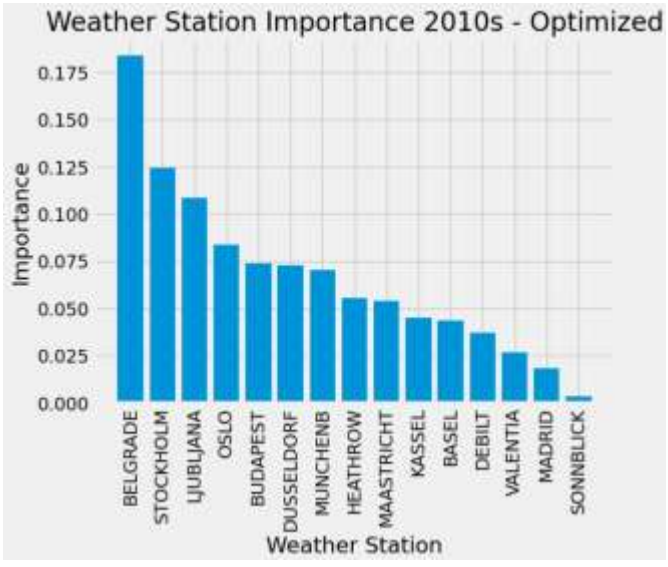
EVALUATING HYPERPARAMETERS

CASE STUDY: CLIMATE DATASET
EXERCISE 2-4

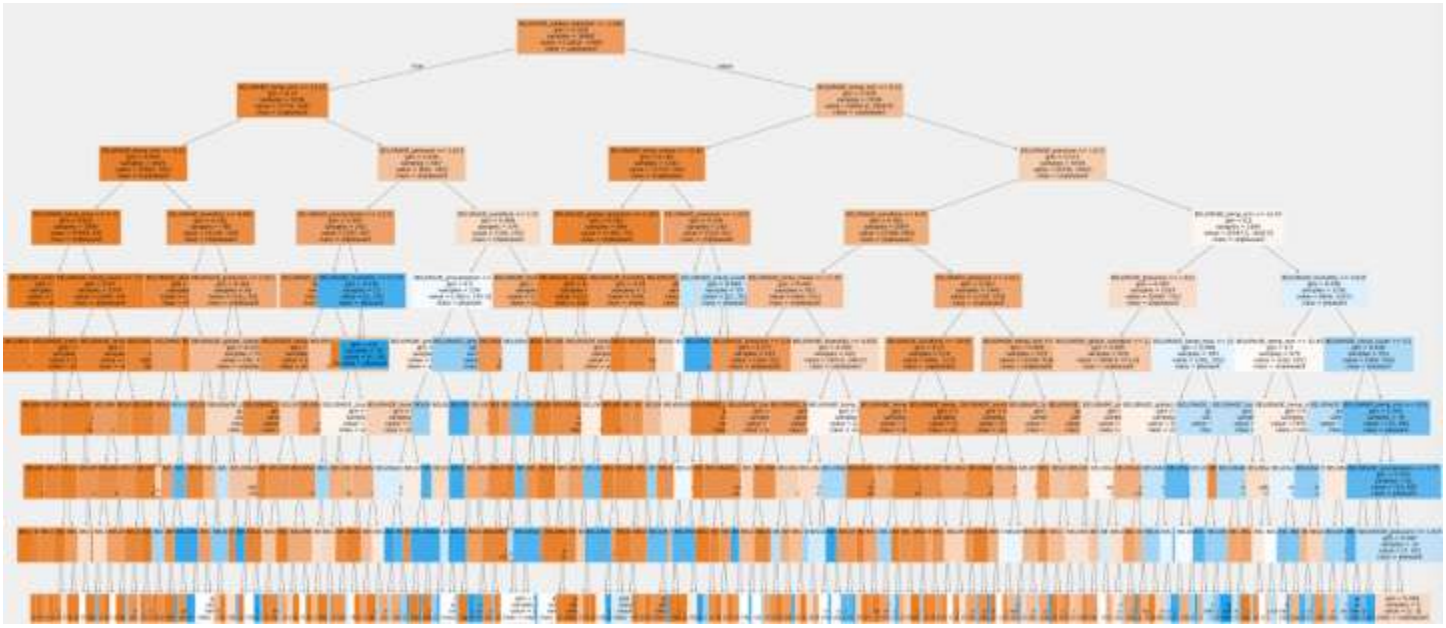
2.4.1 Random Forest:

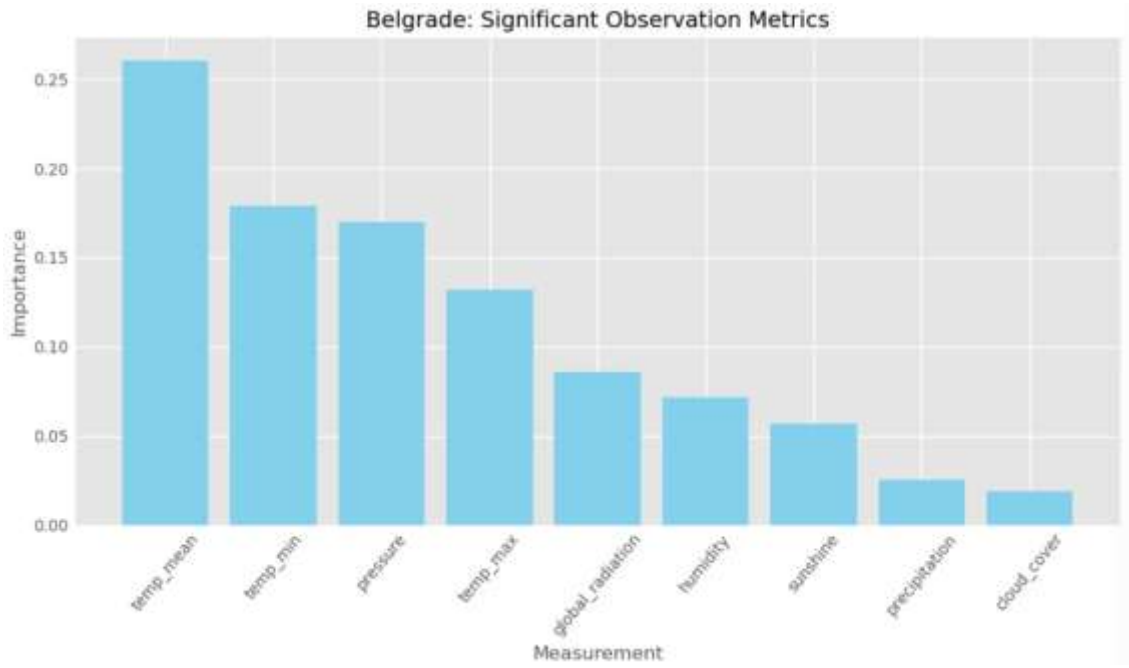
The accuracy of the random forest after optimization was 66.6%, much higher than the 58% in the previous exercise.

The importance table changed a fair bit from the last model as well which is interesting to see.



Accuracy of the random forest for the most important weather station noted in above chart, Belgrade, was found to be 83%. Importance table for this also showed different paraments to be most important.





The importance has now shifted from max temp and precipitation to mean temp and min temp. Belgrade being in cold Serbia, this makes sense.

2.4.2 Deep Learning:

CNN Model:

Before Optimization accuracy: 19.3%

After Optimization accuracy: 85.7%

Confusion Matrix covered 14 of 15 weather stations.

Confusion Matrix:

Pred	BASEL	BELGRADE	BUDAPEST	DEBILT	DUSSELDORF	HEATHROW	KASSEL	\
True								
BASEL	3260	263	21	10	11	8	3	
BELGRADE	45	1045	1	0	0	0	0	
BUDAPEST	22	41	149	2	0	0	0	
DEBILT	16	8	12	46	0	0	0	
DUSSELDORF	4	4	2	4	7	7	0	
HEATHROW	10	3	6	5	3	54	0	
KASSEL	0	4	1	0	1	1	2	
LJUBLJANA	12	8	8	0	0	2	0	
MAASTRICHT	7	1	0	0	0	0	0	
MADRID	24	36	23	6	4	31	0	
MUNCHENB	3	3	0	0	0	0	0	
OSLO	1	0	0	1	0	0	0	
STOCKHOLM	0	0	1	0	0	1	1	
VALENTIA	1	0	0	0	0	0	0	

Pred	LJUBLJANA	MAASTRICHT	MADRID	MUNCHENB	OSLO
True					
BASEL	12	0	94	0	0
BELGRADE	0	0	1	0	0
BUDAPEST	0	0	0	0	0
DEBILT	0	0	0	0	0
DUSSELDORF	1	0	0	0	0
HEATHROW	0	0	1	0	0
KASSEL	2	0	0	0	0
LJUBLJANA	30	0	1	0	0
MAASTRICHT	0	1	0	0	0
MADRID	13	0	321	0	0
MUNCHENB	0	0	1	1	0
OSLO	0	0	0	0	3
STOCKHOLM	0	0	0	0	1
VALENTIA	0	0	0	0	0

2.4.3: Iteration

From the data study so far, the more accurate models start to show importance based on regional weather. I would like to like to segment the data by hot/cold/dry/humid/windy weather by location. Then segment these locations in groups. By also including seasonality into this (should be easy as all of Europe mostly follows the same seasonal patterns in various degrees), we could more accurately predict future conditions based on location and time of year. I would choose Deep learning CNN model as it definitely showed the highest ML accuracy of all the models tested at 85%.

The key variables for Air ambulance to fly safely:

1. Precipitation (for visibility)
2. Wind speed
3. Pressure (sudden changes can be dangerous)