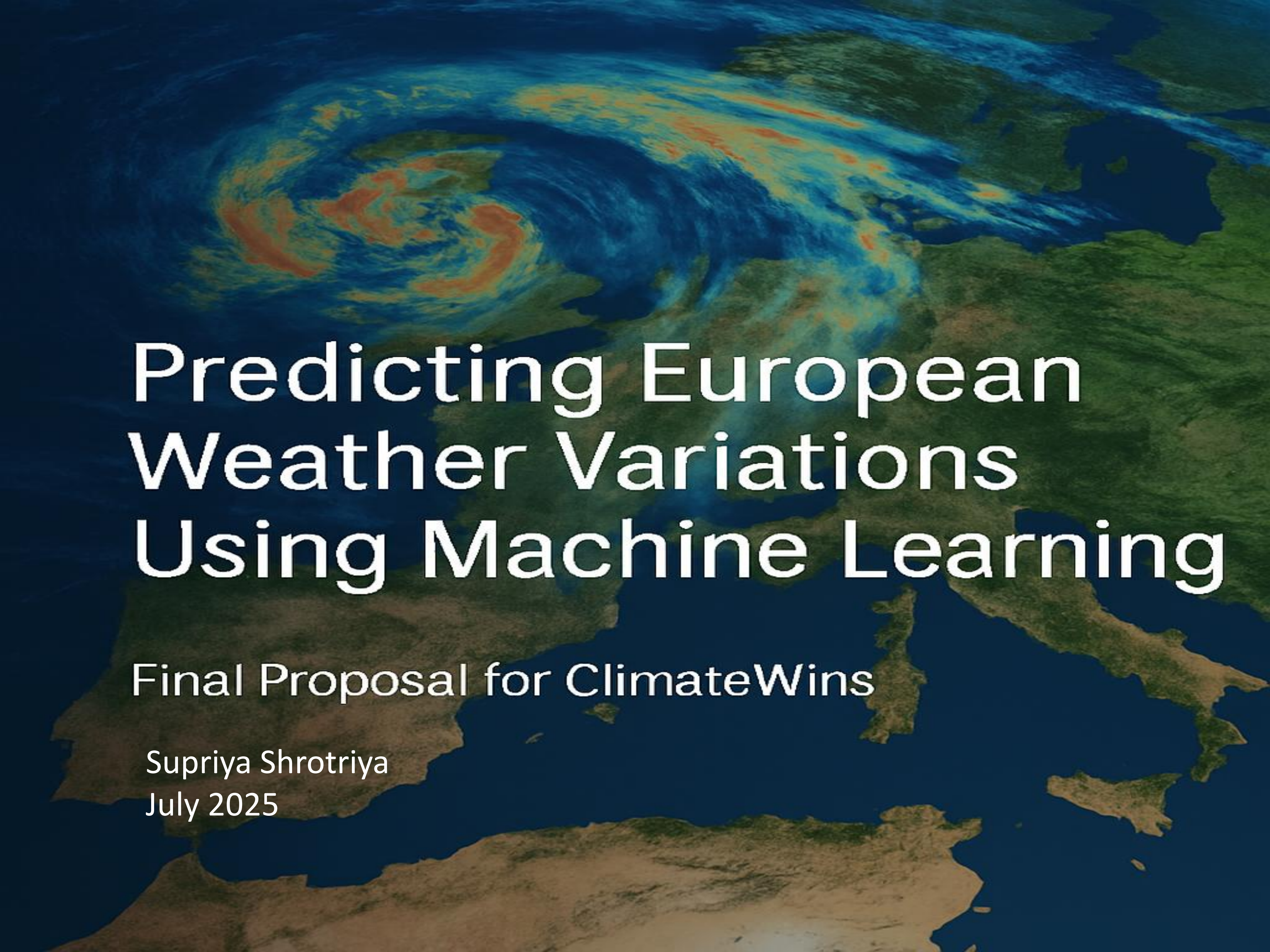


Supriya Shrotriya
July 2025

Project Objectives

- In recent years weather event data has suggested that weather has grown more erratic and harder to predict across the globe, as well as the severity of extreme weather. This could pose a serious risks to human safety and infrastructure. In response to these challenges, ClimateWins is looking toward machine learning models to pursue several key objectives.

