Strategic Proposal for Predictive Weather Analysis using Machine Learning

KACI ERWIN | JUNE 2024

Introduction

Objective

To develop and evaluate machine learning strategies for predicting weather variations to enhance ClimateWins' forecasting accuracy. ClimateWins' goals are:

- Finding new patterns in weather changes.
- Identifying unusual weather patterns.
- Determining increase in unusual weather patterns.
- Predicting future weather conditions.
- Identifying safe places to live in Europe.

Thought Experiments

This presentation will explore three thought experiments designed to harness machine learning technologies for predictive accuracy in weather forecasting.

Machine Learning Approaches

Algorithms Already Explored

Supervised

- Linear Regression
- K-Nearest Neighbors (KNN)
- Decision Tree

Unsupervised

- Principal Component Analysis (PCA)
- Hierarchical Clustering
- Convolutional Neural Networks (CNN)
- Random Forests

Methods To Explore

Algorithms

- Isolation Forests
- Support Vector Machines (SVM)
- Local Outlier Factor (LOF)
- Long Short-Term Memory (LSTM)
- Prophet
- Gated Recurrent Units (GRU)
- Gradient Boosting Machines (GBM)

Approaches

- Data Augmentation with Generative Adversarial Networks (GANs)
- Neural Architecture Search (NAS)
- Hybrid/Multi-Model Approach

Data

Current Data

- The data used by ClimateWins includes a wide range of weather information, such as:
  - **Temperature**
  - **Precipitation**
  - **Wind speed**
  - **Humidity**
- Data is collected from **18 weather stations** across Europe and spans multiple decades.
- An extra dataset indicating daily **weather pleasantness** was used for clustering analysis.

Additional Data

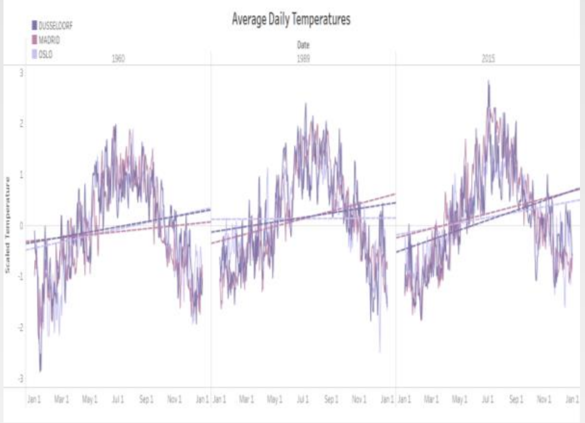
The models could benefit from additional information, such as:

- **real-time climate monitoring data**
- **socio-economic data**
- **geographic information systems (GIS) data**

These would enhance model robustness and predictive accuracy, especially in dynamic weather conditions and diverse regions.

Algorithms Explored – Supervised

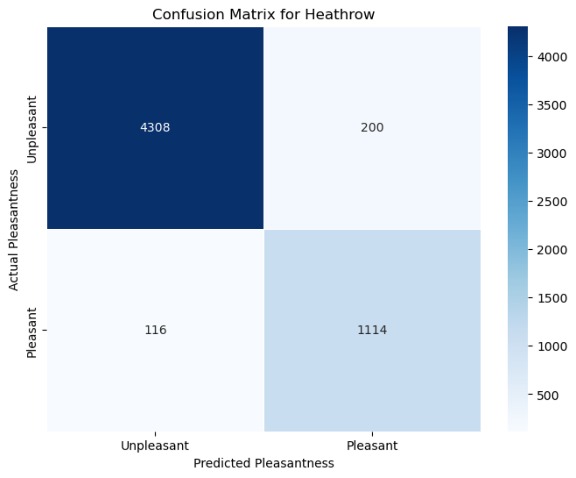
Linear Regression



Findings:

- Positive slopes indicate a consistent warming trend.
- Temperatures are rising annually across various regions in Europe.

