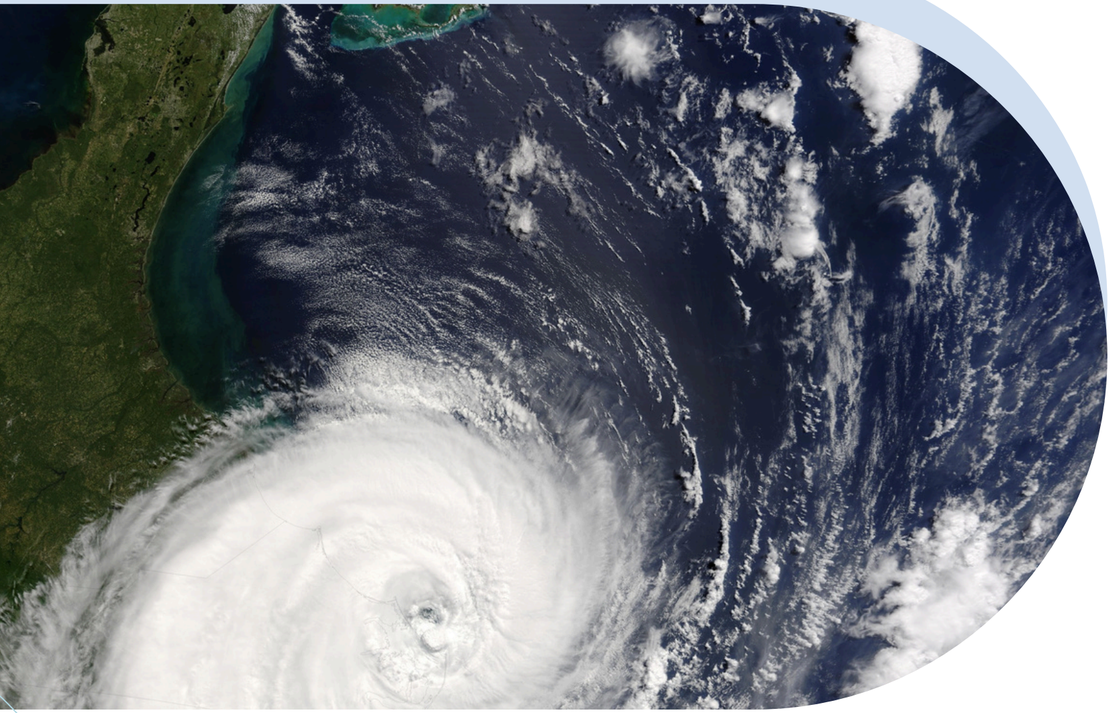


PREDICTING WEATHER VARIATIONS WITH MACHINE LEARNING

Strategy for ClimateWins

THEO LIANG

DECEMBER 18, 2024

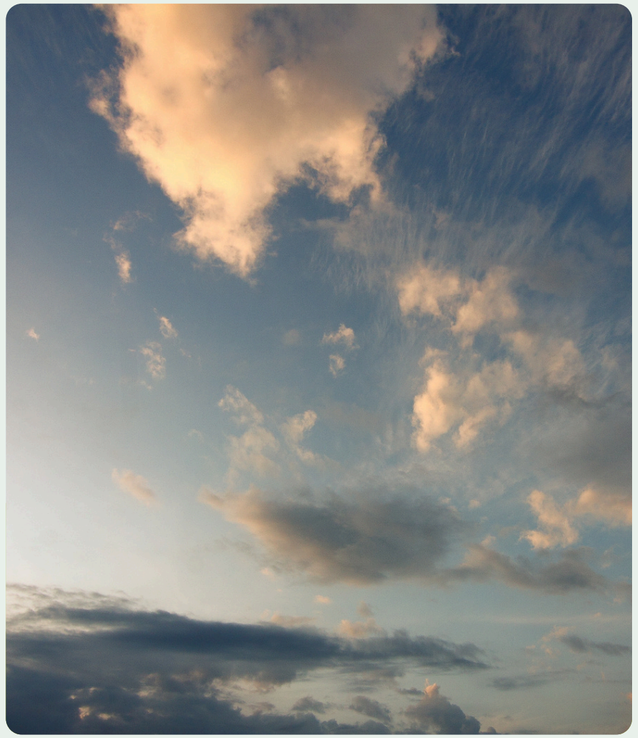
Introduction

Climate change is a global phenomenon characterized by long-term changes in weather patterns due to global warming. The increase in global temperatures triggered by increased greenhouse gas emissions has disrupted the climate system, resulting in irregular rainfall, changing seasons, and an increase in the frequency of extreme weather.

This shift has a direct impact on human life such as food crises, health, and water availability.

ClimateWins 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 25 to 50 years based on current trends.
- Determine the safest places for people to live in Europe over the next 25 to 50 years.



