DEPARTMENT OF INFORMATION TECHNOLOGY

B.TECH III YEAR I SEM

A.Y:2025-26

INNOVATIVE PROJECT REPORT

TITLE OF PROJECT: Water Quality Prediction and Alert System

BATCH NO.	ROLL NO.	NAME	PHONE NO.	EMAIL
11	23891A1202	ADDULA SONY	9440938617	sonyreddyaddula@gmail.coM
	23891A1229	KANDI RAKESH REDDY	9502469558	rakeshreddykandi79@gmail.com
	23891A1252	RAMASWAMY RESHWANTH REDDY	9154640740	reshwanth558@gmail.com
	23891A1253	RAVALKOL MEGHANA	8019917686	ravalkolmeghana@gmail.com

Name of the Guide : Mr. SK. Khaleelullah

Problem Statement :

Access to safe drinking water is a major global challenge, especially in rural and developing regions. Traditional water testing methods require laboratory setups, skilled personnel, and are time-consuming. This creates a need for a fast, cost-effective, and automated solution to evaluate water quality. This system aims to predict whether water is Safe (Potable) or Unsafe (Non-Potable) by analyzing its physicochemical characteristics.

Introduction :

Water is essential for life, and its quality directly impacts human health. Contaminated water leads to severe diseases such as cholera, typhoid, and diarrhea. Due to limited access to lab testing facilities, many communities still consume unsafe water unknowingly.

With advancements in machine learning, it is now possible to analyze water chemically and predict its safety using computational methods.
This project uses a Random Forest Classifier trained on water quality datasets to predict water potability. A Streamlit web interface enables users to input water parameters and instantly check whether the water is drinkable.

The solution is scalable, affordable, and highly efficient, making it suitable for both urban and rural environments.

Objectives :

1. To study the chemical factors affecting water potability.
2. To preprocess and clean water quality data for machine learning.
3. To develop a classification model using the Random Forest algorithm.
4. To evaluate the performance of the model using accuracy metrics.
5. To build an interactive web-based interface using Streamlit.
6. To provide instant water potability predictions based on user inputs.

Key Concepts :

1. Machine Learning :

- A technique that trains models using historical data to make predictions.
- Used here for classifying water as potable or non-potable.

2. Random Forest Classifier :

- An ensemble learning method combining multiple decision trees.
- Provides high accuracy, stability, and reduces overfitting.

3. Data Preprocessing :

- Handling missing values
- Normalization
- Train-test split
- Feature labeling

4. Feature Importance :

- Helps identify which water parameters contribute most to potability prediction.

5. Streamlit Framework :

- Python-based web app tool
- Creates interactive UI for prediction
- Requires no backend or HTML/CSS

Hardware and Software Requirements :

Hardware Requirements :

- A standard PC or laptop
- Minimum 4GB RAM (8GB recommended)
- Processor: Intel i3 or above
- Stable internet connection (for deployment)

Software Requirements :

- **Programming Language :** Python 3.10+
- **Libraries :**
  - ✓ Pandas
  - ✓ Matplotlib
  - ✓ Seaborn
  - ✓ scikit-learn
  - ✓ streamlit
- **Tools / Platforms :**
  - ✓ Jupyter Notebook / VS Code / PyCharm
  - ✓ Streamlit Cloud (for deployment)
  - ✓ GitHub (for version control)
- **Methodology :**

Step 1: Data Collection

- Dataset containing water parameters such as pH, Hardness, Turbidity, Conductivity, etc.

