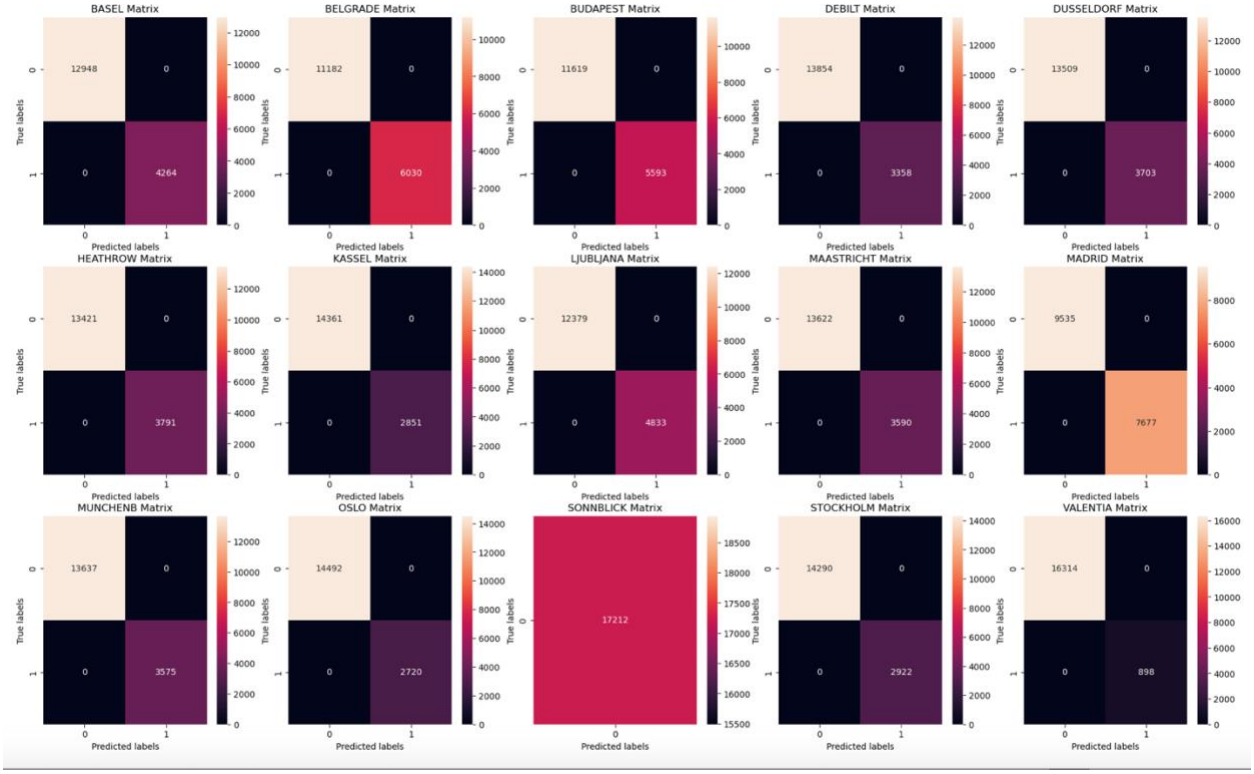
Amrit Gill

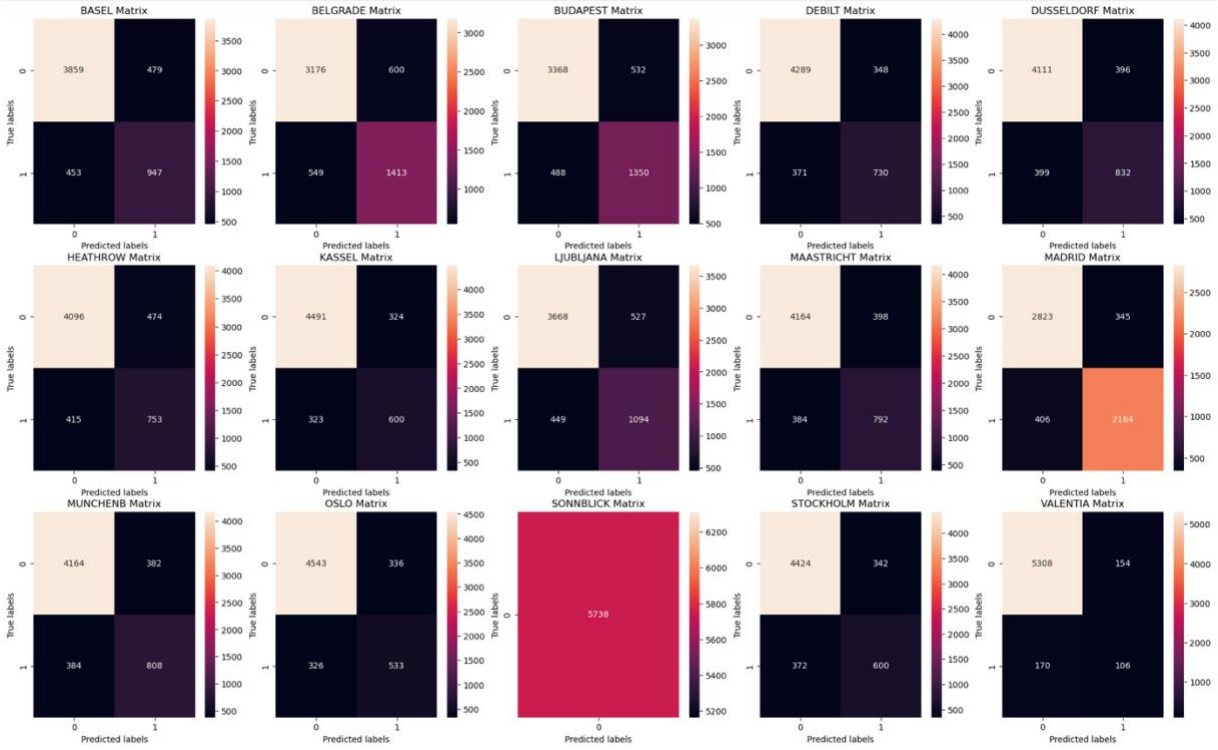
# Exercise 1.5: Supervised Learning Algorithms Part 2

The tree decision model currently illustrates poor accuracy, with only 46% on the training data and 47% on the test data. This minimal difference suggests that the model is not significantly overfitting, but its overall performance remains poor. One likely reason for this is that the tree has become too deep and complex, making it difficult to generalize patterns effectively. To address this issue, pruning is necessary to remove unnecessary branches and simplify the model. By reducing the depth of the tree, pruning can enhance generalization, improve accuracy, and make the model more interpretable.

Decision Tree Training Set:



Decision Tree Test Set:

