Kyle Stanford

Complex Machine Learning Models and Keras pt. 2

The following Random Forest Model was trained and tested on data from the year 2000 to 2009 in the weather data from ClimateWins. It used 100 classifiers with a max depth of 10. Its accuracy was 95.295%:

```
# Create Random Forest classifier
clf = RandomForestClassifier(n_estimators = 100, max_depth = 10)

# Training the model on the training dataset
# Fit function is used to train the model using the training sets as parameters
clf.fit(X_train, y_train)
```

```
RandomForestClassifier(max_depth=10)
```

```
# Performing predictions on the test dataset
y_pred = clf.predict(X_test)

# Using metrics module for accuracy calculation
print("Model Accuracy: ", (metrics.accuracy_score(y_test, y_pred))*100,'%')
```

```
Model Accuracy:  95.29540481400439 %
```

Figure 1 - Random Forest Model with Accuracy Score (95.30%)

The following trees came from the model:

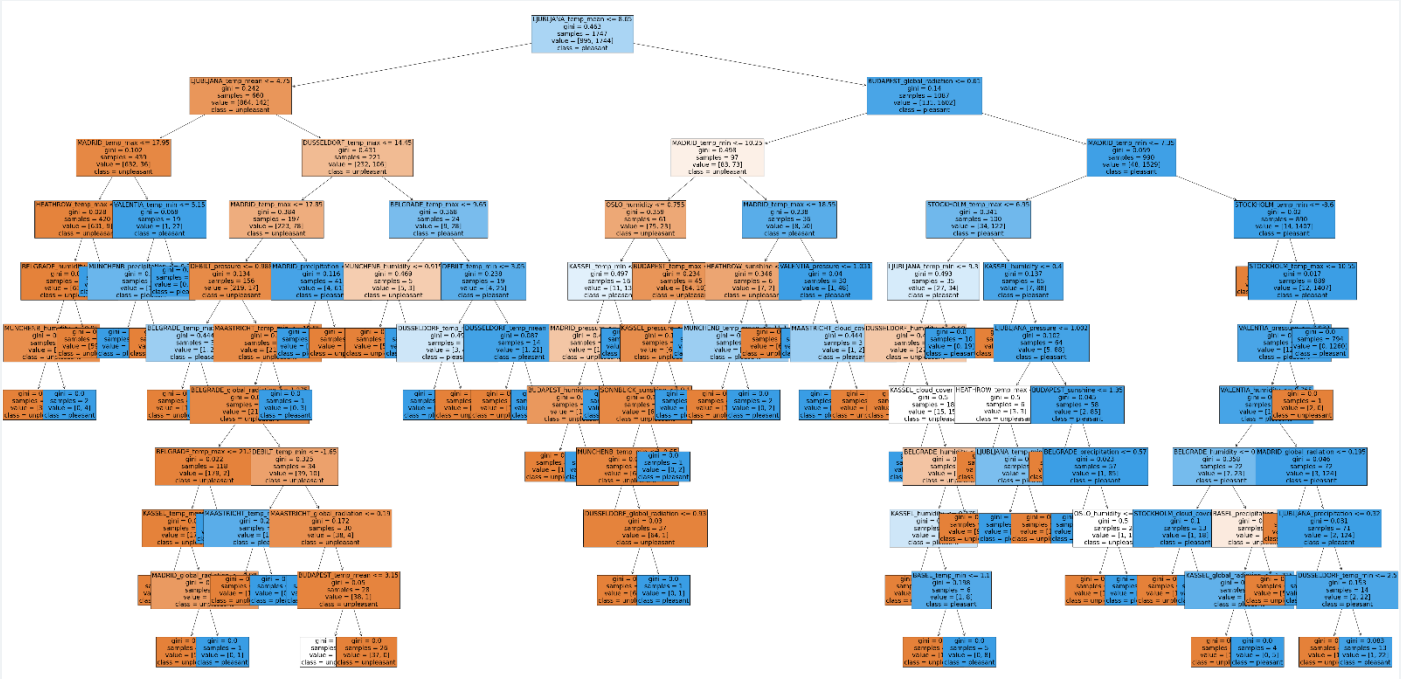

Figure 2 - Decision Tree 9 from Random Forest Classifier

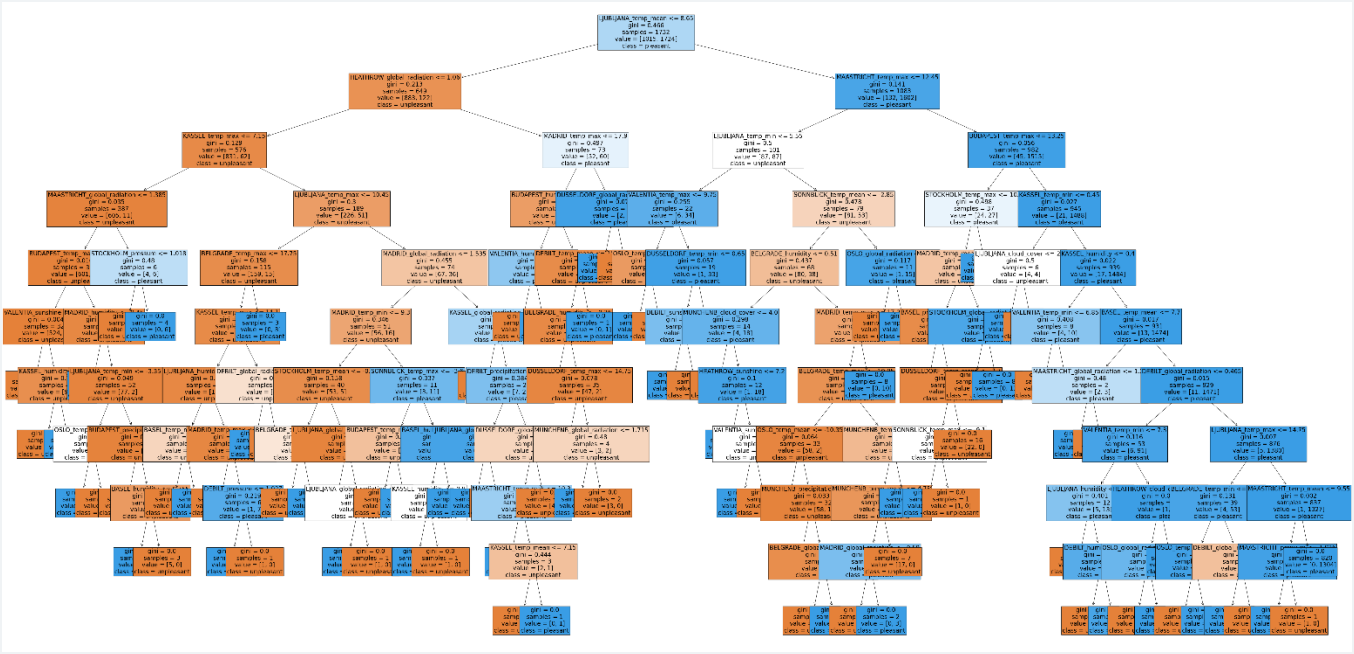

Figure 3 - Decision Tree 18 from Random Forest Classifier

Feature Importances

The 3 weather stations that contributed most significantly to the classification of the data by the Random Forest were **Madrid**, **Budapest**, and **Ljubljana.**

