Analyzing Weather Conditions and Climate Change w/ ClimateWins

Brandon Lu

Introduction

- ClimateWins is interested in using machine learning to **help predict the consequences of climate change around Europe** and, potentially, the world.

- It's been **sorting through hurricane predictions** from The National Oceanic and Atmospheric Administration (NOAA) in the U.S., **typhoon data** from The Japan Meteorological Agency (JMA) in Japan, **world temperatures**, and a great deal of other data.

Hypothesis

MACHINE LEARNING ALGORITHMS WILL BE ABLE TO ACCURATELY PREDICT FUTURE WEATHER CONDITIONS.

PREDICTION ACCURACY WILL VARY BASED ON GEOGRAPHIC LOCATION AND VARIOUS CLIMATE CONDITIONS WITHIN THE REGION.

MACHINE LEARNING CAN IDENTIFY SIGNS OF CLIMATE CHANGE AND ITS ADVERSE EFFECTS.

Data Set Info

Collected by European Climate Assessment & Data Set Project

Based on weather observations from 18 different weather stations across Europe

Date range from late 1800s to 2022

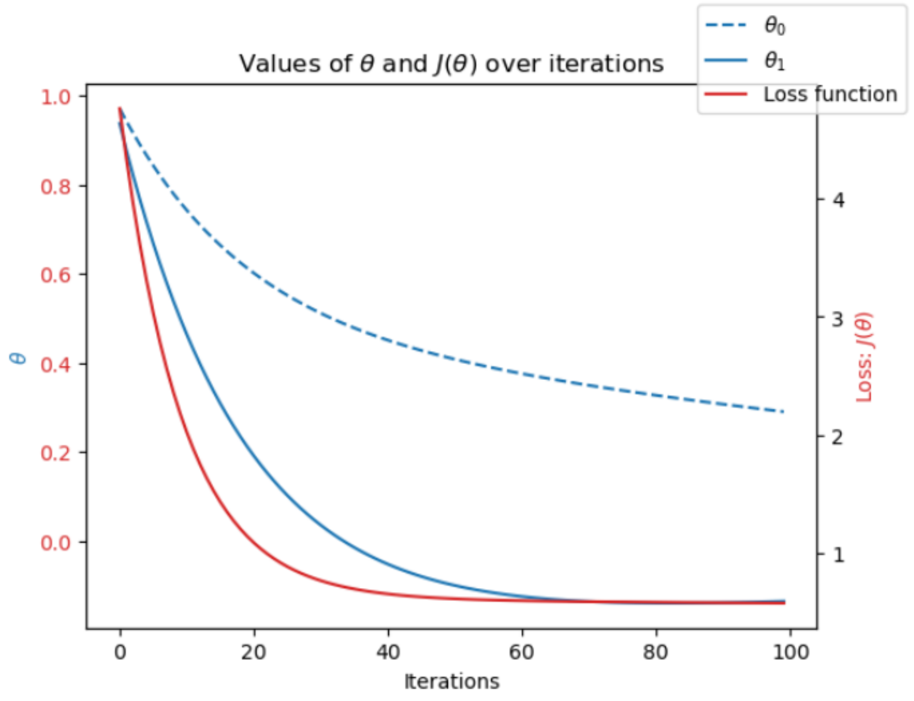
Data points include temperature, wind speed, snow, global radiation, etc

Data Bias & Accuracy

- **Data collection:** Older data may be prone to human error during collection
- **Data diversification bias:** Machine learning models will result in extreme bias towards larger regions and underrepresent smaller regions.
- **Temporal bias:** Over 200 years of data are being analyzed; older data may not reflected current weather conditions accurately
- **Data uniformity:** Data must be uniform for machine learning algorithms to make a proportionate comparison

Data Optimization

- The data was optimized through Gradient Descent.
- Gradient descent:
    - Adjusts parameters to minimize a loss function
    - Makes the model more accurate by reducing this loss through iterative updates to its parameters
    - Uses these adjustments to converge on optimal parameters, in this case, closer to 0



