A PROJECT REPORT

On

AquaAir Insight: Air-Water quality prediction

Submitted by

Shreyansh Shekhar (10800220006)
Rishika Singh (10800220008)
Udayan Choudhary (10800220035)
Sumit Kumar Mihirachariya (10800220042)

Under the Guidance of

Mr. Sudip Kumar De
Assistant Professor, Asansol Engineering College

Information Technology

Asansol Engineering College
Asansol

Affiliated to

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY

June 2024

ASANSOL ENGINEERING COLLEGE

Kanyapur, Vivekananda Sarani, Asansol, Paschim Bardhaman, West Bengal - 713305
Phone: 225-3057, 225-2108 Telefax: (0341) 225-6334
E-mail: principal.aecwb@gmail.com Website: www.aecwb.edu.in

CERTIFICATE

Certified that this project report on **"Aqua-Air Insight: water air quality prediction"** is the bonafide work of **"Shreyansh Shekhar (10800220006), Rishika Singh (10800220008), Udayan Choudhary (10800220035), Sumit Kumar Mihirachariya (10800220042)** who carried out the project work under my supervision.

____

Mr. Sudip Kumar De
Assistant professor
Asansol Engineering College

____

Dr. Anup Kumar Mukhopadhyay
HOD, Information Technology
Asansol Engineering College

Department of Information Technology
Asansol Engineering College
Asansol

ACKNOWLEDGEMENT

It is our great privilege to express our profound and sincere gratitude to our Project Supervisor **Mr. Sudip Kumar De**, Assistant Professor of Information Technology department for providing us with very cooperative and precious guidance at every stage of the present project work being carried out under his supervision. His valuable advice and instructions in carrying out the present study have been a very rewarding and pleasurable experience that has greatly benefitted us throughout our work.

We would also like to pay our heartiest thanks and gratitude to **Dr. Anup Kumar Mukhopadhyay**, HOD, and all the faculty members of the Information Technology Department, Asansol Engineering College for various suggestions being provided in attaining success in our work.

We would like to express our earnest thanks to our colleagues along with all technical staff of the Information Technology Department, Asansol Engineering College for their valuable assistance being provided during our project work.

Finally, we would like to express our deep sense of gratitude to our parents for their constant motivation and support throughout our work.

……………………………………………
Shreyansh Shekhar
(10800220006)

……………………………………………
Rishika Singh
(10800220008)

……………………………………………
Udayan Choudhary
(10800220035)

……………………………………………
Sumit Kumar Mihirachariya
(10800220042)

Date: ----/----/------ **4th Year**
Place: Asansol **Department of Information Technology**

CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: PROJECT SYNOPSIS

The Project Synopsis of this prediction of Water and Air Quality using Machine Learning Technology provides us a brief overview of how water and air parameters can indicate whether water is safe for consumption and usage, and the air we breathe in is safe or not. This project deals with an innovative technology where the water and air parameters combine with machine learning backend (using libraries like NumPy, pandas, seaborn, sklearn). Aqua-Air Insight gives us accurate picture about water and air quality that how pure water and air is to use in a daily human life

Water quality is a critical aspect of public health and environmental sustainability. Contaminated water can lead to a host of health issues, including gastrointestinal illnesses, reproductive problems, and neurological disorders.

can lead to respiratory diseases, cardiovascular problems, and even premature death Air quality is also a crucial factor in public health and environmental sustainability. Poor air quality.

Aqua-Air insight uses advanced algorithm and large dataset to monitor and identify patterns to predict pollution level in both air and water which prevents ecological damage and ensure sustainable living.

Machine learning models can analyse vast amounts of data from multiple sources, such as sensors, satellite imagery, and historical records, to predict water quality and identify potential contamination issues. This technology provides a more proactive approach to water quality management, enabling early detection and intervention.

CHAPTER 2: INTRODUCTION

Aqua-Air is a comprehensive system that focuses on monitoring and predicting the quality of both water and air, crucial elements for human survival and well-being. Water serves as the primary source for delivering energy to cells and regulating bodily functions, with severe dehydration leading to cognitive impairments. The human body contains 70% water, which plays vital role in overall function.

The contamination of water makes it not suitable for any drinking purposes. Aqua-Air Insight makes precise location-based quality predictions for water and air, essential for urban and rural areas facing increased toxicity risks. Early detection through Aqua Air Predictor aids in informed decisions, potentially prompting human movement to safer areas.

Absolutely, Air is an essential for human life and plays a crucial role for human existence. It is a dynamic mixture of gases which makes life possible on earth. But now a days construction of chemical factories and automobile industry which many trees are cut down which makes the air full of toxins. The Air Quality Index (AQI) serves as a numerical indicator of air pollution levels, with higher numbers indicating more severe contamination and health risks.

The Aqua-Air Insight employs Python with libraries such as NumPy, pandas, scikit-learn, and seaborn, leveraging the Random Forest (RF) algorithm and Linear Regression for accurate and reliable quality predictions in water and air parameters.

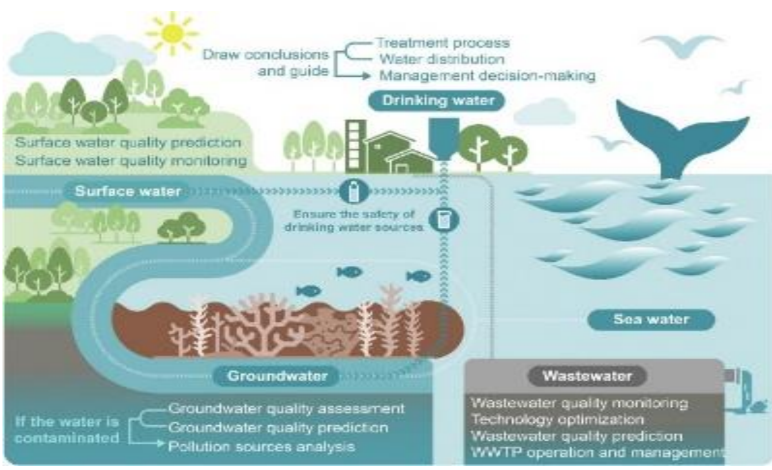

Figure 1: Demonstration of Water Cycle.

