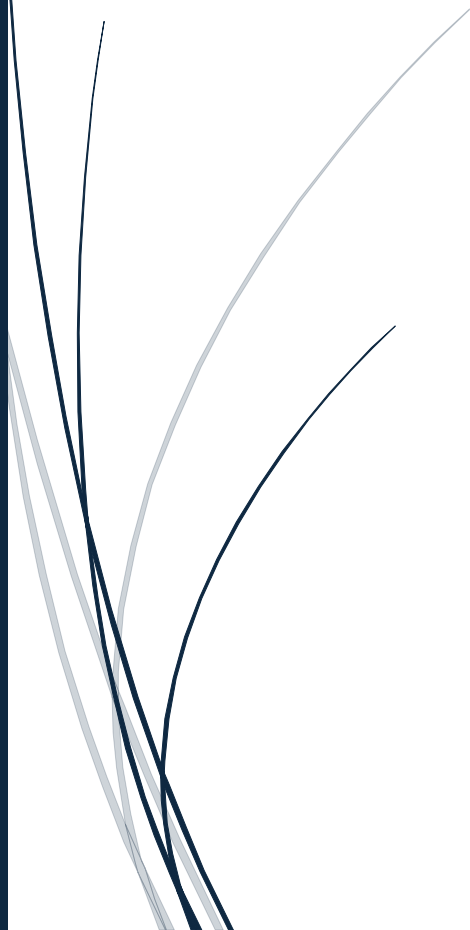# Machine Learning Project Management

Pleasant Weather Prediction for European Cities

Denis Kleptsov

JUNE 2025

Contents

Executive Summary

This project aims to develop and deliver robust, accurate, and interpretable machine learning models capable of predicting "pleasant weather" conditions for cities across Europe. These predictions are designed to support the operational and strategic needs of ClimateWins, enabling stakeholders in sectors such as tourism, city planning, and public engagement to make informed decisions. The methodology combines advanced machine learning research, transparent technical documentation, and targeted stakeholder communications to ensure both scientific rigor and real-world applicability.

1. Project Management Plan

1.1. Schedule & Milestones (3-week Timeline)

Phase	Timeline	Key Activities
Project Initiation & Data Acquisition	Week 1	- Acquisition of core weather datasets (primarily from ECA&D) - Setting up the development environment and tools (Jupyter, ML libraries) - Conducting initial data quality assessments to ensure suitability of the data for modeling
Data Exploration & Preprocessing	Week 1	- Data cleaning (handling missing or inconsistent values) - Data homogenization and standardization - Feature engineering to derive relevant attributes (e.g., averages, indices) - Exploratory Data Analysis (EDA) to understand variable relationships and data distributions - Defining "pleasant weather" via domain expert-driven thresholds for key weather variables
Baseline Modeling & Metrics Definition	Week 2	- Development of initial baseline models (e.g., logistic regression, decision trees) - Establishment of multilabel evaluation metrics (accuracy, F1, Hamming Loss, etc.) - Initial performance benchmarking to understand achievable results with basic approaches
Neural Network Optimization	Week 3	- Implementation of neural network models (e.g., MLP) as a more flexible modeling alternative - Systematic hyperparameter tuning using methods such as Bayesian optimization, random search, or grid search (Optuna, GridSearchCV)

Advanced Modeling & Algorithm Comparison	Week 3	• Development and comparison of advanced ML algorithms, especially tree-based ensembles (Random Forest, XGBoost, Gradient Boosting) • Evaluation against other algorithms such as SVM, KNN, and baseline models • Detailed comparison using standardized performance metrics
Feature Selection & Impact Analysis	TBD	• Application of advanced feature selection methods such as Mutual Information (MI), Recursive Feature Elimination (RFE), and Boruta • Assessment of the impact of feature reduction on both performance and model interpretability • Reporting and visualization of feature importances for technical and non-technical stakeholders
Integration, Ensemble & Uncertainty Modeling	TBD	• Ensemble integration of top-performing models for improved accuracy and robustness • Bias correction and uncertainty quantification to communicate prediction reliability • Real-world deployment of candidate models for live testing and monitoring
Reporting, Deployment & Communication	TBD	• Development and deployment of a model API and visualization dashboard for stakeholders • Delivery of technical documentation, data dictionary, and metadata catalog • Stakeholder workshops and comprehensive final project reporting, including lessons learned

