File Edit Selection View Go Run Terminal Help

Search

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 > Exploratory Data Analysis > df.isnull().sum()

+ Code + Markdown ···

Select Kernel

Hyperparameter Tuning / Model Optimization

DT HPT

```
dt.get_params().keys()
```

[178] Python

```
dict_keys(['ccp_alpha', 'class_weight', 'criterion', 'max_depth', 'max_features', 'max_leaf_nodes', 'min_impurity_decrease', 'min_impurity_split',
'min_samples_leaf', 'min_samples_split', 'min_weight_fraction_leaf', 'presort', 'random_state', 'splitter'])
```

```
#example of grid searching key hyperparametres for logistic regression
from sklearn.model_selection import RepeatedStratifiedKFold
from sklearn.model_selection import GridSearchCV

# define models and parameters
model = DecisionTreeClassifier()
criterion = ["gini", "entropy"]
splitter = ["best", "random"]
min_samples_split = [2,4,6,8,10]

# define grid search
grid = dict(splitter=splitter, criterion=criterion, min_samples_split=min_samples_split)
```

Restricted Mode ⊗ 0 ⚠ 0 Cell 5 of 80

Type here to search 29°C ENG 16:29 19-11-2022

File Edit Selection View Go Run Terminal Help

Search

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 > Exploratory Data Analysis > df.isnull().sum()

+ Code + Markdown ··· Select Kernel

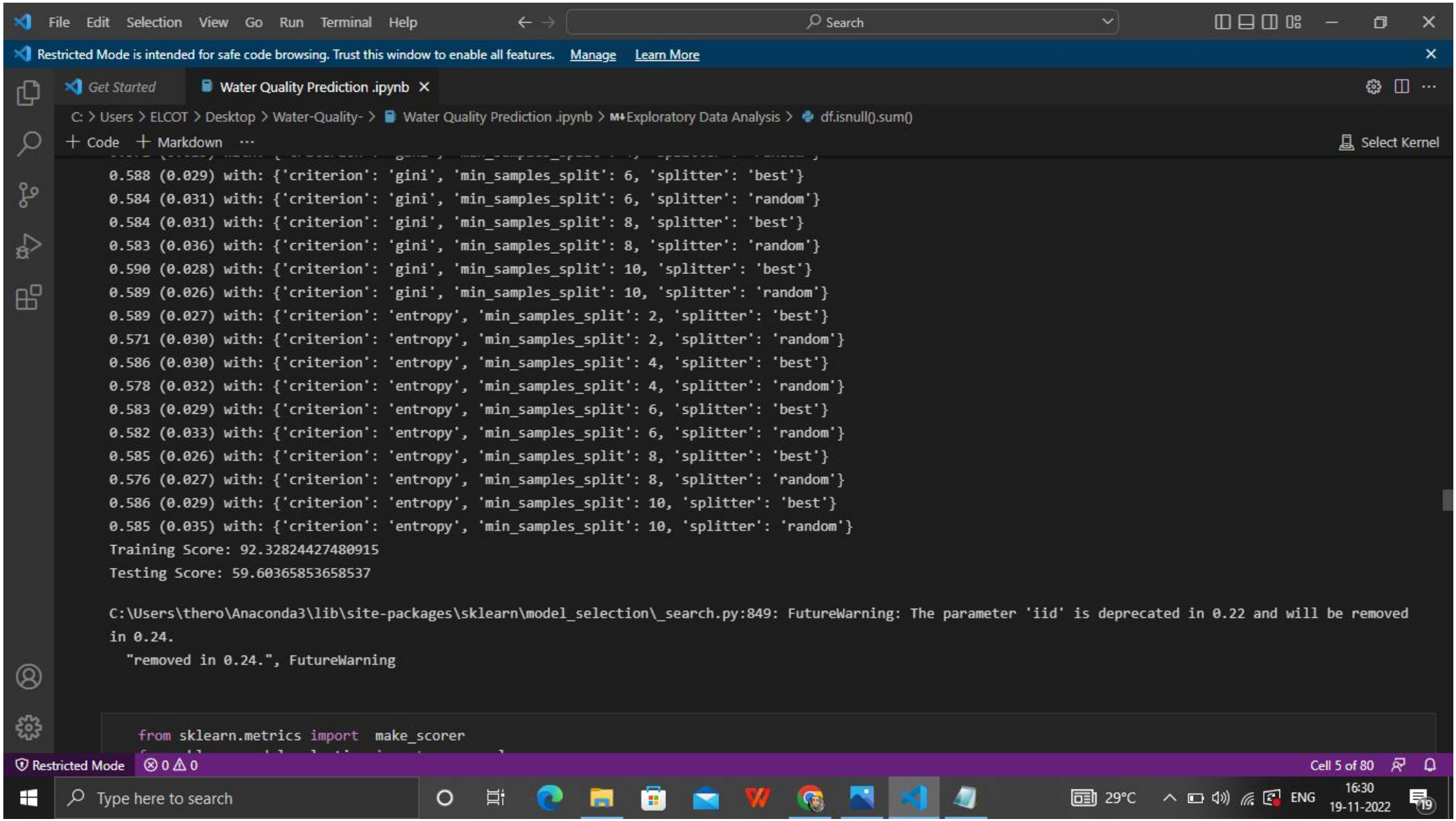
```
# define grid search
grid = dict(splitter=splitter, criterion=criterion, min_samples_split=min_samples_split)
cv = RepeatedStratifiedKFold(n_splits=10, n_repeats=3, random_state=1)
grid_search_dt = GridSearchCV(estimator=model, param_grid=grid, n_jobs=-1, cv=cv,
                              scoring='accuracy',error_score=0, iid=True)
grid_search_dt.fit(X_train, Y_train)

# summarize results
print(f"Best: {grid_search_dt.best_score_:.3f} using {grid_search_dt.best_params_}")
means = grid_search_dt.cv_results_['mean_test_score']
stds = grid_search_dt.cv_results_['std_test_score']
params = grid_search_dt.cv_results_['params']

for mean, stdev, param in zip(means, stds, params):
    print(f"{mean:.3f} ({stdev:.3f}) with: {param}")

print("Training Score:",grid_search_dt.score(X_train, Y_train)*100)
print("Testing Score:", grid_search_dt.score(X_test, Y_test)*100)
```

