Implementation of Data Analysis and Visualization using exploratory data in various aspects of Football

Sundar K[*1], Viswak R B[#2], Diwakaran N[#3], Jaya Surya M[#4]

[*]*Associate Professor, Department of Computer Science Engineering,*
Velammal Engineering College, Ambattur Red Hills Road,
Chennai, Tamil Nadu 600066.
[1]Sundar@velammal.edu.in

[#]*UG Scholars, Department of Computer Science Engineering,*
Velammal Engineering College, Ambattur Red Hills Road,
Chennai, Tamil Nadu 600066.
[2]jayasuryamaha16@gmail.com
[3]diwakarnehrudasan@gmail.com
[4]viswakbalaji2000@gmail.com

Abstract*—. Over the last decade, the emergence of Data Analytics has been growing in every aspect of our lives. And even more recently, it's proven to be an important source in field of sports, especially in Football. But the issue of being able to interpret with data still remains. If the football club is presented with huge streams of numbers and data, but lack the ability to interpret them and derive insights and useful information, then the Data becomes almost pointless. In order to make better decisions, football teams need data, but they also need analytics to make sense of it and derive insights and a visualization module to make use of those available data. To address this issue, this paper presents the brief implementation of Data in various aspects of football and it delves into a deeper analysis of how it could be benefited from data driven decision in various aspects. This paper is classified into 3 major aspects – Financial aspect which includes player recruitment, scouting etc. ,Individual/Player aspect – involves analyzing an individual player, player performance etc., Tactical aspect – involves tactical analysis of the team using data. And also implement visualization modules to present these data. And the technique of web scraping is used to derive data from the opensource. Therefore, this paper gives a brief implementation on Data in the above mentioned aspects and this would significantly improve the game in all those aspects, in terms of financially and improving the performance on the pitch

Keywords—Data Analysis, data visualization, Performance Analysis, web scraping, Tactical analysis.

I. INTRODUCTION

Data Analytics is the method of cleansing and analyzing the raw data to make conclusions about the information, and this conclusion could be presented via data visualization for making data driven decisions.

Especially in football the data could be used in many different ways to come up with conclusion. This usage of data would significantly help the club financially and also performance wise. This combination of Data and Football led to the rise of Liverpool in 2018 and even the Southampton football club emphasizes the usage of Data to make their signings. The business model of the clubs differ from each other, for example Southampton football club's business model is highly depends on the financial aspect. To put it into simpler words they sign a player for a smaller fee with high potential and they sell the player at some point for a significant profit. The business model may differ from club to club, but the underlying principle always remains the same for all the clubs.

This implementation on Financial aspect's main motive is the finance and profit for the club. By the use of Data we can make much better signings at a significantly lower price and this would in turn make the club a lot of profits financially. And the Player aspect is on the individual or the analysis of player performance using data. We can analyze the data and come up to the conclusion regarding a player, his on the pitch performances, his goal contributions, past records and his training performance and by this way it helps to analyze the player individually. Final one is the tactical aspect, this is highly important for football managers and head team coaches to understand their team better and also to make changes to the side. By using data, we could derive tactical conclusions of the game such as the team shape and formation, errors that lead to a goal etc. We use the process of web scraping to scrap the modules from the open source and make a dataset, by this way it's very cost efficient for the smaller clubs.

1.1 WEB SCRAPING:

Data scraping, also known as web scraping, is the process of importing information from a website into a spreadsheet or local file saved on your computer. It's one of the most efficient ways to get data from the web, and in some cases to channel that data to another website.

There are two ways to scrap data, it could be done either manually using Microsoft Excel web query or this process could be automated to Beautiful Soup. Beautiful Soup is a Python module that is able to make sense of the tags inside HTML and XML documents. Some of the steps involved are, we have to inspect the data source and the next step is to decipher the URL. And the next step is to inspect the entire site using developer tools. Now we have to create a virtual environment to scrape the data. A code is imported which

issue a HTTP get request to the site and then it retrieves the HTML data and the server sends it back then we can use the print to .text attribute. It makes the process much easier if the site is a static site.

