Project Report Format

1. INTRODUCTION:

1.1 Project overview:

Forecasting water quality is an essential part of environmental management. With the least amount of input data possible, artificial intelligence (AI) and machine learning (ML) approaches have been used to forecast water quality characteristics such the Water Quality Index (WQI) 1. To choose crucial factors including BOD5, NH4+, PO43-, turbidity, TSS, coliform, and DO 1, the Bayes method (BMA) has been employed.

Machine learning algorithms have been employed in a number of research to forecast the quality of river water. For example, a study that aimed to lower the cost of surface water quality monitoring 2 employed machine learning and deep learning algorithms to calculate the WQI with limited input data. A different study examined the effectiveness of artificial intelligence methods such as support vector machines (AVM), group method of data handling (GMDH), and artificial neural networks (ANN).

In conclusion, water quality metrics have been predicted with minimum input data using AI and ML algorithms. Machine learning algorithms have been utilized in several research to predict the quality of river water. To forecast water quality metrics, these research have employed a variety of techniques, including ANN, GMDH, SVM, and deep learning algorithms.

1.2Purpose:

The goal of river water quality forecasting with AI and ML is to anticipate metrics like the Water Quality Index (WQI) with the least amount of input data possible 1. To choose crucial factors including BOD5, NH4+, PO43-, turbidity, TSS, coliform, and DO 1, the Bayes method (BMA) has been employed.
River water quality has been modelled and water quality metrics have been predicted with minimum input data using machine learning techniques. These methods can be used to estimate pollution levels, identify the primary sources of contamination, and conduct a thorough investigation into the quality of the river water
A review paper published recently emphasizes the cutting edge use of machine learning methods for river water quality predictions. It outlines the several essential methods, benefits, drawbacks, and applications.

2. LITERATURE SURVEY :

2.1 Existing problem:

current issue The management of the environment must include water quality. Hazardous wastes and contaminants have a negative impact on river water quality. Concerns are raised by the fact that water pollution is getting worse and worse, degrading the quality of the water and making it unfit for human usage. 1.

2.2 References:

https://link.springer.com/article/10.1007/s40808-020-01041-z

https://ieeexplore.ieee.org/document/9528216/

https://www.sciencedirect.com/science/article/abs/pii/S002216942030130X

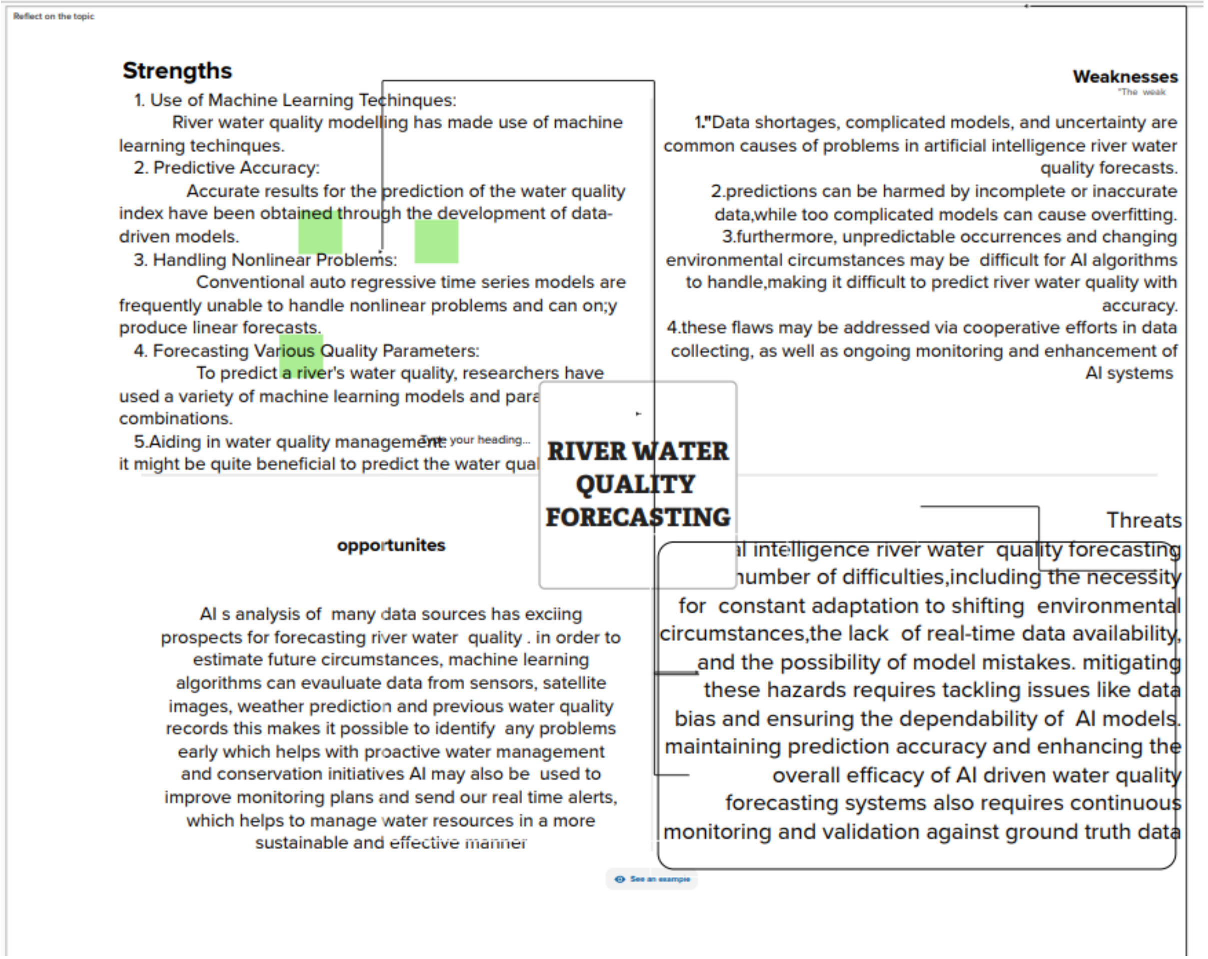
2.3 Problem Statement Definition:

Problem Synopsis An explanation of the AI/ML model used for forecasting river water quality The goal of river water quality forecasting with AI and ML is to anticipate water quality metrics like WQI with the least amount of input data possible. 1. River water quality has been modeled and water quality metrics have been predicted with minimum input data using machine learning techniques. These methods can assist with carrying out.

Maps of Empathy An aid for visualizing the ideas, emotions, and behaviors of clients or users is empathy mapping. It can be utilized to pinpoint possible areas for development and aid in providing a better knowledge of the client experience. 1.

3.IDEATION & PROPOSE SOLUTION:

3.1Empathy Map Canvas:



3.2 Ideation and brain storming:

Ideation and brainstorming for ai,ml model-based river water quality forecasting The process of generating ideas and solutions through various sessions including Brainstorming, Brainwriting, Prototyping, Sketching, Worst Possible Idea, and many more ideation techniques is called Ideation. Additionally, the Design Thinking process's third step is ideation. 2. Teams should gather various viewpoints and come up with ideas for solutions to design problems during the ideation stage. 3.

4. REQUIREMENT ANALYSIS:

4.1Functional Requirements:

Functional prerequisite The system services or functions are described by the functional requirements. One of the functional needs for river water quality forecasting with AI and ML could be the capacity to anticipate metrics like the Water Quality Index (WQI) with little to no input data 1.

4.2 Non- Functional Requirements:

Non-Functional prerequisites for utilizing the AI,ML model to forecast river water quality In contrast to functional requirements, which describe a system's business features and capabilities, non-functional requirements are the system quality attributes for a system. Non-functional requirements for AI and ML-based river water quality forecasts could include the following:

Mobility

Safety

Reliability

Dependability

The ability to scale

Achievement

Returnability

Adaptability

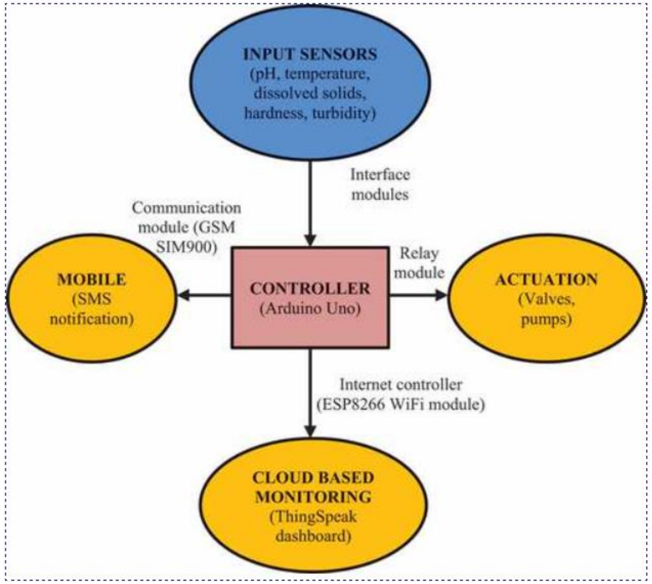
5.PROJECT DESIGN:

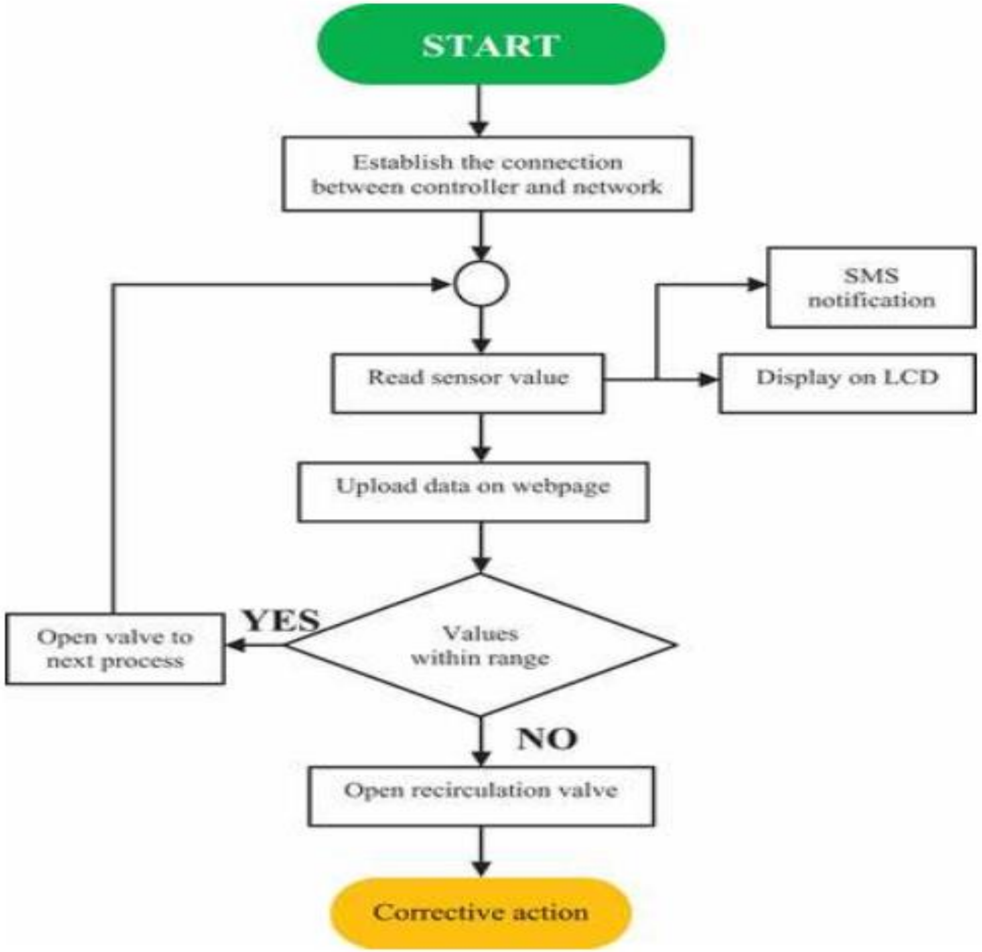
5.1 Data Flow Diagrams and User Stories:

Data Flow Diagrams:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

Example: DFD Level 0 (Industry Standard)


INPUT SENSORS
(pH, temperature,
dissolved solids,
hardness, turbidity)
Interface
modules
Communication
module (GSM
SIM900)
Relay
module
MOBILE
(SMS
notification)
CONTROLLER
(Arduino Uno)
ACTUATION
(Valves,
pumps)
Internet controller
(ESP8266 WiFi module)
CLOUD BASED
MONITORING
(ThingSpeak
dashboard)


