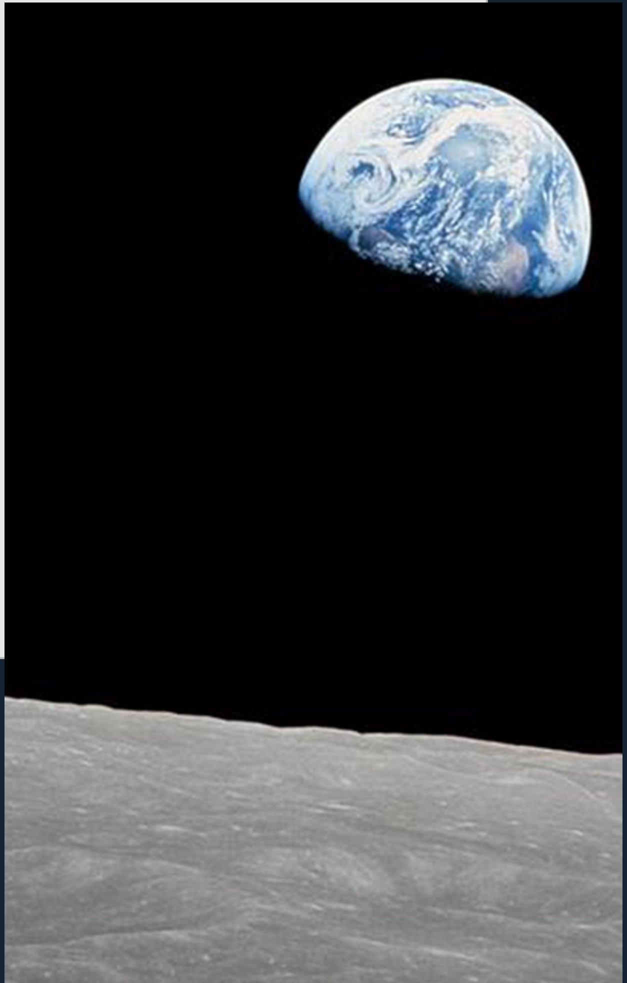
Climate_ Wins

Weather Conditions & Climate Change Analysis

June 2025

Interim Report

Amy Zhang



Objective

Assess **machine learning tools** for enabling **ClimateWins**, a non-profit, to **predict climate change impacts**—beginning with the categorization of mainland Europe weather data.

I.
Data Provenance: Ethical Considerations & Data Integrity

Bias Detection, Label Validity & Ethical Risks in Forecasting

II.
Machine Learning Foundations & Hypotheses for Weather Data

Three Core Hypotheses re: Machine Learning for Weather Analysis

III.
Preliminary ML Model Evaluations for ClimateWin Requirements

Gradient Descent Demo
+
3 Supervised ML Models

I. Data Provenance:
Ethics & Integrity Considerations

- **Date range: 1960 - 2022**
  62 years of observations at per diem granularity

- **18 Weather Stations:**
  Data points include various weather metrics (see right)

- **Dataset Provenance: European Climate Assessment & Data Set Project (ECA&D)**
  - Public initiative est. in 1998 by EUMETNET (collaborative network of meteorological and hydrological services across Europe and Mediterranean), with support from the European Commission.

Full Weather Dataset Overview

Dimension: **date**

PK	date_id	int64
none	date	datetime64
none	month	int64
none	day_of_week	int64
none	day_name	text

Dimension: **weatherstation_loc**

PK	weatherstation_id	int64
none	weatherstation	text
none	country	text

Fact: **weather_metrics**

FK	date_id	int64
FK	weatherstation_id	int64
none	**cloud_cover**	int64
none	**wind_speed**	float64
none	**humidity**	float64
none	**pressure**	float64
none	**global_radiation**	float64
none	**precipitation**	float64
none	**snow_depth**	int64
none	**sunshine**	float64
none	**temp_mean**	float64
none	**temp_min**	float64
none	**temp_max**	float64

‘Pleasant Weather’ Dataset Overview

Dataset of daily, per-station records spanning 1960 to 2022, each labeled with a binary pleasantness score.

