MAPÚA MALAYAN COLLEGES MINDANAO

An Internet of Things and Machine Learning Approach of Predicting Hydrological Risk Intensities in Coastal Areas Within Davao City

John Mark P. Bago-od

Franz Dominic S. Brillante

Rey Joshua M. Morales

Bachelor of Science in Computer Science

Thesis Adviser

Dr. Patrick D. Cerna

College of Computer and Information Science

An Internet of Things and Machine Learning Approach of Predicting Hydrological Risk Intensities in Coastal Areas Within Davao City

by

John Mark P. Bago-od
Franz Dominic S. Brillante
Rey Joshua M. Morales

A Thesis Report Submitted to the College of Computer and Information Science
in Partial Fulfillment of the Requirements for the Degree

Bachelor of Science in Computer Science

Mapúa Malayan Colleges Mindanao
Davao City

May 2023

APPROVAL SHEET

The thesis, entitled **"AN INTERNET OF THINGS AND MACHINE LEARNING APPROACH OF PREDICTING HYDROLOGICAL RISK INTENSITIES IN COASTAL AREAS WITHIN DAVAO CITY"** prepared and submitted by **JOHN MARK P. BAGO-OD**, **FRANZ DOMINIC S. BRILLANTE**, and **REY JOSHUA M. MORALES** in partial fulfillment of the requirements for the degree of **BACHELOR OF SCIENCE IN COMPUTER SCIENCE** is hereby accepted.

DR. PATRICK D. CERNA
THESIS ADVISER

Accepted as partial fulfillment of the requirements for the degree of **BACHELOR OF SCIENCE IN COMPUTER SCIENCE**

RHODESSA J. CASCARO, DIT
DEAN

ACKNOWLEDGEMENT

First, the proponents would like to express our deepest thank you to the institution, Mapúa Malayan Colleges Mindanao, for allowing us to conduct our study within the community. Despite the pandemic, the school has provided great support to us, researchers, in completing our thesis manuscript.

Our hearts are overflowing with appreciation for Mr. Neil P. Magloyuan, our thesis facilitator, who set the bar high for us as senior researchers. Despite the hurdles of this endeavor, he went the extra mile to guide and support us at every turn, urging us to push beyond our limits and achieve our best possible outcomes. We couldn't imagine a thesis facilitator for our research project, who is as firm as him. With this, we are forever grateful for his outstanding contributions.

To our dear advisor, Dr. Patrick D. Cerna, we extend our extreme gratitude for guiding us in writing and developing our study. We sincerely appreciate his motivating words as he always believed in our capabilities and was a pillar of support, especially in times we needed it most. This paper would not have been as comprehensive without him.

We also take this opportunity to express a deep sense of gratitude to our family, for their unending love, support, and understanding, not only throughout this research journey but in our day-to-day lives. The completion of this undertaking would not have been possible without the participation and assistance of a lot of people, some of whose names may not have been enumerated here. All contributions are sincerely appreciated and gratefully acknowledged.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ARTICLE 1

ARTICLE 2

LIST OF ABBREVIATIONS

PAGASA	Philippine Atmospheric, Geophysical and Astronomical Services Administration
NAMRIA	National Mapping and Resource Information Authority
MMCM	Mapúa Malayan Colleges Mindanao
IOT	Internet of Things
XG Boost	Extreme Gradient Boosting
API	Application Programming Interface

Article 1

An Internet of Things Approach Of Monitoring Water Levels and Weather Conditions Using Low-Cost Sensors on Coastal Areas Within Davao City

John Mark P. Bago-od[a*], Franz Dominic S. Brillante[a], Rey Joshua M. Morales[a]

[a]College of Computer and Information Science, Mapúa Malayan Colleges Mindanao, Davao City, Philippines 8000

Emails: jmbagood@mcm.edu.ph, fdbrillante@mcm.edu.ph, rjmorales@mcm.edu.ph

*Corresponding author

College of Computer and Information Science

Mapúa Malayan Colleges Mindanao

Gen. Douglas MacArthur Hwy, Talomo

Davao City, Philippines

Email jmabagood@mcm.edu.ph

Abstract

The Internet of Things (IoT) is a rapidly growing technology that connects devices and sensors to the Internet for real-time monitoring and data collection. Monitoring coastal weather conditions is crucial for disaster prevention and management, as extreme weather phenomena have significant impacts on people's lives. However, traditional weather monitoring systems that are currently being used in Davao City are costly and high maintenance. Meanwhile, IoT devices provide a low-cost and efficient alternative for monitoring weather conditions. The researchers investigated an IoT-based approach to monitoring water level, sea level pressure, temperature, relative humidity, and wind speed in coastal areas using cost-efficient sensors which aimed to know the viability of data collection in coastal areas at Davao City. The study focused on the weather features that are within the financial capacity of the researchers which included the sensors' viability utilized in the weather station. The sensors used were BME280 barometric and humidity sensor, a DHT11 temperature sensor, JSNSR04T waterproof ultrasonic sensor, and a custom-made, hall sensor-driven, cup anemometer to implement the prototype.

Results revealed that the IoT system is capable of monitoring Temperature, Pressure, Water Level, Wind Speed, and Humidity. It also showed that the gathered data from the IoT devices only have 0.04 and 0.86 from wind speed and maximum temperature as the lowest mean absolute error whilst the highest mean absolute error rates are only 4.23 and 3.84 from Humidity and Minimum Temperature. This concludes that the IoT-based monitoring system developed by the researchers proved to be a viable solution for monitoring environmental factors in real time. The deployment in Sta. Ana Wharf showed immense potential for integration with weather-sensitive industries and provided valuable insights and details. The researchers recommend better equipment quality, including cheaper sensors and more robust communication devices, and a longer deployment time to gather more data and identify potential challenges.

Keywords: Internet of Things (IoT), coastal weather conditions, real-time monitoring, disaster prevention, traditional weather monitoring systems, Davao City, low-cost, efficient, sensors, water level, sea level pressure, temperature, relative humidity, wind speed, data collection, weather station, BME280 barometric and humidity sensor, DHT11 temperature sensor, JSNSR04T waterproof ultrasonic sensor, hall sensor-driven, prototype, IoT system, mean absolute error, monitoring system, environmental factors, real time, IoT deployment.

1. Introduction

Throughout the years of mankind, the Internet of Things (IoT) has been developing innovation that interfaces gadgets and sensors to the web, taking into consideration continuous observing and information assortment. The utilization of IoT is unending, and one of them is the observing of atmospheric conditions in waterfront regions. Coastal areas are extremely sensitive to weather fluctuations, making monitoring them vital for disaster prevention and management.

According to David Odesola et al. (2019), frequent temperature variations have a significant impact on our modern way of life. Since the world is changing so quickly, weather stations have also changed. As a result, climate and weather fluctuations, for example, have a significant impact on what fruits and vegetables may be produced. Extreme weather phenomena, such as hurricanes, droughts, fires (forest fires), floods, heat waves or cold snaps, and winter storms, are another major part of weather with significant impact on the lives of people. With this study, the researchers aimed to remotely regulate and monitor the temperature of our surroundings through creating a weather station based on IoT approaches for it to be more convenient, low-cost in maintenance, and can be accessed anywhere in the Philippines through an internet.

The term "internet of things" refers to anything connected to the internet, but it is also used to describe objects that "speak" to one another. Simply described, the Internet of Things is a network of networked computers. The Internet of Things (IoT) connects physical devices to the internet, enabling data collection and analytics. This allows consumers to interact with the global network without needing a keyboard or screen, as their devices and equipment can receive commands from the network with minimal human input. In industries, IoT can provide the same efficiencies as the internet has long provided for intelligence work.

With billions of internet-enabled sensors collecting data worldwide, businesses can gather information about their operations, monitor properties, and automate processes. However, there are privacy and security concerns around researchers collecting data on people's interests and behaviors through IoT (IoT Philippines, n.d)

According to Sharma & Prakash (2021) natural environment issues such as typhoons, floods, global warming, and many more causes several struggles in various fields such agriculture, fisheries, wildlife conservations, and many more but the solutions for these struggles are costly and high maintenance thus IoT approaches became known and became a better alternatives to those high end devices because IoT devices are not costly, low maintenance, and can be operate by most people.

Further, IoT devices are less expensive than traditional weather monitoring systems and enable real-time monitoring of weather conditions, allowing for fast reactions in the event of an emergency. In fact, IoT devices can correctly and effectively gather data, offering significant insights into weather patterns and trends. Furthermore, IoT devices are easy to deploy and operate, making them an excellent choice for weather monitoring in remote or difficult-to-access locations.

Specifically, the researchers of this research project sought to answer the following problems:

1. Is an IoT-based approach viable for the data collection on coastal areas in Davao City?
2. Is having a low-cost budget enough for a functional IoT-based weather station for coastal areas within Davao City?
3. How much maintenance does it need for a low-cost IoT-based weather station for coastal areas within Davao City?

The study on an IoT system in monitoring meteorological conditions in Davao City's coastal areas is critical because of the lack of instruments that focus on collecting data directly on these areas. Given the cost constraints associated with acquiring pre-assembled weather stations, their implementation is often hindered. Therefore, incorporating low-cost sensors becomes imperative in facilitating the adoption of such technology. The findings in the device readings can be used as a means to predict an upcoming hydrological risk.

Furthermore, the general objective of this research study was to develop a cost-effective and efficient system for monitoring water levels and weather conditions in coastal areas of Davao City using Internet of Things (IoT) technology. The following are the specific objectives of this article:

1. Identify the environmental factors like water level and environmental monitoring for coastal risk in Davao City.
2. Develop and Deploy an IoT-based system that measures the wind speed, humidity, temperature, water level, and pressure.
3. Tabulate and visualize the measured water level and environmental monitoring.

This study aimed to develop an Internet of Things (IoT)-based system for gathering data on water level and atmospheric variations in the coastal areas of Davao City. The potential beneficiaries of this research include the following individuals or groups:

NAMRIA. If NAMRIA decides to pursue a project relating to this topic, this research study may help the agency. This may offer them some useful information on the incorporation of weather characteristics into water levels.

PAGASA. This study may also help the government agency by providing informative data that can be transmitted to disaster risk reduction agencies to give them insights about other ways of implementing weather instruments, specifically in combining hydrological measurements with meteorological ones in cheap ways.

Nearby Settlements and Businesses. This research study may also benefit the communities residing in coastal regions by having an idea on creating their own instrument inspired from this research, wherein they would know the cheap ways of acquiring these kinds of instruments and have an idea on how to handle it.

Future Researchers. This study may also be beneficial to future researchers to be more inclined and prepared in investigating this topic. This may help them improve with the given recommendations while also being able to pave the way for more research to come as this topic is crucial for the development of advanced weather data collection techniques.

With this in mind, the scope of this IoT implementation encompasses the following aspects:

Sensor Deployment. The research paper focused on strategically placing the low-cost sensors across coastal areas in Davao City. This deployment ensures comprehensive data collection, enabling meaningful analysis of hydrological risk intensities.

Data Acquisition: The IoT system continuously acquires data from the deployed sensors, ensuring a steady stream of real-time information. The data collected includes temperature, humidity, rainfall, water levels, and wind speed, providing a holistic view of the hydrological conditions.

Data Processing and Analysis. The collected data was processed and analyzed to identify patterns and trends related to hydrological risks, which is further discussed in the next article. This includes the effectiveness of the apparatus in gathering data.

This article only covers the Internet-of-things layer of the project, wherein discussed here are the specifications of the device, the deployment methods, data processing, and comparison of results to PAGASA readings.

2. Materials and Methods

Table 1: Hardware Used Descriptions

Features	Description	Mode of Measurement
Water Level	The level assumed by the surface of a particular body or column of water	**JSN-SR04T**
Humidity	Humidity is the amount of water vapor in the air. It is expressed as a percentage of the amount of moisture the air can hold at a given temperature.	**BME280**
Wind Speed	Wind speed is a measure of the speed at which wind is moving through the air. It is typically measured in units of distance per time, such as miles per hour or kilometers per hour.	**Anemometer**

Temperature	Temperature is a measure of the hotness or coldness of an object or a substance. It is typically measured in degrees on a temperature scale, such as Celsius or Fahrenheit.	**BME280**
Pressure	Atmospheric pressure is the force exerted on the Earth's surface by the weight of the air above it. It is typically measured in units of force per unit of area, such as pounds per square inch or pascals.	**BME280**

The table above shows the details of each microcontroller including its purpose and the kind of parameter it detected.

Raspberry Pi. The Raspberry Pi is a collection of single-board computers that may serve a variety of functions, including those of a general-purpose computer, a robot, and even a hub for IoT devices. Any display could be plugged into an RPI. Raspberry Pi 3 was the main component for the IoT weather collecting device. BME280 is a multi-feature sensor that can measure altitude, barometric pressure, and ambient temperature. It is a sensor especially developed for mobile applications and wearables where size and low power consumption are key design parameters. The unit combined high linearity and high accuracy sensors and is perfectly feasible for low current consumption, long-term stability, and high EMC robustness. The sensor has -+1 degrees Celsius of accuracy in temperature and +-Pa in accuracy in pressure.

Anemometer. Anemometer is an instrument that can detect and measure the wind speed and direction of an area. This is commonly used in weather stations. Anemometers helped the researchers to measure the necessary data regarding the wind speed and the wind direction as these are parameters in detecting hydrological risks. The JSN-SR04T is an ultrasonic distance sensor.

JSN-SR0T4. JSN-SR0T4 is a waterproof ultrasonic distance measurement sensor module with a range of 25cm-450cm non-contact distance measurement. This sensor is used for collecting water level data through sound travel time. The DHT11 sensor is designed to measure temperature and relative humidity with precision, while still being affordable. It incorporates an analog-to-digital converter that transforms analog readings of temperature and humidity into digital signals. The sensor's 8-bit microcontroller handles data conversion, and it delivers dependable results when used in conditions with ambient humidity ranging from 20% to 90% RH and temperatures from 0°C to 50°C. Hence, a suitable option for various everyday applications and measurements in homes and offices.

Hall sensor module. A hall sensor module for Arduino is a type of sensor module that is designed to detect the presence of a magnetic field. It uses a Hall Effect sensor, which is a type of transducer that measures the magnetic field strength. This is used in the anemometer where it reads the signal of 3 neodymium magnets attached in the rotor, this is to get frequency of the signals everytime a magnet gets past to the hall sensor.

2.1. Device Sketch and 3D Models

The following images below are the rough sketches of the IoT device:

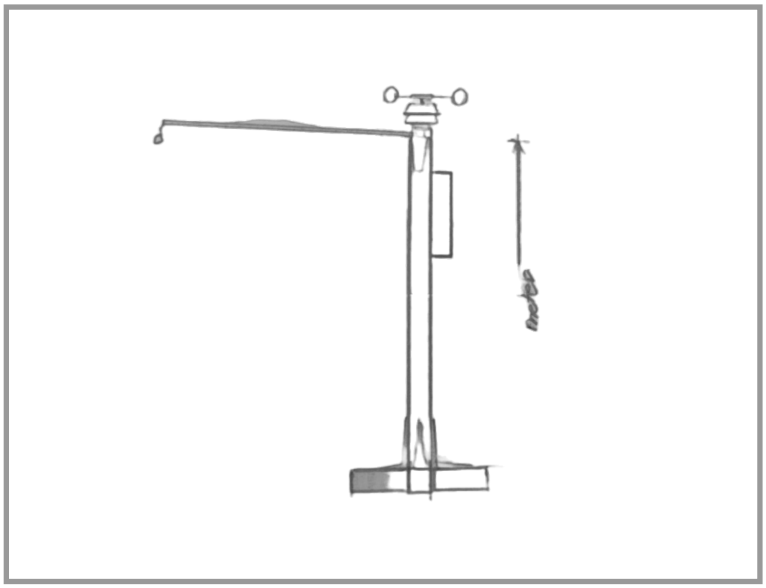


Fig. 1*:* IoT Design and Model. The figure depicts a preliminary outline of the IoT stand's design. It provides a visual representation of the positioning of the devices.

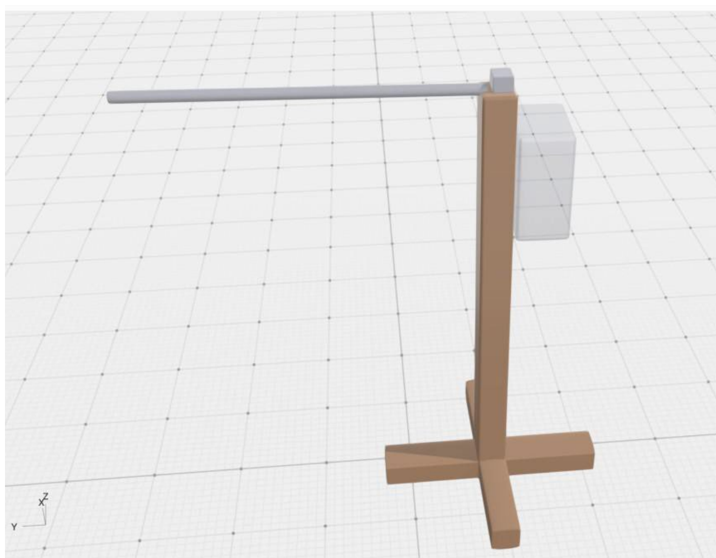


Fig. 2: Stand Design. This figure depicts the 3d model of the IoT stand's design. It provides a visual representation of the positioning of the devices

Below are the actual images of the IoT Device:

Fig. 3-4: IoT actual Design & Model Implementation. Figures presented above, labeled as Figure 3 and 4, shows the actual implementation of the previous model designs.

Table 2: Anemometer Specifications

Name	Quantity
M4x80 threaded rod stainless	3
M4 nuts stainless	3
Washer stainless	3
DHT11	1
BME280	1
8x40mm Round Aluminum	1
Magnets 5x2mm	3
608ZZ Ball Bearing	2
Hall Sensor	1

Table 2 above depicts the anemometer specifications and its quantities. It provides a list of components required for a project or assembly. It includes the quantities and specific items needed. The components listed include M4x80 threaded rod stainless, M4 nuts stainless, washer stainless, DHT11 sensor, BME280 sensor, 8x40mm Round Aluminum, magnets 5x2mm, 608ZZ ball bearing, and a Hall sensor. These components are essential for creating the Anemometer

The images below present the Three-Dimensional (3D) model of the Anemometer:

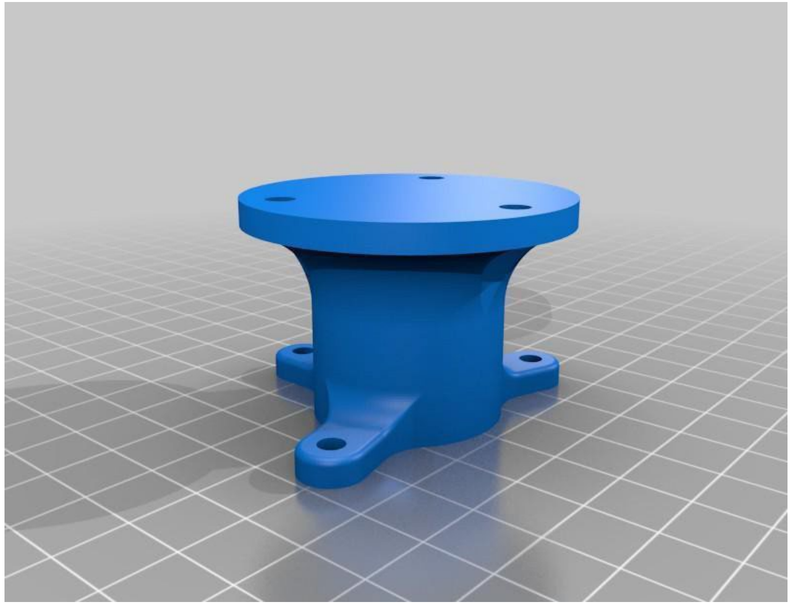


Fig. 5: Bottom Stand of the Anemometer. This is the bottom stand of the anemometer which is the main foundation of the anemometer to be able to stand.

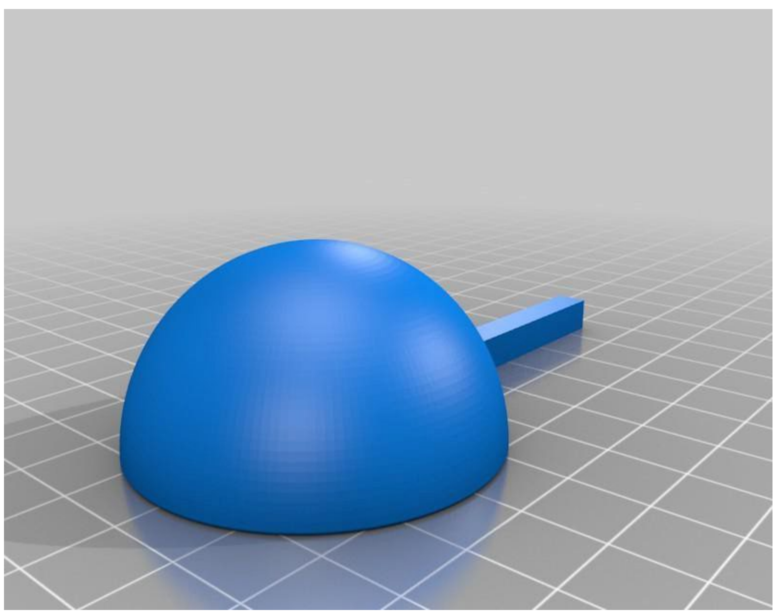


Fig. 6: Vane structure of the Anemometer. An anemometer typically has three or four cups that are evenly spaced and attached to a horizontal axis, which rotates as the wind blows. The rotation of the cups-like shapes or vanes are used to measure the speed of the wind. The cups are usually hemispherical in shape and made of lightweight materials such as plastic or metal.

