WEATHER STATION PREDICTION ANALYSIS

April Boyd

INTRODUCTION

Our project is focused on evaluating weather station data by applying optimization and supervised learning techniques. The goal is to identify meaningful trends, predict future patterns, and assess the accuracy of different models. By combining these approaches, we can not only strengthen our understanding of the data but also provide ClimateWins with reliable insights to guide decision-making around weather predictions, such as determining conditions suitable for activities like picnics.

Key Questions

- How can machine learning be applied to weather and climate data?
- What ethical concerns exist in using AI for climate predictions?
- Can models predict favorable conditions or dangerous extremes that affect people’s lives?

HYPOTHESIS

- Hypothesis 1: Some stations will show consistent patterns that make predictions more accurate than others.

- Hypothesis 2: Optimization methods will highlight features that most strongly influence prediction accuracy.

- Hypothesis 3: Supervised learning models will perform differently across stations, with potential overfitting in more uniform datasets.

DATA SOURCE & BIASES

Source: Historical weather station temperature datasets.

Possible Biases:

- Data gaps (missing years, leap year inconsistencies).
- Regional variability (some stations may have more predictable climates).
- Measurement errors from sensors or human logging.

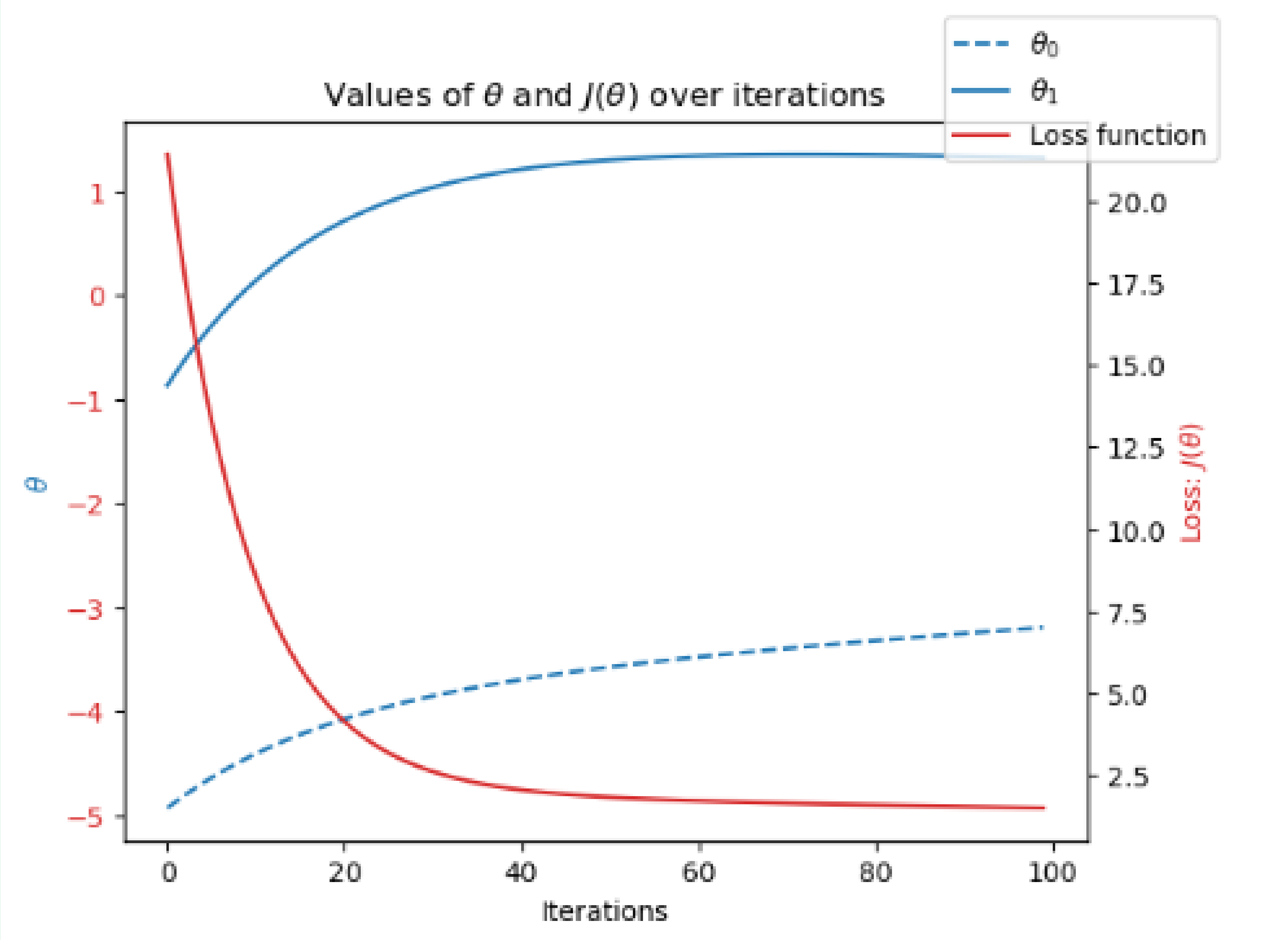
Accuracy: Most stations show strong accuracy (mid-80% to 90%), with Sonnblick at 100% raising overfitting concerns.

