Machine Learning project

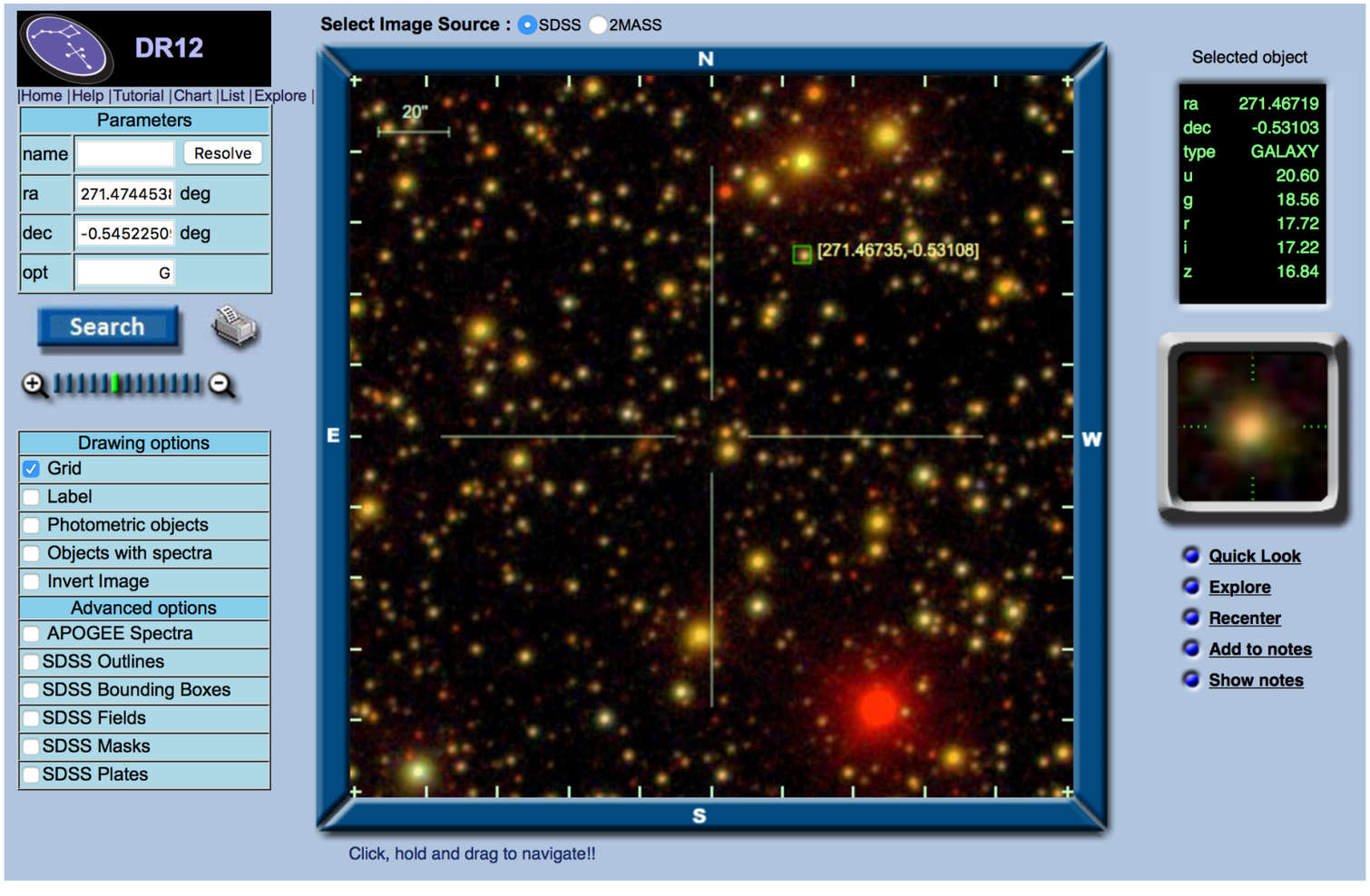
Testing SDSS classification technique

Laura Hangard

17-05-2016

Data and goal

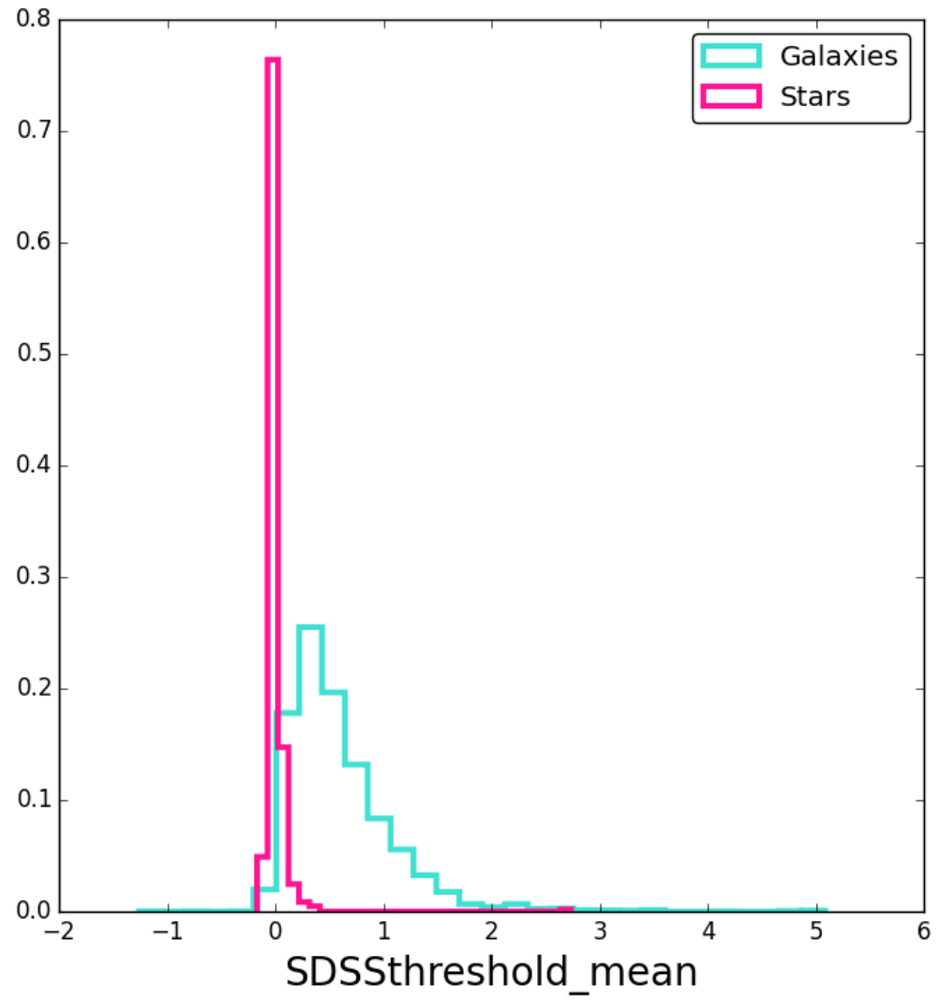
- Data : catalogue of ~5000 galaxies and ~2000 stars
- ~2000 galaxies are misclassified stars



Data and goal

Method used by SDSS :

- **(psfMag - cmodelMag) > 0.145 —> Galaxy**
- **< 0.145 —> Star**



(+ little additional checks, most likely)

Data and goal

Idea :

Can we define a better classification method than this one, using different features?

