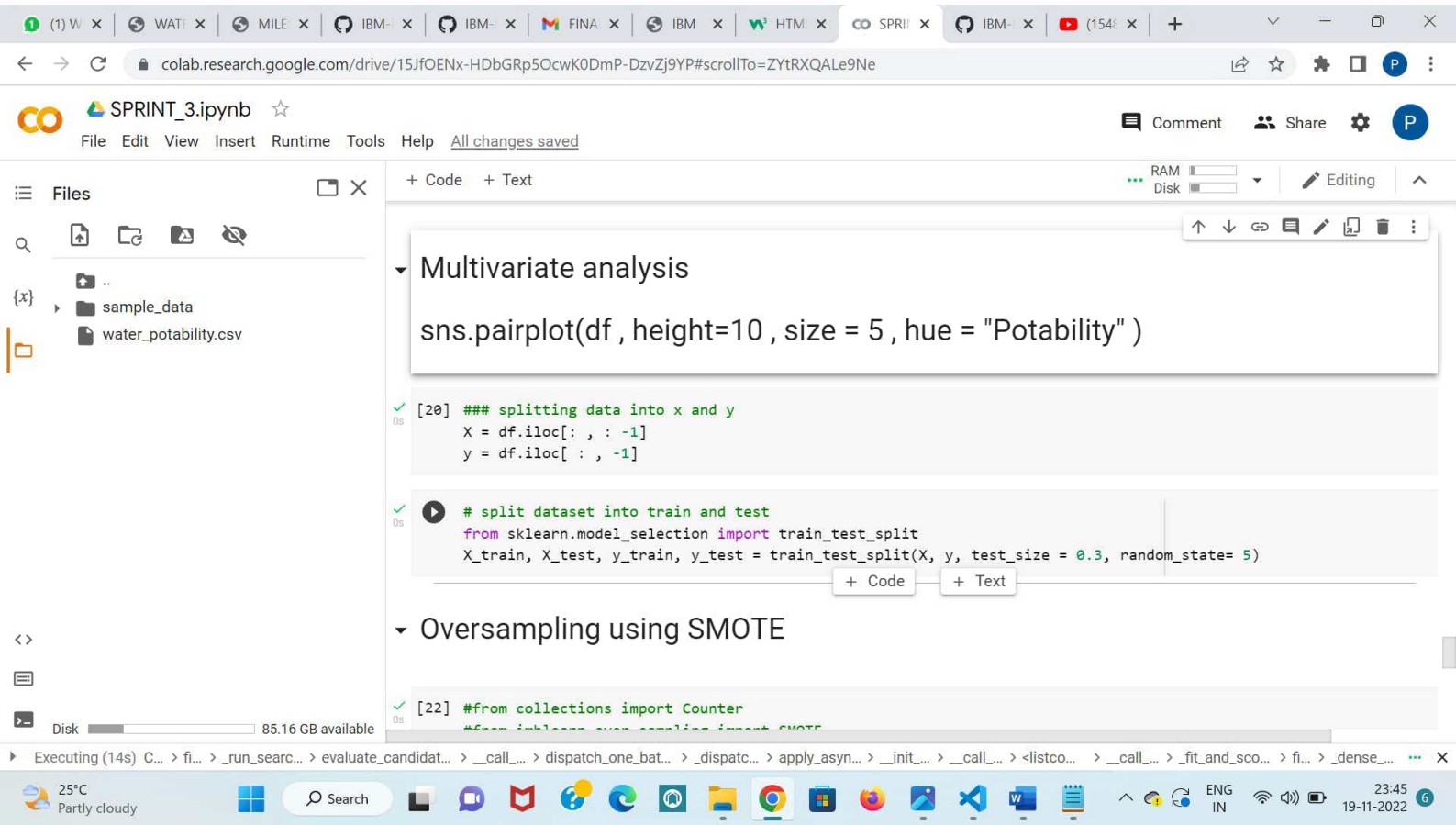
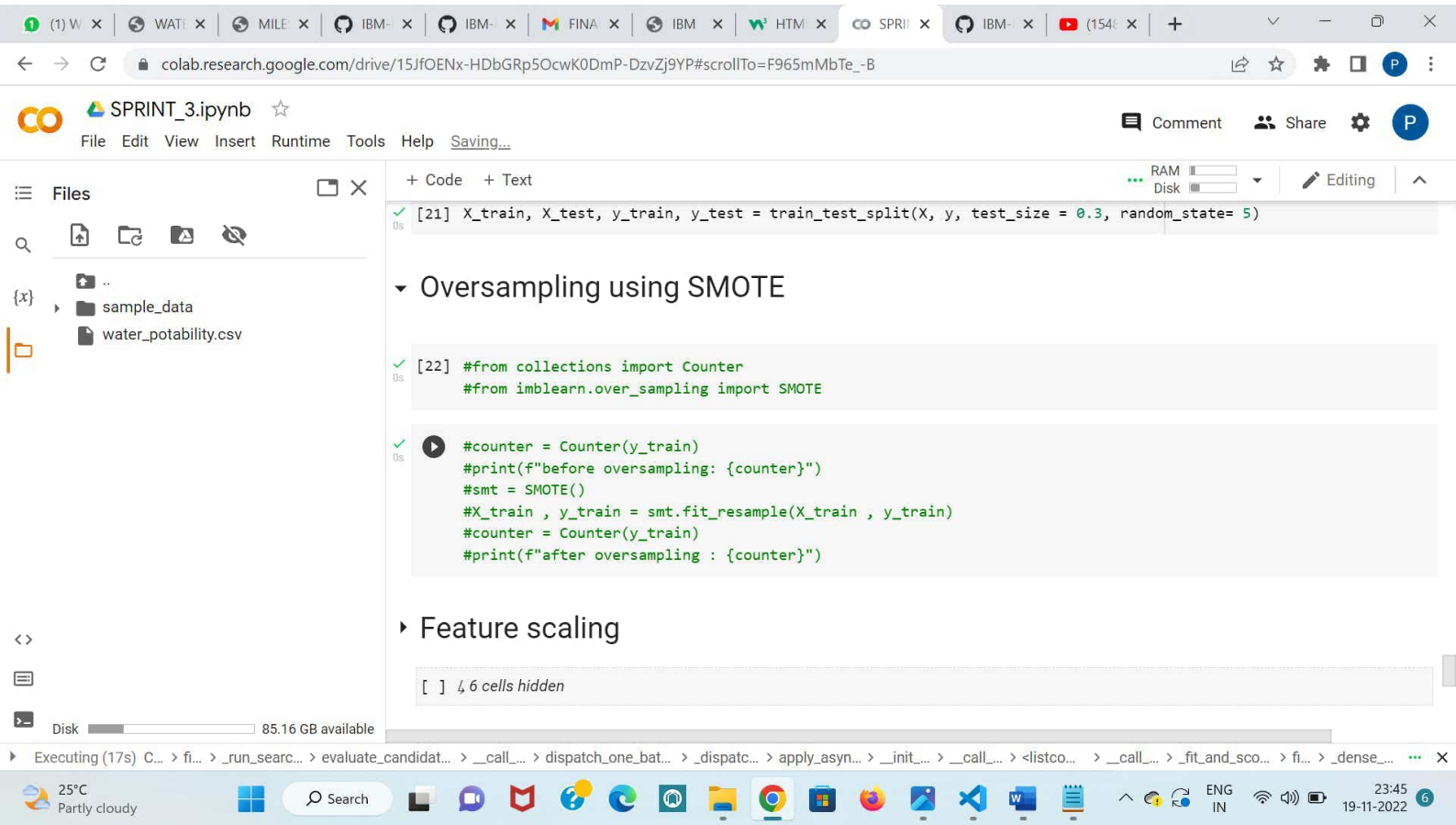
