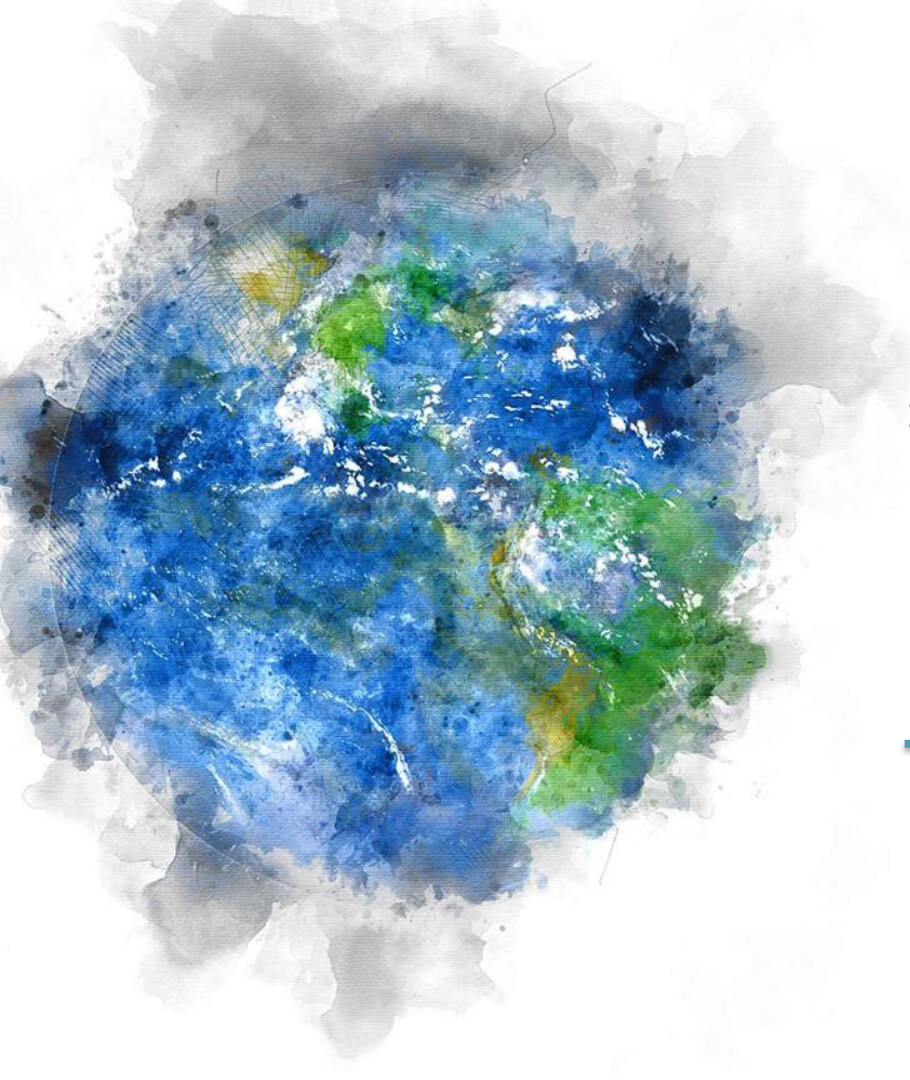


Analyzing Weather Conditions with ClimateWins

An interim report by Lisa Danen

Introduction

- ClimateWins is a fictitious European non-profit organization concerned with the increase in extreme weather events over the past 10-20 years.
- They think that weather extremes could be predicted and planned for.
- They have sorted through a great deal of data, including hurricane predictions from The National Oceanic and Atmospheric Administration (NOAA) in the U.S., and typhoon data from The Japan Meteorological Agency (JMA) in Japan.
- ClimateWins is interested in using machine learning to help predict the consequences of climate change around Europe and, potentially, the world.

Hypotheses

Machine Learning algorithms will be able to predict future weather conditions.

The accuracy of the predictions will vary based on location and regional climate conditions.

Data Set

- Based on observations from 18 weather stations across Europe
- Contains data ranging from the late 1800s to 2022
- Values include temperature, wind speed, snow, global radiation, and more
- Collected by the European Climate Assessment & Data Set project

Accuracy and Bias

Collection bias: Data collection tools have advanced significantly since the 1800s. The older data may be less accurate due to the tools used as well as potential human error.

Temporal bias: The large proportion of historical data compared to the last 10 to 20 years may lead to results that don't accurately reflect current conditions.

Regional bias: The data may be biased towards more developed and populated areas compared to less developed areas due to the location of the stations.

