# Footballer value prediction

*Edited by*


TRANG Thomas
YAYA-OYÉ Ikhlass


May 2021

# Contents

# 1 Context

Football is no longer just a sport, but also a major business. Player transfers are important for the clubs both in a sporting and a business sense. In the market, many millions of dollars are often paid for football players. Up to now in professional football, market values have primarily been estimated by sports magazines, or fans on internet platforms.

Both experts and fans estimates have repeatedly been shown to be highly accurate, but the estimation process is usually extremely complex. Market values are therefore often only updated once or twice a year. Recently, the footballer Kevin De Bruyne negotiated a 83 million pound contract at Manchester City by using data analysts.
Through this project, by using player's available data and artificial intelligence, we have implemented a conjecturing model in order to predict the market value of a player according to his personal football skills and attributes.

**Are the personal skills of a player enough to give a good prediction of his market value ?**

In this paper we will briefly explain our work : some decisions we made, our application (the model and the API) and the limitations of our AI solution.

# 2 Web Scrapping

To predict player's value, we needed data which could permit us to inter compare players. So we thought about personal skills. Given that there is no open source dataset containing the data we were looking for, we decided to collect web data.

## 2.1 Libraries & data source

The data we decided to work on is Fifa statistics on players. In our opinion, fifa statistics are universal to every players so that all players would be evaluated the same way. Furthermore, we did this choice because there are enough statistics on Fifa to highlight each player's abilities.

We chose the *FifaIndex.com* website in which you can find all data corresponding to numerous player even unknown ones. As the pages on which we wanted to scrap needed data are static, we only use requests to get access on a page and BeautifulSoup for retrieve the html of this one from which we extract features.

## 2.2 Collected data

We decided to collect very accurate player attributes in order to get the most exact features possible by categories :

- **Player description** : Age, Market Value, Salary
- **Technical skills** : Ball control, Dribbling
- **Defense**: Man to Man marking, Sliding tackle, Standing tackle
- **Aptitude**: Commitment, Responsiveness, Placem. offensive, Intercept, Vista, Discipline
- **Shots**: Header, Shot strength, Finishing, Long shots, Effect, Penalties, Precision Free kick
- **Physic**: Pace, Endurance, Strength, Balance, Speed, Agility, Vertical jump
- **Passes**: Cross, Short pass, Long pass
- **GoalKeeper Abilities**: Placement, Dive, Hand play, Kick play, Reflexes

# 3 Exploratory Data Analysis (EDA)

**Data processing & creation** : We created 2 more variables in order to do Data Exploration. The first variable deals with the player's level. It classifies the level class of the player based on the number of good ratings a player has out of 100. Another variable we created is the age class of a player, it also classifies the range age class of the player.

**Data exploration** : In this exploration, we showed the correlation and many plots between many variables and the target which is the market value. Among these variables, we saw the impact of the level of the player, his age, his wage and the main attribute of the footballer.

All the details concerning the creation of these variables can be found ***on this notebook of our GitHub (click here to access)***

# 4 Model prediction algorithms

To predict the market value of the player we use several statistical techniques, mathematical functions and algorithms which are parts of Machine Learning : a general term to describe when computers learn from data, recognising pattern and make right decisions.

About Deep learning, it is a sub part of of machine learning, based on artificial neural network which are more complex and can recognise deeper pattern.

## 4.1 Machine Learning

We have applied machine learning-based algorithms that predicts the cost at which a player can be sold in the football world market. We estimated the players' selling price using their personal performances and skills. Tests were carried out in various machine learning models like Linear Regression, Random Forest Regressor, Support Vector Regression (SVR) and Extras Trees Regressor.
In the notebook, you can found out that among these algorithms, Extras Trees regressor and Random Forest regressor gave best results for predicting footballers' prices as you can see in the reporting below.

| | Score R2 |
|---|---|
| **model** | |
| **Random Forest Regressor** | 0.861054 |
| **Extra Trees Regressor** | 0.856780 |
| **Multiple Linear regression** | 0.654429 |
| **SVR** | 0.389134 |

## 4.2 Artificial Neural Network

During our training we tried three different neural networks. The best one was the deepest with 5 hidden layer using Batch Normalisation before the activation of neurons (Relu function) and using a 0,4 Dropout's rate.
With this model we reached a **$R^2$ score of 0.95** which is our best value. However, when we take a look at the prediction made by our neural network on "famous" player, it seems to be less efficient than the random forest model. Due to the numerous "weak" player in the data frame, the model have some complications to be accurate on top player and others at the same time.

# 5 Visualise the model's results

Let's compare the market value of a player to his predicted market value by our algorithm.