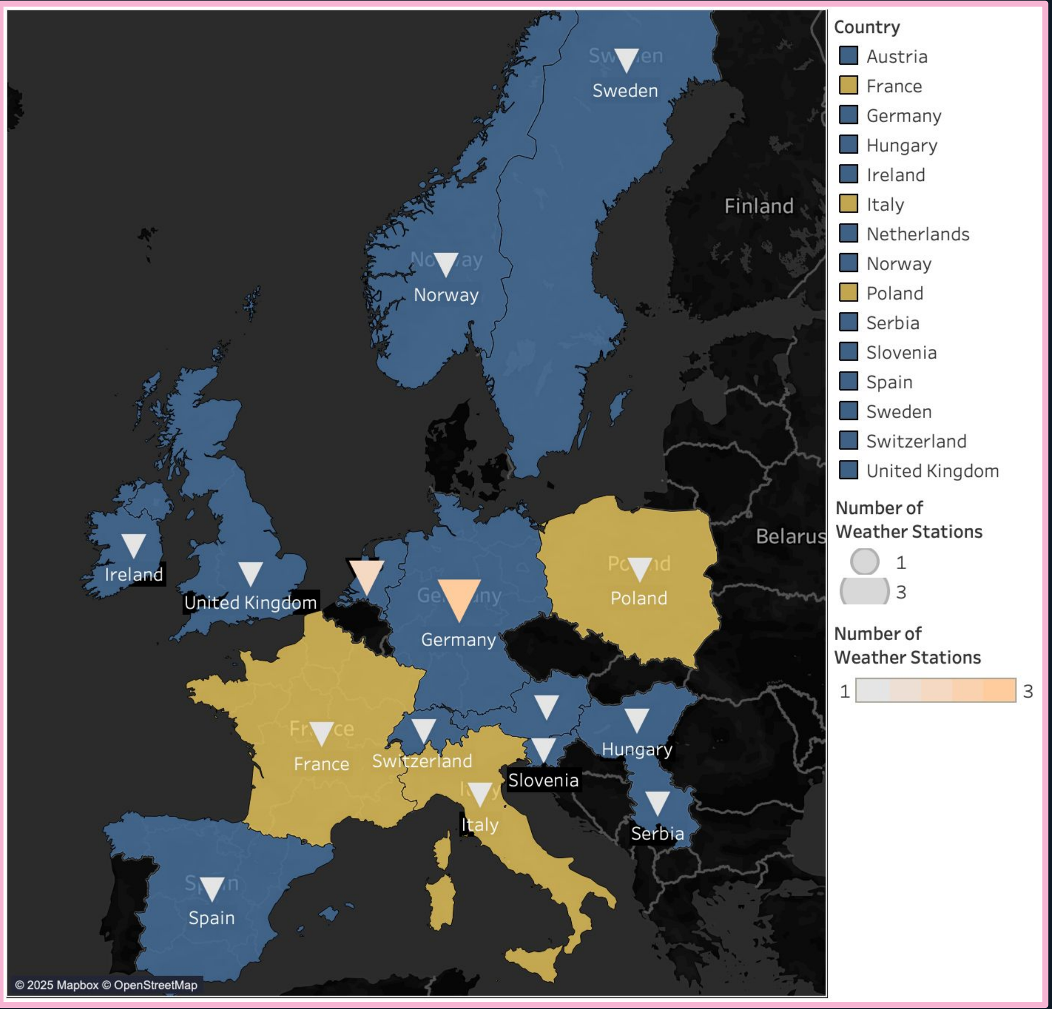
- 1 = Pleasant
- 0 = Unpleasant

I. Data Provenance: Ethics & Integrity Considerations

#1: Geographic Coverage Bias

- **Tours (France), Roma (Italy), and Gdansk (Poland)** are **excluded** from the "pleasant weather" subset. (Each was the **only station** representing its country.)
- Spain is now the sole warm-climate voice.
- Meanwhile, **Germany** has **3 stations**, and the **Netherlands** has **2**.

Geographic gaps in data coverage risk **skewing model learning.**
Models can **overfit to well-sampled regions,** struggling to generalize for others.

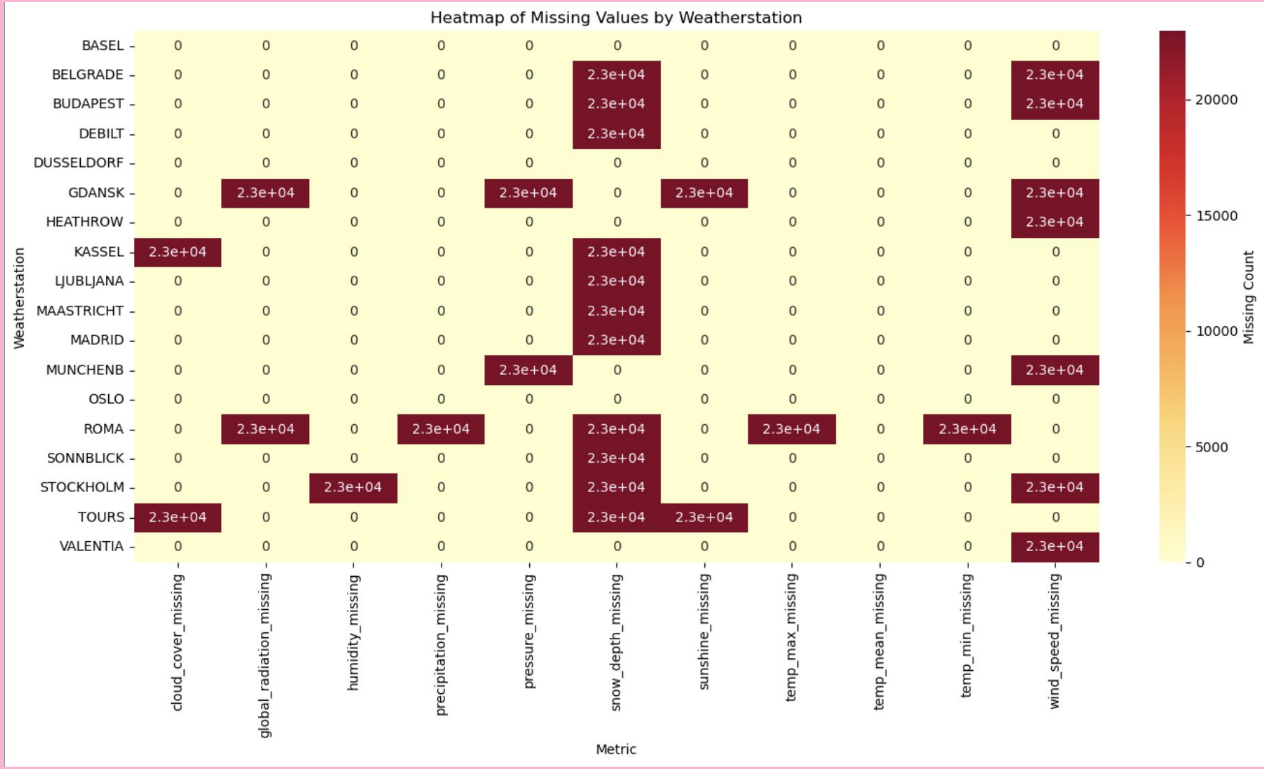


I. Data Provenance:
Ethics & Integrity Considerations

#2: Feature Completeness Bias

- Only **7 out of 18 stations** record *snow depth*.
- **Only Basel, Dusseldorf, and Oslo** have complete weather records.

In trying to make the data ML-compatible, **incomplete features are often dropped**; this can lead to **overrepresentation of simple weather profiles** over more complex or volatile contexts.

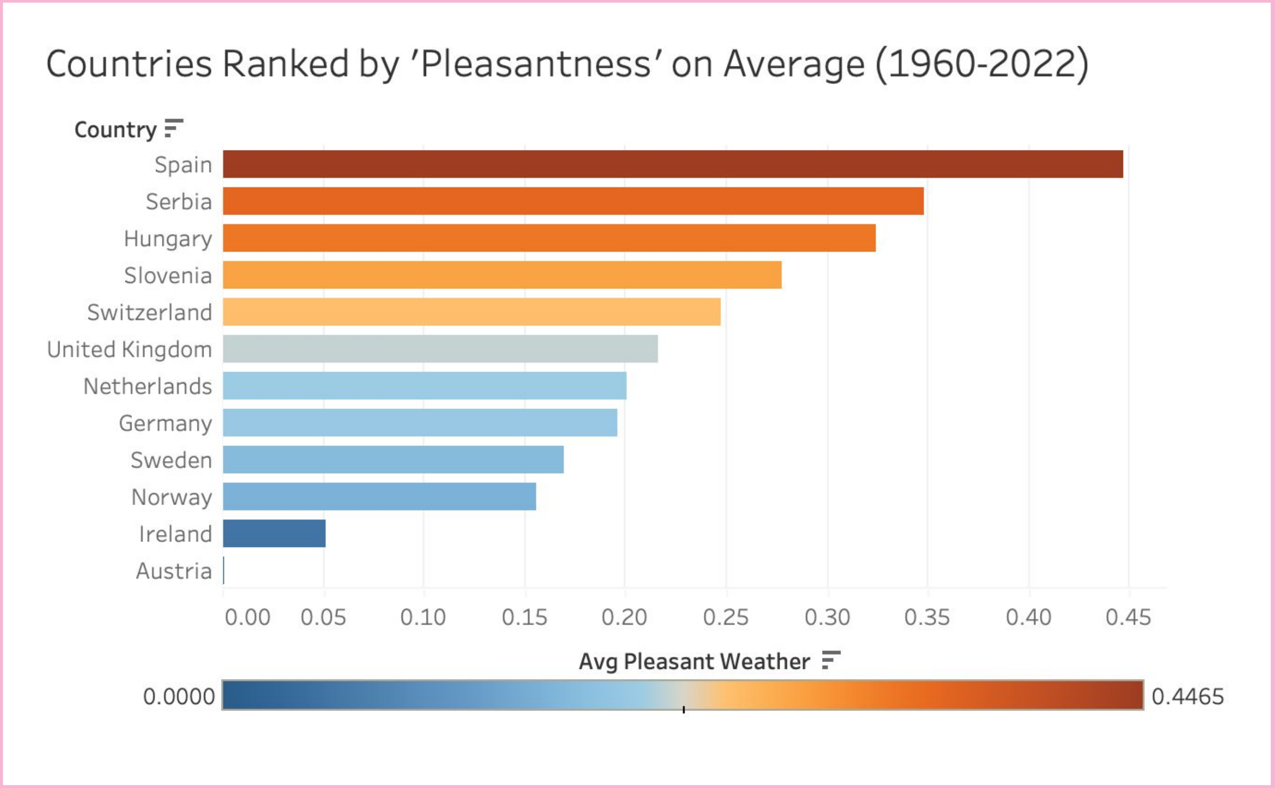


I. Data Provenance: Ethics & Integrity Considerations

#3: Subjectivity in Labeling

- **Sonnblick (Austria):** *0* "pleasant" days
- **Madrid (Spain):** nearly *50/50* split; average 'Pleasantness' ~45%

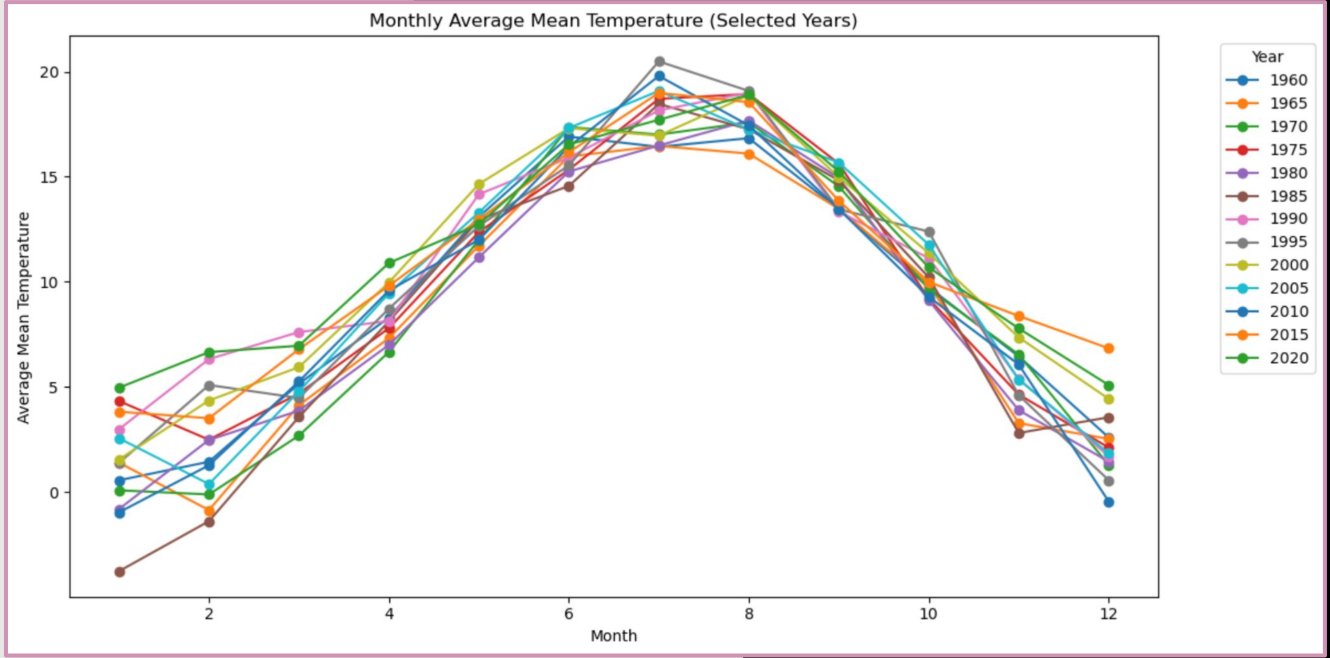
Boolean labels can encode **narrow thresholds** that reflect **implicit normative assumptions** about comfort. These can **shape downstream model behavior,** user expectations (e.g. tourists), and even influence resource allocation.



II. Machine Learning Foundations & Hypotheses for Weather Data

Hypothesis #1: Unlocking Emergent Patterns with ML

- **Traditional analysis** (see right) shows broad seasonal trends.
- But **short-term fluctuations**—intra-annual “spikes and dips”—often defy visual explanation.
- **ML models** can analyze **complex, nonlinear relationships**, identifying subtler climate dynamics.



<u>Hypothesis</u>: The deeper interplay of predictive factors can **enrich time-based models.**

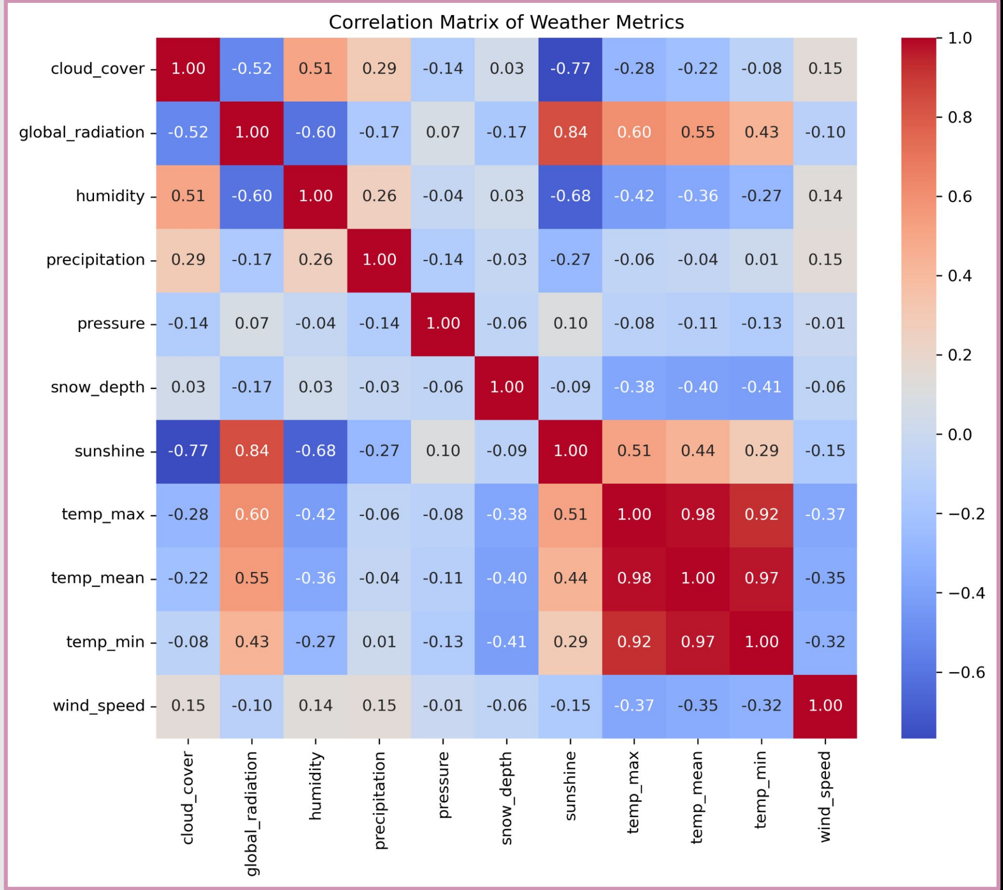
II. Machine Learning Foundations & Hypotheses for Weather Data