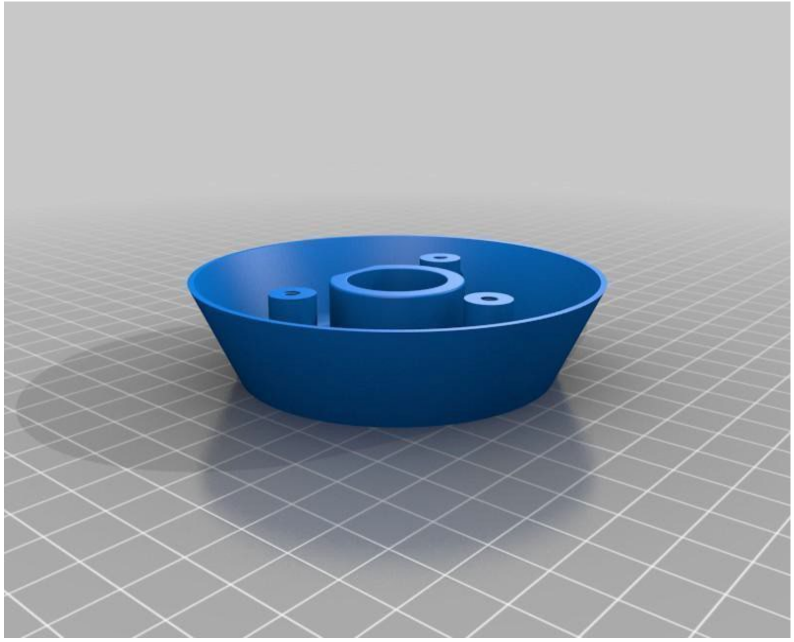


Fig. 7: Top part of the Anemometer. This is a part of the anemometer in which the screws will go through to lock anemometer and to protect the microcontrollers sensors inside.

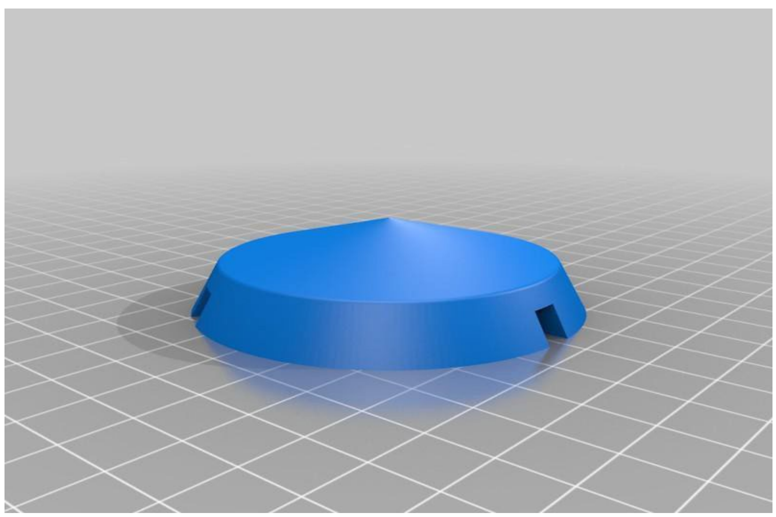


Fig. 8: Rotor of the Anemometer. An anemometer's rotor is the portion of the instrument that rotates in response to wind. It is often made up of many cups or vanes installed on a horizontal shaft and aligned perpendicular to the shaft. When the wind blows against the cups or vanes, it causes the rotor to revolve around the shaft, which is then utilized to compute the wind speed. The rotor is an important part of the anemometer since it is the part of the instrument that transforms wind energy into mechanical motion that can be measured.

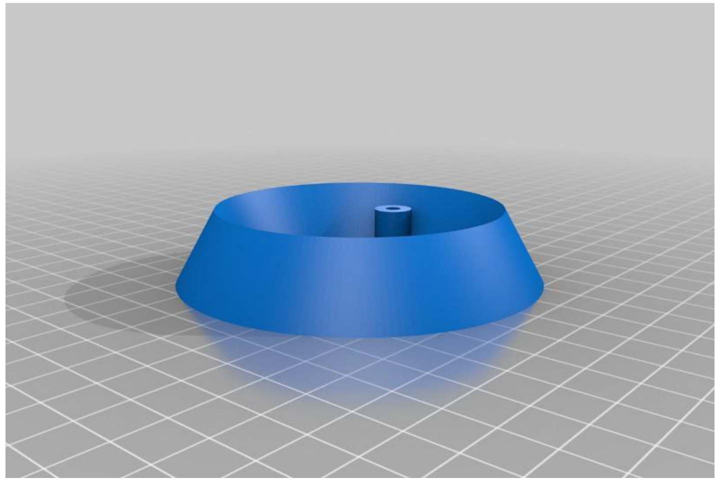


Fig. 9: Middle part of the Anemometer. This is just the middle part of the Anemometer to put the Anemometer in place and this is also a component to protect the microcontrollers that are placed inside.

2.2. Input/Output Diagram

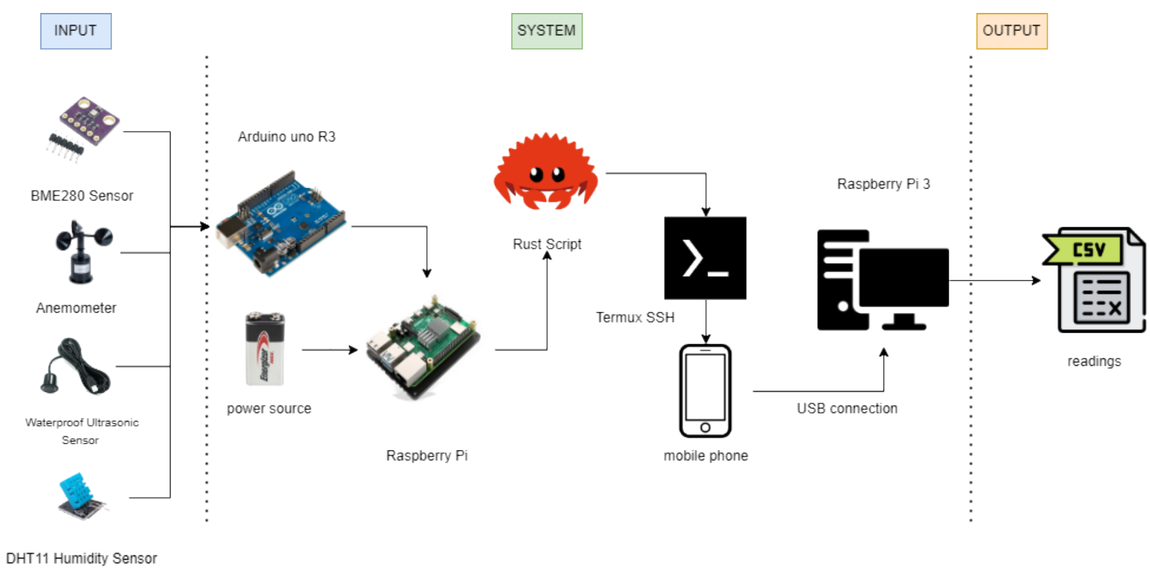


Fig. 10: Input/Output Diagram. The above figure illustrates the path of data flow from input devices to the system where it is processed and transformed into a CSV file. The input devices gathered data which is then collected and delivered to the system for processing. The raw data started from the sensors captured through the Arduino microcontroller. The Raspberry Pi received the data from the microcontroller through a serial connection, and a Rust Lang application was used to create the CSV file. The generated CSV file was obtained through SSH connection using the Termux application in mobile phone and then transferred to the computer through USB connection to the desktop computer and was used for analysis.

Anemometer and JSN SR04T Calculations

The anemometer must employ a factor value of three to offset the resistances and to ach ieve a substantially regularized computation of wind speed. However, due to the complexity and a lack of wind tunnel testers in Davao City, the researchers were only able to generate a substandard anemometer model using basic anemometer calculations. The formula for wind speed is as follows:

$$C = 2 * 3.14 *$$

$$r A = C / 3$$

$$f = 3$$

$$V = F * A * f$$

Where V is the wind speed, F for frequency of the signals, A for arc length from one cup to next, f for the anemometer factor and C for circumference. The arc length was used since anemometer uses 3 neodymium magnets as signals for the hall sensor in which its circumference was divided by three (3) to get the distance from each signal.

On the other hand, getting the water level measurement of *JSN SR04T* Ultrasonic sensor was done by first calculating the travel time of sound clicks from the sensor down to the water body, and then subtracting it to the distance of the sensor to the sea floor. The formula can be described into:

$$d = duration * 0.034 / 2$$

Where *d* is the *distance* between the sensor and obstacle, *duration* is the elapsed time from starting signal to the end signal, and *0.034* is a constant speed of sound in centimeters per microsecond.

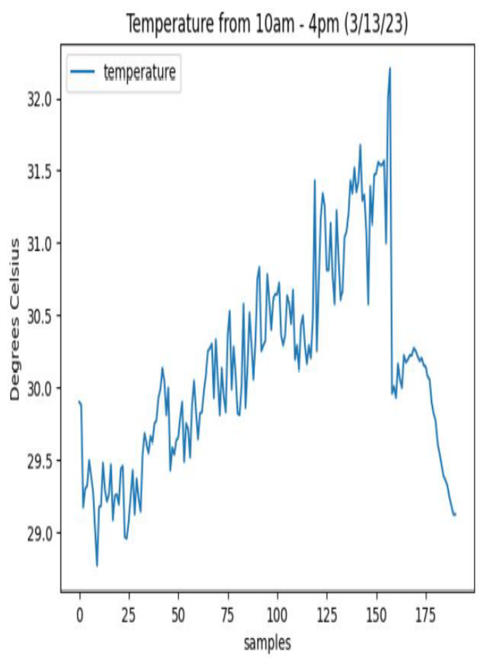
Deployment of the Device

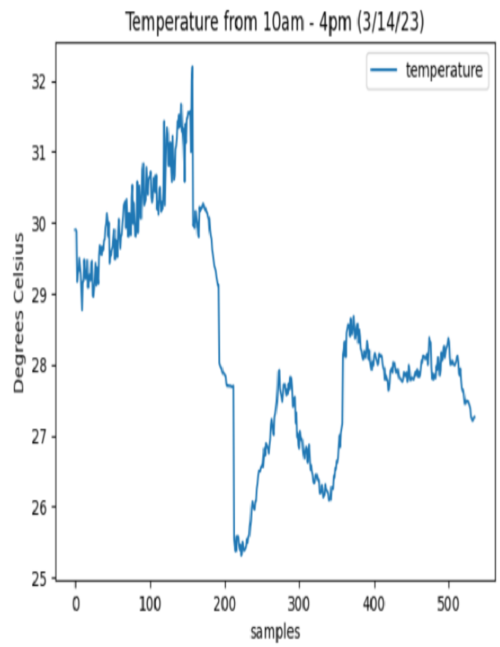
The IoT system is placed in the entrance of Sta. Ana Port Pier 1, after assembling, its frontal bar is aimed at the water body, sticking out the water sensor. The distance is measured with steel tape down from the water sensor to the seabed. To turn the device on, plug the micro-USB cable from the power source to the Raspberry Pi and connect the Arduino into it. This automatically started its readings and was saved through the SD card of the device.

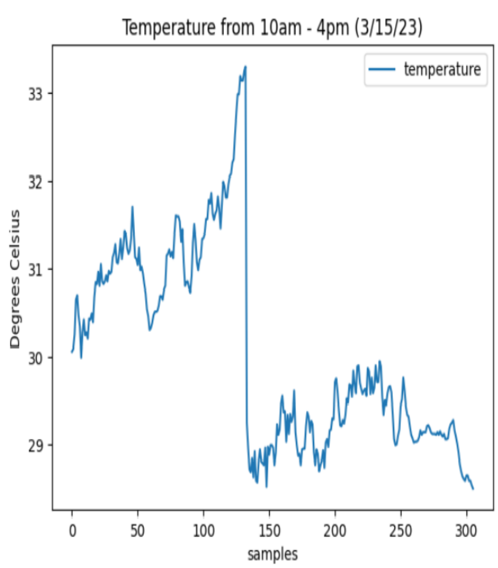
3. Results and Discussion

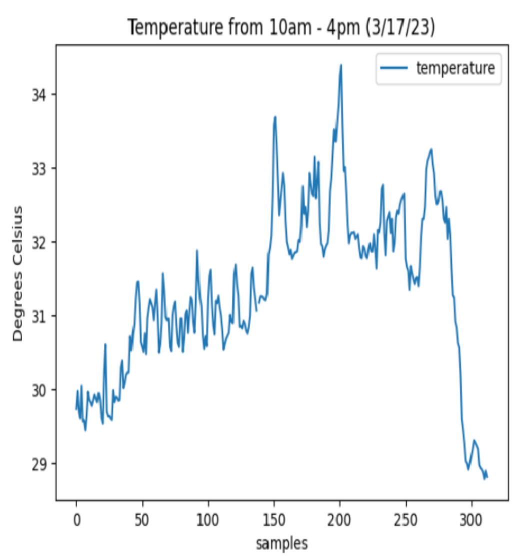
The data from the sensors were collected and analyzed using a Raspberry Pi and an Arduino Uno R3. This includes a collection of values for each parameter, as well as data visualization through graphs to help analyze trends and patterns in the readings.

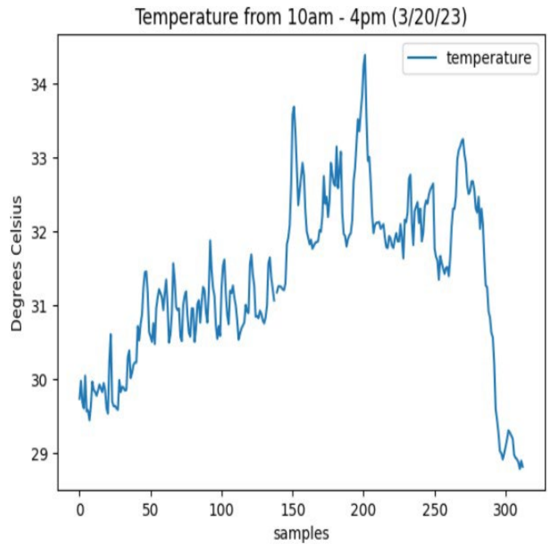
Temperature Results











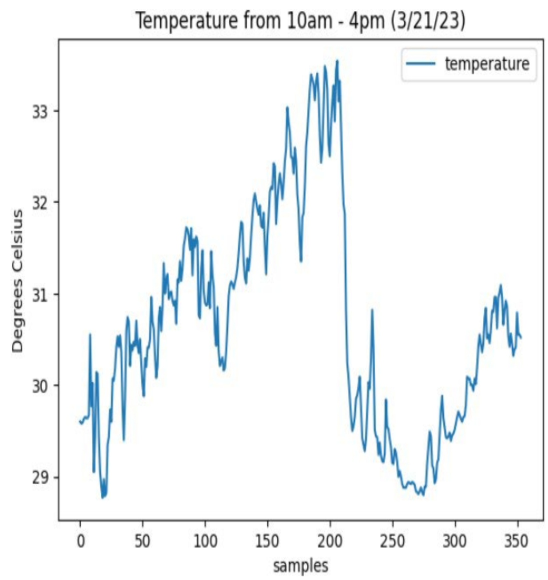


Fig. 11: Temperature Results. The average temperature readings for each day over the course of a week are shown in this graph. The sample is represented by the x-axis, while the temperature in degrees Celsius is shown by the y-axis. The figure shows the temperature variations at Santa Ana Wharf over the seven days that were monitored. You may see differences in daily temperature trends by contrasting the various temperature parameters. It can be noticed that the temperature graph patterns in sta ana port are different, the graphs show that there are sudden drops in temperature throughout the day. The temperature trends at Santa Ana Wharf may be quickly and thoroughly viewed in these graphs, which helped analyze the daily temperature fluctuations and patterns over the course of a week.

Pressure Results

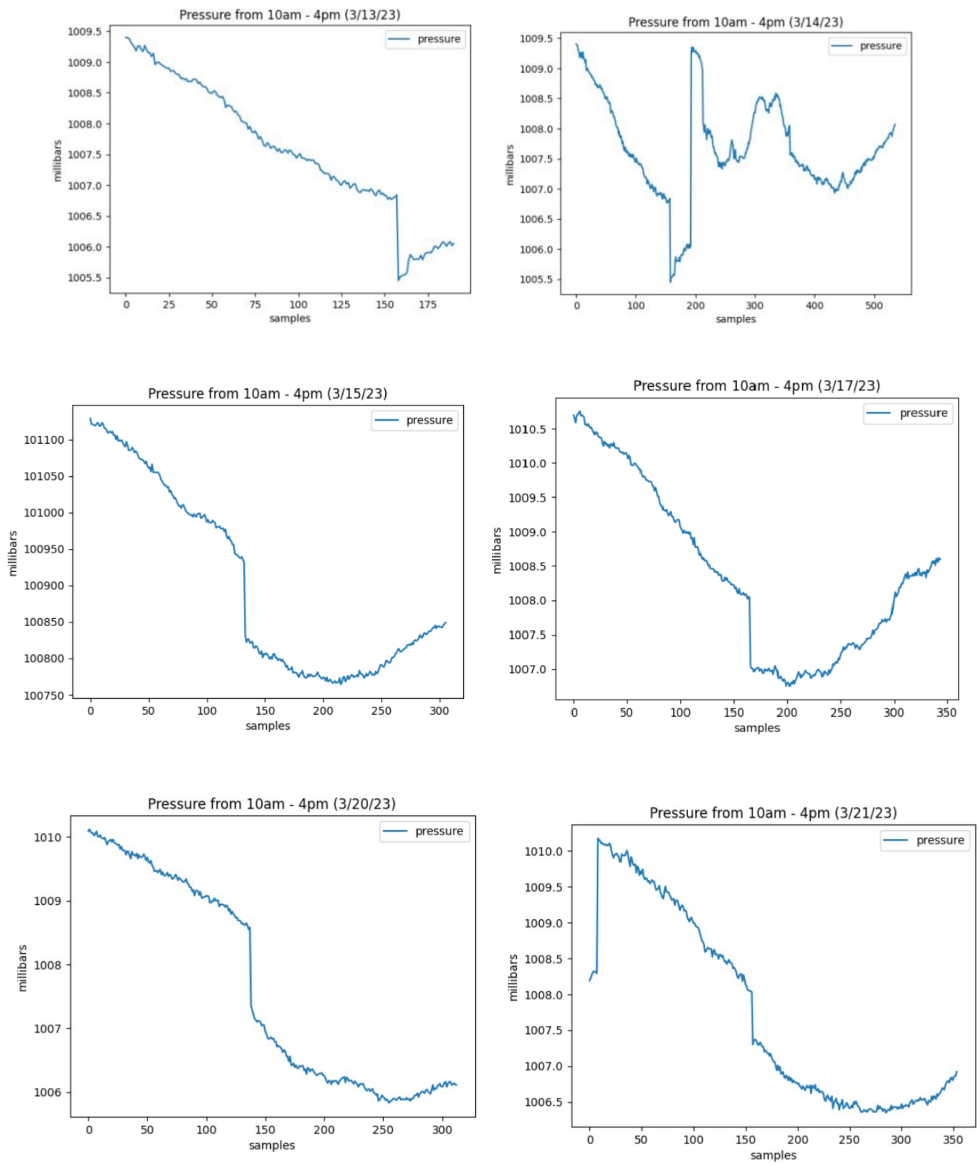


Fig. 12: Pressure Results. In figure 12, we present the average monitored pressure readings for each day over the seven (7) day period. The x-axis represents the sample, while the y-axis represents the pressure monitored in Millibars. The graphs provided a visual representation of the pressure fluctuations at Santa Ana Wharf over the observed seven (7) day period. By comparing the different pressure parameters, you can observe variations in daily pressure patterns. It can be noticed that the daily graph shows different patterns on the first and the second day, whilst the third to the last day show similar rise and fall in the readings. These graphs allowed for a quick and comprehensive overview of the pressure trends at Santa Ana Wharf, providing insights into the daily pressure variations and patterns over the 7-day period.