This paper uses the FIFA 22 Dataset, the player attributes and potential are derived from the latest FIFA and the wages

1.2 PYTHON MODULES:

Python is a key part of Data analysis and visualization. The libraries used here are NumPy, Pandas and Matplotlib. NumPy is a library for Python that adds support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

Pandas is a high-level data manipulation tool that is built on the NumPy package. The key data structure in Pandas is called the DataFrame. DataFrames are incredibly powerful as they allow you to store and manipulate tabular data in rows of observations and columns. NumPy used for storing multi-dimensional arrays and it can be used to perform high level mathematical functions. It is suggested anaconda, as most of the modules will be available. If we have to install separately, pip command can be used. But it's strongly advised to use Anaconda as it comes with modules such as Jupiter web and Spyder which it very effective when it comes to Data analysis and visualizations

1.3 STREAMLIT:

Streamlit is a tool that is used to build interactive web applications using data. It is highly effective when it comes to creating dashboards. And these dashboards could be set up on a localhost or it could be even hosted on a website.

One of the major advantages of this tool is that it's very quick and highly interactive. Slider functionality in Streamlit module is highly valuable. This module could be highly emphasized in the player performance analysis for creating dashboards for interactive analysis.

II. RELATED WORKS

A. Data analytics in football sport to identify gaps for the improvement of quality opportunities throughout world-wide teams :

Proposed by Syed Ali Fathima[1] S J, Sumathi V P, Sumanth S. Football is a widely known sport. Billions watch and play the game around the world. Data Analytics has assumed a huge role in the world of Football. It has transformed how people approach games team formation, player selection etc. Data analytics has enabled teams from around the world to understand their game better and perform better. Data analytics is also used to predict the outcomes of games enabling people to make educated guesses while betting. There is no doubt that Football is worldwide sport. However, there are so many teams worldwide who haven't improved when compared to some of the others. Few teams don't even manage to make into the main tournaments like FIFA. Some countries lack funding and some teams don't have the exposure to standard equipment, coaching opportunities etc. It is very important for a Football enthusiast to know that the game keeps evolving towards a point where there are more quality teams around the world. It is very important for data analytics to move into this direction of finding answers to the question "What can be done to provide quality opportunities to the teams worldwide?". The present paper discusses exactly that and looks to provide an answer to that very question.

B. Process Mining of Football Event Data: A Novel Approach for Tactical Insights Into the Game

Proposed by Pavlina Kröckel and Freimut Bodendorf.[2] The paper explores process mining and its usefulness for analyzing football event data. We work with professional event data provided by OPTA Sports from the European Championship in 2016. We analyze one game of a favorite team (England) against an underdog team (Iceland). The success of the underdog teams in the Euro 2016 was remarkable, and it is what made the event special. For this reason, it is interesting to compare the performance of a favorite and an underdog team by applying process mining. The goal is to show the options that these types of algorithms and visual analytics offer for the interpretation of event data in football and discuss how the gained insights can support decision makers not only in pre- and post-match analysis but also during live games as well. We show process mining techniques which can be used to gain team or individual player insights by considering the types of actions, the sequence of actions, and the order of player involvement in each sequence. Finally, we also demonstrate the detection of typical or unusual behavior by trace and sequence clustering.

C. The Role of Analytics in Assessing Playing Talent using Data Analytics in Football:

By . Dr Bill Gerrard.[3] This chapter considers the role that data analysis should play in decisions requiring an assessment of players whether it be young players in a youth development programme or established first-team regulars. Moneyball has highlighted the possibilities for analytics as a competitive strategy particularly for small-markets teams with relatively limited resources. This chapter will go beyond Moneyball to consider the problems of constructing player rating systems in the invasion-territorial team sports in which player performance is multidimensional. Drawing on decision theory and cognitive psychology, it is argued that the role of statistical analysis is secondary to the expert identification of the characteristics of optimal player performance. It is concluded that effective analytics in sport must always be coach-led.

D. Analysis of Decision Making Process in Moneyball: The Art of Winning an Unfair Game:

Billy Beanes's[4] success in using data-driven decision making in baseball industry is wonderfully written by Michael Lewis in Moneyball. As a general manager in baseball team that were in the bottom position of the league from the financial side to acquire the players, Beane, along with his partner, explored the use of data in choosing the

team's player. They figured out how to determine the worth of every player.The process was not smooth, due to the condition of baseball industry that was not common with using advanced statistic in acquiring players. Many teams still use the old paradigm that rely on experts' judgments, intuition, or experience in decision making process. Moneyball approached that using data-driven decision making gave excellent result for Beane's team. The team won 20 gamessequently in the 2002 season and also spent the lowest cost per win than other teams.This paper attempts to review the principles of Moneyball – The Art of Winning an Unfair Game as a process of decision making and gives what we can learn from the story in order to win the games, the unfair games.

