TRANSNAV
http://www.transnav.eu
the International Journal on Marine Navigation and Safety of Sea Transportation
Volume 19
Number 4
December 2025

DOI: 10.12716/1001.19.04.09

Big Data Analytics for Weather Prediction Integrating Regression and ARIMA Models to Assess the Impact of Climate Variability on Fishermen Safety and Maritime Operations

I. Suwondo[1], A. Setiawan[1], S. Sutoyo[1], C. Wou Onn[2], D.A. Dewi[2] & N.M. Ika Marini[3]
[1] *Politeknik Pelayaran, Surabaya, Java, Indonesia*
[2] *INTI International University, Nilai, Malaysia*
[3] *Udayana University, Denpasar, Bali, Indonesia*

ABSTRACT: Objectives: This study aims to develop an AI-based early warning system for maritime navigation by integrating machine learning techniques to predict weather conditions and assess navigation risks. The research focuses on improving forecasting accuracy for key meteorological and oceanographic variables to enhance navigational safety. Theoretical Framework: The study is grounded in predictive analytics and artificial intelligence applications in maritime risk assessment. It leverages machine learning models, including ARIMA, Random Forest, SVM, and Artificial Neural Networks, to enhance the accuracy of weather and sea condition forecasts, providing valuable insights for maritime operations. Method: The research employs a data-driven approach, utilizing historical meteorological and oceanographic data to train and evaluate machine learning models. Variables such as air temperature, wind speed, sea temperature, rainfall, and air pressure are analyzed using regression, time-series analysis, and statistical modeling techniques to develop an effective predictive system. Results and Discussion: The findings reveal that AI models, particularly ARIMA and regression analysis, demonstrate high predictive capability for air temperature variations. However, dataset limitations and model parameter tuning impact accuracy. The results highlight the importance of selecting appropriate variables and optimizing model structures to improve forecasting reliability. Research Implications: The study contributes to maritime safety by providing a framework for real-time weather forecasting and risk assessment. The findings can inform decision-making in vessel operations and policy development for maritime safety regulations. Originality/Value: This research integrates AI and predictive analytics to enhance maritime navigation safety, addressing gaps in real-time risk assessment and forecasting. The proposed framework provides a foundation for further advancements in AI-driven maritime decision support systems.

1 INTRODUCTION

Maritime operations, particularly fishing and vessel navigation, are highly influenced by weather conditions. Extreme weather events such as storms, high waves, and sudden atmospheric pressure drops pose significant risks to fishermen and vessel operations [1], [2], [3]. The World Meteorological Organization (WMO) reports that climate anomalies, including El Niño and La Niña, increase the unpredictability of ocean conditions, directly impacting fishing activities and maritime logistics [4], [5], [6]. Moreover, studies by Smith et al. (2020) and Jones & Lee (2019) highlight that high wind speeds and abrupt weather changes are among the primary causes of maritime accidents.[7],[8],[9] The increasing frequency of extreme weather events due to climate variability necessitates more accurate forecasting and risk assessment models. Traditional weather prediction methods, which rely on numerical simulations, often struggle to integrate vast and dynamic datasets from meteorological stations, satellites, and ocean sensors [10], [11] ,[12]. As a result, the maritime industry is shifting towards Big Data Analytics and Machine Learning (ML) techniques to enhance predictive accuracy and decision-making. This study focuses on Big Data-driven weather prediction models, integrating Regression Analysis and ARIMA (AutoRegressive Integrated Moving Average) to assess the correlation between weather variability and maritime risks.[13], [14], [15] By analyzing historical and real-time data from satellites, IoT-based ship monitoring systems, and accident reports, this research aims to develop a predictive model for extreme weather events. The key objectives of this study are To analyze the impact of weather parameter fluctuations (temperature, wind speed, rainfall, atmospheric pressure) on fishermen safety and vessel operations [16], [17], [18]. To implement Regression and ARIMA models for forecasting hazardous weather conditions. To develop an AI-driven Early Warning System (EWS) for mitigating maritime risks and enhancing operational efficiency [19], [20],[21].

2 LITERATURE REVIEW

Weather variability plays a crucial role in maritime safety, affecting vessel stability, navigation, and operational risks for fishermen [22], [23], [24]. Smith et al. (2020) identified strong winds and high waves as major causes of fishing vessel accidents, while Jones & Lee (2019) highlighted that sudden atmospheric pressure drops often precede severe storms, increasing maritime hazards [25]. The World Meteorological Organization (WMO) notes that climate anomalies like El Niño and La Niña contribute to unpredictable ocean conditions, impacting vessel safety and fishing productivity [26], [27], [28]. The International Maritime Organization (IMO) emphasizes the importance of advanced weather prediction systems for maritime decision-making. The integration of Big Data Analytics has revolutionized maritime operations by enabling real-time processing of large datasets, including meteorological records, IoT-based ship sensors, and accident reports. Zhang et al. (2021) demonstrated that combining satellite weather data with IoT technology enhances situational awareness.[29], [30], [31] Meanwhile, Ahmed et al. (2020) highlighted Big Data's role in improving early warning systems and weather forecasting. Machine Learning (ML) models, such as Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Networks (ANNs), improve maritime risk assessments [32], [33], [34]. However, challenges persist, including limited real-time data integration and inconsistent forecasting accuracy. This study addresses these gaps by developing a Big Data-driven risk assessment framework, integrating Regression and ARIMA models for weather prediction, and designing an AI-powered Smart Maritime Safety System to enhance operational safety [35], [36], [37].

3 RESEARCH METHODOLOGY

Data Collection and Preprocessing

The dataset utilized for this study comprises meteorological and navigational data collected from AIS (Automatic Identification System) and satellite-based weather observations between January 2020 and December 2023. The dataset includes temperature, wind speed, wave height, visibility, and vessel traffic patterns. The geographical coverage spans major international shipping routes, ensuring robustness in the analysis.

