

```
In [4]: runfile('/Users/noemi/Documents/Uni/Astrostatistics/working/lec20.py', wdir='/Users/noemi/Documents/Uni/Astrostatistics/working')
```

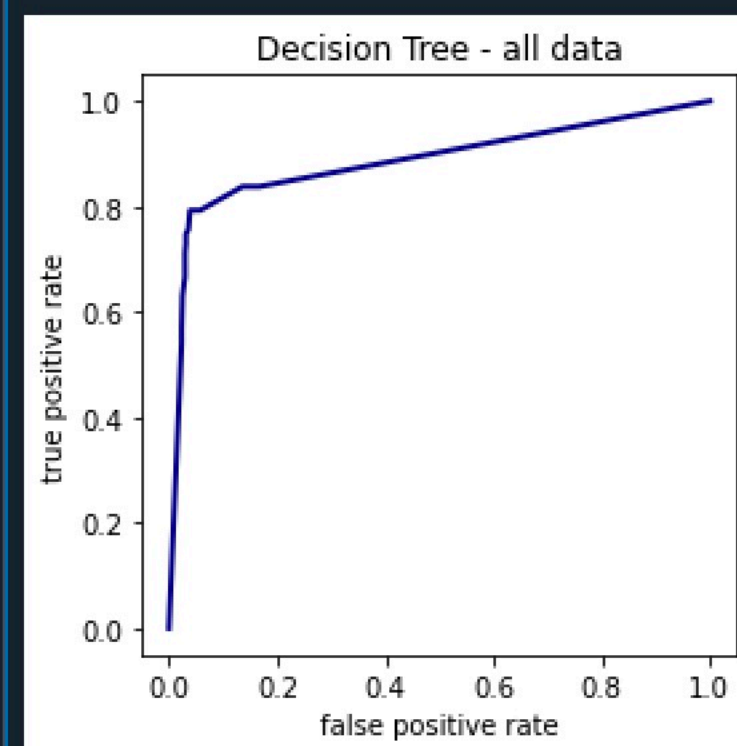
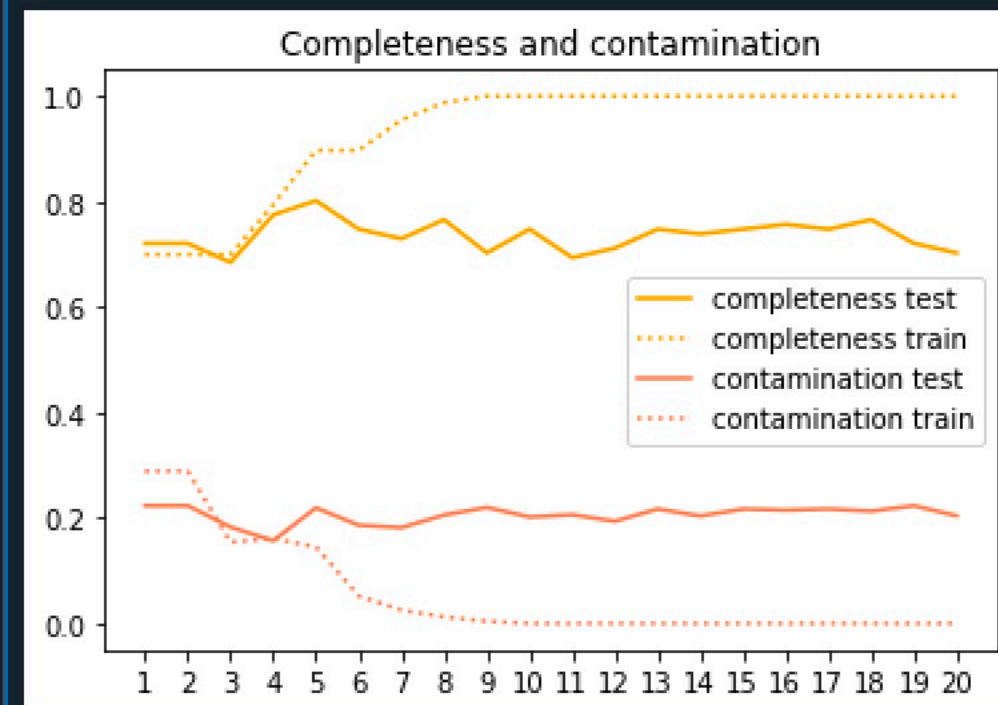
In this analysis I will use 2500 datapoints

