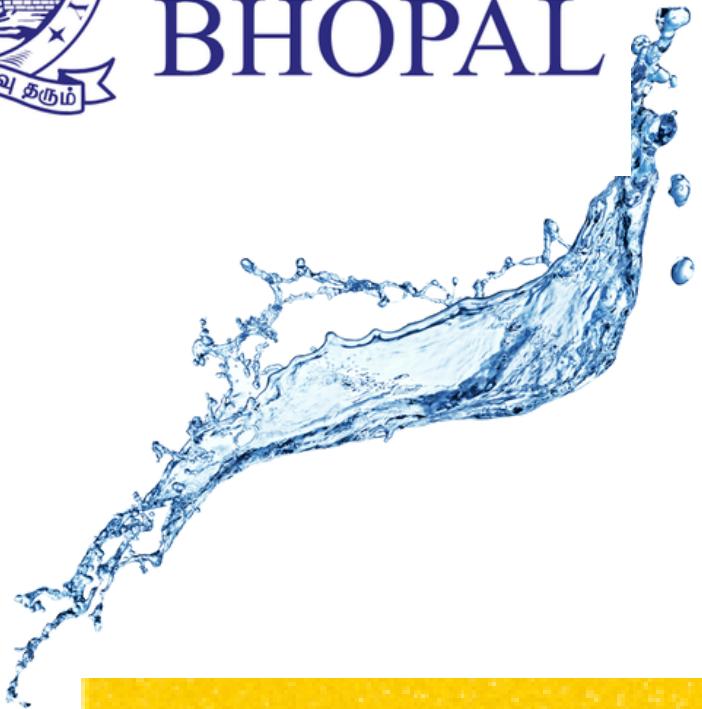
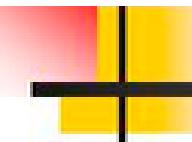




VIT[®]
—
BHOPAL



WATER QUALITY PREDICTION



APPLIED NUMERICAL METHOD
FACULTY
DR. MAMTA AGARWAL

Table of Content

Introduction

- Dataset
- Analysis of Algorithm
- Project Pipeline
- System Design
- Results and Discussion
- Conclusion



Water-Quality-Prediction-machine-learning-python

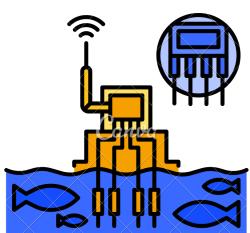




INTRODUCTION

With the population and the ever increasing need for various resources, we are faced with a dilemma on how to manage our lives. In a struggle to do that, we sometimes end up utilizing a poor or contaminated source of water for our use and thus put our health on stake.

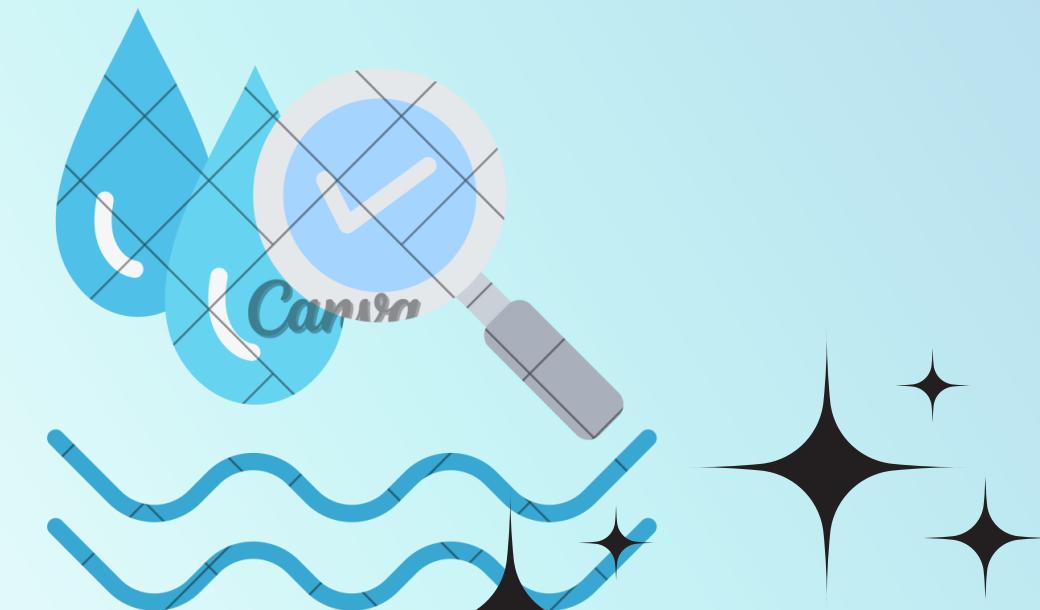
According to a recent survey of World Health Organization (WHO), more than 2.2 billion people in India face problems due to unsafe drinking water and 21% of the diseases are related to impure water.





