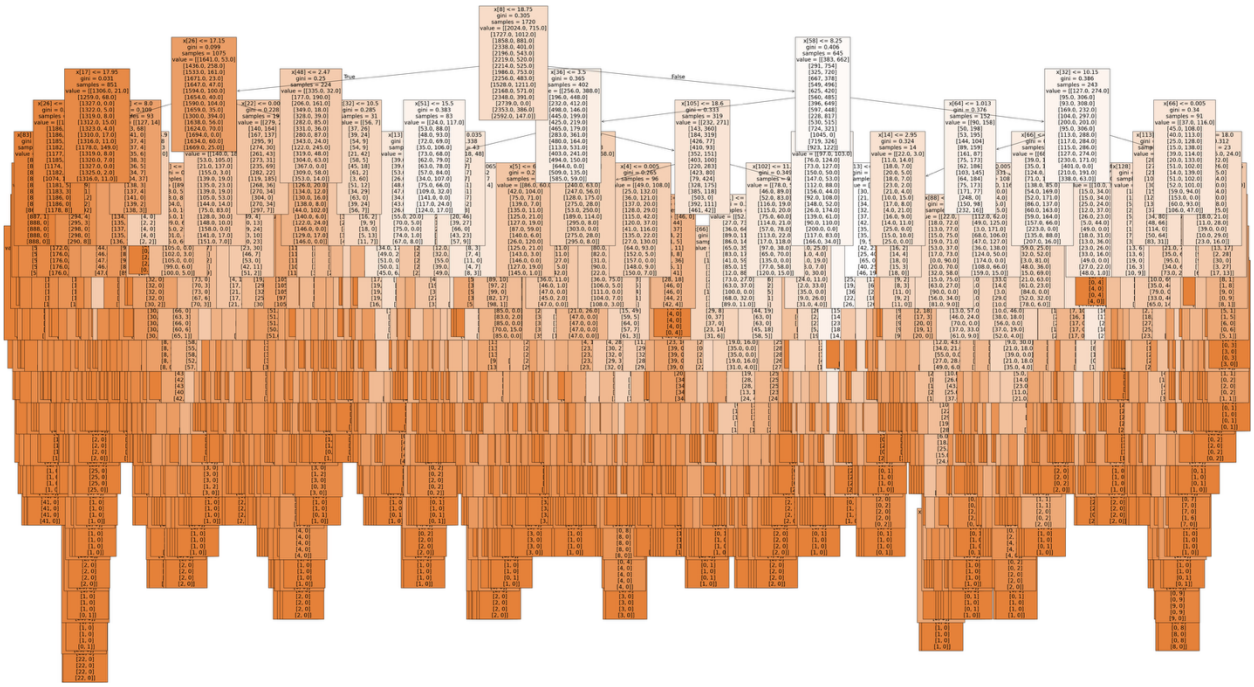
Complex Machine Learning Models and Keras Part 2

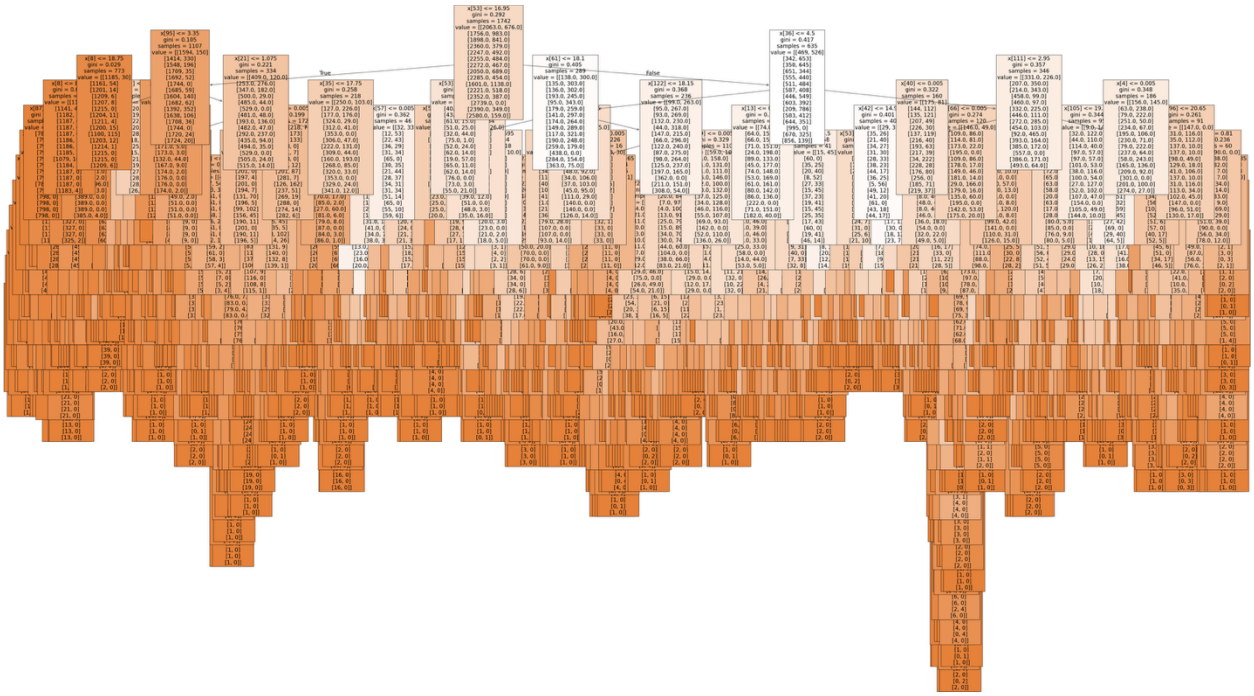
Random Forest Model – All 15 stations, 1960-1969

Model Accuracy: 0.606

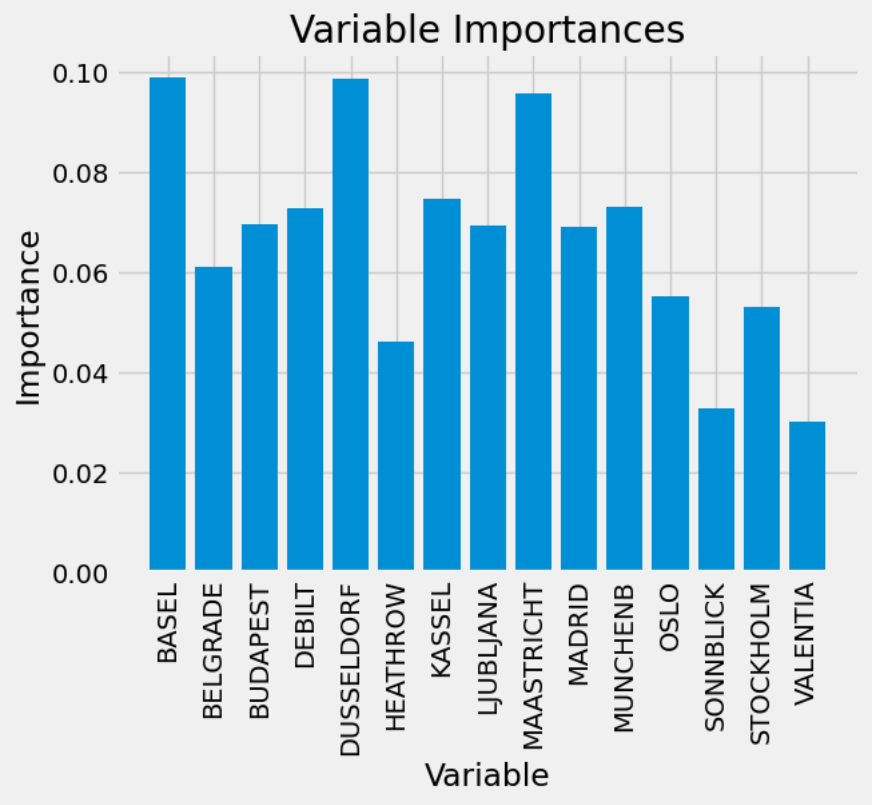
Plot of Tree 1:



Plot of Tree 2:



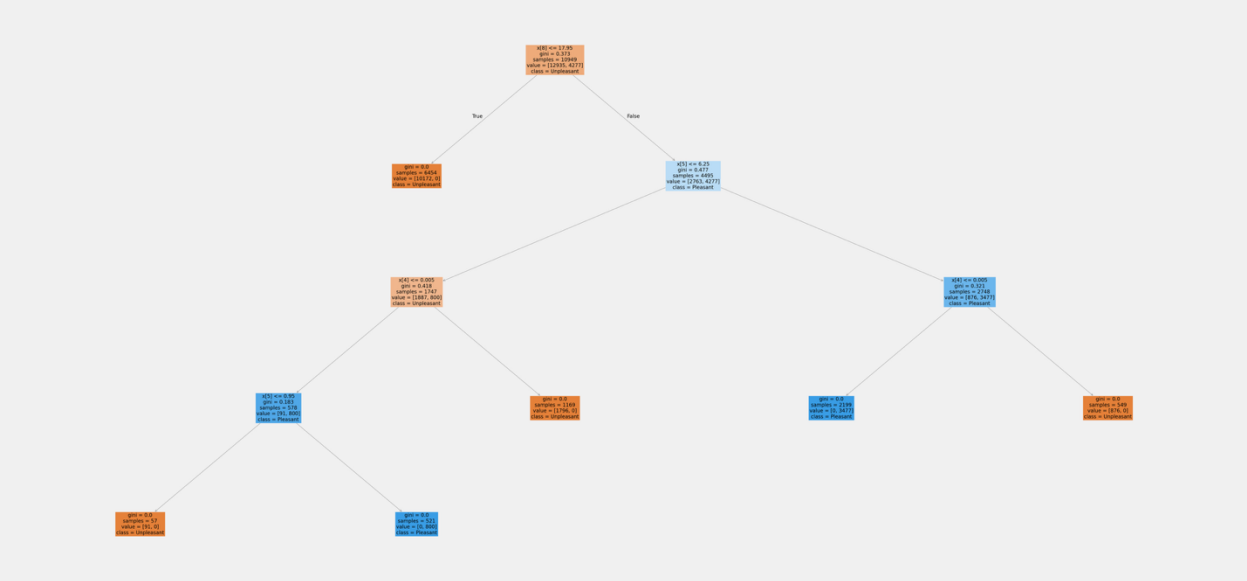
Bar chart of importances (stations)



Top 3 stations in importance: Basel, Dusseldorf, Maastricht

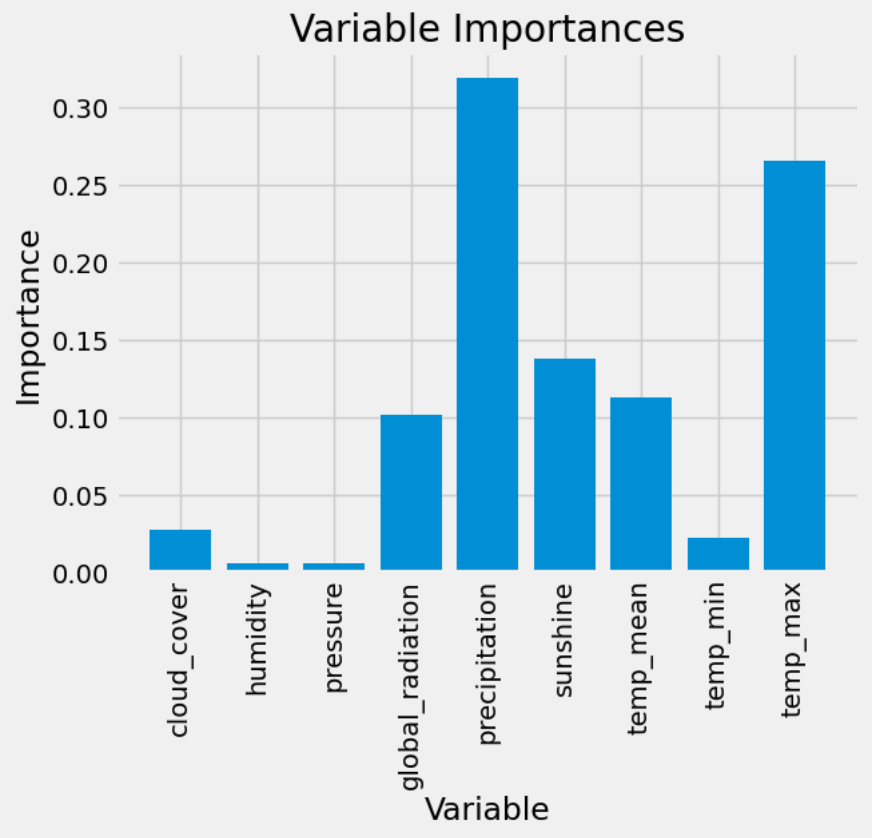
Random Forest Model – BASEL Station, 1960-2022

Plot of Tree 1:



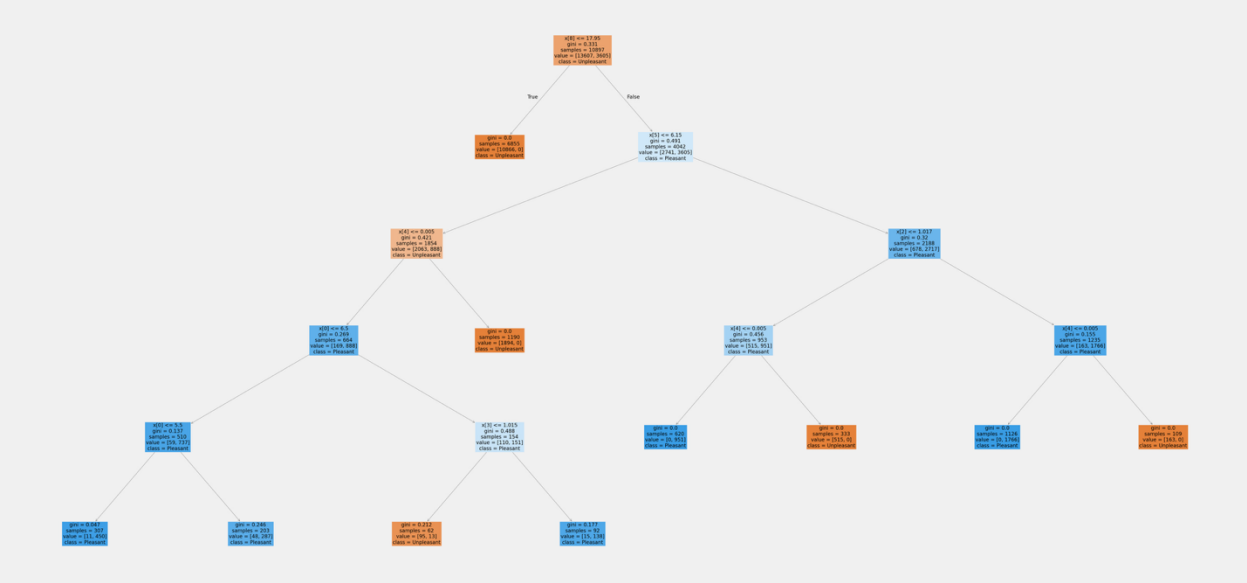
Model Accuracy: 1.0

Bar chart of importances (observations):



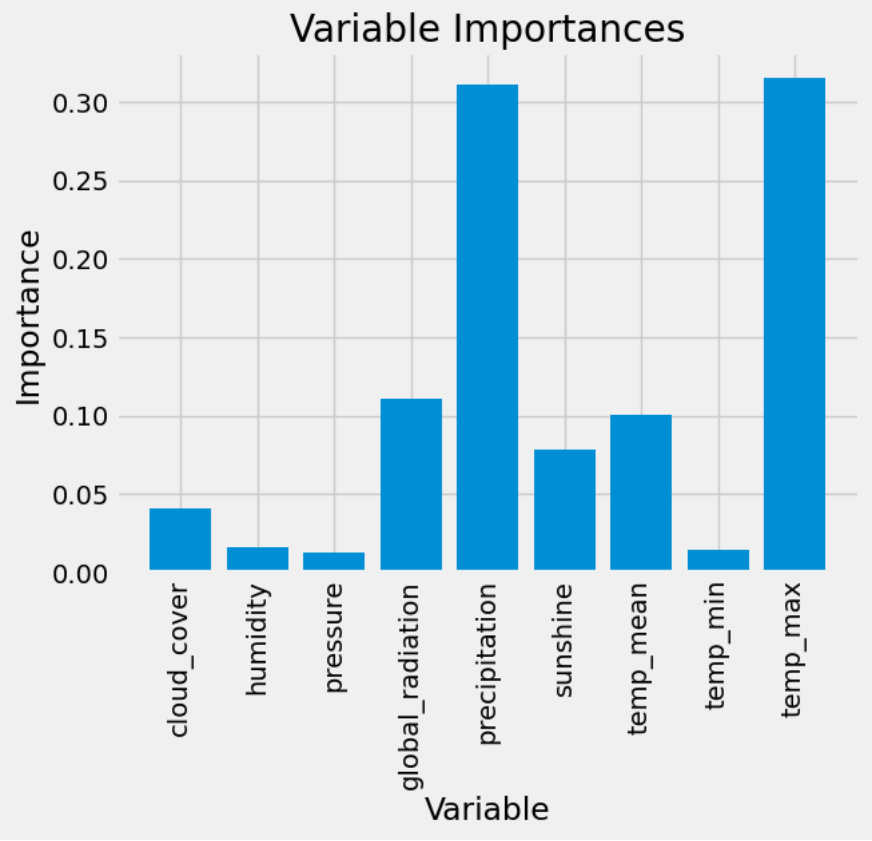
Random Forest Model – DUSSELDORF Station, 1960-2022

Plot of Tree 1:



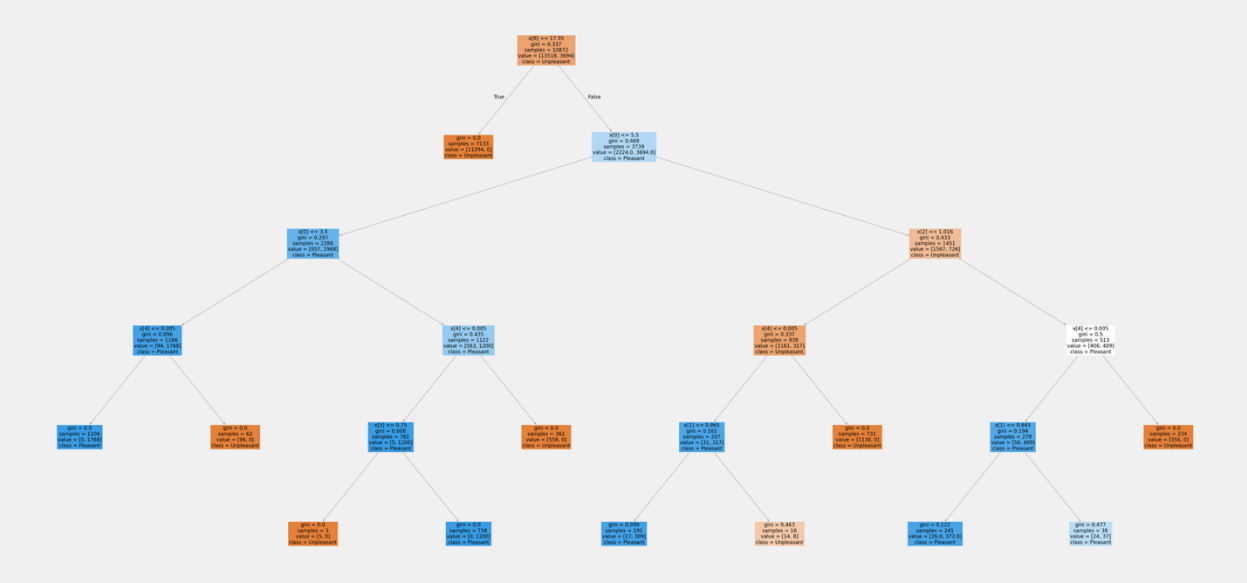
Model Accuracy: 1.0

Bar chart of importances (observations):



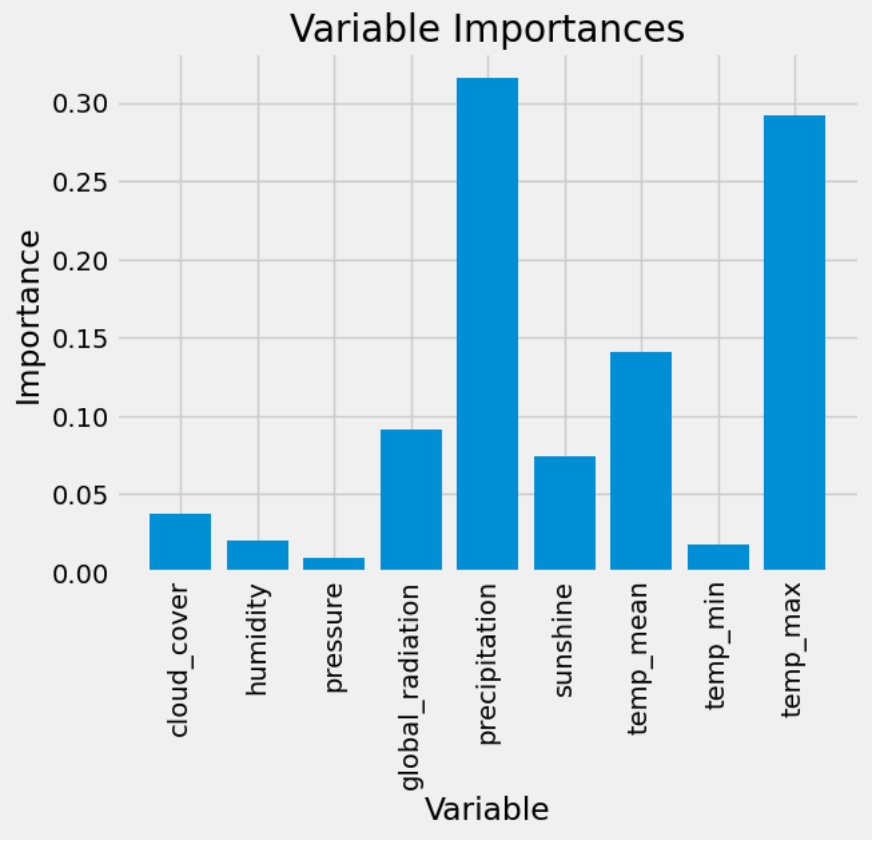
Random Forest Model – MAASTRICHT Station, 1960-2022

Plot of Tree 1:



Model Accuracy: 1.0

Bar chart of importances (observations):



Comparison chart of accuracies and importances

Data subset	Model Accuracy	Top 3 Importances
All stations, 1960 - 1969	0.606	Basel Station Dusseldorf Station Maastricht Station
Basel, 1960 - 2022	1.0	Precipitation Temp_max Sunshine
Dusseldorf, 1960 - 2022	1.0	Temp_max Precipitation Global_radiation
Maastricht, 1960 - 2022	1.0	Precipitation Temp_max Temp_mean

Conclusions

When running the model on all the stations in one decade, the feature importance chart shows us that the three stations **Basel, Dusseldorf** and **Maastricht** have the most influence on the outcomes of the model. But we must consider that the model is not very accurate, and these three stations are in very similar geographical regions (Maastricht is very close to Dusseldorf). If the model is being influenced by these stations more than others, it may not be able to predict accurately for regions that are farther north or farther south.

When we look at each of these three important stations and run the random forest model on them, we find that it predicts at 100%. In evaluating the top observations that influence these very accurate predictions, we see that **precipitation** and **maximum temperature** are the leading indicators.

In summary, tracking maximum temperature and precipitation will be important for predicting future weather events. But as the model may be overfitting to certain stations, we should be wary of using the same model and importances when predicting weather in contrasting regions.