--- ALL DATA ---

- Decision Tree -

From cross validation the best maximum depht is 6

100%|██████████| 20/20 [00:00<00:00, 59.37it/s]



Goodness of the classification

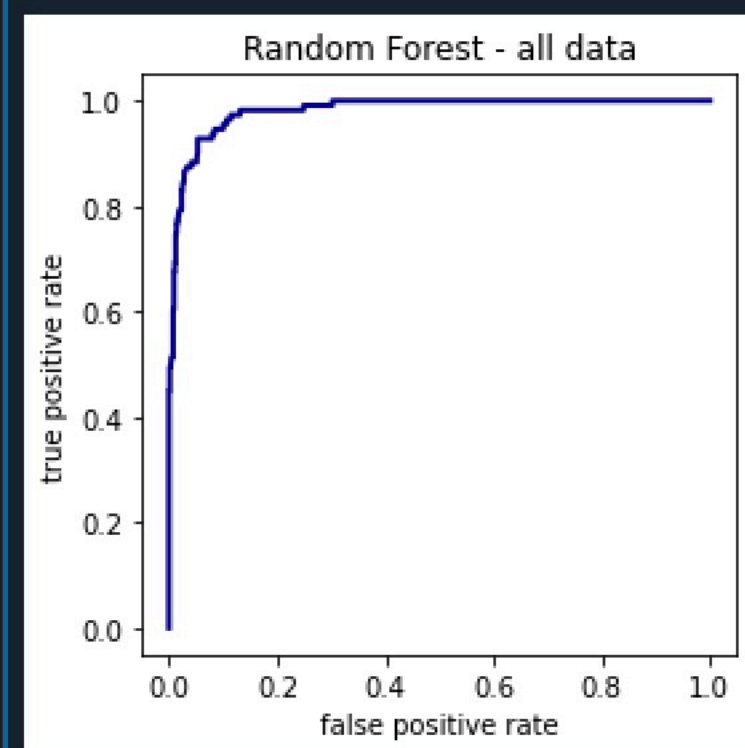
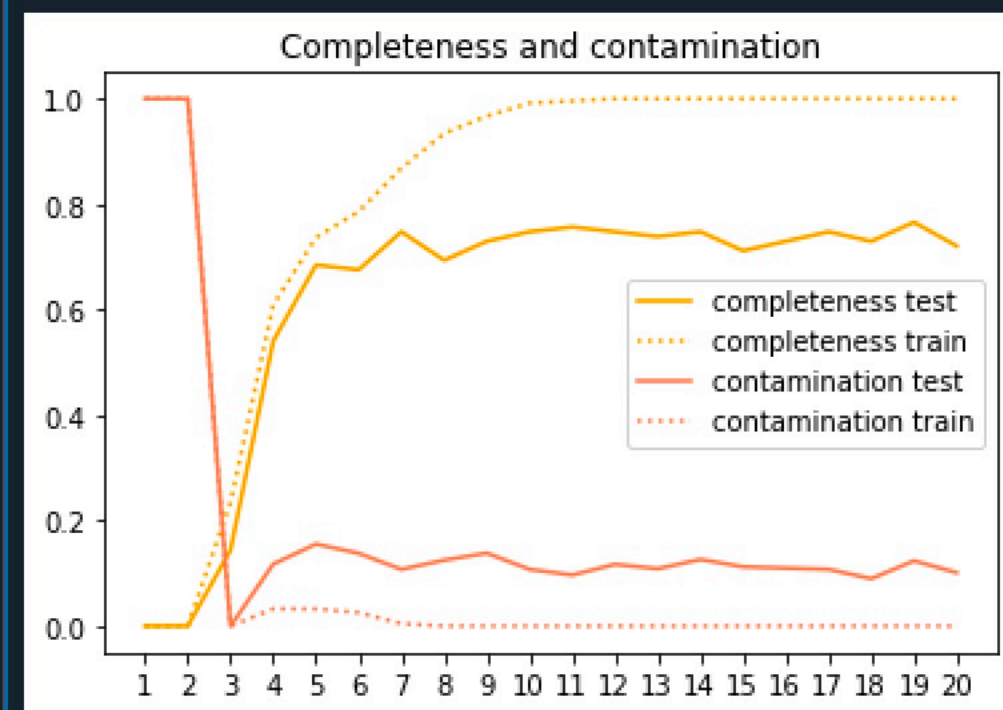
Completeness: 0.7297

