Report | Assignment II - Classification

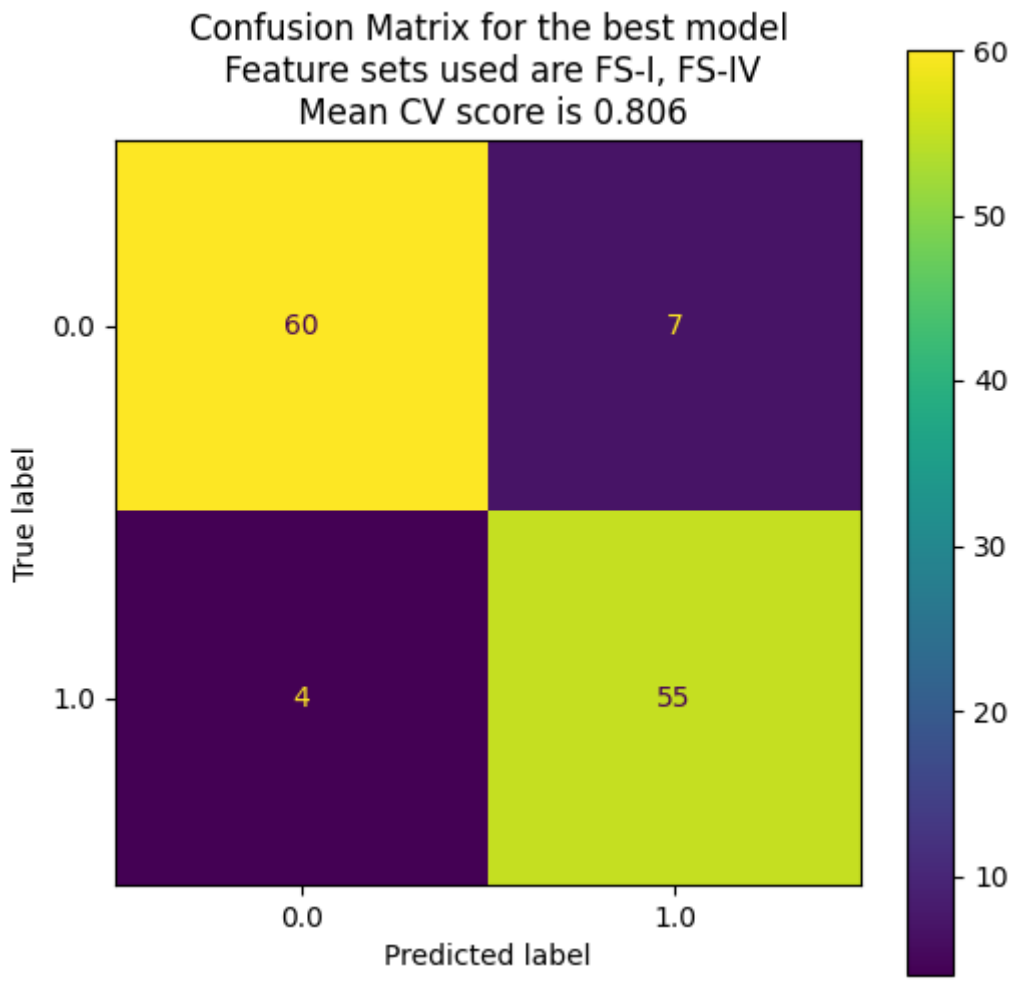
```
Average K-fold mean for each feature set combination (across multiple seeds):
FS-I: mean 0.752, std 0.069
FS-II: mean 0.625, std 0.034
FS-III: mean 0.0, std 0.0
FS-IV: mean 0.759, std 0.063
FS-I, FS-II: mean 0.746, std 0.103
FS-I, FS-III: mean 0.743, std 0.071
FS-I, FS-IV: mean 0.8, std 0.054
FS-II, FS-III: mean 0.619, std 0.033
FS-II, FS-IV: mean 0.759, std 0.088
FS-III, FS-IV: mean 0.762, std 0.058
FS-I, FS-II, FS-III: mean 0.759, std 0.099
FS-I, FS-II, FS-IV: mean 0.756, std 0.097
FS-I, FS-III, FS-IV: mean 0.8, std 0.054
FS-II, FS-III, FS-IV: mean 0.756, std 0.078
FS-I, FS-II, FS-III, FS-IV: mean 0.756, std 0.097
The best feature set is: FS-I, FS-IV
```