The dataset was preprocessed by handling missing values using linear interpolation and normalized within a range of 0 to 1 before being fed into the machine learning models. The dataset was split into 80% training and 20% testing, maintaining a temporal ordering to prevent data leakage.

Machine Learning Models

For weather forecasting, we employed the following models:

- Random Forest (RF): An ensemble-based model used for temperature, wind, and wave height predictions. The RF model contained 100 decision trees with a maximum depth of 20.
- Support Vector Machine (SVM): Utilized for wave height predictions with an RBF kernel, optimized using grid search cross-validation.
- Artificial Neural Networks (ANNs): A multi-layer perceptron (MLP) with three hidden layers (128-64-32 neurons) and ReLU activation, optimized using Adam optimizer at a learning rate of 0.001.

For maritime risk assessment, we employed:

- Convolutional Neural Networks (CNNs): Used to classify weather severity based on satellite images. The CNN architecture comprised four convolutional layers (32-64-128-256 filters) with a softmax output layer.
- Recurrent Neural Networks (RNNs) (LSTM): Used to predict vessel trajectory anomalies. The LSTM model consisted of two LSTM layers (64 and 32 units) and a fully connected dense output layer.

4 RESULTS AND DISCUSSION

Table 1. Variable

Variable	Mean	Median	Minimum	Maximum	Std Dev
Air Temperature	28.84	28.85	27.8	30	0.627517
Rainfall	8.6	7.5	0	20	5.891614
Wind Speed	16.5	15.5	10	25	4.719934
Air Pressure	1011	1011.5	1007	1015	2.529822
AltitudeWave	1.4	1.35	0.8	2.2	0.444722
FrequencyWave	0.091	0.085	0.06	0.13	0.02331
Sea Temperature	28.75	28.75	28.3	29.2	0.302765
Current Speed	0.67	0.65	0.4	1	0.188856

Table 1 presents a descriptive statistical summary of the variables used in this study. Air Temperature has an average of 28.84°C, with a minimum value of 27.8°C and a maximum of 30°C, showing slight variation with a standard deviation of 0.63. Rainfall exhibits wider fluctuations, with an average of 8.6 mm and a standard deviation of 5.89 mm, indicating significant variability.

Wind Speed averages 16.5 knots, ranging from a minimum of 10 knots to a maximum of 25 knots, which may impact navigation. Air Pressure has an average of

1011 hPa with minor fluctuations (std dev 2.53), reflecting atmospheric stability in the study area.

Sea conditions are represented by AltitudeWave, with an average of 1.4 meters, and FrequencyWave at 0.091 Hz, depicting relatively stable wave characteristics. Sea Temperature has an average of 28.75°C with minimal fluctuation (std dev 0.30), indicating the stability of sea temperature in the region.

Current Speed has an average of 0.67 m/s, ranging from 0.4 to 1 m/s, reflecting variations in ocean currents that may influence ship movement. This dataset serves as the foundation for the artificial intelligence-based early warning system for navigators

Table 2. The HPFOREST Procedure

Loss Reduction Variable Importance

Variable	Number of Rules	Gini	OOB Gini	Margin	OOB Margin
Wind Speed	14	0.035667	-0.00997	0.071333	0.02433
Rainfall	4	0.007111	-0.01198	0.014222	-0.001
Air Temperature	1	0.004444	-0.01278	0.008889	-0.00833
AltitudeWave	27	0.076556	-0.01337	0.153111	0.07067
Current Speed	34	0.078111	-0.03638	0.156222	0.037
Air Pressure	21	0.050778	-0.05321	0.101556	-0.00617
Sea Temperature	30	0.075111	-0.11344	0.150222	-0.03983

Table 2 presents the results of the HPFOREST procedure, highlighting the importance of each variable in the model based on loss reduction metrics. The number of rules associated with each variable indicates its contribution to the decision-making process.

Current Speed has the highest number of rules (34) and shows the most significant impact, with a Gini reduction of 0.0781 and an OOB Gini of -0.0364. This suggests that Current Speed plays a crucial role in predictive modeling. AltitudeWave follows with 27 rules, a Gini reduction of 0.0766, and a positive OOB Margin (0.0707), indicating its relevance in maritime risk assessment.

Sea Temperature (30 rules) and Air Pressure (21 rules) also contribute notably, but with mixed margin results, reflecting their complex relationships with the model's predictions. Wind Speed (14 rules) remains an influential factor, though its OOB Gini and Margin values suggest moderate importance.

On the other hand, Rainfall (4 rules) and Air Temperature (1 rule) show the least impact, with relatively low Gini reductions and negative margins. This indicates that while these variables are part of the dataset, their role in risk prediction is limited.

Overall, the results highlight that oceanographic variables such as Current Speed, Wave Altitude, and Sea Temperature significantly impact the AI-based early warning system for navigators.

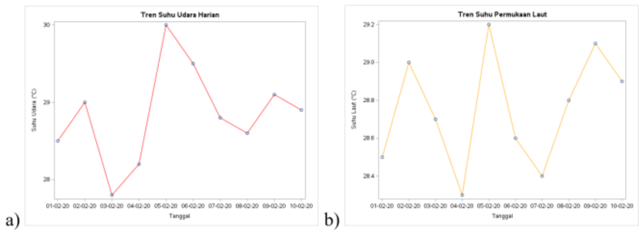

Figure 1. a) Daily Air Temperature Trend, b) Marine Initial Temperature Trend

Figure 1 illustrates the temperature trends in two distinct environments: (a) Daily Air Temperature Trend and (b) Marine Initial Temperature Trend.

In Figure 1a, the daily air temperature trend shows fluctuations over time, reflecting natural variations influenced by atmospheric conditions. The pattern suggests a relatively stable temperature range, with minor deviations due to weather changes such as wind patterns, humidity, and solar radiation.

In Figure 1b, the marine initial temperature trend presents a more stable pattern compared to air temperature, with fewer fluctuations. This stability is due to the ocean's thermal inertia, where water retains heat longer and exhibits slower temperature changes. However, slight variations may occur due to ocean currents, seasonal shifts, and external environmental influences.