Step 2: Data Preprocessing

- Remove missing values
- Normalize the dataset
- Convert data types
- Apply train-test split (80% training, 20% testing)

Step 3: Model Selection

Random Forest Classifier is chosen due to:

- High accuracy
- Low risk of overfitting
- Ability to handle nonlinear data

Step 4: Model Training

- Train using training dataset
- Test using unseen test dataset
- Evaluate using:
    - ✓ Accuracy Score
    - ✓ Confusion Matrix
    - ✓ Classification Report

Step 5: Web App Development (Streamlit)

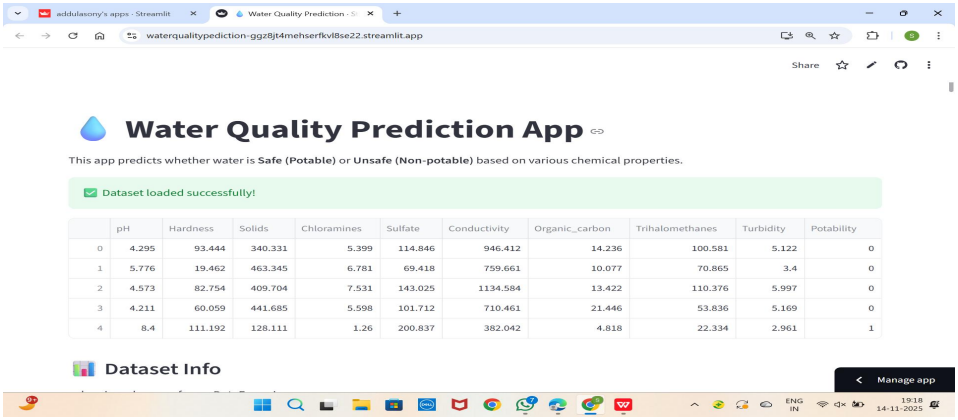
- Create input form for water parameters
- Load trained model
- Display prediction result

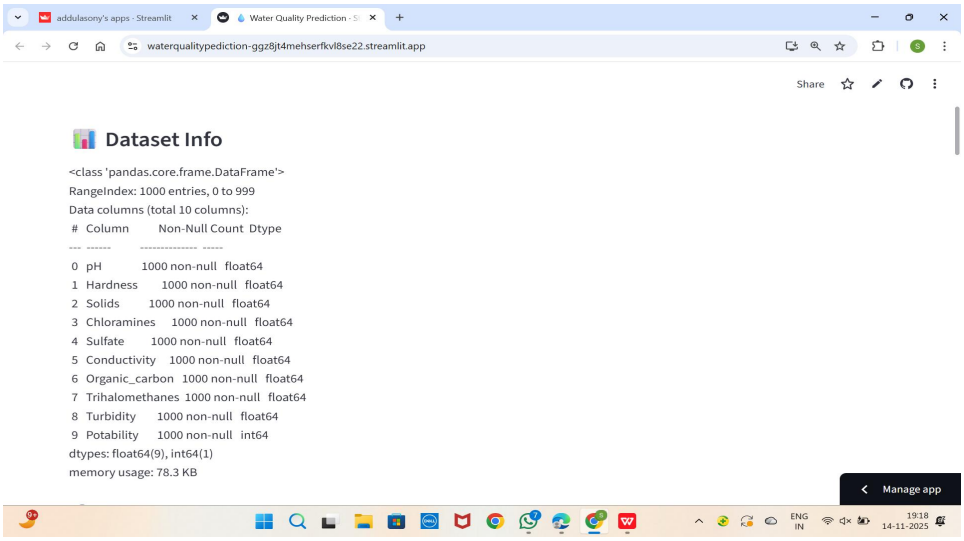
- Visualize graphs (heatmap, feature importance)