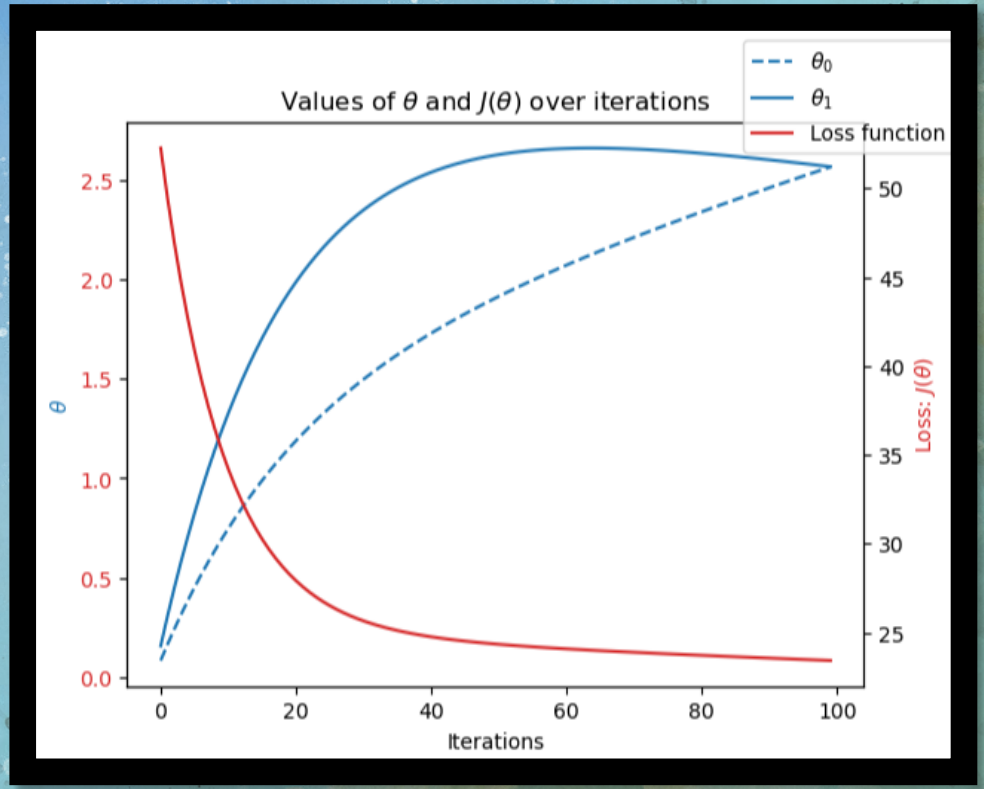
Human bias: It's important to avoid introducing personal assumptions or preconceived ideas into the process, as it may skew results.

Data Optimization

The data from multiple stations was optimized through Gradient Descent.

Gradient descent allows the analyst to adjust parameters to minimize the loss function.

The iterative adjustment of parameters increases accuracy.



Testing Machine Learning Models

Three Machine Learning models were tested to determine which one would be the most accurate at predicting “pleasant” days, when people could get outside and do “normal things”.

Models tested:

- K-Nearest Neighbor
- Decision Tree
- Artificial Neural Network.

K-Nearest Neighbor

This method was used to predict pleasant weather days utilizing the data from 15 of the weather stations.

The K-Nearest Neighbor (KNN) algorithm makes predictions by calculating the distance between new data points and other data points. Nearby data points are grouped based on the number of neighbors in each group.

This model was able to achieve an overall accuracy rate of 88%.

It is likely that overfitting is occurring with this model.

Station	Accurate Predictions		False predictions		Accuracy Rate
	Unpleasant (0)	Pleasant (1)	Unpleasant (0)	Pleasant (1)	
Basel	3907	935	465	431	84.38%
Belgrade	3238	1502	460	538	82.61%
Budapest	3416	1432	406	484	84.49%
Debilt	4346	732	369	291	88.50%
Dusseldorf	4167	800	431	340	86.56%
Heathrow	4161	754	414	409	85.66%
Kassel	4563	607	316	252	90.10%
Ljubljana	3726	1133	410	469	84.68%
Maastricht	4249	819	357	313	88.32%
Madrid	2735	2257	313	433	87.00%
Munchenb	4222	766	426	324	86.93%
Oslo	4624	507	255	352	89.42%
Sonnblick	5738	0	0	0	100%
Stockholm	4449	588	384	317	87.78%
Valentia	5391	108	168	71	95.83%
Overall accuracy					88.15%

