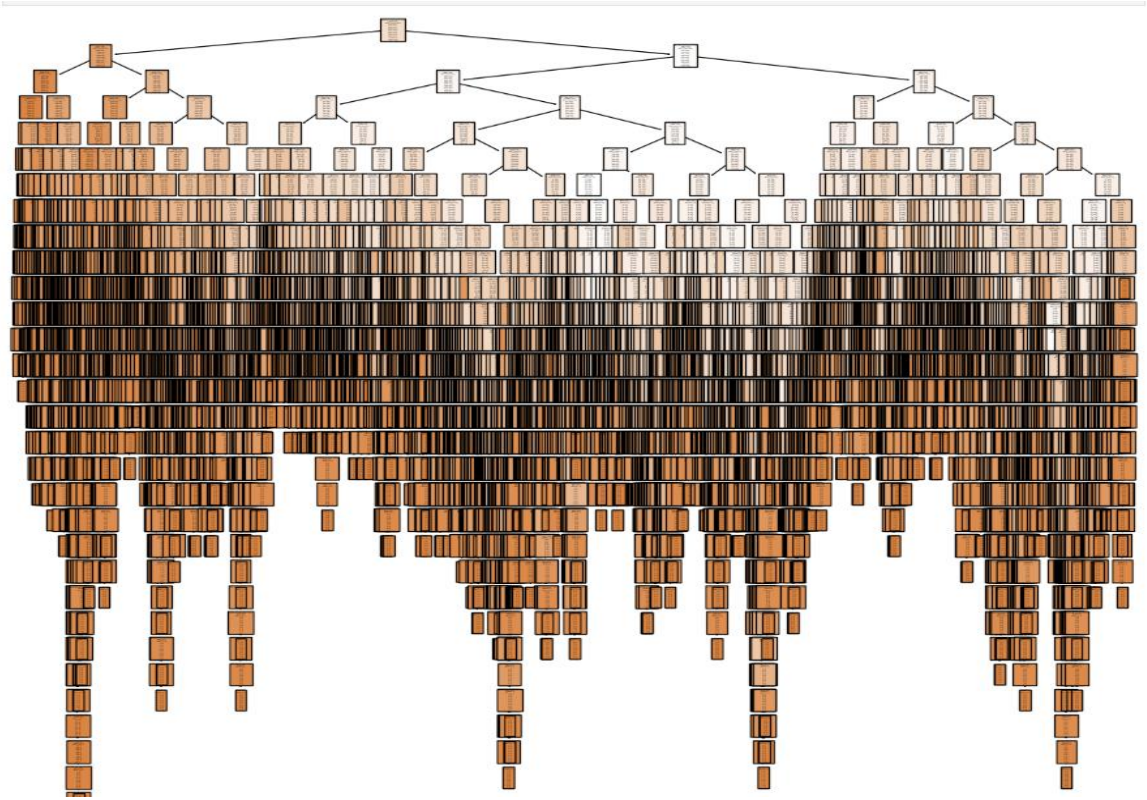
