ClimateWins: Predicting Climate Change with Machine Learning

Presented by Dr Yasemin Mustafa
Climate Wins Data Analyst
04/09/2025

Why Does It Matter?

- Extreme weather is becoming more frequent and severe across Europe and the world.
- ClimateWins, a European nonprofit, is exploring how machine learning can help predict and understand the impacts of climate change.
- This project uses data science to support smarter, more adaptive responses to our changing environment.

Project Objectives

1. Identify unusual weather patterns in Europe
2. Determine if unusual patterns are increasing
3. Generate forecasts for next 25–50 years
4. Identify safest places to live in Europe

Three Thought Experiments

Discover Hidden Weather Patterns

Hierarchical clustering allows us to move past simply labelling weather as typical or atypical, enabling more actionable groupings and insights.

Forecast Future Climate Scenarios

By generating synthetic weather data with a GAN, we can train a CNN to forecast potential climate conditions for the next 50 years.

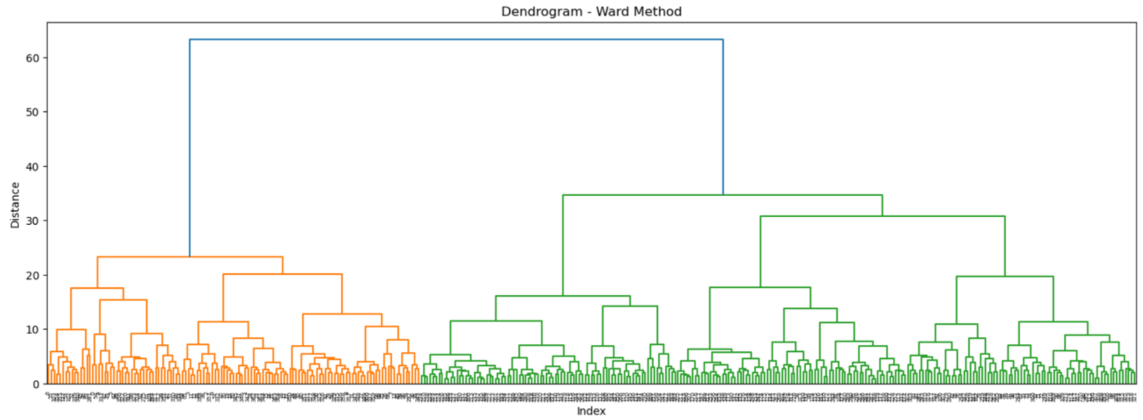
Identify Climate-Safe Regions

By refining a random forest model, we can highlight the most critical weather features and use them to pinpoint Europe’s safest regions for future habitation.

Required Machine Learning Models

Hierarchical Clustering

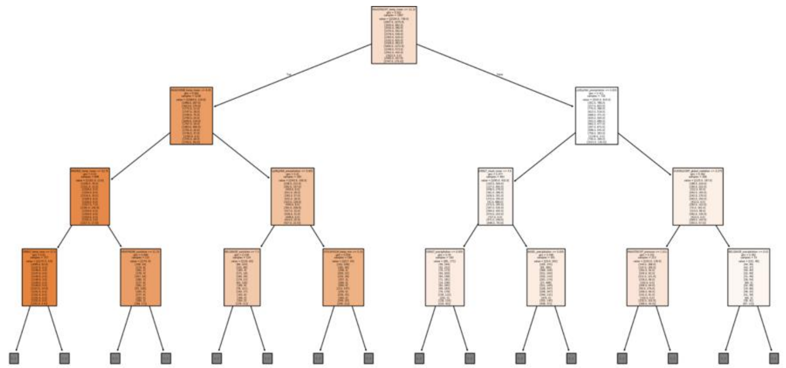
Clusters data by similarity into a hierarchical tree, with categories shown in colour.



- ✓ Creates new groupings directly from the data
- ✓ Works without predefined categories

Random Forest

Classifies data with multiple decision trees, averaging results from random samples for a final prediction.



