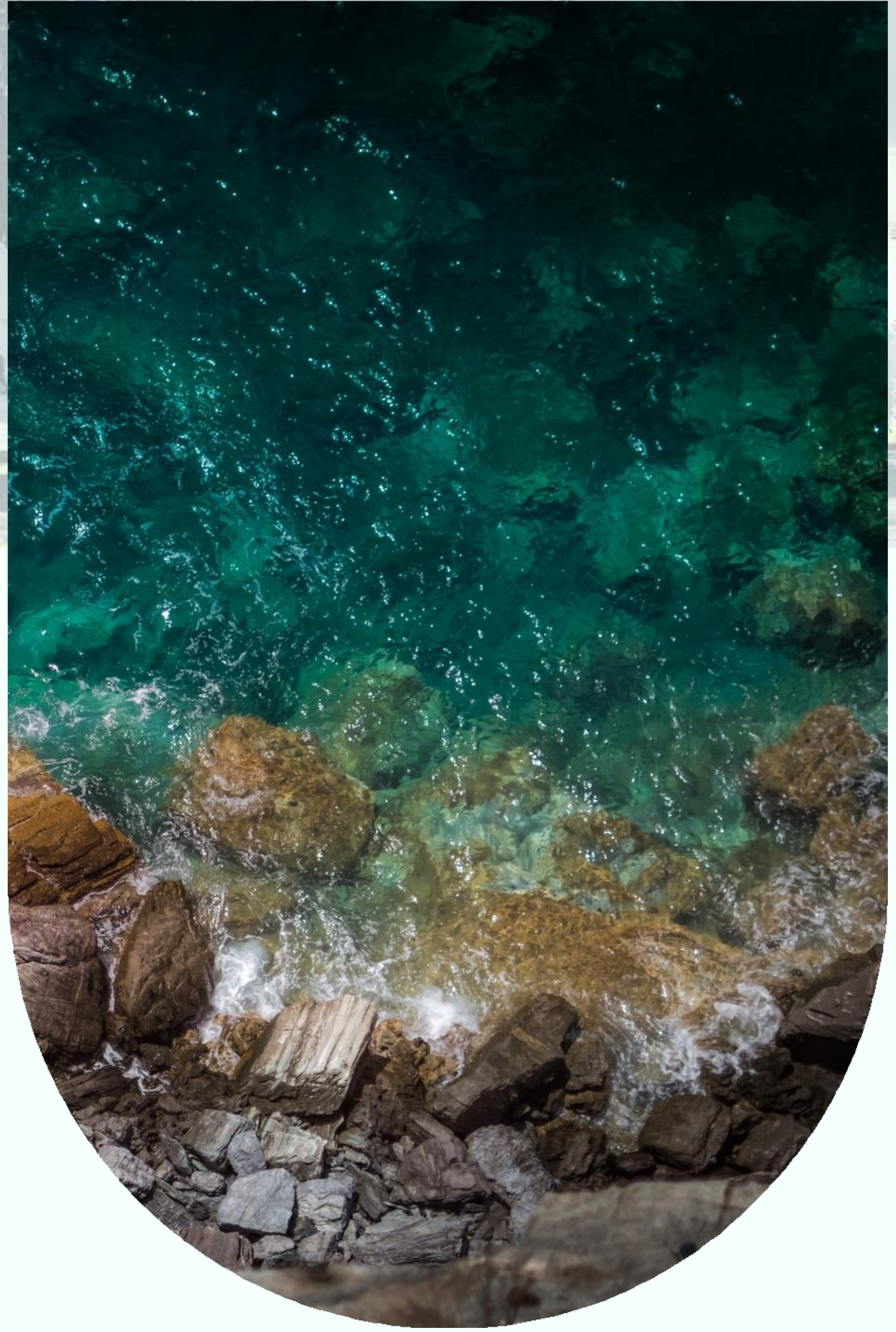
CLIMATE WINS: PREDICTING CLIMATE CHANGE WITH MACHINE LEARNING



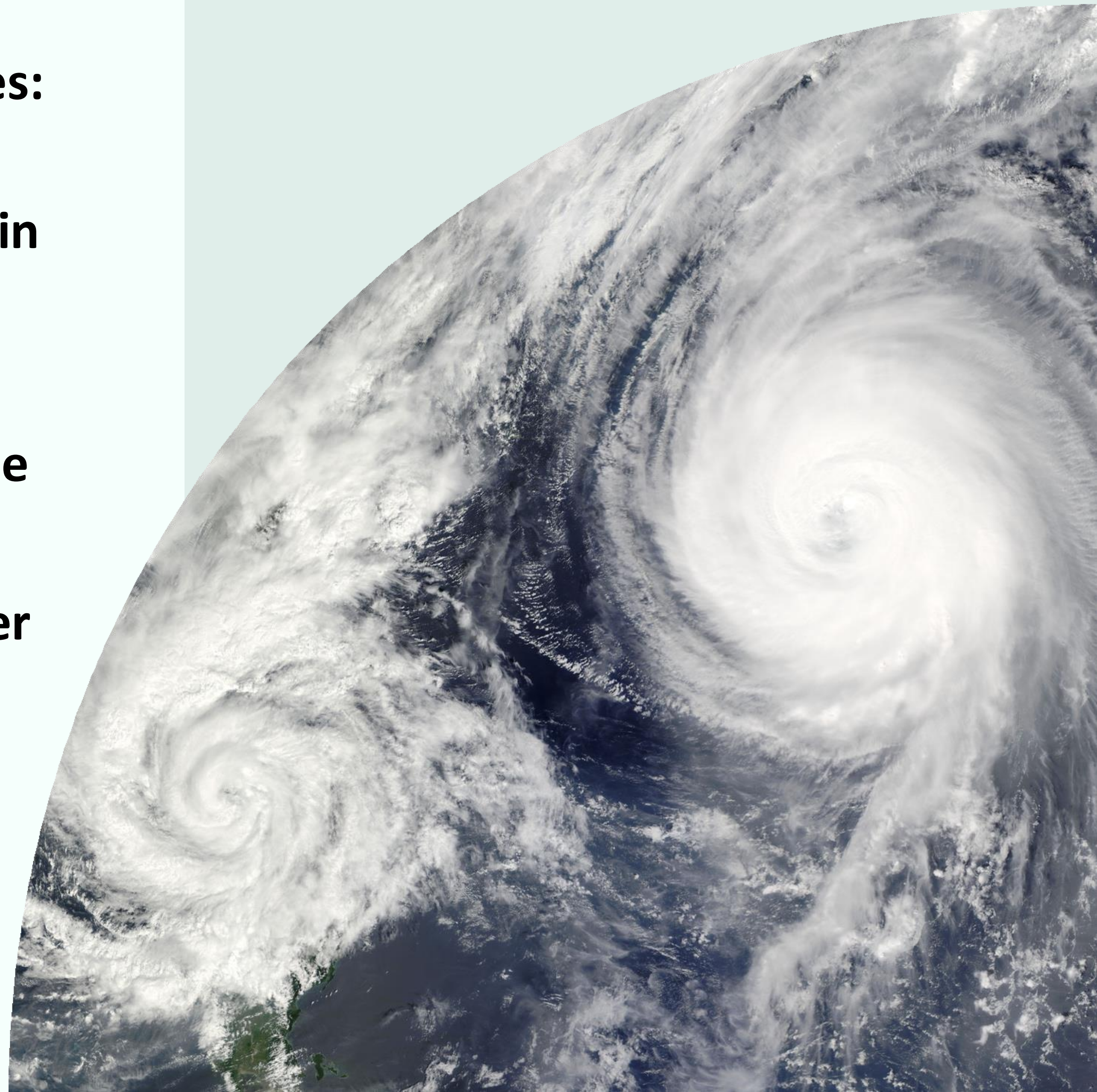
Shaival Mehta

Introduction

As climate change intensifies, global weather patterns have become increasingly erratic, leading to a rise in extreme weather events that pose significant threats to community safety.

To safeguard human safety and well-being, ClimateWins is leveraging machine learning to achieve the following objectives:

- 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



Three Thought Experiments

Classify Unusual Weather

By applying hierarchical clustering, we can move beyond a simple typical-versus-atypical classification to uncover meaningful and actionable weather pattern groupings.

Generate Accurate Predictions

By leveraging a Generative Adversarial Network (GAN) to synthesize realistic weather data, we can train a Convolutional Neural Network (CNN) to forecast potential weather conditions over the next 50 years.