Water Level Results

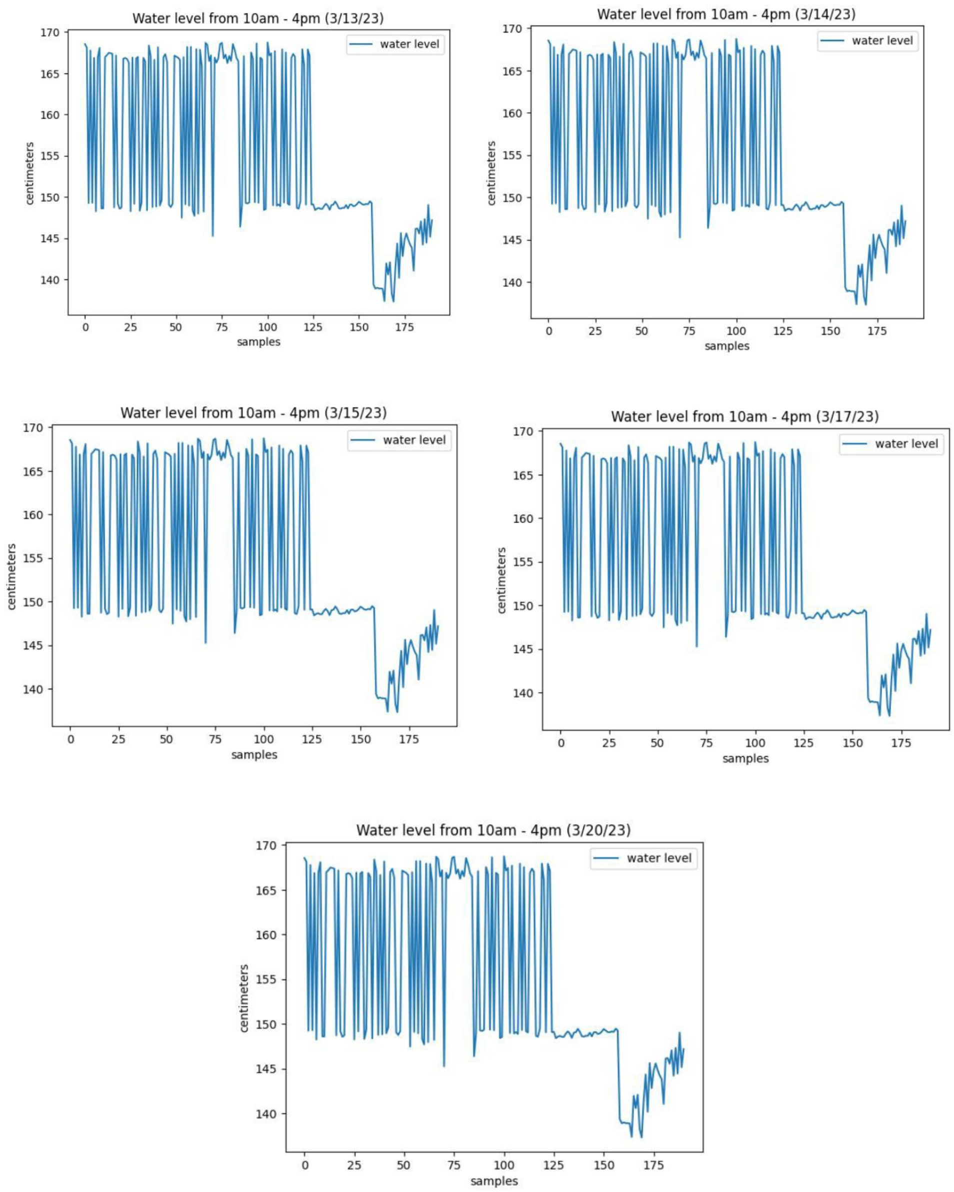


Fig. 13: Water Level Results. Figure 13 shows the changes in the water level at Santa Ana Wharf over the course of seven days. It displays the water level's daily rise and fall over a period of time. For instance, you might see increased water levels on some days because of tide patterns or weather conditions. It can be noticed that the water levels in sta ana have almost the same patterns for 7 days as the water levels rise and fall are attributed to the gravitational pull of the moon which is a constant variable. The graph would show a summary of the overall trends in water level throughout the course of the monitoring period.

Wind Speeds Results

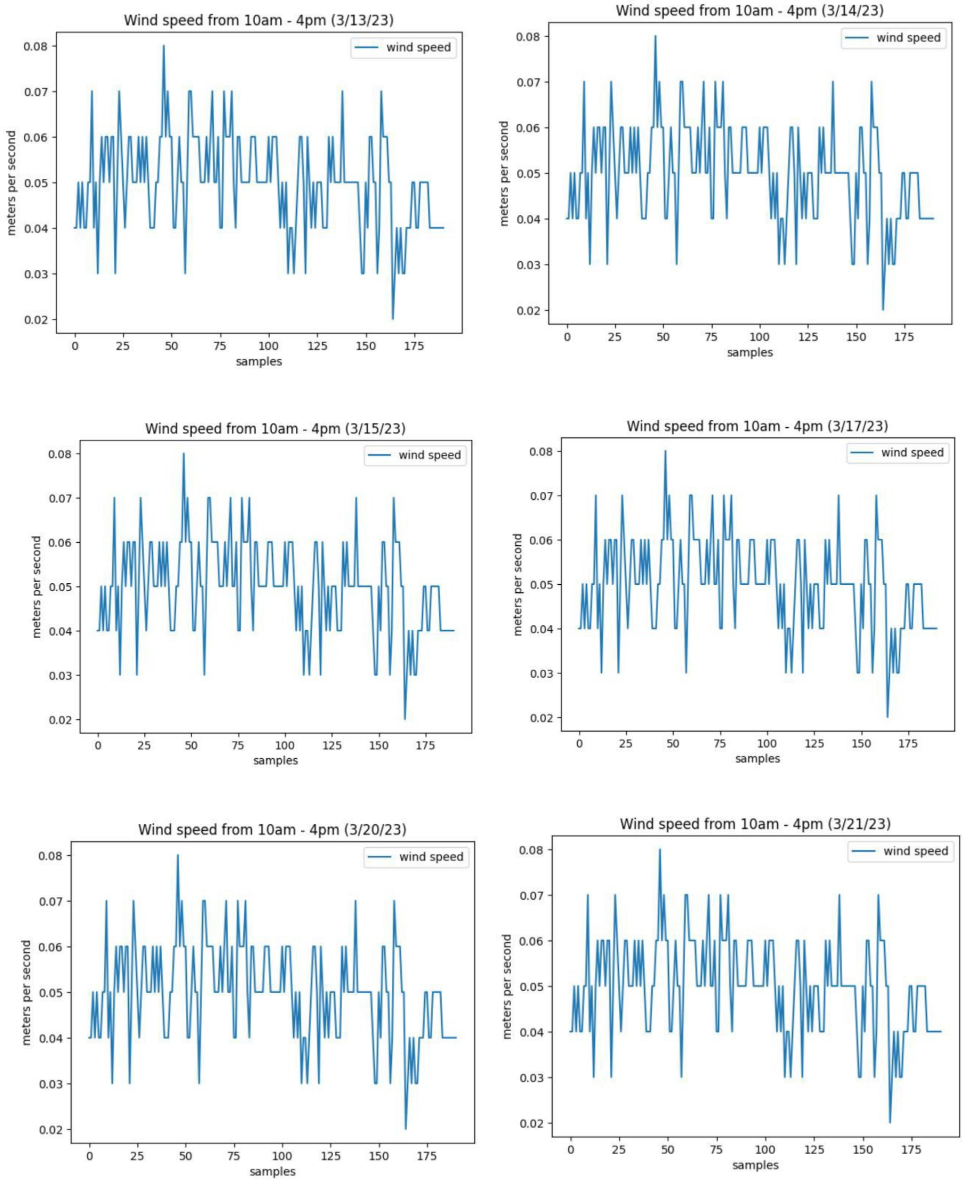
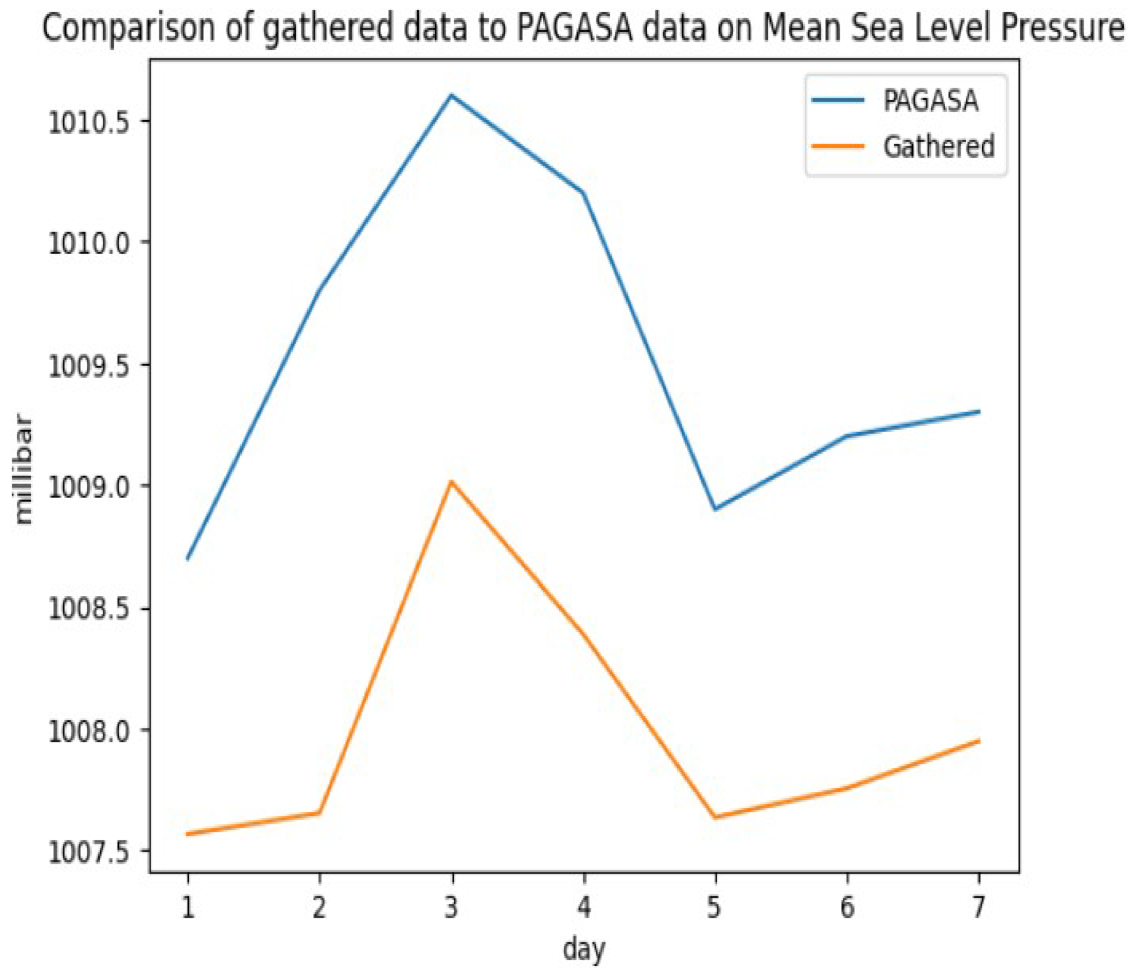
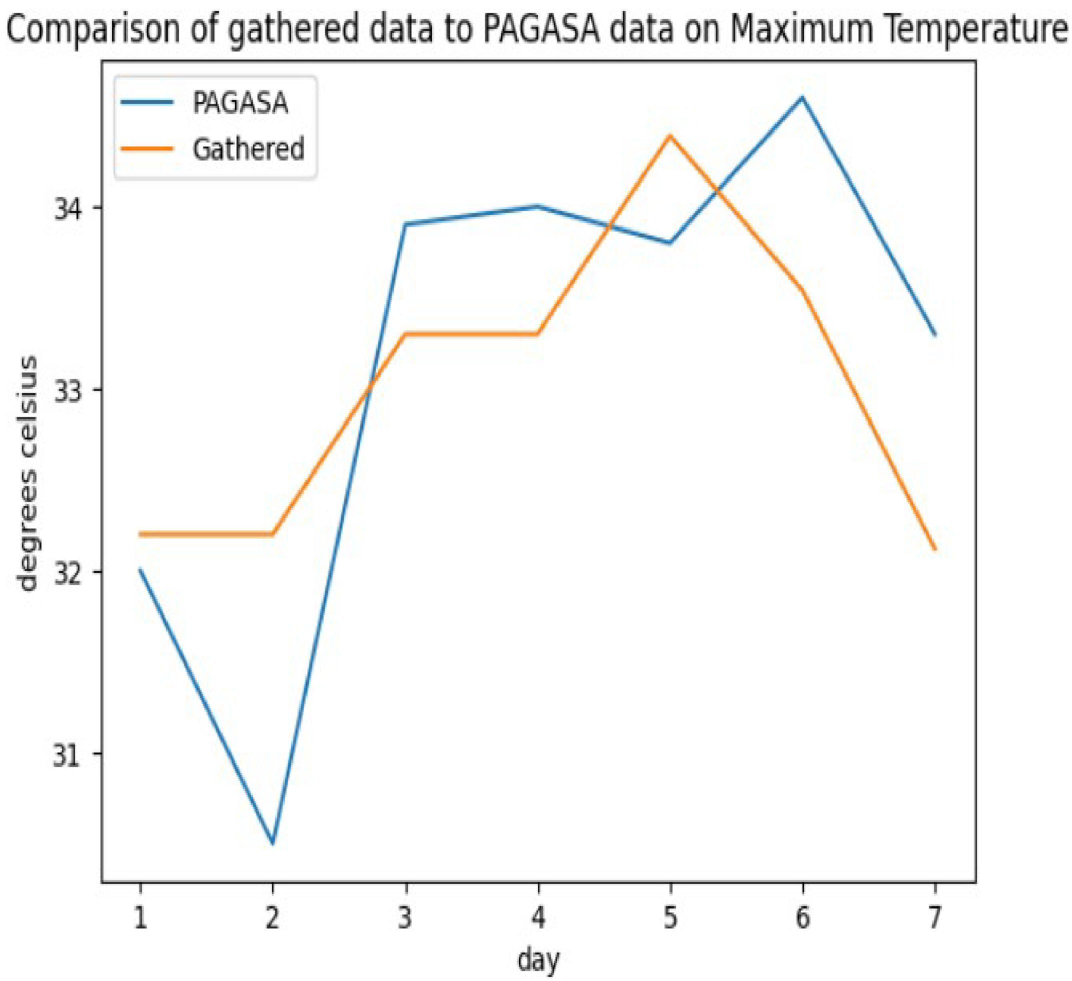


Fig. 14: Wind Speed Results. In this figure 14, the x-axis represents the sample, while the y-axis represents the monitored meters per second. The graphs present visual representations of the wind speed variations at Santa Ana Wharf over the observed seven (7) day period. By comparing the different wind speed parameters, variations in daily wind patterns were observed. For example, the average wind speed showed a sudden drop and increase in strength, while measurements with the same patterns are also depicted in the graphs. This graph allows for a quick and comprehensive overview of the wind speed trends at Santa Ana Wharf, which shows the daily wind variations and patterns over the seven (7) day period.



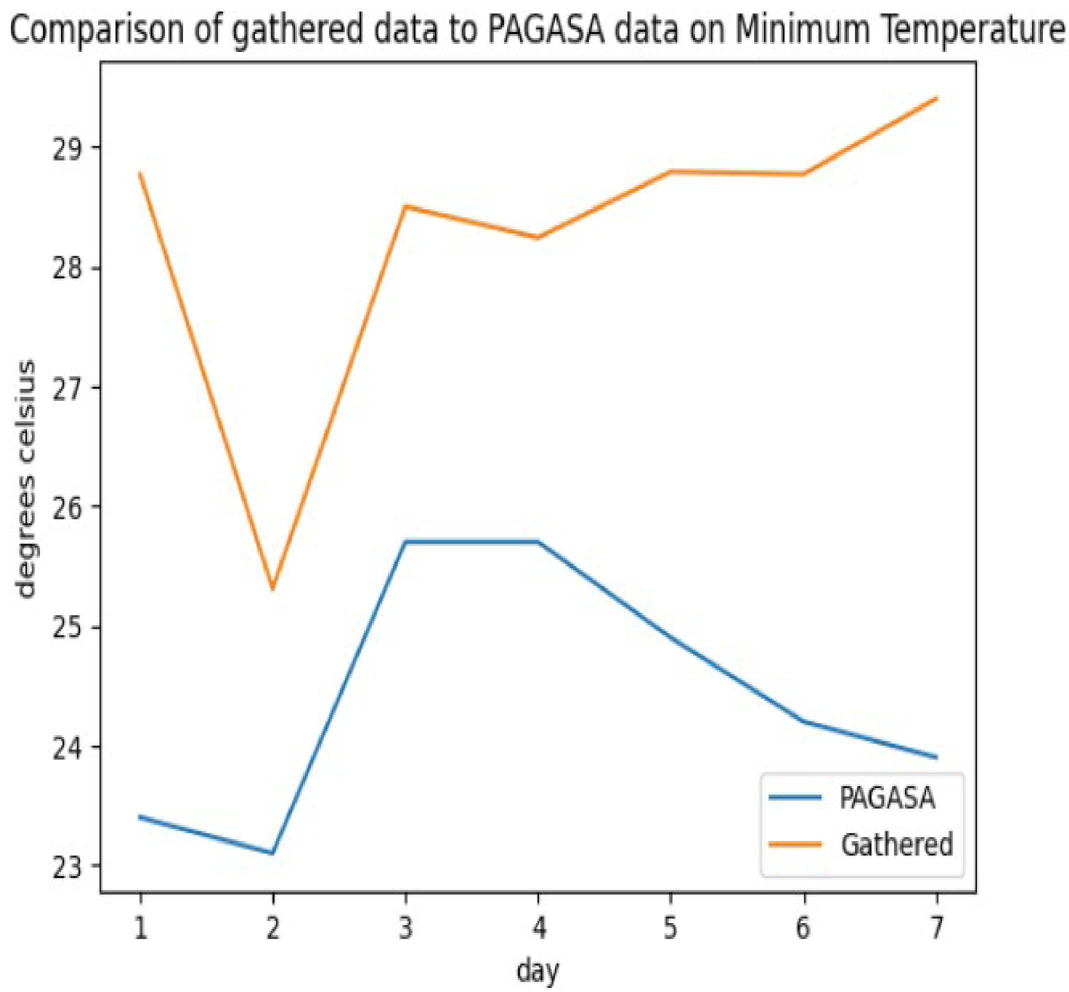

```
mean absolute error: 1.53
mean absolute percentage error: 0.15%
mean squared error: 2.46
root mean squared error: 1.57
```

Fig. 15: Mean Sea Level Pressure Comparison. In this figure 15, the graph shows the comparison of the gathered data to the data from PAGASA on the data of Mean Sea Level Pressure from day 1 to day 7, the PAGASA data starts from between 1008.5 and 1009.0 millibar and rose up to 1010.5 on the third day and went down 1009.0 on the fifth day and went straight up to 1009.0 on sixth and seventh day on the other hand, the gathered data from the device, it started from 1007.5 and rose up to 1009.0 on the third day then went down between 1007.5 and 1008.0 and remained on the same range on the sixth and seventh day. The mean absolute error of 1.53, it has the mean absolute percentage error of 0.15%, it has mean squared error of 2.46, and lastly it has root mean squared error with 1.57.



