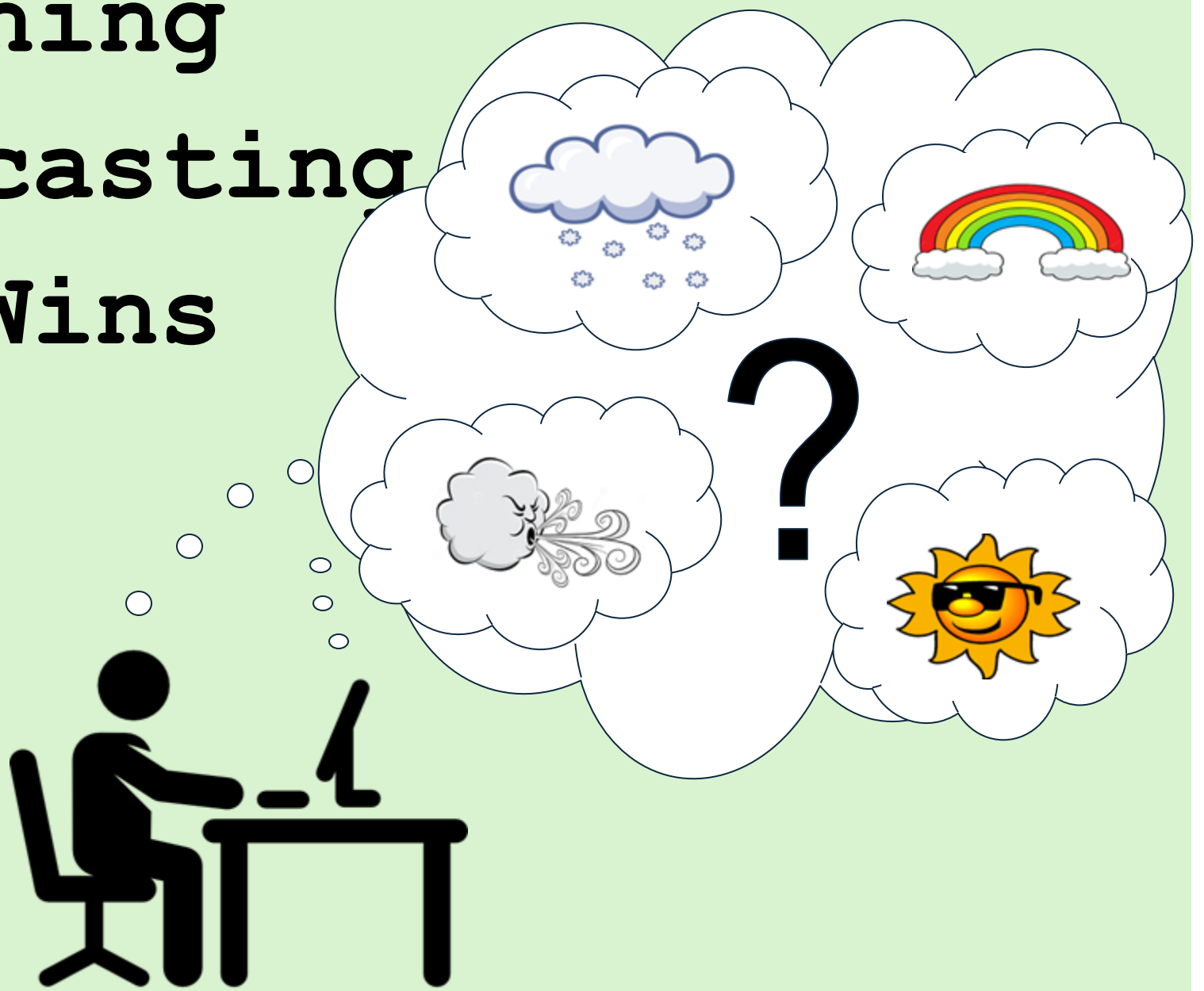
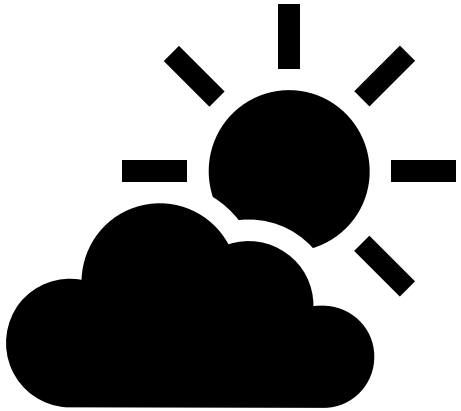
Machine Learning Weather Forecasting With ClimateWins