START
Establish the connection between controller and network
SMS notification
Read sensor value
Display on LCD
Upload data on webpage
YES
Open valve to next process
Values within range
NO
Open recirculation valve
Corrective action

USER STORIES:

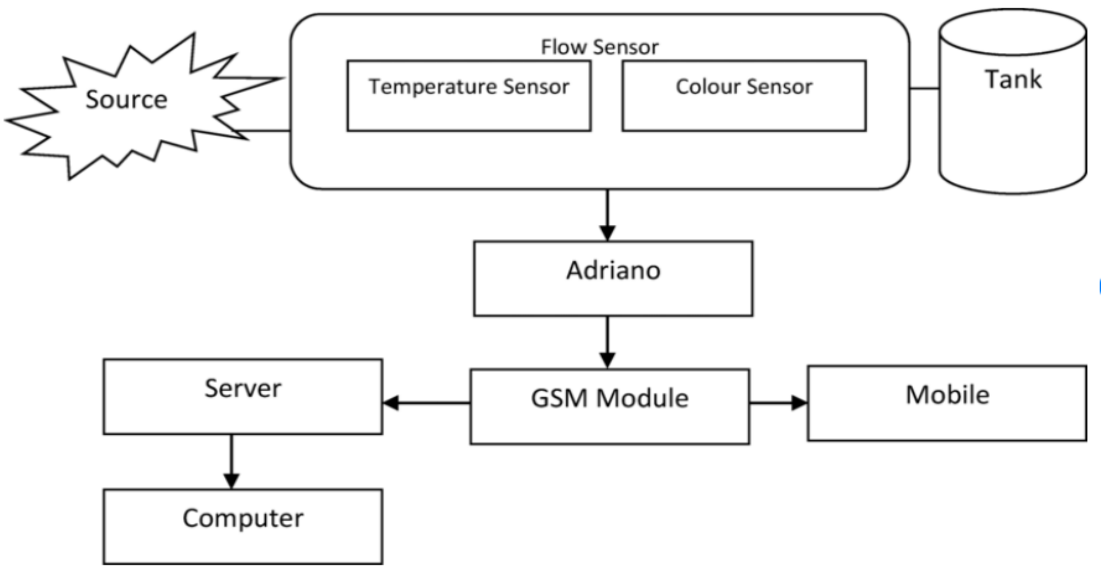
Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
River Water Quality Forecasting Using AI	Story 1	As a user, I want to collect historical river water quality data and relevant environmental factors so that the AI model can learn from this data.	The system successfully collects historical river water quality data and relevant environmental factors. The data is stored in an accessible and secure manner.	High	Sprint-1
River water quality forecasting using AI	Story 2	As a user, I want to clean and preprocess the collected data to make it suitable for training the AI model.	The system successfully cleans and preprocesses the collected data. The preprocessing includes handling missing values, outliers, and any inconsistencies in the data.	Low	Sprint-1
River Water Quality Forecasting Using AI	Story 3	As a user, I want to develop an AI model using suitable machine learning algorithms to predict river water quality.	The system successfully develops an AI model using suitable machine learning algorithms. The model is capable of predicting river water quality based on the input data.	High	Sprint-2
River Water Quality Forecasting Using AI	Story 4	As a user, I want to train the AI model using the preprocessed data to learn patterns and relationships between the data and river water quality.	The system successfully trains the AI model using the preprocessed data. The model is able to learn patterns and relationships between the data and river water quality.	High	Sprint-1
River Water Quality Forecasting Using AI	Story 5	As a user, I want to evaluate the performance of the trained model using appropriate metrics to ensure its accuracy and reliability.	The system successfully evaluates the performance of the trained model using appropriate metrics. The model meets or exceeds the required performance standards.	High	Sprint-1
River Water Quality Forecasting Using AI	Story 6	As a user, I want to optimize the model by tuning its parameters to improve its predictive performance.	The system successfully optimizes the model by tuning its parameters. The optimization improves the model's predictive performance without overfitting the data.	Low	Sprint-1
River Water Quality Forecasting Using AI	Story 7	As a user, I want to deploy the optimized model in a test environment to validate its performance in a real-world scenario.	The system successfully deploys the optimized model in a test environment. The model performs well in the test environment and its predictions closely match the actual river water quality.	High	Sprint-2

