Development of Inter-Island Sailing Safety System Through Machine Learning with IoT-Based Data Integration

Alifya aura balqis[1], Nanta Fakih Prebianto[1]
[1]Electronics Engineering Technology Study Program, Department of Electrical Engineering, Politeknik Negeri Batam, Batam, Indonesia
Email: 4242201082@mahasiswa.polibatam.ac.id

Abstract—Unpredictable weather phenomena pose significant risks to maritime vessels navigating inter-island waterways near Belakang Padang, Batam, particularly for smaller craft. While meteorological services provide regional forecasts, these macroscopic datasets frequently omit localized anomalies critical to navigation safety at particular waypoints. Addressing this information gap, the research proposes an IoT framework enabling continuous, hyper-localized weather parameter measurement from two geographically-separated monitoring stations.

The system architecture comprises two separate units: an onboard Node mounted on the vessel and a shoreside Station. These components communicate using Long Range (LoRa) technology and implement the K-Nearest Neighbor (KNN) classification method to assess sailing feasibility. The Node subsystem focuses on vessel dynamics by measuring tilt angles, while the Station subsystem concurrently captures environmental data specifically temperature, humidity, wind velocity, and precipitation which are displayed on a real-time web interface. Testing confirmed that LoRa communication succeeded across distances reaching 8.67 km. Linear regression calibration of the temperature sensor yielded a high-fidelity model ($R^2 = 0.9998$), reducing measurement error from an initial 4.04% to 2.88%. Using K = 7 and Manhattan distance weighting, the KNN model achieved 98.90% classification accuracy with an average 10-fold cross-validation score of 96.72%. These results demonstrate the platform's capability to monitor maritime weather patterns continuously and categorize seaworthiness with high precision, thereby offering substantial value for reducing shipping incidents.

Keywords: IoT, LoRa, K-Nearest Neighbor, maritime weather monitoring, seaworthiness.

I. INTRODUCTION

Among Indonesia's cities, Batam stands out for its unique geography: approximately 73% of the municipal area comprises water bodies, making it a maritime hub, which places maritime and shipping activities at the heart of the region's connectivity and economy [1] One of its districts, Belakang Padang, is an archipelagic area where sea transport is essential to inter-island mobility. Its shipping lanes carry community travel, fishing trips, and inter-island transport, and in planning any voyage the direction of the current and the height of the waves cannot be ignored [2]

Because activity on these waters depends heavily on weather and oceanographic factors — wind speed and direction, rainfall, wave height, currents, and tides — the swift and hard-to-predict shifts in maritime weather rank among the main triggers of shipping accidents, above all for smaller vessels [3]. The weather information available today tends to be regional and does not always mirror what is happening at a particular navigation point, so a local, real-time monitoring system becomes necessary. On top of that, coverage from mobile networks is frequently limited over the water, which calls for a low-power communication technology that can still reach across long distances[4].

A number of earlier studies have produced IoT-based weather monitoring systems. built a power-efficient LoRa weather station able to send data as far as 10 km [5]. Reddy created a comparable system offering low power draw and broad coverage without a cellular network [6]. developed an IoT weather monitoring system driven by the Naïve Bayes algorithm, reaching 92.49% accuracy, though it was confined to a single land-based site [7]. meanwhile, demonstrated that the K-Nearest Neighbor (KNN) algorithm can classify weather conditions at high accuracy (97%) [8]. evertheless, the preponderance of existing systems remain architecturally constrained to singular geographic nodes and have not yet incorporated seaworthiness classification capabilities into their operational paradigm.

To address these limitations, the present investigation engineered a distributed IoT architecture comprising two geographically-separated monitoring endpoints: one mounted aboard maritime vessels and another stationed on terra firma. Communication between nodes leverages Long Range (LoRa) wireless technology to facilitate data exchange across extended distances. The system applies the KNN algorithm to classify sailing feasibility from weather sensor data. The goal is to design a maritime weather monitoring system capable of measuring weather parameters in real time, transmitting data over long distances via LoRa, and delivering a sailing feasibility classification that supports the mitigation of shipping-accident risk.

