>
<td>9.0884</td>
<td>0.0364</td>
</tr>
<tr>
<td>6,337</td>
<td>7.9958</td>
<td>9.1321</td>
<td>0.0365</td>
</tr>
<tr>
<td>6,338</td>
<td>7.2731</td>
<td>9.1540</td>
<td>0.0366</td>
</tr>
<tr>
<td>6,339</td>
<td>7.8749</td>
<td>9.1103</td>
<td>0.0364</td>
</tr>
</tbody>
</table>

Table 1. Result's data sample

As shown in Table 1, the model successfully detected pH, TDS (in parts per thousand), and salt concentration within the water. Since the data obtained from the model contains more than 6,300 rows, this study decided to reduce the number of rows using time-based reduction. With this method, this study can reduce all rows into minute-based intervals. Thus, this study obtained less dense data that was easier to analyze and illustrate. Utilizing statistical formulas, the model calculated an average pH level of 7.4211. Similarly, for salinity levels, the model identified an average of 8.8212 ppt (TDS) and 0.0353 grams of salt. Consequently, over the course of 24 hours, the water quality was deemed suitable for sustaining Arowana fish. However, relying solely on the average values did not provide a comprehensive understanding of water quality fluctuations throughout the 24-hour period. The model's minimum readings included 7.0002 for...
acidity, 4.4522 ppt for TDS, and 0.0178 grams for salt content. In contrast, the maximum values recorded were 7.9997 for acidity, 9.3945 ppt for TDS, and 0.0376 grams for salt. Subsequently, the next set of results pertained to predictions made using the ARIMA model. The subsequent figures illustrate the outcomes of ARIMA predictions.

As depicted in Figure 6, this study observed unexpected prediction trends in relation to pH and TDS levels. Utilizing the ARIMA model, the pH prediction trend initially exhibited an increase before stabilizing. In contrast, the TDS prediction showed a decrease when compared to the actual data. Based on these findings, the accuracy of the ARIMA model was determined to be 98.12% for pH as shown in Figure 6(a) and 98.68% for TDS as shown in Figure 6(b). To substantiate the accuracy of these findings, additional validation was undertaken, employing mean square error (MSE) to estimate the algorithm's error percentage. The outcomes revealed a minimal 2.943% error in ARIMA’s pH predictions and a corresponding 1.779% error in TDS predictions. These notably low error percentages serve as validation, attesting to the elevated predictive accuracy of ARIMA in determining both pH and TDS levels.

The conclusive set of results provides a focused examination of the predictions produced by the SARIMA model. In delving into the intricacies of these predictions, the ensuing figures play a pivotal role in conveying a comprehensive visual representation of the obtained outcomes. Each figure within this set serves as a nuanced illustration, shedding light on the nuanced insights and patterns extracted through the application of the SARIMA methodology.

As illustrated in Figure 7, the results obtained from SARIMA exhibited distinct behavior when compared to the ARIMA model presented in Figure 6. Notably, both pH and TDS predictions closely resembled the original test data but exhibited a lag. This discrepancy can be attributed to the influence of seasonal patterns. Consequently, the accuracy of these predictions surpassed that of the regular ARIMA model. Specifically, the accuracy for pH prediction shown in Figure 7(a) reached 98.529%, while TDS prediction shown in Figure 7(b) achieved an accuracy rate of 99.890%. Following the determination of SARIMA accuracy, the validation proceeded by assessing MSE. The computed MSE revealed a mere 1.930% error in pH predictions and an exceptionally low 0.034% error in TDS predictions within the SARIMA framework. The consistently lower error percentages further affirm the commendable accuracy of SARIMA predictions.
Within this section, the discussion unfolds by delving into the results acquired earlier, encompassing two pivotal domains. The initial discussion centers on the efficacy of the proposed models, while the subsequent discourse delves into the analysis of predictions stemming from both the ARIMA and SARIMA models. This bifurcated approach aims to provide an exploration of the diverse facets encapsulated in the obtained results.

The first discussion centered on the monitoring capability of the proposed models, which incorporate IoT technology for water quality assessment. This study introduced a model equipped with two sensors to facilitate real-time monitoring of pH and TDS levels in water. Analysis of the data table revealed that the model effectively monitored pH and TDS levels throughout a 24-hour testing period. The average pH level recorded was 7.4, with a minimum value of 7.0 and a maximum of 7.9. Simultaneously, the average TDS level measured 8.8 ppt, with a minimum of 4.45 ppt and a maximum of 9.39 ppt. This initial discussion concludes that the proposed model is capable of real-time water quality monitoring.