CHAPTER 3: PROJECT DETAILS

3.1 SYSTEM REQUIREMENTS

The system requirements for Aqua Air Insight software are similar to those of other advanced applications, with an emphasis on performance and compatibility with modern operating systems. Here are the general recommendations.

Hardware Requirements

Aqua-Air Insight operates efficiently on standard hardware configuration commonly found in contemporary computing environments. In order to ensure smooth development of our project below is the list of fundamental hardware required,

Table 1: Hardware Requirements

Graphics	NVIDIA GeForce GTX 1050ti or above
Processors	4.10 GHz or Faster Processor
RAM	8GB or more
Resolution	1920 X 1080 or Higher Resolution

Software Requirements

Aqua-Air Insight software's requirements give a range of technology to ensure seamless. To ensure smooth development of our project below is the list of fundamental software requirements,

Table 2: Software Requirements

Operating System	Linux(preferred), Windows 10 and above
Front End	HTML, CSS
Libraries	NumPy, Pandas, Matplotlib, Scikit-learn, Pickle, Seaborn and Plotply, Joblib.
Code Editor	Jupyter Notebook, VS Code and Google Collab
Dataset	• https://www.kaggle.com/datasets/adityakadiwal/water-potability • https://mega.nz/file/0n1j3Dja#O4AdTcZeMZgBag6OTRUaz9wPC9pp4xeeG8IpH-HRTo0
Web Browser	Google Chrome, Firefox, Brave, Edge
Models	Linear Regression, Random Forest Classifier

Network & Connectivity

Aqua-Air Insight web interface demands stable internet connection to predict the quality of air and water. While the system is accessible on different networks speed, reliable and reasonable connection is advised for prompt diagnosis of water and air.

3.2 PROPOSED METHODOLOGY

Proposed methodology outlines a comprehensive approach to build and deploy a water and air quality prediction system using machine learning, ensuring accuracy, real-time capabilities, and continuous improvement. Figure 2 illustrates the flow chart representation of water and air quality prediction.

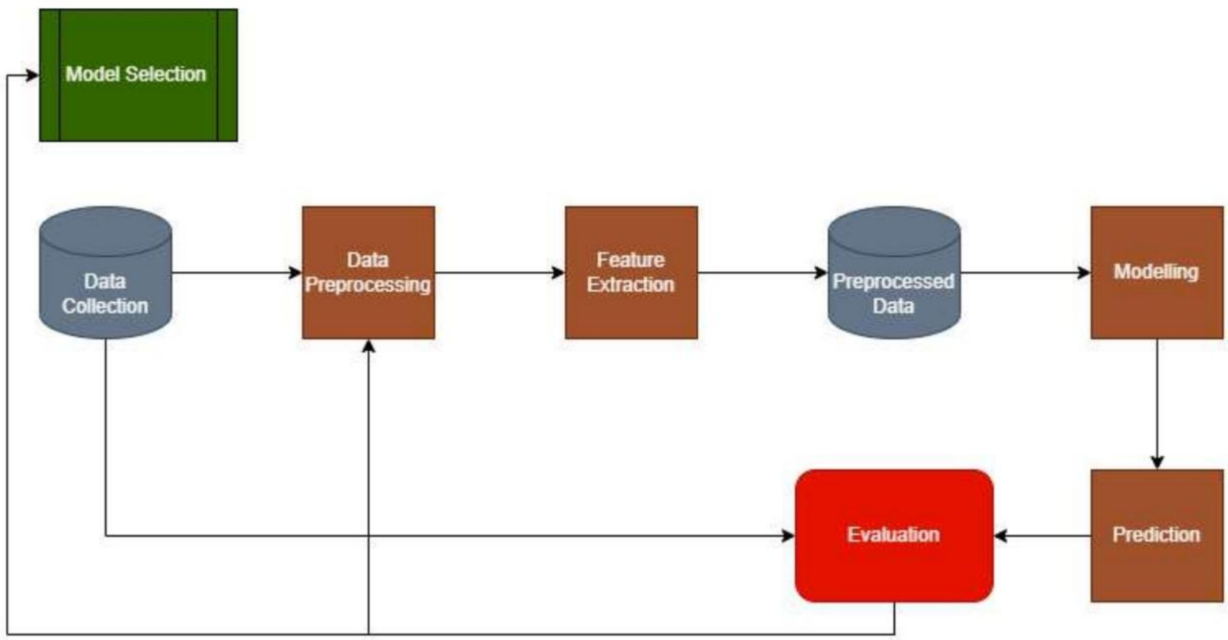

Figure 2: Workflow of Aqua Air Insight.

The steps of the project are mentioned below:

1. **Data Collection:** Gather water and air quality data from online resources (Kaggle), pH, turbidity, PM2.5, CO_2 and NO_2. for air also data is collected from online resources (Kaggle) keeping in mind the factors like SO_2, NO_2, RSPM, SPM.
   Water Dataset - https://www.kaggle.com/datasets/adityakadiwal/water-potability
   Air Dataset - https://mega.nz/file/0n1j3Dja#O4AdTcZeMZgBag6OTRUaz9wPC9pp4GH

2. **Data Pre-processing:** Clean the data by handling missing values, outliers, and noise, and normalize it for consistency.

3. **Exploratory Data Analysis (EDA)**: Perform descriptive statistics, correlation analysis, and visualization to understand data patterns.

4. **Model Selection**: Evaluate regression (e.g., Random Forest, Linear Regression) and classification models (e.g., SVM) based on MAE, RMSE, accuracy, and F1-score.

5. **Random Forest**: Random forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of their predictions for classification or the mean prediction for regression. It is used for water quality prediction in this project.

6. **Linear Regression**: Linear regression is a statistical method that models the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data. It is used for air quality prediction in this project.

