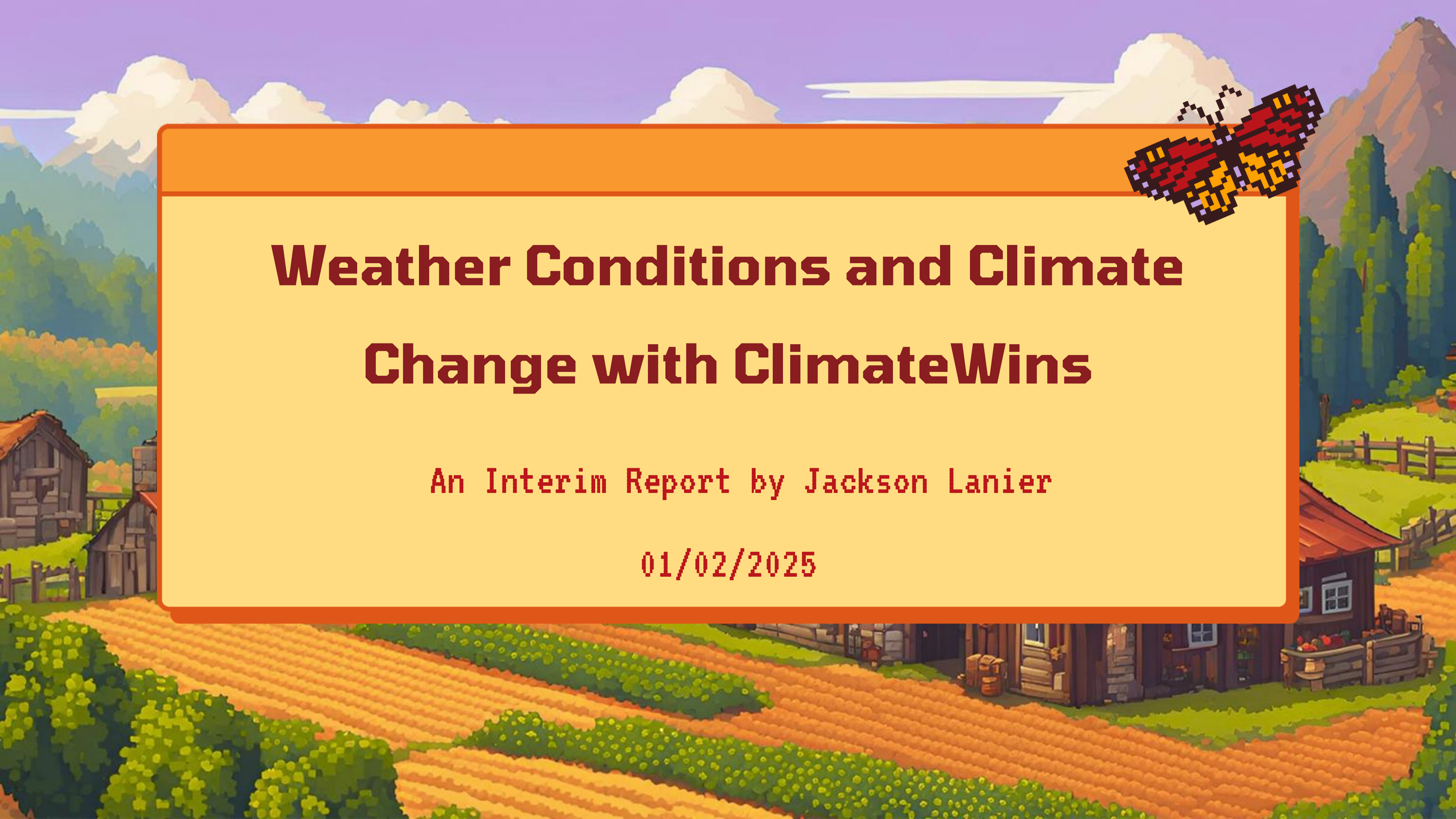


Weather Conditions and Climate Change with ClimateWins

An Interim Report by Jackson Lanier

01/02/2025

Why Should You Care?

Increases in extreme weather events over the last 10 to 20 years have prompted ClimateWins, a European nonprofit organization, to take action.

ClimateWin’s concern for our future has spawned a project that aims to use machine learning to help predict the consequences of climate change.