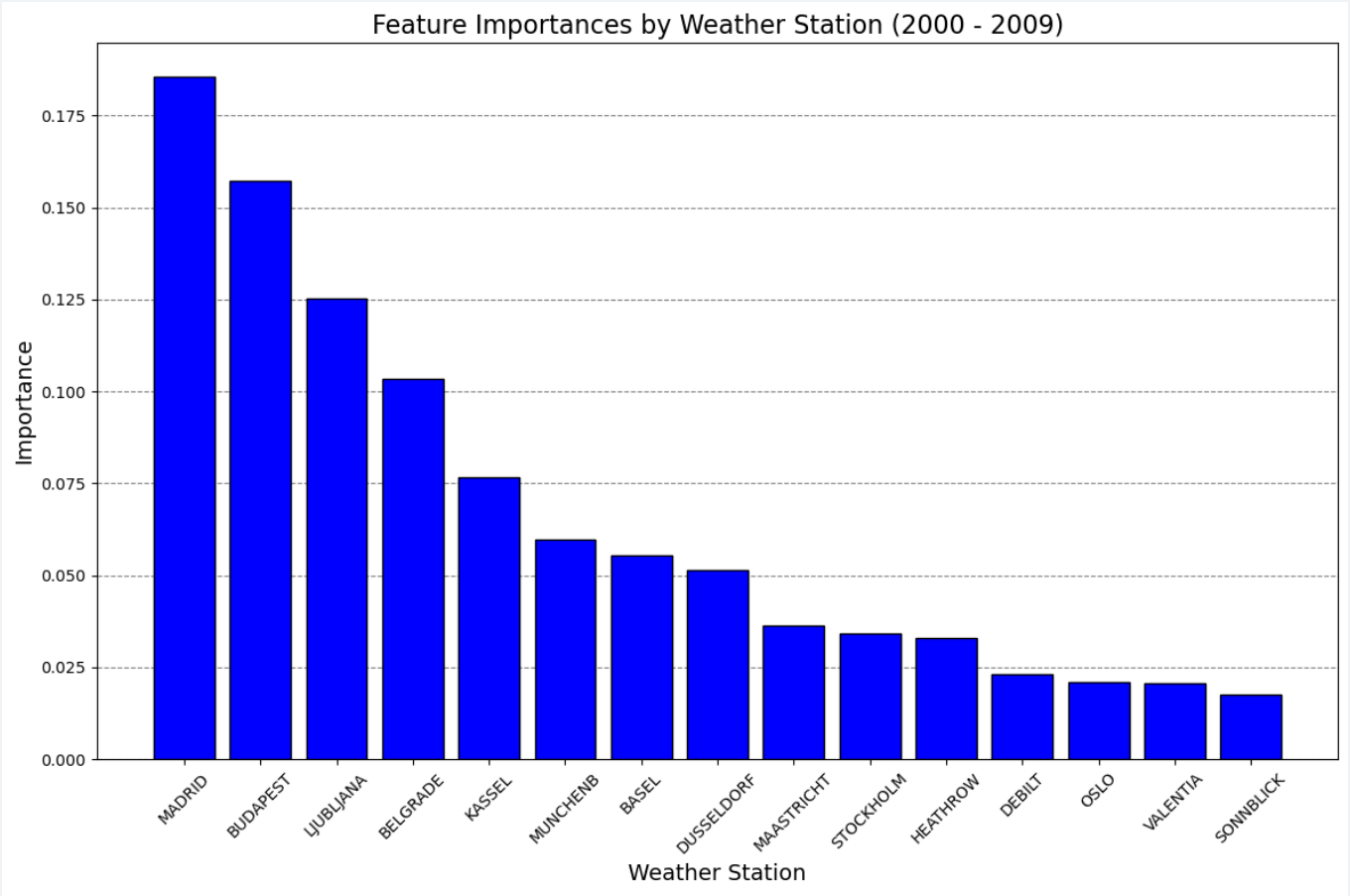

Figure 4 - Feature Importances Plot

Madrid (Accuracy 99.6%):

Madrid's weather data was put run through a random forest model to determine which measurements were most important to predicting pleasant weather. In short, the model showed that **maximum daily temperature** was the most important factor, followed by **average daily temperature**, and then by **precipitation**.