Nancy Kolaski

October, 2024

Objectives:



Identify weather patterns outside the regional norm in Europe.

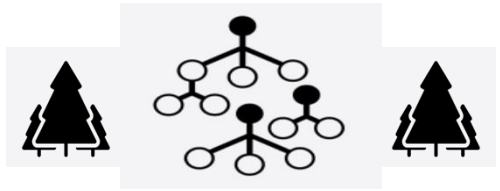
Determine if unusual weather patterns are increasing.

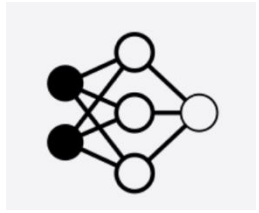
Generate possibilities for future weather conditions over the next 25 to 50 years based on current trends.

Determine the safest places for people to live in Europe over the next 25 to 50 years.

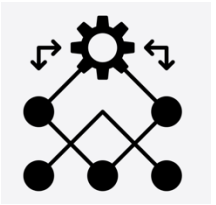
Algorithm Overview: Machine Learning

(Supervised/Unsupervised)





Discriminator



Generator

Random Forest & Clustering:

(using Dendrograms & Decision Trees)

Method

Combines multiple decision trees to make the most accurate predictions. Each decision tree trains on a random sample of the total data, with final prediction made by averaging the predictions of all trees.

Application

Identify and predict key weather features at specific locations across Europe to recognize extreme weather patterns.

Deep Learning with CNN & RNN:

(Convolutional Neural Network & Recurrent Neural Network)

Method

CNNs handle **images & numerical** data as they were inspired by visual cortex processes of the brain. RNNs handle temporal data such as **text, handwriting, and speech**. RNNs also use LSTM (long short-term memory), imitating a brain that forgets unimportant data – updating the significant data as needed)

Application

Many hidden layers learn features of data (edges/colors), with each layer increasing complexity of aspects it recognizes, advancing into the next layer, & so on. This aides in identifying weather trends over time for forecasting climate conditions in Europe.

GANs:

(Generative Adversarial Network)

Method

Use two adversarial neural networks working against each other – generator (creates data) & Discriminator (samples of real & artificial, discriminating which is real). Can create artificial data (text & images)

Application

Simulates weather changes to predict extreme weather to plan for future (safe European locations).

Thought Experiment 1:

Identify extreme weather conditions.

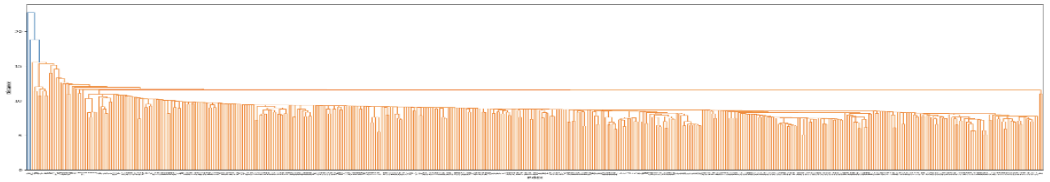
Goal: Use Hierarchical Clustering via Dendrograms and then combine with PCA (Principal Component Analysis) for dimension reduction to interpret large amounts of data to find weather trends