II. METHOD

A. General System Design

The research foundational architecture constitutes an Internet-of-Things infrastructure tailored for maritime meteorological assessment and real-time environmental data acquisition. The platform employs a distributed, bifurcated deployment topology: a seaborne data-collection subsystem (Node) positioned aboard ocean-going vessels functioning in tandem with a terrestrial reception and processing module (Station) located shoreside. Inter-unit communications leverage Long Range (LoRa) radio frequency technology—selected explicitly to overcome the ubiquitous constraint of inadequate or completely absent cellular network coverage in maritime zones. The maritime-deployed Node element carries dual operational functions: measurement and quantification of vessel dynamics (roll, pitch, list) coupled with continuous acquisition of positional coordinates. The computational subsystem of the Node component centers on a LoRa Ray V3 microcontroller development board, incorporating an ATmega processor architecture and an integrated 915 MHz LoRa transceiver for wireless telecommunication. The inertial measurement unit (MPU6050) performs continuous sampling of vessel orientation across multiple axes; simultaneously, a positioning module (GPS) captures and transmits vessel geolocation data. User feedback and operational status indication derive from an onboard I2C-driven liquid crystal display, acoustic alarm signaling apparatus, and a tri-color LED indicator panel (green, yellow, red hues denoting Safe, Caution, and Dangerous operational states respectively). Following acquisition, all telemetry originating from the Node subsystem undergoes transmission to the shoreside Station facility through the LoRa communication pathway.The Station facility, sited on terrestrial ground, functions as the data reception and aggregation nexus for the integrated system architecture. This shoreside component likewise utilizes a LoRa Ray V3 development platform as its computational foundation; however, it interfaces with a dedicated weather instrumentation array: specifically, a thermal-humidity sensor (DHT22) for environmental parameter sampling, a wind velocity transducer (anemometer), and precipitation measurement apparatus (rain gauge and rain drop detection sensors). Concurrent with the reception of maritime telemetry via LoRa telecommunications, the Station apparatus independently logs environmental parameters specific to its geolocation.

The aggregated measurement dataset accumulated by the Station facility undergoes transmission to a microcontroller gateway module (ESP32) serving as the data ingestion portal. Subsequently, this telemetry is conveyed to a cloud-based database infrastructure (Firebase) in near-real-time fashion and concurrently rendered as interactive visualizations on a web-baseduser dashboard. Upon reception at the processing subsystem, the meteorological parameters undergo algorithmic analysis via the K-Nearest Neighbor (KNN) classification methodology, which systematically categorizes the observed sailing conditions according to three operationally-distinct classifications: Safe, Caution, and Dangerous.

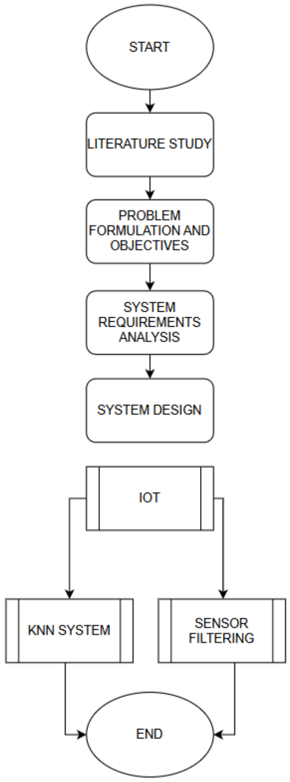


Figure 1. System Flowchart

B. Hardware Design

The hardware is arranged into two independent circuits — one for the Node (vessel) and one for the Station (shore). Both rely on the LoRa Ray V3 board, whose ATmega core and built-in 915 MHz LoRa module remove the need for any separate LoRa transceiver.

1. Node Hardware (Ship)

In the Node circuit, the LoRa Ray V3 board is wired to an MPU6050 sensor over I2C to capture the vessel's roll and pitch, together with a GPS module that reports the ship's location through a serial link. The board also drives an I2C LCD for on-board readouts, a buzzer for audible alerts, and three status LEDs (green, yellow, red) that mark the Safe, Caution, and Dangerous conditions. Power is supplied from a 12V source that is regulated by an LM2596 buck converter. After processing the tilt and position readings, the board forwards them to the Station through its built-in LoRa module. The Node components are listed in Table 1, and the circuit is shown in Figure 1.

