In [56]:

```
import shap
import pandas as pd
import numpy as np
from xgboost import XGBClassifier
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.compose import ColumnTransformer
from sklearn.model_selection import StratifiedKFold, GridSearchCV,
train_test_split
from sklearn.preprocessing import OneHotEncoder, StandardScaler
from sklearn.metrics import accuracy_score, precision_score, recall_score,
f1_score, confusion_matrix, roc_auc_score
```

In [57]:

```
df = pd.read_csv('table.csv')
# replace C with 1 and P with 0
df['Label'] = df['Label'].replace({'C': 1, 'P': 0})
df
```

```
C:\Users\user\AppData\Local\Temp\ipykernel_13268\3825214023.py:3: FutureWarning:
Downcasting behavior in `replace` is deprecated and will be removed in a future v
ersion. To retain the old behavior, explicitly call `result.infer_objects(copy=Fa
lse)`. To opt-in to the future behavior, set `pd.set_option('future.no_silent_dow
ncasting', True)`
  df['Label'] = df['Label'].replace({'C': 1, 'P': 0})
```

Out[57]:

	mean_value_ML	mean_value_AP	mean_distance_ML	mean_distance_AP	mean_dista
0	0.046123	-1.461512	0.385651	0.543167	
1	0.042500	-0.365777	0.532939	0.484075	
2	0.496358	-1.401023	0.364302	0.400104	
3	0.314393	-0.549541	0.573516	0.486970	
4	1.412529	0.186249	0.929037	1.094830	
...	...	...	...	...	
100	2.605348	-2.030942	0.335179	0.524191	
101	0.776783	1.222163	0.541288	0.659628	
102	-0.708497	-0.800798	0.466417	0.711986	
103	0.268076	-1.346882	0.930665	0.949494	
104	3.524215	-1.546033	0.291584	0.580360	

105 rows × 73 columns

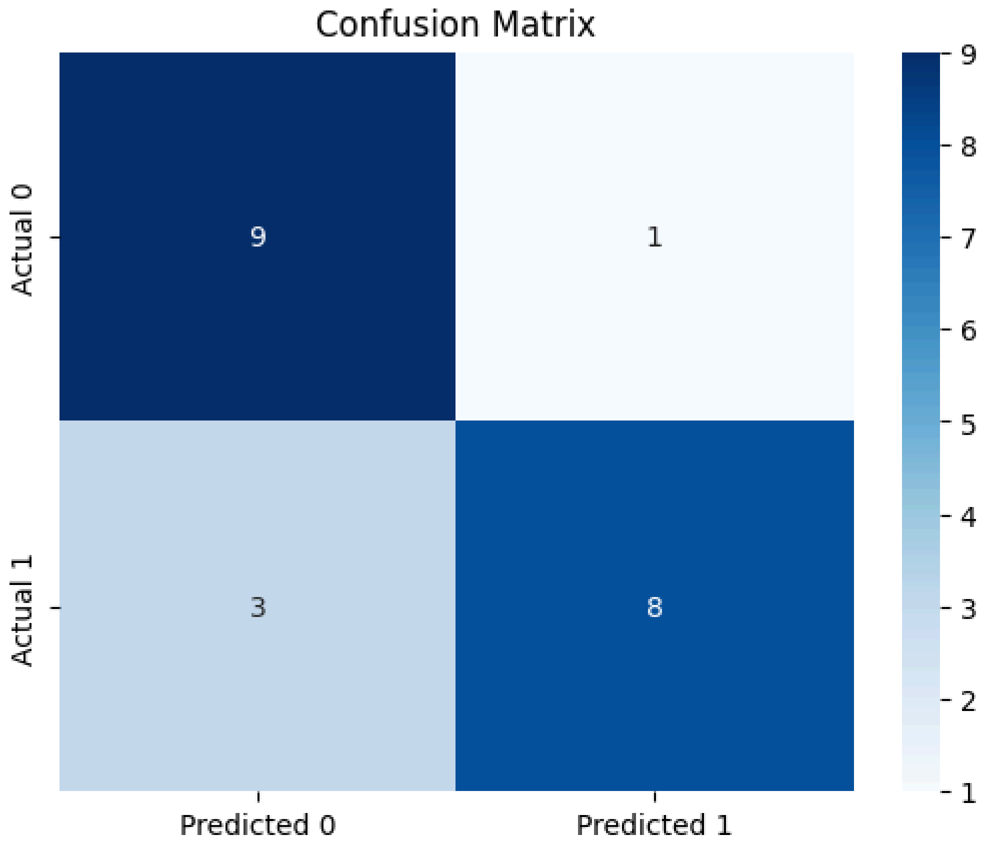
In [58]:

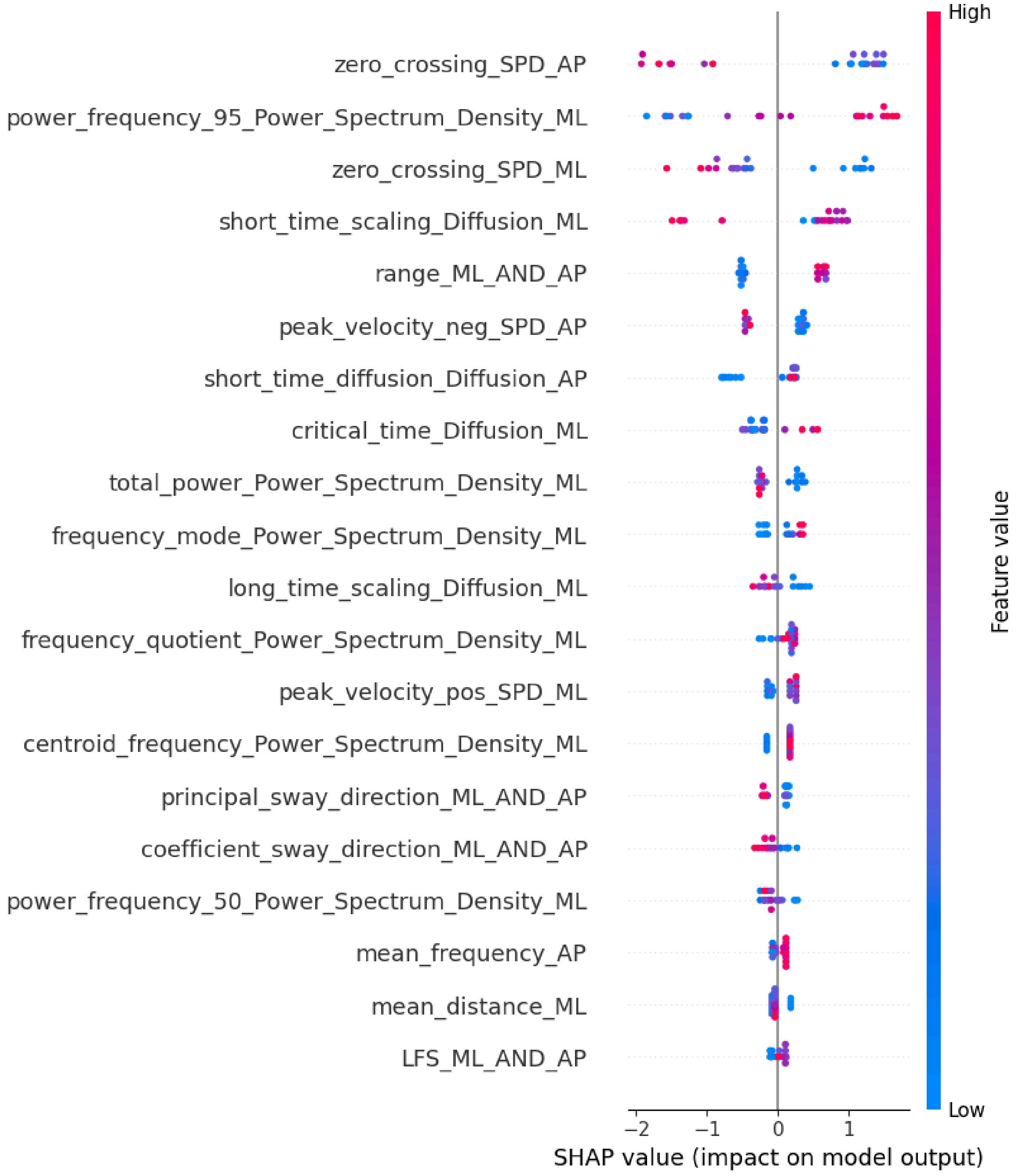
```
X = df.drop(['Label'], axis=1)
y = df['Label']
```

In [59]:

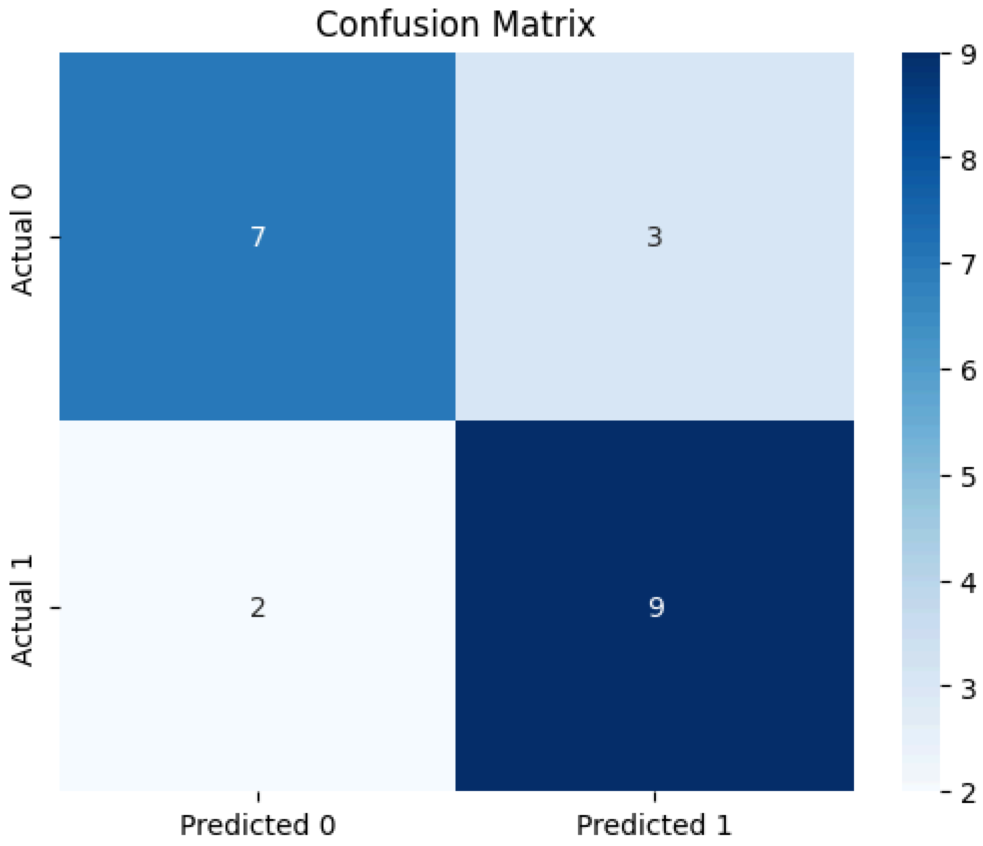
```
cv = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)

for i, (train_index, test_index) in enumerate(cv.split(X, y)):
    print(f'Results for split {i+1}')
    X_train, X_test = X.iloc[train_index], X.iloc[test_index]
    y_train, y_test = y.iloc[train_index], y.iloc[test_index]

    clf = XGBClassifier(eval_metric='logloss', random_state=42)
    clf.fit(X_train, y_train)

    y_pred = clf.predict(X_test)

    accuracy = accuracy_score(y_test, y_pred)
    print(f'Accuracy: {accuracy:.4f}')

    precision = precision_score(y_test, y_pred)
    print(f'Precision: {precision:.4f}')

    recall = recall_score(y_test, y_pred)
    print(f'Recall: {recall:.4f}')

    f1 = f1_score(y_test, y_pred)
    print(f'F1 Score: {f1:.4f}')

    cm = confusion_matrix(y_test, y_pred)
    sns.heatmap(cm, annot=True, fmt="d", cmap="Blues", xticklabels=['Predicted 0', 'Predicted 1'], yticklabels=['Actual 0', 'Actual 1'])
    plt.title('Confusion Matrix')
    plt.show()

    explainer = shap.TreeExplainer(clf)
    shap_values = explainer.shap_values(X_test)

    shap.summary_plot(shap_values, X_test)
```

