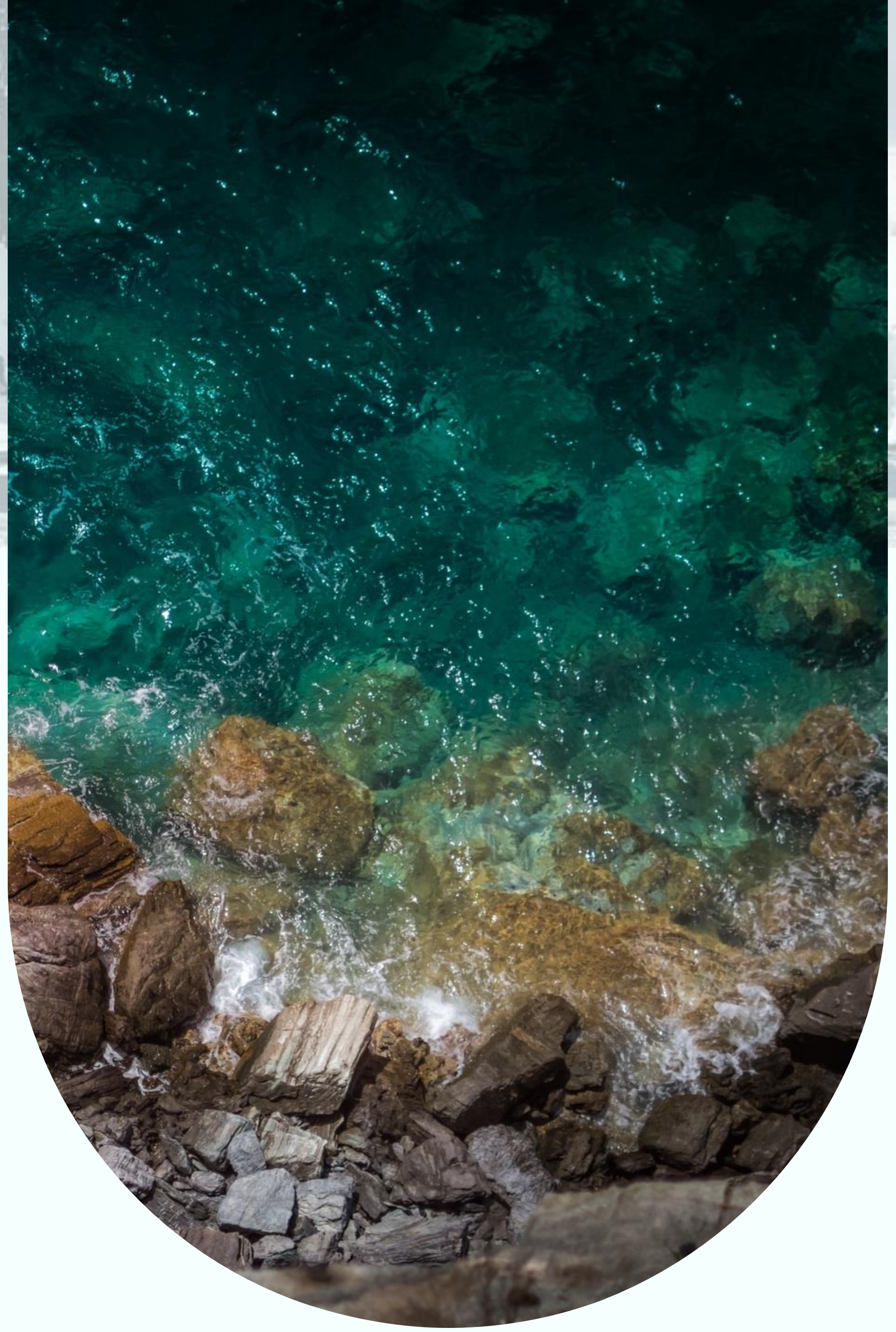
# CLIMATE WINS: PREDICTING CLIMATE CHANGE WITH MACHINE LEARNING



Laura DeCesare

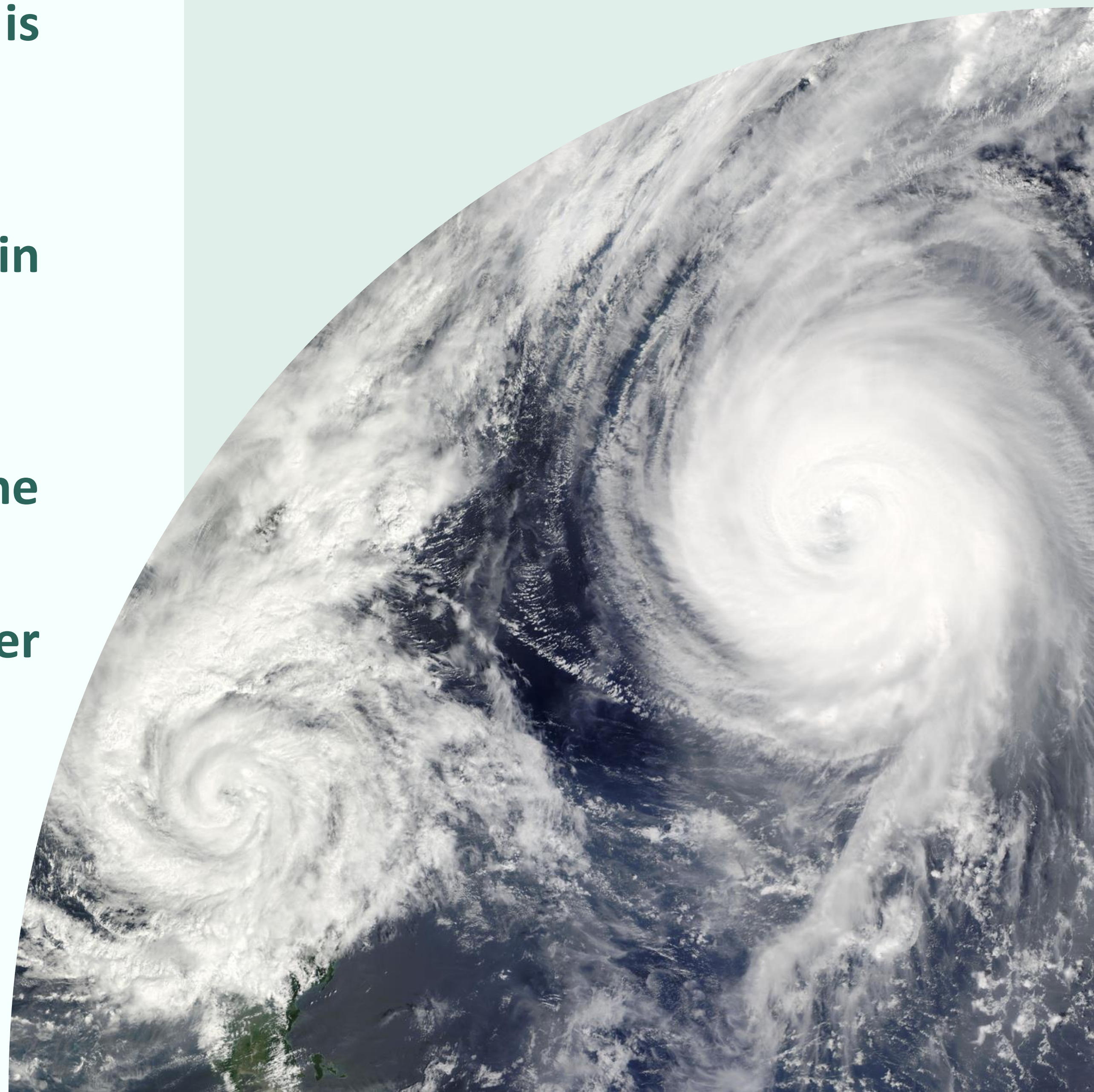
November 25, 2024

# 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 and wellbeing, 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