Table 1. Node Hardware Components

Component	Function	Output
LoRa Ray V3 (ATmega + LoRa)	Controller + LoRa communication	Data processing & transmission
MPU6050	Measures vessel tilt (roll & pitch)	Tilt angle (°)
GPS (GY-GPS7MV3)	Determines vessel position	Coordinates
LCD I2C	Displays data on the vessel	Local display
Buzzer	Audible alert	Alarm
LEDindicator (green/yellow/red)	Condition status indicator	Visuals Safe/Alert/Danger
Buck converter LM2596	Stabilizes 12V supply	Regulated power

The Node circuit schematic is shown in Figure 1.

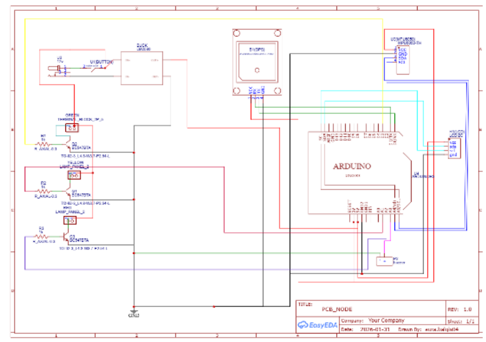


Figure 2. *Node Circuit Diagram*

The board analyzes the tilt and positional data collected by the sensors and then transmits it to the Station via the included LoRa module.

2. Station Hardware (Shore)

In the Station circuit, the LoRa Ray V3 board is connected to a set of weather sensors: a DHT22 for reading temperature and humidity, an anemometer for wind speed, and rain gauge and rain drop sensors for detecting and measuring rainfall. Apart from logging the local weather at its own site, the board also takes in the data sent from the Node over LoRa. All of the collected information is then passed to an ESP32 module, which acts as the internet gateway, uploading the data to the Firebase database in real time so that it can be shown on the web dashboard. The Station components are listed in Table 2.

Table 2. Station Hardware Components

Component	Function	Output
LoRa Ray V3	Controller + LoRa receiver	Data processing & reception
ESP32	Internet gateway (WiFi)	Data to Firebase
DHT22	Measures temperature & humidity	°C & %RH
Anemometer	Measures wind speed	m/s
Rain gauge	Measures rainfall	mm/h
Rain drop	Detects rainfall	Wet/dry

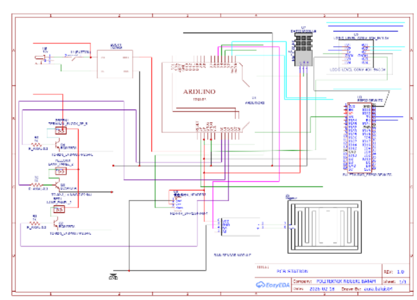


Figure 3. . Station Circuit Schematic

C. K-Nearest Neighbor (KNN) Classification

The weather readings that have been gathered — wind speed, wave height (tilt), temperature, and rainfall — are assembled into feature vectors that act as the input for the K-Nearest Neighbor (KNN) algorithm [9] . Based on these features, the model assigns each sailing condition to one of three categories: Safe, Caution, or Dangerous. The best parameter combination was obtained through Grid Search paired with cross-validation, which settled on K = 7, distance-based weighting, and the Manhattan metric [10]. Working from these features, the model places every sailing condition into one of three groups: Safe, Caution, or Dangerous. To find the strongest parameter set, Grid Search was run together with cross-validation, and the search converged on K = 7 with distance weighting and the Manhattan metric . Building the model involved a sequence of steps — collecting the data, evening out the classes, separating it into

NO	Conditions	DHT22(°C)	Blue Gizmo (°C)	Differences	%Error	Accuracy
1	Air conditioning room	25,2	24,3	0,9	3,70%	96,30%
2	Regular room	29,5	28,4	1,1	3,87%	96,13%
3	Outdoor area	34,5	33,7	0,8	2,37%	97,63%
4	Hot machine	45,7	45,0	0,7	1,56%	98,44%
	Average				**2,88%**	97, 12%

training and testing portions at an 80:20 ratio, carrying out the training, and rounding off with an evaluation via accuracy, precision, recall, and a confusion matrix.The flow of the KNN classification process is presented in Figure 3.

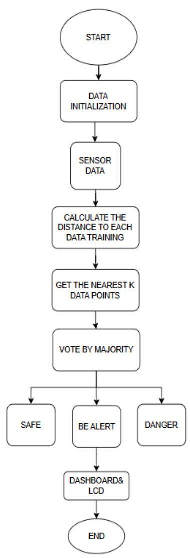


Figure 4. KNN Classification Flowchart

D. Sensor Testing and Calibration

Sensor testing was performed by placing each sensor's output side by side with that of a reference instrument. For the DHT22 temperature sensor, calibration was carried out through linear regression against the reference to improve its accuracy [11], while the readings themselves were smoothed with a Moving Average Filter (MAF) to dampen fluctuations [12]. From this comparison, the error and accuracy of every sensor were then determined. The testing for each sensor moved through three phases: collecting the raw data, applying linearization, and finally implementing the MAF.

E. Testing and Evaluation

Evaluation of the system was organized into three phases. The first phase focused on sensor accuracy, obtained by weighing the sensor readings against reference measuring instruments to establish the error of each one. The second phase examined the LoRa link, measuring how far data could be transmitted based on distance and signal quality (RSSI). The third phase assessed the KNN classification model in terms of accuracy, precision, recall, and 10-fold cross-validation.

III. RESULTS AND DISCUSSION

A. LoRa Communication Range Testing.

LoRa communication assessments were performed to evaluate the data transmission distance between the Node and the Station. The examination outcomes indicate that the system successfully transmitted data over a minimum distance of 8.67 km, with received signal strength indicator (RSSI) values varying from −54 dBm at short range to −113 dBm at the furthest distance. The findings suggest that the integrated LoRa module on the LoRa Ray V3 board can facilitate long-range data transfer, rendering it appropriate for maritime regions with sparse cellular network availability.

B. Temperature Sensor Calibration (DHT22)

The DHT22 temperature sensor was tested by comparing its readings against a reference instrument. The raw readings produced an average error of 4.04%. To improve accuracy, linearization was applied by determining a linear regression equation, resulting in $y = 1.056x - 3.094$ with a coefficient of determination $R^2 = 0.9998$. After applying the linearization, the average error decreased to 2.88%, indicating that the calibration successfully improved the accuracy of the temperature readings.

Table 3. DHT22 Calibration Results

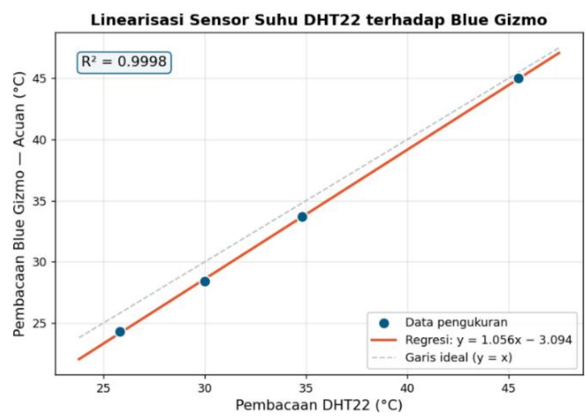
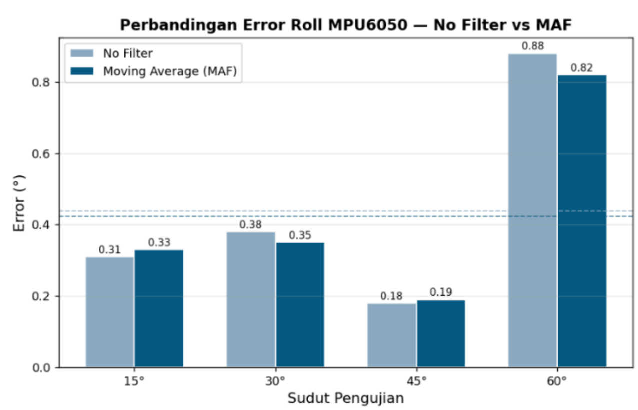
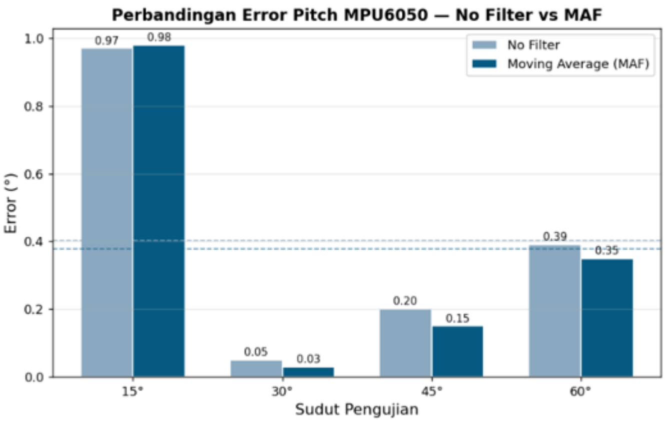


