



edunet  
foundation

# WATER QUALITY PREDICTION

By: AKSHAR NATH GORAIN

Registration ID:

STU6829b3f24d2691747563506

AICTE Internship ID:

INTERNSHIP\_1746416864681834e0e35d8

## Learning Objectives

- To develop an AI-powered system for predicting water pollutants using machine learning.
- To utilize historical station-based monitoring data for environmental analysis and forecasting.
- To implement a multi-output machine learning model for pollutant prediction.
- To deploy the prediction model as a user-friendly web application using Gradio.
- Understand how to use Python libraries like Pandas, NumPy, Scikit-learn for analysis and model building.
- Understand how to use Matplotlib, Seaborn for visualization.
- Understand how to use GitHub for project version control and uploading.
- Understand the web deployment process via Gradio in Google form Colab.



## Tools and Technology used

Tools used :

Coding Language : Python

Environment used : Collab

• Libraries used :

Numpy :For handling arrays & mathematical operations.

Pandas :For data analysis, cleaning and filtering.

Matplotlib :For visualisation.

Seaborn : For making attractive graphs.

Sklearn : Used in ML(Regression , classification).

• Models used :

Linear regression : Used in SML for predicting continuous numeric values.

Random forest regressor :SML algorithm used for regression task, good for overfitting problem.

• Evolution metrices :

RMSE(Root Mean Squared Error), MAE(Mean Absolute Error), R2 Score.

## Methodology

- **Data Collection:** Used PB\_All\_2000\_2021.csv dataset, covering the time period 2000–2021.
- **Data Cleaning:** Cleaned null values and filtered key stations from the dataset.
- **EDA & Visualization:** Conducted exploratory data analysis and visualized pollutant trends.
- **Model Building:** Implemented a MultiOutputRegressor with RandomForestRegressor.
  - **Input Features:** Year, Station ID
  - **Output Pollutants:** O<sub>2</sub>, NO<sub>3</sub>, NO<sub>2</sub>, SO<sub>4</sub>, PO<sub>4</sub>, Cl<sup>-</sup>
- **Model Persistence:** Saved the trained model as pollution\_model.pkl and column features as model\_columns.pkl using Joblib.
- **Web Deployment:** Converted the application to Gradio for web deployment in Google Colab.

## **Problem Statement:**

- Lack of a proactive system to forecast water pollutant levels, hindering effective environmental management.**
- Difficulty in rapidly assessing potential water quality issues based on historical trends and station-specific data.**
- Need for an accessible and user-friendly tool that can provide quick insights into future water quality to aid decision-makers.**

## **Solution:**

- Developed an AI-powered machine learning model capable of predicting six key water pollutants ( $O_2$ ,  $NO_3$ ,  $NO_2$ ,  $SO_4$ ,  $PO_4$ ,  $Cl^-$ ).**
- The model leverages historical year and station\_id data for predictions.**
- Created an interactive web application using Gradio, allowing users to input a year and station ID to get immediate pollutant level predictions.**
- Provided a deployment solution via Google Colab, making the prediction tool easily accessible for environmental analysis and forecasting.**



## Screenshot of Output:

id5ed52bd6da81.gradio.live

### Water Pollutants Predictor

Predict levels of O<sub>2</sub>, NO<sub>3</sub>, NO<sub>2</sub>, SO<sub>4</sub>, PO<sub>4</sub>, and Cl<sup>-</sup> based on year and station ID.

Year  
2022

Station ID  
e.g., 1

Predicted Pollutant Levels

Flag

Clear Submit

## Gradio Interface

```

1 # Install required packages
2 !pip install gradio pandas scikit-learn --quiet
3
4 # Define Gradio app
5 import pandas as pd
6 import joblib
7 import gradio as gr
8
9 # Load model and columns
10 model = joblib.load("pollution_model.pkl")
11 model_cols = joblib.load("model_columns.pkl")
12 pollutants = ['O2', 'NO3', 'NO2', 'SO4', 'PO4', 'CL']
13
14 def predict_pollutants(year, station_id):
15     try:
16         input_df = pd.DataFrame({'year': [year], 'id': [station_id]})
17         input_encoded = pd.get_dummies(input_df, columns=['id'])
18
19         # Align columns with training features
20         for col in model_cols:
21             if col not in input_encoded.columns:
22                 input_encoded[col] = 0
23         input_encoded = input_encoded[model_cols]
24
25         prediction = model.predict(input_encoded)[0]
26         return (p, round(val, 2) for p, val in zip(pollutants, prediction))
27     except Exception as e:
28         return ("Error": str(e))
29
30 # Interface
31 gr.Interface(
32     fn=predict_pollutants,
33     inputs=[
34         gr.Number(label="Year", value=2022, minimum=2000, maximum=2100),
35         gr.Textbox(label="Station ID", placeholder="e.g., 1")
36     ],
37     outputs=gr.JSON(label="Predicted Pollutant Levels"),
38     title="Water Pollutants Predictor",
39     description="Predict levels of O2, NO3, NO2, SO4, PO4, and Cl- based on year and station ID."
40 ).launch(share=True)

