WATER QUALITY ANALYSIS

CONSIDER EXPLORING ANAMALY DETECTION TECHNIQUES TO IDENTIFY UNUSUAL PATTERNS IN WATER QUALITY PARAMETERS

OBJECTIVES

The water quality prediction problem is classified into five categories based on the size of a water quality dataset. The main objectives of this study are summarized as follows:

Objective-1: A first analysis was conducted on the available data to clean, normalize and perform feature selection on the water quality measures, and therefore, to obtain the minimum relevant subset that allows high precision with low cost. In this way, expensive and cumbersome lab analysis with specific sensors can be avoided in further similar analyses.

Objective-2: A series of representative supervised prediction (prediction, classification and regression) algorithms were tested on the dataset worked here. The complete methodology is proposed in the context of water quality numerical analysis.

TECHNIQUES

The contribution is:

- To carry out a systematic literature review in order to ascertain the current ML techniques used for the WQAD (Water Quality Anomaly Detection) problem.
- To highlight the shortcomings and limitations of these current methods
- To propose a hybrid DL-ELM framework in WQAD, which could be investigated further
- To recommend future research directions T

PROGRAM

```
import numpy as np

import pandas as pd


import seaborn as sns;

import matplotlib.pyplot as plt;

import plotly.express as px;

import missingno as msno;


from sklearn.tree import DecisionTreeClassifier;

from sklearn.ensemble import RandomForestClassifier;

from sklearn.model_selection import RandomizedSearchCV, RepeatedStratifiedKFol

d, train_test_split;

from sklearn.metrics import precision_score, confusion_matrix;


from sklearn import tree;


import os

for dirname, _, filenames in os.walk('/kaggle/input'):

    for filename in filenames:

        print(os.path.join(dirname, filename))
```

OUTPUT:

	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic carbon	Trihalomethanes	Turbidity	Potability
0	NaN	204.890455	20791.318981	7.300212	368.516441	564.308654	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