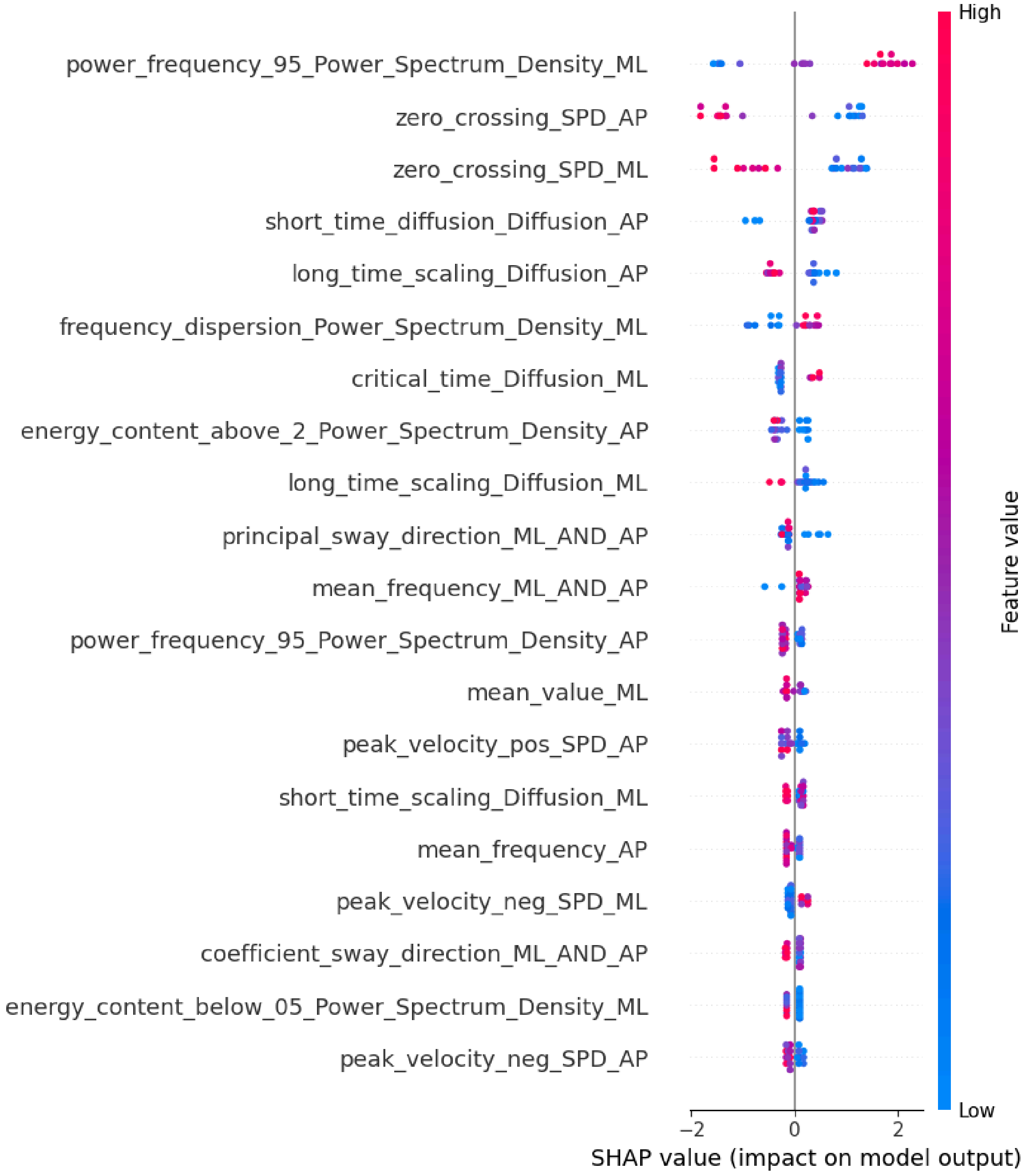
```
Results for split 1
Accuracy: 0.8095
Precision: 0.8889
Recall: 0.7273
F1 Score: 0.8000
```


Confusion Matrix
Actual 0
Actual 1
Predicted 0
Predicted 1
9
1
3
8
9
8
7
6
5
4
3
2
1

