

## SPRINT 3

### Project Deliverables (Model Building Code & Evaluation)

Team ID	PNT2022TMID21875
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 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`, predicts on `X_test_final` to get `y_pred_scv`, and calculates the `accuracy_score` between `y_test` and `y_pred_scv`. The output of this cell is `0.6225`.

The next cell in the notebook prints the `classification_report` for `y_test` and `y_pred_scv`. The output is as follows:

	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." It also shows the server running on `http://127.0.0.1:5000` and a `POST /predict HTTP/1.1` request.

File Edit Selection View Go Run Terminal Help

Water\_quality.ipynb - Water Quality - Visual Studio Code

app.py 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 | ...

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

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

	Actual 0	Actual 1
Predicted 0	345	152
Predicted 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

Type here to search

04:41 PM 11-11-2022

File Edit Selection View Go Run Terminal Help

Water\_quality.ipynb - Water Quality - Visual Studio Code

app.py 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

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

# 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] END	.....C=0.1, gamma=1, kernel=rbf;, score=0.628 total time=	0.2s
[CV 2/5] END	.....C=0.1, gamma=1, kernel=rbf;, score=0.630 total time=	0.2s
[CV 3/5] END	.....C=0.1, gamma=1, kernel=rbf;, score=0.630 total time=	0.1s
[CV 4/5] END	.....C=0.1, gamma=1, kernel=rbf;, score=0.630 total time=	0.1s
[CV 5/5] END	.....C=0.1, gamma=1, kernel=rbf;, score=0.627 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](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

OUTLINE

TIMELINE

Jupyter Server: Local

Cell 2 of 52

04:41 PM 11-11-2022

The screenshot displays a Visual Studio Code editor with a Jupyter Notebook titled "Water\_quality.ipynb". The notebook is open in the "Code" view, showing the following code:

```
from sklearn.model_selection import GridSearchCV
from sklearn.svm import SVC
from sklearn.metrics import r2_score

# Create the SVC model
svc = SVC()

# Define the parameter grid
param_grid = {'C': [0.1, 1, 10, 100], 'gamma': [0.001, 0.01, 0.1], 'kernel': ['rbf']}

# Create the GridSearchCV object
grid_search = GridSearchCV(svc, param_grid, cv=5)

# Fit the model
grid_search.fit(X_train, y_train)

# Print the best parameters
print("Best parameters found: ", grid_search.best_params_)

# Print the best estimator
print("Best estimator: ", grid_search.best_estimator_)

# Predict on the test set
y_pred = grid_search.predict(X_test)

# Calculate the R2 score
r2 = r2_score(y_test, y_pred)
```

The notebook also includes a plot of the cross-validated scores for the GridSearchCV object. The plot shows the mean cross-validated score and its standard error for different values of C and gamma. The best parameters found are C=100 and gamma=0.01.

The terminal output shows a warning message 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 -
```

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 shows a Visual Studio Code editor with a Jupyter Notebook titled 'Water\_quality.ipynb'. The notebook is open to a cell containing 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 is displayed below the code:

```
[45]: ✓ 0.2s
... Confusion matrix of Support vector Machine :
[[362 135]
 [159 144]]
```

The terminal window at the bottom shows the output of a REST client, including a warning about a development server and a successful POST request to /predict HTTP/1.1.

```
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 -
```

Now it contains all the necessary files ..