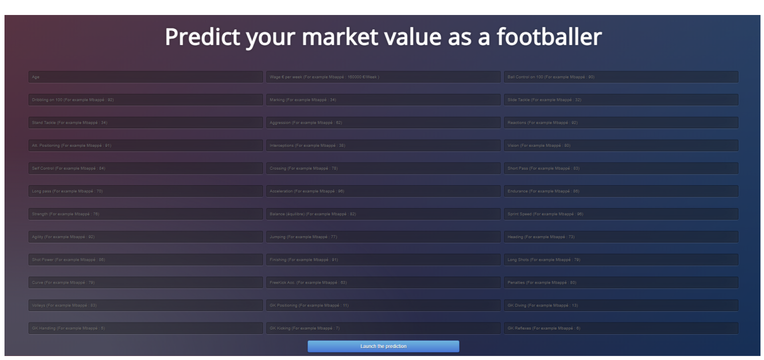
| | Name | Value | Prediction | Difference |
|---|---|---|---|---|
| 0 | Manuel Neuer | 20,500,000.0 | 45,163,263.0 | 24,663,263.0 |
| 1 | Lionel Messi | 103,500,000.0 | 106,100,685.0 | 2,600,685.0 |
| 2 | Jan Oblak | 120,000,000.0 | 56,541,667.0 | 63,458,333.0 |
| 3 | Kalidou Koulibaly | 76,500,000.0 | 67,397,540.0 | 9,102,460.0 |
| 4 | N'Golo Kanté | 78,000,000.0 | 74,661,121.0 | 3,338,879.0 |
| 5 | Alisson | 88,000,000.0 | 84,346,144.0 | 3,653,856.0 |
| 6 | Toni Kroos | 87,500,000.0 | 85,396,846.0 | 2,103,154.0 |
| 7 | Erling Haaland | 122,500,000.0 | 87,795,486.0 | 34,704,514.0 |
| 8 | Keylor Navas | 33,500,000.0 | 41,783,791.0 | 8,283,791.0 |
| 9 | Bruno Fernandes | 121,000,000.0 | 114,094,254.0 | 6,905,746.0 |
| 10 | Karim Benzema | 83,500,000.0 | 81,528,524.0 | 1,971,476.0 |
| 11 | Marc-André ter Stegen | 110,000,000.0 | 93,693,202.0 | 16,306,798.0 |
| 12 | Harry Kane | 123,000,000.0 | 112,268,845.0 | 10,731,155.0 |
| 13 | Joshua Kimmich | 110,000,000.0 | 94,659,516.0 | 15,340,484.0 |
| 14 | Sadio Mané | 92,000,000.0 | 106,175,259.0 | 14,175,259.0 |
| 15 | Robert Lewandowski | 124,500,000.0 | 112,236,500.0 | 12,263,500.0 |

We can see that the Random Forest model has a pretty good quality of prediction except for Goalkeepers. Indeed the model has little difficulties to predict an accurate value for GK because there are not that much features for GK, that's why the model is not that accurate for them. Furthermore the model might under evaluated players like Erling Haaland because it didn't take into account that he is only 19, with that high attributes which is rare.

# 6 Deploying a Flask Application

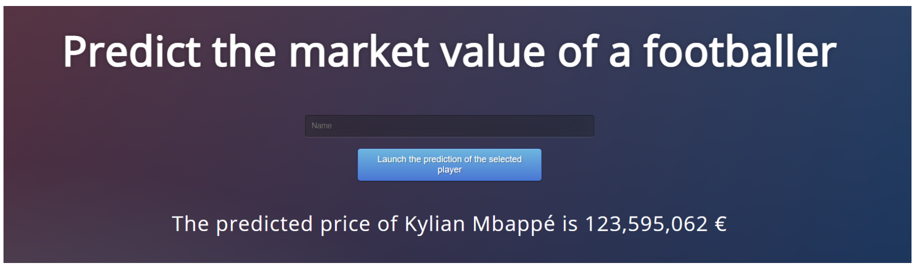
To present our work in a friendly way we implemented our selected model into a Flask API accessible online (hosted by Heroku) with two possible actions for the user.

- **On the first page**, the user can enter all attributes on 100 by himself to predict the market value of a wanted player. For clubs that would like to estimate their footballer's price, they could enter all the attributes of the player and get his predicted price.




**Predict your own market value**

- **On the second page** (accessible with the navigation bar), he can type the name of a real footballer to get a price prediction for this player. There is an auto completion function which proposes known player. On the example below, we selected Kylian Mbappé and see that his predicted price is 123 595 062 million €.




**Choose a footballer and predict his price**

# 7 Conclusion

We finally selected the Random Forest model. Deeper neural network could be efficient but we made sure that our application keeps a good time performance and predicts fast and accurate results within 3 seconds, helping club investors make quick decisions during the mercato. The application can also help clubs estimating their personal players' price by inputting their attributes.

To conclude, this project proved us that players' fifa attributes permitted to predict accurate market prices for players as we saw in the part 5. However we found out that the model have a little bit less accurate predictions for GK and youngsters. Furthermore, in real market transfer a player value is also determined by some marketing factors which are not take into consideration by our model.