ClimateWins

Weather Conditions and Climate Change

Stephanie Ugwuanya - February 2025



Introduction

ClimateWins aims to leverage machine learning to predict extreme weather events across Europe. This phase focuses on **unsupervised learning, deep learning, and composite models** to analyse historical weather data and uncover meaningful insights. The ultimate goal is to identify the most effective tools and algorithms for developing accurate predictive models that assess the impacts of climate change.

Research Objectives:

- Identify new patterns in weather changes over the past **60 years**.
- Detect weather patterns that deviate from regional norms in Europe.
- Assess whether unusual weather patterns are becoming more frequent.
- Forecast potential weather conditions over the next **25 to 50 years** based on current trends.
- Determine the safest locations for people to live in Europe within the next **25 to 50 years**.

Machine Learning Options

Random Forest

A **Random Forest** model can predict weather for **ClimateWins** by analysing historical weather data and identifying patterns. It works by:

1. **Training on Past Data** – The model learns from past weather conditions, such as temperature, humidity, and wind speed.

2. **Making Predictions** – It uses multiple decision trees to predict future weather trends based on current inputs.

3. **Improving Accuracy** – By averaging the predictions of many trees, it reduces errors and provides more reliable forecasts.

CNNs & RNNs

CNNs analyse spatial patterns in weather data (like satellite images).

RNNs process time-series data to predict future weather based on past trends. Together, they improve weather prediction accuracy by handling both spatial and temporal aspects.

GANs (Generative Adversarial Networks)

Generative Adversarial Networks (GANs) can be used for weather prediction by generating realistic simulations of weather patterns based on historical data

1.GANs are trained on past weather data (e.g., temperature, rainfall, wind speed). This data acts as the "real" information the GAN learns from.
2.Two networks: Generator: creates synthetic weather data (predictions) from random noise.
3.Discriminator: evaluates the generated weather data against the real historical data, trying to distinguish between the two.

Additional Datasets Required

To achieve ClimateWins' goals, additional data beyond historical weather records could include:

1. **Real-time Weather Data:** Continuous updates on current conditions (temperature, humidity, pressure, wind speed, etc.) to improve short-term predictions.

2. **Social and Economic Data:** Insights into population density, land use, and infrastructure to assess the impact of weather events on communities and plan for resilience.

3. **Geospatial Data:** Information about the terrain, urban areas, and natural landscapes, which can affect local weather patterns.

Thought Experiment 1

Hypothesis: *A CNN Model can be used as a reliable model for predicting weather conditions by analysis satellite images and by analysing patterns in historical data.*

Approach

- Trained a Convolutional Neural Network (CNN) for classifying weather-related images into different categories - 'Cloudy', 'Rain', 'Shine', 'Sunrise'
- Bayesian optimisation was employed to fine-tune hyperparameters such as the number of neurons, batch size, and learning rate, resulting in improved accuracy."

Conclusion
Hypothesis accepted

CNN Model was able to accurately classify weather images, this could potential be used for classifying satellite images. The CNN model can also be used to find patterns in weather data meaning it could be a powerful tool for making future predictions.

CNNs can also be used to identify patterns in data, and in this case, it worked reasonably well. However, CNNs are optimised for visual data, so an RNN (Recurrent Neural Network), which is better suited for sequential or time-series data, may have been a more appropriate choice.

Results -Thought Experiment 1

The results of optimising the parameters of the CNN Model using the Bayesian Optimiser for identifying patterns in weather data.

Stats Before Optimisation

CNN Model Accuracy: 64.1%

- Stagnant Performance: The model's loss stays at 24.3251, indicating it is not learning or improving during training.
- 14/15 weather stations recognised.

