Predicting Water Levels from Environmental Parameters Using Random Forest Models

Arum Putri Kusuma Anggraini
Sensing Technology Study Program, Faculty of Defense Science and Technology, Indonesia Defense University, Jakarta, Indonesia

Trismadi
Sensing Technology Study Program, Faculty of Defense Science and Technology, Indonesia Defense University, Jakarta, Indonesia

Asep Adang Supriyadi
Sensing Technology Study Program, Faculty of Defense Science and Technology, Indonesia Defense University, Jakarta, Indonesia

Abstract: Real-time monitoring of sea water levels is essential for maritime safety, coastal management, and disaster mitigation. This study addresses the challenges of sensor dependency and environmental vulnerability in traditional monitoring systems by proposing a machine-learning-based soft sensor. A Random Forest model was developed to predict sea water levels using atmospheric parameters such as barometric pressure, temperature, and relative humidity, leveraging data collected over seven months at one-minute intervals from a Marine Automatic Weather Station (AWS) in Tanjung Priok, Indonesia. Data preprocessing included outlier removal, normalization, and temporal feature extraction. The model achieved a high correlation coefficient (R = 0.8415) and low error metrics (MSE = 0.0209, RMSE = 0.1448), demonstrating robust predictive performance. The findings confirm the model's ability to capture tidal patterns and its potential to complement or replace physical sensors in harsh maritime environments. This research contributes to the field by improving monitoring resilience and reducing dependency on hardware sensors. Future work will explore integrating additional environmental variables, temporal modeling techniques, and hybrid approaches to further enhance prediction accuracy and robustness.

Keywords: Soft sensor, Water level predictions, Random Forest, Atmospheric data, Machine learning.

Correspondents Author:
Arum Putri Kusuma Anggraini, Sensing Technology Study Program, Faculty of Defense Science and Technology, Indonesia Defense University, Jakarta, Indonesia
Email : arumputrikusuma@gmail.com

Received Oktober 20, 2024; Revised December 9, 2024; Accepted December 11, 2024; Published December 12, 2024

Introduction

As a country located in a disaster-prone area, such as earthquakes and tsunamis, real-time water level monitoring in Indonesia is one of the crucial aspects that need to be considered regarding availability and accuracy. Accurate tidal or water level data is vital in various sectors, from tsunami disaster mitigation to commercial maritime transportation, fishing activities, and military operations (Gultom et al., 2017; Malik et al., 2023; Sagala et al., 2021). The results of sea level monitoring are usually tidal forecasts, which are used to plan fishing activities, design and maintenance management of coastal infrastructure, and ensure shipping safety (Anwar et al., 2023;Djunarsjah et al., 2023; Martin Miguez et al., 2012; Pugh, 1987).

While the Geospatial Information Agency (BIG) primarily monitors sea level data via three sensor types: radar, bubbler, and submerged pressure sensors(Illigner et al., 2016), Other agencies, including BRIN with the IDSL system and BMKG with Marine AWS, frequently function with a single radar or ultrasonic sensor(A. A. Putra et al., 2021). This single-sensor dependency presents vulnerabilities since radar and ultrasonic sensors are primarily vulnerable to malfunction caused by corrosion, vandalism, and harsh environmental conditions (M. Putra, 2021; Sandro et al., 2020). Such failures highlight the need for alternative or complementary approaches to ensure continuous monitoring.

Nevertheless, advancement and gaps persist in the incorporation of atmosphere data for forecasting seawater levels. Current research frequently neglects the capacity of machine learning models to improve monitoring resilience and diminish hardware reliance. The inverse barometer effect illustrates the correlation between air pressure and sea level fluctuations; nevertheless, this link is inadequately applied to predictive modelling (Han et al., 2024; Li et al., 2016).

This study aims to address these issues by creating a machine-learning-based soft sensor utilising a Random Forest model. This work presents a robust and cost-effective solution for improving seawater level predictions by incorporating atmospheric factors, including barometric pressure, temperature, and relative humidity. This research can reduce the need for hardware sensors susceptible to environmental deterioration, utilise existing AWS infrastructure for cost-effective deployment, and illustrate the effectiveness of machine learning in maritime contexts. This study establishes a basis for subsequent research to investigate hybrid models and further environmental variables to enhance accuracy and reliability.

Literature Review

Monitoring sea levels is essential for marine safety, disaster mitigation, and coastal management. Conventional monitoring systems, exemplified by those utilised by the Geospatial Information Agency (BIG) in Indonesia, depend on many sensors, including radar, bubbler, and submerged pressure sensors (Illigner et al., 2016). These systems are engineered to deliver precise and continuous data; yet, they are frequently vulnerable to problems such as sensor corrosion, vandalism, and environmental deterioration, which might impede data availability (M. Putra, 2021; Sandro et al., 2020). Conversely, organisations such as BMKG and BRIN depend on single-sensor systems, such radar or ultrasonic sensors, which are more susceptible to malfunctions, underscoring the necessity for redundancy and complementary approaches to guarantee data reliability (Fatkhurrohman et al., 2023). Indonesia is very vulnerable to tectonic activity that can cause earthquakes and tsunamis (Matondang et al., 2020; Novianto et al., 2021), so most the sea water level measurement related to the early warning system.

Machine learning has garnered considerable attention in environmental monitoring owing to its capacity to analyse complicated and high-dimensional datasets. Random Forest models have exhibited exceptional efficacy in predicting tasks, particularly water-level forecasting. Research indicates that Random Forest models may efficiently incorporate diverse environmental variables to forecast results with considerable precision (Li et al., 2016). For example, Random Forest models have been utilised to estimate lake water levels, demonstrating superior predictive accuracy relative to conventional statistical techniques (Han et al., 2024). These findings highlight machine learning models' capability to overcome traditional monitoring systems' constraints by utilising current data streams without requiring further hardware.