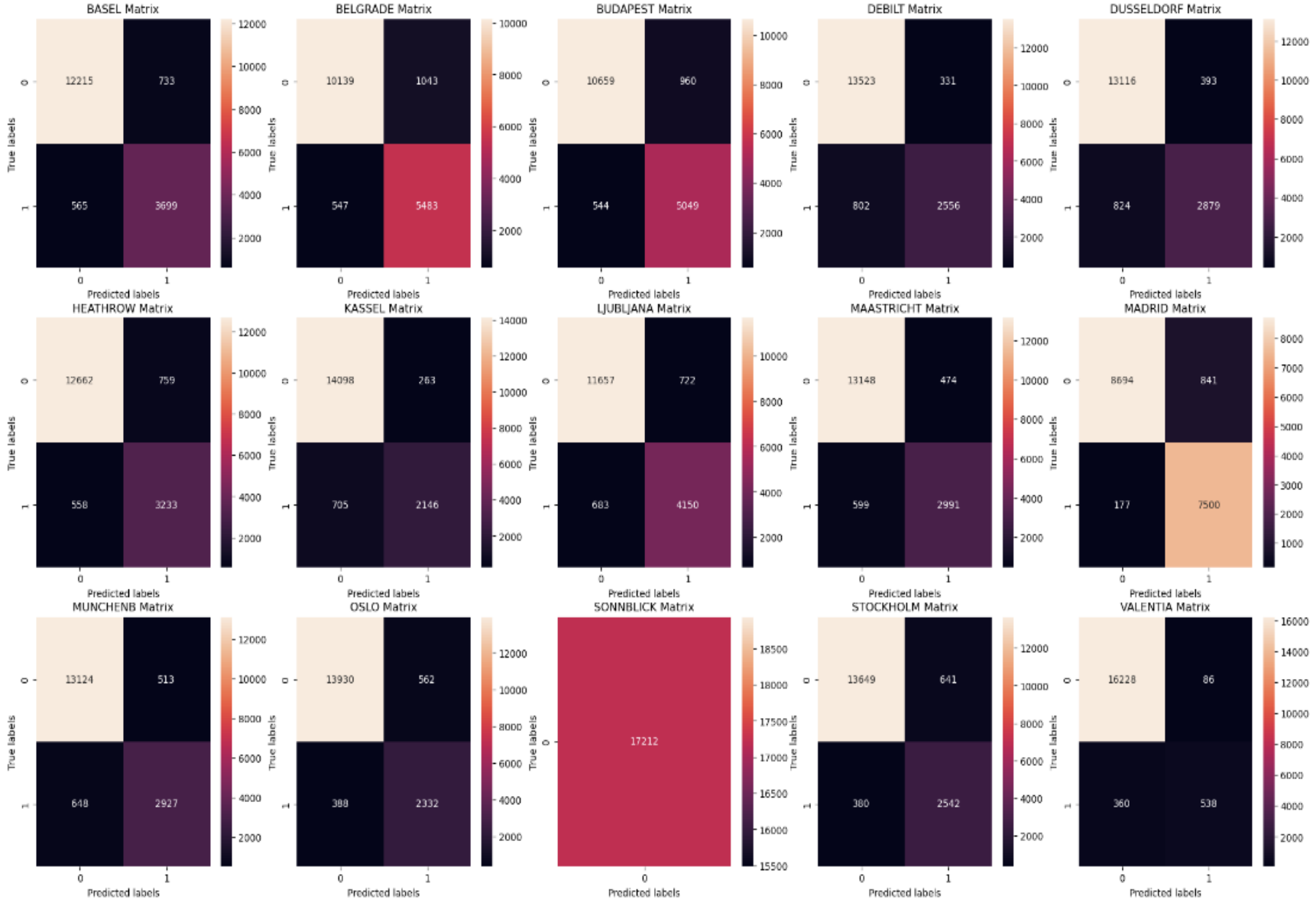
Part 1: Decision Tree