Figure 5. DHT22 Linearization Graph

C. Tilt Sensor Testing (MPU6050)

The MPU6050 sensor was tested at several tilt angles for both roll and pitch axes, comparing readings before and after applying the Moving Average Filter (MAF). For the roll axis, the average error was 0.44° without filtering and 0.42° with MAF. For the pitch axis, the average error was 0.40° without filtering and 0.38° with MAF. Both axes produced errors below 1°, indicating that the tilt sensor provides accurate and stable readings for use as an indicator of wave conditions.





D. KNN Classification Results

The KNN model was tested on a dataset of 454 samples with an 80:20 training-testing split and 10-fold cross-validation. Using Grid Search, the optimal parameters were obtained as K = 7, distance weighting, and the Manhattan metric. The model achieved an accuracy of 98.90% with an average cross-validation score of 96.72% (±3.26%). The small difference between the testing accuracy and the cross-validation score (2.18%) indicates that the model is stable and does not overfit.Based on the confusion matrix, the model correctly classified 90 out of 91 test data. The Safe class achieved 97.83% accuracy (45 of 46 data), while the Caution and Dangerous classes were classified perfectly (100%). The misclassification that occurred was conservative in nature — one Safe sample was classified as Caution — and no critical error occurred in which a Dangerous condition was classified as Safe. This indicates that the model is reliable for supporting sailing safety decisions.

Table 4. KNN Classification Report

Class/Status	Precision	Recall	F1-Score	Number of Test Samples (*Support*)
Secure	1,00	0,98	0,99	46
Alert	0,96	1,00	0,98	24
Dangerous	1,00	1,00	1,00	21
Accuracy	-	-	0,99	91
Macro Avg	0,99	0,99	0,99	91
Weighted Avg	0,99	0,99	0,99	91

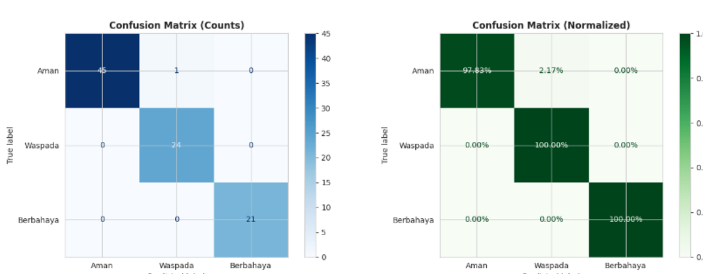


Figure 6. Confusion Matrix

E. Discussion

The high classification accuracy (98.90%) is consistent with the characteristics of the dataset, in which the condition labels are determined based on non-overlapping wave height ranges, making the class boundaries clearly distinguishable by the algorithm. This represents a limitation of the study, as the model was tested on data with clear class separation. Compared to previous studies, the KNN algorithm has been shown to classify weather conditions with high accuracy [8], which is consistent with the results obtained in this study. For more challenging testing, the use of data with transitional (borderline) conditions between categories is recommended.

IV. CONCLUSION

This study successfully developed an Internet of Things (IoT)-based maritime weather monitoring system with a two-node configuration — the Node (vessel) and the Station (shore) — connected through Long Range (LoRa) communication, and applied the K-Nearest Neighbor (KNN) algorithm to classify sailing conditions. According to the conducted tests, the system successfully transmitted data over a minimum distance of 8.67 km, indicating that the integrated LoRa module is appropriate for maritime regions with sparse cellular network availability. The linearization of the DHT22 temperature sensor's calibration yielded a regression equation with a determination coefficient $R^2 = 0.9998$, decreasing the mean error from 4.04% to 2.88%. The KNN classification model, with an optimal K value of 7, achieved an accuracy of 98.90% with an average 10-fold cross-validation score of 96.72%, and correctly classified 90 out of 91 test data with a conservative misclassification pattern.