The correlation between climatic characteristics and sea level fluctuations has been thoroughly examined, with barometric pressure, temperature, and relative humidity identified as key predictors. The inverse barometer effect elucidates the impact of atmospheric pressure on sea levels, whereby elevated air pressure depresses the sea surface, while diminished pressure results in its elevation (Pugh, 1987; Djunarsjah et al., 2023). Furthermore, temperature influences water density and thermal expansion, whereas humidity indirectly affects precipitation and local water levels (Martin Miguez et al., 2012). Notwithstanding these known links, the incorporation of such characteristics into prediction models for sea level monitoring remains inadequately investigated, especially in marine areas susceptible to intricate environmental dynamics such as Indonesia.

Although prior research has established the viability of employing machine learning algorithms for environmental monitoring, deficiencies persist in its utilisation for real-time sea level forecasting. Numerous studies concentrate on particular habitats, such as lakes or confined water bodies, but there is a paucity of research on wide coastal and maritime regions (Han et al., 2024; Li et al., 2016). Moreover, the majority of current systems depend significantly on physical sensors, which are susceptible to deterioration and malfunction in adverse conditions. This reliance highlights the necessity for the development of soft sensors to augment and strengthen the resilience of current infrastructure (Lu & Hardin, 2021).

This study mitigates these deficiencies by employing a Random Forest model to create a soft sensor for forecasting sea water levels based on atmospheric data. This solution lowers dependence on hardware sensors and leverages existing AWS infrastructure, rendering it both economical and scalable. This research enhances predictive modelling in maritime applications by incorporating barometric pressure, temperature, and relative humidity, providing a reliable solution for continuous sea level monitoring in dynamic and demanding conditions. Future study directions involve integrating supplementary elements, including wind speed and historical water level data, and investigating hybrid modelling methodologies to improve prediction accuracy.

Research Method

There are several processes of the research that have been illustrated in Figure 1. The flowchart represents a systematic methodology for predicting sea water levels using atmospheric data and a Random Forest model.

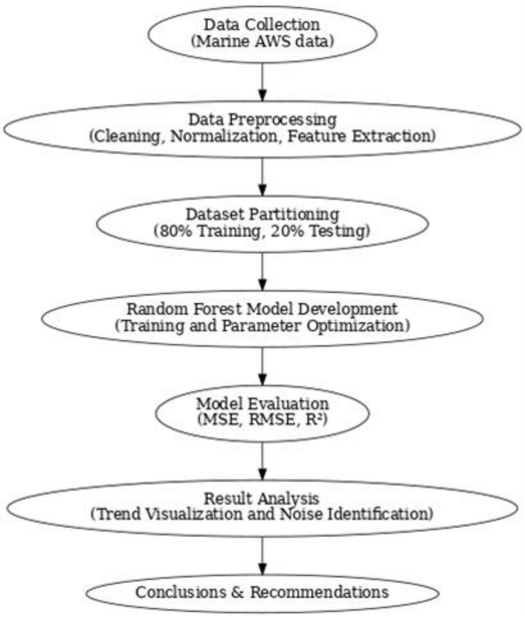

Figure 1. Flowchart of research methodology

The process begins with data collection from the Marine Automatic Weather Station (MAWS) managed by BMKG, which is stationed in Tanjung Priok, close to the Naval Harbour of Pondok Dayung. MAWS provided real-time weather data with a sampling rate of 1 minute (Sugiarto et al., 2020). The AWS functions continuously, utilising a solar power system, and is fitted with a CR1000 X data logger. The sensors comprise the Vegapuls C23 for seawater level measurement, the Vaisala HMP155 for temperature and relative humidity, the RM Young 61302 for barometric pressure, and the SPlite 2 for sun radiation measurement. The dataset spans 7 months, from January to July 2024, with a sample frequency of 1 minute, providing a total of 525,600 records.

Several preprocessing processes are conducted to prepare the data for analysis. Data cleansing includes the management of missing values and anomalies. Missing values can be mitigated via median or mean imputation, and outliers are handled through Z-score filtering. Critical features for the model, such as barometric pressure, temperature, and relative humidity, are chosen, with sea water level designated as the objective variable. The characteristics are normalised by Min-Max scaling to ensure equal contribution to the model training process. Furthermore, temporal attributes, including the hour and day of the week, are extracted from the timestamp data to identify temporal patterns.

The research project focused on developing a Random Forest model to forecast seawater levels. The dataset is partitioned into training and testing sets utilising an 80-20 split ratio, allocating 80% for training and 20% for testing. The Random Forest model is trained on the training set, with hyperparameters optimised by grid search and cross-validation to ascertain the ideal number of trees and the maximum depth of each tree. The significance of features is evaluated to comprehend each feature's contribution to seawater level forecasting. Many indicators assess the model's performance (Chicco et al., 2021). The Mean Absolute Error (MAE) quantifies the average magnitude of discrepancies between predicted and actual data. It represents the mean absolute deviation between the predicted and actual seawater levels. A cohesive approach for estimating prediction error in Random Forests can integrate Mean Absolute Error (MAE) to assess prediction uncertainty and error distribution, improving model reliability (Lu & Hardin, 2021).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{1}$$

The Mean Absolute Error (MAE) can be quantified by initially determining the absolute difference between the actual value (y_i) and the predicted value $(\hat{y}_i)$ For each data point i. The difference is aggregated across all data points and subsequently averaged by dividing by the total number of data points n in the test dataset. MAE indicates the mean of the absolute differences between actual observations and model predictions, quantifying the deviation of predictions from actual values.

