Machine Learning Models for Weather Prediction: ClimateWins

NIRAV BARIYA

04/07/2024

Overview

Objective: This project is aimed at utilize machine learning techniques to predict the impacts and consequences of climate change. Through our project, we wish to recommend resources, algorithms, and data ClimateWins will need to predict weather changes.

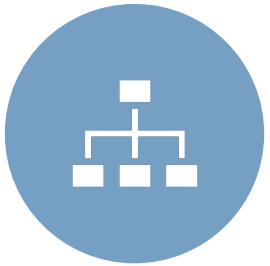
Thought Experiments:

One: Clustering algorithms may help us find changes in weather patterns over a period.

Two: Classification algorithms can be used to find out weather patterns that are outside regional norms and find out important features that may help us predicting extreme weather events.

Three: Generative adversarial networks have the potential to be used to generate images that will help us visualize future extreme weather events.

Machine Learning Algorithms



Hierarchical Clustering is an unsupervised machine leaning algorithm to find clusters in the dataset. We can use them to changes in weather patterns over a period.

Convolutional Neural Network is a deep learning algorithm that works well with spatial data. This algorithm is useful to identify weather unusual weather patterns are increasing.

Random forests are an ensemble learning algorithm that combine multiple decision trees to improve predictive performance. We can use them to identify important features that are related to extreme weather events.

Generative Adversarial Algorithms are used to create synthetics data by training on real data sets using Convolutional Neural Networks. These can be used to generate future weather scenarios.