••••• OPTIMIZATION •••••

Method: Gradient descent applied to extract θ_0 (intercept) and θ_1 (slope) for each station-year dataset.

Why it matters:

- Helps determine the relationship between time and average daily temperatures.
- Identifies warming (positive slope) or cooling (negative slope) patterns.



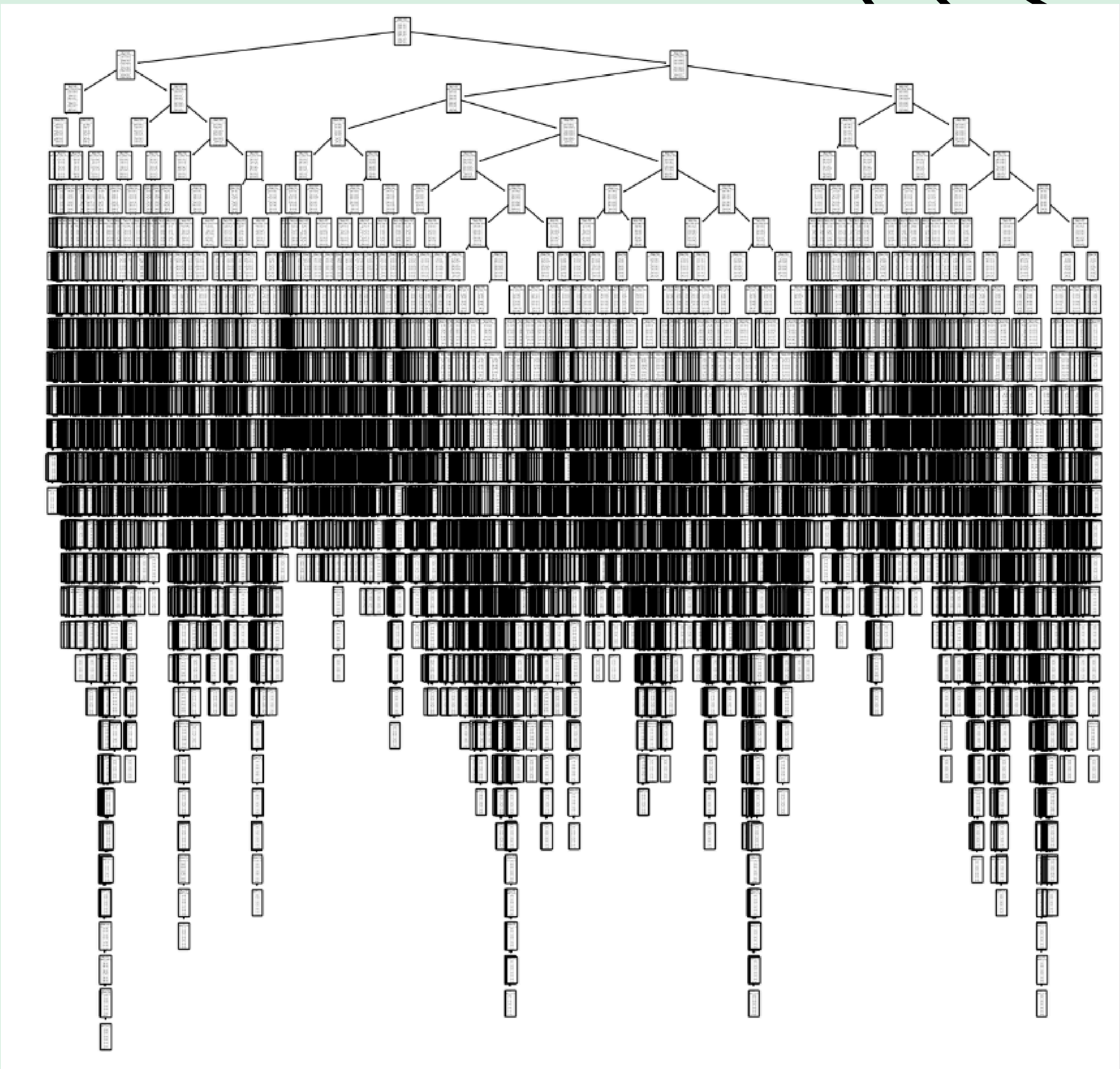
METHOD 1: K-NEAREST NEIGHBOR (KNN)