The Root Mean Squared Error (RMSE) quantifies the square root of the mean of the squared differences between predicted and observed values. It indicates the mean magnitude of the predicted mistakes. Nonetheless, it penalises more substantial errors more severely than minor ones due to the squaring process, and it is extensively utilised in weather-related predictive analyses (Taqiyuddin & Sasongko, 2024). RMSE explains the average size of prediction mistakes and is susceptible to outliers.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{\sum_{i=1}^{n} |y_i - \bar{y}_i|} \quad (2)$$

The R-squared (R^2) metric represents the proportion of variance in the sea water level that is explainable by the features used in the model. It indicates the model's goodness of fit, showing how well its predictions match the actual data. An R^2 value of 1 indicates a perfect fit, where the model explains all the variability in the data, while an R^2 value of 0 means the model explains none of the variability. Negative values can occur if the model performs worse than a simple mean-based model.

A sensitivity analysis is conducted to evaluate the impact of each feature on the model's predictions by perturbing input features and observing changes in the output. To address potential sensor failure issues, simulations are performed with missing data to test the robustness of the Random Forest model. Strategies for integrating redundant sensor data or alternative data sources are explored to ensure continuous and reliable monitoring. Results are visualised through various plots. A prediction versus actual plot compares the model's predictions with observed seawater levels, and a feature importance plot displays the relative significance of each feature in the Random Forest model.

Result and Discussion

The assessment of the Random Forest model for forecasting sea water levels based on atmospheric variables (barometric pressure, temperature, and relative humidity) shows encouraging outcomes. The analysis process is carried out in several stages such as feature importance analysis to see the relationship between parameters, model performance to test the application of the model and scatter plot analysis to see the level of model precision. In the training, validation, and testing, the final results of the model testing are discussed, while temporal trends and analysis of noise in predicted water levels complete the temporal analysis and noise output from the model.

Feature Importance Analysis

Feature analysis shows the relationship between water level and atmospheric variables of air pressure, temperature, and relative humidity (RH). This analysis includes an explanation of the physical relationship between changes in water level and fluctuations in each atmospheric parameter, as well as an interpretation of the patterns that appear in the visualization graph.

The first stage of analysis is to analyze the relationship between water level and air pressure. The graphic visualization shows a clear sinusoidal pattern between changes in air pressure and changes in water level. The pattern that appears shows a close relationship between air pressure fluctuations and variations in water height at the research location.

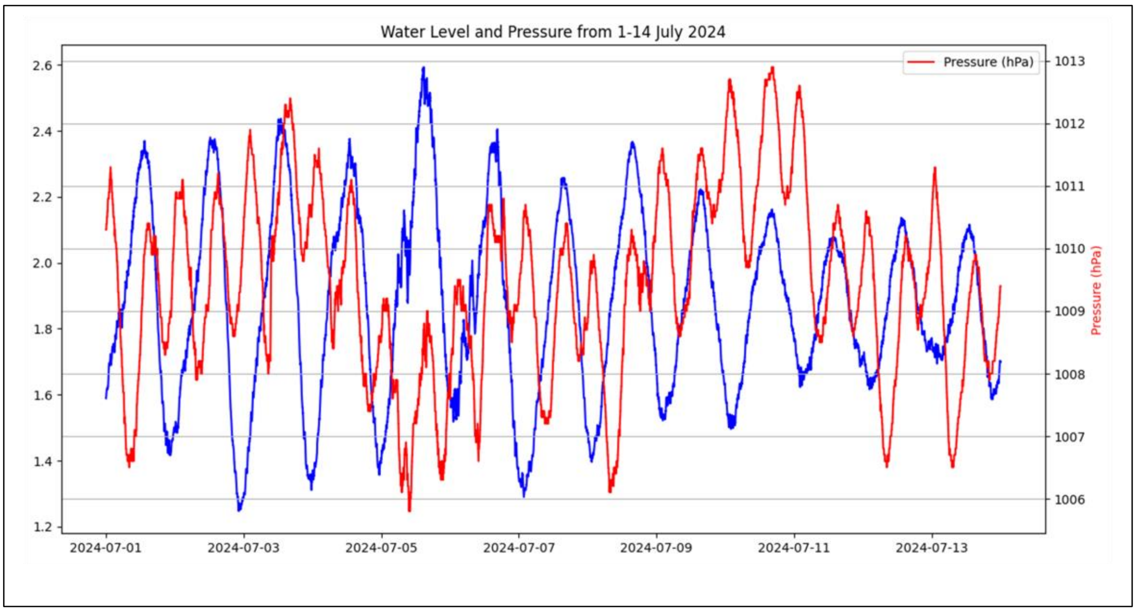

Figure 2. Relationship between water level sensor value and barometer sensor

The results in Figure 2 show the relationship between air pressure fluctuations and water level variations. The mathematical relationship between changes in water level or tides and air pressure can be explained through the basic concepts of atmospheric physics and fluid dynamics. Air pressure plays a role in influencing sea level through an effect known as the inverse barometer effect. The inverse barometer effect happens when air pressure increases; the downward pressure on the sea surface also increases, causing a decrease in sea level or tides. Conversely, when air pressure decreases, the downward pressure decreases, allowing seawater to rise or a rise in tides.

