ClimateWins:
Analyzing Weather Conditions and Climate Change

- Supriya Shrotriya
- June 2025

Project Overview

ClimateWins is a European nonprofit organization, that is interested in using machine learning to help predict the consequences of climate change around Europe and, potentially, the world.

Key Questions :

- How is machine learning used? Is it applicable to weather data?
- Can machine learning reliably predict whether the weather will be pleasant based on historical weather data?

Objective and Hypothesis

Objective:

The goal is to predict daily weather conditions that are pleasant for outdoor activities across various European weather stations using machine learning.

Hypothesis

We hypothesize that different algorithms will have varying success in modeling "pleasant" weather, and this presentation aims to explore which model is most effective.

- Weather prediction accuracy may change depending on the location and climate.
- Machine learning can help identify signs of climate change and its potential effects.
- Supervised learning can be used to predict if a day will be pleasant or not based on historical weather data.

Data

This data is collected by the European Climate Assessment & Data Set project.

European Climate Assessment & Data Set Project

Across Europe, it contains data ranging from the late 1800s to 2022

Data points include temperature, wind speed, snow, global radiation, etc

Based on weather observations from 18 different weather stations across Europe

Data Bias & Accuracy

Bias in data can reduce model accuracy, cause unfair predictions, and mislead strategic decisions.

- Temporal Bias - Data spans from the 1800s to 2022. Earlier data may be inaccurate due to outdated tools and methods.
- Location Bias - Some data comes from extreme environments, "normal" or "pleasant" weather based on unusual conditions, limiting generalizability.
- Selection Bias - The dataset includes only 18 weather stations—mostly from well-developed regions. Underrepresents smaller or less-developed areas, leading to biased predictions that favor certain geographies.
- Historical Bias - Climate change has accelerated recently, but large volumes of older data dominate the dataset.

Data Optimization

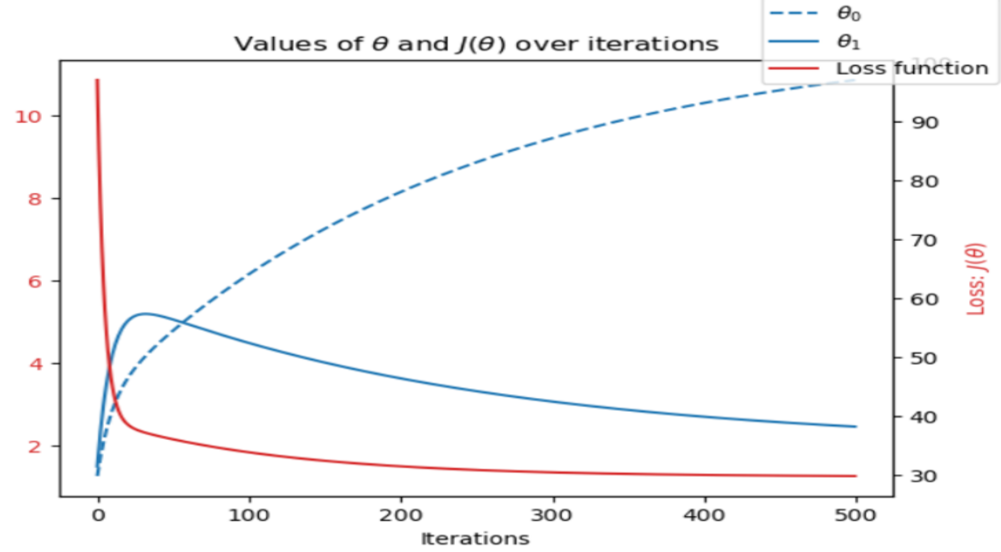
Optimization lowers the risk of error and improves the accuracy of a model, often used to determine which algorithms to use. It helps understand valleys and peaks of the local/global landscape of the data.