7. **Model Training and Test:** Split data into training and testing sets, use k-fold cross-validation, and optimize hyper parameters. The model is split into 80 percent training and 20 percent test set.

8. **Real-time Prediction**: Implement models for real-time data prediction and integrate with a user-friendly dashboard.

9. **Performance Monitoring:** Continuously monitor model performance, which is for public use. Continuous monitoring and measuring procedure will be stablished to address issues and implementing details.

3.3 DEFINITIONS AND THEORIES

DEFINITIONS:
Aqua-Air Insight project is a ground breaking initiative designed to predict the toxicity levels of air and water and timely action taken against those. It provides a user-friendly solution accessible to both individuals and organisation for better habituation.

- **Objective and Scopes:**
  The primary objective of Aqua -air Insight project is to simplify the complex task of predicting the toxins present in our earth realm. The project focuses on creating an interactive web interface to predict toxins and pollutants present in the realm and give swift and accurate analysis.

- **Role of early detection:**
  The project is on the theory that early detection n of the air and water significantly impacts the habituation of the man kind by providing this platform which gives quick and accurate prediction, aqua-air Insight aim to facilitate early intervention, therefore improving analysis for individual or an organisation.

- **Medical Significance:**
  The project gives the snippet of crucial conditions of pollution around us i.e. in air and water and emphasises on early prediction of air and water thereby reducing the risks of airborne and waterborne disease.

- **Classification:**
  Classifications are machine learning techniques used to categorize data into predefined classes or labels. Common methods include decision trees, random forests, support vector machines (SVM), k-nearest neighbour (KNN), and neural networks.

- **Regression**:
  Regression is a statistical method used to model and analyse the relationships between a dependent variable and one or more independent variables. Common types include linear, polynomial, and logistic regression. It is widely used for predicting continuous outcomes, such as sales forecasting, price estimation, and risk assessment, by identifying trends and patterns within the data.

THEORIES:

Machine Learning and Data-Driven Approaches for water-Air quality prediction:

Supervised Learning: This involves training models on labelled datasets, where the input features (e.g., temperature, pH, turbidity, NO_2, CO_2) and the corresponding water quality outcomes (e.g., contaminant levels) are known. Common supervised learning algorithms used in water quality prediction include:

- **Linear Regression**:
  Predicts continuous air quality parameters based on linear relationships between input features and the target variable. Linear regression is a type of supervised machine learning algorithm that computes the linear relationship between the dependent variable and one or more independent features by fitting a linear equation to observed data.

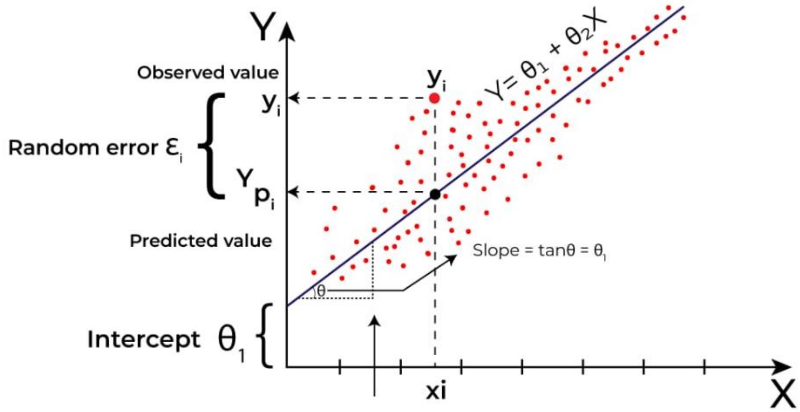

Figure 3: Graphical representation of Linear Regression Model used in Air quality prediction.

- **Random Forests**:
  Random Forest algorithm is a powerful tree learning technique in Machine Learning. It works by creating a number of Decision Trees during the training phase. Each tree is constructed using a random subset of the data set to measure a random subset of features in each partition. This randomness introduces variability among individual trees, reducing the risk of overfitting and improving overall prediction performance. The pictorial presentation of algorithm is explained in figure 3.

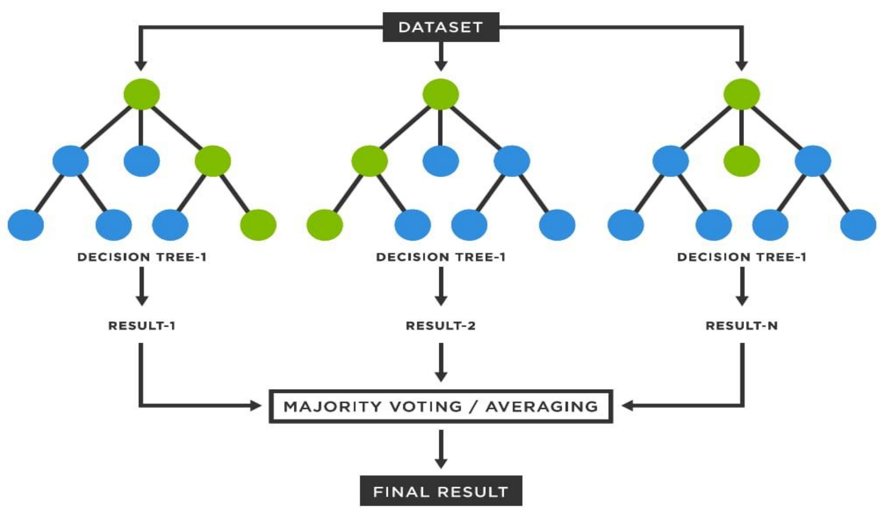

Figure 4: Graphical representation of Random Forest Model used in Air quality prediction

3.4 WORKING FLOW OF THE PROJECT

The working flow of water and air quality prediction using machine learning involves several key steps, from data collection to model deployment and continuous improvement. Here's a detailed outline of the process:

Flowchart

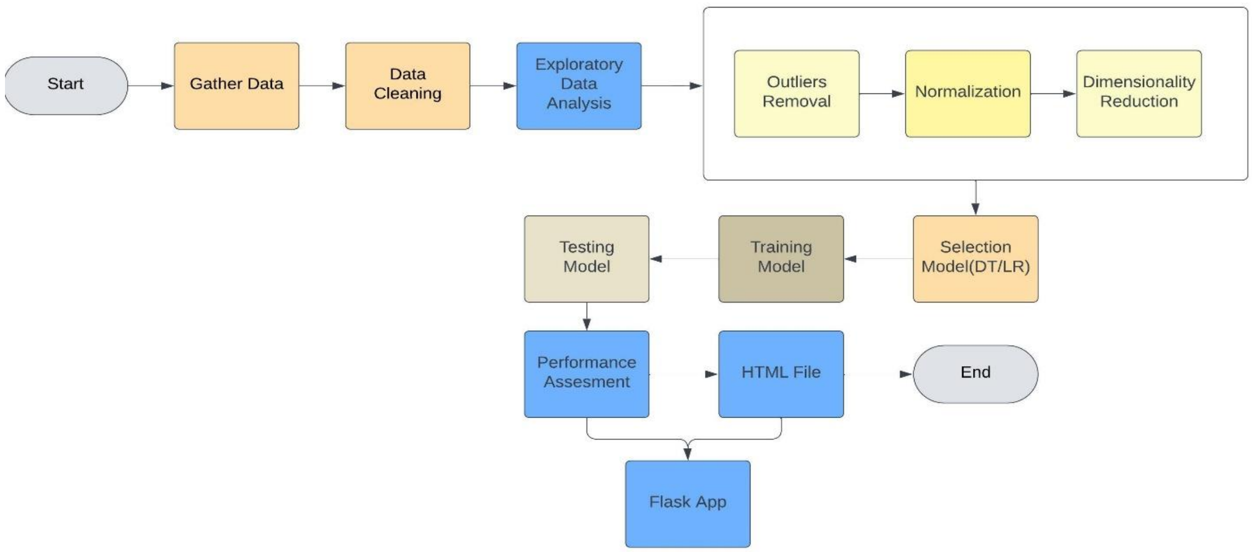

Figure 5: Steps followed in Aqua Air Insight Project

- **Data Collection:** Data collection involves gathering information from various sources to be used for analysis, research, or decision-making.
  Dimension of air quality dataset is: (84239, 4) means 84239 rows and 4 columns.

	stn_code	sampling_date	state	location	agency	type	so2	no2	rspm	spm
0	150.0	February - M021990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	4.8	17.4	NaN	NaN
1	151.0	February - M021990	Andhra Pradesh	Hyderabad	NaN	Industrial Area	3.1	7.0	NaN	NaN
2	152.0	February - M021990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	6.2	28.5	NaN	NaN
3	150.0	March - M031990	Andhra Pradesh	Hyderabad	NaN	Residential, Rural and other Areas	6.3	14.7	NaN	NaN
4	151.0	March - M031990	Andhra Pradesh	Hyderabad	NaN	Industrial Area	4.7	7.5	NaN	NaN

Figure 6: Dataset for Air Quality prediction

Dimension of water quality dataset is: (3276, 10) means 3276 rows and 10 columns.

ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability
NaN	204.890455	20791.318981	7.300212	368.516441	564.308654	10.379783	86.990970	2.963135	0
3.716080	129.422921	18630.057858	6.635246	NaN	592.885359	15.180013	56.329076	4.500656	0
8.099124	224.236259	19909.541732	9.275884	NaN	418.606213	16.868637	66.420093	3.055934	0
8.316766	214.373394	22018.417441	8.059332	356.886136	363.266516	18.436524	100.341674	4.628771	0
9.092223	181.101509	17978.986339	6.546600	310.135738	398.410813	11.558279	31.997993	4.075075	0

Figure 7: Dataset for Water Quality Prediction

- **Data Pre-processing**: Data preprocessing involves cleaning, transforming, and organising raw data into suitable format for analysing and modelling.

- **Exploratory Data Analysis (EDA):** Visualization: Use graphs and plots to understand data distributions, trends, and relationships between variables. Statistical Analysis: Perform summary statistics to gain insights into the data and identify key patterns.

- **Outliers Detection and Data Cleaning**: Remove of missing values Impute missing values. In Figure 8 we have found out that independent variables of solid has more outliers and substituted the mean value. In Figure 9 we have found out that independent variables of soi and noi has more outliers and substituted the mean value.

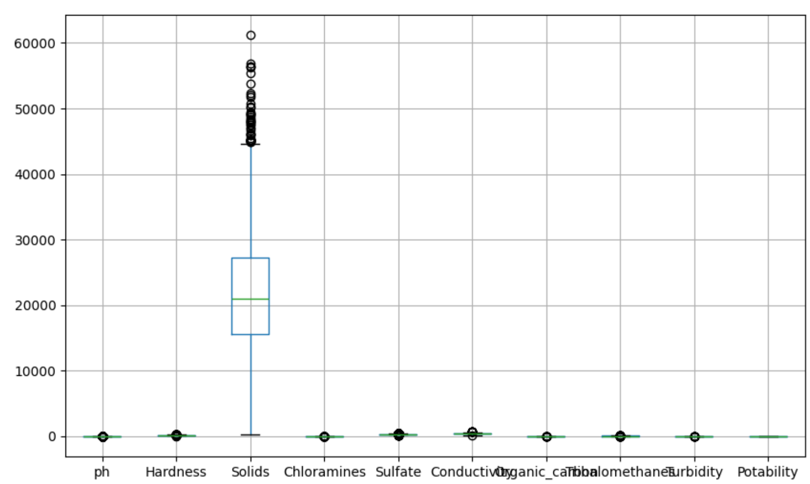

Figure 8: Outliers of independent variables (water Quality prediction)

