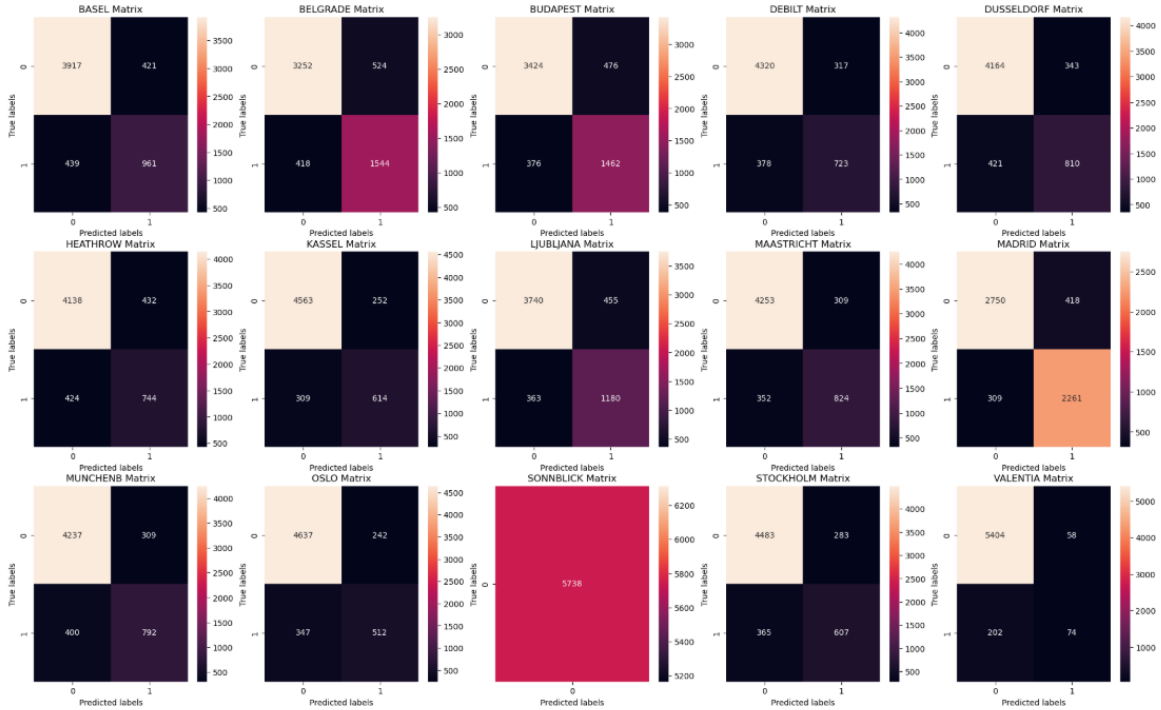
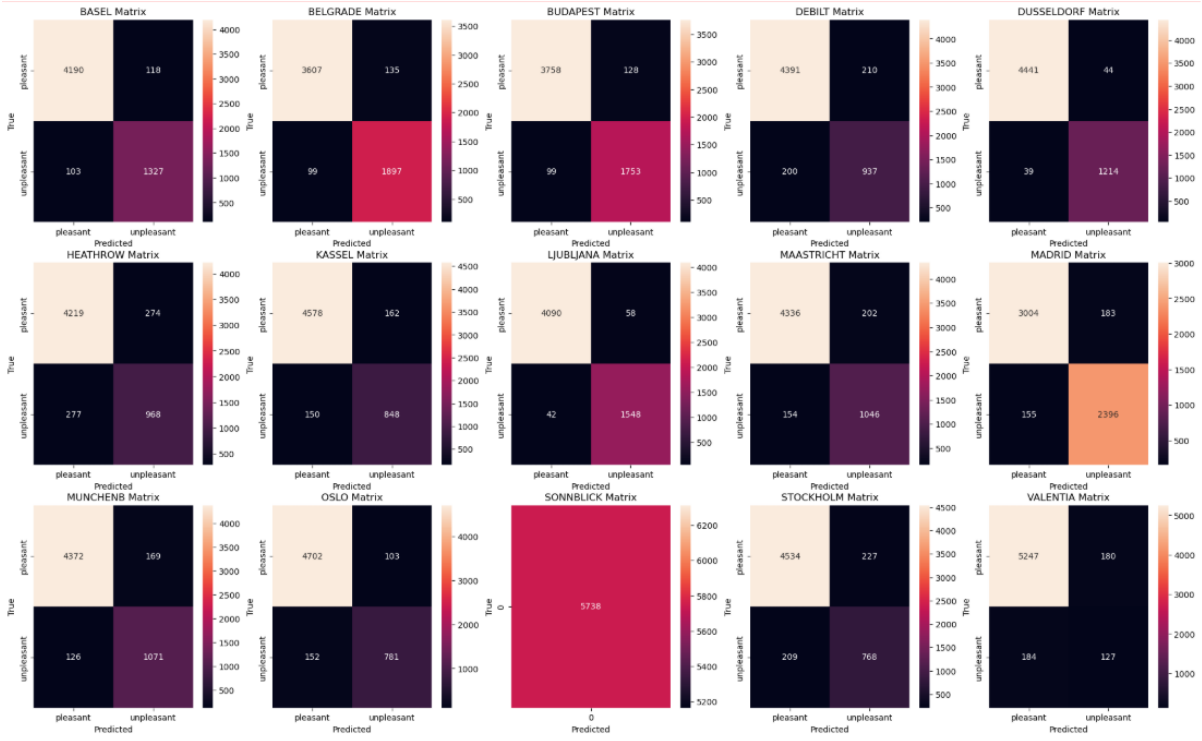
ML Exercise 1.5
Esteban Torres

Supervised Learning Algorithms Pt 2

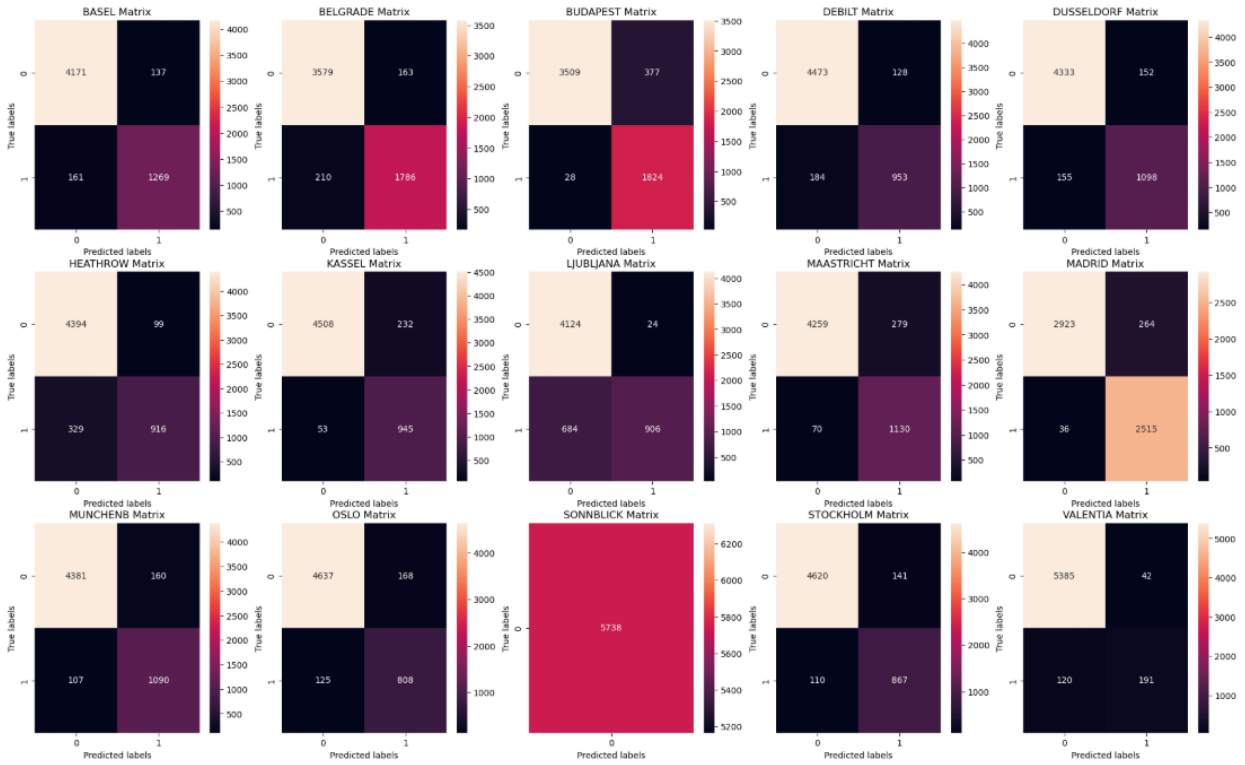
KNN Model



Decision Tree



ANN Model



- Which of these algorithms (including the KNN model from Exercise 1.4) do you think best predicts the current data?

The KNN model has the highest accuracy at 88.5% meaning it predicts the data the best. The decision tree only predicts at 63.9% accuracy and the ANN model at a 62.5% accuracy.

Decision Tree

I do believe the decision tree needs to be pruned. The tree appears overly complex, with many splits that likely capture noise rather than meaningful patterns in the data. This level of complexity increases the risk of overfitting, where the model performs well on training data but poorly on unseen data. Pruning would simplify the tree, improve generalization, and lead to more reliable predictions.

ML Exercise 1.5
Esteban Torres

ANN Model

Scaling improved model performance by stabilizing gradient descent and improving convergence. The scaled model achieved higher and more consistent testing accuracy compared to the unscaled model.

The best-performing model used three hidden layers with 20, 20, and 10 nodes, achieving the highest testing accuracy of 65.8% while maintaining a reasonable gap between training and testing performance. This indicates that the model was able to capture meaningful patterns in the data without severe overfitting. Increasing network depth and complexity improved training accuracy, but gains in testing accuracy began to level off, suggesting diminishing returns beyond this architecture.

- Are any weather stations fully accurate? Is there any overfitting happening?

Only one weather station is classified with 100% accuracy. This is evident from the confusion matrices, which show misclassifications across multiple stations in every scenario. Even in the best-performing model, some stations are confused with others, likely due to overlapping or similar weather patterns. There is **some overfitting**, as training accuracy consistently exceeds testing accuracy across all model configurations.

- Are there certain features of the data set that might contribute to the overall accuracy?

Yes, several features in the dataset contribute strongly to the model's overall accuracy. Variables such as temperature, humidity, wind speed, and atmospheric pressure are key indicators of local weather conditions and help differentiate patterns across stations. Overall, features that are consistently measured provide the strongest signal for accurate predictions.

- Which model would you recommend that ClimateWins use?

The KNN model because of its high accuracy.