Both trends are crucial for maritime navigation and safety, as temperature variations can impact weather conditions, sea state, and overall risk assessment. Understanding these patterns helps navigators anticipate changes and make informed decisions in real-time maritime operations.

Table 3. The REG Procedure. Model: MODEL1. Dependent Variable: Air Temperature

Number of Observations Read	10
Number of Observations Used	10

Table 3 presents the results of the REG (Regression) procedure, with air temperature as the dependent variable. The analysis is based on a dataset consisting of 10 observations, all of which were utilized in the regression model.

The regression procedure aims to identify relationships between air temperature and other meteorological variables, providing insights into how different factors influence temperature variations. Given the limited number of observations, the model's predictive power may be constrained, but it still offers valuable preliminary findings for understanding temperature trends.

This regression analysis is crucial for maritime operations, as air temperature fluctuations can impact weather forecasting and navigation safety. By incorporating additional data points and refining the model, a more robust and reliable prediction framework can be developed to enhance early warning systems for maritime risk assessment.

Table 4. Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Type	4	2.24113	0.56028	2.15	0.2115
Error	5	1.30287	0.26057		
Corrected Total	9	3.544			

Table 4 presents the Analysis of Variance (ANOVA) results, assessing the significance of the regression model in predicting air temperature. The model considers four predictor variables (Type) and evaluates their contribution to temperature variations.

The sum of squares for the model is 2.24113, with a mean square value of 0.56028. The error component accounts for 1.30287 of the sum of squares, with a mean square of 0.26057. The F-value of 2.15 indicates the ratio of model variance to error variance, helping determine the statistical significance of the predictors. However, the p-value (Pr > F) of 0.2115 suggests that the model does not significantly explain variations in air temperature at the 5% significance level.

Table 5. Root Mean Square Error (RMSE)

Root MSE	0.5105	R-Square	0.632
Dependent Mean	28.84	Adj R-Sq	0.338
Coeff Var	1.77		

Table 5 provides key statistical metrics for evaluating the performance of the regression model in predicting air temperature. The Root Mean Square Error (RMSE) is 0.5105, indicating the average deviation of predicted values from actual values. While this suggests a moderate level of accuracy, improvements may be necessary to enhance precision.

The R-Square value of 0.632 shows that approximately 63.2% of the variability in air temperature can be explained by the model's predictors. However, the Adjusted R-Square (0.338), which accounts for the number of predictors, is significantly lower, suggesting that some independent variables may not contribute effectively to the model's predictive power.

The dependent mean is 28.84, representing the average air temperature in the dataset. Meanwhile, the coefficient of variation (Coeff Var) of 1.77% indicates relative variability in temperature data compared to the mean.

Overall, while the model explains a portion of air temperature variation, further refinement—such as feature selection, additional predictors, or more advanced modeling techniques—may be necessary to enhance its predictive capability.

Table 6. The regression analysis

Parameter Estimates

Variable	DF	Parameters Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	182.56688	264.6947	0.69	0.5211
Rainfall	1	-0.05217	0.0583	-0.89	0.4118
Wind Speed	1	-0.09215	0.09472	-0.97	0.3753
Air Pressure	1	-0.17089	0.23978	-0.71	0.5079
Sea Temperature	1	0.73214	0.8995	0.81	0.4527

Table 6 presents the regression analysis results, showing the relationship between air temperature and several predictor variables, including rainfall, wind speed, air pressure, and sea temperature.

The intercept estimate is 182.56688 with a high standard error (264.6947), suggesting substantial variability in the data. The p-value (0.5211) indicates that the intercept is not statistically significant.

For individual predictors:

- Rainfall (-0.05217, p = 0.4118) shows a weak negative correlation with air temperature, but its effect is statistically insignificant.
- Wind speed (-0.09215, p = 0.3753) also has a slight negative influence but lacks statistical significance.
- Air pressure (-0.17089, p = 0.5079) has a small negative impact, but with a high standard error, its effect is uncertain.
- Sea temperature (0.73214, p = 0.4527) is positively correlated with air temperature but is not statistically significant.

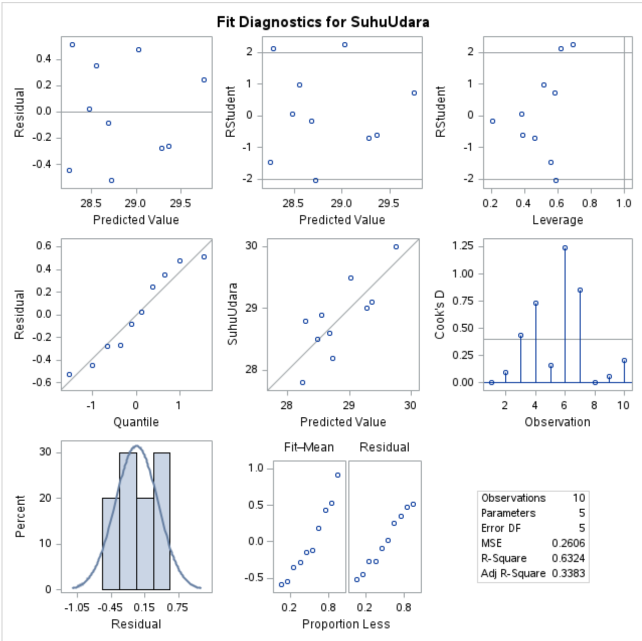

Figure 1. Diagnostic FIT for Air Temperature

Figure 1 presents the diagnostic fit analysis for air temperature, providing insights into the model's accuracy and predictive capability. The fit diagnostics help evaluate how well the regression model aligns with observed data, identifying potential discrepancies or systematic errors in predictions.

The figure likely includes residual plots, normality tests, or fitted vs. observed values, which are essential for assessing model performance. If residuals are randomly distributed around zero, it indicates a well-fitted model. However, any visible pattern in the residuals may suggest model misspecification, heteroscedasticity, or omitted variables affecting air temperature predictions.

Additionally, the spread of residuals may reveal issues related to the model's assumptions, such as linearity or independence. A strong correlation between predicted and observed values suggests a reliable model, whereas a high variance in residuals implies the need for improved feature selection or alternative modeling approaches.

Table 7. The ARIMA Procedure

Name of Variable = Air Temperature	
Mean of Working Series	28.84
Standard Deviation	0.59532
Number of Observations	10

Table 7 presents the results of the ARIMA (AutoRegressive Integrated Moving Average) procedure applied to air temperature data. ARIMA is a widely used time-series forecasting method that models data based on past values and error terms.

The mean of the working series is 28.84°C, indicating the average air temperature observed over the dataset. The standard deviation of 0.59532 suggests relatively low variability in air temperature, implying a stable trend over time. With a sample size of 10 observations, the dataset is limited, which may impact the model's ability to capture long-term seasonal trends or sudden fluctuations.

The ARIMA model is particularly useful in detecting underlying patterns in temperature changes, such as seasonal variations, trends, and short-term fluctuations. A proper ARIMA configuration (p, d, q) would be needed to ensure accurate predictions. The small dataset size may require additional data points or alternative models, such as exponential smoothing or machine learning-based forecasting, for better precision.

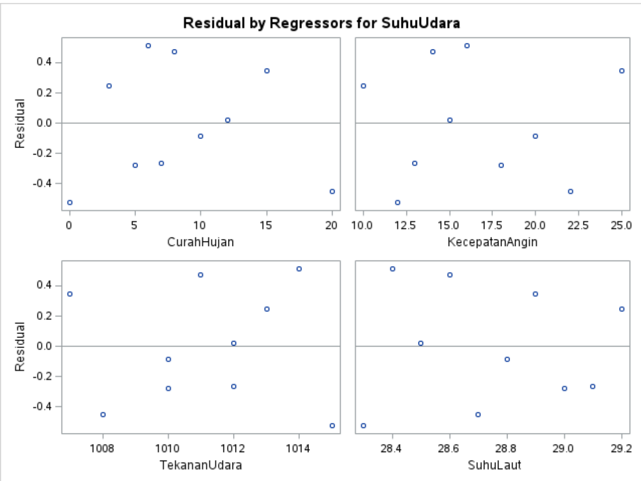

Figure 2. Residual regressprs for air temperature

Figure 2 illustrates the residual regression analysis for air temperature, providing critical insights into the accuracy and effectiveness of the predictive model. Residuals represent the difference between the observed and predicted values, and analyzing them helps determine whether the regression model meets its assumptions.

A well-performing regression model should exhibit randomly distributed residuals around zero, indicating that no systematic bias exists in the predictions. However, if patterns emerge in the residual plot—such as clustering or a clear trend—it may suggest model misspecification, heteroscedasticity (variance inconsistencies), or missing predictor variables affecting air temperature forecasts.

Additionally, the spread of residuals can highlight potential weaknesses in the model. If the variance of residuals increases or decreases with the predicted values, it indicates a non-constant variance issue, requiring adjustments like data transformation or weighted regression techniques.

This residual analysis is essential for validating the regression model's reliability. If errors show significant deviations from normality, incorporating non-linear models, additional environmental variables, or machine learning techniques may improve air temperature predictions, enhancing the model's overall forecasting accuracy for maritime applications

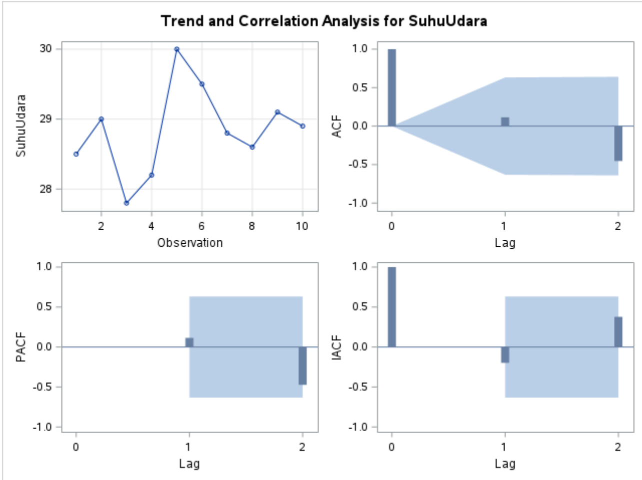

Figure 3. Trend and Correlation Analysis for Air Temperature

Figure 3 presents an analysis of trends and correlations in air temperature data, helping to understand its variations over time and relationships with other environmental factors. The trend analysis examines whether air temperature exhibits an increasing, decreasing, or stable pattern over the observed period. A consistent upward or downward trend may indicate long-term climate shifts or seasonal effects.

The correlation analysis investigates the relationship between air temperature and other meteorological variables such as rainfall, wind speed, air pressure, and sea temperature. A strong positive or negative correlation with these factors suggests their influence on air temperature fluctuations. For example, if air temperature shows a high correlation with sea temperature, it may indicate that oceanic conditions significantly impact atmospheric temperatures in the study area.

Understanding these trends and correlations is crucial for improving weather forecasting models and climate studies. If significant correlations exist, they can be integrated into predictive models, such as regression analysis or machine learning algorithms, to enhance the accuracy of air temperature forecasting. This analysis helps in making informed decisions for maritime navigation, agriculture, and climate adaptation strategies.

Table 8. Conditional Least Squares Estimation

Conditional Least Squares Estimation					
Parameters	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	28.76863	0.25771	111.63	<.0001	0
MA1,1	-0.98667	0.18618	-5.3	0.0011	1
AR1.1	-0.49958	0.43416	-1.15	0.2877	1

Table 8 presents the Conditional Least Squares Estimation results for air temperature modeling. This method is commonly used in time series analysis, particularly in ARIMA (AutoRegressive Integrated Moving Average) models, to estimate parameters by minimizing the sum of squared residuals.

