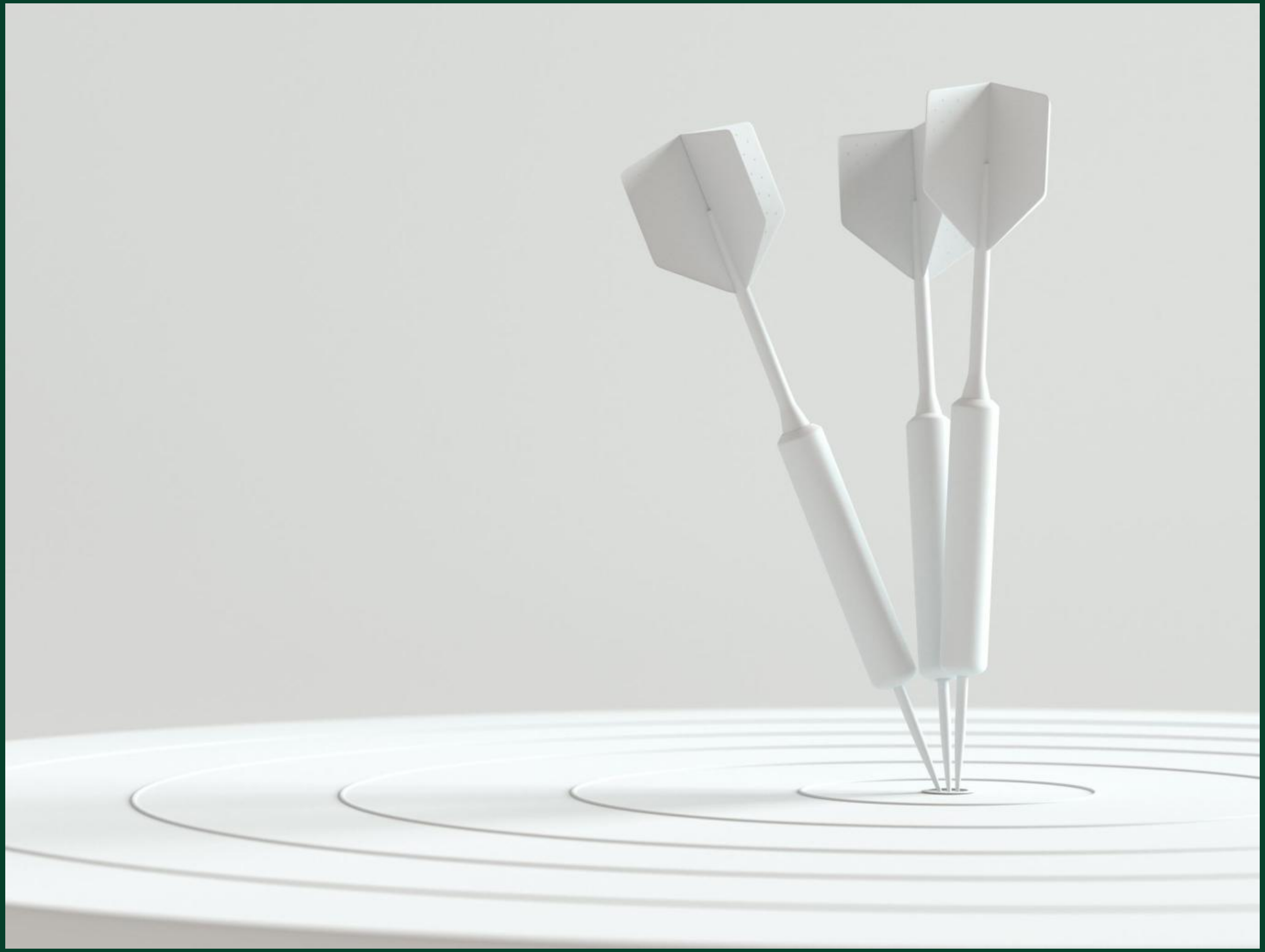
# MACHINE LEARNING FOR WEATHER PREDICTION & CLIMATE INSIGHTS

Technical Assessment for ClimateWins

Jacques van Rensburg August 2025





# OBJECTIVE & HYPOTHESES

Objective: Use machine learning to predict weather patterns and assess climate change effects in Europe.

