

**Machine Learning Specialisation**

# **ClimateWins**

**Weather Conditions and Climate Change**

**Prepared by Ana Lazarevska, November 2024**

# Contents

1. Project Objectives & Thought Experiments
2. Machine Learning Algorithms
3. Data Overview
4. Thought Experiment 1 - Predicting Extreme Weather Events with GANs
5. Thought Experiment 2 - Identifying Unusual Weather Patterns with CNNs
6. Thought Experiment 3 - Identifying Climate Change Hotspots with Random Forests
7. Pros/Cons of the Thought Experiments
8. Recommendations

# Project Objective

- 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

- Predicting Extreme Weather Events with GANs
- Identifying Unusual Weather Patterns with CNNs
- Identifying Climate Change Hotspots with Random Forests



# Machine Learning Algorithms

The machine learning options for achieving ClimateWins' goals.

- **Random Forests**
  - Can be used for both regression (predicting continuous variables like temperature) and classification (predicting categorical variables like weather events)
- **Convolutional Neural Networks (CNNs)**
  - Excellent for analysing spatial data like satellite imagery.
  - Can identify patterns in atmospheric and oceanic processes that influence climate.
- **Recurrent Neural Networks (RNNs)**
  - Well-suited for time series data, such as historical climate records.
  - Can capture temporal dependencies and predict future climate trends.
- **Generative Adversarial Networks (GANs)**
  - Can generate synthetic climate data, helping to fill gaps in historical records and simulate future scenarios.
  - Can be used to create realistic climate projections.

# Data Overview

To achieve ClimateWins' goals, additional data beyond the historical weather set would significantly enhance the accuracy and reliability of predictions.

- **Socioeconomic Data**
  - **Population Density:** Understanding population distribution can help identify areas at risk of extreme weather events.
  - **Infrastructure Vulnerability:** Information on infrastructure like roads, bridges, and power grids can assess potential damage.
  - **Economic Indicators:** Economic data can help evaluate the impact of climate change on various sectors.
- **Land Use and Land Cover Data**
  - **Urbanisation Patterns:** Urban areas often experience heat island effects and increased flood risks.
  - **Forest Cover:** Forests play a crucial role in carbon sequestration and water regulation.
  - **Agricultural Land:** Agricultural practices can influence local climate patterns.
- **Oceanographic Data**
  - **Sea Surface Temperature:** Ocean temperatures influence atmospheric patterns and weather events.
  - **Ocean Currents:** Ocean currents can transport heat and moisture, affecting regional climates.
  - **Sea Level Rise:** Rising sea levels pose significant threats to coastal areas.
- **Atmospheric Composition Data**
  - **Greenhouse Gas Concentrations:** Understanding greenhouse gas emissions and their impact on the climate system is essential.
  - **Aerosol Concentrations:** Aerosols can affect cloud formation and radiation balance.

# **Thought Experiment 1:**

## **Predicting Extreme Weather Events with GANs**



# TE1: Predicting Extreme Weather Events with GANs

## Linked Objective

- Generate possibilities for future weather conditions over the next 25 to 50 years based on current trends.

## Machine Learning Methods

- Generative Adversarial Networks (GANs)

The GAN can generate realistic simulations of extreme weather events, helping to identify potential future scenarios and their impacts.

## Data Needed

- Historical weather data (temperature, precipitation, wind speed, atmospheric pressure)
- Satellite imagery

# TE1: Predicting Extreme Weather Events with GANs

## Methodology

1. **Data Collection:** Gather historical weather data, including temperature, precipitation, wind speed, and atmospheric pressure.
2. **GAN Training:** Train a GAN to generate synthetic weather data that mimics real-world patterns, including extreme events.
3. **Extreme Event Simulation:** Use the trained GAN to simulate various extreme weather scenarios, such as heatwaves, droughts, floods, and storms.
4. **Impact Assessment:** Analyse the simulated events to assess their potential impact on infrastructure, agriculture, and human health.

