**Shigha velloth**

**MACHINE LEARNING ALGORITHM**

**Parametric Algorithm**

**INTRODUCTION:**

Safe drinking water is a fundamental human right, a necessity for good health, and a component of sensible health protection policies. At the national, regional, and municipal levels, this is crucial as a matter of health and development. Since the decreases in unfavourable health consequences and medical expenses outweigh the costs of implementing the interventions, it has been demonstrated that expenditures in water supply and sanitation can produce a net economic gain in some areas.

The chemical, physical, and biological content of water is used to describe water quality. Even when there is no pollution present, the water quality of rivers and lakes varies with the seasons and geographical locations. A single metric cannot be used to define good water quality. For instance, water that complies with drinking water standards can be used for irrigation, but irrigation water may not. Guidelines for water quality give fundamental scientific knowledge about water quality parameters and ecologically applicable toxicological threshold values to safeguard certain water uses.

All plants and animals need water to survive. There can be no life on earth without water. Why is water so important? Because 60 percent of our body weight is made up of water. Our bodies use water in all the cells, organs, and tissues, to help regulate body temperature and maintain other bodily functions. Because our bodies lose water through breathing, sweating, and digestion, it's crucial to rehydrate and replace water by drinking fluids and eating foods that contain water.

Let’s look at all the ways water impacts our lives

1. Water helps by creating saliva
2. It regulates body temperature
3. Water aids cognitive functions
4. Water protects the tissues, spinal cord, and joints
5. Water maximizes our physical performance
6. It helps to boost our energy levels
7. Water prevents overall dehydration

* WHO – World health organisation , they reconmmended permissible limit of some features for quality water
* Water quality standards (WQS) are provisions of state, territorial, authorized tribal or federal law approved by EPA that describe the desired condition of a water body and the means by which that condition will be protected or achieved. Water bodies can be used for purposes such as recreation (e.g. swimming and boating), scenic enjoyment, and fishing, and are the home to many aquatic organisms. To protect human health and aquatic life in these waters, states, territories and authorized tribes establish WQS. WQS form a legal basis for controlling pollutants entering the waters of the United States

**FEATURES DESCRIPTION**:

1. **pH value:** PH is an important parameter in evaluating the acid–base balance of water. It is also the indicator of acidic or alkaline condition of water status. WHO has recommended maximum permissible limit of pH from 6.5 to 8.5. The current investigation ranges were 6.52–6.83 which are in the range of WHO standards.
2. **Hardness:** Hardness is mainly caused by calcium and magnesium salts. These salts are dissolved from geologic deposits through which water travels. The length of time water is in contact with hardness producing material helps determine how much hardness there is in raw water. Hardness was originally defined as the capacity of water to precipitate soap caused by Calcium and Magnesium.
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. These minerals produced un-wanted taste and diluted color in appearance of water. This is the important parameter for the use of water. The water with high TDS value indicates that water is highly mineralized. Desirable limit for TDS is 500 mg/l and maximum limit is 1000 mg/l which prescribed for drinking purpose.
4. **Chloramines**: Chlorine and chloramine are the major disinfectants used in public water systems. Chloramines are most commonly formed when ammonia is added to chlorine to treat drinking water. Chlorine levels up to 4 milligrams per liter (mg/L or 4 parts per million (ppm)) are considered safe in drinking water.
5. **Sulfate**: Sulfates are naturally occurring substances that are found in minerals, soil, and rocks. They are present in ambient air, groundwater, plants, and food. The principal commercial use of sulfate is in the chemical industry. Sulfate concentration in seawater is about 2,700 milligrams per liter (mg/L). It ranges from 3 to 30 mg/L in most freshwater supplies, although much higher concentrations (1000 mg/L) are found in some geographic locations.
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. Generally, the amount of dissolved solids in water determines the electrical conductivity. Electrical conductivity (EC) actually measures the ionic process of a solution that enables it to transmit current. According to WHO standards, EC value should not exceeded 400 μS/cm.
7. **Organic\_carbon**: Total Organic Carbon (TOC) in source waters comes from decaying natural organic matter (NOM) as well as synthetic sources. TOC is a measure of the total amount of carbon in organic compounds in pure water. According to US EPA < 2 mg/L as TOC in treated / drinking water, and < 4 mg/Lit in source water which is use for treatment.
8. **Trihalomethanes**: THMs are chemicals which may be found in water treated with chlorine. The concentration of THMs in drinking water varies according to the level of organic material in the water, the amount of chlorine required to treat the water, and the temperature of the water that is being treated. THM levels up to 80 ppm is considered safe in drinking water.
9. **Turbidity**: The turbidity of water depends on the quantity of solid matter present in the suspended state. It is a measure of light emitting properties of water and the test is used to indicate the quality of waste discharge with respect to colloidal matter. The mean turbidity value obtained for Wondo Genet Campus (0.98 NTU) is lower than the WHO recommended value of 5.00 NTU.
10. **Potability**: Indicates if water is safe for human consumption where 1 means Potable and 0 means Not potable.