Hypothesis #2: Weather as Web – Tapping Feature Interdependencies

Interdependencies exist amongst weather metrics:

- ☁️ Humidity ⇔ Cloud Cover ⇔ Precipitation
- ☀️ Radiation ⇔ Sunshine ⇔ Temperature

<u>Hypothesis:</u> ML models can **leverage correlated feature clusters** as composite predictors. achieving more robust, context-aware forecasts.

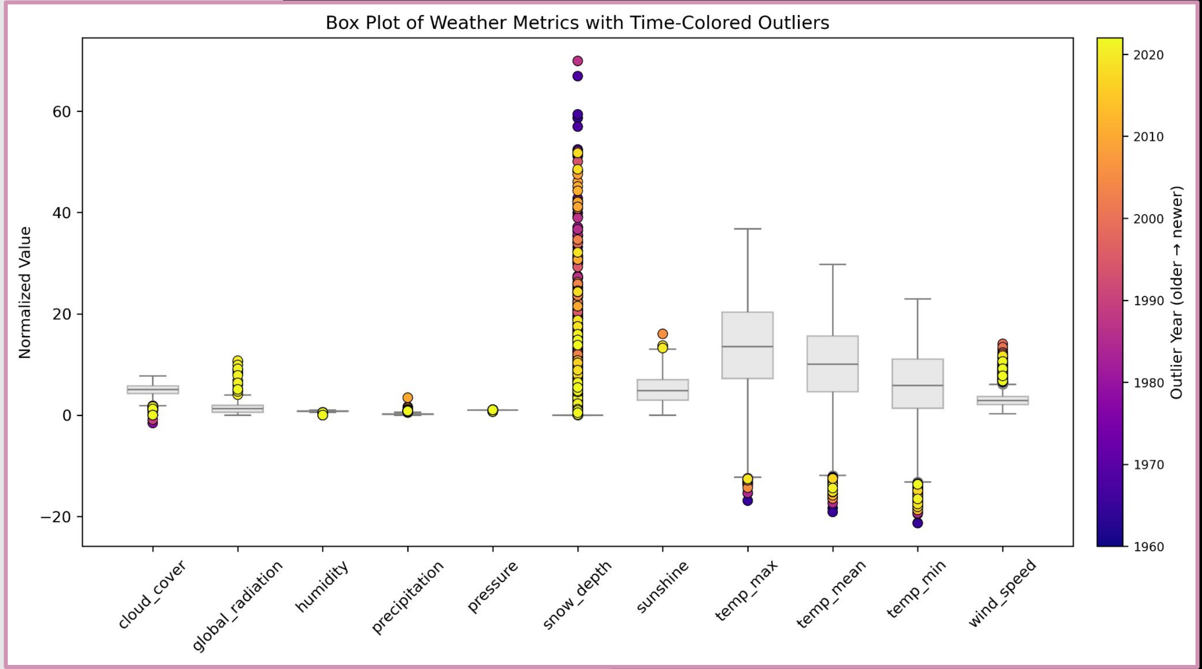


II. Machine Learning Foundations & Hypotheses for Weather Data

Hypothesis #3: Temporal Drift in Weather Dynamics

- Color-coded outliers in weather metric distributions reveal **emerging "new normals."**
- Rising temperatures and reduced snowfall in recent years **shift central tendencies**—making older data points increasingly appear as outliers.

Hypothesis: Machine learning models trained on historical climate data may underperform as key weather variables **drift** over time, reducing predictive accuracy.



Supervised Machine Learning: model evaluations

📈**Linear Regression**
2 univariate models; highlight "learning" as parameter optimization (ex: gradient descent)

🤖**ANN: Artificial Neural Network**

2 ANN – single-layer Logistic Regression & Multi-Layer Perceptron

📍 **KNN: K-Nearest Neighbor**

1 KNN model; highlights "learning" as hyperparameter vs. parameter optimization (ie cross-validation)

🌳**Decision Tree**
1 Decision Tree model; highlights "greedy" vs. gradient-based parameter optimization

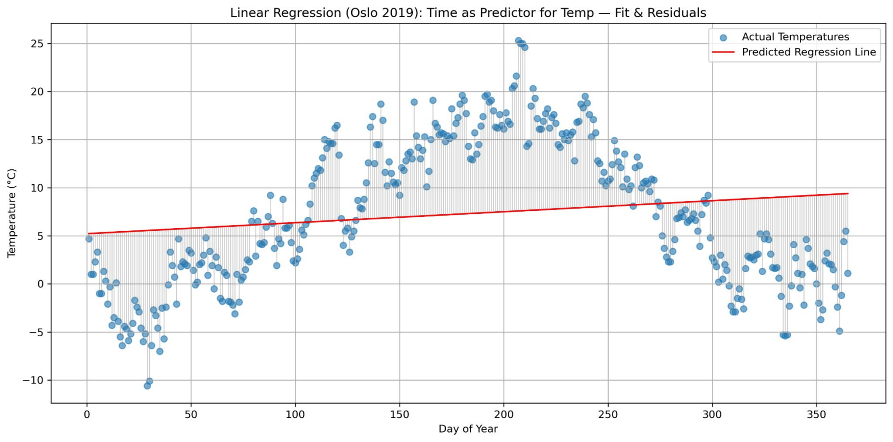
OSLO 2019
(3 features)
Temperature (mean), Time, Global Radiation