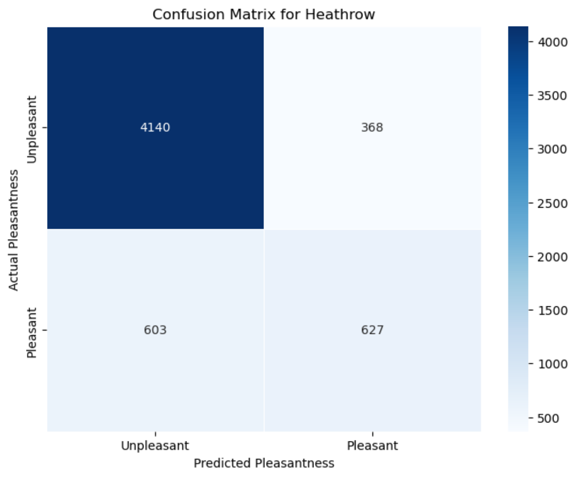
K-Nearest Neighbors



Performance:

- High accuracy (>90%) with single location data.
- Does not perform well with the entire dataset.

Decision Tree

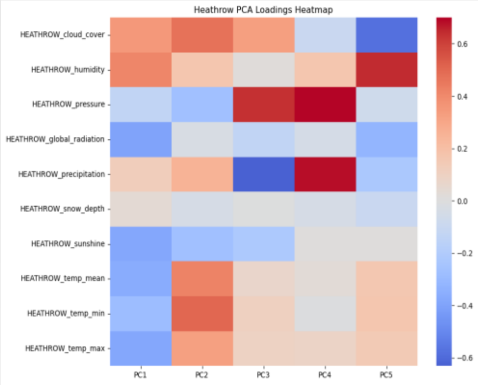


Performance:

- High accuracy (>80%) with single location data.
- Shows bias towards the dominant class, needs data tuning.

Algorithms Explored – Unsupervised

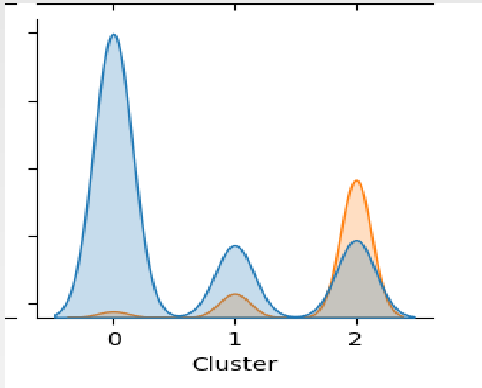
Principal Component Analysis



Findings:

- Determined the variables contributing most to data variance.

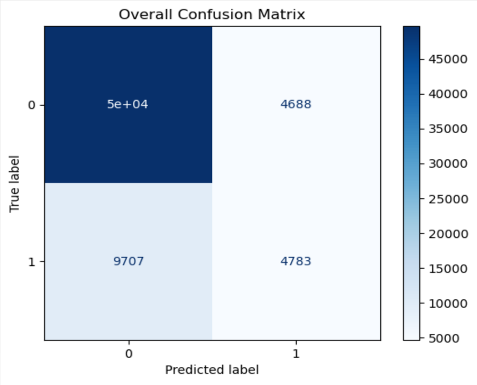
Hierarchical Clustering



Performance:

- Clustering for a full and single location data is incomplete.
- Potential for accurate classification with data tuning.

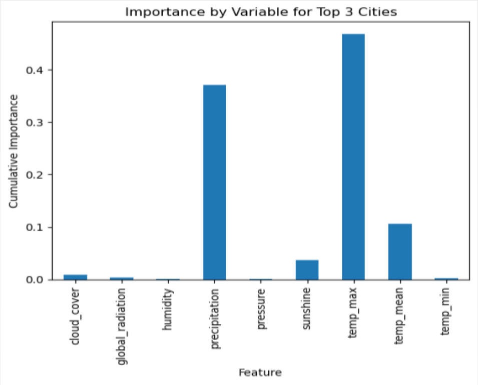
Convolutional Neural Network



Performance:

- High overall accuracy, bias towards the majority class.
- Potential for accurate classification with data tuning.