**About dataset:**

**Column Description:**

1. ph: pH of 1. water (0 to 14).

2. Hardness: Capacity of water to precipitate soap in mg/L.

3. Solids: Total dissolved solids in ppm.

4. Chloramines: Amount of Chloramines in ppm.

5. Sulfate: Amount of Sulfates dissolved in mg/L.

6. Conductivity: Electrical conductivity of water in μS/cm.

7. Organic\_carbon: Amount of organic carbon in ppm.

8. Trihalomethanes: Amount of Trihalomethanes in μg/L.

9. Turbidity: Measure of light emiting property of water in NTU.

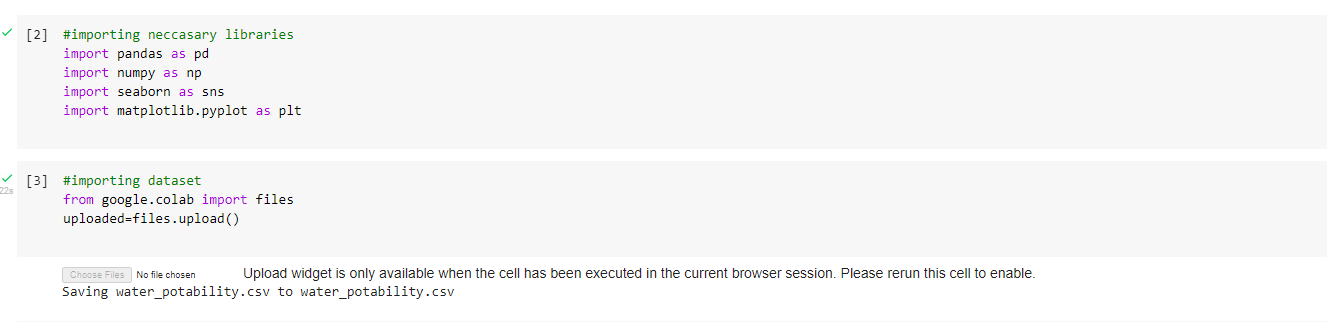
10. Potability: Indicates if water is safe for human consumption. Potable -1 and Not potable -0