At the current scenario, facilities are available for testing the water samples by bringing it to the water authorities. But the process is time consuming as it usually takes several weeks for the reports to be received. This causes dissatisfaction to the users.

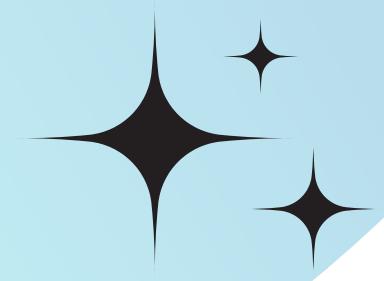
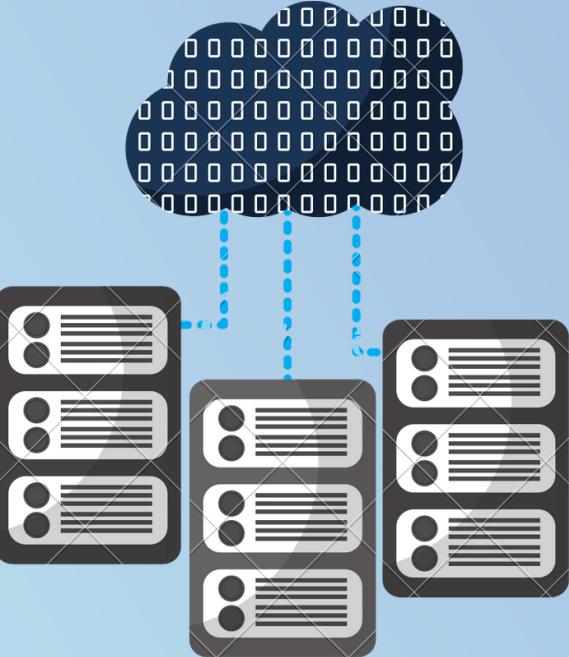
The proposed system aims to provide the solution to the same, by allowing users to monitor the water quality from a given sample of water and predicts whether water is contaminated or not.



DATA SET

Proposed system is implemented using the water potability dataset . The water_potability.csv file contains water quality , 9 features and one class variable.

- Feature lists are pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic Carbon, Trihalomethanes, Turbidity
- Potability is the class label.





CLEANED DATASET

	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability
0	NaN	204.890455	20791.318981	7.300212	368.516441	564.308854	10.379783	86.990970	2.963135	0
1	3.716080	129.422921	18630.057858	6.635246	NaN	592.885359	15.180013	56.329076	4.500656	0
2	8.099124	224.236259	19909.541732	9.275884	NaN	418.606213	16.868637	66.420093	3.055934	0
3	8.316766	214.373394	22018.417441	8.059332	356.886136	363.266516	18.436524	100.341674	4.628771	0
4	9.092223	181.101509	17978.986339	6.546600	310.135738	398.410813	11.558279	31.997993	4.075075	0

First five rows of the dataset is printed in the above figure. The dataset contains null values which is later rectified by preprocessing techniques.