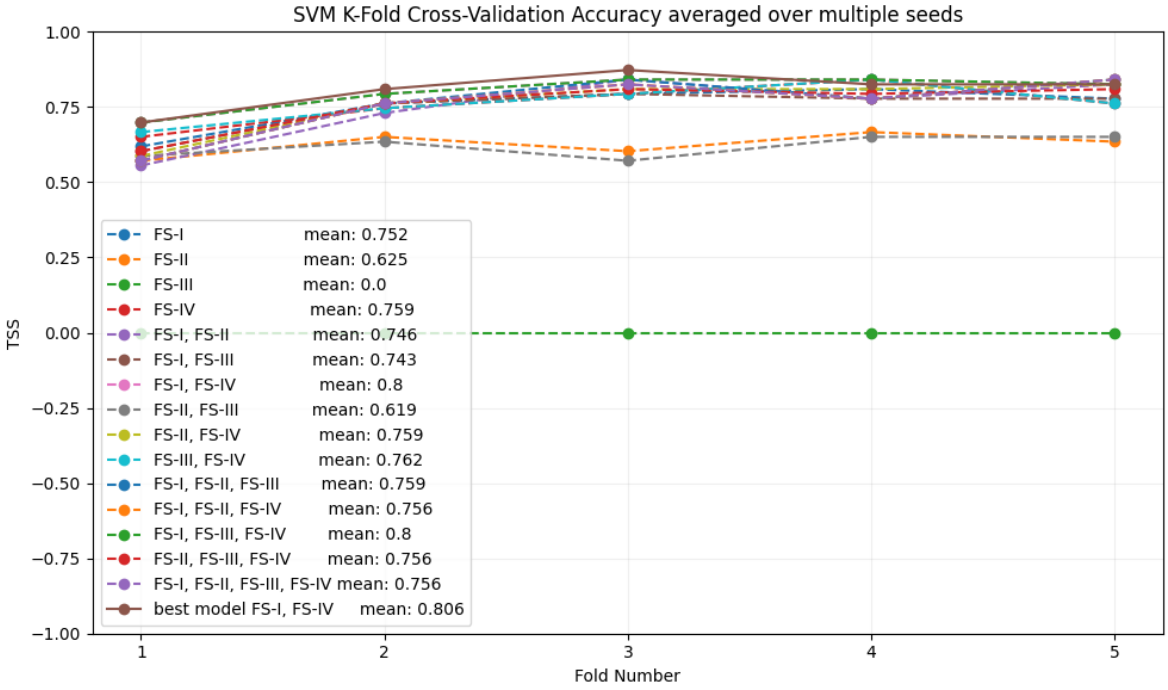
```
Cross-validation scores for best model using the data-2010-15 dataset:
[0.6984127  0.80952381 0.87301587 0.82539683 0.82539683]
Mean CV Score for best model using the data-2010-15 dataset: 0.8063
```

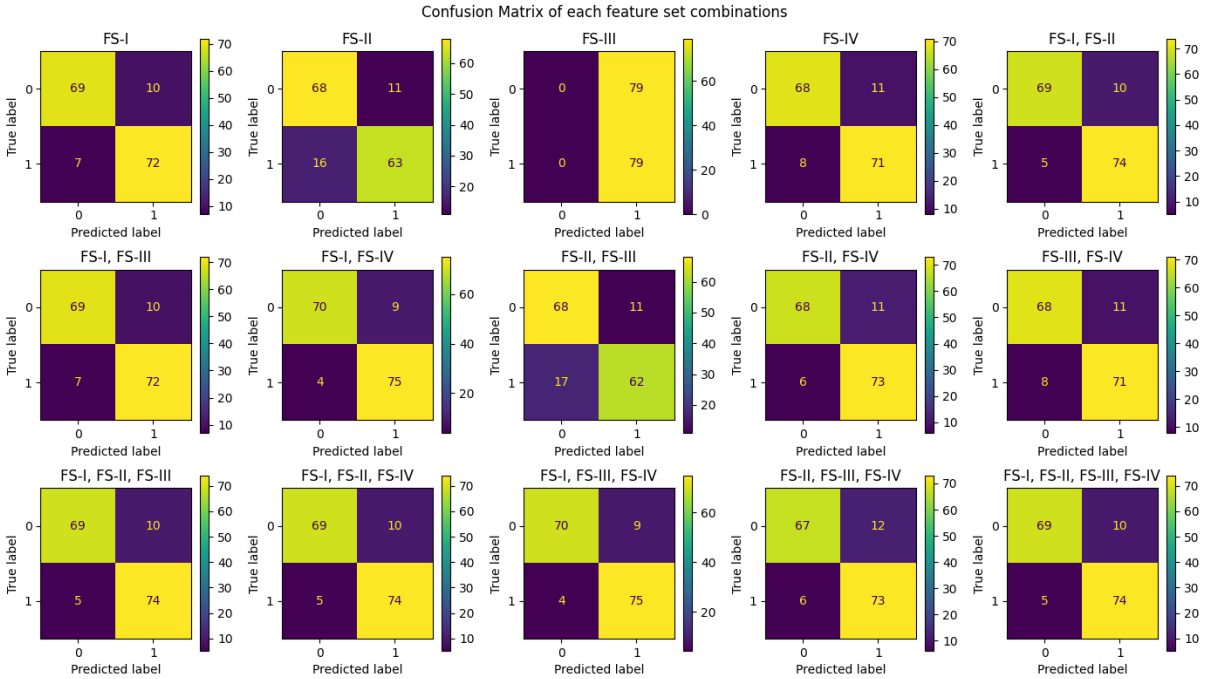
The results presented here are based on the provided data order. Additionally, we have included below the averaged K-Fold Cross-Validation Accuracy outcomes obtained from multiple runs, each initiated with a different random seed and with shuffling activated.

a. Our experiments revealed that the worst-performing feature set combination was FS-III alone, which yielded a True Skill Score (TSS) of 0 across all KFold iterations, indicating no predictive value. In contrast, the best combination was (FS-I, FS-IV), consisting of the Main Feature Set (FS-I) and Max Min Feature (FS-IV), with a TSS of 0.8. After hyperparameter tuning, the best model for this combination achieved an accuracy of 0.806.

b. Our benchmarking results show that adding FS-III to the model does not improve the TSS score. FS-III alone has a TSS of 0 across all KFold iterations, and combinations including it perform similarly to their equivalents without it. This suggests that FS-III does not contribute to the model's predictive power, (the feature is composed uniquely of zeros, indicating zero entropy and therefore zero information content). In contrast, adding FS-IV consistently improves the TSS score, indicating that it provides valuable information for predicting solar flare events. Even as a variation of FS-I, FS-IV slightly outperforms FS-I when used alone, or even when coupled with another feature set, such as (FS-II, FS-IV) which outperforms (FS-I, FS-II). However, the improvement is limited, as some models containing FS-I and not FS-IV still perform within 0.05 of the best model.