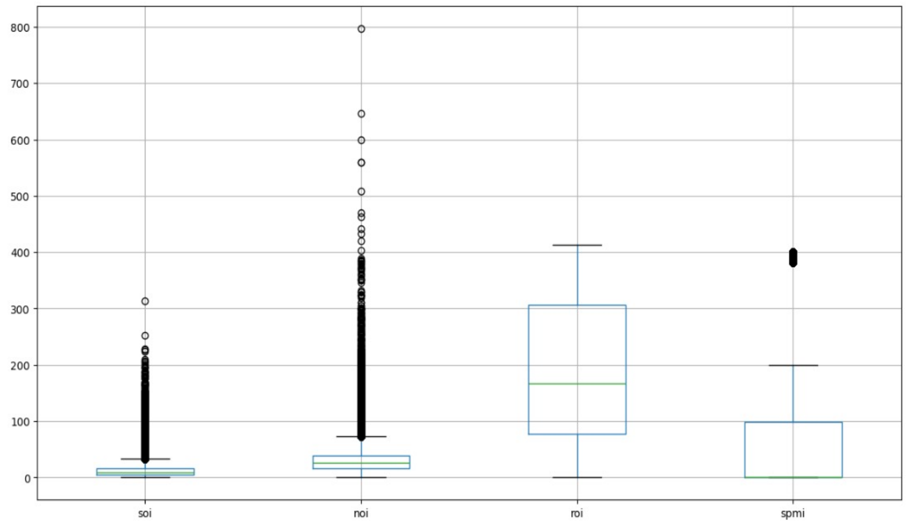

Figure 9: Outliers of independent variables (Air Quality prediction).

- **Normalization**: To assess our water dataset's distribution and detect biases, we will plot the distribution frequency of all independent variables using histograms or density plots. These visualizations as shown in Figure 9 will reveal patterns, skewness, and anomalies, helping us evaluate how well-distributed the data is. This analysis is essential for understanding the dataset's structure and ensuring the accuracy of subsequent analyses.

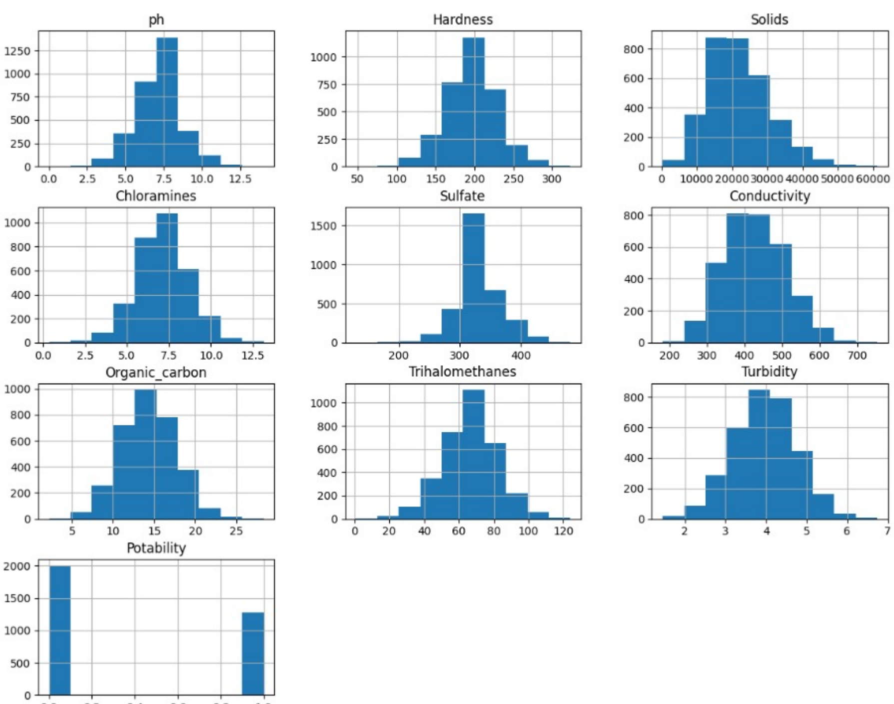

Figure 10: The distribution flow of data of independent variables

- **Dimensionality Reduction**: After verifying the correlations using a heatmap, In Figure 11 Figure 12 we have observed that none of the independent variables are significantly correlated with each other. As a result, we did not perform any dimensionality reduction since there is no redundancy in the features. Each independent variable provides unique information, making all of them valuable for the model.