**Units used:**

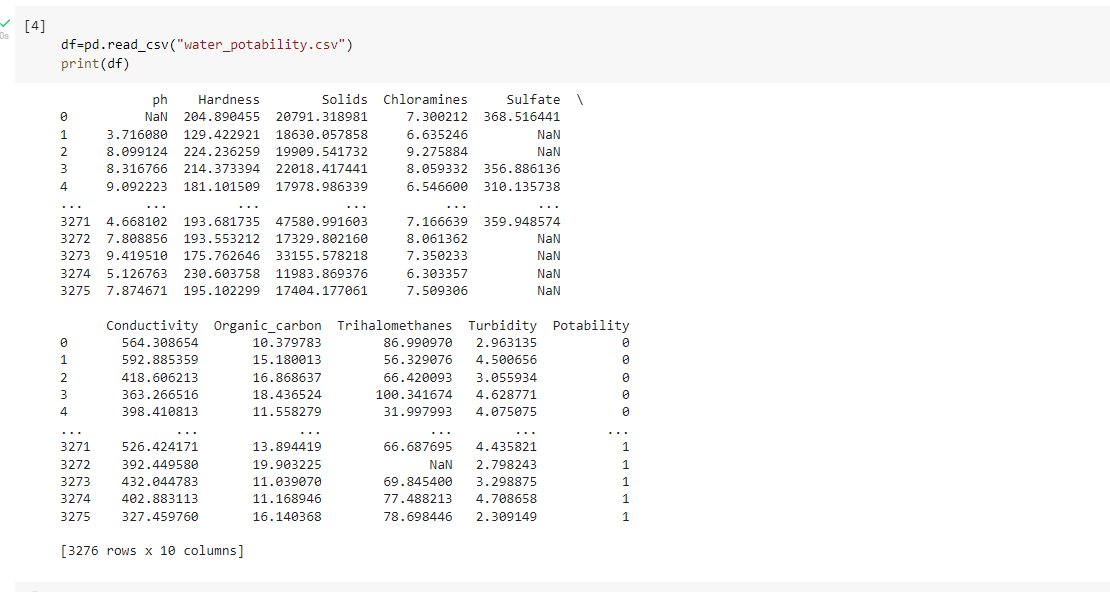
**ppm: parts per million  
μg/L: microgram per litre  
mg/L: milligram per litre**