Contamination: 0.198

Accuracy: 0.9333

- Random Forest -

From cross validation the best maximum depht is 12  
100%|██████████| 20/20 [00:06<00:00, 3.03it/s]

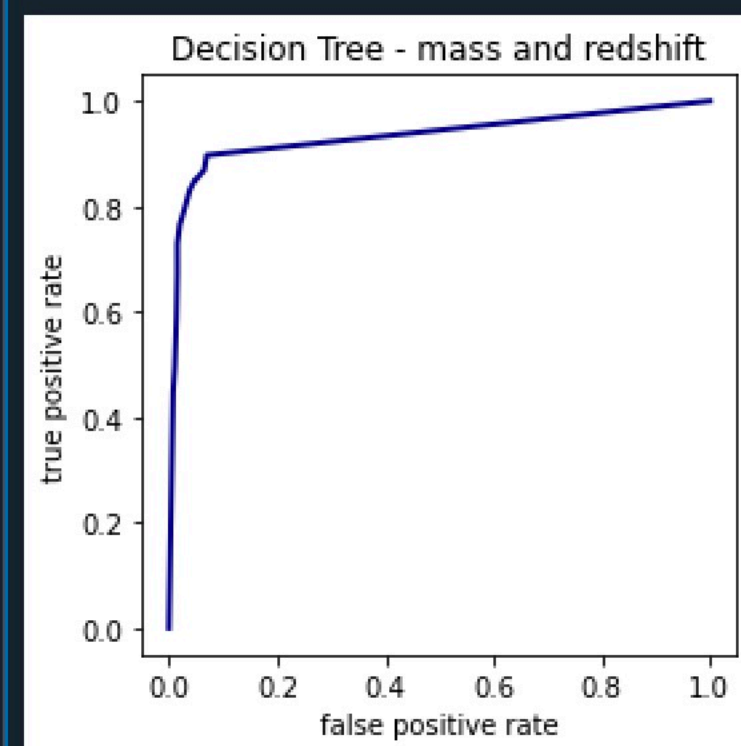
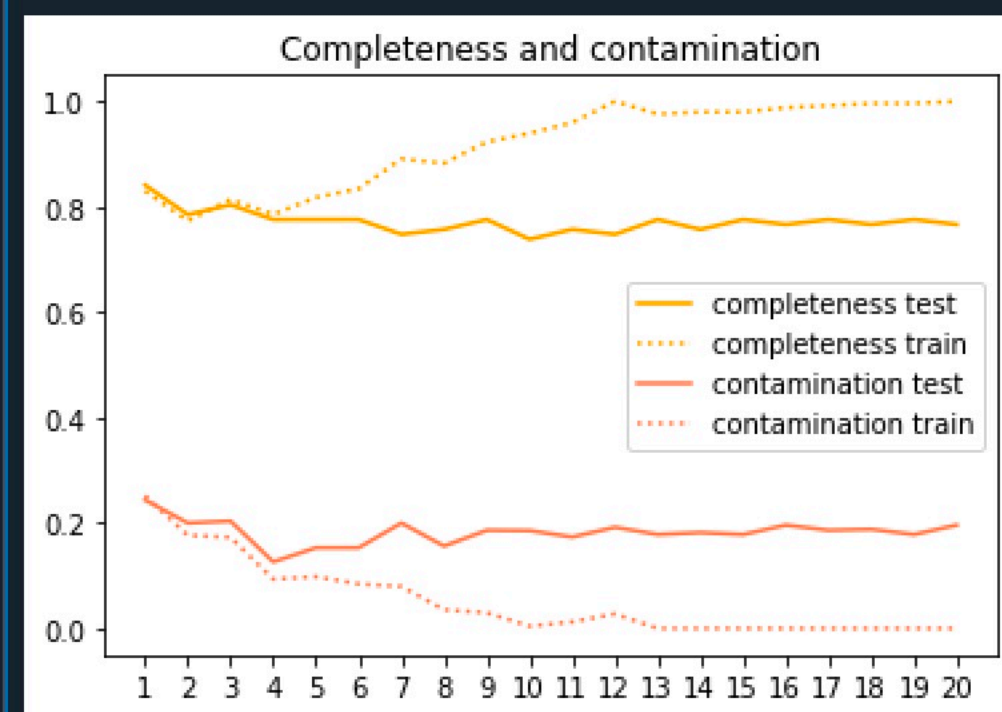