# Three Thought Experiments

### Classify Unusual Weather

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

### 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.

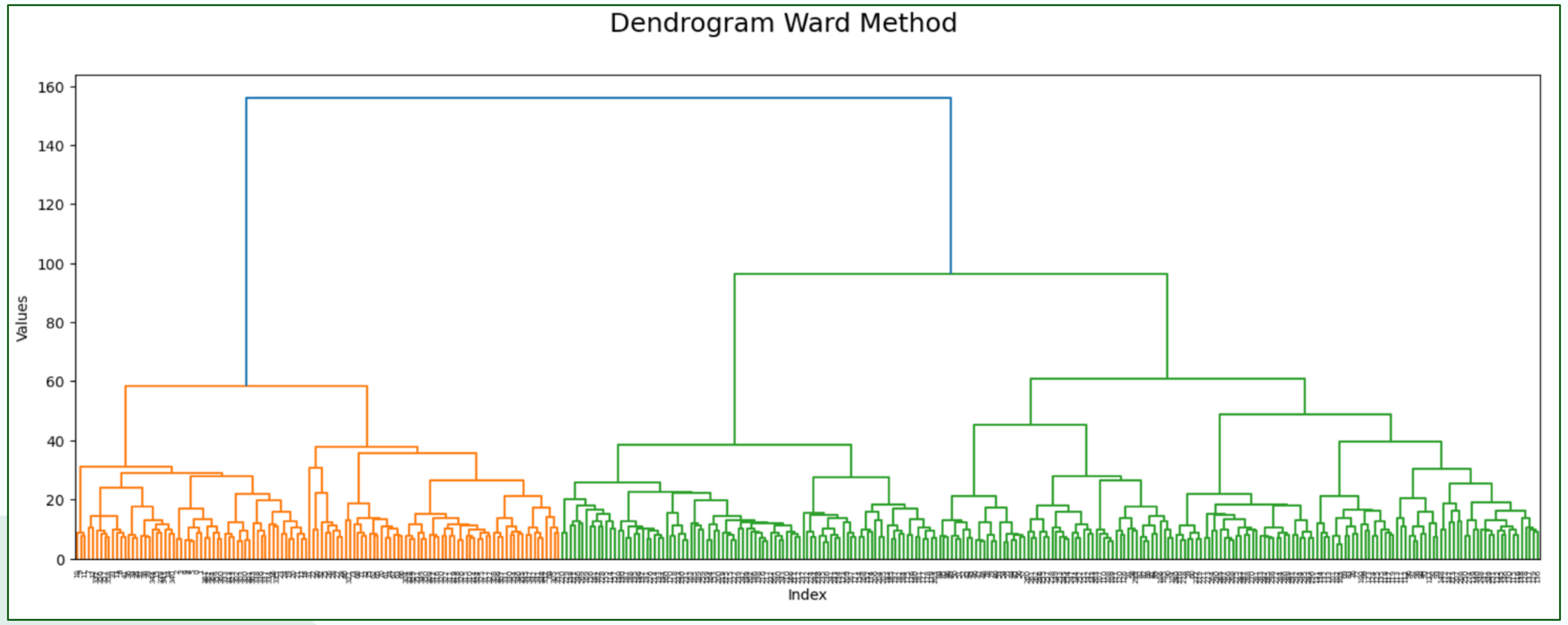
### Identify Safe Living Regions

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.

# Required Machine Learning Models

## Hierarchical Clustering

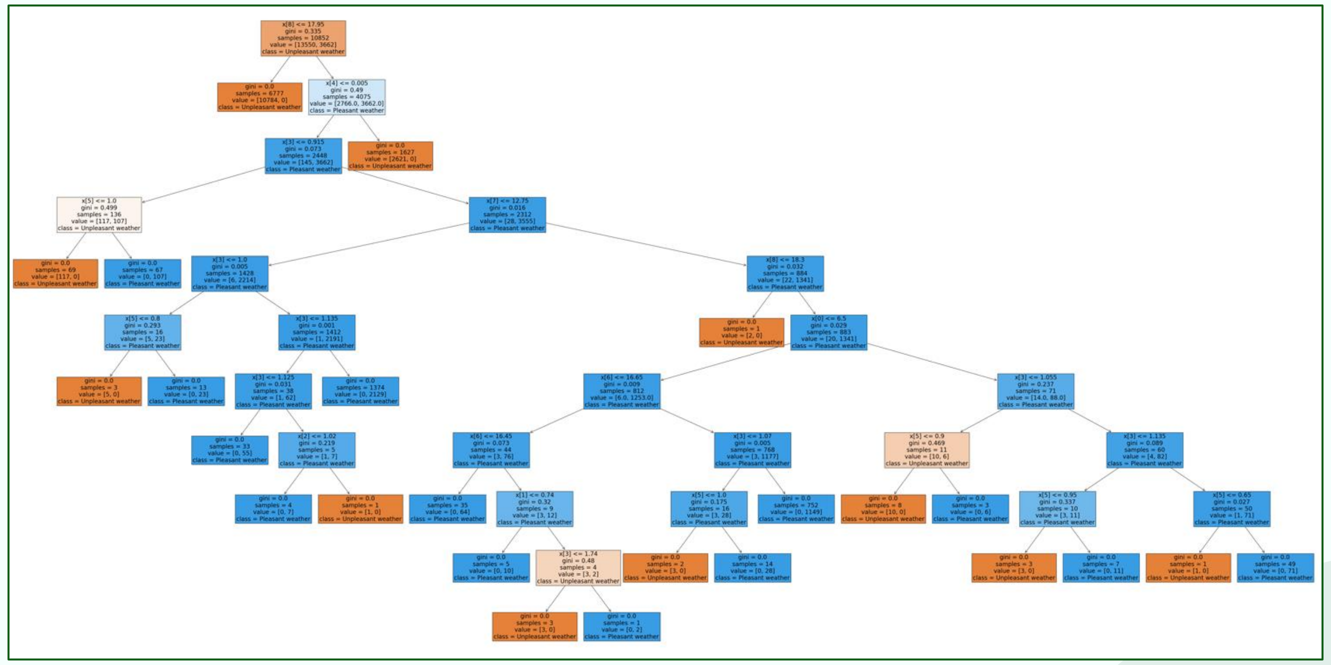
Categorizes each data point, merging data points — "leaves" — into branches based on similarity.  The result resembles an inverted tree with categories designated by color.