**IMPORTING NECESSARY LIBRARIES AND**

**IMPORTIND DATASET**



#READING DATASET



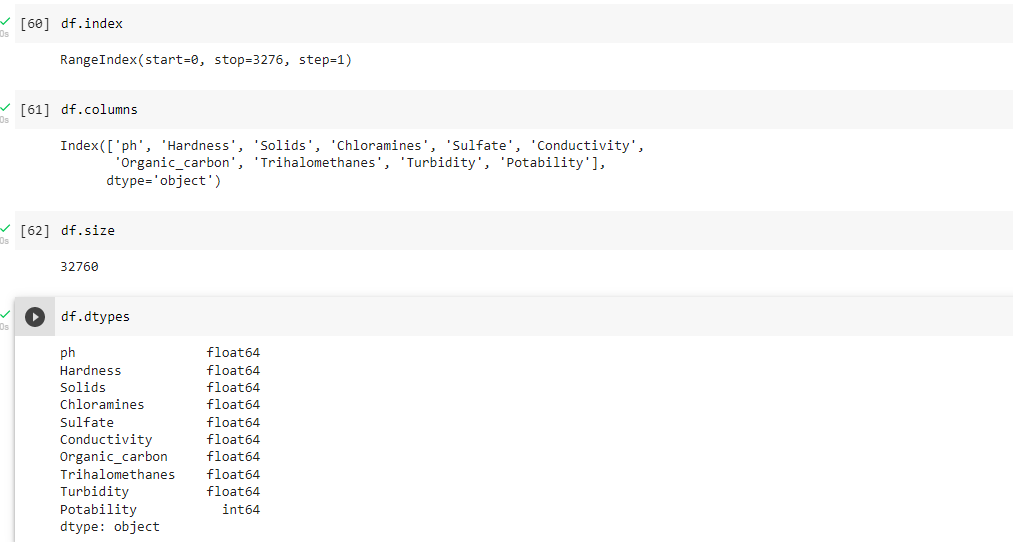
**#UNDERSTNADING DATA:**

Table

Description automatically generated

Graphical user interface, text, application

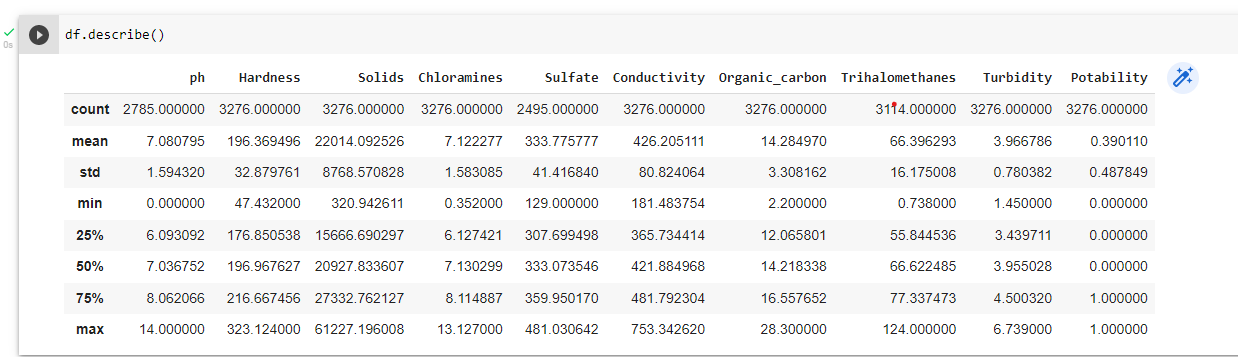
Description automatically generated



#OBTAINING SOME STATS AND INFO OF DATA

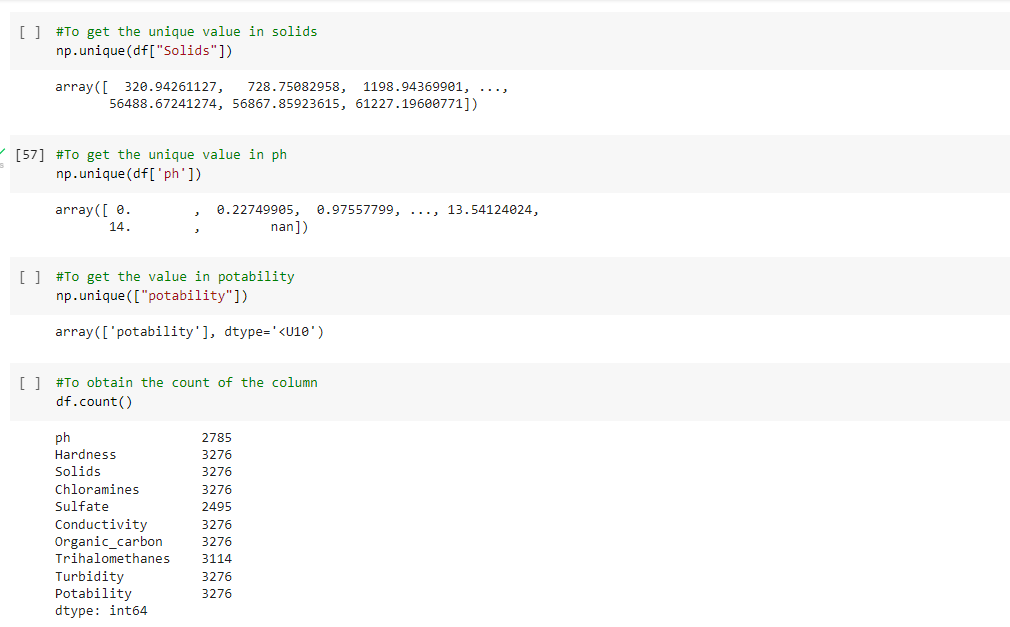
Graphical user interface, application

Description automatically generated



Graphical user interface, text, application

Description automatically generated



**#DATA CLEANING**

**#DETERMININ NULL VALUES AND DEALING WITH IT**

Graphical user interface, text, application

Description automatically generated

We found some missing values in the dataset and dealed or filled the missing values using mean of that particular column ,where

missing values are present and verified that nomore missing values are present in dataset.

