Predicting Weather Patterns in Europe

Machine Learning Analysis - ClimateWins

Kyle Stanford

Updated: November 2nd, 2024

Agenda

1. Project Objectives (and thought experiment overview)
2. Machine Learning Algorithms Overview
3. Thought Experiments
4. Social & Ethical Considerations
5. Summary & Recommendations

Project Objectives

1. Identify weather patterns outside the regional norm in Europe
2. Determine if unusual weather patterns are increasing
3. Generate possibilities for future weather conditions over the next 25-50 years based on current trends
4. Determine the safest places for people to live in Europe over the next 25-50 years

Thought Experiments

1. Unusual Weather Pattern Detection

Use supervised and unsupervised models to identify anomalies and assess trends of unusual patterns.

Objective 1 and 2

2. Forecasting Future Weather Scenarios

Generate plausible future climate conditions across Europe using time-series models.

Objective 3

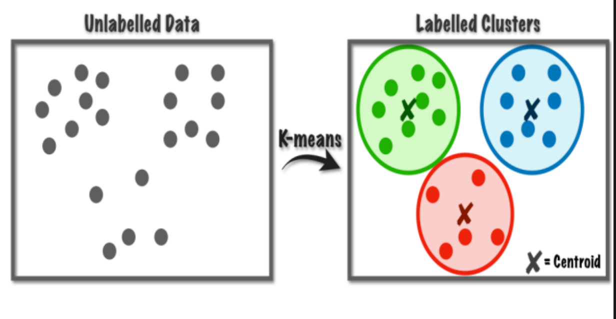
3. Identifying Safe Regions

Combine forecasted weather conditions with classification models to identify safe locations.

Objective 4

Overview: Machine Learning Algorithms

K-Means Clustering



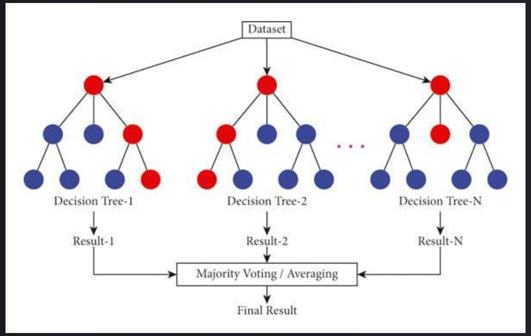
How it works:

Groups similar data points into clusters.

Applications:

Create groupings in weather data based on similar patterns and measurement values. Theoretically, anomalies should also stand out.

Random Forests



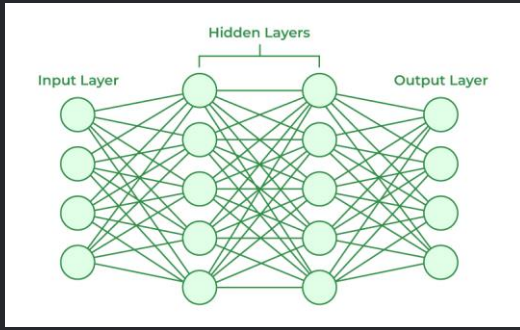
How it works:

Combines results from multiple decision trees to make more accurate predictions.

Applications:

Predict and categorise weather patterns across Europe (and the globe).
Also provides insights into the most important features in its predictions.

Deep Learning (LSTM / RNNs)



How it works:

Interacts with large datasets to extract important and complex patterns.

Applications:

LSTM models can capture long-term dependencies in sequential data, making it ideal for forecasting long-term weather patterns.

1. Unusual Weather Pattern Detection

Idea

Identify and investigate weather patterns that are outside of regional norms to determine if unusual weather is becoming more frequent and/or intense.

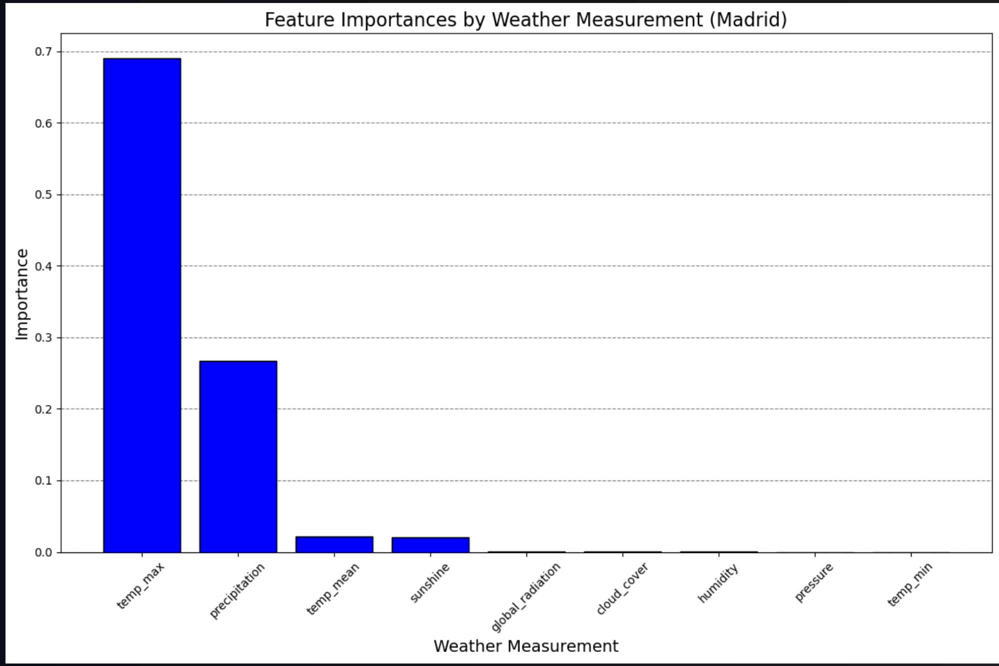
Required Data

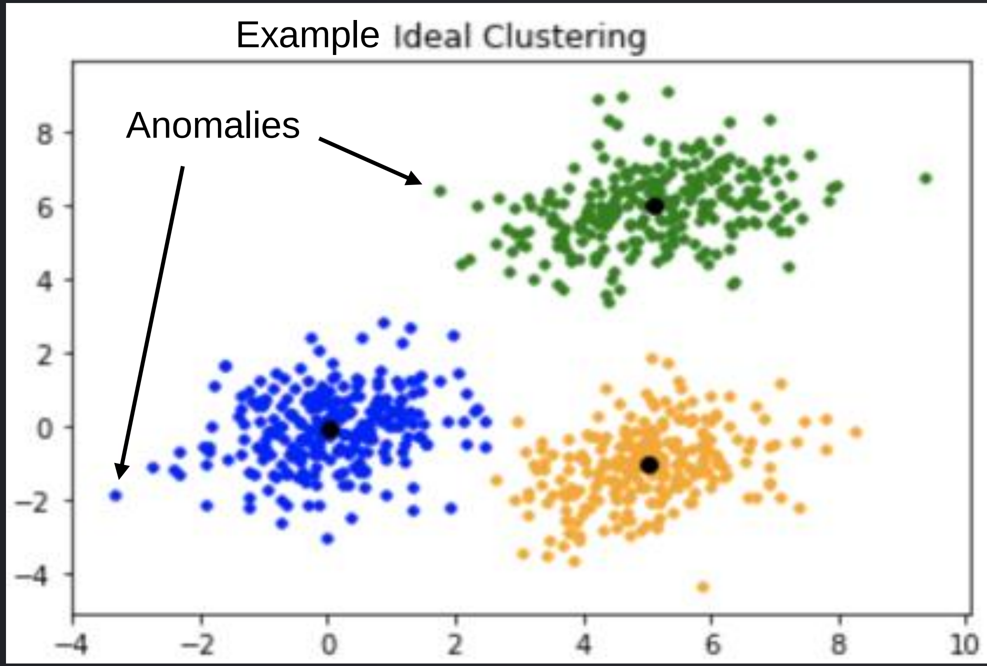
- **Historical Weather Data** (from various regions across Europe for the last 50 to 100 years)

Approach

- **K-Means Clustering:** Use to identify and label similar data (similar weather patterns). Anomalies should also stand out as either being far away from the centre of their assigned cluster or by being clustered in smaller groups.
- **Random Forests:** Use feature importances to assess which features most significantly contribute to identifying anomalies.
- Group anomalies and analyse their occurrences overtime to detect shifting weather patterns.

1. Put into Practice:





Plot of Feature Importances from Random Forest Model trained to recognize ‘pleasant’ and ‘unpleasant’ weather data.

Random Forest Accuracy: 100% (after optimization)

K-means clustering would be used on the historical weather to essentially create labels pertaining to different weather conditions.

Extreme weather could also be better identified by assigning labels based on the distance a data point is away from its centroid.

2. Forecasting Future Weather Scenarios

Idea

Generate plausible future weather scenarios across Europe and analyse the forecast for extreme changes in weather measurements of interest.

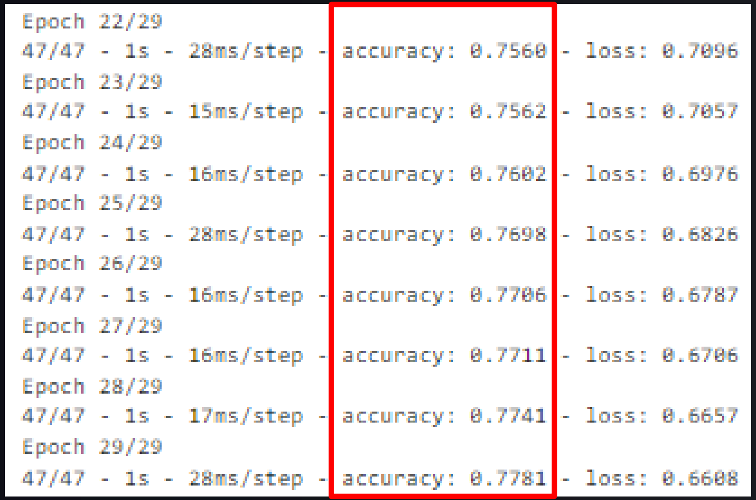
Required Data

- **Historical Weather Data** (from various regions across Europe for the last 50 to 100 years)
- **Climate Data and Projections** (for factors like CO_2 emissions and population growth) (**World Bank**)

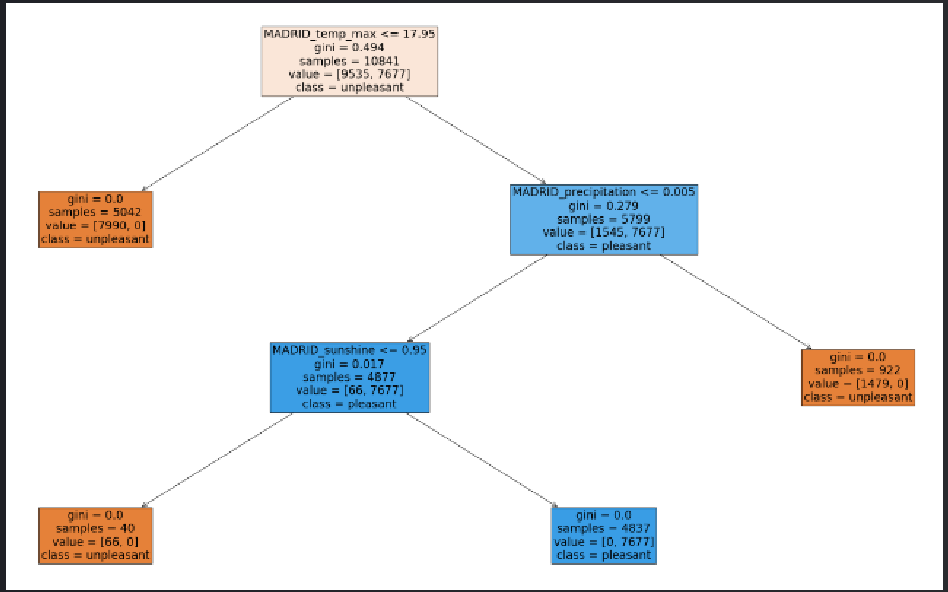
Approach

- **LSTM Network:** train the model on historical weather data and climate projections to allow the model to forecast weather trends for the next 25 to 50 years, also categorising weather as ‘mild’, ‘severe’, or ‘dangerous.’
- **Random Forests:** use to assess which features in the data contribute to categorisation, allowing future analyses to focus on key weather concerns.
- **GANs:** Could supplement LSTM network forecasting, providing diverse weather scenarios using satellite images and historical data.

2. Put into Practice:



```
Epoch 22/29
47/47 - 1s - 28ms/step - accuracy: 0.7560 - loss: 0.7096
Epoch 23/29
47/47 - 1s - 15ms/step - accuracy: 0.7562 - loss: 0.7057
Epoch 24/29
47/47 - 1s - 16ms/step - accuracy: 0.7602 - loss: 0.6976
Epoch 25/29
47/47 - 1s - 28ms/step - accuracy: 0.7698 - loss: 0.6826
Epoch 26/29
47/47 - 1s - 16ms/step - accuracy: 0.7706 - loss: 0.6787
Epoch 27/29
47/47 - 1s - 16ms/step - accuracy: 0.7711 - loss: 0.6706
Epoch 28/29
47/47 - 1s - 17ms/step - accuracy: 0.7741 - loss: 0.6657
Epoch 29/29
47/47 - 1s - 28ms/step - accuracy: 0.7781 - loss: 0.6608
```



The above snippet came from an **optimised LSTM model** that was trained to recognized pleasant and unpleasant weather.

Its training accuracy was **77.81%** but struggled with testing data (more work is needed).

Example of **decision tree** from **Random Forest** showing how it makes its decisions when categorising the weather in Madrid.

One tree looks overly simply, but many decision trees come together to make for a more accurate model in a Random Forest.

3. Identifying Safe Regions

Idea

Using classification models, identify and group the safest European regions based on a combination of safety metrics and forecasted weather data.