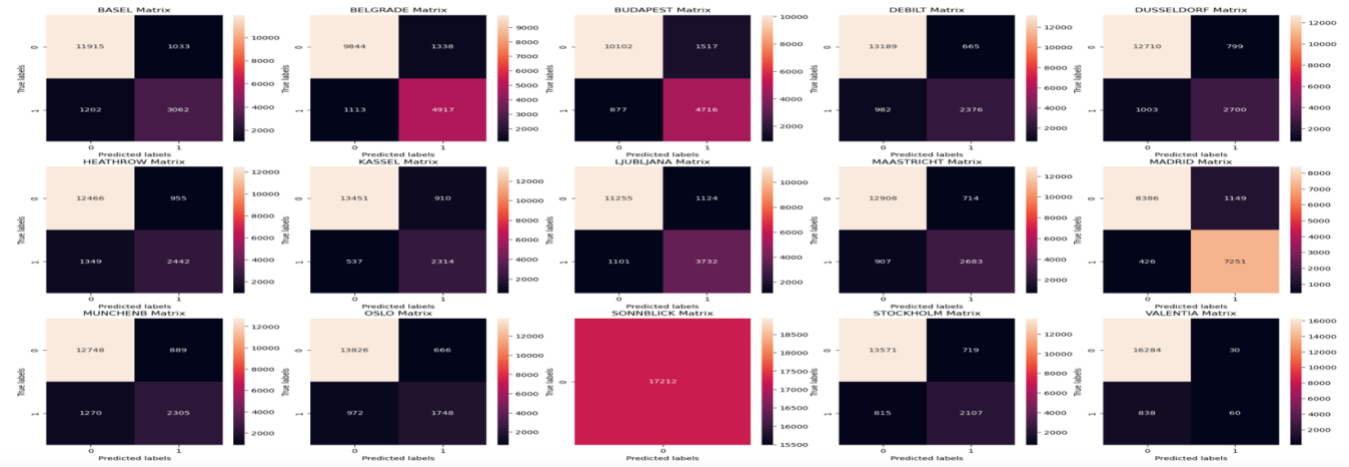



ANN Models:

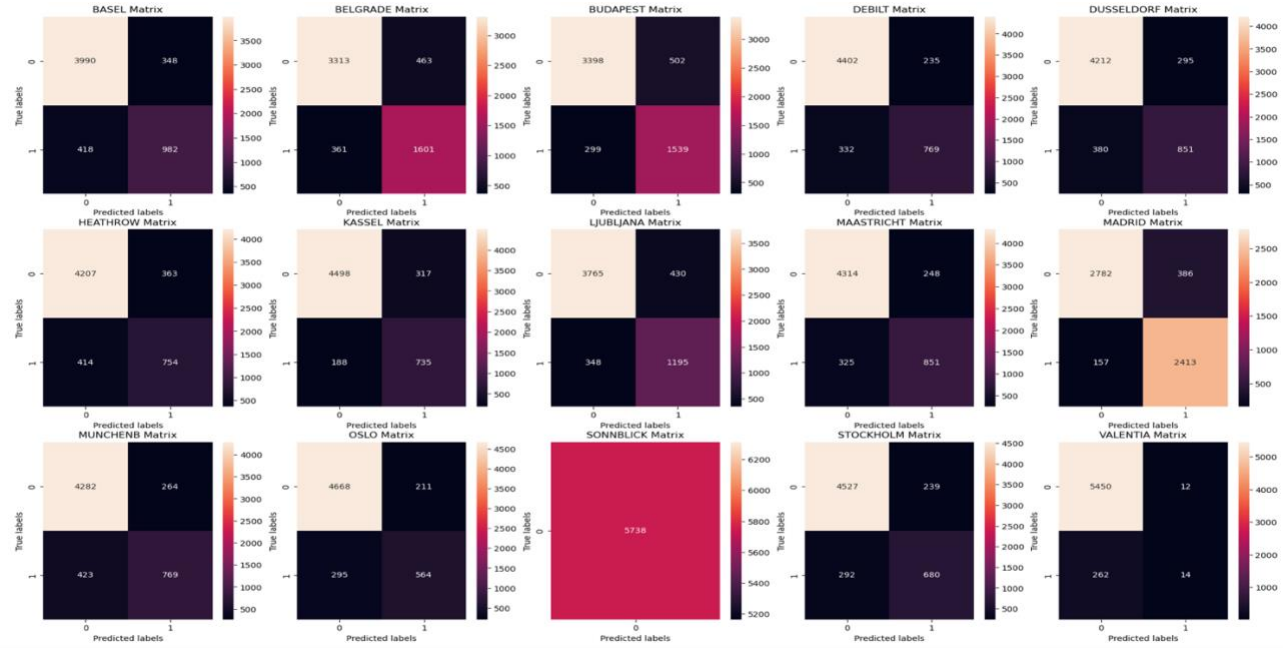
Scenario 1:

```
MLPClassifier
MLPClassifier(hidden_layer_sizes=(10, 5), max_iter=500)
```

Training Data Set: Accuracy for this set is 47%.



