

Project Development Phase Model Performance Test

Date	10November2022
Team ID	PNT2022TMID08775
Project Name	Project– Efficient Water Quality analysis& prediction
Maximum Marks	10Marks

Model Performance Testing:

Project team shall fill the following information in model performance testing template.

S.No.	Parameter	Values	Screenshot
1.	Metrics	RegressionModel: MAE - ,MSE-,RMSE-, R2score- ClassificationModel: Confusion Matrix - , Accuray Score-& ClassificationReport-	Below
2.	Tune theModel	Hyperparameter Tuning - ValidationMethod -	Below

1.Metrics-LinearRegression

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

Splitting data into Train & Test

```

In [43]: from sklearn import neighbors, datasets
data = data.reset_index(level = 0, inplace = False)

In [44]: from sklearn import linear_model

In [45]: from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(x,y,test_size = 0.2,random_state = 10)

In [46]: X_train
Out[46]:
   year
5  2011
6  2012
3  2009
1  2007
0  2006
7  2013
4  2010

In [47]: y_train
Out[47]:
5    75.647013
6    70.060941

```

The notebook is running on a local host (localhost:8888) and is titled 'analysis'. The code is executed in a Python 3 (ipykernel) environment. The output of the last cell shows the training data (X_train) as a series of years from 2006 to 2013.

NT- Efficient Water Quality x wqi.pkl (editing) x analysis - Jupyter Notebook x IBM x IBM x IBM-Project-11589-165933 x +

localhost:8888/notebooks/NT-%20Efficient%20Water%20Quality%20Analysis%20%26%20Prediction/analysis.ipynb#

Gmail YouTube Maps Loading... [Solved] WARNING... IBM IBM Cloud What is Exploratory... Water Quality Analy...

jupyter analysis Last Checkpoint: an hour ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

In [47]: y_train

Out[47]:

5	75.647013
6	78.969041
3	74.085193
1	72.549000
0	71.308824
7	75.009425
4	74.648723

Name: wqi, dtype: float64

In [48]: x_test

Out[48]:

	year
8	2014
2	2008

In [49]: y_test

Out[49]:

8	76.833852
2	72.570943

Name: wqi, dtype: float64

In [50]: reg = linear_model.LinearRegression()
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = 4)

In [51]: reg.fit(x_train, y_train)

Out[51]: LinearRegression()

analysis (3).ipynb Show all

23°C Partly cloudy Search ENG IN 9:01 PM 11/18/2022

NT- Efficient Water Quality x wqi.pkl (editing) x analysis - Jupyter Notebook x IBM x IBM x IBM-Project-11589-165933 x +

localhost:8888/notebooks/NT-%20Efficient%20Water%20Quality%20Analysis%20%26%20Prediction/analysis.ipynb#

Gmail YouTube Maps Loading... [Solved] WARNING... IBM IBM Cloud What is Exploratory... Water Quality Analy...

jupyter analysis Last Checkpoint: an hour ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

In [51]: reg.fit(x_train, y_train)

Out[51]: LinearRegression()

In [52]: from sklearn import neighbors, datasets
data = data.reset_index(level=0, inplace=False)
data

Out[52]:

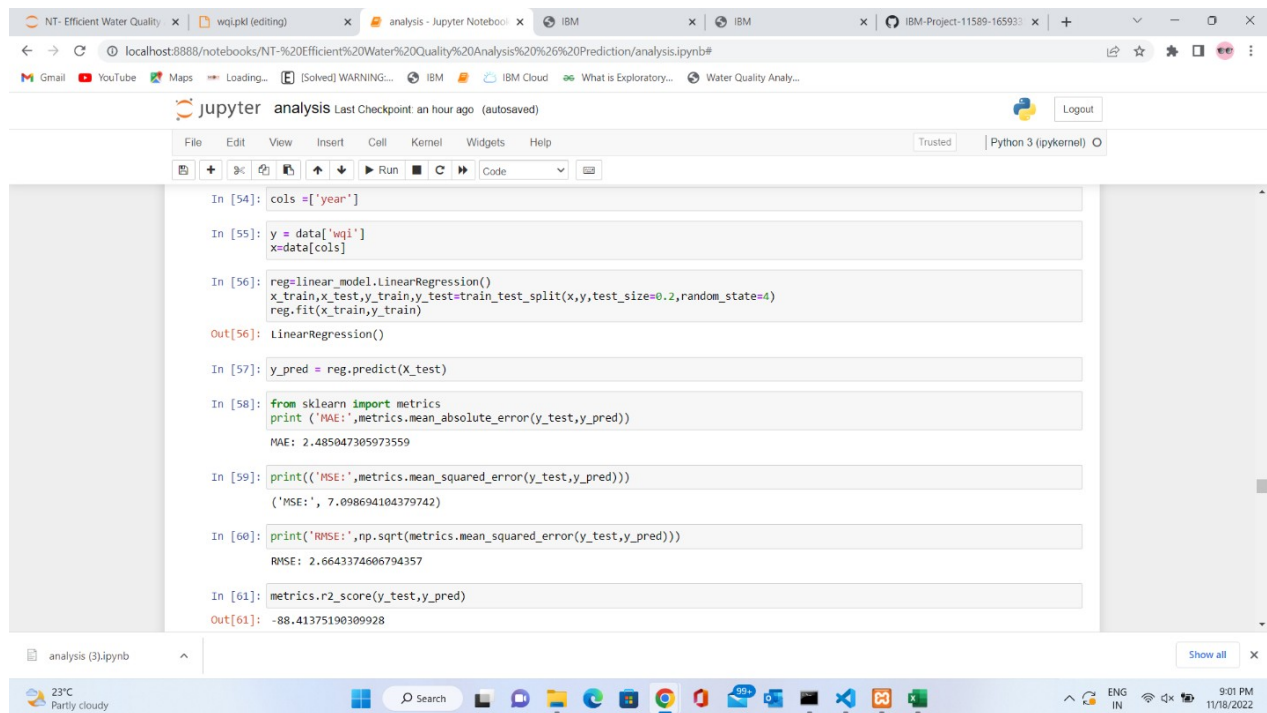
	index	year	wqi
0	0	2006	71.308824
1	1	2007	72.549000
2	2	2008	72.570943
3	3	2009	74.085193
4	4	2010	74.648723
5	5	2011	75.647013
6	6	2012	78.969041
7	7	2013	75.009425
8	8	2014	76.833852

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

In [54]: cols = ['year']

analysis (3).ipynb Show all

23°C Partly cloudy Search ENG IN 9:01 PM 11/18/2022



```
In [54]: cols = ['year']

In [55]: y = data['wqi']
x=data[cols]

In [56]: reg=linear_model.LinearRegression()
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=4)
reg.fit(x_train,y_train)

Out[56]: LinearRegression()

In [57]: y_pred = reg.predict(X_test)

In [58]: from sklearn import metrics
print ('MAE:',metrics.mean_absolute_error(y_test,y_pred))
MAE: 2.485047305973559

In [59]: print(('MSE:',metrics.mean_squared_error(y_test,y_pred)))
('MSE:', 7.098694104379742)

In [60]: print('RMSE:',np.sqrt(metrics.mean_squared_error(y_test,y_pred)))
RMSE: 2.6643374606794357

In [61]: metrics.r2_score(y_test,y_pred)
Out[61]: -88.41375190309928
```

2. Tunethe Model

HyperparameterTuning:

- Thenumberoffeaturesisimportant andshouldbetunedinlinearregression
- Initially all parameters in the dataset are taken as independent values to arrive at thedependent decisionof ExploratoryAnalysisofWaterQualityPrediction
- But the result was not accurate so used only 8 more correlated values as independent valuestoarriveatthedependentdecisionofExploratoryAnalysis ofWaterQuality Prediction.

ValidationMethod:

- Itinvolvespartitioningthetrainingdatasetintosubsets,whereonesubsetis heldouttotesttheperformanceofthemodel.
- This data set is called the validation data set. Cross validation is to use different modelsandidentifythebest:LinearRegressionModelperformancevalues:

NT-Efficient Water Quality x wqi.pkl (editing) x analysis - Jupyter Notebook x IBM x IBM x IBM-Project-11589-16593 x +

localhost:8888/notebooks/NT-%20Efficient%20Water%20Quality%20Analysis%20%26%20Prediction/analysis.ipynb#

Gmail YouTube Maps Loading... [Solved] WARNING... IBM IBM Cloud What is Exploratory... Water Quality Analy...

jupyter analysis Last checkpoint an hour ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (pykernel)

In [54]: cols = ['year']

In [55]: y = data['wqi']
x=data[cols]

In [56]: reg=linear_model.LinearRegression()
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=4)
reg.fit(x_train,y_train)

Out[56]: LinearRegression()

In [57]: y_pred = reg.predict(X_test)

In [58]: from sklearn import metrics
print('MAE:',metrics.mean_absolute_error(y_test,y_pred))
MAE: 2.485047305973559

In [59]: print(('MSE:',metrics.mean_squared_error(y_test,y_pred)))
('MSE:', 7.098694104379742)

In [60]: print('RMSE:',np.sqrt(metrics.mean_squared_error(y_test,y_pred)))
RMSE: 2.6643374606794357

In [61]: metrics.r2_score(y_test,y_pred)

Out[61]: -88.41375190309928

analysis (3).ipynb Show all

23°C Partly cloudy Search ENG IN 9:01 PM 11/18/2022