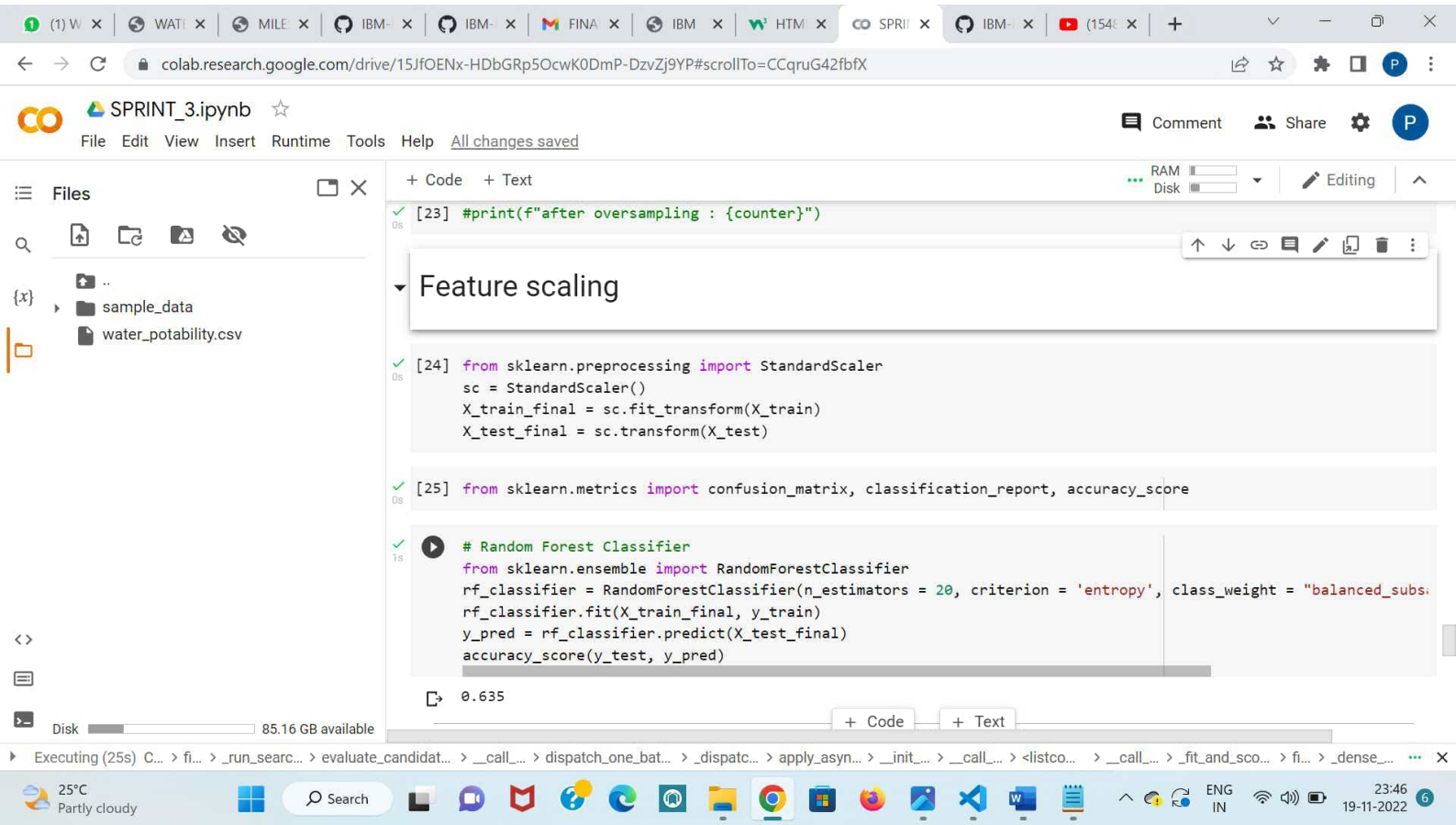
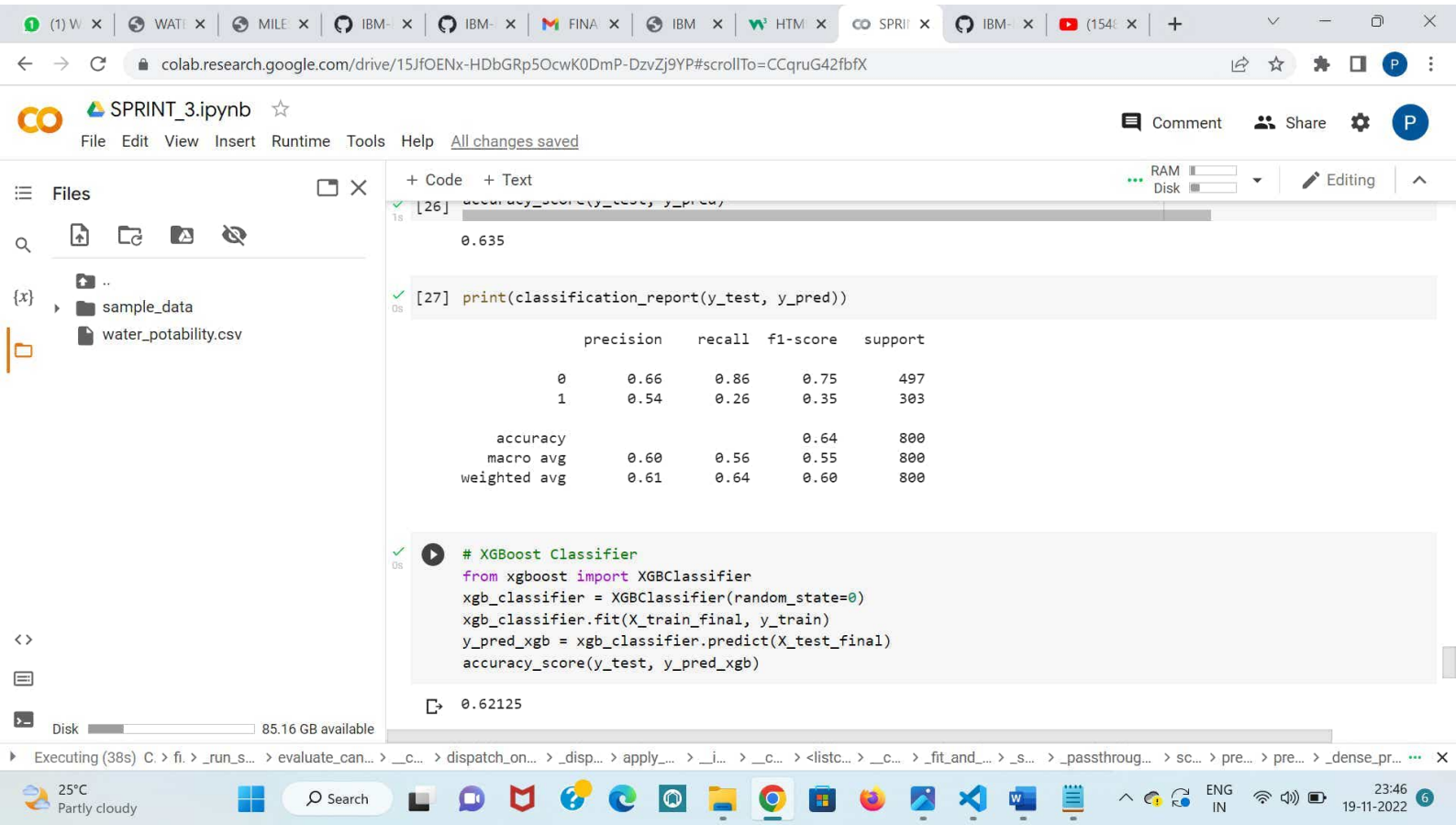

colab.research.google.com/drive/15JfOENx-HDbGRp5OcwK0DmP-DzvZj9YP#scrollTo=ZYtRXQALe9Ne
SPRINT_3.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
sample_data
water_potability.csv
Disk 85.16 GB available
+ Code + Text
RAM
Disk
Editing
Multivariate analysis
sns.pairplot(df , height=10 , size = 5 , hue = "Potability")
[20] ### splitting data into x and y
X = df.iloc[: , : -1]
y = df.iloc[: , -1]
# split dataset into train and test
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, random_state= 5)
+ Code + Text
Oversampling using SMOTE
[22] #from collections import Counter
Executing (14s) C... > fi... > _run_searc... > evaluate_candidat... > __call_... > dispatch_one_bat... > _dispatc... > apply_asyn... > __init_... > __call_... > <listco... > __call_... > _fit_and_sco... > fi... > _dense_...
25°C
Partly cloudy
Search
ENG
IN
23:45
19-11-2022


colab.research.google.com/drive/15JfOENx-HDbGRp5OcwK0DmP-DzvZj9YP#scrollTo=F965mMbTe_-B
SPRINT_3.ipynb
File Edit View Insert Runtime Tools Help Saving...
Comment
Share
Files
sample_data
water_potability.csv
Disk 85.16 GB available
+ Code + Text
RAM
Disk
Editing
[21] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, random_state= 5)
Oversampling using SMOTE
[22] #from collections import Counter
#from imblearn.over_sampling import SMOTE
#counter = Counter(y_train)
#print(f"before oversampling: {counter}")
#smt = SMOTE()
#X_train , y_train = smt.fit_resample(X_train , y_train)
#counter = Counter(y_train)
#print(f"after oversampling : {counter}")
Feature scaling
[ ] 6 cells hidden
Executing (17s)
25°C
Partly cloudy
Search
ENG
IN
23:45
19-11-2022


