ClimateWins Weather Prediction Strategy Using Machine Learning

FEMINA

09/11/2024

Objective

Project Objectives:

- Identify unusual weather patterns in Europe.
- Monitor increasing trends in extreme weather.
- Forecast future climate conditions (25–50 years).
- Assess safe regions in Europe.

Company Context: ClimateWins, a nonprofit focused on data-driven climate solutions.

Three Thought Experiments:

Random Forest for Weather Outcome Prediction

Utilizes multiple decision trees on random data subsets to predict weather events like temperature and precipitation. This ensemble method increases accuracy and helps identify anomalies by examining various environmental factors and their interactions.

Deep Learning for Image-Based Weather Classification

This experiment proposes using a **Convolutional Neural Network (CNN)** to classify radar and satellite images, enabling ClimateWins to better interpret complex weather data visually. The model's performance was enhanced through **Bayesian optimization.**

GANs for Synthetic Climate Scenarios

Uses GANs to create realistic synthetic weather data, augmenting existing datasets and enhancing model robustness. This synthetic data can help simulate future climate scenarios, assisting ClimateWins in preparing for potential long-term climate shifts and identifying safer zones.

Data Requirements

Data Provided: Weather observations from 18 European stations (late 1800s to 2022), including temperature, wind speed, snow, etc.

Additional Data Needed:

- Socioeconomic and infrastructure data.
- Satellite imagery for spatial weather trends.
- Global climate indexes (for anomaly detection and regional comparisons).

Random Forest for Weather Outcome Prediction

Objective:

Detect and predict unusual weather patterns by analyzing historical weather data to identify anomalies.

Machine Learning Method:

Model: Utilized **RandomForestClassifier** to analyze key features like precipitation and temperature across multiple European weather stations.

Hyperparameter Tuning: Applied **RandomizedSearchCV** to fine-tune parameters including n_estimators, max_depth, and min_samples_split, achieving optimized model performance.

