Unraveling Tomorrow: Predicting Europe's Climate Future with Machine Learning

Drew Riedlin

March 14, 2025

Overview

A New Frontier: Explore different advanced machine learning models, and experiment to find which are most useful and accurate in weather prediction

Decoding the Unknown: Uncover hidden patterns in global weather data

Tracing the Signals: Identify meaningful trends within decades of volatile weather data to enhance climate prediction and risk assessment

Data Profile

Weather data gathered from 18 European stations from the late 1800s to 2022, including daily records of temperature, wind speed, snow, and global radiation, sourced from the European Climate Assessment & Data Set project.

Potential Bias

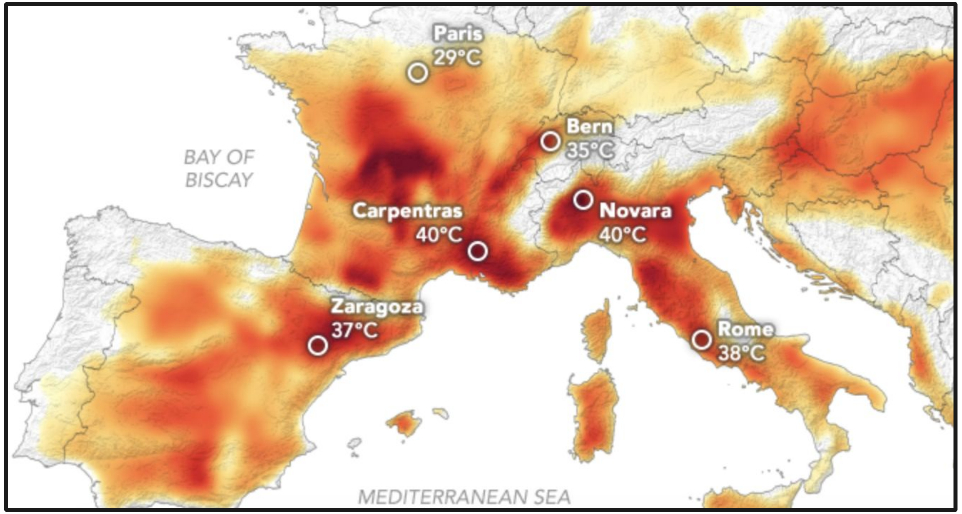
- **Regional:** Machine learning in climate analysis can worsen regional and cultural biases. This dataset only includes data from 18 specific weather stations which may not accurately represent the many climates and regions of Europe. This could lead to inaccurate predictions for unrepresented areas.
- **Systemic:** In climate data, human bias can be amplified in ML models, influencing predictions and policies. Political and economic interests may shape data, and confirmation bias can reinforce selective narratives.
- **Temporal:** The data set includes over 200 years of weather data. If the model relies on incomplete or biased historical data, it may overlook vulnerable regions, underestimate extreme events, or inaccurate predictions that could have serious consequences for disaster preparedness, infrastructure planning, and climate policy decisions.

Exploring Possibilities: Hypothesis

1. Using Machine Learning algorithms, future weather patterns can be predicted with accuracy based on historical weather data.

2. Prediction accuracy will vary depending on geographic location and the diverse climatic conditions present within the region.