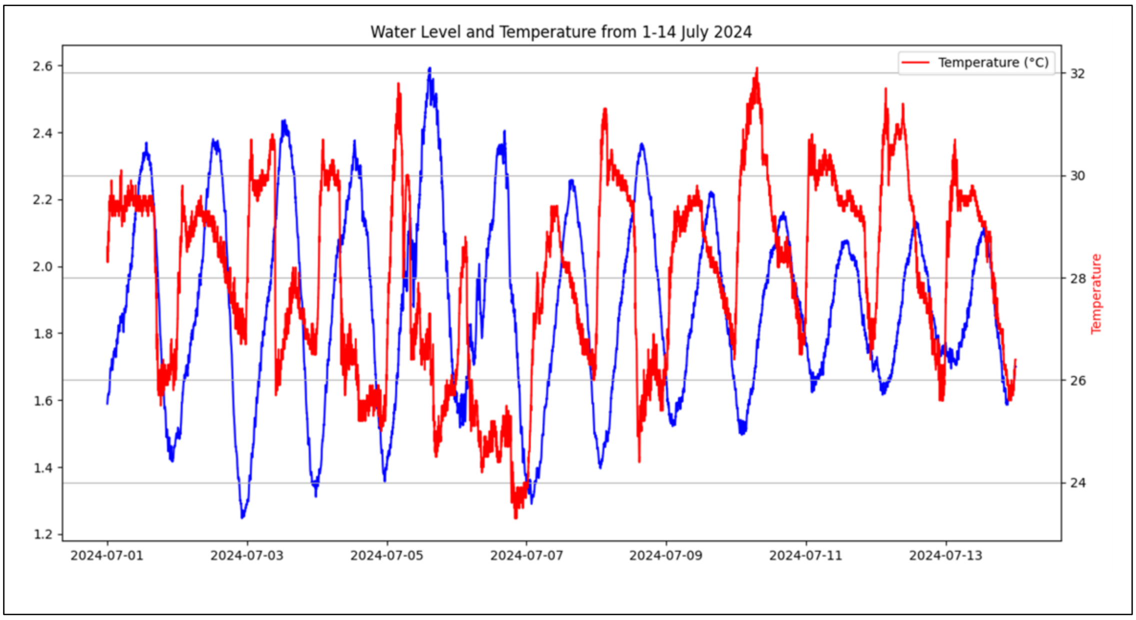

Figure 3. Relationship between water level sensor value and temperature sensor

Figure 3 shows this sinusoidal pattern, indicating a relationship between the two variables during the observation period. Although there are some similarities in the pattern with water level, air temperature does not follow the exact same sinusoidal pattern as the relationship between water level and barometer. Mathematically, the relationship between water level or tides and air temperature can be explained by the effect of temperature on water density and thermal expansion. Increasing air temperature causes thermal expansion of seawater, which slightly increases the water volume and can impact water level rise.

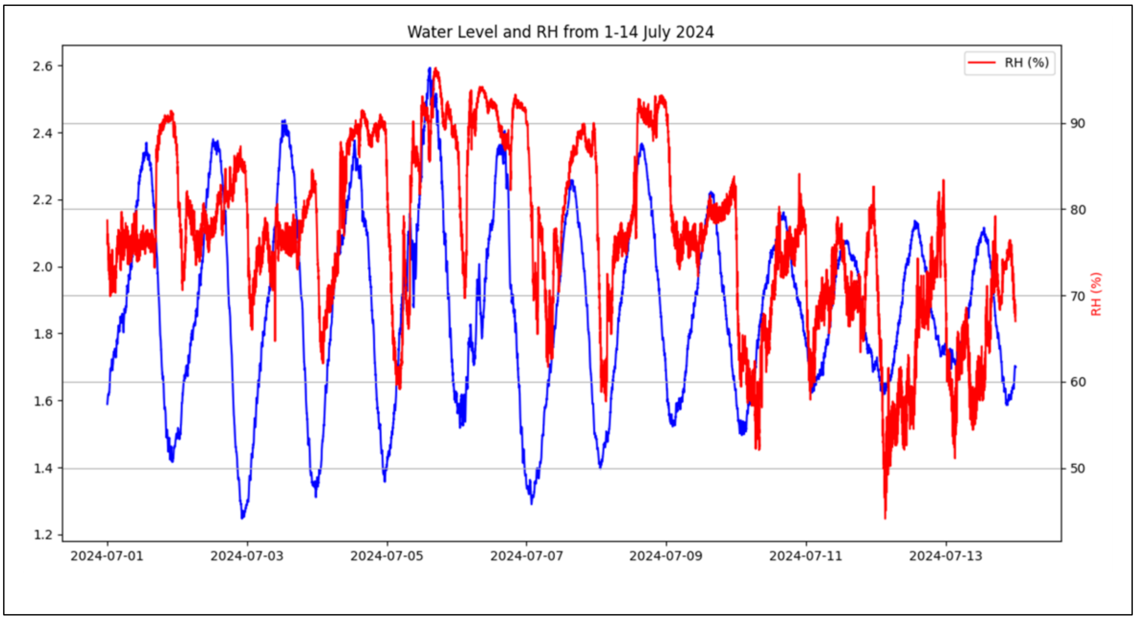

Figure 4. Relationship between water level sensor value and relative humidity sensor

The relative humidity data in Figure 4 also show significant variations, although the fluctuation pattern is not as systematic as the water level with the barometer. The water level fluctuation shows repeated peaks and valleys during the observation period, reflecting the dynamics of the tides at the study site. The relative humidity fluctuates in a more complex way than the water level. Although no pattern is completely consistent, there are several moments where the increase in relative humidity coincides with the increase in water level.

