MODEL BUILDING

Date	15 November 2022
Team ID	PNT2022TMID49672
Project Name	Efficient Water Quality Analysis and Prediction Using Machine Learning
Maximum Marks	8 Marks

Model Building

DT

```python
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score,confusion_matrix,precision_score
dt=DecisionTreeClassifier(criterion= 'gini', min_samples_split= 10, splitter= 'best')
dt.fit(X_train,Y_train)
```
[228]   Python

···   DecisionTreeClassifier(min_samples_split=10)

```python
prediction=dt.predict(X_test)
accuracy_dt=accuracy_score(Y_test,prediction)*100
accuracy_dt
```
[229]   Python

···   58.84146341463414

C: > Users > ELCOT > Desktop > Water-Quality- > Water Quality Prediction .ipynb > M↓ Exploratory Data Analysis > df.isnull().sum()

+ Code   + Markdown   ···

```python
print("Accuracy on training set: {:.3f}".format(dt.score(X_train, Y_train)))
print("Accuracy on test set: {:.3f}".format(dt.score(X_test, Y_test)))
```
[230]

```
Accuracy on training set: 0.923
Accuracy on test set: 0.588
```

```python
accuracy_score(prediction,Y_test)
```
[232]

```
0.5884146341463414
```

```python
print("Feature importances:\n{}".format(dt.feature_importances_))
```
[234]

```
Feature importances:
[0.14979308 0.1306299  0.11639812 0.10910309 0.11691893 0.09302001
 0.10460544 0.10201786 0.07751357]
```

Restricted Mode is intended for safe code browsing. Trust this window to enable all features.   Manage   Learn More

Get Started   Water Quality Prediction .ipynb ✕

C: > Users > ELCOT > Desktop > Water-Quality- > ■ Water Quality Prediction .ipynb > ᴹ↓ Exploratory Data Analysis > ◆ df.isnull().sum()

➕ Code ➕ Markdown ⋯   🖳 Select Kernel

```
···     0.5884146341463414
```

```python
      print("Feature importances:\n{}".format(dt.feature_importances_))
```
[234]   Python

```
···   Feature importances:
      [0.14979308 0.1306299  0.11639812 0.10910309 0.11691893 0.09302001
       0.10460544 0.10201786 0.07751357]
```

```python
      confusion_matrix(prediction,Y_test)
```
[233]   Python

```
···   array([[272, 140],
             [130, 114]], dtype=int64)
```

Prediction on only one set of data

```python
      X_DT=dt.predict([[5.735724, 158.318741,25363.016594,7.728601,377.543291,568.304671,13.626624,75.952337,4.732954]])
```
[240]   Python

🛡 Restricted Mode   ⊗ 0 ⚠ 0   Cell 5 of 80