Hypotheses:

1. Supervised machine learning can predict whether a day will be “pleasant” or “unpleasant” based on historical weather data.

2. Artificial Neural Networks (ANN) will outperform Decision Trees and K-Nearest Neighbors (KNN) for imbalanced weather datasets.

3. Gradient descent optimization can identify seasonal trends and critical features influencing temperature and precipitation.

# DATA SOURCE & BIASES

This Photo by Unknown Author is licensed under CC BY

Data Source:

- European Climate Assessment & Data Set (18 weather stations, late 1800s – 2022)
- Variables: Temperature, wind speed, snow, global radiation, etc.

Potential Biases:

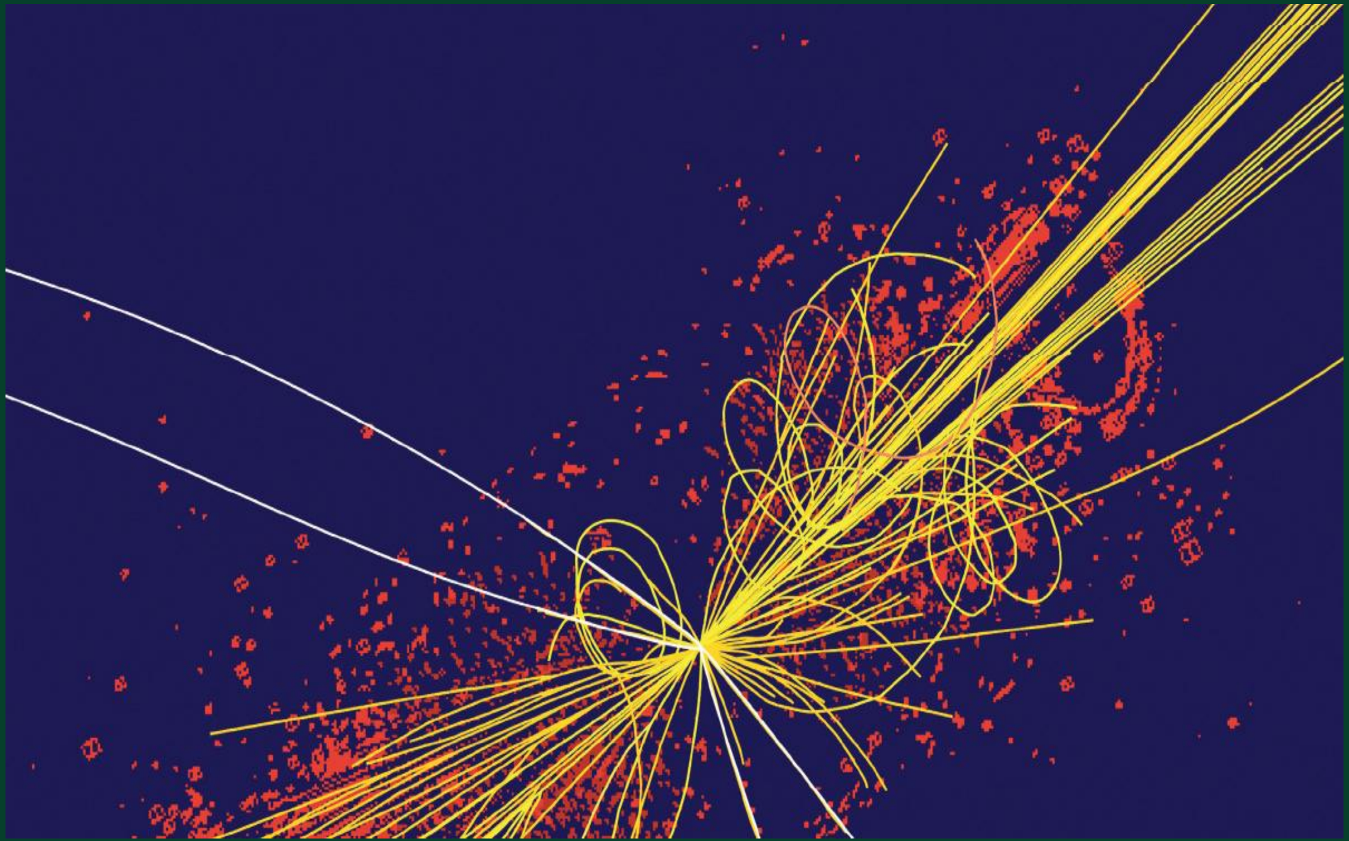
- Temporal bias: Older records may not reflect current climate conditions accurately.
- Regional bias: Larger regions overrepresented; smaller regions underrepresented.
- Labeling bias: "Pleasant" days may have subjective definitions.
- Missing/uneven data: Some stations have incomplete data for rare events.