Determine if unusual patterns are increasing

Identify weather patterns outside the regional norm

Forecast weather conditions over 25–50 years

Determine safest areas to live in Europe in the next 25-50 years.

Thought Experiments

To support these objectives, we've designed **three thought experiments**, each aligned with a key machine learning approach:

1. Climate Forecasting with GANs

- Use GANs to simulate a range of possible future weather scenarios across Europe for long-term planning.

2. Regional Risk Clustering

- Apply clustering algorithms to group European regions based on shared climate vulnerability patterns.

3. Resilience Classification using Random Forest

- Use Random Forest to rank regions by safety and livability using integrated climate, infrastructure, and socio-economic data.

Machine Learning Approaches

GANS: GENERATE SYNTHETIC FUTURE WEATHER SCENARIOS

K-MEANS OR DBSCAN: GROUP REGIONS BY CLIMATE RISK FACTORS

RANDOM FOREST: CLASSIFY REGIONAL SAFETY AND RESILIENCE

LSTM (BACKGROUND): TIME SERIES PREDICTION FOR TREND MODELING

ONE-CLASS SVM (OPTIONAL): ANOMALY DETECTION IN RECENT DATA

Additional Data Needed

- Historical weather data (temperature, rainfall, storms)
- Real-time and satellite data (radar, imagery)
- Topographic data (elevation, coastlines)
- Socioeconomic and infrastructure data
- Climate vulnerability indicators (flood zones, sea level rise)
- Migration and land use trends

Thought Experiment 1: Climate Forecasting with GANs

Hypothesis :Using a GAN to synthesize weather data, we can train a CNN to predict possible conditions over the next 50 years.

Objective :Generate possibilities for weather conditions over the next 25 to 50 years, based on current trends

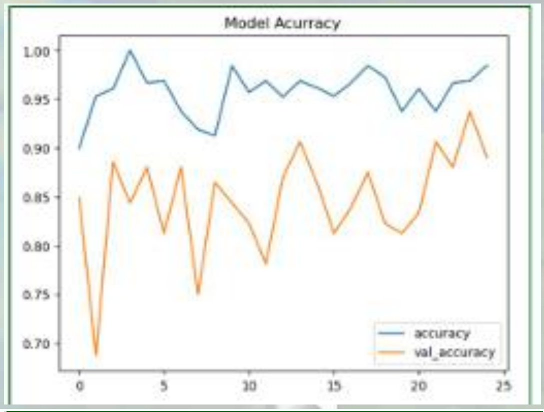
Approach :1. Run Bayesian optimization on a CNN model to evaluate its accuracy 2. Use a GAN to synthesize realistic weather data

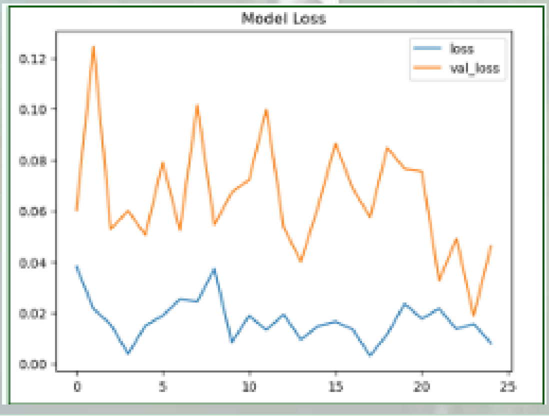
Results :

