FootStats - An Application for Statistics and Predictive Analytics in Football

Alok Rajpurohit
Dept. of Computer Engineering
MPSTME, NMIMS University
Mumbai,India
alok.rajpurohit76@nmims.edu.in

Rushang Phira
Dept. of Computer Engineering
MPSTME, NMIMS University
Mumbai,India
rushang.phira78@nmims.edu.in

Aniket Raman
Dept. of Computer Engineering
MPSTME, NMIMS University
Mumbai,India
aniket.raman79@nmims.edu.in

Shaurya Rawat
Dept. of Computer Engineering
MPSTME, NMIMS University
Mumbai,India
shaurya.rawat83@nmims.edu.in

Prachi Natu
Dept. of Computer Engineering
MPSTME, NMIMS University
Mumbai,India
prachi.natu@nmims.edu

***Abstract*—Data analytics in sports and specifically in football has tremendous potential. At present, an increasing number of teams in top leagues are implementing this in the hopes that this strategy can lead to improved performances on the pitch. This paper outlines a web application that serves to provide detailed statistics on almost all football games with international coverage, along with certain functionalities making use of machine learning. Using this application, users can make predictions for the market values of players with reasonable accuracy. Other functionalities include a player similarity module, where users can check how similar players are to one another based on their abilities, position on the pitch, and their potential for improvement. Such a metric can be useful for finding replacements for players in teams. Finally, an analysis of expected goals is provided based on events within selected games. The data was collected through scraping and data-sets available online while the application software was created using the ReactJS and Flask frameworks.**

***Index Terms*—Football Analytics, Football Statistics, Expected Goals, Player Similarity, Machine Learning in Football**

I. Introduction

Analytics as a field has grown to be instrumental because of the competitive nature of sports. Sports analytics is frequently relied on by players, team managers, coaches and fans before making decisions or developing strategies to win games. Today, there is an expansive amount of statistics available on just about any competitive sport. Data science is not just used in sports to fuel competition between professional players; it also plays a key role in improving game quality, fan experience, and player safety. Students with advanced data science degrees may find themselves delivering critical services to their favourite teams and helping to transform sports for a new generation as a result of these growing applications. Teams in the world of sports and especially football are constantly developing their ability to use sports analytics as a tool to improve their win rate [1].

Sports analytics is defined as the use of data related to any sport or game. This data includes player statistics, weather conditions, a team's previous wins/losses, and so on. Using this data, we can build predictive machine learning models to help managers make educated decisions. The primary goal of sports analysis is to improve team performance and increase the likelihood of winning a game [1][2]. The value of a win speaks volumes and takes on different forms like the fans filling the stadium, television contracts, fan store merchandise, sponsorships, new transfer attractions, player retention, and local pride.

II. Background

Role of Analytics in Football: Historically, coaches could rely on little more than pen, paper, and mental prowess to aid in analysis of a game. , Manchester United were one of the pioneering teams in the '90s to utilize analytics; over the next few decades, it became much more widespread internationally. In recent years, football experienced rapid technological advances, with platforms able to capture and analyze data from training, match play, IoT devices, and wearables. Coaches now rely heavily on metrics and analytics to guide their decision-making and help their teams excel [3]. The use of data and technical analysis has become essential for today's football clubs [4], and it is no longer confined to the biggest and richest teams. Smaller clubs are making use of it as well thanks to software being cheaper and more readily available than in the past. An example of such tool is optical tracking, which can be used to pinpoint the position of players on the pitch 25 times a second, in relation to the ball, opposition, and teammates. Ball-related data such as passes, shots, and turnovers can be used in conjunction with more advanced tools that can analyze defensive stability, pitch control, and off-ball scoring opportunities. Real-time information is of most value to coaches [5]. They can use this data to change the formation of their teams and work on their morale to increase their likelihood of winning [6][5]. These types of insights are gathered from playing logs, video, GPS tracking, and spatially related data.

Teams can also be examined as a whole, using nodes to represent players and lines between the nodes to show interactions, such as passes between teammates. This graph-like structure can assist coaches and data analysts in identifying and encoding various types of interactions and events. This data allows them to identify, change, and test the effectiveness of typical passages of play. Data analytics can also play a huge role in youth development [7]. Having objective and measurable feedback could help both coaches and players track development. This can help make reasonable estimates about players' potential and how to cultivate them [8][9].

Predictive Analytics: Predictive analytics are what can help teams adapt for the future, thus helping coaches understand the consequences of altering a team's structure in-game. Coaches can tailor training, strategy, and individuals' roles according to data about their next opponents.

A metric quickly gaining popularity has been discovered to have room to make an impact is expected goals (xG), which measures the quality of players' shots in attacking play and the probability the shots will result in goals. xG uses algorithms that take into account factors like the distance from goal, angles, and more [10]. Football teams could use this metric to optimize their patterns of play to get their players into the best possible position in order to score goals as well as prevent conceding them. In this case, analytics provide insights into the most effective strategies to apply in different situations. Besides helping with in-game strategies, data can be put towards player recruitment as well. Helping scout players with high potential for the future, finding suitable replacements for outgoing players are some useful applications. Thanks to the exponentially increasing amount of data, these applications can only get easier and more accurate in the future [11].