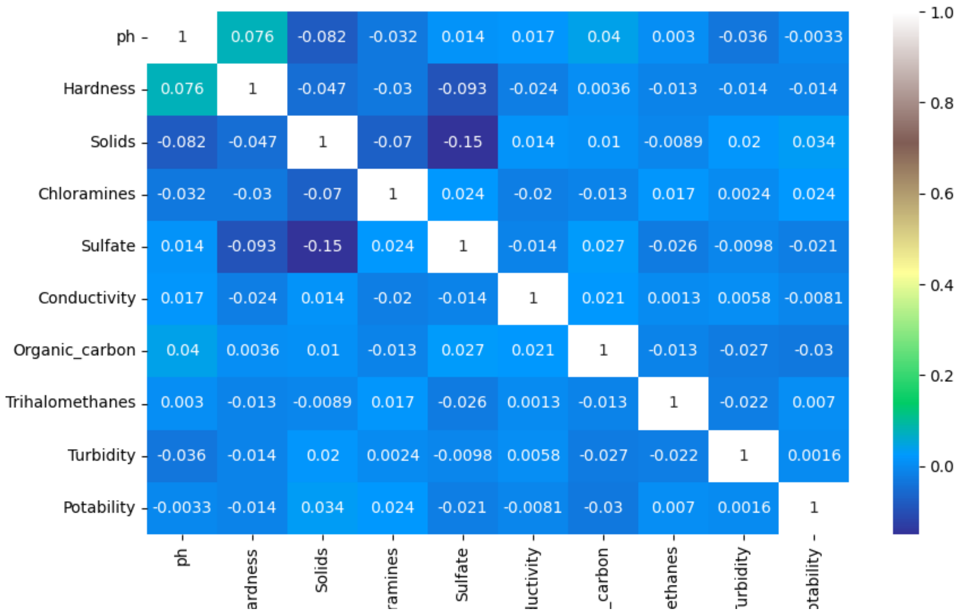

Figure 11: Heat map of independent variables of water quality dataset

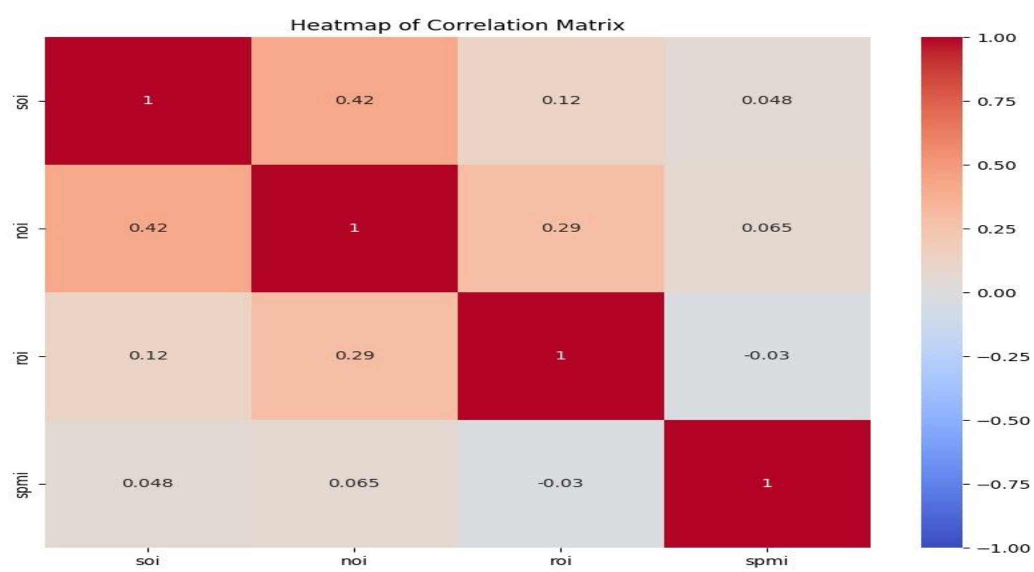

Figure 12: Heat map of independent variables of air quality dataset

- **Data Splitting:** We used the “train_test_split” function from the sklearn. Model selection module to divide our dataset into two parts: a training set and a testing set. We split our dataset into training and testing sets with an 80-20 proportion using the “train_test_split” function. This approach allocates 80% of the data for training the model and 20% for testing its performance.

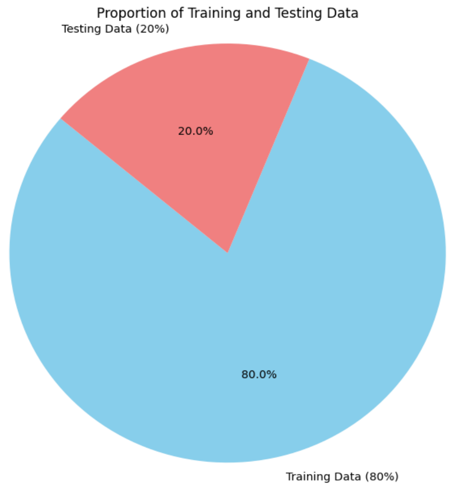

Figure 13: Data proportion of Training and Testing Aqua-Air prediction

- **Model Selection:** In our approach to predict water quality, we have used the Random Forest model, which is a robust classification algorithm. This model is particularly effective for categorizing water quality based on various parameters and features that influence its condition.

  We have utilized the linear regression model for predicting air quality due to its inherent capability in handling regression tasks. This model allows us to estimate air quality metrics by establishing a linear relationship between input variables and the target output.

  The model selection plays a vital role in predictions as we have implemented the models to yield the best accuracy for seamless and precise prediction.

- **Model Training and Testing**: We trained our random forest model using 80% of the dataset and then tested it with the remaining 20%, which was randomly split. For our water quality prediction model, the model achieved an accuracy of 70.43% as shown in Figure 14.
  For our air quality prediction, we used a linear regression model. We trained the model with 80% of the dataset and tested it with the remaining 20%. The model achieved an accuracy score of 90.85% as shown in Figure 15.

```python
# Random forest....
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier()
rf.fit(x_train,y_train)
random_forest = rf.predict(x_test)
random_forest_score = accuracy_score(random_forest,y_test)*100
print(f'Random Forest Accuracy: {random_forest_score:.2f}%')
```

```
Random Forest Accuracy: 70.43%
```

Figure 14: Training and Testing of model for Water quality prediction using Random Forest

```python
linear = LinearRegression()
linear.fit(x_train, y_train)
linear_pred = linear.predict(x_test)
linear_score = linear.score(x_test, y_test)
print('Linear Regression R2 Score:', linear_score*100)
```

```
Linear Regression R2 Score: 90.85330255322454
```

Figure 15: Training and Testing of model for Air quality prediction using Linear Regression

- **Code Implementation**:

Code Implementation of Water prediction: Having serialized our machine learning model using pickle, we proceeded to integrate it with Flask, a versatile web framework. This setup enabled us to develop a web application that interacts with users via HTML forms. Upon receiving input data, Flask routes it to the model for prediction, returning the results seamlessly to the user interface. This integration streamlined the deployment process while providing an intuitive interface for leveraging the model's predictive capabilities. In Figure 16 the code for water quality prediction model is shown In Figure 17 the code base of the flask file which is use to integrate with HTML document is shown. The structure of HTML document is shown in Figure 18.

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import joblib

data = pd.read_csv('D:\water_potability.csv')

data = data.fillna(data.mean())

x = data.drop('Potability',axis = 1)
y = data['Potability']

from sklearn.metrics import accuracy_score,confusion_matrix
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size = 0.2,shuffle=True,random_state=None)

# Random forest....
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier()
rf.fit(x_train,y_train)
random_forest = rf.predict(x_test)
random_forest_score = accuracy_score(random_forest,y_test)*100
print(f'Random Forest Accuracy: {random_forest_score:.2f}%')

joblib.dump(rf, 'model.pkl')
```

Figure 16: Code of Random Forest Model

```python
from flask import Flask, request, jsonify, render_template
import joblib
import numpy as np
model = joblib.load('model.pkl')
app = Flask(__name__)
@app.route('/')
def home():
    return render_template('index.html')


@app.route('/predict', methods=['POST'])
def predict():

    features = [float(request.form['feature1']),
                float(request.form['feature2']),
                float(request.form['feature3']),
                float(request.form['feature4']),
                float(request.form['feature5']),
                float(request.form['feature6']),
                float(request.form['feature7']),
                float(request.form['feature8']),
                float(request.form['feature9'])]


    features_array = np.array(features).reshape(1, -1)
    prediction = model.predict(features_array)
    if prediction[0] == 1:
        result = 'Potable'
    else:
        result = 'Non-Potable'

    return render_template('index.html', prediction_text=f'The water quality is predicted as: {result}')

if __name__ == '__main__':
    app.run(debug=True)
```

Figure 17: Code of Flask Document Water Quality Prediction.

```html
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <link rel="stylesheet" href="../static/air_style.css">
    <title>Document</title>
</head>
<body>
    <div class="container">
        <h1 class="heading">Air Quality Prediction</h1>
        <form action="{{url_for('predict')}}" method="post">
            <p>SO2</p>
            <input type="number" class="input" id="so2" step="0.01" name="so2">

            <p>NO2</p>
            <input type="number" class="input" id="no2" step="0.01"  name="no2">

            <p>RSPM</p>
            <input type="number" class="input" id="rspm" step="0.01" name="rspm">

            <p>SPM</p>
            <input type="number" class="input" id="spm" step="0.01" name="spm">

            <button class="btn">Predict</button>
        </form>
        {% if prediction is not none %}
            <div class="result">
                <h2>Predicted Value: {{ prediction }}</h2>
                <h2>Air Quality: {{ prediction_range }}</h2>

            </div>
        {% endif %}
    </div>
</body>
</html>
```

Figure 18: HTML document of Water Quality Prediction.

Code Implementation of Air prediction: We saved our model with joblib, loaded it in a Flask app, and integrated it with HTML for user input, facilitating predictions via web forms for a seamless user experience. The linear regression model for air quality prediction is shown in Figure 19. Figure 20 shows the flask file which is used to integrate the machine learning model With HTML document. The HTML document of air quality prediction is shown in Figure 21.

```python
linear = LinearRegression()
linear.fit(x_train, y_train)
linear_pred = linear.predict(x_test)
linear_score = linear.score(x_test, y_test)
print('Linear Regression R2 Score:', linear_score)


import joblib
joblib.dump(linear,'air_quality.joblib')
import os
print(os.getcwd())
```

Figure 19: Code of Linear Regression Model.

```python
from flask import Flask, render_template, request, jsonify
import joblib
import numpy as np

model = joblib.load('air_quality.joblib')

app = Flask(__name__, template_folder='templates')

@app.route('/')
def home():
    return render_template('idx.html')

@app.route('/predict', methods=['POST'])
def predict():
    so2 = float(request.form['so2'])
    no2 = float(request.form['no2'])
    rspm = float(request.form['rspm'])
    spm = float(request.form['spm'])
    input_data = np.array([[so2, no2, rspm, spm]])
    prediction = model.predict(input_data)
    prediction_range = get_air_quality(prediction)
    return render_template('idx.html', prediction=prediction[0], prediction_range=prediction_range)

def get_air_quality(aqi_value):

    if aqi_value <= 50:
        return 'Good'
    elif aqi_value <= 100:
        return 'Moderate'
    elif aqi_value <= 150:
        return 'Unhealthy for Sensitive Groups'
    elif aqi_value <= 200:
        return 'Unhealthy'
    elif aqi_value <= 300:
        return 'Very Unhealthy'
    else:
        return 'Hazardous'

if __name__ == '__main__':
    app.run(debug=True)
```

Figure 20: Code of Flask Document Air Quality Prediction.

```html
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <link rel="stylesheet" href="../static/style.css">
    <title>Water Quality Detection</title>
</head>
<body>
    <div class="container">
        <h1>Water Quality Detection</h1>
        <form id="predict-form" action="/predict" method="post">

            <input type="number" id="feature1" name="feature1" placeholder="Ph(0-14)" class="input" step="0.001">
            <input type="number" id="feature2" name="feature2" class="input" placeholder="Hardness(50-300)" step="0.001">
            <input type="number" id="feature3" name="feature3" class="input" placeholder="Solids(350-50K)" step="0.001">
            <input type="number" id="feature4" name="feature4" class="input" placeholder="Chloramines(0-13)" step="0.001">
            <input type="number" id="feature5" name="feature5" class="input" placeholder="Sulphate(150-450)" step="0.001">
            <input type="number" id="feature6" name="feature6" class="input" placeholder="Conductivity(200-750)" step="0.001">
            <input type="number" id="feature7" name="feature7" class="input" placeholder="Organic_Carbon(2-28)" step="0.001">
            <input type="number" id="feature8" name="feature8" class="input" placeholder="Trihalomethanes(0-120)" step="0.001">
            <input type="number" id="feature9" name="feature9" class="input" placeholder="Turbidity(1-6)" step="0.001">

            <button type="submit" id="submit">Predict</button>
        </form>
        <h2 id="result">

            {% if prediction_text %}
                {{ prediction_text }}
            {% endif %}
        </h2>
    </div>