colab.research.google.com/drive/15JfOENx-HDbGRp5OcwK0DmP-DzvZj9YP#scrollTo=CCqruG42fbfX
SPRINT_3.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
sample_data
water_potability.csv
Disk 85.16 GB available
+ Code + Text
RAM
Disk
Editing
[23] #print(f"after oversampling : {counter}")
Feature scaling
[24] from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train_final = sc.fit_transform(X_train)
X_test_final = sc.transform(X_test)
[25] from sklearn.metrics import confusion_matrix, classification_report, accuracy_score
# Random Forest Classifier
from sklearn.ensemble import RandomForestClassifier
rf_classifier = RandomForestClassifier(n_estimators = 20, criterion = 'entropy', class_weight = "balanced_subs
rf_classifier.fit(X_train_final, y_train)
y_pred = rf_classifier.predict(X_test_final)
accuracy_score(y_test, y_pred)
0.635
+ Code + Text
Executing (25s) C... > fi... > _run_searc... > evaluate_candidat... > __call_... > dispatch_one_bat... > _dispatc... > apply_asyn... > __init_... > __call_... > <listco... > __call_... > _fit_and_sco... > fi... > _dense_...
25°C
Partly cloudy
Search
ENG
IN
23:46
19-11-2022


colab.research.google.com/drive/15JfOENx-HDbGRp5OcwK0DmP-DzvZj9YP#scrollTo=CCqruG42fbfX
SPRINT_3.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
sample_data
water_potability.csv
Disk 85.16 GB available
+ Code + Text
RAM
Disk
Editing
[26]
0.635
[27] print(classification_report(y_test, y_pred))
precision recall f1-score support
0 0.66 0.86 0.75 497
1 0.54 0.26 0.35 303
accuracy 0.64 800
macro avg 0.60 0.56 0.55 800
weighted avg 0.61 0.64 0.60 800
# XGBoost Classifier
from xgboost import XGBClassifier
xgb_classifier = XGBClassifier(random_state=0)
xgb_classifier.fit(X_train_final, y_train)
y_pred_xgb = xgb_classifier.predict(X_test_final)
accuracy_score(y_test, y_pred_xgb)
0.62125
Executing (38s)
25°C
Partly cloudy
Search
ENG
IN
23:46
19-11-2022

(1) W | WAT | MILE | IBM- | IBM- | FINA | IBM | HTM | SPRI | IBM- | (154

colab.research.google.com/drive/15JfOENx-HDbGRp5OcwK0DmP-DzvZj9YP#scrollTo=GHPmvMf-fyFA

SPRINT_3.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share P

Files

..
sample_data
water_potability.csv

Disk 85.16 GB available

+ Code + Text

RAM Disk Editing

```
[28] accuracy_score(y_test, y_pred_xgb)
```

```
0.62125
```

```
print(classification_report(y_test, y_pred_xgb))
```

```
              precision    recall  f1-score   support

           0       0.64      0.90      0.75       497
           1       0.50      0.17      0.25       303

    accuracy                           0.62       800
   macro avg       0.57      0.53      0.50       800
weighted avg       0.59      0.62      0.56       800
```

Support vector Machine

```
[30] # Support vector classifier
     from sklearn.svm import SVC
     svc_classifier = SVC(class_weight = "balanced" )
     svc_classifier.fit(X_train_final, y_train)
```

Executing (44s) C... > fi... > _run_searc... > evaluate_candidat... > __call_... > dispatch_one_bat... > _dispatc... > apply_asyn... > __init_... > __call_... > <listco... > __call_... > _fit_and_sco... > fi... > _dense_...

25°C Partly cloudy

Search

ENG IN

23:46 19-11-2022

colab.research.google.com/drive/15JfOENx-HDbGRp5OcwK0DmP-DzvZj9YP#scrollTo=GHPmvMf-fyFA

SPRINT_3.ipynb

File Edit View Insert Runtime Tools Help Saving...

Comment Share

+ Code + Text

RAM Disk

Editing

Files

..

sample_data

water_potability.csv

Disk 85.16 GB available

Support vector Machine

```
[30] # Support vector classifier
     from sklearn.svm import SVC
     svc_classifier = SVC(class_weight = "balanced" )
     svc_classifier.fit(X_train_final, y_train)
     y_pred_scv = svc_classifier.predict(X_test_final)
     accuracy_score(y_test, y_pred_scv)
```

```
0.6225
```

```
print(classification_report(y_test, y_pred_scv))
```

```
              precision    recall  f1-score   support

           0       0.70      0.69      0.70       497
           1       0.50      0.50      0.50       303

    accuracy                           0.62       800
   macro avg       0.60      0.60      0.60       800
weighted avg       0.62      0.62      0.62       800
```

Executing (52s) C... > fi... > _run_searc... > evaluate_candidat... > __call_... > dispatch_one_bat... > _dispatc... > apply_asyn... > __init_... > __call_... > <listco... > __call_... > _fit_and_sco... > fi... > _dense_...

25°C Partly cloudy

Search

ENG IN

23:46 19-11-2022

(1) W | WAT | MILE | IBM- | IBM- | FINA | IBM | HTM | SPRI | IBM- | (154 | +

colab.research.google.com/drive/15JfOENx-HDbGRp5OcwK0DmP-DzvZj9YP#scrollTo=GHPmvMf-fyFA

SPRINT_3.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share P

RAM Disk Editing

+ Code + Text

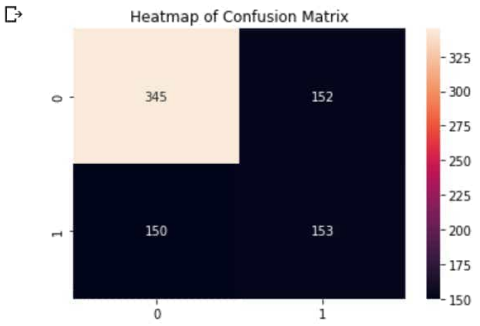
Files

..
sample_data
water_potability.csv

Disk 85.16 GB available

```
[31]
                accuracy                           0.62       800
               macro avg       0.60      0.60      0.60       800
            weighted avg       0.62      0.62      0.62       800
```

```
cm = confusion_matrix(y_test, y_pred_scv)
plt.title('Heatmap of Confusion Matrix', fontsize = 12)
sns.heatmap(cm, annot = True, fmt = "d")
plt.show()
```