Stats After Optimisation

CNN Model Accuracy: 84.47%

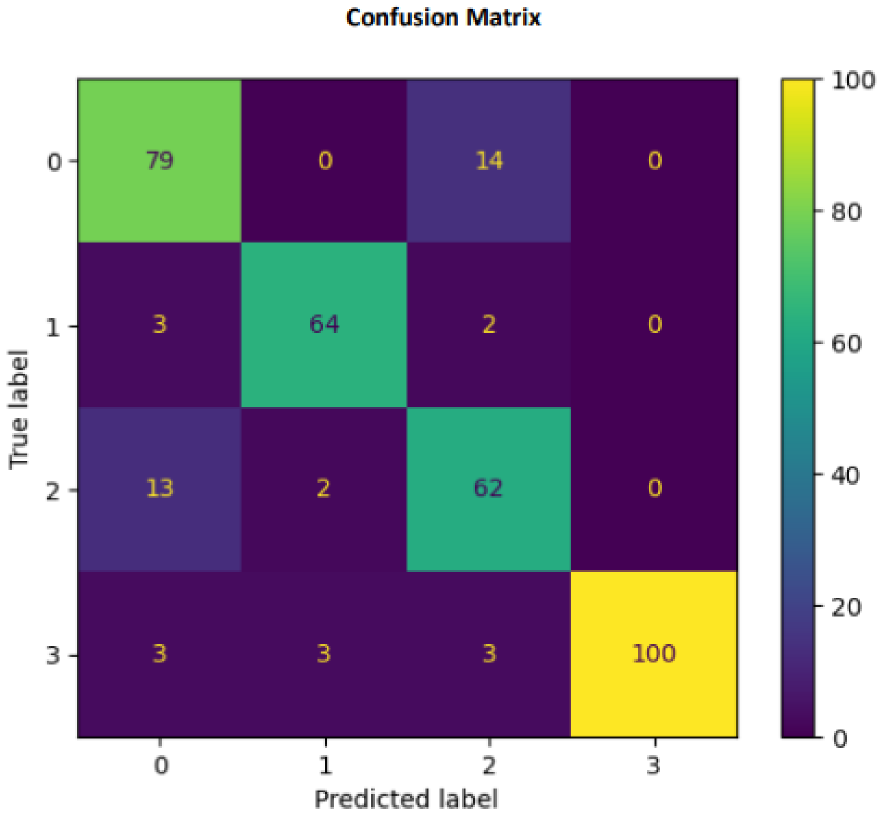
- The model starts with a high loss of 1.6153 at Epoch 1, but the loss decreases over time, showing improvement in accuracy.
- By Epoch 32, the loss reduces to 0.6677.
- 14/15 weather stations recognised.

Optimised Confusion Matrix

Pred True	BASEL	BELGRADE	BUDAPEST	DEBILT	HEATHROW	LJUBLJANA	MADRID
BASEL	3221	328	32	15	14	5	67
BELGRADE	212	817	19	3	3	1	37
BUDAPEST	45	66	68	5	4	1	25
DEBILT	14	18	11	23	8	1	7
DUSSELDORF	7	6	6	2	3	1	4
HEATHROW	22	6	1	2	15	1	35
KASSEL	6	2	2	0	1	0	0
LJUBLJANA	19	8	4	0	0	3	27
MAASTRICHT	9	0	0	0	0	0	0
MADRID	112	48	9	0	2	1	286
MUNCHENB	8	0	0	0	0	0	0
OSLO	3	1	0	0	0	0	1
STOCKHOLM	1	2	1	0	0	0	0
VALENTIA	1	0	0	0	0	0	0

Utilising Bayesian optimization to fine-tune the parameters of the CNN model resulted in a remarkable 20.37% increase in accuracy. This significant improvement highlights the potential of the CNN model as a powerful tool for detecting patterns in weather data, thereby enhancing weather prediction capabilities.

Results -Thought Experiment 1



The CNN model successfully identified 4 out of 6 generated weather images (these images can be seen on the left), indicating that it is effectively trained on the dataset and can be used to identify new satellite images in the future.

Training Accuracy: 96.5% – The model performs very well on the training data, correctly classifying most images illustrated in the confusion matrix above.

This performance suggests that the CNN model is capable of accurately classifying weather patterns from satellite images and can be relied upon for future weather predictions.

Thought Experiment 2

Hypothesis: *A Random Forest model can be used for weather predictions by analysing historical weather data, such as temperature, humidity, and pressure, to identify patterns and accurately forecast future weather conditions.*

Approach

- Developed **a Random Forest** model to identify key weather features and significant weather stations across Europe. Key weather features included: **temperature max, precipitation and sunrise.**
- Applied **Random Search** to optimise hyperparameters, improving model performance.