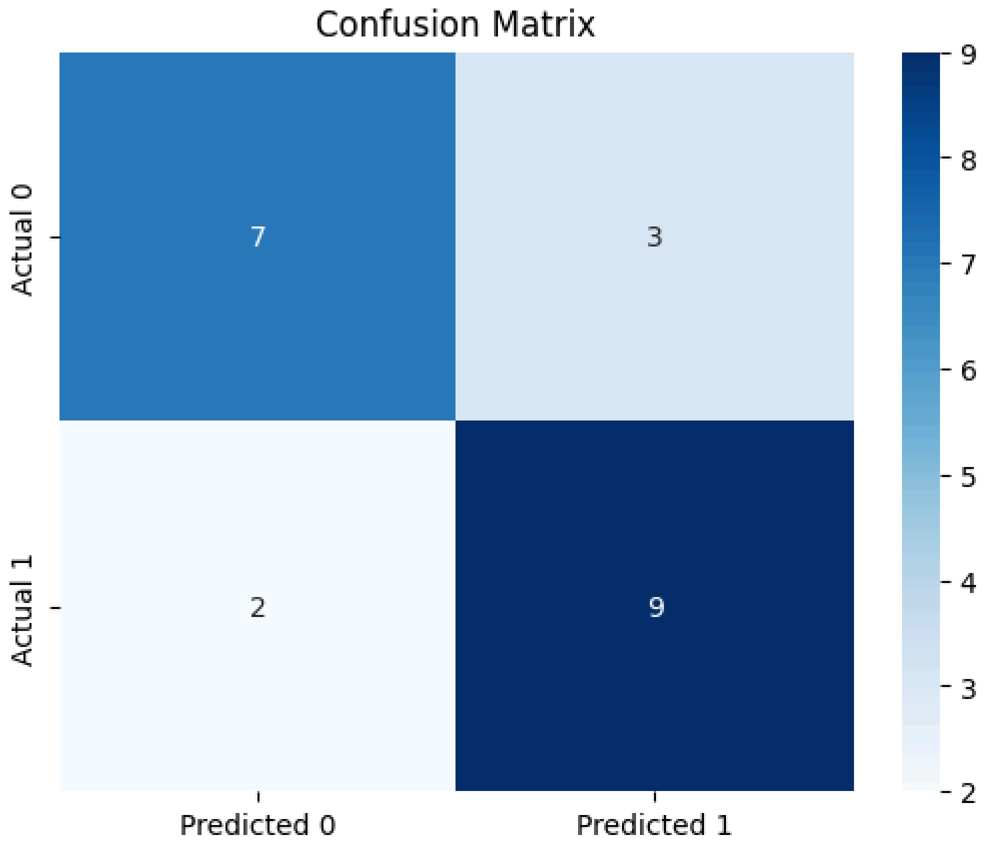


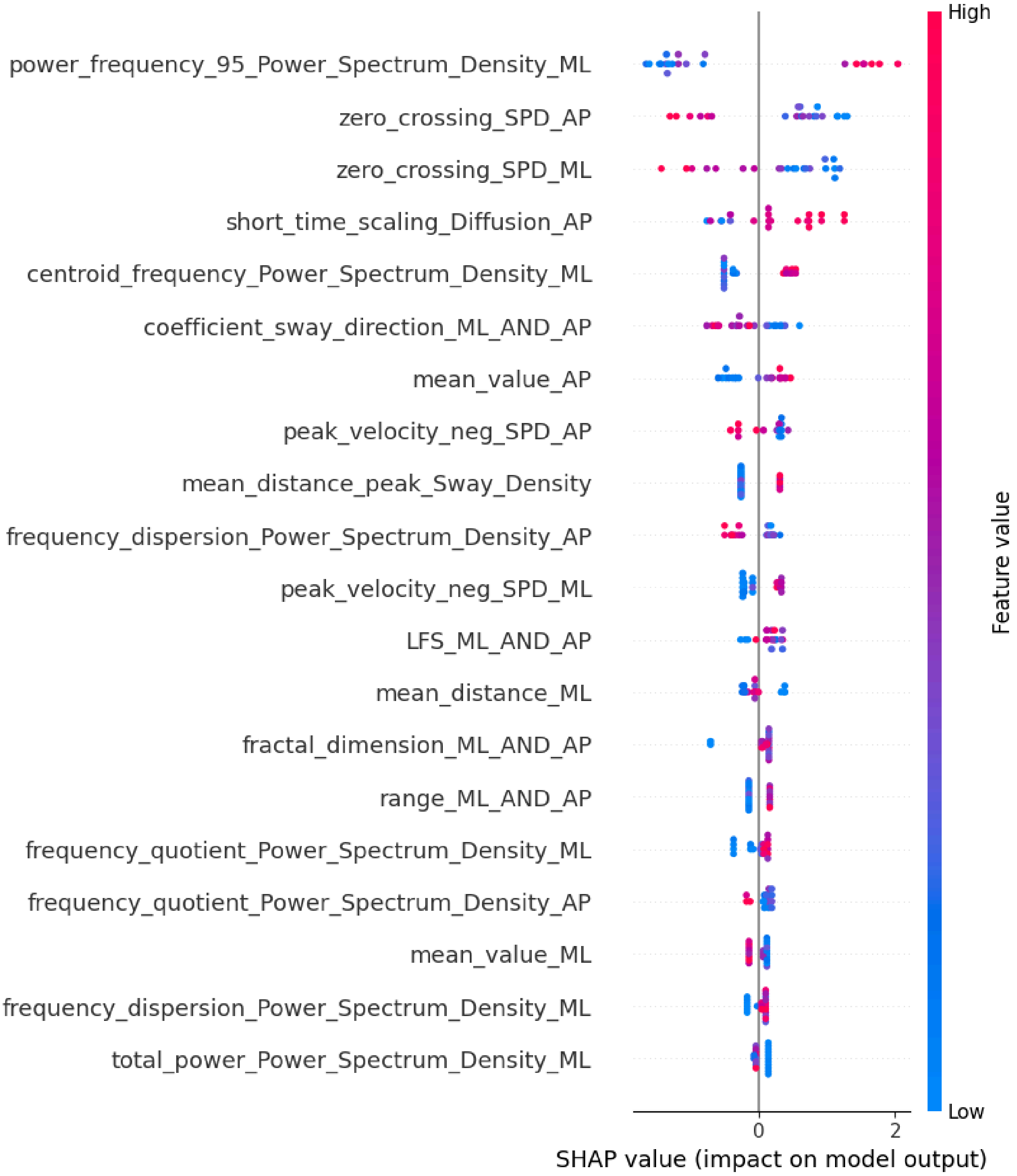
```
Results for split 2
Accuracy: 0.7619
Precision: 0.7500
Recall: 0.8182
F1 Score: 0.7826
```


Confusion Matrix
Actual 0
Actual 1
Predicted 0
Predicted 1
7
3
2
9
9
8
7
6
5
4
3
2