Mean of the feature ph=7.080795

Mean of feature sulfate=333.775777

Mean of feature Trihalomethane=66.396293

**#DATA VISUALIZATION:**

Graphical user interface

Description automatically generated

**There is considerable difference between classes**

**Count of potable water=1278**

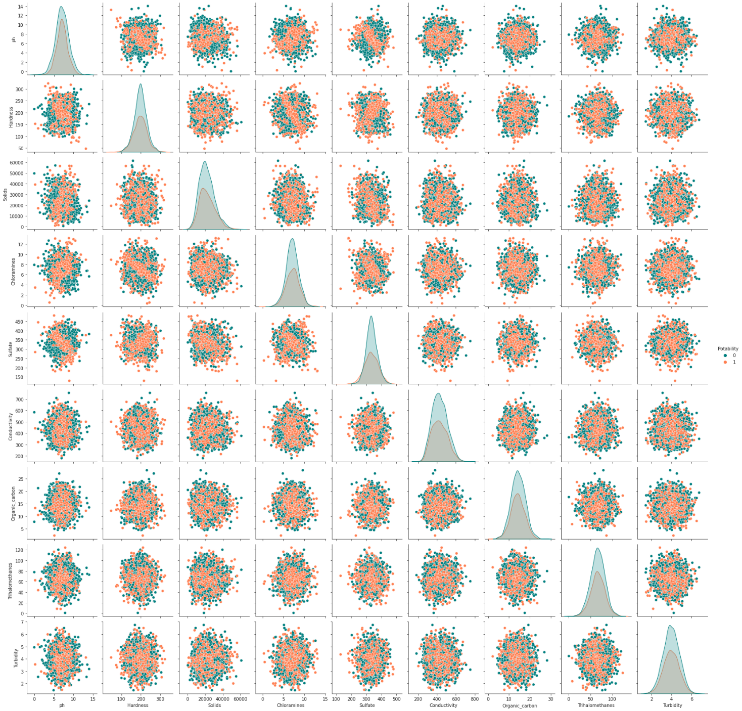
**Count of not potable water=1998**

Graphical user interface, application

Description automatically generated

**Not potable water present more compared to potable**





* **Potable water sample are always clustured in the middle of the plots.**
* **There is no apparent relationship between features**

**#PLOTTING FEATURES BY POTABILITY**

A picture containing chart

Description automatically generated

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

**levels up to 80 ppm is considered\n safe in drinking water**

* A pair plot to find correlation among features

A picture containing chart

Description automatically generated

* None of the field have high correlation

Almost 61% of water sample are not fit for drinking.

A potable water sample has less ph,Hardness,Solids,Chloramines,Sulfate,Conductivity,Organic\_carbon,Trihalomethanes and Turbidity as compared to a not potable water

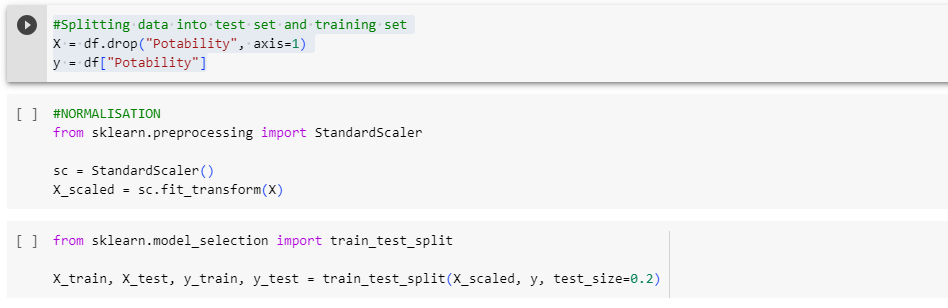
It is apparent that there is not much relation between the features

**Limits desriable for potable water:**

* Recommand limit of pH for potable water is 6.5 to 8.5
* Hardness is mainly caused by calcium and magnesium salts
* Total dissolved solids desirable limit is between 500 mg/L and 1000 mg/L
* Chloramine levels up to 4 mg/L or 4 ppm are considered safe
* Sulfate ranges from 3 to 30 mg/L in most freshwater supplies
* Electrical conductivity value should not exceed 400 μS/cm
* Organic carbon-According to US EPA < 2 mg/L as TOC in treated drinking water
* Trihalomethane levels up to 80 ppm is considered safe in drinking water
* Turbidity depends on the quantity\n of solid matter present in the suspended state

