SPRINT 1

CODE AND TEST CASES

Date	18 October 2022
Team ID	PNT2022TMID52665
Project Name	Project – Efficient Water Quality Analysis and Prediction using Machine Learning
Maximum Marks	8 Marks

Data Collection & Data pre-processing:

```python
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
```
[1]   Python

```python
data = pd.read_csv('water_dataX.csv',encoding='ISO-8859-1',low_memory=False)
```
[2]   Python

```python
data.head()
```
[3]   Python

	STATION CODE	LOCATIONS	STATE	Temp	D.O. (mg/l)	PH	CONDUCTIVITY (Âµmhos/cm)	B.O.D. (mg/l)	NITRATENAN N+ NITRITENANN (mg/l)	FECAL COLIFORM (MPN/100ml)	TOTAL COLIFORM (MPN/100ml)Mean	year
0	1393	DAMANGANGA AT D/S OF MADHUBAN, DAMAN	DAMAN & DIU	30.6	6.7	7.5	203	NAN	0.1	11	27	2014
1	1399	ZUARI AT D/S OF PT. WHERE KUMBARJRIA CANAL JOI...	GOA	29.8	5.7	7.2	189	2	0.2	4953	8391	2014
2	1475	ZUARI AT PANCHAWADI	GOA	29.5	6.3	6.9	179	1.7	0.1	3243	5330	2014
3	3181	RIVER ZUARI AT BORIM BRIDGE	GOA	29.7	5.8	6.9	64	3.8	0.5	5382	8443	2014
4	3182	RIVER ZUARI AT MARCAIM JETTY	GOA	29.5	5.8	7.3	83	1.9	0.4	3428	5500	2014

```python
data.describe()
```
[4]   Python

	year
count	1991.000000
mean	2010.038172
std	3.057333
min	2003.000000
25%	2008.000000
50%	2011.000000
75%	2013.000000
max	2014.000000

```python
data.info()
```
[5]   Python

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1991 entries, 0 to 1990
Data columns (total 12 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   STATION CODE     1991 non-null   object
 1   LOCATIONS         1991 non-null   object
 2   STATE             1991 non-null   object
 3   Temp              1991 non-null   object
 4   D.O. (mg/l)       1991 non-null   object
 5   PH                1991 non-null   object
```

Sprint1.ipynb ✕   Sprint2.ipynb

D: > study > ibm > Sprint1.ipynb > import numpy as np

+ Code + Markdown | ▷ Run All ≡ Clear Outputs of All Cells ↺ Restart | Variables ≡ Outline …   Python 3.8.7 64-bit

```
 7   B.O.D. (mg/l)                      1991 non-null   object
 8   NITRATENAN N+ NITRITENANN (mg/l)   1991 non-null   object
 9   FECAL COLIFORM (MPN/100ml)         1991 non-null   object
 10  TOTAL COLIFORM (MPN/100ml)Mean     1991 non-null   object
 11  year                               1991 non-null   int64
dtypes: int64(1), object(11)
memory usage: 186.8+ KB
```

```python
data.shape
```

[6]   Python

```
(1991, 12)
```

```python
data.isnull().any()
```

[7]   Python

```
ï»¿STATION CODE                     False
LOCATIONS                           False
STATE                               False
Temp                                False
D.O. (mg/l)                         False
PH                                  False
CONDUCTIVITY (Âµmhos/cm)            False
B.O.D. (mg/l)                       False
NITRATENAN N+ NITRITENANN (mg/l)    False
FECAL COLIFORM (MPN/100ml)          False
TOTAL COLIFORM (MPN/100ml)Mean      False
year                                False
dtype: bool
```

Sprint1.ipynb ✕   Sprint2.ipynb

D: > study > ibm > Sprint1.ipynb > import numpy as np

+ Code + Markdown | ▷ Run All ≡ Clear Outputs of All Cells ↺ Restart | Variables ≡ Outline …   Python 3.8.7 64-bit

```python
data.isnull().sum()
```

[8]   Python

```
ï»¿STATION CODE                     0
LOCATIONS                           0
STATE                               0
Temp                                0
D.O. (mg/l)                         0
PH                                  0
CONDUCTIVITY (Âµmhos/cm)            0
B.O.D. (mg/l)                       0
NITRATENAN N+ NITRITENANN (mg/l)    0
FECAL COLIFORM (MPN/100ml)          0
TOTAL COLIFORM (MPN/100ml)Mean      0
year                                0
dtype: int64
```

```python
data.dtypes
```

[9]   Python

```
ï»¿STATION CODE                     object
LOCATIONS                           object
STATE                               object
Temp                                object
D.O. (mg/l)                         object
PH                                  object
CONDUCTIVITY (Âµmhos/cm)            object
B.O.D. (mg/l)                       object
NITRATENAN N+ NITRITENANN (mg/l)    object
FECAL COLIFORM (MPN/100ml)          object
```

```
B.O.D. (mg/l)                          object
NITRATENAN N+ NITRITENANN (mg/l)       object
FECAL COLIFORM (MPN/100ml)             object
TOTAL COLIFORM (MPN/100ml)Mean         object
year                                   int64
dtype: object
```

```python
data['Temp']=pd.to_numeric(data['Temp'],errors='coerce')
data['D.O. (mg/l)']=pd.to_numeric(data['D.O. (mg/l)'],errors='coerce')
data['PH']=pd.to_numeric(data['PH'],errors='coerce')
data['B.O.D. (mg/l)']=pd.to_numeric(data['B.O.D. (mg/l)'],errors='coerce')
data['CONDUCTIVITY (Âµmhos/cm)']=pd.to_numeric(data['CONDUCTIVITY (Âµmhos/cm)'],errors='coerce')
data['NITRATENAN N+ NITRITENANN (mg/l)']=pd.to_numeric(data['NITRATENAN N+ NITRITENANN (mg/l)'],errors='coerce')
data['TOTAL COLIFORM (MPN/100ml)Mean']=pd.to_numeric(data['TOTAL COLIFORM (MPN/100ml)Mean'],errors='coerce')
data.dtypes
```

