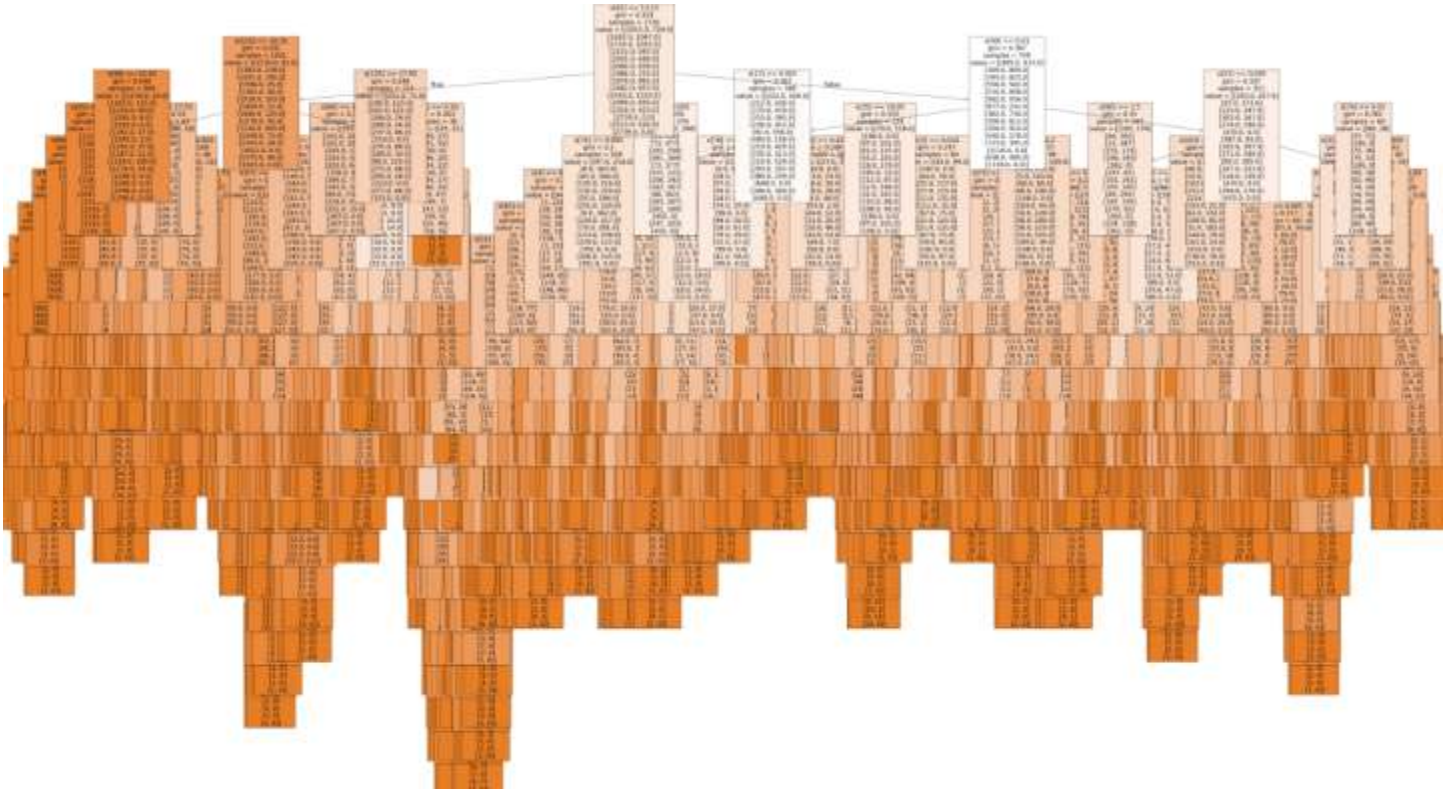
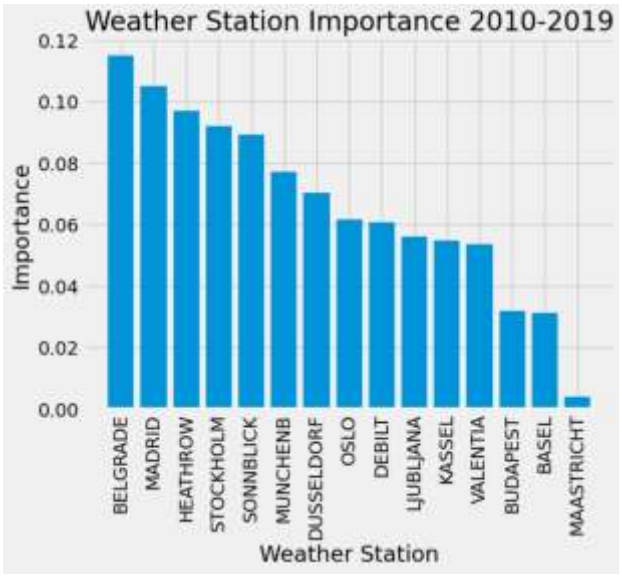
COMPLEX ML & KERAS PART 2