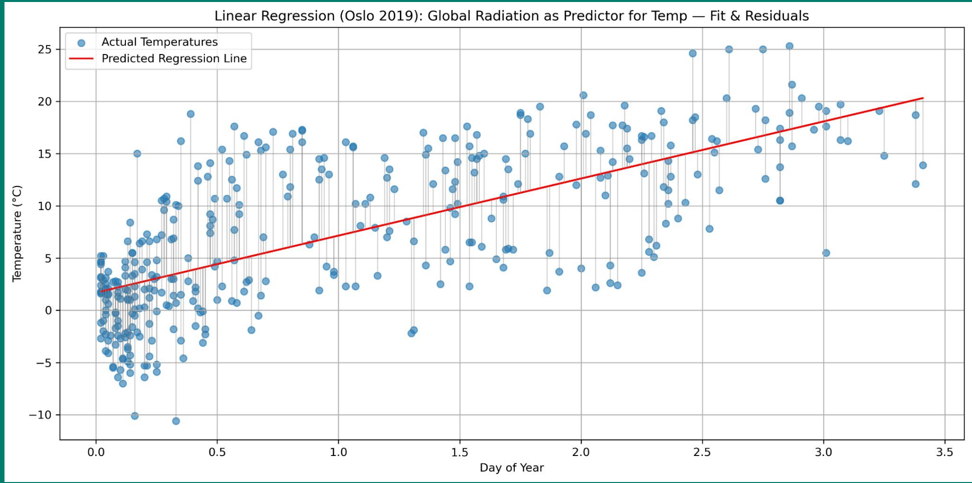
OSLO all years
(6 features)
Temperature (all), Global Radiation, Precipitation, Sunshine

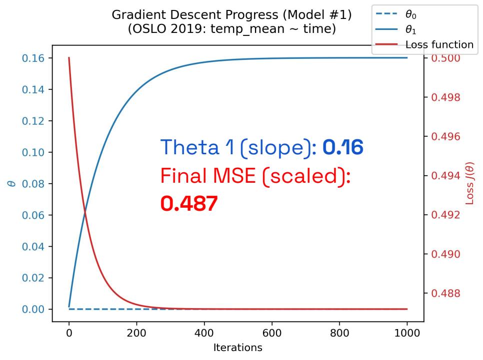
Conclusion
Comparison & Recommendations for Climate Wins

‘Actual v. Predicted’ Comparison: Two Linear Regression Models for OSLO 2019

Both models optimized efficiently. However, as a single-variable predictor, ***global radiation*** proved more relevant due to its **stronger linear relationship** within the underlying data.

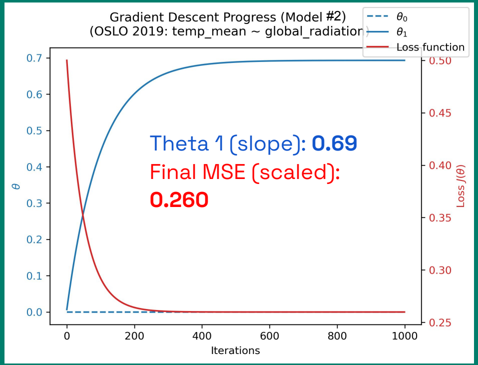






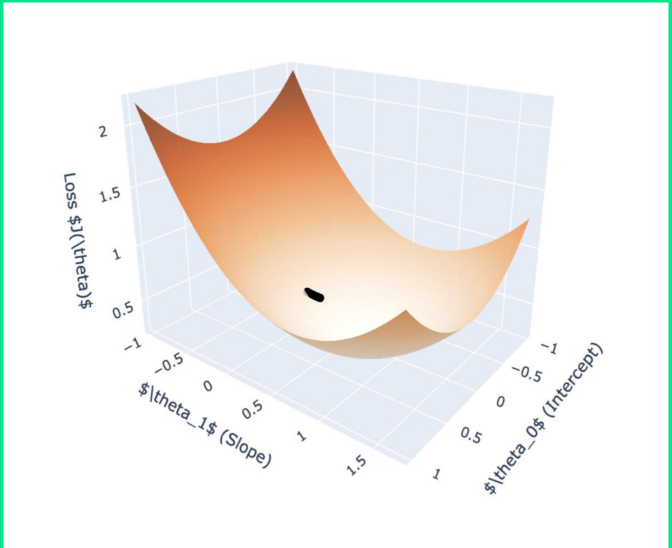
Linear Regression Model #1: **Mean Temperature Over Time**

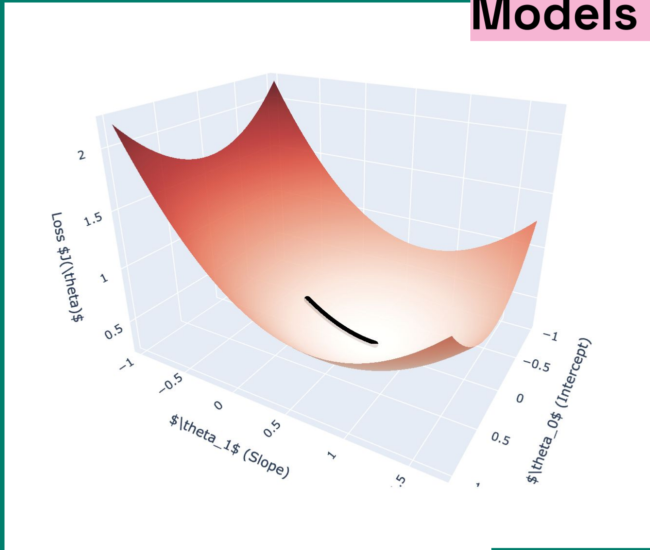
Linear Regression Model #2: **Mean Temperature Over Global Radiation**



📈Surface Plot Comparison: Two Linear Regression Models for OSLO 2019

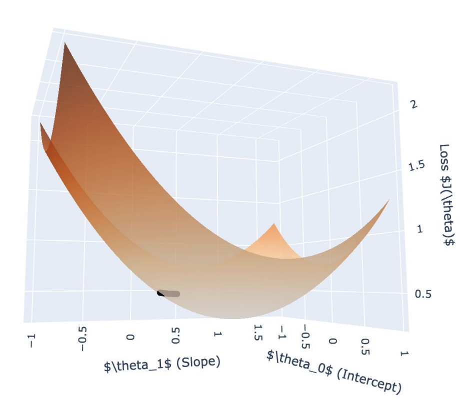
Model 1's trajectory– shorter, misaligned– indexes a flatter, noisier surface: weaker gradient signals.