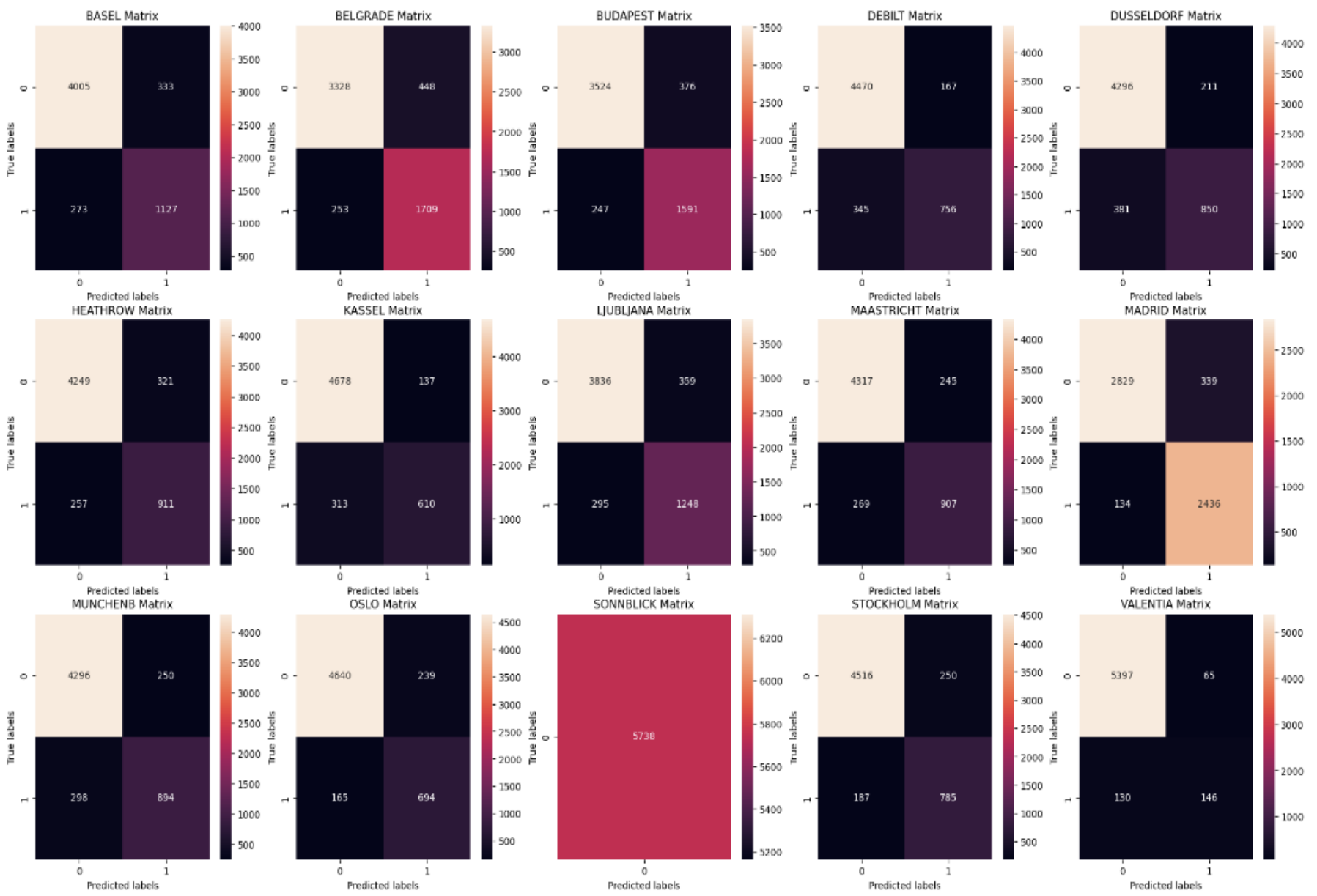


Decision tree should be pruned. It is impossible to read, and very complex. The decision tree had a training accuracy of **46%** and a testing accuracy of **47%**

