BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2019
http://excel.fit.vutbr.cz

Footie: Web Simulator of Football Leagues and Championships

Martin Urbanczyk

Abstract
The goal of this project is to create a web football simulator, which would simulate realistic matches of the football world. The concept of the whole simulator is complex. It generates future results based on real historical match results. However, unpredictable and crazy results are also a part of the fun. Each user has its own simulation and he or she can have completely different results, which are more and more diverse year after year of the simulation. Because of the complexity of the football world, the simulator supports only selected football nations. The main focus of the project is to improve the simulation of matches. Calculation of match results is based on real historical results and it uses neural networks. It is mixed with random factors described in the paper. The project focuses on the group of football fans who would like to only watch simulations of the football world without the need to manage teams and players. This is something that is not available in the current market. Every game focuses on management mainly, not the simulation. The core of the simulator could be also used for trying to predict real football matches because it is based on real data. This paper describes base concepts of predicting a football match. It presents football systems and shows, how the simulator game is different from existing solutions.

Keywords: Football — simulator — machine learning — match prediction

Supplementary Material: *N/A*

*xurban61@stud.fit.vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

The main value of this project is a system predicting match results based on real data (real results). It is connected to a simulating system which includes random factors and provides a unique evolution of the football world, based on current data. These random factors are not just random functions. They are based on some potential motivating factors, which exist in the real sports world. A user who would like to see scenarios of the next football season can have satisfaction in this game same as the user who likes sci-fi scenarios and would like to see generated football world for example in 2100. The core of the prediction can be also used for prediction of match results in the current season in the real world.

The current version of the system simulates the biggest European football nations (England, Spain, France, Italy, and Germany) and the Czech Republic. It also contains the biggest club competitions like Champions League or Europa League, which use the UEFA coefficients[1] for recalculation of nations' ranking after each season. The ranking is used to calculate the number of participating clubs for every nation. Hence, it is possible that some smaller nations might have

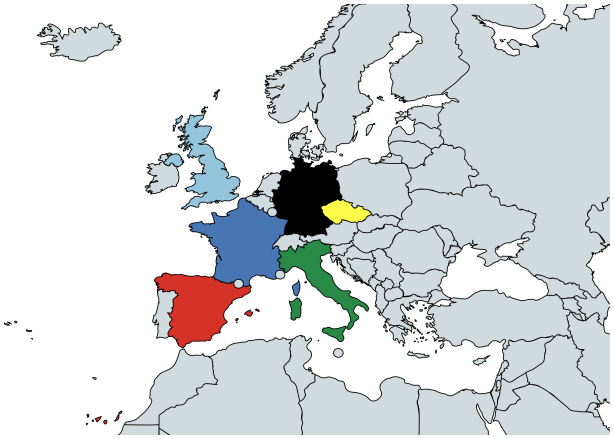

Figure 1. UEFA nations which are included

more teams the competitions in future seasons. The data of all world leagues are included, but the project is pretty complex. I chose to focus on the match prediction. Figure 1 present map of UEFA nations, which are included in detail simulation.

The system needs to be entertaining and playable for football stats maniacs as well as for casual fans who would like to see how their favourite team evolves in the nearby future. The project does not guarantee fun for every user. It focuses on the people who are interested in football and its future. It offers something to either casual fans and bigger experts who have experience with playing manager games and who are interested in deeper stats. Who knows, maybe in your simulation Zbrojovka Brno will play Champions League final in the year 2100. It is also planned to create a community hub where users can share their results and compare the state of their world.

The crucial task of the project was to find the key features of the prediction. I also had to find the *soft* features which might also have an impact on match (e. g., motivation, frustration, form, or match importance).

2. Analysis

Most of my inspiration came from playing management games and detailed watching of the football world. The biggest football manager is the game from SEGA called Football Manager 2019 [2]. It is a pretty realistic game but it focuses more on the management factor. No matter which team you choose, the world automatically changes due to your actions. I took some inspiration from its user interface (Figure 2). I also used to play Czech Soccer Manager [3] in the past. The game is not so huge and complex as FM 2019, but I took some inspiration from it's easy going UI and its simplicity.