[216] Python

```
Best: 0.590 using {'criterion': 'gini', 'min_samples_split': 10, 'splitter': 'best'}
0.584 (0.029) with: {'criterion': 'gini', 'min_samples_split': 2, 'splitter': 'best'}
0.569 (0.030) with: {'criterion': 'gini', 'min_samples_split': 2, 'splitter': 'random'}
0.584 (0.028) with: {'criterion': 'gini', 'min_samples_split': 4, 'splitter': 'best'}
0.571 (0.025) with: {'criterion': 'gini', 'min_samples_split': 4, 'splitter': 'random'}
0.588 (0.029) with: {'criterion': 'gini', 'min_samples_split': 6, 'splitter': 'best'}
0.584 (0.031) with: {'criterion': 'gini', 'min_samples_split': 6, 'splitter': 'random'}
```

Restricted Mode ⊗ 0 ⚠ 0 Cell 5 of 80

Type here to search 29°C ENG 16:29 19-11-2022


File Edit Selection View Go Run Terminal Help
Search
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 > Exploratory Data Analysis > df.isnull().sum()
+ Code + Markdown ···
Select Kernel
0.588 (0.029) with: {'criterion': 'gini', 'min_samples_split': 6, 'splitter': 'best'}
0.584 (0.031) with: {'criterion': 'gini', 'min_samples_split': 6, 'splitter': 'random'}
0.584 (0.031) with: {'criterion': 'gini', 'min_samples_split': 8, 'splitter': 'best'}
0.583 (0.036) with: {'criterion': 'gini', 'min_samples_split': 8, 'splitter': 'random'}
0.590 (0.028) with: {'criterion': 'gini', 'min_samples_split': 10, 'splitter': 'best'}
0.589 (0.026) with: {'criterion': 'gini', 'min_samples_split': 10, 'splitter': 'random'}
0.589 (0.027) with: {'criterion': 'entropy', 'min_samples_split': 2, 'splitter': 'best'}
0.571 (0.030) with: {'criterion': 'entropy', 'min_samples_split': 2, 'splitter': 'random'}
0.586 (0.030) with: {'criterion': 'entropy', 'min_samples_split': 4, 'splitter': 'best'}
0.578 (0.032) with: {'criterion': 'entropy', 'min_samples_split': 4, 'splitter': 'random'}
0.583 (0.029) with: {'criterion': 'entropy', 'min_samples_split': 6, 'splitter': 'best'}
0.582 (0.033) with: {'criterion': 'entropy', 'min_samples_split': 6, 'splitter': 'random'}
0.585 (0.026) with: {'criterion': 'entropy', 'min_samples_split': 8, 'splitter': 'best'}
0.576 (0.027) with: {'criterion': 'entropy', 'min_samples_split': 8, 'splitter': 'random'}
0.586 (0.029) with: {'criterion': 'entropy', 'min_samples_split': 10, 'splitter': 'best'}
0.585 (0.035) with: {'criterion': 'entropy', 'min_samples_split': 10, 'splitter': 'random'}
Training Score: 92.32824427480915
Testing Score: 59.60365853658537
C:\Users\thero\Anaconda3\lib\site-packages\sklearn\model_selection_search.py:849: FutureWarning: The parameter 'iid' is deprecated in 0.22 and will be removed in 0.24.
"removed in 0.24.", FutureWarning
from sklearn.metrics import make_scorer
Restricted Mode
Cell 5 of 80
Type here to search
29°C
ENG
16:30
19-11-2022