Features :

- SDSS method features (PSF mag and model mag)
- Additional features (colours, radius)

Machine learning

- Supervised learning, classification
- Decision Tree
- Random Forest
- SVM
- Boosting classification

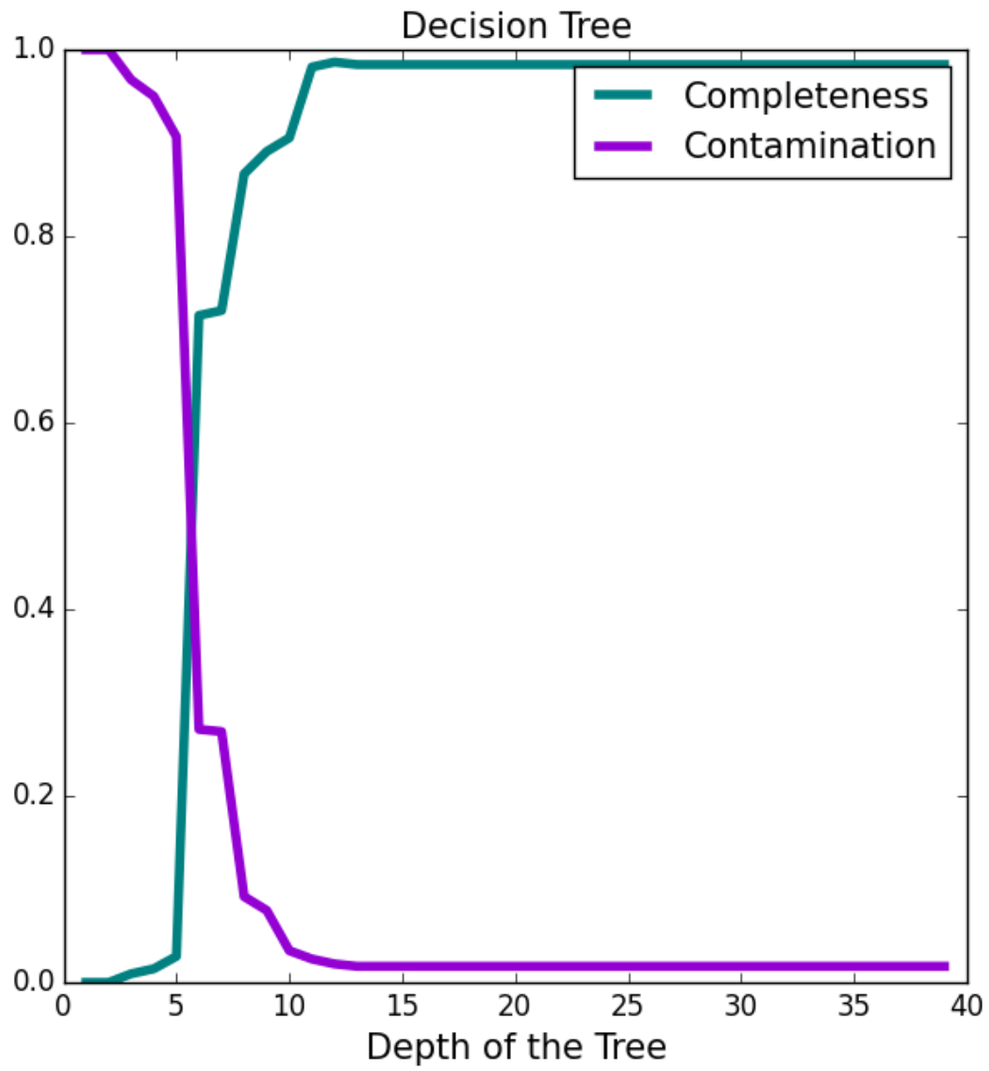
Preprocessing data

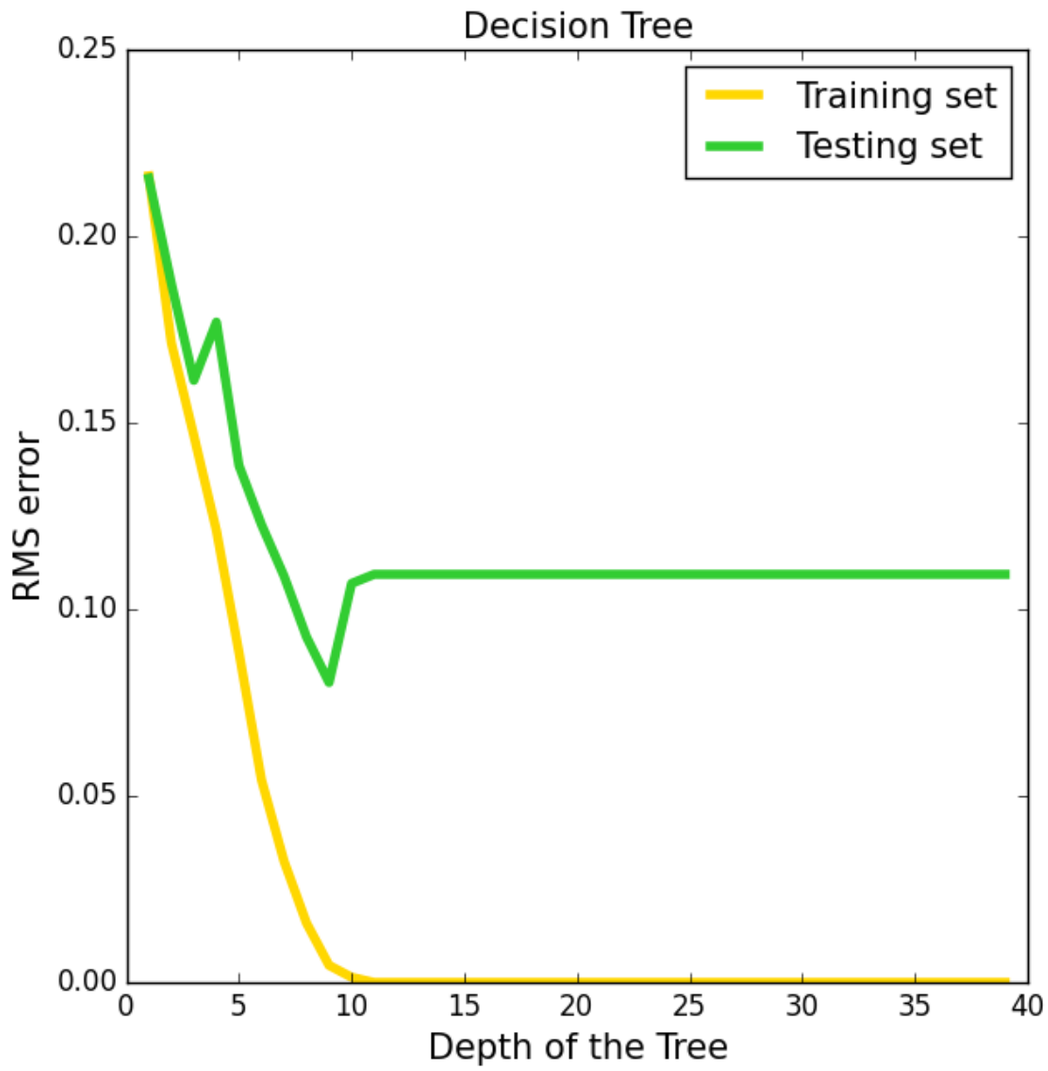
- Make all the features comparable : zero mean and standard deviation of 1
- Split the data : 80% training sample, 20% test sample (for cross-validation)