Goodness of the classification  
Completeness: 0.7477  
Contamination: 0.0978  
Accuracy: 0.9507

--- MASS AND REDSHIFT ---

- Decision Tree -

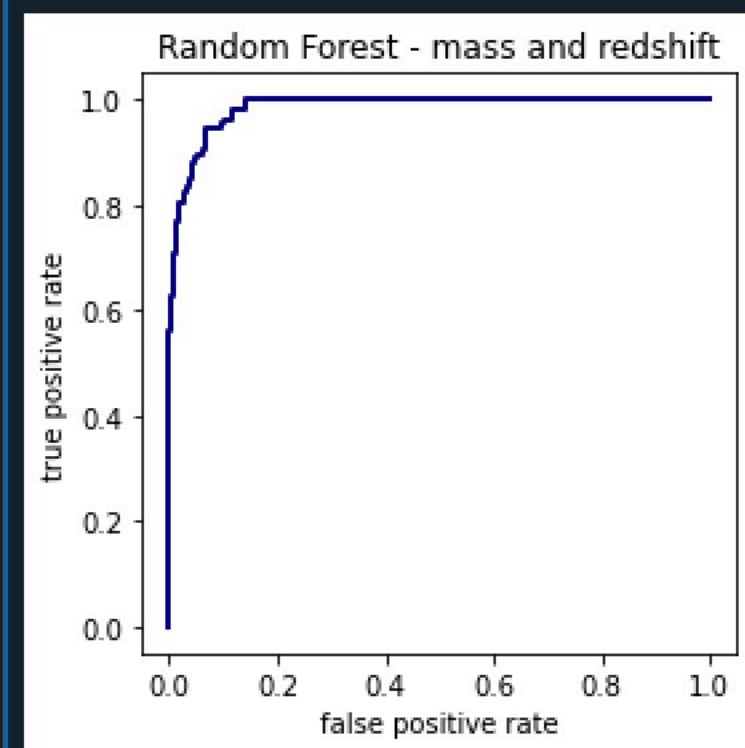
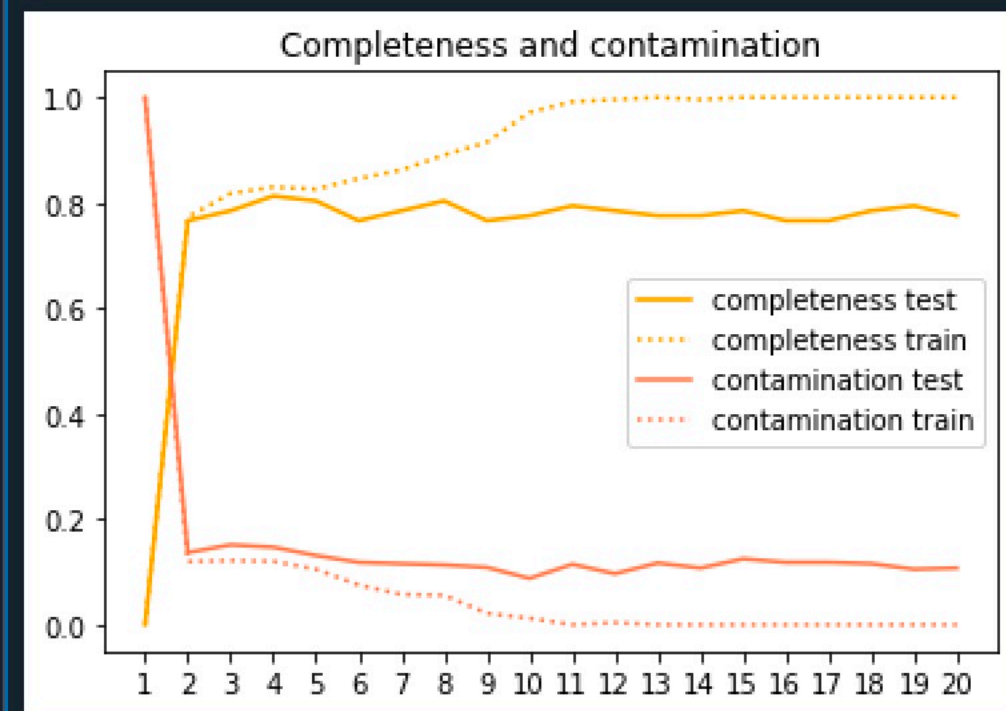
From cross validation the best maximum depht is 6  
100%|██████████| 20/20 [00:00<00:00, 146.62it/s]