Random Forest



Findings:

- Identified key locations and climate variables that contribute most to data variance.

Thought Experiments

I have devised three thought experiments to better understand and predict climate-related phenomena.

Detecting and Analyzing Weather Extremes

Use anomaly detection models on historical climate data.

Predicting Future Weather Patterns

Utilize time series forecasting models like LSTM and Prophet

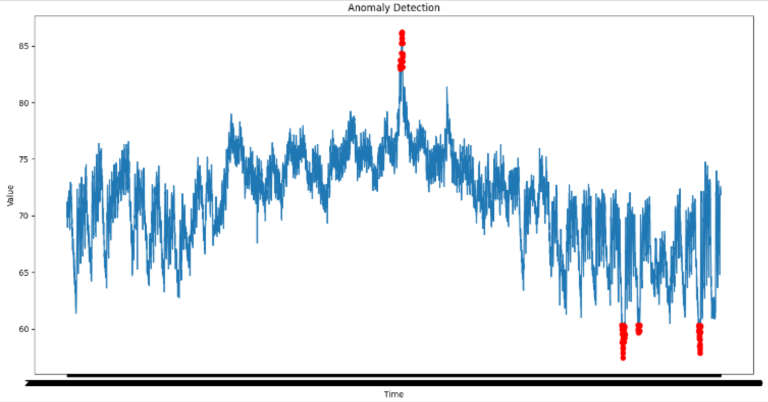
Identifying Safe Living Locations

Apply deep learning models to assess and predict climate safety.

Thought Experiment 1: Detecting and Analyzing Weather Extremes

Objective

- Identify and analyze weather patterns outside the regional norm in Europe.



Sample Anomaly Detection Plot

Approach

- Identify outliers in the data using anomaly-detection models like
  - **Isolation Forests**
  - **One-Class Support Vector Machines (SVM)**
  - **Local Outlier Factor (LOF)**
- Employ **Generative Adversarial Networks (GANs)** to generate synthetic instances of extreme weather conditions to augment the training dataset to provide a more comprehensive basis for anomaly detection.

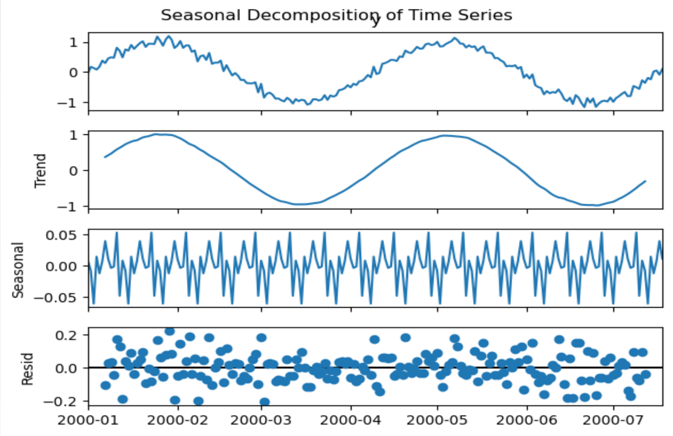
Data

- Use the existing climate dataset and filter for unusual values in key variables defined as those that fall beyond certain statistical thresholds (e.g., above the 95th percentile or below the 5th percentile).
- Augment the dataset with data representing rare but plausible extreme weather scenarios for more robust model training.

Thought Experiment 2: Predicting Future Weather Patterns

Objective

- Generate possibilities for future weather conditions over the next 25 to 50 years.



