Machine Learning Specialisation

ClimateWins

Weather Conditions and Climate Change

Prepared by Ana Lazarevska, November 2024

Contents

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
- Predicting Long-Term Climate Trends with RNNs and Deep Learning
- 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

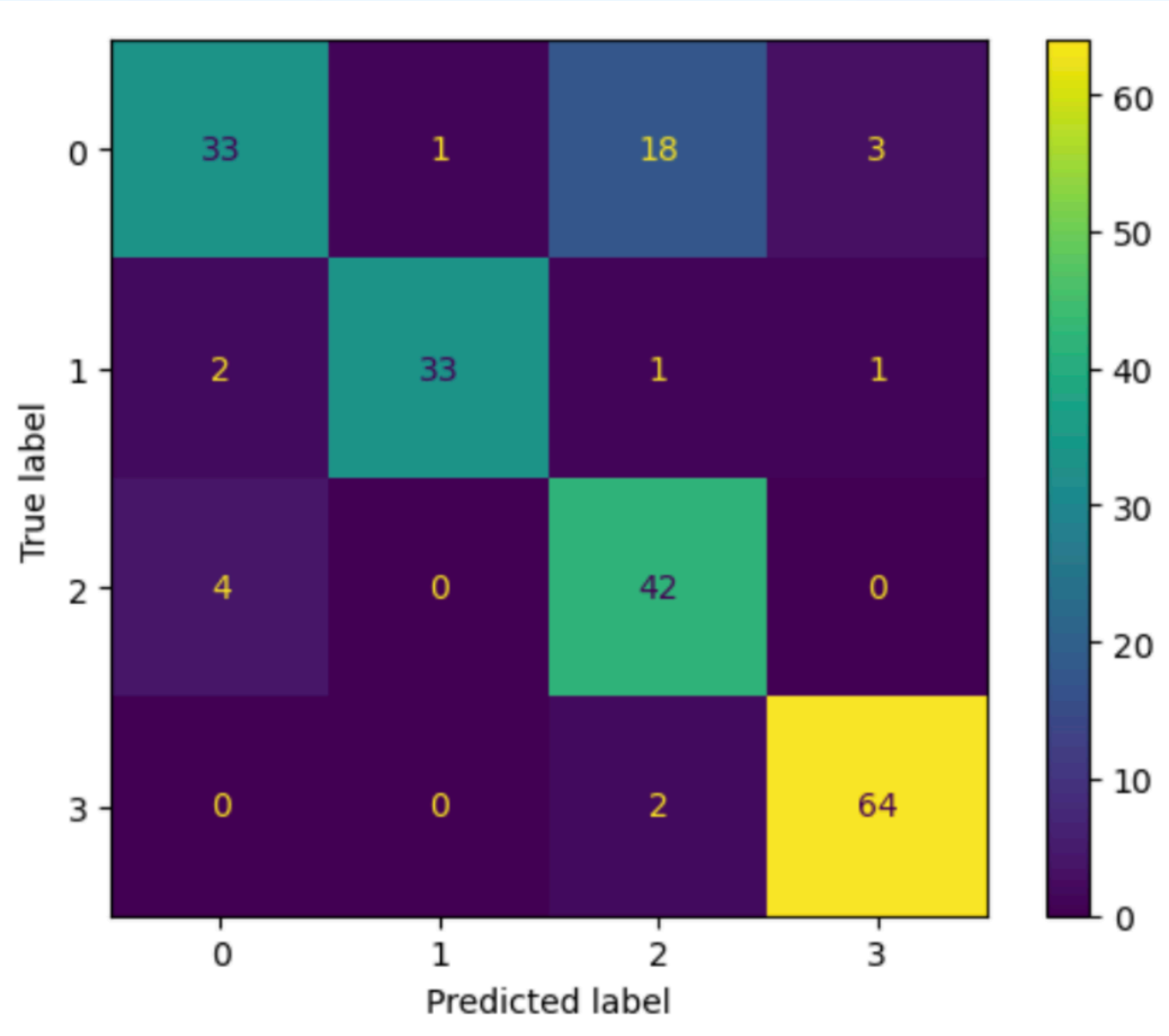
1. **Data Collection:** Gather historical weather data, including temperature, precipitation, wind speed, and atmospheric pressure.
2. **GAN Training:** Train a GAN to generate synthetic weather data that mimics real-world patterns, including extreme events.
3. **Extreme Event Simulation:** Use the trained 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.

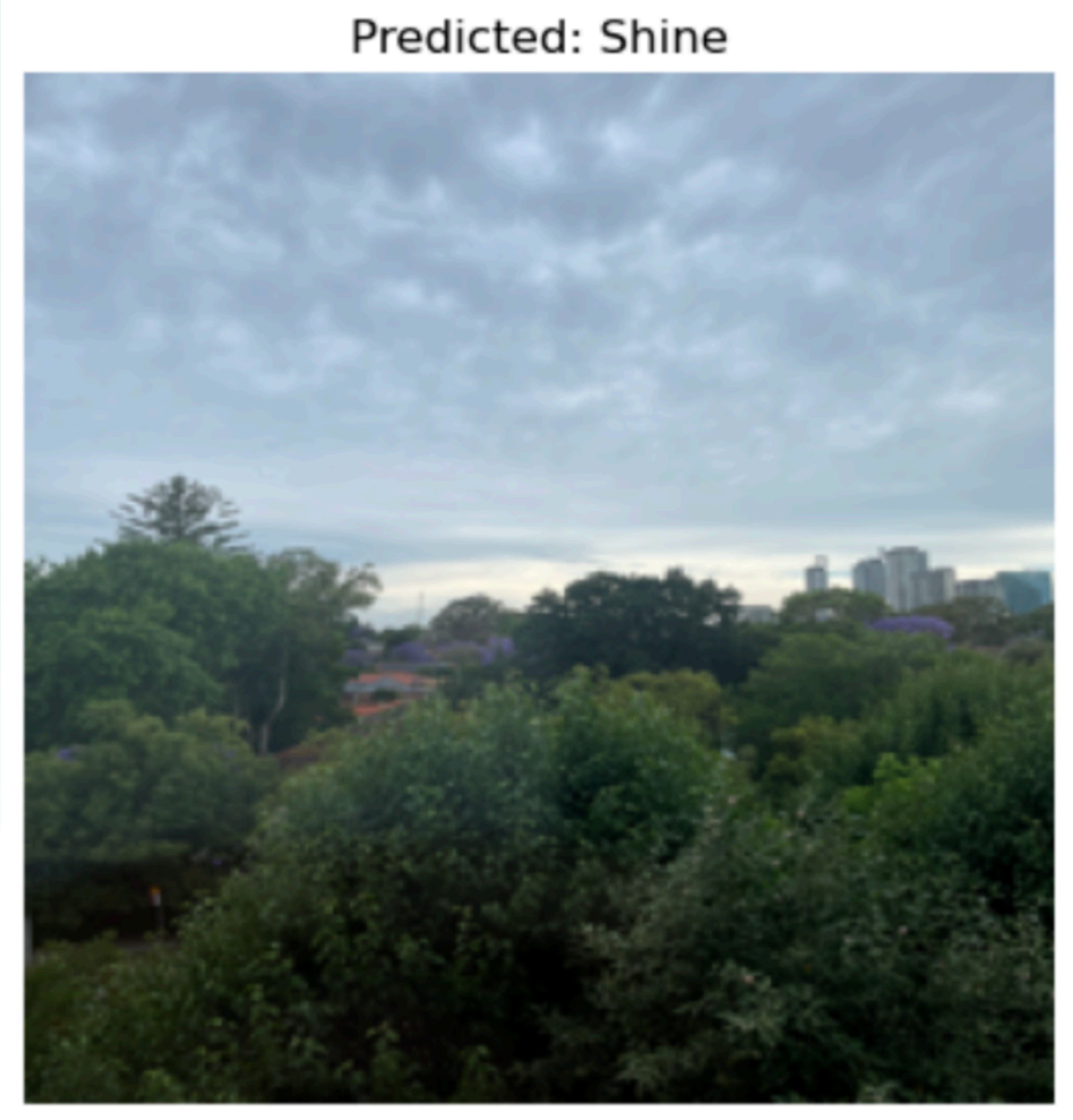
Expected Outcomes

- **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 struggles to identify as cloudy any image that has much brightness. As seen on the image above, the cloudy and overcast day was predicted as sunny (shine).

This is due to the bias in the dataset, where many of the shine images include clouds, but the clouds cover less than 50% of the sky.

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 analysing 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 / True	BASEL	BELGRADE	BUDAPEST	DEBILT	DUSSELDORF	HEATHROW	KASSEL
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 / True	LJUBLJANA	MAASTRICHT	MADRID	MUNCHENB	OSLO
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:

Predicting Long-Term Climate Trends with RNNs and Deep Learning

TE3: Predicting Long-Term Climate Trends with RNNs and Deep Learning

Linked Objective

- 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 Methods

- Recurrent Neural Networks (RNNs), Deep Learning

RNNs and other deep learning techniques can effectively predict long-term climate trends and future weather conditions. The RNN works well with historical data.

Data Needed

- Historical weather data (temperature, precipitation, wind speed, atmospheric pressure)
- Climate model projections (GCMs, RCMs)
- Socioeconomic data (population density, infrastructure vulnerability, economic indicators)
- Land use and land cover data
- Oceanographic data
- Atmospheric composition data

TE3: Predicting Long-Term Climate Trends with RNNs and Deep Learning

Methodology

1. **Data Collection:** Gather historical climate data, including temperature, precipitation, and atmospheric pressure.
2. **Feature Engineering:** Extract relevant features from the data, such as seasonal cycles, long-term trends, and climate indices.
3. **Model Training:** Train a deep learning model, such as a Long Short-Term Memory (LSTM) network, to learn the underlying patterns in the climate data.
4. **Future Prediction:** Use the trained model to generate predictions for future weather conditions, including temperature, precipitation, and extreme events.

Expected Outcomes

- **Climate Change Projections:** Provide insights into future climate change scenarios.
- **Policy and Planning:** Inform climate change adaptation and mitigation policies.
- **Risk Assessment:** Assess the long-term risks associated with climate change.

TE3: Predicting Long-Term Climate Trends with RNNs and Deep Learning

Practical Applications

```
Epoch 47/47
538/538 ━━━━━━━━━━━━━━━━━━━━ 1s 1ms/step - accuracy: 0.9889 - auc: 0.9991 - fn: 682.4174 - fp: 759.2764 - loss: 0.0291 - precision: 0.9729 - recall: 0.9759 - tn: 100751.8906 - tp: 27405.3047 - val_accuracy: 0.9861 - val_auc: 0.9982 - val_fn: 806.0000 - val_fp: 392.0000 - val_loss: 0.0373 - val_precision: 0.9780 - val_recall: 0.9557 - val_tn: 67467.0000 - val_tp: 17405.0000
```

Pred	BASSEL	BELGRADE	BUDAPEST	DEBILT	DUSSELDORF	HEATHROW	KASSEL \
True							
BASSEL	1234	553	324	19	54	34	16
BELGRADE	1	802	87	22	15	6	8
BUDAPEST	2	1	129	2	10	0	5
DEBILT	0	0	0	15	18	4	7
DUSSELDORF	1	0	0	0	4	3	0
HEATHROW	0	1	0	0	0	17	0
KASSEL	0	0	0	0	0	0	5
LJUBLJANA	1	1	0	0	0	0	0
MAASTRICHT	0	0	0	0	0	0	0
MADRID	0	1	0	0	0	0	0
MUNCHENB	0	0	0	0	0	0	0
OSLO	0	0	0	0	0	0	0
STOCKHOLM	0	0	0	0	0	0	0
VALENTIA	0	0	0	0	0	0	0

