

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

|               |   |
|---------------|---|
| Date          | 17 November 2022  |
| Team ID       | PNT2022TMID45606  |
| 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><br>MAE -0.9425563909774494<br>MSE: 5.63627572932331<br>RMSE: 2.374084187497004<br><b>R2 score :</b><br>Training: 0.9948711603144075<br><br>Testing:0.9692766700278257 | Screenshot 1 |
| 2.    | Tune the Model | Hyper parameter Tuning - NIL<br>Validation Method –<br>Split Sample/ Data Validation.  | Screenshot 2 |

## Screenshot 1:

The screenshot shows a Jupyter Notebook interface with the following content:

### 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

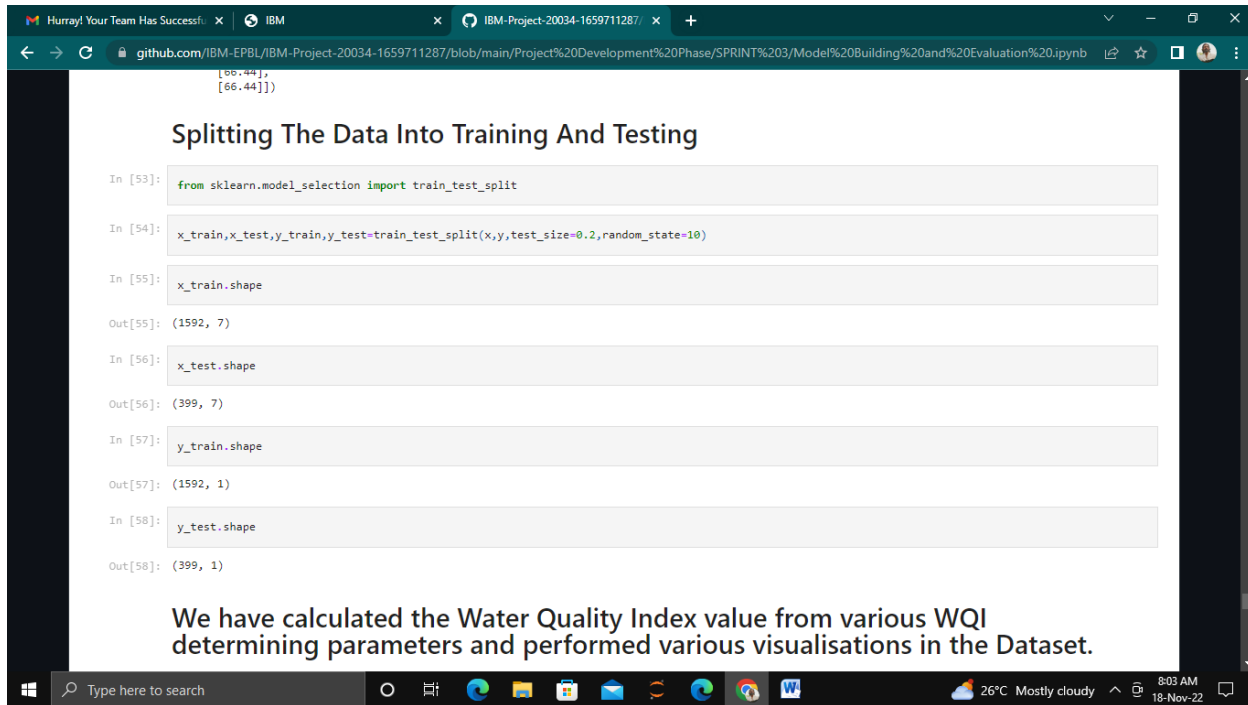
### Saving The Model

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

Out[72]: ['wqi.pkl']

The notebook is displayed in a web browser window with the URL: [github.com/IBM-EPBL/IBM-Project-20034-1659711287/blob/main/Project%20Development%20Phase/SPRINT%203/Model%20Building%20and%20Evaluation%20.ipynb](https://github.com/IBM-EPBL/IBM-Project-20034-1659711287/blob/main/Project%20Development%20Phase/SPRINT%203/Model%20Building%20and%20Evaluation%20.ipynb). The Windows taskbar at the bottom shows the time as 8:03 AM on 18-Nov-22, with a temperature of 26°C and weather condition 'Mostly cloudy'.

## Screenshot 2:



The screenshot shows a Jupyter Notebook interface within a web browser. The browser's address bar displays the GitHub repository URL: `github.com/IBM-EPBL/IBM-Project-20034-1659711287/blob/main/Project%20Development%20Phase/SPRINT%203/Model%20Building%20and%20Evaluation%20.ipynb`. The notebook content includes a title, a code cell with imports and splitting logic, and several output cells showing the shapes of the resulting training and testing datasets.

### Splitting The Data Into Training And Testing

```
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)
```

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

The Windows taskbar at the bottom shows the search bar, taskbar icons for various applications, and system information including the date and time (8:03 AM, 18-Nov-22) and weather (26°C, Mostly cloudy).