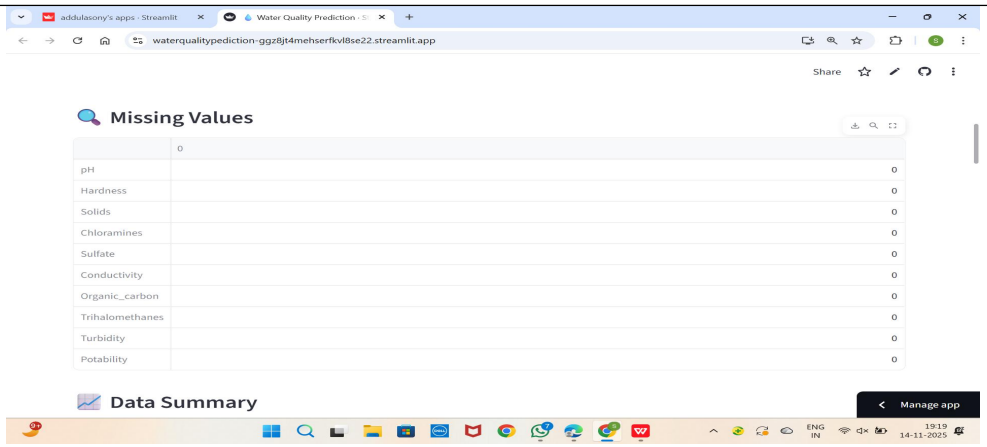
Step 6: Deployment

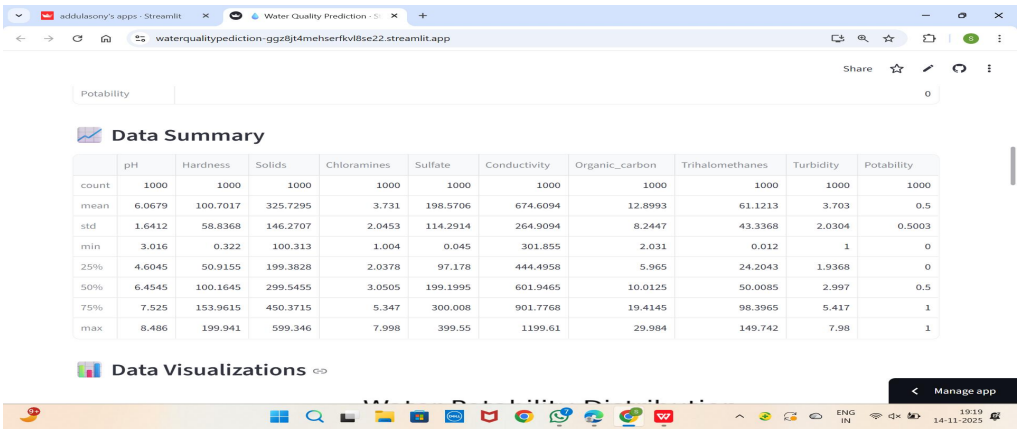
- Upload project to GitHub
- Deploy using Streamlit Cloud
- App becomes accessible on any device via URL

Output :