Goodness of the classification  
Completeness: 0.7757  
Contamination: 0.1531  
Accuracy: 0.948

- Random Forest -

From cross validation the best maximum depht is 7  
100%|██████████| 20/20 [00:04<00:00, 4.76it/s]



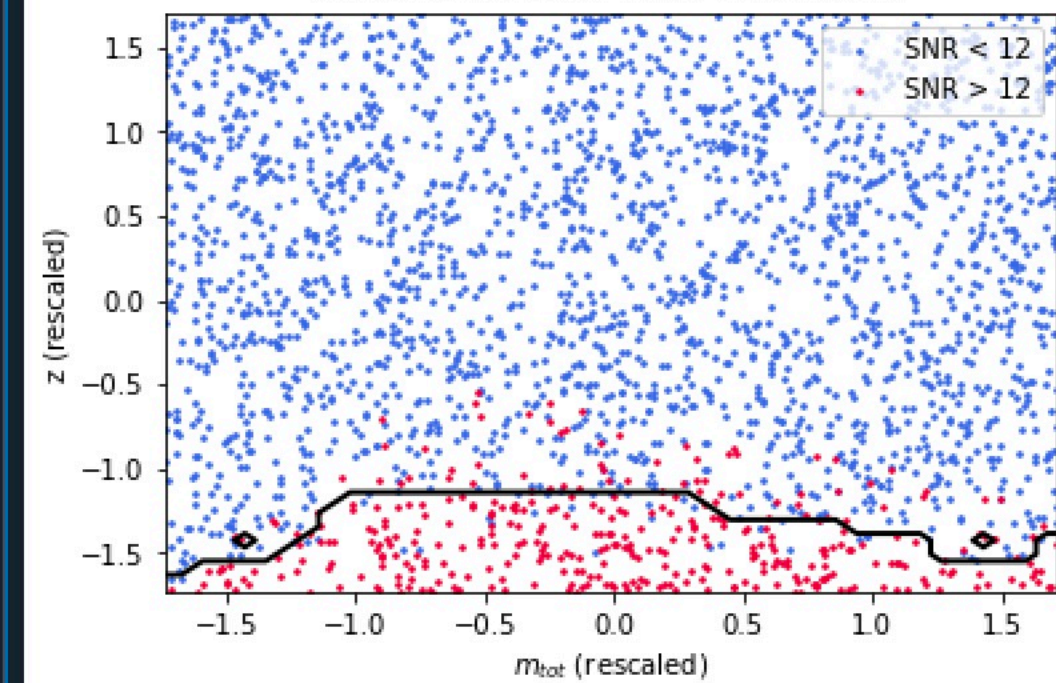
Goodness of the classification

Completeness: 0.7664

Contamination: 0.1183

Accuracy: 0.952

Classification with mass and redshift



--- PCA ---

I will use 4 components

- PCA1 contains 64.1 % of the information

and it's composed by:

mtot: 12.8 %  
 q: 0.3 %  
 i: 14.3 %  
 psi: 14.4 %  
 z: 21.0 %  
 dec: 3.4 %  
 ra: 11.7 %  
 spin 1x: 3.0 %  
 spin 1y: 4.0 %  