SDSS classification method : test

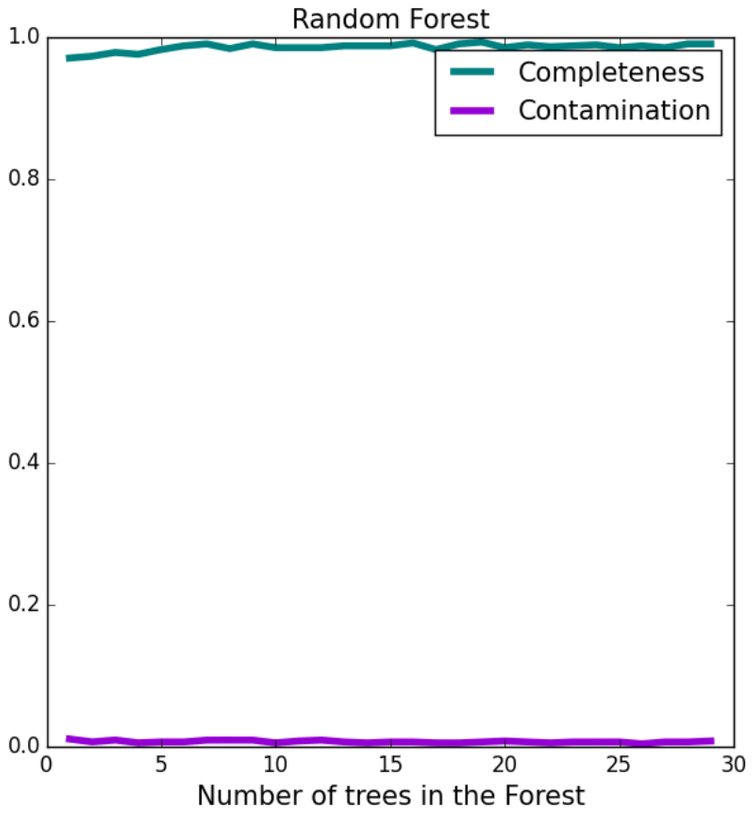
- First step, consider only correctly classified objects
- Features: only SDSS criterion