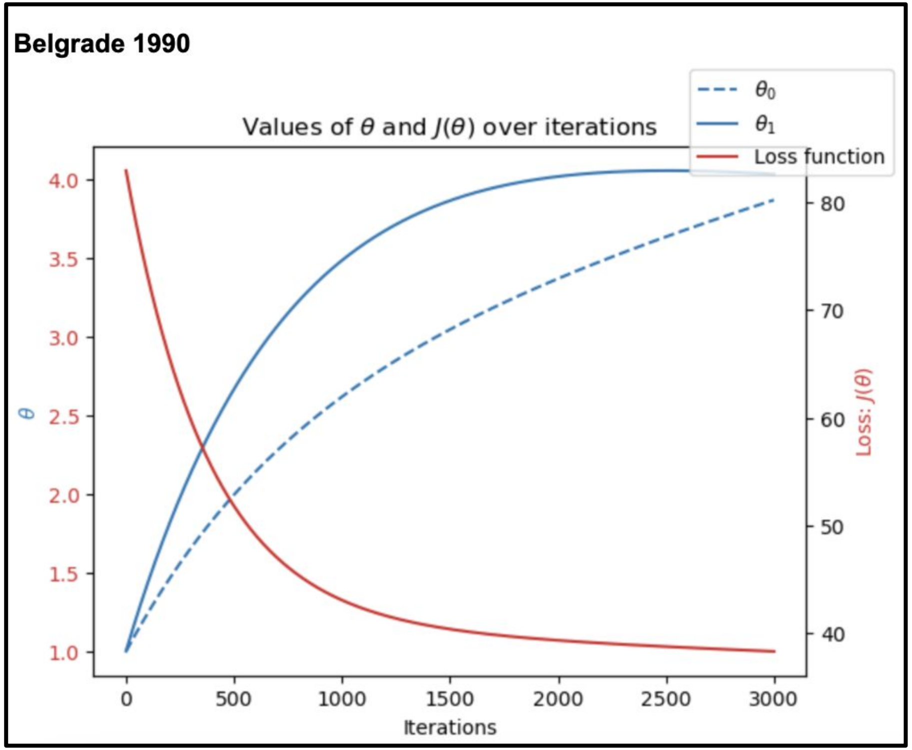
3. Machine learning will help reveal patterns in the climate data that can help detect climate change and uncover its hidden impacts.



Data Optimization

Gradient Descent was used to optimize the data.

- **Minimizes Error**: Iteratively adjusts model parameters to reduce the difference between predictions and actual values.
- **Finds Optimal Solutions**: Navigates the cost function to locate the point of lowest error or loss.
- **Updates Parameters**: Adjusts model weights by calculating gradients to improve prediction accuracy step-by-step.

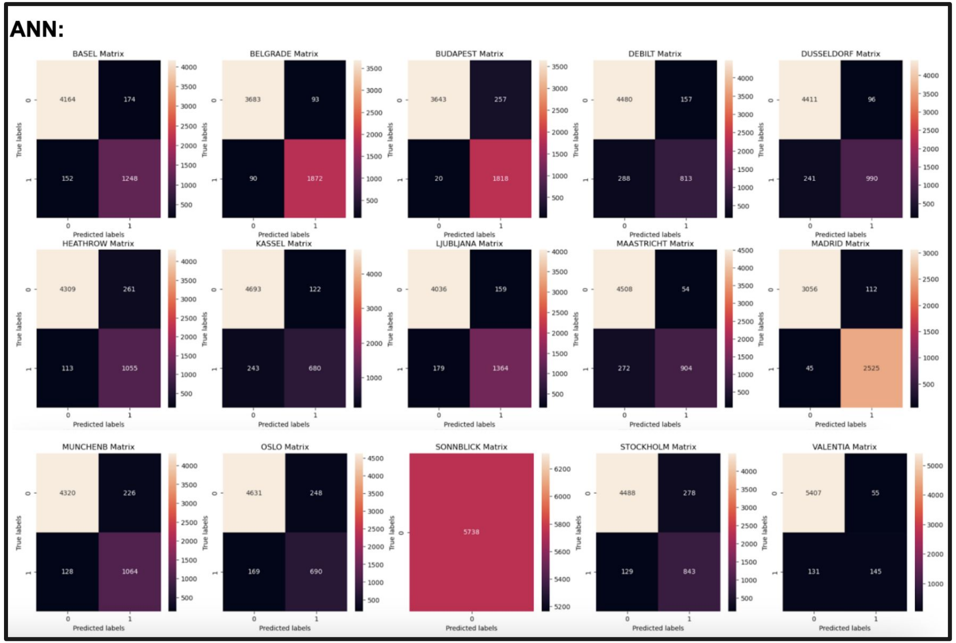


Decision Matrix

Used to evaluate and compare model performance and select the best algorithm based on multiple criteria, in this case pleasant vs unpleasant weather days.