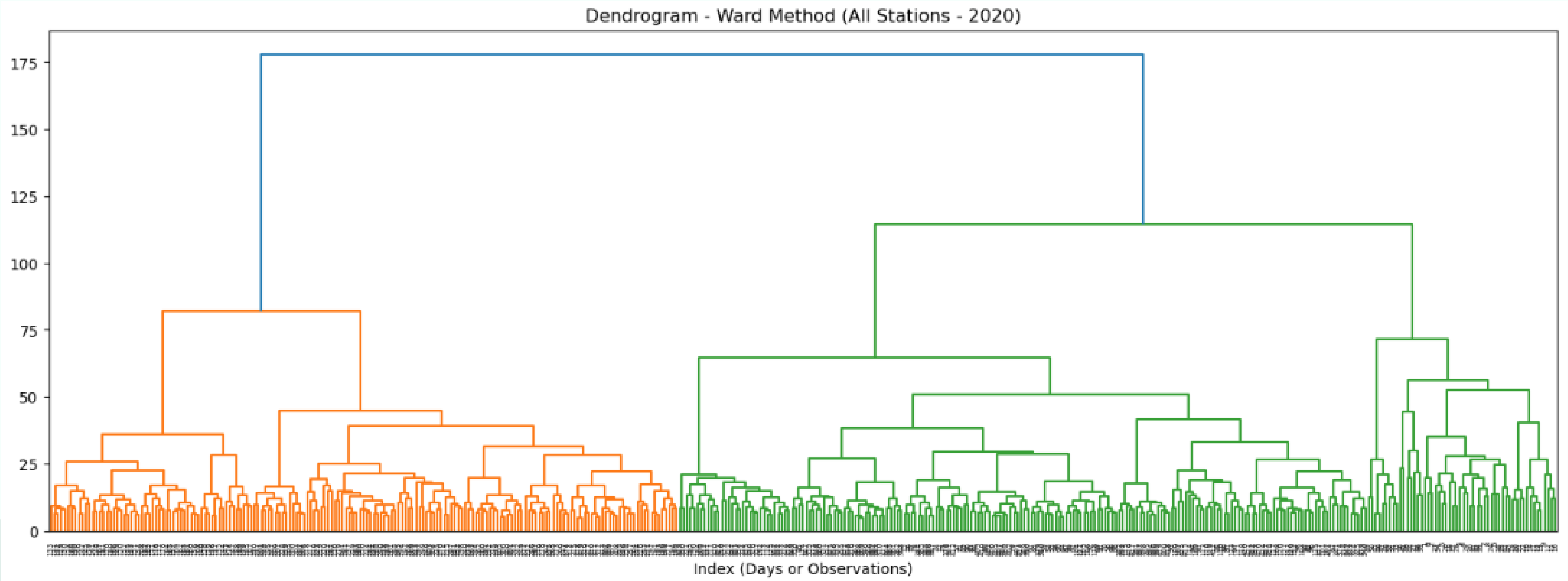
Identify Safe Living Regions

By optimizing a random forest model, we can pinpoint the most influential weather variables and leverage that insight to identify regions with safer living conditions.

Required Machine Learning Models

Hierarchical Clustering

Each data point is categorized by merging similar ones—'leaves'—into branches based on shared characteristics, forming an inverted tree structure where categories are visually distinguished by color.