File Edit Selection View Go Run Terminal Help

Search

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 > Exploratory Data Analysis > df.isnull().sum()

+ Code + Markdown ···

Select Kernel

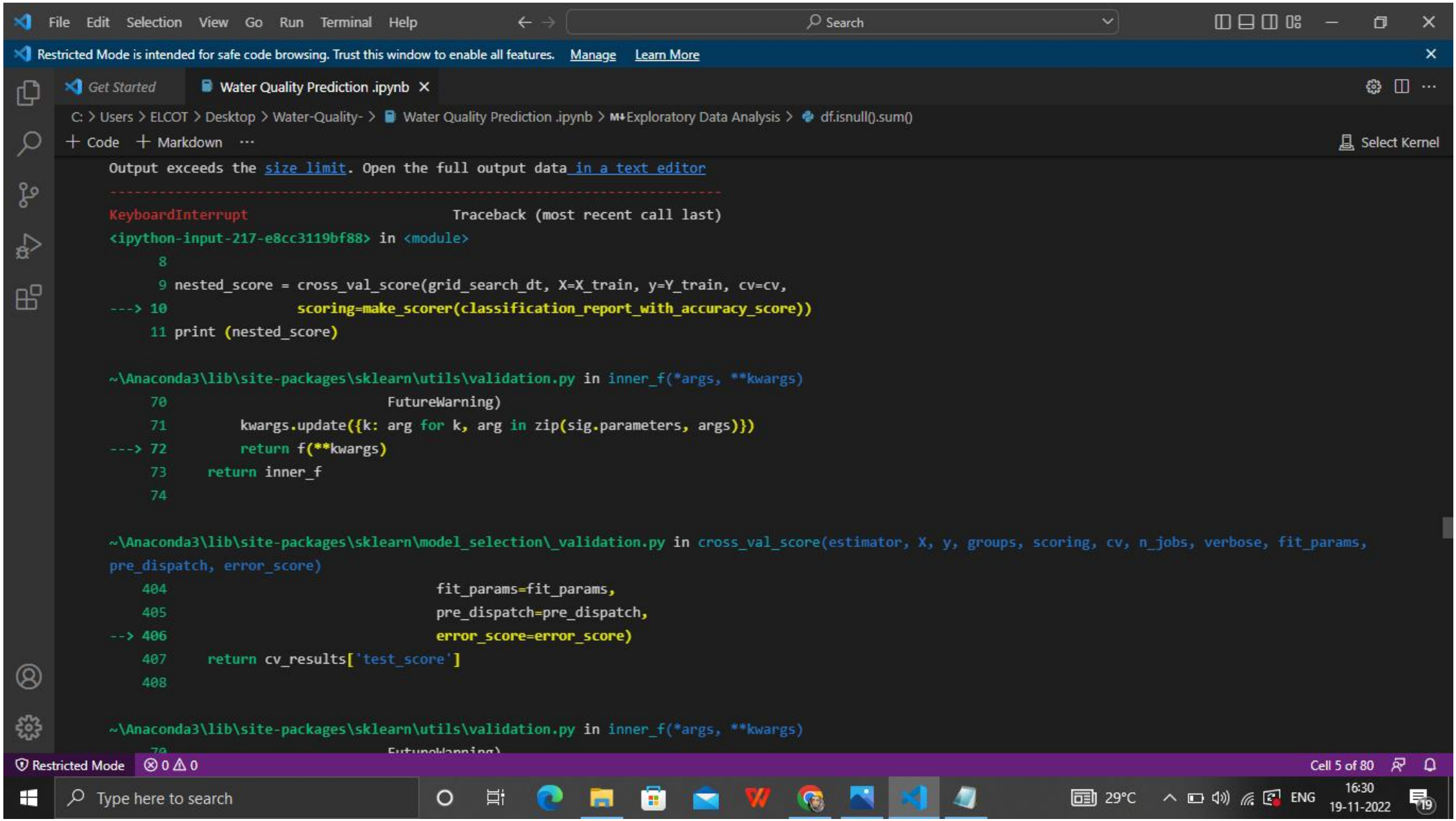
```
from sklearn.metrics import  make_scorer
from sklearn.model_selection import cross_val_score

def classification_report_with_accuracy_score(Y_test, y_pred2):
    print (classification_report(Y_test, y_pred2)) # print classification report
    return accuracy_score(Y_test, y_pred2) # return accuracy score


nested_score = cross_val_score(grid_search_dt, X=X_train, y=Y_train, cv=cv,
                               scoring=make_scorer(classification_report_with_accuracy_score))
print (nested_score)
```