 spin 1z: 4.1 %  
 spin 2x: 2.6 %  
 spin 2y: 4.5 %  
 spin 2z: 4.0 %

- PCA2 contains 14.8 % of the information

and it's composed by:

mtot: 1.0 %  
 q: 3.8 %  
 i: 1.6 %  
 psi: 0.0 %  
 z: 13.7 %  
 dec: 9.6 %  
 ra: 35.3 %  
 spin 1x: 6.3 %  
 spin 1y: 3.8 %  
 spin 1z: 4.8 %  
 spin 2x: 6.5 %  
 spin 2y: 6.8 %  
 spin 2z: 6.7 %



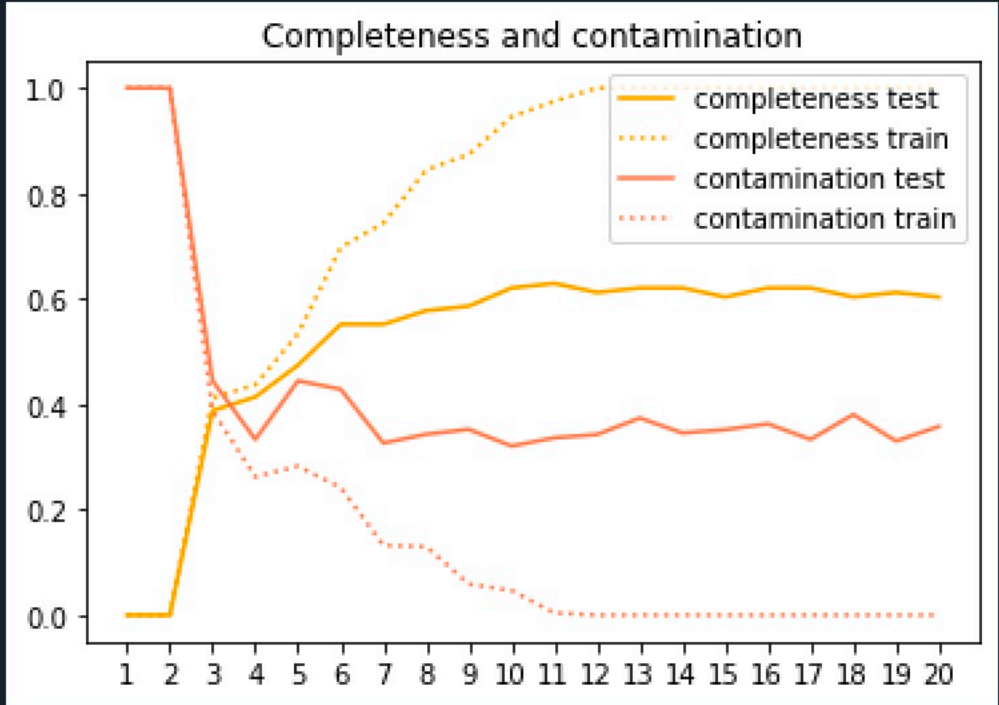
- PCA3 contains 6.9 % of the information  
and it's composed by:  
mtot: 3.3 %  
q: 0.7 %  
i: 34.7 %  
psi: 23.2 %  
z: 8.7 %  
dec: 5.6 %  
ra: 3.8 %  
spin 1x: 2.9 %  
spin 1y: 3.6 %  
spin 1z: 4.0 %  
spin 2x: 4.3 %  
spin 2y: 3.3 %  
spin 2z: 1.8 %

