Premier League Dashboard

Kiran Seetharam, Kaushik Nandan
Prayag Ganesh Prabhu, Tejas Pratap, Pratik Sabne

March 2024

1 Introduction

1.1 Problem Statement

We aim to develop a user-friendly Premier League dashboard that provides real-time match predictions, updating after each match and resetting at the end of the season. This dashboard will feature a dynamic table predictor, ensuring users stay informed with the latest match outcomes and predictions. Additionally, users will have access to various interactive graphs and charts representing key attributes utilized in our prediction model. Our goal is to offer an intuitive platform for football enthusiasts to analyze and engage with Premier League data effectively.

1.2 Inspiration

Due to the growing interest in data-driven decision-making across various industries, including sports, we have decided to propose this idea. Predicting the outcome of sports events has always captivated enthusiasts, and the Premier League, being one of the most popular and competitive football leagues worldwide, offers an exciting challenge for data scientists and has a potential for lots of learning.

1.3 Background

This project builds upon existing work and background research in sports analytics and predictive modeling. Several studies have explored predictive models for sports outcomes, including football. Researchers have investigated various factors such as team statistics, player performance, playing style, and historical results to develop predictive models. However, given the dynamic nature of the Premier League and the interplay of multiple variables, there is a constant need for improved models that capture the complexity of the competition.

2 Method

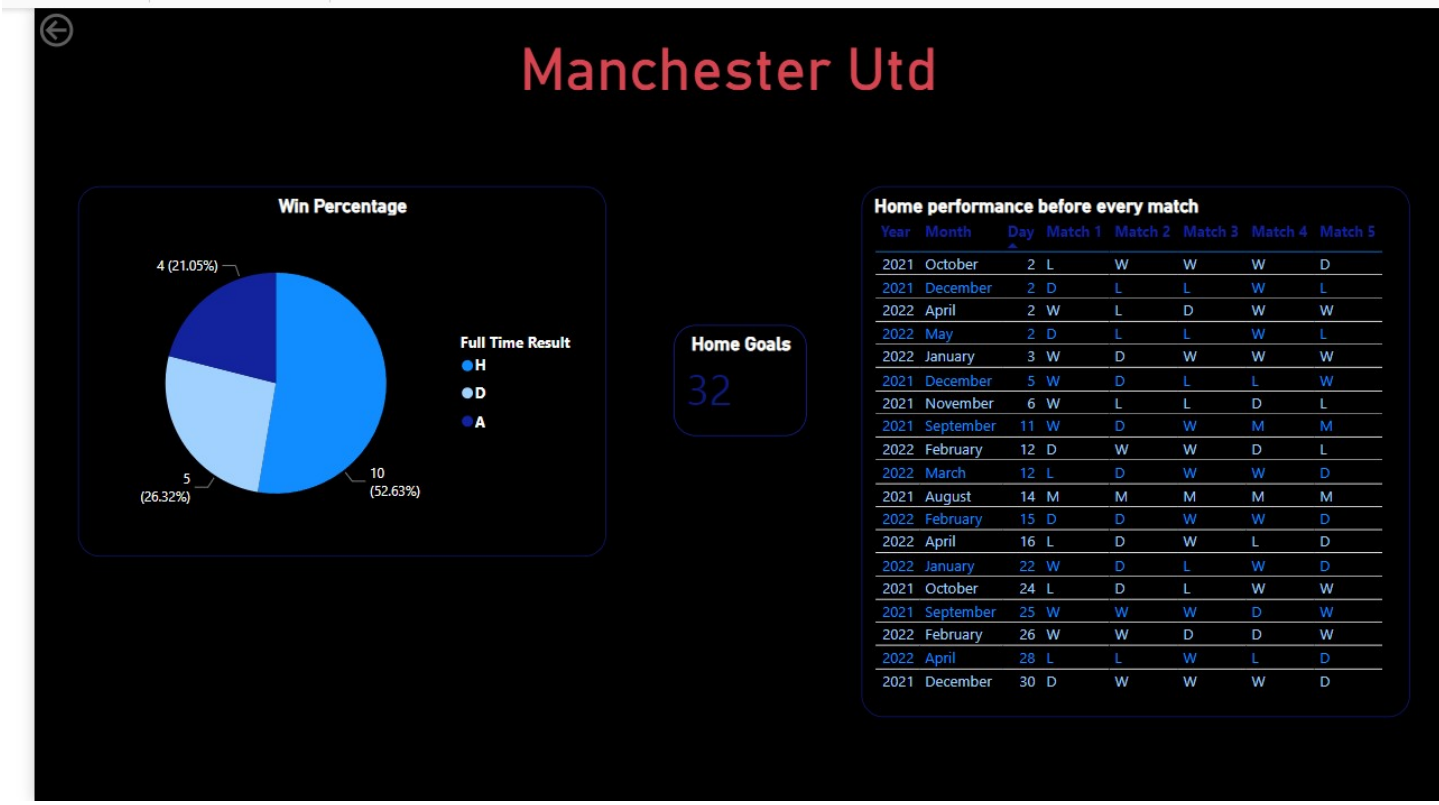
Our methodological approach is delineated below, encompassing a systematic series of steps:

