Water quality analysis using ML Algorithm

Machine Learning Algorithms: That we are Going to use

1. Logistic Regression

2. Decision Tree

3. Random Forest

4. K-Nearest Neighbours

5. Support Vector Machine

Copper (mg/l)	Zinc (mg/l)	Lead (mg/l)	Nickel (mg/l)	Total Microbial count cfu/ml	E.coli /ml	Total coliform bacteria (MPN)/ml
0.05	5	0.01	0.02	--	--	--
1.5	15	no relaxation	no relaxation	20	--	--
0.003	0.014	0.0014	0.0176	108	71	121
0.0012	**0.0098**	**0**	**0.0139**	**14**	**0**	**0**
0.038	0.028	0.0042	0.0253	140	98	167
0.0015	**0.0022**	**0**	**0.0169**	**16**	**0**	**0**
0.001	0.006	0.0002	0.0176	22	25	2
0.0008	0.005	0	0.0158	19	34	9.6

Rishwant/Water-Quality-A | Water quality analysis/ | Water quality analysis(tea

localhost:8888/notebooks/Water%20quality%20analysis/Water%20quality%20analysis(team).ipynb

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Run Code

In [1]:
```
pwd
```

Out[1]:
```
'C:\\Users\\Roshan\\Water quality analysis'
```

In [49]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import plotly.express as px
import matplotlib.pyplot as plt

#sklearn
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split

from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
```

In [3]:
```
df = pd.read_csv("C:\\Users\\Roshan\\Water quality analysis\\water_potability.csv")
df
```

Out[3]:

	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
...	...	...	...	...	...	...	...	...	...	...
3271	4.668102	193.681735	47580.991603	7.166639	359.948574	526.424171	13.894419	66.687695	4.435821	1
3272	7.808856	193.553212	17329.802160	8.061362	NaN	392.449580	19.903225	NaN	2.798243	1
3273	9.419510	175.762646	33155.578218	7.350233	NaN	432.044783	11.039070	69.845400	3.298875	1
3274	5.126763	230.603758	11983.869376	6.303357	NaN	402.883113	11.168946	77.488213	4.708658	1
3275	7.874671	195.102299	17404.177061	7.509306	NaN	327.459760	16.140368	78.698446	2.309149	1

3276 rows × 10 columns

Rishwant/Water-Quality-A | Water quality analysis/ | Water quality analysis(tea

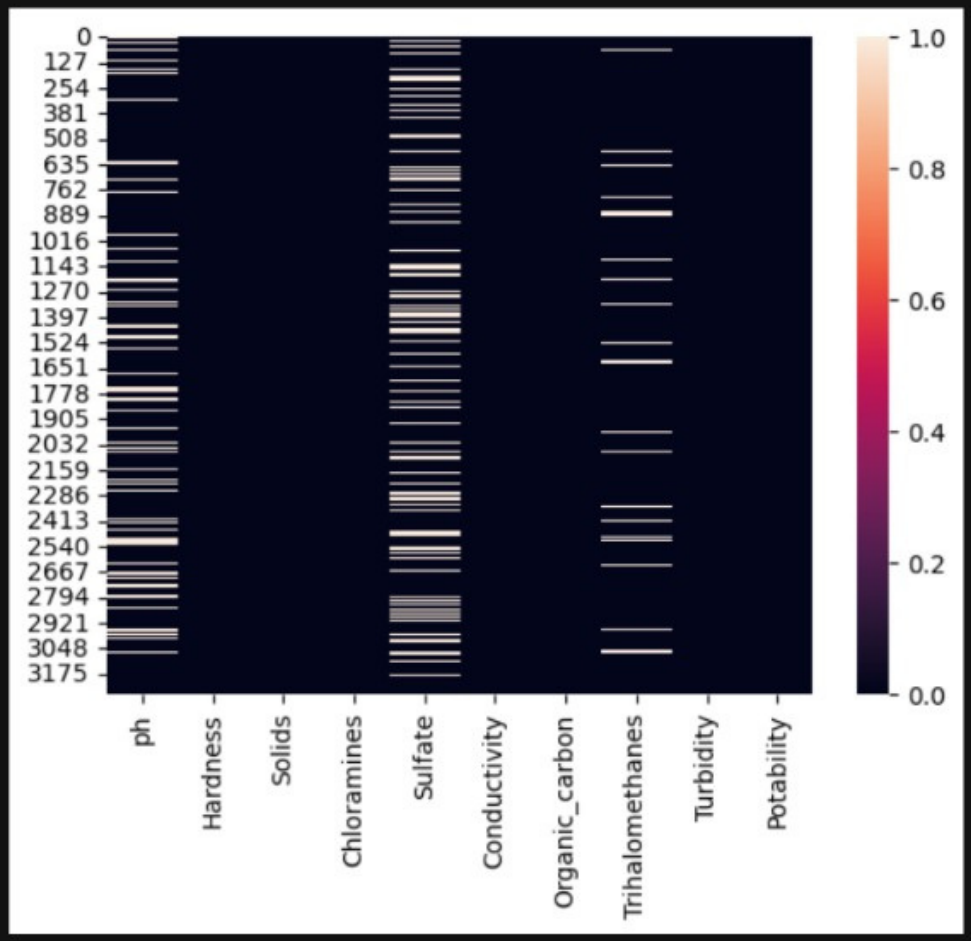
localhost:8888/notebooks/Water%20quality%20analysis/Water%20quality%20analysis(team).ipynb

File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 3 (ipykernel)

Run | Code

In [8]:
```
sns.heatmap(df.isnull())
```

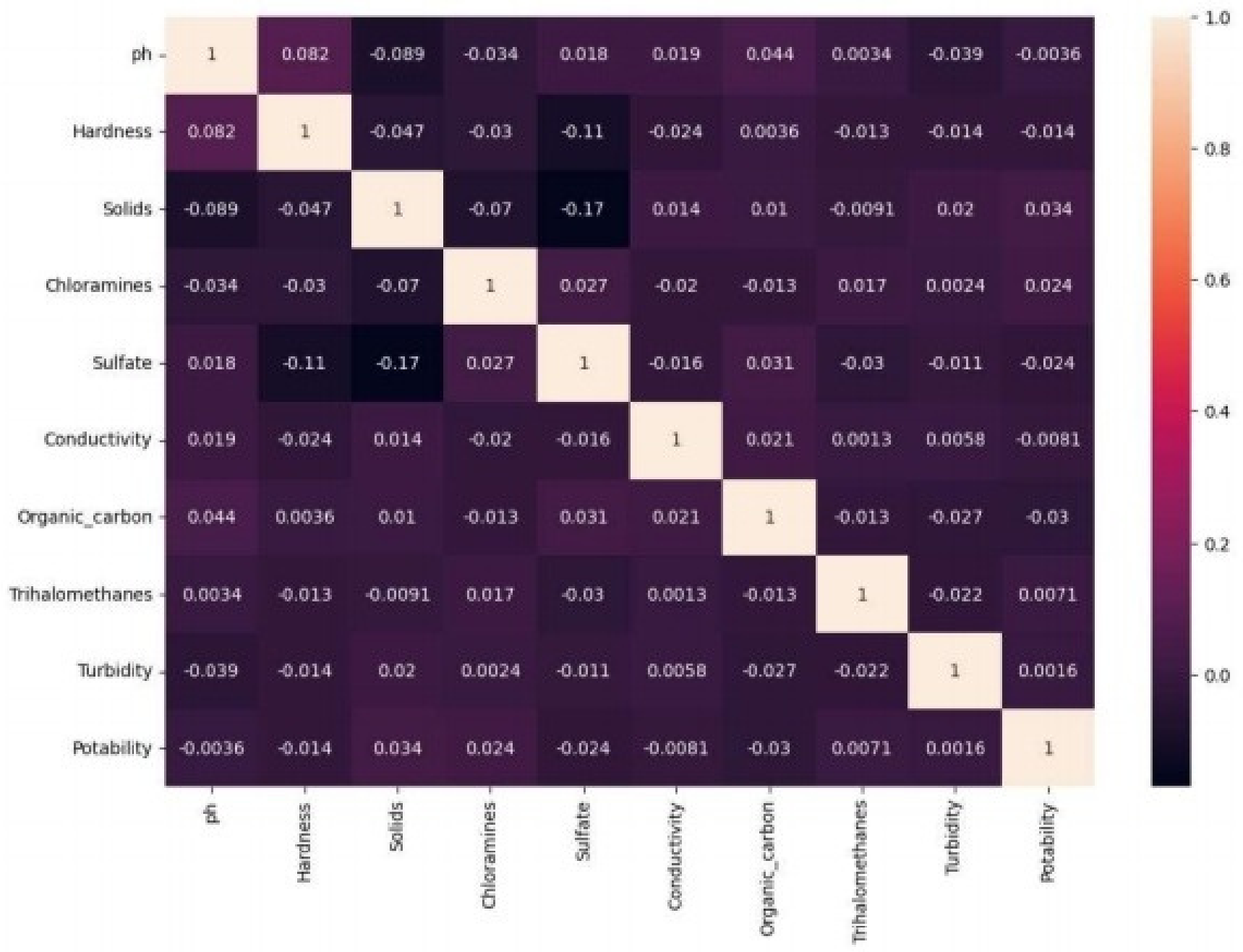
Out[8]: <Axes: >



In [9]:
```
plt.figure(figsize=(12,8))
sns.heatmap(df.isnull())
```

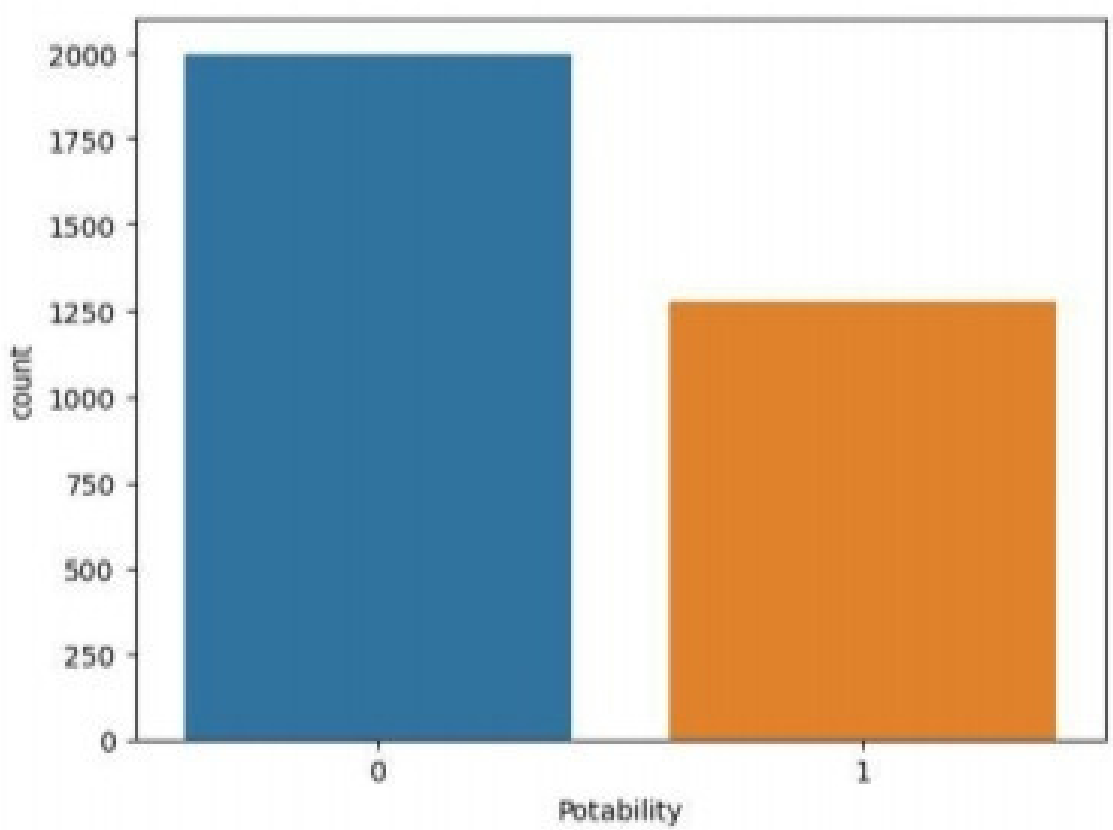
In [10]:
```
plt.figure(figsize=(12,8))
sns.heatmap(df.corr(),annot=True)
```

Out[10]: <Axes: >


ph 1 0.082 -0.089 -0.034 0.018 0.019 0.044 0.0034 -0.039 -0.0036
Hardness 0.082 1 -0.047 -0.03 -0.11 -0.024 0.0036 -0.013 -0.014 -0.014
Solids -0.089 -0.047 1 -0.07 -0.17 0.014 0.01 -0.0091 0.02 0.034
Chloramines -0.034 -0.03 -0.07 1 0.027 -0.02 -0.013 0.017 0.0024 0.024
Sulfate 0.018 -0.11 -0.17 0.027 1 -0.016 0.031 -0.03 -0.011 -0.024
Conductivity 0.019 -0.024 0.014 -0.02 -0.016 1 0.021 0.0013 0.0058 -0.0081
Organic_carbon 0.044 0.0036 0.01 -0.013 0.031 0.021 1 -0.013 -0.027 -0.03
Trihalomethanes 0.0034 -0.013 -0.0091 0.017 -0.03 0.0013 -0.013 1 -0.022 0.0071
Turbidity -0.039 -0.014 0.02 0.0024 -0.011 0.0058 -0.027 -0.022 1 0.0016
Potability -0.0036 -0.014 0.034 0.024 -0.024 -0.0081 -0.03 0.0071 0.0016 1
ph
Hardness
Solids
Chloramines
Sulfate
Conductivity
Organic_carbon
Trihalomethanes
Turbidity
Potability
1.0
0.8
0.6
0.4
0.2
0.0

In [11]:
```
sns.countplot(x="Potability",data=df)
```

Out[11]: <Axes: xlabel='Potability', ylabel='count'>



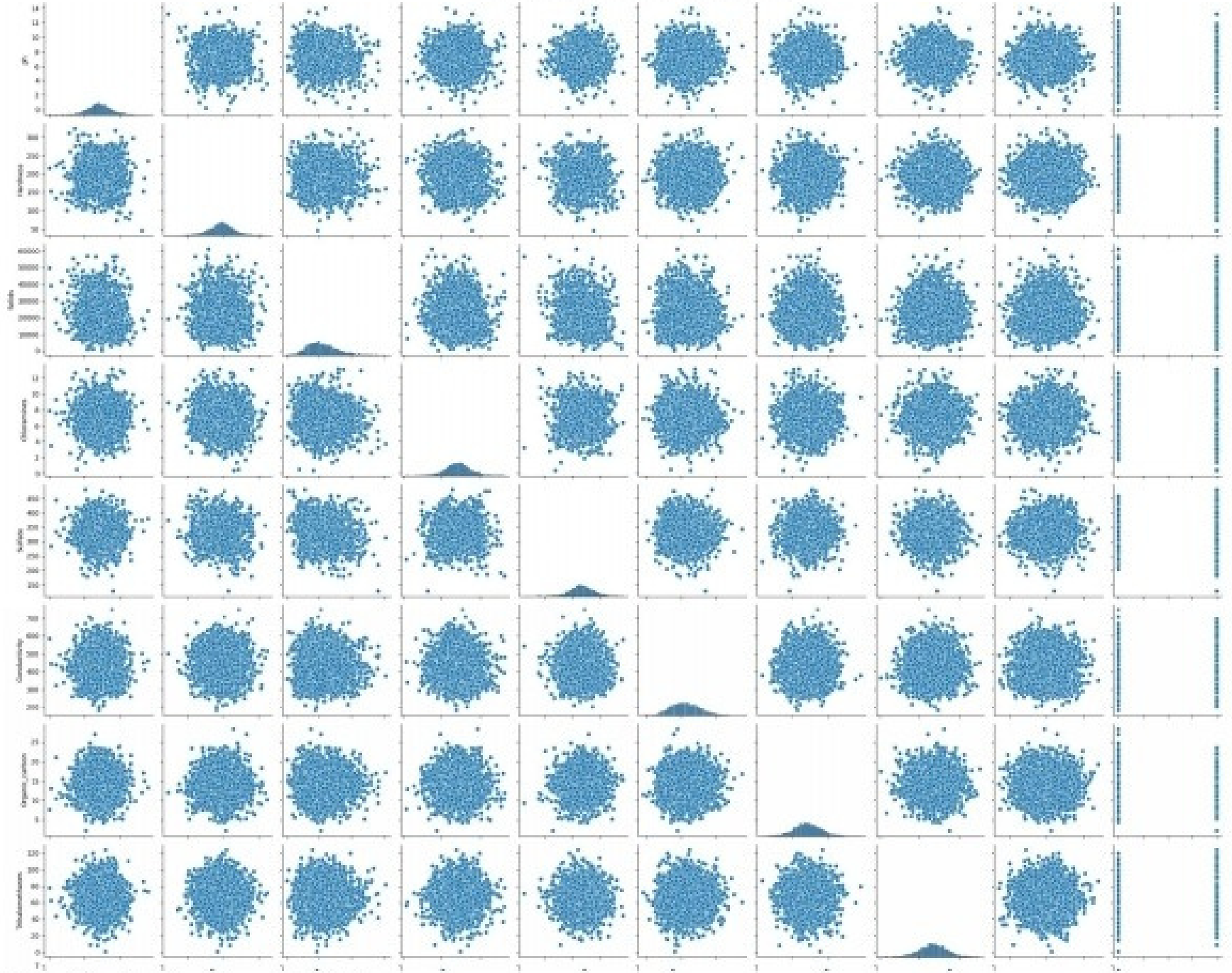
In [13]:
```
df["Potability"].value_counts()
```

Out[13]:
```
Potability
0    1998
1    1278
Name: count, dtype: int64
```

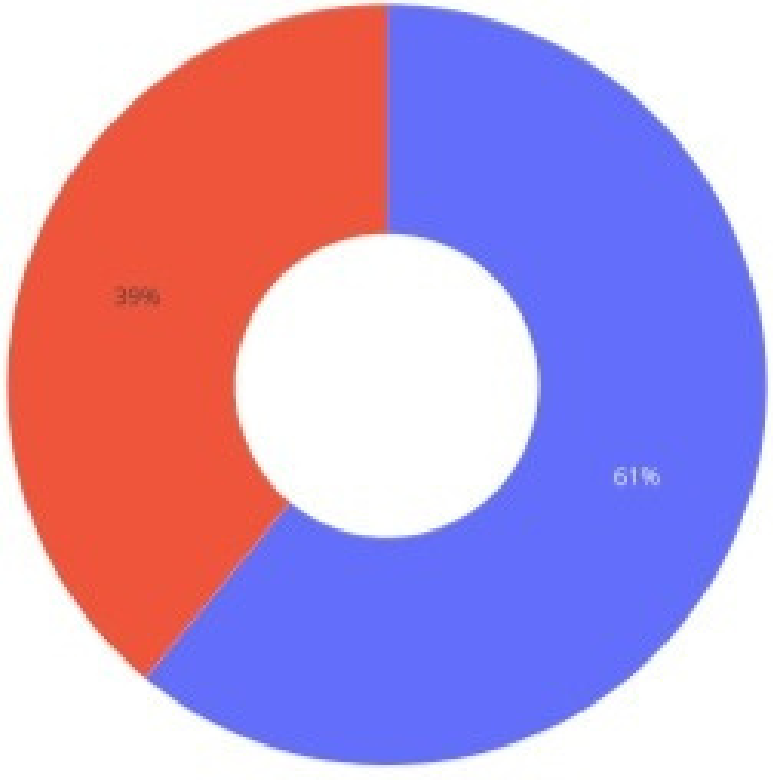
In [15]:
```
sns.pairplot(df)
```

```
C:\Users\Roshan\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to ti
ght
  self._figure.tight_layout(*args, **kwargs)
```

Out[15]:
```
<seaborn.axisgrid.PairGrid at 0x1bcea59a2d0>
```


ph
Hardness
Solids
Chloramines
Sulfate
Conductivity
Organic_carbon
Trihalomethanes

In [22]:
```
fig = px.pie(df,names = "Potability", hole = 0.4,template ="plotly_dark")
fig.show()
```



Continuous Monitoring and Improvement

Monitor the model's performance and gather new data to continually improve the model's accuracy and relevance.