KNN makes predictions by comparing new data points to what it has already seen. It looks at the "nearest neighbors" in the dataset and groups points together based on similarity. For my analysis, this method performed fairly well, with an average accuracy of about 89 percent across stations. To make KNN stronger, it would be important to expand the training data to include more diverse weather conditions across different regions so the model doesn't favor only certain stations.

WEATHER STATION	ACCURATE PREDICTIONS	FALSE POSITIVES	FALSE NEGATIVES	ACCURACY RATE
BASEL	3917	961	421	85%
BELGRADE	3252	1544	524	84%
BUDAPEST	3424	1462	476	85%
DEBILT	4320	723	317	88%
DESSELDORF	4164	810	343	87%
HEATHROW	4138	744	432	85%
KASSEL	4563	614	252	90%
LJUBLJANA	3740	1180	455	86%
MAASTRICHT	4253	824	309	88%
MADRID	2750	2261	418	87%
MUNCHENB	4237	792	309	88%
OSLO	4637	512	242	90%
SONNBLICK	5738	0	0	100%
STOCKHOLM	4483	607	283	89%
VALENTIA	5404	74	50	96%
AVERAGE				**88%**

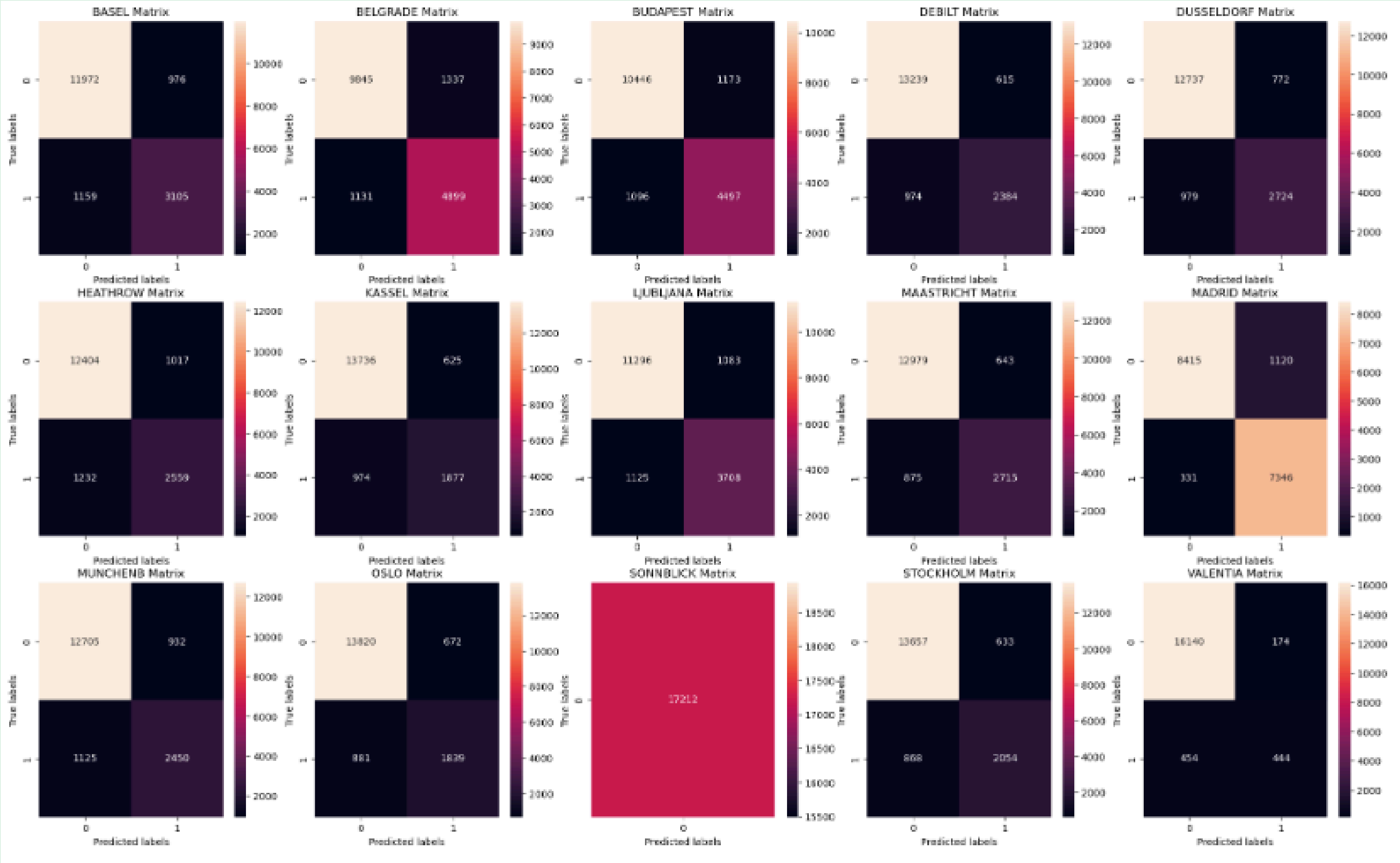
METHOD 2: DECISION TREE

A decision tree works by asking a series of yes-or-no style questions about the features in the data and then branching down until it makes a prediction. While decision trees are easy to understand visually, the one I used ended up being very complex with many branches, which lowered interpretability and increased training time. In this case, the accuracy was around 64 percent, which is not as strong as KNN, and the complexity suggests the tree might be overfitting or struggling with the variability in the dataset.



METHOD 3: ARTIFICIAL NEURAL NETWORK (ANN)