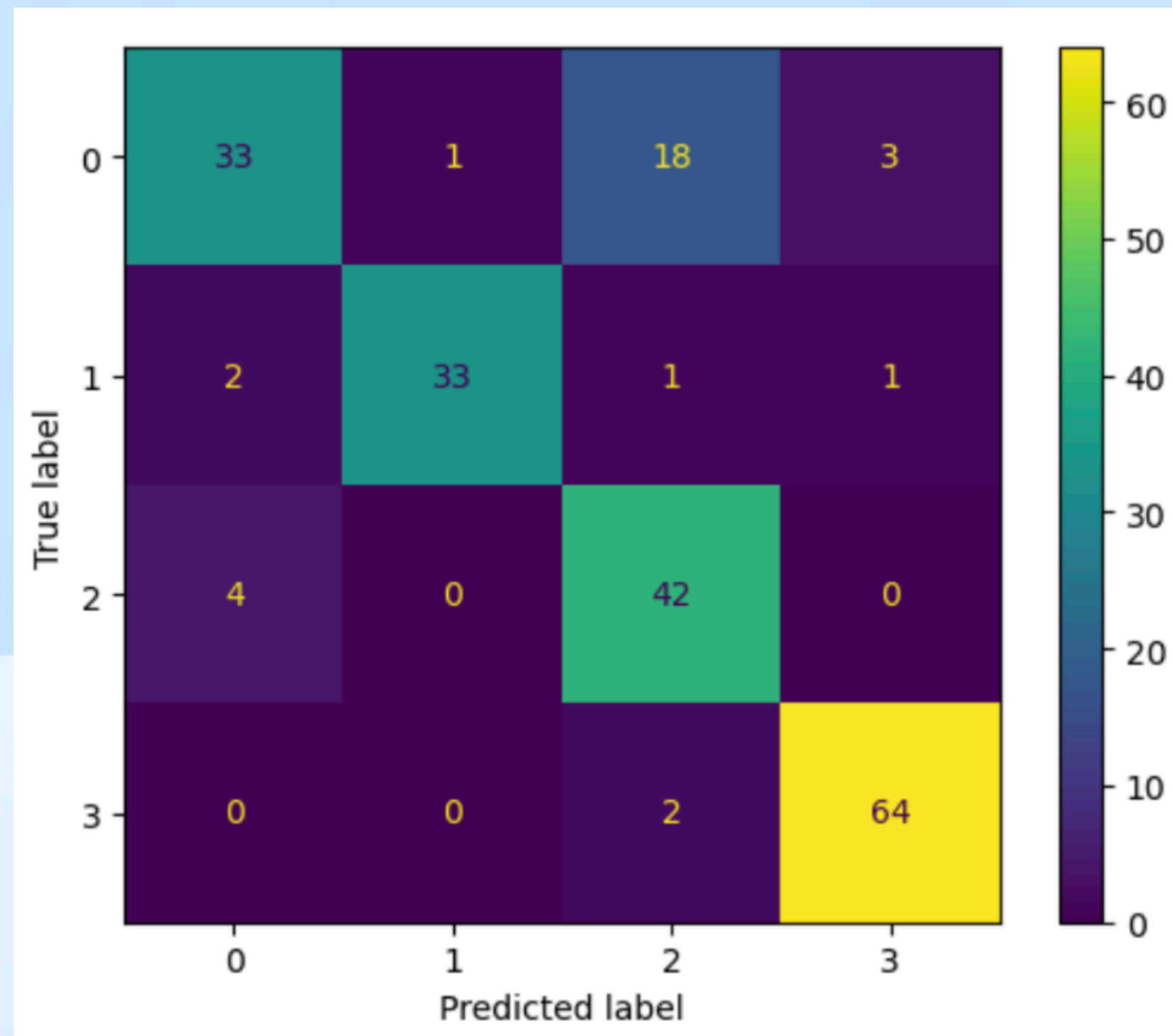
## Expected Outcomes

- **Early Warning Systems:** Identify potential risks and develop early warning systems.
- **Infrastructure Planning:** Inform infrastructure design and adaptation strategies.
- **Resource Allocation:** Optimise resource allocation for disaster response and recovery.