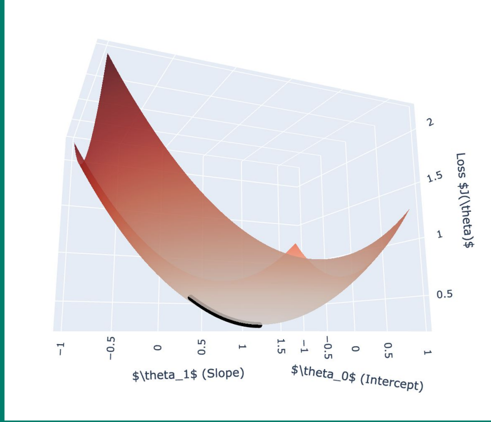




Model 2's path (longer and more curved) reflects a more directional loss surface — strong gradients guiding toward the minimum. **'Global radiation' creates a more informative terrain,**

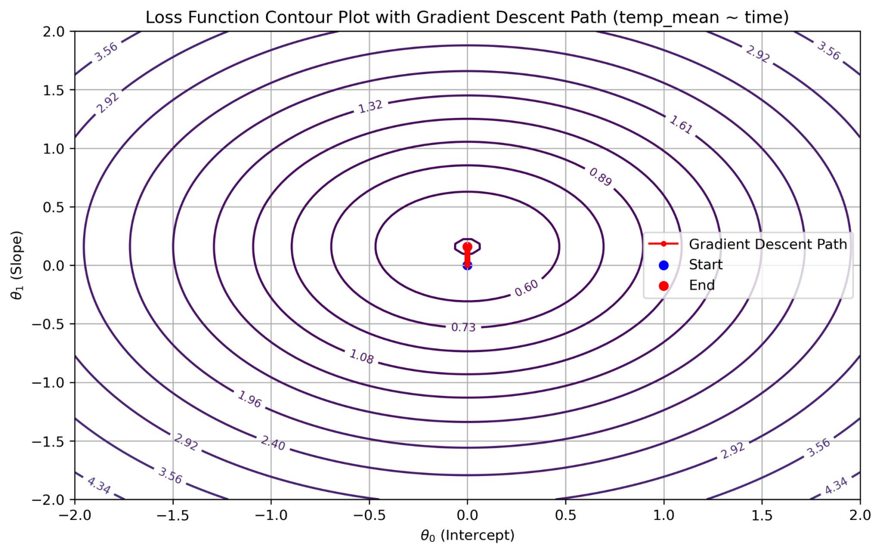
Linear Regression Model #1: **Mean Temperature Over Time**

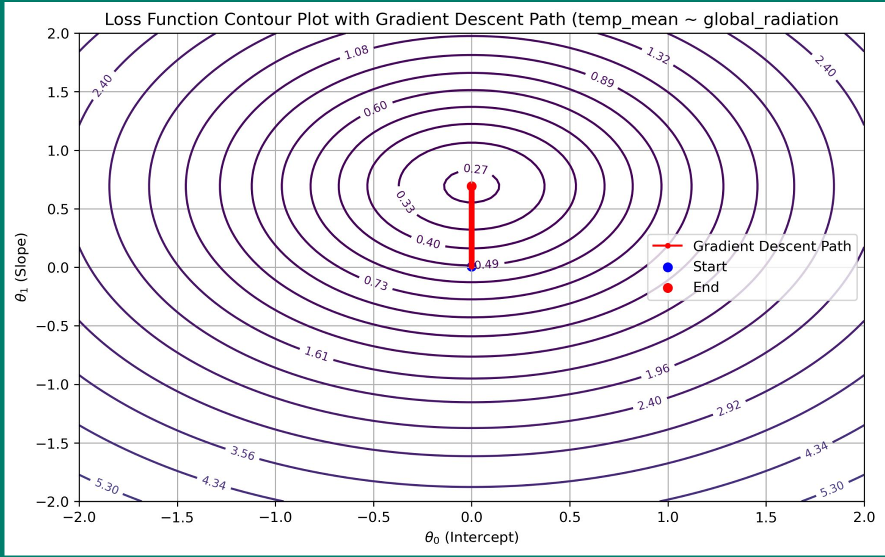




Linear Regression Model #2: **Mean Temperature Over Global Radiation**

📈 Contour Plot Comparison: Two Linear Regression Models for OSLO 2019





Linear Regression Model #1: **Mean Temperature Over Time**

Model 1 (Temp vs. Time):

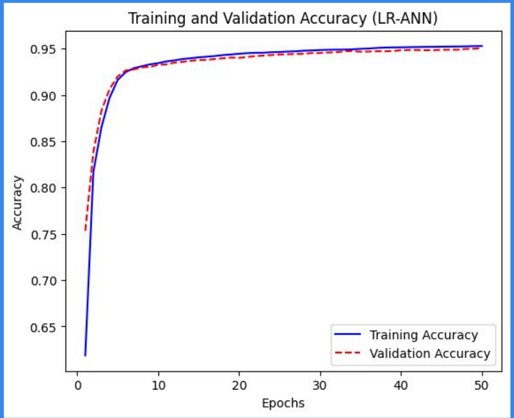
- **Wide, gentle contours** = flatter terrain.
- **Short descent** suggests the model saw little gradient to guide it down.
- Final loss stayed high (~0.60).

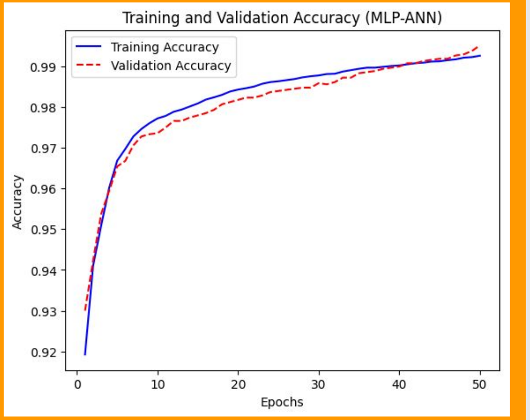
Linear Regression Model #2: **Mean Temperature Over Global Radiation**

Model 2 (Temp vs. Global Radiation):