The Data

- Weather observations from 18 weather stations across Europe
- Collected by the European Climate Assessment & Dataset project
- Daily observations from 1960 through 2022
- Measures include precipitation, humidity, and temperatures
- No missing values
- Downloadable Dataset

Beware of Bias

- **Temporal Bias**: Data collected in the 1960s could suffer from outdated and potentially unreliable tools and methods
- **Historical Bias**: The large volume of historical data may overshadow rapid climate changes occurring in recent years, leading to inaccurate predictions for current and future conditions
- **Selection Bias**: The data might overrepresent more developed regions of Europe due to the location of the data collection stations, potentially excluding or underrepresenting less developed areas

Hypotheses

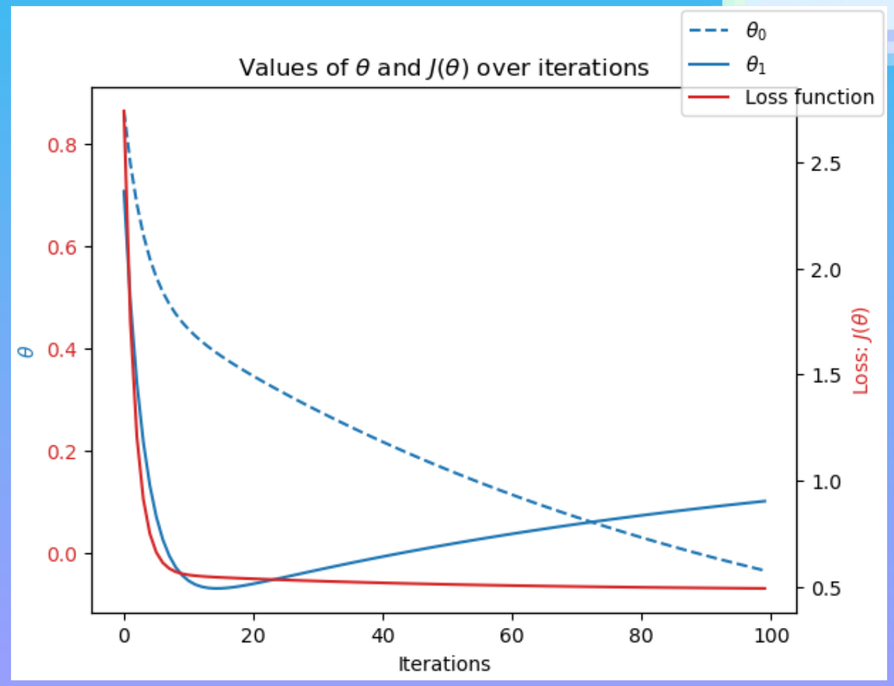
There is a correlation between rising temperatures and extreme weather events (heatwaves, significant precipitation).

Weather prediction accuracy will vary depending on location due to climate conditions in each location.

A single machine learning algorithm will not stand out as a “best choice” for predicting weather.

Optimization

- 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 2021. Notice the loss function's descent towards 0.