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

## Random Forest

Categorizes data using multiple decision tree models, which classifies data points using true/false processes.  A random forest trains each tree on a random sample and averages the results to produce a final prediction.



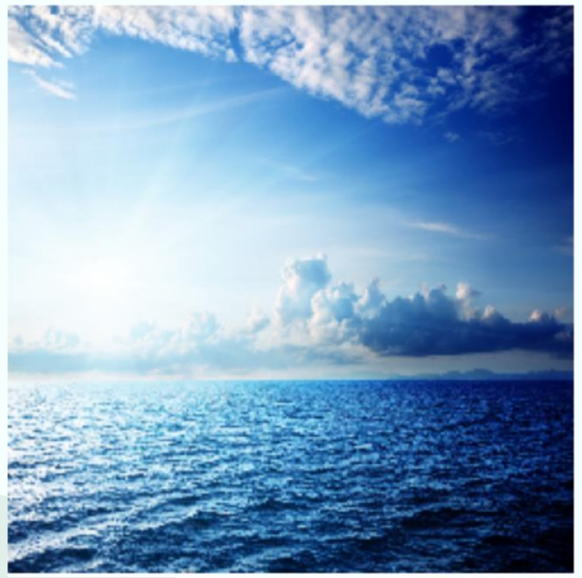
Benefits:
- Averaging process reduces overfitting
- Provides clarity through  visualizations

# Required Machine Learning Models

## General Adversarial Network (GAN)

A generator network produces artificial data based on real data points. A discriminator network samples both and attempts to identify accurate data. The generator then uses these results to create more effective "fakes."

Correct Prediction - class: Shine - predicted: Shine



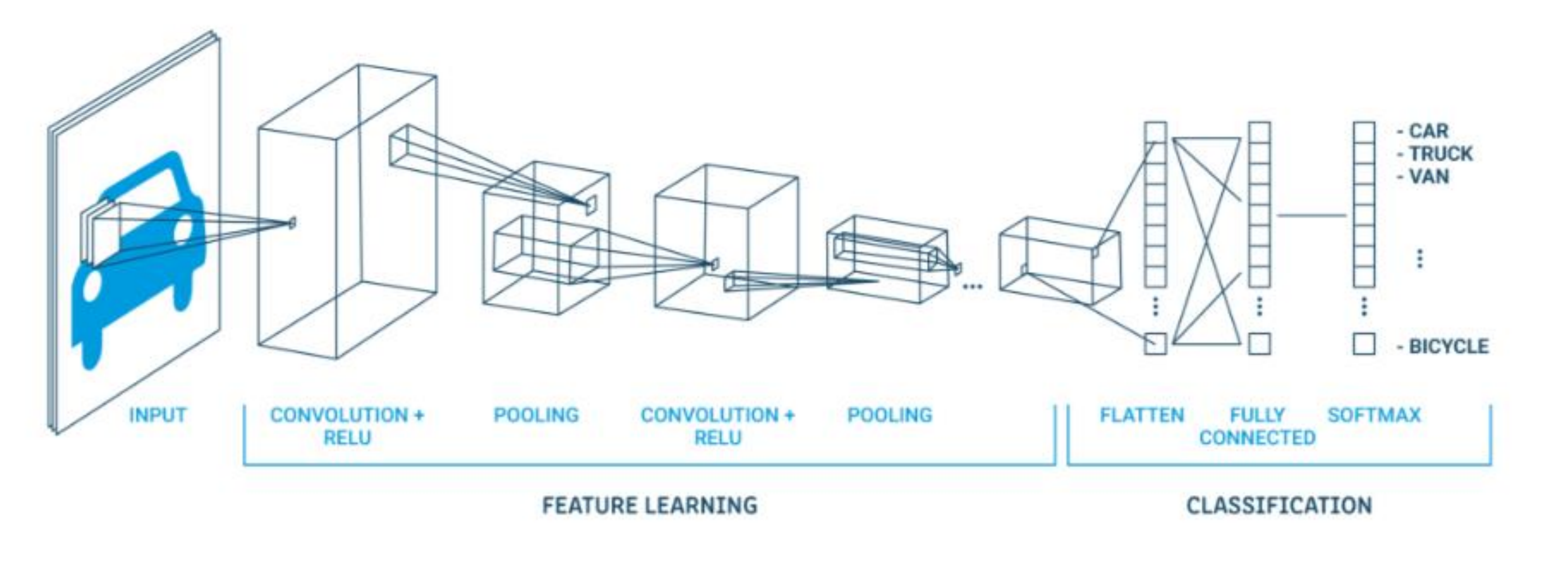
Correct Prediction - class: Cloudy - predicted: Cloudy