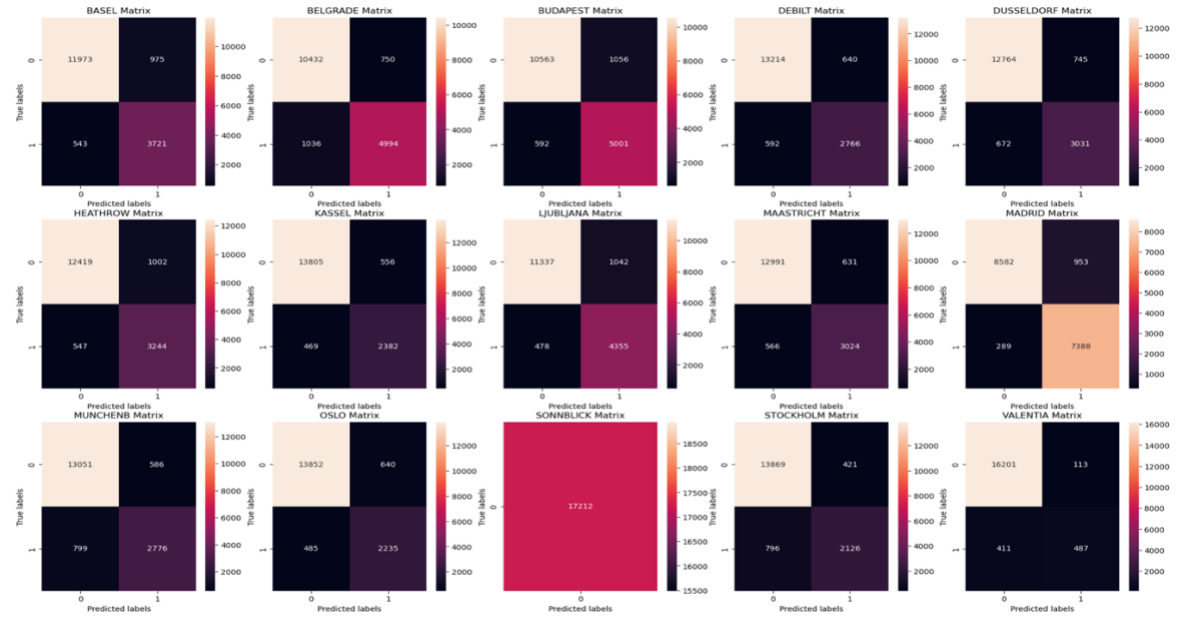
Test Data Set: Accuracy for this set is 47%.



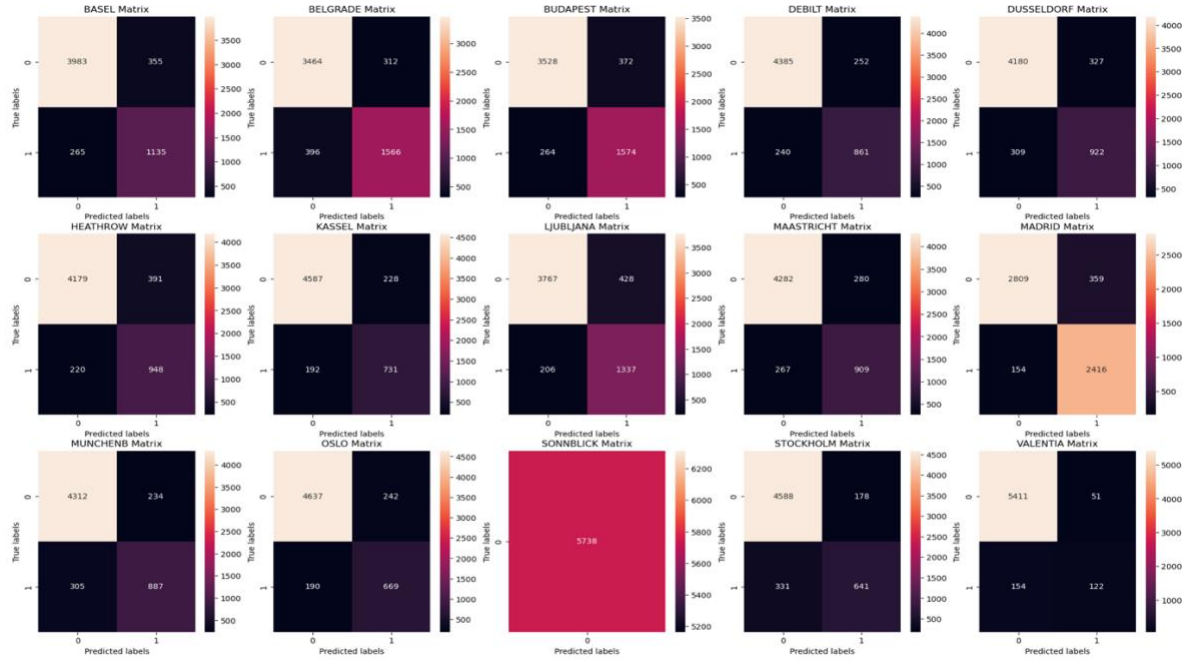

Scenario 2:

```
MLPClassifier
MLPClassifier(hidden_layer_sizes=(70, 60, 60), max_iter=1000)
```

Training Data Set: Accuracy is 54%.



Test Data Set: Accuracy for this set is 49%.

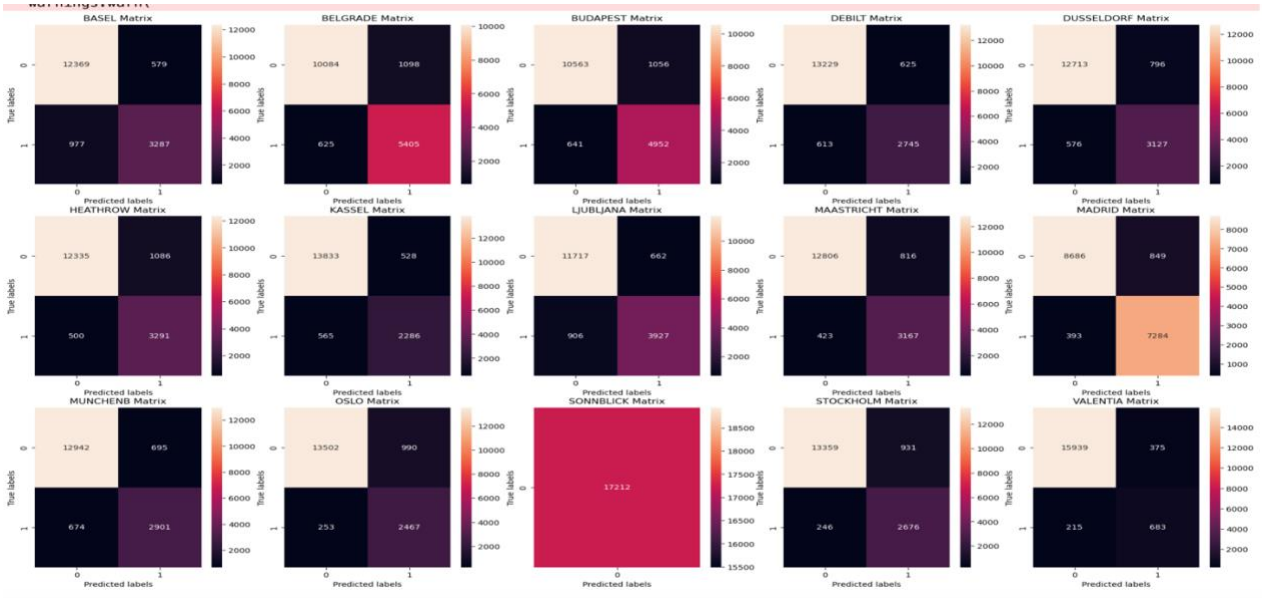



Scenario 3:

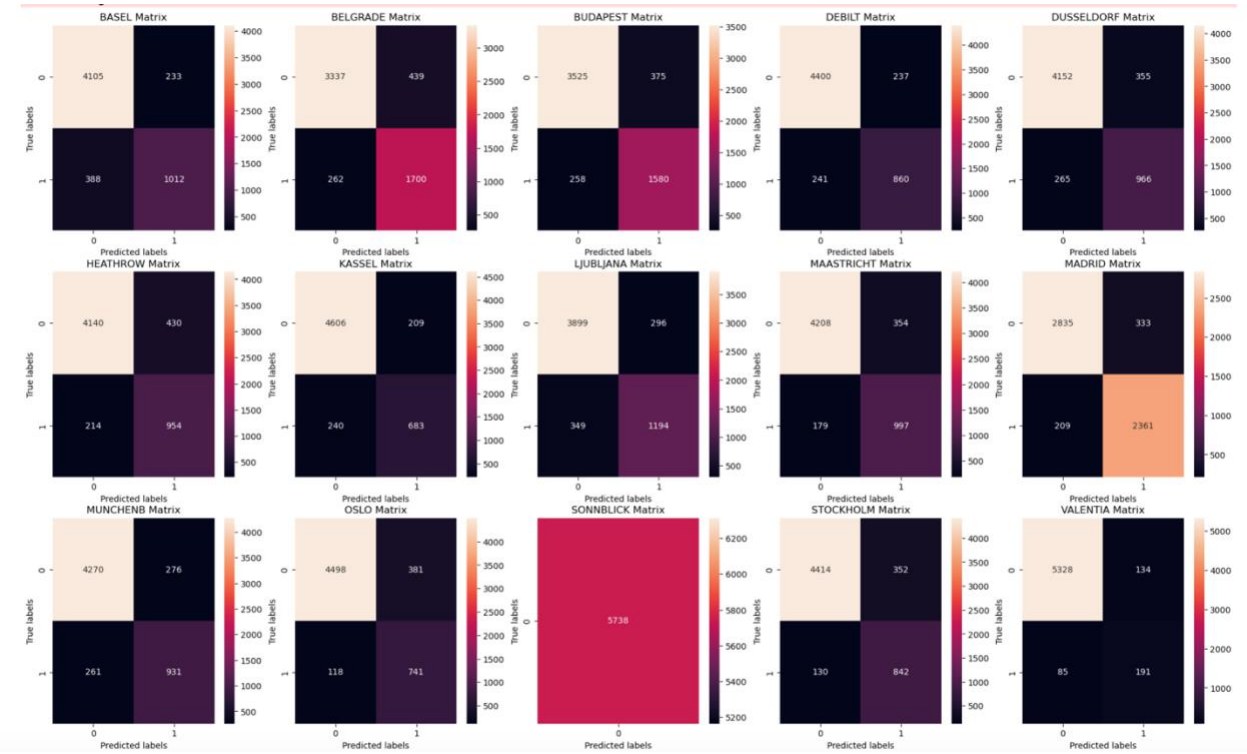
MLPClassifier

```
MLPClassifier(hidden_layer_sizes=(70, 60, 60), max_iter=1000, tol=0.0003)
```

Training Data Set: Accuracy is 53%.



Test Data Set: Accuracy is 48%.




K-Nearest Neighbors (KNN) is the most effective algorithm for ClimateWins to provide accurate predictions of good picnic weather since it effectively identifies patterns and makes reliable predictions.

The Sonnblick station models are probably overfitting. They achieve 100% accuracy by always predicting bad weather, a misleading result. There are too few instances of good weather in the data to suggest that the model is unable to recognize diverse weather patterns effectively. Thus, it may fail on new situations, padding accuracy without enhancing actual weather prediction.