III. Literature Review

In [10], the authors look to examine the credibility of xG and whether it maps well against real goals. The data set used is of shot locations (x and y coordinates) across the 12-13 season from the Premier League and the Bundesliga (18218 values). They then split the pitch into 8 zones, each indicating a shot location, and examined the probability if a shot in that zone would be a goal. The expected and actual goals have a high relationship value and hence the statistic is credible. [12] combines data from two sources (from OPTA for stats and FIFA for player ratings) to create a sophisticated model. Concepts were created manually and with the eye test and validated with a flowchart. The next thing to do was classify shots by their type. All the variables used in the data set were listed. All of them were given a value/weight based on some of the data-set to predict further data. [13] tried to predict games and thereby set odds on these games. The parameters used are player performance, team performance, whether it is a home or away game. Based on the given data, we predict which team has what probability of winning while [14] proposed a prediction approach on the football data by implementing Machine Learning algorithms. The predictions are made for the player's position, the number of goals per season, and the number of shots per match. The predicted results are visualized by bar charts and compared to the real results. It was observed that the accuracy of the predicted results was good in all three experiments. The probability is calculated based on data from all football game results (home win/away win/draw), matches between the two teams in question, whether any player from either side has a decent record vs the other teams.

[15] examines the various approaches to predict the value players. It evaluates all the approaches and concludes the best one. The various data modeling approaches are stated. The crowd-based estimation model works on the parameters on which the public judges the value of the player. The next approach involves multilevel regression analysis, which involved the implementation of several models each considering different features as significant. The third approach involves the use of an index score to assess the player's value. An MLP neural network was implemented on the players' data with various variations of the hyper-parameters. In the end, it is concluded that the multilevel regression approach is the best among the others.

In [16], a model assesses a player's value based on performance data. Two data sources were used: one provides the player game stats and the other provides the estimated market value. A regression model is developed after performing feature extraction where the stats, physical data, and team-related information are included. The performance is compared to other methods. The paper also attempts to check the vice versa effect as to whether the market value affects the performance on the field but no conclusion is stated in this case. A data-driven framework was developed in order to predict the outcome of football matches which would generate meaningful profits by betting in [17]. This paper describes a way to estimate the outcome of football matches based on large data sources and common machine learning algorithms.

Time-series data compose a sequence of data that is collected at regular intervals over a period of time. In the case of [18], it is a set of data built from football match history. Players' performance and manager indices were gathered from the 2014-2015 season of the English Premier League. Only 53.3% accuracy was obtained using SVM, while an 85% was attainable using Neural Networks, according to the paper. Thus, it concludes that SVM is not an appropriate method for this purpose. [19] provides a data science technique for reducing the time required to pick a player for a team by taking cost and player talent into account as limitations. It gives a cost-effective statistical study of player performance based on abilities and skills for a new team by utilising powerBI and Python Pandas. The results suggest that it leads to increased corporate earnings by systematically improving football data-sets.

The study in [20] explored the possibility of using publicly available data to create straightforward mathematical models to predict the most likely outcome of a football match. In their approach, goals scored by a team are predicted almost independently from a team's opposition. We only used game-related data instead of trying to incorporate too many variables such as form, injuries, line-up, game importance, etc.

IV. Predictive Analytics

The end-product of this project is a web application. This detailed web app will provide details about footballing activities, statistics, and functionalities from the top leagues across the world. It will also provide some machine learning based functionalities. These functionalities include – prediction of player's market value, player similarity, and prediction of expected goals in a match. In this section, we have developed the machine learning models for these functionalities. The data for the same is compiled from some of the biggest names in football data collection, one can review and analyze an enormous number of data points, including goals, shots, dribbles, passes, and many more. These data points are available for matches in a host of competitions, including the Premier League, La Liga, Serie A, Bundesliga, Champions League, and Europa League. The detailed description of the data-sets and database used in this project is provided in the next section. The machine learning models are as follows:

A. Market Value Analysis

An important part of this project is predicting the market value of players based on various statistics. Typically, in reality, a younger player will cost more than an older one. If the player is already closer to retirement, a football player has fewer years left to keep performing at the highest level thus depreciating their value whereas a young player is perceived to be a long-term investment [15]. Many clubs make a lot of their income by developing young players and then selling them to a bigger club for a higher amount. Thus, the outcomes of predicting the market value are valuable in the industry. Two other equally important factors are the player's skills and the position on the field they prefer to play in. Attacking players, and those who score a lot of goals or give a lot of assists, tend to be more expensive. Higher-skilled players will have larger transfer fees [21]. Research shows that players who get higher average ratings tend to have higher transfer fees. Thus, we made use of a data set that encapsulates the points mentioned above. As a part of the analysis to be done, an extract is presented here. The data set used consisted of roughly 6000 players and information on their skill level based on the video game "FIFA". It consisted of statistics such as their interpretive statistics such as overall skill level, strength, salary (speculated) as well as real-life metrics like their weight, age, height, dominant foot, etc.

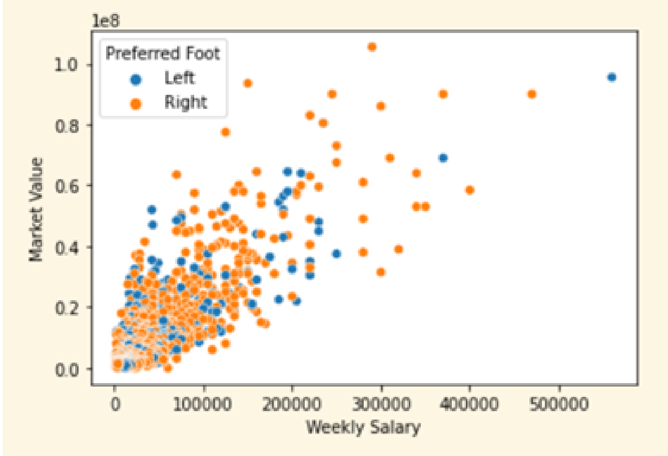

Fig. 1. Distribution of players' dominant foot in the data set

We applied several regression algorithms were applied to the data-set, primarily in order to predict the market values of players. To do so, we dropped attributes that did not contribute towards this goal such as 'Height', 'Weight', 'Preferred Foot', 'Market Value', and 'Player'. Preprocessing steps were also applied like scaling through the MinMaxScaler. The metrics for measuring the efficacy were the mean squared error and R2 score.

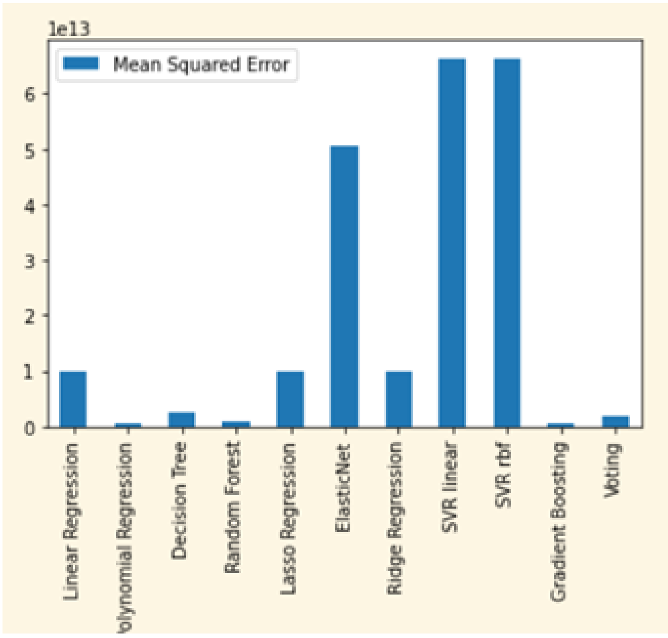

Fig. 2. Comparison of mean squared error of the applied regression algorithms

We find that polynomial regression, random forest and gradient boosting have the least mean squared error values. To get the best algorithm, R2 score comparison is also conducted. We find that all the 3 best algorithms have the same R2 score.

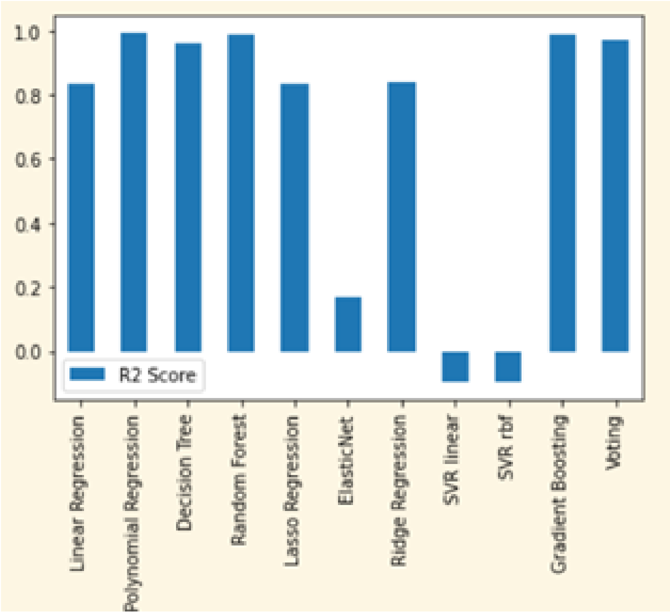

Fig. 3. Comparison of R2 scores of the applied regression algorithms

B. Player Similarity

Seeking out similar alternatives to certain players is a key element of scouting. Oftentimes, a targeted player may be under contract at another club which makes acquiring them nearly impossible. In these cases, finding alternate players with high similarity to these targets would be highly beneficial.

