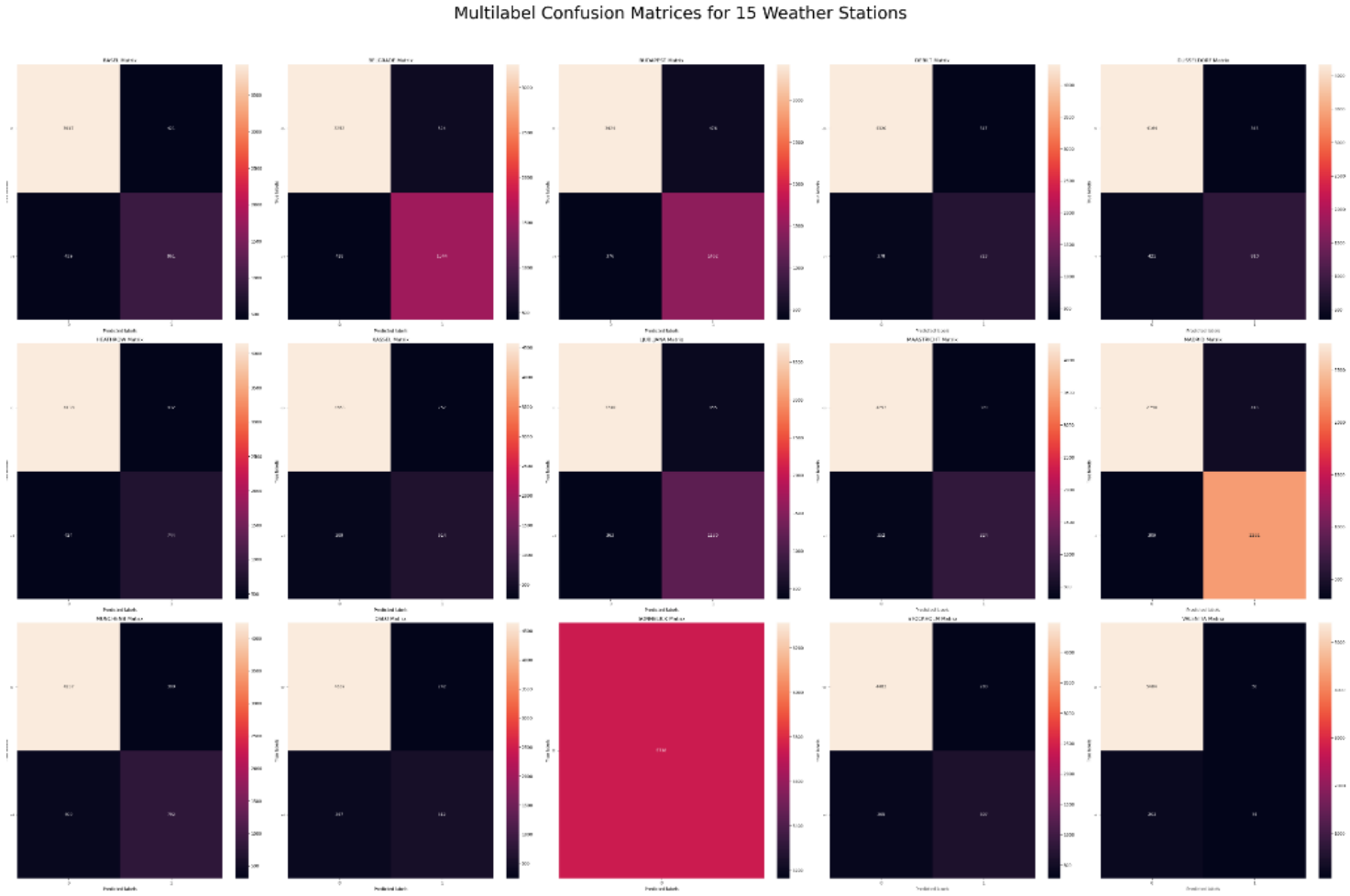
JESSICA GATTA

Weather Prediction Analysis – KNN Model

Analysis conducted by: Jessica Gatta
Date: January 2026



Objective

ClimateWins aims to use European temperature data to predict whether a day will be pleasant enough for outdoor activities. Pleasant days are labeled as 1, and unpleasant days as 0. The goal of this analysis is to evaluate the KNN model's predictive ability and see how it performs across different weather stations.

Model Overview

The analysis uses a K-Nearest Neighbors (KNN) algorithm with a multilabel approach, predicting multiple stations at once. Performance is evaluated by using confusion matrices for each station.

Results by Station

The table below shows the model's performance for each station, including accurate predictions, false positives, false negatives, and overall accuracy:

	Weather Station	Accurate Predictions	False Positive	False Negative	Accuracy Rate
0	BASEL	961 - 3917	421	439	0.85
1	BELGRADE	1544 - 3252	524	418	0.84
2	BUDAPEST	1462 - 3424	476	376	0.85
3	DEBILT	723 - 4320	317	378	0.88
4	DUSSELDORF	810 - 4164	343	421	0.87
5	HEATHROW	744 - 4138	432	424	0.85
6	KASSEL	614 - 4563	252	309	0.90
7	LJUBLJANA	1180 - 3740	455	363	0.86
8	MAASTRICHT	824 - 4253	309	352	0.88
9	MADRID	2261 - 2750	418	309	0.87
10	MUNCHENB	792 - 4237	309	400	0.88
11	OSLO	512 - 4637	242	347	0.90
12	SONNBLICK	0 - 5738	0	0	1.00
13	STOCKHOLM	607 - 4483	283	365	0.89
14	VALENTIA	74 - 5404	58	202	0.95

Analysis

The model shows varying accuracy across stations. Sonnblick achieved perfect accuracy, while stations like Madrid were less precise, especially for predicting pleasant weather. This suggests that the model may perform best where training data patterns are consistent but could struggle to generalize in more variable conditions.
The perfect score at Sonnblick may indicate overfitting, meaning the model memorized the training data instead of learning to generalize. Differences in performance across stations highlight the need for more diverse training data to ensure the model works well in different locations.

Recommendations & Conclusions

To improve the model, the training dataset should include a wider range of weather patterns. Applying techniques like cross-validation or ensemble methods could enhance reliability. Overall, the KNN model performs well in some stations, but its effectiveness varies across Europe. Increasing diversity in training data and validation will help the model generalize better and make more reliable predictions.