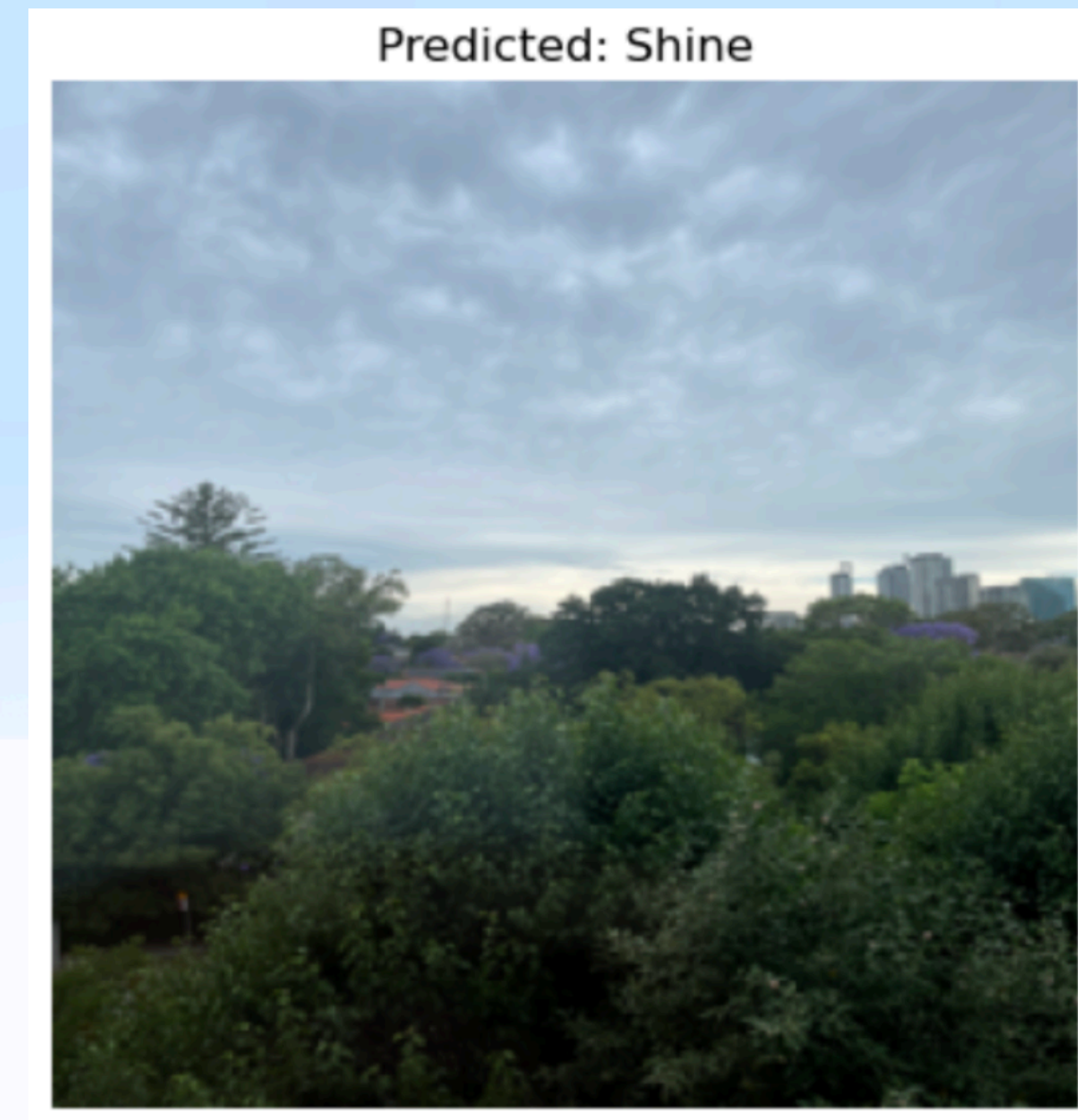


# TE1: Predicting Extreme Weather Events with GANs

## Practical Applications



The confusion matrix indicates that the GAN model correctly classified most of the images for the four weather conditions: cloudy, rain, shine, and sunrise. The biggest exception is that 18 samples of cloudy images were classified as shine.