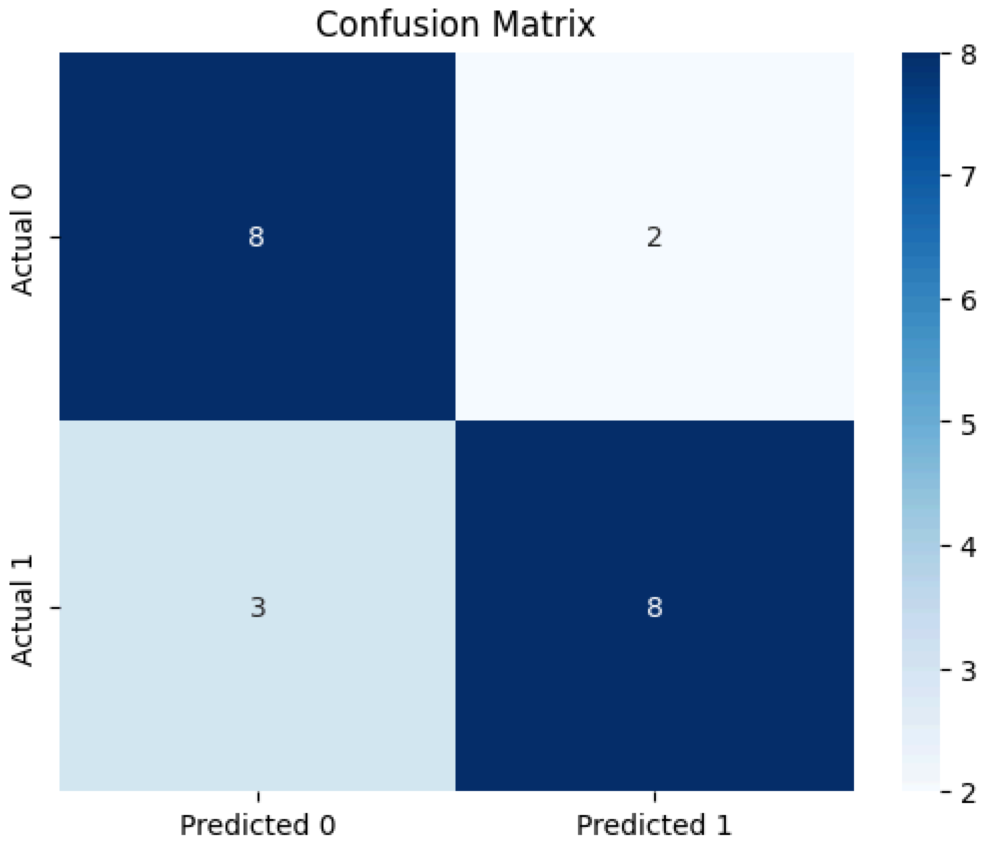


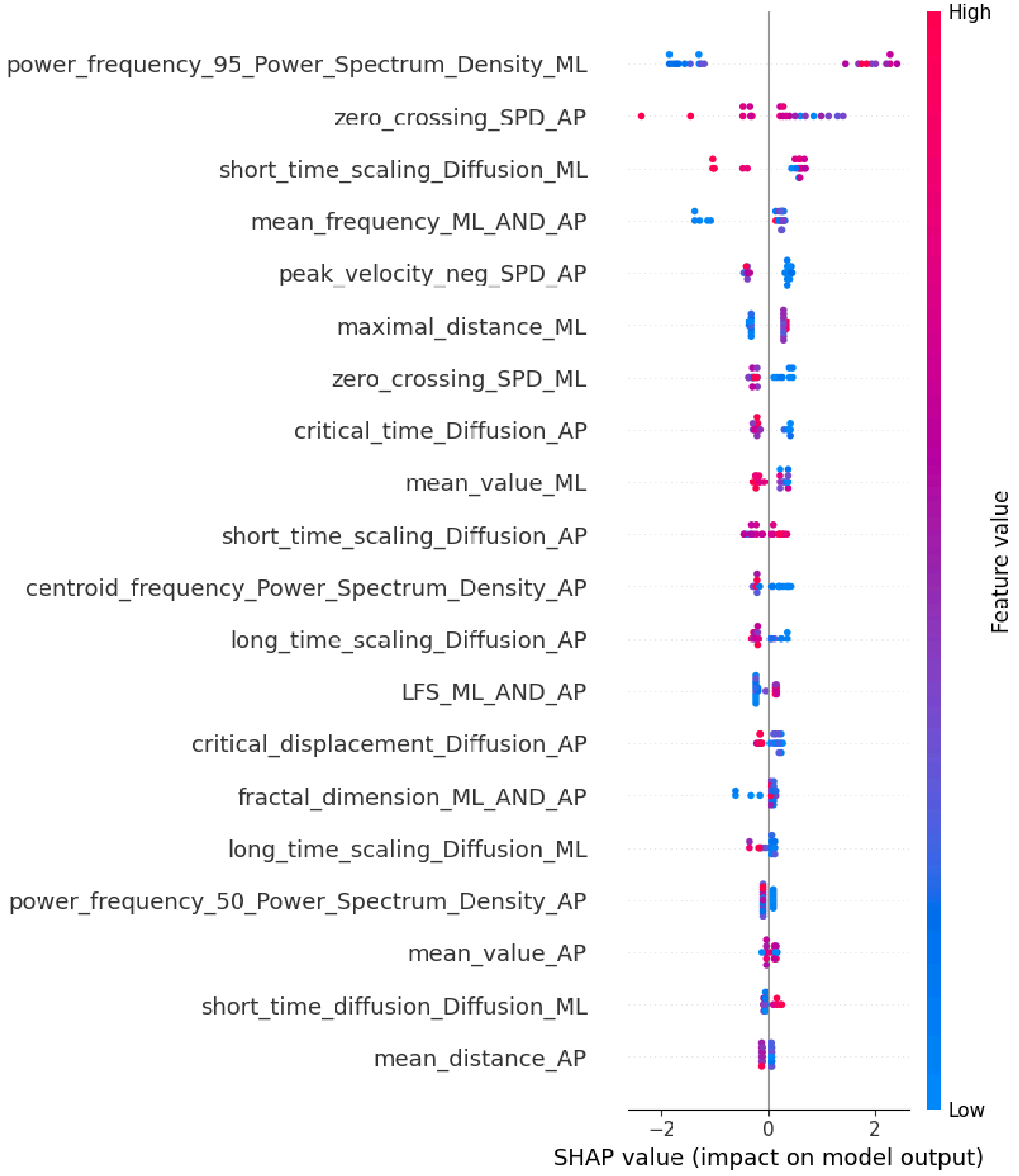
```
Results for split 3
Accuracy: 0.7619
Precision: 0.7500
Recall: 0.8182
F1 Score: 0.7826
```