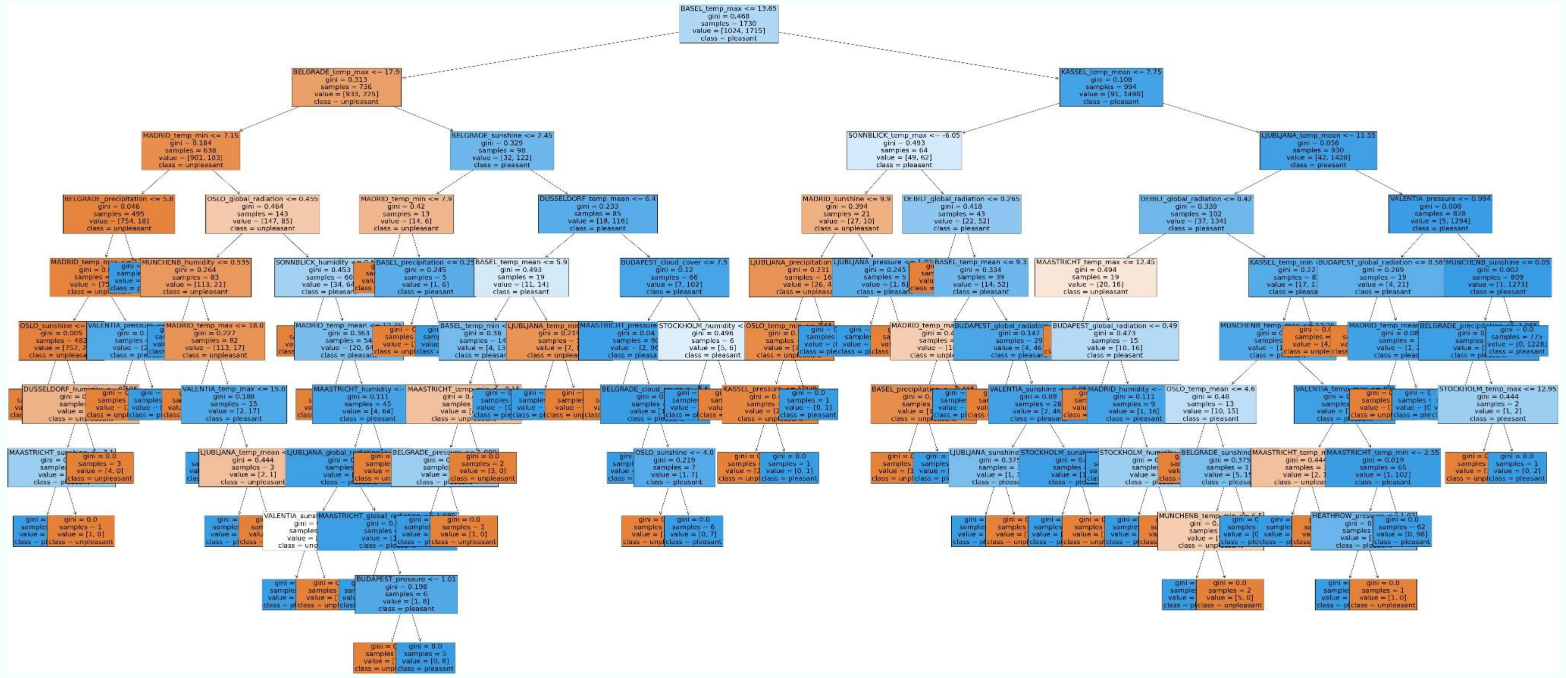


Benefits:

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

Random Forest

Utilizes multiple decision trees to classify data points through a series of true/false splits. A random forest trains each tree on a random subset of the data and combines their outputs to generate a final, more accurate prediction.



Benefits:

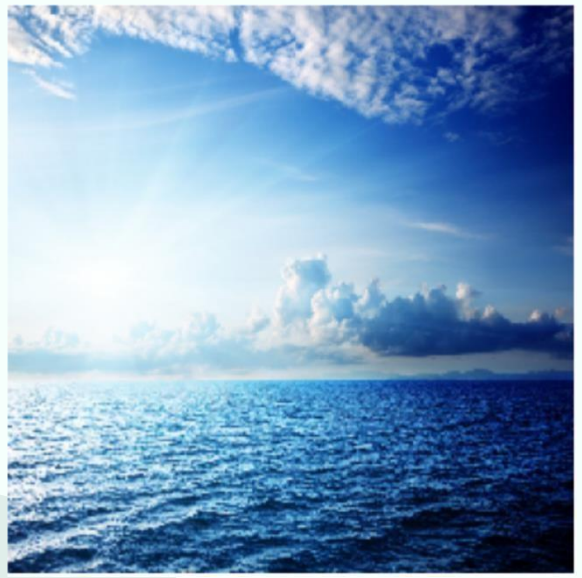
- Averaging process reduces overfitting
- Provides clarity through visualizations

Required Machine Learning Models

General Adversarial Network (GAN)

A generator network creates synthetic data modeled after real examples, while a discriminator network evaluates both real and generated data to distinguish between them. The generator learns from this feedback to produce increasingly realistic 'fakes.'