</body>
</html>
```

Figure 21: HTML document of Air Quality Prediction.

3.5 OUTCOMES OF THE PROJECT

The outcomes of water and air quality prediction using machine learning and can significantly enhance the management and protection of water and air resources. Key outcomes include**:**

- **Accurate Forecasting:** Machine learning models can provide precise predictions of water and air quality parameters (e.g., pH, turbidity, SO2, NO2) for various time frames, ranging from short-term forecasts (days to weeks).

- **Early Warning Systems:** Accurate predictions can provide early warnings of poor air or water quality, enabling public health officials to issue advisories and take preventative measures to protect vulnerable populations, such as the elderly, children, and those with pre-existing health conditions.

- **Disease Prevention:** By predicting potential contamination events in water and air resources, authorities can prevent waterborne and airborne diseases by taking timely action to ensure water and air safety.

- **Real-Time Monitoring:** Immediate detection of changes in air and water quality, enabling swift interventions to safeguard from negative impacts.

- **Recovery Planning:** Post-disaster recovery efforts can be better planned using predictions to restore air and water quality, ensuring safe and healthy living conditions more quickly.

- **Urban Planning**: Air quality predictions influence urban development plans by identifying areas with high pollution levels. This informs decisions on infrastructure placement, traffic management, and green space allocation to mitigate exposure to pollutant

- **Agricultural Planning:** Farmers can use predictions of water quality to make informed decisions about irrigation, reducing crop damage from contaminated water and improving overall agricultural productivity.

3.6 APPLICATION INTERFACE

The Aqua Air Insight application interface developed using HTML, CSS and Joblib, offers a user-friendly experience for accessing functionalities such as accurate water and air quality prediction. Interface of Water Quality Prediction is designed to input parameter of water and it analyse whether the water is potable or not as shown in Figure 21. Interface of Air Quality Prediction is designed to input of air quality parameters to analyse the air quality index as well as the of air as shown in Figure 22.

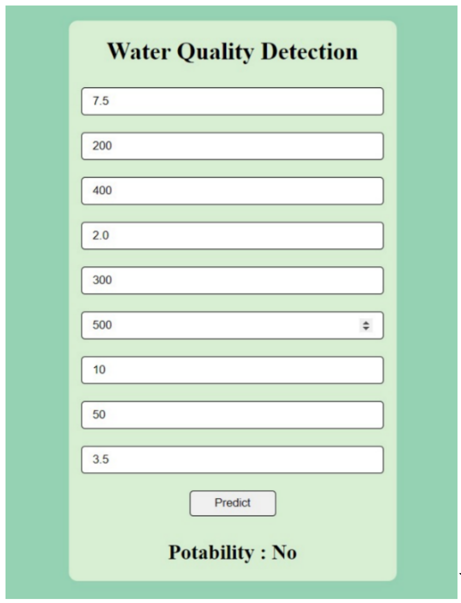

Figure 22: Interface of Water Quality Prediction.

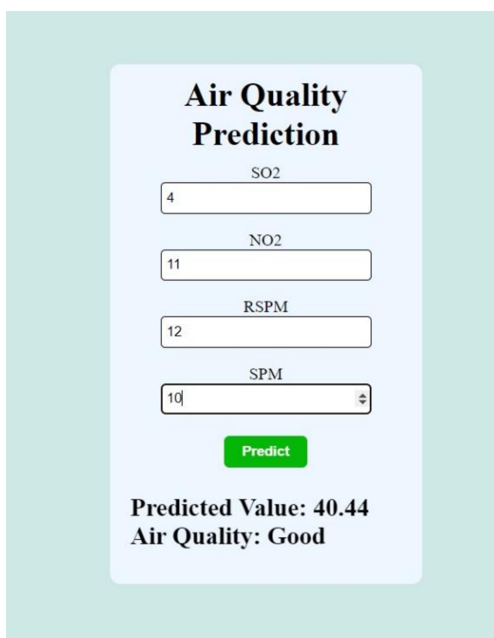

Figure 23: Interface of Air Quality Prediction

CHAPTER 4: CONCLUSION & FUTURE SCOPE

CONCLUSION

As we all know the importance of water for the human body. Knowing the quality of water is essential because consuming unsafe water can lead to water-borne diseases such as Cholera, Typhoid, Giardia, E. Coli, and Hepatitis A. Therefore, assessing water quality is crucial. However, traditional methods of testing water quality require laboratory analysis, which can be costly and time-consuming. This paper proposes an alternative approach using artificial intelligence, specifically the Random Forest algorithm, to predict water quality. The outcomes of this project extend beyond statistical performance metrics. The insights derived from the feature selection analysis contribute to our understanding of the factors influencing Air and Water Quality Prediction. these predictions are valuable for the mankind as it ensures safety from waterborne and airborne diseases.

In Conclusion, the project on Aqua Air Insight using Machine Learning has provided valuable insights into the development of robust predictive models for early and risk detection. The exploration of different machine learning techniques, including Random Forest Classifier, Decision Tree Classifier, Logistic Regression allowed for a comprehensive understanding of their strengths and limitations in the context of Air and Water Quality prediction. Linear Regression and Random Forest used in this project gives comprehensive understanding pf the strengths and limitations in context of Aqua Air Insight.

FUTURE SCOPE

The future scope of water and air quality prediction will be combined for better accuracy with hyper parameter tuning and complex and large dataset will be used to enhance model accuracy through diverse data sources, hyperparameter optimization, and combining algorithms, will yield more sophisticated models. Integrating these models with real-time monitoring systems will enable prompt detection and response, promoting public health and environmental sustainability.

CHAPTER 5: REFERENCE

1. www.google.com :
2. https://www.kaggle.com/datasets/adityakadiwal/water-potability
3. https://mega.nz/file/0n1j3Dja#O4AdTcZeMZgBag6OTRUaz9wPC9pp4xeeG8IpH-HRTo0
4. www.youtube.com
5. Hayder, G.; Kurniawan, I.; Mustafa, H.M. Implementation of machine learning methods for monitoring and predicting water quality parameters. *Biointerface Res. Appl. Chem.* **2020**, *11*, 9285–9295.