Conclusion
Hypothesis accepted

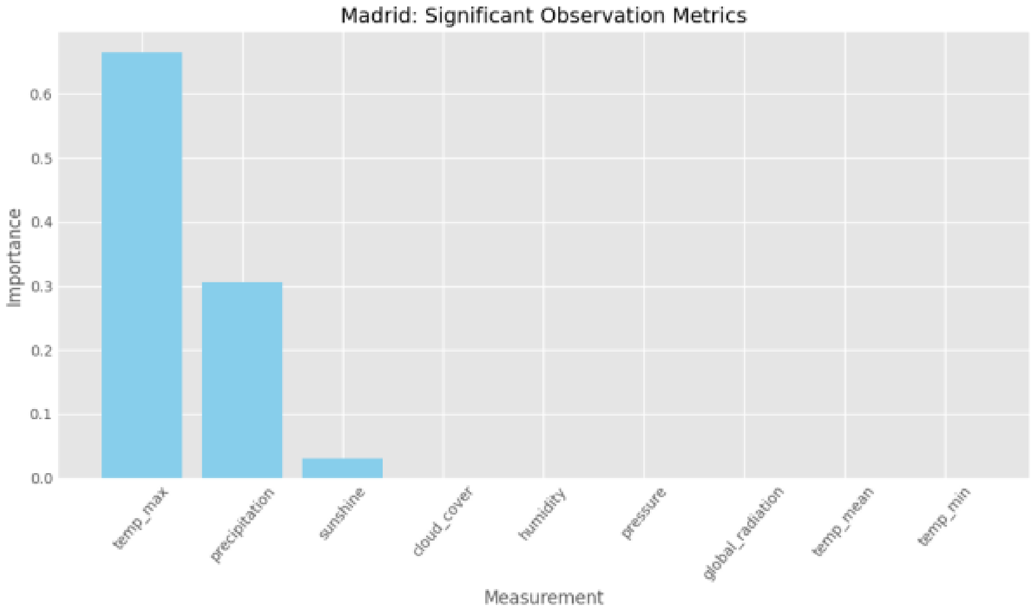
Optimising the Random Forest model led to a significant boost in accuracy, enhancing overall performance to 98% and achieving perfect classification for Madrid. The model identified key weather stations, such as Budapest, Delit, and Stockholm, as the most influential, highlighting their critical role in driving accurate weather predictions. This, coupled with the identification of key features, underscores the model's effectiveness in leveraging historical data to deliver precise and reliable weather forecasts.

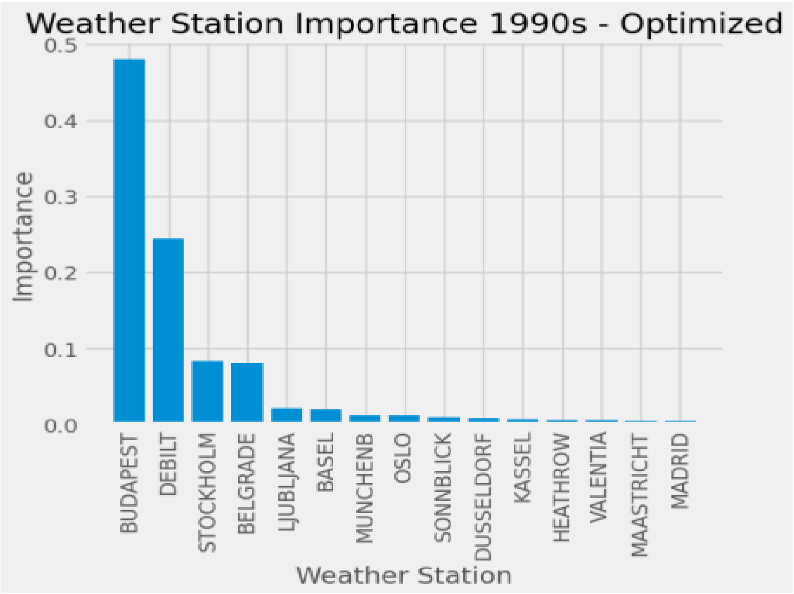
Results -Thought Experiment 2

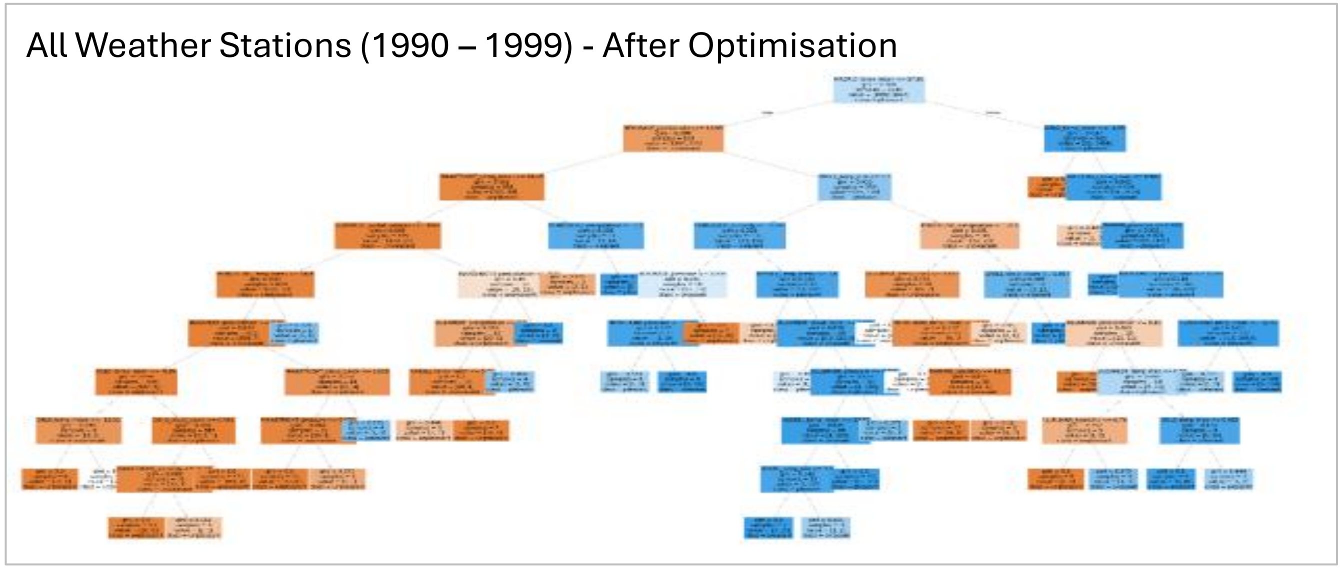
- Before optimisation, the model achieved 90.14% accuracy across all weather stations and 99% accuracy for Madrid (1990–1999).