Brief about the Project

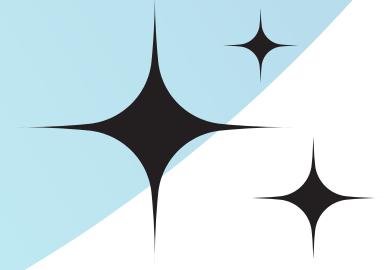
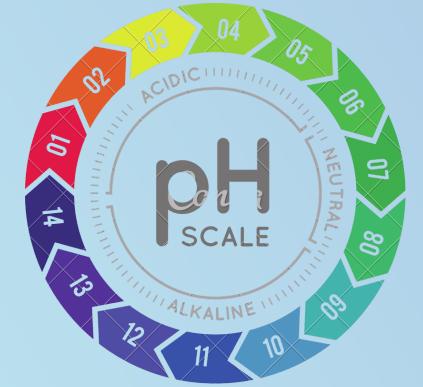
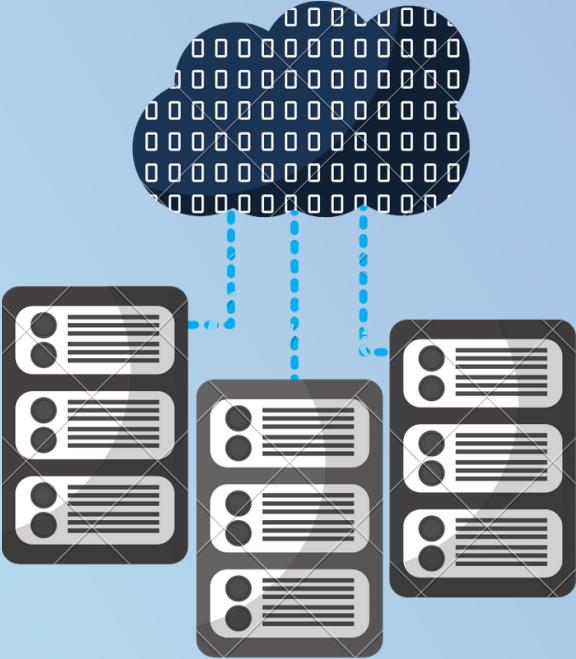
The growing worldwide emphasis on dealing with water quality is giving rise to widespread research and expanding for novel and astute monitoring systems.

Current method: Lab process where samples are taken from water bodies and testing is done in labs.

Method is time consuming, wastage of manpower and not economical.

ANN is used to solve this problem. This method eliminates chemical method of evaluating water quality parameters and is cost effective.

Also finds the measures of alkalinity, chloride and sulphate values of water using known parameters such as pH, electrical conductivity, TDS, etc.