CASE STUDY: CLIMATE DATASET
EXERCISE 2-3

Random Forest Model:

Only data from 2010 to 2019 was used for this study. Using number of estimators at 100, and auto depth, we got an accuracy of 58%. The decision tree in this study was quite complex.

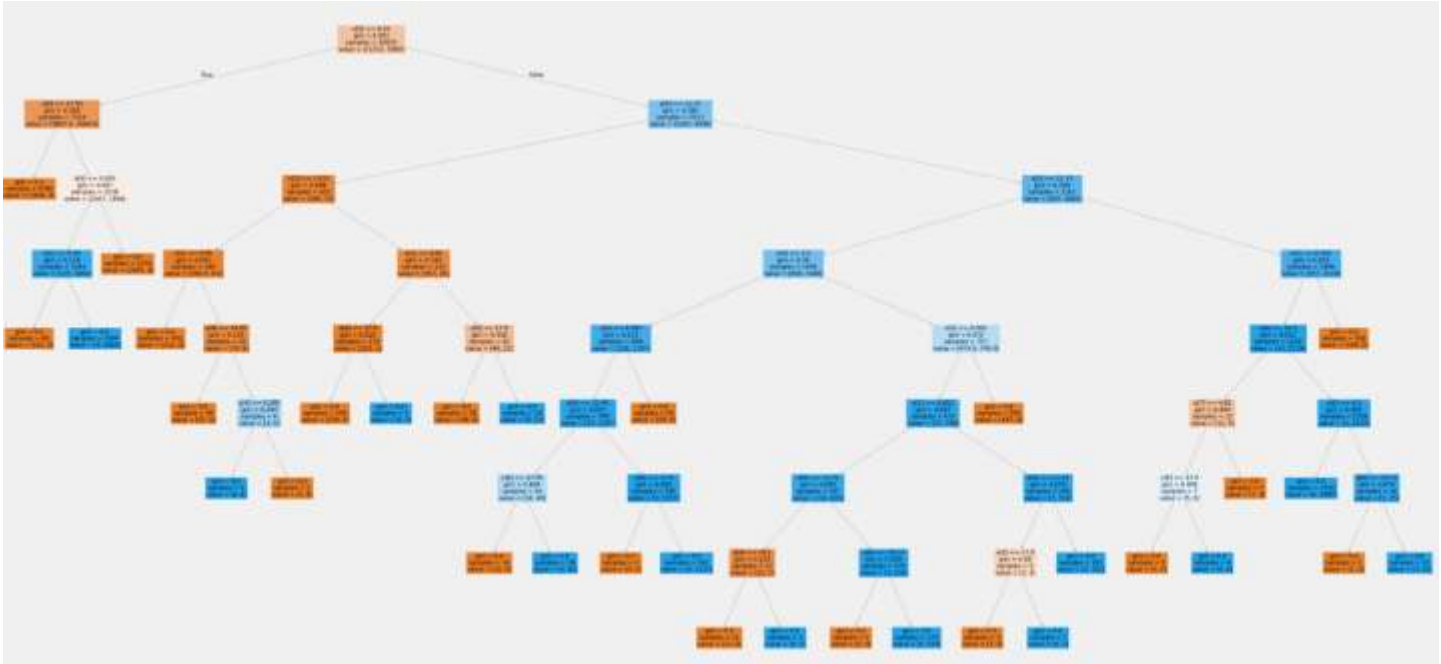


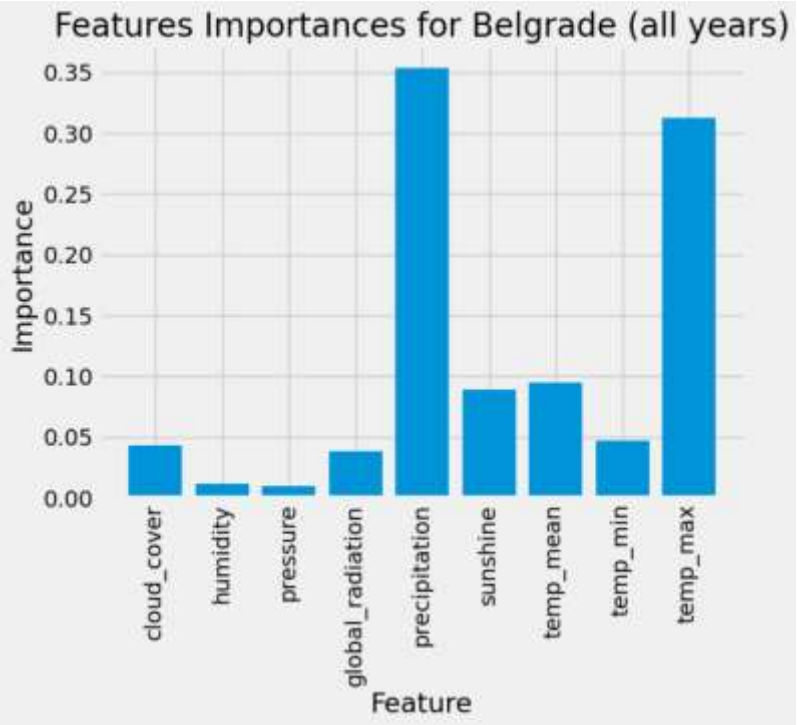


Belgrade, Madrid and Heathrow showed to be the most important features

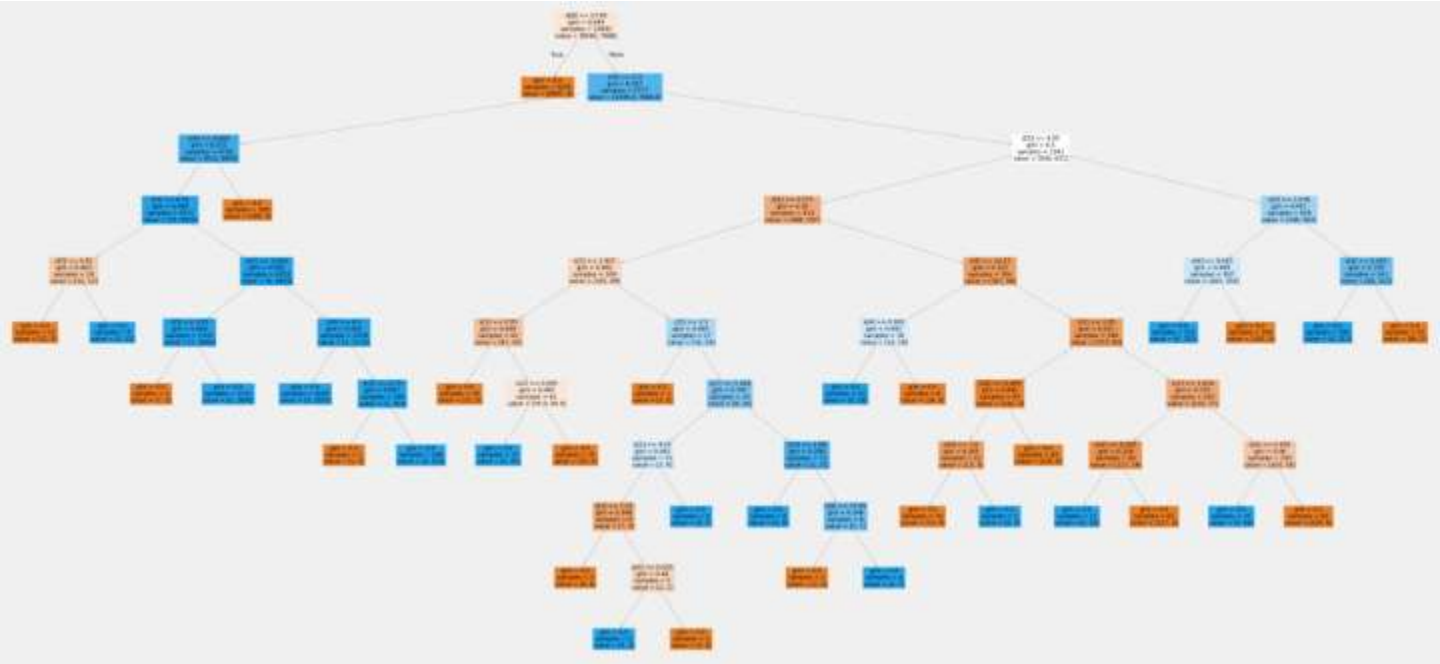
Random Forest Models for the 3 most important features are as below:

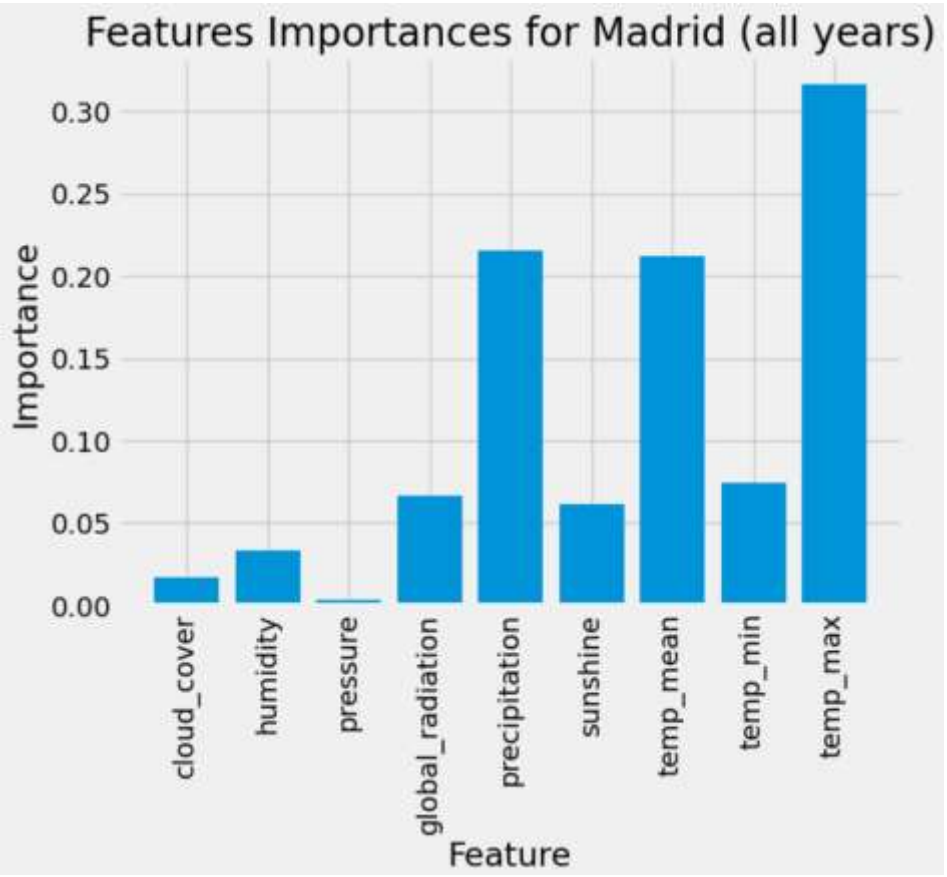
1. Belgrade



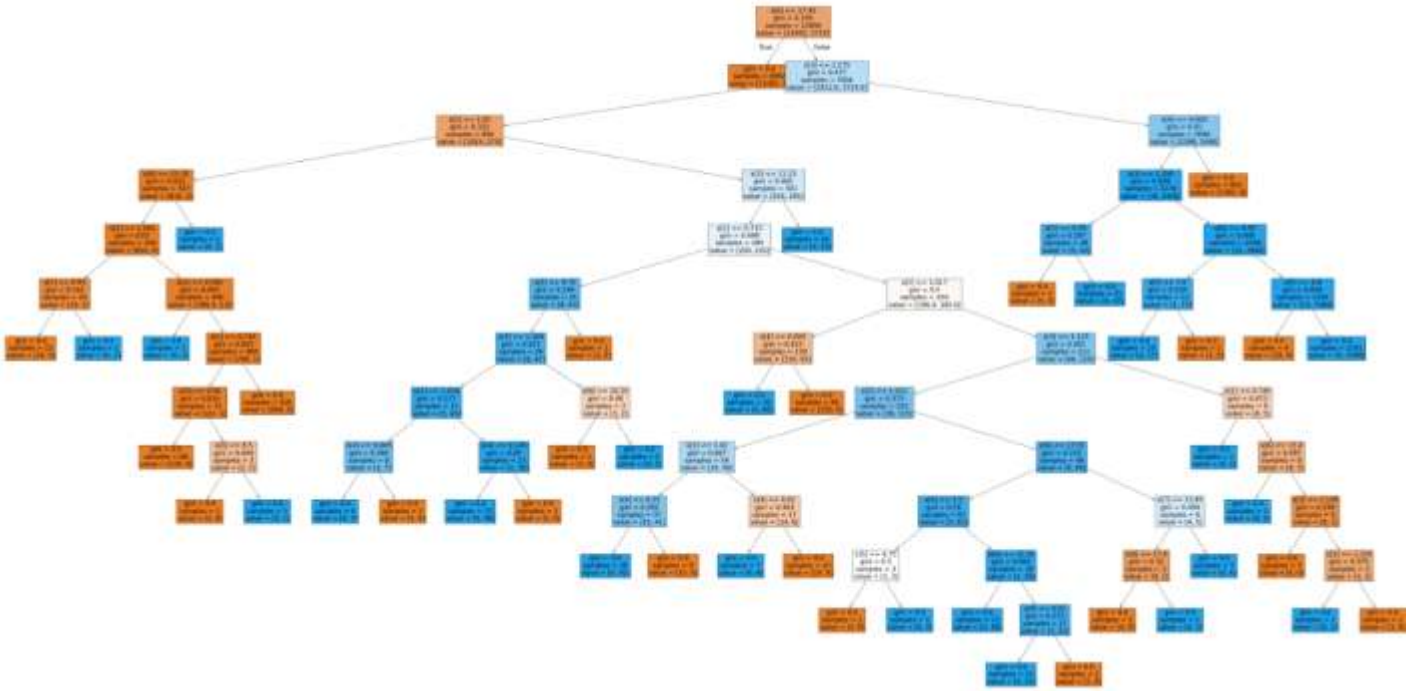


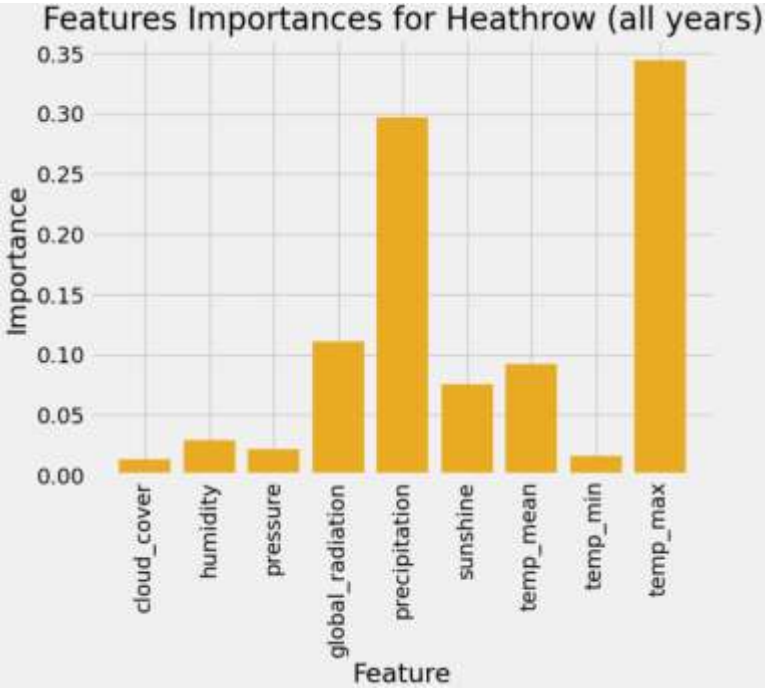
2. Madrid




Features Importances for Madrid (all years)
Importance
0.30
0.25
0.20
0.15
0.10
0.05
0.00
cloud_cover
humidity
pressure
global_radiation
precipitation
sunshine
temp_mean
temp_min
temp_max
Feature

3. Heathrow





Summary:

Max temperature and precipitation show consistently as the two most important features across the three examined weather stations. This means that to predict future climate changes, these two parameters should be considered the most important, with mean temperature and sunshine being a close third and fourth. I would recommend investing in hardware to monitor these four parameters the most.