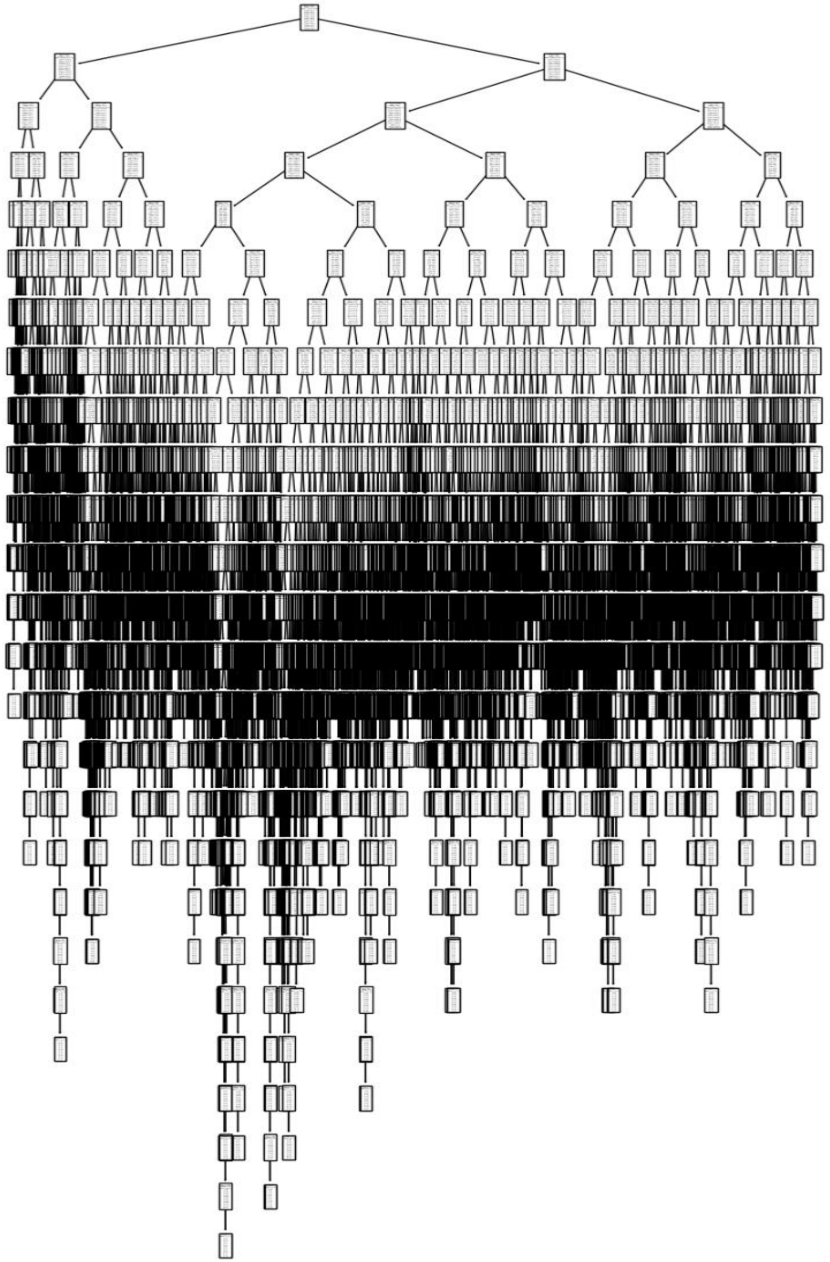
Method 1: K–Nearest Neighbor

- K-nearest neighbor algorithm(KNN) makes predictions by calculating the distance between new data points and all other data points in the training set. Nearby data points are grouped based on the number of neighbors in each group.
- **Average accuracy:** 89%
- To improve accuracy and remove the disparity between weather stations, the training model must incorporate for varied data and include a wider range of weather conditions across more regions.

Weather Station	Accurate predictions		False positive	False negative	Accuracy
BASEL	3981	1016	384	357	87%
BELGRADE	3329	1584	378	447	86%
BUDAPEST	3463	1491	347	437	86%
DEBILT	4388	755	346	249	90%
DUSSELDORF	4215	852	379	292	88%
HEATHROW	4198	768	400	372	86%
KASSEL	4599	655	268	216	91%
LJUBJANA	3808	1198	345	387	87%
MAASTRICHT	4319	855	321	243	90%
MADRID	2799	2300	270	369	89%
MUNCHENB	4259	837	355	287	89%
OSLO	4662	551	308	217	91%
SONNBLICK	5738	0	0	0	100%
STOCKHOLM	4497	656	316	269	90%
VALENTIA	5408	88	188	54	96%

Method 2: Decision Tree

- Decision tree makes predictions by answering a series of questions about the features of the data.
- The tree consists of nodes and branches in which new data points follow the flow of the tree based on its best fit.
- **Accuracy:** 64%
- The decision tree is very complex with too many branches which can lead to a model being too difficult to interpret.
- The tree is also very large, which can take a long time to train.