Bayesian optimization improved the accuracy of CNN modeling from 10.4% to 61.2%

The GAN produced weather data with 98.4% accuracy and 1.1% loss, suggesting high potential for use with the CNN model

Next Steps :• Acquire recommended radar data and run Bayesian optimization • Run the optimized CNN model using the GAN-generated data • Analyze results across years, countries, and regions





Layer (type)	Output Shape	Param #
conv1d_89 (Conv1D)	(None, 15, 61)	610
batch_normalization_32 (BatchNormalization)	(None, 15, 61)	244
dense_442 (Dense)	(None, 15, 61)	3,782
dropout_57 (Dropout)	(None, 15, 61)	0
dense_443 (Dense)	(None, 15, 61)	3,782
dense_444 (Dense)	(None, 15, 61)	3,782
max_pooling1d_89 (MaxPooling1D)	(None, 7, 61)	0
flatten_89 (Flatten)	(None, 427)	0
dense_445 (Dense)	(None, 15)	6,420

Total params: 18,620 (72.73 KB)
Trainable params: 18,498 (72.26 KB)
Non-trainable params: 122 (488.00 B)

Thought Experiment 2: Regional Risk Clustering

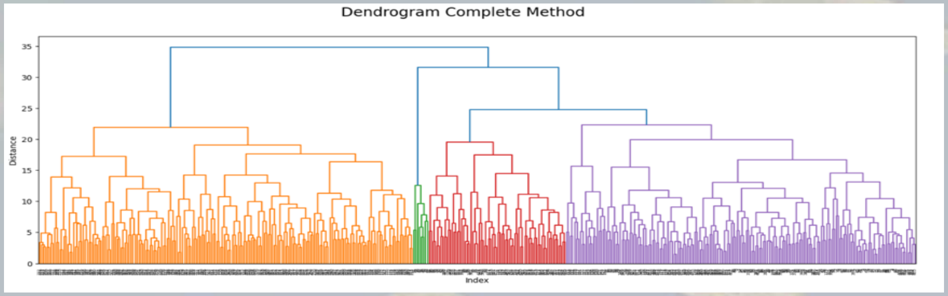
Hypothesis : Using hierarchical clustering, we can go beyond classifying atypical vs typical weather to find actionable categorizations.

Objective :To identify whether unusual weather patterns are occurring

Approach :To assess the viability of this model, we created a dendrogram using existing weather station data. We employed principal component analysis to reduce number of dimensions for optimal resource use.

Results The model produced between two and three clusters consistently, potentially representing binary divisions or low-mid-high divisions.

Next Steps : For this analysis to be valuable, we need to bring in additional types of modeling. We recommend Running this model on multiple years, seasons, and months of data to compare results .Using the results to support other, more conclusive model types



Thought Experiment 3: Resilience Classification

Hypothesis : By optimizing a random forest model, we can identify the most important weather features to track and use that information to identify safe living regions.

Objective :Determine the safest places for people to live in Europe over the next 25 to 50 years

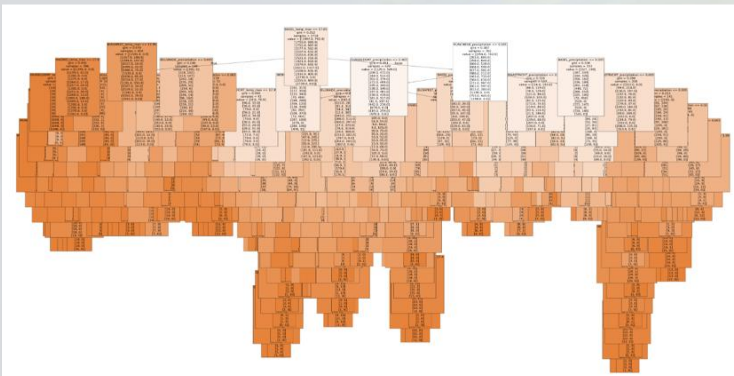
Approach :1. Refine hyperparameters for a random forest model using grid and random search methods 2. Compare unoptimized and optimized results of weather data analysis