Confusion Matrix for the best model
Feature sets used are FS-I, FS-IV
Mean CV score is 0.806
True label
Predicted label
0.0
1.0
60
7
4
55
60
50
40
30
20
10

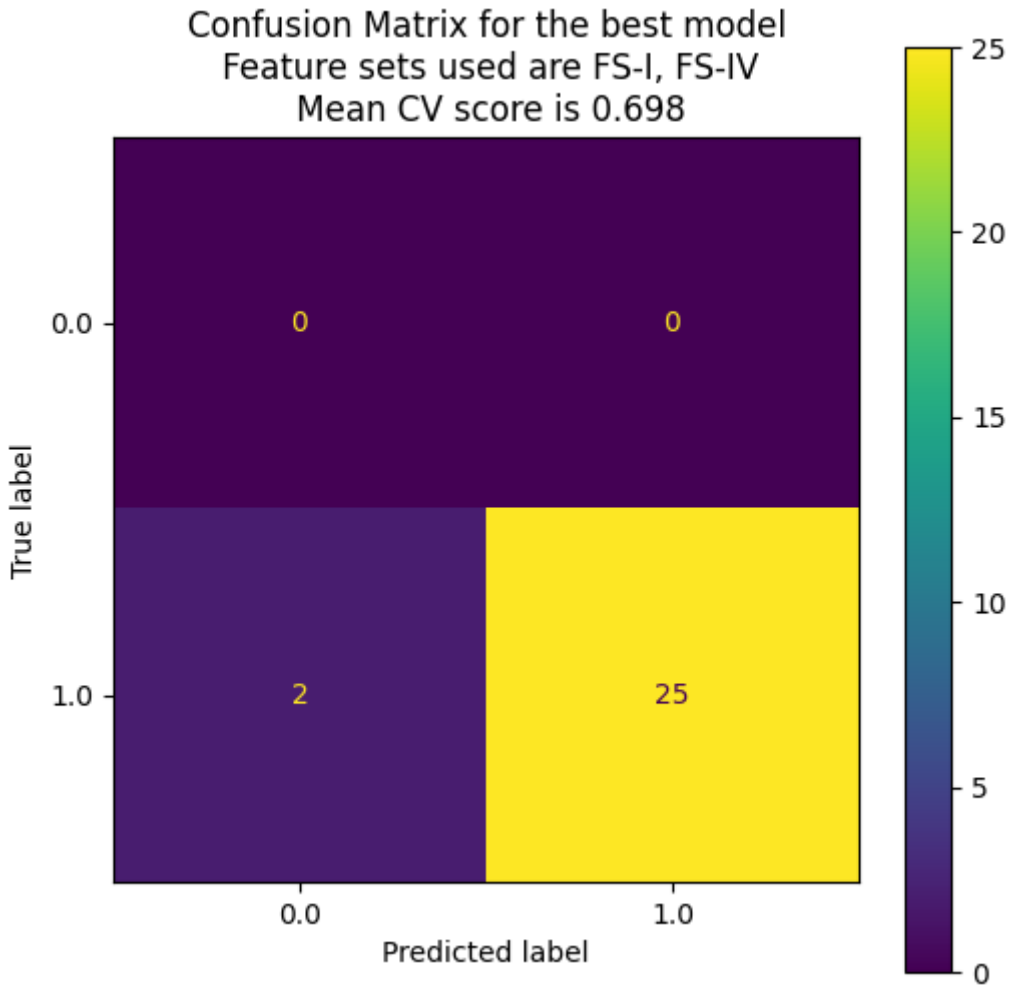

SVM K-Fold Cross-Validation Accuracy averaged over multiple seeds
TSS
Fold Number
FS-I mean: 0.752
FS-II mean: 0.625
FS-III mean: 0.0
FS-IV mean: 0.759
FS-I, FS-II mean: 0.746
FS-I, FS-III mean: 0.743
FS-I, FS-IV mean: 0.8
FS-II, FS-III mean: 0.619
FS-II, FS-IV mean: 0.759
FS-III, FS-IV mean: 0.762
FS-I, FS-II, FS-III mean: 0.759
FS-I, FS-II, FS-IV mean: 0.756
FS-I, FS-III, FS-IV mean: 0.8
FS-II, FS-III, FS-IV mean: 0.756
FS-I, FS-II, FS-III, FS-IV mean: 0.756
best model FS-I, FS-IV mean: 0.806