[217] Python

```
C:\Users\thero\Anaconda3\lib\site-packages\sklearn\model_selection\_search.py:849: FutureWarning: The parameter 'iid' is deprecated in 0.22 and will be removed in 0.24.
  "removed in 0.24.", FutureWarning

              precision    recall  f1-score   support

           0       0.62      0.68      0.65       159
           1       0.42      0.36      0.39       103

    accuracy                           0.55       262
   macro avg       0.52      0.52      0.52       262
weighted avg       0.54      0.55      0.55       262
```

Restricted Mode ⊗ 0 ⚠ 0 Cell 5 of 80

Type here to search 29°C ENG 16:30 19-11-2022


File Edit Selection View Go Run Terminal Help
Search
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 > Exploratory Data Analysis > df.isnull().sum()
+ Code + Markdown ···
Select Kernel
Output exceeds the size limit. Open the full output data in a text editor
KeyboardInterrupt Traceback (most recent call last)
<ipython-input-217-e8cc3119bf88> in <module>
8
9 nested_score = cross_val_score(grid_search_dt, X=X_train, y=Y_train, cv=cv,
---> 10 scoring=make_scorer(classification_report_with_accuracy_score))
11 print (nested_score)
~\Anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*args, **kwargs)
70 FutureWarning)
71 kwargs.update({k: arg for k, arg in zip(sig.parameters, args)})
---> 72 return f(**kwargs)
73 return inner_f
74
~\Anaconda3\lib\site-packages\sklearn\model_selection_validation.py in cross_val_score(estimator, X, y, groups, scoring, cv, n_jobs, verbose, fit_params, pre_dispatch, error_score)
404 fit_params=fit_params,
405 pre_dispatch=pre_dispatch,
--> 406 error_score=error_score)
407 return cv_results['test_score']
408
~\Anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*args, **kwargs)
Restricted Mode
Cell 5 of 80
Type here to search
29°C
ENG
16:30
19-11-2022

File Edit Selection View Go Run Terminal Help

Search

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 > Exploratory Data Analysis > df.isnull().sum()

+ Code + Markdown ··· Select Kernel

```
~\Anaconda3\lib\site-packages\sklearn\utils\validation.py in inner_f(*args, **kwargs)
     70                          FutureWarning)
     71         kwargs.update({k: arg for k, arg in zip(sig.parameters, args)})
...
--> 296                 waiter.acquire()
    297                 gotit = True
    298             else:

KeyboardInterrupt:
```

