Conclusion: We can’t pretend that we can establish a strong betting strategy with the results we found with our analysis. However, we discovered some interesting points comparing the two different models and using different variables. Firstly, we saw that the predictions of the models using the games’ statistics are significantly better than the ones with the odds. It’s important to keep in mind that we used the statistics of played games and we didn’t use time series, but we could imagine that the odds are not perfectly predicted by websites because they not only take account of the statistics but they also take account of the emotional incitement of the players.

Concerning the two classification methods we used, we saw that the logistic regression as always, a better accuracy rate than the decision tree method. However, it’s better to use decision trees when we have a big inequality in the distribution of the outputs, even if we don’t have a good precision and recall, we are able to calculate them with the decision tree when it’s not possible with the logistic regression.

We observe significant results and we are sure that with more time and with a deeper knowledge of data analysis and Python we could improve our model. For example, by adding other variables like the actual market value of each team or the actual ranking of each team. We could also use time series to use the previous statistics. We could also add some conditions to select some specific bets.