Pleasant Weather Prediction

for European Cities

June 2025

Denis Kleptsov

CONTENT

PROJECT OVERVIEW & OBJECTIVE

01 PROJECT BRIEF

The goal of ClimateWins is to develop robust, accurate, and interpretable machine learning models to predict "pleasant weather" in European cities, supporting stakeholders in tourism, city planning, and public engagement.

02 OBJECTIVE

Deliver actionable weather predictions to enable smarter planning and improved public services across Europe.

RESEARCH HYPOTHESES

TESTED

01

Hyperparameter tuning significantly enhances neural network performance for multilabel weather prediction.

TESTED

02

Tree-based ensemble models outperform other algorithms for high-dimensional, correlated multilabel tasks.

PLANNED

03

Feature selection (Mutual Information, RFE, Boruta) improves non-tree model performance and narrows the gap with ensembles.

PREPARATION & DATA

S DATA SOURCE

- European Climate Assessment & Dataset (ECA&D), 1960–2022
- 18 European cities, 170+ features
- 22,950 samples, 165 features (after cleaning)

P PREPARATION STEPS

- Removal of 24 temporal columns
- Excluded 3 stations with incomplete data
- Multilabel targets for each city

D CLASS DISTRIBUTION

"Pleasant weather" rate varies from

- 5.1% (Valentia) to
- 44.6% (Madrid)
- 0% for Sonnblick

FINAL SET

124 features after excluding temporal data

- Temperature: 42 features
- Pressure: 13
- Wind: 8
- Precipitation: 20
- Radiation: 14
- Clouds: 13

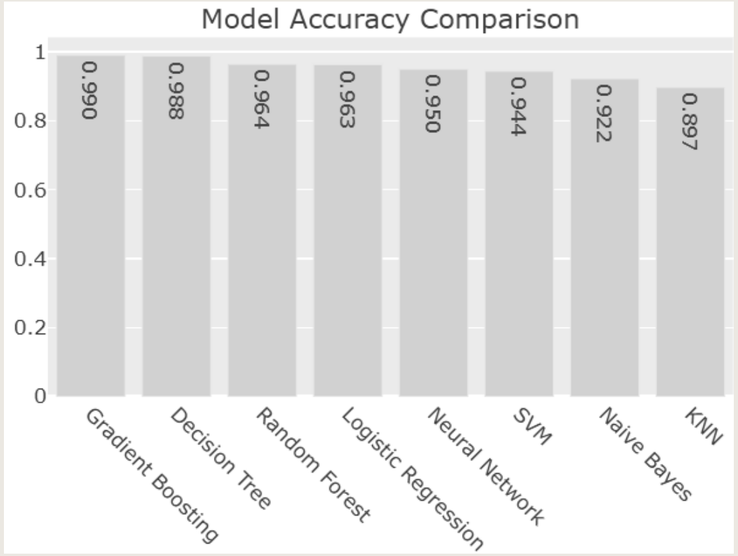
All available features used
(no algorithmic feature selection such as MI/RFE/Boruta implemented for final model)

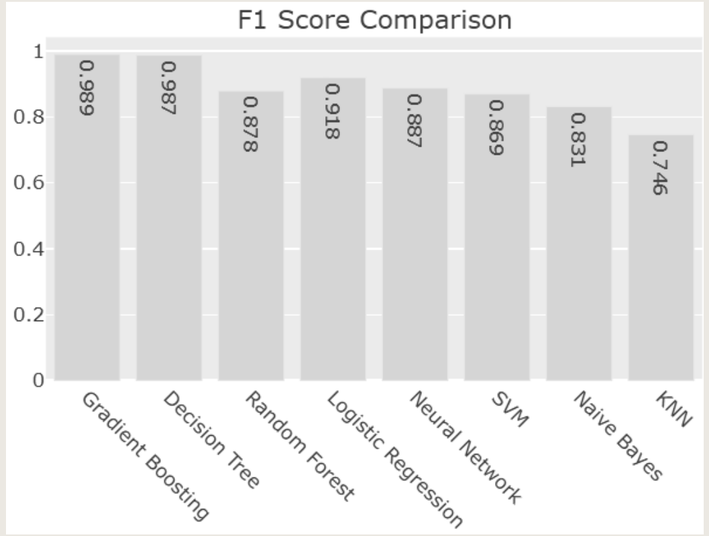
MODEL COMPARISON

Model Comparison

- **Gradient Boosting**: Accuracy 0.9899, F1 0.9889
- **Decision Tree**: Accuracy 0.9875, F1 0.9867
- **Random Forest**: Accuracy 0.9642, F1 0.8783
- **Neural Network**: Accuracy 0.9500, F1 0.8867
- **SVM**: Accuracy 0.9438, F1 0.8689
- **KNN**: Accuracy 0.8971, F1 0.7461

Gradient Boosting and Decision Tree provided the highest accuracy and F1 scores for multilabel ‘pleasant weather’ prediction across cities.





Key Findings & Recommendations

All tested machine learning models were able to predict multilabel pleasant weather with high accuracy for most cities.

Gradient Boosting showed the best overall performance (accuracy: **0.9899**, F1: **0.9889**)

Some cities (e.g., Madrid) are easier to predict than others (e.g., Valentia, due to data imbalance).

Recommendations:

- Deploy the Gradient Boosting model for operational use.
- Monitor performance, especially in cities with low “pleasant weather” rates.
- Consider station-specific tuning and further analysis for low-performing locations.

FUTURE DEVELOPMENT PLAN

25Q3

Interactive Dashboard & API

Develop and deploy dashboards and APIs for real-time prediction access by end users.

25Q4

Algorithmic Feature Selection

Implement advanced feature selection methods (Mutual Information, RFE, Boruta) to optimize models further.

25Q4

Hyperparameter Optimization

Apply systematic hyperparameter tuning (e.g., Optuna, GridSearchCV, Bayesian search) for improved model performance.

26Q1

Model Ensembling & Uncertainty Quantification

Explore stacking, voting, and methods to communicate prediction confidence and reliability.

25Q1

Advanced Visualizations

Create user-friendly, interactive visualizations beyond static charts to better engage stakeholders.

Thank you!

Questions?

Denis Kleptsov | denis.kleptsov@email.com