Analyzing Weather Conditions & Climate Change **with ClimateWins**

Final Report by Kirsten Currie
Career Foundry | 3.6.25

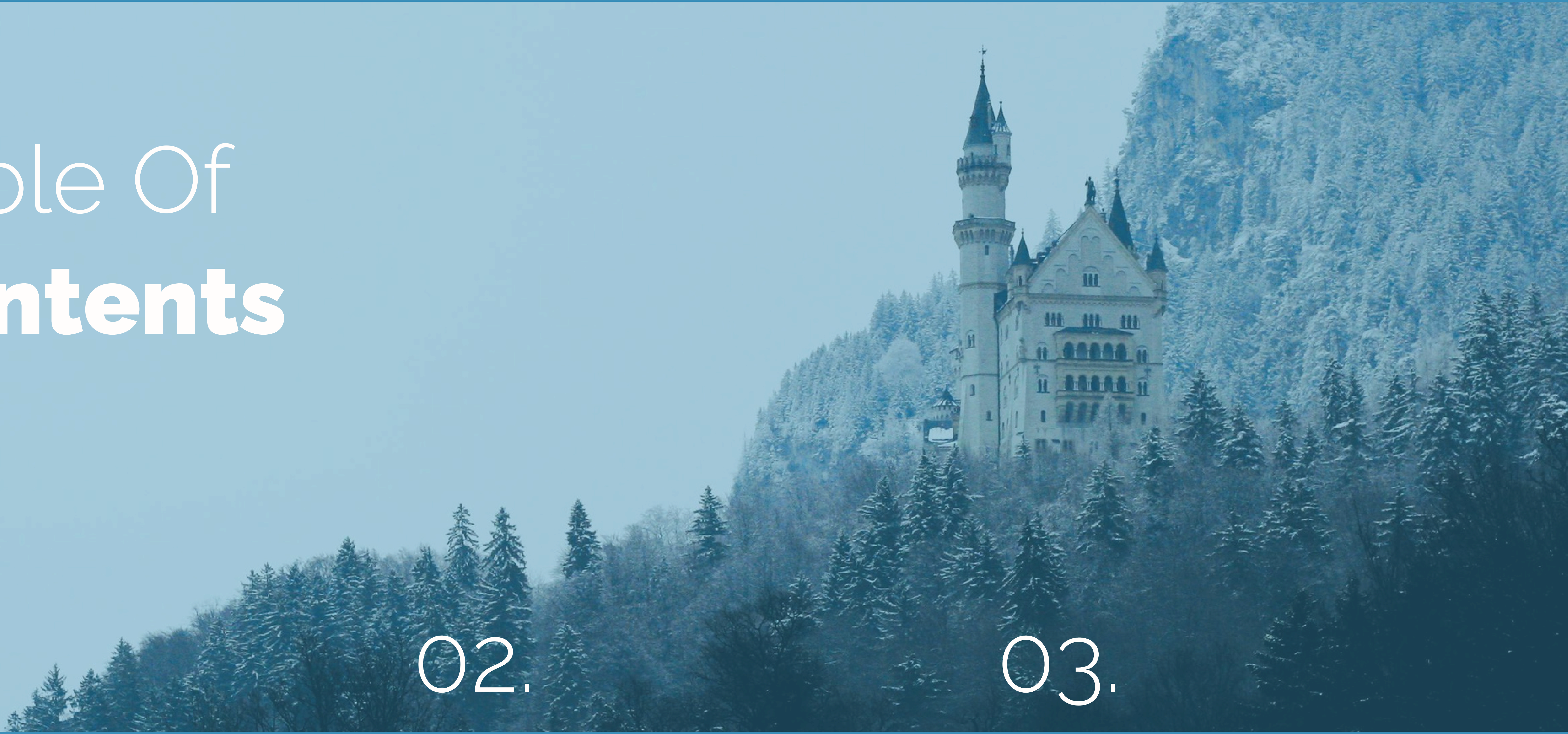


Table Of **Contents**

CLIMATEWINS

CLIMATEWINS

Introduction

EVALUATING CLIMATEWINS GOALS BEFORE SELECTING APPROACH

ClimateWins
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.

CLIMATEWINS

The **Methods**

MACHINE LEARNING APPROACH OVERVIEWS

Machine Learning Process Approaches

CLIMATEWINS

HIERARCHICAL CLUSTERING

Use machine learning to cluster weather data features.