5.2 Solution Architecture:

Solution architecture is a complex process – with many sub-processes – that bridgesthe gap between business problems and technology solutions. Its goals are to:

- Find the best tech solution to solve existing business problems.
- Describe the structure, characteristics, behavior, and other aspects of the software to project stakeholders.
- Define features, development phases, and solution requirements.
- Provide specifications according to which the solution is defined, managed,and delivered.

Solution Architecture Diagram:

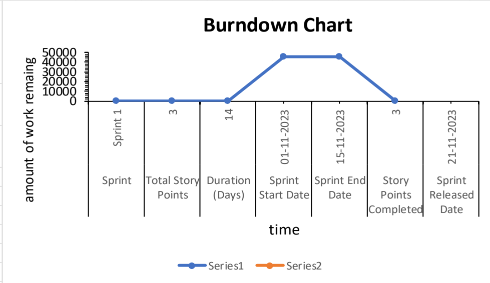


Architecture of the proposed water quality forecasting system

6.PROJECT PLANNING AND SCHEDULING:

6.1 Technical Architecture:

Sprint	Total Story Points	Duration (Days)	Sprint Start Date	Sprint End Date	Story Points Completed	Sprint Released Date
Sprint 1	3	14	01-11-2023	15-11-2023	3	21-11-2023
Sprint 2	5	6	15-11-2023	21-11-2023	5	



6.2 Sprint Planning and estimation:

Sprint	Functional Requirement (Epic)	User Story Number	User Story/Task	Story Points	Priority	Team Members
Sprint1	River Water Quality Forecasting using AI	Story 1	As a user, I want to collect historical river water quality data and relevant environmental factors so that the AI model can learn from this data.	3	High	Dhanya Deepika
Sprint2	River Water Quality Forecasting using AI	Story 2	As a user, I want to clean and preprocess the collected data to make it suitable for training the AI model.	5	Low	Nikhitha
Sprint1	River Water Quality Forecasting using AI	Story 3	As a user, I want to develop an AI model using suitable machine learning algorithms to predict river water quality.	8	High	Bhuvaneswari

Sprint	Functional Requirement (Epic)	User Story Number	User Story/Task	Story Points	Priority	Team Members
Sprint1	River Water Quality Forecasting using AI	Story 4	As a user, I want to train the AI model using the preprocessed data to learn patterns and relationships between the data and river water quality.	8	High	Dhanya Deepika
Sprint1	River Water Quality Forecasting using AI	Story 5	As a user, I want to evaluate the performance of the trained model using appropriate metrics to ensure its accuracy and reliability.	5	High	Bhuvaneswari
Sprint2	River Water Quality Forecasting using AI	Story 6	As a user, I want to optimize the model by tuning its parameters to improve its predictive performance.	8	Low	Nikhitha
Sprint1	River Water Quality Forecasting using AI	Story 7	As a user, I want to deploy the optimized model in a test environment to validate its performance in a real-world scenario.	13	High	Dhanya Deepika

6.3 Sprint Delivery Schedule:

Sprint	Total Story Points	Duration (Days)	Sprint Start Date	Sprint End Date	Story Points Completed	Sprint Released Date
Sprint 1	3	14	2023-11-01	2023-11-15	3	2023-11-21
Sprint 2	5	6	2023-11-15	2023-11-21	5	

7. CODING AND SOLUTIONING:

7.1 Feature 1:

```
import numpy as np
import pandas as pd
mport os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
main_df = pd.read_csv("/kaggle/input/water-potability/water_potability.
csv")
df = main_df.copy()
df.head()
```

7.2 Feature 2:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
import plotly.express as px
import warnings
warnings.filterwarnings('ignore')
print(df.shape)
```

```
print(df.shape)
(3276, 10)
print(df.columns)
Index(['ph', 'Hardness', 'Solids', 'Chloramines', 'Sulfate', 'Conductiv
ity',
       'Organic_carbon', 'Trihalomethanes', 'Turbidity', 'Potability'],
      dtype='object')
df.describe()
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3276 entries, 0 to 3275
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   ph               2785 non-null   float64
 1   Hardness         3276 non-null   float64
 2   Solids           3276 non-null   float64
 3   Chloramines      3276 non-null   float64
 4   Sulfate          2495 non-null   float64
 5   Conductivity     3276 non-null   float64
 6   Organic_carbon   3276 non-null   float64
 7   Trihalomethanes  3114 non-null   float64
 8   Turbidity        3276 non-null   float64
 9   Potability       3276 non-null   int64
dtypes: float64(9), int64(1)
memory usage: 256.1 KB
```

```
print(df.nunique())
```

```
ph                 2785
Hardness           3276
Solids             3276
Chloramines        3276
Sulfate            2495
Conductivity       3276
Organic_carbon     3276
```

```
Trihalomethanes    3114
Turbidity          3276
Potability            2
dtype: int64
```

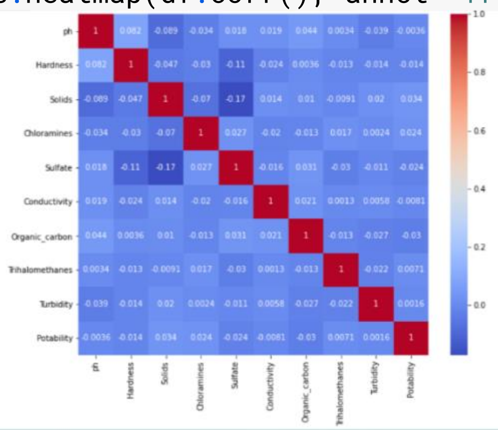
```
print(df.isnull().sum())
```

```
ph                 491
Hardness             0
Solids               0
Chloramines          0
Sulfate            781
Conductivity         0
Organic_carbon       0
Trihalomethanes    162
Turbidity            0
Potability           0
dtype: int64
```

```
df.dtypes
```

```
ph                 float64
Hardness           float64
Solids             float64
Chloramines        float64
Sulfate            float64
Conductivity       float64
Organic_carbon     float64
Trihalomethanes    float64
Turbidity          float64
Potability           int64
dtype: object
```

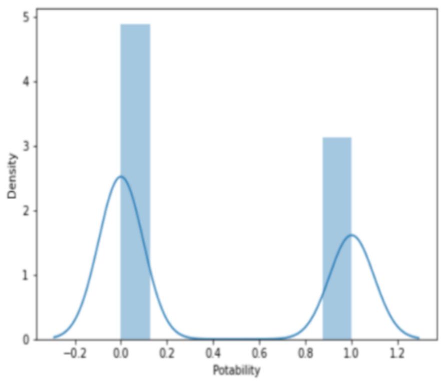
```
plt.figure(figsize=(10, 8))
sns.heatmap(df.corr(), annot= True, cmap='coolwarm')
```




```
corr = df.corr()
c1 = corr.abs().unstack()
```

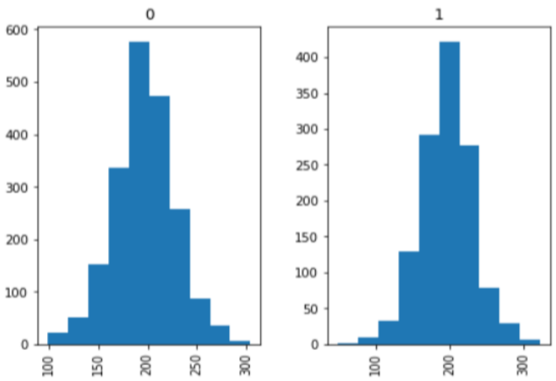
```
c1.sort_values(ascending = False)[12:24:2]
```

```
Hardness  Sulfate          0.106923
ph        Solids           0.089288
Hardness  ph               0.082096
Solids    Chloramines      0.070148
Hardness  Solids           0.046899
ph        Organic_carbon   0.043503
dtype: float64
```