```
ï»¿STATION CODE                        object
LOCATIONS                              object
STATE                                  object
Temp                                   float64
D.O. (mg/l)                            float64
PH                                     float64
CONDUCTIVITY (Âµmhos/cm)               float64
B.O.D. (mg/l)                          float64
NITRATENAN N+ NITRITENANN (mg/l)       float64
FECAL COLIFORM (MPN/100ml)             object
TOTAL COLIFORM (MPN/100ml)Mean         float64
year                                   int64
dtype: object
```

```python
data.isnull().sum()
```

```
ï»¿STATION CODE                        0
LOCATIONS                              0
STATE                                  0
Temp                                   92
D.O. (mg/l)                            31
PH                                     8
CONDUCTIVITY (Âµmhos/cm)               25
B.O.D. (mg/l)                          43
NITRATENAN N+ NITRITENANN (mg/l)       225
FECAL COLIFORM (MPN/100ml)             0
TOTAL COLIFORM (MPN/100ml)Mean         132
year                                   0
dtype: int64
```

```python
data['Temp'].fillna(data['Temp'].mean(),inplace=True)
data['D.O. (mg/l)'].fillna(data['D.O. (mg/l)'].mean(),inplace=True)
data['PH'].fillna(data['PH'].mean(),inplace=True)
data['CONDUCTIVITY (Âµmhos/cm)'].fillna(data['CONDUCTIVITY (Âµmhos/cm)'].mean(),inplace=True)
data['B.O.D. (mg/l)'].fillna(data['B.O.D. (mg/l)'].mean(),inplace=True)
data['NITRATENAN N+ NITRITENANN (mg/l)'].fillna(data['NITRATENAN N+ NITRITENANN (mg/l)'].mean(),inplace=True)
data['TOTAL COLIFORM (MPN/100ml)Mean'].fillna(data['TOTAL COLIFORM (MPN/100ml)Mean'].mean(),inplace=True)
```

```python
data.drop(["FECAL COLIFORM (MPN/100ml)"],axis=1,inplace=True)
```

File Edit Selection View Go Run Terminal Help

Sprint1.ipynb ×   Sprint2.ipynb

D: > study > ibm > Sprint1.ipynb > import numpy as np

+ Code + Markdown | ▷ Run All ≡ Clear Outputs of All Cells ↻ Restart | ▥ Variables ≡ Outline ⋯

Python 3.8.7 64-bit

```python
data=data.rename(columns = {'D.O. (mg/l)': 'do'})
data=data.rename(columns = {'CONDUCTIVITY (Âµmhos/cm)': 'co'})
data=data.rename(columns = {'B.O.D. (mg/l)': 'bod'})
data=data.rename(columns = {'NITRATENAN N+ NITRITENANN (mg/l)': 'na'})
data=data.rename(columns = {'TOTAL COLIFORM (MPN/100ml)Mean': 'tc'})
data=data.rename(columns = {'STATION CODE': 'station'})
data=data.rename(columns = {'LOCATIONS': 'location'})
data=data.rename(columns = {'STATE': 'state'})
data=data.rename(columns = {'PH': 'ph'})
data
```

[14]

Python

	STATION CODE	location	state	Temp	do	ph	co	bod	na	tc	year
0	1393	DAMANGANGA AT D/S OF MADHUBAN, DAMAN	DAMAN & DIU	30.600000	6.7	7.5	203.0	6.940049	0.100000	27.0	2014
1	1399	ZUARI AT D/S OF PT. WHERE KUMBARJRIA CANAL JOI...	GOA	29.800000	5.7	7.2	189.0	2.000000	0.200000	8391.0	2014
2	1475	ZUARI AT PANCHAWADI	GOA	29.500000	6.3	6.9	179.0	1.700000	0.100000	5330.0	2014
3	3181	RIVER ZUARI AT BORIM BRIDGE	GOA	29.700000	5.8	6.9	64.0	3.800000	0.500000	8443.0	2014
4	3182	RIVER ZUARI AT MARCAIM JETTY	GOA	29.500000	5.8	7.3	83.0	1.900000	0.400000	5500.0	2014
...	...	...	...	...	...	...	...	...	...	...	...
1986	1330	TAMBIRAPARANI AT ARUMUGANERI, TAMILNADU	NAN	26.209814	7.9	738.0	7.2	2.700000	0.518000	202.0	2003
1987	1450	PALAR AT VANIYAMBADI WATER SUPPLY HEAD WORK, T...	NAN	29.000000	7.5	585.0	6.3	2.600000	0.155000	315.0	2003
1988	1403	GUMTI AT U/S SOUTH TRIPURA,TRIPURA	NAN	28.000000	7.6	98.0	6.2	1.200000	1.623079	570.0	2003
1989	1404	GUMTI AT D/S SOUTH TRIPURA, TRIPURA	NAN	28.000000	7.7	91.0	6.5	1.300000	1.623079	562.0	2003
1990	1726	CHANDRAPUR, AGARTALA D/S OF HAORA RIVER, TRIPURA	NAN	29.000000	7.6	110.0	5.7	1.100000	1.623079	546.0	2003

1991 rows × 11 columns

⊗ 0 ⚠ 6   ↻ Discovering Python Interpreters

Jupyter Server: Local   Cell 1 of 15   Go Live