

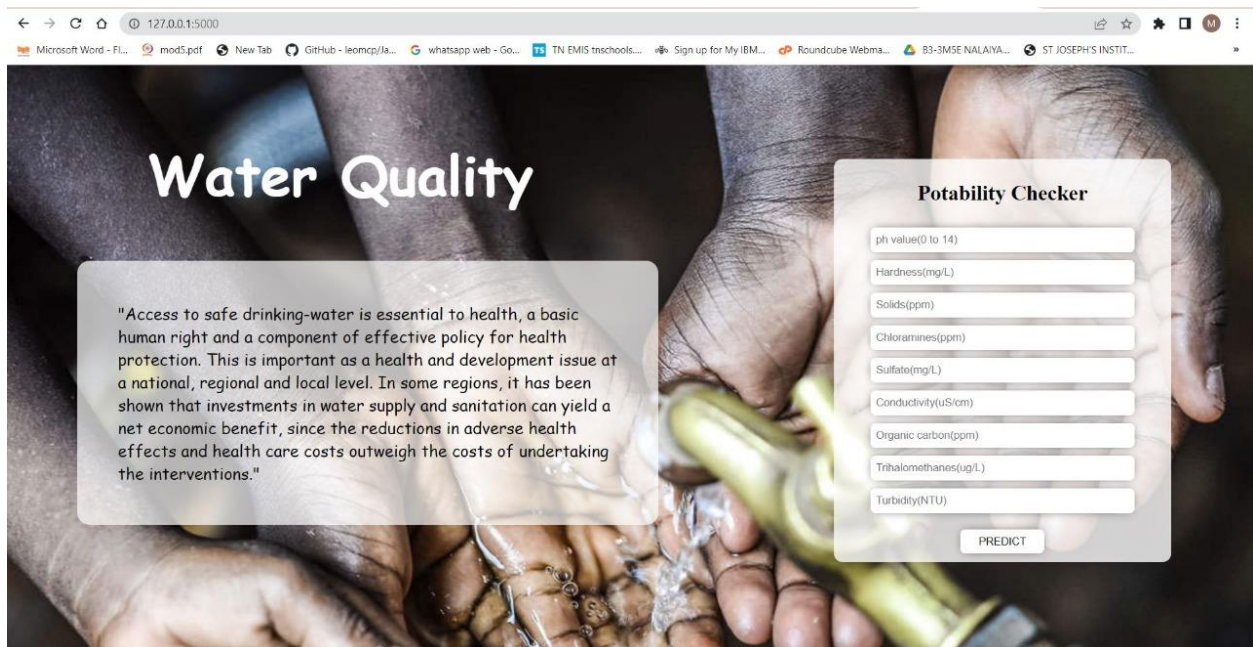
Project Development Phase Model Performance Test

Date	18 November 2022
Team ID	PNT2022TMID28091
Project Name	Efficient Water Quality Analysis & Prediction using Machine Learning
Maximum Marks	10 Marks

Model Performance Testing:

Project team shall fill the following information in model performance testing template.

Dashboard design



Data Responsiveness

```
In [67]: df = pd.read_csv('water_potability.csv')
```

```
In [68]: df.head()
```

```
Out[68]:
```

	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

```
In [69]: df.isnull().sum()
```

```
Out[69]: ph                491
Hardness                0
Solids                  0
Chloramines             0
Sulfate                 781
Conductivity            0
Organic_carbon          0
Trihalomethanes        162
Turbidity               0
Potability              0
dtype: int64
```

```
In [70]: df.drop_duplicates(inplace=True)
```

```
In [71]: df.dropna(how='all', inplace=True)
```

```
In [72]: df.describe()
```

```
Out[72]:
```

	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability
count	2785.000000	3276.000000	3276.000000	3276.000000	2495.000000	3276.000000	3276.000000	3114.000000	3276.000000	3276.000000
mean	7.080795	196.369496	22014.092526	7.122277	333.775777	426.205111	14.284970	66.396293	3.966786	0.390110
std	1.594320	32.879761	8768.570828	1.583085	41.416840	80.824064	3.308162	16.175008	0.780382	0.487849
min	0.000000	47.432000	320.942611	0.352000	129.000000	181.483754	2.200000	0.738000	1.450000	0.000000
25%	6.093092	176.850538	15666.690297	6.127421	307.699498	365.734414	12.065801	55.844536	3.439711	0.000000
50%	7.036752	196.967627	20927.833607	7.130299	333.073546	421.884968	14.218338	66.622485	3.955028	0.000000
75%	8.062066	216.667456	27332.762127	8.114887	359.950170	481.792304	16.557652	77.337473	4.500320	1.000000
max	10.000000	224.236259	19909.541732	9.275884	368.516441	592.885359	15.180013	100.341674	4.628771	1.000000