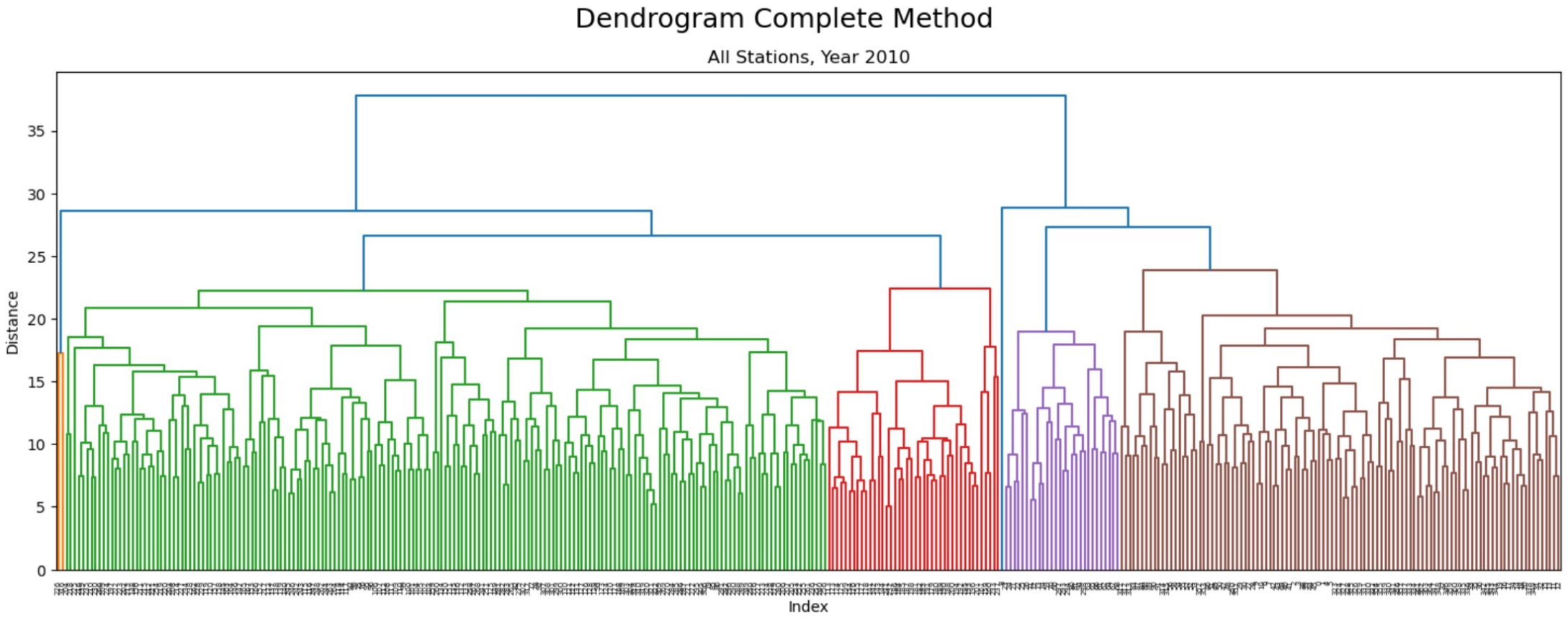
1. Changes in Weather Patterns

Idea: If we are to find new weather patterns over years, we can perhaps group similar weather patterns together for each year. Then, we can analyze each these groups to see if there are any significant changes that has taken place over the span of 60 years.

Data Requirements: To perform the analysis of groups of weather patterns we need to have historical weather data of last 60 years that includes features such as temperature, precipitation, snow fall, sunshine, radiation, and so on.

Algorithms: We are proposing **Hierarchical clustering algorithm** to achieve this objective. This algorithm creates an upside-down tree, called a dendrogram, to create clusters of data points.

Dendrogram Example:



- The above example considered all weather stations across Europe for 2010. We can either focus on all weather stations across Europe at a time or look at individual weather stations at a time. We can also have control or how the between clusters is to be calculated.
- We can then analyze each of the cluster and identify if there are any changes in the weather patterns year over year.

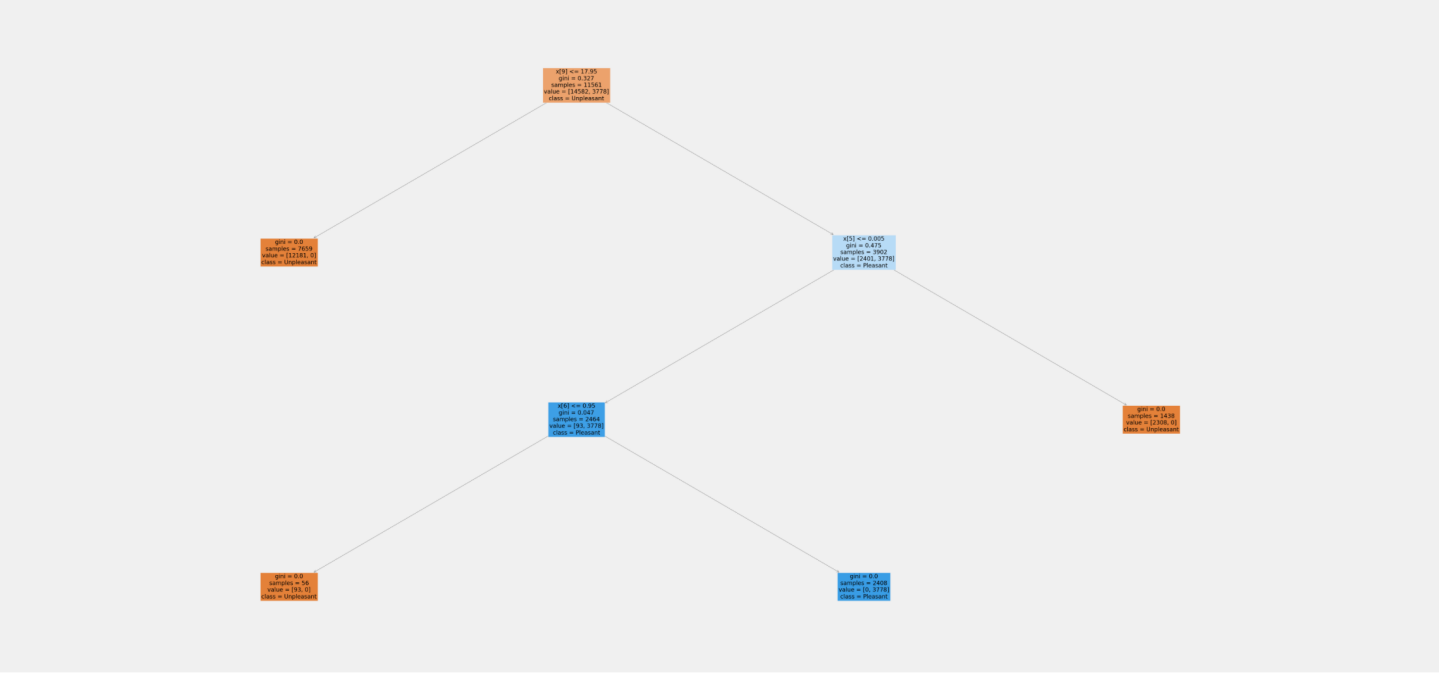
2. Identifying Weather Patterns Outside the Regional Norm in Europe