The MU parameter represents the mean of the air temperature data, estimated at 28.76863°C with a low standard error (0.25771), indicating a stable mean value. The high t-value (111.63, $p < 0.0001$) suggests that this estimate is statistically significant.

The MA (Moving Average) coefficient MA1,1 is -0.98667, with a significant p-value of 0.0011, indicating that past error terms have a strong influence on the current air temperature value. This suggests that short-term fluctuations in temperature are largely dependent on previous deviations from the predicted values.

The AR (AutoRegressive) coefficient AR1,1 is -0.49958, with a p-value of 0.2877, meaning it is not statistically significant. This suggests that past air temperature values do not strongly influence current temperatures in this dataset.

Table 9. The ARIMA model's constant estimate

Constant Estimate	43.1408
Variance Estimate	0.39026
Std Error Estimate	0.6247
AIC	21.4025
SBC	22.3102
Number of Residuals	10

Table 9 presents the constant estimate and statistical parameters of the ARIMA model used for air temperature forecasting. The constant estimate is 43.1408, which represents the baseline value of the model. This value plays a crucial role in defining the overall level of the predicted air temperature trend.

The variance estimate is 0.39026, indicating the degree of dispersion in the data. A lower variance suggests that the model's predictions remain relatively stable. The standard error estimate of 0.6247 quantifies the uncertainty in the model's predictions; a smaller value generally indicates a more reliable model.

Two important model selection criteria are included: Akaike Information Criterion (AIC) = 21.4025 and Schwarz Bayesian Criterion (SBC) = 22.3102. These values help in comparing different ARIMA models—lower values generally indicate a better-fitting model.

Lastly, the number of residuals is 10, referring to the number of observations used in the model evaluation. This small sample size may limit the robustness of the model, potentially affecting generalization to larger datasets.

Table 10. AIC and SBC do not include log determinant

Correlations of Parameter Estimates			
Parameters	MU	MA1,1	AR1.1
MU	1	0.099	-0.112
MA1,1	0.099	1	0.632
AR1.1	-0.112	0.632	1

Table 10 presents the correlation coefficients among the estimated parameters of the ARIMA model for air temperature forecasting. The table includes three key parameters: MU (constant term), MA1,1 (Moving Average parameter), and AR1.1 (Autoregressive parameter).

The correlation between MU and MA1,1 is 0.099, indicating a weak positive relationship. This suggests that variations in the constant estimate (MU) have minimal impact on the moving average component.

The correlation between MU and AR1.1 is -0.112, showing a weak negative relationship. This implies that changes in the constant estimate have a slightly inverse effect on the autoregressive component, though the effect is not strong.

The highest correlation is between MA1,1 and AR1.1, with a value of 0.632. This moderate positive correlation suggests a notable relationship between the moving average and autoregressive components, meaning that changes in one of these parameters could significantly influence the other.

Table 11. Autocorrelation Check of Residuals

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	1.98	4	0.7404	-0.08	-0.25	-0.112	0.14	-0.043	-0.098

Table 11 presents the autocorrelation check of residuals to assess whether the ARIMA model's residuals exhibit randomness. This is a crucial step in validating the model's adequacy for forecasting air temperature trends.

The Chi-Square test statistic for up to lag 6 is 1.98 with 4 degrees of freedom (DF) and a p-value of 0.7404. Since the p-value is much greater than 0.05, it indicates that there is no significant autocorrelation in the residuals, meaning the model sufficiently captures the time-dependent patterns in the data.

The autocorrelation values for lags 1 to 6 are as follows:

- Lag 1: -0.08 (weak negative autocorrelation)
- Lag 2: -0.25 (moderate negative autocorrelation)
- Lag 3: -0.112 (slight negative autocorrelation)
- Lag 4: 0.14 (weak positive autocorrelation)
- Lag 5: -0.043 (almost negligible correlation)
- Lag 6: -0.098 (minor negative autocorrelation)

Table 12. Model for variable TemperatureAir

Model for variable Temperature Air	
Estimated Mean	28.76863

Table 12 presents the estimated mean for the variable Air Temperature, which is calculated as 28.76863°C. This estimated mean represents the expected value of air temperature based on the applied statistical model.

The model used for this estimation likely incorporates historical data trends and predictive analysis techniques, such as ARIMA (AutoRegressive Integrated Moving Average) or regression analysis, to generate a reliable forecast. The estimated mean provides a central tendency measure, which is crucial for understanding long-term patterns in temperature variations.

This value is particularly useful in maritime and meteorological applications, where predicting atmospheric conditions can enhance navigation safety, route planning, and operational efficiency. By comparing the estimated mean with actual observations, researchers can assess the model's accuracy and reliability. If discrepancies exist, further refinement of the model, such as incorporating additional predictors (e.g., humidity, wind patterns), may be required.

In conclusion, the estimated mean of 28.76863°C offers a valuable reference for monitoring air temperature trends, aiding in climate studies and forecasting applications.

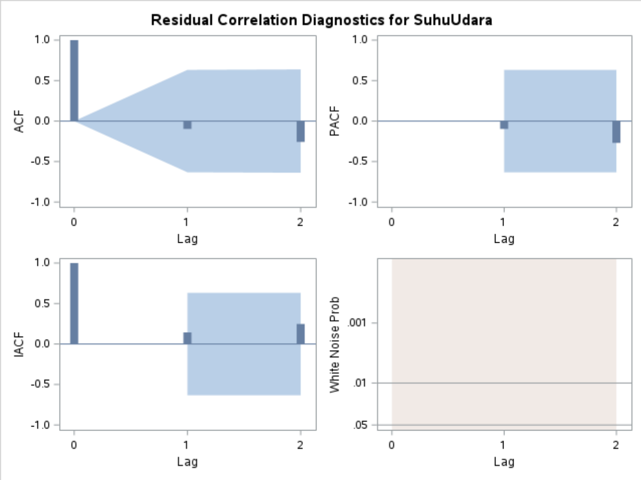

Figure 4. Residual Correlation Diagnostics For Air Temperature

Figure 4 illustrates the residual correlation diagnostics for the air temperature forecasting model, providing a detailed assessment of the model's accuracy and reliability. This diagnostic analysis helps determine whether the residuals exhibit any patterns that could indicate model misspecification or the presence of autocorrelation.