K-Nearest Neighbors

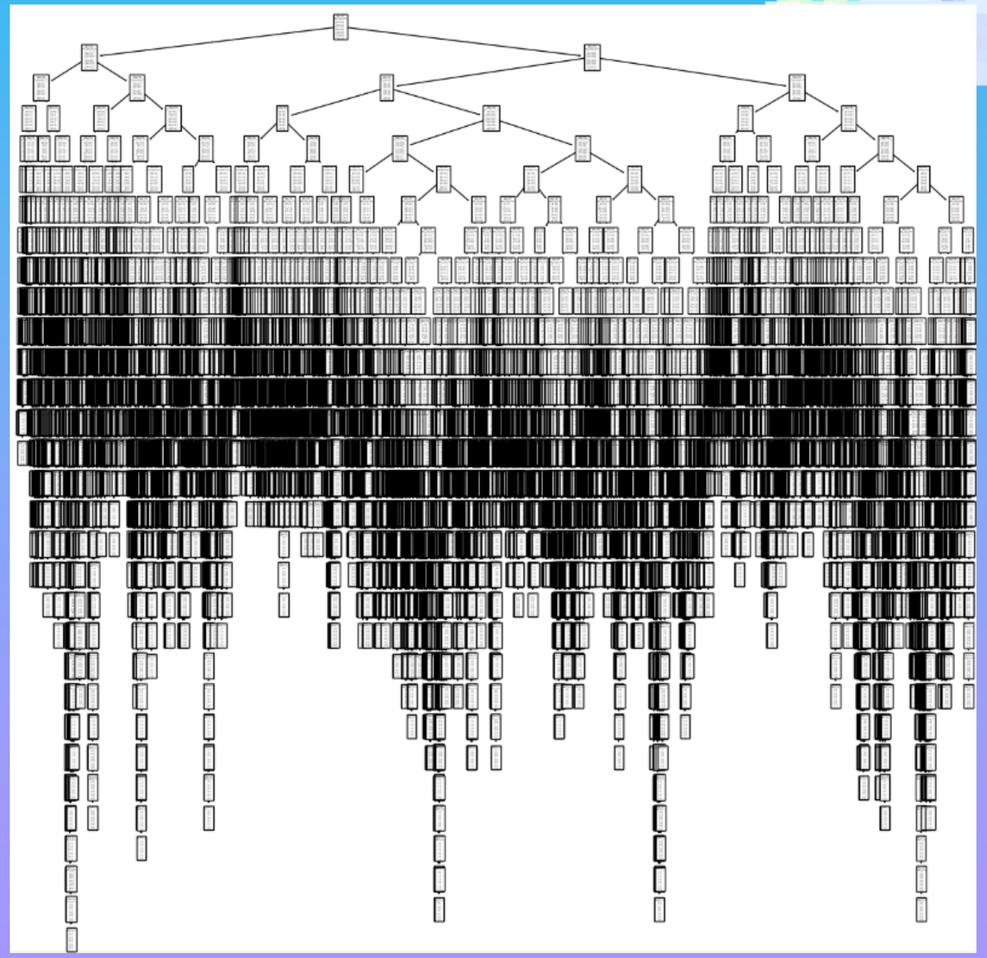
- The k-nearest neighbors (KNN) algorithm makes predictions by measuring the distance between new data points and nearby data points, or neighbors, then groups the new data with nearby data points based on the number of neighbors in each group.
- KNN was used to predict pleasant weather days in 15 of the European weather stations
- Using a k-range value of 3, the model was able to predict pleasant weather days with an average accuracy of 88%
- Overfitting is likely occurring in this model

Weather Station	Accurate Predictions		False Positives	False Negatives	Accuracy Rate
Basel	3907	935	431	465	84%
Belgrade	3238	1502	538	460	83%
Budapest	3416	1432	484	406	84%
De Bilt	4346	732	291	369	88%
Dusseldorf	4167	800	340	431	87%
Heathrow	4161	754	414	409	86%
Kassel	4563	607	252	316	90%
Ljubljana	3726	1133	469	410	85%
Maastricht	4249	819	313	357	88%
Madrid	2735	2257	433	313	87%
Munchen	4222	766	324	426	87%
Oslo	4624	507	255	352	89%
Sonnblick	5738	0	0	0	100%
Stockholm	4449	588	317	384	88%
Valentia	5391	108	71	168	96%
				Average:	88%

Notice Sonnblick's 100% accuracy rate. There are zero pleasant weather days in Sonnblick.