# DATA SOURCE & BIASES



Accuracy Consideration:

- Historical recordings are mostly reliable, but extreme weather events may be underreported.



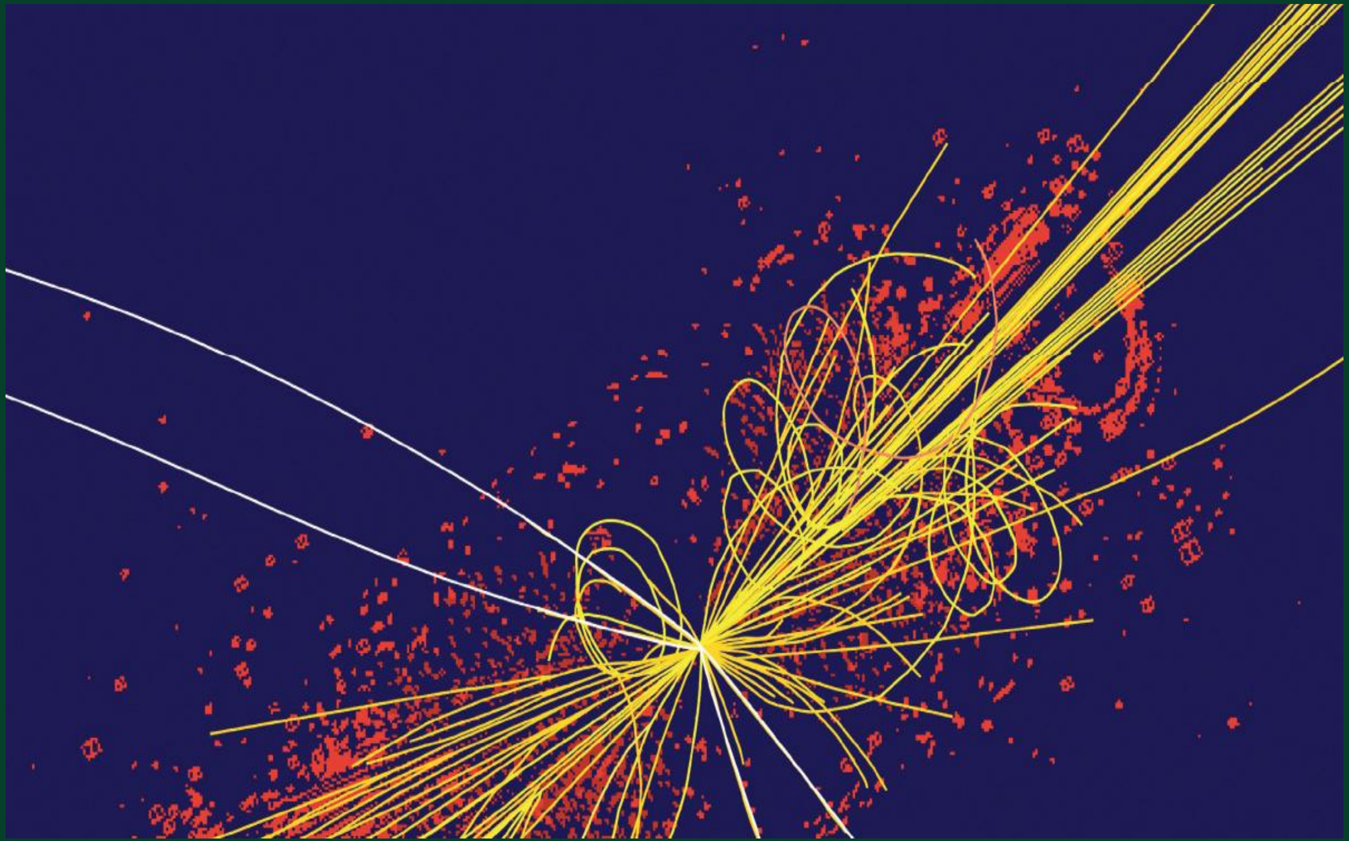


# OPTIMIZATION & FEATURE ENGINEERING

Goal: Prepare and optimize data for supervised learning.

Steps:

- Normalized and scaled numerical features.
- Applied gradient descent to identify patterns in seasonal temperature changes.
- Calculated loss functions to minimize prediction error.
- Evaluated feature importance for variables like temperature, wind, and radiation.





# OPTIMIZATION & FEATURE ENGINEERING

Outcome:

- Gradient descent successfully revealed seasonal trends.
- Key features identified: Temperature, wind speed, and global radiation.


# SUPERVISED LEARNING APPROACHES

Algorithms tested:

1, Decision Tree (DT)

- Simple, interpretable, but can overfit on complex datasets.
- Accuracy: Moderate; struggles with rare “pleasant” days.



# SUPERVISED LEARNING APPROACHES

Algorithms tested:

2, K-Nearest Neighbors

- Predicts by comparing distance to neighbors.
- Accuracy: High on balanced datasets (≈89%).
- Sensitive to feature scaling and imbalance.


# SUPERVISED LEARNING APPROACHES

Algorithms tested:

3, Artificial Neural Network (ANN)

- Multi-layer “neurons” adjust weights through backpropagation.
- Handles non-linear, imbalanced data well.