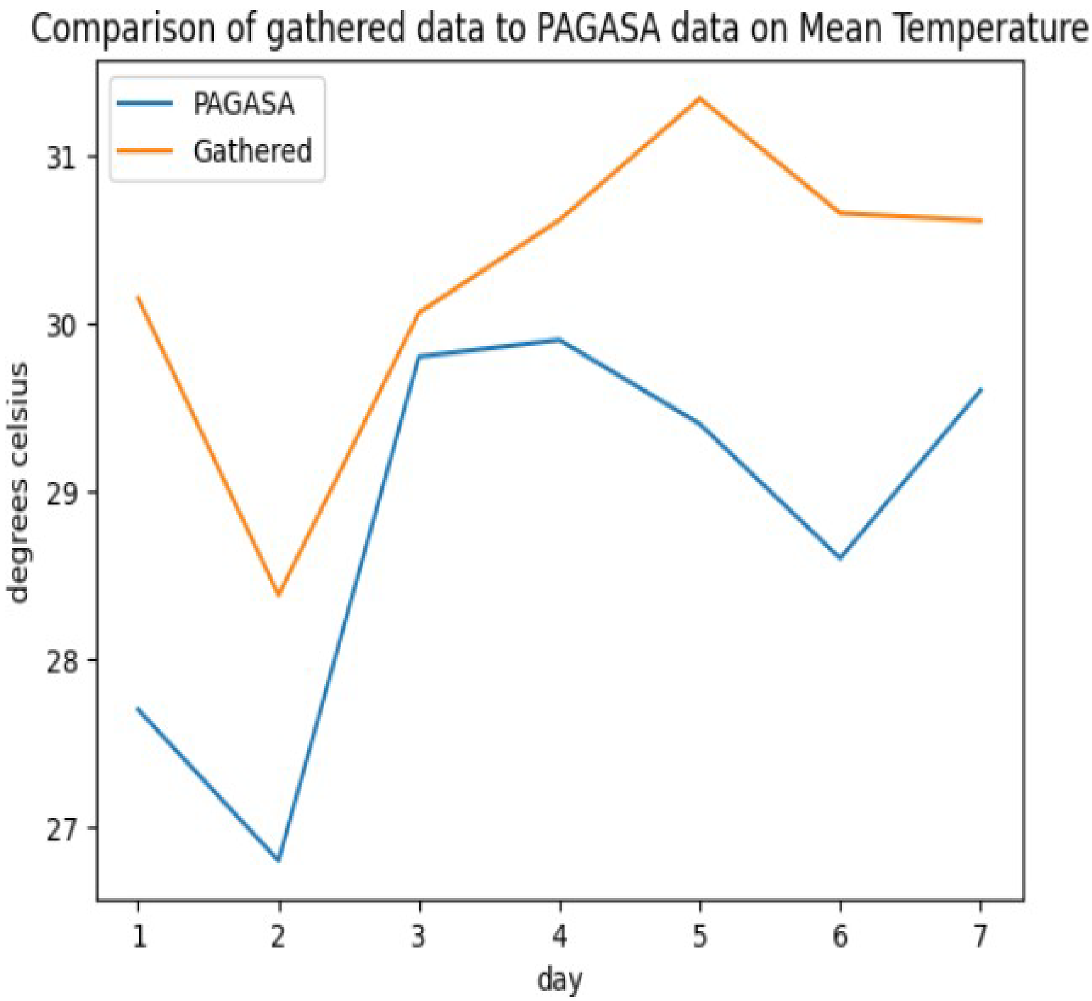

```
mean absolute error: 0.86
mean absolute percentage error: 2.63%
mean squared error: 0.95
root mean squared error: 0.97
```

Fig. 16: Maximum Temperature Comparison. In this figure 16, the graph shows the comparison of the gathered data to the data from PAGASA on the data of Maximum Temperature with the ranges from day 1 to day 7, the PAGASA data starts from between 32 degrees Celsius and then went down to 31 degrees Celsius and below on the second day and rose up to 34 degrees Celsius on the third day while it remained within that range on the fourth and fifth day. It was also observed that after remaining stable for two days, the range however, went straight up above 34 degrees Celsius on sixth day and then went down to 33 degrees Celsius on the seventh day. On the other hand, the data obtained from the device began at 32 degrees Celsius and reached 33 degrees Celsius on the third and fourth days before dropping back to 32 degrees Celsius on the sixth and seventh days. The mean absolute error of 0.86, it has the mean absolute percentage error of 2.63%, it has mean squared error of 0.95, and lastly it has root mean squared error with 0.97.



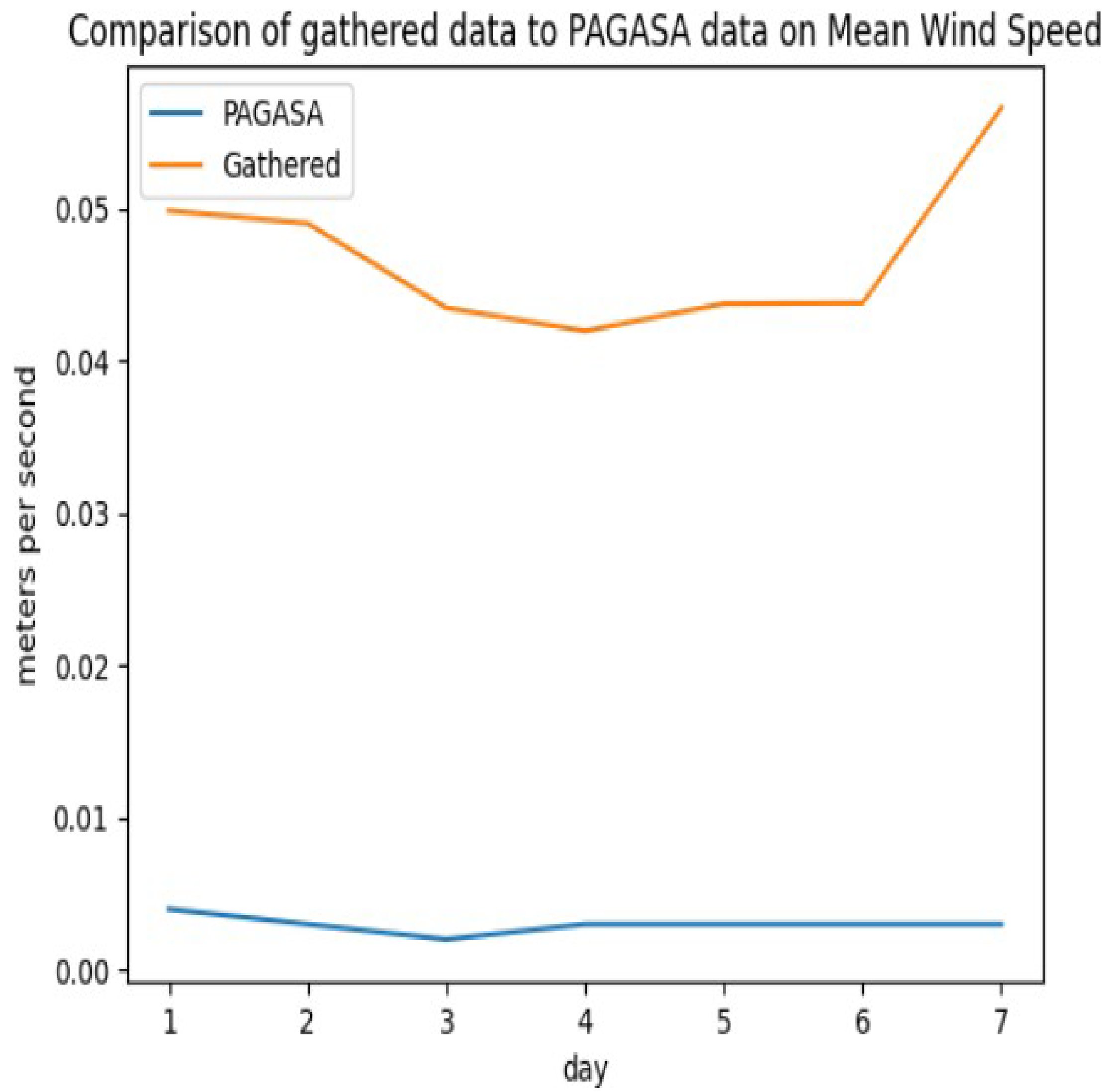

```
mean absolute error: 3.84
mean absolute percentage error: 15.83%
mean squared error: 16.33
root mean squared error: 4.04
```

Fig. 17: Minimum Temperature Comparison. The graph compares the gathered data to PAGASA data on the Minimum Temperature from day one (1) to day seven (7). PAGASA data began between 23 and 24 degrees Celsius for two days, then went up to 26 degrees Celsius and remained between 25 and 26 degrees Celsius on the third until the fifth day. However, this went down to 24 degrees Celsius on the sixth day and seventh day. Meanwhile, the device's data ranged from 29 degrees Celsius to 25 degrees Celsius on the second day. The temperature then increased to between 28 and 29 degrees Celsius on the third day and stayed there for the rest of the week. The mean absolute error of 3.84, it has the mean absolute percentage error of 15.83%, it has mean squared error of 16.33, and lastly it has root mean squared error with 4.04.




```
mean absolute error: 1.43
mean absolute percentage error: 5.03%
mean squared error: 2.58
root mean squared error: 1.61
```

Fig. 18: Mean Temperature Comparison. In figure 18, the graph shows the comparison of the gathered data to the data from PAGASA on the data of mean temperature within seven (7) days. PAGASA data started between 27 degrees Celsius and 28 degrees Celsius for the first day then went down to 27 degrees Celsius on the second day. It then went up to 30 degrees Celsius on the third day to fourth day while remaining within this range on the fifth day. However, it dropped on the sixth day with 29 degrees Celsius but went up again on the seventh day. The data collected from the device, on the other hand, began at 30 degrees Celsius and decreased between 28 and 29 degrees Celsius on the second day before rising to 30 degrees Celsius on the third day. On the fifth day, the temperature rose to 31 degrees Celsius and stayed there for the rest of the week. The mean absolute error of 1.43, it has the mean absolute percentage error of 5.03%, it has mean squared error of 2.58, and lastly it has root mean squared error with 1.61.




```
mean absolute error: 0.04
mean absolute percentage error: 1507.96%
mean squared error: 0.00
root mean squared error: 0.04
```

Fig. 19: Mean Wind Speed Comparison. In this figure 19, the graph shows the comparison of the gathered data to the data from PAGASA on the data of Mean Wind Speed for seven days. Based on the PAGASA data gathered on the deployment time, it can also be noticed that the wind speed readings gathered are highly inaccurate due to significantly high mean absolute error score, this is affected by several factors, a major one would be the location, where winds are specifically persistent at coastal areas. Based on the results, it has the mean absolute error of 0.04, it has the mean absolute percentage error of 1507.96%, it has mean squared error of 0.00, and lastly it has root mean squared error with 0.04.

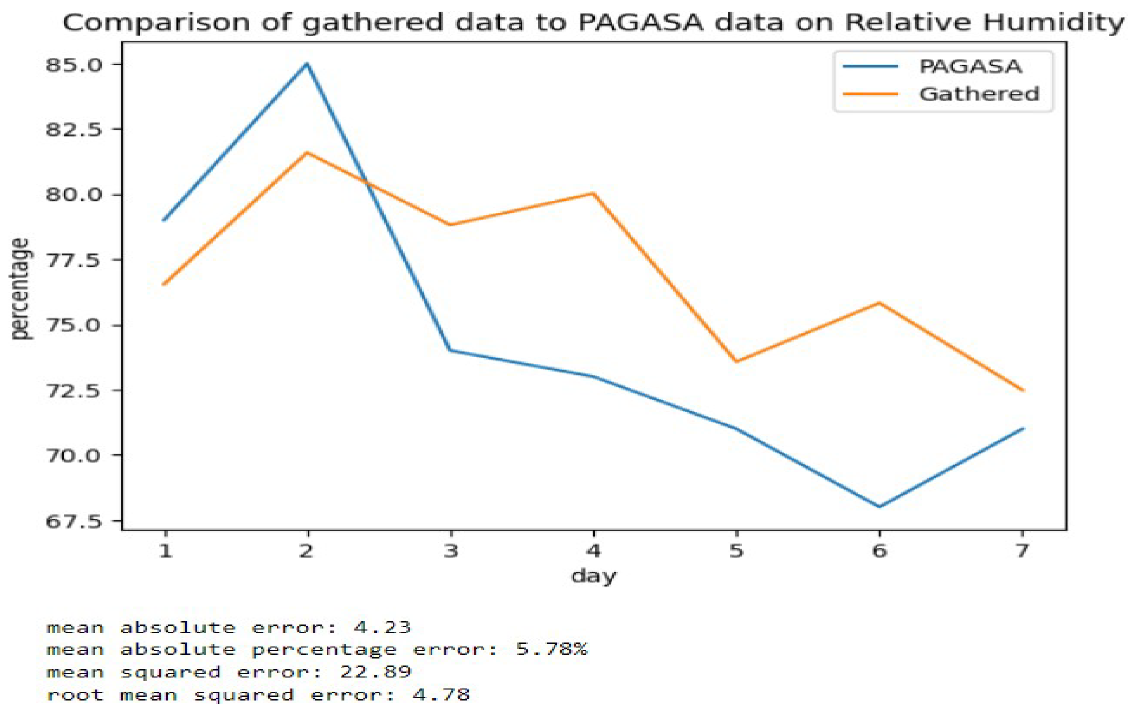


Fig. 20: Comparison of gathered data to PAGASA Data on Relative Humidity. In this figure 20, the graph shows the comparison of the gathered data to the data from PAGASA on the data of Relative for seven days, the PAGASA data starts from between 77.5 percent and 80 percent for the first day then went up to 85 percent on the second day and on the third day went down between 75 percent and 72.5 percent and for fourth day and fifth day, it kept going down from 75 percent and on the sixth day it went down to 67.5 percent but went up to 70 percent on the seventh day on the other hand, the gathered data from the device, it starts between from 75 percent and to 77.5 percent, on the second day it went up to 80 percent then stays within those ranges on the third and fourth day,and it went down between 75 percent and 72.5 percent on the fifth day stays within those ranges on the last two remaining days. The mean absolute error of 4.43, it has the mean absolute percentage error of 5.78%, it has mean squared error of 22.89, and lastly it has root mean squared error with 4.78.

The results show that the proposed IoT system is capable of monitoring Temperature, Pressure, Water Level, Wind Speed, and Humidity. The system provides live monitoring data that can be accessed remotely using a web interface depending on the microcontroller used. The analysis of the data reveals that there is not much of a change in the features: temperature, pressure, wind speed, and humidity, but has substantial values on the water level, which we can directly correlate it to tidal patterns.

4. Conclusion

The IoT System proposed by the researchers for detecting the elements specifically speed, humidity, temperature, water level, and pressure proved to be viable during the deployment in Sta. Ana Wharf with a duration of seven (7) days. This research project successfully achieved its objectives of identifying environmental factors, developing, and deploying an IoT-based system for monitoring, and visualizing the measured water level and environmental parameters. The study highlighted the significance of environmental monitoring, particularly water level, in assessing coastal risk in Davao City. Accurate and timely data collection is crucial for mitigating risks and preventing disasters. The researchers successfully identified the environmental features, which are: humidity, wind speed, temperature, pressure, and water level, then implemented an IoT device with the specific sensors to measure them. All three objectives outlined in this article have been successfully achieved.

The first objective focused on identifying environmental factors, humidity, wind speed, temperature, pressure, and water level, for assessing coastal risk in Davao City. These were identified through literature reading and its availability in second article's PAGASA datasets. The researchers effectively identified these environmental features and implemented specific sensors to measure each factor. The second objective involved the development and deployment of an IoT-based system capable of measuring wind speed, humidity, temperature, water level, and pressure. The researchers successfully created an IoT device for this purpose, as demonstrated in this paper, and the results are presented in graphical form. Lastly, the third objective aimed to tabulate and visualize the measured water level and environmental monitoring data in a clear and straightforward manner. The researchers accomplished this by using a Python Pandas library that displays and visualizes the results programmatically. This tool was used to tabulate and visualize data which are all presented in results section.

The graphs shown in this first article are a comprehensive analysis of temperature, pressure, water level, wind speed, and humidity variations at Santa Ana Wharf over a 7-day monitoring period. The analysis has provided valuable insights into the daily fluctuations and patterns of each parameter, which can help in understanding the dynamic nature of weather patterns at Santa Ana Wharf. The temperature graphs showcased distinct patterns, including sudden drops throughout the day, which could be useful in identifying weather patterns. The pressure graphs exhibited variations, with the initial two days displaying different patterns compared to the subsequent days, which could be indicative of weather changes in the region. The water level graph revealed consistent patterns attributed to the gravitational pull of the moon, which can be useful in predicting tides in the region. The wind speed graphs depicted fluctuations, with sudden drops and increases in strength, which could be useful in predicting wind patterns and potential weather changes. Lastly, the humidity graphs demonstrated significant changes during the early hours and spikes during the late hours, which could be useful in predicting humidity changes and potential weather changes. These visual representations have significantly contributed to the understanding of the dynamic nature of weather patterns at Santa Ana Wharf. The findings presented in this thesis paper lay the foundation for further research and the development of predictive models that can enhance weather forecasting accuracy in this region. There is a proposed method for building a model out from these features which are later discussed in the second article of this paper.

REFERENCES

Odesola, D. F., Olivarez, R., Ramos, B. A., Malolos, D., & Viray, P. (1970, January 1). Internet of things (IOT) based Home Automated Weather Monitoring System. LPU â€" Laguna Journal of Engineering and Computer Studies. https://ejournals.ph/article.php?id=15417

Gaynor, A.-L., & Sevilleno, C. N. (2018). Matina Flood Watch: An Urban Flooding Prediction Application using Random Tree Algorithm. Matina Flood Watch: An Urban Flooding Prediction Application Using Random Tree Algorithm, 2–97. https://doi.org/10.13140/RG.2.2.19397.50408

Granata, F., & Di Nunno, F. (2021, April 9). Artificial Intelligence models for prediction of the tide level in Venice. Stochastic Environmental Research and Risk Assessment, 35(12), 2537–2548. https://doi.org/10.1007/s00477-021-02018-9

Knight, P., Bird, C., Sinclair, A., Higham, J., & Plater, A. (2021, January 8). Testing an “IOT” tide gauge network for coastal monitoring. MDPI. https://www.mdpi.com/2624-831X/2/1/2

Arshad, B., Ogie, R. I., Barthelemy, J., Pradhan, B., Verstaevel, N. R., & Perez, P. (n.d.). Computer vision and IOT-based sensors in flood monitoring and mapping: A systematic review. Research Online. https://ro.uow.edu.au/smartpapers/306/

Yuyun, Y., Zulfaesa, H., & Abdul, L. A. (2020, November 24). Early warning system for flood disasters using the internet of things. Early Warning System for Flood Disasters Using the Internet of Things | Atlantis Press. https://www.atlantis-press.com/proceedings/ijcse-20/125946416

A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari and M. Ayyash, "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications," in IEEE Communications Surveys & Tutorials, vol. 17, no. 4, pp. 2347-2376, Fourthquarter 2015, doi: 10.1109/COMST.2015.2444095.

Hafiz Suliman Munawar a, a, b, management, A., Han, F., Li, Q., Klemens, J. A., Ip, F., Kishorbhai, V. Y., Lohani, A. K., Lu, P., Mohammadimanesh, F., Motta, M., Plate, E. J., Pregnolato, M., Rahman, M., Rahman, M. R., Seuring, S., Shu, C., … Sankaranarayanan, S. (2021, September 9). A review on flood management technologies related to image processing and machine learning. Automation in Construction. https://www.sciencedirect.com/science/article/abs/pii/S0926580521003678

C. -H. Yang, C. -H. Wu, C. -M. Hsieh, Y. -C. Wang, I. -F. Tsen and S. -H. Tseng, "Deep Learning for Imputation and Forecasting Tidal Level," in IEEE Journal of Oceanic Engineering, vol. 46, no. 4, pp. 1261-1271, Oct. 2021, DOI: 10.1109/JOE.2021.3073931.

Andang1, A., Hiron1, N., Chobir1, A., & Busaeri1, N. (2019, July 1). IOPscience. IOP Conference Series: Materials Science and Engineering. https://iopscience.iop.org/article/10.1088/1757-899X/550/1/012018

Dswilan1, S., Harmadi1, & Marzuki1. (2021, April 1). IOPscience. Journal of Physics: Conference Series. https://iopscience.iop.org/article/10.1088/1742-6596/1876/1/012003

An automated smart water level indicator using IOT-an effective ... - IJCSE. (2017, April). http://www.ijcse.net/docs/IJCSE17-06-04-010.pdf

Dibyahash Bordoloi, S. S. (2021). The internet of things (iot)-automated smart water level indicator: A practical application of Smart Irrigation. Webology. https://www.webology.org/abstract.php?id=3995

Article 2

Predicting Hydrological Risks in Coastal Areas Within Davao City Using Extreme Gradient Boosting Algorithm

John Mark P. Bago-od[a*], Franz Dominic S. Brillante[a], Rey Joshua M. Morales[a]

[a]College of Computer and Information Science, Mapúa Malayan Colleges Mindanao, Davao City, Philippines 8000

Emails: jmbagood@mcm.edu.ph, fdbrillante@mcm.edu.ph, rjmorales@mcm.edu.ph

*Corresponding author

College of Computer and Information Science

Mapúa Malayan Colleges Mindanao

Gen. Douglas MacArthur Hwy, Talomo

Davao City, Philippines

Email jmabagood@mcm.edu.ph

Abstract

Coastal Risks caused by Hydrological factors are becoming increasingly severe due to climate change and other weather conditions. Rising atmospheric temperatures contribute to sea-level rise, leading to heightened flooding and erosion in coastal areas, thus, hydrological risks have caused potential health and safety hazards to people. As a result, the researchers used an XGboost-based algorithm that read collected data on the environmental attributes to create a prediction based on the environmental features present. The researchers used data on water level, sea level, pressure, mean max min temperature, humidity, and wind speed in their algorithm which are collected from government facilities such as PAGASA and NAMRIA. This study focused solely on the environmental factors available to the researchers and did not include any inaccessible features or data. The results revealed that the machine learning model was able to predict at +-5% on test data, with Average water level and Pressure with the lowest mean absolute percentage error and Relative humidity and Wind speed with the highest. This denotes that the machine learning algorithm proved viable for predicting hydrological patterns with satisfactory accuracy in a seven (7) days range of prediction, particularly water levels, and pressure. Further research and development, including incorporating additional features and datasets, are recommended for refinement and improvement of the model, which can potentially be a valuable tool for predicting hydrological risks for coastal regions.