- **Data Acquisition**: To ensure the integrity and reliability of our dataset, we meticulously gathered comprehensive Premier League standings, statistics, and match results spanning from the 2000-2001 season to the present campaign. This process entailed leveraging web scraping techniques on the *fbref.com* platform [1], utilizing the powerful Python libraries *Beautiful Soup*, *NumPy*, and *Pandas*.
- **Dataset Composition**: The resultant dataset encapsulates matchwise data from the 2000-01 season onwards, comprising a diverse array of meticulously collected attributes and calculated metrics. Key components of this dataset include team names, manager names, full-time results, match-specific details such as dates, full-time and half-time goals, as well as auxiliary metrics such as teams' head-to-head records, managers' head-to-head records, past 5 match form, and expected goals (xG).
- **Model Development and Prediction**: Leveraging the acquired xG data, we employed the robust *xG Boost Random Forest Regressor Model* [2] to facilitate predictive analysis. This model assimilates historical xG data and employs sophisticated algorithms to generate accurate match outcome predictions, thereby enhancing the dashboard's prognostic capabilities.
- **Dashboard Creation**: In a bid to furnish users with an immersive and intuitive platform for data exploration, all scraped and computed data, in conjunction with the prediction model, were seamlessly integrated into *Power BI*. Leveraging the platform's extensive visualization capabilities, we crafted an

interactive dashboard that offers users a rich and dynamic visual experience. The incorporation of drill-through features enables seamless navigation between different pages, facilitating access to nuanced and meaningful insights within the dataset.

3 Result

Our prediction model yields results with an accuracy rate of 54.5%, a commendable achievement given the dual task of predicting match outcomes and the number of goals scored by teams. This multifaceted approach underscores the complexity of our model and its ability to provide nuanced insights into Premier League matches. The accompanying figures offer a glimpse into the comprehensive nature of our project.
The first figure showcases a page from our Premier League Dashboard, providing users with a visually appealing and intuitive interface for accessing match predictions and insights. Subsequent figures provide glimpses into segments of our extensive dataset, demonstrating the breadth and depth of information incorporated into our analysis. Lastly, the xG dataset, displayed in the final figure, serves as a foundational component for predicting match results and offers valuable insights into the expected goals scored by teams, enriching the predictive capabilities of our model.



Year	Month	Day	Match 1	Match 2	Match 3	Match 4	Match 5
2021	October	2	L	W	W	W	D
2021	December	2	D	L	L	W	L
2022	April	2	W	L	D	W	W
2022	May	2	D	L	L	W	L
2022	January	3	W	D	W	W	W
2021	December	5	W	D	L	L	W
2021	November	6	W	L	L	D	L
2021	September	11	W	D	W	M	M
2022	February	12	D	W	W	D	L
2022	March	12	L	D	W	W	D
2021	August	14	M	M	M	M	M
2022	February	15	D	D	W	W	D
2022	April	16	L	D	W	L	D
2022	January	22	W	D	L	W	D
2021	October	24	L	D	L	W	W
2021	September	25	W	W	W	D	W
2022	February	26	W	W	D	D	W
2022	April	28	L	L	W	L	D
2021	December	30	D	W	W	W	D