Model Performance

The model's mean-squared error (MSE) is 0.0209, and the root-mean-squared error (RMSE) is 0.1448. These metrics indicate that the model can provide predictions with relatively small errors. In contrast, the RMSE provides a more straightforward interpretation in the same unit as the sea water level target variable. Furthermore, the correlation coefficient (R) between the predicted and actual water levels is 0.8415, signifying a strong positive linear relationship and confirming that the model effectively captures the variability in the sea water level data.

Scatter Plot Analysis

Scatter plot illustrating actual versus predicted water level data for February, shwon in Figure 5. The points are close to the red dashed line, signifying an ideal prediction. Despite occasional anomalies, particularly at high values, the overall trend demonstrates the model's strong performance across a broad spectrum of water levels.

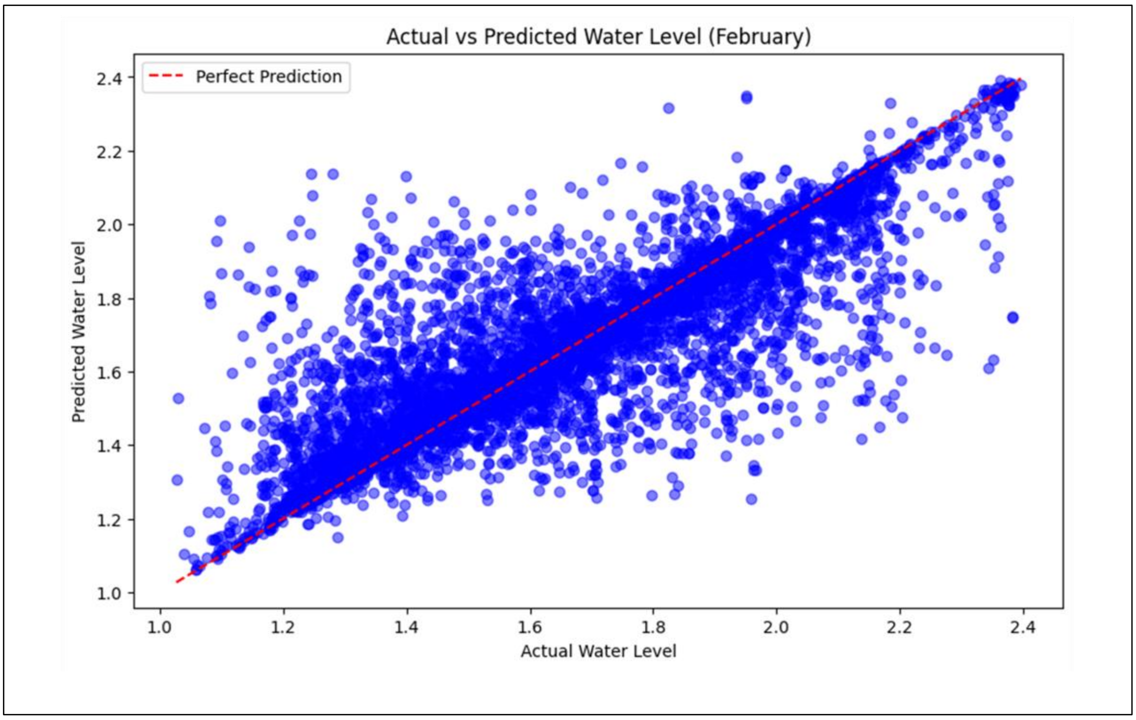

Figure 5. Scatter plot illustrating actual versus predicted water level data

Training, Validation, and Testing Results

Figure 6 provides a comprehensive analysis of model performance across training, validation, and testing datasets, demonstrating the durability of the Random Forest model. The training data resulted in an R^2 of 0.9728, indicating an exceptional match with low overfitting. The validation data obtained an R^2 of 0.8315, indicating that the model generalises effectively to novel data. The test data gave an R^2 of 0.8399, further validating the model's predictive capability. The aggregate R^2 for all datasets is 0.9199, confirming the model's capacity to account for almost 91% of the variance in seawater levels. These findings emphasise the significance of incorporating atmospheric variables as features and the usefulness of the Random Forest model for this application.