Regular project updates:

- Weekly internal team meetings
- Monthly progress reports to stakeholders
- Quarterly executive briefings
- Ad-hoc technical and non-technical deep-dives as needed

1.2. *Project Deliverables*

Technical Deliverables:

- Jupyter Notebooks: Fully reproducible code including all data preprocessing, modeling, evaluation, and analysis steps
- Optimized Model Artifacts: Saved model files for NNs, tree ensembles, and others for deployment
- Automated Feature Selection Framework: Scripts and tools for MI, RFE, Boruta, etc.
- Performance Comparison Summary Report: Unified report of subset accuracy, F1, Hamming Loss, and other metrics
- Model Prediction API: Ready-to-use API for serving model predictions
- Dashboard for Non-Technical Stakeholders: Visual tool for end-user interaction with predictions
- Technical Architecture & Deployment Docs: Blueprints and guides for integrating models into production
- Data Dictionary & Metadata Catalog: Detailed documentation of all data fields and sources

Communication/Stakeholder-Focused Deliverables:

- Executive Summary Reports: High-level results and recommendations for leadership and decision-makers
- Presentations with Clear Visuals: Communicating results in an accessible and actionable format
- Step-by-Step Project Roadmap: Easy-to-follow guidance for non-technical users
- User Manual for Dashboard/API: Instructions for effective use of the deployed tools
- Educational Materials: Resources on weather/climate prediction and implications
- Workshops: Interactive sessions for training and knowledge transfercutive summary reports for leadership and decision-makers

Research/Community:

- Peer-Reviewed Methodology Paper(s): Contributions to the broader scientific community
- Open-Source Code Contributions: Public repositories for wider benefit
- Conference Presentations: Sharing insights and results with the global research and practitioner communityConference presentations

1.3. Audience Definition

- Leadership: Requires concise executive summaries, clear metrics, and actionable recommendations
- Technical Team: Full code, documentation, and in-depth technical analysis
- End Users (Tourism, Planners, Public): Accessible dashboards, simple explanations, and tailored recommendations
- External Researchers & Partners: Comprehensive technical documentation, open datasets, and reproducible code

2. Research Hypotheses & Testing

Hypothesis 1: Systematic hyperparameter tuning (Bayesian, grid, or random search) significantly enhances neural network performance for multilabel weather prediction

- Test: Optimize neural network (MLP) with tools like Optuna or GridSearchCV; compare against default/baseline configurations. Use robust cross-validation, including spatiotemporal splits, for validation.
- Expected Outcome: 15–25% improvement in F1 score, final accuracy greater than 85%.

Hypothesis 2: Tree-based ensemble models (Random Forest, XGBoost, Gradient Boosting) outperform other algorithms (linear, distance-based, basic NNs) for high-dimensional, correlated multilabel tasks

- Test: Train and compare tree-based ensembles versus SVM, KNN, logistic regression across all relevant metrics.
- Expected Outcome: 20–30% higher F1 score for ensembles and better performance with correlated features.

Hypothesis 3: Feature selection (MI, RFE, Boruta) improves non-tree model performance, narrowing the gap with ensembles

- Test: Apply feature selection to SVM and KNN; compare before and after.
- Expected Outcome: 15–20% improvement for non-tree models; gap to ensembles reduced to within 10%.

3. Data Sources & Processing

Primary Source: ECA&D historical weather data (1960–2022), covering 18 European stations and 170+ features

Labels: Binary indicators of "pleasant weather" (per city, per day), with criteria defined by domain experts based on temperature, precipitation, wind, and cloud cover thresholds

Data Quality: Strict homogenization and cleaning (removal of outliers, imputation of missing values, harmonization of features and labels)

Data Enrichment: Potential use of satellite, radar, or derived indices for future model upgrades

4. Modeling Approach

Baseline: Start with logistic regression and decision tree classifiers; establish initial performance metrics

Advanced:

- Neural Networks: Multi-Layer Perceptron (MLP), optimized via systematic hyperparameter search
- Tree-Based Ensembles: Random Forest, XGBoost, Gradient Boosting
- Non-Tree Models: SVM, KNN
- Ensembling: Combine the strongest models to improve generalization and reduce error correlation