None of these games and projects focuses on realistic simulation. You can get some crazy results

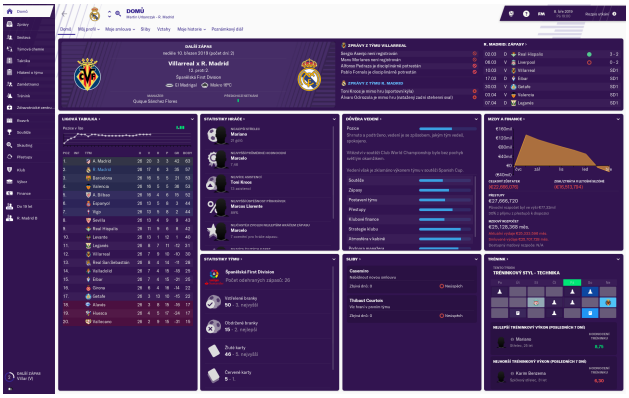

Figure 2. FM 2019 main dashboard [2]

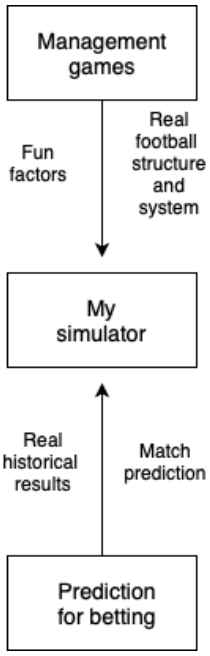

Figure 3. Difference between existing solutions and my simulator

practically immediately. It is alright when you play a game just for fun. On the other hand, there are also some projects which focus just on match prediction mainly for betting [4]. These projects do not focus on the simulation of a hypothetical world. Figure 3 presents how I mixed my project from these types of existing solutions.

3. Design

The solution is on the halfway to both analyzed categories. It uses real prediction of the results. Then, it adds the random factor to make the game more entertaining. Results of the matches are close to the predicted results but they might be different for every player. This helps to make the evolution of the football world diverse and crazy in the future. The simulator, however, does not force players to plan tactics, buy or sell proper players or negotiate about their salaries. It focuses more on the simulation of the games than any

other game does.

For example, if players of Football Manager go for one year lasting holiday (the active management of the team is interrupted), they will see more diverse results in the league they chose. Just because they chose it. For example, you took the manager seat in England. Even if you do nothing, there will be some shocking results that are not so realistic. However, in a league where you are not a manager (for example Germany), results are still going to be similar (for example Bayern Munchen will win 10 titles in 10 next seasons). There will be some unpredictable results but not more in some region just because you choose it.

If the player starts a new game in Football Manager, they will see more diverse results in the league they chose. Just because they chose it. Other leagues evolve too, but the change is far slower. So it is not quite realistic. But it is mainly a game, so it is understandable because players want to go higher fast. For example, when a player begins a game in England, weaker teams can finish higher in England. Although, in Germany or Spain it is different. Every time it takes more than a decade when some weaker team wins the league.

However, I plan to add some extensions in the future, that will allow players to add their teams, players to real leagues, or make some transfers to see how things might evolve (e. g., What would happen if Messi and Ronaldo play in the same team that is on the bottom of the table).

All of the biggest football manager games are available for the only desktop platform. I decided to make a cloud solution, so you can connect everywhere and it play it on (almost) every device, that has an internet browser.

4. Implementation

4.1 Architecture

The architecture of the system is displayed in Figure 4. The system is structured into modules which are not dependent on each other. Each functionality has it separate control (i.e. team screen - controls for players, fixtures, stats). This makes the system extensible and scalable. In the future, it is planned to provide subsets of the modules to users according to their purchase plans. They will be able to choose the modules according to their interests (e. g. leagues, nations).

I implemented the system as a web application in PHP 7.2 (Nette framework) and JavaScript mainly for asynchronous loading and user actions. It has a lightweight user interface. It is mobile friendly and it does not overwhelm users with stats and data (the detailed data are accessible in specific pages). The system uses the MVC architecture and it supports REST API written in node.js for the connection with the database. It is expected that native mobile applications will be implemented in the future.

