1. **LOGISTIC REGRESSION**

I trained three models, two Logistic Regression one simple and another Multinomial, and a Random Forest Model.

The features selected as regressors were: 'koi\_fpflag\_nt', 'koi\_fpflag\_ss', 'koi\_fpflag\_co', 'koi\_fpflag\_ec', 'koi\_period', 'koi\_time0bk', koi\_impact','koi\_duration','koi\_depth','koi\_prad','koi\_teq','koi\_insol','koi\_slogg','koi\_model\_snr', 'koi\_tce\_plnt\_num', 'koi\_steff','koi\_slogg', 'koi\_srad', 'ra', 'dec', 'koi\_kepmag'

In the case of Logistic Regressions both models reach the following scores:

**Logistic** :

* Training Data Score: 0.5088689681480069
* Testing Data Score: 0.5137299771167048

**Multinomial**:

* Training Data Score: 0.5012397482357429
* Testing Data Score: 0.5011441647597255

This models can´t predict with enough certainty which candidates belong to the class CONFIRMED as new exoplanets.

With Grid Search I found that Logistic Regression could improve its score to 0.8191874880793439, with C= 7.742636826811269.

In the case of Logistic Regression Multinomial the model could improve its score to 0.8178523745946977 with C= 59.94842503189409.

I used simple Logistic Regression and Multinomial in order to compare if a Multinomial model could perform better, but the scores are not very different form the scores resulted of using Logistic Regression.

In order to improve their performance, it will be necessary to include significative features as for example features shared with declared planets.

1. **RANDOM FOREST**

The features selected as regressors were: 'koi\_fpflag\_nt', 'koi\_fpflag\_ss', 'koi\_fpflag\_co', 'koi\_fpflag\_ec', 'koi\_period', 'koi\_time0bk', koi\_impact','koi\_duration','koi\_depth','koi\_prad','koi\_teq','koi\_insol','koi\_slogg','koi\_model\_snr', 'koi\_tce\_plnt\_num', 'koi\_steff','koi\_slogg', 'koi\_srad', 'ra', 'dec', 'koi\_kepmag'

In the case of the random forest model, it performs better than Logistic Regression, the general accuracy of this model is Accuracy: 0.7276887871853547 mean accuracy of each observations 0.9006289159432102 with a standard deviation of 0.006486526037407053 which means that our prediction has a very low variance.

With Grid Search our model improve and reach a score of 0.9013923326339882.

The best model to predict classify candidates to exoplanets is in this case the Random Forest.