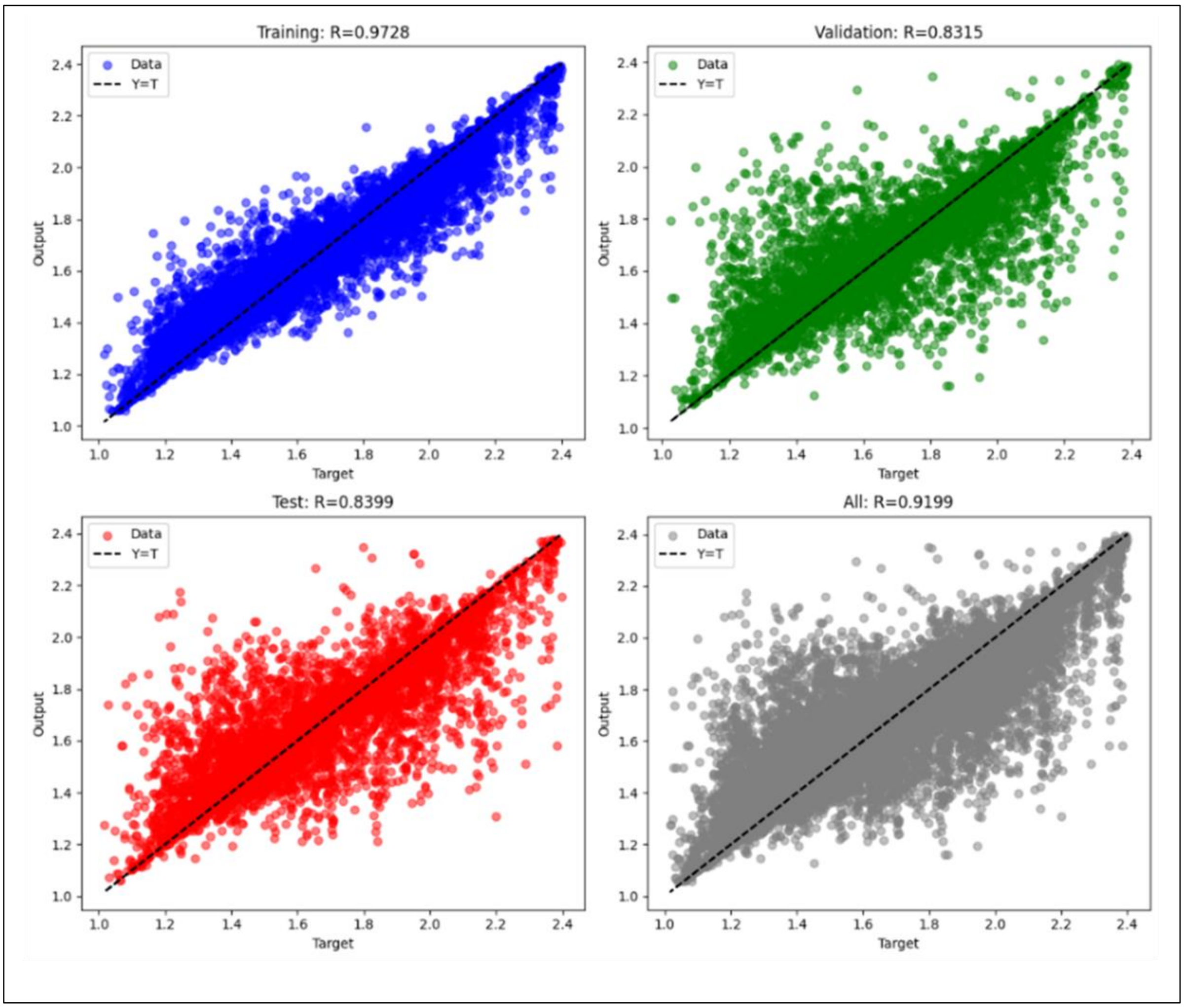

Figure 6. Performance outcomes of the model

Temporal Trends

Temporal trend analysis illustrating the observed versus predicted water levels over time for the initial week of February, illustrated in Figure 7. It demonstrates that the model accurately reflects the natural daily trends in the data. Differences in forecasted water levels become particularly noticeable during periods of rapid fluctuation, such as tidal peaks and troughs. This may indicate constraints in the resolution of the input characteristics or the fundamental complexity of tidal dynamics, which the model could only partially encapsulate.

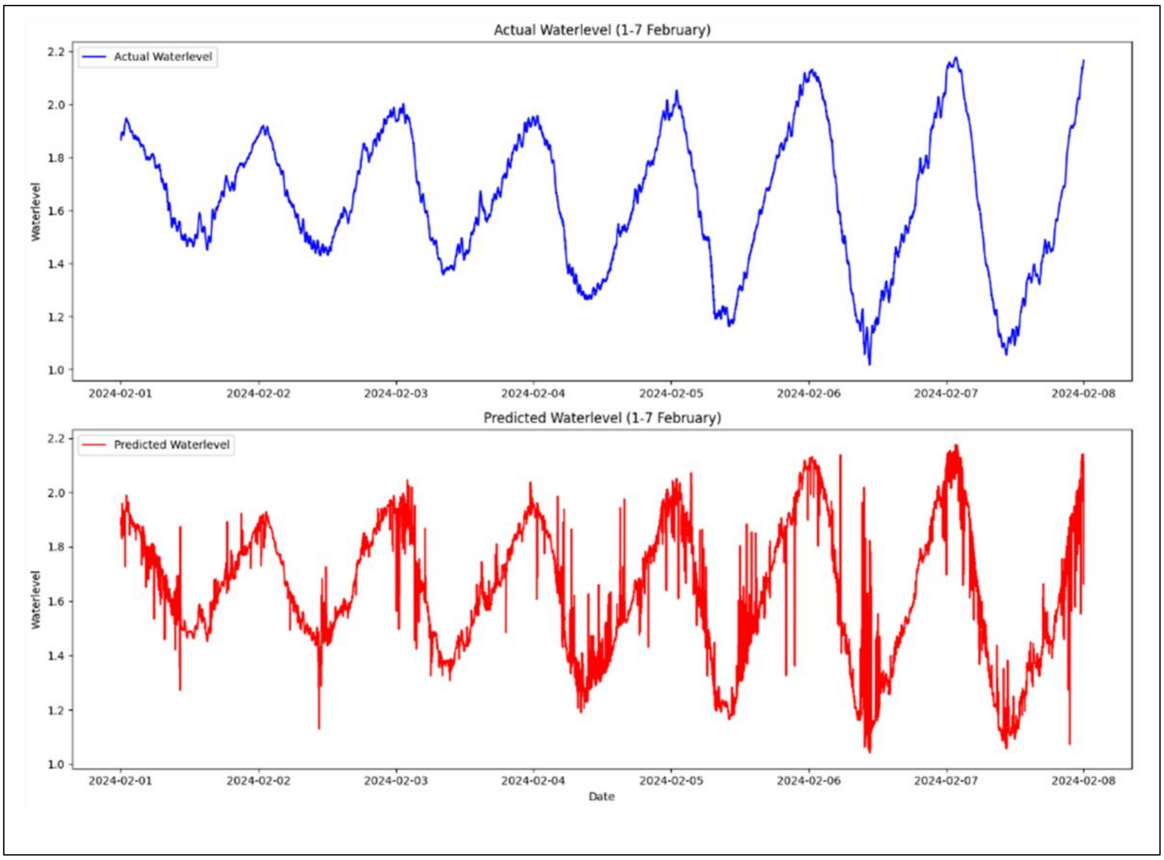

Figure 7. Temporal examination of observed versus predicted water levels

The Random Forest model's robust correlation and minimal error metrics indicate its efficacy in predicting sea water levels. Including atmospheric pressure, temperature, and relative humidity as input variables substantially enhances the model's usefulness. The high R^2 values for training, validation, and testing datasets suggest that the model balances accuracy and generalisation.