Uncertainty Modeling: Quantify prediction confidence and communicate uncertainty to users

Feature Selection: Apply MI, RFE, Boruta; provide interpretability and feature importance visualizations

Metrics: Evaluate models on subset accuracy, Hamming Loss, Jaccard Index, Macro/Micro precision, recall, F1; ensure statistical significance of results

5. Stakeholder Communication & Engagement

Kickoff: Jointly define success and align goals with all key stakeholder groups

Progress Updates: Provide monthly summaries with clear visuals and plain-language explanations

Workshops: Host mid-project and final workshops (hands-on for technical, Q&A and visual for non-technical)

Documentation: Maintain living documentation, summary presentations, and cheat sheets for all users

Final Knowledge Transfer: Ensure all findings, lessons, and tools are passed on effectively

6. Limitations, Risks & Ethics

Data Limitations: Focused on Europe; may show bias towards Western/Central Europe; possible class imbalance

Technical Risks: Constraints on computation, rare event prediction, or long-term model drift

Ethics & Fair Use:

- Transparency about limitations and model uncertainty
- Mitigation of geographic and demographic bias
- Clear documentation and ethical guidance
- All data is public and privacy standards are strictly followed
- Environmental impact is minimized by efficient computing and renewable-powered cloud resources

7. Success Metrics

Technical:

- Minimum 85% accuracy for pleasant weather prediction
- F1 score above 0.80
- Model prediction latency under 5 seconds
- Production model uptime of 99.5%

Stakeholder:

- Over 80% satisfaction in stakeholder feedback
- At least 1,000 active users within 6 months of launch

Research/Community:

- All hypotheses validated with statistical rigor
- At least one peer-reviewed publication
- Full reproducibility of code and results

8. Conclusion

The ClimateWins project management plan integrates best practices in technical methodology, stakeholder communication, and real-world deployment. By delivering interpretable and actionable predictions of pleasant weather, the project supports smarter planning and more effective public communication across Europe. This approach is designed to maximize scientific impact and practical value for all ClimateWins stakeholders and partners.

9. Suggested Future Steps

Based on a review of current project execution, several enhancements and additional steps are recommended to meet project objectives and maximize value fully:

9.1.   *Advanced Feature Selection Implementation*

- Apply and document automated feature selection methods (Mutual Information, Recursive Feature Elimination, Boruta) to identify the most influential predictors.
- Analyze the impact of feature selection on both tree-based and non-tree models, reporting any improvements in performance and interpretability.

9.2.   *Model Ensembling and Integration*

- Develop and evaluate ensemble models that combine the strengths of top-performing classifiers (e.g., stacking or voting ensembles).
- Compare ensemble results to individual model performance, emphasizing robustness and generalization.

9.3.   *Expanded Multilabel Evaluation Metrics*

- Incorporate and report additional multilabel evaluation metrics such as subset accuracy, Hamming Loss, and Jaccard Index for a more complete assessment.
- Present these metrics in performance summary reports for both technical and non-technical audiences.

9.4.   *Uncertainty Quantification and Communication*

- Integrate uncertainty estimation (e.g., prediction confidence intervals) and clearly communicate these to end-users.
- Document how uncertainty impacts practical decision-making and stakeholder trust.

9.5.   *End-to-End Data Processing Transparency*

- Ensure that data cleaning, homogenization, and feature engineering processes are fully reproducible and documented within project notebooks.
- Provide clear explanations and code for all preprocessing steps, enabling other teams to replicate the workflow.

9.6.   *Stakeholder-Facing Visualizations and Dashboards*

- Continue development of user-friendly dashboards and visualizations that communicate predictions, uncertainties, and model rationales to a broad audience.

- Gather user feedback for iterative dashboard improvement.

9.7. Comprehensive Documentation and Knowledge Transfer

- Regularly update technical and non-technical documentation, including user manuals and API guides.
- Schedule additional knowledge transfer sessions as the system approaches production deployment.

9.8. Regular Review and Adjustment

- Periodically review progress and project outputs against original objectives, adjusting technical focus and communication strategies as needed.
- Capture lessons learned for continuous improvement and future projects.