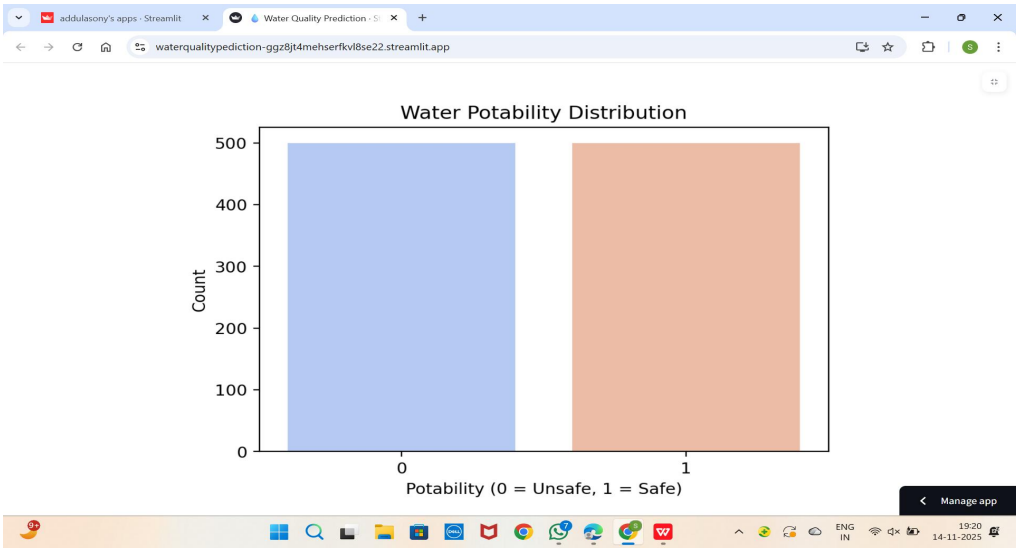




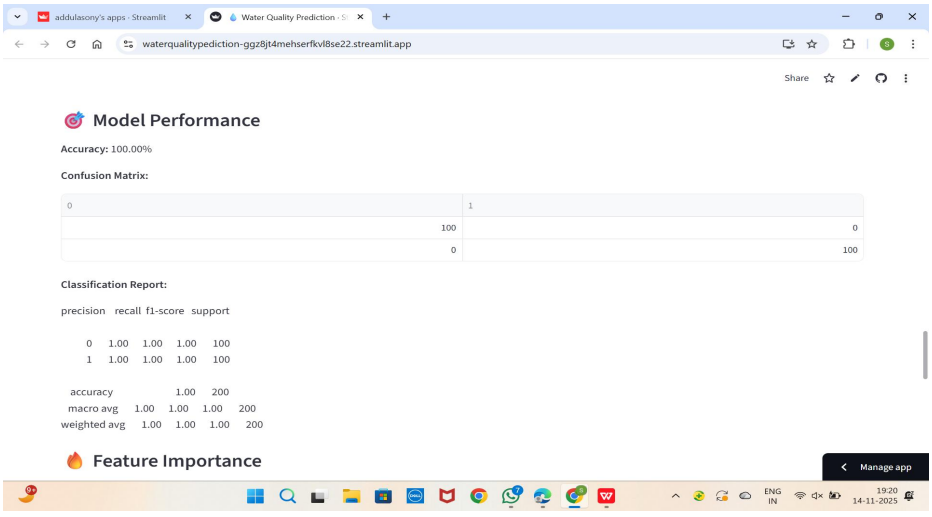


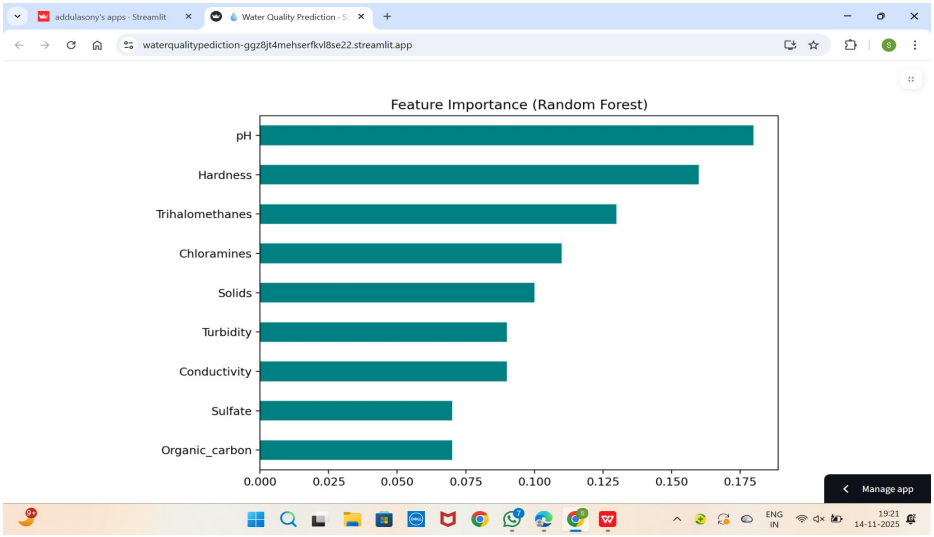


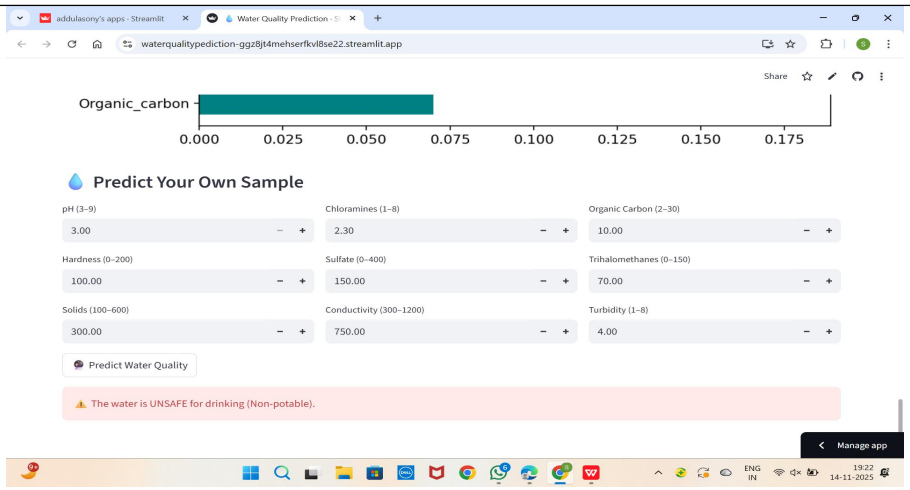
	pH	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability
count	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
mean	6.0679	100.7017	325.7295	3.731	198.5706	674.6094	12.8993	61.1213	3.703	0.5
std	1.6412	58.8368	146.2707	2.0453	114.2914	264.9094	8.2447	43.3368	2.0304	0.5003
min	3.016	0.322	100.313	1.004	0.045	301.855	2.031	0.012	1	0
25%	4.6045	50.9155	199.3828	2.0378	97.178	444.4958	5.965	24.2043	1.9368	0
50%	6.4545	100.1645	299.5455	3.0505	199.1995	601.9465	10.0125	50.0085	2.997	0.5
75%	7.525	153.9615	450.3715	5.347	300.008	901.7768	19.4145	98.3965	5.417	1
max	8.486	199.941	599.346	7.998	399.55	1199.61	29.984	149.742	7.98	1