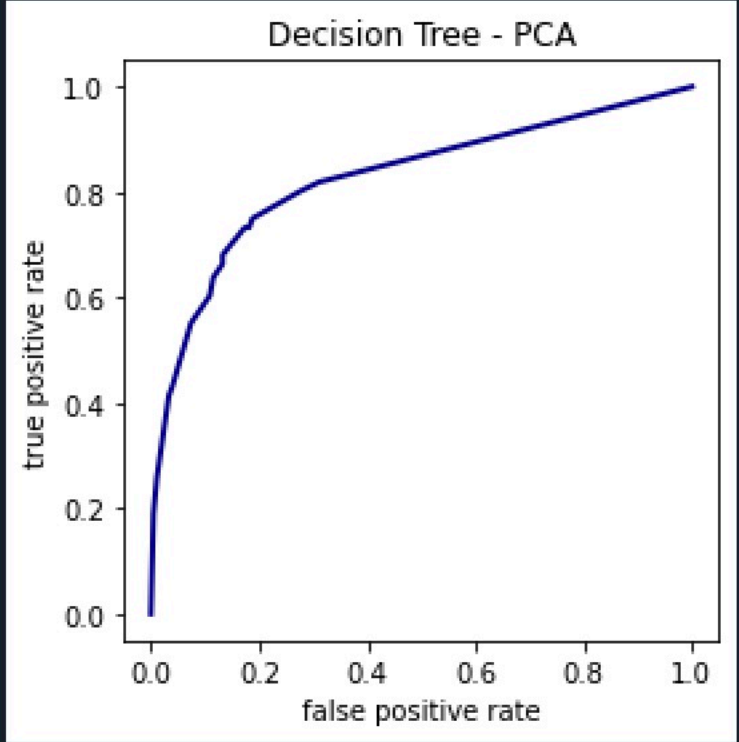
- PCA4 contains 4.7 % of the information  
and it's composed by:  
mtot: 4.6 %  
q: 3.3 %  
i: 6.8 %  
psi: 25.9 %  
z: 24.2 %  
dec: 12.9 %  
ra: 6.2 %  
spin 1x: 1.8 %  
spin 1y: 5.2 %  
spin 1z: 0.3 %  
spin 2x: 2.6 %  
spin 2y: 3.0 %  
spin 2z: 3.2 %

With 4 components I obtain 90.5 % of the information

- Decision Tree -

From cross validation the best maximum depht is 6  
100%|██████████| 20/20 [00:00<00:00, 107.85it/s]