Benefits:

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

## Convolution Neural Network (CNN)

Using hidden layers, the CNN learns the features of a data set and creates classifications. Hidden pooling layers compress data into average and maximum values, reducing required computing power.




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**

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

**Objective**

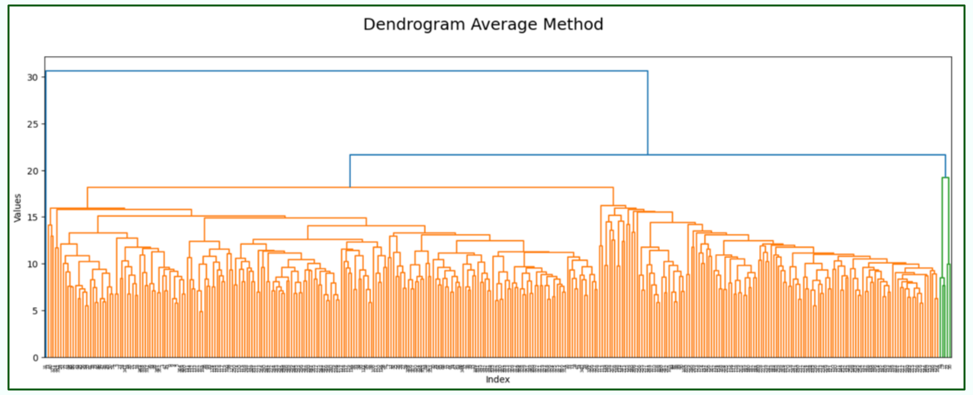
To identify whether unusual weather patterns are occurring

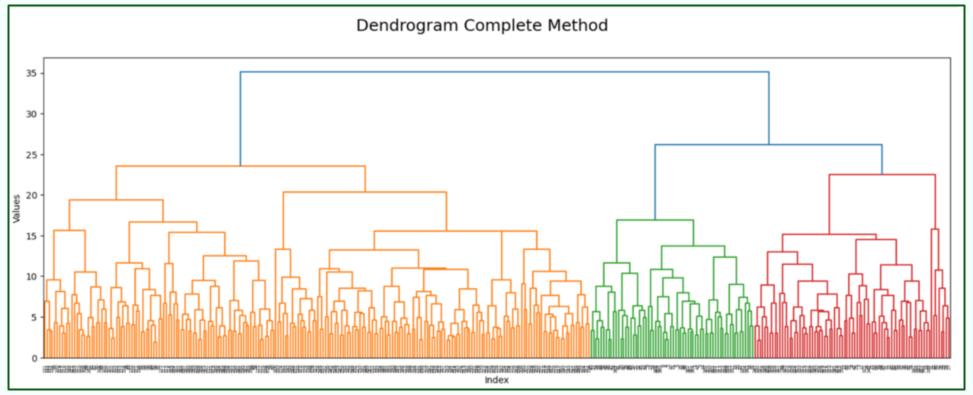
**Approach**

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 #2: 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. Run Bayesian 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 from 10.4% to 61.2%