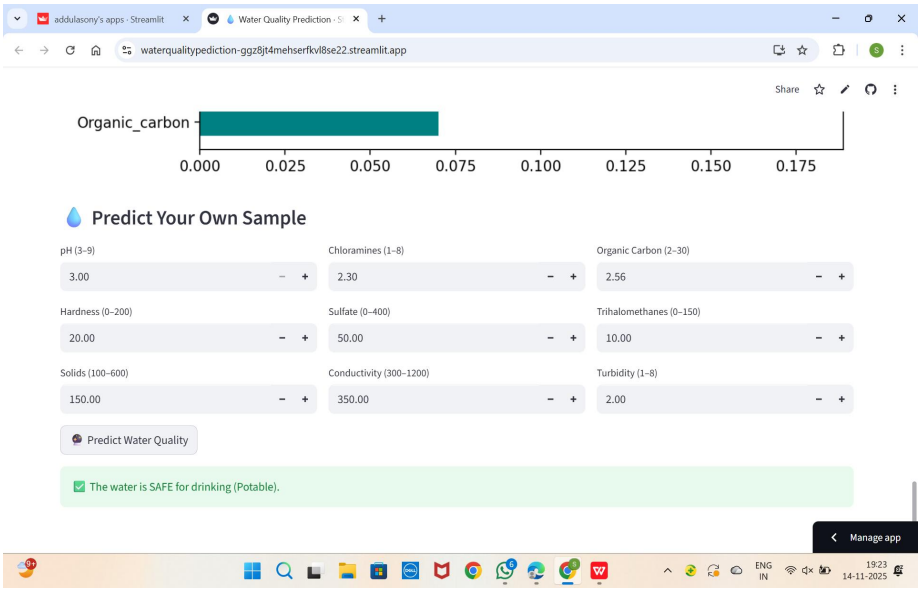


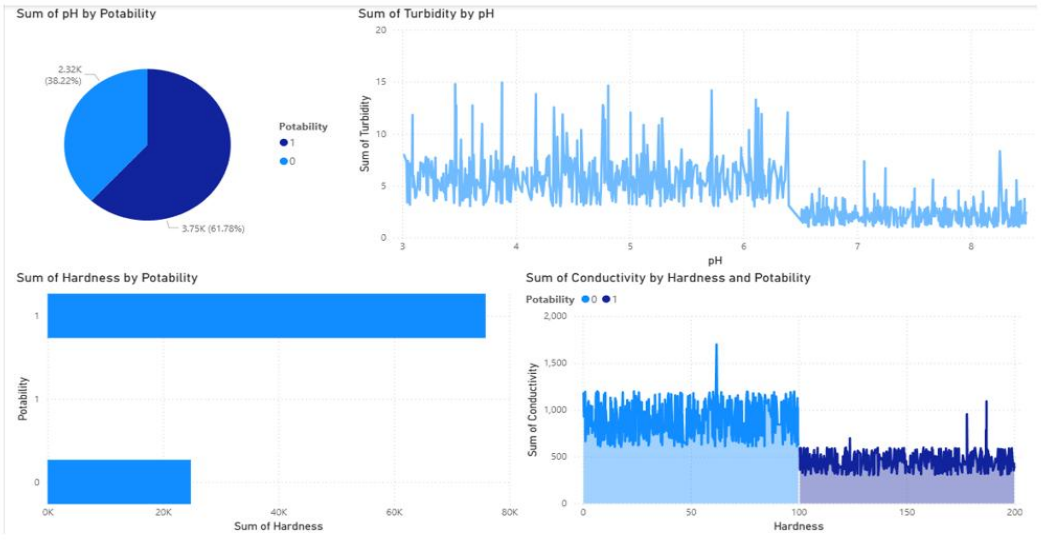

addulasony's apps - Streamlit
Water Quality Prediction
waterqualitypediction-ggz8jt4mehserfkvl8se22.streamlit.app
Feature Correlation Heatmap
pH 1.00 0.76 -0.74 -0.74 0.76 -0.76 -0.72 -0.75 -0.75 0.87
Hardness 0.76 1.00 -0.74 -0.74 0.76 -0.73 -0.72 -0.73 -0.73 0.87
Solids -0.74 -0.74 1.00 0.72 -0.75 0.73 0.71 0.73 0.74 -0.86
Chloramines -0.74 -0.74 0.72 1.00 -0.75 0.73 0.71 0.71 0.73 -0.85
Sulfate 0.76 0.76 -0.75 -0.75 1.00 -0.75 -0.73 -0.74 -0.73 0.87
Conductivity -0.76 -0.73 0.73 0.73 -0.75 1.00 0.72 0.72 0.72 -0.85
Organic_carbon -0.72 -0.72 0.71 0.71 -0.73 0.72 1.00 0.73 0.72 -0.84
Trihalomethanes -0.75 -0.73 0.73 0.71 -0.74 0.72 0.73 1.00 0.71 -0.85
Turbidity -0.75 -0.73 0.74 0.73 -0.73 0.72 0.72 0.71 1.00 -0.85
Potability 0.87 0.87 -0.86 -0.85 0.87 -0.85 -0.84 -0.85 -0.85 1.00
pH Hardness Solids Chloramines Sulfate Conductivity Organic_carbon Trihalomethanes Turbidity Potability
1.00 0.75 0.50 0.25 0.00 −0.25 −0.50 −0.75
Manage app