Idea: To predict extreme or unusual weather patterns in future we can train classification algorithms that accurately identifies the features that are important for weather prediction and train another model using the same features to learn intricate weather patterns.

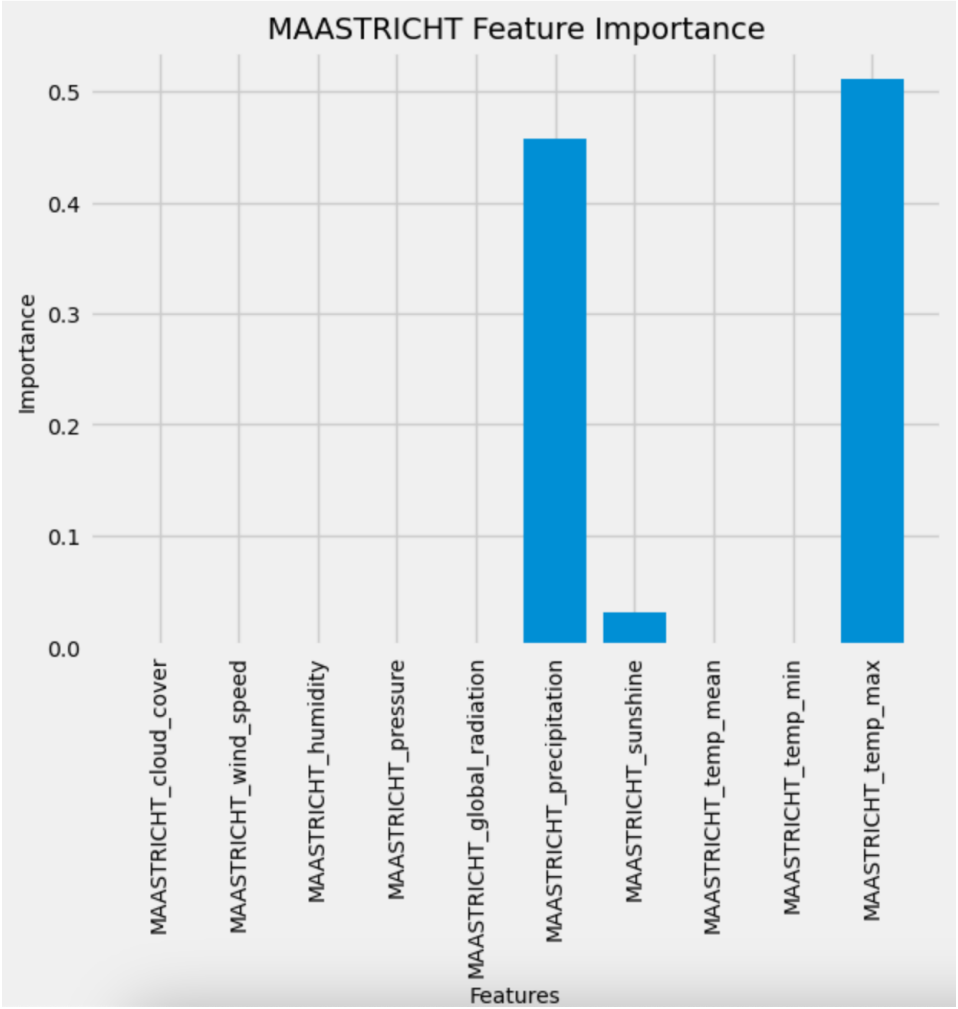
Data Requirements: Historical time-series data of weather data including extreme weather events over last 100 years and images of different weather conditions.