Figure 1: Power Bi Dashboard Page

Date	HomeTeam	AwayTeam	Home Team Manager	Away Team Manager	FTHG	FTAG	FTR	HTH_hom	HTH_away	HTGS
19-08-2000 00:00	Charlton Ath	Manchester City	Alan Curbishley	Joe Royle	4	0	H	0	0	
19-08-2000 00:00	Chelsea	West Ham	Gianluca Vialli	Harry Redknapp	4	2	H	0	0	
19-08-2000 00:00	Coventry City	Middlesbrough	Gordon Strachan	Bryan Robson	1	3	A	0	0	
19-08-2000 00:00	Derby County	Southampton	Jim Smith	Glenn Hoddle	2	2	D	0	0	
19-08-2000 00:00	Leeds United	Everton	David O'Leary	Walter Smith	2	0	H	0	0	
19-08-2000 00:00	Leicester City	Aston Villa	Peter Taylor	John Gregory	0	0	D	0	0	
19-08-2000 00:00	Liverpool	Bradford City	GÃ©rard Houllier	Chris Hutchings	1	0	H	0	0	
19-08-2000 00:00	Sunderland	Arsenal	Peter Reid	ArsÃ¨ne Wenger	1	0	H	0	0	
19-08-2000 00:00	Tottenham	Ipswich Town	George Graham	George Burley	3	1	H	0	0	
20-08-2000 00:00	Manchester Utd	Newcastle Utd	Alex Ferguson	Bobby Robson	2	0	H	0	0	
21-08-2000 00:00	Arsenal	Liverpool	ArsÃ¨ne Wenger	GÃ©rard Houllier	2	0	H	0	0	
22-08-2000 00:00	Bradford City	Chelsea	Chris Hutchings	Gianluca Vialli	2	0	H	0	0	
22-08-2000 00:00	Ipswich Town	Manchester Utd	George Burley	Alex Ferguson	1	1	D	0	0	
22-08-2000 00:00	Middlesbrough	Tottenham	Bryan Robson	George Graham	1	1	D	0	0	
23-08-2000 00:00	Everton	Charlton Ath	Walter Smith	Alan Curbishley	3	0	H	0	0	
23-08-2000 00:00	Manchester City	Sunderland	Joe Royle	Peter Reid	4	2	H	0	0	
23-08-2000 00:00	Newcastle Utd	Derby County	Bobby Robson	Jim Smith	3	2	H	0	0	
23-08-2000 00:00	Southampton	Coventry City	Glenn Hoddle	Gordon Strachan	1	2	A	0	0	
23-08-2000 00:00	West Ham	Leicester City	Harry Redknapp	Peter Taylor	0	1	A	0	0	
26-08-2000 00:00	Arsenal	Charlton Ath	ArsÃ¨ne Wenger	Alan Curbishley	5	3	H	0	0	
26-08-2000 00:00	Bradford City	Leicester City	Chris Hutchings	Peter Taylor	0	0	D	0	0	
26-08-2000 00:00	Everton	Derby County	Walter Smith	Jim Smith	2	2	D	0	0	
26-08-2000 00:00	Ipswich Town	Sunderland	George Burley	Peter Reid	1	0	H	0	0	
26-08-2000 00:00	Manchester City	Coventry City	Joe Royle	Gordon Strachan	1	2	A	0	0	
26-08-2000 00:00	Middlesbrough	Leeds United	Bryan Robson	David O'Leary	1	2	A	0	0	
26-08-2000 00:00	Newcastle Utd	Tottenham	Bobby Robson	George Graham	2	0	H	0	0	
26-08-2000 00:00	Southampton	Liverpool	Glenn Hoddle	GÃ©rard Houllier	3	3	D	0	0	
26-08-2000 00:00	West Ham	Manchester Utd	Harry Redknapp	Alex Ferguson	2	2	D	0	0	
27-08-2000 00:00	Aston Villa	Chelsea	John Gregory	Gianluca Vialli	1	1	D	0	0	
05-09-2000 00:00	Leeds United	Manchester City	David O'Leary	Joe Royle	1	2	A	0	0	

