

SPRINT 3

Project Deliverables (Model Building Code & Evaluation)

Team ID	PNT2022TMID51670
Project Name	Efficient Water Quality Analysis & Prediction using Machine Learning

Using the best accuracy algorithm (SVC) we are going to train our model for deployment:

The screenshot shows a Visual Studio Code window with a Jupyter Notebook titled "Water_quality.ipynb". The notebook is open to a cell containing Python code for training a Support Vector Classifier (SVC) model. The code imports `svm` from `sklearn.svm` and `SVC` from `sklearn.svm`. It then creates an `SVC` object with `class_weight = "balanced"`, fits it to the training data (`X_train_final`, `y_train`), and uses it to predict on the test data (`X_test_final`). The accuracy score is printed, showing a value of 0.6225.

```
# 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)
```

The output of the cell shows the accuracy score: 0.6225.

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

The output of the cell shows the classification report:

	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

The terminal window shows the following output:

```
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
* Restarting with stat
C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:329: UserWarning: Trying to unpickle estimator StandardScaler from version 0.24.0 when using version 1.1.3. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to:
https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
warnings.warn(
* Debugger is active!
* Debugger PIN: 873-600-839
127.0.0.1 - - [11/Nov/2022 13:57:12] "GET / HTTP/1.1" 200 -
C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:443: UserWarning: X has feature names, but StandardScaler was fitted without feature names
warnings.warn(
127.0.0.1 - - [11/Nov/2022 13:57:43] "POST /predict HTTP/1.1" 200 -
```

The bottom status bar shows the Jupyter Server is running on the local machine, and the current cell is 2 of 52. The system clock shows 04:40 PM on 11-11-2022.

Visual Studio Code interface showing a Jupyter Notebook titled "Water_quality.ipynb" with a Python cell. The cell contains code to calculate a confusion matrix and display a heatmap.

```
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()
```

The output of the cell is a heatmap titled "Heatmap of Confusion Matrix". The x-axis is labeled 0 and 1, and the y-axis is labeled 0 and 1. The values in the matrix are:

	0	1
0	345	152
1	150	153

The bottom panel shows the terminal output, including a warning about the development server and a message about the StandardScaler estimator.

WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.

* Running on http://127.0.0.1:5000

Press CTRL+C to quit

* Restarting with stat

C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:329: UserWarning: Trying to unpickle estimator StandardScaler from version 0.24.0 when using version 1.1.3. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to: https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations

warnings.warn(

* Debugger is active!

* Debugger PIN: 873-600-839

127.0.0.1 - - [11/Nov/2022 13:57:12] "GET / HTTP/1.1" 200 -

C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:443: UserWarning: X has feature names, but StandardScaler was fitted without feature names

warnings.warn(

127.0.0.1 - - [11/Nov/2022 13:57:43] "POST /predict HTTP/1.1" 200 -

File Edit Selection View Go Run Terminal Help

Water_quality.ipynb - Water Quality - Visual Studio Code

Python 3.9.1 64-bit

EXPLORER

WATER_QUALITY

static

templates

home.html

app.py

model.pkl

my_scaler.save

Profile

requirements.txt

water_potability.csv

Water_quality.ipynb

Water_quality.ipynb

home.html 2

water_potability.csv

Water_quality.ipynb

Problem Statement

Task 1

Code

Markdown

Run All

Clear Outputs of All Cells

Restart

Variables

Outline

Hyperparameter Tuning with Support vector Machine

```
# 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']}
```

```
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)
```

Output exceeds the size limit. Open the full output data in a text editor

Fitting 5 folds for each of 40 candidates, totalling 200 fits

[CV 1/5] ENDC=0.1, gamma=1, kernel=rbf; score=0.628 total time= 0.2s

[CV 2/5] ENDC=0.1, gamma=1, kernel=rbf; score=0.630 total time= 0.2s

[CV 3/5] ENDC=0.1, gamma=1, kernel=rbf; score=0.630 total time= 0.1s

[CV 4/5] ENDC=0.1, gamma=1, kernel=rbf; score=0.630 total time= 0.1s

[CV 5/5] ENDC=0.1, gamma=1, kernel=rbf; score=0.627 total time= 0.1s

[CV 1/5] ENDC=0.1, gamma=0.1, kernel=rbf; score=0.638 total time= 0.1s

PROBLEMS 2

OUTPUT

DEBUG CONSOLE

TERMINAL

JUPYTER

WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.

* Running on http://127.0.0.1:5000

Press CTRL+C to quit

* Restarting with stat

C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:329: UserWarning: Trying to unpickle estimator StandardScaler from version 0.24.0 when using version 1.1.3. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to: https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations

warnings.warn(

* Debugger is active!