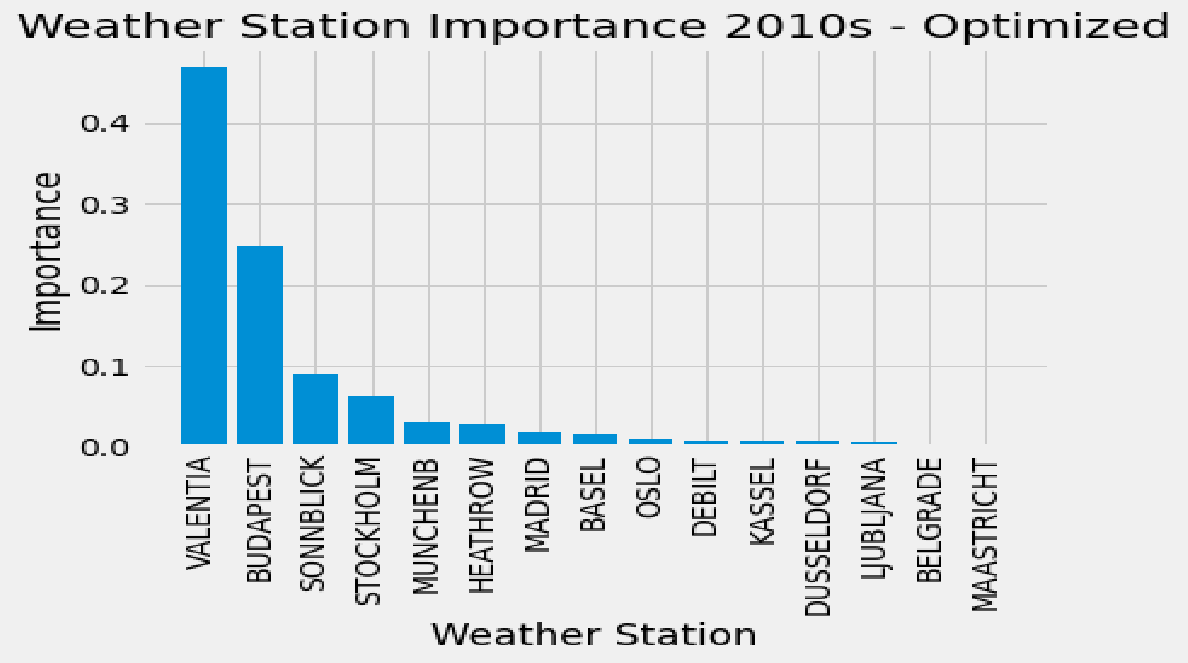
Data Requirements

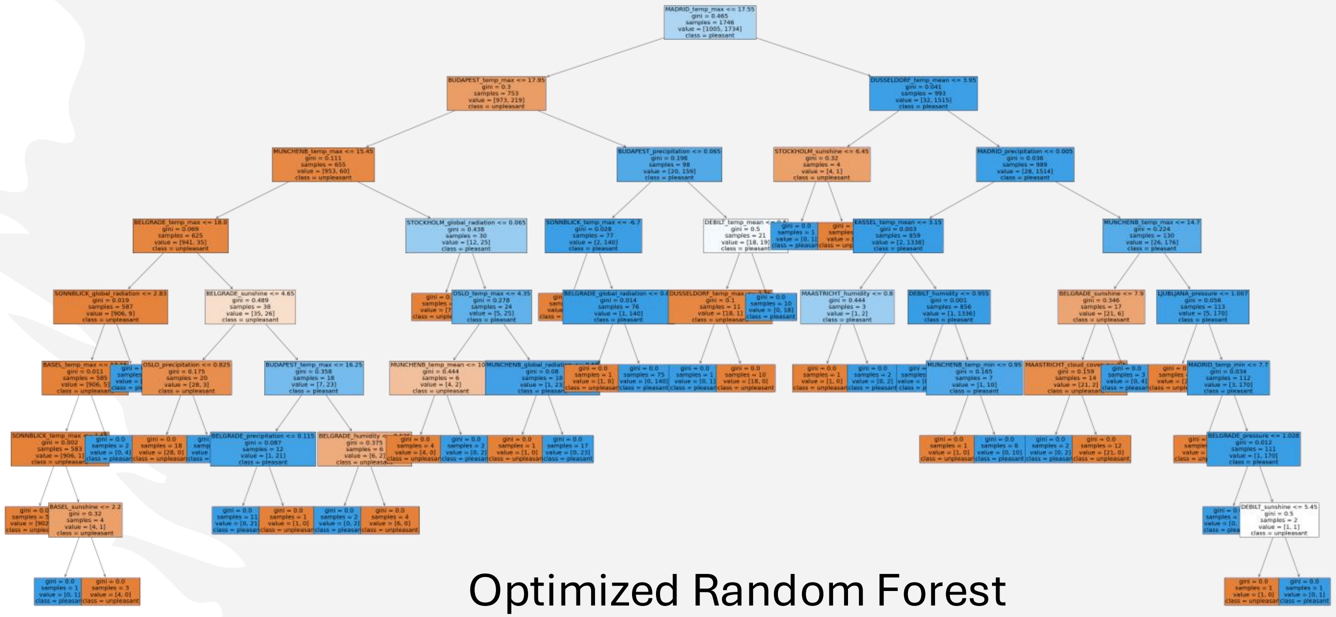
Historical Weather Data: Longitudinal data on temperature, precipitation, wind speed, etc., across European weather stations.

Environmental and Geographic Data: Contextual data (e.g., elevation, proximity to water) to refine predictions based on local geography.

Key Insights

- **Optimized Model**: RandomForestClassifier with tuned hyperparameters.
- **Accuracy**: Improved from 97% initially to approximately 98% post-optimization.
- **Top contributing stations:** Valentia (highest importance at 0.468), followed by Budapest, Sonnblick, and Stockholm.





Optimized Random Forest

Deep Learning for Image-Based Weather Classification

Objective

CNNs can effectively interpret radar and satellite imagery, allowing for more accurate classification of weather conditions and enhancing the prediction of weather trends to aid ClimateWins in identifying changing patterns.

Machine Learning Method:

Model: Developed a **CNN model** to classify radar images into weather categories (e.g., cloudy, rainy, sunny).

Hyperparameter Optimization: Leveraged **Bayesian Optimization** to refine parameters such as the number of neurons, batch size, and learning rate, targeting improved accuracy and model convergence.

Model Performance Before Optimization

- **Initial Accuracy**: Around **25%** across all epochs, indicating poor model performance.
- **Loss and Convergence**: High and fluctuating loss values showed ineffective optimization, with the model biased towards predicting a single class ("Basel").
- **Classification Diversity**: Limited recognition—model identified only 7 out of 15 stations, underscoring restricted learning capability.

Optimized Model Performance

- **Initial Accuracy**: **59.85%** with a loss of **1.37** after the first few epochs.
- **Final Accuracy**: Achieved **92.19%** accuracy by the 47th epoch, with steady improvements per epoch.
- **Loss and Convergence**: Loss values consistently decreased, indicating effective convergence.
- **Enhanced Classification**: Model expanded to recognize 13 out of 15 stations, demonstrating better adaptability and differentiation between varied weather stations.