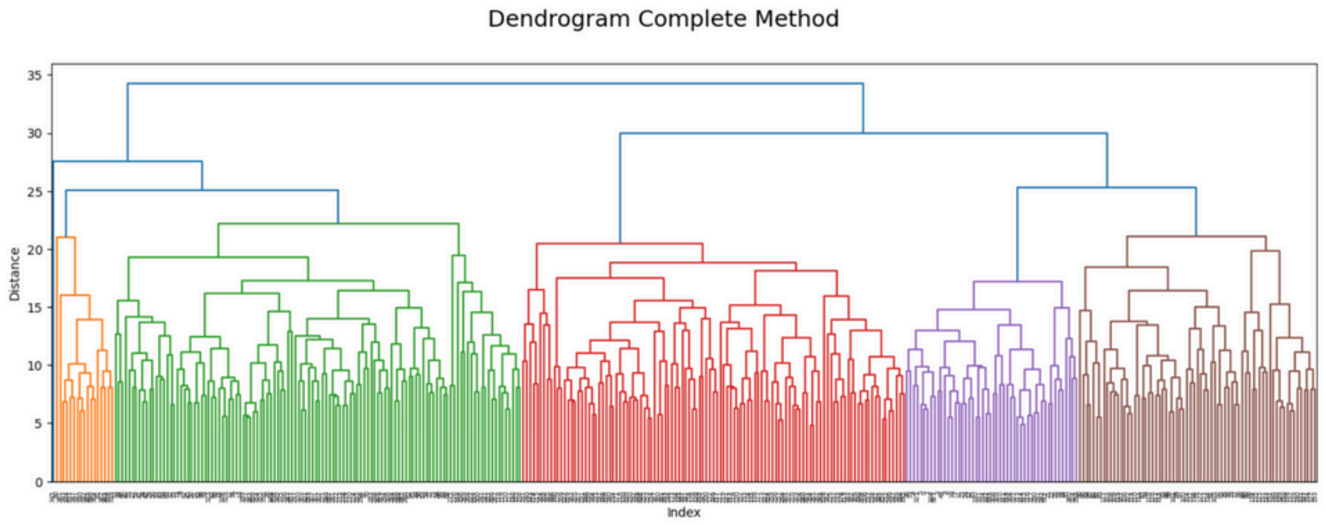
RECURRENT NEURAL NETWORKS

Predict weather patterns over time using RNNs.

RANDOM FORESTS

Determine important weather data features using Random Forests.

Hierarchical Clustering



Example of how hierarchical clustering is used to interpret patterns and potential groupings within weather pattern data as showcased by this dendrogram plot.

CLIMATEWINS

Method: Use hierarchical clustering to identify and classify weather patterns across Europe.

Plan of Execution:

- Preprocess weather station data by relevant location, time, & features
- Apply hierarchical clustering algorithms to group similar weather patterns.
- Analyze the resulting clusters to identify patterns outside the regional norm.
- Track the frequency of unusual patterns over time to determine if they are increasing.
- Use the clustered data to train a predictive model for future weather conditions.

Explanation: Hierarchical clustering can effectively group similar weather patterns, allowing for the identification of outliers or unusual patterns. This method can reveal how weather patterns have changed over time and potentially predict future trends.

Possible Limitations: This model may struggle with multiple data features as all variables must be treated equally, so it might be best to deal with only a few feature at a time.

Recurrent Neural Networks

```
n_hidden = 64

timesteps = len(X_train[0])
input_dim = len(X_train[0][0])
n_classes = len(y_train[0])

model = Sequential([
    Input(shape=(timesteps, input_dim)),
    LSTM(n_hidden),
    Dropout(0.5),
    Dense(n_classes, activation='tanh')
])
```

An example of changing hyperparameters within an RNN model (epochs, batch size, number of hidden layers, etc) in order to produce the ideal training scenario that will decrease model prediction loss and increase accuracy.

CLIMATEWINS

Method: Use Recurrent Neural Networks (RNNs) to analyze and predict weather patterns over time.

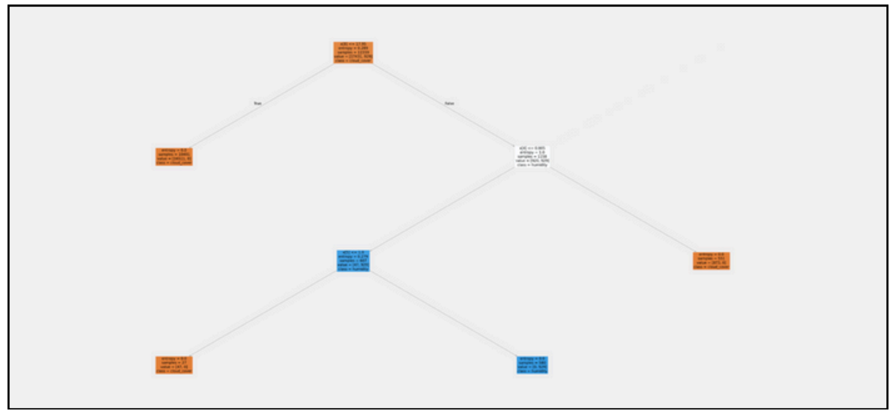
Plan of Execution:

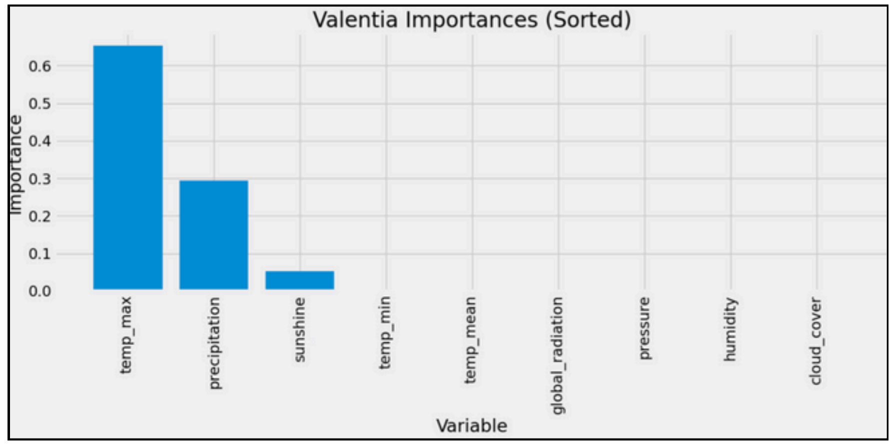
1. Design and train RNN models to capture temporal trends in weather patterns based on past 50 years at each weather station.
2. Use the trained models to identify anomalies in historical data.
3. Analyze the frequency and severity of identified anomalies over time.
4. Generate long-term weather predictions for the next 25-50 years.
5. Determine safety levels of European zones based on frequencies of predicted weather patterns.

Explanation: RNNs are better for time series data than other neural networks (e.g. CNNs) in that they can capture long-term dependencies in weather patterns and have the advantage of being able to consider both previous and present layers within the data (backpropogation). They can be effective in identifying unusual patterns and making future predictions based on historical trends.

Possible Limitations: Could become more challenging to train after a certain number of years of data in what's known as the vanishing gradient problem.

Random Forests





CLIMATEWINS

Evaluating important weather station features from a fine-tuned Random Forest model (minimum temperature, precipitation, and sunshine important in Valentia).

Method: Utilize Random Forests for weather pattern identification, trend analysis, and future prediction.

Plan of Execution:

1. Preprocess weather station data by a determined set of weather patterns across all stations.
2. Use the model to identify top important weather patterns to track against each station.
3. Compare feature importances and prediction accuracies across these models to detect changing patterns over time:
    a. Define safety criteria (e.g., frequency of extreme events, temperature ranges).
    b. Identify the most crucial factors in weather pattern changes and safety assessments.

Explanation: Random Forest is a learning method that constructs multiple decision trees and merges them to get a more accurate and stable prediction. They are particularly useful in providing feature importance rankings, helping identify key factors in weather pattern changes and can capture non-linear relationships in the data, which is crucial for complex weather systems.

Possible Limitations: Achieving optimal performance requires careful tuning of hyperparameters, which can be time-consuming and requires expertise.

CLIMATEWINS

Additional **Data**

- Topographical data to account for geographical influences on weather patterns.
- Solar activity data for long-term climate predictions.
- Ocean current data, particularly for coastal regions.
- Air pollution data to account for human influences on local weather.
- Population density and urban development data to factor in heat island effects.

Summary

CONCLUDE BEST MACHINE LEARNING METHOD
FOR CLIMATEWINS

CLIMATEWINS

Which Method is the **Best**?

Because ClimateWins is a nonprofit and is seeking out ways to conduct this experiment design in the most cost effective and time efficient way, it needs to narrow down on the most suitable method for answering questions related to weather prediction, regional safety, etc.

However, the data represents a variety of factors that could be evaluated--each with features that more or less represent each station. Sometimes not just one model would be suitable for capturing all this information.

Conducting machine learning research **using Random Forests** can be used to determine **which weather data features** (e.g. mean and max temperature, precipitation levels, etc) **will be the most crucial to analyze.** Following this approach, a **Reccurrent Neural Network** (RNN) could **evaluate the time series** of different stations focusing on the top features pre-determine from the Random Forest so as to predict the safest regions and what weather might look like in the next 25 - 50 years.

Thank You

CLIMATEWINS

QUESTIONS?

Please contact me at
kirstenlynncurrie@gmail.com

GITHUB

Feel free to explore the Jupyter Notebook behind this project on my **GitHub** (github.com/kirstencurrie).

Slides Carnival

This presentation template is free for everyone to use thanks to the following:

SlidesCarnival for the presentation template
Pexels for the photos

Happy designing!