Figure 2: Dataset

ATLossStre	HTGD	ATGD	DiffPts	DiffFormP	HS	AS	HST	AST	HHW	AHW	Prev_hth	HTH_hom	HTH_away	season	HM1_map	H
0	0	0	0	0	17	8	14	4	2	1	0	0	0	2000	0	
0	0	0	0	0	17	12	10	5	1	0	0	0	0	2000	0	
0	0	0	0	0	6	16	3	9	0	1	0	0	0	2000	0	
0	0	0	0	0	6	13	4	6	0	0	0	0	0	2000	0	
0	0	0	0	0	17	12	8	6	0	0	0	0	0	2000	0	
0	0	0	0	0	5	5	4	3	0	0	0	0	0	2000	0	
0	0	0	0	0	16	3	10	2	0	0	0	0	0	2000	0	
0	0	0	0	0	8	14	2	7	0	0	0	0	0	2000	0	
0	0	0	0	0	20	15	6	5	2	1	0	0	0	2000	0	
0	0	0	0	0	19	9	9	6	2	0	0	0	0	2000	0	
0	-0.5	0.5	-1.5	-1.5	17	7	12	4	0	0	0	0	0	2000	-1	
0	-0.5	1	-1.5	-1.5	12	14	3	6	0	0	0	0	0	2000	-1	
0	-1	1	-1.5	-1.5	13	15	8	6	0	0	0	0	0	2000	-1	
0	1	1	0	0	12	11	6	4	0	0	0	0	0	2000	1	
0	-1	2	-1.5	-1.5	13	8	8	4	0	0	0	0	0	2000	-1	
0	-2	0.5	-1.5	-1.5	15	9	10	4	0	0	0	0	0	2000	-1	
0	-1	0	-0.5	-0.5	9	10	4	5	0	0	0	0	0	2000	-1	
0	0	-1	0.5	0.5	12	7	4	5	0	0	0	0	0	2000	1	
0	-1	0	-0.5	-0.5	17	4	12	2	1	0	0	0	0	2000	-1	
0	-0.5	2	-1.5	-1.5	18	7	9	4	0	0	0	0	0	2000	1	
0	0.333333	0.333333	-0.33333	-0.33333	8	13	4	8	0	0	0	0	0	2000	1	
0	0.333333	-0.33333	0.666667	0.666667	12	7	9	4	0	0	0	0	0	2000	1	
0	-0.66667	-0.33333	-0.66667	-0.66667	14	9	5	3	1	0	0	0	0	2000	1	
0	-0.66667	-0.33333	0	0	14	9	5	8	1	0	0	0	0	2000	1	
0	0.666667	1	-0.66667	-0.66667	15	16	8	11	0	0	0	0	0	2000	1	
0	-0.33333	0.666667	-0.33333	-0.33333	15	10	6	2	1	1	0	0	0	2000	1	
0	-0.33333	-0.33333	-0.66667	-0.66667	14	9	7	4	2	0	0	0	0	2000	-1	
0	-1	0.666667	-1.33333	-1.33333	17	8	8	5	0	2	0	0	0	2000	-1	
0	0	0	-0.33333	-0.33333	12	11	9	7	0	0	0	0	0	2000	1	
0	1	-0.66667	1	1	6	8	1	3	1	1	0	0	0	2000	1	