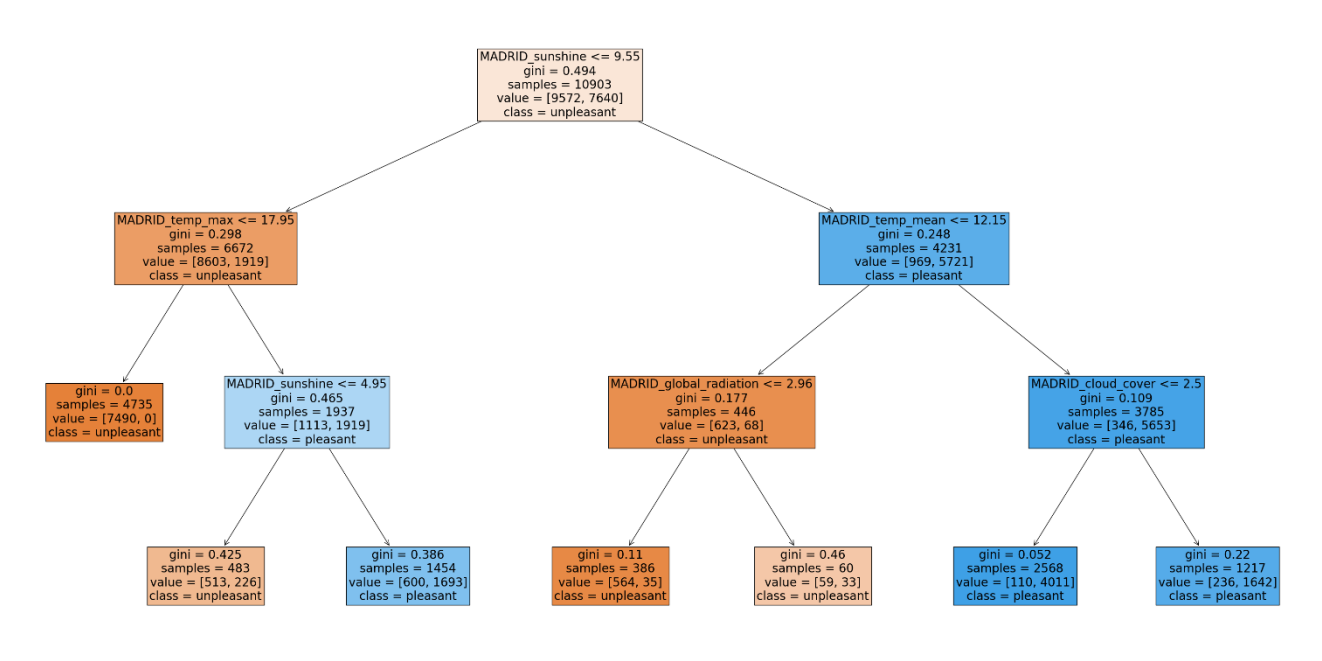

Figure 5 - Decision Tree 27 from Madrid's Random Forest Classifier

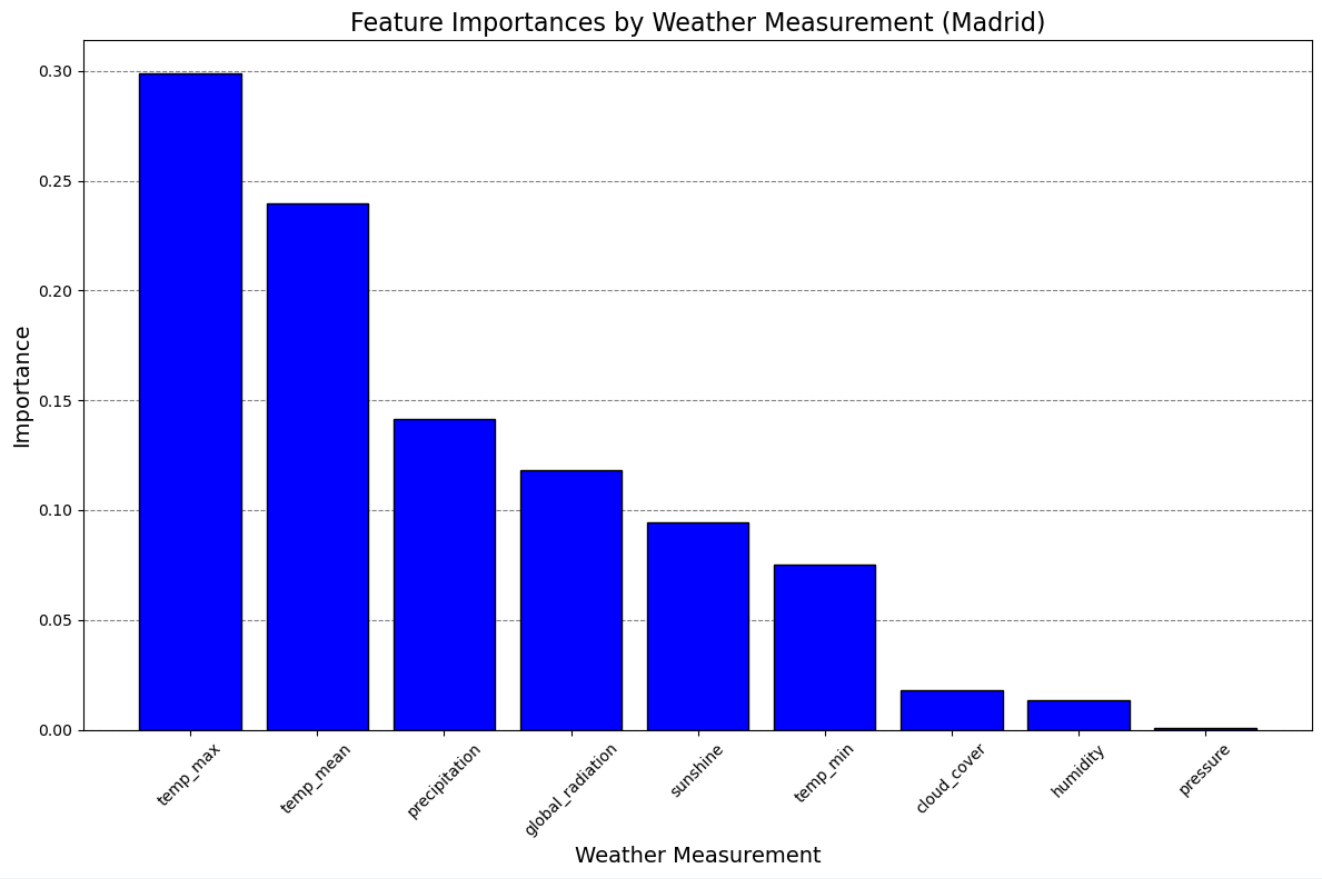

Figure 6 - Feature Importances Plot for Madrid

Budapest (Accuracy 99.8%):

Budapest's weather data was also put run through a random forest model to determine which measurements were most important to predicting pleasant weather. Similar to Madrid, **maximum daily temperature** was the most important factor, followed by **average daily temperature**, and then by **precipitation**.