Thought Experiments

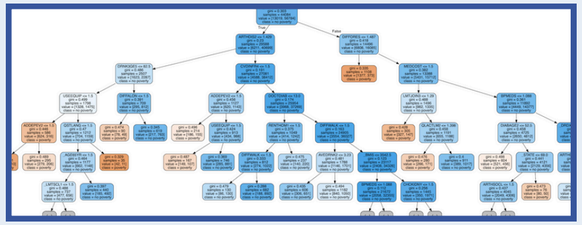
	Experiment 1	Experiment 2	Experiment 3
Objective	*Determine if unusual weather patterns are increasing.*	*Generate possibilities for future weather conditions over the next 25 to 50 years based on current trends.*	*Determine the safest places for people to live in Europe over the next 25 to 50 years.*
Machine Learning Technique	**Random Forest**	**Generative Adversarial Network (GAN)**	**Hierarchical Clustering**
Outcome	**Determine whether extreme weather events, like heatwaves or heavy rainfall, are becoming more common over time.**	**Create realistic future weather scenarios based on current trends, giving a glimpse of what Europe's climate might look like in 25–50 years.**	**Identify areas in Europe that are less likely to experience severe climate impacts, offering guidance on where people might consider relocating.**

Machine Learning Techniques Explained

Random Forest

A collection of decision trees that work together to make predictions. Each tree looks at a random subset of data, and the final prediction is based on what most trees say.

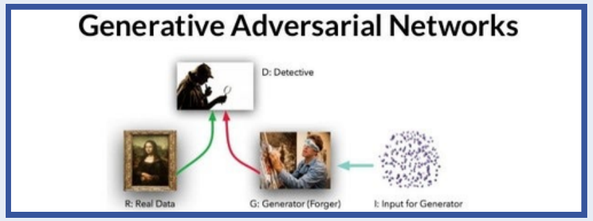
Used for making predictions, like whether a weather event is normal or unusual, or predicting a temperature. It also helps identify which factors are most important for making those predictions.



Generative Adversarial Network

Two neural networks work together—one generates new data (like weather patterns), and the other checks if the data looks real. They improve by competing with each other.

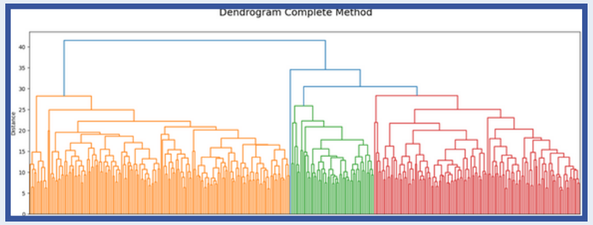
Used to create new data that looks real, like simulating future weather or generating realistic images.



Hierarchical Clustering

This technique groups similar data points into clusters, and it organizes them in a tree structure. It starts with each data point as its own group and gradually combines the closest groups.

Used to find natural groupings in data, like clustering regions with similar weather patterns, or identifying unusual data points.



Thought Experiment 1:

Determining Unusual Weather Patterns with Random Forest

Objective

To determine whether or not unusual weather patterns are occurring.

Hypothesis

Using Random Forest, historical weather data can reveal whether the frequency of unusual events, such as extreme temperatures or precipitation, has increased over time.

Approach

Historical weather data is analyzed using a Random Forest classifier to classify events as "normal" or "unusual" based on key weather features. The model's predictions are examined for trends over time, and feature importance metrics are used to understand which variables most strongly influence the classification of unusual events.

Thought Experiment 2:

Predicting Future Weather Conditions Using Generative Adversarial Network

Objective

To generate future weather scenarios using a GAN

Hypothesis

A Generative Adversarial Network can generate realistic weather scenarios for the next 25 to 50 years by learning patterns and trends from historical weather data.

Approach

The Generative Adversarial Network will be trained on historical weather data to learn existing patterns and trends. The generator will create future weather scenarios, while the discriminator will evaluate their realism, refining the model to produce accurate weather projections for the next 25 to 50 years.

Thought Experiment 3:

Identifying the Safest Places to Live in Europe with Hierarchical Clustering

Objective

Determine the safest places to live in Europe over the next 25 to 50 years using hierarchical clustering.

Hypothesis

Hierarchical clustering can identify regions in Europe that will remain the safest to live in the future, based on factors like climate stability, natural disaster risk, and other safety indicators.

Approach

Data on climate trends, natural disaster frequencies, and safety indicators will be collected and processed. Hierarchical clustering will group regions based on similarities, creating a dendrogram that shows how regions relate to each other. This will help identify areas likely to remain safe in the future based on these factors.

Summary

Random Forest Model

Identify unusual weather patterns by analyzing historical data and classifying events as "normal" or "unusual" based on key weather variables.

Generative Adversarial Network

Generate realistic future weather scenarios by learning from current weather trends to predict conditions over the next few decades.

Hierarchical Clustering

Group regions in Europe based on safety-related factors, identifying areas likely to remain safe in the future.

Recommendation

Using all three techniques will help ClimateWins assess and prepare for future climate challenges. Random Forest can identify changing weather patterns, GANs can generate future scenarios, and Hierarchical Clustering can pinpoint the safest regions for long-term resilience.

Data Needed

Historical Weather Data

Past temperature, precipitation, wind speed, humidity, and other relevant weather variables.

Natural Disaster Data

Historical data on extreme weather events (e.g., floods, heatwaves, storms) to classify unusual events using Random Forest and analyze regional safety in Hierarchical Clustering.

Geographic Data

Elevation, proximity to water bodies, and other geographic features to assess risk and safety in the clustering process.

Next Steps

- **Compare the outputs of Random Forest, GAN, and Hierarchical Clustering to identify trends and inconsistencies in the results.**
- **Analyze how changes in input data or model parameters impact the results to ensure model reliability.**
- **Refine and validate the future scenarios and clustering results with additional data.**

THANK YOU

For more information or to connect, visit the links below or reach out via email at theodore.liang@gmail.com.

GitHub

in

tableau