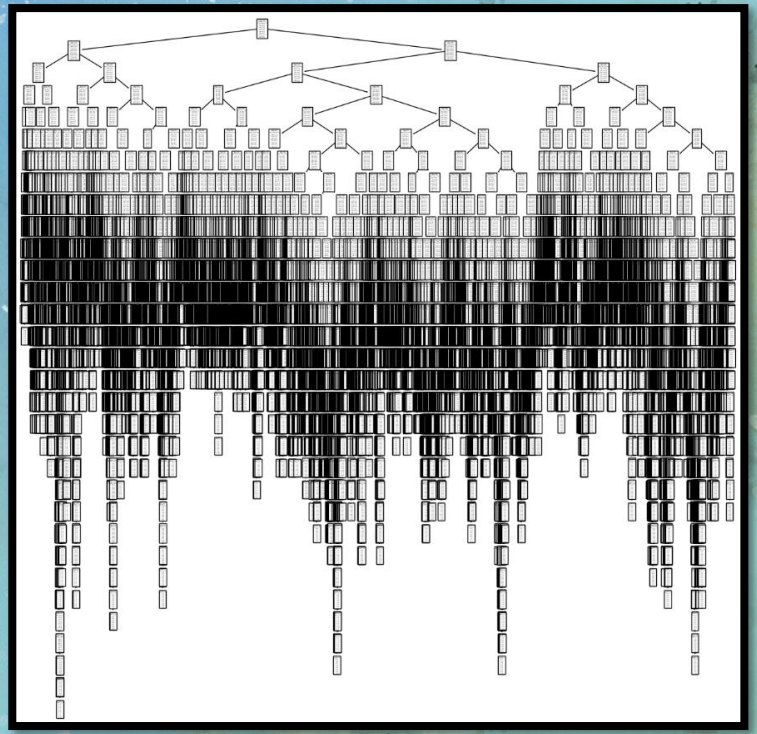
Decision Tree

The Decision Tree model makes predictions by starting answering a series of questions. First is the root questions, which has an “yes” or “no” answer. Below that, each question leads to another question (the branches) the stopping point answer is a leaf.

This model was used to predict pleasant weather days utilizing the data from 15 of the weather stations.

The Decision Tree model achieved 46% accuracy with the training data and 47% accuracy with the testing data.

The Decision Tree is extremely complex, with too many branches and would need to be pruned to be more useful.



Artificial Neural Network

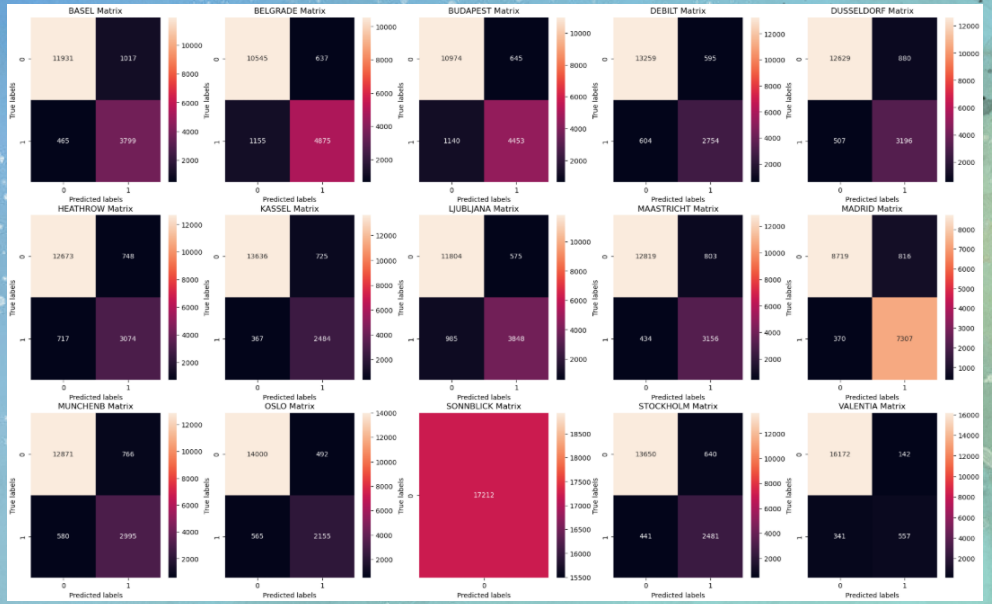
The Artificial Neural Network (ANN) model mimics the way a human brain processes information.

Layers of interconnected nodes learn from each other and learn patterns in the data through training.

This model was also used to predict pleasant weather days utilizing the data from 15 of the weather stations.

Multiple tests were run, adjusting the number of layers, layer sizes, and iterations to affect the model's ability to make accurate predictions.

Maximum accuracy achieved: 55%



A Confusion Matrix displays the accurate and inaccurate predictions produced. The image above shows all 15 matrices from one of the ANN model runs.

Upper left = accurate unpleasant prediction
Upper right = inaccurate unpleasant prediction
Lower left = inaccurate pleasant condition
Lower right = accurate pleasant prediction

Which Model is Best?

- The KNN model had the highest accuracy rate at an average of 88%.
- The Decision Tree was too complex for interpretation and showed a lower accuracy.
- The ANN model returned lower accuracy rates.
- Based on current testing, the KNN model is the best to use for predicting weather, due to its higher accuracy rate.
- It’s important to note, however, that the ANN model could likely be further adjusted to perform better and is generally well suited to complex tasks.

Summary and Next Steps

Summary

- Machine Learning algorithms are able to predict future weather conditions, with an accuracy rate up to 88%.
- Accuracy does vary by location and regional climate conditions.

Next Steps

- Incorporate unsupervised machine learning algorithms in pursuit of higher accuracy in results.
- Include other variables to test the accuracy of predictions based on factors other than temperature.

Thank you!

Questions? Please email me at
lisadanen@outlook.com

Check out my GitHub for scripts and datasets used in this analysis, as well as additional projects