- **True Positive = unpleasant weather predicted and unpleasant was correct**
- **True Negative = pleasant weather predicted and pleasant was correct**
- **False Positive = unpleasant weather predicted but weather was pleasant**
- **False Negative = pleasant weather predicted but weather was unpleasant**

Three supervised machine learning methods were used to compare accuracy and performance.



1. K-Nearest Neighbors (KNN) Algorithm

KNN finds the k closest data points (neighbors) based on a chosen distance metric, assigns the most common label among the K neighbors to the new data point, and then predicts the average value of the K neighbors.

Accuracy: 88%

Precision: 93%

Recall: 93%

F1 Score: 93%

Station	True Positive	True Negative	False Positive	False Negative	Accuracy	Precision	Recall	F1 Score
BASEL	3917	961	421	439	85%	90%	90%	90%
Belgrade	3252	1544	524	418	84%	86%	89%	87%
Budapest	3424	1462	476	376	85%	88%	90%	89%
Debilt	4320	723	317	378	88%	93%	92%	93%
Dusseldorf	4164	810	343	421	87%	92%	91%	92%
Heathrow	4138	744	432	424	85%	91%	91%	91%
Kassel	4563	614	252	309	90%	95%	94%	94%
Ljubljana	3740	1180	455	363	86%	89%	91%	90%
Maastricht	4253	824	309	352	88%	93%	92%	93%
Madrid	2750	2261	418	309	87%	87%	90%	88%
Munchenb	4237	792	309	400	88%	93%	91%	92%
Oslo	4637	512	242	347	90%	95%	93%	94%
Sonnblick	5738	0	0	0	100%	100%	100%	100%
Stockholm	4483	607	283	365	89%	94%	92%	93%
Valentia	5404	74	58	202	95%	99%	96%	98%
Total	63020	13108	4839	5103	88%	93%	93%	93%

2. Decision Tree Algorithm

Divides data into subsets based on the most significant feature, and based on this feature. Branches grow to show outcomes, and leaves represent the final prediction. It follows a top-down approach, making decisions at each step to reach a prediction.

Accuracy: 95%

Precision: 97%

Recall: 97%

F1 Score: 97%

Station	True Positive	True Negative	False Positive	False Negative	Accuracy	Precision	Recall	F1 Score
BASEL	4151	1237	187	163	94%	96%	96%	96%
Belgrade	3646	1855	130	107	96%	97%	97%	97%
Budapest	3740	1689	160	149	95%	96%	96%	96%
Debilt	4412	908	225	193	93%	95%	96%	95%
Dusseldorf	4477	1187	30	44	99%	99%	99%	99%
Heathrow	4268	925	302	243	91%	93%	95%	94%
Kassel	4672	778	143	145	95%	97%	97%	97%
Ljubljana	4138	1506	57	37	98%	99%	99%	99%
Maastricht	4336	993	226	183	93%	95%	96%	95%
Madrid	2994	2425	174	145	94%	95%	95%	95%
Munchenb	4387	1042	159	150	95%	97%	97%	97%
Oslo	4716	709	163	150	95%	97%	97%	97%
Sonnblick	5738	0	0	0	100%	100%	100%	100%
Stockholm	4518	708	248	264	91%	95%	94%	95%
Valentia	5303	109	159	167	94%	97%	97%	97%
Total	65496	16071	2363	2140	95%	97%	97%	97%

3. Artificial Neural Network (ANN)

Divides data into subsets based on the most significant feature, and based on this feature. Branches grow to show outcomes, and leaves represent the final prediction. It follows a top-down approach, making decisions at each step to reach a prediction.

