CareerFoundry ML with Python
November 16, 2025
Charlotte Lin

Achievement 1.4:
Supervised Learning Algorithms Part 1

Pleasant Weather Prediction

This exercise investigates whether supervised learning can predict “pleasant” (1) versus “unpleasant” (0) daily weather conditions in the ClimateWins Dataset. The goal is to evaluate how well historical weather observations predict pleasant-day suitability and whether the model performs consistently across stations. Because each day includes 15 independent station labels, this task is a multi-label classification problem. This is important for model evaluation, as standard accuracy measurements can be misleading.

Data Preparation & Modeling Approach

The “pleasant weather” answer key included labels for 15 of these stations. Three stations without label coverage were removed. Columns DATE and MONTH were excluded from the feature matrix **X**, and DATE was excluded from the label matrix **y**. A standard 80/20 train–test split (random_state = 42) was used to ensure reproducibility.

A **MultiOutputClassifier** was used to handle the multi-label nature of the task. Because each station produces an independent binary label, evaluating the model using `.score()` would incorrectly treat all 15 labels per row as a single combined class. To avoid this error and account for class imbalance, performance was assessed using:

- **Macro F1-score** (primary metric)
- **Classification report** (precision, recall)
- **Per-station confusion matrices** (interpretation only)

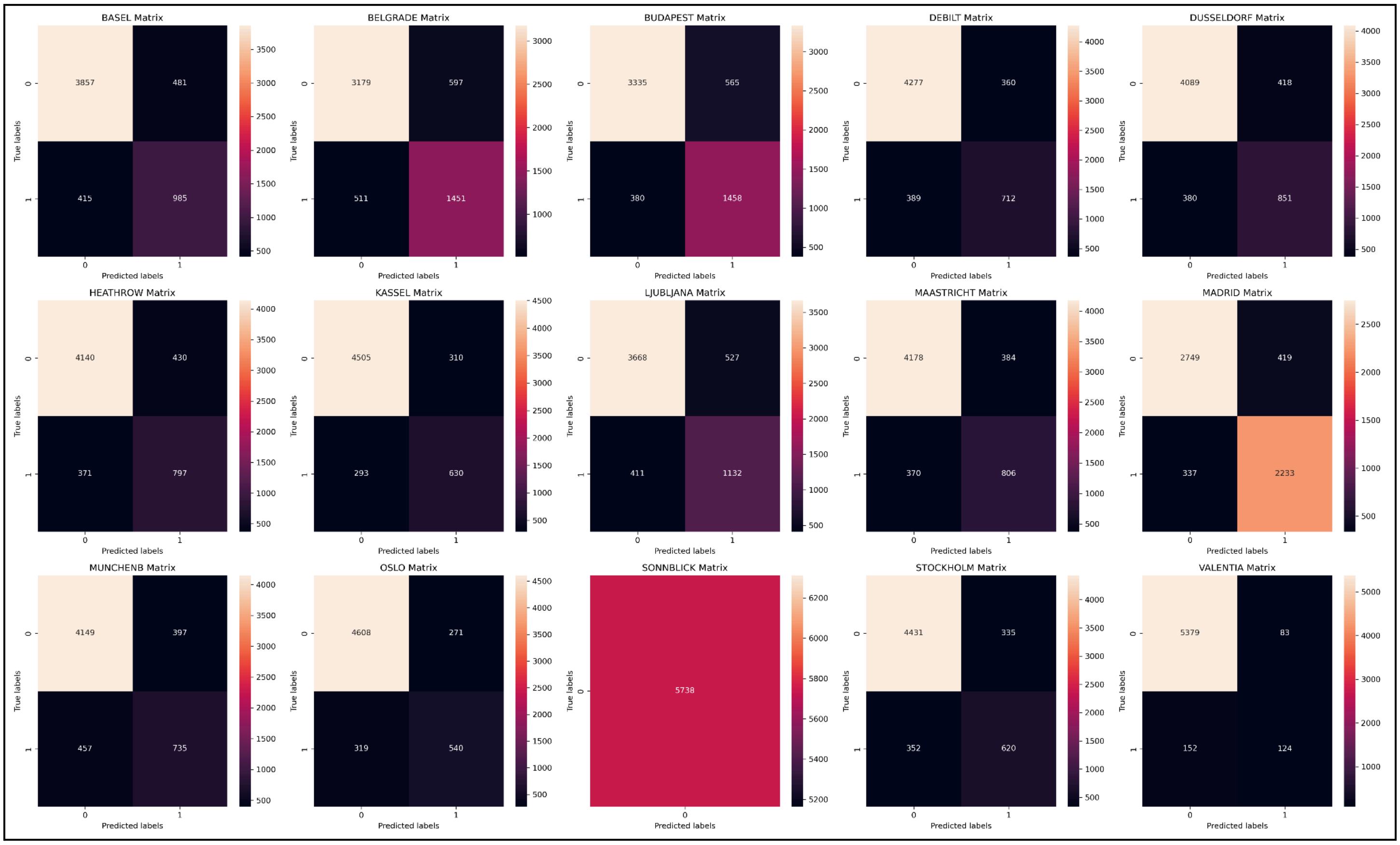
Results

F1-Performance Across Values of k

k	Train F1	Test F1
1	1.00	0.81
2	0.90	0.77
3	0.91	0.82

- k = 1 and k = 3 produced the strongest test F1-scores (~0.81).
- Train–test gap at k = 1 (1.00 vs 0.81) shows moderate overfitting.
- k = 3 slightly reduces overfitting while maintaining strong performance; therefore, k = 3 was selected as the final model.

Confusion Matrices Showing Station-Level Performance



- Most stations show a strong ability to identify unpleasant days, with high true-negative counts (upper-left cell).
- The model is less accurate at detecting pleasant days, which is expected due to class imbalance in the dataset.
- Stations with highly imbalanced labels show skewed matrices where the model predicts mostly one class.

Station-Level Accuracy Summary

	Station	True Negative	False Positive	False Negative	True Positive	Accuracy (%)
0	BASEL	3857	481	415	985	84.38
1	BELGRADE	3179	597	511	1451	80.69
2	BUDAPEST	3335	565	380	1458	83.53
3	DEBILT	4277	360	389	712	86.95
4	DUSSELDORF	4089	418	380	851	86.09
5	HEATHROW	4140	430	371	797	86.04
6	KASSEL	4505	310	293	630	89.49
7	LJUBLJANA	3668	527	411	1132	83.65
8	MAASTRICHT	4178	384	370	806	86.86
9	MADRID	2749	419	337	2233	86.82
10	MUNCHENB	4149	397	457	735	85.12
11	OSLO	4608	271	319	540	89.72
12	SONNBLICK	5738	0	0	0	100.00
13	STOCKHOLM	4431	335	352	620	88.03
14	VALENTIA	5379	83	152	124	95.90

Across all 15 stations, the **average accuracy rate is 87.55%**, which shows that the KNN model performs reasonably well overall at distinguishing pleasant from unpleasant weather conditions. Most stations show balanced performance, with accuracy rates clustered between **83% and 90%**.

A few stations stand out. **Kassel** and **Oslo** achieve among the highest balanced accuracies (≈89–90%), suggesting their local weather patterns align more closely with the training data. In contrast, **Sonnblick** shows a perfect 100% accuracy but predicts *only* unpleasant days, which reflects extreme label imbalance rather than true predictive skill.

Conclusion

The KNN model provides a reasonable first attempt at predicting pleasant-weather suitability across European stations. Future exercises using Decision Trees and Artificial Neural Networks will allow for deeper comparison and may help address limitations identified here.