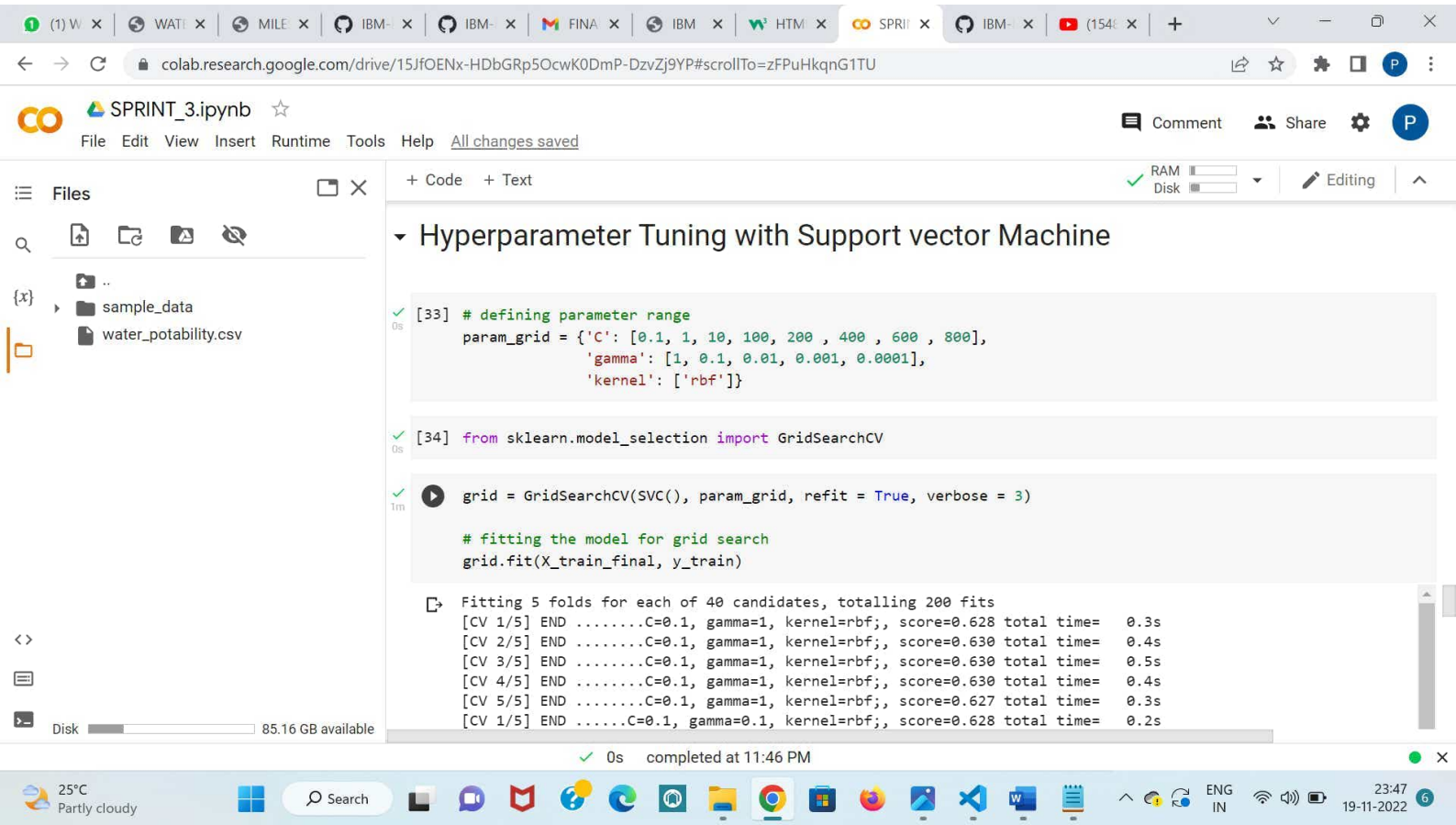
Executing (52s) C... > fi... > _run_searc... > evaluate_candidat... > __call_... > dispatch_one_bat... > _dispatc... > apply_asyn... > __init_... > __call_... > <listco... > __call_... > _fit_and_sco... > fi... > _dense_...