For this implementation, we used a data-set containing information of over 15000 players collected from the FIFA data-set. It involves attributes of players in detail.

Player similarity can be used to find players of a similar profile, especially if the player's age and rating are far apart. It can be used to find players who are significantly younger but have a similar player build as the player being compared to. It can also be used to dismiss the fact that just because two players are of a similar caliber, they must be of a similar type and either of them could replace the other or play in their position.

In Data Mining, similarity measure is the distance with dimensions representing features of the data object in a data set. The lesser the distance, the higher the degree of similarity, but when the distance is large, there will be a low degree of similarity.

Some of the popular similarity measures are – Euclidean Distance, Manhattan Distance, Jaccard Similarity, Minkowski Distance, Cosine Similarity.

Cosine similarity: Cosine similarity is a metric, helpful in determining, how similar the data objects are irrespective of their size. The adjusted cosine similarity measure is a variant of vector-based similarity in which we consider the fact that various users have varying rating methods; in other words, some people may rate objects highly in general, while others may rank items lower as a preference. The cosine similarity is advantageous because, even if two identical data objects are separated by the Euclidean distance due to their size, they may have a lesser angle between them. The greater the resemblance, the smaller the angle. When shown on a multidimensional space, cosine similarity captures the orientation (the angle) of the data objects rather than the magnitude.

Below, we have compared the best-rated player on FIFA 22 with one of the worst players. To the casual observer, they might conclude that there should be no similarity. However, in reality, despite the disparity in rating, there is a resemblance between them, given that the worse player is young has room to improve their attributes. They also play in similar positions on the pitch, making an 80% similarity rating justified.

```
dataSetI = new_data.iloc[0:1,:]
dataSetII = new_data.iloc[19234:19235,:]
dataSetIII= new_data.iloc[1:2,:]
cosine_result_1 = 1 - spatial.distance.cosine(dataSetI, dataSetII)
cosine_result_1

0.8001408605559942
```

Fig. 4. Cosine Similarity between the best and worst player in the data-set

In another example of cosine similarity working well, the players compared are Neymar Jr. and Noa Lang. Although Neymar Jr. is rated considerably higher, they are almost identical to each other. This is because both play the same wide positions on the field. Moreover, their stats relative to their rating are extremely similar in terms of pace, dribbling ability, passing, and more. This is why there is a cosine similarity of approximately 99% between them. In the real world, data like this could cause scouts to earmark Lang as a possible replacement or successor to Neymar Jr., thus helping out in the club's transfer dealings.

```
[ ] data1SetI = data.iloc[2:3,:]
    data1SetII = data.iloc[837:838,:]
    data1SetIII= data.iloc[19234:19235,:]
    cosine_result_1 = 1 - spatial.distance.cosine(data1SetI, data1SetIII)
    cosine_result_1

    0.9984635475822238
```

Fig. 5. Cosine Similarity between Neymar Jr. and Noa Lang

C. Expected Goals

For analysis of expected goals, we made use of a data-set containing shot events for various matches during the years 2012-2017. xG could give teams a better understanding of their performance, regardless of the result. This is because results are not always reflective of performance. Most publicly available football statistics are limited to aggregated data such as Goals, Shots, Fouls, Cards. When assessing performance or building predictive models, this simple aggregation, without any context, can be misleading. A team that produces 10 shots on target from long range, for example, has a lower chance of scoring than a club that produces the same number of shots from inside the box. Metrics derived from this simple count of shots, however, will equally evaluate the two teams. This data set was obtained from Kaggle and as a consequence of web scraping and integration of various data sources. There are 11 different sorts of events, as well as the primary and secondary players involved in those events, as well as numerous more statistics. The data set provides a granular view of 9,074 games, totaling 941,009 events from the biggest 5 European football (soccer) leagues: England, Spain, Germany, Italy, France from the 2011/2012 season to 2016/2017 season as of 25.01.2017. A football game generates much more events and it is very important and interesting to take into account the context in which those events were generated.

```
Confusion Matrix:
[[70773   921]
 [ 6228  2276]]
Report:
              precision    recall  f1-score   support

           0       0.92      0.99      0.95     71694
           1       0.71      0.27      0.39      8504

    accuracy                           0.91     80198
   macro avg       0.82      0.63      0.67     80198
weighted avg       0.90      0.91      0.89     80198
```

Fig. 6. Confusion matrix for XGBoost on expected goals data-set

Now that we have our data set, we can apply the xG model. We started by using one-hot encoding to all the shot parameters (whether a goal was scored, the type of attack, the area the shot was taken from, etc). This would help us learn the data well. We then split the data using 35% for validation and the rest for training. XGBoost, being an ensemble technique, gave us the best results for market value predictions and was also used here. We got a very good accuracy of 91%. Looking at the confusion matrix to our precision, recall, and F1 score were all above 90% too. Similarly, we applied Logistic Regression and Neural Networks too with the former giving us an accuracy of a much lower 72% but still above 90% scores on the confusion matrix. Neural networks had an accuracy of 91.10%, slightly better than XGBoost. We saw there was a massive correlation between xG and goals which suggests our model was working.

