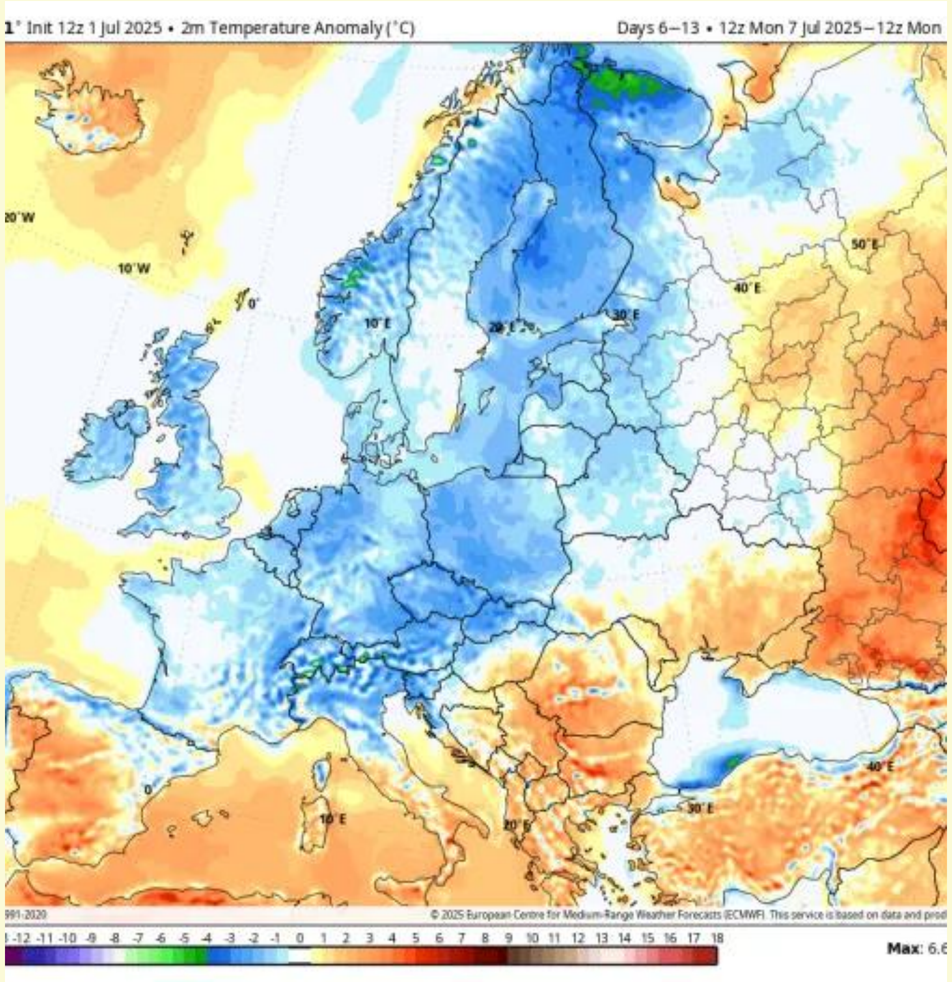
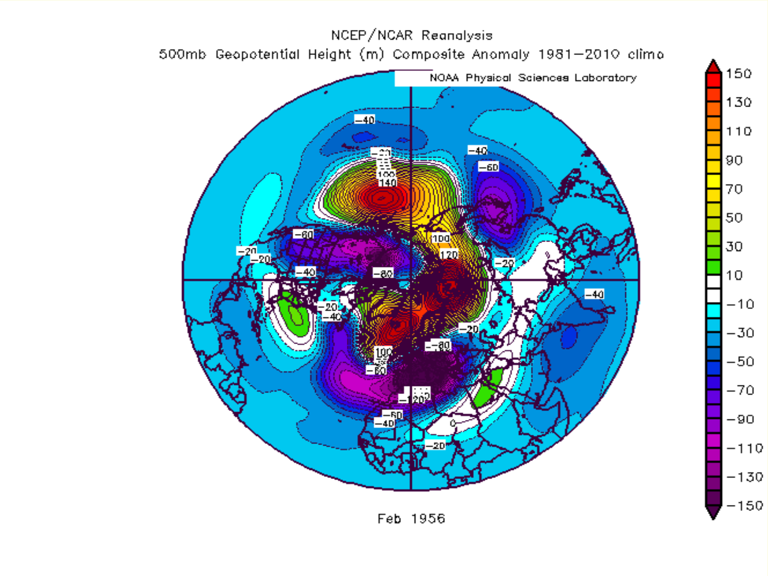




PREDICTING WEATHER VARIATIONS WITH MACHINE LEARNING FOR CLIMATEWINS

Aaron Gilbert July 2024

Introduction

As climate change intensifies, weather patterns worldwide have become increasingly unpredictable. An increase in extreme weather events has threatened the safety of many communities.

To protect human safety ClimateWins is embracing machine learning to achieve the following goals:

- Identify weather patterns outside the regional norm in Europe.
- Determine if unusual weather patterns are increasing.
- Generate possibilities for future weather conditions over the next 20 to 50 years based on current trends.
- Determine the safest places for people to live in Europe over the next 25 to 50 years.

Necessary Additional Data

Radar Imagery

Weather radar data corresponding to regions served by weather stations used in this analysis.

Dangerous Weather Classifications

Indications of what constitutes hazardous weather, to be generated using the above datasets.

Topographical and Climate Zone Data

Essential for understanding how physical geography impacts regional weather patterns.

Real-time Weather Station Data

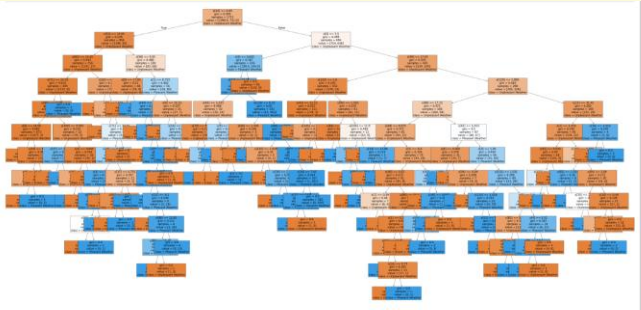
Allows for immediate updates to models, improving the accuracy of forecasts.

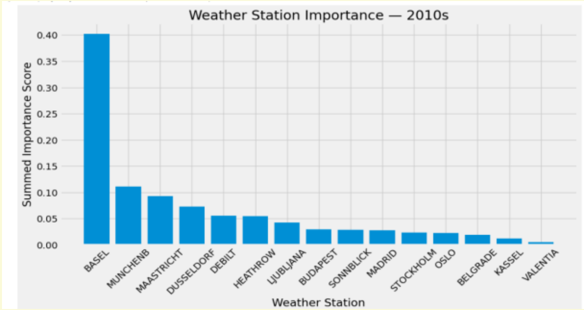
Overview of Machine Learning Approaches

Random Forest Model

Benefits:

- Averaging process reduces overfitting.
- Provides clarity through visualizations.

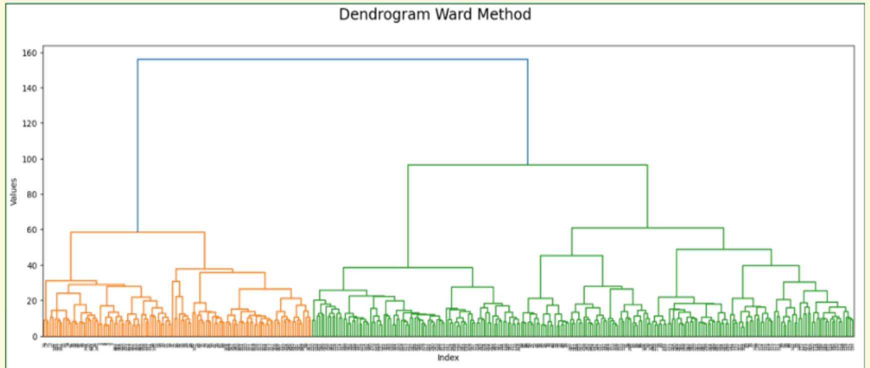




Hierarchical Clustering

Benefits:

- Generates new classifications from existing data.
- Does not require pre-categorization.



General Adversarial network (GAN)

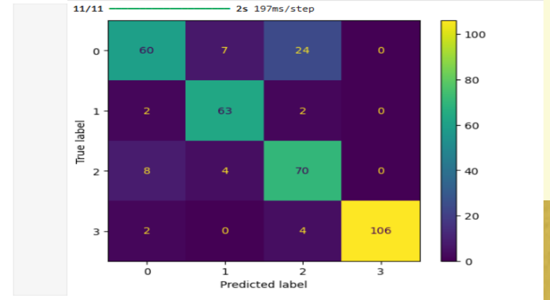
Benefits:

- Can create weather data that looks and behaves like observed conditions.
- Increases the amount of data available for modeling.

Convolution Neural Network (CNN)

Benefits:

- Effectively analyzes images, including radar-generated graphics.
- Capable of analyzing complex data.