Training Set Confusion Matrix



Testing Set Confusion Matrix



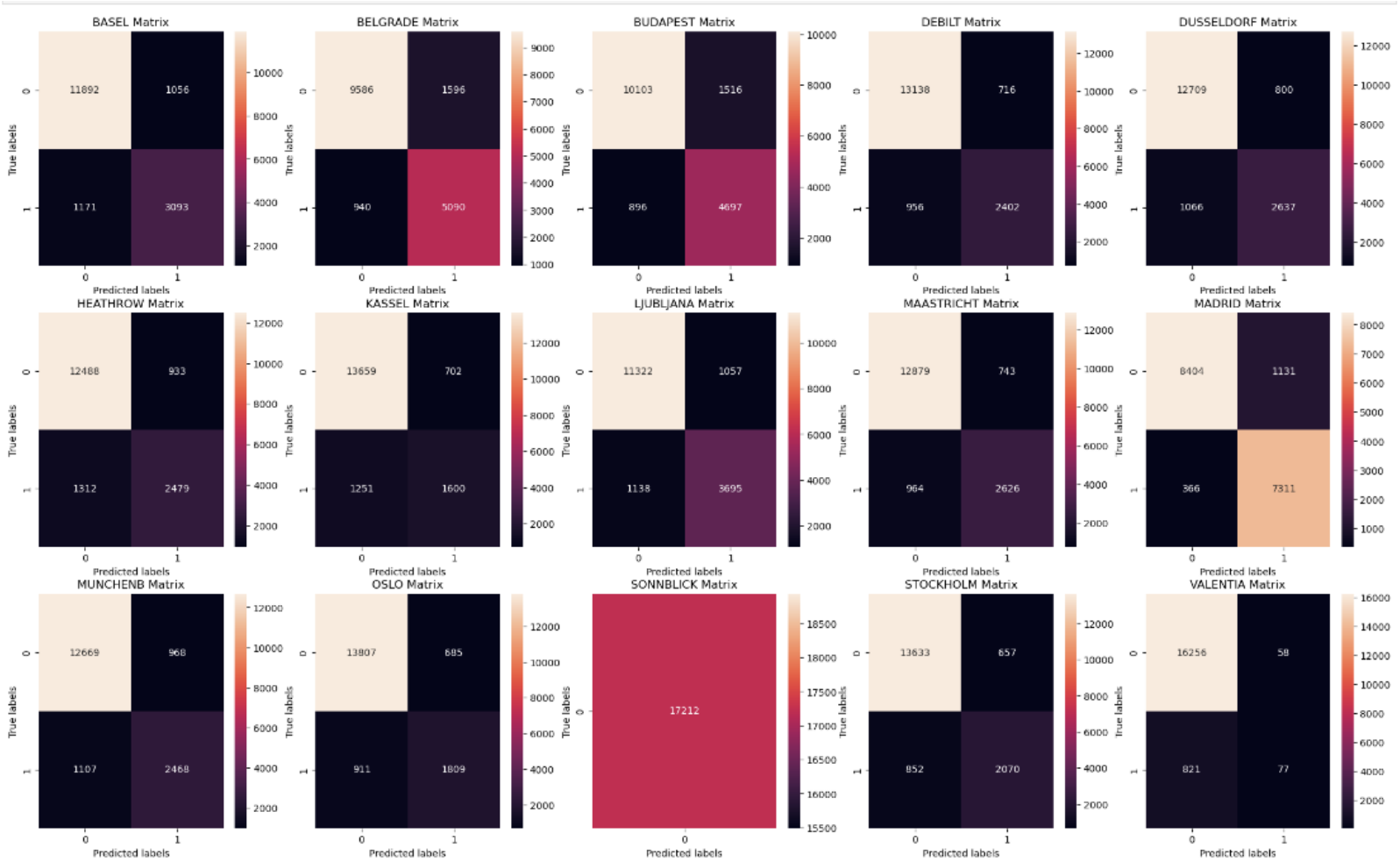
Part 2: ANN

ANN - 2 hidden layers - 5 nodes each - 500 iterations

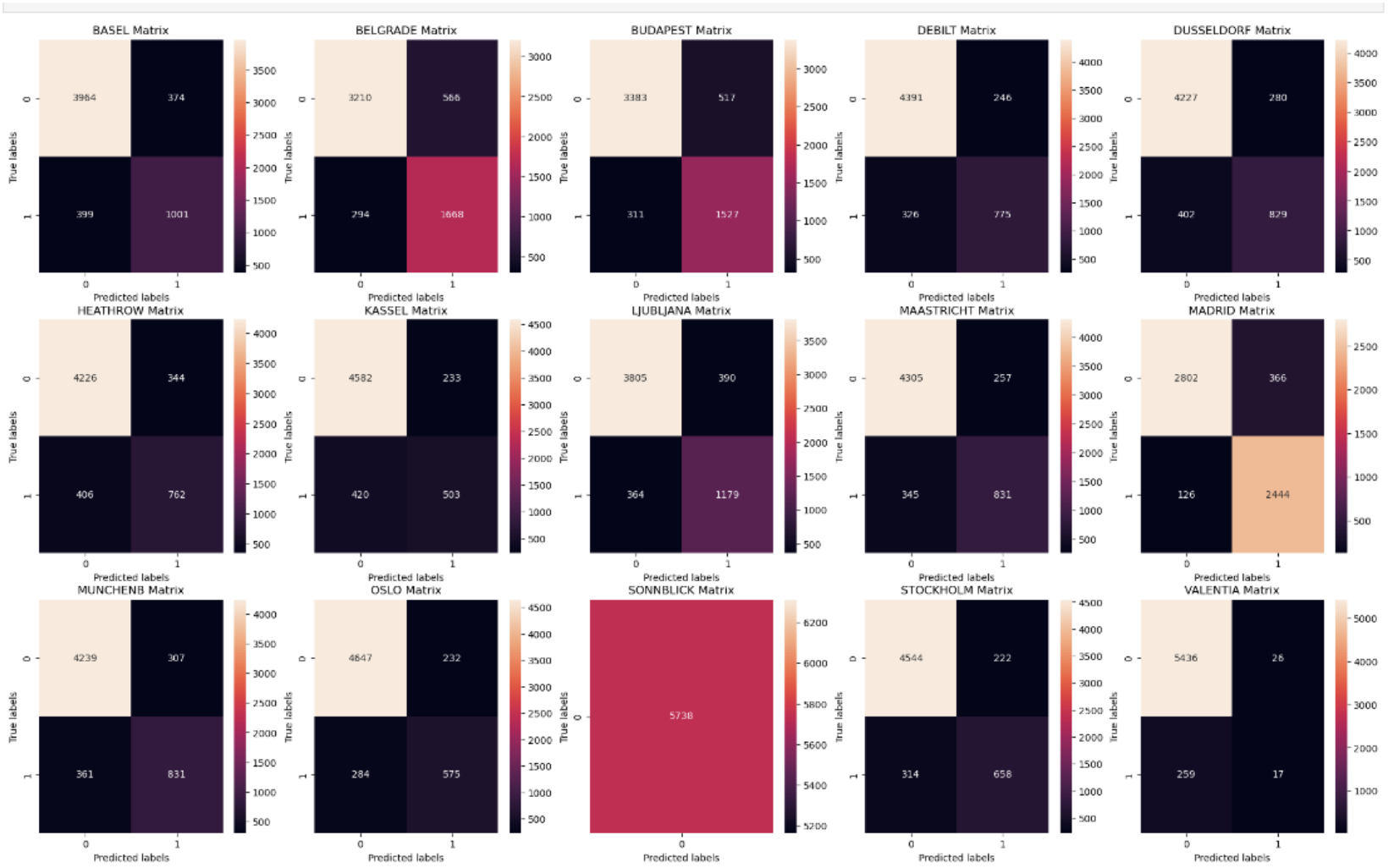
Training Set Accuracy: **47%**

Testing Set Accuracy: **48%**