- Accuracy: Moderate to high; most scalable for future larger datasets.



# SUPERVISED LEARNING APPROACHES

Evaluation:

- Confusion matrices used to assess precision, recall, and F1-score for “pleasant” vs. “unpleasant” days.
- ANN recommended for long-term predictive analysis due to robustness and scalability.

# SUMMARY OF FINDINGS



Hypotheses & Methods:

- ML can predict pleasant/unpleasant days with reasonable accuracy.
- ANN is the most effective supervised model for complex, imbalanced climate data.
- Gradient descent efficiently optimized features and highlighted seasonal trends.

# SUMMARY OF FINDINGS



Next Steps:

- Incorporate additional variables (humidity, extreme weather events).
- Test unsupervised learning to identify hidden climate patterns.
- Expand dataset to other regions and recent years.
- Refine model based on real-world validation.



# ETHICAL CONSIDERATIONS & BIAS MANAGEMENT

Ethical Concerns:

- ML predictions can influence policy and planning; decisions must be validated.
- Transparency: Stakeholders should understand model limitations.

# ETHICAL CONSIDERATIONS & BIAS MANAGEMENT



Bias Mitigation Strategies:

- Use multiple data sources.
- Avoid overgeneralization from small or imbalanced datasets.
- Post-process predictions with human review.



# THANK YOU

Any questions?