An ANN is designed to mimic the way the human brain works, using layers of “neurons” that adjust their connections as they learn from data. It can capture complex patterns, but it also requires careful tuning, like adjusting the number of layers and the maximum iterations. In this project, the neural network reached about 64 percent accuracy, similar to the decision tree. This shows potential, but it would need more optimization and experimentation with different architectures to compete with simpler models like KNN.



MODEL COMPARISON & RECOMMENDATION

- Decision tree is too complex and not interpretable
- KNN gives the highest accuracy (89%), stronger than ANN (60%)
- ANN is more powerful for complex, nonlinear tasks

- ANN scales better for larger datasets and is less sensitive to noise
- Recommendation: Use ANN for weather predictions due to the complexity of climate data

SUMMARY

- Machine learning algorithms can predict weather conditions with strong results, reaching accuracy rates between 84% and 96%.
- Accuracy does vary depending on geographic location and regional climate conditions.
- By comparing weather conditions over time, machine learning can also help identify signals of climate change and its potential impacts.

NEXT STEPS

- Compare performance across multiple supervised learning models to confirm which approach generalizes best.
- Expand the dataset to include additional features such as humidity, wind speed, and precipitation for richer predictions.
- Test different optimization strategies and learning rates to improve stability and convergence.

THANK YOU

https://github.com/aprilboyd12