Having analyzed how good each technique was, we took our xG model to the player level to analyze them better. We analyzed the best finishers, players with high xG, players with high xG - goals, the best headers, left-footed players, right-footed players, and outside-of-the-box shooters. We also took our model to the passing side of the game and analyzed players with the best passes, crosses, and the most unlucky players.

	goals_scored	expected_goals
goals_scored	1.000000	0.976885
expected_goals	0.976885	1.000000

Fig. 7. Correlation between expected goals and actual goals scored

V. Application Development

As previously mentioned, this study aims to create an application that uses ML and AI to predict team and player statistics as well as attributes. In the previous section, the machine learning models for the same were built. In this section, we have developed the application and integrated the models in it along with various other functionalities. It is a completely responsive web application built using ReactJS and Flask.

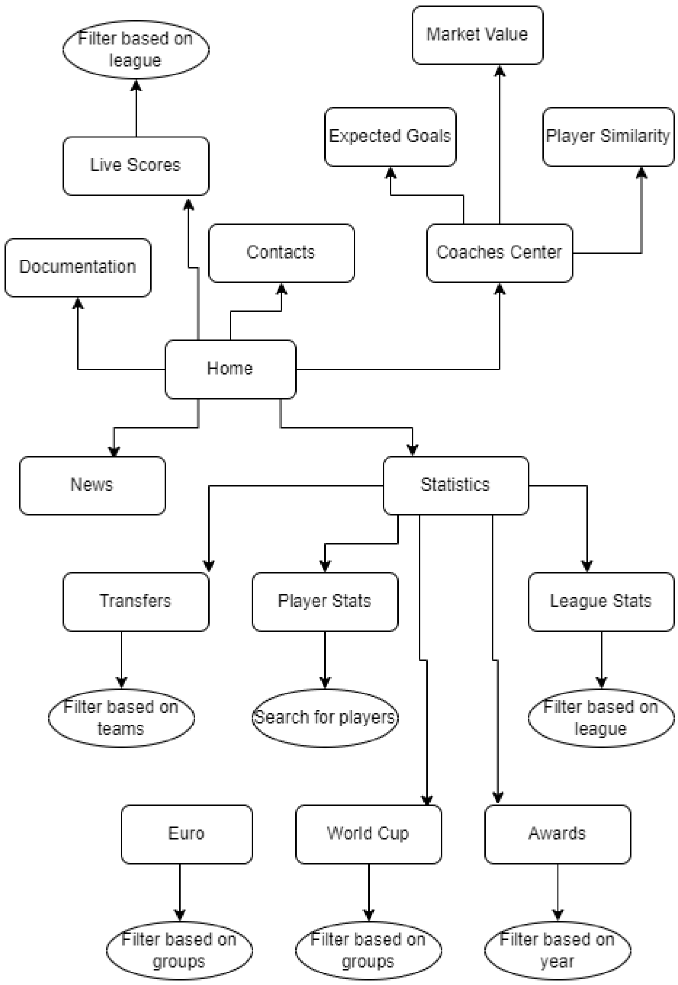

Fig. 8. System Diagram

A. *Planning and Design*

During the planning and designing stage of the project, the requirements gathering was completed first. This was done by discussing internally among our team as well as with other friends and coworkers who actively followed football. We then created system diagrams and module diagrams for the front-end and ER-diagrams for the proposed database on the back-end. We also created wireframes for the entire application front-end. During the creation of these wireframes, we followed all the golden rules of UI/UX design.

Fig. 9. Home page mockup

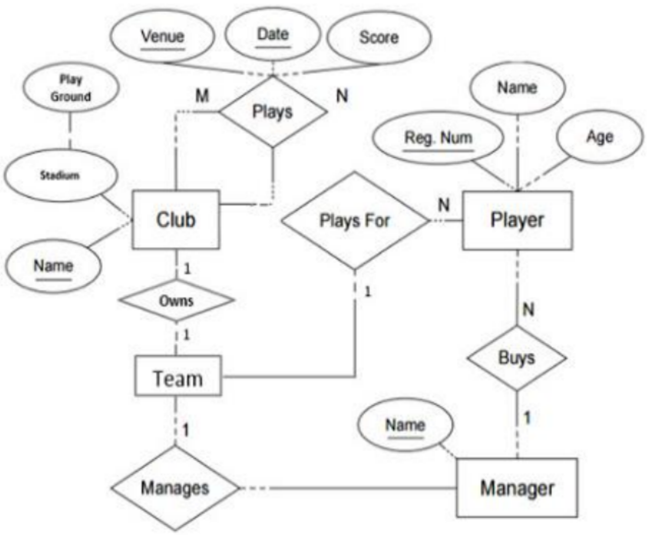

Fig. 10. ER Diagram

B. *Development and Implementation*

1) Front-End: The front-end was created using react. This interactive react app is displayed to the end-user. The app includes 5 modules including home, live scores, news, stats, and coaches center. Live scores from matches across the world are displayed to the user in the live scores modules. These scores can be filtered by leagues. The latest football news is displayed under the news module and various live league and player stats are shown in the stats module. These modules use the data fetched/received from the server through API calls. The coaches center module shows the results from the prediction models developed on the server using machine learning.