These results indicate that the developed system is able to monitor maritime weather conditions in real time and classify sailing feasibility with high accuracy, thereby having the potential to support the mitigation of sailing accident risks in inter-island shipping. For future development, it is recommended to add data on transitional (borderline) conditions between classes to test the model under more challenging scenarios, as well as to test the wind sensor using a stable airflow source (wind tunnel) to improve its measurement accuracy.

REFERENCES

[1] Badan Pusat Statistik Kota Batam, “Kota Batam Dalam Angka 2024,” Batam, 2024.

[2] A. Rawson, M. Brito, Z. Sabeur, and L. Tran-Thanh, “A machine learning approach for monitoring ship safety in extreme weather events,” *Saf. Sci.*, vol. 141, p. 105336, Sep. 2021, doi: 10.1016/j.ssci.2021.105336.

[3] P. Brandt, Z. H. Munim, M. Chaal, and H. S. Kang, “Maritime accident risk prediction integrating weather data using machine learning,” *Transp. Res. D Transp. Environ.*, vol. 136, Nov. 2024, doi: 10.1016/j.trd.2024.104388.

[4] S. Pensieri *et al.*, “Evaluating LoRaWAN Connectivity in a Marine Scenario,” *J. Mar. Sci. Eng.*, vol. 9, no. 11, p. 1218, Nov. 2021, doi: 10.3390/jmse9111218.

[5] E. Murdyantoro, R. Setiawan, I. Rosyadi, A. W. W. Nugraha, H. Susilawati, and Y. Ramadhani, “Prototype weather station uses LoRa wireless connectivity infrastructure,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Nov. 2019. doi: 10.1088/1742-6596/1367/1/012089.

[6] K. V. Reddy, “LORA Based Wireless Weather Station,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 10, no. 6, pp. 2556–2561, Jun. 2022, doi: 10.22214/ijraset.2022.44412.

[7] A. D. Phang, L. Sitanayah, and J. B. Sanger, “Sistem Pemantauan Cuaca Berbasis Internet of Things,” JOINTER-Journal of Informatics, vol. 05, no. 02. pp. 22–29, 2024.

[8] R. Verian, D. Saputra, I. Widaningrum, J. Karaman, and S. L. Tsaqila, “Rezano Verian Dwi Saputra, Ida Widaningrum, Jamilah Karaman, Siti Lathifah Tsaqila. Penggunaan Algoritma K-Nearest Neighbor (K-NN) Untuk Prediksi Cuaca Dengan Data Record Penggunaan algoritma K-Nearest Neighbor (K-NN) untuk prediksi cuaca dengan data record,” Multitek Indonesia: Jurnal Ilmiah, no. 2, pp. 1907–6223, 2024, [Online]. Available: http://journal.umpo.ac.id/index.php/multitek.

[9] S. Zhang, “Challenges in KNN Classification,” *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 10, pp. 4663–4675, Oct. 2022, doi: 10.1109/TKDE.2021.3049250.

[10] Z. Tarek *et al.*, “Wind Power Prediction Based on Machine Learning and Deep Learning Models,” *Computers, Materials & Continua*, vol. 74, no. 1, pp. 715–732, 2023, doi: 10.32604/cmc.2023.032533.

[11] S. Sugeng, T. N. Nizar, D. A. Jatmiko, R. Hartono, and Y. Y. Kerlooza, “Kalibrasi Sensor Monitoring Cuaca pada Area Lokal untuk Meningkatkan Akurasi pada Sensor Biaya Rendah,” *Komputika : Jurnal Sistem Komputer*, vol. 13, no. 2, pp. 277–287, Oct. 2024, doi: 10.34010/komputika.v13i2.13949.

[12] F. Baskoro, B. I. G. P. Asto, M. Rohman, F. D. A. Soeadyfa, and A. P. Nurdiansyah, “Impact of Sample Size Variation on Moving Average Filter Performance for Stability and Accuracy in Ultrasonic Sensor Measurements,” *TEM Journal*, vol. 14, no. 2, pp. 1681–1688, 2025, doi: 10.18421/TEM142-65.