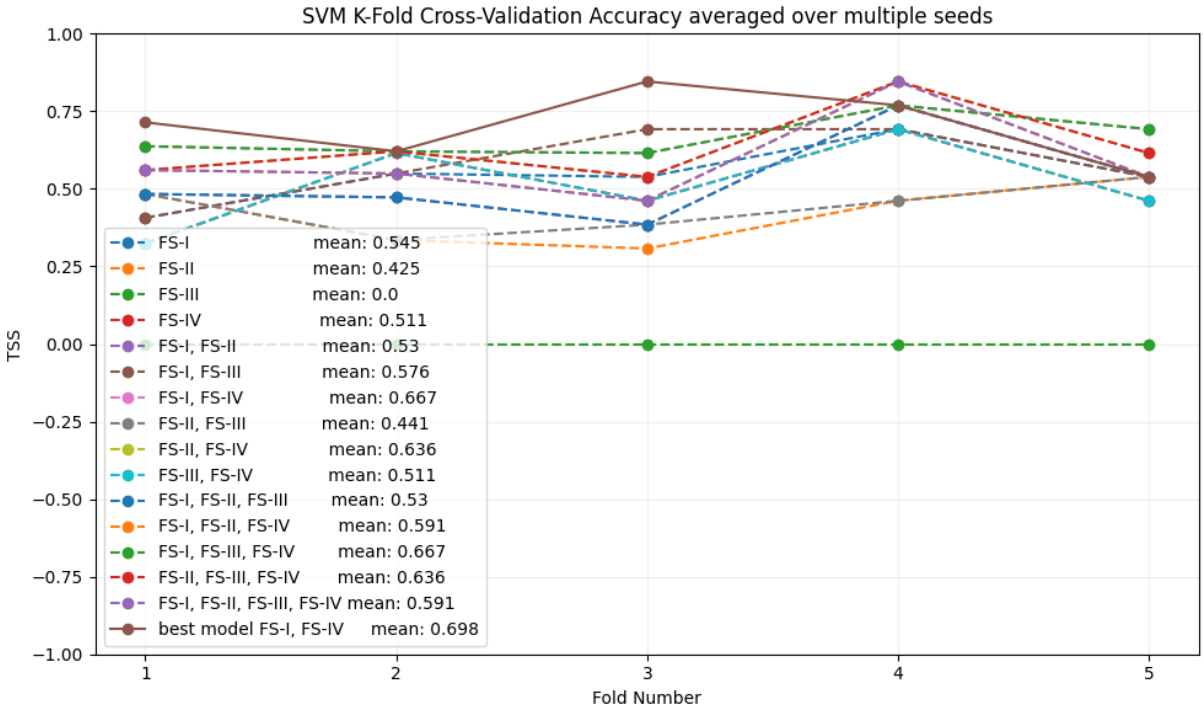


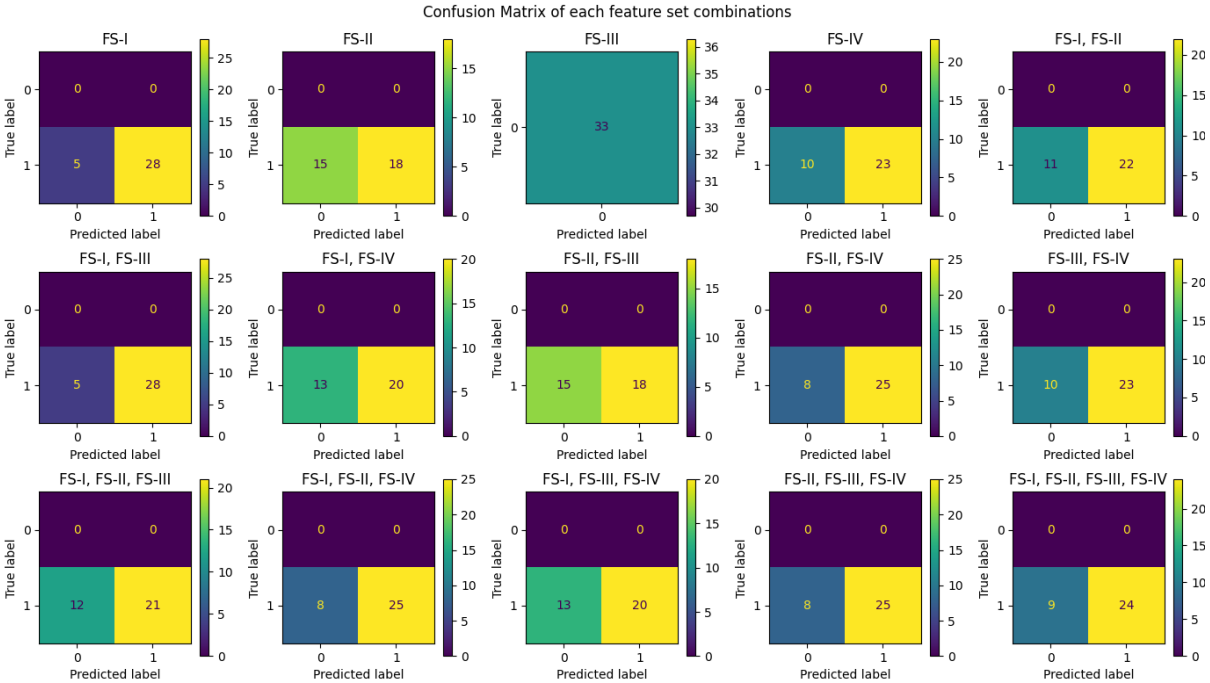
c. We observed that the 2010 dataset yielded a better TSS score of 0.806 compared to the 2020 dataset's score of 0.698. This difference may be attributed to the number of data samples in each dataset. The 2010 dataset has 315 samples per class, whereas the 2020 dataset has only 66 samples per class. This also explains the high fluctuation we can observe in the score of the different folds, ranging from 0.538 to 0.928 for the best model, which is less stable than for the 2010 dataset. This disparity could be due to reduced solar activity during the 2020 dataset's time period or an imbalanced representation of both events, leading to a reduced number of vectors to maintain balance. In addition, we can see that there are only samples from the positive class at the time of prediction. Furthermore, as this represents a significant proportion of them (27 out of 66), it means that the 66 samples of the negative class and the remaining 39 in the positive class are in the training set. This creates an imbalance in the training set which can lead to suboptimal performances.

