

SPRINT 3

Project Deliverables (Model Building Code & Evaluation)

Team ID	PNT2022TMID38291
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 and evaluating a Support Vector Machine (SVC) model. The code imports `svm` from `sklearn` and `SVC` from `sklearn.svm`. It then defines `svc_classifier` with `class_weight = "balanced"`, fits it to `X_train_final` and `y_train`, and predicts on `X_test_final` to get `y_pred_scv`. The accuracy score is calculated and printed. The output of the cell shows a classification report with precision, recall, f1-score, and support for classes 0 and 1, along with overall accuracy and macro/weighted averages.

```
# 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)
```

Output of the cell:

```
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

The terminal at the bottom shows a warning message: "WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead." followed by a message about the StandardScaler estimator version mismatch.

Activate Windows
Go to Settings to activate Windows.

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.ipynb x home.html 2 water_potability.csv

Water_quality.ipynb > M*Problem Statement > M*Task 1

+ Code + Markdown | ▶ Run All | Clear Outputs of All Cells | Restart | Variables | Outline

Water_quality.ipynb

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()

[37] ✓ 1.2s

Heatmap of Confusion Matrix

	0	1
0	345	152
1	150	153

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 -

cmd
Code

Activate Windows
Go to Settings to activate Windows.

Jupyter Server: Local Cell 2 of 52

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

M4Problem Statement

M4Task 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 -

Activate Windows

Go to Settings to activate Windows.

Jupyter Server: Local

Cell 2 of 52

The screenshot shows a Jupyter Notebook titled "Water_quality.ipynb" in Visual Studio Code. The notebook is running on a Jupyter Server. The code in the cell is as follows:

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

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

[41]: SVC(C=100, gamma=0.01, kernel='rbf')
```

The output of the code is a dictionary representing the best parameters found by GridSearchCV:

```
{'C': 100, 'gamma': 0.01, 'kernel': 'rbf'}
```

The terminal output shows a warning from sklearn about unpickling an estimator from an older version, and a message from the Jupyter Server indicating that the server is running on http://127.0.0.1:5000.

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 the project structure, including files like 'app.py', 'model.pkl', 'my_scaler.save', 'Profile', '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 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')
```

Below the code cell, the output shows the confusion matrix:

```
[[362 135]
 [159 144]]
```

The bottom of the interface features a terminal window 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 terminal window displays the 'Activate Windows' watermark and the text 'Go to Settings to activate Windows.'.

Now it contains all the necessary files ..

