

English Premier League Match Predictor

About the Project:

The aim of this project is to predict which Premier League team is going to win a match, whether it is the home team or the away team. The data used for this project was gotten from Football-data.co.uk, the data downloaded was for the previous season 2020/2021 and the current season 2021/2022.

Nature of the Project:

The project is a classification problem, the project is to predict whether Home team, or Away team is going to win a match or whether the match is going to end in a draw. The project is a multiclass classification project, with possible outcome being Home team to win, Away team to win, or the match is going to end in a draw.

Work Done:

Data cleaning was done on the data, fortunately there was no missing values or columns to deal with. Exploratory Data Analysis was performed to gain insights from the data, feature selection was done to prepare the data for machine learning modelling. Machine learning model was built on the data using Random Forest Classifier, and finally local optimization was done to improve the accuracy of the model built.

Insights Gained from Exploratory Data Analysis:

There are 563 entries in the data with 106 columns of data; this means that there are 563 Match records in our data out of this 563 Match records, the Home team won 221 games, the Away team won 210 games, and 132 games ended in a draw between the home team and the away team.

H 221

A 210

D 132

Name: FTR, dtype: int64

Overall from all the Matches the Home team has more wins from the past two seasons, a team playing at home has higher chance of winning the game.

Bookings: Away teams has more bookings than the Home teams; in the 563 games the Away teams has a total of 889 Yellow cards, and the Home teams has a total of 871 Yellow cards. Away teams suffer 42 Red cards in total, while the Home teams has a total of 27 Red cards.

In 500 Games the referee that gave the most Red cards is C. Kavanagh, with a total of 4 Red cards in four games for the Home team.

S/N	Referees' Name	No of Matches with Red Cards in the Last 500 Games for Home team
1	C Kavanagh	4
2	A Taylor	3
3	R Jones	2
4	D Coote	2
5	C Pawson	2
6	P Tierney	2
7	M Atkinson	1
8	M Dean	1
9	A Marriner	1
10	G Scott	1

Name: Referee, dtype: int64

The table above contains the top ten Referees that gave Red cards in the last 500 games for the Home team.

S/N	Referees' Name	No of Matches with Red Cards in the Last 500 Games for Away team
1	J Moss	3
2	A Taylor	2
3	M Dean	2
4	M Oliver	2
5	K Friend	2
6	P Bankes	2
7	A Madley	2
8	S Attwell	1
9	D Coote	1
10	G Scott	1

Name: Referee, dtype: int64

The table above contains the top ten Referees that gave Red cards in the last 500 games for the Away team.

S/N	Referees' Name	No of Matches with Yellow Cards in the Last 500 Games for Home team
1	P Tierney	4
2	C Kavanagh	2
3	M Oliver	2
4	A Taylor	2
5	D Coote	2
6	K Friend	2
7	J Moss	1
8	S Attwell	1
9	M Dean	1
10	G Scott	1

Name: Referee, dtype: int64

The table above contains the top ten Referees that gave Yellow cards in the last 500 games for the Home team.

S/N	Referees' Name	No of Matches with Yellow Cards in the Last 500 Games for Away team
1	C Kavanagh	4
2	P Bankes	3
3	P Tierney	2
4	M Dean	2
5	A Taylor	2
6	C Pawson	2
7	S Attwell	2
8	A Marriner	1
9	S Hooper	1
10	D Coote	1

Name: Referee, dtype: int64

The table above contains the top ten Referees that gave Yellow cards in the last 500 games for the Away team.

Most Officiating Referee: the most officiating referee in the whole of 563 games is M Olivier, he officiated 43 matches. He is followed by A. Taylor with 42 matches.

S/N	Referees' Name	Games Officiated
1	M Oliver	43
2	A Taylor	42
3	M Atkinson	38
4	M Dean	37
5	J Moss	36
6	A Marriner	35
7	P Tierney	35
8	C Pawson	34
9	S Attwell	33
10	D Coote	32

Name: Referee, dtype: int64

The table above contains information about the ten most officiating referees and the number of games officiated.

Season Performance

2021/2022 Goal Performance: so far into the season the Home team performance has been good with 77 wins so far, the Away team has won 57 games and 49 ended in a draw.

Home Team wins	77
Away Team wins	57
Draw	49

Name: FTR, dtype: int64

The table above represents wins by the home teams, away teams, and draws.

Half Time Games: majority of matches at half time ends in a draw, with most wins coming from the second half of the games. 74 matches ended in draw by half time, 64 matches were in favour of the home teams by half time, while 45 matches ended at half time with the away teams taking the lead.

No of games drawn at half time	74
No of games won at half time by home teams	64
No of games won at half time by away teams	45

Name: HTR, dtype: int64

The table above shows game performance at half time.

Shots: the home team has fired a total of 2569 shots this season, while the away team has fired a total of 2135 shots.

Home Teams Shots 2021/2022	2569
Away Teams Shots 2021/2022	2135

Shots on Target: home teams has a total of 889 shots on target, while the away teams has 756 shots on target.

Home Teams Shots on Target 2021/2022	889
Away Teams Shots on Target 2021/2022	756

Goals: home teams has netted 285 goals so far this season, while away teams has scored 233 goals so far this season

Home Teams Goals2021/2022	285
Away Teams Goals 2021/2022	233

So far into the season the performance of teams playing at home is better than the performance of visiting teams.

Previous Season 2020/2021

Game Performance (Wins):

Wins: the previous season of Premier League saw the away teams winning more games than the home teams. The visiting teams won 153 games, the teams playing at home won 144 games, while 83 games ended in a draw.

Away Teams wins	153
Home Teams wins	144
Draw	83

The table above number of wins of each teams.

Goals: home teams scored more goals than the visiting teams in 2020/2021 season. The home teams has total goals of 514, while the away teams netted the goalpost 510 times. The goal difference is 4.

Home Teams Goals2020/2021	514
Away Teams Goals 2020/2021	510

Performance wise the visiting teams performed better than the home teams previous season.

Model Building with Machine learning.

A machine learning modelling was done on the data, to predict which team is going to win a game, whether the home team, the visiting team or the match will end in a draw. The algorithm used was random forest classifier because of its efficiency in predicting a multi class variable, and its easy optimization manually. The model produced an accuracy score of 82.3%, which means that our model will always be right 82.3% of the time.

The notebooks to this project can be accessed via the links below

<https://colab.research.google.com/drive/1A1RB3LBAH0T38obb1Lz7dD3kCeljMI72?usp=sharing#scrollTo=iNgFTZcAYzhb> for the Machine learning model that was built

https://colab.research.google.com/drive/10H4VlnQ96sAzVND66bOPFVM275_JcR8x?usp=sharing#scrollTo=rAiYHr6c9xSw for the data analysis that was performed on the data.

The data used for this project can be gotten via the link below

football-data.co.uk/data.php