Analysis of Noise in Predicted Water Levels

The temporal analysis of actual versus predicted water levels from February 1–7 reveals notable discrepancies, particularly in noise in the predicted values. The fluctuations are apparent in the lower panel of Figure 8, where predicted water levels reveal considerable differences, particularly during swift variations in water levels. The deviations, indicated as fluctuations exceeding 5 cm (marked by orange spots), illustrate probable noise sources and constraints in the model's forecast accuracy.

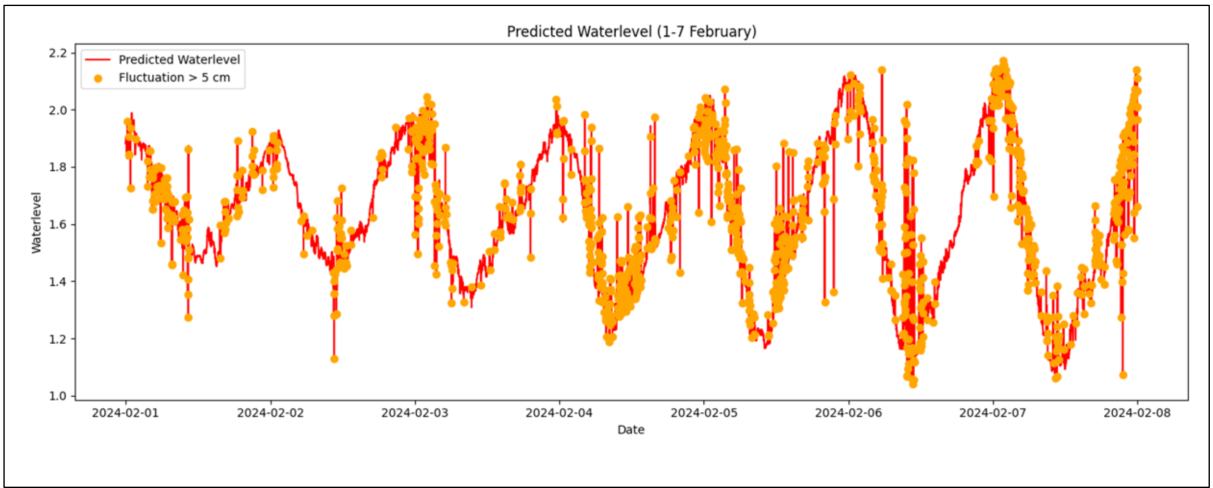

Figure 8. The variability of water level predictions

The noise in the predicted water levels indicates the necessity for more sophisticated modelling approaches or improvements to the existing methodology. Although the Random Forest model demonstrates robust performance overall, mitigating these irregularities through enhanced feature selection, preprocessing, or alternative models could improve its predictive accuracy, especially in scenarios demanding precision under extreme conditions. Future research should use temporal models and supplementary environmental data to mitigate these constraints.

Conclusions

This study illustrates the effectiveness of the Random Forest model in predicting sea water levels based on atmospheric factors, including barometric pressure, temperature, and relative humidity. The model achieved significant performance metrics, featuring a high correlation coefficient (R = 0.8415), a low Mean Squared Error (MSE) of 0.0209, and a Root Mean Squared Error (RMSE) of 0.1448, thereby validating its ability for precise predictions. The research indicates that the model accurately reflects natural daily tidal rhythms, as seen by the temporal trends in predictions. Nevertheless, significant noise was detected in the predicted water levels, especially during swift tidal variations, indicating possible constraints in the feature collection or the model's management of temporal dependencies. This indicates that although the Random Forest model is an effective instrument for general water level forecasting, additional refinements are required to mitigate noise and enhance accuracy under extreme conditions. Future research should include supplementary environmental variables, including wind speed and historical water level data while investigating temporal models such as LSTM or hybrid methodologies to manage sequential dependencies more effectively. This research demonstrates that machine learning-based soft sensors may augment physical sensors in maritime and coastal monitoring, offering a dependable and scalable solution for continuous water level forecasting.