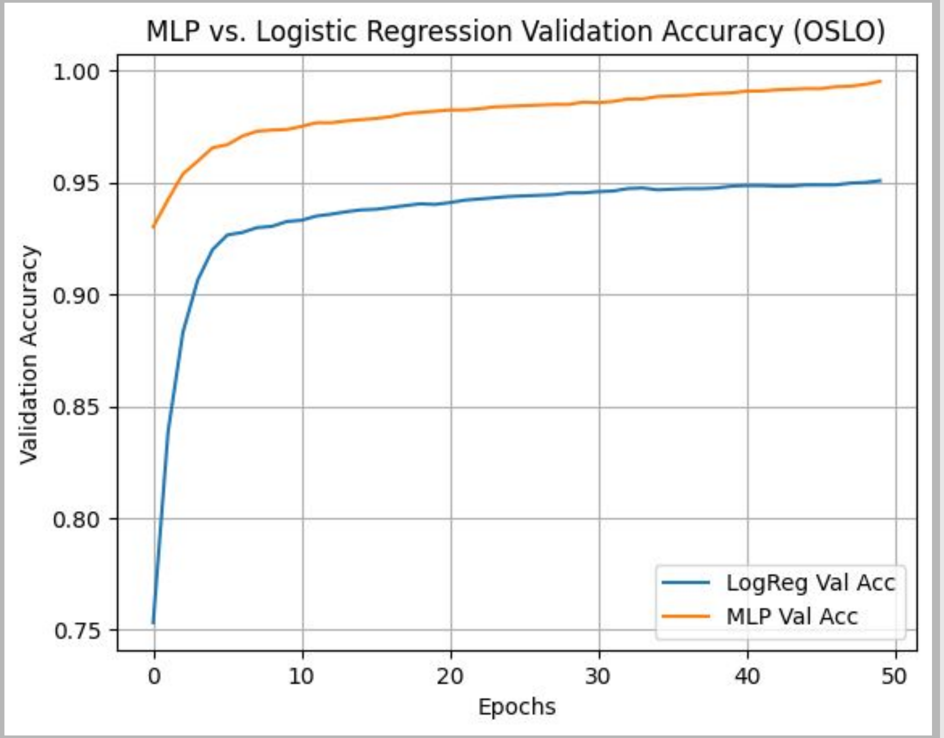
- **Tight, concentric contours** = steep, informative terrain.
- The **longer path** reflects **stronger gradients and greater “depth”**.
- Efficient movement toward the center: a much lower loss (~0.26)

Classification Model #1: 🤖 **Artificial Neural Network**





Now, predicting “pleasant” vs. “unpleasant” — categorical output.



Tested two ANN types:
- **Logistic Regression** (single-layer)
- **Multi-Layer Perceptron (MLP)**

Both optimize with gradient descent & binary cross-entropy

Evaluation via training & validation accuracy

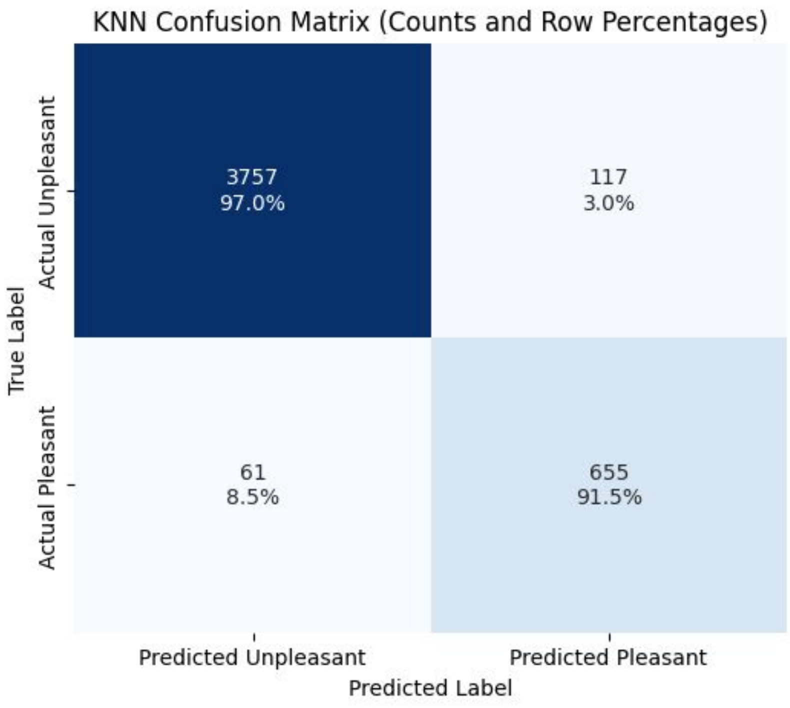
📈 **MLP starts higher, ends higher — outperforms LR in generalization**

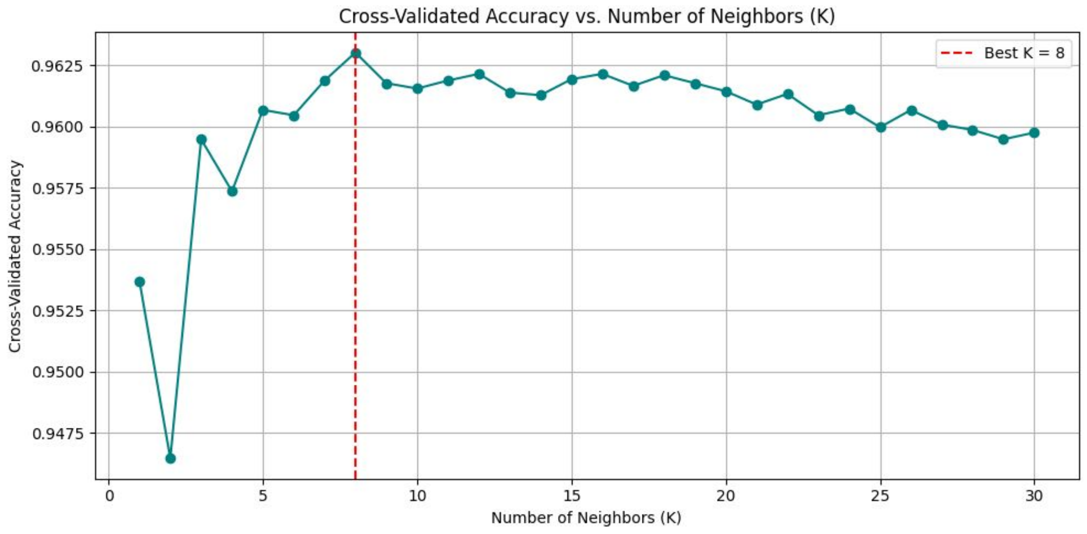
Classification Model #2: 📍 **K-Nearest Neighbor (KNN)**

KNN is an **instance-based, non-inductive model**—it predicts based on proximity, learns mainly via hyperparameter tuning (cross-validation).

KNN achieved **96% accuracy** on unseen data with optimal K=8.