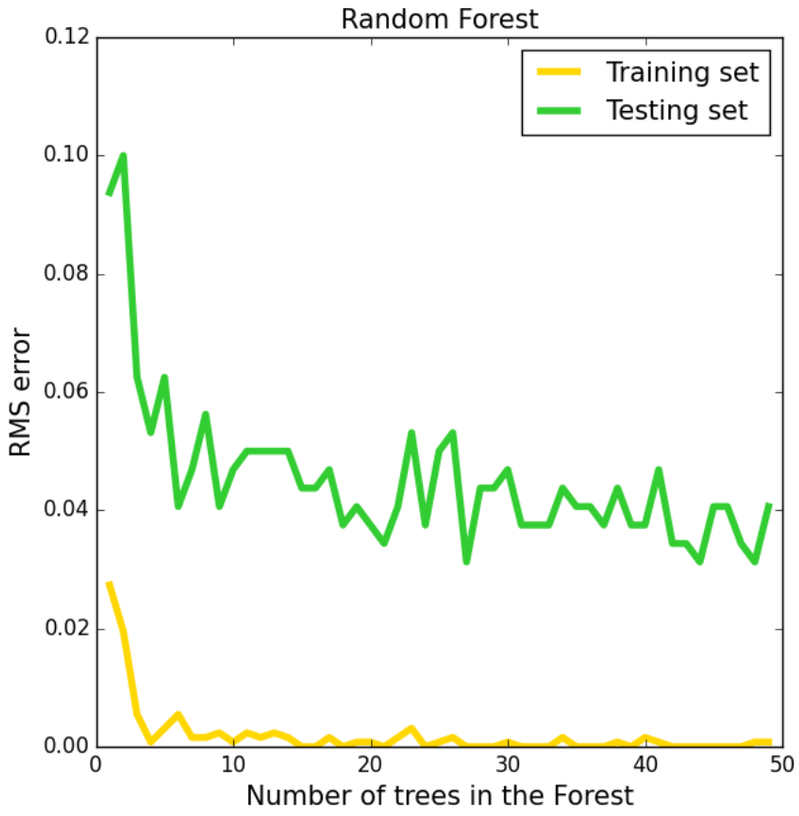
SDSS classification method : test

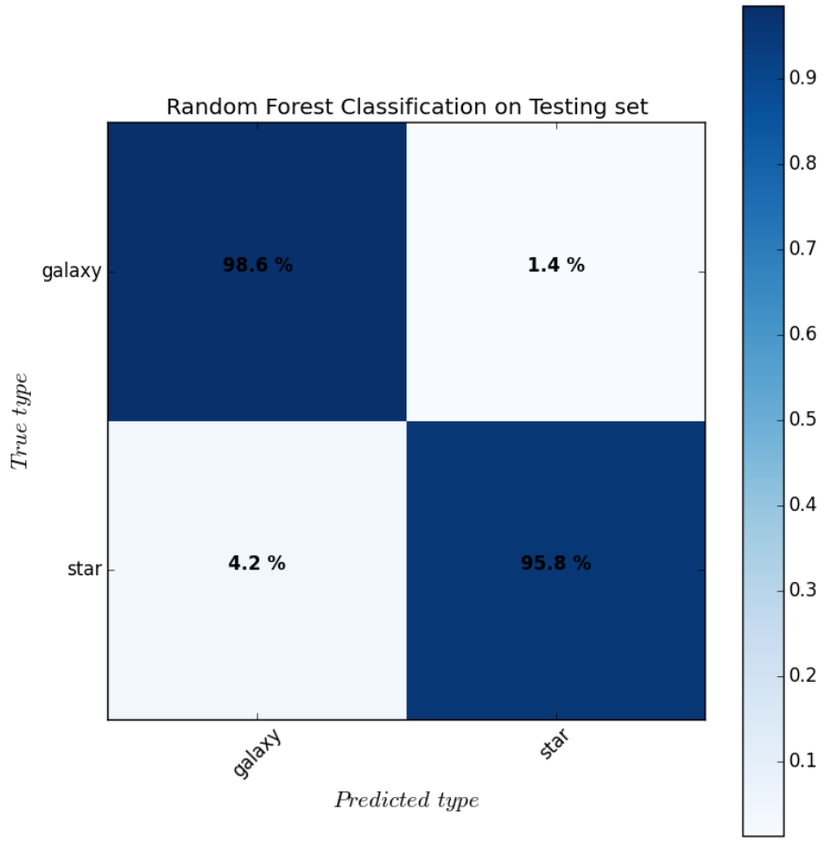




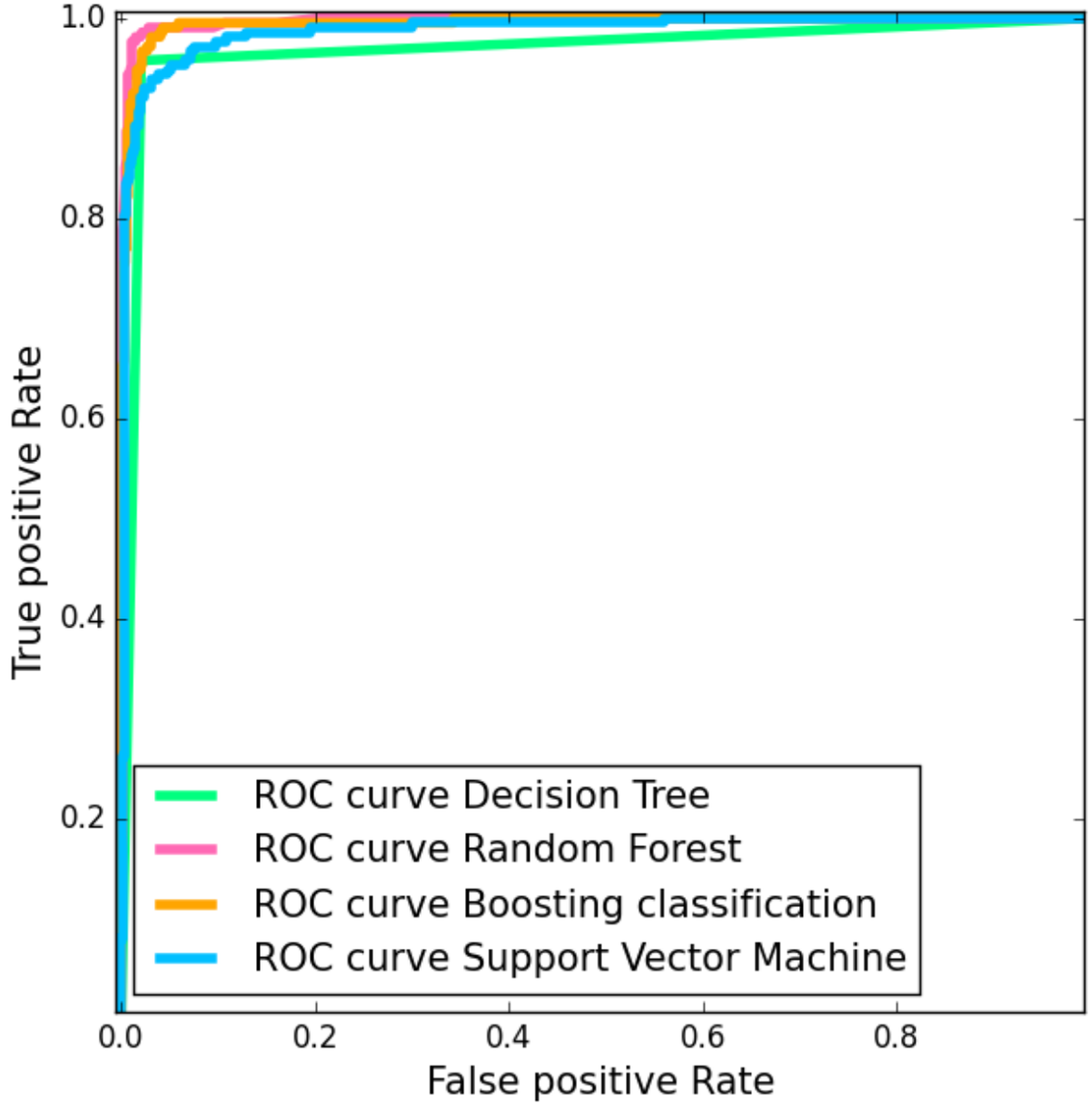
SDSS classification method : test





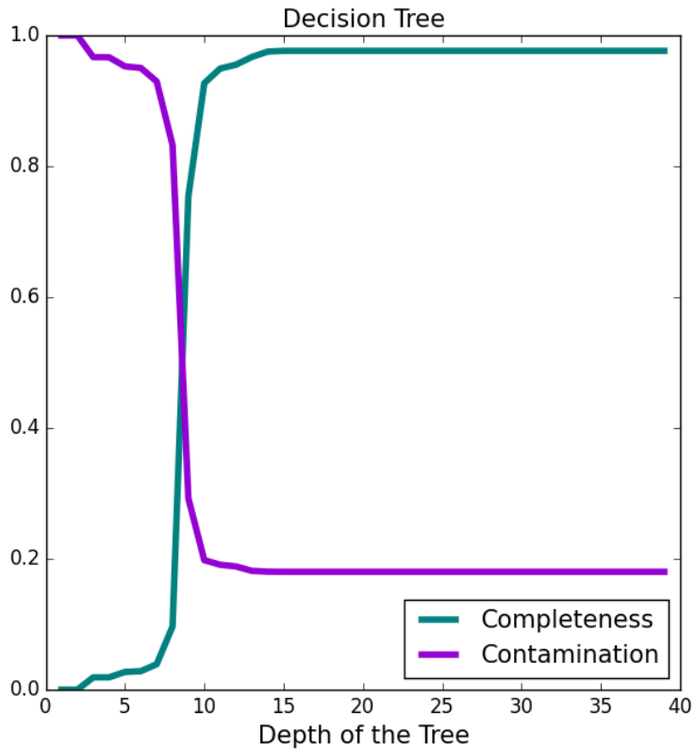


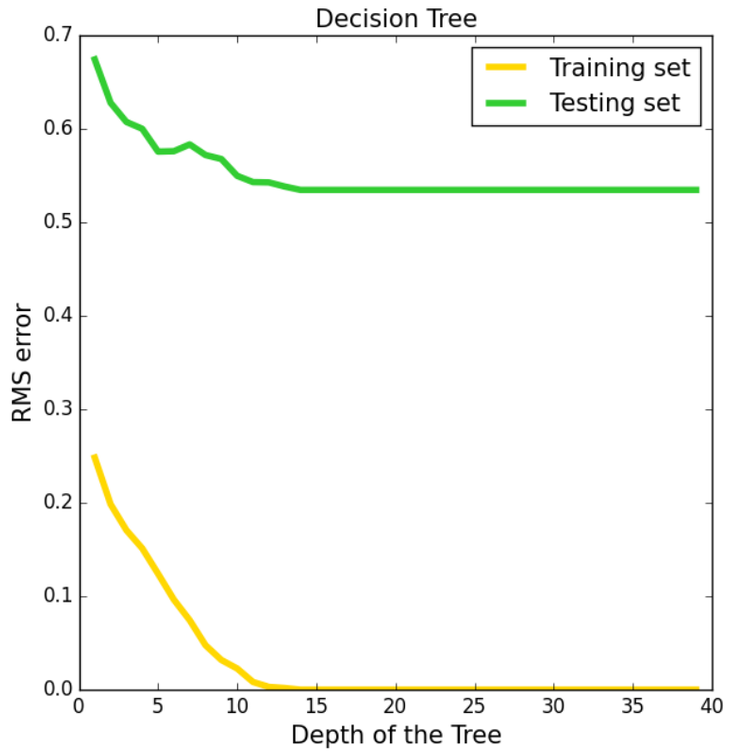
SDSS classification method : test

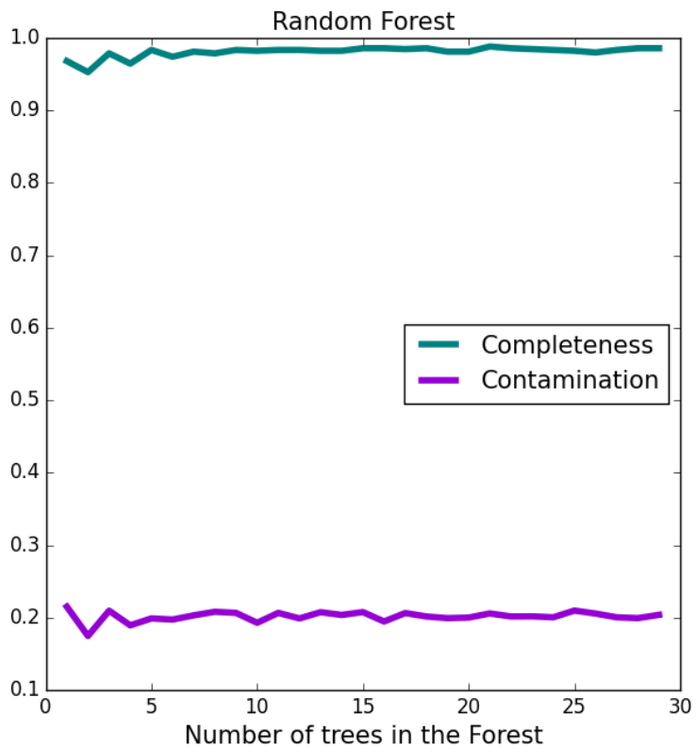


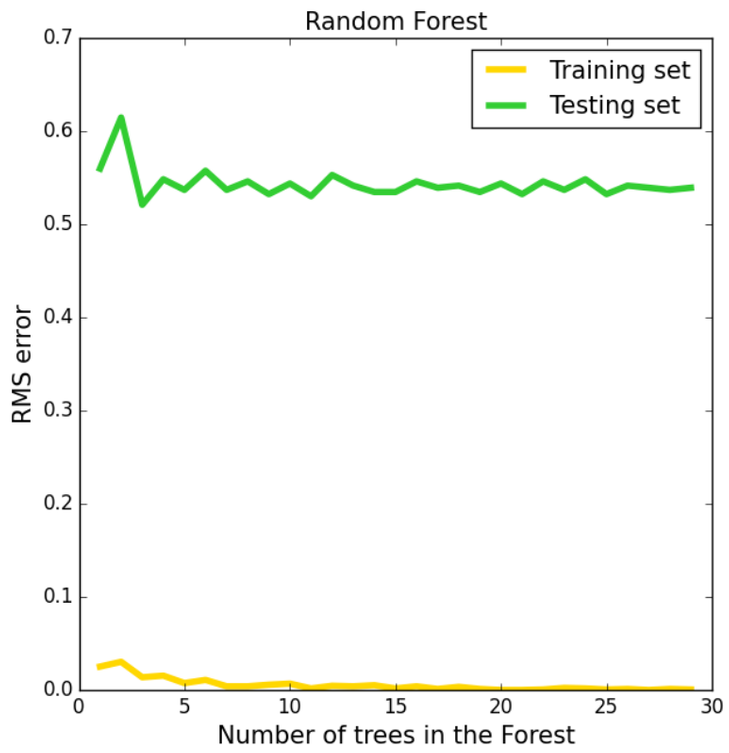
Adding misclassified objects

Data :
+ 2000 stars that SDSS criterion fails at classifying

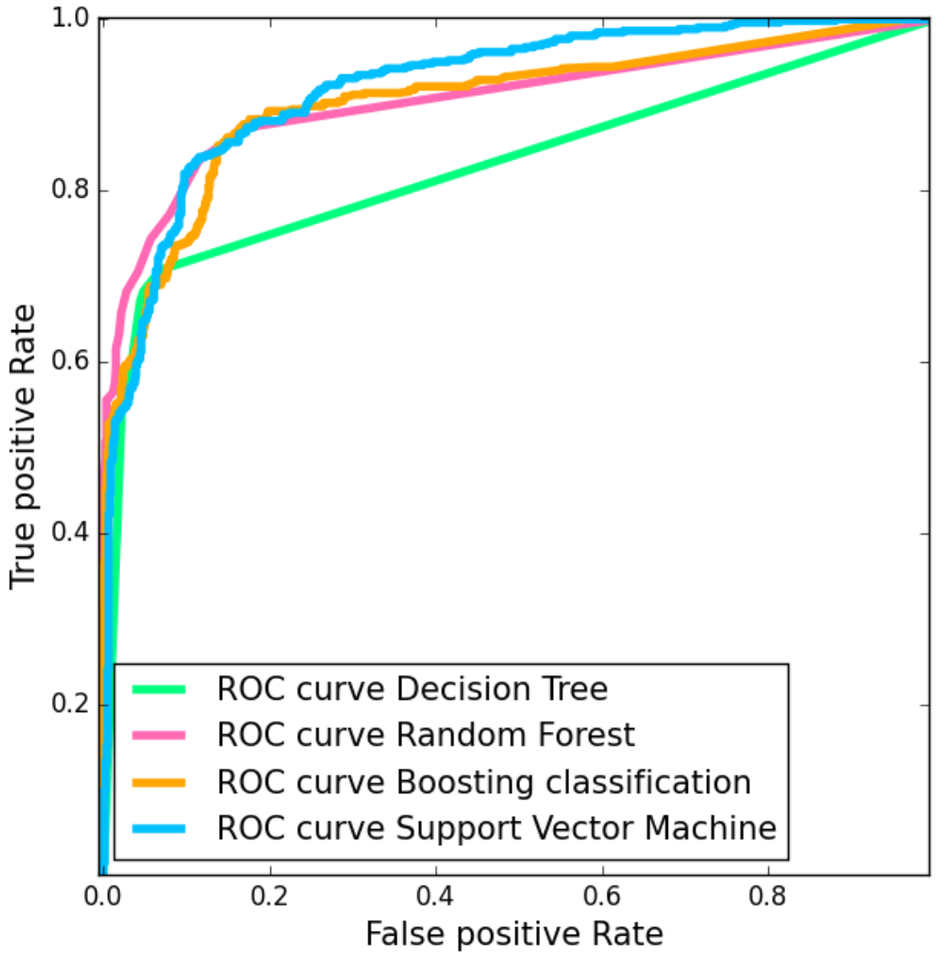