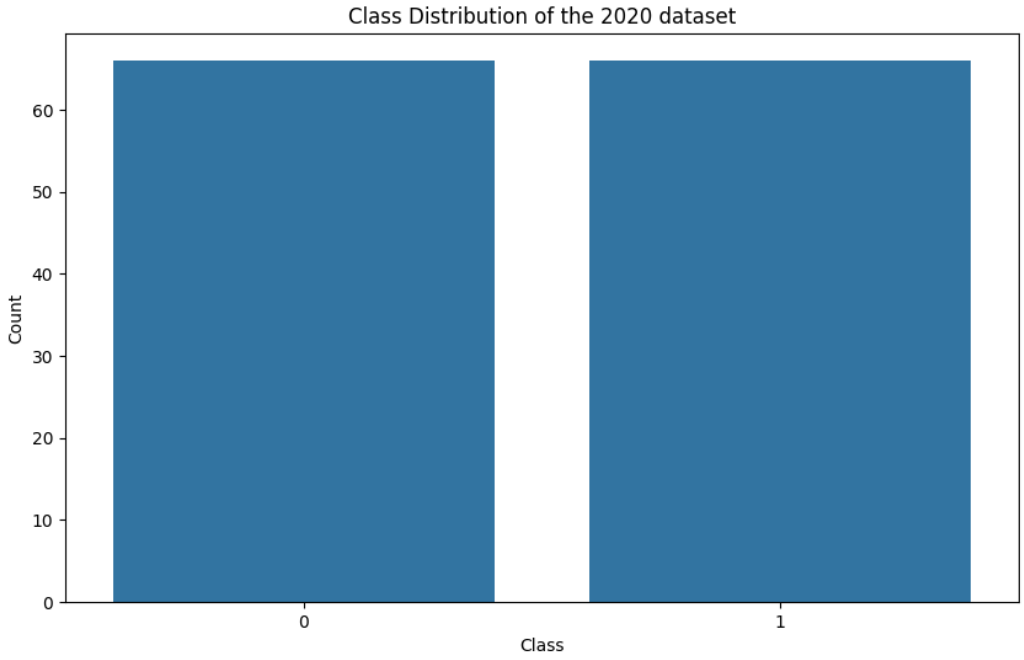
Results for the 2020 dataset

```
Cross-validation scores for best model using the data-2020-24 dataset:
[0.71428571 0.62087912 0.84615385 0.76923077 0.53846154]
Mean CV Score for best model using the data-2020-24 dataset: 0.6978
```


Confusion Matrix for the best model
Feature sets used are FS-I, FS-IV
Mean CV score is 0.698
True label
Predicted label
0.0
1.0
0
0
2
25
25
20
15
10
5
0


SVM K-Fold Cross-Validation Accuracy averaged over multiple seeds
TSS
Fold Number
FS-I mean: 0.545
FS-II mean: 0.425
FS-III mean: 0.0
FS-IV mean: 0.511
FS-I, FS-II mean: 0.53
FS-I, FS-III mean: 0.576
FS-I, FS-IV mean: 0.667
FS-II, FS-III mean: 0.441
FS-II, FS-IV mean: 0.636
FS-III, FS-IV mean: 0.511
FS-I, FS-II, FS-III mean: 0.53
FS-I, FS-II, FS-IV mean: 0.591
FS-I, FS-III, FS-IV mean: 0.667
FS-II, FS-III, FS-IV mean: 0.636
FS-I, FS-II, FS-III, FS-IV mean: 0.591
best model FS-I, FS-IV mean: 0.698

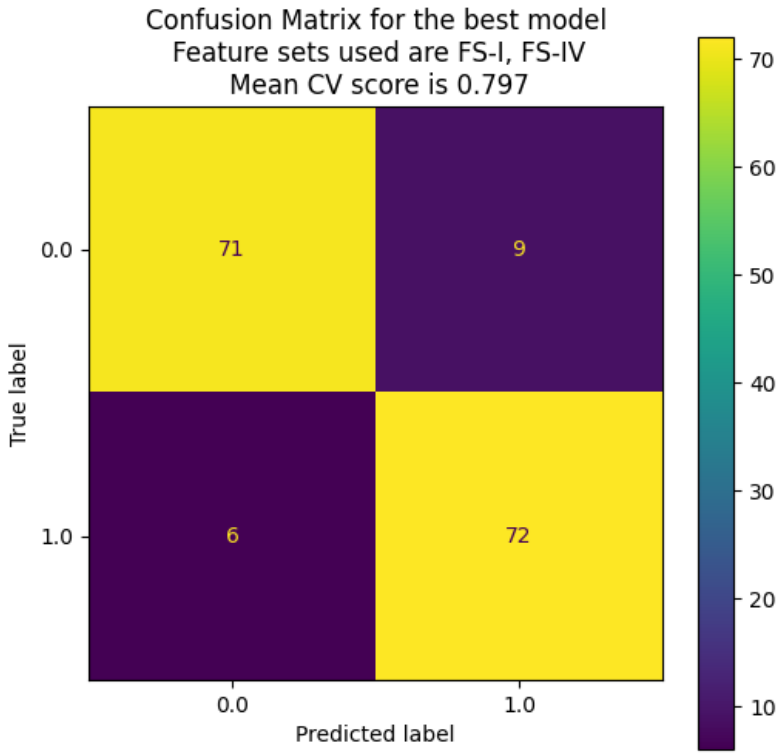

Confusion Matrix of each feature set combinations
FS-I
FS-II
FS-III
FS-IV
FS-I, FS-II
FS-I, FS-III
FS-I, FS-IV
FS-II, FS-III
FS-II, FS-IV
FS-III, FS-IV
FS-I, FS-II, FS-III
FS-I, FS-II, FS-IV
FS-I, FS-III, FS-IV
FS-II, FS-III, FS-IV
FS-I, FS-II, FS-III, FS-IV
True label
Predicted label