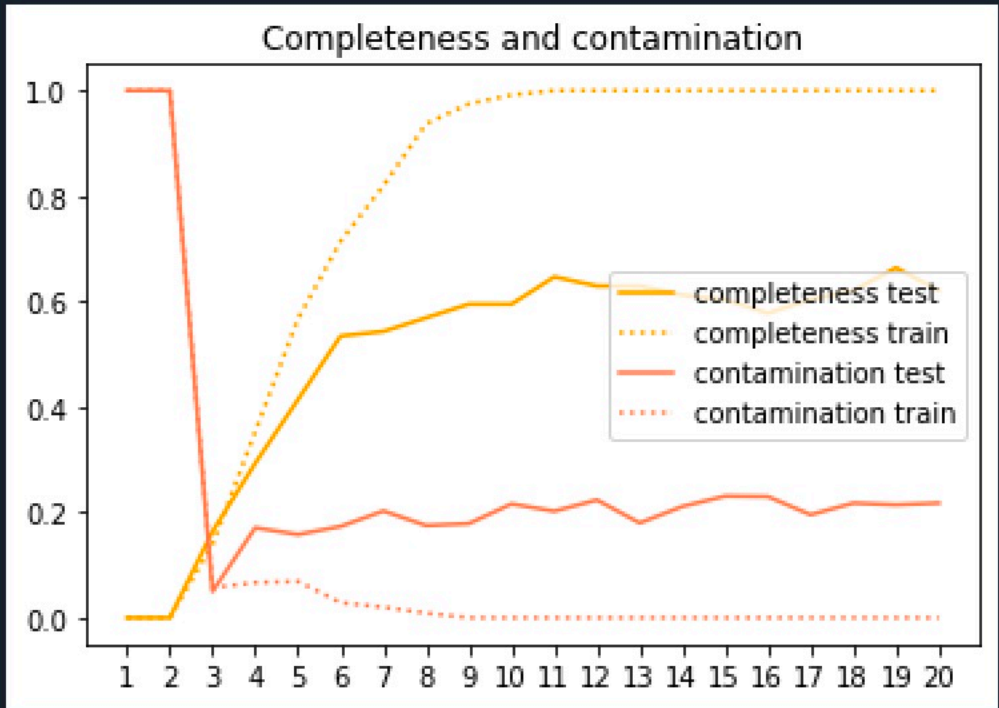


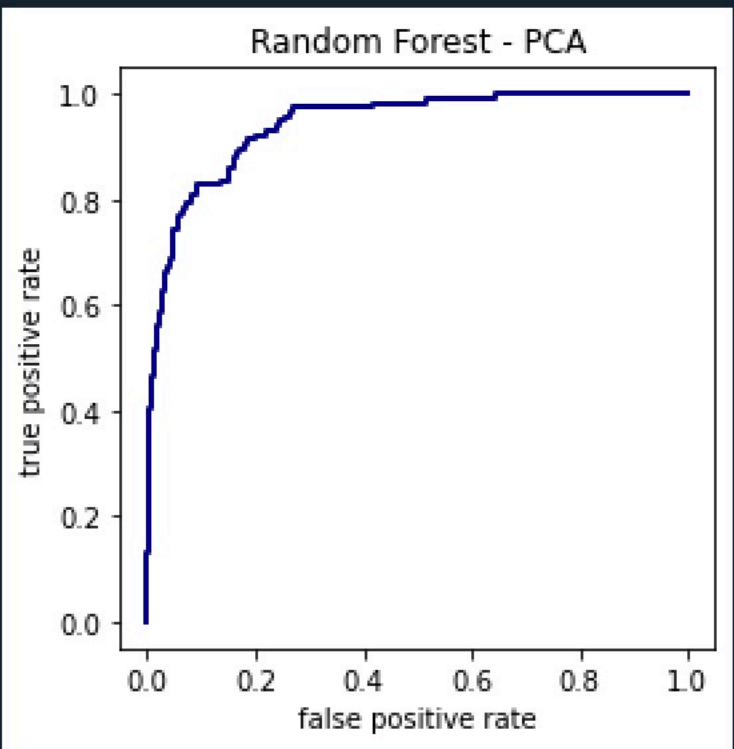


Goodness of the classification  
Completeness: 0.5517  
Contamination: 0.4234  
Accuracy: 0.868

- Random Forest -

From cross validation the best maximum depht is 7  
100% | 20/20 [00:05<00:00, 3.62it/s]





Goodness of the classification  
Completeness: 0.5517  
Contamination: 0.1579  
Accuracy: 0.9147