PROJECT PHASE-0

TEAM MEMBERS (GROUP 8):-

ADARSH R NAIR(AM.EN.U4CSE20102)

BOLLEPALLI RAVI CHANDU(AM.EN.U4CSE20117)

GUTTAPATI RUSHIKESH REDDY(AM.EN.U4CSE20131)

K.K.VENKATA KESAVA SATHVIK(AM.EN.U4CSE20137)

G.V.S KOUSHIK (AM.EN.U4CSE20173)

WATER QULAITY ASSESMENT

ABSTRACT

H2O or water however one would like to call is such an important factor for life to continue. Water is so important that without it a human being can not survive for a span of 3 days. The human body is filled with almost 75 percent of water, it is also estimated that each person on earth requires almost 20 to 50 liter's of clean water everyday in order to perform various task's like drinking, cooking and simple hygiene so from this we can understand the importance of drinking water in our blissful life which when found make a place paradise to live.

To make our life easier we are going to develop a machine learning algorithm or model which is more generalized that is going to take several characteristics into consideration and predict whether drinking found in a place('s) is suitable for drinking or not

We are going to implement this by keeping a threshold value and using several machine learning algorithm('s) like K-NN algorithm and more to predict and ensure that model which we have developed have a good accuracy and precision. In Future if possible we are also planning to collab with a diagnostic center and develop a web application or an app so that even when a person in a distinct place when put a drop of water on the fingerprint sensor can able to know whether the water is fit to consume or not.

Informal description:-

To predict or say whether the water available in our surroundings is consumable or not.

Formal description:-

Task:- To predict if the water is safe to consume or not.

Experience:- Various samples of water are given to the model so that it learns on how to predict when a new sample is given.

Performance:- When given a set of samples what proportion of samples that was safe to consume was actually found safe.

INTRODUCTION

Motivation:- We read in a article that everyday 7 people are dying by drinking contaminated water, so when we read this article we were very disappointed to see such incidents. Water is a property of nature so In some areas it maybe good to consume in some areas it may not. We want to do a project in machine learning regarding some thing but when we read that article we have decided to do project in this global topic and also we can gain experience of developing a project, a bit satisfaction that we have contributed ourself to society.

Benefits of solution:- When people know whether the water around them is not safe to drink directly, they can get a filter or such to drink the water. So by this the percentage of people drinking safe water increases and also proper sanitation this leads to decrease in diseases like cholera and all water bound diseases. When people get to know water around them is not good then they can use several techniques to purify it as water quality is important because when water found in a place is not good it will not only affect the aquatic animals and plants but also the surrounding ecosystem.

Solution use:- Since the model which we are developing is available to each individual, we can see decrease in diseases because a person is drinking safe water which is free from bacteria and several toxic materials. We are even going to update this in future as a webpage or app so that the functionalities of this model increases.

DATA SET FINALIZATION

Dataset 1 :

Link to Dataset : _DATA SET-1_

What is the data about?

This Dataset Contains water quality metrics of different water bodies and predicts if water is safe for drinking by checking different parameters, such as, pH values, Hardness, chloramines, etc. and based on these parameters it detects if water is safe for use.

For this dataset we have 9 features such as pH level, Hardness, Total dissolved solids, Chloramines, Sulfate, Conductivity, organic carbon, Trihalomethanes, and Turbidity.

pH level is an important parameter in evaluating the acid–base balance of water. Hardness checks for the capacity of water to precipitate soap caused by Calcium and Magnesium.

Total dissolved solids checks for if water has ability to dissolve a wide range of inorganic and some organic minerals or salts Chlorine and chloramine are the major disinfectants used in public water systems. Sulfates are naturally occurring substances that are found in minerals, soil, and rocks. Pure water is not a good conductor of electric current rather it's a good insulator. Increase in ions concentration enhances the electrical conductivity of water. TOC is a measure of the total amount of carbon in organic compounds in pure water.

In what all applications the dataset has been used previously, if any.

This Dataset has been used previously for predicting and checking water quality by many users, and also to check if water is safe for consumption.