**note: Dendrograms are similar to Decision Tree with bottom-up approach instead.*

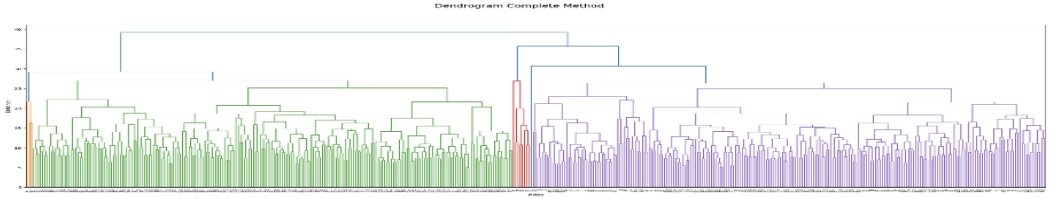
Approach: Compare all stations with Complete method (reduced data) to find trends, then narrow down specific locations to find the extreme weather condition patterns using Single method.

Data: historical weather data from various weather stations in Europe over the last 25-50 years

----------Dendrograms comparing all stations in 2010----------

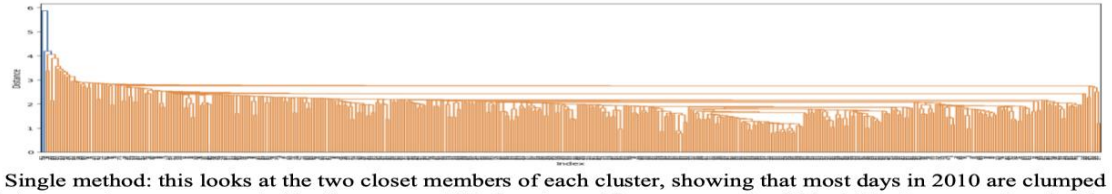


Single method: all stations/days clumped into one category – not useful



Complete method: separates into 4 categories, 2 large and 2 small. This will be interesting to look at as it is almost equally divided into two clusters with two smaller clusters.

----------Dendrograms comparing Madrid and Belgrade in 2010----------



Single method: this looks at the two closet members of each cluster, showing that most days in 2010 are clumped into one category, represented by the orange. This is not providing any meaningful insights. I am guessing that the small group to the left were classified outliers?

Complete method: This is looking at the distance between the farthest members of each cluster, broken down more evenly into 6 categories, a more even distribution for the year. This may be interesting to look at to gain insight from patterns in these clusters.

Dendrogram Complete Method *(after PCA to reduce data and find greatest variance within it)* gives clear, distinct clusters with less noise to find key patterns in the data (all weather stations per year)

Dendrogram Single Method can find outliers, interpreted as extreme weather conditions *when looking at specific weather stations*. It is not as useful when comparing all weather stations

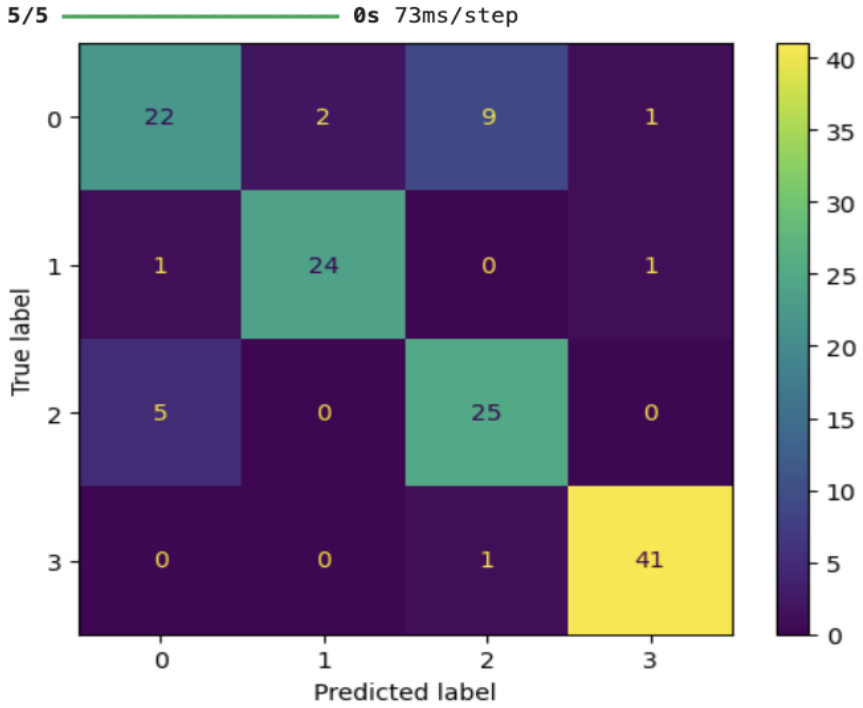
Thought Experiement-2:

Determine if unusual weather patterns are increasing.

Goal: Investigate weather patterns that that could potentially lead to climate change (specifically looking at European weather stations)

Approach: Use **CNNs** and **GAN**s to analyze spatial data to find trends across these regions

Data: historical weather data from various weather stations in Europe over the last 25-50 years



CNN model Confusion Matrix showing 4 classes:
cloudy, rain, sunshine and sunrise
91% Accuracy

Incorrect prediction:
Class cloudy, predicted shine

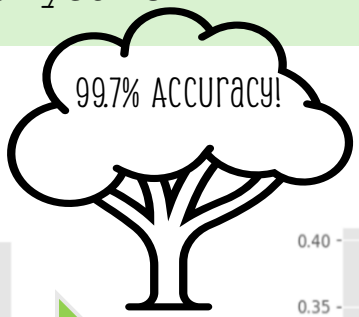
Thought Experiment 3:

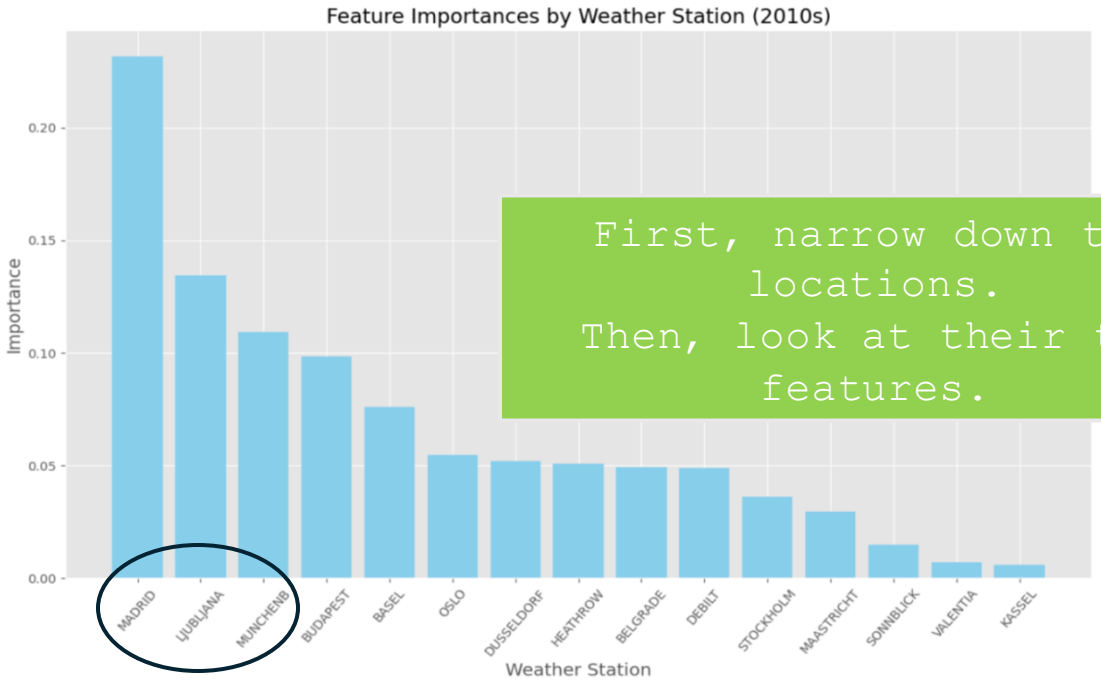
_Determine the safest places for people to live in the future