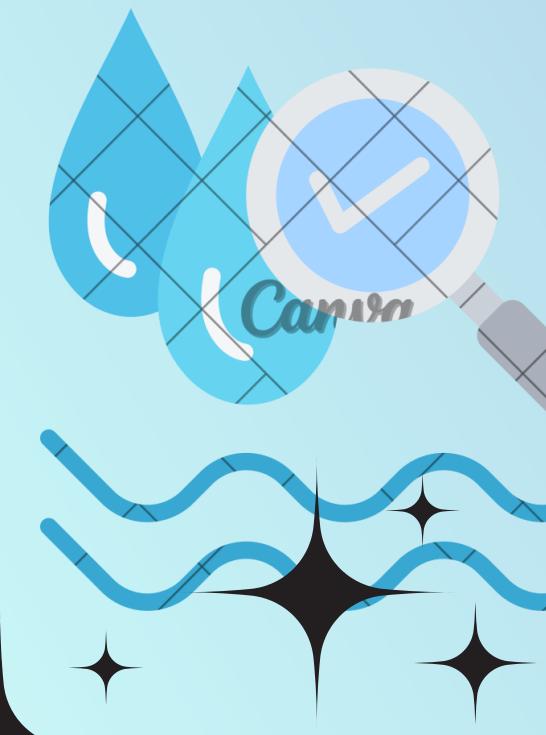




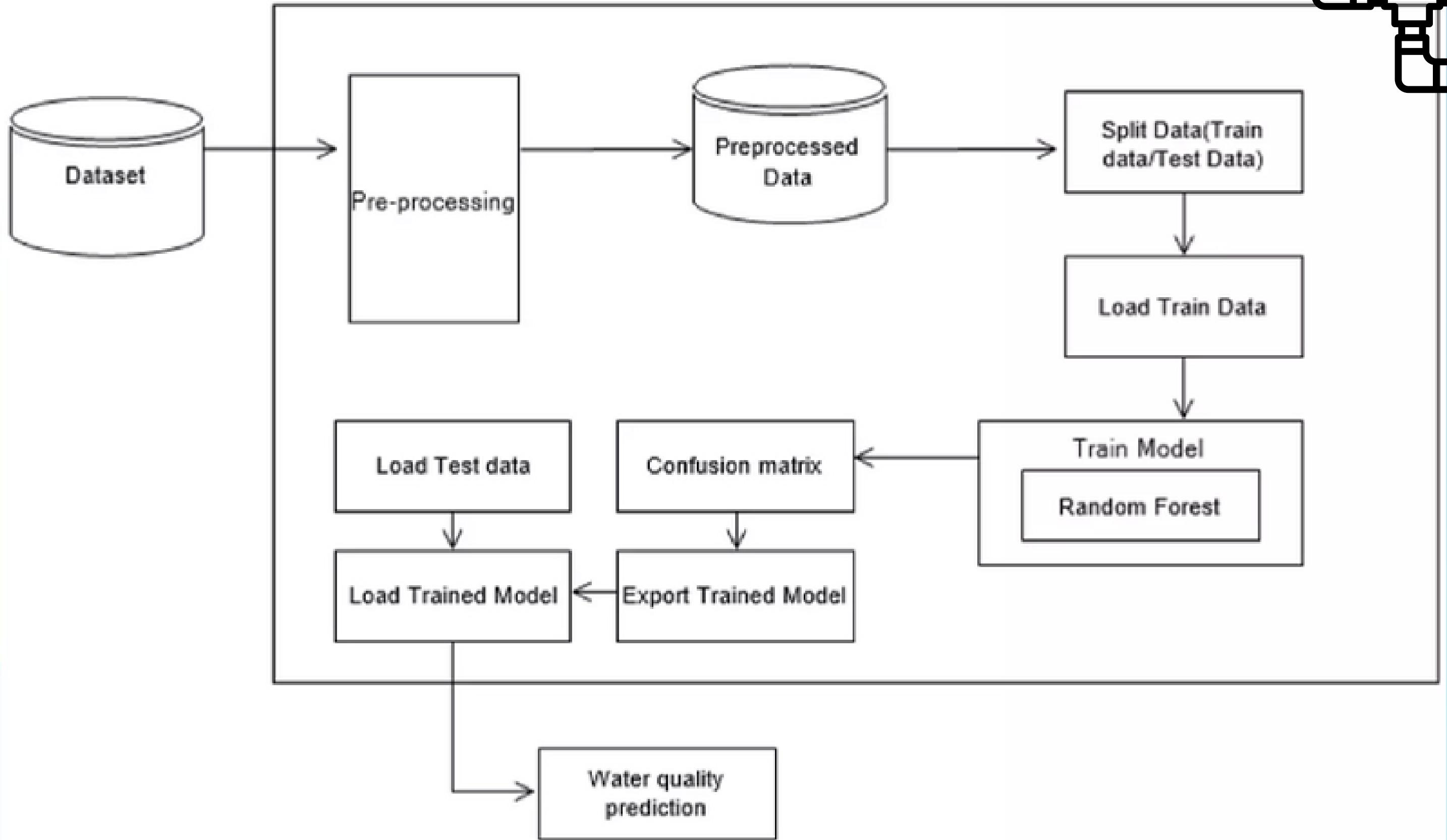
In this experimental study, we use different Machine Learning algorithms to decide the quality of the water within the Rivers and their tributaries along with the Kaggle water potability dataset.