Correct Prediction - class: Shine - predicted: Shine



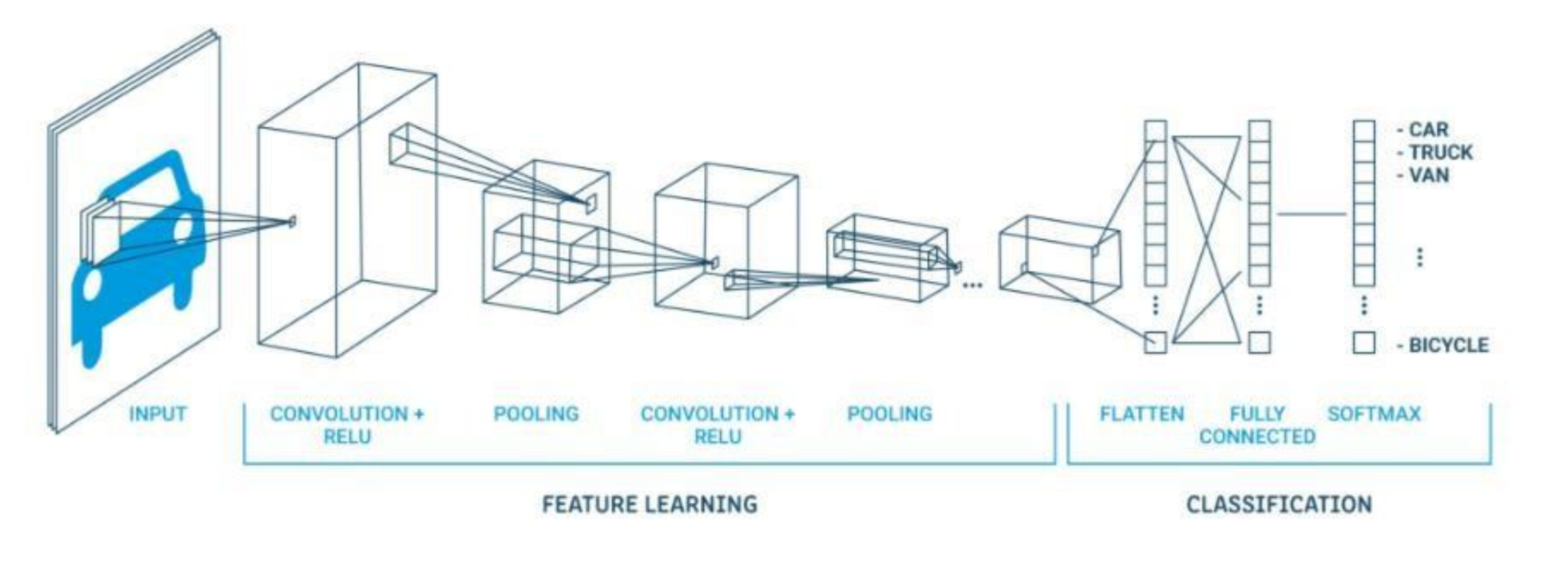
Correct Prediction - class: Cloudy - predicted: Cloudy

Benefits:

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

Convolution Neural Network (CNN)

A CNN uses hidden layers to learn key features from a dataset and generate classifications. Pooling layers within the network compress information by capturing average and maximum values, helping reduce computational load.



Benefits:

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

Image Credit: CareerFoundry

Necessary Additional Data

Weather Event Data

Records of extreme weather events across Europe: Storms, excessive heat, and extreme cold

Radar Imagery

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

Healthcare Data

Information on illnesses, injuries, and deaths attributed to extreme weather

Dangerous Weather Classifications

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

Thought Experiment #1: Classifying Weather Using Hierarchical Clustering

Hypothesis

Hierarchical clustering enables us to move beyond simply labeling weather as typical or atypical, uncovering meaningful and actionable weather categories.

Objective

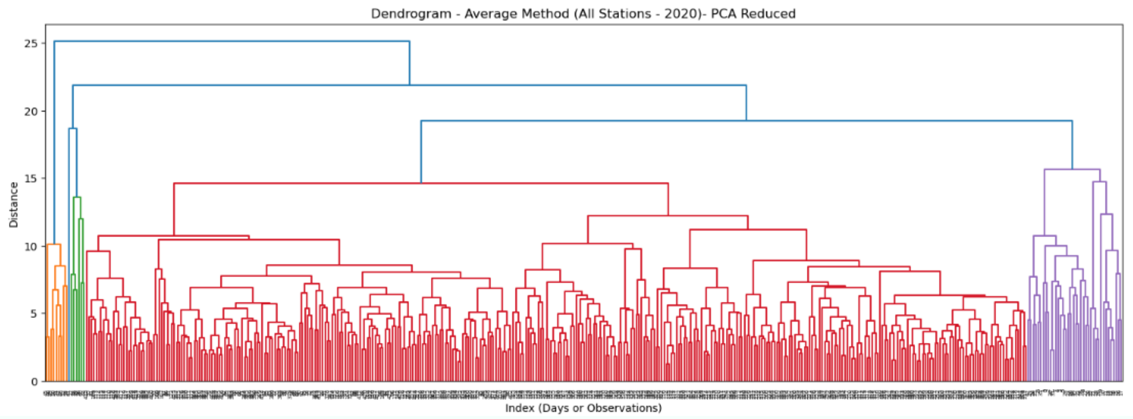
To identify whether unusual weather patterns are occurring.