The system uses the standard MySQL database. The database has two parts. The first part consists of "hard" data which will not be changed in the future and are same for all users. For example, it contains example names of players and teams, the number of teams in each league etc. The second part differs for each user. It consists of actual data of a simulation (e. g., player goals, team history etc.). The whole database contains a lot of data (for example, around 80 000 players).

Also, I needed to do a lot of parsing and merging of data, because of different sources. It was necessary to merge the Football Manager database, my system and database and api-football data. I used mainly Python scripts for this.

4.2 Match Prediction

The match prediction is provided by using the Python scikit-learn library [5]. It uses the neural network which was trained with the data gathered from the api-football.com service [6]. The service provides a REST API for getting real results. The results are mixed with the results of the simulation (in the first years of simulation). The main features of machine learning are:

1. **Team rating**: the actual rating from the database; it is calculated after each season from the results.
2. **Actual team form**: the last 10 matches played by each team.
3. **Form at home/away stadium**: the last 5 matches played at home and away from fields, depending on which team is the home team and which is away team.
4. **Last matches between the two teams**: the matches between the two teams during last five years (not the 5 last matches, because it is not relevant if these two teams played together 15 years ago under completely different circumstances).

I made a dataset of historical results and tried to predict some older games. I have a lot of different data from all over the world. I did a lot of experimentation with a different type of matches. I took also some official odds for comparison. It was provided by the api-football API, but I also did a comparison with Fortuna's odds [7]. I had around 85 percent of

successfully predicted results of the matches containing a strong favorite. The success of prediction means the right prediction of win/draw/loses. It was worse for mixed datasets, where it was around 65 percent of successful predictions. Then, I did a competition with Fortuna on near planned football matches through the whole weekend in the biggest football leagues, which are mentioned upper. The most important comparison here was between my prediction and odd prediction (my prediction versus the lowest odd on each match). Betting company is still a winner, but not too heavily. In each iteration, they beat me just by a few percents (average is 6 percent, which is 3 matches on average—from 49 weekend matches). I chose a polynomial regression model and the rate between training and testing data is 0.8/0.2. The final model is extracted, but it requires preprocessed data from the system and api-football API.

4.3 Prediction with Randomness

The simulation of future matches uses mostly historical simulator's data (it calculates new results from the results of the last 10 matches and last 5 years). Some random factors can have an impact in later years. These main random factors are used mainly to simulate extra motivation, or frustration, or importance of upcoming matches which can produce an unexpected score. It generates the rank of "randomness".

Then, the final result is mixed with the result generated by the neural network. The algorithm works with a tape of numbers representing the random difference from the predicted results. In the beginning, there are mostly the numbers 0 and 1 (but there are also higher numbers). If the level of randomness should be higher than I add higher numbers (2, 3, 4, . . .). Some of the numbers of this tape is chosen for each team. The chosen values can add or remove some goals which were predicted by the neural network. Mostly, the result will be similar to the prediction, but there can be a situation (which happens also in real football world) when a result is shocking. The goal is not to have still the same results, but it is not going to be completely random. Hence, I take the result of the prediction, generate the random value and use the value to change the originally predicted result.

5. Testing

5.1 Target Groups

Last year, before the beginning of the work, I spoke with a group of different people who are involved and deeply interested in football or football management games. I asked them for help with testing the system. I tried to ask both experts and casuals. As an expert, I mean a person who deeply understands football leagues structure and watches more than just his favourite teams. As a casual, I mean a person who is just a fan of some teams and is not really interested in complex statistics etc. The website is going to be free for all, however, if users would like to edit the date for their own there will be a small fee for this addition.

5.2 Actual testing

Firstly, when I finished the UI templates, I asked the users for feedback and they gave me a lot of hints. Currently, they are giving me feedback about the actual functionality of system and prediction. It helped me to fix some bugs.

Many of the users made long-lasting simulation so they gave me feedback on how the football world can evolve in the future. Such testing would be time-consuming for one person. Their simulations completely differ in later years but had some common results in the nearby future. After decades of simulation, some football giants have fallen into lower leagues and new giants were born. It is quite different for everybody, but this is part of the fun. The system is currently not publicly accessible. It is planned before the deadline of the thesis (end of May). So I gave users my laptop for a few days to check the simulator.

6. Plans for the Future

The project is developed as part of a master thesis but I plan to continue with the development afterwards. Firstly, I plan to add all European leagues (at least first and second level). Secondly, other football competitions will be included—e. g., international football (World Cups, European Championships and its qualifiers). And also, as I said before, I plan to add some customizers for users. It will be possible to edit/add their own teams, players or make some transfers so the users can check how the world changes after. In the year 2020, I am going to move the application to other platforms (Windows: desktop application, iOS and Android). It is possible because I made universal back-end using REST API. However, users need to be connected to the Internet on all platforms.

7. Conclusion

This paper provides information about Footie—the web football simulator. The simulator generates match, league and season results and is not limited by a number of seasons. It is practically limited just by the amount of data which can be stored.

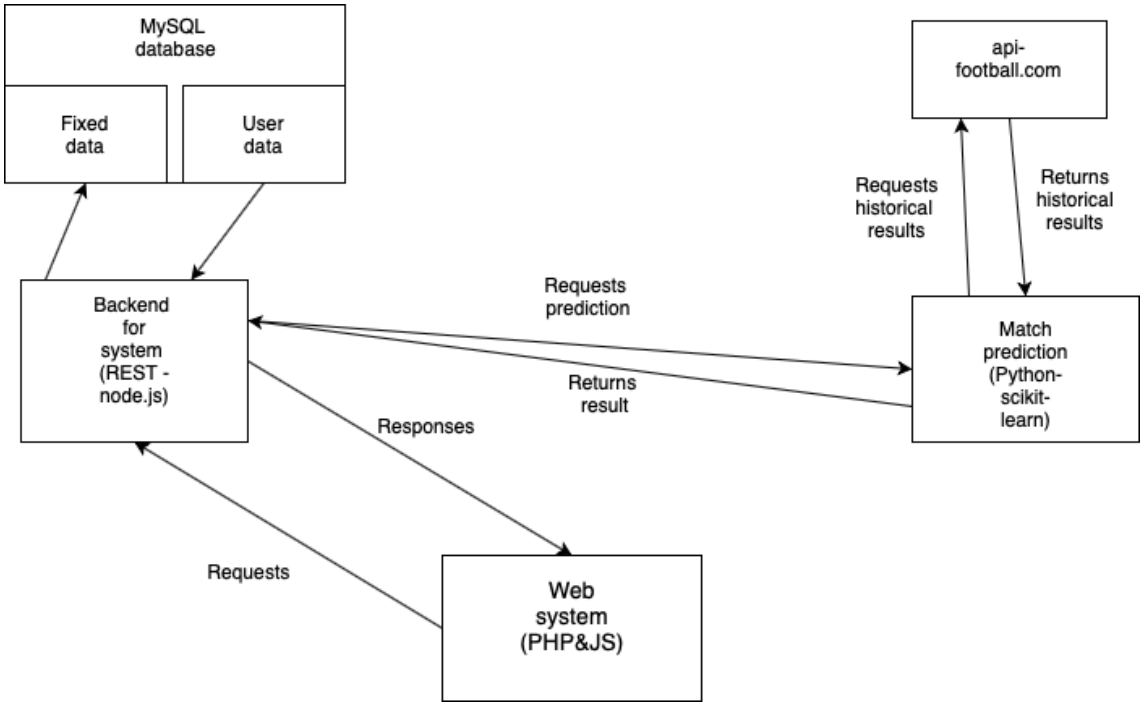

Figure 4. Architecture of the system

In the beginning, it was important to define the main idea of the system. I had a lot of concepts. I tried to compare my concepts with existing solutions (mainly football management games and match prediction for making bets). Then, I designed a which is probably on the halfway to both categories.

Then, I made an architecture that includes multiple modules. The main modules are the web system, back-end API, neural network and API with real historical results. The main goal was to create match predictor which combines the predicted results with regulated random factors in order to bring both realistic but sometimes unpredictable and crazy results (as real football world does). It was tested on both historical and real-time matches with the help of betting odds.

At the current stage, the system includes basic simulation of the biggest European leagues and European club cups. The system is composed of modules, so it is not a problem to add or remove any module. The web application is designed responsively so it is runnable on all devices which has a modern internet browser. Its database is optimized as much as possible because of the amount of data. It included around 280 leagues and 70 000 players. All these leagues are going to be playable in the future. Currently, the app is still in development so it can be only seen locally, but at least its first phase prototype is going to be released by the end of May 2019.