For each algorithm to work, we used a set of parameters by which to measure the quality of the river water. Out of all the algorithms utilized, we found that random forest and K-Nearest Neighbor (KNN) were the best at achieving accurate results, with accuracy ratings of 0.6520 and 0.6469, respectively.

From this data, we may be able to determine which sections of the main River well as which tributaries, are acceptable (or healthy enough) for primary contact recreation use.

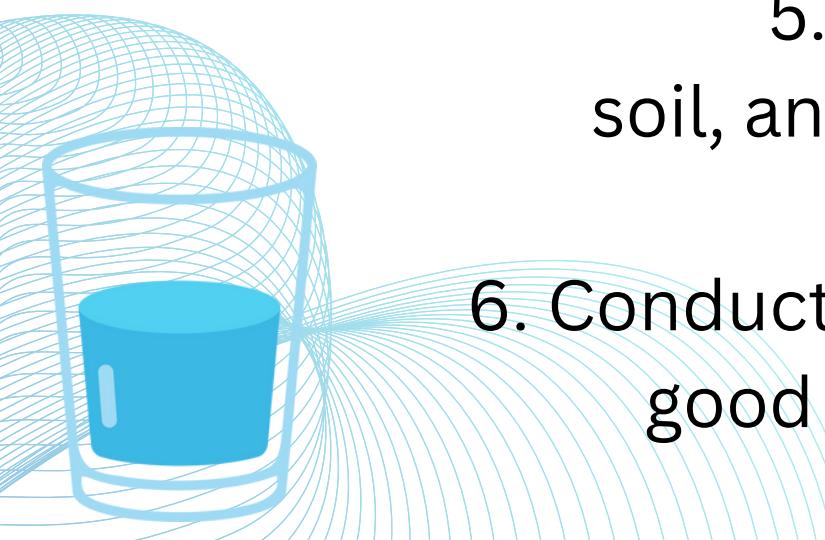


PROJECT PIPELINE



Analysis of feature variables

1. pH value: PH is an important parameter in evaluating the acid-base balance of water.
2. Hardness: Hardness is mainly caused by calcium and magnesium salts. These salts are dissolved from geologic deposits through which water travels.
3. Solids (Total dissolved solids - TDS): Water has the ability to dissolve a wide range of inorganic and some organic minerals or salts such as potassium, calcium, sodium, bicarbonates, chlorides, magnesium, sulfates etc.
4. Chloramines: Chlorine and chloramine are the major disinfectants used in Plblic water systems.
5. Sulfate: Sulfates are naturally occurring substances that are soil, and rocks. They are present in ambient air, groundwater, plan nd foo
6. Conductivity: Pure water is not a good conductor of electric current rather's a good insulator. Increase in ions concentration enhances the electrical conductivity of water.