```
df.hist(column='Hardness', by='Potability')
```

```
df.nunique()
df.nunique()
```



```
df.nunique()
```

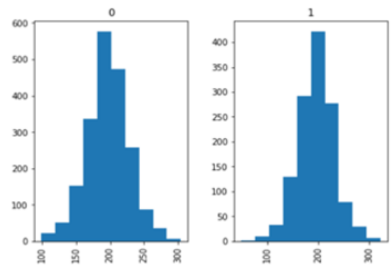
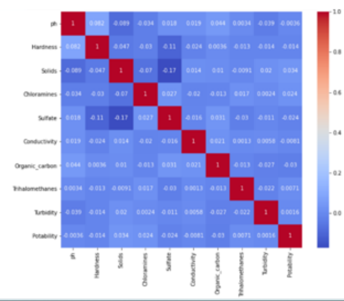
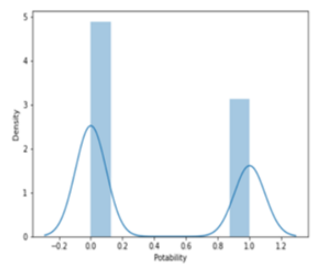
```
ph                 2785
Hardness           3276
Solids             3276
Chloramines        3276
Sulfate            2495
Conductivity       3276
Organic_carbon     3276
Trihalomethanes    3114
Turbidity          3276
Potability            2
dtype: int64
```

```
lg = accuracy_score(y_test, pred_lg)
print(lg)
```

```
0.6284658040665434
```

8.PERFORMANCE TESTING:

8.1 performance Metrics:

S.No.	Parameter	Values	Screenshot
1.	Precision	0.89	
2.	Accuracy	0.6284658040665434	
3.	Histogram		
4.	Correlation Matrix	1(perfect positive correlation)	
5.	Density Plot	0.2	

10.ADVANTAGES:

Efficiency: AI and ML can efficiently and reliably handle vast amounts of data, which can be used to find patterns and trends in data on water quality2.

Accuracy: Water quality metrics may be predicted with high accuracy by AI and ML, which can assist identify possible problems with water quality before they become an issue2.

Economical: By automating the procedure, AI and ML can assist in lowering the cost of monitoring the quality of the water3.

Water quality metrics may be monitored in real-time by AI and ML, which makes it possible to swiftly detect changes in the quality of the water4.

DISADVANTAGES:

Data quality: The caliber of the data used to train the models affects how accurately AI and ML predictions turn out. Reliability issues can result in imprecise forecasts.

Complexity: AI and ML models may be intricate and challenging to comprehend, which may make it challenging to interpret the outcomes.

Absence of transparency: AI and ML models may be challenging to read, which can make it challenging to comprehend how the models generate predictions.

Restricted scope: The quality of AI and ML models depends on the data used to train them. The models might not be able to forecast water quality characteristics with enough accuracy if the data is scarce or lacking.

All things considered, AI and ML can be helpful resources for forecasting water quality metrics, but they are not without drawbacks. Before incorporating these instruments into a program to monitor the quality of the water, it is crucial to carefully weigh the benefits and drawbacks of doing so.

11.CONCLUSION:

To sum up, while AI and ML can be helpful resources for forecasting water quality metrics, they are not without drawbacks. The rapid and accurate processing of vast volumes of data using AI and ML can be used to spot patterns and trends in data on water quality. They can also assist in the highly accurate prediction of water quality metrics, which can aid in the early detection of possible problems with the quality of the water. Furthermore, by automating the process and enabling real-time monitoring of water quality indicators, AI and ML can assist lower the cost of water quality monitoring by enabling the prompt identification of changes in water quality.

12.FUTURE SCOPE:

1. In the future, river water purity forecasts will benefit greatly from AI and ML. The following are some possible domains of utilization:

2.Real-time monitoring: Water quality metrics may be monitored in real-time using AI and ML, which makes it possible to swiftly spot changes in water quality1.

3.Early warning systems: Water quality concerns can be detected early by using AI and ML to create early warning systems that can assist stop the spread of illnesses and other health issues.2.

4.Predictive modeling: By using AI and ML to create models for water quality metrics, it is possible to identify any problems with the quality of the water before they become an issue2.

5.Data analysis: A lot of data on water quality can be analyzed using AI and ML to help spot trends and trends in the statistics on water quality2.

6.Optimization: By using AI and ML to streamline water treatment procedures, water treatment costs can be lowered and water quality can be raised.

13. APPENDIX:

Source code:

Github and Project Demo Link: