2.3: Complex Machine Learning Models and Keras Part 2 December 31, 2024
Isaac Contreras

Random Forest and Unsupervised ML Models

Random Forest Model

Data Subset and Model Setup

We began by reducing the data to a **single decade (2010–2019)** to keep runtime manageable. A **RandomForestClassifier** was then instantiated with:

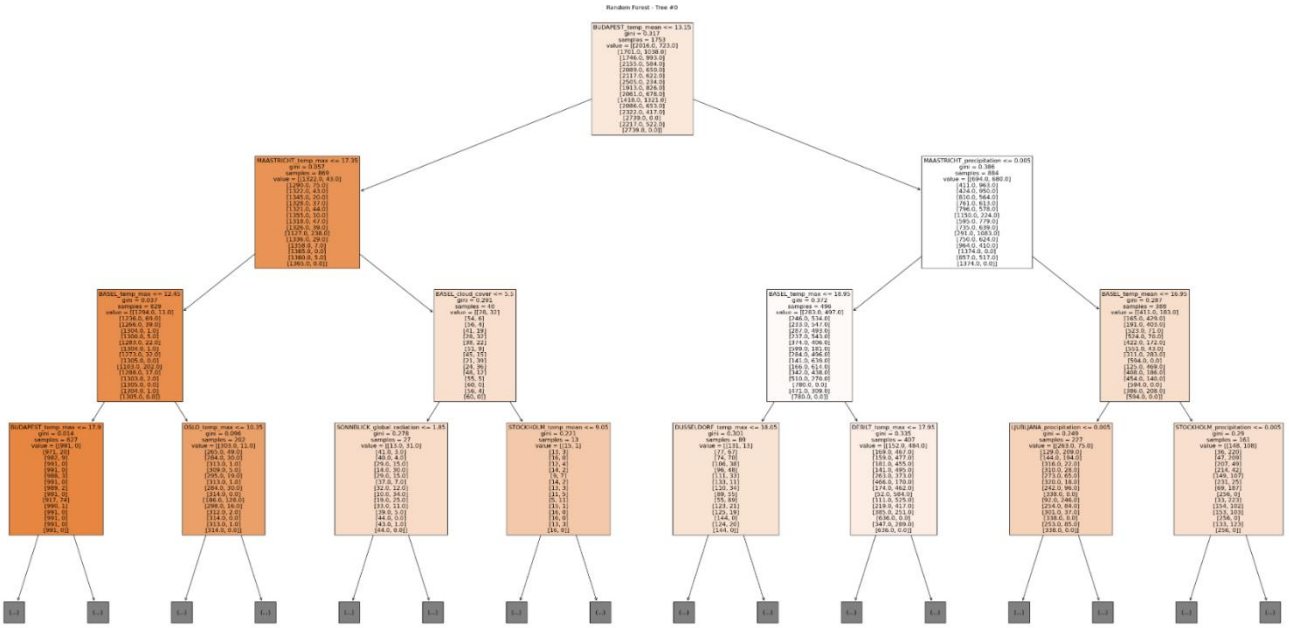
- n_estimators = 100
- random_state = 42 (for reproducibility)
- All other parameters set to defaults (i.e., the maximum depth was automatic).

Unlike some approaches that reduce the target y to a single class (e.g., np.argmax()), we kept **all 15 target columns** in a multi-label format. This naturally leads to a more complex decision tree.

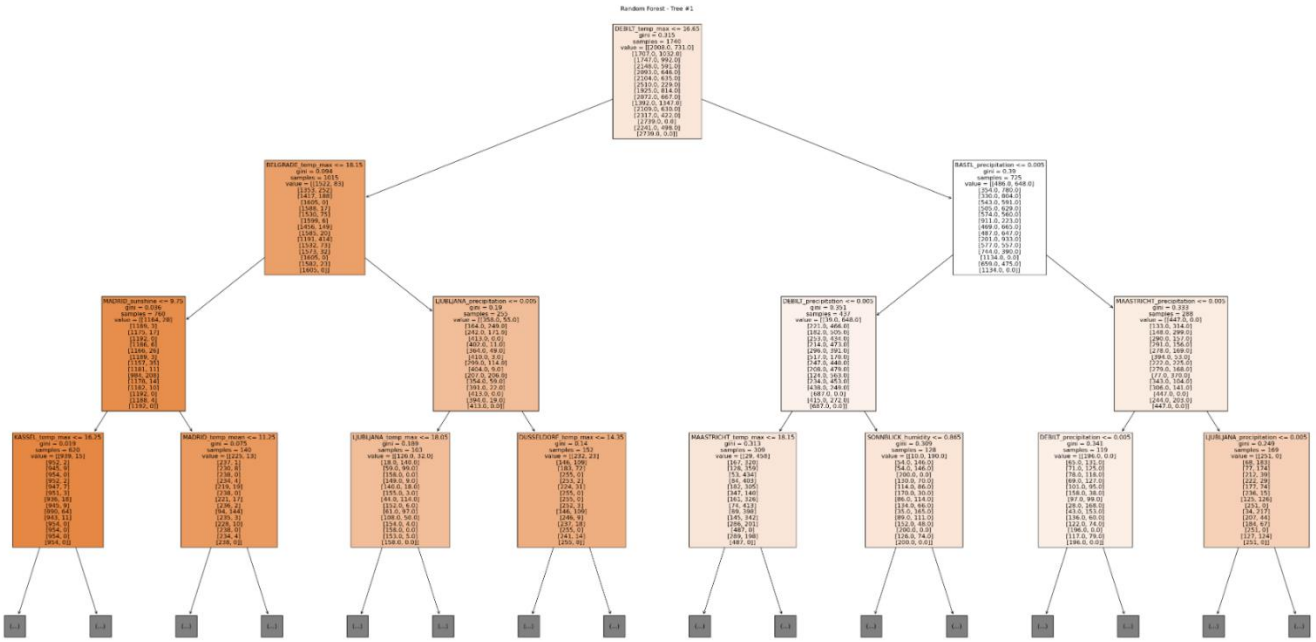
Accuracy and Key Differences

Model Accuracy on the decade data reached **59.1%**

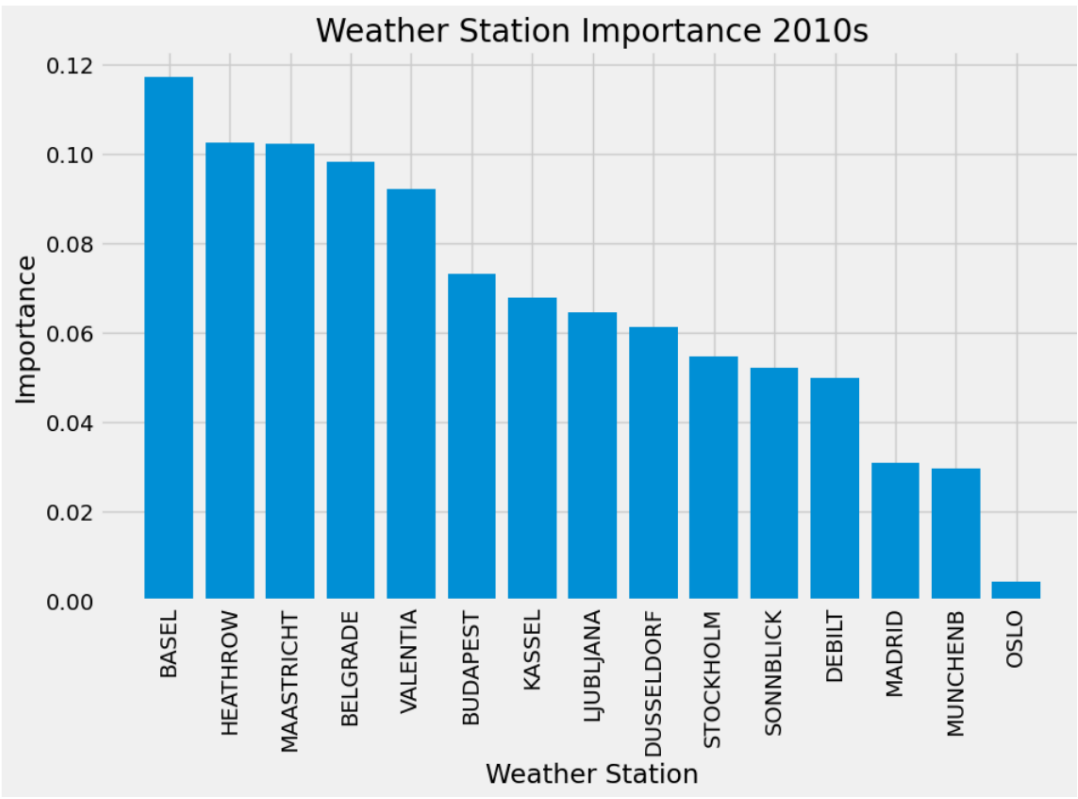
Tree #0



Tree #1



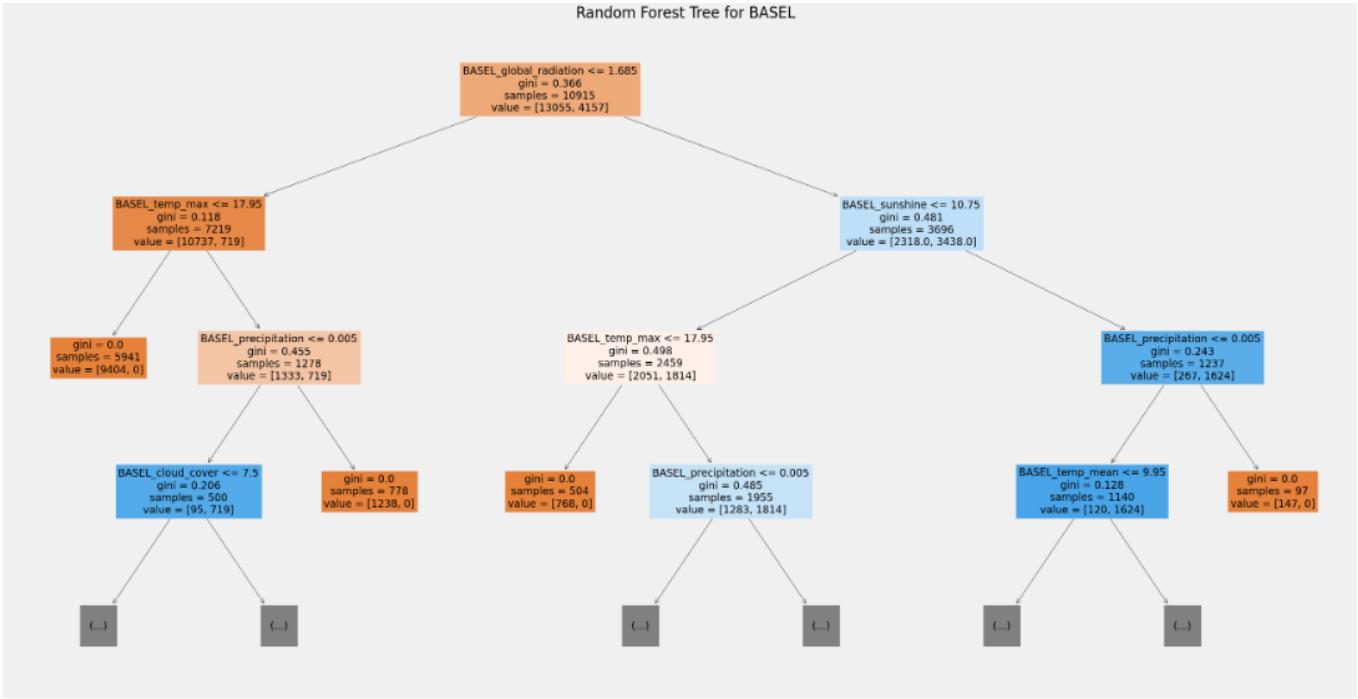
The 3 most important stations are Basel, Heathrow, and Maastricht.

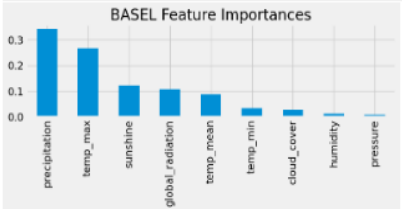


Individual Models for the Top 3 Stations

After identifying Basel, Maastricht, and Heathrow as most important, **we built separate random forest models** using the entire dataset (all years, 1960–2019) for each of these three stations.

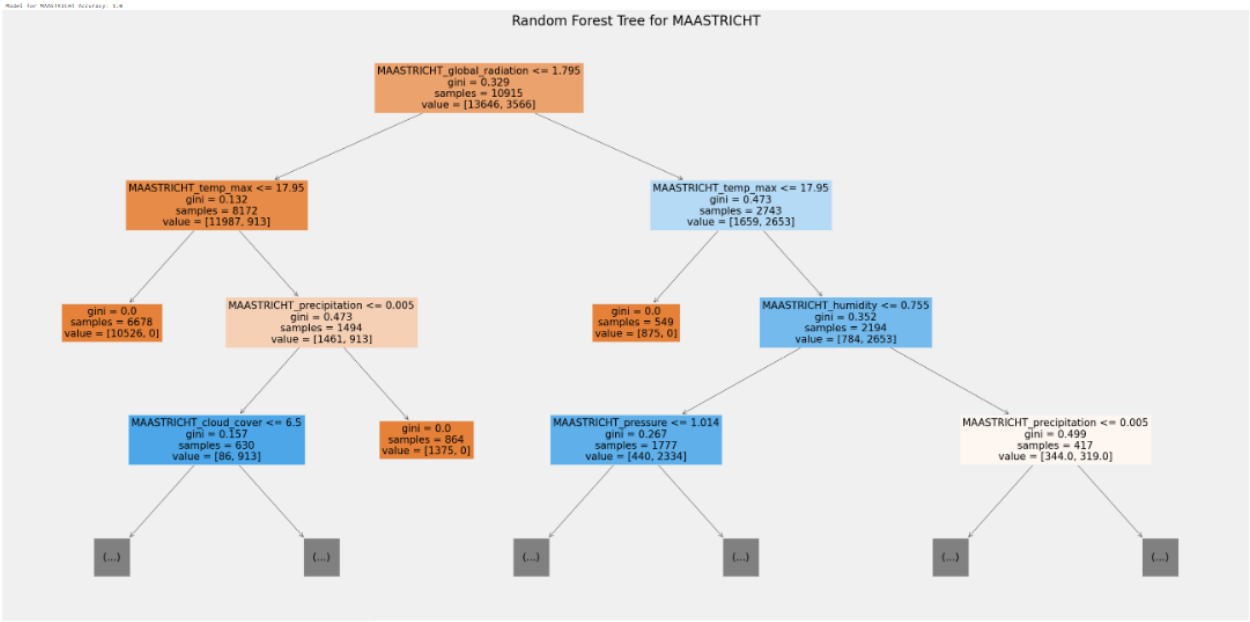
1. **Basel**

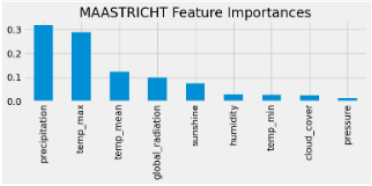




   - **Accuracy**: 100%
   - **Top 3 Features** (in descending importance):
     1. Precipitation
     2. Temp_max
     3. Sunshine

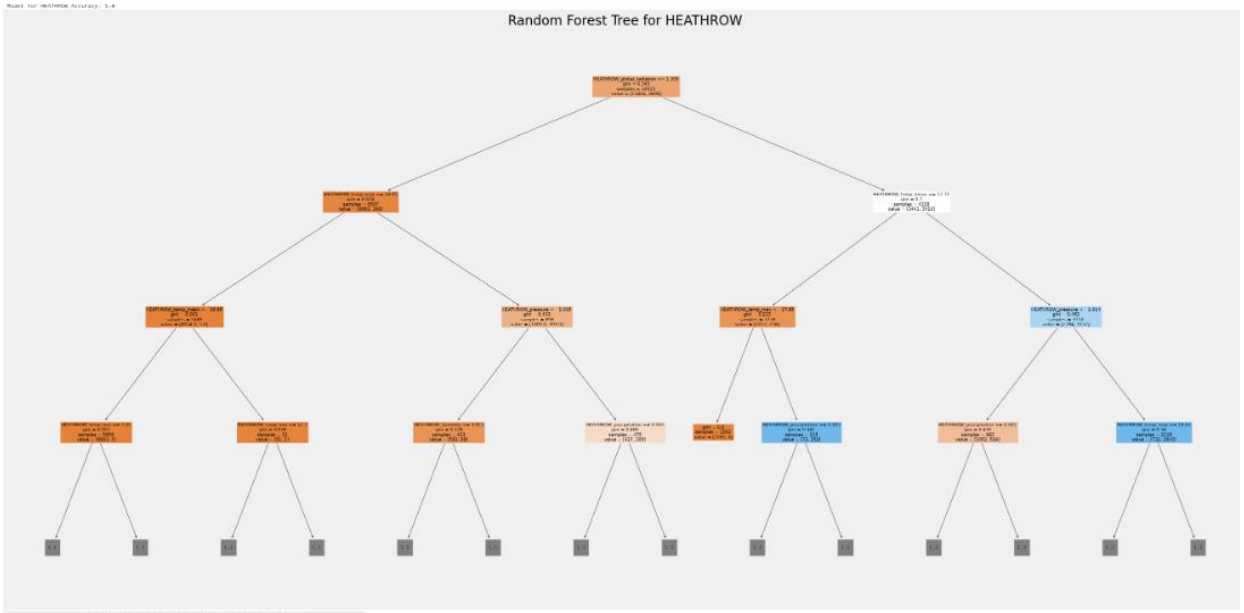
2. **Maastricht**

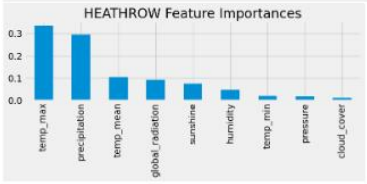




- **Accuracy**: 100%
- **Top 3 Features** (in descending importance):
    1. Precipitation
    2. Temp_max
    3. Temp_mean

3. **Heathrow**





- **Accuracy**: 100%
- **Top 3 Features** (in descending importance):
    1. Temp_max
    2. Precipitation
    3. Temp_mean

Conclusions

The consistent importance of **precipitation** and **temp_max** (and often **temp_mean**) across these three weather stations suggests that these indicators will play a central role in predicting and adapting to future climate variability. As global temperatures continue to rise, we can expect:

- More frequent and extreme **max temperature** events (e.g., heatwaves)
- Possible intensification or shifts in **precipitation** patterns across regions

Given that precipitation and temperature extremes profoundly affect local weather outcomes—floods, droughts, storms, agriculture, etc.—it is vital to maintain robust **climate-tracking systems** focused on

these indicators. Monitoring temperature and precipitation will help both researchers and policymakers anticipate changes and plan for a more variable climate future.

In short, **precipitation** and **temperature** (particularly **temp_max**) stand out as powerful predictors of "pleasant days." This underscores the value of **investing in measurement instruments** and **modeling capabilities** that can pinpoint how these two drivers continue to shape weather outcomes, especially in a warming global climate.

Report Highlights

- **Decade Model**: 59.1% accuracy (multi-label approach with 15 columns).
- **Top 3 Stations**: Basel, Maastricht, Heathrow.
- **Single-Station Models**: 100% accuracy for each; consistent top indicators are precipitation and temperature measures.
- **Future Implications**: Temperature and precipitation measurements are key to anticipating and mitigating climate impacts.