Decision Tree

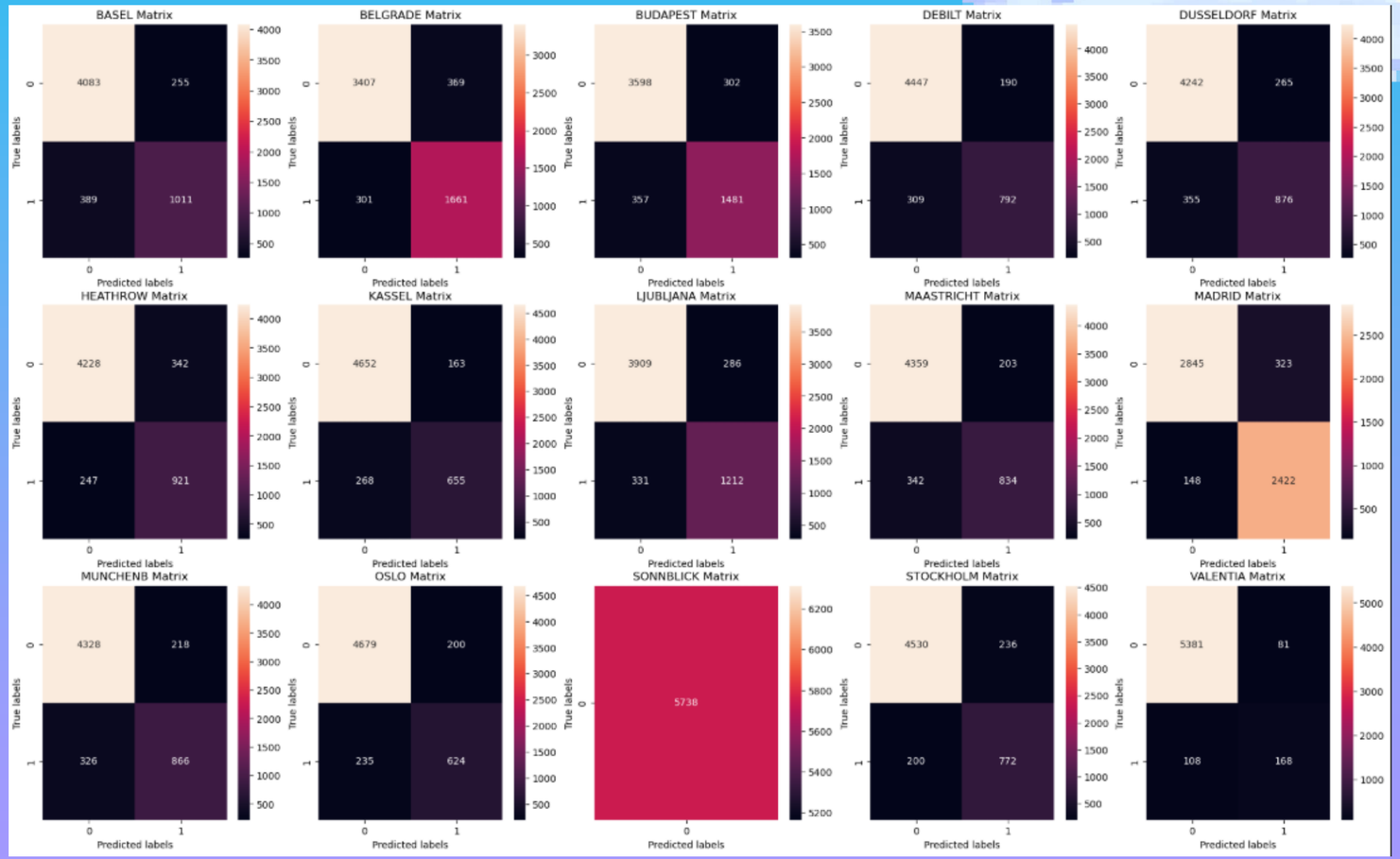
- The decision tree algorithm classifies data by starting with a root, or all of the data, and creating branches based on features of the data, similar to a flow chart. Predictions can be made based on new data's features by following the flow of the tree starting with the root, flowing into branches that best fit the new data's features, and finding an end point, or leaf, for the data.
- The decision tree model was used to predict pleasant weather days in 15 of the European weather stations
- The model performed with training data accuracy of 46.2% and testing data accuracy of 45.8%
- The decision tree needs to be pruned back



Notice the complexity of the tree. It appears to be suffering from overfitting.

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 the inputs are weighted based on their connection to hidden layers. Hidden layers apply functions to the weighted data and send the results to output layers.
- The ANN model was used to predict pleasant weather days in 15 of the European weather stations
- Several tests were run, with each test adjusting hidden layers, node sizes, iterations, and tolerance.
- Maximum testing data accuracy was 51%
- Training data accuracy consistently outperformed testing data accuracy



Above is a series of confusion matrices. 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).

Each weather station has its own matrix. Notice Sonnblick's solid matrix, which represents the ANN's 100% accuracy in predicting only unpleasant weather days for this location.

Which Algorithm is Best?

Based on current accuracy, the KNN algorithm is likely the best algorithm for predicting future weather.

The ANN algorithm can likely be adjusted to perform at higher levels than currently seen in testing.

Summary

Hypotheses

- Correlation between rising temps and extreme weather events
- Prediction accuracy will vary depending on location
- No algorithm will stand out as a “best choice”

Methods

- Focusing on temperature data from across Europe
- Optimization with gradient descent
- Supervised machine learning with KNN, decision tree, and ANN algorithms

Next Steps

- Continue exploring algorithms for improved accuracy
- Incorporate unsupervised machine learning algorithms
- Investigate variables other than temperature
- Ultimately, make predictions on future weather

Thank You!

For questions and further discussion, please contact me at jacksonlanier94@gmail.com

Check out my GitHub (github.com/jacksonlanier) for the Jupyter notebooks behind this work as well as additional projects!

View my video presentation on Youtube!

PowerPoint Template Credits

This presentation template is free for everyone to use thanks to the following:

Slides Carnival for the presentation template

Pexels and Pixabay For the photos