E. The Application of Data Mining Algorithm Based on Association Rules in the Analysis of Football Tactics

Proposed by Wang Puchun Yuxi[5] Normal University, Yuxi, Yunnan, China. In this paper, an improved association rule data mining algorithm is proposed, and by applying this algorithm to the analysis of football tactics, the training level of football players is improved. Firstly, the problem of mining association rules is explained, and the concept of support and confidence is defined in advance. Secondly, for a given transaction database and a target object, top K direct association rules can be extracted using our proposed algorithm. Finally, to verify the effectiveness, we choose the final match in European Cup 2008 (that is Spain vs Germany) as the testing data, and the toolbox Adobe Premiere Pro is exploited to obtain soccer ball possession and pass time, and the football tactical information from the data of two samples is mined based on association rule mining algorithm, and experimental results demonstrate that the proposed association rule data mining algorithm can effectively extract football tactics from the testing data.

F. Implementation of a recruit visualization tool for UVA football:

Proposed by Leah Walter, Andrew Citeria, Knowles[6], University of Virginia from the aforementioned reference papers. Effective recruiting is essential for building a successful football program, as it is the primary method by which college football teams acquire new players. The University of Virginia football recruiting staff works[9] hard to direct its efforts toward recruits that fit Virginia's football culture and can lead to future wins. This paper describes four models and visuals that help the staff determine which recruits to pursue. Recruiters and coaches evaluate players based on the following criteria: 1) football talent, 2) academic performance, and 3) mental toughness, referred to as "grit". By analyzing the staff's processes and metrics, we found areas where data analytics could be effectively applied. A model that forecasts a recruit's college GPA was developed to help predict a recruit's future academic success. To inform grit evaluations, we developed a visual that uses recruits' tweets to show their confidence and toughness[12]. Additionally, to ensure recruiting efforts are directed at players that have a chance of committing to Virginia, we designed a prediction model that outputs the likelihood of a recruit selecting the University of Virginia from the set of schools that have given the recruit an offer to join the team. Furthermore, a supplementary visual known as the competitive landscape plots Virginia against all other schools that have given the recruit an offer, helping the coaches see how they differ from competing schools. Ultimately, these models and visuals were combined into a comprehensive visual, called the Recruit Dashboard, to give the staff a simple way to analyze all metrics needed to evaluate a recruit.

PROPOSED MODEL

As of now, there isn't any recognized Data Analytics and a Visualization model or tool in Football. This is due to various factors such as certain Football clubs have their own Business model and they use their analytics to work depending upon their model. But still there are certain common patterns to all of these these football clubs and their models and there isn't any recognized tool for it. This paper proposes a set of Data Analytics and Visualization tool with the common models of all these football clubs. We created a set of tool with each having it's own purpose, it could be applied in three key aspects, Performance Analysis – Player Recruitment Strategy – Tactical Analysis We used python and data analysis and visualization modules (numpy, pandas, matplotlib) to create these tools which could be applied in the above mentioned areas, which would prove to be beneficial to the clubs. Using the exploratory data that is available we made tool for Data analysis and Visualization and these could be further developed into a business model connected to a live database.

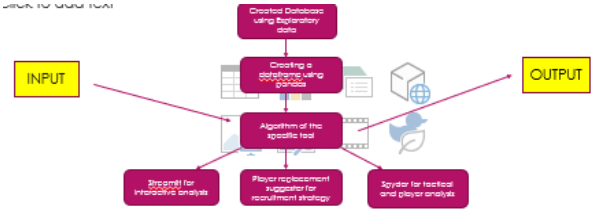

Figure 1 Proposed System Architecture

3.1 IMPLEMENTATION OF DATA IN FINANCIAL ASPECT:

Finance plays a key role in the game of football when it comes to player recruitment, building the stature of the club. Data could be implemented in this aspect which would significantly save the club a lot of money. It's highly important when it comes to the player recruitment part.