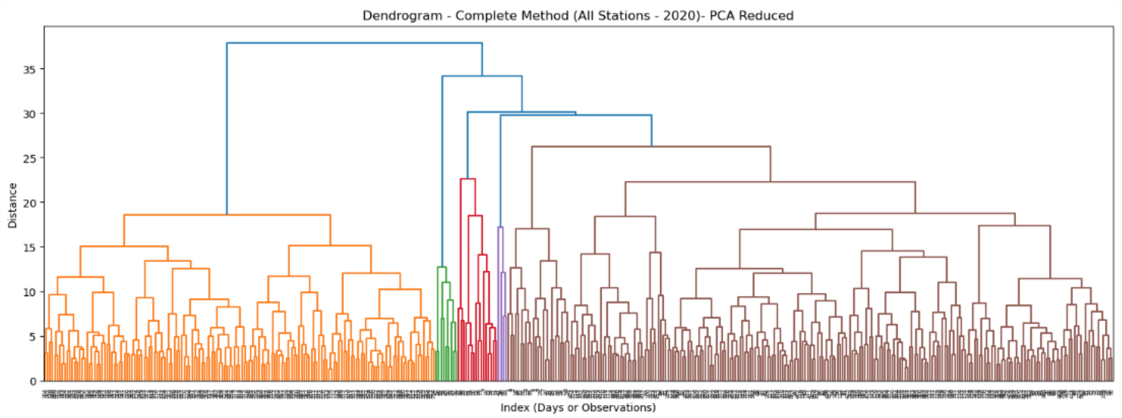
Approach

To evaluate the model's effectiveness, we generated a dendrogram from existing weather station data. Principal component analysis was applied to reduce dimensionality, optimizing resource utilization.

Results

The model consistently generated two to three clusters, which may correspond to binary categories or a low–mid–high classification scheme.





Next Steps

To enhance the value of this analysis, incorporating additional modeling approaches is essential. 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 #2: Synthesizing Data to Improve Predictions

Hypothesis

By using a GAN to synthesize weather data, we can train a CNN to forecast potential weather conditions up to 50 years into the future.

Objective

This approach generates plausible scenarios based on current climate trends for the next 25 to 50 years.

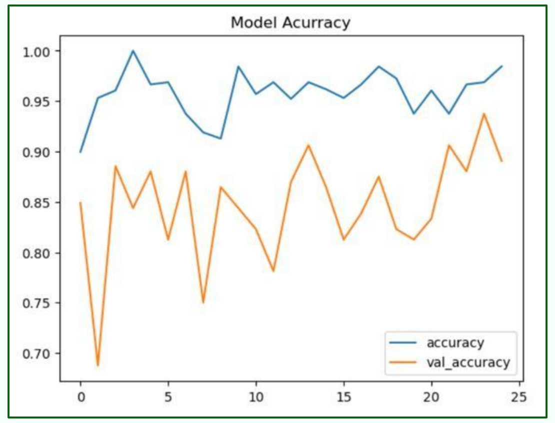
Approach

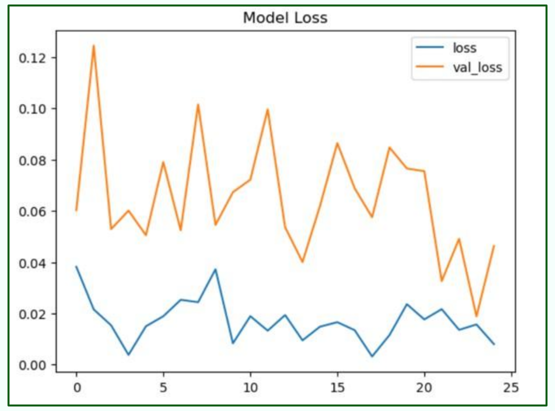
1. Run optimization on a CNN model to evaluate its accuracy
2. Use a GAN to synthesize realistic weather data

Results

Bayesian optimization improved the accuracy of CNN modeling to 66.3%

The GAN produced weather data with 98.4% accuracy and 1.1% 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 #3: Assess Regions Using a Random Forest Model

Hypothesis

Optimizing a random forest model enables us to pinpoint the key weather features to monitor, which in turn helps identify regions with safer living conditions.

