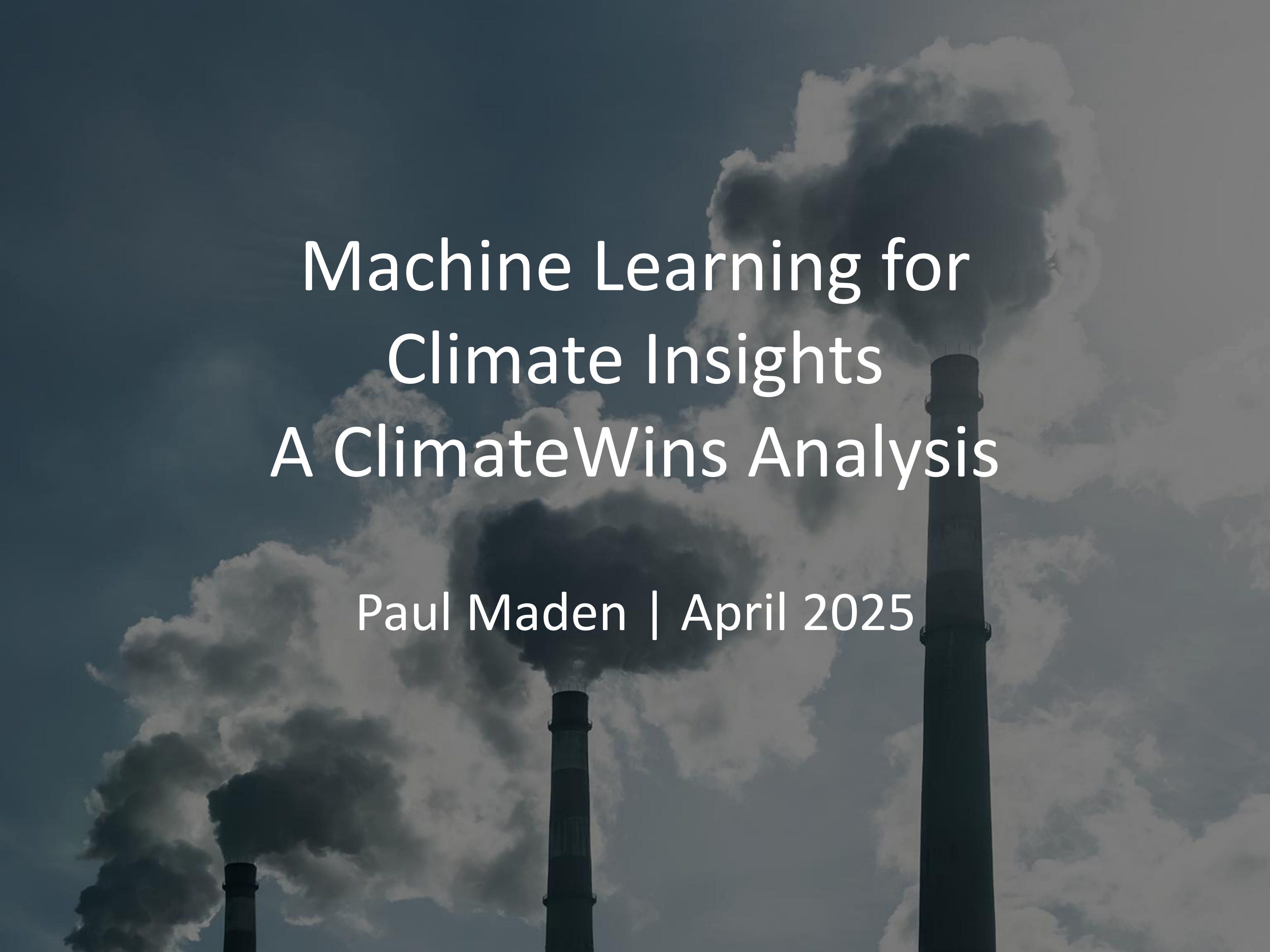

# Machine Learning for Climate Insights
# A ClimateWins Analysis

Paul Maden | April 2025

# Introduction & Background

## Who is ClimateWins?

ClimateWins is a European nonprofit aiming to understand and predict the effects of climate change particularly extreme weather events in Europe.

## Why Machine Learning?

With limited resources and increasing data complexity, ML offers scalable ways to defect patterns and forecast trends critical for preparedness and risk reduction.

## Project Goal

Use supervised machine learning models to explore historical weather data, test predicted algorithms and recommend an optimal model for forecasting climate related conditions.

# Objective & Hypotheses

Objective: Use Machine Learning to predict and understand climate related patterns focusing on extreme weather event and long-term climate trends across Europe

Hypotheses:

1. ML can accurately predict extreme weather events (e.g. storms & heat waves) up to 48 hours in advance.

2. Prediction accuracy will vary by location and climate type, reflecting regional weather complexities.

3. ML can detect long term trends in temperature and precipitation patterns that correlate with effects in climate change.

# Data Sources, Bias & Accuracy

Data Source: European Climate Assessment & Dataset (ECAD), containing historical data across 18 European weather stations (spanning 1880s – 2022). Variables Tracked: Daily Temperature, Wind Speed, Snowfall, Global Radiation, Precipitation, Humidity, etc.

Biases: Uneven station distribution (some regions may be over or underrepresented). There maybe missing values in older record and changes in measurement standards or equipment over time.