Results :• Optimization increased the accuracy of results from 59.1% to 67.3% when looking at all stations • Optimization decreased from 100% to 85% when assessing a single station, suggesting a corrected error from the unoptimized model • Running the optimized and unoptimized models allowed us to weigh feature importance in climate analysis

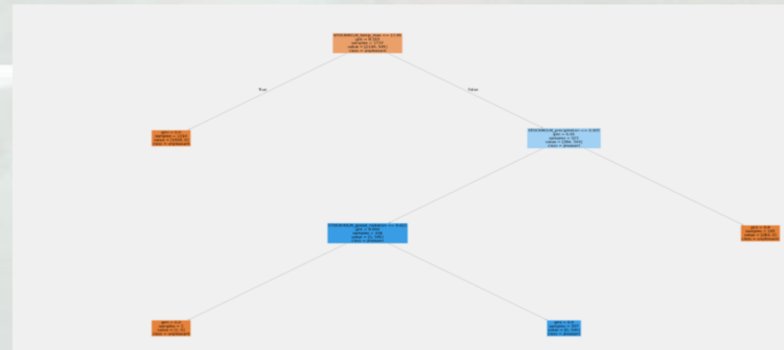
Next Steps :• Run optimization processes on additional stations and decades • Run optimized model on extreme weather data • Compare results to healthcare data to correlate with human wellbeing

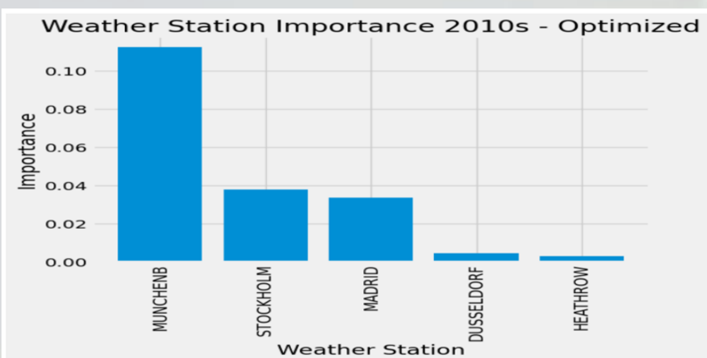
Thought Experiment 3: Resilience Classification

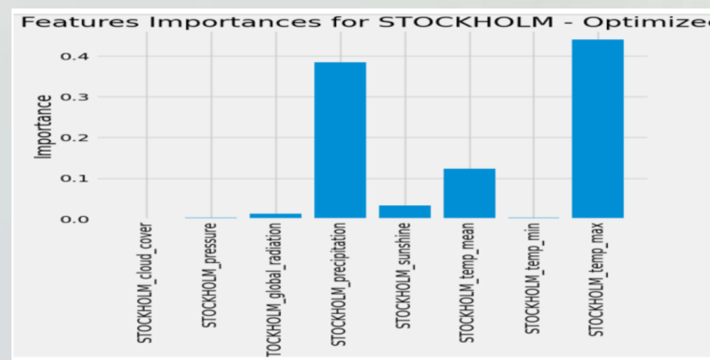
Pre-Optimised



Post-Optimised







Dataset	Accuracy before optimization	Accuracy after optimization
All Weather Stations (2010 – 2019)	58.7%	64.6%
Stockholm (2010-2019)	98.78%	100%

Summary & Recommendations

Recommendations: → GANs for long-term scenario simulation → Clustering for segmenting regional vulnerabilities → Random Forest for practical policy planning

Next Steps: → Data integration and model training → Visualization dashboards (e.g., Tableau) → Validate with real-time weather and climate data

Thank You!

Supriya Kedar Shrotriya

- Contact : supriyapundlik@gmail.com
- GitHub
- LinkedIn