Training Set Confusion Matrix



Testing Set Confusion Matrix

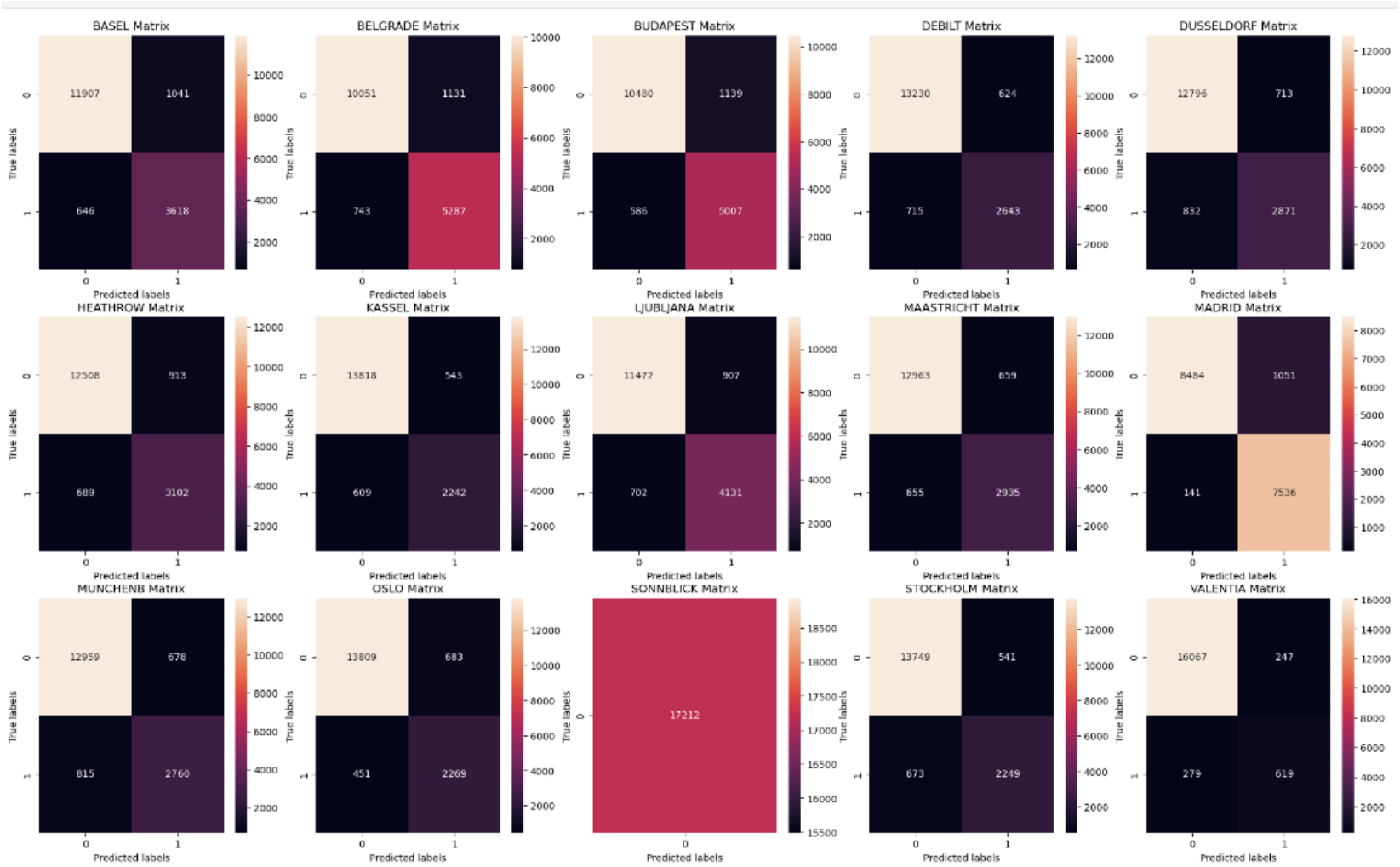


ANN- 3 hidden layers - 25 nodes each - 500 iterations

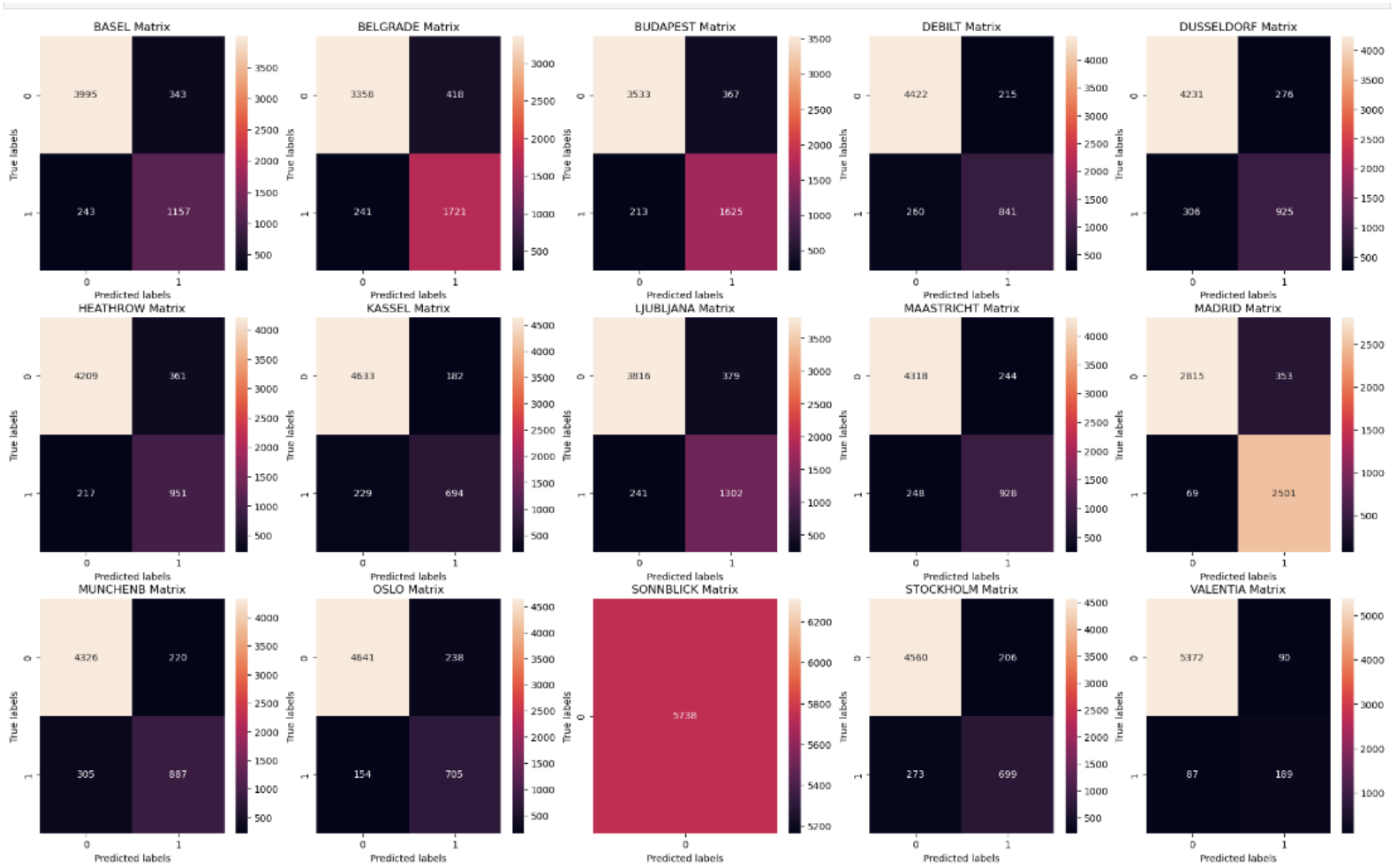
Training Set Accuracy: **53%**

Test Set Accuracy: **52%**

Training Set Confusion Matrix