Pred	LJUBLJANA	MAASTRICHT	MADRID	MUNCHENB	OSLO	STOCKHOLM	VALENTIA
True							
BASSEL	140	231	914	17	56	1	89
BELGRADE	2	6	140	1	1	1	0
BUDAPEST	4	3	55	0	2	0	1
DEBILT	2	3	24	1	6	2	0
DUSSELDORF	1	0	20	0	0	0	0
HEATHROW	1	0	61	0	1	0	1
KASSEL	0	1	5	0	0	0	0
LJUBLJANA	22	0	35	0	1	0	1
MAASTRICHT	0	6	1	1	0	0	1
MADRID	0	0	457	0	0	0	0
MUNCHENB	0	0	0	7	1	0	0
OSLO	0	0	0	0	5	0	0
STOCKHOLM	0	0	0	0	0	4	0
VALENTIA	0	0	0	0	0	0	1

- Using a RNN model with batch_size of 32 and 47 epochs, achieved high level of accuracy. The training accuracy is 99.31%, and the validation (test) accuracy is 99.33%.

- Both the training and validation accuracy have significantly improved compared to the previous results, when I used parameters identical to those used in the CNN model. Reducing the batch size to 32 positively impacted the model's performance.

Thought Experiment 4:

Identifying Climate Change Hotspots with Random Forests

TE4: 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

TE4: Identifying Climate Change Hotspots with Random Forests

Methodology

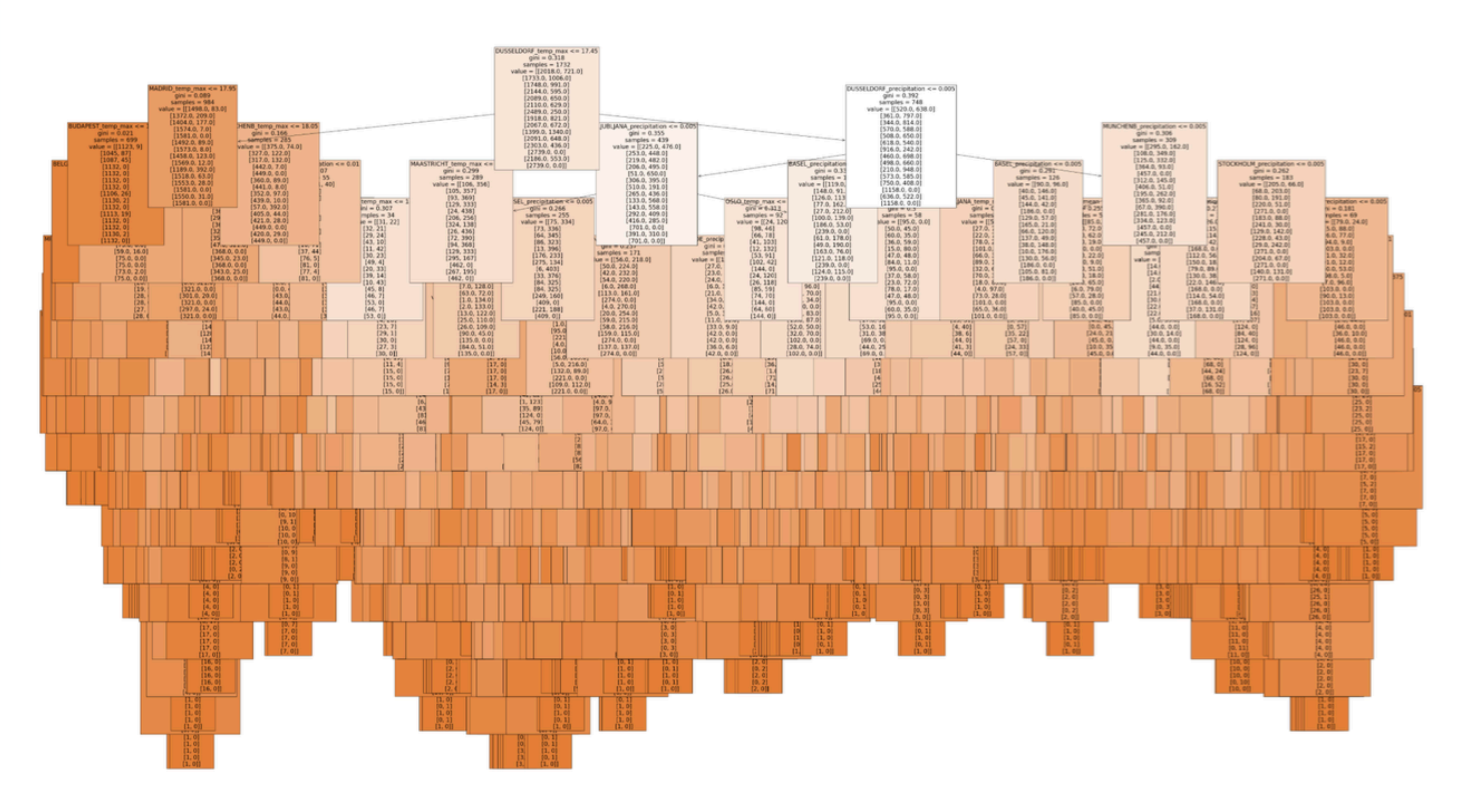
1. **Data Collection:** Gather relevant data from various sources, including meteorological stations, satellite sensors, and government agencies.
2. **Data Preprocessing:** Clean and preprocess the data to handle missing values, outliers, and inconsistencies.
3. **Feature Engineering:** Create relevant 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.