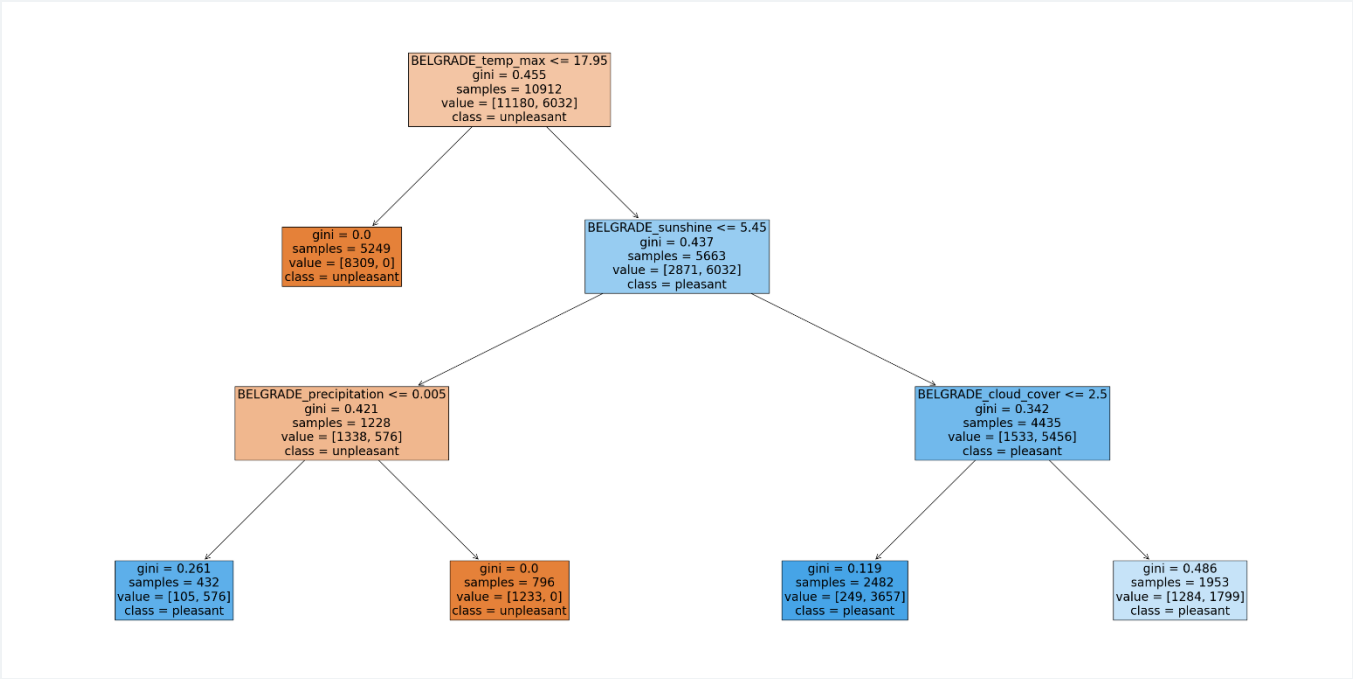

Figure 7 - Decision Tree 27 from Budapest's Random Forest Classifier

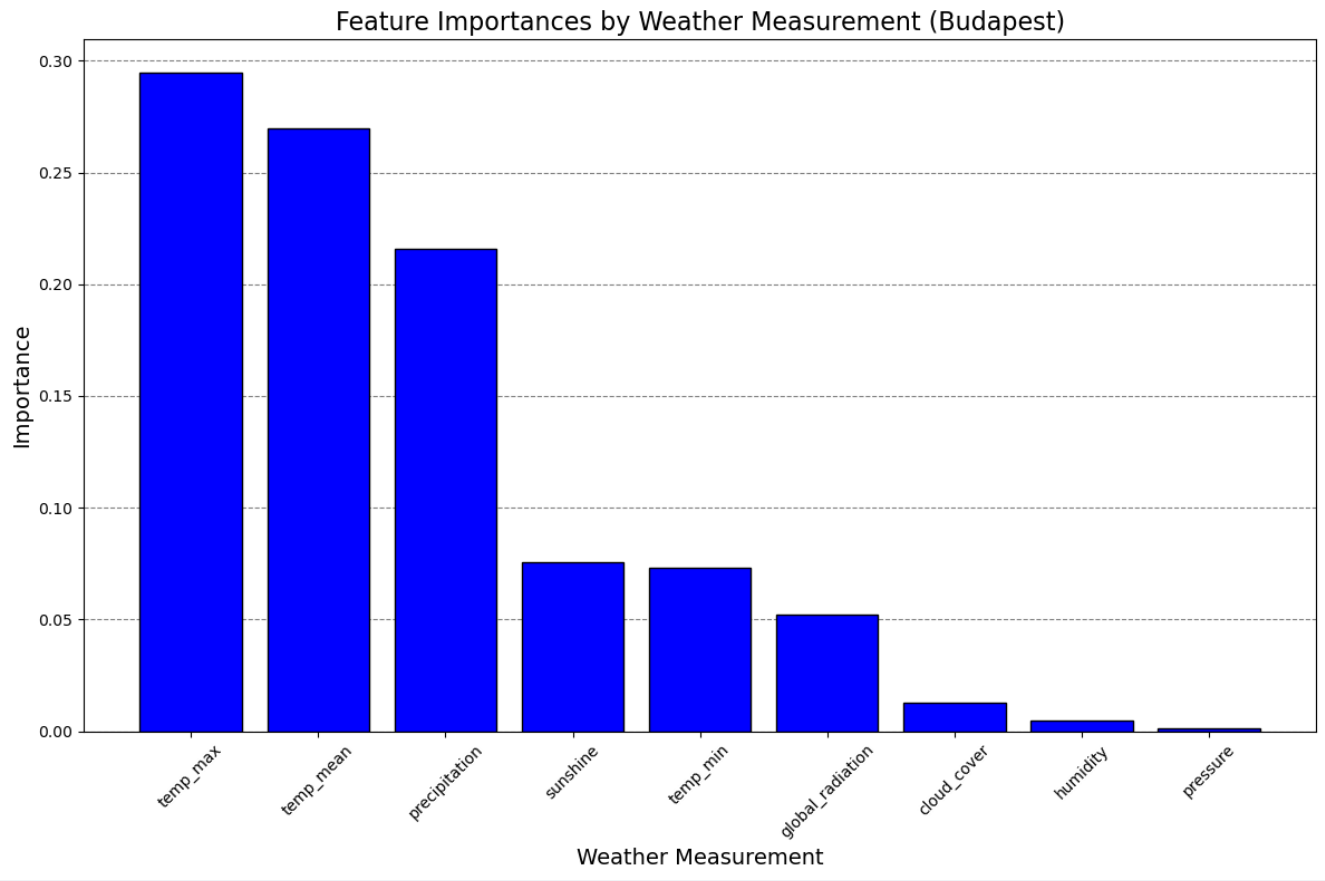

Figure 8 - Feature Importances plot for Budapest

Ljubljana (Accuracy 99.9%):

Ljubljana's weather data was put run through its own random forest model last to determine which measurements were most important to predicting pleasant weather. Similar to Madrid and Budapest, **maximum daily temperature** was the most important factor. However, the relative importances of **average daily temperature** and **precipitation** were swapped when comparing to Madrid and Budapest

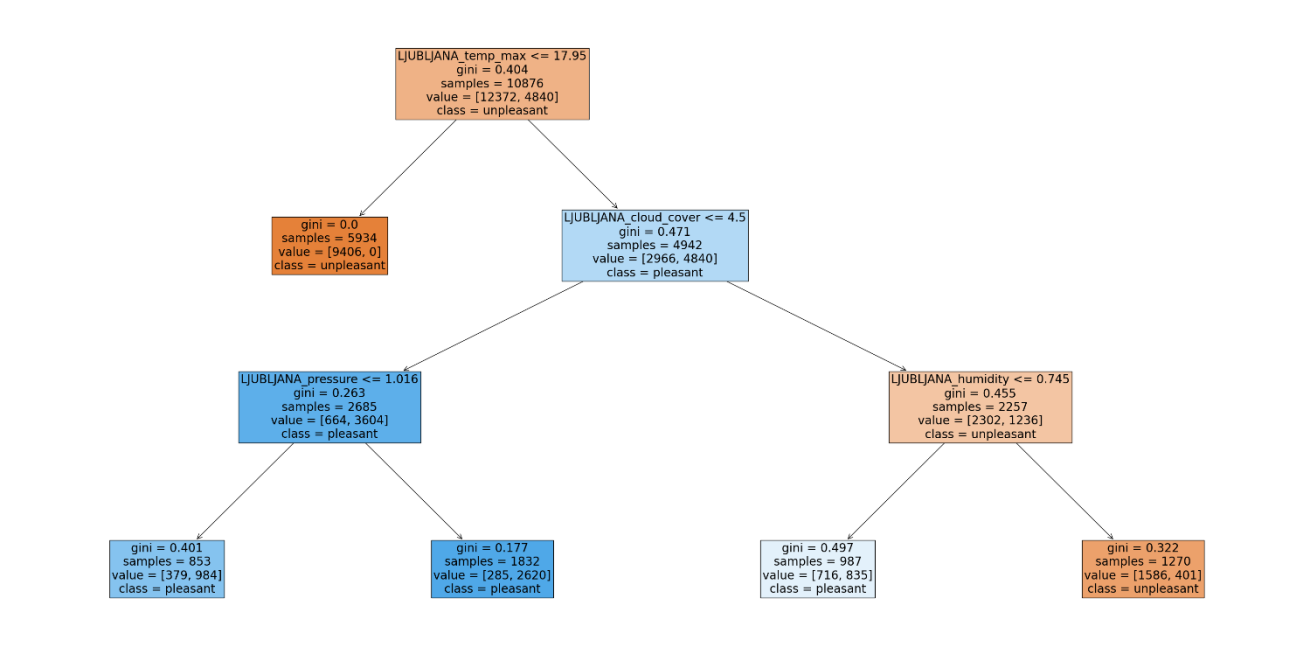

Figure 9 - Decision Tree 27 from Ljubljana's Random Forest Classifier

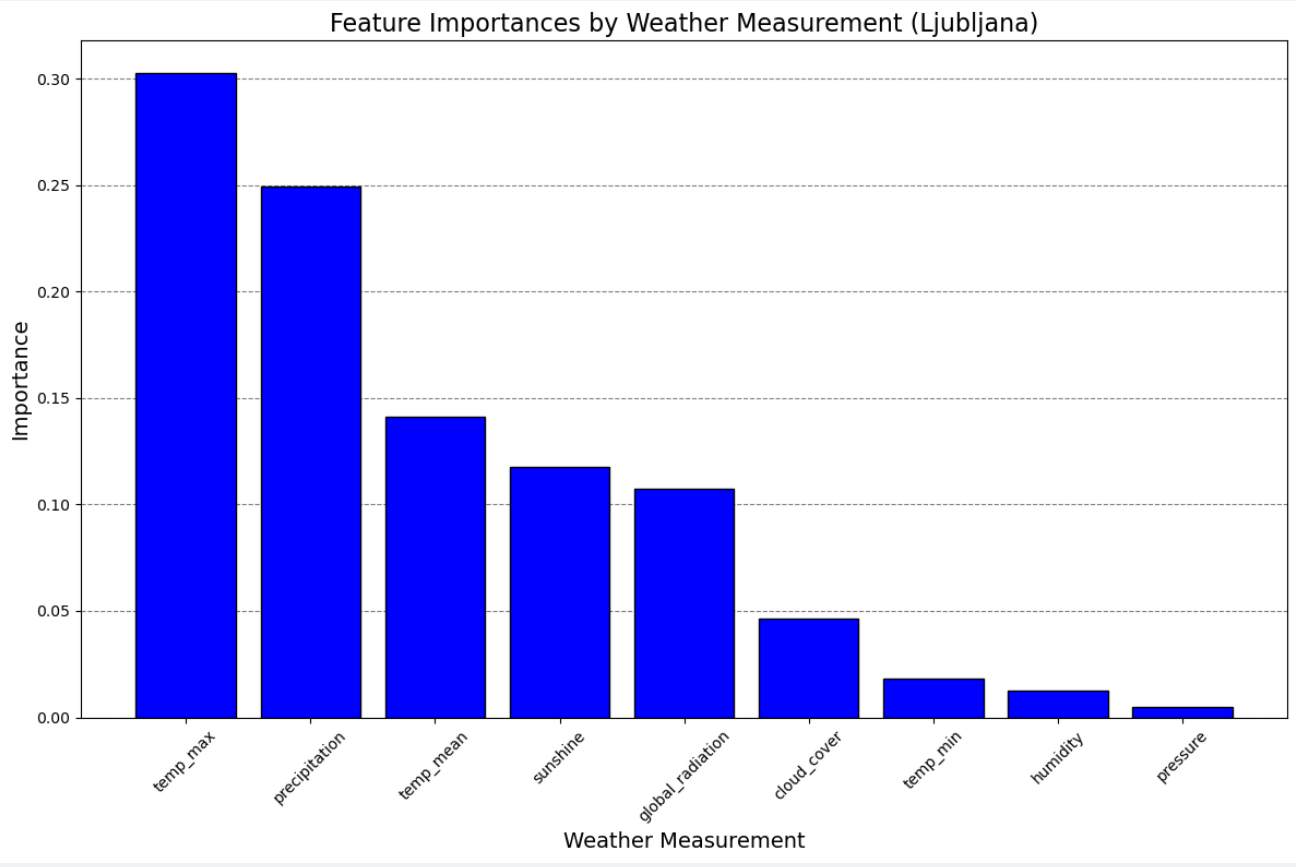

Figure 10 - Feature Importances plot for Ljubljana

Kyle Stanford

Summary

Weather Measurement's Importance to Predicting Whether

Measurement	**Madrid**	**Budapest**	**Ljubljana**
temp_max	1st	1st	1st
temp_mean	2nd	2nd	3rd
precipitation	3rd	3rd	2nd

Maximum daily temperatures, average daily temperatures, and precipitation were consistently the most significant indicators of predicting weather patterns in the analysed weather stations. Although this context, predicting picnic suitability, is somewhat trivial, the constant importance of these factors likely extends to predicting and adapting to more significant future weather events. For example, rising maximum and average temperatures is a well-studied and acknowledged result of climate change. This trend was also previously observed within ClimateWins own weather data. Consequently, focusing on temperature data will likely become even more critical, alongside precipitation, in predicting climate patterns.

Therefore, investing in equipment to better monitor temperature and precipitation should be a top priority for ClimateWins, especially at Madrid, Budapest, and Ljubljana. The impact of more accurate and detailed data relating to temperature and precipitation will be crucial to predicting and preparing for shifting weather patterns and events.