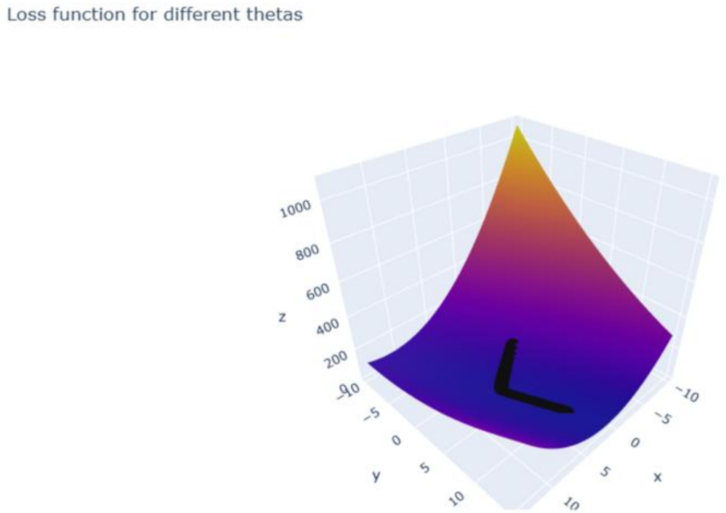
Gradient Descent was used to optimize data on the average temperatures of several weather stations.

Iterative optimization was able to minimize the loss function to nearly zero.

The temperature data can likely be predicted by differentiable functions.

This graph displays a gradient descent loss function for Madrid's average daily temperatures in 2018.