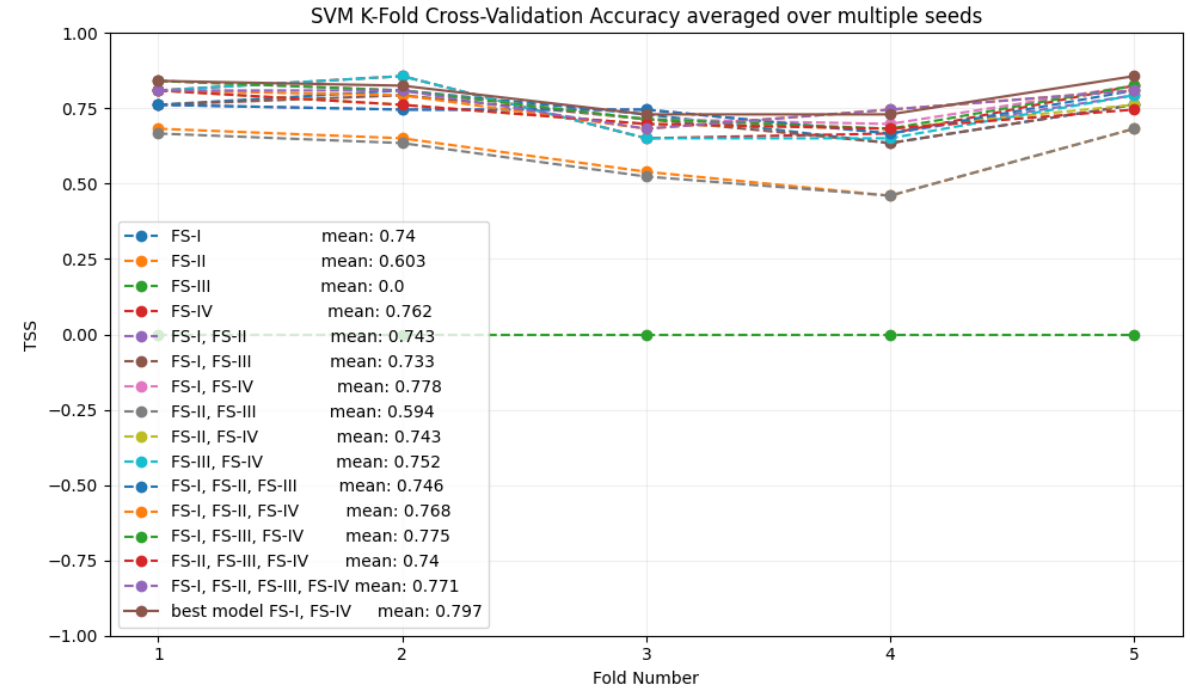

Class Distribution of the 2020 dataset
Count
Class

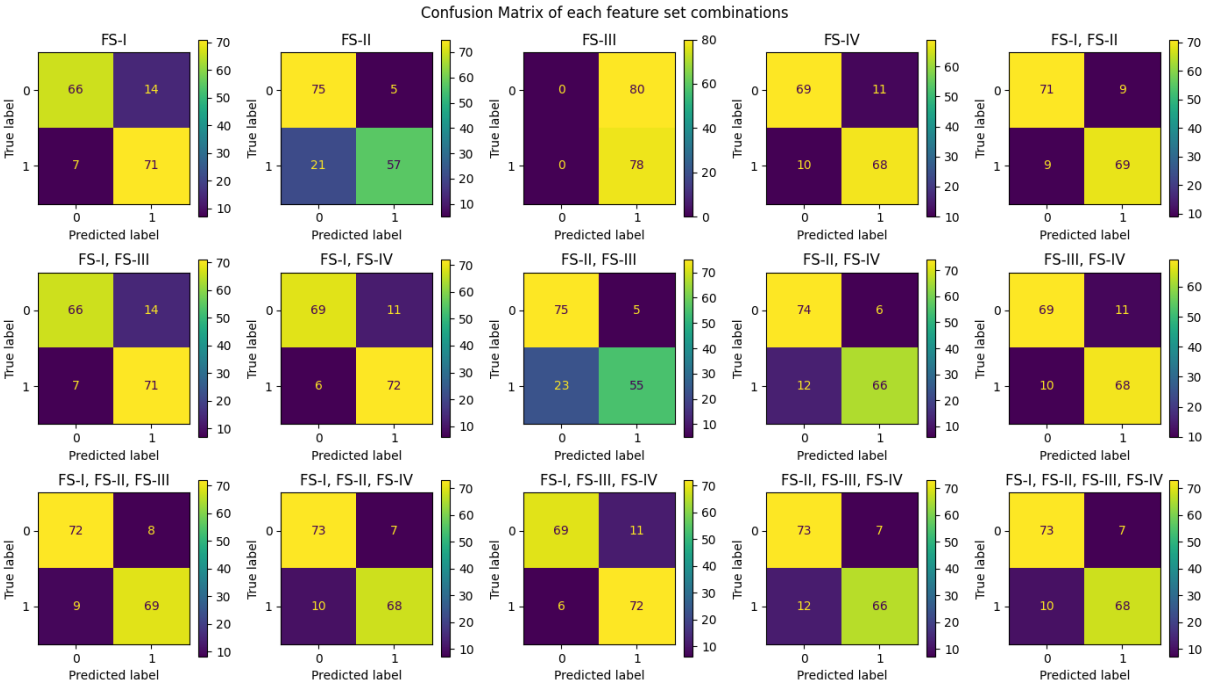
Results for averaged K-fold scores using shuffled data and random state

2010 Dataset

```
Average K-fold mean for each feature set combination (across multiple seeds):
FS-I: mean 0.74, std 0.058
FS-II: mean 0.603, std 0.089
FS-III: mean 0.0, std 0.0
FS-IV: mean 0.762, std 0.086
FS-I, FS-II: mean 0.743, std 0.042
FS-I, FS-III: mean 0.733, std 0.055
FS-I, FS-IV: mean 0.778, std 0.059
FS-II, FS-III: mean 0.594, std 0.087
FS-II, FS-IV: mean 0.743, std 0.046
FS-III, FS-IV: mean 0.752, std 0.086
FS-I, FS-II, FS-III: mean 0.746, std 0.046
FS-I, FS-II, FS-IV: mean 0.768, std 0.049
FS-I, FS-III, FS-IV: mean 0.775, std 0.064
FS-II, FS-III, FS-IV: mean 0.74, std 0.046
FS-I, FS-II, FS-III, FS-IV: mean 0.771, std 0.051
The best feature set is: FS-I, FS-IV
```