The GAN model struggled with identifying as cloudy any image that has much brightness. As seen on the image below, the cloudy and overcast day was predicted as sunny (shine).

# **Thought Experiment 2:**

## **Identifying Unusual Weather Patterns with CNNs**

# TE2: Identifying Unusual Weather Patterns with CNNs

## Linked Objectives

- Identify weather patterns outside the regional norm in Europe.
- Determine if unusual weather patterns are increasing.

## Machine Learning Methods

- Convolutional Neural Networks (CNNs)

CNNs can effectively identify unusual weather patterns by analyzing satellite imagery and other spatial data.

## Data Needed

- Satellite imagery
- Radar data
- Historical weather data



# TE2: Identifying Unusual Weather Patterns with CNNs

## Methodology

1. **Data Collection:** Gather satellite imagery, radar data, and other relevant spatial data.
2. **Feature Extraction:** Use CNNs to extract relevant features from the data, such as cloud patterns, temperature anomalies, and atmospheric disturbances.
3. **Pattern Recognition:** Train a CNN to recognise unusual weather patterns, such as extreme storms, heatwaves, and cold spells.
4. **Trend Analysis:** Analyse the frequency and intensity of unusual weather patterns over time.

## Expected Outcomes

- **Improved Weather Forecasting:** Enhance the accuracy of weather forecasts and early warning systems.
- **Climate Change Monitoring:** Track changes in weather patterns and their potential impact on climate change.
- **Risk Assessment:** Assess the risk of extreme weather events and inform disaster preparedness.

# TE2: Identifying Unusual Weather Patterns with CNNs

## Practical Application

Pred	BASEL	BELGRADE	BUDAPEST	DEBILT	DUSSELDORF	HEATHROW	KASSEL	\
True								
BASEL	3528	48	16	2	8	9	1	
BELGRADE	98	991	1	0	0	0	0	
BUDAPEST	20	20	172	2	0	0	0	
DEBILT	10	4	15	53	0	0	0	
DUSSELDORF	3	0	1	8	8	9	0	
HEATHROW	6	1	2	3	4	65	0	
KASSEL	1	2	1	0	1	0	4	
LJUBLJANA	6	5	4	0	0	7	1	
MAASTRICHT	3	0	0	1	0	1	0	
MADRID	12	13	15	2	3	13	0	
MUNCHENB	5	1	0	0	0	0	0	
OSLO	0	0	0	0	1	0	0	
STOCKHOLM	1	0	1	0	1	0	0	
VALENTIA	0	0	0	0	0	1	0	

Pred	LJUBLJANA	MAASTRICHT	MADRID	MUNCHENB	OSLO
True					
BASEL	4	1	64	1	0
BELGRADE	0	0	2	0	0
BUDAPEST	0	0	0	0	0
DEBILT	0	0	0	0	0
DUSSELDORF	0	0	0	0	0
HEATHROW	0	0	0	0	1
KASSEL	1	1	0	0	0
LJUBLJANA	34	0	3	0	1
MAASTRICHT	0	3	1	0	0
MADRID	7	0	393	0	0
MUNCHENB	0	0	0	2	0
OSLO	0	0	0	0	4
STOCKHOLM	0	0	0	0	1
VALENTIA	0	0	0	0	0

- The accuracy of the CNN model before optimisation was around 12%, with 30 epochs, batch-size of 32, and 128 hidden layers.
- After optimisation the accuracy was steadily increasing to about 92%, epochs increased to 47, the batch size greatly increased to 460, and the number of neurons in the hidden layers reduced to 61. The activation switched from relu to softsign. All of these changes produced a much more accurate model.
- However, the optimised model only recognised 12 weather stations, which makes it less efficient.