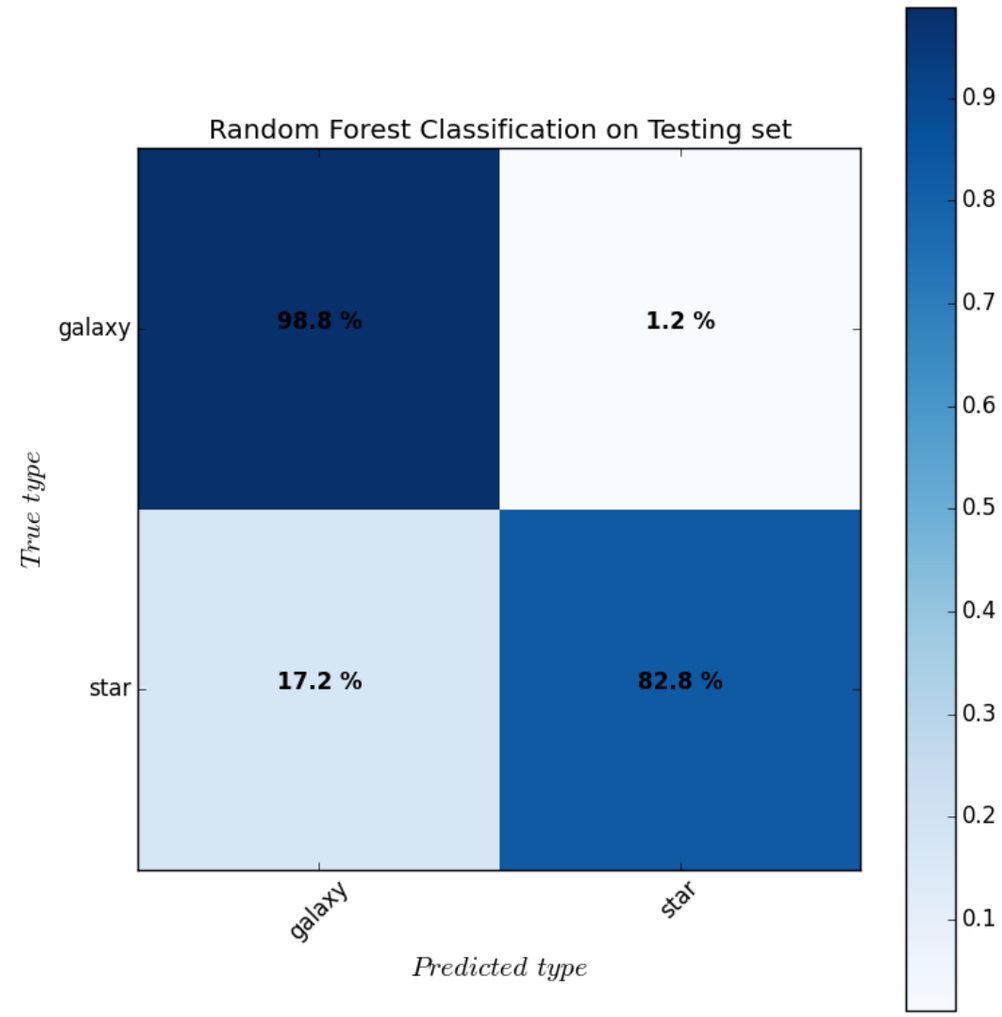




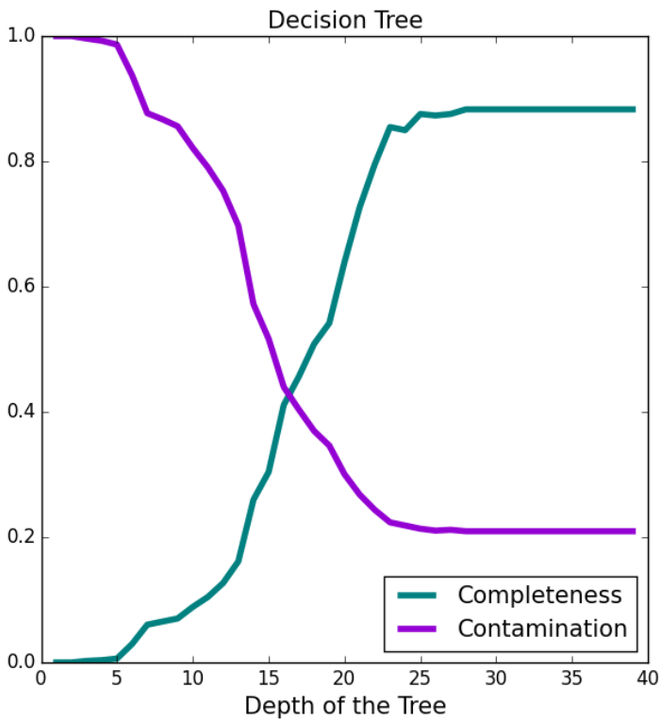


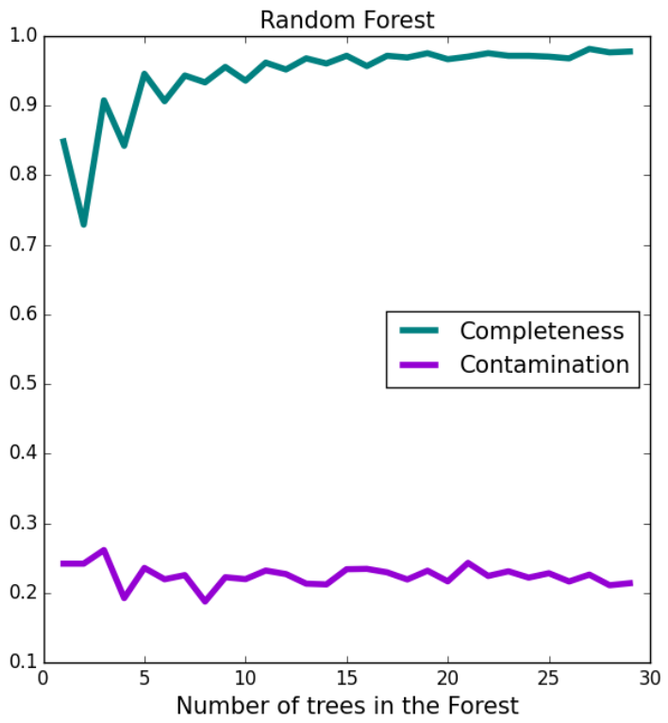
Adding misclassified objects





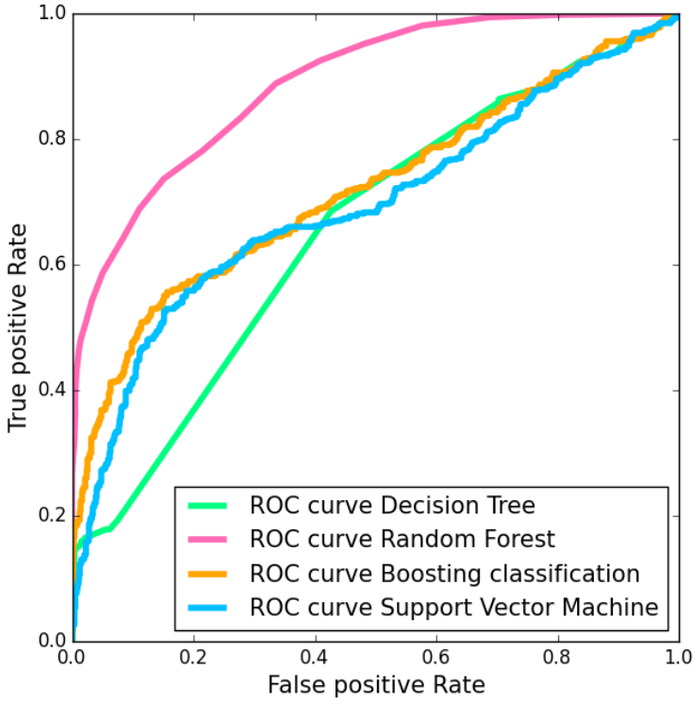
Testing new features