Sample Time Series Forecasting Plot

Approach

- To predict future weather conditions, use time series forecasting models like
  - **Long Short-Term Memory (LSTM)**
  - **Prophet**
  - **Gated Recurrent Units (GRU)**
- Implement **Neural Architecture Search (NAS)** to automatically optimize neural network architectures, enhancing the performance of the time-series forecasting models

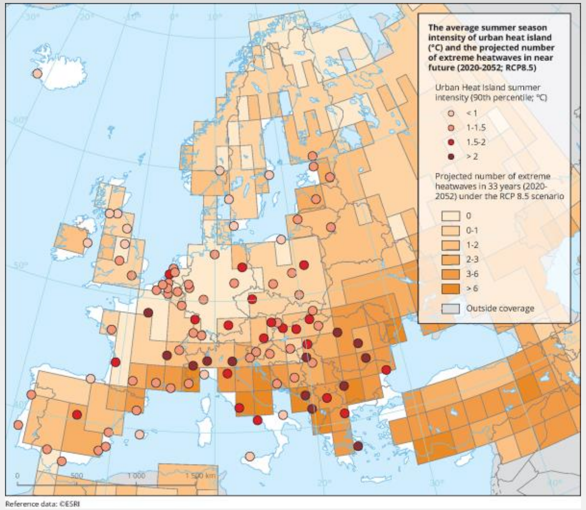
Data

- The current dataset contains sequential data and is appropriate for use with these suggested models.
- Similar climate datasets from other regions can be added to make the models more robust.

Thought Experiment 3:
Identifying Safe Living Locations

Objective

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



Sample European Heatmap

Approach

- Identify key climatic factors for determining safety Using robust algorithms capable of handling complex relationships such as
  - **Random Forests**
  - **Gradient Boosting Machines (GBM)**
  - **Support Vector Machines (SVM)**
  - **Neural Networks**
- Use **ensemble methods** to combine insights from multiple predictive models to assess safety more robustly
- Apply deep learning to model complex interactions between climatic factors and **human-centric metrics** to forecast safe zones.

Data

- Incorporate data on weather-related deaths and weather-related migrations.
- Incorporate data on socio-economic or urbanization factors to provide a broader definition of "safe."

Thought Experiment Conclusions

Summary

Detecting Weather Extremes

- Suggested use of anomaly detection models like Isolation Forests and LOF.
- Potential to augment datasets with GANs for improved anomaly detection.

Predicting Future Weather Patterns

- Suggested time series forecasting models include LSTM, Prophet, and GRU.
- Neural Architecture Search (NAS) recommended to optimize model architectures.

Identifying Safe Living Locations

- Proposed use of Random Forests, Gradient Boosting Machines, and ensemble methods.
- Integration of socio-economic data for comprehensive safety analysis.

Recommendations

Most Promising Experiment: Predicting Future Weather Patterns

RATIONALE

Has the potential to provide actionable insights well into the future and is crucial for long-term planning and adaptation strategies.

IMPACT

Improved forecasting can significantly influence policy-making, urban planning, and disaster preparedness, potentially saving lives and resources.

FEASIBILITY

While Quantum ML is still emerging, the groundwork with NAS and existing machine learning frameworks sets a realistic path for development.

DATA

Leverages existing climate data sets, with the potential to seamlessly integrate new data as it becomes available.

Approach

Initial Model Testing

- Begin with preliminary testing of suggested models using existing climate dataset.
- Evaluate performance and refine models based on initial results.

Data Enrichment

- Plan to incorporate real-time climate data, socio-economic data, and GIS data.
- Enhance robustness and accuracy of predictive models.

Model Optimization

- Utilize Neural Architecture Search (NAS) to optimize model performance.
- Apply transfer learning with similar climate datasets for broader applicability.

Integration & Validation

- Develop a unified platform to integrate different models and validate predictions.
- Continuous monitoring and updating of models based on new data and findings.

Collaboration

- Work closely with climatologists, urban planners, and policymakers.
- Ensure practical applicability and relevance of the models in real-world scenarios.

Thank You!

YOU CAN REVIEW THE SCRIPTS USED FOR THIS PROJECT IN THIS GITHUB REPOSITORY