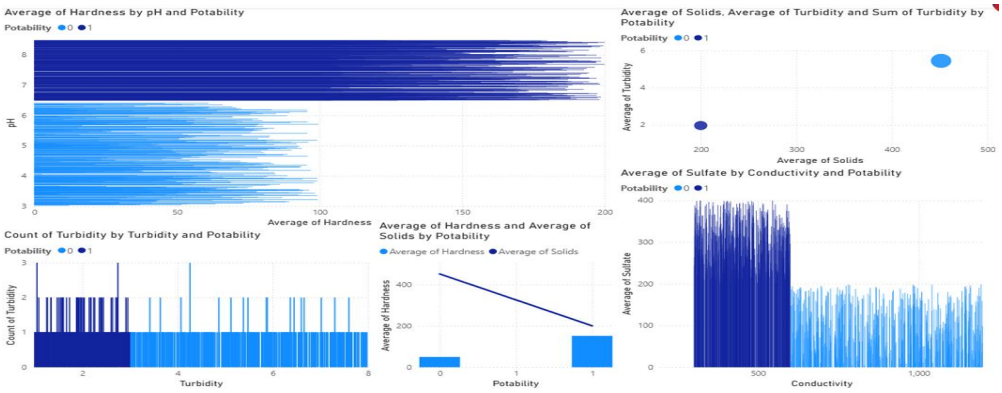

addulasony's apps - Streamlit
Water Quality Prediction
waterqualitypediction-ggz8jt4mehserfkvl8se22.streamlit.app
Share
Model Performance
Accuracy: 100.00%
Confusion Matrix:
0 1
100 0
0 100
Classification Report:
precision recall f1-score support
0 1.00 1.00 1.00 100
1 1.00 1.00 1.00 100
accuracy 1.00 200
macro avg 1.00 1.00 1.00 200
weighted avg 1.00 1.00 1.00 200
Feature Importance
Manage app


addulasony's apps - Streamlit
Water Quality Prediction
waterqualitypediction-ggz8jt4mehserfkvl8se22.streamlit.app
Feature Importance (Random Forest)
pH
Hardness
Trihalomethanes
Chloramines
Solids
Turbidity
Conductivity
Sulfate
Organic_carbon
0.000 0.025 0.050 0.075 0.100 0.125 0.150 0.175
Manage app


addulasony's apps · Streamlit
Water Quality Prediction
waterqualitypediction-ggz8jt4mehserfkvl8se22.streamlit.app
Share
Organic_carbon
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
Predict Your Own Sample
pH (3–9)
3.00
Chloramines (1–8)
2.30
Organic Carbon (2–30)
10.00
Hardness (0–200)
100.00
Sulfate (0–400)
150.00
Trihalomethanes (0–150)
70.00
Solids (100–600)
300.00
Conductivity (300–1200)
750.00
Turbidity (1–8)
4.00
Predict Water Quality
The water is UNSAFE for drinking (Non-potable).
Manage app
ENG
IN
19:22
14-11-2025


addulasony's apps · Streamlit
Water Quality Prediction
waterqualitypediction-ggz8jt4mehserfkvl8se22.streamlit.app
Share
Organic_carbon
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
Predict Your Own Sample
pH (3–9)
3.00
Chloramines (1–8)
2.30
Organic Carbon (2–30)
2.56
Hardness (0–200)
20.00
Sulfate (0–400)
50.00
Trihalomethanes (0–150)
10.00
Solids (100–600)
150.00
Conductivity (300–1200)
350.00
Turbidity (1–8)
2.00
Predict Water Quality
The water is SAFE for drinking (Potable).
Manage app
ENG
IN
19:23
14-11-2025


Sum of pH by Potability
2.32K (38.22%)
3.75K (61.78%)
Potability
1
0
Sum of Turbidity by pH
Sum of Turbidity
pH
Sum of Hardness by Potability
Potability
Sum of Hardness
0K
20K
40K
60K
80K
Sum of Conductivity by Hardness and Potability
Potability 0 1
Sum of Conductivity
Hardness
0
50
100
150
200
500
1,000
1,500
2,000



Conclusion

The Water Quality Prediction System successfully uses machine learning techniques to determine whether water is safe for drinking. The Random Forest model delivers high accuracy and effectively analyzes complex chemical parameters. The Streamlit interface makes the system user-friendly and accessible, enabling instant prediction without laboratory testing.

This project demonstrates the importance of technology in solving real-world environmental problems and provides a practical solution for water safety monitoring. It can be extended to IoT-based real-time monitoring systems in the future.

Project Guide	**Project Coordinator**	**HOD**