Goal: Investigate weather patterns that that could potentially lead to climate change (specifically looking at European weather stations)

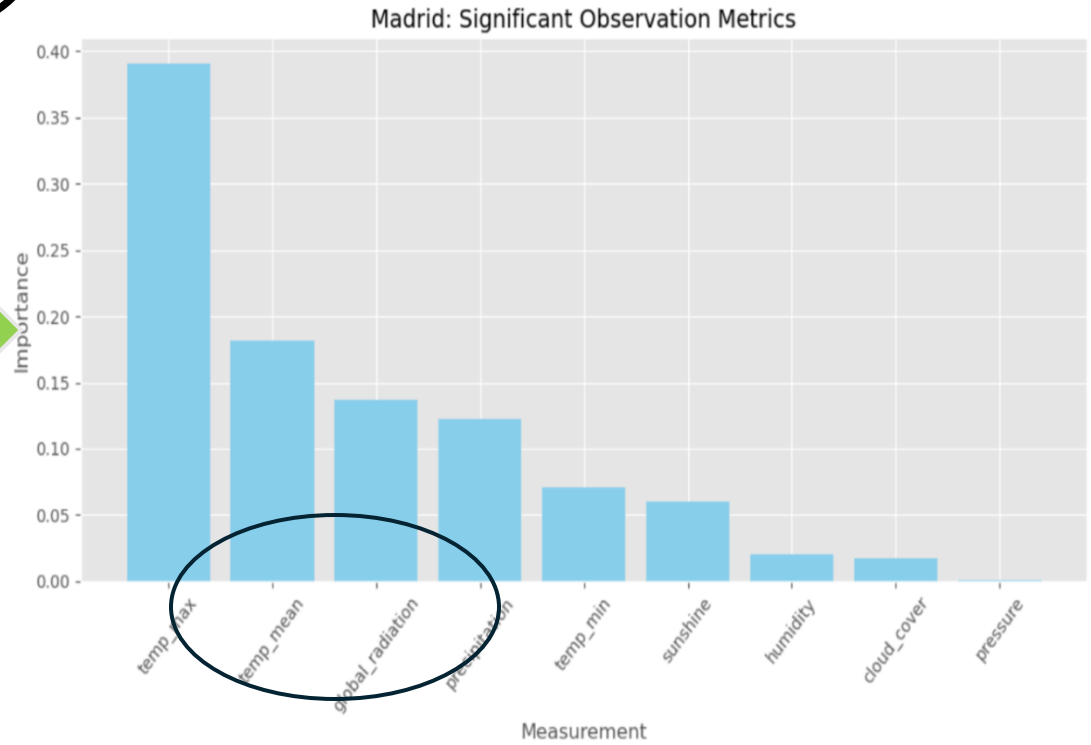
Approach: Use **Random Forest / Decision Trees (for 2010s)** to use predicted weather conditions to determine regional safety. Use the identified features within top locations to gauge future weather predictions.

Data: historical weather data from various weather stations in Europe over the last 25-50 years





First, narrow down top locations.
Then, look at their top features.



Recommendations/Top Method:

Random Forest: Highest Accuracy of 99.7% and most useful for breakdown of data

Data is narrowed down to determine different interpretations of pleasant vs. unpleasant conditions using the top weather features (rather than just temperature). It provided the most insightful and meaningful data specifics to determine detailed weather conditions and regions.

Combining these specific insights with the use of CNNs and GANs, we can narrow down particular trends in rising temperatures that could be aligning with our concern of global warming.

METHOD

1. Start with Random Forest to narrow down the details of regions and features to investigate.
2. Use single method Dendrogram to find extreme outliers (for any valuable insights).
3. Use CNNs & GANs to look for trends on a larger scale to compare insights and interpret predictions. These models can be further narrowed down by using the data from the top features from random forest and plugging those into the confusion matrix and testing predictions.

Thank you for your time and consideration.
Questions?

Nancy Kolaski
Nancykolaski@gmail.com
Github.com/Nancy-Kolaski