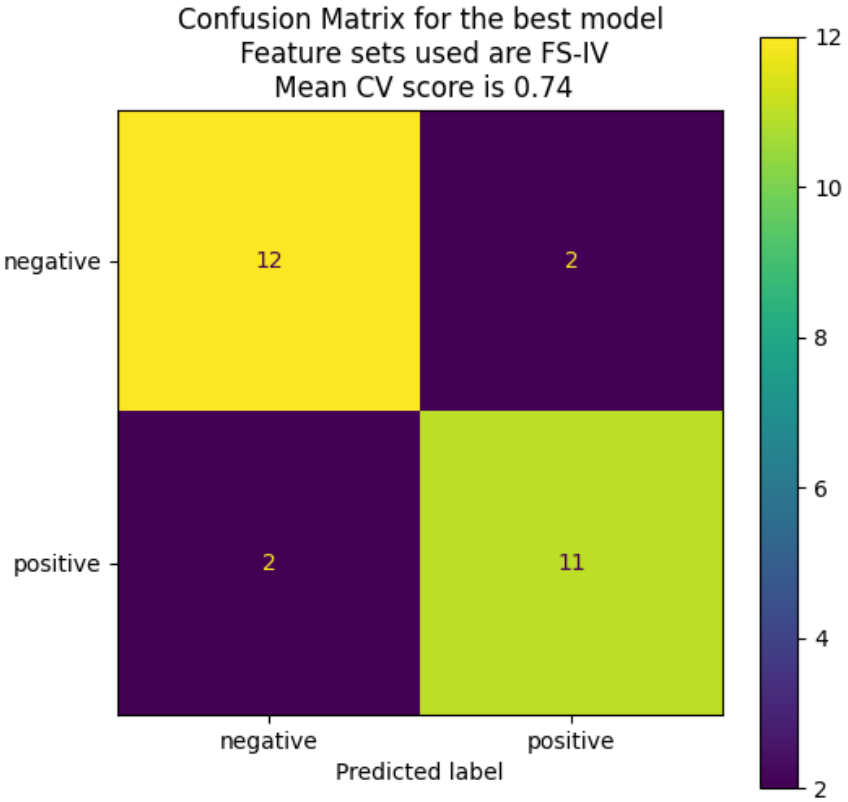


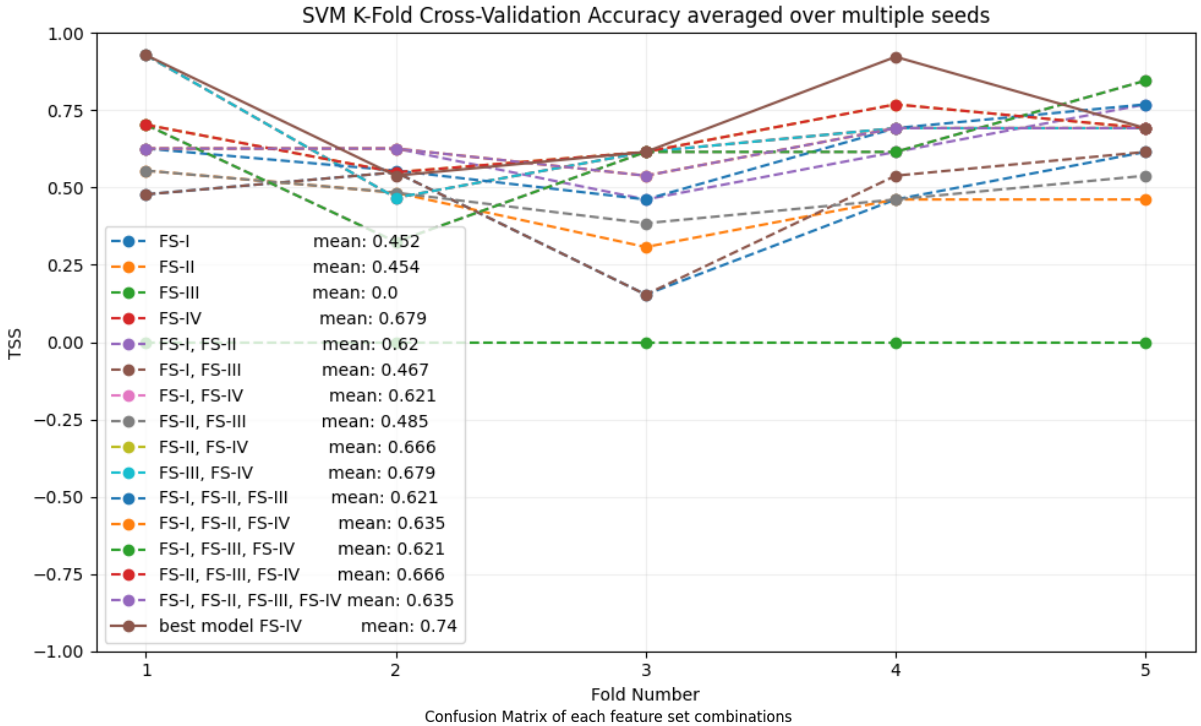

SVM K-Fold Cross-Validation Accuracy averaged over multiple seeds
TSS
Fold Number
FS-I mean: 0.74
FS-II mean: 0.603
FS-III mean: 0.0
FS-IV mean: 0.762
FS-I, FS-II mean: 0.743
FS-I, FS-III mean: 0.733
FS-I, FS-IV mean: 0.778
FS-II, FS-III mean: 0.594
FS-II, FS-IV mean: 0.743
FS-III, FS-IV mean: 0.752
FS-I, FS-II, FS-III mean: 0.746
FS-I, FS-II, FS-IV mean: 0.768
FS-I, FS-III, FS-IV mean: 0.775
FS-II, FS-III, FS-IV mean: 0.74
FS-I, FS-II, FS-III, FS-IV mean: 0.771
best model FS-I, FS-IV mean: 0.797

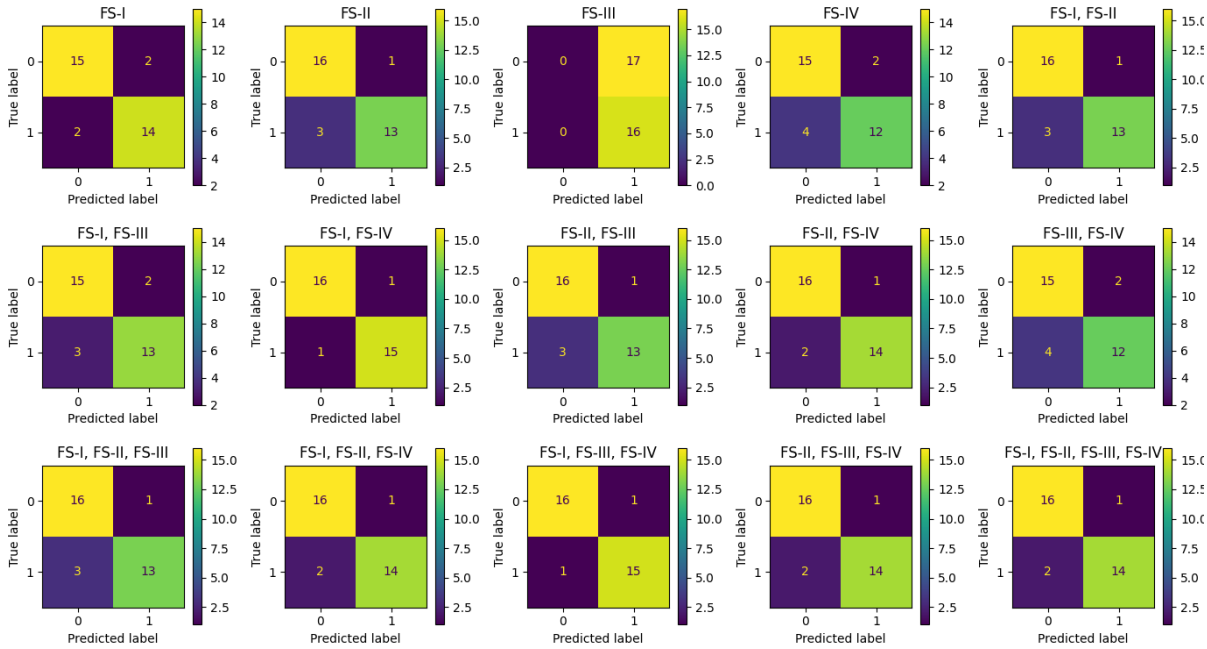

Confusion Matrix of each feature set combinations
FS-I
FS-II
FS-III
FS-IV
FS-I, FS-II
FS-I, FS-III
FS-I, FS-IV
FS-II, FS-III
FS-II, FS-IV
FS-III, FS-IV
FS-I, FS-II, FS-III
FS-I, FS-II, FS-IV
FS-I, FS-III, FS-IV
FS-II, FS-III, FS-IV
FS-I, FS-II, FS-III, FS-IV
True label
Predicted label

2020 Dataset

```
Average K-fold mean for each feature set combination (across multiple seeds):
FS-I: mean 0.452, std 0.159
FS-II: mean 0.454, std 0.081
FS-III: mean 0.0, std 0.0
FS-IV: mean 0.679, std 0.149
FS-I, FS-II: mean 0.62, std 0.097
FS-I, FS-III: mean 0.467, std 0.163
FS-I, FS-IV: mean 0.621, std 0.171
FS-II, FS-III: mean 0.485, std 0.061
FS-II, FS-IV: mean 0.666, std 0.076
FS-III, FS-IV: mean 0.679, std 0.149
FS-I, FS-II, FS-III: mean 0.621, std 0.107
FS-I, FS-II, FS-IV: mean 0.635, std 0.057
FS-I, FS-III, FS-IV: mean 0.621, std 0.171
FS-II, FS-III, FS-IV: mean 0.666, std 0.076
FS-I, FS-II, FS-III, FS-IV: mean 0.635, std 0.057
The best feature set is: FS-IV
```




SVM K-Fold Cross-Validation Accuracy averaged over multiple seeds
TSS
Fold Number
FS-I mean: 0.452
FS-II mean: 0.454
FS-III mean: 0.0
FS-IV mean: 0.679
FS-I, FS-II mean: 0.62
FS-I, FS-III mean: 0.467
FS-I, FS-IV mean: 0.621
FS-II, FS-III mean: 0.485
FS-II, FS-IV mean: 0.666
FS-III, FS-IV mean: 0.679
FS-I, FS-II, FS-III mean: 0.621
FS-I, FS-II, FS-IV mean: 0.635
FS-I, FS-III, FS-IV mean: 0.621
FS-II, FS-III, FS-IV mean: 0.666
FS-I, FS-II, FS-III, FS-IV mean: 0.635
best model FS-IV mean: 0.74


Confusion Matrix of each feature set combinations
FS-I
FS-II
FS-III
FS-IV
FS-I, FS-II
FS-I, FS-III
FS-I, FS-IV
FS-II, FS-III
FS-II, FS-IV
FS-III, FS-IV
FS-I, FS-II, FS-III
FS-I, FS-II, FS-IV
FS-I, FS-III, FS-IV
FS-II, FS-III, FS-IV
FS-I, FS-II, FS-III, FS-IV
True label
Predicted label