* Debugger PIN: 873-600-839

127.0.0.1 - - [11/Nov/2022 13:57:12] "GET / HTTP/1.1" 200 -

C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:443: UserWarning: X has feature names, but StandardScaler was fitted without feature names

warnings.warn(

127.0.0.1 - - [11/Nov/2022 13:57:43] "POST /predict HTTP/1.1" 200 -

2

OUTLINE

TIMELINE

cmd

Code

Activate Windows

Go to Settings to activate Windows.

Jupyter Server: Local

Cell 2 of 52

04:41 PM

11-11-2022

The screenshot displays the Visual Studio Code interface with a Jupyter Notebook titled 'Water_quality.ipynb'. The Explorer sidebar on the left shows the project structure, including files like 'app.py', 'model.pkl', 'my_scaler.save', 'requirements.txt', 'water_potability.csv', and 'Water_quality.ipynb'. The main editor area shows the notebook content, which includes a code cell with the following output:

```
[CV 5/5] END ...C=0.1, gamma=0.001, kernel=rbf, score=0.627 total time= 0.1s
[CV 1/5] END ...C=0.1, gamma=0.001, kernel=rbf, score=0.628 total time= 0.1s
[CV 2/5] END ...C=0.1, gamma=0.001, kernel=rbf, score=0.630 total time= 0.1s
[CV 3/5] END ...C=0.1, gamma=0.001, kernel=rbf, score=0.630 total time= 0.1s
[CV 4/5] END ...C=0.1, gamma=0.001, kernel=rbf, score=0.630 total time= 0.1s
...
[CV 2/5] END ...C=800, gamma=0.001, kernel=rbf, score=0.630 total time= 0.3s
[CV 3/5] END ...C=800, gamma=0.001, kernel=rbf, score=0.630 total time= 0.3s
[CV 4/5] END ...C=800, gamma=0.001, kernel=rbf, score=0.630 total time= 0.3s
[CV 5/5] END ...C=800, gamma=0.001, kernel=rbf, score=0.627 total time= 0.3s
```

Below the code cell, a variable inspector shows the 'GridSearchCV' object with the 'estimator' attribute set to 'SVC'. The next code cell contains the following Python code:

```
# print best parameter after tuning
print(grid.best_params_)

# print how our model looks after hyper-parameter tuning
print(grid.best_estimator_)
```

The output of this cell shows the best parameters and the trained SVC model:

```
{'C': 100, 'gamma': 0.01, 'kernel': 'rbf'}
SVC(C=100, gamma=0.01)
```

The bottom panel of the editor shows the 'TERMINAL' output, which includes a warning from the Jupyter server:

```
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
* Restarting with stat
C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:329: UserWarning: Trying to unpickle estimator StandardScaler from version 0.24.0 when using version 1.1.3. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to: https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
warnings.warn(
* Debugger is active!
* Debugger PIN: 873-600-839
127.0.0.1 - - [11/Nov/2022 13:57:12] "GET / HTTP/1.1" 200 -
C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:443: UserWarning: X has feature names, but StandardScaler was fitted without feature names
warnings.warn(
127.0.0.1 - - [11/Nov/2022 13:57:43] "POST /predict HTTP/1.1" 200 -
```

The bottom status bar indicates the Jupyter server is running locally on port 5000, and the system clock shows 04:41 PM on 11-11-2022.

Now we have successfully trained our model , then we will save our trained model using PICKLE library

Model.pkl file: (this is the file where we will be saving our trained model)

The screenshot displays the Visual Studio Code interface with a Jupyter Notebook titled "Water_quality.ipynb". The Explorer sidebar on the left shows a project structure with files like "app.py", "model.pkl", "my_scaler.save", "requirements.txt", "water_potability.csv", and "Water_quality.ipynb".

The notebook cell contains the following Python code:

```
## 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')
```

The output of the cell shows the execution time [45] ✓ 0.2s and the printed confusion matrix:

```
Confusion matrix of Support vector Machine :
[[362 135]
 [159 144]]
```

The bottom panel shows the Jupyter Terminal with the following output:

```
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
* Restarting with stat
C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:329: UserWarning: Trying to unpickle estimator StandardScaler from version 0.24.0 when using version 1.1.3. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to: https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
  warnings.warn(
* Debugger is active!
* Debugger PIN: 873-600-839
127.0.0.1 - - [11/Nov/2022 13:57:12] "GET / HTTP/1.1" 200 -
C:\Users\HOME\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:443: UserWarning: X has feature names, but StandardScaler was fitted without feature names
  warnings.warn(
127.0.0.1 - - [11/Nov/2022 13:57:43] "POST /predict HTTP/1.1" 200 -
```

The bottom right corner of the screen shows the Windows taskbar with the search bar, taskbar icons, and system tray showing the time as 04:41 PM on 11-11-2022.

Now it contains all the necessary files ..