- After optimisation, accuracy improved to 98% for all weather stations and 100% for Madrid, demonstrating significant performance enhancement.

- **Feature Importance:** After optimisation, the Budapest, Delit, and Stockholm weather stations were identified as the most influential in the model's predictions.
- **Significant Observations:** Temperature max, precipitation and sunshine were identified as most important when making weather predictions in Madrid.







Thought Experiment 3

Hypothesis: *Generative Adversarial Networks (GANs) are essential for advancing weather prediction research, with the potential to significantly enhance forecasting accuracy and support climate modelling.*

Approach

1. Generator: Tries to create fake but realistic-looking data (e.g., weather maps, images).
2. Discriminator: Tries to tell the difference between real data and the fake data created by the generator.

The generator keeps improving until the discriminator can't tell the difference anymore, meaning the GAN has learned to create highly realistic data.

Conclusion

Hypothesis accepted

GANs play a crucial role in weather prediction research. They can generate realistic weather images to enhance training data, improve the resolution of weather maps for better forecasts, and simulate future weather scenarios based on historical patterns. These capabilities highlight GANs' potential to improve the accuracy of weather predictions and aid in climate modelling.

Using GANS for Weather Predictions

Example 1

GANs can be used to create fake weather images that look real, like pictures from satellites or radar maps.

These fake images can help improve machine learning models by providing more data to train them.

Example: If we don't have enough weather data for certain days, we can use a GAN to generate new, realistic weather images from past data. This helps the model learn better and make more accurate predictions

Example 2

Weather forecasts, such as satellite images or weather maps, are often produced at a low resolution to reduce computation time and data size.

This means the images may appear blurry or lack the details needed to make more accurate predictions. A **GAN can help with this by improving the resolution of these images.**

The GAN does this by taking a blurry or low-resolution weather map and learning how to create a clearer, higher-resolution version that looks more realistic and detailed.

Example 3

GANs can simulate future weather conditions based on past data to predict possible weather scenarios.

By learning patterns in temperature, pressure, and other atmospheric data, a GAN can generate multiple potential outcomes for the coming days or weeks.

Example: Train a GAN on historical temperature and pressure data to predict possible weather conditions, like heatwaves or rainstorms, in the upcoming weeks.

Conclusion

- **Random Forest** is a powerful tool that can be used to identify feature importance and significant observation that can help identify patterns and abnormalities in weather data.
- **CNN models** can be used to interpret satellite data and successfully categorise weather images resulting in improved weather forecasting, more accurate identification of weather patterns, and better decision-making for climate-related scenarios.
- **GANs** can create realistic weather images to enrich training datasets, enhance the resolution of weather maps for more precise forecasts, and simulate future weather conditions by analysing historical patterns.

Next Steps/Recommendations

- **Refine Model Accuracy:** Continue optimising the Random Forest model by fine-tuning hyperparameters to further improve prediction accuracy.
- **Explore Additional Data**: Incorporate more diverse weather data sources, such as satellite imagery or real-time weather feeds, to enhance model robustness.
- **Valuate Other Algorithms:** Experiment with other machine learning algorithms e.g., RNNs which are better for which are better suited for analysing time-series data and capturing temporal patterns in weather data. This could help improve predictions for weather events that depend on past conditions and their sequential nature.
- **Focus on developing GANs** for long-term scenario simulation to enhance predictive capabilities and improve future weather modelling.

Thank you

Stephanie Ugwuanya

LinkedIn GitHub