The home module contains three sections - the news section, the scores section, and the transfers section. The news section provides the latest football news from the news page, the scores section provides the latest football scores from the live scores page and the transfers section provides the latest football transfers from the statistics page. All this information is presented in the form of cards - which is the basic component of this application. The news module shows the latest news from the world of football including results, pundit opinions, discussions, transfer rumours, etc. The live scores module shows the latest fixtures and live scores from all the football matches around the world. These matches or cards can be filtered by league. Almost all leagues recognized by FIFA are available in this module. The cards used in this module are bigger and contain more information as compared to the ones used in the home module.

The stats module contains 5 different smaller modules - player stats, league stats, awards, transfers, and international competitions. In the player stats module, all the statistics on every player registered with FIFA are available. These stats include name, age, dob, preferred foot, nationality details, club details, positions, ratings, traits, attributes, etc. Users can search for any player of their choice using the search functionality available in this module. More than 17,000 players are available in this module. The league stats module shows the latest and live tables from the top leagues of the world. Apart from this it also shows the top scorers and top assists from those leagues. The awards module shows all the Ballon d'Or award winners and their details and statistics. The transfers module shows all the transfers from the current season or the latest transfer window. Filter functionality is available in all the above-mentioned modules. The international competitions module shows all the statistics, details, and journeys of the two most followed international tournaments - the latest FIFA World Cup and the latest UEFA Euro.

The coaches center module shows the results and stats from the 3 machine learning models developed in the previous section. These include predicting any player's market value, predicting the similarity between 2 players, and analyzing/predicting the expected goals of a match. The models themselves as well as their working is also shown in this module.

2) Back-End: The server for the app is created using flask. Data scraping is performed on the server, after which we use our own APIs to create/fetch JSON data sets. Using these data sets we have created the machine learning models to predict player similarity, market value, expected goals, etc using python and its various packages. We have deployed these models on the server. Every time the application loads up or

Fig. 11. Home module

any page within it is refreshed, the server is also refreshed as it is connected to the front-end. This provides the live functionality to the news, scores, league stats, and various other features within the application. This is because every time the server is refreshed, the API will be called, web scraping will be performed and the JSON files containing the latest data will be created or loaded.

3) Database: The database for this software consists of data sets and JSON files. There are a total of 10 such data files within the database out of which some are manually created by web scraping every time the application is loaded while the rest are predefined. These data files include:

1) newsData - to store data about the latest football news.
2) scoresData - to store data about the latest football scores.
3) leagueTables - to store data about the league table of the top leagues around the world.
4) leagueStats - to store the information about the top scorers and top assists of the top leagues of the world.
5) leagueConversion - to store the key-name pairs of all the leagues around the world.

