**WATER QUALITY ANALYSIS**

**PHASE 3: DEVELOPMENT PART 1**

**Abstract:**

The invention discloses a data pre processing method of a water quality prediction system. The method comprises the following steps: 1) each monitoring station loads water quality data to a monitoring database, a user sends a user instruction to a network interface, a network server converts the user instruction into a structured query language (SQL) instruction and sends the SQL instruction to the monitoring database and a water quality database, and the data of the monitoring database is returned to the network server; 2) the water quality data is loaded into the memory of the network server in the form of a table in the monitoring database and the water quality database, and the data structure of the data to be analyzed is established; and 3) k-means clustering analysis is firstly performed on the water quality data, then the missing data is interpolated, and finally singular spectrum analysis is performed, thereby reducing the data noise and extracting an oscillation component and a trend component. The method disclosed by the invention reads the related water quality data from the database more quickly, improves the prediction efficiency, provides clean, accurate and concise data to a prediction model, and improves the prediction accuracy and stability.

Code:

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

import plotly as pt

from scipy import stats

import sklearn

from sklearn import preprocessing

from sklearn.impute import SimpleImputer

from math import floor,ceil

import tabulate as tb

from sklearn.preprocessing import StandardScaler

from sklearn.model\_selection import train\_test\_split

from imblearn.over\_sampling import SMOTE

from sklearn.linear\_model import LogisticRegression,RidgeClassifier,SGDClassifier,PassiveAggressiveClassifier

from sklearn.svm import SVC,LinearSVC,NuSVC

from sklearn.neighbors import KNeighborsClassifier

from sklearn.tree import DecisionTreeClassifier

from sklearn.ensemble import RandomForestClassifier,AdaBoostClassifier,GradientBoostingClassifier

from xgboost import XGBClassifier

from sklearn.naive\_bayes import GaussianNB,BernoulliNB

from sklearn.model\_selection import GridSearchCV

from sklearn.metrics import precision\_score,accuracy\_score

from sklearn.metrics import classification\_report

import warnings

warnings.filterwarnings('ignore')

Data set:

1. **ph** : Indicator of acidic or alkaline condition of water, ranging from 1 to 14.

Acceptable Limit: 6.5 to 8.5

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

Acceptable Limit: Upto 500 or 600 mg/L

1. **Solids** : Total dissolved solids (TDS) in ppm. The water with high TDS value indicates that water is highly mineralized.

Acceptable Limit: 500 - 1000 ppm

1. **Chloramines** : Amount of Chloramines in ppm.

Acceptable Limit: Upto 4 ppm

1. **Sulfate** : Amount of Sulfates dissolved in water in mg/L.

Acceptable Limit: Upto 400 mg/L

1. **Conductivity** : Electrical conductivity of water in μS/cm.

Acceptable Limit: Upto 400 μS/cm

1. **Organic\_carbon** : Amount of carbon in organic compounds in ppm.

Acceptable Limit: Less than 2 mg/L

1. **Trihalomethanes**: Amount of Trihalomethanes in μg/L.

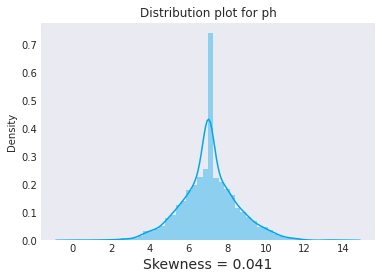
Acceptable Limit: Upto 80 ppm

1. **Turbidity**: Measure of light emiting property of water in NTU.

Acceptable Limit: 5-10 NTU

1. **Potability**: Indicates if water is safe for human consumption.

Here Potable -1 and Not potable -0.

****

In this article I will show the Exploratory Data Analysis (EDA) process with visualization techniques to analyze whether a sample of water with a certain content is potable for drinking or not.

## Data Content

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.

Data form:

| **Parameter** | **Unit** | **WHO** | **Remarks** |
| --- | --- | --- | --- |
| pH |  |  | Optimum: 6.5–8 |
| TDS | mg L−1 | − | Optimum: &lt;1200 |
| Turbidity | NTU | 5 |  |
| Ca2+ | mg L−1 | − | Optimum: &lt;250 |
| Mg2+ | mg L−1 | − | Optimum: &lt;150 |
| Na+ | mg L−1 | − | Optimum: &lt;200 |
| HCO3− | mg L−1 | − | Optimum: &lt;600 |
| SO42− | mg L−1 | − | Optimum: &lt;500 |
| Al | mg L−1 | 0.2 |  |
| As | mg L−1 | 0.01 |  |
| B | mg L−1 | 2.4 |  |
| Ba | mg L−1 | 0.7 |  |
| Cd | mg L−1 | 0.003 |  |
| Cl | mg L−1 | 5 (C) | For total chlorine |
| Cr | mg L−1 | 0.05(P) | For total chromium |
| Cu | mg L−1 | 2 |  |
| Fe | mg L−1 | 0.3 |  |
| Hg | mg L−1 | 0.006 |  |
| K | mg L−1 | − | Optimum: &lt;250 |
| Li | mg L−1 | 0.07 |  |
| Mn | mg L−1 | 0.4(C) |  |
| Mo | mg L−1 | 0.07 |  |
| Pb | mg L−1 | 0.01 |  |
| Rb | mg L−1 | − |  |
| Sb | mg L−1 | 0.02 |  |
| Sr | mg L−1 | − |  |
| Ti | mg L−1 | − |  |
| Tl | mg L−1 | − |  |
| U | mg L−1 | 0.03 (P) |  |
| Zn | mg L−1 | − | O |

**Conclusion:**

**Life without water is not possible. We need it for many things including cleaning, cooking, using the washroom, and more. Moreover, we need clean water to lead a healthy life. We can take many steps to conserve water on a national level as well as an individual level.**

**By preserving and supporting good water quality we benefit the environment, public health, and the protection of water resources for future generations.**