Analysis of class variable

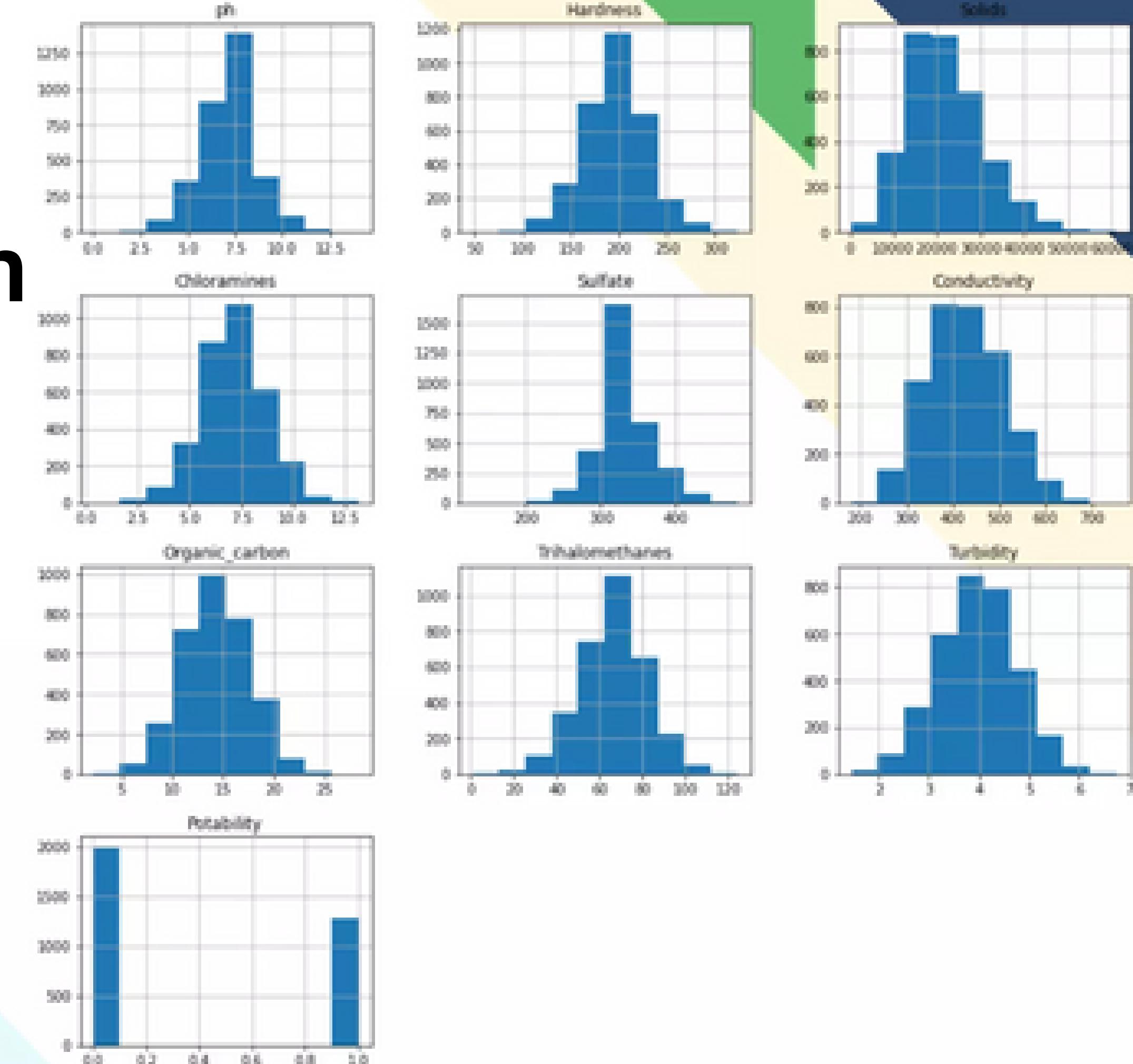
```
[14] df.Potability.value_counts()
```

```
0    1998  
1    1278  
Name: Potability, dtype: int64
```

Figure displays the count of the class variable which is namely potability. The dataset contains 1998 number of 0s which implies contaminated water and 823 number of 1s which implies potable water.