# **Thought Experiment 3:**

## **Identifying Climate Change Hotspots with Random Forests**



# TE3: Identifying Climate Change Hotspots with Random Forests

## Linked Objectives

- Identify weather patterns outside the regional norm in Europe.

## Machine Learning Methods

- Random Forest

Random Forest models can effectively identify regions in Europe that are experiencing significant climate change impacts.

## Data Needed

- Historical weather data (temperature, precipitation, wind speed, atmospheric pressure)
- Satellite imagery
- Socioeconomic data
- Land use and land cover data

# TE3: Identifying Climate Change Hotspots with Random Forests

## Methodology

1. **Data Collection:** Gather relevant data from various sources, including meteorological stations, satellite sensors, and government agencies.
2. **Data Preprocessing:** Clean and preprocess the data to handle missing values, outliers, and inconsistencies.
3. **Feature Engineering:** Create relevant features from the raw data, such as:
  - Climate indices (e.g., temperature anomalies, precipitation anomalies, drought indices)
  - Vegetation indices (e.g., NDVI)
  - Urbanisation indices
  - Socioeconomic indicators
4. **Model Training:** Train a Random Forest model on the prepared dataset. The model will learn to classify regions based on their climate change vulnerability.
5. **Model Evaluation:** Evaluate the model's performance using appropriate metrics, such as accuracy, precision, recall, and F1-score.
6. **Hotspot Identification:** Use the trained model to classify regions into different categories of climate change vulnerability.
7. **Visualisation:** Visualise the results on a map to identify hotspots and potential areas of concern.