Objective

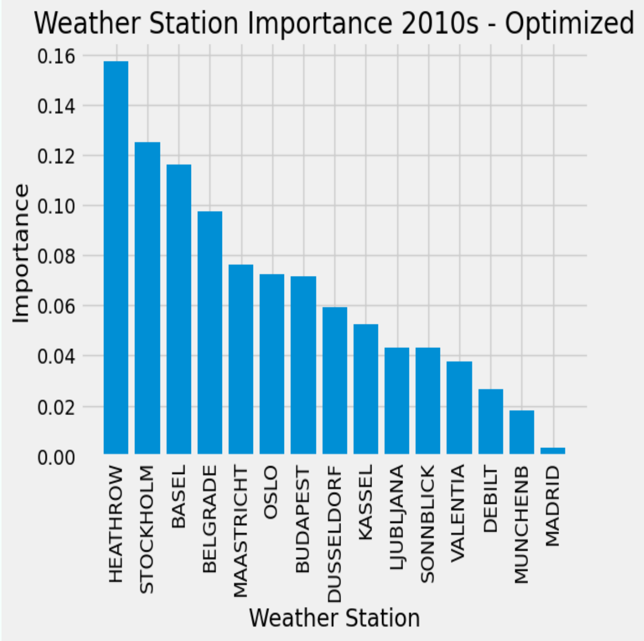
Determine the safest places for people to live in Europe over the next 25 to 50 years

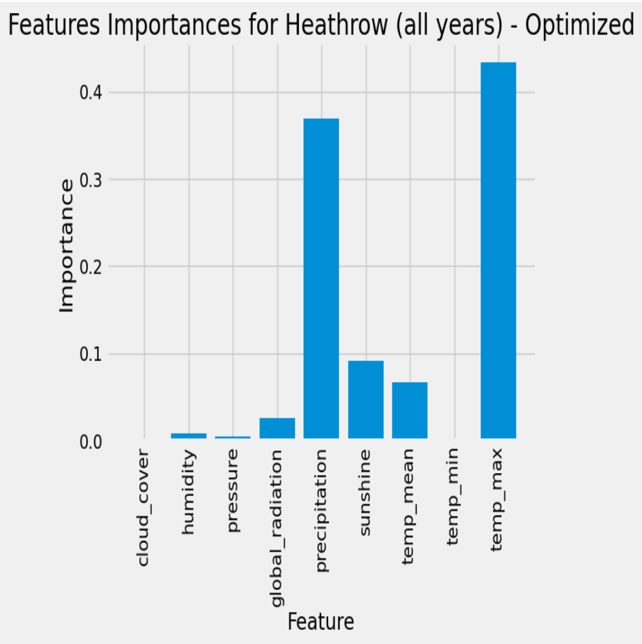
Approach

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 improved overall accuracy across all stations from 58.7% to 66.3%.**
- **However, accuracy for a single station decreased from 100% to 85%, indicating that the optimized model corrected an overfitting issue present in the unoptimized version.**
- **Comparing both models enabled us to better evaluate feature importance in climate analysis.**





Next Steps

We plan to run optimization processes on more weather stations across several decades, apply the optimized model to extreme weather datasets, and correlate these results with healthcare data to assess impacts on human wellbeing.

Recommendations

The data indicates that ClimateWins stands to gain the greatest benefit by focusing resources on optimizing GAN and CNN models for weather pattern prediction:

- **This experiment demonstrated the largest improvements in accuracy and revealed strong potential to meet key project objectives.**
- **Moreover, this approach has a proven track record of enhancing severe weather forecasts in real-world applications.**

Data and Algorithms Required

- CNN, GAN, and Bayesian optimization algorithms
- Extreme weather event data
- Radar imagery
- Dangerous weather classifications

Next Steps

- Refine Bayesian optimization of CNN
- Run CNN with generated data from GAN
- Gather and prepare additional data

THANK YOU!

Scripts: GitHub
Presentation: Vimeo