

**Project Development Phase**  
**Model Performance Test**

Date	17 November 2022
Team ID	PNT2022TMID45622
Project Name	EFFICIENT WATER QUALITY ANALYSIS AND 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	<b>Regression Model:</b> MAE -0.9425563909774494 MSE: 5.63627572932331 RMSE: 2.374084187497004 <b>R2 score :</b> Training: 0.9948711603144075  Testing:0.9692766700278257	Screenshot 1
2.	Tune the Model	Hyper parameter Tuning - NIL Validation Method – Split Sample/ Data Validation.	Screenshot 2

## Screenshot 1:

The screenshot displays a Jupyter Notebook interface with four sections of code and their corresponding outputs. The browser window at the top shows the GitHub repository URL: `github.com/IBM-EPBL/IBM-Project-19207-1659694215/blob/main/Project%20Development%20Phase/SPRINT%203/Model%20Building%20and%20Evaluation%20...`.

### R2 Score on Testing Data

```
In [68]: print(metrics.r2_score(y_test, y_pred))
```

0.9692766700278257

### R2 Score on Training Data

```
In [69]: print(metrics.r2_score(y_train, y_train_pred))
```

0.9948711603144075

### Overall Metrics Value Of This Random Forest Regression Model

```
In [70]: print('MAE:', metrics.mean_absolute_error(y_test, y_pred))
print('MSE:', metrics.mean_squared_error(y_test, y_pred))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
```

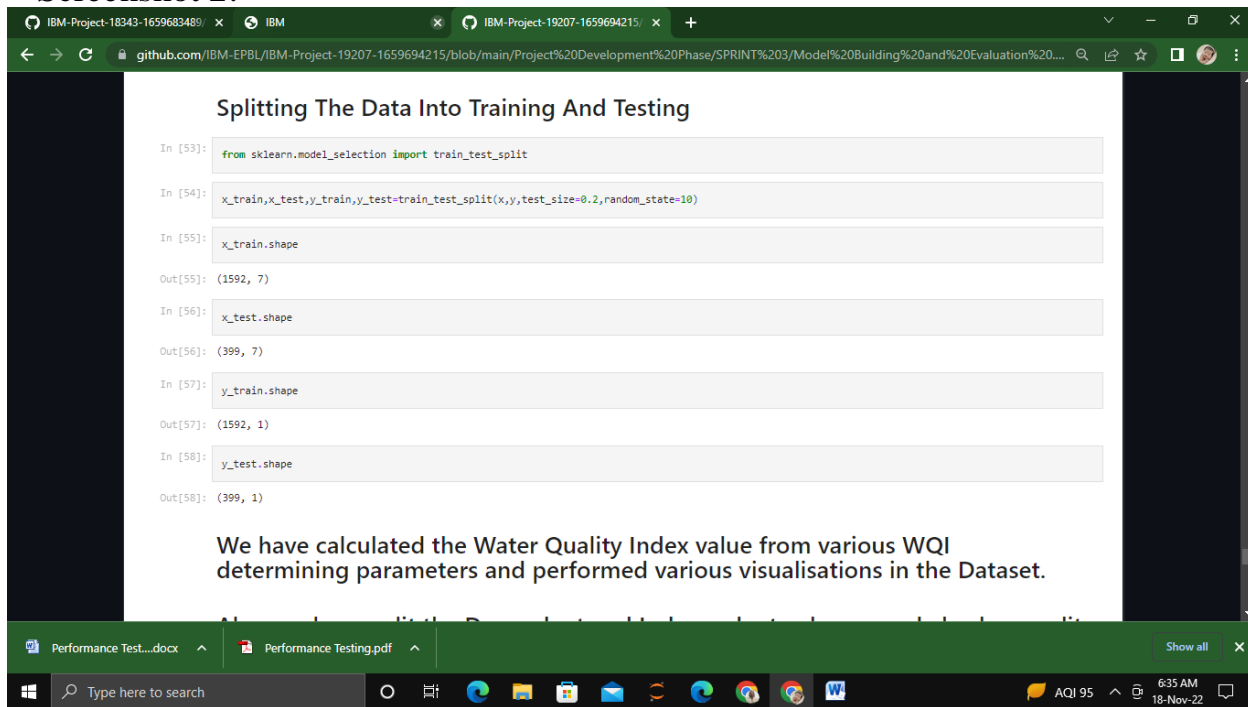
MAE: 0.9425563909774494  
MSE: 5.63627572932331  
RMSE: 2.374084187497004

### Saving The Model

```
In [72]: import joblib
joblib.dump(regressor, 'wq1.pkl')
```

The bottom of the image shows the Windows taskbar with the search bar, task view button, and several application icons. The system tray on the right indicates an AQI of 95 and the date/time as 6:35 AM on 18-Nov-22.

## Screenshot 2:



The screenshot displays a Jupyter Notebook interface within a web browser. The browser's address bar shows a GitHub repository URL. The notebook's title is "Splitting The Data Into Training And Testing". The code cells show the following:

```
In [53]: from sklearn.model_selection import train_test_split
```

```
In [54]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=10)
```

```
In [55]: x_train.shape
```

```
Out[55]: (1592, 7)
```

```
In [56]: x_test.shape
```

```
Out[56]: (399, 7)
```

```
In [57]: y_train.shape
```

```
Out[57]: (1592, 1)
```

```
In [58]: y_test.shape
```

```
Out[58]: (399, 1)
```

Below the code cells, a text cell contains the following paragraph:

We have calculated the Water Quality Index value from various WQI determining parameters and performed various visualisations in the Dataset.

The bottom of the image shows a Windows taskbar with a search bar, several application icons, and a system tray displaying "AQI 95" and the date/time "6:35 AM 18-Nov-22".