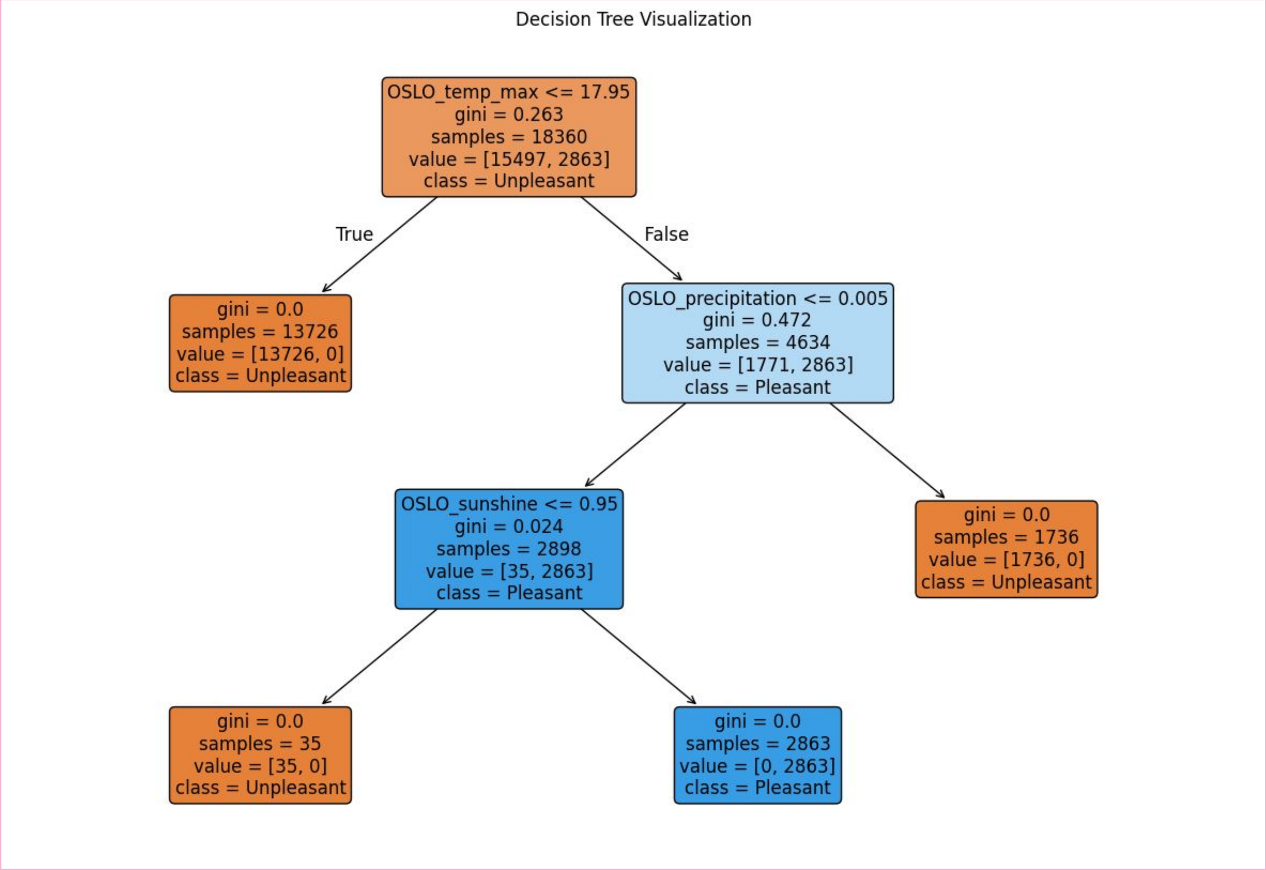
Precision & Recall:
- *Unpleasant* class: Precision 98%, Recall 97%
- *Pleasant* class: Precision 85%, Recall 91% *

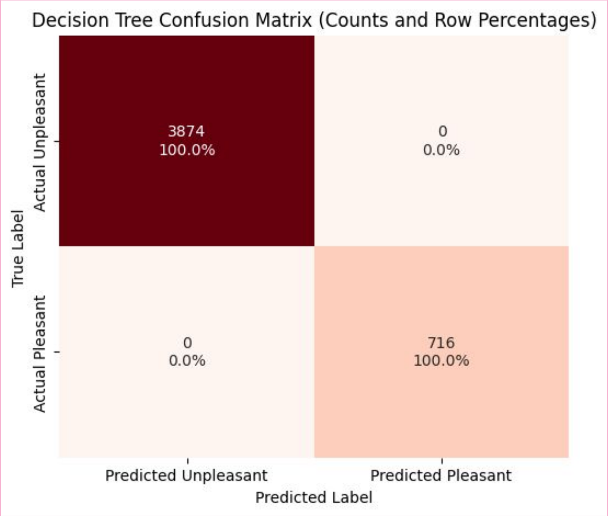




*** Precision < recall on Pleasant class reflects class imbalance—few Pleasant examples make accurate predictions harder despite good coverage.**

Classification Model #3: 🌳 **Decision Tree**





Perfect results reflect data simplicity, not overfitting
Audit confirmed **DT rules (*see below*)** exactly match original labels → **fully separable dataset**

```
|--- OSLO_temp_max <= 17.95
|   |--- class: 0
|--- OSLO_temp_max >  17.95
|   |--- OSLO_precipitation <= 0.00
|   |   |--- OSLO_sunshine <= 0.95
|   |   |   |--- class: 0
|   |   |--- OSLO_sunshine >  0.95
|   |   |   |--- class: 1
|   |--- OSLO_precipitation >  0.00
|   |   |--- class: 0
```

Final Report of Findings

1. 🌳Decision Tree (DT)
   - Best overall metrics: perfect accuracy and recall.
   - Ideal for this dataset because of its **deterministic, rule-based nature**.
   - **Caveat:** Future data might be less clean or more complex—need to clarify how "pleasant weather" is defined.
2. 🤖2️⃣MLP-ANN
   - Strong performance across all metrics with high accuracy and balanced recall.
   - Better suited for **more complex, real-world data patterns** due to its capacity to model nonlinear relationships.
   - **Computational cost is higher** but still manageable. Good candidate if dataset complexity grows beyond simple rules.
3. 📍KNN
   - Solid accuracy and recall, but its **non-inductive, instance-based nature limits adaptability**.
   - Relies heavily on local data distribution, which might **struggle with data drift or feature noise** common in weather data.
4. 🤖1️⃣LR-ANN
   - Lowest computational cost and simplest architecture; acceptable baseline performance
   - Could be a **cost-effective choice for fast prototyping or where interpretability is key**, but may require **hyperparameter tuning or feature engineering** to boost performance.

Metric	🤖1️⃣ Logistic Regression - ANN	🤖2️⃣ MLP - ANN	📍 KNN	🌳 Decision Tree
Accuracy	0.95	0.99	0.96	1.00
Pleasant Recall	0.83	0.99	0.91	1.00
Unpleasant Recall	0.97	1.00	0.97	1.00
Pleasant F1-score	0.84	0.98	0.88	1.00
Macro F1-score	0.91	0.99	0.93	1.00

Climate_ Wins

Final Recommendations

Interim Report
Amy Zhang

1. Choose the Decision Tree (🌳) for this task and dataset—it's accurate and interpretable.

2. Subset data by region to mitigate geographic and feature biases, and to handle subjective definitions of "pleasantness."

3. Plan for future expansion: develop more nuanced "pleasantness" criteria, broaden the dataset, and explore flexible models like MLP when complexity grows