Keywords: Coastal risks, hydrological factors, climate change, Philippines, health and safety hazards, XGboost-based algorithm, humidity, wind speed, PAGASA, NAMRIA, machine learning model, hydrological patterns, prediction range.

1. Introduction

Coastal Risks caused by Hydrological risks are getting higher throughout the year because of climate change and other weather factors. The warming of the atmospheric temperatures contributes to the rising of sea levels, which also leads to increased flooding and erosion in coastal areas. As more vapor is absorbed by the atmosphere, it becomes the fuel that propels destructive and stronger storms. In some areas of the world, the United Kingdom for example, was devastated with a massive 177 billion euros worth of maintenance and damage to properties affected by hydrological risks, creating a potential health and safety hazards to people.

Hydrological risks in coastal areas, on the other hand, are those coastal floods, storm surges, tsunamis, and sea droughts. According to Mondino et al. (2019), hydrological risks are increasing because of the climatic and socioeconomic changes around the world and the awareness and knowledge of the local citizens regarding the preparedness for hydrological risks. The United Nations International Strategy for Disaster Reduction said that around 90% of all disaster losses are related to weather-related hazards. The sustainable and resilient development of the coastal zone needs to consider the nexus between the land and sea, such as the natural flows occurring between realms, to balance and incorporate information from multiple stakeholders, thus to develop regional and national strategies for coastal risk reduction through natural and hybrid infrastructure (Neumann and Unger, 2019; Sutton-Grier et al., 2015).

The precise prediction and assessment of hydrological risks are of importance and coincidentally, Machine Learning models are now replacing traditional models since traditional models are much more complex and computationally expensive.

Machine Learning algorithms are much faster, more accurate, and more precise in reading massive storages of data. It is also better at identifying patterns and can predict hydrological hazards in real time. The application of machine learning algorithms in the prediction and monitoring of hydrological risks especially in coastal areas can provide critical information for emergency response planning and decision-making for our fellow disaster risk reduction members. Accurate and timely hydrological risk predictions allow the authorities to take calculated measures to protect the lives and property of citizens living in hazard risks areas. This also helps the disaster risk reduction team in actions such as evacuating communities, issuing flood warnings, and implementing coastal erosion control measures. Furthermore, the use of ML algorithms can aid in the optimization of resource use and the reduction of costs associated with managing and mitigating hydrological risks.

In the Philippines, hydrological hazards are common such as tropical cyclones, flooding, droughts, etc. due to its location. Cases of storm surges are found everywhere in the country. In fact, way back in 1984 a storm surge from Typhoon Undang devastated hundreds of homes of people living in Basey, Samar. The worst storm surge recorded in the Philippines was Typhoon Yolanda which affected eastern Samar with over 6,000 people casualties affecting millions of lives, the storm surge recorded a maximum of over seven (7) meters in height. Storm Surges are considered the worst part of a hurricane since it is from the ocean that is moved by typhoon winds which are unpredictable and pose a high-level risk.

Specifically, the researchers sought to answer the following:

1. Can wind speed, humidity, temperature, and pressure be incorporated with coastal water levels for the prediction of hydrological risk?
2. Is the Extreme Gradient Boosting algorithm efficient in NAMRIA and PAGASA?
3. Does the Extreme Gradient Boosting algorithm perform well in generating predictions on wind speed, humidity, temperature, pressure, and water levels?

The researchers found out that predicting such hydrological risks intensity is crucial and needed to pay attention as these hydrological risks can cause significant damage and loss to environmental resources, ecosystems and their services, infrastructures, service provision and economic output, property, and livelihoods, and they can result in the loss of human lives, injuries, and other negative health consequences. With this, the researchers created a machine-learning model that can analyze and generate predictions used to assess the possibility of a hydrological risk. Hence, employing XGboost Algorithm to gather and analyze data from NAMRIA, PAGASA, and as well as data received from the sensor components to generate predictions of hydrological risk possibilities.

Also, the following are the specific objectives of the study:

1. Create a machine-learning model with Extreme Gradient Boosting using Datasets from NAMRIA and PAGASA
2. Perform Cross Validation on the Machine learning model created using different evaluation metrics and learning rates.
3. Create a website that displays the results of the predictions made from the Machine Learning Model.

The study was conducted to develop a model that analyzes data from hydrological risks and atmospheric data in the Davao City coastlines. The predictions made by the model was presented in a web application. Furthermore, the results of this study may be beneficial to the following:

NAMRIA. If NAMRIA decides to pursue a project relating to this topic, this research study may help the agency. This may offer them some useful information on the incorporation of weather characteristics into water levels.

PAGASA. This study may also help the government agency by providing informative data that can be transmitted to disaster risk reduction agencies to give them insights about other ways of implementing weather instruments, specifically in combining hydrological measurements with meteorological ones in cheap ways.

Nearby Settlements and Businesses. This research study may also benefit the communities residing in coastal regions by having an idea on creating their own instrument inspired from this research, wherein they would know the cheap ways of acquiring these kinds of instruments and have an idea on how to handle it.

Future Researchers. This study may also be beneficial to future researchers to be more inclined and prepared in investigating this topic. This may help them improve with the given recommendations while also being able to pave the way for more research to come as this topic is crucial for the development of advanced weather data collection techniques.

With this in mind, the delimitation of this machine learning approach includes the following aspects:

Dataset Preparation. The collected IoT data forms the basis for training and validating the machine learning model. The dataset includes historical records of hydrological risk intensities in coastal areas within Davao City, along with corresponding sensor data.

Feature Selection and Engineering. The machine learning model incorporates relevant features from the dataset, such as temperature, humidity, pressure, water levels, and wind speed. Feature engineering techniques are applied to extract meaningful representations of the data, enhancing the model's predictive capabilities.

Model Training and Validation. The XGboost algorithm is trained on the prepared dataset to learn the complex relationships between the selected features and hydrological risk intensities. The model is validated using appropriate evaluation metrics to assess its performance.

Prediction. The trained model is used to predict hydrological risk intensities in coastal areas within Davao City. By leveraging the power of machine learning, the research paper aims to provide accurate and effective prediction methods in aiding decision-making processes for disaster preparedness and response.

2. Methodology

2.1. Research Participants

The participants of this research were Philippine Atmospheric, Geophysical, and Astronomical Services Administration (PAGASA), National Mapping and Resource Information Authority (NAMRIA), the Mapua Malayan Colleges Mindanao (MMCM), Sta. Ana port personnel, the advisers, and the validators.

2.2. Data Collection

The datasets were gathered in two parts: first was collecting data received from NAMRIA (coastal water levels), and second was from PAGASA (wind speed, temperature, pressure, humidity).

2.2.1 NAMRIA Datasets

The National Mapping and Resource Information Authority (NAMRIA) is the central mapping agency of the government of the Philippines. Its primary mandate is to provide accurate and reliable maps, charts, and geospatial information to support national development and security, disaster risk reduction and management, environmental management, and other important sectors of society (NAMRIA, n.d.).

NAMRIA plays a critical role in providing accurate and reliable maps and geospatial information to support national development and security, disaster risk reduction and management, environmental management, and other important sectors of society in the Philippines. The datasets received from NAMRIA were from 2019, 2020 and 2021. The dataset format received is through .LEV files, which is a file format specific to LevelLogger, a proprietary software that NAMRIA uses in conducting their data collection operations.

2.2.2 PAGASA datasets

Philippine Atmospheric, Geophysical, and Astronomical Services Administration

The Philippine Atmospheric, Geophysical, and Astronomical Services Administration (PAGASA) is the national meteorological, climatological, astronomical, and geophysical agency of the Philippines (PAGASA). Its primary mandate is to provide weather forecasts and warnings, as well as climate-related information and services, to the government, the public, and various sectors of society (PAGASA).

PAGASA plays a vital role in ensuring the safety and well-being of Filipinos by providing timely and accurate information about weather and climate conditions, and by supporting disaster risk reduction and management efforts across the country (Department of Science and Technology).

The datasets received from PAGASA Davao branch, near the Old Airport, Sasa, ranged from 2019, 2020 and 2021. The researchers converted the datasets that were manually written into .csv format through manual typing. The data were recorded daily while the features taken were the mean sea level pressure, maximum temperature, minimum temperature, mean temperature, wind speed, and mean relative humidity which are the data needed for training the model. Also, automated sensors were placed at various points throughout the region to collect data. At regular intervals, measurements were taken to capture the weather over a predetermined time frame. The data underwent thorough preparation procedures to confirm and clean it, ensuring its accuracy and dependability.

To find patterns and trends in the meteorological data, statistical techniques were used in the analysis. To protect the integrity of the data, overall values and procedures used in the study were not provided; however, the statistical significance of the results was maintained. In order to preserve data privacy and prevent any identification of specific weather stations, the findings were limited into ten (10) examples.

2.3. Overview of the Datasets

Without disclosing overall values, the following table depicts a general summary of the weather data set:

Table 1: PAGASA Dataset

Day of the Month	Mean sea Level Pressure	Maximum Temp	Minimum Temp	Mean Temp	Mean Relative Humidity	Wind Speed
1	1007.7	33	25	29	74	0.002
2	1008.2	33	25.3	29.1	76	0.002
3	1010.4	29.3	24	26.6	84	0.003
4	1009.2	32.3	24.9	28.6	71	0.004
5	1008	32.1	24	28.1	77	0.004
6	1007.8	31.5	25.2	28.4	76	0.003
7	1008.8	31	25	28	76	0.003
8	1009.1	33.4	24.2	28.8	72	0.002
9	1008.7	33.2	26	28	77	0.004
10	1008.7	32.1	25.4	28.8	84	0.002

Table 2: PAGASA Dataset Specifications

Specifications	***Value***
Dimension	*30-31 x 9*

Rows	*30-31*
Columns	*9*

Table 1 provides a broad perspective on the range of each weather variable. Visualizations, such as graphs and charts, were also utilized to present trends and comparisons while maintaining data privacy. Please note that due to the sensitive nature of weather data and privacy concerns, the exact locations and complete specific measurements were omitted in this example. The focus is on the general trends and patterns observed within the data set to draw meaningful conclusions.

Table 3: NAMRIA Dataset Overview

Time Of Observation	**Day**	**Month**
0-23 Hours	1 - 31	Jan – Dec

The table above provides a general overview of the weather data set without disclosing specific values. The data set used in this study was made up of weather observations gathered from various weather stations positioned in a certain area. The data set contains details like the observational period (0–23 Hours), station name, day, and month. Any information that could be used to identify an individual was kept undisclosed to safeguard their privacy as well as the privacy of the participating weather stations. Automated weather monitoring devices that were deployed at each station were used for data collecting. At regular intervals throughout the day, observations were taken, resulting in a complete dataset spanning a certain amount of time. The data was validated and cleaned during the preprocessing stages to guarantee its accuracy and dependability. Without revealing precise time values or specific dates, the table summarizes the temporal features of the meteorological data collection. The time of observation of (0–23 hours), the day of the month is rendered to (0–31), and the month of the year is (Jan–Dec) which are all indicated.

The data obtained maintained its privacy while enabling broad temporal analysis. It is also worth noting that the station names were omitted from this example to protect the privacy and confidentiality of the concerned weather stations. This example emphasized how to present time-related information inside the data set while maintaining data privacy, therefore precise measurements or weather factors were not included.

2.4 Diagrams

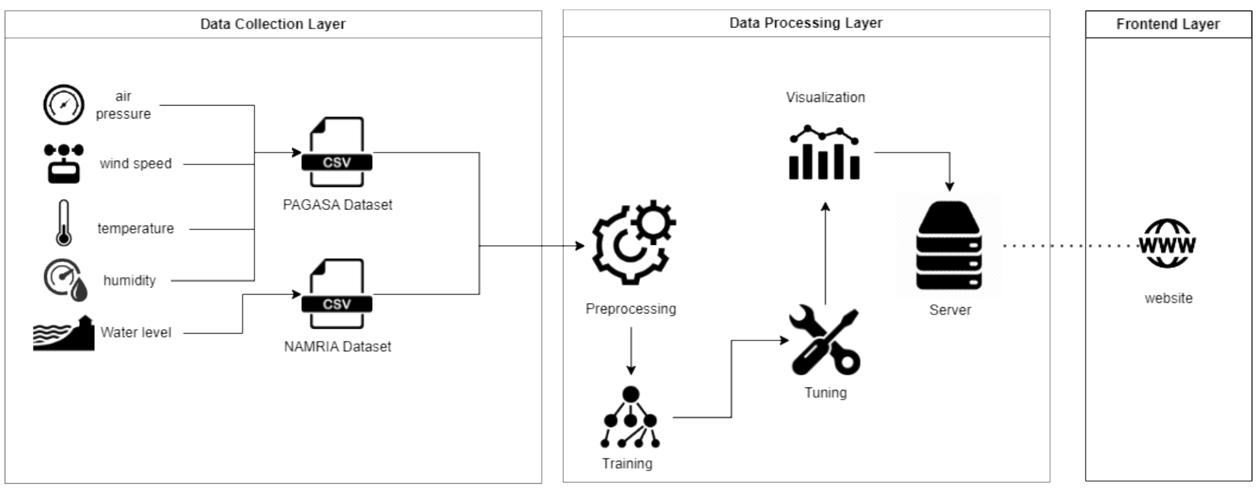


Fig.1: Machine Learning Pipeline. The system architecture of the research project is depicted in Figure 1, illustrating three distinct layers: the Data collection layer, Data processing layer, and Front-end layer. The data collection layer plays a crucial role in acquiring essential information from sensors within the Internet of Things (IoT) network, encompassing variables such as air pressure, wind speed, temperature, humidity, and water level. Subsequently, the collected raw data is transformed into a comprehensive dataset, which is further augmented by integrating it with two additional datasets obtained from PAGASA and NAMRIA. The resulting dataset is then directed to the data processing layer, where it undergoes a series of essential procedures, including preprocessing, training, tuning, cleaning, and visualization.

These processes were employed to enhance the quality and utility of the data. Following the processing phase, the data was transmitted to the server for storage, ensuring its availability for future reference. Eventually, the processed data was presented to users through the Front-end layer, where it was displayed on the project website, enabling convenient access and exploration. The outlined system architecture highlights the sequential flow of data from collection to processing and finally to the front-end presentation. This comprehensive framework ensures effective utilization of the collected data and facilitates the dissemination of valuable information to users.

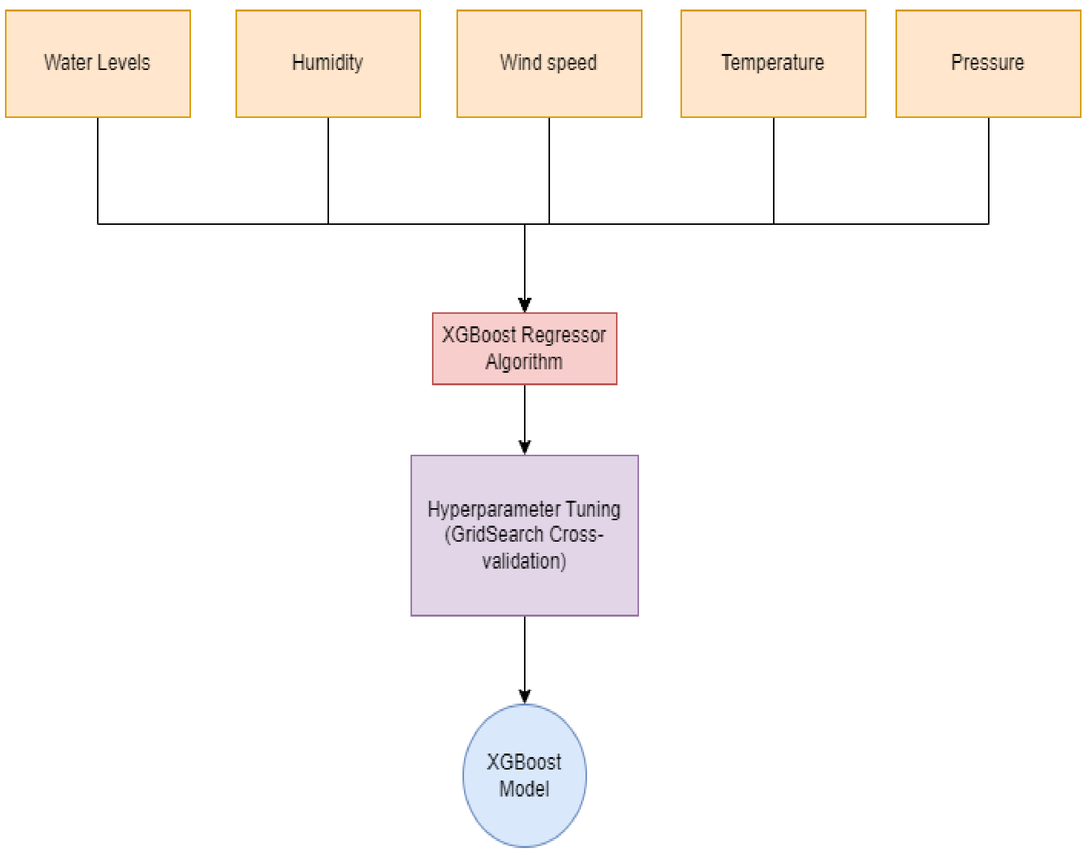


Fig. 2: Machine Learning Model Training Structure. Figure 2 illustrates the representation of five crucial parameters, namely Water levels, Humidity, Wind speed, Temperature, and Pressure, along with the training process of the model.

These mentioned features are subsequently inputted into an Extreme Gradient

Boosting Regressor algorithm to compute the model weights essential for prediction generation. To ensure optimal performance, the researchers employed cross-validation techniques utilizing the GridSearchCV python library. The outcomes of this process can be referred to in the provided table, which presents the results obtained from the model's evaluation and parameter optimization.

Table 4: Hyperparameters

Hyperparameter	Value	Description
colsample_bynode	1	Controls the subsampling ratio of columns on each tree node during training process
colsample_bylevel	1	Controls the subsampling ratio of columns on each level of tree nodes during training process
colsample_bytree	1	Controls the subsampling ratio of columns on each tree during training process
Subsample	1	Subsample ratio of the training instance
Booster	gbtree	Specifies which booster to use
grow_policy	depthwise	Tree growing policy
Gamma	0.2	Minimum loss reduction required to make a further partition on a leaf node of the tree.
Lambda	1	Regularization term on weights (xgb's lambda)
importance_type	gain	-
max_leaves	50	Sets the maximum number of leaves in each tree

max_depth	6	Sets the maximum level of each tree
eval_metric	rmse	Sets the evaluation metric to be used for monitoring the training process for early stopping rounds.
tree_method	hist	Sets the tree method to use
early_stopping_rounds	500	Sets the number of negative rounds to allow before stopping the training process

Table 4 depicts the hyperparameters used in training the XGBoost model. These are the results generated from using the GridSearchCV.

2.5. Data Analysis Plan

Preprocessing. NAMRIA and PAGASA datasets went through preprocessing using the Pandas library in the Python programming language, however, NAMRIA datasets were converted to .csv file using Nom Parser library. The primary objective of this preprocessing was to remove any rows from the datasets that contain missing data. For the PAGASA dataset, only the columns corresponding to the specified features will be retained. The two datasets were then concatenated and paired using datetime objects in order to ensure that both datasets sync together.

Training. The model was developed using the Extreme Gradient Boosting library in the Python programming language with mean squared error validation. Regression was the primary focus of the model. Hyperparameter tuning was also performed using GridSearchCV, a function from the Scikit Learn library, to identify the optimal configuration of parameters for tuning the model.

Testing. To conduct a rigorous analysis, the research protocol required a minimum of seven (7) consecutive days for the data to be collected. To achieve this, the last seven (7) rows of the dataset was used as test data, while the remaining rows were used for training the model. The performance of the model was evaluated by calculating the mean squared error of the test result vectors, a widely accepted metric for measuring the accuracy of regression models.

Analysis of Results. The data was analyzed using the matplotlib library from python and compared the different columns from the actual and predicted readings. The analyzed data was provided as visuals.