Three Thought Experiments

Generate Accurate Predictions

Using a generative adversarial network (GAN) to synthesize weather data, we can train a convolution neural network (CNN) to predict possible future conditions over the next 50 years.

Classify Unusual Weather

Using hierarchical clustering, we can go beyond classifying atypical vs typical weather to find actionable categorizations.

Identify Great Living Regions

By optimizing a random forest model, we can identify the most important weather features to track and use that information to identify safe living regions.

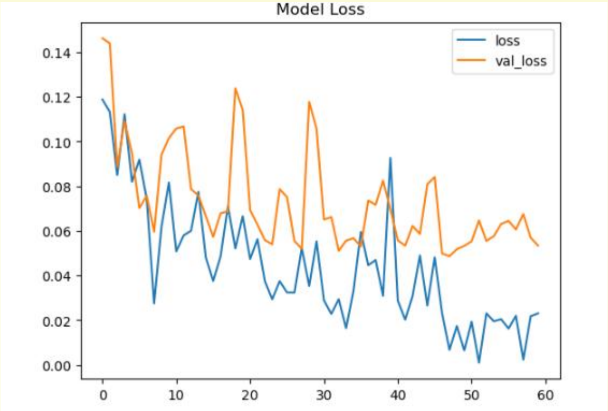
Thought Experiment #1: Synthesizing Data to Improve Predictions

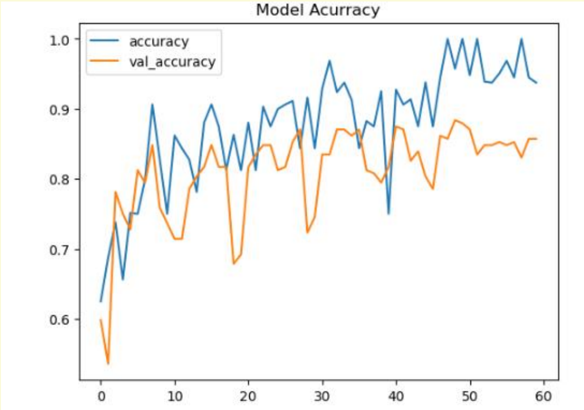
Hypothesis: Using a GAN to synthesize weather data, we can train a CNN to predict possible conditions over the next 50 years.

Objective: Generate possibilities for weather conditions over the next 25 to 50 years, based on current trends.

Approach: 1.Use a GAN to synthesize realistic weather data.

Results: The GAN produced weather data with 93.75% accuracy and 0.02 loss, suggesting high potential for use with the CNN model





Next Steps

- Acquire recommended radar data and run Bayesian optimization.
- Run the optimized CNN model using the GAN-generated data.
- Analyze results across years, countries, and regions.

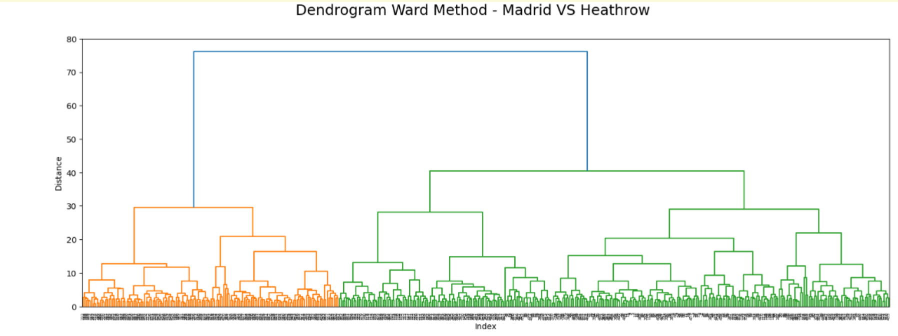
Thought Experiment #2: Classifying Weather Using Hierarchical Clustering

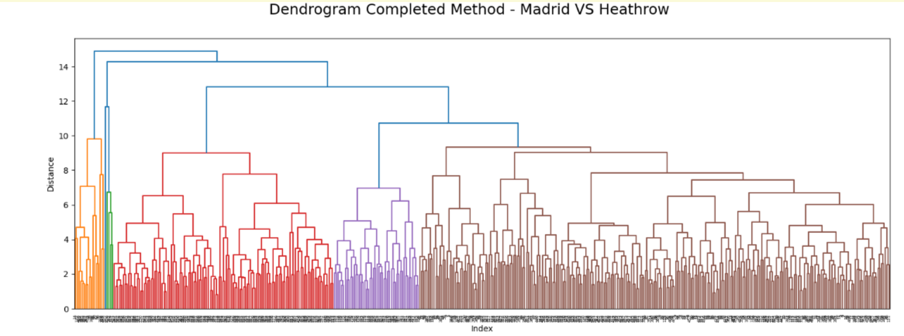
Hypothesis: Using hierarchical clustering, we can go beyond classifying atypical vs typical weather to find actionable categorizations.