References

Anwar, S., Satria, S. D. D., Muzadi, M. F., Zulkarnain, M. R., & Hidayatullah, N. S. (2023). Efektivitas Alat Pendeteksi Pasang Surut Air Laut di Desa Tajung Widoro. *Journal of Engineering and Innovation*, *1*(1), 29–33.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *Peerj Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

Djunarsjah, E., Nusantara, C. A. D. S., Putra, A. P., Wijaya, R. A., Sianturi, S. S., Anantri, N. M. K., Kusumadewi, D., & Julian, M. M. (2023). Prospects and Constraints of Lowest Astronomical Tide (LAT) as Determination of Sea Boundaries in Indonesia. *The Egyptian Journal of Aquatic Research*, *49*(4), 444–451. https://doi.org/10.1016/j.ejar.2023.08.002

Fatkhurrohman, F., Siregar, M. R. T., Sugiarto, S., & Putra, M. (2023). Performance analysis of Sebesi Island automatic water level station as tsunami confirmation system. *AIP Conference Proceedings*, *2865*(1). https://doi.org/10.1063/5.0182372

Gultom, F., Harsono, G., Pranowo, W. S., & Adrianto, D. (2017). Sistem Informasi Pasang Surut Berbasis Android di Wilayah Kerja Pangkalan TNI Angkatan Laut (Studi Kasus Belawan, Tarempa, Sibolga, Natuna dan Cilacap): Android Based Tidal Information System in the Working Area of the Naval Base (Case Study of Belawan,. *Jurnal Chart Datum*, *3*(2), 81–92. https://doi.org/10.37875/chartdatum.v3i2.121

Han, Y., Sun, J., & Pan, Q. (2024). Prediction of drum water level of power plant boiler based on random forest. *2024 3rd International Conference on Energy, Power and Electrical Technology (ICEPET)*, 32–36. 10.1109/ICEPET61938.2024.10626447

Illigner, J., Sofian, I., Abidin, H. Z., Syafi'i, M. A., & Schöne, T. (2016). Coastal sea level monitoring in Indonesia: connecting the tide gauge zero to leveling benchmarks. *IAG 150 Years: Proceedings of the IAG Scientific Assembly in Postdam, Germany, 2013*, 451–457. https://link.springer.com/chapter/10.1007/1345_2015_23

Li, B., Yang, G., Wan, R., Dai, X., & Zhang, Y. (2016). Comparison of random forests and other statistical methods for the prediction of lake water level: a case study of the Poyang

Lake in China. *Hydrology Research*, *47*(S1), 69–83. http://dx.doi.org/10.2166/nh.2016.264

Lu, B., & Hardin, J. (2021). A unified framework for random forest prediction error estimation. *Journal of Machine Learning Research*, *22*(8), 1–41. http://dx.doi.org/10.48550/arXiv.1912.07435

Malik, K., Putra, I. W. S. E., Darwan, D., Wiratama, A. P., Setiawan, F., & Saputro, B. S. (2023). Studi Karakteristik Pasut, Arus Pasut, dan Gelombang pada Perencanaan Pembangunan Pangkalan TNI AL Saumlaki Menggunakan Pemodelan Numerik: Study of The Characteristics of Tidals, Tidal Currents, And Waves in Development Planning of The Saumlaki Naval Base. *Jurnal Chart Datum*, *9*(2), 123–134. https://doi.org/10.37875/chartdatum.v9i2.293

Martin Miguez, B., Testut, L., & Wöppelmann, G. (2012). Performance of modern tide gauges: towards mm-level accuracy. *Scientia Marina*, *76*(S1), 221–228. 10.3989/scimar.03618.18A

Matondang, J. S., Sumargana, L., Adityawarman, Y., & Yuniantoro, R. M. T. (2020). Development of near real-time and archival Tsunami data visualization dashboard for Indonesia. *2020 IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology (AGERS)*, 36–41. https://doi.org/10.1109/AGERS51788.2020.9452771

Novianto, D., Husrin, S., Nugroho, D., Bramawanto, R., Setiawan, A., Permana, S. M., Sufyan, A., Sianturi, D. S. A., Daniel, D., & Suhelmi, I. R. (2021). IDSL (Inexpensive Device for Sea Level) performance analysis for tews (Tsunami Early Warning System) in Sadeng fisheries port. *IOP Conference Series: Earth and Environmental Science*, *860*(1), 012101. 10.1088/1755-1315/860/1/012101

Pugh, D. T. (1987). *Tides, surges and mean sea level*.

Putra, A. A., Soepardjo, A. H., Purwanti, S., Widyantoro, B. T., Efendi, J., & Pramono, G. H. (2021). Validation of Badan Informasi Geospasial Tide Model in the Sunda Strait Waters using Sebesi Island Tide Gauge Data. *IOP Conference Series: Earth and Environmental Science*, *750*(1), 012036. https://iopscience.iop.org/article/10.1088/1755-1315/750/1/012036/meta

Putra, M. (2021). Desain dan Implementasi Automatic Water Level Berbasis Sensor Akustik untuk Pemantauan Tinggi Muka Laut di Selat Sunda. *Seminar Nasional Teknik Elektro*, *6*(1), 65–69.

Sagala, H. A., Pasaribu, R. P., & Ulya, F. K. (2021). Pemodelan Pasang Surut dengan Menggunakan Metode Flexible Mesh untuk Mengetahui Genangan Rob di Pesisir Karawang. *PELAGICUS*, *2*(3), 141–156. http://dx.doi.org/10.15578/plgc.v2i3.10341

Sandro, T., Hairatunnisa, H., & Putra, M. (2020). Analisis kinerja sensor water level-pressure transducer (Studi kasus automatic weather system di Pelabuhan Merak). *Electrices*, *2*(2), 37–42. https://doi.org/10.32722/ees.v2i2.3588

Sugiarto, S., Putra, M., Dharmawan, G. S. B., & Akram, Z. M. (2020). Pengembangan Stasiun Pengamatan Cuaca Maritim Otomatis Berbasis Webserver. *Prosiding Seminar Nasional Teknik Elektro UIN Sunan Gunung Djati Bandung*, 68–76.

Taqiyuddin, M., & Sasongko, T. B. (2024). Prediksi Cuaca Kabupaten Sleman Menggunakan Algoritma Random Forest. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *8*(3), 1683–1693. http://dx.doi.org/10.30865/mib.v8i3.7897