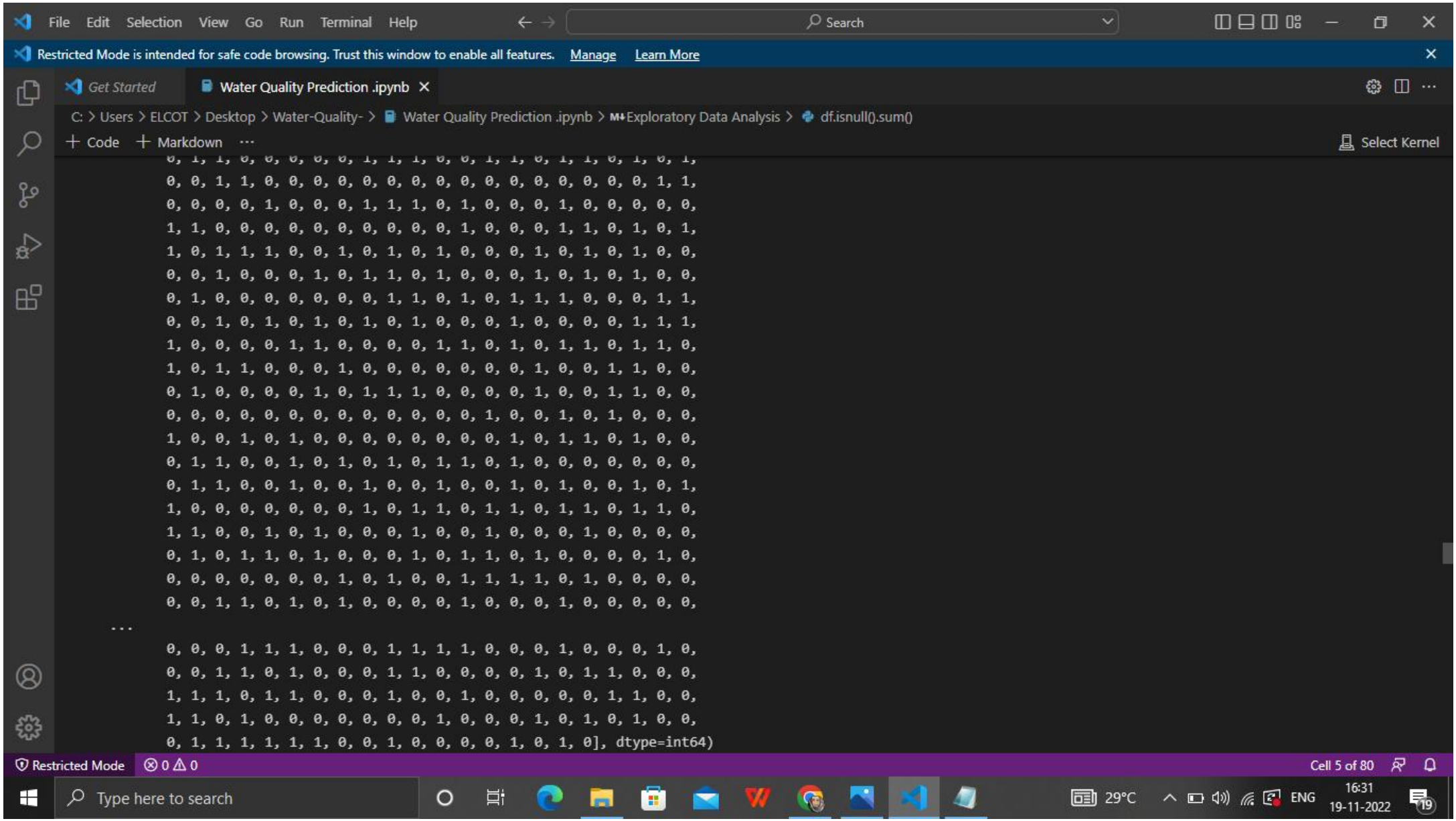
```
dt_y_predicted = grid_search_dt.predict(X_test)
dt_y_predicted
```