This paper proposes a player replacement suggester, where this takes in consideration of the player wages, player's potential, attributes such as defensive and offensive. Of all these attributes taken into consideration, this gives the cheaper and a better replacement suggestion[10] . When the user enter a player they need to replace, this suggester will give them the better options with higher potential for a cheaper price. By this model, player recruitment could be significantly improved by a huge margin. And this model helps the clubs to have better finances when it comes to recruiting big valued players. This method could be implemented on any club despite their business model, as we mentioned before the underlying principles are all the same in this model.

```
Enter player name:  Fred
[6]:
```

	short_name	wage_eur	value_eur	player_positions	overall	age
91	W. Ndidi	120000.0	66500000.0	CDM, CM	85	24
136	F. Kessié	51000.0	49500000.0	CDM, CM	84	24
176	P. Højbjerg	105000.0	43500000.0	CDM, CM	83	25
253	M. Locatelli	74000.0	42000000.0	CDM, CM	82	23
260	Palhinha	18000.0	41000000.0	CDM, CM	82	25
273	D. Rice	69000.0	43000000.0	CDM, CM	82	22
278	T. Souček	81000.0	36000000.0	CDM, CM	82	26

Figure 2 Player Replacement Suggester

3.2 IMPLEMENTATION OF DATA IN PLAYER PERFORMANCE:

Player performance is a key role in the game. Player performance is used in analysis not just for the head coaches, but also for the common audience, when it comes to things such as FPL. This model could be also used to build predictive model for FPL Analysis.

This model is a fantasy football analysis model[7] , which is a player performance analysis model that's built on the streamlit module. This model provides a interactive dashboard for the user. It also comes with the visualizer. The slider feature could be very helpful for the player worth. It provides a scatter chart for the user for better visualization, which involves goal involvements such as goals and assists. It can also show single or multiple results as it has the filter feature based on clubs and player positions. This model could be implemented of player performance analysis and the visualization module[15] helps the user for better understanding of the conclusion drawn from the data.

Figure 3 Implementation of Data in Player performance

3.3 IMPLEMENTATION OF DATA IN TACTICAL ASPECT:

Tactics are the essential game plan for the coaches. It is considered as the key when it comes to winning matches. Tactics differs from coaches to coaches , which includes popular formations[11] such as 4-3-3, 4-4-2 or 5-3-2, also it depends on the playing style such as Gegenpressing, Tiki-taka, or park the bus style. By the use of data it is possible to analyze the game tactically, such as formation throughout the game, team shape[14] . And this makes it easier to visualize how and where the player positions are when the goal is conceded or the goal is scored.

This model consists of plotting the pitch using matplotlib module[8], and it helps to mark the player positions in that pitch. This helps in better visualization for both the coaches and the players. It could be advanced even further into tracking player movements or a certain player or the opposition runs. This model will be very valuable for the head coaches in terms of pre or post game tactical analysis.