References

[1] Bert Kassies. *UEFA Coefficients calculation method.* https://kassiesa.home.xs4all.nl/bert/uefa/calc.html.

[2] Sports Interactive. *Football manager official website.* https://www.footballmanager.com.

[3] *Czech Soccer Manager official website.* http://csmweb.net.

[4] Jaroslav Bezděk. *Aplikace metod strojového učení na predikci výsledků fotbalových zápasů.* 2015.

[5] SciKit-Learn Developers. *scikit-learn user guide. [Online; 16.3.2019].*

[6] *Api football REST API.* https://www.api-football.com.

[7] *Fortuna offical website.* https://www.ifortuna.cz.

8. Screenshots of the application

Figure 5. Main screen - list of nations

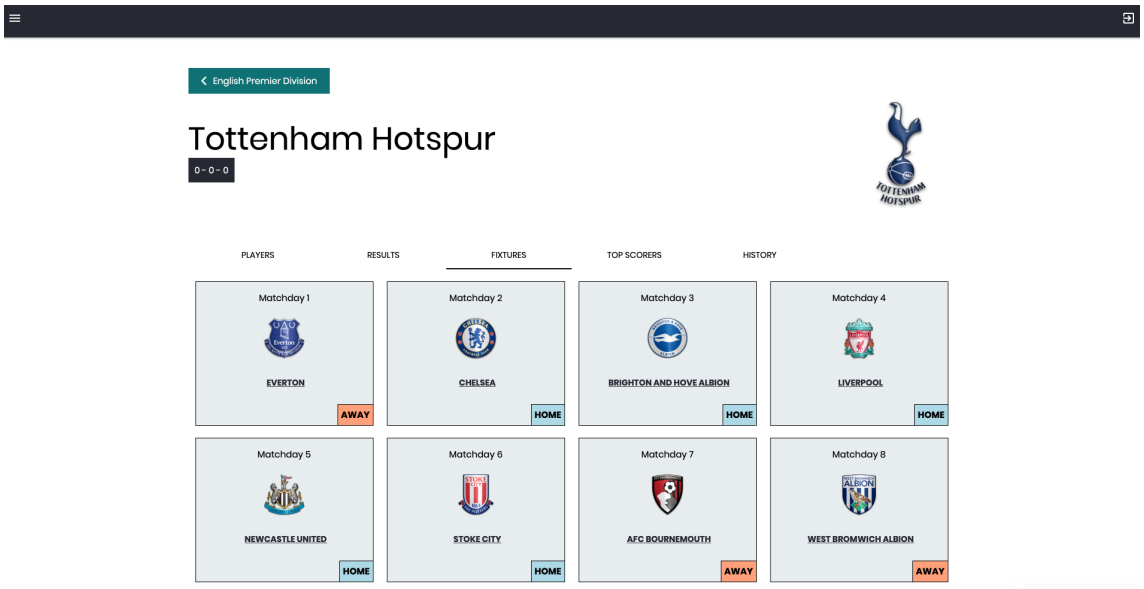

Figure 6. Team screen - upcoming fixtures