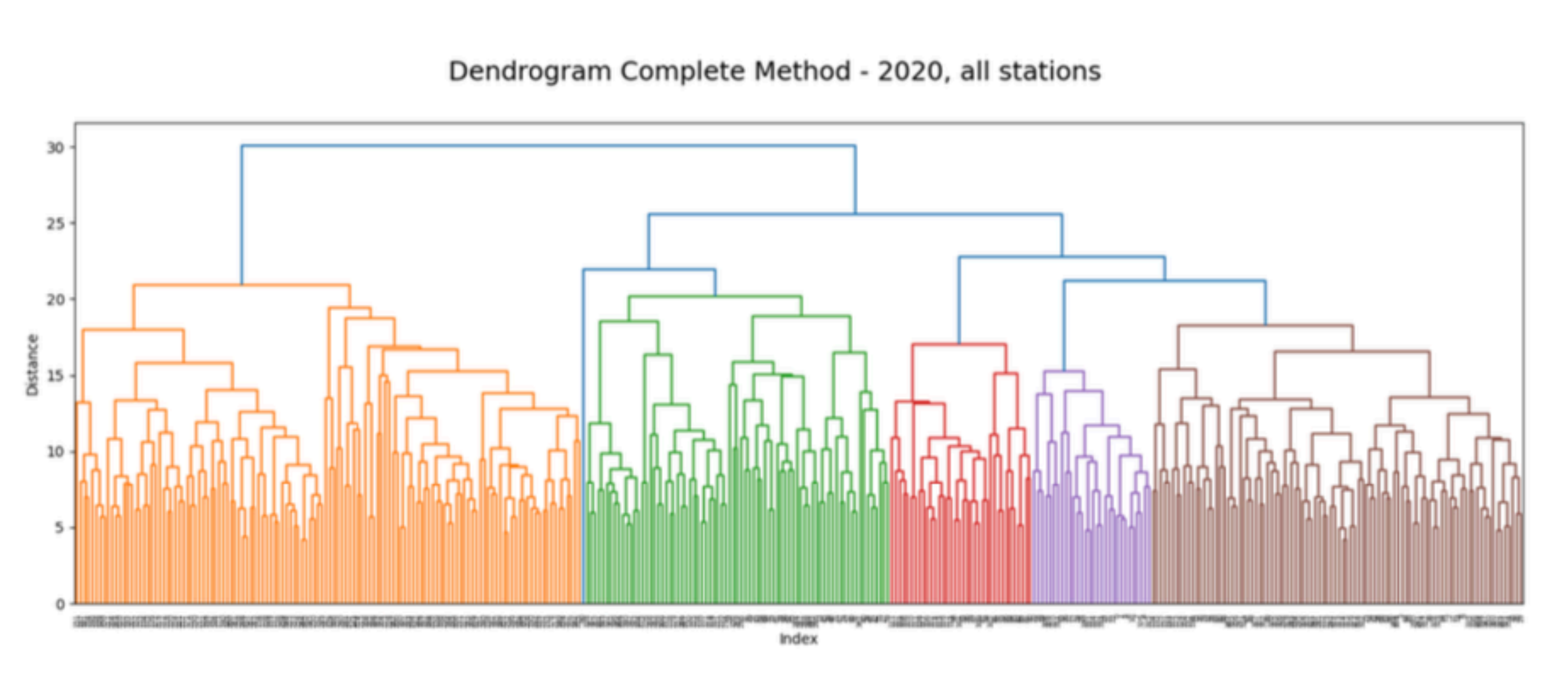
Expected Outcomes

- **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.

TE4: Identifying Climate Change Hotspots with Random Forests

Practical Application





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

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%.

Random Forests can be less accurate when working with less data points, such as in this case when only one decade of data was used.

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:
- Successfully identified unusual weather patterns, particularly those involving brightness and cloud cover.
- The GAN model generated realistic images of various weather conditions.

CONS:
- The model's performance was negatively impacted by the bias in the dataset, leading to misclassifications of cloudy images.

TE2: Identifying Unusual Weather Patterns with CNNs

PROS:
- The optimised CNN model achieved a substantial increase in accuracy.
- Successfully identified and classified different weather patterns.

CONS:
- The model's performance was limited to a specific set of weather stations.

TE3: Predicting Long-Term Climate Trends with RNNs and Deep Learning

PROS:
- Achieved high accuracy in both training and validation sets.
- Successfully captured temporal dependencies in the climate data.

CONS:
- The high accuracy might be due to overfitting, especially with a small dataset.

TE4: Identifying Climate Change Hotspots with Random Forests

PROS:
- Successfully identified clusters of cities with similar climate conditions.
- Provides insights into the underlying factors driving climate patterns.

CONS:
- The Random Forest model achieved moderate accuracy in classifying weather conditions.
- The model's performance is heavily dependent on the quality and relevance of the features.

Recommendations

Based on the provided information, Thought Experiment 2: CNNs for Identifying Unusual Weather Patterns appears to be the most promising.

- The CNN model, after optimisation, demonstrated high accuracy in identifying various weather patterns.
- It directly addresses ClimateWins' objective of identifying unusual weather patterns.
- CNNs can be scaled to handle large datasets and complex patterns.

While the other thought experiments have their strengths, they also have limitations. For example, GANs can be computationally expensive and sensitive to data quality, while the Random Forests may require careful feature engineering and long-term data.

While the RNN model showed promising results in terms of accuracy, there are a few considerations to keep in mind.

- RNNs often require large amounts of sequential data to train effectively. If the dataset is limited or the data quality is inconsistent, the model's performance may degrade.
- RNNs can be computationally expensive to train, especially for long sequences and deep architectures. This can limit their scalability and practicality.
- RNNs can be prone to overfitting, especially when trained on small datasets. This can lead to poor generalisation performance on unseen data.

While the RNN model performed well in this specific experiment, it's important to consider the broader context and potential limitations. The CNN model, on the other hand, demonstrated strong performance and scalability, making it a more robust and reliable choice for identifying unusual weather patterns.

ClimateWins should combining the strengths of both CNN and RNN models. For example, the CNN could be used to extract relevant features from the input data, and then these features can be fed into an RNN to capture temporal dependencies. This hybrid approach can lead to improved performance and better generalisation.

Questions?

Contact: aplazarevska@gmail.com

Thank you for your attention!

Link:

GitHub: https://github.com/AnaLazarevska/climatewins_ml_real_world_applications