Pred True	BASEL	BELGRADE	BUDAPEST	DEBILT	DUSSELDORF	HEATHROW	KASSEL \
BASEL	3559	48	13	4	11	5	2
BELGRADE	294	526	130	15	10	17	5
BUDAPEST	71	4	89	14	4	8	1
DEBILT	23	0	2	39	9	2	0
DUSSELDORF	11	0	0	1	5	6	0
HEATHROW	18	0	0	0	3	38	0
KASSEL	3	0	0	0	0	0	1
LJUBLJANA	15	0	1	0	0	0	0
MAASTRICHT	4	0	0	0	0	1	0
MADRID	89	4	1	0	0	3	0
MUNCHENB	4	0	0	0	0	0	0
OSLO	0	0	0	0	0	0	0
STOCKHOLM	1	0	0	0	0	0	0
VALENTIA	1	0	0	0	0	0	0

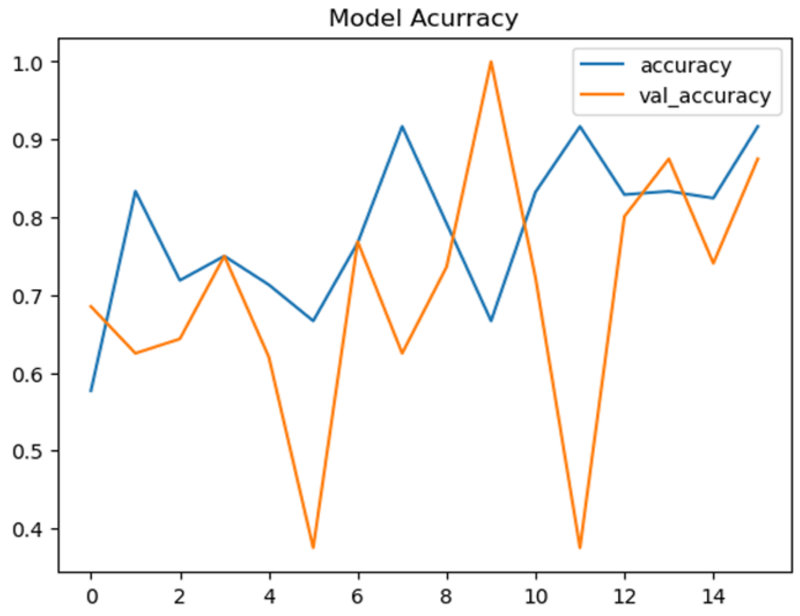
Pred True	LJUBLJANA	MAASTRICHT	MADRID	MUNCHENB	OSLO	STOCKHOLM
BASEL	6	1	23	1	9	0
BELGRADE	11	0	68	2	14	0
BUDAPEST	4	0	12	0	7	0
DEBILT	0	0	3	0	3	1
DUSSELDORF	1	0	4	0	1	0
HEATHROW	1	0	18	0	4	0
KASSEL	2	0	3	1	1	0
LJUBLJANA	35	0	9	0	1	0
MAASTRICHT	0	1	2	0	1	0
MADRID	1	0	353	0	7	0
MUNCHENB	0	0	0	3	1	0
OSLO	0	0	0	0	5	0
STOCKHOLM	0	0	0	1	2	0
VALENTIA	0	0	0	0	0	0

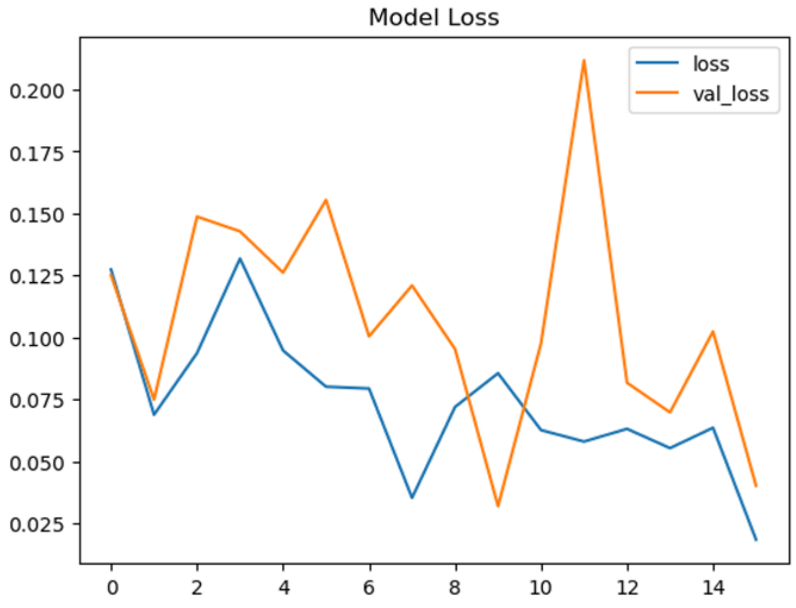
Radar Recognition with Deep Learning

Model Details:

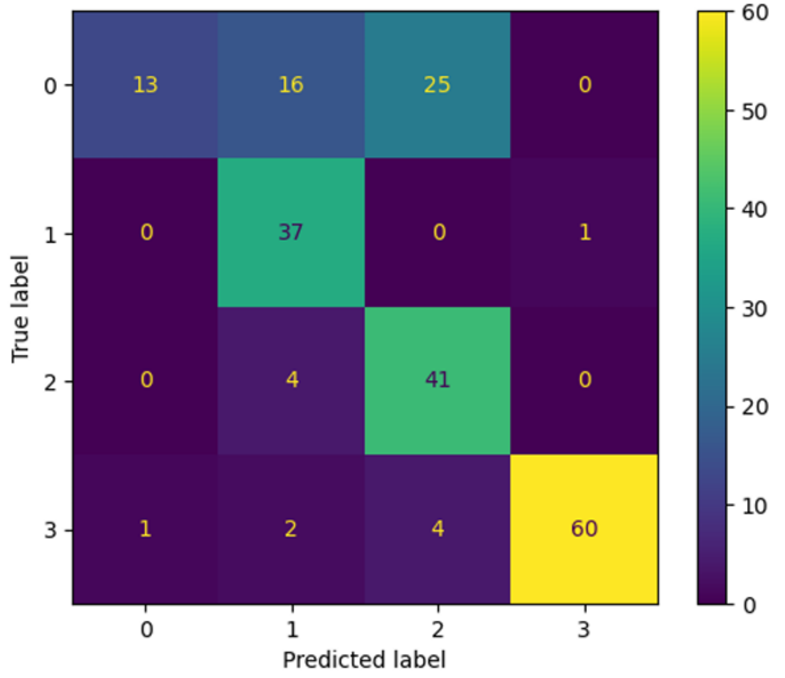
- **Architecture**: CNN for radar-based weather classification
- **Epochs**: 16
- **Accuracy**: 91%
- **Loss**: 0.0184

- High accuracy in predicting most weather types
- Some confusion between similar weather classes, indicating overlapping features
- Strong classification performance overall, with minor misclassification





Confusion Matrix for Weather Condition Classification



GANs for Synthetic Climate Scenarios

Objective:

Create realistic simulations of future weather patterns to explore possible climate scenarios and identify safe regions.

Machine Learning Method:

Generative Adversarial Networks (GANs) - Two networks (generator and discriminator) work in competition to create and validate realistic synthetic data, producing new scenarios that match known data patterns.

Data Needed:

- **Historical Climate Data**: Temperature, precipitation, storm frequency, and extreme weather events as a foundation for GANs to learn from.
- **Synthetic Data Outputs**: Simulated weather data generated by GANs to augment training datasets and provide climate scenario projections.

GAN-generated climate scenarios can help ClimateWins explore a wide range of possible future weather patterns, enabling a proactive approach to climate adaptation and identifying safer locations for living in Europe based on projections for the next 25 to 50 years.

Summary of Thought Experiments

Deep Learning for Image-Based Weather Classification

- Model achieved 91% accuracy with improved class differentiation after tuning.
- Strong potential for visual data analysis, making it suitable for regions with radar coverage; aligns with ClimateWins' need for real-time weather trend analysis.

Random Forest for Identifying Weather Anomalies

- Model improved from 97% to 98% accuracy after optimization, highlighting shifts in key contributing stations.
- High potential for identifying regions with increasing climate anomalies, aligning with ClimateWins' goal of tracking abnormal weather patterns.

GANs for Synthetic Weather Projections

- High potential for long-term climate projection and training data augmentation, which supports ClimateWins' mission of preparing for future weather trends.

Recommendations and Next Steps

Most Promising Approach: Random Forest for Identifying Weather Anomalies due to its high accuracy, feature interpretability, and direct alignment with detecting and understanding climate shifts.

Next Steps for Analysis:

- **Expand Data Collection:** Increase historical and radar data across diverse regions to support all models and enhance robustness.
- **Experiment with Combined Architectures:** Integrate CNN and RNN architectures for spatiotemporal data to improve predictions in weather classification.
- **Synthetic Data Generation with GANs:** Develop GAN-based synthetic weather data to augment datasets for improved model training.

By implementing these next steps, ClimateWins can harness machine learning more effectively to analyze, classify, and predict evolving weather patterns across Europe.

Questions?

Thank you for your attention and interest!

For further information or to discuss next steps, please reach out:

Contact Information:

- **Name:** FEMINA
- **Email:** fjasmin76@gmail.com
- GitHub: Github link