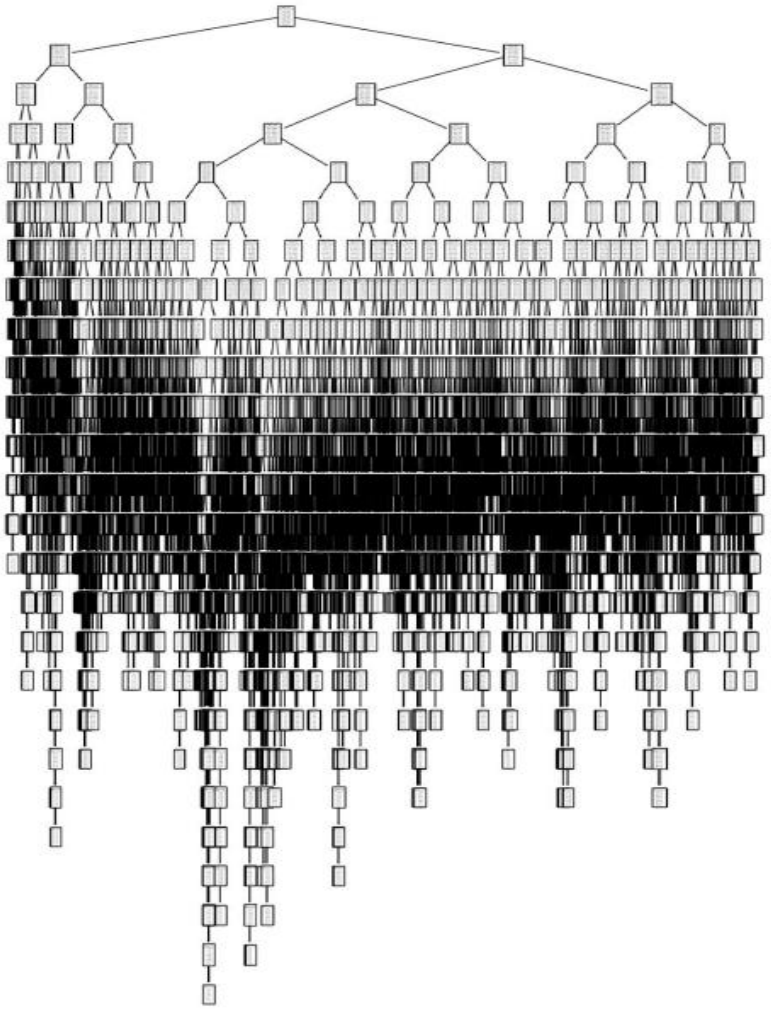




Method 1: K–Nearest Neighbor

- The k-nearest neighbors (KNN) algorithm makes predictions by measuring the distance between new data points and nearby data points, or neighbors, then groups the new data with nearby data points based on the number of neighbors in each group.
- Using a **k-range value of 3**, the model was able to predict pleasant weather days with an **average accuracy of 88%**
- The model's **100% at Sonnblick** may indicate **overfitting**.
- The accuracy variations suggest the training data might not fully represent the range of real-world conditions the model will face.

Weather Station	Accurate	predictions	False Positive	False Negative	Accuracy Rate
Basel	3917	961	421	439	85%
Belgrade	3252	1544	524	418	84%
Budapest	3424	1462	476	376	85%
Debilt	4320	723	317	378	88%
Desseldorf	4164	810	343	421	87%
Heathrow	4138	744	432	424	85%
Kassel	4563	614	252	309	90%
Ljubljana	3740	1180	455	363	86%
Maastricht	4253	824	309	352	88%
Madrid	2750	2261	418	309	87%
Munchemb	4237	792	309	400	88%
Oslo	4637	512	242	347	90%
Sonnblick	5738	0	0	0	100%
Stockholm	4483	607	283	365	89%
Valentia	5404	74	50	202	96%
					88%



Method 2: Decision Tree

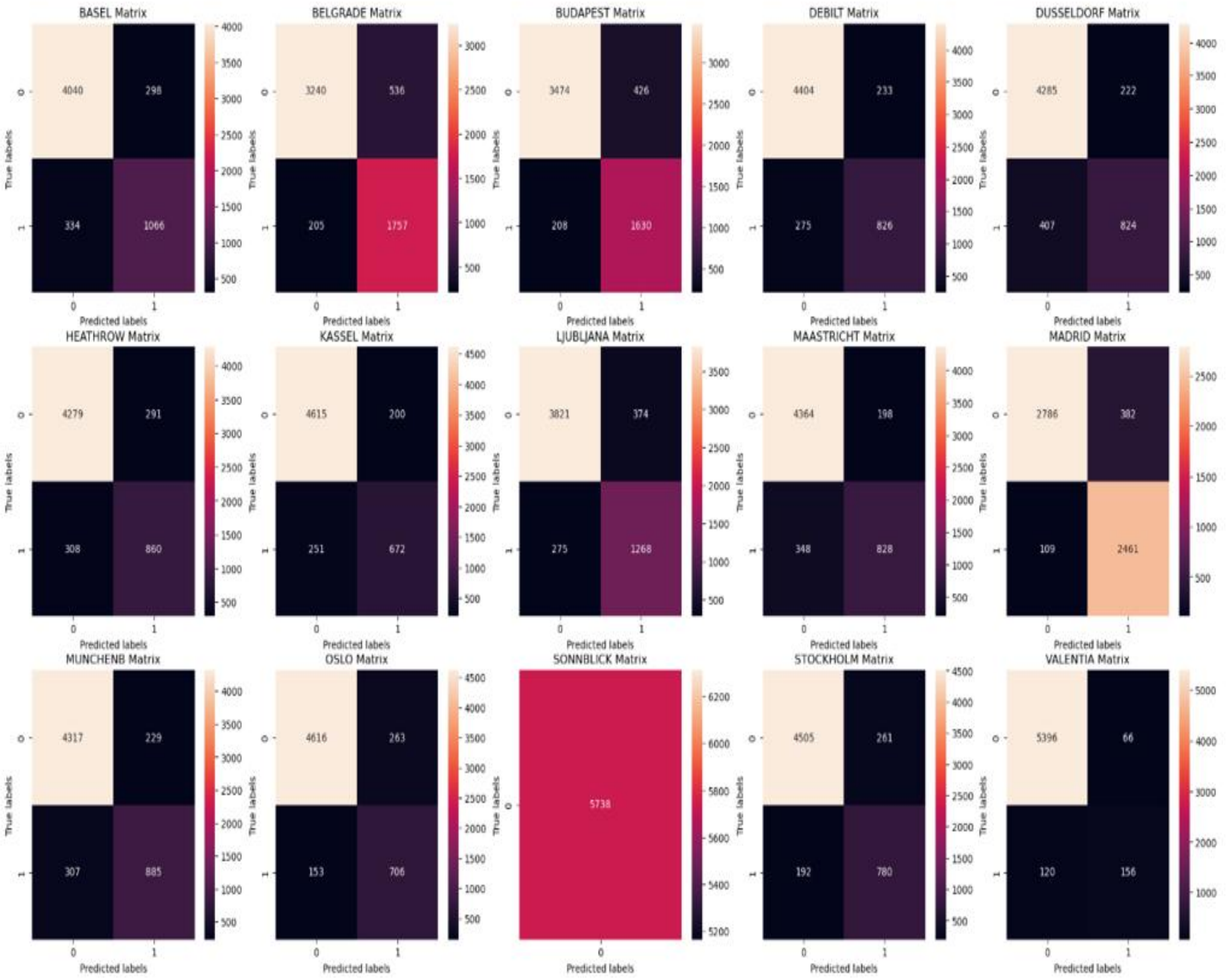
- Decision tree makes predictions by answering a series of questions about the features of the data.
- They have roots, branches, and leaves. The *root* is the first question asked, with yes/no answer. Leading to another question/*branch*. The answer is the *leaf*/stopping point.
- The model performance :