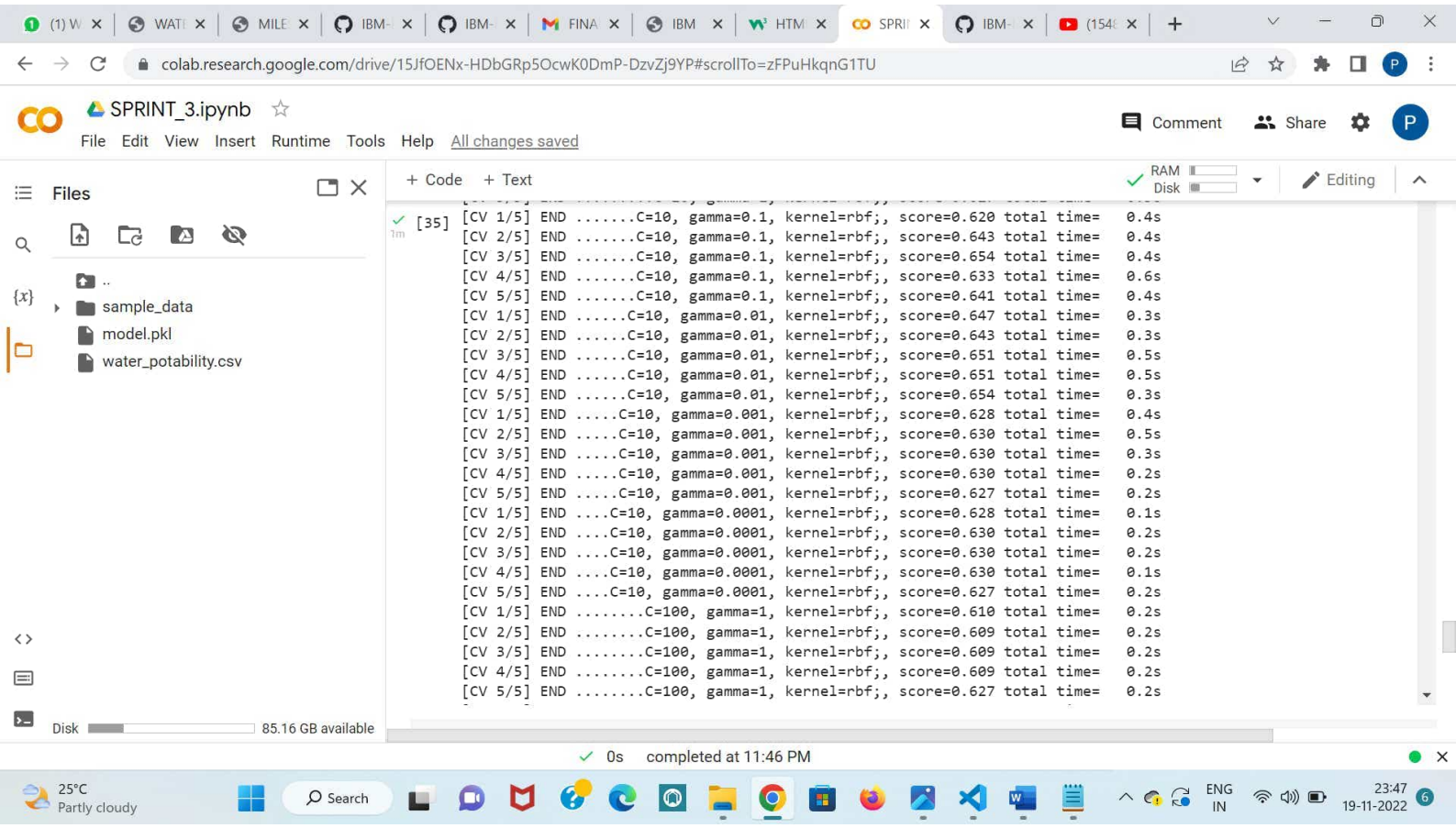
25°C Partly cloudy

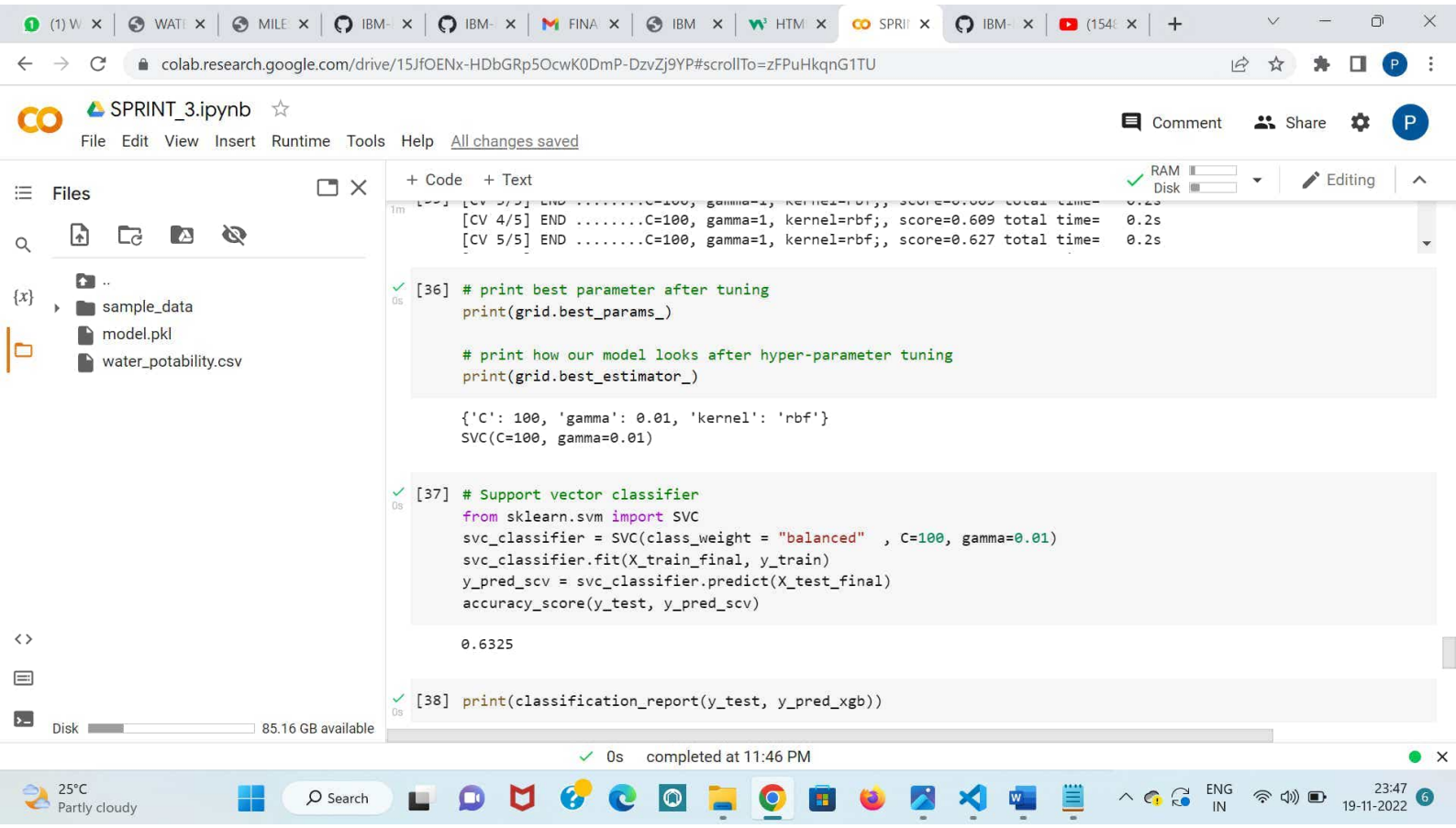
Search

ENG IN

23:46 19-11-2022


colab.research.google.com/drive/15JfOENx-HDbGRp5OcwK0DmP-DzvZj9YP#scrollTo=zFPuHkqnG1TU
SPRINT_3.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
sample_data
water_potability.csv
Disk 85.16 GB available
+ Code + Text
RAM
Disk
Editing
Hyperparameter Tuning with Support vector Machine
[33] # defining parameter range
param_grid = {'C': [0.1, 1, 10, 100, 200 , 400 , 600 , 800],
'gamma': [1, 0.1, 0.01, 0.001, 0.0001],
'kernel': ['rbf']}
[34] from sklearn.model_selection import GridSearchCV
grid = GridSearchCV(SVC(), param_grid, refit = True, verbose = 3)
# fitting the model for grid search
grid.fit(X_train_final, y_train)
Fitting 5 folds for each of 40 candidates, totalling 200 fits
[CV 1/5] END .....C=0.1, gamma=1, kernel=rbf;, score=0.628 total time= 0.3s
[CV 2/5] END .....C=0.1, gamma=1, kernel=rbf;, score=0.630 total time= 0.4s
[CV 3/5] END .....C=0.1, gamma=1, kernel=rbf;, score=0.630 total time= 0.5s
[CV 4/5] END .....C=0.1, gamma=1, kernel=rbf;, score=0.630 total time= 0.4s
[CV 5/5] END .....C=0.1, gamma=1, kernel=rbf;, score=0.627 total time= 0.3s
[CV 1/5] END .....C=0.1, gamma=0.1, kernel=rbf;, score=0.628 total time= 0.2s
0s completed at 11:46 PM
25°C Partly cloudy
Search
ENG IN
23:47 19-11-2022


colab.research.google.com/drive/15JfOENx-HDbGRp5OcwK0DmP-DzvZj9YP#scrollTo=zFPuHkqnG1TU
SPRINT_3.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
sample_data
model.pkl
water_potability.csv
Disk 85.16 GB available
+ Code + Text
RAM
Disk
Editing
[35]
[CV 1/5] END .....C=10, gamma=0.1, kernel=rbf;, score=0.620 total time= 0.4s
[CV 2/5] END .....C=10, gamma=0.1, kernel=rbf;, score=0.643 total time= 0.4s
[CV 3/5] END .....C=10, gamma=0.1, kernel=rbf;, score=0.654 total time= 0.4s
[CV 4/5] END .....C=10, gamma=0.1, kernel=rbf;, score=0.633 total time= 0.6s
[CV 5/5] END .....C=10, gamma=0.1, kernel=rbf;, score=0.641 total time= 0.4s
[CV 1/5] END ......C=10, gamma=0.01, kernel=rbf;, score=0.647 total time= 0.3s
[CV 2/5] END ......C=10, gamma=0.01, kernel=rbf;, score=0.643 total time= 0.3s
[CV 3/5] END ......C=10, gamma=0.01, kernel=rbf;, score=0.651 total time= 0.5s
[CV 4/5] END ......C=10, gamma=0.01, kernel=rbf;, score=0.651 total time= 0.5s
[CV 5/5] END ......C=10, gamma=0.01, kernel=rbf;, score=0.654 total time= 0.3s