Testing Set Confusion Matrix

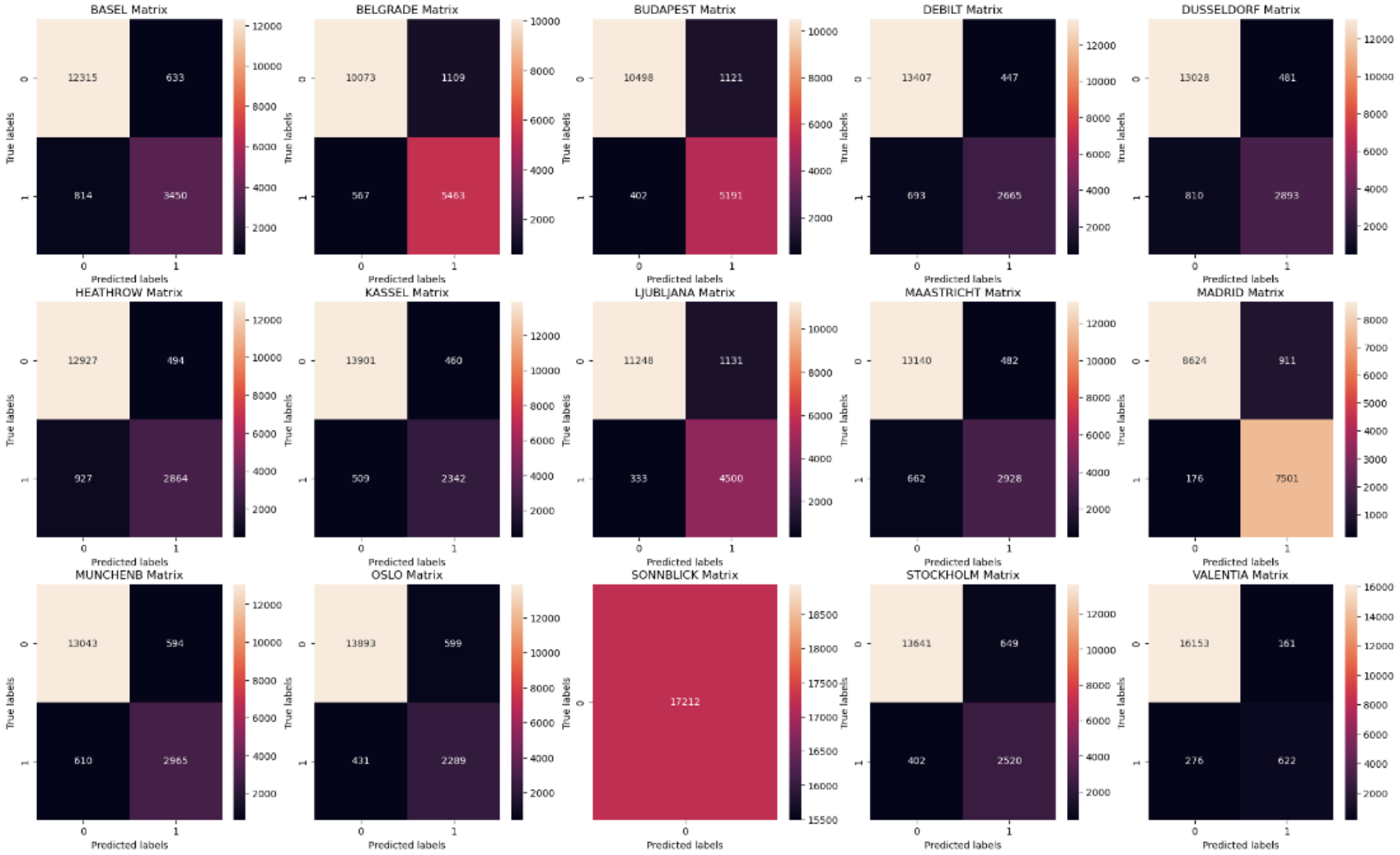


ANN - 5 hidden layers - 50 nodes each - 1000 iterations

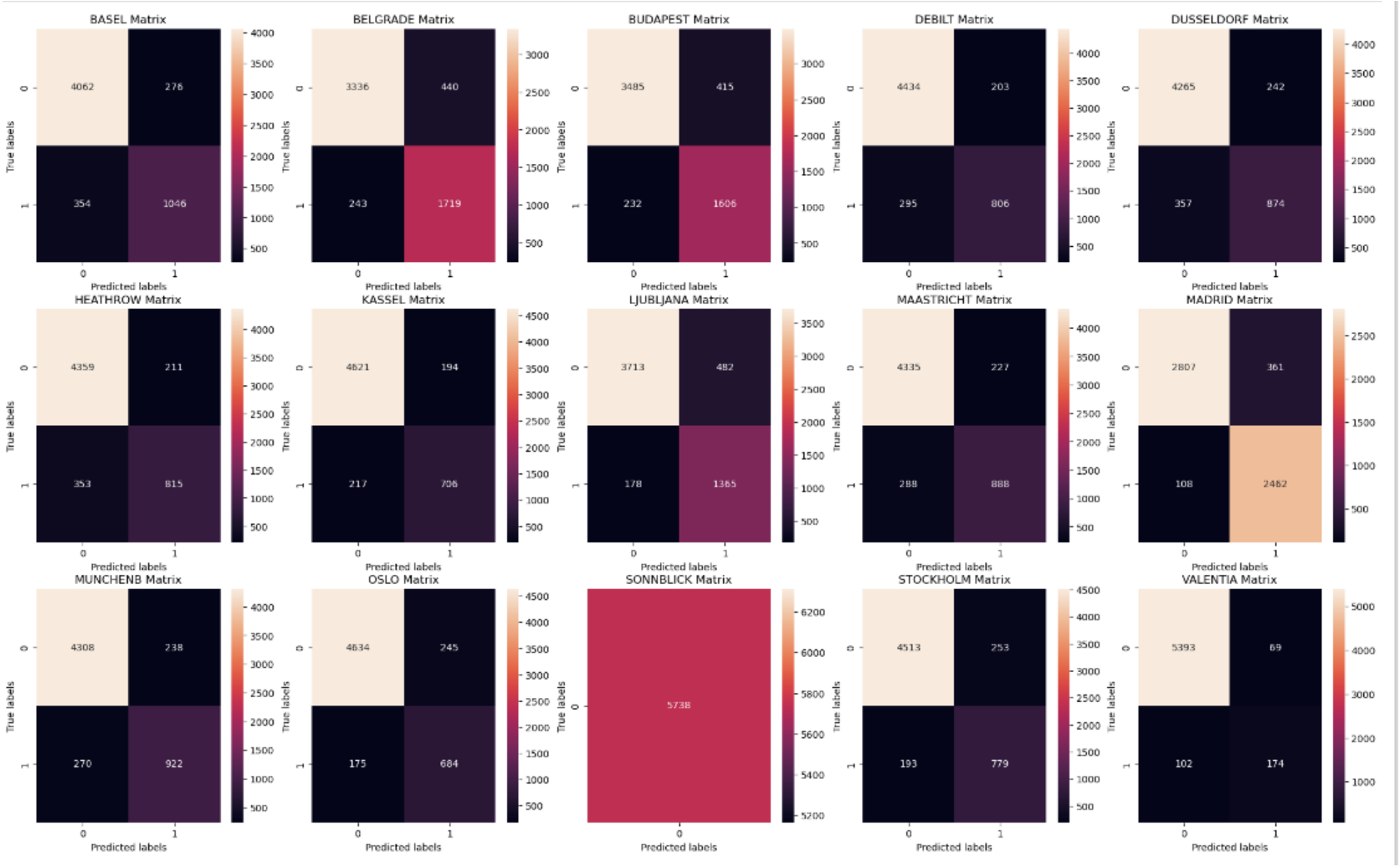
Training Set Accuracy: **58%**

Testing Set Accuracy **51%**

Training Set Confusion Matrix



Testing Set Confusion Matrix



Analysis and Recommendations:

Between the KNN, ANN, and decision tree models, KNN had the highest overall accuracy of 88%. Because of its higher accuracy relative to the other models, I would recommend ClimateWins use the KNN model.

For every model explored, Sonnblick had a 100% accuracy rating. This is suspicious. Models predicting weather for Sonnblick could be overfitted. Additional data for this area is needed to improve the model.

A common issue for both ANN and decision tree models was a false positive for Madrid. This is not seen in the KNN model.