Accuracy: 95%

Precision: 97%

Recall: 97%

F1 Score: 97%

Station	True Positive	True Negative	False Positive	False Negative	Accuracy	Precision	Recall	F1 Score
BASEL	4164	1248	174	152	94%	96%	96%	96%
Belgrade	3683	1872	93	90	97%	98%	98%	98%
Budapest	3643	1818	257	20	95%	93%	99%	96%
Debilt	4480	813	157	288	92%	97%	94%	95%
Dusseldorf	4411	990	96	241	94%	98%	95%	96%
Heathrow	4309	1055	261	113	93%	94%	97%	96%
Kassel	4693	680	122	243	94%	97%	95%	96%
Ljubljana	4036	1364	159	179	94%	96%	96%	96%
Maastricht	4508	904	54	272	94%	99%	94%	97%
Madrid	3056	2525	112	45	97%	96%	99%	97%
Munchenb	4320	1064	226	128	94%	95%	97%	96%
Oslo	4631	690	248	169	93%	95%	96%	96%
Sonnblick	5738	0	0	0	100%	100%	100%	100%
Stockholm	4488	843	278	129	93%	94%	97%	96%
Valentia	5407	145	55	131	97%	99%	98%	98%
Total	65567	16011	2292	2200	95%	97%	97%	97%

Algorithm Summary

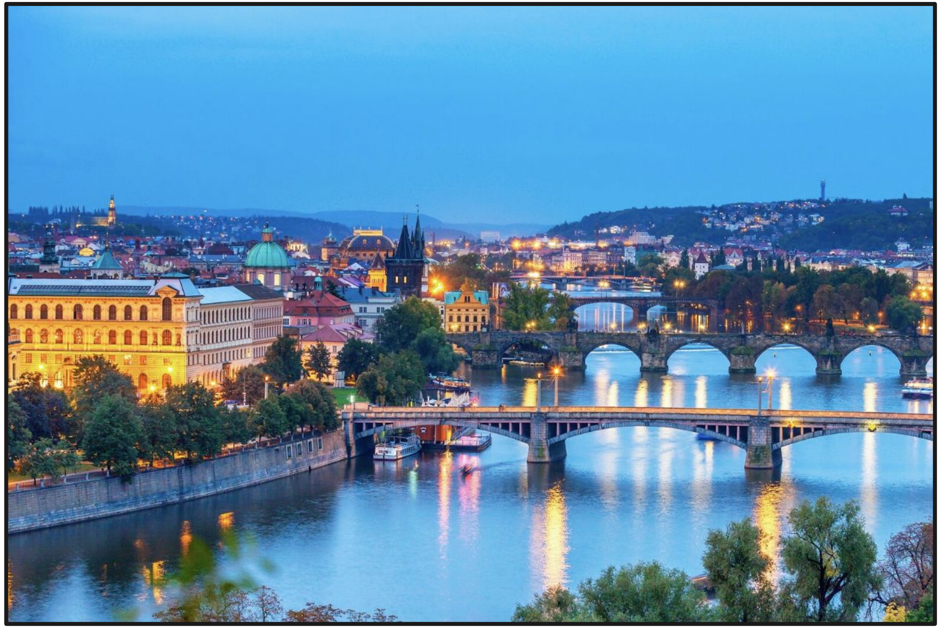
- All models were able to learn and produce results with a high level of accuracy. The Decision Tree and ANN models had the highest level of accuracy with 95% accuracy and a F1 score of 97%.
- **Decision Tree**: complex and would need to be trimmed.
- **KNN**: simple and best for smaller data sets. Not the best choice for this project because it will be computationally expensive to run.
- **ANN**: will be the best choice here because it is better suited to larger data sets and can help us predict the weather with high accuracy.

Hypothesis Recap

1. Machine Learning models are capable of predicting future weather trends with up to 95% accuracy.
2. The accuracy of the predictions made by the different machine learning algorithms fluctuated between the various stations and their different climate types.
    a. Example: All algorithms correctly predicted all days at the Sonnblick to be ‘unpleasant’.
3. Machine learning algorithms helped us reveal patterns in the weather data that may help determine climate change levels in Europe.

Next Steps

- Continue to test unsupervised machine learning algorithms, and experiment with unsupervised algorithms to find the best one for this project.
- Continue to make adjustments to optimize model accuracy.
- Utilize other variables, such as wind speed, humidity, air pressure, etc. to further enhance model accuracy.



Thank you.

For questions or further discussion, please reach out to me at drew.riedlin@gmail.com

Link to GitHub repositories

Check out my portfolio website