Objective: To identify whether unusual weather patterns are occurring

Approach: To assess the viability of this model, we created a dendrogram using existing weather station data. We employed principal component analysis to reduce number of dimensions for optimal resource use.

Results: The model produced between two and three clusters consistently, potentially representing binary divisions or low-mid-high divisions.





Next Steps

For this analysis to be valuable, we need to bring in additional types of modeling. We recommend:

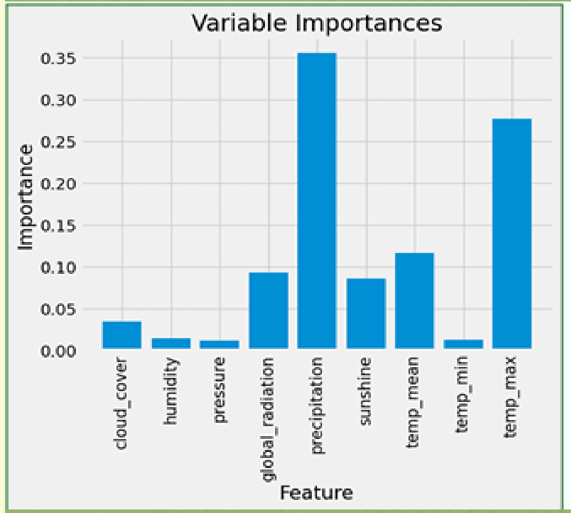
- Running this model on multiple years, seasons, and months of data to compare results
- Using the results to support other, more conclusive model types

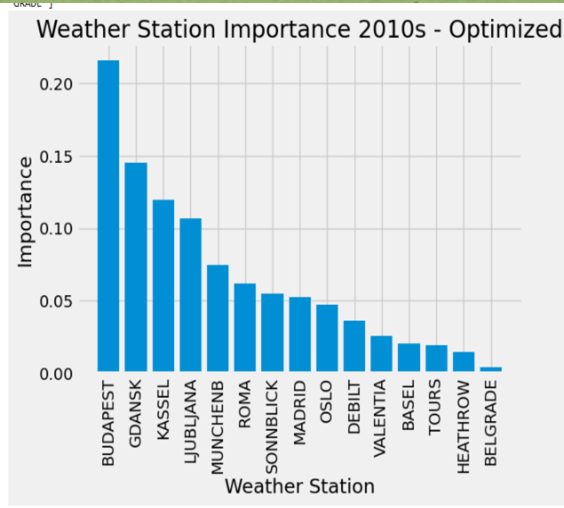
Thought Experiment #3: Assess Regions Using a Random Forest Model

Hypothesis: By optimizing a random forest model, we can identify the most important weather features to track and use that information to identify safe living regions.
Objective: Determine the safest places for people to live in Europe over the next 25 to 50 years.
Approach: 1. Refine hyperparameters for a random forest model using grid and random search methods
2. Compare unoptimized and optimized results of weather data analysis

Results:
- Optimization increased the accuracy of results from 59.1% to 65.93% when looking at all stations
- Optimization decreased from 98.8% to 92% when assessing a single station, suggesting a corrected error from the unoptimized model
- Running the optimized and unoptimized models allowed us to weigh feature importance in climate analysis

Next Steps
- Run optimization processes on additional stations and decades
- Run optimized model on extreme weather data





Recommendations

Data suggests that ClimateWins will derive the most value from investing resources in GAN and CNN optimization to predict weather patterns:

- This experiment showed the highest jumps in accuracy.
- The project showed immediate potential for achieving key goals.
- This approach has proven real-world success in improving severe weather predictions.

Next Steps

- Predictive Modeling: Implement LSTM model for time series forecasting of future climate patterns.
- Recommend adaptive planning based on model outcomes (urban development, migration risk zones, etc.).
- Collect long-term historical and recent climate data (temperature, precipitation, extreme events) across European regions.
- Adjust hyperparameters to optimize predictive accuracy and reduce overfitting.

Thank You!

FOR FURTHER QUESTIONS, PLEASE CONTACT ME AT:

AARON_GILBERT101@YAHOO.COM

TO VIEW THE SCRIPTS AND DATASETS USED, PLEASE VISIT MY GITHUB REPOSITORY FOR THIS ANALYSIS AT:

HTTPS://GITHUB.COM/AARONGILBERT9/CLIMATEWINS-PREDICTION-/TREE/MAIN