[CV 1/5] END .....C=10, gamma=0.001, kernel=rbf;, score=0.628 total time= 0.4s
[CV 2/5] END .....C=10, gamma=0.001, kernel=rbf;, score=0.630 total time= 0.5s
[CV 3/5] END .....C=10, gamma=0.001, kernel=rbf;, score=0.630 total time= 0.3s
[CV 4/5] END .....C=10, gamma=0.001, kernel=rbf;, score=0.630 total time= 0.2s
[CV 5/5] END .....C=10, gamma=0.001, kernel=rbf;, score=0.627 total time= 0.2s
[CV 1/5] END ....C=10, gamma=0.0001, kernel=rbf;, score=0.628 total time= 0.1s
[CV 2/5] END ....C=10, gamma=0.0001, kernel=rbf;, score=0.630 total time= 0.2s
[CV 3/5] END ....C=10, gamma=0.0001, kernel=rbf;, score=0.630 total time= 0.2s
[CV 4/5] END ....C=10, gamma=0.0001, kernel=rbf;, score=0.630 total time= 0.1s
[CV 5/5] END ....C=10, gamma=0.0001, kernel=rbf;, score=0.627 total time= 0.2s
[CV 1/5] END ....C=100, gamma=1, kernel=rbf;, score=0.610 total time= 0.2s
[CV 2/5] END ....C=100, gamma=1, kernel=rbf;, score=0.609 total time= 0.2s
[CV 3/5] END ....C=100, gamma=1, kernel=rbf;, score=0.609 total time= 0.2s
[CV 4/5] END ....C=100, gamma=1, kernel=rbf;, score=0.609 total time= 0.2s
[CV 5/5] END ....C=100, gamma=1, kernel=rbf;, score=0.627 total time= 0.2s
0s completed at 11:46 PM
25°C
Partly cloudy
Search
ENG
IN
23:47
19-11-2022


colab.research.google.com/drive/15JfOENx-HDbGRp5OcwK0DmP-DzvZj9YP#scrollTo=zFPuHkqnG1TU
SPRINT_3.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
sample_data
model.pkl
water_potability.csv
Disk 85.16 GB available
+ Code + Text
RAM
Disk
Editing
[CV 4/5] END .....C=100, gamma=1, kernel=rbf;, score=0.609 total time= 0.2s
[CV 5/5] END .....C=100, gamma=1, kernel=rbf;, score=0.627 total time= 0.2s
[36] # print best parameter after tuning
print(grid.best_params_)
# print how our model looks after hyper-parameter tuning
print(grid.best_estimator_)
{'C': 100, 'gamma': 0.01, 'kernel': 'rbf'}
SVC(C=100, gamma=0.01)
[37] # Support vector classifier
from sklearn.svm import SVC
svc_classifier = SVC(class_weight = "balanced" , C=100, gamma=0.01)
svc_classifier.fit(X_train_final, y_train)
y_pred_scv = svc_classifier.predict(X_test_final)
accuracy_score(y_test, y_pred_scv)
0.6325
[38] print(classification_report(y_test, y_pred_xgb))
0s completed at 11:46 PM
25°C Partly cloudy
Search
ENG IN
23:47 19-11-2022

colab.research.google.com/drive/15JfOENx-HDbGRp5OcwK0DmP-DzvZj9YP#scrollTo=zFPuHkqnG1TU