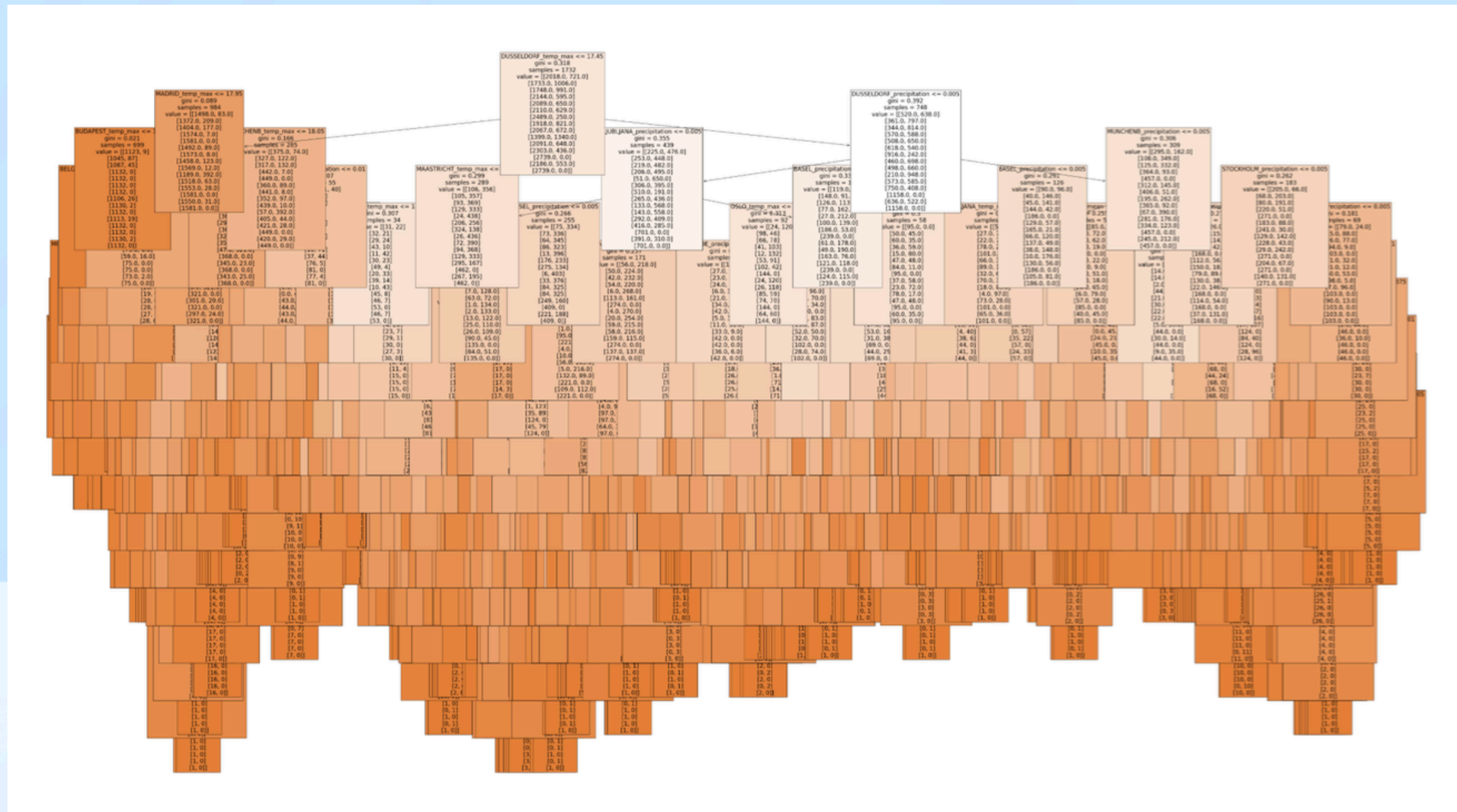
## Expected Outcomes

- **Identification of Climate Change Hotspots:** Pinpoint regions that are particularly vulnerable to climate change impacts.
- **Prioritisation of Adaptation and Mitigation Efforts:** Focus resources on areas with the highest risk.
- **Informed Decision-Making:** Provide data-driven insights to policymakers and stakeholders.
- **Improved Climate Resilience:** Develop strategies to reduce the negative impacts of climate change.