Data Visualization

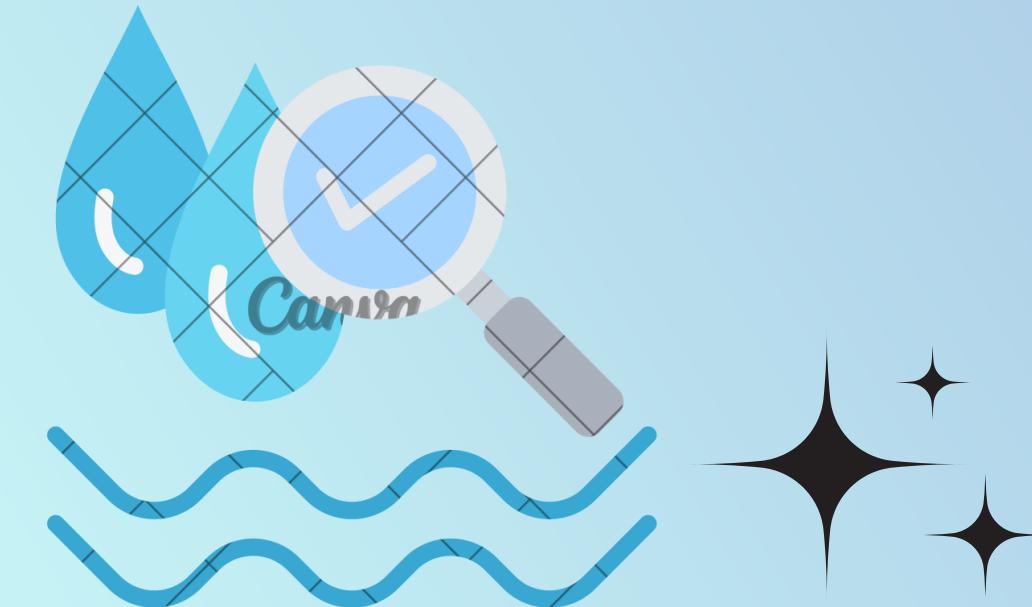
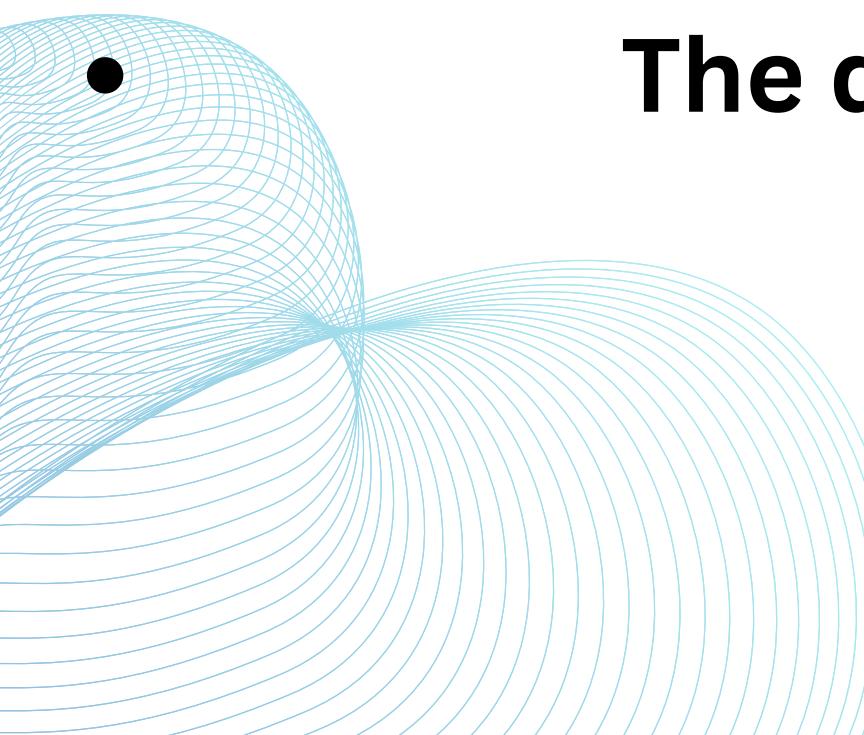
- Histogram representation of all the feature list and the class variable:





RESULTS AND DISCUSSION

- An accuracy score of 63.5% was obtained by repeated training of the model by changing the hyper parameters like criterion, n_estimators, random state etc..
- The criterion was changed from entropy to gini as the accuracy score for entropy was much lesser than gini.
- n_estimators implies the number of trees and it was increased to get the final accuracy.
- The dataset splitting ratio was changed from 3:7 to 2:8.





Thank
You

LIKE 