Dataset 2 :

Link to Dataset : DATA SET-2

What is the data about?

Water quality is measured by several factors, such as the concentration of dissolved oxygen, bacteria levels, the amount of salt (or salinity), or the amount of material suspended in the water

(turbidity). So this Dataset represents approximately 8000 samples of water in an urban environment, with their individual concentration of Different elements and With these values it classifies the sample is safe or unsafe.

What are the number of features and describe each of the features and explain the importance?

The data set has 21 different columns representing the concentration of different elements such as aluminium, lead, ammonia, arsenic, barium, cadmium, chloramine, chromium, copper, fluoride ,bacteria, viruses, nitrates, nitrites, mercur, perchlorate, radium, selenium, silver uranium .All these elements have a threshold above which the sample becomes dangerous or unsafe. We can see water ingredients listed as column names..Idea is to have a binary classification of samples on the basis of the element concentration with their Threshold value given. Below is a description of each element and the minimum amount above which water becomes dangerous.

   Aluminium - Dangerous if greater than 2.8
   Ammonia - Dangerous if greater than 32.5
   Arsenic - Dangerous if greater than 0.01
   Barium - Dangerous if greater than 2
   Cadmium - Dangerous if greater than 0.005
   Chloramine - Dangerous if greater than 4
   Chromium - Dangerous if greater than 0.1
   Copper - Dangerous if greater than 1.3
   Fluoride - Dangerous if greater than 1.5
   Bacteria - Dangerous if greater than 0
   Viruses - Dangerous if greater than 0
   Lead - Dangerous if greater than 0.015
   Nitrates - Dangerous if greater than 10
   Nitrites - Dangerous if greater than 1
   Mercury - Dangerous if greater than 0.002
   Perchlorate - Dangerous if greater than 56

Radium - Dangerous if greater than 5
Selenium - Dangerous if greater than 0.5
Silver - Dangerous if greater than 0.1
Uranium - Dangerous if greater than 0.3
is_safe - class attribute {0 - not safe, 1 - safe}

In what all applications the dataset has been used previously, if any.

The dataset has been used by multiple users to predict the water quality and it's safe and unsafe features and has given adequate results.

Dataset 3 :

Link to Dataset : *DATA SET-3*

What is the data about?

This Dataset Contains water quality measures of different water bodies and predicts if water is safe for drinking by checking different parameters such as station code, location, state, temperature, D.O, PH, conductivity, B.O.D, NITRATE AND NITRITE , fecal coliform, and based on these parameters it detects if water is safe for use or not.

What are the number of features and describe each of the features and explain the importance?

For this dataset we have 10 features that are station code, location, state, temperature, D.O, PH, conductivity, B.O.D, NITRATE AND NITRITE and fecal coliform.
Station code gives the station code of every listed place.
Location gives the address of the place where the data acquired from.
State gives the name of the state where the values are measured.
Temperature gives average values over time of the sample.

D.O that is dissolved oxygen is a measure of how much oxygen is dissolved in the water in mg/l.

PH level is an important parameter in evaluating the acid–base balance of water

Conductivity because of the presence of H+ and OH- ions due to the self-ionization of water.Ultra-pure water has a conductivity of 0.055 μSiemens/cm or a resistivity of 18.18 mΩ·cm at 25°C.

B.O.D biochemical or chemical oxygen demand is a measure of the amount of oxygen required to remove waste organic matter from water in the process of decomposition by aerobic bacteria.

Nitrate and Nitrite : Nitrate-nitrogen occurs naturally in groundwater, usually at concentrations far below a level of concern for drinking water safety. An initial test of a new water supply is needed to determine the baseline nitrate concentration. Therefore, if the water supply has never been tested for nitrate, it should be tested.

Fecal coliform : The presence of fecal coliform bacteria in aquatic environments indicates that the water has been contaminated with the fecal material of man or other animals. ... The presence of fecal contamination is an indicator that a potential health risk exists for individuals exposed to this water.

In what all applications the dataset has been used previously, if any.

This Dataset has been used previously for predicting and checking water quality by many organizations to check if water is safe for consumption.