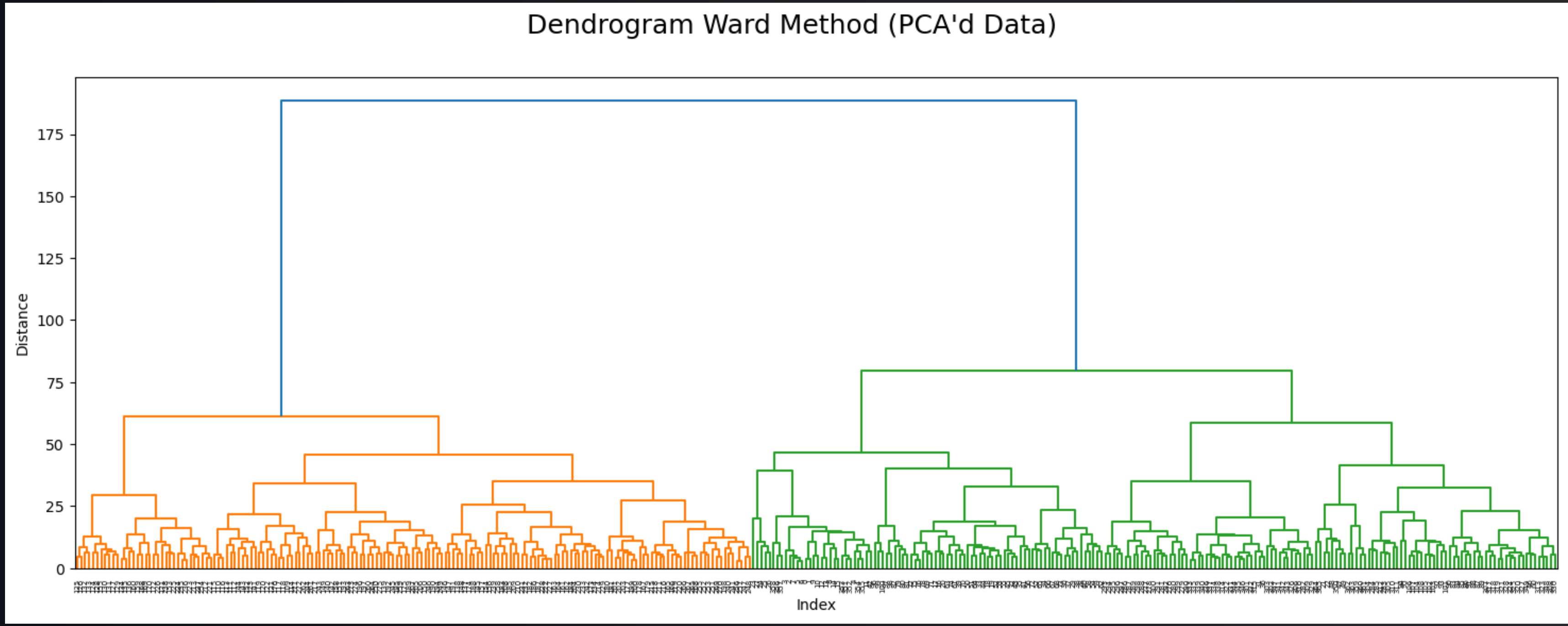
Required Data

- **Historical Weather Data for Europe**
- **Natural Disaster Incidents Data**
- **Infrastructure Resilience Data**

Approach

- **Defining Safety:** This is somewhat interpretable, but an elementary guess would be several scores based on:
  - Frequency/severity of natural disasters
  - Resilience of local infrastructure
  - Erratic weather patterns (documented and forecasted)
- **K-means or Hierarchical Clustering**: Used to compare regions based on their safety scores. This could be done before and after integrating forecasted weather data to observe changes to region safety.

3. Put into Practice:



Dendrogram Ward Method using European weather data after reducing data dimensions by performing **Principal Component Analysis**.

This method shows promise in dividing data into distinct clusters based on similarity, indicating that key patterns and features would be prominent and interpretable.

Social and Ethical Considerations

Transparency

- Clear communication for public trust.
- Make information regarding models, rationale for models used, and limitations available.

Accountability

- Models may influence environmental and government initiatives.
- Models must have regular oversight and be subject to scrutiny.
- ClimateWins must be prepared to take accountability for errors.

Equity

- Models should account for all communities
- Ensure all communities benefit from models’ insights

Summary & Recommendations

Scenario 1: Unusual Weather Pattern Detection

- ✓ Relatively simple models and therefore easiest implementation & interpretation.
- ✓ Historical weather data readily available.
- ✓ Capacity for integration with other models (CNN) and real-time data sources (e.g., satellite images).

- ✕ K-Means Clustering assumes spherical, equally-sized clusters, which may not fit true structure of historical data.
- ✕ Somewhat arbitrary act of pre-defining the number of clusters (potentially leading to inconsistent results).

Scenario 2: Forecasting Future Weather Scenarios

- ✓ Allows for greater understanding of future climate risks.
- ✓ LSTM model is known to handle time-series data effectively for forecasting.
- ✓ GANs allow for multiple scenarios to be explored.

- ✕ LSTM models can be computationally expensive.
- ✕ Model complexity necessitates expertise to ensure correct model infrastructure, tuning, and maintenance.

Scenario 3: Identifying Safe Regions

- ✓ Can be applied to historical weather data to assess safe regions now and in the future (pending results of previous scenarios).
- ✓ Analysis captures multiple safety-factors.

- ✕ Identification of future safe regions is dependent on forecasted data from scenario 2.
- ✕ Potential difficulty deciding where to cut dendrograms to form clusters (affecting interpretability).
- ✕ Hierarchical clustering is sensitive to noise/outliers.

Final Recommendations

Scenario 1: Unusual Weather Pattern Detection

Scenario 2: Forecasting Future Weather Scenarios

Scenario 3: Identifying Safe Regions

1) Immediately develop models to detect unusual weather patterns.

- This is the most achievable project which will also yield short-term results.

2) Start allocating resources to develop a LSTM model for long-term forecasting.

- Comparing results of this model with Scenario 1 can help address model inconsistencies and strengthen short- to mid-term results

3) Once Scenario 1 and 2 gain momentum, begin working on Scenario 3.

- This will require careful stakeholder management and staff dedicated to policy research to ensure social inequities or tensions are not worsened by the identification of “unsafe” regions.

Thank you

Have Any Questions?