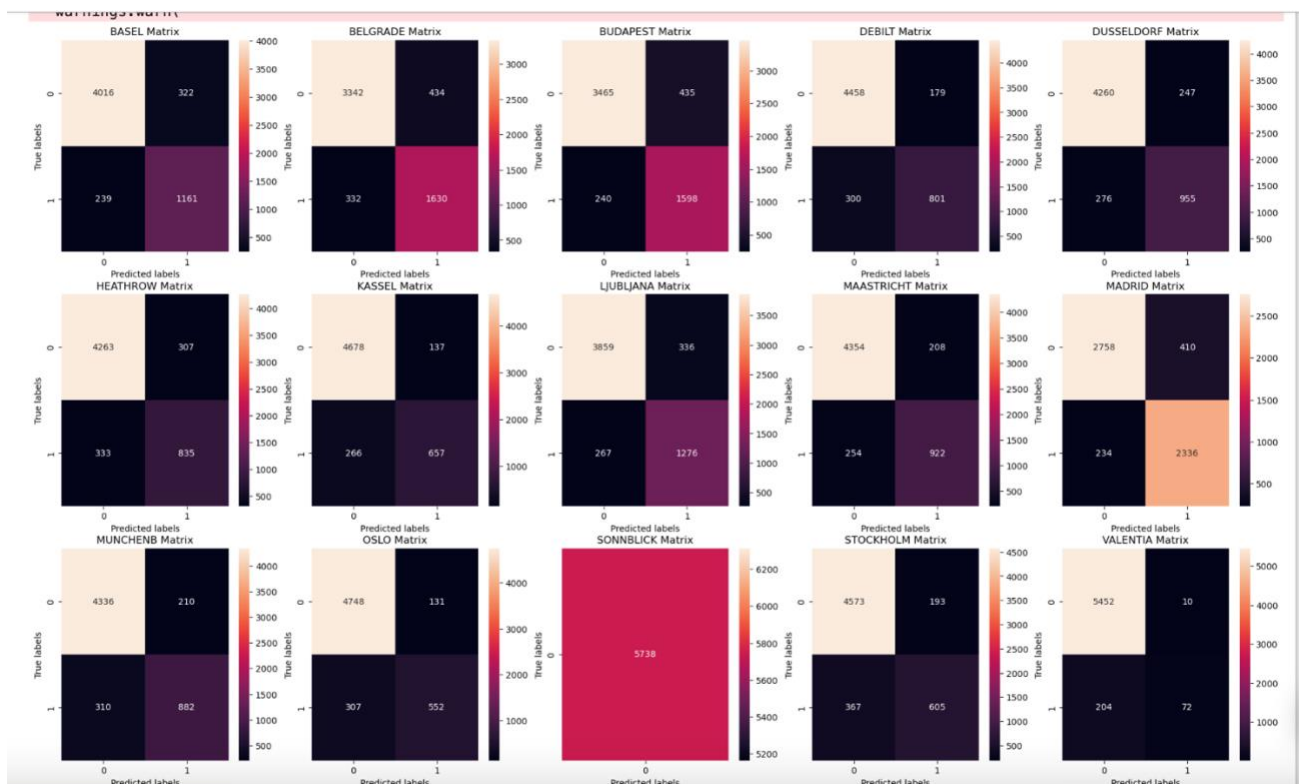


Machine Learning Analysis: Evaluation of KNN, Decision Tree, and ANN

K-Nearest Neighbor Test on Weather and Picnic Suitability Data

- **Starting Parameters:** 1 to 4 neighbors
- **Final Parameters:** 1 to 30 neighbors
- **Accuracy of Training Data:** Drops significantly as the number of neighbors approaches 2, then decreases slightly in accuracy but remains at 0.5 or above
- **Accuracy of Test Data:** Consistently low — below 0.5 — but approaches test accuracy as the number of neighbors increases. By 30, the test and train accuracy have almost converged.

Final Confusion Matrix:



Commentary:

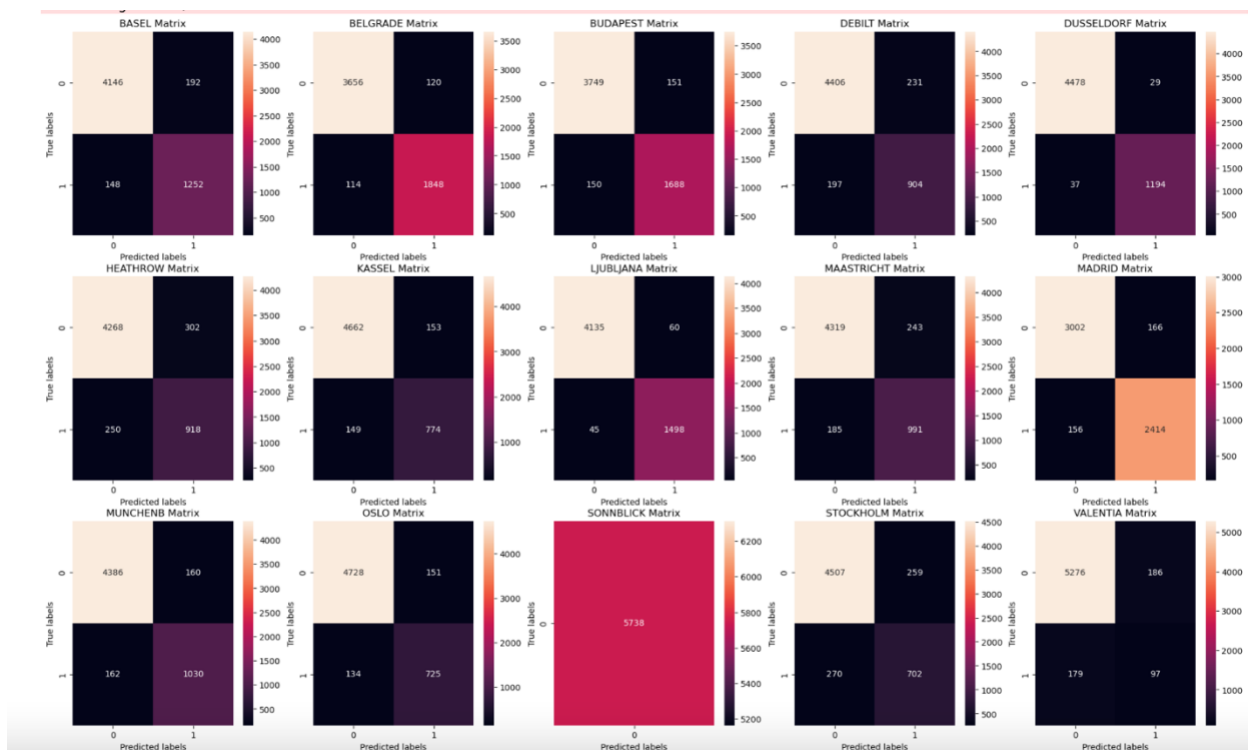
This algorithm is relatively poor at predicting current data. The best fit is between predicted unpleasant and actually unpleasant days. Predicted pleasant and measured pleasant days are a much poorer fit, with more mistakes in both directions (predicted pleasant, actually unpleasant; predicted unpleasant, actually pleasant).

No weather stations are fully accurate, but Madrid comes closest with the fewest mistakes and most accurate pleasant-weather predictions. I don't see anything that could be overfitting, and I'm not certain how to identify underfitting. (High error rates?)

Sonnblick may be decreasing the overall accuracy, but I can't figure out why it's throwing an error. I looked at the actual data and I don't see anything that looks like a red flag, but my lack of experience may be the reason.

I also notice that Valentia has a lot of unpleasant days. It's possible that station may be causing some overfitting, but other stations have errors relatively evenly balanced between pleasant and unpleasant — it's not making significant assumptions on the "unpleasant" side.

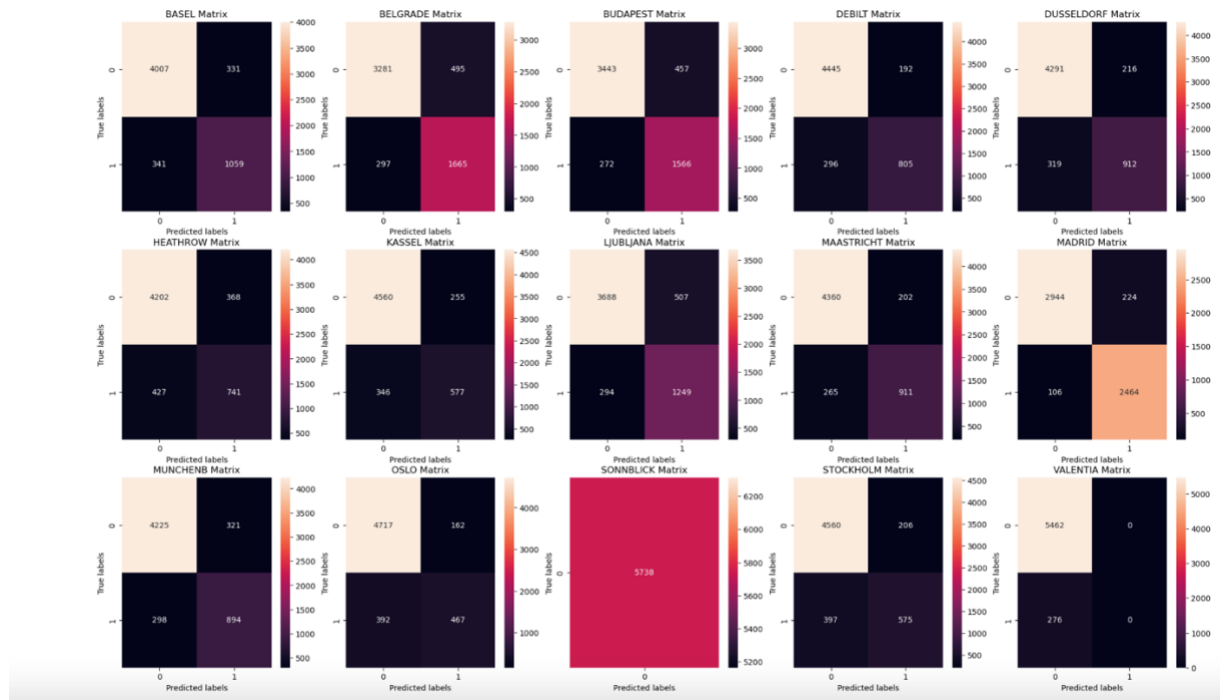
Decision Tree Model: Scaled Data



Artificial Neural Network (ANN)

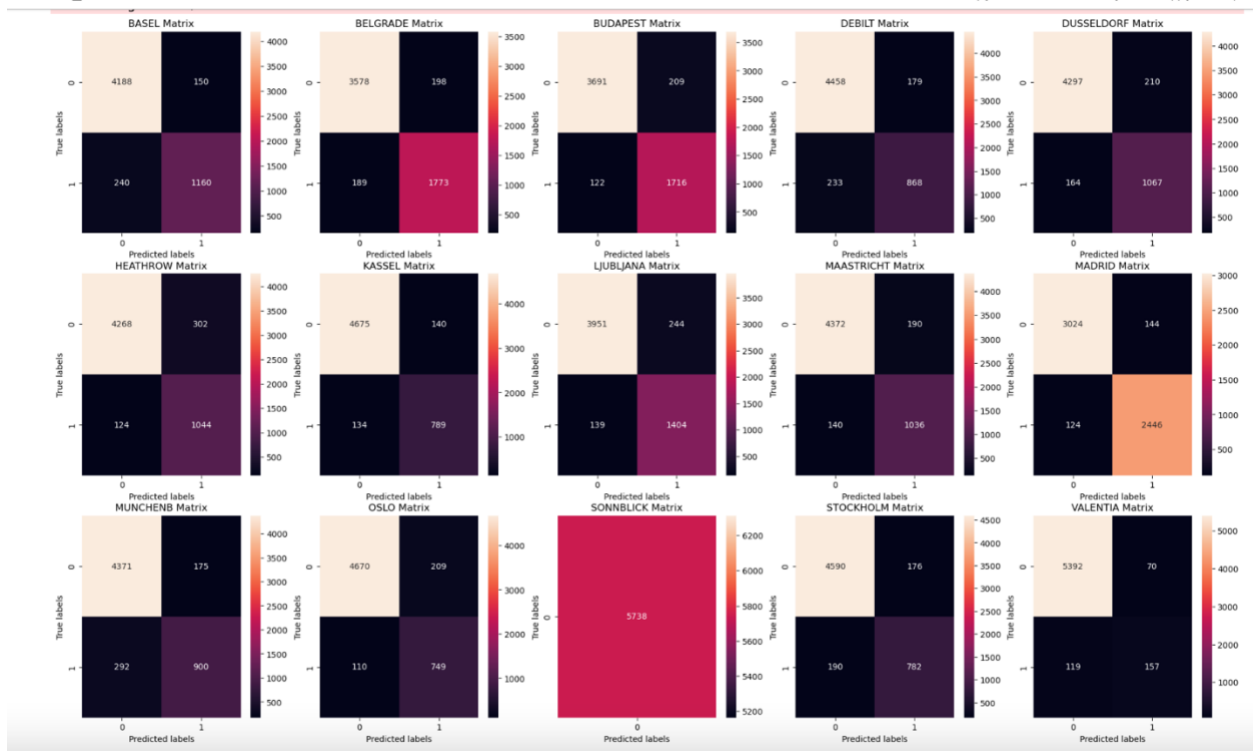
Attempt #1

- Number of hidden layers: 2
- Nodes per layer: 5 each
- Max iterations: 500
- Tolerance: 0.0001



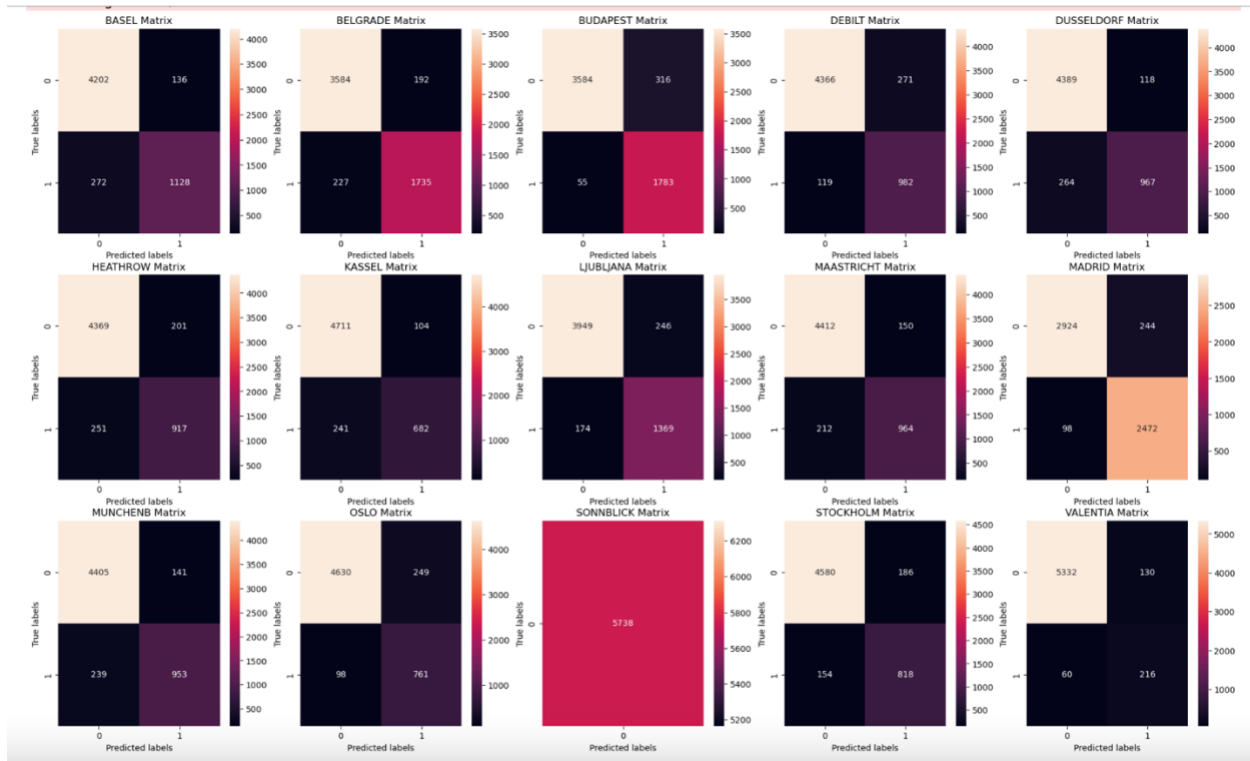
Attempt #2

- Number of hidden layers: 3
- Nodes per layer: 100, 50, 25
- Max iterations: 500
- Tolerance: 0.00001



Attempt #3

- Number of hidden layers: 3
- Nodes per layer: 40, 80, 120
- Max iterations: 5000
- Tolerance: 0.000001



Analysis:

There appears to be minimal measurable difference among the three models. The KNN seems to have the most errors, so I'll advise against that one. Artificial neural network (ANN) seems to have slightly fewer errors than the Decision Tree model, and there seems to be more opportunities to adjust the algorithm. I would cautiously recommend that one but seek a second opinion from someone more experienced.

As of right now, no weather stations are fully accurate, but Valentia seems closest. I don't see any overfitting, but the lack of pleasant days atop the Sonnblick mountain may throw off the model a bit.