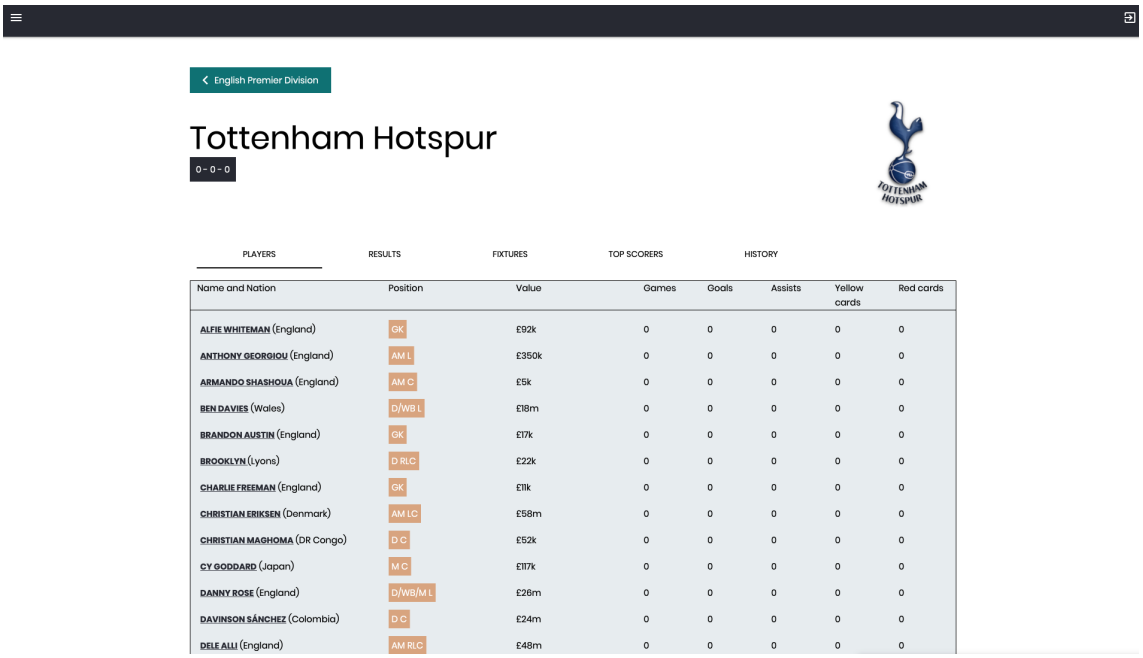

Figure 7. Team screen - tables with the players

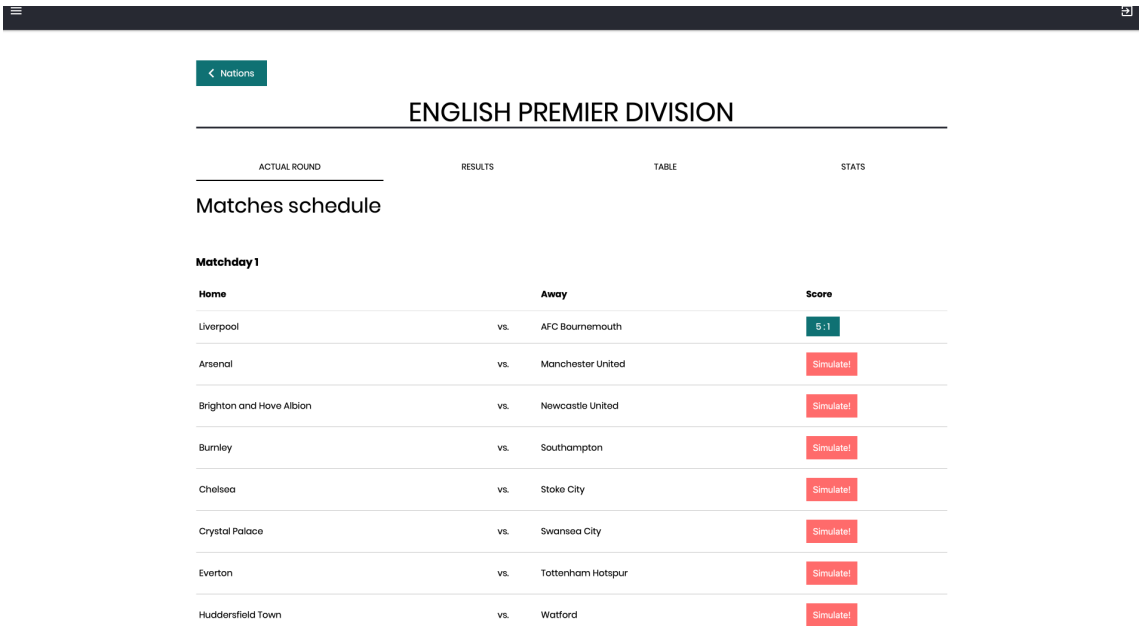

Figure 8. Competition screen - actual round results/fixtures

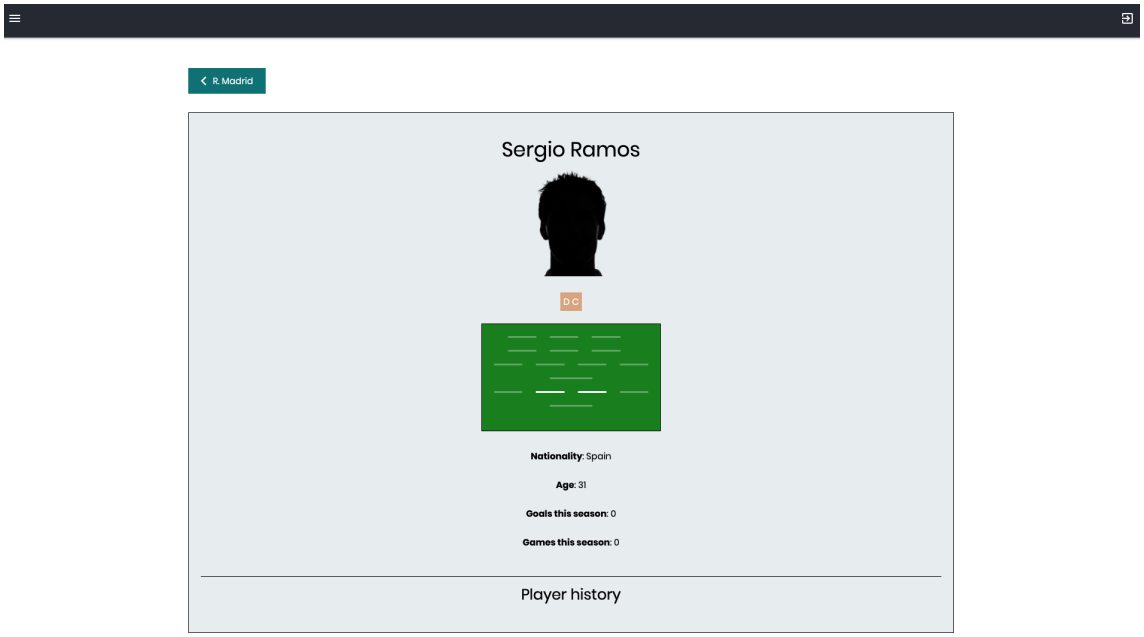

Figure 9. Player detail screen

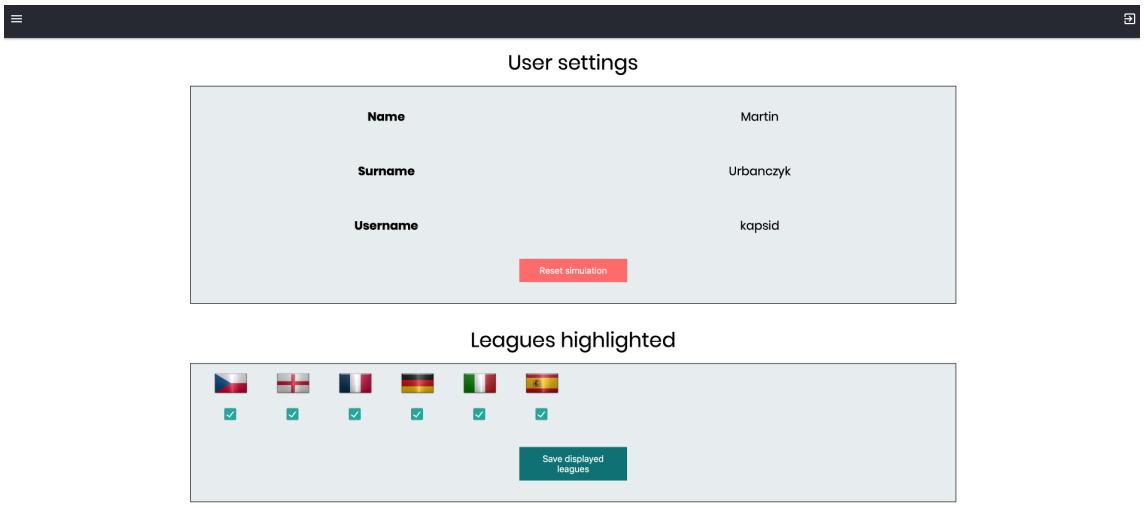

Figure 10. Settings of simulated leagues