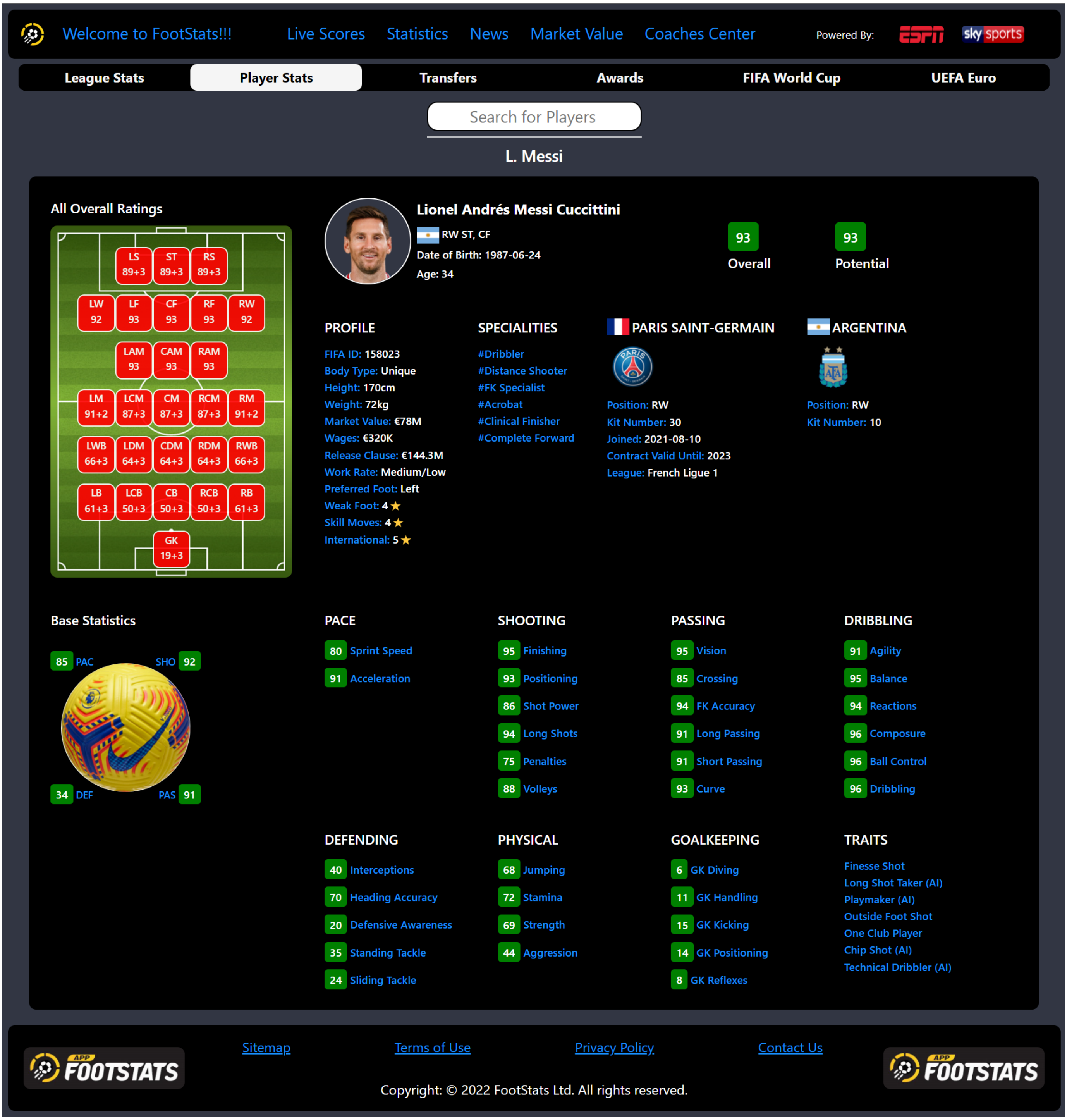

Fig. 12. Statistics module

6) PlayerNames - to store the names of all the players registered with FIFA.
7) PlayerDetails - to store all the statistics about all the players registered with FIFA.
8) awardsData - to store all the data about the Ballon d'Or award from the start of the award to the current year.
9) transfersData - to store the information about all the transfers of the current season or window.
10) contactsData - to store all the contact information about the developers of this application.

C. Advantages

A possible use of this app through its statistics is for users to gauge how they have performed in betting markets, to detect trends, or make betting selections. Although this means the app does not include every possible feature relating to football, this app is ahead of the market in this area, providing a view of fixtures that is not available elsewhere.

Another advantage of it is being able to predict the similarity between players. This will allow the coach to identify an ideal and suitable replacement for any player who leaves the club during the transfer window. The xG prediction and market value prediction done by the app will help the coach pick the best 11 for each match and also prevent the club from paying any extra money for any transfer. The player statistics on the other hand will help the club scout players around the world on an initial level before they go ahead and invest in physical scouting.

D. Algorithms and Technologies

This software contains approximately 4500 lines of code in total, and the code is modular in nature, meaning it is divided into modules or components for re-usability within the application. This reusability of code and components is very important in modern-day software development. ReactJS helps in this by providing the option of components in an application. When measured using lizard.ws, the average complexity score of all modules was 6. A simple search algorithm based on the Greedy Approach was used to retrieve data from the database on the server.

The languages used for the application include HTML, CSS, and JavaScript - front-end; Python and JSON for the back-end. The frameworks used are ReactJS, Node.js, and Flask. The libraries and packages used in the application include jupyter, sklearn, BeautifulSoup, Selenium and AntDesign.

E. Testing and Deployment

White box testing was carried out following the completion of the software implementation. A large number of test cases were created for the same purpose. Approximately 80% of the test cases were successfully passed after white box testing. The remaining tests all yielded unfavorable outcomes in some way. The software code was changed several times until the bugs were fixed and the test cases were successfully cleared in subsequent white box testing rounds. Before deployment/release, the application was also subjected to black-box testing by distributing it to a few friends and coworkers and asking them to try it. All of the bugs they reported were fixed in subsequent code refactoring rounds. Few of the bugs were skipped as they didn't impact the application in any significant way for the release. The entire testing and bug report can be found in the documentation of the software.

The application is currently not deployed on any platform on the internet. But it is hosted on GitHub with its own public repository (URL - https://github.com/alokpurohit18/footstats) that contains all the source code, documentation, packages, and releases. To check the application out, users can download it from this repository and install it on their own systems. The instructions for the installation are mentioned in the readme.md file of the documentation. The application and the GitHub repository is free and open source but under an MIT license.