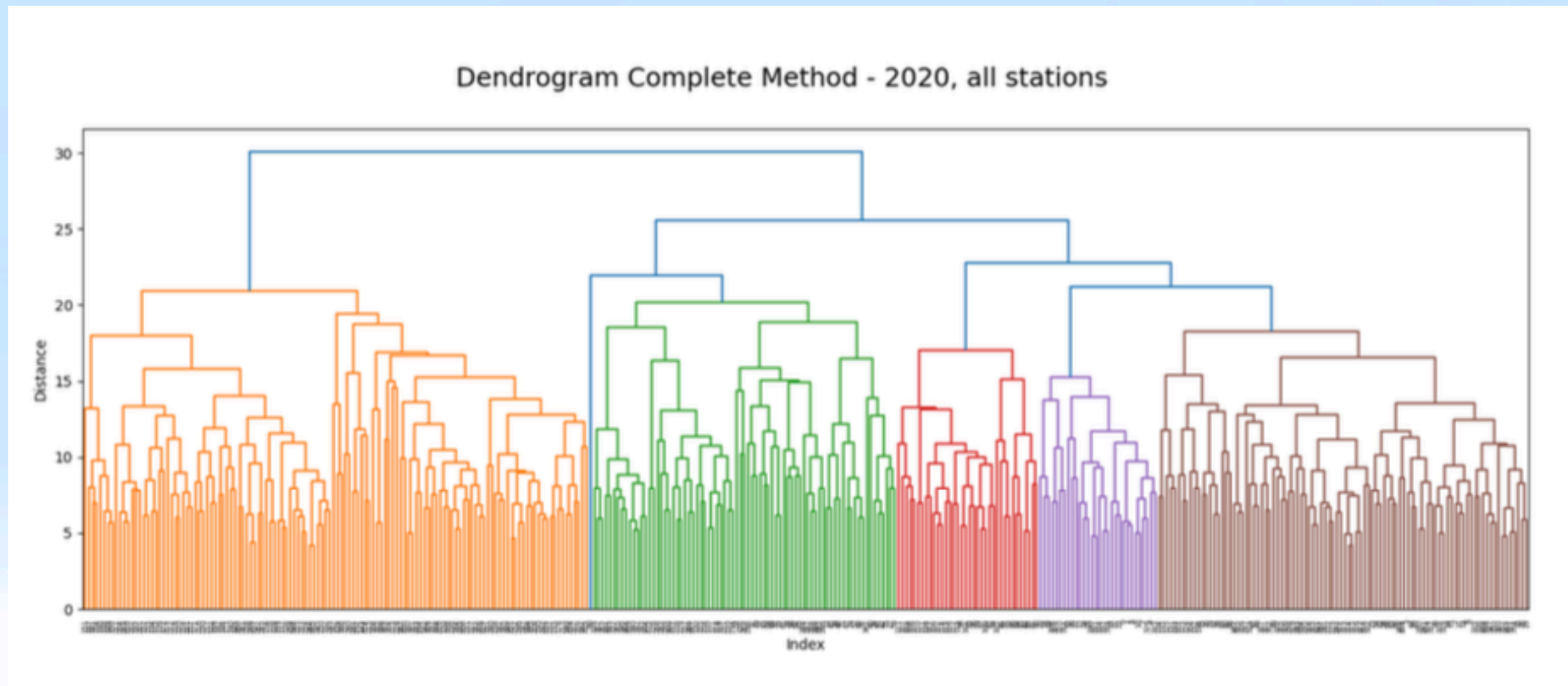


# TE3: Identifying Climate Change Hotspots with Random Forests

## Practical Application



Looking at 15 weather stations in Europe for the 2010s, after the optimisation of the random tree model and taking into consideration pleasant and unpleasant weather data points, the accuracy of the model was only 66.2%.



A dendrogram for all weather stations for the 2020s, using the complete method, grouped the data into 5 clusters.

The large orange cluster probably contains cities with similar weather conditions, such as a temperate climate with moderate temperatures and precipitation.

The smaller clusters might represent cities with more extreme or unique climates, such as a maritime climate with mild temperatures and high humidity, or a continental climate with cold winters and hot summers.



# Pros/Cons of the Thought Experiments

## TE1: Predicting Extreme Weather Events with GANs

### **PROS:**

- GANs can generate diverse and realistic weather scenarios.
- Can identify unusual weather patterns that may not be captured by traditional statistical methods.

### **CONS:**

- Training GANs can be computationally expensive.
- GANs are complex models that require careful tuning and expertise.

## TE2: Identifying Unusual Weather Patterns with CNNs

### **PROS:**

- CNNs are well-suited for analysing spatial data like satellite imagery.
- Can identify complex patterns in weather data.

### **CONS:**

- Requires large amounts of labeled data for training.
- CNNs can be complex and computationally intensive.

## TE3: Identifying Climate Change Hotspots with Random Forests

### **PROS:**

- Random Forest models are relatively easy to interpret.
- Can handle a wide range of data types and is less prone to overfitting.
- Can be used for both classification and regression tasks.

### **CONS:**

- Requires careful feature engineering to extract relevant information.
- Can become complex with a large number of trees.

# Recommendations

The third thought experiment, Identifying Climate Change Hotspots with Random Forests, has the most potential for ClimateWins for the following reasons:

- Directly addresses the goal of identifying regions vulnerable to climate change.
- Provides insights into the factors driving climate change impacts.
- Handles various data types and is less prone to overfitting.
- Can be adapted to different climate change analysis tasks.

By focusing on this thought experiment, ClimateWins can develop a powerful tool for assessing climate risks and informing adaptation strategies.

# Questions?

Contact: [ana.lazarevska@dataanalyst.com](mailto:ana.lazarevska@dataanalyst.com)

# Thank you for your attention!

**Link:**

GitHub: [https://github.com/AnaLazarevska/climatewins\\_ml\\_real\\_world\\_applications](https://github.com/AnaLazarevska/climatewins_ml_real_world_applications)