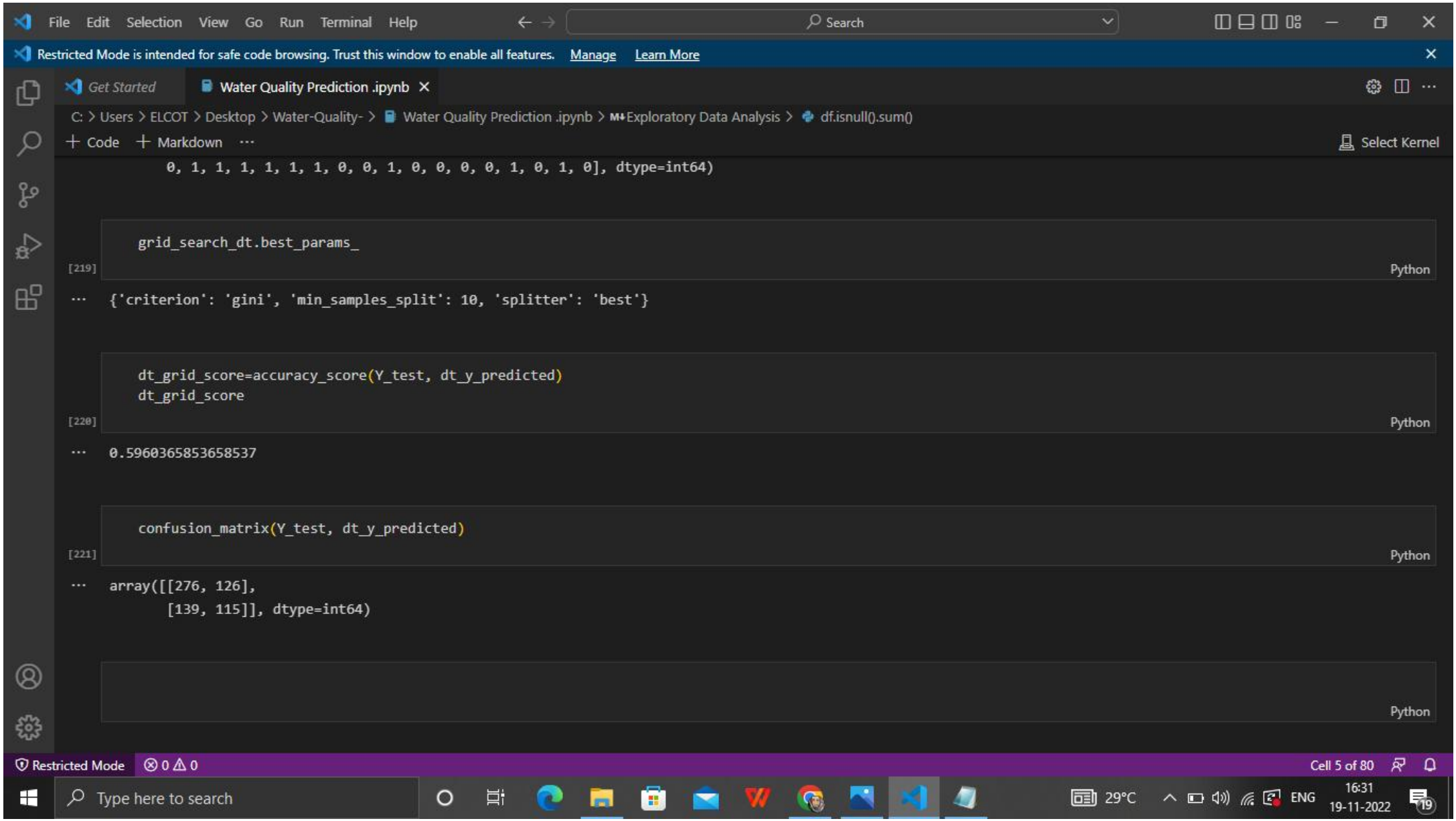
[218] Python

```
Output exceeds the size limit. Open the full output data in a text editor
array([1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1,
       1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0,
       0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1,
       0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1,
       0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0,
       0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1,
       0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
       0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0,
```

Restricted Mode ⊗ 0 ⚠ 0 Cell 5 of 80

Type here to search 29°C ENG 16:31 19-11-2022


File Edit Selection View Go Run Terminal Help
Search
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 > Exploratory Data Analysis > df.isnull().sum()
+ Code + Markdown ···
Select Kernel
0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0,
1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 1,
1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0,
0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0,
0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1,
0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1,
1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0,
1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0,
0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0,
1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0,
0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0,
0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1,
1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0,
1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0,
0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0,
0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0,
0, 0, 1, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0,
...
0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0,
0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0,
1, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0,
1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0,
0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0], dtype=int64)
Restricted Mode
0 0
Cell 5 of 80
Type here to search
29°C
ENG
16:31
19-11-2022


File Edit Selection View Go Run Terminal Help
Search
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 > Exploratory Data Analysis > df.isnull().sum()
+ Code + Markdown ···
Select Kernel
0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0], dtype=int64)
grid_search_dt.best_params_
[219]
Python
{'criterion': 'gini', 'min_samples_split': 10, 'splitter': 'best'}
dt_grid_score=accuracy_score(Y_test, dt_y_predicted)
dt_grid_score
[220]
Python
0.5960365853658537
confusion_matrix(Y_test, dt_y_predicted)
[221]
Python
array([[276, 126],
[139, 115]], dtype=int64)
Python
Restricted Mode
⊗ 0 ⚠ 0
Cell 5 of 80
Type here to search
29°C
ENG
16:31
19-11-2022