VI. Conclusion

This project offers something rarely seen in related applications on the market currently. Applications and websites like "Transfermarkt", "OneFootball" and more do provide comprehensive details about the world of football, predictive analytics is notably missing from these. This makes it an application encompassing several important areas.

The ability to look up any and all results on almost all top leagues quickly thanks to the speed of ReactJS in deployment mode makes this a convenient application to use. This combined with an in-depth analysis of players' abilities completes the statistics part of the application. What makes it really unique, however, is our additions of predicting the market value of players, checking player similarity, and analysis of expected goals which provide a different dimension to an already robust application.

VII. Future Scope

There is significant potential in this application, particularly in the domain of Deep Learning. The metrics for expected goals are evolving constantly. Developing our own xG algorithm using deep learning is an area that could be explored. Using stills from live matches for each shot, we can develop a novel method to determine expected goals based on our preferred metrics. For example, some xG models consider the position of opposition players on the pitch while some consider only the shot angle and distance from the goal. The contentious nature of this concept makes it an interesting challenge to tackle in the future. All these can be provided in future updates and releases of this software through GitHub.

References

[1] P. Cintia, F. Giannotti, L. Pappalardo, D. Pedreschi, and M. Malvaldi, "The harsh rule of the goals: Data-driven performance indicators for football teams," in *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 1–10, IEEE, 2015.

[2] K. Tuyls, S. Omidshafiei, P. Muller, Z. Wang, J. Connor, D. Hennes, I. Graham, W. Spearman, T. Waskett, D. Steel, *et al.*, "Game plan: What ai can do for football, and what football can do for ai," *Journal of Artificial Intelligence Research*, vol. 71, pp. 41–88, 2021.

[3] U. Lichtenthaler, "Mixing data analytics with intuition: Liverpool football club scores with integrated intelligence," *Journal of Business Strategy*, 2020.

[4] M. Mondello and C. Kamke, "The introduction and application of sports analytics in professional sport organizations a case study of the tampa bay lightning," vol. 6, pp. 1–15, SAGAMORE PUBL LLC 1807 N FEDERAL DR, URBANA, IL 61801 USA, 2014.

[5] D. Memmert, "Data analytics in football: positional data collection, modeling, and analysis," *Journal of Sport Management*, vol. 33, no. 574, pp. 2019–0308, 2019.

[6] A. Pearson, "How real-time analytics changes the face of the sports betting industry," vol. 5, pp. 61–75, Henry Stewart Publications, 2017.

[7] V. De Silva, M. Caine, J. Skinner, S. Dogan, A. Kondoz, T. Peter, E. Axtell, M. Birnie, and B. Smith, "Player tracking data analytics as a tool for physical performance management in football: A case study from chelsea football club academy," *Sports*, vol. 6, no. 4, p. 130, 2018.
[8] I. Behravan and S. M. Razavi, "A novel machine learning method for estimating football players' value in the transfer market," vol. 25, pp. 2499–2511, Springer, 2021.
[9] R. Pariath, S. Shah, A. Surve, and J. Mittal, "Player performance prediction in football game," in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 1148–1153, IEEE, 2018.
[10] A. Rathke, "An examination of expected goals and shot efficiency in soccer," vol. 12, pp. 514–529, Universidad de Alicante, 2017.
[11] R. Rein and D. Memmert, "Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science," vol. 5, pp. 1–13, SpringerOpen, 2016.
[12] P. Madrero Pardo, "Creating a model for expected goals in football using qualitative player information," 2020.
[13] J. Stübinger, B. Mangold, and J. Knoll, "Machine learning in football betting: Prediction of match results based on player characteristics," vol. 10, p. 46, MDPI, 2019.
[14] K. Apostolou and C. Tjortjis, "Sports analytics algorithms for performance prediction," in *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pp. 1–4, IEEE, 2019.
[15] D. Patnaik, H. Praharaj, K. Prakash, and K. Samdani, "A study of prediction models for football player valuations by quantifying statistical and economic attributes for the global transfer market," in *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, pp. 1–7, IEEE, 2019.
[16] R. Stanojevic and L. Gyarmati, "Towards data-driven football player assessment," in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pp. 167–172, IEEE, 2016.
[17] J. Stübinger and J. Knoll, "Beat the bookmaker–winning football bets with machine learning (best application paper)," in *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pp. 219–233, Springer, 2018.
[18] C. P. Igiri, "Support vector machine–based prediction system for a football match result," vol. 17, pp. 21–26, 2015.
[19] P. Rajesh, M. Alam, M. Tahernezhadi, *et al.*, "A data science approach to football team player selection," in *2020 IEEE International Conference on Electro Information Technology (EIT)*, pp. 175–183, IEEE, 2020.
[20] A. Partida, A. Martinez, C. Durrer, O. Gutierrez, and F. Posta, "Modeling of football match outcomes with expected goals statistic," vol. 10, 2021.
[21] M. He, R. Cachucho, and A. J. Knobbe, "Football player's performance and market value.," in *Mlsa@ pkdd/ecml*, pp. 87–95, 2015.