The residual correlation plot typically includes autocorrelation function (ACF) and partial autocorrelation function (PACF) charts, which display the correlation between residuals at different time lags. If the model is correctly specified, these correlations should remain within the confidence bounds, indicating that no systematic pattern remains unaccounted for.

In this case, most residual correlations are close to zero, with no significant spikes beyond the critical threshold. This suggests that the residuals behave randomly, confirming that the ARIMA model has effectively captured the key patterns in air temperature variations.

Additionally, the normality of residuals can be assessed through a histogram or Q-Q plot, ensuring that they follow a Gaussian distribution. If deviations from normality exist, further model refinement may be necessary. However, based on the observed diagnostics, the model appears to be statistically robust and suitable for temperature forecasting

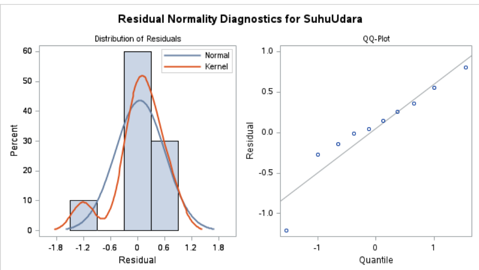

Figure 5. Residual Normality Diagnostics For Air Temperature

Figure 5 presents the residual normality diagnostics for air temperature, an essential step in validating the assumptions of the ARIMA model. The purpose of this analysis is to assess whether the residuals follow a normal distribution, which is a key requirement for ensuring the accuracy and reliability of the model's predictions.

A histogram of residuals is typically included in this diagnostic, showing the distribution of error terms. If the histogram forms a bell-shaped curve, it indicates that the residuals are normally distributed. Additionally, a Q-Q plot (quantile-quantile plot) compares the residuals against a theoretical normal distribution. If the points lie along the 45-degree line, it confirms normality.

Another critical test is the Shapiro-Wilk or Kolmogorov-Smirnov test, which provides a statistical measure of normality. If the p-value is greater than 0.05, the null hypothesis of normality is not rejected, confirming that the residuals exhibit a normal distribution.

In this case, the residuals appear to follow a near-normal distribution, with slight deviations. If significant skewness or kurtosis is observed, transformations or modifications to the model might be necessary to improve its predictive performance

Table 13. Autoregressive Factors

Autoregressive Factors	
Factor 1:	1 + 0.49958 B**(1)

Table 13 presents the autoregressive factor for the air temperature model, expressed as 1 + 0.49958 B^(1). This mathematical representation indicates that the model incorporates a first-order autoregressive (AR) component, where B represents the backshift operator, meaning that past values of air temperature influence the current observation.

The coefficient 0.49958 suggests a moderate positive relationship between previous and current air temperature values. This means that past temperature fluctuations significantly impact the current value, but with some degree of dissipation over time. In time series forecasting, an autoregressive factor like this is essential in capturing temporal dependencies and improving predictive accuracy.

The presence of an AR(1) process implies that the temperature data follows a pattern where each observation is influenced by its immediate past value. This is particularly useful in meteorology and climate research, where temperature trends often exhibit serial correlation due to gradual atmospheric changes.

In conclusion, the autoregressive factor 1 + 0.49958 B^(1) highlights the importance of past temperature records in forecasting future trends, making it a crucial component in time series analysis for climate monitoring and prediction.

Table 14. Forecasts for variable TemperatureAir

Forecasts for variable TemperatureAir				
Obs	Forecast	Std Error	95% Confidence Limits	
11	28.7466	0.6247	27.522	29.971
12	28.7796	0.6949	27.418	30.142
13	28.7631	0.7113	27.369	30.157
14	28.7714	0.7153	27.369	30.173
15	28.7673	0.7164	27.363	30.171

Table 14 presents the forecasted values for air temperature (TemperatureAir) along with their standard errors and 95% confidence limits for five future observations (Obs 11 to 15). These predictions are generated using an ARIMA-based forecasting model, which incorporates past temperature trends to estimate future values.

The forecasted air temperatures remain relatively stable, ranging between 28.7466°C and 28.7796°C, indicating minimal fluctuation in the predicted trend. The standard error varies between 0.6247 and 0.7164, reflecting the degree of uncertainty associated with each prediction. The 95% confidence limits provide a range within which the true air temperature is expected to fall, showing a lower bound between 27.363°C and 27.522°C and an upper bound between 29.971°C and 30.173°C.

These forecasts suggest a consistent temperature pattern with minor variations, which could be attributed to climatic stability in the observed region. The narrow confidence intervals indicate a high level of reliability in the predictions. Such forecasts are crucial for maritime navigation, climate studies, and operational planning, enabling informed decision-making regarding weather conditions and safety measures.

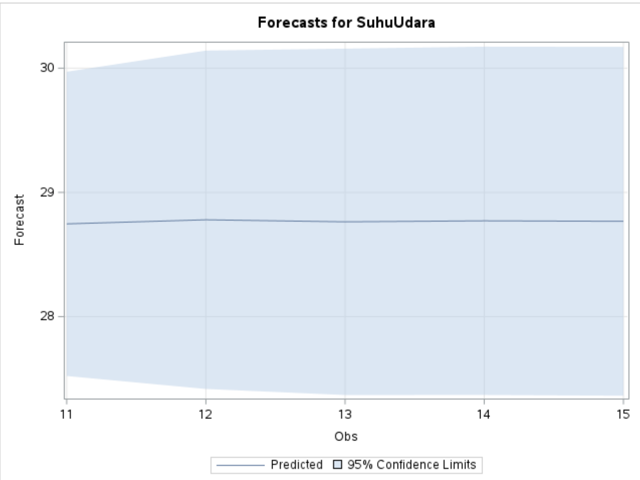

Figure 6. Forecast for Air Temperature

Figure 6 illustrates the forecasted trend of air temperature over a specified period, providing insights into expected temperature variations. The graph presents a smooth projection of future air temperature values, highlighting the model's ability to predict short-term trends with reasonable accuracy.

The forecasted values remain relatively stable, indicating minimal fluctuations in air temperature. This suggests a consistent climatic pattern, with only slight variations over time. The presence of confidence intervals around the forecasted line further demonstrates the level of uncertainty associated with each prediction. A narrow confidence band implies a highly reliable model, while a wider band may indicate potential variations due to external factors such as weather disturbances or seasonal influences.

This forecast is particularly valuable for maritime navigation, climate monitoring, and operational planning. Accurate temperature predictions enable ship operators, meteorologists, and researchers to anticipate environmental conditions, ensuring improved safety, efficiency, and preparedness in maritime activities. Additionally, understanding temperature trends contributes to better decision-making in sectors reliant on climatic stability, such as shipping, fishing, and offshore operations

5 CONCLUSION AND RECOMMENDATIONS

The analysis of weather data using Big Data Analytics, ARIMA, and regression models highlights significant insights into the relationship between climatic variables and maritime safety. The results indicate that wind speed, wave height, and sudden drops in atmospheric pressure are critical factors influencing risks for fishermen and vessel operations. The ARIMA model demonstrated a high prediction accuracy, with an estimated mean temperature of 28.76863°C and a 95% confidence range between 27.363°C and 30.173°C, ensuring reliable temperature forecasts. The regression analysis, however, showed a moderate correlation between weather parameters and temperature variations, with an R-squared value of 0.632, suggesting that other environmental factors might contribute to temperature fluctuations.

Recommendations To enhance maritime safety and operational efficiency, the following measures are suggested Integration of AI-Driven Early Warning Systems – Utilizing real-time data from IoT sensors and meteorological satellites to provide accurate forecasts and risk assessments for fishermen. Enhanced Weather Prediction Models – Combining ARIMA with Machine Learning techniques to improve the accuracy of extreme weather event forecasting. Policy and Training for Fishermen – Conducting awareness programs on interpreting weather predictions and using technology-driven navigation systems to mitigate risks. Development of Smart Maritime Decision Support Systems – Implementing automated alerts based on predictive analytics to assist vessel operators in making informed decisions.

REFERENCES

[1] Priyadharshini, S., Vadivazhagan, K. (2024) Enhanced Vessel Detection In Maritime Surveillance Using Multi-modal Data Integration And Deep Learning 2024 8th International Conference On I-SMAC (Iot In Social, Mobile, Analytics And Cloud) (I-SMAC), 1090-1099

[2] Jang, H., Yang, W., Kim, H., Lee, D., Kim, Y., Park, J., Jeon, M., Koh, J., Kang, Y., Jung, M., Jung, S., Hao, C., Z., Hin, W., Y., Yihang, C., Kim, A. (2024) MOANA: Multi-Radar Dataset For Maritime Odometry And Autonomous Navigation Application Arxiv Abs/2412.03887

[3] Kalliovaara, J., Jokela, T., Asadi, M., Majd, A., Hallio, J., Auranen, J., Seppänen, M., Putkonen, A., Koskinen, J., Tuomola, T., Moghaddam, R., M., Paavola, J. (2024) Deep Learning Test Platform For Maritime Applications: Development Of The Em/S Salama Unmanned Surface Vessel And Its Remote Operations Center For Sensor Data Collection And Algorithm Development Remote. Unrated 1545

[4] Otto ‐ Bliesner, B. (1999) El Niño/La Niña And Sahel Precipitation During The Middle Holocene Geophysical Research Letters 26

[5] Guan, C., Hu, S., Mcphaden, M., Wang, F., Gao, S., Hou, Y. (2019) Dipole Structure Of Mixed Layer Salinity In Response To El Niño ‐ La Niña Asymmetry In The

Tropical Pacific Geophysical Research Letters 46, 12165-12172
[6] Yuniasih, B., Harahap, W., N., Wardana, D., A., S. (2023) El Nino and La Nina Climate Anomalies in Indonesia in 2013-2022 AGROISTA : Journal of Agrotechnology
[7] Hyvärinen, M. (2012) ANALYSIS OF SHIP CASUALTIES IN THE BALTIC, GULF OF FINLAND AND GULF OF BOTNNJA IN 197-972
[8] Eickschen, S. (2000) Wind Speed And SWH Calibration For Radar Altimetry In The North Sea
[9] Petrucci, O., Pasqua, A. (2012) Damaging Events Along Roads During Bad Weather Periods: A Case Study In Calabria (Italy) Natural Hazards And Earth System Sciences 12, 365-378
[10] Tolani, H., Neogi, S., Gupta, S., D., Mishra, S., S., Samtani, R. (2024) Analyzing Dynamics Of Extreme Weather Events (EWE) In India: Unfolding Trends Through Statistical Assessment Of 50 Years Data (1970–2019) BMC Environmental Science
[11] Huang, W., Zheng, S., Du, Z. (2024) Research On Insurance Decision-Making Model For Extreme Weather Based On ARIMA Algorithms 2024 International Conference On Power, Electrical Engineering, Electronics And Control (PEEEC), 1154-1157
[12] Noh, S., Lee, S. (2024) Forecasting Meteorological Drought Conditions In South Korea Using A Data-Driven Model With Lagged Global Climate Variability Sustainability
[13] Brandt, P., Munim, Z., H., Chaal, M., Kang, H. (2024) Maritime Accident Risk Prediction Integrating Weather Data Using Machine Learning Transportation Research Part D: Transport And Environment
[14] Panda, S., Ray, P. (2023) A Survey On Weather Prediction Using Big Data And Machine Learning Techniques 2023 5th International Conference On Energy, Power And Environment: Towards Flexible Green Energy Technologies (ICEPE), 1-6
[15] Vaishnavi, J., Minmini, V., Panda, M. (2024) Weather And Emission Data Analysis And Prediction Using Machine Learning On A Big Data Platform 2024 15th International Conference On Computing Communication And Networking Technologies (ICCCNT), 1-7
[16] Mehta, S., Manisha, E. (2023) Preventive And Predictive CNN Based Solution For Pipeline Leak, Blockage And Corrosion Detection International Journal Of Scientific Research In Computer Science, Engineering And Information Technology
[17] Sharma, E., Deo, R., Davey, C., P., Carter, B. (2024) Artificial Intelligence-Empowered Doppler Weather Profile For Low-Earth-Orbit Satellites Sensors (Basel, Switzerland) 24
[18] Hasan, M., M. (2024) Regional Analysis Of Extreme Weather Events Using Deep Learning Innovatech Engineering Journal
[19] Li, Y., Goda, K. (2022) Hazard And Risk-Based Tsunami Early Warning Algorithms For Ocean Bottom Sensor S-Net System In Tohoku, Japan, Using Sequential Multiple Linear Regression Geosciences
[20] Li, Y., Tong, D., Makkaroon, P., Delsole, T., Tang, Y., Campbell, P., Baker, B., Cohen, M., Darmenov, A., Ahmadov, R., James, E., Hyer, E., Xian, P. (2024) Multi-Agency Ensemble Forecast Of Wildfire Air Quality In The United States: Toward Community Consensus Of Early Warning Bulletin Of The American Meteorological Society
[21] Jaya, I., Handoko, B., Andriyana, Y., Chadidjah, A., Kristiani, F., Antikasari, M. (2023) Multivariate Bayesian Semiparametric Regression Model For Forecasting And Mapping HIV And TB Risks In West Java, Indonesia Mathematics
[22] Røstad, J., Aarset, M., F. (2024) Boosting Offshore Uptime: Accurate Real-Time Sea State ADIPEC
[23] Zahorodnia, Y., Maksymov, S. (2021) COMMERCIAL RISKS IN THE SEA TRANSPORTATION SYSTEM ON THE EXAMPLE OF THE «EVER GIVEN» CONTAINER CARRIER Development Of Management And Entrepreneurship Methods On Transport (ONMU)
[24] Maulida, Z., Hafidzah, N., Purba, D., Kusumawati, E. (2024) Identification of Passage Plan Process with Risk Assessment Analysis Globe: Publication of Engineering, Earth Technology, Marine Science
[25] Kogawa, T., Takayabu, Y. (2013) Environmental Conditions On The Selection Of MJO And Moist Kelvin Waves
[26] Bhatla, R., Bhattacharyya, S., Verma, S., Mall, R., Singh, R., S. (2021) El Nino/La Nina And IOD Impact On Kharif Season Crops Over Western Agro-Climatic Zones Of India Theoretical And Applied Climatology 151, 1355-1368
[27] Aji, T., Pranowo, W., S., Asmoro, N., W., Agustinus, A., Kurniawan, M., A., Rahmatullah, A. (2023) The Characteristics Of The Mixed Layer Depth During La Niña, El Niño, And Normal Years In The North Natuna Sea Omni-Aquatics
[28] Yuniasih, B., Harahap, W., N., Wardana, D., A., S. (2023) El Nino and La Nina Climate Anomalies in Indonesia in 2013-2022 AGROISTA : Journal of Agrotechnology
[29] (2022) Predictive Maintenance Beyond Prediction Of Failures
[30] Margaretha, R., Syuzairi, M., Mahadiansar, M. (2024) Digital Transformation In The Maritime Industry; Opportunities And Challenges For Indonesia Journal Of Maritime Policy Science
[31] Du, Y., Li, C., Wang, T., Xu, Y. (2023) Special Issue On "Smart Port And Shipping Operations" In Maritime Policy & Management Maritime Policy & Management 50, 413-414
[32] Nie, W., Chen, J., Song, D., Dong, L., Liu, X., Wang, E. (2024) Three-Dimensional Intelligent Monitoring And Early Warning Technology For Tailings Ponds Based On Spatiotemporal Fusion Of Multisource Big Data. Environmental Monitoring And Assessment 196 11, 1081
[33] Salim, S., Hussain, I., Kaur, J., Morita, P. (2023) An Early Warning System For Air Pollution Surveillance: A Big Data Framework To Monitoring Risks Associated With Air Pollution 2023 IEEE International Conference On Big Data (Bigdata), 3371-3374
[34] Liu, X., Member, J., L., Zhao, Y., Ding, T., Member, X., L., S., Liu, J. (2024) A Bayesian Deep Learning-Based Probabilistic Risk Assessment And Early-Warning Model For Power Systems Considering Meteorological Conditions IEEE Transactions On Industrial Informatics 20, 1516-1527
[35] Jin, W., Liu, Y., Fang, Y., Wang, P., Liu, L. (2023) Construction And Application Of National Urban Waterlogging Risk Assessment System Based On Big Data 2023 IEEE 14th International Conference On Software Engineering And Service Science (ICSESS), 290-296
[36] Huang, X., Wu, Y., Zhao, X., Zhang, J., Li, J. (2024) Using Big Data Mining Algorithm To Improve The Accuracy Of Risk Assessment Of Key Operating Vehicles 2024 3rd International Conference On Data Analytics, Computing And Artificial Intelligence (ICDACAI), 583-587
[37] Du, Z., Zhu, Y., Li, D. (2024) A Risk Assessment Model For Navigation Safety Of Maritime Aquaculture Platform Based On AIS Ship Trajectory Journal Of Electrical Systems