Algorithms: We propose that we use **Random Forests classifier** to classify extreme weather events and identify important features. Then use these important features with **Long-Short Term Memory** (LSTM) neural network to predict future extreme events.

Random Forests Example



One of the decision trees for weather station Maastricht, Netherlands. The accuracy of the Random Forest Classifier was 100% for predicting pleasant or unpleasant days.



We can identify important features that contributes to the accuracy of Random Forests the most. In this case, maximum temperature, precipitation, and sunshine were important.

Optimized Convolutional Neural Network for Classification

Model Accuracy: 95.96%

Confusion Matrix

True \ Pred	BASEL	BELGRADE	BUDAPEST	DEBILT	DUSSELDORF	HEATHROW	KASSEL
BASEL	2855	49	5	1	0	1	0
BELGRADE	47	825	5	0	0	1	1
BUDAPEST	11	18	145	1	0	1	0
DEBILT	2	5	14	55	1	1	0
DUSSELDORF	5	0	0	5	15	7	0
HEATHROW	3	0	0	0	4	53	0
KASSEL	0	0	2	1	0	0	4
LJUBLJANA	3	1	0	0	0	0	0
MAASTRICHT	1	0	1	0	2	0	0
MADRID	22	0	3	0	0	3	0
MUNCHENB	3	0	0	0	0	1	0
OSLO	2	0	0	0	0	0	0
STOCKHOLM	0	0	0	0	0	0	0
VALENTIA	2	0	0	0	0	0	0

True \ Pred	LJUBLJANA	MAASTRICHT	MADRID	MUNCHENB	OSLO	STOCKHOLM
BASEL	1	0	6	1	0	0
BELGRADE	0	0	0	0	0	0
BUDAPEST	3	0	3	0	0	0
DEBILT	0	0	0	0	0	0
DUSSELDORF	1	1	2	0	0	0
HEATHROW	2	0	7	0	0	0
KASSEL	0	0	1	0	0	0
LJUBLJANA	40	0	4	0	0	0
MAASTRICHT	0	2	1	0	0	0
MADRID	1	0	324	0	0	0
MUNCHENB	0	0	0	0	0	0
OSLO	0	0	0	0	2	0
STOCKHOLM	0	0	0	0	0	1
VALENTIA	0	0	0	0	0	0

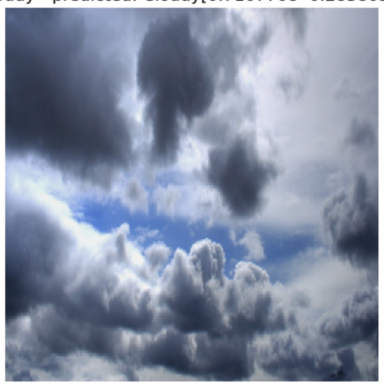
Note: To predict future extreme events, we may use Recurrent Neural Networks(RNN) such as Long-Short Term Memory(LSTM) which works better with time-series data.

Extending CNN for Predicting Weather Events using Images

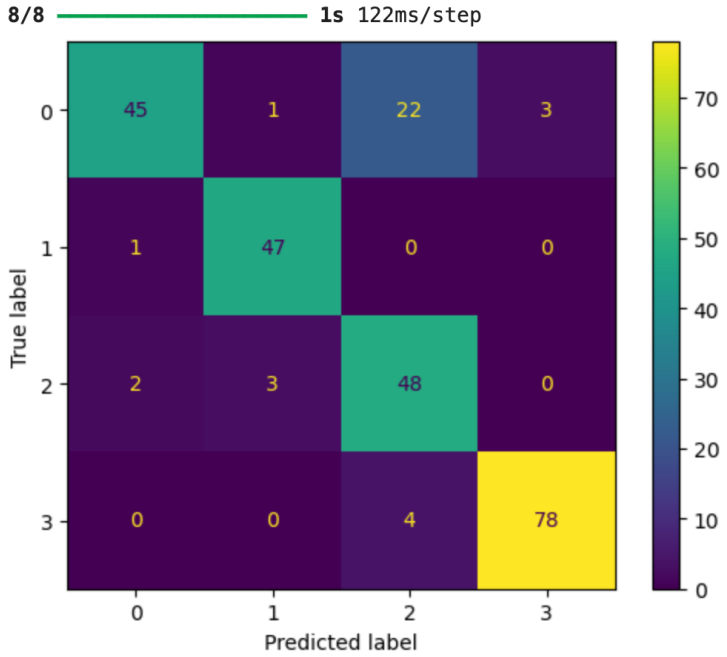
Model Accuracy: 93.33%

Correct Prediction - class: Sunrise - predicted: Sunrise[6.6560646e-12 2.5988825e-12 1.54

