Project Development Phase
Model Performance Test

Date	10 NOvember 2022
Team ID	PNT2022TMID54402
Project Name	Project – 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.

S.No.	Parameter	Values	Screenshot
1.	Metrics	**Regression Model:** MSE - 4.076 ,RMSE - 2.018 R2 score – Testing data: 0.9743 Training data: 0.995 MAE: 0.735	In [48]: metrics.r2_score(y_test, y_pred) #R2 score on testing data Out[48]: 0.9743920713924711 In [49]: metrics.r2_score(y_train,y_train_pred) #R2 score on training data Out[49]: 0.9954846162499059 In [50]: metrics.mean_squared_error(y_test,y_pred) #Mean squared error Out[50]: 4.0760741930827065 In [51]: np.sqrt(metrics.mean_squared_error(y_test,y_pred)) #Root Mean Squared error Out[51]: 2.018928971777538 In [52]: metrics.mean_absolute_error(y_test, y_pred) Out[52]: 0.7353172932330914

2.	Tune the Model	Hyperparameter Tuning - -	

In [54]:
```
from sklearn.model_selection import GridSearchCV
```

In [98]:
```
parameters = { 'n_estimators':[10,50,100],
               'max_depth':[100,200,2000],
               'max_leaf_nodes':[200,300,100,500]}
```

In [99]:
```
clf  = GridSearchCV(RandomForestRegressor(),param_grid=parameters,verbose =2)
```

In [100]:
```
clf.fit(x_train,y_train)
```

```
[CV] END max_depth=2000, max_leaf_nodes=500, n_estimators=100; total time=   0.7s

C:\Users\crpri\AppData\Roaming\Python\Python39\site-packages\sklearn\model_selection\_validation.py:686: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
  estimator.fit(X_train, y_train, **fit_params)
```

In [101]:
```
clf.best_score_
```

Out[101]:
```
0.9765413404788307
```

In [102]:
```
clf.best_params_
```

Out[102]:
```
{'max_depth': 2000, 'max_leaf_nodes': 200, 'n_estimators': 50}
```

In [104]:
```
model = RandomForestRegressor(max_depth= 2000, max_leaf_nodes= 200, n_estimators= 50)
```

In [105]:
```
model.fit(x_train,y_train)
```

```
C:\Users\crpri\AppData\Local\Temp\ipykernel_23784\1002171229.py:1: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
  model.fit(x_train,y_train)
```

Out[105]:
```
RandomForestRegressor
RandomForestRegressor(max_depth=2000, max_leaf_nodes=200, n_estimators=50)
```