  Training data accuracy of **46%**

  Testing data accuracy of **47%**
- The decision tree needs to be pruned as it is too deep and complex to interpret.

Method 3: Artificial Neural Network

- Inspired by the human brain, artificial neural network (ANN) algorithms make predictions by processing data through a set of layers. Input layers receive the data and are weighted based on their connection to hidden layers. Hidden layers apply functions to the weighted data and send the results to output layers.
- ANN provided more balanced results, showing potential for better generalization but still struggled with the complexity of the data
- Train accuracy score 52%
- Test Accuracy Score 49%



Confusion Matrix- Each matrix represents accurate predictions of unpleasant days (upper left), inaccurate predictions of unpleasant days (upper right), inaccurate predictions of pleasant days (lower left), and accurate predictions of pleasant days (lower right).
Sonnblick has a solid matrix, which represents the ANN’s 100% accuracy in predicting only unpleasant weather days for this location.

How Well Did Our Models Perform?

- The **decision tree** model is too complex to interpret.
- The KNN model provides the highest accuracy with an average of 89% across all weather stations compared to an average of 53% with ANN.
- Based on current accuracy, the KNN algorithm is likely the best algorithm, but The ANN algorithm can likely be adjusted to perform at higher levels than currently seen in testing.
- The ANN model **is more powerful and better suited for complex tasks** as compared to the KNN model
- The ANN model is more scalable for larger datasets and less sensitive to noise.
- the ANN model is recommended over the KNN model for weather predictions due to complex and nonlinear nature of the data.

Conclusion and Next Steps

Conclusion:

- While ANN shows promise, model generalization and interpretability must be improved for practical use by ClimateWins.

Recommendation:

- Based on overall performance, **ANN** with further tuning could offer the best balance between training and testing accuracy.

Next Steps

- Diversify data points to include other variables for analysis like hummidity, and air pressure.
- Continue testing adjustments to ANN model to increase model accuracy.
- Observe more weather stations to increase machine learning capabilities and Improve underperforming stations like Roma and Gdansk by filling data gaps.

Thank You for Your Time

If you have any questions, please feel free to contact me.

supriyapundlik@gmail.com