2.6. User Interface of the Web Application

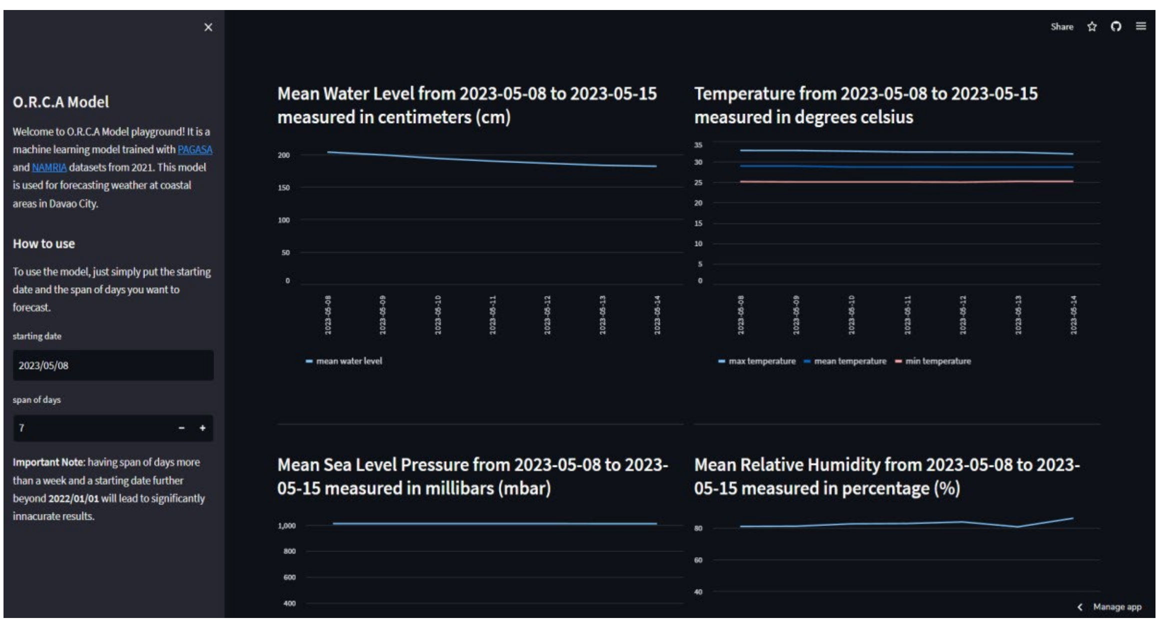


Fig. 3: Website Main Page. This figure displays the results of a 300-day prediction range. It showcases the environmental parameters and their corresponding predictions in the form of graphical representations. Users have the ability to input the desired starting date and specify the desired number of days through the interface located on the left side of the website.

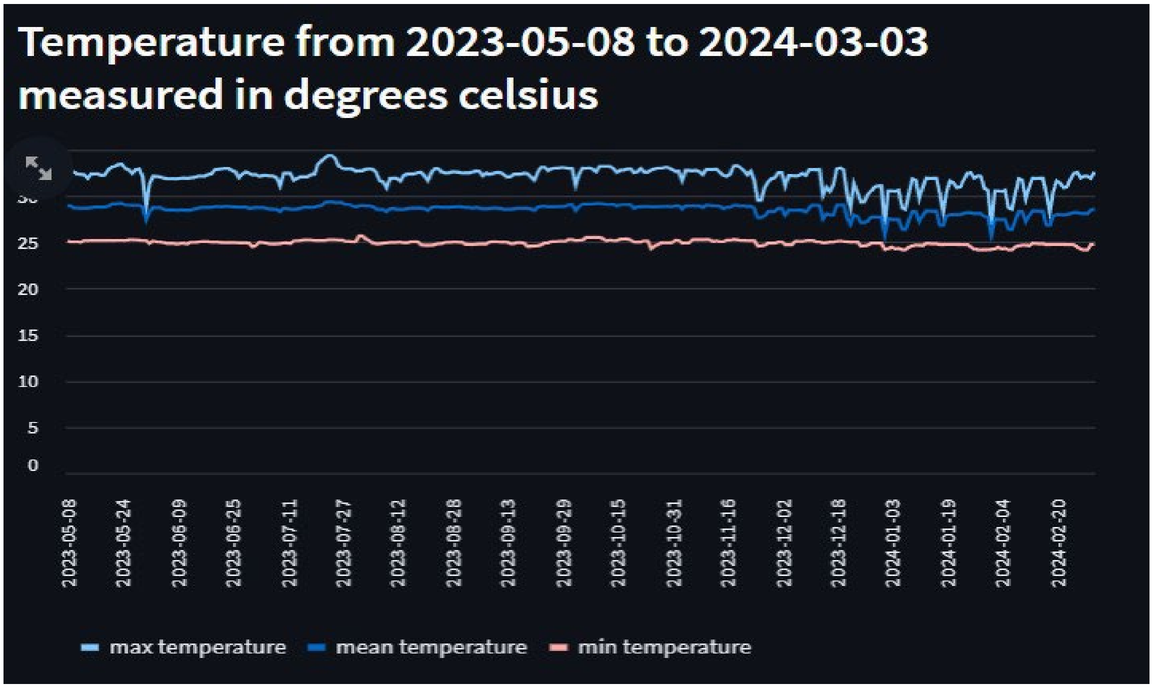


Fig. 4: Sample Predicted Graph. This figure depicts a sample graph representing the predicted temperature readings. This graph specifically focuses on the temperature parameter, displaying data from May 08, 2023, to March 03, 2024, measured in degrees Celsius. The graph distinguishes three lines: a light blue line indicating the maximum temperature, a dark blue line representing the mean temperature, and a pink line denoting the minimum temperature.

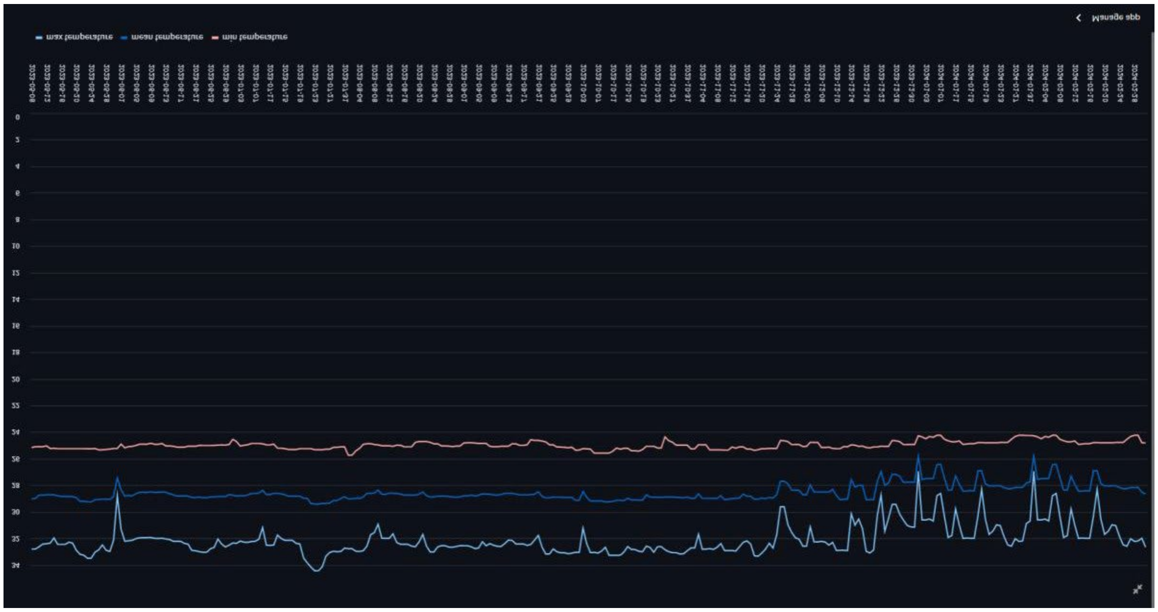


Fig. 5: Zoom Specification Function. This figure showcases that the graphs can be zoomed in to see the specific details of the graphs.

2.7. Datasets for Training and Testing

	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	...	19:00	20:00	21:00	22:00	23:00	station_name	day	month	year	date
0	109	96	106	135	170	207	240	258	258	240	...	284	278	242	198	156	DV	1	1	2021	2021-01-01
1	117	97	95	115	149	187	223	249	261	250	...	279	283	265	229	186	DV	2	1	2021	2021-01-02
2	144	116	102	106	130	168	204	236	257	261	...	259	277	272	246	213	DV	3	1	2021	2021-01-03
3	176	145	123	118	129	151	189	221	247	256	...	231	257	264	253	228	DV	4	1	2021	2021-01-04
4	199	166	143	126	125	141	171	200	228	250	...	204	227	248	250	245	DV	5	1	2021	2021-01-05
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
360	217	199	177	157	147	149	153	169	189	209	...	172	186	203	214	223	DV	27	12	2021	2021-12-27
361	226	218	203	184	166	158	155	162	176	193	...	161	164	175	194	212	DV	28	12	2021	2021-12-28
362	225	231	225	215	199	180	168	158	159	175	...	150	142	143	155	176	DV	29	12	2021	2021-12-29
363	200	216	228	228	220	204	182	164	153	155	...	159	136	126	127	147	DV	30	12	2021	2021-12-30
364	167	192	218	233	237	230	213	186	164	150	...	182	148	121	106	109	DV	31	12	2021	2021-12-31

365 rows × 29 columns

Fig. 6: Preprocessed Data from NAMRIA. Figure 6 shows the preprocessed data obtained from NAMRIA exhibits minimal differences from the original dataset, as the primary objective was to convert the file format into comma-separated values (CSV). To accomplish this, the researchers employed the nom library to construct a parser capable of converting the dataset.

	pressure	max temp.	min temp.	mean temp.	mean rel. humidity	wind speed	date
0	1009.6	26.4	23.7	25.1	86.0	0.002	2021-01-01
1	1009.1	31.3	24.4	27.9	85.0	0.002	2021-01-02
2	1009.3	30.8	24.9	27.7	86.0	0.001	2021-01-03
3	1009.7	29.6	23.3	26.4	86.0	0.002	2021-01-04
4	1008.2	31.2	24.8	28.0	81.0	0.002	2021-01-05
...	...	...	...	...	...	...	...
360	1011.7	30.2	24.1	27.2	76.0	0.003	2021-12-27
361	1012.2	29.1	24.0	26.6	85.0	0.003	2021-12-28
362	1011.5	29.8	25.6	27.7	78.0	0.004	2021-12-29
363	1011.2	33.0	25.0	29.0	77.0	0.003	2021-12-30
364	1011.8	30.4	24.7	27.6	87.0	0.003	2021-12-31

Fig. 7: Preprocessed Data from PAGASA. The original PAGASA dataset lacks clear values, such as dates, and includes various features such as wind direction, dry bulb, dew point, among others.

To address these issues, the dataset underwent transformation by removing unnecessary features and adding a date column to each row. The preprocessing of the dataset was accomplished using the Pandas Library, a Python tool specifically designed for transforming comma-separated values (CSV) data.

	pressure	max temp.	min temp.	mean temp.	mean rel. humidity	wind speed	average water level	day	month	year
0	1009.6	26.4	23.7	25.1	86.0	0.002	178.294118	1	1	2021
1	1009.1	31.3	24.4	27.9	85.0	0.002	174.705882	2	1	2021
2	1009.3	30.8	24.9	27.7	86.0	0.001	176.176471	3	1	2021
3	1009.7	29.6	23.3	26.4	86.0	0.002	183.294118	4	1	2021
4	1008.2	31.2	24.8	28.0	81.0	0.002	189.000000	5	1	2021
...	...	...	...	...	...	...	...	...	...	...
324	1011.7	30.2	24.1	27.2	76.0	0.003	193.058824	27	12	2021
325	1012.2	29.1	24.0	26.6	85.0	0.003	200.823529	28	12	2021
326	1011.5	29.8	25.6	27.7	78.0	0.004	208.941176	29	12	2021
327	1011.2	33.0	25.0	29.0	77.0	0.003	207.470588	30	12	2021
328	1011.8	30.4	24.7	27.6	87.0	0.003	206.882353	31	12	2021

329 rows × 10 columns

Fig. 8: Finalized dataset from NAMRIA and PAGASA. The finalized dataset consists of extracting the minimum, maximum, and mean temperature from PAGASA dataset and getting the average of water level from NAMRIA dataset. The figure above shows the actual dataset to be used for training and testing.

	pressure	max temp.	min temp.	mean temp.	mean rel. humidity	wind speed	average water level	day	month	year
0	1009.6	26.4	23.7	25.1	86.0	0.002	178.294118	1	1	2021
1	1009.1	31.3	24.4	27.9	85.0	0.002	174.705882	2	1	2021
2	1009.3	30.8	24.9	27.7	86.0	0.001	176.176471	3	1	2021
3	1009.7	29.6	23.3	26.4	86.0	0.002	183.294118	4	1	2021
4	1008.2	31.2	24.8	28.0	81.0	0.002	189.000000	5	1	2021
...	...	...	...	...	...	...	...	...	...	...
317	1011.3	33.5	25.7	29.6	71.0	0.003	184.529412	20	12	2021
318	1010.9	29.3	25.1	27.2	81.0	0.003	174.941176	21	12	2021
319	1011.3	27.0	24.0	25.5	89.0	0.002	173.235294	22	12	2021
320	1009.7	32.0	24.9	28.4	74.0	0.003	173.294118	23	12	2021
321	1010.1	30.0	25.0	27.5	80.0	0.003	177.411765	24	12	2021

322 rows × 10 columns

Fig. 9: Training Dataset

	pressure	max temp.	min temp.	mean temp.	mean rel. humidity	wind speed	average water level	day	month	year
322	1010.3	33.5	23.5	28.5	74.0	0.004	179.588235	25	12	2021
323	1011.1	32.5	25.0	28.8	74.0	0.002	185.176471	26	12	2021
324	1011.7	30.2	24.1	27.2	76.0	0.003	193.058824	27	12	2021
325	1012.2	29.1	24.0	26.6	85.0	0.003	200.823529	28	12	2021
326	1011.5	29.8	25.6	27.7	78.0	0.004	208.941176	29	12	2021
327	1011.2	33.0	25.0	29.0	77.0	0.003	207.470588	30	12	2021
328	1011.8	30.4	24.7	27.6	87.0	0.003	206.882353	31	12	2021

Fig. 10: Testing Dataset

The training model consisted of using all the data excluding the most recent, seven (7) days of data, this was to ensure that the data obtained from model was accurate. The splits of the data are shown in figures nine (9) and ten (10).

2.8. Limitations of the Methodology

The primary limitation of the methodology employed in this research was the possibility that the output of the model may not include important features that are presently inaccessible to the researchers. This may be due to financial constraints and the quality of the datasets provided by the relevant authorities, which restricts our control over the necessary scope and parameters and may potentially compromise the accuracy of our findings. Additionally, several factors that can influence water levels and atmospheric changes in coastal areas exist beyond the scope of the datasets.

2.8.1. Security

The security limitations of an IoT monitoring device that relied on Arduino and Raspberry Pi technology extended beyond the vulnerabilities of the platforms themselves. The elements, such as heat, rain, and other environmental factors, might affect these devices. This vulnerability was made even worse by the device compartment's principal material being plastic. The plastic covering might not offer enough protection in bad weather, such as intense heat or heavy rain, potentially jeopardizing the device's performance and, consequently, its security. Electronic components' performance can be impacted by moisture or high heat, which can result in data loss, system errors, or even physical damage. Furthermore, the reliance on plastic as the primary material may make the device more susceptible to the physical tampering of outside forces and other factors such as human interactions, unauthorized access, or theft. In order to mitigate potential risks and maintain the overall security of the system, it is important to take into account these limitations when deploying IoT devices and to take the necessary precautions, such as using weather resistant enclosures or implementing additional physical security measures.

2.8.2. Maintenance

Maintenance of an IoT monitoring device that utilizes Arduino and Raspberry Pi technology comes with inherent limitations, further compounded by potential financial challenges. the need for regular updates and patches to ensure compatibility and security may incur additional costs, especially for devices deployed in large-scale IoT deployments. The physical components of the device, such as sensors, etc., may also require periodic maintenance, constant calibrating, or replacement to maintain accurate readings and analysis, which can impose financial constraints. Financial constraints are also an issue since this research is not funded by any organization. Moreover, in remote or inaccessible locations, the cost of physically accessing and servicing the device may be significant, further burdening maintenance efforts. These financial limitations highlight the importance of budgeting and resource allocation to accommodate ongoing maintenance needs and ensure the longevity and reliability of IoT monitoring devices based on Arduino and Raspberry Pi technology.

2.9. Ethical Considerations

The following ethical considerations were meticulously considered during the development of the system:

Social Value. The results of this research may help government facilities if projects related to the study are to be conducted. This also benefits society since this research may help prevent risks and hazards that may befall the populus near these coastal areas.

Informed Consent. The researchers sent an informed consent to the NAMRIA and PAGASA officers to ask for datasets regarding coastal tidal levels and atmospheric attributes.

Potential Risk. This research does not present any risk to the officers as no personal information was required and no clinical trials were conducted.

Privacy and Confidentiality. The data received from the NAMRIA and PAGASA offices were held confidential and was not divulged to the public as this dataset was strictly for the researchers and the officers of the corresponding offices.

Adequacy of Facilities. The study was conducted on computers and IoT sensor devices that are apparent and available to the researchers.

Transparency. The research was fully accessible to the NAMRIA and PAGASA, no data withheld and was presented to the data source.

Qualifications of the Researchers. The researchers of this study are Computer Science students in Mapua Malayan Colleges Mindanao, the researchers were qualified to conduct this as this was intended for the safety of our fellow Davaoeños. Furthermore, the researchers hoped to help the community avoid future calamities that can stem from high-level coastal tide waters.

Community Involvement. This research may affect the safety of the communities near coastal regions that were affected by sea level rising which can create coastal floods affecting their homes and lifestyle, however, this study may also be conducted in areas near bodies of water such as rivers, among others.

3. Results and Analysis

Table 5: Datasets Results from Extreme Gradient Boosting Algorithm

Metric	Value
Time	40.20 seconds
Mean absolute error	1.96
Mean squared error	11.08
Mean absolute percentage error	7.27%

Table 5 shows the training outcomes derived from the dataset, which contains 322 rows of data and was analyzed using the Extreme Gradient Boosting algorithms. It can be observed that the algorithm was able to perform well even with slow learning rate (0.001) and specific hyperparameters (based on Table 4).

Presented below are the overall results of the predictions:

```
In [12]: actual
```

Out[12]:

	average water level	pressure	max temp.	min temp.	mean temp.	mean rel. humidity	wind speed
1	179.588235	1010.3	33.5	23.5	28.5	74.0	4.0
2	185.176471	1011.1	32.5	25.0	28.8	74.0	2.0
3	193.058824	1011.7	30.2	24.1	27.2	76.0	3.0
4	200.823529	1012.2	29.1	24.0	26.6	85.0	3.0
5	208.941176	1011.5	29.8	25.6	27.7	78.0	4.0
6	207.470588	1011.2	33.0	25.0	29.0	77.0	3.0
7	206.882353	1011.8	30.4	24.7	27.6	87.0	3.0

```
predicted
```

Out[20]:

	average water level	pressure	max temp.	min temp.	mean temp.	mean rel. humidity	wind speed
1	179.315704	1009.023254	28.588589	24.498663	26.246571	85.498390	2.854224
2	186.984848	1009.121521	28.582514	24.220619	26.233917	85.214493	2.953953
3	191.602463	1009.121521	29.844549	24.384916	26.888529	81.482674	2.950247
4	195.797516	1009.121521	30.307766	24.823683	27.621746	82.327164	2.960642
5	199.280441	1008.972168	30.784542	24.435482	27.498632	81.467468	2.952622
6	205.462540	1008.646423	31.385801	24.480093	27.803154	77.675903	2.944982
7	205.462540	1008.646423	31.220478	24.477350	27.711226	77.048210	2.962551

Fig. 11: Predicted and Actual Results Comparison. This figure illustrates the comparison between the actual data results and the predicted results generated by the Internet of Things (IoT) device. The figure provides a visual representation of the accuracy and reliability of the predictive model.

By juxtaposing the actual data and the predicted values, the figure offers a comprehensive overview of the performance of the IoT device in forecasting the desired outcomes. The comparison allows for a detailed examination of any discrepancies or similarities between the predicted and actual results.

Table 6: Overview of Results

Data	Actual	Predicted	Mean Absolute Error (%)
Average Water Level	179.58 - 206.88	179 - 205	1.52%
Pressure	1010.3 - 1011.8	1009 - 1008	0.24%
Max Temperature	33.5 - 30.4	28 - 31	6.13%
Min Temperature	23.5 - 24.7	24.49 - 24.47	2.79%
Mean Temperature	28.5 - 27.6	26.24 - 27.71	3.87%
Humidity	85 – 87	85.49 - 77.04	8.26%
Wind Speed	4 – 3	2.85 - 2.96	8.26%

The weather data utilized in this study was obtained from weather stations located at Sta. Ana Port, spanning a duration of 7 days. Data collection took place between 10:00AM-4:00PM each day, encompassing measurements of mean sea level pressure, maximum temperature, minimum temperature, mean temperature, wind speed, and mean relative humidity. Stringent preprocessing procedures were implemented to ensure the accuracy and dependability of the data.

Moreover, predictive models were developed employing XGBoost based on the collected data, with the aim of generating predictions. These models employed machine learning algorithms to generate anticipated readings for each weather parameter. To assess the accuracy of the utilized model, a comparison was conducted between the actual readings and the predicted readings.

The outcomes of the comparison indicated a close alignment between the predicted and actual readings for mean sea level pressure and water level, with mean absolute errors of 0.24 and 1.52 percent, respectively. However, a slightly higher disparity was observed between the actual and predicted readings for maximum temperature, minimum temperature, mean temperature, wind speed, and humidity, with mean absolute errors ranging from 2.79% to 8.26%. These findings imply that the predictive models exhibited satisfactory performance in forecasting specific weather parameters. Nonetheless, further refinement is required to enhance the accuracy of predictions for other parameters. The comparison between the actual and predicted readings offers valuable insights into the effectiveness of the predictive models and highlights potential areas for improvement in future forecasting endeavors.

