Russia 2018 Prediction Challenge

This article consists of my submission of the *Russia 2018 prediction challenge*. It will include a detailed explanation of the process which I used to come to my conclusions.

In this competition, I resorted to the basic.csv file and added extra data from various sources, for example, I used score power from the fifa website, and the latest results from the 2018 friendly games as I couldn't find enough data from the given csv file I also used the power information for each team from the PES video game (def power, speed power, mid power, attack power…)

During the data preprocessing phase, I transferred the result of each game between the national teams:

Column result:

- 1: If the Home Team won the game
- 0: If it's a tie
- 2: If the Away Team won the game

I created new features which consist of the ratio of power of each team's information.

Then, I trained my model and I used as input the ratio features and as output the result of the previous games (0, 1, 2)

Furthermore, I used the same model to predict the probability of win, draw and lose. I did the same process of each year (2015, 2016, 2017 and 2018)

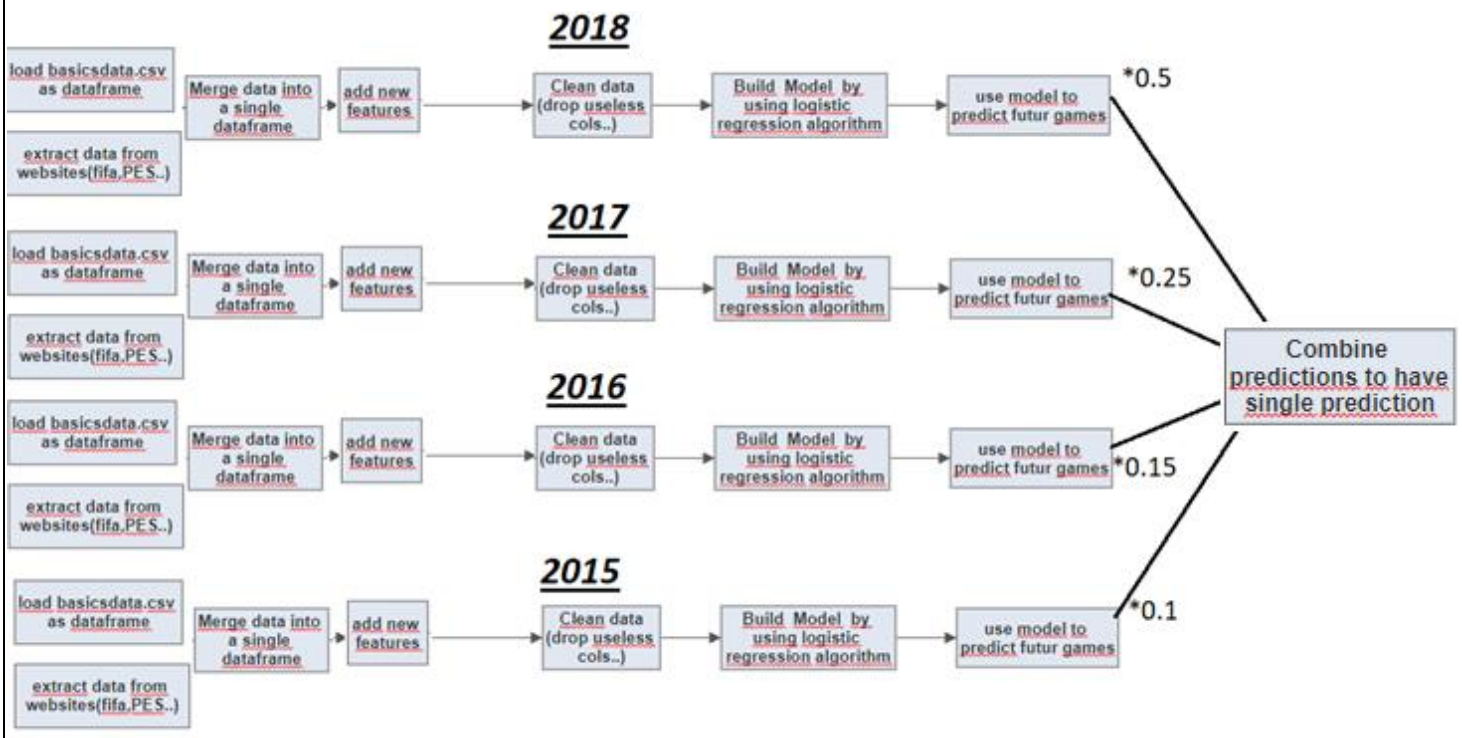
In the end, I used all my submission to build the final submission file.

File_sub =0.1*sub_2015+0.15*sub_2016+0.25*sub_2017+0.5*sub_2018

To conclude, the Russia 2018 Prediction Challenge is my first time participating in a data science competition. I learned a lot in this experience and It was highly enjoyable and interesting.

Sources of data:

- http://www.pesmaster.com/
- https://fifa.com/fifa-world-ranking/ranking-table/men/index.html
- https://www.flashscore.com/football/world/friendly-international/results/


2018
load basicsdata.csv as dataframe
extract data from websites(fifa,PES..)
Merge data into a single dataframe
add new features
Clean data (drop useless cols..)
Build Model by using logistic regression algorithm
use model to predict futur games
*0.5
2017
load basicsdata.csv as dataframe
extract data from websites(fifa,PES..)
Merge data into a single dataframe
add new features
Clean data (drop useless cols..)
Build Model by using logistic regression algorithm
use model to predict futur games
*0.25
2016
load basicsdata.csv as dataframe
extract data from websites(fifa,PES..)
Merge data into a single dataframe
add new features
Clean data (drop useless cols..)
Build Model by using logistic regression algorithm
use model to predict futur games
*0.15
2015
load basicsdata.csv as dataframe
extract data from websites(fifa,PES..)
Merge data into a single dataframe
add new features
Clean data (drop useless cols..)
Build Model by using logistic regression algorithm
use model to predict futur games
*0.1
Combine predictions to have single prediction