Amount Data to Rendered (DB2 Metrics)

```
In [67]: df = pd.read_csv('water_potability.csv')
```

```
In [68]: df.head()
```

```
Out[68]:
```

	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

```
In [69]: df.isnull().sum()
```

```
Out[69]: ph          491
Hardness         0
Solids           0
Chloramines       0
Sulfate          781
Conductivity      0
Organic_carbon    0
Trihalomethanes  162
Turbidity         0
Potability        0
dtype: int64
```

```
In [70]: df.drop_duplicates(inplace=True)
```

```
In [71]: df.dropna(how='all', inplace=True)
```

```
In [72]: df.describe()
```

```
Out[72]:
```

	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability
count	2785.000000	3276.000000	3276.000000	3276.000000	2495.000000	3276.000000	3276.000000	3114.000000	3276.000000	3276.000000
mean	7.080795	196.369496	22014.092526	7.122277	333.775777	426.205111	14.284970	66.396293	3.966786	0.390110
std	1.594320	32.879761	8768.570828	1.583085	41.416840	80.824064	3.308162	16.175008	0.780382	0.487849
min	0.000000	47.432000	320.942611	0.352000	129.000000	181.483754	2.200000	0.738000	1.450000	0.000000
25%	6.093092	176.850538	15666.690297	6.127421	307.699498	365.734414	12.065801	55.844536	3.439711	0.000000
50%	7.036752	196.967627	20927.833607	7.130299	333.073546	421.884968	14.218338	66.622485	3.955028	0.000000
75%	8.062066	216.667456	27332.762127	8.114887	359.950170	481.792304	16.557652	77.337473	4.500320	1.000000

Utilization of Data Filters

```
In [67]: df = pd.read_csv('water_potability.csv')
```

```
In [68]: df.head()
```

```
Out[68]:
```

	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

```
In [69]: df.isnull().sum()
```

```
Out[69]: ph          491
Hardness         0
Solids           0
Chloramines       0
Sulfate          781
Conductivity      0
Organic_carbon    0
Trihalomethanes  162
Turbidity         0
Potability        0
dtype: int64
```

```
In [70]: df.drop_duplicates(inplace=True)
```

```
In [71]: df.dropna(how='all', inplace=True)
```

```
In [72]: df.describe()
```

```
Out[72]:
```

	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability
count	2785.000000	3276.000000	3276.000000	3276.000000	2495.000000	3276.000000	3276.000000	3114.000000	3276.000000	3276.000000
mean	7.080795	196.369496	22014.092526	7.122277	333.775777	426.205111	14.284970	66.396293	3.966786	0.390110
std	1.594320	32.879761	8768.570828	1.583085	41.416840	80.824064	3.308162	16.175008	0.780382	0.487849
min	0.000000	47.432000	320.942611	0.352000	129.000000	181.483754	2.200000	0.738000	1.450000	0.000000
25%	6.093092	176.850538	15666.690297	6.127421	307.699498	365.734414	12.065801	55.844536	3.439711	0.000000
50%	7.036752	196.967627	20927.833607	7.130299	333.073546	421.884968	14.218338	66.622485	3.955028	0.000000
75%	8.062066	216.667456	27332.762127	8.114887	359.950170	481.792304	16.557652	77.337473	4.500320	1.000000

Effective User Story

Scenario 1:

Water Quality

"Access to safe drinking-water is essential to health, a basic human right and a component of effective policy for health protection. This is important as a health and development issue at a national, regional and local level. In some regions, it has been shown that investments in water supply and sanitation can yield a net economic benefit, since the reductions in adverse health effects and health care costs outweigh the costs of undertaking the interventions."

Potability Checker

7
20
4
89
1
43
1
1
4

PREDICT

Water Quality

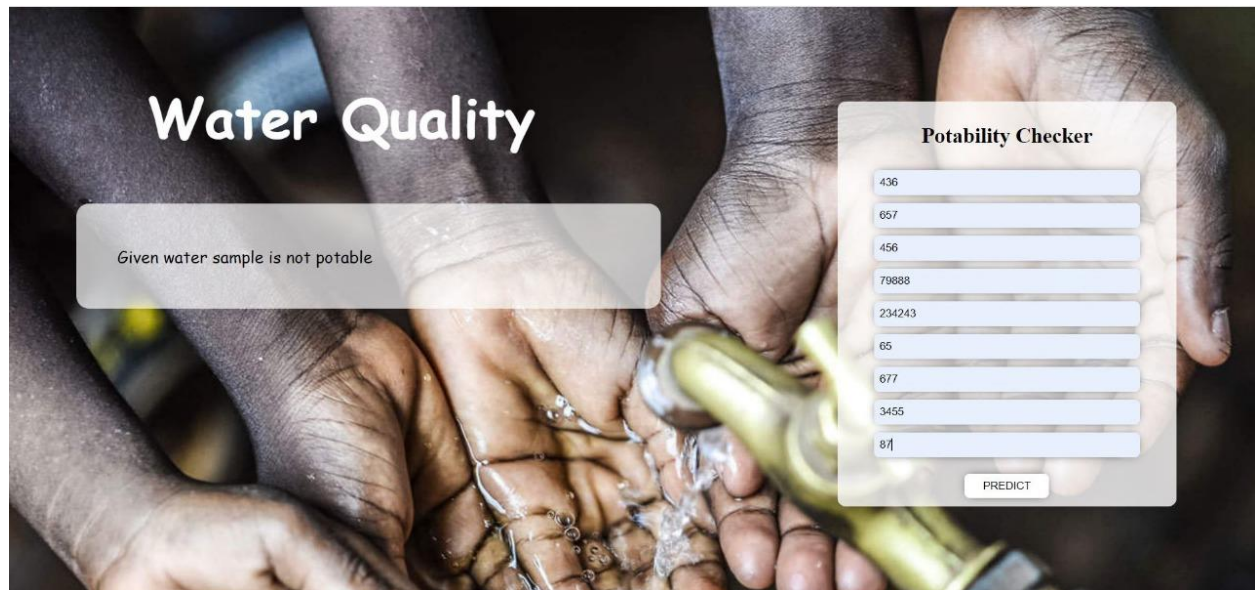
Given water sample is Potable

Potability Checker

ph value(0 to 14)
Hardness(mg/L)
Solids(ppm)
Chloramines(ppm)
Sulfate(mg/L)
Conductivity(uS/cm)
Organic carbon(ppm)
Trihalomethanes(ug/L)
Turbidity(NTU)

PREDICT

Scenario 2:



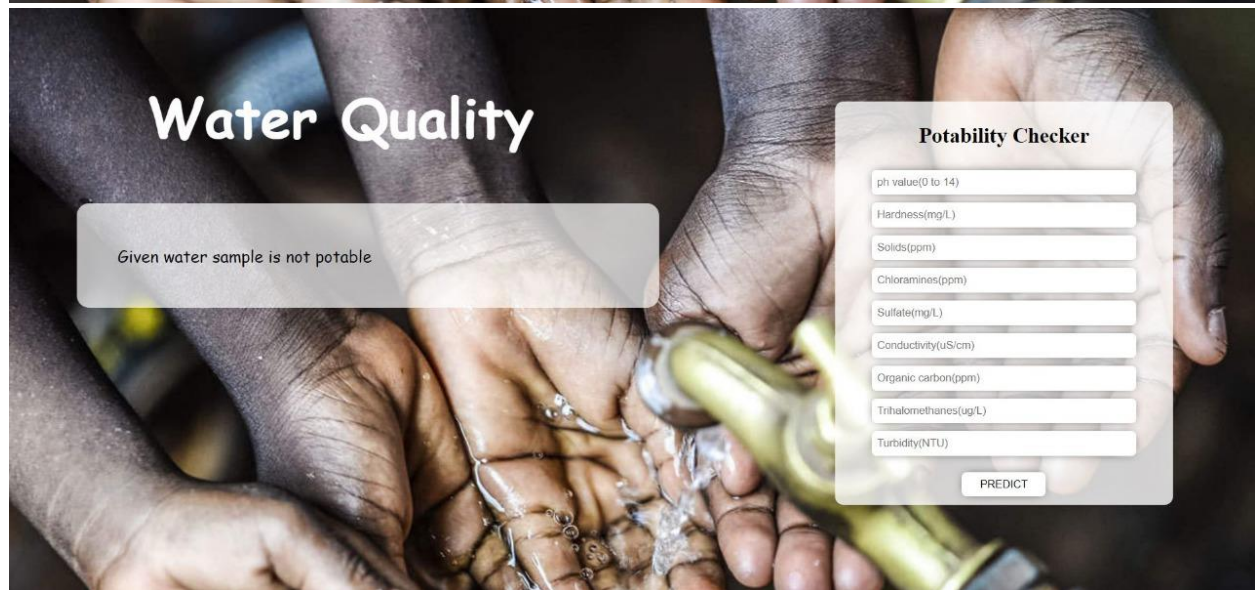
Water Quality

Given water sample is not potable

Potability Checker

436
657
456
79888
234243
65
677
3455
87

PREDICT



Water Quality

Given water sample is not potable

Potability Checker

ph value(0 to 14)
Hardness(mg/L)
Solids(ppm)
Chloramines(ppm)
Sulfate(mg/L)
Conductivity(uS/cm)
Organic carbon(ppm)
Trihalomethanes(ug/L)
Turbidity(NTU)

PREDICT

Descriptive Reports:

