Implementing Machine Learning for Weather Prediction

ClimateWins Project

Susan Wang | November 23, 2025

Exploring ClimateWins' Goals

Objective: ClimateWins, a European nonprofit organization, is interested in using machine learning to help predict the consequences of climate change around Europe.

Thought experiments can assist in formulating methods in the use of machine learning to predict extreme weather events.

1. Location:
Can machine learning models accurately identify future extreme weather events based on weather data from the last 60 years?

2. Timing:
If we determine that extreme weather events are increasing, can we predict when and how frequent these events will occur?

3. Climate Havens:
Can machine learning help forecast where the most climate-safe places to live in Europe over the next 25 or 50 years are?

Machine Learning Options

These machine learning models will likely be most effective in achieving ClimateWin's Goals:

Model/ Algorithm	What is it/ What is it good at	Application for ClimateWins
Autoregressive Integrated Moving Average (ARIMA)	• Regression algorithm for identifying trends in a time series	Identifying weather trends over time
Hierarchical Clustering	• Unsupervised learning model • Agglomerative clustering for classification	Classifying extreme weather events
Apriori	• Unsupervised algorithm that classifies by association • Good for looking for common characteristics between features	Finding associations between different types of weather events (e.g. heatwaves & wildfires)
Artificial Neural Network (ANN)	• Supervised machine learning model • Pattern recognition and intuitive results	Classifying weather events from structured data
Convolution Neural Network (CNN)	• Type of artificial neural network • Good for assessing images and numerical data	Working with weather radar images
Long Short-Term Memory (LSTM)	• Type of recurrent neural network (RNN) • Good for handling temporal data	Forecasting weather events over long period of time
Generative Adversarial Network (GAN)	• Two neural networks, generates new data as the model learns • Good for creating visual record of what you want to predict	Creating images to reproduce and to simulate probability of weather events
Hidden Markor Model (HMM)	• Time-based models for speech, text, and bioinformatics	Predicting probabilities in weather forecasting

Data Wishlist

The historical weather data set provided by ClimateWins alone would not be enough to determine their goals.

Here are some suggested additional data:

ClimateWins DATA SET:

- Collected by the European Climate Assessment & Data Set project
- From 18 weather stations across Europe
- Records from late 1800's to 2022
- Values such as temperature, wind speed, snow, precipitation, global radiation and more

+ Data on Extreme Weather Events:
'Answer' data sets identifying past extreme weather events (wildfires, heatwaves, storms, floods).

+ Data on Population Density:
Population data from European cities near the 18 weather stations from the main data set.

+ Data on Geography:
Geographical features in Europe, including landforms, bodies of water, and forests.

+ Data on Infrastructure:
Information about the infrastructure in risk areas.

+ Weather Radar Images:
Images from weather radars during extreme weather events.

Thought Experiment on Location

Initial Thought:
Can machine learning models accurately identify future extreme weather events based on weather data from the last 60 years?

Follow-up Question:
Would models identify various types of events with equal accuracy? For example, would storms be easier to predict than wildfires?

Also:
Are there correlations between different types of events, such as heatwaves and wildfires? Would they compound risks in certain areas?

Steps for Analysis	Datasets Required	ML Model
Optimize a model to accurately identify extreme weather conditions.	• ClimateWins dataset • Extreme weather events dataset	ANN
Observe correlations between different weather events.	• Extreme weather events dataset	Apriori
Predict level of risk where extreme events will hit, assessing population density as well as geography and infrastructure.	• Population dataset • Infrastructure dataset • Geography dataset • Prediction results from first step	LSTM or HMM

Thought Experiment on Timing

Initial Thought:
If we determine that extreme weather events are increasing, can we predict when and how frequent these events will occur?

Follow-up Question:
How accurate will the predictions be? What time frames can we predict? Within weeks or months?

Also:
If we can predict when and how often these disasters will happen, will we be better prepared? Will there be enough time to make changes?

Steps for Analysis	Datasets Required	ML Model
For each weather event, determine if there is a trend of more frequent events occurring over the years.	• Extreme weather events dataset	ARIMA
Train model to predict the time and frequency of events, varying time frames for accuracy.	• ClimateWins dataset • Extreme weather events dataset	LSTM
Analyze the population and infrastructure by location to determine the optimal time frame for disaster preparation.	• Population dataset • Infrastructure dataset • Location results from trained model	LSTM or HMM

Thought Experiment on Climate Havens

Initial Thought:
Can machine learning help forecast where the most climate-safe places to live in Europe over the next 25 or 50 years are?

Follow-up Question:
Do we have enough information on different geographical areas? Are the available weather stations accurately representing the wide range of locations?

Also:
Will predicting the 'safest' places to live unintentionally spur people to migrate instead of investing in preparation and improving infrastructure where they are living now?

Steps for Analysis	Datasets Required	ML Model
Use weather radar images to train models to predict weather in areas between stations.	• ClimateWins dataset • Weather radar images • Geography dataset	GAN and ANN
Determine the *least* and *most* safe places are in the next 25 or 50 years, assessing by count of each type of natural disaster.	• Expanded ClimateWins dataset from previous step • Extreme weather events dataset	LSTM or HMM
Implement machine learning to predict flow of population into those 'safe' areas over the next 25 to 50 years.	• Population dataset • Extreme weather events dataset • Location results	ANN or LSTM

Conclusion

Summary:

Thought	Data	Algorithms
Location	ClimateWins, weather events, pop, geo, infra	ANN, Apriori, LSTM, HMM
Timing	ClimateWins, weather events, pop, geo, infra	ARIMA, LSTM, HMM
Havens	ClimateWins, weather events, pop, geo, radar	GAN, ANN, LSTM, HMM

Why the thought experiment on **Location** has the most potential:

- Provides a risk assessment for defined locations.
- Timing of events can come secondary after determining high risk locations.
- The process for determining climate-safe havens are less feasible and also carry uncertainty regarding population migration and distribution.

Next Steps:

1 **Collect data to run analysis:**
The additional data sets will need to be collected, cleaned, and prepared for running the machine learning models.

2 **Follow outlined steps and test accuracy:**
Run the machine learning models, optimizing for accuracy.
Make initial predictions for the following year and observe the performance of the model.

Thank you!

Feel free to contact me if you have questions or comments.

Visit the project on Github: