

# Prediction of Betting Odds for LaLiga matches

Cai Badal Regàs  
Machine Learning from Data  
January 2020,  
Barcelona, Spain

**Abstract**—The main aim of this project is to be able to, as precisely as possible, predict the betting odds for any LaLiga game. The odd that is pretended to be predicted is the final odd of the betting house *Bet365* just at the beginning of the match. The main aim of the research is to exploit the variability of the odds according to users bets and track the best bets and moments to bet.

**Index Terms**—LaLiga, Betting Odds, Regression.

## I. INTRODUCTION

Bets and gambling have shown an impressive growth during the last years. According to the *Dirección General de Ordenación del Juego* of Spain, the amount of money used on online (71% of the total) gambling has been multiplied by 4 during the years 2012-2015. As it can be seen in the graph bellow, this stats are lead by sports betting, with 8.563 millions of Euros spent in 2015.

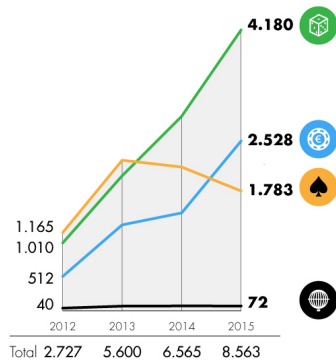


Fig. 1. Online betting in Spain (in millions of Euros) according to the *Dirección General de Ordenación del Juego*

Betting houses offer an amount of price to users depending on the likelihood of the bet to happen. This amount of price are initially determined by the house but move around depending on what users are betting. In this project we will try to predict the final betting odd that the *Bet365* betting house will offer and try to make a decision to find which bets are better and can give bigger benefits. It will be based on the offers of the winning team or draw for LaLiga matches from season 16-17 until the matches played in January of 2020.

### A. What are betting odds

Apart from many other options that won't be taken into account in this project, betting houses offer a betting odd for the 3 possible final results of a match: 'H': the local (home)

team wins, 'A': the visitant (away) team wins or 'D': draw. The offer is a number greater than one which represent the amount of € a user gets for guessing the result for each 1€ bet. The following table shows what bets represent:

Leganés v. Barcelona			
	Home	Draw	Away
Result			
If you bet 10€ you get	10x8.00 = 80€	10x5.25 = 52.5€	10x1.33 = 13.3€
Probability	100/8.00 = 12.5%	100/5.25 = 19.05%	100/1.33 = 75.19 %

Fig. 2. Meaning of betting odds

As it can be seen in the table the sum of all the probabilities is slightly bigger than 100%. This overshoot is the one that ensures that the house gets a benefit. Since the probabilities are inversely proportional to the benefits for users they ensure this way a benefit. In this case, the sum is 106% but it can even go around 109%. However, the benefits of the house are mostly around the 5%.

### B. Odds variability

The way the betting house put values to each bet is by following the statistics of what people is betting. They initially offer some odds for each result chosen by some analysts of the betting house of what they expect the probabilities to get. However, this odds can vary a lot before the match starts depending on what is people betting on: the considered probability of the teams to win is almost the same than what people are betting. This way the betting house also ensures winning by giving money according to what they are earning.

In this project we will try to exploit this variability of the betting odds. What our model will do is to predict where the odd will end to know the final offer of the house for every result. This way, we know how much money we should get paid for every bet before than other users. For example, in the example of the *Figure 2*, if our model says that the Away Team should be paid with 1.40, this bet would bad to do, since our models says that the odd should be higher (it is low payed for the high probability it has).

On the other hand, it is very likely than the Away Team odd that our model is going to be higher than the offered one. This would be it should be a good bet to do. In the project

we will try to find a reasonable criterion to know where a bet is worth or risky enough to take.

## II. DATASET DESCRIPTION

In this project we will generate our own dataset build around 2 existing datasets that will be presented in this section. We are based in LaLiga matches from season 16-17 until the matches played the beginnings of 2020. This dataset can be generalized to any league in the world if the following data is available.

### A. LaLiga results

In order to create the dataset is required LaLiga results from 15-16, a season earlier than what is going to be used to train the models. From all of this matches we require the result, the teams and the date of each match. What is also required and will be used as targets for the model ( $\hat{y}$ ) are the betting odds of the betting house at the start of the match. All this data is available in.

From this data set we will create a dataset with statistics about the past of both home and away team. We will add two types of information: actual season and last season. The first one consists of how is the team doing in the actual season until that match, like for example, the percentage of points, goals scored, conceded goals, etc. For all of them we will find the mean, mode, median and max, and also how each team does as local or visitants, for example, the mean of score goals of the visitor as visitor in that season (*AwayTeam-AwayGoalsMean*).

The second segment of this part of the dataset is the information about how the team did last season, including information like the position in LaLiga, the goals scored and conceded, among many others. This is the reason why we need the information of a year earlier of the start of the generated dataset.

There was conflict for the teams that each year are promoted and descended from the second league in Spain. The solution to this cases was that for teams who had already been in LaLiga the data of that season was used. For the ones that didn't have a past season in the dataset were given the stats of the team they replaced. For example, Girona FC, who finished 2<sup>nd</sup> in the second league and got promoted in season 16-17, it was given the stats for the previous year of CA Osasuna who ended 19<sup>th</sup>.

### B. FIFA Dataset

FIFA is a football video game that has been for years the most popular football simulator. The video game is based on the real players who every year get a numeric rating of their abilities: definition, shooting, tackling, sprint... For each match we will use as inputs this information to let the model know which team has the better players according to this

rating.

The inputs computed are going to be divided per position (defender, midfielder and attacker). For each role we will input the maximum, medium and minimum of each team. We will also add information about the top 11 players of each team which is supposed to show the information of the players on the field.

## III. REGRESSION METHODS

### A. XGBoost

### B. Neural Network

### C. Other methods...

## IV. RESULTS

### A. Metrics

### B. Regression results

### C. Real results

## V. CONCLUSIONS