| Layer (type) | Output Shape | Param # |
|---|---|---|
| conv1d_89 (Conv1D) | (None, 15, 61) | 610 |
| batch_normalization_32 (BatchNormalization) | (None, 15, 61) | 244 |
| dense_442 (Dense) | (None, 15, 61) | 3,782 |
| dropout_57 (Dropout) | (None, 15, 61) | 0 |
| dense_443 (Dense) | (None, 15, 61) | 3,782 |
| dense_444 (Dense) | (None, 15, 61) | 3,782 |
| max_pooling1d_89 (MaxPooling1D) | (None, 7, 61) | 0 |
| flatten_89 (Flatten) | (None, 427) | 0 |
| dense_445 (Dense) | (None, 15) | 6,420 |

**Total params:** 18,620 (72.73 KB)
**Trainable params:** 18,498 (72.26 KB)
**Non-trainable params:** 122 (488.00 B)

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**

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**

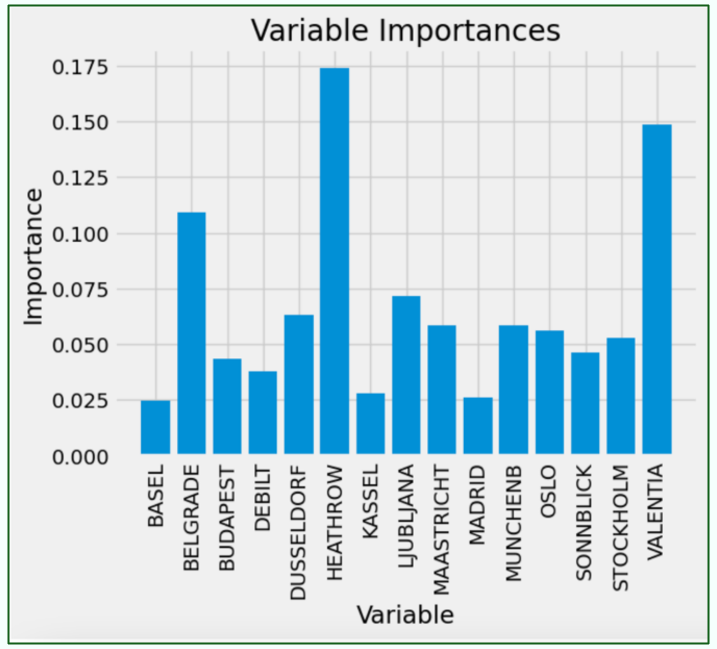
Determine the safest places for people to live in Europe over the next 25 to 50 years

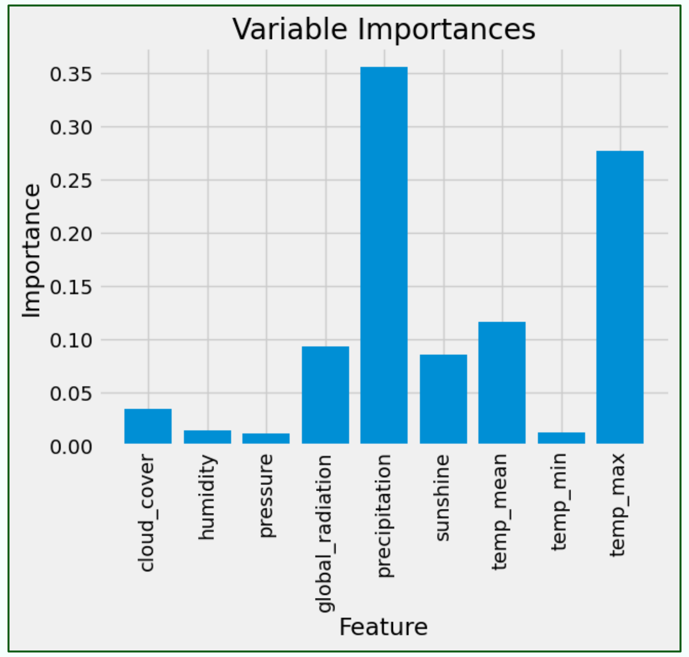
**Approach**

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 67.3% when looking at all stations
- Optimization decreased from 100% to 85% 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
- Compare results to healthcare data to correlate with human wellbeing

# 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

## 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!

Questions: lldecesare@gmail.com
Scripts: GitHub