4. Recommendation for Further Study

The deployment of the IoT-based system demonstrated significant potential for analyzing and providing real-time readings of environmental factors. With further refinement and deployment, there is a promising opportunity to integrate this system into various industries, particularly those that rely on weather-sensitive operations. In conclusion, this project has showcased the potential of IoT in environmental monitoring, analysis, and management. To enhance the system's performance, accuracy, reliability, and durability, the researchers recommend improving the quality of the equipment used. This entails incorporating cheaper sensors and more robust communication devices. Additionally, it is crucial to ensure that the enclosure housing the devices offers better security for environmentally sensitive components.

Furthermore, the researchers suggest extending the deployment time to gather more data and gain a deeper understanding of the system's performance over longer periods. Prolonged deployment allows researchers to identify and address obstacles, challenges, and unexpected issues that may arise. By incorporating these recommendations, the IoT-based monitoring system can achieve improved performance, quality, and sustainability, facilitating its further evolution and advancement.

The machine learning model developed using the Extreme Gradient Boosting (XGBoost) algorithm, in conjunction with datasets from NAMRIA and PAGASA, has demonstrated promising viability and achieved favorable performance scores, particularly in the water level and pressure features. Notably, a significant finding of this study was the successful integration of the two datasets from PAGASA and NAMRIA, enabling the generation of predictions for coastal weather conditions.

The XGBoost algorithm proved to be efficient and speedy in training, with training times as low as less than fifteen seconds for one year's worth of data. This efficiency provides possibilities for iterative daily predictions instead of weekly predictions, as well as the implementation of hyperparameter tuning for improved performance tracking and consistently achieving the optimal combination of parameters in each prediction. Furthermore, the research objectives included creating a machine learning model using XGBoost, performing cross-validation with various evaluation metrics and learning rates, and developing a website to display prediction results. These objectives were successfully accomplished, as evidenced by the creation of the machine learning model using NAMRIA and PAGASA datasets, the execution of cross-validation with metrics and learning rates, and the creation of a website displaying sample graphs and predictions.

The study evaluated the accuracy of predictive models for different weather parameters. The findings revealed that the predicted readings for mean sea level pressure and water level closely aligned with the actual readings, exhibiting low mean absolute errors. However, there were notable deviations between the predicted and actual readings for maximum temperature, minimum temperature, mean temperature, wind speed, and humidity, indicating a need for further refinement to enhance the accuracy of predicting these parameters.

In conclusion, this research contributes valuable insights into the effectiveness and potential applications of the machine learning model developed using XGBoost and the integration of NAMRIA and PAGASA datasets. The study acknowledges the need for ongoing research and development to refine and improve the model, including the exploration of additional features and the incorporation of more datasets. The model and website hold promise as valuable tools for predicting hydrological patterns and mitigating risks for coastal regions, making further refinement of the system imperative to support diverse industries in the future.

REFERENCES

Shen, C., Chen, X., & Laloy, E. (2021, April 1). Editorial: Broadening the use of machine learning in Hydrology. Frontiers. https://www.frontiersin.org/articles/10.3389/frwa.2021.681023/full

Yao, Z., Karmakar, M., Fernandes, M. J., Cazenave, A., Jiang, L., Ozyavas, A., Gharibreza, M., Sharifi, M. A., Zhang, Z.-Z., Han, S.-C., Kusche, J., Chen, J. L., Loomis, B. D., Armitage, T. W. K., Bon, R., Kouraev, A. V., & Karimi, A. A. (2021, January 15). Monitoring of Caspian Sea-level changes using Deep Learning-based 3D reconstruction of Grace Signal. Measurement. https://www.sciencedirect.com/science/article/abs/pii/S0263224121000403

Zhang, J., Zhang, D., Yaseen, Z. M., Yarar, A., Wu, H., Kisi, O., Khatibi, R., Kebede, S., Jiang, Y., Graf, R., Burgsteiner, H., Afan, H. A., Aksoy, H., … Hopfield, J. J. (2020, March 12). Forecasting of water level in multiple temperate lakes using machine learning models. Journal of Hydrology. https://www.sciencedirect.com/science/article/abs/pii/S0022169420302791

Zhu, S., Lu, H., Ptak, M. et al. Lake water-level fluctuation forecasting using machine learning models: a systematic review. Environ Sci Pollut Res 27, 44807–44819 (2020). https://doi.org/10.1007/s11356-020-10917-7

Choi, C., Kim, J., Han, H., Han, D., & Kim, H. S. (2019, December 26). Development of water level prediction models using machine learning in wetlands: A case study of upo wetland in South Korea. MDPI. https://www.mdpi.com/2073-4441/12/1/93

KlompenburgT., V., SirsatM.S., MengL., GorelickN., GaoB.C., AlibabaeiK., BergstraJ., BóbedaG.R., BreimanL., BureauU.C., CaoJ., ChambersR.G., Charoen-UngP., Fournier-VigerP., KhakiS., KolesnikovA., & LouppeG. (2022, September 24). Extreme gradient boosting for yield estimation compared with deep learning approaches. Computers and Electronics in Agriculture. https://www.sciencedirect.com/science/article/abs/pii/S0168169922006548

D. H. Nguyen, X. Hien Le, J. -Y. Heo and D. -H. Bae, (2021) "Development of an Extreme Gradient Boosting Model Integrated With Evolutionary Algorithms for Hourly Water Level Prediction," in IEEE Access, vol. 9, pp. 125853-125867, 2021, doi: 10.1109/ACCESS.2021.3111287.

Yang, X., Xiang, Y., Shen, G., & Sun, M. (2022, November 30). A combination model for displacement interval prediction of concrete dams based on residual estimation. MDPI. https://www.mdpi.com/2071-1050/14/23/16025

Liu, W., Liu, W. D., & Gu, J. (2020, May). Predictive model for water absorption in sublayers using a joint ... Predictive model for water absorption in sublayers using a Joint Distribution Adaption based XGBoost transfer learning method. https://www.researchgate.net/publication/338466741_Predictive_model_for_water_absor ption_in_sublayers_using_a_Joint_Distribution_Adaption_based_XGBoost_transfer_learning_method

Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 785–794. https://doi.org/10.1145/2939672.2939785

Zhang, R., Li, B., & Jiao, B. (2019). Application of xgboost algorithm in bearing fault diagnosis - iopscience. Application of XGboost Algorithm in Bearing Fault Diagnosis. https://iopscience.iop.org/article/10.1088/1757-899X/490/7/072062

Guillou, N., & Chapalain, G. (2021, October). Machine learning methods applied to sea level predictions in the upper part of a tidal estuary. Oceanologia, 63(4), 531–544. https://doi.org/10.1016/j.oceano.2021.07.003

Menéndez, M., & Woodworth, P. L. (2010). Changes in extreme high water levels based on a quasi-global tide-gauge data set. Journal of Geophysical Research: Oceans, 115(C10). https://doi.org/10.1029/2009jc005997

Park, S. J., & Lee, D. K. (2020, August 25). Prediction of coastal flooding risk under climate change impacts in South Korea using machine learning algorithms. Environmental Research Letters, 15(9), 094052. https://doi.org/10.1088/1748-9326/aba5b3

Yuan, Z., Liu, J., Liu, Y., Zhang, Q., Li, Y., & Li, Z. (2022, October). A two-stage modelling method for multi-station daily water level prediction. Environmental Modelling &Amp; Software, 156, 105468. https://doi.org/10.1016/j.envsoft.2022.105468

Storm surge due to 2013 Typhoon Yolanda (Haiyan) in Leyte Gulf, the (n.d.).

Retrieved December 18, 2022, from https://www.researchgate.net/publication/281484963_Storm_Surge_Due_to_2013_Typhoon_Yolanda_Haiyan_in_Leyte_Gulf_the_Philippines

Hamka, Z. Yuyun, W., and Lateif, A. (2020, Novermber.). Early warning system for flood disasters using the internet of things. Retrieved December 18, 2022, from https://www.researchgate.net/publication/346371432_Early_Warning_System_for_Flood_Disasters_Using_the_Internet_of_Things

Huber, F., Yushchenko, A., Stratmann, B., & Steinhage, V. (2022). Extreme Gradient Boosting for yield estimation compared with Deep Learning approaches. Computers and Electronics in Agriculture, 202, 107346. https://doi.org/10.1016/j.compag.2022.107346

Garde, P. (2022, April 4). Internet of Things in the Philippines: Everything You Need to Know. IOT Philippines Inc. https://www.iotphils.com/internet-of-things-in-the-philippines-everything-you-need-t o-know/

Barthélemy J, Verstaevel N, Forehead H, Perez P. Edge-Computing Video Analytics for Real-Time Traffic Monitoring in a Smart City. Sensors. 2019; 19(9):2048. https://doi.org/10.3390/s19092048

R.V.L.N., Sridhar & Rao, M V H & Kalyani, Kutagulla & Bhaskar, K. & Chandran, Anand & Mahajan, Monika & Manja, A. & Gouda, Girish & Jalluri D P V, Tayaramma & Amudha, P. & Kandpal, Madan & Pramod, K. & Viswanath, S. & L V, Prasad & a S, Laxmiprasad & Chakraborty, P. & Kamalakar, J. & G. Dr, Nagendra & Viswanathan, M.. (2015). Lyman Alpha Photometer: A far-ultraviolet sensor for the study of hydrogen isotope ratio in the Martian exosphere. Current science. 109. 1114-1120. 10.18520/v109/i6/1114-1120.

Thingiverse.com. (n.d.). Weather station anemometer by LACRIMOSA 2K2.

Thingiverse. Retrieved March 31, 2023, from https://www.thingiverse.com/thing:4355150

APPENDIX A

REVIEW OF RELATED LITERATURE

Research journals and articles presented in this section were taken from MDPI, IEEE Journals, Research Gate, Springer Link, Science Direct, Wiley, and IOP. This chapter discusses topics related to this research endeavor which include IoT Devices, Coastal Areas, Water Levels, Sensors, and Extreme Gradient Boosting Algorithm.

IoT Devices & Monitoring Devices

The research conducted by Odesola et al. (2019) presented a system that utilize various sensors to collect data on temperature, humidity, pressure, and other weather-related factors, and transmit this data to a cloud server for analysis and display. Moreover, the study depicted a smart weather station based on IoT approaches that can read and analyze meteorological measurements at any distinct areas and shows and relay the finalized data to an online web cloud in real-time. The conclusions from this study have an important implication on this research endeavor as the researchers also utilized low-cost microcontrollers. The hardware used or microcontrollers utilized here is reasonably inexpensive and available locally such as ESP-32, DHT11, BMP180, Rain Sensor, Servo, and Light Dependent Resistor (LDR).

The objectives of their study were achieved and their IoT device was able to detect and transmit accurate data coming from sensors to their cloud-based website. Furthermore, the study also proved that using low-cost microcontrollers or hardware are not hindrance nor can affect the data being transmitted and stored.

In another study conducted by Knight et al. (2021), it revealed that coastal areas are prone to hydrological risks such as coastal flooding, and beaches serve as a natural barrier to prevent these risks. This is why the monitoring of the sediments present in these areas should be monitored and the monitoring techniques should always be updated and dynamic since traditional methods are costly and time-consuming. A recent technique called the "waterline" method created by researchers Bell et al. (2016) involved this method to accurately determine intertidal morphology, which is a time series of spatial and temporal images that are equivalent to digital elevation models (DEMs). When combined with data from an IoT tide gauge network, this method can provide near-real-time data and assist in identifying significant changes in sediment distribution, supplementing regional tidal gauge networks for safe port navigation and tsunami/surge detection. There are various ways to measure this such as fixed sensors, remote/mobile sensors, among others.

Some of the available devices can be very expensive, but inexpensive pressure sensors, such as Measurement Specialties' MS5803, 2-bar, and 5-bar pressure sensors, can be used for long periods and produce reliable data when combined with low-cost microcontrollers. These sensors stored pressure measurements on microSD cards internally, and Lyman et al. (2015) used them with low-power microcontrollers to record high-frequency pressure data for extended periods. Moreover, Arshad et al. (2019) agreed that natural hazards such as floods, tsunamis, storm surges, etc. are posing significant threats to lives anywhere in the world. In fact, without proper monitoring and effective mitigation measures, these natural hazards frequently culminate in disasters with serious economic, social, and environmental consequences. The importance of flood monitoring devices and their improvements are high given uncertainties brought about by climate change and the increase of people living in flood-prone areas.

The previously conducted studies discussed in this chapter depict that image processing and computer vision techniques are plausible in estimating water levels (Barthélemy et al., 2019). The IoT sensor data could be fused with data obtained from camera devices for real-time calibration. Studies also revealed that there is a wide range of applications that are feasible for the mapping of floods. They also showed that computer vision is advantageous in covering a wider range as a sensor in terms of finding water level whilst IoT is more accurate but can only present point-based readings. These ways present their shortcomings but can be complemented together by fusing their presented data, thus improving the accuracy of flood monitoring. It also concluded that IOT-based devices are essential in real-time monitoring as they provide instant information about water levels. With this, the researchers, seek to update on the advancements of technologies for flood mapping and to highlight existing solutions to adapt and better manage coastal lagoons.

On the other hand, in Indonesia, which is located between two countries (Asia & Australia) changes its season every six months alternating from rainy to dry seasons (YunYun et al., 2020). In this country, the rainfall depends on the shape of the terrain, direction, wind direction, among others. These rainfall types have patterns including monsoon atterns, equatorial pattern, and local rainfall patterns. The city of Makasar is categorized as a low area with an elevation of 1-4 mass above sea level. This area is often hit by floods, therefore the researchers deemed it necessary to create an early warning system to reduce casualties from these floods and used Wireless Sensor Network (WSN). The WSN is comprised of an ESP8266 MCU node. The nodes were controlled as a data processing unit that uses e-Lua firmware that has a Lua Programming Language. MCU node also supported Arduino IDE software. Data collected by the MCU node was then sent to and subsequently saved on the online server by utilizing Blink's iCloud. In this study, it revealed that the created prototype detected water levels and sent information to a smartphone thru SMS or email. The tests with different water levels also showed that the application operates properly, obtaining an accuracy rate of 93.93% for the simulation.

A study conducted by Al-Fuqaha et al. (2015) entitled "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications" also revealed that the intrusion detection systems (IDSs) are important in safeguarding computer networks from unauthorized access and malicious attacks. Machine learning algorithms, such as Hidden Markov models (HMMs) and decision tree classifiers (DTCs), are commonly used in IDSs to detect network anomalies and identify security threats. HMMs are statistical models that can capture the temporal dependencies of a process, making them effective in detecting network intrusions. DTCs are easy to interpret and can handle both continuous and categorical data, also making them effective in IDSs. Hence, the combination of both HMMs and DTCs showed improved detection performance.

Further, other studies also proposed hybrid IDSs that use HMMs and DTCs to detect network intrusions, with results showing better detection accuracy compared to using HMMs or DTCs alone. Overall, HMMs and DTCs are widely used techniques in IDSs, and combining these methods can improve detection accuracy. Lee et al. (2015) proposed an IDS that uses HMMs and DTCs to detect network intrusions, achieving high detection accuracy.

Further, Munawar et al. (2021) also claimed that modern technology has dramatically altered the world and the way it operates in recent years. In fact, disaster management is one field that is increasingly embracing cutting-edge technology, with both developed and developing countries grappling with flood risks. Climate change, combined with inadequate flood preparedness in many regions around the world, has been shown in predictive analysis to result in a historic level of flood-related damage. As floods become more common around the world, it is critical to identify effective methods of mitigating the risks associated with these disasters. Flood events have resulted in global losses of lives, crops, infrastructure, and economic resources, with flood risks and losses being greater than any other climatic hazard. Image processing techniques like edge detection, segmentation, and pixel analysis are examples, as are machine learning models like ANN, SVM, MLP, and WNN. UAV imaging, SAR, and remote sensing are the most common methods for acquiring images. Existing techniques tend to concentrate on both the pre-and post-disaster phases. In addition to traditional flood management methods such as Global Positioning System (GPS) and Geographic Information System (GIS), complex image processing and machine learning-based approaches have been investigated.

However, the study identified several research gaps, including the lack of application of hybrid models for flood management that combine image processing and machine learning. According to the authors, the study highlighted cutting-edge technology used at different phases of the disaster management lifecycle, as well as the constraints of each technique. Lastly, advanced technologies have become increasingly important in disaster management, especially flood management. This research showed how image processing and machine learning techniques have the potential to transform flood management. Nonetheless, more studies are required regarding the application of hybrid models that combine image processing and machine learning techniques.

Coastal Areas

According to Dhiman et al. (2018), urban areas are important for economic growth and are innovation drivers, they play a very important role in achieving sustainable development. Most of the urban areas are located along coastlines, which are under increasing pressure from both natural and human-caused factors such as fossil fuel combustion, industrial activities, tourism, climate changes, and marine traffic. India is particularly affected by the result of these human and natural effects, including the unpredictable monsoon season, flash floods, and sea-level rise, due to its large coastline and agricultural-dependent population.

As the effects of climate change become more severe, adaptation measures for coastal cities, which are densely populated and vulnerable to extreme weather events, become increasingly important. Because of the shortage of infrastructure capacity and adaptation measures, developing countries are particularly vulnerable to the effects of climate change at the city level. This review article investigates food-induced risks and adaptation strategies in several Indian coastal cities, identifying potential consequences such as riverine and urban flooding, cyclonic storms, tsunamis, tidal inundation, and failures in urban planning and land-use practices.

Water Sensor - JSN-SR04T

The study conducted by Andang et al. (2019) revealed that in order to minimize flooding, it is important to investigate the usage of the JSN-SR04T ultrasonic sensor for river-level detection. The reason for this was due to the unpredictable weather patterns that caused rainfall to be distributed unevenly, as a result, causing flooding and substantial losses. Numerous efforts were made to address this issue, including the application of various methods like hydrological and hydraulic analysis, multilayer neural networks, early detection systems, and ultrasonic sensors. While ultrasonic sensors were used because of their high accuracy, the JSN-SR04T type ultrasonic sensor is advantageous as it shows an accuracy of 1 cm and an effective measurement range of 25 cm to 4.5 m, making it the ideal equipment to be placed in higher and safer locations. The JSN-SR04T ultrasonic sensor, based on the Atmega328 microprocessor, and the BMP085 pressure sensor were employed in the investigation.

The device worked by transmitting an ultrasonic signal in eight steps at a minimum frequency of 40 kHz and measuring the time difference between that signal's transmission and reception. The sensor was positioned four (4) meters above the surface of the river, and the measurements were shown on an LCD in a meter unit. The study determined that the JSN-SR04T ultrasonic sensor could be used efficiently for river-level detection in flood protection and provided the hardware and software setup of the system.

A study that focused on flood monitoring systems conducted by Sixtinah et al. (2021) employed SIM900A and JSN-SR04T ultrasonic sensors to reduce the frequency and area of floods in Indonesia, which are both growing. The study also depicted that rainfall intensity, which might result in a river's or a drainage channel's capacity being exceeded, is what causes flooding. However, installing floodgates in the river flow allowed for the monitoring of river water level, although measuring the water levels manually revealed to be unnecessary resulting to the need to employ automatic design of river water level. As a result, although HC-SR04 ultrasonic sensors were frequently employed for water level monitoring, their performance is negatively impacted by rainstorm splashes. A better choice is the JSN-SR04T, which can detect objects up to a distance of six meters and can be installed higher and more securely in the event of rain. Furthermore, it created a flood monitoring system that could gauge water levels and provide SMS alerts. The designed tool set for flood disaster mitigation was tested on river flow.

In order to reduce water waste and physical labor, Palaghat Sai et al. (2017) conducted a study that addressed the issue of water scarcity in India and the necessity to optimize irrigation. The suggested system optimized Internet of Things (IoT) to remotely monitor soil moisture levels in the field and tank water levels. The system connected the sensors so they can read and respond using Arduino and the ESP8266 Wi-Fi module. The system also employed a waterproof ultrasonic JSN-SR04T sensor to monitor soil moisture and a soil moisture sensor to detect water levels. It also featured a mobile app that enables farmers to remotely check the water levels and manage the water supply from their smartphones, wherever they are, and without the need for human interaction. The system's goal was to increase irrigation efficiency and decrease water waste to improve farmers' quality of life. The findings revealed that the suggested Internet of Things-based smart irrigation system was a reliable and efficient approach to keeping an eye on water levels and adjusting irrigation.

The research conducted by Bordoloi and Shukla (2021). Plans in creating an IoT-based irrigation system that uses the JSN-SR04T sensor to monitor the moisture content of the water storage container and soil and activate sprinklers when the moisture level drops below a certain threshold. It aimed to address India's water shortage and increase agricultural productivity. Further, it utilized an Arduino board, an ESP8266 Wi-Fi module, and a smartphone app to construct a wireless network, connect to the irrigation system, and control the water supply remotely.

The study suggested a cutting-edge upgrade to the existing irrigation system, which alerts farmers via an alarm when the water level in the tank reaches a certain threshold. The new technique completely avoids the danger of water waste that could result from this strategy if the warning doesn't sound. The proposed system can also control irrigation systems, preventing the land from being over-irrigated and boosting agricultural yield. The waterproof JSN-SR04T sensor is capable of detecting water up to a distance of 4.5 meters. This Internet of Things-based irrigation system was a creative response to India's water deficit crisis and can be applied in other regions with comparable problems.

Hydrological Risk

The research conducted by Chaopeng Shen et al. (2021) demonstrates that the emergence of deep learning has sparked a notable increase in the application of machine learning techniques in various hydrology domains. Specifically, long short-term memory (LSTM) has been utilized as a dynamic modeling tool for predicting soil moisture and streamflow. Machine learning enables the construction and training of models that can effectively capture and replicate spatial and temporal patterns present in datasets. The fundamental principle underlying deep learning is to minimize human involvement in feature design while maximizing information extraction from the available data. The integration of machine learning in hydrology has led to enhanced prediction accuracy, not by incorporating process-based assumptions into the models, but by extensively training them with large volumes of data.

Within the hydrology community, there exists potential for exploring the immense power of vast datasets in different subdomains of the field. This can be achieved through various avenues, including the incorporation of physics-based knowledge into machine learning models, improving the interpretability of these models, developing integrated physics-informed neural networks, quantifying and propagating uncertainty in model outcomes, establishing standardized benchmark training datasets, and establishing a collaborative computational platform for the community. The creation of publicly available benchmark training datasets holds particular importance in advancing machine learning applications in Earth science domains.

Furthermore, there is an opportunity to enhance the utilization of remote sensing products that capture components of the water cycle with increasingly higher spatial and temporal resolutions. To ensure success, it is crucial to leverage open-source resources provided by federal agencies, while also implementing standardized data management practices to enhance data usability.

Water Level

The scholarly work authored by Sorkhabi et al. (2021) underscores the crucial role of water displacement and sea-level monitoring in effective water resource management and environmental conservation. The advent of GRACE satellites in 2002 revolutionized the study of mass transfer on a global scale, with a primary focus on acquiring comprehensive geopotential models and analyzing temporal variations in the geopotential field. This paper aims to investigate changes in the Caspian Sea level (CSL) utilizing GRACE data while mitigating errors arising from signal leakage. To achieve this objective, the authors employ the DL method, with additional validation performed using SA data. Furthermore, the study examines the historical trend of CSL and compares its observations with data obtained from different sensing devices. The implementation of the DL method effectively reduces errors, as evident from the outcomes. Furthermore, through cross-wavelet transform analysis, a significant correlation between annual period phases is revealed. The CSL exhibits a recurring pattern of rising and falling amplitudes, persisting for over two centuries.

The findings indicate that CSL experienced a decrease of approximately 70 0.24 cm between 2005 and 2016. Additionally, the DLRGTS demonstrates better compatibility with other sensing devices compared to FM, with an average SC (statistical coefficient) of 82%. Finally, the Volga River discharge aligns with observations from other sensors and displays consistent annual behavior in tandem with CSL.

Senlin et al. (2020) emphasized the vital importance of lakes as natural resources that significantly contribute to economic and societal development. Water levels in lakes serve as crucial indicators of their physical state, and fluctuations therein have substantial impacts on lake ecosystems. Such fluctuations can lead to detrimental effects on keystone species, facilitate the spread of invasive species, and result in biodiversity loss. Mathematical models, ranging from simple statistical models to complex hydrodynamic models, have been employed to simulate and forecast changes in lake water levels. However, the utility of complex hydrodynamic models is limited in regions with scarce data due to their reliance on extensive input data.

In recent years, the field witnessed the emergence of machine learning models, which have been effectively utilized in hydrological and environmental research to address this data limitation. Deep learning networks, such as the long short-term memory (LSTM) recurrent neural network, have demonstrated successful applications in forecasting various hydrological time series, including lake water level predictions. The study suggests that DL models hold promise for forecasting water levels in specific lakes, such as Egirdir Lake, Tongding Lake, and Vrana Lake. Nevertheless, drawing definitive conclusions regarding the performance of DL models for lake water level forecasting is challenging due to the limited focus of individual studies on a single lake.

Further, Hongfang et al. (2020) emphasized the critical importance of lakes as essential water sources for various purposes, including domestic, industrial, agricultural, flood control, and aquaculture. Fluctuations in water levels within lakes and reservoirs significantly impact their ecosystems and flood control efforts. Over time, several mathematical models have been developed to forecast water levels (WL) in lakes, encompassing physically based, stochastic, and system dynamic models. With the advent of artificial intelligence (AI), machine learning (ML) models have gained popularity for WL forecasting in lakes. However, challenges persist in their implementation and interpretation, necessitating a comprehensive review of their applications to foster a better understanding, communication, and advancement. This study also examined the applications of seven distinct types of ML models for forecasting lake water levels, while addressing challenges such as improving model reliability by determining optimal input combinations and capturing water-level dynamics influenced by climate change and extreme events.

The paper was structured into sections that review ML models, discuss methodological aspects and limitations, present the advantages and disadvantages of established ML models, and culminate in future directions for research. Despite the challenges faced, modeling water levels in lakes remains a significant concern actively pursued by researchers worldwide.

According to Choi et al. (2019), wetlands play a crucial role in environmental preservation and ecological maintenance by regulating water levels, among other functions. However, the survival of wetlands is impacted by factors such as flow rate, water level, and inundation depth. To effectively manage and protect wetlands, it is necessary to measure and predict water levels. Despite recognizing the significance of water level measurement and prediction, obtaining long-term wetland data poses challenges due to inadequate monitoring and limited research budgets. In wetland areas like South Korea's Ramsar Designated Wetlands, water level measurement is particularly constrained, and the existing estimation methods require time-consuming calibration.

To overcome these limitations, machine learning-based data-driven models, such as artificial neural networks (ANN), have been applied for water level forecasting. However, research on water level prediction in wetlands is relatively new compared to other fields, with a predominant focus on the use of ANN. The objective of this study is to develop a water level prediction model specifically tailored for wetlands by combining machine learning models like decision trees (DT), random forest (RF), support vector machines (SVM), and ANN. The researchers utilized water level data, weather data, and information from the Mokpo embankment and Sindang Drainage Pump Station. The data was divided into training and test datasets, and various statistical models including DT, RF, SVM, and ANN were employed to create a predictive model for water levels in the Upo wetland. The performance of each model was evaluated using assessment indicators such as correlation coefficient (CC), Nash-Sutcliffe Efficiency (NSE), root mean square error (RMSE), peak value, and prediction interval (PI).

Extreme Gradient Boosting Algorithm (XGBoost)

Apropos to the this study, Florian Huber et al. (2022) focused on comparing the performance of extreme gradient boosting (XGBoost) with deep learning approaches in yield estimation within the context of precision agriculture. The authors aimed to determine which algorithm exhibits superior accuracy, as well as evaluate the training and prediction times of these algorithms. The study utilizes datasets consisting of plant images paired with corresponding yield data.

In the experimental phase, the researchers assessed the performance of the XGBoost algorithm along with deep learning approaches in two distinct scenarios: end-year prediction and in-year prediction. To extract features from the images, a deep learning algorithm (VGG-16) was employed. These extracted features were utilized as input for training and testing various machine learning algorithms, including XGBoost, as well as three deep learning algorithms (AlexNet, GoogleNet, and VGG-16). The performance of these algorithms is compared based on accuracy, training time, and prediction time.

Furthermore, this research endeavor hoped to provide valuable insights into the comparative performance of XGBoost and deep learning approaches for yield estimation in precision agriculture. The results indicate that the XGBoost algorithm demonstrates high accuracy and efficiency, surpassing state-of-the-art deep learning algorithms. Specifically, XGBoost outperforms the deep learning algorithms in terms of accuracy, yielding an average prediction error of 4.76%. In contrast, the best-performing deep learning algorithm (VGG-16) achieves an average prediction error of 7.44%. Additionally, XGBoost exhibits shorter training and prediction times compared to the deep learning algorithms.

According to Nguyen et al. (2022), floods pose significant risks in cities worldwide, and their complexity has increased due to climate change. As a result, there is a pressing need to develop more effective methods for flood prediction. To address this, the researchers employ hybrid models, specifically GA-XGBoost (Genetic Algorithm) and DE-XGBoost (Differential Evolution), to predict the hourly water level in the Jungrang urban basin situated at the Han River in South Korea. The primary objective of this study is to construct and assess the performance and accuracy of these proposed models.

The research findings showed that accurately predicting water levels remains challenging due to the nonlinear and nonstationary characteristics of such data. Nonetheless, the results demonstrate that GA-XGBoost and DE-XGBoost exhibit similarities in their performance and both outperform alternative models such as RF and CART Models. Notably, for a 6-hour lead time prediction, DE-XGBoost demonstrates superior accuracy compared to GA-XGBoost, as evidenced by the generated values. Moreover, GA-XGBoost outperforms the other models when it comes to separate time step predictions. According to Zhang et al. (2022), traditional models commonly used for dam monitoring are inadequate and severely limited, which compromises their robustness and accuracy. To address this issue, the researchers propose the utilization of XGBoost to effectively assess the lag behavior of dam seepage. They develop a monitoring model that incorporates the time lag effect, specifically targeting the uplift pressure of concrete dams. In order to enhance the XGBoost model, a hybridization technique called Hybridizing Grey

Wolf Optimization (HGWO) is applied. The study concludes that HGWO demonstrates superior performance in terms of stability, accuracy, and resistance against overfitting in predicting uplift pressure in concrete dams. Furthermore, when compared to the trial-and-error method, the proposed approach aligns more closely with the actual present situation, indicating its effectiveness.

To corroborate with the findings of previous studies presented, Liu et al. (2020) emphasized the challenge of accurately understanding water absorption in water injection sublayers, primarily due to significant heterogeneity along the longitudinal direction of reservoirs. The specification of an effective water injection scheme faces substantial difficulties. To address this issue, the researchers suggested the utilization of the Extreme Gradient Boosting (XGBoost) model to construct a water absorption prediction model. The objective was to predict water absorption in sublayers that lack a water injection profile. To achieve this, the researchers employed a technique called Joint Distribution Adaptation-based XGBoost transfer learning (JDA-XGBoost). By training the model with the provided datasets, JDA-XGBoost was employed to predict the water absorption of sublayers. The study highlighted that the proposed approach demonstrates acceptable accuracy in predicting the water absorption of sublayers in injectors without an injection profile. Additionally, the results indicated the applicability of transfer learning modeling. The demonstrations highlight the efficacy of the provided dataset in achieving accurate prediction results using the JDA-based approach for XGBoost.

Chen and Guestrin (2016) demonstrated the effectiveness and widespread utilization of tree boosting as a machine learning approach. They pointed out XGBoost as an end-to-end tree-boosting system that is extensively employed by data scientists to achieve state-of-the-art outcomes across various machine learning tasks.

The XGBoost system introduced innovative techniques to address the challenges of sparse data, including a sparsity-aware algorithm and a weighted quantile sketch for approximate tree learning. Additionally, XGBoost offered insights into cache access patterns, data compression, and sharding, facilitating the development of a scalable tree-boosting system. By combining these insights, XGBoost was capable of scaling beyond billions of examples while utilizing fewer resources compared to existing systems. The study emphasized the significance of XGBoost's sparsity-aware algorithm and weighted quantile sketch in handling sparse data. Moreover, it underscores the importance of cache access patterns, data compression, and sharding in constructing a scalable system. The authors note that the lessons learned from the development of XGBoost can be applied to other machine learning systems. By leveraging these insights, XGBoost has demonstrated its ability to effectively address large-scale real-world problems with minimal resource requirements.

To add, the research conducted by Ghang et al. (2019) emphasized the critical role of rolling bearings as vulnerable components in rotating machinery, contributing to approximately 30% of faults. Vibration data collection and feature extraction, particularly utilizing time-domain parameters like kurtosis, are commonly employed methods for detecting and diagnosing these faults. The study introduced the XGBoost algorithm, which encompasses a range of features including a customizable loss function, tree building and pruning, and support for approximate search of splitting points. The authors utilized XGBoost to diagnose rolling bearing faults and compare its performance with various other tree algorithms. The findings illustrated the superiority of XGBoost in terms of both accuracy and efficiency. It also revealed that XGBoost holds significant potential in this field; however, several challenges still need to be addressed. The application of time-domain parameters for fault identification encounters difficulties arising from diverse noise environments, and the adjustment of model parameters assumes critical importance. Thus, further advancements in this field necessitate research tasks that focus on identifying noise-robust parameters, developing techniques for signal denoising, and implementing effective feature extraction methods. Overall, the study underscores the importance of XGBoost in diagnosing rolling bearing faults and highlights the need for continued research to overcome challenges associated with noise environments and parameter adjustments.

APPENDIX B

DEVICE DEPLOYMENT AND SOURCE CODE

Sta. Ana Deployment

IoT Source Code

```
// lcd
#include <LiquidCrystal_I2C.h>

// i2c (BMP280)
#include "i2c_BMP280.h"

// Humidity Sensor
#include "DHT.h"

#define DHT_TYPE            DHT11
#define DHT11_PIN           4
#define ANEMOMETER_PIN      5
#define JSNSR04T_ECHO_PIN   6
#define JSNSR04T_TRIG_PIN   7


// timing
float serialInterval = 60; // s // change this  for timing serial prints
float updateInterval = 1; // s // change this for timing value updates
float LCDInterval = 300; // this is in millis // ms // change this for LCD interval
unsigned long previousSerialSeconds = 0;
unsigned long previousUpdateSeconds = 0;
```

```
unsigned long previousLCDMillis = 0;

// lcd
LiquidCrystal_I2C lcd(0x27,16, 1);
String outputMessage = "                ";
int pos = 0;


// DHT11
DHT dht(DHT11_PIN, DHT_TYPE);
float humidity;


float avgHumidity;

// BMP280
BMP280 bmp280;
float temperature;
float pascal;
float millibar;


float avgTemperature;
float avgPressure;
```

```
// JSNSR04T
long duration;
float distance;

float avgDistance;

// Anemometer
float windspeed;
float signals = 0;
bool hovered = false;

float avgWindspeed;



void setup()
{
  Serial.begin(9600);

  // lcd
  lcd.init();
  lcd.clear();
  lcd.backlight();
```

```
    lcd.leftToRight();



    // BMP280
    bmp280.initialize();
    bmp280.setEnabled(0);
    bmp280.triggerMeasurement();

    // DHT11
    dht.begin();



    // JSNSR04T
    pinMode(OUTPUT, JSNSR04T_TRIG_PIN);
    pinMode(INPUT, JSNSR04T_ECHO_PIN);

    // Anemometer
    pinMode(INPUT, ANEMOMETER_PIN);



    // // init
    // updateValues();

}
```

```
void loop() {


  // ANEMOMETER
  int status = digitalRead(ANEMOMETER_PIN);


  if (status == 0) {

    if (!hovered) {

      signals++;
      hovered = true;
    }

  } else {

    hovered = false;
  }

  // update values

  if ((millis()/1000) - previousUpdateSeconds >= updateInterval) {
```

```
    updateValues();

    previousUpdateSeconds += updateInterval;
}

// send serial values

if ((millis()/1000) - previousSerialSeconds >= serialInterval) {

    printSerialValues();

    previousSerialSeconds += serialInterval;
}



// LCD
if (millis() - previousLCDMillis >= LCDInterval) {



    lcd.clear();
    lcd.setCursor(0, 0);

    lcd.print(outputMessage.substring(pos,  pos+16));
```

```
        if (pos < outputMessage.length()) {
            pos++;
        } else {
            pos = 0;
        }

        previousLCDMillis += LCDInterval;
    }

}


void printSerialValues() {

        // // Output
    Serial.print("temperature:");
    Serial.print(avgTemperature / serialInterval);

    Serial.print("pressure:");
    Serial.print(avgPressure / serialInterval);



    Serial.print("windspeed:");
```

```
    Serial.print(avgWindspeed / serialInterval);




    Serial.print("waterlevel:");
    Serial.print(avgDistance / serialInterval);
    avgDistance = 0;

    Serial.print("humidity:");
    Serial.print(avgHumidity / serialInterval);

    Serial.print("$"); // marker



    avgTemperature = 0;
    avgPressure = 0;
    avgWindspeed = 0;
    avgDistance = 0;
    avgHumidity = 0;
}


void updateValues() {
```

```
// Anemometer
float circumference = (2 * 3.1415926535 * 0.08);
float arc = (circumference) / 3;
float factor = 3;

windspeed = (signals / serialInterval) * arc * factor; //
avgWindspeed += windspeed;

signals = 0; //reset


// BMP280
bmp280.awaitMeasurement();
bmp280.getTemperature(temperature);
bmp280.getPressure(pascal); // millibar

avgTemperature += temperature;

millibar = pascal / 100;
avgPressure += millibar;

bmp280.triggerMeasurement();
```

```
// DHT11
humidity = dht.readHumidity();

avgHumidity += humidity;


// JSNSR04T
digitalWrite(JSNSR04T_TRIG_PIN, LOW);
delayMicroseconds(3);
digitalWrite(JSNSR04T_TRIG_PIN, HIGH);
delayMicroseconds(10);
digitalWrite(JSNSR04T_TRIG_PIN, LOW);
duration = pulseIn(JSNSR04T_ECHO_PIN, HIGH);
distance = duration*0.034/2; // Get the max range of the sensor


avgDistance += distance;


// lcd update
outputMessage = "                ";
outputMessage += "temperature: ";
```

```
    outputMessage += String(temperature);
    outputMessage += " ";
    outputMessage += "pressure: ";
    outputMessage += String(millibar);
    outputMessage += " ";
    outputMessage += "wind speed: ";
    outputMessage += String(windspeed);
    outputMessage += " ";
    outputMessage += "water level: ";
    outputMessage += String(distance);
    outputMessage += " ";
    outputMessage += "humidity: ";
    outputMessage += String(humidity);
    outputMessage += " ";
}
```

Sta. Port Request Letter

MAPÚA
MALAYAN COLLEGES
MINDANAO

CCIS
COLLEGE OF COMPUTER AND INFORMATION SCIENCE
MAPÚA MALAYAN COLLEGES MINDANAO

February 28, 2023

Coast Guard Station
Monteverde Street, Sta. Ana
Davao City

Greetings!

We have embarked on a study involving coastal areas for our thesis titled as *An Internet of Things and Machine Learning Approach of Predicting Hydrological Risk Intensities in Coastal Areas in Davao City*. The mentioned study will make use of machinelearning techniques and IoT technology necessary for forecasting coastal risks in Davao City coastlines.

In line with this, we would like to request an approval of the deployment of our device on the premises located in Sta. Ana. Our device requires us to be near coastal waters and needs to stay for 10 hours a day for at least 7 days within the month of March to satisfy the requirements of our study's methodology.

Here is a quick rundown of our device:

Specifications	
Width	19 cm
Height	200 cm
Power source	Powerbank (5v 2.4 A)
Internet source	Hotspot

We shall highly appreciate your kind approval. We will count this as a major contribution to the success of our study.

Sincerely,

MMCM Student Researchers

Rey Joshua Morales
09166877241
Joshuarmm900@gmail.com

John Mark Bago-od
09662235724
jmabagood@mcm.edu.ph

Franz Brillante
fdBrillante@mcm.edu.ph

Noted by:

GENEVIEVE A. PILONGO, MIT
Course Facilitator

PATRICK D. CERNA, DIT
Research Adviser

ENGR. NEIL P. MAGLOYUAN, MEE, PCPE
Program Head – Computer Science

PAGASA Request Letter

MAPÚA MALAYAN COLLEGES MINDANAO

CCIS COLLEGE OF COMPUTER AND INFORMATION SCIENCE MAPUA MALAYAN COLLEGES MINDANAO

March 4, 2023

Philippine Atmospheric, Geophysical, and
Astronomical Services Administration
Old Airport Compound, Sasa
Davao City

Greetings!

We are Mapua Malayan Colleges Mindanao students of Bachelor of Science in Computer Science and we are conducting a research entitled as *An Internet of Things and Machine Learning Approach of Predicting Hydrological Risk Intensities in Coastal Areas near Davao City.* This research involves using certain meteorological instruments as its core methodology.

With that being said, we would like to request calibration of our teams' custom-made anemometer. This would require us to monitor some of your production-level anemometers, compare it to ours, and find for anemometer factor for it.

We shall highly appreciate your approval. We believe that your assistance is crucial in our study.

Sincerely,

MMCM Student Researchers

Rey Joshua Morales	John Mark Bago-od	Franz Brillante
09166877241	09662235724	fdBrillante@mcm.edu.ph
Joshuamm900@gmail.com	jmabagood@mcm.edu.ph	

Noted by:

GENEVIEVE A. PILONGO, MIT
Course Facilitator

PATRICK D. CERNA, DIT
Research Adviser

ENGR. NEIL P. MAGLOYUAN, MEE, PCPE
Program Head – Computer Science