Correct Prediction - class: Cloudy - predicted: Cloudy[0.7107708 0.2838036 0.00414516 0.00128037]



Confusion Matrix



We can also train CNN on image data to predict future weather events based on images.

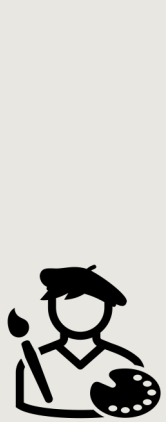
3. Visualize Future Extreme Weather Events

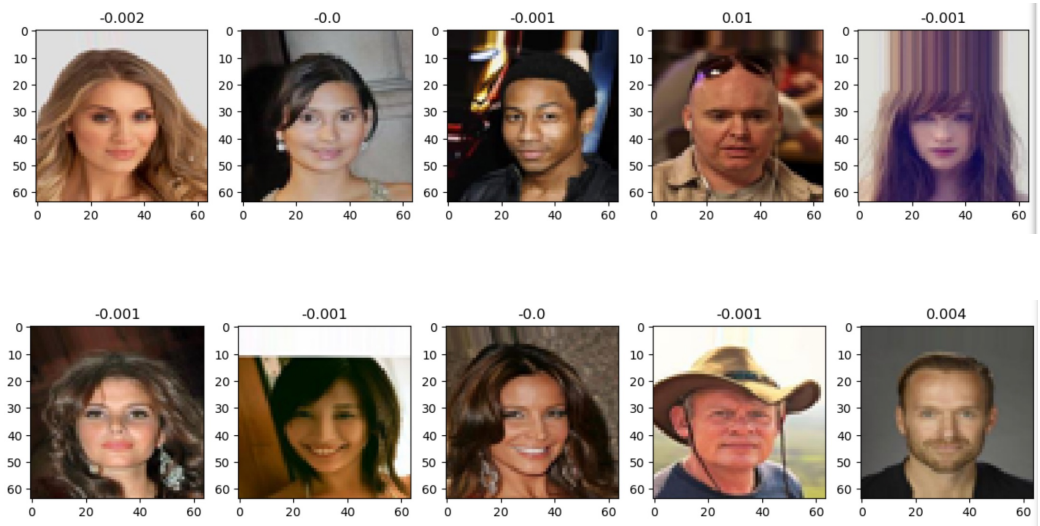
Idea: To simulate what the future extreme weather scenario may look like we can use images of historical scenarios of extreme events. This can help us determine the possible impacts of weather events and prepare for such events.

Data Requirements: Images of historical extreme weather events.

Algorithms: We propose **Generative Adversarial Algorithms(GAN)** to generate these images. These algorithms will be based on **Convolutional Neural Networks.**

Generative Adversarial Networks(GAN) Example





These celebrity images were generated using real celebrity images. A similar approach can be used to visualize future weather events.

The most difficult part for these kind of models is to make them converge.

Conclusion and Recommendations

We saw three different thought experiments using:

- Hierarchical Clustering for identifying changing patterns
- Random Forests and CNN for extreme weather prediction
- GAN for visualizing future events

Recommendation

- All three algorithm fulfill a different object that ClimateWins has for European weather prediction and can be further developed.
- The algorithms with highest accuracy after optimization were Random Forests(100%) and CNN(95.96%).
- We can develop more on the second thought experiment that predicts future events using Random Forests (identifying key features), CNN(capturing intricate patterns), and LSTM for forecasting future events.
- Algorithms can be trained using image data on extreme weather events.

Thank You

Do you have any questions? I'm happy to answer any questions or discuss ideas with you.

Reach me out using the following links:

nkb.bariya@gmail.com

https://github.com/b-nirav

https://www.linkedin.com/in/nirav-bariya/