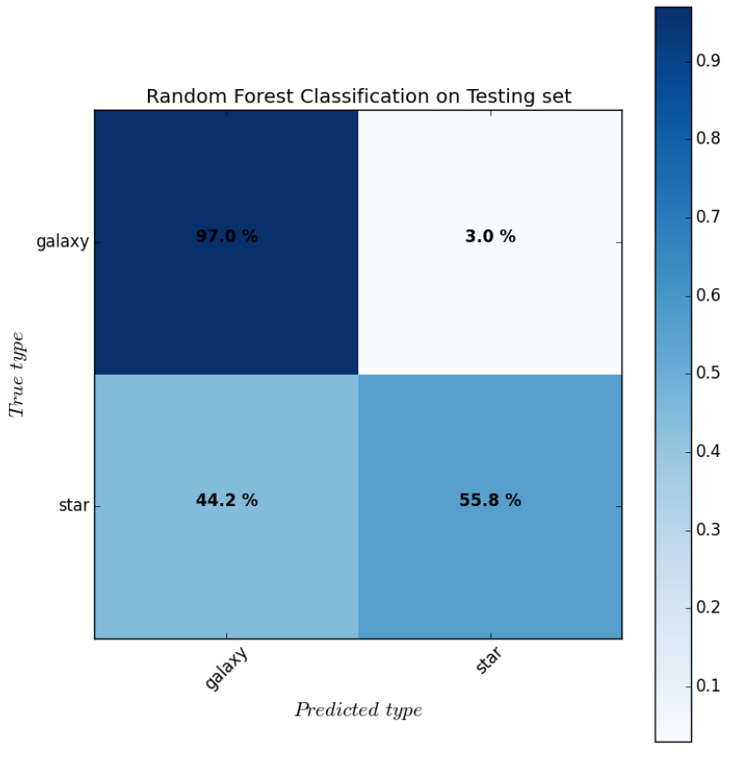




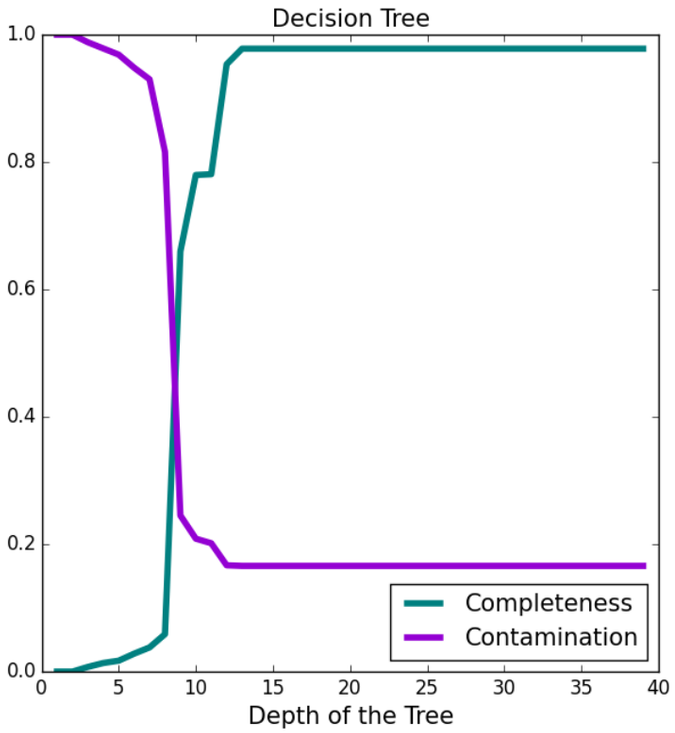
Features:
optical colours + radius

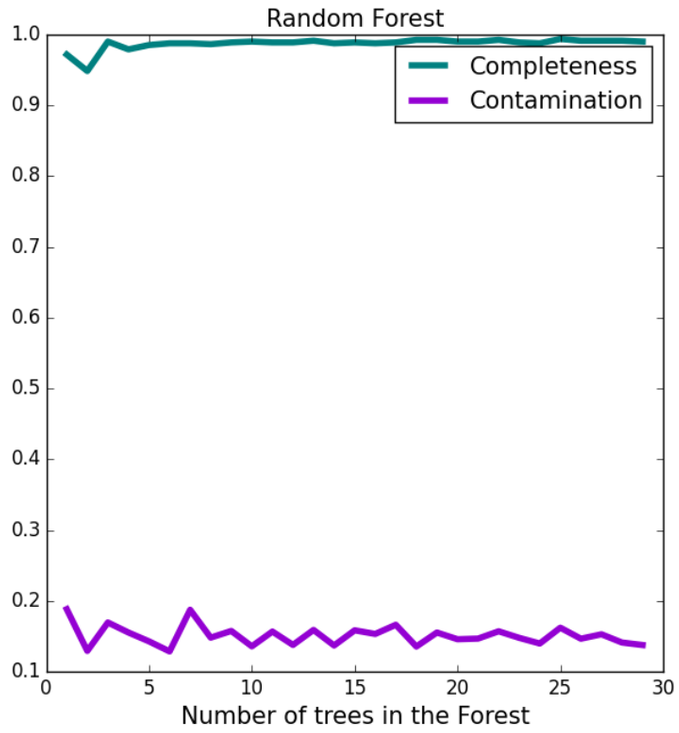
Most important feature :
radius





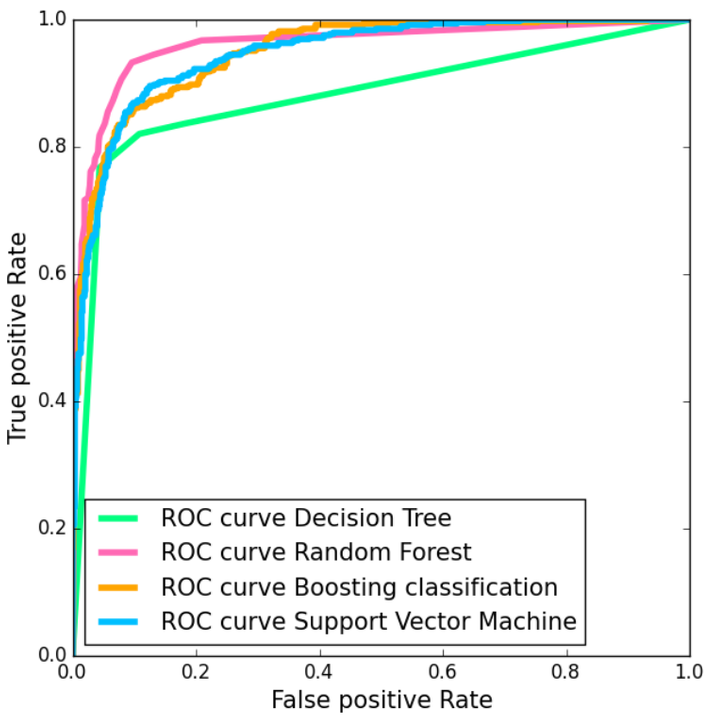
Mixing features