Figure 3: Dataset

Date	Home	xG(Home)	xG(Away)	Away
11-08-2017	Arsenal	2.5	1.5	Leicester City
12-08-2017	Brighton	0.3	1.9	Manchester City
12-08-2017	Southampton	2	0.3	Swansea City
12-08-2017	Everton	0.6	0.4	Stoke City
12-08-2017	Chelsea	1.5	0.6	Burnley
12-08-2017	West Brom	1.3	0.5	Bournemouth
12-08-2017	Crystal Palace	1.1	1.5	Huddersfield
12-08-2017	Watford	2.1	2.6	Liverpool
13-08-2017	Manchester Utd	2.1	0.5	West Ham
13-08-2017	Newcastle Utd	0.8	2.5	Tottenham
19-08-2017	Stoke City	0.7	1.5	Arsenal
19-08-2017	Liverpool	2.5	0.7	Crystal Palace
19-08-2017	Burnley	1.3	0.9	West Brom
19-08-2017	Leicester City	2	0.2	Brighton
19-08-2017	Southampton	2.1	2	West Ham
19-08-2017	Bournemouth	1	2.4	Watford
19-08-2017	Swansea City	0.4	3	Manchester Utd
20-08-2017	Tottenham	0.7	0.7	Chelsea
20-08-2017	Huddersfield	0.3	0.7	Newcastle Utd
21-08-2017	Manchester City	1.1	0.6	Everton
26-08-2017	Manchester Utd	2.8	0.9	Leicester City
26-08-2017	Newcastle Utd	2.3	0.6	West Ham
26-08-2017	Crystal Palace	0.9	0.7	Swansea City
26-08-2017	Watford	0.3	1.1	Brighton
26-08-2017	Huddersfield	1.4	0.7	Southampton
26-08-2017	Bournemouth	0.5	1.4	Manchester City
27-08-2017	Liverpool	3.1	0.6	Arsenal
27-08-2017	Tottenham	2.3	0.9	Burnley
27-08-2017	West Brom	0.8	1.3	Stoke City

Figure 4: xG data used for prediction

4 Conclusion

In conclusion, our project aimed to develop a user-friendly Premier League dashboard that provides real-time match predictions, updating after each match and resetting at the end of the season. This dashboard features a dynamic table predictor and various interactive graphs and charts representing key attributes utilized in our prediction model.

Drawing inspiration from the growing interest in data-driven decision-making, particularly in sports, our project focused on predicting Premier League match outcomes, leveraging existing research in sports analytics and predictive modeling.

Our method involved gathering valid and reliable data from fbref.com using web scraping techniques and Python libraries such as Beautiful Soup, NumPy, and Pandas. The dataset included matchwise data from the 2000-01 season onwards, covering various aspects such as team and manager names, match results, head-to-head statistics, past form, and expected goals (xG).

We utilized the xG data in a Random Forest Regressor Model [2] for match outcome prediction. Additionally, we incorporated all scraped and calculated data, along with the prediction model, into Power BI for creating a visually appealing dashboard. Drill-through features were utilized for easy navigation between different pages, facilitating access to meaningful data.

Our prediction model achieved a decent accuracy rate, considering its ability to predict both match outcomes and the number of goals scored by teams. The Power BI dashboard provides users with a comprehensive tool for analyzing and engaging with Premier League data effectively.

Overall, our project contributes to the advancement of predictive modeling in sports analytics and provides football enthusiasts with an intuitive platform for analyzing Premier League data.

5 Future Work

For future development, the Premier League dashboard and prediction model could undergo significant improvements. Firstly, refining the prediction model is essential, involving the continuous integration of additional features such as player injuries, weather conditions, and team tactics to enhance accuracy and robustness. Secondly, advanced data visualization techniques could be employed to enrich the dashboard's user experience. This could include the incorporation of heat-maps, network graphs for visualizing player connections, and in-

teractive maps displaying match locations and statistics, offering users deeper insights into match dynamics. Thirdly, enhancing user interaction and customization options could be prioritized, enabling users to personalize their dashboard experience by selecting favorite teams, comparing player or team performance, and setting up notifications for match updates and predictions. These improvements would not only elevate the dashboard's functionality but also provide users with a more engaging and tailored platform for analyzing Premier League data.

References

[1] 2023-2024 Premier League Stats, https://fbref.com/en/comps/9/Premier-League-Stats, [Accessed: 21-3-2024].

[2] Random forests(tm) in xgboost, https://xgboost.readthedocs.io/en/stable/tutorials/rf.html, [Accessed: 21-3-2024].