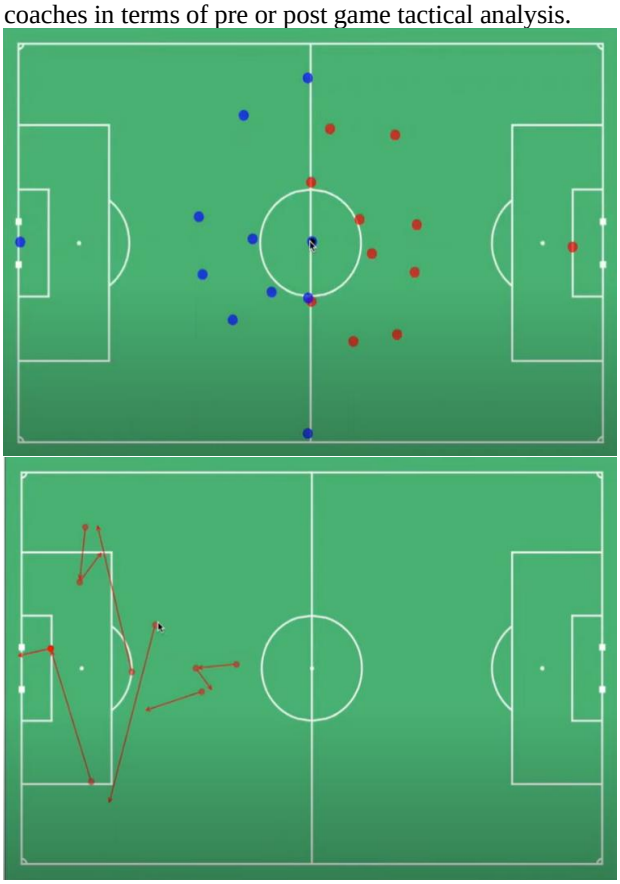

Figure 4 Implementation of Data in tactical aspect

ADVANTANGES IN PROPOSED SYSTEM:

a) Financial module is helpful for the club financially in terms of player recruitment strategy and profit from transfer fee and replacing the key players.

b) Player Analysis module helps the coaches to analyze the player performance on the pitch and even in the training to analyze and evaluate the individual.

c) Data implementation in tactics is valuable for coaches in terms of studying their own team or the creating opposition reports like most used formations and team shapes , even heat maps of players could be derived from data.

III. CONCLUSION

To conclude, the idea is to enhance the game of Football using Data, that's the main aim of this research. This paper gave an in depth insight of how data could be used in different aspects of football, which in turn would be beneficial for all the people. For example the player replacement finder would help the club to save a lot of money, in terms of wages and profit in terms of selling value. The player Analysis and tactical model would help the coach or the manager to produce better results on the pitch. This idea could be enhanced further by connecting to a bigger network or even creating a live database with a pool of Analysts.

ACKNOWLEDGEMENT

We wish to acknowledge with thanks to the significant contribution given by the management of our college Chairman, Dr. M. V. Muthuramalingam, and our Chief Executive Officer Thiru. M. V. M. Velmurugan, for their extensive support.

We would like to thank Dr. S. Sathish Kumar, Principal of Velammal Engineering College, for giving me this opportunity to do this project.

We wish to express my gratitude to our effective Head of the Department, Dr. B. Murugeshwari, for her moral support and for her valuable innovative suggestions, constructive interaction, constant encouragement and unending help that have enabled me to complete the project.

We wish to express my indebted humble thanks to our Project Coordinators, Dr. P. Pritto Paul, Dr. S. Rajalakshmi and Dr. S. Gunasundari, Department of Computer Science and Engineering for their invaluable guidance in the shaping of this project.

We wish to express my sincere gratitude to my Internal Guide, Dr. P. Pritto Paul, Associate Professor, Department of Computer Science and Engineering for her guidance, without her this project would not have been possible.

We are grateful to the entire staff members of the Department of Computer Science and Engineering for providing the necessary facilities and carrying out the project. I would especially like to thank my parents for providing me with the unique opportunity to work and for their encouragement and support at all levels.

Finally, my heartfelt thanks to The Almighty for guiding me throughout the life.

IV. REFERENCES

[1] Data analytics in football sport to identify gaps for the improvement of quality opportunities throughout world-wide teams

[2] Process Mining of Football Event Data: A Novel Approach for Tactical Insights Into the Game

[3] The Role of Analytics in Assessing Playing Talent using Data Analytics in Football:

[4] Analysis of Decision Making Process in Moneyball: The Art of Winning an Unfair Game:

[5] The Application of Data Mining Algorithm Based on Association Rules in the Analysis of Football Tactics

[6] Implementation of a recruit visualization tool for UVA football:

[7] Machine Learning for Position Detection in Football, IEEE [2020]

[8] A Novel Decision Support Model to Discover the Interesting Pattern in Football Match [2019]

[9] Research of association rules in analyzing technique of football match

[10] Constructing Spaces and Times for Tactical Analysis in Football

[11] Measuring efficiency of Thailand's football premier leagues using data envelopment analysis [2021]

[12] Clustering and Evolutionary System Analysis of Data Mining Algorithms in the Field of Football

[13] Sports analytics & risk monitoring based on hana platform: Sports related big data & IoT trends by using HANA In-memory platform [2019]

[14] [The harsh rule of the goals: Data-driven performance indicators for football teams2020]

[15] The Growing Importance of Football AnalyticsTutorial [https://soccerment.com/the-importance-of-football-analytics/]