**MODELLING:**

**PARAMETRIC ALGORITHMS**

****

**A white rectangular object with black text

Description automatically generated**

**A screenshot of a computer program

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

#NEURAL NETWORK

import numpy as np

import tensorflow as tf

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Dense

from tensorflow.keras.optimizers import Adam

from sklearn.metrics import accuracy\_score

# Assuming you have X\_train, y\_train, X\_test, y\_test as your data

# Convert labels to one-hot encoding (if needed)

num\_classes = len(np.unique(y\_train))

y\_train\_onehot = tf.keras.utils.to\_categorical(y\_train, num\_classes)

y\_test\_onehot = tf.keras.utils.to\_categorical(y\_test, num\_classes)

# Create a simple neural network model

model = Sequential()

model.add(Dense(128, input\_shape=(X\_train.shape[1],), activation='relu'))

model.add(Dense(64, activation='relu'))

model.add(Dense(num\_classes, activation='softmax'))

# Compile the model

model.compile(loss='categorical\_crossentropy', optimizer=Adam(lr=0.001), metrics=['accuracy'])

# Train the model

batch\_size = 32

epochs = 10

model.fit(X\_train, y\_train\_onehot, batch\_size=batch\_size, epochs=epochs, validation\_split=0.1)

# Make predictions on the test data

y\_pred\_onehot = model.predict(X\_test)

y\_pred = np.argmax(y\_pred\_onehot, axis=1)

# Convert predictions back to class labels (if needed)

# If your labels are already one-hot encoded, skip this step

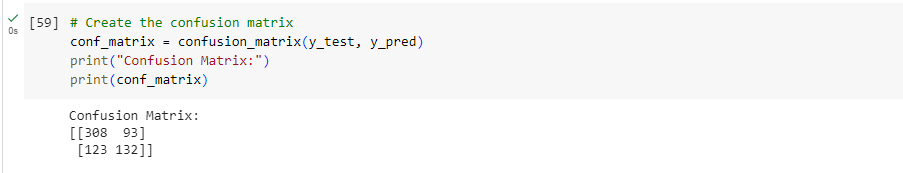
y\_pred = np.argmax(y\_pred\_onehot, axis=1)

# Calculate and print accuracy score

accuracy = accuracy\_score(y\_test, y\_pred)

print("Accuracy: {:.2f}%".format(accuracy \* 100))

****

****

**CONCLUSION:**

HERE NEURAL NETWORK ALGORITHM HAS THE HIGHEST ACCURACY OF 67%

1. Naive Bayes Classifier:

- The Naive Bayes classifier is a probabilistic algorithm based on Bayes' theorem.

- It assumes that features are conditionally independent given the class label.

- It is particularly useful for text classification and simple classification tasks.

- Naive Bayes works well with high-dimensional data and is computationally efficient.

- However, the independence assumption might not hold in complex real-world datasets.

2. Logistic Regression:

- Logistic Regression is a linear model used for binary and multi-class classification tasks.

- It models the probability of the binary outcome using the logistic function (sigmoid).

- Logistic Regression can handle both numerical and categorical features.

- It's interpretable, easy to implement, and can handle large datasets.

- Logistic Regression may not perform well on complex data with non-linear relationships.

3. Neural Network:

- Neural networks are versatile models capable of learning complex patterns from data.

- They consist of interconnected layers of neurons that learn representations from data.

- Deep Learning models, such as neural networks, can automatically extract features.

- Neural networks require more data and computational resources to train effectively.

- Fine-tuning hyperparameters and architectures is essential for optimal performance.

In summary, parametric algorithms such as Naive Bayes and Logistic Regression offer simplicity and interpretability, making them suitable for smaller datasets and tasks where the relationships are relatively simple. However, they may struggle with complex data and non-linear relationships. Neural networks, on the other hand, are powerful models capable of learning intricate patterns, but they require more data and computational resources. The choice of algorithm depends on the specific problem, data characteristics, and the trade-off between model interpretability and performance.

**THANK YOU**