Method 3: Artificial Neural Network

- Artificial Neural Network (ANN) is a computer model that mimics the way the human brain processes information.
- It consists of layers of interconnected "neurons" that learn patterns from data through training.
- By adjusting connections between neurons, it improves its ability to make predictions or classifications.
- Adjustments to layer size and max iteration
- **Best accuracy:** 64%

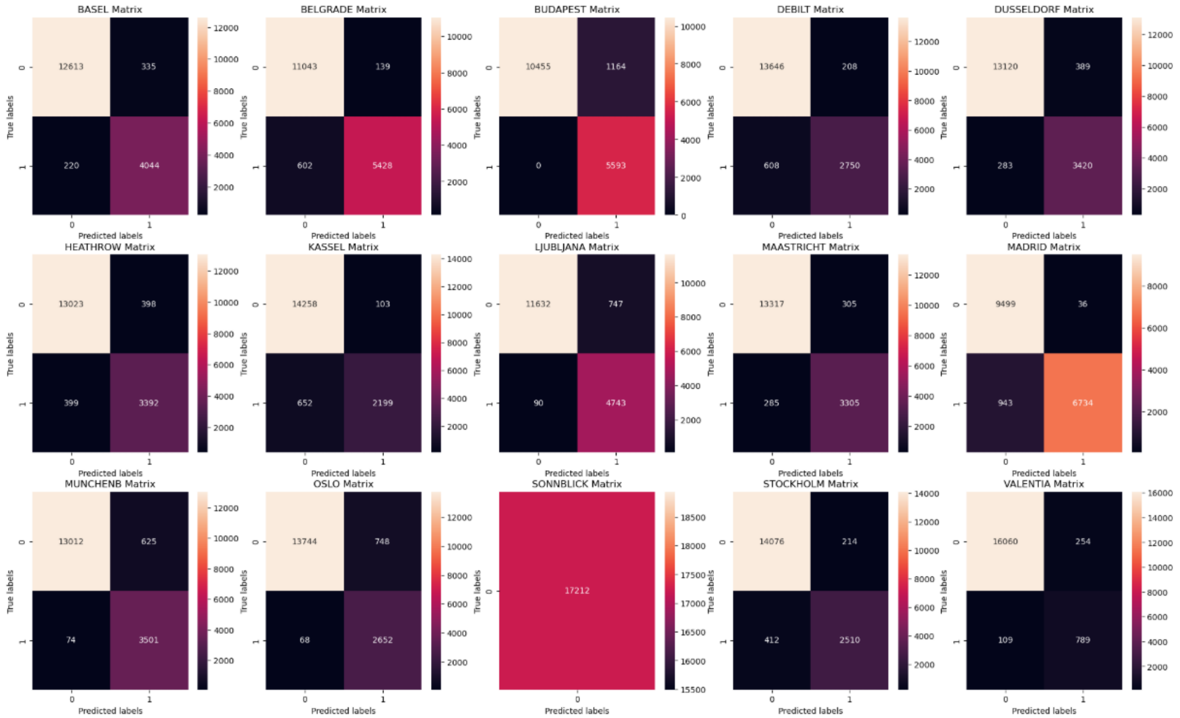


Fig. 1 - Confusion Matrix: Displays the accurate and inaccurate predictions for unpleasant and pleasant weather conditions.

Upper left = accurate unpleasant
Upper right = inaccurate unpleasant
Bottom left = inaccurate pleasant
Bottom right = accurate pleasant

Algorithm Recap

- The **<u>decision tree</u>** model is too complex to interpret.
- The **<u>KNN model</u>** provides the highest accuracy with an average of 89% across all weather stations compared to an average of 60% with ANN.
- The **<u>ANN model</u> is more powerful and better suited for complex tasks** as compared to the KNN model which allows more complex analysis.
- The ANN model is more scalable for larger datasets and less sensitive to noise.
- Because the weather data is complex and non-linear, the ANN model is recommended over the KNN model for weather predictions.

Summary

Machine learning algorithms can accurately predict future weather conditions with some models with accuracy up to 89%.

Prediction accuracy does vary based on geographic location and various climate conditions within the region.

For example, Sonnblick predicted 5738 unpleasant weather conditions compared to Madrid, which only predicted 2799.

Machine learning can identify signs of climate change and its adverse effects by comparing unpleasant weather condition over time.

Next Steps

- Incorporate unsupervised machine learning algorithms
- Diversify data points to include other variables for analysis
- Continue testing adjustments to increase model accuracy
- Observe more weather stations to increase machine learning capabilities

Thank you!

For any questions, please contact me at
brandonlu777@gmail.com

Check out my Github for the Python scripts and dataset used in this analysis