Confusion Matrix
Actual 0
Actual 1
Predicted 0
Predicted 1
7
3
2
9
9
8
7
6
5
4
3
2

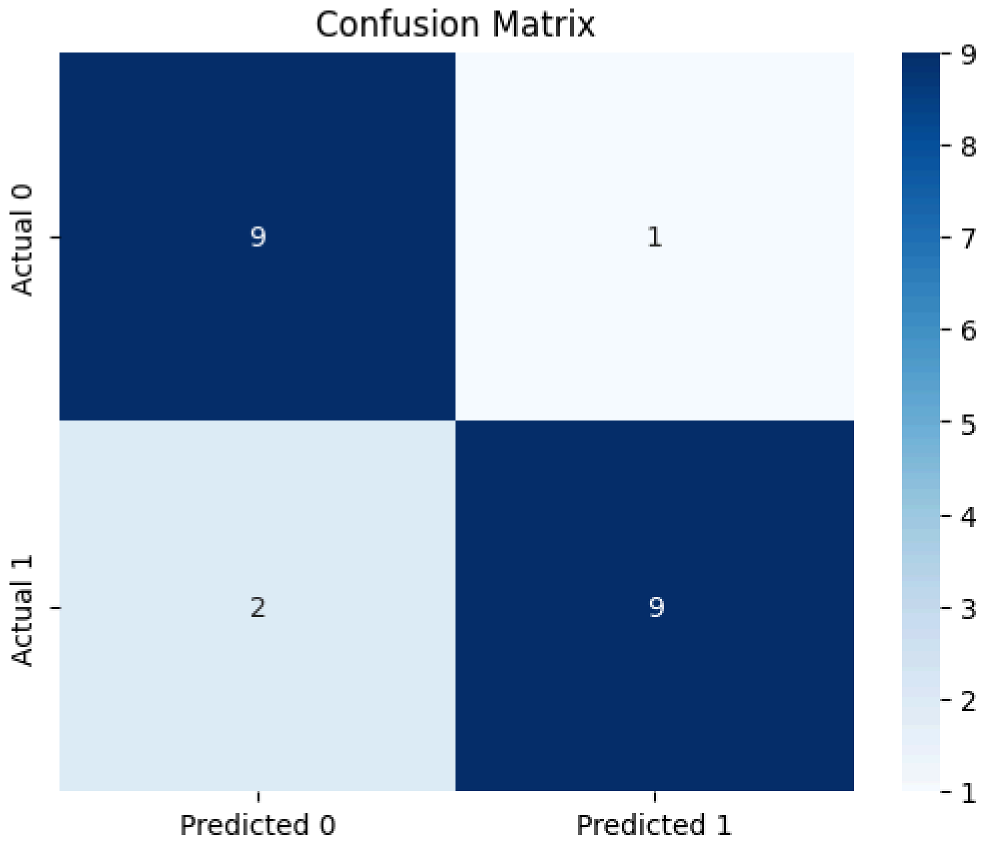


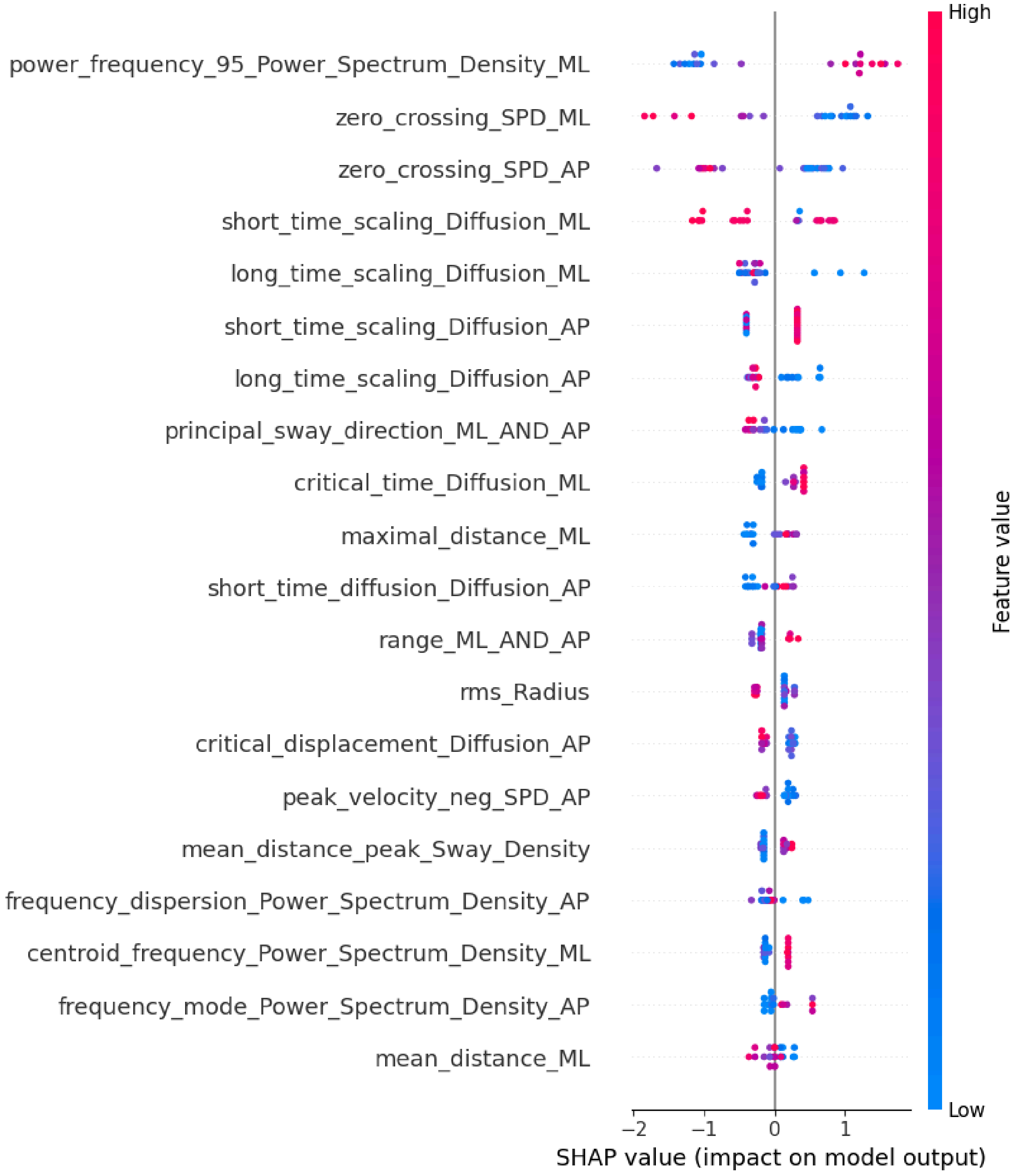
```
Results for split 4
Accuracy: 0.7619
Precision: 0.8000
Recall: 0.7273
F1 Score: 0.7619
```


Confusion Matrix
Actual 0
Actual 1
Predicted 0
Predicted 1
8
2
3
8
8
7
6
5
4
3
2



```
Results for split 5
Accuracy: 0.8571
Precision: 0.9000
Recall: 0.8182
F1 Score: 0.8571
```


Confusion Matrix
Actual 0
Actual 1
Predicted 0
Predicted 1
9
1
2
9
9
8
7
6
5
4
3
2
1


power_frequency_95_Power_Spectrum_Density_ML
zero_crossing_SPD_ML
zero_crossing_SPD_AP
short_time_scaling_Diffusion_ML
long_time_scaling_Diffusion_ML
short_time_scaling_Diffusion_AP
long_time_scaling_Diffusion_AP
principal_sway_direction_ML_AND_AP
critical_time_Diffusion_ML
maximal_distance_ML
short_time_diffusion_Diffusion_AP
range_ML_AND_AP
rms_Radius
critical_displacement_Diffusion_AP
peak_velocity_neg_SPD_AP
mean_distance_peak_Sway_Density
frequency_dispersion_Power_Spectrum_Density_AP
centroid_frequency_Power_Spectrum_Density_ML
frequency_mode_Power_Spectrum_Density_AP
mean_distance_ML
−2
−1
0
1
SHAP value (impact on model output)
High
Feature value
Low