- ✓ Reduces overfitting through averaging
- ✓ Enhances understanding with clear visualisations

Required Machine Learning Models

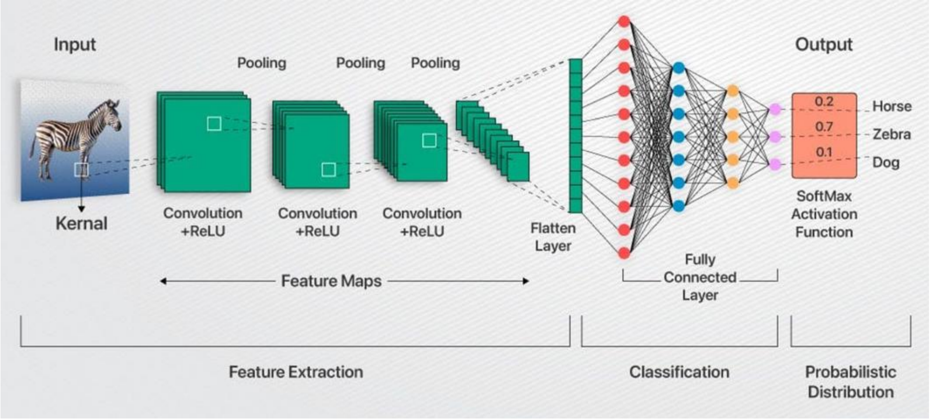
General Adversarial Network (GAN)

A generator creates synthetic data, while a discriminator compares it to real data and flags authenticity. The generator learns from this feedback to improve.

- ✓ Generates realistic weather data
- ✓ Expands datasets for better modelling

Convolution Neural Network (CNN)

CNN hidden layers extract features and classify data, while pooling layers condense information into key values, reducing computation.



- ✓ Accurately analyses radar imagery and complex datasets

Essential Data

Weather Event Data → Documented severe events in Europe: storms, extreme heat, and cold.

Radar Imagery → Radar observations corresponding to areas monitored by the selected stations.

Healthcare Data → Health outcomes linked to extreme weather: sickness, injury, and mortality.

Dangerous Weather Classifications → Indicators of hazardous weather derived from these datasets.

Thought Experiment 1: Weather Classification with Hierarchical Clustering

Hypothesis

If we use hierarchical clustering, then we can classify weather into meaningful, actionable groups rather than just typical or atypical.

Objective

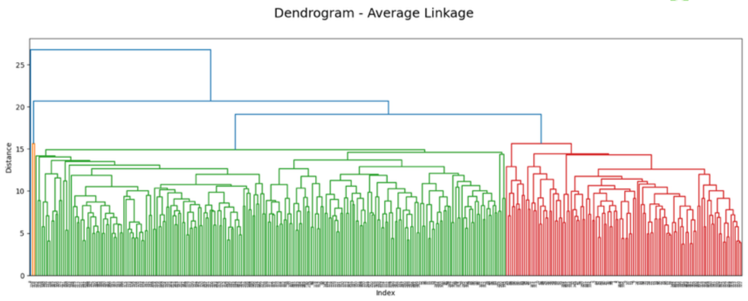
To detect the presence of unusual weather patterns.

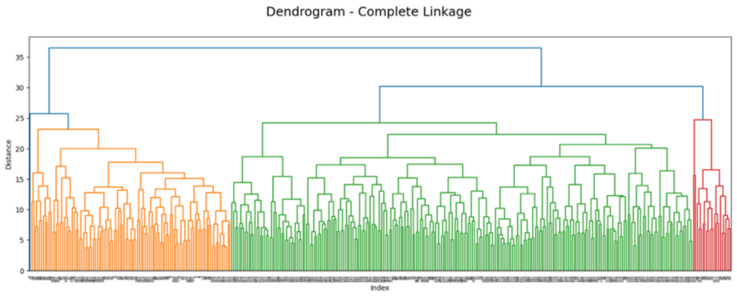
Approach

Model viability was evaluated through a dendrogram of weather station data, with PCA applied to minimise dimensions and maximise efficiency.

Results

The model consistently produced two to three clusters, possibly reflecting binary or low–mid–high divisions.





Next Steps…

Apply this model across multiple years, seasons, and months to compare results.

Thought Experiment 2: Data Simulation to Improve Forecast Accuracy

Hypothesis

If we use a GAN to synthesise weather data, then we can train a CNN to predict possible conditions over the next 50 years.

Objective

Forecast potential weather conditions for the next 25–50 years.

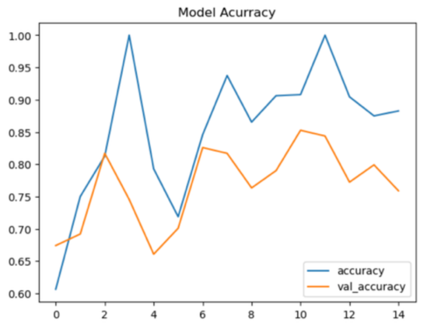
Approach

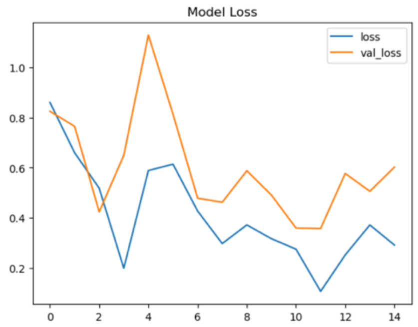
1. Apply Bayesian optimisation to a CNN model to assess accuracy.
2. Use a GAN to generate realistic synthetic weather data.

Results

1. Bayesian optimization reduced CNN accuracy from 60.2% to 56.5%
2. GAN outputs achieved 88.3% accuracy and 29.2% loss.

Dataset	Accuracy before optimisation	Accuracy after optimisation
All Weather Stations and Only a Decade of Data	60.2%	56.5%
Single-Station Full Timeline Dataset	100%	99.9%





Next Steps…

- Run the optimised CNN model using GAN-generated weather data.
- Analyse results across multiple years, countries, and regions.

Thought Experiment 3: Regional Assessment Using Random Forests

Hypothesis

If we optimise a random forest model, then we can identify the most critical weather features, using them to determine safe living regions.

Objective

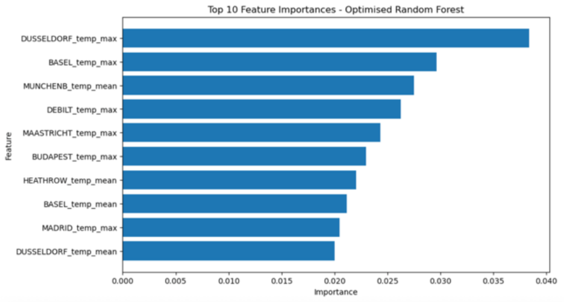
Identify the safest regions for living in Europe over the next 25–50 years.

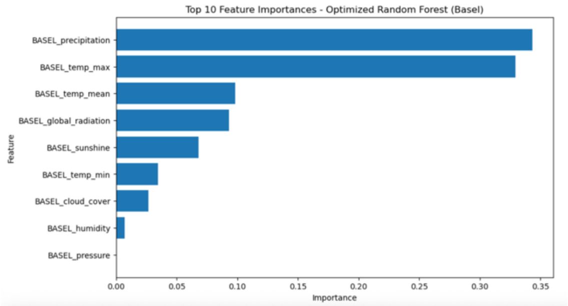
Approach

Optimise random forest hyperparameters and compare pre- vs. post-optimisation results.

Results

1. Single-station accuracy dropped from 100% to 99.9%.
2. Comparing models helps assess feature importance.





Next Steps…

- Apply optimisation to more stations and decades of data.

Summary, Recommendations & Next Steps

Data indicates ClimateWins gains the most by investing in GAN and CNN optimisation for weather prediction:

- Highest accuracy improvements observed
- Immediate potential to achieve key project goals
- Proven effectiveness in real-world severe weather forecasting

Required Data and Algorithms:

- CNN, GAN, and Bayesian optimization
- Extreme weather event records
- Radar imagery
- Hazardous weather classifications

Next Steps:

- Refine CNN with Bayesian optimization
- Run CNN using GAN-generated data
- Collect and prepare additional datasets

Thank You

I appreciate your time and interest in this project.

If you'd like to explore the code, data, and visualisations further, please check out my GitHub:

👉 **https://github.com/YLMustafa/climate_wins_ML**

Feel free to reach out with any questions or feedback!