SPRINT_3.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

Files

.. sample_data model.pkl water_potability.csv

Disk 85.16 GB available

+ Code + Text

RAM Disk Editing

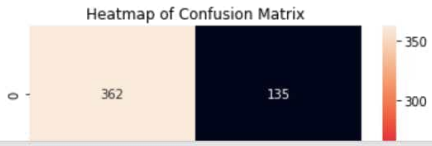
```
print(classification_report(y_test, y_pred_xgb))
```

```
              precision    recall  f1-score   support

           0       0.64      0.90      0.75       497
           1       0.50      0.17      0.25       303

    accuracy                           0.62       800
   macro avg       0.57      0.53      0.50       800
weighted avg       0.59      0.62      0.56       800
```

```
[39] cm = confusion_matrix(y_test, y_pred_scv)
     plt.title('Heatmap of Confusion Matrix', fontsize = 12)
     sns.heatmap(cm, annot = True , fmt = "d")
     plt.show()
```



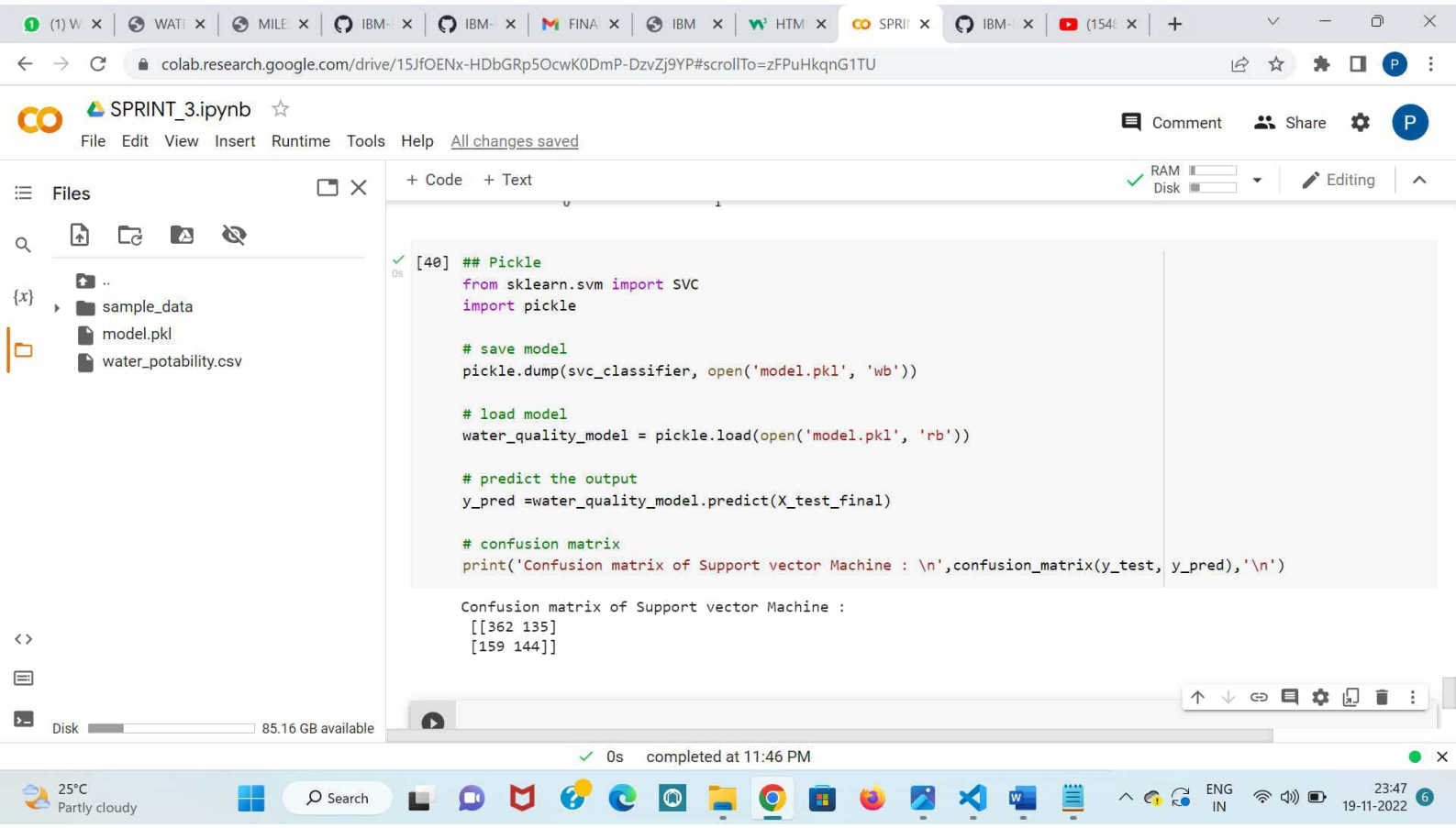
0s completed at 11:46 PM

25°C Partly cloudy

Search

ENG IN

23:47 19-11-2022


colab.research.google.com/drive/15JfOENx-HDbGRp5OcwK0DmP-DzvZj9YP#scrollTo=zFPuHkqnG1TU
SPRINT_3.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
Editing
..
sample_data
model.pkl
water_potability.csv
[40] ## Pickle
from sklearn.svm import SVC
import pickle
# save model
pickle.dump(svc_classifier, open('model.pkl', 'wb'))
# load model
water_quality_model = pickle.load(open('model.pkl', 'rb'))
# predict the output
y_pred =water_quality_model.predict(X_test_final)
# confusion matrix
print('Confusion matrix of Support vector Machine : \n',confusion_matrix(y_test, y_pred),'\n')
Confusion matrix of Support vector Machine :
[[362 135]
[159 144]]
Disk 85.16 GB available
0s completed at 11:46 PM
25°C
Partly cloudy
Search
ENG
IN
23:47
19-11-2022