```

Colab notebook detected. To show errors in colab, \* Running on public URL: <https://eb3f1d5ed52bd6da81.gradio.live>

This share link expires in 1 week. For free pe

Water Quality Prediction

GitHub: <https://github.com/workforakng/WaterQualityPrediction>

To install basic/necessary libraries

```

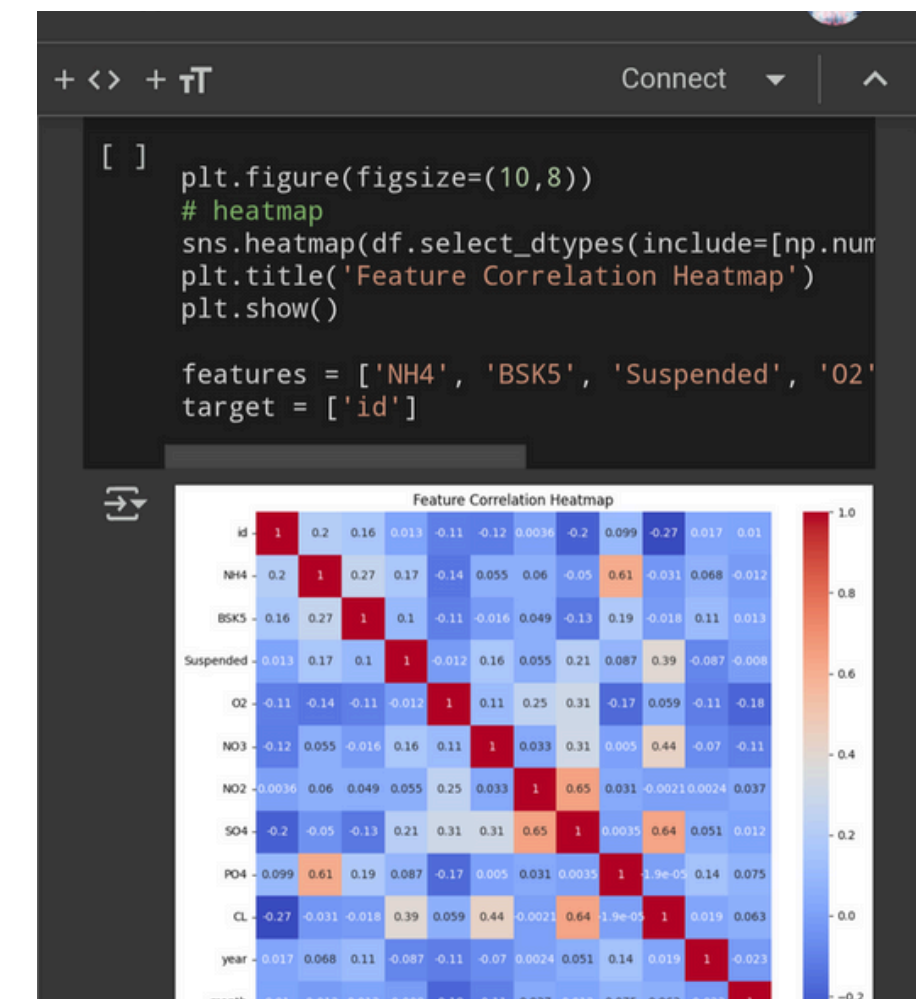
pip install pandas numpy matplotlib seaborn scikit-learn

```

```

[1] 1 # Import necessary libraries
2 import pandas as pd # data manipulation

```



## Conclusion:

- Successfully developed and deployed an AI-powered water quality prediction model using historical monitoring data.
- The Multi-output Random Forest Regressor effectively predicts six major water pollutants ( $O_2$ ,  $NO_3$ ,  $NO_2$ ,  $SO_4$ ,  $PO_4$ ,  $Cl^-$ ).
- The Gradio web application provides an accessible and intuitive tool for environmental analysis and forecasting.
- This project demonstrates the significant potential of machine learning to support proactive environmental management and decision-making for a healthier future.

[Click to open GitHub Link](https://github.com/workforakng/WaterQualityPrediction.git) : <https://github.com/workforakng/WaterQualityPrediction.git>



## GITHUB:

**GitHub Repository:** <https://github.com/workforakng/WaterQualityPrediction.git>

- Find the full project code, dataset, and trained model files on GitHub.

[Click to open](#)

[Repository](#)

## Run the App in Google Colab:

- Click below to launch the prediction app:
-  [Run on Colab](#)

## Run Locally using Gradio:

- `pip install -r requirements.txt`
- `python app.py`