The second discussion delved into the predictions generated by the ARIMA and SARIMA models. Both models demonstrated high prediction accuracy exceeding 98%. In terms of superiority, the SARIMA model exhibited slightly higher accuracy, with a 0.40% difference for pH and 1.20% for TDS. Additionally, each model exhibited distinct prediction trends that characterized their behavior. For instance, the logarithmic trend observed in pH and TDS levels was indicative of the ARIMA model, while SARIMA displayed a moving average trend for both pH and TDS. Consequently, SARIMA outperformed the ARIMA model due to its ability to provide more accurate water quality predictions. This second discussion concludes that the SARIMA model offers superior accuracy in water quality prediction.

4. CONCLUSION

Water quality monitoring emerged as a pivotal factor in ensuring the well-being of Arowana fish. The absence of a balanced pH and appropriate levels of dissolved salt in the water can pose a significant threat to the livelihood of Arowana. Consequently, owners were compelled to resort to manual water quality monitoring to ascertain pH and TDS levels. However, this conventional approach proved less effective when compared to the IoT approach. Numerous prior models, though proficient in monitoring water quality, were often hindered by limitations in their predictive capabilities. Hence, the primary objective of this study was to contribute to the field of water quality monitoring by proposing a model capable of predicting pH and TDS levels. Evaluation of the proposed model over a 24-hour testing course yielded promising results, with an average pH level of 7.4 and a TDS level of 8.8 ppt. Additionally, the ARIMA model achieved prediction accuracies of 98.12% for pH and 98.86% for TDS. Conversely, the SARIMA models outperformed, attaining prediction accuracies of 98.52% for pH and an impressive 99.89% for TDS. The validation process for both algorithms involved assessing MSE, yielding results consistent with anticipated values. For ARIMA, the prediction error rates for pH and TDS stand at 2.943% and 1.779%, respectively. Notably, the introduction of SARIMA further refines accuracy, resulting in reduced error percentages of 1.930% for pH prediction and a minimal 0.034% for TDS prediction. Both models demonstrated exceptional accuracy in predicting water quality, with the SARIMA model exhibiting the highest level of precision compared to the ARIMA model. In conclusion, it is evident that the proposed model effectively monitored and predicted water quality, specifically concerning pH and TDS in the water.

ACKNOWLEDGEMENTS

We extend our sincere gratitude to the Research and Community Service Organization of Semarang University for their invaluable financial support and guidance, which played a crucial role in enabling us to successfully complete this study.

REFERENCES

BIOGRAPHIES OF AUTHORS

April Firman Daru earned his doctorate in computer science from Universitas Kristen Satya Wacana (UKSW) Salatiga, Indonesia. Serving as a lecturer since 2010, he contributes his expertise to the Department of Informatics Engineering at Universitas Semarang (USM). His research endeavors focus on diverse areas, with particular interests lying in internet of things, artificial intelligence, attack detection, and machine learning. His academic journey and professional commitment underscore his dedication to advancing knowledge in these dynamic and evolving fields. He can be contacted at email: firman@usm.ac.id.
Susanto is a Lecturer within the Department of Informatics Engineering at Universitas Semarang (USM), he specializes in instructing fundamental information technology, computer animation, and software quality assurance. With a keen interest in research, he delves into the realms of data mining, web development, and software quality assurance. His commitment to educational excellence is reflected in his targeted teaching areas, ensuring a comprehensive understanding of core information technology concepts. Additionally, his research pursuits contribute valuable insights to the dynamic fields he explores, emphasizing a holistic approach to both teaching, and scholarly endeavors. He can be contacted at email: susanto@usm.ac.id.

Whisnumurti Adhiwibowo currently pursuing a doctorate in data science at Universitas Diponegoro, Semarang, Indonesia, he has been actively engaged as a lecturer in the Department of Informatics Engineering at Universitas Semarang (USM) since 2010. Within the academic realm, his focus encompasses a diverse array of subjects, contributing to the comprehensive education of students. He’s research endeavors delve into the dynamic realms of internet of things, network, and blockchain, reflecting a commitment to advancing knowledge in these evolving fields. His dual roles as a dedicated student and educator underscore a passion for both learning and imparting valuable insights within the academic domain. He can be contacted at email: whisnu@usm.ac.id.