Overall accuracy: The long record span provides a strong historical insight. However, newer data is generally more consistent and complete. Also, preprocessing (e.g. normalization and interpolation) is used to improve readability.

# Ethical Considerations in Machine Learning

<u>Human Oversight Needed:</u> Machine Learning must be reviewed by climate scientists or policymakers to ensure responsible applications.

<u>False Predictions</u>: False positive (e.g. predicting storms that don’t happen v’s false negative (failing to predict real dangers) could influence policy preparedness and public trust.

<u>Ethics in Messaging:</u> Predictions can influence migration, agriculture, insurance, and more messaging must be responsible and transparent.

# Feature Optimization

Used feature *scaling and pruning* for improved accuracy.

ANN configuration tuning: Layers, nodes, tolerance, iterations.

Gradient descent to minimize loss during training.

# Algorithm 1 – K-Nearest Neighbors (KNN)

Accuracy: 83–89%

Simple, non-parametric, effective for climate data.

Sensitive to distance metrics and number of neighbors.

# Algorithm 2 – Decision Tree

Accuracy: 46%

Overfitting likely due to complexity and depth.

Requires pruning for generalization and interpretability.

# Algorithm 3 – Artificial Neural Network (ANN)

Accuracy:

51% training

49.8% testing.

Deep architecture with (50,50,60) performs best.

Requires tuning for tolerance and iteration limits.

# Model Comparison & Insights

**KNN**:

*Best* performer on current data.

(Simple & effective with clean, structured data. Slow for large datasets).

**ANN**:

Better for complex patterns, needs more tuning.

(Great for nonlinear, complex trends).

**Decision Tree:**

Prone to overfitting, lowest accuracy.

(Easy to interpret, but over fits without pruning).

Below is the confusion matrix summarizing the model's classification performance across different weather stations:

| Weather Station | True Negatives (TN) | True Positives (TP) | False Positive (FP) | False Negative (FN) | Accuracy Rate |
|---|---|---|---|---|---|
| **Basel** | 3907 | 935 | 431 | 465 | 85% |
| **Belgrade** | 3239 | 1303 | 537 | 439 | 84% |
| **Budapest** | 3416 | 1410 | 484 | 408 | 85% |
| **Debilt** | 4346 | 732 | 291 | 389 | 88% |
| **Dusseldorf** | 4167 | 800 | 380 | 800 | 87% |
| **Heathrow** | 4160 | 754 | 332 | 424 | 85% |
| **Kassel** | 4635 | 607 | 252 | 316 | 90% |
| **Ljubljana** | 3726 | 1132 | 469 | 411 | 86% |
| **Maastricht** | 4249 | 820 | 313 | 356 | 88% |
| **Madrid** | 2735 | 2250 | 433 | 322 | 87% |
| **Munchenb** | 4222 | 768 | 324 | 400 | 88% |
| **Oslo** | 4623 | 507 | 256 | 352 | 90% |
| **Sonnblick** | 5738 | 0 | 0 | 0 | 100% |
| **Stockholm** | 4450 | 588 | 316 | 308 | 89% |
| **Valentia** | 5391 | 309 | 71 | 167 | 96% |
| **Total** | | | | | **88.53%** |

The table below highlights the false negative rate (FNR) for each station, indicating the proportion of actual positive cases that were misclassified as negatives:

| Weather Station | True Positives (TP) | False Negatives (FN) | False Negative Rate (FNR) |
|---|---|---|---|
| Basel | 935 | 465 | 33.21% |
| Belgrade | 1303 | 439 | 25.20% |
| Budapest | 1410 | 408 | 22.44% |
| Debilt | 732 | 389 | 34.70% |
| Dusseldorf | 800 | 800 | 50.00% |
| Heathrow | 754 | 424 | 35.99% |
| Kassel | 607 | 316 | 34.24% |
| Ljubljana | 1132 | 411 | 26.64% |
| Maastricht | 820 | 356 | 30.27% |
| Madrid | 2250 | 322 | 12.52% |
| Munchenb | 768 | 400 | 34.25% |
| Oslo | 507 | 352 | 40.98% |
| Sonnblick | 0 | 0 | nan% |
| Stockholm | 588 | 308 | 34.38% |
| Valentia | 309 | 167 | 35.08% |

# Summary

KNN RECOMMENDED FOR IMMEDIATE USE (HIGHEST ACCURACY).

TUNE ANN FURTHER FOR COMPLEX, NON-LINEAR PATTERNS.

STANDARDIZE DATA, RETRAIN WITH NEW SAMPLES FOR BETTER GENERALIZATION.

**Recommeded model**: *K-Nearest Neighbours (KNN)* is recommended for initial deployment, due to its superior accuracy and performance on current weather data. Use it for short-term predictions (e.g., next 48–72 hours).

# Next Steps

DEPLOY **KNN** WITH CROSS VALIDATION ACROSS MULTIPLE EUROPEAN REGIONS.

TRAIN **ANN** WITH TUNED PARAMETERS ON LONG-TERM DATASETS.

INTEGRATE PREDICTION OUTPUT INTO AN INTERACTIVE **DASHBOARD** FOR CLIMATEWINS STAKEHOLDERS.

# Thank You!

Questions?

Contact: wallmaden@hotmail.com

Presented by: Paul Maden

GitHub-Repo: Supervised-ML-Weather-Models