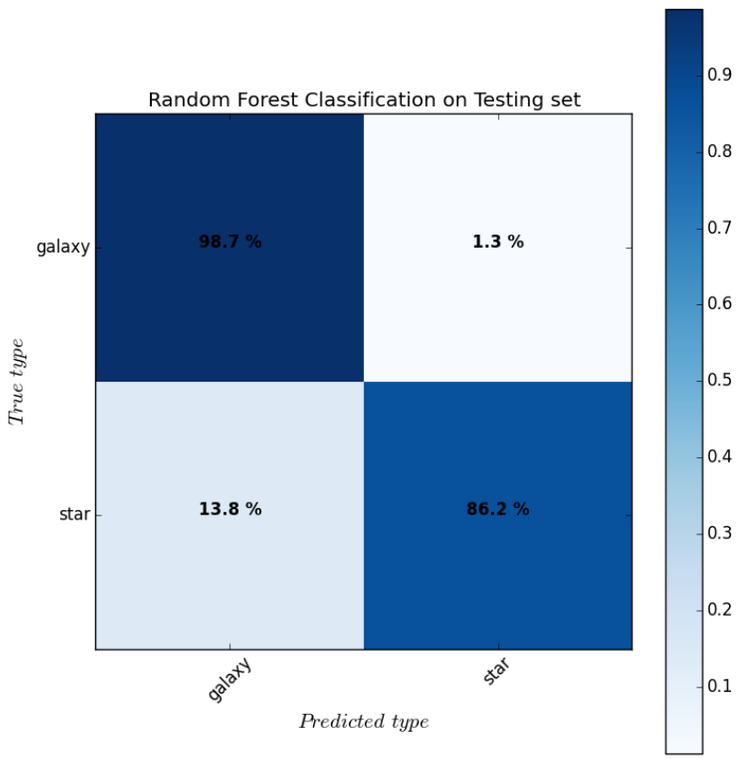




Features:
SDSS criterion
+ colours + radius

Most important feature :
SDSS criterion





Conclusion

- Slightly better with additional features
- Not enough to justify the use a much more complicated model (size of SDSS catalogue)